Phase Transitions in Evolution

When do quasispecies form error thresholds?

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

Complex Systems Seminar

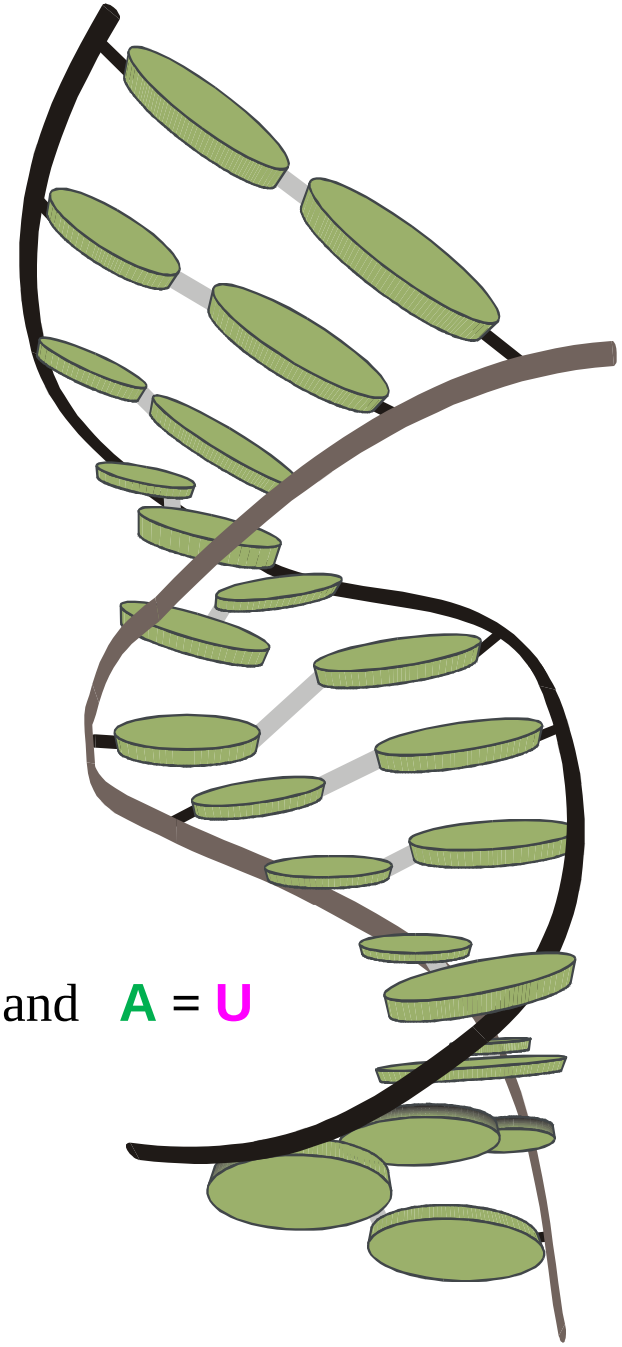
Universität Wien, 21.03.2014

Web-Page for further information:

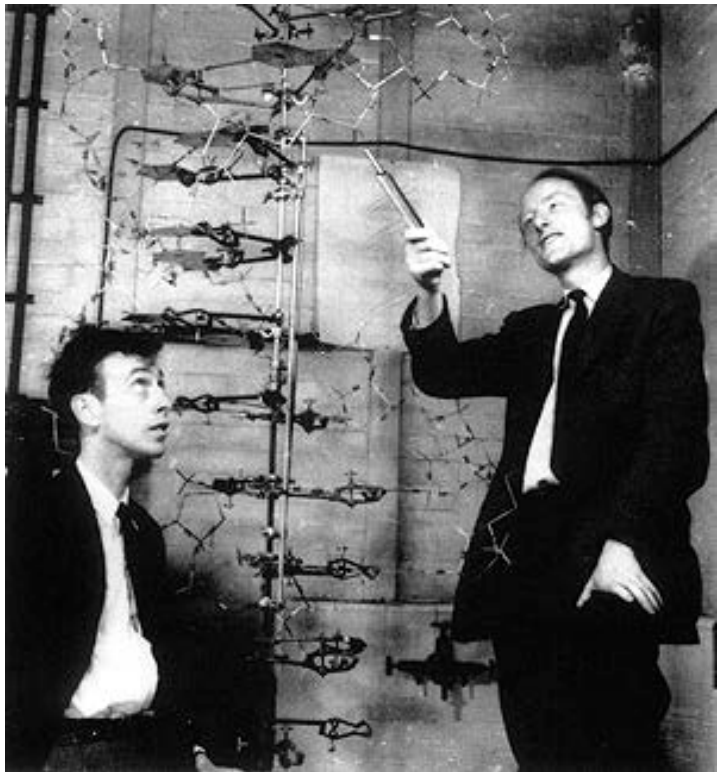
http://www.tbi.univie.ac.at/~pks

1. What is a „quasispecies“?
2. Detection of the „error threshold“
3. Error thresholds on „simple landscapes“
4. Error thresholds and phase transitions
5. „Realistic“ landscapes
6. Neutrality in evolution

1. **What is a „quasispecies"?**
2. Detection of the „error threshold"
3. Error thresholds on „simple landscapes"
4. Error thresholds and phase transitions
5. „Realistic" landscapes
6. Neutrality in evolution



G ≡ C and A = U



James D. Watson, 1928- , and Francis Crick, 1916-2004,
Nobel Prize 1962

The three-dimensional structure of a short double helical stack of B-DNA

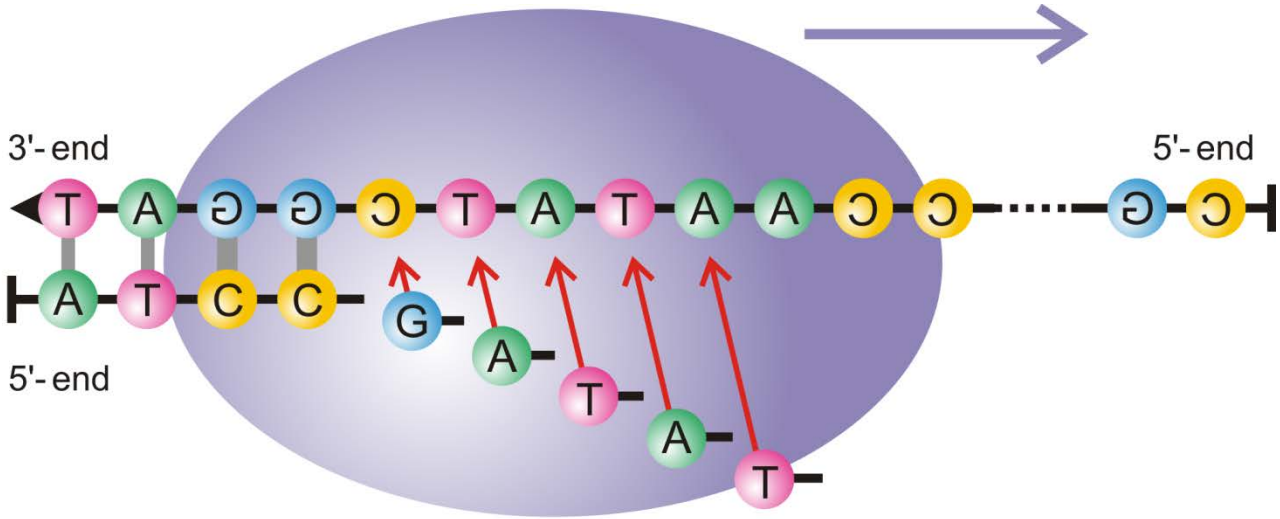
adenine A

thymine T

uracil U

guanine G

cytosine C



correct replication

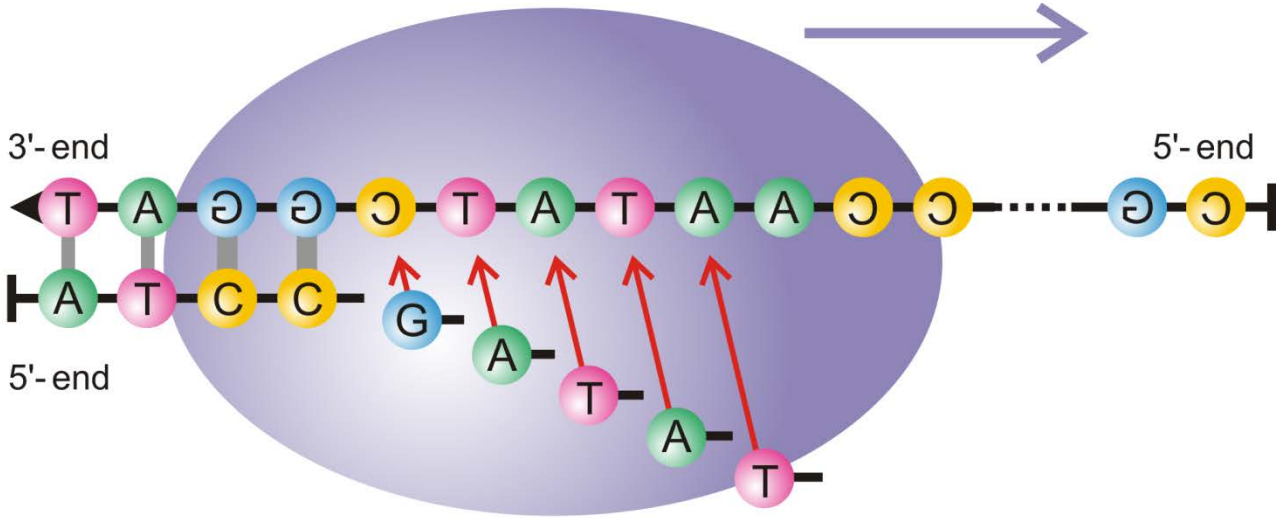
adenine A

thymine T

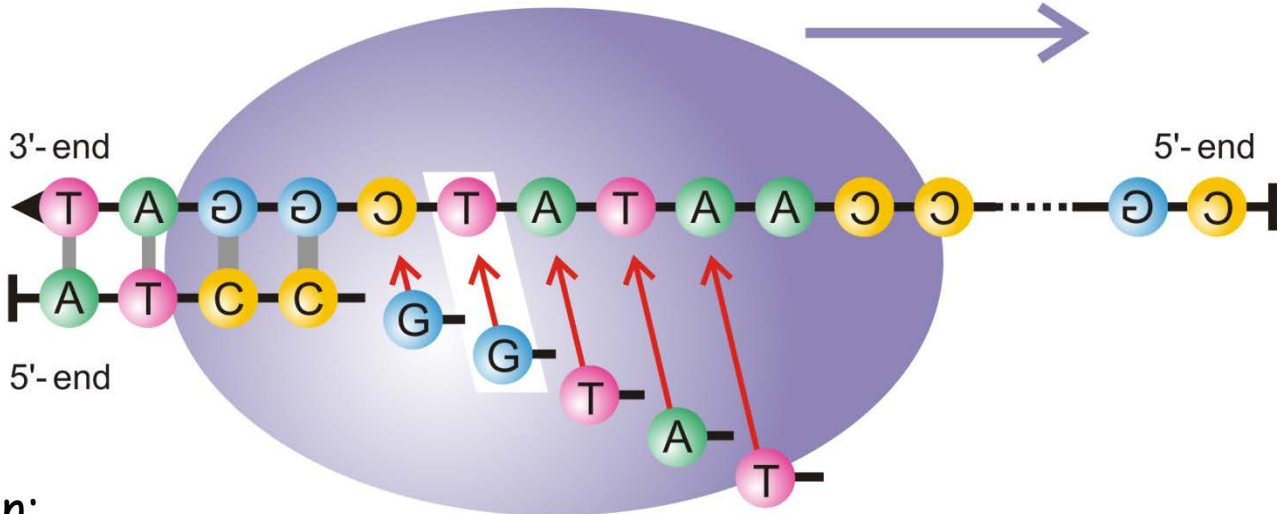
uracil U

guanine G

cytosine C



correct replication



mutation

accuracy of replication:

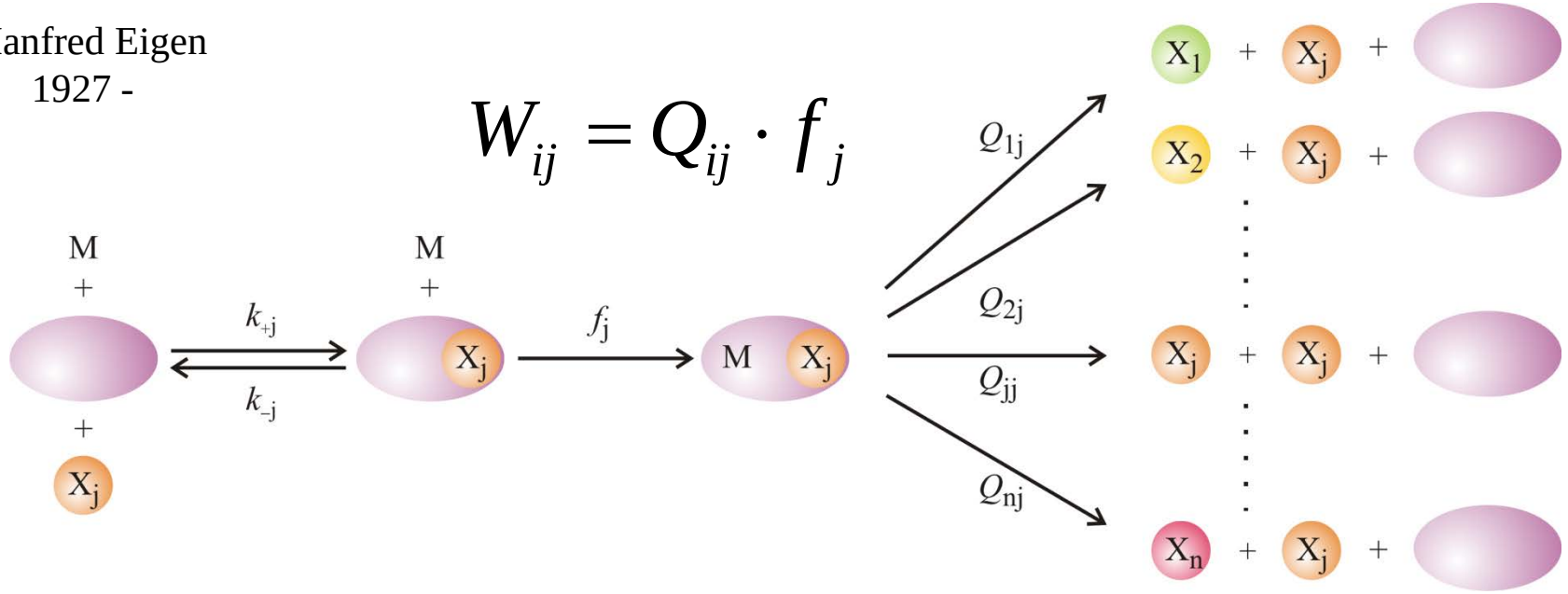
$$Q = q_1 \cdot q_2 \cdot q_3 \cdot q_4 \cdot \ldots$$

Manfred Eigen
1927 -

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

$$\Phi = \sum_{i=1}^{n} f_i\, x_i \Big/ \sum_{i=1}^{n} x_i$$

$$W_{ij} = Q_{ij} \cdot f_j$$



Mutation and (correct) replication as parallel chemical reactions

M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977. *Naturwissenschaften* 64:541, 65:7 und 65:341

$$\frac{\mathrm{d}x_j}{\mathrm{dt}} = \sum_{i=1}^{n} Q_{ji}\, f_i\, x_i - x_j \cdot \phi\,;\ j = 1, 2, \ldots, n \ \text{ with } \ \phi = \sum_{i=1}^{n} f_i x_i = \overline{f}$$

$$z_j(t) = x_j(t) \cdot \exp\left(\int_0^t \phi(\tau) d\tau\right) \ \text{ with } \ \exp\left(\int_0^t \phi(\tau) d\tau\right) = \sum_{i=1}^{n} z_i(t)$$

$$\frac{\mathrm{d}z_j}{\mathrm{dt}} = \sum_{i=1}^{n} W_{ji}\ z_i = \sum_{i=1}^{n} Q_{ji}\, f_i\ z_i\,;\ j = 1, 2, \ldots, n \quad \text{or} \quad \frac{\mathrm{dz}}{\mathrm{dt}} = \mathrm{Q} \cdot \mathrm{F}\ \mathbf{z}$$

W … nonnegative, primitive: W^m … strictly positive

Perron-Frobenius theorem applies

$$\mathrm{B}^{-1} \cdot \mathrm{W} \cdot \mathrm{B} = \Lambda\,, \quad \lambda_0 > |\lambda_1| \geq \ldots \geq |\lambda_n|$$

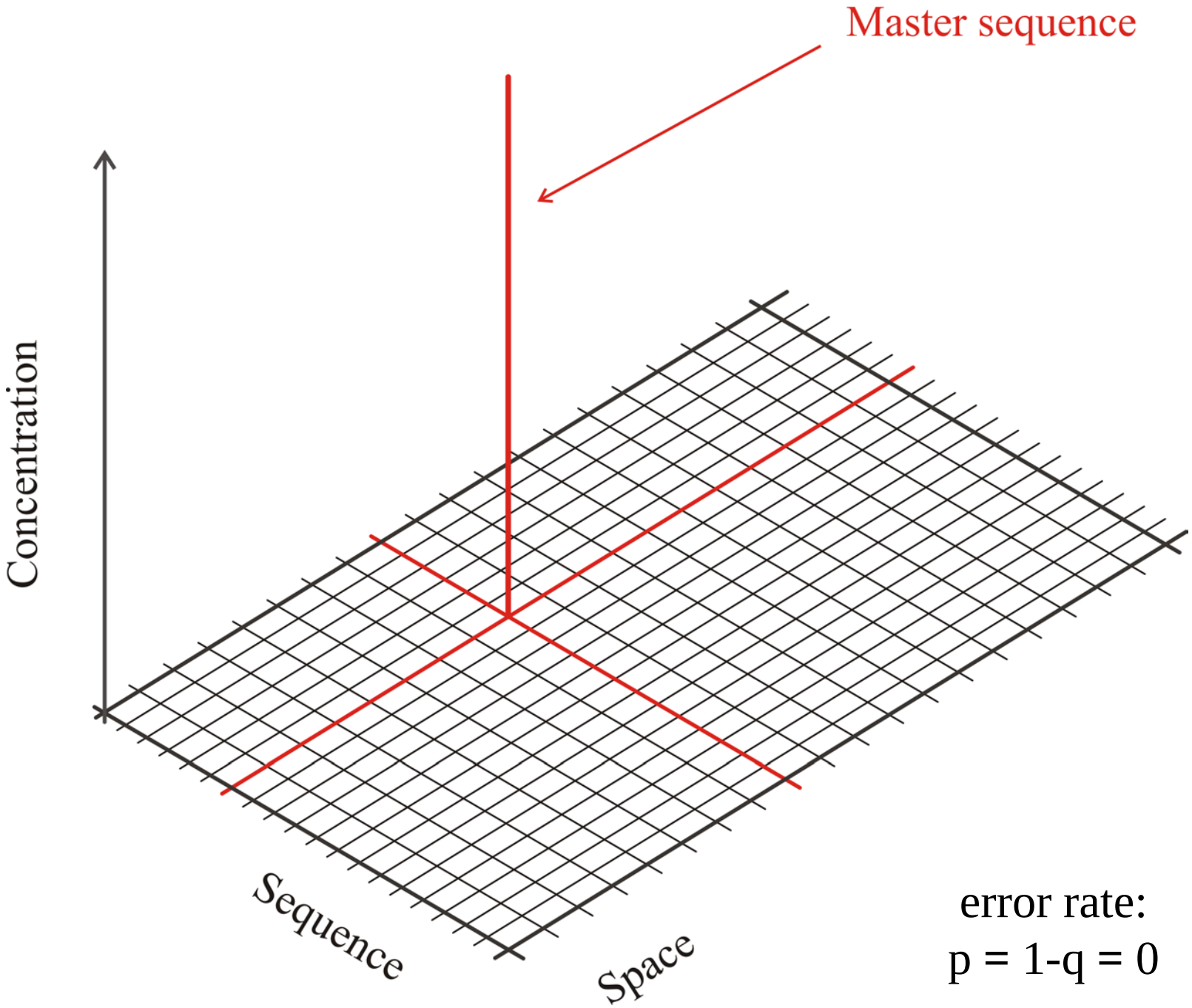
$$\mathrm{B}^{-1} \cdot \mathrm{W} \cdot \mathrm{B} = \Lambda, \quad \lambda_0 > |\lambda_1| \geq \ldots \geq |\lambda_n|$$

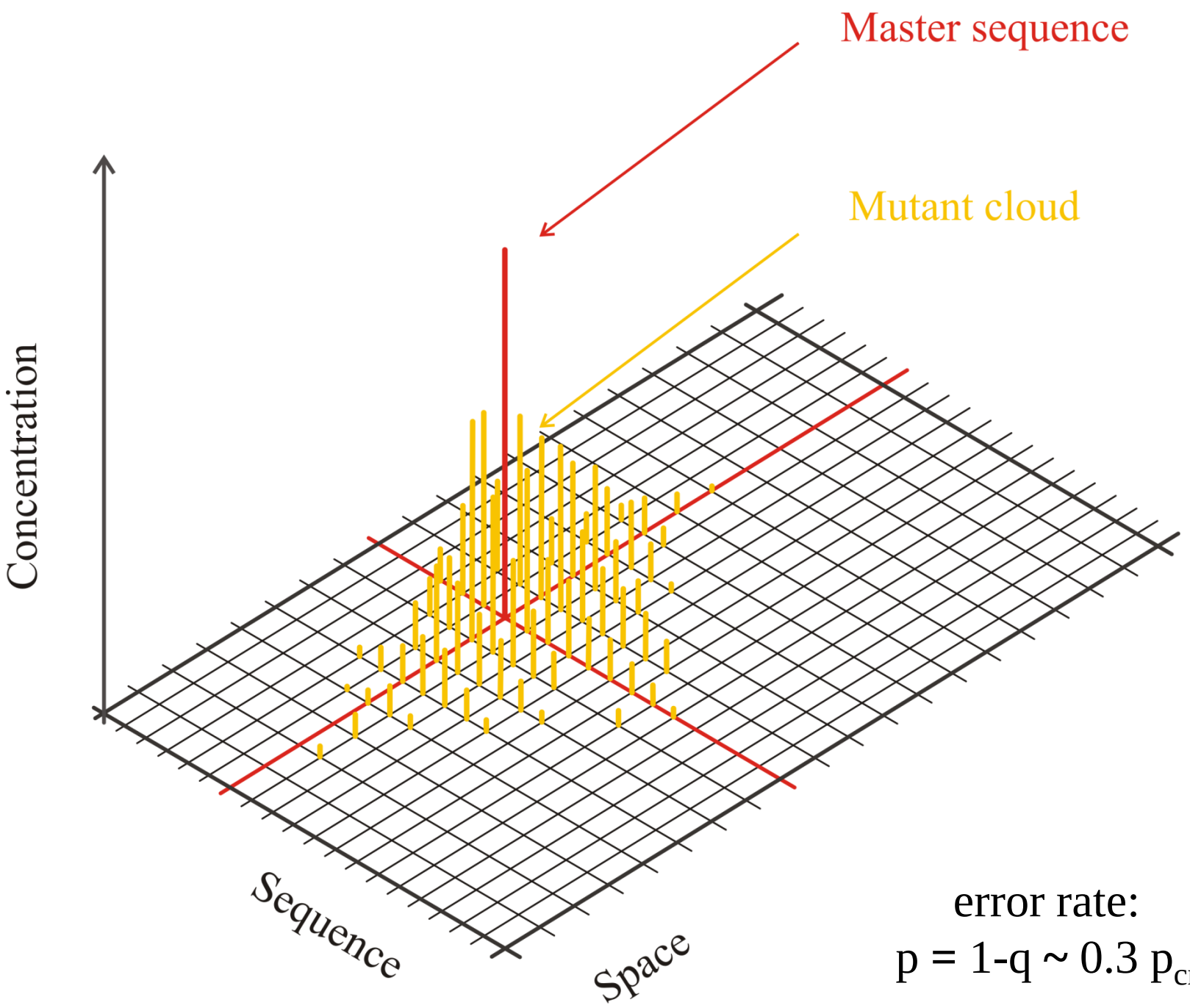
$$\mathrm{B}^{-1} = H \quad \text{and} \quad \beta_k(0) = \sum_{i=1}^{n} h_{ki}\, z_i(0) = \sum_{i=1}^{n} h_{ki}\, x_i(0)$$

$$x_j(t) = \frac{\sum_{k=0}^{n-1} b_{jk}\, \beta_k(0)\, \exp(\lambda_k\, t)}{\sum_{i=1}^{n} \sum_{k=0}^{n-1} b_{ik}\, \beta_k(0)\, \exp(\lambda_k\, t)}$$

$$\bar{x}_j(t) = \frac{b_{j0}\, \beta_0(0)\, \exp(\lambda_0\, t)}{\sum_{i=1}^{n} b_{i0}\, \beta_0(0)\, \exp(\lambda_0\, t)}$$

