# A hybrid stochastic-deterministic approach to explore multiple infection and evolution in HIV

**Jesse Kreger[a], Natalia L. Komarova[a] and Dominik Wodarz[b,a]**

a) Department of Mathematics, University of California Irvine, Irvine CA 92697

b) Department of Population Health and Disease Prevention Program in Public Health
Susan and Henry Samueli College of Health Sciences, University of California, Irvine CA 92697

**Abstract**

To study viral evolutionary processes within patients, mathematical models have been instrumental. Yet, the need for stochastic simulations of minority mutant dynamics can pose computational challenges, especially in heterogeneous systems where very large and very small subpopulations coexist. Here, we describe a hybrid stochastic-deterministic algorithm to simulate mutant evolution in large viral populations, such as acute HIV-1 infection, and further include the multiple infection of cells. We demonstrate that the hybrid method can approximate the fully stochastic dynamics with sufficient accuracy at a fraction of the computational time, and quantify evolutionary end points that cannot be expressed by deterministic models, such as the mutant distribution or the probability of mutant existence at a given infected cell population size. We apply this method to study the role of multiple infection and intracellular interactions among different virus strains (such as complementation and interference) for mutant evolution. Multiple infection is predicted to increase the number of mutants at a given infected cell population size, due to a larger number of infection events. We further find that viral complementation can significantly enhance the spread of disadvantageous mutants, but only in select circumstances: it requires the occurrence of direct cell-to-cell transmission through virological synapses, as well as a substantial fitness disadvantage of the mutant, most likely corresponding to defective virus particles. This, however, likely has strong biological consequences because defective viruses can carry genetic diversity that can be incorporated into functional virus genomes via recombination. Through this mechanism, synaptic transmission in HIV might promote virus evolvability.

## 1 Introduction

The evolution of HIV-1 within patients is an important determinant of the disease process and of treatment outcomes [19, 13, 28]. Evolutionary changes in the virus population over time are thought to contribute to the progression of the infection from the asymptomatic phase towards AIDS [19], involving the evolution of immune escape as well as evolution towards faster replication, increased cytopathicity, and broader cell tropism [19]. The emergence of mutants that are resistant against anti-viral drugs can result in challenges to the long-term control of the infection. While viral evolution is important throughout the course of the disease, extensive virus replication towards relatively high viral loads during the acute phase of the infection presents ample opportunity for the generation of viral mutants that might influence post-acute setpoint virus load, the subsequent disease course, and the response to treatment [26].

Mathematical models have played a key role in defining the principles of within-host dynamics and evolution of HIV [30, 32, 22, 20, 21, 3, 9, 8, 7]. During the acute phase of the infection, however, the number of virus-infected cells can reach very large numbers [10, 27], while some subpopulations of importance may be very small, which presents computational problems. Mutant evolution can be driven by stochastic effects, because mutant viruses initially exist at low population sizes, even though the wild-type population can be very large. Stochastic simulations of the viral evolutionary dynamics thus become computationally costly if the overall viral population size

is large. To get around this, models can assume unrealistically low population sizes of infected cells, together with unrealistically large mutation rates, in the hope that the effects observed in such models scale up to more realistic population sizes and lower mutation rates. The accuracy of such explorations, however, is unclear. Alternatively, deterministic models in the form of ordinary differential equations (ODEs) can be used to approximate the average number of mutants over time as the virus population grows. The disadvantage of this approach is that other important evolutionary measures, such as the number of mutants at a given infected cell population size or the time of mutant generation, are not clearly defined in ODEs. Furthermore, the distribution of mutants at a given time or at a given infected cell population size cannot be determined with ODEs.

An interesting aspect that can influence the viral evolutionary dynamics, especially at large population sizes [25], is the multiple infection of cells [17, 18, 15, 16, 38]. Multiple infection has been documented to occur in HIV infection both in vitro [6, 35] and in vivo from human tissue samples [24], and is especially promoted by direct cell-to-cell transmission of the virus through virological synapses [14, 5, 1, 37]. In contrast to free virus transmission, synaptic transmission typically involves the simultaneous transfer of multiple viruses from the source cell to the target cell. As the virus grows to high levels, minority populations of multiply infected cells, which can be governed by stochastic effects, coexist with a much larger population of singly infected cells, which is again a computationally challenging situation. Interesting evolutionary dynamics can occur as a result of multiply infected cells, especially if different virus strains are present in the same cell. A disadvantageous mutant can gain fitness through complementation [12], and an advantageous mutant might experience fitness reduction due to interference by the wild-type virus [41]. Recombination can be another evolutionary consequence of multiple infection [25, 18, 29, 23].

This paper makes two contributions: (i) We describe a stochastic-deterministic hybrid method, which allows us to simulate the evolutionary dynamics of viruses at large population sizes (but in the presence of small subpopulations of evolutionary importance), including the possibility of multiple infection of cells. (ii) We apply this methodology to investigate the effect of multiple infection on mutant evolution during acute HIV infection. This includes an analysis of how intracellular interactions among viruses, such as complementation and interference, can influence evolutionary trajectories. In this context, special emphasis is placed on the direct cell-to-cell transmission of HIV through virological synapses [14, 5, 1, 37], because it has been shown that synaptic transmission not only promotes multiple infection, but can promote the repeated co-transmission of genetically distinct virus strains from one cell to the next. This in turn can enhance the potential of complementation and interference to impact mutant spread. In this paper, we do not focus on recombination processes, which were analyzed in a previous paper [23].

The paper starts by describing the basic mathematical model under consideration. This is followed by a description of the stochastic-deterministic hybrid methodology and a comparison of simulation results to both fully stochastic simulations and ODEs. Finally, we apply the hybrid methodology to study viral evolutionary dynamics during acute HIV infection, in the presence and absence of multiple infection, focusing on the role of viral complementation and interference. This work has relevance beyond HIV, because multiple infection and intracellular interactions (such as complementation and interference) can occur in other viruses, such as bacteriophages [41, 42]. Therefore, beyond parameter combinations that are relevant to HIV, we also explore wider parameter sets for broader relevance.

## 2 Methods

### 2.1 Mathematical model description

We begin with a deterministic model for HIV-1 infection, which includes both free virus transmission and synaptic cell-to-cell transmission [22]. We assume that cells are sufficiently well-mixed, such that relative spatial locations of cells do not play a significant role in the dynamics. To include the possibility of multiple infection, we let $x_i(t)$ represent the number of cells infected with $i$ copies of the virus at time $t$, where $i$ ranges from 0 (uninfected cells) to $N$ (cells infected with $N$ viruses). Descriptions of the model parameters can be found in Table 1. With only one viral strain, the ODE model (in its simplest formulation) is

$$\begin{aligned}
\dot{x}_0 &= \lambda - \beta Z x_0 - \gamma Z x_0 - d x_0, && (1)\\
\dot{x}_i &= \beta Z(x_{i-1} - x_i) + \gamma Z(-x_i) - a x_i, \text{ if } 0 < i < S && (2)\\
\dot{x}_i &= \beta Z(x_{i-1} - x_i) + \gamma Z(x_{i-S} - x_i) - a x_i, \text{ if } S \le i \le N - S && (3)\\
\dot{x}_i &= \beta Z(x_{i-1} - x_i) + \gamma Z(x_{i-S}) - a x_i, \text{ if } N - S < i < N && (4)\\
\dot{x}_N &= \beta Z(x_{N-1}) + \gamma Z(x_{N-S}) - a x_N, && (5)
\end{aligned}$$

where the number of infected cells is defined as

$$Z(t) = \sum_{i=1}^{N} x_i(t). \tag{6}$$

We assume that $N$, the maximum multiplicity of infection, is large enough to not result in a significant amount of cells near the end of the infection cascade. In the above equations, cell free virus transmission happens at rate $\beta$, and synaptic transmission at rate $\gamma$. Mathematically the two processes are similar, except that free virus transmission involves the entry of one virus into the target cells, while multiple viruses (on average $S$ viruses) can enter the target cell simultaneously during synaptic transmission. Therefore, in this model, the total number of infected cells is independent of transmission mode, and the transmission mode only alters the average infection multiplicity of cells (average number of virus copies in infected cells).

