# Adaptive efficiency of information processing in immune-pathogen co-evolution

Quinn N Bellamy,[1,*] Zachary Montague,[1,*] Luca Peliti,[2] and Armita Nourmohammad[1,3,4,5,†]

[1] *Department of Physics, University of Washington, 3910 15th Avenue Northeast, Seattle, WA 98195, USA*
[2] *Santa Marinella Research Institute, 00058 Santa Marinella, Italy*
[3] *Paul G. Allen School of Computer Science and Engineering, University of Washington, 85 E Stevens Way NE, Seattle, WA 98195, USA*
[4] *Department of Applied Mathematics, University of Washington, 4182 W Stevens Way NE, Seattle, WA 98105, USA*
[5] *Fred Hutchinson Cancer Center, 1241 Eastlake Ave E, Seattle, WA 98102, USA*

Organisms have evolved immune systems that can counter pathogenic threats. The adaptive immune system in vertebrates consists of a diverse repertoire of immune receptors that can dynamically reorganize to specifically target the ever-changing pathogenic landscape. Pathogens in return evolve to escape the immune challenge, forming an co-evolutionary arms race. We introduce a formalism to characterize out-of-equilibrium interactions in co-evolutionary processes. We show that the rates of information exchange and entropy production can distinguish the leader from the follower in an evolutionary arms races. Lastly, we introduce co-evolutionary efficiency as a metric to quantify each population's ability to exploit information in response to the other. Our formalism provides insights into the conditions necessary for stable co-evolution and establishes bounds on the limits of information exchange and adaptation in co-evolving systems.

## I. INTRODUCTION

Organisms, ranging from bacteria to humans, possess immune systems that have evolved to counteract pathogenic threats in their environments. In vertebrates, including humans, the adaptive immune system mounts a flexible and diverse response to neutralize pathogens. This system is shaped by eco-evolutionary dynamics of immune cells during the host's lifetime, as it detects infections and generates and reorganizes its diverse repertoire of immune cells to specifically target and clear pathogens [1]. The CRISPR-Cas immunity in bacteria operates on the same principle to combat infections by phages, which are viruses that attack bacteria [2–4]. On the other hand, pathogens, especially viruses, thrive by transmission through multiple hosts and evolve to escape the hosts' immune defenses over time. In response, the host updates its immune repertoire to battle the escaping pathogens, resulting in an out-of-equilibrium co-evolutionary arms race [5].

Focusing on vertebrates, immune-pathogen co-evolution occurs on multiple scales. Intra-host co-evolution, for example, can shape infections with chronic viruses like HIV [5–11]. In this case, the immune B-cell receptors (BCRs) undergo rapid Darwinian somatic evolution through affinity maturation to accumulate beneficial mutations that improve their affinities for binding to an infecting viral variant [12]. Viruses then evolve within each host to escape this immunity, prompting further affinity maturation of B-cell receptors. At the population level, viruses transmit from one host to another eliciting independent immune responses across individuals. Viruses evolve to escape the resulting population-level immunity, again driving a co-evolutionary arms race, apparent in viral escape during the COVID pandemic or in the evolution of the seasonal influenza virus [13, 14].

In both the intra-host and population-level co-evolutionary processes, the affinity of interactions between immune receptors and viral epitopes is a key molecular phenotype that determines the (evolutionary) fitness of either side. Pathogens escape through epitope mutations that reduce their affinity while immune receptors are selected for their heightened affinities. The map from genotypes to a given phenotype describing molecular function (e.g., protein binding affinity) is highly degenerate, and so is the map from combinations of phenotypes to fitness, as a measure of the reproductive success of an individual. Population genetics models have been developed to study the effects of mutations, selection, and stochasticity on such co-evolutionary processes in both the genotype and the phenotype spaces [5, 13–18]. The picture that emerges from all these descriptions has a lot in common with how a statistical description of physical systems with large number of degrees of freedom, e.g. gas molecules, can be mapped into a thermodynamic model for macroscopic quantities, such as pressure and energy [19–21]. This connection has been explored extensively for populations in equilibrium [22, 23], and more formal analogies have been drawn between stochastic thermodynamics and driven evolutionary processes in time-dependent environments [24–29]. Although compelling, these analogies are not exact. Notably, evolutionary theory is a stochastic process and does not have a conserved quantity resembling the energy in thermodynamics (i.e., it does not satisfy a conservation law similar to the first law of thermodynamics). However, there is

* These authors contributed equally.
† Correspondence should be addressed to: Armita Nourmohammad: armita@uw.com.

a strong correspondence between the second law of thermodynamics concerning entropy production in physical systems and the rules governing the production of information theoretical entropy for evolving populations. Notably, the fitness flux theorem was developed [25] in close analogy with the concept of entropy production in stochastic thermodynamics [30–32]; fitness flux quantifies the extent of out-of-equilibrium adaptation a population undergoes in response to changing environments.

In this work, we focus on co-evolutionary processes, in which the feedback between the two populations—and not an external time-varying environment—is the source of the evolutionary drive. We propose a bipartite discrete-time Markov process for interacting subsystems, as a proxy to model immune-pathogen interactions. While this approach comes at the cost of biological over-simplification, it ensures the tractability and interpretability of our models, offering clear insights into the newly introduced thermodynamics concepts in biological co-evolution. Specifically, we introduce the rates of information exchange and entropy production as a way to identify the leader and the follower in a co-evolutionary system. We introduce the concept of co-evolutionary efficiency, which generically applies to stochastic processes that are not constrained by the first law of thermodynamics (i.e., in which there is no conserved quantity like energy). Co-evolutionary efficiency quantifies the efficacy of each population in exploiting information in response to the other population. It defines conditions for stable co-evolution, and places bounds on the limits of information exchange and adaptation in co-evolving systems.

## II. MODEL

We assume interacting adversarial populations of immune agents and pathogens, which without loss of generality, we refer to as antibodies and viruses, respectively. We assume that the dynamics of the joint system follows a Markov process, whereby the state of the joint system at time $t$ only depends on its state at time $t-1$ and not on its prior history (Fig. 1A).

To describe this process more concretely, let $A$ represent the antibody ensemble and $V$ the viral ensemble. We represent the space of realizable states of $A$ by $\Omega(A) = \{1, \ldots, n_A\}$ and the space of realizable states of $V$ by $\Omega(V) = \{1, \ldots, n_V\}$, where $n_A$ and $n_V$ are the dimensionality of the antibody and virus ensemble respectively. We represent a joint state of the antibody-virus system as $(a, v)$, where $a \in \Omega(A)$ and $v \in \Omega(V)$. For instance, for the case of $n_A = n_V = 2$, all the possible states $(a, v)$ that can be realized are $\{(1,1), (1,2), (2,1), (2,2)\}$.

We describe the stochastic transitions between two arbitrary (joint) states $(a', v') \to (a, v)$ by an $N \times N$ discrete-time Markov transition matrix $W$, where $N = n_A n_V$ is the number of possible joint states. Each entry of $W$ represents the probability of transitioning from one state to another. Specifically, we use the convention in which the matrix element $W_{ji} = W^{v' \to v}_{a' \to a}$ denotes the transition probability from the state $i \equiv (a', v')$ (column) to the state $j \equiv (a, v)$ (row). The diagonal elements of the matrix are set by normalization: $W^{v \to v}_{a \to a} = 1 - \sum_{(a',v') \neq (a,v)} W^{v \to v'}_{a \to a'}$. Moreover, to make the problem analytically tractable, we assume that the system is bipartite, whereby transitions happen exclusively in one or the other population, but never simultaneously. Analogous bipartite treatments have been done for the analysis of information exchange between physical systems within the framework of stochastic thermodynamics [33–36]. The resulting bipartite transition matrix is given by

$$W^{v' \to v}_{a' \to a} = \begin{cases} W^{v}_{a' \to a} & \text{if } v = v', \\ W^{v' \to v}_{a} & \text{if } a = a', \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The fitness of antibodies or viruses depend on their binding abilities: a bound antibody-virus complex can lead to the neutralization of the virus, while an unbound virus is able to evade immunity, causing more infections. In our model, we assume that a state $(i, j)$ is bound if $i = j$, and unbound otherwise (Fig. 1B). With this definition, we implicitly disregard cross-reactivity in interactions, which refers to the ability of a given antibody (or virus) to react with targets other than its cognate one, often with slightly reduced affinity.

