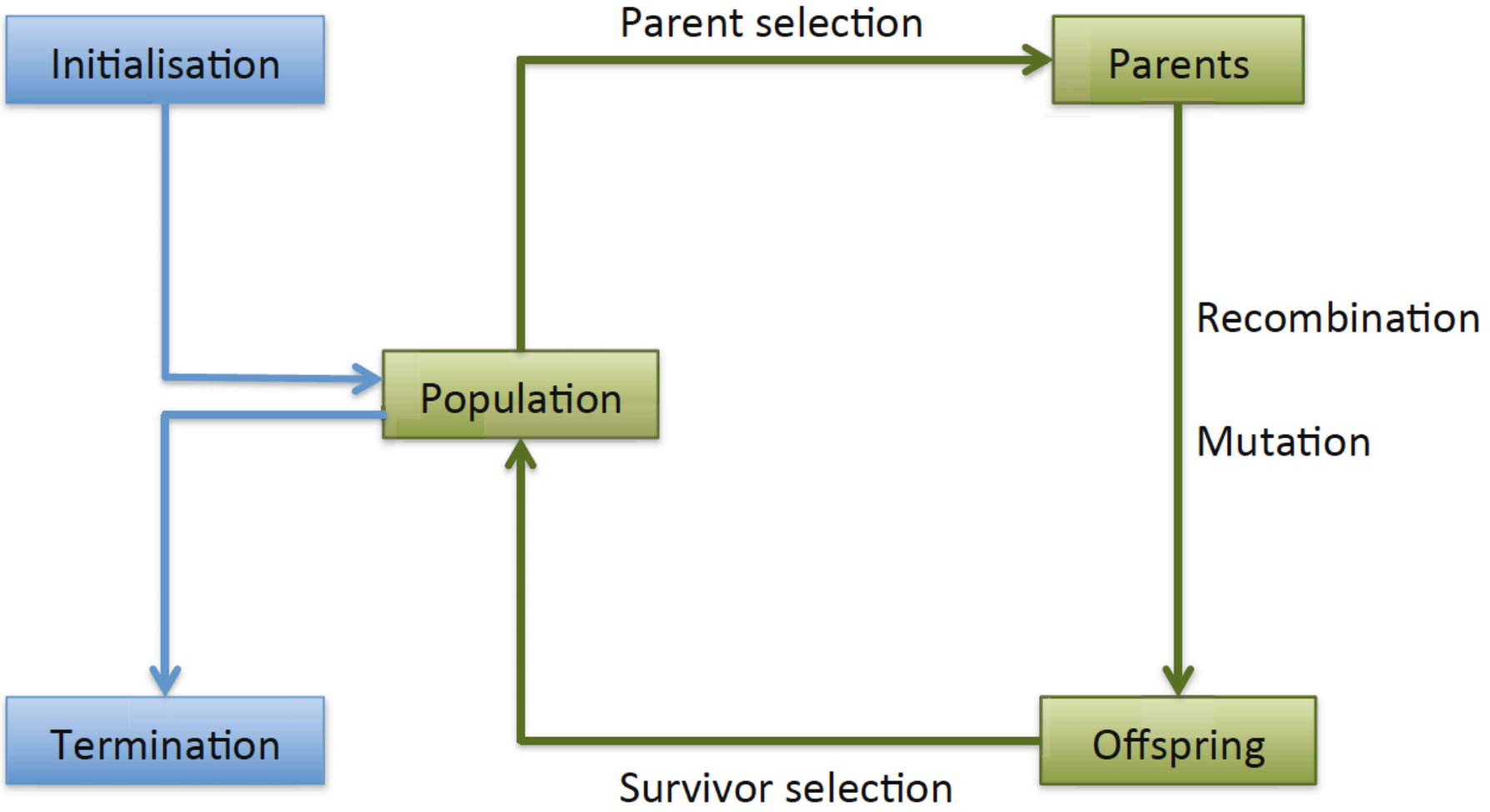
3: Evolutionary Algorithm Basics

- EA flowchart
- Representation
- Evaluation
- Population and selections
- Variation operators
- Textbook Chapter 3.2

