

Genetic Algorithms

Part 1: Introduction

Spring 2009

Instructor: Dr. Masoud Yaghini

Outline

- Evolutionary Computing (EC)
- Biological Background
- Motivations for EC
- What are Genetic Algorithms?
- Simple Genetic Algorithm (SGA)
- References

Evolutionary Computing (EC)



Evolutionary Computing (EC)

- EC is part of computer science
- EC is not part of life sciences/biology
- It draws inspiration from the process of natural evolution
- EC can be applied in biological research

The Main EC Metaphor

EC	Nature
Optimization problem	Environment
Feasible solutions	Individuals living in that environment
Solutions quality	Fitness (Individual's degree of adaptation to its surrounding environment)

Genetic Algorithms: Part 1

Evolutionary Computing Areas

Genetic
Programming

Evolution
Strategies

Genetic
Algorithms

Evolutionary
Programming

Brief History

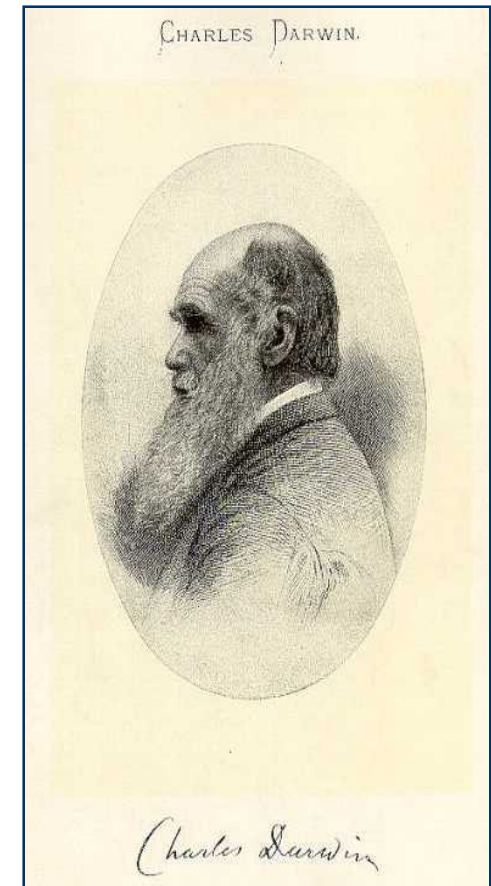
- 1964, Rechenberg introduces **evolution strategies**
- 1965, L. Fogel, Owens and Walsh introduce **evolutionary programming**
- 1975, Holland introduces **genetic algorithms**
- 1992, Koza introduces **genetic programming**

The image features a large green shape on the left side, which has a white, rounded rectangular cutout. The text "Biological Background" is centered within this white area. A dark blue horizontal bar with rounded ends extends from the right side of the green shape across the bottom of the page.

Biological Background

Darwin's principles

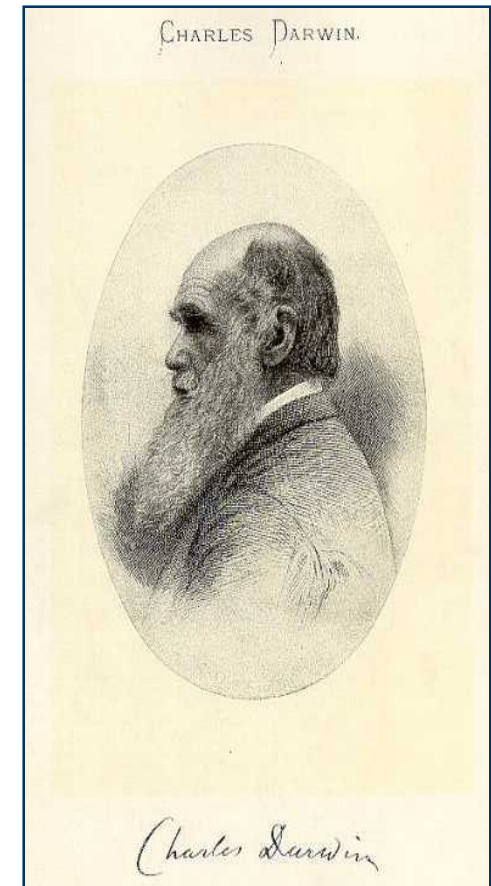
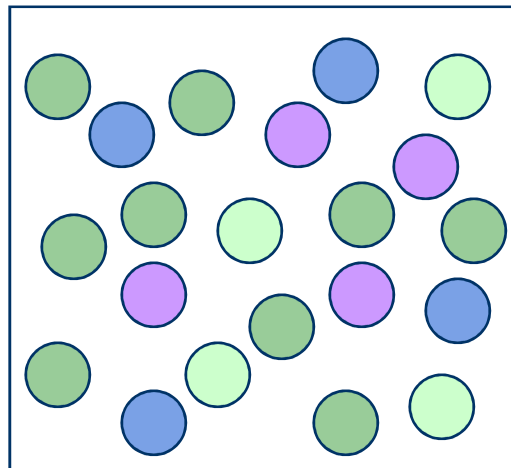
- **Variety** of species individuals within the population
- **Overproduction** of offspring generation
 - Individuals have basic instinct towards reproduction
- **Competition** for limited resources
 - Environment only support a limited number of individuals
- **Survival of the fittest**
 - Those individuals, that are adopted or fit to the environmental condition best, have increased chance of reproduction



