

OPTIMISING THE FREQUENCY ASSIGNMENT PROBLEM
UTILIZING PARTICLE SWARM OPTIMISATION

by

WILLIAM BEZUIDENHOUT

DISSERTATION

submitted in the fulfilment
of the requirements for the degree

MAGISTER SCIENTIAE

in

INFORMATION TECHNOLOGY

in the

FACULTY OF SCIENCE

at the

UNIVERSITY OF JOHANNESBURG

SUPERVISOR: DR G.B. O'REILLY

MAY 2012

Contents

List of Figures	ix
-----------------	----

List of Tables	xiii
----------------	------

I Background on the problem domain and influential algorithms	1
--	----------

1 Introduction	3
-----------------------	----------

1.1 Introduction	3
1.2 Research Questions and Objective	4
1.3 Chapter Breakdown	6
1.3.1 Part I - Background	6
1.3.2 Part II - Implementation	8

2 Cellular Communication Technology	9
--	----------

2.1 Introduction	9
2.2 GSM Networks	10
2.2.1 A Brief History of GSM Networks	11
2.3 Topology of a GSM Network	15
2.3.1 Base Station Subsystem	17
2.3.2 Mobile Switching Centre	20
2.3.3 Network databases	22
2.3.4 GSM Network Management Entities	23
2.4 GSM Interfaces	24
2.5 GSM Channels	25

2.6	Handover	29
2.7	Summary	32
3	The Frequency Assignment Problem	35
3.1	Introduction	35
3.2	NP-Complete	37
3.3	Constraint Handling mechanisms	38
3.4	Frequency Allocation Types	39
3.4.1	Fixed Frequency/Channel Assignment	39
3.4.2	Dynamic Frequency/Channel Assignment	40
3.5	Interference	41
3.6	Frequency Assignment Problem types	46
3.6.1	Minimum Order Frequency Assignment Problem	47
3.6.2	Minimum Span Frequency Assignment Problem	47
3.6.3	Minimum Interference Frequency Assignment Problem	48
3.7	fixed spectrum frequency assignment problem (FS-FAP) Math- ematical Formulation	48
3.8	FAP Benchmarks	50
3.8.1	Philadelphia Benchmarks	50
3.8.2	CELAR	51
3.8.3	COST 259	51
3.9	FAP in the Industry	54
3.9.1	Satellite Communication	54
3.9.2	Wireless Mesh Networks and Wireless Local Area Net- works (WLANs)	54
3.9.3	Military Field Communication	55
3.9.4	Television and Radio Broadcasting	56
3.9.5	Cellular Communication	56
3.10	Summary	57

4	Metaheuristic Algorithms	59
4.1	Introduction	59
4.2	Search Spaces and States	60
4.3	Metaheuristics Algorithms	61
4.4	Tabu Search	64
4.4.1	Introduction	64
4.4.2	Important Tabu Search Characteristics	65
4.4.3	Flow of the algorithm	70
4.4.4	Tabu Search on the frequency assignment problem (FAP)	72
4.5	Simulated Annealing	74
4.5.1	Introduction	74
4.5.2	Important Simulated Annealing Characteristics	76
4.5.3	Flow of the Algorithm	79
4.5.4	Simulated Annealing on the FAP	81
4.6	Genetic Algorithm	82
4.6.1	Introduction	82
4.6.2	Important Genetic Algorithm Characteristics	85
4.6.3	Flow of the Algorithm	89
4.6.4	Genetic Algorithm on the FAP	92
4.7	Summary	94
5	Swarm Intelligence	95
5.1	Introduction	95
5.2	Stigmergy	98
5.3	Ant Colony Optimisation (ACO)	99
5.3.1	Introduction	99
5.3.2	ACO Characteristics	102
5.3.3	Flow of the Algorithm	108
5.3.4	ACO on the FAP	110
5.4	Artificial Bee Colony (ABC) Algorithm	112
5.4.1	Introduction	112
5.4.2	ABC Algorithm Characteristics	116

5.4.3	Flow of the Algorithm	119
5.4.4	ABC algorithm on the FAP	121
5.5	Particle Swarm Optimisation (PSO)	123
5.5.1	Introduction	123
5.5.2	PSO Characteristics	124
5.5.3	Flow of the Algorithm	131
5.5.4	PSO on the FAP	133
5.6	Summary	134
 II Discussion and results of implementing a PSO algo-		
rithm on the FAP		137
 6 Applying the particle swarm optimisation (PSO) to the FAP		139
6.1	Introduction	139
6.2	Position in the Frequency Planning Domain	140
6.3	The Fitness Function	144
6.4	Velocity Function for Frequency Planning	145
6.4.1	Movement in the Frequency Planning Domain	147
6.4.2	Keeping Frequencies Bounded	153
6.4.3	Using Indices instead of Frequencies	156
6.5	Building a Global Best	161
6.6	Keeping History	166
6.7	Summary	170
 7 Results		171
7.1	Introduction	171
7.2	PSO COST 259 Siemens Results	172
7.2.1	Siemens1	173
7.2.2	Siemens2	174
7.2.3	Siemens3	175
7.2.4	Siemens 4	176
7.2.5	Comparison with best results in COST259	177
7.3	The Performance of the PSO	177

7.3.1	Velocity Method 1 vs Method 2	177
7.3.2	Different Global Schemes	178
7.3.3	Population Size	180
7.4	Summary	181
8	Conclusion	183
8.1	Introduction	183
8.2	Research Questions	183
8.3	Research Objective	185
8.4	Chapter Breakdown Revisited	185
8.4.1	Part I - Background	185
8.4.2	Part II - Implementation	187
8.4.3	The Conclusion	188
8.5	Future Work	188
	Acronyms	193
	Bibliography	199

List of Figures

2.1	GSM architecture [53]	16
2.2	Cells with BTSs	18
2.3	Cell Sectorization	19
2.4	TDMA frame and logical channels [83]	26
3.1	Co-channel interference	42
3.2	Adjacent channel interference	42
3.3	Frequency Separation	43
5.1	The shortest path bridge experiment [26]	100
5.2	Visual particle velocity update [33, 34, 90, 96]	128
6.1	The structure of a frequency plan	142

List of Algorithms

1	Basic Tabu Search Algorithm [82,95]	70
2	Basic Tabu Search Algorithm (continued) [82,95]	71
3	Basic Simulated Annealing Algorithm [85,86]	80
4	Basic Genetic Algorithm Algorithm [38,61,119,122]	90
5	Ant System Algorithm [34]	108
6	Ant System Algorithm (continued)	109
7	Basic Artificial Bee Colony Algorithm [59]	119
8	Basic Artificial Bee Colony Algorithm (continued)	120
9	Basic Global Particle Swarm Optimisation Algorithm [34] . .	132
10	The FAP PSO Algorithm	143
11	FAP Cost Function	146
12	Velocity Method 1	149
13	Subtract One Position From Another	150
14	Multiply Position by a Value (Method 1)	151
15	Add One Position to Another (Method 1)	152
16	BoundValue Method	154
17	Velocity Method 2	158
18	MoveIndices	159
19	ApplyVelocity	159
20	CalculateIndexVelocity	160
21	Standard Gbest Selection in FAP PSO	161
22	Building Global Best with Cells	164
23	Building Global Best with Transceivers	164
24	SanitizePosition	167
25	ResolveCollision	167

List of Tables

4.1	Results of applying tabu search (TS) with heuristic manipulation technique (HMT) on COopratiOn europene dans le domaine de la recherche Scientifique et Technique (COST) 259	73
4.2	SA on COST 259 Benchmark	81
4.3	GA on COST 259 Benchmark	92
4.4	Summary of algorithm performance on the COST 259 benchmark	94
5.1	ACO and ant colony optimisation (ACO)* on custom GSM FAP benchmark [68]	112
7.1	Siemens1 results using velocity method 1	173
7.2	Siemens1 results with velocity method 2	173
7.3	Siemens2 results with velocity method 1	174
7.4	Siemens2 results with velocity method 2	174
7.5	Siemens3 result with velocity method 1	175
7.6	Siemens3 result with velocity method 2	175
7.7	Siemens4 results using velocity method 1	176
7.8	Siemens4 results using velocity method 2	176
7.9	FAP PSO results compared to best obtained results	177

Part I

Background on the problem domain and influential algorithms

Chapter 1

Introduction

1.1 Introduction

In the technology age, life is almost unfathomable without the mobile phone. It is hard to believe how the business world managed to function in the pre-mobile phone era¹.

The invention of the mobile phone fulfilled the need to always be connected and within reach of the modern world. This need can actually be attributed to feeling part of something and this something is deemed as being part of a network. This is similar to how mobile phones are able to provide connectivity at almost any location within a country.

Behind the connectivity of mobile phones lies an intricate layer of communication technology which has evolved since the invention of the normal landline phone. This technology is referred to at the top level as wireless communication, which includes inventions such as the radio. To enable communication at the level needed by mobile phones, wireless communication had to evolve and go a step further than normal radio communication. Hence the concept of cellular mobile networks was developed.

Without cellular networks mobile phones have no means of providing communication. The purpose of the cellular network is to provide the necessary functions to connect and facilitate communication between two entities (mobile phones) over a wireless network.

Cellular networks achieve this level of communication through expensive equipment and “smart” algorithms. By “smart” algorithms, what is actually

¹Pre 1980s where the mobile phone did not yet exist

meant is algorithms that utilise artificial intelligence concepts to perform a certain function in an attempt to either automate a task within the network or improve a certain part of the network. Therefore, the correct term for these types of algorithms is actual artificial intelligence (AI) algorithms.

AI algorithms have a wide spread of functions that they can perform ranging from making informed intelligent decisions to optimising the operation of certain processes. In particular computers are indeed very apt at optimising certain procedures. This is because a computer is able to test and evaluate a huge number of different alterations and combinations of a certain procedure in a short amount of time, thereby allowing it to find the best combination out of those tested.

In this dissertation an algorithm is presented which concentrates on cellular phone networks. This particular algorithm falls into the latter part of the AI algorithms discussed, namely an AI optimisation algorithm. Thus the algorithm presented in this dissertation operates on a cellular phone network to optimise a certain part of the network.

In this section a introduction was presented to the dissertation. The main themes of the dissertation were outlined namely, cellular phone networks and optimisation algorithms. In the next section research plan and chapter breakdown is presented which is followed by this dissertation.

1.2 Research Questions and Objective

As previously discussed, an artificial intelligent algorithm is presented in this dissertation that generates plans for use by cellular networks. The purpose of creating such an algorithm is to determine the applicability of modern swarm-based algorithms with regard to finding better solutions to real-world problems that exist in domains such as cellular communication.

The research approach followed in this dissertation was first and foremost to understand the cellular network domain. More importantly, the focus is on how and what frequencies are used to facilitate communication as well as what affects the quality of the communicational link between two entities in a cellular network.

The cellular domain is the problem domain, but to develop a solution in the problem domain a study needs to be conducted in the artificial intelligence domain, namely a study of optimisation algorithms. In order for

a new optimisation algorithm to be developed, other algorithms that have achieved success in the respective optimisation problem domains in which they have been applied need to be investigated.

Therefore in an attempt to develop and apply a modern viable optimisation-based algorithm, a series of research questions has been identified which need to be answered. The research questions are as follows:

- **Question 1** — What is cellular technology, how it got developed and what improvements have been made since its initial development?
- **Question 2** — What is the architecture behind a modern cellular network, how do the various hardware entities within a network communicate with each other and how is a communicational link established between two users of the network?
- **Question 3** — What exactly is the frequency problem and how does it affect modern wireless communication?
- **Question 4** — What variants of the frequency problem exist and which are most applicable to cellular networks?
- **Question 5** — What are the most popular optimisation algorithms and what characteristics make them unique?
- **Question 6** — With algorithms that achieved success in their respective optimisation problems, what particular technique is used by the algorithms that allowed them to achieve better performance?

The PSO algorithm might be short and elegant, but applying it to the FS-FAP required various new techniques in response to the following questions:

- How can a particle best be represented as a frequency plan?
- How can one frequency plan move to another?
- How can particles be prevented from using forbidden frequencies when they fly towards a particular plan?

The above questions were only the preliminary questions and were in fact a problem that had to be addressed for successful application of the PSO to

the FAP. In chapter 6 a discussion is presented on how these problems were solved as well as how other problems were solved that were not anticipated.

Throughout the course of this dissertation the aim is to answer each of these identified research questions. In the next section a chapter breakdown of this dissertation is presented.

The aim of this dissertation is first and foremost to answer the identified research questions. Answering these questions will aid in reaching the research objective of this dissertation. The research objective is therefore to develop and determine the viability of an algorithm based on modern swarm optimisation techniques that will operate on the frequency assignment problem in an attempt to provide better frequency plans for use in cellular networks.

1.3 Chapter Breakdown

1.3.1 Part I - Background

The first part of this dissertation is concerned with the domain and problem the algorithm presented in this dissertation addresses, and finally to also understand the intricate details of how the algorithm operates.

Below is an outline of the chapters in the first part of this dissertation.

Chapter 1

This chapter provides an introduction to the dissertation as well as a broad overview of the topics in the dissertation will discuss.

Chapter 2

This chapter is concerned with providing information on how a modern cellular network functions. Within this chapter a brief history is presented on how cellular network technology was developed. The chapter also provides an overview of the architecture of a cellular network, and each part of the network's intended purpose and function to facilitate wireless communication is discussed.

Chapter 3

This chapter presents the problem that the dissertation addresses, namely the frequency assignment problem. The chapter provides a discussion on why the problem exists, the causes of the problem and what it means for a problem to be NP-Complete. Furthermore the variants of the problem and how they differ depending on the wireless domain that is being considered are also discussed. Finally the chapter also provides a formal definition of the problem which is later utilised by the algorithm developed in this research.

Chapter 4

This chapter marks the beginning of a discussion on various optimisation algorithms in this dissertation. The algorithms presented in this chapter were chosen due to their widespread usage as well as success on NP-Complete problems. Each algorithm is discussed in depth providing an outline of the core features that make the algorithm unique as well as each core feature in detail. For each algorithm the chapter also presents an analysis on related work of the particular algorithm being applied to the frequency assignment problem.

Chapter 5

This chapter is concerned with providing algorithms that are new in the research domain of optimisation algorithms. The algorithms presented in chapter 4 are fairly old and have been applied to a wide variety of problems. Algorithms in this chapter are relatively new in the optimisation domain and have not been applied to the same number of problems as the algorithms in chapter 4. In this chapter swarm algorithms are presented and the algorithms have the particular characteristic that they are based generally on processes observed in nature. Each algorithm is discussed in depth with its core characteristics outlined. Furthermore for each algorithm an analysis is given if the algorithm were to be applied to the frequency assignment problem. Finally it is formally stated what algorithm this dissertation is applied to the frequency assignment problem in this research.

1.3.2 Part II - Implementation

Chapter 6

In this chapter various optimisation problems are presented. The problems were chosen due to their usage in the literature to test other optimisation algorithms. The problems are used to test two variants of the particle swarm optimisation algorithm. The tests were done not only to determine the general viability of the algorithm, but also to understand the algorithm better. The chapter also presents the results obtained by applying the algorithms to the test problems, together with a discussion on the performance of the algorithms.

Chapter 7

This chapter provides a discussion of the algorithm developed to be applied to the frequency assignment problem. Within this chapter an outline is given of the process in developing a specialised particle swarm algorithm for the frequency algorithm. Each specialised technique developed is discussed in depth, along with an explanation of why the technique is needed as well as why it is used by the algorithm.

Chapter 8

This chapter is concerned with providing the results after applying the algorithm to a specialised set of benchmark problems for frequency assignment algorithms. The particular selected benchmark problems were discussed in chapter 2.

Chapter 9

This chapter concludes this dissertation. In this chapter it is determined whether the research goal was reached as well as whether any future work can be done to improve the presented algorithm.

Chapter 2

Cellular Communication Technology

2.1 Introduction

Wireless technology is used by modern electronic devices to establish communication over which information is exchanged [2]. Wireless technology facilitates communication between two devices by transmitting data or voice via a radio frequency [2]. Examples of devices exchanging information wirelessly are radios for audio entertainment; television remotes to change channels; cellular phones for communication; wireless access points to create wireless LANs [2].

The popularity and rapid adoption of wireless technology has made it difficult to plan, manage and operate wireless networks [2].

As the popularity and use of services that use wireless technology increase, the need for these services to use different radio frequencies to communicate becomes greater [83]. This need arises due to an effect called *interference*, which occurs when two or more connections between entities use the same radio frequency to facilitate communication [83]. This effect is discussed in more detail in chapter 3.

A wireless cellular operator is not allowed to operate on just any frequency. A governing body licenses a certain piece of the available wireless spectrum to the operator for use in their network [29]. These frequencies that need to be licensed for use are known as *commercial* frequencies and are very scarce as it is immensely expensive to license [29].

The use of frequencies by services must be carefully considered to avoid interference when entities communicate with each other. This is a problem as the number of devices that communicate far outstrips the amount of frequencies available for communication [83]. Thus assigning frequencies to entities for communication is a difficult problem and is referred to as the FAP [2, 31].

Initially manual techniques were used to assign frequencies in an attempt to solve the FAP [2]. As a result, the assignment of frequencies was either too complex or just too daunting because of the sheer number of devices that need to be assigned frequencies [2]. Also, because of the rapid adoption of wireless technology the assignment of frequencies needed to be dynamic and hence automated [2].

By understanding how a general system for mobile communication (GSM) network operates, especially how communication is facilitated, the reader will have a better basis to understand the problem this dissertation addresses. Therefore, this chapter will start off with an overview of GSM networks and also a brief history of GSM will also be given. A discussion on the topology of GSM will then follow. This chapter concludes with a discussion how the different GSM entities communicate with each other and how a piece of spectrum is divided into frequencies for communication.

2.2 GSM Networks

The GSM is a system for multiservice cellular communication that is capable of providing voice as well as data services [53, 83]. Most cellular networks in operation are GSM based [2, 83]. The primary service that GSM caters for is voice communication, but other data services such as short message service (SMS), multimedia message service (MMS) and internet connectivity services, e.g. general packet radio system (GPRS), enhanced data global evolution (EDGE) and high speed circuit switched data (HSCSD), are becoming more important [31, 53].

GSM is one of the most widely used radio communication technologies, which is why one needs to look at the history behind it in order to understand the domain of radio communication better [53]. A brief history of the GSM network specification will now be presented in the next section.

2.2.1 A Brief History of GSM Networks

In early 1981 a group known as the *Groupe Speciale Mobile* was established to develop a europe-wide radio communication system using the reserved 900 MHz band¹

At the start of the GSM specification in the early 1980s it was initially thought that the system would be analogue based, but this soon changed with the integrated service digital network (ISDN) specification nearing completion [78]. As such the GSM specification started following many of the same design principles and access protocols that ISDN exhibited [78].

After the completion of the ISDN specification, the advantages of switching to digital instead of analogue for communication became clear [78] One of the primary advantages of ISDN is that it is capable of transmitting data at higher speeds [78] GSM would therefore be based on digital transmission and speech would be represented by a digital stream of 16 kbits/s [78].

Before the switch to digital transmission was finalised the Groupe Speciale Mobile first wanted to evaluate the spectral efficiency of analogue and digital-based transmission [78]. Spectral efficiency plays an important part in wireless communication since the radio spectrum is a limited resource and whichever transmission technology is used, the utilisation of the spectrum should be maximised [78].

Maximum utilisation is an important problem that is discussed in detail in later sections of this chapter. After an evaluation of spectral efficiency it was decided that the GSM system would be digitally based using time division multiple access (TDMA) [74,78]. TDMA is discussed in section 2.5.

By the early 1990s GSM became an evolving standard and the first GSM-based network was demonstrated in 1991² [31,53]. The following year a number of GSM networks were operating in Europe due to mobile terminals and equipment capable of operating on the networks becoming more widely available to the general public [31,78]. In the same year an operator in Australia became the first non-European operator to implement a GSM-based network [31].

¹In 1990 the United Kingdom requested that 1800 MHz band be added to the scope of the Groupe Speciale Mobile standard group. This variant of the GSM specification became known as the digital cellular system (DCS1800) [2,78].

²Near the end of 1991 the GSM group was renamed *Speciale Mobile Group* (SMG) to eliminate confusion between the standard and the group

The collective subscriber base of GSM networks surpassed the million subscriber mark in 1993 [78]. Due to this phenomenal growth in GSM network use, numerous extensions were made to the GSM specification. Some of the extensions that were made are the following [53, 78]:

- Half rate speech codec
- Improved SMS
- Line identification
- Call waiting
- Call holding

The specification with these extensions is known as GSM Phase 2+. As the world shifted towards more digital and data-intensive services it became difficult to deliver these services over GSM networks. This difficulty was due to the restriction that data could only be transmitted at 9.6 kbps [2, 78].

This restriction also applies to voice, as voice is encoded into data that is transmitted over the network [2, 78]. Before the extensions were made, voice was encoded using the Full Rate codec [53]. The Full rate codec was the first codec used to encode voice in the GSM standard and required the full 9.6 kbps that was available [53]. With half rate speech telephony, only half of the 9.6 kbps is required for the same voice quality to be delivered [53].

The new specification defined new technologies such as GPRS, EDGE and HSCSD, which were designed with the primary goal of making more bandwidth available for data transmission [2, 53]. Data transmission was improved by these technologies by enabling more than one GSM slot to be used for a terminal or service at a time [2, 53].

If more than one GSM slot is to be used by a terminal or service, transceivers are required to have a higher signal-to-noise ratio (SIR) ratio [53, 74]. This requirement affects radio interfaces as there is a higher likelihood that interference might occur; hence it makes it more difficult to generate a low interference frequency plan [31, 74].

The actual SIR at a receiver is dependent on a number of factors that include [2, 53]:

- Frequency used at the transceiver

- Strength of the signal
- Weather conditions
- Shape of the surrounding environment
- Direction of the transmission

Even taking these factors into account, the calculation of the SIR at a transceiver is not trivial. This calculation of the SIR as well as its impact on mobile radio frequencies is discussed in section 3.5.

As the GSM standard matured as a cellular technology, industry experts began specification of the next generation of cellular networks, which would in time replace the GSM cellular system.

The universal mobile telecommunications system (UMTS) is considered the third generation (3G) of cellular networks [31, 109]. UMTS was designed from the beginning to operate in parallel with the legacy GSM system. The first standard of UMTS was issued in the beginning of 2000 and subsequently most modern networks are based on it or are migrating their networks to it [31, 109].

UMTS is a major improvement over the GSM in two areas, namely data transmission bandwidth and frequency planning due to UMTS utilising direct sequence collision detection multiple access (DS-CDMA) and wide band collision detection multiple access (WCDMA) [31, 109]. The higher data transmission speed (2 Mb/s) can be attributed to UMTS using the DS-CDMA scheme [31, 109]. The scheme also allows more users to be served than previous generations of networks [31, 109].

DS-CDMA and WCDMA sends data over a wide band spectrum of 5 MHz. The wider spectrum allows for more frequencies to be utilised, is that the FAP is easier to solve compared to GSM [31, 109]. A more in depth discussion on how a range of spectrum is divided into frequencies for communication in GSM networks is discussed in section 2.5. DS-CDMA and WCDMA will not be discussed in depth as this discussion focusses on GSM. The interested reader is directed to [109] and [83] for more information.

Code division multiple access (CDMA) is a technology primarily used in broadband systems [53]. Users do not gain access to only a small portion of the bandwidth, but rather use the entire band for the duration of a connection [53]. Users also do not gain exclusive access to the whole band, but

instead share the usage of the bandwidth with other users simultaneously, hence the name *multiple access* [53].

With CDMA a user's signal is not transmitted as its original signal. Instead the signal is spectrally spread over a multiple of its original bandwidth using a spreading factor [53]. The spreading factor fluctuates between values of 10 and 1 000 [53]. Using these spreading factors less interference and fewer disturbances are encountered because the broadband signal is generated from a narrowband signal [53]. UMTS may be a major improvement, but its adoption does not spread very far from busy city centres where there is a large concentration of clients in a small geographical area [53]. The reason for this is that, as mentioned previously, UMTS caters for larger data usage and therefore more clients can be serviced simultaneously [53].

Most network operators do not implement entirely new backbone architecture for UMTS to operate on, but instead utilise the same backbone used for GSM and GPRS [53]. This not only extends the lifetime of previous infrastructure investment by the operator, but also builds upon the redundancy provided by the GSM network [53]. Thus even with new technological improvements such as UMTS, GSM as a wireless technology is still used for communication and is therefore still relevant today [53].

The most up-to-date cellular network technology is referred to as the fourth generation of cellular networks (4G) [123]. The purpose of 4G is to improve upon 3G and bring true broadband speeds that is delivered over the cellular network [123]. There are currently two competing technologies that aim to provide 4G capabilities to cellular networks which is long term evolution (LTE) and Mobile worldwide interoperability for microwave access (WiMAX) [123]. Both technologies provide broadband speeds in the range of 14 - 40 Mbits/s [123]. According to the 4G standard both technologies cannot be considered as "true" 4G technologies as the standard defines that broadband speeds in excess of 100 Mbits/s must be provided to be considered true LTE [123].

In this section a brief overview on the history of the GSM network was presented. In the next section an explanation of the topology of GSM network is given. This will broaden the understanding of GSM networks.

2.3 Topology of a GSM Network

A GSM network consists of a variety of different subsystems to realise the goal of establishing a radio communication link between two parties. The hierarchy of systems and their respective connections to each other are illustrated in figure 2.1. Each subsystem will now be discussed.

Mobile Station

A mobile station (MS) as it is defined in the GSM specification refers to any mobile device that is capable of making and receiving calls on a GSM network. The MS is the main gateway for a user to gain access to the GSM network [31, 53]. Typical devices that fall under the category of MS are cellular phones, smart phones and [point of sale (POS) devices. The MS has two features that play an important role throughout the GSM network, namely:

subscriber identification module (SIM) — Inserted into a mobile device. The SIM contains the international mobile subscriber identity (IMSI) and is used throughout the network for authentication as well as being a key part in providing encrypted transmissions [31].

international mobile equipment identity (IMEI) — Used to identify mobile station equipment. Primarily used in the denial of service to equipment that has been blacklisted³ and tries to gain access to the network [31].

³Equipment can be blacklisted for a variety of reasons, e.g. theft

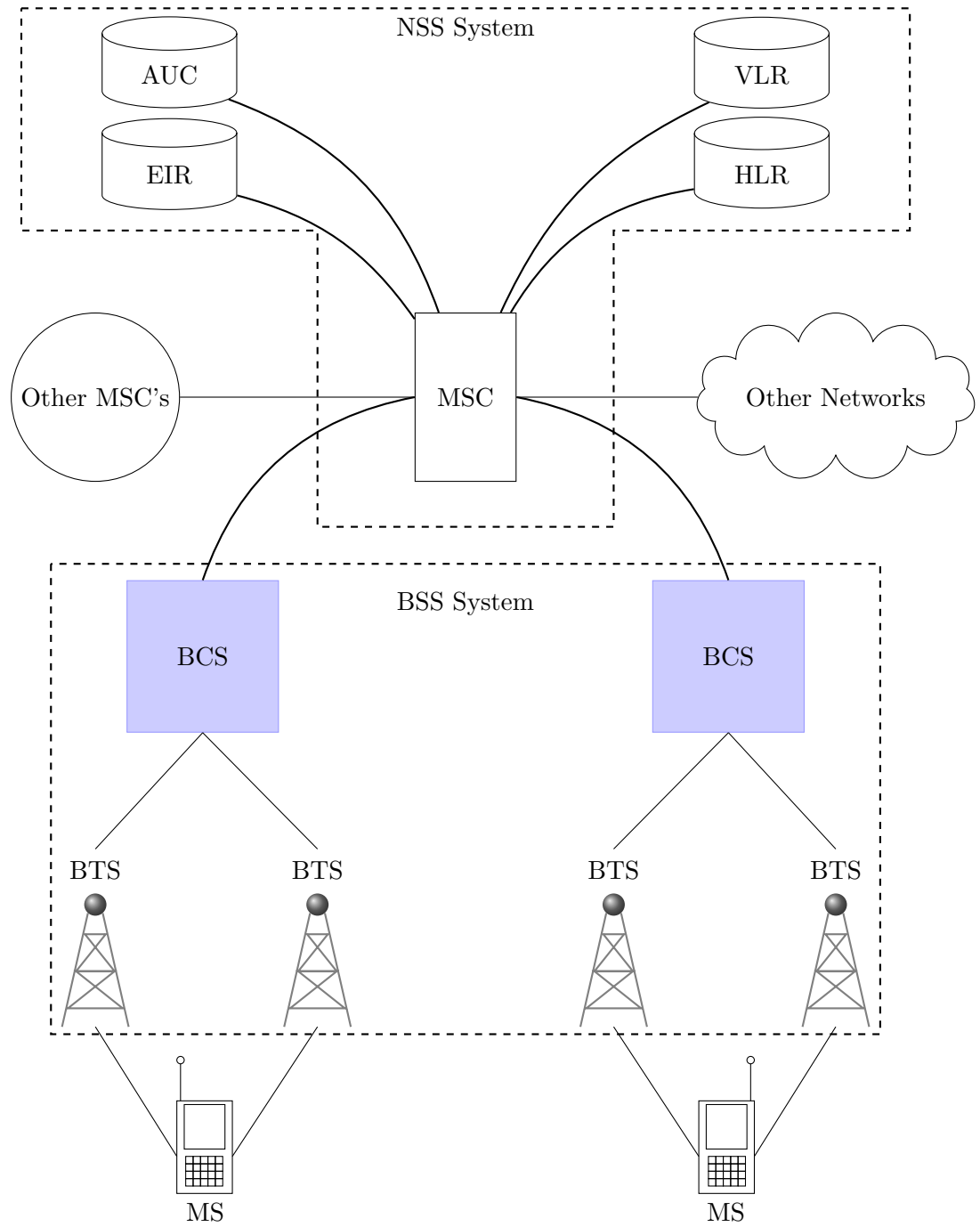


Figure 2.1: GSM architecture [53]

The MS has the capability to change the transmission power it uses from 0.8, 2, 5, 8 MW to a maximum value of 20 MW [74]. The change in transmission power is automatically set to the lowest level by the base transceiver stations (BTS) to ensure reliable communication after evaluating the signal strength as measured by the MS [53, 74]. The power adjustment also minimises co-channel interference because the transmitted signal is less powerful and therefore less likely to interfere with other signals [74]. Co-channel interference is discussed in chapter 3

2.3.1 Base Station Subsystem

According to the GSM Phase 2+ specification, this system is viewed by the mobile switching centre (MSC) through an Abis radio interface. The base station subsystem (BSS) is responsible for communicating with mobile stations in a particular location area [31]. The BSS usually consists of one base station controller (BSC) with one or more BTSs that it controls [31]. A BTS has similar equipment to that of a MS [74]. Both have transceivers, antennae and the necessary functions to perform radio communication.

The communication link between the MSC and BSC is called the A-interface. The interface between the BSC and BTS is called the Abis interface [31]. The A and Abis interface as well as other interfaces between other GSM architecture components are discussed in section 2.4.

In a GSM network a service area is defined as a geographical area where the network wants to deliver its cellular network service to potential clients [2, 53]. The service area is subdivided into location areas (LAs) which are then further divided into smaller radio zones called cells [91]. When a cellular network is modelled, cells are modelled as hexagonal shapes. Each cell in the modelled network is served by only one BTS and is usually regarded to be in the centre of a cell as can be seen in figure 2.2 [53].

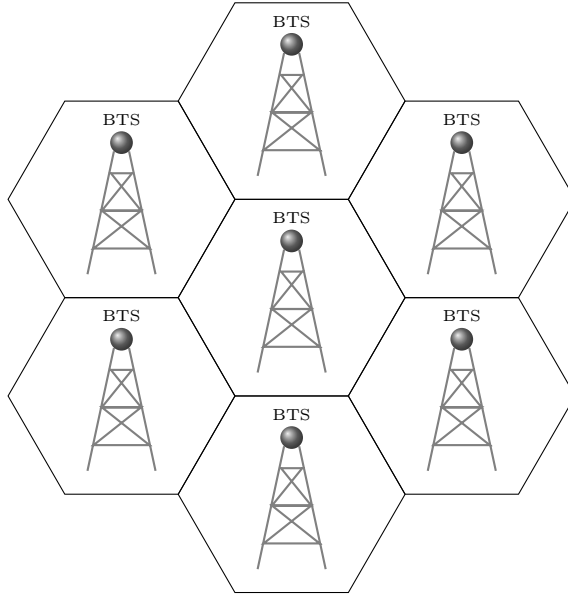


Figure 2.2: Cells with BTSs

z

Even though cells are modelled as being hexagons (see figure 2.2), the actual coverage area of a cell has no predefined regular shape [53].

For the purpose of this research a BTS is considered to be assigned channels which are used for communication. No discussion will be presented on the electro-magnetic wavelength properties of channels as it has no effect on the results presented nor problem being investigated.

With the network modelled as a series of interconnecting hexagons it allows one to more easily take constraints into account [31]. For a cell to serve its geographical area, it needs to be allocated frequencies to operate on. Therefore, for each cell i in the modelled network a subset S_i of frequencies from the total frequencies F allocated to the GSM network is assigned [53]. Neighbouring cells must at all costs avoid having the same subset of frequencies allocated to them, since such a scenario would lead to severe interference on any communication and thus degrade quality [53].

Since the number of cells in a network greatly outnumber the available of subset frequencies available, one is forced to start reusing frequency subsets in cells [53]. To ensure that the reused frequency subsets do not interfere with their neighbouring cells, a reuse distance D is defined [53]. The reuse distance means that a certain number of cells must be between the

cell already assigned the frequency subset S_i and the cell to be assigned a frequency subset [53]. The amount of cells is the distance value D .

The size of a cell determines the amount of potential traffic that the cell will be required to handle [31, 53, 78]. Therefore, if the size of a cell is chosen to be small, fewer channels will be required to be allocated to that cell as it would not be required to handle as much traffic as a larger cell [53].

By making the size of cells smaller, the network operator is required to invest more into its network infrastructure. Smaller cells have a direct consequence that more cells would be required to serve the same geographical area, and more cells means that more BTSs etc. need to be built and maintained, more locations need to be rented [53]. Hence, making a cell smaller has a compounding effect on the amount of infrastructure needed to support it.

Fortunately all the extra infrastructure investment by the operator can be greatly scaled down if cells are divided into sectors. Each sector performs the exact same function as a traditional cell and is therefore regarded to also be a cell, just smaller in size and not omnidirectional [53, 74, 78].

A cell is divided into 3 to 6 service sectors and each sector is allocated an antenna/transceiver [74]. Figure 2.3 illustrates a cell divided into sectors.

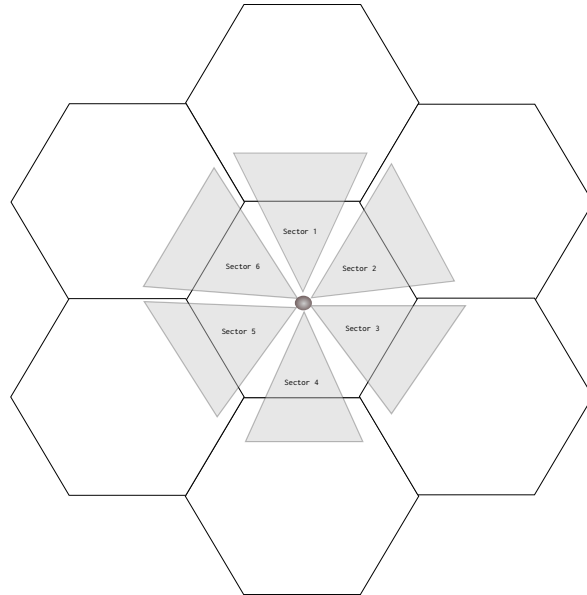


Figure 2.3: Cell sectorisation [74].

Depending on how many sectors are at a cell, the operating angles of the antennae need to be adjusted accordingly to ensure 360 degree service. If there is only one sector, an omnidirectional antenna is used, otherwise the antennae operating angles are adjusted to $\frac{360^\circ}{n}$ where n is the number of antennae [31].

By dividing the cell into sectors the amount of co-channel interference that would occur in a cell is greatly reduced [53]. It is important to note that the reduction of co-channel interference is only applicable when the angle of the antenna by which transmission occurs is restricted [53].

Suppose, if a cell using an omnidirectional transceiver is assigned 6 channels. If the cell were to be divided into three sectors, where the sectors' antennae are 120° apart, the number of interfering co-channels shrinks from 6 to 2 and from 6 to one in the case when the cell is divided into 6 sectors [53, 74, 78].

Each sector operates one or more elementary transceivers called transceiver (TRX). The number of TRX per sector is determined by the expected peak traffic demand that the cell must be able to handle. Each TRX can handle 7 to 8 communication links or calls in parallel except the first TRX, which handles fewer calls than normal because it is responsible for transmitting cell organisation and protocol information [31]. TRXs are able to handle 7-8 calls in parallel due to the use of frequency division multiplexing (FDM) and time division multiplexing (TDM) schemes.

TRXs are assigned channels, which enable them to provide conversion between digital traffic data on the network side and the radio communication between MSs and the GSM network [15, 68]. The channels that are used by a cell for communication are discussed in section 2.5.

2.3.2 Mobile Switching Centre

The MSC is at the heart of a cellular switching system and forms part of the network switching subsystem (NSS). The MSC is responsible for the setting up, routing and supervision of calls between GSM subscribers [74, 78]. The MSC has interfaces to communicate with GSM subscribers (through the BSS) on the one hand and with external networks on the other [78]. The MSC interfaces with external networks to utilise their superior capability in data transmission as well as for the signalling and communication with

other GSM network components [78].

The most basic functions that an MSC is responsible for in a network are the following [83]:

- Voice call initialization, routing, control and supervision between subscribers
- Handover process between two cells
- Location updating
- MS authentication
- SMS delivery
- Charging and accounting of services used by subscriber
- Notification of other network elements
- Administration input or output processing functions

To achieve most of these functions the MSC has an integrated visitor location register (VLR) database that stores call setup information for any MS that is currently registered for service with the MSC [78, 83].

The VLR retrieves this information from the home location register (HLR) that contains all the registered GSM subscriber information for the network. This information enables the MSC to quickly retrieve the necessary information to set up a call between two clients that want to communicate with each other [74, 91].

A requirement for being able to communicate with other network elements such as public switching telephone network (PSTN) is the ability to multiplex and demultiplex signals to and from such network elements. This operation is a necessity, since the incoming or outgoing connection bitrate from the source entity might be either too low or too high for the receiving entity.

A typical scenario where this operation proves vital is when a mobile subscriber makes a call to a subscriber on a PSTN. The connection bitrate needs to be changed at the MSC from a wireless connection bitrate to a bitrate suitable for transmission over a PSTN. In chapter 3 section 3.5 more information regarding bit rates is presented.

2.3.3 Network databases

The HLR, authentication centre (AUC) and equipment identity register (EIR) are the three ‘back-end’ databases which store and provide information for the rest of the GSM network. In this subsection each one of the databases that form part of the ‘back-end’ is discussed briefly and a description is given of the core functions that each database performs in the network.

Home location register — The HLR is a database that permanently stores information pertaining to a given set of subscribers. It needs to store a wide range of subscriber parameters because it is the reference source for anything GSM subscriber related in the network [74].

Subscriber parameters that are stored in the database include billing information, routing information, identification numbers, authentication parameters and subscribed services [74]. The following information is also stored, but the information is of a temporary nature and can change at any time: The current VLRs and MSC the subscriber is registered with and whether the subscriber is roaming [74].

Authentication centre — The AUC is the entity in the GSM network that performs security functions and thus stores information that enables it to provide secure over-the-air communication [74, 78]. The information that is stored contains authentication information as well as keys that are used in ciphering information [74, 78].

During an authentication procedure no ciphering key is ever transmitted over the air; instead a challenge is issued to the mobile that needs to be authenticated. This challenge requires the mobile station to provide the correct signed response (SRES) with regard to the random number generated by the AUC [74, 78]. The random number and ciphering keys that are used change with each call that is made; thus an attacker would gain nothing by intercepting a key, since it will change with the next call [74].

Each mobile that is registered in the HLR database needs to be authenticated and each call that is instantiated needs to retrieve keys from the AUC to establish a secure communication link [74, 78]. The AUC is sometimes included with HLR to allow for fast communication between the two databases [74].

Equipment identity register — The EIR is a database that stores the IMEI numbers of all registered mobile equipment that has accessed the network [74]. Only information about the mobile equipment is stored, nothing about the subscriber or call is stored in the database [74].

Typically there is only one EIR database per network and interfaces to the HLR databases contained in the network [74]. The IMEIs are grouped into three categories: *white list*, *black list* and the *gray list* [74]. The white list contains only the IMEI numbers of valid MSs; the black list stores the IMEI numbers of equipment that has been reported stolen and the gray list stores the IMEI numbers of equipment that has some fault (faulty software, wrong make of equipment) [74].

2.3.4 GSM Network Management Entities

In a GSM network most of the elements that form part of and make the network function are often distributed in a wide geographical area to provide the best network coverage for the customer [74].

For a network to function properly and efficiently network engineers need to be kept up to date on the state of the network and be alerted if *any* problems occur [74]. For this purpose there are two systems in the GSM network architecture that allow for this functionality required by network engineers [74].

One system is called the operations and management centre (OMC) which is responsible for centralised regional and local operational and maintenance activities [74]. The other system is called the network management centre (NMC) and unlike the OMC it provides global and centralised management for operations and maintenance of the network supported by the OMCs [74].

In the following paragraphs more in depth discussions on the critical functions the OMC and NMC perform is presented.

Operational and Management Centre — The OMC is capable of communicating with GSM network components using two protocols, namely SS7 and X.25 [74]. The SS7 protocol is usually used when the OMC is communicating within the GSM network over short and medium distances [74]. The

X.25 protocol is used for large external data transfers [74]. All communication where the OMC is involved occurs over fixed line networks and/or leased lines. The OMC is usually used for day-to-day operation of a network [74].

The OMC has support for alarm handling [74]. An alarm in a GSM network goes off whenever a predefined expected condition does occur. Engineers are able to define the severity of an alarm, which defines who or what is further alerted and if the alarm needs to be escalated to a higher level [74].

The OMC is also capable of fault management in the GSM network [74]. It is able to activate, deactivate, remove and restore a service manually or automatically on network devices [78]. Various tests can be run and diagnostic information can be retrieved on the network devices to detect any current or future defects [74].

Network Management Centre — The NMC is similar to the OMC but it is not restricted to only regional GSM network components as it is in charge of all the GSM network components in the network [74]. The NMC provides traffic management for the global network and also monitors high priority alarms such as overloaded or failed GSM network components [74]. It is usually used in long-term planning of a network, but it has the capability to perform certain OMC functions when an OMC is not staffed.

2.4 GSM Interfaces

In the GSM network all the network components communicate with each other through predefined interfaces. In this section an overview is given of these interfaces between the components.

Um interface — This interface is the link between an MS device and a BTS and is also referred to as the *Air* interface since communication occurs wirelessly. The primary protocol used on this interface is the link access protocol D channel modified (LAPDm), which is an extension of the ISDN LAPD protocol to accommodate the mobile nature of MS devices as well as for the shorter TDMA frames which are used in GSM networks [83, 91].

Abis interface — Between the BTS and BSC the interface used for communication is known as the Abis interface. The only messages that the BTS is interested in are those that have to do with management of radio resources [83, 91]. All other messages are left alone and merely pass through the BTS to the BSC transparently.

A interface — The interface between the BSC and MSC is known as the A interface. This interface is used for the transfer of information, which is used by the MSC to manage BSSs, control connections and manage the mobility of MS in its area of service [53, 83].

Other Interfaces — The MSC has interfaces going from itself to databases and other external networks. Each interface connects to a specific database, MSC or network and is therefore very specific as to what function it performs [53, 83]. An interface going from one MSC to another will convey information regarding handling the administration of handover of an MS device leaving one administrative area to another MSC's administrative area. The administrative area of an Msc is the section of BSS components the Msc is responsible for. The handover procedure is a very delicate process which is described in section 2.6.

In this section a brief overview was given of the most critical interfaces used by all the network components of the GSM network involved. In the next section the difference between a logical channel and a frequency is described. Additionally an outline and overview of all the logical channels defined in the GSM is presented.

2.5 GSM Channels

GSM defines a series of logical channels, which are used for communication over these interfaces. A distinction needs to be made between channels and frequencies. As discussed earlier, a network is licensed in a certain section of the wireless spectrum for use for commercial communication. This piece of spectrum is referred to as bandwidth and is measured in Hz, therefore W Hz, where W denotes the allocated bandwidth [120].

This bandwidth W is then divided into N smaller chunks of bandwidth called narrowband chunks. Each N narrowband chunk is a channel and has

a width of W/N Hz [120]. This channel of W/N Hz width is also referred to as a frequency.

Using TDMA the GSM system is able to provide additional transmission capacity by dividing the frequency into eight equal timeslots [83]. As can be observed from figure 2.4 each TDMA frame has a series of consecutive timeslots. Each timeslot can be used for both uplink and downlink transmission. A GSM *channel* is a logical channel and refers to a single timeslot within a TDMA frame [53, 83].

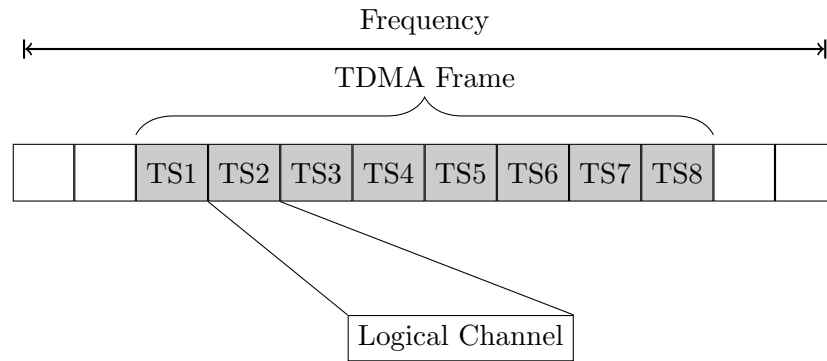


Figure 2.4: TDMA frame and logical channels [83]

The GSM system is therefore able to use the same physical frequency in eight different timeslots without interference as these *logical* channels are used at different times. Therefore using TDMA the available channels that can be used for communication in GSM are increased eightfold [83].

Frequencies are assigned to the uplink and downlink portion of the connection with a duplex separation of 45 Mhz in the frequency band to avoid interference between uplink and downlink. There are two types of channels, traffic channels and control channels. traffic channels (TCHs) primary purpose is to enable communication of user speech and data and therefore carry no control information [53].

A TCH is assigned to an MS device when the device indicates that it needs to communicate with another device either with speech or data. When an MS has finished with the TCH the allocated TCH is reclaimed for use by other MS devices on the network. This request by the MS device occurs using the control channels [53].

Control channels are much more actively used in a GSM network since

they are the primary means by which control and management of the network occurs [53]. These channels are used even when the MS has no active connection and is in idle mode. This constant activity on the control channels is to keep the network updated with information such as the position of the MS (location updating) and signal strength [31, 53, 74].

The control channels are divided into three main channel groups namely broadcast channel (BCH), common control channel (CCCH) and dedicated control channel (DCCH) [53]. Each of these channel groups contains other channels that aid in the control and management of the network. Each group along with the associated channels will now be briefly discussed.

The first group, BCH, consists of three channels:

broadcast control channel (BCCH) — This channel is broadcast using the very first frequency assigned to a cell. Using this frequency, the channel broadcasts information regarding the network. This information includes radio channel configuration of the current and neighbouring cells, synchronisation information, registration identifiers and most importantly the format of the CCCH used by the local BTS [53].

frequency control channel (FCCH) — Synchronisation information is broadcast to the MSs to enable them to perform frequency correction on the transmission. Typical synchronisation information on this channel, for instance, is the exact frequency the local BTS is using for transmission to enable the MSs to attune themselves to the same frequency [53].

synchronisation channel (SCH) — On this channel, identifying information regarding the BTS is transmitted. Also on this channel information regarding synchronisation of frames is sent which aids an MS to, for example, structure the time frames of TDMA frames.

The FCCH and SCH are always broadcast with the BCCH since these channels are needed for the operation of the radio subsystem [53]. The CCCH is a point-to-point signalling channel that is used to localise an MS through the use of paging [53].

When a MS is paged in a GSM network, a CCCH message is sent to all the cells in the surrounding area of the last known cell the MS was recorded to be at [53]. The message is sent to surrounding cells in case the MS

changed location since it was last communication with the network [53]. The MS receives the paging through the CCCH and responds if it is intended for the particular MS [53]. The channel is also used to assign dedicated channels [53]. The CCCH is made up of the following channels:

random access channel (RACH) — The RACH forms the uplink portion of the CCCH and is randomly accessed by the MS to request a dedicated channel for a single signalling transaction [53].

access grant channel (AGCH) — The AGCH forms the downlink part of the CCCH. It is used by the radio subsystem to assign a stand-alone dedicated control channel (SDCCH) or TCH to an MS [53].

paging channel (PCH) — The PCH also forms part of the downlink portion of the CCCH. This channel is used by the radio subsystem to page specific MSs, which aids in the process of locating an MS [53].

notification channel (NCH) — This channel is used to inform an MS of any incoming group calls or calls that are being broadcast [53].

The last group of signalling channels is referred to as dedicated/associated control channel (D/ACCH). This group of channels has the characteristic that it is a bidirectional point to point channel [53].

Stand-alone dedicated control channel (SDCCH) — This channel is used for communication between the BSS and MS even when there is no active connection [53]. Hence the ‘stand-alone’ since it means that there need not be a TCH assigned for communication to occur between the BSS and MS [53].

slow associated control channel (SACCH) — When a TCH or SDCCH is assigned, an accompanying SACCH is also assigned. The SACCH is used to transmit information for optimal radio operation, which can include information on power control of the radio transmitter and synchronisation information [53]. Packets must be continuously sent over the SACCH as it is used as proof that there is still a physical radio connection [53]. When the MS has finished using the SACCH channel, it transmits a report regarding the current results of the radio signal level, which is continuously measured [53].

fast associated control channel (FACCH) — When more bandwidth is required for signalling purposes, the signal of the TCH is modified using dynamic pre-emptive multiplexing. The additional bandwidth comes at the expense of the user data transport [53]. When a channel is created in this manner it is called a FACCH [53].

Finally, one last channel is defined namely the cell broadcast control channel (CBCH), which shares the same physical channel that the SDCCH uses. On this channel messages of the short message service cell broadcast are broadcast [53].

In this section an overview was given of how *logical* channels are used for communication between an MS and BSS/MSC. These three groups of channels collectively enable the GSM network to facilitate wireless communication, a very important function for a telecommunication network. The following section deals with how the network is able to keep a connection to an MS alive and allow the MS to make calls while the device is moving around geographically within the network.

2.6 Handover

The handover process in a GSM network is initiated when an MS with an active call moves outside the coverage area of a cell, BSS or MSC [31, 53, 83]. A handover might also be initiated because of measurements indicating bad channel quality [53].

Various information needs to be migrated across and shared between the components handling the calls to ensure a smooth handover. Not only the GSM architecture components that need to continuously share information but also the MS. The MS is required to continuously observe and measure signal strength of up to 6 neighbouring cells. The MS does this by monitoring the BCCH [53, 83]. This information is of course relayed to the MSC and BTS. The information plays a critical role in the decision process as to which entity will be the best in taking over the administration of the active call [53, 83].

An MS can in some cases receive the same BCCH from different cells, which are most likely neighbours [53]. This problem of duplicate BCCH from different cells can be attributed to the frequency reuse in the cellular

network as well as to the smaller sector cells forming clusters and therefore overlapping coverage area [53]. As discussed in section 2.3.1, cells are divided into sectors, which lessen the problem of co-channel interference. This division into sectors can cause clustering of cells and overlapping of coverage area.

This creates a problem for the MS since it needs to distinguish between the two cell measurements as it does not know which measurement belongs to which cell [53]. To distinguish between different cells, the MS also tries to determine the identity of each cell it is monitoring. Only cells whose identity can be determined reliably are included in the report sent to the BTS [31, 53, 83]⁴. Using this report the handover algorithm is able to decide how to handle the handover, and which cell needs to take over the call [31, 53, 83].

Once the cell has been selected, the actual handover process starts. Which has to take into account that the frequency allocated to the incoming call from the other cell does not interfere with other present active calls in the cell receiving the handover [31, 53, 83]. A frequency interfering with calls currently active in the cell can cause users to hear other conversations from the interfering call, or experience their call being “dropped” i.e. disconnected [31].

There are three core handover procedures when a handover needs to occur in a GSM network: *intra-BSC*, *inter-BSC* and *inter-MSC*, each of which involves different GSM network components [83].

Intra-BSC — Also known as *intercell handover*, this handover is concerned with the transfer of an active call/connection from an MS to another cell which is controlled by the same BSC as the current cell [53, 83]. The current BTS that manages the active call of the MS constructs a report containing measurement information from the MS, as well as measurements the BTS has taken on signal strength and error bit ratio of the current connection [53, 83]. The error bit ratio is defined as how often part of the data being sent over a connection needs to be retransmitted because an error occurred [53, 83]. An error can be caused by interference, faulty hardware or the signal strength is fading [53, 83].

⁴When cell identity cannot be determined this can be due to environmental factors like signal strength or to packets getting lost/dropped

The constructed report is forwarded to the managing BSC. The report is analysed to determine the necessity of handing over the active call to another BTS. If a handover is deemed necessary, the BSC starts by initialising the BTS to prepare it to handle the new connection [53,83].

The BSC then notifies the MS through the old BTS of the new BTS identity, and properties of the new connection such as TCH frequency, power output etc. After receiving the information about the new connection, the MS makes the necessary adjustments for it to continue operating and handling the call on the new connection [53,83].

Once all the adjustments have been made the MS sends a confirmation to the BSC of the successful handover through the new BTS. The BSC instructs the old BTS that it must relinquish the use of the TCH and its associated SDCCH used by the old MS call. The BSC notifies the MSC of the handover as it is used in network operation reports [53,83].

Inter-BSC — In an inter-BSC handover a call of an MS that is being managed by a BTS is transferred to another BTS which has a *different* controlling BSC. Thus, the call is essentially moved between two BSCs, and it is up to the new control BSC to select a suitable BTS that will actually handle the call of the MS being handed over [53,83].

This handover occurs because the MS is moving or is about to move into a cell that is not controlled by the current BSC. The current BSC detects this and therefore takes the necessary precautions to ensure that the new BSC is able to make suitable provision to assume control of the call within one of its BTSs [53,83].

The BSC informs the managing MSC that a handover to another BSC must occur. The request sent to the MSC contains the identity of the cell managed by a different BSC. The MSC determines the managing BSC of the cell in question and notifies it that it must select and prepare the cell for handover. The new BSC informs the cell to create a new connection, which will be used by the MS once the handover is complete [53,83].

The new BSC informs the MSC of the new connection details which the cell will use to handle the handover and maintain the active connection of the MS [83]. The MSC forwards the connection information to the old BSC, which forwards it to the MS [83]. The MS makes the

necessary adjustments for it and then moves on to the new connection. The MS informs the new BSC that the handover has been completed successfully [83]. The new BSC informs the MSC, which then instructs the old BSC to relinquish the old connection used by the MS [53, 83].

Inter-MSC — With an inter-MSC handover, control and management of an active call on an MS must be transferred to another BTS, which resides in a different area that is managed by a different MSC [83]. The handover process follows the same basic formula as the intra-BSC and inter-BSC handovers once the handover request is made [53, 83]. The BSC managing the BTS to which the MS is going to be handed over is to be determined by the MSC [83]. The BSC is notified by the new managing MSC that certain BTSs must bring a new connection online for the incoming MS; the network components upstream⁵ are notified [53, 83].

The new MSC informs the old MSC of the new connection details. The old MSC transfers the connection information to the MS that is going to be a handover to another BTS [83]. The MS makes the corresponding adjustments and then starts operating on the new connection. The MS informs the BSC of the successful handover [53, 83].

The BSC in turn informs the new MSC, which in turn informs the old MSC that the handover was successful. The MSC then instructs the BSC to ensure that the old connection resources are relinquished by the BTS.

In this section a discussion was presented on the handover process in GSM network. The effect this has on channel selected as well as the different handover procedures were highlighted.

2.7 Summary

In this chapter a broad discussion was given of modern cellular technology, specifically GSM cellular network technology. A brief history on how GSM was developed to be the most widely used cellular technology in use

⁵components upstream are the network components which are higher up in the managing structure. In this instance, BTS – BSC – MSC.

today was provided. The GSM architecture components followed. In the GSM architecture all the entities present in a modern GSM network were identified.

Following the discussion of the GSM entities, a broad overview was given of the communication interfaces used between the GSM network components to communicate with each other. This was followed by a definition of the GSM channels which are used on the interfaces to communicate information.

The chapter concluded with the handover process which is used to allow an MS device to move freely geographically within the network.

Chapter 3

The Frequency Assignment Problem

3.1 Introduction

The FAP¹ is a generalisation of the graph-colouring problem and is consequently an NP-Complete problem [29]. The FAP is an NP-Complete problem due to fact that only a finite number of frequencies can be assigned to TRXs, where the number of transceivers to be assigned frequencies greatly outweighs the number of available frequencies [29]. A more thorough definition of what it means for a problem to be NP-Complete is given in section 3.2.

In wireless communication a huge concern is interference which occurs when frequencies used for communication are close to each other in the frequency spectrum [2]. Interference and its effects is discussed in detail in section 3.5. Essentially for the FAP the primary concern is to develop an approximate plan on assigning frequencies in such a way that interference is kept to a minimum.

Using exact algorithms to find a solution is not practical since the time to find a solution is polynomial. Generally metaheuristic algorithms are used to find optimal solutions to NP-Complete problems [68]. In the chapter 4, a discussion is presented on algorithms that are used to find solutions to NP-Complete problems.

¹Also known as automatic frequency planning (AFP) or channel assignment problem (CAP) [68]

As discussed in chapter 2 network operators are licensed a range of frequencies from the available spectrum. A licensed piece of spectrum contains a series of consecutive frequencies as well as gaps. Gap frequencies are barred from being used by any device within the network as they may have already been allocated to another operator for use [14]. By barring frequencies, a scenario is avoided where the different networks' equipment interferes with their respective operations [14].

Due to the whole spectrum not being available to network operators and only a subset being available for commercial communication as per the frequencies allocated to them, networks opt to reuse their frequencies [14]. The networks do this to maximise the use of their allocated frequencies and to minimise their licensing fees, since if the network needs more frequencies, they need to be licensed [29].

It is not always possible to simply allocate more frequencies to a network even if the network pays the associated fees. The whole commercial spectrum may already have been licensed to various entities. Hence, licensed frequencies are a very valuable and scarce commodity [14, 29].

The chapter is organised as follows. The first section the follows provides a discussion on NP-Complete problems. In section 3.3 constraint handling mechanisms are presented. Section 3.4 presents an overview of different methods for allocated frequencies. Interference is discussed in depth in section 3.5. After the interference discussion a section is presented on the different types of frequencies assignment problems in section 3.6. A mathematical formulation of the fixed spectrum frequency assignment problem is presented in section 3.7. Section 3.8 provides an overview of the different benchmarks used to evaluate frequency assignment problems. After the discussion on the different benchmarks that are available in the domain a section on the where the FAP is encountered in the industrial domain is presented. This chapter concluded with a section that summarises the contents presented.

In the next section an overview of what it means when a problem is NP-Complete is given.

3.2 NP-Complete

The term NP stems from the field known as complexity analysis. Algorithms are typically measured for their worst running time using $O(n)$ notation. The field of complexity analyses is more interested in measuring complexity of a problem than the running time of an algorithm [100].

The field of complexity analysis makes a keen distinction between problems that can be solved by algorithms in polynomial time and problems that cannot be solved in polynomial time using any available algorithm [100]. An algorithm is said to be of polynomial time if the number of steps it needs to perform to solve the problem is given by $O(n^k)$ where k is a nonnegative integer and n refers to the complexity of the input [100].

A distinction needs to be made between *finding* a possible solution and determining whether a result is a valid solution to an NP problem [100]. Verifying whether some result is indeed a solution to an NP problem is a quick operation [100]. Finding a solution through the use of an algorithm is what polynomial time refers to.

Algorithms that are able to solve problems in polynomial time usually have worst case running times of $O(n)$, $O(\log n)$ and $O(n^2)$ [100]. These classes of problems are easy to solve using modern algorithms available today. Problems that can be easily solved with these kind of running times are classified as being in the P range of complexity problems [100].

Another range of problems consists mostly of problems that cannot be solved in polynomial time [100]. Hence, if an algorithm were to try each and every possible solution, it would take an arbitrarily long time, which cannot be determined, which is why these problems also have the characteristic of being non-deterministic [100]. Problems in this range are referred to as being in the NP range of complexity problems [100].

There is another range of NP problems that are a subset of NP problems, which are referred to as the “most extreme” problems within NP. These problems are collectively known as NP-Complete problems and are the most difficult problems to determine feasible solutions for in the NP problem range [100]. The FAP is one such problem [30, 133].

3.3 Constraint Handling mechanisms

Not only is the FAP an NP-Complete problem but it is also a constraint problem. Constraint problem restrict the search space of possible solutions with boundaries. These boundaries are referred to as constraints. The FAP domain has specific constraints defined for each of the different individual problems. Before the various ways in which frequencies can be allocated is discussed a brief discussion needs to be presented on constraint handling mechanisms.

Constraints define the boundaries of the search space. Solutions found to be violating the constraints are considered to be outside of the search space. Various methods exist how to treat solutions found to be outside of the defined search space. These methods are known to be *constraint handling mechanisms* and are discussed below.

In research done by Engelbrecht [34] the following constraint handling mechanisms are defined.

- **Reject infeasible solutions** — Solutions that violate the defined constraints are not even considered and rejected.
- **Penalty function methods** — Modifies the objective function to add a penalty that is enforced if a solution violates constraints. By adding a penalty the infeasible solutions are discouraged.
- **Convert the constrained problem to an unconstrained problem** — By solving the unconstrained problem better solutions can be found although not guaranteed.
- **Preserving feasibility methods** — Candidate solutions are initialised in the search space and satisfy all constraints. These solutions are continuously moved or transformed with specialised operators who ensure that the solutions continue to satisfy all the constraints. The operators ensure that all solutions stay within the bounded search space.
- **Pareto ranking methods** — Solutions are ranked based on the severity of their constraint violations.

- **Repair methods** — Operators are used to transform solutions that violate constraints to solutions that adhere to all the constraints and are thus feasible solutions.

Constraint handling methods need to be used when solving constrained problems like the FAP. Without these methods, the search for solutions are undirected and solutions that cannot be used are presented as most optimal. Reference to these methods will be made where applicable.

In the next section, 2 methods are discussed that define how frequencies are allocated.

3.4 Frequency Allocation Types

In this section the different methods used to allocate frequencies to cells in a cellular network are discussed. Furthermore the method that relates to the specific FAP variant in this dissertation is described.

Within the FAP domain there are different types of FAP, which have emerged over the years as the domain of wireless communications matured and technological requirements changed. These FAP variants are discussed in section 3.6.

There are a variety of FAPs in the wireless communication domain but each individual problem can be classified into one of the following two categories based on the way frequencies are assigned to cells:

- *fixed frequency allocation (FFA)* is where frequencies assigned to cells are static; therefore they cannot be changed until a new assignment plan is calculated [110].
- *dynamic frequency allocation (DFA)* is the process of allocating frequencies to cells as required to meet the current traffic demand imposed on them by clients [110].

3.4.1 Fixed Frequency/Channel Assignment

FFA is the process of permanently assigning frequencies to cells [110]. The frequencies assigned are fixed and cannot be changed immediately while the network is active, since the frequencies assigned to the cell form part of

a delicate frequency plan designed to keep interference on communication links to a minimum [110].

When the channel used by a particular cell for communication is suddenly changed, the cell might start interfering with neighbouring cells' communication links. If the assigned frequencies of the neighbouring cells are close to each other on the frequency spectrum. Hence, if the cell is sectorized (refer to section 2.3.1 in chapter 2 for a discussion on cell sectorisation) it can interfere with a minimum of three and up to a maximum of six neighbouring cells [110].

When an FFA plan is created, cells are assigned frequencies based on the estimated traffic that cell will be expected to handle during peak network usage. FFA is ideally suited for macro cellular networks since the nature of the traffic encountered in such networks has the characteristic of being homogeneous, stationary and predictable [110]. Cellular networks can be classified as being in the macro cellular group of wireless networks.

With FFA, networks are able to permanently allocate a certain subset of frequencies to cells since the nature of the traffic on their network allows them to predict with reasonable certainty the call blocking probability [110]. A call is blocked on the network when a cell has no available frequencies to use when establishing a communication link [110].

In situations where the nature of the traffic is neither homogeneous nor stationary, using the FFA allocation scheme is not feasible as its use of available frequencies is grossly inefficient [110].

A problem that occurs with FFA when all assigned frequencies are in use is that any new call or call handed over will be blocked [110]. The call will be blocked even if adjacent cells have suitable capacity to handle the call [110].

3.4.2 Dynamic Frequency/Channel Assignment

DFA is a channel allocation scheme where frequencies assigned to cells are not permanent but rather assigned to cells as the need arises [110]. Therefore all the frequencies licensed by a particular network are available to each and every cell to establish a communication link as long as the channel does not violate the co-channel reuse constraint [110].

The co-channel reuse constraint must be adhered to otherwise the amount of interference occurring on the communication link will be too much [110]. This constraint forms part of the electromagnetic constraints, which are described in section 3.5.

The DFA allocation scheme is ideally suited for micro cellular wireless networks. The traffic on these networks have the characteristic of being immensely unpredictable as traffic demand varies constantly [40, 103, 110].

As the name indicates, micro cellular wireless networks have much smaller cell sizes than macro cellular networks. Thus a cell in a micro cellular network must handle a lot more handover requests than a cell in a macro cellular network, since an MS with an active connection is much more likely to move out of the coverage area of a micro cell than a macro cell [40, 103, 110].

Since a micro cellular network has increased handover requests between cells compared with a macro cellular network, a DFA scheme must rapidly allocate frequencies to requesting cells that must handle the handovers [40, 103, 110].

DFA is much more efficient than FFA when the amount of mobile traffic on the network is relatively low [40, 110]. On the other hand, when the network is under heavy mobile traffic load, the FFA scheme outperforms the DFA scheme [40]. The DFA scheme is outperformed because it allocates frequencies to cells in an inefficient arrangement that might affect the amount of interference encountered on the network [103].

Finally DFA inherently requires a great deal more computational power than FFA, since the frequencies need to be selected and allocated with great speed, otherwise the cell requesting a channel will not be able to handle the call and will therefore block the call or drop the call [40, 103, 110].

This concludes the discussion on the different allocation schemes used in modern cellular networks. In the next section a description is given of what interference is and why it is important for cellular networks. An overview will also be given of when interference occurs.

3.5 Interference

Interference can be defined as any unwanted signal that is received along with a signal of interest [35]. The unwanted signal is said to *interfere* with

the original signal and as a consequence degrades the original signal quality with unwanted information [35].

Interference usually occurs when two or more entities communicate independently on the same channel or on adjacent frequencies [35, 40]. Other external factors can also contribute to interference on a communication link, such as machines, which inherently produce some sort of electromagnetic distortion, for instance a car's ignition or a big turbine [35, 40]. endfigure Interference that occurs when two signals operate on the same channel can be seen in figure 3.1 and interference that occurs as a consequence of two signals operating on adjacent frequencies can be seen in figure 3.2.

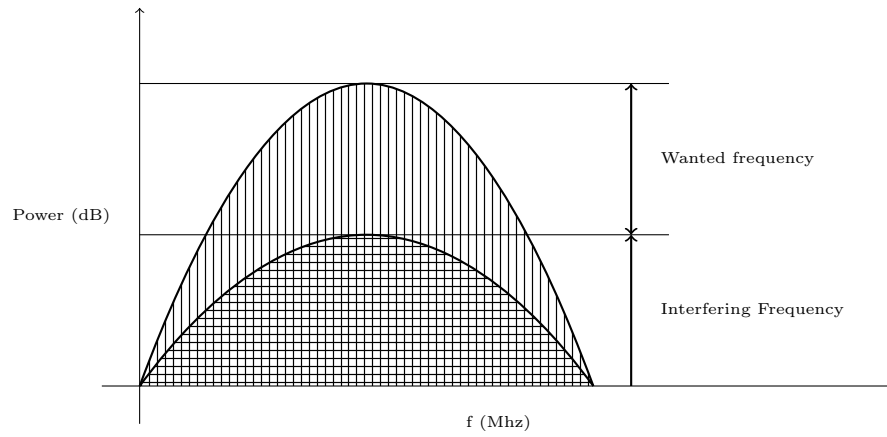


Figure 3.1: Co-channel interference

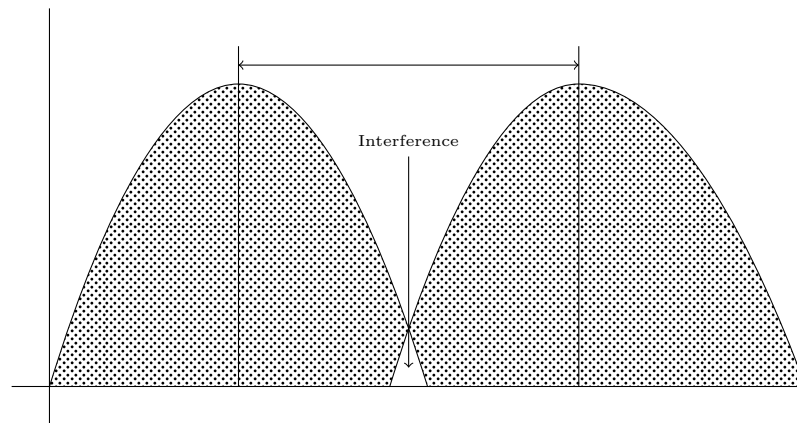


Figure 3.2: Adjacent channel interference

The impact interference will have on an entity that has established a connection and that operates on the same channel or adjacent channel as another entity decreases as the geographic distance between them increases [101]. Therefore, to minimise the impact interference will have on communication links a *separation* is defined [40]. More specifically, this separation is known as the channel reuse or frequency reuse distance within wireless networks [40].

This separation is defined as the minimum number of cells (which must all use different frequencies) between one cell, which has been allocated a channel, and another cell before a cell is allowed to reuse the same channel that another cell has been allocated [40, 101].

The separation can be depicted visually as in figure 3.3 where f_a, f_b, f_c, f_d are different frequencies that are assigned to the specific cells. The frequency f_a is allowed to be reused since the two cells it is assigned to are separated by 3 cells (shaded in gray) because the separation for this network was set to 3.

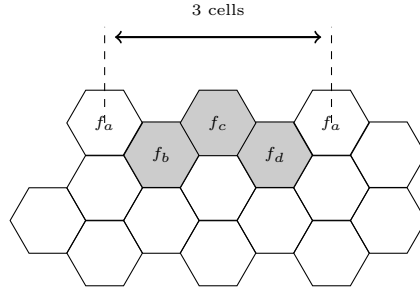


Figure 3.3: Frequency Separation

As discussed earlier, cellular networks are forced to reuse their licensed frequencies multiple times to keep costs to a minimum. Therefore, the design of a cellular network is limited to the defined separation distance between cells as it defines the size of cells that will be in the network [31, 101]. Smaller cells can lead to a larger separation distance compared with when cells are larger [35, 40].

Cellular networks use the amount of interference on their networks as a qualitative measure for their quality of service (QoS) [40]. A network with high interference would experience a lot of dropped connections/calls, which occurs when the interference is too high to sustain a connection or

call for communication; consequently their QoS degrades as interference increases [35, 40].

Even though interference can cause a call or connection to be lost, i.e. dropped, there are other situations where a call can be dropped [74]. For instance, a call can be dropped when a handover procedure occurs between two cells and one cell receiving the call is at full utilisation of its allocated frequencies [35, 74]. A call can also be dropped due to weather conditions which brings forth natural interference or the caller entering a building which drastically reduces cellular reception [35].

In the literature a variety of methods are used to calculate the amount of interference in a network. The SIR ratio is the recommended way of calculating the potential interference at a certain point [2].

The SIR equation is actually based on the signal to interference noise power ration (SINR) but since cellular networks are interference limited, the noise is not considered in the interference calculation [40]. Noise can be disregarded since the power of interference is much larger than the power of noise [35].

A formulation of the SINR and SIR is as follows:

$$sSINR = \frac{P_r}{N_0 + P_I} \quad (3.1)$$

$$SIR = \frac{P_r}{P_I} \quad (3.2)$$

Where P_r is the power of the received signal and P_I is the power associated with interference from within a cell (intracell interference) and interference from outside a cell (intercell interference) [40].

This calculation can be considered a best guess as it models the environment, weather and other factors which may influence the potential interference at a point with a Gaussian distribution with standard deviation for noise represented by the N_0 [2].

Using the SIR formula cellular networks are able to determine the bit error rate (BER) users on the network will experience on their connections [35]. The BER is defined as the probability that a received bit on the connection will be incorrect [103].

As the BER increases voice quality on the connection decreases since more bits that are used to describe the voice information are incorrectly

received [35]. SIR and BER probability are interlinked. As SIR increases, i.e. less interference is encountered on the communication link, the probability that bits will be received incorrectly decreases [35].

Whether precise measurements are taken or the interference is calculated based on the SIR formula, the end result of both methods is that all the calculated or measured values are put into a matrix to produce an *interference matrix* [68].

An interference matrix consists of a number of cell pairs (i,j) , where i is the cell receiving interference and j the cell whose allocated channel is providing the interference [2]. Each cell pair in the matrix has two corresponding values that indicate the level of interference if the *electromagnetic constraints* are violated [30, 68].

Primarily interference occurs when the electromagnetic constraints are violated. These constraints are defined as:

Co-channel — As discussed earlier, when cell i and cell j operate on the same channel interference will occur [74, 110]. When this type of interference occurs it is referred to as *co-channel* interference.

Adjacent channel — When cell i and cell j operate on adjacent frequencies, their allocated frequencies differ by one, i.e. cell i operates on channel f , then if cell j operates on either channel $f - 1$ or $f + 1$, then interference will occur [74, 101]. This type of interference is referred to as *adjacent channel* interference.

The electromagnetic constraints defined above are applicable in any wireless network. With regard to mobile telecommunication networks, such as cellular networks, there are additional constraints that are imposed due to technological requirements, availability, location and size of area with unacceptable interference [2, 30, 31]. These constraints are defined as the following:

Co-site — If cell i and cell j are located at the same site, then their allocated channel ranges must differ by a certain distance in the frequency domain. This distance is known as the reuse distance where cell i and cell j serve different sectors [32, 133]. In the benchmarks which are discussed in section 3.8 this distance is also referred to as the *separation variable*.

Co-cell — Channels used on the same antennae of a cell must differ by a certain number. This is typically set to 3 but can be any number greater than 0 that the network operator deems necessary to avoid unwanted interference [2, 30, 31].

Handover — This constraint means that frequencies must differ by a pre-defined margin, i.e. 2 or 3, when one cell hands over a call to another cell. If this constraint is violated a mobile subscriber will experience a dropped call since the handover between cells fails [2, 30, 31].

Within the licences of wireless networks there are two hard constraints which forbid networks from using certain frequencies. Hard constraints means that under no circumstances are these constraints allowed to be violated.

The first set of hard constraint frequencies is known as *globally blocked frequencies*. Frequencies that are in the set of globally blocked frequencies are usually frequencies that have been licensed to other networks [2, 101].

The second set of hard constraint frequencies is known as *locally blocked frequencies*. These frequencies are not allowed at certain geographic areas but are free for use at any other area [101]. A typical area where certain frequencies will be forbidden to be used is near a country border [101]. The locally blocked frequencies are most likely in use by another network resident to the bordered country.

In this section a description was given of what interference is and what the consequences are of too much interference in a network. This section further elaborated on the circumstances in which interference can occur in a wireless network. In the next section an overview is given of the various different subproblems in the FAP domain.

3.6 Frequency Assignment Problem types

In this section each of the problem variants for the FAP is discussed, starting with one of the first and oldest problems in the FAP domain. This section will conclude with a discussion on the particular variant of FAP focussed on in this research.

3.6.1 Minimum Order Frequency Assignment Problem

The minimum-order frequency assignment problem (MO-FAP) was the first FAP that emerged in the 1970s. The MO-FAP is concerned with assigning frequencies to transmitters while interference is minimised as well as minimising the number of different frequencies that are used [2].

In MO-FAP channel reuse is prioritised and the usage of a channel has a certain cost associated with it. The reason for this is that when the wireless network industry started, operators were billed according to the number of different frequencies they used. In the beginning of commercial cellular networks frequencies were not cheap since they were sold per unit [81].

Over the years as the law governing the wireless spectrum changed and new technology as well as standards emerged, MO-FAP lost its relevancy [2, 81]. Companies are no longer billed according to the different frequencies they use, but they purchase licences from a regulatory body [2, 81]. This licence usually stipulates what channel band the network is allowed to use.

In some instances a certain band of frequencies is put up for auction by a regulatory body, on which interested parties can bid to own the specified spectrum [2, 81]. Due to the shift in how frequencies are allocated to networks, neither the regulatory bodies nor the network operators care about the number of different frequencies are used [2, 81].

3.6.2 Minimum Span Frequency Assignment Problem

The minimum span frequency assignment problem (MS-FAP) is a problem that is very relevant today, especially when network operators want to deploy a new network in a region [2]. The MS-FAP is concerned with keeping the interference below a certain level during assignment as well as minimising the span [99]. The interference threshold used is specified by the network designer as the minimum allowable interference on the network [99].

The span is defined as an interval on the frequency domain. This interval is the difference between the maximum and minimum frequencies used during assignment [2, 99]. With the span value, network operators are able to request certain frequency bands and know their network will be able to operate at suitable interference levels [2, 99].

The MS-FAP and MO-FAP are two very similar problems, the only difference being that MO-FAP focuses on minimising different frequencies

and MS-FAP focuses on minimising the interval of frequencies used during assignment [2]. The Philadelphia benchmark is usually used to gauge how well the algorithm performs.

3.6.3 Minimum Interference Frequency Assignment Problem

The minimum interference frequency assignment problem (MI-FAP) or FS-FAP is typically encountered after the network operator has obtained a frequency band from a regulatory body [2]. Other problems use matrices to forbid certain frequencies from being used by certain transmitters [81].

Unlike the previous problems, in MI-FAP any available channel in the allocated band may be used even though it produces interference. The other problems are concerned with the frequencies used, even though they might be violating some constraints that incur a huge amount of interference [44, 81]. The interference value does not play a large role in their respective objective functions [31, 44]. In MI-FAP the objective is to minimise the total amount of interference on the network. It is important to note that this amount of interference might not necessarily be zero [2, 31].

The MI-FAP is the problem currently most encountered in cellular networks, since there are more operating networks than new networks being designed in the cellular industry today. This particular problem forms the focus of this research.

Since MI-FAP is very close to real-world instance problems, authors tend to use real-world instances or benchmarks to test the quality and efficiency of their algorithms [31, 81]. The quality and efficiency of the solution in this research is benchmarked against the COST 259 benchmark, which is discussed in section 3.8.

In the following section a formal mathematical definition for the fixed spectrum MI-FAP is set out. The definition is important as it forms the basis for the objective/cost function that the algorithm in this research uses.

3.7 FS-FAP Mathematical Formulation

A Mathematical definition of the FAP is given in this section. The mathematical definition is used by the algorithm discussed in this dissertation to

evaluate the amount of interference that generated frequency plans exhibit.

The FAP can be represented as a graph colouring problem and is known to be NP-Complete. Before a mathematical definition can be formally given for the FAP, some symbols and their respective definitions need to be introduced.

$$G = (V, E) \quad (3.3)$$

$$V = \{v_0, v_1, \dots, v_i\} | i \in \mathbb{N} \quad (3.4)$$

$$E = \{v_0v_1, v_0v_2, \dots, v_iv_j\} | v \in V, \forall ij \in \mathbb{N}, i \neq j \quad (3.5)$$

$$D = \{d_{01}, d_{02}, \dots, d_{ij}\} | \forall \{i, j\} \in E, \exists d_{ij} \in \mathbb{N}^+ \quad (3.6)$$

$$P = \{\{p_{00}, p_{01}\}, \{p_{10}, p_{11}\}, \dots, \{p_{i0}, p_{i1}\}\} | \forall \{i, j\} \in E, \exists p_{ij} \in \mathbb{N}^+ \quad (3.7)$$

$$F = \{0, 1, 2, 3, \dots, k\} | \forall k \in \mathbb{N}, \forall v \in V \exists f \in F \quad (3.8)$$

$$d_{ij} < |f(i) - f(j)|, \forall ij \in \mathbb{N}, i \neq j \quad (3.9)$$

Let G (see equation 3.3) be a weighted undirected graph, where V (see equation 3.4) is a set of vertices [81]. Each $v \in V(G)$ represents a transmitter in the FAP [81].

E (see equation 3.5) is a set of edges [81]. An edge consists of two vertices v_i and v_j that are joined because there is a constraint on the frequencies that can be assigned between the two vertices or transmitter [81]s. Each edge has two associated labels d_{ij} and p_{ij} [15, 82].

The label d_{ij} that is part of the set D (see equation 3.6) denotes the maximum separation that is required between frequencies assigned to two transmitters v_i and v_j . $f(i)$ denotes the frequency assigned to i . Using equation 3.9 the amount of interference that is generated between transmitters v_i and v_j can be determined [15, 82].

The other label, p_{ij} , forms part of the set P (see equation 3.7) which is referred to as the interference matrix² [31]. Each label p_{ij} contains two values which represent interference:

- p_{i0} represents the value for co-channel interference [15, 82].
- \bar{p}_{i1} represents the value for adjacent channel interference [15, 82].

²Discussed in section 3.5

Finally the set F (see equation 3.8) denotes a set of consecutive frequencies for every transmitter in V [15, 82].

Formally the FS—FAP can now be defined as a 5-tuple FS-FAP = (V, E, D, P, F) with a required mapping of $f : V \rightarrow F$ [82]. The objective of the FS-FAP is to find an assignment of frequencies to transmitters that minimise the sum of total interference (see equation 3.11). The collective interference value for a particular transmitter is represented by $c(p_i)$ which is based on equation 3.10. In equation 3.10 the co-channel interference value is represented by p_{i0} and the adjacent-channel interference value is represented by \bar{p}_{i1} .

$$c(p_i) = \begin{cases} p_{i0} & , \text{if } |f(i) - f(j)| = 0 \\ \bar{p}_{i1} & , \text{if } |f(i) - f(j)| \leq d_{ij} \\ 0 & , \text{if } |f(i) - f(j)| > d_{ij} \end{cases} \quad (3.10)$$

$$TotalInterference = \sum_{i=0}^{|P|} c(p_i), p \in P \quad (3.11)$$

In the following section a brief discussion on the different FAP benchmarks that exist is provided and the benchmark against which the algorithm developed in this research is evaluated is defined.

3.8 FAP Benchmarks

Some of the most used benchmarks in the FAP domain are now discussed. The first benchmark was introduced in the 1970s.

3.8.1 Philadelphia Benchmarks

The Philadelphia benchmarks are derived from an instance that was introduced in 1973 by Anderson [5]. Each instance is a hexagonal grid of cells that overlaps the area of interest. At the centre of each cell there is a transmitter. Past approaches used these hexagonal systems to model modern cellular networks [2, 71].

In this benchmark interference is measured by a co-channel reuse distance [2]. This distance stipulates that the difference between the frequencies

assigned to two cells must be greater than or equal to a certain value d . A channel cannot be assigned to a cell if it violates this minimum distance [71].

These benchmarks are typically used to test algorithms developed for MS-FAP, since there is no concept of cost or penalty for interference incurred by violating constraints.

3.8.2 CELAR

In 1994 European Cooperation on the Long term in the Defense (EUCLID) introduced a project called CALMA, which was a combined effort by several European governments that were part of EUCLID to investigate algorithms for military applications [2]. The project was granted to 6 research groups. Within the project 36 instances were made available by the *Centre d'Electronique de l'Armement* (CELAR) for radio link frequency assignment [2, 28].

All the CELAR instances have the constraint that the difference between frequencies assigned to interfering radio links must be greater than a certain predefined distance in the frequency domain [2]. This is a soft constraint and may be violated. Another constraint in the CELAR instances is that each pair of parallel links must differ by an exact predefined distance [2]. This constraint is a hard constraint and may not be violated [28].

These instances were initially not available to the general public as they were contained to be within the CALMA project [1]. In 2001 the CELAR launched the International ROADEF challenge, where certain instances from the CALMA project were made available for the research teams taking part in the challenge [1]. The instances made available had been modified to take polarisations and controlled relaxations of certain EMC constraints [49].

3.8.3 COST 259

The COST 259 is a set of real-world GSM instances made available by the European Union. The instances are publicly available and can be downloaded for free at <http://fap.zib.de/> (FAP Web 2011). The website also contains the most recent results obtained by researchers using these instances [2, 31].

The instances are difficult due to the large number of transmitters (900 - 4 000) that need to be assigned frequencies, with a relatively small num-

ber of spectrum of frequencies. The most important characteristic of these benchmarks are that they resemble real-world GSM network data. Due to these benchmarks' real-world applicability, they were selected as the main benchmarks to evaluate the algorithm presented in this dissertation.

More specifically this research concentrates on a small subset of the instances that are available, namely Siemens1, Siemens2, Siemens3 and Siemens4. In the paper by Montemanni and Smith [82] the same subset of problems was used and to date their algorithm has produced some of the best results. The characteristics of each instance will now be discussed.

As discussed in section 3.3 constraint handling mechanisms need to be used for problems like the FAP. Each of the Siemens benchmarks define a set of globally or locally blocked frequencies which are hard constraints. Due to this constraint one of the mechanisms that will be used in the algorithm to handle solutions violating this constraint is *to reject any solutions that is found to violate this hard constraint*.

Siemens1

The Siemens1 instance resembles a GSM network that follows the GSM900 standard. This particular network has been allocated a spectrum set of frequencies $F = \{16, 17, 18, \dots, 90\}$ which are allowed to be assigned to cells.

Not all 74 frequencies are available to be used by the network. The allocated frequency block is split into two blocks. According to the problem instance frequencies ranging from $F = \{36, 37, 38, \dots, 67\}$ are globally blocked. Only frequencies ranging from $F = \{16, 17, \dots, 35\}$ and $F = \{68, 69, 70, \dots, 90\}$ therefore available for assignment.

This problem instance finally also defines this network as consisting of a total of 506 cells where on average each cell has 1.84 transceivers that need to be assigned a frequency. The co-site separation is stated to be 2 and the co-cell separation is stated to be 3.

Siemens2

The Siemens2 problem instance describes a GSM network based on GSM900 and has 86 sites. The problem specifies that the network consists of 254 cells

where each cell has on average 3.85 transceivers that need to be assigned frequencies.

For this problem, the network has been allocated two blocks of frequencies: one block of 4 frequencies ranging from $F = \{42, 43, 44, 45, 46\}$ and a second block of frequencies ranging from $F = \{53, 54, 55, \dots, 124\}$. The frequencies allocated to the network have been split into two blocks because frequencies ranging from $F = \{47, 48, 49, 50, 51, 52\}$ are globally blocked. Finally the problem specifies that the co-site separation must be set to 2 and the co-cell separation must be set to 3.

Siemens3

The Siemens3 problem describes a network based on GSM900. This network has been allocated a continuous set of frequencies $F = \{681, 682, 683, \dots, 735\}$. Thus the network has 55 frequencies, which can be allocated to transceivers in its network.

The problem defines the network as consisting of 366 sites and 894 cells. On average each cell has 1.82 transceivers that need to be allocated a frequency to handle communication.

Siemens4

The Siemens4 instance is similar to a GSM network that follows the GSM900 standard. According to this instance this network has been allocated 39 continuous frequencies starting at 56, thus $F = \{56, 57, 58, \dots, 94\}$. No frequencies are said to be globally or locally blocked in this network.

According to this problem instance this network has 276 sites and consists of 760 cells. Where each cell is said to have on average 3.66 transceivers. The Co-site separation is set to be 2 and the co-cell separation must be 3.

In the next section a general overview is given on the different industries where the FAP is encountered. The purpose of this discussion is to get a better understanding about how far reaching the problem is and the different forms the problem is encountered in.

3.9 FAP in the Industry

FAP is a real-world problem that is encountered in industries that make use of wireless technology for wireless communication [2]. For each industry listed, a brief overview is given of how the problem differs compared with other industries.

3.9.1 Satellite Communication

The FAP in the satellite communication domain occurs in the ground terminals that transmit and receive signals via a satellite [2]. One would assume that the problem includes the satellite as it handles a variety and order of magnitude more connections than a ground station at any given moment. In fact, the problem is only concerned with the frequencies that the ground terminals use [2].

In Satellite communication, a signal is transmitted to one or more satellites via an uplink from a ground terminal [2]. Ground terminals communicate to the signal is received by the recipient satellite and relayed via downlink to those ground terminals which have expressed interest in certain signals from other ground terminals [2].

A large distance in the frequency domain separates the frequencies used by the ground terminals for uplink and downlink communication. The typical distance is much larger than the bandwidth. When frequencies are assigned to transmitters, downlink transmitters are ignored and only uplink transmitters are considered [2].

A radical difference with regard to the use of frequencies compared with the standard FAP in cellular networks is that frequencies are only allowed to be used once [2]. This is specific to the satellite domain to avoid interference [2].

3.9.2 Wireless Mesh Networks and Wireless Local Area Networks (WLANs)

Wireless mesh networks and WLANs³ are the most recent applications where the FAP is encountered.

³Both applications use the same standard and encounter similar problems in their respective domains.

Multiple WLANs are increasingly being used to provide backbone support for large fixed line networks, enterprise networks, campuses and metropolitan areas [111]. To be able to provide backbone support for these networks, a primary design goal when designing and deploying these networks is capacity [111]. A limiting factor for WLAN capacity is interference, which affects multihop hop settings. Thus the overall network interference needs to be minimized to increase the capacity of the network [111].

Typical approaches allocating frequencies include using DFA and FFA⁴. DFA is not very popular because the dynamic switching of frequencies lowers the response time on commodity hardware since there is a delay in milliseconds when switching frequencies. Typical packet transmission times are in microseconds. To guarantee uptime and high responsiveness, FFA is the preferred approach [111].

The FAP in wireless mesh networks and WLANs differs from the standard problem in that it introduces an extra constraint. Channels assigned to links on a node cannot be more than the available interfaces on that particular node. This constraint is known as the *interface constraint* [111]. Another aspect to consider is the placement of access points (APs) in the network, which is similar to the problem cellular networks face with regard to base station placement [2].

3.9.3 Military Field Communication

In a military context the FAP is a very difficult problem to be solved due to its dynamic nature [1]. During deployment, connections need to be established rapidly between nodes which guarantee that the nodes will stay static at locations. Usually nodes are military field phones or can be any transceiver device [1, 28].

Due to the nature of the problem the DFA scheme is used to allocate frequencies to nodes. The military FAP has the property that any of the nodes is mobile and can move at any moment to a new location, potentially interfering with another connection [1, 28]. This property differs from the traditional FAP where the nodes that are allocated frequencies are stationary [1]. Two frequencies need to be assigned to each connection that is established, one for each direction of communication. These allocated fre-

⁴Discussed in section 3.4

quencies must also differ by a certain distance in the frequency domain to prohibit alternating directions of communication from interfering [1, 28].

A lot of literature can be found on Military field communication. This is due to two organizations CELAR and EUCLID making data available to various research groups and allowing them to develop algorithms for frequency assignment [1, 28].

3.9.4 Television and Radio Broadcasting

The FAP encountered in broadcasting very closely resembles the problem domain found in cellular networks [2]. The only notable difference is that the required distance by which allocated frequencies must differ in the frequency domain are larger in broadcasting than in cellular networks [2].

Since the problem resembles the problem found in cellular networks, there are few articles that specifically discuss frequency assignment in broadcasting as a main topic. Research that specifically discusses FAP in broadcasting has been conducted by Idoumghar and Schott [52].

3.9.5 Cellular Communication

Cellular communication⁵ can be considered the main driving force behind research in the frequency assignment domain. As new standards are developed and used in 3G networks, in general an FAP still needs to be solved since these newer technologies still use GSM as their backbone architecture, as discussed in section 2.2.1. With new networks being deployed or current networks being expanded, standard GSM is used as it is cheaper than using the latest 3G technology. Therefore, standard GSM is still relevant and in use in modern networks.

There is a wealth of research that concentrates on the FAP within cellular networks. This is because cellular networks are used by millions of people around the world and as such this presents an interesting notion to produce better results since viable solutions have the possibility to impact millions of people. Most of the literature concentrates on this domain and one can find a lot of research in the literature presenting viable algorithms that produce real-world solutions [31].

⁵An overview of cellular communication technology called GSM is presented in Chapter 2.

Because the FAP problem is NP-Complete most presented algorithms are either of the metaheuristic type or more recently of the swarm intelligence type. Both of these algorithmic types are discussed in chapters 4 and 5 respectively.

3.10 Summary

In this chapter a discussion was presented of the problem this dissertation is based upon. The problem was defined as being the FAP and is categorised as being part of the set of NP-Complete problems. The NP-Complete nature of the problem is an important concept to understand.

Within the FAP domain there are two different techniques when assigning frequencies. The two different techniques were discussed in section 3.4.

An important concept that needs to be understood to comprehend why the FAP exists is the concept of interference. This was discussed in depth in section 3.5.

The FAP is not just one problem but consists of various subproblems that have different goals for the resulting frequency plan. Some problems are concerned with the number of frequencies used, others are more concerned with the amount of interference that is generated on the network because of the assignment.

The various FAP subproblems were outlined and discussed in section 3.6.

A formal mathematical definition of the MI-FAP was also presented. Finally, the chapter concluded with an overview of the various benchmarks problems which are used to test the viability of frequency assignment algorithms.

Chapter 4

Metaheuristic Algorithms

4.1 Introduction

Metaheuristics is a subdomain of the artificial intelligence (AI) domain [100]. It evolved out of a need for more efficient search techniques with regard to hard problems.

Metaheuristics forms part of a collective body of algorithms that use heuristics to search a particular domain's problem space for the most optimal solution [100, 132]. Some problems explicitly define constraints to which the produced solution must adhere to. Problems that define such constraints are referred to as constraint optimisation problems [33]. Problems where there are either no constraints or just a boundary constraint defined are referred to as unconstrained problems [33].

Constraints can be either hard or soft. A hard constraint is defined as a certain condition an algorithm or potential solution is not allowed to violate [2, 31, 100, 132]. A soft constraint is allowed to be violated but there is some sort of penalty or cost involved which is imposed onto the potential solution, which lowers its desirability [2, 31, 100, 132].

An optimal solution would therefore be any solution that violates no hard constraints and violates no or a minimum number of soft constraints [2, 31, 100, 132].

Algorithms that are classified as being part of the collective body of algorithms known as metaheuristic algorithms are Tabu Search [79, 87], Simulated Annealing [112, 121] and Genetic Algorithm [93, 126].

The above-mentioned algorithms are not the only algorithms to form part of this subdomain, but they are the algorithms that have received the most attention in the literature and produce good results [76].

The main focus of this chapter is on each of the above algorithms. Before each of the algorithms is discussed, a brief overview is given of the various characteristics that metaheuristic algorithms exhibit.

AI search algorithms operate in search spaces, where they occupy various states while they are searching for a solution or goal state in a search space. Before these algorithms can be properly introduced, these concepts need to be discussed briefly.

4.2 Search Spaces and States

A search space is defined as a set of candidate goal states which may or may not be a solution to a problem [100]. The candidate states are defined by the problem definition [100]. The possible states can given by a successor function which generates new states adhering to the problem defined constraints if any [100].

The search space can also be explicitly defined with constraints to which as solution must adhere to [100]. When constraints are defined it is up to the algorithm how it moves from one state to another in the search space [100]. Each state the algorithm moves to is then checked whether it adheres to the constraints [100].

A state is a position in the search space [100]. This position represents a configuration that describes a possible solution to the defined problem [100]. When a algorithm is said to be moving to a new state, the algorithm moves to a new position in the search space [100]. This new position represents a different solution than the previous position that was occupied [100].

When dealing with optimisation problem there can be more than one solution [100]. Which is why with optimisation problems not only does an algorithm have to find a solution, but also the most optimal solution among the candidate solutions [100]. This subset of solutions in the search space is referred to as the *solution* space.

Both concepts of search spaces and states have now been defined. All search algorithms operate in search spaces and occupy states. The domain

of search contains a wide variety of algorithms. In the next section a Metaheuristics Algorithms is discussed as this body of algorithms are used to solved optimisation problems.

4.3 Metaheuristics Algorithms

NP-Complete¹ problems have been proven to not be solvable in polynomial time by traditional uninformed search methods such as Breadth First Search and Depth-First Search [100].

Uninformed algorithms are not able to distinguish whether a particular goal state is more “correct” than any other goal state [100].

Uninformed search algorithms search spaces are represented by a tree structure [100]. A search is performed in the search space by starting at a root node, the algorithm expands successive nodes based on a strategy [100]. Bread-first search continuously expands all successive nodes and Depth-First search expands the deepest node first [100].

The path taken to a final goal state node represents a solutioncite [100]. In a worst-case scenario, an uninformed search will expand every single node in its search space, thus each and every single possible solution is evaluated [100].

It is not always viable to test every possible solution in a given problem search space, especially in NP-Complete problems, since their search spaces are usually huge or infinite. This is why traditional algorithms are not able to produce optimal solutions in polynomial time [100].

With NP-Complete problems, the path taken from an initial possible solution to a final optimal solution is irrelevant [100]. Only the final optimal solution is needed. Algorithms that disregard the path taken to a solution and that only care about the final solution are classified as *local search* algorithms [100].

Local search algorithms tend to use a small amount of memory since the algorithm is only concerned with the current state and only moves to a neighbouring state when searching for a potential solution. This makes local search algorithms a good fit for optimisation problems as an optimised

¹A discussion of NP-Complete problems is appears in section 3.2

solution is just a solution which constitutes to the current best state found by the algorithm [48, 100].

In its most basic form, a local search algorithm is just an algorithm that moves from one state to another. How it generates neighbouring states to move to and how and why it moves to a certain state makes the local search algorithm unique [34, 48, 100].

Metaheuristic algorithms are considered to be *general-purpose* algorithms and can thus be applied to a wide variety of optimisation problems with only small modifications that need to be made to the algorithm model [69].

Metaheuristic algorithms do not dictate all aspects of the search procedure but define guidelines and aspects which direct the search [89]. Therefore a metaheuristic can be defined as follows: *the algorithm selects candidate solutions from a neighbourhood with respect to one or more current solutions of which the candidates are either rejected or accepted* [89].

As can be gathered from the name, a metaheuristic algorithm uses some sort of heuristic. A heuristic is a decision rule. Algorithms use heuristics to make decisions by applying the heuristic to the data the algorithm is currently working on [100, 132].

When an algorithm utilises a heuristic it forms part of a group known as informed search methods. As the name implies the algorithm is more “informed” using the heuristic the algorithm gains additional information about its search space. With this additional information the algorithm is able to perform a better search for a possible solution [100].

A heuristic is able to dictate what an algorithm must do for its next iteration by evaluating the current internal state of the algorithm, i.e. whether it should move to a different point in the solution space, generate new data, or select the current data as the most optimal solution [100, 132].

In general “meta” means *beyond* or *higher level* [100, 132]. A metaheuristic therefore refers to a heuristic that is more complex with regard to the decisions it is able to make compared with a standard heuristic [100, 132]. With a standard heuristic only the current state is considered [100]. A metaheuristic takes additional information into account when considering a current state [132]. This additional information can be previous states, or states that are considered to be close in the search space [100, 132].

Metaheuristic algorithms do not search the solution space statically by

testing and evaluating every possible permutation in the solution space. This is inefficient and a normal path-based algorithm can be used instead [9].

Metaheuristic algorithms make use of strategies and heuristics (specific to the problem domain) to search the solution space intelligently [9]. Heuristics used by the algorithms take more information into account based on the current state of the algorithm which allows the algorithm to make a more informed decision to move and occupy a solution in the search space [100]. Strategies used by these algorithms can include recording good/bad solutions, introducing new search spaces or even adding additional rewarding states [34, 100, 132].

As these algorithms explore the search space, neighbouring states are selected non-deterministic making their search procedures of a stochastic nature [89, 132]. Heuristics are used to guide the search to move to more desirable states in the search space where there is a higher probability of obtaining high quality candidate solutions [76, 82].

Metaheuristic-based search methods are not guaranteed to find the most optimal solutions in the solution space; instead these methods are usually used to find near-optimal solutions. Thus most algorithmic development in the metaheuristic domain focuses on developing new techniques that will increase the probability that a good solution will be obtained in difficult combinatorial problems [9].

Similarly, metaheuristics is not guaranteed to find suitable solutions or perform well in each problem domain it is applied to. The quality of the solution and performance of the metaheuristic is very much dependent upon on the expertise of the algorithm designer [130].

Although heuristics plays a key role in the performance of metaheuristic algorithms, it is not the only factor that has an impact on performance and results. Algorithms also use techniques and concepts from other system paradigms like multi-agent systems [75].

In multi-agent systems, multiple agents have to communicate with each other and the system as a whole has to perform some sort of autonomous self-organisation [75].

These social and self-organisation concepts enable these systems to be distributed, robust and flexible. This is why in metaheuristic algorithms that are population-based, hybrid and/or distributed, these concepts of social and self-organisation are used to better exploit the solution space [75].

In this section the characteristics of metaheuristics that set these algorithms apart from the conventional algorithms used on difficult problems was introduced. In the next section of this chapter the tabu search algorithm is discussed.

4.4 Tabu Search

4.4.1 Introduction

TS was first proposed by Glover [39] as a new searching technique to help algorithms avoid getting trapped in local optima in combinatorial and optimisation problems [95]. Since Glover introduced the algorithm in the 1980s, tabu search has been applied to a wide range of problems such as the vehicle routing problem [79], FAP [82], capacitated-lot sizing problem [43], Nurse Scheduling [27] and the Resource Constrained Assignment Problem [95].

Even though the problems mentioned differ by a large margin, the algorithm has been successful in most optimisation problems it has been applied to. If one observes the results obtained in research [82, 109], it can be inferred that tabu search has on average obtained the best results compared with previous attempts with other algorithms.

TS resembles in its most basic form the hill-climbing search algorithm [113]. The hill-climbing search algorithm starts from an initial solution and then iteratively moves from the current solution to a neighbouring solution [100]. Each neighbour is rated based on its attractiveness as a possible optimal solution that the algorithm is being applied to [100].

The hill-climbing algorithm moves to the neighbour with the highest rating without considering whether the neighbour might lead the algorithm astray, to a position where the neighbours are in fact *worse* than previously encountered possible solutions [100].

The TS algorithm addresses this shortcoming by introducing the concept of memory [113]. The memory of the algorithm is actually a history of previous solutions that the algorithm has moved to in its search for the problem space for a solution [113].

General search algorithms like hill-climbing, random-restart² or scatter

²Random-restart is a search algorithm where once a certain trend of repeated moves is noticed, the algorithm restarts by generating a new initial solution to start from and then

search tend to get trapped in local optima [100]. The local optima might be a very attractive solution and thus general search algorithms will not move to better solutions since, according to the algorithm's built-in strategy, it has found the best solution.

In actual fact the solution that was found is the best solution in the *local* search space, but not in the *global* search space [34, 100]. Therefore an important characteristic of algorithms being applied to optimisation problems is breaking out of local optima [34, 100].

The next section discussed some of the characteristics that make the TS algorithm unique.

4.4.2 Important Tabu Search Characteristics

Various characteristics are important to the TS algorithm. The first characteristic is exactly how the TS algorithm iteratively improves upon the initial start solution.

Initial Solution Generation

The core feature of the TS algorithm is to sequentially improve an initial solution [135]. An initial possible solution is a point in the problem space where the TS algorithm will *start* exploring in search of a more optimal solution [100, 135].

An important consideration one has to make is how initial solutions are generated for the TS algorithm to start on [100, 135].

Random initial solutions might seem to be a good starting point, but by introducing randomisation it becomes hard to control the quality of the end solution [135]. Hence the generation of starting solutions must be controlled to limit the infeasibility of potential solutions [135].

Control of the randomly generation solutions can be achieved by simply constraining the random solution generator to only generate initial starting points in a bounded subset of the entire search space. For example: instead of letting the random initial starting point be any number between positive infinity and negative infinity, the random number generator is constrained to only generate numbers between 5 and -5.

continues its search process from that generated solution [100].

Neighbourhood Search

The following discussion on neighbourhood search is not meant to be an exhaustive survey on the different methods and how they differ under different problems. Instead the discussion is meant as a general overview to get an idea of neighborhood generation in the TS algorithm context.

The following neighbourhood discussion will be based on the assumption that the underlying problem the TS is applied to, has a search space with defined boundaries that is suitable for neighborhood generation.

TS uses a neighbourhood local search process to explore the solution space. There is no set process of how neighbourhood candidate solutions are selected as it is problem dependant. The overall quality of the solution produced by TS is also dependent on the neighbourhood search strategy used [135].

The neighbourhood search phase is the first operation performed after the algorithm has been initialised, which is to say the algorithm has generated an initial starting solution from which the exploration process can start.

The neighbourhood search phase is the primary means for the TS algorithm to search the solution space for an optimal solution. It is within this phase that new possible solutions must be presented for the TS heuristic to allow the algorithm to decide to which solution it must move next.

The new possible solutions that are generated are called neighbouring solutions; hence the TS algorithm always moves to a neighbouring solution. When the TS algorithm moves to a neighbouring solution, the current solution is replaced by the neighbouring solution. Therefore, in the next iteration, neighbours for the new solution need to be generated.

Generation of new neighbours can range from a simple increment option to a complex operation that incorporates additional intelligence by means of a more heuristic approach to generate new neighbours.

The TS algorithm is not limited to just one neighbourhood search strategy. In the paper by Gopalakrishnan et al. [43] five neighbourhood move strategies are developed and are used interchangeably; in some cases a strategy is used three times in a row due to stagnation in the search space.

Stagnation occurs when the algorithm does not move to a better solution; instead it opts to stay on the current solution, as no neighbouring solution

is better than the current one.

Other neighbourhood strategies developed is that by N. A. Wassan [125]. Wassan used a neighbourhood selection strategy that exchanges route nodes from initial vehicle routes for the vehicle routing problem. This route exchange enables the TS algorithm to search much more broadly due to the constant supply of different solutions.

Since initial solutions are constantly modified, it enables the TS procedure to be a very fined-grained process, because often a small change in a potential solution can have a big impact on the overall proposed solution by the TS algorithm.

In the research done by Zhang et. al. [135] a neighbourhood selection scheme called *dynamic penalty* is developed. When the algorithm moves onto an infeasible solution a penalty is imposed. By dynamically changing the penalty that is imposed the “feasibility” of solutions produced is influenced.

Therefore, if and when the algorithm continually produces infeasible solutions, the penalty imposed is increased to guide the algorithm to produce more feasible solutions. Finally, when the algorithm becomes trapped at local optima, the penalty is reduced, which allows the algorithm to consider moving onto infeasible solutions thus escaping local optima.

Considering all the research done to develop new neighbourhood selection strategies that improve TS to search the solution space more efficiently and produce better solutions, TS still has some drawbacks, especially with problems that have very large solution spaces [8].

TS is an iterative algorithm, executing a set of operations sequentially until a stopping criterion is met [8, 79]. At each iteration the algorithm has to determine feasibility of the immediate neighbourhood candidate solutions [8, 79].

Therefore each candidate must be evaluated by some function, which may be a costly operation in terms of computational cycles as well as in terms of time [8, 79]. This constant evaluation can drastically reduce the overall performance of the algorithm, since it is spending more time calculating feasibility than actually searching the solution space [8, 79].

Memory Structures of Tabu Search

The hill-climbing and random-restart algorithms are able to break out of local minima, but there is nothing stopping these algorithms from avoiding the local optima with their second or n-pass in the search space. TS addresses the shortcoming of these algorithms by incorporating an important concept: the notion of memory.

In its most basic form TS keeps a local memory of all its recent best moves, and puts them into a *tabu list* that has a predefined size. In the literature the Tabu list is also referred to as the *tabu tenure* [43,56,125,135]. The algorithm is not allowed to move to any solution that is in the tabu list unless a solution that is *tabu* is better than any current moves available in the immediate search neighbourhood [43, 56, 125, 135]. The process of overriding a solution's tabu status in the tabu tenure is called the *aspiration criterion* [43,56,125,135]. With the use of the tabu tenure and the aspiration criterion, the algorithm is able to avoid cycling, local optima as well as searching in a too narrow region [6,88].

Research done by Ashish Sureka and Peter R. Wurman makes an important distinction with regard to the memory scheme that is used in the TS algorithm. Two memory schemes are discussed: *explicit memory* and *attribute-based memory* [113,114]. Of the two memory schemes the explicit memory scheme is the most used in the literature [79].

With explicit memory the algorithm stores a complete solution in the tabu tenure; hence the algorithm is prohibited from moving to that position in the solution for as long as the solution is in the Tabu tenure [113,114]. With attribute-based memory the algorithm stores the *operation* used to move from the previous solution to the current solution [113,114]. Therefore with attribute-based memory the tabu tenure intended function is changed from prohibiting certain solutions already encountered to rather prohibiting making changes to the current solution that would lead to solutions already present in the tabu tenure [113,114].

In research conducted by D. M. Jaeggi, G.T. Parks, T. Kipouros and P. J. Clarkson [54], the authors add two additional memory structures called medium term memory (MTM) and long term memory (LTM) besides the standard short term memory (STM), typically referred to as the tabu List [54]. Each additional structure remembers a different set of solutions for use

by the diversification and intensification phases in the algorithm.

STM is similar to the traditional tabu list: to store the most recent solutions produced by the algorithm. MTM is designed to remember optimal or near-optimal solutions. These solutions are therefore used later in the intensification phase. Finally, the LTM structure stores all the solutions that the algorithm has already explored and is thus used in the diversification phase of the algorithm [54].

Search Phases

As TS searches through the search space, it goes through two cycles of search phases called *diversification* and *intensification* [17, 36, 48, 56].

The diversification phase in the TS algorithm is the phase where the algorithm is directed to areas in the search space that has not yet been explored. The algorithm usually applies diversification as mechanisms monitoring the memory; note that solutions being produced are being repeated [36, 125].

Research by Fescioglu-Unver and Kokar [36] provides a strategy that consists of two components namely the *observer* and the *diversifier*. The goal of the observer is to continually monitor the best solution obtained by the algorithm as to whether it violates the *stagnation period*. The stagnation period is defined as the number of iterations where the current best obtained solution has not changed and the algorithm has not moved to a new solution [36].

As soon as the current solution exceeds the stagnation period the observer component activates and transfers the necessary information needed by the diversifier component. The diversifier component dynamically changes the size of the tabu tenure based on the information the observer gathered. The diversifier mainly targets older moves to diversify, but for short bursts of time it decreases the tabu list size to a very small value in an attempt to combine new and old moves [36].

The specific mechanism used to define a new position where the algorithm can continue to search, should ideally select areas in the search space that have not been explored yet. Therefore, the diversification phase typically makes extensive use of the knowledge present in the long-term memory structures as an indication of what areas of the search space have been previously explored and which areas have not [17, 36, 48, 56].

Intensification is usually the first phase of the TS algorithm, since it is responsible for building up a history in memory on which the diversification phase can act. Fescioglu-Unver and Kokar also present an intensification strategy based on control theory in their research [36]. The authors identify repetition length as a critical value for their intensification strategy to be based upon. The repetition length is a control measure that defines how many times the algorithm can occupy the same solution within a span of iterations.

In the following section, an overview is presented of the flow the TS algorithm along with pseudo code describing the TS algorithm.

4.4.3 Flow of the algorithm

In this section the general flow of the TS algorithm is described using the algorithm 1 presented in pseudo code as a reference point.

Algorithm 1 Basic Tabu Search Algorithm [82,95]

```

1: Initialize parameters
2:  $x_0 \leftarrow$  Initialize starting solution
3: while stopping criteria not met do
4:    $y_i \leftarrow$  Determine  $x_i$  neighbourhood solutions
5:   Evaluate neighbouring solutions with fitness function  $f(y_i)$ 
6:    $z_i \leftarrow$  Select best neighbour from  $y_i$ 
7:   if Move to  $z_i$  is Tabu then
8:     if  $z_i$  meets Aspiration Criterion then
9:        $x_i \leftarrow z_i$ 
10:    end if

```

Algorithm 2 Basic Tabu Search Algorithm (continued) [82, 95]

```

11:   else
12:       Add  $x_i$  to Tabu List
13:        $x_i \leftarrow z_i$ 
14:       if  $x_i$  repeated  $\geq$  max repeats then
15:           diversify()
16:       else
17:           intensify()
18:       end if
19:   end if
20: end while
21: Return  $x_i$  as best found solution

```

Before the algorithm can actually start searching, it first needs to initialise various parameters. These parameters include, but are not limited to, the tabu list size, the aspiration criterion and the starting solution. The initialisation can be observed to occur from lines 1 - 2.

Once all the various parameters that are needed by the algorithm have been initialised, the algorithm is ready to enter the actual search phase, which ranges from lines 3 - 21.

The search phase starts off by first generating possible solutions that neighbour the current solution x_i as can be observed in line 4. Generating neighbouring solutions are a critical process in the TS algorithm as they are the means by which the algorithm is able to move from one possible solution to the next in the search space.

After all the solutions that neighbour the current possible solution have been generated, the algorithm needs to decide which of the possible neighbours is the most rewarding. The algorithm therefore determines the fitness of each neighbour y_i by applying a fitness function $f(y_i)$.

Once all the neighbours have been evaluated, the algorithm selects the best neighbour that not only has the best fitness out of all the generated neighbours, but also has a better fitness than the current solution held by the algorithm. The best neighbour selection can be seen to occur in line 6.

The algorithm has now determined a possible neighbour z_i to move towards. Before moving on to the next iteration, it first needs to perform a series of checks that will aid it in the search process.

The first check that needs to be performed is whether the neighbour z_i is in the tabu list and this occurs in line 7.

If the neighbour z_i is in the tabu list, then another check is performed where the aspiration criterion is calculated as can be seen in lines 8 – 10. The aspiration criterion determines whether the algorithm can make neighbour z_i its current solution once more even though it is tabu.

If the aspiration criterion has been met, the algorithm makes neighbour z_i its current solution x_i .

In the algorithm, if a neighbour z_i is found not to be in the tabu list, the algorithm then adds the currently held solution x_i to the tabu list. The current solution is added to prohibit future movements to the same solution in an attempt to avoid cycling of solutions. After x_i has been added to the tabu list, the algorithm makes z_i the current solution x_i . This process can be observed from lines 12 – 14.

Before the algorithm continues to the next iteration, it performs one last final check. The purpose of this check is to determine whether the algorithm is repeating solutions. As can be observed from lines 15 – 19, the algorithm calculates whether the new selected solution has been repeated for a certain number of iterations.

If the solution has indeed been repeated for a predetermined number of iterations, the algorithm activates its diversification strategy or intensifies its search.

The section that follows presents an overview of the TS being applied to the FAP.

4.4.4 Tabu Search on the FAP

In a study conducted by Robert Montemanni and Derek Smith [82] the TS algorithm is used on the FS-FAP. The authors had to make some alterations to the algorithm to suit their needs as well as to make the algorithm more efficient in exploring in the FAP solution space.

The TS algorithm used by the authors is the multistart TS algorithm, which randomly starts on different initial solutions [82].

The authors developed a technique called HMT. HMT first monitors an underlying heuristic being used on the problem by the algorithm [82]. It then identifies certain characteristics that good solutions exhibit. In the

FAP, it is transmitters that are assigned different frequencies which results in an overall lower interference value [82].

The HMT then uses the identified characteristics to add *additional* constraints to the problem [82]. By adding constraints, the search space is reduced. However, by reducing the search space, other near-optimal solutions which might be far better are excluded [82]. It is for this reason that the authors opted not to add the constraints permanently.

The authors applied their TS algorithm together with HMT to the COST 259 family of benchmarks, specifically the Siemens1, Siemens2, Siemens3 and Siemens4 problems. The results are presented in table 4.1. The values presented are scalar and indicate the total interference of the frequency plan. The lower the interference value is, the better the frequency plan.

Problem instance	TS with HMT	Best COST 259
Siemens1	2.7692	2.200
Siemens2	14.9360	14.280
Siemens3	6.6496	5.19
Siemens4	110.9725	81.89

Table 4.1: Results of applying TS with HMT on COST 259

As can be observed from the results obtained by the authors, the TS algorithm with HMT produces results that rank very favourably against other algorithms also applied to the COST 259.

When critically reasoning about the TS algorithm with regard to applying it to the FAP, the following disadvantages in theory can be identified:

Search based on a single solution — The TS algorithm at any moment in time only searches in the vicinity of *one* current solution for possible neighbours that might be the current solution for the next iteration. FAPs typically have huge search spaces due to their NP-Complete nature. Therefore, only searching for possible rewarding neighbours from only potential solutions seems to be terribly inefficient. A better strategy would be to use the notion of population-based algorithms and have multiple solutions from which more rewarding neighbours are searched.

Neighbourhood generation — The TS algorithm defines no set process for generating a neighbouring solution given a starting solution. Generating neighbours from a solution is a critical process in the TS algorithm, for it is the only means by which the algorithm considers other solutions, i.e. it is the mechanism by which the algorithm searches. Generating a new neighbour can be as simple as changing only one value from the current solution or it can be very complex and incorporate other algorithms together with mathematics formulae. Regardless of the complexity of the neighbour generation that is used, care must be taken to ensure that the algorithm is able to produce a wide diversity of neighbours and is also able to intensify on the most optimal solution.

Tabu lifetime — The TS algorithm only operates on a single solution at a time and at most only considers one potential neighbour as its next possible current solution. Therefore a difficult choice needs to be made as to how long a solution stays tabu. In the FAP a solution might be entered into the tabu list early in the search process of the algorithm. A large majority of the neighbours of this solution are vastly superior solutions compared with any of the current solutions produced by the algorithm. Due to the solution with these neighbours being in the tabu list, these neighbours will not be reconsidered until much later when the solution is removed from the list. A possible option to allow the TS to reconsider the tabu solutions is to increase the aspiration criterion. Increasing the aspiration criterion does have its risks. A high aspiration criterion and the algorithm might be too eager to select just any solution even though it is Tabu. A low aspiration criterion and the algorithm will be too strict in selecting a tabu solution.

In the next section the simulated annealing algorithm is discussed.

4.5 Simulated Annealing

4.5.1 Introduction

simulated annealing (SA) is a metaheuristic search technique proposed in the 1980s by Kirkpatrick to solve combinatorial optimisation problems. The technique is based on a natural process which is known in metallurgy as annealing [62, 86, 112, 121]. Kirkpatrick was the first to use SA to solve

optimisations problems but the basic algorithm structure was defined by Metropolis et al. in 1953 [85, 121].

Annealing is the natural process of crystallisation when a solid is heated to a high temperature and then systematically cooled to a lower temperature to reach a crystallised form [4, 65, 77, 121]. The crystallised form of the solid is known to be the global minimum of the solid's internal energy state.

When the solid is rapidly cooled from a high temperature, the molecules have no time to reach a thermodynamic equilibrium stage [4, 65]. Therefore the molecules of the solid have high energy and the resultant structure has no real crystalline form; thus the solid energy is at a local minimum [65, 77, 121]. When the solid is slowly cooled in a controlled manner, the molecules are able to reach a thermal equilibrium at each temperature [77, 86, 121].

In the algorithm the energy state is the *cost function* that needs to be minimised, and the molecules are the *variables*, which represent the solutions, and thus their state needs to be optimised to reach the desired energy state.

The SA algorithm is able to purposely move to a worse solution in the search space [129]. In the research when an algorithm moves to a worse solution it is classified as an *uphill* move [100]. Similarly, when a move is made by the algorithm which results in a better solution the move is classified as an *downhill* move [100].

The following equation is the standard probability function that is used to determine when an uphill move is performed by the algorithm. This function is known in the literature as the *metropolis criterion*.

$$M_{AC} = \begin{cases} 1, & \text{if } f(y) \leq f(x) \\ e^{-\frac{\Delta E}{T_k}}, & \text{otherwise} \end{cases} \quad (4.1)$$

The function f is the objective function or a function that determines the state of a given position in solution space [129]. The parameter T_k is the temperature of the algorithm at iteration k [129]. Finally, ΔE is the change in “energy” between two solutions x and y [129].

The main purpose of the SA algorithm (like most optimisation algorithms) is to minimise or maximise the cost function [112]. This cost function typically evaluates a solution desirability compared with other solutions in the immediate *neighbourhood* of the algorithm's current position [22].

The immediate neighbourhood of solutions is generated based on some heuristic implemented by the algorithm designer [100]. This heuristic, as with the TS algorithm, can be simple or complex. The neighbourhood generated is the first step right after the starting solution has been generated by the algorithm.

Typically a neighbouring solution is only selected as the new best state if its desirability ranks higher than the current solution. When the algorithm moves to a better solution from the previous solution, the move is typically referred in the literature as a *downhill* move [121].

The best state is not always selected; in some cases the algorithm is also able to move to solutions that are worse than the current solution. A worse solution is only selected based on some probability which is controlled by the *annealing temperature* of the algorithm [22].

At a high annealing temperature the probability that the algorithm will select a bad solution is very good. As the annealing temperature decreases so does the probability that a bad solution will be selected [121]. When the algorithm moves to a worse solution, the move is typically referred to in the literature as an *uphill* move [121]. Uphill moves allow the algorithm to break out of local minima and can lead the algorithm down a different path, which may ultimately result in obtaining the global optimum [112].

As with the TS algorithm, the standard SA algorithm does not define a set neighbourhood generation mechanism; instead it is up to the algorithm designer to implement a suitable generation mechanism that will allow the algorithm to adequately explore the search space [94].

The following section presents a discussion on various characteristics of the SA algorithm.

4.5.2 Important Simulated Annealing Characteristics

There are four characteristics of the SA algorithm that make the algorithm unique. One of the most important is the cooling schedule.

Cooling Schedule

The cooling schedule/annealing Schedule is the most defining characteristic of the SA algorithm. It is the procedure where the natural annealing process

is mimicked. The temperature of the SA algorithm is a control parameter that defines how much the algorithm moves around in the solution space.

After each iteration, whether the algorithm has selected a new best solution or not, the temperature is reduced by a certain amount. This amount is determined by the cooling schedule.

In general, when the SA algorithm temperature has a very high value most solutions that are produced from the neighbourhood are accepted [70]. Thus the algorithm moves freely in the solution space with little constraint. As the temperature decreases, the probability that the algorithm will select a bad or just any solution decreases. When the temperature is very low, the SA algorithm is similar to a greedy algorithm in the sense that it only accepts downhill movements [70].

The cooling schedule provides the SA algorithm with the critical ability to control the rate the algorithm transitions from the diversification phase (high temperature) to the intensification phase (low temperature).

In the literature there are three annealing schedules in common use, namely *the logarithmic schedule*, *the geometric schedule* and *the Cauchy schedule* [85, 112].

The standard and most commonly used schedule is known as the logarithmic schedule and is based on Boltzmann annealing [85]. The main disadvantage of this schedule is that is slow due to its logarithmic nature [85]. It also requires moves to be generated from a Gaussian distribution for it to be able to reach the global minimum [112]. The logarithmic annealing function has the following form:

$$T_k = \frac{T_0}{\ln(k)}, \text{ where } k \text{ is the iteration value and } k \neq 0 \quad (4.2)$$

Where T_k is the temperature at iteration k .

The Cauchy schedule is faster than the logarithmic schedule. Similar to the logarithmic, this schedule also has a movement requirement. Moves must be generated from a Cauchy distribution for the algorithm to be able to reach the global minimum [85, 112]. The Cauchy schedule is also typically referred to as fast annealing [85]. The schedule has the following form:

$$T_k = \frac{T_0}{k}, k \neq 0 \quad (4.3)$$

Finally, the fastest annealing schedule is known as the geometric or exponential annealing schedule [112]. The schedule has no move generation

requirement to reach the global minimum, since there is no rigorous proof in the literature [112]. By using the geometric schedule the SA temperatures are rescaled which is called *re-annealing* [85].

The geometric schedule has the following form:

$$T(k) = T_0 e^{-Ck}, \text{ where } C \text{ is a constant} \quad (4.4)$$

Initial Temperature

The initial temperature is a very important parameter to define in the SA algorithm, since it defines a point from which the cooling schedule will start [94]. Therefore, depending on what the initial value of the temperature, is the final result that the algorithm will produce can be influenced [107, 127].

When the initial temperature is set to a very high value, the algorithm takes a long time to reach a result since the search space is being explored more [94, 127]. More exploration is favourable for SA as it lets the algorithm be less susceptible to local minimum.

If the initial temperature is set to a very low temperature, the algorithm might converge too quickly and thus produce a result which may be the local minimum [94, 107, 127].

The initial temperature together with the cooling factor allows the algorithm designer to define the time window for the algorithm to escape local minima, as well as the rate of convergence to an optimum solution [94, 127].

A low initial temperature together with a low cooling factor makes the time window for the algorithm to leave a local optimum very small [127]. With a high initial temperature and cooling factor value that is almost 1, the time window for the algorithm to leave the local optimum is much larger [127].

When the algorithm is near a global optimum, a low initial temperature and low cooling factor will allow the algorithm to reach the optimum faster in the solution space [127]. In contrast, if a high temperature and a very low cooling factor are used, the algorithm will take longer to reach the optimum even though it is near the global optimum [127].

Move Generation

Most of the research done on the SA algorithm focuses on the annealing schedule and not so much on the move/solution/neighbourhood generation.

Typically an initial solution is generated and then small changes are made to the solution to represent a new solution. The solution is said to be perturbed to the next solution.

Move generation is the phase where neighbouring solutions to the current solution are generated. It is the ideal section for an algorithm designer to embed domain-specific knowledge which will allow the algorithm to generate better possible solutions.

In research done by Tseung and Lin [121] an initial solution is not modified, but a move generation technique known as *pattern* search is used. Pattern search has two forms of movement, namely the exploratory move and the patten move. The exploratory move continually changes the certain variables of a solution [121]. This is done so that the technique can rapidly find and identify a “downhill” move. The pattern move uses the information gathered by the exploratory move to move towards the minimum of the function [121].

Algorithm Efficiency

The algorithm is also efficient with regard to CPU cycles when compared with the genetic algorithm. SA only has to evaluate a certain number of moves each iteration, instead of a whole population each iteration. The genetic algorithm is discussed in section 4.6.

Unlike TS, the basic SA algorithm does not keep any memory and is therefore memory efficient, but in contrast suffers the risk that the solution will cycle. The more iterations spent at a temperature, the longer the algorithm spends at a certain temperature and therefore the higher the probability that solutions may cycle.

In next section the flow of the SA algorithm is discussed and pseudo code for the SA algortihm is presented.

4.5.3 Flow of the Algorithm

In an attempt to better understand how the SA algorithm operates, a general discussion on the flow of the algorithm will now be given using algorithm 3 as a reference point.

Algorithm 3 Basic Simulated Annealing Algorithm [85, 86]

```

1: Initialize parameters
2: Set starting temperature  $T(0)$ 
3:  $x_0 \leftarrow$  Generate initial starting solution
4: while Stopping criterion not met do
5:    $y_i \leftarrow$  Generate neighbouring solutions to  $x_i$ 
6:   Evaluate  $y_i$  neighbours with fitness function  $f(y_i)$ 
7:   Calculate probability  $p_i$  of  $y_i$  neighbours with equation 4.1
8:    $x_i \leftarrow$  Select  $y_i$  neighbour based on probability  $p_i$ 
9:   Reduce temperature  $T(i)$  based on cooling schedule
10: end while
11: Return best solution  $x_i$ 

```

From lines 1 – 3, the SA algorithm is initialised. The most important step here is setting the starting temperature for the annealing process to start. As mentioned in the introduction, the temperature of the annealing process plays a critical role in the potential solution selection process.

After the algorithm has been initialised the search phase of the algorithm starts which ranges from lines 4 – 10. Like the TS algorithm, the SA algorithm starts the search phase by generating a number of neighbours to the current solution held by the algorithm as can be observed in line 5.

Before selecting a neighbour the algorithm first needs to evaluate the generated neighbours. It evaluates each neighbour by applying a fitness function $f(y_i)$ in order to determine its fitness. Once the fitness of all the generated neighbours has been determined, the algorithm uses equation 4.1 to calculate the probability of selecting a particular neighbour for all the generated neighbours as well. The probability calculation can be observed to occur in line 7. The algorithm then selects the neighbour with the highest probability to be the current solution, as observed in line 9.

Before the algorithm advances to the next iteration the temperature needs to be lowered. As discussed, the temperature is lowered according to a particular cooling schedule. In the algorithm the process of lowering the temperature occurs in line 10.

In the section that follows a discussion is given on when SA is applied to the FAP.

4.5.4 Simulated Annealing on the FAP

The SA algorithm, as with the TS algorithm, has achieved good results in other optimisation problems, as mentioned in section 4.5.1. Due to its success on other NP-Complete optimisation problems, the SA algorithm has also been applied to the FAP.

In literature by Carlo Mannino and Gianpaolo Oriolo [72] the SA algorithm is applied to the FAP. The resulting SA algorithm was benchmarked on the COST 259 Siemens benchmark instances. The results obtained by the authors is presented in table 4.2. Values represent the total interference generated by the frequency assignment.

Problem instance	SA	COST 259 (old)	COST 259 (new)
Siemens1	22.96	23.00	2.200
Siemens2	14.72	14.75	14.280
Siemens3	52.43	52.55	5.19
Siemens4	80.96	80.80	81.89

Table 4.2: SA on COST 259 Benchmark

Note that at the time the above results were obtained the best results obtained for the siemens problem were different from the latest results.

An overview of the SA algorithm along with its unique characteristics have now been given. Utilising the knowledge that was gained from understanding how the SA algorithm operates a critical evaluation can be presented.

In the following paragraphs, a theoretical critical evaluation is presented if the SA algorithm would be applied to the FAP.

Cooling Schedule — Depending on the cooling schedule selected, the algorithm might converge too quickly. As discussed previously the cooling schedule reduces the temperature. The temperature plays a large part in the determination of whether a particular solution will be moved to or not in an iteration. Thus early on the algorithm will explore a lot more (diversification) and later on will exploit more (intensification). In the FAP, the algorithm must not only be able to explore and exploit, but also be able to return to an exploration phase if need be. As the temperature becomes

colder the SA algorithm exploits more and therefore will not easily move to a worse off solution. In the FAP, it might be desirable to rather move a worse off solution later on, as the particular current solution is a local minimum and yields bad neighbours as potential next solutions. With the cooling schedule this is simply not possible, unless the temperature and schedule are reset. Resetting the temperature and schedule is not ideal, since the algorithm keeps no history and might risk making the same faults as before the reset.

Neighbourhood generation — The SA algorithm, as with the TS algorithm, has no set process that defines how neighbours should be generated. As discussed, neighbourhood generation is the primary means by which the SA algorithm moves about the search space in search of an optimal solution. Therefore applying the SA algorithm would require a custom neighbourhood generation scheme. A desirable neighbourhood generation scheme would be one that keeps track of where the algorithm has been previously. By keeping history, the algorithm will be able to avoid previously explored areas in the search space.

Single solution based search — The SA algorithm is similar to the TS algorithm in the sense that it only searches from one solution per iteration. It searches by generating neighbours around the current solution of the algorithm. The FAP search space is huge; hence it would be more efficient to have multiple current solutions from which neighbours are generated. This enables the algorithm to explore the search space much more efficiently at the expense of more computational resources.

4.6 Genetic Algorithm

4.6.1 Introduction

The genetic algorithm (GA) is a stochastic search method that is based on the natural process of genetic evolution and the Darwinian concept of “survival of the fittest” [38,46,61,122]. The genetic algorithm (GA) was first proposed by Fraser, but it was not till the research presented by Holland that GA’s became popular [34]. Holland initially applied the algorithm

to adaptive systems but has been widely used in the optimisation field of study due to its success on multidimensional problems [38, 122, 124]. The GA developed by Holland is referred to in the literature as the Canonical GA (CGA) [34].

The wide use of the GA can also be attributed to its generic algorithm structure as well as the ease of implementation [61, 122]. The GA defines generic operators that select, create and mutate individuals for the next population of the next generation [34]. It should be noted that initially the mutation operator was not a required to be part of the GA. Only after successive implementations of the GA showcased the explorative power that the mutation operator brings tot he search capability of the GA was the mutation operator considered to be important [34].

The GA search procedure involves searching the solution space through artificial evolution and natural selection [55, 105, 122]. An individual or point in the solution space is known as a *chromosome* [18]. An initial set of chromosomes (referred to in the research as the *population*) is randomly generated to form the starting population [46, 55, 105, 122].

A chromosome consists of a sequence of smaller parts called *genes* [34]. The sequence upon which these genes appear in the chromosome determines the characteristics of an individual [34]. In the GA a single gene represents a single variable of a solution [46, 122]. A whole chromosome therefore represents a solution [46, 122]. Exactly how best to represent a chromosome as a solution is problem dependent [34].

According to the evolution theory proposed by Darwin individuals of a population with the best chromosome have the best chance to survive and to reproduce [34]. These individuals are reffered to as the fittest individuals of the population. In a GA population each individual is “rated” to determine how good the solution its chromosome represents [34]. The rating of a chromosome is referred to as its fitness value and is also problem dependent [34].

Determining the fitness value of a chromosome is achieved by means of a *fitness function*. The fitness function is a mathematical function which maps the chromosome representation to a scalar value [34]. An optimisation problem has a objective function which calculates how good a solution is, therefore in the GA the fitness function represents the objective function [34].

When the fitness of individuals in a population have been determined

the GA probabilistically selects individuals for the next generation [34]. The selection probability is referred to as the *selective pressure* [34]. The individuals that are selected by the operator enter the reproduction phase of the GA where offspring are created [34].

Offspring is created by combining parts of one or more chromosomes to form a new chromosome this procedure is called the *crossover* [34]. The chromosomes used in the production of the offspring is referred to as the parent chromosomes [34].

With regard to how offspring and parents are handled, there are two forms of the GA [122]. One form is called the *generational* GA and the other form is known as the *steady-state* GA [122,126].

With the generational GA the offspring are not immediately used in the next generation; instead they are kept in a pool until the pool reaches a required size [122]. The offspring are then used to replace the parents entirely in the next generation [122]. In the steady-state GA once offspring has been created a selection operator is applied once more. The operator selects individuals from the old population and from the offspring to form the new population for the next generation [122,126].

The GA search process moves around in the search space using probabilistic rules rather than deterministic rules [122]. The probabilistic transition rules aid the algorithm to avoid local optima regions in the search space [55]. Note that by even using the transition rules there is no guarantee that the GA will completely avoid local optima [34].

The GA makes no assumptions about the search space and primarily works on the information provided by the chromosomes and the representation of solutions used [34,55,98].

Unlike the SA and TS algorithms, the GA if properly initialized, starts with a number of search points that cover a wide area of the search space. Depending on the operators use, diversity can be quickly lost and the population can become homogeneous very quickly [38,55,122].

As discussed earlier the defined operators consists of a selection operator, crossover operator and mutation operator [105,119]. In the section that follows each operator is discussed.

4.6.2 Important Genetic Algorithm Characteristics

The various operators used by the GA makes the procedure unique and is one of its defining characteristics. The first is the characteristic to be discussed is the selection operator.

Selection Operator

The selection operator is the first operator to be applied to the population after each generation. The operator determines which individuals will be used to create offspring [84, 98, 102].

The individuals are selected based on the selective pressure and are moved to the “mating pool”. Individuals that are selected by selection operator are used by the crossover and mutation operators (in the reproduction phase), to generate offspring from the mating pool [46, 61].

By favouring high fitness individuals above low fitness individuals the operator is said to have a high selective pressure [34]. Care must be taken if the operator uses a high selective pressure. Since high fitness individuals are preferred diversity among the individuals in the population will deteriorate and thus result in premature convergence [34, 98].

If the search space is known to have only one optimum, then a high selection pressure is feasible [98]. The selection operator directs the search into a gradient-based direction that converges on a single optimum solution [98].

In contrast, with a search space that is known to have multiple optima a low selection pressure is more beneficial to the GA [98]. Low selection pressure allows the population of the GA to explore the search space more [98].

The means by which the selection operator selects individuals is referred to as the selection scheme. The most widely adopted selection scheme is known as the *roulette wheel* selection scheme [98, 102, 126, 134]. With this scheme an individual is selected based on a probability defined by the fitness of the individual divided by the collective fitness of the population [126].

Depending on how complicated the objective function is and how large the population is, the selection phase may be the most computationally expensive as well as time consuming [46].

The interested reader is directed to the survey by Engelbrecht [34] for a more wide-ranging discussion on selection operators.

The selection process as described below can also be applied by the crossover operator. This process was opted to be split from the crossover operator as to provide a better overview of how the algorithm works

After the offspring has been generated, a selection operator is applied again [34]. The purpose of this selection operator is to determine the population of the next generation [34]. Within this selection process replacement policies can *optional* be applied [34]. These replacement policies define which individuals of the current population should be replaced by the newly generated offspring [34]. For instance, the policy can define that the entire current population should be replaced by the offspring [34]. Using this policy is not ideal as some generated offspring could have much worse fitness than their parents but due to the replacement the solutions represented by the good parents are lost [34]. Other replacement policies include replacing the worst individual in the current population with offspring on the premise that the offspring has a better fitness [34]. By replacing older chromosomes with better performing chromosomes in the population replacement strategy, the GA achieves an hill-climbing ability.

The reader that is interested in more information on different selection operators is directed to the survey by Engelbrecht [34].

Crossover Operator

The crossover operator is the first operator applied to the population in the reproduction phase. The crossover operates exclusively on the chromosomes in the mating pool.

Crossover works by interchanging genes from one or more parent chromosomes. The parent chromosomes are selected with a probability from the mating pool to produce a single chromosome known as the offspring [18, 102, 122]. The probability is known as the *crossover probability* and is problem dependent [34].

By defining a high crossover probability, good genes from the current population are retained in the next population [34]. With good genes being retained more historical information is passed onto the next generations population [122].

There are a variety of ways in which genes are interchanged between chromosomes in the crossover operation i.e. fixed point crossover, two point crossover, uniform crossover and gaussian crossover [34]. Fixed point crossover operates on binary parents whereby a point is selected in one parent and then all other bits are replaced by the other parents' bits [18].

Two point crossover generates two random indices which dictate a certain segment in the one parent to be interchanged with the other parent [98]. Where a segment consists of a subset of genes from the chromosome.

With a uniform crossover a chromosome is represented by a byte [34]. Each byte is made up of sequence of bits. As discussed each chromosome is made out of genes and therefore in the uniform crossover each bit is a gene [34]. The uniform crossover is the most basic of all crossovers since it randomly selects bits from one parent to be replaced by another parent's bits and is usually used when a large solution space must be search [124, 126].

Finally, the Gaussian crossover also uses the byte representation. The Gaussian crossover interchanges bits between parents based on a Gaussian distribution [124, 126]. Depending on the state of the algorithm, crossover operators can also be interchanged or even paired if the algorithm needs better search performance for large or small solution spaces [3, 124].

Caution should be exercised by the crossover operator when selecting chromosomes for reproduction. It is possible for the operator to select the same chromosome twice to be the parent of a single offspring [34]. The operator is also at risk of selecting the same chromosome multiple times for reproduction [34]. The operator must therefore incorporate a test to detect unenecessary repeated usage of a chromosome [34].

Finally, it is important to note that there exist strategies for both Gaussian crossover and mutation based operators for use with continuous-valued representations [33].

Mutation Operator

The mutation operator is a probabilistic operator and thus it is applied to an existing individual with a probability referred to as the mutation rate [34].

The purpose of the mutation operator is to increase the diversity of the genes of a individual chromosomes [34]. By introducing new genes into

an individual the diversity of the populations characteristics are increased [46, 98, 102].

The mutation operator has no previous information on the chromosome it is mutating; thus it is entirely possible that the mutation may modify the chromosome for the worse [46]. A worse solution might lead the algorithm out of local optima or lead it down a new path to find the global optima, but this is not always the case [61, 98, 122]. It is for this reason that it is recommend that the mutation rate be set to a low value to ensure good solutions are not distorted too much [34].

In a survey done by Engelbrecht [34] another mutation operator is discussed. Instead of mutating a small part of randomly selected chromosomes, this operator generates new offspring to be inserted back into the population. The operator randomly generates a new chromosome and then uses any of the previously discussed crossover operators (see page 86).

The mutation operator is not always a basic random replacement of genes operation. In research done by Il-kwon Jeong and Ju-jang Lee [61], a mutation operator is presented that uses the SA algorithm to determine the genes that need to be replaced. The SA mutation operator generates a new chromosome from which genes are used to replace in the chromosome being mutated [61].

Addition mutation operators are discussed in Engelbrecht [34]

Initial Population Generation

Initial population generation is the very first activity that the GA performs. Out of this population potential mating candidates are selected based on their fitness, which indicates desirability. Generally the initial population is generated by means of randomisation [119]. Since the algorithm searches multiple points simultaneously in the search space, it is desirable that the initial population have a wide diversity with regard to the solution it represents [38, 84]. By controlling the initial population generation we can control, to a small degree, the amount of exploration the algorithm does initially [84]. Therefore care must be taken in the selection of the particular randomisation scheme that will be used to generate chromosomes for individuals.

In a survey done by Andrea Reese [97], two randomisation schemes were defined, namely pseudo random number generators (PRNGs) and quasi ran-

dom number generators (QRNGs). PRNGs were found to be heavily problem dependent, improving the search efficiency in some instances and in other instances having no considerable impact. QRNGs, on the other hand, were shown to significantly improve the final solution produced by the GA as well as lower the number of generations for the solution to be obtained [97].

Algorithm Efficiency

The GA is a powerful, yet simple algorithm and tends to find good solutions given enough time, it does have its disadvantages. One of the major disadvantages occurs when the GA is applied to problems that have very large solution spaces. In these problems, the population size is a very sensitive parameter [3, 61, 92, 108]. If the population is too small the algorithm will not have enough diversity to search and will tend to converge prematurely.

A large population is preferred in large search spaces in order to get good chromosome diversity among individuals. Hence, the population size must be fine-tuned to achieve optimal performance in large problem spaces [34, 61]. Note an increase in population does not guarantee good chromosome diversity among the population. As discussed in the initial population generation it is also dependant on the random generation scheme used as well.

In the next section the pseudo code for the GA algorithm is presented and the flow of the GA algorithm is discussed

4.6.3 Flow of the Algorithm

The core concepts of the genetic algorithm were introduced in the previous section. To better understand the algorithm, a general overview of the algorithm be presented in this section using algorithm 4 as a reference point.

Algorithm 4 Basic Genetic Algorithm Algorithm [38, 61, 119, 122]

```

1:  $pop_n \leftarrow$  Initialize population
2: while Stopping criteria is not met do
3:   Evaluate individuals of population with fitness function  $f(i)$ 
4:    $y_k \leftarrow$  Select parent individuals from population using selection op-
       erator
5:   repeat
6:     for Each chromosome  $g_i$  in  $y_{k-1}$  do
7:        $c_i \leftarrow$  calculate crossover probability for  $g_i$ 
8:       if  $c_i \geq$  Crossover threshold then
9:          $g_i \leftarrow$  Apply crossover operator to  $g_i$ 
10:      end if
11:       $o_i \leftarrow g_i$ 
12:       $m_i \leftarrow$  Calculate Mutation probability for  $o_i$ 
13:      if  $m_i \geq$  Mutation threshold then
14:        Apply mutation operator to  $o_i$ 
15:      end if
16:      Add offspring chromosome  $o_i$  to  $new_{pop}$ 
17:    end for
18:  until  $size(new_{pop}) = size(pop_n)$ 
19:   $pop_n \leftarrow$  select new population from  $pop_n$  and  $new_{pop}$ 
20: end while
21:  $x_i \leftarrow$  Determine best chromosome in  $pop_n$ 
22: Return best solution  $x_i$ 

```

The GA algorithm is a population-based algorithm and therefore needs to initialise its population. Each individual of the population represents a potential solution. Population initialisation occurs in line 1 in algorithm 4.

The amount of individuals in a population to be generated is predefined and is known as the population size. In algorithm 4 the population size is represented by the subscript n in pop_n , where $n > 0$.

Before the algorithm can start *evolving* its population, it first needs to determine each individual in the population's fitness. The fitness of an individual is calculated using a fitness function $f(x_i)$.

After each individual within the initial population has been evaluated, the algorithm is able to start its searching process, which starts at line 3

and ends at line 17 of algorithm 4.

Since each individual has a fitness value after being evaluated, the selection operator is applied. The selection operator used on line 4 determine which individuals will form part of the parents used to generate offspring for the next population.

Once the selection operator has selected the parent individuals needed for the next population, the algorithm is ready to enter the reproduction phase, which ranges from lines 5 – 17 of algorithm 4.

In the reproduction phase the crossover and mutation operators are applied probabilisticly. For the crossover operator a crossover probability is calculated on line 7. Depending if the calculated probability satisfies the crossover probability the crossover operator is applied on line 9. Depending on the crossover used, offspring is generated using one or more individuals as parents. The resulting offspring from the crossover operation is assigned to g_i .

On line 11 the value of g_i is assigned to the variable o_i which represents the offspring. From lines 7 – 11 it can be observed that a particular individual does not have to part take in a crossover operation to be carried over to the new population. Thus knowledge gained from the current population is persisted based on the crossover probability.

The second step of the reproduction phase is where the mutation operator is applied. For each of offspring a mutation probability is calculated. If the calculated probability for a particular offspring is high enough, the algorithm enters the mutation phase. In the mutation phase an individual is selected and the mutation operator is applied. Application of the mutation operator can be seen to occur in lines 13 – 15 of algorithm 4.

Regardless of whether the offspring has been mutated or not, the resulting offspring is added to the new population. The reproduction phase continually loops, until the new population equals the size of the initial starting population. Once the amount of individuals in the new population has reached the population size n , the algorithm moves on to its next iteration.

After the algorithm has completed the reproduction phase the algorithm selects the new population to be used in the next generation. The new population is selected from the current population and the created offspring.

The algorithm continually generates a new population for each generation until a predetermined stopping criterion has been met. Once the

criterion has been met, the algorithm selects the individual with the best fitness in the current population as its most optimal solution.

The following section discusses literature where the GA has been applied to the FAP.

4.6.4 Genetic Algorithm on the FAP

Continuing the trend of the SA and TS algorithms, the GA has also been applied to a wide variety of optimisation problems and has on average obtained good results as can be observed in the following research [7, 108, 130]. Due to the success of the GA on other difficult problems, it has therefore also been applied to the FAP.

Colombo and Allen [21] have developed a GA to be applied on the FAP. The authors decomposed the FAP into smaller subproblems. On average the solution quality is improved by using the technique but at the expense of more complex and taxing evaluations that have to be performed [21].

In table 4.3 the results obtained by authors on the COST 259 benchmarks are compared. The listed values are scalars representing the total interference generated by the frequency plan.

Problem instance	GA	COST 259
Siemens 1	2.60	2.20
Siemens 2	16.34	14.280
Siemens 3	6.37	5.19
Siemens 4	84.08	81.89

Table 4.3: GA on COST 259 Benchmark

As per the results in table 4.3 the GA produces good results coming close to the best obtained results in the benchmark. By critically reasoning about the GA if it were applied to the FAP in theory, the following disadvantages can be identified:

Diversity — The GA continually operates on a set population that was randomly initialised at the beginning of the algorithm. It therefore only has this set of generated genes in the initialised population to therefore evolve successive populations. If one disregards mutation, the GA is a process by

which the optimal combinations of the starting genes are found. Thus, the GA purifies the starting population genes in an attempt to find those individual genes, which if combined into a single individual, will produce an optimal individual, i.e. solution. Therefore the GA is very reliant on the quality of the random generator used. The probability of the algorithm finding a particular desirable gene that is exhibited by the starting population is directly related to the random number generator used to initialise the population.

Crossover — The crossover operation in the GA is the only means by which successive populations are generated and can therefore be regarded as the primary means by which the algorithm performs its search. As the crossover is defined in the standard algorithm, certain parts of both parents are copied and combined to form a new individual. With regard to the FAP, if each individual represents a frequency plan, the crossover operation would copy certain cells from the two parent plans. This is not desirable, since a single channel within a cell can generate major interference which overshadows the rest of the channels that generate low interference in the cell. Thus, for the GA to generate high quality solutions on the FAP, the algorithm would be better off utilising a crossover operation which works on individual channels assigned rather than cells. Crossover operation is also a memory and computationally expensive operation since individuals need to be constantly created and values need to be copied to these new individuals from the respective parents.

Mutation — Mutation is a means by which more diversity is introduced into the chromosomes of the individuals. As discussed, the mutation operator introduces new genes to existing chromosome which can lead to an excellent solution being distorted and becoming one of the worst solutions. With regard to the FAP, the low probability of mutation is not desirable, as the FAP search space is huge and therefore requires constant diversity to be introduced to accurately explore it. A possible good mutation would be one that is slightly more intelligent than the standard mutation operator, which just randomly modifies a selected individual. An intelligent mutation would be one that takes into account the recent history of the individual as well as the history of the population and, based on the collective knowledge alters

or *mutates* a particular individual. Another means of applying the mutation operator is to set the mutation rate proportional to an individual's fitness. By modifying the mutation rate in this manner, mutating good individuals can be avoided.

4.7 Summary

In this chapter a description was given of metaheuristic algorithms. What it means for an algorithm to be classified as being of a metaheuristic nature was explained and the characteristics these algorithms must exhibit were identified.

Three metaheuristic algorithms were discussed in this chapter. For every algorithm discussed an explanation was given of how the algorithm works as well as the various characteristics that make the algorithm unique.

For each algorithm, a brief overview of studies using the particular algorithm was given as well as some of the disadvantages or challenges that would be faced when applying the particular algorithm to the FAP.

The first algorithm discussed was the tabu search algorithm and the second was the simulated annealing algorithm. The chapter concluded with the genetic algorithm. In the next chapter the class of algorithm that is relatively new to optimisation research, namely the swarm intelligence class of algorithms, is discussed.

In the table 4.4, the algorithms discussed in this chapter and their performance on the COST 259 set of benchmarks have been summarised.

Problem instance	TS	SA	GA	COST 259
Siemens1	2.7692	23.00	2.60	2.20
Siemens2	14.9360	14.75	16.34	14.280
Siemens3	6.6496	52.55	6.37	5.19
Siemens4	110.9725	80.80	84.08	81.89

Table 4.4: Summary of algorithm performance on the COST 259 benchmark

Chapter 5

Metaheuristic Algorithms: Swarm Intelligence

5.1 Introduction

The research field of artificial intelligence stands a lot to gain by the study of the inner workings of nature itself; this is why there is a branch of artificial intelligence that incorporates some of nature's processes, like evolution, which can be seen being applied in practice in the GA (see section 4.6).

There are other approaches in artificial intelligence which also have their routes in nature, for instance animal learning or the study of how dogs learn [13]. These approaches only look at a single agent thought process when an agent percepts (senses) are mapped to actions in the agents particular environment [13].

A percept can be said to be a process that is specifically designed to take data from the surrounding environment, process it and present it as information upon which decisions can be made [13, 100]. For instance, eyes are percepts, which take visual data presented by the environment [13, 100]. The visual data is processed by the brain into information to enable decisions to be made on navigating the environment [13, 100].

The research field of swarm intelligence is an approach more concerned with the underlying processes and behaviour patterns when multiple agents (insects, animals) come together and perform a task as one collective entity [13, 20, 67, 100]. The study of animals and insects and how they socially interact within their local environments in groups has already contributed to

artificial intelligence as a whole in the form of Swarm Intelligence [20,67]. In the field of Swarm Intelligence, animals and insects are represented by agents which are simple stimulus-response agents and can only perceive changes in their local environments. A group of agents is referred to as a *swarm*.

Swarm intelligence works on a key aspect observed in nature, the notion of emergent behaviour [16,33,34]. Emergent behaviour is when an individual in the swarm achieves success due to the it obtaining more knowledge through exploration about its local environment [16,33,34].

In swarm intelligence, when an agent exhibits emergent behaviour, the outcome of the behaviour needs to be shared with the whole swarm in order for the swarm to adapt and use the newly gained knowledge [16,33,34,67].

In a swarm, the agents communicate with each other about knowledge gained as a consequence of the changes in their local environments [16,33,34,67]. Changes can be caused by emergent behaviour by the agent itself, other agents or the environment [16,33,34,67]. Communication among agents facilitates *knowledge sharing* [16,33,34,67].

Social interaction among agents is certainly not the only means of interacting, but with regard to biological inspired systems it is the most prominent [34].

Using this social interaction between the agents of the swarm can influence their own local environment to move towards more promising space in the search space [16,33,34,67]. Due to the social interaction between agents the swarm is also referred to as a *social swarm* [16,33,34,67]. Thus each agents of a swarm contributes to the swarm as a whole to locate and produce better solutions [20].

As discussed the behaviour propagates from one agent to another through social interaction, which brings forth information exchange [16]. Social interaction is but one component of self-organisation. Other components that form part of self-organisation are [115]:

- Positive and negative feedback [115]
- Increasing fluctuations of random events [115]

The means by which agents facilitate indirect communication with each other is known as *stigmergy*. Stigmergy as well as the different forms of stigmergy is discussed in 5.2.

Swarm intelligence algorithms are also meta-heuristic algorithms, with the distinction being made that swarm intelligence algorithms use multiple agents as a collective entity of knowledge to search the problem space [16, 20, 33, 34, 67].

The initial algorithms developed with regard to swarm intelligence, were based on the coordination and behaviour exhibited by schools of fish and flocks of birds. The newer generation of algorithms include [16, 20, 67]:

- ACO [16].
- artificial bee colony (ABC) [20].
- PSO [67].
- bacterial foraging optimisation [34].
- Firefly optimisation [34].
- Fish school optimisation [34].

Swarm intelligence based algorithms are able to achieve good results since they have simple individuals searching in their own local environments for more optimal solutions [33, 34]. A direct consequence of multiple individuals in a swarm searching is that the algorithms are able to explore multiple locations within the defined search space [33, 34]. Traditional single agent based metaheuristic algorithms like TS (section 4.4) and SA (section 4.5) only have in essence one “individual” searching for a solution [33, 34].

With algorithms like TS and SA, the information is not shared since there is only one individual search [33, 34, 112, 135]. The information is kept to influence future decisions [79, 82, 100, 121, 129]. Swarm intelligence algorithms also store information for use by the various individuals that make up the swarm to make more informed decisions as the solution space is explored [33, 34].

NP-Complete optimisation problems are but one of the fields where swarm intelligence algorithms have been adapted to. Other fields where swarm intelligence has been applied include neural network training [34], vehicle routing [24], clustering [47], search engines and electrical power systems [10]. Swarm intelligence thus seems more suited towards problems with combinatorial complexity [23].

The chapter is organised as follows. Before the algorithms are discussed an overview of stigmergy is presented in section 5.2. Starting with section 5.3 the first swarm intelligence algorithm is discussed namely, Ant Colony Optimisation. Artificial bee colony is discussed in section 5.4. Section 5.5 is a discussion about the Particle Swarm Optimisation algorithm. This chapter concludes with section 5.6 that summarises the chapter.

5.2 Stigmergy

Stigmergy is defined as the method used by animals and insects to facilitate indirect communication [12, 26, 33, 34]. Through the use of stigmergy animals or insects are able to socially interact with their own species to convey information to each other [12, 26, 33, 34].

Interaction occurs through signals that the individuals receive which might require them to perform a specific action [12, 26, 34].

Two forms of stigmergy can be observed in nature. One form, *sematectonic stigmergy*, is a direct and physical form of interaction since it relies on altering the environment [34].

Examples of this type of stigmergy are nest building and brood sorting by ants [34]. Schools of fish also use this type of stigmergy to communicate direction and speed by visually observing their closest partner in the school. Besides using visual information, birds use sound to communicate with and alert each other [16].

The other form, *sign-based stigmergy*, is an indirect form of interaction, where communication occurs through some sort of signal mechanism [34]. Ants use sign-based stigmergy to communicate with each other. More on how ants communicate with this type of stigmergy is discussed in section 5.3.1.

Other species that use sign-based stigmergy are bees [45]. When a bee determines that an entity poses a threat to the hive, it might decide to sting the entity. The sting of a bee not only injects a toxin into the entity, but also releases a pheromone [45]. This pheromone alerts nearby other bees from the hive of the presence of an entity that is a potential danger to the hive [45].

The other bees of the hive pick up this pheromone that is released by the initial bee's stinger and attack the entity by also stinging it [45]. As

more bees sting the entity, more bee stingers emit the danger pheromone identifying the entity [45]. Hence the pheromone is reinforced and becomes stronger, which persuades more bees into action [45].

Stigmergy is a powerful mechanism that is able to alter the behaviour of a collective entity efficiently, as can be gathered from the above-mentioned examples of stigmergy in nature [12, 26, 34]. Stigmergy is therefore a core concept upon which swarm intelligence algorithms are based as these communication techniques are exploited to aid the algorithm in finding better solutions [12, 26, 34].

In the forthcoming sections three swarm intelligence algorithms is discussed. Each section is divided into four subsections.

First, an overview of the algorithm is given, where basic concepts about the algorithm are introduced as well as a general outline given of the search process the algorithm uses. The second subsection will give an in-depth discussion of some of the core characteristics that make the algorithm unique. The third subsection will provide a step-by-step discussion of the algorithm using pseudocode as a reference.

Finally, for each algorithm studies using the algorithm on the FAP are mentioned and the various considerations that need to be made to apply the algorithm to the FAP are identified.

5.3 Ant Colony Optimisation (ACO)

5.3.1 Introduction

ACO is a class of algorithms incorporating different behavioural aspects that ants exhibit when they perform certain activities, i.e. gather food, build nests and construct cemeteries [26, 34]. The first ACO algorithms that were developed were based on the foraging behaviour that was exhibited by ants when finding the most optimal path towards a food source. Deneubourg noticed the foraging behaviour when he performed the bridge experiment [26, 34].

The bridge experiment outlined by Deneubourg placed a food source a certain distance away from the nest [26, 34]. Two bridges of equal length were established towards the food resource. The ants initially selected a path randomly with no clear distinction of the more dominant path to take

to retrieve food from the food source [26,34]. After a finite amount of time, one of paths to the food source became the preferred route for the ants even though both paths were of equal length. The preference of the one bridge by the ants is due to the bridge having a higher concentration of pheromones due to the bridge being chosen randomly more by ants.

The bridge experiment was extended to have two bridges that differ in length. The extended bridge experiment is presented in figure 5.1 and is known as the shortest path bridge experiment [33]. In this experiment the ants started to prefer the shortest bridge within a finite amount of time [33]. The conclusion was made that the ants preferred the shorter bridge because ants return to the nest quicker and therefore the path is reinforced with pheromones faster than on the longer path [33].

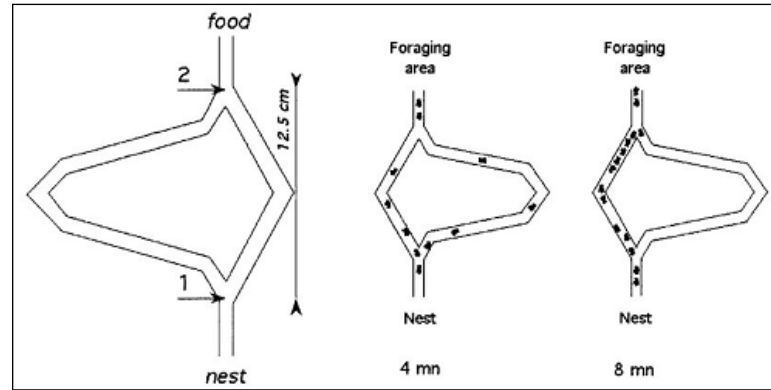


Figure 5.1: The shortest path bridge experiment [26]

Deneubourg concluded that ants utilise pheromones to communicate to the rest of the foraging ants the shortest path towards a food source. [33] By using pheromones to communicate with other ants it can be concluded that ants use sign-based stigmergy (discussed in section 5.2) when they retrieve food [12, 26, 34]. As the ant moves along a particular path, it marks the path with a chemical signal that alerts other ants to the desirability of the path [34]. The chemical signal that ants use to indicate optimal paths is called *pheromones* [26, 34].

Pheromones used by the ACO go through two phases. The first phase is where pheromones are deposited on a particular path by an ant whether there exists previous pheromones or not. The second phase is where pheromones evaporate. Pheromones are not permanent and deteriorate over time [33].

By letting pheromones evaporate, ants can “forget” previous decisions made by the colony [33]. It can be concluded that the more pheromones evaporate the less influence the ants will have on their path selection, therefore promoting exploration [33].

The concept of pheromones and how the ACO proceeds in updating the pheromones is a critical concept of ACO. Hence, an in-depth discussion on pheromones is provided in subsection 5.3.2.

The ACO class of algorithms have a *core requirement* about the problem they are applied to [33]. The problem must be able to be modelled as a graph. The reason behind this requirement is that each individual ant in the ACO algorithm constructs a path through the graph [33]. The path constructed represents a solution.

A path through a graph is made up of a series of links between nodes [42,100]. A link between two nodes represents a movement from one node to the other [42,100]. Therefore, a path can be considered as the traversal of the interlinked nodes, from a starting node to some final node [42,100]. A path differs from another path by the order in which the nodes are interlinked between a start and end node [42,100].

The ACO class of algorithms has been applied to a wide range of problems that include single machine scheduling [50], weapon target assignment [63], flow shop scheduling [19] and image thresholding [136]. Variants of the standard algorithm have been developed, but all of the algorithms still follow the core structure of the ACO algorithm [33,34]. The first algorithm developed based on the foraging behaviour of ants is known as the simple ant colony optimisation (SACO) and was proposed by Dorigo in 1992 [34]. The algorithm provided the basis for how pheromones are used and updated. The SACO is an algorithmic implementation of the double bridge experiment.

The first algorithm to improve upon the SACO is the ant system (AS) [12,34]. The AS included heuristic information into the probability that an ant chooses to move towards a node. The AS also added memory to the AS by using a tabu list and also incorporated pheromone evaporation. The improvements made enabled the AS to better explore the search space and produce better results [12,34].

The AS algorithm has achieved relatively good success in the problems it has been applied to, but it does have some disadvantages [24,131]. One

of the primary disadvantages of AS is that it tends to premature converge on local optima [33, 131]. The premature convergence can be attributed to due to the ants exploiting the high concentration of pheromones on good solution paths too quickly [33]. With the ants focusing only on the good solutions less exploration occurs in the search space leading to local optima being produced as the best solution [33].

Subsequently, various algorithms have been developed that improve on the AS algorithm. These improved algorithms include the ant colony system (ACS), min-max ant system (MMAS), Ant-Q, fast ant system, AntTabu, AS-rank and iapproximate non-deterministic tree search (ANTS) [12, 34]. The discussion in this section is focused on providing an introduction to the general ACO algorithm concepts and not to discuss various improvements made by variants of the core algorithm. In the forth coming sections reference is made to these algorithms and their respective improvements.

The AS algorithm is the base algorithm upon which all other ACO class algorithms are based upon. Therefore in the forth coming sections when a reference is made to the ACO algorithm it is directed at the base AS algorithm.

In this section the concepts upon which the core of ACO is based upon were briefly introduced. The concepts include pheromone trails, pheromone evaporation, pheromone updating and state transition rules. In the next section a discussion is given on each of the concepts.

5.3.2 ACO Characteristics

In this section characteristics that are important and unique to the ACO class of algorithms is discussed. The discussion is focused upon three core characteristics namely pheromone trials, pheromone evaporation, pheromone updates and state transition rules.

Pheromone Trail

The pheromone technique used by ants forms part of the core methodology used by the ACO algorithm [37]. As an ant moves it lays down pheromones to mark the path it is walking.

With the use of pheromones ants are able to communicate the best and shortest link between nodes [26, 34, 37]. The more ants following a pre-

ferred link the more pheromones would be laid on that specific link. This increases the strength of the pheromones [131]. The increase in strength of the pheromones on a link would thus let ants more clearly distinguish between links they should and should not take [131]. Therefore, a pheromone provides positive feedback to the colony [26, 34, 37].

Initially, all the ants will choose random a link to a node [26, 34, 37]. After all the ants have completed their paths, each path is evaluated using a cost function defined by the problem domain [34]. The amount of pheromone marking the links contained in a path in the standard ACO is related to the cost function [26, 34, 37]. Therefore, a low cost function value will have a high pheromone dosage indicating the link the ant took from each node, and a high cost function value will have a low dosage [34].

By incorporating the cost of a particular path into the amount of pheromone deposited the colony is able to influence future decisions [34]. In terms of minimisation a path with a low cost will have a high pheromone value making the links of path more likely to be selected by future ants [34].

In the iterations following the initial one, the ants will at each node decide based on a probability whether it should to a particular neighbouring node. The higher the pheromone intensity is at a neighbouring node, the higher the probability that the ant will choose to move towards that node [26, 34, 37]. The probability with which ants choose links to neighbouring nodes will defined and discussed in the next section.

Due ants choosing links to node based on a probability, it is still possible for the ant to choose a random link towards an other node. Thus the ACO algorithms are considered stochastic search procedures due to the ants ability to choose links randomly when exploring the search space [24, 131].

The pheromone trail was initially developed with only one colony in mind [34]. In the research done by Tiwari et al. [118] pheromones in multiple colonies are considered. The basic principle of how pheromones are used by the ants stays the same, but the meaning of the pheromone changes if an ant of another colony encounters the pheromone trail [26, 34, 37]. The ant will not follow or even consider the pheromone trail since any pheromone encountered from other colonies repulses the ant [118]. Thus pheromones only provide positive feedback if the ant is from the same colony, otherwise the pheromone gives negative feedback, in a way warning the ant to stay away [118]. This repulsion strategy promotes exploration among the

multiple colonies [118].

The probability with which a link towards a particular node is chosen by an ant forms part of the *state transition rules* which is discussed in the section that follows.

Pheromone Evaporation

Initially when the pheromone concept was first implemented the ants of the colony rapidly converged on a solution. The search space was not adequately explored and the produced solution was of local optima. To combat this premature convergence and force the ants to explore the search space more, the concept of *pheromone evaporation* was introduced [12, 25, 26, 34].

Real pheromones used by ants to mark a particular link to a node is not permanent and over time the strength of the pheromone deteriorates until it eventually disappears. Pheromone deteriorating is known in the literature as pheromone evaporation. The pheromone will not completely evaporate as long there are ants traversing the defined link reinforcing the pheromone. The evaporation of pheromones is modelled in the ACO by equation 5.1 [12, 25, 26, 34]:

$$\tau_{ij}(t) = (1 - p)\tau_{ij}(t), p \in [0, 1] \quad (5.1)$$

The constant p defines the rate at which the pheromone evaporates. If $p = 1$ the pheromone completely evaporates every iteration. With no pheromone on a link towards a node the ants take no knowledge gained from the previous iteration into account and therefore select a link randomly [26, 34]. Thus, the amount of exploration done by the algorithm can be controlled by the constant p [26, 34].

Equation 5.1 was first introduced in the AS [12, 25, 26, 34]. Most subsequent algorithms that are a form of the ACO class of algorithms also use the concept of pheromone evaporation, but they either use the standard equation or develop their own variant [12, 25, 26, 34].

A more aggressive form of pheromone evaporation is added to the AS discussed in the research done by Gambardella et. al [37]. The more aggressive form works beside the already present pheromone evaporation, but this form seeks to add an additional search phase called *diversification* [37]. The aim of the diversification phase is to lead the algorithm into another

direction of the search space [37]. This is done in an attempt to avoid local minima and stagnation [37].

In the ant system developed by Gambardella et. al. the algorithm continually monitors the current best solution and keeps a history of recent best solutions [37]. If the algorithm starts to notice that solutions are cycling or that the current best solution has not changed for a certain number of iterations, the algorithm activates the diversification phase [37]. The algorithm is forced to re-search the search space to create new solutions, as it cannot rely on previous historical information provided by the pheromone trails [37].

State Transition Rules

The intention for this section is not to give an exhaustive survey of different transition rules in the literature. Therefore, only the first transition rule that was developed is discussed, since most of the other rules can simply be considered derivatives of the first.

As discussed previously, the ants select which link to follow towards a node next based on a probability. This probability is also known as the *transition probability* and is formulated by equation ref5.2.

$$p_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t)\eta_{ij}^\beta}{\sum_{u \in N_i^k(t)} \tau_{iu}^\alpha(t)\eta_{iu}^\beta(t)}, & \text{if } j \in N_i^k(t) \\ 0, & \text{if } j \notin N_i^k(t) \end{cases} \quad (5.2)$$

The transition probability is used by individual ants of the AS algorithm [25, 33, 34, 37, 118]. An ant k uses this equation to decide with what probability it will move from node i to node j [33, 34, 37, 118]. τ_{ij} is the amount of pheromone on the link between nodes i and j [26, 33, 34, 37, 118]. Heuristic information is incorporated into the equation through the symbol η_{ij} , which is the desirability of the link from node i to node j as evaluated by a heuristic function [26, 33, 34, 37, 118].

Each ant starting at a source node moves from one node to another based on the defined transition probability until the ant reaches the final node.

Through the use of parameters α to represent pheromone intensity and β to represent heuristic information the algorithm is able to achieve a good balance between exploration and exploitation when $\alpha = \beta$ [33, 34, 37, 118]. When $\alpha = 0$ no pheromone is taken into account; hence, any history that

the algorithm has on the link between node i and node j is neglected and the algorithm degrades to a stochastic greedy search procedure. If $\beta = 0$ then the algorithm does not take into account the amount of desirability of the link between node i and node j as dictated by the problem-specific heuristic function.

The set $j \in N_i^k(t)$ contains all the valid neighbourhood moves ant k is allowed to make when moving from node i to node j . A tabu list is kept by each ant to trim the set of moves already performed previously, and thus cycling is prevented.

The interested reader that wants more information about state transition rules is directed to the survey by Engelbrecht [33].

Pheromone Update

Pheromones start to evaporate over time, and so the link marked by the pheromone trail becomes less attractive to the ants. Therefore, a path that represents a good solution needs its pheromone trail to be continuously updated. Certain rules govern when and by how much pheromones are updated.

Most of the variants that have been developed differ in what pheromone update rules they employ. In the literature pheromone update rules are classified into two groups [34]. One group is called the global update rule. The other group is called the iteration-based or local update rule [34].

The first local pheromone update rule was first introduced in the AS algorithm [25, 26, 34]. The ants would retrace their path after each iteration, depositing pheromones on each link that makes the complete path. The following equation is used to update the pheromone:

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \Delta\tau_{ij}(t), \quad (5.3)$$

$$\text{where } \Delta\tau_{ij} = \sum_{k=1}^{n_k} \Delta\tau_{ij}^k(t)$$

In equation 5.3 $\tau_{ij}(t+1)$ represents the amount of pheromone that will be on the link for the next time step $(t+1)$. τ_{ij} represents the amount of pheromone currently on the link (i, j) . $\Delta\tau_{ij}$ is the actual amount of pheromone that needs to be added to the current pheromone τ_{ij} .

Pheromone update rules that are in the global update group only allow the pheromone trail of the path representing the best found solution by the algorithm since the first iteration to be updated [34]. Thus the global rule favours intensification where the algorithm exploits the global knowledge gained by the ants to find a better solution. By updating a pheromone the concentration of the particular pheromone is reinforced.

ACS was the first to use both the global update rule and local update rule together [34]. By using both types of rules the algorithm is able to efficiently exploit the history provided by the pheromones [34]. The global update rule used by the ACS is formulated in the following equation [34]:

$$\begin{aligned} \tau_{ij}(t+1) &= (1 - p_1)\tau_{ij}(t) + p_1\Delta\tau_{ij}(t), \\ \text{where } \Delta\tau_{ij} &= \begin{cases} \frac{1}{f(x^+(t))} & \text{if } (i, j) \in x + (t) \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5.4)$$

The parameter $f(x^+(t))$ represents the best/shortest path found so far by the algorithm [34]. p_1 is the variable that controls the rate of evaporation. $\Delta\tau_{ij}(t)$ is the amount of pheromone at the current time step t for the link ij .

By using the global update rule the algorithm is able to direct the search more, which is to say the algorithm exploits the search space more. Exploitation is achieved since the best path is continually used in the update of the pheromone as can be observed in equation 5.4 [33, 34].

As can be seen in the following equation, the ACS uses a slight variant of the local update rule first used in AS [34]:

$$\tau_{ij}(t) = (1 - p_2)\tau_{ij} + p_2\tau_0 \quad (5.5)$$

In the above equation τ_0 is a small constant and $p_2 \in [0, 1]$ is the constant that defines the rate of evaporation [34]. With the local update rule, the algorithm is able to explore more as the individual ant constructed path is used to update the pheromone and no information from the best path found in the colony is incorporated, as with equation 5.4 [33, 34].

The MMAS algorithm as discussed also improves on the AS. The global update rule used by AS has a disadvantage in the sense that the search might concentrate too quickly on a particular good solution (the global best path) [33]. MMAS address this disadvantage by using the global update rule on an iteration basis [33].

When considering only on an per iteration basis, the best path found by the algorithm can differ from one iteration to the next [33]. Therefore the MMAS a different path will be updating with the global update rule each iteration [33]. Using this approach the algorithm is allowed to explore the search space more [33].

Another short coming of the AS is that pheromone concentrations can become extremely high leading to less exploration by the algorithm [33]. The MMAS algorithm addresses this short coming by enforcing an maximum and minimum amount of pheromone that can exist on a path [33]. By defining a maximum the algorithm is prevents the algorithm from settling on one particular solution i.e, stagnation [33]. On the other hand, defining a minimum on all possible links between nodes ensures that links will be continously considered for possible inclusion into a solution [33]. In addition to providing boundaries for the pheromones MMAS also uses a smoothing strategy to even out the difference between high and low pheromones [33].

5.3.3 Flow of the Algorithm

In this section the process the AS algorithm uses to explore the search space is described using algorithm 5 as a reference point.

Algorithm 5 Ant System Algorithm [34]

- 1: Initialize τ_{ij} with small starting values
 - 2: $t \leftarrow 0$
 - 3: Place n_k ants on starting node
 - 4: **while** stopping condition not reached **do**
 - 5: **for** each ant $k \leftarrow 0$ to number of ants n_k **do**
 - 6: $p^k(t) \leftarrow$ Initialize path p^k for time step t
 - 7: **repeat**
 - 8: Select next node based on probability equation 5.2
 - 9: Add link (i,j) to path $p^k(t)$
 - 10: **until** Final node reached
-

Algorithm 6 Ant System Algorithm (continued)

```

11:    $x^k(t) \leftarrow$  Remove loops from path  $p^k(t)$ 
12:   Calculate length of path  $f(p^k(t))$ 
13:   end for
14:   for each link  $(i, j)$  in graph do
15:      $\tau_{ij} =$  Reduce pheromone of link  $(i, j)$  with equation 5.1
16:   end for
17:   for each ant  $k = 0$  to number of ants  $n_k$  do
18:     for each link  $(i, j)$  in  $p^k(t)$  do
19:        $\Delta\tau_{ij} = \frac{1}{f(p^k(t))}$ 
20:       Update the pheromone  $\tau_{ij}$  with equation 5.4
21:     end for
22:   end for
23:    $t \leftarrow t + 1$ 
24: end while
25: return path  $x^k(t)$  with the smallest  $f(x^k(t))$  as the solution

```

The ACO algorithm initialises by creating a set population of ants and placing them on random starting nodes as well as initialising the pheromones to starting values as can be observed from algorithm 5, lines 1 – 3. The main purpose of the ant is to explore the solution space and to ultimately produce a solution that might be optimal. The ant explores the solution space by performing a series of moves from one node to another. Each move is a link that is added to the path. This process can be seen in lines 5 – 10.

The ant selects which node to move to next based on a probability. The probability is calculated taking into account the amount of pheromone that is on the current link representing the movement from the current node to the next node [33, 34]. This decision process can be seen to occur in line 8.

As the ant moves it records each link between the nodes it traverses until it reaches the final node. All the links the ant has traversed represent a path taken through the solution space [33, 34]. Thus, as the ant is moving it is actively building an optimal solution.

Before the ant deposits pheromone on the links it traversed to construct its solution, the pheromones first need to be decayed. This is why in lines 14 – 16, the algorithm traverses all links that contain pheromones and reduces the amount of pheromones by applying equation 5.1.

Once an ant has constructed a path and the pheromone evaporation has occurred, the ant is ready to inform the rest of the ants of what movements it made to construct its solution. The ant needs to share this movement information in order for the rest of the colony to know which movements worked well and which did not. The ant therefore needs to signal the other ants, which is accomplished with pheromones. Therefore, in the next phase of the algorithm, pheromones are deposited on all the links that make up the path the particular ant constructed. In the algorithm pheromones are deposited on lines 17 – 22 in algorithm 6.

After all the ants have deposited pheromones on all the links represented by each individual ant's constructed solution, the algorithm is ready to continue to its next iteration. This process occurs until some defined stopping criterion occurs.

5.3.4 ACO on the FAP

ACO has been applied to a wide number of problems and has produced good results. As discussed in chapter 3, the FAP can be modelled as a graph and therefore the ACO has also been applied to it.

When using the ACO algorithm on the FAP the ants need to construct a path that represents a frequency plan and has low interference. With the ACO, a node is a cell that has a unique set of channels assigned to it. Thus the same cell may exist in the search space, but will have a different set of channels assigned to it, and will therefore represent an entirely different node to the ACO.

As an ant moves in the frequency planning domain, it is actually moving between two cells that are said to interfere. The interference between two cells occurs as a consequence of the channels that have been assigned to them.

As an ant completes a movement from one cell to another, i.e. it assigns channels to the cell, it measures the interference that occurs due to the assignment. The measured interference information is incorporated into the pheromone, which the ant will deposit on the link between the two cells.

An optimal frequency plan would therefore be a path through all the interfering cells marked with a high dosage of pheromone. As discussed in the previous sections, the pheromone indicates the desirability of a particular

path. In the FAP, a desirable path would be one where interference is low; thus a path with a high dosage of pheromones would be the frequency plan with the lowest interference found by the algorithm.

When analysing the basic ACO algorithm 5 one can identify the following possible disadvantages if the algorithm were applied to the FAP:

Memory Usage — The algorithm requires a fair amount of memory. The memory is used to keep track of each permutation of a particular cell and its allocated frequencies until the pheromone that links to the cell has decayed enough to be discarded. As an ant moves from one cell to another, it might not select the previous cell (due to probability) to move to, but rather generates an entirely new cell to move towards. This newly generated cell would then be linked to the previous cell, and therefore the algorithm needs to keep track of the pheromone on that link until it has completely been decayed away. The algorithm needs to keep track of these pheromones on the links even if the new link to the generated cell is not even close to optimal and has very high interference.

Building a solution — The ACO *builds* an optimal solution. Therefore, early decisions made by the ants still influence the plan later for better or for worse. A good decision might seem to be good early on, but later the algorithm might be better off with a slightly worse decision. In the FAP, a cell can have multiple interfering cells, but a particular ant only knows about one link between two cells and not about the other interfering cells. Thus an ant will find the optimal path on the first interfering link between two cells, in other words it will optimise the channels allocated to these cells so that interference is low. The first interfering link is now optimised, and subsequent ants will reinforce this channel allocation since the interference is low. When the ants later reach the other cells that also interfere with the first cell that has been optimised, they will have difficulty changing the assignments that have already been made, since the pheromone representing that assignment is too strong to disregard.

The above disadvantages have only been identified by critically evaluating the algorithm as a possible point of interest to produce an optimal solution for the FAP for this dissertation. Even with these disadvantages the ACO has achieved success in producing high quality optimal solutions

Time limit	ACO*	ACO
120s	104719.72	91140.04
600s	103752.12	89703.44
1 800s	103781.86	88345.94

Table 5.1: ACO and ACO* on custom GSM FAP benchmark [68]

for the FAP.

In research conducted by Luna et al. [68] an ACO algorithm was applied to a custom cellular network instance. This network instance had 711 sectors with 2 612 transceivers, which needed to be assigned frequencies. For their particular network, only 18 channels were available for assignment. The channels started at 134 and ended at 151 [68].

The authors presented two versions of the algorithm. The first version used no heuristic information (henceforth referred to as ACO*) and the other version used heuristic information to update the pheromone laid down by the artificial ants [68].

With regard to the heuristic updating of the pheromone trails, the authors opted to increase the pheromone by some magnitude [68]. This magnitude was handtuned to be 100. The heuristic only increases a certain path's pheromone if the frequencies assigned to the transceivers represented by this path differ enough so as to not cause significant interference [68]. Thus, the heuristic aims to amplify good choices made previously by the algorithm for the next iteration of the algorithm.

In table 5.1 the results obtained by Luna et. al [68]. By evaluating the results obtained, one can clearly see that the ACO version that incorporates heuristic information to reinforce pheromone trails outperforms the version that does not [68]. The values depicted in the table represent the amount of interference that will result if the plan is used in the network [68].

5.4 Artificial Bee Colony (ABC) Algorithm

5.4.1 Introduction

The ABC algorithm is the most recently presented algorithm in the literature discussed in this chapter [59, 60, 106]. The algorithm was first proposed by

Karaboga in 2005 who wanted to mimic the foraging behaviour exhibited by bees [59, 60, 106]. Like ants, bees need to gather food to support the colony. To understand how the ABC algorithm tries to mimic the foraging behaviour of bees, this behaviour of real bees needs to be described first [59].

In a bee colony there are numerous bees, each with a specific role that dictates what actions a bee can perform. There are bees that protect the queen, maintain the colony, scout for resources and gather food, i.e the worker bees. The most important bees for foraging are those that scout and gather food [59].

The scout bees are sent out and as their role implies, are responsible for exploring the surroundings of the hive to find suitable food sources [59]. If a scout bee has found a food source it needs to return to the colony to share the information with the worker bees [59]. When the bee enters the colony it needs to communicate to the other bees by using some form of stigmergy (see section 5.2) [59].

The scout bee accomplishes this communication by performing a dance known as the *waggle dance* in the colony for all the bees to see [59]. This is not a dance as in the traditional sense, since through certain movements the bee is able to communicate a variety of characteristics about the food source including [59]:

- How far the food source is from the colony
- Quality of the food source
- Path towards the food source

It can be concluded that foraging bees use *sematectonic* stigmergy (discussed in section 5.2). This is deduced from the dance which is a physical form of communication.

The dance is observed by *onlooker* worker bees [11, 59]. These onlooker bees are initially *unemployed* in the colony [11, 59]. Once the information of the scout has been transferred to the onlooker bees, the onlooker bees become *employed* bees [11, 59]. They become employed bees when they operate on a particular food source to gather food [11, 59]. Thus it is the job of the worker bees to *exploit* the information provided by the *exploration* done by the scout bees [59, 60].

Worker bees gather food from the designated food source, until the food source reaches a certain quantity with regard to nectar content [59,60]. Each time the bee returns to the colony it evaluates the current food source versus other food sources discovered [59,60]. If a better food source is found, the bee abandons the previous and starts gathering food from the new source [59,60]. On the other hand, if the food source has been exhausted, meaning there is no more nectar content to gather, the bee returns to the colony and becomes “unemployed” again [59,60].

In the ABC algorithm, possible solutions are considered to be food sources [59,60]. Each food source has an employed bee associated with it. Onlooker bees either wait for new food sources to be communicated to them or become employed bees by moving to another, more attractive food source [59,60].

A food source might be more attractive to a bee because its defined nectar content is more than that of the current food source the bee is operating on [59,60]. The nectar content of a food source can be considered to be the fitness, which is determined using the fitness function of the specific problem domain [59,60].

As with real honey bees, a *waggle dance* is performed to all the onlooker bees by employed bees that provide information on the nectar amount (fitness value) that they represent [11,59,66]. The onlooker bees choose food sources depending on the nectar amount [11,59,66]; therefore as the nectar amount of a food source increases, the probability that more onlooker bees will choose the source increases [11,59,66]. How and what affects the probability is discussed in the next subsection.

Bees can transition to different roles depending on their situation [59,60]. An onlooker bee becomes employed when assigned to a food source and an employed bee can become a scout if its initial food source becomes exhausted [11,59,66]. Note that not all employed bees of a food source become scouts; only the first employed bee of a food source transitions to a scout [11,59,66]. Scout bees are sent to randomly generated food sources [11,59,66].

The more onlooker bees a food source attracts, the more the neighbourhood will be explored since the onlooker bees move to the food source and choose an immediate neighbouring food source to be employed upon [59,60]. Thus, this can be considered exploitation and the algorithm is therefore performing a local search [59,60,66]. Finally, the number of onlooker bees a food

source has also indicates its desirability. A very good solution will have the majority of onlooker bees choosing it and searching for nearby better food sources [59,60,66]. More bees are lured towards a particular food source due to the high nectar amount that has been communicated to them by other employed bees [59,60,66].

When a food source is abandoned, the previous bee that occupied the food source transitions to a scout bee [59,60]. The scout bee is responsible for replacing the abandoned food source by finding a new one, and a new food source is generated and communicated back to the colony [11,59,60]. The generation of food sources is discussed in the next subsection.

Karaboga was not the first to base an algorithm on the above foraging behaviour. Other bee foraging inspired algorithms have been developed such as the BeeHive algorithm, bee colony optimisation (BCO) and bee swarm optimisation (BSO) [60,73,116].

The BeeHive algorithm is based on the dance communication used inside the colony of bees. In BCO solutions are randomly generated and assigned to bees [60,73]. The solutions are then progressively modified using certain strategies. Finally, BSO solutions are iteratively constructed by forager (worker) bees and the best solution is communicated to the rest of the colony by performing a dance [60,73].

Another bee algorithm is the virtual bee algorithm (VBA) which, like the previous algorithms, is also based on the foraging behaviour of bees, but it differs in that it is not designed for combinatorial problems [60]. Instead the VBA is a variant of the standard ABC algorithm which is designed for numerical function optimisation [60]. In VBA bees move around in the search space communicating to each other any target nectar food sources that are found [60]. Good food sources are function evaluations of particular coordinates in the numerical search space which produce low function evaluation values in the case of minimisation [60].

Karaboga developed the ABC algorithm based on the previous research done on bee colony optimisation and the above algorithms. The ABC algorithm is designed to be a multivariable optimisation algorithm and has to date been applied to the job scheduling problem, clustering [73], neural network training and reconfiguration of distribution networks [66]. Due to the nature of the algorithm being similar to that of the ACO, the ABC algorithm will most likely also be applied to a whole host of other of problems.

In the next subsection the general flow of the ABC algorithm is described, which will aid in the understanding of how the algorithm searches a particular problem space.

5.4.2 ABC Algorithm Characteristics

Various characteristics of the ABC algorithm define the algorithm and make it unique. The first characteristic is how food sources are handled in the algorithm. The second is how information is communicated to the colony.

Food Sources

As discussed previously, food sources represent solutions to the problem the ABC algorithm is being applied to. When the algorithm starts, there are no defined food sources for the bees to evaluate and report on, and therefore initially a finite number of food sources are randomly generated [11, 20, 41, 59, 60, 66]. Since each food source needs an employed bee to evaluate the nectar amount of the source, the parameter that defines the number of food sources also defines the number of employed bees [11, 59, 60, 106].

Employed bees evaluate these food sources by determining their nectar amount [11, 59, 60, 106]. The nectar amount is directly related to the fitness value calculated using a domain specific cost function [59, 66, 106]. After the amount is determined the employed bee advertises the food source to the colony by performing the waggle dance. In algorithm 7 lines 4 – 9 are where the nectar amount of a particular solution is calculated.

Onlooker bees witness a number of dances from a variety of employed bees [11, 20, 41, 60, 66]. They therefore need to select a food source that is the most attractive while maintaining some diversity in the pool of solutions. Thus, an onlooker bee selects a food source based on a probability function which is formulated in equation 5.6 [59]:

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (5.6)$$

The parameter p_i is the i th food source under consideration by the onlooker bee. As discussed above, the fit_i parameter represents the value of the cost function and is directly related to the nectar amount of food source i . The parameter SN is the maximum amount of food source and hence the maximum employed onlooker bees [59].

In algorithm 7 the waggle dance can be seen being applied in line 11 where the probability of all the employed bee' solutions are calculated using equation 5.6. From lines 12 – 16 the onlooker bees evaluate the solutions of the employed bees based on the probability p_i that was calculated. P_i can be seen as the rating of the waggle dance that was performed by the employed bee.

Employed and Onlooker Bees

As previously outlined, when recruited onlooker bees reach the advertised food source that is stored in memory, they do not occupy the same food source [59,60]. Instead the bees explore the immediate neighbourhood of the food source that was communicated to them [20,41,66]. They seek to find a food source that improves on the previous one [20,59,60,106]. Equation 5.7 is used by the bees to generate new food sources in the neighbourhood of food source x_i [20,41,59].

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (5.7)$$

The subscripts $k \in \{1, 2, \dots, SN\}$ and $j \in \{1, 2, \dots, D\}$ are indices which are randomly chosen. D is the maximum dimensionality of the vector a solution represents. The index k has a constraint tied to it – whatever value is randomly assigned to k *must* differ from the value i . The position of the new food source in the neighbourhood of x_{ij} is controlled by the ϕ_{ij} parameter, which is a bounded random value between $[-1, 1]$.

From equation 5.7 it can be concluded that the randomness of the food source position decreases as the difference between $x_{ij} - x_{kj}$ decreases. Thus, as the algorithm moves closer to an optimal solution the finer grained the search process of the algorithm becomes [11,59,60].

After a new solution v_i is produced, the bee takes the new and old solutions from memory to compare their respective nectar contents. If the new solution is found to have higher quality nectar, the bee replaces the old solution in memory with the new solution [11,20,59,66]. Otherwise, the bee abandons the new solution and keeps the old solution in memory [59,60,106]. Thus, the bee seeks to always move towards a better solution and therefore uses a greedy selection process [59,66,106].

One of the problems with the above approach is that little information about the food source is used in generating a neighbouring food source.

In the research by Singh [106] a slight variation is proposed to generating food source neighbours by using more global information. The author adds a constraint to the algorithm that all neighbouring solutions generated by *employed* bees must be unique.

When an employed bee generates a neighbour and an identical solution already exists in the system, a *collision* is said to have occurred. A collision is solved by letting the employed bee transition to a scout bee so that a completely random solution can be generated [106]. Scout and onlooker bee generated solutions are not checked if they collide with other solutions in the system since the aim for them is exploring and not exploiting as with employed bees [20, 59].

Scout Bee

The artificial bees are modelled on the behaviour of real bees. Thus an employed bee can also abandon certain food sources when it has outlived its usefulness. Abandonment of a food source can occur for the following reasons [11, 20, 60]:

- The employed bee has reached the maximum allowed cycles to improve the nectar amount. The maximum cycles spent on a food source allow the algorithm to avoid local optima [11, 59, 60].
- The bee cannot improve the solution represented by the food source any further [11, 59, 60].

When a food source in the algorithm is abandoned, it needs to be replaced by a new food source [11, 20, 59]. Note that a food source is not abandoned when it represents the globally best found solution. An employed bee undergoes a role transition when it abandons a food source from an employed bee to a scout bee [11, 59, 60].

It is the responsibility of the scout bee to replace the abandoned food source with a new randomly generated one [11, 20, 59]. The scout bee uses equation 5.8 to produce a new food source that will replace food source x_i .

$$x_{ij} = x_{yj} + rand[0, 1](x_{zj} - x_{yj}) \quad (5.8)$$

The subscript y represents the minimum value of i and the subscript z represents the maximum value of i .

In research done by Gómez-Iglesias et al. [41] an extension is made to the scout bees. The scout bee individuals are divided into two types of bees, namely *rovers* and *cubs* bees [41].

- Rover bees are similar to traditional scout bees and hence use diversification strategies to explore the solution space.
- Cub bees explore the solution space relative to a good solution found by a rover by randomly changing configuration parameters.

By using two different scout bees a good balance is achieved when searching the solution space in the beginning where diversity is preferred and late in the algorithm where intensification is preferred [41].

5.4.3 Flow of the Algorithm

Most of the concepts that are used in the ABC algorithm have been explained. The general search process of the algorithm will now be discussed using algorithm 7 as a reference point.

Algorithm 7 Basic Artificial Bee Colony Algorithm [59]

```

1:  $b_n \leftarrow$  Initialize bees
2:  $s_n \leftarrow$  Initialize starting solutions
3: Evaluate starting solutions with fitness function  $f(s_n)$ 
4:  $t \leftarrow 0$ 
5: while stopping criteria not met do
6:   for each employed bee  $eb_i = 0$  to max bees  $b_n$  do
7:      $v_i \leftarrow$  Generate new solution with equation 5.7
8:     Evaluate with fitness function  $f(v_i)$ 
9:     Apply greedy selection between  $v_i$  and the current solution  $s_i$  of
       bee  $eb_i$ 
10:  end for
11:  Calculate probability  $p_i$  for solutions  $s_i$  in  $s_n$  using equation 5.6
12:  for each onlooker bee  $ob_i \leftarrow 0$  to max bees  $b_n$  do
13:     $x_i \leftarrow$  Select solution  $s_i$  based on  $p_i$ 
14:     $v_i \leftarrow$  Generate new solution with  $x_i$  and  $p_i$ 
15:    Evaluate  $v_i$  with fitness function  $f(v_i)$ 
16:    Apply greedy selection between  $v_i$  and bee  $ob_i$  current solution
17:  end for

```

Algorithm 8 Basic Artificial Bee Colony Algorithm (continued)

```

18:   if there is an abandoned solution for a scout bee then
19:       Replace with solution generated with equation 5.8
20:   end if
21:    $t \leftarrow t + 1$ 
22: end while

```

The algorithm starts off by generating a set number of possible solutions. The number of solutions is equal to the number of employed bees. Each starting solution is evaluated using a fitness function. The operations that perform these functions can be observed to occur from lines 1 – 3.

At first each bee is assigned to one of the initial generated solutions (food sources). Hence the bees start off as employed bees and each bee has in its memory a particular possible solution with an associated nectar amount. The algorithm can now be considered to be initialised, and therefore the algorithm enters the next phase, which is where the actual optimisation and search procedure occurs. This phase stretches from lines 5 – 22.

From lines 6 – 10, each employed bee modifies its particular solution based on local information, which is also referred to as visual information in the literature. The modified solution is then tested to determine its nectar amount, i.e. the fitness of the generated solution is calculated. The employed bee then compares the newly generated nectar amount with the nectar amount of the solution in the bee's memory. If the newly generated solution has a better nectar amount, the bee replaces the current solution in its memory with the newly generated solution.

After all the employed bees have determined whether to keep the newly generated solution or keep the one in memory, they then need to communicate to the rest of the bee hive the nectar amount of the food sources that they occupy. This phase is where the waggle dance occurs and can be observed in algorithm 7 from lines 12 – 17.

Each onlooker bee then selects which food source it will move towards based on a probability. The probability takes into account the nectar amount that was communicated through the waggle dance by a particular employed bee. The probability is calculated using equation 5.6 which is discussed in section 5.4.2.

Once an onlooker bee has selected a food source based on the calculated

probability, it then starts to search the immediate neighbourhood of the selected food source for other food sources. The neighbouring food sources are generated using equation 5.7 which is discussed in section 5.4.2. This procedure of generating neighbouring food sources by an onlooker bee can be observed in line 14 in algorithm 7.

The onlooker bee then applies the same procedure as an employed bee with regard to determining if the newly generated food source should be remembered or discarded. The bee does this by evaluating each generated neighbouring food source to determine its nectar amount, which is then compared to the nectar amount of the food source in the bee's memory.

In the last phase of the algorithm (before the next iteration starts) the algorithm determines which food sources have been abandoned by the bees. A food source in the algorithm can be abandoned if, for a certain number of iterations the food source has not improved, meaning its nectar content has not increased. When a food source is abandoned the employed bee that occupied the particular food source transitions to a scout bee.

A scout bee aims to replace the abandoned food source with a new food source. In algorithm 7 this occurs in lines 18 – 22. The scout bee uses equation 5.8 to generate a new food source. It then transitions to an employed bee and occupies the newly generated solution. The newly generated food source will now also be evaluated to determine its nectar amount as the rest of the employed bees do at the start of the next iteration.

5.4.4 ABC algorithm on the FAP

The ABC algorithm and all its variants are relatively new. To date it has only been applied to a select few problems such as the traveling salesman problem.

As yet, no research has been done to apply the ABC algorithm to the FAP. The following critique is based on a theoretic implementation of the ABC algorithm on the FAP. Based on this evaluation the following obstacles can be identified if one were to apply the algorithm to the FAP:

Food source representation — Each food source can either represent a frequency plan or it can be a particular cell and a collective of food sources represents a frequency plan. If each food source is a frequency plan the

algorithm will require a fair amount of memory, since as onlooker bees select it, they will start searching for neighbouring solutions. These neighbouring solutions are *also* frequency plans. However, if each food source is a cell, this would require less memory. The problem with the latter approach is that the bees would then single out one cell as the optimum, since they do not know that all food sources collectively represent a plan and each cell is actually unique.

Scout bee generation — When a food source is abandoned a scout bee needs to generate a new food source to take its place. In particular with the FAP, the newly generated solution cannot be completely random. The scout bee needs to incorporate knowledge already gained by the colony operating on different food sources, otherwise a completely random solution might not be even nearly lucrative enough for the rest of the bee colony to consider if it contains no knowledge gained by the algorithm.

Knowledge sharing — As a food source becomes more popular due to its high nectar amount, more onlooker bees will select it. By selecting the food source the onlooker bee will then proceed to search in its neighbourhood for better solutions. Therefore the bees are disregarding previously gained knowledge while searching for neighbouring sources on *other* food sources. If a food source represents a complete frequency plan, a previous food source a bee operated on might have had one or more cells that were assigned to their optimal frequencies. Due to the larger majority of the cells not being optimised, these *good* cells are overshadowed. Thus due to the *bad* cells overshadowing the good cells, the food source nectar amount becomes lower. As a consequence of the low nectar amount by the food source the bees abandon the food source and those optimal cells are lost.

The above obstacles present real relevant challenges that would require new techniques to be developed. As the algorithm has not been applied to a wide variety of problems and taking into account the above obstacles, it is difficult to gauge if the ABC is well suited to be applied. The algorithm first needs to be applied to a wider set of problems. Once the algorithm has matured in the research behind it the algorithm can be revisited and applied to the FAP.

In this section the various obstacles one would encounter when applying

the ABC class of algorithms to the FAP were identified. This concludes the discussion on the ABC algorithm. The next section deals with the particle swarm optimisation algorithm.

5.5 Particle Swarm Optimisation (PSO)

5.5.1 Introduction

PSO is population-based stochastic search technique that was developed by Kennedy and Eberhart in 1995 [104]. The basic model of the algorithm is based on simulations done to recreate the natural behaviour of a flock of birds [128].

In the early stages of the particle swarm development, simulations were developed to closely model the stigmergy (see page 98 for a discussion on stigmergy) exhibited when a flock of birds cohesively move as one and are able to suddenly change direction in a unpredictable, graceful manner, only to regroup as one observed entity [57].

As the “leading” bird of the flock changes its movements the information is shared with all birds in the immediate vicinity of the leading bird. As the information is shared locally among birds, each bird modifies his own movement to that of the leading bird’s movement [57].

Because birds obtain information by observing their neighbouring birds, the stigmergy can be deemed to be of a physical nature; therefore the particular stigmergy used by birds is sematectonic stigmergy (see page 98).

The simulations based on this behaviour of the flock allowed researchers to discover the underlying patterns that governed the way birds are able to share information about the general movement of the flock. Based on these patterns and simulations, the particle swarm algorithm emerged into an optimisation algorithm [34].

In the algorithm a particle is an individual. A group of particles, referred to in the literature as a swarm, are moved through the search space of the problem the algorithm is being applied to. Each particle changes its movement based on information shared with it by neighbouring particles in the swarm [33, 34].

As information is shared among particles, the success of one particle ripples through the rest of the swarm and each particle is able to utilise

shared information that leads to success of another particle. Thus, each particle's own personal experience and knowledge of the search space has an effect on its neighbouring particles [33, 34].

Two variants of the initial PSO algorithm were developed namely the global PSO and local PSO. The only difference between the two algorithms is how they go about sharing information with the rest of the swarm. The sharing models of these two algorithms along with other sharing models is discussed in the PSO characteristics subsection [90].

In the swarm, each particle is a potential solution that is represented by a D-dimensional vector [57, 96]. As a particle moves through the solution space, it continually evaluates its current position and adjusts it accordingly to move in the general direction of its own personal best position and the position of the best particle in its neighbourhood.

A particle evaluates its current position by using a heuristic function, or in more evolutionary algorithm terms, a fitness function. The fitness value indicates to the particle how far it is from an optimal position [34].

As a particle moves through the solution space it keeps a memory of the personal best position it has achieved since the start of the algorithm. In the literature and in the algorithm this personal best position is referred to as *pbest* [90].

Most of the research done on PSO has focused on the convergence of the algorithm as well as improving the diversity [33]. Some of these improvements and modifications is discussed in the subsection on PSO characteristics.

5.5.2 PSO Characteristics

The defining characteristics of the PSO algorithm that are discussed in this section include *neighbourhood topology*, *the particle swarm*, *movement of particles* and *keeping particle velocities in check*.

The Particle Swarm

The initialising of particles in a swarm are the same as traditional population-based evolutionary algorithm use to initialise their respective populations [133]. At the start of the algorithm the swarm is initialised by randomly generating possible solutions that will represent the position of particles [34].

The PSO algorithm in some aspects resembles evolutionary algorithms like the genetic algorithm since it also has a population that operates in the problem space in search of an optimal solution. However unlike the GA the PSO does not continually generate new solutions to be reinserted into the swarm to increase diversity [117]. Thus the swarm size needs to be adjusted to get an optimal representation of the search space because as particles move in the swarm, the diversity among particles decreases rapidly as information is shared [33, 34].

Diversity decreases as the swarm moves because, with each iteration, the swarm converges towards the neighbourhood best position [33, 34, 57] since the velocity equation directs the general movement of a particle in the direction of the neighbourhood best [33, 34, 57]. The velocity equation is discussed below [33, 34, 57]. If the best position in the neighbourhood does not change as the algorithm processes more iterations, a larger portion of the swarm will soon occupy the a position which is a weighted average between the neighbourhood and each individual particles personal best [33, 34, 57].

Diversity among the particles in the swarm is important and at least initially must be maintained for the search space to be explored adequately. Depending on the neighbourhood topology used diversity can be increased. In the next section neighbourhood topologies is discussed.

Neighbourhood topologies

The neighbourhood topology (also called social structure) used by a PSO algorithm dictates how information is shared among the particles of the swarm. According to Engelbrecht [33] there are 6 neighbourhood topologies. Each topology will now be listed and a short description will also be given.

- star** — With the star topology all particles within a swarm are interconnected with each other. Any one particle can communicate with any other particle in the swarm [33]. Using this topology each particle is attracted towards the best solution found globally by the swarm [33].
- ring** — Using the ring topology each particle communicates with its immediate adjacent neighbour [33]. Each particle aims to mimic the best solution found within its neighbourhood [33]. Neighbourhoods are allowed to overlap in the ring topology [33]. Overlapping enables

information sharing among neighbourhoods and facilitates the swarm in converging to a single solution [33].

wheel — In the wheel topology there is a central particle with which all other particles in the swarm are connected to [33]. The other particles are isolated and can only share information with the central particle [33]. The central particle utilises information about all the particles in its neighbourhood and adjusts its own best position accordingly [33]. Depending on whether the new best position represents a better solution than previously, the central particle shares the information to its neighbours [33]. A consequence of the limited sharing that occurs in the wheel topology, propagation of good solutions through the swarm is slowed [33].

pyramid — The pyramid topology interconnects particles so that their connected structure resembles a three-dimensional pyrami [33]d.

four clusters — With this topology four clusters are formed. Each cluster containing interconnected particles [33]. Between each cluster there exists only two connections to other clusters [33].

von Neumon — Using this topology all particles are connected in a structure that resembles a grid [33]. This structure has been shown to enable the swarm to produce better results than other neighbourhood topologies [33]

Based on the neighbourhood topologies defined the global PSO uses a star neighbourhood to allow for information to be shared all particles in the swarm. The particle whose position in the search space indicates the best solution found by the swarm is denoted as *gbest* [33, 34, 90].

In contrast, the local PSO follows the process of natural birds more closely and uses the ring neighbourhood for information sharing [33, 34, 90]. Hence, particles only share information with their immediate neighbourhood and not with the whole swarm. The best particle in the local PSO is denoted as *lbest* [33, 34, 90].

Movement of particles

A particle moves with a certain velocity through the solution space. As the information is shared the particle must take advantage of the newly

gained knowledge and therefore needs to adjust its own velocity to match the movement of the swarm. The particle updates its own velocity to move in the direction of the *gbest* shared position, its own *pbest* position and its current heading.

A particle needs to systematically explore the solution space; therefore when the particle needs to update its personal velocity, it does not use all the information it has available, otherwise it will start to cycle solutions. The particle uses a certain amount of global information together with a certain amount of local information to produce a direction and new velocity [33, 34, 90, 96].

The amount of global knowledge is referred to as the *social* component [33, 34, 90, 96]. The amount of personal information used by a particle is referred to as the *cognitive* component [33, 34, 90, 96].

The velocity calculation of each particle is where the optimisation procedure occurs in the PSO algorithm. It is the only means by which the PSO algorithm searches the solution space, since it is the only means by which particles are moved to new positions in the search space [34].

The velocity update is where the personal experience of a particle and the knowledge gained through social sharing are incorporated. By updating the velocity of a particle, the particle is steered into a more promising direction. The velocity of a particle is updated based upon on equation 5.9.

$$\vec{v}_i(t+1) = \vec{v}_i(t) + c_1\vec{\phi}_1(t)[pbest_i - \vec{x}_i(t)] + c_2\vec{\phi}_2(t)[gbest_i - \vec{x}_i(t)] \quad (5.9)$$

$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t+1) \quad (5.10)$$

where $\vec{v}_i(t+1)$ is the new velocity of particle i for the next time step $t+1$. The cognitive component is represented by the term $c_1\vec{\phi}_1(t)[pbest_i - \vec{x}_i(t)]$ where c_1 is the cognitive coefficient. The social component is represented by the term $c_2\vec{\phi}_2(t)[gbest_i - \vec{x}_i(t)]$ where c_2 is the social coefficient (discussed on page 127) [33, 34]. Each of the respective components c_1 and c_2 controls how much neighborhood information is used in the calculation of the new velocity.

The variables $\vec{\phi}_1$ and $\vec{\phi}_2$ are vectors containing random scalar values in the range $[0, 1]$. The current position of a particle in the solution space at time step t is represented by parameter $\vec{x}_i(t)$ [33, 34]. After the new velocity is calculated the position of the particle is updated for time step $t+1$ using

equation 5.10 [33,34]. The velocity update can be visually depicted as shown in figure 5.2.

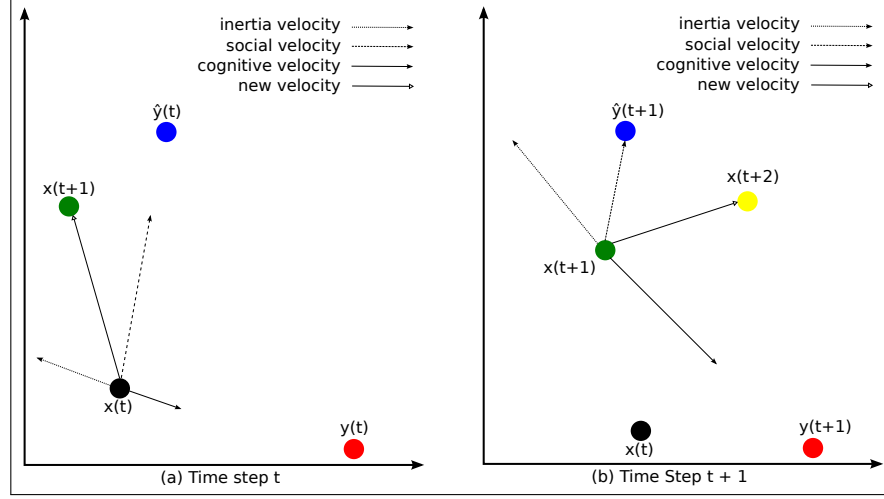


Figure 5.2: Visual particle velocity update [33, 34, 90, 96]

As discussed earlier gbest is the best position the particle has occupied since the start of the algorithm. In the literature there are two defined methods of determining gbest. The most common method used is where gbest is the best position obtained by a particle in the swarm since the start of the algorithm; thus long-term knowledge dictates the best position found which favours exploitation [33, 34].

The second method of determining the gbest is the best particle position occupied by any particle in the swarm, in the *current* iteration of the algorithm; thus short-term knowledge dictates the best position found which favours exploration [33, 34].

Most of the literature has concentrated on the velocity of the particle because it is the main function performing the optimisation. In research done by Ratnaweera et al. [96] particle positions in the solutions space are continually monitored. If the particle appears to be stagnant in the solution space, the velocity is first updated, and then the particle is reinitialised with a random position. The new position of the particle is then updated with the new velocity, thus knowledge of the discarded particle is retained by using the velocity it had [96].

In research done by Kalivarapu et al. [58] a PSO algorithm is developed

that seeks to incorporate the pheromone notion of ACO algorithms into the velocity updating of particles. The premise of the method is to allow greater sharing of information about promising areas between particles. The algorithm developed by the authors achieved promising results, with it in certain cases finding solutions faster and also better solutions than other PSO algorithms [58].

Other research done by Monson and Seppi [80] is more concerned with how the particle is presented. In the general PSO algorithm, particles have no physical form or volume and so particles in the swarm move through each other. The authors changed this in their algorithm by letting each particle have a radius around itself. This means that as particles move through the solution space and another particle at a certain time step occupies the same space, the particles are said to collide. As one would expect, when a collision occurs both particles are deflected into random directions [80]. At a greater expense of computational time due to constant collision detection, the PSO gains greater exploration in the solution space.

Finally, in the research by Lenin and Monan a PSO algorithm is developed that is called the attract repulse particle swarm optimisation (ARPSO). The algorithm continually monitors the solutions in the swarm. If it picks up that a certain percentage of the swarm is stagnating, it activates the repulse state. In the repulse state particles are repelled from other particles in the swarm, which facilitates greater exploration. After a certain number of iterations, the algorithm returns to its default state, where particles attract each other. The state of attraction facilitates exploitation [64].

Keeping velocity in check

As can be observed in equation 5.9 the new velocity is added to the old velocity. The velocity of particles can get very large, especially for those particles that are far from the pbest and gbest positions. Large velocities are necessary for early exploration.

Velocities should be kept in check since if a particles velocity becomes too large, it can overstep the search spaces boundaries and produce infeasible solutions [33]. Thus the velocity of a particle needs to be bounded to ensure that its movement within the search space stays within acceptable bounds. One of the means to bound the velocity of a particle is by applying equation

5.11. The equation is applied on the velocity before its position is updated [33].

$$\vec{v}_i(t+1) = \begin{cases} \vec{v}_i'(t+1), & \text{if } \vec{v}_i'(t+1) < \vec{V}_{max} \\ \vec{V}_{max}, & \text{if } \vec{v}_i'(t+1) \geq \vec{V}_{max} \end{cases} \quad (5.11)$$

$$\vec{V}_{max} = \delta(\vec{x}_{max} - \vec{x}_{min}) \quad (5.12)$$

In equation 5.11 \vec{v}_i' is calculated using equation 5.9 for the global PSO. Where \vec{V}_{max} is the maximum allowed velocity and $\delta \in (0, 1]$. The values \vec{x}_{max} and \vec{x}_{min} are the respective minimum and maximum position vectors in the domain the algorithm is being applied to [33]. The value of δ is very problem dependent and must be carefully chosen to maximise the exploration-exploitation trade-off [33]. The use of velocity clamping is not mandatory and should be considered only if the problem requires it [33]. Finally there is no guarantee that velocity clamping will prohibit velocities becoming too large [33]. There is still a chance, just to a smaller extent [33].

Velocity clamping is not the only developed means by which the exploration-exploitation trade-off of the PSO can be control. Consider the case when an object moves with a certain velocity it carries momentum. If the object were to suddenly change direction, momentum would for a certain period still move the object in the previous direction. Inertia weight seeks to add type of behaviour to the particles of the PSO algorithm. The velocity update equation with added inertia is formulated in equation 5.13 [33].

$$\vec{v}_i(t+1) = w\vec{v}_i(t) + c_1\vec{\phi}_1(t)[pbest_i - \vec{x}_i(t)] + c_2\vec{\phi}_2(t)[gbest_i - \vec{x}_i(t)] \quad (5.13)$$

Inertia (w in equation 5.13) was added to the general velocity update equation in an attempt to control the exploration and exploitation of the PSO as well as eliminate the need for velocity clamping [33]. Although the inertia component did succeeded in enabling the control of the PSO's exploration-exploitation in the search space, the need for velocity clamping could not be eliminated [33].

For values of $w > 1$, the inertia of the particle is increased. With increased the particle will explore more but it also more likely to leave the boundaries of the search space leading to infeasible solutions [33].

When $w < 1$ and depending on the values of c_1 and c_2 each time a particles velocity is updated a certain amount of momentum is lost. The

particle seems to slow down, allowing it to exploit the current solution space in finer detail [33]. This is not always the case, as $w < 1$ can also lead to the particles in the swarm diverging never to converge on an optimal solution.

To allow for a greater trade-off between exploration and exploitation, the inertia value can be made dynamic. Exploration is favoured early on in an optimisation algorithm and exploitation later on the algorithm when it is near an optimum; thus various methods that are either linear decreasing or non-linear decreasing have been developed that modify the inertia component as the algorithm moves around in the solution space [33, 34].

Finally, a similar inertia type component was developed from the analysis of particle dynamics [33]. This new component is called the *constriction coefficient* and, like the inertia above, also modifies the velocity update equation slightly [33, 34, 80].

This modification can be observed in equation 5.14, which is the standard velocity equation with the constriction coefficient. The constriction coefficient is formulated in equation 5.15 [33, 34, 80].

$$v_i(t+1) = \chi[v_i(t) + c_1\phi_1(t)[pbest - x_i(t)] + c_2\phi_2(t)[gbest - x_i(t)]] \quad (5.14)$$

$$\chi = \frac{2\kappa}{|2 - \phi - \sqrt{\phi^2 - 4\phi}|} \quad (5.15)$$

The constriction coefficient is represented by the value *phi*. The constriction coefficient evaluates to an ever-decreasing value between $[0, 1]$. By using the constriction coefficient the PSO algorithm is also guaranteed to converge for values of $\phi \geq 4$ and $\kappa \in [0, 1]$. As with the inertia discussed above, high values of κ allow for greater exploration and slow convergence, whereas low values of κ force the algorithm to exploit the solution space and converge quickly [33, 34, 80].

5.5.3 Flow of the Algorithm

The general concepts that are evident in the PSO algorithm have been covered. Using these concepts a general overview will now be given of the PSO algorithm flow using algorithm 9 as a reference point. Note that for the algorithm presented a start neighbourhood topology is used and the algorithm is applied on a minimisation problem.

Algorithm 9 Basic Global Particle Swarm Optimisation Algorithm [34]

```

1: Initialize  $s_n$  swarm
2: while Stopping condition not met do
3:   for each particle  $p_i \leftarrow 0$  in  $s_n$  do
4:     Evaluate particle with fitness function  $f(p_i)$ 
5:     if  $f(p_i) \leq pbest(p_i)$  then
6:       personal best of  $p_i$  to  $f(p_i)$ 
7:     end if
8:     if  $f(p_i) \leq f(gbest)$  then
9:        $gbest \leftarrow f(p_i)$ 
10:    end if
11:  end for
12:  for each particle  $p_i \leftarrow 0$  in  $s_n$  do
13:    update velocity of  $p_i$  with equation 5.9
14:    update position of  $p_i$  with equation 5.10
15:  end for
16: end while

```

The PSO algorithm starts off by initialising the swarm of particles. Each particle is randomly assigned a certain position in the problem space. After the swarm has been initialised the algorithm enters the optimization or search phase, which starts in line 2.

Before the swarm can start moving around in the problem space, it first needs to determine the gbest particle as well as each particle's own pbest position. Therefore as can be observed in line 3, each particle's current fitness $f(p_i)$ is determined using a problem-specific fitness function. The fitness determines the lucrativeness of the current position a particle occupies in the problem space.

Once the fitness of a particle's position is calculated, the algorithm needs to determine whether the current position of the particle is its pbest since the algorithm started. This comparison can be seen occur in line 5.

If the fitness of the currently held position is indeed better than the previous personal best of the particle, then the new position is remembered as the personal best for that particular particle, as can be observed in line 6.

Regardless of whether the personal best of a particle has been updated or

not, the algorithm performs another comparison also utilising the calculated fitness of the current position of the particle. The algorithm uses this fitness to also determine whether the current position of the particle is the best in the *entire* swarm, i.e. whether it is the *global* best (gbest). This comparison occurs in line 8.

If the position of the particle is indeed the best position in the entire swarm, the algorithm replaces the current gbest with the position of the current particle being evaluated, as seen in line 9.

After the swarm has been evaluated, each particle should have a personal best and the swarm should have a global best. The swarm is therefore ready to move around in the problem space, which occurs in algorithm 9 from lines 12 – 15.

For each particle in the swarm the algorithm determines the particle's new velocity, as can be observed in line 13. The velocity of a particle is calculated using equation 5.9.

Once the velocity of a specific particle has been calculated, the particle is ready to move to a new position. Moving a particle from its current position to a new position using the calculated velocity is done by applying equation 9 and occurs in line 14 in algorithm 9.

After the whole swarm has been moved, the algorithm continues to the next iteration to evaluate the new positions. This process occurs until certain stopping criteria are met.

5.5.4 PSO on the FAP

The PSO algorithm is also a relatively new algorithm and has been applied to only a handful of NP-Complete problems, including the FAP. In this dissertation the PSO algorithm will be utilized on the FS-FAP to try and produce optimal solutions.

Only two groups have conducted research where the PSO has been applied to the FAP to produce a near optimal solution. The research concentrated on the MS-FAP variant of the FAP, and so the aim of their algorithm was to reduce the span of frequencies used. The problem this dissertation is concerned with is the FS-FAP where the amount of interference generated needs to be minimised.

To date, no PSO algorithm has been designed to operate on the FS-FAP variant. Therefore, the interest in the research mentioned above is more to do with how the authors went about encoding a particular frequency plan as a position for a particle, than with the actual optimisation procedure.

In the research presented by Elkamchouchi et al. [32] a PSO algorithm is applied to produce optimal solutions for the MS-FAP. The way the authors went about assigning frequencies in their algorithm is known as frequency exhaustive assignment (FEA). This method works by first generating a list of calls, called a *call list*, denoting calls that occur in the system [32].

The FEA method then iterates over the calls in the list and assigns the lowest possible frequencies to the calls without violating interference constraints [32]. The authors note that the specific frequency that is assigned to a particular call depends heavily on the order the calls are in the list [32].

Because of the success of the PSO on the MS-FAP, for this dissertation the PSO algorithm was selected as the primary means by which to address the FAP.

The algorithm also makes extensive use of knowledge gained by the various particles as they search the problem space. Depending on the values used for w , c_1 and c_2 , a particle does not only keep personal history (with $pbest$), but the swarm as a whole keeps a history of the best particle (neighbourhood best). Thus with regard to FAP, it is possible that even though a particle might be in an overall bad position, it might have some small bit of good knowledge being overshadowed by bad knowledge. Through the extensive use of historical knowledge good information is more likely to be shared or kept slightly longer in the algorithms collective knowledge.

For this current research the PSO was applied to the FS-FAP on the COST 259 benchmark problems. The approach by the authors in the above literature could not be used. FS-FAP is concerned with interference generated and there are some constraints which cannot be broken which are mentioned in section 3.8.3 page 51. In contrast with MS-FAP, the performance measure is explicitly the number of constraints violated not interference.

5.6 Summary

In this chapter three swarm intelligence algorithms were discussed. The general flow of the ant colony optimisation algorithm was described with the

help of a pseudo code as well as how the algorithm came about. The defining characteristics of the algorithm were identified and a literature review was given of the ACO being applied to the FAP.

The second swarm intelligence algorithm was the artificial bee colony optimization algorithm. How the algorithm was developed and how it performs its search in a problem space were explained. A diagram also outlined the general flow of the ABC algorithm.

A series of defining characteristics was explained. Each characteristic is a defining attribute of the algorithm that makes it unique with regard to other algorithms. No literature is available on the algorithm being applied to the FAP since to date no research has been conducted on such an ABC algorithm.

This chapter concluded with the most important algorithm, which is used in this dissertation on the FAP, namely the particle swarm optimization algorithm. The flow of the PSO algorithm was described in 9. Furthermore, characteristics that make the algorithm unique were explained, and a literature review was given of the PSO algorithm being applied to the FAP.

Part II

Discussion and results of implementing a PSO algorithm on the FAP

Chapter 6

Applying the PSO to the FAP

6.1 Introduction

PSO, as discussed previously (see page 123), is an algorithm that is largely based on the flying behaviour exhibited by a flock of birds. This is why the core of the algorithm is based upon vector mathematics, with new positions and velocities being calculated after each iteration of the algorithm. Thus a D-dimensional vector represents each particle position and is then simulated by flying through the D-dimensional space using the velocity equation (see section 5.5.2 on page 127).

Most of the problems to which PSO has been applied to date have been problems where the position of particles has a constant D-dimensional space. In formal terms, therefore, *the dimensionality of a particle position in its entirety is constant*. Therefore, each particle has a position which is defined in a set dimension like 2D where a position is represented at x and y coordinate pairs.

This constant dimensionality introduces an intriguing problem if one wants to apply the PSO to an inherent multidimension problem like the FAP. This chapter deals with how the PSO was applied to the FAP.

Firstly, the particle position is represented in the frequency planning domain is defined. This definition of the particle position is important because it plays a central part in the movement of particles through the Frequency

planning domain. A description is then given of how each position is evaluated as well as the fitness function that the PSO will use in the FAP domain.

Arguably the most important part of the swarm is how the velocity of a particle is calculated and then moving it to a new position in the problem space. The velocity update is important, as it is the primary means by which the algorithm searches problem space.

As was discussed in section 5.5.4, applying the PSO to the FAP introduces a variety of challenges. One of the challenges is how exactly one “flies” a frequency plan towards another frequency plan. This is an important question that needs to be addressed as the PSO algorithms have no other way of searching the problem space by any other means.

To be able to allow the PSO to operate in the FAP space custom velocity functions had to be developed to enable the particles of the swarm to move. These velocity functions that were developed will expanded on in section 6.4.

Developing custom velocity functions for the PSO was simply not enough to achieve good results with the PSO . Therefore more innovations needed to be made to improve the solution quality of the PSO . In section 6.5 a new mechanism is presented for selecting the global best which enabled the PSO to get better fitness values and therefore direct the swarm more towards better solutions.

Finally, the chapter will conclude with how the swarm utilises history to produce better results to enable the PSO to further improve the solution quality.

6.2 Position in the Frequency Planning Domain

In this section a description is given of what a position is in the frequency planning domain. First a frequency plan is defined and the general structure to represent such a plan is provided. The section will conclude with the hard and soft constraints and how the constraints aid in creating a frequency plan that is suitable for a network.

A frequency plan, is almost exactly what the name implies: A plan that outlines frequency usage for a mobile telecommunication network. The benchmark problems that were used to test the developed PSO all pertained to cellular phone networks and were presented in section 3.8. For cellular

networks, the frequency plan outlines which frequency must be allocated to which transceiver.

With this basic definition, the problem is deceptive as one naturally assumes that there are an infinite number of frequencies that can be used or the number of frequencies available for assignment is more than the number of transceivers in the network.

The reality is that there are only a finite number of frequencies available for cellphone transmissions, as was discussed in chapters 2 and 3. Hence a regulatory body needs to assign wireless spectrum to cellphone network operators for use in their networks. A regulatory body is needed because, if a network operator uses just any frequency it wants, it is bound to interfere with someone else also utilising the same frequency.

A network is not assigned to the entire wireless spectrum for wireless communication, but rather only a subset is assigned to the network. If one observes the FAP benchmark problems the PSO was applied to (see section 3.8) for instance Siemens1, the allotted spectrum is from frequency 16 to 90. Which gives the network operator 74 frequencies to use in its network without considering other constraints.

Besides the electromagnetic constraints that are also applicable here, there are regulatory constraints, for instance frequencies in the spectrum that are by no means allowed to be used. These frequencies are referred to as globally blocked frequencies and are hard constraints. An in-depth discussion of these constraints was given in section 3.5.

As discussed in chapters 2 and 3, a cellphone network is divided into a number of cells, and each cell requires a certain number of transceivers to service its corresponding area.

The number of transceivers is based on the expected volume of traffic that a particular cell will experience at peak network usage. Some cells might be located in highly populous areas, which means the potential traffic that cell might need to handle during peak network usage is very high and thus the cell has more than one transceiver to handle the potential traffic. With cells that are located in areas that have a low population, the potential traffic the cell might experience during peak network usage is low and thus the cell only has one transceiver to handle potential traffic.

Based on the amount of traffic a cell needs to handle, the number of transceivers differs; thus in a frequency plan not all cells have the same

number of transceivers, otherwise a frequency plan can be modelled as a series of constant D-dimensional vectors, where the D represents the number of transceivers. As can be seen in figure 6.1 a cellular network can have

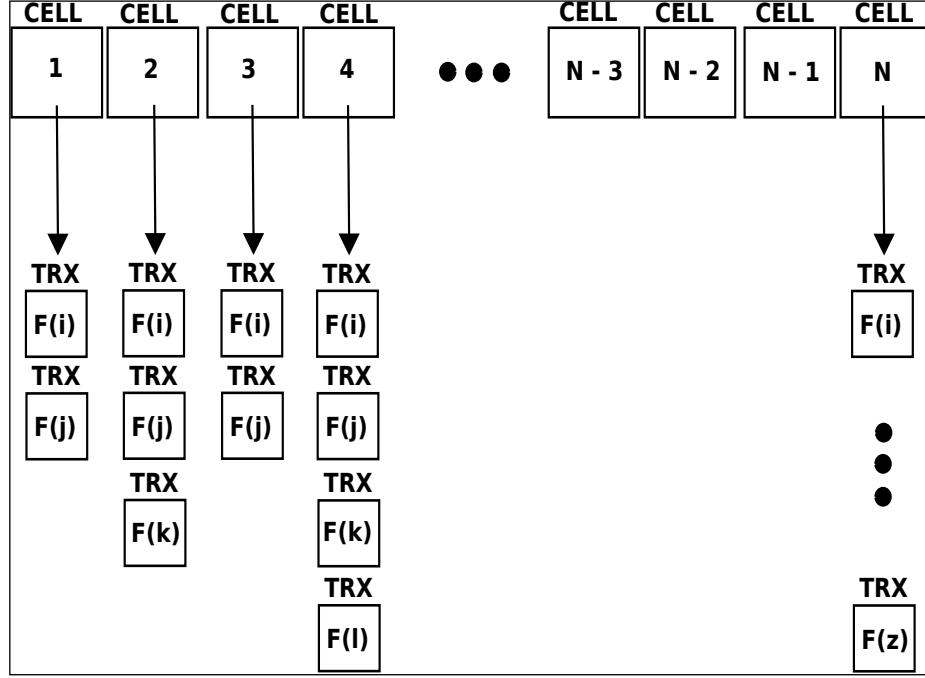


Figure 6.1: The structure of a frequency plan

any number (N in the figure) of cells to attain the desired coverage over the geographical landscape. In the COST 259 benchmark problems the cellular networks have a large number of cells that range from 500 to more than 1 000.

The most important part of the plan is the actual transceivers within each cell. In figure 6.1 it can be clearly seen how the number of transceivers (TRXs) varies from one cell to the next. $F(i)$ is a frequency at position i from the available usable spectrum.

Based on the structure of the plan depicted in figure 6.1 there is no concept of which cell interferes with which other cell and if there is indeed interference, the extent of this interference. Not all this information is part of the plan. Instead this information, for the purpose of this research is supplied by the COST 259 benchmark.

The interference information is referred to as the interference matrix. A definition of the structure of an interference matrix was given in section 3.5.

As discussed, each entry references two cells' entries: Cell A and Cell B. Along with the entry the amount of interference that occurs when Cell B interferes with Cell A is also listed¹.

A frequency plan is a possible solution to the FAP. Therefore in the PSO that was developed each particle's position in the solution space is represented by a frequency plan. As illustrated in figure 6.1, a frequency plan is just a series of cells, where each cell has a set of transceivers; thus in the PSO algorithm a plan is actually represented as an array of cells. This enables the algorithm to access particular cells in a plan by index as can be observed in listed algorithms 12 and 17.

Before the particles can actually start to move around in the FAP space, they first need to be assigned positions. In the developed PSO listed in algorithm 6.2 line 1 the first operation that the algorithm executes is to initialise all the particles in the swarm. A particle position in the algorithm is initialised by assigning it a random position; thus a frequency plan (representing a position) is randomly generated by the algorithm.

Algorithm 10 The FAP PSO Algorithm

```

 $s_n$  = Initialize Swarm  $s_n$ 
while Termination criterion not met do
    EvaluateSwarm( $s_n$ )
    UpdateGlobalBest( $s_n$ )
    UpdateSwarmMovement( $s_n, gbest$ )
end while

```

The position is purely random in that the only considerations made by the position generator are that valid frequencies are assigned to transceivers installed at cells. Thus the generator does not check whether a frequency has already been assigned in the current cell or any other considerations. The intended purpose of the generator is just to place a particle in the problem space, not the premature start of the optimisation process.

Since particles are able to occupy positions in the FAP space the PSO algorithm is now able to move them around in the problem space. As mentioned previously, moving particles through the frequency plan solution space introduces an interesting problem due to the multidimensionality

¹Interference occurs based on the electromagnetic constraints as defined in chapter 3

of a plan. A discussion of how particles are moved from one position to another through the solution space is provided in section 6.4

In the next section an explanation is given on the fitness function that determines the desirability of a particular particle’s position or rather the frequency plan its position represents.

6.3 The Fitness Function

The fitness function rates the desirability of a particular particle’s position in the problem space.

As discussed in the previous section, the COST 259 benchmark problems provide an interference matrix that lists the total amount of interference that occurs when a pair of cells interfere. As outlined in the structure definition (see section 3.5) each entry in the interference matrix defines a pair of cells that are said to interfere, along with two additional values. The first value is referred to as co-channel interference and is the total amount of interference that will occur on the communication link when the allocated frequency of one transceiver is equal to a transceiver in the other cell that is listed in the interference matrix.

The second value is called adjacent channel interference and it is the total amount of interference that will occur on the communication link when the allocated frequency of a transceiver in one cell differs by 1 from another frequency allocated to the transceiver from the other cell that is listed in the interference matrix.

Particles move towards other particles because the other particles have indicated (through information sharing) that the positions they occupy are very lucrative and thus they have found potentially good solutions. The only way particles can know the lucrateness of the position they occupy is if the position is evaluated with a fitness function. Thus the lucrateness of a position is actually the fitness value obtained from the fitness function.

Since a particle position is defined as a frequency plan, a procedure is needed that calculates the fitness of a frequency plan. With the FS-FAP the primary concern is to keep interference to a minimum. Therefore in the PSO that was developed the fitness value is the total amount of interference generated by all the cells with their allocated frequencies.

The evaluation procedure goes through each pair of cells defined in the interference matrix where it looks up both cells in the frequency plan. The second cell is said to interfere with the first cell. Therefore each transceiver in the first cell is checked against all the transceivers of the other cell. Depending on whether the frequencies differ from each other, the fitness procedure adds either co-channel or the adjacent channel interference to a summing variable. This procedure is mathematically defined in chapter 3 (see page 50 for the formal equation) and algorithm 11 is the pseudocode of the implement equation used by the PSO .

As can be seen in algorithm 11 not all interference values are added to the total amount of interference variable. The COST 259 benchmarks define a minimum tolerable interference variable. This means that if a given interference value is either equal to or less than this defined value the interference generated is negligible and can be disregarded as it will not have a noticeable impact on the communication link.

In the next section a description is given of how particles are moved from one iteration to the next using frequency plans as positions in the solution space.

6.4 Velocity Function for Frequency Planning

The elocity function is arguably the core of the PSO algorithm. It is the procedure by which particles in the swarm move from one point to another in the solution space.

The velocity function does not blindly move a particle from one point to another, but instead it takes the particle history into account as well as the best particle in the swarm. Therefore the velocity function is the core means by which the swarm explores the solution space. A more thorough explanation is provided in section 5.5.2.

The development of a velocity function that is suitable for particles to move from one frequency plan to another is covered next. The section will start off with the first velocity function that was developed. With each method discussed, the problems associated with it will also be mentioned.

This section will conclude with the second method that was developed and that is also the primary method the developed PSO uses.

Algorithm 11 FAP Cost Function

Require: normalCell**Require:** interferingCe;;

```

1:  $totalInterference \leftarrow 0$ 
2: for Each TRX  $trx_i$  in interferingCell do
3:   for Each TRX  $trx_j$  in normalCell do
4:      $interference \leftarrow 0$ 
5:      $difference \leftarrow |trx_i - trx_j|$ 
6:     if difference is 0 then
7:       if coChannelInterference  $\leq$  minInterferenceThershold then
8:          $interference \leftarrow interference + 0$ 
9:       else
10:         $interference \leftarrow interference + coChannelInterference$ 
11:      end if
12:    else
13:      if difference is 1 then
14:        if adjChannelInterference  $\leq$  minInterferenceThershold
15:        then
16:           $interference \leftarrow interference + 0$ 
17:        else
18:           $interference \leftarrow interference + adjChannelInterference$ 
19:        end if
20:      end if
21:    end if
22:     $totalInterference \leftarrow totalInterference + interference$ 
23:  end for

```

6.4.1 Movement in the Frequency Planning Domain

The standard velocity equation works on the basis of vector mathematics. Each particle has a velocity and position, which is represented by a standard mathematical vector. The standard equation alters the direction of the particle to move to a more promising position in the solution space that is in the general direction of the global best particle and a previous personal best position the particle held..

Vector mathematics has standard basic operations defined for adding, subtracting and multiplying; hence applying the PSO to problems that are either mathematical functions or problems that map well to the vector domain is a defined process. With regard to the frequency planning domain an important question needs to be answered: How can one multidimension frequency plan be moved to another?

With any difficult problem it is better to break the problem down into its most basic constructs and then solve each piece individually until the problem as a collective is solved. This technique is also commonly known as divide and conquer. This technique was first applied to the nature of a frequency plan.

A frequency plan is a plan that consists of a series of different cells that are in use in the network. The plan specifies the frequencies that each individual transceiver installed at a cell must use for communication. Thus a frequency plan can be broken up into three important constructs:

1. A plan is a list of different cells.
2. Each cell in a plan has a list of transceivers that it has installed.
3. Each installed transceiver has a single number allocated to it called the frequency. This frequency is used for communication.

Now that a frequency plan has been broken up into its constructs, the question of how to move one frequency plan to another can be rephrased. How does one move a frequency allocated to a *transceiver* in a particular *cell* of one frequency plan to another frequency of the *same* transceiver and cell in another *different* plan? An important realisation needs to be noted here. In the PSO at any one time the algorithm is only considering two positions and for the FAP the two positions are frequency plans. Both plans

are *identical* except for the specific that frequencies transceivers use. Thus a cell that exists in the one plan, also exists in another plan. Both the cells have exactly the same number of transceivers installed; only the frequencies each individual transceiver uses for communication differ.

Using this realisation, the conclusion can be made that the velocity equation can only work with the frequencies assigned to transceivers. Therefore a potential velocity equation mechanism needs to operate on the finest granularity of a frequency plan, i.e. the frequencies.

The principle on which the first velocity method developed is based, is for the movement of the swarm to be at a much finer granularity and hence movement is based on frequencies. Therefore when a particle needs to move towards a global best particle, the velocity procedure goes into the intricate details of the particle wanting to move and the global best particle. The procedure goes into each cell defined in the frequency plan represented by the standard particle as well as the global best particle to be able to access each installed transceiver.

To be able to move from one frequency plan to another by utilising the standard velocity equation, the equation needs to be broken up into its smaller operations. In this way, small operations can be developed that perform the same function as the individual parts.

The velocity equation in section 5.5.2 can be broken up into the following parts:

- **Subtraction** — $pbest - x_i(t)$ — (SubtractionResultPbest) and $gbest - x_i(t)$ — (SubtractionResultGbest)
- **Multiplication** — $c_1\phi_1 * \text{SubtractionResultPbest}$ — (MultiPbestResult) and $c_2\phi_2 * \text{SubtractionResultGbest}$ — (MultiGBestResult)
- **Addition** — $v_i(t) + \text{MultiPbestResult} + \text{MultiGBestResult}$

There are no mathematical constructs that define how two frequency plans are added together or subtracted, let alone multiplied. As discussed earlier, a frequency plan is just a series of cells that have frequencies and these frequencies are numbers that internally are just integers and there are mathematical constructs that define how two integers should be added, subtracted or multiplied. Both velocity methods that were developed utilise the basic

principle that on a fine granularity of a frequency plan one is merely working with integers.

The first velocity method is listed in algorithm 12. Velocity method 1 works on the principle of moving one cell in a particular frequency plan to the same cell in a different frequency plan. As noted earlier, the cells are the *same*, but the frequencies that have been allocated to each transceiver within a cell differs. Thus in velocity method 1, each cell has an array of transceivers. The array of transceivers contains the individual frequency numbers that have been allocated to a cell.

Algorithm 12 Velocity Method 1

Require: currentParticle

Require: globalBestParticle

```

1:  $pbest \leftarrow$  currentParticle position
2:  $gbest \leftarrow$  globalBestParticle position
3:  $localCoeff \leftarrow getLocalCoefficient()$ 
4:  $globalCoeff \leftarrow getGlobalCoefficient()$ 
5:  $pBestSubtractResult \leftarrow$  Subtract current Particle position from pbest
6:  $gBestSubtractResult \leftarrow$  Subtract current Particle position from gbest
7:  $a \leftarrow$  Multiply localCoeff with pBestSubtractResult and a random value
8:  $b \leftarrow$  Multiply globalCoeff with gBestSubtractResult and a random value
9: if first time velocity is calculated for current Particle then
10:    $currentParticle.Velocity \leftarrow$  Add a and b
11: else
12:    $abAdditionResult \leftarrow$  Add a and b
13:    $inertia \leftarrow getInertia()$ 
14:    $newVelocity \leftarrow$  Add abAdditionResult and currentParticle.Velocity
15:    $currentParticle.Velocity \leftarrow$  Multiply inertia with newVelocity
16: end if
17:  $currentParticle.Position \leftarrow$  Add currentParticle.Velocity and current-
   Particle.Position
18: SanatizePosition(currentParticle.Position)

```

Velocity method 1 actually moves one array of transceivers to another array of transceivers. This is why, if one observes lines 5 – 17 in algorithm 12, each particle position is passed to methods that are named Subtraction, Multiplication and Addition. As discussed in the previous section, each

position of a particle is actually a frequency plan.

The first method that is used to calculate the velocity of a frequency plan, is the first basic operation defined in the velocity equation, namely subtraction, as can be observed from algorithm 13. The particular algorithm expects two positions to be given to it: the position it must be moved from and the position it must be move to. Because each position is a frequency plan that contains an array of cells, a for-loop is started on line 1 that allows the algorithm to go through each cell that is defined in the frequency plan. Each cell has a number of transceivers, each with an assigned frequency.

Algorithm 13 Subtract One Position From Another

(Method 1)

Require: fromPosition

Require: toPosition

```

1: for Each cell  $c$  in fromPosition do
2:   for Each transceiver  $f_{trx}$  in  $c$  do
3:      $t_{trx} \leftarrow$  Get same  $f_{trx}$  in toPosition
4:      $f_{trx} \leftarrow f_{trx} - t_{trx}$ 
5:   end for
6: end for
```

To be able to subtract two transceivers from each other, the algorithm needs access to a particular transceiver of a cell. The algorithm gains this access by entering another for-loop based on the number of transceivers a cell has, as can be observed to occur in line 2. Within this for-loop the actual subtraction of transceivers occurs.

The algorithm first obtains the exact same transceiver in the toPosition frequency plan. This operation is quick, as the two plans are identical except for the frequencies assigned to transceivers for each cell. Thus the algorithm is able to refer to the cell and specific transceiver in the toPosition plan by the same index it utilises to access the cell and transceiver in the fromPosition.

Once the toPosition plan transceiver frequency value has been obtained, the algorithm performs a standard integer subtraction as seen in line 4 of algorithm 13. After the subtraction algorithm is completed, a position is returned which is the result of subtracting the fromPosition from the toPosition. Using this result the velocity method 1 algorithm is now ready to

apply the next operation in the velocity update equation, namely multiplication which is depicted in algorithm 14.

Algorithm 14 Multiply Position by a Value (Method 1)

Require: fromPosition

Require: value

Require: mustRandom

```

1: for Each cell  $c$  in fromPosition do
2:   for Each transceiver  $trx$  in  $c$  do
3:     if must multiply with random number then
4:        $trx \leftarrow trx * value * random()$ 
5:     else
6:        $trx \leftarrow trx * value$ 
7:     end if
8:   end for
9: end for

```

The multiplication algorithm is where the local Coefficient and the global coefficient² defined for the PSO are multiplied into a position, i.e. frequency plan. As with the subtraction algorithm the multiplication algorithm enters two for-loops. The first for-loop is to access each cell defined in the frequency plan. The second loop allows the algorithm to access each transceiver within a particular cell. These for-loops can be observed to occur in algorithm 14 in lines 1 and 2.

Once the algorithm is able to access each individual transceiver and get its assigned frequency, it is able to perform the multiplication operation. As can be observed in the algorithm in line 3 the algorithm checks whether it needs to multiply the frequency by an additional random number besides the *value* passed to the algorithm. The reason this is done is that multiplication is required in the PSO in the following cases:

- The inertia case — $0.5 * v(t + 1)$, where 0.5 is the inertia value and $v(t + 1)$ is velocity that has already been calculated for a particular particle t . Note, a velocity that has been calculated is still a frequency plan.

²Local coefficient is the cognitive component and the global coefficient is the social component. These components are discussed in section 5.5.1

- Standard velocity calculation randomisation case — As can be observed from the velocity equation 5.9 and also from the multiplication bullet point on page 148, the equation requires that a position be multiplied by a coefficient c_1 or c_2 and then a random value ϕ_1 or ϕ_2 .

Regardless of which case is executed, the actual operation that is performed is integer multiplication, which therefore means that even though the inertia and random numbers are decimal, the fractional component of the result is discarded. Channels are integers so the loss of the fractional component is warranted as it is of no use.

As per the last bullet point on page 148 the last basic operation that occurs in the velocity equation is an integer addition operation. Algorithm 15 is the addition procedure used by velocity method 1 (algorithm 12).

Algorithm 15 Add One Position to Another (Method 1)

Require: fromPosition

Require: toPosition

```

1: for Each cell  $c$  in fromPosition do
2:   for Each transceiver  $f_{trx}$  in  $c$  do
3:      $t_{trx} \leftarrow$  Get same  $f_{trx}$  in toPosition
4:      $f_{trx} \leftarrow f_{trx} + t_{trx}$ 
5:     Bound( $t_{trx}$ )
6:   end for
7: end for
```

In the addition algorithm, as with multiplication and subtraction, two for-loops are started as can be observed in lines 1 – 2. The algorithm first starts iterating through all the cell’s present in the fromPosition frequency plan and then within the cell for-loop starts another for-loop which iterates through each cells transceivers. Inside the transceiver for-loop the algorithm actually performs the addition of two frequencies.

Two frequencies are added based on integer addition. Once the fromPosition cell transceiver is accessed, the algorithm retrieves the exact transceiver and cell in the toPosition in order to get its allocated frequency value. After both values have been retrieved the algorithm performs the addition as can be observed in line 4 of algorithm 15.

The addition operation is the last operation that occurs in the velocity equation and is also the only operation that occurs when the resultant velocity is applied to the current position of a particle as in equation 5.10. Since the addition operation is also used in the final position update of the particle, its last operation in the transceiver loop is a bound operation.

The purpose of the bound operation is to keep the frequencies within valid value ranges.

6.4.2 Keeping Frequencies Bounded

In the previous section the first velocity method that was developed was discussed. The velocity method is important as it calculates the direction and next position of a particle in the problem space, where the problem space is the FAP and a position of a particle is a possible frequency plan. With the velocity method defined, the swarm is now able to move around in the problem space, but this alone is not enough. The swarm has no concept of the constraints³ that are imposed inherently by the domain as well as the network for which a frequency plan is being created.

Due to the swarm not having a concept of the constraints a constraint handling mechanism needs to be used. Each of the methods discussed in this section fall under the category of using the *repair method* which is discussed in section 3.3

The first method that was developed was the BoundValue method and is depicted in algorithm 16.

³Constraints in the FAP are discussed in chapter 3

Algorithm 16 BoundValue Method

Require: Position

```

1: for Each cell  $c_i$  in Position do
2:   for Each transceiver  $trx_j$  in  $c_i$  do
3:      $trx_j = |trx_j|$ 
4:     if  $trx_j \geq maxChannelValue$  then
5:        $trx_j = minChannelValue + (trx_j \bmod maxChannelValue)$ 
6:     else
7:       if  $trx_j \leq minChannelValue$  then
8:          $BoundValue(trx_j + minChannelValue)$ 
9:       end if
10:    end if
11:  end for
12: end for

```

The constraints that are imposed are usually some part of the spectrum that may under no circumstances be used anywhere in the network, which as discussed in chapter 3 is known as globally blocked frequencies. Some frequencies that are only allowed to be used in certain parts of the network are known as locally blocked frequencies. Then of course the swarm also needs to take into account the electromagnetic constraints⁴.

The velocity function used by the PSO needs to be altered to make the swarm in some sense more aware of the domain it is operating in and hence keep the particle positions bounded within the allowable search space. By not bounding particle positions in the FAP problem space, transceivers of some cells might be assigned a frequency that is not allowed or not even allocated to the network. The PSO will accept this assignment since the fitness function operates on the assumption that under no circumstances will these invalid frequencies be assigned and thus does not penalise invalid assignments.

To keep assigned frequency values to transceivers in the allocated spectrum of the network a boundary check needs to be added to the velocity method. The purpose of the boundary check is to validate all assignments in a position, i.e. the frequency plan that a particle currently occupies, and if any assignments violate the defined boundary constraints then it must

⁴Discussed in section 3.5

take the violating value and modify it to be in the acceptable value range.

The boundary check that is used by velocity method 1 operates on the basis that for any range of values there is a defined lower bound (a minimum value) and upper bound (a maximum value). The frequency boundary check is only applied when one of the following conditions is met after the calculated velocity has been applied to the current position:

- If a frequency allocated to a transceiver is above the maximum allowable frequency (upper bound) given to the network.
- If a frequency allocated to a transceiver is below the minimum allowable frequency (lower bound) given to the network.

As can be seen in line 5 of algorithm 16, a mod operation is applied to the value to bring it within the allowable range. The integer mod operation is similar to integer division, the only difference being in the result that is produced. Division produces the result of two numbers being divided. Mod produces the remainder of two numbers being divided. If two numbers divide perfectly into one another there will not be any remainder; if the numbers do not divide perfectly into one another there will be a remainder.

$$10 \bmod 50 = 10 \quad (6.1)$$

$$60 \bmod 50 = 10 \quad (6.2)$$

$$50 \bmod 50 = 0 \quad (6.3)$$

$$100 \bmod 50 = 0 \quad (6.4)$$

$$35 \bmod 50 = 35 \quad (6.5)$$

As can be seen in the above example mod operations, any value that is modded will be kept in the range $[0, 50]$. With regard to frequencies, the following occurs: If, for instance, the maximum allowable frequency is 50, and the transceiver has a frequency value (after velocity) of 56, the 56 value is modded by 50 to produce a value of 6. This modded value is then added to the minimum allowable frequency. In essence, the value is wrapped around to always be within acceptable range.

The difficult case is when the frequency value is lower than the minimum frequency given to the network. This is because modding the frequency value has no effect. For example, if the lowest allowable frequency is 20 and the

transceiver value after movement is 15, modding the transceiver value of 15 by 20 has no effect. To solve this, the following options are then considered:

1. First subtract the lower value from the minimum allowable frequency. Then add the result to the minimum allowable frequency. The resultant value is checked again as to whether it oversteps the bounds of the maximum allowable frequency and is bounded accordingly.
2. Add the lower value to the minimum allowable frequency. The resultant value is checked as to whether it oversteps the bounds of the maximum allowable frequency and is bounded accordingly.
3. Repeatedly subtract the lower value from the maximum allowable frequency until the resultant frequency is within the acceptable frequency range.

An important notion to consider is that, based on the velocity equation, it is entirely within the realm of possibility that a frequency value after movement might contain a negative value. As can be seen in line 3 of algorithm 16 the algorithm solves this problem by first taking the absolute value of the negative frequency value. The boundary check then treats the now positive frequency value as a normal value that needs to be bounded.

It is better to use frequency index values rather than actual frequency values for frequency plans, and this is explained below.

6.4.3 Using Indices instead of Frequencies

As discussed in section 6.4 and as can be observed from algorithm 12, the first velocity method that was developed for the PSO worked with raw frequency values. This is not ideal since upon closer inspection the frequency range that the swarm used to move around was indeed incorrect. The bound value algorithm only keeps frequencies within a minimum and maximum allowable frequency range, but globally blocked frequencies and locally blocked frequencies can be in between this minimum and maximum frequency range.

With velocity method 1 and the PSO using raw frequency values, the swarm increasingly moved towards allocating these blocked frequencies to transceivers since the fitness function does not penalise the use of blocked or invalid frequency values. This is partly due to the fact that these values

are under no circumstances allowed to be used and thus the fitness function is not designed to check for these values explicitly to impose a penalty.

Frequency plans that utilise these blocked frequencies are invalid and cannot be used. If a network were to use a plan that uses blocked frequencies, it could cause unexpected interference to other services and the network could be fined by the governing body that controls the spectrum. A bare minimum requirement then is that the PSO must generate valid frequency plans and hence swarm particles can only occupy valid positions. The following options were presented to solve the problem of particles moving towards invalid positions and hence having invalid frequency plans:

1. Modify the fitness function to penalise a frequency plan if it uses any globally blocked frequencies or locally blocked frequencies.
2. Instead of letting the swarm work with raw frequency values, rather let it work with indices of an array. The array index values indicate positions in an array that has been pre-filled with only *valid* frequencies. Thus the swarm then moves around in a range from 0 to F , where F is the size of the frequency array.

With the first solution, the fitness function would have to be modified to impose a penalty if a prohibited frequency value were used. The first proposed solution was disregarded because it introduces complexity which can be completely avoided with the second proposed solution.

With the second solution the fitness function does not have to be modified and the boundary check is simplified since there is no need to check for a lower bound. The boundary check now only has to check for negative index values and whether the upper bound, which is now the size of the array, is violated.

The second method is also a method known in constraint handling as *preserving feasibility* (see section 3.3 for a discussion on constraint handling methods). By using only valid frequencies the search space is even more constricted to only feasible solutions.

Working with index values rather than with raw frequency values led to a second velocity method. Algorithm 17 is the pseudocode for the second velocity method. The second velocity method differs from velocity method 1 due to it being designed to only work with indices of an array.

Algorithm 17 Velocity Method 2

Require: currentParticle**Require:** globalBestParticle

```

1:  $currPos \leftarrow$  currentParticle position
2:  $pbestPos \leftarrow$  currentParticle best position
3:  $gbestPos \leftarrow$  global best particle position
4: for Each cell  $c$  in  $currPos$  do
5:    $c_{pbestPos} \leftarrow$  Same cell in  $pbestPos$ 
6:    $c_{gbestPos} \leftarrow$  Same cell in  $gbestPos$ 
7:    $movedIndices \leftarrow$  MoveIndices( $c_{pbestPos}, c, c_{gbestPos}$ )
8:   if First time velocity is calculated then
9:     ApplyVelocity( $c, movedIndices$ )
10:  else
11:     $newVelocity \leftarrow$  CalculateIndexVelocity( $currPos.velocity, movedIndices$ )
12:    ApplyVelocity( $c, newVelocity$ )
13:  end if
14: end for
15: SanatizePosition( $currPos$ )

```

Velocity method 2 also differs from method 1 due to the manner in which it applies the velocity equation. Velocity method 1 applies the velocity equation in stages. Each stage is applied to the entire position, i.e. frequency plan before applying the next stage. With velocity method 2 the algorithm first enters a for-loop as in line 6. Within this for-loop the algorithm obtains the current cell, the cell stored in its previous best-held position $pbestPos$ and the same cell stored in the global best position $gbestPos$.

Once all the correct cells have been obtained the algorithm calls a method named *MoveIndices* (which is presented in algorithm 18) to which all the cells are passed.

In the MoveIndices algorithm three cells are passed to it. The current cell of the velocity algorithm that is being moved, as is the same cell that is stored in the previous best-held position of the particle and also the same cell in the global best position. The algorithm starts off by iterating through all the transceivers installed at the current cell as can be seen in line 1 of algorithm 18. In lines 2 – 3 the algorithm obtains the frequency values from the $pbestCell$ and $gbestCell$.

Algorithm 18 MoveIndices

Require: pbestCell**Require:** currCell**Require:** gbestCell

```

1: for Each  $trx_i$  in currCell do
2:    $pbestTRX_i = \text{Get } trx_i \text{ in pbestCell}$ 
3:    $gbestTRX_i = \text{Get } trx_i \text{ in gbestCell}$ 
4:    $r1 \leftarrow \text{Random double number}$ 
5:    $r2 \leftarrow \text{Random double number}$ 
6:    $localCoeff \leftarrow \text{getLocalCoefficient}()$ 
7:    $globalCoeff \leftarrow \text{getGlobalCoefficient}()$ 
8:    $a \leftarrow localCoeff * r1 * (pbestTRX_i - trx_i)$ 
9:    $b \leftarrow globalCoeff * r2 * (pbestTRX_i - gbestTRX_i)$ 
10:   $trx_i \leftarrow a + b$ 
11: end for

```

After the MoveIndices algorithm has obtained the frequency values it is ready to apply the velocity equation. As can be seen in lines 8 – 10 the algorithm applies the standard velocity equation 5.9 formulated in chapter 5.

Once the MoveIndices algorithm has completed, the currCell variable now stores the calculated velocity. Algorithm 17 uses the return calculated velocity to update the cell position of the current particle. The algorithm accomplishes this by first determining whether the particle has had a previously calculated velocity. If it is the first time the velocity has been calculated, the algorithm simply applies the velocity by using algorithm 19.

Algorithm 19 ApplyVelocity

Require: currPos**Require:** velocity

```

1: for Each index  $c_i$  in currPos do
2:    $c_i \leftarrow c_i + velocity_i$ 
3: end for

```

If the particle does have a current velocity, the algorithm first executes algorithm 20 as can be observed to occur in line 11 of algorithm 17 before applying the calculated velocity with algorithm 19. Velocity method 2 first

executes algorithm 20 to apply the concept of inertia⁵.

Algorithm 20 CalculateIndexVelocity

Require: currVelocity

Require: newVelocity

- 1: **for** Each value v_i in currVelocity **do**
 - 2: $inertia \leftarrow getInertia()$
 - 3: $v_i \leftarrow v_i + (inertia * newVelocity_i)$
 - 4: **end for**
-

As can be observed in algorithm 20 the CalculateIndexVelocity method loops the current velocity. Within the loop the algorithm retrieves the inertia value that needs to be applied, after which the inertia value is multiplied by the newVelocity value.

After the CalculateIndexVelocity algorithm has finished executing, the concept of inertia has been applied to the velocity. Velocity method 2 therefore applies the velocity using the ApplyVelocity method listed in algorithm 19 to the current cell found in the particle position to move it in order to obtain a new position.

The observant reader might have noticed that in all the presented and associated velocity algorithms the procedures operate and store the result in the current particle position that is currently being worked on. Naturally one might think that this overwrites the position information currently in that position. In the developed PSO algorithm when velocity is calculated the respective velocity methods operate on clones of the original position. The original position is only used once the velocity has been calculated and the particle has actually moved.

Another important note, is that if one analyses velocity method 1, the switch to using index values instead of raw frequencies does not affect the method’s operation. Since the method only requires explicit knowledge on the data to operate on in the BoundValue algorithm, the algorithm is still able to calculate velocity.

It is up to the algorithm designer to update the bound value method to use the array index bounds rather than the raw lower and upper bounds of

⁵Inertia is discussed in chapter 5 in the subsection entitled “Inertia Weight” of section 5.5.2.

the frequencies. The BoundValue algorithm needs to be updated since it is the primary means by which velocity method 1 ensures valid positions.

Both the velocity methods that are utilised by the developed FAP PSO algorithm have now been explained with corresponding pseudocode. Why the developed PSO was modified to operate on frequency index values rather than frequency values was also explained. All the algorithms that enable the main algorithm to accomplish particle movement with indices were also discussed. The PSO is hence able to move the particles in the FAP using two different methods, but for a particle to be moved it needs a personal best and most importantly, a global best to move towards. The next section deals with on how the developed PSO algorithm differs from the standard PSO with regard to selecting a global best.

6.5 Building a Global Best

Selection of the global best particle by the swarm is a very important procedure. After the swarm has determined which particle has achieved the best position, the swarm enters the velocity function phase.

As discussed previously each particle position is then modified to move in the general direction of the global best and personal best position. Therefore the global best acts as a beacon for the rest of the swarm in the solution space to indicate where good solutions seem to be for the rest of the swarm.

Algorithm 21 Standard Gbest Selection in FAP PSO

Require: swarm

Require: gbest

```

1:  $gbestCost = \text{Evaluate}(gbest)$ 
2: for Each particle  $p_i$  in swarm do
3:    $cost = \text{Evaluate}(p_i)$ 
4:   if  $cost \leq gbestCost$  then
5:      $gbestCost = cost$ 
6:      $gbest = p_i$ 
7:   end if
8: end for return gbest

```

Initially the FAP PSO algorithm used the standard method for selecting the global best particle from the swarm and did not differ at all from the

traditional global PSO algorithm. The standard global best selection is listed in algorithm 21.

As can be observed in lines 2 – 8 the FAP PSO algorithm loops through all the particles in the swarm and applies the fitness function to evaluate the fitness of the particle’s position. The fitness value is also referred to as the cost. In the FAP PSO the cost or fitness value of a particle position is the amount of interference the frequency plan that represents the particles position generates.

A low cost value is preferred to a high cost value, since a low cost value indicates low interference. In lines 4 – 6 of algorithm 21 the FAP PSO algorithm determines whether the current particle position has a lower cost value than the current global best particle position.

If the current particle position evaluates to a lower cost value than the stored global best, the algorithm replaces the current global best with the current particle being evaluated, which in the algorithm is p_i .

Selecting the global best by evaluating the position as a whole seems to be a natural fit. As outlined in the critical evaluation of each algorithm in chapters 4 and 5, some of the algorithms had a problem with regard to some cells or even transceivers overshadowing better cells or transceivers.

For this dissertation, overshadowing is a term that describes a scenario where a bad value of one part of the frequency plan is so large that it causes other smaller values within the frequency plan not to be considered.

As per the following example a few cells in a frequency plan might have the worst possible frequencies assigned to their respective transceivers, and other might have the best. Now the few cells with the worst frequencies generate a great deal of interference, whereas the cells with the best frequencies generate almost nothing.

When the example frequency plan is evaluated, the bad cells push up the cost value. The high cost value of the frequency plan causes the PSO algorithm to disregard the whole plan. By discarding the whole plan the FAP PSO algorithm loses the knowledge gained on the few cells that had their best frequencies assigned to their respective transceivers.

In the traditional method of selecting the global best, a particle is actually selected as the swarm best because it contains fewer overshadowing cells or transceivers, and potentially good frequency assignments are lost.

The FAP PSO therefore needed to exploit the knowledge that the fitness function exposes much more thoroughly. The information exposed by the fitness function allows one to see what effects certain frequency assignments have on the interference of the cell when assigned to the individual transceivers. To make better use of this fitness information two methods were developed for the FAP PSO , each one being finer grained than the other.

1. Besides the particle storing its fitness or cost, the particle also needed to store the interference generated by an entire cell due to the frequencies allocated to its installed transceivers.
2. Besides the particle storing the total fitness, it also needed to store the interference generated by a frequency allocated to a particular transceiver of a cell.

With both these methods, the global best selection scheme needs to be changed to allow the swarm to take advantage of this newly exposed information. As discussed, initially the FAP PSO used the standard global best selection scheme listed in algorithm 21, but now with these new methods, a global best position is no longer selected, but built.

Before the FAP PSO is able to build a global best, the way a particle stores its evaluated fitness needs to change. For the standard global selection scheme, the particle only needs to store one fitness value that is a result of evaluating the whole frequency plan. To be able to build a global best as described above, the fitness value cannot simply be one lump sum representing interference. Instead in the FAP PSO algorithm the interference generated by every transceiver is stored.

The FAP PSO is able to know the performance of every single frequency allocated to a particular transceiver and also compare the allocation with other similar transceivers in other frequency plans.

Each global best scheme developed for the FAP PSO is more finely grained than the other with regard to what the scheme uses to build a gbest. Algorithm 22 uses interference information of cells to build a gbest. Algorithm 23 uses the interference generated by each transceiver installed at a cell to build a gbest. Since each cell has transceivers, the second algorithm is therefore more finer grained than algorithm 22.

Algorithm 22 Building Global Best with Cells

Require: gbest**Require:** swarm

```

1: for Each particle  $p_i$  in swarm do
2:   for Each cell  $c_j$  in  $p_i$  do
3:      $gbestcell_j = \text{Get cell } c_j \text{ in gbest}$ 
4:      $cellCost = \text{Get total interference for cell } c_j$ 
5:      $gbestCellCost = \text{Get total interference for cell } gbestcell_j$ 
6:     if  $cellCost \leq gbestCellCost$  then
7:        $gbestcell_j = c_j$ 
8:     end if
9:   end for
10: end for

```

Algorithm 23 Building Global Best with Transceivers

Require: gbest**Require:** swarm

```

1: for Each particle  $p_i$  in swarm do
2:   for Each cell  $c_j$  in  $p_i$  do
3:     for Each transceiver  $trx_k$  in  $c_j$  do
4:        $gbestTrx_k = \text{Get } trx_k \text{ in } c_j \text{ in gbest}$ 
5:        $trxCost = \text{Get total interference for } trx_k$ 
6:        $gbestTrxCost = \text{Get total interference for cell } gbestTrx_j$ 
7:       if  $trxCost \leq gbestTrxCost$  then
8:          $gbestTrx_k = trx_k$ 
9:       end if
10:    end for
11:   end for
12: end for

```

Each algorithm will now be discussed since the difference between them is subtle. Algorithm 22 was the first global best building scheme which was developed and is discussed first. The algorithm starts off in line 1 by iterating through all the particles in the swarm. For each particle the algorithm enters another loop as can be observed in line 2 of algorithm 22.

The second loop of the algorithm iterates through all the cells in the

frequency plan that represent a position of the particle. Once the algorithm has obtained a cell it looks for the same cell in the gbest frequency plan. Since the algorithm now has the current cell and the gbest cell it is able to compare the two cells based on the interference their respective transceivers generate.

The interference generated by the cell is retrieved and then compared as can be observed to occur in line 6. If the current cell has a lower interference (cost) value than the same cell in the global best plan, then the algorithm replaces the cell in the global best with the current cell.

The second algorithm is the finer grained algorithm and was developed because, after analysing the algorithm using cells, it was concluded that it is possible that a single bad frequency allocation to a transceiver within a cell can overshadow other potentially good frequency allocations to other cells within the cell.

Algorithms 22 and 23 have very similar flow, except that the second algorithm has an extra for-loop. In line 3 of algorithm 23 a third for-loop is started, which iterates through all the transceivers a particular cell c_j has installed. The algorithm then obtains the frequency allocated to the same transceiver in the global best frequency plan.

Once both transceivers have been obtained, the algorithm determines the interference (cost) their respective frequency allocations generated. Using the cost values the algorithm determines whether the current transceiver frequency allocation generates less interference than the frequency allocated to the same transceiver in the global best frequency plan.

If the current transceiver frequency generates less interference, the algorithm then proceeds to replace the transceiver frequency in the global best with the current transceiver frequency. Thus it can be seen that algorithm 23 utilises individual transceivers to build a global best.

Initially when the FAP PSO algorithm was tested using both of these global best schemes, the PSO did not produce noticeably better results. This was due to the algorithm at each iteration discarding the interference or cost information calculated in that iteration and making it zero. Making the cost values zero does initially seem correct, but effectively what is happening is that the algorithm is discarding knowledge gained by that iteration.

To enable this information to direct the swarm a bit more, the FAP PSO algorithm was modified to not reset the interference values for every

transceiver and cell to 0. Instead, the interference values for an iteration are now added to the previous iteration interference values stored by the cell and transceiver.

By letting interference values compound after each iteration the PSO becomes much more aggressive. This is because as the interference compounds bad decisions made by the swarm for a particular particle become progressively worse as the swarm progresses through more iterations.

With compounding interference values the FAP PSO was able to produce much better positions and had lower total interference (cost) than all previously generated positions by previous FAP PSO algorithms.

The FAP PSO algorithm is able to produce better results by allowing particles to keep a history of their previous movements. This is covered in the next section.

6.6 Keeping History

In the traditional PSO history is kept by using the particle personal best position to direct the next movement of the particle. Other methods such as inertia also allow history to direct the movement of the particle. With regard to the developed PSO on the FAP, the algorithm also uses these concepts. But these concepts are not able to effectively exploit the history of a particle since they have no concept of what combinations of frequency values have previously been used in a cell.

In the FAP PSO algorithm more historical information is kept. The algorithm accomplishes this by incorporating the concept of tabu lists from the TS algorithm. Using tabu lists a particle will be able to better exploit the problem space it currently finds itself in. In the FAP PSO algorithm, tabu lists were incorporated by adding to each cell a list which keeps track of each frequency value that has been assigned to the transceivers in the cell for 20 iterations.

Initially the FAP PSO algorithm calculated the velocity of a particle and then applied this to the current position of the particle. This moved the particle to its next position in the problem space. With tabu lists this movement step becomes more complicated.

Tabu lists are there to prevent cycling of movements to the same position. Thus to stop the particle from moving to a position that was previously

occupied, an extra check has to occur before the particle can occupy a new position. As can be seen in the two developed velocity methods in algorithms 12 and 17 the last step that occurs in both algorithms is that the SanitizePosition method is called.

The SanitizePosition method is listed in algorithm 24. Within this algorithm a particle's future position is first checked and sanitised before the particle is allowed to move to that position. The main purpose of this algorithm is to check if the future position has been occupied previously and hence is in the tabu list.

Algorithm 24 SanitizePosition

Require: currPosition

```

1: for Each cell  $c_i$  in currPosition do
2:    $tbList = \text{Get Tabu List of currPosition}$ 
3:    $\text{ResolveCollision}(c_i, tbList)$ 
4:    $\text{AdhereToSeparation}(c_i)$ 
5: end for

```

In the FAP PSO the tabu list check works slightly differently from what one would expect. As can be observed in algorithm 24, it enters a loop which iterates through all the cells in the position of the particle. Note that the position passed to the SanitizePosition algorithm is a *future* position; thus the particle does not yet occupy the position yet. Within the for-loop in line 2 the method *ResolveCollision* is called which is listed in algorithm 25.

Algorithm 25 ResolveCollision

Require: cell

Require: tabuList

```

1: for Each  $trx_i$  in cell do
2:   while  $trx_i$  exists in TabuList do
3:      $trx_i = \text{Generate random frequency}$ 
4:     if Collision not resolved after 20 attempts then
5:       Break out of while loop
6:     end if
7:   end while
8: end for

```

As can be observed in algorithm 25, when a frequency value is found

to exist in the tabu list a collision is said to occur. In the FAP PSO a collision means that the specific frequency value that has been assigned to a transceiver for a particular cell was found in the tabu list. Once a collision occurs, the algorithm tries to generate a new random frequency that can be assigned to the transceiver as can be seen to occur within the while-loop in lines 2 – 3.

The algorithm generates a new random frequency value and then checks to see if the generated value collides with the tabu list. If collision still occurs, the algorithm will generate another random frequency. As long as a collision occurs the algorithm will continually generate a new random frequency until it has attempted 20 random frequencies with no frequency colliding.

After 20 attempts the algorithm just accepts the last generated frequency as the new frequency. The maximum number of 20 attempts was selected through testing and can be increased at the expense of more computational time.

The resolution of collisions can be seen as a mechanism to increase the exploration of the PSO algorithm as well as to increase the diversity. By making certain frequency assignments to transceivers tabu the algorithm is forced to try new frequency assignments and thus explore more of the problem space.

Care must be taken to select a maximum size of the tabu List since one wants to keep enough history so that the problem space can be adequately exploited. The maximum tabu List size must be less than the number of available frequencies otherwise the algorithm will not be allowed to make any assignments.

Finally the maximum tabu List cannot be too large, since the checks the algorithm has to do to see if a value is tabu are very expensive. The operation is expensive, since the tabu list needs to be iterated through for each potential value to see if the frequency value is tabu.

By incorporating tabu lists and the collision resolving procedure, the efficiency of the algorithm reduces dramatically. To increase efficiency of the operations in the algorithm, the FAP PSO algorithm utilises parallelisation. Since the collision resolving procedure is very expensive it was one of the first operations to be parallelised. Other procedures that were also parallelised to

increase efficiency were the velocity and any other procedures which involved constraint checks.

By parallelising these operations the efficiency of the algorithm increased and it was able to produce results significantly faster. This is because parallelisation is a good fit to the now standard multicore CPUs in desktop computers.

With the parallelisation of the procedures a slight side effect was noticed. The randomness of the random number generator decreased. This effect was noticed because during testing the counter variable of the collision resolver was displayed on the console. When the value was being displayed on the console the FAP PSO algorithm produced much better results.

The reason for this is that outputting the variable inherently introduces a delay and therefore the random number generators in other threads have different seed values. Hence with a delay in each parallel thread the numbers generated by the random number generator are more distinct.

Due to how parallel threads are scheduled by the operating system, some threads might start off with similar seed values because in the FAP PSO algorithm the current time is used as a seed value⁶.

Keeping the delay counter variable displayed on the console introduced a delay in the collision resolving procedure. The reason the particular procedure was selected was that it was where the effect of delay was first noticed. After performing tests with delays of 5 milliseconds (ms), 10 ms, 15 and 20 it was found that 20 ms was the best-suited delay, as it gave just enough time for a reasonable distinction to be made between seed values used by other parallel threads.

In this section a discussion was presented on how the FAP PSO keeps additional history. The reason why the FAP PSO needs to keep more history was discussed as well as what mechanism the algorithm uses to store this information, namely tabu lists. Also covered was how the algorithm deals with collisions, which occurs when positions are in the tabu list. Finally collision resolution was explained with the aid of pseudocode of the algorithm that is utilised.

⁶This is the default behaviour of the .Net 4.0 random number generator

6.7 Summary

In this chapter an algorithm was presented based on the standard particle swarm optimisation algorithm to operate on the frequency assignment problem encountered in cellular networks.

At the beginning of the chapter it was explained how a frequency plan is represented by the algorithm for use internally. Reasons were given for choosing the particular representation in the algorithm.

One of the most important phases of the PSO algorithm is velocity calculation. The problem was outlined as to why the standard velocity calculation was unsuitable for the FAP. The customised velocity calculation used by the algorithm developed in this research was presented along with suitable pseudocode.

The chapter concluded with small additions made to the algorithm to improve performance and, most important of all, improve solution quality.

Chapter 7

Results

7.1 Introduction

In the previous chapter the FAP PSO algorithm was discussed. This algorithm was developed by modifying the standard PSO algorithm to operate on the FS-FAP. Thus far, no other PSO algorithms have been attempted on the FS-FAP. The only PSO algorithm that has been attempted in the FAP domain was discussed in chapter 5 and was not relevant to the study in this dissertation, as the PSO was applied to an entirely different FAP variant (MS-FAP). The PSO on the MS-FAP was not relevant because the performance measures and what the algorithms optimise differ.

With the FS-FAP, the main performance measurement is interference and the PSO aims to allocate frequencies in an optimal way to internally produce a frequency plan. On the other hand, the MS-FAP is concerned with the span of frequencies used and the performance measurement is based on the calls dropped. The main purpose of the PSO in the MS-FAP is to minimise the span of frequencies used and keep the number of dropped calls to a minimum.

In the previous chapter all the modifications that were made to the standard PSO were discussed. The modifications were made to enable to PSO to operate on the FAP. Two velocity methods were developed and in addition to the standard global best selection scheme, two additional global best selection schemes were put forward. The algorithm presented was benchmarked against the Siemens instances of the COST 259 benchmark.

In this chapter the results are given of applying the FAP PSO algorithm with its different velocity methods as well as different global selection schemes. This is followed by how the different velocity methods affected the PSO performance as well as how the global selection schemes affected the final produces results.

7.2 PSO COST 259 Siemens Results

The PSO was applied to the benchmarks on the following machine and frameworks:

- 4 GB RAM
- Windows 7
- Intel Quad Core CPU
- C# using .Net 4 Framework with Parallel Extensions

The FAP PSO was applied to Siemens1, Siemens2, Siemens3 and Siemen4 of the COST 259 benchmark suite. For more information about the nature of the benchmarks, the reader is directed to section 3.8.3.

For each benchmark, 12 results are presented. The following changes were made to the FAP PSO to obtain 12 different results:

- The two velocity methods that were developed for the PSO were tested.
- Each velocity method also used the three different global best mechanisms that were developed.
- The swarm size was set to 20, 50, 100, 500.
- Inertia was set to 0.5.
- Cognitive coefficient was set to 0.4.
- Social coefficient was set to 0.5.
- Each variant of FAP PSO was allowed 15 minutes on each of the benchmarks.

The velocity methods and global selection schemes were discussed in the previous chapter.

In the following sections the results obtained are tabulated.

7.2.1 Siemens1

In table 7.1 the results obtained using the FAP PSO with velocity method 1. In table 7.2 presents the results for the FAP PSO that used velocity method 2.

GBest selection	Population	Interference
Standard	20	106.855
Built with cells	20	39.945
Built with TRXs	20	114.121
Standard	50	102.666
Built with cells	50	34.689
Built with TRXs	50	116.150
Standard	100	104.649
Built with cells	100	38.252
Built with TRXs	100	126.630
Standard	500	106.132
Built with cells	500	44.171
Built with TRXs	500	118.911

Table 7.1: Siemens1 results using velocity method 1

GBest selection	Population	Interference
Standard	20	817.161
Built with cells	20	336.102
Built with TRXs	20	312.120
Standard	50	637.030
Built with cells	50	258.049
Built with TRXs	50	217.247
Standard	100	517.347
Built with cells	100	252.827
Built with TRXs	100	197.421
Standard	500	272.717
Built with cells	500	123.427
Built with TRXs	500	141.060

Table 7.2: Siemens1 results with velocity method 2

7.2.2 Siemens2

Below the results obtained by applying FAP PSO on the siemens2 benchmark is presented in table 7.3 and table 7.4.

GBest selection	Population	Interference
Standard	20	76.199
Built with cells	20	55.051
Built with TRXs	20	82.262
Standard	50	78.181
Built with cells	50	55.488
Built with TRXs	50	81.076
Standard	100	77.511
Built with cells	100	55.436
Built with TRXs	100	82.857
Standard	500	77.104
Built with cells	500	54.880
Built with TRXs	500	86.794

Table 7.3: Siemens2 results with velocity method 1

GBest selection	Population	Interference
Standard	20	217.025
Built with cells	20	316.679
Built with TRXs	20	111.535
Standard	50	240.186
Built with cells	50	197.572
Built with TRXs	50	103.896
Standard	100	228.486
Built with cells	100	195.921
Built with TRXs	100	100.298
Standard	500	236.228
Built with cells	500	96.605
Built with TRXs	500	96.832

Table 7.4: Siemens2 results with velocity method 2

7.2.3 Siemens3

Table 7.5 and table 7.6 below present the results obtained by applying the FAP PSO algorithm to the siemens3 benchmark.

GBest Selection	Population	Interference
Standard	20	135.078
Built with cells	20	50.543
Built with TRXs	20	141.538
Standard	50	138.418
Built with cells	50	51.300
Built with TRXs	50	144.444
Standard	100	138.774
Built with cells	100	50.652
Built with TRXs	100	157.552
Standard	500	141.105
Built with cells	500	54.417
Built with TRXs	500	148.643

Table 7.5: Siemens3 result with velocity method 1

GBest Selection	Population	Interference
Standard	20	1104.006
Built with cells	20	415.866
Built with TRXs	20	460.403
Standard	50	2197.760
Built with cells	50	1014.373
Built with TRXs	50	1182.665
Standard	100	764.894
Built with cells	100	311.646
Built with TRXs	100	316.998
Standard	500	361.118
Built with cells	500	167.185
Built with TRXs	500	186.705

Table 7.6: Siemens3 result with velocity method 2

7.2.4 Siemens 4

GBest Selection	Population	Interference
Standard	20	528.162
Built with cells	20	287.665
Built with TRXs	20	555.445
Standard	50	542.143
Built with cells	50	283.473
Built with TRXs	50	557.154
Standard	100	541.617
Built with cells	100	293.238
Built with TRXs	100	551.930
Standard	500	538.812
Built with cells	500	302.441
Built with TRXs	500	582.992

Table 7.7: Siemens4 results using velocity method 1

GBest Selection	Population	Interference
Standard	20	2300.40
Built with cells	20	1091.827
Built with TRXs	20	2966.514
Standard	50	1877.781
Built with cells	50	2197.760
Built with TRXs	50	1014.373
Standard	100	1182.665
Built with cells	100	814.950
Built with TRXs	100	930.631
Standard	500	1934.260
Built with cells	500	631.668
Built with TRXs	500	723.772

Table 7.8: Siemens4 results using velocity method 2

7.2.5 Comparison with best results in COST259

Below the best results achieved by the FAP PSO is compared to the best results obtained with each siemens problem as published on the FAP website [51].

Algorithm	Siemens1	Siemens2	Siemens3	Siemens4
FAP PSO	34.689	54.880	54.417	283.473
k-thin FAP	2.200	14.271	5.129	77.246
Dynamic Tabu Search	—	14.275	5.186	81.876
SAG FAP Tool	2.301	14.751	5.259	80.967

Table 7.9: FAP PSO results compared to best obtained results

7.3 The Performance of the PSO

In this section the effects of the changes made to the FAP PSO algorithm in terms of the results is discussed.

Firstly, the effect of the two developed velocity methods in terms of the results is described. In section 7.3.2 the three global schemes that are used by the algorithm are discussed. Finally this section concludes with the effect of a larger population size on solution quality rather than a small population.

7.3.1 Velocity Method 1 vs Method 2

The FAP PSO algorithm is able to utilise two different velocity methods to move the swarm around in the FAP space. The algorithms that implement these two methods were presented in section 6.4 and section 6.4.3.

By analysing the results, it becomes abundantly clear that velocity method 1 is by far the superior method for moving in the problem space. In each of the results, when comparing end fitness values produced by algorithm variants that use method 1 one can easily come to the conclusion that method 1 performs better than method 2.

As discussed in chapter 6, method 1 uses a stage-based approach when applying the velocity function whereas method 2 applies the velocity function as is without it being broken up. Based on the results, using a stage-based approach to apply the velocity function is far better than applying the velocity equation directly to the transceiver in the plan.

Method 1 works by moving the whole swarm through to each stage before applying the next stage in the velocity equation. Thus after each stage, the whole swarm is at the same phase of the equation which keeps the swarm structured.

With method 2, the whole velocity equation is applied to whatever value is exposed to be operated on. Thus when method 2 moves a particle the frequency plan is moved piece by piece to some destination in the problem space.

By applying the velocity equation it is difficult to control the algorithm search process. With the FAP PSO control is necessary as there are various constraints which must not only be avoided but also adhered to for the generated plan to be usable. With method 2 adding domain knowledge is difficult, since after the velocity equation has been calculated the particle is very close to being moved to a new position. All that still needs to be done, before a particle is moved, is to apply inertia, which means there is a little check that can be done to ensure that all the frequency values are within acceptable bounds.

By breaking the velocity equation up into smaller parts (stages) using method 1 the algorithm is able to direct and ensure that the swarm is moving in the general direction of valid frequency allocations.

Also the algorithm is able to embed domain knowledge earlier into the calculation of the velocity and is therefore able to intercept early on movements that will result in invalid frequency allocations at each stage of the velocity equation.

7.3.2 Different Global Schemes

In the previous chapter three global selection schemes were identified. The first global selection scheme uses the standard PSO selection and the particle with the best fitness is the global best. This scheme is called “Standard GBest”.

As discussed in section 6.5, using the standard gbest selection scheme is not preferable as it can lead to the swarm losing out on good frequency allocations due to overshadowing of frequencies¹.

¹Overshadowing is discussed in chapter 6

Even with overshadowing the standard global scheme does not produce bad results, which seems to indicate that overshadowing of frequency allocations does not impact the frequency plans as significantly as thought initially.

In addition to the standard gbest selection scheme, two other selection schemes were tested. It is actually incorrect to call these schemes selection schemes of gbest, since they build global best rather than select them.

By far the worst performing scheme is where the global best is built from transceivers. In every benchmark performed where this scheme was paired with a velocity method and population size, the algorithm was simply not able to produce any relatively good solutions. All possible solutions had high interference values, making them undesirable.

The bad performance of the build from transceivers scheme can be attributed to the granularity it uses to build a global best. As outlined in section 6.5 the scheme only considers the interference generated by a single frequency allocated to a transceiver. This would have worked well if there were some sort of guarantee that a particular transceiver would only be interfered with by one other transceiver.

In reality and in the Siemens4 benchmarks this is definitely not the case. More often than not, transceivers are interfered with by more than one other transceiver. Thus by only concentrating on a single case-by-case basis of frequencies allocated to transceivers, the scheme is discarding all other possible interferences.

It might select a frequency at one point as the best, since in that scenario, the interference generated with the only other transceiver that is considered at that point is low. But this particular frequency is too close on the spectrum to another frequency allocated to some other transceiver that also interferes. Due to the algorithm only considering individual cases, this potential interference with the other transceiver will not be noticed by the algorithm and it will go ahead in selecting the frequency as the best for the transceiver.

By analysing the results produced by the various FAP PSO algorithms, it can be concluded that the best global best selection scheme is by far the one in which cells are used to build a global best. With the cell selection scheme, the algorithm does not suffer the pitfall that is the reason for the transceiver gbest building scheme's bad performance.

As discussed the build from cells scheme uses cells to build a gbest, and thus each cell stores the interference that the frequencies allocated to its transceivers generate by interfering with other cells. As a cell interferes with other cells, the interference generated is added to the cell causing the interference.

After the PSO has calculated the fitness of all positions, each cell will contain the interference it personally has caused throughout the network to other cells. A cell with low interference means the frequencies that have been allocated to this particular cell are the best combination which causes the least amount of interference. Therefore, with the build gbest from cells scheme, the algorithm is able to make informed choices when selecting a cell to be included in the global best.

7.3.3 Population Size

As discussed in chapter 5 the population parameter of the PSO is a sensitive parameter. For problems with big search spaces, it is better to have a large population. A large population increases diversity, meaning more particles occupy different positions in the problem space and hence the space is better explored, which internally increases the likelihood of finding a better solution. Thus, in the results the population size was also varied.

In the results using a larger population did not significantly improve the possible solutions. In only one benchmark, namely Siemens2, the algorithm was able to produce a better possible solution than an algorithm that used a lower population size. In Siemens1, Siemens3 and Siemens4 the algorithm using a lower population was able to produce the best overall results.

Due to the complexity of the FAP it is difficult to use the PSO with a large population. The algorithm's efficiency greatly decreased with the increase in population, taking significantly longer to iterate through the same number of iterations as the algorithms with lower populations. The decrease in efficiency was not unexpected due to the function evaluations, movements and considerations increasing fivefold over the lower population size. What was unexpected was the degree to which the efficiency degraded.

7.4 Summary

In this chapter the results produced by the algorithm discussed in chapter 6 were presented. The FAP PSO algorithm was applied to four COST 259 benchmarks namely Siemens1, Siemens2, Siemens3 and Siemens4. These four benchmarks were discussed in detail in chapter 3. For each of the benchmarks, 12 different variants of the FAP PSO algorithm were tested. Each variant used a different velocity function, global best selection scheme or population size. The chapter concluded with critical analyses of each of the different algorithms developed in this study to enable the PSO to operate in the FAP space.

Chapter 8

Conclusion

8.1 Introduction

The aim of this dissertation was to develop a modern artificial intelligence optimisation algorithm which would be applied to a problem in the cellular technology domain. This aim was achieved: an algorithm was developed based on the standard PSO algorithm which was then applied to operate on the FAP.

In current research no similar PSO algorithm to date has been presented that has been applied to the FS-FAP using interference as a performance measurement. In chapter 6 the innovation and new techniques that made it possible for an algorithm based on the PSO to operate in the FS-FAP domain were discussed.

8.2 Research Questions

At the start of the dissertation a series of research questions were outlined. These questions were identified and served as a roadmap to understand the problem domain as well as to gather enough background information on optimisation algorithmic techniques to allow innovation.

Six questions were identified in total. Each original question will now be listed together with a short discussion on how the question was answered and to which chapter it is related to.

- **Question 1** — *What is cellular technology, how was it got developed and what improvements have been made since its initial development?*

This question was answered in chapter 2, which provided the history of cellular technology and also highlighted improvements that have been made since the inception of cellular networks.

- **Question 2** — *What is the architecture behind a modern cellular network, how do the various hardware entities within a network communicate with each other and how is a communicational link established between two users of the network?* This question was answered in chapter 2 where the architecture of a modern-day cellular network was outlined and discussed in depth. Each entity used by the network to facilitate communication was identified and discussed in detail.
- **Question 3** — *What exactly is the frequency problem and how does it affect modern wireless communication?* This question was answered in chapter 3, which contains a description of the general problem of the frequency assignment problem as well as exactly what happens in a cellular network for this problem to occur.
- **Question 4** — *What variants of the frequency problem exist and which are most applicable to cellular networks?* The variants that currently exist in the problem domain were listed in chapter 3, together with their respective history and what domain they affect. The particular variant concentrated on in this dissertation concentrated on was formally stated.
- **Question 5** — *What are the most popular optimisation algorithms and what characteristics make them unique?* This question was answered in two chapters namely chapters 4 and 5. In these chapters the most popular and modern optimisation algorithms were evaluated.
- **Question 6** — *With algorithms that achieved success in their respective optimisation problems, what particular technique is used by the algorithms that them it to achieve better performance?* Chapters 4 and 5 covered modern optimisation algorithms, which were also evaluated. Exactly what made the algorithm unique and allowed it to achieve success in the domain it was applied to was also dealt with

8.3 Research Objective

As stated in the introduction, the objective of this research was to determine whether it is viable and possible to apply a modern-day swarm-based algorithm to the frequency assignment problem.

The question therefore is whether this research objective was achieved. The answer is yes. An algorithm based on the PSO algorithm was presented that operated on the frequency assignment problem. The only part still left is the viability of applying this algorithm to modern cellular networks.

Before the viability analysis is presented, the chapter breakdown in the introduction is revisited in the next section.

8.4 Chapter Breakdown Revisited

The chapter breakdown presented in the introduction is provided once more but with a summary of exactly what was discussed in each particular chapter.

8.4.1 Part I - Background

Chapter 1

This chapter provided an introduction to the research as well as a broad overview of the topics the dissertation discussed.

Chapter 2

In this chapter a broad discussion was given on modern cellular technology, specifically the GSM cellular network technology. A brief history on how GSM was developed to be the most widely used cellular technology in use today was presented in section 2.2.1. This was followed by various GSM architecture entities.

A broad overview followed in section 2.4 on the various communication interfaces used between the various GSM entities to communicate with each other. The different GSM channels which are used on the interfaces to communicate information were covered.

The chapter concluded with the handover process which is used to allow an MS device to move freely geographically within the network.

Chapter 3

The problem defined in this dissertation is the frequency assignment problem (FAP). The problem where the problem was categorised as being part of the set of NP-Complete problems. The NP-Complete nature of the problem is an important concept to understand.

Within the FAP domain there exists are different techniques when assigning frequencies, discussed in section 3.4.

The FAP is not just one problem but consists of various subproblems that have different goals for the resulting frequency plan. Some problems are concerned with the number of frequencies used, others are more concerned with the amount of interference that is generated on the network because of the assignment. The various FAP subproblems were outlined in section 3.6.

A formal mathematical definition of the MI-FAP was also given. Finally, the chapter concluded with the various benchmark problems which are used to test the viability of frequency assignment algorithms.

Chapter 4

In this chapter a definition was given of what it means for an algorithm to be classified as metaheuristic and also what characteristics these algorithms must exhibit.

Three metaheuristic algorithms were discussed in this chapter. Each algorithm was accompanied by a flow chart depicting the general flow of the algorithm. How the algorithm works as well as the various characteristics that make the algorithm unique were described.

For each algorithm, a brief overview of literature using the particular algorithm was given as well as some of the disadvantages or challenges faced when applying the particular algorithm to the FAP.

Chapter 5

In this chapter a discussion was presented on three swarm intelligence algorithms. For each of the three algorithms a flow chart was given to allow the reader a more visual view of the operation of the algorithm.

Each algorithm was analysed in detail to identify its strengths and the techniques it uses to achieve good results. Finally the algorithm was critically evaluated in the terms of being applied to the FAP.

8.4.2 Part II - Implementation

Chapter 6

In this chapter a series of mathematical optimisation problems were presented. Each problem was mathematically formulated as well as categorised as to whether the particular problem was multimodal and separable. To get a better idea of how the particle swarm optimisation (PSO) algorithm operates and performs, two PSO algorithms were developed. Finally, this chapter concluded with the results obtained by the algorithms.

Chapter 7

This chapter contained an algorithm based on the standard PSO algorithm to operate on the FAP encountered in cellular networks.

All the various problems as well as how they were solved during the development of the main algorithm of this dissertation were described. For each new or old technique used by the algorithm pseudocode was given for a better idea of how the algorithm uses it.

The chapter concluded with small additions made to the algorithm to improve performance and, most important of all, improve solution quality.

Chapter 8

In this chapter results produced by the algorithm discussed in chapter 6 were presented. The FAP PSO algorithm was applied to four COST 259 benchmarks namely Siemens1, Siemens2, Siemens3 and Siemens4. These four benchmarks were discussed in detail in chapter 3. For each of the benchmarks, 12 different variants of the FAP PSO algorithm were tested. Each variant used a different velocity function, global best selection scheme or population size. The chapter concluded with a critical analysis of each of the different algorithms developed for this dissertation to enable the PSO to operate in the FAP space.

8.4.3 The Conclusion

In chapter 6 new velocity methods were created to enable the PSO swarm to move around in the problem space. To further improve the performance of the PSO additional techniques were developed.

The additional techniques were mainly concerned with how the gbest for the swarm is selected. The algorithms for these methods and techniques were also listed in the chapter. Not all the techniques used in the FAP PSO were entirely new; some techniques that were used can be easily identified like the use of tabu lists. Other techniques are a bit more difficult to identify. The building of the gbest actually borrows a concept used by ants in the ACO algorithm to build solutions.

The FAP PSO algorithm uses random collision resolution when a particle position is already in the tabu list. The random collision resolution works by randomly selecting a new channel until it is found not to be tabu. This procedure is very similar to the mutation operation used by the GA algorithm.

In chapter 7 the results of the FAP PSO algorithm applied to the COST 259 Siemens benchmark were discussed. By critically evaluating the results one can conclude that the best performing PSO variant is the particular algorithm using the first developed velocity method, which utilises cells to build the global best. The results achieved by this variant of the algorithm greatly outweigh those of the other algorithms, but when compared with the best achieved in the literature the algorithm still has some way to go.

Being so far off the best achieved results in the literature is not surprising, as applying the PSO to the FS-FAP and to the COST 259 benchmark is a first. Nonetheless this research has shown that it is indeed viable to apply the PSO to the FS-FAP, and with more research it is possible that the PSO algorithm might be able to either come near the best presented results or actually improve upon them.

8.5 Future Work

Most of the techniques developed for the FAP PSO, aim to stay true to the standard PSO algorithm. The next step is to hybridise the FAP PSO algorithm. A good candidate for hybridisation would be the GA as the

algorithm naturally “purifies” results in finding the right genes that make up a good possible solution.

In the PSO a good entry point for the GA to be incorporated would be twofold. The first point would be to take a certain global best selected with the standard procedure and then mate it with successive global bests, with the offspring global best being a best for a future iteration. With this method the PSO is able to use the history of the algorithm from the start.

The second method takes a swarm of a certain iteration and then mates all the particles with each other for a certain number of iterations. This method can be seen as a means to clean the swarm of inefficient genes and could serve as an intensification phase for the algorithm.

Another point of interest is to disregard the PSO and rather try and produce an ABC algorithm on the FS-FAP. As mentioned, the ABC algorithm was not chosen for this research, since it is new and has not been applied to a wide variety of problems.

By using some of the techniques developed in this research for instance the selection schemes, a viable ABC algorithm could be developed. The ABC algorithm is not that specific as to how new solutions are generated, as with the PSO which relies on vector mathematics, which allows an algorithm designer considerably more freedom.

Acronyms

ABC artificial bee colony. 97, 112–116, 119, 121–123, 135

ACO ant colony optimisation. xiii, 97, 99–104, 109–112, 115, 129, 135

ACS ant colony system. 102, 107

AFP automatic frequency planning. 35

AGCH access grant channel. 28

AI artificial intelligence. 59, 60

ANTS approximate non-deterministic tree search. 102

ARPSO attract repulse particle swarm optimisation. 129

AS ant system. 101, 102, 104–108

AUC authentication centre. 22, 23

BCCH broadcast control channel. 27, 28, 30

BCH broadcast channel. 27

BCO bee colony optimisation. 115

BER bit error rate. 44

BSC base station controller. 17, 25, 31, 32

BSO bee swarm optimisation. 115

BSS base station subsystem. 17, 21, 25, 28, 29

BTS base transceiver stations. 17–19, 25, 27, 30–32

- CAP** channel assignment problem. 35
- CBCH** cell broadcast control channel. 29
- CCCH** common control channel. 27, 28
- CDMA** code division multiple access. 13, 14
- COST** COoperation europene dans le domaine de la recherche Scientifique et Technique. xiii, 48, 51, 73, 81, 92, 94, 134
- D/ACCH** dedicated/associated control channel. 28
- DCCH** dedicated control channel. 27
- DCS1800** digital cellular system. 11
- DFA** dynamic frequency allocation. 39–41, 55
- DS-CDMA** direct sequence collision detection multiple access. 13
- EDGE** enhanced data global evolution. 10
- EIR** equipment identity register. 22, 23
- EUCLID** EUropean Cooperation on the Long term in the Defense. 51, 55
- FACCH** fast associated control channel. 29
- FAP** frequency assignment problem. v, vi, xi, xiii, 10, 13, 33, 35, 37–39, 46, 48–50, 52–57, 64, 72–74, 80–82, 92–94, 99, 110–112, 121–123, 133–135, 139–141, 143, 147, 153, 154, 161–163, 165–170
- FCCH** frequency control channel. 27, 28
- FDM** frequency division multiplexing. 20
- FEA** frequency exhaustive assignment. 134
- FFA** fixed frequency allocation. 39–41, 55
- FS-FAP** fixed spectrum frequency assignment problem. iv, 47–49, 72, 133, 134, 144
- GA** genetic algorithm. 82–86, 88–90, 92–95, 125

- GPRS** general packet radio system. 10
- GSM** general system for mobile communication. 10–15, 17, 18, 20–30, 32, 33, 51–53, 56
- HLR** home location register. 21–23
- HMT** heuristic manipulation technique. xiii, 72, 73
- HSCSD** high speed circuit switched data. 10
- IMEI** international mobile equipment identity. 15, 23
- IMSI** international mobile subscriber identity. 15
- ISDN** integrated service digital network. 11, 25
- LA** location area. 17
- LAPDm** link access protocol D channel modified. 25
- LTE** long term evolution. 14
- LTM** long term memory. 68, 69
- MI-FAP** minimum interference frequency assignment problem. 47, 48, 57
- MMAS** min-max ant system. 102, 107, 108
- MMS** multimedia message service. 10
- MO-FAP** minimum-order frequency assignment problem. 46, 47
- MS** mobile station. 15, 17, 20, 21, 23, 25, 27–33, 40
- MS-FAP** minimum span frequency assignment problem. 47, 50, 133, 134
- MSC** mobile switching centre. 17, 21, 22, 25, 29–32
- MTM** medium term memory. 68, 69
- NCH** notification channel. 28
- NMC** network management centre. 23, 24

- NSS** network switching subsystem. 21
- OMC** operations and management centre. 23, 24
- PCH** paging channel. 28
- POS** [point of sale. 15
- PRNG** pseudo random number generator. 88, 89
- PSO** particle swarm optimisation. vi, xi, 97, 124–135, 139–141, 143–145, 147, 151, 154, 156, 157, 160–163, 165–170
- PSTN** public switching telephone network. 21, 22
- QoS** quality of service. 43
- QRNG** quasi random number generator. 88
- RACH** random access channel. 28
- SA** simulated annealing. 74–82, 84, 88, 92, 94, 97
- SACCH** slow associated control channel. 29
- SACO** simple ant colony optimisation. 101
- SCH** synchronisation channel. 27, 28
- SDCCH** stand-alone dedicated control channel. 28, 29, 31
- SIM** subscriber identification module. 15
- SINR** signal to interference noise power ration. 44
- SIR** signal-to-noise ratio. 12, 13, 43, 44
- SMS** short message service. 10
- SRES** signed response. 23
- STM** short term memory. 68
- TCH** traffic channel. 27–29, 31

- TDM** time division multiplexing. 20
- TDMA** time division multiple access. 11, 25, 26, 28
- TRX** transceiver. 19, 20, 35
- TS** tabu search. xiii, 64–74, 76, 79–82, 84, 92, 94, 97, 166
- UMTS** universal mobile telecommunications system. 13, 14
- VBA** virtual bee algorithm. 115
- VLR** visitor location register. 21, 22
- WCDMA** wide band collision detection multiple access. 13
- WiMAX** worldwide interoperability for microwave access. 14

Bibliography

- [1] Karen Aardal, Cor Hurkens, Jan Karel Lenstra, and Sergey Tiourine. Algorithm for radio link frequency assignment: The calma project. *Operations Research*, 50(6):968–980, Nov - Dec 2002.
- [2] Karen I. Aardal, Stan P. M. van Hoesel, Arie M. C. A. Koster, Carlo Mannino, and Antonio Sassano. Models and solution techniques for frequency assignment problems. *4OR: A Quarterly Journal of Operations Research*, 1(4):261–317, December 2004.
- [3] E. Alba, F. Luna, A. J. Nebro, and J. M. Troya. Parallel heterogeneous genetic algorithms for continuous optimization. *Parallel Comput.*, 30(5-6):699–719, 2004.
- [4] Mahmoud H. Alrefaei and Sigrún Andradóttir. A simulated annealing algorithm with constant temperature for discrete stochastic optimization. *Management Science*, 45:748–764, 1999.
- [5] L. G. Anderson. A simulation study of some dynamic channel assignment algorithms in a high capacity mobile telecommunications system. *IEEE Transactions on Communications*, 21:1294–1301, 1973.
- [6] S. Areibi and A. Vannelli. Circuit partitioning using a tabu search approach. In *Circuits and Systems, 1993., ISCAS '93, 1993 IEEE International Symposium on*, pages 1643 –1645, 3-6 1993.
- [7] M.B. Aryanezhad and Mohammad Hemati. A new genetic algorithm for solving nonconvex nonlinear programming problems. *Applied Mathematics and Computation*, 199(1):186 – 194, 2008.

- [8] A. Augugliaro, L. Dusonchet, and E. R. Sanseverino. An evolutionary parallel tabu search approach for distribution systems reinforcement planning. *Advanced Engineering Informatics*, 16(3):205 – 215, 2002.
- [9] I. Badarudin, A.B.M. Sultan, M.N. Sulaiman, A. Mamat, and M.T.M. Mohamed. Metaheuristic approaches for optimizing agricultural land areas. In *Data Mining and Optimization, 2009. DMO '09. 2nd Conference on*, pages 28 –31, 27-28 2009.
- [10] Hua Bai and Bo Zhao. A survey on application of swarm intelligence computation to electric power system. In *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*, volume 2, pages 7587 –7591, 0-0 2006.
- [11] T.R. Benala, S.D. Jampala, S.H. Villa, and B. Konathala. A novel approach to image edge enhancement using artificial bee colony optimization algorithm for hybridized smoothening filters. In *Nature Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*, pages 1071 –1076, 9-11 2009.
- [12] Christian Blum. Ant colony optimization: Introduction and recent trends. *Physics of Life Reviews*, 2(4):353 – 373, 2005.
- [13] Bruce M. Blumberg. *Exploring Artificial Intelligence in the New Millennium*, chapter D-Learning: what learning in dogs tells us about building characters that learn what they ought to learn, pages 37–67. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.
- [14] R. Borndörfer, A. Eisenblätter, M. Grötschel, and A. Martin. Frequency assignment in cellular phone networks. *Annals of Operations Research*, (76):73–93, 1998.
- [15] Ralf Borndörfer, Andreas Eisenblätter, Martin Grötschel, and Alexander Martin. The orientation model for frequency assignment problems. Technical report, Konrad-Zuse-Zentrum Berlin, 1998.
- [16] Jeffrey E. Boyd, Gerald Hushlak, and Christian J. Jacob. Swarmart: interactive art from swarm intelligence. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 628–635, New York, NY, USA, 2004. ACM.

- [17] L. Cavique, C. Rego, and I. Themido. Subgraph ejection chains and tabu search for the crew scheduling problem. *The Journal of the Operational Research Society*, 50:608–616, 1999.
- [18] A. Chawla, S. Mukherjee, and B. Karthikeyan. Characterization of human passive muscles for impact loads using genetic algorithm and inverse finite element methods. *Biomechanics and Modeling in Mechanobiology*, 8(1):67–76, 2009.
- [19] Ruey-Maw Chen, Shih-Tang Lo, Chung-Lun Wu, and Tsung-Hung Lin. An effective ant colony optimization-based algorithm for flow shop scheduling. In *Soft Computing in Industrial Applications, 2008. SMCia '08. IEEE Conference on*, pages 101 –106, june 2008.
- [20] Chin Soon Chong, Appa Iyer Sivakumar, Malcolm Yoke Hean Low, and Kheng Leng Gay. A bee colony optimization algorithm to job shop scheduling. In *WSC '06: Proceedings of the 38th conference on Winter simulation*, pages 1954–1961. Winter Simulation Conference, 2006.
- [21] G. Colombo and S.M. Allen. Problem decomposition for minimum interference frequency assignment. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, pages 3492 –3499, sept. 2007.
- [22] Agnieszka Debudaj-Grabysz and Zbigniew J. Czech. Theoretical and practical issues of parallel simulated annealing. In *PPAM'07: Proceedings of the 7th international conference on Parallel processing and applied mathematics*, pages 189–198, Berlin, Heidelberg, 2008. Springer-Verlag.
- [23] B. Denby and S. Le Hgarat-Masclé. Swarm intelligence in optimisation problems. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 502(2-3):364 – 368, 2003. Proceedings of the VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research.
- [24] Marco Dorigo and Christian Blum. Ant colony optimization theory: A survey. *Theoretical Computer Science*, 344(2-3):243 – 278, 2005.

- [25] Marco Dorigo and Christian Blum. Ant colony optimization theory: A survey. *Theoretical Computer Science*, 344(2-3):243 – 278, 2005.
- [26] Marco Dorigo, Eric Bonabeau, and Guy Theraulaz. Ant algorithms and stigmergy. *Future Gener. Comput. Syst.*, 16(9):851–871, 2000.
- [27] K A Dowsland and J M Thompson. Solving a nurse scheduling problem with knapsacks, networks and tabu search. *J Oper Res Soc*, 51(7):825–833, 07 2000.
- [28] Audrey Dupont, Andrea Carneiro Linhares, Christian Artigues, Dominique Feillet, Philippe, and Michel Vasquez. The dynamic frequency assignment problem. *European Journal of Operational Research*, 195:75–88, 2009.
- [29] A. Eisenblätter, M. Grötschel, and A. Martin. Frequency planning and ramifications of colouring. Technical report, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2000.
- [30] Andreas Eisenblätter. Assigning frequencies in gsm networks. Technical report, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), 2001.
- [31] Andreas Eisenblätter. *Frequency Assignment in GSM Networks: Models, Heuristics, and Lower Bounds*. PhD thesis, Technische Universität Berlin, Berlin, Germany, 2001.
- [32] H.M. Elkamchouchi, H.M. Elragal, and M.A. Makar. Channel assignment for cellular radio using particle swarm optimization. In *Radio Science Conference, 2006. NRSC 2006. Proceedings of the Twenty Third National*, volume 0, pages 1 –9, 14-16 2006.
- [33] Andries P. Engelbrecht. *Fundamentals of Computational Swarm Intelligence*. John Wiley & Sons, 2006.
- [34] Andries P. Engelbrecht. *Computational Intelligence: An Introduction*. Wiley Publishing, 2007.
- [35] Dr. Kamilo Feher. *Wireless Digital Communications: Modulation & Spread Spectrum Applications*. Prentice Hall, 1995.

- [36] N. Fescioglu-Unver and M.M. Kokar. Application of self controlling software approach to reactive tabu search. In *Self-Adaptive and Self-Organizing Systems, 2008. SASO '08. Second IEEE International Conference on*, pages 297–305, 20-24 2008.
- [37] L. M. Gambardella, ÉD Taillard, and M. Dorigo. Ant colonies for the quadratic assignment problem. *The Journal of the Operational Research Society*, 50(2):167–176, Feb 1999.
- [38] Gautam Garai and B. B. Chaudhuri. A distributed hierarchical genetic algorithm for efficient optimization and pattern matching. *Pattern Recogn.*, 40(1):212–228, 2007.
- [39] Fred Glover. Tabu search - part i. *INFORMS Journal on Computing*, 1(3):190–206, 1989.
- [40] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.
- [41] Antonio Gómez-Iglesias, Miguel A. Vega-Rodríguez, Francisco Castejón, Miguel Cárdenas-Montes, and Enrique Morales-Ramos. Artificial bee colony inspired algorithm applied to fusion research in a grid computing environment. In *PDP '10: Proceedings of the 2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, pages 508–512, Washington, DC, USA, 2010. IEEE Computer Society.
- [42] Michael T. Goodrich. *Data Structures and Algorithms in Java*. John Wiley & Sons, 2005.
- [43] Mohan Gopalakrishnan, Ke Ding, Jean-Marie Bourjolly, and Srimathy Mohan. A tabu-search heuristic for the capacitated lot-sizing problem with set-up carryover. *Manage. Sci.*, 47(6):851–863, 2001.
- [44] J.S Graham, R. Montemanni, J. N J. Moon, and D. H. Smith. Frequency assignment. multiple interference and binary constraints. *Wireless Networking*, 14:449–464, 2008.
- [45] C. Grosan and A. Abraham. *Stigmergic optimization: inspiration, technologies and perspectives*, volume 31 of *Studies in Computational Intelligence*, chapter 1, pages 1–24. Springer Berlin / Heidelberg, 2006.

- [46] Timo Hämäläinen, Harri Klapuri, Jukka Saarinen, Pekka Ojala, and Kimmo Kaski. Accelerating genetic algorithm computation in tree shaped parallel computer. *J. Syst. Archit.*, 42(1):19–36, 1996.
- [47] Julia Handl and Bernd Meyer. Ant-based and swarm-based clustering. *Swarm Intelligence*, 1(2):95–113, December 2007.
- [48] Abdel-rahman Hedar and Masao Fukushima. Tabu search directed by direct search methods for nonlinear global optimization. *European Journal of Operational Research*, 170:329–349, 2006.
- [49] Alan Herz, David Schindl, and Nicolas Zufferey. Lower bounding and tabu search procedures for the frequency assignment problem with polarization constraints. *JOR: A Quarterly Journal of Operations Research*, 3:139–161, 2005.
- [50] O. Holthaus and C. Rajendran. A fast ant-colony algorithm for single-machine scheduling to minimize the sum of weighted tardiness of jobs. *The Journal of the Operational Research Society*, 56(8):pp. 947–953, 2005.
- [51] <http://fap.zib.de/>. Fap web, 2010.
- [52] Lhassane Idoumghar and René Schott. Two distributed algorithms for the frequency assignment problem in the field of radio broadcasting. *IEEE Transactions on Broadcasting*, 55:223–229, 2009.
- [53] C. Bettstetter J. Eberspächer, H.-J. Vögel and C. Hartmann. *GSM - Architecture, Protocols and Services*. Wiley Publishing, third edition, 2009.
- [54] D.M. Jaeggi, G.T. Parks, T. Kipouros, and P.J. Clarkson. The development of a multi-objective tabu search algorithm for continuous optimisation problems. *European Journal of Operational Research*, 185(3):1192 – 1212, 2008.
- [55] A.A. Javadi, R. Farmani, and T.P. Tan. A hybrid intelligent genetic algorithm. *Advanced Engineering Informatics*, 19(4):255 – 262, 2005. Computing in Civil Engineering.

- [56] I. Jovanoski, I. Chorbev, D. Mihajlov, and I. Dimitrovski. Tabu search parameterization and implementation in a constraint programming library. In *EUROCON, 2007. The International Conference on Computer as a Tool*, pages 459–464, 9-12 2007.
- [57] Wei jun Xia and Zhi ming Wu. A hybrid particle swarm optimization approach for the job-shop scheduling problem. *The International Journal of Advanced Manufacturing Technology*, 29(3):360–366, June 2006.
- [58] Vijay Kalivarapu, Jung-Leng Foo, and Eliot Winer. Improving solution characteristics of particle swarm optimization using digital pheromones. *Structural and Multidisciplinary Optimization*, 37:415–427, 2009.
- [59] Dervis Karaboga and Bahriye Akay. A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*, 214(1):108–132, 2009.
- [60] Dervis Karaboga and Bahriye Basturk. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *J. of Global Optimization*, 39(3):459–471, 2007.
- [61] Il kwon Jeong and Ju jang Lee. Adaptive simulated annealing genetic algorithm for system identification. *Engineering Applications of Artificial Intelligence*, 9(5):523–532, 1996.
- [62] Sergio Ledesma, Miguel Torres, Donato Hernández, Gabriel Aviña, and Guadalupe García. Temperature cycling on simulated annealing for neural network learning. In *MICAI 2007: Advances in Artificial Intelligence*, pages 161–171, 2007.
- [63] Zne-Jung Lee, Chou-Yuan Lee, and Shun-Feng Su. An immunity-based ant colony optimization algorithm for solving weapon-target assignment problem. *Applied Soft Computing*, 2(1):39–47, 2002.
- [64] K. Lenin and M.R. Mohan. Attractive and repulsive particle swarm optimization for reactive power optimization. *Journal of Engineering and Applied Sciences*, 1(4):288–292, 2006.

- [65] Yuming Liang and Lihong Xu. Mobile robot global path planning using hybrid modified simulated annealing optimization algorithm. In *GEC '09: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, pages 309–314, New York, NY, USA, 2009. ACM.
- [66] Nguyen Tung Linh and Nguyen Quynh Anh. Application artificial bee colony algorithm (abc) for reconfiguring distribution network. In *Computer Modeling and Simulation, 2010. ICCMS '10. Second International Conference on*, volume 1, pages 102 –106, 22-24 2010.
- [67] Hongbo Liu, Ajith Abraham, and Maurice Clerc. Chaotic dynamic characteristics in swarm intelligence. *Applied Soft Computing*, 7(3):1019 – 1026, 2007.
- [68] Francisco Luna, Christian Blum, Enrique Alba, and Antonio J. Nebro. Aco vs eas for solving a real-world frequency assignment problem in gsm networks. In *GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pages 94–101, New York, NY, USA, 2007. ACM.
- [69] H. Mahmoudzadeh and K. Eshghi. A metaheuristic approach to the graceful labeling problem of graphs. In *Swarm Intelligence Symposium, 2007. SIS 2007. IEEE*, pages 84 –91, 1-5 2007.
- [70] S. Mallela and L.K. Grover. Clustering based simulated annealing for standard cell placement. In *Design Automation Conference, 1988. Proceedings., 25th ACM/IEEE*, pages 312 –317, 12-15 1988.
- [71] Vittoria Maniezzo and Roberto Montemanni. An exact algorithm for the min-interference frequency assignment problem. Technical report, Department of Computer Science, University of Bologna, 2000.
- [72] Carlo Mannino and Gianpaolo Oriolo. Solving stability problems on a superclass of interval graphs. Technical report, Centro Vito Volterra, 2002.
- [73] Y. Marinakis, M. Marinaki, and N. Matsatsinis. A hybrid discrete artificial bee colony - grasp algorithm for clustering. In *Computers Industrial Engineering, 2009. CIE 2009. International Conference on*, pages 548 –553, 6-9 2009.

- [74] Asha Mehrotra. *GSM System Engineering*. Artech House, Inc, 1997.
- [75] David Meignan, Jean-Charles Créput, and Abderrafaa Koukam. A cooperative and self-adaptive metaheuristic for the facility location problem. In *GECCO '09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 317–324, New York, NY, USA, 2009. ACM.
- [76] George Jiri Mejtsky. The improved sweep metaheuristic for simulation optimization and application to job shop scheduling. In *WSC '08: Proceedings of the 40th Conference on Winter Simulation*, pages 731–739. Winter Simulation Conference, 2008.
- [77] P.R.S. Mendonca and L.P. Caloba. New simulated annealing algorithms. In *Circuits and Systems, 1997. ISCAS '97., Proceedings of 1997 IEEE International Symposium on*, volume 3, pages 1668 –1671 vol.3, 9-12 1997.
- [78] Marie-Bernadette Pautet Michel Mouly. *The GSM System for Mobile Communications*. Cell & Sys, 1992.
- [79] Qi Ming-yao, Miao Li-xin, Zhang Le, and Xu Hua-yu. A new tabu search heuristic algorithm for the vehicle routing problem with time windows. In *Management Science and Engineering, 2008. ICMSE 2008. 15th Annual Conference Proceedings., International Conference on*, pages 1648 –1653, 10-12 2008.
- [80] Christopher K. Monson and Kevin D. Seppi. Adaptive diversity in pso. In *GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 59–66, New York, NY, USA, 2006. ACM.
- [81] Roberto Montemanni. *Upper and Lower bounds for the fixed spectrum frequency assignment problem*. PhD thesis, School of Technology, University of Glamorgan, 2001.
- [82] Roberto Montemanni and Derek H. Smith. Heuristic manipulation, tabu search and frequency assignment. *Comput. Oper. Res.*, 37(3):543–551, 2008.

- [83] Garry Mullet. *Wireless telecommunications systems and networks*. Thomsan Delmar Learning, 2006.
- [84] Amit Nagar, Sunderesh S. Heragu, and Jorge Haddock. A combined branch-and-bound and genetic algorithm based approach for a flow-shop scheduling problem. *Annals of Operations Research*, 63(3):397–414, June 1996.
- [85] B. Natrajan and B.E. Rosen. Image enhancement using very fast simulated reannealing. In *Image Analysis and Interpretation, 1996., Proceedings of the IEEE Southwest Symposium on*, pages 230 –235, 8-9 1996.
- [86] Lin Ni and Hong-Ying Zheng. An unsupervised intrusion detection method combined clustering with chaos simulated annealing. In *Machine Learning and Cybernetics, 2007 International Conference on*, volume 6, pages 3217 –3222, 19-22 2007.
- [87] Koji Nonobe and Toshihide Ibaraki. A tabu search approach to the constraint satisfaction problem as a general problem solver. *European Journal of Operational Research*, 106(2-3):599 – 623, 1998.
- [88] E. Nowicki and S. Zdrzalka. Single machine scheduling with major and minor setup times - a tabu search approach. *The Journal of the Operational Research Society*, 47:1054–1064, 1996.
- [89] Sigurdur Ólafsson. Chapter 21 metaheuristics. In Shane G. Henderson and Barry L. Nelson, editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pages 633 – 654. Elsevier, 2006.
- [90] Michael O’Neill and Anthony Brabazon. Self-organising swarm (soswarm). *Soft Comput.*, 12(11):1073–1080, 2008.
- [91] Thomas La Porta Patrick Traynor, Patrick McDaniel. *Security for Telecommunications Networks*, volume 40 of *Advances in Information Security*. Springer US, 2008.
- [92] Pedro Pereira, Fernando Silva, and Nuno A. Fonseca. Bioired - a genetic algorithm for pattern detection in biosequences. In *2nd Interna-*

- tional Workshop on Practical Applications of Computational Biology and Bioinformatics (IWPACBB 2008)*, pages 156–165, 2009.
- [93] Jean-Yves Potvin. Genetic algorithms for the traveling salesman problem. *Annals of Operations Research*, 63:337–370, 1996.
 - [94] Nilesh B. Prajapati, Rupal R. Agravat, and Mosin I. Hasan. Comparative study of various cooling schedules for location area planning in cellular networks using simulated annealing. *Networks & Communications, International Conference on*, 0:146–150, 2009.
 - [95] Abraham P. Punnen and Y. P. Aneja. A tabu search algorithm for the resource-constrained assignment problem. *The Journal of the Operational Research Society*, 46(2):214–220, Feb 1995.
 - [96] A. Ratnaweera, S.K. Halgamuge, and H.C. Watson. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *Evolutionary Computation, IEEE Transactions on*, 8(3):240 – 255, june 2004.
 - [97] Andrea Reese. Random number generators in genetic algorithms for unconstrained and constrained optimization. *Nonlinear Analysis: Theory, Methods & Applications*, 71(12):e679 – e692, 2009.
 - [98] D. J. Reid. Genetic algorithms in constrained optimization. *Mathematical and Computer Modelling*, 23(5):87 – 111, 1996.
 - [99] Angelos N. Rouskas, Michael G. Kazantzakis, and Miltiades E. Anagnostou. Minimizing of frequency assignment span in cellular networks. *IEEE Transactions on Vehicular Technology*, 48:873–882, 1999.
 - [100] Stuart Russel and Peter Norvig. *Artificial Intelligence A Modern Approach*. Prentice Hall, second edition, 2003.
 - [101] P. Demestichas E. Tzifa S. Kotrotsos, G. Kotsakis and V. Demesticha. Formulaton and computationally efficient algrithms for an interference-orientated version of the frequency assignment problem. *Wireless Personal Communications*, 18:289–317, 2001.
 - [102] Mohsen Saemi and Morteza Ahmadi. Integration of genetic algorithm and a coactive neuro-fuzzy inference system for permeability predic-

- tion from well logs data. *Transport in Porous Media*, 71(3):273–288, February 2008.
- [103] Mischa Schwartz. *Mobile Wireless Communications*. Cambridge University Press, 2005.
- [104] Matthew Settles and Terence Soule. Breeding swarms: a ga/pso hybrid. In *GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation*, pages 161–168, New York, NY, USA, 2005. ACM.
- [105] Patrick Siarry, Alain Pétrowski, and Mourad Bessaou. A multipopulation genetic algorithm aimed at multimodal optimization. *Adv. Eng. Softw.*, 33(4):207–213, 2002.
- [106] Alok Singh. An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem. *Applied Soft Computing*, 9(2):625 – 631, 2009.
- [107] J. R. Slagle, A. Bose, P. Busalacchi, and C. Wee. Enhanced simulated annealing for automatic reconfiguration of multiprocessors in space. In *IEA/AIE '89: Proceedings of the 2nd international conference on Industrial and engineering applications of artificial intelligence and expert systems*, pages 401–408, New York, NY, USA, 1989. ACM.
- [108] K.G. Srinivasa, K.R. Venugopal, and L.M. Patnaik. A self-adaptive migration model genetic algorithm for data mining applications. *Information Sciences*, 177(20):4295 – 4313, 2007.
- [109] Marc St-Hilaire, Steven Chamberland, and Samuel Pierre. A tabu search algorithm for the global planning problem of third generation mobile networks. *Comput. Electr. Eng.*, 34(6):470–487, 2008.
- [110] Gordon L. Stuiüber. *Principles of Mobile Communication*. Kluwer Academic Publishers, 1996.
- [111] Anand Prabhu Subramanian, Himanshu Gupta, Samir R. Das, and Jing Cao. Minimum interference channel assignment multiradio wireless mesh networks. *IEEE Transactions on Mobile Computing*, 7(7):1459–1473, December 2008.

- [112] B. Suman and P. Kumar. A survey of simulated annealing as a tool for single and multiobjective optimization. *The Journal of the Operational Research Society*, 57(10):1143–1160, 2006.
- [113] Ashish Sureka and Peter R. Wurman. Applying metaheuristic techniques to search the space of bidding strategies in combinatorial auctions. In *GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation*, pages 2097–2103, New York, NY, USA, 2005. ACM.
- [114] Ashish Sureka and Peter R. Wurman. Using tabu best-response search to find pure strategy nash equilibria in normal form games. In *AA-MAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1023–1029, New York, NY, USA, 2005. ACM.
- [115] Peter Tarasewich and Patrick R. McMullen. Swarm intelligence: power in numbers. *Commun. ACM*, 45(8):62–67, August 2002.
- [116] Dusan Teodorovic, Panta Lucic, Goran Markovic, and Mauro Dell'Orco. Bee colony optimization: Principles and applications. In *Neural Network Applications in Electrical Engineering, 2006. NEUREL 2006. 8th Seminar on*, pages 151 –156, 25-27 2006.
- [117] T.O. Ting, M.V.C. Rao, and C.K. Loo. A novel approach for unit commitment problem via an effective hybrid particle swarm optimization. *Power Systems, IEEE Transactions on*, 21(1):411 – 418, feb. 2006.
- [118] Raj Gaurang Tiwari, Mohd. Husain, Sandeep Gupta, and Arun Pratap Srivastava. A new ant colony optimization meta-heuristic algorithm to tackle large optimization problem. In *COMPUTE '10: Proceedings of the Third Annual ACM Bangalore Conference*, pages 1–4, New York, NY, USA, 2010. ACM.
- [119] Vedat Toğan and Ayşe T. Daloğlu. An improved genetic algorithm with initial population strategy and self-adaptive member grouping. *Comput. Struct.*, 86(11-12):1204–1218, 2008.
- [120] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

- [121] Hsien-Yu Tseng and Chang-Ching Lin. A simulated annealing approach for curve fitting in automated manufacturing systems. *Journal of Manufacturing Technology Management*, 18:202 – 216, 2007.
- [122] Roger L. Wainwright. A family of genetic algorithm packages on a workstation for solving combinatorial optimization problems. *SIGICE Bull.*, 19(3):30–36, 1994.
- [123] Haohong Wang, Lisimachos Kondi, Ajay Luthra, and Song Ci. *4G Wireless Video Communications*. Wiley Publishing, 2009.
- [124] Z. G. Wang, Y. S. Wong, and M. Rahman. Development of a parallel optimization method based on genetic simulated annealing algorithm. *Parallel Comput.*, 31(8+9):839–857, 2005.
- [125] N. A. Wassan. A reactive tabu search for the vehicle routing problem. *The Journal of the Operation Research Society*, 57(1):111–116, Jan 2006.
- [126] Xian-Huan Wen, Tina Yu, and Seong Lee. Coupling sequential-self calibration and genetic algorithms to integrate production data in geo-statistical reservoir modeling. In *Geostatistics Banff 2004*, volume 14 of *Quantitative Geology and Geostatistics*, pages 691–701. Springer Netherlands, 2005.
- [127] Dennis Weyland. Simulated annealing, its parameter settings and the longest common subsequence problem. In *GECCO '08: Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pages 803–810, New York, NY, USA, 2008. ACM.
- [128] Andreas Windisch, Stefan Wappler, and Joachim Wegener. Applying particle swarm optimization to software testing. In *GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pages 1121–1128, New York, NY, USA, 2007. ACM.
- [129] Lihua Wu and Yuyun Wang. An introduction to simulated annealing algorithms for the computation of economic equilibrium. *Computational Economics*, 12:151–169, 1998.
- [130] Yiliang Xu, Meng Hiot Lim, and Yew-Soon Ong. Automatic configuration of metaheuristic algorithms for complex combinatorial optimization.

- tion problems. In *Evolutionary Computation, 2008. CEC 2008. (IEEE World Congress on Computational Intelligence). IEEE Congress on*, pages 2380 –2387, 1-6 2008.
- [131] Jingan Yang and Yanbin Zhuang. An improved ant colony optimization algorithm for solving a complex combinatorial optimization problem. *Appl. Soft Comput.*, 10(2):653–660, 2010.
- [132] Xin-She Yang. *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, 2008.
- [133] Dominic C O'Brien Yangyang Zhang. Fixed channel assignment in cellular radio networks using particle swarm optimization. In *Proceedings of the International Symposium of Industrial Electronics*, June 2005.
- [134] Quan Yuan, Feng Qian, and Wenli Du. A hybrid genetic algorithm with the baldwin effect. *Information Sciences*, 180(5):640 – 652, 2010.
- [135] L. Zhang, S. Guo, Y. Zhu, and A. Lim. A tabu search algorithm for the safe transportation of hazardous materials. In *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, pages 940–946, New York, NY, USA, 2005. ACM.
- [136] Xin Zhao, Myung-Eun Lee, and Soo-Hyung Kim. Improved image thresholding using ant colony optimization algorithm. In *Proceedings of the 2008 International Conference on Advanced Language Processing and Web Information Technology*, pages 210–215, Washington, DC, USA, 2008. IEEE Computer Society.