Our model can allow for an imbalanced number of viral and antibody states, i.e., $n_V \neq n_A$. In the case that $n_V > n_A$ there exists states that are inaccessible to the antibodies (i.e., $v = n_A + 1, \ldots, n_V$) and if the virus transitions to any of them, no change in the antibody state can result in binding. As a result, the virus can simply hide in the inaccessible states, leading to the extinction of the host—and the breakdown of the co-evolutionary dynamics—as no immune response can counter infections. On the other hand, if $n_A > n_V$, for each viral state there exists more antibody states that do not bind to the virus, compared to the balanced case of $n_A = n_V$. This results in an evolutionary distraction for antibodies, which reduces the efficiency of the antibody response to viruses but does not qualitatively impact the co-evolutionary process. Although mathematically interesting, the imbalanced case $n_A \neq n_V$ does not offer additional biological insight into the co-evolutionary dynamics. Therefore, for the clarity of our derivations, we will consider the balanced case of $n_A = n_V = n$ in the main text and only offer discussions about the imbalanced case in Appendix A.

In our bipartite model, we assume that the transition probabilities between different states in each subpopulation only depend on whether the ensuing phenotypic change is beneficial or deleterious for the evolving (transitioning) subpopulation. As such, we set that for the antibodies, beneficial changes (transitions from unbound to bound) occur with a probability $A_+$, deleterious changes (transitions from bound to unbound) occur with a prob-

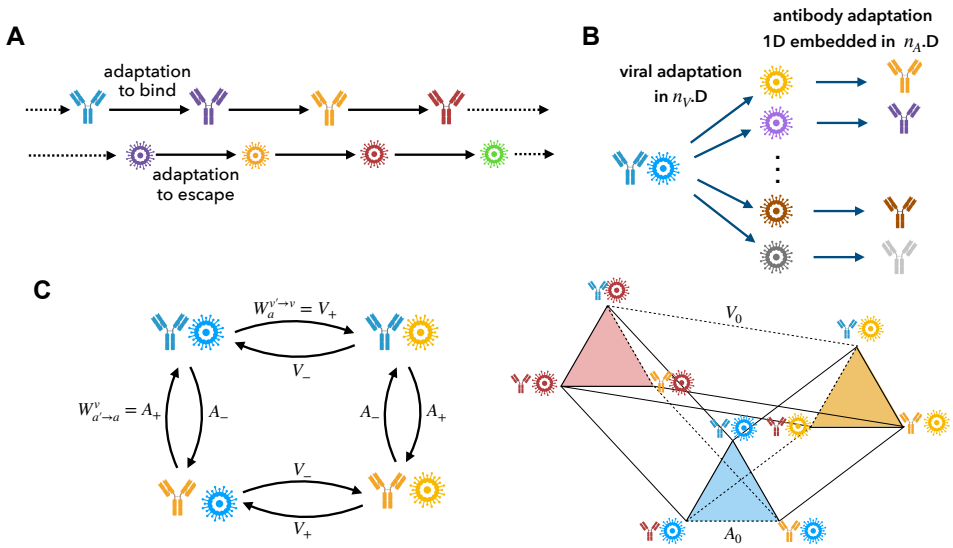


Figure 1. **Bipartite stochastic model of immune-pathogen co-evolution.** **(A)** Antibodies adapt to bind to the circulating viruses (match the color), while viruses adapt to escape the antibodies (mismatch the color). **(B)** We assume that viruses can escape to $n_V - 1$ states (out of $n_V$) that do not match the color of the circulating antibody. To bind to and neutralize a given virus, antibodies have only a single choice to evolve to match the virus (i.e., the state with the same color). As a result, the beneficial substitution of antibodies to bind to a given virus happens along 1 dimension, embedded in an $n_A$ dimensional space of possible antibodies. For most of our analyses, we assume a balanced evolutionary space where $n_A = n_V = n$. **(C)** We assume that the system is bipartite, whereby transitions happen exclusively in one or the other population, but never simultaneously (eq. (1)). The bipartite graph for $n = 2$ (left) and $n = 3$ (right) are shown. We assume that the transition probabilities between different states depend on whether the new state is beneficial or deleterious for the transitioning subpopulation. For antibodies, beneficial transitions with probability $W^v_{a'\to a} = A_+$ result in a matching color with the circulating virus ($a = v$), and deleterious transitions with probability $W^v_{a'\to a} = A_-$ result in a mismatched color ($a \neq v$). For $n \geq 3$, neutral transitions can occur, whereby the system remains unbound after the transition. Neutral transitions occur with probability $A_0$ for antibodies, and with probability $V_0$ for viruses, and are indicated by dashed lines for the case with $n = 3$ (right). Beneficial and deleterious transitions are indicated by full lines.

ability $A_-$, and neutral changes that do not modify the binding phenotype of the antibody occur with a probability $A_0$ (Fig. 1C). Similarly, for viruses, we assume that the beneficial changes (transitions from bound to unbound) occur with a probability $V_+$, deleterious changes (transitions from unbound to bound) occur with a probability $V_-$, and neutral changes occur with a probability $V_0$ (Fig. 1C). The resulting non-zero off-diagonal elements of the transition matrix $W$ can be expressed as

$$W^v_{a'\to a} = \underbrace{(1-\delta_{a',v})\delta_{a,v}}_{\text{unbound}\to\text{bound}} A_+ + \underbrace{\delta_{a',v}(1-\delta_{a,v})}_{\text{bound}\to\text{unbound}} A_- + \underbrace{[\delta_{a',v}\delta_{a,v} + (1-\delta_{a',v})(1-\delta_{a,v})]}_{\text{no phenotypic change}} A_0, \quad (2)$$

$$W^{v'\to v}_a = \underbrace{\delta_{a,v'}(1-\delta_{a,v})}_{\text{bound}\to\text{unbound}} V_+ + \underbrace{(1-\delta_{a,v'})\delta_{a,v}}_{\text{unbound}\to\text{bound}} V_- + \underbrace{[\delta_{a,v'}\delta_{a,v} + (1-\delta_{a,v'})(1-\delta_{a,v})]}_{\text{no phenotypic change}} V_0. \quad (3)$$

where $\delta_{x,y}$ is the Kronecker delta function, taking the value 1 when $x = y$, and 0, otherwise. These transition probabilities should be interpreted as substitutions in a population, which is the probability for a variant carrying a specific mutation to grow and take over the population (i.e., fix) by natural selection. Even though the baseline mutation rates should not depend on the phenotypic effect of a mutation, the substitution rates do. Beneficial mutations are exponentially more likely to fix in a population than their deleterious counterparts [37]. As such, we assume that $A_+ > A_0 > A_-$, and $V_+ > V_0 > V_-$. Lastly, our model implies that both the antibody and viral populations are monomorphic, meaning that they are comprised of a single type at a given time, which can be substituted by another type by the transition probability $W$. A more realistic evolutionary model should include the effect of diversity and polymorphism in each population.

The bound and unbound states form the two relevant phenotypes for evolutionary transitions. For the case with a balanced number of antibody and viral states ($n_A = n_V = n$), we can argue by symmetry that the diagonal terms of the transition matrix $W^{v\to v}_{a\to a}$ fall into two classes: the ones associated with bound states ($a = v$), which we denote by $W_{\mathbf{B}}$, and the ones associated with unbound states $a \neq v$, which we denote by $W_{\mathbf{U}}$. These

diagonal elements are the probabilities that the system remains in the state that it starts at. These two probabilities are given by

$$W_{\mathbf{B}} = 1 - (n-1)\left(V_+ + A_-\right), \tag{4}$$
$$W_{\mathbf{U}} = 1 - \left[V_- + A_+ + (n-2)(A_0 + V_0)\right]. \tag{5}$$

The diagonal element in each case is the complement of the probability to leave the specified state. For example, the probability of remaining in a bound state $W_{\mathbf{B}}$ is the complement of the sum of probabilities for all the transitions out of a bound state. Since our model allows for only one bound state for a given antibody and virus, i.e., the state $a = v$, all transitions from a bound state would end in an unbound configuration. There are $n-1$ such transitions that can be reached either by beneficial changes in viruses or deleterious changes in antibodies, with substitution probabilities $V_+$ and $A_-$, respectively. Therefore, $W_{\mathbf{B}}$ can be expressed by eq. (4). Similarly, $W_{\mathbf{U}}$ in eq. (5) is evaluated by accounting for all the possible transitions out of an unbound state, i.e., the transition to a single bound state by either antibodies or viruses, and the neutral transitions to any of the other $n-2$ unbound states (with rates $A_0$ and $V_0$ for antibodies and viruses); see Appendix A for a discussion on the imbalanced case.