This model has a virus free steady state,

$$\begin{aligned}
x_0 &= \frac{\lambda}{d}, && (7)\\
Z &= 0, && (8)
\end{aligned}$$

and an infection steady state,

$$\begin{aligned}
x_0 &= \frac{a}{\beta + \gamma}, && (9)\\
Z &= \frac{\lambda}{a} - \frac{d}{\beta + \gamma}. && (10)
\end{aligned}$$

The stability of these steady states depends on the basic reproductive ratio, $R_0 = \frac{\lambda(\beta+\gamma)}{ad}$. If $R_0 < 1$, the virus free steady state is stable and if $R_0 > 1$ then the infection steady state is stable. Therefore, when considering the total virus population, the properties of this model are identical to those in standard virus dynamics models [30, 22].

This model can be adapted to describe competition among different virus strains, and mutational processes that give rise to mutant viral strains, thus allowing us to study the evolutionary dynamics of the virus. The basic assumptions of the model remain the same as described above,

| Notation | Description | Units (if applicable) |
|---|---|---|
| $\lambda$ | production rate of uninfected cells | days$^{-1}$ |
| $\beta$ | rate of free virus transmission | days$^{-1}$ |
| $\gamma$ | rate of synaptic cell-to-cell transmission | days$^{-1}$ |
| $d$ | death rate of uninfected cells | days$^{-1}$ |
| $a$ | death rate of infected cells | days$^{-1}$ |
| $x_0(t)$ | number of uninfected cells at time $t$ | NA |
| $x_i(t)$ | number of cells infected with $i$ copies of the virus at time $t$ | NA |
| $Z(t)$ | sum of all infected populations at time $t$, $Z(t) = \sum_{i=1}^{N} x_i(t)$ | NA |
| $Z_i(t)$ | sum of fraction of subpopulations infected with $i^{\text{th}}$ strain | NA |
| $N$ | maximum infection multiplicity | NA |
| $S$ | number of viruses transferred per synapse | NA |
| $\mu$ | mutation rate | NA |
| $\mathcal{M}$ | hybrid algorithm size threshold | NA |
| $F_i$ | fitness of the $i^{\text{th}}$ strain | NA |

Table 1: Description of model parameters and units (if applicable).

and the multi-strain model is described in detail in the Supplementary Information Section 1.1. For neutral mutants, the rate of virus transmission from a multiply infected cells is proportional to the fraction of the virus strain in the infected cell. For advantageous or disadvantageous mutants, this also applies. Fitness differences are modeled by modifying the probability of the virus strain that has been chosen for infection to successfully enter the new target cell (note that the basic formulation (1-5) assumes that viruses are 100% successful in infecting the target cell). For example, a disadvantageous mutant is assumed to have an increased probability that successful infection fails. Hence, fitness differences are expressed at the level of entry into the new target cells. Mutations are assumed to occur during the infection process, corresponding to mutations that occur during reverse transcription in HIV infection. We refer to "mutants" as virus strains with a specific characteristic, such as a drug-resistant virus strain, an immune escape strain, or another specific phenotype. We refer to the virus population that does not share this characteristic as the non-mutant or wild-type population, even though RNA virus populations tend to exist as a quasi-species, due to reduced replication fidelity [39]. The ODE model that includes mutations and competition among strains of different fitness is given by Equations (3-4) in Section 1.1 of Supplementary Information.

In computer simulations, we will concentrate on parameters that are relevant for acute HIV infection, characterized by a basic reproductive ratio $R_0 = 8$. The assumed model parameters are based on the literature and explained in the Supplementary Information Section 1.2. Since the model is applicable to viruses other than HIV, we also vary parameters more broadly to investigate dynamics for lower values of $R_0$, where we expect to see larger effects of stochasticity.

## 2.2 Hybrid algorithm

Here, we describe a stochastic-deterministic hybrid algorithm that simulates the dynamics of small mutant populations and small populations of multiply infected cells stochastically, while describing the majority populations deterministically. This allows us to run computationally efficient simulations of viral evolutionary processes at large population sizes, without losing the effects arising from the stochastic dynamics of minority subpopulations. This methodology is based on our previous work in the context of tumor cell evolution [34], which in turn is related to work in the

field of chemical kinetics [36, 4, 43]. The hybrid algorithm is based on the idea that if a cell population is sufficiently large, an ODE representation can provide a good approximation of most stochastic trajectories of the population. We can write the ODE system as a single vector equation $d\mathbf{V}/dt = \mathbf{F}(\mathbf{V})$, where $\mathbf{V}$ is a vector that contains all the cell subpopulations. Let $\mathcal{M}$ be a given population size threshold. We classify the cell population $x_i$ as small at time $t$ if $x_i(t) < \mathcal{M}$, or large otherwise. We simulate the small populations stochastically using the Gillespie algorithm and use the ODEs for the large populations. Further details of the hybrid method are given in the Supplementary Information Section 2.

### 2.2.1 Implementation

The size threshold $\mathcal{M}$ is a very important parameter in the hybrid algorithm. If $\mathcal{M} = 0$, then at each time point every non-zero population is classified as large and the hybrid algorithm is identical to the deterministic solution of the ODEs. If $\mathcal{M}$ is very large, that is larger than all populations for the duration of the time-span of interest, then the hybrid algorithm is the same as the completely stochastic Gillespie simulation of the model and can be extremely computationally inefficient. For intermediate $\mathcal{M} > 0$, the hybrid algorithm is computationally efficient and the averages over many hybrid simulations go from approximating the deterministic predictions to converging to the stochastic averages as $\mathcal{M}$ increases. Therefore, in order to efficiently approximate the completely stochastic implementation of the model, we need to choose an intermediate $\mathcal{M}$ such that the results are close to the fully stochastic implementation.

We can achieve this by comparing the hybrid averages over many simulations to completely stochastic averages over many simulations for simplified models, such as assuming a constant large number of uninfected cells or using parameter values that result in smaller and more computationally manageable population sizes. For these models, completely stochastic simulations can be carried out and allow us to determine what size threshold $\mathcal{M}$ is reasonable for the related models. Specifically, since the averages over many hybrid simulations start from the deterministic prediction ($\mathcal{M} = 0$) and converge to the completely stochastic average, similarly to [34] we i) set some difference threshold $\varepsilon > 0$, ii) test multiple size thresholds $\mathcal{M}$, and iii) choose the smallest $\mathcal{M}$ such that the hybrid average is within $\varepsilon$ of the completely stochastic average.

Table 2 contains approximate computer simulation run times for the completely deterministic ODE system, the hybrid method, and the completely stochastic Gillespie algorithm. Each system is run for the single mutation, double mutation, and triple mutation models. All simulations include only free virus transmission with limited multiple infection ($N = 3$), represent established infections only (we ignore stochastic simulations in which the infection dies out), and are stopped once the infected cell population reaches $10^8$ cells. The times for the ODE and hybrid simulations also depend on the ODE solution method and the step size, $h$ (here $h = 10^{-5}$ with Euler method). In general, with $k$ possible mutations, the number of strains per model is $2^k$ and the number of equations (subpopulations) per model is $\binom{N+2^k}{2^k}$.

Because the parameters chosen for the simulations in Table 2 correspond to $R_0 = 8$, a relatively small size threshold $\mathcal{M}$ gives a good approximation of the fully stochastic simulations. Simulations with lower $R_0$ require higher values of $\mathcal{M}$ and hence take longer to run.

