Performance of Adaptive Satellite Antenna Array Processing and Comparison with Optimal Multi-User Communications

Oliver Lücke

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

Performance of Adaptive Satellite Antenna Array Processing and Comparison with Optimal Multi-User Communications

Oliver Lücke

Vorsitzender des Promotionsausschusses:	Prof. Dr.-Ing. Lindenmeier
1. Berichterstatter:	Prof. Dr.-Ing. K. Tröndle
2. Berichterstatter:	Prof. Dr.-Ing. B. Lankl
3. Berichterstatter:	Prof. Dr.-Ing. E. Lutz

Tag der Prüfung 11.11.2004

Mit der Promotion erlangter akademischer Grad:
Doktor-Ingenieur
(Dr.-Ing.)

Vorwort

Diese Arbeit entstand zum größten Teil in den Jahren 2002/03 während meiner letzten beiden Jahre als wissenschaftlicher Mitarbeiter am Institut für Kommunikation und Navigation des Deutschen Zentrums für Luft- und Raumfahrt e.V. (DLR) in Oberpfaffenhofen.

Mein Dank gilt daher meinem damaligen Abteilungsleiter Prof. Erich Lutz, der mir im letzten Jahr meiner DLR-Tätigkeit den nötigen Freiraum gelassen hat, diese Arbeit im entscheidenden Maß voran zu bringen.

Ausdrücklich bedanken möchte ich mich bei Prof. Tröndle für die Betreuung der Arbeit als meinen Doktor-Vater. Die vielen Stunden der Diskussion mit ihm und seine stetige Ermahnung zur klaren Darstellung (und zum klaren Denken!) haben die vorliegende Arbeit deutlich geprägt. Ebenso bedanken möchte ich mich bei Prof. Lankl und Prof. Lutz dafür, dass sie sich als Berichterstatter für diese Arbeit zur Verfügung gestellt haben.

Von meinen ehemaligen Kollegen beim DLR sind vor allem Matthias Holzbock und Axel Jahn zu nennen, die mir immer wieder Antrieb und Ermutigung gegeben haben. Besonders möchte ich auch Harald Ernst dafür danken, dass er sich oft Zeit für lange Diskussionen genommen hat, auch wenn ich ihn für ”nur 10 Minuten” hatte stören wollen. Allen anderen Kolleginnen und Kollegen aus Spanien, Italien, Frankreich oder Deutschland sei für das gute Arbeitsklima in der Abteilung gedankt.

Ein Dank an Renate und Michael und auch an Bärbel und Heinz, die meine durch schlechtes Vorankommen an dieser Arbeit immer wieder mal getrübte Laune toleriert haben. Meinen Patenkindern Miriam und Leon und natürlich auch Rebekka und Jakob möchte ich sagen, dass ich mich bemühen werde, dass aus ihnen was Rechtes wird (möglichst Dipl.-Ing. natürlich).

Ich möchte mich auch bei Nina bedanken, dass sie es mehr oder weniger ohne zu murren hinnehmen konnte, dass in den letzten Monaten vor der Fertigstellung dieser Arbeit manchmal etwas wichtiger war als sie (die vorliegende Arbeit nämlich).

Meinen Eltern Susanne und Hans-Karl habe ich mit dieser Arbeit bewiesen, denke ich, dass auch ohne Latein oder Altgriechisch eine wissenschaftliche Arbeit gelingen kann. Danke für die stetige Unterstützung!

Oliver Lücke

Gilching 23.12.2004

Contents

Chapter 1

Introduction

Wireless communications systems, either satellite based or terrestrial networks (e.g. GSM), require utilisation of various resources[1] in order to send information at a desired rate from the transmitters (sources) to the receivers (sinks) of the communication system.

The physical resources being considered relevant in this thesis are, firstly, the portion of the *frequency spectrum* of certain bandwidth that is required to serve the desired number of users in the communications system and, secondly, the *power* that has to be applied by the user terminals to reliably transmit information to the receiver (cf. [Jah99]).

Transmit power is limited due to several reasons, such as maximum available output power of amplifiers, health protection for the users of the communications equipment (especially for wireless communications with hand-held units), limitation of available energy (for battery operated devices), and, finally, signal power may also be limited due to system and regulatory aspects, e.g. to avoid excessive interference both for the respective communications system itself (intra-system interference) as well as for other communications systems (inter-system interference).

Also availability of the frequency spectrum resource generally has to be considered as limited. However, for cable based communications (by transmission of electrical or optical signals) this has little relevance, because if for a given communication service the number of users exceeds what can be served with the available bandwidth, additional capacity can be gained by simply installing new cables. Transmitting information in the same portion of the frequency spectrum, but over different cables can be considered as very efficient space division multiple access (SDMA), where the required spatial separation of any two given communication links (=cables) to avoid interference is virtually zero (however influence of cross-talk cannot be neglected in general, cf. [Mat00]).

In contrast, the frequency spectrum that is occupied (possibly time-variant) by a particular wireless communications system, offering its services in a certain geographical region, cannot be reused easily by another system in the same area, because of the mutual interference that would arise and eventually degrade the signal quality too much.[2] Therefore it is of great importance to implement any wireless communications system with a *spectrum efficiency* as high as possible to accommodate as many systems, services, and users in the available frequency spectrum as possible.

[1] Here, *resources* denotes all physical, technical, and economical means that have to be provided and possibly are consumed to create the intended communications services [Jah99].

[2] In [Ber77] the *spectral space* is introduced as a measure of the occupation of the spectrum resource in space and time, being defined by the product of occupied bandwidth, time and space.

This is particularly true for satellite communications systems as they have to compete about the limited resource of available frequency spectrum with the providers of wireless terrestrial communications systems, who express an ever-growing bandwidth demand to the spectrum regulatory authorities (e.g. European Conference of Postal and Telecommunications Administrations (CEPT)).

One method of achieving a high spectrum efficiency is the coordinated reuse of frequencies in geographical regions sufficiently spatially separated [LWJ00]. Nowadays satellite systems achieve such a *frequency reuse* by employing multi-beam antennas which create a pattern of fixed *spot beams*. Users located within different spot-beams may reuse the same spectrum resource at the same time. The frequency reuse is limited by the resulting mutual interference, caused by the users sharing the same spectrum resource at the same time (*co-channel users*). The *co-channel interference* generally increases as the spatial separation between co-channel users decreases, and the maximum frequency reuse is limited by the minimum co-channel user separation that still yields a tolerable level of interference.

Frequency reuse and, in consequence, spectrum efficiency can be maximised by employing means to more efficiently cope with the interference caused by closely spaced co-channel users. In the available literature one can identify two fundamentally different approaches to better cope with the interference from co-channel users.

The first approach is based on decoding concepts suggested by the information theoretic description of the multiple access channel (MAC)[3]. The multiple access channel is encountered when $M > 1$ sources send information to a common receiver, in our case the satellite, over a shared medium (cf. Fig. 1.1) [CT91]. In nowadays satellite communications systems the usual method of

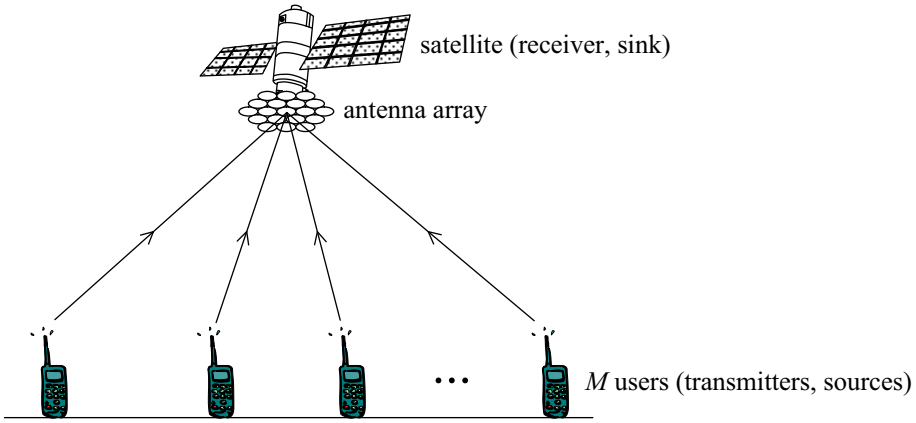


Figure 1.1: Schematic representation of the multiple access channel (MAC) comprising M sources transmitting information to a single common receiver.

decoding the M source signals at the common receiver is *independent decoding*, where each signal is decoded considering interference from co-channel users as unknown noise. Information theory of the MAC proves that independent decoding of the source signals possibly allows only to achieve a fraction of the information rates that can be achieved by optimal *joint decoding*. However, the related decoder complexity for joint decoding can be considered prohibitive (at least with the computing power available today) as the decoding complexity grows exponentially with the number of sources. Therefore *successive decoding* is an attractive alternative, which has a complexity

[3]Note the the acronym MAC sometimes refers to *medium access control*, however, in this work only the multiple access channel shall be meant.

growing only linear with the number of sources and can with certain restrictions achieve the same information rates as joint decoding. Fixed beamforming with subsequent successive decoding in satellite communications scenarios is discussed in, e.g., [Ern99, Moh00, Ern01].

The second approach tries to efficiently exploit the spatial distribution of the interfering co-channel users by employing an *antenna array* with subsequent adaptive beamforming at the satellite (again, cf. Fig. 1.1). This is sometimes also referred to as a *smart antenna*. By this it is possible to adapt the beam pattern of the antenna array to the particular co-channel user distribution by means of (digital) processing of the received source signals. Therefore interference is suppressed more efficiently, such that even closely spaced sources can use the same spectrum resource without causing excessive mutual interference. Because of the source signal separation based on the different source locations in space, this is referred to as space division multiple access (SDMA).[4] This approach is discussed in a number of publications dealing with the particular case of satellite communications, e.g. [CLW92, Bjö93, GG95, LL96, Yu96, Lüc98, LC00, Lüc00, Gay02].

Note that a receiver employing adaptive beamforming is transparent [Bjö93] and interference can be suppressed independent of the particular signal format (e.g. modulation scheme). Therefore also interference from arbitrary sources (e.g. microwave ovens [Jah99]) can be suppressed. With joint and successive decoding this is not possible.

Finally, both approaches can be combined in a receiver that employs adaptive beamforming as well as subsequent successive decoding. Of course, this receiver performs better than a receiver relying purely on one or the other interference mitigating technique.

1.1 Problem Formulation

The following questions are raised in the framework sketched above:

- Is it possible to describe the receiver types introduced above, employing fixed or adaptive beamforming, and independent or successive decoding, in a unified way?
- Is there a simple solution to the resource allocation problem of allocating transmit powers to the sources in order to achieve equal information rates for all sources?
- How do the receivers employing adaptive beamforming and/or successive decoding perform in comparison with the receiver relying on fixed beamforming and independent decoding being standard today?
- How close is the performance of a receiver employing both adaptive beamforming and subsequent successive decoding to a receiver employing only either adaptive beamforming or successive decoding?
- Is fixed beamforming with successive decoding performing better than adaptive beamforming with independent decoding or vice versa?
- How do the answers to the above questions possibly depend on the particular source distribution and the resulting interference scenario?

An outline of the content of the 5 chapters where the answers to the above questions will be derived is provided in the following.

[4] Note that also the frequency reuse obtained with fixed beams can be considered as SDMA.

1.2 Chapter Outline

Chapter 2: Satellite Communications As a basis for the subsequent chapters central concepts and terms related to satellite communications are introduced here, such as orbit heights etc. Also the characteristics of the wireless propagation channel encountered in satellite communications will be addressed.

Chapter 3: Adaptive Array Antennas and Satellite Scenarios A systematic treatment of the posed questions requires to define a detailed antenna array signal model. For this firstly basic antenna properties and related definitions are introduced. Further, since this work is concerned with satellite communications, satellite antennas are treated in detail, in particular multi-beam antennas and the related frequency reuse. A signal model for the MAC in satellite communications is defined. Finally, two particular satellite communications systems are introduced and described in detail, which will serve as sample scenarios to evaluate and compare the performance of the various receiver options introduced above.

Chapter 4: Resource Allocation for the Fading Vector Multiple-Access Channel In a communications system it is usually required that the sources are supposed to transmit information at a certain rate. Then the procedure of allocating transmit powers to the sources depends on whether the receiver employs independent or successive decoding. We will for both cases address the problem of optimally allocating transmit powers to all sources required to achieve the required information rates. To ease access to the topic, the well known classical fading MAC will be treated at first, where again independent and successive decoding will be looked at. Then independent and successive decoding are investigated for the case that an antenna array with adaptive beamforming is employed. As fixed beamforming is the usual technique today an according signal model is defined and the information theoretic implications are addressed.

Chapter 5: Receiver Structures for the Fading Vector MAC in Satellite Scenarios Various receiver options will be compared in the two satellite scenarios introduced in Chap. 3. The main focus is on four receiver options, namely combinations of fixed or adaptive beamforming with independent or successive decoding.

Chapter 6: Implementation Considerations In the last chapter of this thesis implementation aspects of the receiver variants are discussed for comparison.

Drahtlose Kommunikationssysteme, basierend auf Satelliten oder terrestrischen Netzen (z.B. GSM), bedürfen verschiedener Ressourcen[5], damit Informationen mit einer gewünschten Rate von den Sendern (Quellen) zu den Empfängern (Senken) des Kommunikationssystems übertragen werden können.

Die für die vorliegende Arbeit relevanten physikalischen Ressourcen sind erstens die Bandbreite des Frequenzspektrums, welche benötigt wird, um die gewünschte Teilnehmerzahl im Kommunikationssystem bedienen zu können und zweitens die Leistung, die von den Sendegeräten aufgebracht werden muss, um Informationen zuverlässig zum Empfänger übertragen zu können (vgl. [Jah99]).

