Percolation of a bit-string model

S. Taneri[a]

Feza Gürsey Institute PO Box 6, 81220 Çengelköy, Istanbul, Turkey

Abstract. We investigate the effect of mutations on adaptability in a bit-string model of invading species in a random environment. The truncation-like fitness function depends on the Hamming distance between the optimal (wild)-type at each site and the invading species for a square lattice. We allow invasion if the relative fitness is above or equal to an adjustable threshold. We have also allowed for the decay and extinction of a species at a site that it has already invaded. We find that the invading species always percolates through regions of arbitrary size, for all threshold values, with a time parameter which depends on the threshold and the size in the absence of decay. If decay is introduced then there is a critical value of the threshold variable beyond which the invading species is confined. Radius of gyration and average population of a colony of mutants have a power-law dependence with time and relevant fractal dimensions are calculated for percolation.

PACS. 02.60.Lj Ordinary and partial differential equations; boundary value problems – 05.10.Ln Monte Carlo methods – 05.45.Df Fractals

1 Introduction

It is well known that neutral mutations play a very important role in evolution [1], as they allow the quasi species to explore the nearby genetic phase space and thereby to adapt to new environmental conditions which may select for traits differing from the wild type or the consensus genome. The extremely high mutation rates encountered in bacteria or viruses, on the other hand, make it very difficult to combat them, as new breeds prove resistant to the antibodies developed. It has recently been conjectured [2] that the HIV infection proceeds not via a suppression of the immune system, but due to the extremely rapid mutation of the virus, so that resistant breeds keep appearing as soon as new antibodies are generated.

In this paper we introduce a model which may be regarded as either a geographical landscape or an organism being invaded by an array of neutral mutants of some species. The different mutants as well as the local ideal-type are represented by random bit-strings consisting of 1's and 0's, which can be written $\boldsymbol{v}_i(0, 1, ...)$ and $\boldsymbol{h}_i(0, 1, ...)$ respectively, where they indicate lattice sites. The fitness function is defined as $f(d_{\boldsymbol{ij}}) = 1 - d_{\boldsymbol{ij}}$ where

$$d_{\boldsymbol{ij}} = \frac{1}{l} \sum_{\alpha=1}^{l} |v_{\boldsymbol{i}\alpha} - h_{\boldsymbol{j}\alpha}| \qquad (1)$$

is the distance between the genome of the organism, $\boldsymbol{v}_i$ at site $\boldsymbol{i}$, and the ideal type, $\boldsymbol{h}_j$ at site $\boldsymbol{j}$ which are nearest

neighbors. The site $\boldsymbol{j}$ is invaded by the organism occupying that nearest neighbor of $\boldsymbol{j}$ which satisfies the condition $f(d_{\boldsymbol{ij}}) \geq r$ nearest neighbor bit-string characteristics being carried, and is left vacant if no nearest neighbor satisfies this condition. The parameter r may be regarded as a barrier height, or, alternatively $1 - r$ the maximum allowed relative distance from the ideal type at site $\boldsymbol{j}$. At a given constant mutation rate μ, the neighboring organism may eventually suffer such mutations as to bring it closer to the target site, in which case it will eventually occupy it. We consider two variants of this model: A) an organism which can not surmount the surrounding barriers is allowed to stagnate at that site. B) The organism at any given site is killed off (decays) with a probability equal to the mutation rate or with the probability to decay. The cases are denoted as WOD (without decay) and WD (with decay) respectively, in what follows.

