SI Appendix

Robert Marsland III, Owen Howell, Andreas Mayer, and Pankaj Mehta
(Dated: November 30, 2020)

I. DETAILS OF OUR MATHEMATICAL MODEL

We begin by considering a more mechanistic, biologically plausible model of Treg-mediated adaptive immunity. As in the main text, we will always use the convention that T cells to refer to non-regulatory T cells (i.e. conventional T cells). The basic elements of our model are as follows:

A. Assumptions and basic dynamics of the model

- λ_i are clone sizes of "conventional" T cells (specifically, $CD25^-CD4^+$ helper T cells).
- w_α are clone sizes of Tregs.
- v_x is the abundance of antigen-presenting cells (APC's) displaying antigen x. In general, x can represent any antigen displayed by a class II MHC, including neoantigens, but in this work we focus on the scenario where all displayed antigens are self-peptides.
- IL_x is the average local concentration of interleukin 2 (IL-2) in the vicinity of APC's displaying antigen x.
- p^c_{ix} ("cross-reactivity function") is the probability that a conventional T cell from clone i that encounters an APC displaying antigen x will bind and activate. The cross-reactivity function is determined by the binding affinity ΔG_{ix} between the T cell receptor and the antigen, via some model of the binding and activation kinetics. For example, a simple two-state equilibrium model would give $p^c_{ix} = 1/(1+e^{-\Delta G_{ix}/k_B T})$. In the present work, we do not attempt to relate p^c_{ix} to ΔG_{ix}, but instead sample p^c_{ix} directly from one of three probability distributions described in Section IV below.
- $p^r_{\alpha x}$ is the probability that a Treg from clone α that encounters an APC displaying antigen x will bind and activate. See above for explanation of relationship to binding affinity.
- IL-2 stimulates proliferation of both Tregs and normal T cells, with the growth rate some saturating functions $g_r(\mathrm{IL}_x),\ g_c(\mathrm{IL}_x)$ of the local IL-2 concentration IL_x.
- T cells deplete IL-2 at a rate proportional to the level of growth stimulation, with constants of proportionality ϵ_c^{-1} and ϵ_r^{-1} (notation comes from analogy with the efficiency of resource conversion into biomass).
- Activated T cells produce IL-2 at rate a.
- Activated Tregs directly suppress growth of nearby activated T cells, with each Treg cell decreasing the growth rate of T cells in its vicinity by an amount b.
- Tregs only suppress conventional T cells bound to the same APC, as suggested by the experiments of [1].
- Each antigen x is displayed on a small fraction $f \ll 1$ of the total population of APC's. This implies that a Treg binding to antigen x suppresses a conventional T cell bound to a different antigen y only on a much smaller fraction $f^2 \ll f$ of APC's. Under these conditions, we can neglect cross-antigen suppression, and consider only the suppression that occurs between Tregs and conventional T cells that are activated by the same antigen x.
- Both Tregs and conventional T cells circulate rapidly through the body, so that the total populations (including both activated and unactivated cells) are evenly distributed over all APC's.
- In the absence of extracellular IL-2, activated T cells proliferate at a basal rate ρ.
- Extracellular IL-2 is degraded by some external mechanisms, and has a lifetime τ in the absence of T cells.

These statements result in the following set of differential equations:

$$\frac{d\lambda_i}{dt} = \lambda_i \sum_x v_x p^c_{ix} \left[\rho + g_c(\mathrm{IL}_x) - b \sum_\alpha w_\alpha p^r_{\alpha x} \right] \tag{S1}$$

$$\frac{dw_\alpha}{dt} = w_\alpha \sum_x v_x p^r_{\alpha x} \left[g_r(\mathrm{IL}_x) - m \right] \tag{S2}$$

$$\frac{d\mathrm{IL}_x}{dt} = a \sum_i \lambda_i p^c_{ix} - \epsilon_c^{-1} \sum_i \lambda_i p^c_{ix} g_c(\mathrm{IL}_x) - \epsilon_r^{-1} \sum_\alpha w_\alpha p^r_{\alpha x} g_r(\mathrm{IL}_x) - \tau^{-1} \mathrm{IL}_x. \tag{S3}$$

B. Considering limit of fast interleukin dynamics yields minimal model in main text

To derive the minimal model in the main text, we assume that interleukin dynamics is fast compared to Treg and T cell proliferation. In this case, we can make a quasi-adiabatic approximation by setting $d\mathrm{IL}_x/dt = 0$. We further assume that the responses of the T cells and Tregs to interleukin concentrations are far from saturation so that we can approximate the growth rates using linear functions $g_c(\mathrm{IL}_x) = c_c\mathrm{IL}_x, g_r(\mathrm{IL}_x) = c_r\mathrm{IL}_x$. With these two assumptions, one gets that steady-state concentration of interleukins near antigen x takes the form

$$\mathrm{IL}_x = \frac{a\sum_i \lambda_i p_{ix}^c}{\tau^{-1} + \epsilon_c^{-1} c_c \sum_i \lambda_i p_{ix}^c + \epsilon_r^{-1} c_r \sum_\alpha w_\alpha p_{\alpha x}^r}. \tag{S4}$$

We will work in the regime where Tregs are the main sink for IL-2, and dominate the denominator of this expression. This results in the following dynamics for the two populations:

$$\frac{d\lambda_i}{dt} = \lambda_i \sum_x v_x p_{ix}^c \left[\rho + \frac{c_c\epsilon_r a}{c_r\sum_\alpha w_\alpha p_{\alpha x}^r}\sum_j \lambda_j p_{jx} - b\sum_\alpha w_\alpha p_{\alpha x}^r\right] \tag{S5}$$

$$\frac{dw_\alpha}{dt} = w_\alpha \sum_x v_x p_{\alpha x}^r \left[\frac{\epsilon_r a}{\sum_\beta w_\beta p_{\beta x}^r}\sum_j \lambda_j p_{jx}^c - m\right]. \tag{S6}$$

If we additionally assume that c_c/c_r small (consistent with the assumption of Treg-dominated interleukin consumption), we can ignore the positive feedback term in the first equation, yielding the system of equations:

$$\begin{aligned}\frac{d\lambda_i}{dt} &= \lambda_i \sum_x v_x p_{ix}^c \left[\rho - b\sum_\alpha w_\alpha p_{\alpha x}^r\right] \\ \frac{dw_\alpha}{dt} &= w_\alpha \sum_x v_x p_{\alpha x}^r \left[\frac{\epsilon_r a}{\sum_\beta w_\beta p_{\beta x}^r}\sum_j \lambda_j p_{jx}^c - m\right],\end{aligned} \tag{S7}$$

which are identical to the dynamics in the main text (where for notational simplicity we write $\epsilon_r a$ simply as a).

