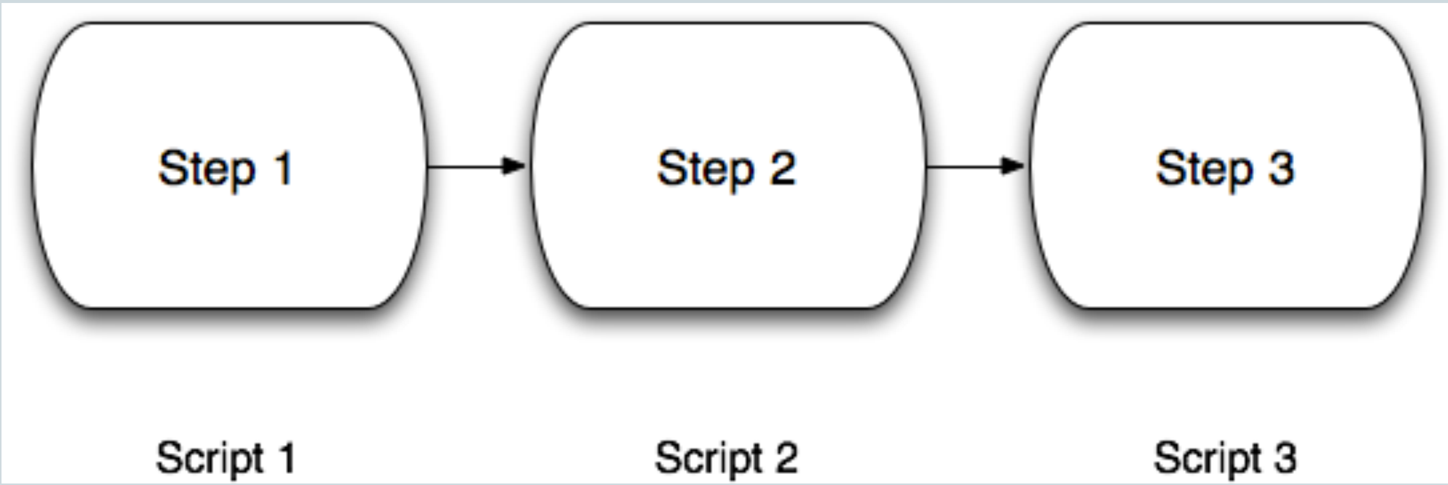
Making sense of data

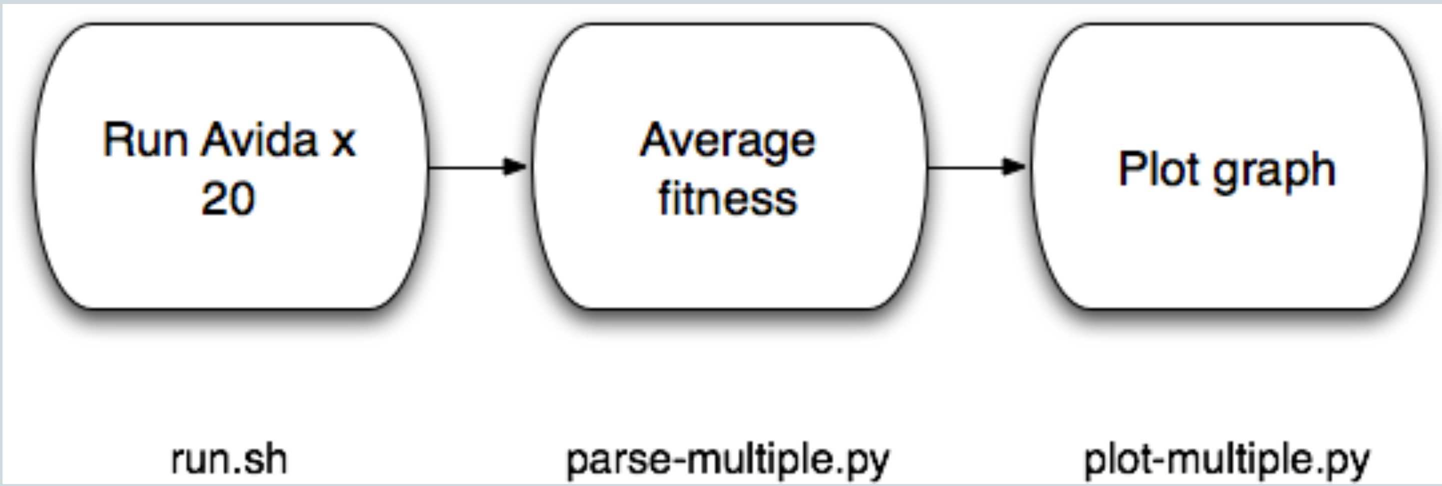
Why pipeline scripts?