Interestingly, the fact that the probabilities of remaining in a given state (diagonal elements) must be between zero and one sets an upper bound for the beneficial substitution probability of viruses that strongly depend on the dimensionality of the system (from eq. 4),

$$V_+ \leq \frac{1-(n-1)A_-}{n-1} \tag{6}$$

Similarly, eq. 5 sets the upper bound for the beneficial substitution probability of antibodies as $A_+ \leq 1 - (V_- + (n-2)(A_0+V_0))$. The neutral substitution probabilities, $A_0$ and $V_0$, scale as $1/N_{\text{eff}}$ with the effective population size of each system $N_{\text{eff}}$ [37], which is likely to be much larger than the dimensionality of the system $n$, which follows $(n-2)(A_0+V_0) \ll 1$. Therefore, given that the deleterious substitution probabilities are exponentially small $V_- \ll 1$, the upper bound on the beneficial substitution probability for antibodies can be approximated as $A_+ \lesssim 1$, with no strong dependence on the dimensionality of the system.

## III. RESULTS

### A. Stationary state statistics

**Stationary state probabilities.** The stationary state probability of all the possible $n^2$ states, which we denote by the vector $P_{\text{st}}$, should follow from $P_{\text{st}} = WP_{\text{st}}$, i.e., $P_{\text{st}}$ is the eigenvector of the transition matrix associated with eigenvalue 1. To solve this eigenvector problem, let us denote the stationary state probability associated with the bound state $(a, v)$ by $P_{\text{st}}(a, v; \mathbf{B})$, which can be expressed as

$$P_{\text{st}}(a,v;\mathbf{B}) = W_{\mathbf{B}} P_{\text{st}}(a,v;\mathbf{B}) + \sum_{a' \neq a, v} W^{v}_{a' \to a} P_{\text{st}}(a',v;\mathbf{U}) + \sum_{v' \neq v, a} W^{v' \to v}_{a} P_{\text{st}}(a,v';\mathbf{U}). \tag{7}$$

The first term accounts for remaining in the bound $(a, v)$ state with probability $W_{\mathbf{B}}$ (diagonal term from eq. (4)). The other term accounts for the transitions from all the accessible $n-1$ unbound states to the bound state $(a, v)$ through either beneficial antibody substitutions with probabilities $W^{v}_{a' \to a} = A_+$ (second term), or deleterious viral substitutions with probabilities $W^{v' \to v}_{a} = V_-$ (third term). By symmetry, we can assume that all the $n$ bound states ($\forall a = v$) have the same probability $P_{\text{st}}(a, v; \mathbf{B}) \equiv \pi_{\mathbf{B}}$, and all the $n(n-1)$ unbound states ($\forall a' \neq v$) have the same probability $P_{\text{st}}(a', v; \mathbf{U}) \equiv \pi_{\mathbf{U}}$. We can thus simplify eq. (7) and, by requiring the probabilities to be normalized, arrive at the expressions for the stationary state probabilities associated with being in a given bound or an unbound state,

$$\pi_{\mathbf{B}} = \frac{A_+ + V_-}{Z}, \qquad \pi_{\mathbf{U}} = \frac{A_- + V_+}{Z}, \tag{8}$$

where $Z = n(A_+ + V_-) + n(n-1)(A_- + V_+)$ is the normalization constant.

**Probability current**. The probability current $J^{v' \to v}_{a' \to a}$ between states $(a', v') \to (a, v)$ quantifies the change in the probability density:

$$\frac{\mathrm{d}}{\mathrm{d}t} P(a,v) \equiv \sum_{a',v'} J^{v' \to v}_{a' \to a} = \sum_{a',v'} P(a',v') W^{v' \to v}_{a' \to a} - P(a,v) W^{v \to v'}_{a \to a'}. \tag{9}$$

Although in the stationary state (i.e., when $\mathrm{d}_t P(a, v) = 0$), the total probability current to any state is zero, the individual elements of the probability current matrix $J^{v' \to v}_{a' \to a}$ can still be non-zero, consistent with the fact that the system is out of equilibrium, i.e., it does not satisfy detailed balance. In our bipartite system, the probability current splits into partial components associated with changes in the antibodies $J^{v}_{a' \to a} = P(a', v) W^{v}_{a' \to a} - P(a, v) W^{v}_{a \to a'}$ and in the viral subsystem $J^{v' \to v}_{a} = P(a, v') W^{v' \to v}_{a} - P(a, v) W^{v \to v'}_{a}$, conditioned on the state of the other subsystem.

By inspection, one can see that the elements of the probability current (and the conditional components) are asymmetric, i.e., $J^{v' \to v}_{a' \to a} = -J^{v \to v'}_{a \to a'}$, and their magnitudes depend on the phenotypic effect of the transition. Therefore, given the bipartite nature of the system, the problem of computing the probability current elements reduces to computing the favorable and the neutral conditional currents; the currents associated with the unfavorable transitions are the negative of the inverted favorable

transitions. Using the definition of probability current in eq. (9), we can evaluate the conditional currents associated with (i) favorable changes in antibodies from the unbound state $(a', v)$ to the bound state $(a, v)$, i.e., $J^{v}_{a':\mathbf{U}\to a:\mathbf{B}}$, (ii) favorable changes in viruses from a bound to an unbound state, i.e., $J^{v':\mathbf{B}\to v:\mathbf{U}}_{a}$, and (iii) neutral changes in antibodies and viruses, denoted by $J^{v}_{a':\mathbf{U}\to a:\mathbf{U}}$ and $J^{v':\mathbf{U}\to v:\mathbf{U}}_{a}$, respectively. The elements of the probability current are thus given by

$$\begin{aligned} J^{v}_{a':\mathbf{U}\to a:\mathbf{B}} &= -J^{v}_{a:\mathbf{B}\to a':\mathbf{U}} = \pi_{\mathbf{U}} A_{+} - \pi_{\mathbf{B}} A_{-} = \frac{\sigma}{Z}, \\ J^{v':\mathbf{B}\to v:\mathbf{U}}_{a} &= -J^{v:\mathbf{U}\to v':\mathbf{B}}_{a} = \pi_{\mathbf{B}} V_{+} - \pi_{\mathbf{U}} V_{-} = \frac{\sigma}{Z}, \\ J^{v}_{a':\mathbf{U}\to a:\mathbf{U}} &= J^{v':\mathbf{U}\to v:\mathbf{U}}_{a} = 0, \end{aligned} \tag{10}$$

where $\sigma = A_{+}V_{+} - A_{-}V_{-}$, and $Z$ is the normalization constant from eq. (8). The probability current matrix takes a simple form in which the elements associated with favorable, unfavorable, and neutral transitions are $\sigma/Z$, $-\sigma/Z$, and 0, respectively.

### B. Fitness and transfer flux as rates of entropy production and information exchange in co-evolving systems

The two quantities of fitness and transfer flux have been previously introduced to measure the extent of non-equilibrium drive in evolving populations [5, 24, 25]. To characterize the out-of-equilibrium drive in our bipartite co-evolutionary system, we start with fitness flux, which measures the rate of adaptation in a population. Fitness flux was introduced in ref. [25] in analogy to the entropy production rate in non-equilibrium stochastic thermodynamics [30, 31]. Specifically, fitness flux is a measure for the irreversibility of information entropy through adaptation of a population, e.g., for evolution in time-dependent environments. The applicability of fitness flux to co-evolutionary adaptation has been specifically discussed in ref. [5]. As shown in Appendix B, the total rate of entropy production in our bipartite system can be expressed as

$$\begin{aligned} \Phi &= \sum_{a'\geq a, v'\geq v} J^{v'\to v}_{a'\to a} \ln \frac{W^{v'\to v}_{a'\to a}\ P(a', v')}{W^{v\to v'}_{a\to a'}\ P(a, v)} \\ &= \underbrace{\sum_{a'\geq a, v} J^{v}_{a'\to a} \ln \frac{W^{v}_{a'\to a}\ P(a', v)}{W^{v}_{a\to a'}\ P(a, v)}}_{=\Phi^{A}} \\ &\quad + \underbrace{\sum_{a, v'\geq v} J^{v'\to v}_{a} \ln \frac{W^{v'\to v}_{a}\ P(a, v')}{W^{v\to v'}_{a}\ P(a, v)}}_{=\Phi^{V}}, \end{aligned} \tag{11}$$

