



Combinatorial decision-making driven by multicomponent surface condensates

Aidan Zentner^a, Ethan V. Halingstad^{b,c,d} , Cameron Chalk^e, Michael P. Brenner^{a,f} , Arvind Murugan^{d,g}, Erik Winfree^e, and Krishna Shrinivas^{b,c,d,1}

Affiliations are included on p. 11.

Edited by Ned S. Wingreen, Princeton University, Princeton, NJ; received September 30, 2025; accepted May 18, 2026 by Editorial Board Member Paul Chaikin

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.

condensates | molecular computation | biophysics | self-organization

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 (19–24) or RNA scaffolds (25, 26), and signaling condensates that are membrane-localized (27–29). 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 (30–40). Leveraging this, recent theoretical (41, 42) and experimental (43–49) work highlights the potential of condensates to perform functions beyond compartmentalization, including computations. A key question is what capabilities and constraints govern the ability of multicomponent biomolecular fluids to perform computations through condensation.

Significance

In living cells, biomolecules self-organize into compartments called condensates that form by phase transitions, often around surfaces such as DNA and membranes. Typically viewed as concentrating mechanisms, we show that the physics underlying condensation can also encode decision-making. Specifically, we design multicomponent fluids that discriminate between subtly different surfaces by condensing distinct molecular compositions on each surface. For instance, genomic loci could nucleate activating or repressive condensates depending on DNA features, directly achieving gene-regulatory decisions through condensation. These fluids can sculpt complex decision boundaries, making them expressive, and can be reprogrammed to solve new tasks without altering designed molecular interactions, making them adaptable. Our results suggest an expanded view of condensates beyond compartmentalization, with direct roles in information-processing and computation.

The authors declare no competing interest.

This article is a PNAS Direct Submission. N.S.W. is a guest editor invited by the Editorial Board.

Copyright © 2026 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: krishna@northwestern.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2527873123/-/DCSupplemental>.

Published July 1, 2026.

In this paper, we explore how biomolecular fluids can perform classification by selectively assembling distinct condensates on different surfaces. 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 such as 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 concentrations 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, 27). 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 3D region that contains both the biological scaffold (such as a membrane or a DNA locus) and the neighborhood with which it interacts. In the model, species are partitioned into three types. First, each surface is characterized by the composition of input species that are localized to it; these species do not exchange with the reservoir. The reservoir consists of output and hidden species that, unlike input species, exchange freely with the surface.

We note two assumptions underlying our minimal formulation of the surface and reservoir (expanded upon in the next

section and *SI Appendix, Supplementary Note 7*). First, the physical structure of the scaffold is not explicitly considered. Instead, the effective volume V is simply defined as a region of scaffold with a well-mixed composition of input species. Second, we posit that the reservoir is maintained at a constant chemical potential, without describing its bulk behavior/composition explicitly. As a consequence of these assumptions, we do not model how a condensate adheres to or reshapes a lower-dimensional scaffold and cannot directly distinguish whether surface condensates undergo prewetting or wetting (22). We focus only on the composition of the condensate that forms at a well-mixed surface equilibrating with a reservoir.

With this model framework, 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 multiple types of condensates, i.e., multiple phases where each 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 ultrasensitive switches in the recruitment of different output molecules by designing phase boundaries in the space of input composition (42, 50) (Fig. 1B).

Model Formulation. Toward 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 v 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 $\phi_i = n_i v / V$. 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 $\vec{\phi} \equiv \vec{\phi}_{\text{in}} \circ \vec{\phi}_{\text{out}} \circ \vec{\phi}_{\text{h}}$, where $\circ$ indicates vector concatenation and

$$\phi_{\text{in}} = (\phi_{\text{in},1}, \dots, \phi_{\text{in},N_{\text{in}}}) \quad [1]$$

$$\phi_{\text{out}} = (\phi_{\text{out},1}, \dots, \phi_{\text{out},N_{\text{out}}}) \quad [2]$$

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

The subvectors denote the input, output, and hidden composition vectors, respectively, and the total volume fraction of (nonsolvent) 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 do not prescribe any specific model of the reservoir (*SI Appendix, Supplementary Note 7*) and assume that it can maintain output and hidden species at a constant chemical potential described by

$$\vec{\mu}_{\text{res}} = (\mu_{\text{out},1}^{\text{res}}, \dots, \mu_{\text{out},N_{\text{out}}}^{\text{res}}, \mu_{\text{h},1}^{\text{res}}, \dots, \mu_{\text{h},N_{\text{h}}}^{\text{res}}). \quad [4]$$

The nondimensionalized free energy density of such a surface is

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

where $\vec{\phi}_{\text{oh}} = \vec{\phi}_{\text{out}} \circ \vec{\phi}_{\text{h}}$ and $\beta = 1/k_B T$ 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. f is the internal free energy density of the surface,

