Information processing driven by multicomponent surface condensates

Aidan Zentner[1], Ethan V. Halingstad[2,3], Cameron Chalk[4], Michael P. Brenner[1,5], Arvind Murugan[6], Erik Winfree[4], and Krishna Shrinivas[2,3,*]

[1]School of Engineering and Applied Sciences, Harvard University, 29 Oxford St, Cambridge,MA 02138
[2]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208
[3]Center for Synthetic Biology, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208
[4]Computation and Neural Systems, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125
[5]Department of Physics, Harvard University, 17 Oxford St, Cambridge, MA 02138
[6]Department of Physics, University of Chicago, 929 E 57th St, Chicago, IL 60637
[*]correspondence to: krishna@northwestern.edu

Living organisms rely on molecular networks, such as gene circuits and signaling pathways, for information processing and robust decision-making in crowded, noisy environments. Recent advances show that interacting biomolecules self-organize by phase transitions into coexisting spatial compartments called condensates, often on cellular surfaces such as chromatin and membranes. In this paper, we demonstrate that multicomponent fluids can be designed to recruit distinct condensates to surfaces with differing compositions, performing a form of surface classification by condensation. We draw an analogy to multidimensional classification in machine learning and explore how hidden species, analogous to hidden nodes, expand the expressivity and capacity of these interacting ensembles to facilitate complex decision boundaries. By simply changing levels of individual species, we find that the same molecular repertoire can be reprogrammed to solve new tasks. Together, our findings suggest that the physical processes underlying biomolecular condensates can encode and drive adaptive information processing beyond compartmentalization.

Introduction

Living organisms process information through networks of interacting constituents spanning molecular to ecological scales. In cells, classic examples include gene regulatory circuits and signal transduction pathways where molecular features such as binding and copy number combine to drive biological decisions such as discrimination, feedback control, adaptation, and bistability (1–4). Although biological pathways are often described as modular (2), where a dedicated decision-making module drives distinct downstream events, some computational capability is embedded in processes that appear to serve different cellular tasks. For instance, the very act of building a macromolecular assembly can encode and interpret high-dimensional inputs to trigger context-specific outcomes (5–10). As another example, while genetic control circuits can be engineered to reduce fluctuations in molecular concentrations (11, 12), the same control naturally emerges from the thermodynamics that underlies single-species phase separation (13). More generally, this kind of embedded and distributed computational power is often quite robust due to the underlying collective physics that drives it.

Recently, biomolecular condensation has emerged as a conserved mechanism for spatially organizing the cellular milieu across the tree of life (14–16). Rather than being well-mixed, molecules in cells often self-organize to form dozens of coexisting compartments called condensates. These compartments condense multiple biomolecules through phase transitions (14, 17, 18), typically around intracellular surfaces. Prominent examples span gene regulatory condensates that form at specific DNA (19–23) or RNA scaffolds (24, 25), and signaling condensates that are membrane-localized (26–28). At many surfaces, a particular combination of surface-resident molecules (i.e. “inputs”) like DNA-bound transcription factors or membrane-localized receptors facilitates assembly of specific multicomponent condensates. These condensates, in turn, selectively recruit biomolecules (i.e. “outputs”) like polymerases or signaling messengers from the cellular milieu to drive surface-specific downstream functions—like, for instance, activating certain genes but not others. In multicomponent fluids such as biomolecular condensates, the mapping from molecular parameters to emergent high-dimensional phase behavior is typically nonlinear (29–39). Leveraging this, recent theoretical (40, 41) and experimental (42–48) work highlights the potential of condensates to perform computations beyond compartmentalization. There is interest to understand the design principles and constraints that accompany biomolecular condensate-mediated computations.

In this paper, we explore the computational abilities of biomolecular fluids to assemble surface-specific condensates, i.e., a form of surface classification by condensation. First, we model the exchange of molecules between a surface—characterized by its composition of surface-resident input species—with the broader cellular milieu, or “reservoir”. By exploiting differentiable methods, we tune molecular parameters like intermolecular interactions and reservoir makeup to imbue fluids with desired phase behavior. With this framework, we demonstrate that designed fluids can deploy distinct condensates on surfaces that only subtly differ in their input compositions. This surface classification is driven by the formation of condensates that recruit to certain surfaces, but not others, high concentra-

tions of an output molecule necessary for executing specific downstream functions. The addition of extra hidden species that can interact with all other species but cannot functionally substitute output molecules enhances the capacity to sculpt complex decision boundaries. We show that this expanded expressivity is driven by encoding novel phases that are distinct in hidden species composition but recruit the same outputs. Once designed, we show that simply adjusting hidden species levels in the reservoir enables the same molecular repertoire to classify new tasks. Together, our study suggests that the physics underlying multicomponent condensates offers flexible and versatile mechanisms for information processing in living and synthetic systems.

Model Framework

Motivation

Surface condensation plays a key role in regulating intracellular processes, such as the formation of activating or silencing condensates on distinct genetic loci or varying signaling condensates on the plasma membrane (Fig. 1A) (20, 26). Typically, the combination of loci-specific DNA-bound transcription factors (or surface-localized "input" species) facilitates the assembly of particular condensates. These loci-specific condensates, in turn, selectively recruit either gene-activating polymerases (an example of an "output" species, Fig. 1A, green) or gene-silencing repressors (an example of another "output" species, Fig. 1A, pink) that drive distinct downstream functions. Beyond input and output species, transcriptional cofactors and chromatin remodelers ("hidden" species) often regulate phase behavior and molecular recruitment but ultimately do not directly drive output response. Surface-localized receptor combinations (inputs), downstream messengers or transcription factors (outputs), and adaptors/kinases (hidden species) play analogous roles in membrane-localized condensation. Although these represent different biological pathways, they share similarities in that surface-specific properties enable the assembly of function-specific condensates, a form of classification by condensation.

This motivates a minimal model for the surface condensation of molecules from a complex cellular milieu. The cellular milieu is modeled as an infinite molecular reservoir that exchanges molecules with a surface of volume V. Here, V describes an *effective* volume occupied by the biological scaffold and adjacent interacting molecules, and it can generically describe 2D membranes or 3D DNA loci. In the model, species are partitioned into three types. Input species are localized to the surface at a fixed composition, and distinct surfaces differ in the combination of input species they localize. Unlike the input species, both the output and hidden species freely exchange between the surface and the reservoir.

With this model, our goal is to design a molecular network such that one specific output molecule is recruited to surfaces with specific combinations of input molecules, and a different output molecule is recruited to surfaces with other input combinations. The recruitment of distinct outputs to surfaces with specific combinations of input molecules is possible when the molecular network encodes for multiple types of condensates, i.e., multiple phases where each phase is enriched in only one output species. The formation of one condensate over another in response to subtle differences in input combinations represents a sharp phase transition that can, in principle, be exploited to engineer for ultra-sensitive switches in the recruitment of different output molecules by designing phase boundaries in the space of input composition (41, 49) (Fig. 1B).

Model formulation

Towards this goal, we model a multicomponent fluid with N solute species and an additional solvent species. These N solutes consist of N_{in} input, N_{out} output, and N_{h} hidden species ($N = N_{\text{in}} + N_{\text{out}} + N_{\text{h}}$). For simplicity, the sizes of all species are assumed to be equal to the volume ν of the solvent molecule, and the mean volume fraction of species i is therefore related to the absolute number of particles n_i within the surface by $\phi_i = n_i\nu/V$. We work in the mean-field limit and assume that the surface remains well-mixed. The surface is therefore described by its mean composition vector, labeled as $\vec{\phi} \equiv \phi_{\text{in}} \circ \phi_{\text{out}} \circ \phi_{\text{h}}$, where $\circ$ indicates vector concatenation and

$$\phi_{\text{in}} = (\phi_{\text{in},1}, ..., \phi_{\text{in},N_{\text{in}}}) \quad \textbf{(1)}$$

$$\phi_{\text{out}} = (\phi_{\text{out},1}, ..., \phi_{\text{out},N_{\text{out}}}) \quad \textbf{(2)}$$

$$\phi_{\text{h}} = (\phi_{\text{h},1}, ..., \phi_{\text{h},N_{\text{h}}}) \, . \quad \textbf{(3)}$$

The subvectors denote the input, output and hidden composition vectors, respectively, and the total volume fraction of (non-solvent) species is $\phi_T = \sum_{i=1}^{N} \phi_i$. The surface only exchanges hidden and output species with the infinite reservoir. Within our framework, we don't prescribe any specific model of the reservoir (see SI Note 7) and assume that it can maintain output and hidden species at a constant chemical potential described by

$$\vec{\mu}_{\text{res}} = \left(\mu^{\text{res}}_{\text{out},1}, \ldots, \mu^{\text{res}}_{\text{out},N_{\text{out}}}, \mu^{\text{res}}_{\text{h},1}, \ldots, \mu^{\text{res}}_{\text{h},N_{\text{h}}}\right) \, . \quad \textbf{(4)}$$

The non-dimensionalized free energy density of such a surface is

$$\Omega_{\text{surface}} = \beta\nu f(\vec{\phi}, \chi) - \beta\vec{\mu}_{\text{res}} \cdot \vec{\phi}_{\text{oh}} \quad \textbf{(5)}$$

where $\vec{\phi}_{\text{oh}} = \vec{\phi}_{\text{out}} \circ \vec{\phi}_{\text{h}}$ and $\beta = 1/k_BT$ is the inverse temperature. The quantity $\beta\vec{\mu}_{\text{res}} \cdot \vec{\phi}_{\text{oh}}$ therefore describes the external coupling of the output and hidden species to the reservoir. Additionally, f is the *internal* free energy density of the surface, approximated in Flory-Huggins theory as

$$\beta\nu f(\vec{\phi}, \chi) = \sum_{i=1}^{N} \phi_i \log \phi_i + (1 - \phi_T) \log (1 - \phi_T) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi_i \chi_{ij} \phi_j \quad \textbf{(6)}$$

