UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR/1003 - 6

SUMMER COLLEGE IN CONDENSED MATTER ON
" STATISTICAL PHYSICS OF FRUSTRATED SYSTEMS "

(28 July - 15 August 1997)

" The Darwinian paradigm "
" Finite populations "
and
" Coevolution"

presented by:

L. PELITI
Università di Napoli "Federico II"
Dipartimento di Scienze Fisiche
Mostra d'Oltremare, Pad. 19
80125 Napoli
Italy

---

1. The Darwinian Paradigm

Part 1

Introduction to the Statistical Theory of Darwinian Evolution

Thanks to:
Paul G. Higgs (Manchester)

1. The Darwinian Paradigh
   - Fitness landscapes
   - The Quasispecies equation
   - The Error Threshold
2. Finite Populations
   - Adaptative walks
   - Muller's ratchet
   - Complex landscapes: the infernal cycle
3. Recombination
   - Quasispecies models with recombination
   - Species formation models
4. Coevolution
   - Host-Parasite model
   - The NKC model
   - SOC models
5. Extinction
   - The debate over extinction rates
   - Sibani-Alstrøm model
   - Newman model

Host - Parasite Interaction

III 2

J. Maynard Smith
Evolutionary Genetics

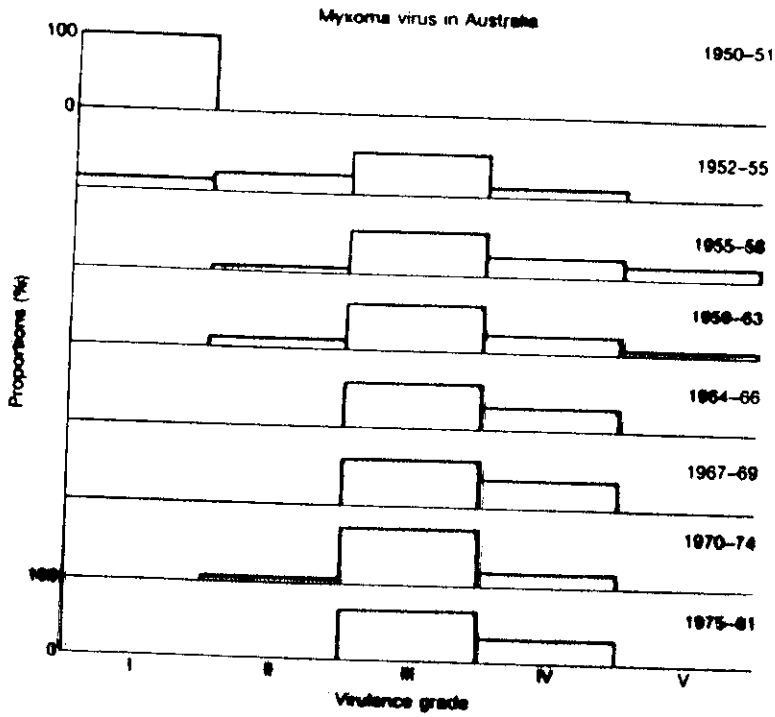


FIG. 9.8. Proportions of the various grades of myxoma virus in wild populations of rabbits in Australia, from 1950–81 (data from Fenner 1983, quoted by May and Anderson 1983).

Prey - Predator Interaction

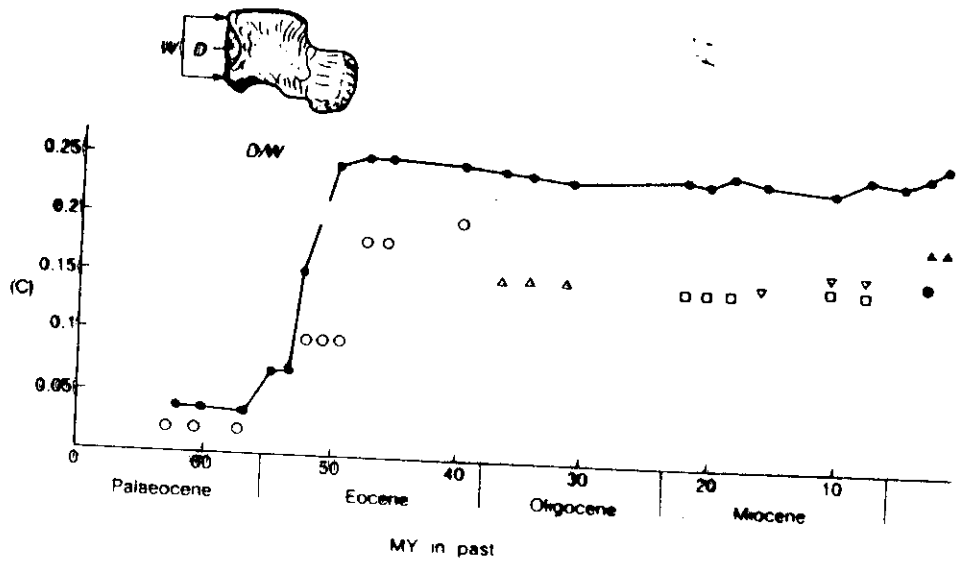


FIG. 14.12. Coevolution of carnivores and ungulates. The index *D/W* measures the depth of the groove in the astragalus, which indicates the degree to which the limb was constrained to move in a single plane. The full line represents a series of ungulates: they are not a phylogenetic series. The single points are carnivores: ○, mesonychids; △, hyaenodontids; □, amphicyonids; ▽, borophagines; ●, hyaenids; and ▲, canines. (After Bakker, in Futuyma and Slatkin 1983.)

3. COEVOLUTION

COEVOLUTION: Effects of mutual interaction among different species under evolutionary dynamics

INTERACTIONS:

1. Competition: The presence of each species inhibits the population growth of the other.
2. Exploitation: The presence of species A stimulates the growth of species B, and the presence of species B inhibits the growth of A.
3. Mutualism: The presence of each species stimulates the growth of the other.

SCOPE:

In the small: Coevolution of few-species systems (e.g. host-parasite)

In the large: Evolution of ecosystems

Part 3

3. Coevolution

SUMMARY

- Finite populations ⇒ disordered system
- "Drift", stochastic escape, Muller's ratchet
- Selective pressure depends on population site (peripatric speciation, Mayr)

Y. C. ZHANG, Phys. Rev. E55, R3817 (1997)

Q.S. equation with noise

$$\frac{dx_i}{dt} = \sum_j' M_{ij} x_j + [w_i - \langle w \rangle_t] x_i + \eta_i(t) \sqrt{x_i}$$

$\Rightarrow$ One has up to d_H mutants, where

$$d_H \sim \ln H / |\ln \mu|$$

Best fitness peak within the region

$$W_H \sim \sqrt{\ln M \ln N / |\ln \mu|}$$

$\sim$ adaptive walk

$$l \sim \ln(\ln M \ln N / |\ln \mu|)/(2 \ln 2)$$

There are $\sim \exp(d_H N)$ genotypes in the explored region

$\Rightarrow$ LOCAL OPTIMUM

ZHANG neglects stochastic escape.

$\Rightarrow$ Better optimum by noise-assisted process

Waiting time $t_d M \mu^d \sim 1$

$$t_d^{-1} \exp(t - t_d) w' \sim \exp(t w)$$

w': new "best" fitness; w: "old" best fitness

Shortest time

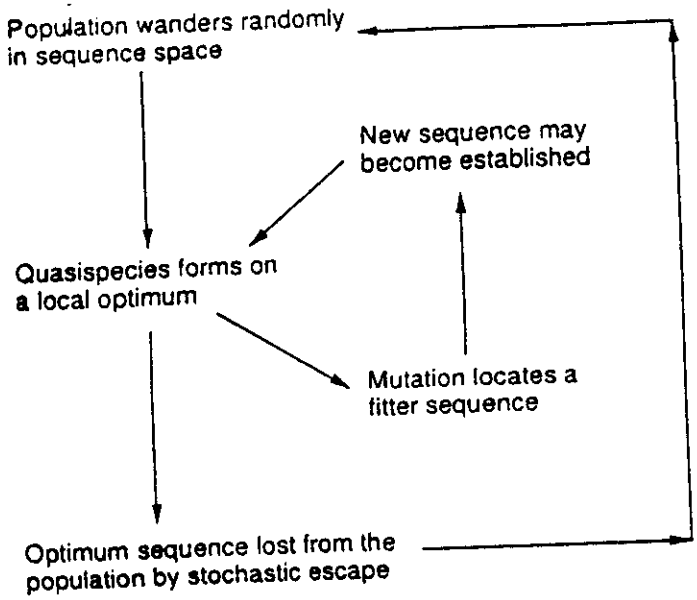
$$t \sim (1/M) \mu^{-\ln N / 4 |\ln \mu|^2}$$

$$w' \sim w + \frac{1}{2} \ln N / |\ln \mu|$$

Higgs & Woodcock

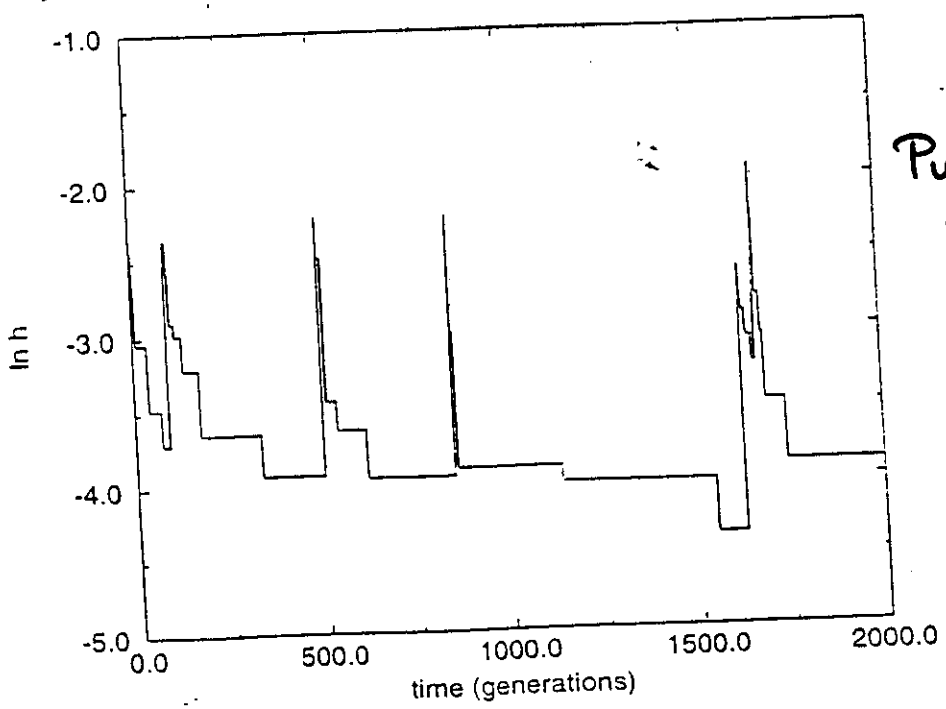
II 15

The Evolutionary Loop

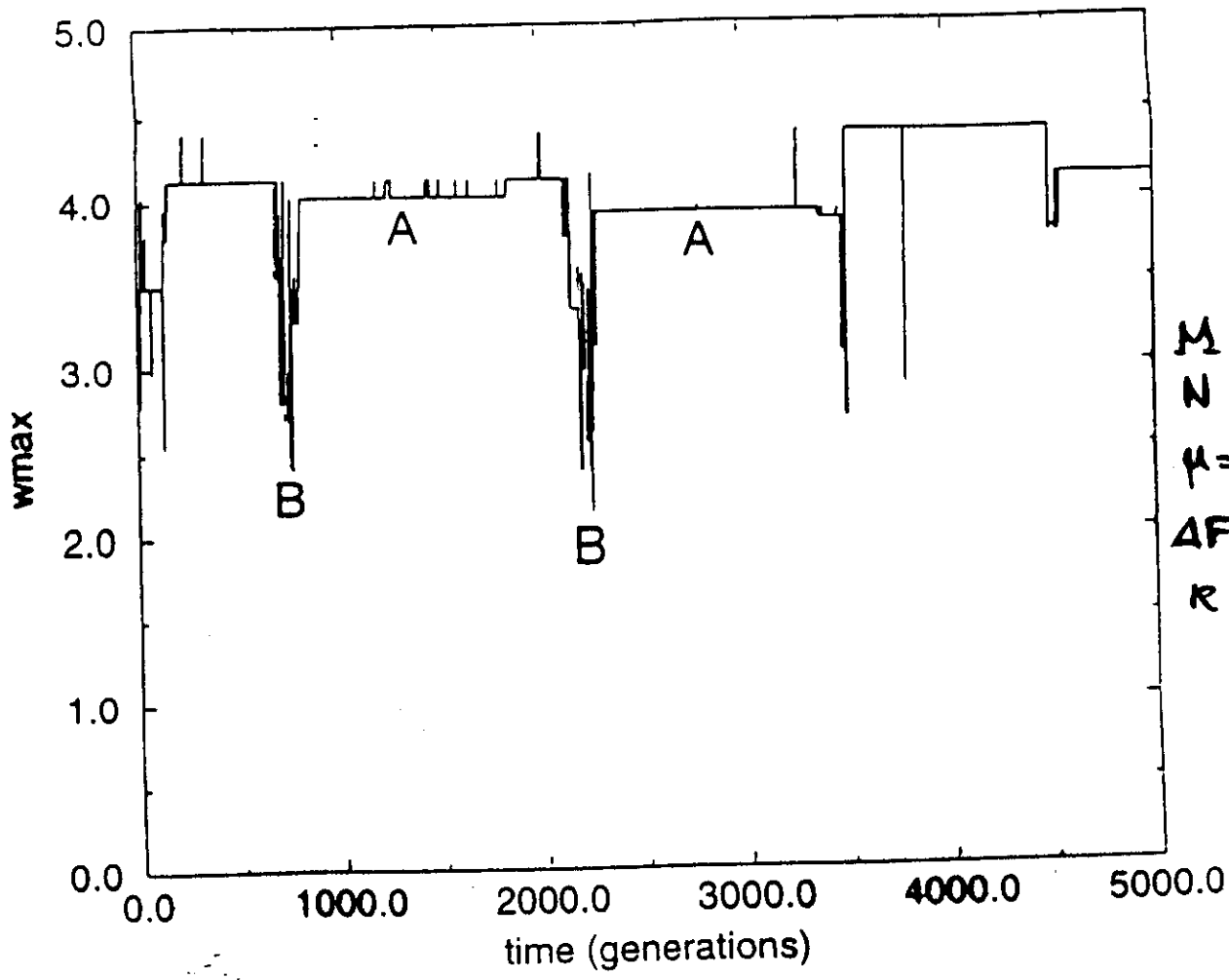


$$h_s = \int_{w_s}^{\infty} dw' \, \mathcal{F}(w') \; ; \; h = \max_s h_s$$

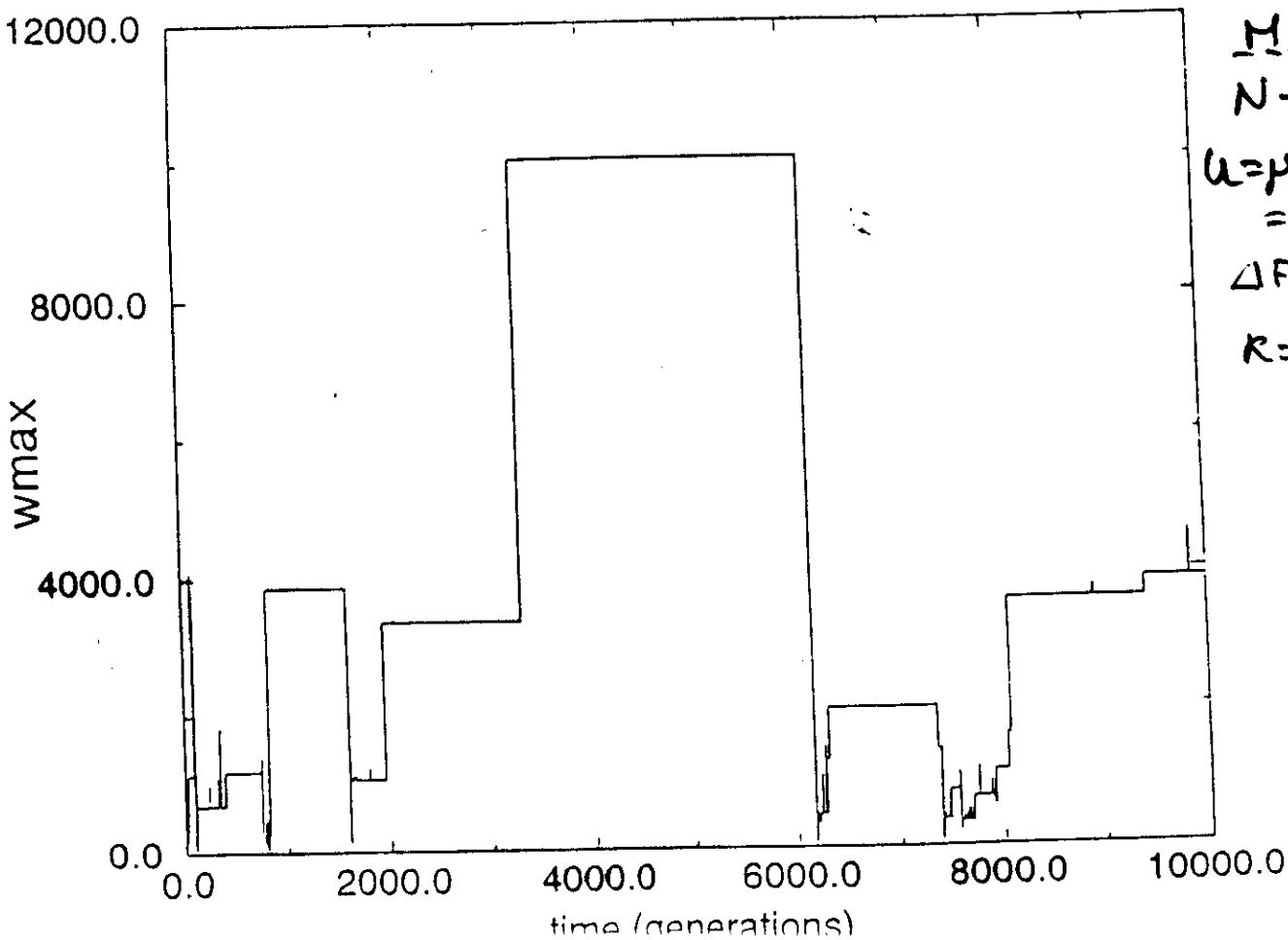
$M = 50, \; u = 2.1$



Punctuated equilibria



$M = 100$
$N = 15$
$\mu = 0.015$
$\Delta F^2 = N$
$R = 0.1$



$M = 10$
$N \to \infty$
$u = \mu N$
$= 2.0$
$\Delta F^2 = 1$
$R = 2.$

ADAPTIVE WALKS AND STOCHASTIC ESCAPE

Woodcock & Higgs, 1996

The Strong Selection Limit (only fittest reproduce)

