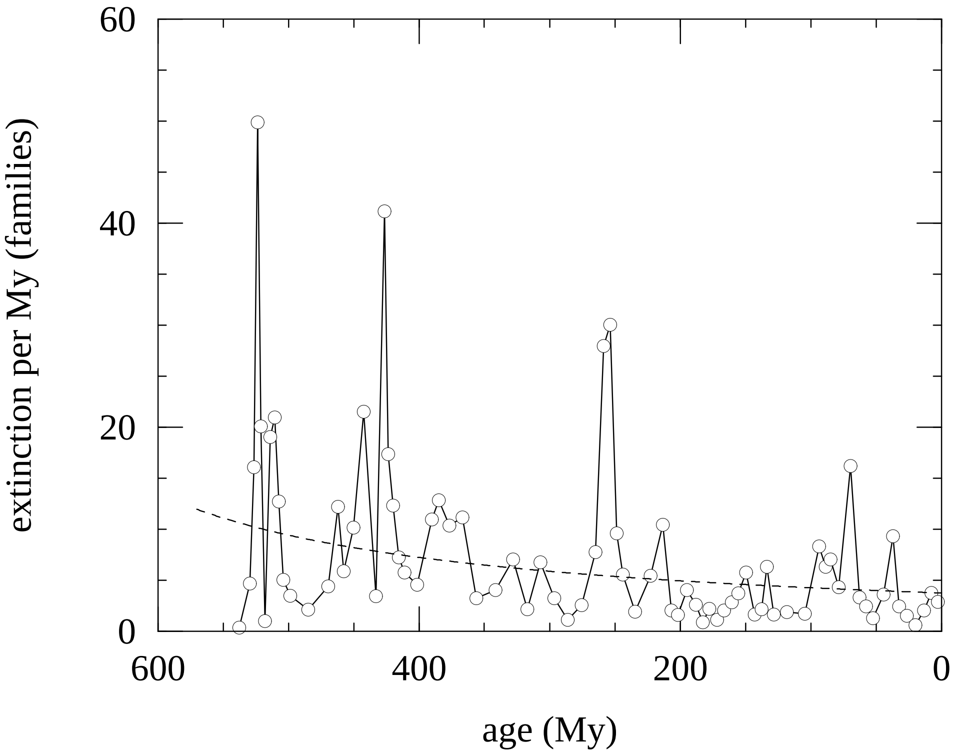
Regimes of asexual evolution in rugged fitness landscapes

- Macroevolution and microevolution
- The adaptive landscape
- Sequence space, fitness and Fisher-Wright sampling
- Regimes of asexual population dynamics
- Evolutionary trajectories in the quasispecies model

Joint work with Kavita Jain JSTAT (2005) P04008, q-bio.PE/0606025, q-bio.PE/0508008
Supported by DFG within SFB/TR 12 “Symmetries and Universality in Mesoscopic Systems”
& SFB 680 “Molecular Basis of Evolutionary Innovations”

Macroevolution: Episodic patterns in the fossil record

- Number of extinct families of known marine organisms (Sepkoski, 1992)
- Extinction rate is intermittent with power law event size distribution
- Decline of the average extinction rate ⇒ **non-stationary** dynamics



Newman & Eble (1999)

Microevolution: Adaptation of viral populations

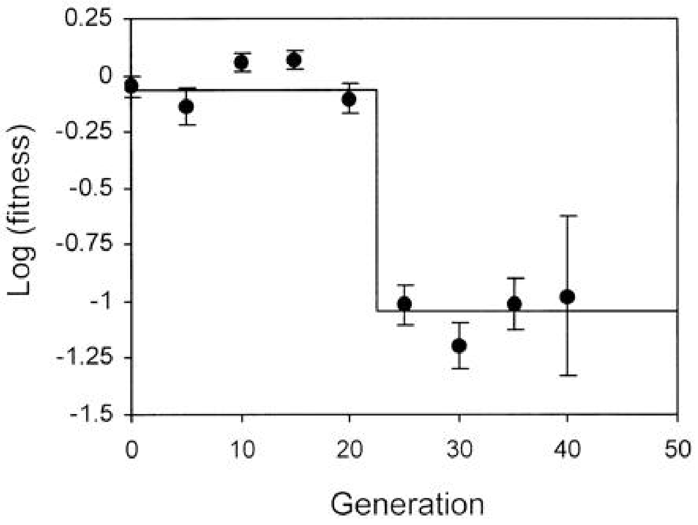
- Fitness decline induced in populations of bacteriophage $\phi 6$ by successive population bottlenecks

- Fitness recovery through increasing population size

- Step-like fitness changes reflect single deleterious or beneficial (compensatory) mutations

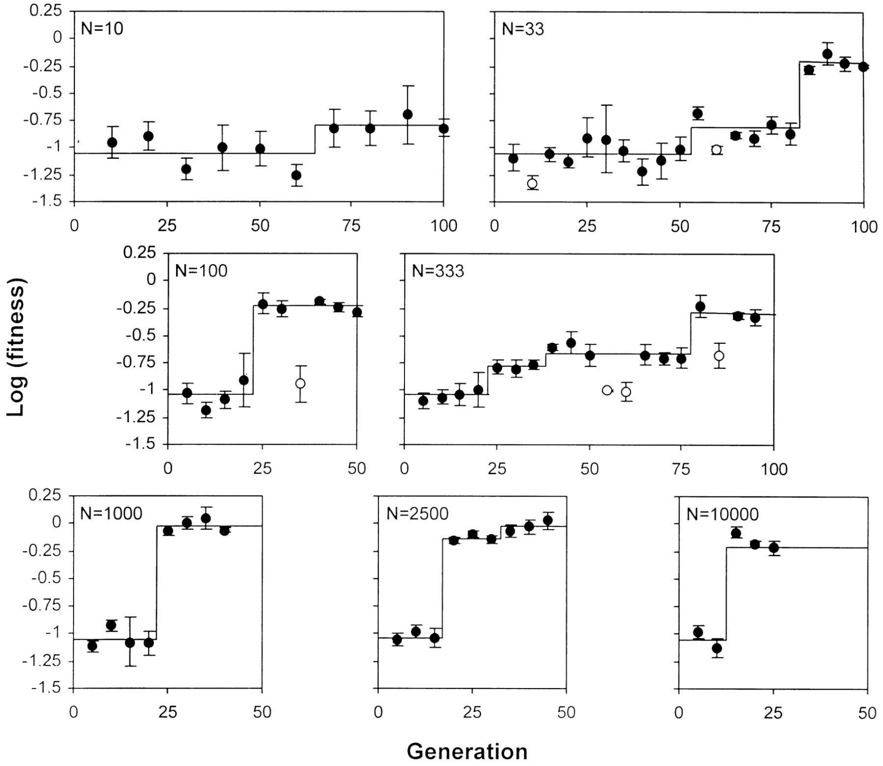
- Evidence for ruggedness of the fitness landscape

C.L. Burch, L. Chao,
Genetics **151**, 921 (1999)



Microevolution: Adaptation of viral populations

- Fitness decline induced in populations of bacteriophage $\phi6$ by successive population bottlenecks
- Fitness recovery through increasing population size
- Step-like fitness changes reflect single deleterious or beneficial (compensatory) mutations
- Evidence for ruggedness of the fitness landscape

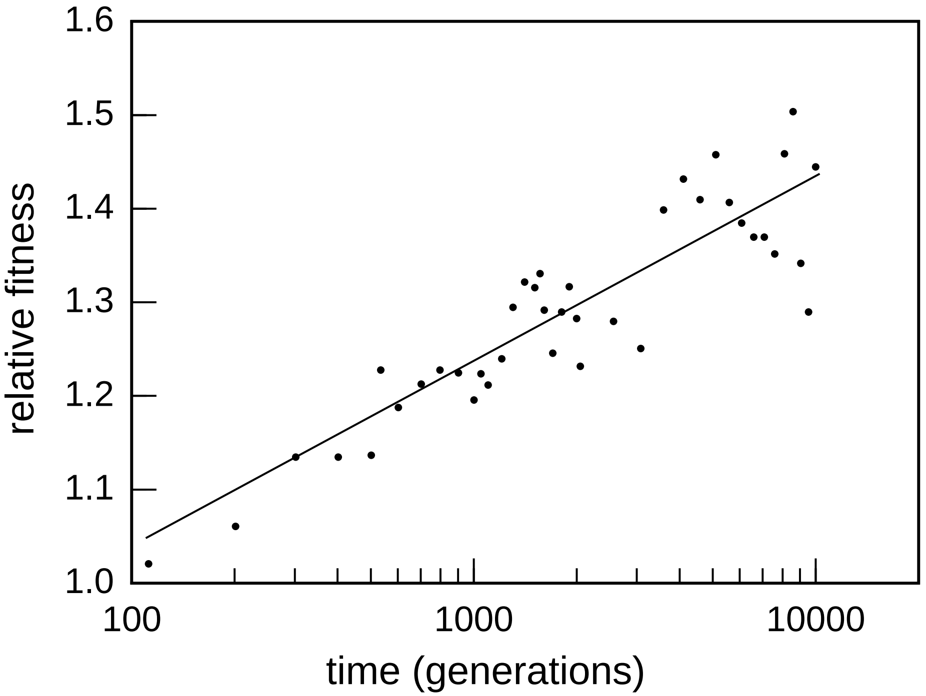


C.L. Burch, L. Chao,
Genetics **151**, 921 (1999)

Microevolution: Adaptation of bacterial populations

S.F. Elena, R.E. Lenski, Nature Reviews Genetics **4**, 457 (2003)