### 2.2.2 Choosing a size threshold $\mathcal{M}$

We have developed an analytical method for finding a lower bound on size threshold $\mathcal{M}$, which is based on the notion of $R_0$. This method does not depend on the number of mutations, infection multiplicity, fitness landscape, etc. The basic reproductive ratio, $R_0$, is the average number of

| Model | single mutation $k = 1$ 2 strains, 10 equations | double mutation $k = 2$ 4 strains, 35 equations | triple mutation $k = 3$ 8 strains, 165 equations |
|---|---|---|---|
| Full ODEs | $< 1$ second | 4 seconds | 12 minutes |
| Hybrid, $\mathcal{M} = 10$ | $< 1$ second | 4 seconds | 13 minutes |
| Hybrid, $\mathcal{M} = 10^3$ | $< 1$ second | 4 seconds | 13 minutes |
| Hybrid, $\mathcal{M} = 10^5$ | $< 1$ second | 6 seconds | 15 minutes |
| Hybrid, $\mathcal{M} = 10^7$ | 1 minute | 7 minutes | 30 hours |
| Full Gillespie | 12 minutes | 100 minutes | 1 week |

Table 2: Approximate average run times for a single simulation for the completely deterministic ODE system (Euler method with step-size $h = 10^{-5}$), the hybrid method with different threshold values ($\mathcal{M}$), and the completely stochastic Gillespie algorithm (rows). Each system is run for the single mutation, double mutation, and triple mutation models (columns). In each system we assume all strains are neutral ($F_i = 1$ for all $i$). The other parameters are $N = 3,\ \mu = 3 \times 10^{-5},\ \lambda = 1.59 \times 10^7,\ \beta = 3.60 \times 10^{-9},\ \gamma = 0,\ d = 0.016$, and $a = 0.45$.

newly infected cells generated per single infected cell at the beginning of the infection. Therefore, infections with larger $R_0$ will lead to quicker and more successful growth of the overall virus population. While in a deterministic system, infections with $R_0 > 1$ will never go extinct, in the stochastic setting, even if $R_0 > 1$, a single infected cell can die out before successfully infecting other cells. The rate at which infections stochastically go extinct is given by $\frac{1}{R_0}$ [44, 2]; in other words, infection will successfully spread with probability $\Phi^\infty = 1 - 1/R_0$. Moreover, one can show that an infection will increase until size $K$ (before possibly going extinct) with probability

$$\Phi^K = \frac{1 - \frac{1}{R_0}}{1 - \left(\frac{1}{R_0}\right)^K}. \tag{11}$$

Setting the size-threshold to a given value $\mathcal{M}$ essentially means that we assume that a population that has reached that size will no longer go extinct, because its subsequent dynamics are described by ODEs. Let $\delta > 0$ be some small difference threshold. We define the lower bound size threshold, $\hat{\mathcal{M}}$, as the smallest natural number $\mathcal{M}$ such that

$$|\Phi^{\mathcal{M}} - \Phi^\infty| < \delta,$$

which gives the estimate

$$\hat{\mathcal{M}} = \lceil \ln\left(1 + \frac{R_0 - 1}{\delta R_0}\right) / \ln R_0 \rceil, \tag{12}$$

where $\lceil . \rceil$ denotes the ceiling function. Note that $\hat{\mathcal{M}}$ is a lower bound, and depending on the details of the model, it is possible that a larger $\mathcal{M}$ is needed to get accurate descriptions of mutant dynamics. In general, we can always confirm that a chosen $\mathcal{M}$ is large enough using the $\varepsilon$ test described in the preceding section and in [34].

# 3 Results

## 3.1 Comparing and contrasting ODE versus stochastic / hybrid simulations

Here, we show that the hybrid simulations allow us to obtain a number of evolutionary measures that cannot be obtained accurately from ODEs. While ODE simulations can serve as a good approximation of the average population sizes (both wild-type and mutant) over time in a stochastic process, an important measure that cannot be predicted by ODEs is the number of mutants at a

given population size of infected cells. If we determine the number of mutants in ODE simulations once the infected cell population size in the ODE has reached a threshold $N$ (say, at time $t_N$), we are effectively determining the average number of mutants over different stochastic trajectories, which all correspond to different infected cell population sizes. This is because at time $t_N$, while the average trajectory reaches size $N$, some stochastic trajectories at that time will be lower and some higher than $N$. Additional evolutionary measures that cannot be predicted by ODEs are the time of first mutant generation, the distribution of mutant numbers at a given infected cell population size, and the probability that a mutant is present at a given population size.

To underline these points, in this section we compare ODE predictions to outputs from the stochastic simulations, in the context of the evolution of neutral, advantageous, and disadvantageous mutants. Here we focus on relatively simple scenarios, considering the exponential growth phase of the virus population and only including free virus transmission; synaptic transmission and infection peak dynamics are studied in the next section. While parameter sets explored here are relevant to HIV, we also include broader parameter sets for comparison, especially those where the basic reproductive ratio is lower. In these regimes, the dynamics are governed by stochasticity to a larger extent.

#### 3.1.1 The average number of mutants at a given infected cell population size

We start by determining the average number of neutral mutants once the number of infected cells has reached a threshold size in the purely stochastic process (we discard simulations in which the infection goes extinct stochastically before reaching the threshold size). We then compare this to the number of mutants predicted by the ODE at the time when the average infected cell population size is the same threshold. To be able to run fully stochastic simulations, we determine the number of mutants at a relatively low infected cell population size of $10^4$.

Figure 1(a) shows the results for a neutral mutant, assuming different values for the basic reproductive ratio of the virus, $R_0$. The lower the value of $R_0$, the higher the discrepancy between the average of the stochastic simulations and the ODE results. For $R_0 = 8$, which is characteristic of HIV infection [33, 31], the discrepancy is minimal. The reason is that for relatively large values of $R_0$, the variation of the infected cell population size at a given time is reduced. Figures 1(b) and 1(c) show equivalent plots for advantageous and disadvantageous mutants, respectively. Again, the extent of the discrepancies increases with lower values of $R_0$. Discrepancies tend to be larger than for neutral mutants, and are apparent even for higher values of $R_0$ (e.g. $R_0 = 8$).

While ODEs cannot accurately describe the average behavior of the stochastic model, the hybrid method (with a sufficient size threshold) is able to do so, as is demonstrated in Figure S4(a).

#### 3.1.2 The timing of mutant emergence

Another important measure is the time at which the first copy of a given mutant is generated, and the infected cell population size at which this mutant is generated. The closest measure in the ODE is the the time and infected cell population size at which the average number of mutants crosses unity. As shown in Figure S3, however, significant discrepancies exist between this ODE measure and the accurate prediction of stochastic simulations, and this discrepancy increases with a larger number of mutation events required to generate this mutant (i.e. 1-hit, 2-hit. 3-hit mutants etc). The hybrid method, however, provides an accurate approximation (Figure S4(b)).

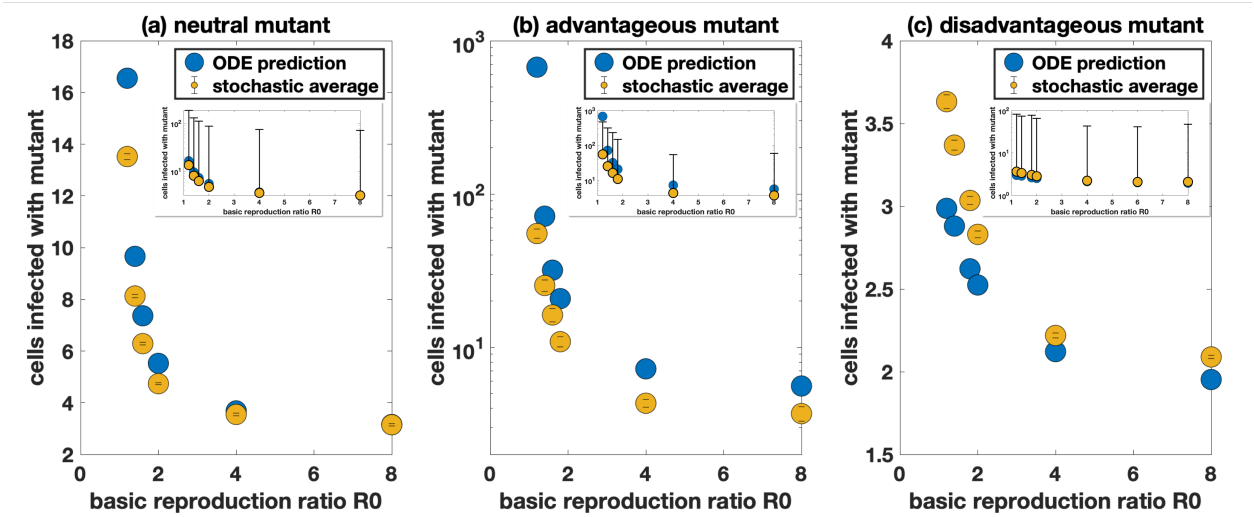


Figure 1: Comparison of the deterministic prediction and stochastic average of the number of cells infected with the mutant with free virus transmission only. The deterministic predictions are in blue and the stochastic hybrid simulations with $\mathcal{M} = 10^4$ (infected populations always treated completely stochastically) are in yellow. Standard error bars are included and the inserts show standard deviation bars. (a) Neutral mutant, $F_{\text{mutant}} = 0.9$. Each yellow dot represents the average taken over at least $2 \times 10^6$ simulations. (b) Advantageous mutant with 10% advantage, $F_{\text{mutant}} = 0.99$. Each yellow dot represents the average taken over at least $1.1 \times 10^3$ simulations. (c) Disadvantageous mutant with 10% disadvantage, $F_{\text{mutant}} = 0.81$. Each yellow dot represents the average taken over at least $3.5 \times 10^6$ simulations. We have $R_0 = \frac{\lambda\left(\beta F + \gamma(1-(1-F)^S)\right)}{ad}$, and the parameters are $F_{\text{wild-type}} = 0.9$, $N = 3$, $\mu = 3 \times 10^{-5}$, $\lambda = 1.59 \times 10^7$, $\beta = 4 \times 10^{-9}$, $\gamma = 0$, and $d = 0.016$. The infected cell death rate $a$ is adjusted to achieve the required $R_0$.