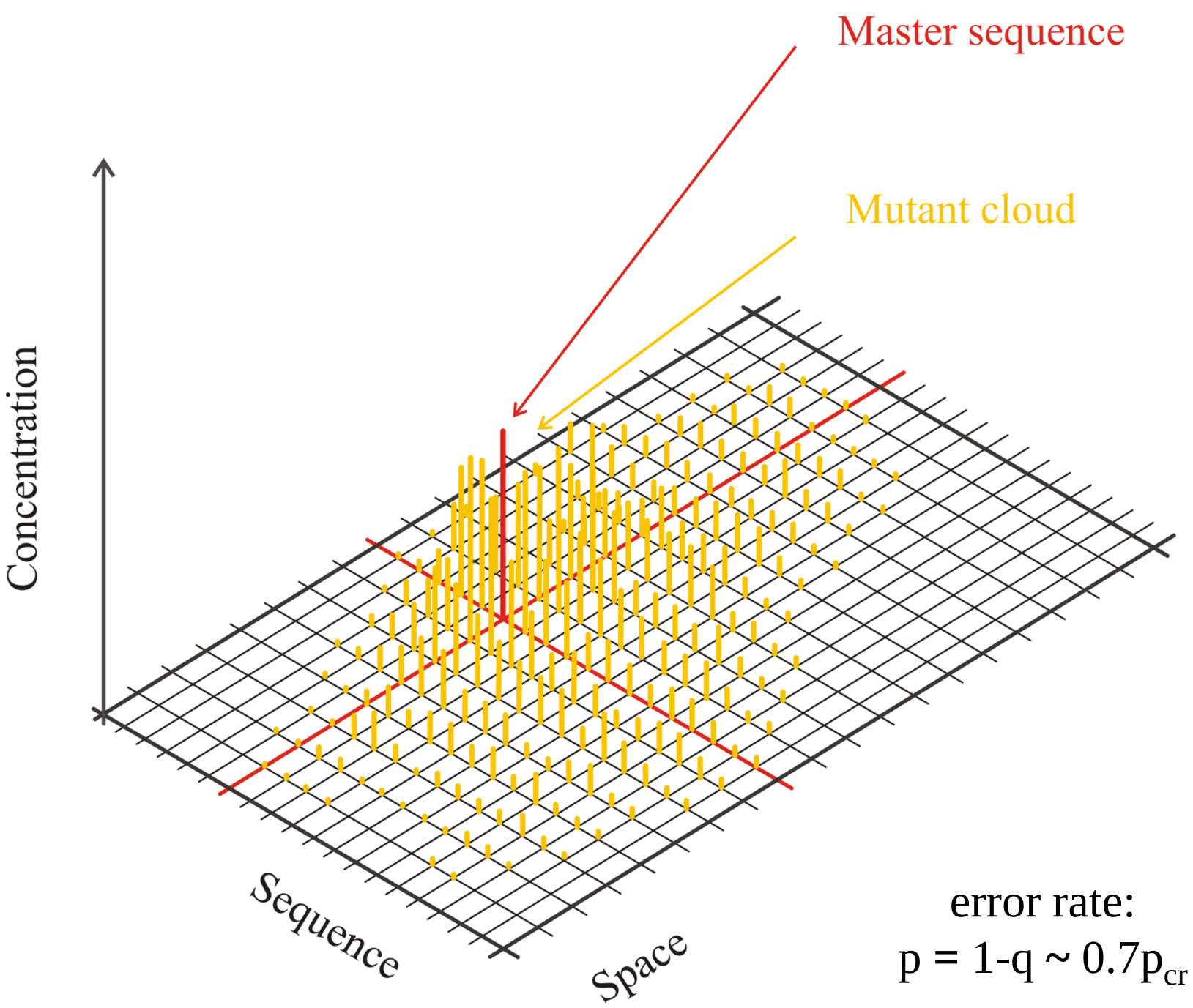
quasispecies


Master sequence
Concentration
Sequence
Space
error rate:
p = 1-q = 0



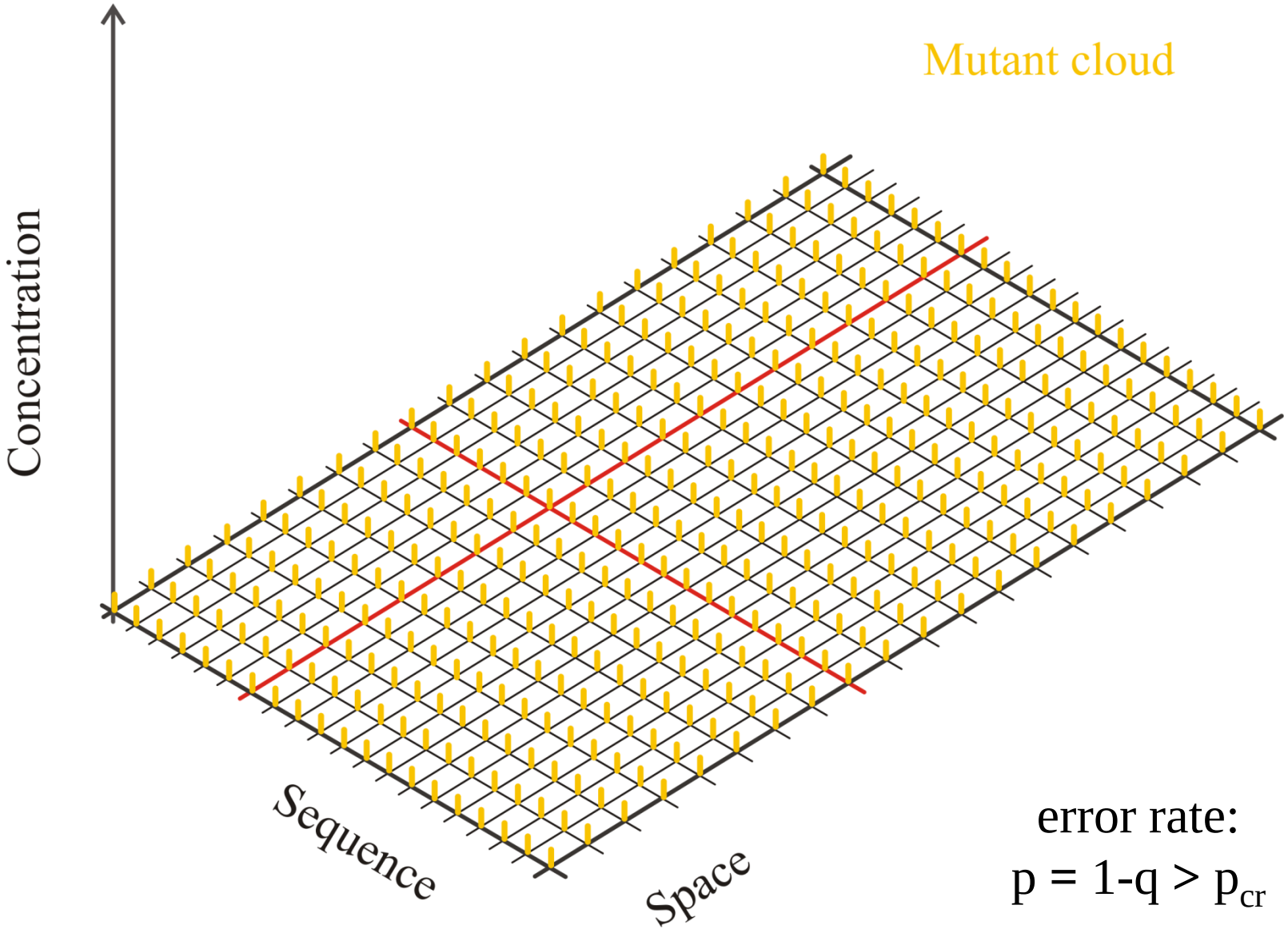
error rate:

$p = 1\text{-}q \sim 0.3\ p_{cr}$



error rate:

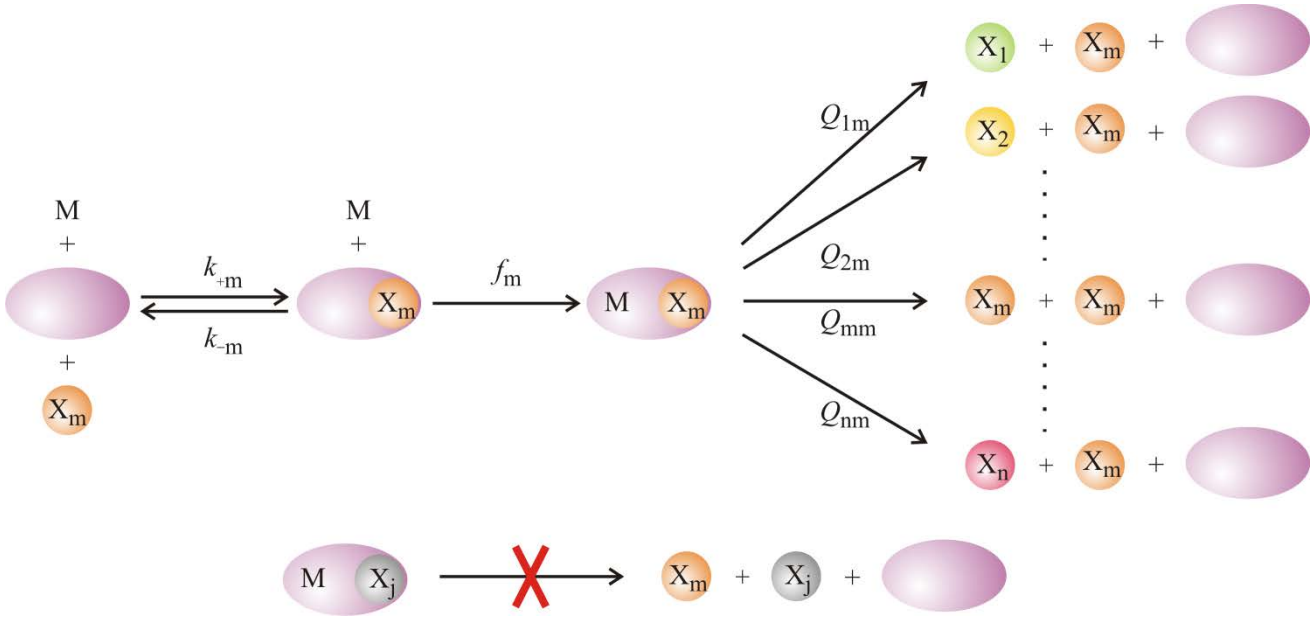
$p = 1\text{-}q \sim 0.7p_{cr}$



error rate:

$p = 1\text{-}q > p_{cr}$

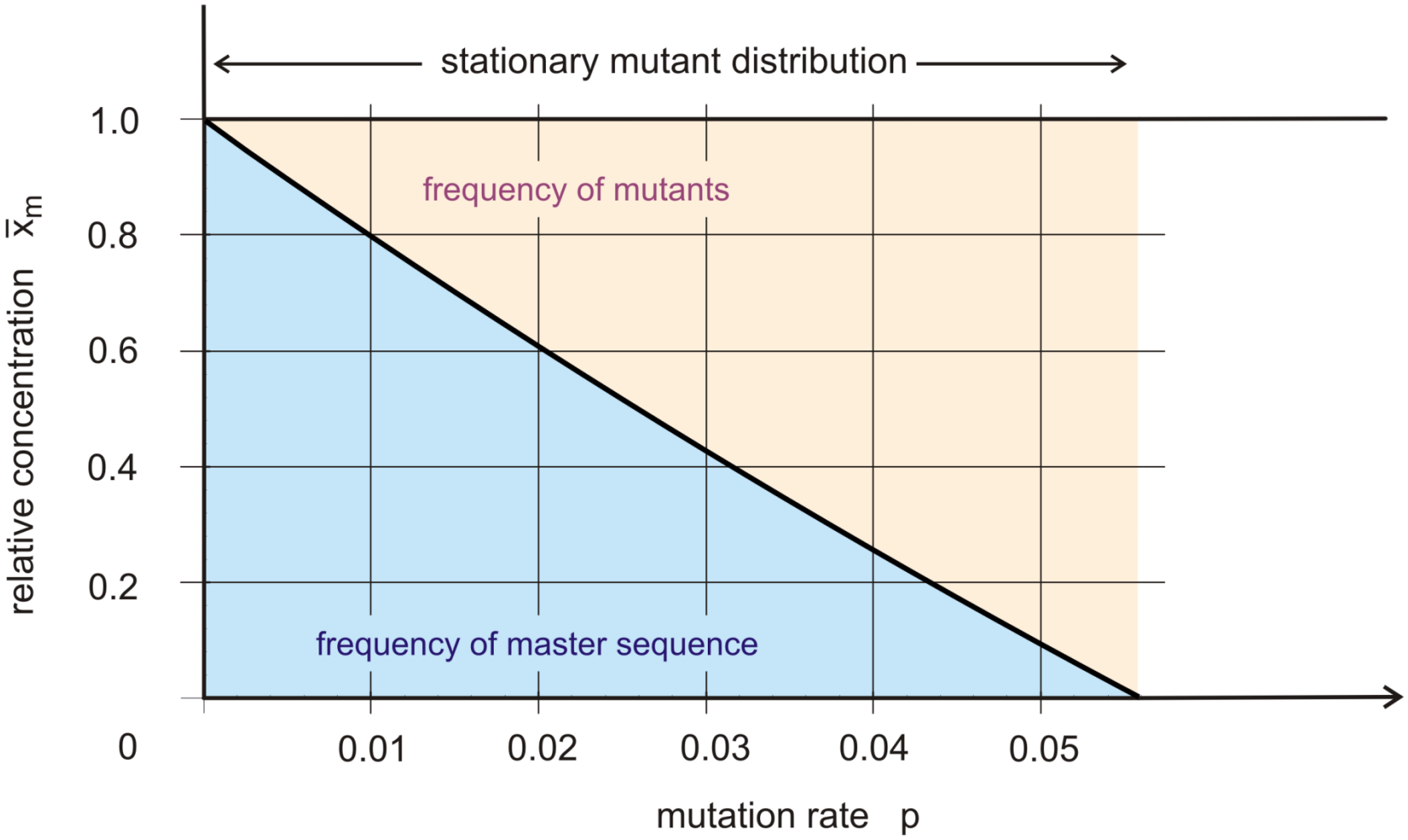
1. What is a „quasispecies"?
2. **Detection of the „error threshold"**
3. Error thresholds on „simple landscapes"
4. Error thresholds and phase transitions
5. „Realistic" landscapes
6. Neutrality in evolution

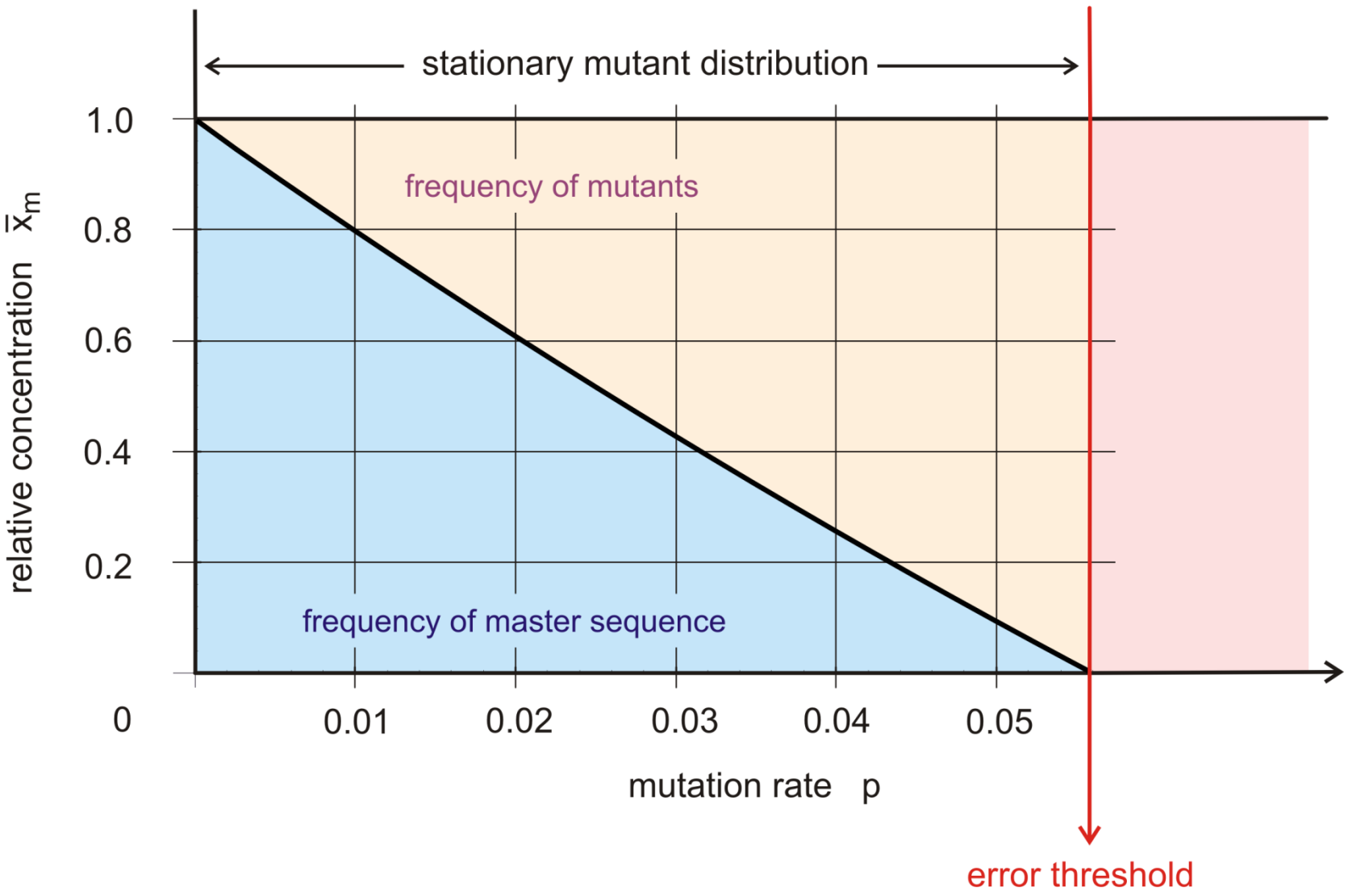


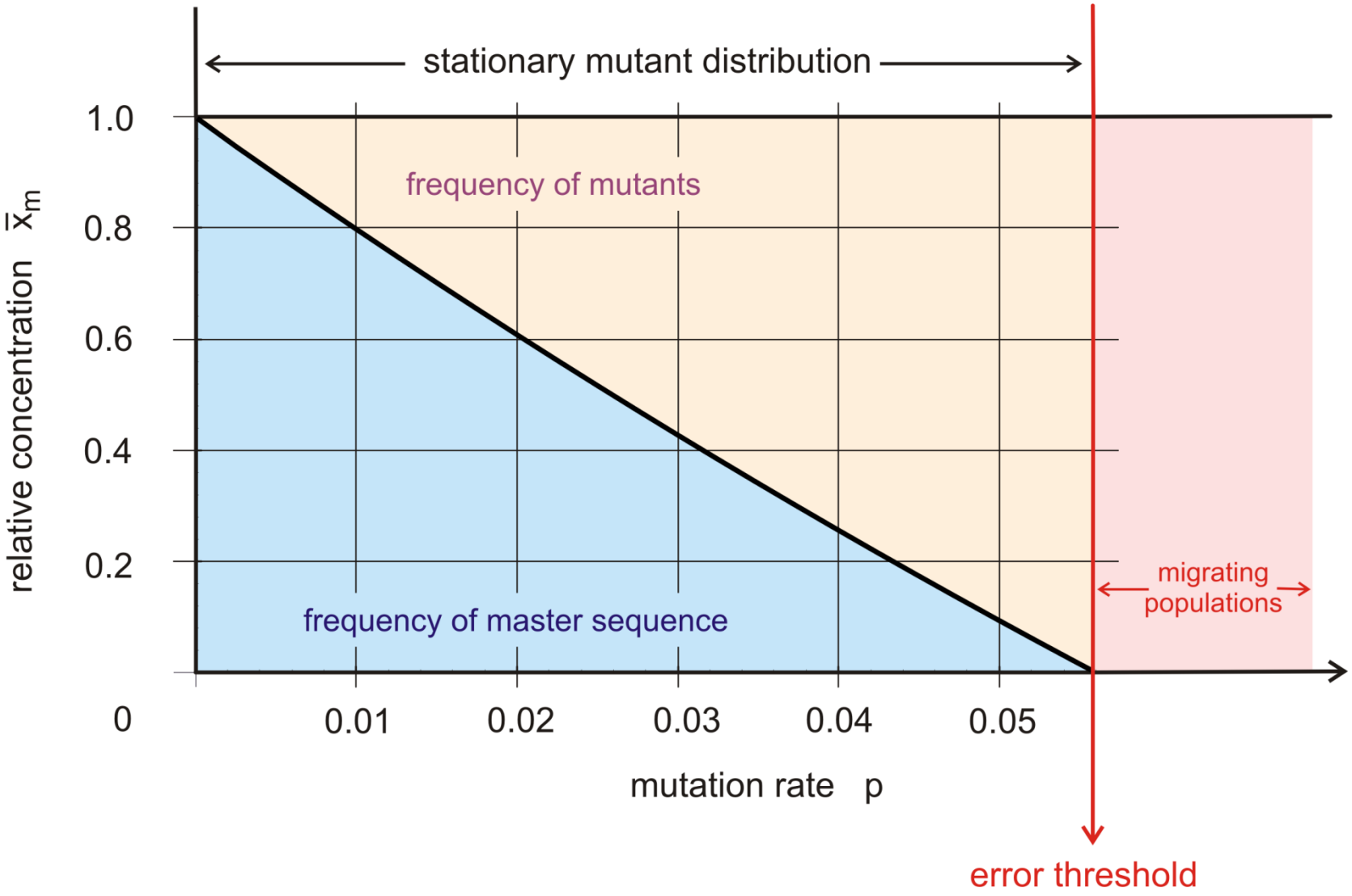
$$\frac{dx_m}{dt} = Q_{mm}\, f_m\, x_m - x_m \Phi = \left(Q_{mm}\, f_m - \Phi\right) x_m\,,\ \Phi = \sum_{i=1}^{n} f_i\, x_i \Big/ \sum_{i=1}^{n} x_i$$