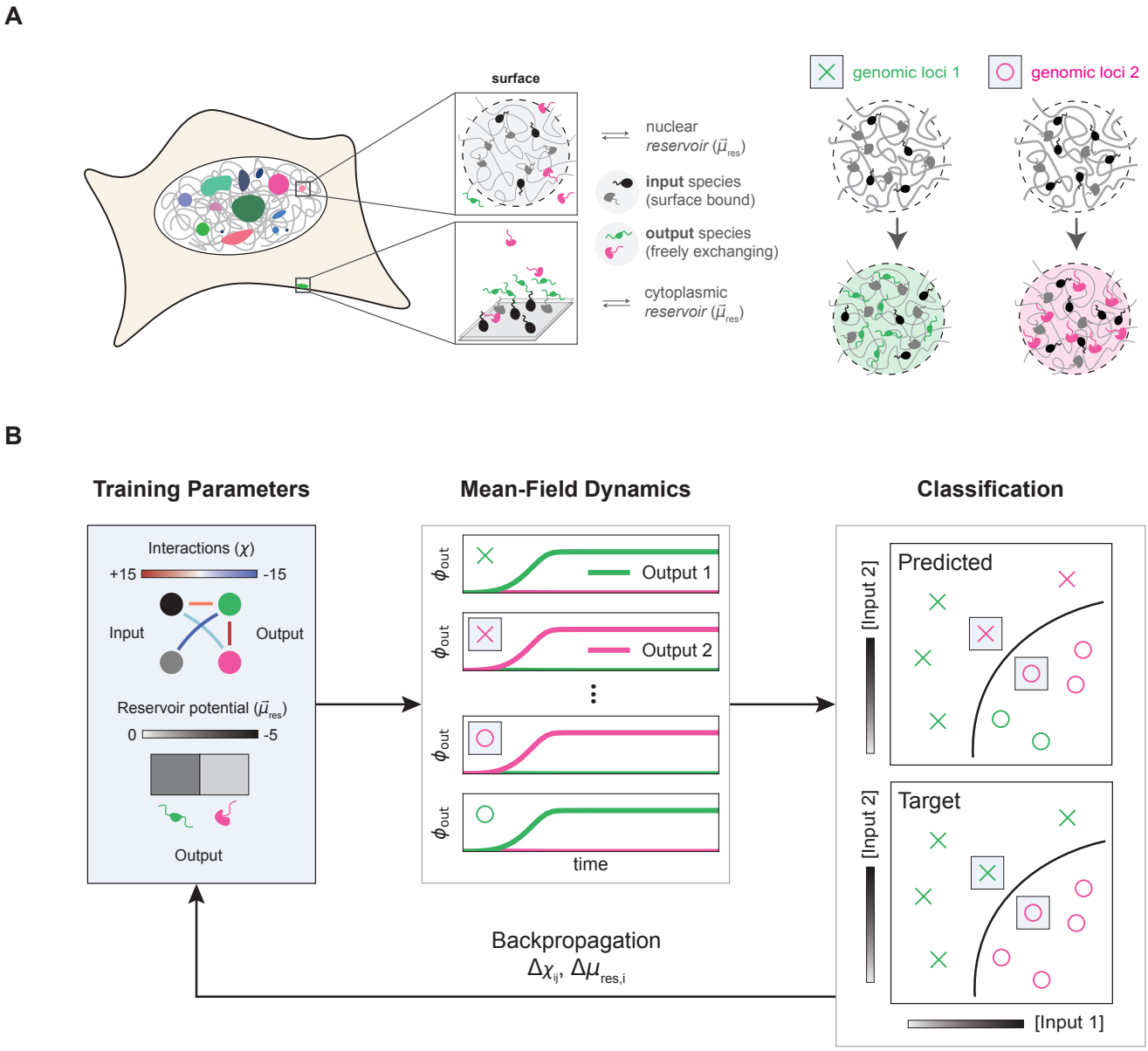

Fig. 1. **(A)** The model is motivated by multiple cellular condensates that form on surfaces such as DNA and bilayers. Species that are localized primarily to the surface, such as transcription factors (DNA) and membrane proteins (bilayer), are modeled as input species (black and gray). Other species, such as coactivators (DNA) or kinases (bilayer), freely exchange between the surface and the cellular environment, or reservoir. Output species (green, pink), in particular, are freely exchanging molecules that can drive a particular downstream function—for example, polymerases (DNA) that turn on genes or allosteric activators (bilayer) that can translocate to complete signal transduction. Polymerases are recruited to active genes (green) and repressors to silenced genes (pink). **(B)** Motivated by (A), we consider a simplified model in which surfaces characterized by the presence of different combinations of input species recruit distinct output species from an infinite reservoir. (Left) The key parameters of the model are the interactions χ between the species and the reservoir chemical potentials $\vec{\mu}_{\text{res}}$. (Middle) We consider the evolution of surfaces in the well-mixed, mean-field limit. (Right) The recruitment of distinct outputs is accomplished through forming multiple phases in the parameter space of input concentrations. The coexistence line across which the system undergoes an abrupt phase transition functions as a decision boundary in a classification of surfaces. (Loop) We use JAX to iteratively tune χ and $\vec{\mu}_{\text{res}}$ with the goal of recruiting the desired output species (based on the phase label, x vs o) for each training data point.

where χ is the effective interaction matrix,

$$\chi_{ij} = z\beta\left(\epsilon_{ij} - \frac{\epsilon_{ii}+\epsilon_{jj}}{2}\right) \tag{7}$$

and z is the number of nearest interacting neighbors, with ϵ_{ij} being the microscopic nearest-neighbor contact energy between species i and species j. Note that since $\chi_{ii} = 0$ by definition, the $i = j$ terms do not contribute to the above equation, and further the equation assumes negligible effective solute-solvent interactions.

We next write a dynamical model to probe the steady-state composition of a surface characterized by a fixed input species composition $\vec{\phi}_{\text{in}}$. The volume fractions of all non-input species evolve over time due to the exchange with the reservoir until a steady-state is reached. We treat these compositional dynamics as near-equilibrium relaxation that, to a first approximation, is driven by linear gradients of the free energy with respect to the surface's composition (50). Here, we assume that solvent molecules have much faster dynamics than solutes, which improves nu-

merical stability of the optimization but does not affect the steady-state (Fig. S6). Thus, the temporal evolution of the surface composition $\vec{\phi}_{\text{oh}}$ of the exchanging output and hidden species is written as (SI Note 1)

$$\frac{d\vec{\phi}_{\text{oh}}}{dt} \approx -\mathbf{D}\frac{d\Omega_{\text{surface}}}{d\vec{\phi}_{\text{oh}}} = -\beta\mathbf{D}(\vec{\mu} - \vec{\mu}_{\text{res}}) \quad \textbf{(8)}$$

where $\beta\mu_i = \partial(\beta\nu f)/\partial\phi_{\text{oh},i}$ is the intrinsic (non-dimensionalized) chemical potential of exchanging species i. $\mathbf{D}$ is the $(N_{\text{out}} + N_{\text{h}}) \times (N_{\text{out}} + N_{\text{h}})$ mobility matrix that sets the rate of exchange between the surface and reservoir and is chosen, for simplicity, to be diagonal, identical for solutes, and consistent with Fick's law at dilute equilibrium conditions (36) (SI Note 1). At steady state, the surface and reservoir must have identical chemical potentials in the non-input species but can have distinct compositions—a feature of multiphase systems that we aim to exploit.

Designing multiphase classifiers

With this forward model, our goal is to identify an effective interaction matrix χ and reservoir chemical potential $\vec{\mu}_{\text{res}}$ (at $\beta = 1$) such that, for a surface defined by a given input vector $\vec{\phi}_{\text{in}}$, the steady state is enriched in the desired output species and depleted in all other outputs (Fig. 1B). This output convention is akin to "one-hot" representations common in machine learning. To train this model, we employ a differentiable implementation of the above dynamical description amenable to gradient-based optimization methods that minimize a loss function (51).

In particular, we require that the following criteria be captured by our loss function: first, the final concentration of the desired output species should be above some threshold value $\phi_{\text{max}} = A/N$, where A is a value to be specified. Second, the final concentrations of the undesired output species should be below some threshold $\phi_{\text{min}} = B/N$, where B is a value to be specified. These two criteria in turn enforce that the ratio of desired to undesired outputs should be above a set threshold A/B, and that this ratio is attained with a sufficiently enriched output species. We choose $A = 1.1$ (mild enrichment above $1/N$), and $B = 0.25$ (significant depletion below $1/N$).

We find empirically that the following loss function gives the best performance in optimizing for these two criteria:

$$\mathcal{L}(\chi, \vec{\mu}_{\text{res}}) = \frac{1}{n_{\text{batch}}} \sum_{a=1}^{n_{\text{batch}}} l_{j(a)}\left(\chi, \vec{\mu}_{\text{res}}; \vec{\phi}_a\right) \quad \textbf{(9)}$$

where the sum is over n_{batch} data points in the training set, data point a corresponds to a surface that reaches steady-state concentrations $\vec{\phi}_a$, and $j(a)$ is the index of the desired output species for data point a. We define the function

$$l_j(\chi, \vec{\mu}_{\text{res}}) = \log(1 + Np_j) + \sum_{\substack{k=1 \\ (k\neq j)}}^{N_{\text{out}}} \log(1 + Nq_k) \geq 0 \quad \textbf{(10)}$$

where

$$p_j = \max(0, \phi_{\text{max}} - \phi_{\text{out},j}) \quad \textbf{(11)}$$

$$q_k = \max\left(0, \phi_{\text{out},k} - \phi_{\text{min}}\right). \quad \textbf{(12)}$$

The term l_j is therefore at a global minimum when $p_j = q_k = 0$, and $\mathcal{L}$ is at a global minimum when this condition is satisfied for all data points a.

We minimize $\mathcal{L}$ with respect to χ and $\vec{\mu}_{\text{res}}$ over several thousand training epochs using an RMSProp algorithm from the Optax library (51, 52). Once trained, we evaluate the performance of the classifier using a success criterion that follows from the definition of the loss function: given a test point a of input concentrations, the surface's steady-state composition must be enriched above ϕ_{max} in the $j(a)$'th output and depleted below ϕ_{min} in all other outputs for the point to be considered successfully classified. The classification success is therefore

$$S_c = \frac{1}{n_{\text{set}}} \sum_{a=1}^{n_{\text{set}}} \left[1 - \Theta(l_{j(a)})\right] \quad \textbf{(13)}$$

where $n_{\text{set}} = 500$ is the number of points in the validation/test set and $\Theta(x) = 0$ if $x = 0$ and is 1 otherwise.

In training the system over χ and $\vec{\mu}_{\text{res}}$, we impose several constraints. First, since we are modeling liquid phases, we require that energies be of order k_BT and therefore enforce that each entry of the chi matrix has $|\chi_{ij}| < 15$, which is $\mathcal{O}(z)$. Second, since we are designing surfaces to only enrich one particular output species, we require that output-output interactions be repulsive, with $\chi_{ij} > 10$ for distinct output species i and j. Third, we enforce that all output species have the same reservoir chemical potential as a design criterion, which is meant to mimic the surface choosing from outputs that are at "identical" potentials in the reservoir. Finally, since input-input interactions and input chemical potentials don't affect steady-state behavior in the mean-field limit, they are omitted from the model and not treated as free design parameters.

Results

Tuned molecular networks drive linear classification

Armed with this model, we first aim to create linear classifiers. In a simple mixture comprising only 2 input and 2 output species, our target is to design surface condensates that recruit a specific output molecule (green or pink) depending on which input species is at higher concentrations (Fig. 2A), i.e., an ultrasensitive ratiometric sensor. With this objective in mind, we initialize a non-interacting liquid mixture and train the molecular interactions and reservoir potential over multiple epochs (Fig. 2B). The learned interaction matrix broadly matches physical intuition, with each input preferring to mix with the cognate output and demix from the non-cognate output. Upon testing, our model exhibits a sharp switch in composition across the boundary (Fig. 2C). This switch is consistent with a first-order phase