### 3.1.3 Probability distributions of mutant numbers

The probability distribution of the number of mutants at a given infected cell population size, or at a given time, is a measure that has no equivalent in ODEs, yet these measures have strong biological relevance. For example, it is important to understand the likelihood that certain mutants exist at various stages during virus growth, such as virus strains resistant against one or more drugs or against one or more immune cell clones. The hybrid method provides a good approximation of the results from stochastic simulations, as shown in Figure S1. This also applies to simulations that assume relatively low values of $R_0$ (Figure S2), although larger size thresholds $\mathcal{M}$ are required for smaller values of $R_0$.

## 3.2 Impact of multiple infection on mutant evolution

In this section, we apply the above-described hybrid method to explore how multiple infection can affect virus evolution during an exponential growth phase and near the peak infection, with particular relevance to the acute phase of HIV infection, during which the infected cell population grows to large sizes. Multiple infection can influence viral evolution in a variety of ways. On a basic level, the ability of viruses to enter cells that are already infected increases the target cell population and allows the virus to undergo more reverse transcription events, thus increasing the effective rate at which mutations are generated. In addition, viral fitness can be altered in multiply infected cells through viral complementation or inhibition [12], which again has the potential to influence the evolutionary dynamics. In the context of HIV infection, direct cell-to-cell transmission through virological synapses (synaptic transmission) increases the complexity of these processes. Synaptic transmission typically results in the transfer of multiple viruses from the source cell to the target cell, thus increasing the level of multiple infection [14, 5, 1, 37]. In addition, synaptic transmission can lead to the repeated co-transmission of different virus strains [6, 24] which can amplify the effect of viral complementation or inhibition. To explore these dynamics, the hybrid method is important because multiple infection becomes increasingly prevalent at large population sizes, where both mutant viruses and multiply infected cells exist as relatively small populations compared to the larger populations of wild-type viruses and singly infected cells. We will focus on basic evolutionary processes that do not involve recombination.

### 3.2.1 The effect of multiple infection on the spread of neutral mutants

We start with the most basic scenario: the effect of multiple infection on the presence of neutral mutants during the growth phase of the virus. For simplicity, we concentrate on free virus transmission only. Because this analysis is done with HIV in mind, we set $R_0 = 8$. Figures S7 and 2 show the distributions of cells infected with neutral single and double mutants in the presence and absence of multiple infection. Figure S7 shows that at relatively low virus loads, the average number of mutants is the same, whether multiple infection is assumed to occur or not. At larger population sizes that are close to peak virus load, however, we observe a pronounced difference, Figure 2. In these simulations, we recorded the number of mutants at $6 \times 10^8$ infected cells, as it is close to the peak and almost all stochastic simulations reached this threshold. We can see that multiple infection results in a 2-fold or larger increase in the average number of mutants, both for single-hit (Figure 2(a)) and double-hit mutants (Figure 2(b)). The reason is that larger number of infection events occur in the presence of multiple infection, thus raising the number of mutants that are generated. We further note that multiple infection not only increases the average number of mutants at high viral loads, but that it also leads to a larger variation in mutant numbers, shown by a larger standard deviation of mutant numbers in the presence of multiple infection (Figure 2).

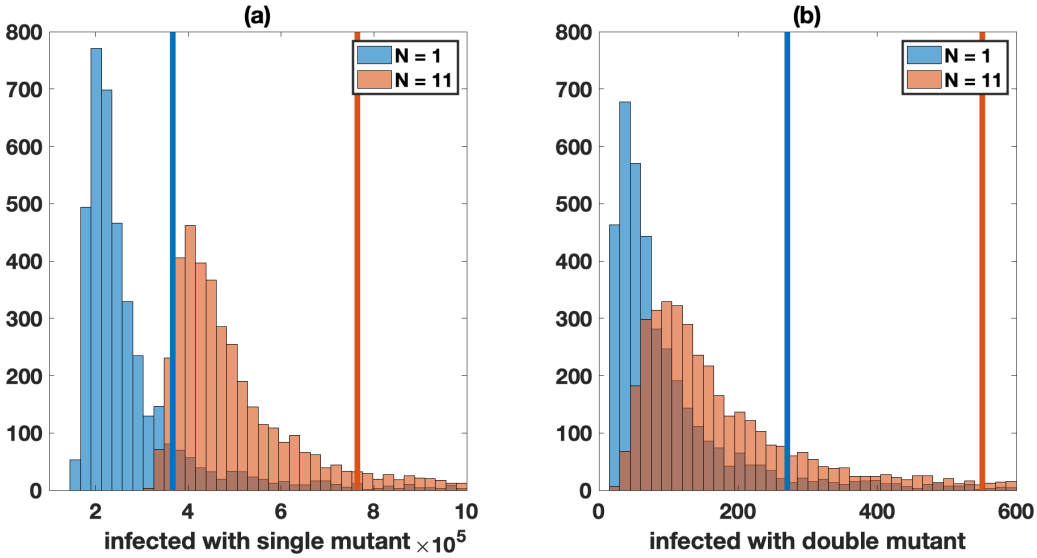


Figure 2: Neutral mutant evolution in the absence of synaptic transmission, comparing simulations with single infection only ($N = 1$, blue) and in the presence of multiple infection ($N = 11$, red). The mean values are shown by vertical lines. For both panels, the Kolmogorov-Smirnov test between the two distributions gives a $p$-value less than $10^{-6}$. (a) Number of cells infected with one of the single mutant strains. The average for single infection is approximately $3.7 \times 10^5$ and for multiple infection is approximately $7.6 \times 10^5$. (b) Number of cells infected with the double mutant strain. The average for single infection is approximately 271 and for multiple infection is approximately 551. Distributions represent $4 \times 10^3$ hybrid simulations with size threshold $\mathcal{M} = 50$. Simulations in which infections are not established (or in the rare case a simulation does not reach the infected size threshold) are discarded. Simulations are stopped when the infected cell population is close to peak infection ($6 \times 10^8$ cells). The other parameters are similar to Figure 1 ($F_{\text{wild-type}} = 1,\ F_{\text{mutant}} = 1,\ \mu = 3 \times 10^{-5},\ \lambda = 1.59 \times 10^7,\ \beta = 3.60 \times 10^{-9},\ \gamma = 0,\ a = 0.45,\ d = 0.016$, and $R_0 = 8$.)

These trends are not particular to neutral mutants because we focus on exponential, or nearly-exponential, virus growth. Similar trends are observed for advantageous or disadvantageous mutants (see Supplementary Information Section 4 and Figure S8).

While computationally more costly, we also examined the prevalence of neutral triple-hit mutants, because such mutants can be important for simultaneously escaping three immune response specificities or three drugs. We found that even near peak virus load, the probability that a triple mutant exists is relatively low (Figure S9). In other words, such mutants are unlikely to exist even at the peak of primary HIV infection. Nevertheless, multiple infection results in an almost 2-fold increase in the probability that neutral triple mutants exist around peak infection. Such an increase in mutant generation could be important for virus persistence in the face of mounting immune responses during the acute phase of the infection.