[5] Hier werden mit dem Begriff *Ressourcen* alle physikalischen, technischen, oder ökonomischen Mittel benannt, die bereit gestellt werden müssen und evtl. verbraucht werden, um die gewünschten Kommunikationsdienste zu ermöglichen [Jah99].

Die Sendeleistung ist aus verschiedenen Gründen begrenzt, etwa durch die maximale Ausgangsleistung der Sendeverstärker, aus Gründen des Gesundheitsschutzes für die Nutzer der Kommunikationsgeräte (dies gilt speziell für drahtlose Kommunikation mit Handgeräten), wegen der begrenzt verfügbaren Energie (für batteriebetriebene Geräte) und schließlich kann die Sendeleistung auch durch System- oder regulatorische Aspekte beschränkt sein, z.B. um Interferenzen sowohl für das eigene (Intra-System Interferenz), als auch für andere Kommunikationssysteme (Inter-System Interferenz) zu vermeiden.

Auch die Ressource Frequenzspektrum muss generell als nur begrenzt verfügbar erachtet werden. Für kabelbasierte Kommunikation (mittels Übertragung von elektrischen oder optischen Signalen) kommt diese prinzipielle Begrenzung der Spektrumsressource kaum zu Tragen: falls für einen Kommunikationsdienst die Zahl der potenziellen Nutzer das übersteigt, was mit der verfügbaren Bandbreite bedient werden kann, dann kann zusätzliche Kapazität in einfacher Weise durch Installation weiterer Kabel gewonnen werden. Die Übertragung von Information im selben Frequenzbereich, jedoch über verschiedene Kabel, kann man als sehr effizienten Raummultiplex (space division multiple access (SDMA)) betrachten. Dabei ist die benötigte räumliche Trennung zweier Kommunikationsverbindungen (=Kabel) zur Vermeidung von Interferenz nahezu verschwindend (jedoch kann der Einfluss von Übersprechen im Allgemeinen nicht vernachlässigt werden, vgl. [Mat00]).

Im Gegensatz dazu kann das Frequenzspektrum, das von einem System zur drahtlosen Kommunikation in einer bestimmten geographischen Region und evtl. zeitvariant belegt wird, nicht einfach durch einen anderen Teilnehmer, Dienst oder ein anderes System in derselben Region wieder verwendet werden, da in der Regel die Trennung durch Antenne und Funkausbreitung nicht ausreicht und Interferenz entsteht, die die Signalqualität zu stark beeinträchtigen würde.[6] Deshalb ist es von großer Bedeutung, dass ein System zur drahtlosen Kommunikation mit einer möglichst hohen spektralen Effizienz (*spectrum efficiency*) implementiert wird, um eine möglichst große Zahl von Systemen, Diensten und Teilnehmern im verfügbaren Frequenzspektrum zu ermöglichen.

Dies gilt insbesondere für Satellitenkommunikationssysteme, weil diese um die begrenzt verfügbare Ressource des Frequenzspektrums mit den Anbietern drahtloser terrestrischer Kommunikationsdienste konkurrieren, welche einen stets wachsenden Bedarf an Bandbreite bei den Regulierungsbehörden anmelden (z.B. bei der European Conference of Postal and Telecommunications Administrations (CEPT)).

Eine Methode eine hohe spektrale Effizienz zu erzielen, ist die koordinierte Wiederverwendung von Frequenzen in ausreichend voneinander entfernten geographischen Regionen [LWJ00]. Heutige Satellitensysteme erzielen eine solche Frequenzwiederverwendung (*frequency reuse*) durch den Einsatz von Mehrkeulenantennen (*multi-beam antennas*), die eine Gruppe von schmalen Antennenkeulen (*spot beams*) erzeugen. Nutzer, die sich in verschiedenen, in der Regel jedoch nicht benachbarten Antennenkeulen befinden, können gleichzeitig dieselben Frequenzen nutzen. Die Frequenzwiederverwendung ist durch die resultierende gegenseitige Interferenz begrenzt, die durch diese Gleichkanalnutzer (*co-channel users*) verursacht wird. Die Gleichkanalinterferenz (*co-channel interference*) nimmt tendenziell zu, wenn die räumliche Trennung zwischen den Gleichkanal-Nutzern abnimmt, und der maximale Grad an Frequenzwiederverwendung ist durch den minimalen Abstand zwischen Gleichkanal-Nutzern bestimmt, der gerade noch ein tolerierbares Maß an Interferenz erzeugt.

[6] In [Ber77] wird der *Spektralraum* als Maß für die räumlich und zeitlich belegte Spektrumsressource eingeführt, definiert durch das Produkt von belegter Bandbreite, Zeit und Raum.

Die Frequenzwiederverwendung und damit die spektrale Effizienz können durch den Einsatz von Maßnahmen maximiert werden, die die Interferenz von nahe benachbarten Gleichkanal-Nutzern effizienter unterdrücken. In der Literatur kann man zwei fundamental unterschiedliche Ansätze finden, die eine bessere Unterdrückung der Interferenz von Gleichkanal-Nutzern ermöglichen.

Der erste Ansatz fußt auf Dekodierungsmethoden, die durch die informationstheoretische Beschreibung des Mehrfachzugriffskanals (*multiple access channel (MAC)*[7]) nahegelegt werden. Den Mehrfachzugriffskanal trifft man an, wenn $M > 1$ Quellen Information über ein gemeinsam genutztes Medium zu einem einzelnen Empfänger übertragen, in unserem Fall der Satellit (vgl. Abb. 1.2) [CT91]. In heutigen Satellitenkommunikationssystemen ist die üblicherweise

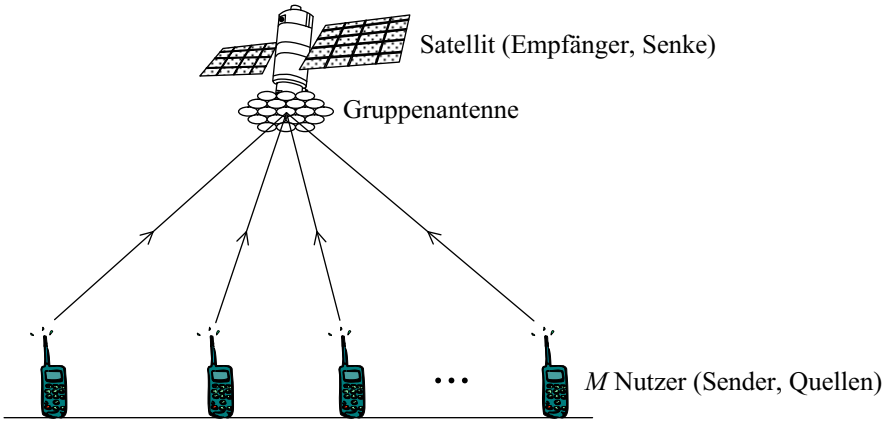


Figure 1.2: Schematische Darstellung des Mehrfachzugriffskanals, mit M Quellen, die Information zu einem einzelnen gemeinsamen Empfänger übertragen.

am gemeinsamen Empfänger verwendete Methode zur Dekodierung der M Quellensignale die unabhängige Dekodierung (*independent decoding*). Jedes Signal wird dekodiert, indem die Interferenz der Gleichkanal-Nutzer als unbekannte Störung betrachtet wird. Die Informationstheorie des Mehrfachzugriffskanals zeigt, dass mit unabhängiger Dekodierung möglicherweise nur ein Bruchteil der Informationsraten erzielt werden kann, wie sie mittels optimaler Verbunddekodierung (*joint decoding*) erreicht werden können. Jedoch verbietet bisher der für die Verbunddekodierung nötige Implementierungsaufwand die praktische Anwendung (zumindest mit der heute verfügbaren Rechenleistung), da der Dekodieraufwand exponentiell mit der Zahl der Quellen ansteigt. Deshalb stellt sukzessive Dekodierung (*successive decoding*) eine attraktive Alternative dar, weil die Komplexität einerseits nur linear mit der Zahl der Quellen ansteigt und weil mit gewissen Einschränkungen dieselben Informationsraten erreicht werden können wie mit Verbunddekodierung. Sukzessive Dekodierung in Verbindung mit Mehrkeulenantennen in Szenarien der Satellitenkommunikation wird z.B. in [Ern99, Moh00, Ern01] diskutiert.

Der zweite Ansatz versucht die räumliche Verteilung der interferierenden Gleichkanal-Nutzer effizient zu nutzen, indem eine Gruppenantenne (*antenna array*) zusammen mit adaptiver Strahlformung (*adaptive beamforming*) am Satelliten eingesetzt wird (vgl. Abb. 1.2). Diese Kombination wird auch als intelligente Antenne (*smart antenna*) bezeichnet. Damit ist es möglich, die Richtcharakteristik (*beam pattern*) der Gruppenantenne mittels (digitaler) Verarbeitung der empfangenen Quellensignale an die jeweilige konkrete Verteilung der Gleichkanal-Nutzer anzupassen. Auf diese Weise wird Interferenz effizienter unterdrückt, so dass selbst nah benachbarte Quellen

[7] Beachte, dass das Akronym MAC auch für den Begriff *medium access control* gebräuchlich ist, in dieser Arbeit soll jedoch ausschließlich der Mehrfachzugriffskanal, engl. multiple access channel, gemeint sein.

dieselbe Frequenz nutzen können, ohne dass zu viel Interferenz erzeugt wird. Weil die Trennung der Quellensignale auf der räumlichen Verteilung der Quellen basiert, wird dies als Raummultiplex bezeichnet (*space division multiple access (SDMA)*).[8] Dieser Ansatz wird in einigen Veröffentlichungen speziell im Zusammenhang mit Satellitenkommunikationssystemen betrachtet, z.B. [CLW92, Bjö93, GG95, LL96, Yu96, Lüc98, LC00, Lüc00, Gay02].

Es ist anzumerken, dass ein Empfänger, der adaptive Strahlformung anwendet, *transparent* ist [Bjö93] und somit Interferenz unabhängig vom jeweiligen Signalformat (z.B. Modulationsschema) unterdrückt werden kann. Deshalb können auch von beliebigen Quellen (z.B. Mikrowellen-Öfen [Jah99]) verursachte Interferenzen unterdrückt werden. Mit Verbund- oder sukzessiver Dekodierung ist dies nicht möglich.

Schließlich können beide Ansätze in einem Empfänger kombiniert werden, der dann sowohl adaptive Strahlformung als auch sukzessive Dekodierung anwendet. Gewiss übertrifft die Leistungsfähigkeit eines solchen Empfängers die eines Empfängers, der ausschließlich nur auf der einen oder der anderen Technik zur Interferenzunterdrückung basiert.

Fragestellungen

Folgende Fragestellungen ergeben sich innerhalb des oben beschriebenen Rahmens:

- Ist es möglich, die oben beschriebenen Empfängervarianten, die entweder feste oder adaptive Strahlformung, bzw. entweder unabhängige oder sukzessive Dekodierung anwenden, in einer einheitlichen Weise zu beschreiben?
- Gibt es eine einfache Lösung für das Problem der Ressourcenzuteilung, die Sendeleistungen den Quellen zuzuordnen, welche notwendig sind, um eine bestimmte Informationsrate für alle Quellen zu erzielen?
- Wie ist die Leistungsfähigkeit eines Empfängers, der adaptive Strahlformung und/oder sukzessive Dekodierung anwendet, im Vergleich zu einem Empfänger, der feste Strahlformung und unabhängige Dekodierung einsetzt, wie es heutzutage der Standard ist?
- Wie nahe liegt die Leistung des Empfängers, der sowohl adaptive Strahlformung, als auch sukzessive Dekodierung verwendet, an der des Empfängers, der nur entweder adaptive Strahlformung oder sukzessive Dekodierung verwendet?
- Ist der Empfänger, der feste Strahlformung mit sukzessiver Dekodierung implementiert, besser als der, der adaptive Strahlformung mit unabhängiger Dekodierung implementiert, oder umgekehrt?
- Wie hängen die Antworten auf obige Fragen ab von der jeweiligen Quellenverteilung und dem daraus resultierenden Interferenzszenario?

[8] Beachte, dass man auch die Frequenzwiederverwendung, die mit fester Strahlformung erzielt wird, als SDMA betrachten kann.

Übersicht über die Kapitel

Kapitel 2: Satelliten Kommunikation Als Grundlage für die folgenden Kapitel werden wichtige Begriffe und Größen aus der Satellitenkommunikation eingeführt, wie z.B. Orbithöhen. Die Eigenheiten des Übertragungskanals, wie er in der Satellitenkommunikation angetroffen wird, werden erläutert.

Kapitel 3: Adaptive Gruppenantennen und Satellitenszenarien Eine systematische Abhandlung der aufgeworfenen Fragen bedarf der Definition eines detaillierten Gruppenantennen-Modells. Dazu werden zunächst grundlegende Eigenschaften von Antennen und entsprechende Definitionen eingeführt. Weil sich die vorliegende Arbeit mit Satellitenkommunikation befasst, werden Satellitenantennen im Detail behandelt, insbesondere Mehrkeulenantennen und die damit verbundene Frequenzwiederverwendung. Ein Signalmodell für den Mehrfachzugriffskanal in der Satellitenkommunikation wird definiert. Zuletzt werden zwei konkrete Satellitenszenarien eingeführt und im Detail beschrieben, die als Beispielszenarien dienen werden, um die Leistungsfähigkeit der verschiedenen oben beschriebenen Empfängervarianten zu beurteilen und zu vergleichen.