- $N \to \infty$, mutation probability u
- Rugged landscape $p(w)$

  RARENESS: $h(w) = \int_w^\infty dw'\, p(w')$

$\Rightarrow$ Probability that child is better than parent

$$hu$$

$\Rightarrow$ Probability that at least ONE fitter genotype is hit

$$q(h) = 1 - (1-hu)^M$$

$\Rightarrow$ Probability of stochastic escape $\quad p = u^M$

TWO REGIMES:
- Adaptive walk $\quad q(h) < p$
- Stasis (quasispecies) $\quad p \sim q(h)$

Probability of escape after T generations

$$P_{esc}(T) = p(1-p)^{T-1}$$

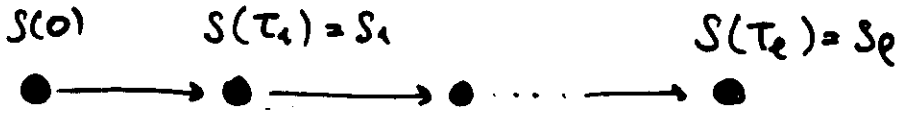
Distribution of rareness after T generations

$$P_{best}(h|T) = (1-h)^{MuT-1}\, h^{MuT}$$

$$\Rightarrow P_{best}(h) = \sum_{T=1}^{\infty} P_{best}(h|T)\, P_{esc}(T)$$

$$= \frac{p\, Mu\, (1-h)^{Mu-1}}{[1-(1-p)(1-h)^{Mu}]^2}$$

$$\Rightarrow \langle h \rangle_{best} = \int_0^1 P_{best}(h)\, h\, dh \sim \frac{u^{M-1}}{1-u^M} \ln(1/u)$$



$$\overline{1 - F(S_{k+1})} = \frac{1}{2}\left(1 - F(S_k)\right) \qquad (*)$$

$$\overline{1 - F(S_\ell)} \sim \frac{1}{N} \sim \frac{1}{2^\ell} \Rightarrow \bar{\ell} \simeq \frac{\ln N}{\ln 2}$$

Duration of walks

Because of (*), at each step the probability that the next step is the last doubles, and the probability per unit time that it will be taken is halved

Hence: the probability Q_t that the whole walk lasts t generations

$$Q_t \sim \frac{1}{\bar{t}} e^{-t/\bar{t}}$$

But:

$$Q_0 = \frac{1}{N} \Rightarrow \bar{t} = N$$

$$\Rightarrow Q_t \sim \frac{1}{N} e^{-t/N}$$

Exercise:

For the Fujiyama landscape $\bar{\ell} = N$

$$\bar{t} \simeq N \ln N$$

ADAPTIVE WALKS

Kauffman 1987

- Finite population
- Strong selection limit
- Small mutation rate

$\Rightarrow$ At any generation, the population is localized at one point in genotype space: $S = S(t)$

(a) Choose a new genotype S' such that

$$d_H(S(t), S') = 1 \quad \text{(one-mutant)};$$

(b) If $F(S') > F(S)$ then $S(t+1) = S'$

else $S(t+1) = S(t)$

$T = 0$ Monte-Carlo ∇_0

S.A. Kauffman and S. Levin, J. theor. Biol., 128, 11 (1987)

H. Flyvbjerg and B. Lautrup, Phys. Rev. A46, 6714 (1992)

REM model: $F(S)$ independent, identically distributed "RUGGED LANDSCAPE"

$p(F)$ does not matter $\Rightarrow$

F uniformly distributed between 0 and 1

II 10 bis

WOODCOCK & HIGGS

J. theor Biol. 1995

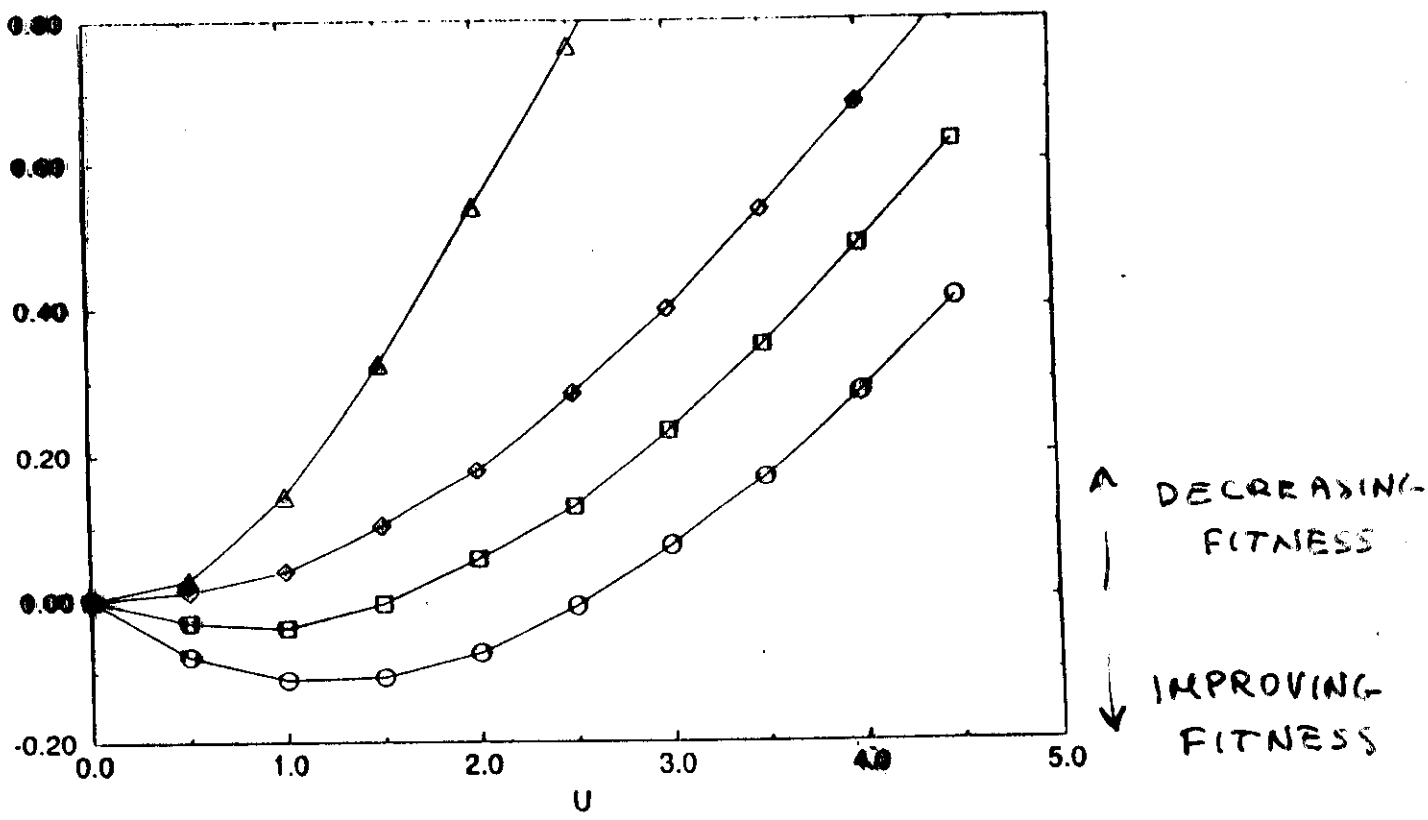
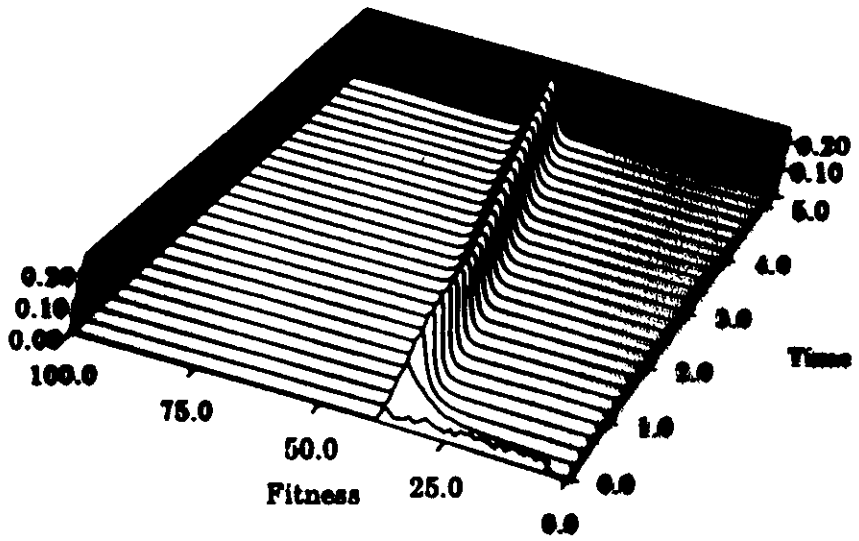


Figure 6. The rate of accumulation of mutations R shown as a function of U for several values the fraction p of favourable mutations. Positive R means decreasing fitness, and negative R means increasing fitness. R is always positive for $p < p_c$, whilst for $p > p_c$ R is negative at small U. These curves are simulation results with N = 100 and s = 0.2. The values of p are 0.0 (triangles), 0.1 (diamonds), 0.12 (squares) and 0.14 (circles). The critical value p_c lies between 0.1 and 0.12.

II 10



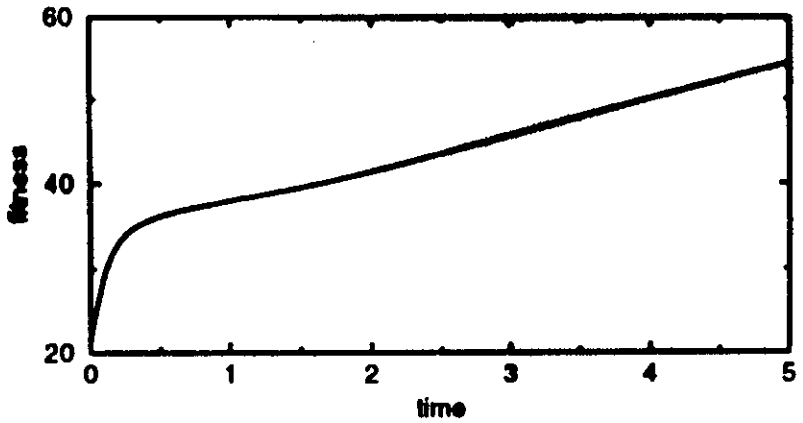


FIG. 2. Evolution of the viral colony concentration in fitness space starting from random distribution within a range $\{5.0 < r < 18.0\}$ in the framework of Eq.5. Diffusion constant is $D = 1$, $p_{tot} = 1$, and threshold concentration $p_c = 5 \cdot 10^{-4}$: a - space-time diagram $p(r,t)$; b - average fitness $\bar{r}$ versus time

$r \leftrightarrow w$

$$\frac{\partial p(r,t)}{\partial t} = \theta(p - p_c)(r - \bar{r})\, p(r,t) + D\frac{\partial^2 p(r,t)}{\partial r^2} \qquad (5)$$

L. Tsimring, H. Levine, D. A. Kessler, RNA virus evolution via a fitness-space model, Phys. Rev. Lett. 76, 4440 (1996)

BUT... fitness may increase

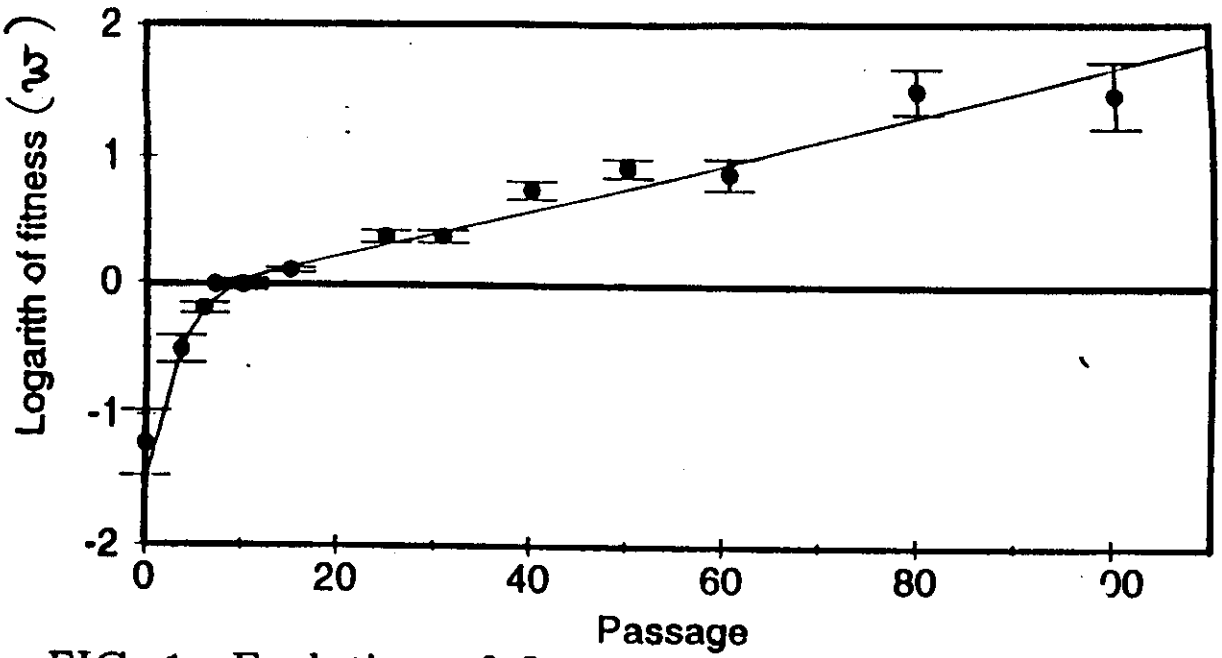


FIG. 1. Evolution of fitness of MARM clone during the transmission series of 80 passages on HeLA cells (Fig. 2b of [1] with permission from PNAS)

I.S. Novella, E.A. Duarte, S.F. Elena, A. Moya, E. Domingo, J.J. Holland, Proc. Natl. Acad. Sci. USA 92, 5841 (1995)

Assuming $u \gg 1$

$$p_k = 1 - \delta_k \; ; \qquad \delta_k = e^{-u} \sum_{m=0}^{k-1} \frac{u^m}{m!}$$

$$R = \sum_{k=1}^{\infty} (p_k)^M \simeq \sum_{k=1}^{r} (p_k)^M$$

N: $\delta_k < 1/M$ for $k < r$

Approximating the Poissonian with a Gaussian

$$\frac{1}{M} = \delta_r \simeq \int_{-\infty}^{r} \frac{du'}{\sqrt{2\pi u}} \exp\left(-\frac{(u'-u)^2}{2u}\right)$$

$$= \frac{1}{2} \operatorname{erfc}\left(\frac{(u-r)}{\sqrt{2u}}\right)$$

$$\Rightarrow \quad R \simeq u - c(M)\sqrt{2u}$$

$$\operatorname{erfc}(c(M)) = 2/M$$

Neutral case [KIMURA]

$$R = u$$

The strong selection limit

$$s \to 1$$

$\Rightarrow$ Only fittest individuals have offspring

k_{min} # mutations of fittest individual

The probability that an individual has no less than k mutations is given by

$$p_k = \sum_{m=k}^{\infty} f(m)$$

$f(m)$ probability that an individual has m mutations

e.g. $f(m) = \frac{e^{-U}}{m!} U^m \Rightarrow p_1 = 1 - e^{-U}$

The probability that ALL individuals have no less than k mutations is given by p_k^M

$\Rightarrow$ The probability that

$$k_{min}(t+1) - k_{min}(t) = k$$

is given by

$$p_k^M - p_{k+1}^M$$

Mutation rate:

$$R = \sum_{k=1}^{\infty} k \left(p_k^M - p_{k+1}^M \right) = \sum_{k=1}^{\infty} p_k^M$$