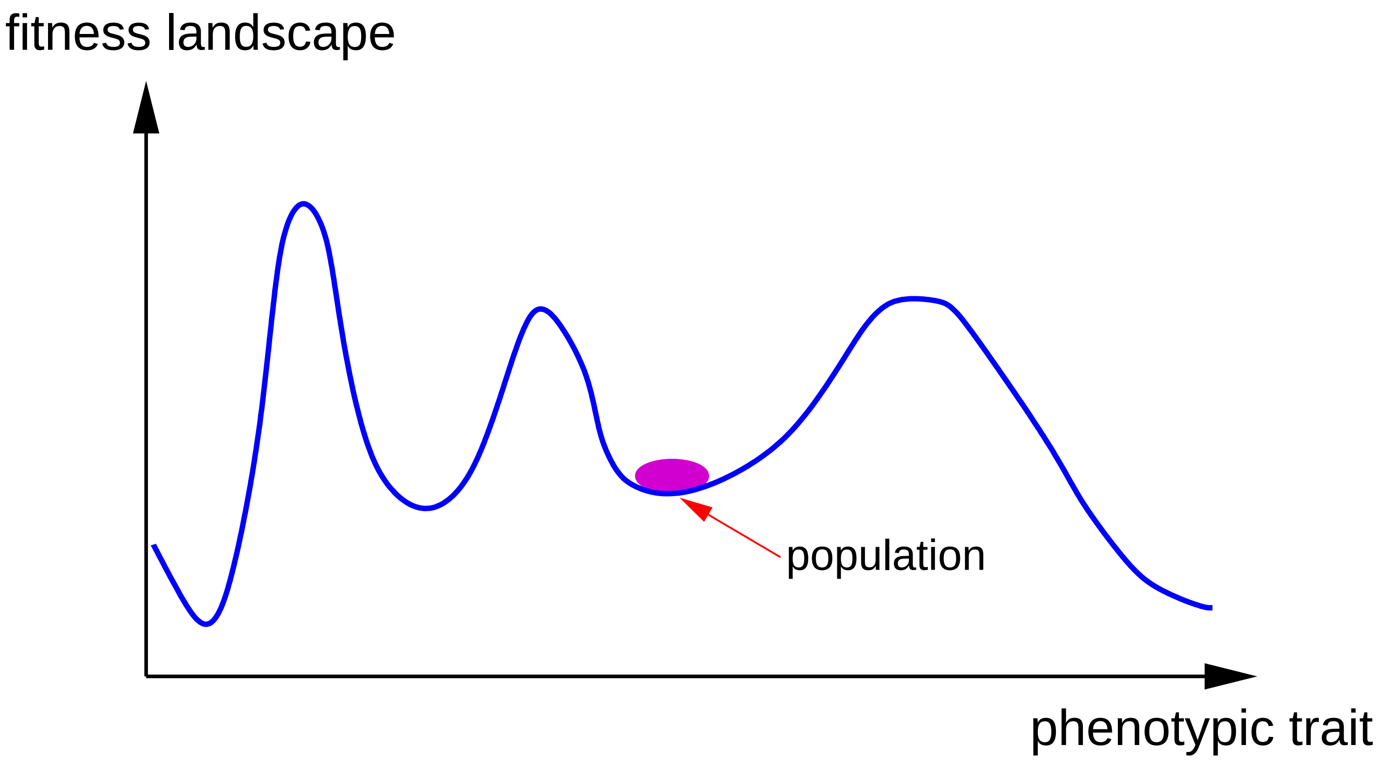
- 12 populations of *E. coli* propagated in identical environments
- Step-wise increase of cell size and fitness reflects selection of rare beneficial mutations; 4 - 6 steps in 10000 generations
- Fitness increase slows down, but rate of genetic change does not
- Parallel and divergent changes



fit: Sibani, Brandt & Alstrøm (1998)

