nature communications

Article

https://doi.org/10.1038/s41467-025-61873-0

Limits on the computational expressivity of non-equilibrium biophysical processes

Received: 15 November 2024

Accepted: 3 July 2025

Published online: 05 August 2025

Check for updates

Carlos Floyd[1,2] ✉, Aaron R. Dinner[1,2,3], Arvind Murugan[2,4] & Suriyanarayanan Vaikuntanathan[1,2,3] ✉

Many biological decision-making tasks require classifying high-dimensional chemical states. The biophysical and computational mechanisms that enable classification remain enigmatic. In this work, using Markov jump processes as an abstraction of general biochemical networks, we reveal several unanticipated and universal limitations on the classification ability of generic biophysical processes. These limits arise from a fundamental non-equilibrium thermodynamic constraint that we have derived. Importantly, we show that these limitations can be overcome using common biochemical mechanisms that we term input multiplicity, examples of which include enzymes acting on multiple targets. Analogous to how increasing depth enhances the expressivity and classification ability of neural networks, our work demonstrates how tuning input multiplicity can potentially enable an exponential increase in a biological system's ability to classify and process information.

To survive, cells must understand and respond effectively to their chemical and physical environments. This information-processing task relies on intricate chemical coding systems[1–9]. A simple example of a system that decodes a chemical signal is the Goldbeter-Koshland push-pull circuit[10], which transitions between binary states as the activity of an enzyme varies (Fig. 1A, B). Phase separation in the cell can similarly lead to sharp boundaries in the space of molecular concentrations[11]. Biochemical modules such as the p53 tumor suppression pathway enable cells to classify environmental stresses[12], and recent advances in synthetic biology have made it possible to recapitulate and engineer the classification capabilities of such systems within cells[13,14]. Finally, the so-called glycan code can be viewed as a rich encoding of a high-dimensional cell state into hundreds of different discrete states (classes): enzyme activities in the Golgi apparatus act as inputs by attaching varying amounts of different sugar molecules to proteins which then embed in the plasma membrane and serve as signaling molecules which encode the cell state (Fig. 1C, D)[15–17]. These biochemical systems draw decision boundaries through their input spaces, demarcating them into regions that map to classes. Training of these systems presumably occurs over evolutionary time to yield sets of kinetic rates and chemical conditions that allow them to perform their computational tasks precisely. How these systems, with energetics and kinetics constrained by thermodynamic laws, are able to classify potentially high-dimensional chemical and physical states into one of many discrete choices is not well understood and remains an important open question.

Previous works have studied aspects of computation in physical networks[18–20], specific chemical model systems[6,21–25], and notably in competitively interacting molecular networks at equilibrium[5,6,9,26–28]. Although such studies have illustrated an analogy between neural networks and biochemical networks, it is currently not clear what constraints on the amount of information that can be encoded (i.e., the expressivity) are introduced through the use of molecular activities as representations. In addition, to our knowledge, a general investigation of the classification ability of non-equilibrium biological processes has not been carried out.

In this work, we use tools developed to describe far-from-equilibrium systems to investigate this central question. Our results reveal strong and surprising limits on the ability of non-equilibrium biological systems—modeled as general non-equilibrium Markov state networks—to perform classification tasks. These constraints are derived from a class of non-equilibrium response limitations recently

[1]The Chicago Center for Theoretical Chemistry, The University of Chicago, Chicago, IL, USA. [2]The James Franck Institute, The University of Chicago, Chicago, IL, USA. [3]Department of Chemistry, The University of Chicago, Chicago, IL, USA. [4]Department of Physics, The University of Chicago, Chicago, IL, USA. ✉e-mail: csfloyd@uchicago.edu; svaikunt@uchicago.edu

https://doi.org/10.1038/s41467-025-61873-0

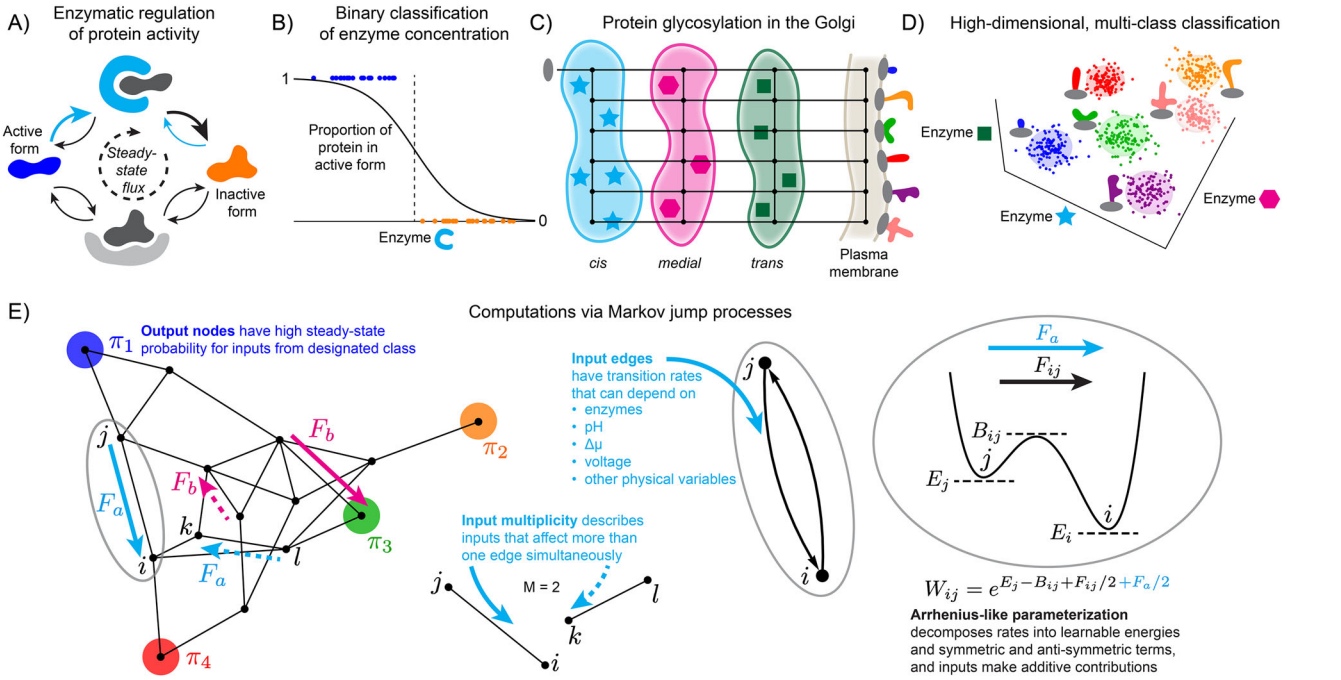

Fig. 1 | Classification tasks performed by biochemical networks. A The push-pull circuit of enzyme activation. The input here is the activity of activating enzyme, shown in cyan, which affects the colored transition rates in the corresponding Markov network. **B** Schematic graph of the binary (active vs. inactive) classification task, which computes a soft threshold on the activity of the activating enzyme. Colored points represent desired outputs, which are approximated by the learned function shown in black. **C** Schematized representation of the process of protein glycosylation in the Golgi apparatus, adapted from the model in ref. 16. Proteins shown as gray ellipses traverse through many cisternae, and the state of the cell dictates the set of glycosyltransferase enzymes found in each cisternae and, in turn, the sugars attached to the proteins. A decorated protein ends up in one of many distinct glycan forms on the plasma membrane, where it serves as an encoding of the cell state. **D** Schematic graph of how protein glycosylation yields many output states, which cluster based on the set of enzymes in the Golgi cisternae. The colors of data points represent the output glycan identities at a given point in enzyme space, and the colored ellipsoids represent decision boundaries that approximately achieve this desired classification. **E** Drawing of a random Markov graph with 15 nodes and 25 edges. The output nodes are labeled, and the input forces (with positive orientation) are drawn labeled with arrows. In classification tasks using this network, the solid arrows are always used as inputs, and the dashed arrows are used when $M = 2$. Input edge driving, input multiplicity, and Arrhenius-like parameterization of the edge rates are illustrated.

reported by some of us in ref. 29. We show how these constraints can be systematically lifted using commonly found biochemical mechanisms such as enzymes acting on multiple targets. Analogous to the way increasing depth and width of artificial neural networks increases their expressivity, tuning input multiplicity may enable an exponential increase in the ability of a biological process to classify and process information. We further show that sharp classification transitions are enabled by input multiplicity along with certain topological conditions, and that the form of computations performed by Markov networks is related to those of transformer architectures. These results offer insights into the mechanisms by which high-dimensional multi-class classification tasks are performed by cells. Our work establishes fundamental design principles underlying biological systems that perform complex computational tasks.

Results

Classification tasks using Markov jump processes

Cells are frequently required to make discrete decisions that require integrating from many different input signals. Examples include decisions made in processes such as chemotaxis, transcription regulation in response to heat shock, quorum sensing, and many others[1,2]. These decisions are made using networks of biochemical components based on complex combinations of input signals from the environment. Can these biochemical networks compute arbitrarily complicated functions of their input signals, or, if not, what ingredients are needed to allow for more complicated decision making?

To address this question, we work with a general mesoscale Markov state characterization of biological processes (Fig. 1E). Nodes in the Markov network are coarse representations of the state of the system. Edges encode rates of transitions between the states and can be functions of, for example, temperature, pH, enzyme activities, and chemical potential gradients. This class of physical models is commonly used to represent kinetic schemes of chemical reaction networks[30–37]. We model inputs to the system as modulating the rates along designated edges of the Markov state network. The output is encoded in the steady-state properties of the network, and we first consider representing the output specifically by the occupancy of a few designated output nodes. Our main results rely on non-equilibrium thermodynamic descriptions of the steady state and its response to perturbations, and we obtain several general limits on how effectively the Markov state networks can classify inputs and how sharp the decision boundaries drawn by this physical system can be[30,38]. We describe this effectiveness with the term expressivity, referring to the notion in machine learning of a model's ability to account for and represent complex features in a dataset[39].