Evolution

How does it work?

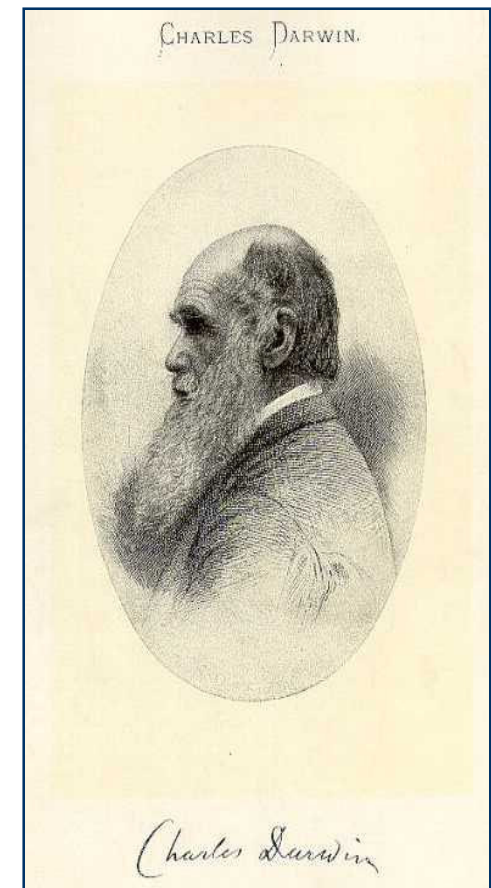
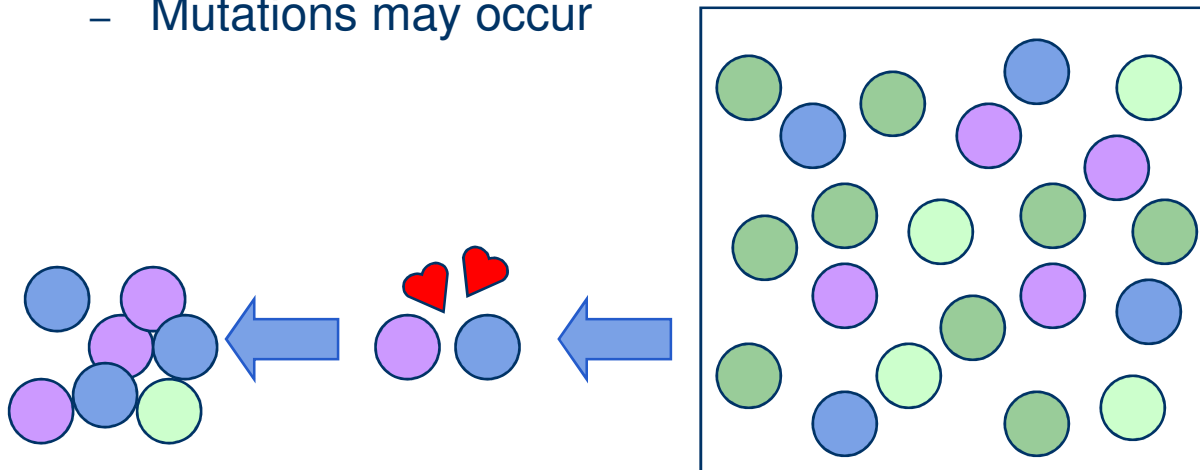
- Initial population
 - Variety of shapes, colors, behaviors
 - Each individual fits differently to the environment



Evolution

How does it work?