The radius of gyration of a growing colony of mutants behaves as

$$R_g^2(t) \sim t^\gamma \qquad (2)$$

while the total number of occupied sites $n(t)$, goes as

$$n(t) \sim t^\beta \qquad (3)$$

where the exponents take different values over the initial, intermediate and late stages of the growth. In these different regimes,

$$n \sim R_g^{d_f} \qquad (4)$$

yields the relation $d_f = 2\beta/\gamma$ [3–7] for the fractal dimension of the clusters. Our simulation results for the time

[a] e-mail: taneri@gursey.gov.tr

Table 1. WOD case γ and β values for various fitness threshold and mutation rates.

r	μ	γ	β
0.0625	0.00	1.996 ± 0.000	1.996 ± 0.000
0.125	0.00	1.996 ± 0.000	1.996 ± 0.000
0.1875	0.00	1.996 ± 0.000	1.996 ± 0.000
0.25	0.00	1.996 ± 0.000	1.996 ± 0.000
0.3125	0.00	1.996 ± 0.001	1.996 ± 0.000
0.375	0.00	2.002 ± 0.003	2.000 ± 0.002
0.4375	0.00	2.018 ± 0.002	2.016 ± 0.002
0.5	0.00	1.70 ± 0.08	1.60 ± 0.08
0.5	0.01	2.021 ± 0.001	2.022 ± 0.008
0.5625	0.01	2.05 ± 0.04	2.08 ± 0.02
0.625	0.01	2.06 ± 0.02	2.06 ± 0.02
0.6875	0.01	2.09 ± 0.04	2.11 ± 0.04
.....	.....	.....	.....

Table 2. WD case γ and β values for various fitness threshold and mutation rates.

r	μ	γ	β
0.5	0.01	2.016 ± 0.007	2.03 ± 0.01
0.5	0.5	2.020 ± 0.006	2.019 ± 0.008
0.5625	0.01	2.06 ± 0.01	2.047 ± 0.009
0.5625	0.5	2.04 ± 0.02	2.04 ± 0.03

Table 3. WOD case fractal dimensions for various fitness threshold and mutation rates calculated by the first method.

r	μ	d_{f_1}
....	....	....
0.3125	0.00	2.000 ± 0.001
0.375	0.00	1.998 ± 0.005
0.4375	0.00	1.998 ± 0.004
0.5	0.00	1.88 ± 0.09
0.5	0.01	2.001 ± 0.004
0.5625	0.01	2.03 ± 0.06
0.625	0.01	2.00 ± 0.04
0.6875	0.01	2.02 ± 0.08
.....	.....	......

Table 4. WD case fractal dimensions for various fitness threshold and mutation rates calculated by the first method.

r	μ	d_{f_1}
0.5	0.01	2.01 ± 0.02
0.5	0.5	2.00 ± 0.01
0.5625	0.01	1.99 ± 0.02
0.5625	0.5	2.00 ± 0.05

dependence of the radius of gyration and the total population are shown in Tables 1 and 2.

A second method via which we may determine the fractal dimension d_f of $M \sim L^{d_f}$ where L is the length of the sample lattice size, is the box-counting method, which is also implemented for comparison (see Tabs. 5 and 6).

Initially, for a totally random organism at a random site, if n is the number of 1's that is the number of mismatches between the genome of the organism and genome of the ideal type at the site, the fitness will be $f = 1 - \frac{n}{l}$ where l is the length of the bit-strings. On the other hand,

Table 5. WOD case fractal dimensions for various fitness threshold and mutation rates calculated by the second method.

r	μ	d_{f_2}
0.375	0.00	2.0000 ± 0.0000
4375	0.00	1.9991 ± 0.0004
0.5	0.00	1.87 ± 0.05
0.5	0.01	2.00 ± 0.00
0.5625	0.01	2.00 ± 0.00
0.625	0.01	2.00 ± 0.00
0.6875	0.01	2.00 ± 0.00
......	.....	.....

Table 6. WD case fractal dimensions for various fitness threshold and mutation rates calculated by the second method.

r	μ	d_{f_2}
0.5	0.01	1.9992 ± 0.0000
0.5	0.5	1.9999 ± 0.0000
0.5625	0.01	1.9874 ± 0.0004
0.5625	0.5	1.9956 ± 0.0002

the probability that the number of mismatches will be n is,

$$P(n) = \frac{1}{2^l} \frac{l!}{(l-n)!n!}. \tag{5}$$

If the fitness tolerates a minimum number m of matching alleles, the fitness threshold is then $r = \frac{m}{l}$. For the probability of fitness to be greater than or equal to the fitness threshold, namely the probability that $l - n \geq m$ is,

$$g_1(r) = \frac{1}{2^l} \sum_0^{l-m-1} \frac{l!}{(l-n)!n!} \quad \text{for} \quad \frac{m}{l} \leq r < \frac{m+1}{l}. \tag{6}$$