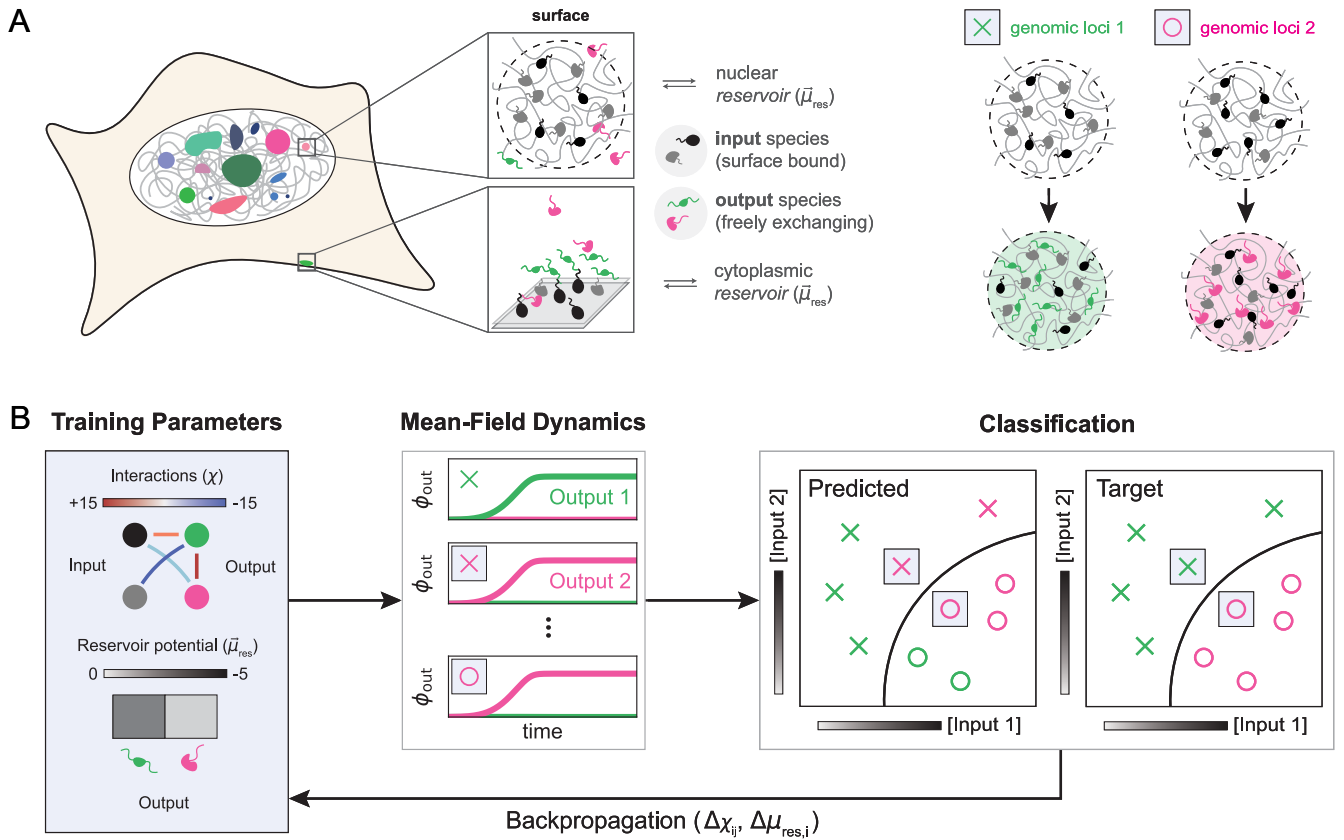


Fig. 1. (A) The model is motivated by cellular condensates that form on surfaces such as DNA and bilayers. Surface-localized species (e.g., transcription factors on DNA) are treated as input species (black, gray), while others freely exchange with a reservoir. Output species (green, pink) are exchanging molecules that drive downstream functions—for example, polymerases (DNA) or signaling activators (membranes). *Right* panel highlights this where 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 $\bar{\mu}_{\text{res}}$. (*Middle*) We consider the evolution of surfaces in the well-mixed, mean-field limit. (*Right*) The recruitment of distinct outputs is accomplished through forming surface-specific condensates. The phase boundary across which the composition abruptly changes functions as a decision boundary to classify surfaces. (Loop) We use JAX to iteratively tune χ and $\bar{\mu}_{\text{res}}$ with the goal of recruiting the desired output species (based on the phase label, x vs. o) for each training data point.

approximated in Flory–Huggins theory as

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

where χ is the effective interaction matrix,

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

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 Eq. 6. Further, we assume negligible effective solute–solvent interactions such that $\chi_{i0} = 0$ for all $i = 1, 2, \dots, N$, where 0 indexes the solvent (see *SI Appendix, Supplementary Note 1* for an extended discussion of this assumption and implications for experimental realization).

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 noninput

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 (51). Here, we assume that solvent molecules have much faster dynamics than solutes, which improves numerical stability of the optimization but does not affect the steady-state (*SI Appendix, Fig. S1*). Thus, the temporal evolution of the surface composition $\vec{\phi}_{\text{oh}}$ of the exchanging output and hidden species is written as (*SI Appendix, Supplementary Note 1*)

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

where $\beta \mu_i = \partial(\beta v f) / \partial \phi_{\text{oh},i}$ is the intrinsic (nondimensionalized) 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 (37) (*SI Appendix, Supplementary Note 1*). At steady state, the surface and reservoir must have identical chemical potentials in the noninput 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. Tuning these parameters allows for the construction of free energy landscapes for which the input concentration $\vec{\phi}_{\text{in}}$ of a surface determines the free energy minimum to which it relaxes. In this case, surfaces enriched in different output species are separated in the input space by a discontinuous phase boundary that behaves functionally as a decision boundary.

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 (52). We require that our loss function captures the following physical criteria at any surface. First, we require the desired output species to be enriched at concentrations exceeding $\phi_{\text{max}} = A/N$. Second, we require all undesired outputs to be depleted at concentrations below $\phi_{\text{min}} = B/N$. These enrichment and depletion criteria, when satisfied, enforces that the desired output is recruited at high concentrations and exceeds undesired outputs by a factor of at least A/B . We choose $A = 1.1$ (mild enrichment), and $B = 0.25$ (significant depletion).

Training the system entails considering a set of surfaces (or data points), each defined by an input concentration vector and a specification of the desired output species to be recruited. We find empirically that the following loss function gives the best performance in optimizing for the enrichment and depletion criteria:

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

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 a function that is monotonic in the extent that $\vec{\phi}_a$ fails to satisfy the enrichment/depletion criteria for a surface:

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

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

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

The term l_j is therefore at a global minimum when $p_j = q_k = 0$, and $\mathcal{L}$ is at a global minimum of 0 when this condition is satisfied for all data points a . Because the loss saturates once the thresholds are met, optimization only requires modest enrichment and depletion. However, two surfaces that are highly similar but lie on opposite sides of the decision boundary requires a sharp flip in the enriched output. The optimization therefore leverages the collective physics of phase separation, which produces discontinuous switches in composition across a phase boundary, and yields much larger enrichment ratios in practice.

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 (52, 53). Once trained, we evaluate the

performance of the classifier using the classification success S_c , defined as the fraction of surfaces whose steady states fully satisfy the enrichment/depletion criteria:

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

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 $k_B T$ and therefore enforce that each entry of the chi matrix has $|\chi_{ij}| < 15$, which is $\mathcal{O}(z = 18)$. Hence, from Eq. 7, the typical strength of an interaction energy is $|\epsilon_{ij}| \leq k_B T$ and comparable to thermal scales, as expected in liquid-like mixtures. 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 do not affect the steady-state 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 concentration (Fig. 2A), i.e., an ultrasensitive ratiometric sensor. With this objective in mind, we initialize a noninteracting 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 noncognate output. Upon testing, our model exhibits a sharp switch in composition across the boundary (Fig. 2C and SI Appendix, Fig. S2A). This switch is consistent with a first-order phase transition, exhibiting a discontinuous jump in the output composition across the phase boundary. As expected, this discontinuity disappears when interactions are weakened (SI Appendix, Fig. S2B).

To understand how the decision boundary emerges from molecular parameters, we develop a simple analytical approach (SI Appendix, Supplementary 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 (SI Appendix, Fig. S3A and Supplementary Note 3). More generally, our theory predicts that liquids with only input and output species can only typically encode linear decision boundaries in input space (SI Appendix, Supplementary Note 3). Consistent with this prediction, we find that our model still sharply classifies higher-dimensional linear manifolds (SI Appendix, Fig. S3B).