A Markov jump process can be represented by a graph with N_n nodes and N_e edges and a probability vector $\mathbf{p}(t)$ over this set of nodes. The rate of jumping from node j to i is denoted $W_{ij} = e^{E_j - B_{ij} + F_{ij}/2 + F_a/2}$, where E_j, $B_{ij} = B_{ji}$, and the non-equilibrium forces $F_{ij} = -F_{ji}$ are learnable parameters (Fig. 1E). We add an input F_a to the value of F_{ij} if edge ij has been assigned as an input edge. We represent the input variables as a D-dimensional vector $\mathbf{F}$, and we represent the $N_n + 2N_e$ learnable parameters $\{E_j\}_{j=1}^{N_n} \cup \{B_{ij}, F_{ij}\}_{ij \in \mathcal{E}}$, with $\mathcal{E}$ the set of edges, as a vector $\boldsymbol{\theta}$ (see the Methods for physical interpretation of these parameters). Under the master equation dynamics $\dot{\mathbf{p}}(t) = \mathbf{W}(\mathbf{F}; \boldsymbol{\theta})\mathbf{p}(t)$, we view the steady state $\boldsymbol{\pi}(\mathbf{F}; \boldsymbol{\theta}) \equiv \lim_{t\to\infty}\mathbf{p}(t)$ as performing a parameterized

https://doi.org/10.1038/s41467-025-61873-0

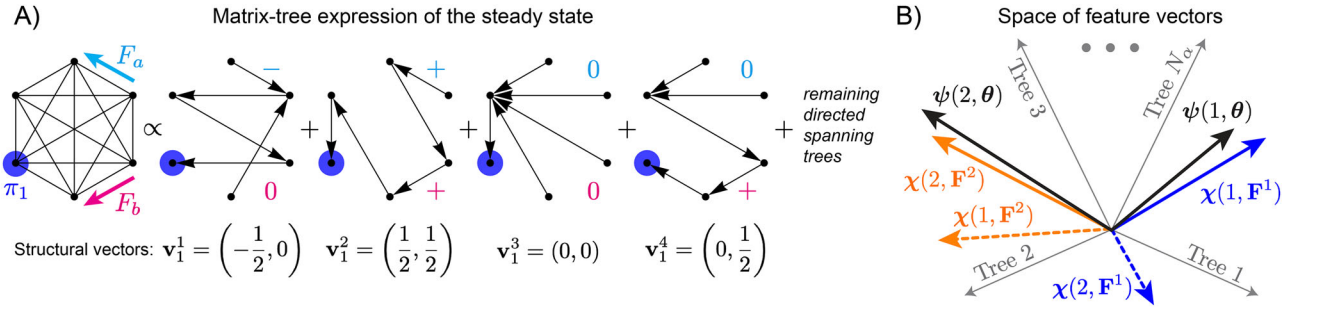

Fig. 2 | The matrix-tree theorem. A Computing the steady-state occupancy π_1 by summing weights over directed spanning trees. Directed spanning trees are subgraphs containing all graph nodes but no cycles, with edges oriented toward a root node. In each directed spanning tree, the input forces make a positive, negative, or zero contribution to the tree weight. The structural vectors $\mathbf{v}_i^\alpha$ are shown below each tree; these quantities enter into Equation (3) below. **B** Schematic illustration of the high-dimensional space of feature vectors $\boldsymbol{\psi}(i;\boldsymbol{\theta})$ and $\boldsymbol{\chi}(i,\mathbf{F})$. The depicted arrangement of vectors could solve a binary classification problem.

computation on the inputs $\mathbf{F}$. Specifically, we typically use a one-hot encoding in which the values π_ρ at designated output nodes should be near 1 when inputs $\mathbf{F}^\rho$ from the corresponding class ρ are presented. We discuss issues of selecting input edges and output nodes in the Methods. We later generalize this setup by optimizing the mutual information between the input and output distributions, without imposing a one-hot encoding scheme.

Computational expressivity from the matrix-tree theorem

Here we review an analytical formula for $\boldsymbol{\pi}(\mathbf{F};\boldsymbol{\theta})$ based on the matrix-tree theorem[38,40] and describe how it can be recast into two equivalent formulations, which we subsequently leverage to highlight its computational expressivity. Specifically, we show how it can be formulated as a rational polynomial with learnable scalar coefficients $\zeta_\mu^i(\boldsymbol{\theta})$ and as a linear attention-like function with non-linear learnable feature vectors $\boldsymbol{\psi}(i;\boldsymbol{\theta})$ and input feature vectors $\boldsymbol{\chi}(i,\mathbf{F})$.

We first restate the well-known matrix-tree theorem expression for the steady-state probability $\boldsymbol{\pi}(\mathbf{F};\boldsymbol{\theta})$[38,40]:

$$\pi_i(\mathbf{F};\boldsymbol{\theta}) = \frac{\sum_{T^\alpha \in \mathcal{T}} w\left(T_i^\alpha, \mathbf{F};\boldsymbol{\theta}\right)}{\sum_k \sum_{T^\alpha \in \mathcal{T}} w\left(T_k^\alpha, \mathbf{F};\boldsymbol{\theta}\right)} \tag{1}$$

Here, $\mathcal{T}$ represents the set of N_α spanning trees, T_i^α represents the α^{th} spanning tree whose edges have been directed to point toward node i as a root so as to connect every other node once to i, and the directed tree weight $w(T_i^\alpha)$ represents the product of all rate matrix elements W_{lm} corresponding to the directed edges $l \leftarrow m$ in T_i^α (Fig. 2A). This formula thus constructs the steady state for node i by summing over all possible kinetic pathways into node i and then normalizing with respect to all nodes.

We define the input multiplicity M as the number of edges affected per input variable, which we assume to be the same for each input. To focus on the functional way in which the input driving enters the steady-state probabilities, the driving contributions can be factored out in the algebraic expressions for the numerator and denominator of Equation (1). This has been previously been used to make analytical progress for $M = D = 1$ in, for example, refs. 29–31. This equivalent formulation of Eq. (1) suggests that steady states of Markov jump processes implement a rational polynomial function of exponentiated input variables. Defining $y_a \equiv e^{F_a/2} > 0$, we rewrite the matrix-tree expression for π_i for general D and M

$$\pi_i(\mathbf{F};\boldsymbol{\theta}) = \frac{\sum_\mu \zeta_\mu^i(\boldsymbol{\theta}) y^\mu(\mathbf{F})}{\sum_\mu \bar{\zeta}_\mu(\boldsymbol{\theta}) y^\mu(\mathbf{F})}. \tag{2}$$

We use the multi-index $\mu = \{\mu_a\}_{a\in\mathcal{A}}$, where $\mathcal{A}$ is the set of D input labels and each component μ_a of the multi-index runs over the values $\{-M, -(M-1), \ldots, M-1, M\}$, to enumerate the $(2M+1)^D$ monomials $y^\mu \equiv \prod_{a\in\mathcal{A}} y_a^{\mu_a}$. These monomials $y^\mu(\mathbf{F})$ in Equation (2) combinatorially depend on the different mixtures μ of input driving, representing a net total μ_a of signed contributions from the input force F_a, μ_b such contributions for F_b, and so on for each input. The coefficients $\zeta_\mu^i(\boldsymbol{\theta})$, which are functions of the parameters $\boldsymbol{\theta}$, are the sums of weights over all directed spanning trees rooted at node i which have the corresponding mixture μ of signed input contributions. The monomial coefficients $\zeta_\mu^i(\boldsymbol{\theta})$ thus represent learnable amplitudes of each polynomial basis function $y^\mu(\mathbf{F})$. The coefficients in the denominator are defined as $\bar{\zeta}_\mu(\boldsymbol{\theta}) \equiv \sum_{k=1}^{N_n} \zeta_\mu^k(\boldsymbol{\theta})$. Classification will be successful if, for $\mathbf{F}^\rho$ drawn from class ρ, the coefficients $\zeta_\mu^\rho(\boldsymbol{\theta})$ and monomials $y^\mu(\mathbf{F}^\rho)$ are large for the same μ. In the subsequent sections of the paper and in the Supplementary Information we use the formulation in Equation (2) to show how the classification ability of a non-equilibrium Markov processes may be systematically modulated.

We show in the Supplementary Information how Equation (1) can alternatively be written as

$$\pi_i(\mathbf{F};\boldsymbol{\theta}) = \frac{\boldsymbol{\psi}(i;\boldsymbol{\theta}) * \boldsymbol{\chi}(i,\mathbf{F})}{\sum_k \boldsymbol{\psi}(k;\boldsymbol{\theta}) * \boldsymbol{\chi}(k,\mathbf{F})}. \tag{3}$$

We interpret $\boldsymbol{\psi}(i;\boldsymbol{\theta})$ as a learnable feature vector with elements $\psi_\alpha(i;\boldsymbol{\theta}) = e^{\mathbf{u}_i^\alpha \cdot \boldsymbol{\theta}} > 0$ corresponding to the trees T_i^α; similarly, $\boldsymbol{\chi}(i,\mathbf{F})$ is an input feature vector with elements $\chi_\alpha(i,\mathbf{F}) = e^{\mathbf{v}_i^\alpha \cdot \mathbf{F}} > 0$. The operation $*$ is a dot product over trees. The structural vectors $\mathbf{u}_i^\alpha \in \mathbb{R}^{N_n + 2N_e}$ encode the topology, i.e., which elements of $\boldsymbol{\theta}$ enter exponentially into the tree weights $w(T_i^\alpha)$ and their signs; $\mathbf{v}_i^\alpha \in \mathbb{R}^D$ records similar information for $\mathbf{F}$. The learnable feature vector $\boldsymbol{\psi}(i;\boldsymbol{\theta})$ is therefore a non-linear encoding of the parameters $\boldsymbol{\theta}$, while the input feature vector $\boldsymbol{\chi}(i,\mathbf{F})$ is a non-linear encoding of the input force $\mathbf{F}$. The goal of training is to adjust $\boldsymbol{\theta}$ so that when $\mathbf{F}^\rho$ is drawn from the class assigned to node ρ, $\boldsymbol{\psi}(\rho;\boldsymbol{\theta})$ has a larger overlap (dot product) with $\boldsymbol{\chi}(\rho,\mathbf{F}^\rho)$ than any other $\boldsymbol{\psi}(\rho';\boldsymbol{\theta})$ has with $\boldsymbol{\chi}(\rho',\mathbf{F}^\rho)$ for $\rho' \neq \rho$ (Fig. 2B). As we show below, this functional form illustrates the role of non-equilibrium affinity in enabling classification. It also sheds light on the potential learning modalities accessible with Markov state networks.