Adaptation as a hill-climbing process

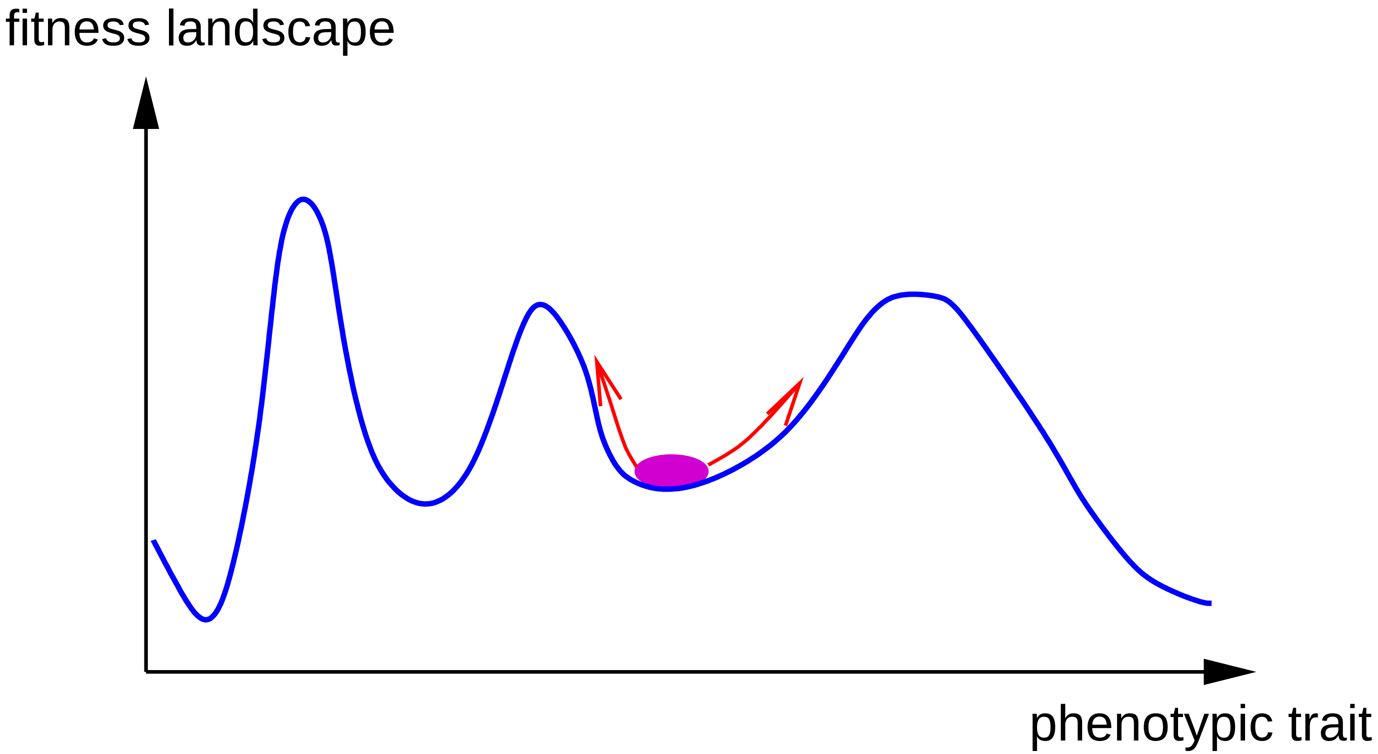
S. Wright, 1932



- Initial population placed in a new environment

Adaptation as a hill-climbing process

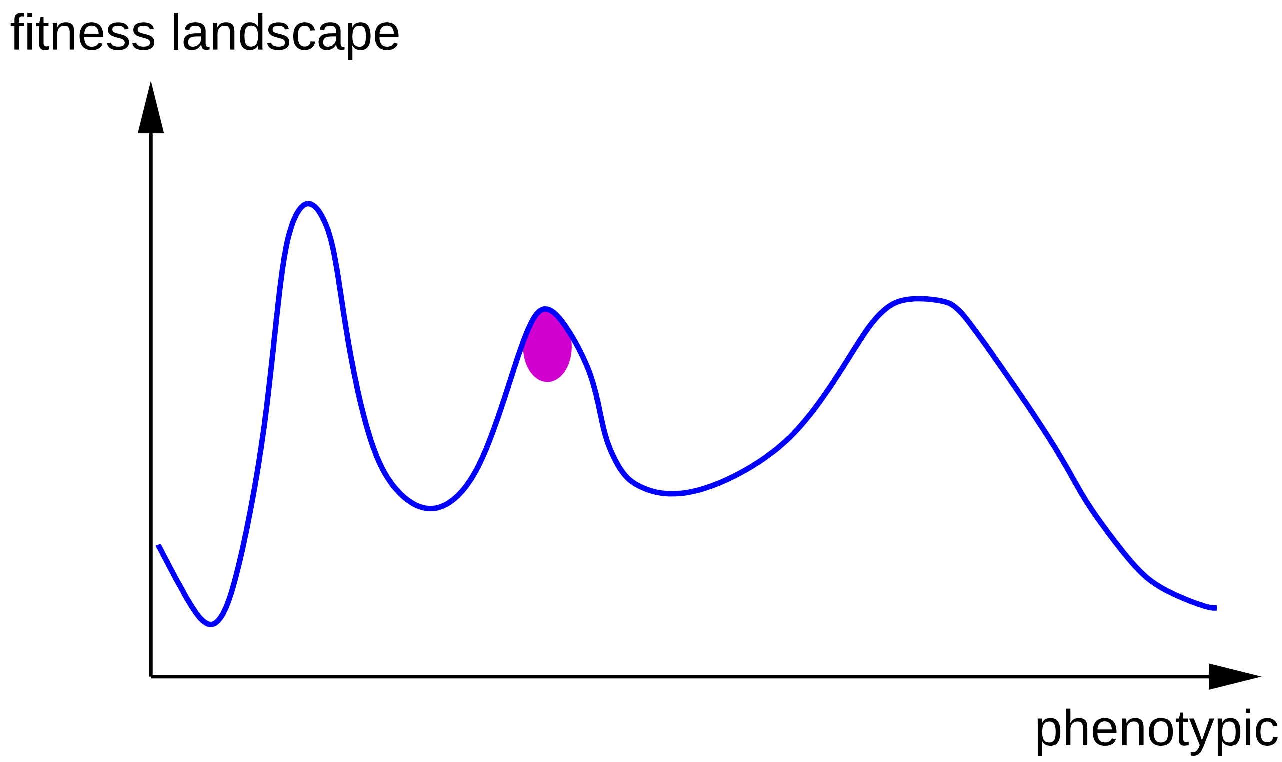
S. Wright, 1932



- Rapid adaptation to a nearby fitness maximum

Adaptation as a hill-climbing process

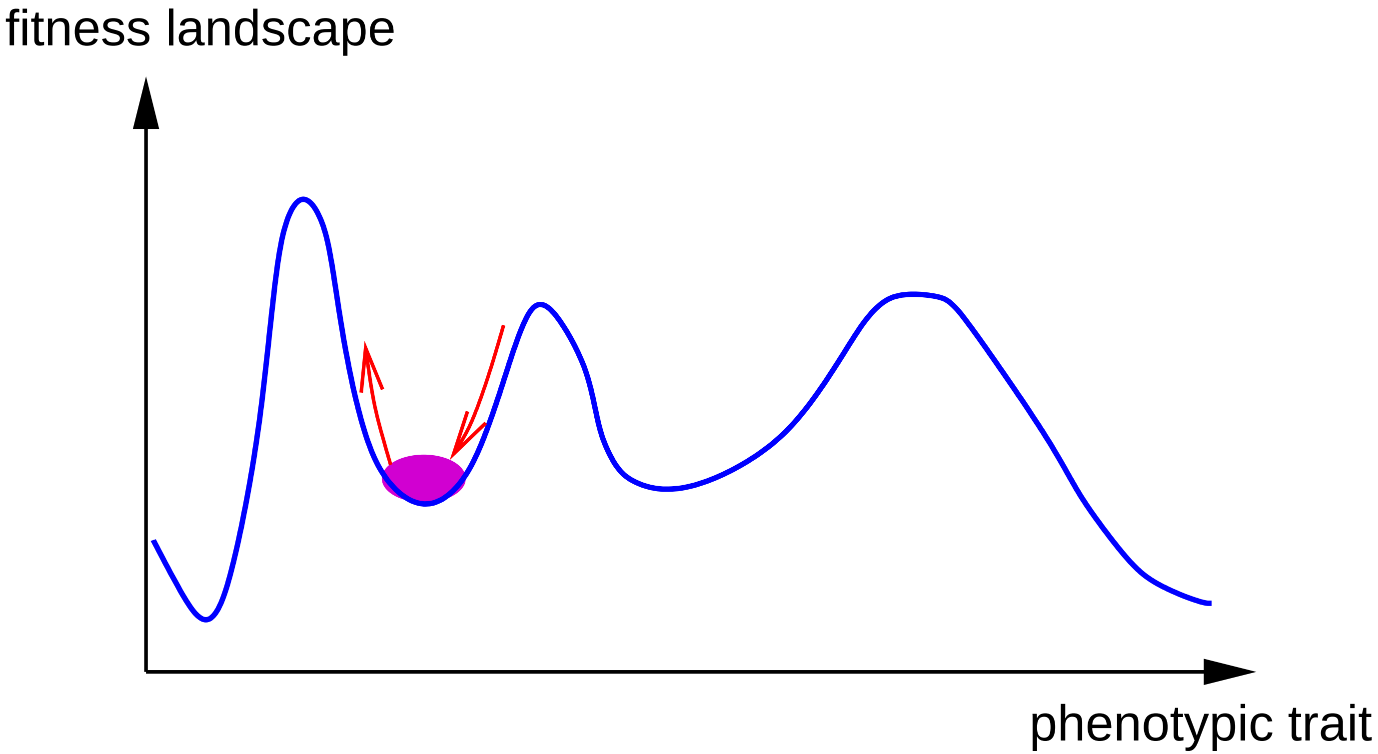
S. Wright, 1932



- Stabilizing selection at the fitness peak; mutation-selection balance

Adaptation as a hill-climbing process

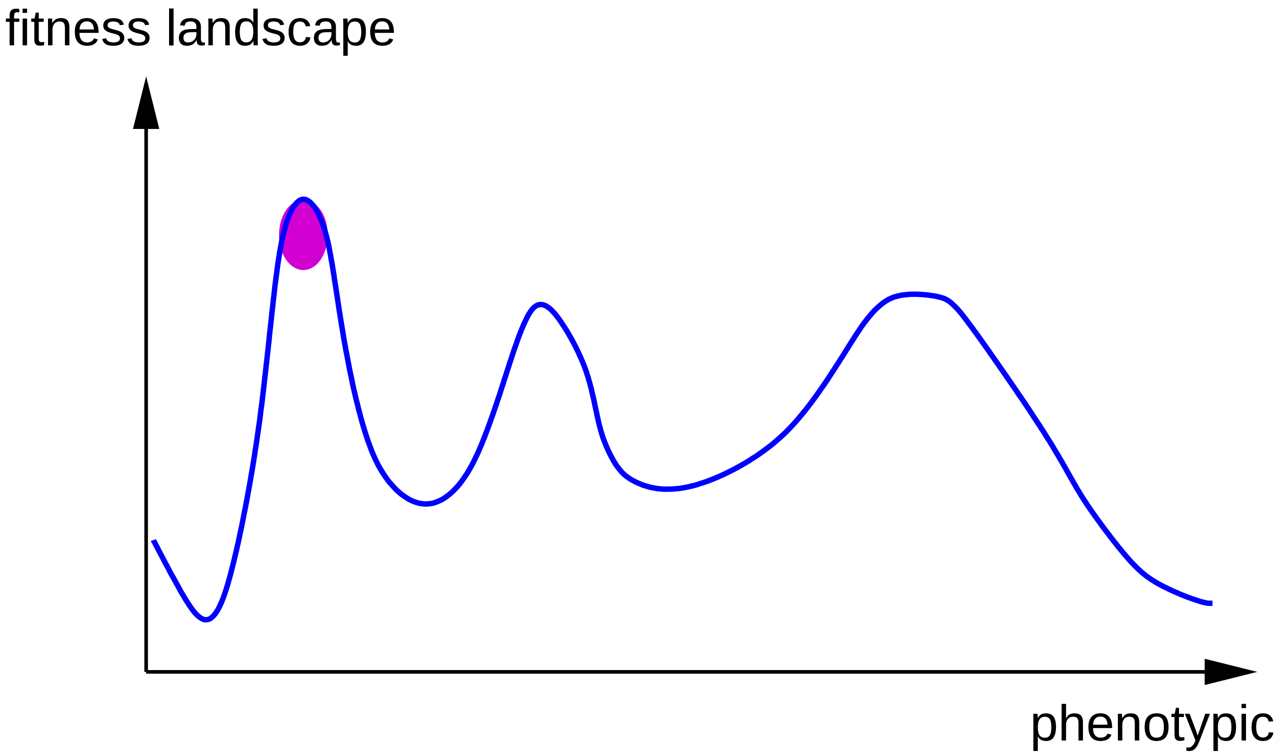
S. Wright, 1932



- Rare, rapid transition through non-adaptive zone to a higher fitness peak:

Adaptation as a hill-climbing process

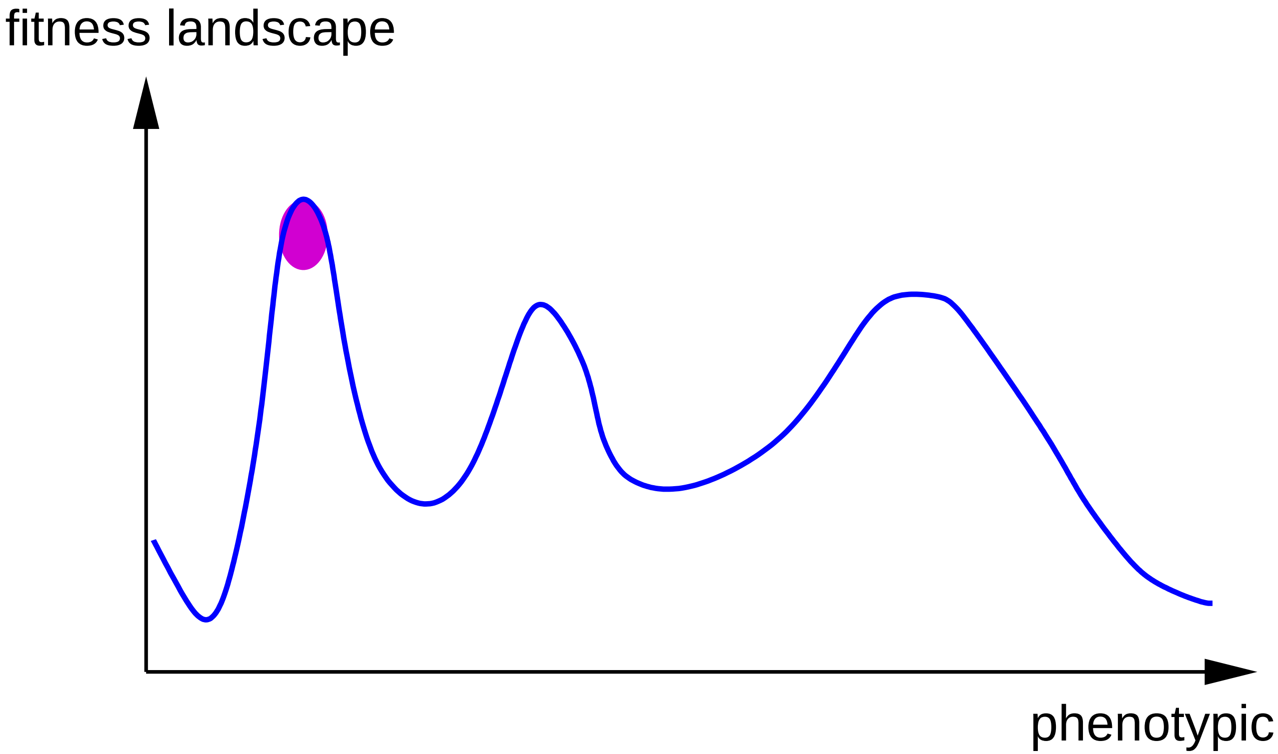
S. Wright, 1932



- Rare, rapid transition through non-adaptive zone to a higher fitness peak: **Quantum evolution** (G.G. Simpson, 1944)

Adaptation as a hill-climbing process

S. Wright, 1932



- **Goal:** To turn this mental imagery into a theory of adaptation

Sequence space

- Each individual carries a genetic sequence of length L

  $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_L)$ with $\sigma_i = \mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}$ **genotype**

  Simplification: Binary sequences $\sigma_i = 0, 1$

- **Point mutations** change individual letters in the sequence:

$$\sigma_i \quad \rightarrow \quad 1 - \sigma_i$$

- The **Hamming distance** between two sequences σ, σ' is the number of letters in which they differ:

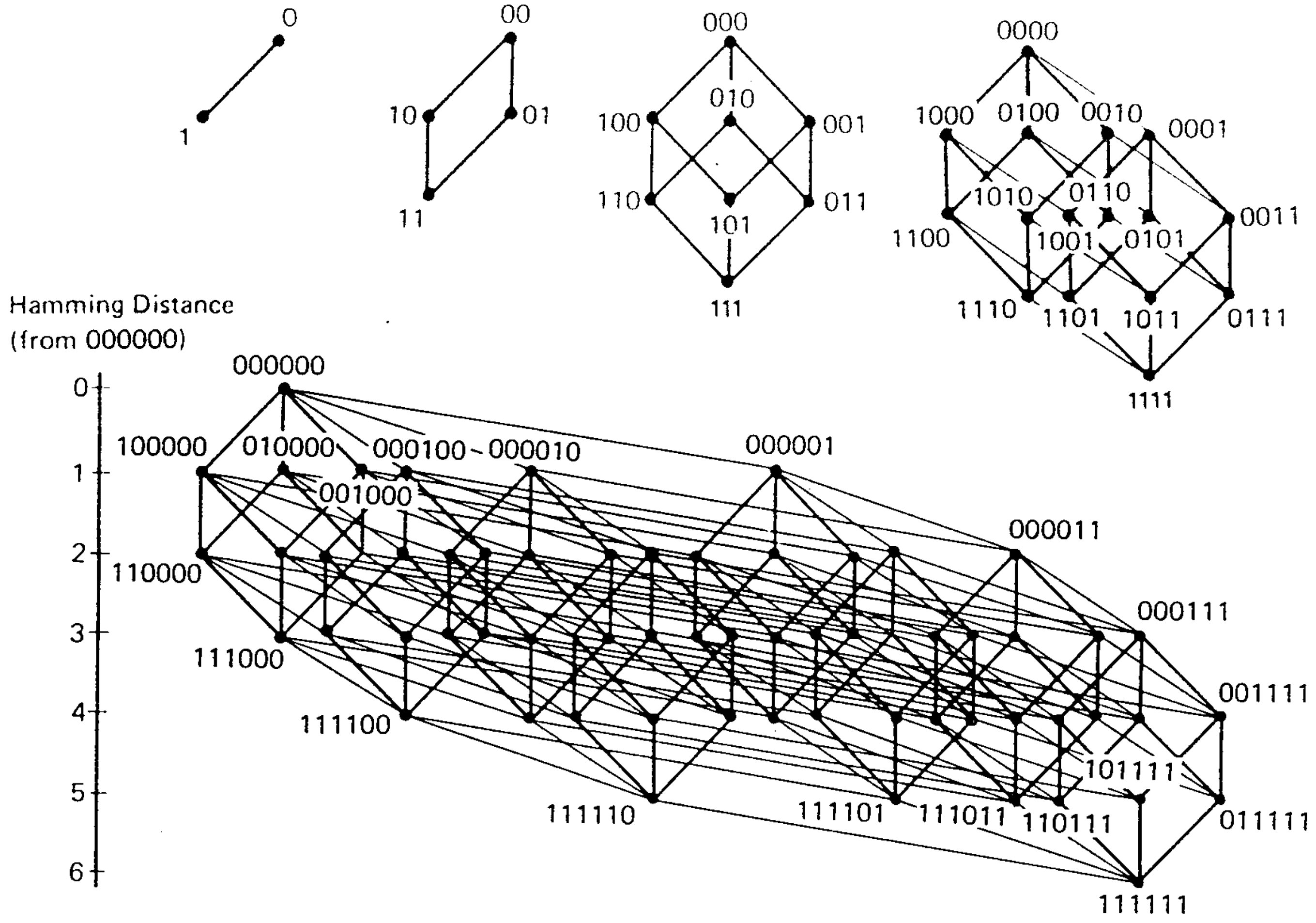
$$d(\sigma, \sigma') = \sum_{i=1}^{L} (\sigma_i - \sigma_i')^2$$

- Total number of sequences: $S = 2^L$ **volume**

  Maximal distance between two sequences: $L = \log_2(S)$ **diameter**

  $\Rightarrow$ infinite dimensionality!

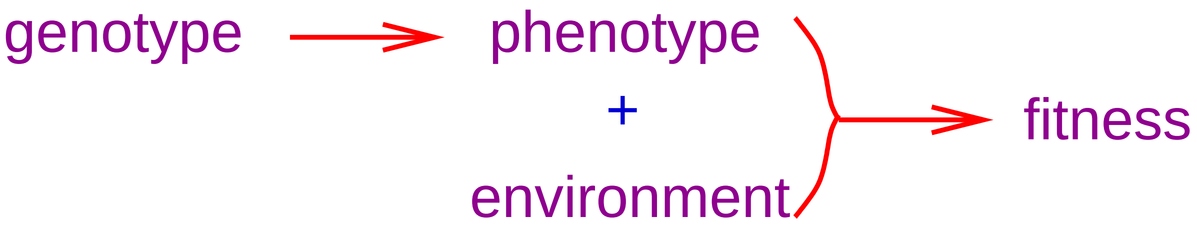
L-dimensional hypercubes/Hamming graphs



Fitness

- The fitness $W(\sigma)$ of genotype σ is the **expected number of offspring** of an individual carrying σ

- The mapping $\sigma \rightarrow W(\sigma)$ is very complicated:



genotype → phenotype + environment → fitness

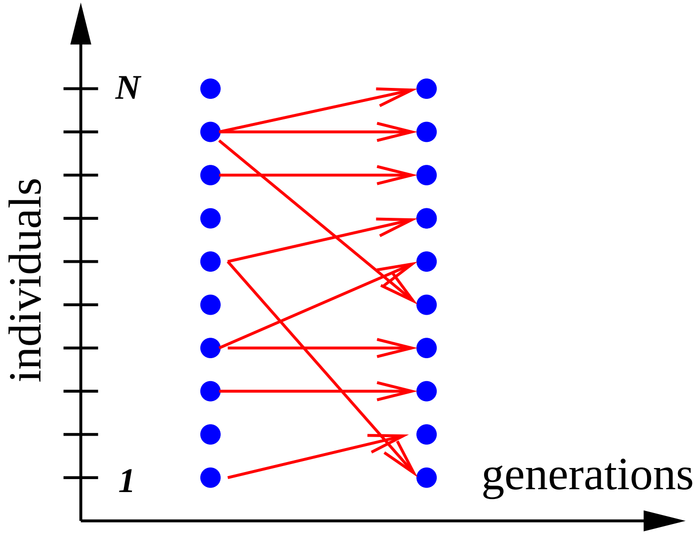
- Single peak landscapes: $W(\sigma) = w(d(\sigma, \sigma_0))$ σ_0: master sequence

- Maximally rugged fitness landscape: Fitnesses $W(\sigma)$ are **uncorrelated quenched random variables** drawn from a common distribution

  $\Rightarrow$ random energy model (REM) of spin glasses (Derrida, 1981)
  $\Rightarrow$ house of cards model of population genetics (Kingman, 1977)

Evolution of asexual populations

Basic model: Stochastic Fisher-Wright sampling of a finite population



- Each individual choses an ancestor from the preceding generation
- Genotype σ is chosen with probability $\sim W(\sigma)$
- Point mutations occur with probability μ per site and generation
- Interplay of disorder with two distinct sources of fluctuations $(\sim 1/N, \mu)$

Fixation

- In the absence of mutations $(\mu = 0)$ the population becomes genetically homogeneous (monomorphic) for $t \to \infty$

- When a single mutant σ' is introduced into a monomorphic population with genotype σ, the outcome for $t \to \infty$ is either fixation (all σ') or loss of the mutation (all σ)

- Fixation of deleterious mutations $(W(\sigma') < W(\sigma))$ is exponentially unlikely for large N

- Beneficial mutations $(W(\sigma') > W(\sigma))$ are fixed with probability

$$\Pi(\sigma', \sigma) = 1 - \frac{W(\sigma)}{W(\sigma')}$$

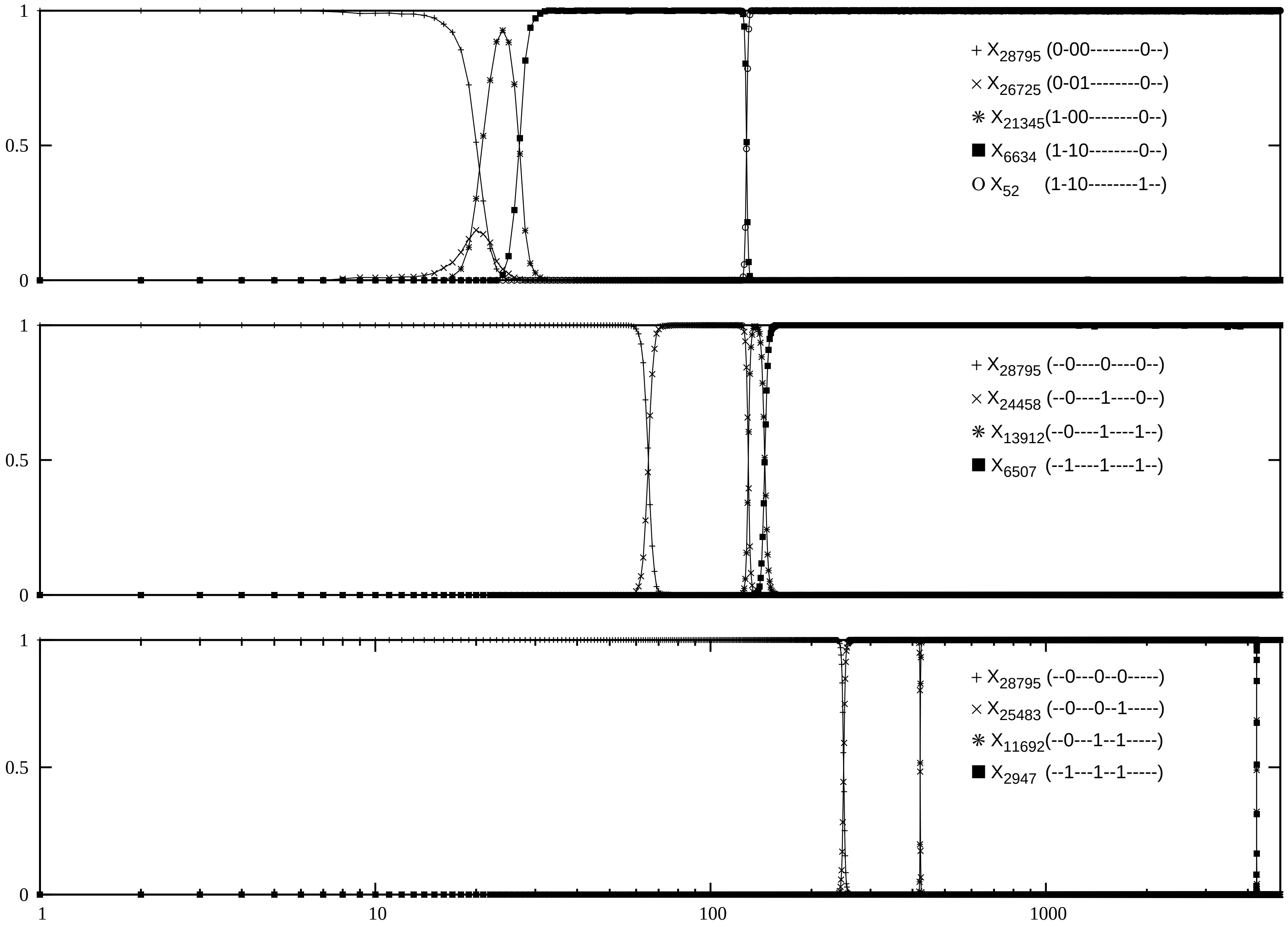
Neutral evolution and adaptive walks

- In a flat fitness landscape $[W = \mathrm{const.}]$ the population diffuses randomly in sequence space (genetic drift)
- Average genetic distance between two individuals: (Derrida & Peliti 1991)