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

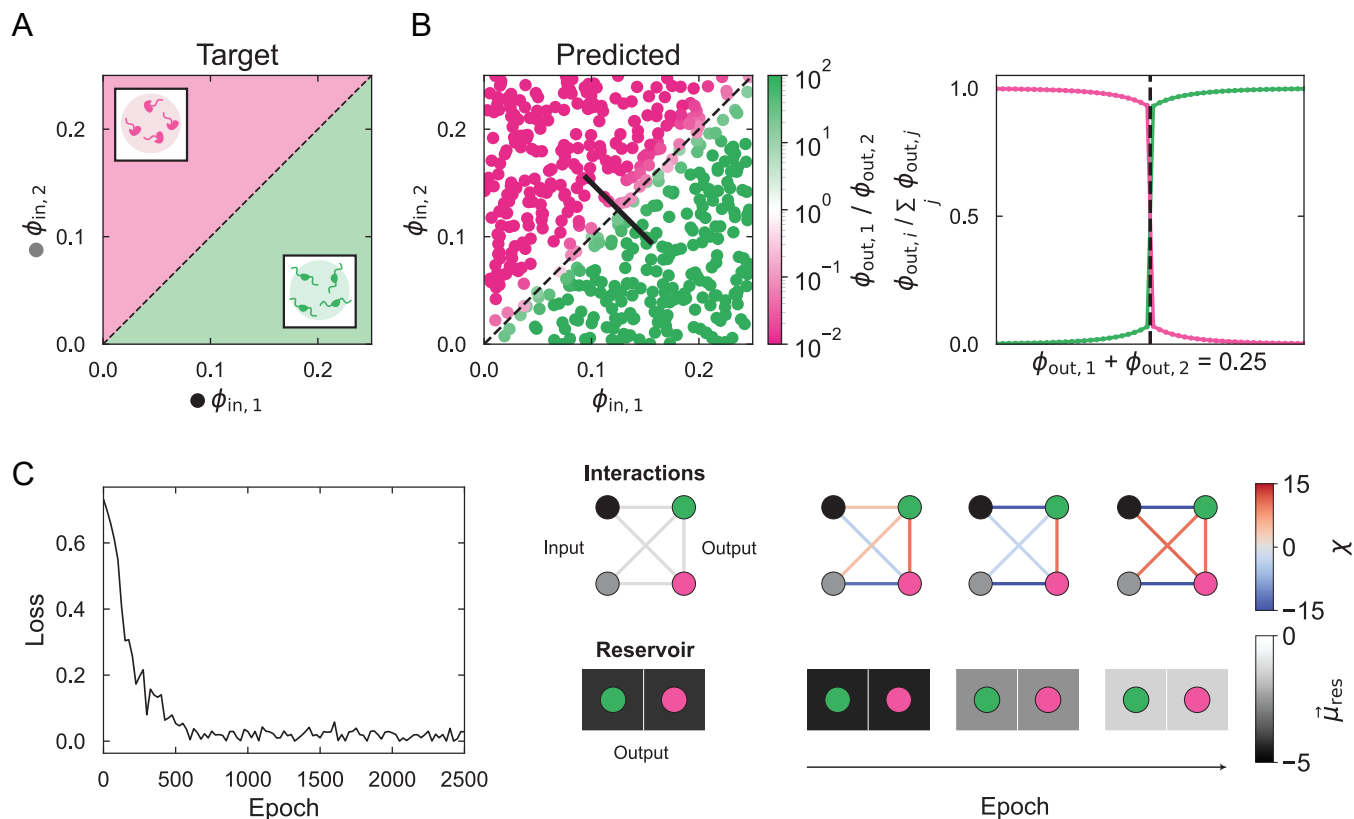


Fig. 2. (A) The target linear decision boundary is shown, with each axis being the concentration of one of the input species. Green and pink denote regions where we desire condensates enriched in the green and pink component, respectively. (B) Predictions from the trained model for different input compositions 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 Appendix, Supplementary Note 3*).

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 (*SI Appendix, Fig. S4A*).

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 (42, 54).

Extending our analytical approach, we find that the addition of a single hidden species offers enough flexibility to encode decision boundaries of varying curvatures (*SI Appendix, Fig. S3C and Supplementary Note 3*). We thus explore classification of complex, high-dimensional decision boundaries by including multiple hidden species (Fig. 3A).

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 essentially binary responses

in the output species recruited to the surface (Fig. 3C and *SI Appendix, 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, discontinuous switches in the recruited species across the boundary consistent with first-order phase transitions (*SI Appendix, Fig. S5*). The trained parameters for each decision boundary are shown in *SI Appendix, Fig. S6*.

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-

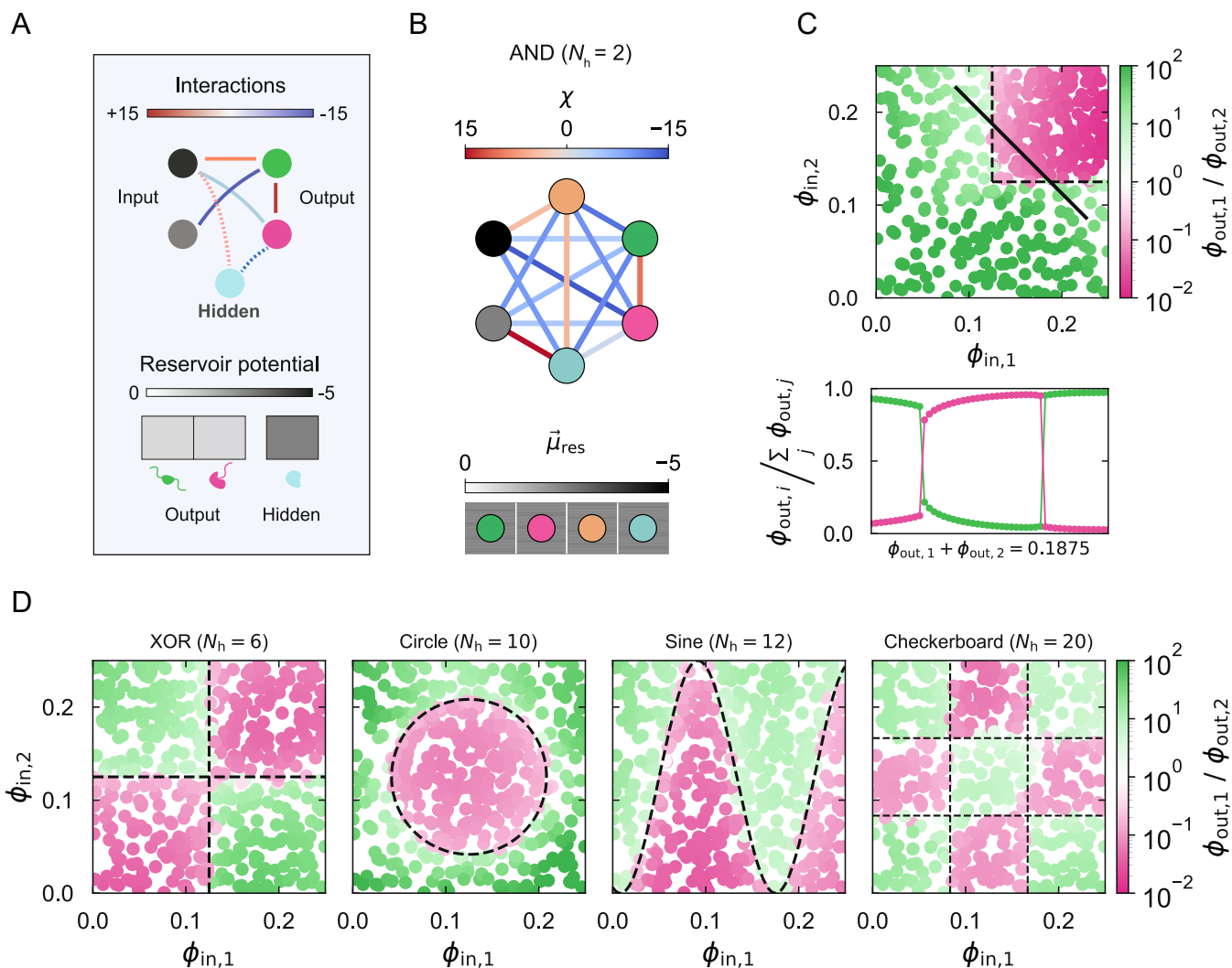


Fig. 3. (A) Hidden species, depicted in cyan in the interaction matrix and analogous to hidden nodes in Boltzmann machines, shape emergent overall phase behavior by interacting with input and output species but cannot directly drive output function. (B) Using only 2 hidden species (gold and cyan), we 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.