We moreover require that the organism dies off with a probability $g_2(r) = 1 - g_1(r)$ in the WD model. We define this quantity as the decay probability. Clearly, at later stages of the evolution of the interface, $g_1(r)$ will acquire a history and site dependence. Neglecting this dependence as we do here, gives rise to an effective Mean Field (MF)-like approximation.

2 Monte Carlo simulation

We performed discrete time simulations of our model on a square lattice. Organisms $\boldsymbol{v_i}$ are allowed to start from the center of a 513×513 lattice in our model. The templates or "wild types" associated with each site $\boldsymbol{h_j}$ as well as the organisms $\boldsymbol{v_i}$ consist of randomly chosen bit-strings of length $l = 16$. At each time step, the bit-strings of the organisms occupying the sites are mutated with a probability μ per bit, according to $\mod(x+1, 2)$ to simulate the mutations where x is 0 or 1. An already occupied site can not be occupied by a fitter neighbor. The results are averaged over 1000 different runs with different initial conditions (different initial $\boldsymbol{v_i}$) in each case for the simulations.

Even in the absence of mutation, percolation was achieved for fitness threshold $r \leq 0.5$ in our model and

Table 7. Percolation times for the method without decay (*WOD*) case for various fitness threshold and mutation rates. t^{SA} stands for semi-analytic calculation (Sect. 3) and t^E stands for the entropy approach (Sect. 4).

r	$g_1(r)$	μ	MCS Time	t^{SA}	t^E
0.0625	1.000	0.00	256 ± 0	256	256
0.125	0.998	0.00	256 ± 0	256	257
0.1875	0.989	0.00	256.0 ± 0.2	256	261.7
0.25	0.962	0.00	256.9 ± 0.7	258	275.8
0.3125	0.895	0.00	260.4 ± 1.1	261	312.2
0.375	0.773	0.00	270.9 ± 1.7	271	381.2
0.4375	0.598	0.00	298.2 ± 3.0	295	473.5
0.5	0.402	0.00	493.7 ± 63.7	361	$512(g_{1eff}(r) = 0.5)$
0.5		0.01	324.3 ± 2.7		
0.5625		0.01	408.8 ± 4.3		
0.625		0.01	609.8 ± 8.8		
0.6875		0.01	1152.0 ± 17.9		
...		...	...		

no matter what the mutation rate is there was no percolation beyond the fitness threshold $r = 0.5625$ for *WD* case in the presence of mutations. There is always percolation as long as there is mutation for *WOD* case. It is hard to make an analytical calculation for the cases with mutation and percolation happens only for $m \leq 9$ in *WD* cases as it was pointed out earlier.

3 Semi-analytical approach

We define the percolation time as the time it takes for the organism to reach one of the edges of the lattice. We do the following to estimate this time. One can write the following for the population density at site (x, y) at time t as,

$$\frac{\partial P(x, y, t)}{\partial t} = D\nabla^2 P(x, y, t) - \Gamma_1 P(x, y, t) \qquad (7)$$

where,

$$D = \lim_{\Delta x \to 0 \, \Delta t \to 0} g_1(r)\frac{\Delta x^2}{\Delta t} \qquad (8)$$

is the diffusion coefficient and

$$\Gamma_1 = \lim_{\Delta t \to 0} \frac{g_2(r) - 4g_1(r)}{\Delta t} \qquad (9)$$

is the effective decay rate, within the *MF*-like approximation introduced in the last section. Then,

$$P(x, y, t) = \frac{1}{4\pi Dt}e^{-\frac{x^2+y^2}{4Dt} - \Gamma_1 t} \qquad (10)$$