Kapitel 4: Ressourcenzuteilung für den Vektor-Mehrfachzugriffskanal mit Signalschwund In einem Kommunikationssystem wird üblicherweise gefordert, dass die Quellen Information mit einer geforderten Rate übertragen. Die Prozedur, die den Quellen entsprechende Sendeleistungen zuteilt, hängt davon ab, ob der Empfänger unabhängige oder sukzessive Dekodierung anwendet. Für beide Fälle wird das Problem behandelt, den Quellen optimale Sendeleistungen zuzuteilen, so dass die geforderten Informationsraten erzielt werden. Um den Zugang zu dieser Thematik zu erleichtern, wird zunächst ein Überblick über den bekannten klassischen Mehrfachzugriffskanal mit Signalschwund (*classical fading MAC*) gegeben, wobei auch unabhängige und sukzessive Dekodierung betrachtet werden. Dann werden unabhängige und sukzessive Dekodierung für den Fall untersucht, dass eine Gruppenantenne mit adaptiver Strahlformung eingesetzt wird. Weil feste Strahlformung den heute üblichen Ansatz darstellt, wird ein entsprechendes Signalmodell definiert und die informationstheoretischen Implikationen diskutiert.

Kapitel 5: Empfängerstrukturen für den Vektor-Mehrfachzugriffskanal mit Signalschwund Verschiedene Empfängervarianten werden in den zwei in Kapitel 3 eingeführten Satellitenszenarien verglichen. Das Hauptaugenmerk liegt dabei auf den vier Varianten, die durch die Kombination von fester oder adaptiver Strahlformung, bzw. unabhängiger oder sukzessiver Dekodierung entstehen.

Kapitel 6: Überlegungen zur Implementierung Im letzten Kapitel dieser Arbeit werden Implementierungsaspekte der Empfängervarianten vergleichend diskutiert.

Chapter 2

Satellite Communications

A general outline of the principles of satellite communications will be provided in this chapter. This comprises the basic parameters describing the topology of a satellite communications system, and further the characteristics of the multipath propagation environment encountered in such a system.

2.1 Topology

The movement of any satellite around Earth is described by the satellites orbit parameters. In fact many important parameters of a satellite communication system introduced below are determined by the orbit of the respective satellites [LWJ00].

For circular orbits the satellite height over ground h_S is approximately constant (assuming an ideal spherical shape for Earth) and linked to the orbit period T_o according to

$$h_\mathrm{S} = \sqrt[3]{\gamma_0 M_\mathrm{E} \left(\frac{T_\mathrm{o}}{2\pi}\right)^2} - R_\mathrm{E}, \tag{2.1}$$

with the gravitational constant $\gamma_0 = 6.672 \cdot 10^{-11}\ \frac{\mathrm{m}^3}{\mathrm{kg\,s}^2}$, Earth mass $M_\mathrm{E} = 5.974 \cdot 10^{24}\ \mathrm{kg}$ and mean radius $R_\mathrm{E} = 6378.144\ \mathrm{km}$.

Generally, three circular orbits are distinguished according to the respective h_S [ITU90]:

- low earth orbit (LEO) for $500\ \mathrm{km} \leq h_\mathrm{S} \leq 2100\ \mathrm{km}$, orbit period $T_\mathrm{o} \approx 2\ \mathrm{h}$,
- medium earth orbit (MEO) (intermediate circular orbit (ICO) is used synonymous) for $5300\ \mathrm{km} \leq h_\mathrm{S} \leq 13200\ \mathrm{km}$, $T_\mathrm{o} \approx 4\ \mathrm{h} \ldots 6\ \mathrm{h}$, and
- geostationary orbit (GEO) for $h_\mathrm{S} = 35800\ \mathrm{km}$. The GEO is a special case of the geosynchronous orbit (GSO). The orbit period of a GSO equals the period of rotation of Earth (= 1 sidereal day = 23 h, 56 min, 4 s). A GEO satellite remains over a fixed longitude at zero latitude (i.e. the satellite remains fixed with respect to Earth coordinates). In contrast to the GEO, which lies in the equatorial plane of Earth, the GSO is an inclined orbit.

An example for a non-circular orbit is the highly elliptical orbit (HEO). Because of its particular importance in providing regional services (e.g. for Japan) at high minimum elevation $\varepsilon_{\min}$, employment of HEO satellites is discussed for future systems (cf., e.g., [NYZ+01]).

Selection of orbit determines important satellite system parameters: number of satellites required for coverage area (eventually global), signal propagation delay, signal attenuation, etc.

The *coverage area* is the geographical region for which a satellite can provide its services to the terrestrial communication partners (cf. Fig. 2.1). The maximum extension of the coverage area is limited by the minimum elevation $\varepsilon_{\min}$ under which the satellite is seen from a terminal with respect to the local horizon. Then the coverage area of a satellite is a spherical segment.

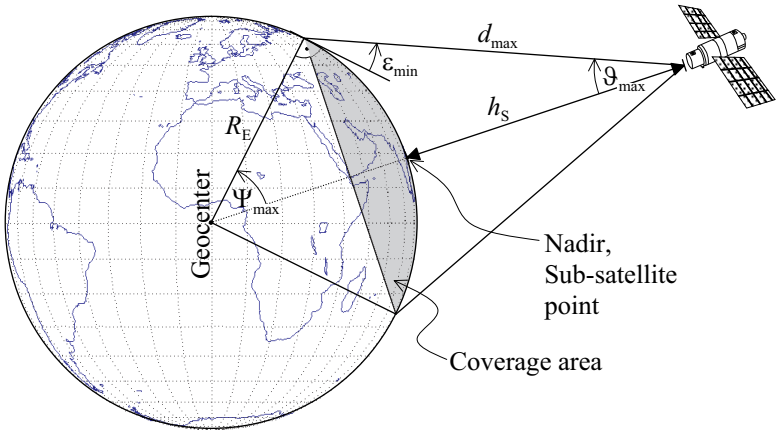


Figure 2.1: Relationship between minimum elevation $\varepsilon_{\min}$, coverage area, maximum Earth centre angle $\Psi_{\max}$, maximum slant range $d_{\max}$ and satellite height over nadir.

Given the satellites and the terminals position in Earth (spherical) coordinates (radius, longitude, latitude (r_S, λ_S, δ_S) and (r_T,λ_T, δ_T), respectively), the elevation ε can be calculated from

$$\varepsilon = \arcsin\left(\frac{r_S\left(\cos\delta_S\cos\delta_T\cos\left(\lambda_S-\lambda_T\right)+\sin\delta_S\sin\delta_T\right)-r_T}{d}\right), \tag{2.2}$$

with $r_S = R_E + h_S$ and d being the *slant range* (distance terminal-satellite) according to

$$d = \sqrt{r_S^2 + r_T^2 - 2r_S r_T\left(\cos\delta_S\cos\delta_T\cos\left(\lambda_S-\lambda_T\right)+\sin\delta_S\sin\delta_T\right)}. \tag{2.3}$$

Relative to the satellite and with respect to the connecting line satellite-nadir the terminal appears under the nadir angle given by

$$\vartheta = \arcsin\left(\frac{R_E}{r_S}\cos\varepsilon\right). \tag{2.4}$$

Together with the equation for the Earth centre angle

$$\Psi = \frac{\pi}{2} - \varepsilon - \arcsin\left(\frac{R_E}{r_S}\cos\varepsilon\right) \tag{2.5}$$

very compact formulas for d and ϑ can be provided:

$$\vartheta = \frac{\pi}{2} - \varepsilon - \Psi, \tag{2.6}$$

$$d = R_{\mathrm{E}} \frac{\sin \Psi}{\sin \vartheta}. \tag{2.7}$$

Depending on the type of satellite service, the required value for $\varepsilon_{\min}$ may vary. For fixed satellite services (i.e. the terrestrial user terminals are stationary) from a GEO satellite, a typical value is $\varepsilon_{\min} = 5°$, while for mobile services higher values of $\varepsilon_{\min}$ are desirable to reduce the probability of frequent signal blocking caused by buildings, vegetation, and geographical factors.

Further, the characteristics of the satellite antenna and the link budget define a lower limit for $\varepsilon_{\min}$, e.g. because smaller minimum elevation angle means larger slant range d leading to higher basic free-space transmission loss, and the coverage area is then the region in which the satellite can provide sufficient signal strength to allow information transmission to the user terminal and vice versa.

On the other hand, $\varepsilon_{\min}$ is upper bounded by the fact that for given h_{S} the coverage area becomes smaller with increasing $\varepsilon_{\min}$. For non-GSO satellite systems a compromise between a tolerable $\varepsilon_{\min}$ and the coverage area, which is directly related to the minimum number of satellites N_{S} and orbit planes required for a complete global coverage, must be found. N_{S} is given by [Jah99]

$$N_{\mathrm{S}} = \left\lceil \frac{4\pi R_{\mathrm{E}}^2}{6R_{\mathrm{E}}^2 \left(2 \arctan\left(\frac{\sqrt{3}}{\cos \Psi_{\max}}\right) - \frac{2\pi}{3}\right)} \right\rceil. \tag{2.8}$$

For the LEO systems Iridium and Globalstar $\varepsilon_{\min} = 10°$, $N_{\mathrm{S}} = 66$ and $\varepsilon_{\min} = 20°$, $N_{\mathrm{S}} = 48$, respectively, has been chosen, further, for the MEO system ICO again $\varepsilon_{\min} = 10°$, $N_{\mathrm{S}} = 10$ (cf. Fig. 2.2).

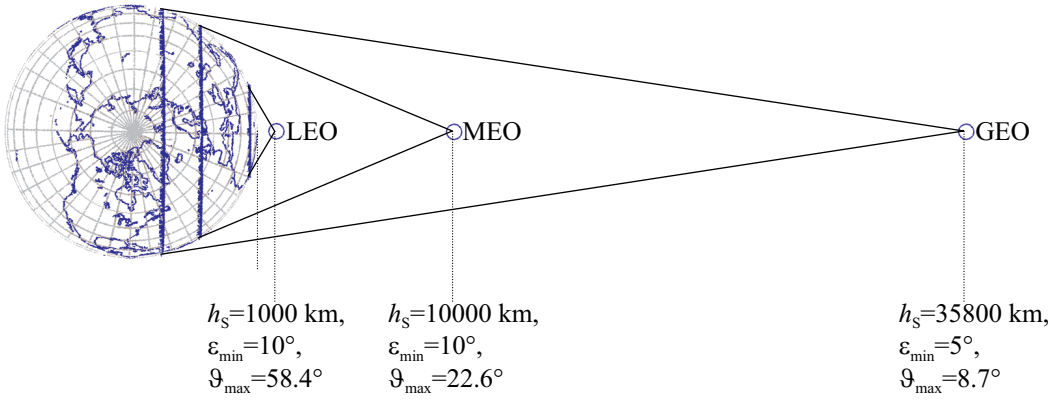


Figure 2.2: Comparison of coverage area for LEO, MEO, and GEO for typical minimum elevation $\varepsilon_{\min}$ and orbit height h_{S}. Also shown are the corresponding maximum nadir angles $\vartheta_{\max}$.

As already mentioned, satellites in HEO can be employed to provide local services and the minimum elevation can be as high as $\varepsilon_{\min} \approx 60°$, which allows satellite visibility even in urban environments with a high probability.

Finally, definitions regarding the links between user terminal, satellite, and gateway have to be introduced (see Fig. 2.3). The link from gateway to user terminal is called the *forward link*, while the link user terminal-gateway is the *return link*. Generally, the link from user terminal or gateway to satellite is the *uplink*, while the *downlink* designates the link from satellite to user terminal and, respectively, the gateway.

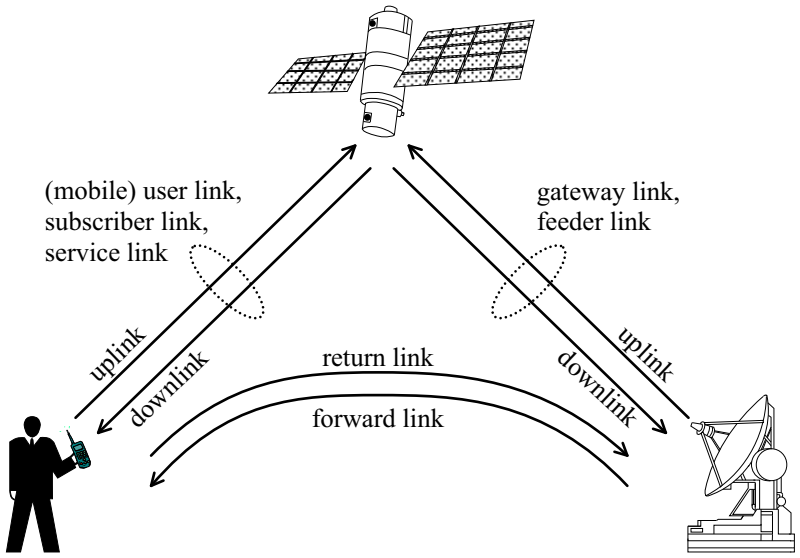


Figure 2.3: Terms for the designation of the links between user terminal, satellite, and gateway station.

2.2 Characteristics of the Propagation Channel for Satellite Systems

To avoid confusion, the term *propagation channel*, *physical channel* or simply *channel* will be used to summarise the (usually linear) distortions that are inflicted on a signal that is transmitted over a generic medium to a receiver, i.e. the physical channel includes attenuation, Doppler shift, delay, multipath effects (Doppler spread, delay spread) etc. In contrast, a portion of spectrum, a time slot or code (i.e. for frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA) or combinations) used to transmit information will be termed *communication channel*. Synonym with *communication channel*, we will also use the term *frequency slot* or *time slot* in case of FDMA and TDMA, respectively.

Just as in terrestrial wireless communications, the dominating property of the propagation channel between a satellite and a user terminal in the uplink and the downlink is the propagation of the transmitted signals on several independent signal paths. Then the propagation channel is said to be a *multipath channel*. Multipath propagation causes *fading*, i.e. fluctuation of the signal level at the receiver.

