

# Simulation Of A Simple Evolutionary System

Mahendra Duwal Shrestha

# Outline

Background

Question 1: Distance between finite population and infinite population

Question 2: Oscillation in finite population

Question 3: Oscillation in finite population under violation in mutation

Question 4: Oscillation in finite population under violation in crossover

Conclusion

# Terms

Population: Collection of binary string members of fixed length  $l$

Population vector: Population represented as population vector  $\mathbf{p}$  where  $p_j$  is the proportion of string  $j$  in the population.

Let population  $P = \{00, 01, 01, 10, 11, 11\}$ . Then  $p_3 = 2/6 = 1/3$

$\mathcal{R}$  denotes set of length  $l$  binary strings

Let  $l = 4$ ,  $x = 1101 \in \mathcal{R}$

Addition and multiplication of elements in  $\mathcal{R}$  are bitwise operations modulo 2

For example,

$$x = 1101, y = 1010$$

$$x + y = 1101 + 1010 = 0111$$

$$xy = 1101 \cdot 1010 = 1000$$

$$\bar{x} = 0010$$

# Operators

Linear fitness used with proportionate selection

Crossover and mutation both are mask based

Mask m is binary string of length  $l$

Operations are under  $\mathcal{R}$

Crossover : Choose two parents u and v , and exchange bits according to mask m following rule:

$$u' = um + \bar{vm}, v' = \bar{um} + vm$$

Ex:  $u = 11001011, v = 11011111, m = 11110000$

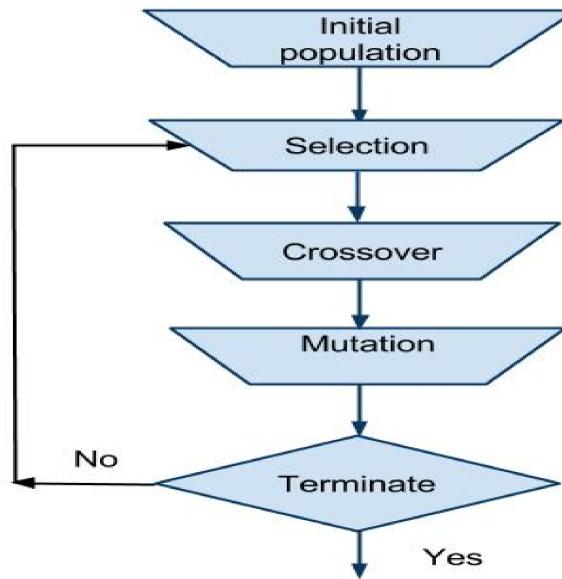
$$\textcolor{red}{11001011} + \textcolor{blue}{11011111} \rightarrow \{\textcolor{blue}{11001111}, \textcolor{green}{11011011}\}$$

Mutation: Flip bits of a string with some mask

Ex:  $11001001 + \underline{01000000} \rightarrow \textcolor{blue}{10001001}$

mutation mask

# Finite Population GA



Two parents  $u$  and  $v$  selected to produce two offspring  $u'$  and  $v'$  through crossover and mutation

Keep one of  $u'$  and  $v'$  with uniform probability

Replace  $P$  by new generations formed

Process is repeated until system stops to improve or threshold is reached

# Infinite Population Model

Population modeled as population vector  $\mathbf{p}$

$\mathcal{G}$  is the function mapping  $\mathbf{p}$  to the next generation

$\mathcal{G}(\mathbf{p})_j$  = the probability that string  $j$  occurs in the next generation

The evolution of  $\mathbf{p}$  in infinite population model is the sequence

$$\mathbf{p} \rightarrow \mathcal{G}(\mathbf{p}) \rightarrow \mathcal{G}(\mathcal{G}(\mathbf{p})) \rightarrow \dots$$

# History

Haldane, in 1932, summarized population genetics basic models : Wright, Fisher and Haldane

Several people working with evolution-inspired algorithms in the 1950s and the 1960s – Box (1957), Friedman(1959), Bledsoe (1961), Bremermann (1962), and Reed, Toombs and Baricelli (1967)

In 1960s and 1970s, Holland and colleagues formalized and promoted population based algorithms with crossover and mutation

Vose (1999) presented efficient methods for computing with a haploid model

# Random Heuristic Search

$\tau$  is stochastic transition rule that maps  $p$  to  $p'$

Finite population sequence  $p, \tau(p), \tau^2(p), \dots$  forms Markov chain

$\tau(p)$  cannot be predicted with certainty

$G(p)$  is the expected next generation  $E(\tau(p))$

The variance of the next generation is

$$E(\|\tau(p) - G(p)\|^2) = \frac{1 - \|G(p)\|^2}{r}$$

# Question 1:

Distance Between Finite Population and Infinite Population

# Distance: Chebyshev's Inequality

Let  $\epsilon = f(r)/\sqrt{r}$ , where  $f(r)$  grows arbitrarily slowly and

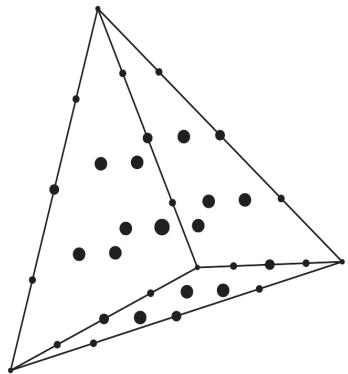
$$\lim_{r \rightarrow \infty} f(r) = \infty$$

From Chebyshev's inequality,

$$\lim_{r \rightarrow \infty} P(\|\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})\| \geq \epsilon) \leq \lim_{r \rightarrow \infty} \frac{1 - \|\mathcal{G}(\mathbf{p})\|^2}{f(r)^2} = 0$$

This suggests distance between  $\tau(\mathbf{p})$  and  $\mathcal{G}(\mathbf{p})$  might decrease as  $1/\sqrt{r}$

# Distance: Population Points



Finite populations are represented by dots

Infinite population can be anywhere in the space

Distance between finite population and infinite population is  $O(1/\sqrt{r})$

Figure 1.2: Population points

This suggests distance between  $\tau(\mathbf{p})$  and  $\mathcal{G}(\mathbf{p})$  might decrease as  $1/\sqrt{r}$

# Distance: Jensen's Inequality

Let  $\eta$  be the random variable  $\|\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})\|$  and convex function be  $\Phi(x) = x^2$

Then from Jensen's Inequality,

$$\mathcal{E}(\|\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})\|) = \mathcal{E}(\eta) \leq \sqrt{\mathcal{E}(\eta^2)} = \frac{\sqrt{1 - \|\mathcal{G}(\mathbf{p})\|^2}}{\sqrt{r}}$$

This also suggests distance might decrease as  $1/\sqrt{r}$

# Diploid Population Model

Diploid genome  $\alpha = \langle \alpha_0, \alpha_1 \rangle$

Haploid genome length  $l$

$q^n \rightarrow$  population at generation  $n$        $q_\alpha^n \rightarrow$  prevalence of diploid  $\alpha$