where we defined the conditional entropy production rates associated with changes in the antibody subsystem $\Phi^{A}$ and the viral subsystem $\Phi^{V}$, conditioned on the state of the other subsystem. The entropy production rate measures the irreversibility of a stochastic process by a Kullback-Leibler divergence between the probability of the forward path (here $W^{v'\to v}_{a'\to a}\ p(a', v')$) and that of the backward path (here $W^{v\to v'}_{a\to a'}\ p(a, v)$). Therefore, the entropy production rate $\Phi$, the conditional rates $\Phi_{A}$ and $\Phi_{V}$, and by association fitness flux, are non-negative and only zero at equilibrium, where reversibility requires the probabilities of the forward and backward trajectories to be equal. In our bipartite model of co-evolution, the conditional entropy production rates quantify the response of each subsystem to the changes caused by the evolution of the other subsystem. Importantly, this response contains the effect of feedback on the adaptive dynamics of each subsystem during co-evolution. Specifically, the effect of feedback is captured by considering the joint state probability $P(a, v)$, when calculating the entropy production rate in eq. (11). If we were to consider the antibody subsystem evolving in an independently time-varying viral environment (i.e., by neglecting the feedback), we would have expressed the probability for a transition $a' \to a$ in the environment $v$ by $W^{v}_{a'\to a}P(a')$, where $P(a')$ is the stationary state probability for the system to be in state $a'$, independently of the state of the environment. Similar modification applies to the backward trajectories, and to the transitions in the viral subsystem. Therefore, we can express the contribution from co-evolutionary feedback to the conditional entropy production rate as

$$\begin{aligned} \Phi^{A} &= \sum_{a'\geq a, v} J^{v}_{a'\to a} \ln \frac{W^{v}_{a'\to a}\ P(v|a')P(a')}{W^{v}_{a\to a'}\ P(v|a)P(a)} \\ &= \underbrace{\sum_{a'\geq a, v} J^{v}_{a'\to a} \ln \frac{W^{v}_{a'\to a}\ P(a')}{W^{v}_{a\to a'}\ P(a)}}_{\equiv\Phi^{A}_{0}} - \underbrace{\sum_{a'\geq a, v} J^{v}_{a'\to a} \ln \frac{P(v|a)}{P(v|a')}}_{=\delta I^{A}}, \end{aligned} \tag{12}$$

where $\Phi^{A}_{0}$ is the entropy production of the antibody subsystem without considering the effect of feedback. Interestingly, the remaining quantity that captures the effect of feedback, $\delta I^{A}$, is the conditional mutual information flux [36] due to the change in antibodies, conditioned on the state of viruses (see Appendix B for derivation); analogous quantities can be defined for viruses. The conditional mutual information flux $\delta I^{A}$ quantifies how changes in antibodies impact the phenotypic state and consequently the fitness of viruses; this measure is similar to transfer flux in co-evolving populations, previously introduced in ref. [5].

In our bipartite co-evolutionary model, the conditional entropy production rates (i.e., fitness flux) and the conditional mutual information flux associated with each sub-

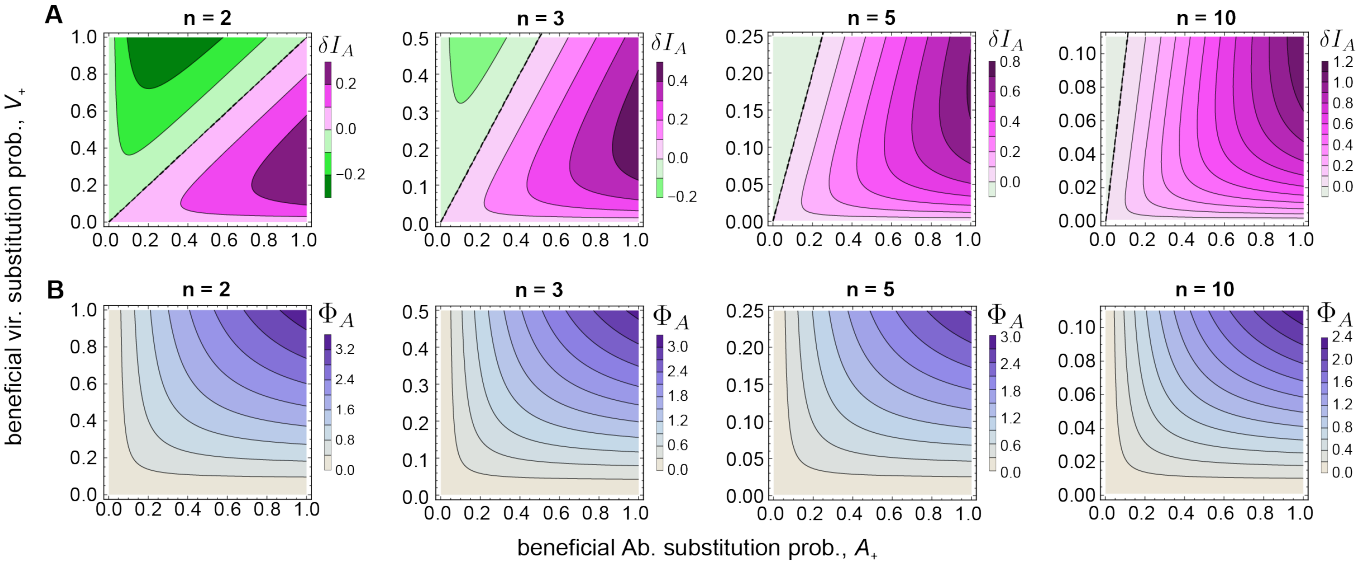


Figure 2. **Co-evolutionary information transfer and fitness flux.** Heatmaps show **(A)** the conditional information flux due to antibody evolution $\delta I_A$ (eq. (14)), which is the measure of the transfer flux from antibodies to viruses [5], and **(B)** the conditional entropy production rate of antibodies $\Phi_A$ (eq. (13)), which is the measure of antibodies' fitness flux [5, 25] in response to viral escape, as a function of the antibody and viral beneficial substitution probabilities, $A_+$, and $V_+$, respectively. The deleterious substitution probabilities are kept fixed $A_- = V_- = 10^{-3}$. Panels show results for different number of states in the balanced system $n_A = n_V = n$. The upper bound for the viral beneficial substitution rate is set using eq. 6 to assure that the transition probabilities are appropriately bounded (eq (4)). The dashed lines indicate the diagonal on which $A_+ = V_+$.

system (i.e., transfer flux) are

$$\Phi^A = n(n-1)\frac{\sigma}{Z}\left[\ln\frac{A_+}{A_-} + \ln\frac{\pi_{\mathbf{U}}}{\pi_{\mathbf{B}}}\right] \geq 0,$$
$$\Phi^V = n(n-1)\frac{\sigma}{Z}\left[\ln\frac{V_+}{V_-} + \ln\frac{\pi_{\mathbf{B}}}{\pi_{\mathbf{U}}}\right] \geq 0, \tag{13}$$

$$\delta I^A = -\delta I^V = n(n-1)\frac{\sigma}{Z}\ln\frac{\pi_{\mathbf{B}}}{\pi_{\mathbf{U}}}, \tag{14}$$

where $\sigma = A_+V_+ - A_-V_-$, and $Z$ is the normalization constant as in eq. (10). Fig. 2 shows the conditional information flux (A) and the entropy production rate (B) due to antibody changes for varying levels of antibody and viral substitution probabilities. Notably, the magnitude of both fluxes increases with the number of states, scaling as $n(n-1)$, which corresponds to the total number of possible transitions.

As expected, in the steady state, the change in the total mutual information, i.e., $\delta I = \delta I^A + \delta I^V$ is zero. Notably, the sign of the conditional mutual information fluxes, determined by the ratio of the steady-state probabilities $\pi_{\mathbf{B}}/\pi_{\mathbf{U}}$, distinguishes between the leader and the follower in this co-evolutionary arms race. When $\pi_{\mathbf{B}} > \pi_{\mathbf{U}}$, being in a given bound state is more likely than being in a given unbound state, not withstanding the fact that the number of unbound states is larger. In this case, viral escape destroys the mutual information $\delta I^V < 0$ by introducing a disorder in an otherwise bound configuration. Antibodies, on the other hand, increase the mutual information $\delta I^A > 0$ by learning and aligning themselves to newly evolved viral states, restoring the bound state. In this case, the virus is "leading" the evolutionary trajectory while the immune system is "following." On the other hand, if $\pi_{\mathbf{B}} < \pi_{\mathbf{U}}$, then the signs and roles are reversed, and changes in viruses create information, while antibodies erase information.