The main factors that affect signal quality at the receiver are [Jah99, LWJ00]:

Free Space Propagation Loss Denotes the decrease of power flux-density (cf. Sec. 3.1) of a transmit signal with increasing distance from the signal source.

Absorption Caused by ions, atmospheric gases (e.g. oxygen, water vapour) and rain (*rain fading*) in ionosphere, stratosphere and atmosphere, respectively, yielding additional signal attenuation.

Diffraction Caused by density fluctuation of the different atmospheric layers. This can lead to a pointing error for terrestrial terminals with directional antennas.

Depolarisation Faraday-rotation of linearly polarised electromagnetic waves may cause a polarisation mismatch at a likewise linearly polarised antenna. In L- (1610–$1626.5\ \mathrm{MHz}$) and S-band (2483.5–$2500\ \mathrm{MHz}$) circular polarisation may be used as a countermeasure, while depolarisation is negligible at higher frequencies.

Scintillation Rapid variation of the local density of the atmosphere causes varying absorption and deflection along the signal path.

Multipath Propagation and Shadowing Topographical or morphological factors (e.g. hills, vegetation, buildings) cause multipath propagation and shadowing (interruption/shadowing of the line-of-sight between satellite and terrestrial terminal).

Interference Caused by signal sources of the same considered communication system or other communication services (e.g. terrestrial radio relay), further man-made noise (e.g. household appliances).

Additive White Gaussian Noise The cause of additive white gaussian noise (AWGN) is thermal noise. The main sources of thermal noise are, firstly, the receiver noise (here, the main contribution comes from the low-noise amplifier (LNA) in the first stage of the receiver and from losses before this LNA) and, secondly, noise received by the antenna from natural sources of electromagnetic radiation (e.g. Sun, Earth). Thermal noise is characterised by the one-sided *thermal noise power spectral density* (constant over frequency)

$$N_0 = \mathrm{k}T_\mathrm{e}, \tag{2.9}$$

with $\mathrm{k} = 1.38 \cdot 10^{-23}\ \frac{\mathrm{J}}{\mathrm{K}} = -228.6\ \mathrm{dBWs/K}$ being the Boltzmann constant, and the overall *effective noise temperature* of the receiver plus antenna T_e, being defined according to (neglecting losses between antenna and first receiver stage) [MB98, LWJ00]

$$T_\mathrm{e} = T_\mathrm{A} + T_\mathrm{R}, \tag{2.10}$$

where T_A is the *antenna noise temperature* (taking care of the noise received from sky and surroundings) and T_R is the effective noise temperature of the receiver without the antenna.

In the following a detailed characterisation of the multipath propagation channel will be provided.

2.2.1 Temporal and Spectral Fading Characteristics

If not otherwise stated, any signals introduced below are understood as complex baseband signals [Pro95].

As depicted in Fig. 2.4, a transmitted signal arrives at the receiver via different paths from different directions, weighted according to the receive antenna gain (see Sec. 3.2.4) and summed, finally leading to the in general frequency selective and (for mobile receivers and/or transmitters) time-selective fading [Par92].

Depending on the characteristics of the antenna (e.g. omnidirectional or high-gain), the fading is more or less distinct.

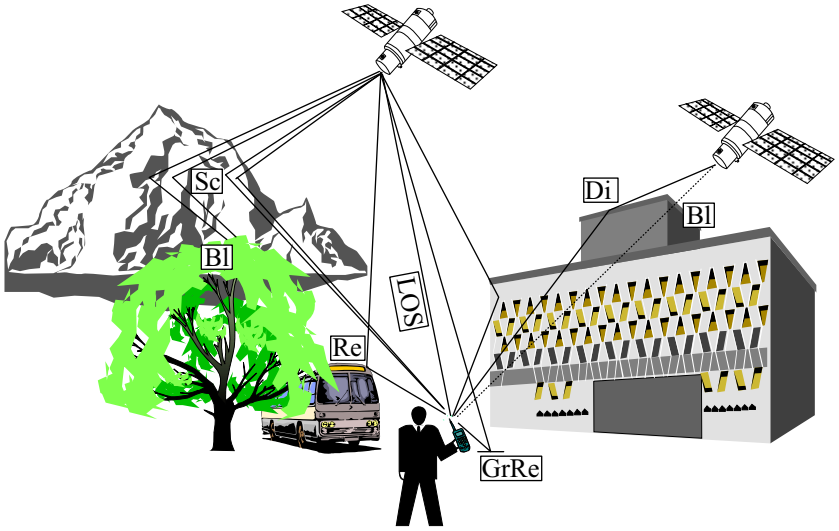


Figure 2.4: Generic multipath scenario and relevant physical effects for the LMS channel. Bl: Blocking, Di: Diffraction, GrRe: Ground Reflection, LOS: Line-of-Sight/Direct Path, Re: Reflection, Sc: Scattering.

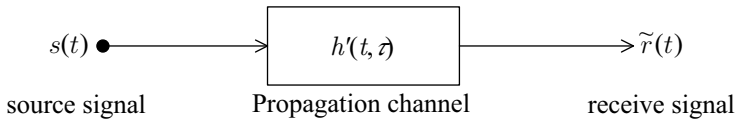


Figure 2.5: Block diagram of propagation channel.

The propagation channel can be characterised by a linear time-variant channel impulse response $h'(t,\tau)$ [Par92] (cf. Fig. 2.5). The channel impulse response $h'(t,\tau)$ is the output of the channel at time t, if a Dirac impulse was transmitted as input to the channel at time $(t-\tau)$. Usually, t is termed *observation time*, while τ is the *delay*.

The receive signal $\tilde{r}(t)$ resulting from a transmit signal $s(t)$ is given by the convolution

$$\tilde{r}(t) = s(t) * h'(t,\tau) = \int_{-\infty}^{\infty} h'(t,\tau)s(t-\tau)d\tau. \tag{2.11}$$

The time-variant impulse response $h'(t,\tau)$ is one of four equivalent system functions describing the propagation channel. Another system function is the *time-variant transfer function*

$$T'(t,f) = \int_{-\infty}^{\infty} h'(t,\tau)\mathrm{e}^{-j2\pi f\tau}d\tau = \mathfrak{F}_\tau\left\{h'(t,\tau)\right\}, \tag{2.12}$$

where $\mathfrak{F}_\tau\{\cdot\}$ denotes the Fourier transformation with respect to τ.

Generally speaking, the propagation channel is characterised by its variability with respect to t, i.e. how fast the channel impulse response varies with t, and what delays τ occur on the channel. The occurring delays τ determine the spectral characteristics of the propagation channel, i.e. variability of $T'(t,f)$ with respect to f.

If the wide-sense stationary uncorrelated scattering (WSSUS) assumption holds for the propagation channel, then the according characteristic parameters *Doppler spread* B_{f_D} and *delay spread* Δ_τ can be defined [Par92].

The Doppler spread B_{f_D} is related to the *coherence time* T_c according to [EB98]

$$T_\mathrm{c} \approx \frac{1}{B_{f_\mathrm{D}}}. \tag{2.13}$$

Doppler spread B_{f_D} and coherence time T_c are a measure for the variability of the channel impulse response with respect to time t.

The delay spread Δ_τ is linked to the *coherence bandwidth* B_c according to [EB98]

$$B_\mathrm{c} \approx \frac{1}{\Delta_\tau}. \tag{2.14}$$

Delay spread Δ_τ and coherence bandwidth B_c are a measure for the variability of the channel transfer function with respect to frequency f.

Now assume that symbols of duration T_s are transmitted. For the most relevant modulation methods in satellite communications, e.g. binary phase shift keying (BPSK), quaternary phase shift keying (QPSK), the signal bandwidth B_s is approximately $B_\mathrm{s} \approx 1/T_\mathrm{s}$.

The propagation channel is said to be *frequency non-selective*, if $B_\mathrm{s} \ll B_\mathrm{c}$ (equivalently: $T_\mathrm{s} \gg \Delta_\tau$). This is also known as the *narrowband assumption*.

Then the channel impulse response can be written as

$$h'(t,\tau) = h'(t)\delta(t-\tau) \quad \text{(narrowband assumption)} \tag{2.15}$$

and the convolution in (2.11) yields

$$\tilde{r}(t) = s(t)h'(t) \qquad \text{(frequency non-selective channel)}, \tag{2.16}$$

therefore this is also known as *multiplicative fading*, and we call $h'(t)$ the *multiplicative time-variant channel impulse response*.

Further, the channel is *time non-selective* if it holds $T_\mathrm{s} \ll T_\mathrm{c}$. In particular, this means that the fading is constant during one symbol period T_s. If channel estimation is to be considered, fading correlations during transmission of several symbols of duration T_s allows channel estimation via pilot symbols and interpolation of the fading process between the pilot symbols.

For satellite communications systems, the narrowband assumption holds for $B_\mathrm{s} \leq 1\ \mathrm{MHz}$ in the L-band, and even up to $B_\mathrm{s} \leq 30\ \mathrm{MHz}$ in the V-band ($40/50\ \mathrm{GHz}$ also denoted as EHF-band) [LWJ00].

Therefore, it will furthermore be assumed that the narrowband assumption holds. It is also assumed that the channel is time non-selective such that viability of channel estimation can be expected.

Fading Statistics

The temporal variability of the multiplicative time-variant channel impulse response $h'(t)$ is considered to occur on two different timescales.

Firstly, there is a possibly fast-changing component, which results from the Doppler spread B_{f_D} introduced by multipath propagation and movements of transmitter or receiver. This describes the short-term fluctuations of received signal power. The coherence time $T_\mathrm{c} \approx 1/B_{f_\mathrm{D}}$ introduced in the last section (cf. 2.13) as a measure for the short-term variability of the fading channel provides the timescale for the fast-fading.

Secondly, a slowly variable attenuation factor describes variations of the medium- and long-term average received power on a larger timescale. This attenuation factor results from changes in the geometry of the multipath propagation environment that occur whenever transmitter or receiver cover a distance large enough to alter the relevant propagation paths (shadowing by vegetation/building, changing free-space loss,...) [Jah99]. In the following it is assumed that time periods are considered during which this attenuation is constant.

Therefore, we split the fading factor $h'(t)$ in a fast-changing fading factor $h(t)$ and a quasi-constant link attenuation factor $\sqrt{\mu}$ according to (cf. [Jah99, LWJ00])

$$h'(t) = \frac{h(t)}{\sqrt{\mu}}, \tag{2.17}$$

with

$$\frac{1}{\mu} \triangleq \mathrm{E}\left\{|h'(t)|^2\right\}. \tag{2.18}$$

Hence, $h(t)$ is the channel impulse response $h'(t)$ normalised by $\sqrt{\mu}$ to unit power (therefore we call $h(t)$ the *normalised multiplicative time-variant channel impulse response*), i.e.

$$\mathrm{E}\left\{|h(t)|^2\right\} = 1. \tag{2.19}$$

The average received power is given by (multiplicative fading, cf. (2.16))

$$P_\mathrm{R} = \mathrm{E}\left\{|\tilde{r}|^2\right\} = \mathrm{E}\left\{|s(t)|^2\right\}\mathrm{E}\left\{|h'(t)|^2\right\} = \frac{P_\mathrm{T}}{\mu}, \tag{2.20}$$

where $P_\mathrm{T} = \mathrm{E}\left\{|s(t)|^2\right\}$ is the average transmit power, whereas the attenuation factor μ is assumed constant.

The reason for splitting the fading factor $h'(t)$ in a fast-changing and a quasi-constant component is that we will assume that a transmitters has knowledge of the attenuation factor $\sqrt{\mu}$, such that the transmit power can be accordingly controlled. On the other hand, the transmitter has no knowledge of the fading factor $h(t)$, and therefore the transmitter cannot control the transmit power to counteract fading. This assumption is motivated by the difficulties of a reliable estimation of the fading in the satellite scenario considered here [Dav97].

As depicted in Fig. 2.4, the signal propagation involves eventually a direct path (the line-of-sight (LOS) path) and, further, several reflected and scattered multipath components. In the following we will shortly address the resulting statistical properties of the fading-factor h.

We restrict to a description of the statistical properties using the corresponding probability density function (PDF) of $|h|$ and $|h|^2$, respectively. Therefore, correlations are not included in the consideration, and are also not required as we will later assume ideal interleaving, such that fading is indeed independent for each code symbol [LWJ00].

If a LOS path component is present (i.e. the LOS is unblocked) the absolute value $|\tilde{r}|$ is described by a Rice distribution.

On the other hand, if the LOS path is blocked, such that propagation occurs only via the indirect multipath components (non-line-of-sight (nLOS)), the absolute value of received signal $|\tilde{r}|$ follows a Rayleigh-distribution [Par92, LWJ00].

The power of the signal component received via the LOS path is denoted with P_{LOS}, while the total power of the diffuse multipath signal components is P_{mp}.

The Rice factor c_{R} is defined as the direct-to-multipath signal power ratio according to [Par92]

$$c_{\mathrm{R}} = \frac{P_{\mathrm{LOS}}}{P_{\mathrm{mp}}}, \tag{2.21}$$

Further, the average received power P_{R} is given by

$$P_{\mathrm{R}} = P_{\mathrm{LOS}} + P_{\mathrm{mp}} = P_{\mathrm{LOS}}(1 + 1/c_{\mathrm{R}}), \tag{2.22}$$

and it follows from (2.20)

$$\mu = \frac{P_{\mathrm{T}}}{P_{\mathrm{LOS}}} \frac{c_{\mathrm{R}}}{1 + c_{\mathrm{R}}}, \tag{2.23}$$

where the ratio $P_{\mathrm{T}}/P_{\mathrm{LOS}}$ will be provided in Sec. 3.2.5.