is a solution with initial condition $P = \delta(x)\delta(y)$ at time equals zero. This solution can be used to calculate the time dependence of radius of gyration by integrating the square of radius of the site weighted by the population density. The result to the first order is simply,

$$R_g^2 \sim t^2. \qquad (11)$$

The velocity of the wave fronts is given by an equation similar to equation (7) known as the two dimensional form of the Fisher-Kolmogorov-Petrosvky-Piscounov equation,

$$\frac{\partial P_{wf}(x, y, t)}{\partial t} =$$
$$D\nabla^2 P_{wf}(x, y, t) - \Gamma_1 P_{wf}(x, y, t) - P_{wf}(x, y, t)^3 \quad (12)$$

where $P_{wf}(x, y, t)$ is the evolution of population density on the wave front [8]. Neglecting the cubic term and using symmetry of the growth, the traveling wave solution,

$$P_{wf}(x, y, t) \sim \exp(-k(x - vt) - k(y - vt)) \qquad (13)$$

can be substituted into the equation where v is the velocity. The discriminant of the second order equation in k should be negative to have a traveling wave solution rather than pure decay. This condition together with $|v| < 1$ condition for the velocity (the probability of percolating to the nearest neighbor site is less than unity at each unit time step) yields $g_1(r) > 0.432$ for *WD* case. Remembering that $g_1(0.5) = 0.402$ one can choose $r < 0.5$. Thus, $r = 0.5, \mu = 0.0$ is the critical point. One should conclude that the differential equation does not represent the physical phenomenon of the fractal growth at the critical point but serves as to identify the upper bound for normal growth. Considering the microscopic cutoffs at this point is irrelevant here, since the solutions and the results of simulations are treated for large times and flat fronts.

Time to reach the edge of the lattice, i.e. a percolation time, can be estimated numerically as shown in Table 7. Here, a discretized form of the differential equation (8) is solved for lattice size $L = 513$ as a fast simulation method,

$$P(x, y, t + \Delta t) = P(x, y, t) + g_1(r)(P(x + \Delta x, y, t)$$
$$+ P(x - \Delta x, y, t) + P(x, y + \Delta y, t)$$
$$+ P(x, y - \Delta y, t)) - g_2(r)P(x, y, t), \quad (14)$$

which serves as sort of a hybrid model for the Monte Carlo Simulation *(MCS)*. Such hybrid models have been considered before in the literature [9] for reaction diffusion processes which is similar to the case we study here. Thus,
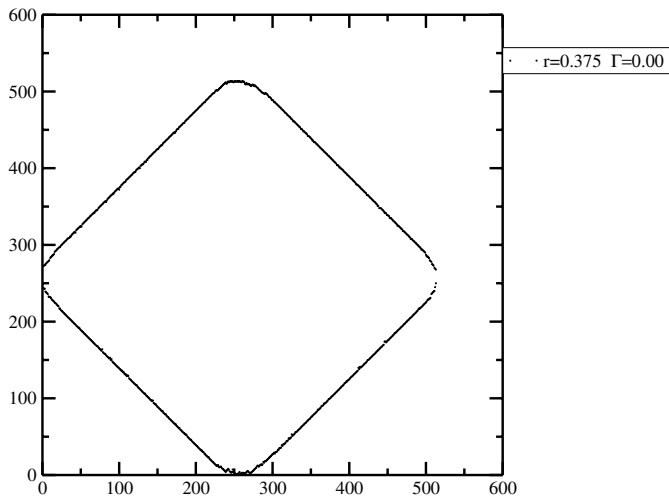
Fig. 1. Surface of population for fitness treshold $r = 0.375$ and mutation rate $\mu = 0.00$. The growth starts from the center of point of the lattice in all subsequent figures.
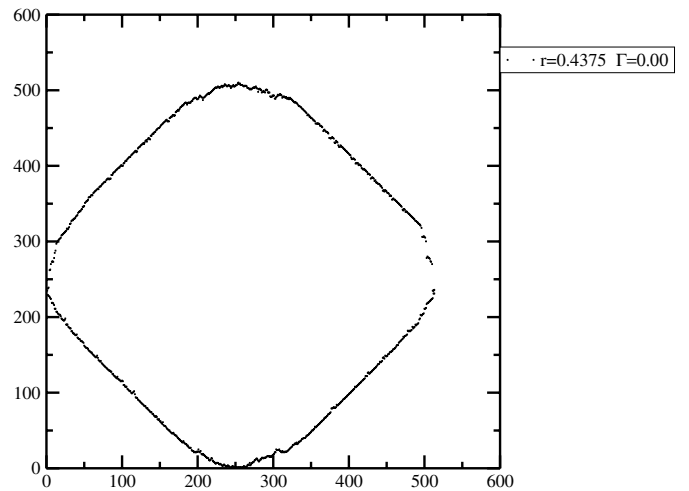