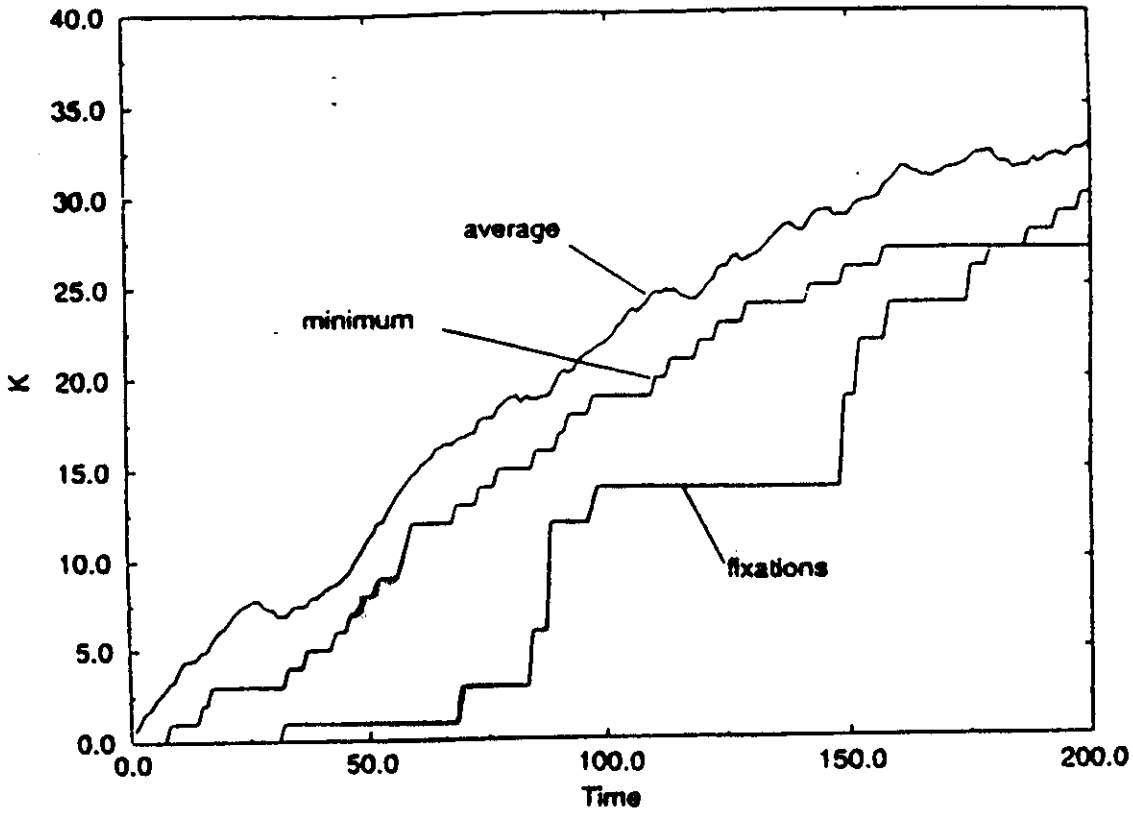
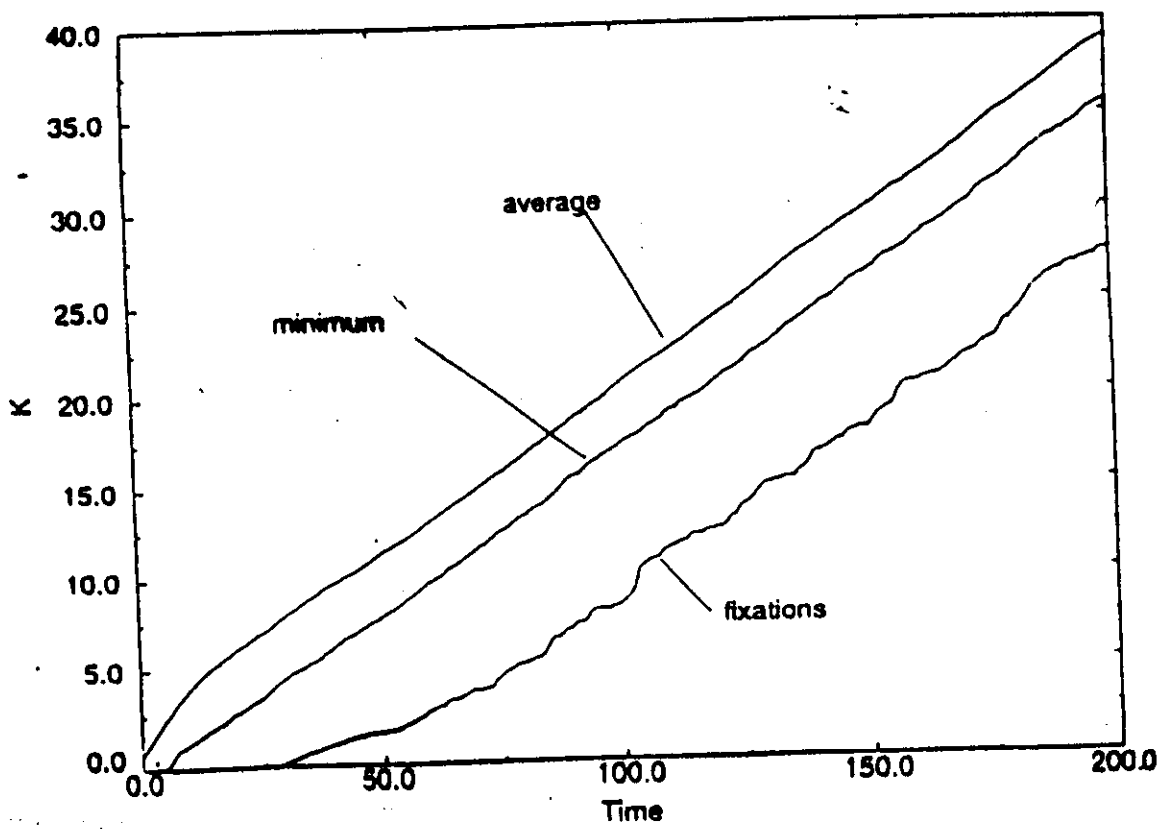


Fig. 2. The progress of the mean number of mutations per individual $\langle k \rangle$, the fittest surving class k_{min}, and the number of fixations k_{fix} as a function of time for a single population of $M = 100$, $U = 0.5$ and $s = 0.01$. Notice that k_{min} increases in small steps, usually only one notch of the ratchet at a time, whereas k_{fix} increases in large steps, indicating that there are large bursts of simultaneous fixations



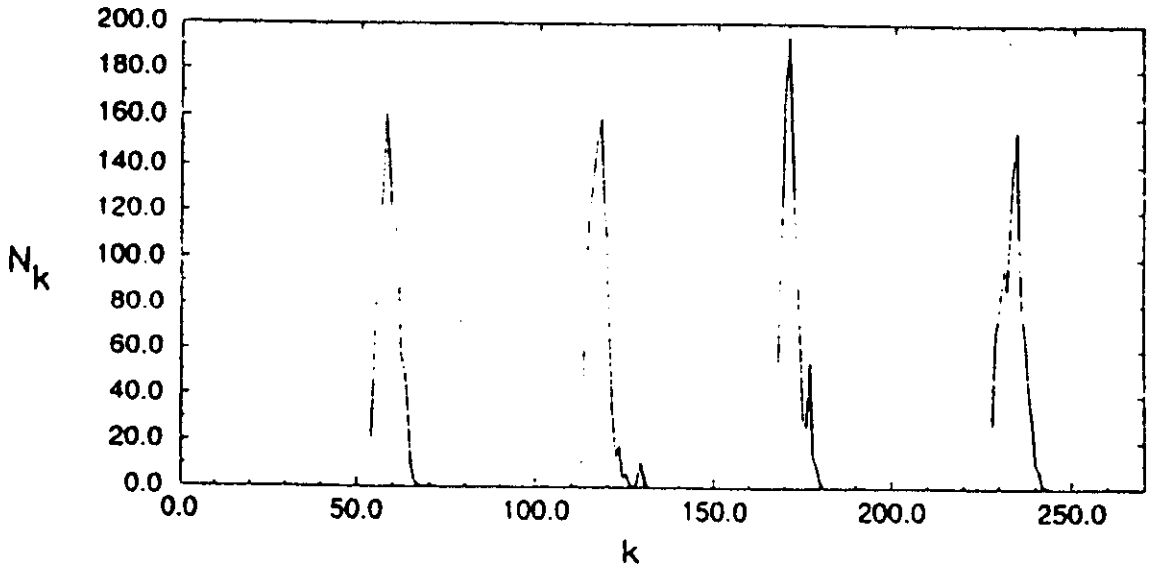


Fig. 1a. The distribution of the number of individuals N_k with k mutations at four points in time with $M = 1000$, $u = 0.1$ and $s = 0.01$

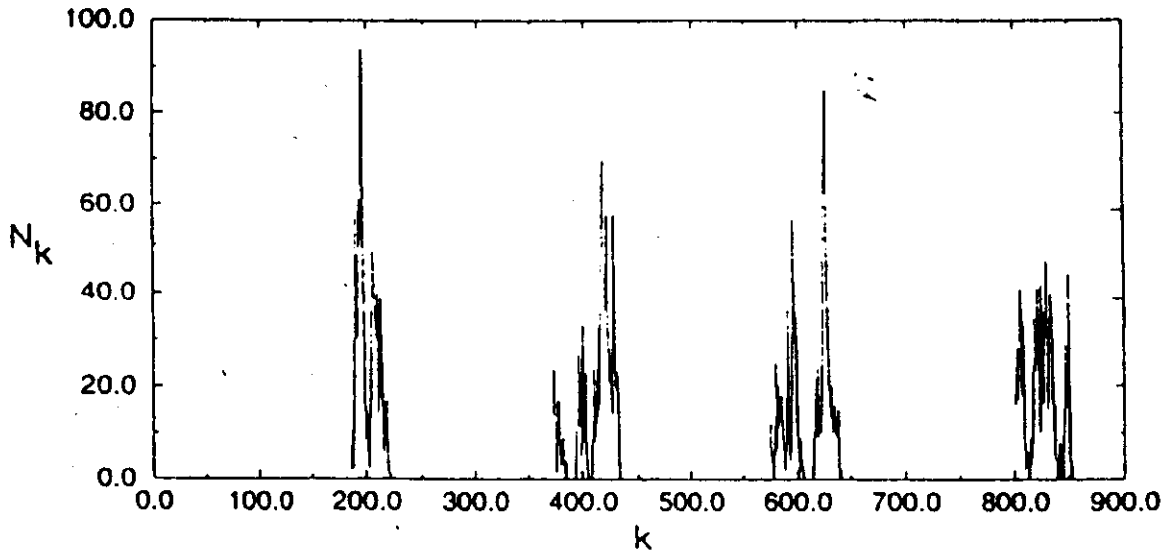


Fig. 1b. The same in the neutral case $s = 0$. Fluctuations are extremely large in the neutral case, since the distribution is non-self-averaging

- FINITE POPULATIONS $\Rightarrow$ DISORDERED SYSTEMS
  [Lack of self-averaging]

Two kinds of averages

- Population average $\langle \ldots \rangle = \frac{1}{M} \sum_{\alpha}' \ldots$
- Process average $\overline{\ldots}$

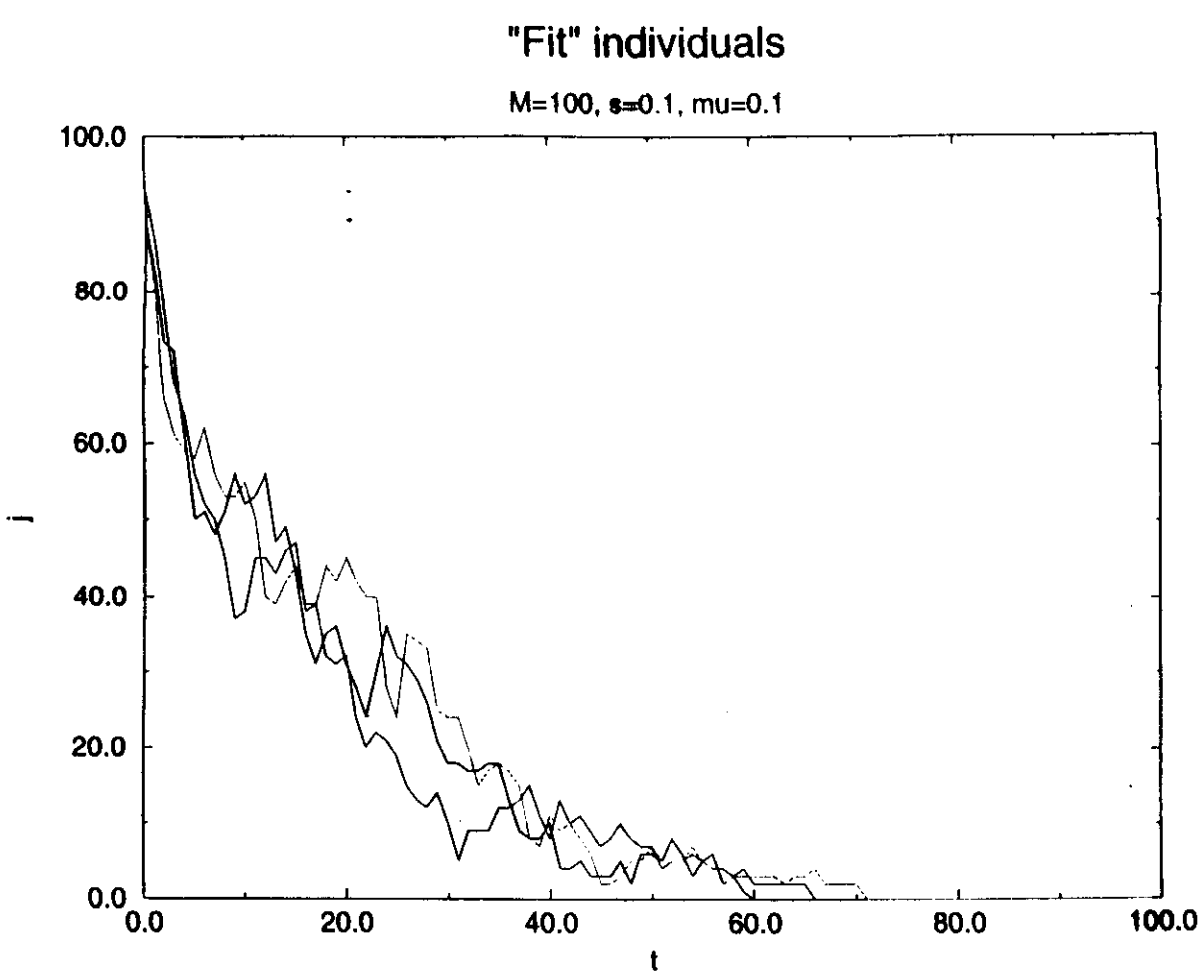
E.g. $\overline{j(t)} = \sum_{j=0}^{M} j \, W_{jk_0}^{t} \propto e^{-t/\tau}$

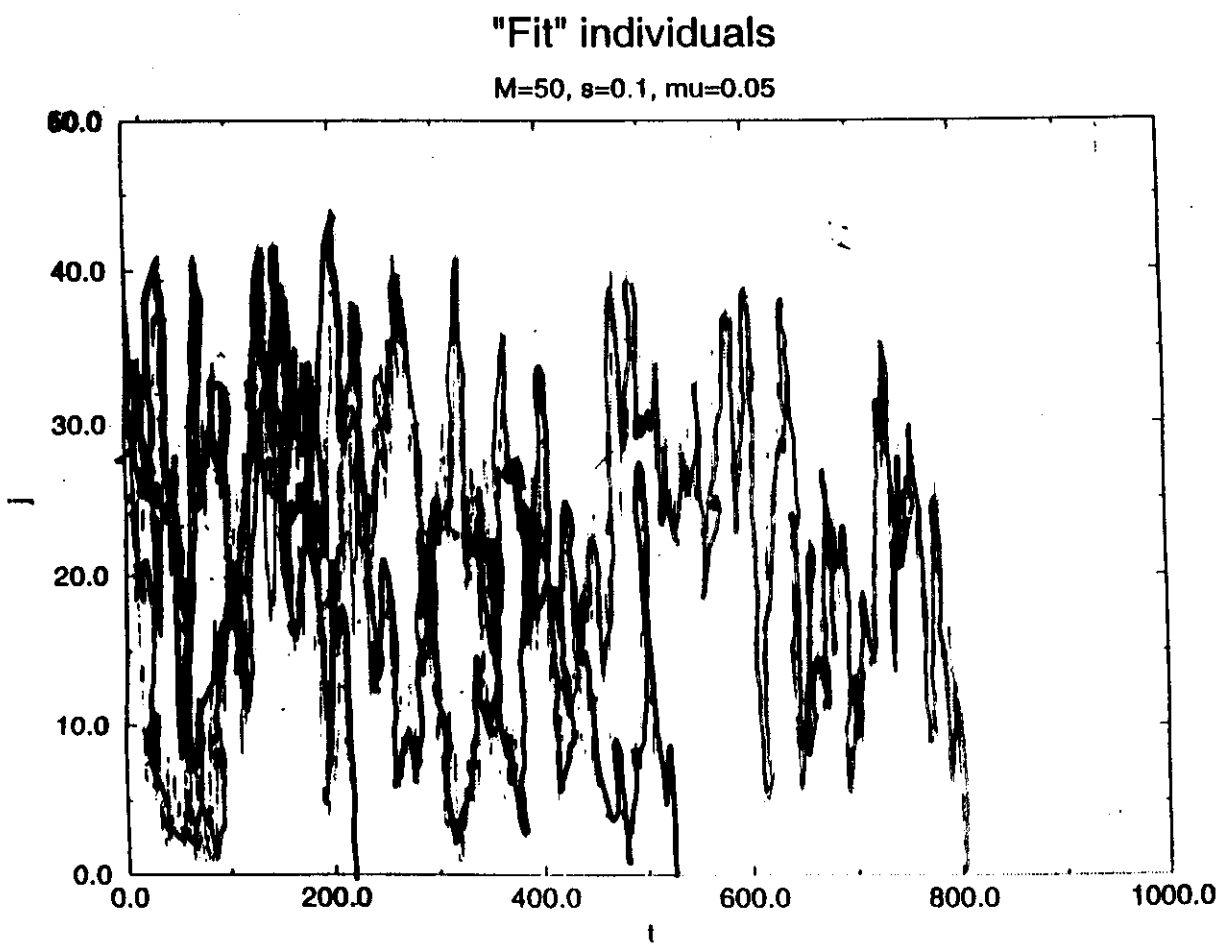
$\neq$ "Typical value" $\sim \begin{cases} M(1-\mu/s) \\ 0 \end{cases}$