- Each script is easily reusable, by you and others;
- Each script can be developed in isolation;
- Each script can be tested in isolation;
- Each script is easier to understand;
- If you plug & play properly, you can use one script in multiple pipelines.
- **Each script can be in a different language.**

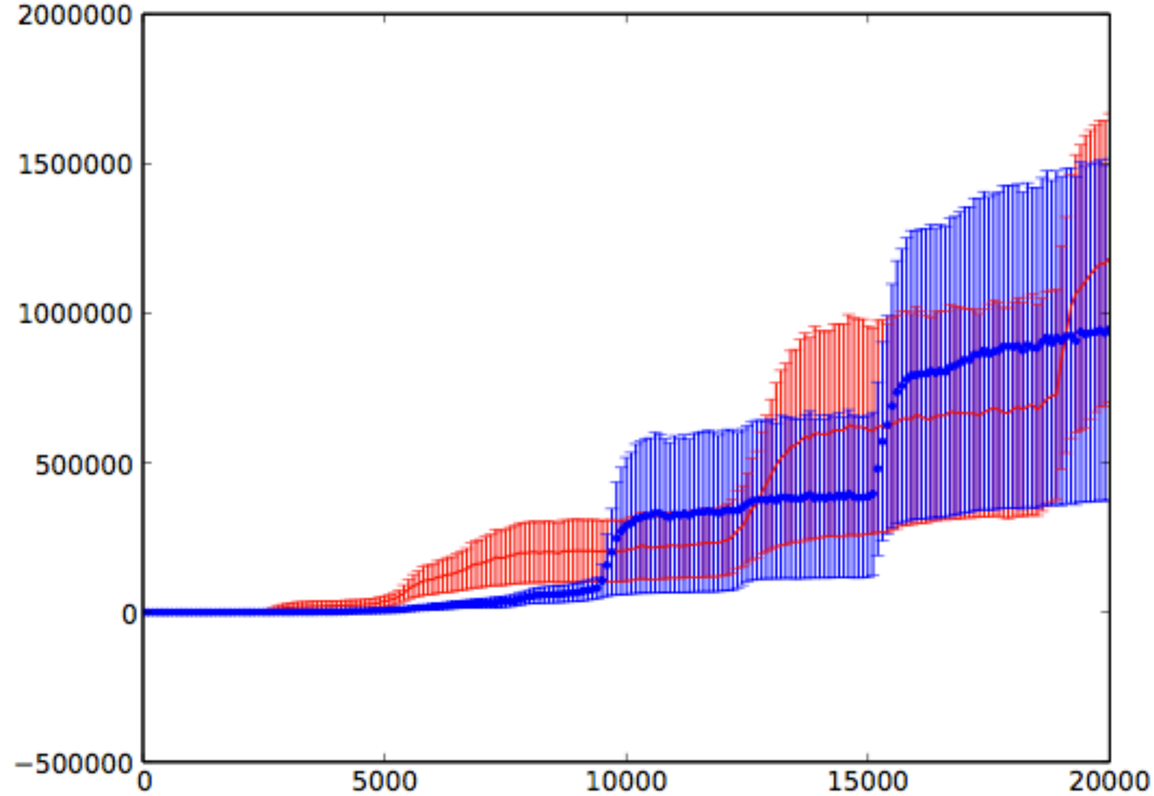