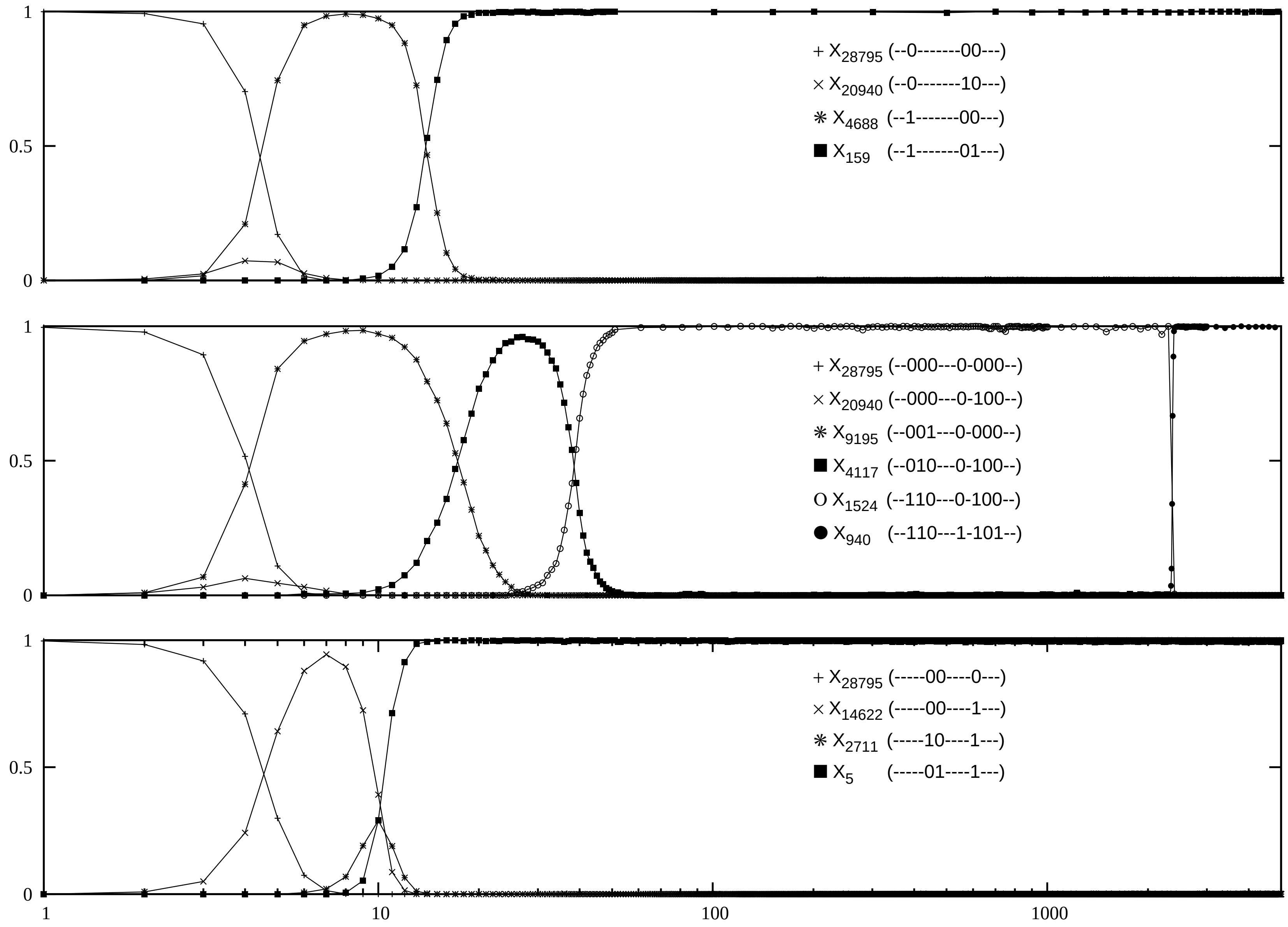
$$\langle d(\sigma,\sigma')\rangle = \frac{L}{2}\left(\frac{4\mu N}{1+4\mu N}\right)$$

- $L\mu N \ll 1 \quad \Rightarrow \quad \langle d\rangle \ll 1 \quad \Rightarrow$ population is monomorphic most of the time and evolves by fixation of rare neutral mutations
- In the presence of selection $[W \neq \mathrm{const.}]$ the population performs an uphill adaptive walk by fixation of rare beneficial mutations (Gillespie, 1984)
- Adaptive walks terminate at local fitness maxima

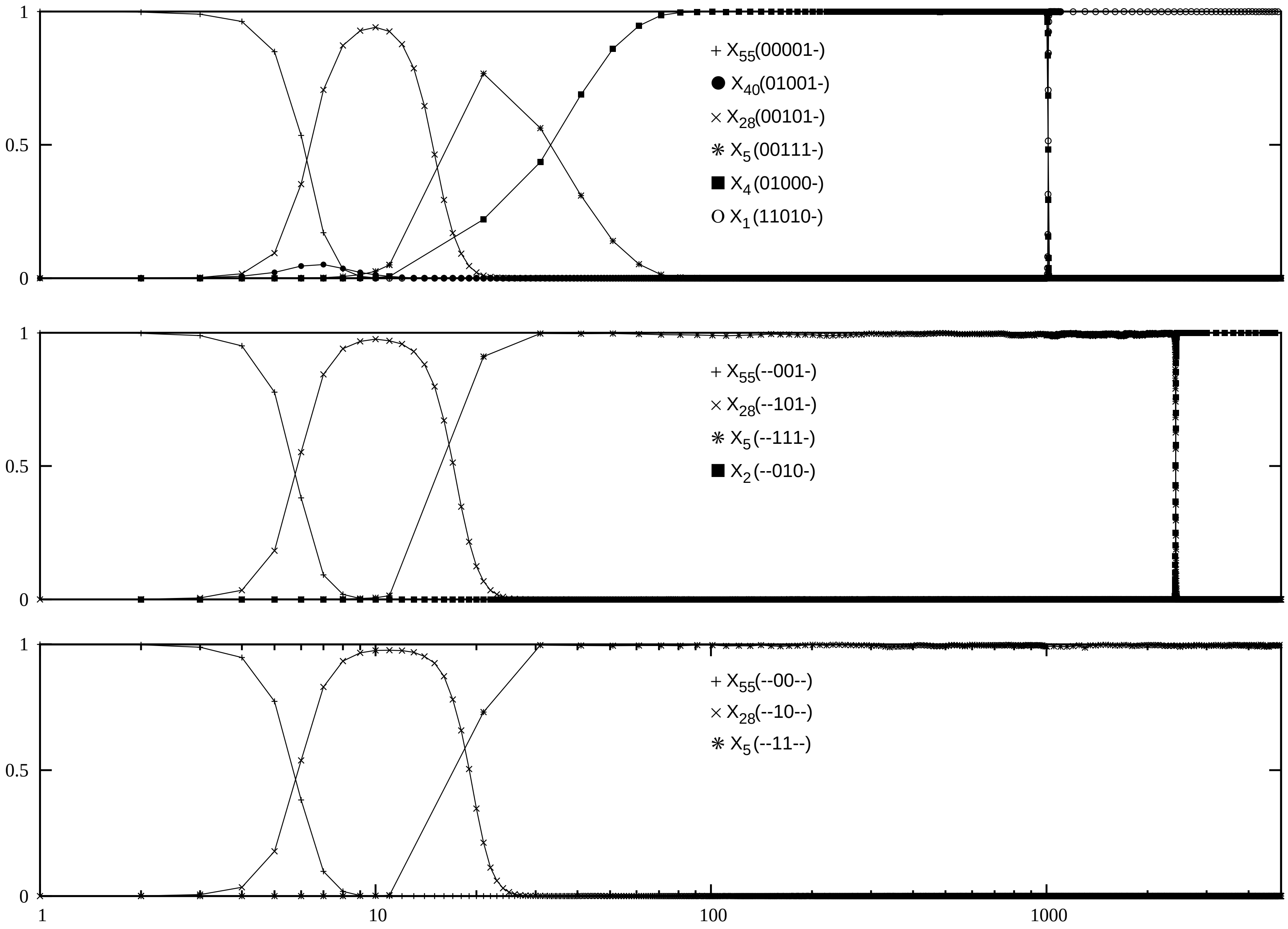
Adaptive walk: $L = 15, N = 1024, \mu = 10^{-5}, 10^{-6}, 10^{-7}$



“Clonal interference”: $L = 15, N = 1024, \mu = 10^{-4}$



Locally deterministic evolution: $L = 6, N = 16384, \mu = 10^{-4}$



Classification of evolutionary regimes

K. Jain, JK, q-bio.PE/0606025

Key parameters: Population size N, mutation probability μ per site & generation, sequence length L

- $LN\mu$: Number of mutants produced per generation
- $LN\mu \ll 1$: Adaptive walk of a monomorphic population
- $1 \ll LN\mu \ll L$: Stochastic regime with clonal interference
- $LN\mu > L$: Deterministic evolution within a shell of size

$$d_{\text{eff}} = \frac{\ln N}{|\ln \mu|}$$

  around the dominating genotype

- $d_{\text{eff}} \to L$: Fully deterministic evolution $\Rightarrow$ quasispecies dynamics