$t_\alpha(g) \rightarrow$  probability that gamete  $g$  is produced from parent  $\alpha$

$$q_\gamma^{n+1} = \sum_{\alpha} q_\alpha^n t_\alpha(\gamma_0) \sum_{\beta} q_\beta^n t_\beta(\gamma_1)$$

# Diploid Model Reduction to Haploid Model

$$q_{\langle \gamma_0, \gamma_1 \rangle}^n = p_{\gamma_0}^n p_{\gamma_1}^n$$

The weighted count of haploid  $g$  in generation  $n$  is,

$$p_g^n = \frac{1}{2} \sum_{\alpha_0, \alpha_1} q_{\langle \alpha_0, \alpha_1 \rangle}^n ([g = \alpha_0] + [g = \alpha_1])$$

Evolution equation in terms of haploid distributions  $p$ ,

$$p_{\gamma_0}^{n+1} = \sum_{\alpha_0, \alpha_1} p_{\alpha_0}^n p_{\alpha_1}^n t_{\langle \alpha_0, \alpha_1 \rangle}(\gamma_0)$$

Matrix form:

$$p'_g = p^T M_g p \quad \text{where} \quad (M_g)_{u,v} = t_{\langle u,v \rangle}(g)$$

# Specialization to Vose's Haploid Model

Mutation:  $\mu_i = (\boldsymbol{\mu})^{\mathbf{1}^T i} (1 - \boldsymbol{\mu})^{\ell - \mathbf{1}^T i}$

Crossover:  $\chi_i = \begin{cases} \chi^{c_i} & \text{if } i > 0 \\ 1 - \chi + \chi^{c_0} & \text{if } i = 0 \end{cases}$

For uniform crossover,

$$c_i = 2^{-\ell}$$

$$t_{\langle u, v \rangle}(g) = \sum_{i \in \mathcal{R}} \sum_{j \in \mathcal{R}} \sum_{k \in \mathcal{R}} \mu_i \mu_j \frac{\chi_k + \chi_{\bar{k}}}{2} [k(u+i) + \bar{k}(v+j) = g]$$

$$u \in \mathcal{R}, v \in \mathcal{R}$$

# Walsh Basis

$$W_{n,t} = N^{-1/2} (-1)^{n^T t}$$

$$\widehat{A} = WAW$$

$$\widehat{w} = Ww$$

Mixing matrix in Walsh basis

$$\widehat{M}_{u,v} = 2^{\ell-1} [uv = \mathbf{0}] \widehat{\mu}_u \widehat{\mu}_v \sum_{k \in \overline{u+v}\mathcal{R}} \chi_{k+u} + \chi_{k+v}$$

Evolution eqn in Walsh basis

$$\widehat{p}'_g = 2^{\ell/2} \sum_{i \in g\mathcal{R}} \widehat{p}_i \widehat{p}_{i+g} \widehat{M}_{i,i+g} \quad \text{where} \quad g\mathcal{R} = \{gi \mid i \in \mathcal{R}\}$$

# Computational Advantages

Specialization simplifies computation, which otherwise for diploid case would have been impractical

Only one mixing matrix as opposed to  $2^l$  is needed to compute next generation

For  $l = 14$ , using  $2^{14}$  mixing matrices with each having  $2^{14} \times 2^{14}$  entries would require 32 TB of memory, whereas one mixing matrix requires only 2 GB

# Distance Computation

Naive computation

$$\|f - q\|^2 = \sum_{\alpha} (f_{\alpha} - q_{\alpha})^2 \longrightarrow 2^{\ell} \cdot 2^{\ell} \text{ terms}$$

Our implementation

$$S_f = \{\alpha \mid f_{\alpha} > 0\}$$

$$\begin{aligned} \|f - q\|^2 &= \sum_{\alpha \notin S_f} (f_{\alpha} - q_{\alpha})^2 + \sum_{\alpha \in S_f} (f_{\alpha} - q_{\alpha})^2 \\ &= \sum_g (p_g)^2 + \sum_{\alpha \in S_f} f_{\alpha}(f_{\alpha} - 2q_{\alpha}) \end{aligned} \longrightarrow 2^{\ell} + |S_f| \text{ terms}$$

# Distance: Convergence

$$\begin{aligned}\chi &= 0.1 \\ \mu &= 0.001\end{aligned}$$

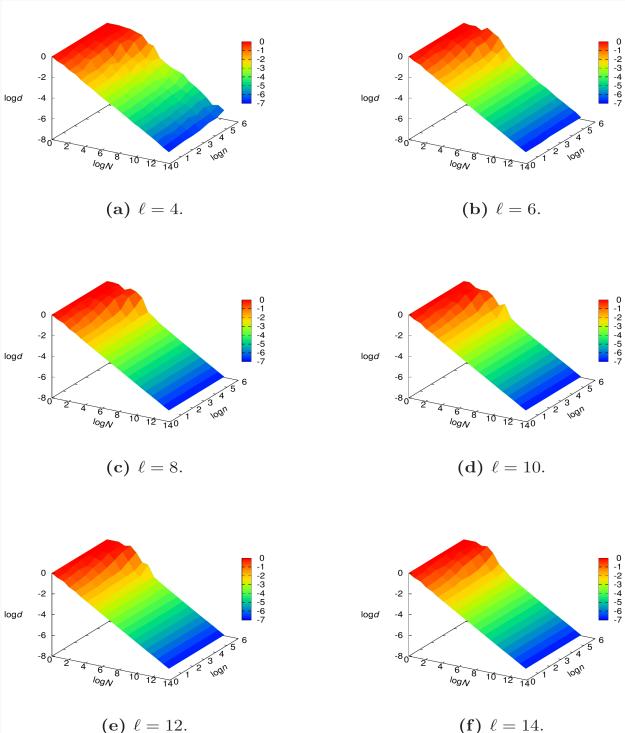
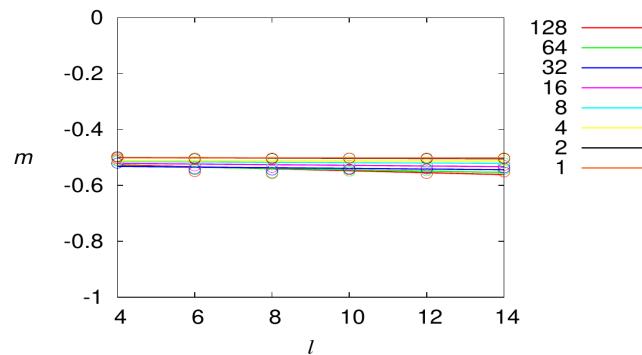


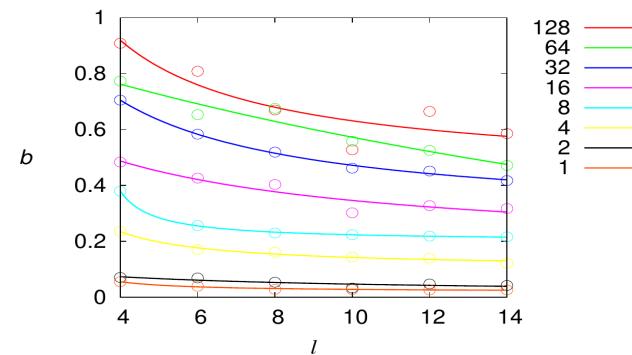
Figure: Convergence of finite population behavior

# Distance: Regression

$$\log d = m \log N + b$$



(a) Slope  $m$ , genome length  $\ell$ .