- AN EXAMPLE: FUJIYAMA LANDSCAPE
  MULLER'S RATCHET
  G. WOODCOCK, P. G. HIGGS

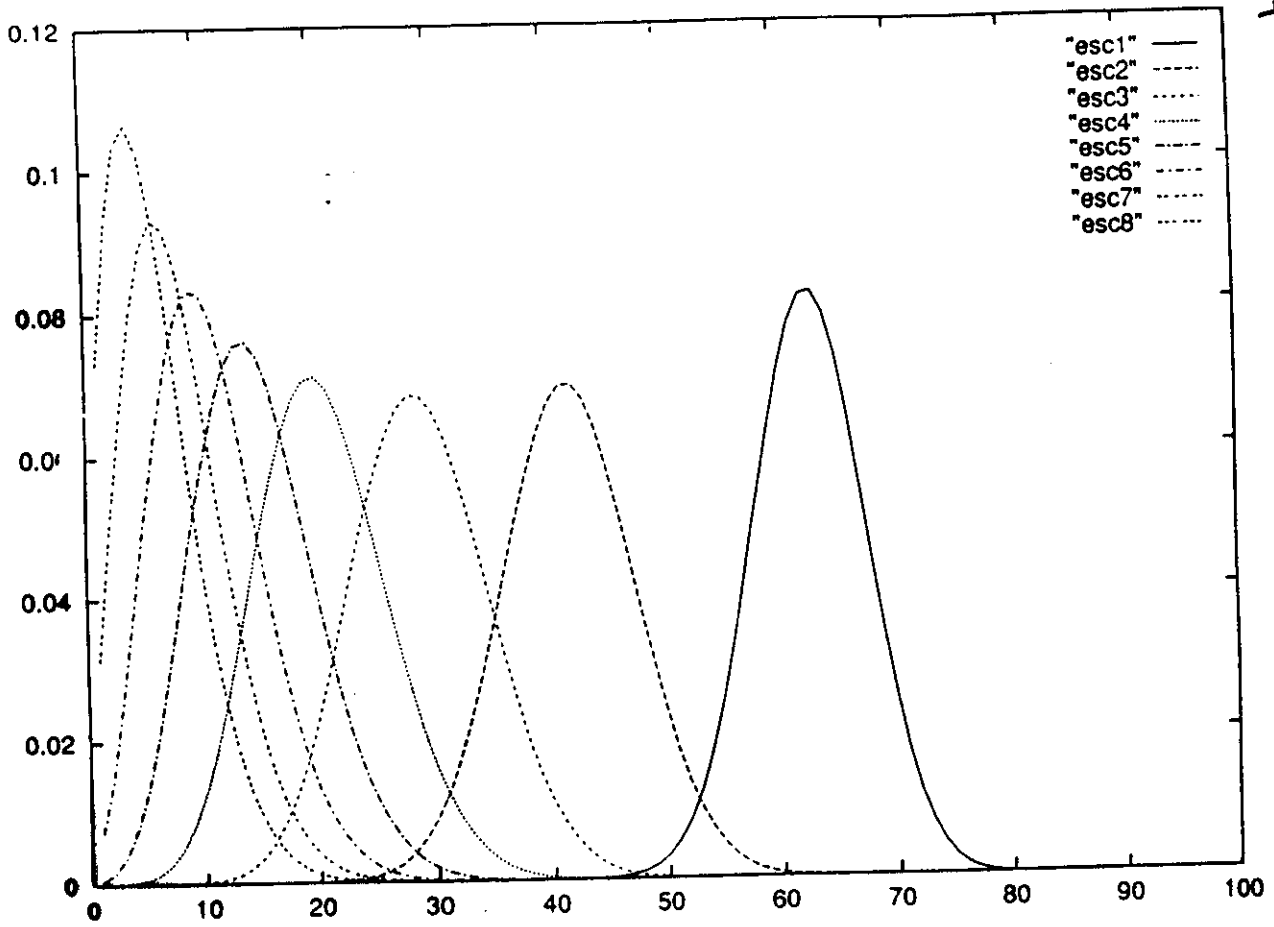
k: # mutations from optimal sequence

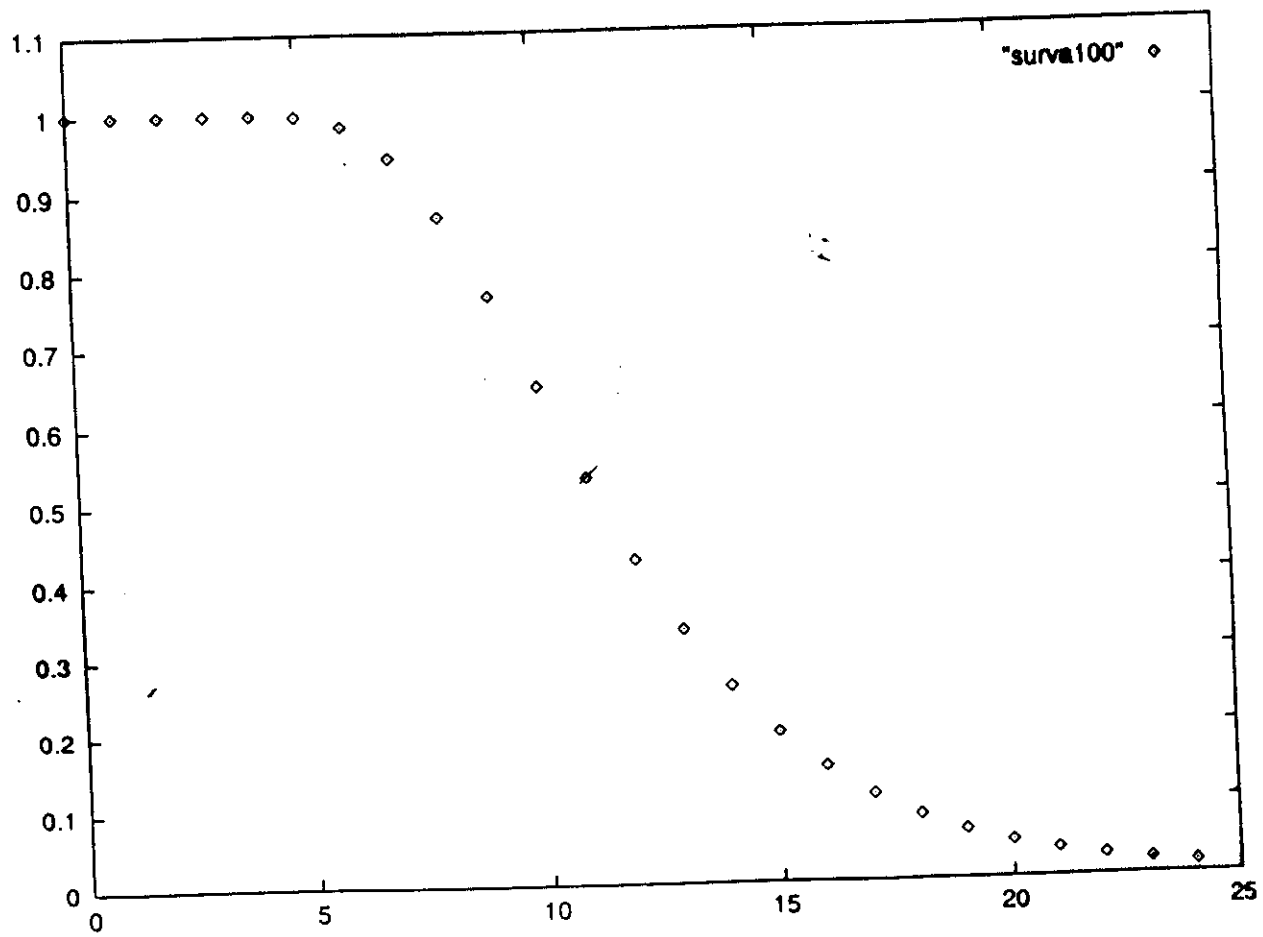
$w_k = (1-s)^k$

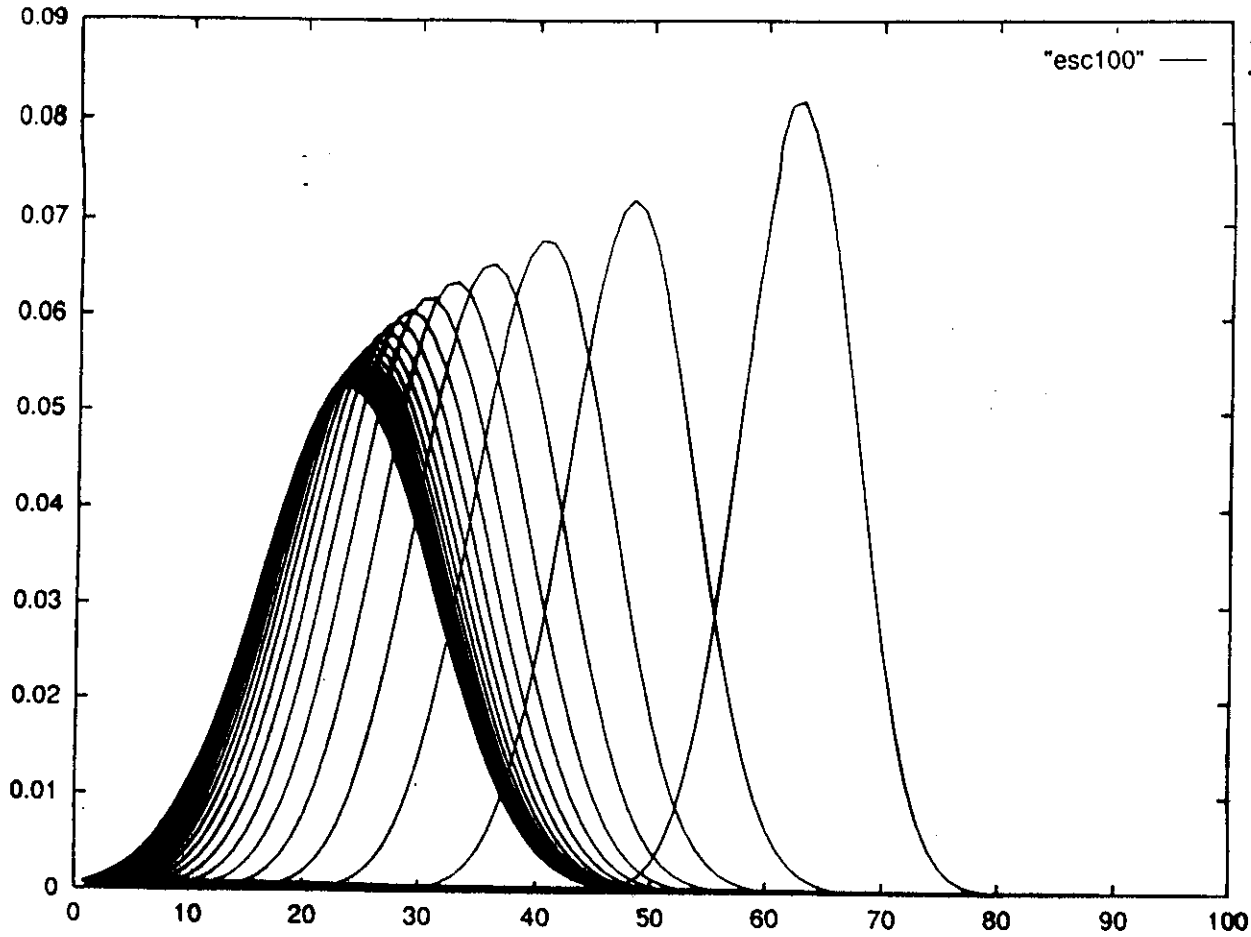

"Fit" individuals
M=100, s=0.1, mu=0.1
100.0
80.0
60.0
40.0
20.0
0.0
i
0.0
20.0
40.0
60.0
80.0
100.0
t

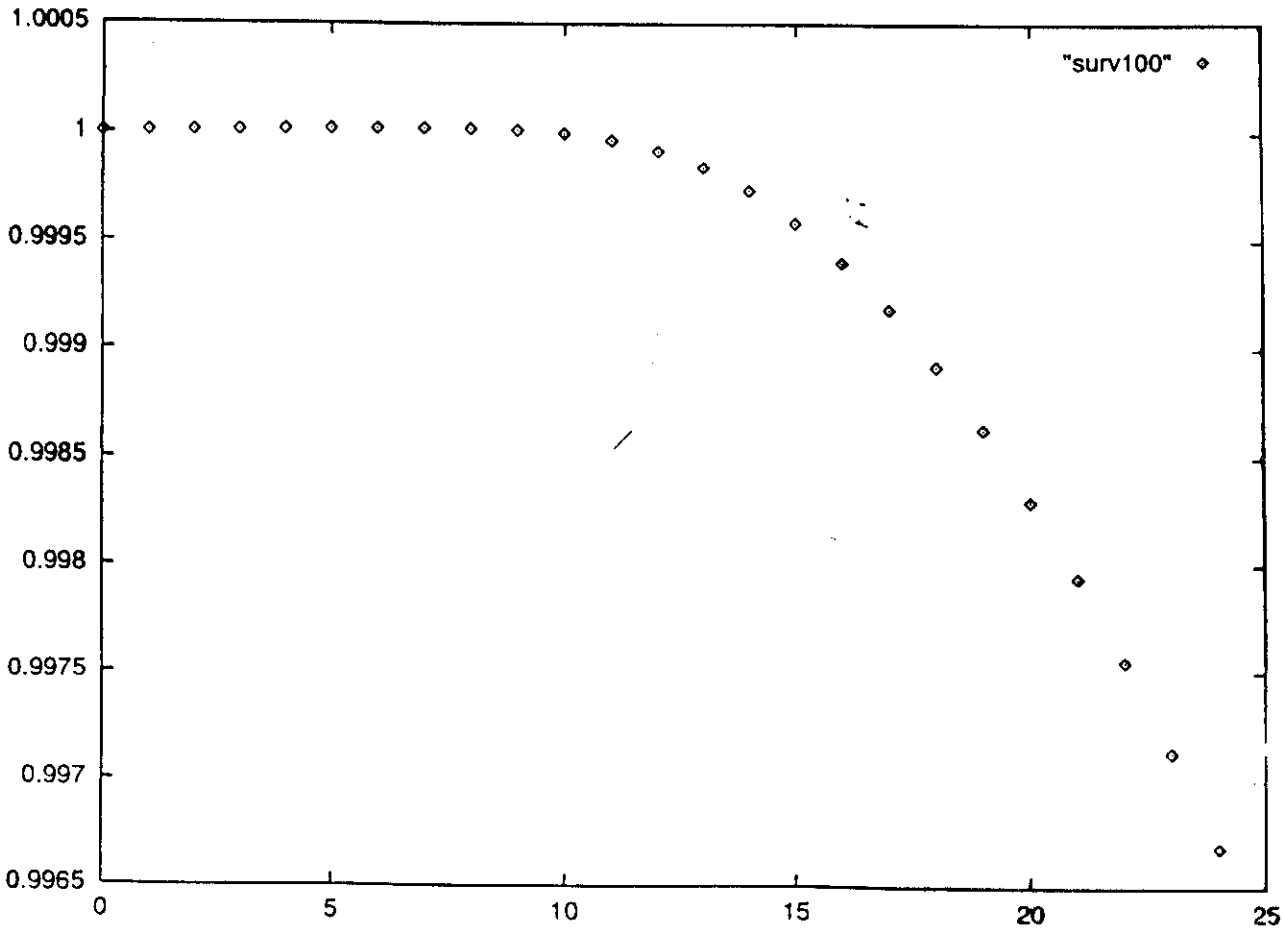

"Fit" individuals
M=50, s=0.1, mu=0.05
60.0
40.0
30.0
20.0
10.0
0.0
i
0.0
200.0
400.0
600.0
800.0
1000.0
t