Fig. 2. Surface of population for fitness treshold $r = 0.4375$ and mutation rate $\mu = 0.00$.

our semi-analytical approach can be considered as an interpolation to the growth rule which we implement via a discrete lattice MC simulation as described before. The initial conditions are $P(0,0,0) = 1$ at the origin (at the center) of the lattice, and $P(x,y,0) = 0$ elsewhere. The population density at each site is updated at each iteration of the algorithm according to equation (11). Since the argument of the exponential factor $\Gamma_1 t$ in the solution (Eq. (10)) pushes the population density to values larger than 1 which is prohibited according to the growth rule, what we do is to enforce the restriction $P(x,y,t) \leq 1$ by hand in the algorithm. This is equivalent to implementing equation (11) given a particular interface configuration until the value unity is attained at any one of the neighboring sites, and then freezing that value unless it becomes smaller than unity once again because of the nonzero decay probability.

The percolation time is measured as the first instant that the value of population density reaches 1 at some site on the edges of the lattice. This analysis gives the percolation times in agreement with MCS results when r is smaller than 0.5 (See Tab. 7). The outcome patterns of surface growth for the fast simulation method introduced in this section overlaps with those of Figures 1 and 2. The cases for $r < 0.375$ are always diamond like figures and not shown here.

4 The entropy approach

That the percolation is achieved for fitness threshold $r \leq 0.5$ in the absence of mutations can be proven by the following argument. In terms of number states and probability, entropy can be written as,

$$S = k \ln \Omega = -k \sum_i P_i \ln P_i \qquad (15)$$

where Ω is the weighted phase space summation, and the characteristic time of the system (considering the fact that time will be proportional to the number of sites visited in the configuration space) is then

$$t \sim \Omega = \prod_i P_i^{-P_i}. \qquad (16)$$

We approximate the partial phase space volume Ω_{ij} for two adjacent sites, ij, by

$$\Omega_{ij} \sim P_i^{-P_i} P_j^{-P_j}. \qquad (17)$$

Within the MF-like approach where we have assumed the probability for invasion P_i of a site to be given by $g_1(r)$ (independently of i), the hopping time t_{hop} becomes, to the leading order,

$$t_{hop} \sim (g_1(r)^{-g_1(r)})^2. \qquad (18)$$

On the other hand, we estimate the time during which the interface will be frozen at an occupied site, with an empty neighboring site, by