(b) Intercept  $b$ , genome length  $\ell$ .

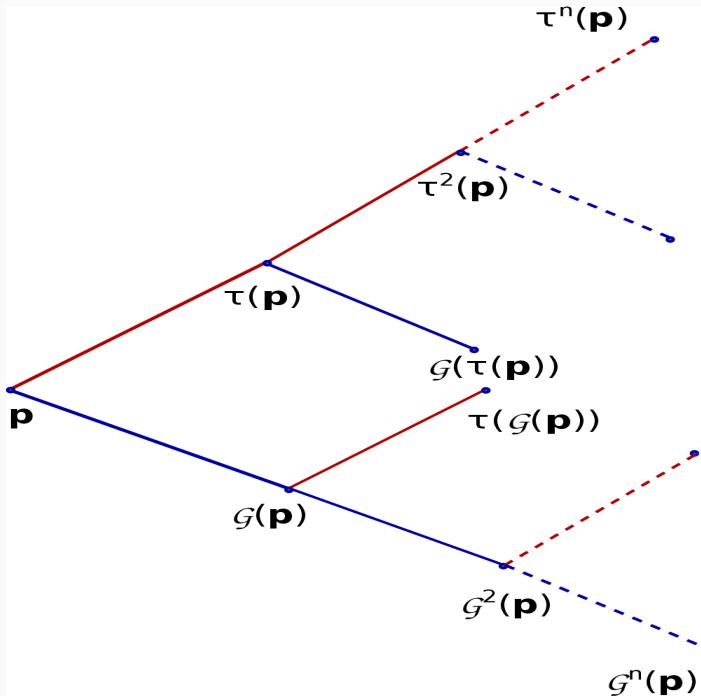
Figure: Regression parameters for generation  $n$  in  $\{1, 2, 4, 8, 16, 32, 64, 128\}$

$$d \approx N^m e^b$$

From figure (a) above,  $m \approx -0.5$

$$d \approx k / \sqrt{N}$$

## Distance: Non-Linearity



## Distance: Conclusion

Vose's infinite population model makes computation in diploid case efficient by reducing to the haploid case

Distance between finite diploid population and infinite diploid population can decrease like  $1/\sqrt{N}$

# Question 2:

Oscillation in Finite Population Evolution

# Limits

The sequence  $p$ ,  $\mathcal{G}(p)$ ,  $\mathcal{G}^2(p)$ , ... may converge to a fixed point

$$\mathcal{G}(\omega) = \lim_{n \rightarrow \infty} \mathcal{G}^n(p) = \omega$$

But under some circumstances, the sequence converges to a periodic orbit that oscillates between two fixed points,  $p^*$  and  $q^*$

# Necessary and Sufficient Conditions to Converge to a Periodic Orbit

For any  $g \in \mathcal{R}$ ,  $g \neq 0$

$$-1 = \sum_j (-1)^{g^T j} \mu_j$$

$$1 = \sum_{k \in \bar{g}\mathcal{R}} \chi_{k+g} + \chi_k$$

Infinite populations converge to a periodic orbit.

Do finite populations also exhibit oscillation from random initial populations?

## Previous Related Works on Oscillation

Akin (1982) proved existence of cycling for continuous time 2-bit diploid model

Hasting (1981) studied cycling in populations with infinite 2-bit diploid population model

Wright and Bidwell (1997) provided examples when cycles in an infinite haploid population model occur with crossover and mutation for 3 bit and 4 bit populations

Wright and Agapie (2001) described cyclings in infinite populations for up to 4 bits, and also presented data for cyclings in finite populations

# Difference From Previous Works

Akin's study was continuous time model and ours is discrete time model

Hastings' study was limited to two bits length, and includes only crossover but not mutation

Examples provided by Wright and Bidwell were for specific set parameter values for crossover, mutation and fitness

We study cyclic behavior for fixed fitness function and random: initial population, mutation and crossover distribution

Wright and Agapie used dynamic mutation that depends upon where population is in the population space while we use static mutation

Another difference between Wright and Agapie's work and ours is fitness

We study oscillation for higher bit length (up to 14), and actually oscillation conditions are independent of string length

We study oscillation for both haploid and diploid populations, and for both finite and infinite populations

We also visualize oscillation

# Simulation

Simulations were run for both haploid and diploid populations

Random initial population

$$I \in \{8, 10, 12, 14\}$$

$$N = \{4096, 40960, 81920\}$$

To visualize oscillation, distance between fixed points and population are plotted