General scheme of EAs



Main components of the design of an EA

- Representation (definition of individuals)
- Evaluation function (or fitness function)
- Population
- Parent selection mechanism
- Variation operators, recombination and mutation
- Survivor selection mechanism (replacement)

Pseudo-code for typical EA

```
BEGIN
  INITIALISE population with random candidate solutions;
  EVALUATE each candidate;
  REPEAT UNTIL ( TERMINATION CONDITION is satisfied ) DO
    1 SELECT parents;
    2 RECOMBINE pairs of parents;
    3 MUTATE the resulting offspring;
    4 EVALUATE new candidates;
    5 SELECT individuals for the next generation;
  OD
END
```

Representations

- Candidate solutions (individuals) exist in *phenotype space*
- They are encoded in chromosomes, which exist in *genotype space*
    - Encoding: phenotype -> genotype
    - Decoding: genotype -> phenotype (one to one)
- Chromosomes contain genes, which are in (usually fixed) positions called loci and have a value (allele)

In order to find the global optimum, every feasible solution must be represented in genotype space

Evaluation (fitness) function

- Represents the requirements that the population should adapt to
  - a.k.a. *quality* function or *objective* function
- Assigns a single real-valued fitness to each phenotype which forms the basis for selection
- Typically we talk about fitness being maximized
  - Some problems may be best posed as minimization problems

Implicit fitness

- For many of the approaches, an explicit fitness function is used. Why?
  - better determine the kinds of good solutions
  - useful to solve particular problems
- Can have fitness be determined implicitly
  - competition in some form for resources, where better competitors are more fit
  - Tierre, Avida, mostly Artificial Life

Population

- Holds (representations of) possible solutions
- Usually has a fixed size and is a *multiset* of genotypes
- Some sophisticated EAs also assert a spatial structure on the population, i.e. a grid
- Selection operators usually take whole population into account, i.e reproductive probabilities are relative to *current* generation
- Diversity of a population refers to the number of different fitnesses / phenotypes / genotypes present (note not the same thing)

Parent selection mechanism

- Assigns variable probabilities of individuals acting as parents depending on their fitnesses
- Usually **probabilistic**
  - High quality solutions more likely to become parents than low quality ones
  - but not guaranteed
  - even worst in current population usually has non-zero probability of becoming a parent
- This *stochastic* nature can aid escape from local optima but slow you down

Variation operators

- Role is to generate new candidate solutions
    - From parents to child (offspring)
- Usually divided into two types according to their arity (number of inputs):
    - Arity 1: mutation operator
    - Arity > 1: recombination operator
    - Arity = 2: typically called crossover
- There has been much debate about relative importance of recombination and mutation
    - Nowadays most EAs use both
    - Choice of particular variation operators is representation dependent

Mutation

- Act on one genotype and delivers another
- Element of randomness is essential and differentiates it from other unary heuristic operators
- Importance ascribed depends on representation and dialect:
    - Binary GAs - background operator responsible for preserving and introducing diversity
    - Continuous variables EP - sole search operator
- May guarantee connectedness of search space and hence coverage proofs

Recombination

- Merges information from parents into offspring
- Choice of what information to merge is *stochastic*
- Most offspring may be worse, or the same as the parents
    - Hope is that some are better by combining elements of genotypes that lead to good traits
- In nature, higher-level organisms reproduce sexually while lower-level ones reproduce asexually - recombination is the superior form of reproduction

Survivor selection (replacement)

- Sometimes also called environmental selection
- Most EAs use fixed population size, and need a way of going from (parents + offspring) to next generation
- Often deterministic (different from parent selection!)
    - Fitness based: e.g., ranking parents + offspring and save the best
    - Age based: e.g., use offspring to replace worst individuals in the previous generation
- Elitism

Initialization and termination

- Initialization usually done at **random**
  - Need to ensure even spread and mixture of possible allele values
  - Can include existing solutions, or use problem-specific heuristics to "seed" the population
- Termination condition checked every generation
  - Reaching some (known/hoped for) fitness
  - Reaching some allowed limit number of generations
  - Reaching some minimum level of diversity
  - Reaching some specified number of generations without fitness improvement (stagnation)
  - Anytime algorithm

General scheme of EAs