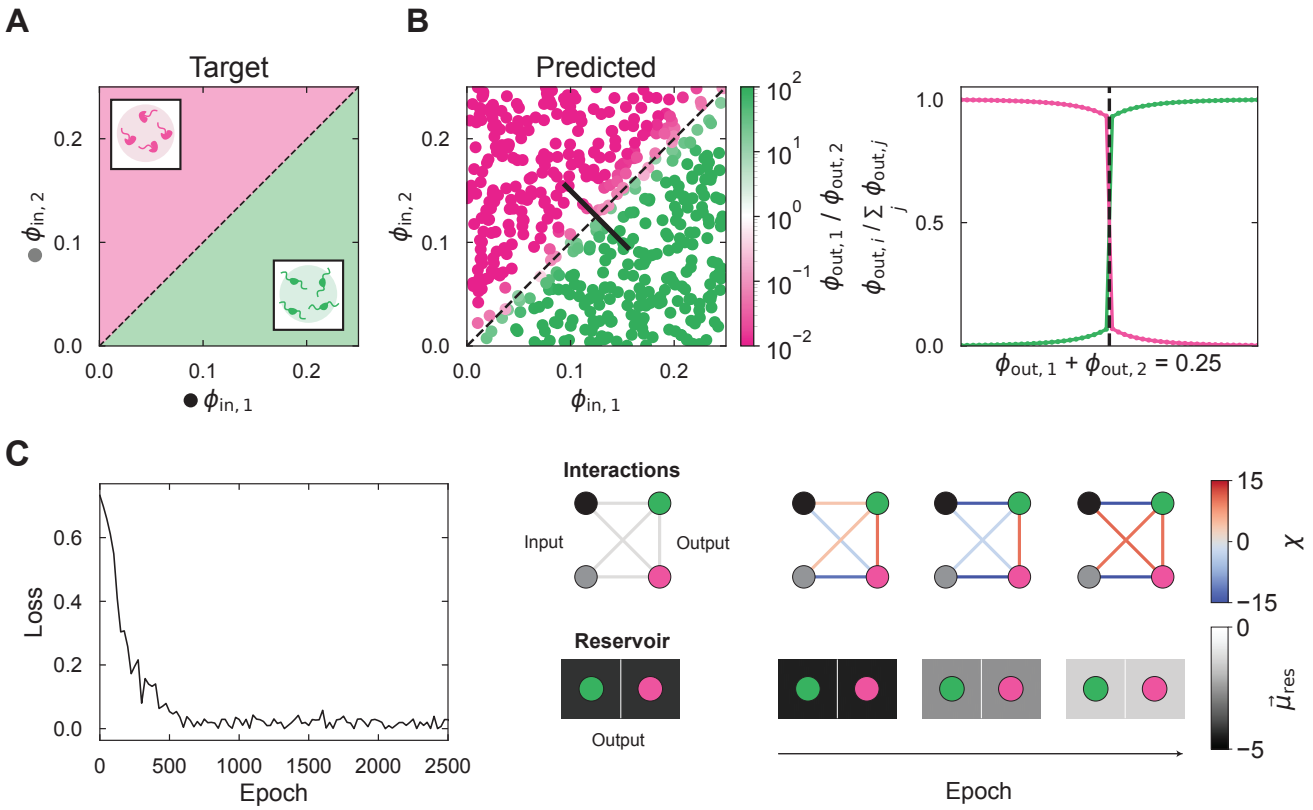

Fig. 2. (A) The target linear decision boundary is shown, with each axis being the concentration of one of the input species. Green and pink denote regions where we desire condensates enriched in the green and pink component, respectively. **(B)** Predictions from the trained model for different input compositions in the test set. The axes depict the input concentrations while each dot is a test input condition, colored by the ratio of the two output species at steady-state (displayed on a log-scale). Along the solid black line, the system undergoes a discontinuous transition in mean-field composition across the boundary, as shown in the right-most panel. **(C)** Evolution of training loss and parameters over the optimization. The training parameters converge to a solution that is analytically consistent with the formation of a linear decision boundary (SI Note 3).

transition (Fig. 2C) that is characterized by a temperature-dependent discontinuity in output recruitment (Fig. S1B).

To understand how the decision boundary emerges from molecular parameters, we develop a simple analytical approach (SI Note 3). We first define the decision boundary as the manifold where all output species are recruited at identical levels. We find that the expressivity (or repertoire of encodable manifolds) of mixtures with 2 inputs and 2 outputs is limited to linear boundaries, and this theoretical prediction is supported by simulation (Fig. S2A, SI Note 3). More generally, we show that liquids with input and output species can only typically encode linear decision boundaries in input space (SI Note 3). Consistent with this prediction, we find that our model still sharply classifies higher-dimensional linear manifolds (Fig. S2B).

To test our model's prediction that purely input-output mixtures cannot classify nonlinear boundaries, we train a 2 input and 2 output mixture to separate an elementary nonlinear manifold: an upper quadrant AND-like distribution, in which one output is recruited only when both inputs are present at high concentrations; otherwise, the other output is recruited. After training, we find that input-output mixtures fail to encode this nonlinear decision boundary, instead showing a best-fit linear approximation (Fig. S4).

Hidden species expand capacity for nonlinear complex decisions

The inability to form nonlinear decision boundaries with simple input-output networks motivates the introduction of hidden species. In our model, hidden species are similar to output species in that they can interact with all molecules and be recruited to surfaces from the reservoir, thereby influencing the steady-state condensate that forms. However, their steady-state concentrations are taken to be irrelevant in performing the classification of the surface; they therefore play a role analogous to that of hidden nodes in a neural network (53).

We explore how adding hidden species to our model could enhance classification (Fig. 3A). Extending our analytical approach, we find that the addition of a single hidden species offers enough flexibility to encode decision boundaries of varying curvatures (Fig. S2C, SI Note 3). We thus explore classification of complex, high-dimensional decision boundaries by including multiple hidden species.

First, we demonstrate the effectiveness of hidden species by programming an AND-like upper quadrant decision boundary with two additional hidden species (gold and cyan in Fig. 3B). Analyzing the trained molecular network reveals a complex interplay of interactions that leads to es-

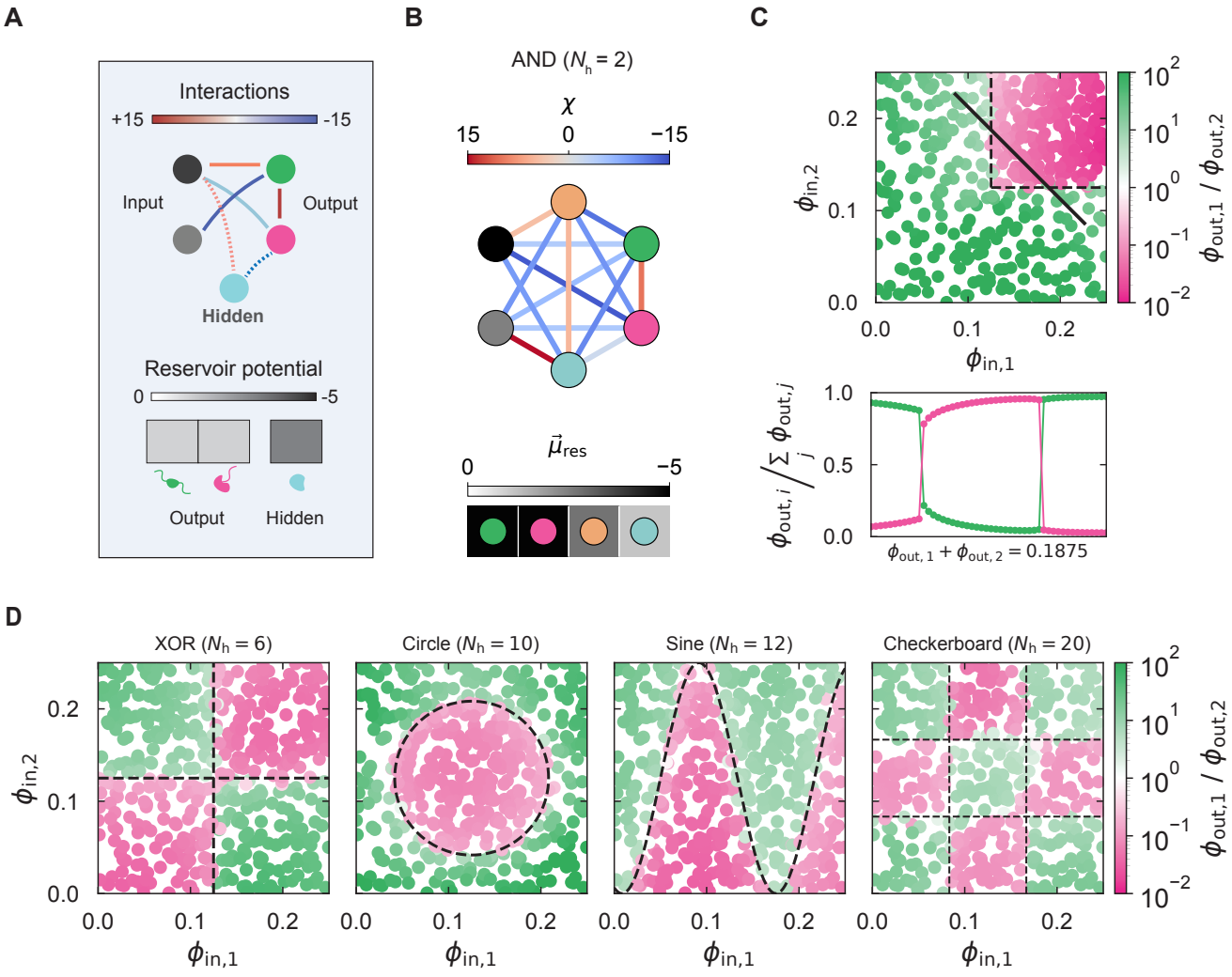

Fig. 3. **(A)** Hidden species, depicted in cyan in the interaction matrix and analogous to hidden nodes in Boltzmann machines, shape emergent overall phase behavior by interacting with input and output species but cannot directly drive output function. **(B)** Using only 2 hidden species (gold and cyan), we train for parameters to form an AND-like upper quadrant decision boundary in the mean-field limit. **(C)** Predictions from the trained classifier for different input compositions in the test set, which resembles the desired decision boundary. Along the solid black line, the system undergoes a discontinuous transition in mean-field composition across the boundary. **(D)** Test predictions for models with other nonlinear decision boundaries in the mean-field limit. (left to right) A XOR boundary trained with 6 hidden species, a circular boundary trained with 10 hidden species, a sine curve boundary trained with 12 hidden species, and a checkerboard boundary trained with 20 species.

sentially binary responses in the output species recruited to the surface (Fig. 3C, Fig. S4). Like the linear classifier, and consistent with a phase transition, our trained AND system exhibits a sharp switch in composition across the decision boundary (Fig. 3C). With the addition of more hidden species, the model can encode increasingly nonlinear decision boundaries such as XOR, circle, sinusoidal, and checkerboard patterns (Fig. 3D). Similar to the AND boundary, each of these systems exhibit sharp, near-discontinuous switches in the recruited species across the boundary (Fig. S5). The trained parameters for each decision boundary are shown in Fig. S7.

Hidden species expand capacity by encoding multiple modular, encrypted phases

To understand how hidden species enhance expressivity, we trained mixtures with varying numbers of hidden species to solve a range of decision boundaries, and evaluated the classification success (as defined in eq. 13). We find that the addition of hidden species improves classification but saturates beyond a decision-boundary specific threshold (Fig. 4A, left). While surface condensates correctly enrich the pertinent output, we find that surfaces with the same output molecules often recruit varying concentrations of hidden species. To better understand this, we estimated how many distinct phases were formed as defined by the *overall* composition of hidden and output species on surfaces. Collecting the compositions across multiple surfaces (n_{set} test points into a matrix of size $n_{set} \times (N_{out} + N_h)$), we perform principal component analysis and use a Marchenko-Pastur (54) based threshold to estimate the number of distinct phases from the significant eigenmodes. We then perform hierarchical clustering to identify the average composition of each phase (see SI