With these definitions and in case that a LOS path component is present, the PDF for the squared absolute value $|h|^2$, normalised to unit power, is given by the Rice distribution [Par92]

$$p_{\mathrm{LOS}}\left(|h|^2\right) = (1 + c_{\mathrm{R}}) \exp\left(-\left(c_{\mathrm{R}} + (1 + c_{\mathrm{R}})\,|h|^2\right)\right) \mathrm{I}_0\left(2\sqrt{|h|^2 c_{\mathrm{R}}(1 + c_{\mathrm{R}})}\right), \quad |h|^2 \geq 0 \tag{2.24}$$

with I_0 being the modified Bessel function of the first kind and order zero.

For $\lim_{c_{\mathrm{R}} \to \infty} p_{\mathrm{LOS}}$ the Rice distribution results in the AWGN channel with $h = 1$, whereas for $c_{\mathrm{R}} = 0$ the fading amplitudes h are Rayleigh distributed, and (2.24) becomes an exponential distribution according to

$$p_{\mathrm{nLOS}}\left(|h|^2\right) = \exp\left(-|h|^2\right), \quad |h|^2 \geq 0, \tag{2.25}$$

and (2.23) becomes $\mu = P_{\mathrm{T}}/P_{\mathrm{mp}}$.

The Rice factor c_{R} depends not only on the multipath propagation environment (e.g. many scatterers in an urban environment, open environment with only few scatterers in a rural scenario), but also on the terminal antenna [LWJ00]. For an omnidirectional antenna the multipath signal components are all equally superposed at the antenna. With a directive antenna the multipath signal components, which typically arrive from low elevation angles, are correspondingly attenuated by the reduced off-boresight antenna gain.

Further, the Rice factor c_{R} depends on the satellite elevation and on the carrier frequency. In [LWJ00] typical values for the Rice factor are given as $c_{\mathrm{R}} = 5\ \mathrm{dB} \ldots 15\ \mathrm{dB}$ for the L-band and a hemispherical antenna characteristic. In contrast the Rice factor is typically in the range $c_{\mathrm{R}} = 13\ \mathrm{dB} \ldots 22\ \mathrm{dB}$ for the EHF-band and for a high-gain antenna.

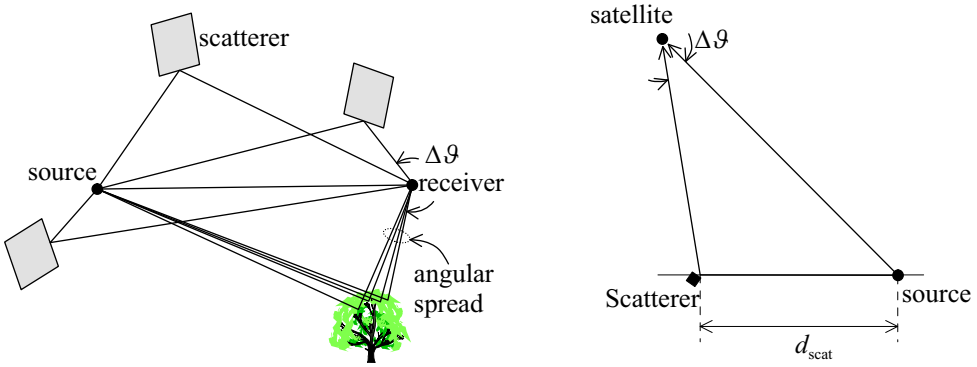
2.2.2 Spatial Characteristics of the Satellite Communication Channel

Antenna arrays allow to efficiently exploit the spatial distribution of the incident signals (or the signals transmitted), as will be discussed in great length in Chap. 3 and 4.

With the upcoming interest in adaptive antennas for application in wireless *terrestrial* communication systems, it was required to extend the up to this time established channel models to include the *spatial* characteristics of the signal propagation as well (see [ECS+98] for a synopsis of various spatial channel models).

While before with omnidirectional or sectorised antennas it was sufficient to describe the physical channel by means of power delay profiles and Doppler spectra to explain and reproduce the observed fading in time and frequency domain, it was now required to provide models for the angular distribution of the signal echoes arriving at the receiving antenna from different directions.

In a terrestrial communication system, both communication partners, i.e. the user terminal and the base station, are located on ground. The spatial characteristics of the multipath channel are governed by the distribution of the scatterers close to the user terminal and the base station [BBJ95]. Therefore, it is well known that one characteristic of the terrestrial wireless communications channel is that the signals of a single source (e.g. mobile terminal) may arrive at the receiver (e.g. base station) from a large angular range as depicted in Fig. 2.6(a). Further, a single propagation path may show a distinct *angular spread*, i.e. this path cannot be associated with a single discrete incident direction [Far97].



(a) Terrestrial channel: signals arrive from a large angular range $\Delta\vartheta$ at the receiver, single paths can eventually be spatially resolved. Due to nearby scatterers eventually significant angular spread is introduced. The angular spread cannot be spatially resolved, but must be considered in the beamformer design.

(b) Satellite channel: the typical maximum distance d_{scat} to relevant scatterers is $100\ \text{m} \ldots 300\ \text{m}$, therefore the angular range $\Delta\vartheta$ within which the source signal arrives at the receiver is negligible.

Figure 2.6: Illustration of the spatial characteristics of wireless terrestrial and satellite communications channel. Single paths experience angular spread due to scatterers.

Obviously, a completely different situation is encountered in satellite communications scenario, which is depicted in Fig. 2.6(b). Considering that the distance between satellite and terrestrial terminal (roughly $1000\ \text{km}$ to $40000\ \text{km}$) is much bigger than the typical maximum distance to relevant scatterers ($100\ \text{m} \ldots 300\ \text{m}$ [Jah99]), it is stated that the signals from any terrestrial source arrive at the satellite from the same direction (the angular range is less than $0.1°$ for a LEO satellite,

and reduces further for higher orbits). From this it follows further that the signal of a terrestrial source arrives at the satellite in a plane wavefront.

As in this work only the satellite case is considered it is henceforth assumed that a signal from a source arrives at the satellite from a single discrete direction-of-arrival (DOA).

Chapter 3

Adaptive Array Antennas and Satellite Scenarios

An introduction to antennas in general and antenna arrays in particular will be presented in this chapter. For this, it is required to review some basic concepts regarding the propagation of an electromagnetic wave and, further, parameters that are used to describe the fundamental properties of antennas and antenna arrays.

The definitions in this chapter follow the recommendations of the International Telecommunication Union (ITU), further [Kra88, MG86, MB98] are used.

3.1 Electromagnetic Fundamentals

The field components $\mathbf{E}$ (electric field vector) and $\mathbf{H}$ (magnetic field vector) of an electromagnetic wave emitted by a generic source (e.g. an antenna) are described in spherical coordinates (cf. Fig. 3.1) by the real vectors

$$\mathbf{E}(\vartheta,\varphi,r,t)=\begin{pmatrix} E_\vartheta(\vartheta,\varphi,r,t)\\ E_\varphi(\vartheta,\varphi,r,t)\\ E_r(\vartheta,\varphi,r,t)\end{pmatrix},\quad \mathbf{H}(\vartheta,\varphi,r,t)=\begin{pmatrix} H_\vartheta(\vartheta,\varphi,r,t)\\ H_\varphi(\vartheta,\varphi,r,t)\\ H_r(\vartheta,\varphi,r,t)\end{pmatrix}, \tag{3.1}$$

where $E_\vartheta(\vartheta,\varphi,r,t)$, $E_\varphi(\vartheta,\varphi,r,t)$ and $E_r(\vartheta,\varphi,r,t)$ are the components of the electric field in direction of the base vectors $\mathbf{e}_\vartheta$, $\mathbf{e}_\varphi$ and $\mathbf{e}_r$ as depicted in Fig. 3.1. The magnetic field components are defined accordingly.

The Poynting vector$\mathbf{S}(\vartheta,\varphi,r,t)$ is given by the cross product of the E- and H-field vectors, such that

$$\mathbf{S}(\vartheta,\varphi,r,t)=\mathbf{E}(\vartheta,\varphi,r,t)\times\mathbf{H}(\vartheta,\varphi,r,t) \tag{3.2}$$

and describes both direction and amount of power per unit area transported by an electromagnetic wave in that direction.

The absolute value of (3.2) is termed as *power flux-density* and is accordingly defined as

$$S(\vartheta,\varphi,r,t)=\left|\mathbf{S}(\vartheta,\varphi,r,t)\right|. \tag{3.3}$$

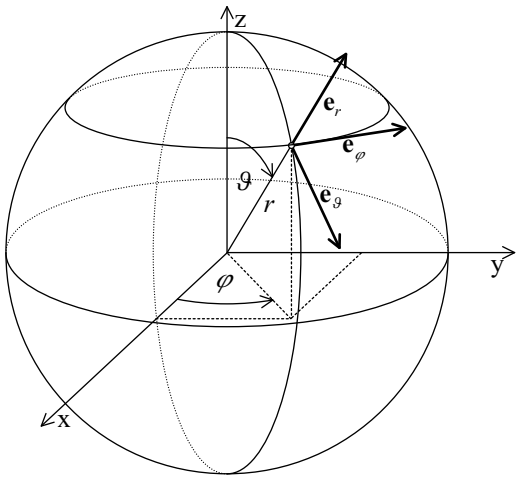


Figure 3.1: Definition of spherical coordinates (ϑ, φ, r) and base vectors $\mathbf{e}_\vartheta$, $\mathbf{e}_\varphi$, $\mathbf{e}_r$.

Only the far-field of the electromagnetic field induced by the antenna will be considered furthermore.

The far-field assumption implicates that:

- Strictly, the far-field condition is valid for $r \to \infty$, but in practice it suffices that the distance r between the antenna and the point in space where the electromagnetic field components are considered is much larger than the wavelength λ *and* the maximum antenna dimension d_0 (for a reflector antenna d_0 would be equal to the diameter of the reflector), i.e.

$$r \gg \lambda \quad \text{and} \quad r \gg d_0. \tag{3.4}$$

  Equation (3.4) is the very condition that justifies the far-field assumption.

- The electromagnetic field components $\mathbf{E}(\vartheta, \varphi, r, t)$ and $\mathbf{H}(\vartheta, \varphi, r, t)$ are in-phase and are orthogonal to each other, and to the propagation vector (which means there are no radial field components along the direction of propagation, i.e. $E_r(\vartheta, \varphi, r, t) = H_r(\vartheta, \varphi, r, t) = 0$),

- Locally the electromagnetic wave propagates with a plane wavefront. Here, "locally" means that the wavefront can be considered plane with respect to the dimensions of a receiving aperture. Considering a multipath channel, for this property it is also required that there are no scatterers close to the receiver. The superposition of the multipath signal components leads to a non-planar phase front.

- The ratio of the magnitude of the electric and the magnetic field components is given by

$$Z_\mathrm{I} = \frac{|\mathbf{E}(\vartheta, \varphi, r, t)|}{|\mathbf{H}(\vartheta, \varphi, r, t)|} = \sqrt{\frac{\mu}{\varepsilon}} = \sqrt{\frac{\mu_\mathrm{r}\mu_0}{\varepsilon_\mathrm{r}\varepsilon_0}}, \tag{3.5}$$

  where Z_I is the intrinsic impedance, μ is the magnetic permeability and ε the electric permittivity of the propagation medium, ε_r and μ_r is the relative permittivity (also called the dielectric constant of the specific medium) and relative permeability, respectively. Further, ε_0 is the permittivity and μ_0 the permeability of the vacuum. With $\varepsilon_0 = 8.85 \cdot 10^{-12} \frac{\mathrm{F}}{\mathrm{m}}$, $\mu_0 = 4\pi \cdot 10^{-7} \frac{\mathrm{H}}{\mathrm{m}} = 1.257 \frac{\mathrm{H}}{\mathrm{m}}$, the intrinsic impedance for free-space (vacuum) is given by $Z_0 = 376.7\Omega \approx 120\pi\ \Omega$).

- The Poynting vector has a radial component only such that (3.2) yields

$$\mathbf{S}(\vartheta, \varphi, r, t) = S(\vartheta, \varphi, r, t)\mathbf{e}_r = |\mathbf{E}(\vartheta, \varphi, r, t)|\,|\mathbf{H}(\vartheta, \varphi, r, t)|\,\mathbf{e}_r \tag{3.6}$$

Polarisation

The orientation of the electric and magnetic field vectors $\mathbf{E}(\vartheta, \varphi, r, t)$ and $\mathbf{H}(\vartheta, \varphi, r, t)$ defines the polarisation of the electro-magnetic wave [ST81]. Because the magnetic field vector is orthogonal to the electric field vector, polarisation is unambiguously defined by one of the two field vectors. Typically the electric field vector is used for definition of the polarisation.

In general, the electric field vector describes an ellipse. Special cases of the polarisation ellipse result in *linear* and *circular* polarisation.

With linear polarisation the electric field vector remains constantly within a fixed plane, the *polarisation plane*. The polarisation plane is spanned by the electric field vector and the Poynting vector.

With circular polarisation the electric field vector describes a spiral with constant radius along the propagation direction. Depending on the direction of rotation, the wave can be either right hand circularly polarised (RHCP) or left hand circularly polarised (LHCP).

The polarisation of an antenna is defined by the polarisation of the wave radiated by the antenna in a given direction. For optimal reception (i.e. maximising the receive signal) of an electro-magnetic wave of particular polarisation, the polarisation of the receiving antenna must be matched to that of the incident electro-magnetic wave.

In the remainder of this work it is always assumed that the polarisation of any antenna is matched to the polarisation of the electro-magnetic wave that is received by the antenna, and this will not be explicitly stated anymore in the following.

3.2 Basic Antenna Properties and Definitions

3.2.1 Reciprocity of Antennas

Eqn. (3.6) is the central equations building the basis for the definition of fundamental antenna parameters. In the presentation of following definitions, arguments partly assuming transmit antennas, partly assuming receive antennas will be used. However, due to the law of reciprocity, all definitions derived for the receiving case are valid for the antenna acting as a transmitter as well and vice versa [ST81].

