Precision of adaptation

Chris Watkins
Department of Computer Science
Royal Holloway

Nature's no. 1 learning algorithm: asexual evolution

Nature's no. 2 learning algorithm: sexual evolution

Machine learning algorithms ….

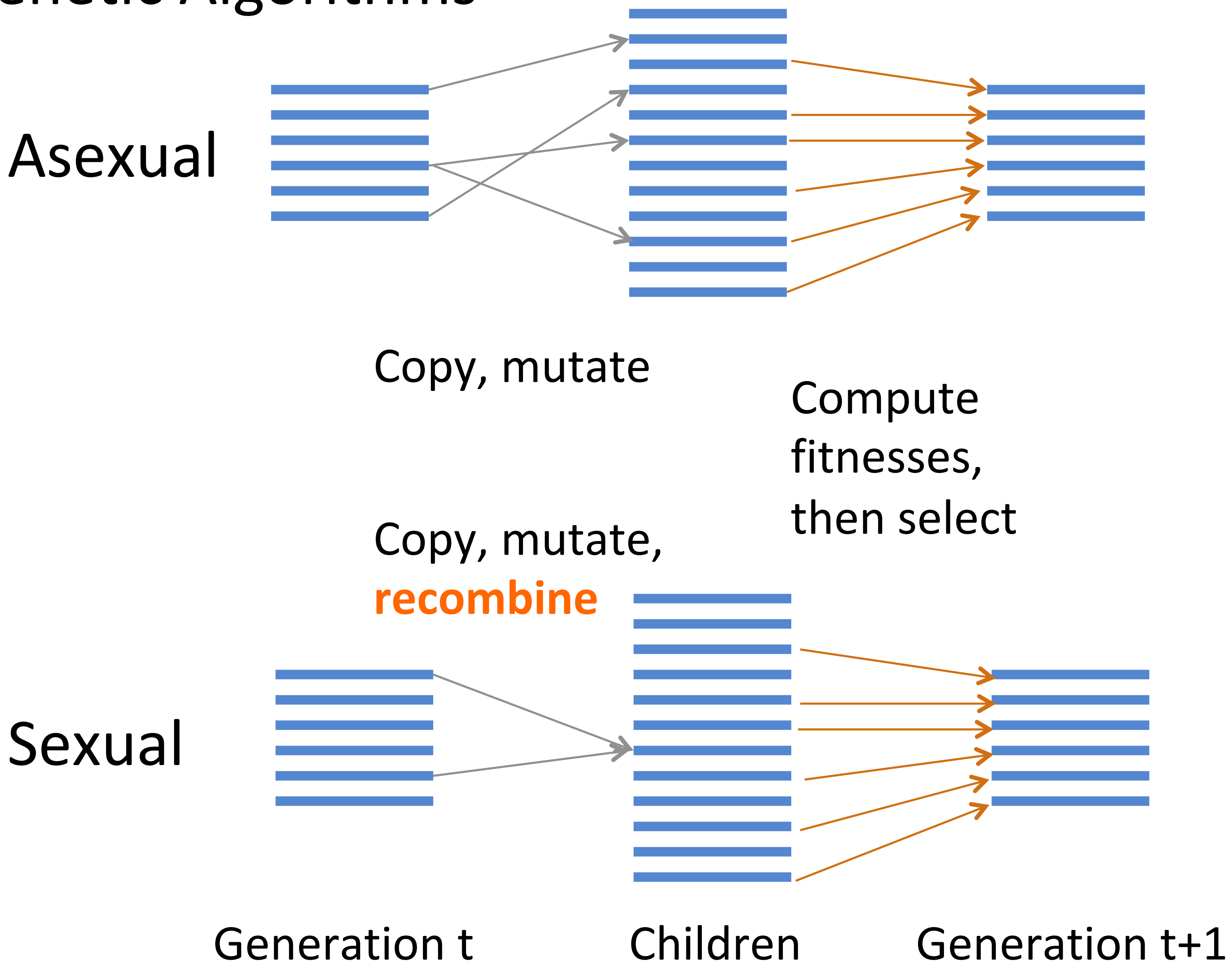
… any connection?

Evolution ≅ nonparametric Bayesian MCMC

Upper bounds on evolvability via a communication game argument

- Sexual evolution has *much* higher channel capacity than asexual
- To use this channel capacity requires distributed genetic codes

Genetic Algorithms

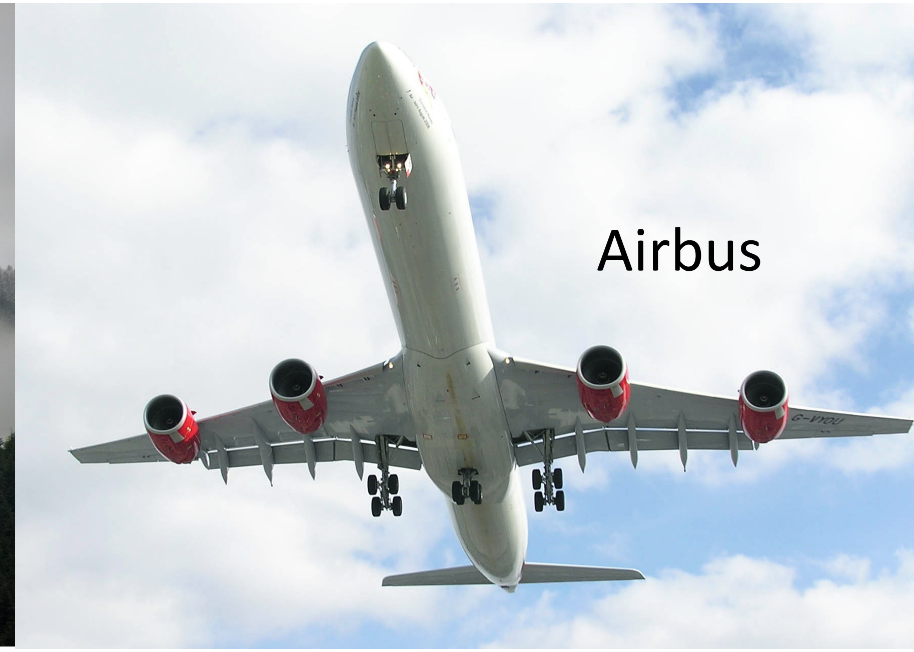


Differences between asexual and sexual organisms

Each difference is by multiple orders of magnitude

	Asexual	**Sexual**
Genome size:	Small	Large
Population size:	Large	Small
Generation time:	Short	Long
Gene regulation:	Simple	Complex
Development:	None	Complex

Horses and donkeys diverged 4 million years ago. Mules are stronger than both.



Can we breed a Boeing with an Airbus by randomly recombining the design files using cut and paste, with ~100 random changes?

Human-designed machines and biological organisms are specified in **very** different ways.

The ‘Omnigenic’ hypothesis

Complex traits – such as height – that are the result of long processes of development are affected by genetic variation at millions of loci spread all over the genome.

(Loosely) nearly every part of the genome affects every complex trait.

An Expanded View of Complex Traits: from Polygenic to Omnigenic, Boyle, Li, and Pritchard, Cell 2017

Limitative informational properties of sexual and asexual evolution: a simple model

Consider genomes as binary sequences of fixed length L.

