Genetic & Evolutionary Algorithms: Further Investigation on Selection

Sep. 30, 2019

Prof. Chang Wook Ahn

Meta-Evolutionary Machine Intelligence (MEMI) Lab.

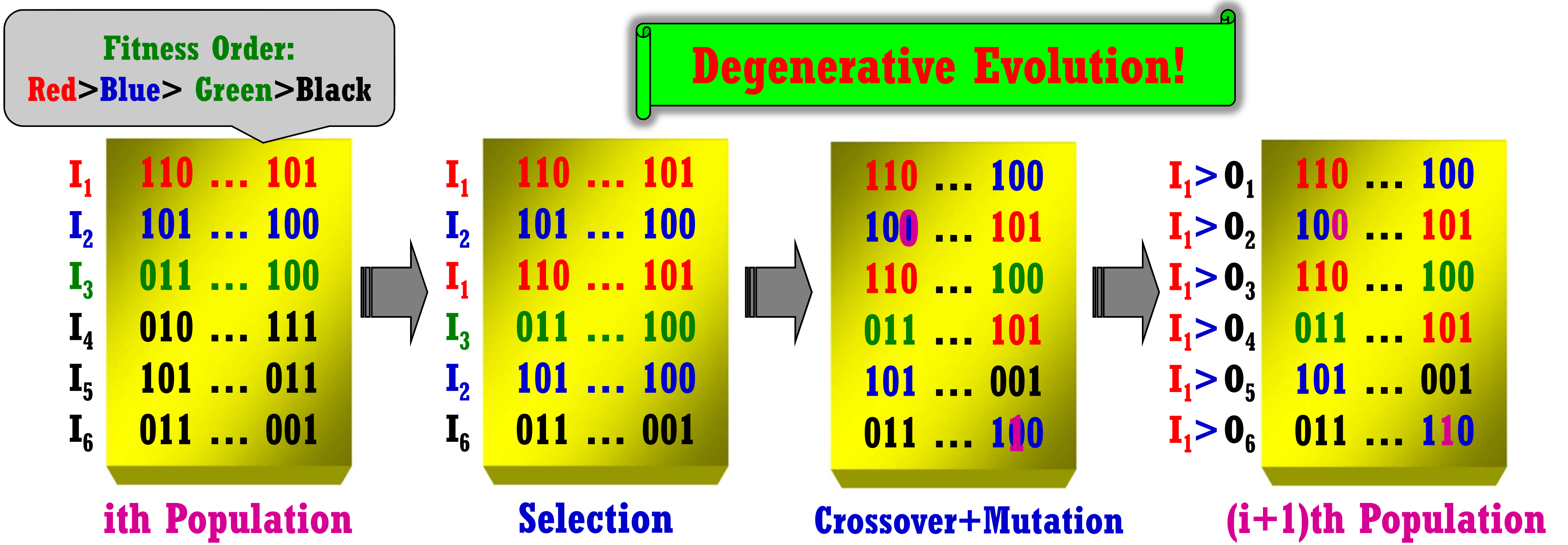
Electrical Eng. & Computer Sci.

Gwangju Inst. Sci. & Tech. (GIST)

Further Detailed Investigation on Selection(2)

New Population (1)

- **Now, isn't there any problem at all in Original GA?**
  - Original GA uses a generational replacement strategy
    - ✓ Selection, crossover, mutation are applied to N individuals (i.e., population)
    - ✓ This set becomes a new population
  - This process seems somewhat weird in view of optimization
    - ✓ We have spent big effort, getting a good solution, to run the risk of throwing it away

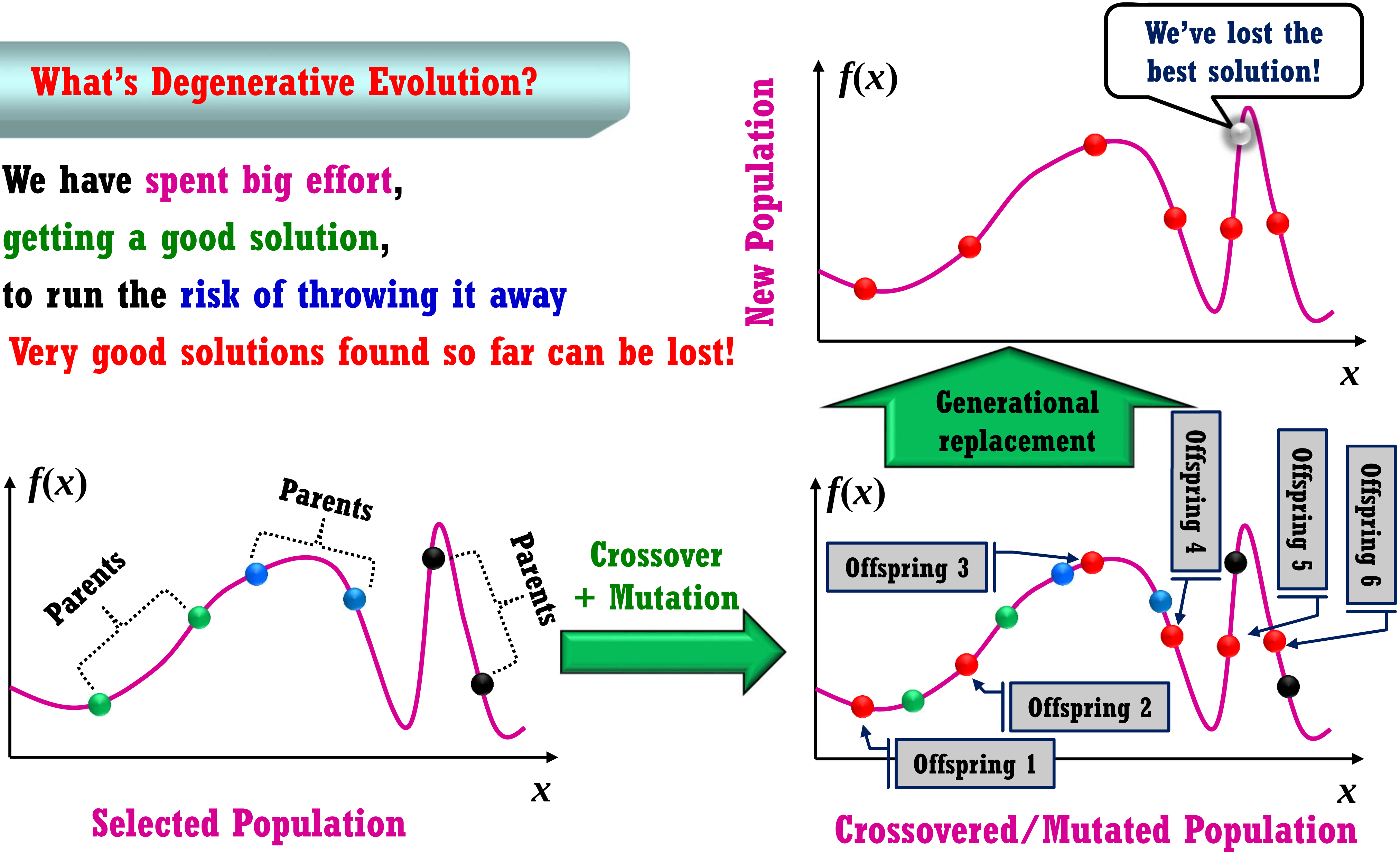


New Population (2)

What's Degenerative Evolution?

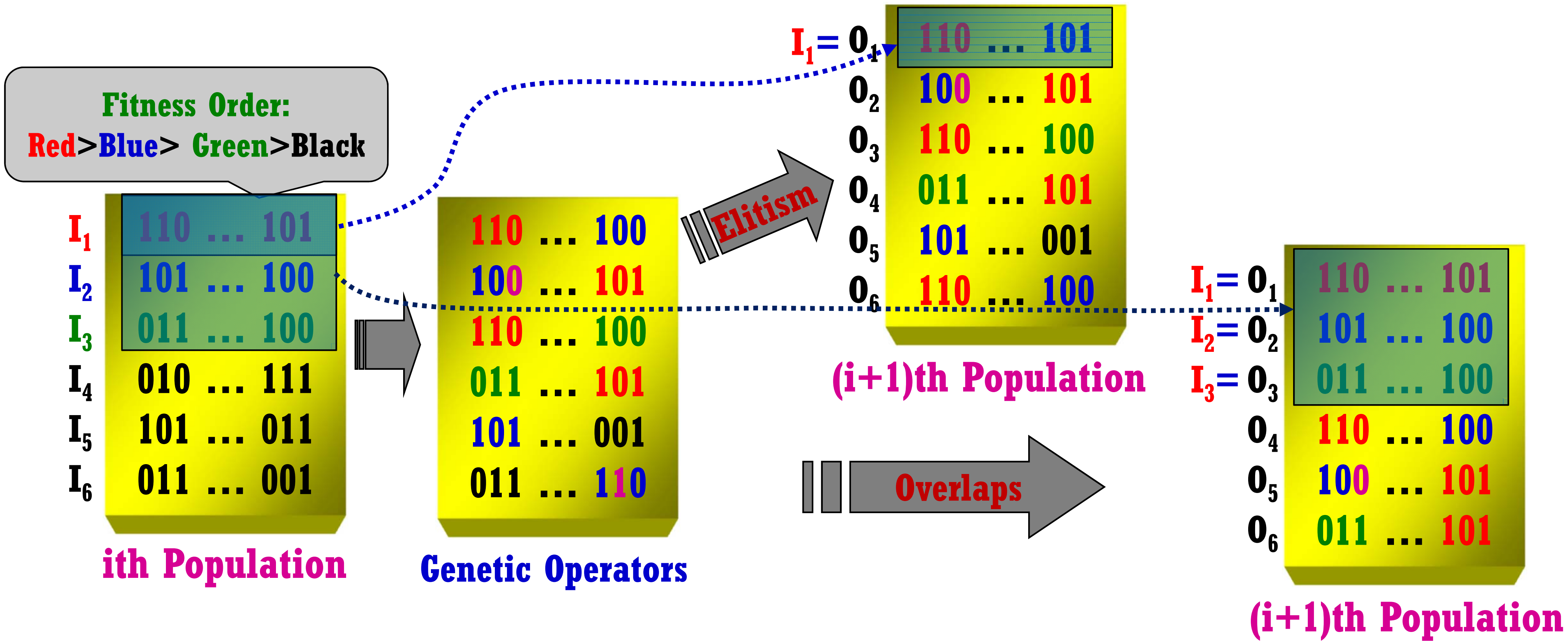
✓ We have spent big effort,
getting a good solution,
to run the risk of throwing it away

➔ Very good solutions found so far can be lost!



New Population (3)

- **Any Idea for Resolving Degenerative Evolution?**
  - Use the concept of *Elitism* and *Population Overlaps*
    - ✓ Elitism ensures the survival of the best individual
    - ✓ Population overlaps replace only a fraction G (generation gap) of the population

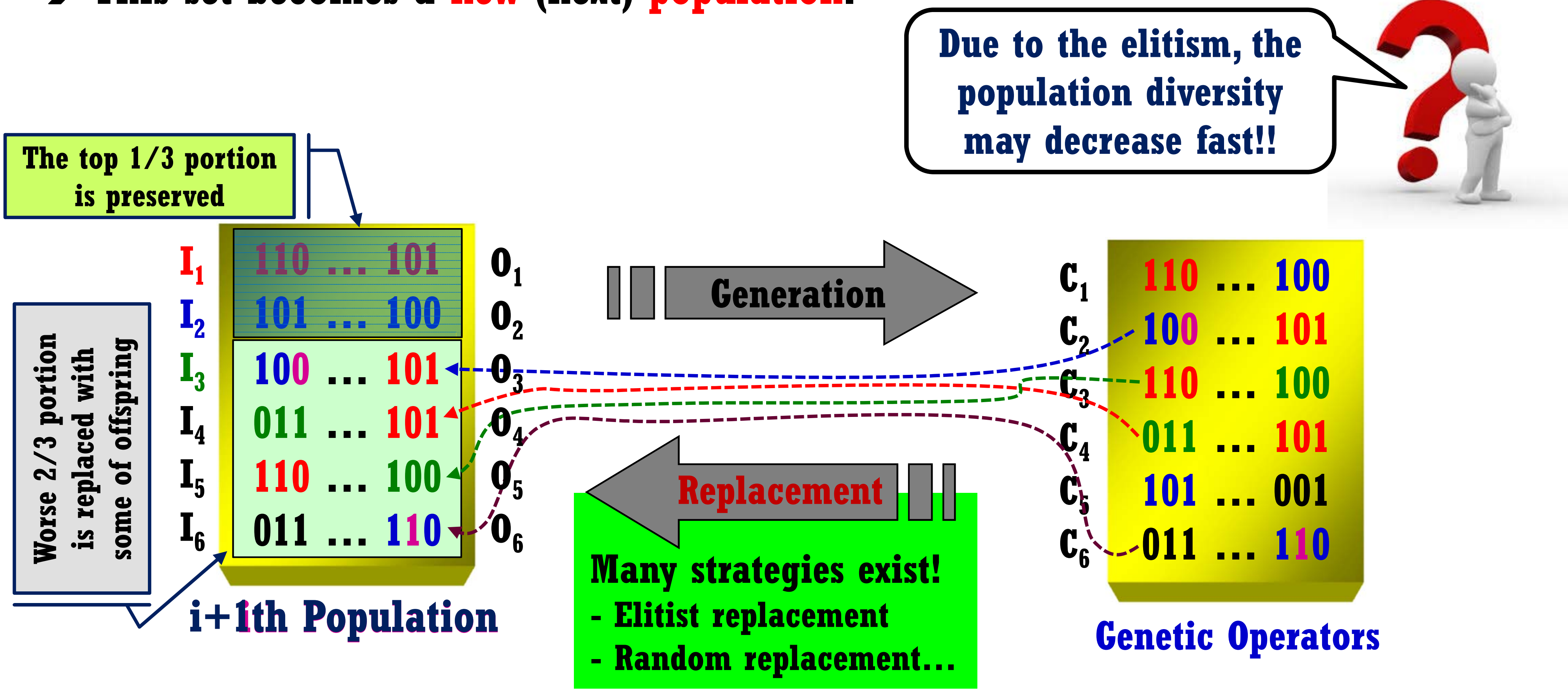


New Population (4)

❖ More General Elitism/Overlap-type Selection?

- The **top τ portion** of the parents are **preserved**
- **Worse (1- τ) portion** of the population are **replaced** with **some of offspring**

➔ This set becomes a **new** (next) **population**.



New Population (5)

❖ Why Population Diversity Important?

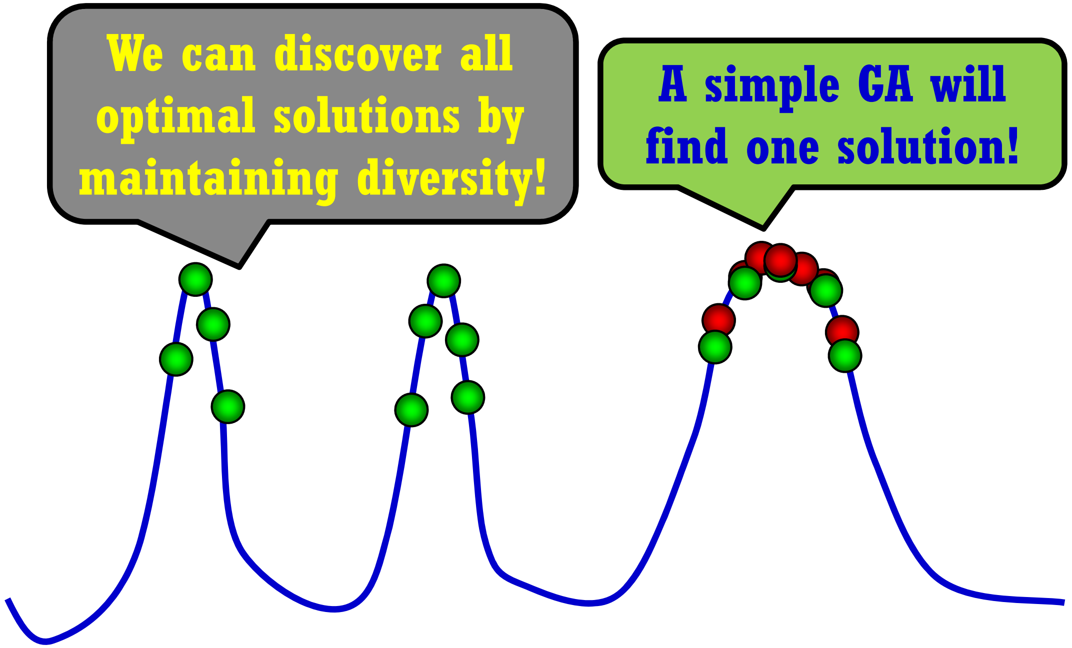
1. Necessary to **cope with** the **premature convergence**
2. Essential to **discover all solutions** in **multimodal** problems

Remind Premature Convergence!

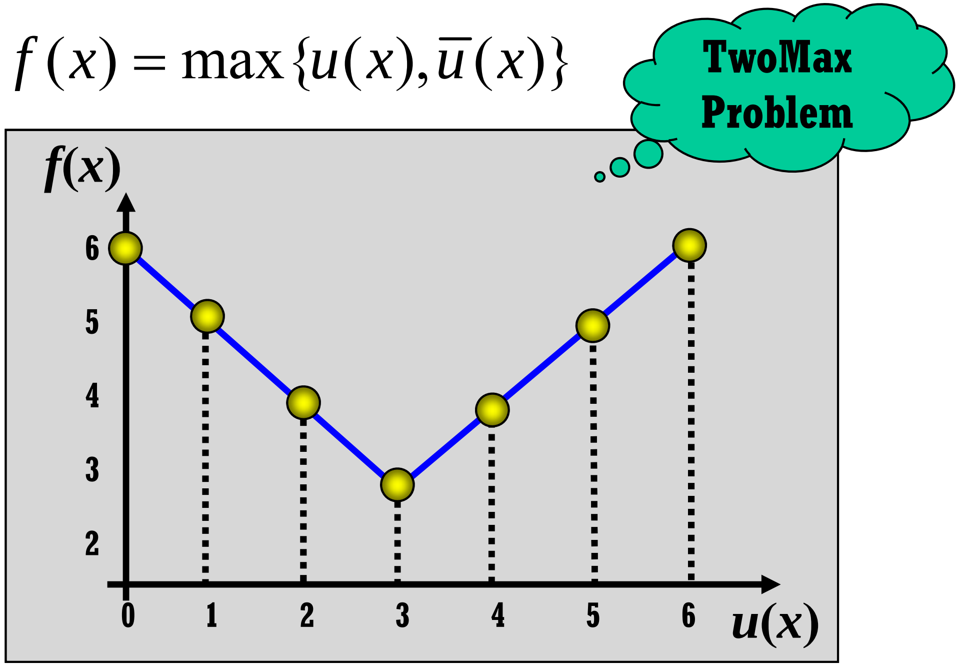
count x_i=1 count x_i=0

$$u(x)=\sum x_i \;,\; \bar{u}(x)=\sum(1-x_i),\ \text{where } x_i\in\{0,1\}$$

$$f(x)=\max\{u(x),\bar{u}(x)\}$$



Multimodal problem:
- It has **three optimal** solutions



A multimodal problem $f(x_1,..,x_6)$

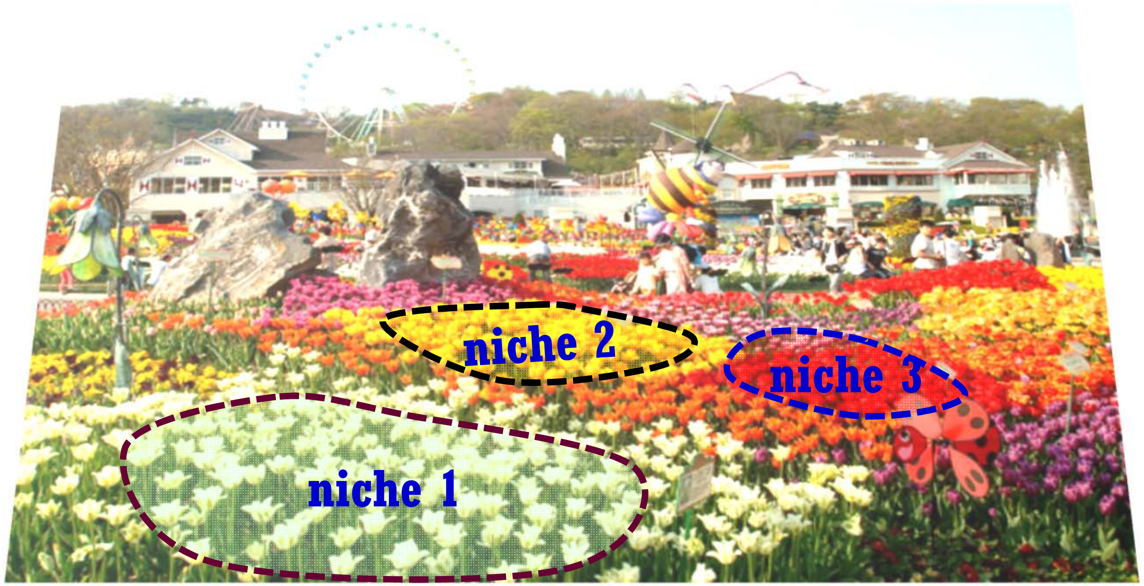
New Population (6)

❖ **Any Idea for Maintaining the Population Diversity?**

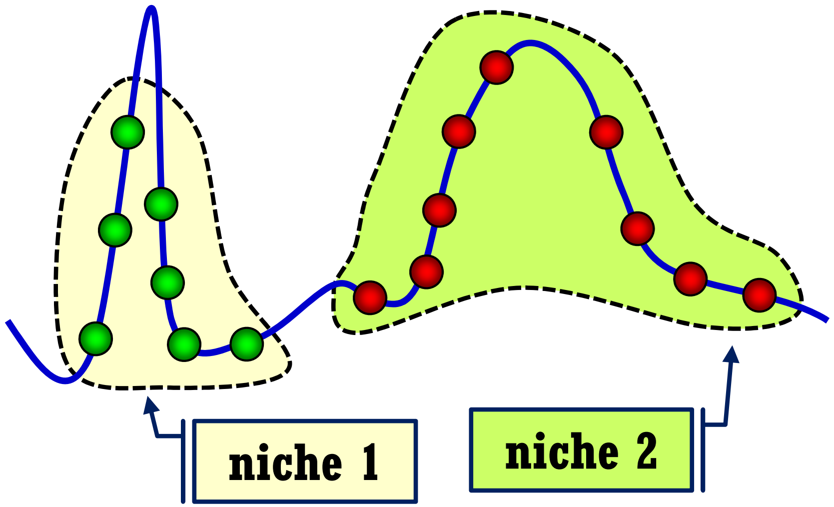
Remind Premature Convergence!

- Another aspect of generating a new population is the diversity maintenance
  - ✓ The concepts of *crowding* and *niching* can be employed:

    A niche in the nature is a set of conditions to which a species is well adapted.
- In the GA, a niche can be treated as a set of similar individuals
  - ✓ <u>IDEA:</u> A newly generated individual replaces one in its own niche!



A Real World Example: Tulips



A GA Example

New Population (7)

❖ Crowding Method

- ➢ A subset of individuals randomly chosen is constructed (as a niche)
- ➢ The closest one for the new individual is replaced with it

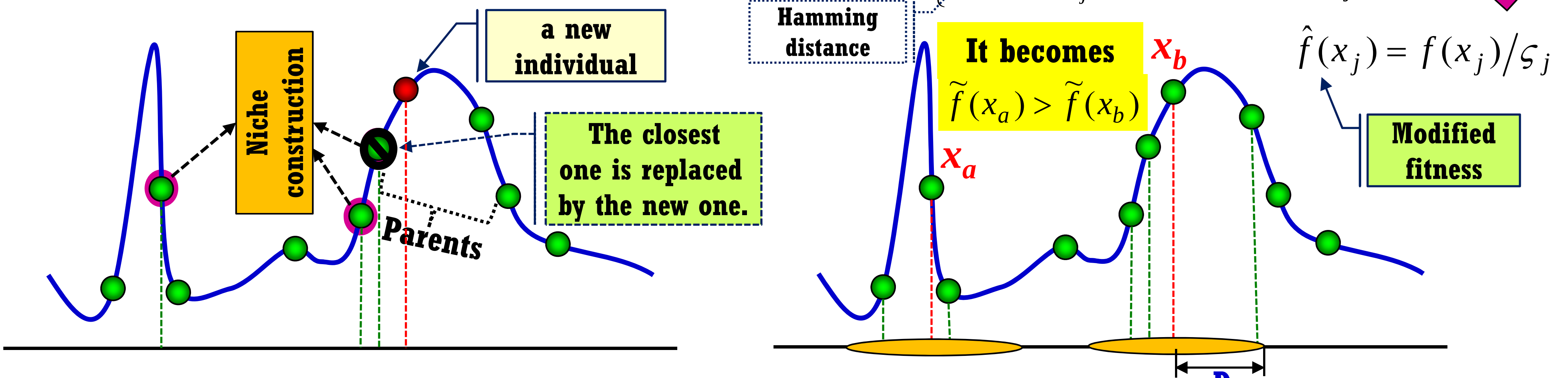
❖ Sharing Method

- ➢ A sharing function, which adjusts the raw fitness function, is used; thereby degrading individuals that occur in clusters comparing to those fairly isolated.

Replacement is done when the new fitness is better; called "Restricted Tournament Replacement".

$$h(d_{ij}) = \begin{cases} 1 - d_{ij}/D, & d_{ij} < D \\ 0, & d_{ij} \geq D \end{cases} \Rightarrow \varsigma_j = \sum_{i \neq j} h(d_{ij})$$

$$\hat{f}(x_j) = f(x_j)/\varsigma_j$$



Crowding Method

- Apt for handling premature convergence

Sharing Method

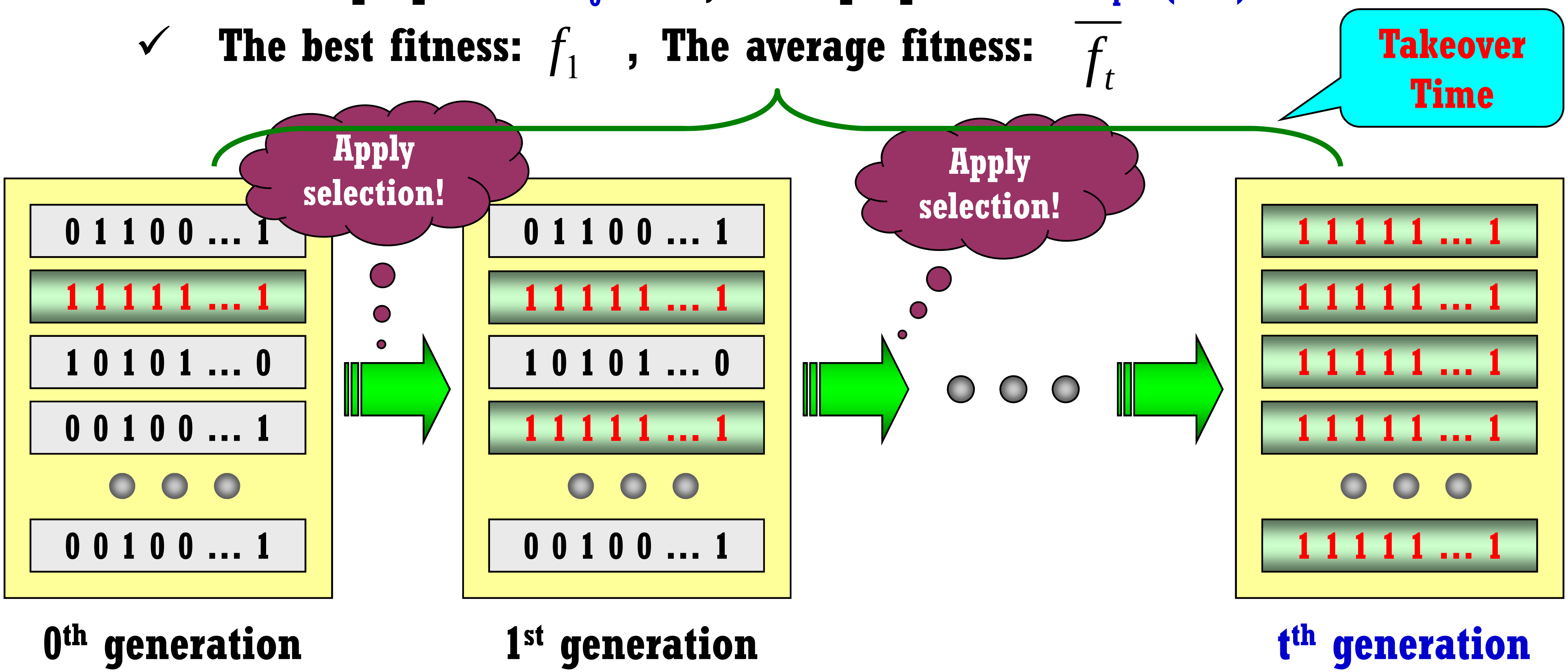
- Good at solving multimodal problems

Further Studies on Selection(3)

Takeover Time (1)

- **Takeover Time** (Convergence Analysis by Applying Selection Alone)
  - Time from (an) initial best individuals until the population is converged
  - Assumption:
    - ✓ Two alternatives: 0 and 1
    - ✓ Initial proportion $P_0=1/N$, Final proportional $P_f=(N-1)/N$
    - ✓ The best fitness: f_1, The average fitness: $\overline{f_t}$

Worst-case Scenario: There is only one best individual!



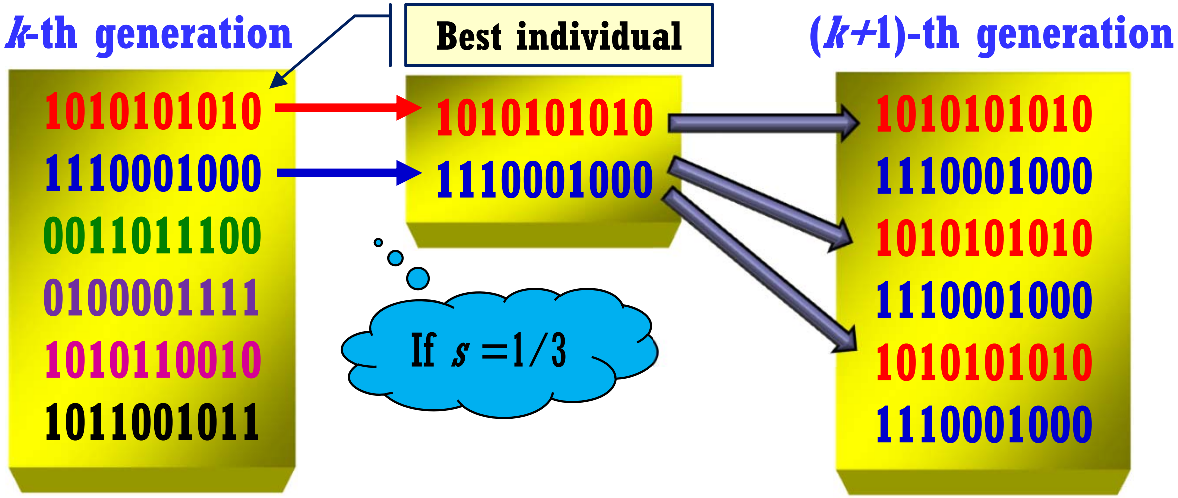
Takeover Time (2)

Truncation Selection

- Truncation selection is easy to analyze: the top $1/s$ individuals are given s copies each!
- If the proportion P_t of best individuals in a population is less than $1/s$, the growth is geometric:

$$P_{t+1} = sP_t$$

- Once the proportion reaches or exceeds $1/s$ ($P_t \geq s^{-1}$), it saturates and the final proportion is one: $P_{t+1}=1$



- Thus, the proportion of best individual under truncation selection alone can be written as:

$$P_t = sP_{t-1} = s^2 P_{t-2} = \cdots = s^t P_0$$

- Assuming $P_0=1/N$, $P_f=1$, and calculating the number of generations to takeover:

$$1 = s^{t^*} \frac{1}{N} \rightarrow s^{t^*} = N \rightarrow t^* \ln s = \ln N$$

- Thus, we have

$$t^* = \frac{\ln N}{\ln s} = \boxed{O(\ln N)}$$

A population is taken to converge as little as O(log N) generations!

Takeover Time (3)

Proportional Selection

You don't have to follow up all the derivations in detail.

- Suppose we have some number of distinct individuals with objective function values $f_j, j \in J$

- The proportion of the ith individual at time t ($P_{i,t}$) is related to the initial proportion of the other individuals and their function values as follows:

$$P_{i,t} = \frac{f_i^t P_{i,0}}{\sum_{j \in J} f_j^t P_{j,0}}$$

- We restrict ourselves to the unit interval and track the proportion of the best individuals, $P_{Best,t}$, where Best={x: 1 – 1/N ≤ x ≤ 1}

$$P_{Best,t} = \frac{\int_{1-N^{-1}}^{1} f^t(x) p_0(x) dx}{\int_0^1 f^t(x) p_0(x) dx}$$

- With p_0=constant, and $f(x)=x^c$, we get

$$P_{Best,t^*} = 1 - (1 - 1/N)^{ct^*+1}$$

- Assuming a final proportion of best individuals $P_f = (N-1)/N$, we have

$$t^* = \frac{1}{c}(N \ln N) = \boxed{O(N \ln N)}$$

Takeover Time (4)

Tournament Selection — i.e., Ordinal Selection

- τ-wise tournament selection is considered:
  Draw τ individuals, and Select the best one!

	ith Population
I_1	101 … 110
I_1	101 … 110
…	… … …
I_N	101 … 011

P_t → P_{t+1} equals the probability that the best individual is picked and survives (by τ-wise tournament) under P_t

- The case the best individual is copied into the next generation:
  If the best individual is drawn at least one time among τ times,

$$P_{t+1} = 1-(1-P_t)^{\tau}$$

$$1-P_{t+1} = (1-P_t)^{\tau}$$

- We use the complementary proportion $Q_t = 1 - P_t$

$$Q_{t+1} = Q_t^{\tau} \Rightarrow Q_t = \overbrace{((((Q_0)^{\tau})^{\tau}\cdots)^{\tau})^{\tau}}^{t \text{ times}} = Q_0^{\tau^t}$$

- By the complementary proportion, $Q_0=(N-1)/N$, $Q_f=Q_{t^*}=1/N$ and taking the natural log:

$$\ln(1/N) = \tau^{t^*}\ln((N-1)/N)$$

- Recognizing that $\ln(1-x) \cong -x$ for small x

$$-\ln N = \tau^{t^*}\ln(1-1/N)$$

$$\ln N = \tau^{t^*}(1/N)$$

- Taking the natural log again:

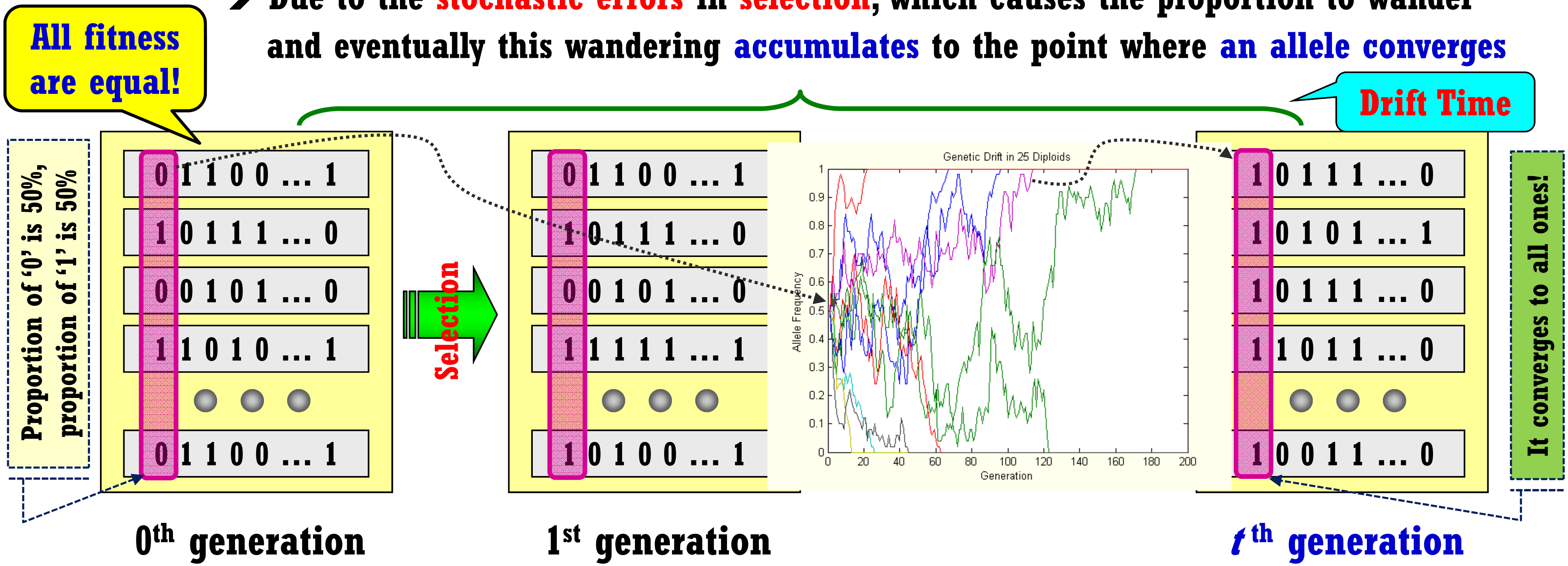
$$\ln\ln N = t^*\ln\tau - \ln N$$

$$t^* = \frac{\ln N + \ln\ln N}{\ln\tau} = \boxed{O(\ln N)}$$

Last But Not Least Issues on Selection(3)

Drift Time

- **Drift Time (Convergence when Selection chooses for No Reason)**
  - **Selection can cause convergence in finite population with little fitness difference**
    - ✓ Assume a population initialized with 50 ones and 50 zeros where $f_1=f_0$.
    - ✓ In a random trial, the population (allele) wanders about from the initial condition until it finally comes to converge to zeros
    - ✓ In another trial, the population (allele) is equally likely to converge to all ones
    - ➔ Due to the stochastic errors in selection, which causes the proportion to wander and eventually this wandering accumulates to the point where an allele converges



Hitchhiking

❖ Hitchhiking Problem

- In general, GA suffers from the Hitchhiking problem
  - ✓ Once a high-fitness schema (i.e., salient genes) is discovered, the unfit genes, especially those next to the fit parts, spread along with the fit ones
  - ✓ This slows the discovery of good schemata in those positions
  - ✓ Early convergence to wrong schemata limits the effectiveness of crossover

➔ This degrades the performance of GA w.r.t. optimality as well as convergence

Fitness Order: $I_1 > I_2 > I_3 > I_4 > I_5 > I_6$

Salient genes : f(****101*****) gives a big contribution to the fitness

Hitchhikers : f(*******001**) gives a small contribution

I_1	1	1	0	1	1	0	1	0	0	1	1	0
I_2	1	0	1	1	1	0	1	0	1	1	0	0
I_3	0	1	0	1	1	1	1	0	1	1	0	0
I_4	0	1	1	1	1	0	1	0	0	1	1	1
I_5	1	0	1	0	1	0	0	1	0	1	1	0
I_6	0	1	1	1	1	0	0	1	0	0	0	1

t-th Population

Selection ➔

I_1	1	1	0	1	1	0	1	0	0	1	1	0
I_1	1	1	0	1	1	0	1	0	0	1	1	0
I_2	1	0	1	1	1	0	1	0	1	1	0	0
I_3	0	1	0	1	1	1	1	0	1	1	0	0
I_4	0	1	0	1	1	0	1	0	0	1	1	1
I_6	0	1	1	1	1	0	0	1	0	0	0	1

(t+1)-th Population

Crossover will do nothing much on them!

Summary

- There are **Two Selection Categories**!
  - **Proportional Selection**; e.g., Roulette-Wheel selection, Scaling
  - **Ordinal Selection**; e.g., Ranking selection, Tournament selection, etc.
- Generally, **Proportional Selection** tends to have **Premature Convergence**
  - Thus, the **scaling method** has been employed.
  - But, it **needs** to do **re-scaling** at every generation.
- **Ordinal Selection** is quite **robust** in this regard.
  - It can adjust **selection pressure** at a **constant level** what we want.
  - But, the **ranking selection** is somewhat **restricted**.
  - **Tournament selection does not have** such constraints.
- In terms of **takeover time** (i.e., convergence with selection only)
  - **Proportional Selection** has **O(NlnN)**, but **Ordinal Selection** has **O(lnN)**.

Thus, we conclude that the **Tournament Selection is the most promising choice**!