Pastur (55) 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 (SI Appendix, Supplementary Note 5). We find that for each decision boundary, the number of steady-state phases, averaged across a collection of molecular networks corresponding to different minima of the loss function, grows with hidden species (Fig. 4 A, Right). This suggests that encoding multiple phases plays an important role in improving the 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 phases 2 and 4) that recruit different hidden species but the same output species. The concentration vectors within a phase show very little variance, as evidenced by their high cosine similarity. 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 (SI Appendix, Figs. S7–S11). We next explore whether we can repurpose this modular multiphase representation learned by hidden species for other tasks.

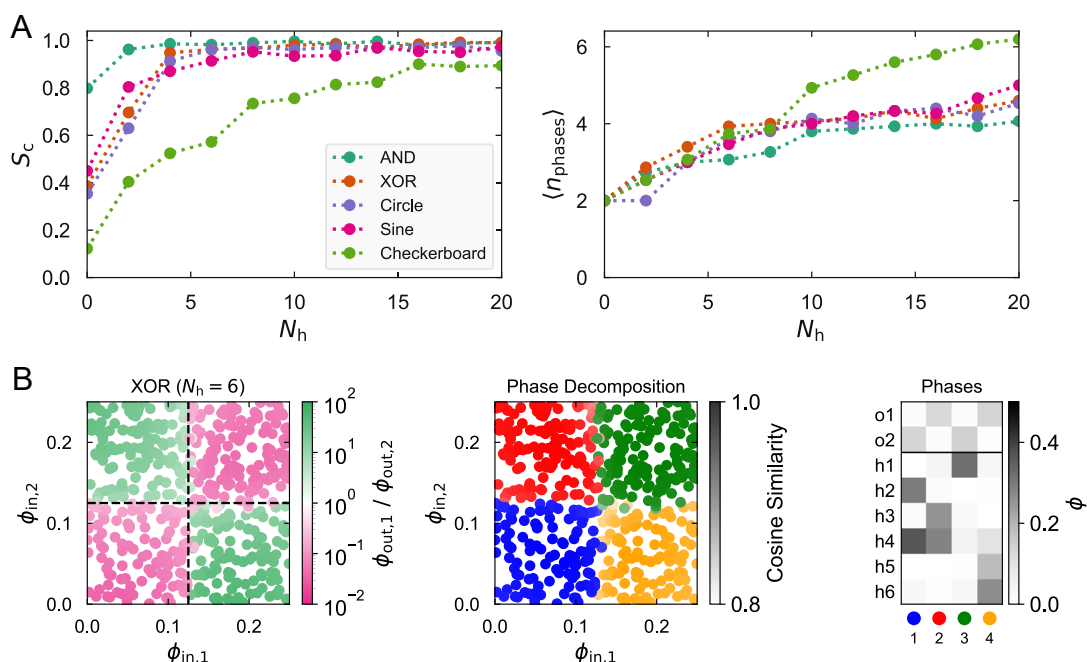


Fig. 4. (A) The scaling of the test set classification success S_c (Left) and the number of phases, averaged across molecular networks corresponding to different minima of the loss function (Right), as a function of the number of hidden species in the model. (B) When defined in terms of output species 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. The color of a point indicates the phase that the surface exhibits, while the color intensity of the point indicates the cosine similarity between the surface's phase vector and the mean concentration vector of all surfaces within the phase (Right). The compositions of the output and hidden species presented in the right-most panel represents the mean composition of surfaces that cluster into a phase (SI Appendix, Supplementary Note 5).

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 (56, 57). 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 finetune 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

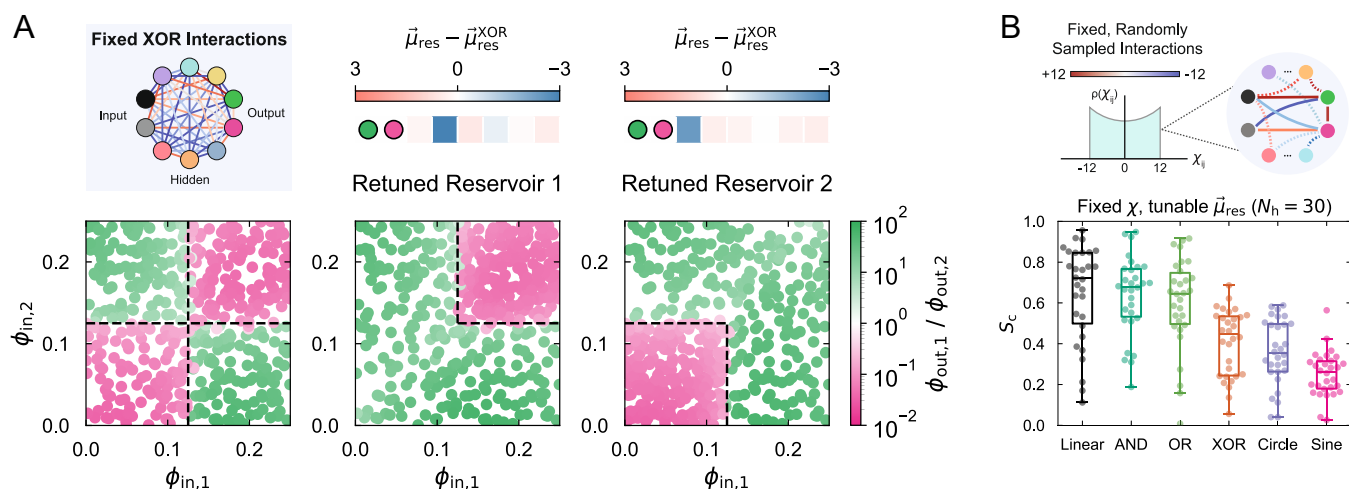


Fig. 5. (A) With the $N_h = 6$ mixture previously trained to solve a XOR boundary (Left), the same interaction network can also solve an AND decision boundary (Middle) and OR decision boundary (Right) by selectively tuning the reservoir potentials. (B) The capacity for classification by liquids with random interaction networks. (Left) A random interaction matrix with $N_h = 30$ is sampled from a probability distribution (SI Appendix, Supplementary 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.

