

Evolutionary Algorithms: **Building Blocks & Linkage Identification**

October 14, 2019

Prof. Chang Wook Ahn



Meta-Evolutionary Machine Intelligence (MEMI) Lab.
Electrical Eng. & Computer Sci.
Gwangju Inst. Sci. & Tech. (GIST)



Further
Studies on
Building-Blocks (2)



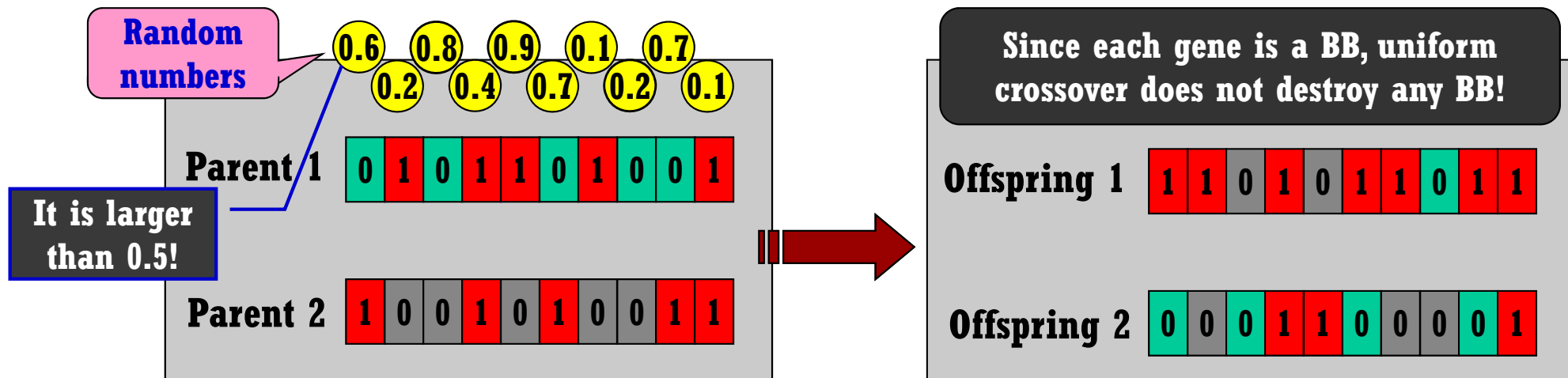
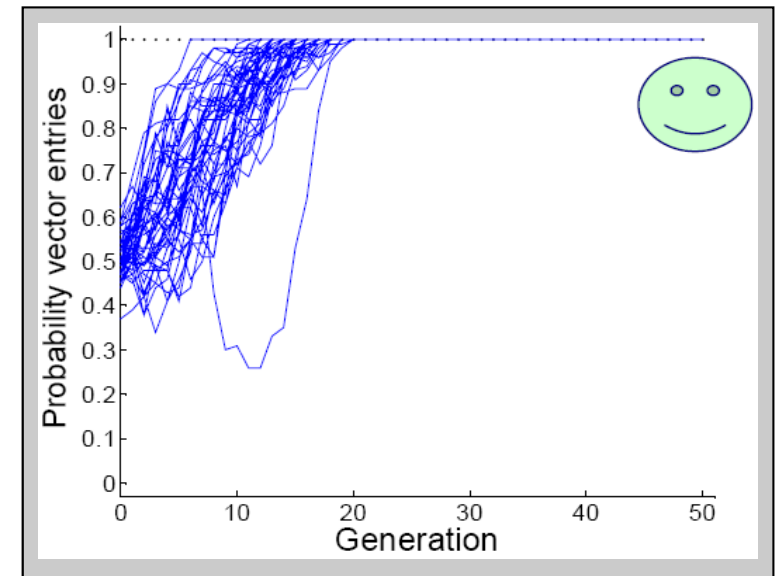


Building-Blocks: BBs (1)

❖ One-Max Problem (i.e., Order-1 BB)

- It is the most **simplest** test problem since **each gene comes under a BB**.
 - ✓ The optimum is $(x_1, x_2, \dots, x_n) = 11\dots 1$
 - ✓ 1 outperforms 0 on average; that is, $f(1####...) > f(0####...)$
- Thus, **uniform crossover** is the best choice!

$$\max f_{OneMax}(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i \quad \forall x_i \in \{0,1\}$$



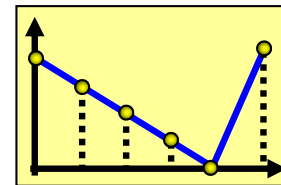


Building-Blocks: BBs (2)

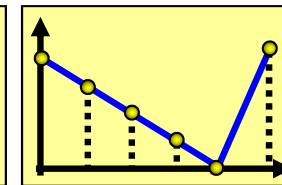
❖ Deceptive Problem (i.e., Order-k BB)

- It is a quite **hard** problem since each group (i.e., a set of genes) contributes via **Trap**.
 - ✓ The optimum is $(x_1, x_2, \dots, x_n) = 11\dots 1$
 - ✓ Single bits are **misleading**; i.e., $f(1####\dots) < f(0####\dots)$
 - ✓ **BB** must always be **preserved**!
- Thus, **BB-wise uniform crossover** is the best choice!

Ex) 5-bit Deceptive Problem;
i.e., Concatenated Traps
 $F(X) = \sum f(B)$

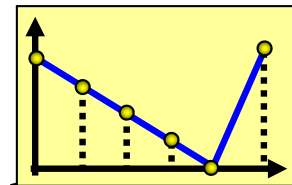


Trap1



Trap2

...



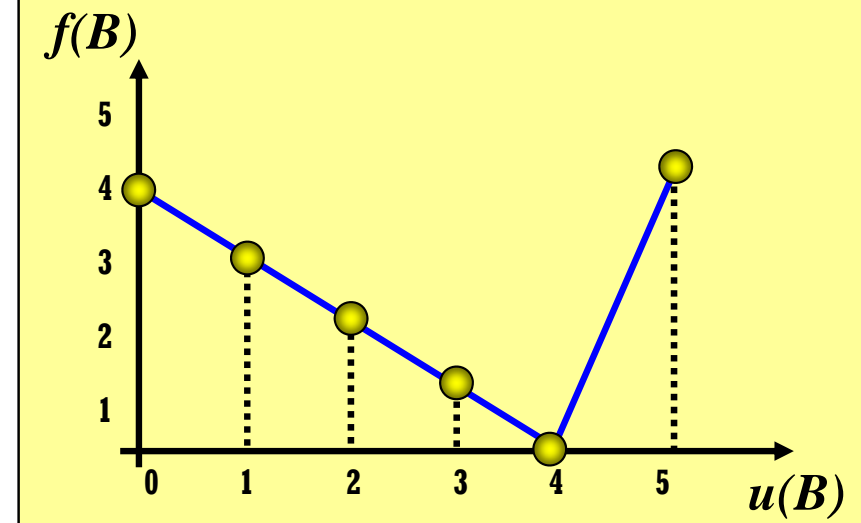
Trapk

$$\max f_{DP}(x_1, \dots, x_n) = \sum_{i=1}^{n/k} f_{trap}(u_i)$$

$$f_{trap}(u_i) = \begin{cases} k, & \text{if } u_i = k \\ k - 1 - u_i, & \text{otherwise} \end{cases}$$

$$u_i = \sum_{j=1}^k x_j \quad \forall x_j \in \{0,1\}$$

Ex) 5-bit Trap Function;
if $u(B) = 5$, then $f(B) = 5$;
else $f(B) = 4 - u(B)$;