II 26bis


"esc1"
"esc2"
"esc3"
"esc4"
"esc5"
"esc6"
"esc7"
"esc8"
0.12
0.1
0.08
0.04
0.02
0
0
10
20
30
40
50
60
70
80
90
100


"surva100"
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25


"esc100"
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50
60
70
80
90
100


"surv100"
1.0005
1
0.9995
0.999
0.9985
0.998
0.9975
0.997
0.9965
0
5
10
15
20
25

2. FINITE POPULATIONS

- Sharp peak landscape, infinite genome limit
  Higgs & Woodcock

$P_k(t)$: probability that the number n_0 of individuals with preferred genotypes equals k ("fit")

$$P_j(t+1) = \sum_{k=0}^{M} W_{jk} P_k(t)$$

$$W_{jk} = \binom{M}{j} \xi_k^j (1-\xi_k)^{M-j}$$

$$\xi_k = \frac{(1-\mu)\cdot k}{k + (1-s)(N-k)}$$

Probability of choosing a "fit" individual, and that it does not mutate

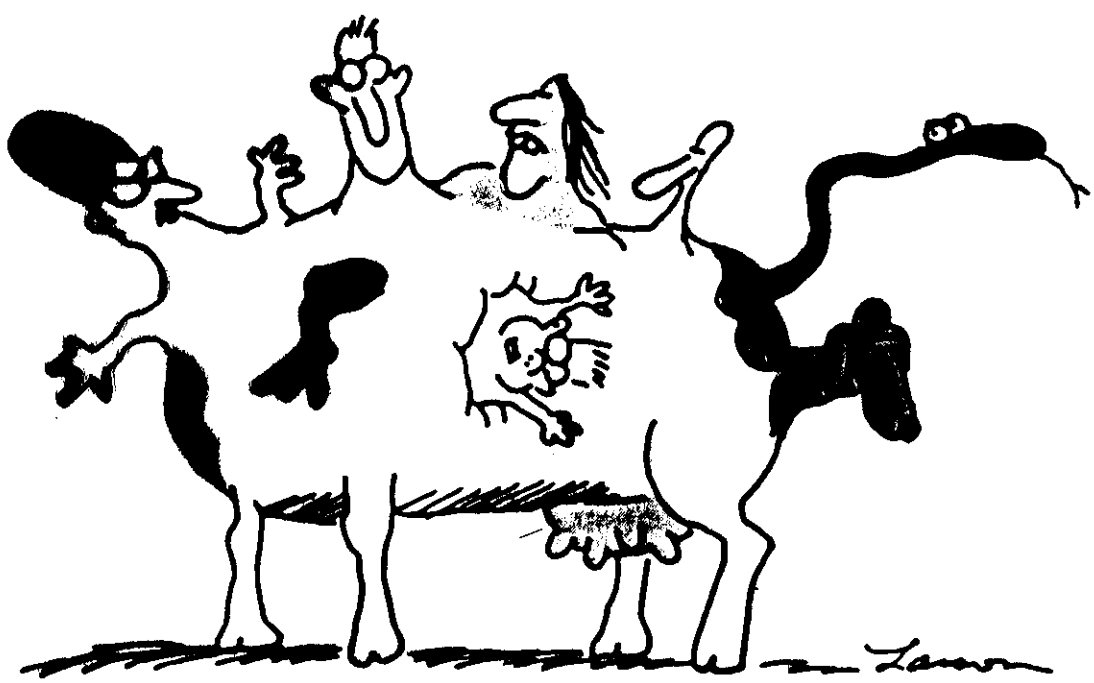
$\lim_{t\to\infty} P_j(t) = \delta_{j0}$ with probability 1

Survival probability $\nu(t) = \sum_{j=0}^{i} P_j(t)$

$$\nu(t) \propto e^{-t/\tau}, \quad \tau \propto \exp(\alpha M)$$

$\Rightarrow$ Stochastic escape

Part 2



2. Finite Populations

References

J. Maynard Smith Evolutionary Genetics
Oxford U.P. (1989)
an introduction

M. Eigen, J. McCaskill, P. Schuster, The Molecular Quasispecies, Adv. Chem. Phys. 75, 149 (1989)

R Tarazona, Error threshold as phase transitions, Phys. Rev. A45 6038 (1992)

E. Baake, M. Baake, H. Wagner, The Ising quantum chain and the quasispecies theory Phys. Rev. Lett. 78, 559 (1997)

P.G. Higgs, Error threshold and stationary mutant distribution, Gen. Res. (Cambridge) 63, 63 (1994)

S. A. Kauffman, The Origins of Order
Oxford V. P. (1993)

S. Franz, L. Peliti, Error threshold in simple landscapes, J. Phys. A: in press

D. Alves, J.F. Fontanari, Population genetics approach to the quasispecies model, Phys. Rev. E54, 4048 (1996)

CONCLUSIONS

- ERROR THRESHOLD
- QUASISPECIES THEORY INCONSISTENT IN THE FREE PHASE
- FINITE POPULATION SIZE?

IN A RUGGED FITNESS LANDSCAPE

S. Franz, L.P., M. Sellitto

$$\overline{F(S)}^2 = \frac{N}{2}$$

$\mathcal{N}(E,q)$: # of sequences with overlap q from a given one, with $F(S) = E$

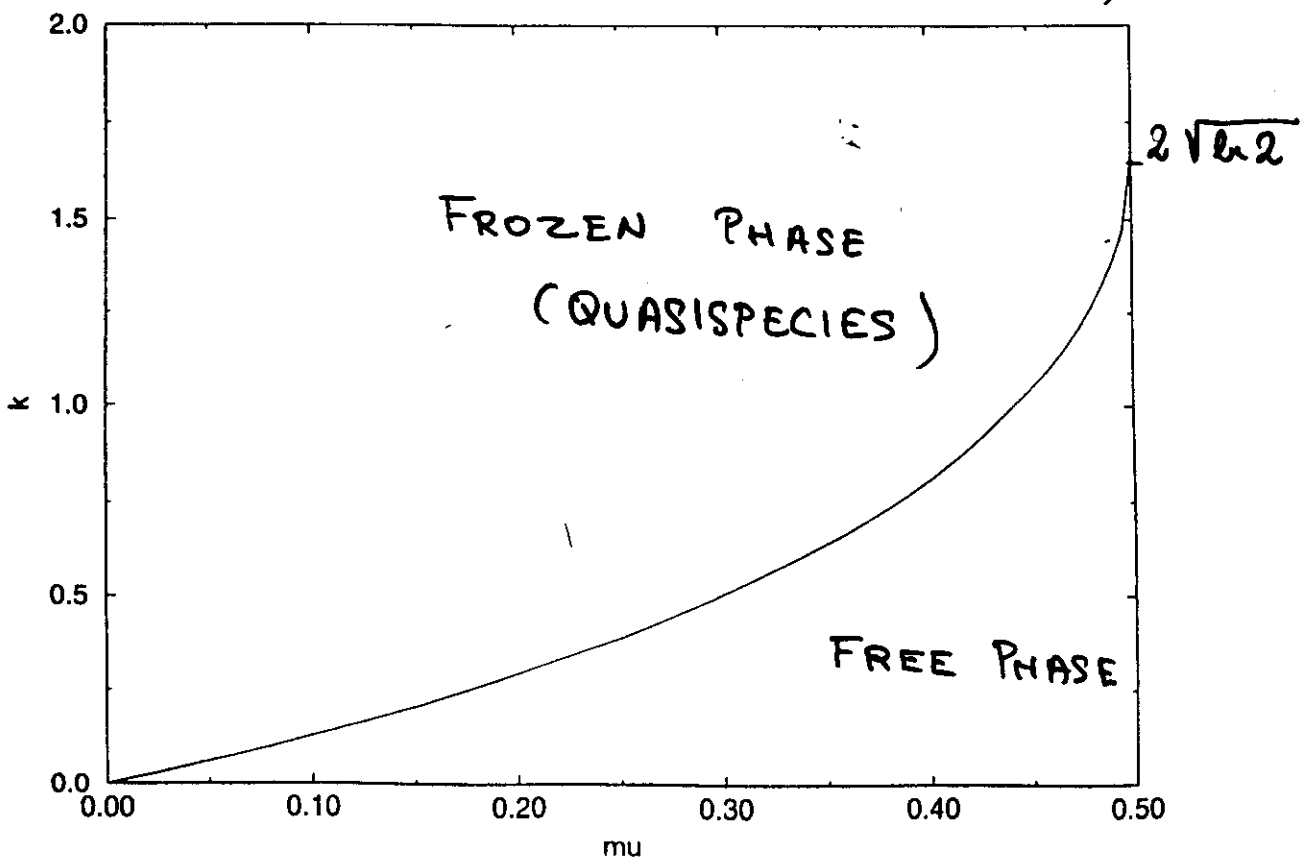
$$\sim \exp\left(N\mathcal{S}(q) - E^2/N\right)$$

$$\mathcal{S}(q) = \ln 2 - \frac{1}{2}\left[(1+q)\ln(1+q) + (1-q)\ln(1-q)\right]$$

$$W = \int_{(N\mathcal{S}(q) - E^2/N) > 0} dE\, dq\, \mathcal{N}(E,q) \exp(kE) \left(\frac{1-\mu}{\mu}\right)^{Nq/2}$$

Transition $\mathcal{N}(E,0) \sim O(1)$

$$k_c = 2\left(\sqrt{\ln 2} - \sqrt{\ln 2 + \ln(1-\mu)}\right)$$



The rugged fitness landscape

- Overlap:

$$q(S,S') = \frac{1}{N}\sum_{i=1}^{N} S_i S'_i = 1 - \frac{2}{N} d_H(S,S')$$

- Reproductive success = complex optimization
  ⇒ Many optima?

- Fitness landscape modelled by RANDOM FUNCTION (Anderson 1983)

$$F(S) = \sum_{\{i_1,\dots,i_p\}} J_{\{i_1,\dots,i_p\}} S_{i_1} \cdots S_{i_p}$$

$J_{\{i_1,\dots,i_p\}}$: independent, identically distributed, random coeff.
(Kauffman 1987, NK model)

$$\overline{F(S)F(S')} = \sum_{\{i_1,\dots,i_p\}} \Delta J^2 (S_{i_1} S'_{i_1}) \cdots (S_{i_p} S'_{i_p})$$

$$\simeq \frac{N^p}{p!} \Delta J^2 \, q^p$$

Fujiyama ⟨≈⟩ $p=1$

$p \to \infty$ REM (Derrida 1981)

$$\overline{F(S)F(S')} \propto \delta_{SS'}$$

$$x_0^* = \frac{1}{W^*} e^{-u} x_0^* \quad \Rightarrow \quad W^* = e^{-u}$$

$$x_1^* = \frac{1}{W^*} \left[u e^{-u} x_0^* + e^{-u}(1-s) x_1^* \right]$$

$$= u x_0^* + (1-s) x_1^*$$

To first order in u, s $\qquad x_1^* = \frac{u}{s} x_0^*$

$$x_2^* = \frac{u^2}{2} x_0^* + u(1-s) x_1^* + (1-s)^2 x_2^*$$

$$\Rightarrow \quad 2s\, x_2^* \simeq u\, x_1^* \qquad x_2^* = \frac{1}{2}\left(\frac{u}{s}\right)^2 x_0^*$$

$$\cdots \quad x_k^* = \frac{1}{k!}\left(\frac{u}{s}\right)^k x_0^*$$

$$\Rightarrow \quad x_k^* = e^{-u/s} \frac{(u/s)^k}{k!}$$

HAIGH, HIGGS

No error threshold

Exercise:

$$w_j \propto e^{-s j^\alpha}$$

$$\Rightarrow \quad x_k^* = \frac{(u/s)^k}{(k!)^\alpha} x_0^*$$

The Fujiyama landscape (SMOOTH LANDSCAPE)

$$w = e^{-\kappa F(s)}$$

$$F(s) = \sum_i s_i s_i^\circ$$

"Multiplicative fitness landscape"
"No epistatic interaction"

$$\Rightarrow \quad w_j \propto (1-s)^j$$

j: # mutations

$$1 - s = e^{-\kappa}$$

Quasispecies equations:

$$x_j(t+1) = \frac{1}{W(t)} \sum_k M_{jk} w_k x_k(t)$$

$$W(t) = \sum_j w_j x_j(t)$$

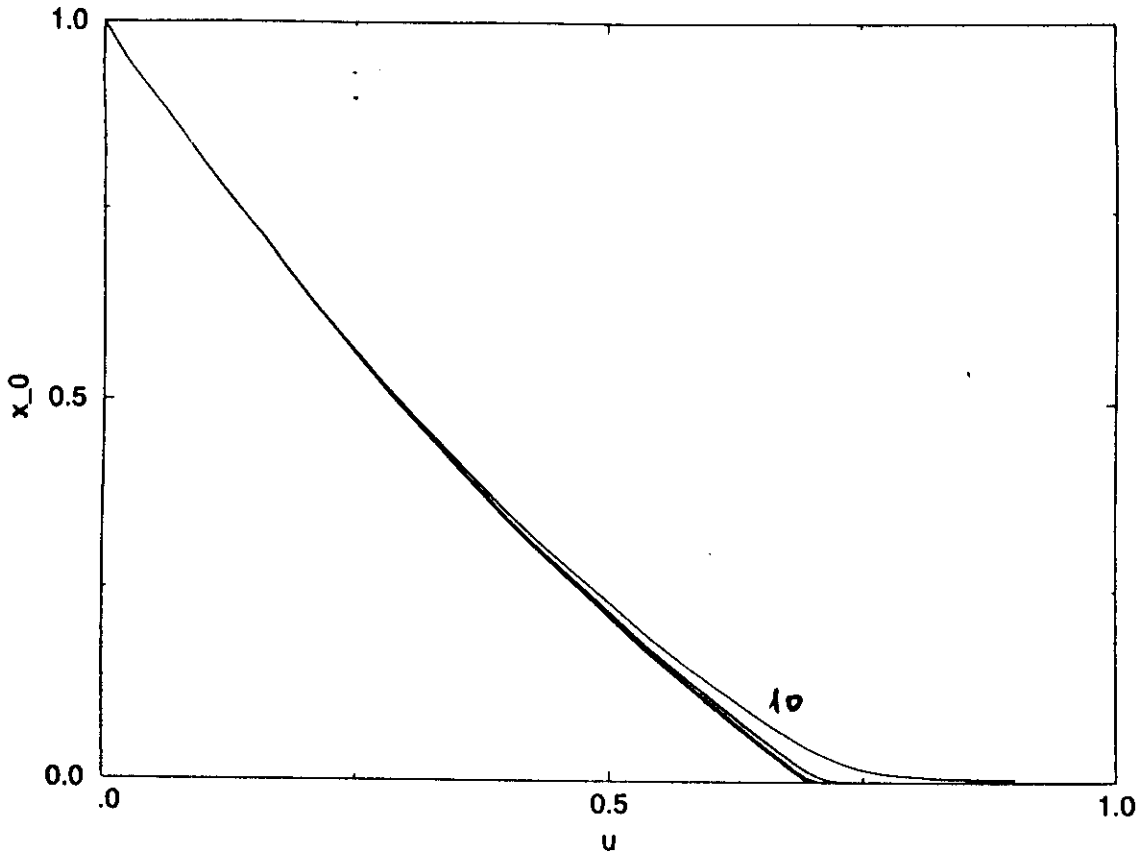
M_{jk}: mutation matrix

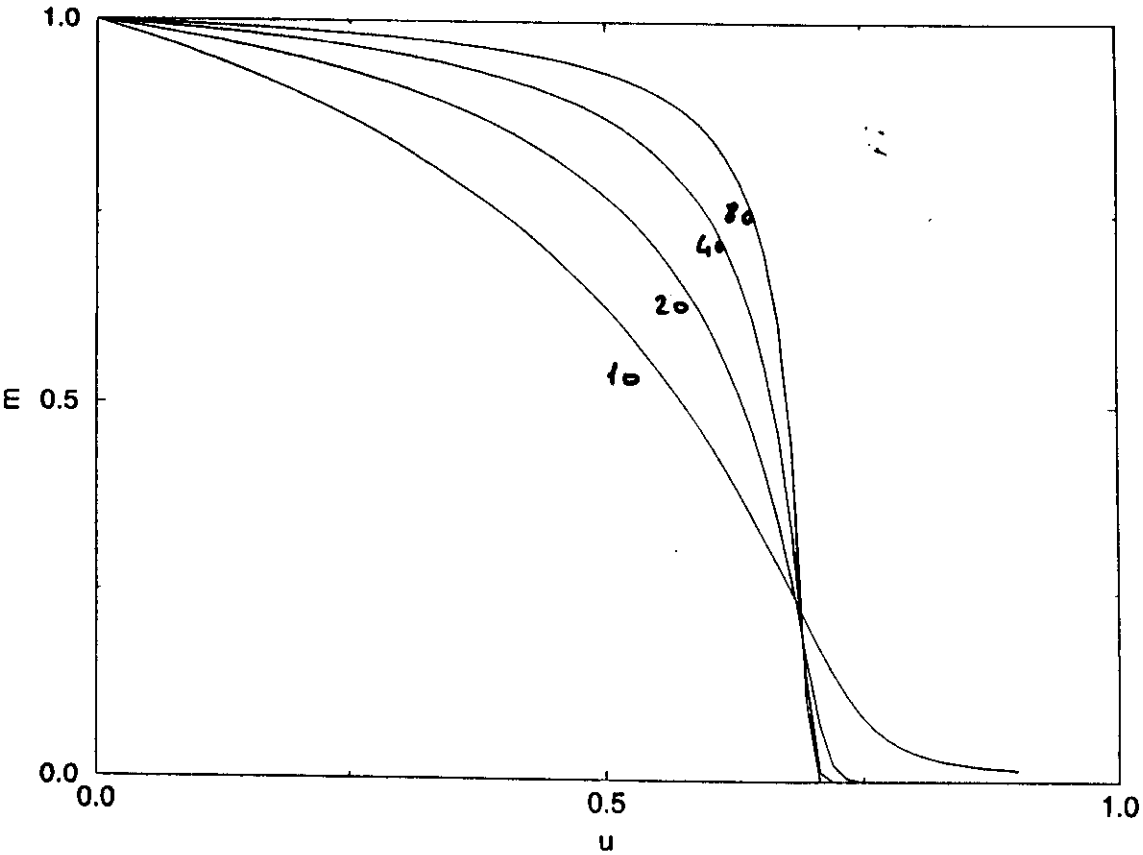
$$M_{jk} = \binom{N-k}{j-k} \mu^{j-k} (1-\mu)^{N-j} \qquad (k \le j \le N)$$