The distinction between the two scenarios leading to the different drivers of the co-evolutionary process is a matter of terminology (i.e., what is called host vs. virus or bound vs. unbound) rather than reflecting different biologically relevant regimes. As shown in Fig. 2A, the conditional information flux of antibodies $\delta I_A$ is anti-symmetric (i.e., the leader becomes the follower) with respect to the switch in the prominence of the bound vs. unbound state ($\pi_{\mathbf{B}}/\pi_{\mathbf{U}}$); this is captured in Fig. 2 by varying the relative beneficial substitution probabilities $A_+/V_+$, while keeping the deleterious rates constant; see eq. (8). As the number of states $n$ increases, the number of possible beneficial states a virus can transition to grows to $(n-1)$, while it remains fixed at one for antibodies (Fig. 1B). As a result, the upper bound on the beneficial viral substitution probability $V_+ \leq \frac{1-(n-1)A_-}{n-1}$ (eq. 6) becomes smaller, reducing the size of the region with permutational symmetry, as seen in Fig. 2A. From our biological intuition, the number of possible escape routes for the virus is generally large $n \gg 2$. Moreover, immune repertoires (e.g., antibodies) are diverse enough that for any given pathogenic variant a desired immune

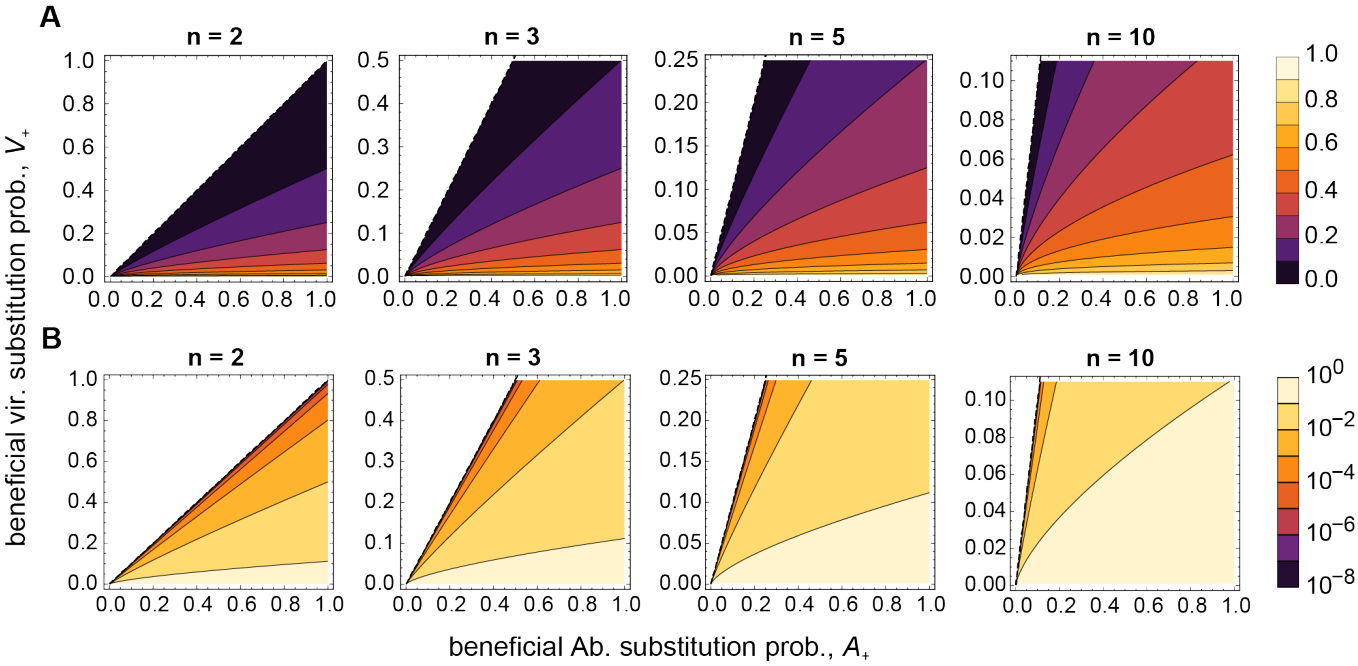


Figure 3. **Co-evolutionary efficiency.** **(A)** The efficiency of antibodies $\eta_A$ (eq. (15)) and **(B)** the total efficiency of the coevolving system $\eta = \eta_A \eta_V$ (eq. 17) are shown as a function of the antibody and viral beneficial substitution probabilities, $A_+$, and $V_+$, respectively, while keeping the deleterious substitution probabilities fixed $A_- = V_- = 10^{-3}$. Panels show results for different number of states in the balanced system $n_A = n_V = n$. The upper bound for the viral beneficial substitution rate is set using eq. 6 to assure that the transition probabilities are appropriately bounded (eq. (4)). The dashed lines indicate the $A_+ = V_+$ diagonal. We consider the case where the probability associated with a bound configuration is larger than that of an unbound configuration $\pi_{\mathbf{B}} > \pi_{\mathbf{U}}$, i.e., the region below the dashed lines.

solution can be found. Therefore, in the steady state of our co-evolutionary dynamics, and prior to the escape of the virus, the bound configuration is the more likely state. Thus, we will focus on the case of $\pi_{\mathbf{B}} > \pi_{\mathbf{U}}$ to better interpret our results (below the dashed lines in Fig. 2A). However, similar arguments apply to the opposite case after switching the virus and antibody labels.

### C. Efficiency of co-evolutionary information processing

Let us consider the case in which viruses lead the co-evolutionary arms race by escaping the predominantly bound configurations, and antibodies follow them. Given that $\delta I^A > 0$, eq. (12) suggests that the coupling (feedback) between antibodies and viruses leads to a reduction in the rate of adaptation (fitness flux) of antibodies in the co-evolving system compared to that of a system without feedback, i.e., $\Phi^A < \Phi_0^A$. In other words, having knowledge about the state of the viral ensemble decreases the extent of necessary adaptation (and entropy production) by antibodies. We can define the efficiency of antibodies in utilizing the co-evolutionary information by the relative reduction in their entropy production (fitness flux),

$$\eta_A = \frac{\Phi_0^A - \Phi^A}{\Phi_0^A} = \frac{\delta I^A}{\Phi^A + \delta I^A} = \frac{\ln \frac{\pi_{\mathbf{B}}}{\pi_{\mathbf{U}}}}{\ln \frac{A_+}{A_-}} \leq 1. \tag{15}$$

In the regime where the evolutionary rates of the antibodies are much larger than that of viruses $A_+, A_- \gg V_+, V_-$, we will have $\eta_A \to 1$, implying that the antibodies are very efficient in sensing and adapting to the viral state (Fig. 3A). As the number of states $n$ increases, the upper bound for the beneficial viral substitution probability becomes smaller (eq. 6), resulting in a broader region for efficient antibody responses (Fig. 3A).

Viruses erase the information generated by antibodies and adapt to escape the immune system. We can define the efficiency of the virus as the amount of information erasure $|\delta I^V|$ per unit of viral adaptation $\Phi^V$,

$$\eta_V = \frac{|\delta I^V|}{\Phi^V} = \frac{\ln \frac{\pi_{\mathbf{B}}}{\pi_{\mathbf{U}}}}{\ln \frac{V_+}{V_-} + \ln \frac{\pi_{\mathbf{B}}}{\pi_{\mathbf{U}}}} \leq 1. \tag{16}$$

In the regime where the evolutionary rates of viruses are much larger than that of the antibodies $V_+, V_- \gg A_+, A_-$, we will have $\eta_V \to 1$, implying that viruses are very efficient in erasing information.

The efficiency of the joint system can be defined as

$$\eta = \eta_A \eta_V = \frac{1}{\left(1 + \frac{\ln \frac{V_+}{V_-}}{\ln \frac{\pi_{\mathbf{B}}}{\pi_{\mathbf{U}}}}\right) \frac{\ln \frac{A_+}{A_-}}{\ln \frac{\pi_{\mathbf{B}}}{\pi_{\mathbf{U}}}}} \leq 1. \tag{17}$$

Note that when the probability of being in an unbound state approaches that of a bound state $\pi_{\mathbf{U}} \to \pi_{\mathbf{B}}$, the efficiency of the joint process goes to its minimum $\eta \to 0$; see Fig. 3B. In this regime, all the adaptive advantage that antibodies gain are erased by viral escape, driving the two populations into a continuous evolutionary arms race. Indeed, a balanced adaptive pressure of antibodies and viruses was previously suggested as a criterion to achieve a stable co-evolutionary arms race in ref. [5].