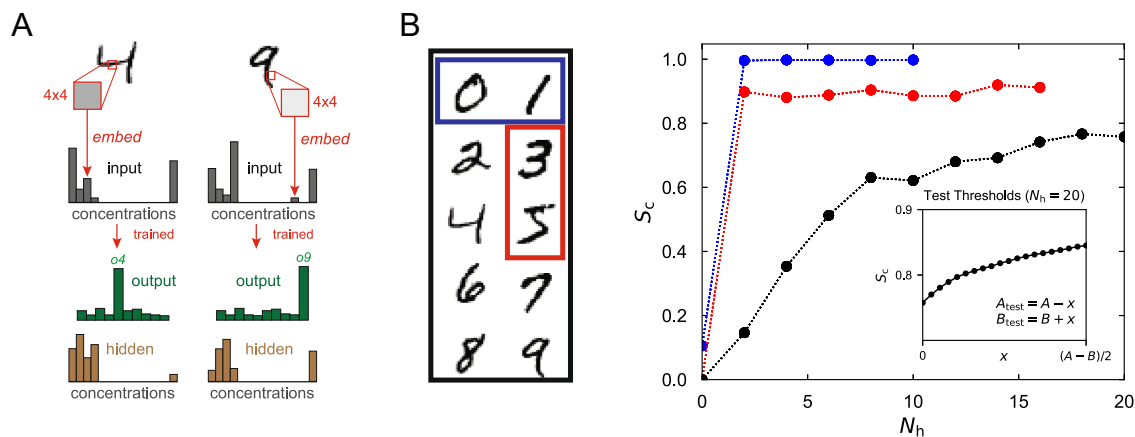


Fig. 6. (A) Solving classification of the MNIST dataset involves embedding pixel grayscale values into many input species concentrations, then training interaction parameters and reservoir concentrations for the formation of condensates enriched in 1 of 10 output species. (B) Classification success S_C 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 to 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 $B_{\text{test}} = B + x$, 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 success is approximately $\sim 85\%$.

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 Appendix, Supplementary 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 Appendix, Supplementary 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 the context of BMP signaling and dimerization networks (4, 58) 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 understand how cells might process chemical stimuli 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 (59), ranging from molecular concentrations to mechanical forces (60, 61). Molecular examples include winner-take-all reaction networks (62), self-assembly with Hebbian-like interactions (8, 10), and multicomponent liquids (41, 42). 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 (*SI Appendix, Fig. S12*).

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 $\phi_{in,i}$ of an input species to its corresponding pixel value (x_i) by $\phi_{in,i} = \phi^0 (x_i/255)$, where $\phi^0 = 0.5/49 \approx 0.01$. We train the mixture to initially discriminate between two digits, achieving strong performance with just 2 hidden nodes. 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 $\sim 85\%$. Generally, the ability to design condensates with large numbers of species for high-dimensional classification is improved with hidden nodes but typically saturates.

These results are generally consistent with recent findings from ref. 42, where the authors develop a 3D lattice condensate model and train it with a probabilistic learning algorithm derived

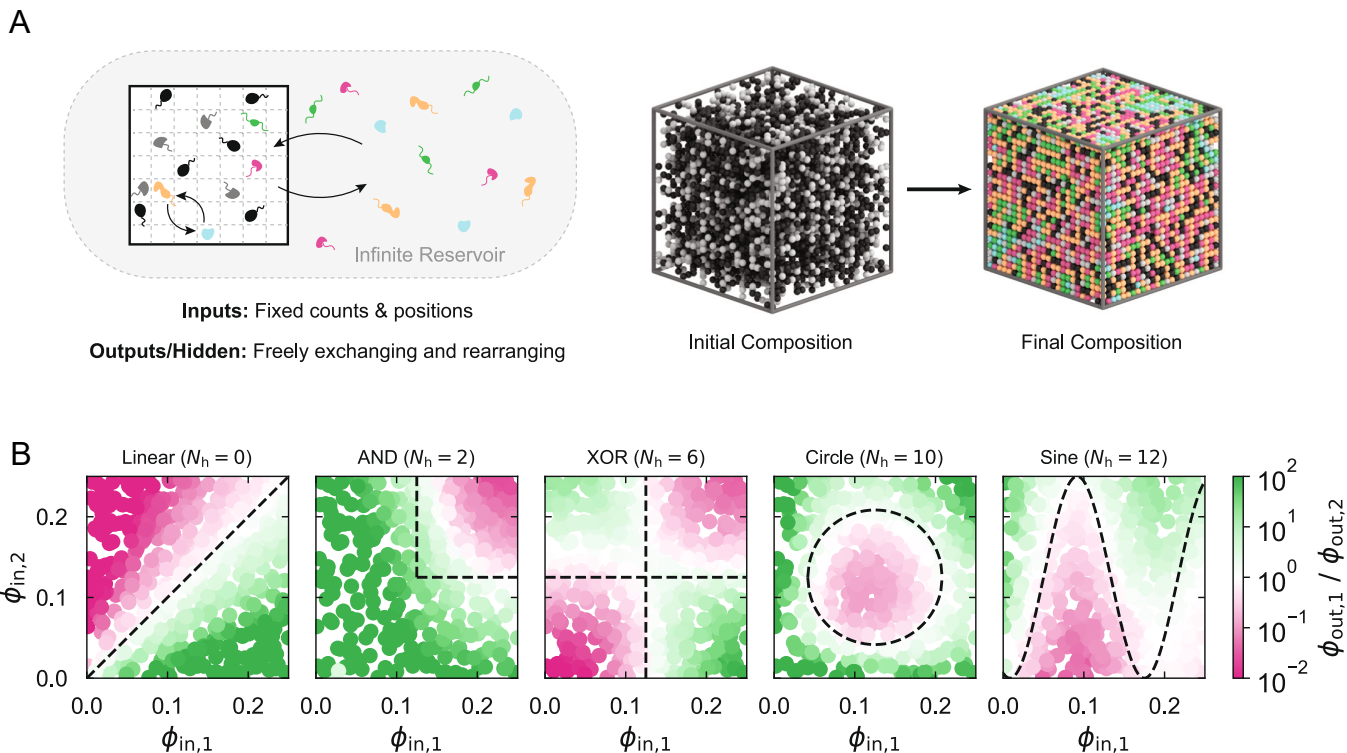


Fig. 7. (A) Schematic of 3D lattice liquid. Inputs are clamped to lattice positions, and output and hidden species can exchange freely. On the *Right* is the evolution of a lattice ($L = 24$) from initial to final configurations. (B) 3D lattice liquid ($L = 24$) classification using mapped mean-field parameters: (Left to Right) Linear boundary with 0 hidden species, AND boundary with 2 hidden species, XOR boundary with 6 hidden species, Circle boundary with 10 hidden species, and Sine boundary with 12 hidden species.

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 ref. 10, which explores crystalline self-limited assembly, MNIST digit classification is similarly demonstrated in a theoretical model with ~ 85 to 90% accuracy depending on the design constraints. Together, these results highlight the potential of multicomponent mixtures to classify high-dimensional decision boundaries despite differences 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 (32, 42, 63), 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 Appendix, Supplementary 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 nonexchanging 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. We sample this mixed-ensemble model through parallelized Monte-Carlo simulations to ensure sufficiently equilibrated thermodynamic properties and compositions (SI Appendix, Supplementary Note 4 and Fig. S13B). We simulate the lattice liquid with trained parameters from the

mean-field model and evaluate its 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 (SI Appendix, Fig. S13A), and all test data shown are at $\beta = 2$ in Fig. 7.