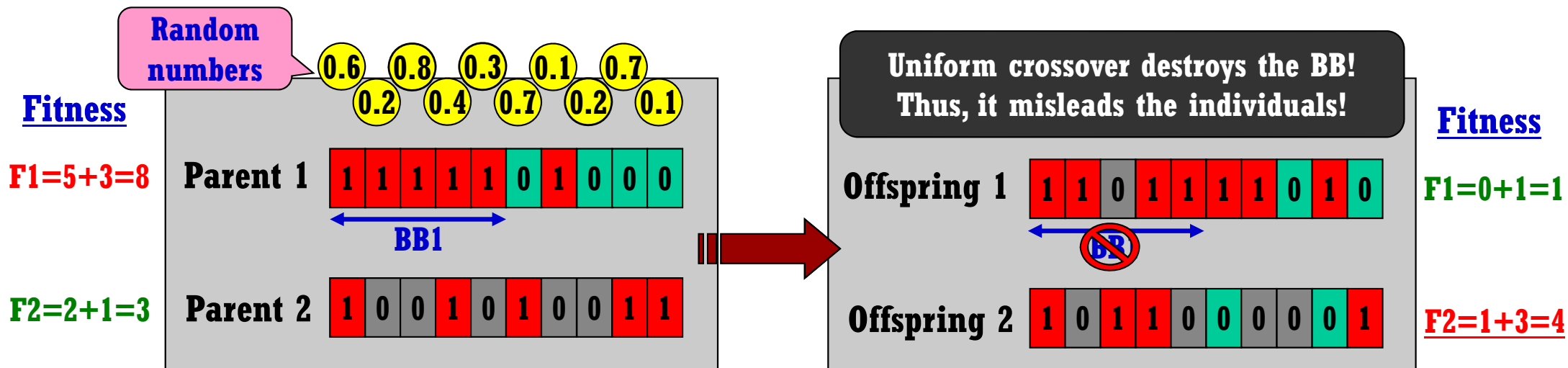
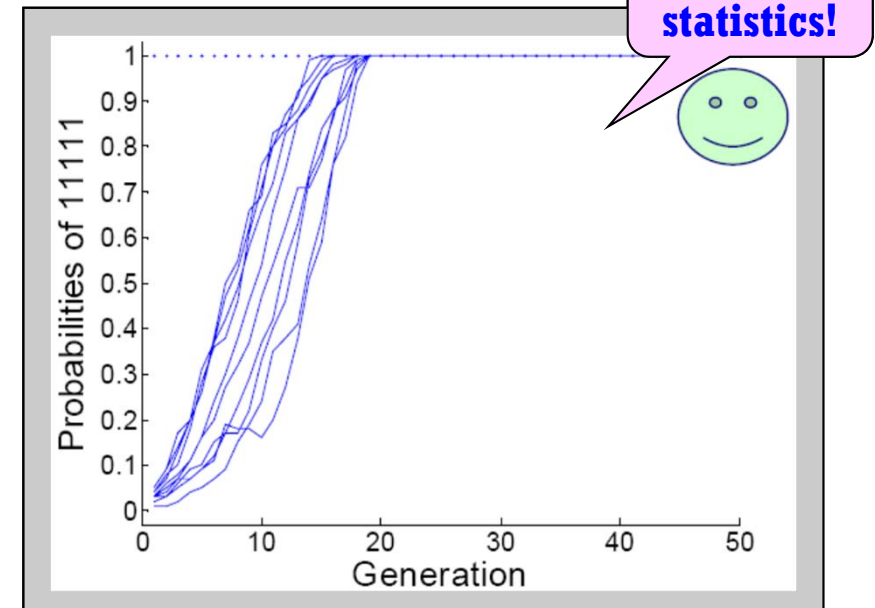


Building-Blocks: BBs (3)

❖ Why Higher-Order Statistics for Deceptive Problem?

➤ Ex) 5-bit Deceptive Problem

- ✓ It consists of 5-bit Trap Function: whose optimum is '11111'
- ✓ But, $f(1****)=1.375$, $f(0****)=2$;
- ✓ Single bits (i.e., 1-bit statistics, uniform crossover) are misleading!
- ✓ Thus, 5-bit statistics is required; thereby BB-wise uniform crossover





Building-Blocks: BBs (4)

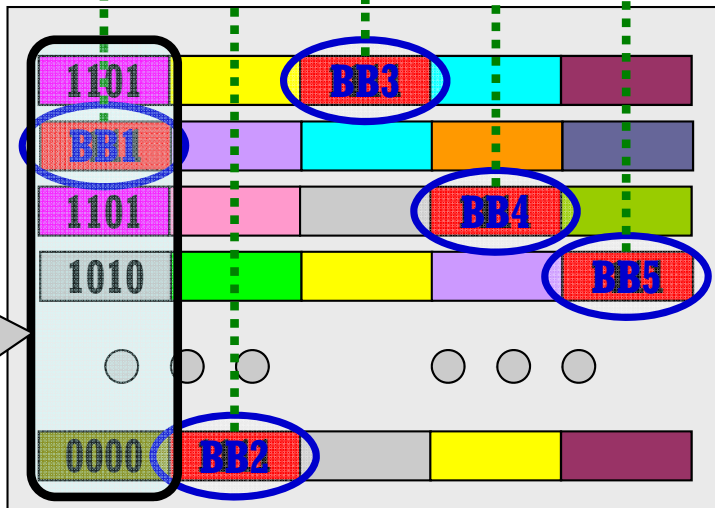
❖ Can We Recognize BBs in the process of Individuals?