neglecting back mutations

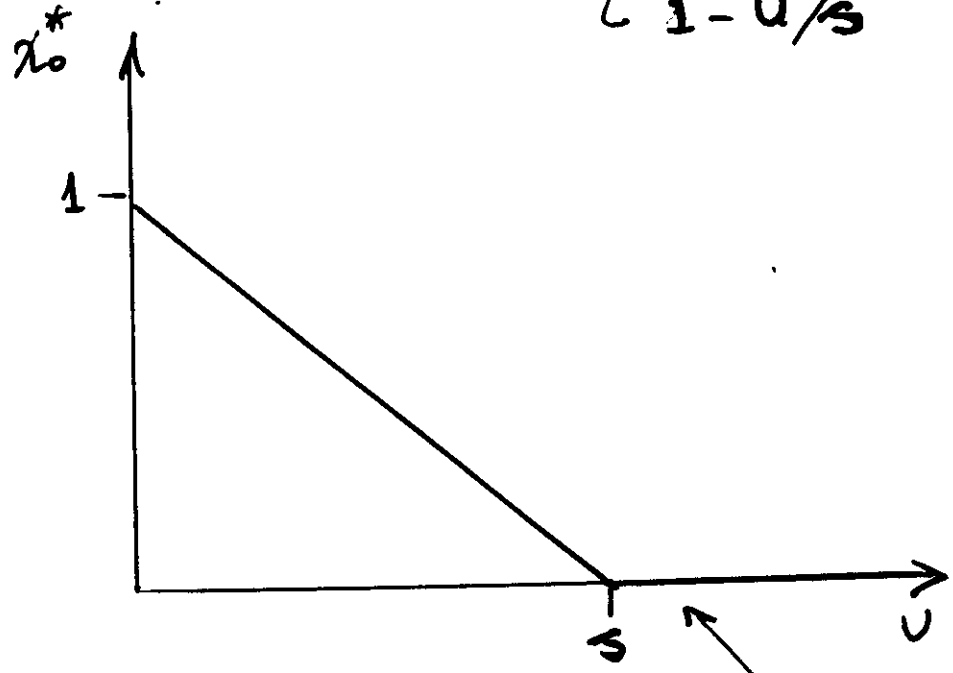
$N \to \infty$, $\mu N = u$

$$M_{jk} \longrightarrow e^{-u} \frac{u^{j-k}}{(j-k)!}$$

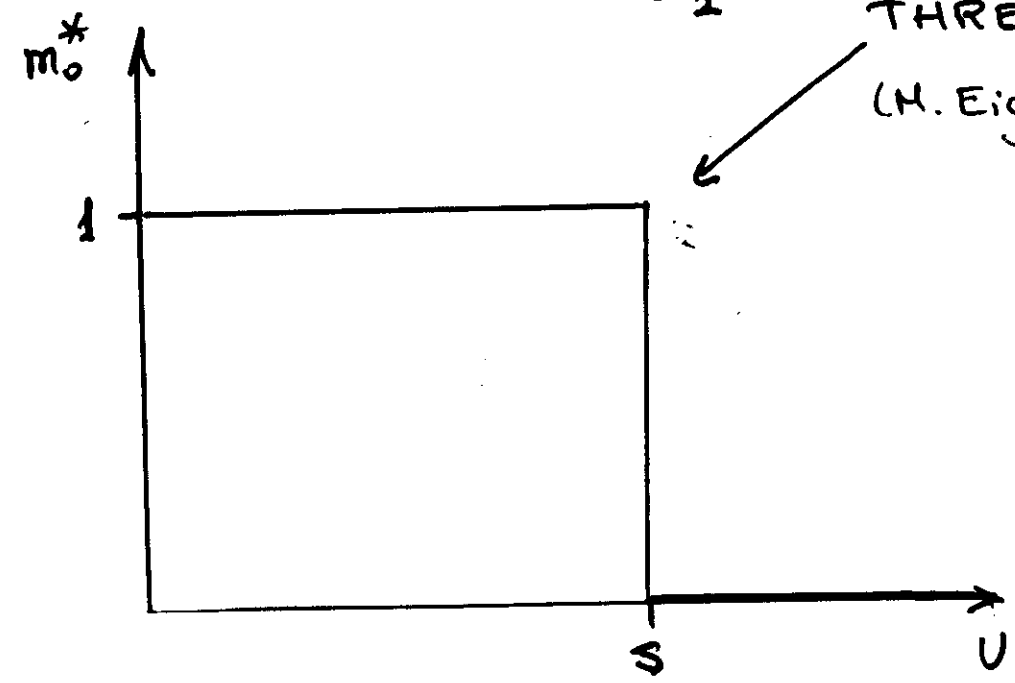

1.0
0.5
0.0
x_0
.0
0.5
1.0
u
10


1.0
0.5
0.0
E
0.0
0.5
1.0
u
80
40
20
10

$$\lim_{t\to\infty} x_0(t) = x_0^* = \begin{cases} 0 \\ 1 - U/S \end{cases}$$



$$\lim_{t\to\infty} m_0(t) = m_0^* = \begin{cases} 0 \\ 1 \end{cases}$$



ERROR THRESHOLD (M. Eigen, 1971)

Stationary state

$$\lim_{t\to\infty} x_k(t) = x_k^*$$

$$x_0^* = \begin{cases} 0 , & v > s \\ 1 - v/s , & v \le s \end{cases}$$

$$x_k^* = \frac{1}{1-s}\left(\frac{v(1-s)}{s(1-v)}\right)^k , \quad k>0$$

More detailed solution for arbitrary N

S. Galluccio, cond-mat/9705020

Order parameter

$$m^0 = \frac{1}{N}\sum_i S_i^0 \langle S_i \rangle$$

Sharp-peak fitness landscape

$$W_S = \begin{cases} 1, & S = S^0 \\ 1-s, & S \neq S^0 \end{cases}$$

Infinite genome limit

$$\lim_{N\to\infty} \mu N = U$$

+ only single mutations are allowed

$x_0 = \langle \delta_{SS^0} \rangle$ fraction of individuals with preferred genotype S^0

x_k: fraction of individuals with k mutations

Quasispecies equation

: neglecting sampling fluctuations (limit of infinite population size)

HIGGS

$$x_0(t+1) = \frac{(1-U)\, x_0(t)}{1 - s + s\, x_0(t)}$$

$$x_1(t+1) = \frac{U x_0(t) + (1-U)(1-s) x_1(t)}{1 - s + s\, x_0(t)}$$

$$x_2(t+1) = \frac{U(1-s) x_1(t) + (1-U)(1-s) x_2(t)}{1 - s + s\, x_0(t)}$$

$$\vdots$$

A model of an evolving population

- Population of M individuals

  $\alpha = 1, \ldots, M$

- Genotype $S^\alpha = (S_1^\alpha, \ldots, S_N^\alpha)$, $S_i^\alpha = \pm 1$

- Discrete (nonoverlapping) generations, one-parent (asexual) reproduction

  At each reproduction step, for each individual α of the new generation, one chooses its parent α' with a probability proportional to its fitness $w(S^{\alpha'})$

  $w(S) \geq 0 \Rightarrow$

  SIGN!

  $w(S) = e^{k F(S)}$

  k: Inverse SELECTION TEMPERATURE

  Only fitness ratios are relevant

  $F(S)$ is determined up to a constant

- Mutations affect randomly (but with small rate) the genotype

  $S_i^\alpha = \epsilon_i^\alpha S_i^{\alpha'}$; $\epsilon_i^\alpha = \pm 1$;

  $\overline{\epsilon_i^\alpha} = 1 - 2\mu$

A Commentary:

"So we see, in physics, disorder growing inexorably in systems isolated from their surroundings; and in biology, fitness increasing steadily in populations struggling for life. Ascent here and degradation there — almost too good to be true."

K. SIGMUND

A Question:

How is variability maintained?

MUTATIONS

Answers:

Fisher: the fitness function changes with time

Wright: populations are "small" (demes)

Kimura: many differences in genotype (mutations) do not change the fitness

... ALL ANSWERS ARE TRUE...

The Fundamental Theorem

R. E. Fisher

- The number ν_α of offspring of an individual α is proportional to its fitness $w_\alpha > 0$

  Fluctuations are neglected

- The fitness of an individual is the same as its parent's one

  One-parent reproduction

  Mutations are neglected

$$\langle w \rangle_t := \frac{1}{M} \sum_{\alpha=1}^{M} w_\alpha$$ Generation t

$$\langle w \rangle_{t+1} := \frac{1}{M} \sum_{\alpha=1}^{M} w_\alpha \nu_\alpha$$

$$x_\alpha = \frac{\nu_\alpha}{M} \propto w_\alpha$$

$$\langle w \rangle_{t+1} = \sum_{\alpha=1}^{M} w_\alpha x_\alpha = \frac{\sum_{\alpha=1}^{M} w_\alpha^2}{\sum_{\alpha=1}^{M} w_\alpha}$$

$$= \frac{\langle w^2 \rangle_t}{\langle w \rangle_t} > \langle w \rangle_t$$

1. The Darwinian Paradigm

- Living forms appear in populations of "like" individuals (species)
- No two individuals are exactly alike (variability)
- Part of this variation is inheritable (heritability)

  Def: The set of inheritable characteristics of an individual is its genotype.

- Variability implies different reproductive power (fitness)

  Def: The expected number of offspring of an individual is proportional to his fitness

- Individuals with higher reproductive power tend to have more offspring, themselves with higher reproductive power

$\Longrightarrow$ FITNESS INCREASES ∇_0 (Adaptation)

Table 14.5. The gene-for-gene model

III 3

		Parasite genotype	
		B_1	B_2
Host genotype	A_1	frequency $=pr$ host survives, $1-t$ dies, transmits B_1, t	frequency $=p(1-r)$ host survives, $1-s$ dies, transmits B_2, s
	A_2	frequency $=(1-p)\,r$ host survives, $1-s$ dies, transmits B_1, s	frequency $=(1-p)(1-r)$ host survives, $1-t$ dies, transmits B_2, t

$$p(n+1) = (1-t)\,p(n)\,r(n) + (1-s)\,p(n)\,(1-r(n)) + (1-s)\,(1-p(n))\,r(n) + (1-t)(1-p(n))(1-r(n))$$

$$r(n+1) = t\,p(n)\,r(n) + \mu s\,p(n)\,(1-r(n)) + (1-\mu)\,s\,(1-p(n))\,r(n)$$

CLARKE 1976 $\mu = 1$

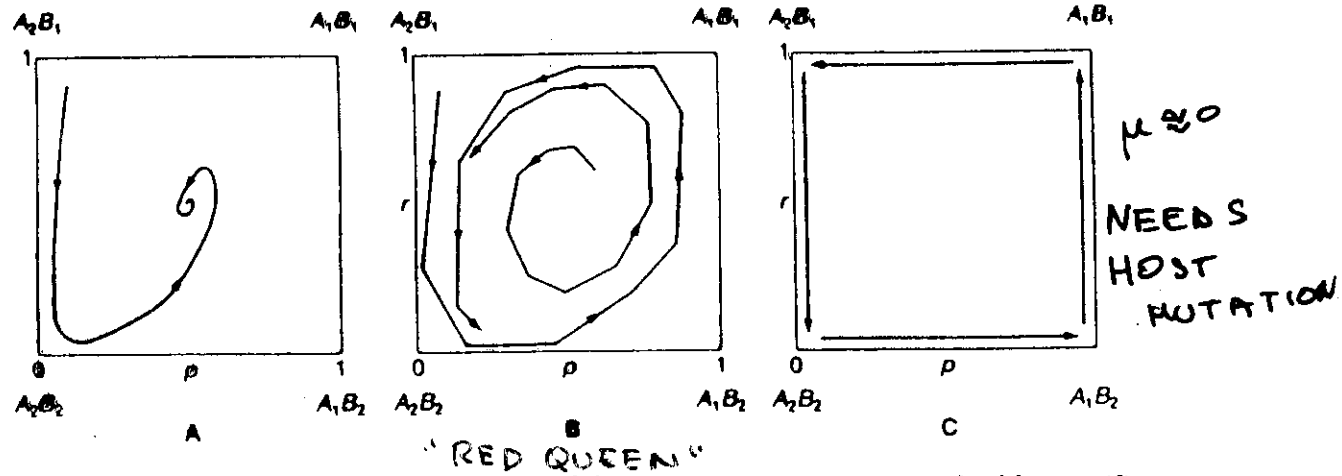


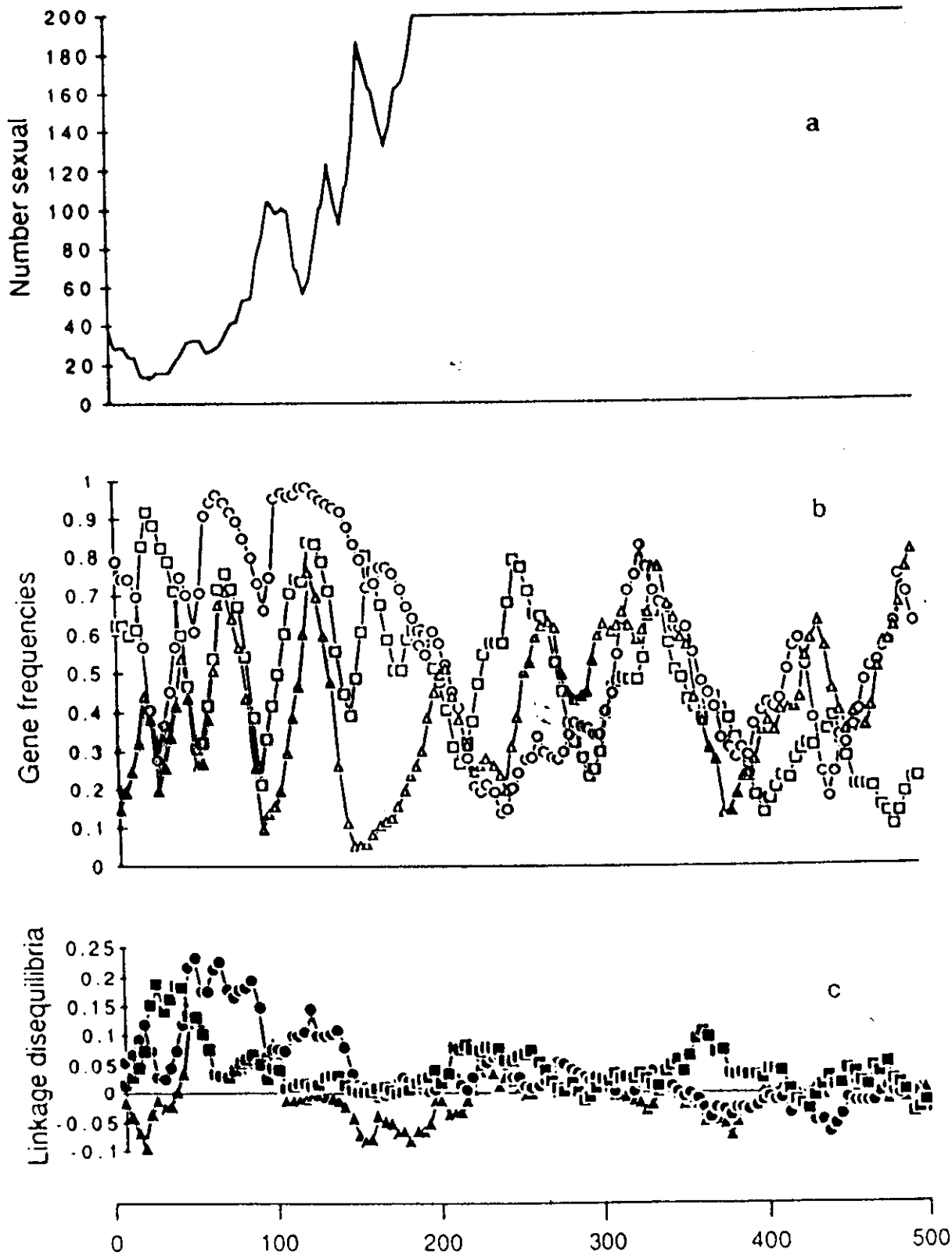
FIG. 14.13. Coevolution in a gene-for-gene model. A Parasite genotypes not highly specific; $s=0.25$, $t=0.6$. B Parasite genotypes more specific; $s=0.1$, $t=0.75$. C No evolution of parasites within resistant hosts. Note the transition from a stable polymorphism (A) to a limit cycle (B), and then to evolution to one corner of the state space, from which escape will occur by mutation.