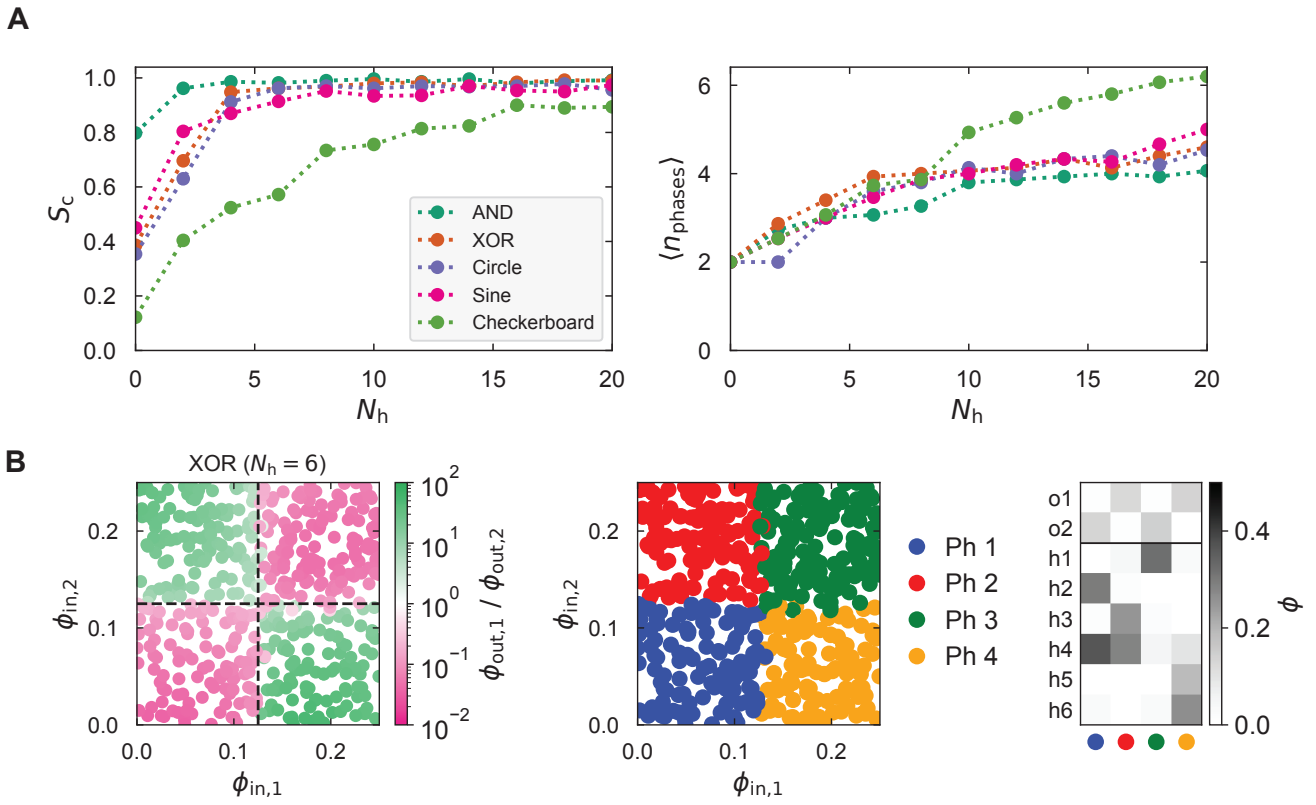

Fig. 4. (A) The scaling of the test set classification success S_c (left) and the number of phases averaged across multiple training trajectories (right) as a function the number of hidden species in the model. **(B)** When defined in terms of output species composition alone, the XOR liquid (left) shows two distinct phases, each repeated twice in the four quadrants of input space. However, each of these quadrants corresponds to a distinct phase (middle) if phases are distinguished by the composition of hidden species (right).

Note 5). We find that the number of steady-state phases with distinct compositions grows with hidden species (Fig. 4A, right). This suggests that encoding multiple phases plays an important role in improving expressivity of multicomponent condensates.

To explore this deeper, we consider the trained XOR liquid with 6 hidden species (Fig. 4B). Compositional analysis reveals that the XOR decision boundary is achieved through 4 *distinct phases*. For example, areas with high output 1 (green output) are encoded by 2 distinct phases (e.g., red and yellow phases, or Ph2 and Ph4) that recruit different hidden species but the same output species. Identifying each point with an independent surface, our model shows that multiple surfaces that condense the same output recruit distinct hidden species, and thus vary in phase composition. Biologically, such a solution might look like condensates that drive gene activation at different DNA loci by recruiting high concentrations of the functional polymerase but varying concentrations of coactivator molecules. Thus, the encoding of multiple *encrypted* phases, which differ in hidden species but recruit similar output molecules, is the primary mechanism by which hidden species improve expressivity.

In the XOR liquid, we find that the 4 distinct encrypted phases modularly partition the input space into quadrants, such that groups of related inputs drive condensation of a particular phase. When we extend this analysis to other nonlinear decision boundaries, we find that hidden species generally learn modular representations of related input surfaces (Figs S8-S12). We next explore whether we can repurpose this modular multiphase representation learned by hidden species for other tasks.

Changing reservoir composition of hidden-species drives solution of new classification tasks

Motivated by the modularity of encoded phases, we hypothesized that once trained with sufficient hidden species, the same molecular ensembles could be adapted to solve new decision tasks by simply tuning the reservoir of hidden species without changing interactions. This idea is analogous to machine-learning architectures comprising modules where an upstream (typically randomly-wired) network remains fixed and solutions to new tasks are achieved by training only the parameters of a small downstream network (55, 56). To demonstrate this idea, we revisit the trained XOR liquid and ask whether it can be repurposed to solve AND or OR decision boundaries only by changing reservoir composition. We find that changing the potential of a few key hidden species is sufficient to fine-

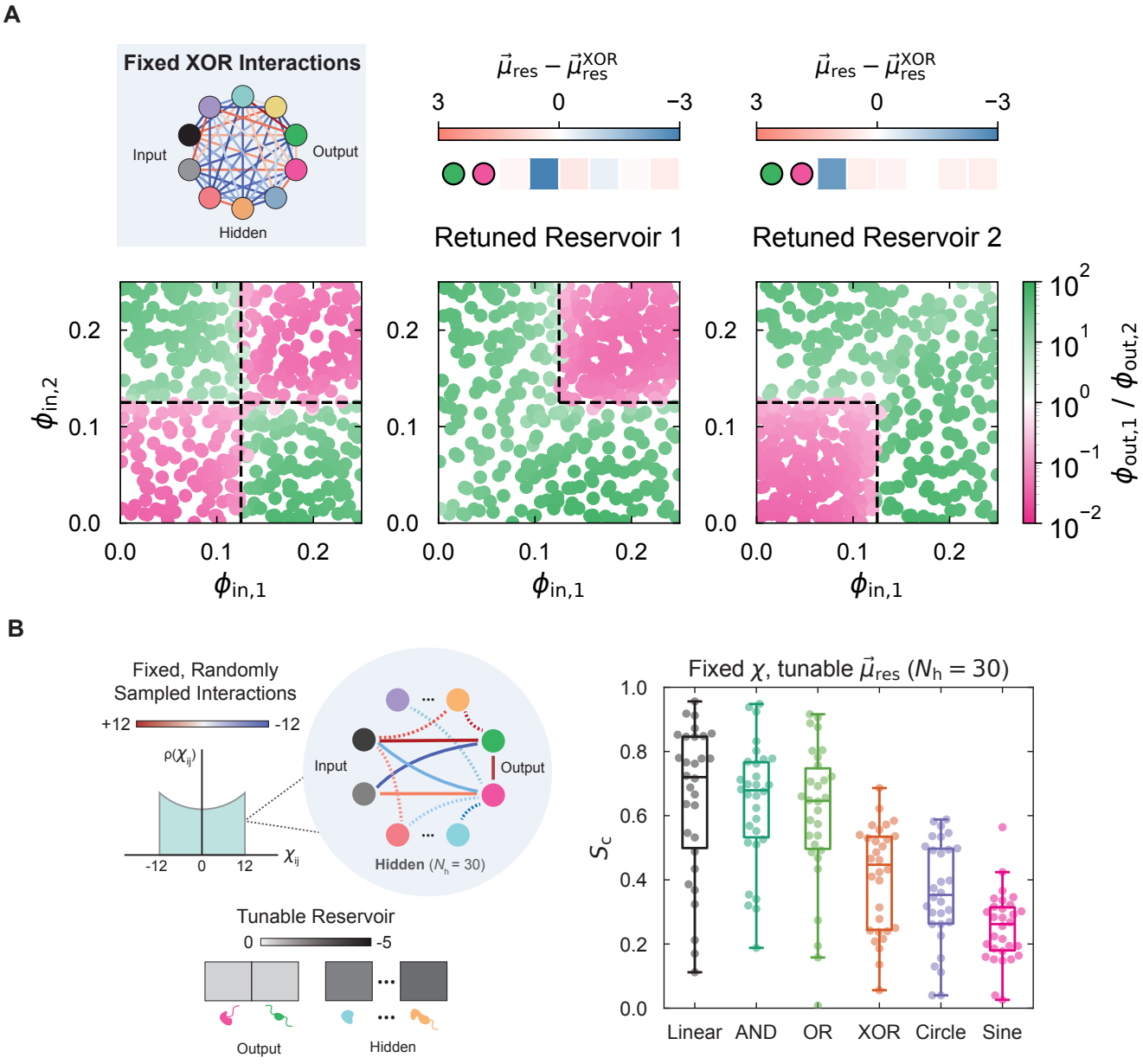

Fig. 5. (A) Given the previously-trained interaction network with $N_h = 6$ and reservoir potential vector that solve a XOR decision boundary (left), the same interaction network can also solve an AND decision boundary (middle) and OR decision boundary (right) by selectively tuning the reservoir potentials. **(B)** The capacity for classification by liquids with random interaction networks. (left) A random interaction matrix with $N_h = 30$ is sampled from a probability distribution (see SI Note 5) and remains immutable; subsequently, reservoir potentials are tuned to solve a classification task. (right) The success rate S_c across multiple trajectories (30 per task, where each has a different random interaction network) for various classification tasks.

tune the same molecular mixture to perform distinct tasks (Fig. 5A).

Given this finding, we next explored whether liquids *without* designed interactions, e.g., with randomly chosen molecular interactions χ_{ij}, could nevertheless be trained to classify surfaces through fine-tuning the reservoir alone. To test this, we generated liquids with 2 inputs, 2 outputs, and a large number of hidden species ($N_h = 30$, SI Note 6). The interactions between species were sampled from a near-uniform distribution such that $|\chi_{ij}| \lesssim 12$; for a fixed decision boundary, we report the distribution of model performance over $n = 30$ different interaction networks (SI Note 6). Through training only the reservoir makeup, we show that liquids with randomly chosen and fixed interactions χ contain the ability to model both linear and nonlinear decision boundaries, albeit with decreasing performance as we increase the complexity of the decision boundaries that we seek to approximate (Fig. 5B).

Our results suggest that rather than constantly redesigning or evolving new interactions, the physics of surface condensation provides a flexible mechanism to redeploy the same molecular repertoire to solve new tasks by adjusting compositions of the reservoir. An analogous idea has been explored previously by Elowitz and coauthors in

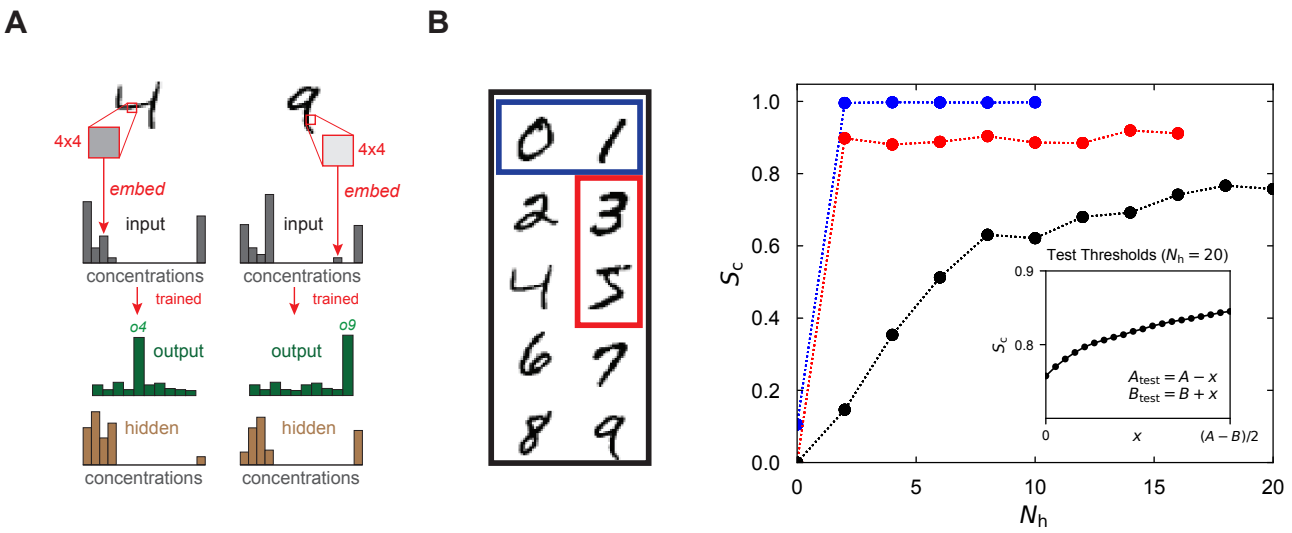

Fig. 6. (A) Solving classification of the MNIST dataset involves embedding pixel grayscale values into many input species concentrations, then training interaction parameters and reservoir concentrations for the formation of condensates enriched in 1 of 10 output species. **(B)** Classification success S_c for 0 vs 1 (blue), 3 vs 5 (red), and 10-digit MNIST (black) tasks as a function of number of hidden species. For the 10 digit MNIST dataset, the classification success reaches a plateau below 1 at approximately 15-20 hidden species. The inset shows the plateau value when the success criterion is made more lenient by decreasing $A_{\text{test}} = A - x$ and increasing $B_{\text{test}} = B + x$ in the success criterion, where $A = 1.1$ and $B = 0.25$ are the threshold values defined in the loss function used to train the classifier. In the extreme case where $A_{\text{test}} = B_{\text{test}}$, the maximum success is approximately $\sim 85\%$.

the context of BMP signaling and dimerization networks (4, 57) where they show that tuning stoichiometries but not binding affinities in dilute molecular ensembles can facilitate solving distinct tasks. Together, this highlights that the physics embedded in collective molecular networks permits flexible computations at distinct hierarchies.