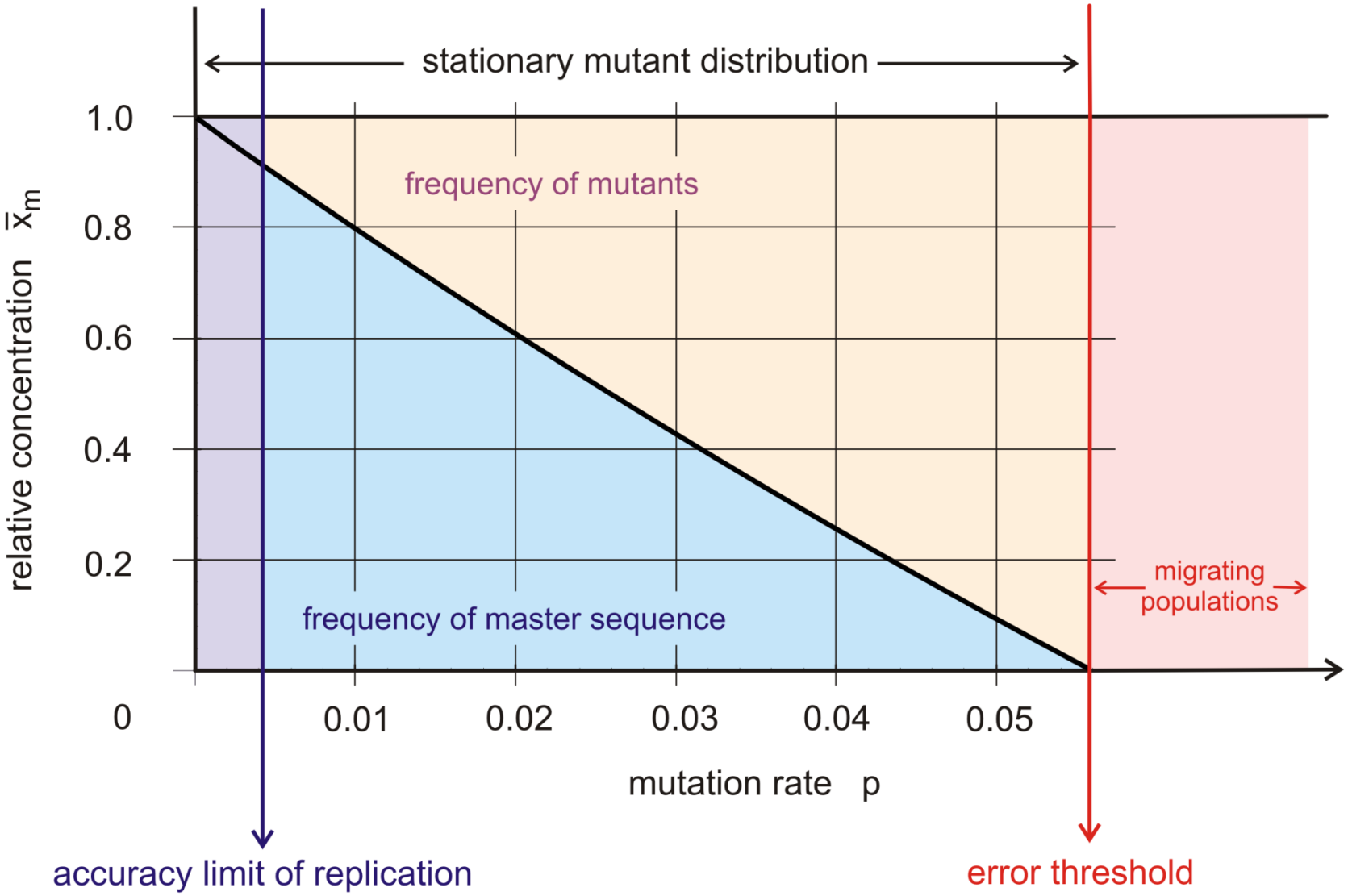
$$\bar{x}_m = \frac{Q_{mm} - \sigma_m^{-1}}{1 - \sigma_m^{-1}}\,,\ \ \sigma_m = f_m / \bar{f}_{-m}\,,\ \ \bar{f}_{-m} = \sum_{i=1, i\neq m}^{n} f_i\, x_i \Big/ 1 - x_m$$

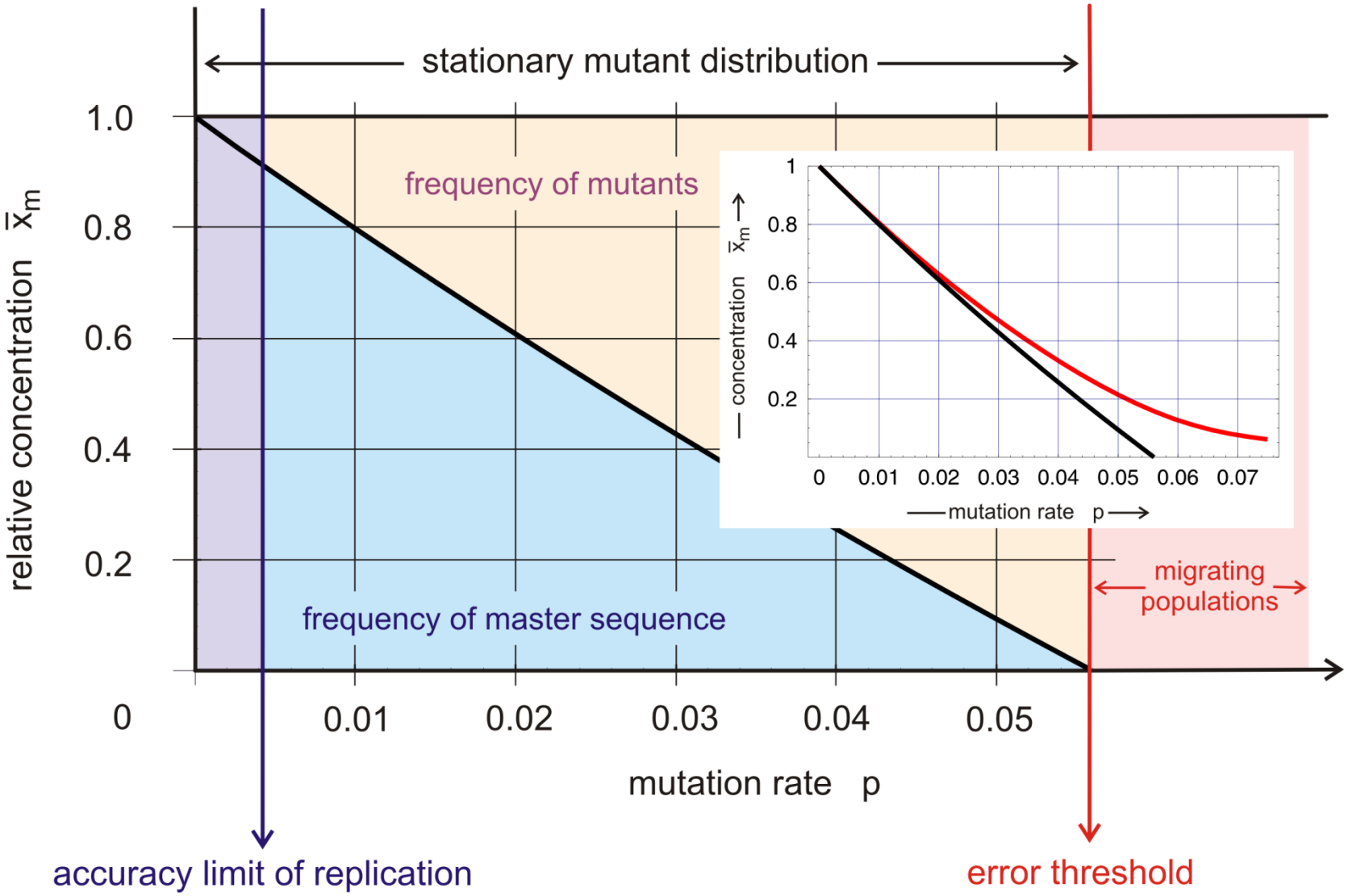
no mutational backflow


stationary mutant distribution
1.0
0.8
0.6
0.4
0.2
0
relative concentration $\bar{x}_m$
frequency of mutants
frequency of master sequence
0.01
0.02
0.03
0.04
0.05
mutation rate p


stationary mutant distribution
1.0
0.8
0.6
0.4
0.2
0
relative concentration $\bar{x}_m$
frequency of mutants
frequency of master sequence
0.01
0.02
0.03
0.04
0.05
mutation rate p
error threshold


stationary mutant distribution
1.0
0.8
0.6
0.4
0.2
0
relative concentration $\bar{x}_m$
frequency of mutants
frequency of master sequence
migrating populations
0.01
0.02
0.03
0.04
0.05
mutation rate p
error threshold


stationary mutant distribution
1.0
0.8
0.6
0.4
0.2
0
relative concentration $\bar{x}_m$
frequency of mutants
frequency of master sequence
migrating populations
0.01
0.02
0.03
0.04
0.05
mutation rate p
accuracy limit of replication
error threshold


stationary mutant distribution
1.0
0.8
0.6
0.4
0.2
0
relative concentration $\bar{x}_m$
frequency of mutants
frequency of master sequence
migrating populations
0.01
0.02
0.03
0.04
0.05
mutation rate p
accuracy limit of replication
error threshold
concentration $\bar{x}_m$
1
0.8
0.6
0.4
0.2
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
mutation rate p

concentrations of entire mutant classes

$$[\Gamma_k] = y_k(p)\,,\ k = 1, 2, \ldots, m$$

$$y_k(p) = \sum_{i=1, d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_m)=k}^{n} x_i(p)\,,\ |\Gamma_k| = \binom{n}{k}$$

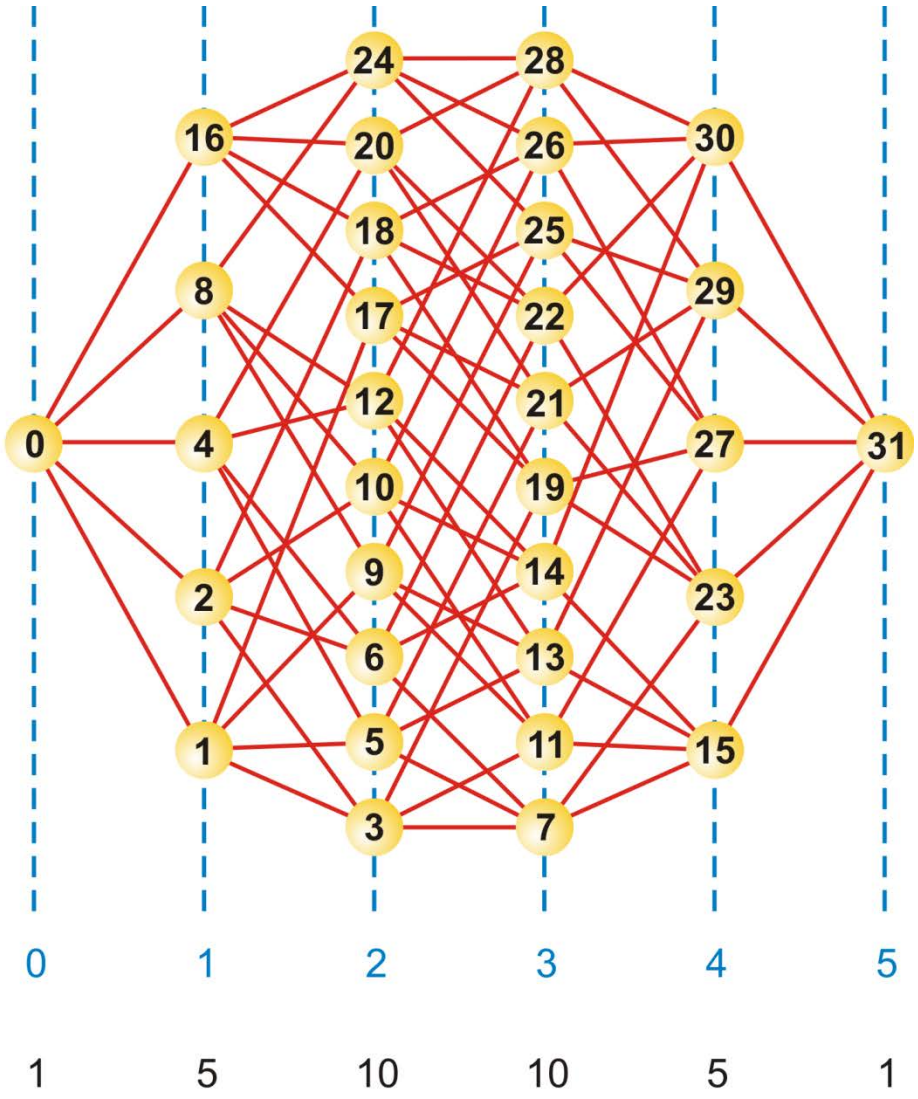
binary sequences are encoded
by their decimal equivalents:

C = **0** and G = **1**, for example,

"0" ≡ **00000** = CCCCC,

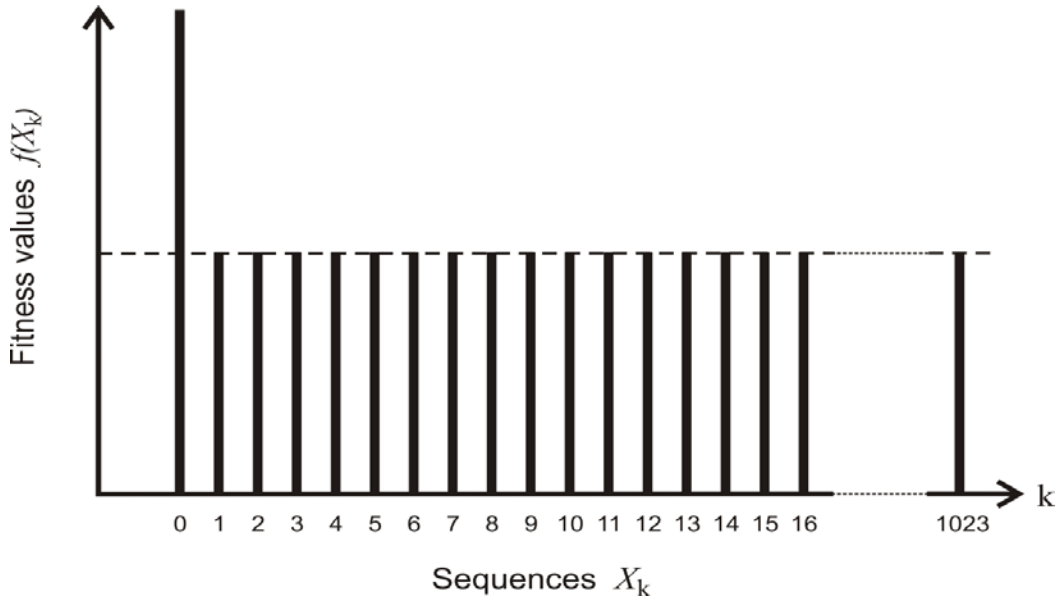
"14" ≡ **01110** = CGGGC,

"29" ≡ **11101** = GGGCG, etc.

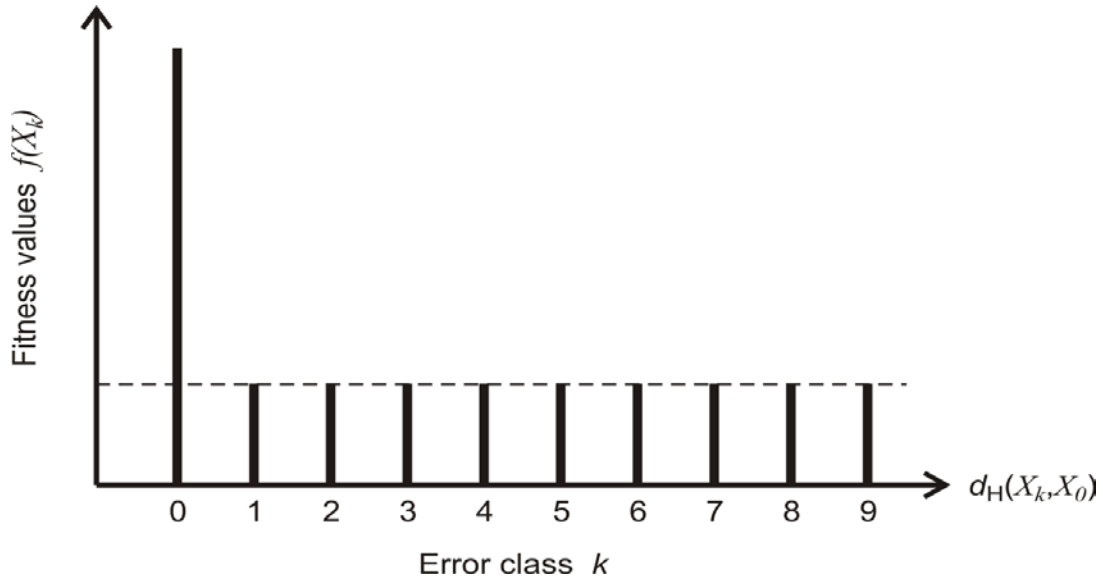


sequence space of dimension m = 5

single peak landscape



single peak landscape



single peak fitness landscape

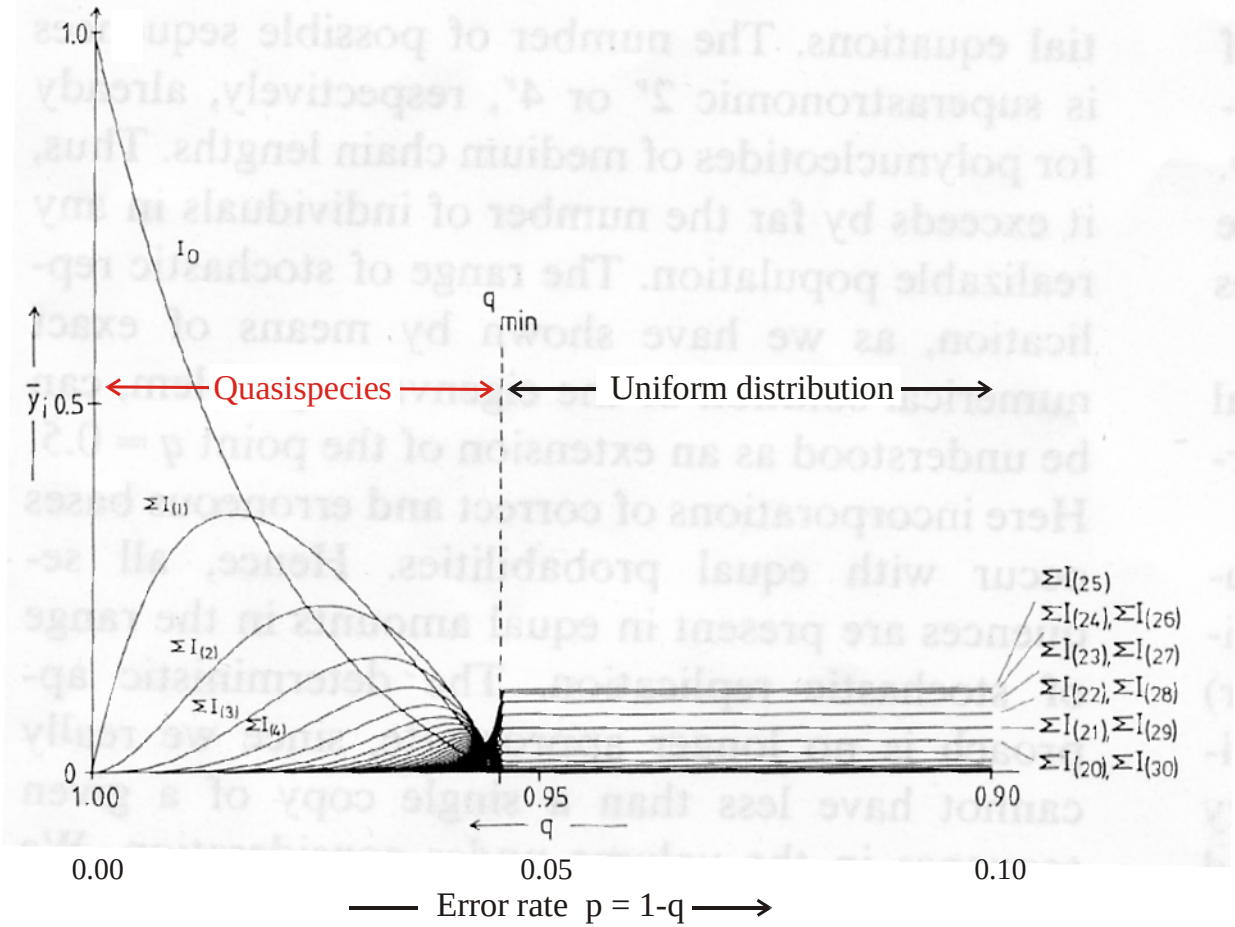
Biophysical Chemistry 16 (1982) 329–345
Elsevier Biomedical Press