- It is **impossible** to exactly Capture BBs themselves!
- But **BBs** must be **assembled** in any way!
- Only **information** we can know is the BB information, i.e., **linkage of variables**
- The **best way** we can do is to apply **uniform crossover at the level of BBs**;
this is the **BB-wise Uniform crossover**

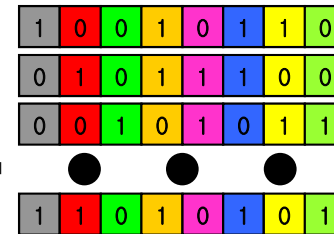
Optimum

1111 1111 1111 1111 1111

Population



Population



BB1

x_1 x_2

BB2

x_5 x_7 x_8

It is possible to assemble BBs!

BBs are always preserved!

BB1

###

BB-wise Crossover

@@@

BB2

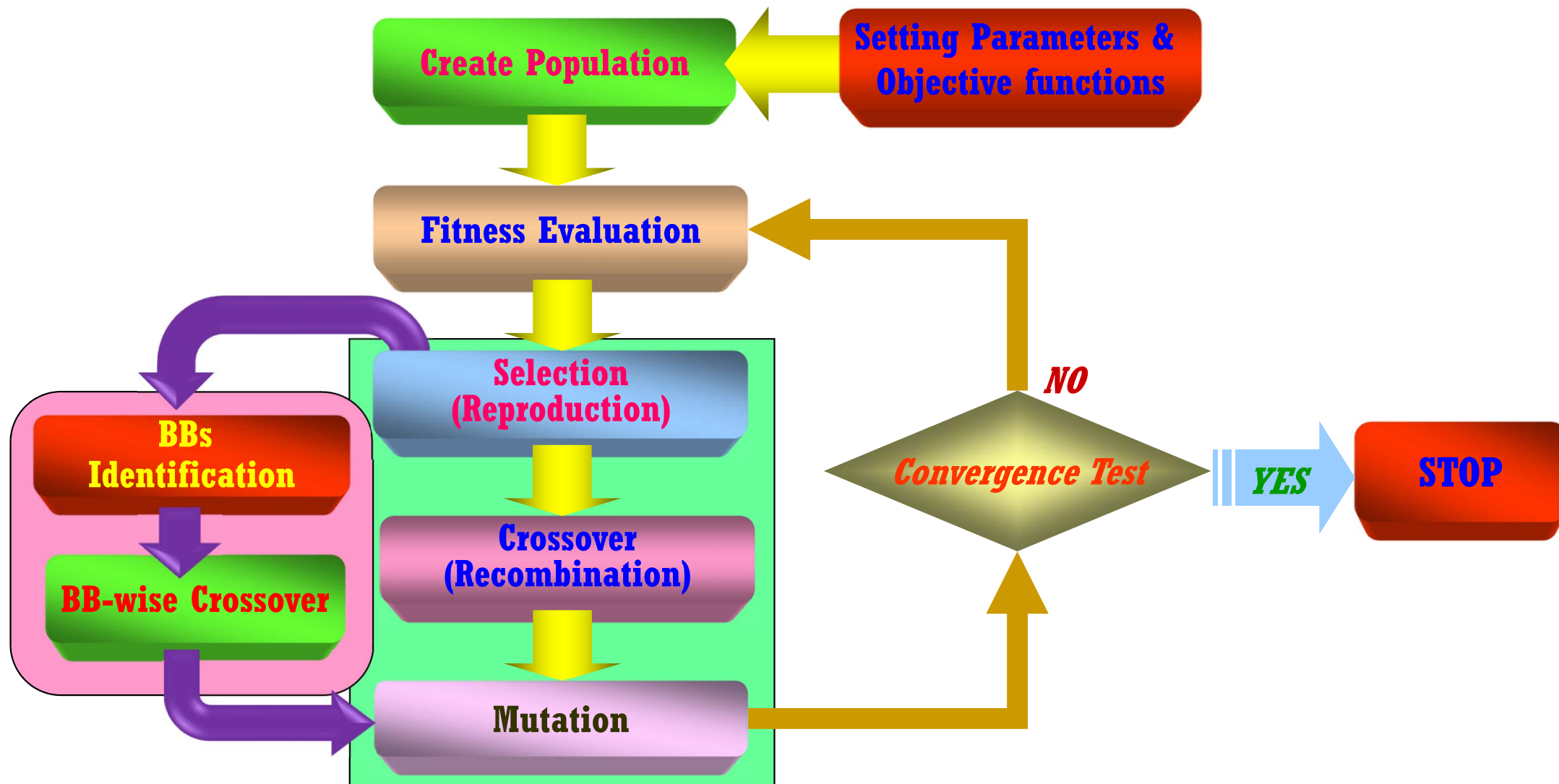
Linkage (or BBs)
Identification -
Brief Introduction





Linkage Identification (1)

❖ Where to Place the BBs (i.e., Linkage) Identification?