Surface condensates classify high-dimensional datasets

Our motivation for physically embedded computation in phase separation is to process chemical stimuli in cells through concentration-dependent condensation, not to build a general-purpose classifier for arbitrary domains (e.g., distinguishing cat vs. dog images). In the same spirit, related work has evaluated physical systems as classifiers of physical stimuli in many domains (58), ranging from molecular concentrations to mechanical forces (59, 60). Molecular examples include winner-take-all reaction networks (61), self-assembly with Hebbian-like interactions (8, 10), and multicomponent liquids (40, 41). Nevertheless, to evaluate expressivity of these physical systems on high-dimensional inputs in a standardized way, we follow this literature and use symbolic ML datasets as benchmarks, not as an end in themselves: each feature is reinterpreted as a molecular concentration and presented to the system as a physical stimulus.

We start by classifying the near-linear Seaborn Iris dataset, which comprises 4 analog flower features (petal and sepal length and width) and 3 output labels (flower species). The value of the j'th feature x_{aj} of the a'th data point x_a is encoded as an input concentration according to the linear, scaled mapping $\phi_{aj} = \phi^0 \left(\frac{x_{aj} - \min_a x_{aj}}{\max_a x_{aj} - \min_a x_{aj}} \right)$, where $\min_a x_{aj}$ and $\max_a x_{aj}$ denote the minimum and maximum x_{aj} across all data points a, respectively, and $\phi^0 = 0.5/4 = 0.125$, such that the input species can occupy a maximum of half of the volume. We incorporate 1 output species per label to mimic a species-specific molecule. Once trained, we demonstrate that this 7-component mixture can directly classify the IRIS dataset without the use of hidden nodes (Fig. S13).

Next, we turn to the higher-dimensional MNIST dataset, a collection of labeled hand-drawn images of digits, to study how our model generalizes to larger interaction networks. We first coarse-grain each grayscale image from 28×28 to 7×7 by averaging pixel values in a 4×4 block and assign each pixel in the reduced image to an input species. Then, we map the volume fraction $\phi_{\text{in},i}$ of an input species to its corresponding pixel value (x_i) by $\phi_{\text{in},i} = \phi^0(x_i/255)$, where $\phi^0 = 0.5/49 \approx 0.01$. We train the mixture to initially discriminate between two digits, achieving strong performance with just a few hidden nodes ($\sim$2-3). However, digits that are traditionally harder to distinguish (Fig. 6B, red, 3 vs 5) reached lower performance levels compared to easier ones (Fig. 6B, blue, 0 vs 1). Extending the model to simultaneously classify all ten digits requires more hidden nodes (~ 15) and saturating performance is lower (Fig. 6B, black). As we relax the classification stringency by requiring lower and lower excess of the desired output species over the undesired ones without retraining the system (Fig. 6B inset), the success in classifying MNIST increases from $\sim 75\%$ to a saturating test success of $\sim 85\%$. The confusion matrices for each of these three cases is shown in (Fig S13). More generally, the ability to design condensates with large numbers of species for high-dimensional

capacity is improved with hidden nodes but typically saturates.

These results are generally consistent with recent findings from (41), where the authors develop a 3D lattice condensate model and train it with a probabilistic learning algorithm derived from classical Boltzmann machines to classify MNIST digits with $\sim 75\%$ accuracy. Their lattice liquid with a "semipermeable membrane" is conceptually equivalent to our approximation of "surface-localized inputs". In (10), which explores crystalline self-limited assembly, MNIST digit classification is similarly demonstrated in a theoretical model with $\sim 85-90\%$ accuracy depending on the design constraints. Together, these results highlight the potential of multicomponent interacting mixtures to effectively classify high-dimensional decision boundaries despite the different choices in microscopic physics, training algorithms, design constraints, and problem encodings.

Mean-field solutions translate to successful classifiers in 3D lattice liquids

We next aim to understand whether the mean-field design of liquids transfers to a more detailed 3D model that explicitly captures spatial correlations. Following earlier work (31, 41, 62), we adopt a lattice liquid formulation in which we treat a surface as a lattice of length $L \times L \times L$ with 1 molecule per site. Interactions between 18 nearest neighbors, i.e., those within a $\sqrt{2}$ lattice distance, contribute to the overall energy of the system, which thus depends on the spatial configuration of molecules (SI Note 4). To mimic our mean-field treatment of surface-localized and well-mixed input species, we fixed their counts and positions on the lattice, thereby treating them as immobile and non-exchanging in the canonical ensemble (Fig. 7A). Output and hidden species are allowed to exchange with the reservoir at a fixed chemical potential, i.e., in the grand-canonical ensemble. Finally, we sample this mixed-ensemble model through parallelized Monte-Carlo simulations to ensure sufficiently equilibrated thermodynamic properties and compositions (SI Note 4, Fig. S14B).

Unlike (41), we do not train molecular parameters using this lattice liquid; instead, we simulate the lattice liquid with trained parameters from the mean-field model and evaluate it's ability to classify surfaces. The designed mean-field interactions are rescaled to account for the number of nearest-neighbors to parameterize this lattice liquid. We find empirically that decreasing temperature (or increasing β) sharpens the decision boundary in the lattice model (Fig. S14A), and all test data shown is at $\beta = 2$ in Fig. 7.

Using liquids trained on a range of decision boundaries reported in Figs 2-3, we parameterize and sample the equilibrium configurations of the 3D lattices. Overall, we find that lattice liquids broadly encode similar classification boundaries as their mean-field counterparts (Fig. 7B) with a few key differences. Near the decision manifolds, we find that lattice liquids exhibit more continuous variation unlike the abrupt jumps in mean-field liquids - likely arising from coexisting but spatially isolated pockets of both output species. Away from the boundary, output species ratios still reach $10-100\times$ ratios of correct over incorrect species (see Fig. S15). Finally, we find that as the decision boundary increases in complexity , and thus requires more hidden species, the asymptotic classification success in the 3D liquid typically decreases (Fig. 7B, Fig S14A). Together, the broad concordance between mean-field and 3D lattice liquids supports the generality of our results and motivates direct avenues for experimental testing.

Discussion

Across the tree of life, biomolecules in cells can self-organize into membraneless organelles called condensates that regulate biological pathways. Motivated by this fact, we explore the computational capabilities that are embedded in and arise from the physical processes shaping condensation in multicomponent mixtures. We find that multicomponent liquids can recruit distinct molecules (and thus condensates) to surfaces that differ only subtly in their composition of surface-resident "input" molecules. This high-dimensional surface classification is offered as a model of how cells might assemble transcriptionally active condensates at certain genetic loci (with a particular combination of DNA-bound transcription factors) but repressive ones at other DNA surfaces (with a different combination of transcription factors). Together, our work suggests that emergent condensation in multicomponent liquids like the cellular milieu can drive computations and information processing that may be necessary for regulating complex biological functions.

We show that inclusion of hidden species—molecules that shape condensation but do not drive downstream function—expands *expressivity* (63), i.e., the ability to encode increasingly complex classification boundaries. We find that hidden species improve expressivity through encoding novel phases that differ in composition of hidden molecules but still recruit the same *functional* output species. The role of such species could be played by different coactivators that recruit the same polymerases to drive gene activity (64), varying co-receptors and adaptor proteins that recruit the same downstream kinase to membranes to propagate signaling cascades (26), and more generally by regulatory molecular cascades. In addition, hidden species simultaneously facilitate *adaptability* by allowing reuse of the same molecular interaction networks, including even purely random ones , to perform distinct tasks (Fig. 5A-B) simply by changing makeup of the cellular milieu. This adaptability loosely mimics cell-type specific expression, in which cellular compositions can use the same genetically-encoded molecular ensemble to drive different gene programs with the same functional molecular output species—a feature that emerges in other multicomponent biomolecular networks (4, 65, 66). More generally, the features of multicomponent phase separation naturally provide cells with regulatory knobs such as

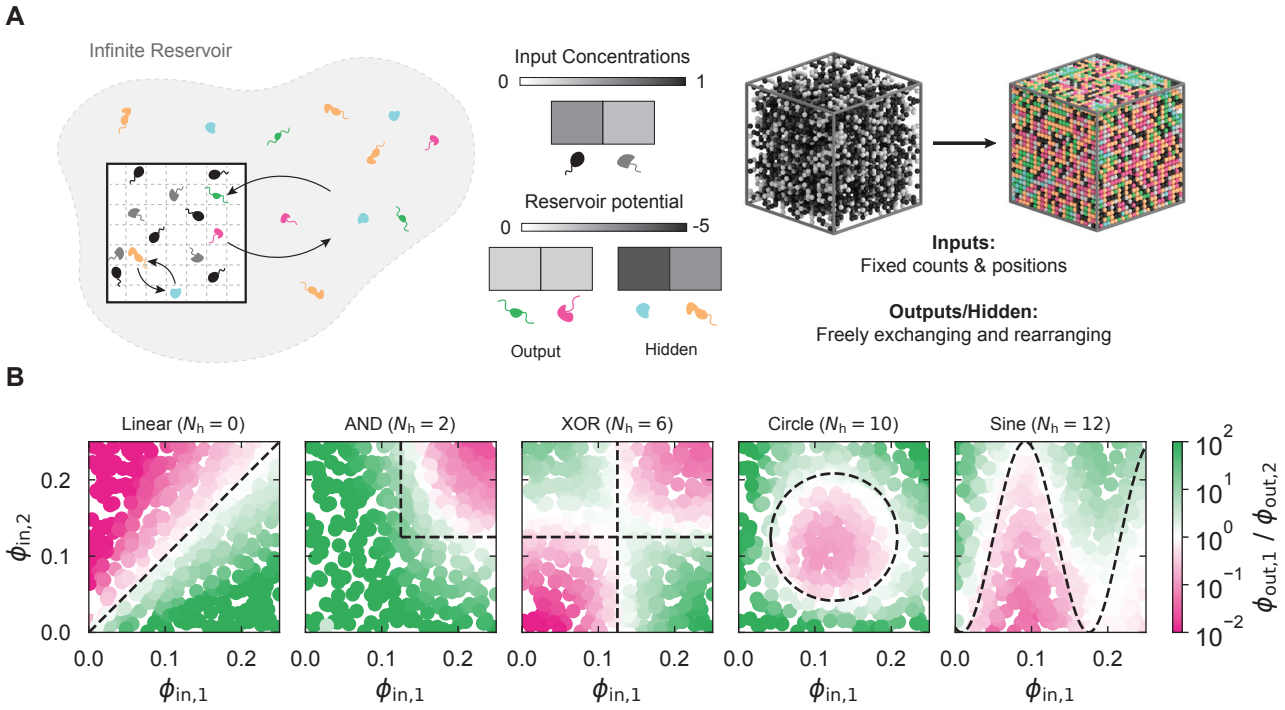

Fig. 7. (A) Schematic of 3D lattice liquid. Inputs are clamped to lattice positions, and output and hidden species can exchange freely. On the right is the evolution of a lattice from initial to final configurations. **(B)** Analog 3D lattice liquid classification using mapped mean-field parameters: (left to right) A linear boundary with 0 hidden species, an AND boundary with 2 hidden species, a XOR boundary with 6 hidden species, a circular boundary with 10 hidden species, and a sine boundary with 12 hidden species.

changing composition (by expression) or interactions (by post-translational modifications) to leverage condensate-mediated computations. Finally, our results emphasize an expanded view of biological condensates through the "hidden-output" axis: since condensates *in vivo* are typically characterized by visualizing only a subset of hidden and/or output species, it is possible that (a) condensates that appear similar (by hidden species) could carry out distinct functions (by recruiting distinct output species that are not visualized), and vice versa (b) condensates that appear distinct (by hidden species) could still perform similar functions (by recruiting similar output species that are not visualized).