SELF-REPLICATION WITH ERRORS

A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *

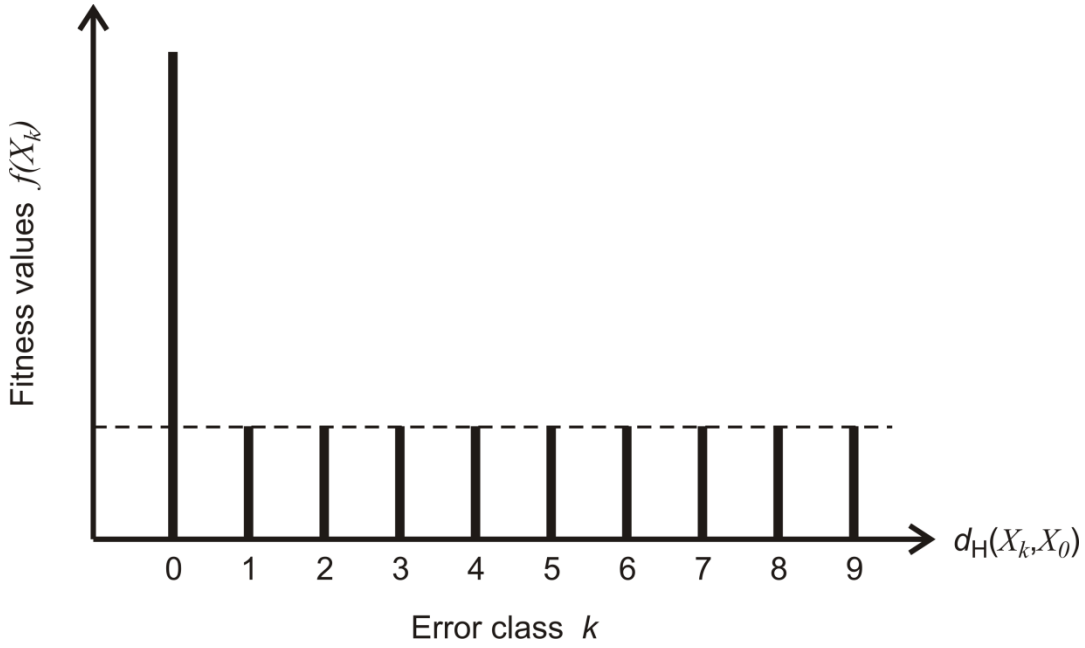
Institut für Theoretische Chemie und Strahlenchemie der Universität, Währingerstraße 17, A-1090 Wien, Austria



stationary population or quasispecies as a function of the mutation or error rate *p*

1. What is a „quasispecies"?
2. Detection of the „error threshold"
3. **Error thresholds on „simple landscapes"**
4. Error thresholds and phase transitions
5. „Realistic" landscapes
6. Neutrality in evolution

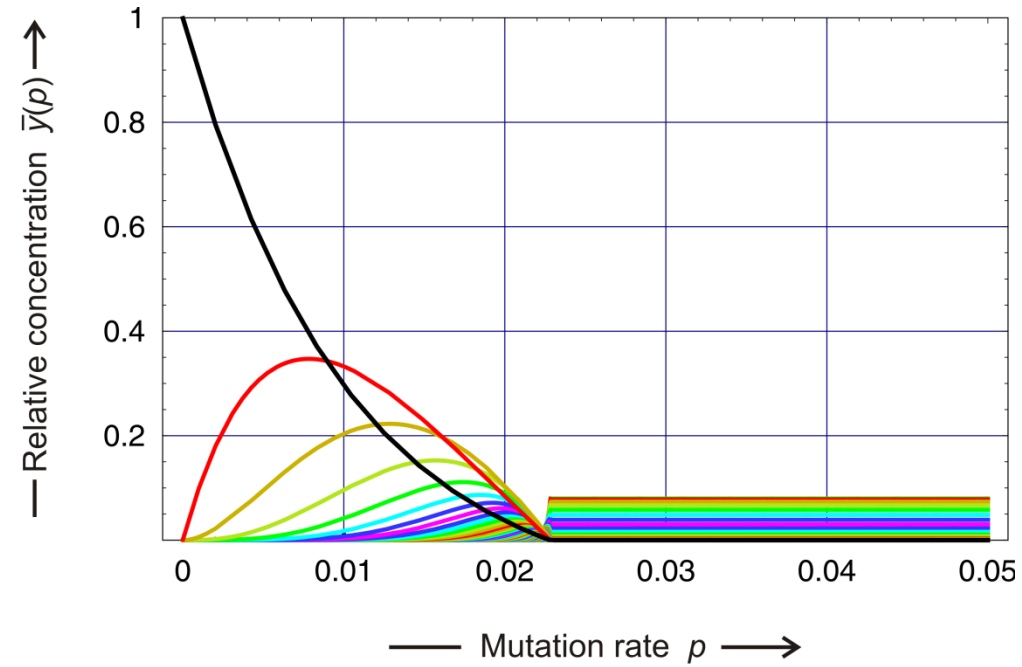
single peak landscape

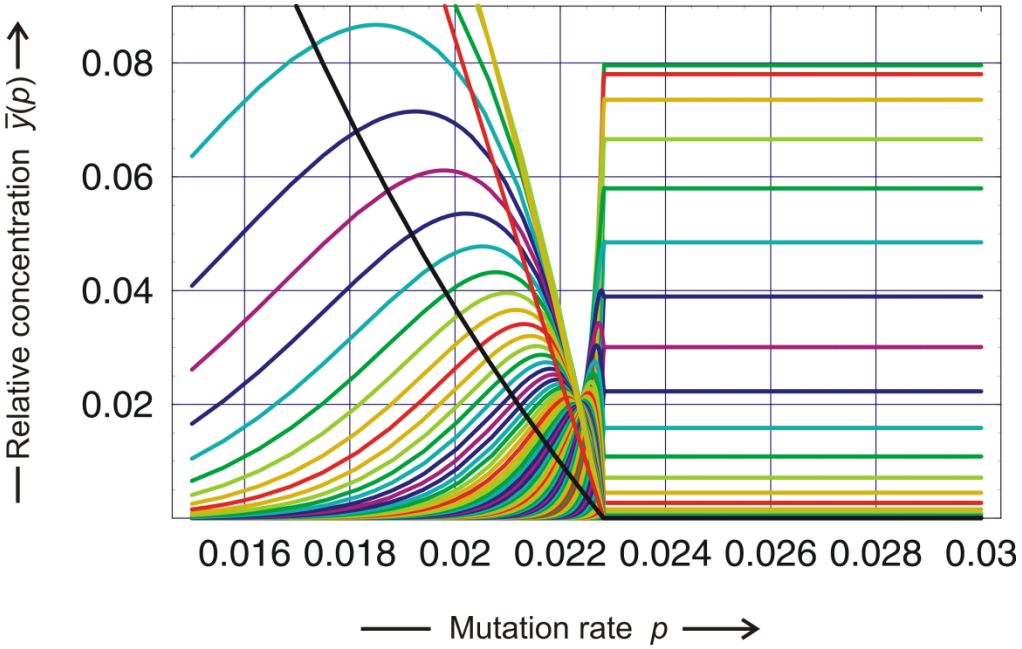


step linear landscape

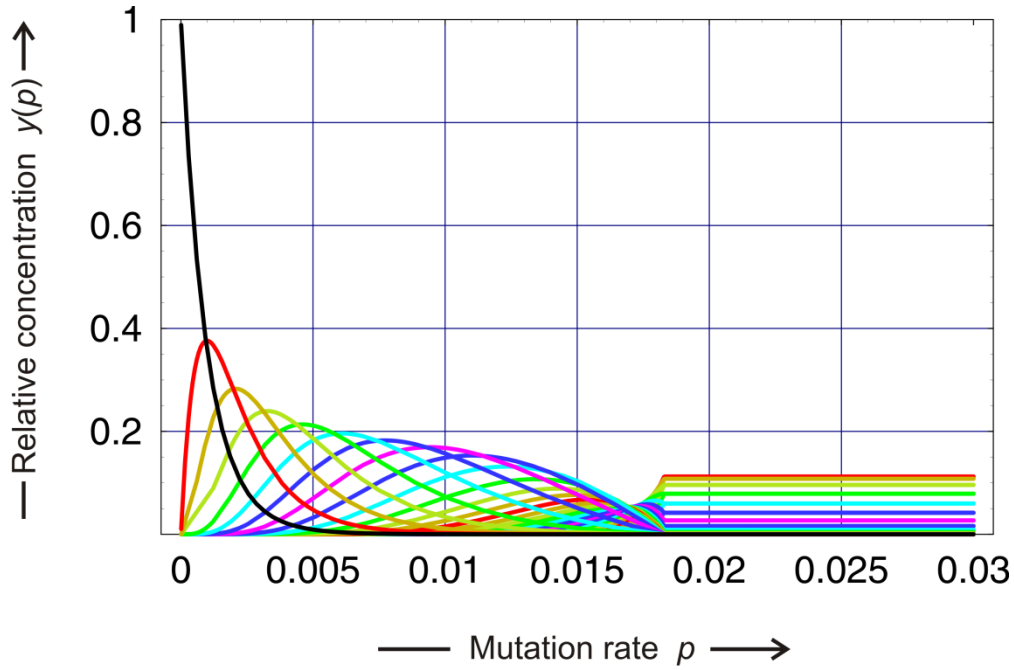
model fitness landscapes I

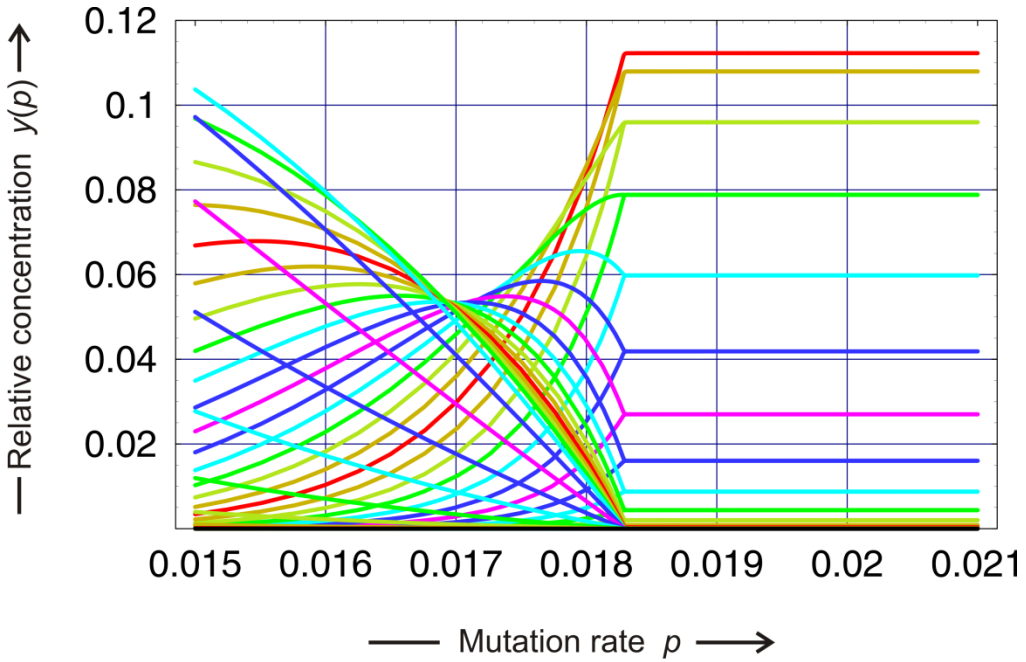
error threshold on the
single peak landscape





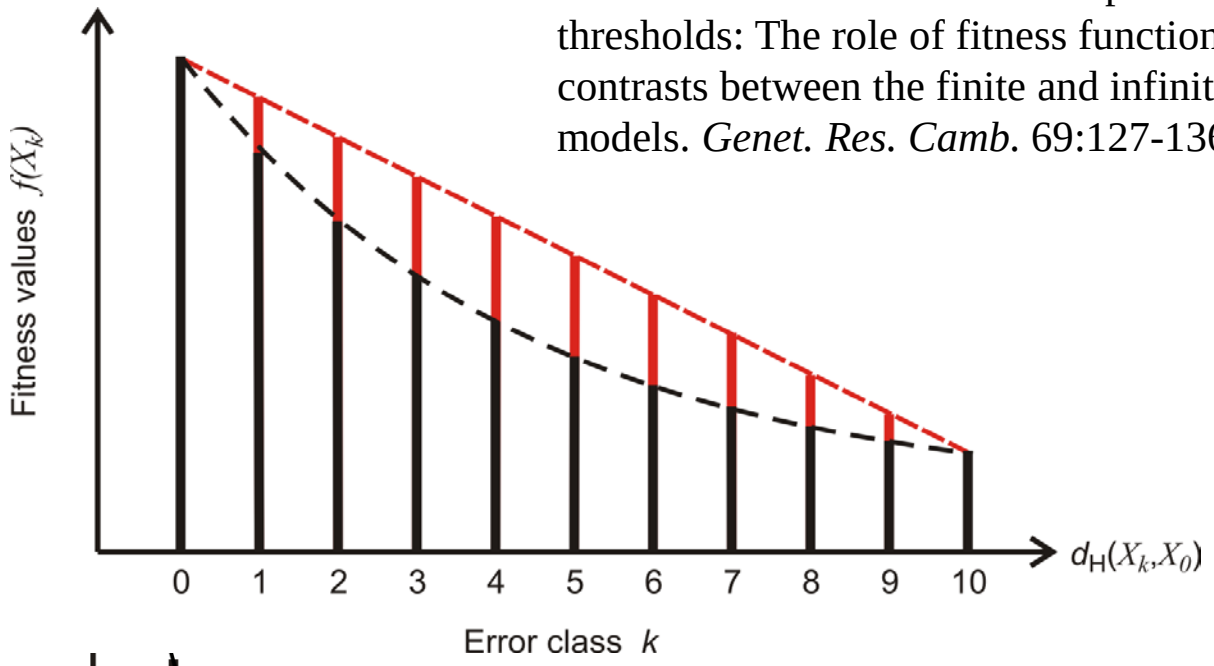
error threshold on the
step linear landscape



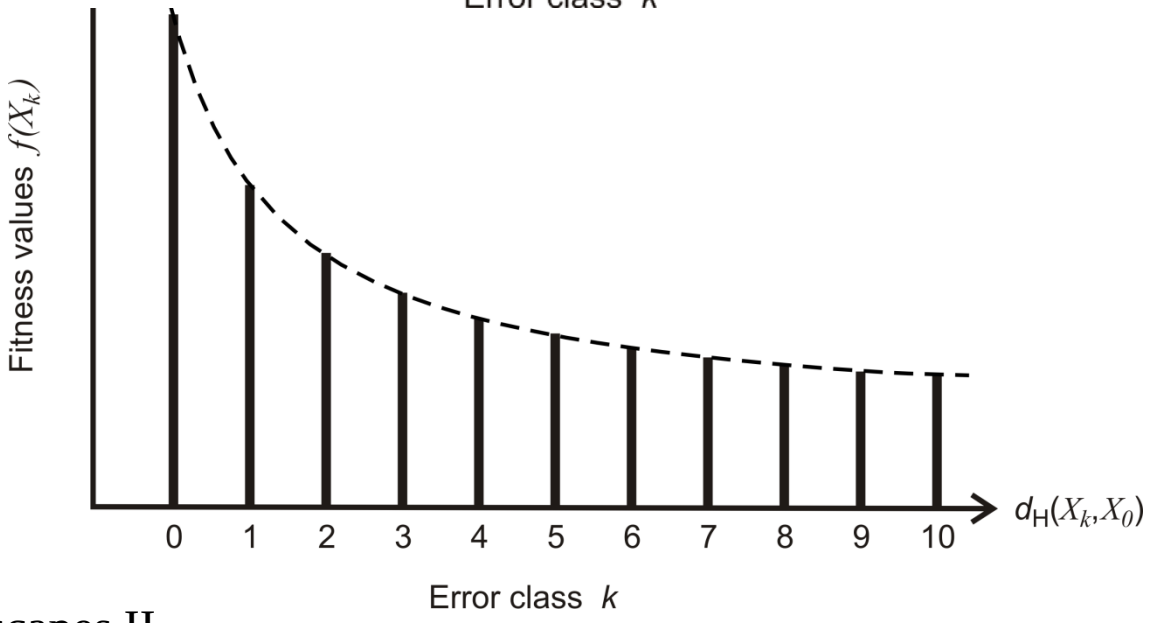


Thomas Wiehe. 1997. Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models. *Genet. Res. Camb.* 69:127-136

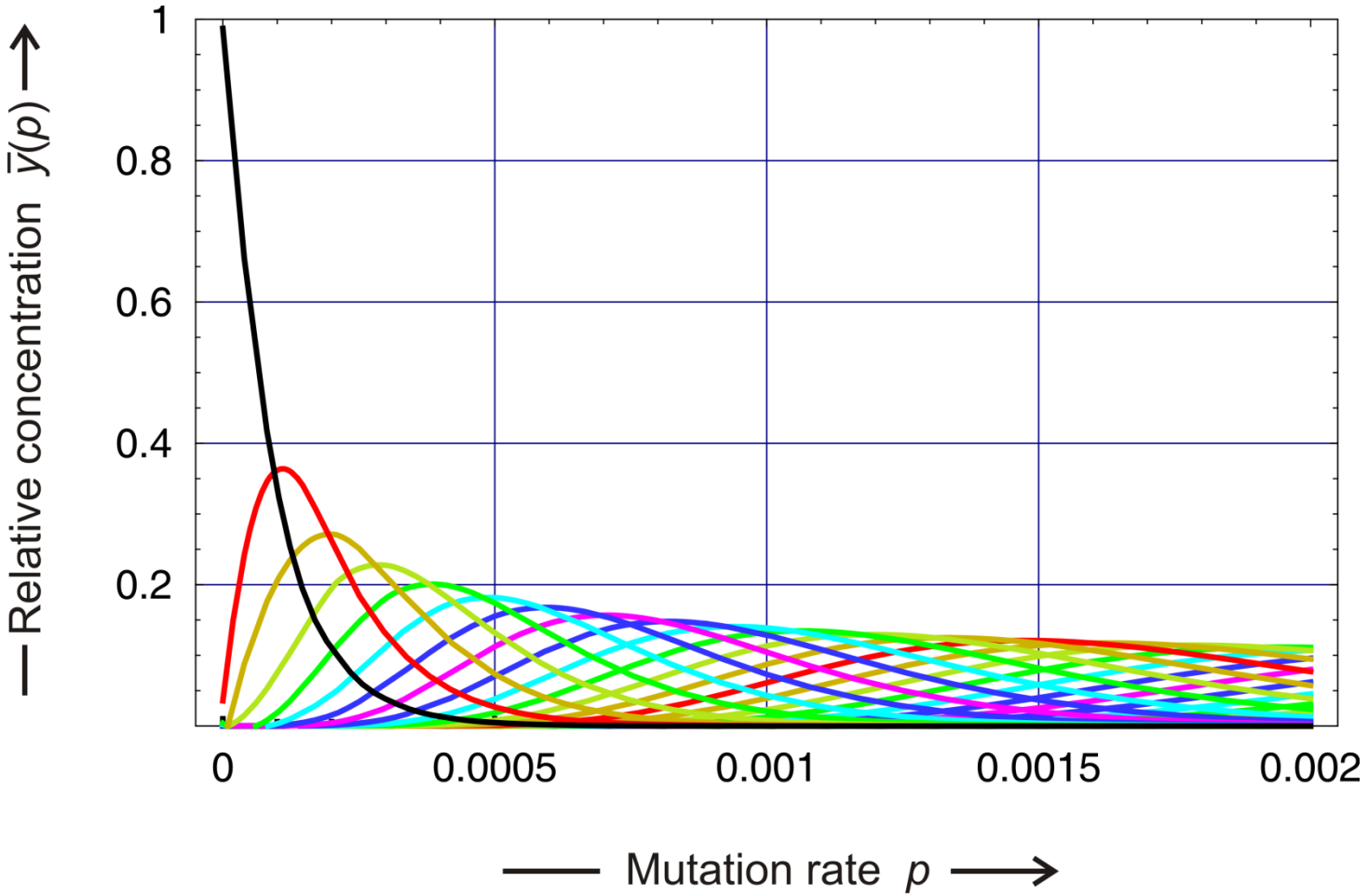
linear and multiplicative



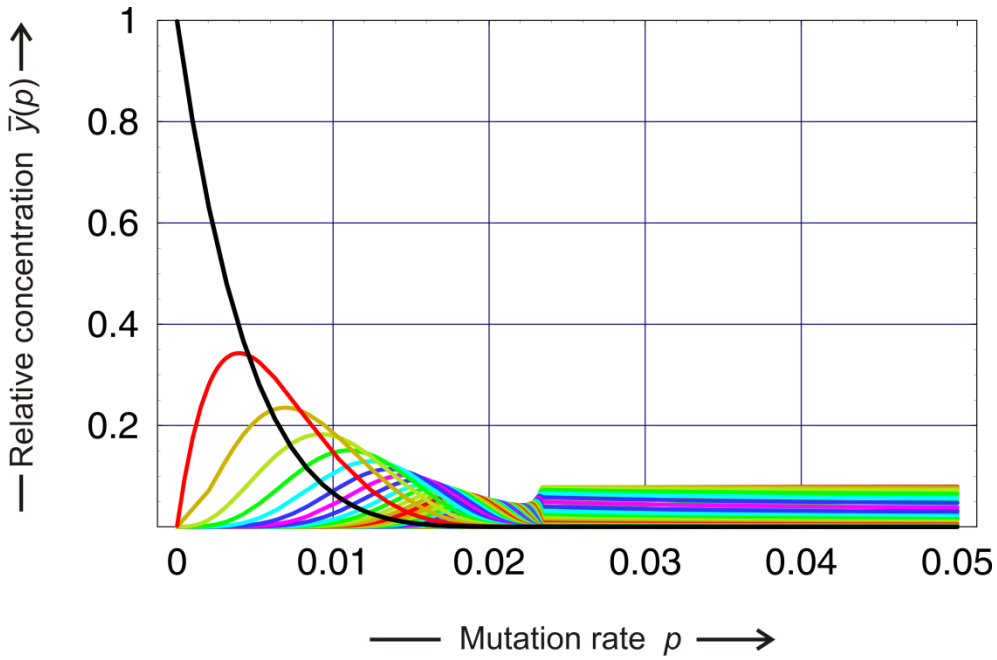
hyperbolic

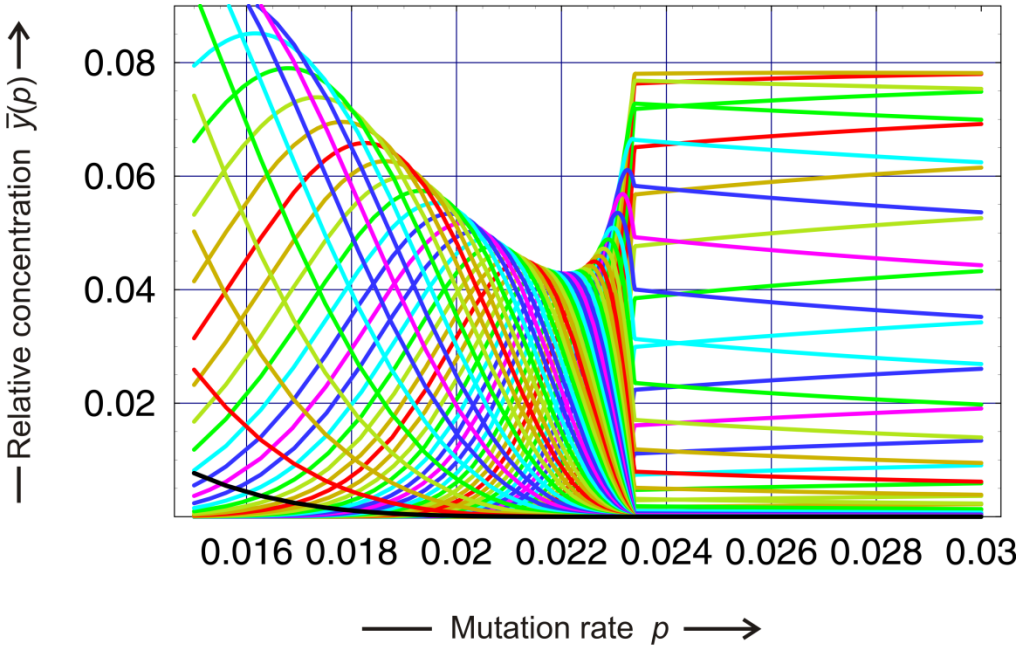


model fitness landscapes II



the linear fitness landscape shows no error threshold





error threshold on the hyperbolic landscape

The error threshold can be separated into three phenomena:

1. **Steep decrease** in the concentration of the master sequence to very small values.