$$t_{frozen} \sim g_1(r)^{-g_1(r)} g_2(r)^{-g_2(r)} \qquad (19)$$

within the same approximation. If $t_{hop} \leq t_{frozen}$ there will be chance to spread and thus percolation. This condition yields $g_1(r) \geq 0.5$ for the percolation threshold in the absence of mutations. From Table 7 we see that this condition is satisfied for $r < 0.5$, and the interface will certainly reach one of the edges of the lattice. Here, the MF-like semi-analytical approach discussed in Section 3 can successfully be accomplished, while $r = 0.5$ case may or may not reach there as it is observed in our MC simulations. Thus, the MF approach breaks down in the vicinity of the critical point (percolation threshold).
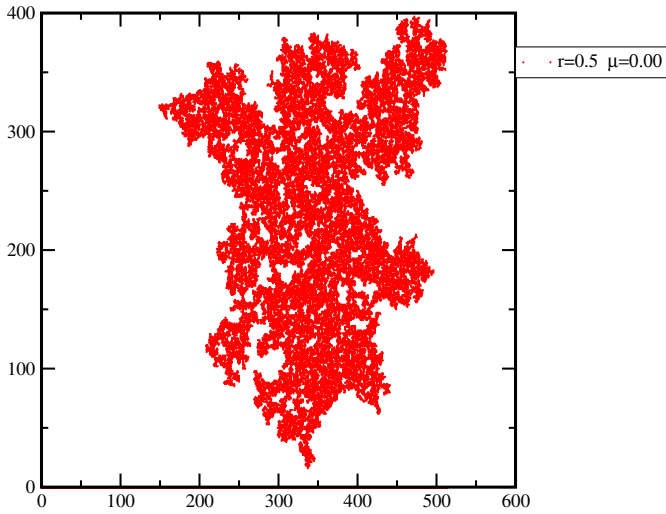
Fig. 3. Population for fitness treshold $r = 0.5$ and mutation rate $\mu = 0.00$.
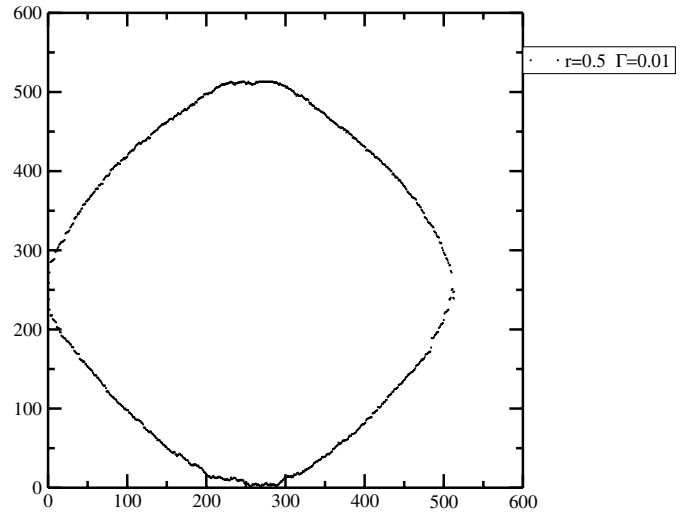
Fig. 4. Surface of population for the case without decay for fitness treshold $r = 0.5$ and mutation rate $\mu = 0.01$.

We will now estimate the time of percolation to the edge in case it reaches there. The time to reach the edge of the lattice for $r = 0.5$ can be calculated by

$$t_{percolation} = t_{hop} \times \frac{L-1}{2} \qquad (20)$$

We find, $t_{percolation} = 512$ (with the 0.5 effective value of $g_{1eff}(0.5)$) if $L = 513$ to be compared with the earlier results in Table 7 as the second way of estimating the percolation time. (Our result seems to be off by about 18 units, however for orthogonal growth case one finds 511.3 ± 80.0 for MCS, which is off by an amount less than 1 unit. Here, we mean the consideration of arrival to the upper edge or the righthand edge of the lattice by the orthogonal growth case.) This type of calculation gives the best estimate for percolation of $r = 0.5$ and $g_{1eff}(r) = 0.5$ (it has also been checked for different lattice sizes), since in this case, the probability for invading a neighboring site is equal to the probability that this is prohibited ($\Omega_{hop}^{-1} = \Omega_{frozen}^{-1}$). That is, the system is equally likely to be found in any one of its accessible states. Therefore, the above approach can be applied successfully, due to the definition of the micro-canonical ensemble.

5 Results

Figures 1–4 depicts the surface of invaded sites at the time the percolation is achieved for the MCS. The values of γ and β for large times and, WOD and WD cases are given in Tables 1 and 2. The values of γ and β for $r = 0.5$ and $\mu = 0.00$ are measured for small times ($t \leq 100$) , since the slope of the lines in Figure 5 has meaning in this case as not all the samples are supposed to reach the edges of the lattice for large times.