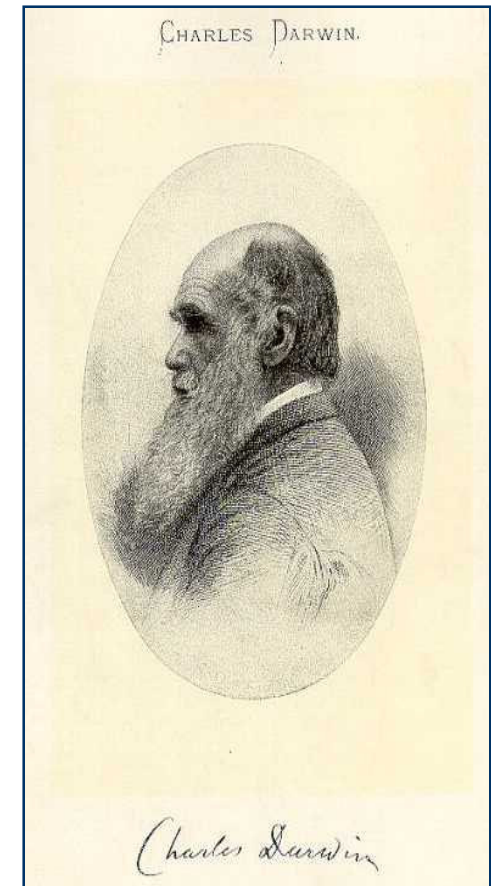
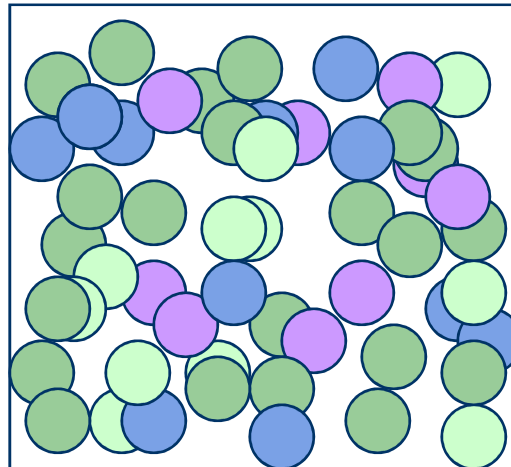
- Initial population
- Reproduction
 - Offspring combines both parents properties
 - Siblings may differ in properties
 - Mutations may occur



Evolution

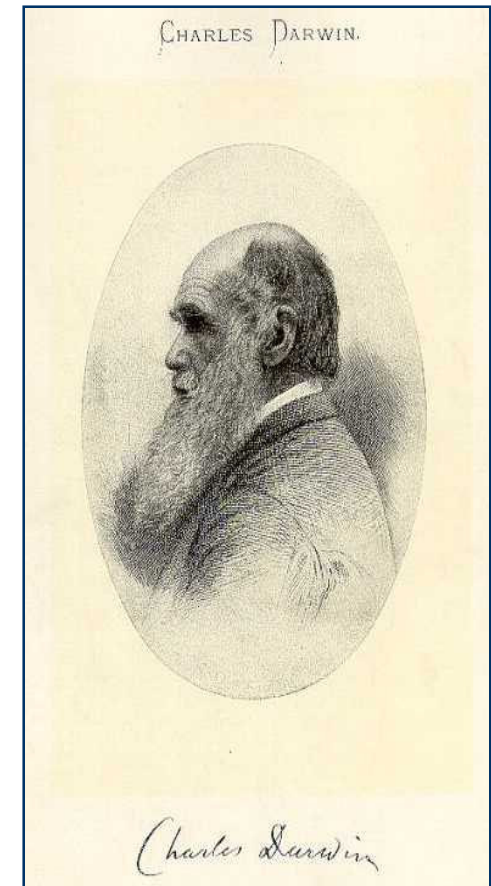
How does it work?

- Initial population
- Reproduction
- Limited environmental resources
 - Only a portion of the individuals survive
 - Survival chances – according to fitness measure



