OPTIMIZATION FOR AI

GLOBAL AND MULTI-OBJECTIVE OPTIMIZATION

Luca Manzoni

EVOLUTION STRATEGIES

EVOLUTION STRATEGIES: IDEAS

- Invented in the ’60
- Some similarities with GA:
  - There is a population of solutions
  - There are offsprings derived from mutation
  - There is a selection process

EVOLUTION STRATEGIES: IDEAS

- However, they have some key differences:
  - There is (usually) no crossover
  - The most used selection is truncated selection
  - Usually the individuals represent floating points values (which is also possible with GA)

ES PARAMETERS

λ

Number of children
that are generated

Number of
selected individuals

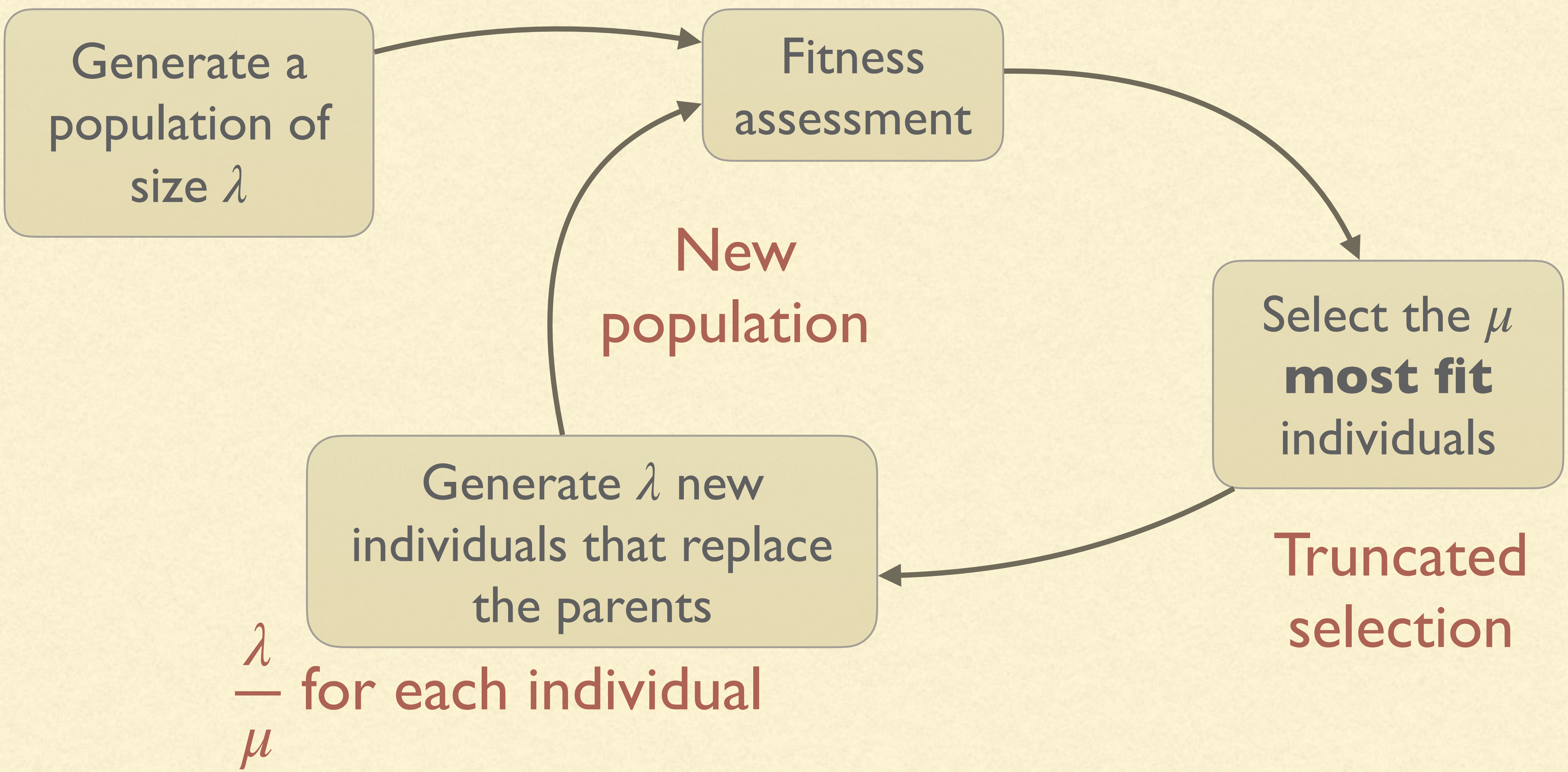
μ

Two different kinds of ES:

$$(\mu, \lambda) - ES \qquad (\mu + \lambda) - ES$$

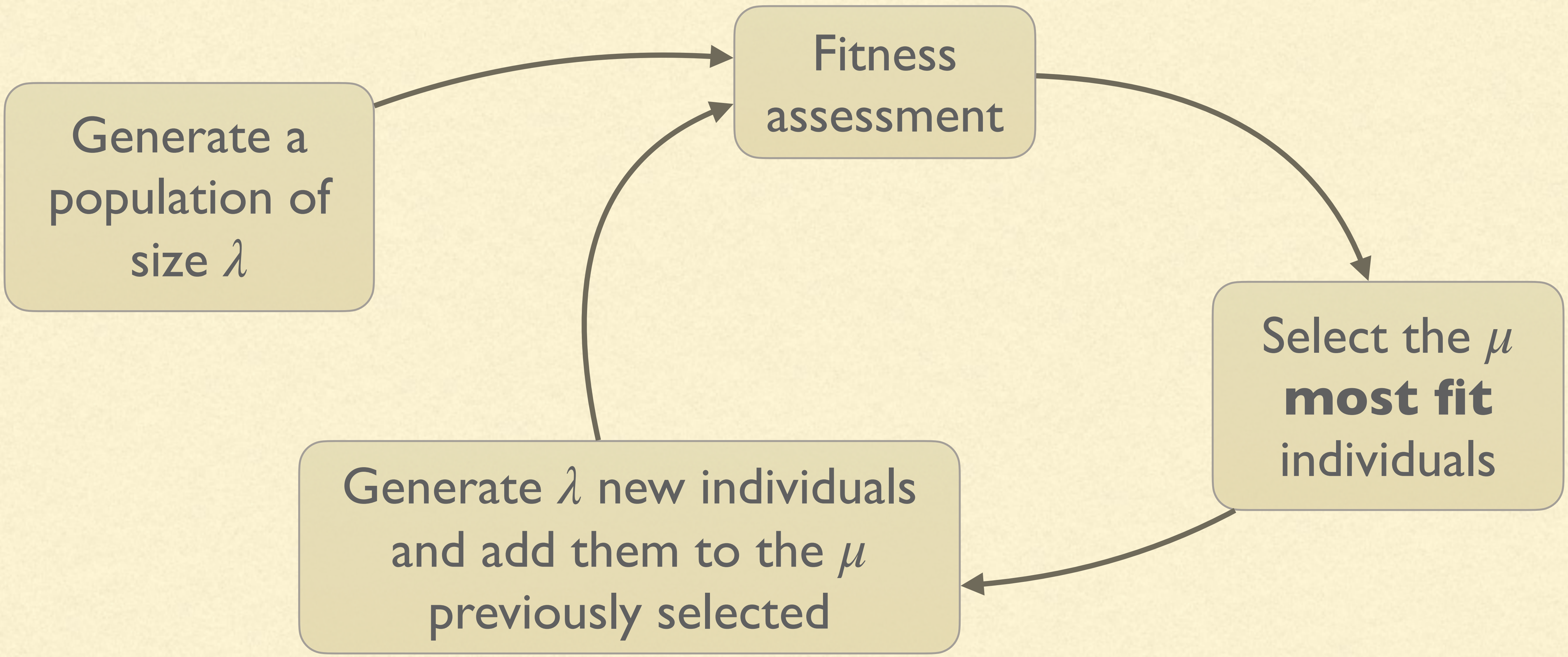
THE **ES** CYCLE

$(\mu, \lambda) - ES$



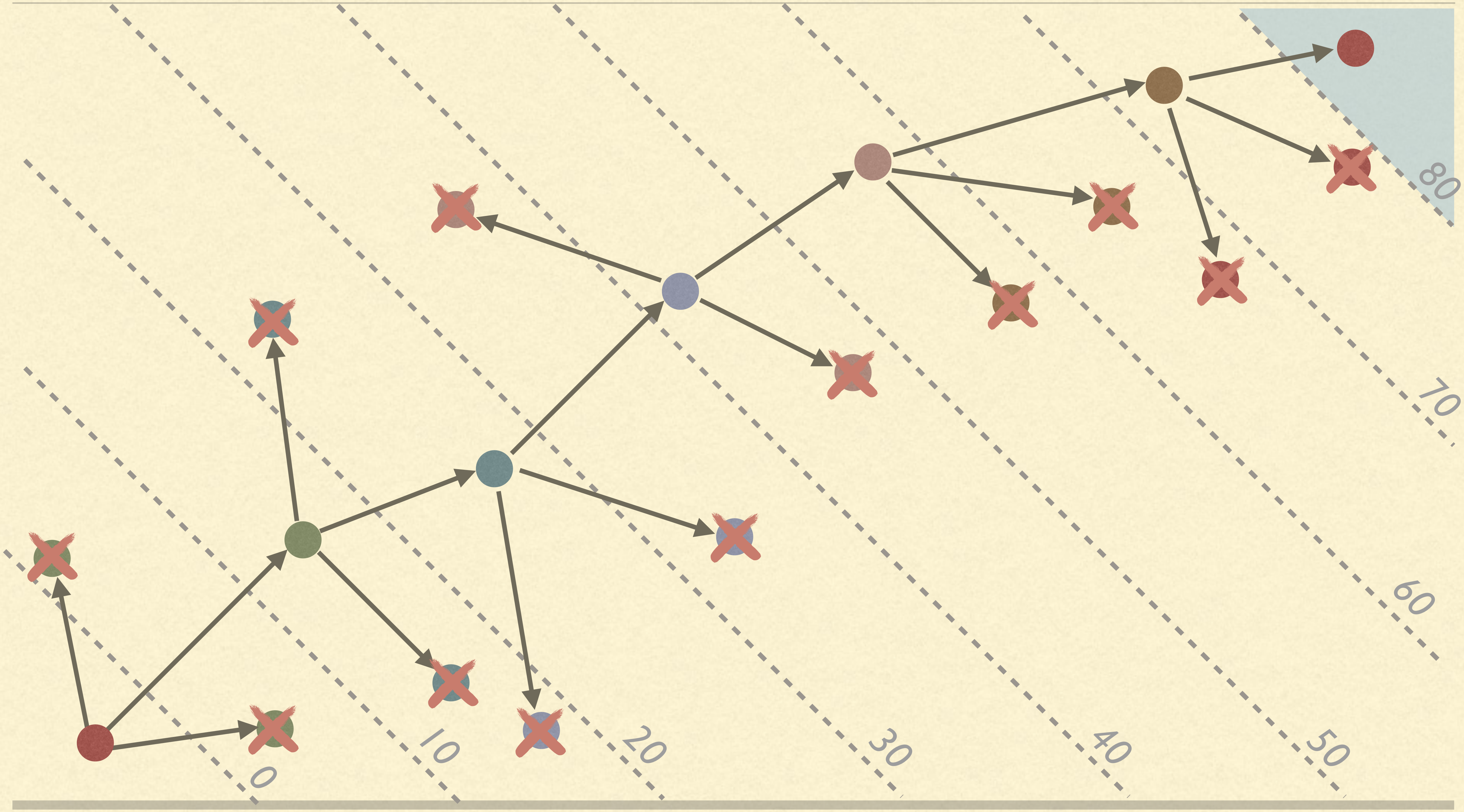
THE **ES** CYCLE

$(\mu + \lambda) - ES$



The parents are added together
with the children to the new population

EXAMPLE OF (1,3) ES



PROPERTIES OF A GOOD MUTATION OPERATOR

- **Reachability**: each area of the search space can be reached in a finite number of steps
- **Unbiasedness**: mutation should not use any information deriving from the fitness (that's the role of selection)
- **Scalability**: the "strength" of the mutation should be adaptable to the specific fitness landscape/search space (e.g., by deciding how much the mutation changes an individual)