A fundamental limit on classification expressivity from non-equilibrium thermodynamics

As a first illustration of how physical constraints can limit expressivity of such systems, we train the network shown in Fig. 1E to perform a series of binary classification tasks (Fig. 3A–C). In each case, we assign node 1 to blue points and node 2 to orange points, and we train the network, as described in the Methods section, to obtain a set of learned parameters $\boldsymbol{\theta}^*$. We indicate the results by drawing contours at $\pi_1(F_a, F_b; \boldsymbol{\theta}^*) = 1/2$ and $\pi_2(F_a, F_b; \boldsymbol{\theta}^*) = 1/2$. In the Supplementary

https://doi.org/10.1038/s41467-025-61873-0

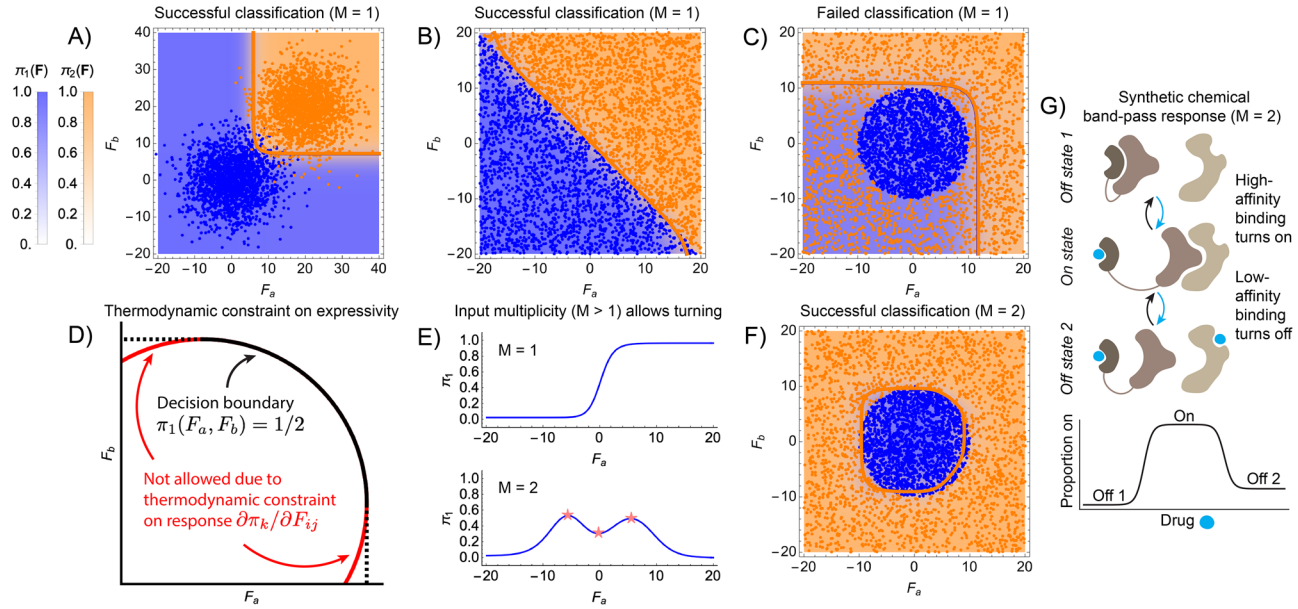

Fig. 3 | Overcoming inflexible decision boundaries by increasing the input multiplicity hyperparameter *M*. **A** Plot of the learned classification functions $\pi_1(\mathbf{F})$ and $\pi_2(\mathbf{F})$ shown as colored density plots over the input force space. On top of this, scatter plots show the dataset, colored by assigned class, which was used to train the network. Solid lines show the contour $\pi_1(\mathbf{F}) = 1/2$ in blue and $\pi_2(\mathbf{F}) = 1/2$ in orange; note that these are approximately overlapping. The network shown in Fig. 1E is used for all classification tasks in this figure. **B**, **C** Same as **A**, but for different classification tasks. **D** Schematic illustration of the monotonicity constraint. **E** Plots illustrating that increasing M from 1 to 2 allows for non-monotonic dependence of a steady-state occupation on an input driving force. **F** Same as panel C, but for the network in Fig. 1E, which also includes driving along the dashed arrows ($M = 2$). **G** Schematic illustration of a recently designed synthetic chemical band-pass system using multiple input binding[42]. A drug binds through a high-affinity pathway to activate a protein and through a second, low-affinity pathway to deactivate the protein, leading to a non-monotonic dependence of activation on the drug.

Information we show examples of the learned parameters in trained networks. The network successfully classifies the points in Fig. 3A, B but not 3C. The failure in Fig. 3C is not a limitation of the training protocol, and it does not improve as N_n is increased. Rather, it emerges from a fundamental constraint on the response of non-equilibrium steady states as the forces F_a and F_b are tuned.

Specifically, for any choice of i, j, and k, and with other parameters held fixed, the derivative $\partial\pi_k/\partial F_{ij}$ has a fixed sign across the entire range of F_{ij}[29]; in other words, $\pi_k(F_{ij})$ is a strictly monotonic function. Thus, for fixed F_b, $\pi_1(F_a, F_b; \boldsymbol{\theta}^*)$ must be a monotonic function of F_a which implies that it can take the value 1/2 at most once along any line drawn parallel to $F_b = 0$ (Fig. 3D). By symmetry, the function $\pi_1(F_a, F_b; \boldsymbol{\theta}^*)$ must also be a single-valued function of F_b along any line parallel to $F_a = 0$. We refer to this limitation on the flexibility of the decision boundary as the monotonicity constraint, which implies that the learnable decision boundaries are not invariant to a rotation of the input space. This corresponds to a specific failure mode of computations by non-equilibrium biophysical systems modeled as Markov jump processes.

Improving expressivity by increasing input multiplicity

Biologically, F_a can be interpreted, for example, as depending on the chemostatted activity of an enzyme (see the Methods). In biochemical kinetics, it is common for some species to be involved in multiple reactions simultaneously, making it plausible for F_a to drive multiple edges[31]. We find that allowing for input multiplicity improves classification expressivity, and one way this happens is by lifting the monotonicity constraint. We assume for simplicity that each of the D input variables $\{F_a\}_{a\in\mathcal{A}}$, where $\mathcal{A}$ is the set of input labels, affects the same number M of edges. Setting $M > 1$ lifts the monotonicity constraint because the condition for $\pi_k(F_{ij})$ to be a monotonic function is that all other edge parameters are held fixed; with $M > 1$ this is no longer true since several edge parameters change simultaneously as an input is varied.

To better understand the gain in the decision boundary's flexibility allowed by setting $M > 1$, in the Supplementary Information we analyze the steady-state representation in the rational polynomial form of the matrix-tree expression, Eq. (2). Considering the case $D = 1$ and identifying turning points as roots of $\partial\pi_i/\partial F_a$, we show that the maximum number R of such roots obeys

$$R = \begin{cases} 0 & M = 1 \\ 2M - 1 & M > 1, \end{cases} \tag{4}$$

which is a direct measure of the classifier's expressivity; see Fig. 3E for an illustration and the Supplementary Information for a numerical verification up to $M = 4$. A proof of the scaling $2M - 1$ for rational polynomials with non-negative coefficients can be found in ref. 41. Thus, once $M > 1$, π_i is no longer subject to the monotonicity constraint and behaves like a non-negative rational polynomial of degree up to $2M$. Input multiplicity thus allows the non-equilibrium biological process to be more expressive and draw out decision boundaries that can classify more complex data structures. Indeed, returning to the previously failed classification with $M = 1$ (Fig. 3C), we see that setting $M = 2$ allows the same network to now learn a decision boundary which successfully encloses the data assigned to class 1 (Fig. 3F). This implies that classifying a finite band of input signal levels (like a band-pass filter) requires setting $M > 1$ along the corresponding input dimension. A recent development in synthetic biology has in fact shown in a specific example that drug binding to receptor molecules via two distinct binding pathways can be used to design band-pass-like responses to the drug (Fig. 3G)[42].

To quantify the binary classification ability for arbitrary M and D, we consider a classic measure called the Vapnik-Chervonenkis (VC) dimension[43]. This represents the largest number N_{VC} of points which, for at least one fixed configuration of the points in the input space, a set of classifiers can correctly classify for any of the $2^{N_{VC}}$ assignments of binary labels to the points. A theorem by Dudley[44,45] states that if a classifier

$h(\mathbf{F})$ (whose sign determines the predicted binary label) belongs to a vector space $\mathcal{H}$ of real scalar-valued functions, then the VC dimension of the set of all classifiers in $\mathcal{H}$ is equal to the dimension of $\mathcal{H}$. Given the representation of the contour $\pi_i(\mathbf{F};\boldsymbol{\theta}) = 1/2$ in the rational polynomial form of the matrix-tree expression, Eq. (2), we see that its vector space is spanned by the $(2M+1)^D$ coefficients $\zeta^i_\mu(\boldsymbol{\theta}) - \bar{\zeta}_\mu(\boldsymbol{\theta})/2$. We thus estimate the VC dimension of this classifier as

$$N_{VC} \le (2M+1)^D. \tag{5}$$

This should be viewed as an upper bound in two senses. First, for $M = 1$, the monotonicity constraint imposes that N_{VC} is strictly less than $(2M+1)^D = 3^D$. Second, even for $M > 1$ the $N_n(2M+1)^D$ coefficients $\zeta^i_\mu(\boldsymbol{\theta})$ are not all independent degrees of freedom, as we illustrate in the next section.

These findings suggest that input multiplicity M significantly increases the complexity (measured by VC dimension) of the classification tasks a biochemical circuit can perform, scaling roughly as $\sim M^D$. Input multiplicity is a known feature of many biochemical networks: for example, many transcription factors[3] as well as glycosyltransferases in the Golgi apparatus[15–17] are known to act on several targets. Other input variables, such as temperature, voltage, or chemical potential gradients, may also affect multiple edge rates simultaneously. Measured binding affinities between glycan molecules and their receptors (called lectins) reveal that these interactions are highly non-specific, corresponding to a high input multiplicity[46]. Additionally, theoretical work on gene regulation has shown in specific examples that an equivalent notion of input multiplicity can allow for increased channel capacity of the regulatory motif[47]. Our work thus provides a potentially unifying description of how input multiplicity could enable biological processes to perform more expressive computations.