The quasispecies model

M. Eigen, Naturwissenschaften **58**, 465 (1972); K. Jain, JK, q-bio.PE/0508008

- Unnormalized population fraction $Z(\sigma,t)$ satisfies linear time evolution

$$Z(\sigma,t+1)=\sum_{\sigma'}W(\sigma')M^{(\mu)}(\sigma',\sigma)Z(\sigma',t)$$

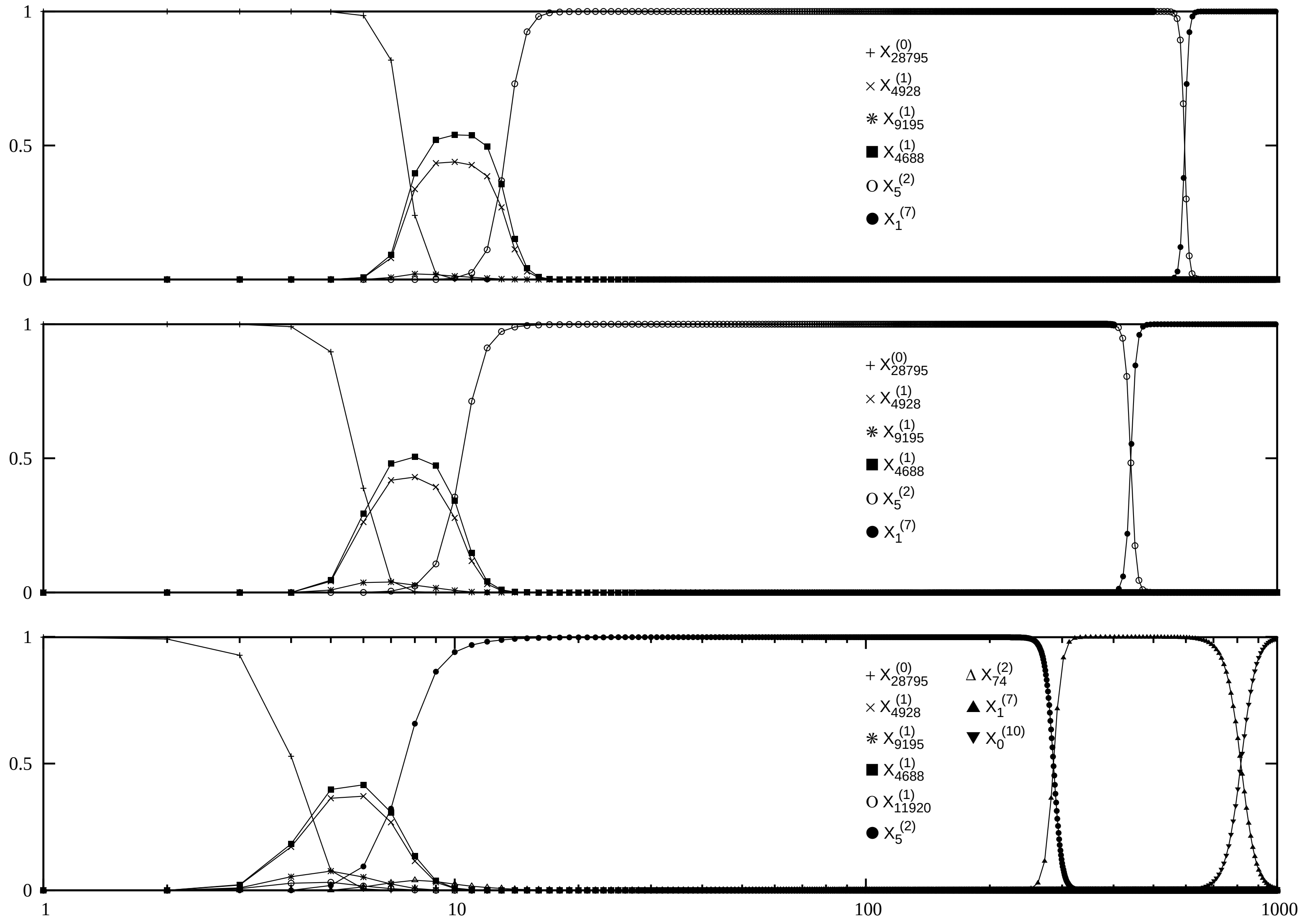
$$\vec{Z}(t+1)=\hat{E}^{(\mu)}\vec{Z}(t),\quad \hat{E}^{(\mu)}=\hat{W}\hat{M}^{(\mu)},\quad \hat{W}$$: diagonal, random

- Mutation matrix: $M^{(\mu)}(\sigma',\sigma)=\mu^{d(\sigma,\sigma')}(1-\mu)^{L-d(\sigma,\sigma')}\approx\mu^{d(\sigma,\sigma')}$

- Initial condition $Z(\sigma,0)=\delta_{\sigma,\sigma_0}$ $\Rightarrow$ after one generation $Z(\sigma,1)\sim\mu^{d(\sigma,\sigma_0)}$
  $\Rightarrow$ population fraction drops to $1/N$ at distance

$$d\approx d_{\text{eff}}=\frac{\ln N}{|\ln\mu|}$$

- Evolutionary trajectories are deterministic and independent of μ:



$$L = 15, \mu = 10^{-8}, 10^{-6}, 10^{-4}$$

The strong selection limit

JK, C. Karl, Physica A 318, 137 (2003)

- For $\mu \to 0$ mutations are irrelevant after the first generation:

$$\vec{Z}(t) = \{\hat{E}^{(\mu)}\}^t \, \vec{Z}(0) \;\;\rightarrow\;\; \{\hat{E}^{(0)}\}^{t-1} \, \vec{Z}(1)$$

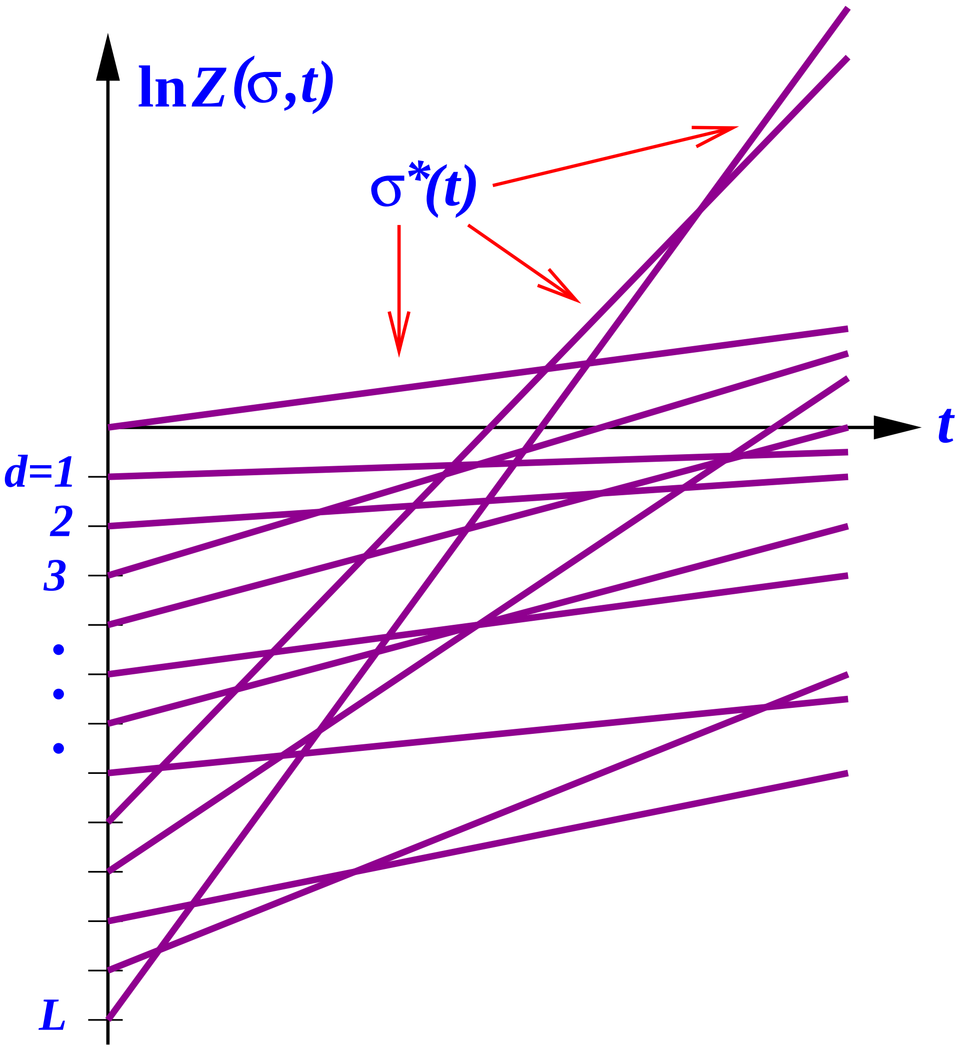
$$\Rightarrow \;\; Z(\sigma,t) \approx W(\sigma)^{t-1} W(\sigma_0) \mu^{d(\sigma,\sigma_0)}$$

- Logarithmic population variables evolve linearly in time:

$$\ln Z(\sigma,t) \approx \ln W(\sigma)(t-1) - |\ln\mu| d(\sigma,\sigma_0)$$

- Change in μ is compensated by rescaling of time with $|\ln\mu|$
- Most populated genotype $\sigma^*(t)$ is determined by maximizing $\ln Z(\sigma,t)$ with respect to σ

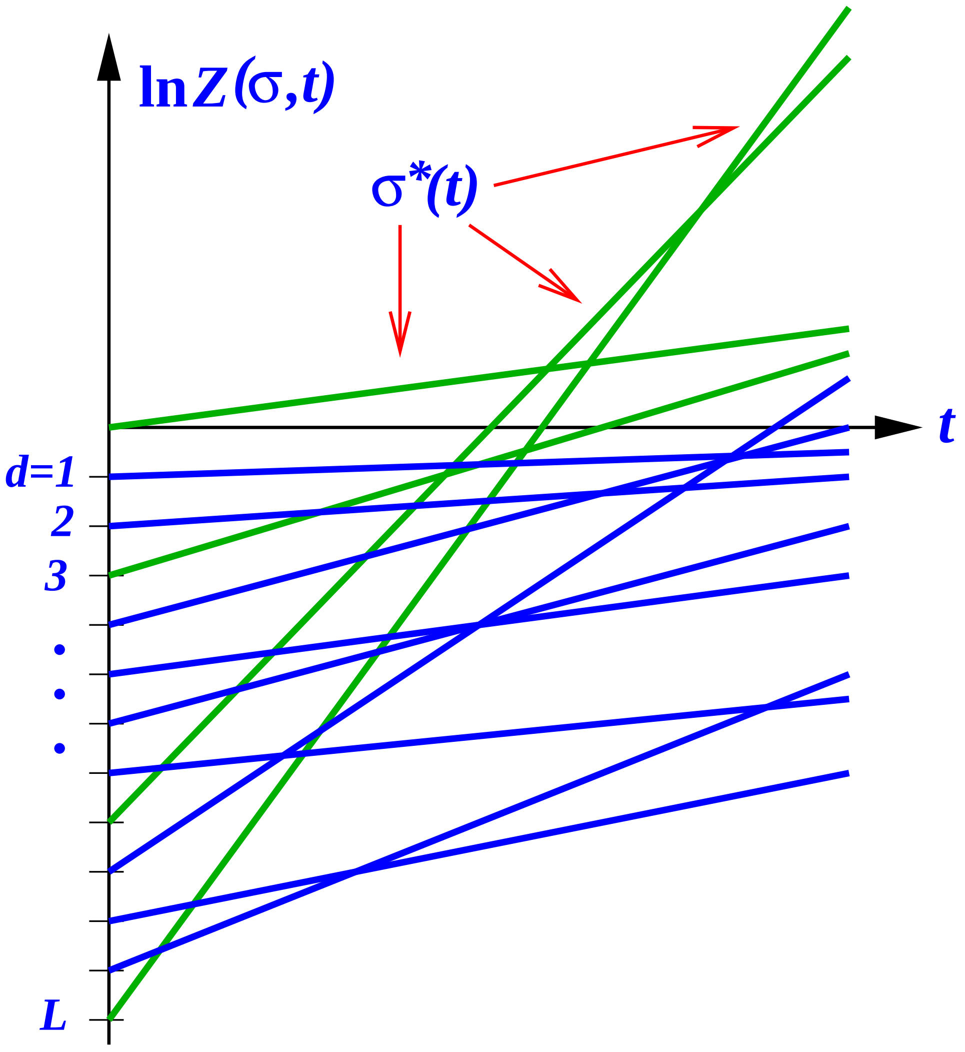
Geometry of the evolutionary race



$$\ln Z(\sigma,t) = \ln W(\sigma)(t-1) - d(\sigma_0,\sigma)$$

- Only the most fit mutant in each **shell** of constant Hamming distance $d(\sigma,\sigma_0)$ needs to be considered ⇒ system size $2^L \rightarrow L$