#### 3.2.2 Evolutionary dynamics in more complex settings: complementation, interference, and the role of synaptic transmission

Multiple infection becomes especially important for viral evolutionary dynamics if different virus strains interact with each other inside the same cell. One type of such interactions is complementation, where a disadvantageous mutant gains in fitness in a coinfected cell [12]. Another example is interference, where an advantageous mutant can lose the fitness advantage when together with a wild-type virus in the same cell [41]. We will use our hybrid methodology to investigate the evolution of disadvantageous and advantageous mutants, and the effect of complementation and interference, respectively. We start by examining the dynamics assuming free virus transmission, and then compare results to simulations that assume virus spread through synaptic transmission. Synaptic transmission can be especially relevant here because it can promote the repeated co-transmission of genetically distinct virus strains. For example, if a disadvantageous mutant is repeatedly co-transmitted with a wild-type virus, and if the disadvantageous mutants benefits from complementation, then synaptic transmission can significantly enhance the spread potential of the mutant.

As before, the fitness difference is modeled at the level of the infection process. For example, for a disadvantageous mutant, there is a chance that infection of a new cell is unsuccessful. In this case, complementation means that the wild-type virus can provide a product that enhances the infectivity of the mutant. Similarly, for interference, it is assumed that the chance of infection by an advantageous mutant is reduced if the offspring mutant was generated in a coinfected cell.

**Effect of viral co-transmission on mutant spread (in the absence of mutations).** To assess to what extent the co-transmission of different virus strains influences viral evolution, we consider computer simulations in the absence of mutant production. Instead, we start with one infected cell that contains both one wild-type and one mutant virus, and simulate the spread of the virus population until a threshold number of infected cells is reached. The purpose of excluding mutant production is to fully quantify to what extent synaptic transmission enhances the spread potential of a mutant.

Complementation: First, consider viral complementation. We study an extreme case where a mutant has zero fitness by itself, but has an infectivity identical to the wild-type virus if the mutant offspring virus is produced in a cell coinfected with a wild-type virus. In this parameter regime, the mutant virus cannot spread at all in the absence of complementation, whether spread occurs by free virus or synaptic transmission. The occurrence of complementation, however, allows virus spread due to the elevated viral fitness in coinfected cells. For free virus transmission, this effect is modest (Figure 3(a)). A limited amount of mutant spread can occur, but the average number of mutants at peak infection levels is still less than one, indicating that mutants largely fail to spread

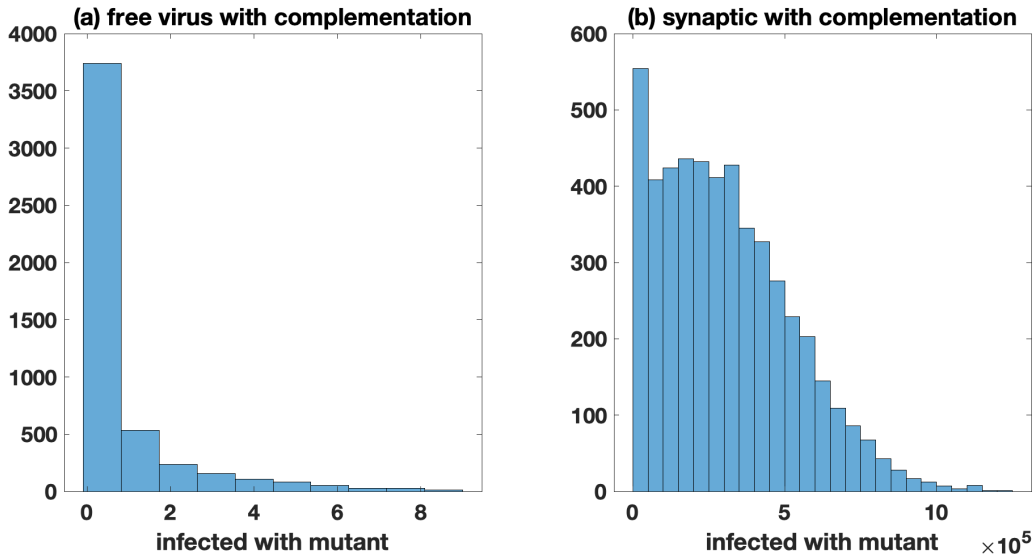


Figure 3: Zero fitness mutants, comparing the effect of complementation for free virus and synaptic transmission. All simulations start with a single infected cell coinfected with a single copy of both the wild-type and mutant, and mutation is turned off ($\mu = 0$). (a) Only free virus transmission ($\beta = 3.60\times10^{-9},\ \gamma = 0,\ N = 11$) with complementation. The average number (standard deviation) of cells infected with the mutant is 0.71 (1.73). (b) Only synaptic transmission ($\beta = 0,\ \gamma = 3.60 \times 10^{-9},\ N = 25$) with complementation. The average number (standard deviation) of cells infected with the mutant is $3.1 \times 10^5$ ($2.2 \times 10^5$). Histograms represent $5 \times 10^3$ hybrid simulations with size threshold $\mathcal{M} = 50$. Simulations in which infections are not established (or in the rare case a simulation does not reach the infected size threshold) are discarded; simulations are stopped when the infected cell population is close to peak infection ($5 \times 10^8$ cells). The fitness of the wild-type is fixed at $F_{\text{wild-type}} = 0.9$ and $F_{\text{mutant}} = 0$. The other parameters are as in Figure 1 ($\lambda = 1.59 \times 10^7,\ a = 0.45$, and $d = 0.016$).

in this setting. In simulations with synaptic transmission, however, we observe extensive mutant spread in the presence of complementation (Figure 3(b)). Around peak infection, the number of cells infected with the mutant is of the order of $10^5$. This shows that synaptic transmission can play a crucial role at promoting the spread of disadvantageous mutants through complementation.

Interference: Next, consider viral interference. Assume an advantageous mutant, which has a significant fitness advantage by itself (10%), but has an infectivity identical to the wild-type virus if the mutant offspring is produced in a cell coinfected with the wild-type. Under free virus transmission (Figure S11(a-b))), coinfection does not play a significant role, and therefore interference only decreases the expected number of mutants by a small percentage. Interestingly, for synaptic transmission (Figure S11(c-d)), interference only plays a marginally larger role compared to the dynamics under free-virus transmission. The reason for this relatively mild effect of interference under purely synaptic transmission is rooted in an inherent reduction of fitness differences due to repeated infection events in synaptic transmission. We elaborate on this later on in the context of dynamics with mutations.

**Evolutionary dynamics in the presence of mutant production.** Here, we repeat this analysis assuming that mutant production occurs. The mutant dynamics are now influenced by two factors: (i) as before, mutant viral replication and mutant fitness influence spread; (ii) mutation

processes generate mutant viruses from wild-type, which also contributes to the increase of mutant numbers. We consider both viral complementation and inhibition.

Complementation: We first focus on a mutant that has zero fitness if it is by itself in a cell. If mutant numbers are measured at relatively low virus loads (Figure 4(a,b)), complementation makes no difference for simulations that assume free virus transmission only (panel (a)). For simulations assuming synaptic transmission only, however, a larger difference between mutant numbers with and without complementation is observed (approximately 2-fold, panel (b)), resulting from the frequent co-transmission of different virus strains, which occurs even at lower virus loads. Even more striking is the difference in the distribution of mutant numbers with and without complementation, under synaptic transmission (panel (b)). The long distribution tail in the presence of complementation is a result of early mutation events, which are extremely rare, but give rise to unusually high numbers of mutants at the threshold size. These events are similar to the so-called "jack-pot" event that have recently attracted attention in the context of mutant evolution in expanding cell populations [11, 45].

If the number of mutants is measured at higher virus loads, near peak, we find that complementation makes a modest difference if only free virus transmission is assumed (Figure 4(c)). This occurs because mutants that are generated at high virus loads will have a substantial chance to enter a cell that also contains a wild-type virus, leading to enhanced mutant spread at high virus loads. If we assume that the virus spreads only through synaptic transmission (panel (d)), complementation makes a larger difference, but the effect of complementation is only slightly larger than that at low virus loads (panel (b)). The reason is that the probability for wild-type and mutant viruses to be co-transmitted does not depend strongly on virus load.

We note that in the models with mutant generation, the effect of complementation on mutant numbers is much less pronounced than in simulations without mutation processes, even if the virus is assumed to only spread through virological synapses. The reason is that in the absence of mutational processes, the initially present mutant virus cannot spread without complementation, whereas it can do so in the presence of complementation. In the presence of mutational processes, however, even zero-fitness mutant numbers can rise over time without complementation, due to mutant production by wild-type viruses. Because the population size at peak virus load is large relative to the inverse of the mutation rate, mutant generation is a significant force that drives mutant numbers over time, limiting the difference that mutant replication in coinfected cells can make on the mutant population size.