Storing more classes by increasing input multiplicity

We now generalize the binary classification task and ask how many different classes can be stored as a function of the hyperparameters M and D. Classifying many different classes is crucial in biology. For example, deciphering the glycan code, which specifies one of several hundred different cell states, or recognizing previously encountered antigens during an immune response both require choosing among large numbers of possibilities[15,48,49]. How these biochemical systems achieve these complex classification tasks (e.g., through microscopic sensing events like estimating antigen binding affinities) remains an important and open question.

For $M = 1$, a simple geometric argument suggests that up to 2^D classes could in principle be separated. Each class could be placed in one of the 2^D orthants of the input space, and a classifier with monotonicity along each axis could in principle assign a unique response to each. In Fig. 4A we attempt to distinguish four classes by placing them in the four quadrants of the (F_a, F_b) plane, but we find, contrary to this expectation, that no network can separate them all with $M = 1$.

This failure stems from the constrained functional form of the steady-state distribution in our Markov network model. Each component $\zeta^i_\mu(\boldsymbol{\theta})$ entering the matrix-tree solution (Eq. (2)) is a polynomial function of the $2N_e$ edge rates $\{W_{ij}\}_{ij\in\mathcal{E}}$, and there exist equality constraints among these functions (Fig. 4B)[29]. These constraints reduce the number of independently adjustable degrees of freedom, limiting the network's capacity to implement many decision boundaries.

In the Supplementary Information we count constraints to show that the maximum number of degrees of freedom $n_{\text{d.o.f.}}$ among the $3N_n$ functions $\{\zeta^i_\mu\}_{i=1}^{N_n}$ for $M = 1$, $D = 1$ is

$$n_{\text{d.o.f.}} = 2N_n. \tag{6}$$

We also show that the maximum number of degrees of freedom among the $9N_n$ functions $\{\zeta^i_\mu(\boldsymbol{\theta})\}_{i=1}^{N_n}$ for $M = 1$, $D = 2$ is

$$n_{\text{d.o.f.}} = \min(2N_e, 3N_n), \tag{7}$$

where the minimum reflects the fact that the number of tunable parameters cannot exceed the number of edge rates.

To classify each input region correctly, the output probabilities must satisfy a number of inequality conditions. For example, the probability $\pi_2(\mathbf{F})$ in the top right quadrant of Fig. 4A must be greater than the probability at all other nodes. For $D = 1$, the number of such conditions scales as $2N_n$, matching the number of available degrees of freedom. But for $D = 2$, the number of conditions grows as $4N_n$, while the number of degrees of freedom caps at $3N_n$. Thus, four-class classification appears to become infeasible for $M = 1$, $D = 2$, a result we verify through detailed analysis in the Supplementary Information. This reveals a new failure mode: equality constraints among classifier coefficients can limit expressivity even when enough parameters appear to be present.

As with binary classification, increasing M improves performance. For $M = 2$, the number of ζ^i_μ functions increases enough that even after accounting for equality constraints, the network typically attains the full $2N_e$ degrees of freedom. This suffices to separate four classes in

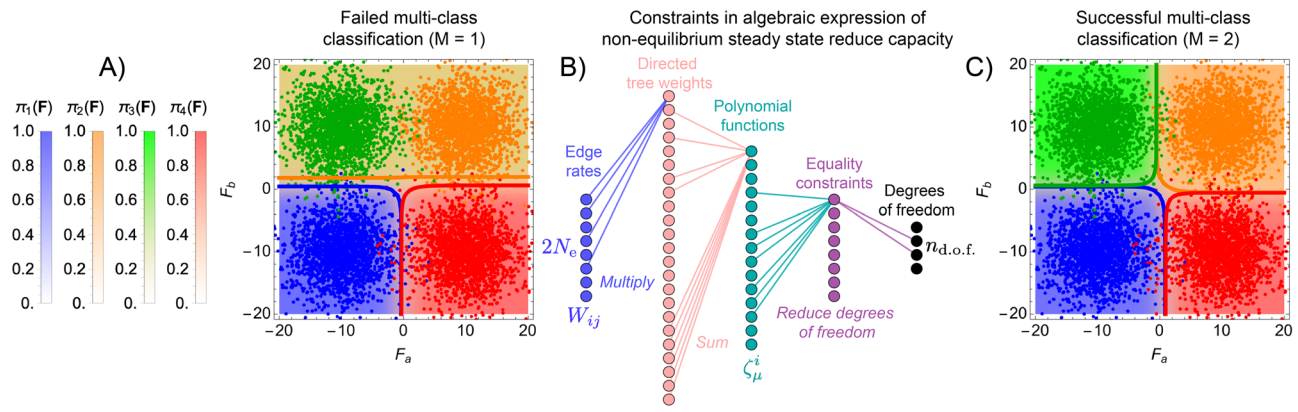

Fig. 4 | Input multiplicity allows overcoming reduced degrees of freedom and increases multi-class capacity. A Plot of the learned classification functions $\pi_1(\mathbf{F})$, $\pi_2(\mathbf{F})$, $\pi_3(\mathbf{F})$, and $\pi_4(\mathbf{F})$ for the network shown in Fig. 1E, with only the solid arrows used for inputs ($M = 1$). **B** Schematic illustration of how the learnable parameters W_{ij} (the edge rates on the left) are first multiplied within directed spanning trees into products called directed tree weights, which are then summed together to yield the polynomial functions $\zeta^i_\mu(\boldsymbol{\theta})$ appearing in Equation (2). Although each function $\zeta^i_\mu(\boldsymbol{\theta})$ is uniquely defined, there exist equality constraints owing to the physics of Markov networks which reduce the effective number of degrees of freedom below the number needed to solve the four-class classification task in **A**. **C** By including driving along the dashed arrows in Fig. 1E (setting $M = 2$), there are sufficiently many degrees of freedom to solve the four-class classification task.

two dimensions (Fig. 4C). These results suggest that input multiplicity may be a key component for the remarkable feats of multi-class classification used in biological processes like adaptive immunity or deciphering the glycan code.

Expressivity requires non-equilibrium driving

A hallmark of biophysical processes is that they are sustained far from thermodynamic equilibrium through continual consumption of chemical free energy. We now explain how this feature is a necessary ingredient for some of the aforementioned computational abilities. In the absence of any non-equilibrium driving, either through the learned parameters F_{ij} or the input variables F_a, the steady-state distribution is a Boltzmann form $\pi_i \propto e^{-E_i}$ and does not depend on the B_{ij} parameters. Beating this restrictive functional form and achieving non-trivial classification expressivity thus requires non-equilibrium driving. In the Supplementary Information, we use the linear attention-like form of the matrix-tree expression, Eq. (3), to show how the non-equilibrium parameters F_{ij} allow for the greatest flexibility in positioning the learnable feature vectors $\boldsymbol{\psi}(i;\ \boldsymbol{\theta})$, thereby enabling expressive computations.

To demonstrate this numerically, we measure how classification accuracy depends on the amount of allowed non-equilibrium driving. To do this, we consider an input modality in which input variables B_a present additive contributions to the B_{ij} parameters along input edges rather than the F_{ij} parameters along those edges. In this way, the only non-equilibrium driving in the system comes from the learned F_{ij} parameters, which we then constrain in magnitude. We train the network in Fig. 1E for the classification task shown in Fig. 3A for several values of $F_{\max}$, which we impose during training as a ceiling on the absolute value of any learned F_{ij} parameter. In Fig. 5 we plot the classification accuracy of the trained networks, showing a continuous increase in performance as a function of $F_{\max}$. This implies that under the linear dynamics of Markov jump processes, it is necessary to break detailed balance to perform non-trivial computations. We provide further theoretical support for this claim by analyzing the learnable vectors $\boldsymbol{\psi}(i;\boldsymbol{\theta})$ in the Supplementary Information.

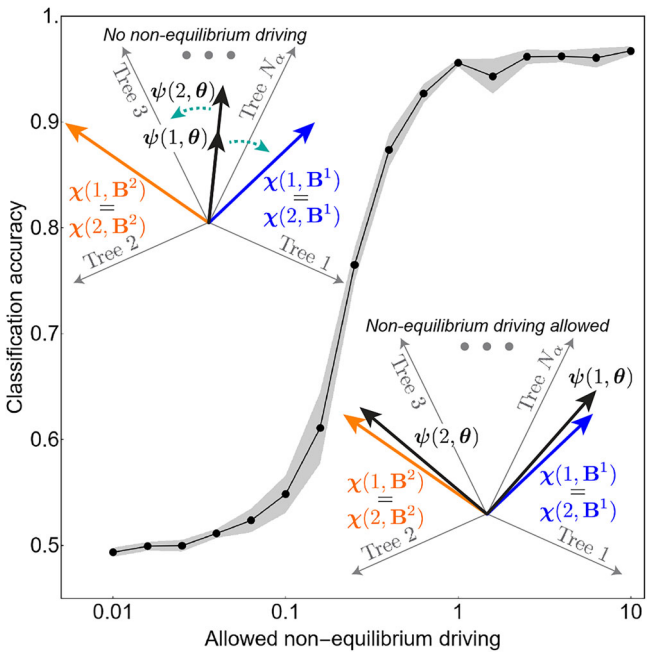

Fig. 5 | Accurate classification requires non-equilibrium driving. This plot corresponds to the classification task in Fig. 3A using B_a and B_b as inputs instead of F_a and F_b. The classification accuracy, defined as the average of $\pi_\rho(\mathbf{F}^\rho;\boldsymbol{\theta})$ over 10^3 randomly drawn samples of $\mathbf{F}^\rho$ from classes $\rho = 1, 2$, is shown as a function of $F_{\max}$, the maximum absolute value of F_{ij} that is allowed on any edge. Five training trials for each value of $F_{\max}$ were performed, and the gray area illustrates the standard deviation of accuracy over these trials. Insets schematically illustrate the feature vectors in each regime.