Choose a fixed binary sequence as a ‘target’ or ‘ideal’ genome

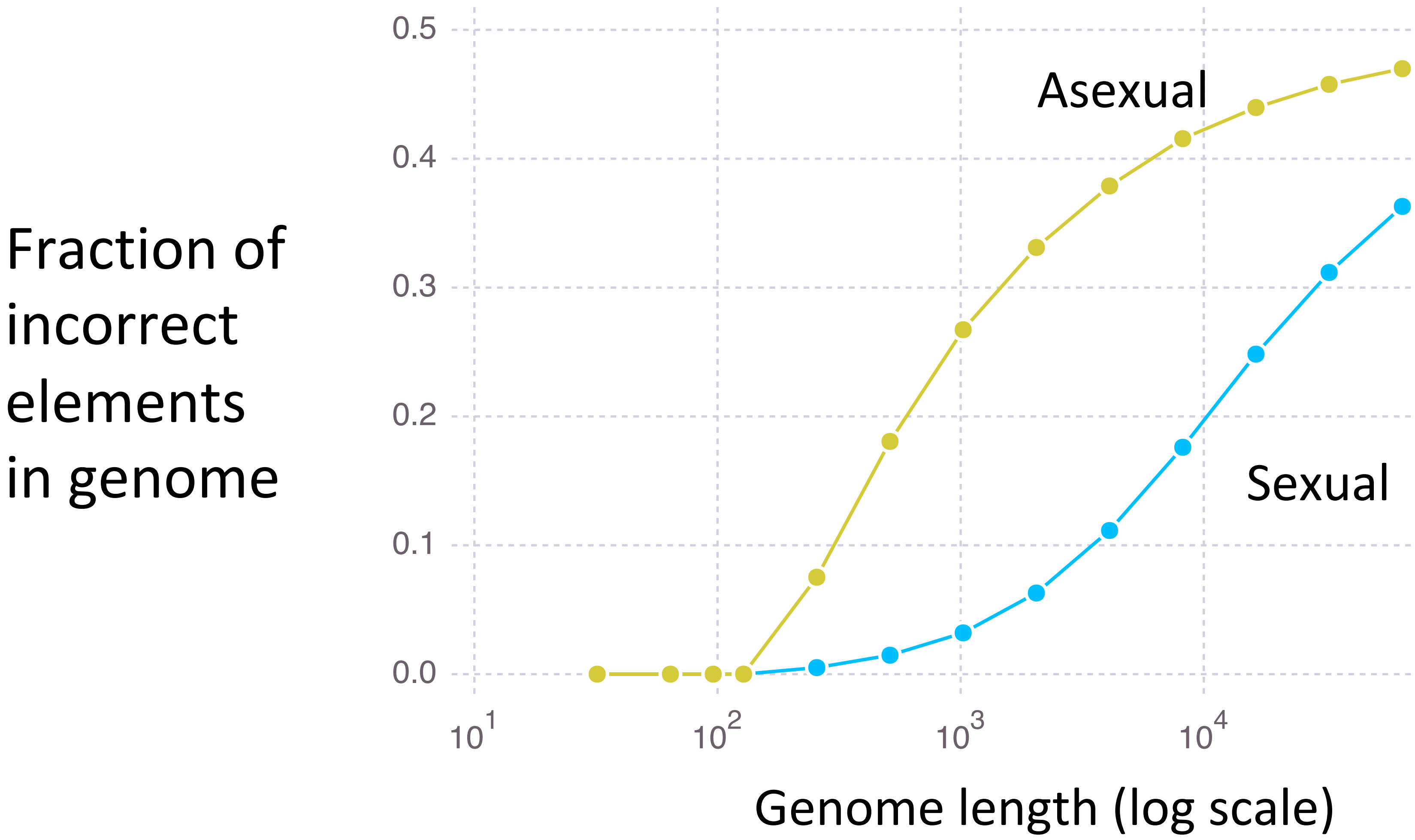
In each generation, breed 2N ‘children’, and then select the N that are closest to the target string.

Not intended to be a realistic model: it is a limitative model.

Every assumption is optimistic: in this way, we can get *upper bounds* on achievable complexity.

Bit-error-rate vs genome length

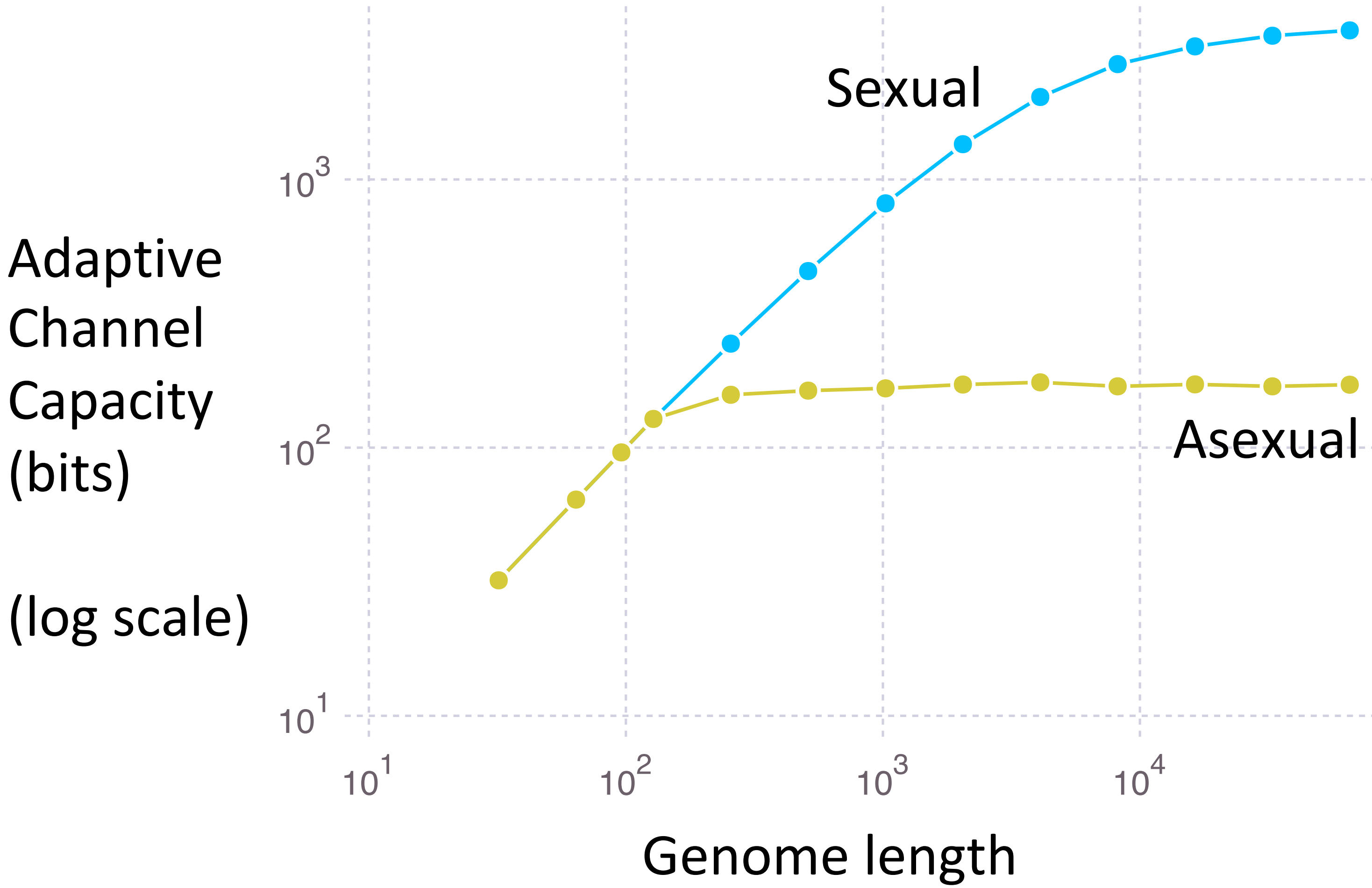
Population size: N = 100 Mutation rate: u = 0.01