# Results: Oscillation



Figure: Oscillation for  $l = 10$

# Results: Oscillation

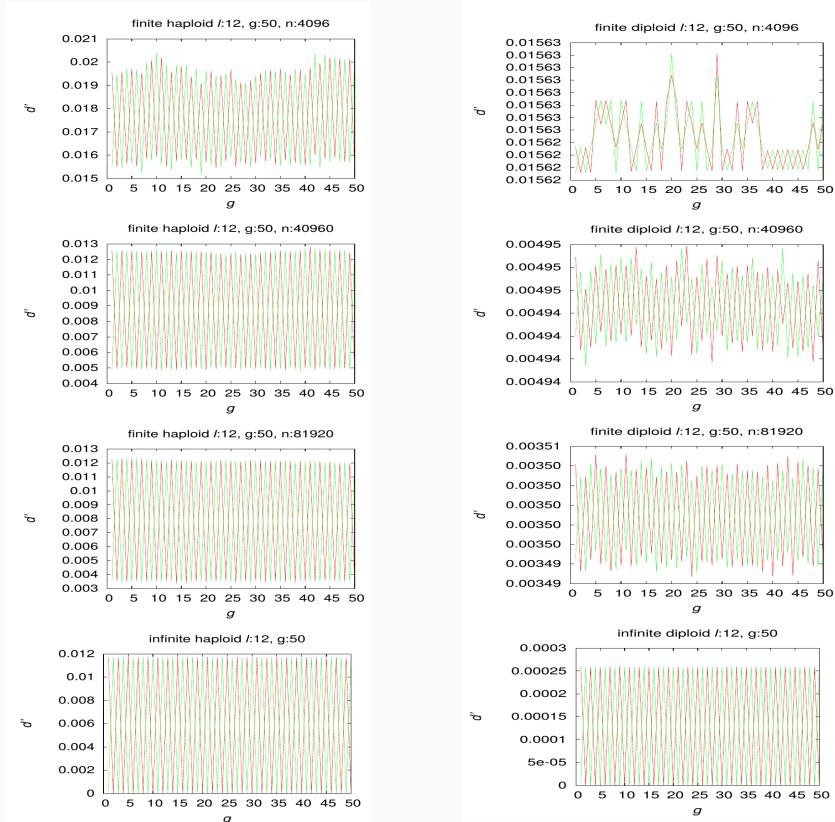


Figure: Oscillation for  $l = 12$

# Results: Oscillation

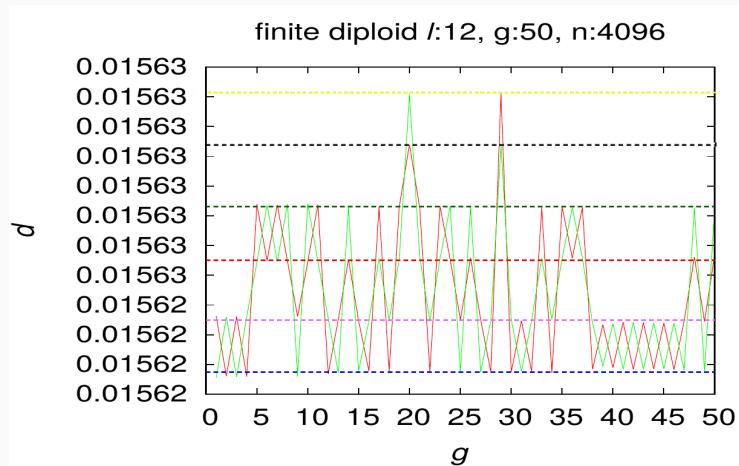


Figure: Oscillation between different fixed points

# Results: Oscillation Amplitude

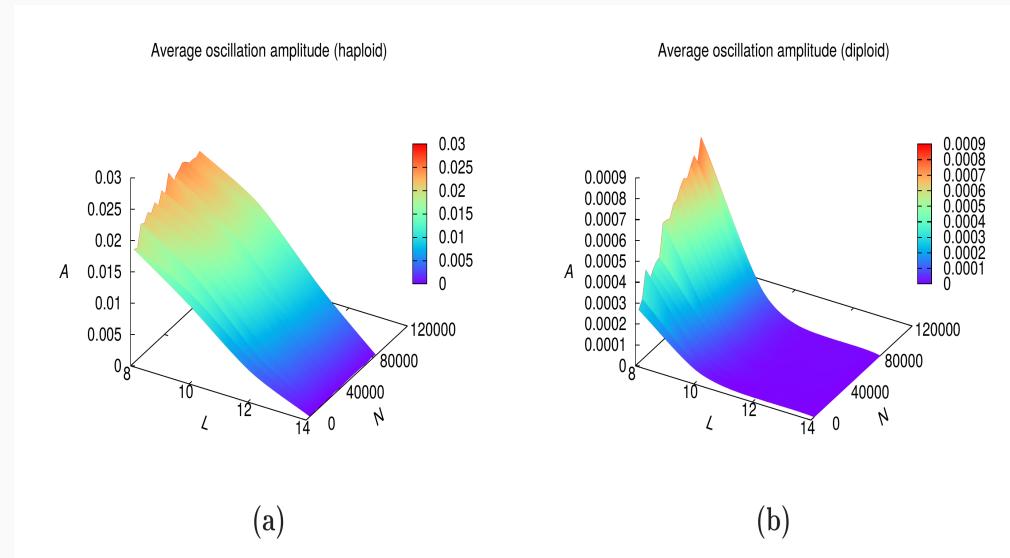


Figure: Average oscillation amplitude

# Oscillation: Conclusion

Finite population evolution exhibit approximate oscillation

As  $\lambda$  increases, oscillation amplitude decreases

As population size increases, oscillation amplitude increases and randomness decreases

Finite population also oscillates between different pairs of fixed points for diploid population of smaller size and larger  $\lambda$

# Question 3:

Oscillation Under Violation in Mutation

# Robustness of Finite Population

A Markov chain is said to be *irreducible* if it is possible to get to any state from any state

A Markov chain is *aperiodic* if it can return to state  $i$  at irregular times

Markov chain is *regular* if it is both irreducible and aperiodic

Steady state distribution exists if Markov chain is regular

No periodic orbit exists for finite population

Can finite population exhibit approximate oscillation?

# Violation in Mutation

Violation  $\epsilon$  is introduced in  $\mu$

$$\mu_0 = \epsilon$$

$$\mu_i := (1 - \epsilon)\mu_i$$

This modification makes the Markov chain regular

No periodic orbits for infinite population

# Simulation

$$\epsilon = \{0.01, 0.1, 0.5\}$$

$$l \in \{8, 10, 12, 14\}$$

$$N = \{4096, 40960, 81920\}$$

Distances of population to limits  $p^*$  and  $q^*$  without violation are plotted