Pipelining scripts



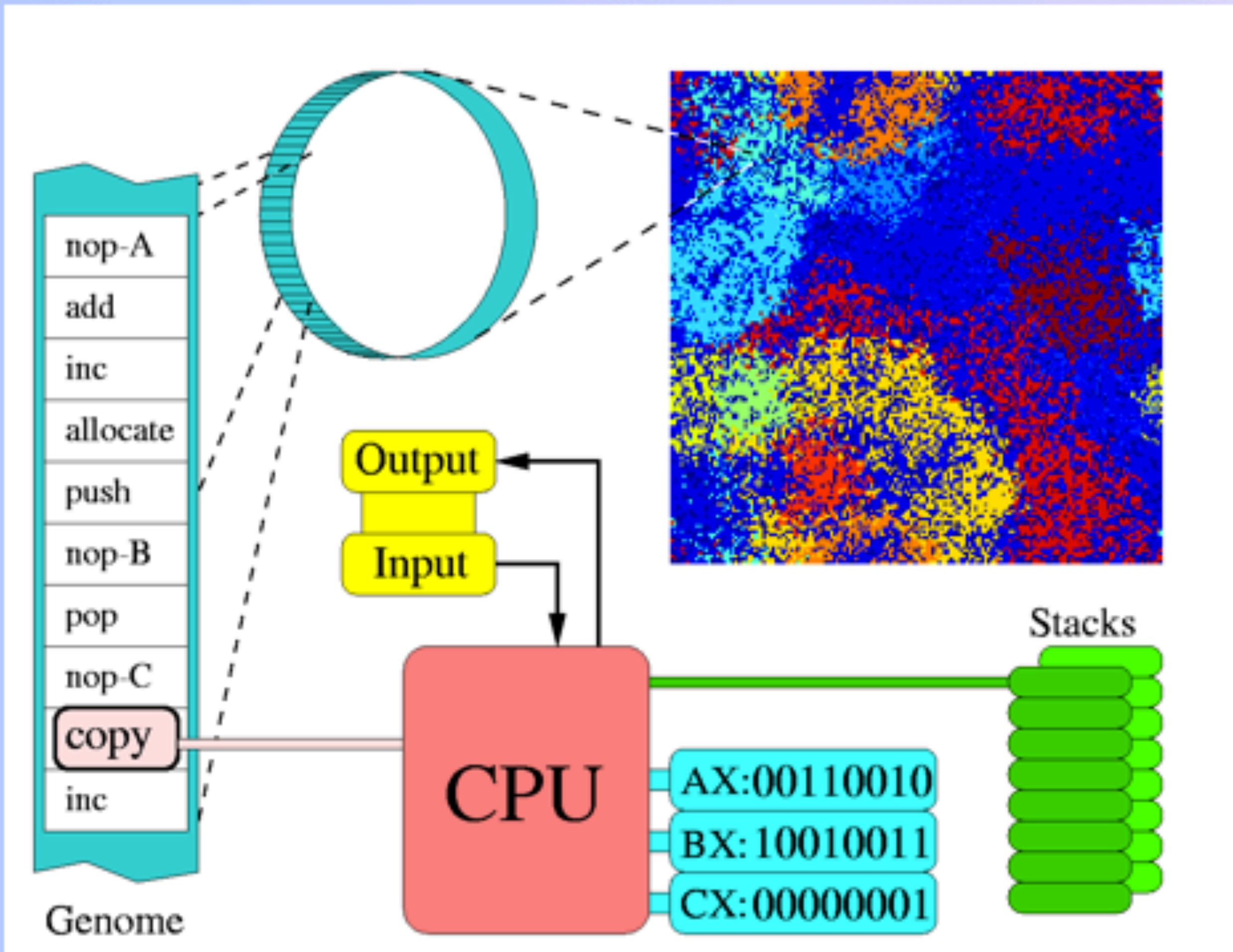
Pipelining scripts



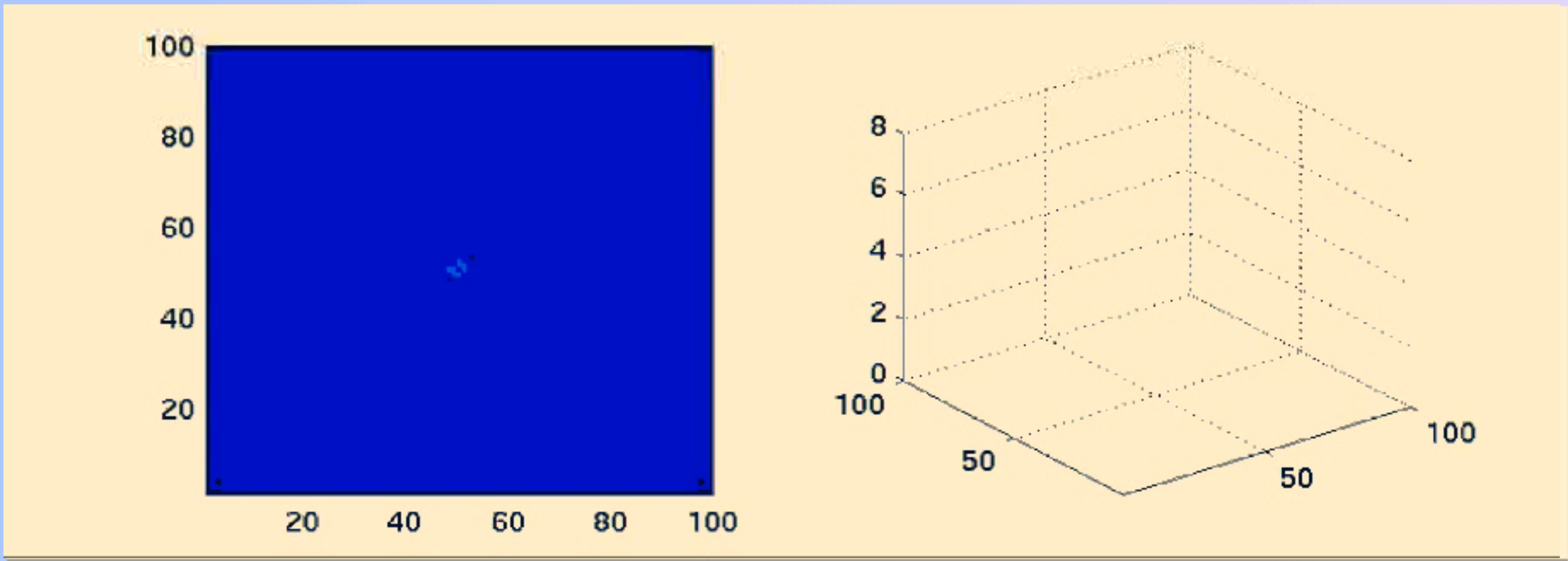
Plotting average fitness doesn't make sense…