An additional perspective on non-equilibrium Markov networks trained for classification can be gained from recent work exploring the analogies between transformers, which implement softmax-based filters over key, query, and value vectors, and models of dense Hopfield networks, in which relaxational dynamics in the landscape of a softmax-like energy function allows for storage of a far greater number of associative memories compared with the usual quadratic Hopfield energy functions[50,51]. The linear attention-like form of the matrix-tree expression, Eq. (3), motivates consideration of the corresponding energy-like function that describes Markov steady states, which we identify in analogy with ref. 51 as $\mathcal{F} \equiv -\ln \sum_k \boldsymbol{\psi}(k;\boldsymbol{\theta})^* \boldsymbol{\chi}(k;\mathbf{F})$, where $\boldsymbol{\psi}(k;\boldsymbol{\theta})$ and $\boldsymbol{\chi}(k;\mathbf{F})$ are the non-linear feature vectors in the attention function. Plotting this function over the input space (F_a, F_b) in trained and untrained graphs reveals a landscape characterized by flat, sloping basins delimited by creases of high curvature. In trained graphs some of these creases co-localize with the learned decision boundaries, separating the input space into regions in which different subsets of trees dominate the contribution to $\nabla_{\mathbf{F}}\mathcal{F}$. Some of these creases also represent topological features of the graph which cannot be removed by training, such as mismatched limits at the corners of the input domain (e.g., $F_a \to \infty$ followed by $F_b \to \infty$ or vice versa). In the Supplementary Information we expand on this analogy between the steady states of Markov processes and dense associative memories, but we leave a full treatment of this connection to future work. We note in addition that the form of the classification function (Eqs. (2) and (3)) might support comparisons to other existing machine learning architectures, such as kernel-based classifiers[39], besides transformers.

In the Supplementary Information, we show that the trees contributing to the steady state in networks trained for classification are more localized in weight space than in untrained ones. The activities of the spanning trees additionally cluster according to the input class, a pattern partially driven by the input itself and present even in untrained graphs. Training appears to tighten clustering of activity in graphs with many spanning trees but loosens it in graphs with fewer, suggesting that evolved reaction networks may alter the spread of reactive flow over the set of pathways, subject to constraints from topology. The generality of this behavior and its dependence on task structure remain open questions. We note that similar changes in collective response dimensionality have been observed in trained elastic networks[52] and in associative memory models undergoing a feature-to-prototype transition[53,54].

Higher input multiplicity enhances information processing: the emergence of one-hot encoding

We have so far considered one-hot encodings of the classification output, where certain nodes are preselected to exhibit high steady-state probability when an input from the corresponding class is presented. This encoding is standard in neural network-based classifier architectures, but there is no a priori reason for a biophysical system to adopt this convention. A more general approach would encode the output across the entire steady-state probability distribution $\boldsymbol{\pi}$ in a manner that contained the most information about the distribution of input driving forces. In this case, it is natural to ask whether using $M > 1$ can still improve information processing capacity.

To address this question, we reformulate the training setup to maximize the mutual information $I(\mathcal{I};\mathcal{F})$ between the random variables $\mathcal{I}$, representing the steady-state occupancies at the N_n node indices, and $\mathcal{F}$, the input driving force. Given a fixed graph (Fig. 6A), an input edge, and a probability distribution $p_{\mathcal{F}}(F_a)$ for a single input force, we optimize the mutual information using conjugate gradient ascent with respect to the edge rates W_{ij}. In the Supplementary Information we provide explicit formulas for this calculation.

https://doi.org/10.1038/s41467-025-61873-0

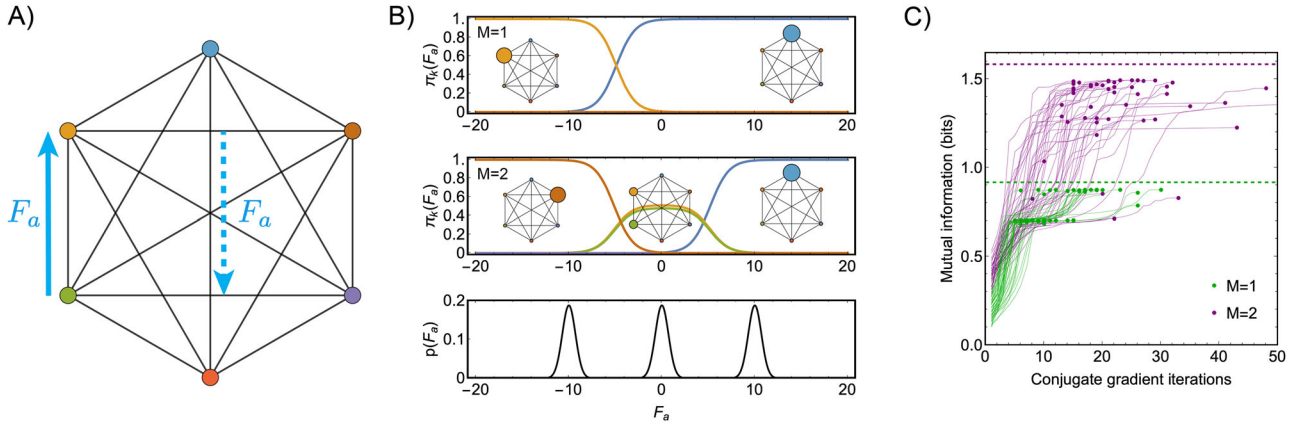

Fig. 6 | Optimizing mutual information recovers one-hot encodings and improves with greater input multiplicity. A Drawing of a fully connected graph with $N_n = 6$, with one ($D = 1$) input force F_a applied along the solid arrow for $M = 1$ and along both arrows for $M = 2$. **B** *Top*: For $M = 1$, plots of the components of the optimized $\boldsymbol{\pi}(F_a)$ colored as in the graph drawing in **A**. Insets schematically show the steady state of the graph for at the corresponding value of F_a. Note that several components are close to zero for all value of F_a. *Middle*: Same as top, but with $M = 2$. *Bottom*: Plot of the input distribution $p(F_a)$, composed of three Gaussian peaks, used in this example. **C** Trajectories of the mutual information between the input distribution $p(F_a)$ and the output distribution $\boldsymbol{\pi}(F_a)$ as it is optimized via updates to the W_{ij} parameters using the Fletcher-Reeves conjugate gradient algorithm. Random initial conditions in the range [0, 1] for the W_{ij} parameters are used for each trajectory. The dashed lines indicate theoretical upper bounds for the entropies of the discrete distributions {1/3, 2/3} (green) and {1/3, 1/3, 1/3} (purple). The final parameters of the trajectories which best optimized the mutual information for each value of M are used for the plots in panel B.

Using a distribution $p_{\mathcal{F}}(F_a)$ consisting of three Gaussian peaks, we find that for $M = 1$, the network cannot fully distinguish among the three peaks (top row of Fig. 6B), achieving a mutual information value of $\mathcal{I} \approx 0.9$ bits. This is roughly the entropy of the discrete distribution {1/3, 2/3} corresponding to the grouped sets of Gaussian peaks. However, for $M = 2$, the trained network exploits the allowed non-monotonicity of the steady-state response to distinguish all three peaks (middle row of Fig. 6B), reaching $\mathcal{I} \approx 1.5$ bits, which is roughly the 1.58 bits required to differentiate among three equally probable options. Although the conjugate gradient optimization reaches different stopping points for the different random initial conditions, the trend that higher mutual information is achievable with $M = 2$ is clear (Fig. 6C). Related works on noisy gene regulatory network elements also show through specific examples that higher information processing capacity is available when the effective input multiplicity is increased[47,55].

Interestingly, networks trained to optimize mutual information emergently learn encodings resembling one-hot encodings, in which separate nodes are assigned to read out each of the Gaussian peaks in $p_{\mathcal{F}}(F_a)$. An exception occurs in the plotted result for for $M = 2$ in Fig. 6B, where the network assigns two nodes to read out the peak near $F_a = 0$. In the Supplementary Information, we provide theoretical arguments supporting these findings. Specifically, for peaked input distributions, if each node reads out a single peak then mutual information is not lost even if multiple nodes correspond to the same peak, since observing any of these nodes uniquely identifies the peak. However, if one node reads out multiple peaks, then the corresponding peak cannot be uniquely determined when the node is observed.

Network topology dictates the sharpness of the decision boundary

The flexibility of the decision boundary may be distinguished from another important feature relevant for biological discrimination, which is the boundary's sharpness. We can quantify this as the norm of the gradient $\nabla_{\mathbf{F}}\pi_i(\mathbf{F}; \boldsymbol{\theta})$ evaluated at a location separating the discrimination regimes. Sharpness and the related measure of steady-state amplification are topics which have received much research focus, particularly in models of cooperative binding, cellular sensing, ultrasensitive covalent modifications, and kinetic proofreading[10,30–33,36,37,56–59]. A consensus among these works is that greater sharpness requires greater expenditure of chemical free energy; this idea is often expressed in the form of inequalities reminiscent of the thermodynamic uncertainty relations[60–62]. Here, we extend this line of research by explicitly framing it within the context of a computational classification problem. We demonstrate systematic methods to sharpen the decision boundary by increasing the number of input-driven transitions that occur serially along a reaction pathway (multiple forms of an intermediate molecule), rather than through parallel reaction pathways (multiple different intermediate molecules). These serially driven transitions tend to yield directed spanning trees with greater net input driving. Serially driven edges occur biologically, for instance, in common models of how ligands drive cooperative binding reactions and how ATP or GTP drives multi-step kinetic proofreading[3,30]. Parallely driven edges occur in some types of general enzymatic schemes[30,58,63], and both kinds of driven edges occur in common models of the flagellar motor[31]. Additionally, we show that it is important to consider not only the extremes of the trees' net input driving but also the multiplicity of trees with intermediate net input driving, as a large number of such intermediate trees can reduce sharpness.

For simplicity, we study sharpness with $D = 1$ but note that the argument extends straightforwardly to $D > 1$, because to compute the gradient norm $\nabla_{\mathbf{F}}\pi_i(\mathbf{F}; \boldsymbol{\theta})$ we merely sum the one-dimensional terms $(\partial\pi_i/\partial F_a)^2$ for $a \in \mathcal{A}$. We first consider a classic biological motif, the Goldbeter-Koshland push-pull network (Fig. 7A), in which a substrate is shuffled between a non-phosphorylated (S) and phosphorylated (S^*) form by competing kinase (E^A) and phosphatase (E^B) enzymes. Our input is the chemostatted activity of kinase, which we assume modulates the transition rates $E^AS \leftarrow S$ and $E^AS \leftarrow S^*$ by the same affinity F. We suppress the subscript on F in this one-dimensional problem.