For reasons of simplicity of notation, the dependency on time t will be omitted in the notation in the following.

3.2.2 Radiation Pattern

A widely used concept in antenna theory is the hypothetical (i.e. technical not realisable) *isotropic* or *omnidirectional* antenna. An isotropic antenna radiates power equally in all directions in space, i.e. independent of ϑ, φ; the surfaces of equal phase are spheres. Hence, the power flux-density for the omnidirectional source is given by

$$S_{\mathrm{i}}(r) = \frac{P_{\mathrm{rad}}}{4\pi r^2}, \tag{3.7}$$

where P_{rad} is the total power radiated by the antenna.

For a generic transmit antenna it is obvious that the power flow through a generic surface enclosing the antenna must be constant and equal to the overall radiated power P_{rad}. To simplify matters and without loss of generality (w.l.o.g.), integration over a sphere with arbitrary radius r can be chosen [ST81], such that

$$P_{\mathrm{rad}} = \iint_A \mathbf{S}(\vartheta, \varphi, r) d\mathbf{A} = \int_{\varphi=0}^{2\pi} \int_{\vartheta=0}^{\pi} S(\vartheta, \varphi, r) r^2 \sin(\vartheta) d\vartheta d\varphi \Big|_{r=\mathrm{const.}}. \tag{3.8}$$

In contrast to the isotropic radiator, the power flux-density induced by arbitrary antennas generally varies with changing direction in space. The term *radiation pattern* is used to indicate this dependency both loosely qualitative, as well as quantitative (also, some publications use this term in particular for the graphical representation of the radiation characteristics [ST81]).

By normalisation of the power flux-density with respect to its maximum value, the dependency on r is dropped, leading to the definition of the *power pattern*

$$|F(\vartheta, \varphi)|^2 \triangleq \frac{S(\vartheta, \varphi, r)}{S_{\max}(r)} \Big|_{r=\mathrm{const.}}, \tag{3.9}$$

where

$$S_{\max}(r) = \max_{\vartheta, \varphi} S(\vartheta, \varphi, r) \tag{3.10}$$

and $F(\vartheta, \varphi)$ is the normalised *field (magnitude) pattern.*

At the satellite we will consider planar antenna arrays, and it will always be assumed that the considered antenna is oriented in the spherical coordinate system, such that the antenna boresight[1] is at $\vartheta = 0$ (cf. Fig.3.2).

3.2.3 Antenna Directivity

The directivity $D(\vartheta, \varphi)$ in a given direction (ϑ, φ) is defined as the ratio of the intensity of radiation (defined as the power per unit solid angle (steradian)), in that direction, to the radiation intensity averaged over all directions. In turn, the radiation intensity averaged over all directions

[1]The boresight of an antenna is the direction of the maximum gain [LWJ00]. Considering planar antennas, this usually corresponds to a vector perpendicular to the array plane.

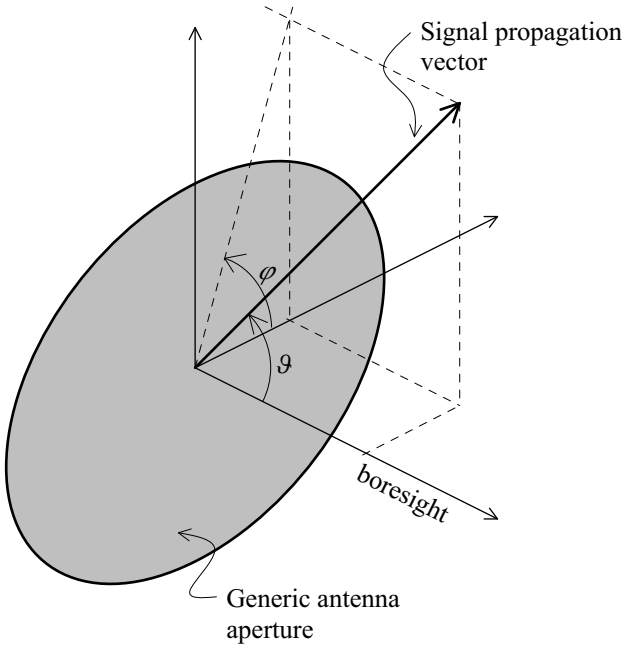


Figure 3.2: Definition of boresight, off-boresight angle ϑ and azimuth angle φ.

is equivalent to the radiation density induced by the isotropic radiator as given by (3.7) [ITU95]. This definition yields

$$\begin{aligned} D(\vartheta,\varphi) &= \left.\frac{S(\vartheta,\varphi,r)}{\frac{P_{\mathrm{rad}}}{4\pi r^2}}\right|_{r=\text{const.}} \\ &= \left.\frac{r^2 S(\vartheta,\varphi,r)}{\frac{1}{4\pi}\int_{\varphi'=0}^{2\pi}\int_{\vartheta'=0}^{\pi} S(\vartheta',\varphi',r) r^2 \sin(\vartheta') d\vartheta' d\varphi'}\right|_{r=\text{const.}} \\ &= \frac{|F(\vartheta,\varphi)|^2}{\frac{1}{4\pi}\int_{\varphi'=0}^{2\pi}\int_{\vartheta'=0}^{\pi} |F(\vartheta',\varphi')|^2 \sin(\vartheta') d\vartheta' d\varphi'} \\ &= \frac{4\pi}{\Omega_{\mathrm{A}}} |F(\vartheta,\varphi)|^2 , \end{aligned} \tag{3.11}$$

where in the last step the definition of the *beam solid angle* Ω_{A} is used with

$$\Omega_{\mathrm{A}} = \int_{\varphi'=0}^{2\pi}\int_{\vartheta'=0}^{\pi} |F(\vartheta',\varphi')|^2 \sin(\vartheta') d\vartheta' d\varphi'. \tag{3.12}$$

The maximum directivity D of an antenna is accordingly given by

$$D \triangleq \max_{\vartheta,\varphi} D(\vartheta,\varphi) = \frac{4\pi}{\Omega_{\mathrm{A}}}. \tag{3.13}$$

It is usual to denote both the direction dependent directivity $D(\vartheta,\varphi)$, as well as the maximum directivity D simply as *directivity*, because a distinction is mostly possible by the context.

3.2.4 Antenna Gain

The *antenna gain* is defined as the ratio of the power required at the input of a lossless reference antenna (e.g. an isotropic antenna isolated in space or a half-wave ($\lambda/2$) dipole) to the power

supplied to the input of the (lossy) antenna under consideration to produce, in a given direction, the same power flux-density $S(\vartheta, \varphi, r)$ at the same distance [ITU90]. Usually, the term gain refers to both the maximum gain, as well as to the direction dependent definition of the antenna gain $G(\vartheta, \varphi)$ (the isotropic radiator will be used as reference consistent to the definition of directivity in the last section, see (3.11).

Hence,

$$G(\vartheta, \varphi) = \left. \frac{4\,\pi\, r^2 S(\vartheta, \varphi, r)}{P_{\mathrm{T}}} \right|_{r=\mathrm{const.}} = D(\vartheta, \varphi)\eta_{\mathrm{r}}, \tag{3.14}$$

where P_{T} is the total input power to the antenna and

$$P_{\mathrm{rad}} = \eta_{\mathrm{r}} P_{\mathrm{T}}, \quad 0 \leq \eta_{\mathrm{r}} \leq 1, \tag{3.15}$$

where η_{r} is the antenna *radiation efficiency*, which takes care of ohmic and dielectric losses, further losses due to mismatching in the antenna feed (cf. Fig. 3.3). If the gain is given in dB and the

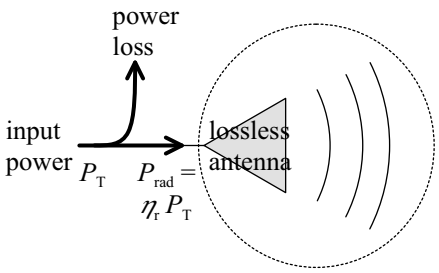


Figure 3.3: On the definition of total input power P_0 and radiation efficiency η_{r}.

reference is the isotropic antenna, it is usual to use dBi instead of dB.

The maximum gain of an antenna is given by

$$G = \frac{A_{\mathrm{eff}}\, 4\,\pi}{\lambda^2}, \tag{3.16}$$

where λ is the wavelength of the transmitted signal and A_{eff} if the effective aperture area of the antenna. The relation between A_{eff} and the geometric area A_{geo} for aperture antennas is given by

$$A_{\mathrm{eff}} = \eta_{\mathrm{A}} \eta_{\mathrm{r}} A_{\mathrm{geo}} \quad \text{(for aperture antennas)}, \tag{3.17}$$

where η_{r} was defined in (3.15) and the *aperture efficiency* η_{A} is made up of several factors, e.g. illumination efficiency (w.r.t. uniform illumination), spill-over efficiency, surface finish efficiency etc.

The product $\eta_{\mathrm{a}} = \eta_{\mathrm{A}} \cdot \eta_{\mathrm{r}}$ (overall antenna efficiency, or simply *antenna efficiency*) is typically between 0.55 and 0.65, depending on the design of the antenna [MB98, LWJ00].

Sometimes description on basis of signal powers may not be sufficient. Then the antenna is described by a transfer function, which becomes a constant phase shift and attenuation if the transmission behaviour is frequency-independent. This is the case considered in the remainder of this work.

Therefore the *complex magnitude pattern* $g(\vartheta, \varphi)$ of an antenna is introduced, which is a complex weight by which an incident signal is multiplied (i.e. it is assumed that the antenna performs equal over frequency). The following relation between antenna gain and the complex magnitude pattern holds

$$g(\vartheta, \varphi) = \sqrt{G(\vartheta, \varphi)} e^{j\alpha(\vartheta,\varphi)}, \tag{3.18}$$

where $\alpha(\vartheta, \varphi)$ is the phase shift introduced by the antenna.

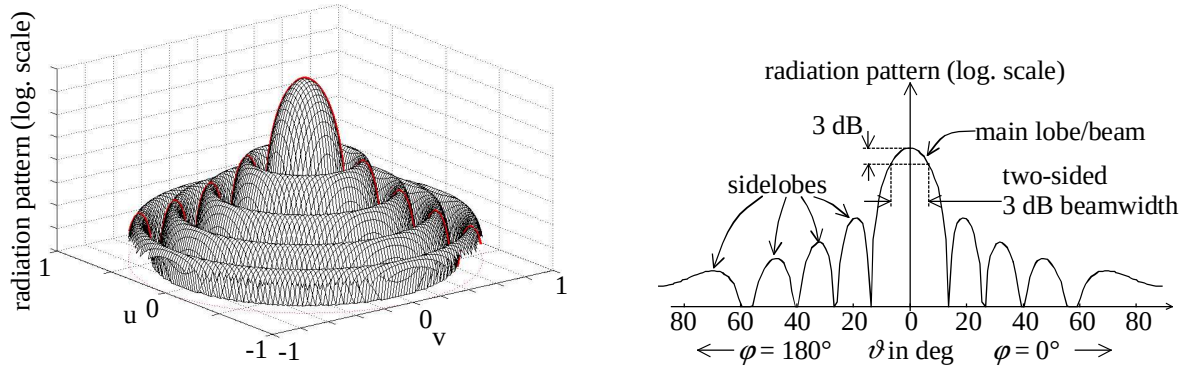
A generic radiation pattern introducing important terms describing the various sections of the pattern is shown in Fig. 3.4. In Fig. 3.4(a) we use the *sine space* coordinates (u, v), defined according to

$$\begin{aligned} u &= \sin\vartheta \cos\varphi \\ v &= \sin\vartheta \sin\varphi \end{aligned} \tag{3.19}$$

which are often used to display 2-dimensional radiation patterns. Accordingly it holds

$$\vartheta = \arcsin\sqrt{u^2 + v^2}, \qquad \varphi = \arctan\frac{v}{u}, \tag{3.20}$$

further note that for real $\vartheta \leq \frac{\pi}{2}$ it must hold $u^2 + v^2 \leq 1$.



(a) 2-dimensional Radiation pattern shown in u-v-coordinates.

(b) Radiation pattern in a cutting plane.

Figure 3.4: Radiation pattern of a circular aperture (logarithmic scale).

3.2.5 Power Transmission Formula

The received signal power for the LOS path alone is P_{LOS} (cf. Sec. 2.2.1) and can be calculated with (for a detailed derivation see [ST81])

$$P_{\mathrm{LOS}} = \frac{P_{\mathrm{T}} G_{\mathrm{T}}(\vartheta_{\mathrm{T}}, \varphi_{\mathrm{T}}) G_{\mathrm{R}}(\vartheta_{\mathrm{R}}, \varphi_{\mathrm{R}})}{L_{\mathrm{bf}}}, \tag{3.21}$$

where $G_{\mathrm{R}}(\vartheta_{\mathrm{R}}, \varphi_{\mathrm{R}})$ and $G_{\mathrm{T}}(\vartheta_{\mathrm{T}}, \varphi_{\mathrm{T}})$ are the gains of the receive and transmit antenna, respectively. Further, L_{bf} is the *free-space basic transmission loss* with

$$L_{\mathrm{bf}} = \left(\frac{4\pi r}{\lambda}\right)^2, \tag{3.22}$$

and the product $P_{\mathrm{T}}G_{\mathrm{T}}(\vartheta_{\mathrm{T}},\varphi_{\mathrm{T}})$ is the equivalent isotropically radiated power (EIRP) of the transmitter [LWJ00].