# Results: Violation in Mutation

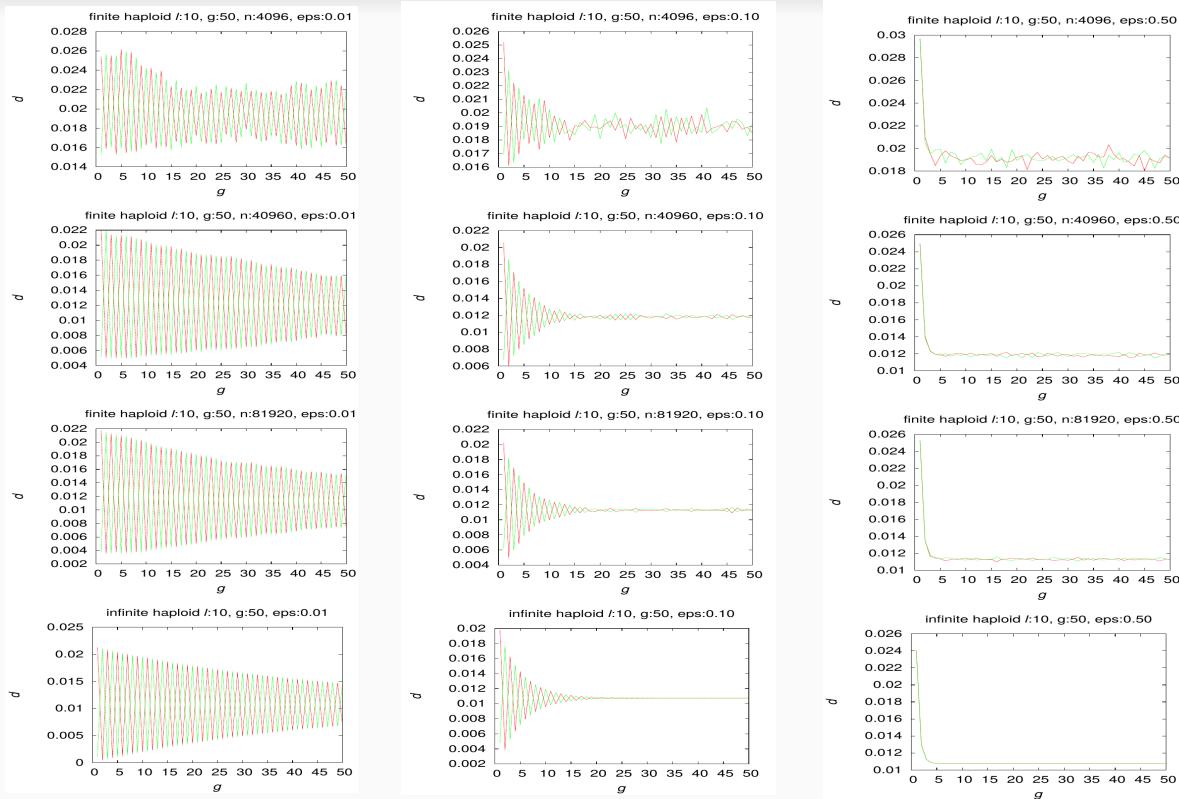


Figure: Oscillation under violation in mutation for haploids,  $l = 10$

# Results: Violation in Mutation

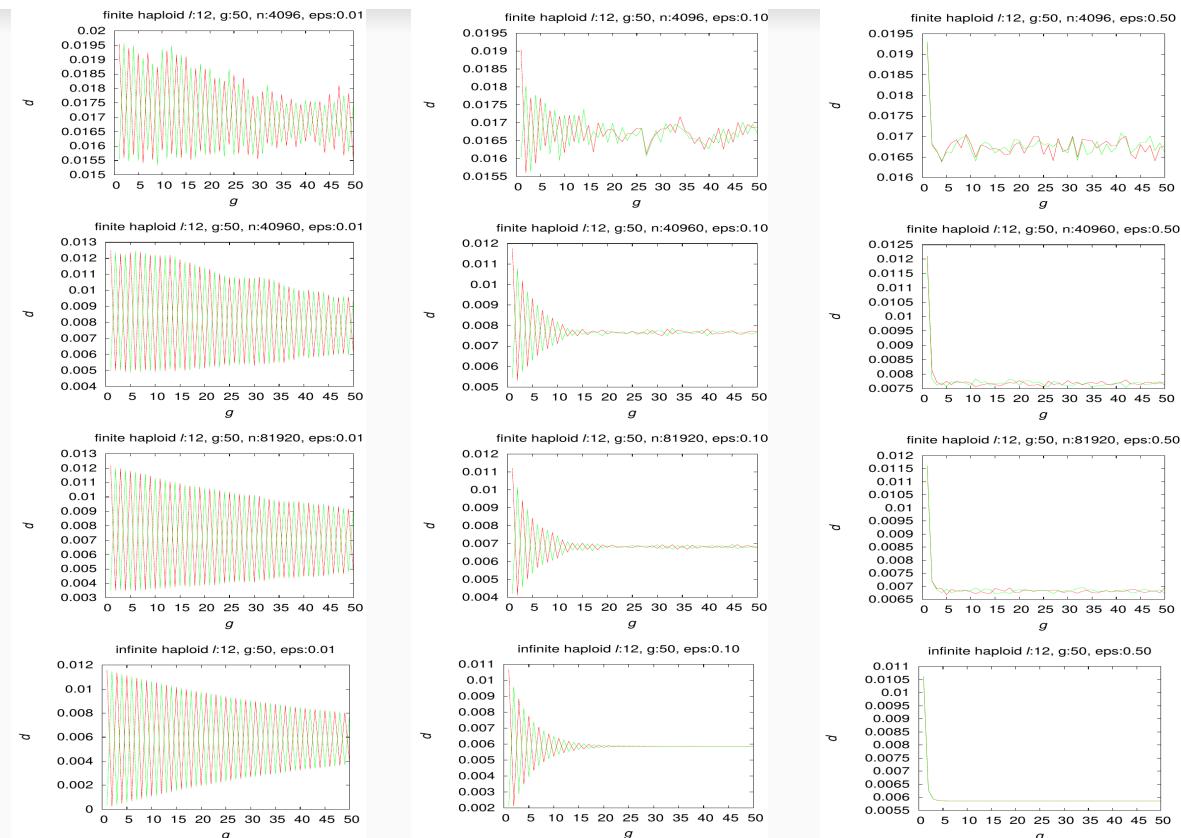


Figure: Oscillation under violation in mutation for haploids,  $I = 12$

# Results: Violation in Mutation

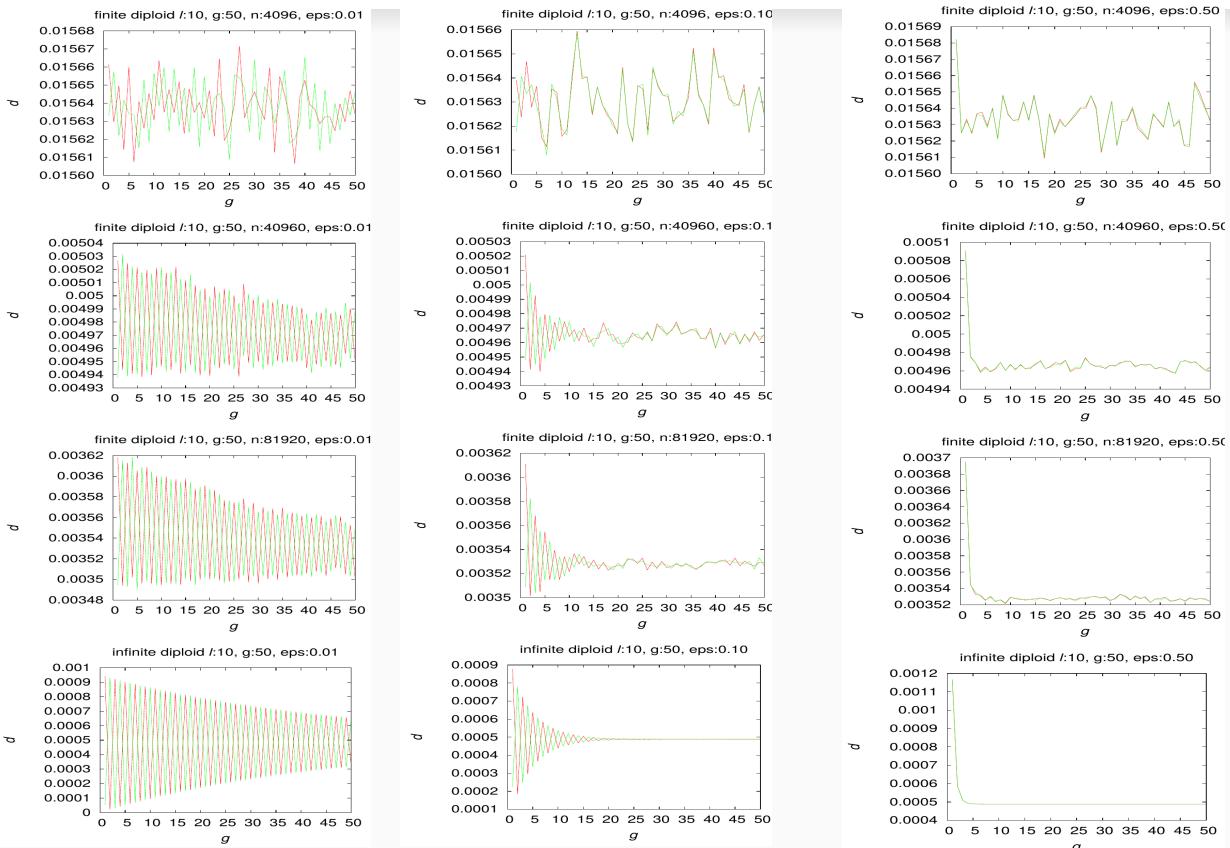


Figure: Oscillation under violation in mutation for diploids,  $l = 10$

# Results: Violation in Mutation

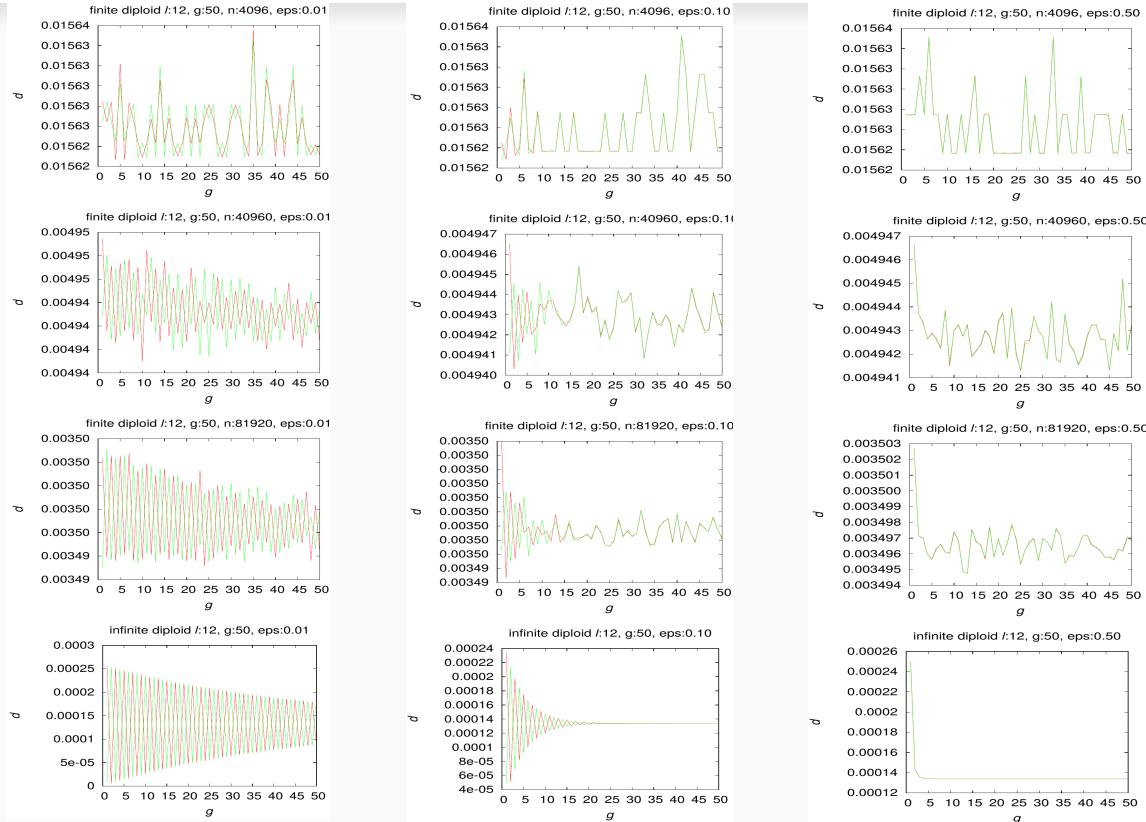


Figure: Oscillation under violation in mutation for diploids,  $\ell = 12$

## Violation in Mutation: Conclusion

Finite population evolution exhibits approximate oscillation even if Markov chain is regular when violation level is small

As string length increases, oscillation degrades

# Question 4:

Oscillation under Violation in Crossover

# Robustness of Finite Population

Violation in crossover condition means no periodic orbit exists for infinite population

But we don't know if Markov chain is regular

Can finite population exhibit approximate oscillation?

## Violation in Crossover

Violation  $\epsilon$  is introduced in  $\chi$

$$\chi_i := (1 - \epsilon)\chi_i$$

$$\chi_j = \epsilon \quad j \text{ is chosen where } \chi_j = 0$$

## Simulation

$\epsilon = \{0.01, 0.1, 0.5\}$

$I = \{8, 10, 12, 14\}$

$N = \{4096, 40960, 81920\}$

Distances of population to limits  $p^*$  and  $q^*$  without violation are plotted