We train this network to classify inputs $F \in (-5, 0)$ with high π_S probability and inputs with $F \in (0, 5)$ with high π_{S^*} probability, with results shown in Fig. 7B. The learned $\pi_S(F; \boldsymbol{\theta})$ curve has the right qualitative features but is not very sharp. To sharpen this transition, we consider systematically adding driven edges (increasing M) in one of two ways, either in parallel (Fig. 7B) or in serial (Fig. 7C) with the original driven edges. Training each of these networks with increasing numbers n of additional pairs of driven edges (also adding undriven edges on the bottom for symmetry), we see that adding edges in parallel fails to sharpen the transition, while adding edges in serial succeeds.

https://doi.org/10.1038/s41467-025-61873-0

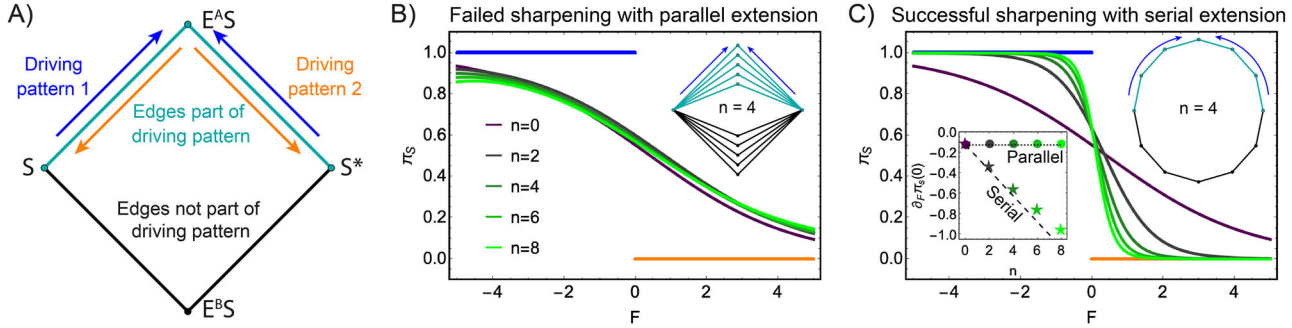

Fig. 7 | Sharpening decision boundaries for a biochemical motif. A Labeled illustration of the four-species push-pull network, with driving patterns drawn as blue and orange arrows. Plots of the trained $\pi_S(F;\boldsymbol{\theta})$ curves for increasing nodes in the parallel (**B**) and serial (**C**) extension. The inset shows the slopes $\partial_F\pi_S(F=0)$ for the serial and parallel extensions as a function of n. The dashed and dotted lines show the bound obtained from $M_R^{\mathcal{O}}$ for the enumerated trees in serial and parallel extensions.

To explain this difference, in the Supplementary Information we use the 1D rational polynomial form of the matrix-tree expression, Equation (3), to maximize $\partial\pi_S/\partial F$ with respect to the learnable coefficients ζ_m^S and $\bar{\zeta}_m$ (treated for now as free and independent). The multi-index μ used in Equation (2) simplifies here to the single index m. We show that

$$\max_{\{\zeta_m^S\},\{\bar{\zeta}_m\}}\left|\frac{\partial\pi_S(F_0;\boldsymbol{\theta})}{\partial F}\right| = \frac{M_R}{8} \tag{8}$$

where the derivative is evaluated at the location of the decision boundary F_0 and $M_R = M_{\max} - M_{\min}$ is the range in exponential powers of $e^{F/2}$ among all directed spanning trees. This result shows that the sharpness of the classifier is fundamentally limited by the structure of the network. A tighter approximation to the bound can be obtained by replacing M_R with $M_R^{\mathcal{O}} \le M_R$, which is the range in exponential powers among only the directed spanning trees rooted on the output nodes. We further explain in the Supplementary Information that the directed spanning trees of the parallelly extended push-pull networks prevent $M_R^{\mathcal{O}}$ from scaling with n, whereas the spanning trees for serially extended networks allow $M_R^{\mathcal{O}} \sim n$, which enables increasingly sharp transitions as more edges are added. In serially extended networks, the structure allows this range $M_R^{\mathcal{O}}$ to grow with the number of added edges n, leading to increasingly sharp transitions. In contrast, parallelly extended networks constrain all output-rooted spanning trees to use the same number of driven edges, keeping $M_R^{\mathcal{O}}$ fixed and preventing sharper transitions.

Finally, even when the bound in Eq. (8) is large it may not be achieved in practice (see Supplementary Information for details). Saturating the bound requires that the coefficients ζ_m^S be concentrated on trees with either the smallest or largest possible net input drive. However, in networks such as those with a ladder-like architecture, many spanning trees make intermediate contributions, and equality constraints among the functions ζ_m^S prevent the network from assigning large weights solely to the extremal trees. As shown in Fig. 4B, overlapping spanning trees entangle the coefficients and reduce the effective degrees of freedom. This structural limitation suggests that sharp decision boundaries may be inherently inaccessible in densely interconnected biochemical networks. This finding resonates with, though is technically distinct from, recent results in refs. 30,31,33,36,37.

Discussion

We have explored the computational expressivity of classifiers implemented in trained non-equilibrium biochemical networks, which we model as Markov jump processes. An analytical solution for the steady states of these systems can be written in several equivalent ways, highlighting complementary interpretations of the classifier as computing a linear softmax operation using learnable, as computing a rational polynomial function with learnable scalar coefficients (Eq. (2)), and nonlinear feature vectors (Eq. (3)). The feature vectors and coefficients are themselves complicated functions of the tree weights of the physical network, and because of this dependency they are significantly constrained relative to abstract parametric classifiers having the same functional form as the matrix-tree expression. We identified several limitations to expressivity, including monotonic responses $\pi_k(F_{ij})$ and a reduction in degrees of freedom of the classifier function. We further showed that increasing input multiplicity (setting $M > 1$) helps mitigate these limitations, by creating additional turning points of $\pi_k(F_{ij})$ and allowing the number of degrees of freedom in the graph to saturate at $2N_e$. With even modest input multiplicity, chemical reaction-based classifiers prove to be capable of solving difficult classification tasks, demonstrating non-linear information processing performance reminiscent of neural networks[1,2].

Key biological implications follow from the sensitive dependence of computational expressivity on the input multiplicity hyperparameter M, which we define as the number of edges driven by a single input variable. Biologically, $M > 1$ occurs when a single input variable, such as activity of an enzyme, temperature, or chemical potential gradient, simultaneously affects more than one chemical transition. Input multiplicity in a biochemical network may at first glance seem counterproductive because it decreases the network's modularity[3], but our results show that it serves to significantly expand a network's computational capabilities. In the context of cooperative binding, there is also a relationship between M and the Hill coefficient, which determines the sharpness of switch-like input responses[31]. We hope to connect our general findings to specific biochemical systems in the future. For example, systems like the glycan code (Fig. 1E) are known to involve promiscuous enzymes which attach sugar molecules to proteins[15,17,64] as well as high levels of cross-talk in the receptor-ligand interactions mediating cellular communication[46]; our results suggest that these forms of high input multiplicity may play a crucial role in enabling efficient molecular information processing[65,66].

Generalizing the results of this paper beyond the particular chemical dynamics and definition of classification that we have adopted is an important avenue for future work. Although the (pseudo) first-order reaction networks modeled by Markov jump processes have less rich dynamics than non-linear chemical kinetics, there are still many biochemical systems to which the matrix-tree theorem, which underlies our results, can be applied[30,31]. For example, approximations based on timescale separations can in some cases be used to create an effective linear system out of non-linear kinetic schemes[63]. We leave to future work general analyses based on chemical reaction network theory, which is feasible in the future using recent theoretical

developments[4,67–69]. In the Supplementary Information, however, we preliminarily extend the matrix-tree theorem approach to non-linear reaction schemes to show that expressivity increases as bimolecular reactions are included. This agrees with our expectation that increasing the reaction order effectively increases the input multiplicity M by allowing for additional ways for the input to affect the reaction rates. Such non-linear systems were recently used to, for example, approximate arbitrary dynamics in a recurrent neural network-like construction[70] and to perform classification and regression tasks through competitive binding interactions[6,8,9] and in a reservoir computing-like setup[71]. Future work should also better characterize the topological and structural features of reaction networks which enable expressive computations, possibly by identifying minimal motifs which can perform computational sub-routines.

Additionally, we adopted here the common convention used in machine learning of one-hot encoding to specify classification outcomes, but biologically, it may be more realistic to specify whole profiles of chemical concentrations as computational outputs. In previous work[29] we showed analytically that the ratio $(\partial\pi_k/\partial F_{ij})/(\partial\pi_{k'}/\partial F_{ij})$ is independent of F_{ij} for any k and k', which can be shown to imply that the monotonicity constraint holds under any linear mapping $\tilde{\boldsymbol{\pi}} = \mathbf{R}\boldsymbol{\pi}$. Thus, the expressivity limitations identified in this work should at least hold for output profiles that are arbitrary linear transformations of one-hot encoded outputs. By considering the encoding-agnostic objective of maximizing the mutual information between the output distribution $\boldsymbol{\pi}$ and the input distribution, we have provided preliminary evidence that the expressive capacity of the system increases with M, independent of specific encoding schemes.

Finally, an additional aspect of physics that deserves attention in the future is that chemical dynamics are inherently stochastic, and fluctuations about the steady-state mean are important, especially when copy numbers are small. Decision-making under fluctuations has often been treated using the framework of information theory[59,72–74]. A general trend from this line of research is that maximizing information flow requires reducing fluctuations, which in turn requires greater expenditure of chemical free energy. On the other hand, some forms of chemical computation harness stochasticity to generatively model probability distributions[26]. Combining insights from these works with our results on classification expressivity can help paint a full picture of how biochemical systems use fuzzy logic to make decisions.

Methods

Network inputs

To present an input to the Markov network, we adjust the parameters along designated input edges (Fig. 1E). In this paper, we mostly present inputs via additive contributions F_a to the F_{ij} parameters along the edges assigned to input a, although we also consider presenting additive contributions B_a to the corresponding B_{ij} parameters. What the edge inputs F_{ij} and B_{ij} represent physically depends on the specific model system that one has in mind, but we next elaborate on their general physical properties.