We show that increasing hidden species generally improves the precision of classification but eventually saturates. In principle, the physics underlying our multicomponent surface condensation model is flexible enough to universally approximate arbitrary decision boundaries through scaling the number of hidden nodes (SI Note 3, Fig. S3), but physical and/or numerical factors, such as those described below and in SI Note 3, are likely to constrain this flexibility. For example, the saturation in precision we observe could arise from a limitation in our optimization formulation, including in our choice of loss function or parameter constraints, that may make it difficult to find global minima of the loss landscape. Second, the saturation could point to a more fundamental limit that arises from two competing physical constraints in our model: (a) with more species, there is an overall dilution that increases the entropic cost of condensation, and (b) the requirement of liquid-like condensates, i.e., energy scales of order k_BT, limits the enthalpic stabilization that is possible to encode in our simple model of pairwise interactions. While not captured in our simple thermodynamic model, biology points to the need for more complex models that may expand the scope of computations possible through condensation—including through leveraging higher-order interactions such as discrete sticker-spacers or excluded-volume interactions that expand capacity of the underlying free-energy landscape (17, 67, 68), multimerization domains that function as sinks to reduce entropic costs of demixing (69), and more generally, out-of-equilibrium reaction cascades that provide additional axes for tunable multiphase behavior.

The balance of entropy-energy trade-offs direct surfaces with differing input compositions to recruit distinct condensates and behave as a classifier. Our model has partial parallels to well-known architectures in inference—for example, the free energy governing phase behavior in our model resembles that of a Hopfield network (70). Our model more closely resembles Boltzmann machines (53) in that we exploit hidden species to encode more complex stimuli-response behaviors, i.e., higher expressivity (49, 71). While we focus on classification, emerging studies argue for broader computational capabilities embedded in multicomponent liquids. For example, (40) explore the capacity of condensates to store and retrieve patterns as stable phases (or memories) analogous to

Hopfield models trained with the Hebbian rule. More recently, (41) use a simple wake-sleep learning algorithm, based on competition between Hebbian learning and anti-Hebbian unlearning as in classical Boltzmann machines (53), to train molecular parameters of 3D lattice liquids to form complex spatial architectures and to perform general probabilistic inference, including for MNIST digit classification. Further, our work finds that MNIST classification saturates ($\sim 85\%$)—potentially hinting at limitations in the physics of condensation and/or in the choice of data encoding/representation. More generally, it would be valuable to delineate and contrast the principles and limits of computations performed by different physical systems with and in addition to condensation—for example, dimerization networks (4), self-assembly (8–10, 72), mechanical systems (73, 74), and stochastic biomolecular reactions (75–77).

We characterize the computational capabilities of programmed multiphase fluids that are trained through gradient-descent based global optimization routines. While we focus on classifier function, the ways by which molecular networks can learn, potentially autonomously, or be trained represents an important area for investigation. For instance, (41) use sleep-wake training rules that are spatially local to train lattice liquids. In conjunction with other recent studies, these suggest that molecular networks can be trained *in situ* through physical learning rules that directly modify parameters like interactions or composition (41, 58, 78, 79). In particular, we show that only changing levels of hidden and output species in the reservoir—a variable amenable to modification in living systems—of trained fluids enables adaptation to new tasks (Fig. 5A). If the levels of reservoir species could be directly regulated by condensate formation—for example through engineered genetic feedback circuits where condensation of output species alters gene expression of reservoir species—this would permit learning over longer time-scales. Together, these hint at biologically plausible mechanisms for autonomous and continual learning in biomolecular fluids without any electrical computers in the loop.

Finally, we demonstrate concordance of our mean-field designs with function in a 3D lattice model that explicitly captures spatial correlations that are relevant *in vitro*—motivating opportunities for experimental testing and refinement. Promising avenues include (a) designed DNA (10, 42, 80) molecules, along with surface-functionalized or immobilized DNA strands, and (b) emerging synthetic biology approaches (81–84) that combine genetic reporter systems with coexpression of phase-separation proteins. More generally, the confluence of machine-learning, physics-based models, and multiplexed experimental techniques will inform future opportunities to dissect as well as design biological computation through condensation.

Limitations of the study

In this paper, we explore how the emergent physics underlying condensation in multicomponent liquids can classify surfaces with distinct compositions. Towards this, we introduce a simple mean-field description of liquids comprising molecules of identical size with pairwise interactions. As a consequence, we are unable to explore the computational capacity and constraints that are afforded through explicit consideration of complex molecules - including effects of polydispersity, higher-order interactions, and anisotropic molecular architectures that all typify biomolecules. We focus on mean-field surface condensation from a large (infinite) cellular reservoir that we posit maintains any learned chemical potential. Thus, a limitation of the model is that molecules are maintained at well-mixed compositions inside the surface and intra-surface demixing is not explicitly studied. Thus, further studies are required to explicitly study the effects of finite copy numbers, surface-surface competition, and dynamics of nucleation. Since our model does not explicitly specify the mechanisms by which the reservoir can be maintained, either in biological or physical systems, new models that explicitly consider specific reservoir models will provide insights on how to realize them.

Acknowledgments

We thank Francesco Mottes, Ryan Krueger, Mason Rouches, William Jacobs, Suriyanarayanan Vaikunthanathan, David Zwicker, Serena Carra, and members of the Murugan, Brenner, Winfree, and Shrinivas labs for helpful discussions on this manuscript. E.V.H. acknowledges support from NSF NRT 2021900 - Synthesizing Biology Across Scales. A.Z. and M.P.B. acknowledge support from NSF AI Institute of Dynamic Systems 2112085. K.S. acknowledges funding and support from Northwestern University. E.W. and C.C. acknowledge support from NSF CCF/FET 2008589 and 2212546. A.M acknowledges support from the National Science Foundation through the Center for Living Systems (grant no. 2317138) and DMR-2239801. K.S. acknowledges helpful discussions related to this paper at the Kavli Institute for Theoretical Physics (KITP) workshop: Physical Principles Shaping Biomolecular Condensates supported in part by NSF PHY-2309135 and the Gordon and Betty Moore Foundation Grant No. 2919.02. The computations in this paper were in part run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University. This research was supported in part through the computational resources and staff contributions provided for the Quest high performance computing facility at Northwestern University.

Code availability

We make all mean-field code available via the following GitHub repository: https://github.com/shrinivaslab/2025_zentner_multiphase_classification