Same data replotted as channel capacity vs genome length

$$\text{Channel Capacity} \geq L(1 - h_2(p))$$

where p is error rate, and $h_2(p)$ is entropy in bits



The Drosophila Game: a thought-experiment

Suppose two geneticists – A and B – wish to communicate.
First, they agree on a code-book of varieties of Drosophila

- could use binary traits: red/black eyes, long/stubby wings, smooth/hairy, etc
- could use a binary code on polymorphic SNPs

A decides on the message to send.
She then breeds a population of N Drosophila in her office; after each generation she has 2N, and she squashes half of them, and continues to breed the rest.

By keeping the flies most similar to the variety she wants, A can breed her population to have the binary trait values she desired.

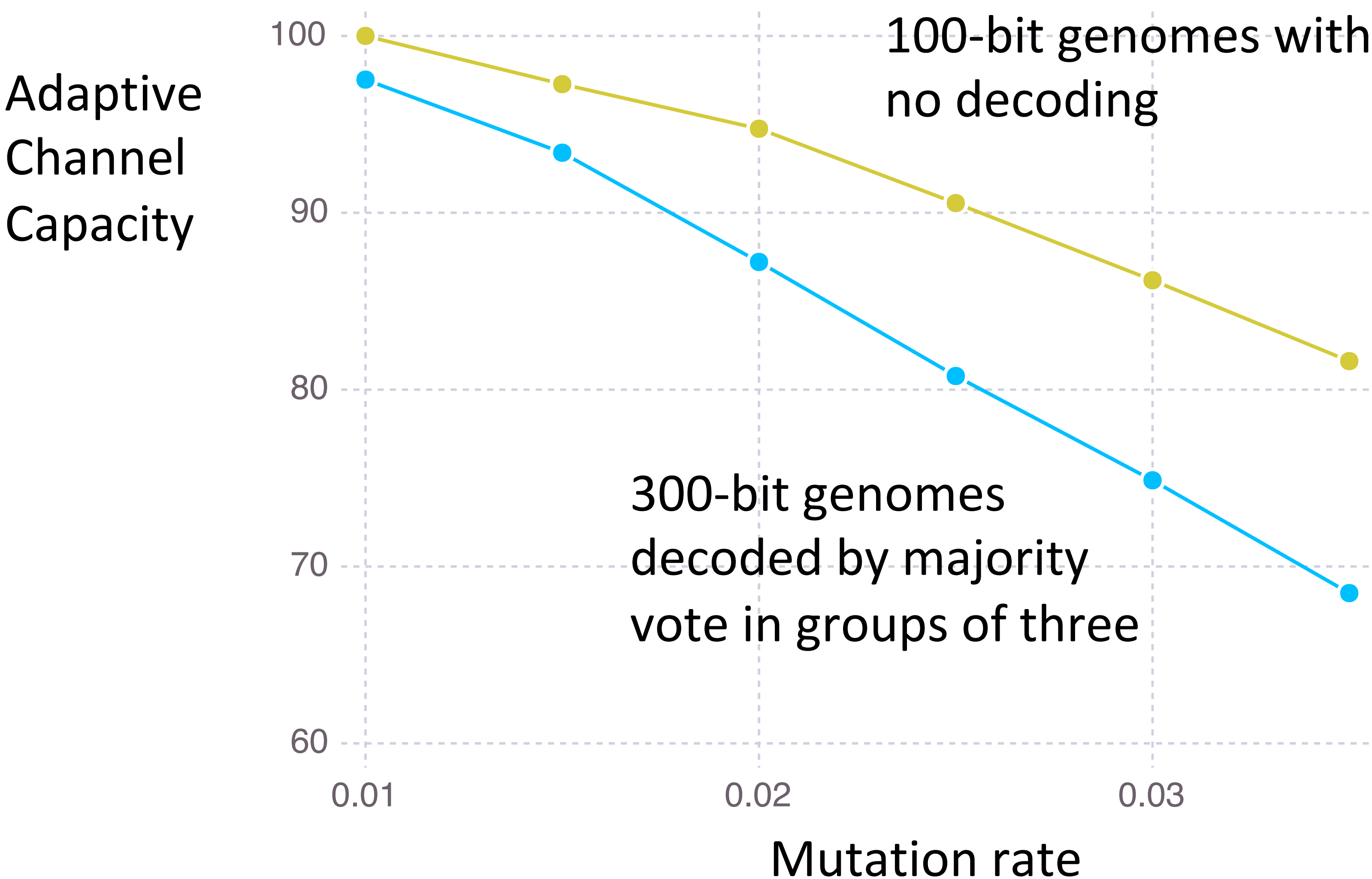
Finally A sends her message: she releases her final selected population of N flies, and B catches one.

B examines (or sequences) the fly, and determines which variety in the code-book it belongs to, and so decodes A's message.