2. **Sharp change** in the stationary concentration of the quasispecies distribuiton.

3. Transition to the **uniform distribution** at small mutation rates.

All three phenomena coincide for the quasispecies on the single peak fitness lanscape.

1. What is a „quasispecies"?
2. Detection of the „error threshold"
3. Error thresholds on „simple landscapes"
4. **Error thresholds and phase transitions**
5. „Realistic" landscapes
6. Neutrality in evolution

PHYSICAL REVIEW A VOLUME 45, NUMBER 8 15 APRIL 1992

Error thresholds for molecular quasispecies as phase transitions: From simple landscapes to spin-glass models

P. Tarazona
*Institut für Theoretische Chemie der Universität Wien, A-1090 Wien, Austria and Departamento de Fisica de la Materia Condensada, Universidad Autonoma de Madrid, E-28049, Madrid, Spain**
(Received 19 June 1991)

The correspondence between Eigen's model [Naturwissenschaften **58**, 465 (1971)] for molecular quasispecies and the equilibrium properties of a lattice system proposed by Leuthäusser [J. Chem. Phys. **84**, 1884 (1986); J. Stat. Phys. **48**, 343 (1987)] is used to characterize the error thresholds for the existence of quasispecies as phase transitions. For simple replication landscapes the error threshold is related to a first-order phase transition smoothed by the complete wetting of the time surface. Replication landscapes based on the Hopfield Hamiltonian for neural networks allow for the tuning of the landscape complexity and reveal the existence of two error thresholds, bracketing a region of spin-glass quasispecies between the simple quasispecies and the fully disordered mixture of sequences.

PACS number(s): 87.10.+e, 64.60.Cn, 05.50.+q

Ira Leuthäusser. Statistical mechanics of Eigen's evolution model.
J. Statist. Phys. 48:343-360, 1987

Ricard V. Solé, Susanna C. Manrubia, Bartolo Luque, Jordi Delgado, Jordi Bascompte.
Phase transitions and complex systems.
Simple nonlinear models capture complex systems at the edge of chaos.
Complexity 1(1):13-26, 1996

$$\frac{\mathrm{d}x_j}{\mathrm{dt}} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

$$\Phi = \sum_{i=1}^{n} f_i\, x_i \Big/ \sum_{i=1}^{n} x_i$$

$$\frac{\mathrm{d}\mathbf{X}}{\mathrm{dt}} = \left(\mathrm{W} - \Phi \cdot \mathbf{1}\right)\mathbf{X}; \quad \Phi = \sum_{i=1}^{n} f_i\, x_i \Big/ \sum_{i=1}^{n} x_i\,; \quad \mathbf{X} = \left(x_1, x_2, \ldots, x_n\right)^{\mathrm{t}}$$

$$\mathbf{X}_n = \mathrm{W}^n \cdot \mathbf{X}_0\,; \mathbf{X}_i = \left(x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\right)^{\mathrm{t}}; x_k = \left[S_k\right]$$

$$S_k = \left(s_1 \bullet s_2 \bullet \cdots \bullet s_\ell\right)\,; \; s_i = \left\{\pm 1\right\}$$

replication-mutation dynamics and spin lattices

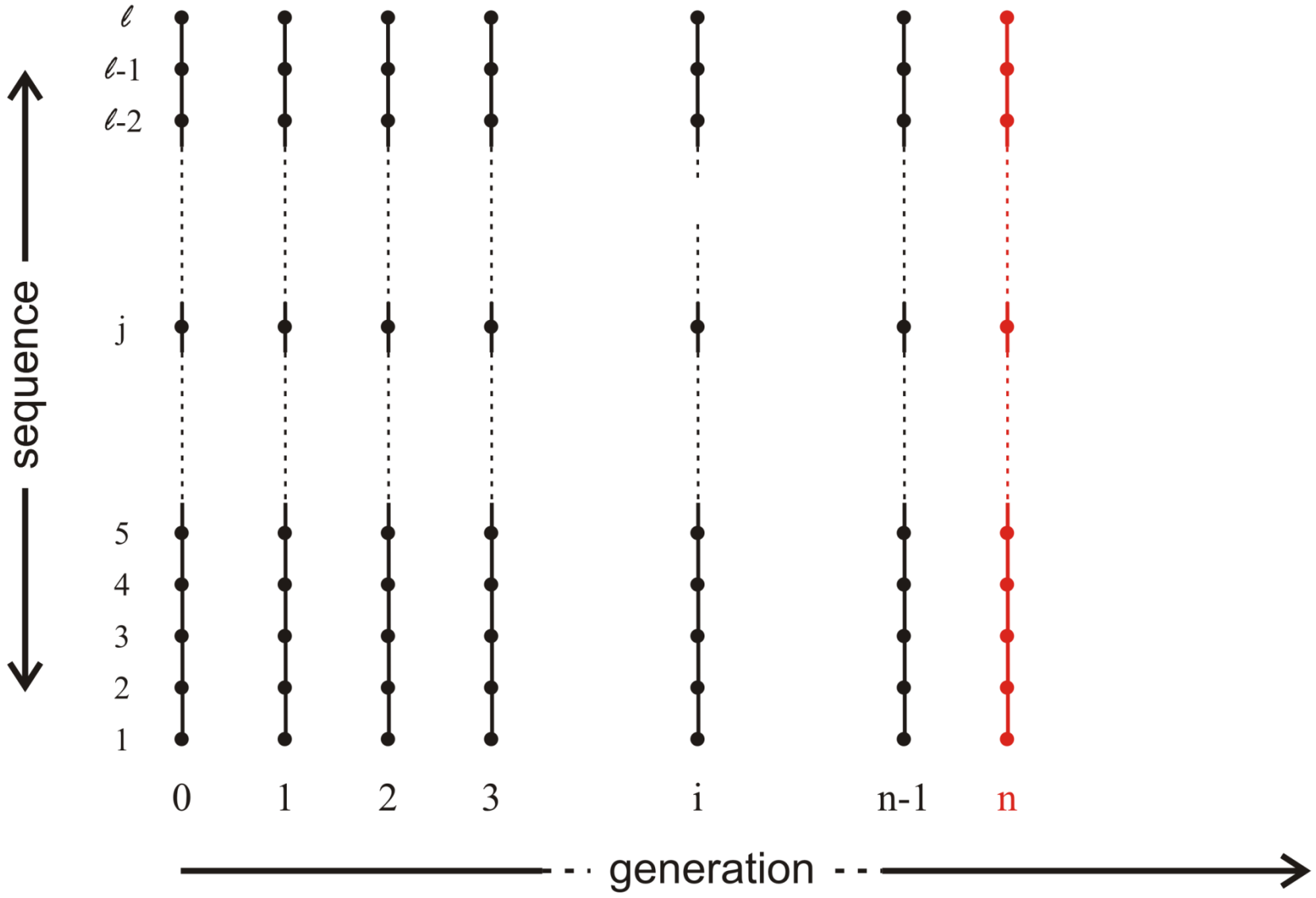
$$W_{ji} = e^{-\beta h(S_j, S_i)} \text{ with } \beta = 1/k_B T$$

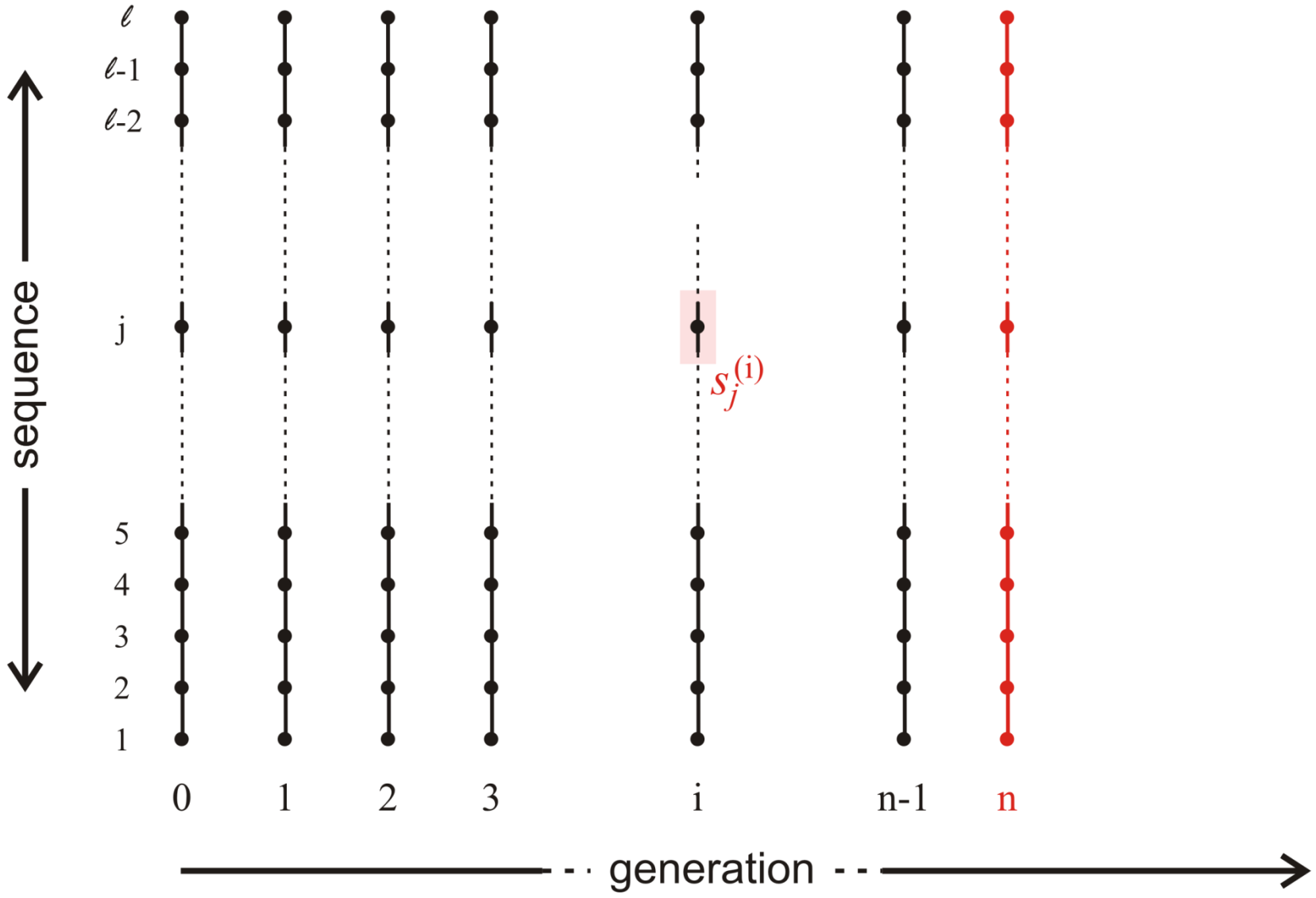
$$W_{ji} = Q_{ji}\ f_i\ =\ q^{\ell}\ \varepsilon^{d_H(S_j, S_i)}\ f_i\ ;\ \varepsilon = \frac{p}{1-p};\ d_H(S_j, S_i) = \frac{1}{2}\left(\ \ell - \sum_{k=1}^{\ell} s_k^{(j)} s_k^{(i)}\right)$$

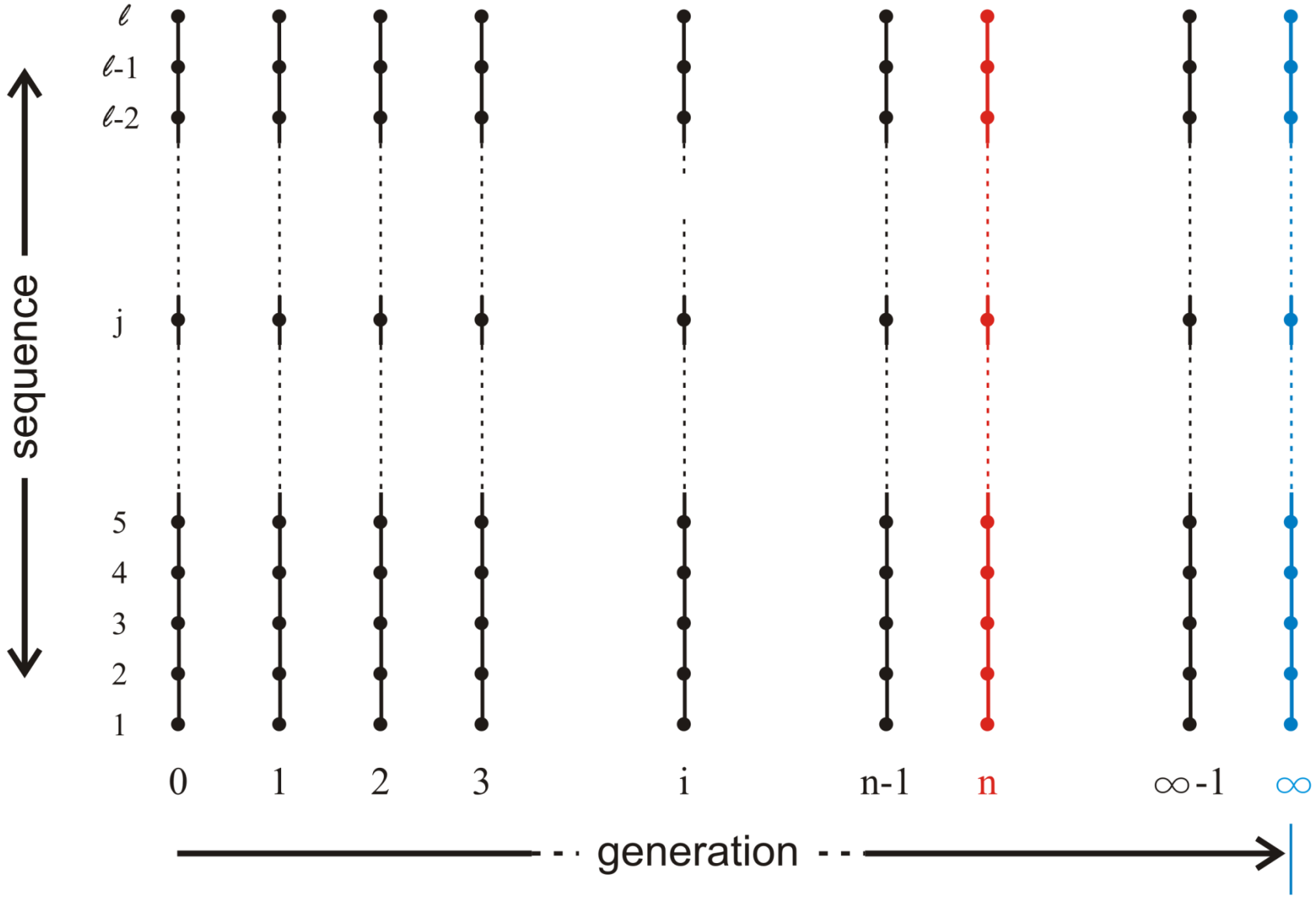
$$-\beta\ H = \sum_{i=0}^{n-1}\left(\ \beta \sum_{k=1}^{\ell} s_k^{(i)}\ s_k^{(i+1)} + \ln\ f_i\ \right)\ + \frac{n\,\ell}{2} \ln\left(p\,(1-p)\right)$$

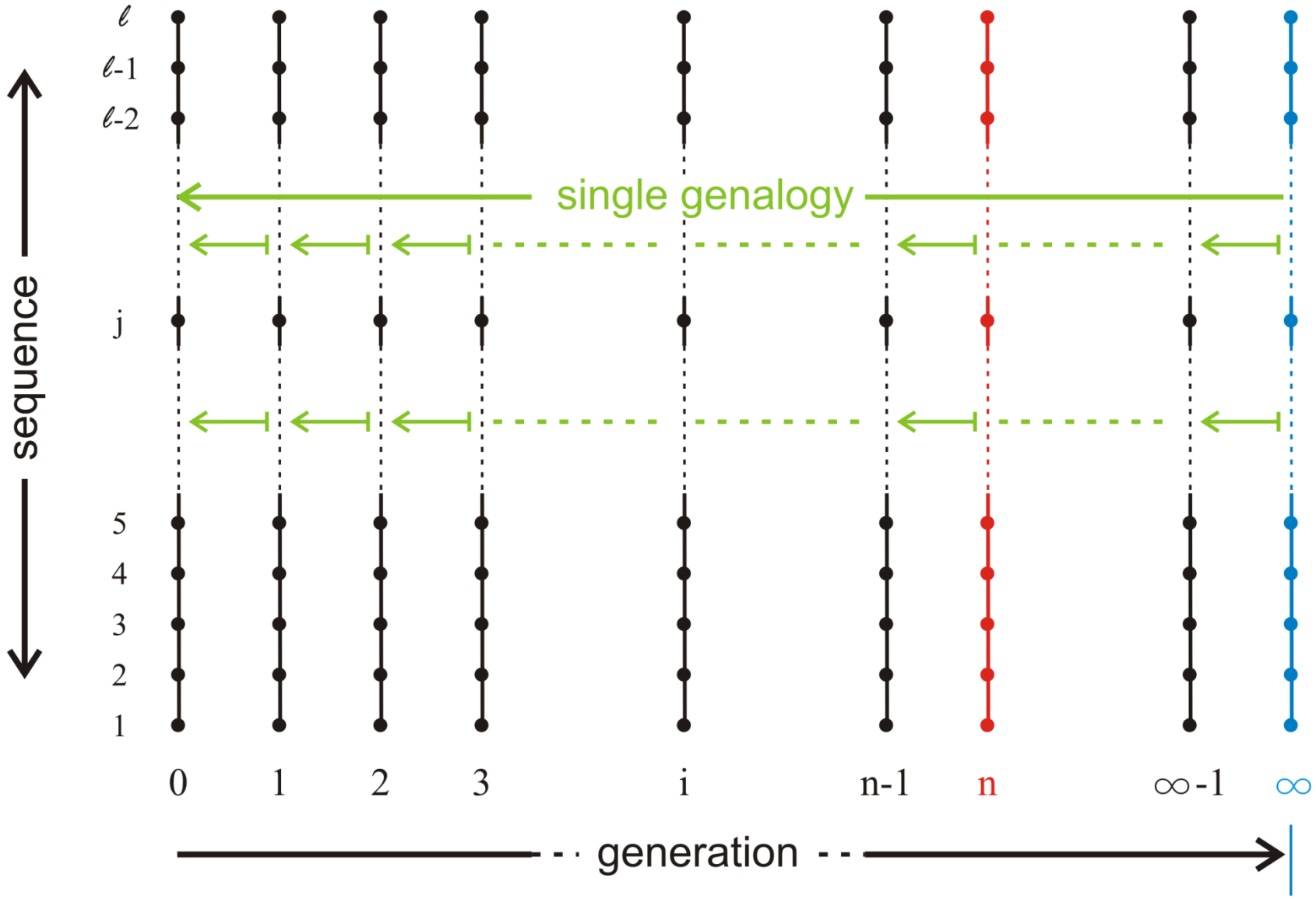
$$\text{temperature}: \quad \mathrm{T}^{-1} = k_B \ln \sqrt{p(1-p)}$$

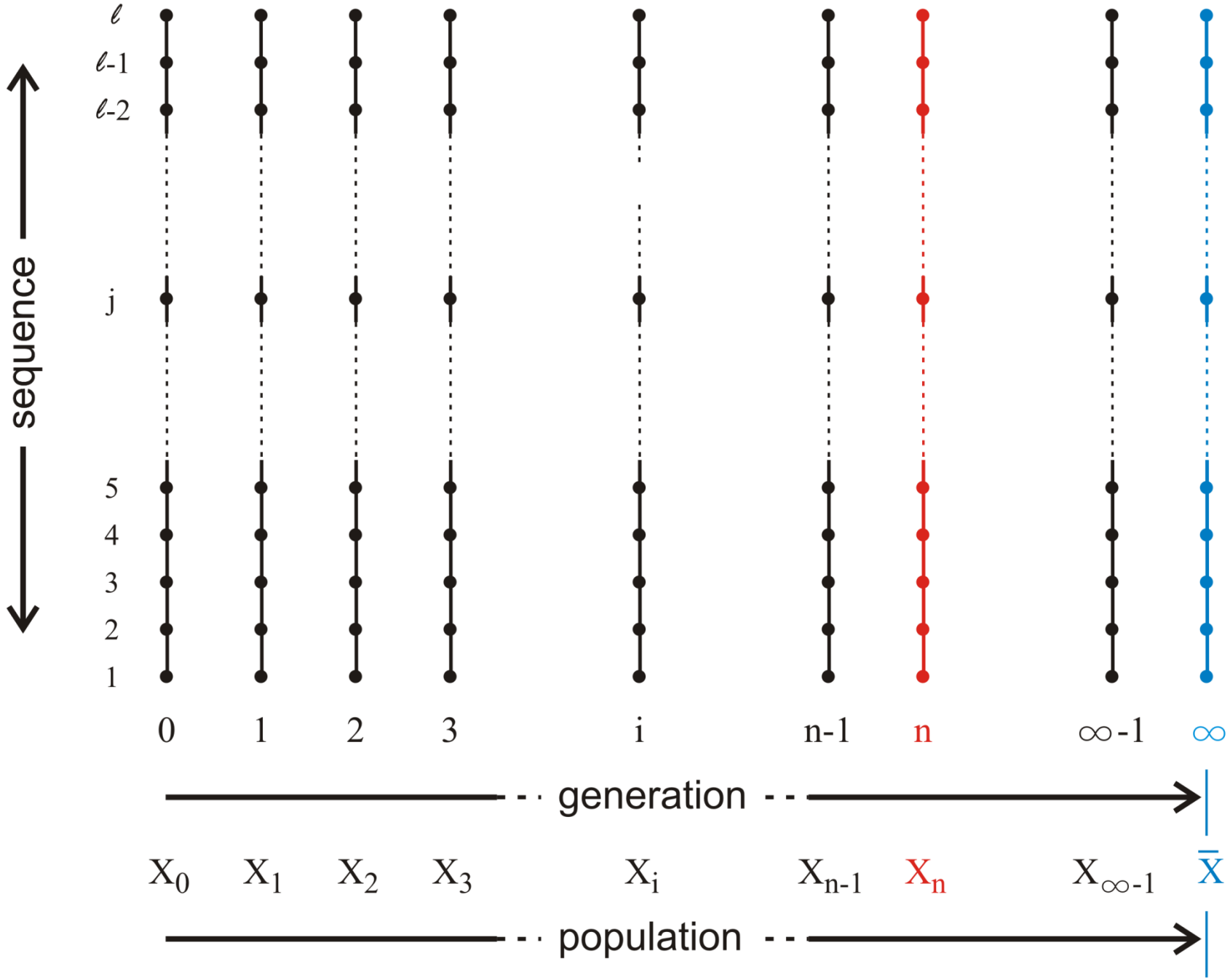
$$p = 0,\ p = 1\text{:}\quad \mathrm{T} \to -\infty \quad \text{and} \quad p = \frac{1}{2}\text{:}\quad T \to \max$$