Bibliography

1. Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. CRC Press, July 2019. ISBN 978-1-000-00132-7. Google-Books-ID: Lg3MDwAAQBAJ.
2. Leland H. Hartwell, John J. Hopfield, Stanislas Leibler, and Andrew W. Murray. From molecular to modular cell biology. *Nature*, 402(6761):C47–C52, December 1999. ISSN 1476-4687. doi: 10.1038/35011540. Publisher: Nature Publishing Group.
3. Wendell A. Lim, Connie M. Lee, and Chao Tang. Design Principles of Regulatory Networks: Searching for the Molecular Algorithms of the Cell. *Molecular Cell*, 49(2):202–212, January 2013. ISSN 1097-2765. doi: 10.1016/j.molcel.2012.12.020. Publisher: Elsevier.
4. Jacob Parres-Gold, Matthew Levine, Benjamin Emert, Andrew Stuart, and Michael B. Elowitz. Contextual computation by competitive protein dimerization networks. *Cell*, 188(7):1984–2002.e17, April 2025. ISSN 1097-4172. doi: 10.1016/j.cell.2025.01.036.
5. M. Conrad. Self-assembly as a mechanism of molecular computing. In *Images of the Twenty-First Century. Proceedings of the Annual International Engineering in Medicine and Biology Society,*, pages 1354–1355 vol.4, November 1989. doi: 10.1109/IEMBS.1989.96236.
6. Johann Elbaz, Oleg Lioubashevski, Fuan Wang, Françoise Remacle, Raphael D. Levine, and Itamar Willner. DNA computing circuits using libraries of DNAzyme subunits. *Nature Nanotechnology*, 5(6):417–422, June 2010. ISSN 1748-3395. doi: 10.1038/nnano.2010.88. Publisher: Nature Publishing Group.
7. Arvind Murugan, Zorana Zeravcic, Michael P. Brenner, and Stanislas Leibler. Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures. *Proceedings of the National Academy of Sciences*, 112(1):54–59, January 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1413941112.
8. Weishun Zhong, David J. Schwab, and Arvind Murugan. Associative Pattern Recognition Through Macro-molecular Self-Assembly. *Journal of Statistical Physics*, 167(3):806–826, May 2017. ISSN 1572-9613. doi: 10.1007/s10955-017-1774-2.
9. Damien Woods, David Doty, Cameron Myhrvold, Joy Hui, Felix Zhou, Peng Yin, and Erik Winfree. Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. *Nature*, 567(7748):366–372, March 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1014-9. Publisher: Nature Publishing Group.
10. Constantine Glen Evans, Jackson O'Brien, Erik Winfree, and Arvind Murugan. Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly. *Nature*, 625(7995):500–507, January 2024. ISSN 1476-4687. doi: 10.1038/s41586-023-06890-z. Publisher: Nature Publishing Group.
11. Jeff Hasty, David McMillen, and J. J. Collins. Engineered gene circuits. *Nature*, 420(6912):224–230, November 2002. ISSN 1476-4687. doi: 10.1038/nature01257. Publisher: Nature Publishing Group.
12. Timothy Frei, Ching-Hsiang Chang, Maurice Filo, Asterios Arampatzis, and Mustafa Khammash. A genetic mammalian proportional–integral feedback control circuit for robust and precise gene regulation. *Proceedings of the National Academy of Sciences*, 119(24):e2122132119, June 2022. doi: 10.1073/pnas.2122132119. Publisher: Proceedings of the National Academy of Sciences.
13. A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, and C. Zechner. Phase separation provides a mechanism to reduce noise in cells. *Science*, 367(6476):464–468, January 2020. doi: 10.1126/science.aav6691. Publisher: American Association for the Advancement of Science.
14. Salman F. Banani, Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews. Molecular Cell Biology*, 18(5):285–298, May 2017. ISSN 1471-0080. doi: 10.1038/nrm.2017.7. Publisher: Nature Publishing Group.
15. Yongdae Shin and Clifford P. Brangwynne. Liquid phase condensation in cell physiology and disease. *Science*, 357(6357):eaaf4382, September 2017. doi: 10.1126/science.aaf4382. Publisher: American Association for the Advancement of Science.
16. Andrew S. Lyon, William B. Peeples, and Michael K. Rosen. A framework for understanding the functions of biomolecular condensates across scales. *Nature Reviews Molecular Cell Biology*, pages 1–21, November 2020. ISSN 1471-0080. doi: 10.1038/s41580-020-00303-z. Publisher: Nature Publishing Group.
17. Jeong-Mo Choi, Alex S. Holehouse, and Rohit V. Pappu. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics*, 49(Volume 49, 2020):107–133, May 2020. ISSN 1936-122X, 1936-1238. doi: 10.1146/annurev-biophys-121219-081629. Publisher: Annual Reviews.
18. Frank Jülicher and Christoph A. Weber. Droplet Physics and Intracellular Phase Separation. *Annual Review of Condensed Matter Physics*, 15 (Volume 15, 2024):237–261, March 2024. ISSN 1947-5454, 1947-5462. doi: 10.1146/annurev-conmatphys-031720-032917. Publisher: Annual Reviews.
19. Denes Hnisz, Krishna Shrinivas, Richard A. Young, Arup K. Chakraborty, and Phillip A. Sharp. A Phase Separation Model for Transcriptional Control. *Cell*, 169(1):13–23, March 2017. ISSN 0092-8674. doi: 10.1016/j.cell.2017.02.007.
20. Krishna Shrinivas, Benjamin R. Sabari, Eliot L. Coffey, Isaac A. Klein, Ann Boija, Alicia V. Zamudio, Jurian Schuijers, Nancy M. Hannett, Phillip A. Sharp, Richard A. Young, and Arup K. Chakraborty. Enhancer Features that Drive Formation of Transcriptional Condensates. *Molecular Cell*, 75(3):549–561.e7, August 2019. ISSN 1097-2765. doi: 10.1016/j.molcel.2019.07.009. Publisher: Elsevier.
21. Woonyung Hur, James P. Kemp, Marco Tarzia, Victoria E. Deneke, William F. Marzluff, Robert J. Duronio, and Stefano Di Talia. CDK-Regulated Phase Separation Seeded by Histone Genes Ensures Precise Growth and Function of Histone Locus Bodies. *Developmental Cell*, 54(3):379–394.e6, August 2020. ISSN 1534-5807. doi: 10.1016/j.devcel.2020.06.003.
22. Jose A. Morin, Sina Wittmann, Sandeep Choubey, Adam Klosin, Stefan Golfier, Anthony A. Hyman, Frank Jülicher, and Stephan W. Grill. Sequence-dependent surface condensation of a pioneer transcription factor on DNA. *Nature Physics*, 18(3):271–276, March 2022. ISSN 1745-2481. doi: 10.1038/s41567-021-01462-2. Publisher: Nature Publishing Group.
23. Agnieszka Pancholi, Tim Klingberg, Weichun Zhang, Roshan Prizak, Irina Mamontova, Amra Noa, Marcel Sobucki, Andrei Yu Kobitski, Gerd Ulrich Nienhaus, Vasily Zaburdaev, and Lennart Hilbert. RNA polymerase II clusters form in line with surface condensation on regulatory chromatin. *Molecular Systems Biology*, 17(9):e10272, 2021. doi: https://doi.org/10.15252/msb.202110272. _eprint: https://www.embopress.org/doi/pdf/10.15252/msb.202110272.
24. Charles S. Bond and Archa H. Fox. Paraspeckles: nuclear bodies built on long noncoding RNA. *The Journal of Cell Biology*, 186(5):637–644, September 2009. ISSN 0021-9525. doi: 10.1083/jcb.200906113.
25. Wilton T. Snead, Mary K. Skillicorn, Krishna Shrinivas, and Amy S. Gladfelter. Immiscible proteins compete for RNA binding to order condensate layers. *Proceedings of the National Academy of Sciences*, 122(32):e2504778122, August 2025. doi: 10.1073/pnas.2504778122. Publisher: Proceedings of the National Academy of Sciences.
26. Lindsay B. Case, Jonathon A. Ditlev, and Michael K. Rosen. Regulation of Transmembrane Signaling by Phase Separation. *Annual Review of Biophysics*, 48(1):465–494, 2019. doi: 10.1146/annurev-biophys-052118-115534. _eprint: https://doi.org/10.1146/annurev-biophys-052118-115534.
27. Yan G. Zhao and Hong Zhang. Phase Separation in Membrane Biology: The Interplay between Membrane-Bound Organelles and Membraneless Condensates. *Developmental Cell*, 55(1):30–44, 2020. ISSN 18781551. doi: 10.1016/j.devcel.2020.06.033. Publisher: Elsevier Inc.
28. Wilton T. Snead. Condensate-membrane interactions shape membranes, tune cytoskeletal assembly, and localize mRNAs. *Current Opinion in Cell Biology*, 95:102540, August 2025. ISSN 0955-0674. doi: 10.1016/j.ceb.2025.102540.
29. Richard P. Sear and José A. Cuesta. Instabilities in Complex Mixtures with a Large Number of Components. *Physical Review Letters*, 91(24):245701, December 2003. doi: 10.1103/PhysRevLett.91.245701. Publisher: American Physical Society.
30. Peter Sollich. Predicting phase equilibria in polydisperse systems. *Journal of Physics: Condensed Matter*, 14(3):R79, December 2001. ISSN 0953-8984. doi: 10.1088/0953-8984/14/3/201.
31. William M. Jacobs and Daan Frenkel. Predicting phase behavior in multicomponent mixtures. *Journal of Chemical Physics*, 139(2):024108, July 2013. ISSN 00219606. doi: 10.1063/1.4812461. Publisher: American Institute of PhysicsAIP.
32. William M. Jacobs. Self-Assembly of Biomolecular Condensates with Shared Components. *Physical Review Letters*, 126(25):258101, June 2021. doi: 10.1103/PhysRevLett.126.258101. Publisher: American Physical Society.
33. Fan Chen and William M. Jacobs. Programmable phase behavior in fluids with designable interactions. *The Journal of Chemical Physics*, 158(21):214118, June 2023. ISSN 0021-9606. doi: 10.1063/5.0147211.
34. David Zwicker and Liedewij Laan. Evolved interactions stabilize many coexisting phases in multicomponent liquids. *Proceedings of the National Academy of Sciences*, 119(28):e2201250119, July 2022. doi: 10.1073/pnas.2201250119. Publisher: Proceedings of the National Academy of Sciences.
35. Krishna Shrinivas and Michael P. Brenner. Multiphase coexistence capacity in complex fluids, October 2022. Pages: 2022.10.19.512909 Section: New Results.
36. Krishna Shrinivas and Michael P. Brenner. Phase separation in fluids with many interacting components. *Proceedings of the National Academy of Sciences*, 118(45), November 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2108551118. Publisher: National Academy of Sciences Section: Biological Sciences.
37. Yicheng Qiang, Chengjie Luo, and David Zwicker. Scaling of phase