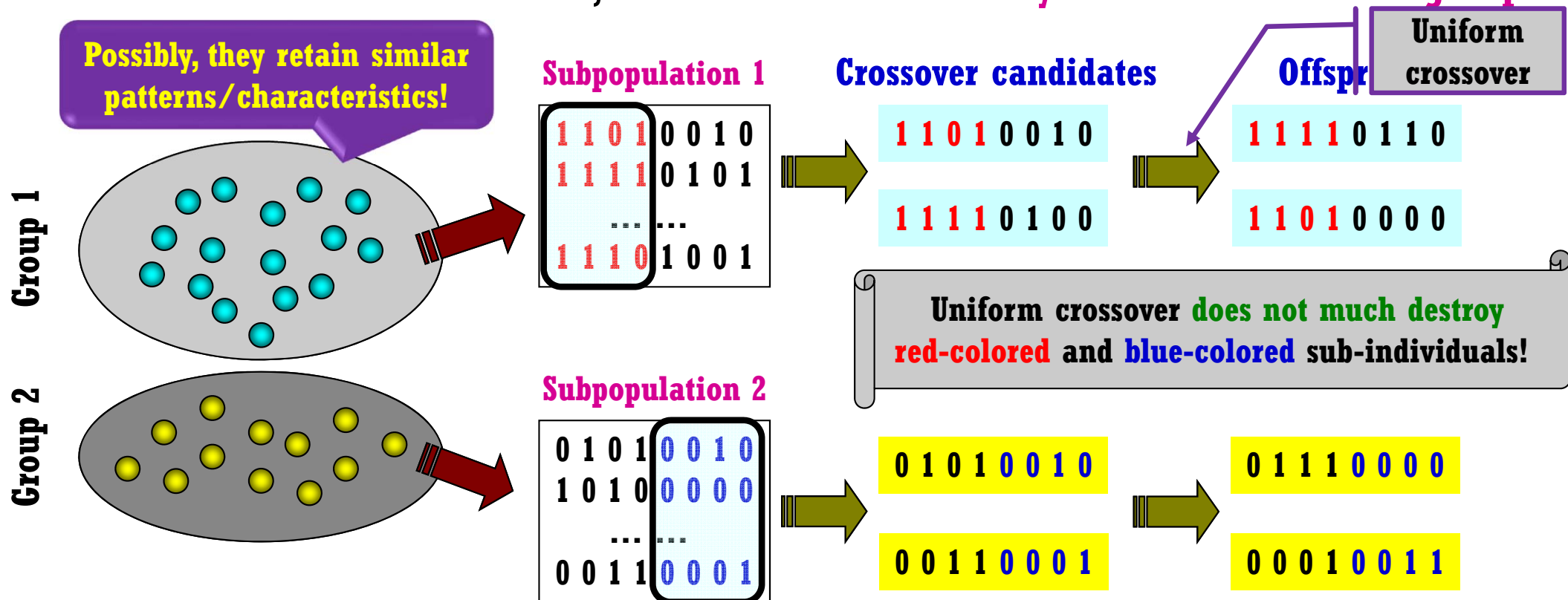




Linkage Identification (2)

❖ BB-Preserving Crossover using Clustering

- The most **simplest** approach to preserving BBs when performing crossover
- It uses **clustering** techniques **in the decision space** (i.e., decision variable)
- The individuals belonging to the same cluster have **some similar patterns**.
- Thus, **crossover** is performed on the individuals **within the same group**.
- It can **somewhat treat BBs**, **but** still need to **carefully recombine different group**.





Linkage Identification (3)



❖ Nonlinearity Check by Perturbation

- It identifies linkage groups by **detecting nonlinearity by (bit) perturbations**
 - ✓ If an arbitrary nonlinearity is detected by perturbations in a pair of genes, they are included in a linkage group
- It is **not efficient** for the **quasi-decomposable** or **noisy** fitness function, and so on. Among other things, it requires **very high computation costs** by perturbations;

A large number of fitness evaluations

Consider an individual $x = (x_1, x_2, \dots, x_n)$, $\forall x_i \in \{0,1\}$

Let $\tilde{x}_i = 1 - x_i$, and $f(x)$ is the fitness function

$$\Delta f_i(x) = f(\dots \tilde{x}_i \dots) - f(\dots x_i \dots)$$

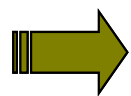
$$\Delta f_j(x) = f(\dots \tilde{x}_j \dots) - f(\dots x_j \dots)$$

$$\Delta f_{ij}(x) = f(\dots \tilde{x}_i \dots \tilde{x}_j \dots) - f(\dots x_i \dots x_j \dots)$$

Note:

If $\Delta f_{ij}(s) = \Delta f_i(s) + \Delta f_j(s)$

Then, x_i and x_j are not members of a linkage set, or they are linked but **linearity** exists!
(It is safe for them to be **unlinked**!)



If $\Delta f_{ij}(s) \neq \Delta f_i(s) + \Delta f_j(s)$

x_i and x_j are members of the same linkage group!



Linkage Identification (4)



❖ Nonlinearity Check by Perturbation [Example]

➤ **5-bit Trap function**, its variables $x=(x_1, x_2, x_3, x_4, x_5)$

✓ All the variables are included in a linkage set

Population

1	1	0	1	0
1	1	1	1	0
...
1	1	1	1	1

For the pair of (x_1, x_2) at the 1st individual,

$$\Delta f_1(x) = f(01010) - f(11010) = 2 - 1 = 1$$

$$\Delta f_2(x) = f(10010) - f(11010) = 2 - 1 = 1$$

$$\Delta f_{12}(x) = f(00010) - f(11010) = 3 - 1 = 2$$

$$\Delta f_{12}(x) = \Delta f_1(x) + \Delta f_2(x)$$

Linkage set:
{ }

For the pair of (x_1, x_3) at the 1st individual,

$$\Delta f_1(x) = f(01010) - f(11010) = 2 - 1 = 1$$

$$\Delta f_3(x) = f(11110) - f(11010) = 0 - 1 = -1$$

$$\Delta f_{13}(x) = f(01110) - f(11010) = 1 - 1 = 0$$

$$\Delta f_{13}(x) = \Delta f_1(x) + \Delta f_3(x)$$

Linkage set:
{ }



For the pair of (x_1, x_5) at the 2nd individual,

$$\Delta f_1(x) = f(01110) - f(11110) = 1 - 0 = 1$$

$$\Delta f_5(x) = f(11111) - f(11110) = 5 - 0 = 5$$

$$\Delta f_{15}(x) = f(01111) - f(11110) = 0 - 0 = 0$$

$$\Delta f_{15}(x) \neq \Delta f_1(x) + \Delta f_5(x)$$

Linkage set:
{ x_1, x_5 }

For the pair of (x_1, x_2) at the nth individual,

$$\Delta f_1(x) = f(01111) - f(11111) = 0 - 5 = -5$$

$$\Delta f_2(x) = f(10111) - f(11111) = 0 - 5 = -5$$

$$\Delta f_{12}(x) = f(00111) - f(11111) = 1 - 5 = -4$$

$$\Delta f_{12}(x) \neq \Delta f_1(x) + \Delta f_2(x)$$

Linkage set:
{ x_1, x_2, x_5 }

For the pair of (x_1, x_3) at the nth individual,

$$\Delta f_1(x) = f(01111) - f(11111) = 0 - 5 = -5$$

$$\Delta f_3(x) = f(11011) - f(11111) = 0 - 5 = -5$$

$$\Delta f_{13}(x) = f(01011) - f(11111) = 1 - 5 = -4$$

$$\Delta f_{13}(x) \neq \Delta f_1(x) + \Delta f_3(x)$$

Linkage set:
{ x_1, x_2, x_3, x_5 }

For the pair of (x_1, x_4) at the nth individual,

$$\Delta f_1(x) = f(01111) - f(11111) = 0 - 5 = -5$$

$$\Delta f_4(x) = f(11101) - f(11111) = 0 - 5 = -5$$

$$\Delta f_{14}(x) = f(01101) - f(11111) = 1 - 5 = -4$$

$$\Delta f_{14}(x) \neq \Delta f_1(x) + \Delta f_4(x)$$

Linkage set:
{ x_1, x_2, x_3, x_4, x_5 }





Linkage Identification (5)

❖ Covariance-based Dependency Discovery

➤ **Covariance** of X and Y: $\text{COV}(X,Y) = E[XY] - E[X]E[Y]$

✓ Paris of **independent** r.v.s have **covariance zero!**

➤ **Correlation Coefficient**: $\rho(X,Y) = \text{COV}(X,Y) / \sigma_X \sigma_Y$

✓ Normalized covariance, i.e., $|\rho(X,Y)| \leq 1$

✓ If X and Y are **independent**, $|\rho(X,Y)|$ goes to '0'

➤ We take **the pair** as a linkage set that returns **the highest** correlation coefficient.

Expectation of r.v. X

$$E[X] = \sum x_i p(x_i)$$

[Algorithm]

[Step1] Compute all $\rho(x_i, x_j)$ for all pairs of (x_i, x_j)

[Step2] Find the pair (x_i, x_j) of the highest value
and $|\rho(x_i, x_j)| > \Delta$

[Step3] Contain x_i and x_j as a linkage set

[Step4] Merge x_i and x_j into an another variable

[Step5] Iterate the above procedures
until no more linkage is found.

Population	x y	
	1	1
	0	1
	1	0
	0	0
	1	1
	0	0

$$p(x=0)=1/2, p(x=0,y=0)=2/6$$

$$p(x=1)=1/2, p(x=0,y=1)=1/6$$

$$p(y=0)=1/2, p(x=1, y=0)=1/6$$

$$p(y=1)=1/2, p(x=1,y=1)=2/6$$

$$E[x] = 0*(1/2) + 1*(1/2) = 1/2$$

$$E[y] = 0*(1/2) + 1*(1/2) = 1/2$$

$$E[xy] = 0*(2/6) + 0*(1/6) + 0*(1/6) + 1*(2/6) = 1/3$$

$$\sigma_x = \sqrt{\{(0-0.5)^2*(1/2) + (1-0.5)^2*(1/2)\}} = 1/2$$

$$\sigma_y = \sqrt{\{(0-0.5)^2*(1/2) + (1-0.5)^2*(1/2)\}} = 1/2$$

$$\text{COV}(x,y) = 1/3 - (1/2)*(1/2) = 1/12$$

$$|\rho(x,y)| = |(1/12) / (1/2 * 1/2)| = 1/3 = 0.333$$



Linkage Identification (6)



❖ Covariance-based Dependency Discovery [Example]

- At k-th generation, we obtain the following selected population.
- We employ the **corr. coeff.** with $\Delta=0.4$

Selected Population

x_1 x_2 x_3 x_4

1	1	1	1
0	1	0	1
1	0	1	1
0	0	1	1
1	1	1	0
1	1	0	1
1	1	0	0
0	0	1	1
1	0	1	1
1	1	0	1

$$\begin{aligned} p(x_1=0) &= 3/10, p(x_1=1) = 7/10 \\ p(x_2=0) &= 4/10, p(x_2=1) = 6/10 \\ p(x_3=0) &= 4/10, p(x_3=1) = 6/10 \\ p(x_4=0) &= 2/10, p(x_4=1) = 8/10 \end{aligned}$$

$$\begin{aligned} p(x_1=0, x_2=0) &= 2/10, p(x_1=0, x_3=0) = 1/10 \\ p(x_1=0, x_2=1) &= 1/10, p(x_1=0, x_3=1) = 2/10 \\ p(x_1=1, x_2=0) &= 2/10, p(x_1=1, x_3=0) = 3/10 \\ p(x_1=1, x_2=1) &= 5/10, p(x_1=1, x_3=1) = 4/10 \end{aligned}$$

$$\begin{aligned} p(x_1=0, x_4=0) &= 0/10, p(x_2=0, x_3=0) = 0/10, p(x_2=0, x_4=0) = 0/10, p(x_3=0, x_4=0) = 1/10 \\ p(x_1=0, x_4=1) &= 3/10, p(x_2=0, x_3=1) = 4/10, p(x_2=0, x_4=1) = 4/10, p(x_3=0, x_4=1) = 3/10 \\ p(x_1=1, x_4=0) &= 2/10, p(x_2=1, x_3=0) = 4/10, p(x_2=1, x_4=0) = 2/10, p(x_3=1, x_4=0) = 1/10 \\ p(x_1=1, x_4=1) &= 5/10, p(x_2=1, x_3=1) = 2/10, p(x_2=1, x_4=1) = 4/10, p(x_3=1, x_4=1) = 5/10 \end{aligned}$$

$$\begin{aligned} E[x_1] &= 7/10, E[x_2] = 6/10, E[x_3] = 6/10, E[x_4] = 8/10 \\ E[x_1 x_2] &= 0*(2/10) + 0*(1/10) + 0*(2/10) + 1*(5/10) = 1/2 \\ E[x_1 x_3] &= 0*(1/10) + 0*(2/10) + 0*(3/10) + 1*(4/10) = 2/5 \\ &\dots\dots\dots \\ \sigma x_1 &= 0.45483, \sigma x_2 = 0.4899, \dots\dots\dots \end{aligned}$$

$$|\rho(x_1, x_2)| = |\{0.5 - (0.7*0.6)\} / (0.4583*0.4899)| = 0.3563$$

$$|\rho(x_1, x_3)| = 0.0891 \quad |\rho(x_1, x_4)| = 0.3273$$

$$|\rho(x_2, x_3)| = 0.6667 \quad |\rho(x_2, x_4)| = 0.4082$$

$$|\rho(x_3, x_4)| = 0.1021$$

For pairs, $\rho(x_2, x_3)$ is the highest & it is greater than 0.4.
Thus, **x_2 and x_3 are linked!**
The similar procedures continues!



Linkage Identification (7)



❖ Information Theoretic Identification

- **Mutual information** can be used.
- It considers every pair of genes, and takes the pair as a linkage set, which returns the **highest mutual information**.

$$I(x_i, x_j) = \sum_{x_i, x_j} P(x_i = a, x_j = b) \log \frac{P(x_i = a, x_j = b)}{P(x_i = a)P(x_j = b)}$$
$$= H(x_i) - H(x_i | x_j) = H(x_j) - H(x_j | x_i)$$

Mutual Information:
Relative entropy between
the joint distribution $p(x_i, x_j)$ &
the product distribution $p(x_i)p(x_j)$

[Algorithm]

[Step1] Compute all $I(x_i, x_j)$ for all pairs of (x_i, x_j)

[Step2] Find the pair (x_i, x_j) of the highest value
and $I(x_i, x_j) > \Delta$

[Step3] Contain x_i and x_j as a linkage set

[Step4] Merge x_i and x_j into an another variable

[Step5] Iterate the above procedures
until no more linkage is found.

Population	x y	
	1	1
	0	1
	1	0
	0	0
	1	1
	0	0

$$p(x=0)=1/2, p(x=0, y=0)=2/6$$

$$p(x=1)=1/2, p(x=0, y=1)=1/6$$

$$p(y=0)=1/2, p(x=1, y=0)=1/6$$

$$p(y=1)=1/2, p(x=1, y=1)=2/6$$

$$I(x;y) = (2/6)\log(4/3) + (1/6)\log(2/3) + (1/6)\log(2/3) + (2/6)\log(4/3) = \mathbf{0.0566}$$



Linkage Identification (8)



◆ What is the **Entropy**?

- ✓ A measure of **uncertainty** of a r.v. in terms of bits (\log_2) or nats (\log_e).

$$H(X) = -\sum_{x_i} p(x_i) \log p(x_i) = E\left[\log \frac{1}{p(X)}\right]$$

◆ What is the **Conditional Entropy**?

- ✓ The expected entropy of conditional distr., averaged over the conditioning r.v.

$$\begin{aligned} H(Y | X) &= \sum_{x_i} p(x_i) H(Y | X = x_i) \\ &= -\sum_{x_i} \sum_{y_i} p(x_i, y_i) \log p(y_i | x_i) \end{aligned}$$

◆ What is the **Relative Entropy**?

- ✓ A measure of the **distance** between two probability distributions
- ✓ **Non-negative, zero if $p(X)$ & $q(X)$ are same!**

Also, it is referred to as **Kullback Leibler distance**

$$D(p(X) \parallel q(X)) = \sum_{x_i} p(x_i) \log \frac{p(x_i)}{q(x_i)}$$

◆ What is the **Mutual Information**?

- ✓ **Relative entropy** between joint distr. $p(X, Y)$ & product distr. $p(X)p(Y)$
- ✓ **Reduction in the uncertainty of one r.v. due to the knowledge of the other!**

$$\begin{aligned} I(X; Y) &= D(p(X, Y) \parallel p(X)p(Y)) \\ &= \sum_{x_i, y_i} p(x_i, y_i) \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} \\ &= H(X) - H(X | Y) \end{aligned}$$



Linkage Identification (9)



❖ Information Theoretic Identification [Example]

- At k-th generation, we obtain the following selected population.
- We employ **mutual info.** with $\Delta=0.1$

Selected Population

$x_1 \ x_2 \ x_3 \ x_4$

1	1	1	1
0	1	0	1
1	0	1	1
0	0	1	1
1	1	1	0
1	1	0	1
1	1	0	0
0	0	1	1
1	0	1	1
1	1	0	1

$$\begin{aligned} p(x_1=0) &= 3/10, p(x_1=1) = 7/10 \\ p(x_2=0) &= 4/10, p(x_2=1) = 6/10 \\ p(x_3=0) &= 4/10, p(x_3=1) = 6/10 \\ p(x_4=0) &= 2/10, p(x_4=1) = 8/10 \end{aligned}$$

$$\begin{aligned} p(x_1=0, x_2=0) &= 2/10, p(x_1=0, x_3=0) = 1/10 \\ p(x_1=0, x_2=1) &= 1/10, p(x_1=0, x_3=1) = 2/10 \\ p(x_1=1, x_2=0) &= 2/10, p(x_1=1, x_3=0) = 3/10 \\ p(x_1=1, x_2=1) &= 5/10, p(x_1=1, x_3=1) = 4/10 \end{aligned}$$

$$\begin{aligned} p(x_1=0, x_4=0) &= 0/10, p(x_2=0, x_3=0) = 0/10, p(x_2=0, x_4=0) = 0/10, p(x_3=0, x_4=0) = 1/10 \\ p(x_1=0, x_4=1) &= 3/10, p(x_2=0, x_3=1) = 4/10, p(x_2=0, x_4=1) = 4/10, p(x_3=0, x_4=1) = 3/10 \\ p(x_1=1, x_4=0) &= 2/10, p(x_2=1, x_3=0) = 4/10, p(x_2=1, x_4=0) = 2/10, p(x_3=1, x_4=0) = 1/10 \\ p(x_1=1, x_4=1) &= 5/10, p(x_2=1, x_3=1) = 2/10, p(x_2=1, x_4=1) = 4/10, p(x_3=1, x_4=1) = 5/10 \end{aligned}$$

$$I(x_1; x_2) = (2/10)\log(20/12) + (1/10)\log(10/18) + (2/10)\log(10/14) + (5/10)\log(50/42) = 0.0633$$

$$I(x_1; x_3) = (1/10)\log(10/12) + (2/10)\log(20/18) + (3/10)\log(30/28) + (4/10)\log(40/42) = 0.004$$

$$I(x_1; x_4) = (0/10)\log(-) + (3/10)\log(10/8) + (3/10)\log(10/7) + (5/10)\log(50/56) = 0.1173$$

$$I(x_2; x_3) = (0/10)\log(-) + (4/10)\log(10/6) + (4/10)\log(10/6) + (5/10)\log(20/36) = 0.1148$$

$$I(x_2; x_4) = (0/10)\log(-) + (4/10)\log(10/8) + (2/10)\log(10/6) + (4/10)\log(40/48) = 0.1185$$

$$I(x_3; x_4) = (1/10)\log(10/8) + (3/10)\log(30/32) + (1/10)\log(10/12) + (5/10)\log(50/48) = 0.0051$$