The the ratio $P_{\mathrm{T}}/P_{\mathrm{LOS}}$ is readily obtained from (3.21) according to

$$\frac{P_{\mathrm{T}}}{P_{\mathrm{LOS}}} = \frac{L_{\mathrm{bf}}}{G_{\mathrm{T}}(\vartheta_{\mathrm{T}},\varphi_{\mathrm{T}})G_{\mathrm{R}}(\vartheta_{\mathrm{R}},\varphi_{\mathrm{R}})},$$

and the attenuation factor μ preliminary defined in (2.23) is finally given by

$$\mu = \frac{L_{\mathrm{bf}}}{G_{\mathrm{T}}(\vartheta_{\mathrm{T}},\varphi_{\mathrm{T}})G_{\mathrm{R}}(\vartheta_{\mathrm{R}},\varphi_{\mathrm{R}})}\frac{c_{\mathrm{R}}}{1+c_{\mathrm{R}}}. \tag{3.23}$$

3.3 Satellite Antennas

After the characteristic parameters of an arbitrary antenna have been introduced above, we can now turn in particular to satellite antennas.

3.3.1 Spot Beams

To allow a compact design for the terrestrial terminals by employment of small, low gain terminal antennas, an accordingly high gain antenna has to be employed at the satellites side. By increasing the satellite antenna diameter d_{a} to achieve a higher gain, the beamwidth decreases, since for the relation between one-sided 3 dB-beamwidth $\vartheta_{3\mathrm{dB}}$ and aperture diameter d_{a} it holds [ST81,LWJ00]

$$\vartheta_{3\mathrm{dB}} \approx 0.61\frac{\lambda}{d_{\mathrm{a}}} \qquad \text{[radians]}, \tag{3.24}$$

where the factor 0.61 is a proportionality term assuming a typical tapered aperture power density distribution [LWJ00].

Usually $\vartheta_{3\mathrm{dB}}$ becomes significantly smaller than $\vartheta_{\max}$ (the maximum off-boresight angle defined by the extent of the coverage area of the satellite, cf. Fig. 2.2). In consequence the coverage area of the satellite has to be illuminated by several *spot beams* [LWJ00] (cf. Fig. 3.5).

The spot beams create a pattern of possibly overlapping cells on Earth's surface, where the boundary of each cell is typically defined by the 3 dB or 4.3 dB below maximum gain contour of the corresponding spot beam, i.e. the gain at edge-of-cell (EOC) is 3 dB or 4.3 dB below the maximum. Choosing a 4.3 dB EOC gain drop maximises the *absolute* antenna gain at EOC for a given cell size. However, a value of 3 dB EOC gain drop may also be appropriate for other reasons, e.g. to limit the required dynamic range of the power amplifiers of terrestrial terminals, and in fact this value is often encountered when looking at actual satellite system designs (e.g. ICO [MS98]). For a complete coverage while obtaining minimum cell overlap, the cell centres are located on a triangular grid [LWJ00].

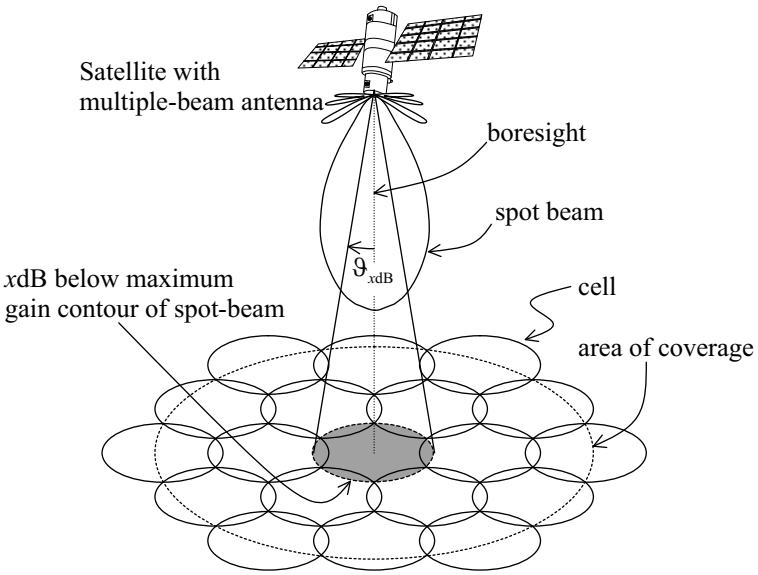


Figure 3.5: Cell pattern with cell centres located on triangular grid. Each cell corresponds to an individual spot beam, created by a multiple beam antenna. Further, a generic spot beam gain pattern is shown to illustrate the relation between EOC and beamwidth $\vartheta_{x\,\mathrm{dB}}$ ($x = 3\ \mathrm{dB}, 4.3\ \mathrm{dB}$).

3.3.2 Multiple-Beam Antennas

Because several spot beams have to be created simultaneously by the satellite antenna, such antennas are called multiple-beam antenna (MBA) and various MBA designs can be used to create the spot beams [LL96].

Fundamental trade-offs in the system and particular antenna design for a given coverage area and for constant wavelength λ are

- With increasing antenna gain the beamwidth decreases and, hence, the cell size decreases as well (cf. (3.24)). Further, the required antenna diameter increases.
- The number of cells required to cover the coverage area increases with decreasing cell size. In addition, for non-GSO satellite systems smaller cells lead to a higher inter-cell handover rate.
- With increasing number of cells (spot beams) and increasing antenna diameter, the antenna complexity and mass increase as well. Finally, if the MBA is implemented using analog beamforming networks a higher number of spot beam comes along with an increased signal degradation. For this reason current satellite systems that feature a high number of spot beams employ digital beamforming to produce a high number of fixed spot beams [MS98].

So, the demand for a high gain antenna is essentially limited by satellite antenna mass and complexity, and a maximum inter-spot beam handover rate which finally may lead to an intolerable amount of signalling traffic. Frequent handovers increase also the risk of call-dropping as in the newly entered spot beam all channels could be occupied already. In Chap. 3 of [DP02] an inter-spot beam handover rate of around $1/\min$ is mentioned.

Because of the demand for broadband satellite services and the ever ongoing competition between terrestrial and satellite service providers over the limited bandwidth resource, several proposals for future satellite systems suppose to use frequencies in the Ka or EHF band [Ega99, MPPL+98]. The use of even higher frequencies is limited by mainly too high an attenuation due to rain/snow and other absorption effects (clouds/fog/water vapour, molecular absorption, tropospheric turbulences) as the availability of the communication link can drop significantly. Satellite communication to aeronautical terminals (i.e. weather conditions and cloud cover have less or even no impact depending on the aircraft altitude) being an exception where even optical links are conceivable [GHLW03].

3.3.3 Frequency Reuse

Besides smaller user terminals there is another advantage of great importance which is achieved by overlaying the coverage area with a number of spot beams: the possibility of *frequency reuse* among users in different spot beams [LWJ00, MB98] (a twofold frequency reuse is also possible by means of antenna polarisation discrimination).

For CDMA, the same carrier frequency can be used in all cells, as all source signals can be separated by employing nearly orthogonal code sequences [LWJ00].

In contrast, for FDMA or TDMA, adjacent cells must use different frequency bands, since signals transmitted simultaneously in the same frequency band would lead to strong mutual interference (co-channel interference (CCI)). However, users whose mutual interference is sufficiently suppressed by the gain characteristics of the involved spot beams can use the same frequency band. Such users are then termed *co-channel users*. It is the beamwidth and the sidelobe level of the spot beam gain patterns that here come into wear (cf. Fig. 3.6).

Although the signals of co-channel users share the same spectrum they can nevertheless be discriminated at the receiver, if the spatial separation of the signals is sufficient, whereby the spot beams act as spatial filters. Hence, one can talk in this case of SDMA and in Chap. 5 methods will be described and analysed that exploit the spatial dimension of the signals in a most efficient way. SDMA leads to an increased spectrum efficiency and, consequently, to a higher system capacity for a given bandwidth available to the satellite system.

A group of neighbouring cells in that different frequency bands have to be employed to avoid too high a CCI is termed *cluster*, and their number *cluster size* (cf. Fig. 3.7). For CDMA a cluster size of 1 can be chosen, which means that the same frequencies are used in every cell. For FDMA and TDMA typical values of the cluster size in satellite systems are 4 or 7 [LWJ00], usually being significantly less than the total number of cells.

If we denote the bandwidth allocated to a single cell to serve the traffic inside this cell with B_{c}, then an overall bandwidth B_{sat} must be allocated to the satellite according to

$$B_{\mathrm{sat}} = B_{\mathrm{c}} K_{\mathrm{s}}, \tag{3.25}$$

where K_{s} is the cluster size [LWJ00]. The *frequency reuse factor* says how often a particular frequency band is reused (on average) in the service area of a satellite or satellite constellation.

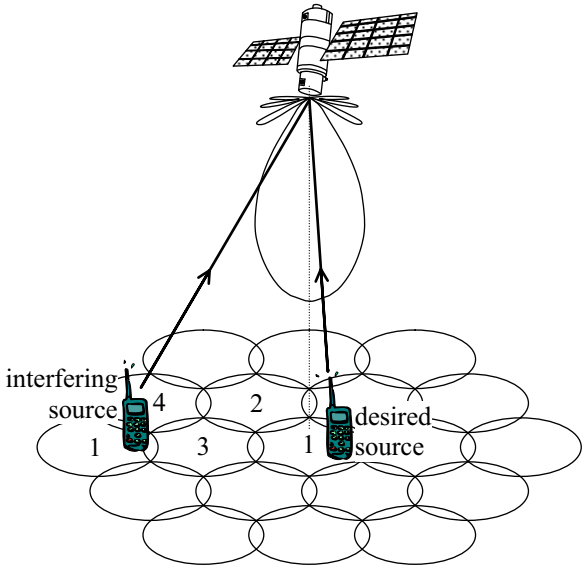


Figure 3.6: When sources in co-channel cells are active, interference is caused by receiving the signal of co-channel sources through sidelobes or even the main lobe (depending on the assumed cluster size) of the spot beam associated with the desired source signal. Depicted is the interference scenario for the uplink, further, numbers in cells denote the allocated frequency bands.

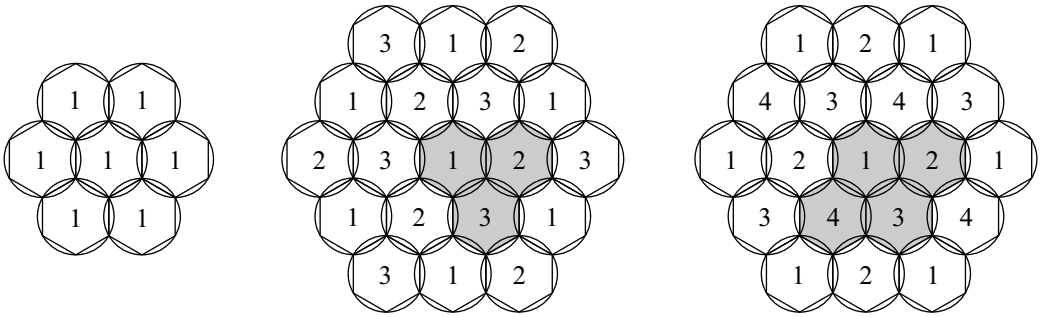


Figure 3.7: Cell pattern with indication of cluster sizes 1, 3, and 4. Numbers in cells denote the frequency bands.

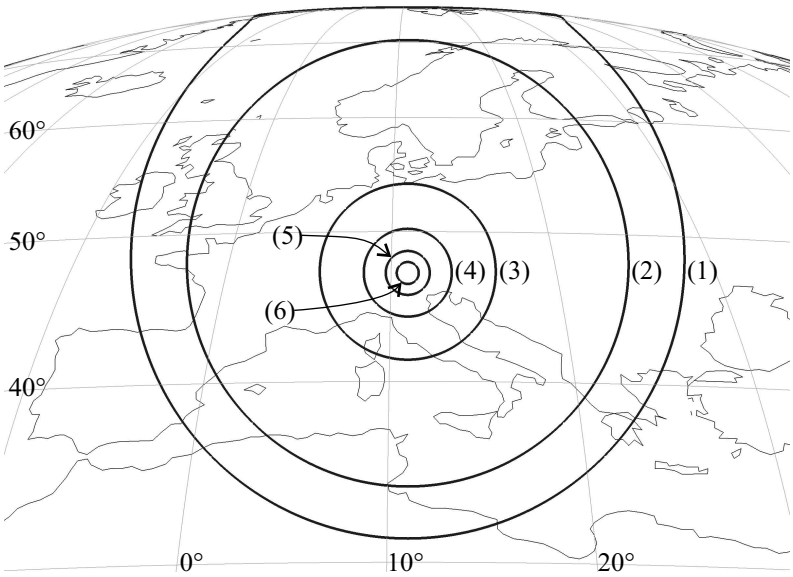


Figure 3.8: Comparison of spot beam size (3 dB below maximum contour is shown) for a GEO satellite at (11°East, 0°). Diameter of satellite antenna is $d_\mathrm{a} = 3.5$ m, further $\eta_\mathrm{a} = 0.6$. (1) $f = 1.6$ GHz, $\vartheta_\mathrm{3dB} = 1.88°$, $G = 33.2$ dB; (2) $f = 2$ GHz, $\vartheta_\mathrm{3dB} = 1.5°$, $G = 35.1$ dB; (3) $f = 5$ GHz, $\vartheta_\mathrm{3dB} = 0.6°$, $G = 43$ dB; (4) $f = 10$ GHz, $\vartheta_\mathrm{3dB} = 0.3°$, $G = 49.1$ dB; (5) $f = 20$ GHz, $\vartheta_\mathrm{3dB} = 0.15°$, $G = 55.1$ dB; (6) $f = 40$ GHz, $\vartheta_\mathrm{3dB} = 0.075°$, $G = 61.1$ dB.