Evolution

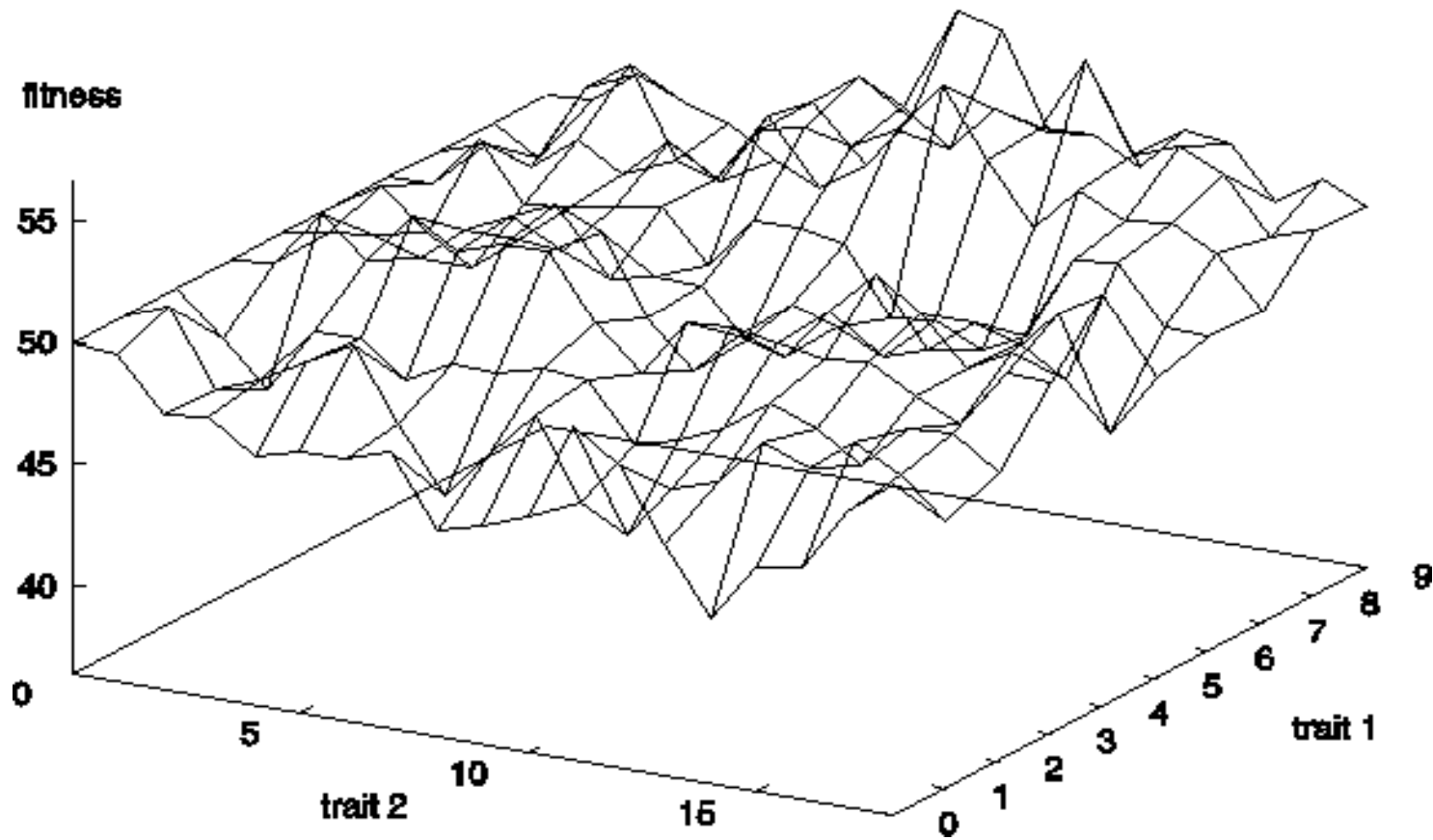
- **Phenotypic traits:**
 - Behaviour / physical differences that affect response to environment
 - Partly determined by inheritance, partly by factors during development
 - Unique to each individual, partly as a result of random changes
- If phenotypic traits lead to higher chances of reproduction
 - Can be inherited
 - Then they will tend to increase in subsequent generations
 - Leading to new combinations of traits ...



Landscape metaphor

- The height dimension belongs to fitness
- The other two (or more) dimensions correspond to biological traits
- The x-y-plane holds all possible trait combinations
- Therefore, each different individual (**phenotype**) represents a single point on the landscape
- Population is therefore a “cloud” of points, moving on the landscape over time as it evolves - adaptation

Example with two traits

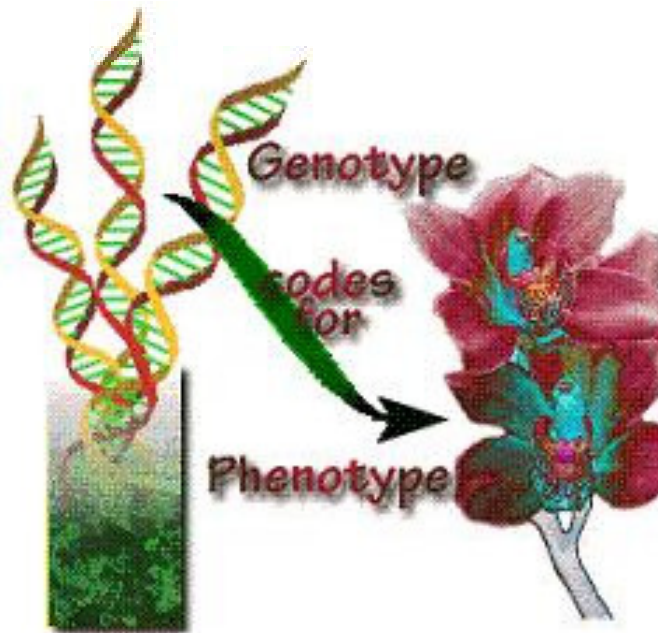


Landscape metaphor

- Selection “pushes” population up the landscape
- There are a number of points that are better than all their neighbouring solutions, we call each of these points a **local optimum**
- The highest of these points is called **global optimum**
- Random variations in feature distribution (+ or -) arising from sampling error can cause the population down hills, thus crossing valleys and leaving local optima

Natural Genetics

- The information required to build a living organism is coded in the **DNA** of that organism
- **Genotype** (DNA inside) determines **phenotype**



Natural Genetics

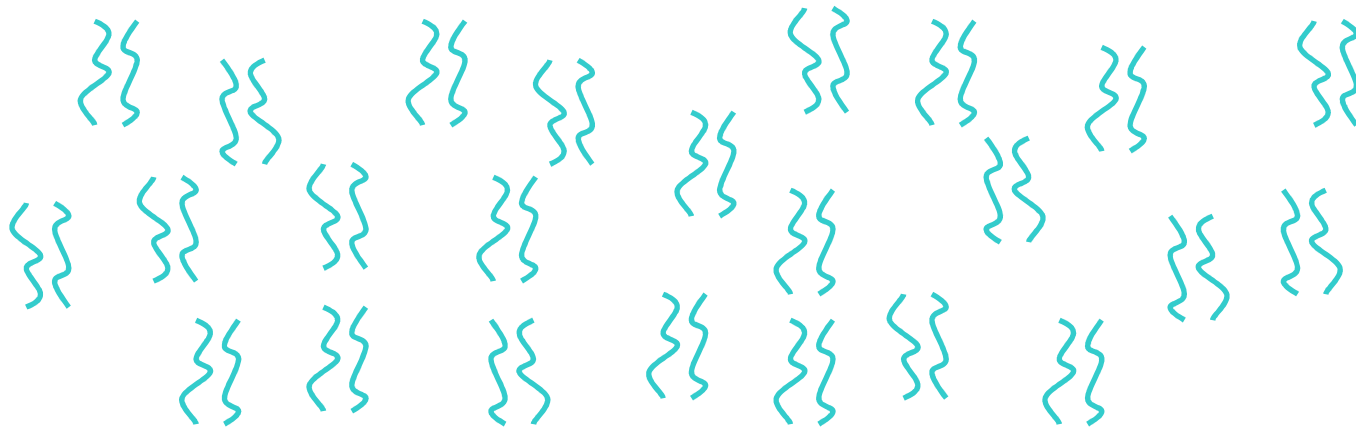
- **Genes** are encoding phenotypic characteristics
 - One gene may affect many traits
 - Many genes may affect one trait
- Small changes in the genotype lead to small changes in the organism (e.g., height, hair colour)
- The possibilities of the **genes** for one property is called **Allele**
- Genotypic variations are consequences of:
 - **Recombination** of genes by sexual reproduction
 - **Mutation** of genes

Genes and the Genome

- The complete genetic information in an individual's genotype is called the **Genome**
- Genes are encoded in strings of DNA called **Chromosomes**
- In most **cells**, there are two copies of each chromosome, called **Diploid**
- Within a species, most of the genetic material is the same

Example

- **Human DNA** is organised into chromosomes
- **Human body cells** contains **23 pairs** of chromosomes which together define the physical attributes of the individual:

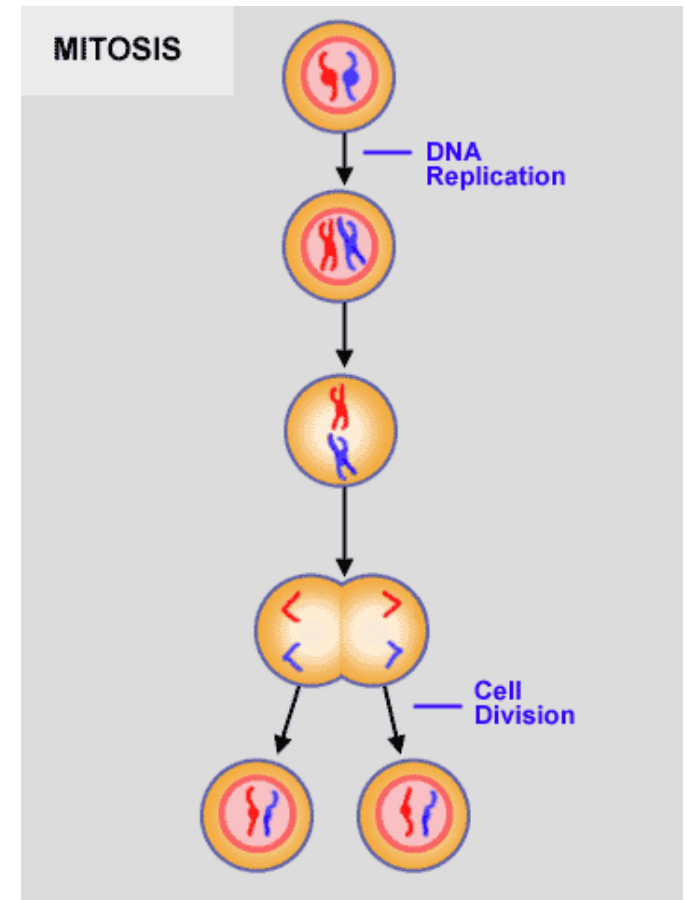


Reproductive Cells

- **Gametes** (i.e, **sperm** and **egg cells**) contain 23 individual chromosomes **rather than** 23 pairs
- Cells (gametes) with only one copy of each chromosome are called **Haploid**
- The haploid sperm cell **merges** with the haploid egg cell and forms a diploid cell, called **Zygote**
- The new organism develops from this zygote by the process named **Ontogenesis**
- All body cells contain the **same** genetic information as the zygote it original form

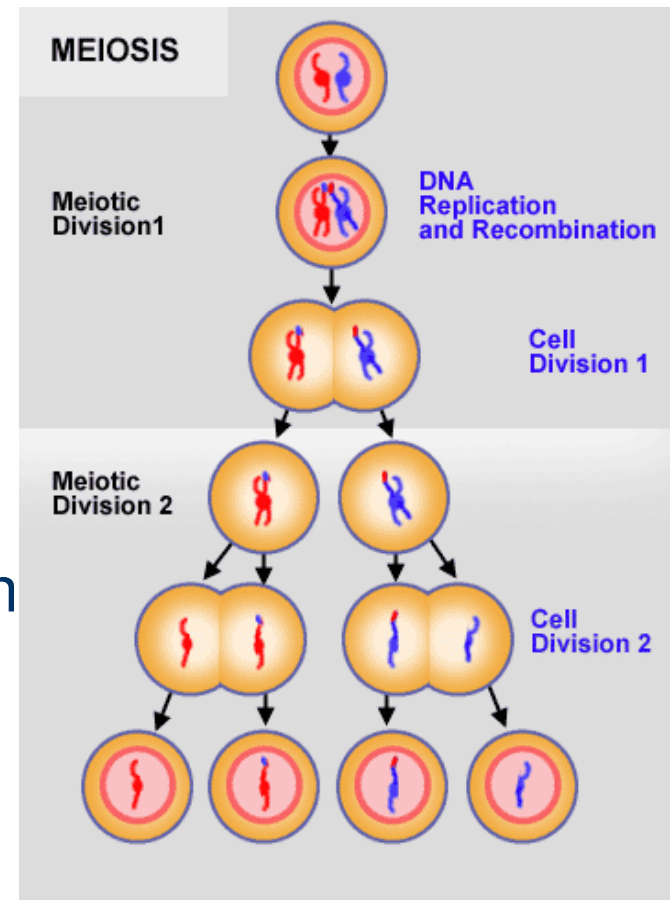
Mitosis

- **Mitosis** is copying the same genetic information to new offspring
- Mitosis is the normal way of growing of multicell structures



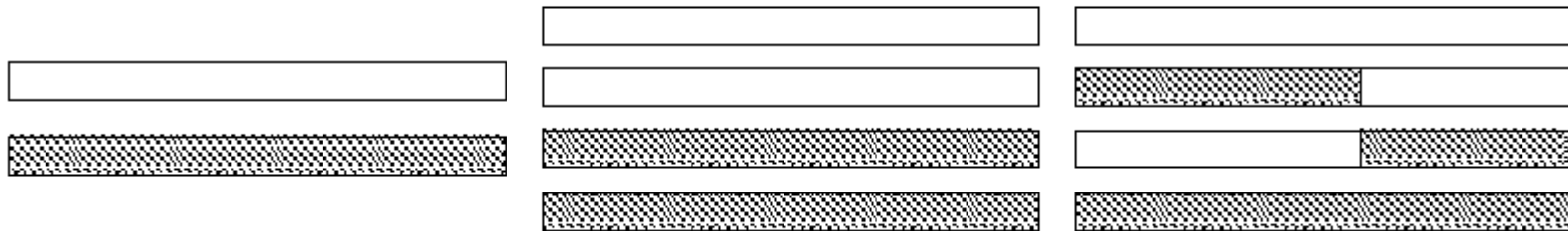
Meiosis

- **Meiosis** is the basis of sexual reproduction
- After meiotic division, gametes appear in the process
- Hence genetic information is shared between the parents in order to create new offspring
- During meiosis the pairs of chromosome undergo an operation called **Crossing over**



Crossing-over during meiosis

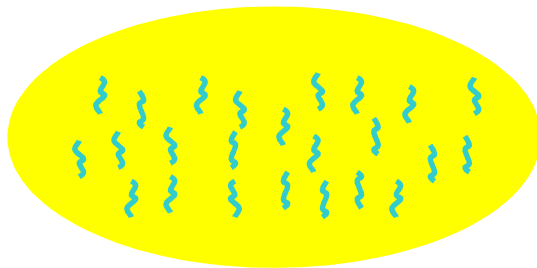
- Chromosome pairs align and duplicate
- Inner pairs link exchange parts of themselves



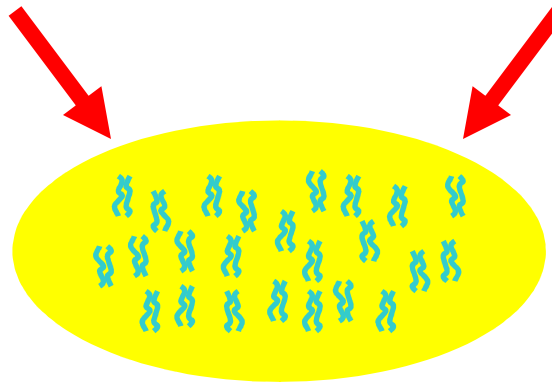
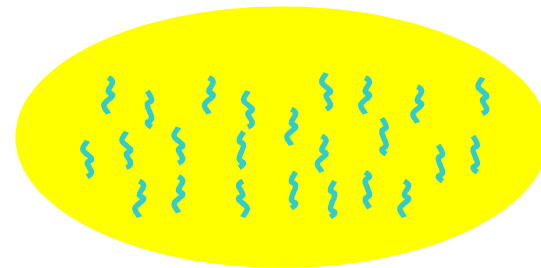
- Outcome is one copy of **maternal/paternal** chromosome plus two entirely new combinations
- After crossing over one of each pair goes into each gamete

Fertilisation

Sperm cell from Father



Egg cell from Mother



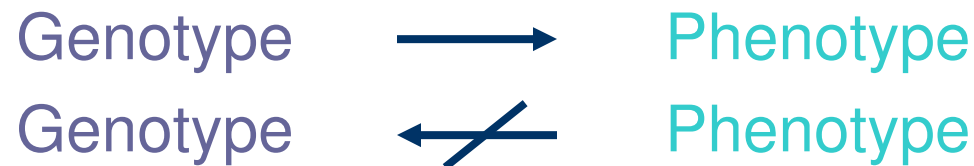
New person cell (zygote)

After fertilisation

- **New zygote** rapidly divides etc creating many cells all with the same genetic contents
- Although all cells contain the same genes, depending on, for example where they are in the organism, they will behave differently
- This process of differential behaviour during development is called **ontogenesis**
- All of this uses, and is controlled by, the same mechanism for decoding the genes in DNA

Transcription, translation

A central claim in molecular genetics: only one way flow



Lamarckism (saying that acquired features can be inherited) is thus wrong!

Mutation

- Occasionally some of the genetic material changes very slightly during this process
- This means that the child might have genetic material information **not inherited** from either parent
- This can be
 - **Disastrous**: offspring is not viable (most likely)
 - **Neutral**: new feature does not influence fitness
 - **Advantageous**: strong new feature occurs



Motivations for EC



Motivations for GA

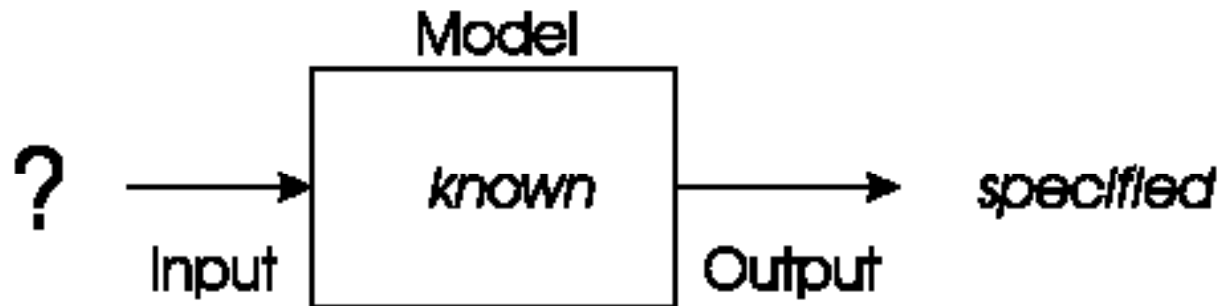
- Nature has always served as a **source** of inspiration for engineers and scientists
- The best problem solver known in nature is:
 - **the (human) brain** that created “the wheel, New York, wars and so on”
 - **the evolution mechanism** that created the human brain

Motivations for EC

- Developing, analyzing, applying problem **solving methods (algorithms)** is a central theme in mathematics and computer science
- **Complexity** of problems to be solved **increases**
- Consequence: **Robust problem solving** technology needed
 - Which do not need much tailoring for specific problems, and
 - Deliver good (not necessarily optimal) solutions within acceptable time
- EC do all this

Optimizations

- We have a model of our system and seek inputs that give us a specified goal



- e.g.
 - Crew scheduling problem, Train timetabling, & ...

What is a Genetic Algorithm?



Genetic Algorithm

- A population of individuals exists in an environment with limited resources
- **Competition** for those resources causes selection of those **fitter individuals** that are better adapted to the environment
- These individuals act as seeds for the generation of new individuals through recombination and mutation
- The new individuals have their fitness evaluated and compete for survival.
- Over time **natural selection** causes a rise in the fitness of the population

What are Genetic Algorithms?

- They are **Bio-Inspired artificial intelligence class, stochastic, population-based** algorithms
- Typically applied to:
 - hard problems with a large search space
 - discrete optimization
- Developed by **John Holland**, USA in the 1970's



Pseudo-code for typical GA

```
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
```

Simple Genetic Algorithm (SGA)



Simple Genetic Algorithm (SGA)

- Holland's original GA is now known as the simple genetic algorithm (SGA)
- Other GAs use different:
 - Representations
 - Mutations
 - Crossovers
 - Selection mechanisms

Genetic Algorithms: Part 1

SGA summary

Representation	Binary strings
Recombination	1-point crossover
Mutation	bit-flipping with fixed probability
Parent selection	Fitness-Proportionate
Survivor selection	All children replace parents
Speciality	Emphasis on crossover

Genetic Algorithms: Part 1

Simple example – $f(x) = x^2$

- Finding the maximum of a function:
 - $f(x) = x^2$
 - Range $[0, 31]$ → Goal: find max ($31^2 = 961$)
 - Binary representation: string length 5 = 32 numbers (0-31)

genotype	0 0 1 0 1	
	<hr/>	
	2^4 2^3 2^2 2^1 2^0	
mapping	16 8 4 2 1	
	<hr/>	
phenotype	$0*16+0*8+1*4+0*2+1*1 = 5$	
	<hr/>	
fitness	25	= $f(x)$

x^2 example

- **x^2 example:**
 - Representation: Binary code
 - Population size: 4
 - Recombination: 1-point crossover
 - Mutation: Bit-flipping with fixed probability
 - Parent selection: Fitness-Proportionate
 - Initialization: Random
- We show one generational cycle done by hand

Genetic Algorithms: Part 1

x^2 example: selection

String no.	Initial population	x Value	Fitness $f(x) = x^2$	$Prob_i$	Expected count	Actual count
1	0 1 1 0 1	13	169	0.14	0.58	1
2	1 1 0 0 0	24	576	0.49	1.97	2
3	0 1 0 0 0	8	64	0.06	0.22	0
4	1 0 0 1 1	19	361	0.31	1.23	1
Sum			1170	1.00	4.00	4
Average			293	0.25	1.00	1
Max			576	0.49	1.97	2

Genetic Algorithms: Part 1

X² example: crossover

String no.	Mating pool	Crossover point	Offspring after xover	x Value	Fitness $f(x) = x^2$
1	0 1 1 0 1	4	0 1 1 0 0	12	144
2	1 1 0 0 0	4	1 1 0 0 1	25	625
2	1 1 0 0 0	2	1 1 0 1 1	27	729
4	1 0 0 1 1	2	1 0 0 0 0	16	256
Sum					1754
Average					439
Max					729

X² example: mutation

String no.	Offspring after xover	Offspring after mutation	x Value	Fitness $f(x) = x^2$
1	0 1 1 0 0	1 1 1 0 0	26	676
2	1 1 0 0 1	1 1 0 0 1	25	625
2	1 1 0 1 1	1 1 0 1 1	27	729
4	1 0 0 0 0	1 0 1 0 0	18	324
Sum				2354
Average				588.5
Max				729

The simple GA

- SGA Shows many **shortcomings**:
 - Representation is too restrictive
 - Mutation & crossovers only applicable for bit-string & integer representations
 - Selection mechanism sensitive for converging populations with close fitness values
 - Generational population model (step 5 in SGA) can be improved with explicit survivor selection



References



References

- Eiben and Smith. **Introduction to Evolutionary Computing**, Springer-Verlag, New York, 2003.
- J. Drevo A. Petrowski, P. Siarry E. Taillard, **Metaheuristics for Hard Optimization**, Springer-Verlag, 2006.



The End