The fractal dimensions, obtained in two different ways, in equation (4) and the box-counting method, are given in Tables 3–6. The fractal dimension for $r = 0.5$ and $\mu = 0.00$
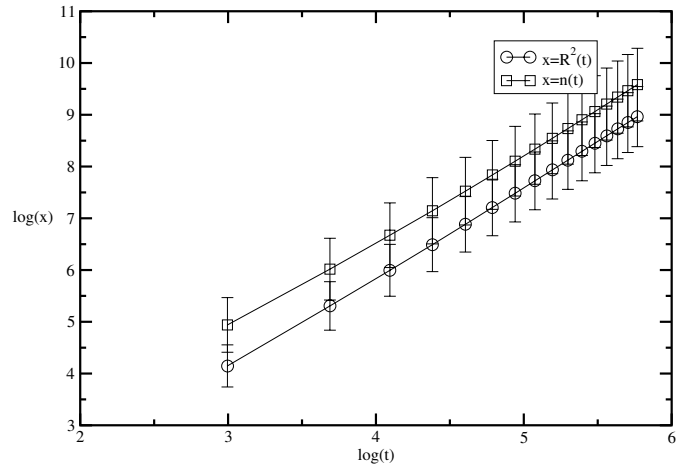
Fig. 5. Time dependence of square of radius of gyration and population, averaged over 1000 samples for $r = 0.5$ and $\mu = 0.00$.

Table 8. Percolation times for the model with decay *(WD)* for various fitness treshold and mutation rates.

r	μ	MCS Time
0.5	0.01	330.2 ± 3.1
0.5	0.5	326.4 ± 2.9
0.5625	0.01	465.7 ± 7.8
0.5625	0.5	445.7 ± 6.5

is 1.88 ± 0.09 by the first method and is 1.87 ± 0.05 by the second method which are close to the invasion percolation fractal dimension 1.89. The rest have the fractal dimension equal to 2 and these are compact or closed packed sets [6].

The Tables 7 and 8 give the percolation times (time to reach the edge of the lattice) for WOD and WD cases. Note that the percolation time for WD case is larger than the time for WOD case for the same fitness threshold and mutation rates since decay is not a desired trait for percolation.

6 Discussion

The growing patterns are observed to be compact except for the $r = 0.5$ and $\mu = 0.00$ case, which serves as the critical point beyond which percolation is impossible when there is no mutation. The pattern yields a fractal with a fractal dimension of the invasion percolation for this case. The most interesting feature in WD case is that, the fitness threshold value $r = 0.5625$, which is larger than the value $r = 0.5$ at the critical point allows percolation in the presence of mutations, though the growing pattern is not a fractal but compact. So, even a small amount of mutation may play a vital role for the invasion of the species at larger values of fitness threshold for which percolation would not be possible at the absence of it.

The amount of error made during the calculation of fractal dimension by the exponential scaling method is larger than the calculation by the box-counting method, which proves to be more trustworthy. On the other hand, the errors in the MCS is large only for the critical point case, where the fluctuations are large in the time calculation.

References

1. J. Maynard Smith, *Evolutionary Genetics* (Oxford University Press, Oxford, 1999)
2. R.M. Zorzenon dos Santos, S. Couthinho, Phys. Rev. Lett. **87**, 168102 (2001)
3. I. Jensen, H.C. Fogeby, R. Dickman, Phys. Rev. A **41**, 3411 (1990)
4. C.F. Ferrara, J.F. Fontanari, Phys. Rev. E **65**, 021902-1 (2002)
5. T. Vicsek, Fractal Growth Phenomena (Utopia Press, Singapore, 1989)
6. R.N. Onody, R.A. Zara, Phys. Rev. E **56**, 2548 (1997)
7. R.A. Zara, R.N. Onody, Int. J. Mod. Phys. C **10**, 229 (1999)
8. E. Brunet, B. Derrida, Phys. Rev. E **56**, 2597 (1997)
9. E. Moro, Phys. Rev. E **69**, 060101(R) (2004)