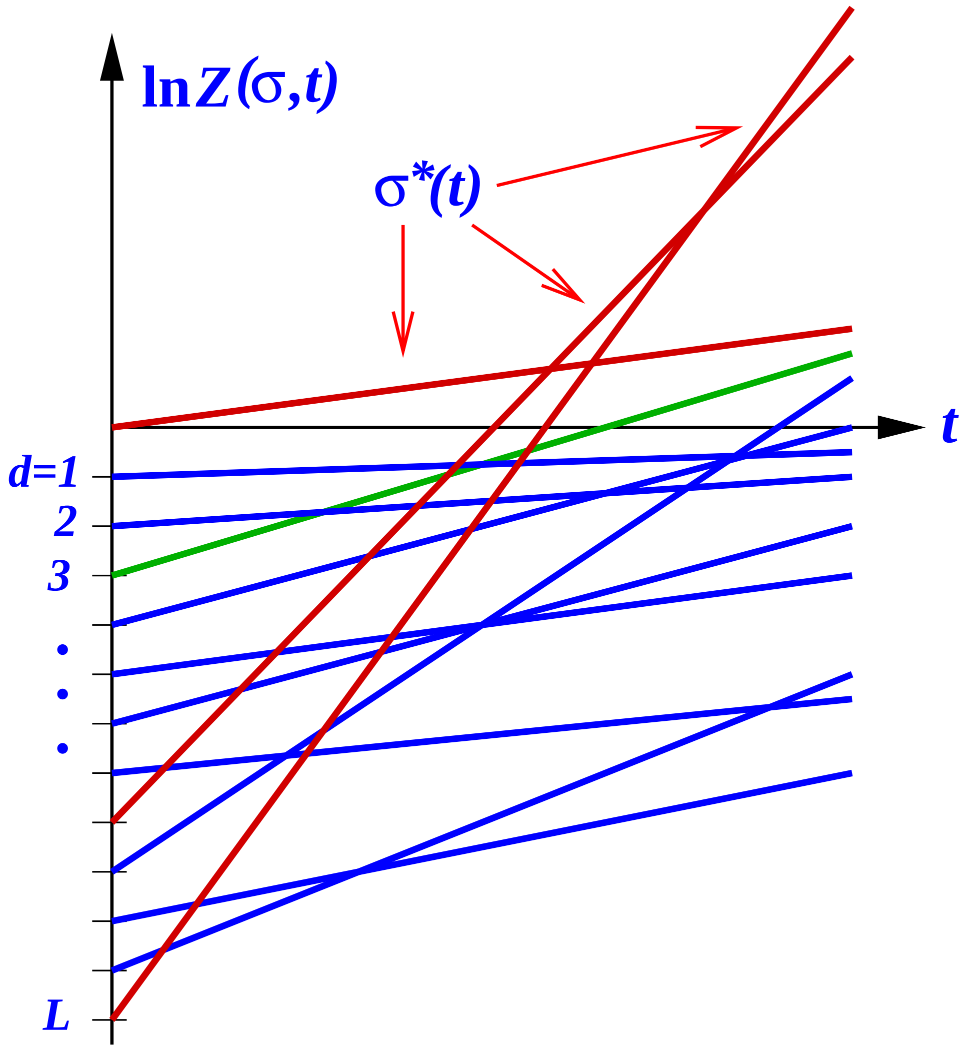
Geometry of the evolutionary race



$\ln Z(\sigma,t) = \ln W(\sigma)(t-1) - d(\sigma_0,\sigma)$

- **Contenders** and **spectators**: Contenders are fitness records σ_k with $W(\sigma_{k+1}) > W(\sigma_k)$

Geometry of the evolutionary race

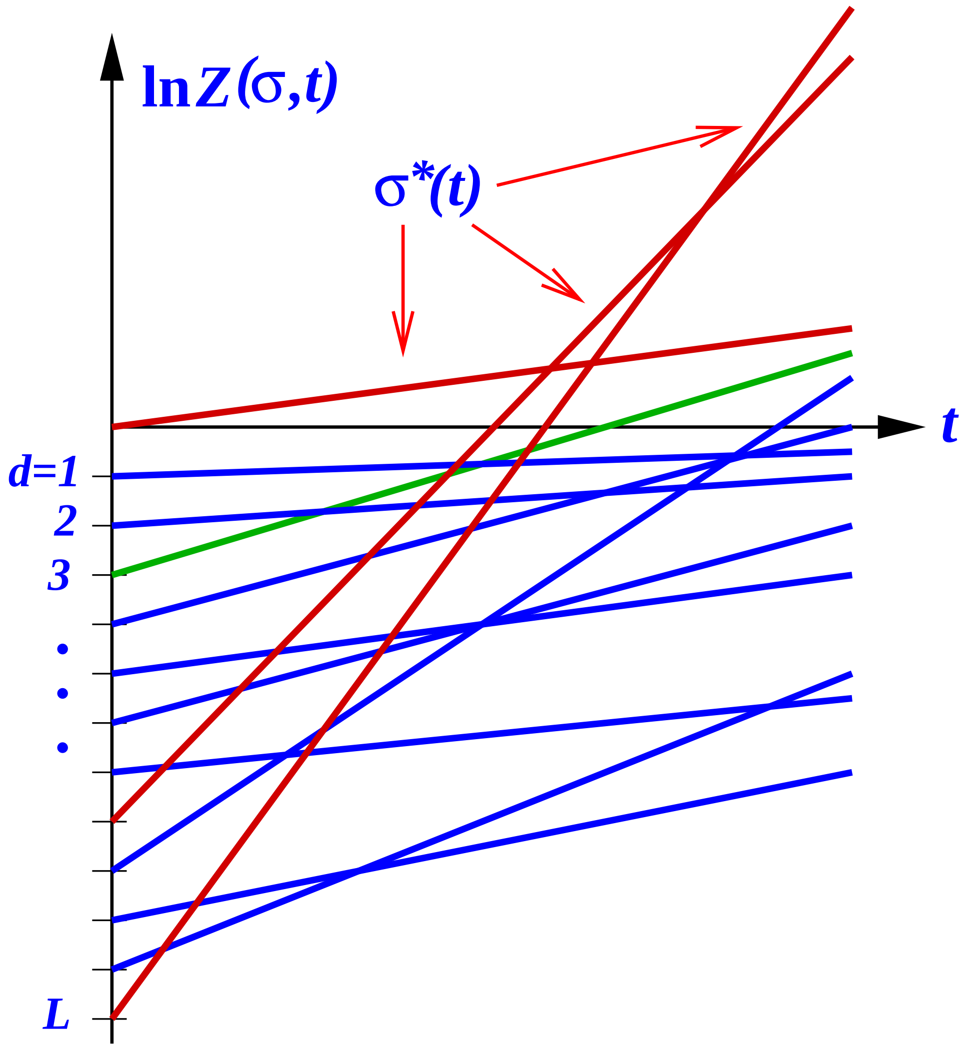


$$\ln Z(\sigma,t) = \ln W(\sigma)(t-1) - d(\sigma_0,\sigma)$$

- **Contenders** and **spectators**: Contenders are fitness records σ_k with $W(\sigma_{k+1}) > W(\sigma_k)$

- Not all contenders are **winners**; some are bypassed

Geometry of the evolutionary race



$$\ln Z(\sigma,t) = \ln W(\sigma)(t-1) - d(\sigma_0,\sigma)$$

- **Contenders** and **spectators**: Contenders are fitness records σ_k with $W(\sigma_{k+1}) > W(\sigma_k)$
- Not all contenders are **winners**; some are bypassed
- The evolutionary trajectory is composed of **non-bypassed fitness records**; how many?

Bypassing for i.i.d. shell fitnesses

- Simplification: Replace logarihmic shell fitnesses

$$F_k = \max_{d(\sigma,\sigma_0)=k} \ln W(\sigma)$$

by i. i. d. random variables from a distribution $P(F)$

- The number of records is $\ln L$ with Poisson distribution

- Numerical conjecture: Bypassing probability $1-\beta$ with $\beta < 1$ depending on the tail of $P(F)$

JK, C. Karl (2003)

(i) exponential-like: $\beta = 1/2$
(ii) power law: $\beta = (\delta - 1)/(2\delta - 1) \quad P(F) \sim F^{-1-\delta}$
(iii) bounded: $\beta = (2+\nu)/(3+2\nu) \quad P(F) \sim (F_{\max} - F)^{\nu}$

- Fluctuations in the number of non-bypassed records are sub-Poissonian

- Analytic proof by first-passage techniques

C. Sire, S.N. Majumdar, D.S. Dean, J. Stat. Mech. (2006) L07001

Bypassing in sequence space

K. Jain, JK, JSTAT (2005) P04008

- Number of sequences at distance d from σ_0 is $\binom{L}{d}$

- Probability to find a new fitness record at d: (M.C.K. Yang, 1975)

$$P(d) = \frac{\binom{L}{d}}{\sum_{k=1}^{d} \binom{L}{k}} \approx \frac{1-2d/L}{1-d/L} \quad \text{for} \quad L, d \to \infty,\ d/L < 1/2$$