C. Rewriting our dynamics in terms of overlap kernels

We now describe the approximation mentioned in the main text that is required to rewrite the above dynamics in terms of overlap kernels. As stated in the main text, the overlap kernels are defined by (see Fig. S1):

$$\begin{aligned}\phi_{\alpha\beta} &= \sum_x v_x p_{\alpha x}^r p_{\beta x}^r \\ \phi_{i\alpha} &= \sum_x v_x p_{\alpha x}^r p_{ix}^c \\ r_i &= \sum_x p_{ix}^c v_x.\end{aligned} \tag{S8}$$

Rearranging Eq. S7 yields a set of equations where the cross-reactivity function and antigen concentrations almost always appear within an overlap expression:

$$\frac{d\lambda_i}{dt} = \lambda_i \left(\sum_y v_y p_{iy}^c\right)\left[\rho - b\frac{\sum_{\alpha,x} v_x p_{\alpha x}^r p_{ix}^c w_\alpha}{\sum_y v_y p_{iy}^c}\right] \tag{S9}$$

$$\frac{dw_\alpha}{dt} = w_\alpha \left[\epsilon_r a \sum_{j,x} \frac{v_x p_{\alpha x}^r p_{jx}^c \lambda_j}{\sum_\gamma w_\gamma p_{\gamma x}^r} - m\sum_{\beta,x} \frac{v_x p_{\alpha x}^r p_{\beta x}^r w_\beta}{\sum_\gamma w_\gamma p_{\gamma x}^r}\right]. \tag{S10}$$

The final step requires an uncontrolled approximation, whereby we ignore the correlations between the numerators and denominators in the dynamics of w_α, and sum over x separately for both sides of the fraction. This approximation

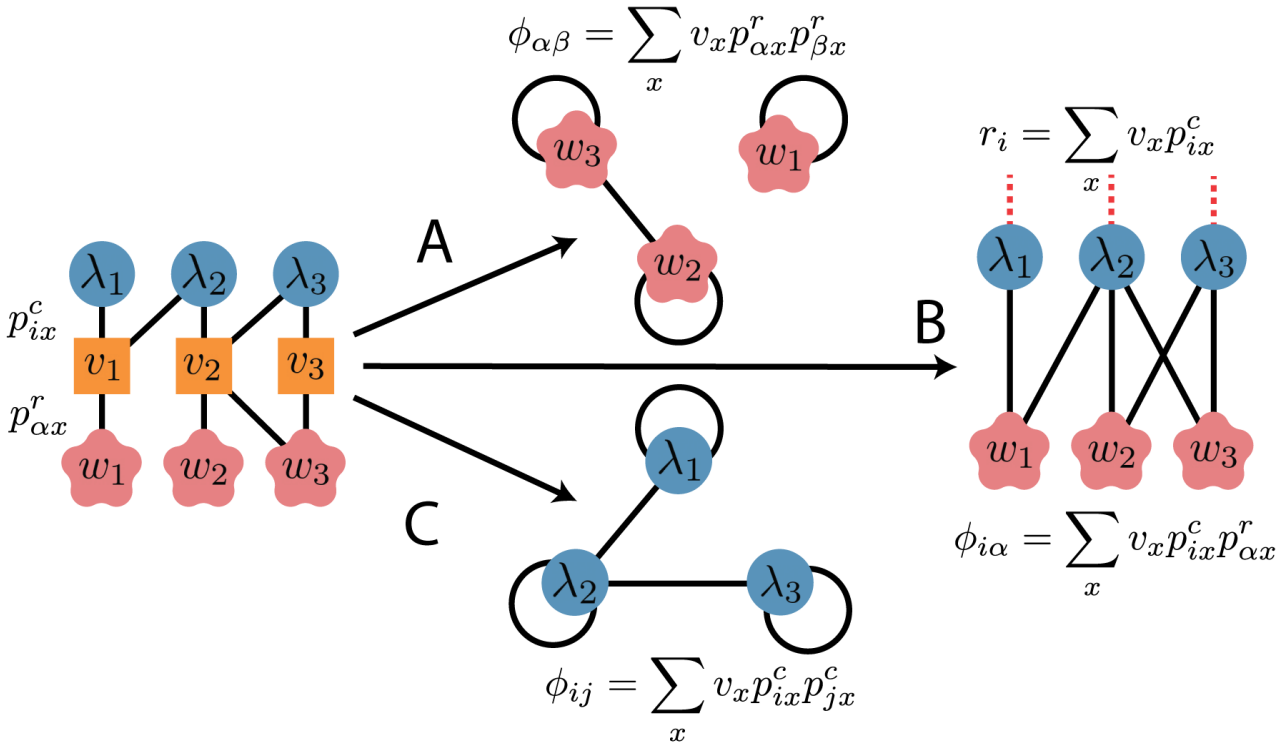

FIG. S1. **Defining the overlaps.** The cross-reactivity functions $p^r_{\alpha x}$ and p^c_{ix} define a network of interactions, with edges connecting Tregs and conventional T cells to antigens that can bind to their TCR, and edge weights representing the affinity of the interaction. The strength of the indirect interaction between two T cells can be quantified in terms of the product of their affinities for the same antigen, summed over all antigens and weighted by the antigen abundance. This procedure gives rise to three "overlap kernels": (A) $\phi_{\alpha\beta}$ for (competitive) effective interactions between Tregs, (B) ϕ_{ij} for (mutualistic) effective interactions between conventional T cells, and (C) $\phi_{i\alpha}$ for effective interactions between conventional T cells and Tregs. Note that ϕ_{ij} only appears in the positive feedback term of the full dynamical model defined in the first section of the SI. This term may be important at later stages of the immune response, when conventional T cell populations become large, but is neglected in the present analysis of the initial proliferation dynamics.

is strictly justified only in the emergent tiling regime, where the denominator $\sum_\gamma p_{\gamma x} w_\gamma$ is the same (equal to ρ/b) for all x. But in numerical simulations it appears to work well even outside of this regime, as well as during the transient on the way to an emergent tiling fixed point (see Fig. S2).