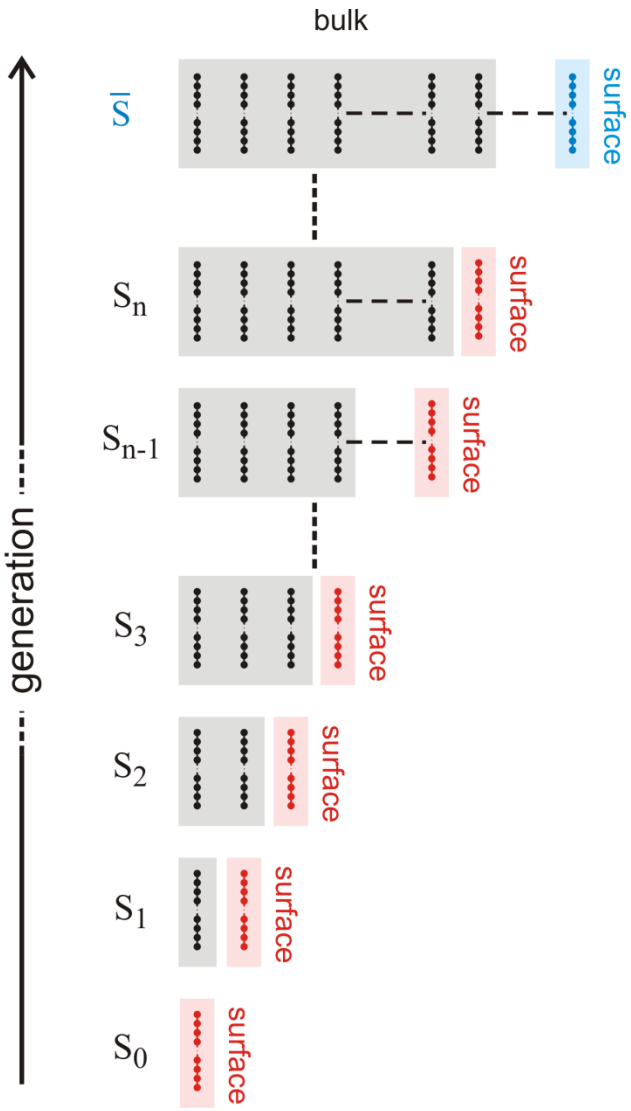

ℓ
ℓ-1
ℓ-2
j
5
4
3
2
1
sequence
0
1
2
3
i
n-1
n
generation


ℓ
ℓ-1
ℓ-2
j
5
4
3
2
1
sequence
$s_j^{(i)}$
0
1
2
3
i
n-1
n
generation


ℓ
ℓ-1
ℓ-2
j
5
4
3
2
1
sequence
0
1
2
3
i
n-1
n
∞-1
∞
generation


ℓ
ℓ-1
ℓ-2
single genalogy
j
5
4
3
2
1
sequence
0
1
2
3
i
n-1
n
∞-1
∞
generation


sequence
ℓ
ℓ-1
ℓ-2
j
5
4
3
2
1
0
1
2
3
i
n-1
n
∞-1
∞
generation
X0
X1
X2
X3
Xi
Xn-1
Xn
X∞-1
X̄
population



only the surface layer is relevant
for evolutionary dynamics

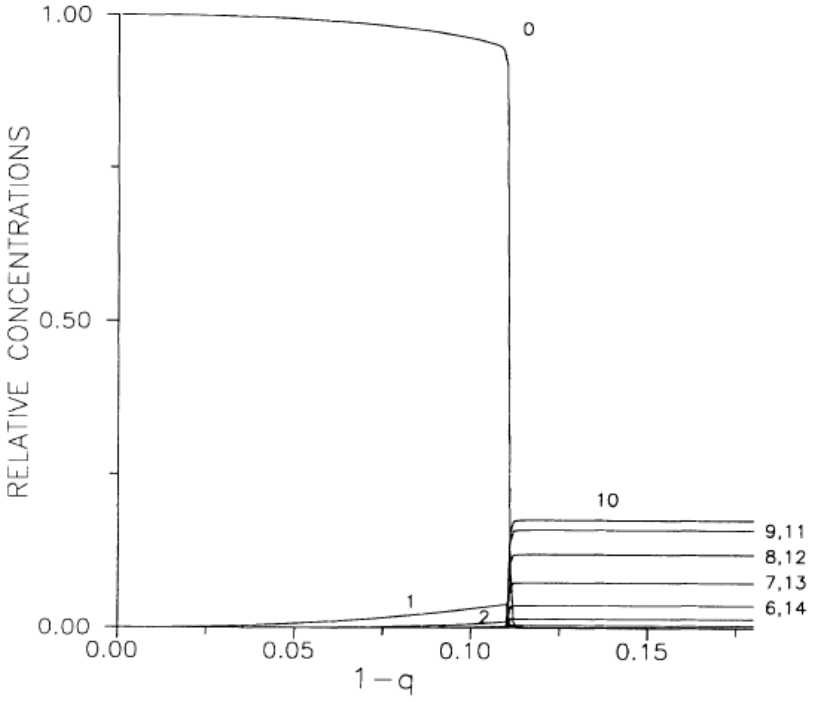


FIG. 2. Relative concentration of the sequences at given Hamming distance from the master sequence, as a function of the error rate $1-q$. The data correspond to the bulk distribution for a sequence with $N=20$ and a single-peak landscape, as in Eq. (8), with $A_0/A_1=10$.

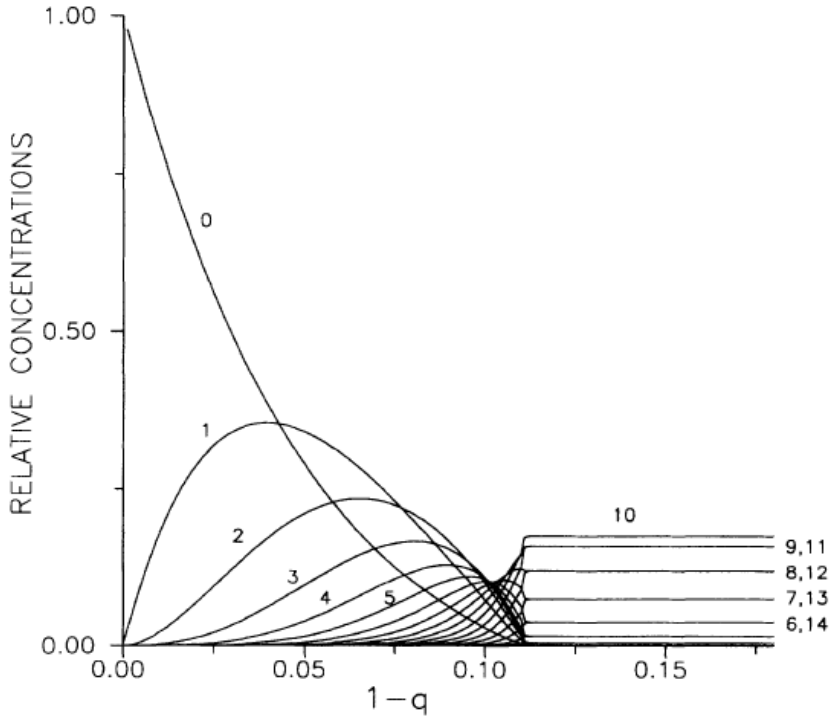


FIG. 3. Distribution of the steady-state population, given by the relative concentrations for the same system as in Fig. 2, but obtained from the surface layer.

Pedro Tarazona. Phys. Rev. A 45:6038-6050, 1992

order parameter

$$m = \frac{1}{\ell} \sum_{k=1}^{\ell} s_k^{(m)} \langle s_k \rangle$$

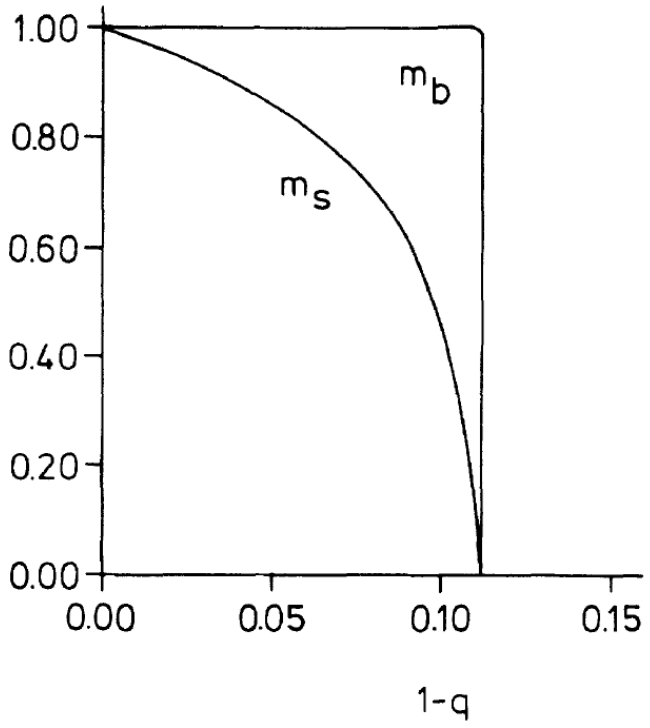
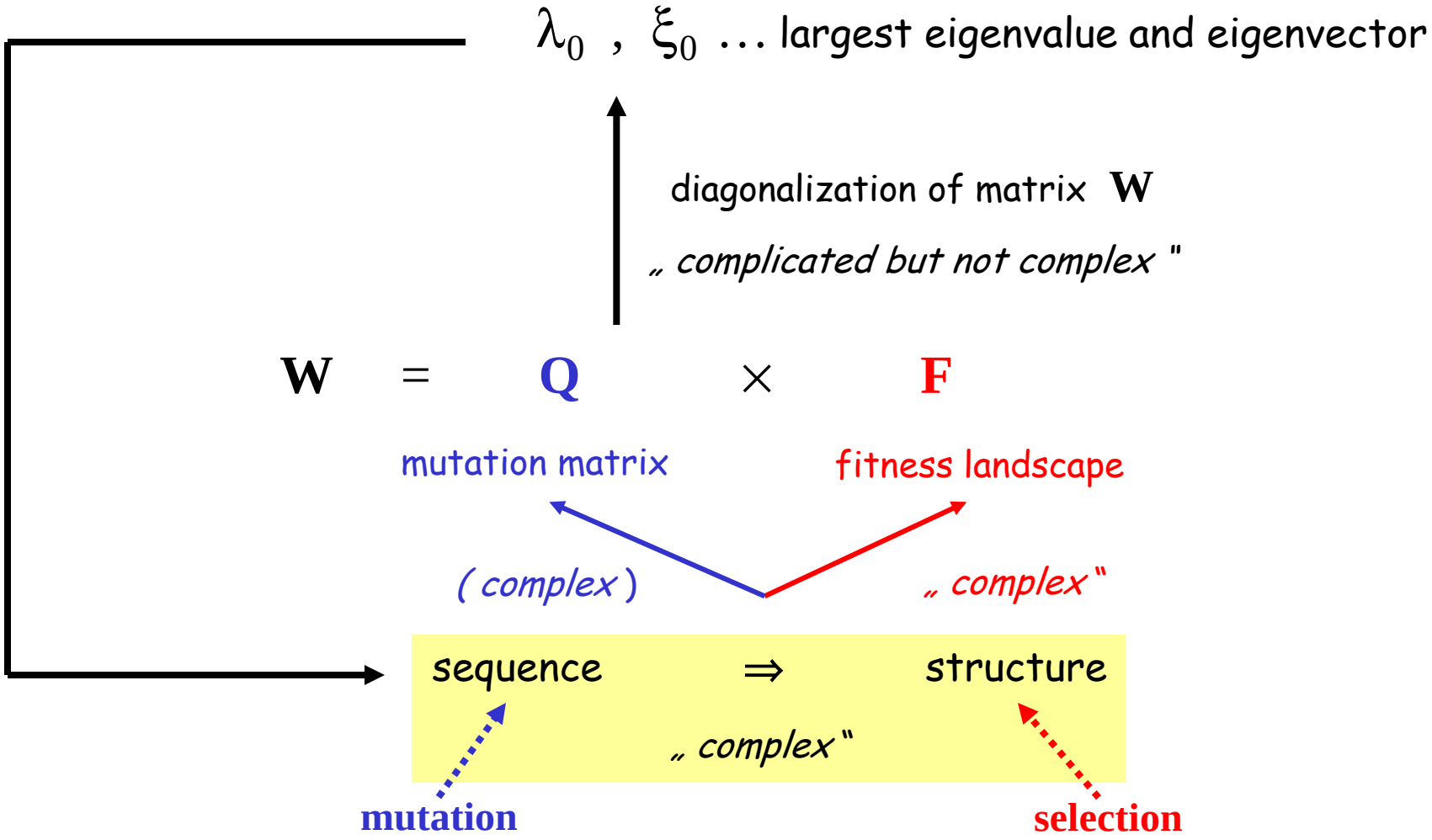


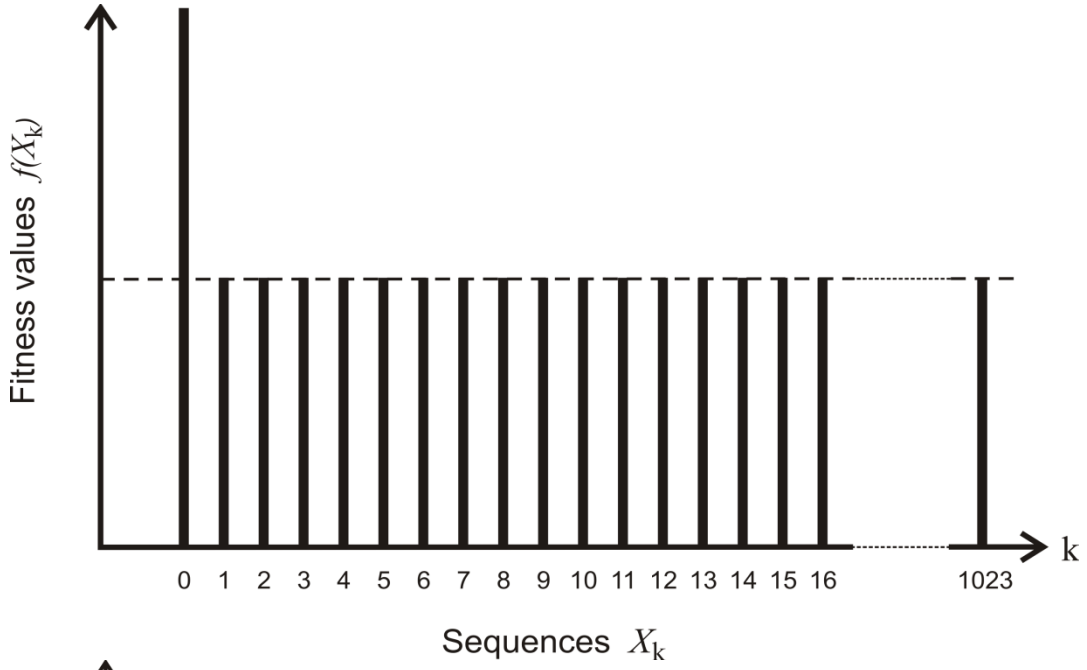
FIG. 4. The order parameter as defined in Eq. (10), for the bulk, m_b, and the surface, m_s, of the same system as in Figs. 2 and 3.

1. What is a „quasispecies"?
2. Detection of the „error threshold"
3. Error thresholds on „simple landscapes"
4. Error thresholds and phase transitions
5. **„Realistic" landscapes**
6. Neutrality in evolution



complexity in molecular evolution

single peak landscape

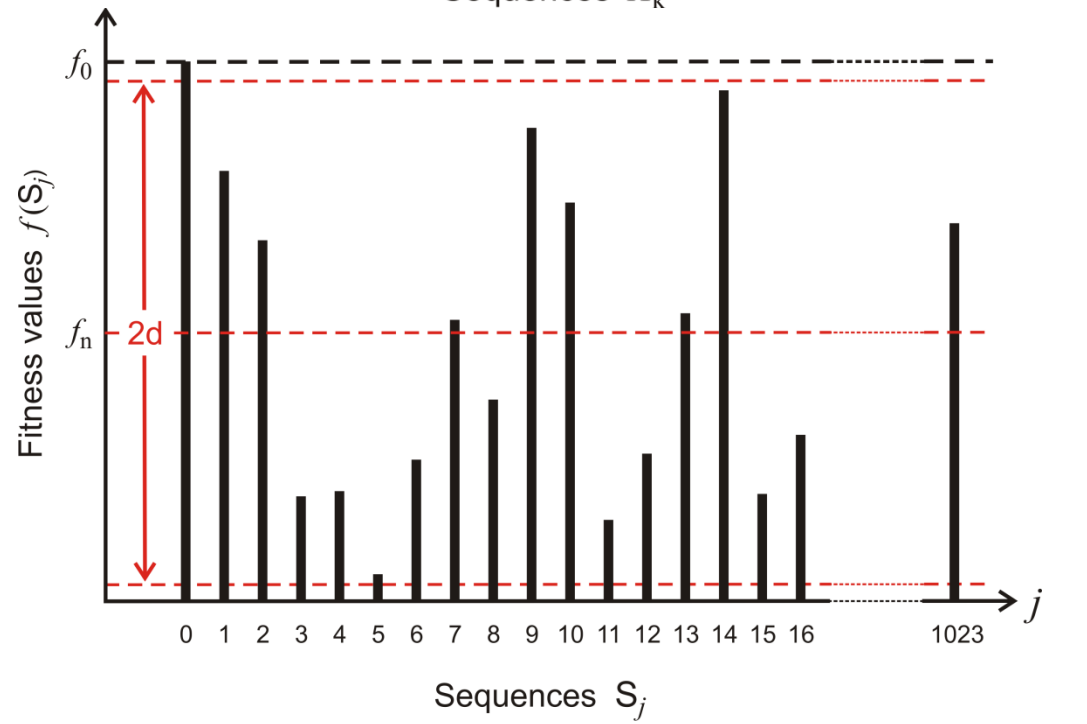


$$f(\mathsf{S}_j) = f_n + 2\,d\,(f_0 - f_n)\left(\eta_j^{(s)} - 0.5\right)$$

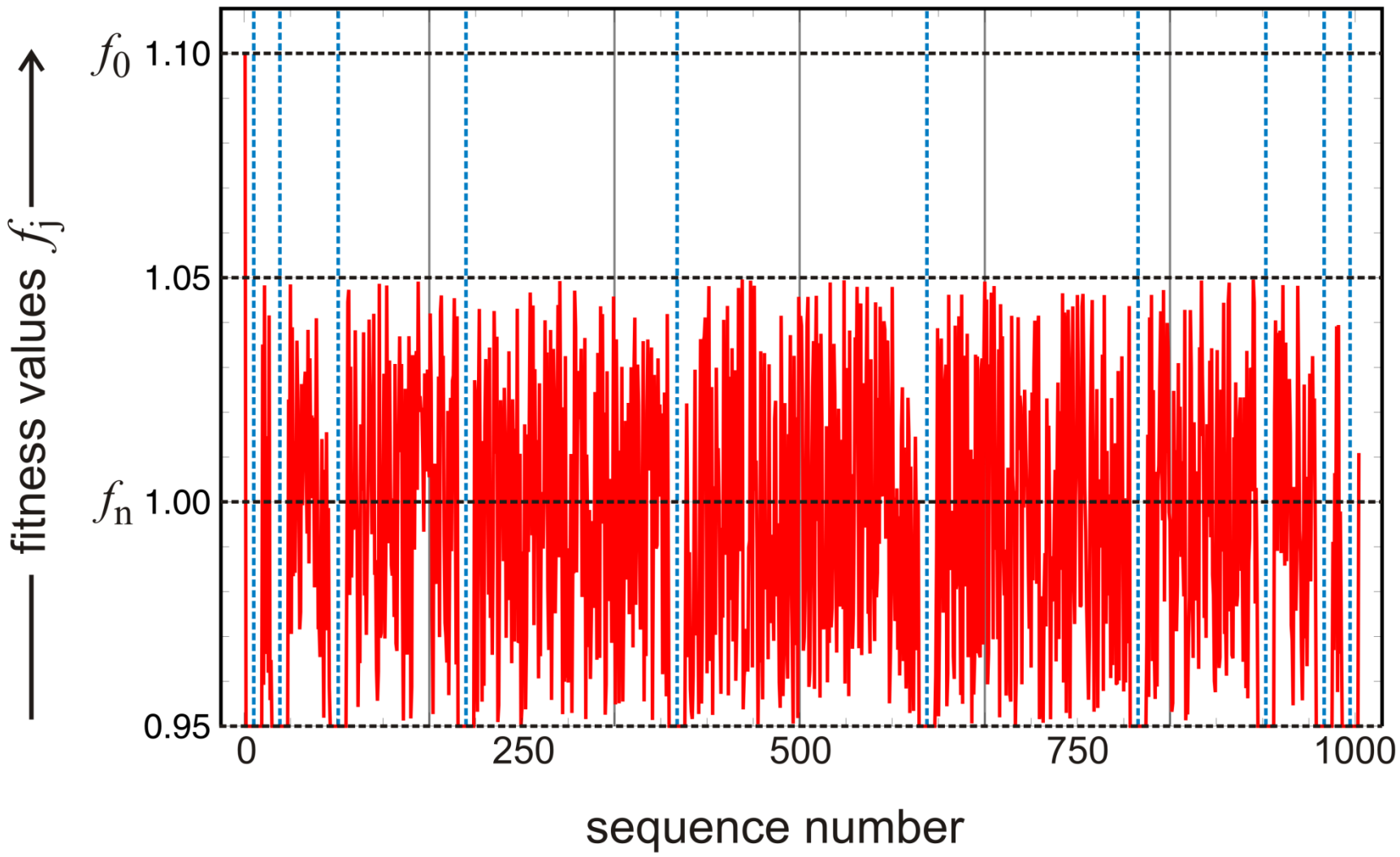
$$j = 1, 2, \ldots, N\,;\; j \neq m\,,$$