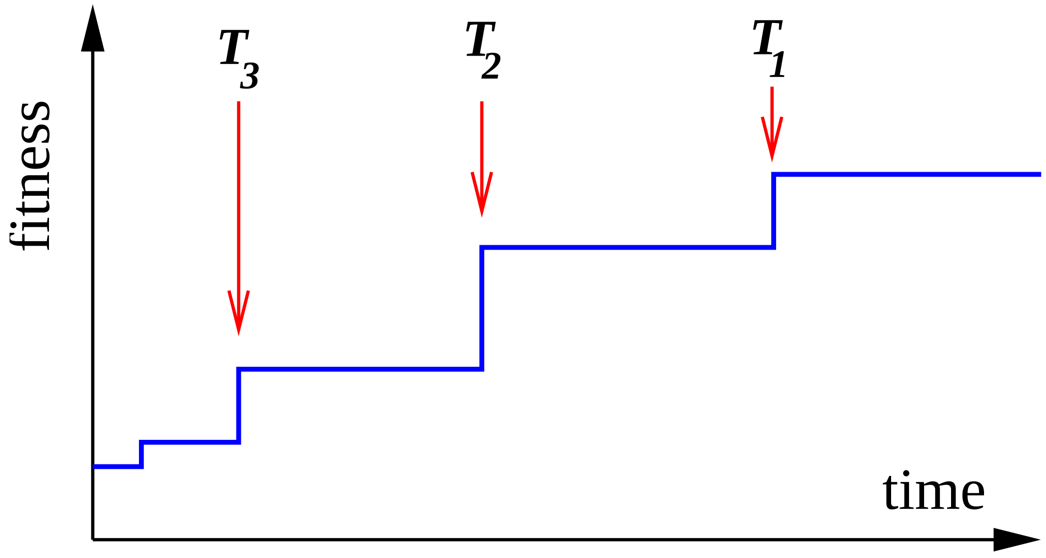
independent of the fitness distribution

- Total number of records $\mathscr{R}$ has mean and variance

$$\langle \mathscr{R} \rangle = (1 - \ln 2)N \approx 0.307\ldots N, \quad \langle \mathscr{R}^2 \rangle - \langle \mathscr{R} \rangle^2 = (3\ln 2 - 2)N \approx 0.0794\ldots N$$

- Bypassing reduces this to $\mathscr{O}(\sqrt{L})$ for exponential-like fitness distributions, and to $\mathscr{O}(1)$ for power-law fitness distributions

Distribution of evolution times



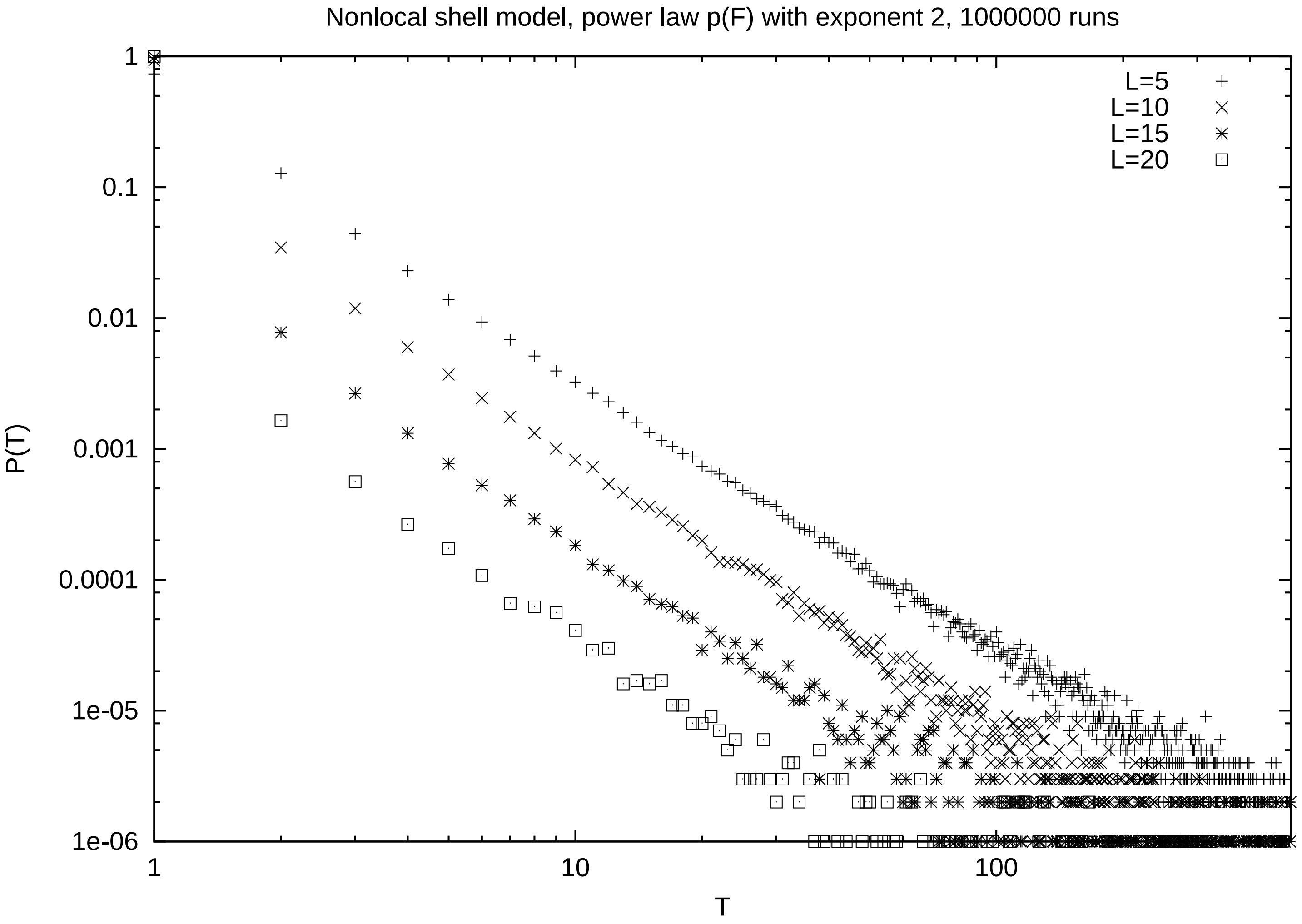
- Global fitness maximum is reached at time T_1
- **Universal** power law tail of distribution of time T_k of the k'th last jump:

$$P_k(T_k) \sim T_k^{-(1+k)}$$

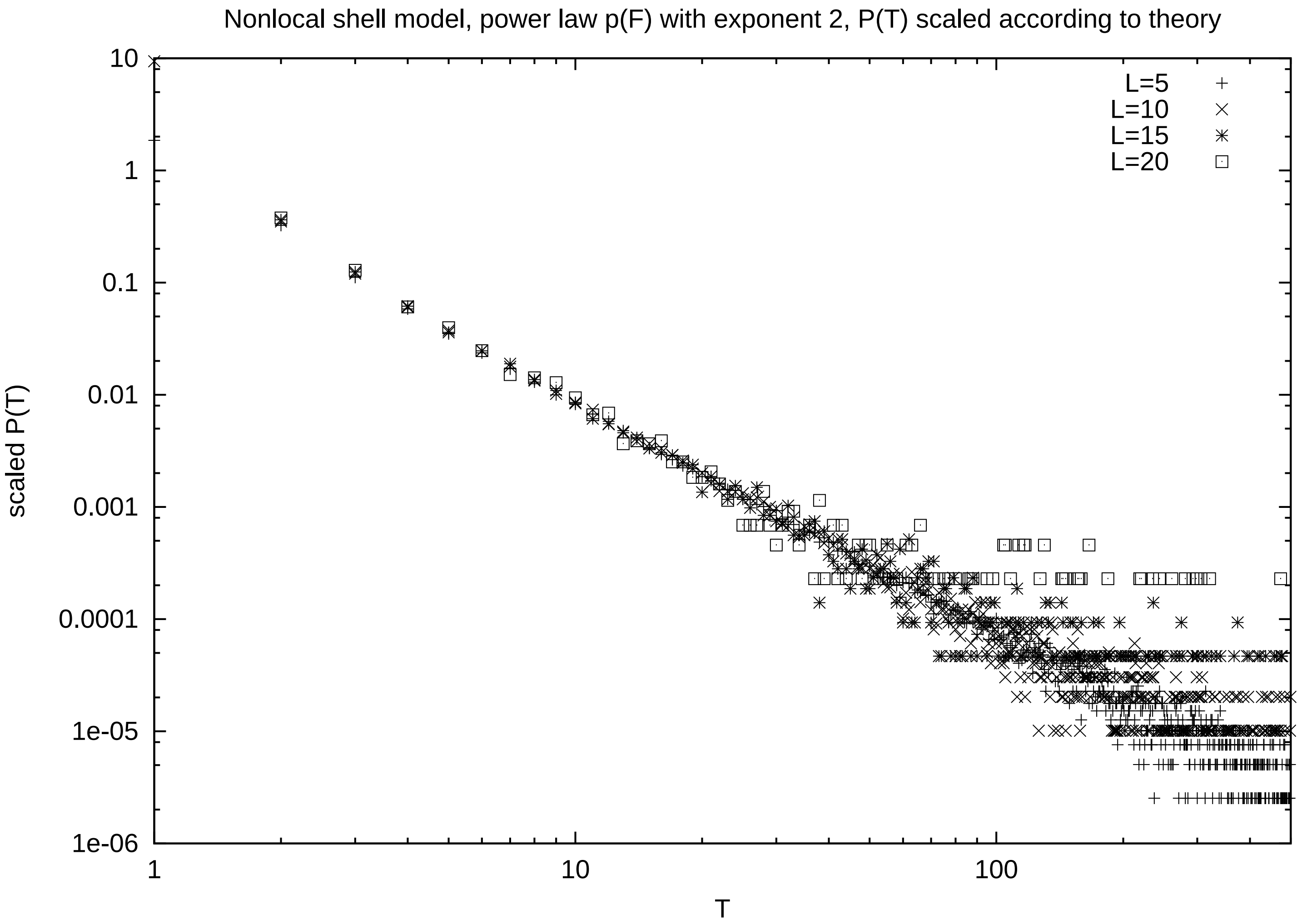
with prefactor depending on fitness distribution and sequence length.

- Expected evolution time is **infinite**: $\langle T_1 \rangle = \infty$

- Exponential fitness distribution: $P_1(T) \sim \sqrt{L}/T^2$
- Power law fitness distribution: $P(F) \sim F^{-(\delta+1)} \Rightarrow P_1(T) \sim \sqrt{L}\, 2^{-L/\delta}/T^2$



- Exponential fitness distribution: $P_1(T) \sim \sqrt{L}/T^2$
- Power law fitness distribution: $P(F) \sim F^{-(\delta+1)} \Rightarrow P_1(T) \sim \sqrt{L}\, 2^{-L/\delta}/T^2$



$\Rightarrow$ All intermediate maxima are bypassed in the power law case!

Summary

- Evolutionary biology & population genetics a rich source of stochastic many-body problems with disorder

- State of the art largely restricted to neutral evolution, small number of sites ($L = 1, 2$), non-interacting (=multiplicative) fitness landscapes

- Distinct fluctuations from mutations & sampling noise

- Different kinds of “thermodynamic limits”:
  $N \to \infty$ $\Rightarrow$ quasispecies model
  $L \to \infty$ $\Rightarrow$ infinite sites model

- Future directions (in SFB 680)
  (i) Speed of adaptation in smooth landscapes
  (ii) Timing of adaptive events in the infinite sites model
  (iii) Modes of reproduction in diploids (sexual, parthenogenetic, selfing)