Using this approximation along with the overlap definitions provided above, we obtain the dynamics stated in Eq. 4 of the main text:

$$\begin{aligned}\frac{d\lambda_i}{dt} &= \lambda_i r_i \left[\rho - b r_i^{-1} \sum_\alpha \phi_{i\alpha} w_\alpha\right] \\ \frac{dw_\alpha}{dt} &= \frac{w_\alpha}{\sum_\beta w_\beta \bar{p}_\beta} \left[\epsilon_r a \sum_j \phi_{j\alpha} \lambda_j - m \sum_\beta \phi_{\alpha\beta} w_\beta\right]\end{aligned} \tag{S11}$$

where $\bar{p}_\beta \equiv \sum_x p^r_{\beta x}$.

II. TREG DYNAMICS AS OPTIMIZATION

We now show the dynamics above have a natural interpretation in terms of constrained optimization. To do so, we make use of the duality between the steady states of the equations above and constrained optimization. For completeness, we briefly explain this duality here. Please see our earlier papers [2, 3] for a detailed discussion.

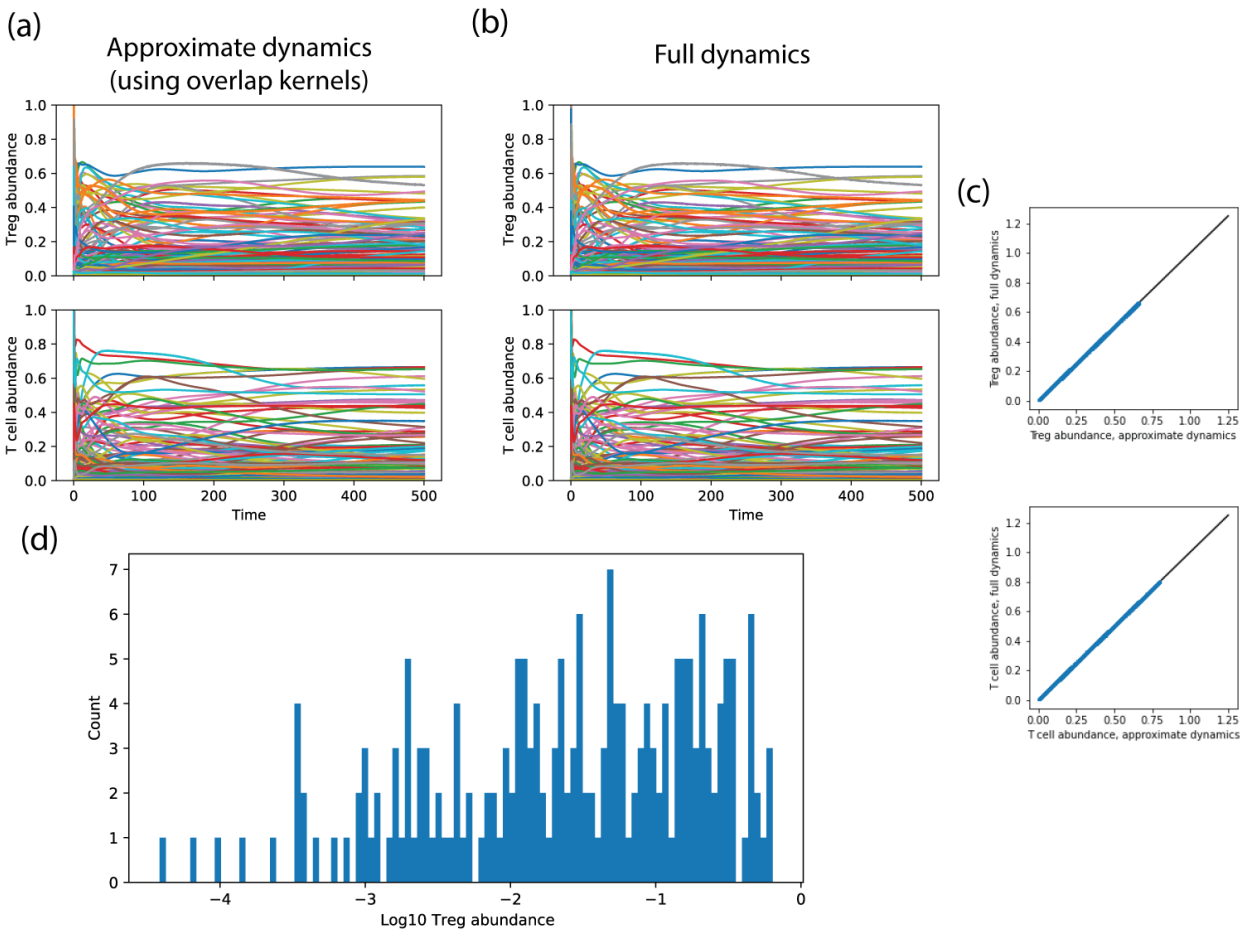

FIG. S2. **Comparing full dynamics and overlap kernel approximation.** *(a)* Sample trajectories of Treg abundances w_α and conventional T cell abundances λ_i using the approximate dynamics of Eq. S11. Cross-reactivity functions were generated using the one-dimensional shape space described in the final section of the SI ("Details on numerical simulations"), with cross-reactivity width $\sigma = 8$, and the other parameters set to $N_a = 100, N_c = 100, N_r = 50, \rho = a = b = m = 1$. Initial abundances of Tregs and conventional T cells were sampled from a lognormal distribution with logarithmic mean 0 and logarithmic standard deviation $\sigma = 2$. *(b)* Trajectories of the full dynamics of Eq. S7, using the same cross-reactivity functions, parameter values and initial conditions as the previous panel. *(c)* Comparison of exact and approximate dynamics from the previous two panels. Each point represents the abundance of a single Treg or conventional T cell lineage at a single time point in the two simulations. *(d)* Histogram of final Treg abundances in the simulation of the full dynamics from panel b. Note that the horizontal axis is the base-10 logarithm of the abundance.