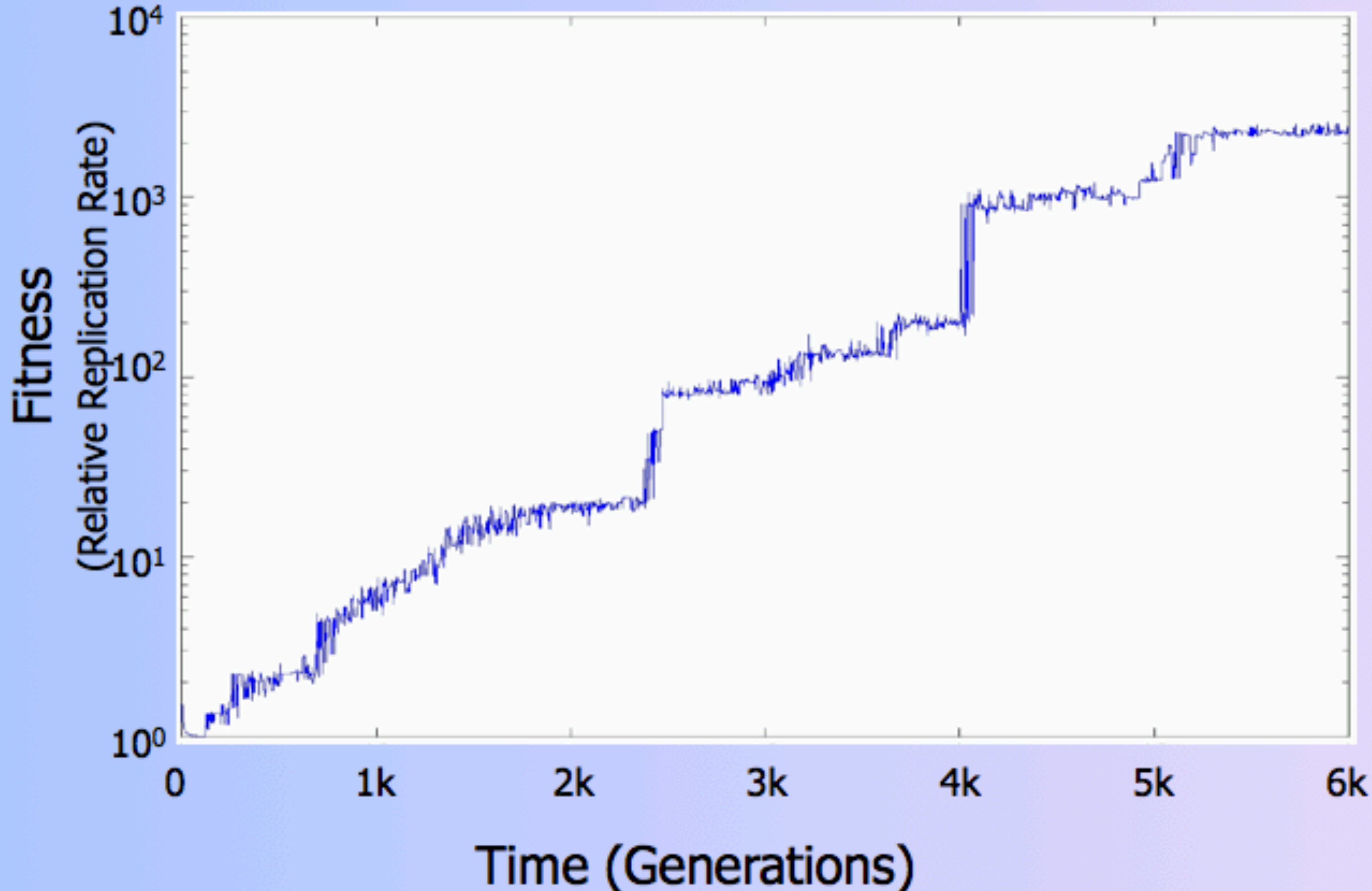
Physical World in Avida



A Typical Avida Experiment

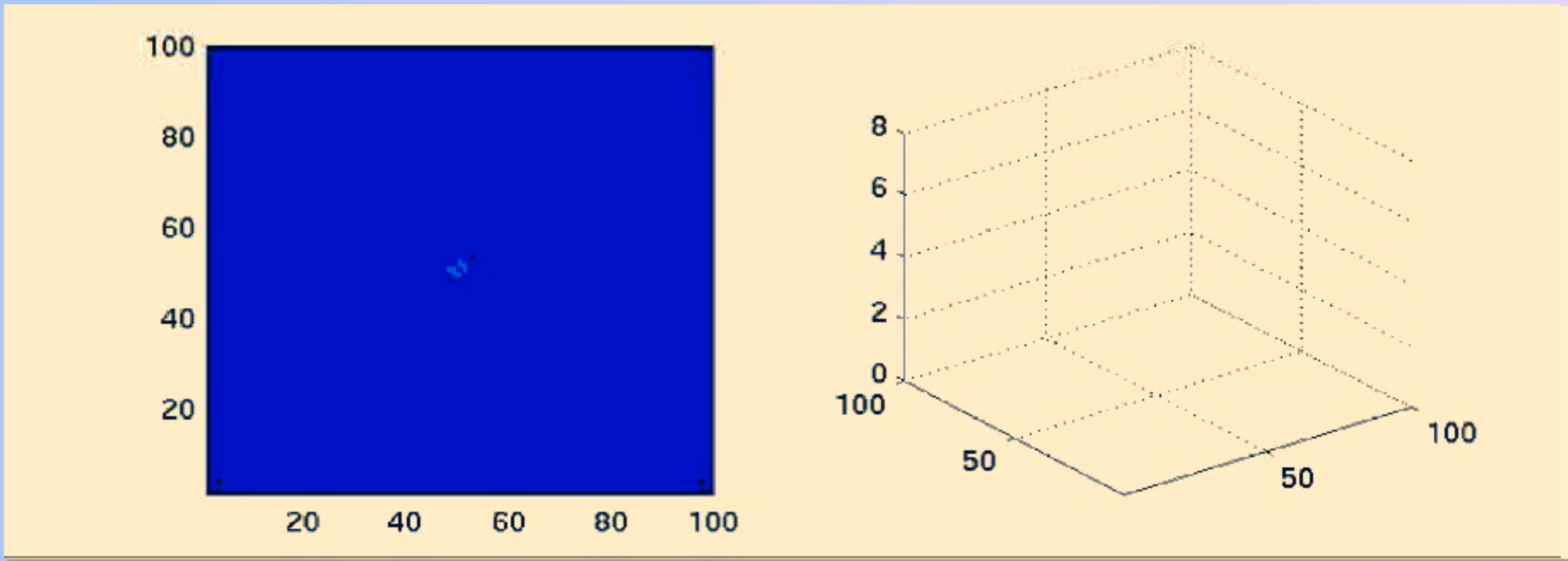


Fitness

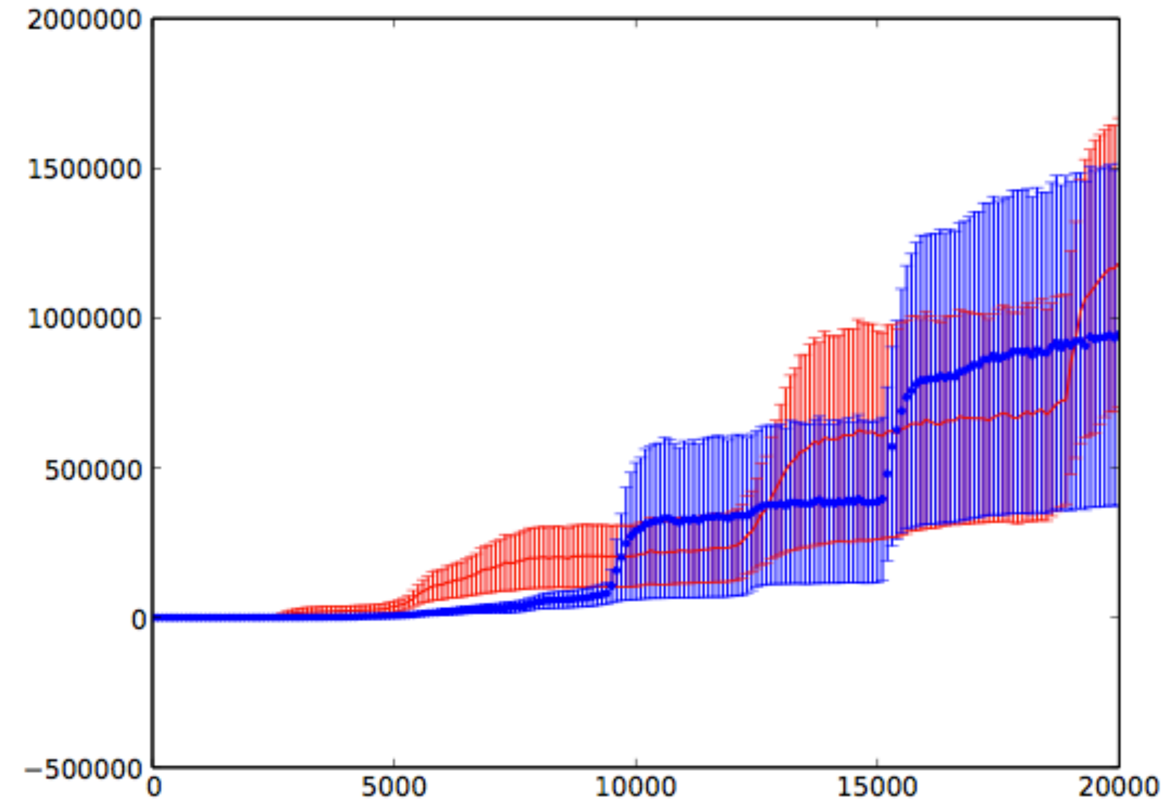


Fitness
(Relative Replication Rate)
10^4
10^3
10^2
10^1
10^0
0
1k
2k
3k
4k
5k
6k
Time (Generations)

A Typical Avida Experiment



Plotting average fitness doesn't make sense…



What *does* make sense?

Comments and notes

- Often while exploring your data, you must invent ad hoc "data analysis" techniques, especially if it's a new kind of data, or a new kind of data gathering technique.
- What's a good interval in time for measuring jumps? What's a good fitness interval for determining whether a jump has happened?
- The trick is to eventually reach an understanding of how these ad hoc choices *should* be made.
- But first, spend some time "playing" with your data.

Measuring properties of fitness progression

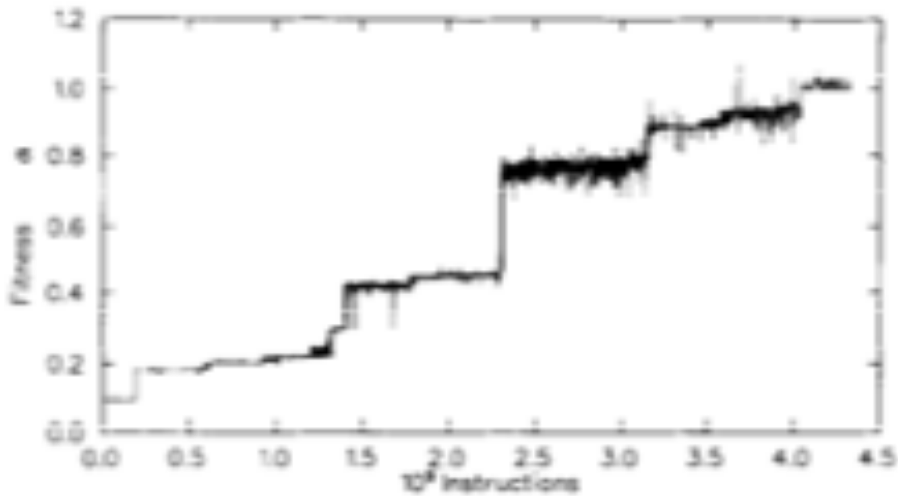


Fig. 1. Fitness curve for a typical run. The fitness parameter α of the most successful (i.e., most populous) genotype is plotted as a function of time, measured every million instructions for a mutation rate $R = 0.5 \times 10^{-8}$.

Adami, Phys. Lett. A, 1994

Fitness curve is a *power spectrum*

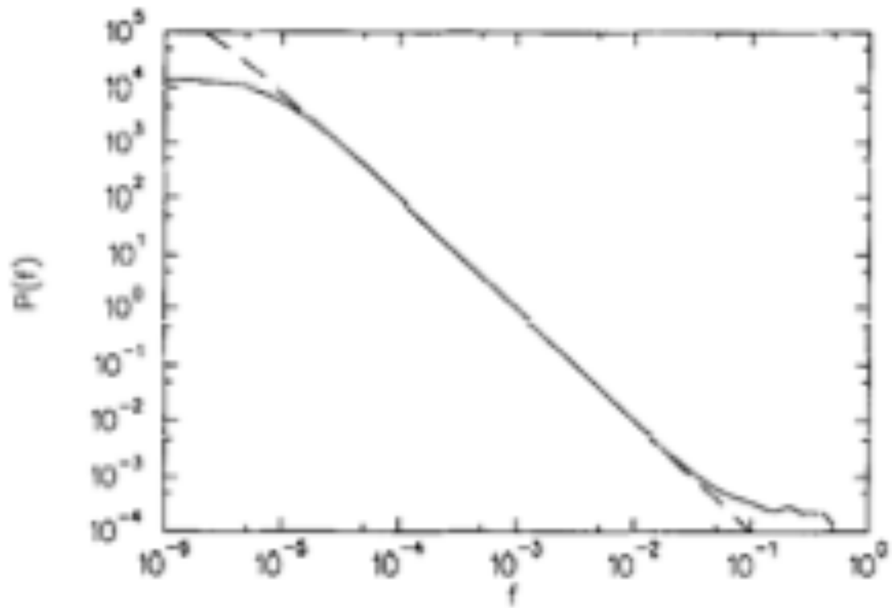


Fig. 2. Power spectrum $P(f)$ of a typical fitness curve $\alpha(t)$ (Fig. 1). The dashed line is a fit to $P(f) \sim f^{-\beta}$ with $\beta = 2.0 \pm 0.05$.

Adami, Phys. Lett. A, 1994

So is time between fitness transitions

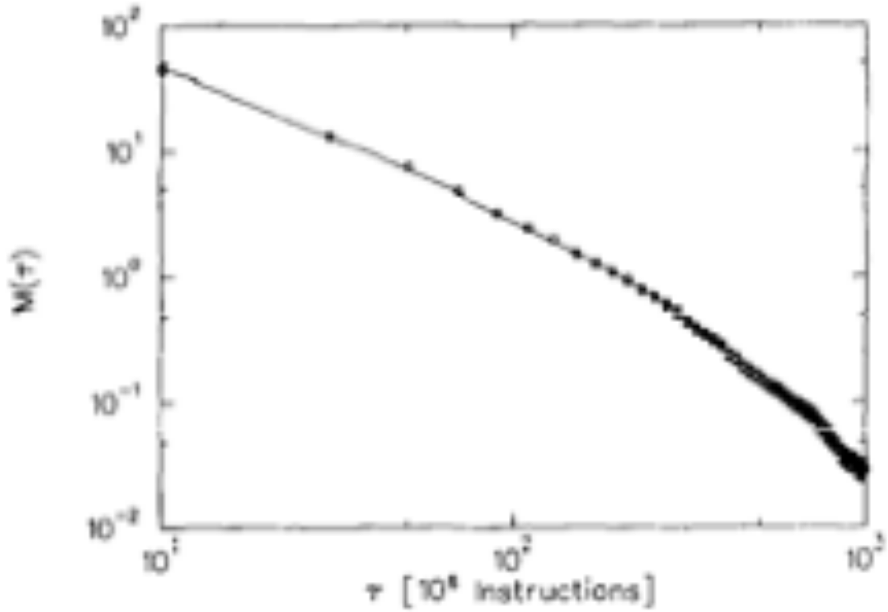


Fig. 3. Integrated distribution of times between phase transitions τ (length of epoch). The solid line is a fit to the incomplete gamma function with $\alpha = 0.6 \pm 0.1$ and a cutoff parameter $T = 540 \pm 40$ modeling finite-size effects.

Adami, Phys. Lett. A, 1994