# Results: Violation in Crossover

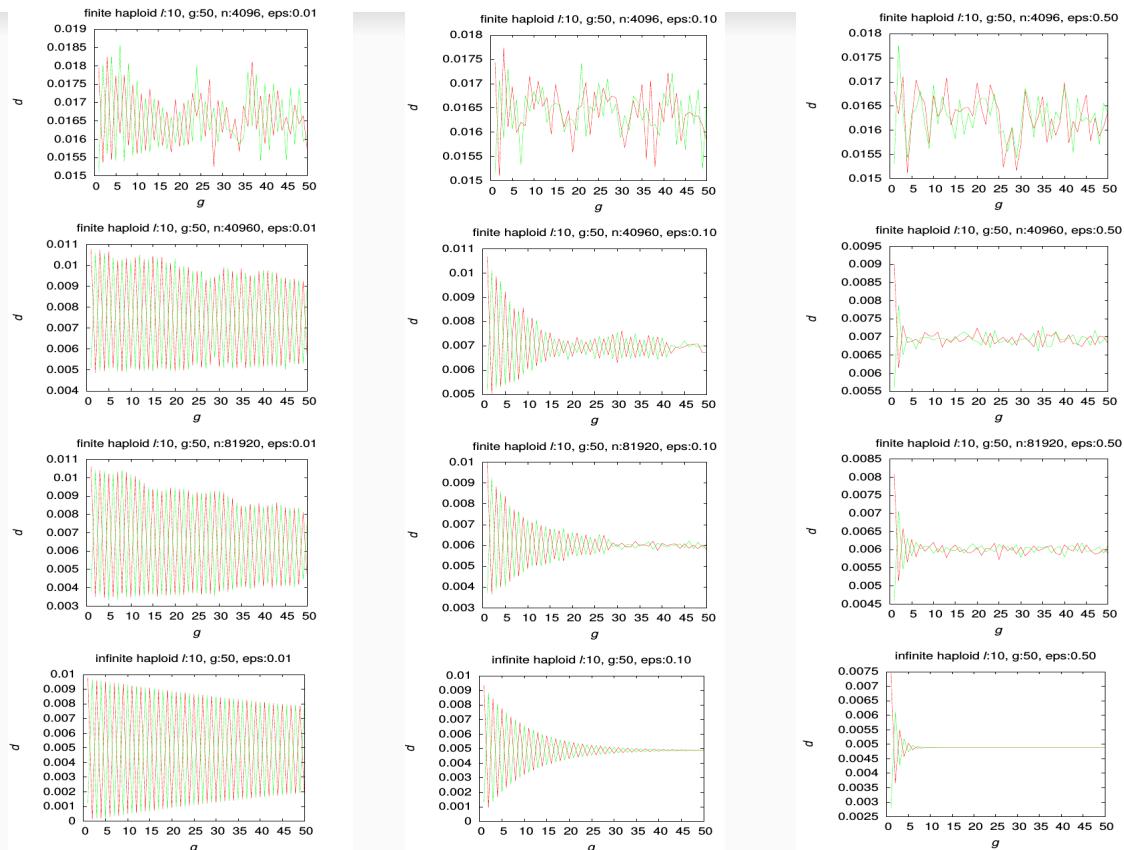


Figure: Oscillation under violation in crossover for haploids,  $\lambda = 10$

# Results: Violation in Crossover

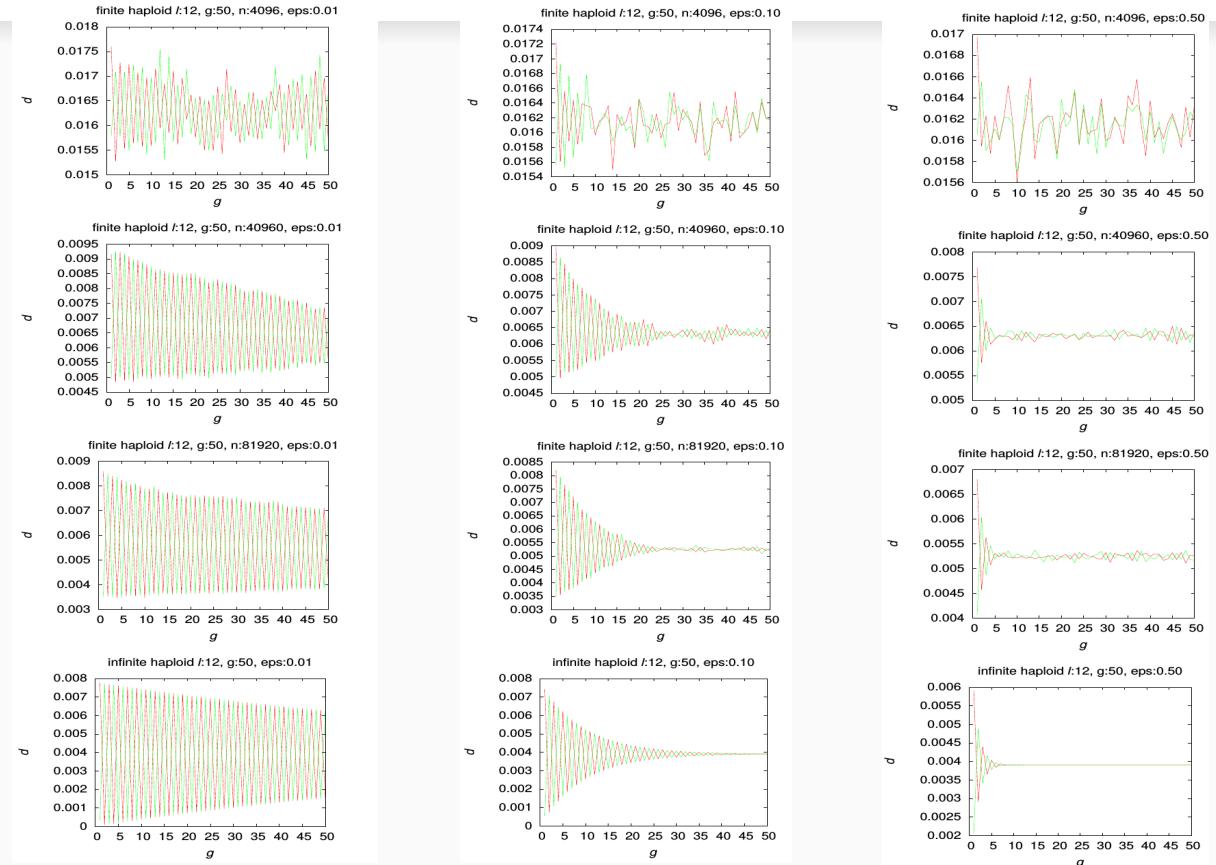


Figure: Oscillation under violation in crossover for haploids,  $l = 12$

# Results: Violation in Crossover

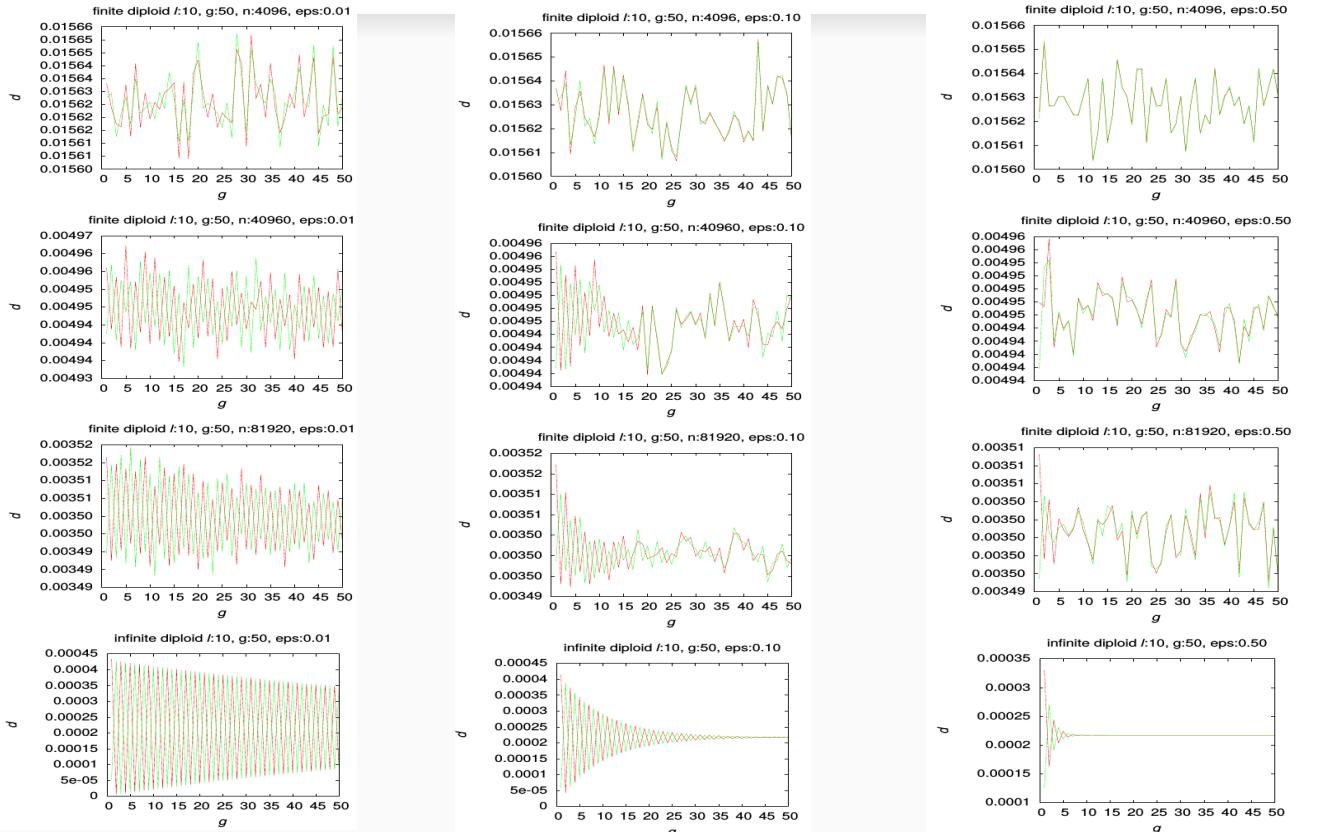


Figure: Oscillation under violation in crossover for diploids,  $l = 10$

# Results: Violation in Crossover

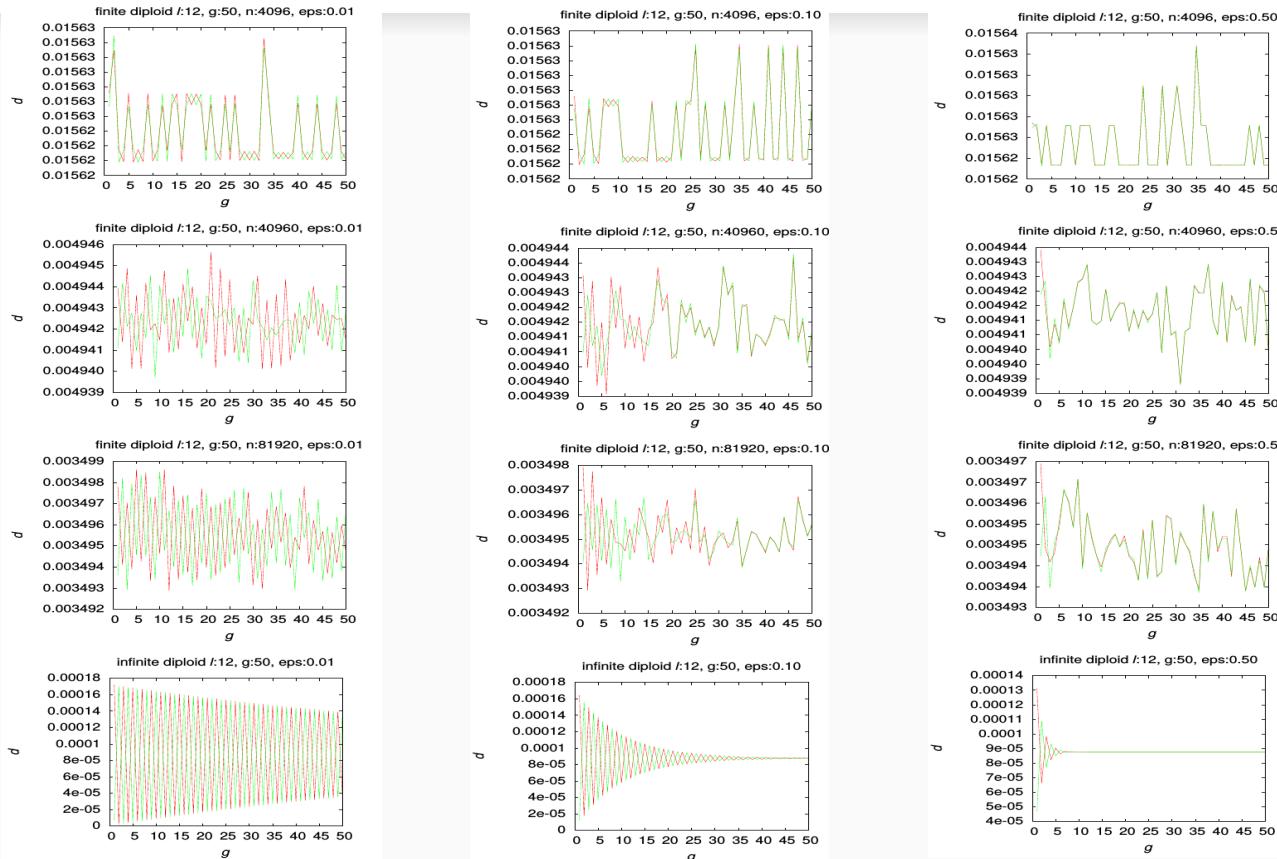


Figure: Oscillation under violation in crossover for diploids,  $I = 12$

## Violation in Crossover: Conclusion

Finite population evolution exhibits approximate oscillation if the violation level is small

If violation is large, then finite population ceases to oscillate

Dying out of oscillation is slower than in violation in mutation case

More randomness are observed with violation in crossover than in mutation

# Conclusion

# Conclusion

Vose's haploid model makes computation efficient in diploid case given uniform fitness by reducing to haploid case

Distance between finite population and infinite population can decrease like  $1/\sqrt{N}$

When necessary and sufficient conditions for infinite population to converge to a periodic orbit, finite population evolution also exhibits approximate oscillation

Even when Markov chain is regular, finite population exhibits approximate oscillation for smaller level of violation

When condition for crossover distribution is violated, Markov chain may not be regular

Finite population exhibits approximate oscillating behavior in case of violation of crossover distribution for smaller level of violation