In the regime where $\pi_{\mathbf{B}} \gg \pi_{\mathbf{U}}$ with $A_+/A_- \gg V_+/V_-$ antibodies can efficiently neutralize viruses, resulting in the maximum efficiency of the joint process that is bounded by $\eta_{\max} = \frac{\log A_+/A_-}{\log A_+/A_- + \log V_+/V_-}$. The quantities $A_+/A_-$ and $V_+/V_-$ define the adaptive potential for antibodies and viruses, respectively. In such imbalanced scenario antibody adaptation becomes the dominant process, leading to the elimination of viruses.

It should be noted that the definition of efficiency introduced here is distinct from that of ref. [36], which considers a physical bipartite system in which the relationship between work, heat dissipation, and (information) entropy production defines the extent of irreversibility in the system's trajectories. The out-of-equilibrium co-evolutionary process here however does not satisfy a constraint similar to the first law of thermodynamics, and thus physical work is not required for the system to process information, as suggested by the Landauer's principle [39]. The measure of efficiency introduced here is suitable for non-physical stochastic processes, in which the adaptive potential in one system can determine its efficiency in processing the information generated by the other.

## IV. DISCUSSION

In this study, we introduced a novel framework to model immune-pathogen co-evolution, using a bipartite Markov process to capture the interactions between antibodies and viruses. Our model provides a simplified yet powerful approach to understanding the evolutionary arms race that governs the dynamics of host-pathogen systems. By considering the processes of beneficial, deleterious, and neutral substitutions within this bipartite structure, we established key metrics like information flux, fitness flux, and co-evolutionary efficiency, which enable us to quantify the out-of-equilibrium behavior underlying immune-pathogen co-evolution, driven by a continuous exchange of information.

A key finding is the identification of co-evolutionary fitness flux and information flux as measures of adaptation and interaction in the coupled immune-pathogen system. The asymmetry in the information flux, measured by the changes in the conditional mutual information due to the evolution of one subpopulation, allows us to distinguish between the leader and follower of a co-evolutionary arms race. In scenarios where viruses rapidly mutate to evade immune detection, they act as the leaders, with antibodies adapting to restore a bound configuration to neutralize them, thus following the viral evolution.

Our model highlights the use of co-evolutionary efficiency in determining how each population exploits the information to adapt more effectively towards their desired states. Specifically, we define efficiency as the relative reduction in the information entropy production of each population due to the feedback from the other. Antibodies, for example, benefit from this feedback, reducing their rate of adaptation when they have information about the state of the viral population.

Prior work has focused on how viral mutation rates are set by the balance between the benefits of increased mutation for immune evasion and the costs of harmful off-target mutations, known as genetic load [40–42]. Our framework suggests an alternative approach to understanding how mutations rates, or more precisely substitution rates, are to be modulated. We show that a co-evolutionary system's overall efficiency is determined by the balance between the adaptation rates of the virus and the antibody populations. Specifically, when the relative beneficial to deleterious substitution rates for both populations are comparable the joint system reaches it minimum efficiency. In this regime, the adaptive advantage that antibodies gain is erased by viral escape, driving the two populations to maintain a constant evolutionary arms race.

While it is challenging to directly measure the co-evolutionary efficiency of viral-immune systems, insights can be gained through the study of endogenous viral elements (EVEs). EVEs are segments of viral DNA that have become integrated into the genomes of their hosts and are transmitted across generations, serving as fossil records of ancient viruses over millions of years. Once integrated, EVEs remain highly stable and can be used to trace the evolutionary origins of different viruses through phylogenetic analyses of host populations. For instance, although HIV-1 was recently introduced into the human population, its origins can be traced back to endogenous lentiviruses in rabbits and other mammals, with its close relative, the simian immunodeficiency virus, dating back over 10,000 years.

Analysis of sequence divergence of mammalian viruses reveals that over short timescales viral evolution tends to exhibit rapid rates of sequence divergence. In contrast, when divergence is measured from EVEs over much longer timescales, the observed evolutionary rates are significantly slower, approaching that of the host's genome [38]. The inherent biases in these data– e.g., sequences that can be aligned over long evolutionary times ought to be similar– make it difficult to draw definitive conclusions about the evolution of viral substitution

rates. Nonetheless, this pattern raises the question as whether viruses that have established long-term coexistence with their hosts have evolved to balance their beneficial and deleterious substitution rates with that of the host's immune updates. Our measure of co-evolutionary efficiency may guide studies of EVEs in host genomes, especially from ancient DNA, to address these questions.

Prior work on stochastic thermodynamics has extended the concept of efficiency for information exchange [36, 43]. However, they consider physical systems, in which the energy is conserved (first law of thermodynamics), and therefore, quantities like work and heat are well-defined. Immune-pathogen coevolution is a specific example of an interacting non-physical stochastic system with information exchange that does not follow the conservation law for an energy-like quantity. Therefore, the notion of efficiency introduced here can be used more broadly for non-physical stochastic systems like those in economics, financial markets, linguistics, and artificial neural networks.

Our model presents a highly simplified description of immune-pathogen co-evolution. Combining this framework with a more realistic description of the evolution and ecology of interacting populations could shed light on how factors such as population diversity, demographic structures, immune memory and cross-immunity influence the co-evolutionary process. Additionally, while we focused on the symmetric case where the number of antibody and viral states are balanced, the imbalanced case could be explored further to understand evolutionary distraction for antibodies.

## V. ACKNOWLEDGMENT

This work has been supported by the CAREER award from the National Science Foundation (grant No: 2045054), and the MPRG funding from the Max Planck Society to A.N.. This work is also supported, in part, through the Department of Physics and the College of Arts and Sciences at the University of Washington.

[1] C. A. Janeway, Jr, P. Travers, M. Walport, and M. J. Shlomchik, *Immunobiology* (Garland Science, 2001).
[2] H. G. Hampton, B. N. J. Watson, and P. C. Fineran, Nature **577**, 327 (2020).
[3] S. Bradde, A. Nourmohammad, S. Goyal, and V. Balasubramanian, Proc. Natl. Acad. Sci. U. S. A. **117**, 5144 (2020).
[4] B. N. J. Watson, J. A. Steens, R. H. J. Staals, E. R. Westra, and S. van Houte, Cell Host Microbe **29**, 715 (2021).
[5] A. Nourmohammad, J. Otwinowski, and J. B. Plotkin, PLoS Genet. **12**, e1006171 (2016).
[6] H.-X. Liao, R. Lynch, T. Zhou, F. Gao, S. M. Alam, S. D. Boyd, A. Z. Fire, K. M. Roskin, C. A. Schramm, Z. Zhang, J. Zhu, L. Shapiro, J. C. Mullikin, S. Gnanakaran, P. Hraber, K. Wiehe, G. Kelsoe, G. Yang, S.-M. Xia, D. C. Montefiori, R. Parks, K. E. Lloyd, R. M. Scearce, K. A. Soderberg, M. Cohen, G. Kamanga, M. K. Louder, L. M. Tran, Y. Chen, F. Cai, S. Chen, S. Moquin, X. Du, M. G. Joyce, S. Srivatsan, B. Zhang, A. Zheng, G. M. Shaw, B. H. Hahn, T. B. Kepler, B. T. M. Korber, P. D. Kwong, J. R. Mascola, and B. F. Haynes, Nature **496**, 469 (2013).
[7] S. Luo and A. S. Perelson, Proceedings of the National Academy of Sciences **112**, 11654 (2015), https://www.pnas.org/doi/pdf/10.1073/pnas.1505207112.
[8] S. Wang, J. Mata-Fink, B. Kriegsman, M. Hanson, D. J. Irvine, H. N. Eisen, D. R. Burton, K. D. Wittrup, M. Kardar, and A. K. Chakraborty, Cell **160**, 785 (2015).
[9] A. Nourmohammad, J. Otwinowski, M. Łuksza, T. Mora, and A. M. Walczak, Mol. Biol. Evol. **36**, 2184 (2019).
[10] N. Strauli, E. K. Fryer, O. Pham, M. Abdel-Mohsen, S. N. Facente, C. Pilcher, P. Pennings, S. Pillai, and R. D. Hernandez, bioRxiv , 646968 (2019).
[11] A. Mazzolini, T. Mora, and A. M. Walczak, Philos. Trans. R. Soc. Lond. B Biol. Sci. **378**, 20220056 (2023).
[12] G. D. Victora and M. C. Nussenzweig, Annu. Rev. Immunol. **40**, 413 (2022).
[13] P. Barrat-Charlaix and R. A. Neher, eLife (2024), 10.7554/elife.97350.1.
[14] M. Meijers, D. Ruchnewitz, J. Eberhardt, M. Łuksza, and M. Lässig, Cell **186**, 5151 (2023).
[15] J. Marchi, M. Lässig, A. M. Walczak, and T. Mora, Proc. Natl. Acad. Sci. U. S. A. **118** (2021), 10.1073/pnas.2103398118.
[16] O. H. Schnaack and A. Nourmohammad, Elife **10** (2021), 10.7554/eLife.61346.
[17] O. H. Schnaack, L. Peliti, and A. Nourmohammad, Phys. Rev. X **12**, 021063 (2022).
[18] V. Chardès, A. Mazzolini, T. Mora, and A. M. Walczak, Proc. Natl. Acad. Sci. U. S. A. **120**, e2307712120 (2023).
[19] N. H. Barton and H. P. de Vladar, Genetics **181**, 997 (2009).
[20] A. Nourmohammad, S. Schiffels, and M. Lässig, J. Stat. Mech. **2013**, P01012 (2013).
[21] A. Nourmohammad, T. Held, and M. Lässig, Curr. Opin. Genet. Dev. **23**, 684 (2013).
[22] G. Sella and A. E. Hirsh, Proc. Natl. Acad. Sci. U. S. A. **102**, 9541 (2005).
[23] V. Mustonen and M. Lässig, Proc. Natl. Acad. Sci. U. S. A. **102**, 15936 (2005).
[24] V. Mustonen and M. Lässig, Trends Genet. **25**, 111 (2009).
[25] V. Mustonen and M. Lässig, Proc. Natl. Acad. Sci. U. S. A. **107**, 4248 (2010).
[26] T. Held, A. Nourmohammad, and M. Lässig, J. Stat. Mech. **2014**, P09029 (2014).
[27] T. J. Kobayashi and Y. Sughiyama, Phys. Rev. Lett. **115**, 238102 (2015).
[28] Y. Sughiyama and T. J. Kobayashi, Phys. Rev. E **95**, 012131 (2017).
[29] A. Nourmohammad, J. Rambeau, T. Held, V. Kovacova, J. Berg, and M. Lässig, Cell Rep. **20**, 1385 (2017).