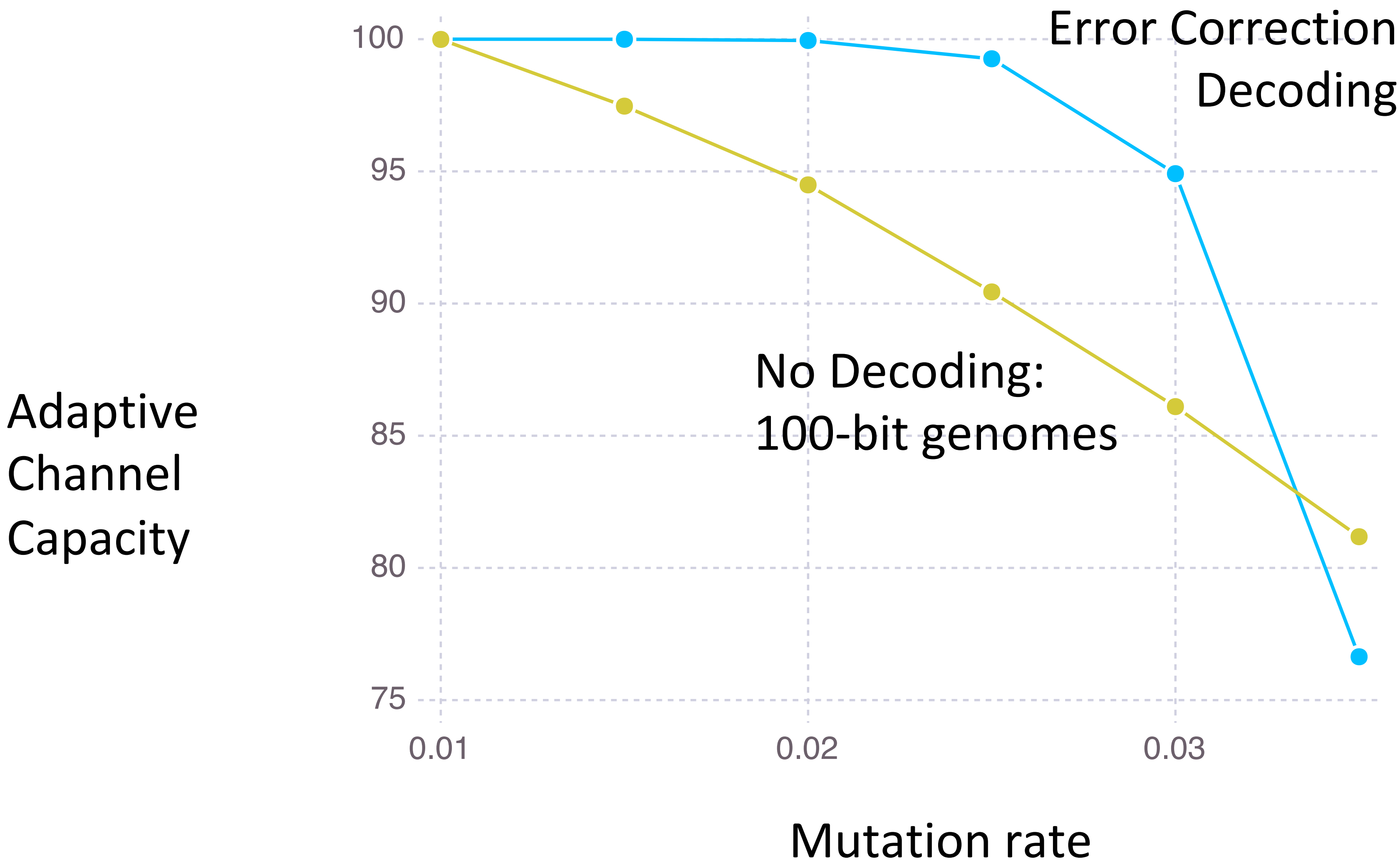
If A can reliably set, say, 10 different binary traits, she can produce 1024 different varieties of fly, and so send 10 bits.

The **channel capacity** measures how precisely the flies can adapt to A's selective breeding

Simple redundancy: 300 bit genome, decoded to 100 bits by majority votes on 100 separate groups of 3 bits.

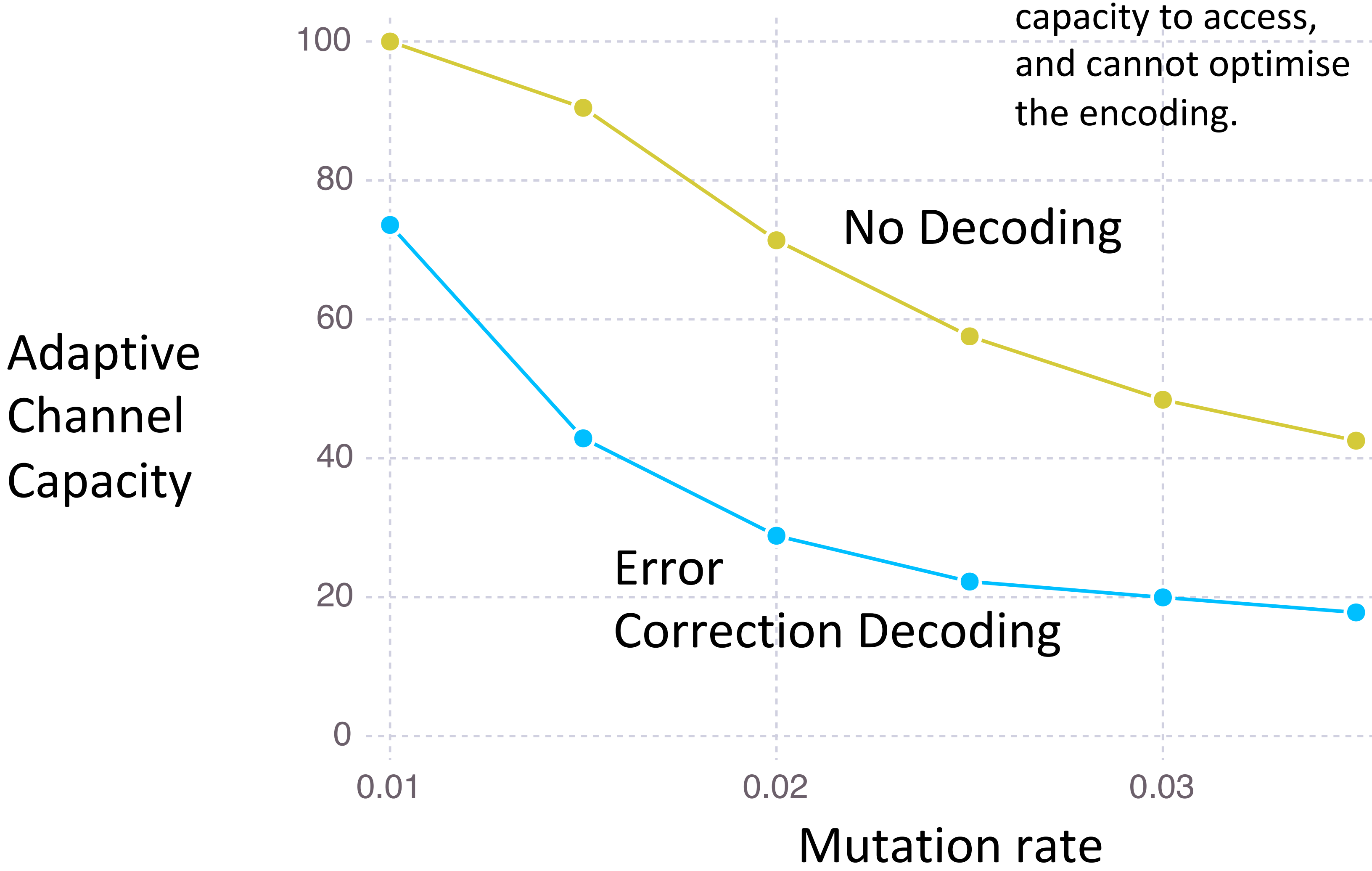


Non-linear decoding: 300 bit genomes, decoded to 100 bits using much-simplified LDPC decoding



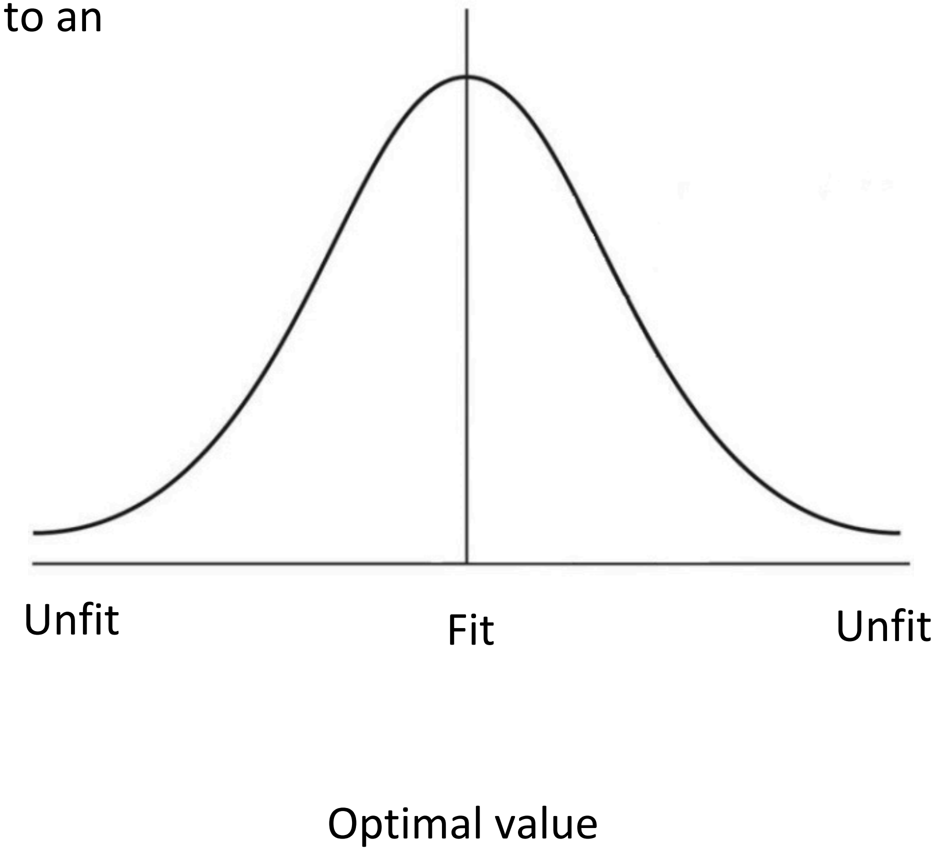
LDPC decoding compared with no decoding for **Asexual GA.**

Asexual GA has no adaptive channel capacity to access, and cannot optimise the encoding.



Stabilising selection: an important case

Selection that maintains a 'quantitative trait' – a design parameter – close to an optimal value



Decoding from a genome to a number

Simplest model: linear mapping

View genome as a vector g

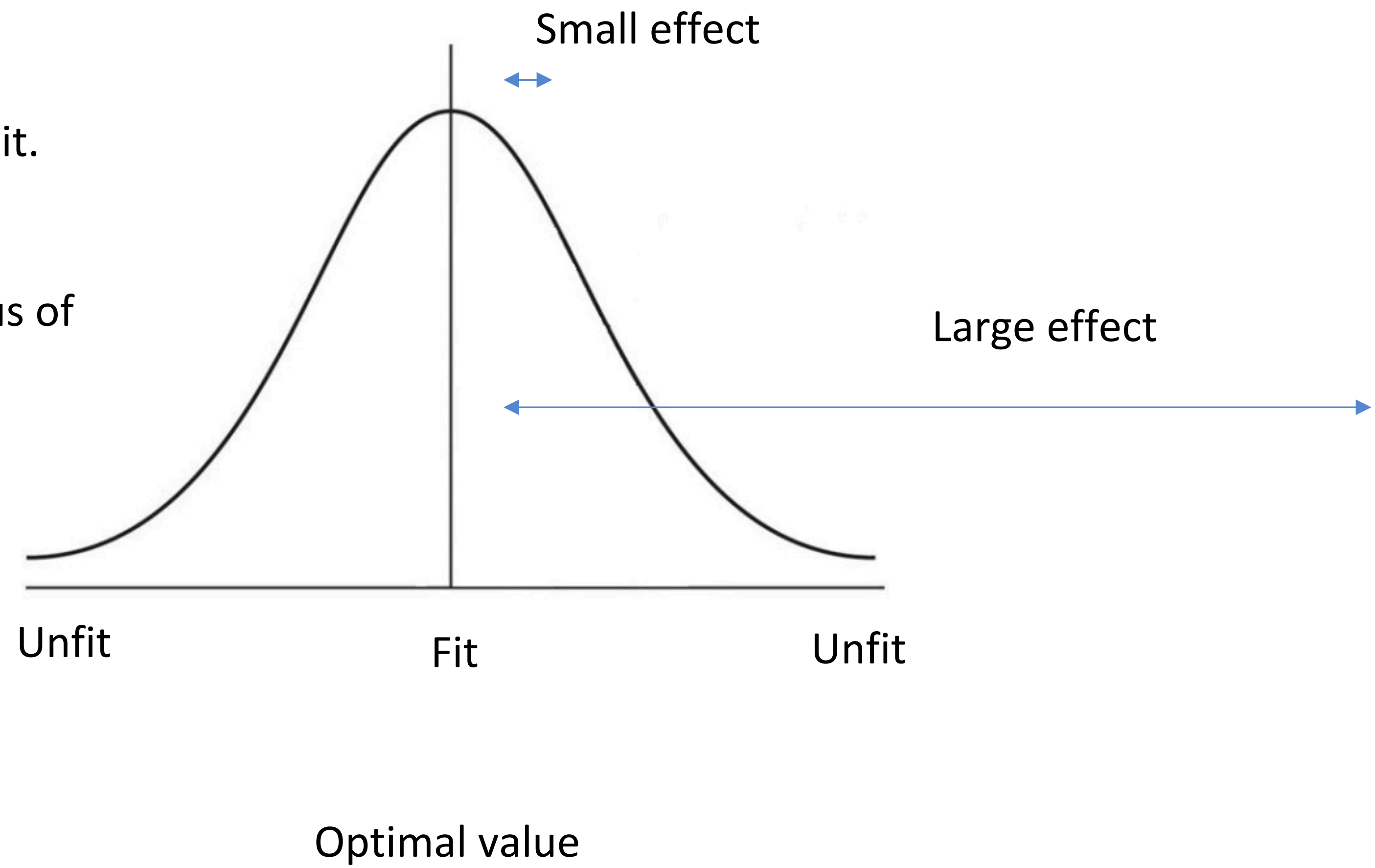
A quantitative trait x is determined as the sum of effects of each element of genome

$$x = \sum_{i=1}^{L} g_i \delta_i$$

Stabilising selection: small and large effects

A 'small' effect is smaller than the natural range of variation of the trait.

Hard to select one value for gene locus of small effect!



Stabilising selection for two values

Population size: 100
Genome length: 400

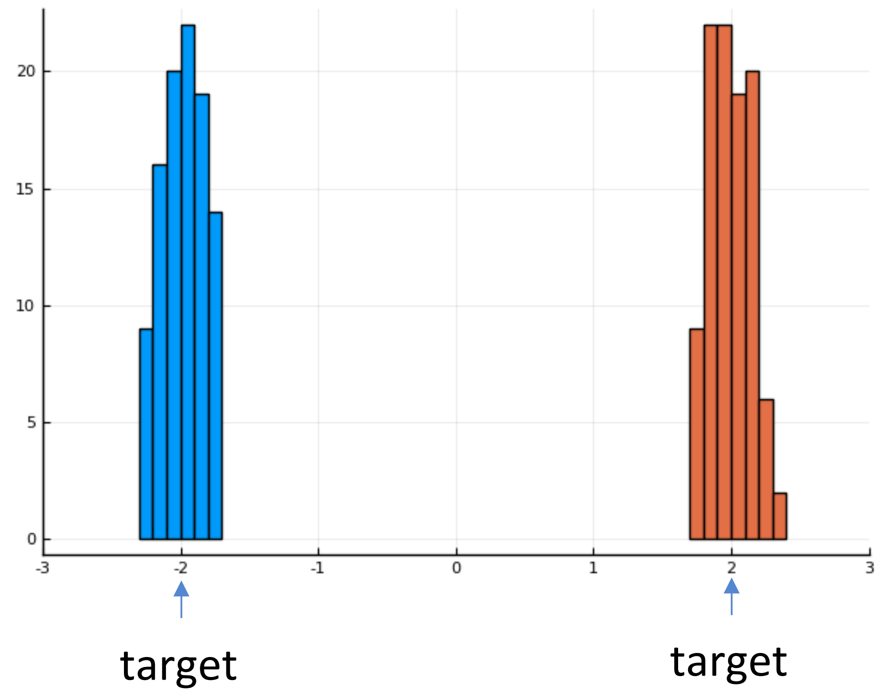
Mutation rate: 0.01

Truncation fraction: 1/3

Linear effects are normally distributed, scaled to produce range of variation shown.

Maximum range of adaptation:
-3 to 3, as shown

Precision of adaptation: approx 6 bits



Histogram of values of the two traits for members of a population at equilibrium.

Better than linear decoding? Spiral decoding

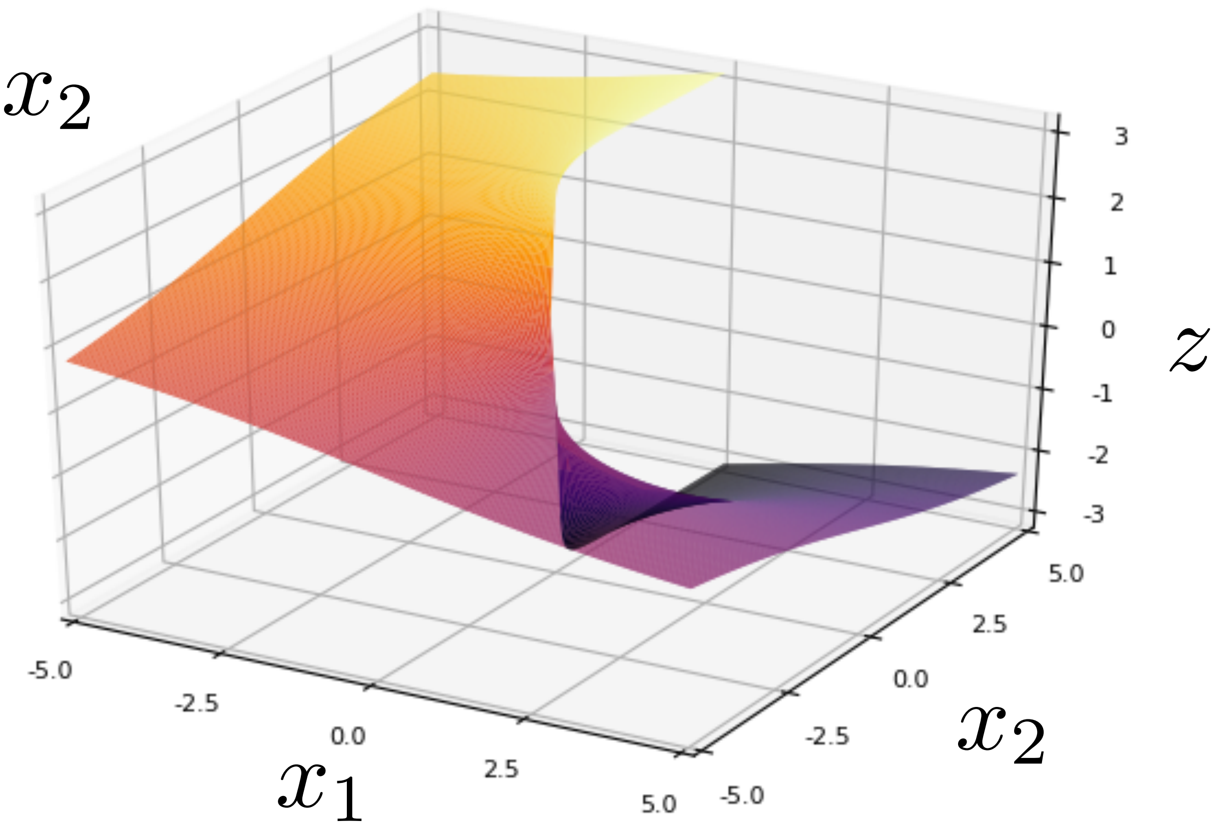
Idea: linearly decode

genome to two

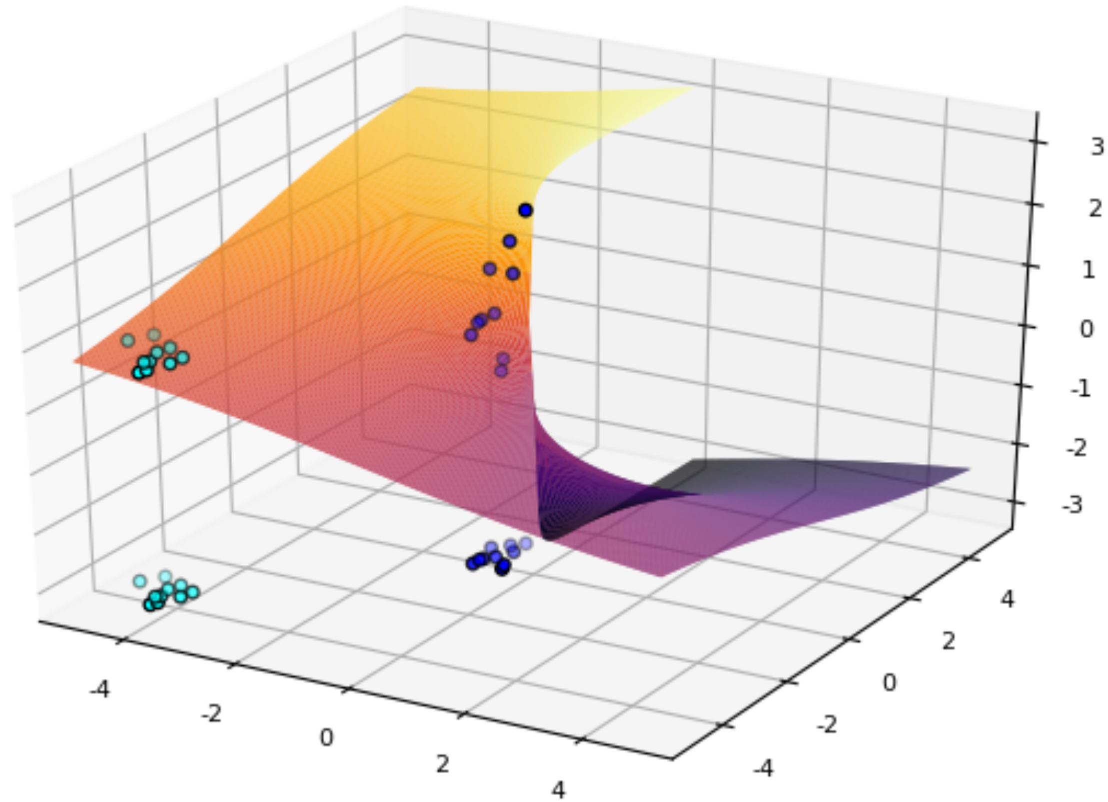
numbers x_1 and x_2

then compute

$$z = f(x_1, x_2)$$



Imprecise values of x_1 and x_2 can yield a more precise value of z



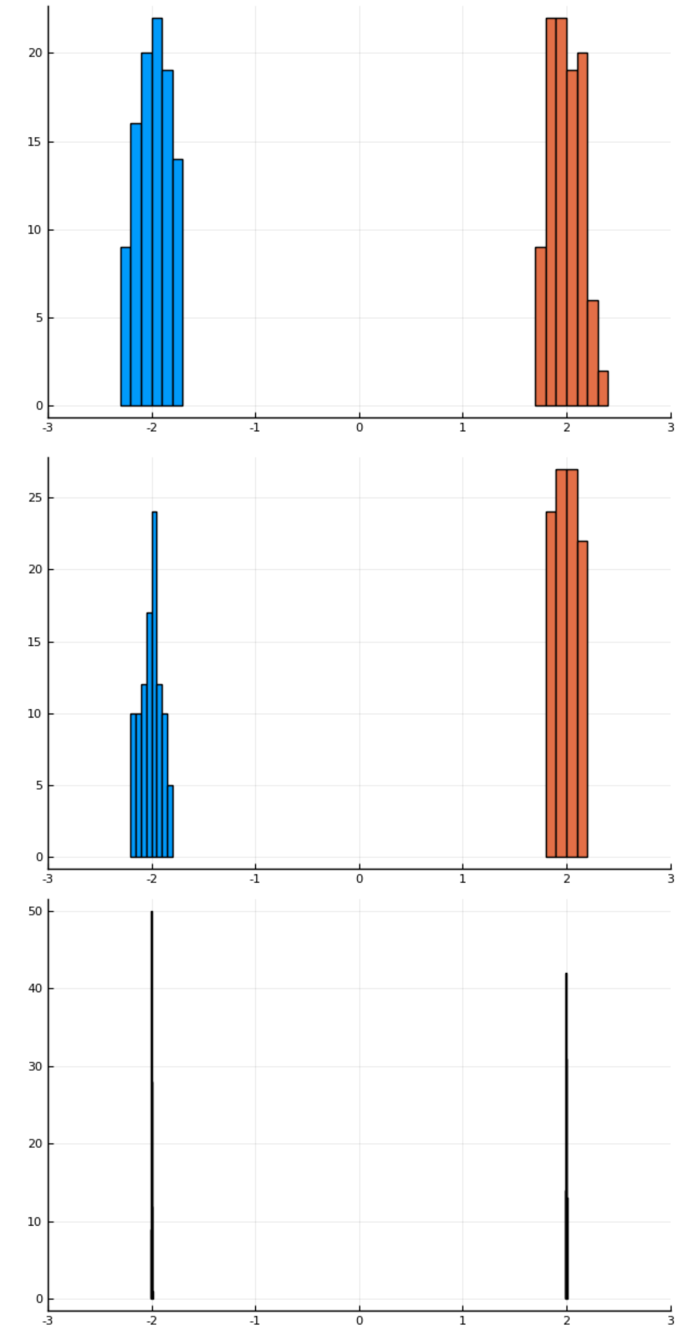
Linear decoding
(small effects)

~ 6 bits

'Spiral' decoding

6- stage 'Spiral' decoding

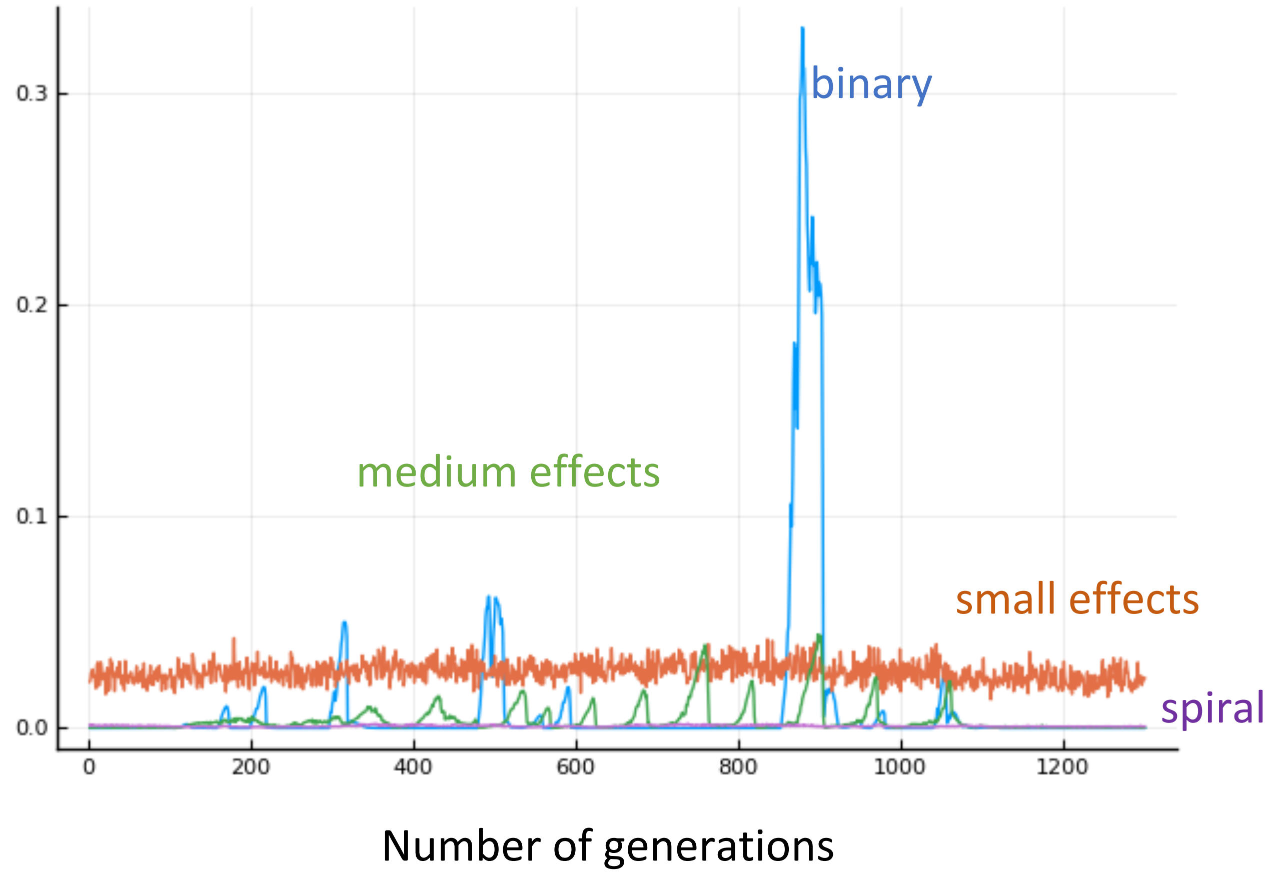
~ 14 bits



Why not just use binary? A problem with large effects.

In binary, 0111 and 1000 are neighbouring values, but cannot evolve directly from one to the other.

RMS tracking error for stabilizing selection on a slowly varying target value



Is there a unifying simplicity?

A field of numerous 'bio-inspired' genetic algorithms, highly abstracted from genetics.

Each genetic algorithm specifies a Markov chain of populations: these chains are usually intractable to analyse.

In some sense, there has been 'evolution of evolvability' – but in what sense?

Is there a framework that puts genetic algorithms within standard Markov-chain Monte Carlo?

Can we construct a genetic algorithm with a Markov chain of populations that satisfies detailed balance, and for which the stationary distribution can be written in closed form?

Can we relate genetic algorithms to existing ideas in machine learning?

Can we put evolutionary and individual learning within the same model?

EBT: Exchangeable Breeding with Tournaments

A tool for studying mutation-selection equilibrium

'Genetic algorithm' that satisfies detailed balance.

Generation t | Exchangeably sample one genome | Reject one genome | Generation t+1

$$g_{N+1} \sim p_B(\cdot \mid g_1, \ldots, g_N)$$

Reject genome k with probability $$\frac{\frac{1}{f_k}}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}}$$

Exchangeable Sampling

“Breeding” is by sampling a genome, conditional upon existing genomes. Joint “breeding distribution” should be exchangeable:

$$p_B(g_1, \ldots, g_N) = p_B(g_{\sigma_1}, \ldots, g_{\sigma_N})$$

For a ‘Genetic Algorithm’, use a breeding distribution that is a product of Beta-Bernoulli distributions (or Dirichlet Processes).

That is, each element of a new genome is sampled from an independent Dirichlet process: the ‘genome’ consists of a vector containing a sample from each DP.

Design decisions for an evolutionary model to satisfy detailed balance

1. Propose (i.e. 'breed') one genome at a time
2. Sample each part of the new genome from the whole population, not just from 2 parents
3. Mutations cannot depend on previous value: a 'mutation' is an 'immigrant' value sampled from an underlying gene pool.
4. During breeding, all members of population contribute equally
5. After proposing a new genome, a genome is selected for rejection: less fit genomes more likely to be rejected. Differences in fitness modelled as differences in lifespan, not fecundity.

Tournament Selection by ‘loser ticket’

If genomes i and j have a ‘tournament’

$$Pr(\text{j wins tournament against i}) = \frac{f_j}{f_i + f_j}$$

Suppose there is one ‘loser ticket’. After many tournaments in which the genome currently holding the loser ticket ‘fights’ another genome, and the loser gets the ticket, limit distribution is:

$$Pr(\text{j holds loser ticket}) = \frac{\frac{1}{f_j}}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}}$$

Key point: Stationary Distribution of EBT can be written in closed form

EBT is a Markov Chain of populations.

Stationary distribution is:

$$\pi(g_1, \ldots, g_N) \propto p_B(g_1, \ldots, g_N) f(g_1) \cdots f(g_N)$$

Analogous to a Bayesian posterior, with 'breeding distribution' as prior, and fitness as likelihood.

Proof of detailed balance of EBT:

Let $G = \{g_1, g_2, \ldots, g_{N+1}\}$

To prove: $\pi(G_{\setminus N+1})\, T(G_{\setminus N+1} \to G_{\setminus i}) = \pi(G_{\setminus i})\, T(G_{\setminus i} \to G_{\setminus N+1})$

Proof:

$$\pi(G_{\setminus N+1})\, T(G_{\setminus N+1} \to G_{\setminus i}) =$$

$$p_B(G_{\setminus N+1})\ f_1 \cdots f_N\ \ p_B(g_{N+1} \mid G_{\setminus N+1})\ \frac{\frac{1}{f_i}}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}}$$

$$= p_B(G)\ \ \frac{f_1 \cdots f_{N+1}}{f_i f_{N+1}}\ \ \frac{1}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}}$$

which is symmetric between g_i and g_{N+1}.

Genetic algorithm with population size 1

Breed **two** new genomes,
discard existing genome

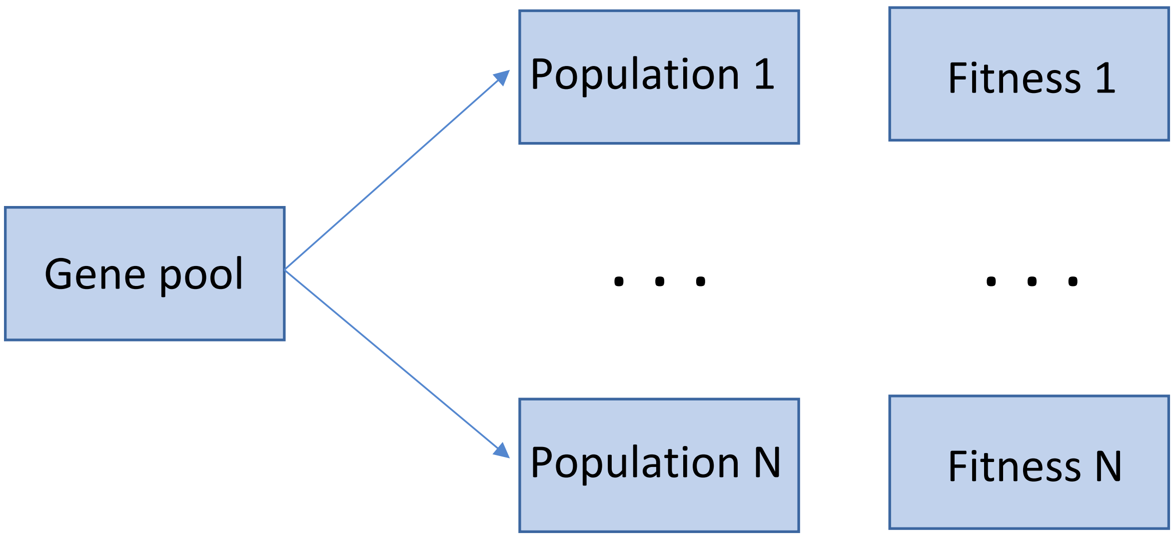
Select between two new genomes.
Bad algorithm!

EBT with population size 1

Sample **one** new genome,
keep existing genome

Metropolis-Hastings (original version)

Evolution of evolvability: a static hierarchical Bayesian view



Populations with migration from the same gene pool must be similar, and

Hyper-prior

Latent variables

Likelihood models

migration

Conclusions and Further Questions

A suggestive difference between the asexual and sexual genetic algorithms is that recombination gives orders of magnitude more ‘adaptive channel capacity’.

A small demonstration that ‘adaptive channel capacity’ can be accessed with suitable decoder.

Exchangeable Breeding with Tournaments (EBT) is an alternative formulation of GAs, that satisfies detailed balance.

The mechanisms of non-parametric Bayesian MCMC using priors based on Dirichlet distributions can be plausibly interpreted as ‘genetic algorithms’.

Links GAs and MCMC, and enables integrated probability models of evolution and individual learning.

Learning, intelligence, agency…

For all except a very few highly sophisticated animals,

learning and cognitive development consist of unpacking genetic information into behaviour,

using as little experience as possible.

With the MCMC approach, we can put evolutionary and individual learning into one probability model.

Evolutionary roles of ‘entry-level’ learning

Situation-adaptive	Organisms born into **different environments**, and must learn from environment to succeed
Decompression	Learning decodes a compact description of behaviour, e.g. as a reward system.
Error-correction	Learning resolves inconsistencies in genetic specification of behaviour, enabling genetic encoding of behaviour of greater complexity

Decompression and error correction valuable even if all organisms in **same environment**