COEVOLUTION AND SEX

III 4

Hamilton et al. PNAS 87 3566 (1990)

7 parasite species, 5 resistance loci (independent)



KAUFFMAN'S MODEL OF COEVOLUTION (NKC)

S. A. Kauffman, The Origins of Order, Oxford U. P. 1993, Chap. 6

- S interacting species, with genome length N
- Fitness of each species $F^{(\alpha)}$:

$$F^{(\alpha)} = F_k^{(\alpha)}(S^{(\alpha)}, S_{i_1}^{(\alpha_1)}, \ldots, S_{i_k}^{(\alpha_k)})$$

N.B. $K = N - 1 \Rightarrow$ R. E. M.

- Adaptive walk (sequential update)

NASH EQUILIBRIA

III 6

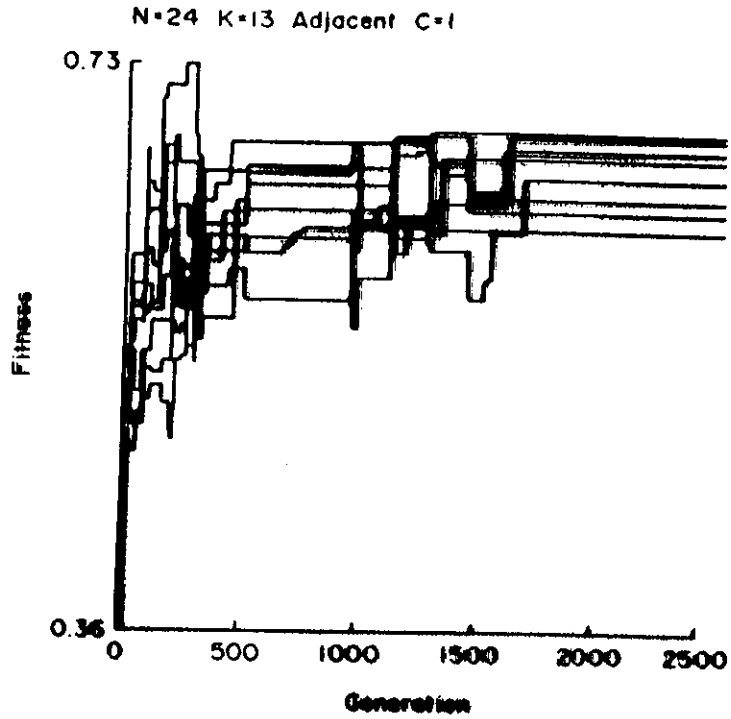
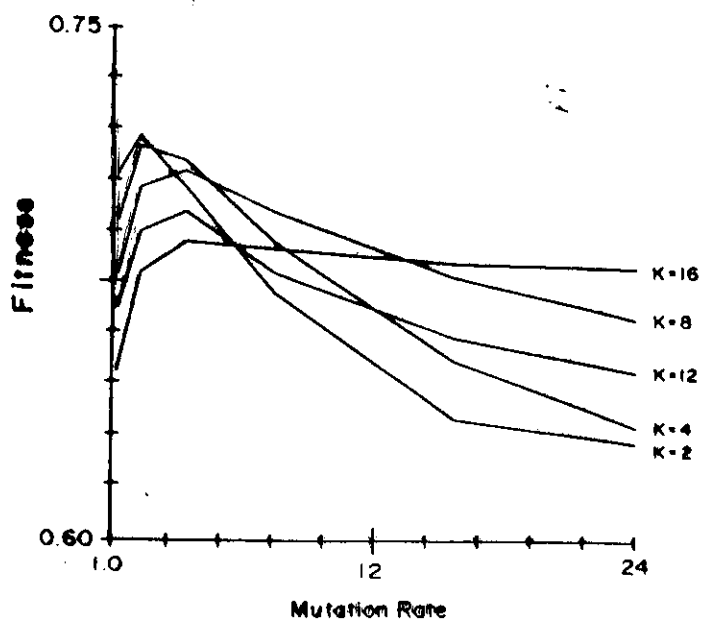
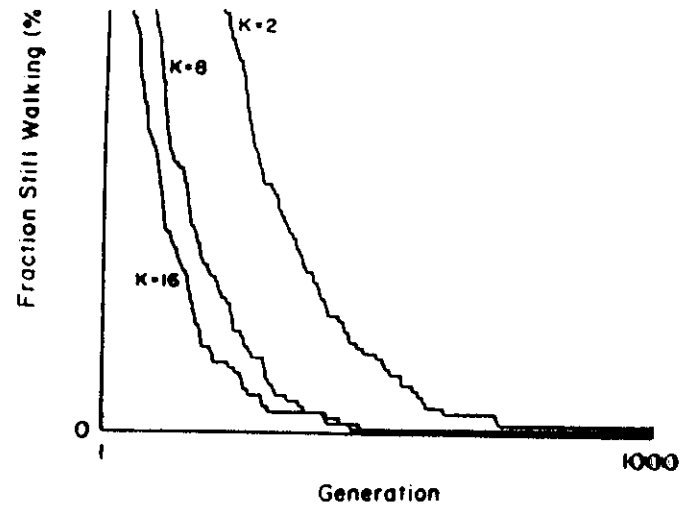


Figure 6.2 Coevolution among eight species, each governed by an *NK* landscape. Each of the *N* traits in each species is affected by $C = 1$ trait in each of the seven other species. System reaches a steady state about generation 1600. Note that mean fitness in the absence of selection is .5.





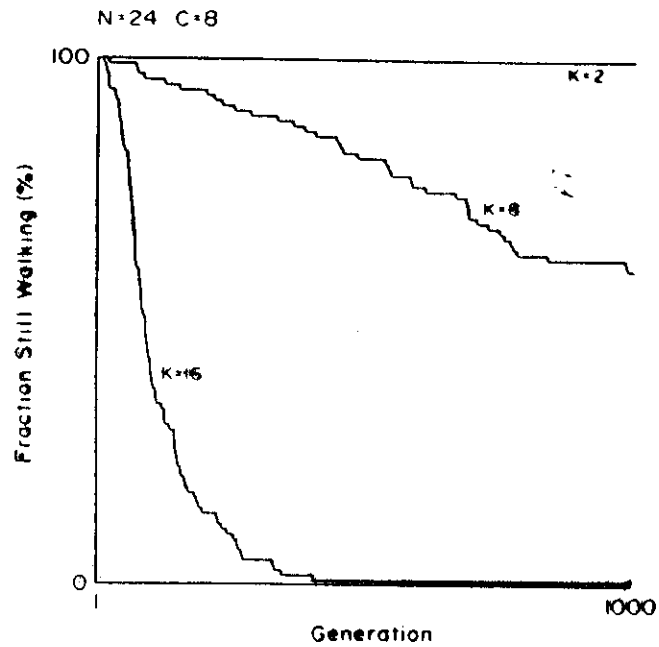
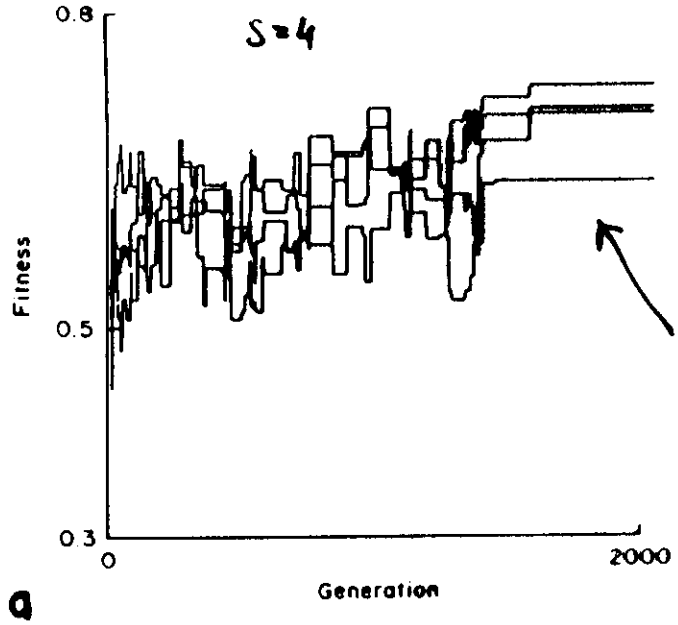


Figure 6.3 Fraction of 100 coevolving pairs of species which have not yet encountered a Nash equilibrium and hence are still walking, as a function of generations elapsed.

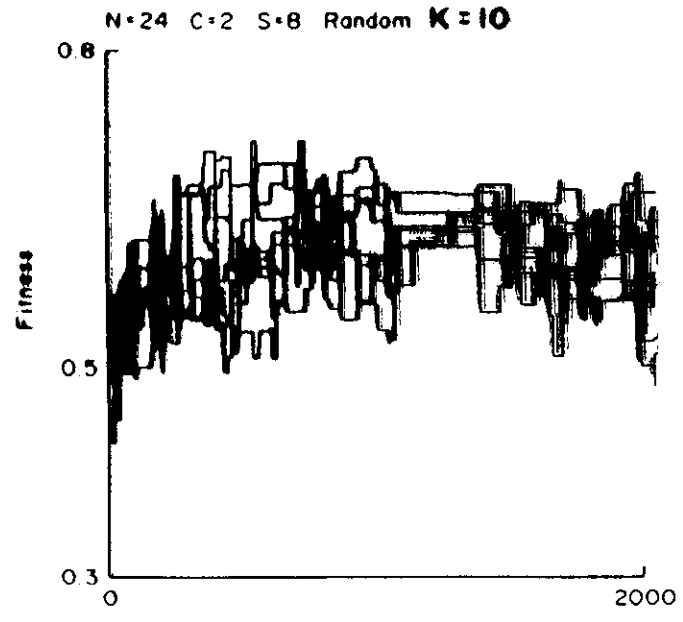
S: # of interacting species

III 8

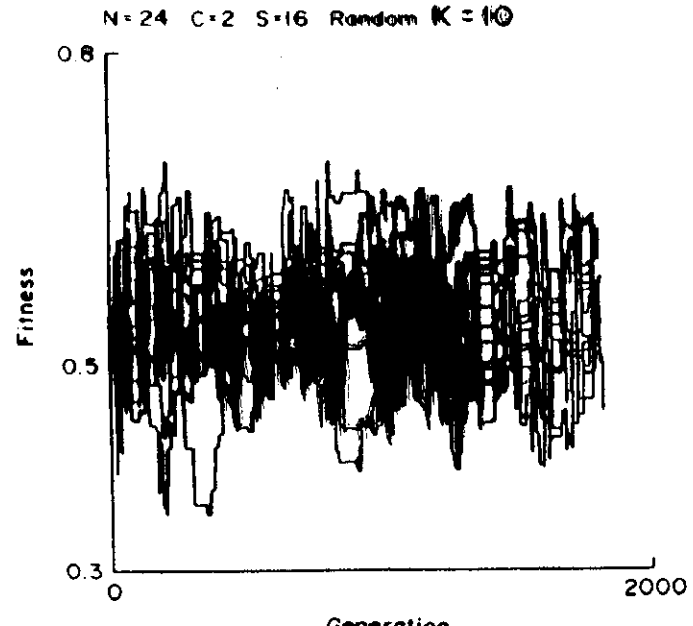


Nash equilibrium!

a



b



c

A SOLVABLE MODEL

P. Bak, H. Flyvbjerg, B. Lautrup

Phys. Rev. A 46, 6724 (1992)

- "Rugged fitness landscape" F random
- Large number of interacting species $S \to \infty$
- Genome length $N \gg 1$
- Annealed approximation: the C interacting species are chosen anew at each mutation step

$\rho_M(F;t)$ relative number of species with fitness F and M less fit mutants

$A(t)$ probability that a mutation is accepted activity

$$A(t) = \sum_{M=0}^{N} \left(1 - \frac{M}{N}\right) \int_0^1 dF\ \rho_M(F;t)$$

$\Phi(F;t)$ probability that a mutation is accepted and results in fitness F

$$\Phi(F;t) = \int_0^F dF' \ \frac{1}{1-F'} \sum_{M=0}^{N} \left(1 - \frac{M}{N}\right) \rho_M(F';t)$$

Master equation for $\rho_M(F;t)$

$$\frac{\partial}{\partial t}\rho_M(F;t) = -\left(1-\frac{M}{N}\right)\rho_M(F;t) + B_{M,N}(F)\,\Phi(F;t)$$
$$-\frac{C}{N}A(t)\rho_M(F;t) + \frac{C}{N}A(t)\,B_{M,N}(F)$$

$B_{M,N}(F)$ probability that M out of N mutants of genome with fitness F have fitness lower than F

$$B_{M,N}(F): \quad \frac{N!}{M!\,(N-M)!}\,F^M\,(1-F)^{N-M}$$

- Nash equilibrium
$$\rho_M(F;t) = \delta_{M,N}\,\rho(F)$$

- "Red Queen" phase

μ_1: # genes changed in adaptive walk of an isolated species

$$\mu_1 = \ln N + 0.09913\ldots + O(N^{-1})$$

If this change, on average, makes more than one species to evolve $\Longrightarrow$ CHAIN REACTION

$$\Rightarrow \quad C_{crit} = N/\mu_1 \simeq N/\ln N$$

Stationary solution

$$\rho^*_M(F) = \frac{N}{N-M+CA^*} B_{M,N}(F)\left[cA^* + \Phi^*(F)\right]$$

$$A^* = A^*[\rho^*], \quad \Phi^* = \Phi^*[\rho^*], \quad c = C/N$$

Assume A^* constant, multiply by $(1-M/N)/(1-F)$ and sum over M

$$\frac{d}{dF}\Phi^*(F) = g(F; cA^*)\left[cA^* + \Phi^*(F)\right]$$

$$g(F;x) = \frac{1}{1-F}\sum_{M=0}^{N-1} \frac{N-M}{N-M+Nx} B_{M,N}(F)$$

Def:

$$G(F;x) = \int_0^F dF'\, g(F';x)$$

$$G(x) = \int_0^1 dF\, e^{G(F;x)}$$

$$\Rightarrow \quad \Phi^*(F) = cA^*\left(e^{G(F;cA^*)} - 1\right)$$

$$\Rightarrow \quad A^* = cA^*\left[-1 + G(cA^*)\right]$$

$$A^* = \begin{cases} 0 \\ c^{-1} = -1 + G(cA^*) \end{cases}$$

III 11

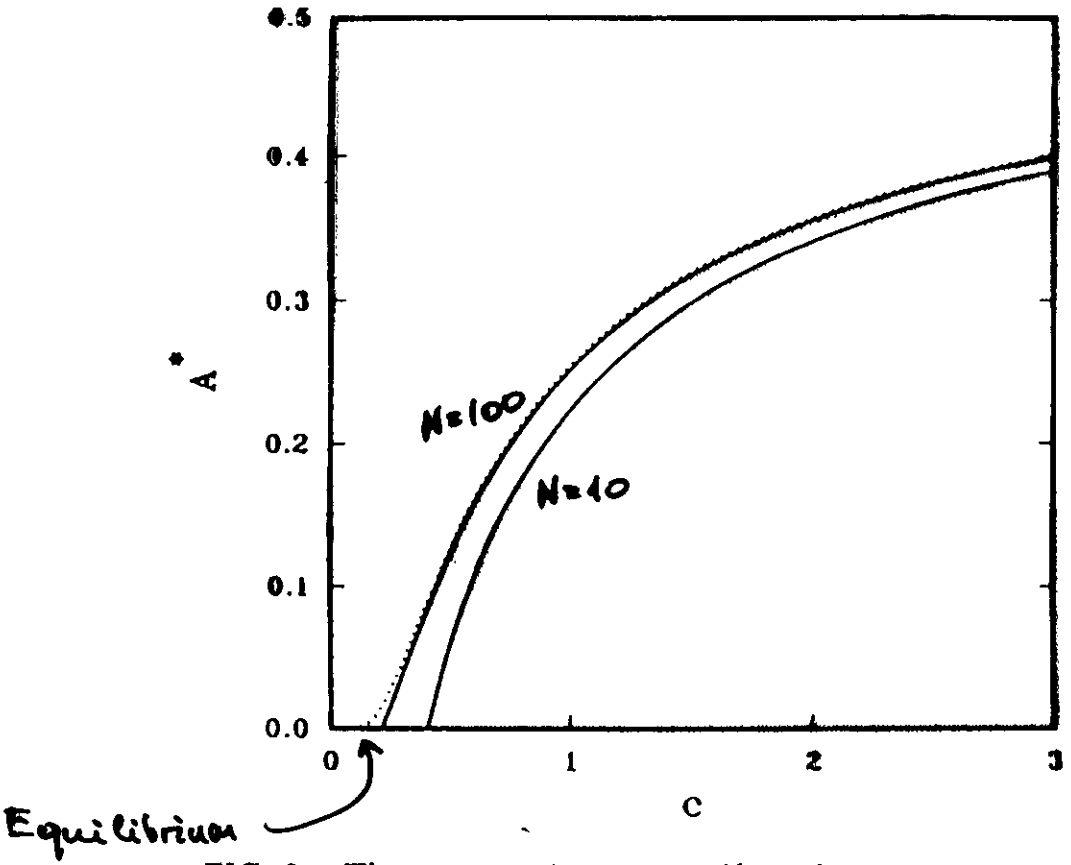


FIG. 2. The asymptotic activity A^* vs the connectivity c for $N = 10$ and 100 according to Eq. (19) (full curves) and Eq. (21) (dotted curve).