Notice that the steady states of Eq. S13 satisfy the following equations

$$0 = \lambda_i \left[\rho - b r_i^{-1} \sum_\alpha \phi_{i\alpha} w_\alpha \right]$$
$$0 = \left[\epsilon_r a \sum_j \lambda_j \phi_{j\alpha} - m \sum_\beta w_\beta \phi_{\alpha\beta} \right]. \tag{S12}$$

Let us define $\tilde{\lambda}_i = \frac{\epsilon_r a r_i}{mb}\lambda_i$. Then we can rewrite the equation above as

$$0 = \tilde{\lambda}_i \left[\rho - b r_i^{-1} \sum_\alpha \phi_{i\alpha} w_\alpha\right]$$
$$0 = \left[-\sum_j \tilde{\lambda}_j b r_j^{-1} \phi_{j\alpha} - \sum_\beta w_\beta \phi_{\alpha\beta}\right]. \tag{S13}$$

If we define the functions

$$g_i(\{w_\alpha\}) = \left[\rho - b r_i^{-1} \sum_\alpha \phi_{i\alpha} w_\alpha\right] \tag{S14}$$

and

$$f(\{w_\alpha\}) = \frac{1}{2}\sum_\beta w_\beta \phi_{\alpha\beta} w_\alpha, \tag{S15}$$

then the steady-state equations take the form

$$0 = \sum_j \tilde{\lambda}_j \frac{\partial g_j(\{w_\alpha\})}{\partial w_\alpha} - \frac{\partial f(\{w_\alpha\})}{\partial w_\alpha}$$
$$0 = \tilde{\lambda}_j g_j(\{w_\alpha\}), \tag{S16}$$

We recognize the equations above as precisely the Karush-Kuhn-Tucker (KKT) conditions for constrained optimization with $f(\{w_\alpha\})$ the function being optimized and the functions $g_j(\{w_\alpha\})$ specifying the constraints. Thus, the steady-states of the equations above coincide with the solutions of the following constrained optimization problem:

$$\underset{\mathbf{w}}{\text{argmin}}\ \frac{1}{2}\sum_{\alpha,\beta} w_\alpha \phi_{\alpha\beta} w_\beta$$
$$\text{subject to :}$$
$$r_i^{-1}\sum_\alpha \phi_{i\alpha} w_\alpha \geq \frac{\rho}{b}. \tag{S17}$$
$$w_\alpha \geq 0 \tag{S18}$$

As an aside, this is very nearly single-class SVM, with training data $r_i^{-1}\phi_{i\alpha}$. It finds a hyperplane that separates all the data from the origin while maximizing the distance between the plane and the origin. It would be exactly a single-class SVM if $\phi_{\alpha\beta}$ were the identity matrix and the $w_\alpha \geq 0$ condition was not enforced. The $w_\alpha \geq 0$ slightly changes the geometrical interpretation of the 1-class SVM. Specifically, the requirement that $w_\alpha \geq 0$ forces the simplex $\sum_\alpha w_\alpha \phi_\alpha - p$ to have all positive coordinate intercepts. See [4] for more details on this interpretation.

III. EMERGENT TILING AS SOLUTION TO OPTIMIZATION PROBLEM

In the previous section, we stated the optimization problem in terms of the overlap kernels $\phi_{i\alpha}$ and $\phi_{\alpha\beta}$, which integrate over the whole antigen space. The emergent tiling conditions in the main text, however, involve the antigen space explicitly (Eq. 6). In this section, we exploit dual formulations of the optimization problem to highlight the role of the antigen space and make the connection to the emergent tiling conditions more transparent.

Before proceeding, it will be useful to collect some of the basic definitions that we stated above through out the text: Denote the dimension of the T-cell space N_c and the dimension of the Treg space N_r. Furthermore, denote the "naive" dimension of antigen space by N_a. This is essentially the number of x we sum over. However, if the matrix p^c_{ix} and $p^r_{\alpha x}$ are structured so that there are lots of correlations between the different antigens x, than this naive counting might be quite misleading, and we should really thing about the effective dimension of the antigen space N_a^{eff}. If all the x are uncorrelated, then of course $N_a^{\text{eff}} = N_a$.

Also, in this section we will drop the tilde from $\tilde{\lambda}_i$ (defined in the previous section), with the understanding that the solutions obtained for λ_i must be multiplied by $mb/(\epsilon_r a r_i)$ in order to give the actual T cell populations.

We will now write the optimization in Eq. S18 in a slightly different way, and in the process gain more physical insights about what we mean by these conditions. We begin by noting that plugging in the first line of Eq. S8 and flipping signs results in the trivial rewriting

$$\begin{aligned} &\underset{\mathbf{w}}{\text{argmax}} \; -\frac{1}{2}\sum_x v_x \left(\sum_\alpha p^r_{\alpha x} w_\alpha\right)^2 \\ &\text{subject to :} \\ &r_i^{-1}\sum_\alpha \phi_{i\alpha} w_\alpha \geq \frac{\rho}{b}. \\ &w_\alpha \geq 0. \end{aligned} \tag{S19}$$

A. Reformulation 1

It is now straightforward to check that

$$\underset{\mathbf{w}}{\text{argmax}} \; -\frac{1}{2}\sum_x v_x \left(\sum_\alpha p^r_{\alpha x} w_\alpha\right)^2 \tag{S20}$$

is the same as the following max-min optimization

$$\underset{\{s_x\}}{\text{argmin}} \;\; \underset{\mathbf{w}}{\text{argmax}} \sum_x \frac{s_x^2}{2v_x} - \sum_{\alpha,x} s_x p^r_{\alpha x} w_\alpha \tag{S21}$$

To see this note that we can differentiate this with respect to s_x and set this expression to zero to get that the optimum over the new auxiliary variable is

$$s_x^{\text{opt}} = \sum_\alpha p^r_{\alpha x} w_\alpha v_x, \tag{S22}$$

and plugging this into Eq. S21 gives the original optimization problem Eq. S19. Thus, we see that s_x^{opt} just measures the total coverage of antigen x by Tregs.

Another useful manipulation is to note that

$$\sum_\alpha \phi_{i\alpha} w_\alpha = \sum_x p^c_{ix} \sum_\alpha v_x p^r_{\alpha x} w_\alpha = \sum_x s_x^{\text{opt}} p^c_{ix} \tag{S23}$$

Combining this with the last line of Eq. S19 we can rewrite the constraint $r_i^{-1}\sum_\alpha \phi_{i\alpha} w_\alpha \geq \frac{\rho}{b}$ as

$$\sum_x p^c_{ix} v_x \left(\sum_\alpha p^r_{\alpha x} w_\alpha - \frac{\rho}{b}\right) = \sum_x p^c_{ix} \left(s_x^{\text{opt}} - v_x \frac{\rho}{b}\right) \geq 0 \tag{S24}$$

Notice that one way of satisfying this constraint is by requiring $s_x^{\text{opt}}/v_x = \sum_\alpha p^r_{\alpha x} w_\alpha = \frac{\rho}{b}$. Here we see that the idea is that we will make sure that each site x gets the same amount of coverage, set by ρ/b. This is precisely the emergent tiling we are seeking.