A contribution to the anti-symmetric term F_{ij} generally exists due to a broken time-reversal symmetry[75], or under coarse-graining, whereby the degrees of freedom of at least two baths with a potential gradient between them are not explicitly modeled in the system dynamics[76]. A pertinent example is the chemical potential difference $\Delta\mu$ between ATP and its hydrolysis products. The assumption that the concentrations of these species are chemostat away from equilibrium implies that their concentrations do not enter as model variables, and they instead break detailed balance by appearing as a contribution to the parameter $F_{ij} \sim \Delta\mu$ along the coarse-grained transition $i \leftarrow j$, which hydrolyzes ATP and releases its products. Transitions that couple to baths of different temperatures, voltage, or osmotic pressure could also have non-zero F_{ij} parameters. Additionally, reaction networks in which a chemical potential gradient F_a can be accessed through multiple pathways will have rates depending on F_a for all such transitions. For example, both the transitions $E^A S \leftarrow S$ and $E^A \leftarrow S^*$ in Fig. 7A are driven by the same F_a depending on enzyme concentration $[E^A]$, which is assumed to be held fixed during the system dynamics and controlled as an input variable.

Contributions to B_{ij} represent symmetric changes to the transition rates between two states. We give two specific biological examples but note that many others are possible. First, we consider a coarse-graining scheme in which the enzymatic reaction $E + S \rightleftharpoons ES \rightleftharpoons E + S^*$ is assumed to be very fast relative to the other dynamics involving S and S^*, and in which the enzyme activity is fixed. It is then possible to show that an effective reaction $S \rightleftharpoons S^*$ has the first order rate constants

$$k_{S\to S^*} = [E]\frac{k_{E+S\to ES}k_{ES\to E+S^*}}{k_{ES\to E+S^*} + k_{ES\to E+S}} \tag{9}$$

and

$$k_{S^*\to S} = [E]\frac{k_{ES\to E+S}k_{E+S^*\to ES}}{k_{ES\to E+S^*} + k_{ES\to E+S}}. \tag{10}$$

The ratio $k_{S\to S^*}/k_{S^*\to S}$ is independent of $[E]$, but $[E]$ symmetrically scales both rates, thus appearing as a contribution $B_{ij} \sim \ln[E]$ in the Markov network model. We refer to refs. 58,63,77 for additional details on how enzymatic reactions may be coarse-grained into Markovian descriptions using the so-called linear framework. Second, we consider a tension-gated ion channel in which the channel's dynamics of opening and closing are fast, and its probability of being open is a function of membrane tension, which we view as an input variable. We can take the ion concentrations on either side of the membrane as among the coarse-grained model variables, and the transition rate through the channel depends on the probability of it being open[78]. This dependence symmetrically scales both directions of ion flow, so that the tension effectively modulates the B_{ij} parameter along the Markovian transition from ions inside the membrane to those outside the membrane.

Structural compatibility

We find that for a given graph structure, not every classification problem (i.e., an assignment of input edges, output nodes, and sets of input data) can be solved. For example, we may assign an input F_a along edge $i \leftarrow j$ and assign node j to be large when $F_a > 0$; this will be very difficult to achieve because all of the spanning trees into node j which involve edge $j \leftarrow i$ will have an exponentially small contribution from the input force, making the input feature vector $\boldsymbol{\chi}(j, \mathbf{F})$ in Equation (3) small; this cannot be helped no matter how we learn the parameters $\boldsymbol{\theta}$. If we flipped the sign of the input F_a in the assignment, then the problem may become solvable. We refer to this mismatch between input force assignment and achievable output node assignment as structural incompatibility. For a fixed set of hyperparameters (i.e., number of nodes, edges, input edges, output nodes, etc.), some problems will be structurally compatible and some will not be. This issue is thus separate from more intrinsic properties of computational expressivity that depend on hyperparameters like M and D, but it implies that we cannot define classification problems completely arbitrarily. We leave to future work a dedicated study of what determines structural compatibility, which may be posed as determining a feasible region as in constrained optimization[79]. Throughout this paper, we bypass this issue to focus on other constraints on expressivity, but we note that structural compatibility represents one limitation on using chemical reaction networks as classifiers. We also note that optimizing the mutual information does not involve specifying output nodes by hand and thus helps to bypass issues of structural compatibility.

Training

We train Markov networks for classification tasks using an approximation to gradient descent, as discussed in refs. 18,80,81. This method requires a variational quantity $\mathcal{L}(\mathbf{p};\mathbf{F},\boldsymbol{\theta})$ which is minimized by the steady-state distribution $\boldsymbol{\pi}(\mathbf{F},\boldsymbol{\theta}) = \text{argmin}_{\mathbf{p}}\mathcal{L}(\mathbf{p};\mathbf{F},\boldsymbol{\theta})$. Two considerations lead to equivalent choices of $\mathcal{L}$: one is that $\mathcal{L}$ should be the KL divergence $\sum_k p_k \ln(p_k/\pi_k)$, which has been shown to act as a Lyapunov function for the evolution of $\mathbf{p}(t)$ and is minimized to zero at steady state[38]. The other consideration is based on the observation that, from Equation (3), $\pi_i \propto e^{-\Phi_i(\mathbf{F};\boldsymbol{\theta})}$ with $\Phi_i = -\ln\left(\sum_{T^\alpha \in \mathcal{T}} w(T_i^\alpha)\right)$ is a Boltzmann-like distribution which must hence maximize a constrained entropy functional. We show in the Supplementary Information that both of these considerations lead to equivalent update rules which require numerically estimating the vectors $\partial_{\boldsymbol{\theta}}\pi_i$ during training.

Optimizing the mutual information

To maximize $I(\mathcal{I};\mathcal{F})$ we numerically evaluate the gradient $\nabla_{\boldsymbol{\theta}} I$ using finite differences, taking as $\boldsymbol{\theta}$ the set of edge rates W_{ij} (softly clipped to be positive) for simplicity. Optimizing with respect to the larger set of parameters E_i, B_{ij}, and F_{ij} is also possible. We use the gradient $\nabla_{\boldsymbol{\theta}} I$ in the Fletcher-Reeves variant of the non-linear conjugate gradient algorithm[79]. We stop the optimization when the difference in $\mathcal{I}$ between iterations falls below 10^{-4} in magnitude. Initial conditions for each W_{ij} are drawn randomly from the interval $[0, 1]$.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

No datasets were generated in this study.

Code availability

Mathematica code used to generate the results in the manuscript is available at https://github.com/csfloyd/NonEqExpressivity.

References

1. Bray, D. Protein molecules as computational elements in living cells. *Nature* **376**, 307 (1995).
2. Bray, D. *Wetware: A Computer in Every Living Cell* (Yale University Press, 2009).
3. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, 2019).
4. Avanzini, F., Freitas, N. & Esposito, M. Circuit theory for chemical reaction networks. *Phys. Rev. X* **13**, 021041 (2023).
5. Murugan, A., Zeravcic, Z., Brenner, M. P. & Leibler, S. Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures. *Proc. Natl Acad. Sci.* **112**, 54 (2015).
6. Evans, C. G., O'Brien, J., Winfree, E. & Murugan, A. Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly. *Nature* **625**, 500 (2024).
7. Antebi, Y. E. et al. Combinatorial signal perception in the BMP pathway. *Cell* **170**, 1184 (2017).
8. Klumpe, H. E., Garcia-Ojalvo, J., Elowitz, M. B. & Antebi, Y. E. The computational capabilities of many-to-many protein interaction networks. *Cell Syst.* **14**, 430 (2023).
9. Parres-Gold, J., Levine, M., Emert, B., Stuart, A. & Elowitz, M. B. Contextual computation by competitive protein dimerization networks. *Cell* **188**, 7 (2025).
10. Goldbeter, A. & Koshland Jr, D. E. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci.* **78**, 6840 (1981).
11. Yoo, H., Triandafillou, C. & Drummond, D. A. Cellular sensing by phase separation: Using the process, not just the products. *J. Biol. Chem.* **294**, 7151 (2019).
12. Hafner, A., Bulyk, M. L., Jambhekar, A. & Lahav, G. The multiple mechanisms that regulate p53 activity and cell fate. *Nat. Rev. Mol. Cell Biol.* **20**, 199 (2019).
13. Chen, Z. et al. A synthetic protein-level neural network in mammalian cells. *Science* **386**, 1243 (2024).
14. Yang, X. et al. Engineering synthetic phosphorylation signaling networks in human cells. *Science* **387**, 74 (2025).
15. Varki, A. et al. *Essentials of Glycobiology* (Cold Spring Harbor Library Press, 2009).
16. Yadav, A., Vagne, Q., Sens, P., Iyengar, G. & Rao, M. Glycan processing in the Golgi as optimal information coding that constrains cisternal number and enzyme specificity. *Elife* **11**, e76757 (2022).
17. Jaiman, A. & Thattai, M. Golgi compartments enable controlled biomolecular assembly using promiscuous enzymes. *Elife* **9**, e49573 (2020).
18. Stern, M., Hexner, D., Rocks, J. W. & Liu, A. J. Supervised learning in physical networks: From machine learning to learning machines. *Phys. Rev. X* **11**, 021045 (2021).
19. Anisetti, V. R., Scellier, B. & Schwarz, J. M. Learning by non-interfering feedback chemical signaling in physical networks. *Phys. Rev. Res.* **5**, 023024 (2023).
20. Dillavou, S., Stern, M., Liu, A. J. & Durian, D. J. Demonstration of decentralized physics-driven learning. *Phys. Rev. Appl.* **18**, 014040 (2022).
21. Hjelmfelt, A., Weinberger, E. D. & Ross, J. Chemical implementation of neural networks and Turing machines. *Proc. Natl Acad. Sci.* **88**, 10983 (1991).
22. Magnasco, M. O. Chemical kinetics is Turing universal. *Phys. Rev. Lett.* **78**, 1190 (1997).
23. Chen, H.-L., Doty, D. & Soloveichik, D. Deterministic function computation with chemical reaction networks. *Nat. Comput.* **13**, 517 (2014).
24. Chan, C.-H., Shih, C.-Y. & Chen, H.-L. On the computational power of phosphate transfer reaction networks. *N. Gener. Comput.* **40**, 603 (2022).
25. Lakin, M. R. Design and simulation of a multilayer chemical neural network that learns via backpropagation. *Artif. Life* **29**, 308 (2023).
26. Poole, W. et al. Chemical Boltzmann machines, in *DNA Computing and Molecular Programming: 23rd International Conference, DNA 23, Austin, TX, USA, September 24–28, 2017, Proceedings 23* (Springer, 2017) pp. 210–231.
27. Zhong, W., Schwab, D. J. & Murugan, A. Associative pattern recognition through macro-molecular self-assembly. *J. Stat. Phys.* **167**, 806 (2017).
28. Su, C. J. et al. Ligand-receptor promiscuity enables cellular addressing. *Cell Syst.* **13**, 408 (2022).
29. Floyd, C., Dinner, A. R. & Vaikuntanathan, S. Local imperfect feedback control in non-equilibrium biophysical systems enabled by thermodynamic constraints. arXiv preprint arXiv:2507.07295 (2025).
30. Owen, J. A., Gingrich, T. R. & Horowitz, J. M. Universal thermodynamic bounds on nonequilibrium response with biochemical applications. *Phys. Rev. X* **10**, 011066 (2020).
31. Owen, J. A. & Horowitz, J. M. Size limits the sensitivity of kinetic schemes. *Nat. Commun.* **14**, 1280 (2023).
32. Fernandes Martins, G. & Horowitz, J. M. Topologically constrained fluctuations and thermodynamics regulate nonequilibrium response. *Phys. Rev. E* **108**, 044113 (2023).
33. Aslyamov, T. & Esposito, M. Nonequilibrium response for Markov jump processes: Exact results and tight bounds. *Phys. Rev. Lett.* **132**, 037101 (2024).
34. Harunari, P. E., Dal Cengio, S., Lecomte, V. & Polettini, M. Mutual linearity of nonequilibrium network currents. *Phys. Rev. Lett.* **133**, 047401 (2024).