Next, we assume that the mutant is no longer a zero-fitness type, but can be transmitted independently of the wild-type virus, although with a 10% fitness cost. In other words, if an infection event is attempted, it succeeds with a probability that is 10% smaller than that for the wild-type virus: $F_{\text{mutant}} = 0.9F_{\text{wild-type}}$. If the mutant virus is in the same cell as the wild-type, however, this fitness cost is assumed to disappear and the mutant is neutral with respect to the wild-type virus. We focus on mutant numbers at high virus loads. We find that the number of mutants is only increased by a small amount, both if we assume that the virus spreads only by free virus transmission (panel (e)) or only by synaptic transmission (panel (f)); the difference is slightly larger for simulations that assume synaptic virus transmission, approximately 1.4 fold in Figure 4(f)).

The relatively small increase in mutant numbers brought about by complementation is surprising in the context of synaptic transmission. Intuitively, even though the disadvantageous mutant virus in Figure 4(f) can spread alone, the assumed 10% fitness cost, which is overcome by complementation, is still substantial. The reason for the limited impact of complementation is that in the presence of synaptic transmission, the actual fitness disadvantage of the mutant is reduced.

The fitness cost is implemented by assuming that upon transfer to the new target cell, each virus has an increased probability to fail successful completion of infection. With synaptic transmission, it is assumed that there are $S$ infection attempts (in our simulation $S = 3$). This increases the likelihood that the cell will become infected (i.e. that at least one of the attempts is successful). Through this process, the effective fitness disadvantage of the mutant ends up being less than the 10% cost assumed per virus, which explains the modest effect of complementation on mutant numbers. The notion that the simultaneous transfer of multiple viruses per synapse reduces the effective relative fitness cost of a mutant has important implications that go beyond the scope of the current paper, and is explored in detail in a separate study. This analysis indicates that viral complementation might only make a substantial impact on the number of disadvantageous mutants if the disadvantage is very large. Therefore, biologically, complementation might be most relevant to defective virus particles, and this effect is more pronounced under synaptic compared to free virus transmission.

Interference: Here we consider an advantageous mutant that loses fitness advantages in cells that contain both the mutant and the wild-type virus. This is implemented similarly to the simulations with disadvantageous mutants. To model the advantage, we assume that a mutant virus, upon transfer, succeeds in infecting the target cell with the probability that is 10% larger than that of the wild-type virus: $F_{\text{mutant}} = 1.1F_{\text{wild-type}}$. As with complementation, Figure 4(g,h) shows that interference has a modest impact on the number of advantageous mutants at the size threshold (close to peak infection levels). Interference lowers the number of advantageous mutants to a slightly stronger degree if we assume synaptic (panel (h)) rather than free virus transmission (panel (g)), although the difference is relatively small in both cases, which is reminiscent of a similarly small effect of interference under synaptic transmission, observed in the absence of mutations, Figure S11(c-d)). The small effect for free virus transmission is explained by the absence of significant co-transmission of mutant and wild-type viruses, which limits the occurrence of the intracellular interactions among the two viral strains. For the simulations with synaptic transmission, the small effect is again explained by a reduction in the effective fitness difference between mutant and wild-type strains as a result of multiple, simultaneous infection events during synaptic transmission. Therefore, these results suggest that interference is unlikely to have a major impact on the dynamics of advantageous mutants, unless the advantage is very large, which would be biologically unrealistic (the simulations shown in Figure 4(g,h) already assume a 10% fitness advantage of the mutant).

# 4 Discussion

In this paper, we described a hybrid stochastic-deterministic algorithm to simulate viral evolutionary dynamics at large population sizes, including the occurrence of multiple infection of cells. The coevolution of relatively small populations (mutants and multiply infected cells) with larger populations (wild-type and singly infected cells) renders stochastic computer simulations computationally costly and not feasible when the virus population rises to higher levels. Ordinary differential equations can only predict the average number of mutants over time, but fail to accurately describe the number of mutants at a given infected cell population size, the mutant number distributions, or the timing of mutant generation. The hybrid method described here, however, provides an accurate approximation of the true stochastic dynamics, at a fraction of the computational cost. This method therefore can serve as a practical tool to simulate complex viral evolutionary processes at large population sizes.

At the same time, however, the hybrid method can also run into computational limitations, depending the assumptions underlying the exact model formulation. While the hybrid method is capable of handling a large number of subpopulations, the number of "reactions" included in the

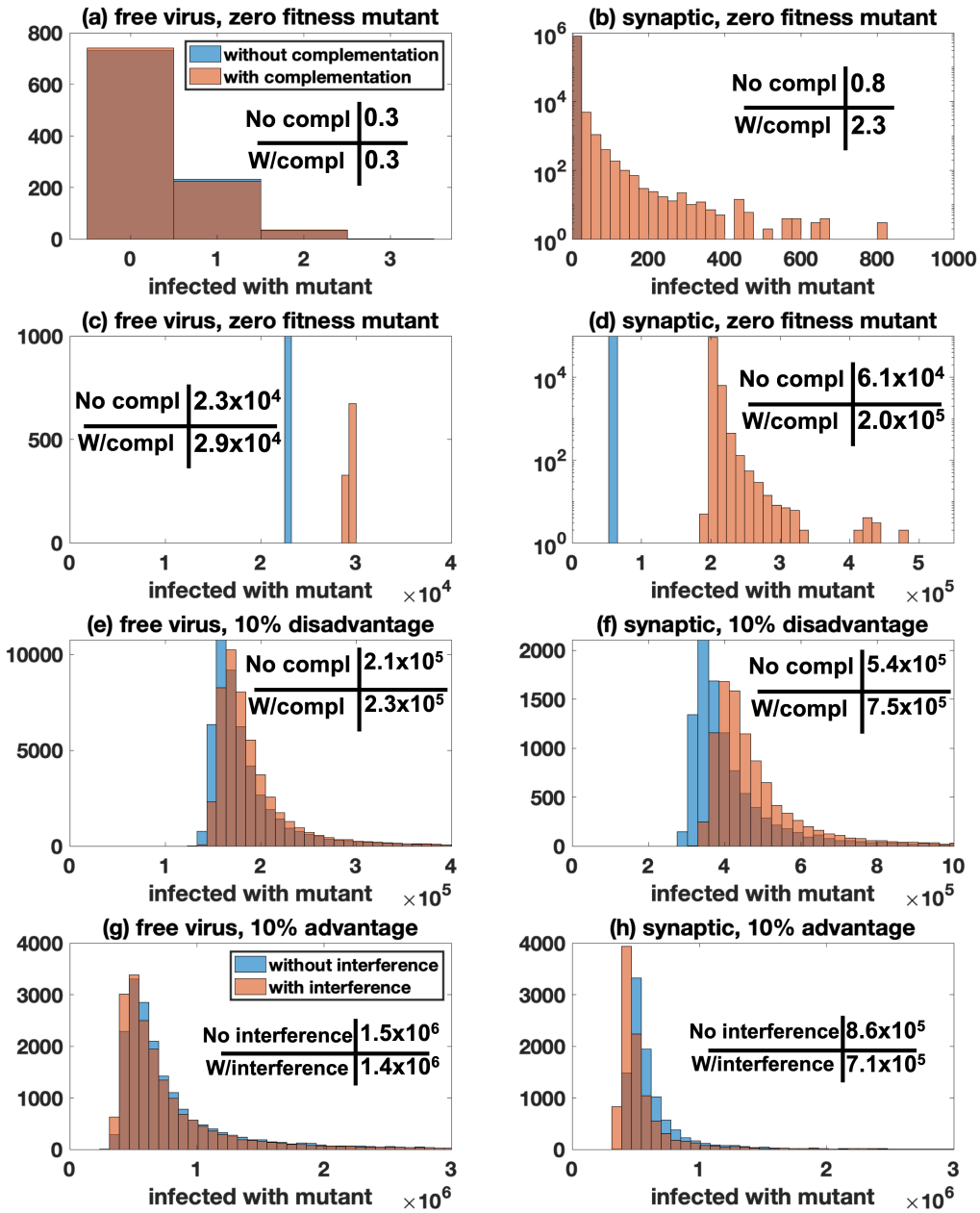


Figure 4: Mutant evolution under different scenarios with 100% free virus transmission (left panels: $\beta = 3.6 \times 10^{-9}$, $\gamma = 0$, $N = 11$) or 100% synaptic transmission (right panels: $\beta = 0$, $\gamma = 3.6 \times 10^{-9}$, $S = 3$, $N = 25$). Panels (a) and (b) record the number of cells infected with the mutant at $10^4$ infected cells, for all other panels it is $5 \times 10^8$ infected cells. For all panels, the blue bars represents simulations without complementation/interference and the red bars represents simulations with complementation/interference. The mean values are presented in each panel. For panels (b)-(h), $p < 10^{-6}$ by the Kolmogorov-Smirnov test. (a-d) Zero fitness mutant ($F_{\text{mutant}} = 0$). (e,f) Disadvantageous mutant ($F_{\text{mutant}} = 0.81$). (g,h) Advantageous mutant ($F_{\text{mutant}} = 0.99$). For all simulations, we fix $\mathcal{M} = 50$, $F_{\text{wild-type}} = 0.9$ and the other parameters are as in Figure 1 ($\mu = 3 \times 10^{-5}$, $\lambda = 1.59 \times 10^7$, $a = 0.45$, and $d = 0.016$).

stochastic part of the algorithm increases with (i) the number of different virus strains, (ii) the maximum multiplicity $N$, and (iii) the number of virus transferred per synapse $S$. If these parameters are too large, the number of reactions for the Gillespie algorithm can become too high to be computationally feasible (even if only small populations are handled stochastically). In general, the number of strains per model is $2^k$ and the number of differential equations (subpopulations) per model is $\binom{N+2^k}{2^k}$. If we model only free virus transmission, the number of infection events is the number of strains multiplied by the number of subpopulations eligible to be infected, but when synaptic transmission is included, there are many more infection events, which is correlated with the number of ways to partition $S$ into $2^k$ non-negative integers that sum to $1, 2, \ldots, S$. When the number of reactions is on the order of $10^4$, each simulation becomes very computationally expensive, which happens, for example, if we consider triple mutants in the presence of synaptic transmission.

We used the hybrid stochastic-deterministic method to study how multiple infection and intracellular interactions among virus strains influence the evolutionary dynamics of mutants in the acute phase of HIV infection, during which the number of infected cells can rise to high levels, of the order of $10^8$ infected cells across the lymphoid tissues [10]. We showed that these processes can shape mutant evolution, but also found that this effect is restricted to select circumstances. On a basic level, the models confirmed the intuitive idea that multiple infection accelerates mutant evolution due to the larger number of mutation events during reverse transcription, when already infected cells become super-infected.

The model predictions about the ability of viral complementation to enhance the spread of disadvantageous mutants was more complex. According to the model, synaptic transmission is required to enhance disadvantageous mutant spread through complementation because it allows the repeated co-transmission of different virus strains; at the same time, however, this effect of complementation is only sizable if the selective disadvantage of the mutant is substantial, which most likely corresponds to a defective virus. The reason is that in the model studied here, synaptic transmission reduces the effective fitness difference between mutant and wild-type virus. This is because during a synaptic transmission event, multiple viruses are assumed to attempt infection of the target cells, thus increasing the chance that the cell will become infected with at least one of them. Even though we assumed a 10% lower probability of successful infection per mutant virus, in the context of our assumption that three viruses attempt infection per synapse, the overall chance that the cell becomes infected with a mutant is only 0.01% lower than the chance that it will become infected with a wild-type virus (the effective fitness difference). With a reduced effective fitness difference, complementation can only accelerate mutant growth by a modest amount.

Even if the effect of complementation is only pronounced for defective viruses, this still has strong biological significance. The maintenance of virus variants with zero or very low fitness during viral spread could be important for the evolvability of HIV in patients. The low fitness virus variants can potentially carry other mutations in their genomes, such as drug resistance or immune escape mutations. If these low fitness variants are repeatedly present in the same cell as wild-type viruses, recombination can transfer the mutation in question onto the wild-type genome, thus accelerating the rate of virus evolution. If the low fitness variants are not maintained, due to lack of complementation, however, this effect would not occur and could lead to a slower rate of virus evolution. Hence, maintenance of defective virus variants through complementation, and the consequent enhanced evolvability of the virus, could be one mechanism underlying the evolution of synaptic transmission in HIV infection. Recombination can be built into the models presented here to explore these dynamics in the future.

Another intracellular interaction that we considered was viral interference, where we track an advantageous mutant that loses fitness when together with a wild-type virus in an infected cell.

As with complementation, for the fitness loss to be a driving event, the repeated co-transmission of wild-type and mutant virus is required through virological synapses. For the same reason as explained above, however, the multiple virus transfer events that occur during synaptic transmission reduce the fitness difference between the two virus strains, thus reducing the impact of interference on mutant numbers. To see a more significant effect would require a very substantial fitness advantage of the mutant, which is biologically unrealistic. According to our results, we therefore expect that viral interference is unlikely to significantly reduce the number of advantageous mutants.

According to the model studied here, viral complementation is not expected to play a significant role for mutant evolution in the absence of a transmission mechanism that involves the simultaneous transfer of multiple viruses from the infected cell to the target cell. It is important to remember, however, that the model presented here assumes well mixed virus and cell populations. If, in contrast, viruses spread in spatially structured cell populations with limited mixing, the spatial restriction could force the repeated co-transmission of different virus strains from one cell to another, even in the context of free virus transmission (simply because only a limited number of target cells are located in the immediate neighborhood of an infected cell). Therefore, spatial restriction during free virus transmission could have a similar effect as synaptic transmission during HIV infection. Indeed, computational modeling work has shown that similar to synaptic transmission, spatially restricted virus growth can lead to higher infection multiplicities, even at lower virus loads [40]. The correspondence between the properties of synaptic transmission in HIV infection and spatially restricted free virus spread remains to be established in more detail, and has relevance for a range of viral infections, importantly bacteriophage infections.

## References


[1] Luis M Agosto, Pradeep D Uchil, and Walther Mothes. Hiv cell-to-cell transmission: effects on pathogenesis and antiretroviral therapy. *Trends in microbiology*, 23(5):289–295, 2015.

[2] L.J.S. Allen and P. van den Driessche. Relations between deterministic and stochastic thresholds for disease extinction in continuous- and discrete-time infectious disease models. *Mathematical Biosciences*, 243(1):99 – 108, 2013.

[3] Ani Asatryan, Dominik Wodarz, and Natalia L Komarova. New virus dynamics in the presence of multiple infection. *Journal of theoretical biology*, 377:98–109, 2015.

[4] Yang Cao, Daniel T Gillespie, and Linda R Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of chemical physics*, 124(4):044109, 2006.

[5] Ping Chen, Wolfgang Hübner, Matthew A Spinelli, and Benjamin K Chen. Predominant mode of human immunodeficiency virus transfer between T cells is mediated by sustained Env-dependent neutralization-resistant virological synapses. *Journal of virology*, 81(22):12582–12595, 2007.

[6] Armando Del Portillo, Joseph Tripodi, Vesna Najfeld, Dominik Wodarz, David N Levy, and Benjamin K Chen. Multiploid inheritance of hiv-1 during cell-to-cell infection. *Journal of virology*, pages JVI–00231, 2011.

[7] DS Dimitrov, RL Willey, H Sato, L-Ji Chang, R Blumenthal, and MA Martin. Quantitation of human immunodeficiency virus type 1 infection kinetics. *Journal of virology*, 67(4):2182–2190, 1993.

[8] Narendra M Dixit and Alan S Perelson. Multiplicity of human immunodeficiency virus infections in lymphoid tissue. *Journal of virology*, 78(16):8942–8945, 2004.

[9] Narendra M Dixit and Alan S Perelson. HIV dynamics with multiple infections of target cells. *Proceedings of the National Academy of Sciences*, 102(23):8198–8203, 2005.

[10] Jacob D Estes, Cissy Kityo, Francis Ssali, Louise Swainson, Krystelle Nganou Makamdop, Gregory Q Del Prete, Steven G Deeks, Paul A Luciw, Jeffrey G Chipman, Gregory J Beilman, et al. Defining total-body aids-virus burden with implications for curative strategies. *Nature medicine*, 23(11):1271, 2017.