This might not always be possible since in general the dimension of antigen space, N_a may be larger than the dimension of the T-cell space N_c and the dimension of the Treg space N_r. However, if the matrix $p^r_{\alpha x}$ is structured such that the number of Tregs is much larger than the effective dimension of antigen space N_a^{eff} then this can be inverted. More generally, the existence of a solution is governed by a Gardner like transition analogous to that of perceptrons [5–7].

B. Reformulation 2: Langrange multipliers instead of inequality constraints

In this section, we will reformulate the problem again and get even more insight into how we can view the problem in the antigen space. This will also lead to important clues about where some unexpected properties of this solution

come from. Let us start with Eq. S26 and rewrite it in terms of p^c_{ix}, $p^r_{\alpha x}$, and v_x using expressions in Eq. S8:

$$\underset{\{\lambda_i\}}{\text{argmax}} \quad \underset{\mathbf{w}}{\text{argmin}} \sum_x \frac{v_x}{2}\left(\sum_\alpha p^r_{\alpha x} w_\alpha\right)^2 - \sum_x v_x \left(\sum_i \lambda_i p^c_{ix}\right)\left(\sum_\alpha p_{x\alpha} w_\alpha\right) + \sum_i \lambda_i \; r_i \frac{\rho}{b} \tag{S25}$$

$$\text{subject to} : \lambda_i \geq 0, \; w_\alpha \geq 0 \tag{S26}$$

Let us focus on the quantity we are optimizing. Notice that by completing the square and changing sign we can rewrite this as

$$\frac{v_x}{2}\left[\left(\sum_\alpha p^r_{\alpha x} w_\alpha\right) - \left(\sum_i \lambda_i p^c_{ix}\right)\right]^2 - \frac{v_x}{2}\left(\sum_i \lambda_i p^c_{ix}\right)^2 \tag{S27}$$

Let us introduce two new auxiliary variables A_x and B_x that will couple to each of these square terms. Then notice the expression above can be written as

$$\begin{aligned}
&\frac{v_x}{2}\left[\left(\sum_\alpha p^r_{\alpha x} w_\alpha\right) - \left(\sum_i \lambda_i p^c_{ix}\right)\right]^2 - \frac{v_x}{2}\left(\sum_i \lambda_i p^c_{ix}\right)^2 \\
&= \underset{\{A_x\}}{\text{argmax}} \frac{-A_x^2}{2v_x} - A_x\left[\left(\sum_\alpha p^r_{\alpha x} w_\alpha\right) - \left(\sum_i \lambda_i p^c_{ix}\right)\right] - \frac{v_x}{2}\left(\sum_i \lambda_i p^c_{ix}\right)^2 \\
&\underset{\{A_x\}}{\text{argmax}} \underset{\{B_x\}}{\text{argmin}} \frac{-A_x^2}{2v_x} - A_x\left(\sum_\alpha p^r_{\alpha x} w_\alpha - \sum_i \lambda_i p^c_{ix}\right) + \frac{B_x^2}{2v_x} - B_x \sum_i \lambda_i p^c_{ix}
\end{aligned} \tag{S28}$$

With all these manipulations we can rewrite the original optimization problem as

$$\underset{\{A_x\}}{\text{argmax}} \underset{\{B_x\}}{\text{argmin}} \underset{\{\lambda_i\}}{\text{argmax}} \underset{\{w_\alpha\}}{\text{argmin}} \sum_x \left[\frac{-A_x^2}{2v_x} - A_x\left(\sum_\alpha p^r_{\alpha x} w_\alpha - \sum_i \lambda_i p^c_{ix}\right) + \frac{B_x^2}{2v_x} - B_x \sum_i \lambda_i p^c_{ix}\right] + \sum_i \lambda_i \; r_i \frac{\rho}{b}$$

$$\text{subject to} : \lambda_i \geq 0, \; w_\alpha \geq 0 \tag{S29}$$

It doesn't look like we have done much right now. But one nice thing about this new optimization function is that it is *linear* in the w_α and λ_i. We can then take derivatives with respect to all four quantities to get a set of optimization equations. Taking the derivative with respect to B_x yields

$$B_x^{\text{opt}} = \sum_i v_x p^c_{ix} \lambda_i^{\text{opt}}. \tag{S30}$$

In other words, B_x^{opt} just measures the coverage of antigen site x by all conventional T cells. Taking the derivative with respect to A_x gives

$$\begin{aligned}
A_x^{\text{opt}} &= \sum_i v_x p^c_{ix} \lambda_i^{\text{opt}} - \sum_\alpha v_x p_{x\alpha} w_\alpha^{\text{opt}} \\
&= B_x^{\text{opt}} - \sum_\alpha v_x p_{x\alpha} w_\alpha^{\text{opt}}
\end{aligned} \tag{S31}$$

This equation say that A_x^{opt} is just the difference between the Tcell and Treg coverages at antigen site x. Taking the derivative with respect to w_α yields

$$\sum_x p_{x\alpha} A_x^{\text{opt}} = 0. \tag{S32}$$

To understand the meaning of this equation, it is useful to combine this with Eq. S31 and use Eq. S8 to rewrite this as

$$\sum_i \phi_{i\alpha} \lambda_i^{\text{opt}} - \sum_\beta \phi_{\alpha\beta} w_\beta^{\text{opt}} = 0, \tag{S33}$$

which is simply the statement that the growth rate of Treg α must be zero. Finally, taking the derivative with respect to λ_i gives the equation

$$\sum_x (B_x^{\text{opt}} - A_x) p_{ix}^c = r_i \frac{\rho}{b}. \tag{S34}$$

Plugging in Eq. S31 and using Eq. S8 it is easy to see that this equation just states that the Tcell growth rates should be zero.

C. Ansatz for solution to optimization problem

Thus, far we haven't gained much. But we will focus on what particularly interesting set of potential solutions to this problem. If the number of antigens N_a is very large compared to the number of T cells N_i, then we can in general easily find solutions to $\sum_x B_x^{\text{opt}} p_{ix}^c = r_i \frac{\rho}{b}$. However, we will make an even stronger ansatz. Notice that $r_i = \sum_x p_{ix}^c v_x$ so that the following ansatz is a solution to Eq. S34

$$\begin{aligned} B_x^{\text{opt}} &= v_x \frac{\rho}{b} \\ A_x &= 0. \end{aligned} \tag{S35}$$

In order for these to be good solutions, from Eq. S31 and Eq. S30 we must have that there exist solutions for w_α and λ_i satisfying the following set of equations.

$$\begin{aligned} \sum_\alpha p_{\alpha x}^r w_\alpha^{\text{opt}} &= \frac{\rho}{b} \\ \sum_j p_{jx} \lambda_j^{\text{opt}} &= \frac{\rho}{b}. \end{aligned} \tag{S36}$$

In general, these equations may not be solvable since the naive number of antigens N_a could be larger than the number of T cells, N_c, or number of Tregs N_r. But as long as the "effective" dimensionality of the antigen space (accounting for correlations between antigen binding affinities) satisfies $N_a^{eff} \ll N_c, N_r$, then we should be able to find such a solution. In fact, such a criteria has recently been derived in the statistical physics literature [7]. When the cross-reactivities are i.i.d, in the thermodynamic limit where $N_a, N_r, N_c \gg 1$, there is a phase transition between a regime where such a solution exists and does not depending on the ratios of N_r/N_a and N_c/N_a. This phase transition corresponds exactly to the Gardner solution to the perceptron problem [5–7].