[30] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[31] G. E. Crooks, Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics **60**, 2721 (1999).
[32] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[33] T. Sagawa and M. Ueda, Phys. Rev. Lett. **109**, 180602 (2012).
[34] T. Sagawa and M. Ueda, New J. Phys. **15**, 125012 (2013).
[35] A. C. Barato and U. Seifert, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **90**, 042150 (2014).
[36] J. M. Horowitz and M. Esposito, Phys. Rev. X **4**, 031015 (2014).
[37] M. Kimura, Nature **217**, 624 (1968).
[38] P. Simmonds, P. Aiewsakun, and A. Katzourakis, Nat. Rev. Microbiol. **17**, 321 (2019).
[39] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[40] J. J. Bull, R. Sanjuán, and C. O. Wilke, J. Virol. **81**, 2930 (2007).
[41] K. Koelle and D. A. Rasmussen, Elife **4**, e07361 (2015).
[42] R. Swanstrom and R. F. Schinazi, Science **375**, 497 (2022).
[43] M. P. Leighton and D. A. Sivak, Phys. Rev. Lett. **130**, 178401 (2023).

### Appendix A: Transition matrix for imbalanced number states for the antibodies and viruses

Because the viral and antibody states can be imbalanced, i.e., $n_V \neq n_A$ in general, we distinguish between two types of unbound states. In the case that $n_V > n_A$, there exists states that are inaccessible to the antibodies (i.e., $v = n_A + 1, \ldots, n_V$), and if the virus transitions to any of them, no change in the antibody state can result in binding, i.e., the virus fully hides from antibodies. On the other hand, if $n_A > n_V$, for each viral state there exists more antibody states that do not bind to the virus compared to the balanced case of $n_A = n_V$, resulting in more evolutionary distraction for antibodies. To reflect these differences, it is useful to introduce three phenotypic states: bound $\mathbf{B}$, unbound and accessible $\mathbf{U}$, and unbound and inaccessible $\not\mathbf{U}$ (only relevant for the imbalanced case). Defining, $n_- = \min(n_A, n_V)$, and $n_+ = \max(n_A, n_V)$, a system has $n_-$ bound states, $n_-(n_- - 1)$ accessible unbound states, and $n_+ n_- - n_-^2$ inaccessible unbound states.

Similar to the balanced case, we can invoke symmetry to argue that the diagonal terms of the matrix $W_{a\to a}^{v\to v}$ can be put into three classes, the ones associated with a bound state $a = v$, which we denote by $W_{\mathbf{B}}$, the ones associated with accessible unbound states ($a \neq v \wedge a, v \leq n_-$), which we denote by $W_{\mathbf{U}}$, and those associated with inaccessible unbound states ($a \neq v \wedge a | v > n_-$), which we denote by $W_{\not\mathbf{U}}$. These three classes of diagonal elements are given by

$$
\begin{aligned}
W_{\mathbf{B}} &= 1 - \left[(n_V - 1)V_+ + (n_A - 1)A_-\right], \\
W_{\mathbf{U}} &= 1 - \left[V_- + A_+ + (n_A - 2)A_0 + (n_V - 2)V_0\right], \\
W_{\not\mathbf{U}} &= 1 - \begin{cases} V_- + (n_A - 1)A_0 + (n_V - 2)V_0, & \text{if } n_v > n_A \\ A_+ + (n_A - 2)A_0 + (n_V - 1)V_0, & \text{if } n_A > n_V. \end{cases}
\end{aligned}
\tag{A1}
$$

In evaluating these diagonal elements, one needs to account for all the transitions out of the specified state. For example, for $W_{\mathbf{B}}$ one should account for all the transitions out of a bound state. Since our model allows for only one bound state for a given antibody and virus, i.e., the state $a = v$, all transitions from a bound state would end in an unbound configuration. There are $(n_V - 1)$ such changes through beneficial changes in viruses, each with rate $V_+$, and $(n_A - 1)$ such transitions through deleterious changes in antibodies with a rate $A_-$, as reflected in the expression for $W_{\mathbf{B}}$. Similarly, $W_{\mathbf{U}}$ and $W_{\not\mathbf{U}}$ can be calculated by accounting for all the possible transitions, including the neutral transitions to another unbound state (with antibody and viral rates $A_0$ and $V_0$), as shown in the expression for these diagonal terms.

In this general case with an imbalanced number of states ($n_A \neq n_V$), the boundedness of the transition probabilities implies an upper bound for the beneficial substitution probability of viruses, $V_+ \leq \frac{1-(n_A-1)A_-}{n_V-1}$. Similarly, the upper bound for the beneficial substitution probability of antibodies is $A_+ \leq 1 - (V_- + (n_A - 2)A_0 + (n_V - 2)V_0)$. The neutral substitution probabilities, $A_0$ and $V_0$, scale as $1/N_{\text{eff}}$ with the effective population size of each system $N_{\text{eff}}$ [37], which is likely to be much larger than the dimensionality of the subsystems, $n_A$ and $n_V$, yielding $(n_A - 2)A_0 + (n_V - 2)V_0 \ll 1$. Given that the deleterious substitution probabilities are exponentially small $V_- \ll 1$ [37], the upper bound on the beneficial substitution probability for antibodies can be approximated as $A_+ \lesssim 1$, with no strong dependence on the dimensionality of the system.

### Appendix B: Entropy production and information exchange for a bipartite co-evolutionary system

**Probability flux.** We express the probability flux to the state $(a, v)$ by

$$\frac{\mathrm{d}}{\mathrm{d}t}P(a,v) = \sum_{a',v'} J^{v'\to v}_{a'\to a} \tag{B1}$$

$$= \sum_{a'\neq a; v'\neq v} P(a',v')W^{v'\to v}_{a'\to a} - P(a,v)W^{v\to v'}_{a\to a'} \tag{B2}$$

$$= \sum_{a'\neq a} \underbrace{P(a',v)W^{v}_{a'\to a} - P(a,v)W^{v}_{a\to a'}}_{J^{v}_{a'\to a}} + \sum_{v'\neq v} \underbrace{P(a,v')W^{v'\to v}_{a} - P(a,v)W^{v\to v'}_{a}}_{J^{v'\to v}_{a}}, \tag{B3}$$

where $J^{v'\to v}_{a'\to a}$ is the flux from the state $(a', v')$ to the state $(a, v)$, and $J^{v}_{a'\to a}$ and $J^{v'\to v}_{a}$ are the conditional probability fluxes due to the changes in the antibodies and viruses, respectively.