count in multicomponent liquids, May 2024. arXiv:2405.01138 [cond-mat, physics:physics].
38. Mason Rouches, Sarah L. Veatch, and Benjamin B. Machta. Surface densities prewet a near-critical membrane. *Proceedings of the National Academy of Sciences*, 118(40), October 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2103401118. Publisher: National Academy of Sciences Section: Physical Sciences.
39. Sheng Mao, Derek Kuldinow, Mikko P. Haataja, and Andrej Košmrlj. Phase behavior and morphology of multicomponent liquid mixtures. *Soft Matter*, 15 (6):1297–1311, 2019. doi: 10.1039/C8SM02045K. Publisher: Royal Society of Chemistry.
40. Rodrigo Braz Teixeira, Giorgio Carugno, Izaak Neri, and Pablo Sartori. Liquid Hopfield model: Retrieval and localization in multicomponent liquid mixtures. *Proceedings of the National Academy of Sciences*, 121(48):e2320504121, November 2024. doi: 10.1073/pnas.2320504121. Publisher: Proceedings of the National Academy of Sciences.
41. Cameron Chalk, Salvador Buse, Krishna Shrinivas, Arvind Murugan, and Erik Winfree. Learning and Inference in a Lattice Model of Multicomponent Condensates. In *DROPS-IDN/v2/document/10.4230/LIPIcs.DNA.30.5*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi: 10.4230/LIPIcs.DNA.30.5.
42. Gabrielle R Abraham, Aria S Chaderjian, Anna B N Nguyen, Sam Wilken, and Omar A Saleh. Nucleic acid liquids. *Reports on Progress in Physics*, 87(6):066601, May 2024. ISSN 0034-4885. doi: 10.1088/1361-6633/ad4662. Publisher: IOP Publishing.
43. Sungho Do, Chanseok Lee, Taehyun Lee, Do-Nyun Kim, and Yongdae Shin. Engineering DNA-based synthetic condensates with programmable material properties, compositions, and functionalities. *Science Advances*, 8(41): eabj1771, October 2022. doi: 10.1126/sciadv.abj1771. Publisher: American Association for the Advancement of Science.
44. Jaimie Marie Stewart, Shiyi Li, Anli A. Tang, Melissa Ann Klocke, Martin Vincent Gobry, Giacomo Fabrini, Lorenzo Di Michele, Paul W. K. Rothemund, and Elisa Franco. Modular RNA motifs for orthogonal phase separated compartments. *Nature Communications*, 15(1):6244, July 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-50003-x. Publisher: Nature Publishing Group.
45. Giacomo Fabrini, Nada Farag, Sabrina Pia Nuccio, Shiyi Li, Jaimie Marie Stewart, Anli A. Tang, Reece McCoy, Róisín M. Owens, Paul W. K. Rothemund, Elisa Franco, Marco Di Antonio, and Lorenzo Di Michele. Co-transcriptional production of programmable RNA condensates and synthetic organelles. *Nature Nanotechnology*, 19(11):1665–1673, November 2024. ISSN 1748-3395. doi: 10.1038/s41565-024-01726-x. Publisher: Nature Publishing Group.
46. Jing Gong, Nozomi Tsumura, Yusuke Sato, and Masahiro Takinoue. Computational DNA Droplets Recognizing miRNA Sequence Inputs Based on Liquid–Liquid Phase Separation. *Advanced Functional Materials*, 32(37): 2202322, 2022. ISSN 1616-3028. doi: 10.1002/adfm.202202322. _eprint: https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adfm.202202322.
47. Masahiro Takinoue. DNA droplets for intelligent and dynamical artificial cells: from the viewpoint of computation and non-equilibrium systems. *Interface Focus*, 13(5):20230021, August 2023. doi: 10.1098/rsfs.2023.0021. Publisher: Royal Society.
48. Hirotake Udono, Jing Gong, Yusuke Sato, and Masahiro Takinoue. DNA Droplets: Intelligent, Dynamic Fluid. *Advanced Biology*, 7(3):2200180, 2023. ISSN 2701-0198. doi: 10.1002/adbi.202200180. _eprint: https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adbi.202200180.
49. Arvind Murugan, David Zwicker, Charlotta Lorenz, and Eric R. Dufresne. Could Living Cells Use Phase Transitions to Process Information? *arXiv*, July 2025. doi: 10.48550/arXiv.2507.23384. arXiv:2507.23384 [physics].
50. P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, 49(3):435–479, July 1977. doi: 10.1103/RevModPhys.49.435. Publisher: American Physical Society.
51. James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs. 2018. URL: http://github.com/google/jax.
52. DeepMind, Igor Babuschkin, Kate Baumli, Alison Bell, Surya Bhupatiraju, Jake Bruce, Peter Buchlovsky, David Budden, Trevor Cai, Aidan Clark, Ivo Danihelka, Antoine Dedieu, Claudio Fantacci, Jonathan Godwin, Chris Jones, Ross Hemsley, Tom Hennigan, Matteo Hessel, Shaobo Hou, Steven Kapturowski, Thomas Keck, Iurii Kemaev, Michael King, Markus Kunesch, Lena Martens, Hamza Merzic, Vladimir Mikulik, Tamara Norman, George Papamakarios, John Quan, Roman Ring, Francisco Ruiz, Alvaro Sanchez, Laurent Sartran, Rosalia Schneider, Eren Sezener, Stephen Spencer, Srivatsan Srinivasan, Miloš Stanojević, Wojciech Stokowiec, Luyu Wang, Guangyao Zhou, and Fabio Viola. The DeepMind JAX Ecosystem. 2020. URL: http://github.com/google-deepmind.
53. David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147–169, January 1985. ISSN 0364-0213. doi: 10.1016/S0364-0213(85)80012-4.
54. V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, April 1967. ISSN 0025-5734. doi: 10.1070/SM1967v001n04ABEH001994. Publisher: IOP Publishing.
55. Ali Rahimi and Benjamin Recht. Random Features for Large-Scale Kernel Machines. In *Advances in Neural Information Processing Systems*, volume 20, pages 1177–1184. Curran Associates, Inc., 2007.
56. Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. Recent advances in physical reservoir computing: A review. *Neural Networks*, 115:100–123, July 2019. ISSN 0893-6080. doi: 10.1016/j.neunet.2019.03.005.
57. Christina J. Su, Arvind Murugan, James M. Linton, Akshay Yeluri, Justin Bois, Heidi Klumpe, Matthew A. Langley, Yaron E. Antebi, and Michael B. Elowitz. Ligand-receptor promiscuity enables cellular addressing. *Cell Systems*, 13 (5):408–425.e12, May 2022. ISSN 2405-4712, 2405-4720. doi: 10.1016/j.cels.2022.03.001. Publisher: Elsevier.
58. Menachem Stern and Arvind Murugan. Learning Without Neurons in Physical Systems. *Annual Review of Condensed Matter Physics*, 14(Volume 14, 2023):417–441, March 2023. ISSN 1947-5454, 1947-5462. doi: 10.1146/annurev-conmatphys-040821-113439. Publisher: Annual Reviews.
59. Menachem Stern, Daniel Hexner, Jason W. Rocks, and Andrea J. Liu. Supervised Learning in Physical Networks: From Machine Learning to Learning Machines. *Physical Review X*, 11(2):021045, May 2021. doi: 10.1103/PhysRevX.11.021045. Publisher: American Physical Society.
60. Menachem Stern, Chukwunonso Arinze, Leron Perez, Stephanie E. Palmer, and Arvind Murugan. Supervised learning through physical changes in a mechanical system. *Proceedings of the National Academy of Sciences*, 117(26): 14843–14850, June 2020. doi: 10.1073/pnas.2000807117. Publisher: Proceedings of the National Academy of Sciences.
61. Kevin M. Cherry and Lulu Qian. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature*, 559(7714):370–376, July 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0289-6. Publisher: Nature Publishing Group.
62. William M Jacobs and Daan Frenkel. Phase transitions in biological systems with many components. *Biophysical Journal*, 112(4):683–691, 2017.
63. Vijay V. Vazirani. *Approximation Algorithms*. Springer Science & Business Media, March 2013. ISBN 978-3-662-04565-7. Google-Books-ID: bJmqCAAAQBAJ.
64. Gaofeng Pei, Heankel Lyons, Pilong Li, and Benjamin R. Sabari. Transcription regulation by biomolecular condensates. *Nature Reviews Molecular Cell Biology*, 26(3):213–236, March 2025. ISSN 1471-0080. doi: 10.1038/s41580-024-00789-x. Publisher: Nature Publishing Group.
65. Yaron E. Antebi, James M. Linton, Heidi Klumpe, Bogdan Bintu, Mengsha Gong, Christina Su, Reed McCardell, and Michael B. Elowitz. Combinatorial Signal Perception in the BMP Pathway. *Cell*, 170(6):1184–1196.e24, September 2017. ISSN 10974172. doi: 10.1016/j.cell.2017.08.015. Publisher: Cell Press.
66. Heidi E. Klumpe, Matthew A. Langley, James M. Linton, Christina J. Su, Yaron E. Antebi, and Michael B. Elowitz. The context-dependent, combinatorial logic of BMP signaling. *Cell Systems*, 13(5):388–407.e10, May 2022. ISSN 2405-4712. doi: 10.1016/j.cels.2022.03.002.
67. Kamal Bhandari, Michael A. Cotten, Jonggul Kim, Michael K. Rosen, and Jeremy D. Schmit. Structure–Function Properties in Disordered Condensates. *The Journal of Physical Chemistry B*, 125(1):467–476, January 2021. ISSN 1520-6106. doi: 10.1021/acs.jpcb.0c11057. Publisher: American Chemical Society.
68. Marco Cappa, Francesco Sciortino, and Lorenzo Rovigatti. A phase-field model for solutions of DNA-made particles. *The Journal of Chemical Physics*, 162(19):194901, May 2025. ISSN 0021-9606. doi: 10.1063/5.0257265.
69. Ushnish Rana, Ke Xu, Amal Narayanan, Mackenzie T. Walls, Athanassios Z. Panagiotopoulos, José L. Avalos, and Clifford P. Brangwynne. Asymmetric oligomerization state and sequence patterning can tune multiphase condensate miscibility. *Nature Chemistry*, 16(7):1073–1082, July 2024. ISSN 1755-4349. doi: 10.1038/s41557-024-01456-6. Publisher: Nature Publishing Group.
70. J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79 (8):2554–2558, April 1982. doi: 10.1073/pnas.79.8.2554. Publisher: Proceedings of the National Academy of Sciences.
71. Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the Expressive Power of Deep Neural Networks. *arXiv*, June 2017. doi: 10.48550/arXiv.1606.05336. also appears in: Proceedings of the 34th International Conference on Machine Learning.

72. Samuel L. Foley and Margaret E. Johnson. Membrane-associated self-assembly for cellular decision making. *arXiv*, 2025.
73. Daniel A Fletcher and R Dyche Mullins. Cell mechanics and the cytoskeleton. *Nature*, 463(7280):485–492, 2010.
74. Hiromi Yasuda, Philip R. Buskohl, Andrew Gillman, Todd D. Murphey, Susan Stepney, Richard A. Vaia, and Jordan R. Raney. Mechanical computing. *Nature*, 598(7879):39–48, October 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03623-y. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7879 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Materials for devices;Mechanical engineering;Mechanical properties;Structural materials Subject_term_id: materials-for-devices;mechanical-engineering;mechanical-properties;structural-materials.
75. Ho-Lin Chen, David Doty, and David Soloveichik. Deterministic function computation with chemical reaction networks. *Natural Computing*, 13(4):517–534, 2014.
76. David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, 2008.
77. Carlos Floyd, Aaron R Dinner, Arvind Murugan, and Suriyanarayanan Vaikuntanathan. Limits on the computational expressivity of non-equilibrium biophysical processes. *Nature Communications*, 16(1):7184, 2025.
78. William Poole, Andrés Ortiz-Muñoz, Abhishek Behera, Nick S. Jones, Thomas E. Ouldridge, Erik Winfree, and Manoj Gopalkrishnan. Chemical Boltzmann Machines. In Robert Brijder and Lulu Qian, editors, *DNA Computing and Molecular Programming*, pages 210–231, Cham, 2017. Springer International Publishing. ISBN 978-3-319-66799-7. doi: 10.1007/978-3-319-66799-7_14.
79. William Poole, Thomas Ouldridge, Manoj Gopalkrishnan, and Erik Winfree. Detailed Balanced Chemical Reaction Networks as Generalized Boltzmann Machines. *arXiv*, May 2022. doi: 10.48550/arXiv.2205.06313. arXiv:2205.06313 [q-bio].
80. William M. Jacobs and W. Benjamin Rogers. Assembly of Complex Colloidal Systems Using DNA. *Annual Review of Condensed Matter Physics*, 16 (Volume 16, 2025):443–463, March 2025. ISSN 1947-5454, 1947-5462. doi: 10.1146/annurev-conmatphys-032922-113138. Publisher: Annual Reviews.
81. Yifan Dai, Lingchong You, and Ashutosh Chilkoti. Engineering synthetic biomolecular condensates. *Nature Reviews Bioengineering*, 1(7):466–480, July 2023. ISSN 2731-6092. doi: 10.1038/s44222-023-00052-6. Publisher: Nature Publishing Group.
82. Zhi-Gang Qian, Sheng-Chen Huang, and Xiao-Xia Xia. Synthetic protein condensates for cellular and metabolic engineering. *Nature Chemical Biology*, 18(12):1330–1340, December 2022. ISSN 1552-4469. doi: 10.1038/s41589-022-01203-3. Publisher: Nature Publishing Group.
83. Ming-Tzo Wei, Yi-Che Chang, Shunsuke F. Shimobayashi, Yongdae Shin, Amy R. Strom, and Clifford P. Brangwynne. Nucleated transcriptional condensates amplify gene expression. *Nature Cell Biology*, 22(10):1187–1196, October 2020. ISSN 1476-4679. doi: 10.1038/s41556-020-00578-6. Publisher: Nature Publishing Group.
84. Heankel Lyons, Reshma T. Veettil, Prashant Pradhan, Christy Fornero, Nancy De La Cruz, Keiichi Ito, Mikayla Eppert, Robert G. Roeder, and Benjamin R. Sabari. Functional partitioning of transcriptional regulators by patterned charge blocks. *Cell*, 186(2):327–345.e28, January 2023. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2022.12.013. Publisher: Elsevier.
85. Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, January 1989. ISSN 0893-6080. doi: 10.1016/0893-6080(89)90020-8.
86. Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley & Sons, November 2012. ISBN 978-1-118-58600-6. Google-Books-ID: Br33IRC3PkQC.
87. Franz Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, September 1991. ISSN 0360-0300. doi: 10.1145/116873.116880.
88. R. O. Duda and H. Fossum. Pattern Classification by Iteratively Determined Linear and Piecewise Linear Discriminant Functions. *IEEE Transactions on Electronic Computers*, EC-15(2):220–232, April 1966. ISSN 0367-7508. doi: 10.1109/PGEC.1966.264302.
89. Jongmin Kim, John Hopfield, and Erik Winfree. Neural Network Computation by In Vitro Transcriptional Circuits. In *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004.
90. Zibo Chen, James M. Linton, Shiyu Xia, Xinwen Fan, Dingchen Yu, Jinglin Wang, Ronghui Zhu, and Michael B. Elowitz. A synthetic protein-level neural network in mammalian cells. *Science*, 386(6727):1243–1250, December 2024. doi: 10.1126/science.add8468. Publisher: American Association for the Advancement of Science.
91. Anthony J. Genot, Teruo Fujii, and Yannick Rondelez. Computing with Competition in Biochemical Networks. *Physical Review Letters*, 109(20):208102, November 2012. doi: 10.1103/PhysRevLett.109.208102. Publisher: American Physical Society.
92. Anthony J. Genot, Teruo Fujii, and Yannick Rondelez. Scaling down DNA circuits with competitive neural networks. *Journal of The Royal Society Interface*, 10(85):20130212, August 2013. doi: 10.1098/rsif.2013.0212. Publisher: Royal Society.
93. Jared A. M. Bard and D. Allan Drummond. Chaperone regulation of biomolecular condensates. *Frontiers in Biophysics*, 2, May 2024. ISSN 2813-7183. doi: 10.3389/frbis.2024.1342506. Publisher: Frontiers.
94. Arjun Narayanan, Anatoli Meriin, J Owen Andrews, Jan-Hendrik Spille, Michael Y Sherman, and Ibrahim I Cisse. A first order phase transition mechanism underlies protein aggregation in mammalian cells. *eLife*, 8:e39695, February 2019. ISSN 2050-084X. doi: 10.7554/eLife.39695. Publisher: eLife Sciences Publications, Ltd.
95. Josiah Willard Gibbs. On the Equilibrium of Heterogeneous Substances. Other, 1879. Volume: 2.