D. Relation to Gardner Transition in Perceptrons

Here, we give a brief overview of the Gardner transition in perceptrons and discuss the relationship to the problem at hand (see books by Nishimore and Engel for details about perceptrons and statistical learning) [6, 8]. The perceptron problem is concerned with "storing" P boolean patterns ξ^u, with each pattern μ consisting of N input bits $S_i^\mu = \pm 1$ for $1 \leq i \leq N$. Each pattern is assigned a value $R^\mu = \pm 1$ according to the rule $R^\mu = \text{sgn}(\sum_i J_i S_i^\mu)$. A natural question one can ask is what is the maximum number of patterns such a function can classify correctly, where we are allowed to choose the N parameters J_i,

It was found that if the patterns were chosen randomly, that the maximum number of patterns that any function of this form could classify correctly was exactly equal to $P = 2N$. For $P > 2N$ generically, there existed no choice of J that will classify all P patterns correctly. This "phase transition" marks a boundary to the regime where there are no solutions for the N variables J_i to the P equations

$$R^\mu = \text{sgn}(\sum_i J_i \epsilon_i^\mu). \tag{S37}$$

If $P < 2N$, there exists a set of J_i that solve these equations. On the other hands, if $P > 2N$, there exists no solutions to this problem

As can be clearly seen, this system of equations are isomorphic to the kinds considered in our immunological problems with N playing the role of the Treg dimension, P the antigen dimension, and w_α^{opt} playing the roles of

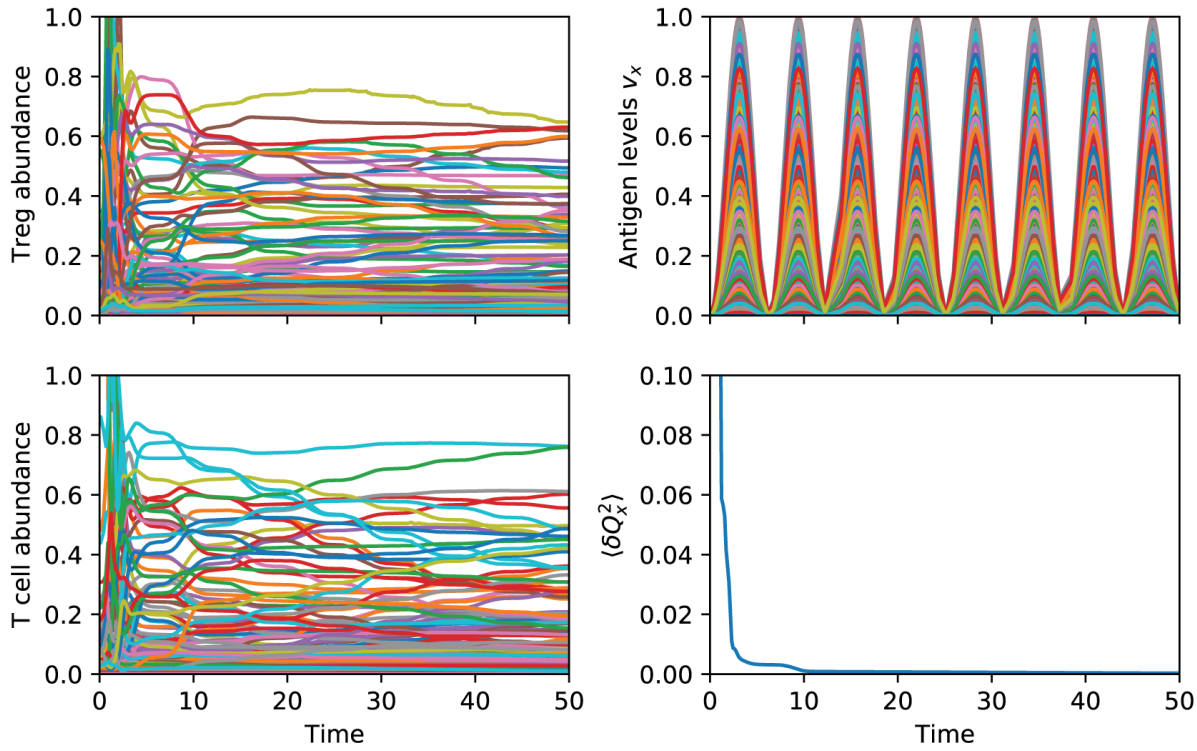

FIG. S3. **Simulation with rapidly oscillating v_x.** The full dynamics of Eq. (S7) were integrated with the antigen abundances v_x varying in time as $v_x = v_x^0 \sin^2(\omega t)$. The oscillation amplitudes v_x^0 were sampled from a uniform distribution between 0 and 1, and the oscillation frequency was $\omega = 0.5$. The other parameters and initial conditions sampling were identical to those of Figure S2b above. Abundances of conventional T cells and Tregs are shown as a function of time, along with the time-varying antigen abundances v_x and the Treg coverage nonuniformity $\langle \delta Q_x^2 \rangle$.

J_i. Thus, if the Treg dimension is at least twice the antigen dimension, we will have solutions to the equations for emergent tiling derived above. Since technically, we need to solve two such set of equations, we also need the diversity of conventional Tcells to be at least twice that of the antigen dimension. However, we make the implicit, biologically realistic assumption that the conventional Tcell diversity is always larger than the Treg diversity. This correspondence between perceptrons and ecology was also noted in a replica calculation in [7].

E. Insensitivity to antigen concentrations

Such a solution if it exists has some amazing properties, namely the T cell and Treg growth rates are insensitive to the antigen concentrations v_x ensuring that the Tregs exhibit an emergent tiling over the T cells:

- It balances conventional T cell and Treg activity at every antigen site x independently since $A_x = 0$.
- The growth rates of the T cells and Tregs become insensitive to changes in the v_x.

To see this latter point, notice that we can define "growth rates" for T cells g_i and Tregs g_α as

$$\frac{d\lambda_i}{dt} \equiv \lambda_i g_i = \frac{1}{b}\lambda_i \left[r_i \rho - \sum_\alpha \phi_{i\alpha} w_\alpha \right] \tag{S38}$$

$$\frac{dw_\alpha}{dt} \equiv \frac{w_\alpha}{\sum_\beta w_\beta \bar{p}_\beta} g_\alpha = \frac{w_\alpha b}{\rho} \left[\epsilon_r a \sum_j \lambda_j \phi_{j\alpha} - m \sum_\beta w_\beta \phi_{\alpha\beta} \right]. \tag{S39}$$

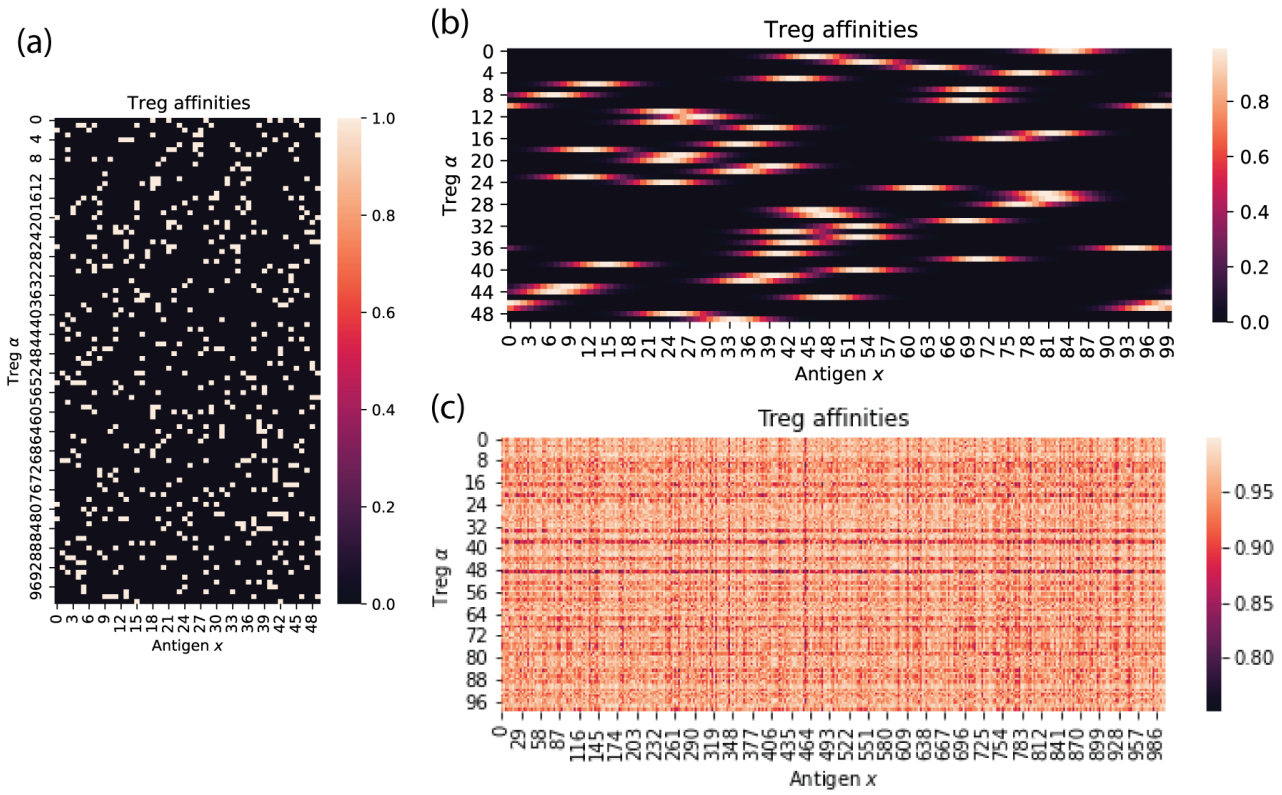

FIG. S4. **Random sampling of cross-reactivity functions p^c_{ix} and $p^r_{\alpha x}$.** Representative samples from the three cross-reactivity function ensembles described in SI Section IV are shown as heat maps. In all simulations, cross-reactivity functions for conventional T cells and Tregs were drawn from the same distribution. *(a)* Bernoulli distribution, in which a randomly chosen T cell and antigen have a probability p (here equal to 0.1) of interacting, and all interactions are assigned independently. *(b)* One-dimensional shape space, where a given T cell can bind to antigens whose shape coordinate is within a tolerance σ (here equal to 4) of a randomly assigned optimal shape. *(c)* High-dimensional shape space, where antigens and TCR's are randomly assigned coordinates in a shape space of specified dimension (here equal to 5), and the binding probability is determined by the pairwise distances.

A straightforward calculations using the definitions above then yields:

$$\begin{aligned}\frac{\partial g_i}{\partial v_x} &= \sum_x p^c_{ix}\left[\frac{v_x\rho}{b} + A_x - B_x\right] = 0\\ \frac{\partial g_\alpha}{\partial v_x} &= \sum_x p^r_{\alpha x} A_x = 0\end{aligned} \tag{S40}$$

IV. DETAILS ON NUMERICAL SIMULATIONS

To generate Figure 3 of the main text, we generated random cross-reactivity functions according to two different protocols, and then used the optimization formulation of the equilibrium conditions (Eqs. 5 of the main text) to efficiently obtain equilibrium populations of Tregs and conventional T cells for each realization. Scripts for generating the matrices, solving the optimization problem and plotting the results can be found in the accompanying github repository https://github.com/Emergent-Behaviors-in-Biology/immune-svm. Optimization was performed using the Python package CVXPY [9].

In the first protocol, the elements of $p^r_{\alpha x}$ and p^c_{ix} were sampled from Bernoulli distributions, with success probability equal to 0.1, 0.2 or 0.3. 30 realizations were generated for 20 values of N_r ranging from 100 to 300. The other parameters were $N_a = 100, N_c = 1,000, \rho = m = a = b = c_r = v_x = 1$.