**Change in the Shannon entropy.** The Shannon entropy of the joint system is $S = -\sum_{a,v} P(a,v)\ln P(a,v)$. The change in the Shannon entropy is given by

$$\begin{aligned}
\mathrm{d}_t S &= \mathrm{d}_t\left[-\sum_{a,v} P(a,v)\ln P(a,v)\right] \\
&= -\sum_{a,v} \mathrm{d}_t P(a,v)\ln P(a,v) + \cancelto{0}{\sum_{a,v} \mathrm{d}t P(a,v)} \\
&= -\sum_{a,v,a',v'} J^{v'\to v}_{a'\to a}\ln P(a,v) \\
&= -\frac{1}{2}\sum_{a,v,a',v'}\left(J^{v'\to v}_{a'\to a} - J^{v\to v'}_{a\to a'}\right)\ln P(a,v) \\
&= -\frac{1}{2}\sum_{a,v,a',v'} J^{v'\to v}_{a'\to a}\left[\ln P(a,v) - \ln P(a',v')\right] \\
&= \sum_{a<a',v<v'} J^{v'\to v}_{a'\to a}\ln\frac{P(a',v')}{P(a,v)},
\end{aligned} \tag{B4}$$

where we used the fact that probability flux is asymmetric $J^{v'\to v}_{a'\to a} = -J^{v\to v'}_{a\to a'}$.

**Entropy production rate.** The total entropy production rate $\Phi$ is equal to the sum of the change in the Shannon entropy of the system $\mathrm{d}_t S$ (eq. (B4)) and the total rate of the change in the environment's entropy, which is equal to $\dot{S}_{\mathrm{env.}} = \sum_{a<a',v<v'} J^{v'\to v}_{a'\to a}\ln\left(W^{v'\to v}_{a'\to a}/W^{v\to v'}_{a\to a'}\right)$. Therefore, the total entropy production rate follows

$$\begin{aligned}
\Phi &= \mathrm{d}_t S + \dot{S}_{\mathrm{env.}} \\
&= \sum_{a<a',v<v'} J^{v'\to v}_{a'\to a}\ln\frac{P(a',v')}{P(a,v)} + \sum_{a<a',v<v'} J^{v'\to v}_{a'\to a}\ln\frac{W^{v'\to v}_{a'\to a}}{W^{v\to v'}_{a\to a'}} \\
&= \sum_{a<a',v<v'} J^{v'\to v}_{a'\to a}\ln\frac{P(a',v')\,W^{v'\to v}_{a'\to a}}{P(a,v)\,W^{v\to v'}_{a\to a'}}
\end{aligned} \tag{B5}$$

$$= \underbrace{\sum_{a<a',v} J^{v}_{a'\to a}\ln\frac{P(a',v)\,W^{v}_{a'\to a}}{P(a,v)\,W^{v}_{a\to a'}}}_{\Phi^A} + \underbrace{\sum_{a,v<v'} J^{v'\to v}_{a}\ln\frac{P(a,v')\,W^{v'\to v}_{a}}{P(a,v)\,W^{v\to v'}_{a}}}_{\Phi^V}, \tag{B6}$$

where we used the bipartite property of the system to define the conditional entropy production rates $\Phi^A$ and $\Phi^V$ due to changes in the antibody and viral ensembles, respectively.

We can evaluate the steady-state conditional entropy production rates in our antibody-viral system. Let us start with $\Phi^A$. By inspecting eq. (B6), we can see that the argument in the sum is symmetric under the state swap $a \leftrightarrow a'$. Given that the sum is avoiding the repetition in the indices and the fact that the probability current $J$ is only non-zero between bound and unbound states, we can rewrite the sum in the phenotypic space by conditioning one index to be associated to the bound state and the other index to all the unbound states. Thus, the conditional entropy production $\Phi^A$ associated with antibody changes is given by

$$\begin{aligned}
\Phi^A &= \sum_{a' \geq a, v} J^v_{a' \to a} \ln \frac{W^v_{a' \to a}\ P(a', v)}{W^v_{a \to a'}\ P(a, v)} \\
&= \sum_{v, a:\mathbf{B}, a':\mathbf{U}} J^v_{a':\mathbf{U} \to a:\mathbf{B}} \ln \frac{W^v_{a':\mathbf{U} \to a:\mathbf{B}}\ P(a', v; \mathbf{U})}{W^v_{a:\mathbf{B} \to a':\mathbf{U}}\ P(a, v; \mathbf{B})} \\
&= n(n-1) \frac{\sigma}{Z}\ \ln \frac{A_+\ \pi_{\mathbf{U}}}{A_-\ \pi_{\mathbf{B}}},
\end{aligned} \tag{B7}$$

where we have used the stationary state probabilities associated with bound and unbound states $\pi_{\mathbf{B}}$, and $\pi_{\mathbf{U}}$ (eq. (8)). The prefactor $n(n-1)$ counts for the summation over the $n$ different viral states and the $(n-1)$ possible antibody transitions from bound to unbound states for each of the viral states.

With a similar approach, we can evaluate the conditional entropy production rate associated with changes in the viral ensemble, obtaining

$$\begin{aligned}
\Phi^V &= \sum_{a, v' \geq v} J^{v' \to v}_a \ln \frac{W^{v' \to v}_a\ P(a, v')}{W^{v \to v'}_a\ P(a, v)} \\
&= \sum_{a, v':\mathbf{B}, v:\mathbf{U}} J^{v':\mathbf{B} \to v:\mathbf{U}}_a \ln \frac{W^{v':\mathbf{B} \to v:\mathbf{U}}_a\ \pi_{\mathbf{B}}}{W^{v:\mathbf{U} \to v':\mathbf{B}}_a\ \pi_{\mathbf{U}}} \\
&= n(n-1) \frac{\sigma}{Z}\ \ln \frac{V_+\ \pi_{\mathbf{B}}}{V_-\ \pi_{\mathbf{U}}}
\end{aligned} \tag{B8}$$

**Change in mutual information.** The mutual information flux for a bipartite system can be calculated in the following way:

$$\begin{aligned}
\mathrm{d}_t I &= \mathrm{d}_t \left[ \sum_{a,v} P(a,v) \ln \frac{P(a,v)}{P(a)P(v)} \right] \\
&= \sum_{a,v} \mathrm{d}_t P(a,v) \ln \frac{P(a,v)}{P(a)P(v)} + \cancelto{0}{\sum_{a,v} \mathrm{d}_t P(a,v) - \mathrm{d}_t P(a) - \mathrm{d}_t P(v)} \\
&= \sum_{a,v,a',v'} J^{v' \to v}_{a' \to a} \ln \frac{P(a,v)}{P(a)P(v)} \\
&= \frac{1}{2} \sum_{a,v,a',v'} \left( J^{v' \to v}_{a' \to a} - J^{v \to v'}_{a \to a'} \right) \ln \frac{P(a,v)}{P(a)P(v)} \\
&= \frac{1}{2} \sum_{a,v,a',v'} J^{v' \to v}_{a' \to a} \left( \ln \frac{P(a,v)}{P(a)P(v)} - \ln \frac{P(a',v')}{P(a')P(v')} \right) \\
&= \frac{1}{2} \sum_{a,v,a'} J^{v}_{a' \to a} \left( \ln \frac{P(a,v)}{P(a)P(v)} - \ln \frac{P(a',v)}{P(a')P(v)} \right) + \frac{1}{2} \sum_{a,v,v'} J^{v' \to v}_{a} \left( \ln \frac{P(a,v)}{P(a)P(v)} - \ln \frac{P(a,v')}{P(a)P(v')} \right) \\
&= \underbrace{\sum_{a<a',v} J^{v}_{a' \to a} \ln \frac{P(v|a)}{P(v|a')}}_{\delta I^A} + \underbrace{\sum_{v<v',a} J^{v' \to v}_{a} \ln \frac{P(a|v)}{P(a|v')}}_{\delta I^V}
\end{aligned} \tag{B9}$$