[11] Diana Fusco, Matti Gralka, Jona Kayser, Alex Anderson, and Oskar Hallatschek. Excess of mutational jackpot events in expanding populations revealed by spatial luria–delbrück experiments. *Nature communications*, 7(1):1–9, 2016.

[12] Huub C Gelderblom, Dimitrios N Vatakis, Sean A Burke, Steven D Lawrie, Gregory C Bristol, and David N Levy. Viral complementation allows hiv-1 replication without integration. *Retrovirology*, 5(1):1–18, 2008.

[13] Vanessa M Hirsch. Evolution of the fittest ends in tragedy. *Nature medicine*, 5(5):488, 1999.

[14] Wolfgang Hübner, Gregory P McNerney, Ping Chen, Benjamin M Dale, Ronald E Gordon, Frank YS Chuang, Xiao-Dong Li, David M Asmuth, Thomas Huser, and Benjamin K Chen. Quantitative 3d video microscopy of hiv transfer across t cell virological synapses. *Science*, 323(5922):1743–1747, 2009.

[15] Lina Josefsson, Martin S King, Barbro Makitalo, Johan Brännström, Wei Shao, Frank Maldarelli, Mary F Kearney, Wei-Shau Hu, Jianbo Chen, Hans Gaines, et al. Majority of CD4+ t cells from peripheral blood of HIV-1–infected individuals contain only one HIV DNA molecule. *Proceedings of the National Academy of Sciences*, 108(27):11199–11204, 2011.

[16] Lina Josefsson, Sarah Palmer, Nuno R Faria, Philippe Lemey, Joseph Casazza, David Ambrozak, Mary Kearney, Wei Shao, Shyamasundaran Kottilil, Michael Sneller, et al. Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. *PLoS pathogens*, 9(6):e1003432, 2013.

[17] Andreas Jung, Reinhard Maier, Jean-Pierre Vartanian, Gennady Bocharov, Volker Jung, Ulrike Fischer, Eckart Meese, Simon Wain-Hobson, and Andreas Meyerhans. Multiply infected spleen cells in hiv patients. *Nature*, 418(6894):144–144, 2002.

[18] Andreas Jung, Reinhard Maier, Jean-Pierre Vartanian, Gennady Bocharov, Volker Jung, Ulrike Fischer, Eckart Meese, Simon Wain-Hobson, and Andreas Meyerhans. Recombination: Multiply infected spleen cells in hiv patients. *Nature*, 418(6894):144, 2002.

[19] Jason T Kimata, LaRene Kuller, David B Anderson, Peter Dailey, and Julie Overbaugh. Emerging cytopathic and antigenic simian immunodeficiency virus variants influence aids progression. *Nature medicine*, 5(5):535, 1999.

[20] Natalia L Komarova, David N Levy, and Dominik Wodarz. Effect of synaptic transmission on viral fitness in hiv infection. *PloS one*, 7(11):e48361, 2012.

[21] Natalia L Komarova, David N Levy, and Dominik Wodarz. Synaptic transmission and the susceptibility of hiv infection to anti-viral drugs. *Scientific reports*, 3:2103, 2013.

[22] Natalia L Komarova and Dominik Wodarz. Virus dynamics in the presence of synaptic transmission. *Mathematical biosciences*, 242(2):161–171, 2013.

[23] Jesse Kreger, Natalia L Komarova, and Dominik Wodarz. Effect of synaptic cell-to-cell tranmission and recombination on the evolution of double mutants in HIV. *Journal of the Royal Society Interface*, 17(164):104–115, 3 2020.

[24] Kenneth M Law, Natalia L Komarova, Alice W Yewdall, Rebecca K Lee, Olga L Herrera, Dominik Wodarz, and Benjamin K Chen. In vivo hiv-1 cell-to-cell transmission promotes multicopy micro-compartmentalized infection. *Cell reports*, 15(12):2771–2783, 2016.

[25] David N Levy, Grace M Aldrovandi, Olaf Kutsch, and George M Shaw. Dynamics of hiv-1 recombination in its natural target cells. *Proceedings of the National Academy of Sciences*, 101(12):4204–4209, 2004.

[26] Jeffrey D Lifson, Martin A Nowak, Simoy Goldstein, Jeffrey L Rossio, Audrey Kinter, Gabriela Vasquez, Theresa A Wiltrout, Charles Brown, Douglas Schneider, Linda Wahl, et al. The extent of early viral replication is a critical determinant of the natural history of simian immunodeficiency virus infection. *Journal of Virology*, 71(12):9508–9514, 1997.

[27] Joseph J Mattapallil, Daniel C Douek, Brenna Hill, Yoshiaki Nishimura, Malcolm Martin, and Mario Roederer. Massive infection and loss of memory cd4+ t cells in multiple tissues during acute siv infection. *Nature*, 434(7037):1093–1097, 2005.

[28] Meriet Mikhail, Bin Wang, Philippe Lemey, Brenda Beckthold, Anne-Mieke Vandamme, M John Gill, and Nitin K Saksena. Role of viral evolutionary rate in hiv-1 disease progression in a linked cohort. *Retrovirology*, 2(1):41, 2005.

[29] Laure Moutouh, Jacques Corbeil, and Douglas D Richman. Recombination leads to the rapid emergence of hiv-1 dually resistant mutants under selective drug pressure. *Proceedings of the National Academy of Sciences*, 93(12):6106–6111, 1996.

[30] Martin Nowak and Robert M May. *Virus dynamics: mathematical principles of immunology and virology*. Oxford university press, 2000.

[31] Martin A Nowak, Alun L Lloyd, Gabriela M Vasquez, Theresa A Wiltrout, Linda M Wahl, Norbert Bischofberger, Jon Williams, Audrey Kinter, Anthony S Fauci, Vanessa M Hirsch, et al. Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. *Journal of virology*, 71(10):7518–7525, 1997.

[32] Alan S Perelson. Modelling viral and immune system dynamics. *Nature Reviews Immunology*, 2(1):28, 2002.

[33] Ruy M Ribeiro, Li Qin, Leslie L Chavez, Dongfeng Li, Steven G Self, and Alan S Perelson. Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *Journal of virology*, 84(12):6096–6102, 2010.

[34] Ignacio A. Rodriguez-Brenes, Natalia L. Komarova, and Dominik Wodarz. The role of telomere shortening in carcinogenesis: A hybrid stochastic-deterministic approach. *Journal of Theoretical Biology*, 460:144 – 152, 2019.

[35] Rebecca A. Russell, Nicola Martin, Ivonne Mitar, Emma Jones, and Quentin J. Sattentau. Multiple proviral integration events after virological synapse-mediated hiv-1 spread. *Virology*, 443:443, 2013.

[36] Howard Salis and Yiannis Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *The Journal of chemical physics*, 122(5):054103, 2005.

[37] Quentin Sattentau. Avoiding the void: cell-to-cell spread of human viruses. *Nature Reviews Microbiology*, 6(11):815, 2008.

[38] Alex Sigal, Jocelyn T Kim, Alejandro B Balazs, Erez Dekel, Avi Mayo, Ron Milo, and David Baltimore. Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature*, 477(7362):95, 2011.

[39] David A Steinhauer and JJ Holland. Rapid evolution of rna viruses. *Annual Reviews in Microbiology*, 41(1):409–431, 1987.

[40] Bradford P Taylor, Catherine J Penington, and Joshua S Weitz. Emergence of increased frequency and severity of multiple infections by viruses due to spatial clustering of hosts. *Physical biology*, 13(6):066014, 2017.

[41] Paul E Turner and Lin Chao. Sex and the evolution of intrahost competition in rna virus $\varphi$6. *Genetics*, 150(2):523–532, 1998.

[42] Paul E Turner and Lin Chao. Prisoner's dilemma in an rna virus. *Nature*, 398(6726):441–443, 1999.

[43] E Weinan, Di Liu, and Eric Vanden-Eijnden. Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *Journal of computational physics*, 221(1):158–180, 2007.

[44] P. Whittle. The outcome of a stochastic epidemic—a note on Bailey's paper. *Biometrika*, 42(1-2):116–122, 06 1955.

[45] Dominik Wodarz and Natalia L Komarova. Mutant evolution in spatially structured and fragmented expanding populations. *Genetics*, 216(1):191–203, 2020.