$c = C/N$

Eq. (21): $C^{-1} = -1 + (1+cA^*) \ln [1 + (cA^*)^{-1}]$

$(N \gg 1)$

EFFECTS OF VARIABILITY ?

- Within a quasi-species approach?

  UNEXPLORED: Probably transition equilibrium → Red Queen

- Finite populations?

  Muller's ratchet DESTABILIZES Nash equilibria

  ⇒ "Red queen" also in systems with small N and C

- Extremal dynamics: only the species most likely to mutate effectively mutates → chain reaction

  BAK - SNEPPEN

THE BAK-SNEPPEN MODEL

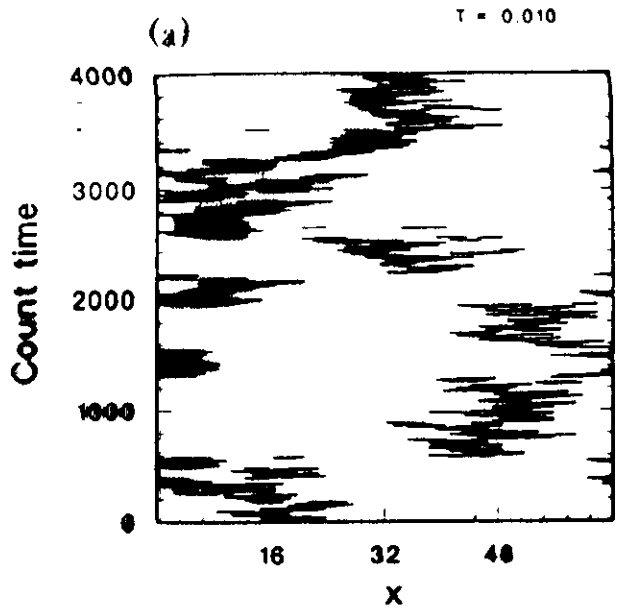
P. Bak, K. Sneppen, PRL 71, 4083 (1993)

- Species arranged in a lattice
- To each species a barrier height x_i is assigned
- The species with the smallest value of x_i "mutates"
- The "mutating" species plus its neighbors draw anew their x_i's

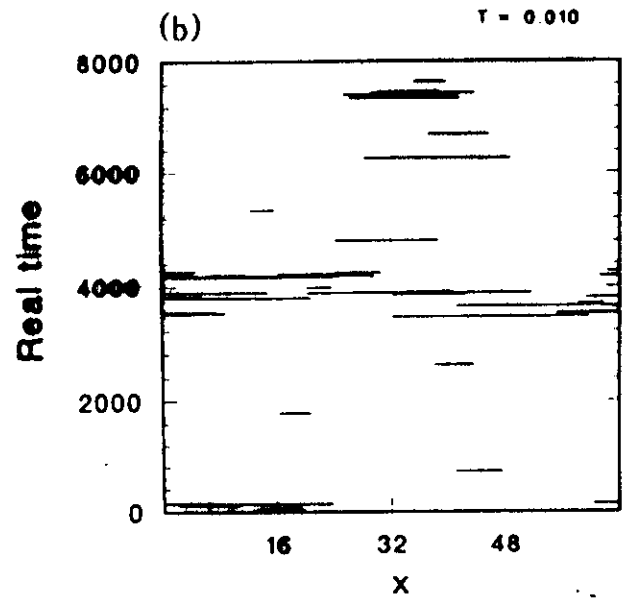
"Real time" $\propto \exp(x_a/T)$

N.B. The model holds provided Nash equilibria are stable (C and N not too large, populations of finite size)

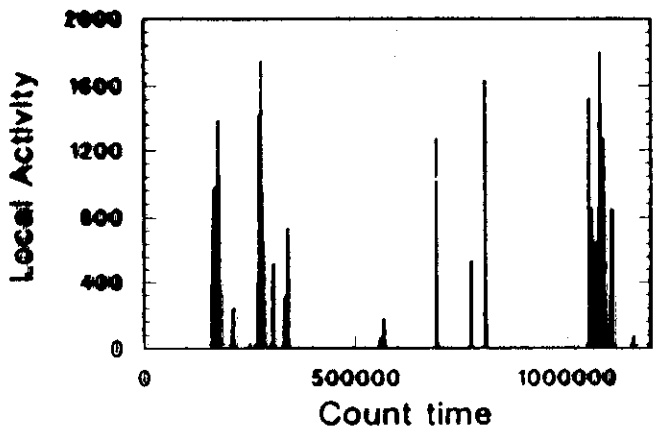
III 14

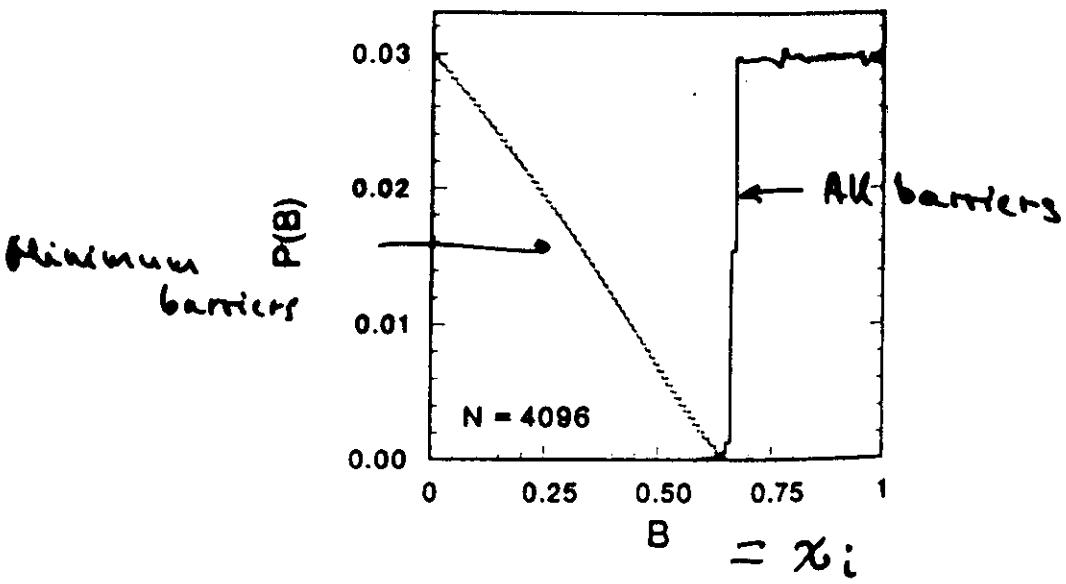


Activity

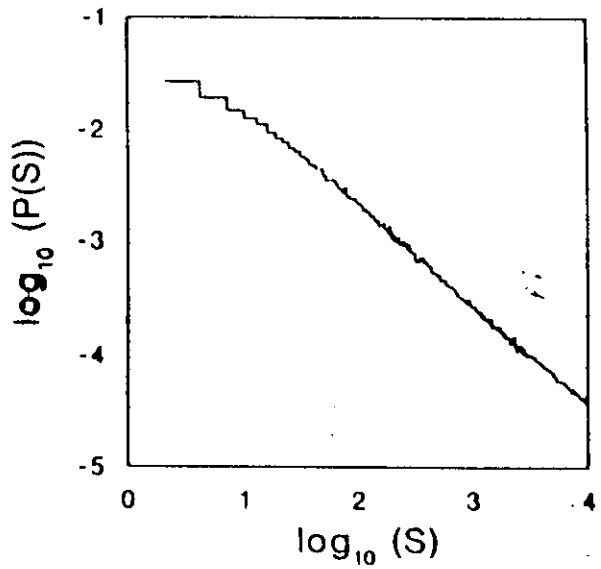


Activity





$= x_i$



Avalanche size

$N(s) \simeq s^{-0.9 \pm 0.1}$

EXACTLY SOLVABLE VERSION

H. Flyvbjerg, K. Sneppen, P. Bak, PRL 71 4087 (1993)

J. de Boer, B. Derrida ... PRL 73 901 (1994)

- Annealed random connectivity (as in the solvable version of the NKC model)

$P_n(t)$ probability that n out of N species in the "ecosystem" have barrier height $x_i < \lambda$

Connectivity $C = 2$

Master equation for $P_n(t)$

$$P_n(t+1) = \sum_{m=0}^{N} M_{n,m} P_m(t), \qquad (1)$$

for $n \geq 1$

$$M_{n+1,n} = \lambda^2 - \lambda^2 (n-1)/(N-1),$$
$$M_{n,n} = 2\lambda(1-\lambda) + (3\lambda^2 - 2\lambda)(n-1)/(N-1),$$
$$M_{n-1,n} = (1-\lambda)^2 + \frac{(-3\lambda^2 + 4\lambda - 1)(n-1)}{N-1},$$
$$M_{n-2,n} = (1-\lambda)^2 (n-1)/(N-1), \qquad (2)$$

$n = 0$

$$M_{0,0} = (1-\lambda)^2, \quad M_{1,0} = 2\lambda(1-\lambda), \quad M_{2,0} = \lambda^2. \qquad (3)$$

Limit $N \to \infty$: Biased random walk with boundary

$$P_0(t+1) = (1-\lambda)^2 [P_0(t) + P_1(t)]$$
$$P_1(t+1) = 2\lambda(1-\lambda)[P_0(t) + P_1(t)] + (1-\lambda)^2 P_2(t)$$
$$P_2(t+1) = \lambda^2 [P_0(t) + P_1(t)] + 2\lambda(1-\lambda) P_2(t) + (1-\lambda)^2 P_3(t)$$
$$\vdots$$
$$P_n(t+1) = \lambda^2 P_{n-1}(t) + 2\lambda(1-\lambda) P_n(t) + (1-\lambda)^2 P_{n+1}(t)$$

Asymptotic solution

(a) $\lambda < \frac{1}{2}$

$$P_0 = 1 - 2\lambda$$
$$P_1 = (1-2\lambda)\left[(1-\lambda)^{-2} - 1\right]$$
$$\vdots$$
$$P_n = (1-2\lambda)\,\lambda^{2n-2}\,(1-\lambda)^{-2n}$$
$$\vdots$$

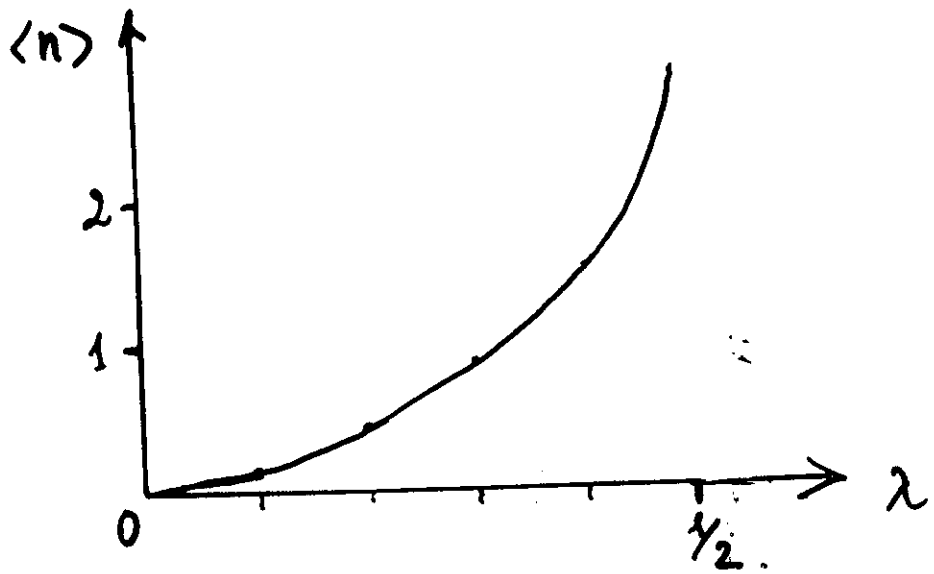
} independent of N

(b) $\lambda > \frac{1}{2} \Rightarrow \forall P_n \to 0$

(in fact the distribution is peaked around $n = (2\lambda - 1)N$)

$\lambda < \frac{1}{2}$

$$\langle n \rangle = \frac{\lambda(-2+3\lambda)}{-1+2\lambda}$$



Avalanches

Def.: λ-avalanche: evolution taking place between successive times where the number n vanishes

$Q_n(t)$ probability that there are n species with $x_i < \lambda$, given that the avalanche started t time steps ago

$q(t)$ probability that the avalanche has duration t

$Q_n(t)$ obeys the same equation as $P_n(t)$, but with $M_{0,n} = 0$

$$q(t) = (1-\lambda)^2 \left[Q_1(t-1) + \frac{1}{N-1} Q_2(t-1) \right]$$

$$Q_1(1) = 2\lambda(1-\lambda) \; ; \; Q_2(1) = \lambda^2 \; ; \; Q_n(1) = 0 \quad (n \geq 3)$$

$$\Rightarrow \quad Q_n(t) = \frac{2n\,(2t+1)!}{(t+n+1)!\;(t-n+1)!}\,\lambda^{t+n-1}\,(1-\lambda)^{t-n+1}$$

$$\Rightarrow \quad q(t) = \frac{(2t)!}{(t+1)!\;t!}\,\lambda^{t-1}\,(1-\lambda)^{t+1}$$

$$\langle t \rangle = (1-2\lambda)^{-1}$$

$$q(t) \simeq \frac{(1-\lambda)\,[4\lambda(1-\lambda)]^t}{t^{3/2}\,\lambda\sqrt{\pi}}$$

Scaling

$N \to \infty$, $\lambda = \frac{1}{2} + \frac{\alpha}{\sqrt{N}}$

$$P_n = \frac{1}{\sqrt{N}} f\left(\frac{n}{\sqrt{N}}\right)$$

$$Q_n(t) = \frac{1}{\sqrt{N}} g\left(\frac{n}{\sqrt{N}}, \frac{t}{N}\right)$$

$$\frac{1}{4}\frac{d^2 f}{dx^2} + (x - 2\alpha)\frac{df}{dx} + f = 0$$ truncated Gaussian

$$\frac{\partial g}{\partial \tau} = g + (x - 2\alpha)\frac{\partial g}{\partial x} + \frac{1}{4}\frac{\partial^2 g}{\partial x^2}$$

III 1920

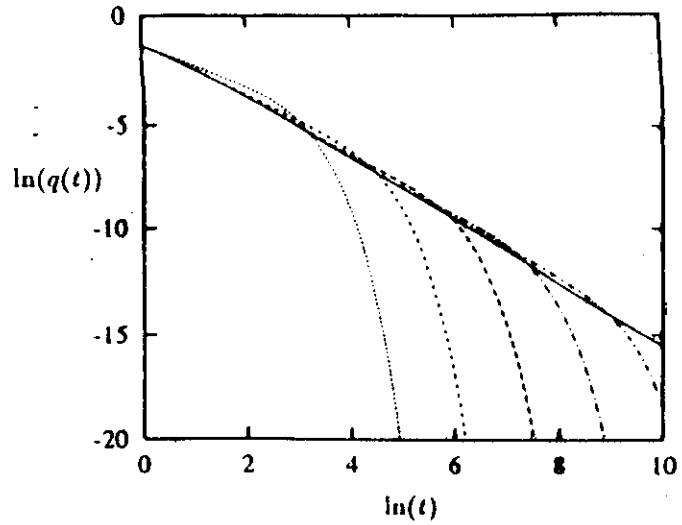


FIG. 1. Dotted and dashed lines: distribution of avalanche lifetimes $q(t)$ at critical point for $N = 5$, 25, 125, 625, and 3125. Full line: analytical expression (26). All cases have $K = 2$ and $\lambda = 1/2$. The results for finite N are exact, obtained by iterating the master equation (1)–(3) numerically.

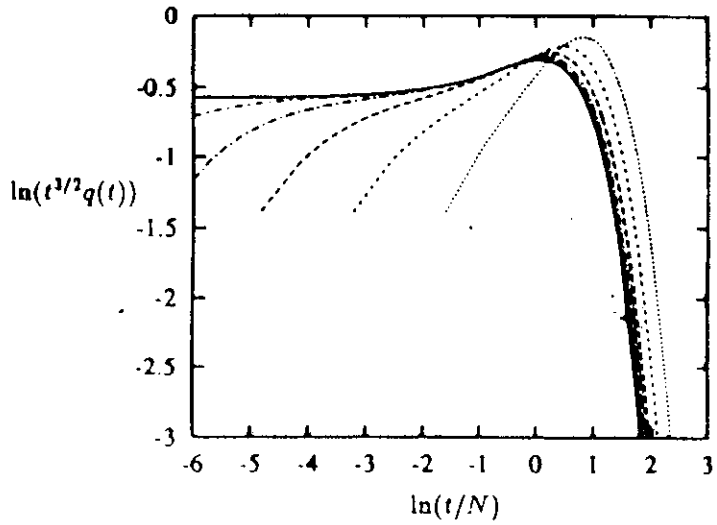


FIG. 2. Same results as Fig. 1, except that we show $t^{3/2}q(t)$ versus t/N so that the pure $t^{-3/2}$ power law behavior corresponds to a straight horizontal line.

- Punctuated equilibria from coevolution
- Self-organized steady state ?
- Speciation (branching) and <u>extinction</u>

$\Rightarrow$ the RESET model

Newman's model