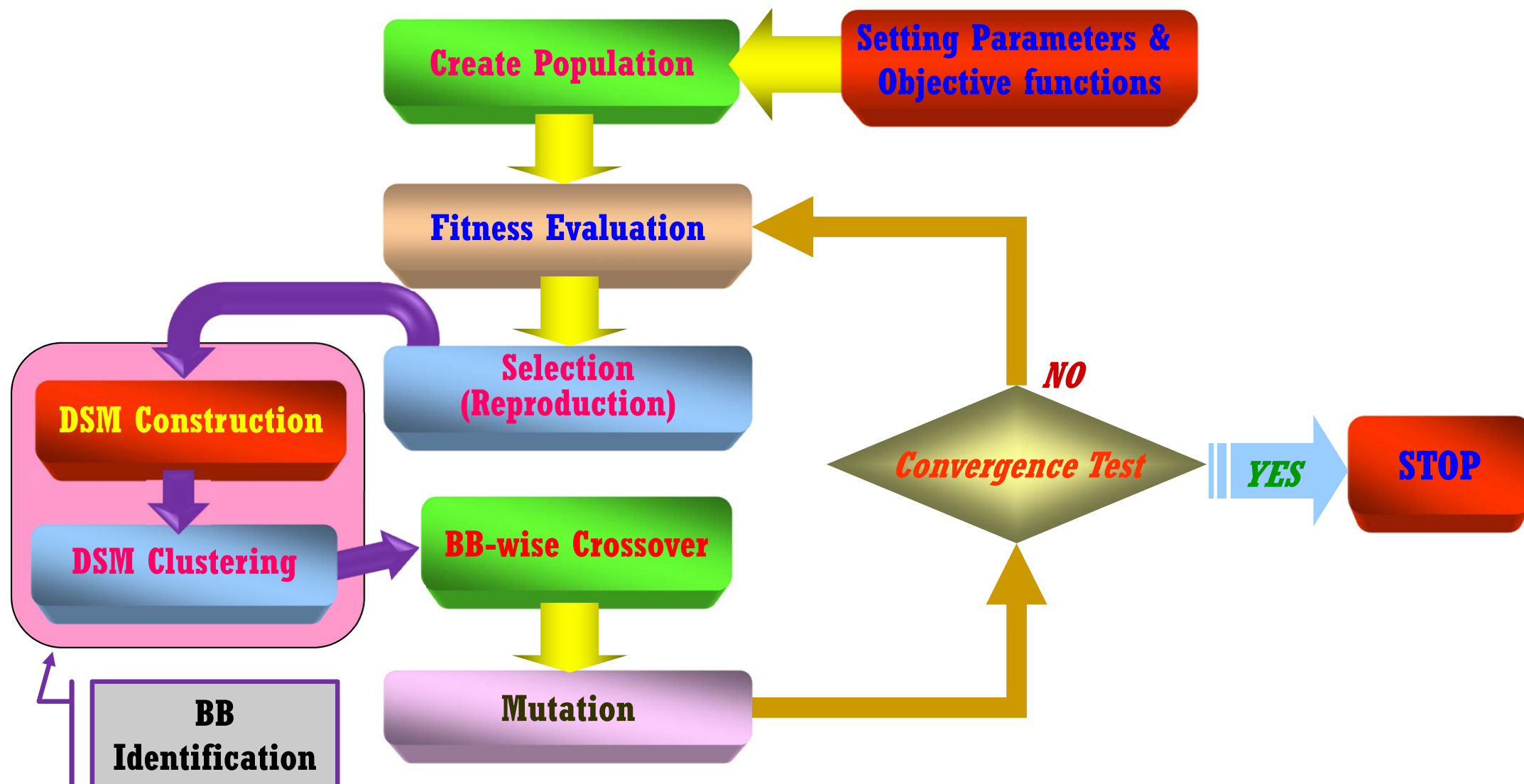
For pairs, $I(x_2; x_4)$ is the highest & it is greater than 0.1.

Thus, x_2 and x_4 are linked!
The similar procedure continues!



DSMGA (1)

❖ More General Approach: Dependency Structure Matrix GA (DSMGA)

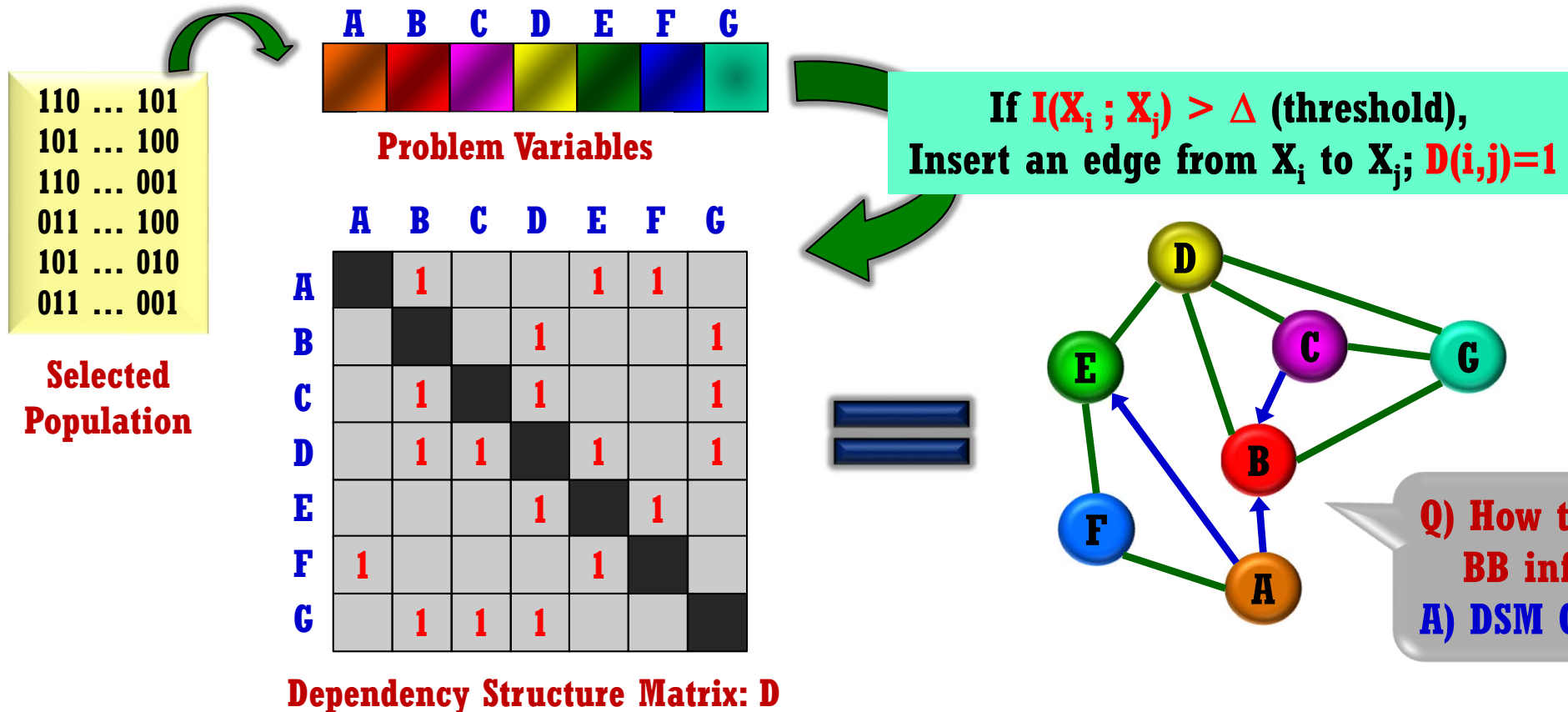




DSMGA (2)

❖ Key Idea of DSMGA

- Utilize a **matrix representation** of an **interaction graph** to decompose the problem
 - ✓ **Construct DSM** by using the **mutual information** from the selected population
 - ✓ After clustering the DSM, perform the **BB-wise uniform crossover**





DSMGA (3)

❖ DSM Clustering

➤ Find a **clustering arrangement** such that...