η ... random number ; s ... seeds

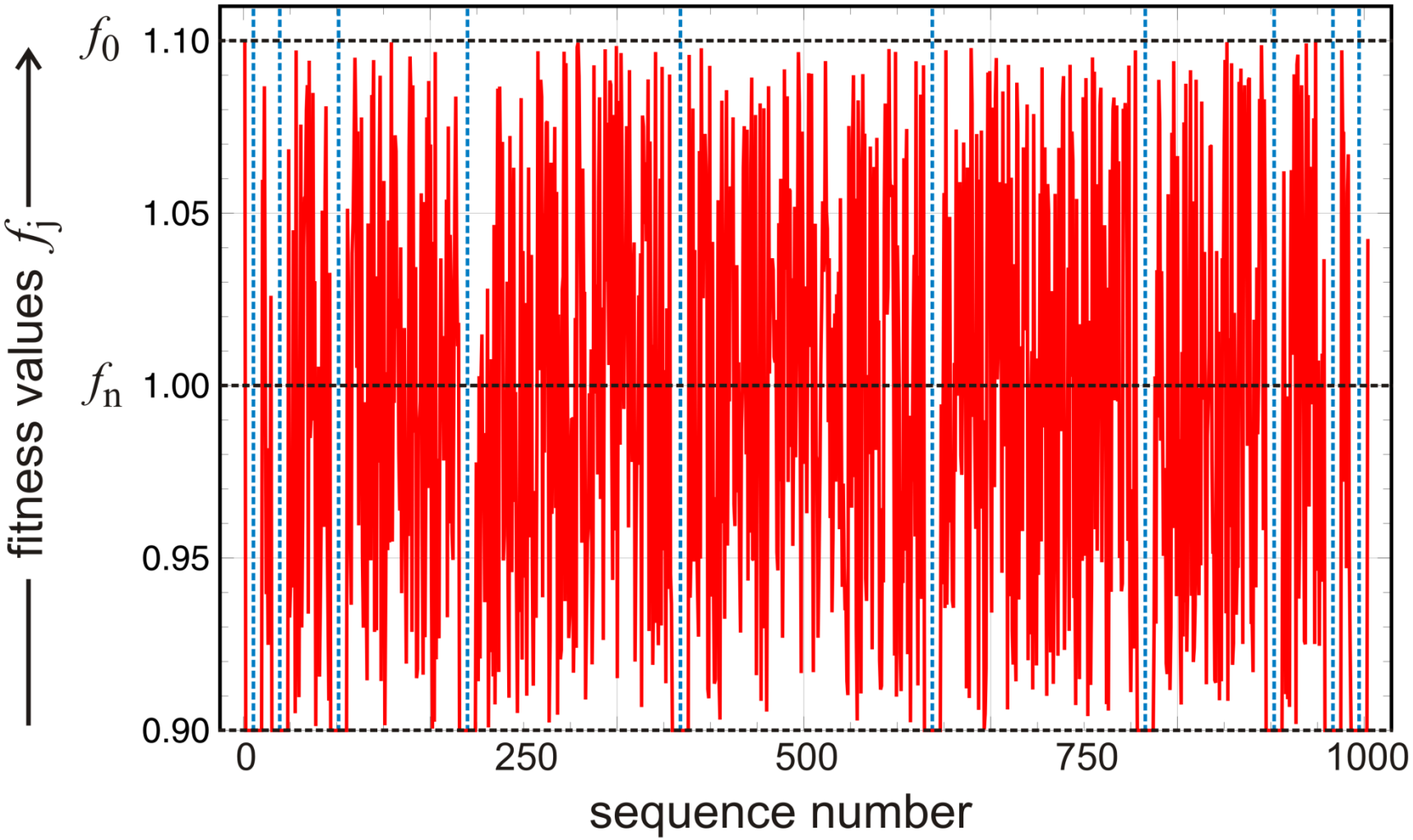
„realistic“ landscape



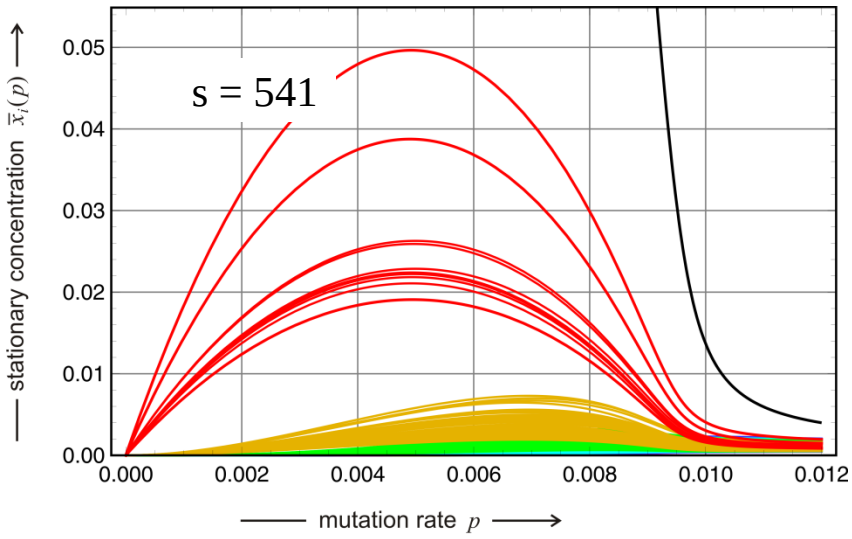
rugged fitness landscapes
over individual binary sequences
with n = 10

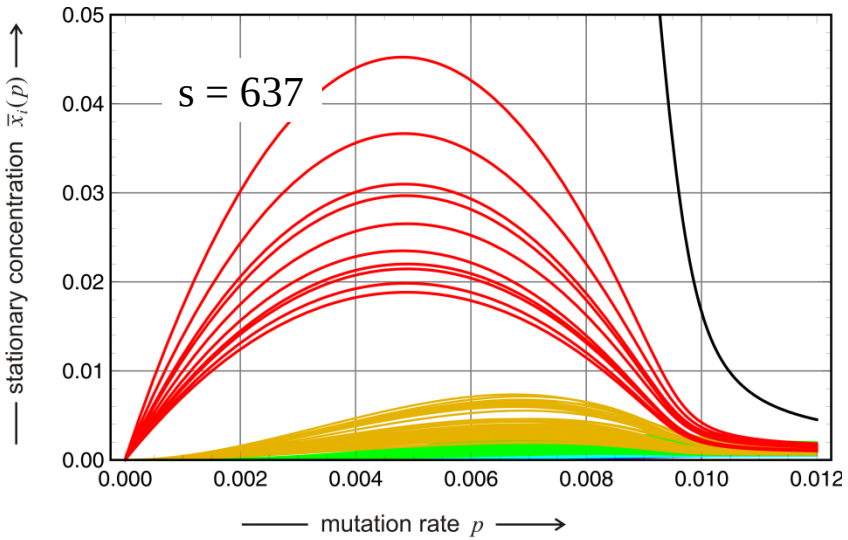


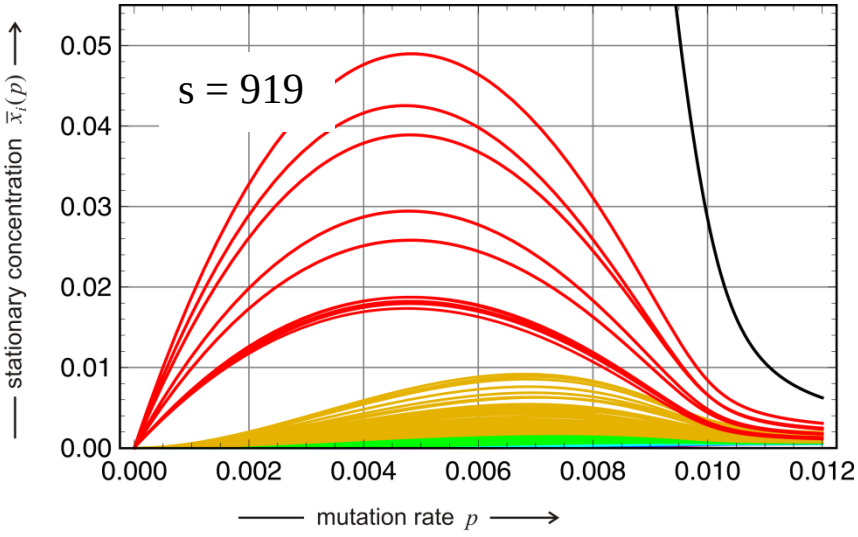
random distribution of fitness values: $d = 0.5$ and $s = 919$



random distribution of fitness values: d = 1.0 and s = 637

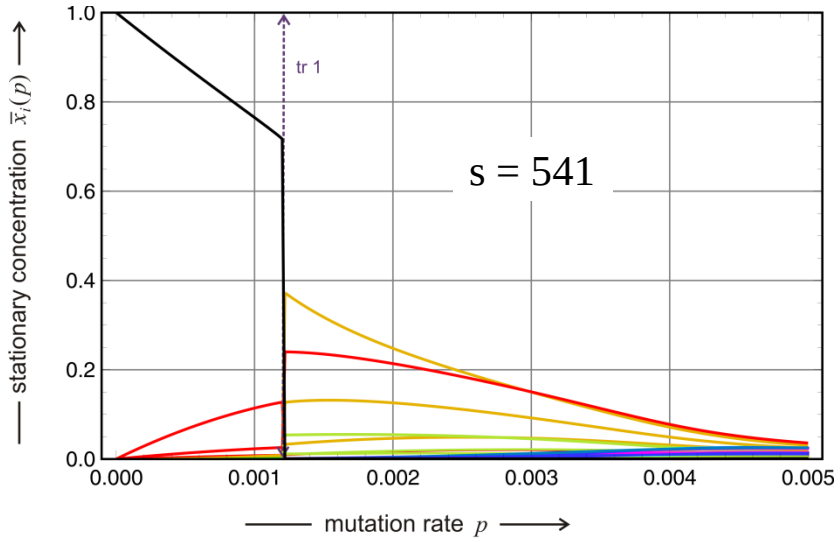


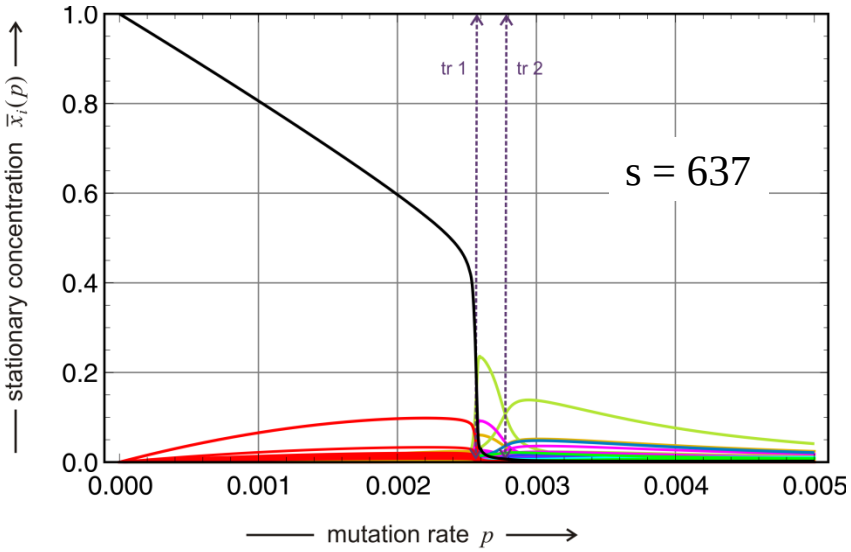


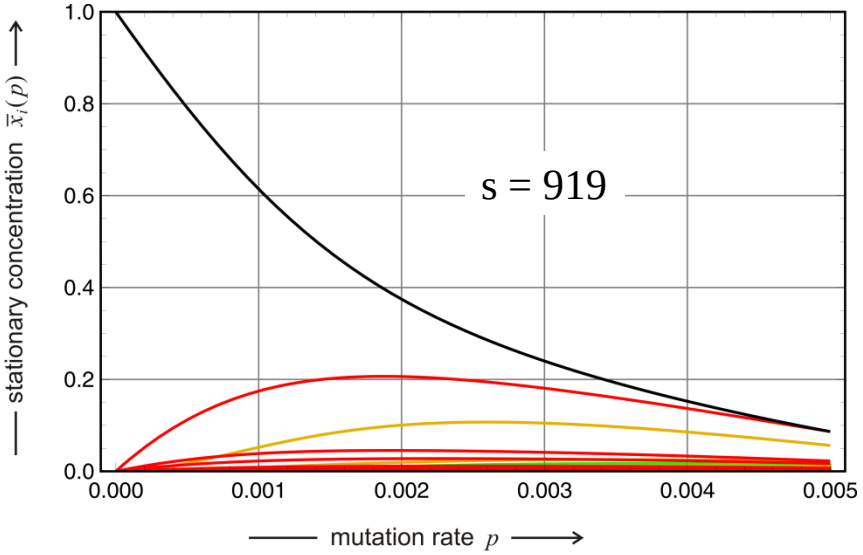


error threshold on ‚realistic' landscapes

n = 10, $f_0 = 1.1$, $f_n = 1.0$, d = 0.5

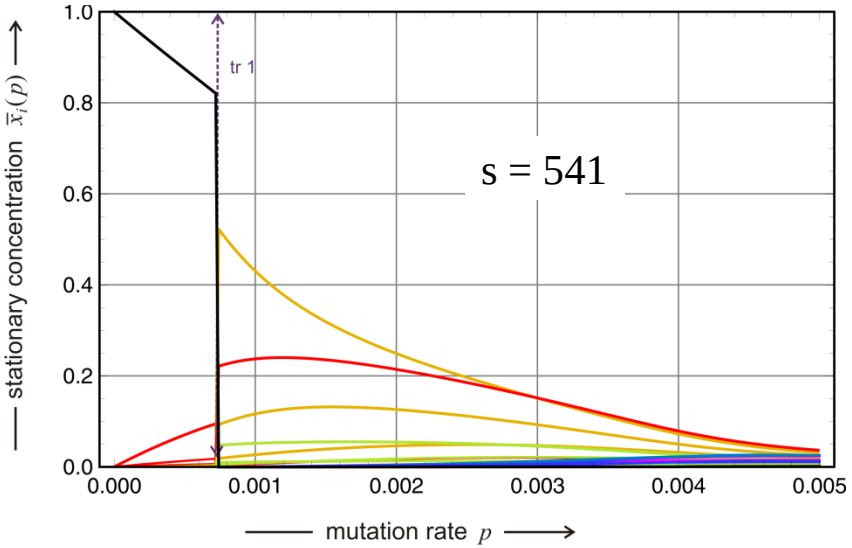


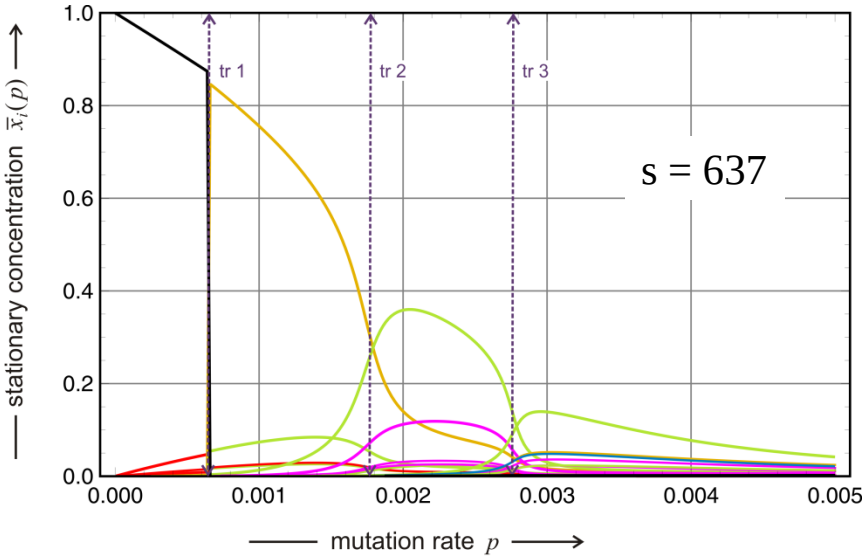


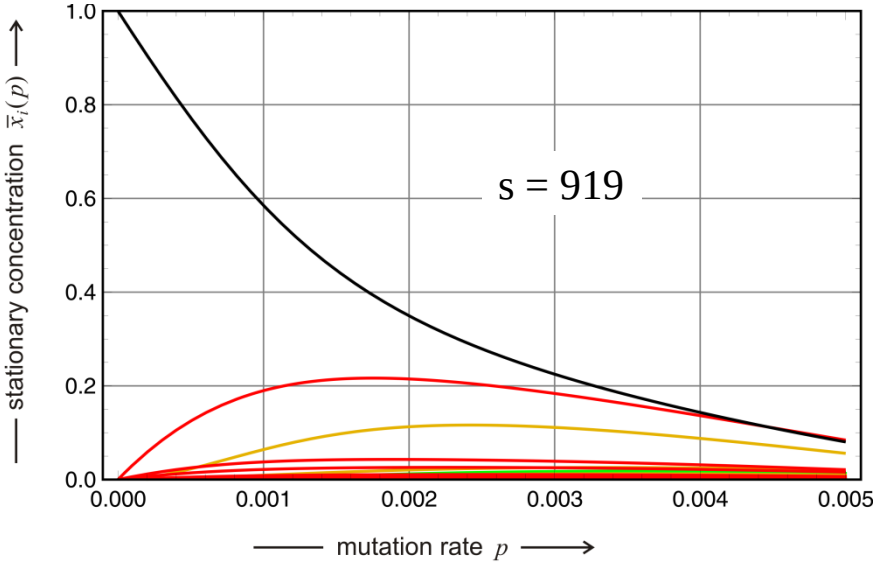


error threshold on ‚realistic' landscapes

n = 10, $f_0 = 1.1$, $f_n = 1.0$, d = 0.995

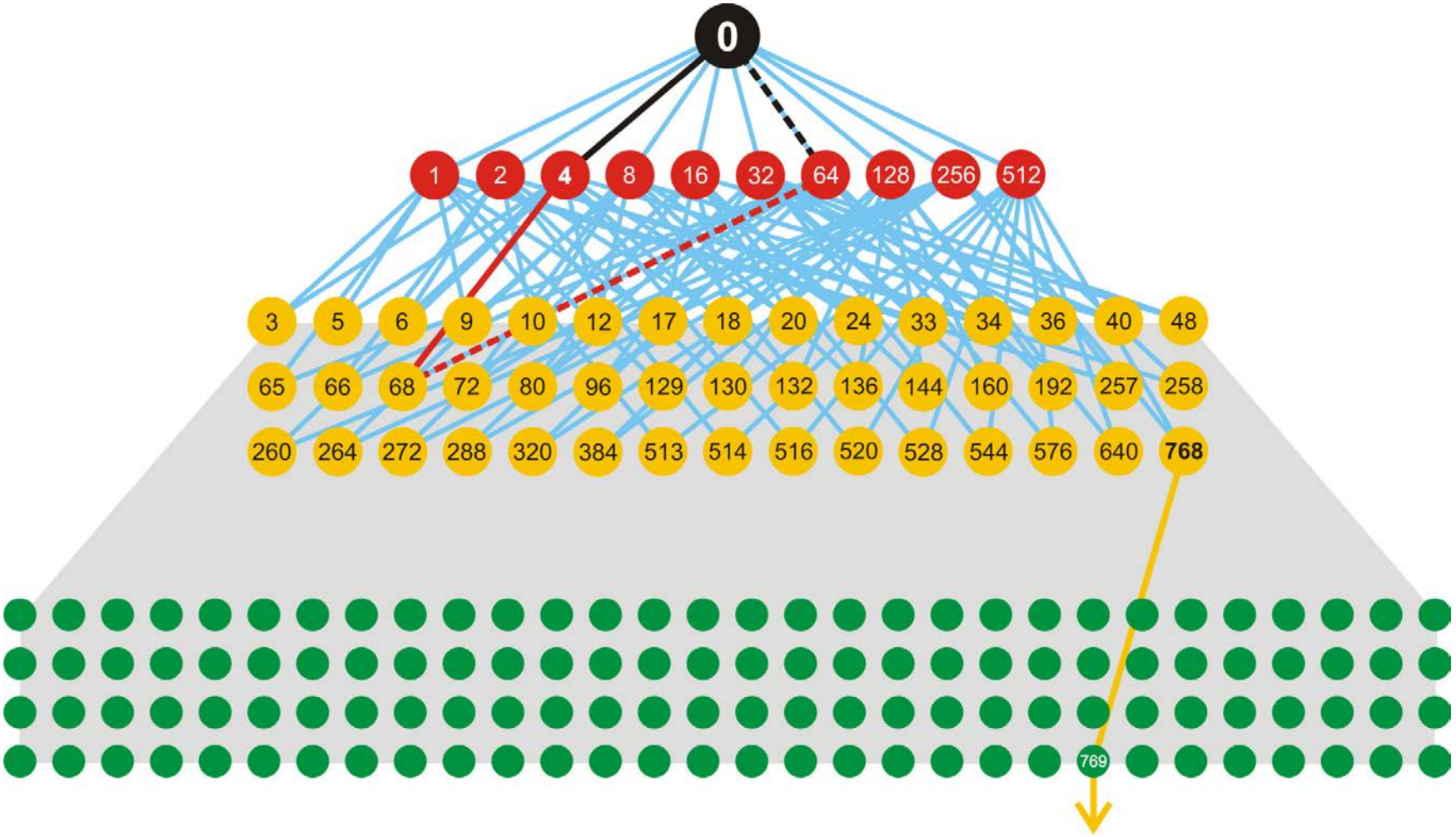




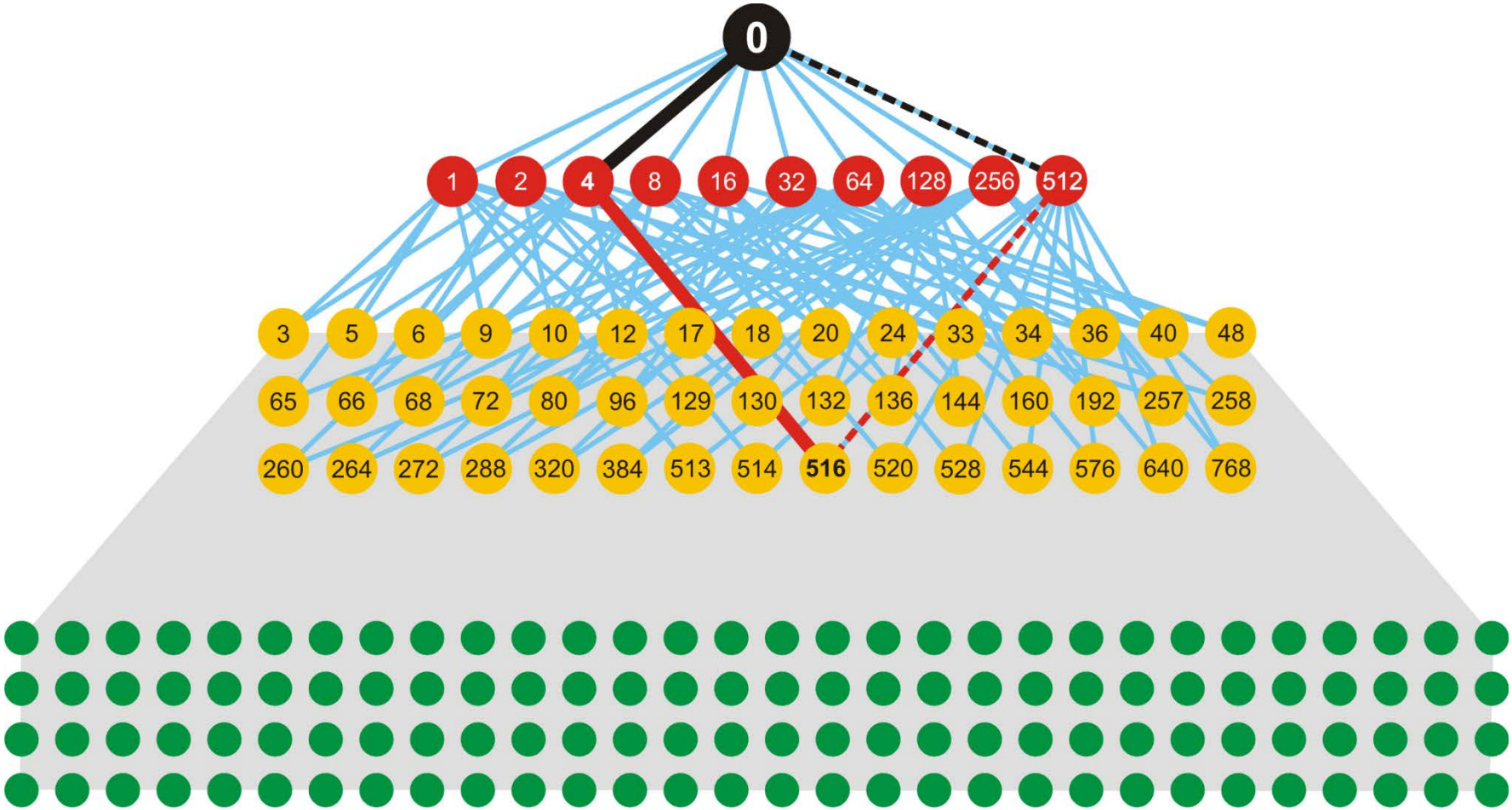


error threshold on ‚realistic‘ landscapes

n = 10, $f_0 = 1.1$, $f_n = 1.0$, d = 1.0



determination of the dominant mutation flow: $d = 1$, $s = 613$



determination of the dominant mutation flow: $d = 1$, $s = 919$

1. What is a „quasispecies"?
2. Detection of the „error threshold"
3. Error thresholds on „simple landscapes"
4. Error thresholds and phase transitions
5. „Realistic" landscapes
6. **Neutrality in evolution**

Motoo Kimura, 1924 - 1994

Motoo Kimura's population genetics of neutral evolution.

