Cell Formation Problem

1.1 Introduction

In manufacturing systems, GT is a manufacturing philosophy based on organizing and grouping common tasks to improve the productivity of the system [5]. One of the most important applications of GT is CM, which is the grouping of machines and parts so that each family of parts is processed within a machine cell [6]. Many benefits have been reported for CM, including reducing material handling costs, setup times, expediting costs, in-process inventories, part makespan, and improving human relations and operator expertise [5]. The core problem in designing a cellular manufacturing system (CMS) is the cell formation problem.

The CFP in CMSs is an important issue in the operational research literature [7, 8]. It consists of decomposing an entire production system into a set of manufacturing cells, assigning machines, and allocating parts to those cells. Some constraints and objectives must be taken into account to produce the most manageable and independent cells during this decomposition [3]. The cell formation problem is known to be a non-polynomial (NP)-hard problem [6]; therefore, the development of efficient machine grouping algorithms has always been the center of interest in CMS design, which has led to a wide range of research [9].

This chapter is organized as follows: In section 1.2, we define the cell formation problem. In section 1.3, we present related work. In section 1.4, we give a formulation of the Generalized Cubic Cell Formation Problem and its mathematical model. Finally, in section 1.5, we conclude.

1.2 Definition of the Cell Formation Problem

1.2.1 Basic Cell Formation Problem

In the basic cell formation problem, the only provided information is the incidence matrix of parts and machines. The incidence matrix is a binary matrix, where machines and parts are represented in rows and columns; each cell in this matrix may contain a 0 or 1 as a value. 0 means that the part in the column does not

	P1	P2	P3	P4	P5	p6	P7
M1	1	1	0	1	0	0	0
M2	0	1	1	1	1	1	1
M3	0	0	1	0	1	0	1
M4	0	0	1	1	0	1	0
M5	1	1	0	0	0	0	0
M6	1	1	0	0	1	0	0

Figure 1.1: Incidence matrix

	P3	P4	P6	P5	P7	P1	P2
M2	1	1	1	1	1	0	1
M4	1	1	1	0	0	0	0
M3	1	0	0	1	1	0	0
M1	0	1	0	0	0	1	1
M5	0	0	0	0	0	1	1
M6	0	0	0	0	1	1	1

Figure 1.2: The resulting groupement

need the machine in the row. Otherwise, the cell is filled with the value 1. Figure 1.1 shows an example of an incidence matrix. The cell formation problem's output is a configuration that specifies the nature of cells must be built, the cell to which each machine is assigned, and the cell to which each part is affected (see Figure 1.2).

1.2.2 Generalized Cubic Cell Formation Problem

In this work, we consider a variant of the CFP, known as the GCCFP. The basic CFP considers that each part has a single route. However, in real situations, a part may have more than one process routing (e.g., part p_i maybe processed on machines m_1 and m_3 or it may be processed on m_1, m_2, and m_4). The CFP that considers many potential process routings is called Generalized Cell Formation Problem (GCFP) [10]. Although cells' formation by considering the parts and the machines is the essence of group technology, its full advantages cannot be achieved without including the human factor [11]. Because of the workers' essential role, grouping workers with similar expertise and skills to produce similar families of parts can improve the CMS design quality. When the worker's dimension is considered in addition to the part and the machine dimensions, the problem is transformed into a GCCFP [3].

1.3 Related Work

In the literature, a wide variety of CFPs have been described, and many techniques and algorithms have been proposed to solve them, including heuristics, meta-heuristics, exact methods, etc. [8, 7]. The use of exact methods to solve CFPs provides the best existing solutions. However, due to these combinatorial problems' NP-hard nature, as the problem's dimensions increase, these exact methods become incredibly costly in time and memory consumption. For that reason, meta-heuristic techniques are considered more convenient to solve

NP-hard problems and to produce acceptable solutions in a reasonable time. The following review of related work is established in [3].

1.3.1 Basic Cell Formation Problem

Adil et al. [12] have developed a new non-linear mathematical programming model for CFP. The model's objective is to minimize the sum of voids and exceptional elements within and between cells. To simultaneously identify families of parts and groups of machines, the authors developed an Assignment Allocation Algorithm (AAA) and a Simulated Annealing (SA) algorithm. The authors modified voids' weights and exceptional elements to allow multiple configurations, thus providing designers with multiple possible solutions. Tavakkoli-Moghaddam et al. [13] used the SA algorithm to solve the CFP. The authors considered two types of cells: common or general cells and specific cells. The difference between these two types is that common cells can produce different products (parts). However, only one type of product can be processed by a specific cell. Neto and Gonçalves Filho [14] suggested a multi-objective approach. The objective of this approach is to build cells to minimize simultaneously three contradictory objectives, namely (i) the level of the work-in-process, (ii) the inter-cell moves, and (iii) the total machinery investment. The authors used the Genetic Algorithm (GA) to solve the CFP. They adopted the Pareto optimality principle in the solution procedure to cope with the conflicting objectives. Shiyas and Pillai [15] proposed a new mathematical model for designing manufacturing cells. The model considers two contradictory objectives, such as the inter-cell moves and the cells' heterogeneity. To solve the model, the authors developed a GA-based method. In order to propose alternative cell configurations to decision-makers, a weighting parameter is assigned to the heterogeneity of the objective function of the model. Hafezalkotob et al. [16] proposed a hybrid algorithm to solve the CFP. The algorithm is a combination of Discrete Particle Swarm Optimization (DPSO) and SA. The purpose of coupling these two algorithms is to ensure rapid convergence by DPSO and bring out the search from the local optimum by SA. Danilovic and Ilic [17] has developed a new hybrid algorithm called Cell Formation OPTimization (CFOPT) to solve the CFP. The algorithm's strategy is to use the specificity of the input instances to reduce the set of possible solutions to increase the optimization process's efficiency. Mahmoodian et al. [18] presented a new algorithm based on Particle Swarm Optimization (PSO). The algorithm integrates the self-organization map neural networks to the PSA algorithm. Karoum and Elbenani [19] combined a local search mechanism with the cuckoo search algorithm to intensify the search and improve solutions' grouping efficacy.

1.3.2 Generalized Cell Formation Problem

The GA is used in many works [20, 21, 22, 23]. The SA algorithm is also widely used to solve the problem [21, 24, 25]. To solve the GCFP, Vin et al. [20] proposed a solution entitled Multi-Objective Grouping Genetic Algorithm (MOGGA), which is the combination of GA with an integrated heuristic. The authors used the GA to solve the routing selection problem (to select the most appropriate parts' processing plan). However, the integrated heuristic is combined to address the CFP simultaneously. Ameli et al. [26] used a mathematical programming method named Branch and Bound (B&B) to solve the GCFP. This method, like other exact methods, is not capable of effectively solving large-scale problems. Wu et al. [21] considered a GCFP model that takes as input a binary incidence matrix. This matrix indicates each part's process plans, where each plan mentions the different machines required by the concerned part without establishing an order between them. To solve the model, the authors combined GA with the SA algorithm. In contrast, in the GCFP model solved by Chung et al. [27], the authors consider the sequences of operations, the alternative routing of the

parts, and the reliability of machines (machine failure). Taking machine failures into account during the design of CMS contributes in improving the system's overall performance. The authors combined two techniques to solve the model: Tabu Search (TS) and GA's mutation operator. The use of this operator is justified by its ability to escape local solutions and prevent premature convergence. Jouzdani et al. [25] have extended the model presented in [26] to consider set-up costs. The authors have applied a meta-heuristic method, which is the SA algorithm, to solve the GCFP. Karoum and Elbenani [28] proposed a method entitled Hybrid Selection Algorithm-Generalized Cell Formation (HCSA-GCF). The method aims to reduce the costs of intra-cellular part movements and machine breakdowns. The authors compared the obtained results with those provided by B&B under LINGO software. Hazarika and Laha [23] used a GA heuristic to solve the GCFP with multiple process routes, operation sequences, and parts volume. Five benchmark problems have been used to show the performance of the method.

1.3.3 Cubic Cell Formation Problem

Before 1993, all studies were on two-dimensional manufacturing CFP. Despite the importance of the human dimension, most studies only considered the dimensions of parts and machines. The central issue has been the grouping of similar parts into part families and machines into machine cells [3]. The cubic CFP, which includes the worker (operator) as a third dimension, was first introduced by MIN and SHIN [11]. The authors considered that the group technology cell workers must be multi-disciplined and highly-skilled to work on different machines and perform various tasks. Li [29] presented a new algorithm to solve cubic CFP. This algorithm's added value is to organize all incidence matrices of the problem that links parts to machines, machines to workers, and workers to parts into a single symmetrical incidence matrix. Mahdavi et al. [30] introduce an integer mathematical programming model for the cellular manufacturing system design in a dynamic environment. In the model, the authors took into account multi-period production planning and dynamic reconfiguration of the system. Nikoofarid and Aalaei [31] presented a new mathematical model for a CFP in production planning in a dynamic virtual cellular manufacturing system. The proposed model includes the worker dimension and considers as objectives the minimization of the holding and backorder costs and the management of machines and workers over a specific planning horizon. Mahdavi et al. [32] used the B&B method of the LINGO software package to solve the CFP model. In the model, the authors considered two objectives: the minimization of voids and exceptional elements. The mathematical model catches workers' skills in performing different tasks. Aalaei and Shavazipour [33] defined an integer mathematical programming model for designing the cellular manufacturing systems under data envelopment analysis. The authors tried to minimize the costs of backorders and inter-cellular movement costs produced by exceptional elements. Bootaki et al. [34] proposed a new multi-objective mathematical model for cubic binary CFP. In the model, the authors introduced a new objective called "Quality Index". This objective measures the quality of the parts produced. To calculate the value of this index, data measuring different workers' skills in producing particular parts on special machines must be performed. The authors developed a hybrid genetic algorithm, AUGMented ε-CONstraint (GA-AUGMECON) method to solve the model. Bootaki et al. [35] developed a new multi-objective mathematical model to design dynamic cubic binary CFP. In the model, the authors consider the machine and the concept of worker utilization. To solve the model, the authors have developed a new goal programming method called Percentage Multi-Choice Goal Programming (PMCGP). Motivated by the inefficiency of exact methods to solve large-sized test problems, Sahin and Alpay [36] proposed a GA to solve cubic binary CFP. Taguchi's method was used as a statistical optimization technique to define the parameters' level. Feng et al. [37] consider that the human factor is essential for successfully implementing cellular manufacturing systems. In the proposed model, in opposite to [34] and [35], the author considered operation sequences and alternative process routings. The author also included in

the model the simultaneous consideration of production scheduling, lot splitting, workload balancing between cells, and worker over-assignment to multiple cells. A hybrid approach combining Combinatorial Particle Swarm Optimization and Linear Programming (CPSO-LP) has been proposed to solve the model's real-sized problems efficiently. Bagheri et al. [38] presented a multi-period CFP in a dynamic environment to maximize the total value of grouping efficacy and minimize the total costs and total non-interest workers in cells. The principal idea is to improve the cells' efficiency by assigning workers who have a mutual interest in working with each other. The authors did not consider sequences of operations. A Revised Multi-Choice Goal Programming (RMCGP) method was used to solve the proposed multi-objective mathematical model.

1.4 Generalized Cubic Cell Formation Problem Formulation

We adopt the assumptions and the formulation provided in [3].

1.4.1 Assumptions

The GCCFP is studied according to the following hypotheses :

- The number of cells, the upper, and the lower number of machines in each cell are known and considered as parameters.
- The quality of treating each part on each machine by each worker is specified using a three-dimensional matrix. This matrix values are integers between 1 and 5 (representing very bad, bad, medium, well, very well). The value 0 indicates that a given part cannot be processed on a given machine by a given worker. These values can be estimated by analyzing the historical acquired data and worker errors. During the initial designing of the production system layout, the same quality level can be assigned to each worker. It is also possible to estimate the level of quality by analyzing his qualifications and experience. After that, a continuous evaluation can be planned to acquire the necessary data on the workers' ability to produce parts and handle machines. These data may help later for a future update of the system configuration.
- The machines' and workers' capacity is not considered.
- Each part type has at least one processing route. Exactly one route will be set up to produce this part.
- There are a single machine and a single worker of each type.
- Parts may move within and between cells. The inter-cellular movement is produced if two consecutive operations are executed on the selected route of a given part in two different cells. However, the intracellular movement is incurred when two consecutive operations of a part are performed in the same processing cell.
- The material Handling cost of a given design is the sum of the intercellular and the intracellular movement of the parts.
- The inter-cellular movement of workers is calculated according to the availability or the absence of the workers in the processing cells.
- A worker can work on several machines.

- A part may be processed by multiple workers, but an operation of a part is assigned to a single worker, and it is performed on a single machine.

1.4.2 The constants

To formulate GCCFP, these notations are used:

C	the total number of cells.
T	the set of cells, T = {1,...,C}.
M	the total number of machines.
P	the total number of parts.
W	the total number of workers.
R_p	the total number of process routes of part p.
Op_{pr}	the total number of the operations in route r of part p.
k	the index of cells, k= 1,2, ...,C.
p	the index of parts, p=1,2, ...,P.
m	the index of machines, m=1,2, ...,M.
w	the index of workers, w=1,2, ...,W.
r	the index of process routes.
s	the index of operations within routes.
UM	the maximum cell size.
LM	the minimum cell size.
CO_p	the cost of moving part p to an outer cell.
CI_p	the cost of moving part p inside the same cell.
CW_w	the cost of moving worker w from a cell to another one.
ap_{rsm}	a binary parameter indicating whether operation s in route r of part p may be processed on machine m.
b_{mw}	a binary parameter indicating whether worker w can use machine m.
c_{wp}	a binary parameter indicating whether worker w can process part p.
q_{pmw}	quality obtained for part i when it is processed on machine m by worker w.

The GCCFP resolution consists of four decisions to be taken :

1. The selection of a single route for each part.
2. The assignment of each machine to a single cell.
3. The allocation of each worker to a single cell.
4. The specification of which worker will perform a given operation of a given part, on which machine, and within which cell.

1.4.3 The decision variables

$$\mathrm{R}_{pr} = \begin{cases} 1 & if\ part\ p\ is\ processed\ according\ to\ process\ route\ r \\ 0 & otherwise \end{cases}$$

$$\mathrm{Y}_{mk} = \begin{cases} 1 & if\ machine\ m\ is\ assigned\ to\ cell\ k \\ 0 & otherwise \end{cases}$$

$$\mathrm{Z}_{wk} = \begin{cases} 1 & if\ worker\ w\ is\ assigned\ to\ cell\ k \\ 0 & otherwise \end{cases}$$

$$\mathrm{X}_{prsmwk} = \begin{cases} 1 & if\ operation\ s\ of\ part\ p\ along\ route\ r\ is\ processed \\ & on\ machine\ m\ by\ worker\ w\ in\ cell\ k \\ 0 & otherwise \end{cases}$$

1.4.4 The mathematical model

$$\begin{cases} min & InterCMHC\ +\ IntraCMHC\ +\ InterCWM \\ max & Quality \end{cases} \tag{1.1}$$

$$InterCMHC = \sum_{k \epsilon T} \sum_{k' \epsilon T \setminus \{k\}} \sum_{p=1}^{P} CO_p * \left[\sum_{r=1}^{R_p} \sum_{s=1}^{Op_{pr}-1} \left[(\sum_{m=1}^{M} \sum_{w=1}^{W} X_{prsmwk}) * (\sum_{m=1}^{M} \sum_{w=1}^{W} X_{prs+1mwk'}) \right] \right] \tag{1.2}$$

$$IntraCMHC = \sum_{k=1}^{C}\sum_{p=1}^{P} CI_p * \left[\sum_{r=1}^{R_p}\sum_{s=1}^{Oppr-1}\left[(\sum_{m=1}^{M}\sum_{w=1}^{W} X_{prsmwk}) * (\sum_{m=1}^{M}\sum_{w=1}^{W} X_{prs+1mwk})\right]\right] \tag{1.3}$$

$$InterCWM = \sum_{k=1}^{C}\sum_{p=1}^{P}\sum_{r=1}^{R_p}\sum_{s=1}^{Op_{pr}}\sum_{m=1}^{M}\sum_{w=1}^{W} CW_w * X_{prsmwk} * (1 - Z_{wk}) \tag{1.4}$$

$$Quality = \sum_{k=1}^{C}\sum_{p=1}^{P}\sum_{r=1}^{R_p}\sum_{s=1}^{Op_{pr}}\sum_{m=1}^{M}\sum_{w=1}^{W} q_{pmw} * X_{prsmwk} \tag{1.5}$$

Subject to:

$$X_{prsmwk} \leq R_{pr} * a_{prsm} * Y_{mk} * b_{mw} * c_{wp} \qquad \forall(p, r, s, m, w, k) \tag{1.6}$$

$$\sum_{k=1}^{C}\sum_{m=1}^{M}\sum_{w=1}^{W} X_{prsmwk} = R_{pr} \qquad \forall(p, r, s) \tag{1.7}$$

$$\sum_{r=1}^{R_p} R_{pr} = 1, \qquad \forall p \tag{1.8}$$

$$\sum_{k=1}^{C} Z_{wk} = 1, \qquad \forall w \tag{1.9}$$

$$\sum_{k=1}^{C} Y_{mk} = 1, \qquad \forall m \tag{1.10}$$

$$\sum_{m=1}^{M} Y_{mk} \leq UM, \qquad \forall k \tag{1.11}$$

$$\sum_{m=1}^{M} Y_{mk} \geq LM, \qquad \forall k \tag{1.12}$$

$$X_{prsmwk}, Y_{mk}, Z_{wk}, R_{pr} \in \{0, 1\} \qquad \forall(p, r, s, m, w, k) \tag{1.13}$$

The objective function of the model is given in equation 1.1. It minimizes the material handling cost and the inter-cellular movement of workers and maximizes the produced parts' quality index. The formulas of calculating the inter-cellular material handling cost (InterCMHC), the intra-cellular material handling cost (IntraCMHC), the inter-cellular worker movement (InterCWM), and the quality index are respectively given

in equations 1.2, 1.3, 1.4, and 1.5. The purpose of the model is to find a better compromise between these objectives.

Equation 1.6 imposes that an operation s will be executed on machine m by worker w within cell k only if:

- The route, to which s belongs, is set up to produce the concerned part p (R_{pr}).
- Machine m is required to execute the operation s (a_{prsm}).
- Machine m is already assigned to cell k because machines cannot be moved between cells (Y_{mk}).
- Worker w can use machine m (b_{mw}).
- Worker w can process part p (c_{wp}).

Equation 1.7 guarantees that an operation s is performed at most on a single machine by a single worker in a single cell and will only be executed if the route to which s belongs is set up to produce the part concerned. Equation 1.8 means that only one route is established for each part. Constraint 1.9 confirms that each worker is assigned precisely to one cell. Constraint 1.10 ensures that each machine is assigned to one and exactly one cell. Constraints 1.11 and 1.12 present the minimum and the maximum number of machines that a cell can contain. The last constraint represents logical binary requirements on the decision variables.

1.4.5 Linearisation of the model

The non-linearisation in the proposed model is caused by the first three terms of the objective function (InterCMHC, IntraCMHC, InterCWM), and the constraint 1.6 of the model. The following linearisation is performed by [3]. To linearize the model, four auxiliary binary variables are used:

- $F_{prsmm'ww'kk'}$ = X_{prsmwk} * $X_{prs+1m'w'k'}$ $\quad k \neq k'$, s $\leq$ Op_{pr}-1
- $I_{prsmm'ww'k}$ = X_{prsmwk} * $X_{prs+1m'w'k}$ $\quad s \leq Op_{pr}$-1
- J_{prsmwk} = X_{prsmwk}*(1-Z_{wk})
- L_{prmk} = R_{pr} * Y_{mk}

Thus, the first three terms of the objective function are computed like that:

$$InterCMHC = \sum_{k \epsilon T} \sum_{k' \epsilon T \setminus \{k\}} \sum_{p=1}^{P} CO_p * \sum_{r=1}^{R_p} \sum_{s=1}^{Op_{pr}-1} \sum_{m=1}^{M} \sum_{m'=1}^{M} \sum_{w=1}^{W} \sum_{w'=1}^{W} F_{prsmm'ww'kk'} \tag{1.14}$$

$$IntraCMHC = \sum_{k=1}^{C}\sum_{p=1}^{P} CI_p \sum_{r=1}^{R_p}\sum_{s=1}^{Op_{pr}-1}\sum_{m=1}^{M}\sum_{m'=1}^{M}\sum_{w=1}^{W}\sum_{w'=1}^{W} I_{prsmm'ww'k} \tag{1.15}$$

$$InterCWM = \sum_{k=1}^{C}\sum_{p=1}^{P}\sum_{r=1}^{R_p}\sum_{s=1}^{Op_{pr}}\sum_{m=1}^{M}\sum_{w=1}^{W} CW_w * J_{prsmwk} \tag{1.16}$$

The constraint 1.6 is replaced by these three constraints:

$$X_{prsmwk} \leq a_{prsm} * b_{mw} * c_{wp} * L_{prmk} \qquad \forall(p,r,s,m,w,k) \tag{1.17}$$

$$2 * L_{prmk} \leq R_{pr} + Y_{mk} \qquad \forall(p,r,m,k) \tag{1.18}$$

$$L_{prmk} + 1 \geq R_{pr} + Y_{mk} \qquad \forall(p,r,m,k) \tag{1.19}$$

The following additional constraints are used to restrict the introduced variables ($F_{prsmm'ww'kk'}$, $I_{prsmm'ww'k}$, J_{prsmwk}):

$$2 * F \leq X_{prsmwk} + X_{prs+1m'w'k'} \qquad \forall(p,r,s,m,m^{'},w,w^{'},k,k^{'}), k \neq k^{'} \tag{1.20}$$

$$F + 1 \geq X_{prsmwk} + X_{prs+1m'w'k'} \qquad \forall(p,r,s,m,m^{'},w,w^{'},k,k^{'}), k \neq k^{'} \tag{1.21}$$

$$2 * I \leq X_{prsmwk} + X_{prs+1m'w'k} \qquad \forall(p,r,s,m,m^{'},w,w^{'},k) \tag{1.22}$$

$$I + 1 \geq X_{prsmwk} + X_{prs+1m'w'k} \qquad \forall(p,r,s,m,m^{'},w,w^{'},k) \tag{1.23}$$

$$2 * J \leq X_{prsmwk} + 1 - Z_{wk} \qquad \forall(p,r,s,m,w,k) \tag{1.24}$$

$$J \geq X_{prsmwk} - Z_{wk} \qquad \forall(p,r,s,m,w,k) \tag{1.25}$$

1.5 Conclusion

In this chapter, an overview of the Cell Formation Problem is presented. Initially, we have defined its basic version. After, we have presented a definition of the version that we will consider in our study, which is the Generalized Cubic Cell Formation Problem. Next, a study of the related work is provided. Finally, a mathematical formulation of the Generalized Cubic Cell Formation Problem is given.

Our problem belongs to the NP-hard class. The problems of this class are algorithmically solvable but computationally intractable. There is no exact method that can find the optimal global solutions to NP-hard problems in polynomial time. Fast approximate heuristics and meta-heuristics are the popular approaches to search for practical solutions. In our study, we will use the genetic algorithm, which is one of the most popular meta-heuristics, often used to solve complex large-scale optimization problems. So in the next chapter, we will give an overview of the genetic algorithm.

Chapter 2

Genetic Algorithms

2.1 Introduction

In many real-life settings, high-quality solutions to hard optimization problems are required in a short amount of time. Due to the practical importance of the combinatorial optimization problems for industry and science, many algorithms to tackle them have been developed [39]. In combinatorial optimization (CO), algorithms can be classified as either exact or approximate algorithms. In approximate methods such as metaheuristics, we sacrifice the guarantee of finding optimal solutions for the sake of getting good solutions in a significantly reduced amount of time. Thus, the use of metaheuristics has received more and more attention in the last decades.

The term metaheuristic was first introduced in [40]. A metaheuristic is an iterative generation process that guides a subordinate heuristic by combining intelligently different concepts to explore and exploit the search space to find efficiently near-optimal solutions [41]. Metaheuristics may be classified into methods that perform a single solution vs. population-based search. This classification refers to the number of solutions used by a metaheuristic at any time. Generally, algorithms that work on a single solution at any time are referred to as trajectory methods. They all share the property that the search process describes a trajectory in the search space (e.g., tabu search, iterated local search, and simulated annealing). Population-based metaheuristics deal at each algorithm iteration with a set of solutions rather than with a single one. From this set of solutions, the next iteration population is produced by the application of some operators. Population-based metaheuristics provide a natural, intrinsic way for the exploration of the search space. However, the final performance strongly depends on the way the population is manipulated. The most studied population-based methods are evolutionary computation (EC) and ant colony optimization (ACO) [39].

EC can be regarded as a metaphor for building, applying, and studying algorithms based on Darwinian natural selection principles. The instances of algorithms based on evolutionary principles are called Evolutionary Algorithms (EA) [42]. EAs can be characterized as computational models of evolutionary processes. There has been a variety of slightly different EAs proposed over the years. In our work, we will use an evolutionary algorithm to solve the Generalized Cubic Cell Formation Problem. This evolutionary algorithm is the Genetic Algorithm.

In this chapter, we briefly introduce the genetic algorithms. In section 2.2, we give some definitions and terminology. In section 2.3, we exhibit the basic genetic algorithms. In section 2.4, we discuss the genetic algorithm operators. Finally, in section 2.5, we conclude.

2.2 Definitions and Terminology

GA are stochastic search methods that combine two main search strategies: exploiting better solutions and exploring the global search space. These algorithms are based on the principles of natural selection proposed by Darwin and natural genetics.

GA was initially introduced by John Holland, his colleagues, and his students at the University of Michigan [4]. Their research goals have been twofold: (i) to abstract and rigorously explain the adaptive processes of natural systems, and (ii) to design artificial systems software that retains the important mechanisms of natural systems. This approach has led to important discoveries in both natural and artificial systems science. Goldberg presented the fundamentals of GAs and described its usual form [43].

GAs have been successfully applied to many optimization problems in different disciplines that are difficult to solve by classical mathematical programming [14, 15, 20, 22, 23, 34, 36, 44, 45, 46, 47, 48, 49]. In the following sections, some important terminology and concepts of GA are presented.

2.2.1 Genes and Chromosomes

The gene is the basic component of the GA. A string of genes is called a chromosome. Chromosomes can be encoded as binary strings, as strings of real numbers, etc.

2.2.2 Populations and Generations

A population is a set of chromosomes. GA begins with a set of randomly created individuals (chromosomes). This set is called the **initial population**. The iterations of GA are called generations. Each iteration involves selecting individuals with closely related characteristics and recombining them until a new generation is created to replace the old one [50].

2.2.3 Parents and Children

The selection of chromosomes from one generation to another consists of choosing individuals in a probabilistic method [50]. Those with high fitness values have a high probability of being selected to undergo crossover and produce new chromosomes called children or offsprings. The crossover happens with a priori fixed probability called crossover rate. It includes a random selection of the parent chromosomes' crossover points, where the mixing of parent's genetic information should be happening.

2.2.4 Mutation

The mutation is a process by which many new points are introduced into the search space. It ensures that aggressive selection does not result in a suboptimal solution. In other words, it prevents premature convergence

to a local optimum. It is achieved by randomly changing some chromosome characteristics and is carried out at very low probability values (mutation rate).

2.2.5 Fitness

The objective function that defines the optimization purpose is called the fitness function. It indicates "goodness" or "badness" for each individual.

2.2.6 Elitism

To improve GA's performance, the best individuals must always participate in reproduction. However, such individuals can be lost if they are destroyed by crossover or mutation operators. Thus, the first best chromosome or the few best chromosomes are copied into the new population [51].

2.3 A Basic Genetic Algorithm

In general, a genetic algorithm must be able to achieve six basic tasks [52] :

1. Encoding the solution elements in the form of genes.
2. Create a string of genes to form a chromosome.
3. Initialize a starting population by generating a set of specific chromosomes, usually randomly.
4. Evaluate and assign fitness values to individuals in the population.
5. Perform reproduction by the fitness weighted selection of individuals of the population.
6. Perform recombination and mutation to produce individuals of the following generation.

A GA, then, is an iterative optimization method that simulates the adaptation and evolution of a single kind of organism. Using a chromosomal mapping system, the GA starts with a large number of possible design configurations. The range of potential configurations is defined by the limitations of the problem and the method of encoding all configuration information into the chromosome [50, 53].

A typical GA is represented in Figure 2.1

To start the optimization, the GA selects a set of configurations, almost always at random. This set is called the initial population, just as in biology. The GA evaluates the performance of each individual of the population using a cost function that compares the individual's performance to the desired or ideal performance and returns to the GA a single number that is a measure of its fitness. As in the evolutionary process of "survival of the fittest", high-quality strings combine and produce offspring, while low-quality strings are removed from the population[52]. Offspring can be generated by many different methods, each of which is essentially a method of combining information from two or more parent chromosomes to form a child with the potential to surpass

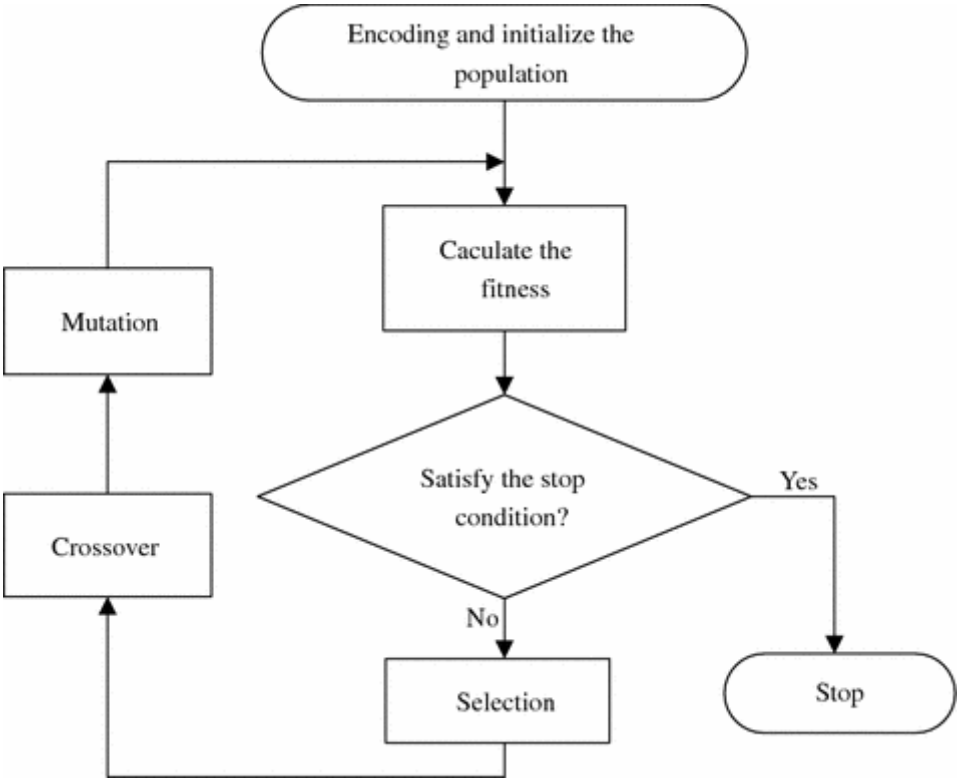


Figure 2.1: The basic process of genetic algorithm

its parents. With succeeding generations, the individuals' quality is continuously improved, and an optimized solution is finally reached. "Champions" will have many offsprings, while those who do not perform well will die without offspring. In this way, after some generations, a good solution is usually achieved [54].

2.4 GA Operators

The tasks that a genetic algorithm must complete and that were outlined in the previous section guide to the presence of three phases in the genetic algorithm optimization [50].

- Initiation;
- Reproduction;
- Generation replacement.

2.4.1 Initiation

Initiation means filling the initial population with encoded parameter strings or chromosomes, usually generated randomly. The coding is a mapping from parameter space to chromosome space [54].

2.4.1.1 Encoding

Encoding is a process of representing individual genes. The process can be performed using bits, numbers, trees, arrays, lists, or other objects. The encoding depends mainly on solving the problem [55].

An encoding function is used to represent the object variables' mapping to a string code. The mapping of string code to its object variable is achieved through the decoding function, as shown in Figure 2.2 [56].

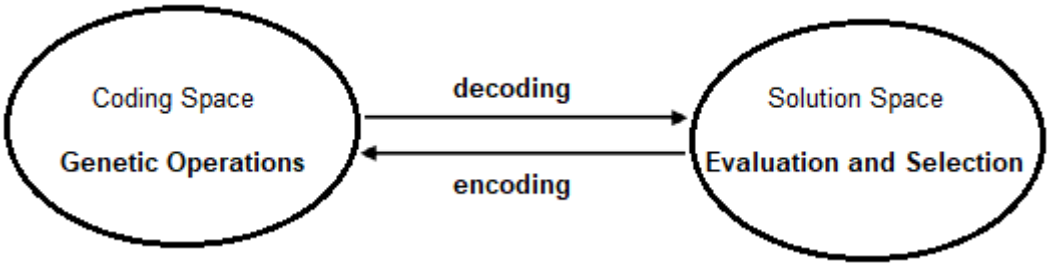


Figure 2.2: Encoding – Decoding method

1. **Binary Encoding**

   The most common way of encoding is a binary string, which would be represented as in Figure 2.3.

   Each chromosome is encoded in the form of a binary string. Every bit in the string may represent some characteristics of the solution. Each string, therefore, is a solution but not necessarily the best solution. Another possibility is that the entire string may represent a number.

   Binary encoding gives many potential chromosomes with a smaller number of alleles [55].

Chromosome 1	1	1	0	1	0	0	0	1
Chromosome 2	0	1	1	1	1	1	1	1

Figure 2.3: Binary encoding

2. **Hexadecimal Encoding**

   This encoding uses a string composed of hexadecimal numbers (0–9, A–F).

Chromosome 1	9	C	E	7	0	A	2	F
Chromosome 2	3	D	B	5	1	4	D	6

Figure 2.4: Hexadecimal encoding

3. **Real Number Encoding**

   The real number encoding is usually used for ordering issues. In this type of encoding, each chromosome represents a sequence of reals; for example, in the traveling salesman problem, the string of numbers represents the sequence of cities visited by the salesman [56]. Figure 2.5 shows an example of the real number encoding.

Chromosome 1	1	5	3	2	6	4	7	9
Chromosome 2	8	5	6	7	2	3	1	4

Figure 2.5: Real number encoding

There are other sorts of encoding, such as octal encoding, value encoding, tree encoding, etc. For more information about these types, an illustrated is provided in [55].

2.4.1.2 Fitness Function

A major problem in optimization is the formulation or the choice of an appropriate fitness function that determines the selection criterion in particular problems. For minimizing a function using genetic algorithms, a simple way to create a fitness function is to use the simplest form $F = A - y$, where A is a large constant ($A = 0$ is sufficient if the fitness is not required to be non-negative) and $y = f(x)$. The objective is to maximize the fitness function and then minimize the objective function $f(x)$. Alternatively, for a minimization problem, we can define a fitness function $F = 1/f(x)$, but it can have a singularity when $f \to 0$. There are many different ways to define a fitness function [57].

The appropriate form of the fitness function will ensure that solutions with higher fitness are selected efficiently. A poor fitness function may result in wrong or meaningless solutions.

2.4.2 Reproduction

It consists of three principal operators: selection, crossover, and mutation. These operators are discussed in the following.

2.4.2.1 Selection Strategies

The selection consists simply of choosing the best individuals to crossover. It aims to take advantage of these individuals' good characteristics by considering their fitness values, which is a measure of "goodness". In theory, there are many selection strategies; however, the most commonly used schemes are described in what follows [54].

1. **Population Decimation**

   This scheme relates to the so-called deterministic strategies [52]. The idea behind this method is simply the survival of the fittest with the elimination of the weakest fit. Since the population is decimated before being replaced by reproduction, this method is called population decimation. An arbitrary minimum fitness is chosen as the cutoff point, and any individual with a lower fitness is eliminated from the population. As an example of population decimation, consider the individuals in Figure 2.6 who are ranked, then population decimation consists of keeping the 50% best individuals. Thus, the results in the lower table of Figure 2.6 are obtained. Notice that the individuals whose values are below the threshold value are all rejected, which is a disadvantage because these individuals may possess some good characteristics that could have been obtained by crossover and/or mutation processes in the following generations.

   The advantage of population decimation is its simplicity. However, it has the disadvantage that: once an individual is eliminated from the population, any unique characteristics that this individual holds are lost [50]. For this reason, stochastic selection techniques have been developed.

2. **Proportionate Selection**

   This method is also known as the roulette wheel. Its philosophy is that individuals are selected based on a selection probability given by equation 2.1 [50, 52, 53].

$$P_{selection} = \frac{f(parent_i)}{\sum_i f(parent_i)} \tag{2.1}$$

   Where :

   - $P_{selection}$ is the probability of an individual parent being selected.
   - $f(parent_i)$ is the fitness value of $parent_i$.

   Initially, individuals are sorted according to their fitness. Then, the probabilities of the different individuals are calculated using equation 2.1. These probabilities are classified into a vector that contains the cumulative sums of these probabilities. A random number (between 0 and 1) is "thrown" (as a die roll), and depending on its value, will choose the individual who will participate in the crossover. Figure 2.7 illustrates the proportionate selection [58], while Table 2.1 shows this method's application to the individuals in Figure 2.6.

   It should be noted that the most qualified individuals have a higher probability of being selected to mate in this selection strategy. This drives to the problem of one or more individuals will dominate the next generations. Finally, the algorithm will saturate, i.e., at a certain generation. Only a group of individuals that are all the same will be found. The following selection strategy overcomes this problem [54].

Individual	Chromosome	Fitness
I_1	010010	5
I_2	101100	22
I_3	101111	11
I_4	001010	3
I_5	010101	6
I_6	101010	17
I_7	110110	1
I_8	000111	9

Rank	Individual	Chromosome	Fitness
1	I_2	101100	22
2	I_6	101010	17
3	I_3	101111	11
4	I_8	000111	9
5	I_5	010101	6
6	I_1	010010	5
7	I_4	001010	3
8	I_7	110110	1

Rank	Individual	Chromosome	Fitness	
1	I_2	101100	22	keep
2	I_6	101010	17	keep
3	I_3	101111	11	keep
4	I_8	000111	9	keep
5	I_5	010101	6	discard
6	I_1	010010	5	discard
7	I_4	001010	3	discard
8	I_7	110110	1	discard

Figure 2.6: An illustrative example of the population decimation

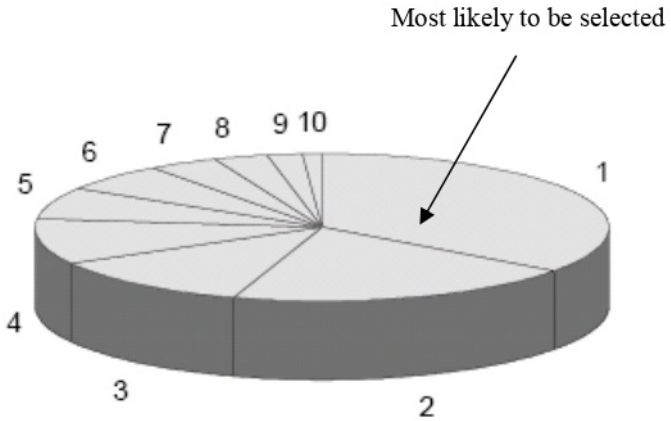


Figure 2.7: Proportionate selection represented as a roulette wheel

Rank	Individual	Chromosome	Cost	probability
1	I_2	**101100**	**22**	**0.282**
2	I_6	**101010**	**17**	**0.218**
3	I_3	**101111**	**11**	**0.141**
4	I_8	**000111**	**9**	**0.115**
5	I_5	**010101**	**6**	**0.077**
6	I_1	**010010**	**5**	**0.064**
7	I_4	**001010**	**3**	**0.038**
8	I_7	**110110**	**1**	**0.013**

Table 2.1: Application of proportionate selection to the individuals in Figure 2.6

3. **Rank Selection**

   The Roulette wheel will have a problem when the fitness values differ very much. If the best chromosome fitness is 90%, its circumference occupies 90% of the Roulette wheel, and then other chromosomes have too few chances to be selected. Rank Selection ranks the population, and every chromosome receives fitness from the ranking. The worst has fitness 1, and the best has fitness N. It results in slow convergence but prevents too quick convergence. It also keeps up selection pressure when the fitness variance is low. It preserves diversity and hence leads to a successful search. In effect, potential parents are selected, and a tournament is held to decide which of the individuals will be the parent. There are many ways this can be achieved, and two suggestions are [55],

   (a) Select a pair of individuals at random. Generate a random number, R, between 0 and 1. If $R < r$ use the first individual as a parent. If $R >= r$, then use the second individual as the parent. Moreover, this is repeated to select the second parent. The value of r is a parameter to this method.

   (b) Select two individuals at random. The individual with the highest evaluation becomes the parent. Repeat to find a second parent.

4. **Tournament Selection**

   An ideal selection strategy should be able to adjust its selective pressure and population diversity to fine-tune GA search performance. Unlike the Roulette wheel selection, the tournament selection strategy provides selective pressure by holding a tournament competition among N_u individuals.
   The best individual from the tournament is the one with the highest fitness, which is the winner of N_u. The winner is then inserted into the mating pool. The tournament competition is repeated until the mating pool for generating new offspring is filled. The mating pool comprising of the tournament winner has higher average population fitness. The fitness difference provides the selection pressure, which drives GA to improve the fitness of the succeeding genes [55].

2.4.2.2 Crossover Strategies

The crossover operator is usually the primary operator working only as a mechanism to introduce variety into the population. The schemes differ from binary to real number encoding.

1. **Binary GA Crossover**

For binary encoding GA, there are many ways to perform a crossover. The selected parents simply interchange parts of their chromosomal structure according to one or more randomly set interchange points, called crossover sites. The number of exchange points is left to the programmer's discretion[54].

- **Single Point Crossover**

After reproduction, the crossover can be done in two steps. First, members of the recently reproduced chromosomes in the mating pool are mated randomly. Then, each pair of chromosomes undergoes crossover as follows: an integer position k on the chromosome is selected uniformly, randomly chosen between 1 and the length of the chromosome minus one. Two new chromosomes are generated by swapping all the genes between k+1 and the length of the chromosomes [43] (see Figure 2.8).

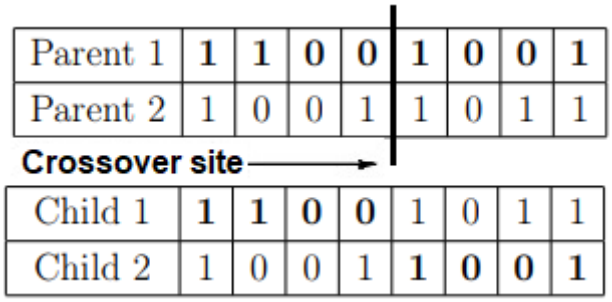


Figure 2.8: Single point crossover

- **Two-Point Crossover**

Two-point crossover is very similar to single-point crossover, except that two cutoff points are randomly generated instead of one (see Figure 2.9).

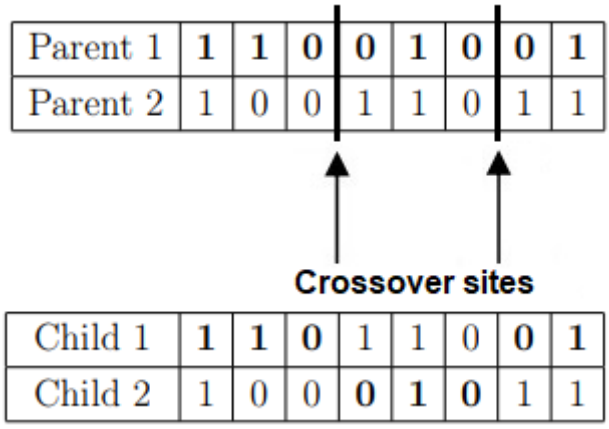


Figure 2.9: Two-point crossover

- **Multi-Point Crossover (N-Point Crossover)**

  There are two cases in this crossover. The first one is an even number of crossover sites; they are randomly selected around a circle, and information is exchanged. The other case is an odd number of crossover sites.

- **Uniform Crossover**

  Uniform crossover is very different from the multi-point crossover. Each gene in the offspring is generated by copying the corresponding gene from parents according to a randomly created binary crossover mask of the same length as the chromosomes. When there is a 1 in the crossover mask, the gene is copied from the first parent, and when there is a 0 in the mask, the gene is copied from the second parent. A new crossover mask is randomly generated for each pair of parents. Offsprings, therefore, contains a mixture of genes from each parent. The number of active crossover points is not fixed but is the average of L/2 (where L is the length of the chromosome) [55].
  In Figure 2.10, new children are produced using the uniform crossover method. It can be seen that in the production of child 1, when there is a 1 in the mask, the gene is copied from parent 1, otherwise from parent 2. In the production of child 2, when there is a 1 in the mask, the gene is copied from parent 2. Else the gene is copied from parent 1.

Parent 1	**1**	**0**	**1**	**1**	**0**	**0**	**1**	**1**
Parent 2	0	0	0	1	1	0	1	0
Mask	**1**	**1**	**0**	**1**	**0**	**1**	**1**	**0**

Child 1	**1**	**0**	0	**1**	1	**0**	**1**	0
Child 2	0	0	**1**	1	**0**	0	1	**1**

Figure 2.10: Uniform crossover

- **Three Parent Crossover**

  In this crossover technique, three parents are chosen at random. Each bit of the first parent is compared with the bit of the second parent. If the two are identical, the bit is taken for the offspring. If not, the bit of the third parent is taken for the offspring. This concept is illustrated in Figure 2.11.

Parent 1	**1**	**0**	**1**	**1**	**0**	**0**	**1**	**1**
Parent 2	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
Parent 3	1	1	0	1	0	1	1	0

Child	1	**0**	0	**1**	0	**0**	**1**	0

Figure 2.11: Three parent crossover

2. **Crossover Techniques in Order Coded GA**

Binary crossover techniques are not applicable to order coded GA. For example, in Figure 2.12, by applying binary single-point crossover, the obtained offspring chromosomes are not valid.

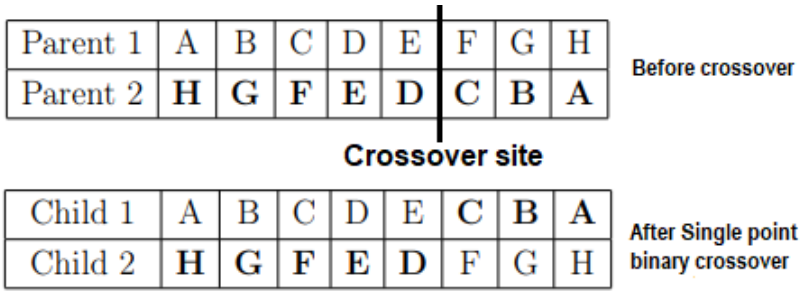


Figure 2.12: Crossover in order coded GA

Since the sequence of gene values is important, Binary crossover techniques are not applicable to order coded GA.

- **Order Crossover (OX)**

  - **Single Point Order Crossover**

    With two parents and a random crossover point. The single point order crossover behaves as follows:
    Child 1 inherits its left section from parent 1, and child 2 inherits its left section from parent 2. For the right section of child 1, copy the gene value from parent 2 in the same order as they appear but not already present in the left section. For the right section of child 2, copy the gene value of parent 1 in the same order as they appear but are not already present in the left section (see Figure 2.13).

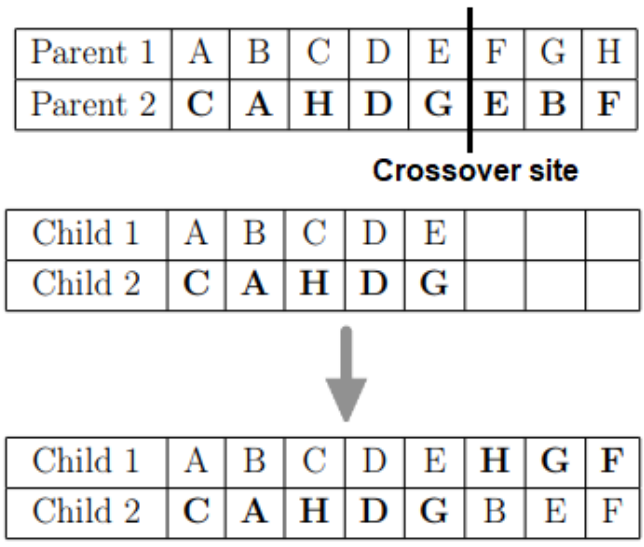


Figure 2.13: Single-point order crossover

– **Two Point Order Crossover**

In the presence of two chromosomes as parents, and two random crossover points are selected, separating them into a left, middle, and right portion. The ordered two-point crossover behaves as follows:
Child 1 and child 2 inherit their middle section from parent 1 and parent 2, respectively. The left and right sections of child 1 are determined by the genes of the left and right sections of parent 1 in the order in which the values appear in parent 2. A similar process is applied to determine child 2. The process is illustrated in Figure 2.14.

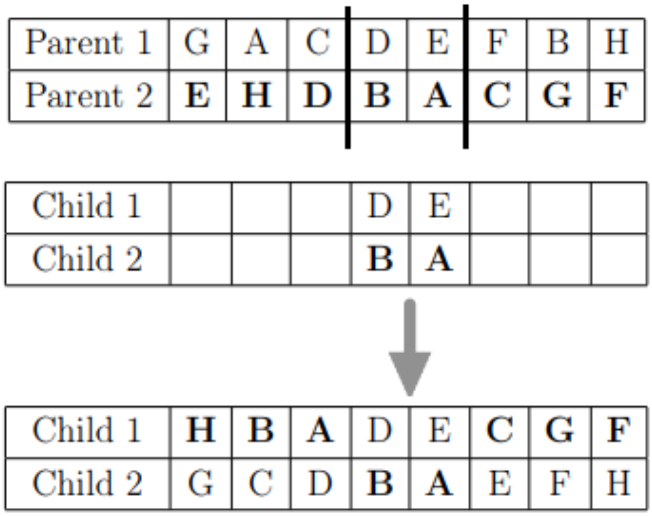
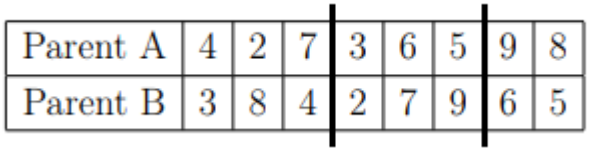


Figure 2.14: Two-point order crossover

- **Partially Matched Crossover (PMX)**

In Partially Matched Crossover, two strings are aligned, and two crossover points are selected uniformly at random along the length of the strings. The two crossover points give a matching selection, which is used to affect across through position by position exchange operations [55].
Consider two strings:



Two crossover points were selected at random, and PMX proceeds by position-wise exchanges. In-between the crossover points, the genes get exchanged, i.e., the 3 and the 2, the 6 and the 7, the 5 and the 9 exchange places. This is by mapping parent B to parent A. Now mapping parent A to parent B, the 7 and the 6, the 9 and the 5, the 2 and the 3 exchange places. Thus after PMX, the offspring produced as follows:

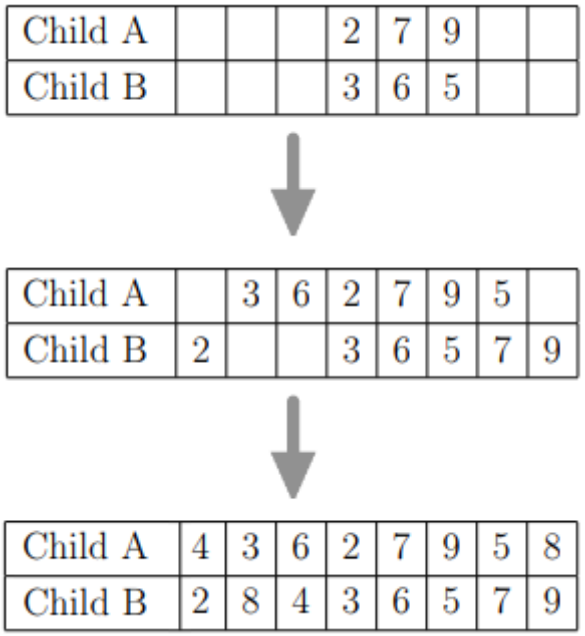


Figure 2.15: Partially matched crossover

- **Precedence Preservative Crossover (PPX)**

  PPX is illustrated in Figure 2.16, for a problem consisting of six genes A–F. The operator works as follows:

  - Create a vector of length equal to chromosomes' length, randomly filled with elements from the set {1,2}. This vector defines the order in which the operations are successively drawn from parent 1 and parent 2.
  - We can also consider the parent and offspring permutations as lists, for which the operations 'append' and 'delete' are defined.
  - First, we start by initializing an empty offspring.
  - The leftmost operation in one of the two parents is selected in accordance with the order of parents given in the vector.
  - After an operation is selected, it is deleted in both parents.
  - Finally, the selected operation is appended to the offspring.
  - This step is repeated until both parents are empty, and the offspring contains all operations involved [55].

2.4.2.3 Mutation Strategies

The Mutation is a background operator that produces spontaneous random changes in various chromosomes. A simple way to achieve mutation would be to alter one or more genes. In GA, mutation serves the crucial role

Parent 1	A	B	C	D	E	F
Parent 2	C	A	B	F	D	E

Select parent (1 or 2)	1	2	1	1	2	2

Offspring permutation	C	A	B	F	D	E

Figure 2.16: Precedence preservative crossover

of either replacing the genes lost from the population during the selection process or providing the genes that were not present in the initial population.

The mutation probability is defined as the percentage of the total number of genes in the population. The mutation probability controls the probability with which new genes are introduced into the population for trial. If it is too low, many genes that would have been useful are never tried out. However, if it is too high, there will be much random perturbation, the offspring will start losing their resemblance to the parents, and the algorithm will lose the ability to learn from the history of the search [59].

1. **Binary GA Mutation**

   It is simply changing a 1 to 0 or vice-versa depending on a probability. In general, the probability of mutation is very small (typically less than 0.2) to avoid losing the chromosomes' good properties (see Figure 2.17).

Before mutation	1	1	**0**	1	0	0	1	0
After mutation	1	1	**1**	1	0	0	1	0

Figure 2.17: Binary mutation

2. **Real GA Mutation**

   Up to now, several mutation operators have been proposed for real numbers encoding.

   - Random mutation*:* operators such as uniform mutation, boundary mutation, and plain mutation belong to the conventional mutation operators, which simply replace a gene with a randomly selected real number with a specified range.
   - Dynamic mutation (non-uniform mutation): is designed for fine-tuning capabilities to achieve high precision, which is classified as the arithmetical mutation operator.

- Directional mutation*:* operator is a kind of direction-based mutation, which uses the gradient expansion of the objective function. The direction can be given randomly as a free direction to avoid the chromosomes jamming into a corner. If the chromosome is near the boundary, the mutation direction given by some criteria might point toward the close boundary, and then jamming could occur.

Several mutation operators for integer encoding have been proposed [59].

- Inversion mutation*:* selects two positions within a chromosome at random and then inverts the substring between these two positions (see Figure 2.18).
- Insertion mutation*:* selects a gene at random and inserts it in a random position (see Figure 2.19).
- Displacement mutation*:* selects a substring of genes at random and inserts it in a random position. Therefore, insertion can be viewed as a particular case of displacement (see Figure 2.20).
- Reciprocal exchange mutation*:* selects two positions random and then swaps the genes on the positions (see Figure 2.21).

2.4.3 Generation Replacement

Once the new offspring solutions are created using crossover and mutation, we need to introduce them to the parental population. There are many ways we can approach this. Bear in mind that the parent chromosomes have already been selected according to their fitness, so we hope that the children (which includes parents who did not undergo crossover) are among the fittest in the population. So we would hope that the population will gradually, on average, increase their fitness. Some of the most common techniques are outlined below [60].

- Delete-all: This technique deletes all the current population individuals and replaces them with the same number of chromosomes that have just been created. This is probably the most common technique and will be the choice technique for most people due to its relative ease of implementation. It is also parameter-free, which is not the case for those listed below.
- Steady-state: This technique deletes **n** old individuals and replaces them with **n** new individuals. The number to delete and replace, n, at any one time is a parameter to this deletion technique. Another consideration for this technique is deciding which individuals to delete from the current population. Do you delete the worst individuals, pick them at random, or delete the chromosomes that you used as parents?
- Steady-state-no-duplicates: This is the same as the steady-state technique, but the algorithm checks that no duplicate chromosomes are added to the population. This adds to the computational overhead but can mean that more of the search space is explored.

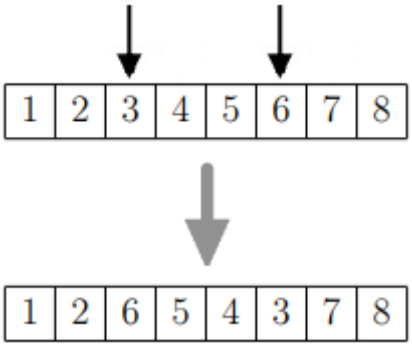


Figure 2.18: Inversion mutation

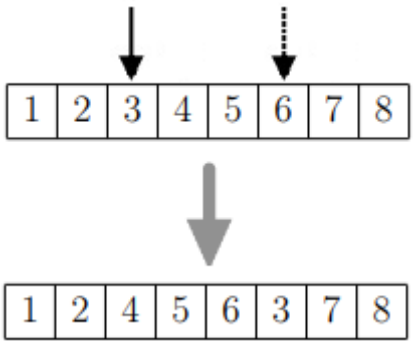


Figure 2.19: Insertion mutation

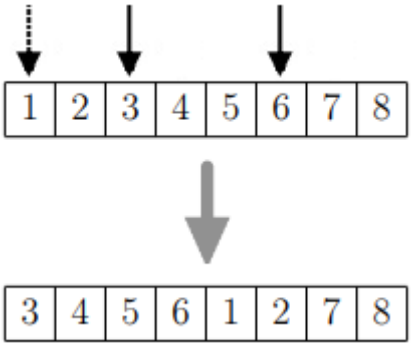


Figure 2.20: Displacement mutation

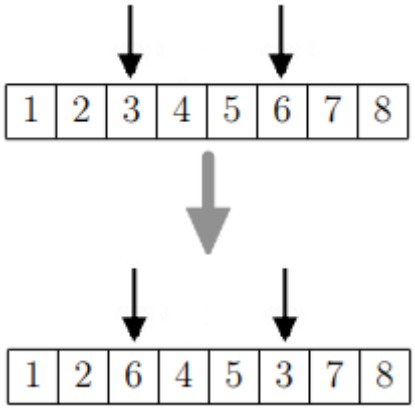


Figure 2.21: Reciprocal exchange mutation

2.4.4 Stopping Criteria

The stopping criteria is an important issue in evolutionary modeling. Early termination may generate poor solutions, whereas late termination might cause high time consumption. The proposed GA is terminated if one of the following cases is reached [61]:

- Maximum number of generations;
- Acceptable fitness reached;
- The maximum number of generations allowed without replacing the fittest reached chromosome.

2.5 Conclusion

In this chapter, an overview of the genetic algorithm and its appearance is provided. Initially, we have given the important definitions and terminology. After we have presented the basic process of genetic algorithms and the tasks must be able to achieve. Next, the genetic algorithm operators are detailed. Finally, the stopping criteria are given.

In the next chapter, we will apply the Genetic Algorithm to the problem already described in chapter 1, which is the Generalized Cubic Cell Formation Problem.