Using liquids trained on a range of decision boundaries reported in Figs. 2 and 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 than the abrupt jumps observed in mean-field liquids. By analyzing the phase structure within each surface, we find that this blurring likely arises from coexisting, spatially isolated pockets of phases that each comprise different output species (SI Appendix, Supplementary Note 5 and Fig. S14). Away from the boundary, output species ratios still reach 10 to $100\times$ ratios of correct over incorrect species (SI Appendix, 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 and SI Appendix, Fig. S13A). The broad agreement between mean-field and 3D lattice liquids supports the generality of our results and motivates 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 that recruits RNA Polymerases) but repressive ones at other DNA surfaces (with a different combination of transcription factors that recruits repressors). The function of a condensate is therefore driven by the recruitment of a single surface-specific “output” molecule. We note that our framework can be easily adjusted to model other pathways where condensates may need to recruit more complex compositions of functional output species, as illustrated in *SI Appendix, Fig. S16*. Together, our work suggests that emergent condensation in multicomponent liquids like the cellular milieu can drive 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 (64), 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 (65), varying coreceptors and adaptor proteins that recruit the same downstream kinase to membranes to propagate signaling cascades (27), 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. 5 A and 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, 66, 67). More generally, the features of multicomponent phase separation naturally provide cells with regulatory knobs such as changing composition (by expression) or interactions (by posttranslational modifications) to leverage condensate-mediated computations. Finally, 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). Together, our work advocates for an expanded view, including through the “hidden-output” axis, of how and where function can be encoded by condensates in biological systems.

We show that increasing hidden species generally improves the precision of classification. To better contextualize this computational power, we outline in *SI Appendix, Supplementary Note 3* an idealized construction, subject to key assumptions, that maps the essential physics underlying our surface condensation model to well-understood computational structures, namely, piecewise linear machines combined with winner-take-all dynamics and their corresponding Voronoi cells (68–70). Within this construction, we argue that surface condensates can universally approximate arbitrary decision boundaries through scaling the number of

hidden nodes (*SI Appendix, Fig. S17*). Systematically dissecting and exploring the assumptions that underlie this construction is an important area for future work. In our paper, we also find that classification power eventually saturates with more hidden nodes (*Fig. 4*). This 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 $k_B T$, 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, 71, 72), multimerization domains that reduce entropic costs of demixing (73), 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 (41, 74). Our model more closely resembles Boltzmann machines (42, 54) in that we exploit hidden species to encode more complex stimuli-response behaviors, i.e., higher expressivity (50, 75). A key distinction, more generally, is that condensation is typically collectively encoded and does not directly enable sequential information processing characteristic of modern deep learning networks. However, unlike these networks, condensation, and many other forms of equilibrium molecular computation, typically expend very little energy and are compatible with massively parallel computation—the same milieu of molecules can classify multiple spatially separated surfaces.

While we focus on classification, recent studies argue for broader computational capabilities embedded in multicomponent liquids. For example, ref. 41 explores the capacity of condensates to store and retrieve compositions as stable phases (or memories) analogous to Hopfield models trained with the Hebbian rule, identifying key roles for three-body interactions for effective memorization. More recently, Chalk et al. (42) train molecular parameters of 3D lattice liquids for multiple tasks, showing that equilibrium multicomponent liquids can form complex spatial architectures, drive Hopfield-like retrieval of memories upon contacting a patterned surface, and perform MNIST digit classification. In the MNIST classification task, our model performance saturates (~85%), potentially hinting at limitations in the physics of condensation and/or in the choice of data encoding. 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 (6–10, 76), mechanical systems (77, 78), and stochastic biomolecular reactions (79–81).

We characterize the computational capabilities of programmed multiphase fluids that are trained through gradient-descent based global optimization routines. While we focus on the system's

capacity for classification and not the biological plausibility of the learning rule, the ways by which molecular networks can learn, potentially autonomously, or be trained represents an important area for investigation. For instance, Chalk et al. (42) train lattice liquids using sleep–wake training rules, a set of spatially local rules based on competition between Hebbian learning and anti-Hebbian unlearning as in classical Boltzmann machines (54). 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 (42, 59, 82, 83). In particular, we show that only changing levels of hidden and output species in the reservoirs of trained fluids—a variable amenable to modification in living systems—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 (84)—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. In our framework, the surface is treated implicitly, characterized instead by the host of surface-resident inputs. Recent theories have begun to relate the explicit physical properties of biological scaffolds, both DNA sequence aspects (20, 22) and reversible membrane binding (39, 85), to the resulting phase behavior. These models describe how condensates can wet (condense on a lower-dimensional surface as well as in the bulk), prewet (condense only on a lower-dimensional surface), or drive coupled phase transitions between the scaffold along with the soluble proteins. Designing over such models, in turn, could enable more expressive computations by directly optimizing surface features and using complex phase behaviors. Toward experimental realization, promising avenues include a) designed DNA (10, 43, 86) molecules, along with surface-functionalized or immobilized DNA strands, and b) emerging synthetic biology approaches (87–90) 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. Toward 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 intrasurface 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 reservoirs will provide insights and offer pathways to design molecular networks more amenable to experimental instantiation.

Data, Materials, and Software Availability. We make code and figure data available via the following GitHub repository: https://github.com/shrinivaslab/2025_zentner_multiphase_classification (91).

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. and K.S. acknowledge support from NSF NRT 2021900–Synthesizing Biology Across Scales and Northwestern University. A.Z. and M.P.B. acknowledge support from NSF AI Institute of Dynamic Systems 2112085. C.C. and E.W. acknowledge support from NSF CCF/FET 2008589 and 2212546. A.M. acknowledges support from the NSF 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 workshop: Physical Principles Shaping Biomolecular Condensates supported in part by NSF PHY-2309135 and the Gordon and Betty Moore Foundation Grant No. 2919.02. This research was supported in part by grants from the NSF (DMS-2235451) and Simons Foundation (MPS-NITMB-00005320) to the NSF-Simons National Institute for Theory and Mathematics in Biology (NITMB). The computations in this paper were, in part, run on the Cannon cluster supported by the Faculty of Arts and Sciences Division of Science Research Computing Group at Harvard University, and in part, through the computational resources provided by the Quest high performance computing facility at Northwestern University.

Author affiliations: ^aSchool of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; ^bDepartment of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208; ^cCenter for Synthetic Biology, Northwestern University, Evanston, IL 60208; ^dNSF-Simons National Institute for Theory and Mathematics in Biology, Chicago, IL 60611; ^eComputation and Neural Systems, California Institute of Technology, Pasadena, CA 91125; ^fDepartment of Physics, Harvard University, Cambridge, MA 02138; and ^gDepartment of Physics, University of Chicago, Chicago, IL 60637

Author contributions: A.Z., E.V.H., C.C., M.P.B., A.M., E.W., and K.S. designed research; A.Z., E.V.H., C.C., A.M., E.W., and K.S. performed research; A.Z., E.V.H., C.C., A.M., E.W., and K.S. contributed new reagents/analytic tools; A.Z., E.V.H., E.W., and K.S. analyzed data; and A.Z., E.V.H., C.C., A.M., E.W., and K.S. wrote the paper.