In the second protocol, the elements of $p^r_{\alpha x}$ and p^c_{ix} were chosen in a correlated way, encoding a one-dimensional "shape space." Specifically, we generated a Gaussian cross-reactivity shape centered at the midpoint of the shape

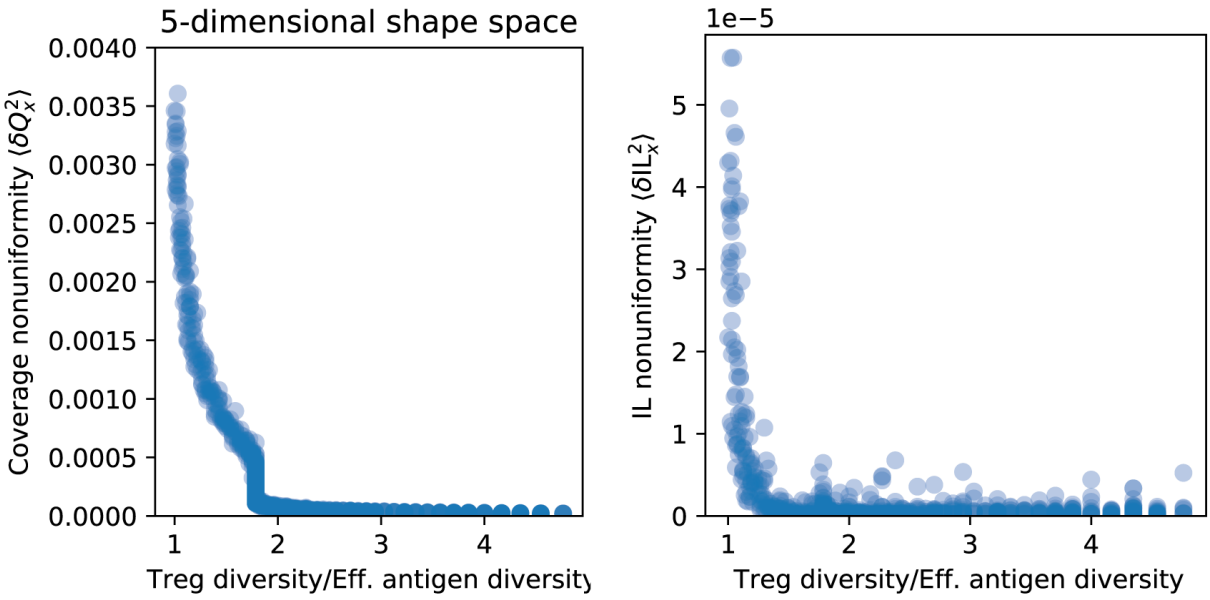

FIG. S5. **Emergent tiling transition in high-dimensional shape space.** Same as Figure 3 of the main text, but with the cross-reactivity functions $p^r_{\alpha x}$ and p^c_{ix} sampled using a five-dimensional shape space as illustrated in Figure S4 above and described in SI Section IV. The effective antigen diversity was defined as the number of singular values of $p_{\alpha x}$ exceeding a cutoff threshold of $\epsilon = 10^{-5}$.

space, given by

$$p_x = e^{-(x-N_a/2)^2/2\sigma^2} \tag{S41}$$

for a given cross-reactivity width σ, and then shifted it by a random offset x_α for each Treg and x_i for each conventional T cell. The shifts were performed with periodic boundary conditions, so that all Tregs and T cells still had the same overall binding capacity. 10 realizations were generated for each of 100 values of σ, ranging from 10 to 100. For each sampled cross-reactivity matrix $p^r_{\alpha x}$, we defined an effective number of distinguishable antigens N_a^{eff} by counting the number of singular values above an empirically determined numerical cutoff threshold of $\epsilon = 10^{-6}$. The horizontal axis in the right-hand panels of Fig. 3 is given by N_r/N_a^{eff}. The true number of antigens was $N_a = 5,000$, and the other parameters were $N_c = 500, \rho = m = a = b = c_r = v_x = 1$.

We also ran simulations with cross-reactivity functions generated in a higher-dimensional shape space. Following [10], we assigned shape coordinates $\mathbf{a}_x$ and $\mathbf{r}_i$ (or $\mathbf{r}_\alpha$ for Tregs) in a space of dimension D to the antigens and TCR's, respectively. We then calculated the cross-reactivity functions as

$$p^c_{ix} = e^{-\frac{||\mathbf{r}_i - \mathbf{a}_x||^2}{2\sigma^2}} \tag{S42}$$

$$p^r_{\alpha x} = e^{-\frac{||\mathbf{r}_\alpha - \mathbf{a}_x||^2}{2\sigma^2}} \tag{S43}$$

where a small distance in shape space corresponds to a good fit between TCR and antigen, while larger distances produce bad fits that do not bind. The parameter σ sets the radius of the region of shape space that is compatible with a given TCR. We randomly sampled all the shape coordinates ($\mathbf{r}_i$, $\mathbf{r}_\alpha$ and $\mathbf{a}_x$) from unit normal distributions. For the simulations shown in Figure S5, we sampled $N_r = 100$ Tregs, $N_c = 100$ conventional T cells and $N_a = 1000$ antigens. We chose a shape space of dimension $D = 5$, following dimensionality estimates derived from analysis of hemagglutination inhibition assays for influenza [11]. We varied the cross-reactivity width σ from 5 to 20 in order to sweep the ratio of the Treg diversity to effective antigen diversity through the emergent tiling threshold.

[1] T. Takahashi, Y. Kuniyasu, M. Toda, N. Sakaguchi, M. Itoh, M. Iwata, J. Shimizu, and S. Sakaguchi, International Immunology **10**, 1969 (1998).

[2] P. Mehta, W. Cui, C.-H. Wang, and R. Marsland III, Physical Review E **99**, 052111 (2019).
[3] R. M. III, W. Cui, and P. Mehta, The American Naturalist **in press** (2020).
[4] O. L. Howell, W. Cui, R. M. III, and P. Mehta, Journal of Physics A: Mathematical and Theoretical **in press** (2020).
[5] E. Gardner, Journal of physics A: Mathematical and general **21**, 257 (1988).
[6] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, New York, NY, 2001).
[7] S. Landmann and A. Engel, Europhysics Letters **124**, 18004 (2018).
[8] A. Engel and C. Van den Broeck, *Statistical mechanics of learning* (Cambridge University Press, 2001).
[9] S. Diamond and S. Boyd, Journal of Machine Learning Research **17**, 1 (2016).
[10] A. Mayer, V. Balasubramanian, T. Mora, and A. M. Walczak, Proceedings of the National Academy of Sciences **112**, 5950 (2015).
[11] A. Lapedes and R. Farber, Journal of theoretical biology **212**, 57 (2001).