35. Mahdavi, S., Salmon, G. L., Daghlian, P., Garcia, H. G. & Phillips, R. Flexibility and sensitivity in gene regulation out of equilibrium. *Proc. Natl Acad. Sci.* **121**, 46 (2024).
36. Liang, S., De Los Rios, P. & Busiello, D. M. Thermodynamic bounds on symmetry breaking in linear and catalytic biochemical systems. *Phys. Rev. Lett.* **132**, 228402 (2024).
37. Arunachalam, E. & Lin, M. M. Information gain limit of biomolecular computation. *Phys. Rev. Lett.* **134**, 148401 (2025).
38. Schnakenberg, J. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.* **48**, 571 (1976).
39. Hart, P. E. et al. *Pattern Classification* (Wiley Hoboken, 2000).
40. Hill, *T. Free Energy Transduction in Biology: The Steady-State Kinetic and Thermodynamic Formalism* (Elsevier, 2012).
41. Bardsley, W. & Wood, R. Critical points and sigmoidicity of positive rational functions. *Am. Math. Month.* **92**, 37 (1985).
42. Shui, S., Scheller, L. & Correia, B. E. Protein-based bandpass filters for controlling cellular signaling with chemical inputs. *Nat. Chem. Biol.* **20**, 586 (2024).
43. Vapnik, V. *The Nature of Statistical Learning Theory* (Springer Science & Business Media, 2013).
44. Dudley, R. M., Kunita, H., Ledrappier, F. & Dudley, R. A course on empirical processes, in *Ecole d'été de Probabilités de Saint-Flour XII-1982* (Springer, 1984) pp. 1–142.
45. Ben-David, S. & Lindenbaum, M. Localization vs. identification of semi-algebraic sets. *Mach. Learn.* **32**, 207 (1998).
46. Bojar, D. et al. A useful guide to lectin binding: machine-learning directed annotation of 57 unique lectin specificities. *ACS Chem. Biol.* **17**, 2993 (2022).
47. Rieckh, G. & Tkačik, G. Noise and information transmission in promoters with multiple internal states. *Biophys. J.* **106**, 1194 (2014).
48. Mayer, A., Balasubramanian, V., Mora, T. & Walczak, A. M. How a well-adapted immune system is organized. *Proc. Natl. Acad. Sci.* **112**, 5950 (2015).
49. Mayer, A., Balasubramanian, V., Walczak, A. M. & Mora, T. How a well-adapting immune system remembers. *Proc. Natl Acad. Sci.* **116**, 8815 (2019).
50. Ramsauer, H. et al. Hopfield networks is all you need. *International Conference on Learning Representations* https://openreview.net/forum?id=tL89RnzIiCd (2021).
51. Lucibello, C. & Mézard, M. Exponential capacity of dense associative memories. *Phys. Rev. Lett.* **132**, 077301 (2024).
52. Stern, M., Liu, A. J. & Balasubramanian, V. Physical effects of learning. *Phys. Rev. E* **109**, 024311 (2024).
53. Krotov, D. & Hopfield, J. J. Dense associative memory for pattern recognition. *Adv. Neural Inf. Process. Syst.* **29** https://doi.org/10.5555/3157096.3157228 (2016).
54. Boukacem, N. E. et al. Waddington landscape for prototype learning in generalized Hopfield networks. *Phys. Rev. Res.* **6**, 033098 (2024).
55. Tkačik, G. & Walczak, A. M. Information transmission in genetic regulatory networks: a review. *J. Phys.: Condens. Matter* **23**, 153102 (2011).
56. Murugan, A., Huse, D. A. & Leibler, S. Speed, dissipation, and error in kinetic proofreading. *Proc. Natl Acad. Sci.* **109**, 12034 (2012).
57. Murugan, A., Huse, D. A. & Leibler, S. Discriminatory proofreading regimes in nonequilibrium systems. *Phys. Rev. X* **4**, 021016 (2014).
58. Owen, J. A., Talla, P., Biddle, J. W. & Gunawardena, J. Thermodynamic bounds on ultrasensitivity in covalent switching. *Biophys. J.* **122**, 1833 (2023).
59. Lammers, N. C., Flamholz, A. I. & Garcia, H. G. Competing constraints shape the nonequilibrium limits of cellular decision-making. *Proc. Natl Acad. Sci.* **120**, e2211203120 (2023).
60. Barato, A. C. & Seifert, U. Thermodynamic uncertainty relation for biomolecular processes. *Phys. Rev. Lett.* **114**, 158101 (2015).
61. Gingrich, T. R., Horowitz, J. M., Perunov, N. & England, J. L. Dissipation bounds all steady-state current fluctuations. *Phys. Rev. Lett.* **116**, 120601 (2016).
62. Horowitz, J. M. & Gingrich, T. R. Thermodynamic uncertainty relations constrain non-equilibrium fluctuations. *Nat. Phys.* **16**, 15 (2020).
63. Gunawardena, J. A linear framework for time-scale separation in nonlinear biochemical systems. *PloS One* **7**, e36321 (2012).
64. Biswas, A. & Thattai, M. Promiscuity and specificity of eukaryotic glycosyltransferases. *Biochem. Soc. Trans.* **48**, 891 (2020).
65. Zwicker, D., Murugan, A. & Brenner, M. P. Receptor arrays optimized for natural odor statistics. *Proc. Natl Acad. Sci.* **113**, 5570 (2016).
66. Qin, S., Li, Q., Tang, C. & Tu, Y. Optimal compressed sensing strategies for an array of nonlinear olfactory receptor neurons with and without spontaneous activity. *Proc. Natl Acad. Sci.* **116**, 20286 (2019).
67. Feinberg, M. *Foundations of Chemical Reaction Network Theory* (Springer Cham, 2019).
68. Dal Cengio, S., Lecomte, V. & Polettini, M. Geometry of nonequilibrium reaction networks. *Phys. Rev. X* **13**, 021040 (2023).
69. Rao, R. & Esposito, M. Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics. *Phys. Rev. X* **6**, 041064 (2016).
70. Dack, A., Qureshi, B., Ouldridge, T. E., & Plesa, T. Recurrent neural chemical reaction networks that approximate arbitrary dynamics. arXiv preprint arXiv:2406.03456 (2024).
71. Baltussen, M. G., de Jong, T. J., Duez, Q., Robinson, W. E. & Huck, W. T. Chemical reservoir computation in a self-organizing reaction network. *Nature* **631**, 549 (2024).
72. Wang, T.-L., Kuznets-Speck, B., Broderick, J. & Hinczewski, M. The price of a bit: energetic costs and the evolution of cellular signaling. *bioRxiv*, https://www.biorxiv.org/content/10.1101/2020.10.06.327700v3.full (2020).
73. Tkačik, G., Walczak, A. M. & Bialek, W. Optimizing information flow in small genetic networks. *Phys. Rev. E* **80**, 031920 (2009).
74. Voliotis, M., Perrett, R. M., McWilliams, C., McArdle, C. A. & Bowsher, C. G. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc. Natl Acad. Sci.* **111**, E326 (2014).
75. Raz, O., Subaşı, Y. & Jarzynski, C. Mimicking nonequilibrium steady states with time-periodic driving. *Phys. Rev. X* **6**, 021022 (2016).
76. Peliti, L. & Pigolotti, *S. Stochastic Thermodynamics: An Introduction* (Princeton University Press, 2021).
77. Nam, K.-M., Martinez-Corral, R. & Gunawardena, J. The linear framework: using graph theory to reveal the algebra and thermodynamics of biomolecular systems. *Interface Focus* **12**, 20220013 (2022).
78. Haswell, E. S., Phillips, R. & Rees, D. C. Mechanosensitive channels: what can they do and how do they do it? *Structure* **19**, 1356 (2011).
79. Nocedal, J. & Wright, S. J. *Numerical Optimization* (Springer, 1999).
80. Scellier, B. & Bengio, Y. Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Front. Comput. Neurosci.* **11**, 24 (2017).
81. Scellier, B. & Bengio, Y. Equivalence of equilibrium propagation and recurrent backpropagation. *Neural Comput.* **31**, 312 (2019).

Acknowledgements

We wish to thank Hector Manuel Lopez Rios, Serena Debesai, Menachem Stern, Martin Falk, Kristina Trifonova, Tarek Tohme, Aleksandra Walczak, and Thierry Mora for helpful discussions. This work was mainly supported by the National Institute of General Medical Sciences of the NIH under Award No. R35GM147400 by funding to SV and CF. AM was funded by the National Institute of General Medical Sciences of the NIH under Award No. R35GM151211. We acknowledge support from the National Science Foundation through the Physics Frontier Center for Living Systems (PHY-2317138). CF acknowledges support from the

University of Chicago through a Chicago Center for Theoretical Chemistry Fellowship. The authors acknowledge the University of Chicago's Research Computing Center for computing resources.

Author contributions

C.F., A.R.D., A.M., and S.V. conceived the research, C.F. carried out the research, and C.F., A.R.D., A.M., and S.V. contributed to writing and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-61873-0.

Correspondence and requests for materials should be addressed to Carlos Floyd or Suriyanarayanan Vaikuntanathan.

Peer review information *Nature Communications* thanks Daniel Maria Busiello, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.