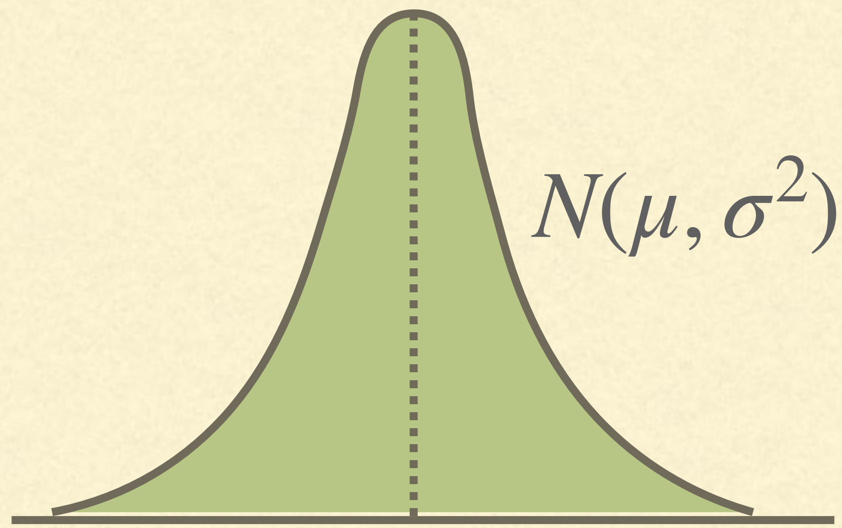
MUTATION

For binary values: the same as GA

In the case of real values the mutation is usually performed by adding a gaussian noise to the coordinates

Gaussian

$$x_i +$$



$\mu = 0$ seems natural, but how to select the variance?

SELF-ADAPTIVITY IN ES

- It is common do have self-adaptive ES, where a series of parameters (e.g., the variance) are modified during the evolution
- You can think of every individual of being a pair $\langle x, s \rangle$ where x is the actual solution and s a set of parameters of the operators used for mutation
- In some cases s itself is modified as part of the evolutionary process with the same operators it controls

ONE-FIFTH RULE

- An empirical rule for self-adaptation of the variance of the mutation operator
- Introduced by Info Rechenberg in the 70s
- If less than 1/5 of the children are fitter than their parents then decrease the variance
- If more than 1/5 of the children are fitter than their parents then increase the variance

ONE-FIFTH RULE

- Two parameters, $k \in \mathbb{N}$ and $c \in (0,1]$ (usually $0.817 < c < 1$)
- p_S is the probability of having a successful mutation
- Every k generations:
  - If $p_S > 1/5$ then set $\sigma = \sigma / c$
  - If $p_S < 1/5$ then set $\sigma = \sigma \cdot c$
  - Otherwise leave σ unchanged

ES WITH RECOMBINATION

- It is possible to extend ES with a recombination step (in addition to mutation) using ρ parents
- The notations are $(\mu/\rho, \lambda)$-ES and $(\mu/\rho + \lambda)$+ES
- It means that to generate each of the λ children, ρ individuals are randomly selected (without reinsertion) from the population of size μ

ES WITH RECOMBINATION

- Two main kinds of recombination:
  - **Discrete/dominant recombination**: for each position select randomly from one of the ρ individuals
  - **Intermediate recombination**: given the values $x_{1,j}, x_{2,j}, \ldots, x_{\rho,j}$ for each position j of the parents the offspring will contain the average $\frac{1}{\rho}\sum_{i=1}^{\rho} x_{i,j}$ of all that values