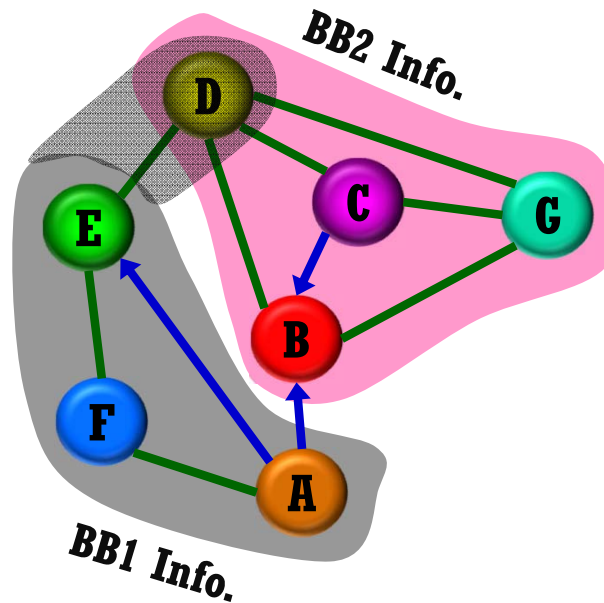
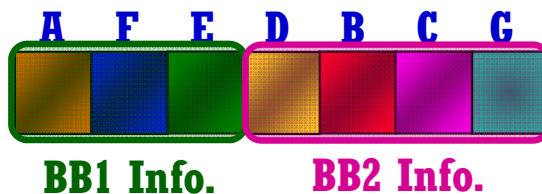
✓ **BB modules minimally interact** with each other while **components within a BB module maximally** interact with each other → **Optimization problem**

♣ This problem can be solved by a **Hill-climber** with **MDL metric**

The issue will be studied later!

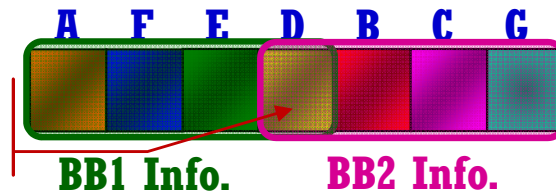
Clustered DSM

	A	F	E	D	B	C	G
A		1	1		1		
F	1		1				
E		1		1			
D			1		1	1	1
B				1		1	1
C				1	1		1
G				1	1	1	



Alternative Clustering

	A	F	E	D	B	C	G
A		1	1		1		
F	1		1				
E		1		1			
D			1		1	1	1
B				1		1	1
C				1	1		1
G				1	1	1	



Overlapped!



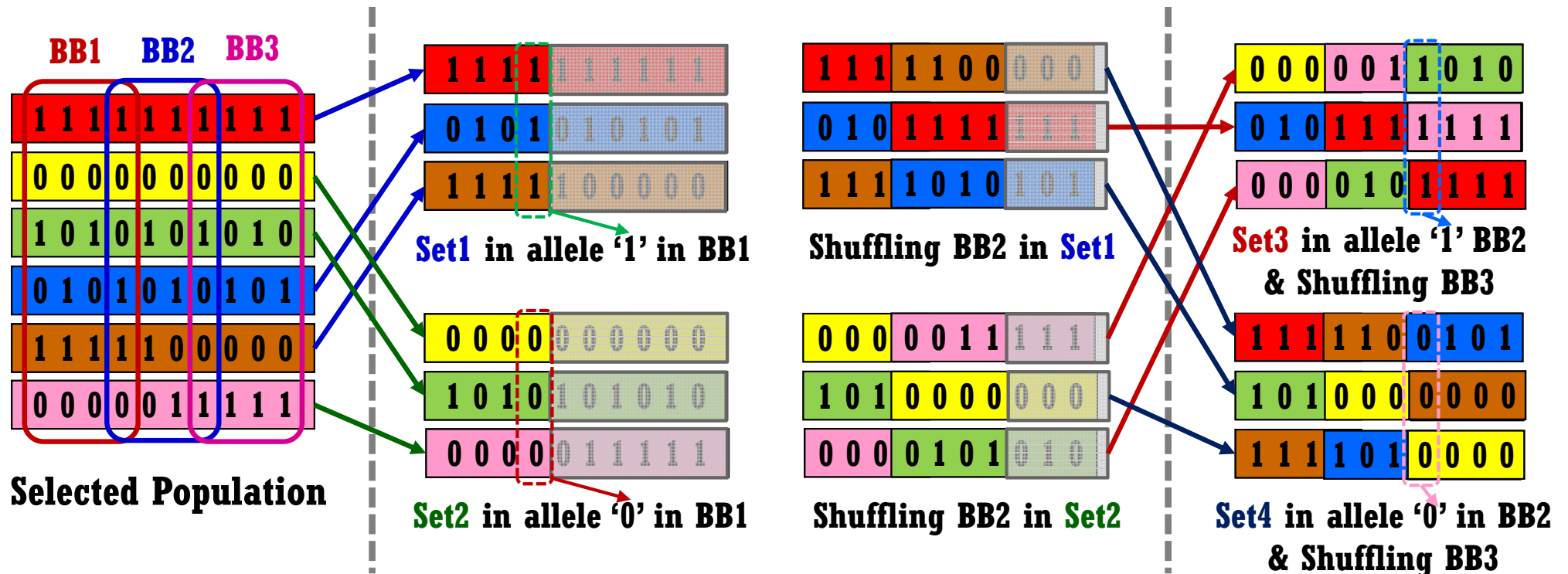
❖ BB-wise Uniform Crossover

➤ **Maximize the search capability without destroying BBs**

✓ **For non-overlapping BBs → Straightforward!**

✓ **As for overlapping BBs → Consider the alleles of the overlapped genes**

1. Partition the population by the alleles of the overlapped genes
2. Population-wise shuffling can be performed between the BBs within each set
3. Similar procedures are carried out for other overlapping BBs





Summary



- ❖ To Find the **Optimum**, **BBs** must be **Preserved** and **Fertilized**.
 - But, We **cannot distinguish** **BBs** themselves from **other partial-solutions**
 - Only the information we can know is the BB information; **Linkage of variables**
 - Crossover must be performed **at the level of BBs**; BB-wise uniform crossover
- ❖ **BBs (i.e., Linkage)** can be **Discovered** by Various Ways
 - Detection from fitness changes by Perturbations – Not recommendable!
 - Detection from **correlation coefficient** by Probability Dist. – **Recommended!**
 - Detection from **mutual information** by **Information Criteria** – **Recommended!**
 - ➔ This is **the most important**, but very difficult, issue in GA research.

Although we don't know exact BBs, they have to be assembled somehow.

It can be done by BB-wise uniform crossover.

To use it, BB-structure (i.e., linkage) must be discovered by proper techniques.

The information theoretic approaches are the most recommendable.

Discovering BB information is a still growing research topic!