Evolutionary rate at the molecular level. *Nature* **217**: 624-626, 1955.

The Neutral Theory of Molecular Evolution. Cambridge University Press. Cambridge, UK, 1983.

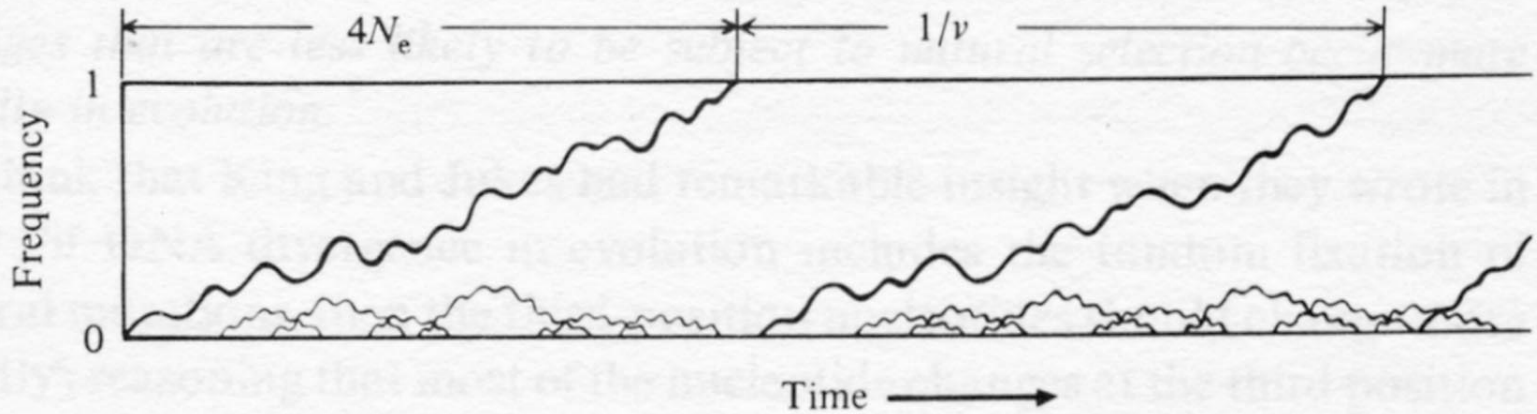
THE NEUTRAL THEORY
OF MOLECULAR EVOLUTION

MOTOO KIMURA
National Institute of Genetics, Japan

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

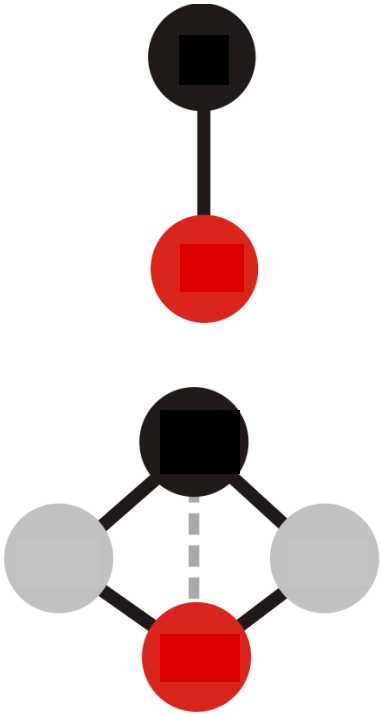
CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

Fig. 3.1. Behavior of mutant genes following their appearance in a finite population. Courses of change in the frequencies of mutants destined to fixation are depicted by thick paths. N_e stands for the effective population size and v is the mutation rate.



Motoo Kimura

Is the Kimura scenario correct for frequent mutations?



$\mathbf{d_H = 1}$

$$\lim_{p\to 0} x_1(p) = x_2(p) = 0.5$$

$\mathbf{d_H = 2}$

$$\lim_{p\to 0} x_1(p) = \alpha/(1+\alpha)$$

$$\lim_{p\to 0} x_2(p) = 1/(1+\alpha)$$

$\mathbf{d_H \geq 3}$

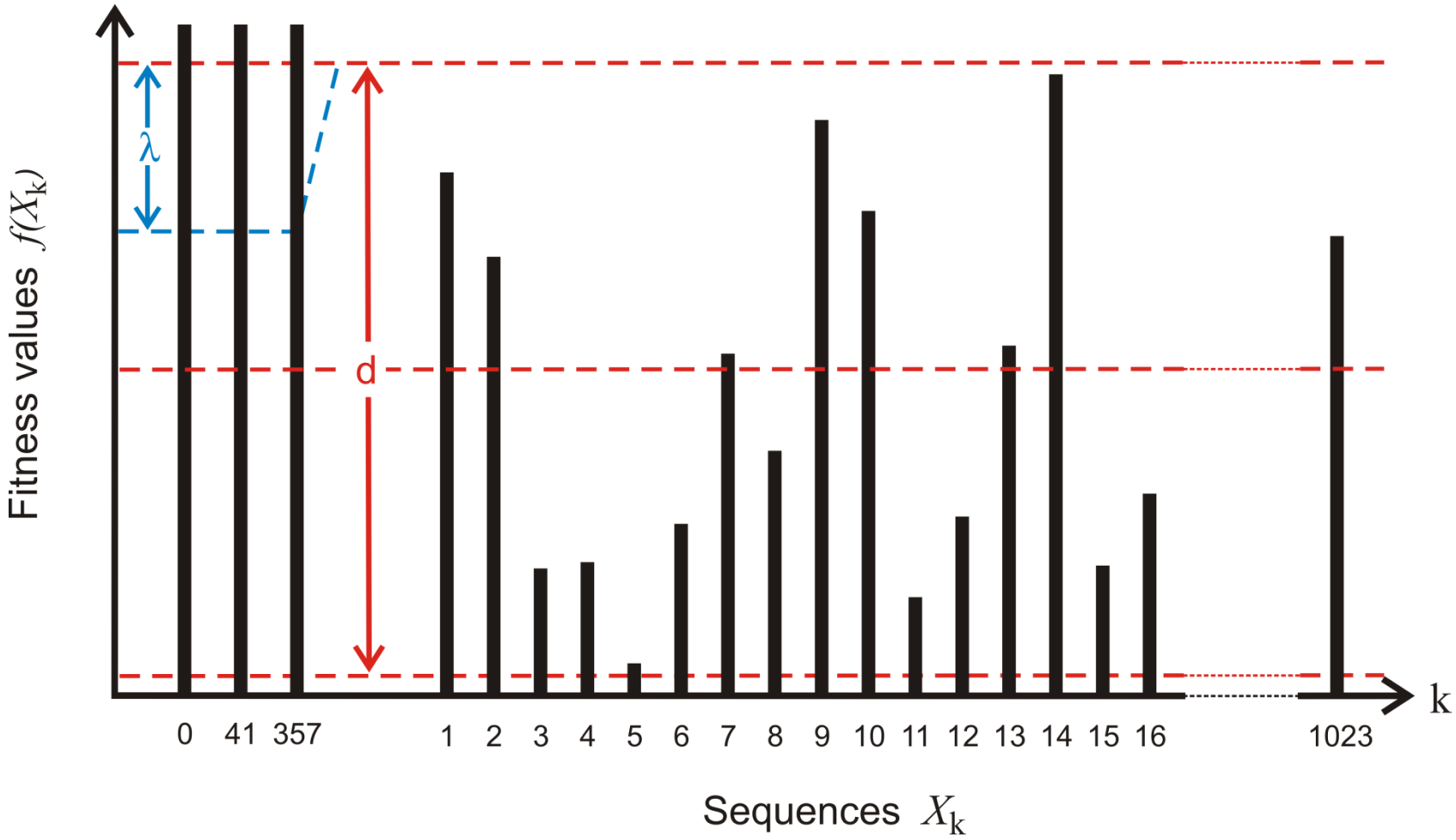
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

$$\lim_{p\to 0} x_1(p) = 0, \lim_{p\to 0} x_2(p) = 1$$

pairs of neutral sequences in replication networks

random fixation in the sense of Motoo Kimura

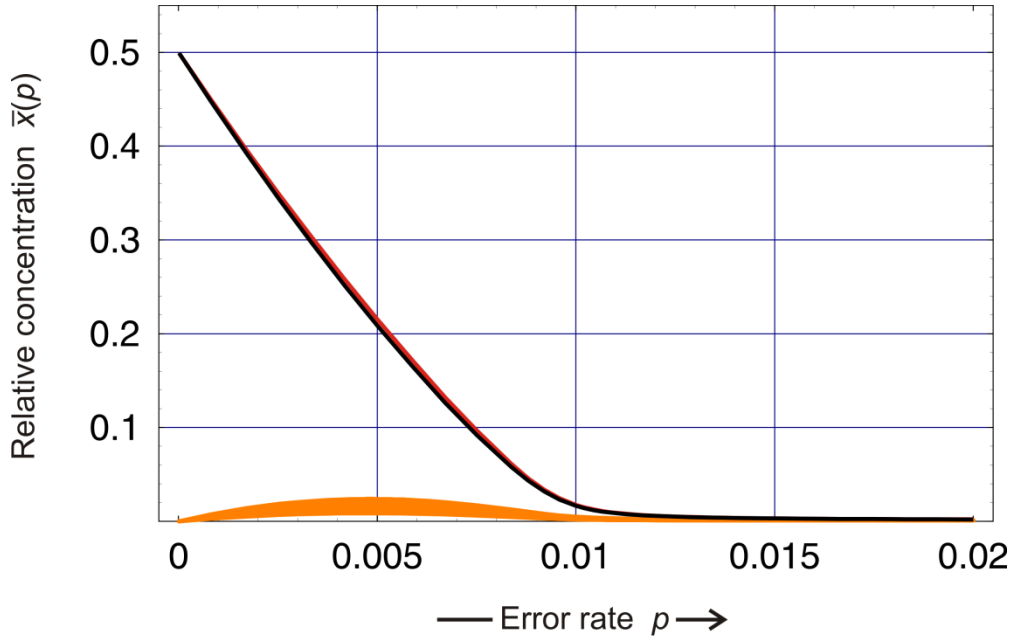
P. Schuster, J. Swetina. 1988. Bull. Math. Biol. 50:635-650

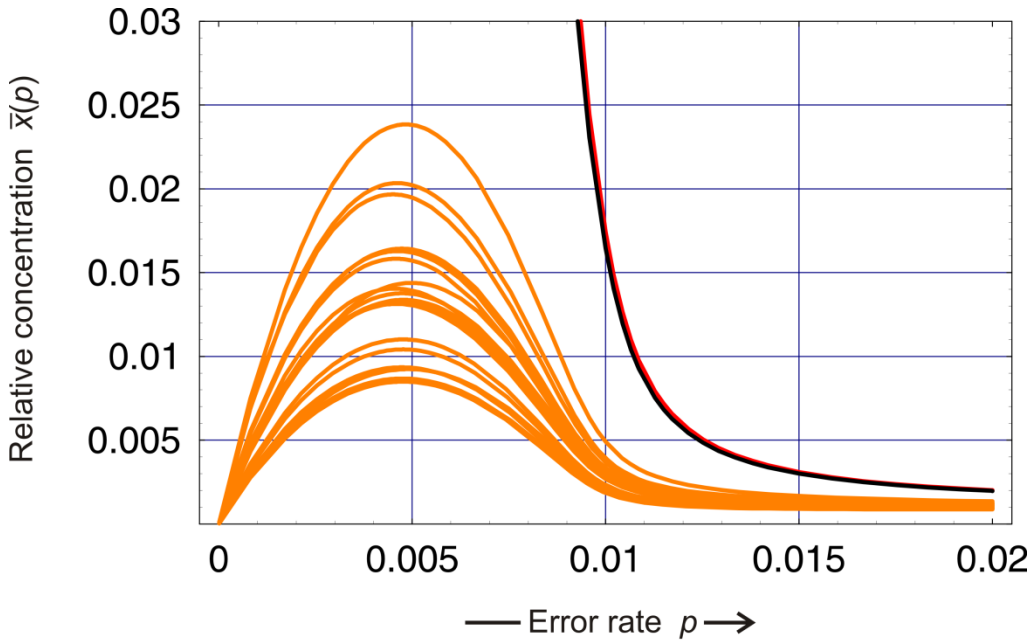


a fitness landscape including neutrality

Neutral network

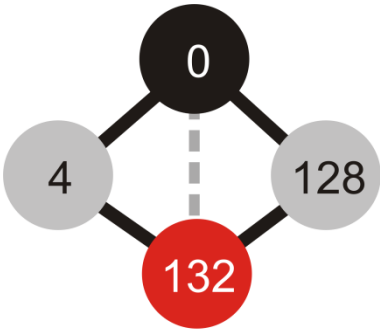
$\lambda = 0.01,\ s = 367$





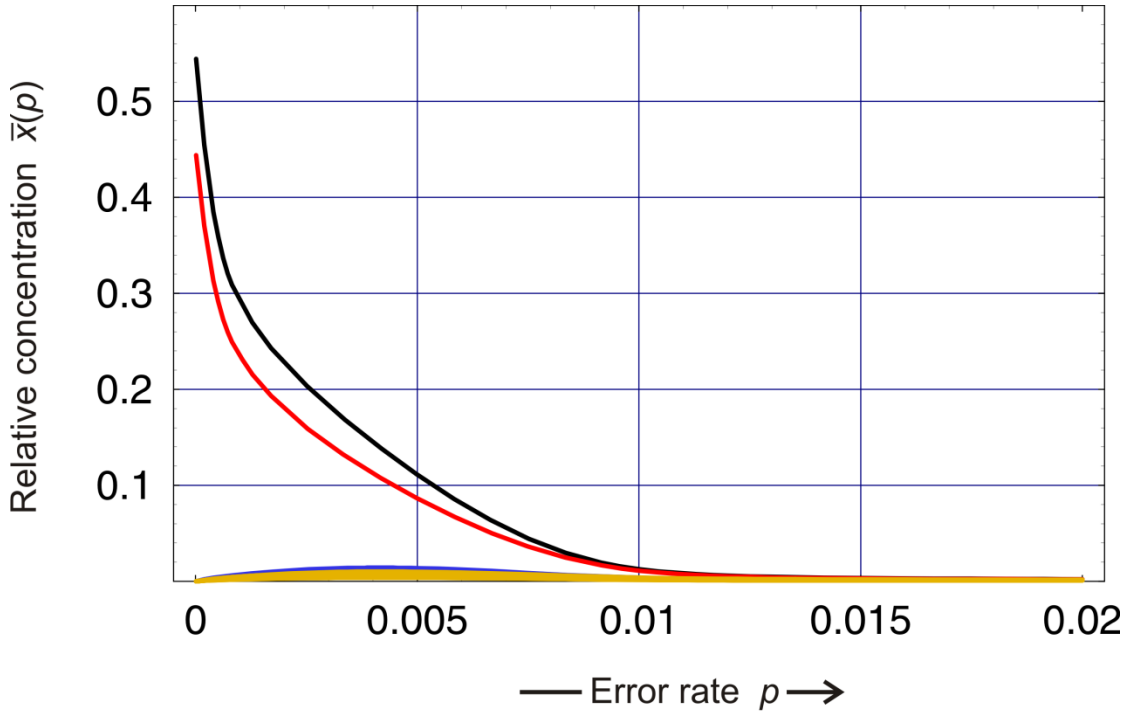
neutral network: individual sequences

$n = 10,\ \sigma = 1.1,\ d = 1.0$



Neutral network

$\lambda = 0.01,\ s = 877$



neutral network: individual sequences

$n = 10,\ \sigma = 1.1,\ d = 1.0$

```
..... ACAUGCGAA .....
..... AUAUACGAA .....
..... ACAUGCGCA .....
..... GCAUACGAA .....
..... ACAUGCUAA .....
..... ACAUGCGAG .....
..... ACACGCGAA .....
..... ACGUACGAA .....
..... ACAUAGGAA .....
..... ACAUACGAA .....
```

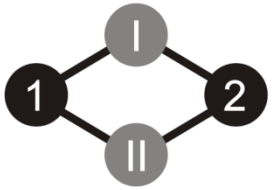
master sequence 1

master sequence 2

..... ACAU$^{G}_{A}$CGAA ..... consensus sequence

```
..... ACAGUCAGAA .....
..... ACAGUCCGAA .....
..... AUAAUCCGAA .....
..... ACAGUCAGCA .....
..... GCAGUCAGAA .....
..... ACAGUCAUAA .....
..... ACAGUCAGAG .....
..... ACAACCCGAA .....
..... ACGGUCAGAA .....
..... ACAGUGAGAA .....
..... ACAAUCAGAA .....
..... ACAAUCCGAA .....
```

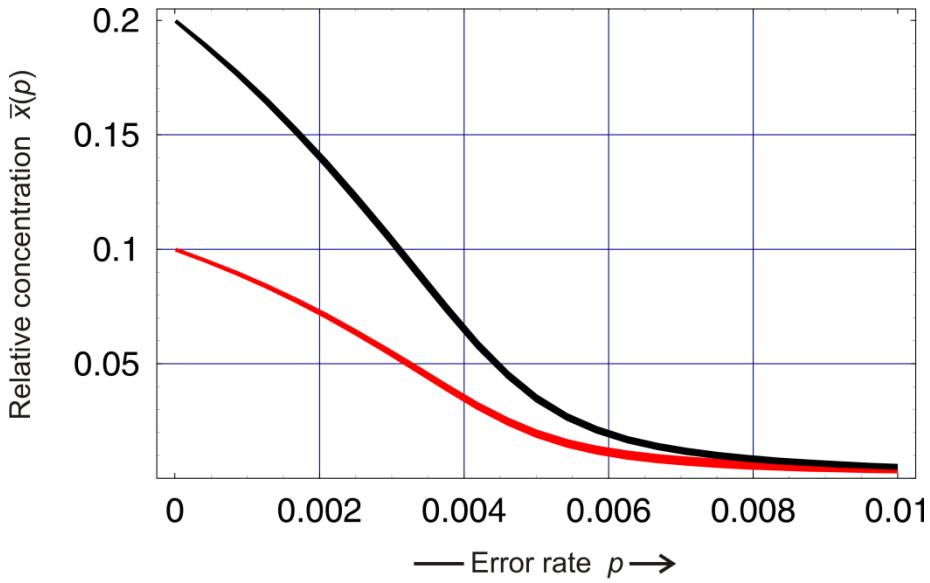
master sequence 1
intermediate I



intermediate II
master sequence 2

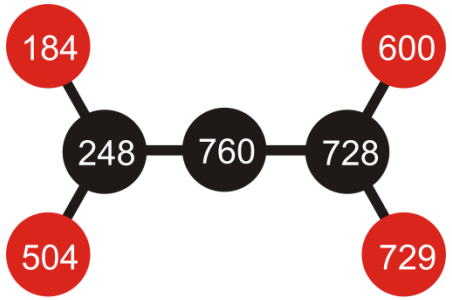
..... ACA$^{G}_{A}$UC$^{A}_{C}$GAA ..... consensus sequence

consensus sequences of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i,X_j) = 1$ and 2.



Perturbation matrix W

$$\mathrm{W} = \begin{pmatrix} f & 0 & \varepsilon & 0 & 0 & 0 & 0 \\ 0 & f & \varepsilon & 0 & 0 & 0 & 0 \\ \varepsilon & \varepsilon & f & \varepsilon & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & f & \varepsilon & 0 & 0 \\ 0 & 0 & 0 & \varepsilon & f & \varepsilon & \varepsilon \\ 0 & 0 & 0 & 0 & \varepsilon & f & 0 \\ 0 & 0 & 0 & 0 & \varepsilon & 0 & f \end{pmatrix}$$



Neutral network

λ = 0.10, s = 229

Adjacency matrix

Largest eigenvector of W

$$\xi_0 = (0.1, 0.1, 0.2, 0.2, 0.2, 0.1, 0.1) .$$

neutral networks with increasing λ: λ = 0.10, s = 229

Coworkers

Universität Wien

Peter Stadler, **Bärbel M. Stadler**, Bioinformatik**,** Universität Leipzig, GE

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Sebastian Bonhoeffer, Theoretical Biology, ETH Zürich, CH

Christian Reidys, Mathematics, University of Southern Denmark, Odense, DK

Christian Forst, Southwestern Medical Center, University of Texas, Dallas, TX

Thomas Wiehe, Institut für Genetik, Universität Köln, GE

Ivo L.Hofacker, Theoretische Chemie**,** Universität Wien, AT

Kurt Grünberger, Ulrike Mückstein, **Jörg Swetina**, **Manfred Tacker**, **Andreas Wernitznig**, Theoretische Chemie, Universität Wien, AT

Acknowledgement of support

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)
Projects No. 09942, 10578, 11065, 13093
13887, and 14898

Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)
Project No. Mat05

Jubiläumsfonds der Österreichischen Nationalbank
Project No. Nat-7813

European Commission: Contracts No. 98-0189, 12835 (NEST)

Austrian Genome Research Program – GEN-AU: Bioinformatics Network (BIN)

Österreichische Akademie der Wissenschaften

Siemens AG, Austria

Universität Wien and the Santa Fe Institute

Universität Wien

Thank you for your attention!

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks