NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

DIVISION OF INFORMATION TRANSMISSION SYSTEMS

AND MATERIAL TECHNOLOGY

RESOURCE OPTIMIZATION IN MULTI-BEAM SATELLITES

DIPLOMA THESIS

of

ALEXIOS ARAVANIS

Supervisor: Prof. Panayotis Cottis

Athens, 2012

Acknowledgments:

The present diploma thesis was carried out in the Interdisciplinary Centre for Security Reliability and Trust (SnT) of the University of Luxembourg under an internship agreement between the National Technical University of Athens and the University of Luxembourg. This was also under the auspices of the European Space Agency ARTES 5.1 Programme (www.esa.int) with the cooperation of the EADS Astrium Satellites, the SES Astra Techcom and the University of Luxembourg.

During this internship a number of people contributed to the development of the material for this thesis and I have benefited significantly from their feedback and suggestions. I wish to particularly thank Bhavani Shankar MR, Grégoire Danoy and Prof. Björn Ottersten from the University of Luxembourg and Pantelis-Daniel Arapoglou and Prof. Panayotis Cottis from the National Technical University of Athens for their support and guidance. Without them this dissertation could not have been written.

Abstract

Multi-beam satellite systems offer higher throughput due to frequency reuse and multiple levels of flexibility and allow the use of smaller earth terminals due to their higher directivity. At a high-level multi-beam satellite operation resembles to that of cellular mobile communications.

As the lifetime of a satellite is about 15 years efficient communication satellites must take into account current as well as future demands. To improve satellite efficiency and reduce operational risk, design flexibility is imperative. Among others, multi-beam satellites, offer significant flexibility both with regard to coverage area and resources allocation. Coverage area flexibility is achieved by adjusting the position of the spots adaptively to communications traffic. Resource allocation flexibility is made possible by adaptively adjusting the power and bandwidth to the various beams according to traffic.

In the present thesis, we focus on resource allocation flexibility offered by multi-beam satellites. In particular, for a specific satellite payload (number of beams, available power, bandwidth etc.,) the first part of the thesis deals with the problem of allocating power to different beams to satisfy the demand as closely as possible. Based on a model relating the power allocated to the data rate offered, we explore the suitability of using/ modifying existing optimization algorithms. Among exact methods as well as approximate methods that are initially considered, a scheme is selected based on complexity, convergence, scalability and other issues. The selected algorithm is then used to optimize the resource allocation. The advantages obtained over other existing methods are presented.

Next, an attempt is made to minimize the DC power consumption of a multi-beam satellite. In this course, the optimized resource allocation obtained in the first part of the thesis is further enhanced, with the aim to minimize the system power consumption. This task is deal with employing a multi-objective optimization algorithm that aims at meeting the traffic demand as closely as possible while minimizing the system power consumption. The benefits from using such multiple objective optimizations in payload design are demonstrated via appropriate simulation results.

KEY WORDS: Multi-beam Satellites, Payload Flexibility, Resource Allocation, Resource Optimization, Metaheuristics, Multi-objective Optimization

Chapter 1
Introduction to Satellite Communications

1.1 Introduction

The remarkable advancement of telecommunications over the last decades has its roots in the ever increasing demand for a plethora of services and throughput. In this course the development of Satellite Communication (SatCom) Systems plays a key role, its contribution being increasing over the years. Current satellite systems provide TV, video and sound broadcasting applications as well as fixed and mobile multimedia interactive Internet and data services.

The primary concept in satellite communications is the use of satellite carrying transponders providing coverage over areas where traditional terrestrial communications may be ineffective. This makes satellite communications ideal for remote and/or low density areas where penetration of terrestrial technologies is low. The necessity for satellite communications depending on the geography and the landscape of the area as well as their specialized role in current telecommunication systems have led to important advances in satellite technology. Over the years the cost and size of the necessary equipment for both the ground and the space segment have been significantly reduced, improving the competitiveness of satellite systems in the markets. This development along with the inherent advantages of satellite communications enabled the evolution of a healthy satellite services sector.

The primary advantages of the satellite communications are:

- **Cost Effectiveness –** The cost of satellite capacity does not increase with the number of users or their distance provided they are located within the coverage (footprint) of the satellite.
- **Ubiquity –** Rural and remote geographical regions having no access to terrestrial infrastructure can rely on satellite communications.
- **Wide Coverage –** Satellites cover wide geographical areas. Three geostationary satellites suffice to cover the Earth's surface (excluding polar areas).
- **Resource Efficiency –** Broadcast and multicast satellites are able to deliver information to an unlimited number of end-users optimizing the use of system resources.
- **Reliability –** Satellite communications operate regardless of man-made or physical problems that may render terrestrial systems non-operational.
- **Scalability –** Existing satellite systems are easy to up-scale using the ground-based equipment.
- **Versatility –** Satellite systems support various services and applications, from TV to multimedia interactive services.

Furthermore, due to their reliability and ease of deployment, satellites are used either for *last mile*[1] operation or for *trunking*[2]. At the same time, the offered bitrate (and the associated cost) can change upon demand.

However, several limitations still have to be overcome in satellite communications. These limitations are:

- The high capital expenditure (CAPEX) required for a satellite system, due to the high cost of the satellite and the supporting ground segment.
- The limited lifetime of the spacecraft and the implicit replacement cost.
- The propagation delay due to the long propagation path of the signal, especially for geostationary satellites.
- The extreme free space loss that renders satellite systems power limited.

1.2 Satellite Orbits

Satellite systems are classified according to their orbit. The most common type of orbit is the geosynchronous orbit (GSO), where the satellite has an orbital period equal to the rotational period of earth (i.e. 23 hours, 56 minutes, 4.091 seconds). The altitude of the orbit is determined at 35786km above mean sea level whereas the velocity of the satellite is 3075m/s. The ground-track[3] of such a satellite is a roughly elliptical or more accurately a figure-eight shape and is formed during one rotational period. The ground track of a geosynchronous orbit is depicted in Fig. 1.1.

Fig 1.1 Ground track of a GSO

A geosynchronous orbit with of zero eccentricity and inclination, namely a circular orbit directly above the Earth's equator, following the direction of the Earth's rotation from West to East is called a geostationary orbit (GEO) and a satellite in this orbit geostationary satellite. As the orbital period of such a satellite is again that of the Earth's rotational period, its ground track is a fixed point on the equator. As a result a geostationary satellite looks still at a fixed position in the sky to the ground terminals located in the coverage area. Thus, tracking of the satellite is not necessary whereas relative immobility of the satellite nullifies the Doppler effect and simplifies the interference prediction.

---

[1] The term "last mile" refers to the segment of the telecommunication network connecting the retail customers to the network.

[2] Trunk is a line or link designed to handle data streams connecting major nodes of a communication network. The use and management of trunks is known as trunking and aims at the minimization of the client connections.

[3] Ground track is the projection of the satellite's orbit onto the Earth's surface.

Moreover, geostationary satellites provide a wide coverage. In particular, a GEO satellite can provide coverage encompassing one or two continents whereas a constellation of three equidistant geostationary satellites can cover the whole Earth's surface (latitudes between $\varphi = \pm 75°$), excluding the poles. Furthermore, GEO satellite systems are described as fast deployment systems since short time is needed after their launch for the systems to become fully operational. For all the above reasons, most commercial communication and broadcast satellites have been installed on the geostationary orbit.

However geostationary orbits have also a number of disadvantages. Due to their high altitude, propagation delay to earth from such GEO satellites is extremely long (approx. 0.125s) to support real time services or mobile satellite communications. Moreover, GEO satellites are placed in orbit over the equator and as a result the sun (an intense source of noise) is often within the beamwidth of the ground terminal. Thus, geostationary systems are subjected to severe solar interference[4].

GEO satellites have a high deployment cost and specific lifetime. Also, due to their low inclination and the resulting inability to cover the poles, GEO satellite coverage is not truly global. Furthermore, a major disadvantage is the complexity of the ground terminal, imposed by the necessity to constantly select the GEO satellite to receive the highest possible signal power. This complexity comes also in the routing process. Finally, covering a wide area between the poles leads to the waste of the satellite resources over the oceans and other uninhabited areas. On the other hand this wide coverage is exploited by international transportation (e.g. planes and boats) that constitute a major market of SatComs.

Some of these drawbacks of GEO satellites are overcome employing different satellite orbits. The most common classification used to classify the orbits is by their altitude. These orbits are the Low Earth orbit (LEO), Medium Earth orbit (MEO) and Highly Elliptical orbit (HEO). The various satellite orbits are depicted in Fig.1.2. Satellite systems operating on these orbits share a number of characteristics due to their varying position in the sky. These satellites are visible to a ground terminal during only small time intervals. Thus, a constellation of satellites must be launched in order to accomplish hand over of the satellite connections. The *handover* process must be completed without any deterioration or interruption of the offered services. The lower the mean altitude of the orbit the more satellites are needed for the same coverage area.

Both the *handover* process that requires the use of high complexity systems and the high number of satellites required to achieve successive coverage increase the deployment cost of such systems. Furthermore, the relative motion of the satellite with respect to the ground terminals cause Doppler shift and jitter.

[4] The consequences of the sun to satellite alignment become fully apparent twice a year, when the sun aligns directly with satellites and the receiving ground terminals raising the noise floor above the satellite signal power causing temporary interruptions.

A Low Earth orbit (LEO) is positioned at altitudes varying between 160km-2000km above mean sea level. The velocity of satellites placed on LEO orbits is 7800m/s whereas their orbital period is 90min, significantly smaller than the period of GEO satellites. Because of their small orbital period LEO satellites revisit the same area multiple times per day which allows them to perform high sampling measurements of the area within a day. This is an important feature both in weather forecast and in military applications, making the Low Earth orbit ideal for meteorological and military satellites. Finally, another interesting advantage of LEO satellites is the relevant small propagation delay. On the other hand, a major disadvantage of LEO satellites is the limited lifetime of the spacecraft since the lowest the orbit the shorter the satellite lifetime.

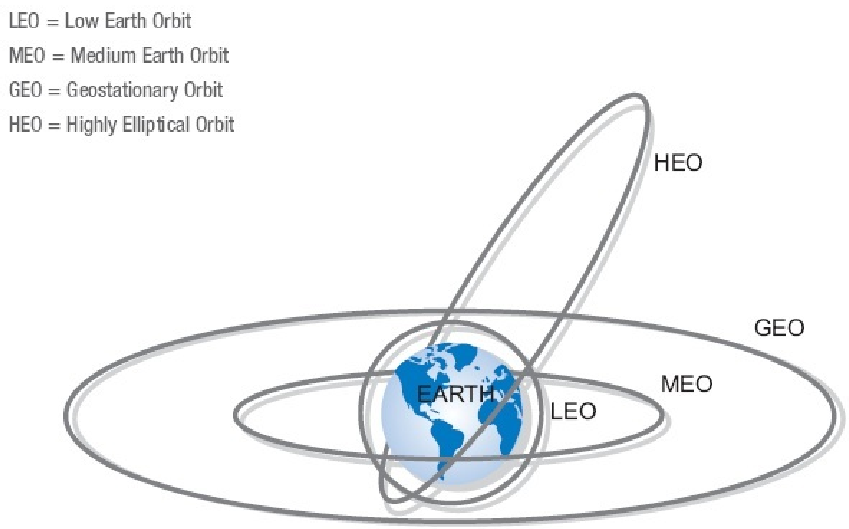


Fig. 1.2 Satellite Orbit Types

A Medium Earth orbit (MEO) is generally positioned at altitudes between 5000km – 20000km. The velocity of MEO satellites ranges between the velocities of GEO and LEO satellites. MEO satellites are used in constellations for telecommunication applications, the most common examples being the radio location systems: GPS (USA), Glonass (Russia) and Galileo (Europe).

A Highly Elliptical orbit (HEO) is an elliptic orbit with a low-altitude perigee (1000km) and a high-altitude apogee exceeding the geostationary orbit. HEO satellites appear motionless for a long time when in the apogee, whereas their velocity seems to increase rapidly when in the perigee. In particular, the ratio of velocity at apogee to velocity at perigee is equal to ratio of orbit radius at perigee to orbit radius at apogee. Thus, depending on the orbital location of a HEO satellite the satellite velocity changes and the signal is subjected to varying propagation delay. Therefore the accurate tracking of the satellite by a functional tracking system is essential for the seamless operation of the system. The Sirius Satellite Radio service operating in North America and the Ondas Media project in Europe employ HEO satellites to provide radio services.

1.3 The family of DVB standards for satellite communications

A benchmark in the evolution of satellite technology was the development of the family of DVB-S standards [1.1] (Digital Video Broadcasting via Satellite) in 1993. This development responded to the need for compatibility between the commercial standards of terrestrial communications and the duplex broadband satellite services. The DVB network architecture comprises a geostationary satellite, a number of Return Channel Satellite Terminals (RCST) and the Network Control Center (NCC), responsible for monitoring and controlling the network (Fig 1.3). The DVB-S or the DVB-S2 standards are then employed in the forward link, namely the link from the NCC hub to the users, providing unicast or multicast data to the users. On the other hand the return link, namely the link from the users to the hub, employs the DVB-RCS standard, where users initiate their connection, send their capacity requests as well as their data.

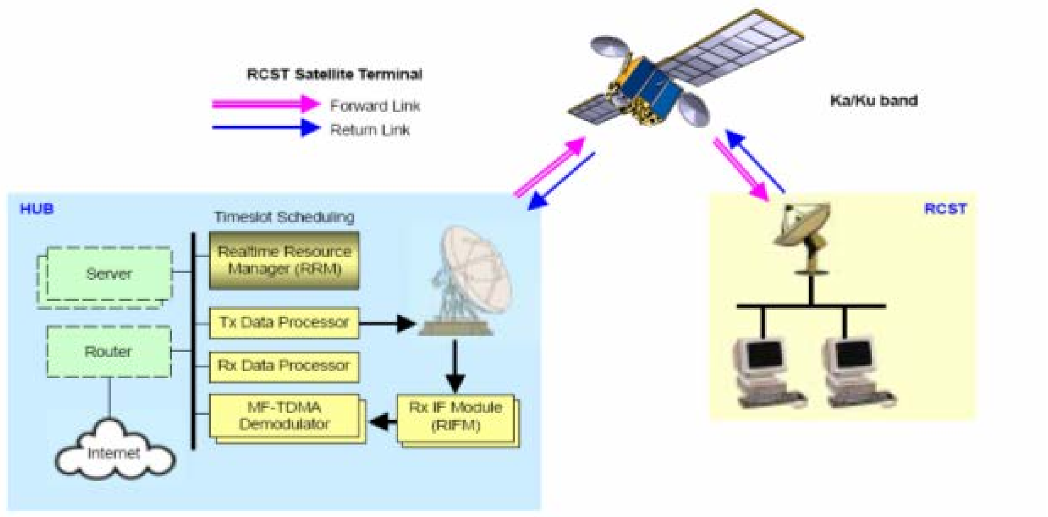


Fig 1.3 DVB network architecture

1.3.1 The DVB-S standard

The DVB-S standard was originally designed to provide Direct to Home (DTH) Television broadcast with enhanced flexibility with regard to the supported services due to the adaptation of Time Division Multiplexing (TDM) - adopted also by the MPEG-2 standard - which allows simultaneous transmission of multiple channels per carrier. Moreover DVB-S uses QPSK modulation and adaptive coding (Forward Error Correction) with code rates: 1/2, 2/3, 3/4, 5/6, and 7/8. The Time Division Multiple Access (TDMA) is used for satellite access protocol.

1.3.2 The DVB-S2 standard

The DVB-S2 standard introduced in 2003 as the successor of DVB-S, deals with a wide variety of broadband services and applications, such as standard Television, High Definition TeleVision (HDTV) and various interactive services (Internet). The main novelty of the DVB-S2 compared to the DVB-S standard is the use of Adaptive Coding and Modulation (ACM) to mitigate propagation impairments along satellite link. Depending on the modulation scheme selected (QPSK, 8PSK, 16APSK, 32APSK), the code rates used are the following: 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, 9/10.

1.3.3 The DVB-RCS standard

The DVB-RCS standard (Digital Video Broadcasting- Return Channel via Satellite) was introduced in 2001 by the DVB consortium. This standard is a specification for interactive on-demand multimedia satellite communication systems. It allows for efficient bandwidth utilization, by setting the general framework for resource allocation to users. Even though the same modulation and coding schemes as in DVB-S is adopted, the DVB-RCS allows the use of robust Turbo coding with the following code rates: 1/3, 2/5, 1/2, 2/3, 3/4, 4/5, 6/7. The uplink of DVB-RCS uses Multi-Frequency Time Division Multiple Access, (MF-TDMA) whereas the downlink uses the DVB-S2 standard.

1.4 Satellite Radio Spectrum

The allocation of the radio spectrum is coordinated on a global basis by the International Telecommunications Union (ITU) [1.2]. ITU allocates designated frequency bands to specific services aiming at efficiently exploiting the limited spectrum. In this course the frequency allocation of standardized wireless systems is dedicated to specific segments of the spectrum and the available spectrum for SatComs is limited to the respective segments. However, the advent of multimedia interactive services over the last decades caused congestion with regard both to satellite spectrum utilization and to GEO positions. Furthermore, the ever increasing demand for broadband access and new broadband applications aggravated the congestion problem. To mitigate the severe spectrum congestion the use of higher frequency bands is imperative. In this course the Ku and Ka bands were adopted in the satellite communications as shown in Table 1.1. Also the V band is also considered for future use.

Table 1.1 [1.4] tabulates the frequency bands assigned to satellite communications for both the downlink (connecting the satellite to the ground terminal) and the uplink (connecting the ground terminal to the satellite) along with the corresponding services of each band. As shown in Table 1.1 the spectral distance between the uplink and the downlink is large enough to facilitate seamless operation of the satellite receiver and transmitter. Moreover, the frequencies used in the uplink are generally higher than those used in the downlink. The use of lower frequencies in the downlink reduces the relevant free space losses; hence, the available DC power of the satellite is more efficiently used. Moreover, the operation of the ground terminals (uplink) at higher frequencies reduces the interference with the terrestrial systems that generally operate at lower frequencies and given the high power of the ground terminal transmitted signals the mitigation of their interference is imperative.

Table 1.1: Satellite Frequency Bands and corresponding Satellite Services

Frequency Band	Down-link Frequency	Up-link Frequency	Telecommunication Services
L	1 GHz	2 GHz	Mobile Satellite Service (MSS) Land Mobile Satellite Service (LMSS)
S	2 GHz	4 GHz	Mobile Satellite Service (MSS) Space Research Service
C	4 GHz	8 GHz	Fixed Satellite Service (FSS) for commercial communications
X	8 GHz	12 GHz	Fixed Satellite Service for military communications
Ku	12 GHz	18 GHz	Fixed Satellite Service (FSS) for commercial communications Broadcast Satellite Service (BSS)
K	18 GHz	27 GHz	Fixed Satellite Service (FSS) for commercial communications Broadcast Satellite Service (BSS)
Ka	27 GHz	40 GHz	Fixed Satellite Service (FSS) for commercial communications Broadcast Satellite Service (BSS)

Apart from the frequency bands and the corresponding satellite services tabulated in Table 1.1, ITU has also assigned frequency slots to certain ad hoc services, such as the Inter-Satellite Service (ISS), the Radio Determination Satellite Service (RDSS), the Radio Navigation Satellite Service (RDSS) and the Maritime Mobile Satellite Service.

1.5 Multibeam Satellite

The inherent advantages of the satellite communications described above as well as the development of the DVB-S/DVB-RCS pair of standards allowed SatCom systems to provide reliable and efficient Broadcast, Fixed and Mobile satellite services. However, to remain competitive in the continuously evolving markets, current satellite systems are facing additional challenges. In particular, due to the advances of the ever expanding terrestrial technology, the preference for satellite services is drastically dependent on the users cost for subscription, usage and terminal purchase. Moreover, current satellite systems have to keep pace with the advent of multimedia interactive services, primarily by adopting appropriate adaptive and flexible systems.

Satellite services have to move towards saving client-side cost by using low-cost terminals. At the same time, enhanced satellite bandwidth utilization and system flexibility are also warranted. In this course [1.3], multibeam satellites can play a key role. Multibeam satellites employ large on-board multibeam antennas. The resulting high gain offered by these antennas compensates for the performance degradation exhibited by low-cost user terminals, at the same time minimizing the satellite transmit power. Moreover, multibeam antennas can provide the necessary flexibility by offering frequency reuse and beam traffic reconfiguration responding to the varying traffic demand over the coverage area of the satellite. This is due to the capability of multibeam antennas to implement multispot coverage similar to terrestrial cellular application as depicted in Fig. 1.4.

The key requirements of multibeam antennas deal with [1.3]: operation frequencies and bandwidths, Rx/Tx operation, polarization characteristics (e.g. linear or circular, single or dual), frequency reuse, number of beams, size and shape of beam and the respective coverage region. The appropriate setting of the above enables multibeam antennas to play a key role with regard to all four aspects of payload flexibility:

- **Power Flexibility:** Total RF power allocation based on real traffic demands.
- **Flexibility in Reconfiguring the Frequency Plan:** Totally independent and flexible management of uplink and downlink frequencies.
- **Coverage Flexibility:** The ability to in-orbit modify the number of beams, the beams shapes and the beam footprint on the Earth.
- **Connectivity & Routing Flexibility:** Selection among multiple downlink channels connected to the receive antenna output.

The flexibility provided by multibeam SatCom systems allows dynamic allocation of system resources, namely the available bandwidth and the transmit power of the satellite, accomplishing their efficient exploitation. Given the scarcity of the available system resources in SatComs as well as their extremely high cost due to physical, technological and regulatory constraints, the flexibility offered by multibeam satellites is essential. However, to fully exploit the benefits offered, the system resources have to be optimally allocated based on the real traffic demands taking into account the inter-beam interference caused due to frequency reuse. In this course, an optimization module must be employed to perform the allocation tasks, satisfying the Quality of Service objectives (i.e. throughput maximization) as well as achieving efficient resource utilization. The use of such a module however, requires proper elaboration of the optimization process based on the system characteristics and available degrees of freedom. This necessitates the use of efficient optimization techniques for resource allocation of fixed satellite systems employing multiple spot beams.

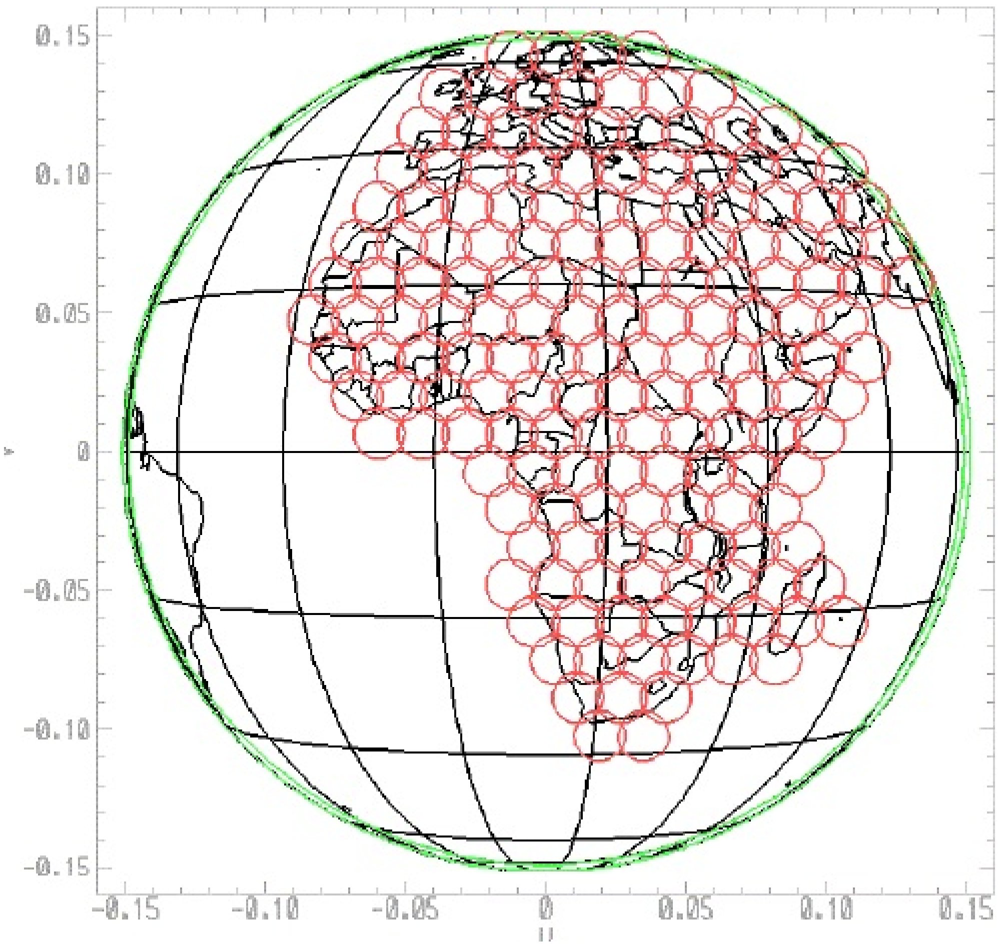


Fig. 1.4 Cellular–like Multibeam Antenna Coverage

An overview of the available optimization methods is performed with emphasis out on their suitability for the particular optimization problem; the theoretical analysis suggests the use of metaheuristics; consequently a systematic performance study of various well known metaheuristics is conducted.

Though several metahueristics have been considered for resource optimization, a systematic study leading to the selection of an appropriate metaheuristic technique was not available in the literature. Thus, the study carried out provides hints regarding which technique should be used in different occasions of the problem, for the first time in literature. The results of this study were published in the framework of the 2nd ESA Workshop on Advanced Flexible Telecom Payloads, 17-19 April 2012, Noordwijk, The Netherlands (Appendix 1). Having conducted a thorough metaheuristics study, the best performing technique is selected and enhanced with the benefits of a different technique to improve the optimization results providing better results than the standalone original method. Thus, the system performance is increased in terms of data throughput. This hybrid approach was also presented in the 2nd ESA Workshop on Advanced Flexible Telecom Payloads, 17-19 April 2012, Noordwijk, The Netherlands (Appendix 1), and was employed for the "Operational Optimization of a Ku Multibeam Flexible Payload" by Astrium Satellites, European Space and Technology Centre (ESTEC) of the European Space Agency (ESA), SES-ASTRA and University of Luxembourg (Appendix 3).

Finally, the optimization problem solved is viewed as a multi-objective optimization problem with respect to two objectives, namely throughput maximization and minimization of the system power consumption. The best power allocation determined via single-objective optimization is input to the multi-objective algorithm to be further enhanced with regard to the power consumption, i.e. the power consumed by the determined allocation will be minimized. As a result the power consumption of the best allocation is significantly reduced without deteriorating the throughput performance of the system. Furthermore, the multi-objective approach offers the trade-off curve between throughput and power consumption, which is a valuable tool for the system operator.

The multi-objective approach aiming at efficiently utilizing the available DC power is also new to the literature and constitutes an enhancement to existing payload optimizers. This approach was presented in the 30th AIAA International Communications Satellite Conference (ICSSC), 24-27 September, 2012, Ottawa, Canada (Appendix 2).

1.6 Outline

The rest of this thesis is structured as follows. Chapter 2 provides the general framework of resource allocation, details the four aspects of payload flexibility and formulates the multibeam fixed satellite optimization problem. Chapter 3 provides an overview of the available optimization techniques along with the rationale for the selection of the most appropriate for the solution to the optimization problem in hand. Chapter 4 describes the simulator used for the multibeam satellite link budget and presents the simulation results confirming the theoretical assumptions of chapter 3 and the benefits offered by the optimized allocation of the satellite resources. Chapter 5 presents the new hybrid optimization technique appropriate for data throughput maximization along with the simulation results demonstrating the improvement provided by the hybrid approach over the standard techniques. Chapter 6 presents an outlook of multi-objective optimization. Then a state of the art multi-objective technique is proposed and subsequently adopted for the optimization of the problem with respect to two objectives, namely throughput maximization and power consumption minimization. The results of this approach are also presented in chapter 6 verifying the performance improvement provided by the multi-objective approach.

References:

[1.1] www.dvb.org

[1.2] www.itu.int

[1.3] P. Angeletti, G. Toso., "*Advances in Multibeam Antennas for Satellite Applications*". European Space and Technology Centre (ESTEC) of the European Space Agency (ESA), AZ Noordwijk, The Netherlands.

[1.4] IEEE Std. 521-2002, IEEE Standard Letter Designations for Radar-Frequency Bands

Chapter 2
Resource Allocation

2.1 Introduction

Wireless networks must appropriately exploit the system resources available to meet the network QoS objectives. By resources we refer to the available power and bandwidth. The latter can be equivalently expressed in the form of either time, polarization or code resources whereas the possibility of considering the spatial dimension as an extra resource has also been considered over the last years. Optimal allocation of the available bandwidth and power is of utmost importance especially in satellite communications because of the scarcity and the extremely high cost related to physical, technological and regulatory constraints (e.g. uncongested spectrum and orbital slots). In this course, the chapter two presents the latest trends concerning resource allocation in satellite communication systems.

The most typical system paradigm in satellite communications is the broadcasting service (mostly of TV programs), for which a static resource allocation approach has been followed. Since system changes are not frequent, the system is dimensioned to accommodate the worst-case user based on a fixed protection margin. However, the advent of demanding multimedia based interactive services (like the Internet) pushed satellite communication systems to become adaptive, which potentially offers substantial capacity gains [2.8]. The characteristic indication of this adaptability is the introduction of adaptive coding and modulation (ACM) in the DVB pair of standards for the forward (DVB-S2) and the return link (DVB-RCS), respectively. Therefore, a new era in the field of dynamic resource allocation, where a network entity allocates resources based on the varying channel and traffic conditions has become a new trend in satellite communications [2.9]. In modern multibeam systems supporting a large number of beams, the spatial domain provides an extra degree-of-freedom, on the other hand, the resources must be allocated taking into account the effect each beam has on the rest (interbeam interference). The module in charge of carrying out the allocation tasks is embedded either in the gateway station (GW), in case of transparent repeaters, or on-board the satellite, in case of regenerative repeaters.

2.2 Frequency and Polarization Resources

Following the example of cellular systems, multi-beam satellites have achieved a tremendous increase in capacity by reusing spectrum in non adjacent beams. Intercell interference is the main factor constraining cellular systems; the same constraint applies to multi-beam satellites, where cells are substituted by beams. According to the well-known four color theorem: ’the regions of any simple planar map can be colored with only four colors in such a way that any two adjacent regions have different colors’ [2.2]. Thus, considering an analogy between the planar map and the coverage area of multi-beam satellite systems, as to the terms adjacent regions and beams as well as colors and frequency sub-bands, the application of the four-color theorem in communication terms is interpreted as: segmenting the total available system bandwidth into four bands is sufficient to prevent adjacent cells from utilizing the same frequencies. In multibeam satellite systems, a frequency reuse factor of four over the coverage area ensures high spatial isolation between beams sharing the same sub-bands, although more aggressive reuse policies might be adopted if the polarization domain is also employed, e.g. a frequency reuse factor of two and a polarization reuse factor of two. Another reason encouraging higher reuse schemes in modern satellite systems is the adoption of advanced waveforms techniques, allowing for the reduction of the required signal to interference-plus-noise ratio (SINR) at the receiver, leading to a more efficient frequency reuse.

2.3 Flexible Satellite Payload Resource Allocation

In the last years there has been an increased interest in designing and manufacturing payload equipment supporting different degrees of flexibility. This flexible satellite payload and its degrees-of-freedom will be briefly described in the following, highlighting the evolution from a conventional fixed payload to an advanced fully flexible one. In addition to optimally allocating system resources in real time to best serve interactive multimedia services, equipment flexibility on board the satellite is also intended for a number of reasons related to the ability of efficient communication satellites to respond to future trends and demands:

- To follow the evolution of the market over the long lifetime of communication satellites, thereby minimizing the business risks of satellite operators.
- To support different services in the new era of multi-mission satellites.
- To re-locate the satellite to a different orbital slot.
- To cope with changes in spectrum allocation status over particular countries or regions imposed by International Coordination agreements.

Throughout this section, satellite payload will refer to the electronic equipment performing the necessary carriers processing between the receiving and the transmitting antenna. The payload usually consists of several channels (also called transponders), which are dedicated to the sub-bands the total system bandwidth is separated into. Following the trend in existing satellite systems, this investigation is focused on transparent repeaters, similar to the one depicted in Fig.2.1, and does not consider other technologies related to regenerative on board processing. Depending on the specified mission and the technological constraints, the main tasks of a transparent satellite payload are [2.1] (see Fig.2.1).

- Low noise amplification of the very weak uplink carriers
- Uplink to downlink frequency conversion
- Channelization of the repeater to separate the wide band satellite signal into transponders
- High power amplification

- Recombining different channels (transponders) and connecting them with a specific beam (in case of multiple beams)

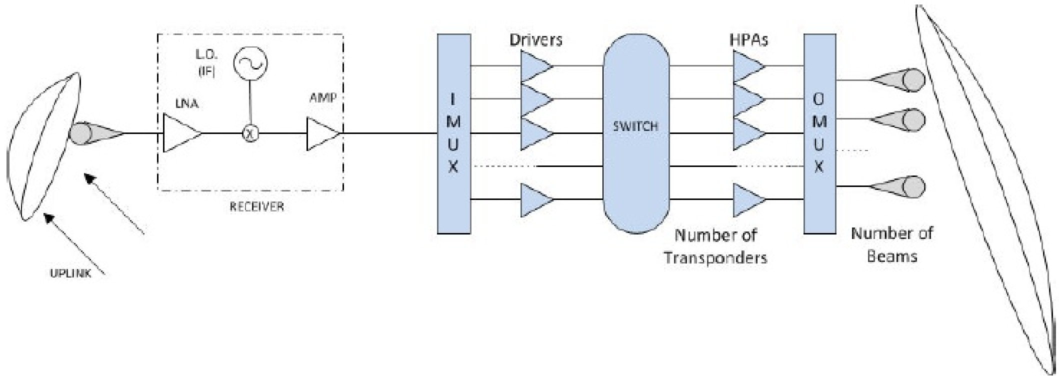


Fig. 2.1 Block diagram of satellite payload

Elaborating on the above process, the first stage consists of a low noise amplifier (LNA) to deal with the very weak signal received on the uplink and minimize the noise contribution of the mixer that follows in the chain performing uplink to downlink frequency conversion. Frequency conversion plus a first level of amplification can be performed either in a single stage or in two stages using an intermediate frequency –lower than the downlink frequency–, if it is difficult to attain the required output power level by a single stage amplification only.

After frequency conversion, the signal is further amplified as it moves through the amplification stages. However, moving the operating point from the linear to the non-linear region gives rise to intermodulation products. This limits any further amplification and dictates channelization of the wide band satellite signal of several hundreds MHz into sub-bands of smaller bandwidth. Since there is a smaller number of carriers in each sub-band, the resulting intermodulation products are less important compared to those coming up when the whole satellite bandwidth is amplified. Hence, the advantages of channelization are twofold [2.1]:

- Power amplification with limited intermodulation products due to the small number of carriers per amplifier
- Increase in the available total power since each transponder benefits from the maximum power available from each amplifier.

Transponder separation in the frequency domain is carried out by a set of bandpass filters called input multiplexers (IMUX). Typical transponder bandwidths are 30, 36, 50, 72 and 120 MHz. The carriers included in the transponder bandwidth are amplified by a high power amplifier (HPA). Next, the various sub-bands are recombined with the help of the output multiplexer (OMUX). In case of multi-beam satellites (dealt with in this chapter), a switch associates payload channels with the transmitting antenna, specifically with the relevant beam feeds (see Fig. 2.1).

2.3.1 Power Flexibility

RF power flexibility is essential for optimum resource allocation in multibeam satellite communications. To understand whether and how this is possible, a brief summary on state-of-the-art satellite High Power Amplifier (HPA) equipment is given. In fact, an HPA is the device that determines the output power of each channel and the Effective Isotropically Radiated Power (EIRP). The most frequently employed satellite HPA is the traveling wave tube amplifier (TWTA) which is based on the interaction between an electron beam and the radio wave within a tube. Typical non-flexible satellite systems transmit the same EIRP in each beam; in practice this corresponds to a TWTA operating with a given backoff. A second option, not as popular but definitely worth mentioning, are the so-called solid-state power amplifiers (SSPAs), where amplification is achieved by connecting FET transistors in parallel. SSPAs provide a very attractive power-to-mass ratio, as well as a very good linearity. Still, their drawback of exhibiting a very low DC to RF conversion efficiency and the low output power left them behind in the competition with TWTAs. Table 6.1 summarizes the basic characteristics of TWTA and SSPA [2.3].

Table 6.1: Summary of TWTA and SSPA characteristics.

Characteristics	TWTA	SSPA
Typical frequency band	C, Ku, Ka	L, C
Saturated output power	20-250 W	20-40 W
Gain at saturation	55 dB	70-90 dB
DC to RF efficiency	50-65%	30-45%
Mass	1.5-2.2 kgr	0.8-1.5 kgr

An ideal reconfigurable payload should allow the redistribution of the available RF power based on real traffic conditions. The desired on-board power flexibility can be obtained by implementing a common power pool wherefrom each channel can adaptively draw the power it requires. To provide this kind of flexibility, new generation of tubes with in-orbit adjustable saturated output power are being developed. These flexible TWTAs have the advantage of significantly reduced power consumption compared to the conventional TWTAs since, on average, they operate in saturation mode for smaller time periods. Specifically, the saturated output power level of flexible TWTAs may vary within a pre-determined range (e.g. around 4 dB). A drawback of current equipment is that the required power level must be set by a telemetry command, which implies that it can accommodate only medium and long term demand changes and cannot implement adaptive power allocation.

The amplification option that provides the highest flexibility is multiport power amplifiers (MPAs) [2.4], where the total available power of a set of amplifiers can be flexibly distributed amid different channels and, in case of multibeam payloads, different beams. MPAs –sometimes also referred to as hybrid or Buttler matrix amplifiers– are composed of three stages: an input network, a set of power amplifiers and an output network. In particular an MPA is composed of an array of n HPAs in parallel and a pair of complementary nxn Buttler matrix networks, i.e. the input and the output network, that consist of 90° hybrid networks. A 90° hybrid network is a four port device equally splitting an input signal with a resultant 90° phase shift between the output ports. Thus, the signal at each input in the MPA is divided into n signals with particular phase relationships. These signals are amplified separately in each HPA, therefore, each amplifier operates on all input signals which enables the amplifiers to operate in their linear region and have equal gain and phase shift and then, the amplified signals are recombined in the output Buttler matrix. In this way, the signal at each input is amplified by all the HPAs, but assembled at the corresponding output [2.12]. The operation principle of a 4x4 MPA is illustrated in Fig.2.2.

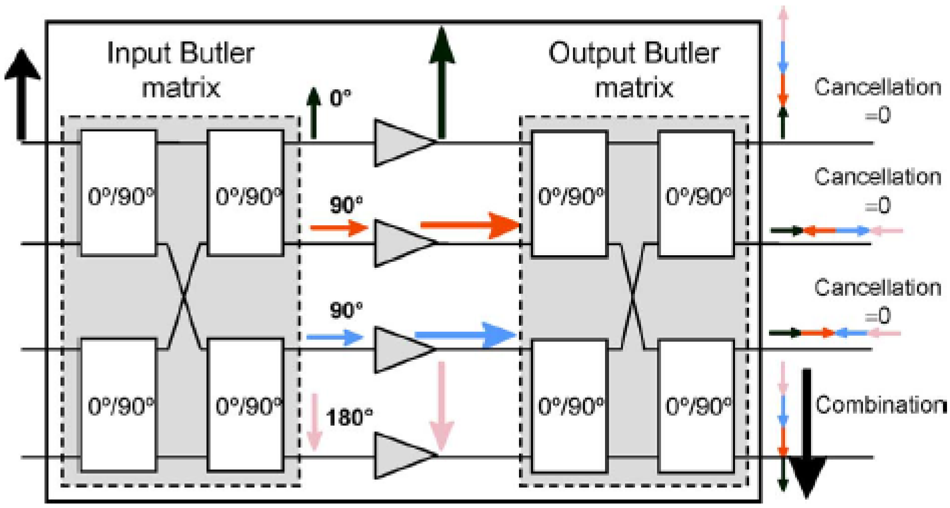


Fig. 2.2 4x4 MPA operation principle. Combination/cancelation principle for one input/output pair.

The main features of MPAs are the following [2.5]:

- The RF power assigned to any beam is a percentage of the total available RF power and all amplifiers contribute to each individual beam.
- The failure of one HPA does not cause the total loss of a beam.
- It is possible to flexibly distribute the total available RF power amid the antenna beams by properly adjusting the input signals.

An important drawback of the MPA is related to isolation losses between channels due to the different electrical characteristics of each path.

2.3.2 Flexibility in Reconfiguring the Frequency Plan

The need for modifying the frequency plan including both the uplink and downlink bands, channel frequencies and bandwidths may result due to a number of factors like the need to transfer the satellite to a different orbital position, a change in the status of spectrum allocation in different countries or regions and the ability to support different services operating in different bands. A first level of frequency plan flexibility is achieved when it is possible to downconvert the uplink to several downlink frequencies. Full frequency flexibility refers to the totally independent and flexible management of uplink and downlink frequencies.

Overcoming the limitations of analogue mechanical switches which in conventional payloads are combined with converters, the use of an on-board digital signal processor (DSP) has the advantage of individually filtering multiple sub-bands within a given band with a very fine granularity. DSPs can be further separated into transparent or regenerative depending on whether they operate on baseband digital signals or not. The generic transparent processor architecture is based on a fully flexible switching network which routes channels between a set of uplink and a set of downlink beams. For this purpose, the signals on each uplink beam are channelized using a digital frequency demultiplexer of fine granularity. The channels can then be routed between beams and mapped onto different frequencies at this fine granularity.

Another means for obtaining flexibility concerning the frequency plan is through the MPA. Specifically, as recombination of power from different HPAs is done within the MPA structure, the use of an OMUX with specific fixed bandwidth per channel can be avoided. Using, instead, a wideband output network offers flexible managing of the downlink frequency plan and differentiates the bandwidth allocated to each beam.

2.3.3 Coverage Flexibility

A multibeam antenna system usually consists of three components [2.6]:

1. A power distribution network, commonly called beamforming network (BFN), necessary when multiple feeds are attached to each beam
2. An array of feeds
3. An optical system to focus the radiated power increasing the antenna gain.

Coverage flexibility is related to the ability to in orbit modify the beam shape, the number of beams or the beam location on Earth and is primarily related to satellite antenna reconfigurability. Such flexibility can be provided: to a limited extent by passive antennas (steerable, zoomable or electro-mechanically reconfigurable); to a larger extent by semi-active or active antennas like array fed reflectors (ARF) or direct radiating arrays (DRAs) which, in addition, can provide a high degree of dynamic adaptation to changing traffic demands.

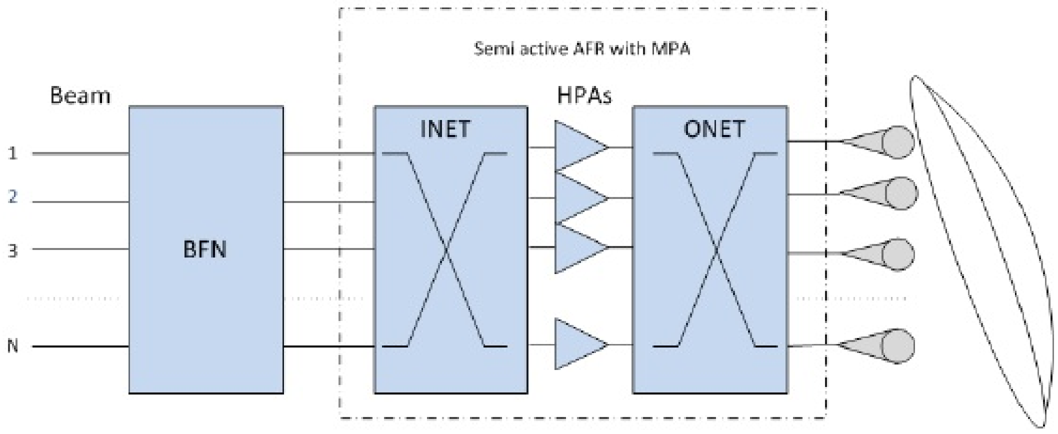


Fig. 2.3 Semi-active AFR with MPA

The AFR payload front-end is widely employed in satellite communication systems. It generally employs a BFN with a limited number of feeds per beam able to partially share the feeds among beams which illuminate a reflector in a focal configuration. When each HPA is directly connected to each feed, the AFR is active. It can also be semi-active when there is a stage of appropriately connected input and output networks respectively before and after the HPAs, i.e. when it is connected as an MPA. Fig. 2.3 depicts an example of this type of AFR implementation. On the other hand, the DRA is characterized by full sharing of feeds per beams without the adoption of a reflector. A simple DRA block diagram is depicted in Fig. 2.4. Note that active antennas may operate with either analogue or digital processing depending on the reconfigurability requirements and this of course will also affect the equipment complexity.

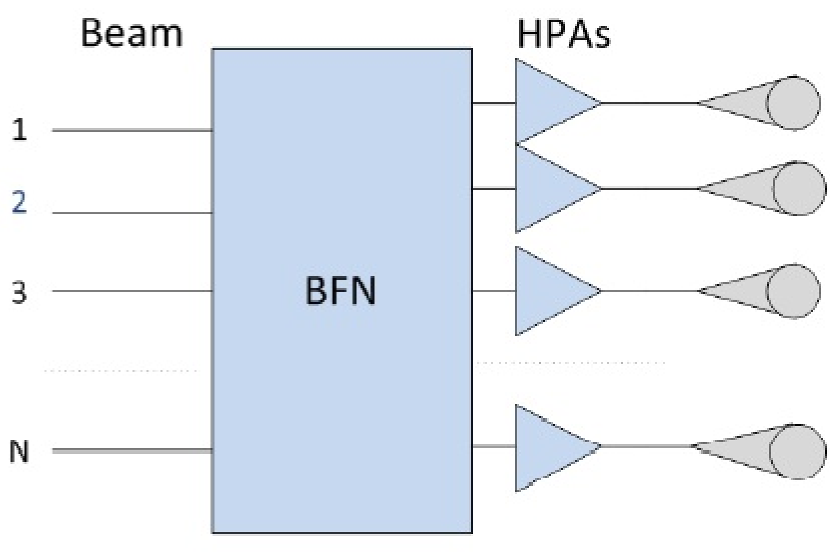


Fig 2.4 DRA

If the array of radiating elements is fed from the same signal with a specific amplitude and phase distribution, a shaped beam is obtained. The distribution is obtained by means of a set of phase shifters, couplers and power splitters, i.e. the BFN[5] [2.7]. Specifically, a BFN is defined as a device having N input ports corresponding to the number of channels (transponders) and M output ports (feeds). The device is able to generate P distinct beams which are orthogonal to each other in the sense that each channel can be independently routed to only one beam. In the simplest case P and M are equal, i.e. each feed generates one beam. In the general case, each beam is generated by 1 up to M feeds. Beamforming techniques can be either analogue or digital. In turn, analogue beamforming can be implemented either before or after the HPA stage. Connecting the BFN after the HPA would be preferred when the number of beams is smaller than the required number of feeds –thus fewer HPAs are required. If the weights on the paths of the BFN are variable, e.g. by employing adjustable gain and phase shifters, then the desired coverage flexibility is obtained. In the case of digital beamforming the antenna must be coupled via a DSP in a DRA or AFR configuration. Digital beamforming offers all the benefits of reconfigurable analogue beamforming, but with enhanced flexibility.

2.3.4 Connectivity and Routing Flexibility

When each uplink channel is always connected to the same downlink beam, or each uplink beam is always connected to the same downlink channel there is no flexibility with regard to connectivity. That is, the selection of the target region is achieved by choosing the uplink carrier frequency so that, after frequency conversion, it falls within the band of one of the channels allocated to the target region (beam). In contrast, when any receive antenna output is connected to several downlink channels and any downlink channel is connected to several transmit antenna inputs there is a high connectivity and routing flexibility. Moving from the (telecommand-controlled) conventional mechanical switches –whose operation is limited to simple channel re-routing and changing the number of channles/beam–, a DSP based solution for the switching matrix could provide a high degree of flexible interconnectivity between transmission channels or subchannels.

2.4 Formulation of the Multibeam Fixed Satellite Optimization Problem

2.4.1 General

The previous section presented the four axes of payload flexibility. In particular flexibility can be viewed with regard to coverage, power, frequency planning and higher layer functionalities (e.g. routing and switching). In exploiting these degrees of flexibility, multibeam satellites can play a key role since their design can easily be extended to support re-configurability of power and frequency plans as well as routing and switching functionalities. Thus, multibeam systems provide the necessary

[5] The term beamforming in the context of this section refers to the electronic equipment either in space or on the ground that implements the corresponding communication technique analyzed in other chapters and in the literature.

flexibility in order to dynamically reconfigure the power and frequency plan in response to the spatiotemporal variations of traffic demand.

However, due to frequency reuse, multibeam systems are subjected to inter-beam interference. Hence a dynamic allocation of system resources in order to satisfy the varying traffic demand while mitigating the resulting interference seems imperative for efficient multibeam systems. These system resources are the bandwidth and the transmit power. The resource allocation is an optimization problem that aims at maximizing the throughput of each beam taking into account the current traffic demand.

Since, in general, bandwidth allocation is done in terms of carriers of fixed bandwidth, the bandwidth allocation problem can be transformed into a power allocation problem where power is either allocated to the respective carrier or not. Furthermore, recent advances in flexible TWTA and particularly in MPA technology allow for efficient power splitting. Hence, the resource allocation problem turns into a power allocation problem where power is efficiently apportioned across beams to match the varying traffic demand.

2.4.2 Multibeam Architecture and System Resources

Assume a multibeam satellite system consisting of N beams employing a typical four color reuse pattern following the four color theorem [2.2]. In order to efficiently utilize the available bandwidth and mitigate interference, the four colors can be arranged as two colors in frequency and two in polarization. The available total downlink system bandwidth B_{TOT} is equally divided among the four colors. It is further assumed that the available bandwidth reused over the different colors is equally divided among the beams of each color and each of these beams accommodates four carriers. A carrier represents the elementary system entity for conveying different streams of information. The bandwidth of each beam is equally divided among its carriers. The multi-beam system in hand is depicted in Fig. 2.5.

Having appropriately allocated B_{TOT} among colors, beams and carriers, the resource allocation turns into the appropriate allocation of the total available system power, P_{TOT}. P_{TOT} is a function of the platform total DC power on board the satellite and must be appropriately allocated to each beam, so that the offered bit rate to each beam meets the relevant user requirements. The power allocated to each beam, P_b, will then be equally divided among the N_c carriers of the beam (i.e. $P_{b,c} = P_b / N_c$). The reason behind the uniform beam power allocation over the respective carriers is the fact that MPAs and TWTAs work on a very high number of carriers and adjusting the power of each of these carriers is a highly involved task.

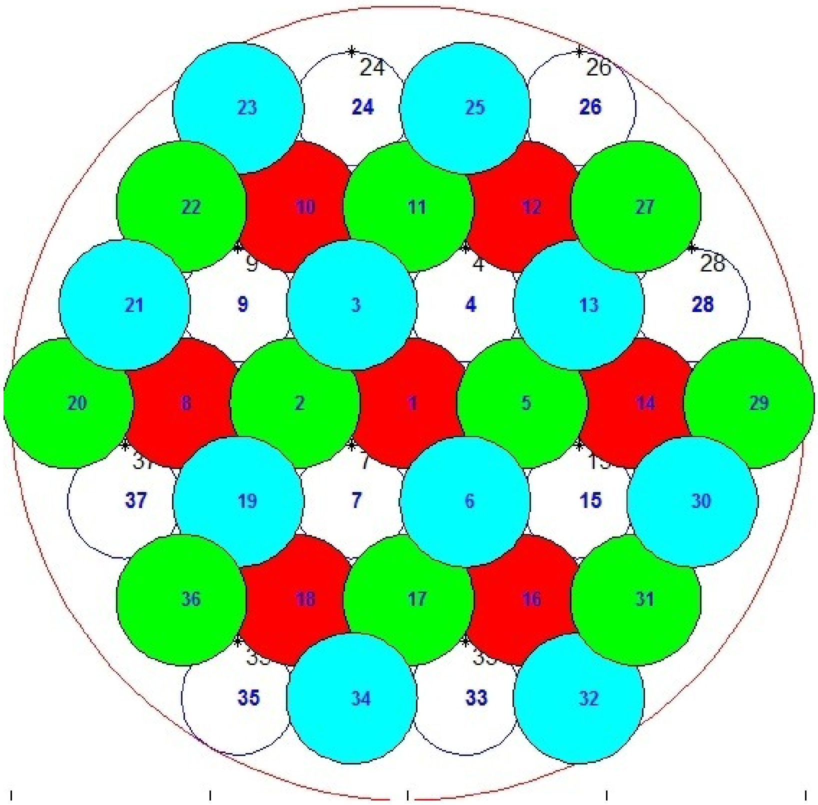


Fig. 2.5 Four color reuse pattern, for a 37 circular beam layout

2.4.3 Resource Optimization

The optimization target is to determine the beam power P_b, that:

Maximizes the system capacity.

subject to:

- The power constraints of the system
- The power constraints of each beam

Thus, the optimization variable of the problem is the vector x = (P_1, …, P_N), where each vector element represents the continuous power value of the respective beam. To maximize system capacity, the following four candidate cost functions have been employed in past studies. Let $R_{b,req}$ denote the bit rate requested by the users in beam b and $R_{b,off}$ denote the offered, bit rate by beam b.

1.Differential System Capacity (DSC) [2.10]

$$\text{minimize } f$$

$$f = \sum_{b=1}^{N} \mathrm{abs}\{R_{b,req} - R_{b,off}\} \tag{2.1}$$

According to (2.1), the performance of the system is evaluated by the absolute difference between the offered and requested beam capacities.

2.Unmet System Capacity (USC) [2.10]

$$\text{minimize } f$$
$$f = \sum_{b=1}^{N} \max\{R_{b,req} - R_{b,off}, 0\}, \qquad (2.2)$$

In (2.2), offered beam capacities exceeding the required do not contribute to the figure of merit and do not affect the optimization.

Satisfaction Factor (SF) [2.11]

$$\text{maximize } f$$
$$f = \frac{\sum_{b=1}^{N} \min\{R_{b,req}, R_{b,off}\}}{\sum_{b=1}^{N} R_{b,req},} \qquad (2.3)$$

The offered beam capacities exceeding the required capacity do not contribute to the figure of merit defined in (2.3). But in this case the objective value is scaled by the cumulative required capacity.

Aggregate Fitness (AF)

$$\text{maximize } f$$
$$f = \sum_{b=1}^{N} \left[1 + \text{abs}\{R_{b,req} - R_{b,off}\}\right]^{-1} \qquad (2.4)$$

According to (2.4), the performance of the system is again evaluated by the absolute difference between the offered and requested beam capacities, but these results are scaled down on a beam bases and the optimization is driven according to the sum of the individual beam fitness.

Subject to:

System power constraint (P_{TOT}):

$$\sum_{b=1}^{N} P_b \leq P_{TOT} \qquad (2.5)$$

Beam power constraints ($P_{b,con}$):

$$P_b \leq P_{b,con}, \quad b=1, \ldots, N \qquad (2.6)$$

$R_{b,req}$ is the output of the traffic model according to the statistical parameters that depend on the size of the beam, population covered by the beam, type of terminals within the beam, busy hours and the GDP per inhabitant. Furthermore, $R_{b,off}$ denotes the cumulative bit rate of all carriers in beam b, i.e.

$$R_{b,off} = \sum_{c=1}^{N_c} B * f_{DVB-S2}(SNIR_c) \qquad (2.7)$$

The function $f_{DVB\text{-}S2}(SNIR_c)$ quantifies the spectral efficiency of the various modulation and coding schemes employed by DVB-S2 as a function of the SNIR of carrier c.

2.4.4 SNIR Evaluation

The SNIR of carrier c characterized by transmission power $P_{b,c}$ and bandwidth B is determined from [2.1]:

$$\mathrm{SNIR}_c = \frac{a_b^2 P_{b,c}(\mathrm{OBO})}{N_0(a_b)B + \sum_{q\in\Phi} \alpha_q^2 P_{q,c}(\mathrm{OBO}) + I_{adj_{ch}}(B,\mathrm{XPD}) + I_{adj_{sat}} + I_{inter}(\mathrm{OBO}, C_b, \mathrm{Mod})} \quad (2.8)$$

Φ: is the set of co-channel beams in the coverage area having active carriers overlapping with the bandwidth of carrier c (co-channel interference with intended beam b)

a_q: is a gain incorporating the effect of (Output of link budget module):

- satellite antenna beam gain toward the intended (covered) region
- terminal receive antenna gain
- free space loss
- clear sky attenuation (according to ITU-R Recommendations)
- rain attenuation (according to ITU-R Recommendations)

a_b: is the beam gain incorporating the effect of the above parameters.

N_0: is the noise power spectral density which is a function of a_b because of the increase in noise temperature under rain fading conditions that affect a_b

I_{adj_ch}: accounts for adjacent channel interference due to filter imperfections (function of B), including spillover from the beams into orthogonal polarizations if both polarizations are employed (function of XPD)

I_{adj_sat}: Inter-system interference caused by adjacent satellites operating closely in orbit with the serving satellite

I_{inter}: Intermodulation interference. It is a function of the OBO, number of carriers through the HPA and the modulation scheme employed.

The notation $P_{b,c}(\mathrm{OBO})$ denotes the dependence on the OBO appropriate for the modulation scheme employed

Following the problem formulation an appropriate optimization technique must be adopt to solve the problem. In this course an overview of the available optimization techniques is performed in the next chapter in order to determine an efficient optimization algorithm for the optimization problem under consideration.

References:

[2.1] G. Maral and M. Bousqouet, *Satellite Communications Systems*, 5th ed. John Wiley & Sons, 2009.

[2.2] G. Gontlier, *“Formal proof the four-color theorem”*. Notices of the AMS, vol.55, no. 11, p. 13821393, November 2008.

[2.3] J. Weekley and B. Mangus, *“TWTA versus SSPA: A comparison of on-orbit reliability data”*. IEEE Trans. Electron. Devices, vol. 52, no. 5, pp. 650-652. 2005.

[2.4] A. Mallet, A. Anakabe, J. Sombrin, and R. Rodriguez, *“Multiport-amplifier-based architecture versus classical architecture for space telecommunication payloads”*. IEEE Trans. Microwave Theory Techn., vol. 54, no. 12, pp. 4353-4361, 2006.

[2.5] M. Aloisio, P. Angeletti, and M. Lisi, *“Mutliport amplifiers for satellite applications: Historical perspective and future directions,”* in IEEE Int. Vacuum Electron. Conf.(IVEC 2009), 2009, pp. 201–201.

[2.6] M. Sabbadini, *“Introduction to the design of antennas for space applications,”* March 2008.

[2.7] P. Angeletti, N. Alagha, and S. D. Addio, *“Space/ground beamforming techniques fo satellite communications,”* in 2010 IEEE Antennas and Propagation Society International Symposium (APSURSI), 2010.

[2.8] R. Rinaldo and R. Gaudenzi, *“Capacity analysis and system optimization for the forward link of multi-beam satellite broadband systems exploiting adaptive coding and modulation,”* International J. Satell. Commun. Network., vol. 22, no. 3, pp. 401–423, 2004.

[2.9] A. Morello and V. Mignone, *“DVB-S2: the second generation standard for satellite broad-band services,”* Proceedings of the IEEE, vol. 94, no. 1, pp. 210–227, 2005.

[2.10] Gabellini P., Gatti N., Gallinaro G., Vecchi G., Della Croce F., Rinaldo R., Angeletti P., *“Proposed Architecture of a System and Antenna/Payload Co-Design Tool for Multi-Beam Broadband Satellite Systems”*, ESA Workshop on Advanced Flexible Telecom Payloads, 18-20 Nov. 2008.

[2.11] Alberti X., Cebrian, J. M., Del Bianco A., Katona Z., Lei J., Vasquez-Castro M. A., Zanus A., Gilbert L., Alagha N., *“System Capacity Optimization in Time and Frequency for Multibeam Multi-media Satellite Systems”*, 5th Advanced Satellite Multimedia Systems Conference, Sep. 2010

[2.12] Mallet A., Anakabe A., Sombrin J., Rodriguez R., “Multiport-Amplifier-Based Architecture Versus Classical Architecture for Space Telecommunication Payloads”, IEEE transactions on microwave theory and techniques, vol.54, no.12, 2006

Chapter 3
Optimization Techniques

3.1 Introduction

A *mathematical optimization problem*, or just *optimization problem*, has the form:

$$\text{minimize } f_o(x) \tag{3.1}$$
$$\text{subject to } f_i(x) \le b_i \text{ , } i = 1, \ldots, m. \tag{3.2}$$

The vector $\mathbf{x} = (x_1, \ldots, x_n)$ is the *optimization variable* of the problem whereas the function $f_o : R^n \to R$ is the *objective function*, the functions $f_i : R^n \to R$, $i = 1, \ldots, m$, are the *constraint functions* and the constants b_i, $i=1,\ldots, m$, are the limits or bounds, for the constraints. A vector $\mathbf{x}^*$ is called *optimal* or a solution to the problem, if it has the smallest objective value among all vectors that satisfy the constraints. That is, for any z satisfying Equation 3.2 $f_o(z) \ge f_o(x^*)$ [3.1].

The optimization problem described above is an abstraction of the problem of choosing the best possible vector in R^n from a set of candidate choices. The set of candidate choices defines the search space of the problem. The problem search space is determined by the constraints $f_i(x) \le b_i$. These constraints represent either specifications or limitations of the problem. A choice belonging to the problem search space is represented by variable vector x and the respective performance is evaluated via the objective function f_o In particular, the objective value $f_o(x)$ represents the cost of choosing x[6]. Hence, the solution of the optimization problem corresponds to a choice that has the minimum cost or maximum utility among all choices of the defined search space.

The above optimization problem, known as a global optimization aims at obtaining the globally optimal solution x^* which minimizes the objective function over all choices in the search space. However depending on the specific characteristics of the problem, a compromise to avoid searching for the globally optimal and search for a locally optimal solution may be necessary. A locally optimal solution minimizes the objective function among feasible solutions that are near it but does not guarantee the lowest objective value among every feasible solution of the search space. Optimization schemes leading to a locally optimal solution achieve local optimization and constitute a widely used branch of mathematical optimization for problems and applications where a good, if not the best, solution is desirable.

A wide variety of problems encountered in practice can be formulated as mathematical optimization problems. Therefore, mathematical optimization is used in many diverse areas such as engineering design, bioinformatics, telecommunications,

[6] Alternatively the objective function - $f_o(x)$ can be employed representing the utility of choosing x, in which case the solution to the optimization problem (3.1) corresponds to a choice that has maximum utility.

aeronautics, transportation and finance. For most of these cases a human decision maker, system designer or system operator is necessary to supervise the process carrying out any actions suggested by the optimization method. However, the advances of micro-processor technology have enabled *embedded optimization*, namely real-time optimization without human intervention carried out by computers embedded in commercial devices. Thus, the use of optimization modules is pervasive in current systems.

As every process is likely to be optimized [3.2] the applications of mathematical optimization are numerous. In this course, a relation between a set of inputs (optimization variables) and a set of permissible outputs (optimization objectives) must be defined. Once such a relation is defined, it must be formulated in the form of a mathematical optimization problem as described above. First the objective function (or cost function) evaluating the performance of the optimization variable needs to be defined. Second the constraint functions have to be formulated taking into account the limitations of the problem and the set of permissible outputs. This concludes the *encoding* of the problem. An appropriate *optimization method* must be selected next.

3.2 Overview of Optimization Methods

An optimization method is an algorithm that evaluates a solution to an instance of an optimization problem under a specified accuracy. The classification of classical optimization methods, given in literature is shown in Fig.3.1 [3.2].

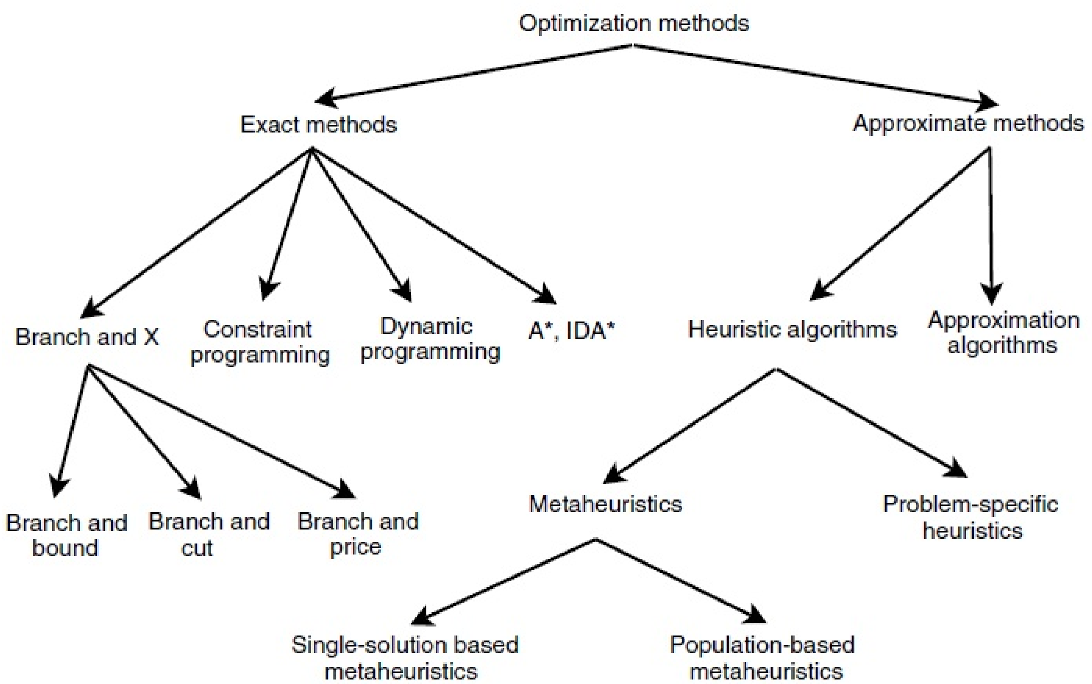


Fig. 3.1 Classification of classical optimization methods

As shown in Fig.3.1 the main differentiation of optimization methods is between exact and approximate methods. *Exact methods* guarantee the optimality of the solution, whereas *approximate methods* may determine a high-quality solution but do not guarantee to find the global solution. Evidently, optimization methods providing exact solutions are favored. However, a large number of real-life optimization problems are complex and difficult to solve. The inherent complexity of such problems lies in a number of factors, such as the form of the objective and constraint functions, the number of variables and constraints as well as the sparsity[7] of the problem [3.1]. As a result such problems cannot be solved exactly within reasonable time, necessitate the use of approximate methods.

Approximate methods [3.2] can be further classified into two classes: *Approximation Algorithms* and *Heuristic Algorithms*. Approximation algorithms provide guarantees on the distance of the obtained solution from the global optimum. Thus, if the complexity of the problem allows the use of an approximation algorithm, the solution provided may not be exact but it will be within a certain distance from the exact solution. If, however, this is not the case the main alternative to solve such problems is to use *metaheuristics,* a subclass of heuristic algorithms, which are general purpose algorithms that can be tailored to solve any optimization problem unlike specific heuristics that are designed and tailored to solve specific problems. Therefore, metaheuristics can be used to solve complex problems with large solution search space. This is done by reducing the effective size of the search space and exploring this reduced space efficiently. Thus, complex problems can be solved faster and satisfactorily with no guarantee, however, on the optimality of the obtained solution.

Metaheuristics are further classified according to the number of solutions used for the efficient exploration of the problem search space. The relevant categories are the single solution based (or trajectory based)[8] and the population based metaheuristics. The main difference between these classes is the exploration pattern they use in order to reduce the effective size of the search space. This notion will be made clear later on, when the concepts of exploration and exploitation are introduced.

All classes of optimization methods mentioned above are described in the following pages with emphasis put on their suitability to solve the resource allocation problem in hand. The analysis demonstrates the suitability of population based metaheuristics.

3.2.1 Exact Methods

Exact methods [3.2] are used to solve complex problems by breaking them down into simpler sub-problems and then combining the solutions to reach an overall solution. Those methods can be viewed as tree search algorithms where the search is carried out over the whole search space of interest, but only certain sub-trees of the search

[7] A problem is considered sparse when each constraint function depends on a small number of the variables.

[8] The terms “single solution based metaheuristics” and “trajectory based metaheuristics” are used interchangeably.

space are examined when solving the corresponding sub-problem. The key idea behind this approach is that many of these sub-problems are often the same; thus by solving them only once, the computation effort is drastically reduced. This property is extremely useful when dealing with large instances of difficult problems, hence with a large number of recurrent sub-problems. The efficiency of an exact method, however, does not depend solely on the number of the redundant sub-problems but, also, on the structure of the problem, making ideal for solving large instances of difficult problems if their structure is appropriate[9] [3.2].

Exact algorithms have been widely used and the most popular exact optimization strategies, like Dynamic Programming, Constraint Programming and Branch and X family of algorithms[10], find applications in many real world problems like the Graph Coloring, Sequential Ordering and Quadratic Assignment problems. However, it is obvious that there is at least one key attribute a problem must have so that exact algorithms are applicable, namely having an optimal substructure. A problem has an optimal substructure when the solution to the optimization problem can be obtained combining optimal solutions to its sub-problems. Furthermore, exact algorithms perform better in discrete and combinatorial optimization[11] problems, where the set of feasible solutions is discrete or can be reduced to discrete. A common problem involving combinatorial optimization is the Travelling Salesman Problem (TSP).

However the satellite resource optimization problem does not have an optimal substructure and therefore exact algorithms cannot be applied. The substructure of the problem is not optimal due to the interdependence of all beams due to the inter-beam interference as well as to the dependence of beam powers on the total available system power. Furthermore, the resource allocation problem cannot be classified as a combinatorial optimization problem, in case the optimization variables of the problem are continuous and the set of candidate solutions infinite. That is the case of the problem, described in Section 2.4.3, where each optimization variable represents the continuous power value of the respective beam. The continuous optimization is opted to avoid quantization errors inherently present in discrete optimization.

[9] Small instances of a specific problem may not be solved by an exact algorithm whereas large instances may be solved by the very same algorithm. Hence, recurrent sub-problems the algorithm failed to solve for some small instances are solved because of the problem structure. The following table tabulates examples found in the literature regarding different instances of popular optimization problems (Sequential Ordering Problem, Quadratic Assignment Problem, Graph Coloring)

Optimization Problem	**SOP**	**QAP**	**GC**
Size of some unsolved instances	53	30	125
Size of some solved instances	70	36	561

[10] Branch and X family of algorithms: (Branch and Bound, Branch and Cut, Branch and Price).

[11] Combinational optimization designates the procedure of finding an optimal solution from a finite set of candidate solutions.

3.2.2 Approximate Methods

When exact methods are not suitable for solving the problem in hand, approximate methods [3.2] have to be used instead. There are two kinds of approximate methods, namely *Heuristic algorithms* and *Approximation algorithms*. To minimize the objective function $f_o(x)$ of the problem as described in Section 3.1, a heuristic algorithm will search over x in some systematic way whereas an approximation algorithm is a model based approach that will use an easily minimized approximation to $f_o(\ )$ to guide its search [3.3]. Moreover, the approximation algorithm will provide a guarantee on the distance of the obtained solution from the global optimum.

An ε-*approximation algorithm* [3.2] determines an approximate solution α which is

- not less than a factor ε times the global optimum solution s:

$$\varepsilon \cdot s \leq \alpha \tag{3.3}$$

$$\text{if } \varepsilon < 1 \tag{3.4}$$

- not greater than a factor ε times the global optimum solution s:

$$\alpha \leq \varepsilon \cdot s \tag{3.5}$$

$$\text{if } \varepsilon > 1 \tag{3.6}$$

where the ε factor can be either a constant or a function of the input instances. An absolute performance guarantee is also possible if the following property holds:

$$(s-\varepsilon) \leq \alpha \leq (s+\varepsilon) \tag{3.7}$$

Such ε-approximation algorithms are the Quadratic Interpolation as well as the Newton-Raphson and Quasi-Newton [3.3]. However, to apply such techniques, an approximation to $f_o(\ )$ that will guide the search is necessary . In case of the Quasi-Newton technique, for example, this guide is the gradient of the function, so the function needs to be continuous and differentiable, to define the approximation to $f_o()$. Thus, it is obvious that an approximation to $f_o(\ )$ is not always available and therefore, not every optimization problem can be solved following these approaches.

This is also the case of the problem discussed in Chapter 2. In particular, on the one hand, the evaluation of SNIR is very complicated and on the other hand look-up tables have to be employed to determine the spectral efficiency for specific SNIR values. Thus, SNIR is related to the corresponding spectral efficiency via a non-continuous relation. That is, determining an approximation to the problem is a highly involved task. In this course, the main alternative is the use of heuristic methods.

Heuristic methods [3.2] (Greek: "εὑρίσκω", "find") refer to problem solving procedures that utilize experimental, especially trial and error, methods. In contrast to the approximation algorithms, though, Heuristics do not have an approximation guarantee on the obtained solution but they deliver satisfactory solutions for large size problem instances. Heuristic methods can be further classified into two classes, namely *Specific Heuristics* and *Metaheuristics.*

Specific Heuristics are designed and tailored to solve a specific problem or instance. To enhance their performance and efficiency such problem specific heuristics can be combined and guided by higher level strategies which are called metaheuristics. Due to their efficiency metaheuristics are selected to solve the problem under consideration (Section 2.4).

3.3 Metaheuristics

The metaheuristics [3.2] are general-purpose algorithms that can be applied to solve almost any optimization problem. The suffix meta (Greek: "μετά" , after) means "later or more highly organized or specialized form of something", so the term Meta-heuristic refers to upper level general methodologies (templates). Those can be used as guiding strategies in designing underlying heuristics to solve specific optimization problems. Metaheuristics, like every Heuristic method, do not provide any guarantee on the optimality of their solution or even on the distance of the obtained solution from the optimal. However, metaheuristics can search very large search spaces of candidate solutions, in a reasonable time from a practical point of view, providing satisfactory solutions. Moreover, metaheuristics make few or no assumptions about the optimization problem, providing high quality solutions to complex problems like the one studied in this Thesis. On the contrary other optimization methods require that certain preconditions are valid. As a result, the application of metaheuristics has found its way into a number of areas including: [3.2]

- Engineering design: topology and structural optimization in electronics and VLSI, aerodynamics, fluid dynamics, telecommunications and robotics.
- Machine learning and data mining in bioinformatics and computational biology and finance.
- System modeling, simulation and identification in chemistry, physics and biology; control, signal and image processing.
- Planning in routing problems, robot planning, scheduling and production problems, logistics and transportation, supply chain management.

In all the above cases, different metaheuristic methods are used depending on the characteristics of the problem to solve. Specifically, once the use of metaheuristics is decided, two contradictory aspects must be traded off, namely *exploration* of the search space and *exploitation* of the best solution(s). Exploration consists in an extensive search to make sure that all regions of the search space have been explored and the search is not confined to a certain region. Upon completion of the exploration of the search space, the regions containing the best solutions are identified. These regions are the most likely to contain a satisfactory solution. Exploitation, on the other hand, is an intensive and exhaustive search within a region of the search space, likely to contain a satisfactory solution, in the attempt to discover the best solution of this region.

In case the problem under consideration necessitates the exploration of the search space, the search has to be diverse, whereas if exploitation is necessary an intensive local search must be carried out. In fact, the more diverse the search the less intensive it is and vice-versa. Therefore, the nature of the problem must be determined in advance, so that either a diverse or an intensive metaheuristics technique is selected. Population-based metaheuristics focus on the diversification of the search while trajectory-based metaheuristics on the intensification. This also constitutes the main classification criterion for metaheuristics[12].

3.3.1 Trajectory-based Metaheuristics

Trajectory-based techniques [3.2] aim at improving a single solution by following a certain search trajectory through the search space. This trajectory is determined applying an iterative procedure that chooses the next solution based on the current solution. At each step of the procedure, a set of candidate solutions are generated applying a certain operator to the current solution. This new set of solutions, ideally, located in the vicinity of the current solution, constitutes the neighborhood of the current solution.

The properties of the neighborhood, especially its locality[13], are of great importance to the convergence of the algorithm. A strong locality ensures the intensity of the search and the exploitation of the current solution whereas a weak locality could result in a random search within the whole search space. Hence, the operator generating the neighborhood has to be chosen carefully. After a new neighborhood has been determined, a neighbor has to be selected to replace the current solution and become the next station of the search trajectory[14]. The kind of solution that will replace the current one depends on the optimization technique used and the way the trajectory is planned. Even points with a higher objective function value than the current one – in case of a minimization problem - can be accepted, if the trajectory planning allows so, e.g. in order to escape out of a local optimum.

After the previous solution has been replaced by a new one, the step is completed. The same procedure iterates and all steps concatenate into forming the search trajectory of the problem, until one of the stopping criteria is met. These criteria can be either a priori known (e.g. a predefined number of iterations) or a posteriori (e.g. a number of non-improving iterations). At this point the process is terminated and the current solution constitutes the result of the optimization.

[12] Many classification criteria can be found in the literature, (Nature inspired/Non-nature inspired, Memory usage/Memoryless methods, Deterministic/Stochastic, Iterative/Greedy) but the classification under population-based/trajectory-based is the main criterion when selecting a metaheuristic technique, for the problem of interest. As a matter of fact, algorithms belonging to each of these two classes use the same search mechanisms, so this type of classification is helpful.

[13] The locality of the neighborhood is determined by the Euclidean norm for a continuous problem and by the Hamming distance, as defined in information theory, for a discrete.

[14] The generation and the replacement of the current solution could be a memoryless process, in which case every operation and decision is based on the current solution, or a memory based process, where, a memory stored, search history is being used for every operation and decision to be made.

3.3.2 Population-based Metaheuristics

Unlike trajectory-based techniques, population based metaheuristics [3.2] do not start the search from a single solution but from an initial population of solutions. This diversity of initial solutions is the very reason why an extensive exploration of the search space is possible with this family of techniques. In any case, the initial population must be diverse enough, so that the method reaches its full potential. Otherwise, the search will be confined to the region encompassing adjacent solutions of the initial population.

Because of the importance of the initialization to the convergence of the algorithm, a number of strategies have been proposed to deal with this[15]. However, the random generation is usually applied to the majority of the problems [3.2]. Although the algorithmic generation of independent random numbers is impossible, the random generation does not have to be performed by pseudo-random numbers. Quasi-random sequences of numbers can be used instead, where emphasis is put on both the independence and the dispersion of the successively generated numbers that constitute the initial population, ensuring the diversity of the initial random population.

After the initialization of the population an iterative procedure is performed. Specifically:

1. In each iteration a new population of solutions is generated
2. Individuals, from the new and the current population, are selected, based on their performance. i.e. their objective value.
3. Selected individuals replace the current population[16].
4. The procedure keeps iterating until one of the stopping criteria is met.

At this point the best individual of the population (i.e. the best solution out of the population of solutions) is selected and returned as the outcome of the optimization.

3.3.4 Metaheuristic Techniques

The classification of metaheuristics under trajectory-based and population-based techniques is given in Fig.3.2, where some of the well-studied algorithms of each class appear. Both classes of metaheuristic methods along with the relevant algorithms are applicable to the problem in hand; one of these algorithms must be selected to solve the problem. However, there is no general rule to follow when choosing an algorithm from the metaheuristics family and, as also suggested by the “No Free Lunch” theorem[17] [3.4], no universal criterion can prove the superiority of a specific technique over others. For this reason, a general intuitive criterion may be adopted to provide hints regarding the suitability of each metaheuristic algorithm for

[15] Random Generation, Sequential Diversification, Parallel Diversification, Heuristic Initialization.

[16] The generation and replacement can again be a memory less or a memory based procedure.

[17] “No free lunch” theorem,(2005), derivative of the “no free lunch theorems for optimization” by Wolpert and Macready, states that any two optimization algorithms are equivalent, when their performance is averaged over all possible optimization problems.

the problem in hand. This criterion associates the search pattern of each algorithm with the landscape of the problem search space. Given that algorithms belonging to the same class of metaheuristics share, to some extent, the same search mechanism, an early differentiation between trajectory-based and population-based techniques is possible. Indeed, the foregoing description of the two classes denotes that population-based methods employ an exploratory search pattern whereas trajectory based methods employ an intensive search pattern focusing on exploitation. However, high dimensionality[18] and the large size of the search space of the problem in hand suggest the use of an exploratory search pattern to ensure the dispersion of the search over the large search space. This infers the suitability of the population-based techniques over the trajectory-based for the current resource allocation problem.

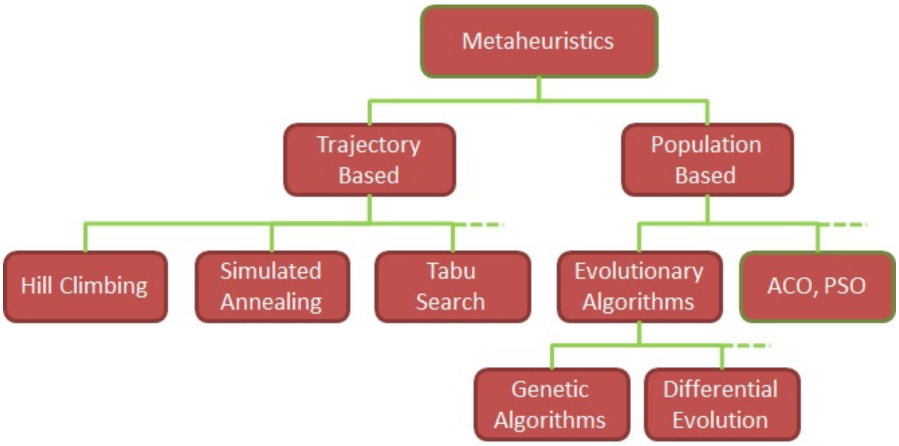


Fig. 3.2 Classification of metaheuristics

(*ACO: Ant Colony Optimization, *PSO: Particle Swarm Optimization)

In addition to the large size, the multimodal landscape of the problem search space suggests the use of an *Evolutionary Algorithm*, among the population-based techniques, particularly the use of a *Genetic Algorithm*. This is due to the fact that the crossover operator of Genetic Algorithms (as well as the recombination operator of Differential Evolution) provides long (random) jumps in the space of possibilities, thus providing a way out of local optima of the multimodal landscape [3.5].

The perspective obtained from this early approach is extremely useful in providing a rough outline of the suitability of each algorithm for the current problem. However, the suggestions of this early intuitive approach must be corroborated in practice. For this reason a systematic study is performed to check the validity of the intuitive approach leading to the selection of an appropriate technique. In this course, the performance of the *Genetic Algorithms* and *Differential Evolution* from the class of Evolutionary Algorithms, the performance of yet another population-based technique, namely the *Particle Swarm Optimization* and the performance of a trajectory-based

[18] The high dimensionality refers to the large number of beams and consequent large number of variables of the problem. Evident of the high dimensionality is the fact that multi-beam systems can have hundreds of beams.

technique, namely the *Simulated Annealing* are compared. The results of this study are presented in the next chapter.

The standard metaheuristic techniques systematically compared in the next chapter, namely the Simulated Annealing, Genetic Algorithms, Differential Evolution and Particle Swarm Optimization are presented in the following pages.

3.3.4.1 Simulated Annealing (SA)

Simulated annealing (SA) [3.2] was developed in the 1980s by the pioneering works of S. Kirkpatrick [3.6] and V. Cerny [3.7] and was originally applied to graph partitioning and VLSI design. In these early approaches SA was used in solving combinatorial optimization problems. Soon, it was extended to deal with continuous optimization problems. This method models the physical process of heating a material and, then, of slowly lowering the temperature to create a strong crystalline structure, thus minimizing system energy. The strength of the structure depends on the rate of cooling. If the initial temperature is not sufficiently high or fast cooling is applied, imperfections come up and, due to the structural defects of the material, minimization of the system energy is not attained. Hence, the initial temperature and the rate of temperature reduction are essential parameters of the annealing process.

The SA is a trajectory-based optimization technique imitating the aforementioned physical process. That is, the SA algorithm simulates the energy changes in a system subjected to a cooling process until it converges to an equilibrium state. This physical procedure is analogous to an optimization method. In particular, the objective function of the problem is analogous to the energy state of the system and the solution of the optimization problem corresponds to a system energy state. The decision variables associated with a solution are analogous to the molecular positions within the material creating its crystalline structure and the global optimum corresponds to the ground state of the system energy.

As in all trajectory-based techniques the search starts from a single point defined in advance. In each iteration of the SA algorithm, a new point is randomly generated. The distance of the new point from the previous point, determining the extent of the search, is based on a probability distribution with a scale proportional to the temperature. The function used to define the extent of the search (i.e. the size of the *neighborhood*) for the next iteration is the *annealing function* and is an essential operator of every trajectory-based technique. Two annealing functions used widely in real-life problems are *Fast Annealing* and *Boltzmann Annealing*. In case of the fast annealing the distance of the new point from the current point must not exceed the value of the temperature whereas in Boltzmann annealing the distance must not exceed the square root of the temperature.

The new points of the trajectory have, generally, a smaller objective value (i.e. a better performance) than the previous points, but under a certain probability, points raising the objective value may be accepted as new points. By accepting points that raise the objective, the algorithm is able to escape out of local minima. This feature of SA is shown in Fig.3.3. The new solution x′ can be accepted by the algorithm in spite raising the objective value. The probability of accepting the worst performing x′ depends, on the one hand, on the objective value of x′ and on the other hand, on the temperature. In fact, at the early stages of the search of Fig.3.3 when the temperature is higher the probability of accepting a move raising the objective value is also higher whereas at a given temperature, the lower the objective value of x′ the higher the probability of accepting it. Evidently a new point with a lower objective value than the previous point is always accepted.

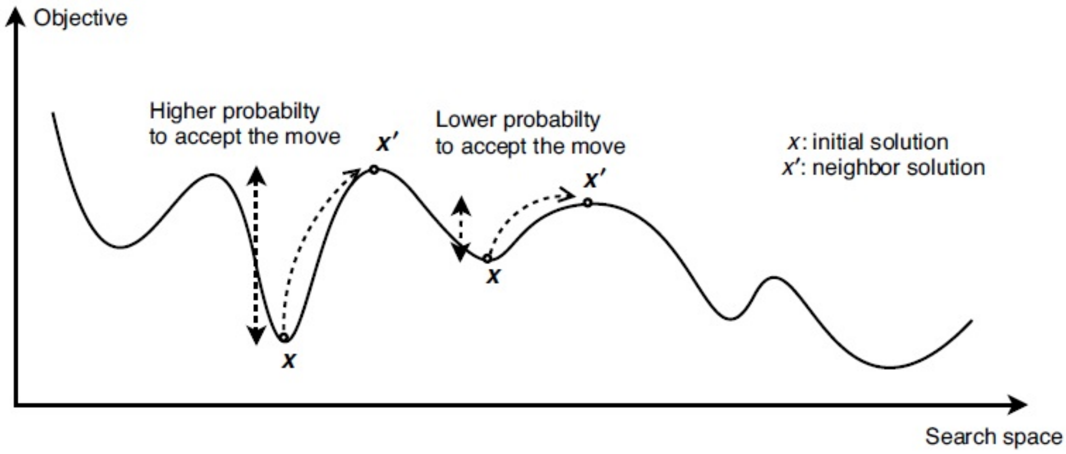


Fig 3.3 Simulated annealing escaping from local optima

Subsequently to the acceptance of a new point, an annealing schedule is selected to systematically reduce the temperature as the algorithm evolves. This task is performed by the *temperature update function,* and as the temperature is reduced, the algorithm reduces the extent of its search to converge to a minimum. Two common temperature update functions found in literature are the *Exponential Temperature Update* and the *Logarithmic Temperature Update.* The exponential temperature update denotes an exponential dependence of the temperature on the number of iterations. Let T_{max} be the initial temperature and let k denote the iteration number. The temperature in the k^{th} iteration T_k is equal to:

$$T_k = T_{max} \cdot (0{,}95 \wedge k) \tag{3.8}$$

Whereas the logarithmic update denotes a logarithmic dependence of the temperature on the number of iterations. Thus, the temperature in the k^{th} iteration is equal to:

$$T_k = \frac{T_{max}}{\ln(k)} \tag{3.9}$$

The SA algorithm described above is the standalone SA found in the literature and as used in the systematic study of standard techniques in the next chapter. However, a new feature of the SA, diverging from the standard theory of trajectory-based techniques - hence not included in the stand alone SA is the concept of *re-annealing*.

Re-annealing raises the temperature after the algorithm accepts a certain number of new points and starts the search again at the higher temperature. Thus, the algorithm avoids getting trapped at local optima.

3.3.4.2 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) [3.2] is a population-based metaheuristic, originally designed for continuous optimization problems and was first introduced by R.C. Eberhart and J. Kennedy in 1995 [3.8]. PSO mimics the social cooperative and competitive behavior of swarms. As described by Eberhart and Kennedy, the PSO algorithm is an adaptive algorithm based on social psychological metaphor; a population of individuals (referred to as *particles*) adapts by returning stochastically towards previously successful regions [3.9].

Like every population-based technique the N *particles* of the *swarm* flying around in a D-dimensional search space have to be randomly initialized. Each particle i of this swarm represents a candidate solution to the problem and is parameterized by its *position* x_i and *velocity* v_i in the decision space. The move of each particle is influenced by its personal success and the success of its peers. During each generation each particle is accelerated toward the best position visited by itself p_i and the global best position p_g, visited by its peers. At each iteration a new velocity value for each particle is calculated based on its current velocity v_i, the distance from its previous best position $(x_i – p_i)$ and the distance from the global best position $(x_i – p_g)$. The new velocity value is then used to calculate the next position of the particle in the search space. This process is iterated until some stopping criteria are met. The velocity update and position update are the two primary operators of PSO.

According to the previous description a particle must be composed of the following three vectors in order to update its velocity and position:

- The x-vector, recording the current position of the particle in the search space.
- The p-vector, recording the location of the best solution found so far by the particle.
- The v-vector, containing the velocity (i.e. the direction) of the particle in the search space.

As well as two fitness values (i.e. objective values): The x-fitness recording the fitness of the x-vector and the p-fitness recording the fitness of the p-vector.

Then at each iteration, each particle will apply the following operations:

Update velocity:

$$v_i(t) = w \cdot v_i(t-1) + \rho_1 \cdot C_1 \cdot (p_i – x_i(t-1)) + \rho_2 \cdot C_2 \cdot (p_g – x_i(t-1)) \qquad (3.10)$$

In this expression ρ_1 and ρ_2 are random variables $\epsilon[0,1]$. The constant C_1 represents the cognitive factor, namely the attraction of the particle toward its own success and the constant C_2 represents the social factor, namely the attraction toward the best success of the swarm. Finally the inertia weight w controls the impact of the particle momentum to the velocity of the next iteration. In fact the inertia weight represents the trade-off between the explorative and exploitative search mode, since a large inertia weight encourages a global exploration whereas a small inertia weight encourages the local exploitation.

However the velocity v_i is limited by the permissible values of the system. Hence if the velocity exceeds the maximal value V_{max}, ($v_i > V_{max}$, $v_i < V_{max}$, respectively) it will be reset to V_{max}. ($-V_{max}$ respecively).

Update position:

$$x_i(t) = x_i(t-1) + v_i(t) \tag{3.11}$$

Thus each particle updates its position according to its updated velocity. The position update is shown in Fig.3.4.

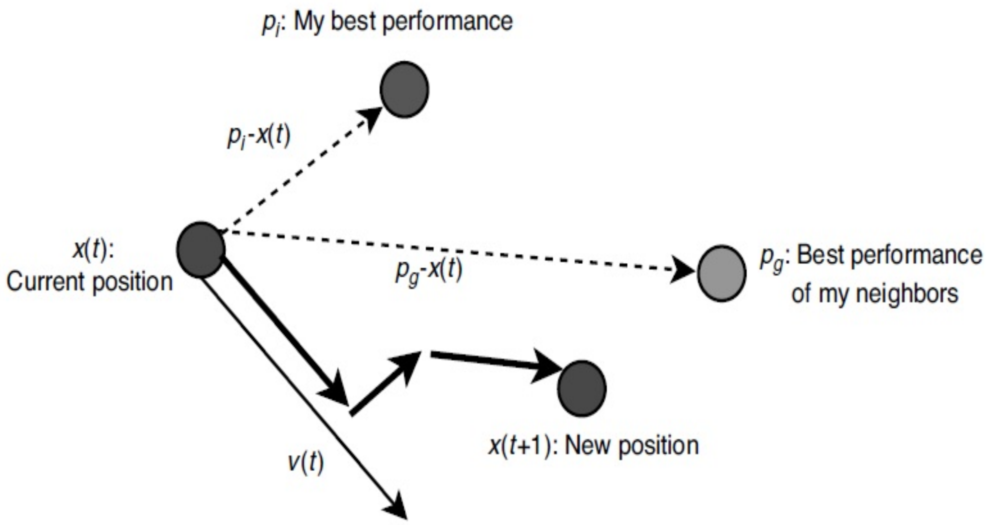


Fig 3.4 Particle Position Update

Update best found particles:

Each particle has a potential of updating the global best solution, therefore:

$$p_g = x_i \tag{3.12}$$

$$\text{if } f(x_i) < p_g \tag{3.13}$$

As well as the local best solution:

$$p_i = x_i \tag{3.14}$$

$$\text{if } f(x_i) < p_i \tag{3.15}$$

At the end of the process the global best solution p_g returns as the result of the optimization. An algorithm describing the PSO is Algorithm 3.1.

Algorithm 3.1 Particle Swarm Optimization algorithm.

Random initialization of the whole swarm ;
Repeat
    Evaluate $f(x_i)$;
    For all particles i
        Update velocities:
            $v_i(t) = v_i(t - 1) + \rho_1 \cdot (p_i - x_i(t - 1)) + \rho_2 \cdot (p_g - x_i(t - 1))$;
        Move to the new position: $x_i(t) = x_i(t - 1) + v_i(t)$;
        If $f(x_i) < f(pbest_i)$ **Then** $pbest_i = x_i$;
        If $f(x_i) < f(gbest)$ **Then** $gbest = x_i$;
        Update(x_i, v_i) ;
    EndFor
Until Stopping criteria

3.3.4.3 Differential Evolution (DE)

Differential Evolution (DE) [3.2] is a population-based metahueristic inspired from natural biological evolution and is one of the most successful approaches for continuous optimization. DE was developed by K. Price [3.10] in 1995 in an attempt to solve the Chebycheff polynomial fitting problem posed by R.Storm. The main idea behind DE is the use of vector differences to shake up the vector population and diversify the exploration of the search space. In this direction a *recombination operator* (based on the vector differences of two or more random individuals of the population) is applied providing random leaps in the space of possibilities. Moreover a self-referential *mutation operator* is applied to drive the search toward "good" solutions.

Like every population-based technique DE generates a random initial population of size k ($k \geq 4$). Each individual i of the population is a D-dimensional vector x_{ij} of floating-point elements. Each element of the i-th vector is generated within the acceptable bounds of each variable as follows:

$$x_{ij} = x_j^{lo} + rand_j[0,1] \cdot (x_j^{hi} - x_j^{lo}),\ \ i \in [1, k], j \in [1, D] \tag{3.16}$$

where x_j^{lo} denotes the lower bound and x_j^{hi} the upper bound of the j-th element and $rand_j$ is a uniformly distributed random variable in the range [0, 1].

Once the initial population is generated the recombination operator is applied. The recombination operator is a linear combination of usually three random individuals. Given a parent i and three randomly selected individuals r_1, r_2, r_3. Each dimension j of the offspring of i will be a linear combination of the j-th dimension of the three selected individuals according to a certain probability. This probability depends on the recombination probability CR $\in[0, 1]$ and the random dimension j_{rand} to ensure that at least the j_{rand} dimension will be evolved regardless of the value of CR. That means that when ($rand_j[0, 1] < CR$ or $j = j_{rand}$) the j-th element of the offspring of i will be the linear combination of the j-th element of r_1, r_2, r_3. Otherwise the offspring variable j will inherit the value of its parent. The aforementioned linear combination is

$$\text{Offspring: } u_{ij} = x_{r3j} + F \cdot (x_{r1j} - x_{r2j}) \quad (3.17)$$

Where the parameter F represents a scaling factor F ϵ[0, 1] controlling the amplification of the difference between the individuals r_1 and r_2 to avoid stagnation of the search process. The recombination operator algorithm described above will be made clear with the Algorithm 3.2.

Algorithm 3.2 Recombination operator in DE.

Input: Parent i, three randomly selected individuals r1, r2, r3, i ≠ r1 ≠ r2 ≠ r3.
$j_{rand} = int(rand_i[0, 1] \cdot D) + 1$;
For (j = 1, j ≤ D, j + +) Do
  If ($rand_j[0, 1]$) < CR) or ($j = j_{rand}$) **Then**
    $u_{ij} = v_{ij} = x_{r3j} + F \cdot (x_{r1j} - x_{r2j})$;
  Else
    $u_{ij} = x_{ij}$;
Output: Offspring u_i.

After the recombination operator is applied to every parent (i) of the population an elitist approach is adopted to determine which of the offspring will further contribute to the population evolution and which will be discarded. As a result of this elitism the offspring replaces its parent if its objective value is better or equal to the objective value of the parent. That is:

$$x_i(t+1) = \begin{cases} u_i(t+1) & \text{if } f\big(u_i(t+1)\big) \leq f\big(x_i(t)\big) \\ x_i(t) & \text{otherwise} \end{cases} \quad (3.18)$$

Thus each individual (i) of the population is iteratively evolved and the evolution process consists of two steps.

- *Mutation and Recombination:* In this step randomly selected members of the population create a variant solution. Next this variant solution is recombined with the individual i creating a trial solution.
- *Replacement:* The trial solution generated is compared to the original individual i and the best performing solution replaces the other in the population.

This process iterates for a number of generations and at the end of the process the global best solution x_i of the population returns as the result of the optimization. The whole DE algorithm described above is clearly presented in the following algorithm (3.3).

Algorithm 3.3 Differential Evolution algorithm.

Input: Parameters: F (scaling factor), CR (crossover constant).
Initialize the population (uniform random distribution) ;
Repeat
    For (i = 1, i ≤ k, i + +) Do /* Each individual */
        Mutate and Recombine:
            $j_{rand} = int(rand_i[0, 1] \cdot D) + 1$;
            For (j = 1, j ≤ D, j + +) **Do**
                If ($rand_j[0, 1]$) < CR) or ($j = j_{rand}$) **Then**
                    $u_{ij} = v_{ij} = x_{r3j} + F \cdot (x_{r1j} - x_{r2j})$
                Else
                    $u_{ij} = x_{ij}$
            Replace:

$$x_i(t+1) = \begin{cases} u_i(t+1) & \text{if } f(u_i(t+1)) \le f(x_i(t)) \\ x_i(t) & \text{Otherwise} \end{cases}$$

    End For
Until Stopping criteria /* exceed: a given number of generations */
Output: Best population or solution found.

The obtained solutions from the variation operators (i.e. recombination and mutation) may exceed the bounds of the system. In those cases a repair strategy has to be followed to reset the exceeded values and in this course *extreme repair strategies* as well as *intermediate repair strategies* are found in literature. The first extreme strategy resets the variable to the limit it exceeds. However this strategy decreases the diversity of the population. In order to ensure the diversity of the population another extreme strategy has been proposed, reinitializing the offending value to a random value, but evidently this strategy is extreme in terms of diversity maintenance and decreases the exploitation of the search hindering the convergence of the algorithm.

In order to deal with the drawbacks of the extreme strategies intermediate strategies may be applied instead. An example of intermediate strategy consists in reinitializing the exceeded value to an intermediate point between its previous value (before variation) and the bound it exceeds. That is:

$$u_{ij}(t+1) = \begin{cases} \frac{x_{ij}(t)+x_j^{lo}}{2} & \text{if } u_{ij}(t+1) < x_j^{lo} \\ \frac{x_{ij}(t)+x_j^{hi}}{2} & \text{if } u_{ij}(t+1) > x_j^{hi} \\ u_{ij}(t+1) & \text{otherwise} \end{cases} \tag{3.19}$$

Because of the simplicity and effectiveness of DE in solving continuous optimization problems DE had a major impact on the field of continuous optimization. Furthermore DE is very easy to tune and a number of empirical rules simplify the tuning process. The tuning of DE as also the tuning of every technique presented herein will be discussed in the next chapter.

3.3.4.4 Genetic Algorithms (GA)

A Genetic algorithm (GA) [3.2] is a population-based technique that mimics the process of natural evolution. This method was developed by J. Holland in the 1970s in an attempt to rigorously and systematically describe the adaptive processes of natural systems [3.11]. Not long after that, artificial systems retaining these important natural mechanisms were designed and, in the 1980s, GAs had already been applied to optimization and machine learning. The reason for that was the adaptability of the method, which provided a good learning mechanism for research in Artificial Intelligence, as well as a good search mechanism tailored to the landscape of the search space. This key feature of GAs is what makes them a fitting optimization technique for rough, multimodal, and large search spaces, or even for search spaces not well understood [3.12].

GAs were originally associated only with combinatorial optimization, the reason for that being mainly historical. The systematic use of binary encodings[19] by Holland and his research group set an example that was followed by the majority of the researchers in the years to come. However, GAs were extended to continuous optimization. In fact, empirical comparisons between binary encodings and continuous encodings have shown better results for the continuous encodings [3.12].

Like any other population-based technique, a GA starts the search from a randomly generated initial population of solutions (*chromosomes*) and iteratively replaces the current population with a new one (the population of the next *generation*) according to the performance of each individual. The performance of each individual is evaluated by its objective value (or *fitness value*) indicating the probability of an individual to be selected for reproduction in a way similar to the natural selection, where the genome of fittest individuals is more likely to survive to the next generation through reproduction. Thus, individuals with higher fitness function values have a higher probability of been selected and, consequently, are expected to breed more offspring. However, since the fitness value is only indicative, even a highly evaluated individual may not be selected at all or, on the contrary, it may be selected more times than once.

The evolution of the population is described by the following iterative process. Two parents are selected according to their fitness and to the *selection function* employed by the algorithm. Then, two genetic operators are sequentially applied under a specific probability, namely the *crossover* operator that exchanges portions of the parent solution vectors and the *mutation* operator that randomly modifies an individual.

Finally, the offsprings generated are evaluated and returned back to the population. The evolution process described above is depicted in Fig. 3.5.

[19] The discrete set of feasible solutions of a combinatorial optimization problem can be represented in a binary way.

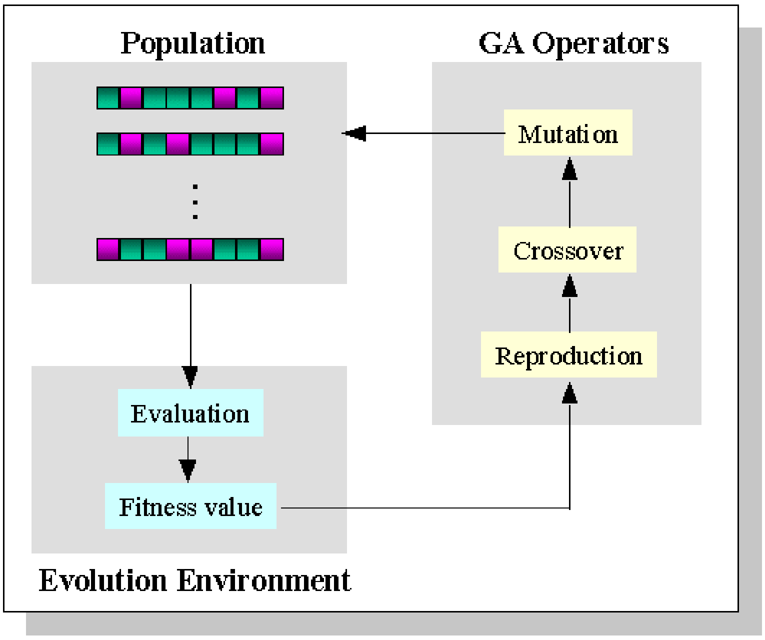


Fig 3.5 The GA evolution process

Selection:

Selection is the process of choosing the individuals that will be subjected to the GA operators in the attempt to further evolve the current population. The purpose of selection is, on the one hand, to drive the search towards the best individuals hoping that their offspring will have an even higher fitness value and, on the other hand, to maintain the diversity of the population preserving the explorative nature of the GA.

The most important selection methods found in literature are: *the roulette wheel selection*, *the stochastic universal selection*, and *the tournament selection*.

The "roulette wheel" sampling chooses parents by simulating a roulette wheel. The population total fitness score (F) is represented by a roulette wheel and cyclic slices of the wheel are assigned to the member of the population. The area of each cyclic slice is proportional to the fitness of the corresponding individual. This fitness wheel is spun as many times as the size of the population (N) and after every spin the individual of the point the wheel stops is selected. Though this stochastic method, in general returns the statistically expected results, it could perform poorly in case of a small population. A series of unfortunate spins could allocate all offspring to the worse individual. For this reason the stochastic universal sampling has been introduced.

The stochastic universal selection method lays out a line where each parent corresponds to a section of the line of length proportional to its fitness value. A pointer moves along the line in steps of equal size, (typically equal to the mean fitness). At each step, the pointer lands on a section selecting the corresponding parent of the section. The first step (r) of the pointer is a uniform random number less than the step size. An example of the stochastic universal selection is depicted in Fig. 3.6.

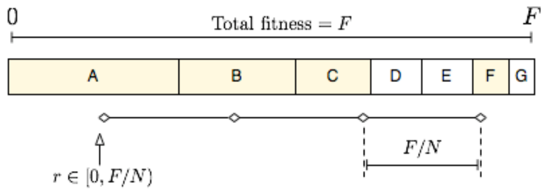


Fig. 3.6 "Stochastic Universal sampling" example

The N repeated random samplings used by the roulette wheel selection are replaced by a single random sampling (i.e. the first step (r)). Thus, the negative effect of the random samplings on the selection process is nullified.

Tournament selection selects the individuals that will be subjected to the GA operators by choosing two Tournament players at random from the population and finding the best individual out of them. Then, a random number between 0 and 1 is chosen and if it exceeds a certain parameter (e.g. 0.8), the best individual is selected to be evolved by the GA operators; otherwise, the worse individual is selected. The two Tournament players are then returned to the original population and the Tournament selection is applied again until selecting the desirable number of individuals. Unlike roulette wheel and stochastic universal selection, tournament selection does not have to pass through the population at each generation to compute the population total fitness score (F). Tournament selection passes through the population once holding as many tournaments as the population size. Hence, it is computationally more efficient and more amenable to parallel implementation.

Finally, *Elitism* is an important supplementary selection method that preserves a number of the best individuals, which otherwise could be lost, if not selected or if destroyed by crossover and mutation. The improvement to the GA's performance by the use of elitism has been verified by numerous researches over the years [3.12].

Crossover:

Crossover is a genetic recombination operator responsible for the reproduction of a part of the population. The percentage of the population that will be subjected to crossover is determined by the *crossover rate.* During crossover, portions of the chromosomes are exchanged between the individuals in order to explore new points of the search space. Because of the diversity of the population in the early stages of the process, the application of the crossover operator causes a significant change in the population allowing the exploration of new regions of the search space. This, however, will settle down in future generations enabling the convergence of the algorithm.

The primary advantage of the GAs comes from the crossover operator[20]. According to Holland [3.5] the respective three explanations for that are:

1. Crossover provides long (random) jumps in the space of possibilities, thus providing a way off of local minima.
2. Crossover repairs mutational damage by sequestering deleterious mutations in some offspring while leaving other offspring free of them.
3. Crossover recombines building blocks.

The most commonly used methods for population crossover are *the single point crossover, the two-point crossover, the uniform crossover.* In single point crossover, a crossover point is chosen at random in the solution vector-matrix. Offspring are then created by exchanging the vector elements after the crossover point between the parents. In case of binary encoding, the offsprings are created by exchanging the bits after the crossover point between the parents. To illustrate the above procedure the single point crossover of binary encoding is depicted in Fig 3.7.

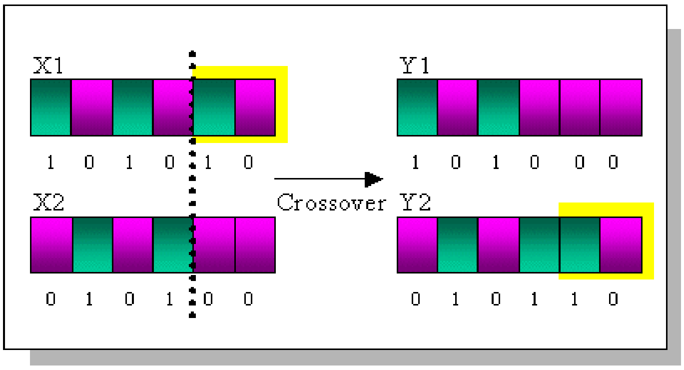


Fig. 3.7 Single point Crossover

In two-point crossover, the ends of each chromosome are joined together to form one ring. Two crossover points are then chosen at random, splitting the ring into two segments. Offspring are then created by exchanging the respective segments between the parents. The same approach has also been extended to multi point crossovers where the ring shaped chromosomes are cut at more than two points creating more segments that are exchanged between the parents.

Finally, the uniform crossover uses a randomly generated crossover mask, which designates the vector elements that will be exchanged between the parents. In case of binary encoding the crossover mask designates the bits that will be exchanged between the parents. To illustrate this procedure the uniform crossover of binary encoding is depicted in Fig.3.8. In this figure the bits of the crossover mask equal to 1 designate the bits that will be exchanged. The respective bits (the 1st and the 5th) are then exchanged between the parents X_1 and X_2 to create the new individuals Y_1 and Y_2.

[20] The paramount importance of the crossover is also apparent in natural systems, where it is known that the crossover rate for mammals is 6 order of magnitude greater than the mutation rate.

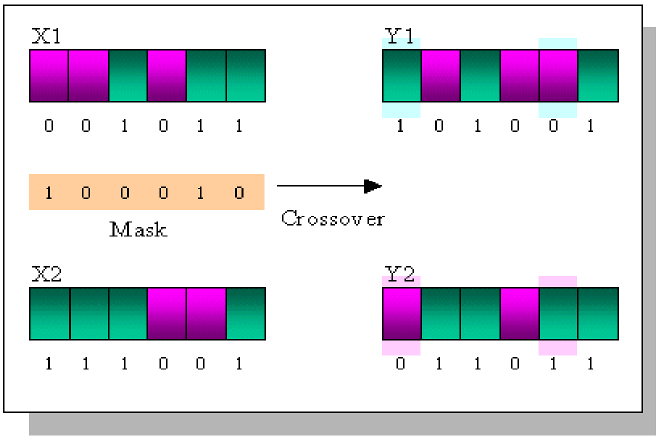


Fig 3.8 Uniform Crossover

Mutation:

Mutation is the modification of a randomly selected vector element. Each element of a vector can be subjected to mutation under a small probability, namely the *mutation rate*. The value of the elements subjected to mutation will be replaced by a random number uniformly selected from the range of the population. In case of binary encoding, the mutated bit will be reversed. The binary mutation is depicted in Fig.3.9.

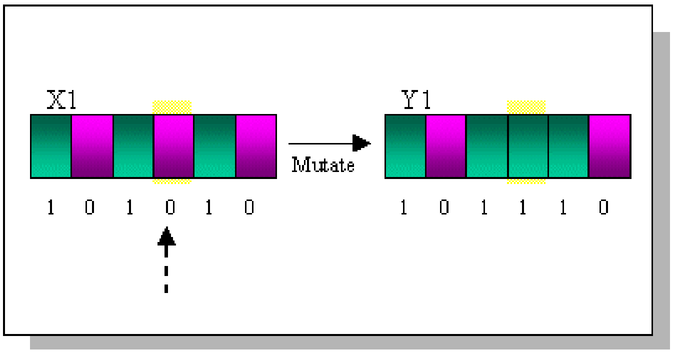


Fig 3.9 Mutation

According to the original GA theory introduced by Holland, the mutation operator plays only a background role, behind the main instrument of variation which is the crossover operator. In the original approach, the mutation operator was nothing more than an evasion mechanism of local optima or of any other locus that the algorithm could get stuck in. But actually, mutation is the only operator which is responsible for the introduction of new genetic material into the population, especially in the case of continuous optimization where the mutated children may develop variable values that do not exist anywhere in the population up to the application of mutation. Therefore, the role of mutation in solving complex continuous problems is crucial.

3.3.5 Statistical analysis of Metaheuristics

Most of the metaheuristics in general and all the metahuristics discussed above belong to the class of nondeterministic (or stochastic) algorithms. As a result, these algorithms have to be averaged over a number of trials to assure statistical confidence. Moreover, a number of measures such as the mean, minimum, maximum, standard deviation may also be of interest in the statistical analysis. The minimum number of trials that must be carried out to produce reliable statistical results is 10 [3.2]. In this course, all results presented herein are averaged over 15 independent runs as the high computational cost of the process (involving the comparison of many metaheuristics, performing an extremely high number of fitness evaluations, to obtain high quality solutions) does not allow for a larger number of trials.

Moreover, the comparison between two sets of metaheuristic independent trials imposes the use of a statistical hypothesis test for assessing whether one of the two sample distributions is stochastically greater[21] than the other. In case the data in the two sets are independent samples from identical continuous distributions with equal medians (null hypothesis) no conclusion regarding the stochastic dominance of the sample distributions can be drawn and the sets cannot be compared. Alternatively if the data are independent samples from distributions that do not have equal medians (alternative hypothesis) one of the two sample distributions is stochastically greater than the other and the sets of metaheuristics can be compared.

Therefore in order to compare the stochastic results of metaheuristics the null hypothesis has to be rejected and for that reason the Wilcoxon rank sum test is performed [3.13]. The Wilcoxon test indicates a rejection of the null hypothesis with a certain confidence level and as a result if the null hypothesis is rejected with a high confidence level (>0.95) the comparison of the results is coherent, whereas if that is not the case the results cannot be compared and the simulations have to be repeated. The statistical confidence in all comparisons done in framework of the present work is verified by performing the Wilcoxon test with standard confidence level 0.95.

[21] The term stochastically greater refers to the concept of stochastic dominance. Stochastic dominance is a form of stochastic ordering and is used in decision theory in situations where a probability distribution over possible outcomes can be ranked as superior to another.

References:

[3.1] Stephen Boyd. *Convex Optimization.* Cambridge University Press. 2004.

[3.2] El-Ghazali Talbi. *Metaheuristics from design to implementation.* John Wiley & Sons, Inc. 2009.

[3.3] Paul Schrimpf. *Matlab – Optimization and Intergration.* http://www.web.mit.edu 2009.

[3.4] Wolpert, D.H., Macready, W.G. *"No Free Lunch Theorem for Optimization"*, IEEE Transactions on Evolutionary Computation. 1997.

[3.5] J. H. Holland. *Building Blocks, Cohort Genetic Algorithms and Hyperplane – Defined Functions*. The University of Michigan, Ann Arbor, MI 48109

[3.6] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. "*Optimization by simulated annealing".* Science, 220(4598):671-680. 1983.

[3.7] V. Cerny. "*A thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm"*. Journal of Optimization Theory and Applications, 45:41-51. 1985.

[3.8] J. Kennedy and R.C. Eberhart. "*Particle swarm optimization"*. In IEEE International Conference of Neural Networks, Perth Australia, pp. 1942-1948. 1995.

[3.9] J. Kennedy and R.Eberhart. "*Swarm Intelligence"*. Morgan Kaufmann Publishers, Inc. San Francisco,CA, 2001.

[3.10] R.M. Storn, and K.V. Prince. *Differential Evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces.* Technical Report TR-95-012, Int CS Institute, University of California, Mar 1995.

[3.11] J.H. Holland. *Adaptation in Natural and Artificial Systems.* The University of Michigan press, Ann Arbor, MI. 1975.

[3.12] Melanie Mitchell. *An introduction to Genetic Algorithms.* The MIT press. 1998.

[3.13] F. Wilcoxon,"Individual comparisons by ranking methods," Biometrics Bulletin Vol. 1, No.6, 1945, pp. 80–83.

Chapter 4
Simulation Results

4.1 Simulator

Following the problem formulation of section 2.4, the described multi-beam satellite system has been simulated and the proposed metaheuristics have been applied in an attempt to determine an optimization technique appropriate for the power allocation problem.

In this course a 37-beam GEO satellite system described in Fig.4.1 has been simulated. The satellite serves fixed users via the multiple beams. Each user position within a beam corresponds to an off-axis angle θ with respect to the boresight (i.e. $\theta = 0°$) and the antenna radiation pattern is approximated employing the Bessel function. In particular, for the satellite antenna beam gain the following expression is employed:

$$G(\theta) = G_{max}\left(\frac{J_1(u)}{2u} + 36\frac{J_3(u)}{u^3}\right)^2 \tag{4.1}$$

where, $u = 2.07123\sin\theta/\sin\theta_{3db}$, and J_1, J_3, are the Bessel functions of the first kind, of order one and three respectively [4.1]. The link budget is calculated assuming one user per beam located at a random position of the beam edge (worst case position). The entries and outputs of the link budget are given in Table 4.1.

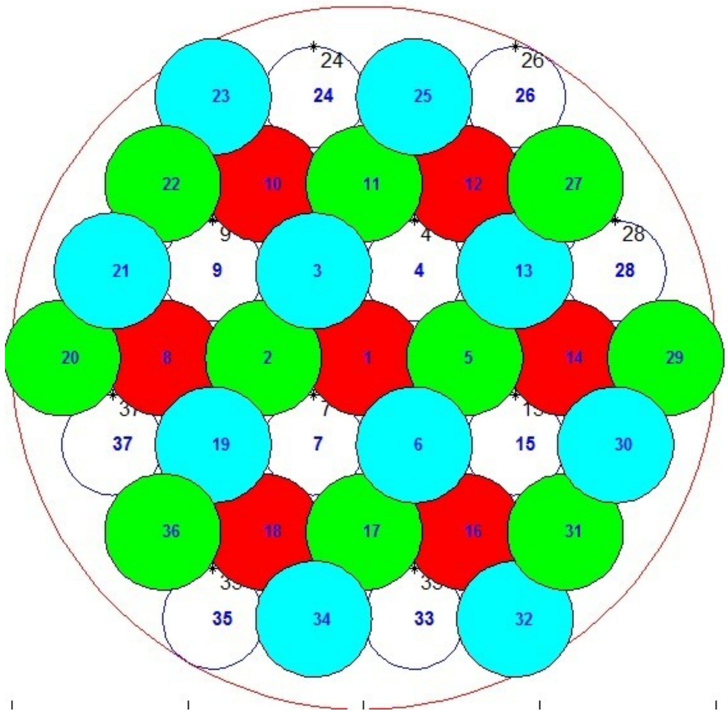


Fig. 4.1 Multibeam satellite layout and antenna gain with respect to the elevation angle for 2 beam sizes

The total downlink bandwidth available is B_{TOT} = 3000MHz, the total available system power is P_{TOT} = 2350W and the beam power constraint is $P_{b,con}$ = 100W.

Table 1 Link Budget Parameters

Link Budget Parameters For Capacity Results	
Parameter	Value
Frequency Band	Ku
User Link Bandwidth $\mathbf{B_u}$	46.875 MHz
HPA saturation Power $\mathbf{P_\tau}$	80 W
Max satellite antenna gain $\mathbf{G_T}$	52 dBi
Output Back Off **OBO**	5 dB
Satellite **EIRP**	66 dBW
Free Space Loss **L**	212 dB
Terminal Antenna Gain $\mathbf{G_R}$	41.7 dBW
Terminal noise Temperature **T**	207 K
Receive **C/N**	20.2 dB
External $\mathbf{(C/I)_{EXT}}$ (Inter-satellite)	30.0 dB

The standard metaheuristic techniques described in Chapter 3 are compared with regard to their performance in maximizing the system capacity for a specific traffic demand. In this course, an appropriate objective function, out of the equations 2.1 up to 2.4, needs to be selected to quantify the satisfaction of the traffic demand. Thus, the best performing technique as well as the best performing objective function need to be identified.

In this course, the first objective function proposed, namely the differential system capacity given by (2.1), will be employed for the performance study of the various metaheuristics and based on the results of this study the best performing metaheuristic technique will then be adopted for the identification of the best objective function out of the Eq. 2.1-2.4.

The performance study of the various metaheuristics will then be repeated, employing the best objective function for a different traffic demand scenario, to ensure that the performance ranking of the various metaheuristics is independent of the problem instance and apply to different search spaces, namely different traffic demands.

Hence, the optimization problem for the initial performance study is the following:

minimize f(x)

$$f(x) = \sum_{b=1}^{37} \mathrm{abs}\{R_{b,req} - R_{b,off}\}, \quad (4.2)$$

subject to:

$$\sum_{b=1}^{37} P_b \leq 2350\mathrm{W} \quad (4.3)$$

$$P_b \leq 100\mathrm{W}, \quad b = 1, 2, \ldots, 37 \quad (4.4)$$

where $R_{b,req}$ denotes the bit rate requested by the user in beam b and $R_{b,off}$ denotes the bit rate offered by beam b. The optimization variable of the problem is the vector x = (P_1, …, P_{37}) where each element denotes the transmit power of the respective beam. The requested bit rate $R_{b,req}$ is shown in Fig.4.2, along with the bitrate provided by the uniform power allocation $R_{b,unif}$. In uniform power allocation the total available system power P_{TOT} is equally distributed among the beams (i.e. $P_b = 2350W/37 = 63.5W$). Uniform power allocation is done by non-flexible satellite systems that transmit the same EIRP over each beam. Therefore, the performance of the uniform power allocation option is used for performance comparison between the proposed optimized power allocation scheme and that of typical non-flexible satellite power allocation.

To drive the optimization towards improving the performance of non-flexible satellite systems, the uniform power allocation is input as an initial solution to the optimization problem, i.e. as a member of the initial population of solutions for all the population-based techniques explored in this Chapter. Moreover, all results reported in this chapter are obtained by averaging over 15 independent runs whereas statistical confidence in the comparisons is assured performing the Wilcoxon test with standard confidence level 0.95, as suggested by the statistical metaheuristics analysis of Chapter 3.

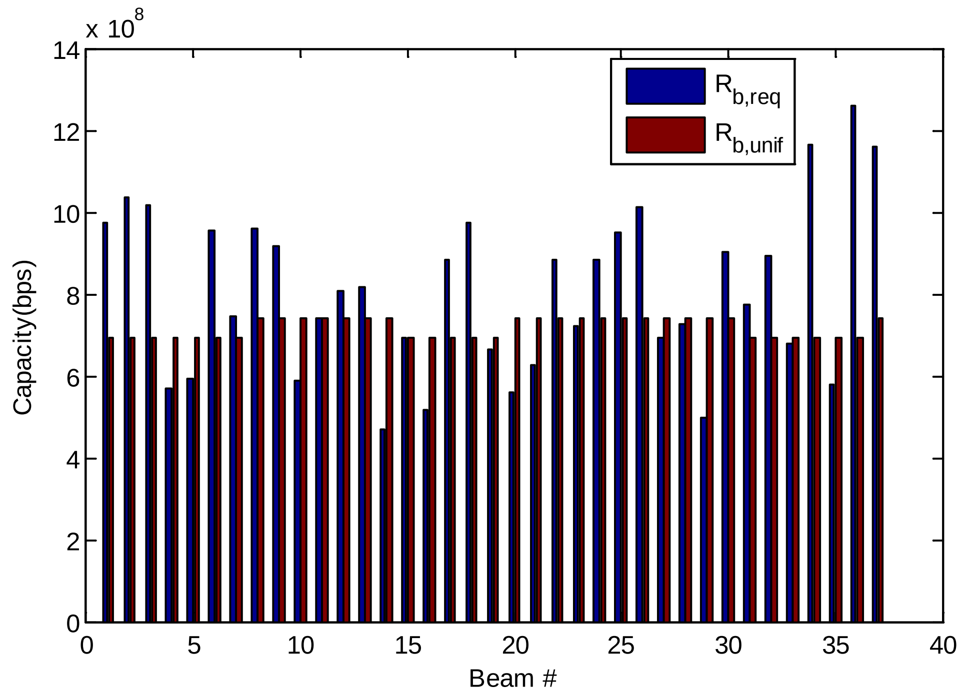


Fig. 4.2 Bitrate requested by users in a beam ($R_{b,req}$) under scenario 1 vs the bitrate provided by non-flexible satellite systems ($R_{b,unif}$)

To conduct the performance assessment of the compared metaheuristics the quality of their solutions is compared with regard to their associated computational effort. Thus, in order to compare results their associated computational cost, expressed by the times the objective function is evaluated, must be the same. In this course, all techniques explored herein perform the same number of objective function (fitness) evaluations and this number is determined by the tuning process.

Following the theoretical approach presented in Chapter 3, the technique expected to yield the best performance is the GA. Therefore, GA is the first technique to be tuned and the results of the tuning process will determine, among other GA parameters, the number of fitness evaluations. Since all optimization techniques must perform the same number of fitness evaluations for their comparison to be fair, the number determined by the GA tuning will also be adopted for the rest of the techniques. Thus, the GA tuning will influence all optimization techniques of this study and therefore the GA tuning process will be thorough.

4.2 GA Tuning

Tuning a GA aims at determining the parameters of the GA as well as the GA functions described in Chapter 3. The GA parameters are the following:

- Population Size
- Number of Generations (Iterations)
- Crossover Rate
- Mutation Rate
- Elite Individuals

The GA functions are the following:

- Selection function
- Crossover function

At the beginning of the tuning process an empirical understanding of the GA parameters provides the initial values. In fact, the significance of the crossover operator suggests that, to enhance the population evolution, a significant portion of the population should be subjected to crossover, i.e. a high crossover rate close to unit must be employed (~0.9). Moreover, an empirical rule concerning mutation sets as initial mutation rate the inverse of the number of the problem decision variables, in this case 1/37 = 0.027.

Having determined the initial values of the crossover and mutation rates, the interdependent parameters of population size, number of generations and elite individuals must be determined via a trial and error approach, i.e. by the convergence plot of different population sizes. The population size providing the best convergence is selected and the number of generations is set equal to the ratio of the number of fitness evaluations required for the convergence of the algorithm over the population size. The elite individuals are subsequently determined as a small percentage of the population size.

The trial and error approach described above will be made clear through the GA tuning of the problem under consideration. After setting the crossover rate equal to 0.9 and the mutation rate equal to 0.027, the population size, number of generations and elite individuals has to be determined. The selection of the appropriate GA functions will ensue, but until then the uniform (scattered) crossover function and the tournament selection function are employed. For the specific entries the convergence of the GA under different population sizes is compared. GAs with population sizes 100-200-400-800-1600-3200-6400-12800 are executed until reaching the algorithm convergence.

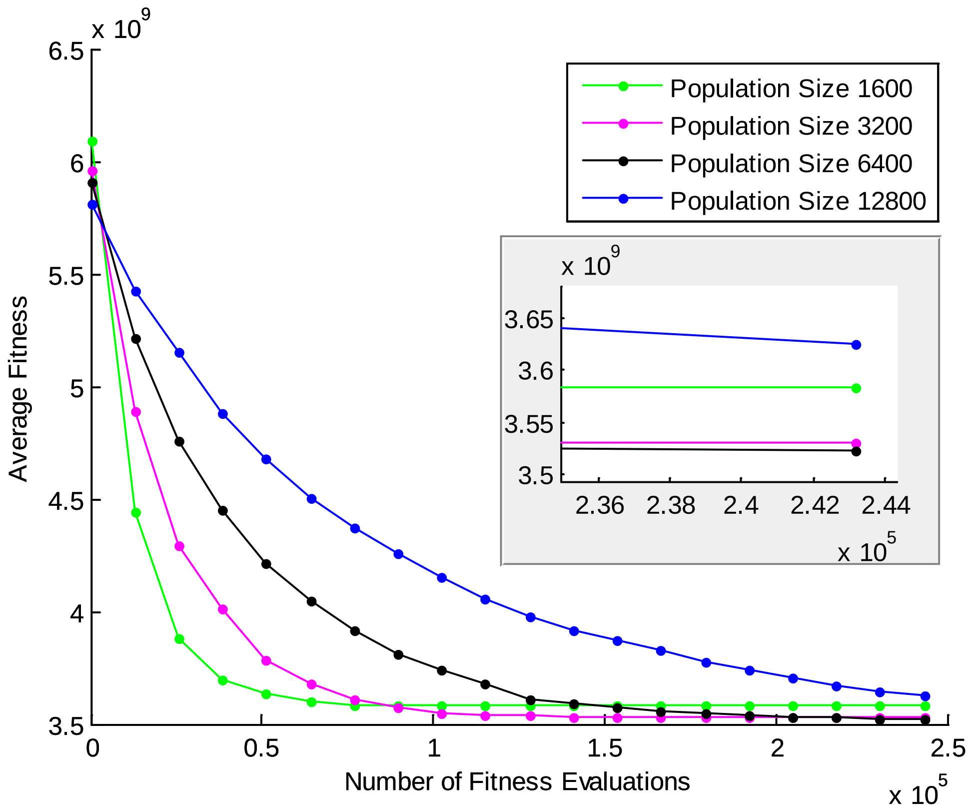


Fig. 4.3 Population size comparison.

Table 4.1 Convergence test with regard to the Population size

Population Size	Average Fitness Value (DSC in Gbps)
1600	3.5826
3200	3.5309
6400	**3.5223**
12800	3.6259

The convergence plot of GAs using 1600-3200-6400-12800 individuals is shown in Fig.4.3. The y-axis of the plot corresponds to the average fitness value (evaluated based on (4.1)), of the best performing individual of the population and the x-axis corresponds to the number of fitness evaluations performed. The increase in the population size results in a steady performance improvement and the four higher - and best performing - population sizes are depicted in Fig.4.3. The values of the final convergence of these four population sizes are shown in Table 4.1.

The population size of 6400 individuals provides the best average fitness value after 243200 fitness evaluations, namely 243200/6400 = 38 Generations. The convergence plot of the GA of 12800 individuals suggests that a further decrease of the average fitness value would be possible for an additional number of fitness evaluations. However, this improvement would have been obtained at an extremely high computational cost, given the population size of 12800 individuals. Hence, the population size selected is 6400 individuals and the required number of generations is 38. The elite individuals used are usually less than ten; however, given the very high population size required to achieve convergence to a better solution a number of 30 elite individuals is assumed.

Following the selection of the GA parameters, the GA functions providing the best convergence must be selected. In this course, the performance of the crossover functions, namely single-point, two-point and uniform (scattered) crossover functions, are compared, at the same time using the tournament selection function and a population size of 6400 and 30 elite individuals. To isolate the impact of the crossover functions considered on the convergence of the algorithm and make the relevant comparison, the mutation operator is neutralized by setting the mutation rate 0; thus, the whole population is subjected to crossover, i.e. crossover rate is equal to 1. The results of this comparison are shown in Fig.4.4 and the numerical results in Table 4.2.

The crossover function providing the smaller average fitness value is the scattered crossover. The results of this comparison demonstrate also the importance of the mutation operator for the convergence of the algorithm. In particular, the GA employing the scattered crossover function of Fig.4.4 stagnates after 60000 fitness evaluations providing an average fitness value of 4.9383 Gbps, due to the neutralization of the mutation operator. On the contrary, the GA of 6400 individuals of Fig.4.4, also employing the scattered crossover function, evolves for 234200 fitness evaluations providing an average fitness value of 3.5223 Gbps. Thus, the performance of the algorithm, employing the exact same GA parameters and functions, decreased by 40% in the absence of mutation!

Finally, the selection function that will be employed in the simulations is the tournament selection function. That is since the very high population size selected necessitates the employment of the computationally efficient selection method, amenable to parallel implementation [4.2].

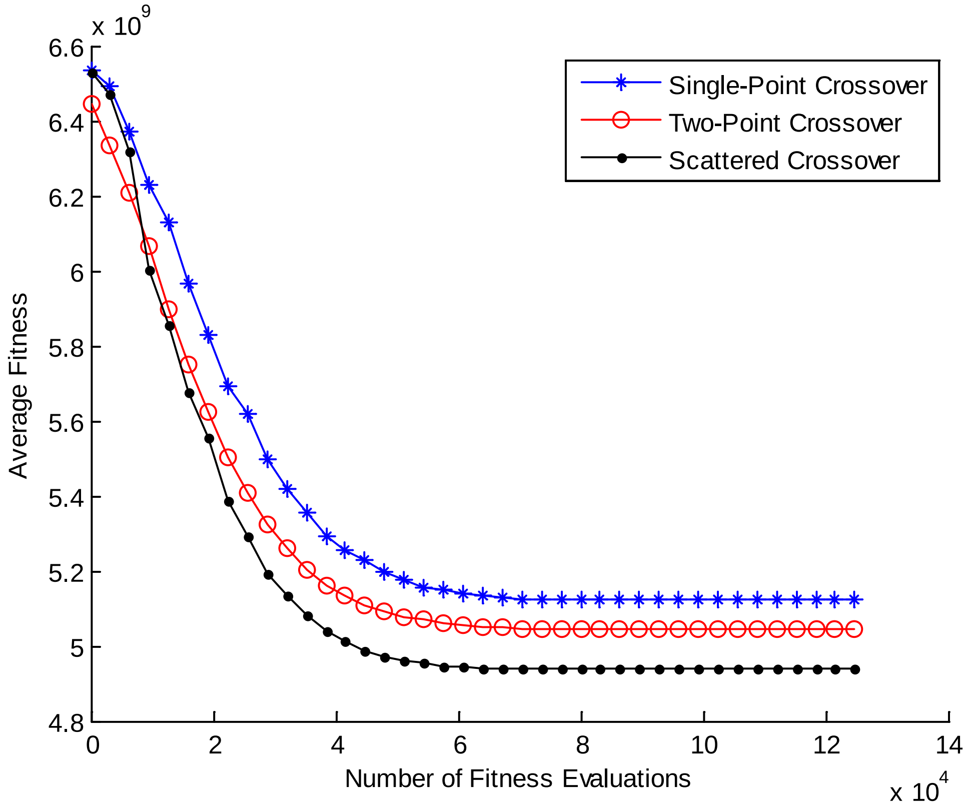


Fig. 4.4 Crossover Function Comparison

Table 4.2 Crossover Function Numerical Results

Crossover Function	Average Fitness Value (DSC in Gbps)
Single-Point	5.124
Two-Point	5.043
Scattered	**4.9383**

Concluding, the parameters and functions selected after the tuning process are given in Table 4.3.

Table 4.3 GA parameters and functions

Parameter / Function	Value / Selection
Crossover rate	0.9
Mutation Rate	0.027
Fitness evaluations	243200
Population Size	6400
Generations	38
Elite individuals	30
Crossover function	Uniform
Selection function	Tournament

The above entities provide an early approach to the GA tuning. In this approach the crossover rate, mutation rate and number of elite individuals were determined according to empirical understandings but each of these parameters was determined individually without reference to the rest of the parameters. However, the royal road experiments [4.2] have demonstrated that it is not the choice between crossover or mutation - through the selection of the crossover rate and mutation rate - that is important but the balance among these operators and their complementary evolution of the population. Thus, the most promising prospect is the use of a GA adapting its own mutation and crossover rates during a search.

In this course, it was noted that a GA could be used to optimize the parameters of another GA![4.2] Following this approach a meta-level GA will be used to optimize the parameters of the GA, providing the final GA tuning. This meta-level GA evolved a population of 24 GA parameter sets for 30 generations, employing the uniform crossover function and the tournament selection function. Each individual encoded four of the GA parameters: Crossover Rate, Mutation Rate, Population Size, Elite individuals and following the results of the early GA tuning each individual was evolved for 250000 fitness evaluations to evaluate its fitness. The range of the parameter values used in the meta-GA is: Crossover Rate [0:1], Mutation Rate [0:0.15], Population Size [64:8192] and Elite individuals [0:0.1]*Population Size.

The parameters and functions selected after the final tuning are given in Table 4.4.

Table 4.4 GA parameters and functions

Parameter / Function	Value / Selection
Crossover rate	0.95
Mutation Rate	0.05
Fitness evaluations	234240
Population Size	5856
Generations	40
Elite individuals	30
Crossover function	Uniform
Selection function	Tournament

And the performance of the GA using these parameters is shown in the performance comparison of the different techniques.

At this point should be noted that the computational cost of the meta-level GA was extremely high since 24(GAs/Generation)*30(Generations)*250.000(Fitness evaluations/GA) = 180.000.000 fitness evaluations had to be performed. This did not allow for the averaging of the used GAs over at least 10 independent runs and as a result the relation between the optimization variables (i.e. the GA parameter sets) and their fitness value (DSC) is not deterministic. Notwithstanding this drawback the results provided by the meta-level GA provided a significant improvement of the GA performance.

4.3 SA Tuning

The tuning of the SA aims at determining the following parameters and functions of SA (see Chapter 3).

- Initial Temperature
- Re-annealing Interval
- Annealing Function
- Temperature Update Function

For a fair comparison of different techniques, the same number of fitness evaluations must be performed. Hence, the necessary fitness evaluations are 5856*40 = 234240. Having decided the number of fitness evaluations, the SA functions are selected. In this course, the Initial Temperature (IT) is arbitrarily set equal to 100 and the Re-annealing Interval (RI) equal to 100. The annealing function selected is fast annealing. Keeping the specific set of SA parameters the exponential and logarithmic temperature update functions are compared. The relevant comparison shown in Fig.4.5 and in Table 4.5 indicate that the exponential temperature update must be selected for the SA algorithm.

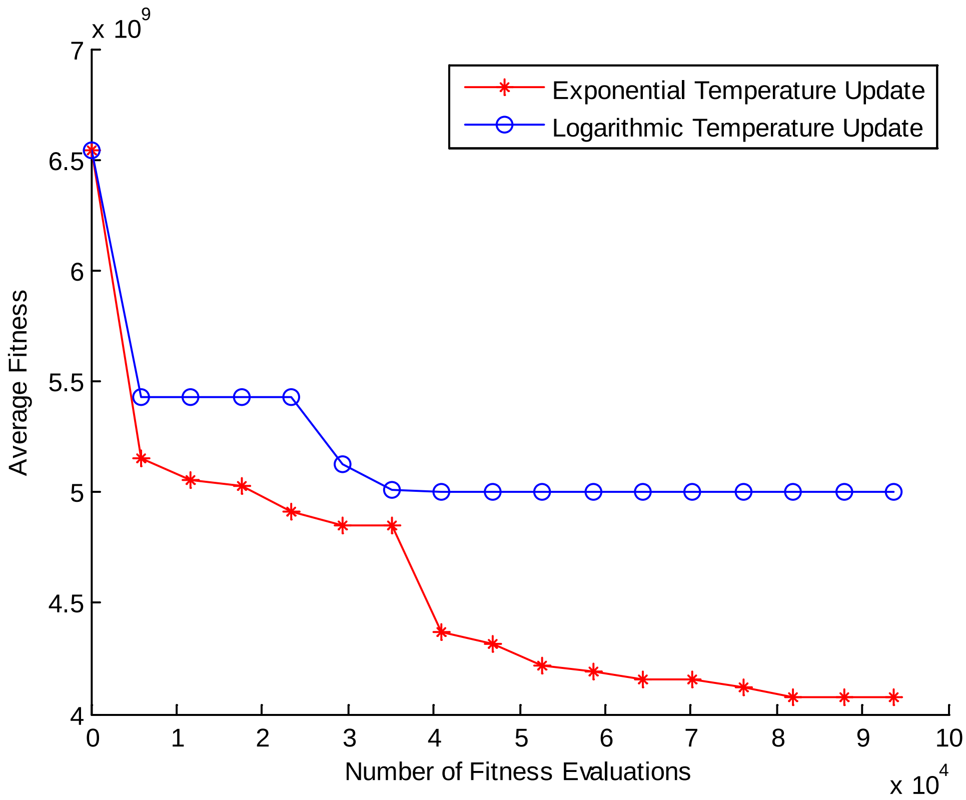


Fig. 4.5 Temperature Update Function Comparison

Table 4.5 Temperature Update Function Comparison

Temperature Update Function	Average Fitness Value (DSC in Gbps)
Exponential Temperature Update	**4.0790**
Logarithmic Temperature Update	4.9973

Following the comparison of the temperature update functions, the performance of fast and Boltzmann annealing is compared employing the exponential temperature update and for IT = 100, RI =100. The results of this comparison are shown in Fig.4.6 and Table 4.6.

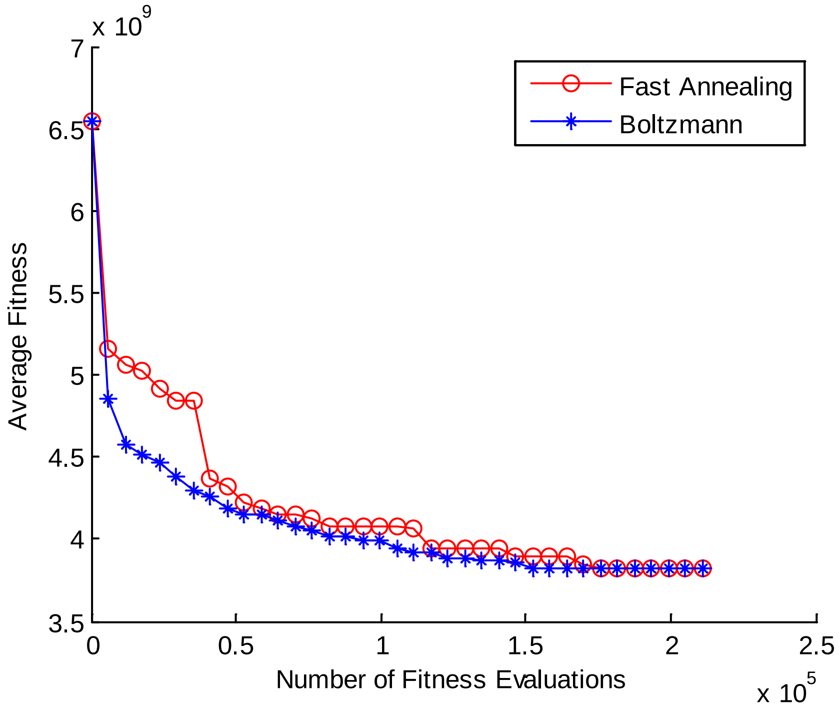


Fig. 4.6 Annealing Function Comparison

Table 4.6 Comparison between Fast and Boltzmann annealing

Annealing Function	Average Fitness Value (DSC in Gbps)
Fast Annealing	3.821
Boltzmann Annealing	**3.817**

From the convergence plot and the numerical results it is evident that Boltzmann annealing outperforms fast annealing with regard to the speed of convergence and to the average fitness value.

Next the Initial Temperature and Re-annealing Interval are determined having selected the exponential temperature update and the Boltzmann annealing function. For RI = 100 the average fitness under four IT values was determined. The results of this process are shown in Table 4.7.

Table 4.7 Selecting the Initial Temperature value

Initial Temperature	Average Fitness Value (DSC in Gbps)
50	3.753
100	3.948
200	3.707
400	**3.665**

The best performing IT value is 400. Setting IT = 400 the average fitness under five RI values is determined. The relevant results been shown in Table 4.8 wherefrom it is verified that the best performing re-annealing interval is equal to 100. Concluding, the parameters and functions selected after the SA tuning are tabulated in table 4.9.

Table 4.8 Selecting the Re-annealing Interval value

Re-annealing Interval	Average Fitness Value (DSC in Gbps)
100	**3.665**
500	4.061
1000	4.312
5000	4.627
10000	5.200

Table 4.9 SA parameters and functions

Parameter / Function	Value / Selection
Initial Temperature	400
Re-annealing Interval	100
Temperature Update Function	Exponential Update
Annealing Function	Boltzmann Annealing

The above parameters are used in all SA simulations that follow. However, in the performance comparison of the standalone techniques, the re-annealing option is not used since it does not belong to the standard SA algorithms.

4.4 DE Tuning

The DE parameters that need to be defined according to the DE theory presented in Chapter 3, are the following:

- Crossover Rate (CR)
- Scaling Factor (F)
- Population Size

The tuning of a DE algorithm however is very easy and is performed according to the guidelines available in the literature for the design of DE algorithms. The population size is set at 10 times the number of decision variables of the problem, namely 10*37 = 370 for the problem in hand. The scaling factor is initialized at F = 0.8 and the crossover constant at CR = 0.9. To refine the results yielded by these values, the population size can be changed and the scaling factor F can be slightly adjusted around 0.8. The parameter adjusted is the scaling factor F and not the crossover CR since DE is much more sensitive to F rather than to CR [4.3]. Following this rational, the performance of DE is examined for several parameter combinations. The relevant results been tabulated in Table 4.10 wherefrom it is deduced that the parameter set yielding the best performance is that of Table 4.11.

Table 4.10 DE Parameters Comparison

Parameter Set	Average Fitness Value (DSC in Gbps)
CR=0.9, F=0.8, Population Size=370	3.601
CR=0.9, F=0.8, Population Size=150	3.601
CR=0.9, F=0.8, Population Size=100	3.616
CR=0.9, F=0.8, Population Size=40	**3.589**
CR=0.9, F=0.9, Population Size=40	3.604

Table 4.11 DE parameters

Parameter	Value
Crossover	0.9
Scaling Factor	0.8
Population Size	40
Iterations	5856

The above parameters are used for all the DE simulations performed that follow. The foregoing numerical results, however, demonstrate that, as long as the general guidelines of DE tuning concerning CR and F are followed the performance of the algorithm does not vary significantly; changes in the population size or adjustments of the scaling factor provide only a refinement of the results.

4.5 PSO Tuning

According to the theory of Chapter 3 for a fixed number of fitness evaluations the PSO tuning includes the setting of the following parameters:

- Cognitive Factor C_1
- Social Factor C_2
- Population Size
- Inertia w

To select an appropriate population size the following arbitrary values are assigned to the other parameters: C_1=0.25, C_2=0.85 and w=1. The performance comparison of various population sizes for the arbitrarily selected parameters is tabulated in Table 4.12.

Table 4.12 PSO Population Size Comparison

Population Size	Average Fitness Value (DSC in Gbps)
11712	4.709
5856	4.620
370	**4.410**
40	5.058

Thus, the best performing population size is 370. Next the inertia w is defined as a linear function of the iteration number (k) so that the inertia is reduced as the optimization evolves driving the algorithm from performing a wide exploratory search to been confined to a local search for the last generations allowing the convergence of the algorithm. Thus, the inertia function is formulated as [4.4]:

$$w = \frac{(k_{max} - k) * (w_{start} - w_{end})}{k_{max}} + w_{end} \tag{4.5}$$

where k_{max} is the total number of iterations of the algorithm, k is the number of the current iteration, w_{start} the inertia at the beginning of the algorithm and w_{end} the inertia at the end of the algorithm. According to [4.4] w_{start}=0.9 and w_{end}=0.4.

According to performance studies found in literature the cognitive and social factors are both set equal to 2. Concluding, the parameters selected after the PSO tuning and applied to all PSO simulations that follow are tabulated in Table 4.13.

Table 4.13 PSO parameters

Parameter / Function	Value / Selection
Cognitive Factor	2
Social Factor	2
Population Size	370
Inertia	$w = \frac{(633 - k) * (0.5)}{633} + 0.4$

4.6 Metaheuristics Performance Study

After the tuning of the algorithms, the selected parameters tabulated in Tables 4.4, 4.9, 4.11, 4.13 are adopted for the performance comparison of the standard techniques and are summarized in Table 4.14.

Table 4.14 Optimization Parameters

<table>
<tr><td rowspan="2">GA</td><td>Crossover Rate: 0.95</td><td>Elite Individuals: 30</td><td>Population Size: 5856 Generations:40</td><td rowspan="2">Fitness Evaluations: 234240</td></tr>
<tr><td>Mutation Rate:0.05</td><td>Uniform Crossover Function</td><td>Tournament Selection Function</td></tr>
<tr><td>SA</td><td>Initial Temperature: 400</td><td>Exponential Temperature Update</td><td>Boltzmann Annealing</td><td>Fitness Evaluations: 234240</td></tr>
<tr><td>DE</td><td>Crossover: 0.9</td><td>Scaling Factor: 0.8</td><td>Population Size: 40 Generations: 5856</td><td>Fitness Evaluations: 234240</td></tr>
<tr><td>PSO</td><td>Cognitive Factor C_1: 2</td><td>Social Factor C_2: 2</td><td>Population Size: 370 Generations: 633</td><td>Fitness Evaluations: 234240</td></tr>
</table>

The convergence plot of the standard techniques for the traffic demand scenario of Fig.4.2 is shown in Fig.4.7 and the numerical results concerning the performance of the metaheuristic solutions are shown in Table 4.15. The performance of the solutions is evaluated by the DSC and Table 4.15 tabulates the average fitness value of the results obtained by the 15 independent runs of each metaheuristic. Apart from their average fitness value the standard deviation of their fitness value is also tabulated since the deviation tendencies of the metaheuristic results must be considered. In particular, techniques providing consistency in their results are favored over techniques exhibiting high deviation tendencies and unforeseen behavior. Finally the overall best DSC over the 15 independent runs is also provided, whereas for the uniform power allocation no statistical analysis is required and the fitness value (DSC) of the uniform power allocation is tabulated.

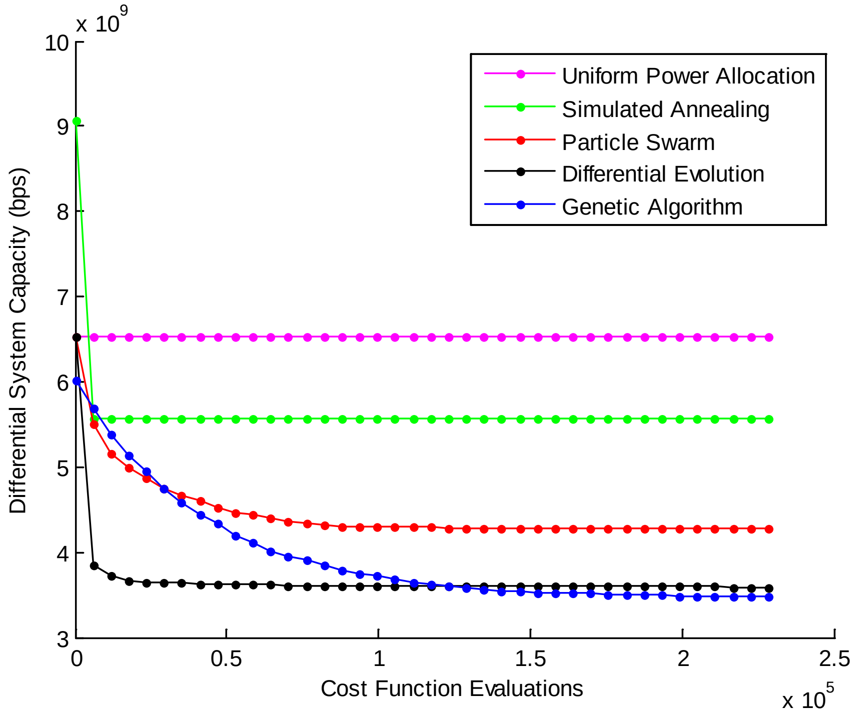


Fig. 4.7 Convergence plot of metaheuristic standard techniques (scenario 1)

Table.4.15- Numerical Results (DSC in Gbps)

Algorithm	Fitness Value (Average of 15 runs)	Standard Deviation (of 15 runs)	Overall Best (over 15 runs)
Simulated Annealing	5.5634	0.35583	4.8916
Particle Swarm	4.2843	0.36050	3.8605
Differential Evolution	3.5854	0.03320	3.5480
Genetic Algorithm	**3.4773**	**0.01679**	**3.4530**
Uniform Power Allocation	Fitness Value		
	6.5388e+009		

According to the above results, the adoption of a GA provides an improvement of 46.8% in the cost function over the uniform allocation case. The improvement provided by the GA over the uniform power allocation is evident in Fig.4.8, where the beam capacity offered by a system transmitting the power suggested by the GA optimization (R_{GA}) is compared to the beam capacity offered by a non-flexible system ($R_{b,unif}$) and to the demanded capacity ($R_{b,req}$).

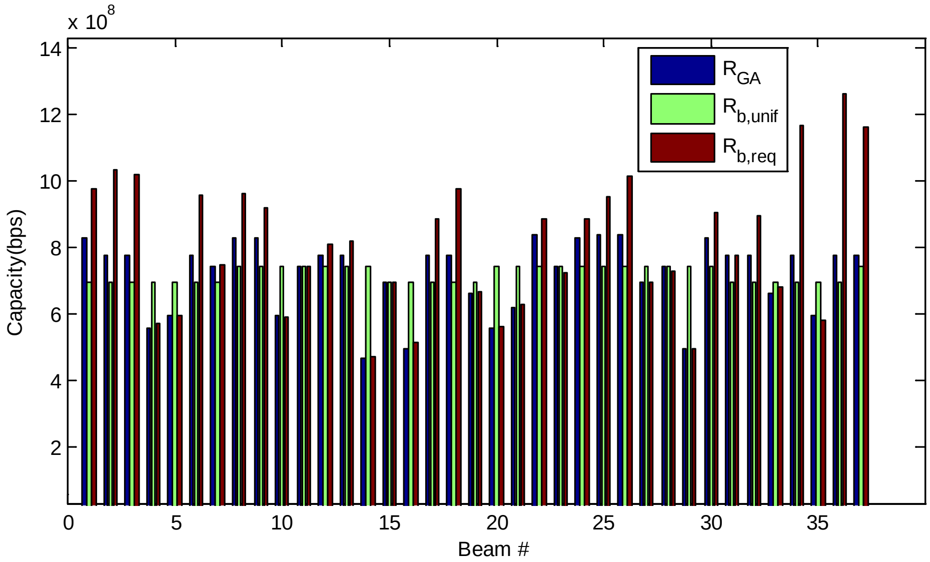


Fig. 4.8 Traffic Demand Satisfaction
of GA-optimized and Uniform power allocation

The convergence plot of Fig.4.7 demonstrates the convergence speed of the various metaheuristics. DE, which is one of the most successful techniques for continuous optimization problems, provided the best speed of convergence and a final result close to the GA-provided result. DE is therefore the best option among the ones compared for systems where a good solution (i.e. a power allocation satisfying the traffic demand) must be reached in minimum time.

The best result, however, is acquired applying the GA. GA is the slowest converging technique but returns the best results with a smaller standard deviation, thus a higher consistency over the independent runs, than DE. Among the standard techniques found in the literature and implemented in the framework of this thesis, GA and DE exhibit the best behavior, indicating a trade-off between convergence speed and quality of solution.

Another interesting conclusion is that the simulation results are consistent with the intuitive suggestions of Chapter 3. In particular, the trajectory-based technique (Simulated Annealing) is outperformed by the population-based ones, whereas the Evolutionary Algorithms outperform the Particle Swarm Optimization and the Genetic Algorithms outperform the Differential Evolution.

However, to ensure that the performance ranking of the various metaheuristics and the substantial benefits acquired from the use of GA are independent of the problem instance and they also apply to different search spaces, namely different traffic demands, the performance study is also carried out for a different traffic demand. The traffic demand of this scenario is depicted in Fig.4.9.

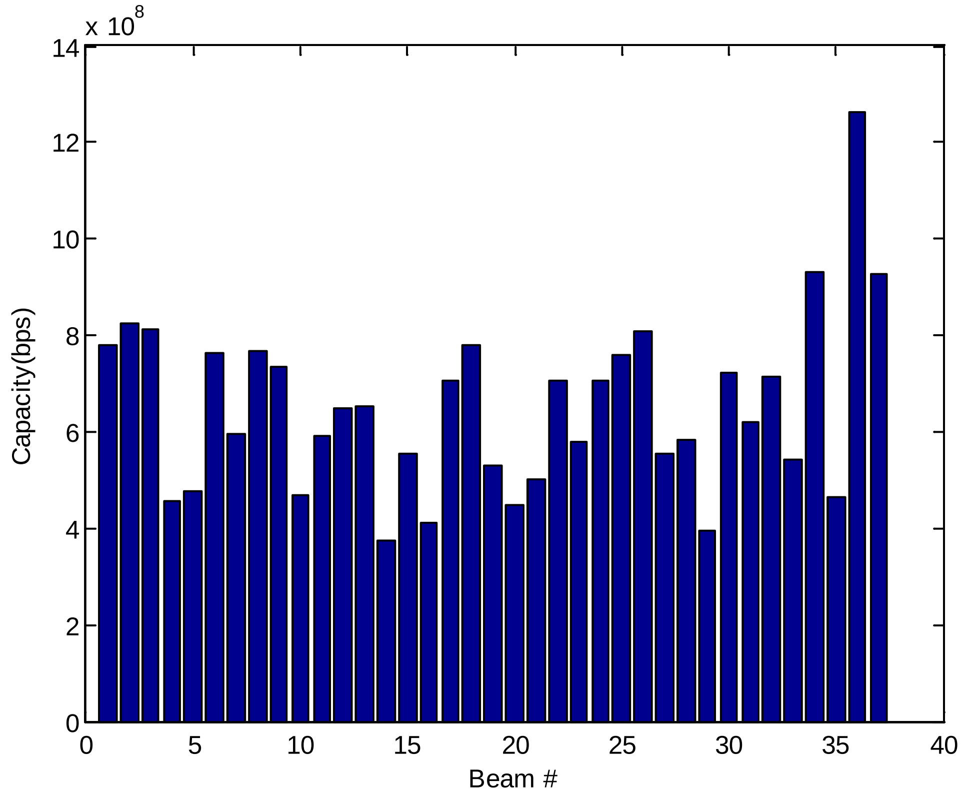


Fig. 4.9 Traffic Demand (scenario 2)

In this course and since a first verification on the suitability of GA for the solution to the problem in hand is provided by the simulations of scenario 1, before the performance study of the metaheuristics following scenario 2, a GA is adopted for the performance assessment of the cost functions presented in Chapter 2 (i.e. Equations 2.1-2.4). The results of this assessment are tabulated in Table 4.16 where the available cost functions (Differential System Capacity (DSC), Unmet System Capacity (USC), Satisfaction Factor (SF) and Aggregate Fitness (AF)) are compared against the unsatisfied capacity under the new traffic demand of scenario 2 (Cumulative Unmet Column).

The numerical results presented in Table 4.16 demonstrate the suitability of the USC as cost function, since USC minimizes the unsatisfied capacity. This result is not surprising since USC focuses on the minimization of the unmet capacity and beam capacities exceeding the required do not contribute to the figure of merit of USC.

Table 4.16 Cost Function Performance Assessment

Cost Function	Capacity [Gbps]		
	(Average of 15 runs)		
	Cumulative Required	Cumulative Offered	Cumulative Unmet
DSC	24.155	23.588	0.801
USC	24.155	25.605	**0.746**
SF	24.155	23.946	0.748
AF	24.155	23.547	0.842
	(Overall Best over 15 runs)		
DSC	24.155	23.593	0.778
USC	24.155	24.879	**0.709**
SF	24.155	24.879	0.709
AF	24.155	23.588	0.802

Hence, hereafter USC is adopted as the best performing cost function and the performance study of the metaheuristics on the new scenario will aim at minimizing the USC

$$\text{minimize } f(x)$$

$$f(x) = \sum_{b=1}^{37} \max\{R_{b,req} - R_{b,off}, 0\} \tag{4.6}$$

for the traffic demand scenario depicted in Fig.4.8. The convergence plot of the new instance is presented in Fig.4.10 and the respective numerical results are tabulated in Table 4.12.

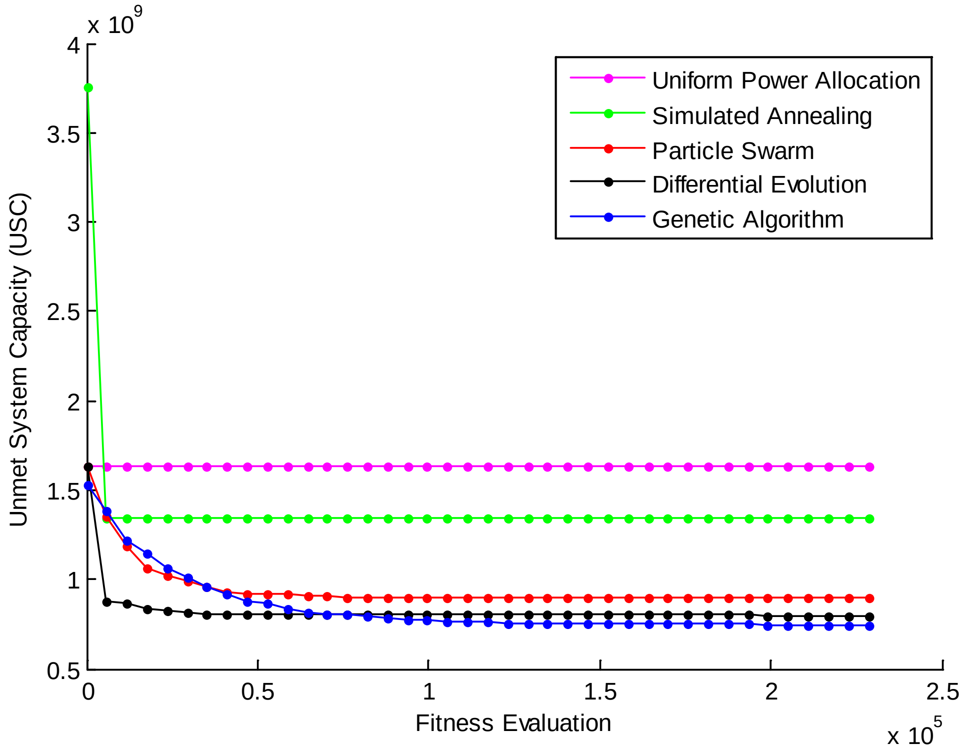


Fig. 4.10 Convergence plot of metaheuristic standard techniques (scenario 2)

Table.4.12- Numerical Results (USC in Gbps)

Algorithm	Fitness Value (Average of 15 runs)	Standard Deviation (of 15 runs)	Overall Best (over 15 runs)
Simulated Annealing	1.3484	0.16253	1.2171
Particle Swarm	0.89785	0.04360	0.80874
Differential Evolution	0.79582	0.04188	0.72260
Genetic Algorithm	0.74579	0.01972	0.70850
Uniform Power Allocation	Fitness Value		
	1.6368		

As in the scenario 1, GA provides a significant improvement over the uniform allocation in the course of cost function minimization (minimizing USC for scenario 2). Moreover, the behavior of the convergence plot remains the same, with the DE providing the best speed of convergence and the GA the best results. Therefore, it is evident that the performance ranking of the metaheuristic techniques is independent of the traffic demand and that the GA is the best performing technique providing a substantial improvement over the performance of non-flexible systems. In particular, the improvement provided by GA over the uniform power allocation for the second scenario is 56.7%.

The improvement provided by the GA over the uniform power allocation under scenario 2 is observable in Fig.4.11, where the beam capacity offered by a system transmitting the power suggested by the GA optimization (R_{GA}) is compared to the beam capacity offered by a non-flexible system ($R_{b,unif}$) and to the demanded capacity ($R_{b,req}$).

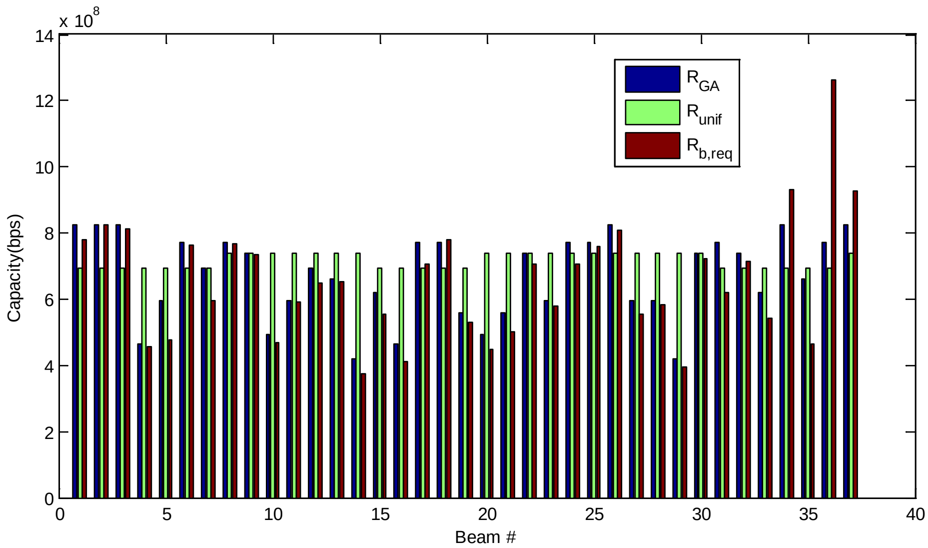


Fig. 4.11 Traffic Demand Satisfaction of GA-optimized and Uniform power allocation (scenario 2)

As the system performance is improved by 46.8% under the first scenario and by 56.7% under the second scenario compared to the performance of the non-flexible system, the previous results demonstrate how satellite multibeam broadcasting can benefit from dynamic power allocation schemes, in general, and from the use of the proposed techniques, in particular. Moreover, the systematic study of the various metaheuristic techniques provides a comparative overview of their performance in the framework of optimizing the power allocation per se and of their performance when various parameters of the problem, such as the quality of the solution, the consistency of the results and the speed of convergence are taken into account. A multibeam satellite system employing DE techniques, for instance, can deal with rapidly changing traffic requirements, a situation where fast convergence is imperative. On the other hand, if the quality of the results is of interest, GA approaches must be followed. Furthermore, GA approaches provide the smaller standard deviation, ensuring the consistency of the obtained results over independent runs, in spite the stochastic performance of metaheuristics.

References:

[4.1] D. Christopoulos, S. Chatzinontas, G. Zeng, J. Grotz, B. Ottersten, *"Linear and non-Linear Techniques for Multibeam Joint Processing in Satellite Communications"*. EURASIP J. on Wirel. Commun. and Networking 2012, 2012:162

[4.2] Melanie Mitchell. *An introduction to Genetic Algorithms.* The MIT press. 1998.

[4.3] El-Ghazali Talbi. *Metaheuristics from design to implementation.* John Wiley & Sons, Inc. 2009.

[4.4] Mathew Settles. *An Introduction to Particle Swarm Optimization.* Department of Computer Science, University of Idaho, November 2005.

Chapter 5
Hybrid Genetic Algorithm Approach

5.1 Hybridization

The simulation results of the previous chapter demonstrate the superiority of the population-based techniques compared over the trajectory based Simulated Annealing. These results are consistent with the intuitive approach of Chapter 3. According to this approach the large size and multimodality of the problem search space imply the use of an explorative technique to solve the problem. However, following this intuitive approach, once the sufficient exploration of the search space is done the exploitation of the solutions provided may enhance them even further.

This proposition relies on the complementary nature of the exploration and exploitation search strategies. In particular, an explorative technique performs a diverse search over the search space in order to discover regions that are promising with regard to encompassing satisfactory solutions. Complementarily to this search an exploitative technique can perform an intensive search within these promising regions, finding locally optimal solutions of higher quality than the ones provided by the initial exploration. It is, therefore, evident that a technique exploiting the benefits of both search strategies might yield a significant improvement in the results obtained.

To formulate such a technique the exploitative nature of trajectory based and the explorative nature of population based techniques may be utilized. The utilization of different techniques in order to combine their key-components and advantages is known as *hybridization* and can be advantageous in several cases. The hybridization of metaheuristics, in particular, has manifold possibilities; a large number of publications document the benefits and successes of such hybrids [5.1]. The documented success of metaheuristic hybridization has also motivated the hybridization of a population-based technique and a trajectory based technique following to the previously mentioned intuition.

In this course the Genetic Algorithm (the best performing out of the population based techniques tested) and the Simulated Annealing can be used complementary to outperform the best performing techniques of the previous chapter and improve system performance even further. This technique is new to the literature and was presented in the 2nd ESA Workshop on Advanced Flexible Telecom Payloads, 17-19 April 2012, Noordwijk, The Netherlands (Appendix 1). Moreover, the proposed technique was employed for the “Operational Optimization of a Ku Multibeam Flexible Payload” by Astrium Satellites, European Space and Technology Centre (ESTEC) of the European Space Agency (ESA), SES-ASTRA and University of Luxembourg (Appendix 3).

5.2 Hybrid GA-SA Approach

The GA-SA model developed relies on the sequential application of the GA and the SA techniques. Specifically:

1. GA is applied for a number of generations.
2. The best solution provided by the GA, along with a number of random individuals are extracted from the last population.
3. These individuals are used to initiate a number of SA optimizations.
4. SA executions are performed for a predefined number of predefined iterations, enhancing the GA solution.
5. These solutions enhanced by the application of SA are input back to the last GA population replacing the extracted individuals.
6. GA resumes the evolution of the enhanced GA population for a number of generations.
7. The algorithm goes back to step 2.

This process iterates for a number of fitness evaluations.

At this point it should also be noted that the SA adopted by the hybrid model is not the standard SA but the re-annealing SA, as described in Chapter 3. This allows for the efficient exploitation of the GA solutions and the effective intensification of the search.

5.3 GA-SA Performance Study

To demonstrate the potential benefits of the GA-SA hybrid approach with regard to throughput maximization, this technique is compared to the two best performing techniques of the previous chapter, specifically the GA and the DE. The parameters of the above three techniques, following the tuning of the algorithms performed in Chapter 4, are presented in Table 5.1.

The comparison of the techniques should be done with regard to the minimization of the best performing fitness function, namely the USC (see (4.6)). However, since the performance of the standard metaheuristics examined in Chapter 4 was assesd based on the minimization of DSC (see (4.2)) for the first traffic demand scenario and on the minimization of USC for the second scenario, the same approach will be adopted in this chapter to enable comparison with the corresponding results of the previous chapter.

The convergence plots of the GA, DE and the hybrid GA-SA techniques under the first traffic demand scenario of Fig.4.2 is plotted in Fig.5.1 and the corresponding numerical results are tabulated in Table 5.2. The results reported in the present chapter were acquired by averaging over 15 independent runs. Also statistical confidence for the comparisons is verified by performing the Wilcoxon test with standard confidence level 0.95, as suggested by the statistical analysis of metaheuristics of Chapter 3.

Table 5.1 Optimization Parameters

<table>
<tr><td rowspan="2">GA</td><td>Crossover Rate: 0.95</td><td>Number of Elite Individuals: 30</td><td>Population Size: 5856 Generations:40</td><td rowspan="2">Number of Fitness Evaluations: 234240</td></tr>
<tr><td>Mutation Rate:0.05</td><td>Uniform Crossover Function</td><td>Tournament Selection Function</td></tr>
<tr><td>DE</td><td>Crossover: 0.9</td><td>Scaling Factor: 0.8</td><td>Population Size: 40 Generations: 5856</td><td>Number of Fitness Evaluations: 234240</td></tr>
<tr><td colspan="5">GA-SA Optimization Parameters</td></tr>
<tr><td rowspan="3">GA</td><td>Crossover Rate: 0.95</td><td>Population Size: 5856</td><td rowspan="3">Number of Evaluations before SA: 35136

Generations before SA: 6</td><td rowspan="5">Number of exchanged individuals: 4 1best-3random

GA-SA Loops: 5

Number of Fitness Evaluations: 234240</td></tr>
<tr><td>Mutation Rate:0.05</td><td>Number Elite Individuals: 30</td></tr>
<tr><td>Uniform Crossover Function</td><td>Tournament Selection Function</td></tr>
<tr><td rowspan="2">Re- Annealing SA</td><td>Initial Temperature: 400</td><td>Re-Annealing Interval: 100</td><td rowspan="2">Number of Evaluations before GA: 11712</td></tr>
<tr><td>Exponential Temperature Update</td><td>Boltzmann Annealing</td></tr>
</table>

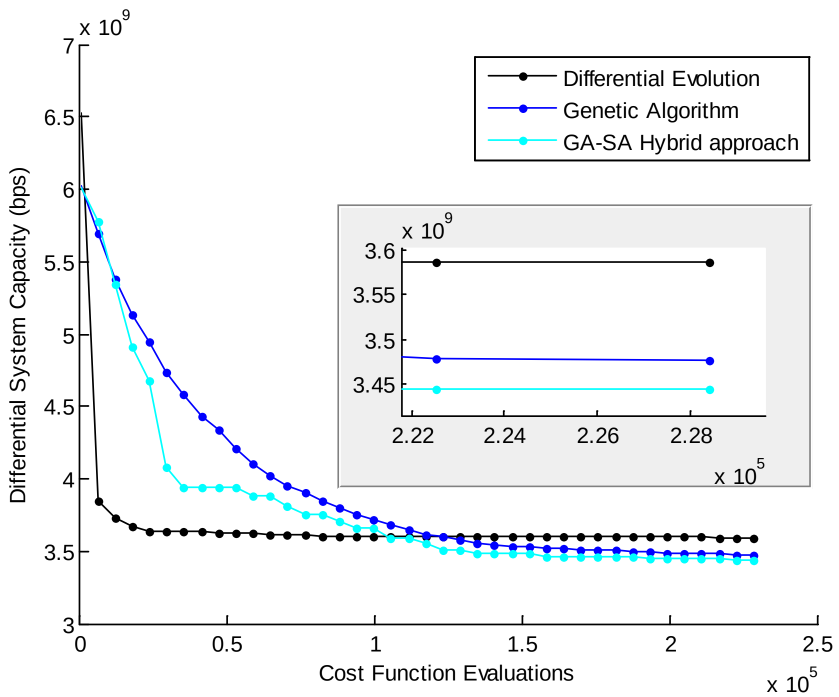


Fig. 5.1 Performance comparison of GA, DE and GA-SA under scenario 1

Table 5.2 Numerical Results (DSC in Gbps)

Algorithm	Fitness Value (Average of 15 runs)	Standard Deviation (of 15 runs)	Overall Best (over 15 runs)
Differential Evolution	3.5854	0.03320	3.5480
Genetic Algorithm	3.4773	0.01678	3.4530
GA-SA	**3.4473**	**0.01007**	**3.4444**

The convergence plots of the GA, DE and the hybrid GA-SA techniques under the second traffic demand scenario of Fig.4.8 are plotted in Fig.5.2 and the respective numerical results are tabulated in Table 5.3.

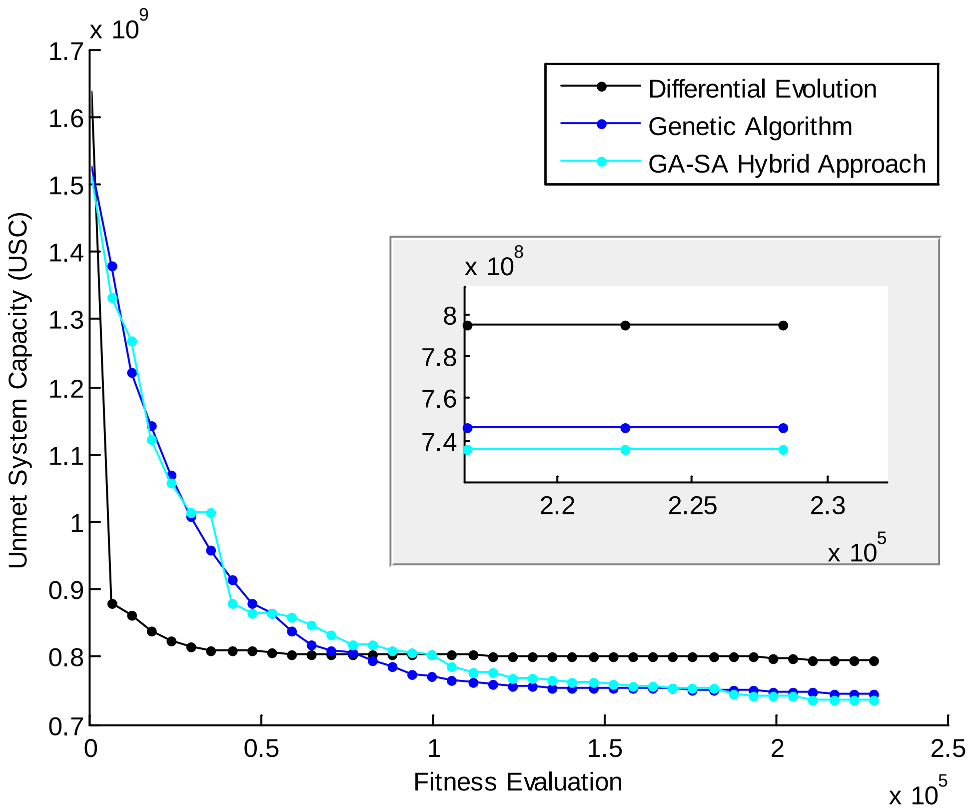


Fig. 5.2 Performance comparison of GA, DE and GA-SA under scenario 2

Table 5.3 Numerical Results (USC in Gbps)

Algorithm	Fitness Value (Average of 15 runs)	Standard Deviation (of 15 runs)	Overall Best (over 15 runs)
Differential Evolution	0.79582	0.04188	0.72260
Genetic Algorithm	0.74579	0.01972	0.70850
GA-SA	0.73612	0.01937	0.70850

The results presented verify the intuitive approach regarding the benefits obtained from the hybridization of explorative and exploitative techniques since the proposed hybrid approach outperforms both GA and DE for both traffic demand scenaria. Specifically, the results provided by the GA-SA improve the GA results by 0.9% in the case of the first scenario and by 1.3% in the case of the second scenario. Moreover, the GA-SA results have a smaller standard deviation ensuring the consistency of the results and the increased reliability of the algorithm with regard to the deviation tendencies over independent runs. Thus, the proposed hybrid GA-SA technique outperforms the GA in every aspect and provides the best results for the problem in hand.

The performance of the system employing the proposed hybrid technique, in the case of the first traffic scenario was improved by 47.3% compared to the performance of the non-flexible system. This improvement is apparent in Fig.5.3, showing the performance in terms of satisfaction of the traffic demand of the GASA-optimized and the non-flexible system. Furthermore, the improvement accomplished by the hybrid approach demonstrates the benefits from the hybridization of explorative and exploitative techniques in general, setting an example for future work in the field.

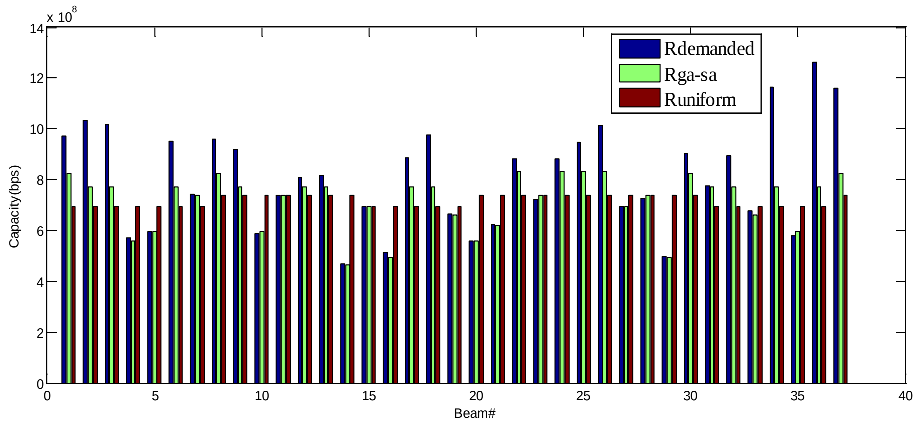


Fig. 5.3 Traffic Demand Satisfaction
of GASA-optimized and Uniform power allocation (scenario 1)

References:

[5.1] G.R. Raidl. *"A unified view on hybrid metaheuristics"*. In F. Almeida, M.J.B. Aguilera, C. Blum, J M.M. Vega, M.P. Perez, A. Roli, and M. Sampels, editors, Hybrid Metaheuristics, vol. 4030 of Lecture Notes in Computer Science, p. 1-12. 2006.

Chapter 6
Multi-objective Optimization Approach

6.1 Multi-objective Optimization

A *multi-objective optimization problem* is formulated in the form:

$$\text{minimize } F(x) \tag{6.1}$$

$$F(x) = (f_1(x), f_2(x), \ldots., f_k(x)) \tag{6.2}$$

$$\text{subject to } g_i(x) \leq b_i \text{ , } i = 1, \ldots. , m. \tag{6.3}$$

where k ($k \geq 2$) is the number of objectives, $\mathbf{x} = (x_1, \ldots, x_n)$ is the vector representing the *optimization variables* of the problem, $F(x) = (f_1(x), f_2(x), \ldots., f_k(x))$ is the vector of objectives to be optimized, $g_i(x)$ are the *constraint functions* and b_i , i = 1, …. , m, are the limits or bounds related to the constraints [6.1].

The development of the multi-objective optimization was triggered by the fact that practical optimization problems are rarely single-objective. In general, a number of conflicting objectives are involved. For instance, during the design of a new product, the manufacturer must maximize its quality, minimize its cost and minimize the environmental impact of the product. Thus, a number of conflicting objectives must be taken into account and the manufacturer has to make an acceptable compromise between conflicting objectives of a multi-objective optimization problem (MOP) [6.1].

It is, therefore, evident that, in contrast to to the single objective optimization, the optimal solution of a MOP is not a single solution optimizing a single objective but a set of solutions representing the possible compromises between conflicting objectives. From this set of solutions the decision maker can select the most satisfactory solution according to his preferences and the priorities set for the various objective. This set of solutions is defined as *Pareto optimal solutions* and comprises all solutions that cannot be improved as to one objective without deteriorating at least another objective. These solutions are called *non-dominated* solutions of the problem.

The concept of *dominance* is essential in MOPs in the course to define the order relation between the solutions. A solution dominates another if it is strictly better in one objective and better or equal with regard to the rest of the objectives. Graphically a solution dominates another if it is located in the box defined by the projections of F(x) on the axes. This is made clear in Fig.6.1 where the solution 1 is located within the box defined by the projections of 2 on the axes; therefore, solution 2 is dominated by solution 1. Thus, the target of multi-objective optimization is to provide a set of all the best performing solutions, i.e. the non-dominated solutions, to the decision maker. This set of solutions is known as *Pareto optimal set* and the image of this set in the objective space is denoted as the *Pareto front*.

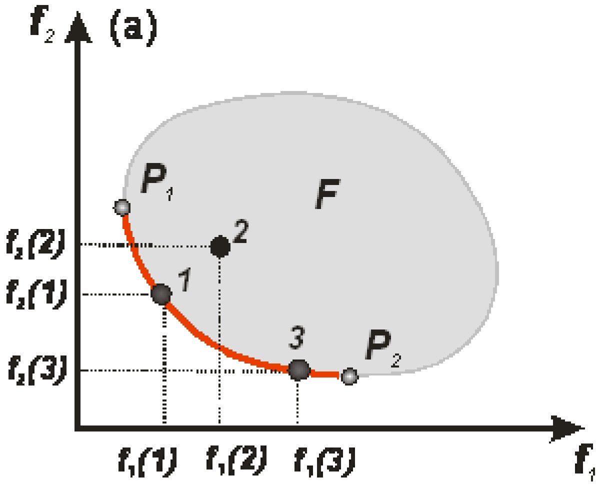


Fig. 6.1. The concept of dominance: solution 2 is dominated by solution 1

In the analysis presented in the previous chapters resource allocation of multi-beam systems was focused on the dynamic allocation of the system power with as the traffic distribution changes. In this course the proposed optimization methods yield substantial capacity gains and significant system performance enhancement. However, such allocations tend to utilize the maximum DC power available whereas its minimization is also desirable due to the scarcity of the satellite resources. Thus, the efficient power allocation in a multi-beam system should be considered as a MOP aiming at maximizing the available system throughput at the same time minimizing the system power consumption.

6.2 Power utilization in multi-beam systems

Efficient power utilization and minimization of DC power consumption in flexible multi-beam systems has not been sufficiently explored in the literature. An early approach in the direction of simultaneously modifying the DC/RF power consumption and the available system throughput is proposed in [6.2] where a number of cost functions are proposed to quantify the response to the traffic demand. To take the power consumption into account, these cost functions incorporate a DC consumption parameter and at the same time maximize the available throughput of the system in response to the traffic demand. Thus, applying this technique the different objective functions related to the satisfaction of the traffic demand and the efficient utilization of the DC power are combined into a single-objective function in a linear way, transforming the MOP into a single-objective problem. This approach is known as scalar approach and the MOP is transformed into the following single-objective optimization problem:

$$\text{minimize } F(x) \tag{6.4}$$

$$F(x) = \sum_{j=1}^{k} \lambda_j f_j(x) \tag{6.5}$$

$$\text{subject to } g_i(x) \leq b_i \text{ , } i = 1, \dots, m. \tag{6.6}$$

where k (k ≥ 2) is the number of objectives and λ_j , j = 1, …, k ($\lambda_j \geq 0$) are the weights for the linear representation of the various objective functions $\sum_{j=1}^{k} \lambda_j = 1$.

The scalar approach, however, requires a priori knowledge of the considered problem objective space, in order to select appropriate weights λ_j, providing the desirable solution (e.g. the solution of the Pareto front providing the maximum throughput). Elaborating on this, the optimal solution **x** obtained from the scalar approach is the point of tangency of the hyperplane F and the feasible space of the MOP as shown in Fig.6.2(a). Thus, to obtain the desirable **x** on the Pareto border, the latter must be known in advance in order to select weights determining a hyperplane F with a gradient parallel to the tangent of the Pareto border at **x.** If, however, the Pareto border is not known in advance, i.e. without an a priori knowledge of the problem, the appropriate weights providing the desirable solution **x** cannot be determined.

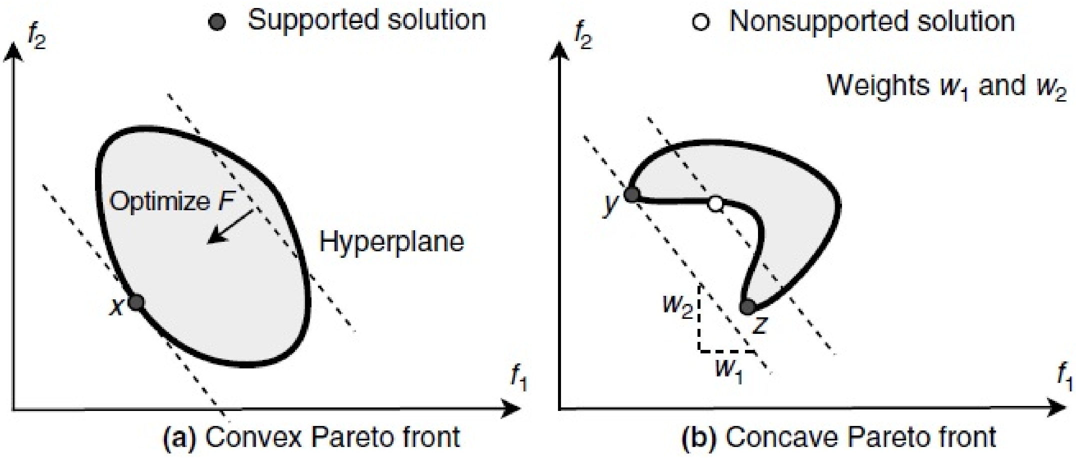


Fig. 6.2 Scalar approach optimization for convex and concave Pareto front

In most cases an a priori knowledge of the problem is not available and therefore the scalar approach cannot be employed. However, in numerous cases the weights are determined based on a common misconception that the weights selected for the linearization of the MOP are representative of the relative importance of the different objectives [6.2]. In these cases the objective space hyperplane F as determined by these weights intersects randomly with the feasible space of the problem providing random solutions of the Pareto border, possibly unacceptable by the decision maker.

An alternative approach to determine the appropriate weights for the linear representation of the scalar approach is the use of the weights as optimization variables of the problem that will be optimized in parallel along with the problem. However, the computational cost of this approach is significant and cannot guarantee to provide all Pareto optimal solutions since the latter are limited to the convex hull of the Pareto front (i.e. solutions in concave regions of the Pareto front cannot be found as shown in Fig. 6.2(b). The solutions that cannot be found are called non-supported solutions).

Concluding, the transformation of a MOP into a single-objective problem provides only one solution of the Pareto optimal solutions. On the other hand dominance based approaches provide all non-dominated solutions of the set of Pareto solutions and the trade-off curve of different objectives. This useful information regarding the systems objective space is suppressed by the scalar approaches and the decision maker is deprived of all the alternative solutions.

To overcome these drawbacks the present work explores the power allocation problem as a strictly multi-objective problem (MOP) and not as a linear function of the various objectives. Formulating this problem in a dominance based multi-objective setting is novel. In this course, a state-of-the-art multi-objective genetic algorithm (MOGA), the non-dominated sorting genetic algorithm II (NSGA-II) [6.1] , to be presented next, is employed. This novel approach was presented in the 30th AIAA International Communications Satellite Conference (ICSSC), 24-27 September, 2012, Ottawa, Canada (Appendix 2).

6.3 Non-dominated Sorting Genetic Algorithm II (NSGA-II)

NSGA-II [6.1] is the most popular and referenced dominance based multi-objective algorithm in the literature. It is a GA employing all typical genetic operators of GA described in Chapter 3 (i.e. selection, crossover, mutation). Its differentiation from standard GAs lies on the fitness assignment of the individuals since multiple objectives are considered and the non-dominated solutions are provided at the end of the optimization. As a GA, NSGA-II belongs to the metaheuristic techniques; also, since metaheuristics do not guarantee the optimality of the solution obtained, the goal of this technique is to provide an approximation of the Pareto optimal set having two necessary properties: convergence to the Pareto optimal solutions and uniform diversity, i.e. a good distribution of the solutions obtained around the Pareto optimal front.

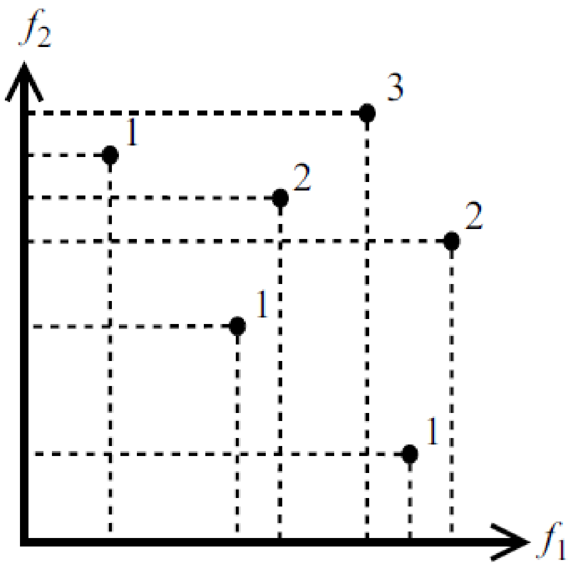


Fig. 6.3 Dominance depth ranking method

In this course, the fitness assignment of all individuals of the population is based on their ranking with respect both to convergence and diversity. In particular, the individuals of the population are initially ranked with respect to their convergence; all non-dominated individuals (solutions) are assigned rank 1.The individuals that are not dominated except by individuals of rank 1 receive rank 2. In general, rank k is assigned to individuals that are dominated only by solutions of ranks 1, 2, …, k-1. This ranking method of NSGA-II is called *dominance depth* and is depicted in Fig.6.3.

In practice the dominance based ranking is performed by the following iterative process. All non-dominated individuals are assigned rank 1 and then they are removed from the population in order to assign rank 2 to the current non-dominated individuals. This process is repeated until the population is empty. Then individuals of the same rank are ranked according to their diversity, namely their distance from surrounding individuals. Individuals with higher distances (i.e. good distribution) are assigned higher rank than others. In practice the diversity ranking is implemented based on the concept of crowding distance. The crowding distance is defined as the circumference of the rectangle defined by the left and right neighbor of an individual and infinity if there is no neighbor. The concept of crowding distance is illustrated in Fig.6.4(b) where the individuals a, b, c and d belonging to rank 1 are ranked according to their diversity. The individuals with the higher diversity are a and d with a crowding distance equal to infinity, then follows the solution b located in the bigger cuboid and then c.

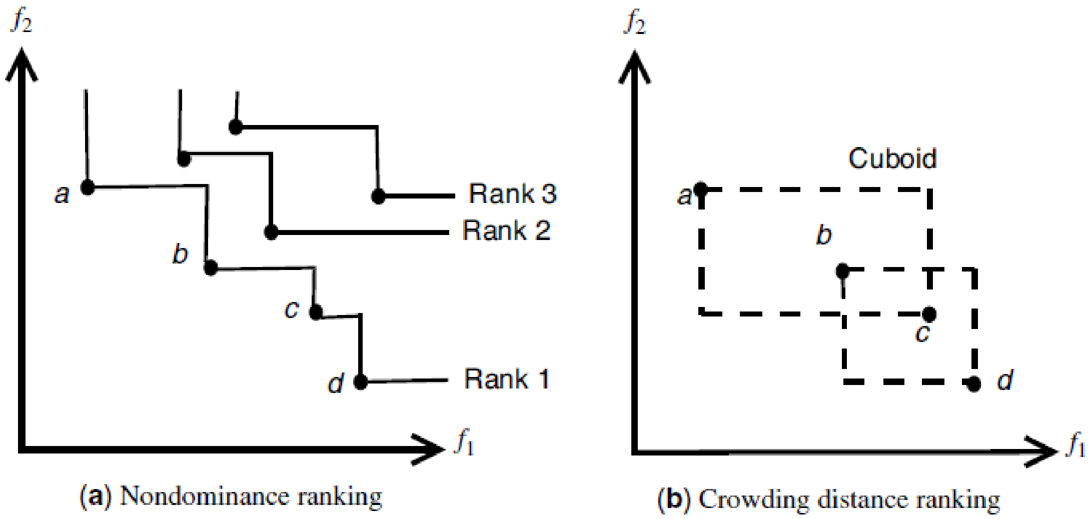


Fig. 6.4 Ranking method in NSGA-II

Following to the non-dominance ranking and the crowding distance ranking of the population the tournament selection is used, to determine the best among two individuals based on their non-dominance rank. If the two individuals belong to the same rank the best among the two individuals is selected based on their crowding distance rank. The selected individuals are then evolved by the typical genetic operators and at the end of the optimization the non-dominated individuals having a good distribution form the Pareto front.

6.4 Simulation Results

The multi-objective optimization performed hereafter aims at the enhancement of the solutions provided in previous chapters in terms of power utilization. In this course, the best performing individual, i.e. the result of the GA-SA approach, is used as member of the initial population of solutions of the NSGA-II algorithm. Thus, the Pareto front provided by the NSGA-II will include a solution providing the best performance in terms of throughput maximization whereas this solution will probably be enhanced by the NSGA-II even further, with regard to minimizing the consumed power.

Moreover, the best performing cost function, namely the USC, is chosen as the cost function. Thus, the optimization problem takes the following form:

$$\text{minimize } f_1 \tag{6.7}$$

$$f_1 = \sum_{b=1}^{37} \max\{R_{b,req} - R_{b,off}, 0\} \tag{6.8}$$

$$\text{minimize } f_2 \tag{6.9}$$

$$f_2 = \sum_{b=1}^{37} P_b \tag{6.10}$$

subject to:

$$\sum_{b=1}^{37} P_b \leq 2350\text{W} \tag{6.11}$$

$$P_b \leq 100\text{W} \tag{6.12}$$

where $R_{b,req}$ denotes the bit rate requested by the users in beam b and $R_{b,off}$ denotes the bit rate offered, by beam b. The best performing result, namely the result of the hybrid GA-SA approach is used as the initial seed of the optimization. The fitness value of this result with regard to the fitness functions of the MOP f_1 (6.8) and f_2 (6.10) is:

$$\sum_{b=1}^{37} \max\{R_{b,req} - R_{b,off}, 0\} = 0.7085 \text{ Gbps} \tag{6.13}$$

$$\sum_{b=1}^{37} P_b = 2030.3\text{W} \tag{6.14}$$

The results presented involve averaging of 15 independent runs.The parameters of the optimization are shown in Table 6.1

Table.6.1 NSGA-II Optimization Parameters

Parameter	Value
Population Initialisation	Random
Population Size	6000
Crossover Function	Uniform
Crossover Rate	0.9
Mutation Rate	0.05
Selection Function	Tournament
Fitness Evaluations	240000

The Pareto front obtained after 15 runs depicted in Fig.6.5 provides complete information concerning the trade-off between the offered capacity and the power requirements, whereas the result of the GA-SA approach used as initial seed is further enhanced as expected.

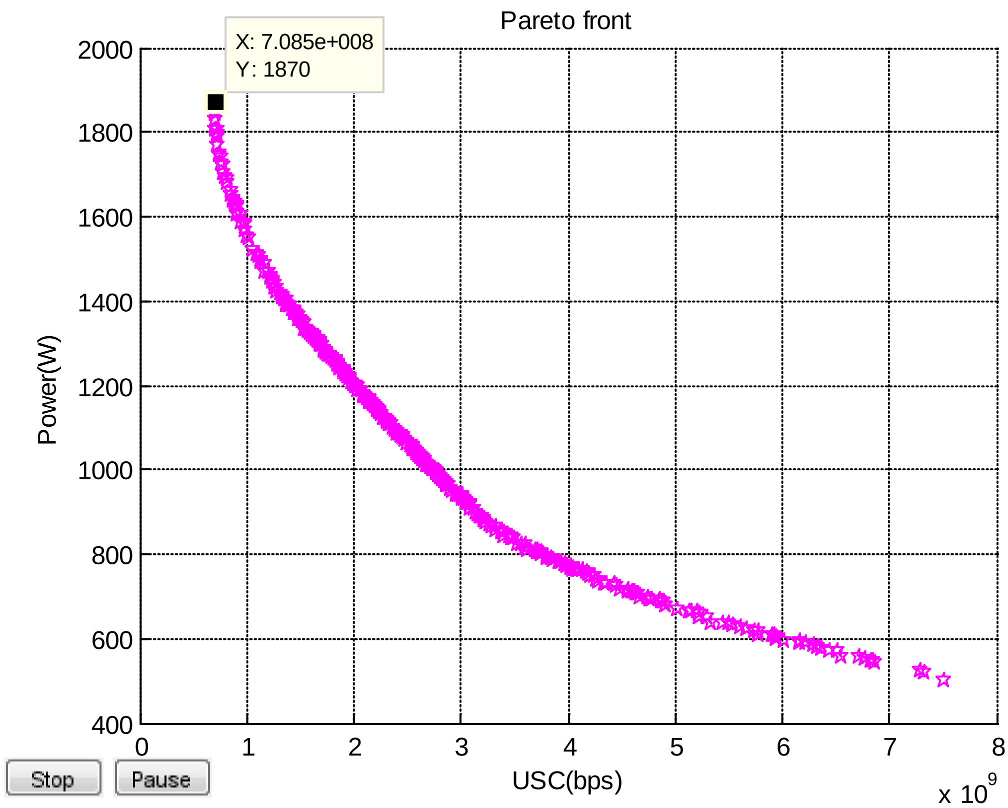


Fig. 6.5. Pareto front encompassing the enhanced result of the single-objective optimization.

From the Fig.6.5, it is deduced that the Pareto front obtained provides a large number of non-dominated solutions as well as a very good diversity, since the Pareto front appears continuous. It is also evident that the result of the single-objective, GA-SA approach, (based on (6.13)-(6.14)) has been significantly improved with regard to the total consumed Power (f_2 (6.10)). In particular, the allocation obtained applying the

GA-SA approach, namely the one consuming 2030.3W for an USC of 0.7085Gbps, drops to a power consumption of 1870W for the same USC. In other words, the multi-objective approach satisfies the same traffic demand as the best single-objective approach, namely the GA-SA, at 7,9% lower power consumption.

Furthermore, the slope of the left part of the Pareto front is steep enough to allow for even more power saving, if the USC is allowed to slightly deteriorate. Fig. 6.6 details the upper left part of the Pareto front to demonstrate the trade-off between the satisfaction of the traffic demand and the relevant power requirements. In this case, the satellite operator can save up to 12% power by decreasing the USC by only 2.7%. It is evident that the Pareto front obtained provides complete information as to the trade-off between the two objectives, enabling the decision maker, i.e. the satellite operator, to adjust the level of satisfaction of the current traffic requirements according to the priority of the two objectives.

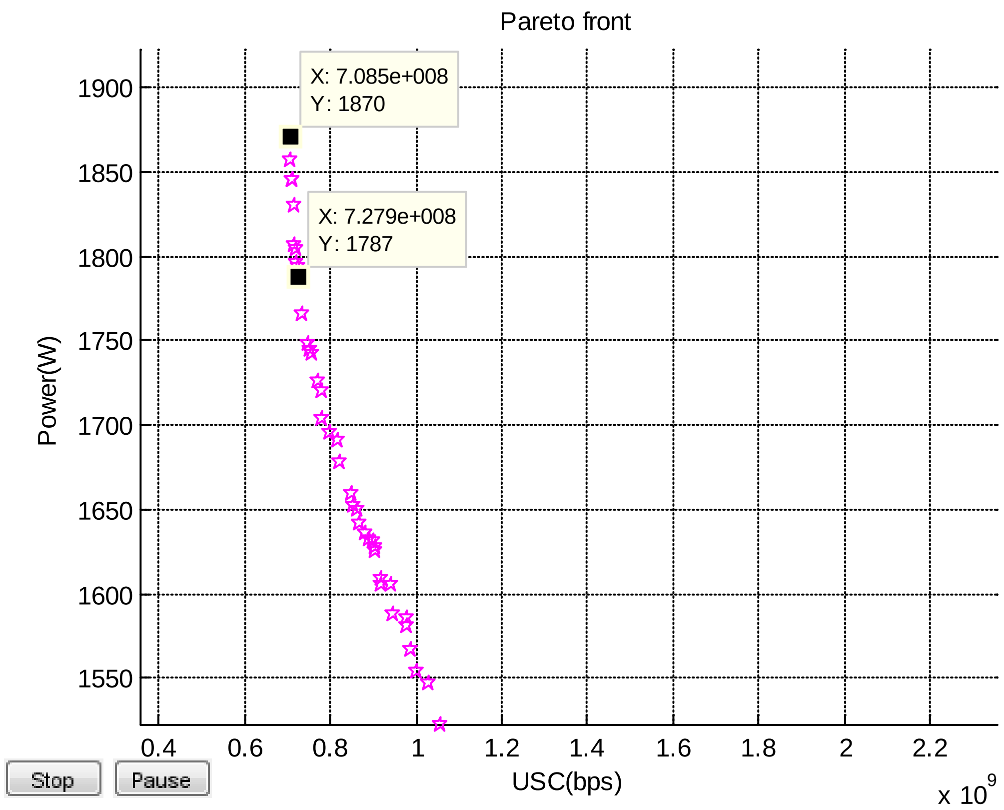


Fig.6.6 Pareto front, zoom in

Thus, the previous analysis demonstrates the potential benefits from using multiple objective optimization in payload design as the addition of another objective provides gains over the single objective approach. Furthermore, the trade-off curve obtained from this approach can provide significant insight and help in system design rendering the proposed approach an enhancement to existing payload optimizers.

References:

[6.1] El-Ghazali Talbi. *Metaheuristics from design to implementation.* John Wiley & Sons, Inc. 2009.

[6.2] Gabellini P., D'Agristina L., Alagha N., Angeletti P. *"On the Cost Function for the Performance Optimization of Multi-Beam Broadband Satellite Systems,"* 2nd ESA Workshop on Advanced Flexible Telecom Payloads, 17-19 April 2012

Appendix – Publications

Appendix 1

2nd ESA Workshop on Advanced Flexible Telecom Payloads, 17-19 April 2012, Noordwijk, The Netherlands

Power allocation in Multibeam satellites

A hybrid-Genetic Algorithm approach

Alexios Aravanis[1,2], Bhavani Shankar[1], Gregoire Danoy[1], Pantelis-Daniel Arapoglou[3], Panayotis Cottis[2], Bjorn Ottersten[1]

[1]University of Luxembourg
Campus Kirchberg, 6, rue Richard Coudenhove-Kalergi
L-1359 Luxembourg, Luxembourg
alexios.aravanis@uni.lu, bhavani.shankar@uni.lu, gregoire.danoy@uni.lu, bjorn.ottersten@uni.lu

[2]National Technical University of Athens
Zografou Campus, Heroon Polytechniou 9,
15780 Zografou, Greece
pcottis@central.ntua.gr

[3]European Space Research and Technology Centre
TEC-ETC Section
2200 AG Noordwijk, The Netherlands
pantelis-daniel.arapoglou@esa.int

INTRODUCTION

To compete with the ever-expanding terrestrial technology, current satellite systems need to keep pace with the evolution of new applications, including mobile video broadcasting and multimedia interactive services. Apart from the services offered, satellite systems also need to consider the deployment of low-cost terminals to reduce the cost of the customer premises equipment compared to terrestrial systems.

Multibeam satellites can play a key role toward these objectives. On the one hand the high gain offered by multibeam antennas may compensate for performance degradation faced by low-cost user terminals. On the other hand, multibeam systems can ensure the competitiveness of satellite systems in the evolving markets, by providing the necessary flexibility. Flexibility in design is an essential issue for the competitiveness of satellites and involves the ability to reconfigure the mission to meet the unexpected market changes during the lifetime of the spacecraft [1]. Flexibility can be viewed from various aspects because multibeam systems are able to support reconfigurability of power and frequency plans, routing and switching functionalities in response to traffic demand, exhibiting both temporal and spatial variability.

In order to rip the benefits of the flexibility provided by the multibeam systems, a system entity has to reconfigure the power and frequency plan according to the traffic demand and the channel variability. However, due to frequency reuse, multibeam systems are subjected to inter-beam interference. Hence a dynamic allocation of system resources in order to satisfy the varying traffic demand, while mitigating the interferences seems imperative for efficient multibeam systems. These system resources are the bandwidth and the transmit power and the resource allocation is an optimization problem that aims at maximizing the throughput of each beam subject to the current traffic demand.

Several publications have studied resource allocation in satellite communications. The power and bandwidth allocation problem studied in this paper has been formulated in [2] and it is based on the use of advanced equipment such as flexible amplifiers and bandwidth processors. Subsequent efforts toward optimization have been largely guided by the ESA ARTES activities. In [2], Unmet System Capacity (USC) is proposed as a figure of merit for the optimization of a system and antenna/ payload design co-design tool. A Genetic Algorithm (GA) [7] based optimization strategy is proposed in [3] for the co-design tool involving iterative antenna and system optimizations. The system optimization involves the power and frequency plans and employs the following steps :(1) an initial screening of the space of solutions is done employing GA , (2) a further single step optimization starting from the solution given by the previous step consists of a Variable Neighborhood Search (VNS) and (3) a refinement of the latter solution is given via multiple iterations of an Iterated Local Search (ILS). The use of a similar architecture is considered in the optimization of resources (illumination, power and/ or carrier) in a Beam Hopped (BH) system [4, 5]. Flexible and BH transmission schemes are also considered in [6] to provide flexibility in frequency (through bandwidth allocation) and time domain (through time slot allocation) respectively. Power and bandwidth (or time slot allocation) are optimized to enhance the Satisfaction Factor. A heuristic algorithm that iteratively optimizes these resources is provided.

Though several techniques have been considered for resource optimization, a systematic study leading to the selection of the appropriate optimization technique is not available. Focusing on the power optimization problem in multibeam systems, the present paper highlights its high dimensionality and multimodal landscape which motivates the use of metaheuristic techniques. Various well-known metaheuristic techniques comprising trajectory based (such as Simulated Annealing (SA) [7]) or population based (such as Particle Swarm (PSO) [7] or Evolutionary Algorithms) methods are explored in a systematic way. Then, simulations are performed to classify the algorithms with regard to their performance, thereby providing hints about which technique should be used in different occasions. Subsequently, a better performing hybrid GA scheme is proposed, based on the hybridization of a GA with a trajectory based technique (SA) to further enhance the GA solutions.

Since, in general, bandwidth allocation is done in terms of carriers of fixed bandwidth, the bandwidth allocation problem can be transformed into a power allocation problem. Furthermore, recent advances in Multi-Port Amplifier (MPA) technology allow for efficient power splitting whereas no publicly available technology for bandwidth sharing is known to the authors. These motivate the focus on power optimization in this work.

The remainder of this paper is organized as follows. Section 2 presents the formulation of the power allocation problem. Then the considered optimization techniques are described in section 3. Experimental results using the proposed approaches are proposed and analyzed in section 4. Finally section 5 contains our conclusions and perspectives.

2. PROBLEM FORMULATION

Assume a multibeam satellite system with $b = 1,\ldots,N$ beams, employing a typical four color reuse pattern (two colors in frequency and two in polarization), where the available total downlink bandwidth of the system B_{TOT} is equally distributed among the four colors. It is further assumed that the available bandwidth reused over the different colors is equally divided into the beams of each color and the beam bandwidth equally divided into carriers. A carrier represents the elementary system entity for conveying different streams of information.

Given this bandwidth allocation, the problem pertains to the appropriate allocation of the total available power of the system P_{TOT}, which is a function of the platform total DC power on board the satellite, to each beam P_b, so that the offered beam bit rate meets the user requirements in the same beam. The power allocated to each beam P_b, will then be equally divided among the N_c carriers of the beam $P_{b,c} = P_b / N_c$.

Thus, the problem is formulated as an optimization problem aiming at minimizing the following cost function:

$$\text{Min f} = \sum_{b=1}^{N} abs\left[R_{b,req} - R_{b,off}\right], \quad \text{while} \sum_{b=1}^{N}[P_\text{b}] \le P_\text{TOT} \text{ and } P_\text{b} \le P_\text{b,con} \tag{1}$$

where, $R_\text{b,req}$ denotes the bit rate requested by the users in beam b, $P_\text{b,con}$ is the power constraint of each beam and

$$R_\text{b,off} = \sum_{c=1}^{N_\text{c}} \text{B} \times f_\text{DVB-S2}(SNIR_\text{c}), \tag{2}$$

denotes the offered, cumulative bit rate of all N_c carriers in beam b. The function $f_\text{DVB-S2}(SNIR_\text{c})$ quantifies the spectral efficiency of the various modulation and coding schemes employed by DVB-S2 as a function of the SNIR_c of carrier c, and B denotes the bandwidth of each carrier.

The $SNIR_c$ of each carrier c, transmitted with power $P_{b,c}$ and bandwidth B is calculated as follows:

$$SNIR_c = \frac{a_b^2 P_{b,c}(OBO)}{N_0(a_b)B + \sum_{q\in\Phi} \alpha_q^2 P_{q,c}(OBO) + I_{adj_{ch}}(B, XPD) + I_{adj_{sat}} + I_{inter}(OBO, C_b, Mod)} \tag{3}$$

Φ: is the set of co-channel beams in the coverage area with active carriers overlapping with the bandwidth of the intended carrier c (co-channel interference into intended beam b)

α_q: is a gain factor encompassing the effect of: satellite antenna beam q, terminal receive antenna gain, free space loss, clear sky attenuation and rain attenuation

N_0 is the noise power spectral density which is a function of a_b because of the increase in noise temperature under rain fading conditions

I_{adj_ch}: accounts for adjacent channel interference due to filter imperfections (function of B), including spillover from the beams in orthogonal polarization if both polarizations are employed (function of XPD)

I_{adj_sat}: Inter-system interference caused by adjacent satellites

I_{inter}: Intermodulation interference
The notation $P_{b,c}(OBO)$ denotes the dependence on the appropriate OBO according to the modulation scheme employed
It is evident that the $SNIR_c$ of each carrier is not characterized by a one to one relation with the power used in beams of the same color. Moreover, the corresponding relation to spectral efficiency is non-linear and is typically obtained employing look up tables. Hence, the previous analysis indicates a non-linear and non-continuous dependence of the cost function on the optimization variables (power).

3. OPTTIMIZATION TECHNIQUES

3.1 Metaheuristics

The non-linear and non-continuous dependence of the cost function on the optimization variables referred to above suggest the use of general-purpose algorithms, like Metaheuristics. Metaheuristics make few or no assumptions about the optimization problem, allowing the optimization of complex problems, like the one studied in the present paper. However, no general axiomatic rule to follow exists when choosing a metaheuristic technique and no universal criterion can be proposed in order to prove the superiority of a specific technique over others, as also suggested by the "No Free Lunch" theorem [8]. The "No Free Lunch Theorems for Optimization" state that two optimization algorithms are equivalent, when their performance is averaged over all possible optimization problems. For this reason and in order to specify a suitable technique for the problem under consideration, a systematic performance study has been carried out of various well known metaheuristics. These techniques are three population based (GA, Differential Evolution (DE) [7], Particle Swarm Optimization (PSO)) and one trajectory based technique (SA).
There is, however one general intuitive criterion, that can be used as a guide, at the beginning of the study, before the simulations and the classification of the techniques. This criterion is matching the landscape of the problem's search space to one of the two available search strategies: The exploration of the search space and the exploitation of the best solution.
On the one hand, the exploration is an extensive search of the search space, to make sure all regions of the search space have been explored and the search is not confined to one location. This search is converging to the most promising regions, in terms of containing a satisfactory solution. Exploitation on the other hand is an intensive search within the promising regions in order to discover the "best" solution of the region. Trajectory-based metaheuristics are more exploitation oriented, whereas population based metaheuristics are more diversification oriented.
The complexity of the problem as well as the large number of variables (multibeam systems can have hundreds of beams) and the perceived multimodal landscape of the problem's search space suggest the use of an explorative technique, namely a population based metaheuristic. This deduction however, may be used as an early intuitive approach, but not as a reliable criterion. For that reason the systematic study of the aforementioned metaheuristics is carried out. A brief introduction to these techniques follows.

3.2 Trajectory Based Metaheuristics

Trajectory based techniques aim at improving a single solution following a specific search trajectory over the search space. This trajectory is determined by applying an iterative procedure that chooses the next station according to the current solution. In each iteration of the process, a set of candidate next stations is defined applying a certain operator on the current solution. This new set of solutions is called a *neighborhood* and, once it is defined, an element of the set (a *neighbor*) is selected to replace the current solution and becomes the next station of the search trajectory. The properties of the *neighbor* that will replace the current one depend on the optimization technique and the way the trajectory is planned. It is possible that points having a cost function value worse than the current one are accepted, provided that the trajectory planning allows so (e.g. in order to escape from some local optimum). The same process iterates and all steps concatenate into forming the search trajectory of the problem until some stopping criteria are satisfied.

3.2.1 Simulated Annealing
SA is a typical trajectory based metaheuristic that mimics the physical process of metal annealing, where the gradual reduction of *temperature* leads the system to a point of minimum energy. In SA, the *temperature* is the operator defining the properties of the *neighborhood*, i.e. the size of the *neighborhood* around the current solution, as well as the probability of accepting a *neighbor* with higher cost function value.

3.3 Population Based Metaheuristics

Unlike trajectory based metaheuristics, population based metaheuristics do not start the search from a single solution, but from an initial population of solutions. This diversity of the initial solutions is the very reason why an extensive exploration of the search space is possible with this family of techniques. After initializing the population, an iterative procedure is invoked: In each iteration, operators are applied to the current population to generate a new population. Individuals from this new population are inserted in the current population, using some selection policy. This whole process keeps iterating until the stopping criterion is satisfied.

3.3.1 Particle Swarm Optimization

PSO is a population based metaheuristic that mimics the social cooperative and competitive behavior of swarms. Each particle in the swarm represents a candidate solution and is represented by a position and velocity (i.e. flying direction). Each particle move is influenced by its personal success (cognitive factor C1) and the success of the whole swarm or of a specific neighborhood (social factor C2).

3.3.2 Differential Evolution

DE is a population based metaheuristic inspired from natural biological evolution. The operator used to evolve the initial population of solutions is a vector difference between 3 individuals of the population. The possibility of each dimension of the vector to be evolved is defined by the parameter $CR \in [0,1]$, whereas a scaling factor $F \in [0,1]$ controls the amplification of the differences. The generated new solution is inserted in the population if improving the parent one.

3.3.3 Genetic Algorithms

GA is a population based metaheuristic, based on the Darwinian evolutionary model, which consists in the following iterative process. Two parents are first selected from the whole population with a given selection criterion. Then two genetic operators are sequentially applied with some probability, namely recombination that exchanges portions of the parents solution vectors, and mutation that randomly modifies the individual. Finally the generated offsprings are evaluated and inserted back into the population following a given criterion.

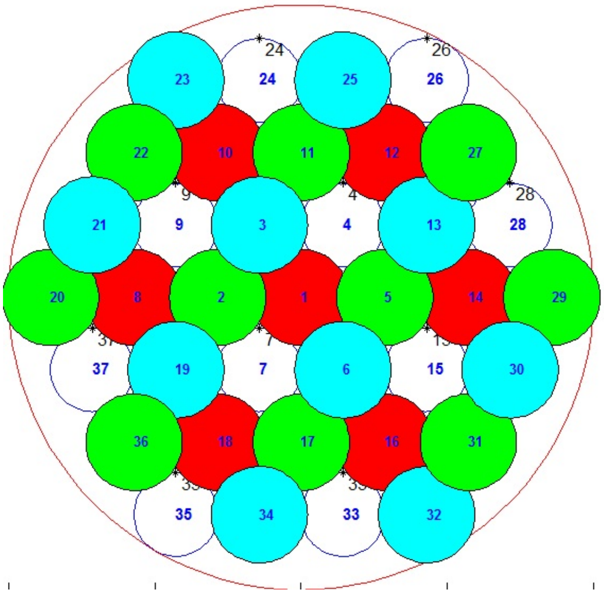
3.4 Hybrid GA-SA Approach

The experimental verification of the suitability of explorative techniques for the optimization of the problem, (provided in the next section), motivated the hybridization of the best performing population based metaheuristic (GA) with the trajectory based one (SA). Indeed, metaheuristics hybridization permits to exploit and combine the advantages of both approaches. Our GA-SA model applies the SA after every *n* generation of the GA and for a predefined number of steps. The solutions sent to the SA consist of the current best solution plus a random one. The improved solutions are then inserted back in the GA that iterates for *n* new generations.

The SA considered includes a re-annealing process that raises the *temperature* of the system after a predefined number of iterations. This permits to efficiently exploit the “good” solutions provided by the GA and avoids fast convergence to local optima.

4. EXPERIMENTAL RESULTS

For the simulations presented herein, a 37-beam system has been simulated, where the antenna pattern is modeled with the help of a Bessel function and the link budget is calculated based on one user per beam, located at the beam edge (worst case position). The parameters of the link budget as well as the beam layout are presented in Fig. 1.



LINK BUDGET PARAMETERS FOR CAPACITY RESULTS

Parameter	Value
Frequency Band	Ka (26.5 GHz)
User Link Bandwidth B_U	500 MHz
HPA saturation power Pt	80 W
Max satellite antenna gain GT	52 dBi
Output Back Off OBO	5 dB
Satellite $EIRP$	66 dBW
Free Space Loss L	212 dB
Terminal Antenna Gain GR	41.7 dBW
Terminal noise Temperature T	207°K
Receive C/N	20.2 dB
External $(C/I)_{EXT}$	30.0 dB

Fig.1 Beam Layout and Link Budget Parameters

Simulating the above system, the standard versions of the metaheuristics techniques described in Section 3 are compared with regard to their performance in minimizing (1). The parameters chosen for the simulations are depicted in Table. 1. All algorithms have been experimentally tested with population sizes ranging from 40 to 6000. The proposed parameters are the best performing ones.

Table. 1 Optimization Parameters

<table>
<tr><td>SA</td><td>Initial Temperature:
400</td><td colspan="2">Temperature Function:
Exponential Update</td><td>Evaluations:
234240</td></tr>
<tr><td>PSO</td><td colspan="2">C1 = C2 = 2</td><td>Population Size:
370</td><td>Evaluations:
234240</td></tr>
<tr><td>DE</td><td colspan="2">CR = 0.9, F = 0.8</td><td>Population Size:
40</td><td>Evaluations:
234240</td></tr>
<tr><td>GA</td><td colspan="2">Crossover rate = 0.95, Mutation rate = 0.05
Elite Individuals = 30</td><td>Population Size:
5856</td><td>Evaluations:
234240</td></tr>
<tr><td colspan="5">GA-SA Optimization Parameters</td></tr>
<tr><td>GA</td><td>Crossover rate = 0.95,
Mutation rate = 0.05
Elite Individuals = 30</td><td rowspan="2">GA-SA
Interval:
5</td><td>Population Size:
5856</td><td>Evaluations before SA:
35136</td></tr>
<tr><td>Re-anneal SA</td><td>Initial Temperature:
400</td><td>Re-annealing Interval:
100</td><td>Evaluations before GA:
11712</td></tr>
</table>

The comparison of the standard techniques available in literature is shown in Fig. 2, where the average of the minimum values of the cost function (1) is depicted, for every 5856 cost function evaluations. The results reported involve averaging over 15 independent runs, as the computational time needed for 234240 evaluations did not allow the use of a bigger sample. The statistical confidence in the comparisons is assessed by performing the Wilcoxon test [9]. The numerical results of the cost function averaged over 15 runs are shown in Table. 2

Table. 2 Numerical Results

Algorithm	Solutions Average	Standard Deviation	Overall Best (over 15 runs)
Simulated Annealing	5.5634e+009	3.5583e+008	4.8916e+009
Particle Swarm	4.2843e+009	3.6050e+008	3.8605e+009
Differential Evolution	3.5854e+009	3.3195e+007	3.5480e+009
Genetic Algorithm	3.4773e+009	1.6783e+007	3.4530e+009

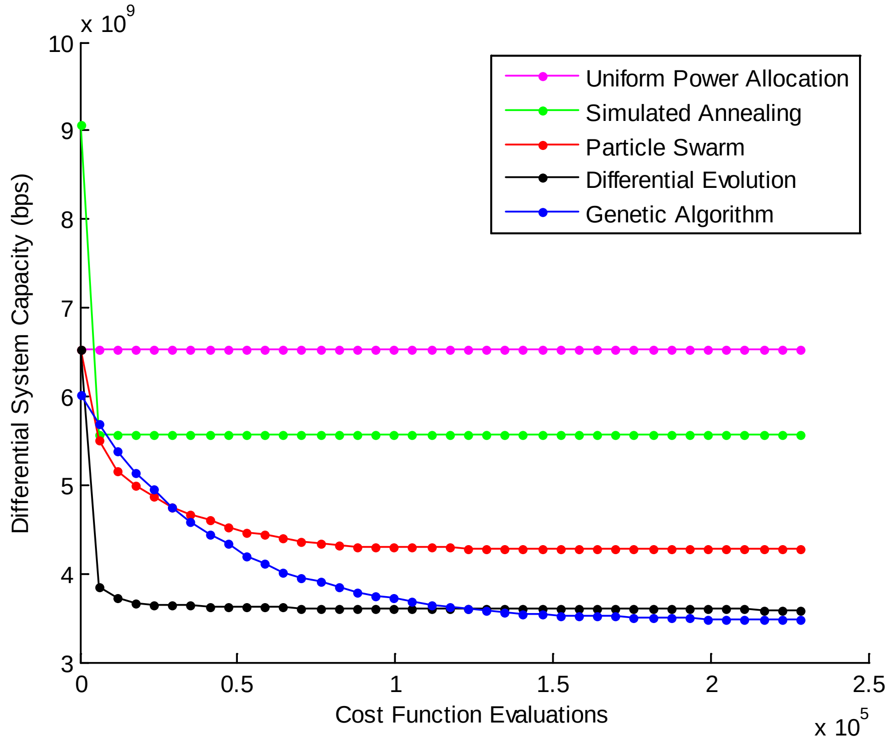


Fig. 2 Comparison of standard techniques available in literature

At the outset, the use of optimization provides an improvement to the tune of 46.8% in the cost function over the static uniform allocation case. DE, one of the most successful techniques for continuous optimization problems provided the best speed of convergence and a final result close to the best result, provided by the GA. DE is therefore the best option (out of the ones compared) for systems where good solutions must be reached in minimum time.

The best result, however is provided by the GA. GA is the slowest converging technique, but is returning the best and more consistent results with a smaller standard deviation than DE. Among the standard techniques found in literature and implemented, GA and DE exhibit the best behavior, indicating a trade-off between convergence speed and quality of solution.

Another interesting conclusion is that the intuitive approach described in section 3 has been verified, as the standard SA performs poorly compared to the population based techniques. The suitability of explorative approaches for the optimization of the problem, motivated the hybridization of the population based technique with an exploitive technique in order to further exploit the provided solutions, as described in section 3.4. The complementary search strategy was applied to the more stable of the two best performing techniques and the one providing the best results, namely the GA, in order to evaluate the improvement yielded to the best solution. The parameters chosen for this hybrid optimization are depicted in Table. 1.

The result of the proposed approach compared to the two best performing techniques GA and DE is depicted in Fig. 3 and the numerical results of the cost function averaged over the 15 independent runs are presented in Table. 3. Statistical confidence in the comparison is verified by Wilcoxon.

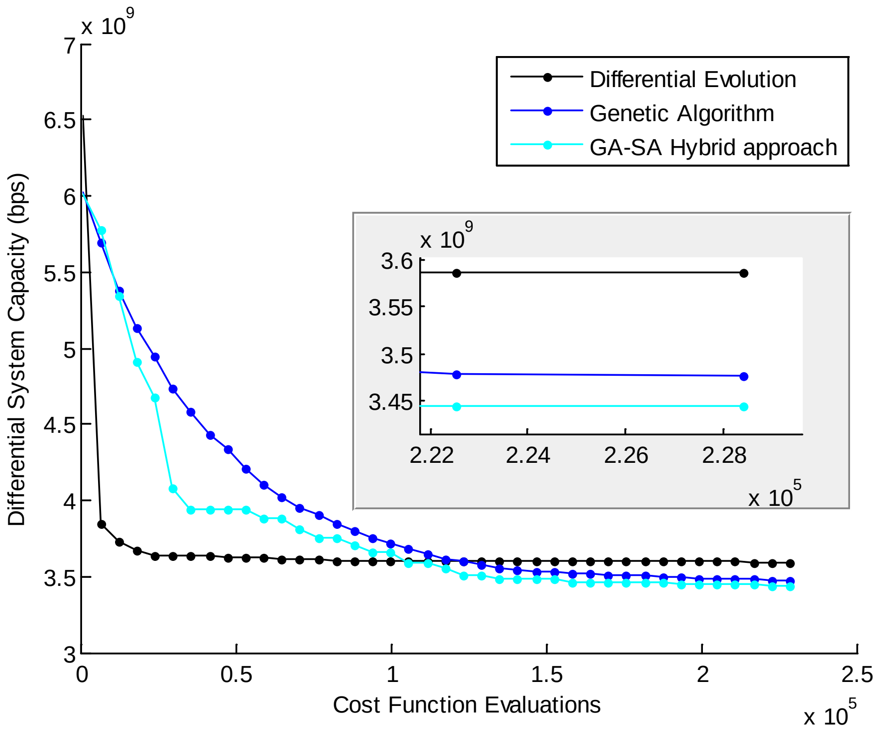


Fig. 3 Performance Comparison of GA-SA, GA and DE

Table. 3 Numerical Results

Algorithm	Solutions Average	Standard Deviation	Overall Best (over 15 runs)
GA-SA	3.4473e+009	1.0070e+007	3.444e+009

The hybrid approach proposed outperforms the best performing techniques. The final result provided by GA-SA improves the one provided by the GA by 1%. In addition GA-SA has a smaller standard deviation, than the results of GA, ensuring the consistency of the results provided in different runs. Furthermore the speed of convergence of the hybrid approach is better than the one of GA. Bottom line the proposed hybrid approach outperforms GA in every aspect and even bridges the gap between the speed of convergence and the quality of solution. The performance of the GA-SA is presented in Fig. 4, also plotted for comparison is the achieved capacity due to uniform power allocation.

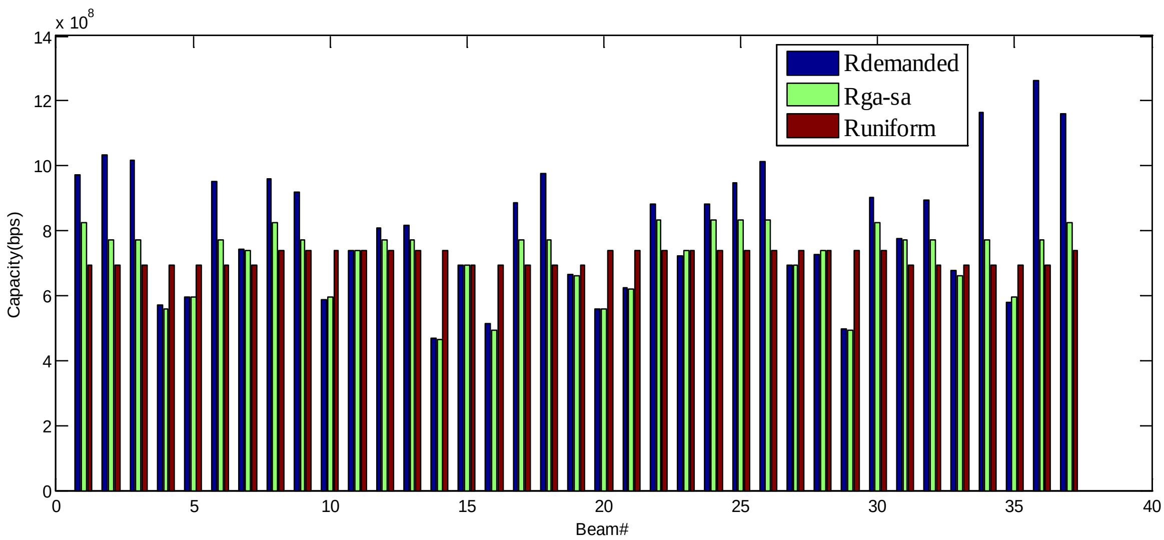


Fig. 4 GA–SA Performance vs Uniform power Performance

5. CONCLUSION

This paper demonstrates how satellite multibeam broadcasting can benefit both from dynamic power allocation schemes, in general, and from the use of the proposed techniques, in particular, since the differential capacity of the system decreased by 46.8% in the case of the GA and by 47.3% in the case of the hybrid GA-SA, compared to the system's performance for uniform power allocation. Moreover the systematic study of the various metaheuristic techniques provides an overview of their performance at the optimization of the power allocation per se and their performance when different parameters of the problem, like the quality of the solution, the variation and consistency of the results and the speed of convergence, have to be taken into account. A multibeam satellite system employing DE techniques, for instance, may deal with rapidly changing traffic requirements, a situation where the convergence speed is of interest, whereas if the quality and the variation of the results is of interest, GA and preferably GA-SA hybrid approaches can be used. Also the improvement accomplished by the hybrid approach proposed, demonstrates the benefits from the hybridization of explorative and exploitive techniques and sets an example for future work in the field.

6. REFERENCES

[1] C. Balty, J-D. Gayrard, P. Agnieray, "Communication Satellites to enter new age of flexibility", Acta Astronautica, vol. 65, pp. 75-81, April 2009.

[2] R. Rinaldo, X. Maufroid, R. C. Garcia "Non-Uniform Bandwidth and Power Allocation in Mult-beam Broadband Satellite Systems", In Proceedings of AIAA 2005, Rome.P. Gabellini, N. Gatti, G. Gallinaro, A. Vernucci, G. Vecchi, F. Della Croce, and R. Rinaldo, "An Introductory Approach to System and Payload/Antenna Co-Design for Multi-Beam Broadband Satellite Systems", Proc. 29th ESA Antenna Workshop, Noordwijk (The Netherlands), April 2007.

[3] P. Gabellini, N. Gatti, G. Gallinaro, G. Vecchi, F. Della Croce, R. Rinaldo, P. Angeletti, "Proposed Architecture of a System and Antenna/Payload Co-Design Tool for Multi-Beam Broadband Satellite Systems", ESA Workshop on Advanced Flexible Telecom Payloads, Noordwijk, Netherlands, 18-20 Nov. 2008.

[4] P. Angeletti, D. Fernandez Prim and R Rinaldo, "Beam Hopping in Multi-Beam Broadband Satellite Systems: System Performance and Payload Architecture Analysis",, Proceedings of the AIAA, San Diego,2006.

[5] J. Anzalchi, A. Couchman, P. Gabellini, G. Gallinaro, L. D'Agristina, N. Alagha, P. Angeletti, "Beam Hopping in Multi-Beam Broadband Satellite Systems: System Simulation and Performance Comparison with Non-Hopped Systems," in the 5th Advanced satellite multimedia systems conference (ASMS) and the 11th signal processing for space communications workshop (SPSC), pp. 248-255, Cagliari, Italy, 13-15 Sept. 2010.

[6] X. Alberti, J. M. Cebrian, A. Del Bianco, Z. Katona, J. Lei, M. A. Vazquez- Castro, A. Zanus, L. Gilbert, and N. Alagha, "System Capacity Optimization in Time and Frequency for Multibeam Multi-media Satellite Systems," in Proc. of the 5th Advanced Satellite Multimedia Systems Conf. and the 11th Signal Processing for Space Commun. Workshop, Cagliari, Italy, pp. 226-233, Sep. 2010.

[7] E.G. Talbi. *Metaheuristics from design to implementation.* John Wiley & Sons, Inc. 2009.

[8] D.H. Wolpert, W.G. Macready., *"No Free Lunch Theorems for Optimization", IEEE Transactions on Evolutionary Computation.* 1997.

[9] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin 1 (6) pp. 80–83. 1945.

Appendix 2

30th AIAA International Communications Satellite Conference (ICSSC), 24-27 September, 2012, Ottawa, Canada

Multi-objective Optimization Approach to Power Allocation in Multibeam Systems

Alexios Aravanis [1]
National Technical University of Athens, 15780 Zografou, Greece
University of Luxembourg, L-1359Kirchberg, Luxembourg

Bhavani Shankar M. R[2] and Gregoire Danoy[3]
University of Luxembourg, L-1359 Kirchberg, Luxembourg

Pantelis-Daniel Arapoglou[4]
National Technical University of Athens, 15780 Zografou, Greece

Panayottis G. Cottis[5]
National Technical University of Athens, 15780 Zografou, Greece

and

Bjorn Ottersten[6]
University of Luxembourg, L-1359 Kirchberg, Luxembourg

Current satellite systems employ multi-beam technology that allows for the dynamic allocation of the system resources (i.e. bandwidth and power), based on changing traffic distribution to provide substantial capacity gains. Often such an allocation tends to utilize the maximum DC power available, whereas a minimization of the same is desirable. This paper deals with the power allocation with respect to two seemingly conflicting objectives, namely the maximization of the available system throughput and the minimization of the system power consumption. A multi-objective

---

[1] Student, School of Electrical and Computer Engineering, Iroon Polytechniou 9, non-member.
[2] Research Associate, Interdisciplinary Centre for Security Reliability and Trust (SnT), Campus Kirchberg, 6, Rue Richard-Coudenhove Kalergi, non-member.
[3] Research Associate, Computer Science and Communications Research Unit (CSC), Campus Kirchberg, 6, Rue Richard-Coudenhove Kalergi, non-member.
[4] Research Associate, School of Electrical and Computer Engineering, Division of Information Transmission Systems and Material Technology, Iroon Polytechniou 9, non-member.
[5] Professor, School of Electrical and Computer Engineering, Division of Information Transmission Systems and Material Technology, Iroon Polytechniou 9, non-member.
[6] Director, Interdisciplinary Centre for Security Reliability and Trust (SnT), Campus Kirchberg, 6, Rue Richard-Coudenhove Kalergi, non-member.

optimization approach is proposed for power allocation to handle the aforementioned objectives. A state-of-the-art multi-objective genetic algorithm (MOGA), i.e. the non-dominated sorting genetic algorithm II (NSGA-II), is used as optimization approach. The set of obtained solutions is presented in the form of a Pareto front, which provides complete information to the user concerning the trade-off between the traffic demand and power requirements.

Nomenclature

N	=	number of beams
N_c	=	number of carriers
B_{TOT}	=	total bandwidth
P_{TOT}	=	total power
P_b	=	beam power
$P_{b,c}$	=	carrier power
B	=	carrier bandwidth
$P_{b,con}$	=	beam power constraint
$R_{b,req}$	=	bit rate requested by user in beam b
$R_{b,off}$	=	offered cumulative bit rate of all N_c carriers in beam b

I. Introduction

THE advent of new applications, including fixed and mobile multimedia interactive services, necessitate that satellite systems become more flexible. Design flexibility is essential to ensure that satellite systems can deal with emerging traffic requirements and unexpected market changes or the change in orbital location during the lifespan of the spacecraft. Flexibility could be viewed in four aspects with regard to coverage, power, frequency planning and higher layer functionalities (e.g. routing, switching)[1]. In exploiting these degrees of flexibility, multibeam satellites can play a key role since their design can easily be extended to support reconfigurability of power and frequency plans as well as routing and switching functionalities. This feature of multibeam systems along with the recent advances in Multi-Port Amplifier (MPA) and flexible Traveling Wave Tube Amplifier (TWTA) technology allow for efficient beam power allocation in response to the spatiotemporal variations of traffic demand.

In some previous work, optimization routine to efficiently apportion power across the beams to match the traffic demand has been considered.[5] Such an optimization is considered to be subject to a maximum RF power constraint, which, in turn, assumes a maximum DC power consumption. Since DC power consumption is a critical factor, the possibility of minimizing DC power is also warranted. However efforts toward modifying the DC/ RF power consumption have not been explored in detail in literature. Therefore, the present paper explores the power allocation problem with respect to both seemingly conflicting objectives: maximization of the available system throughput and minimization of the system power consumption, subject to the aforementioned constraints.

In order to handle these objectives, the appropriate cost (or objective) functions to quantify the performance of each objective have to be selected. In the case of the system power consumption this function is just the overall system power consumption, but in the case of the available system throughput four different cost functions are found in literature, namely the differential and unmet system capacities[7], satisfaction factor[8] and aggregate fitness. For the identification of the most suitable cost function a preliminary study, a comparison of the four functions is conducted. Following the results of Ref. 5, where a systematic study of the appropriate optimization techniques for the maximization of the available throughput is performed, a Genetic Algorithm (GA)[3] is adopted in the preliminary study. The results of the four cost functions employing the GA are then presented and compared and the best performing cost function is selected. Then, in order to enhance the system utilization, this cost function is used together with the minimization of the overall system power consumption as cost functions of a multi-objective optimization problem (MOP).

A MOP deals with the optimization of multiple conflicting objective functions. In such MOPs the order relation between solutions relies on the concept of dominance. A solution dominates another if it is strictly better in one objective and better or equal with regard to all other objectives. The optimal solution then, is not a single solution like in single-objective optimization problems, but a set of non-dominated solutions defined as *Pareto optimal solutions*. This set of solutions represents the compromise between the conflicting objectives and allow for the selection of the best solution, according to the preferences of the decision maker and the priority of each objective.

Subsequently a state-of-the-art multi-objective genetic algorithm (MOGA), the non-dominated sorting genetic algorithm II (NSGA-II)[3] is employed for the resolution of this MOP. The set of solutions obtained is then presented in the form of a Pareto front (i.e. the image of the Pareto solutions in the objective space), providing complete information concerning the trade-off between the traffic demand and power requirements. The results are then analysed, demonstrating the potential benefits that arise from the use of the multi-objective approach.

The remainder of this paper is organized as follows. Section 2 presents the formulation of the resource optimization problem for multibeam architecture as a MOP. In the same section the optimization techniques applied to the problem in previous studies are presented and compared with the approach proposed in the present paper. Then the considered optimization techniques (GA for the preliminary study and NSGA-II for the multi-objective optimization) are described in section 3. The preliminary study and the experimental results using the proposed multi-objective approach are presented and analysed in section 4. Finally section 5 contains our conclusions and perspectives.

II. Resource Optimization in Multibeam Architecture

Multibeam Architecture and System Resources

Assume a multibeam satellite system with b = 1,…,N beams, employing a typical four colour reuse pattern (for e.g., two colours in frequency and two in polarization), where the available total downlink bandwidth of the system B_{TOT} is equally distributed among the four colours. It is further assumed that the available bandwidth reused over the different colours is equally divided into the beams of each colour and each of these beams accommodates four carriers. A carrier represents the elementary system entity for conveying different streams of information and the bandwidth of each beam is equally divided among its carriers. The described layout could be visualized in Fig. 1, where the beam layout adopted for this paper's simulations is depicted.

Having allocated B_{TOT} equally among the carriers, the resource allocation pertains to the appropriate allocation of the total available system power $P_{TOT.}$ P_{TOT} is a function of the platform total DC power on board the satellite, and must be allocated appropriately to each beam, so that the offered beam bit rate meets the user requirements in the same beam. The power allocated to each beam P_b, will then be equally divided among the N_c carriers of the beam $P_{b,c} = P_b / N_c$. Furthermore minimizing the overall power consumption while allocating the available P_{TOT} is also warranted.

B. Resource Optimization

Following this analysis the resource optimization problem is formulated as a MOP. The target of this optimization is to determine the beam power P_b, for which:

1. The system capacity is maximized (cost function $f_{1)}$.
2. The overall power consumption is minimized (cost function f_2).

Subject to:

- The power constrains of the system
- The power constrains of each beam

For the system capacity maximization the following four candidate cost functions have been employed in past studies:

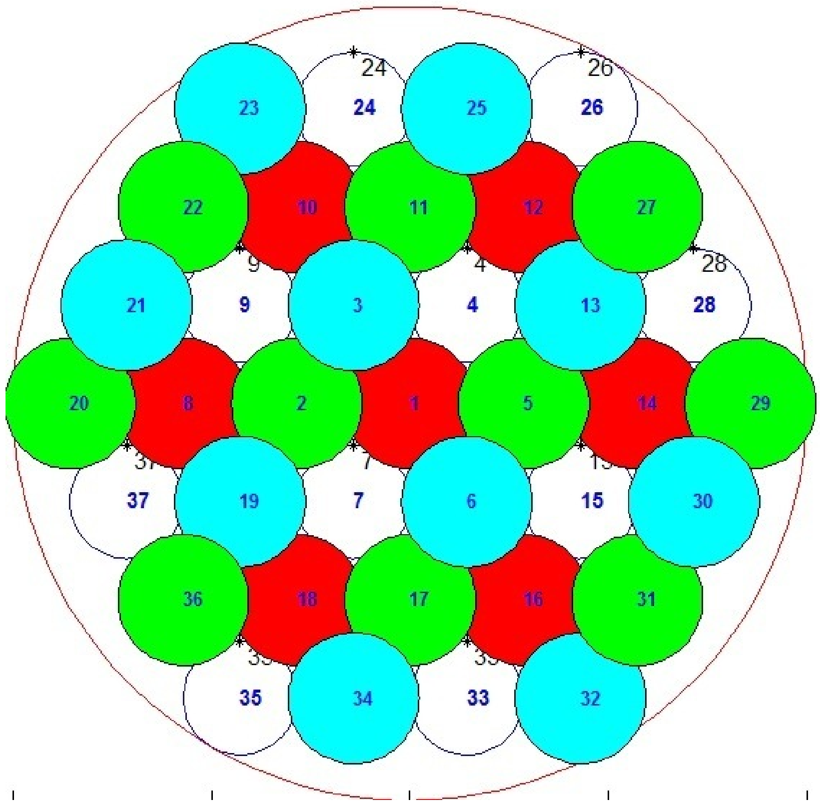


Figure. 1 Four colour reuse pattern, for a 37 circular beam layout

Differential System Capacity (DSC)

$$\text{Min } f_1 = \sum_{b=1}^{N} \text{abs}\left[R_{b,req} - R_{b,off}\right] \tag{1}$$

According to Eq. (1), offered beam capacities exceeding or being less than the required beam capacities, both contribute to the figure of merit and drive the optimization closer to the requirements.

Unmet System Capacity (USC)

$$\text{Min } f_1 = \sum_{b=1}^{N} \max\left[R_{b,req} - R_{b,off}, 0\right], \tag{2}$$

In Eq. (2), offered beam capacities exceeding the required do not contribute to the figure of merit and do not affect the optimization.

Satisfaction Factor (SF)

$$\text{Max } f_1 = \frac{\sum_{b=1}^{N} \min\left[R_{b,req}, R_{b,off}\right]}{\sum_{b=1}^{N} R_{b,req},} \tag{3}$$

The offered beam capacities exceeding the required capacity do not contribute to the figure of merit defined in Eq. (3). But in this case the recessive beam capacities are scaled by the cumulative required capacity.

Aggregate Fitness (AF)

$$\text{Max } f_1 = \sum_{b=1}^{N} \frac{1}{1+\text{abs}\left[R_{b,req} - R_{b,off}\right]} \tag{4}$$

According to Eq. (4), offered beam capacities exceeding or being less than the required beam capacities, both contribute to the figure of merit, like in the case of the DSC, but these results are scaled down on a beam bases and the optimization is driven according to the sum of the individual beam fitness.

For the power consumption minimization the cost function employed is the following:

$$\text{Min } f_2 = \sum_{b=1}^{N}[P_b] \tag{5}$$

And the optimization is subject to the following constrains:

System power constraint

$$\sum_{b=1}^{N}[P_b] \leq P_{TOT} \tag{6}$$

Beam power constraints:

$$P_b \leq P_{b,con} \tag{7}$$

The offered, bit rate $R_{b,off}$ is calculated as follows:

$$R_{b,off} = \sum_{c=1}^{N_c} B \times f_{DVB-S2}(SNIR_c) \tag{8}$$

The function $f_{DVB-S2}(SNIR_c)$ quantifies the spectral efficiency of the various modulation and coding schemes employed by DVB-S2 as a function of the $SNIR_c$ of carrier c.

1. SNIR Calculation

The $SNIR_c$ of each carrier c, transmitted with power $P_{b,c}$ and bandwidth B is calculated as follows[4]:

$$SNIR_c = \frac{a_b^2 P_{b,c}(OBO)}{N_0(a_b)B + \sum_{q\in\Phi} \alpha_q^2 P_{q,c}(OBO) + I_{adj_{ch}}(B,XPD) + I_{adj_{sat}} + I_{inter}(OBO,C_b,Mod)} \tag{9}$$

Φ: is the set of co-channel beams in the coverage area with active carriers overlapping with the bandwidth of the intended carrier c (co-channel interference into intended beam b)
a_q: is a gain factor encompassing the effect of: satellite antenna beam q, terminal receive antenna gain, free space loss, clear sky attenuation and rain attenuation
N_0: is the noise power spectral density which is a function of a_b because of the increase in noise temperature under rain fading conditions
I_{adj_ch}: accounts for adjacent channel interference due to filter imperfections (function of B), including spillover from the beams in orthogonal polarization if both polarizations are employed (function of XPD)
I_{adj_sat}: Inter-system interference caused by adjacent satellites
I_{inter}: Intermodulation interference
The notation $P_{b,c}(OBO)$ denotes the dependence on the appropriate OBO according to the modulation scheme employed

C. Single-objective Transformation of the Considered MOP

An early approach in the direction of modifying the DC/RF power consumption along with the available system throughput is proposed in Ref. 6. In this approach a number of cost functions to quantify the satisfaction of the traffic demand are proposed. These cost functions include a DC consumption parameter, in order to take the power consumption into account, while maximizing the available throughput of the system. This technique known as *scalar approach* is transforming a MOP into a single-objective problem, by combining different objective functions (i.e. the traffic demand satisfaction and the DC consumption minimization) into a single-objective function in a linear way.

This type of approach however, requires a priori knowledge on the considered problem, in order to find the appropriate weights for the linear representation of the different objective functions. That is because the Pareto optimal solution is the node of the objective space hyperplane, defined by the weight vector and the feasible space of the problem. Thus the Pareto border has to be known in order to select the corresponding weights to obtain the desirable optimal solution.

A common misconception, is that the weights selected for the linearization of the MOP are representative of the relative importance of the different objectives. In these cases the objective space hyperplane defined by these weights, could intersect randomly with the feasible space of the problem, providing undesirable solutions.

Alternatively multiple weights can be used, that will be optimized in parallel along with the problem. However the computational cost of this process is significant and even this process cannot guarantee to provide all Pareto optimal solutions, since these are limited to the convex hull of the Pareto front (i.e. solutions in concave regions of the Pareto front cannot be found).

Furthermore the provided optimal solution of these scalar approaches is a single solution unlike the set of Pareto solutions, provided by the dominance based approaches, depriving the decision maker of the alternative solutions, suppressing useful information regarding the system's objective space.

The considered problem however is not known a priori. Hence, the present paper explores the power allocation problem as a strictly multi-objective problem (MOP) and not as a linear function of the various objectives. Formulation of the problem in a dominance based multi-objective setting seems to be novel in literature and in this course a state-of-the-art multi-objective genetic algorithm (MOGA), the non-dominated sorting genetic algorithm II (NSGA-II)[3],presented next, is employed for its resolution.

III. Optimization Techniques

A. Genetic Algorithm (GA)

The GA adopted in the preliminary study of the paper, is a population based metaheuristic, where the search for a "good" solution starts from an initial population of acceptable solutions. This technique is based on the Darwinian evolutionary model, which consists in the following iterative process. Two parents are first selected from the whole population with a given selection criterion. Then two genetic operators are sequentially applied with some probability, namely recombination (or crossover) that exchanges portions of the parents solution vectors, and mutation that randomly modifies the individual. Finally the generated offsprings are evaluated and inserted back into the population following a given criterion.

B. Non-dominated Sorting Genetic Algorithm II (NSGA II)

NSGA-II is the most popular and referenced dominance based multi-objective algorithm in literature. It is a GA with a non-structured population, that is used to obtain the new population after applying the typical genetic operators (i.e. selection, crossover, mutation). As a GA belongs to the metaheuristic techniques, and since metaheuristics do not guarantee the optimality of the obtained solution the goal of this technique is to provide an approximation of the Pareto optimal set with two necessary properties: convergence to the Pareto optimal solutions and uniform diversity (i.e. a good distribution of the obtained solutions around the Pareto optimal front).

In this course, all individuals of the initial population are ranked with respect to both convergence and diversity. All non-dominated solutions are assigned to the rank 1 and are subsequently removed from the population in order to assign the current non-dominated solutions to the rank 2. This process is iterated for all individuals and then individuals within the same rank are ranked according to the distance from their surrounding individuals. Individuals with higher distances (i.e. good distribution) are ranked higher than others.

The individuals are then sorted according to their rank and the typical genetic operators are applied iteratively.

IV. Experimental Results

For the simulations presented herein, the 37-beam system described in Fig. 1 has been simulated, where the antenna pattern is approximated by employing the Bessel function and the link budget is calculated assuming one user per beam, located at the beam edge (worst case position). The parameters of the link budget, are presented in Table 1.

Table 1 Link Budget Parameters

Link Budget Parameters For Capacity Results	
Parameter	Value
Frequency Band	Ku
User Link Bandwidth $\mathbf{B_u}$	46.875 MHz
HPA saturation Power $\mathbf{P_\tau}$	80 W
Max satellite antenna gain $\mathbf{G_T}$	52 dBi
Output Back Off **OBO**	5 dB
Satellite **EIRP**	66 dBW
Free Space Loss **L**	212 dB
Terminal Antenna Gain $\mathbf{G_R}$	41.7 dBW
Terminal noise Temperature **T**	207 K
Receive **C/N**	20.2 dB
External $\mathbf{(C/I)_{EXT}}$	30.0 dB

A. Problem Instance Setup

Simulating the above system, the cost functions defined in Section 2 are compared, based on their performance concerning the maximization of the available system throughput. The technique employed for the comparison of the cost functions is the most appropriate technique out of the standard metaheuristics techniques compared in Ref. 5, namely the GA. The solutions are encoded as 37-dimensional vectors, where each dimension represents the R/F power of the respective beam. The optimization parameters selected for the simulations are shown in Table. 2.

Table 2. GA Optimization Parameters

Parameter	Value
Population Initialisation	Random
Population Size	5856
Crossover Function Crossover Rate	Uniform 0.95
Mutation Function Mutation Rate	Uniform 0.05
Selection Function	Tournament
Elite Individuals	30
Fitness Evaluations	234240

B. Single Objective Results (Preliminary Study)

The performance assessment of the different cost functions is shown in Table. 3. The results reported involve averaging over 15 independent runs, as the computational time needed for 234240 evaluations did not allow the use of a bigger sample. The statistical confidence in the comparisons is assessed by performing the Wilcoxon test[2], with standard confidence level (0.95).

Table 3. Cost Function Performance Assessment

Cost Function	**Capacity [Gbps]**			**Power [W]**
	(Average of 15 runs)			
	Required	Offered	Unmet	
DSC	24.155	23.588	0.801	1758.2
USC	24.155	25.605	0.746	2145.7
SF	24.155	23.946	0.748	2010.4
AF	24.155	23.547	0.842	1752.5

	(Overall Best over 15 runs)			
DSC	24.155	23.593	0.778	1743.9
USC	24.155	24.879	0.709	2030.3
SF	24.155	24.879	0.709	2030.3
AF	24.155	23.588	0.802	1777.9

The numerical results presented above demonstrate the suitability of the USC as a cost function, since the USC provides the lower unmet capacity of all. This result is not surprising since USC focuses on the minimization of the unmet capacity, ignoring the impact of the exceeding beam capacities on the optimization. Therefore the performance of the SF, which also ignores the impact of the exceeding beam capacities is similar, with that of USC, obtaining also the same overall best solution. The performance of the best USC result in the satisfaction of the traffic demand is depicted in Fig. 2.

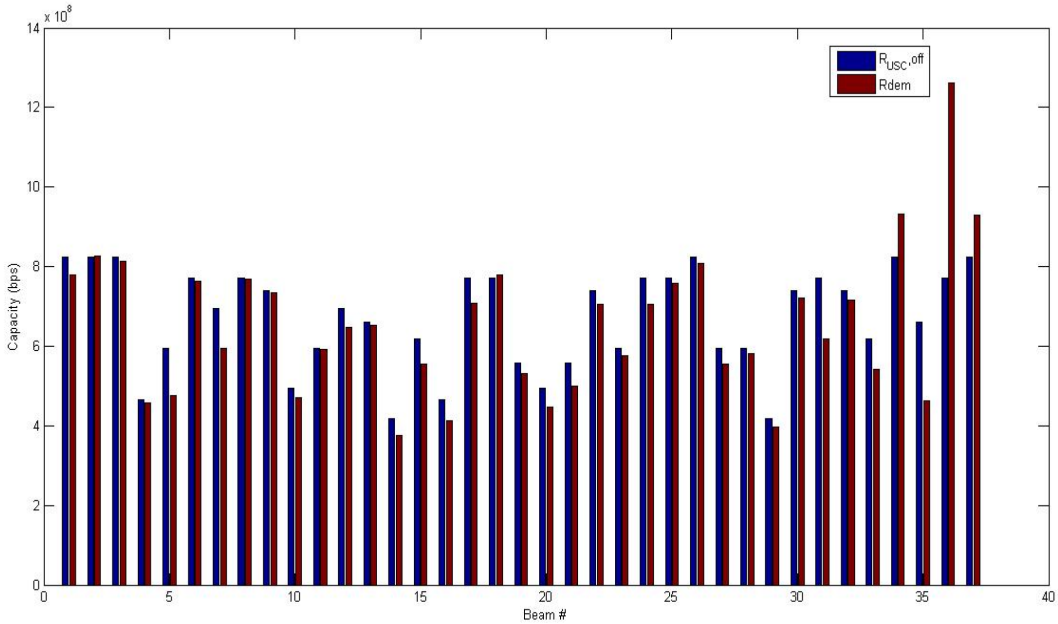


Figure.2 Single objective optimization performance (USC)

C. Multi-objective optimization results

The result presented in Fig.2 is the best result provided for the maximization of the system throughput. This can be used as initial seed for the multi-objective approach ensuring that the result providing the best performance in terms of the capacity maximization will be included in the Pareto optimal set of solutions and possibly will be further enhanced.

Subsequently the obtained solution is used with the same encoding described above as an individual of the initial population of the NSGA-II. The target of the optimization employing the NSGA-II is to determine the power P_b for which the Eq. (2) and (5), are minimized. The results reported from this optimization involve averaging of 15 independent runs and the optimization parameters selected are shown in Table. 4.

Table.4 NSGA-II Optimization Parameters

Parameter	Value
Population Initialisation	Random
Population Size	6000
Crossover Function	Uniform
Crossover Rate	0.9

Mutation Function Mutation Rate	Uniform 0.05
Selection Function	Tournament
Fitness Evaluations	240000

The Pareto front obtained from the 15 runs is depicted in Fig. 3 providing complete information concerning the trade-off between offered capacity and power requirements, whereas the initial population provided by the single-objective approach is further enhanced.

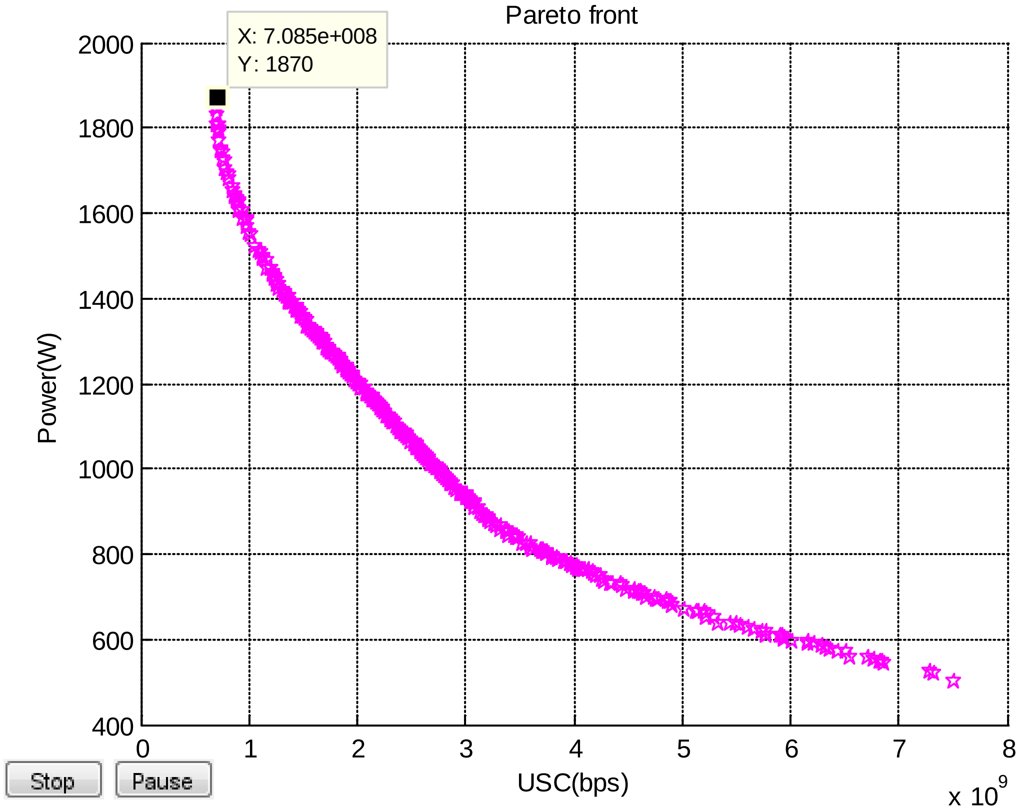


Figure. 3. Pareto front encompassing the enhanced result of the single-objective optimization.

From the above figure, it appears that the obtained Pareto front provides a large number of non-dominated solutions as well as a very good diversity. It is also evident that the result of the single-objective optimization, shown in Table 3 has improved significantly in terms of Power. In particular the allocation obtained by the single-objective approach, consuming 2030.3W for an USC of 0.709Gbps dropped to a power consumption of 1870W for the same USC. In other words the multi-objective approach allows for the system to perform the same in terms of satisfying the traffic demand while saving 7,9% power.

Furthermore the slope of the left part of the Pareto front is steep enough, to allow for even more significant power savings, in case the user is willing to let the USC deteriorate by little. Fig. 4 shows the left part of the Pareto front, demonstrating the trade-off between the satisfaction of the traffic demand and the power requirements. In this case the user can save up to 12% power by decreasing the USC by 2.7%. It is evident that the Pareto front obtained provides complete information regarding the trade-off between the two objectives, enabling the decision maker (in this point of view, the satellite operator) to adjust to the emerging traffic requirements according to the priority of the two objectives.

V. Conclusion

The present paper demonstrates how current multibeam satellite systems can benefit from the use of multi-objective approaches for the allocation of their resources, which is novel approach in literature. The interesting aspect is that the two objectives are seemingly conflicting: maximization of the available system throughput and the minimization of the system power consumption due to their relation to the power and yet the proposed technique can improve one objective by 7,9%, without deteriorating the other. Moreover complete information to the satellite operator regarding the trade-off of the different objectives is provided, enabling a full exploitation of the available flexibility of the system yielding significant capacity gains and power efficiency, which results to the prolongation of the spacecraft's life expectancy.

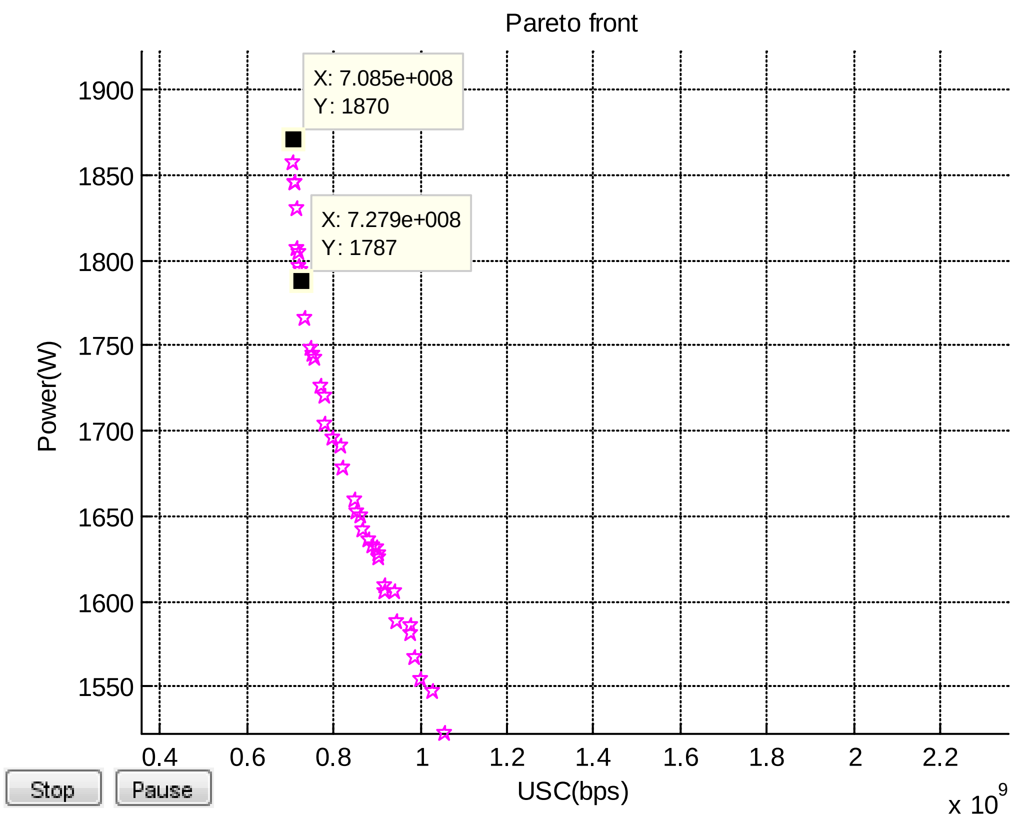


Figure.4 Pareto front, zoom in

References

[1]Balty C., Gayrard J-D., Agnieray P., "Communication Satellites to enter new age of flexibility," *Acta Astronautica*, Vol. 65, Apr. 2009, pp. 75-81.

[2]Wilcoxon F., "Individual comparisons by ranking methods," *Biometrics Bulletin* Vol. 1, No.6, 1945, pp. 80–83.

[3]Talbi E.G., *Metaheuristics from design to implementation*, John Wiley & Sons, Inc., New Jersey, 2009.

[4]Bousquet M., Maral G., *Satellite Communications Systems*, John Wiley & Sons, Inc., United Kingdom, 2009.

[5]Aravanis A., Shankar B., Danoy G., Arapoglou D., Cottis P., Ottersten B., "Power allocation in Multibeam satellites - A hybrid-Genetic Algorithm approach," *2nd ESA Workshop on Advanced Flexible Telecom Payloads*, 17-19 April 2012.

[6]Gabellini P., D'Agristina L., Alagha N., Angeletti P., "On the Cost Function for the Performance Optimization of Multi-Beam Broadband Satellite Systems," *2nd ESA Workshop on Advanced Flexible Telecom Payloads*, 17-19 April 2012

[7]Gabellini P., Gatti N., Gallinaro G., Vecchi G., Della Croce F., Rinaldo R., Angeletti P., "Proposed Architecture of a System and Antenna/Payload Co-Design Tool for Multi-Beam Broadband Satellite Systems", *ESA Workshop on Advanced Flexible Telecom Payloads*, 18-20 Nov. 2008.

[8]Alberti X., Cebrian, J. M., Del Bianco A., Katona Z., Lei J., Vasquez-Castro M. A., Zanus A., Gilbert L., Alagha N., "System Capacity Optimization in Time and Frequency for Multibeam Multi-media Satellite Systems", *5th Advanced Satellite Multimedia Systems Conference*, Sep. 2010.

Appendix 3

2nd ESA Workshop on Advanced Flexible Telecom Payloads, 17-19 April 2012, Noordwijk, The Netherlands

Operational Optimization of a Ku Multibeam Flexible Payload

2nd ESA Workshop on Advanced Flexible Telecom Payloads

17-19 April 2012

ESA/ESTEC
Noordwijk, The Netherlands

Nathalie Metzger(1), Coline Morel(1), Maria Zapata(1), Stéphane Mourgues (1), Nicolas Jacquey(1), Vincent Tugend(1), Dominique Galinier(1), Francesc Coromina(2), Christophe Loeillet(3), Bhavani Shankar(4)

(1) Astrium Satellites
31 Avenue des Cosmonautes - 31402 Toulouse Cedex 4 - France
nathalie.metzger@astrium.eads.net

(2) ESTEC
Keplerlaan 1 – 2201 AZ Noordwijk – The Netherlands
Francesc.Coromina@esa.int

(3) SES-ASTRA
L-6815 Chateau de Betzdorf – Luxemburg
Critsophe.Loeillet@ses.com

(4) University of Luxembourg
162a, Avenue de la Faiencerie L-1511 LUXEMBOURG
Bhavani.Shankar@uni.lu

INTRODUCTION

The use of multibeam systems for broadband services in Ka-band has shown a significant increase of the satellite capacity, due to beam size reduction and an increase in frequency re-use. Such a concept can be used in Ku-band to increase the capacity of systems in regions with high rain rate, where the use of Ka-band is difficult. Ku-band ground equipments are readily available and a large quantity of consumer DTH receivers, as well as some interactive terminals employing mature Ku-band technology, has been installed.

This paper presents the work performed for the ESA study “Techniques and technologies for multi-spot beam Ku-band Satellite networks”.
The advantages brought by the introduction of multi-beams and associated flexibility on Ku-band payloads are assessed, together with an investigation on how system and payload flexibility can help satellite operators to address the variation of service demand over the lifetime of a telecommunications satellite, e.g. evolving markets area, increased market penetration, variability between services, rapid evolution of services.

Realistic system and flexible payload architectures have been defined for a Ku multibeam mission proposed by SES.
Different payload architectures are proposed using flexible payload equipments chosen according to the considered study medium term timeframe, among them Flex-MPM, flexible input section equipment, MPAs and transparent processors.
A major goal of the study is also the development of system performance evaluation software centred on multi-dimensional link budget calculations and allowing the optimization of the payload configuration based on power and bandwidth allocation algorithms. The tool is based on Astrium’s closely connected background system and payload tools, to which an optimizer is added, helping to define the best flexible payload configuration meeting the capacity demand.

MISSION AND SYSTEM REQUIREMENTS

The mission has been defined by SES. The requirements are to be able to embark on the same spacecraft both a Ku multibeam mission over Africa and another DTH mission covering Europe. The multibeam mission is expected to use roughly half of the satellite resources.
The main objective of multibeam proposed mission scenario is the delivery of broadband Internet service to residential and professional users in Ku-band over the African and Middle Eastern area.
Fig.1 gives the forward frequency plan, with Ka-band gateways and for user downlink a "nominal" 4-color frequency plan with 500 MHz per user beam, with a possible "extended" frequency plan allowing reaching up to 1 GHz per user beam.

Traffic demand scenarios have been defined in terms of capacity, with their evolution with time for both forward and return links. Several models of evolution with time have been considered, linked to the evolution of the population density, of the demand per market segment or of the market penetration.
Fig. 2 gives an example of forward traffic demand distribution for 2019.

In order to evaluate the potential benefits of a given architecture, compared to another, a figure of merit has been defined. The figure of merit is "a parameter used to characterize the performance of a system". It has to be a combination of measurable quantities. The system performance shall be, as far as possible, characterized from a commercial point of view. In this study, the proposed figure of merit is a quantity proportional to the operator financial return on the space segment investment.

FLEXIBLE PAYLOAD ARCHITECTURES

The beam layout has been defined considering the traffic demand geographical distribution as starting point. A strong mission requirement was also to be able to cover as much as possible the "white areas" of lower traffic demand.
The beam sizes and number has been chosen to fulfill the constraints to accommodate the multibeam mission on half of a Eurostar 3000 spacecraft, in terms of available mass, power, dissipation and possible number of equipment. Fig. 3 gives the defined layout.
Hot spots are identified with the highest capacity demand.

A first optimization by design has been made to choose the best configuration of TWTs power and bandwidth per beam, to match as much as possible the maximum traffic demand of all the defined scenarios. This optimization process was not extensive as the idea was to use flexible equipment, so that the payload can cope with an evolving traffic evolution.

Architectures with two beams per TWT are considered for most of the beams, in order to decrease the number of required TWTs.

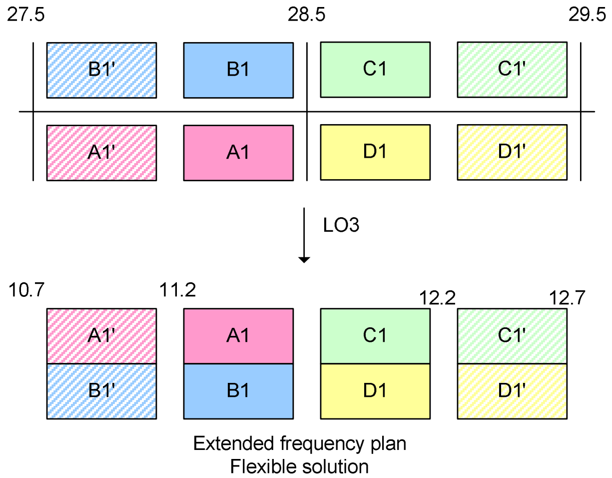


Fig. 1. Forward link frequency plan

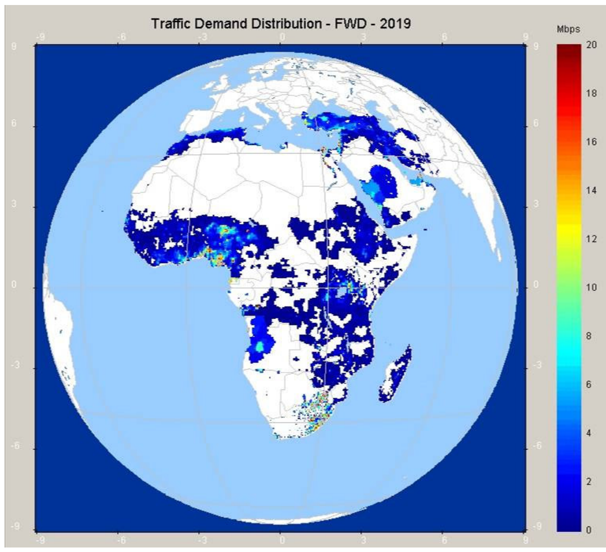


Fig. 2. Forward traffic demand distribution for 2019

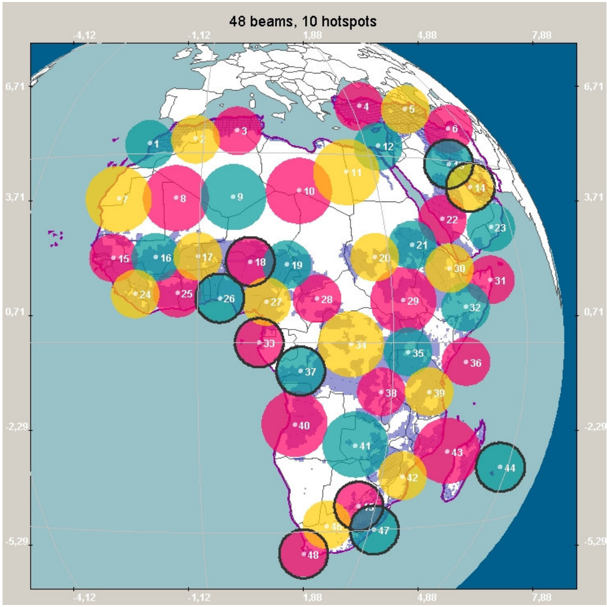


Fig. 3. Beam layout

The following system and payload configurations have been worked:

- a conventional non-flexible architecture with:
  - 48 beam coverage
  - fixed bandwidth allocated per beam and fixed power: between 54 MHz and 1 GHz to be allocated per beam according to the capacity demand
- 3 flexible architectures with:
  - 48 beam coverage and different flexibility options
  - flexibility to start with 6 gateways and a restricted frequency plan (500 MHz max per beam) and to deploy more gateways to reach an extended frequency plan with more bandwidth allocated to each beam (1 GHz max per beam)
  - flexible allocation of bandwidth and power to beam, using Flex-MPMs or MPAs
- an architecture enabling a very high level of flexibility in the allocation of gateways to beams, using a processor on a sub-set of beams.

Fig. 4 gives one flexible payload architecture, based on Flex-MPMs for power management, and flexible input section to save gateway spectrum by adaptation to the demand requirement.

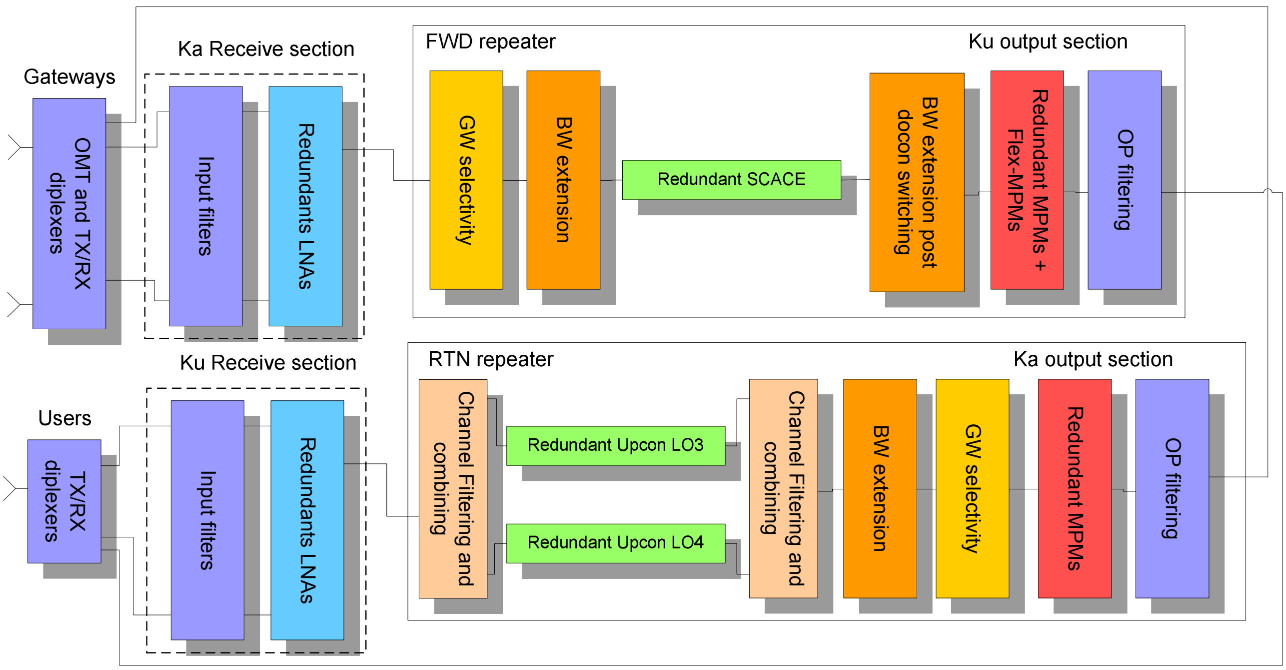


Fig. 4. Flexible architecture with Flex-MPMs and SCACEs

Flex-MPM are used to adapt the power to the used spectrum, in order to get the EIRP/MHz required to match the availability requirement and optimize the link efficiency.
The Flex-MPM models of Tesat-Spacecom GmbH & Co. KG, Backnang, Germany, are used, giving an optimized efficiency for a given saturation point and back-off.
The idea is to be able to save some DC power on beams enabling by design more EIRP/MHz than required, and to use this unused DC power to increase the RF power for beams having a lack in capacity and an RF output power margin. Because of their improved efficiency when used in back-off, the Flex-MPM enable to save more power than classical MPMs.

The flexible input section is made with SCACE, which are flexible down converters enabling to change the frequency conversion and including selectable filtering. These equipment are modified versions for broadband applications of the first SCACE equipment already flying.
Fig.5 gives the principle of the SCACE. One SCACE down converts two beams (beams 39 and 7 for SCACE 1 in Fig.5) of the gateway uplink and the total bandwidth of the two beams can be selected according to the capacity demand, the SCACE enables a filtering bandwidth choice matched to the used bandwidth.

One flexible payload is proposed with 8x8 MPAs and allows sharing power between 16 beams. This offers a very high level of flexibility and enables to handle very heterogeneous demand.

For the processed architecture, a subset of the beams and a part of their bandwidth is managed by a digital processor for a very flexible gateway to beams bandwidth allocation.
A transparent processor could be used for two different functions:

1. for a mesh connectivity on a given number of beams and bandwidth,
2. to process a part of each beam bandwidth.

The mesh connectivity is a service that was not expected by the mission requirements, and not covered by the other proposed flexible payload architectures, then could not be part of the comparison with the different payloads.
It has then been decided to implement option 2, on the forward link only, and for 12 beams.
Four beams of 3 gateways will have a processed bandwidth. Fig.6 gives the processed bandwidth location in the user beams. Each processed part of the beam can come from any gateway processed bandwidth.

TOOL DEFINITION AND DEVELOPMENT

The developed system performance simulation tool allows analysing the system and payload performance by computing the defined Figure Of Merit and performance indicators. Fig.7 shows a functional block diagram of the tool.

The performance simulation tool includes the following input parameters:

- system parameters and constraints e.g. number of beams, available bandwidth, carrier bandwidth granularity, service availability constraints, platform power and dissipation constraints, …
- payload architecture and corresponding repeater and antenna parameters and constraints e.g. RF output power per beam constraints, frequency allocation flexibility constraints, feeder/user link connectivity constraints, mapping beam/TWTs …
- link availability constraints in term of percentage of time,
- requested capacity over the coverage for all the beams.

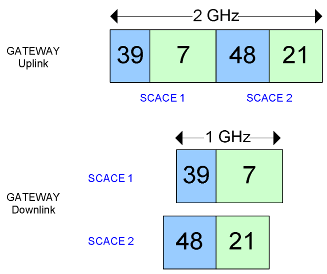


Fig. 5. Example of SCACE use

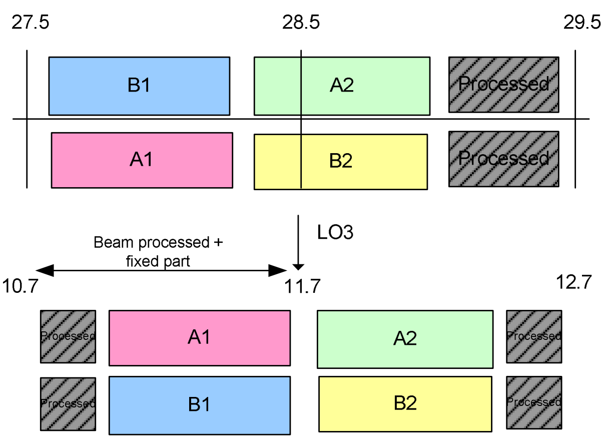


Fig. 6. Processed bandwidth in user beams

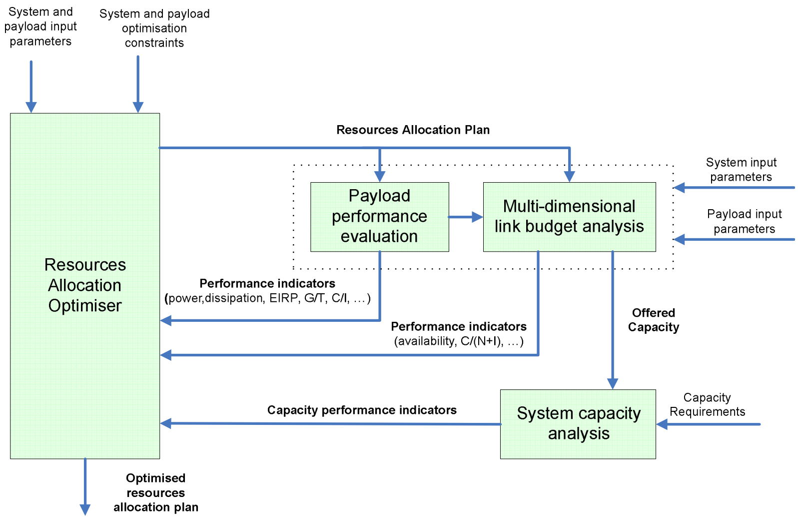


Fig.7 System performance simulation tool functional block diagram

The software model evaluates for a given resources allocation plan:

- RF payload performance over the specified coverage (EIRP, G/T, C/Ico-channel, C/Icross-pol, …) by means of modelling of relevant payload sub-systems,
- payload DC power/dissipation for a given resources allocation plan over the specified coverage,
- multi-dimensional link budgets in order to evaluate system and payload offered capacity and actual served capacity for given traffic; link availability constraints are also checked at that level, taking into account atmospheric propagation aspects.

The optimisation engine is used to optimise bandwidth and power resources against the following defined cost function.

$$\text{Unmet System Capacity (USC)} = \sum_{b=1}^{N_b}\left[R_{b,req} - R_{b,off}\right]^2 \qquad (1)$$

Where $R_{b,req}$ is the bit rate requested by the users in beam b and $R_{b,off}$ is the bit rate offered by the system in beam b according to the output of the optimization process.
$R_{b,req}$ is the output of the traffic model assuming a certain distribution with statistical parameters that depend on: the size of the beam, the population of the beam, the type of terminals within the beam, busy hours and the GDP per inhabitant.

$R_{b,off} = \sum_{c=1}^{C_b} R_{b,c,off}$ corresponds to the cumulative bit rate of all carriers in beam b.

The optimization problem is a Mixed Discrete-Continuous Variable Optimization. Several approaches have been evaluated and a Genetic Algorithm method has been chosen for implementation. This method is described in [1].

PAYLOAD SIMULATIONS AND TRADE-OFFS

Tab.1 gives the main characteristics of the simulated payloads.

	Gateway pattern	**Filter**	**Power allocation**	**Use of SCACE**	**Use of MPA**
Reference	Extended	Standard (1GHz)	Standard with MPM		
Flexible 1	Flexible	Standard (1GHz)	Standard with MPM		
Flexible 2	Flexible	Flexible with SCACE	Flexible with Flex-MPM	X	
Flexible 3	Flexible	Flexible with SCACE	Flexible with MPA	X	X

Tab. 1. Simulated payloads main characteristics

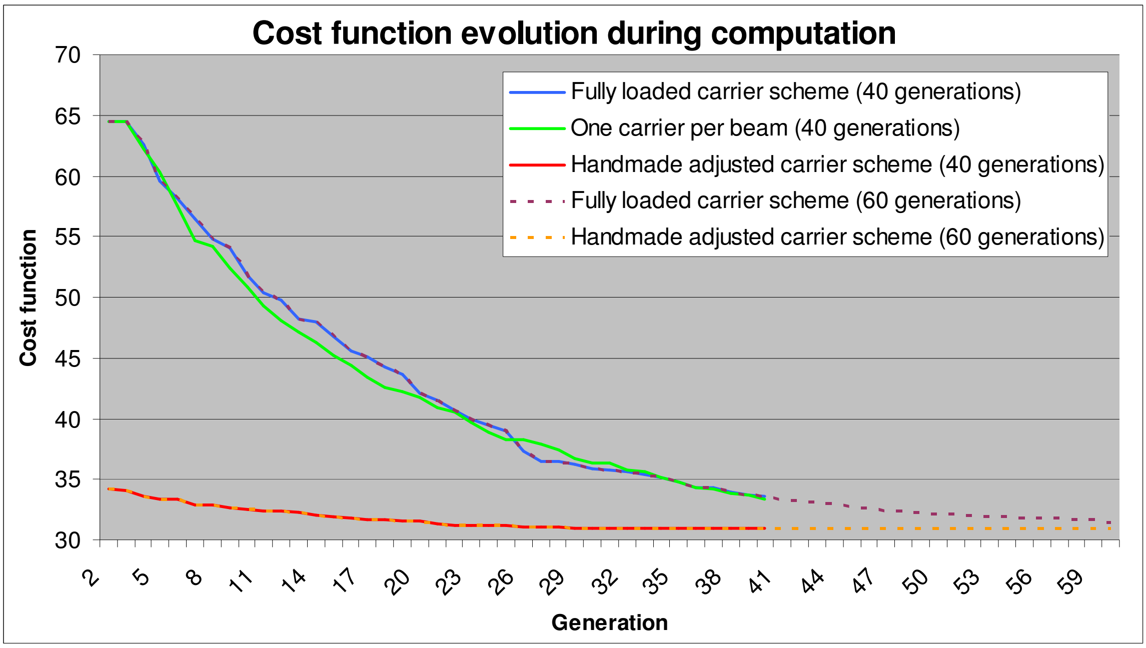


Fig. 8. Computation of cost function for different initial conditions

Fig. 8 gives an example of the optimizer functioning and test:

- the "handmade adjusted carrier scheme" takes, as initial conditions to the optimizer, the results of a basic optimization algorithm,
- the scenario one carrier per beam takes, as initial conditions to the optimizer, one carrier in each beam, the fully loaded one takes the maximum number of carriers per beam,
- 40 (or 60) generations scenarios consider 40 (or 60) iterations of the genetic algorithm, each one considering a population size of 1000.

We can see that whatever the initial conditions are, the optimizer converges to the same value, and there is an optimum in the number of iterations to reach the best cost function.

Fig. 9 gives an example of optimization results, where it remains still difficult to manage the hot spots.

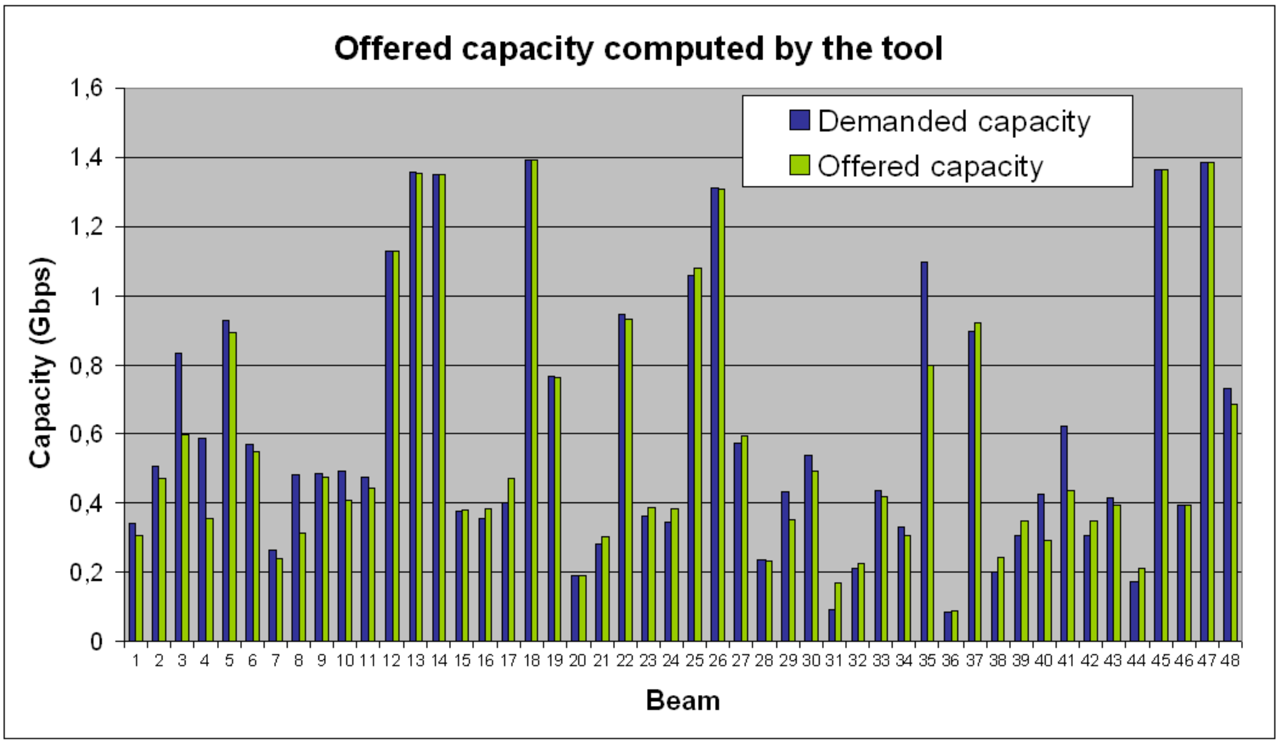


Fig. 9. Optimization results for one flexible architecture and one traffic demand scenario

Several ways to truncate the hot spot capacity demand have been studied together with their influence on the optimization results. The truncation considered in Fig.9 truncates the capacity with the maximum offered capacity of the beam when the system is fully loaded, that means when the whole spectrum allocated to the beam is used.

The simulation software development is finished and under final validation.
The first simulations results show the interest of Flex-MPMs and MPAs architectures, but the final trade-offs are still to be made.
Also, in most cases the best cost function is achieved while the total capacity is not at its maximum value, showing that a big fixed maximum total capacity is not always the answer to meet at best the traffic demand.

CONCLUSION

This paper gives the main outcomes of the ESA study "Techniques and technologies for multi-spot beam Ku-band Satellite networks". The study will end in June 2012, and it remains work for software exploitation.
The simulations results, over several traffic demand scenarios and times, will enable to assess the best flexible payload architecture.
A satellite accommodation of the payload has already been made and will be refined with the chosen best payload.
However, the main goal of the study is already achieved with the system performance evaluation software development, enabling a big step in the optimization of a multibeam flexible payload.
The software, more developed for an operational optimization of an already designed flexible payload, can also help to optimize the design of the payload architecture.

ABBREVIATIONS AND ACRONYMS

MPA	Multiport Amplifier
MPM	Microwave Power Module
Flex-MPM	Flexible Microwave Power Module, MPM with commandable saturated output power
SCACE	Single Channel Agile Converter Equipment

REFERENCES

[1] A. Aravanis, B. Shankar, G. Danoy, D. Arapoglou Pantelis, P. Cottis, B. Ottersten, "Power allocation in Multibeam satellites - A hybrid-Genetic Algorithm approach," *2nd ESA Workshop on Advanced Flexible Telecom Payloads*, 17-19 April 2012.