- U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC Press, 2019).
- L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
- W. A. Lim, C. M. Lee, C. Tang, Design principles of regulatory networks: Searching for the molecular algorithms of the cell. *Mol. Cell* **49**, 202–212 (2013).
- J. Parres-Gold, M. Levine, B. Emert, A. Stuart, M. B. Elowitz, Contextual computation by competitive protein dimerization networks. *Cell* **188**, 1984–2002.e17 (2025).
- 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*, Y. Kim, F. A. Spelman, Eds. (IEEE, 1989), vol. 4, pp. 1354–1355.
- E. Winfree, “Algorithmic self-assembly of DNA,” PhD thesis, California Institute of Technology (1998).
- A. Murugan, Z. Zeravcic, M. P. Brenner, S. Leibler, Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 54–59 (2015).
- W. Zhong, D. J. Schwab, A. Murugan, Associative pattern recognition through macro-molecular self-assembly. *J. Stat. Phys.* **167**, 806–826 (2017).
- D. Woods et al., Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. *Nature* **567**, 366–372 (2019).
- C. G. Evans, J. O’Brien, E. Winfree, A. Murugan, Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly. *Nature* **625**, 500–507 (2024).
- J. Hasty, D. McMillen, J. J. Collins, Engineered gene circuits. *Nature* **420**, 224–230 (2002).
- T. Frei, C. H. Chang, M. Filo, A. Arampatzis, M. Khammash, A genetic mammalian proportional-integral feedback control circuit for robust and precise gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2122132119 (2022).

13. A. Klosin *et al.*, Phase separation provides a mechanism to reduce noise in cells. *Science* **367**, 464–468 (2020).
14. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
15. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
16. A. S. Lyon, W. B. Peeples, M. K. Rosen, A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 1–21 (2020).
17. J. M. Choi, A. S. Holehouse, R. V. Pappu, Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
18. F. Jülicher, C. A. Weber, Droplet physics and intracellular phase separation. *Annu. Rev. Condens. Matter Phys.* **15**, 237–261 (2024).
19. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
20. K. Shrinivas *et al.*, Enhancer features that drive formation of transcriptional condensates. *Mol. Cell* **75**, 549–561.e7 (2019).
21. W. Hur *et al.*, CDK-regulated phase separation seeded by histone genes ensures precise growth and function of histone locus bodies. *Dev. Cell* **54**, 379–394.e6 (2020).
22. J. A. Morin *et al.*, Sequence-dependent surface condensation of a pioneer transcription factor on DNA. *Nat. Phys.* **18**, 271–276 (2022).
23. A. Pancholi *et al.*, RNA polymerase II clusters form in line with surface condensation on regulatory chromatin. *Mol. Syst. Biol.* **17**, e10272 (2021).
24. K. Shrinivas *et al.*, “Dewdrops on the genome: Regulation of gene expression by biomolecular phase separation,” PhD thesis, Massachusetts Institute of Technology (2020).
25. C. S. Bond, A. H. Fox, Paraspeckles: Nuclear bodies built on long noncoding RNA. *J. Cell Biol.* **186**, 637–644 (2009).
26. W. T. Sneed, M. K. Skillicorn, K. Shrinivas, A. S. Gladfelter, Immiscible proteins compete for RNA binding to order condensate layers. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2504778122 (2025).
27. L. B. Case, J. A. Ditlev, M. K. Rosen, Regulation of transmembrane signaling by phase separation. *Annu. Rev. Biophys.* **48**, 465–494 (2019).
28. Y. G. Zhao, H. Zhang, Phase separation in membrane biology: The interplay between membrane-bound organelles and membraneless condensates. *Dev. Cell* **55**, 30–44 (2020).
29. W. T. Sneed, Condensate-membrane interactions shape membranes, tune cytoskeletal assembly, and localize mRNAs. *Curr. Opin. Cell Biol.* **95**, 102540 (2025).
30. R. P. Sear, J. A. Cuesta, Instabilities in complex mixtures with a large number of components. *Phys. Rev. Lett.* **91**, 245701 (2003).
31. P. Sollich, Predicting phase equilibria in polydisperse systems. *J. Phys. Condens. Matter* **14**, R79 (2001).
32. W. M. Jacobs, D. Frenkel, Predicting phase behavior in multicomponent mixtures. *J. Chem. Phys.* **139**, 024108 (2013).
33. W. M. Jacobs, Self-assembly of biomolecular condensates with shared components. *Phys. Rev. Lett.* **126**, 258101 (2021).
34. F. Chen, W. M. Jacobs, Programmable phase behavior in fluids with designable interactions. *J. Chem. Phys.* **158**, 214118 (2023).
35. D. Zwicker, L. Laan, Evolved interactions stabilize many coexisting phases in multicomponent liquids. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2201250119 (2022).
36. K. Shrinivas, M. P. Brenner, Multiphase coexistence capacity in complex fluids. bioRxiv [Preprint] (2022). <https://doi.org/10.1101/2022.10.19.512909> (Accessed 25 February 2026).
37. K. Shrinivas, M. P. Brenner, Phase separation in fluids with many interacting components. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2108551118 (2021).
38. Y. Qiang, C. Luo, D. Zwicker, Scaling laws for phase coexistence in multicomponent mixtures. *Phys. Rev. Res.* **7**, 043008 (2025).
39. M. Rouches, S. L. Veatch, B. B. Machta, Surface densities prewet a near-critical membrane. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103401118 (2021).
40. S. Mao, D. Kulindow, M. P. Haataja, A. Košmrlj, Phase behavior and morphology of multicomponent liquid mixtures. *Soft Matter* **15**, 1297–1311 (2019).
41. R. Braz Teixeira, G. Carugno, I. Neri, P. Sartori, Liquid Hopfield model: Retrieval and localization in multicomponent liquid mixtures. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2320504121 (2024).
42. C. Chalk, S. Buse, K. Shrinivas, A. Murugan, E. Winfree, “Learning and inference in a lattice model of multicomponent condensates” in *DROPS-IDN/v2/document/10.4230/LIPLCS.DNA.30.5*, S. Seki, J. M. Stewart, Eds. (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024).
43. G. R. Abraham, A. S. Chaderjian, A. B. N. Nguyen, S. Wilken, O. A. Saleh, Nucleic acid liquids. *Rep. Prog. Phys.* **87**, 066601 (2024).
44. S. Do, C. Lee, T. Lee, D. N. Kim, Y. Shin, Engineering DNA-based synthetic condensates with programmable material properties, compositions, and functionalities. *Sci. Adv.* **8**, eabj1771 (2022).
45. J. M. Stewart *et al.*, Modular RNA motifs for orthogonal phase separated compartments. *Nat. Commun.* **15**, 6244 (2024).
46. G. Fabrin *et al.*, Co-transcriptional production of programmable RNA condensates and synthetic organelles. *Nat. Nanotechnol.* **19**, 1665–1673 (2024).
47. J. Gong, N. Tsumura, Y. Sato, M. Takinoue, Computational DNA droplets recognizing miRNA sequence inputs based on liquid-liquid phase separation. *Adv. Funct. Mater.* **32**, 2202322 (2022).
48. M. Takinoue, DNA droplets for intelligent and dynamical artificial cells: From the viewpoint of computation and non-equilibrium systems. *Interface Focus* **13**, 20230021 (2023).
49. H. Udono, J. Gong, Y. Sato, M. Takinoue, DNA droplets: Intelligent, dynamic fluid. *Adv. Biol.* **7**, 2200180 (2023).
50. A. Murugan, D. Zwicker, C. Lorenz, E. R. Dufresne, Could living cells use phase transitions to process information? *Phys. Rev. X* **16**, 020501 (2026).
51. P. C. Hohenberg, B. I. Halperin, Theory of dynamic critical phenomena. *Rev. Mod. Phys.* **49**, 435–479 (1977).
52. J. Bradbury *et al.*, JAX: Composable transformations of Python+NumPy programs. GitHub. <http://github.com/google/jax>. Accessed 31 December 2021.
53. DeepMind *et al.*, The DeepMind JAX Ecosystem. GitHub. <http://github.com/google-deepmind>. Accessed 12 August 2025.
54. D. H. Ackley, G. E. Hinton, T. J. Sejnowski, A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9**, 147–169 (1985).
55. V. A. Marčenko, L. A. Pastur, Distribution of eigenvalues for some sets of random matrices. *Math. USSR Sbornik* **1**, 457 (1967).
56. A. Rahimi, B. Recht, *Random Features for Large-Scale Kernel Machines in Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2007), vol. 20, pp. 1177–1184.
57. G. Tanaka *et al.*, Recent advances in physical reservoir computing: A review. *Neural Netw.* **115**, 100–123 (2019).
58. C. J. Su *et al.*, Ligand-receptor promiscuity enables cellular addressing. *Cell Syst.* **13**, 408–425.e12 (2022).
59. M. Stern, A. Murugan, Learning without neurons in physical systems. *Annu. Rev. Condens. Matter Phys.* **14**, 417–441 (2023).
60. M. Stern, D. Hexner, J. W. Rocks, A. J. Liu, Supervised learning in physical networks: From machine learning to learning machines. *Phys. Rev. X* **11**, 021045 (2021).
61. M. Stern, C. Arinze, L. Perez, S. E. Palmer, A. Murugan, Supervised learning through physical changes in a mechanical system. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14843–14850 (2020).
62. K. M. Cherry, L. Qian, Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature* **559**, 370–376 (2018).
63. W. M. Jacobs, D. Frenkel, Phase transitions in biological systems with many components. *Biophys. J.* **112**, 683–691 (2017).
64. V. V. Vazirani, *Approximation Algorithms* (Springer Science & Business Media, 2013).
65. G. Pei, H. Lyons, P. Li, B. R. Sabari, Transcription regulation by biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* **26**, 213–236 (2025).
66. Y. E. Antebi *et al.*, Combinatorial signal perception in the BMP pathway. *Cell* **170**, 1184–1196.e24 (2017).
67. H. E. Klumpe *et al.*, The context-dependent, combinatorial logic of BMP signaling. *Cell Syst.* **13**, 388–407.e10 (2022).
68. R. O. Duda, H. Fossum, Pattern classification by iteratively determined linear and piecewise linear discriminant functions. *IEEE Trans. Electron. Comput.* **EC-15**, 220–232 (1966).
69. F. Aurenhammer, Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.* **23**, 345–405 (1991).
70. W. Maass, On the computational power of winner-take-all. *Neural Comput.* **12**, 2519–2535 (2000).
71. K. Bhandari, M. A. Cotten, J. Kim, M. K. Rosen, J. D. Schmit, Structure-function properties in disordered condensates. *J. Phys. Chem. B* **125**, 467–476 (2021).
72. M. Cappa, F. Sciortino, L. Rovigatti, A phase-field model for solutions of DNA-made particles. *J. Chem. Phys.* **162**, 194901 (2025).
73. U. Rana *et al.*, Asymmetric oligomerization state and sequence patterning can tune multiphase condensate miscibility. *Nat. Chem.* **16**, 1073–1082 (2024).
74. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558 (1982).
75. M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, J. Sohl-Dickstein, “On the expressive power of deep neural networks” in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup, Y. W. The, Eds. (PMLR, 2017), vol. 70, pp. 2847–2854.
76. S. L. Foley, M. E. Johnson, Membrane-associated self-assembly for cellular decision-making. *Phys. Rev. Res.* **8**, 023280 (2026).
77. D. A. Fletcher, R. D. Mullins, Cell mechanics and the cytoskeleton. *Nature* **463**, 485–492 (2010).
78. H. Yasuda *et al.*, Mechanical computing. *Nature* **598**, 39–48 (2021).
79. E. Winfree, “Chemical reaction networks and stochastic local search” in *International Conference on DNA Computing and Molecular Programming*, C. Thachuk, Ed. (Springer, 2019), pp. 1–20.
80. D. Soloveichik, M. Cook, E. Winfree, J. Bruck, Computation with finite stochastic chemical reaction networks. *Nat. Comput.* **7**, 615–633 (2008).
81. C. Floyd, A. R. Dinner, A. Murugan, S. Vaikuntanathan, Limits on the computational expressivity of non-equilibrium biophysical processes. *Nat. Commun.* **16**, 7184 (2025).
82. W. Poole *et al.*, “Chemical Boltzmann machines” in *DNA Computing and Molecular Programming*, R. Brijder, L. Qian, Eds. (Springer International Publishing, Cham, 2017), pp. 210–231.
83. W. Poole, T. Ouldridge, M. Gopalakrishnan, E. Winfree, Detailed balanced chemical reaction networks as generalized Boltzmann machines. arXiv [Preprint] (2022). <https://doi.org/10.48550/arXiv.2205.06313> (Accessed 19 August 2025).
84. R. Zhang *et al.*, Phase separation to buffer growth-mediated dilution in synthetic circuits. *Cell* **188**, 7413–7427 (2025).
85. X. Zhao, G. Bartolucci, A. Honigsmann, F. Jülicher, C. A. Weber, Thermodynamics of wetting, prewetting and surface phase transitions with surface binding. *New J. Phys.* **23**, 123003 (2021).
86. W. M. Jacobs, W. B. Rogers, Assembly of complex colloidal systems using DNA. *Annu. Rev. Condens. Matter Phys.* **16**, 443–463 (2025).
87. Y. Dai, L. You, A. Chilkoti, Engineering synthetic biomolecular condensates. *Nat. Rev. Bioeng.* **1**, 466–480 (2023).
88. Z. G. Qian, S. C. Huang, X. X. Xia, Synthetic protein condensates for cellular and metabolic engineering. *Nat. Chem. Biol.* **18**, 1330–1340 (2022).
89. M. T. Wei *et al.*, Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol.* **22**, 1187–1196 (2020).
90. H. Lyons *et al.*, Functional partitioning of transcriptional regulators by patterned charge blocks. *Cell* **186**, 327–345.e28 (2023).
91. A. Zentner, K. Shrinivas, 2025_zentner_multiphase_classification. GitHub. https://github.com/shrinivaslab/2025_zentner_multiphase_classification. Deposited 15 June 2026.