The *spectral efficiency* (also: *spectrum efficiency*) denotes the efficiency of a system to provide the required traffic capacity C_s (in Erlang) to the communications system while occupying a certain bandwidth $B_\mathrm{sat} = B_\mathrm{c} K_\mathrm{s}$ and is defined according to [LWJ00]

$$\eta_\mathrm{s} = \frac{C_\mathrm{s}}{B_\mathrm{c} K_\mathrm{s} A_\mathrm{s}}, \tag{3.26}$$

where A_s is the geometric area of the coverage area. Accordingly the unit of spectrum efficiency is $\mathrm{Erl}/(\mathrm{MHz\ km^2})$ [LWJ00]. Therefore, spectrum efficiency increases as the cluster size decreases.

In Fig. 3.8 the cell sizes for different frequencies for constant antenna diameter $d_\mathrm{a} = 3.5$ m are shown for a GEO satellite at (11°East,0°) (antenna boresight is aiming at (11°East,47°North)). Assuming the 3.5 m diameter satellite antenna, it is obvious that regional satellite services employing a GEO satellite can achieve a considerable frequency reuse only at frequencies in the Ku band and beyond. To obtain sufficiently small spot beam widths at lower frequencies the required antenna diameter lies beyond $d_\mathrm{a} = 10$ m (at approx. $d_\mathrm{a} = 12$ m) for L-band frequencies (1610–1626.5 MHz) as is shown in Fig. 3.9. In fact, the operative Thuraya satellite system (2 GEO satellites, subscriber links in L-band) employs 12 m diameter deployable satellite antennas [SDR+02], creating a 256 cell pattern with a spot beam width of $\vartheta_\mathrm{3dB} \approx 0.5°$, proving the technological feasibility of antennas of such size.

In [CCF+92] it is envisaged that in the L-band and with $d_\mathrm{a} = 3$ m only three beams are required to cover Western Europe (reusing the frequencies in the two outer cells), while at 40 GHz and with a $d_\mathrm{a} = 0.5$ m diameter satellite antenna 40 beams are required providing the possibility for a much higher frequency reuse.

Finding optimal reuse patterns (sets of cells that can use the same frequencies) that minimise the cluster size and hence maximise the spectrum efficiency while guaranteeing a tolerable level of

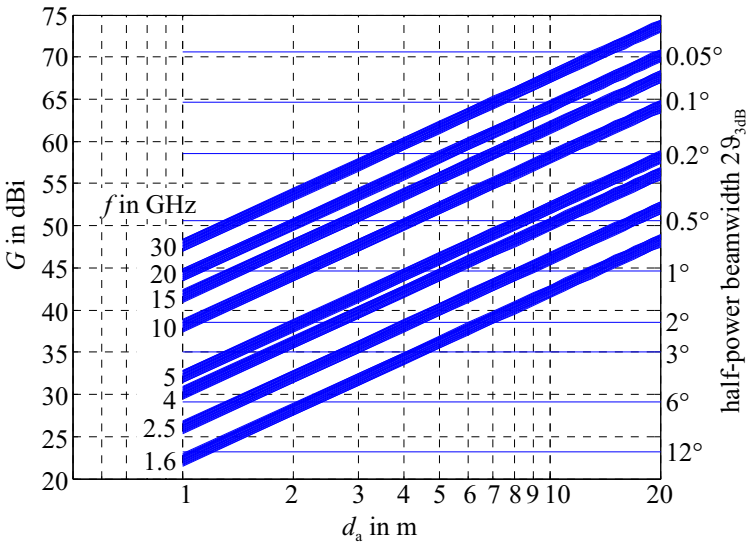


Figure 3.9: Maximum antenna gain G and antenna diameter d_a for antenna efficiencies $0.5 \leq \eta \leq 0.7$ and various carrier frequencies f. Further shown is the corresponding approximate two-sided beamwidth $2\vartheta_{3\,\mathrm{dB}}$ according to (3.24).

co-channel interference is not a trivial task for the general case. In [Jah99] stochastic optimisation using genetic algorithms has been successfully used for solving this problem.

3.4 Satellite Antenna Array Model

In the foregoing sections parameters characterising an arbitrary antenna have been introduced, further, peculiarities of satellite antennas arising from the considered satellite scenario have been discussed, such as multiple-beam antenna and the related potential for frequency reuse. However, no particular satellite antenna design was required, rather it was sufficient to consider the beamwidth and the antenna diameter alone. Finally, we have now to focus on a particular type of satellite antennas, the direct radiating array (DRA) [LL96].

Most generally speaking, for an antenna array the receive signals of 2 or more spatially distributed antennas (in the following denoted as *array elements*) are combined to produce the array output [ST81].

There are many options for how to combine the element signals and Fig. 3.10 shows schematically an array where the receive signals from the different array elements are filtered before the output signal is created by summation.

The process of creating an output from the receive signals of the array elements is termed *beamforming*. The radiation pattern created by the beamforming will be referred to simply as a *beam*. Finally, the part of the receiver that is responsible for processing and combining the array element signals is called *beamformer*.

At this point an important property of antenna arrays becomes evident, which is that the radiation pattern of the array created by the beamforming can be controlled by introducing variable filter functions (e.g. phase shifts) before summation of the array element signals (cf. Fig. 3.10). Note

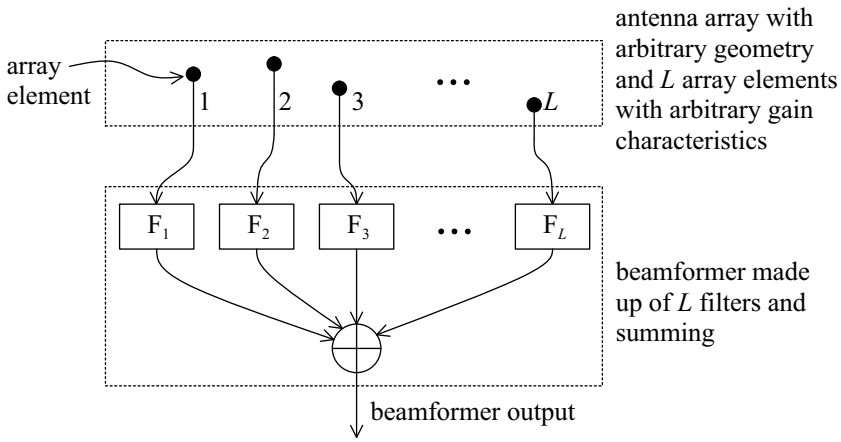


Figure 3.10: A generic antenna array with L array elements. The receive signals from the array elements are passed through L filters before the array output is generated by summation.

that different beams can be created simultaneously from the array element receive signals if several beamformers are employed. If the beamformers are implemented via digital signal processing, no signal degradation occurs for an arbitrary number of beams.

Depending on the way the output signal is generated, different types of antenna arrays can be defined. Firstly, an antenna array can be used to produce fixed beams. Secondly, an antenna array can be used to provide various degrees of adaptivity [ST81]:

Phased Array An array of many antenna elements with the phase and, generally, the amplitude of each element being variable. This allows control of the main beam direction and beam pattern shape.

Adaptive Array An *adaptive array* is a phased array which is controlled using knowledge of the signal environment, i.e. steering the main beam toward a desired signal while placing nulls in the direction of undesired interfering signals.

Smart Antenna An adaptive array where the output signals of the array elements are sampled, digitised and, subsequently processed in the digital domain (e.g. by a computer, digital signal processor etc.). This is the most versatile implementation of an adaptive antenna, because the control of the beam pattern is mainly limited by the particular signal processing algorithm implemented on the digital processor and the available processing power (e.g. measured in floating point operations per second (FLOPS)).

The emphasis in this work will be put on smart antennas, i.e. in particular it will be assumed that beamforming is implemented by digital signal processing.

With smart antennas, it is possible to exploit most efficiently the spatial distribution of the incident signals. This means, it is possible to adapt the array gain pattern to the directions from which the various signals (wanted or interfering) arrive. In particular, the antenna gain in the direction of the interfering signal can be reduced, while the antenna gain for the desired signal is kept at a high level.

Thus, the fundamental advantage of a smart antenna is the ability to increase the signal-to-interference-and-noise-power-ratio (SINR) by reducing the degradation of the desired signal due

to interference while maintaining a high gain for the desired signal. (In general, the higher link quality achieved by an increased SINR may not be the intended goal of using adaptive antennas, but may be traded for an increase in the total capacity of a communication system.)

Generally, it can be stated from what was said above in this section that an antenna array is characterised by the following factors:

- Array element characteristics.
- Antenna array geometry.
- Structure of the beamformer.

In the following we will focus on these points, primarily aiming at a short review of the derivation of directivity and gain for antenna arrays. In particular, the influence of the array geometry and array element gain patterns on the array performance will be discussed.

To simplify the discussion we will consider the case that an unmodulated carrier of frequency f_c is received by the array, and accordingly the wavelength of the carrier will be denoted as λ_c. Further, the far-field assumption is applied (cf. (3.4)), i.e. the signal arrives at the array in a plane wave front. The case of non-zero signal bandwidth, as it is encountered for modulated information bearing signals, will be discussed in Sec. 3.5.2.

In the following it will become obvious that the characterisation of an antenna array in terms of directivity and gain, respectively, is in a sense not appropriate for receiving arrays. This is because the definition of directivity does not take care of any noise sources that may be present *before* the beamforming. This is in contrast to the standard assumption in array signal processing [MM80].

In fact, although this discrepancy between the underlying antenna array models is well known, it seems that this is still an area under investigation, as quite recent publications on this topic suggest, e.g. [SC01]. The difference and similarities between both views of antenna arrays will be pointed out in the following section.

3.4.1 Array Element Radiation Pattern

The antenna array is assumed to consist of L identical array elements. Each array element is characterised by the same gain pattern $G_e(\vartheta, \varphi)$ or, respectively, complex magnitude pattern $g_e(\vartheta, \varphi)$ (cf. Sec. 3.2.4).

The gain pattern $G_e(\vartheta, \varphi)$ of the array elements is further assumed to be that of a uniformly illuminated circular aperture of diameter d_e. Therefore the gain pattern is circular symmetric and, w.l.o.g., the coordinate system is chosen such that $G_e(\vartheta, \varphi) = G_e(\vartheta)$ (accordingly $g_e(\vartheta, \varphi) = g_e(\vartheta)$) holds. The boresight of the array element is at $\vartheta = 0$.

Then the gain pattern $G_e(\vartheta)$ is given by (follows from (3.11) and (3.14), where the expression for the field pattern $F(\vartheta, \varphi)$ can be found in, e.g., [ST81])

$$G_e(\vartheta) = |g_e(\vartheta)|^2 = \eta_{r,e} \left(\frac{\pi d_e}{\lambda} \right)^2 \left(\frac{2 \mathrm{J}_1 \left(\frac{\pi d_e}{\lambda} \sin(\vartheta) \right)}{\frac{\pi d_e}{\lambda} \sin(\vartheta)} \right)^2 , \tag{3.27}$$

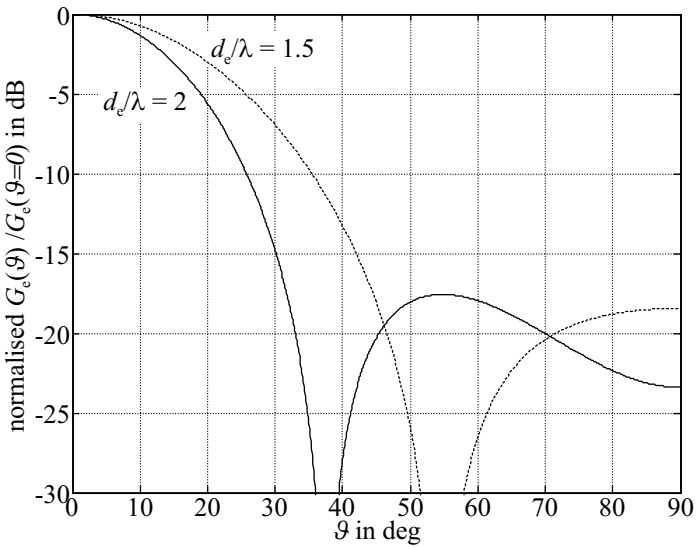


Figure 3.11: Normalised array element gain pattern $G_{\mathrm{e}}(\vartheta)/G_{\mathrm{e}}(\vartheta = 0)$ for $d_{\mathrm{e}}/\lambda = 1.5$ and $d_{\mathrm{e}}/\lambda = 2$.

where $\eta_{\mathrm{r,e}}$ denotes the radiation efficiency of the array element as defined in (3.15), and $\mathrm{J}_1(x)$ is the Bessel function of the first kind and first order.

The gain pattern $G_{\mathrm{e}}(\vartheta)$ is plotted in Fig. 3.11, normalised to the maximum $G_{\mathrm{e}} = \eta_{r,e}(\pi d_{\mathrm{e}}/\lambda)^2$ obtained for boresight at $\vartheta = 0$. Finally, the directivity of an array element will be denoted with $D_{\mathrm{e}}(\vartheta)$, and it holds $D_{\mathrm{e}}(\vartheta) = G_{\mathrm{e}}(\vartheta)/\eta_{\mathrm{r,e}}$ according to (3.14).

3.4.2 Antenna Array Geometry

In this work only planar DRA antennas are considered where the array elements are all located in the same plane and no reflectors (e.g. parabolic) are employed [LL96].

Without loss of generality it is assumed that the array elements are located in the x-y-plane. The geometry of the planar array is then given by the element positions

$$\tilde{\mathbf{p}}_{\mathrm{e},l} = \begin{pmatrix} p^{\mathrm{x}}_{\mathrm{e},l} \\ p^{\mathrm{y}}_{\mathrm{e},l} \\ 0 \end{pmatrix} \in \mathbb{R}^3, \qquad l = 1, 2, \ldots, L, \tag{3.28}$$

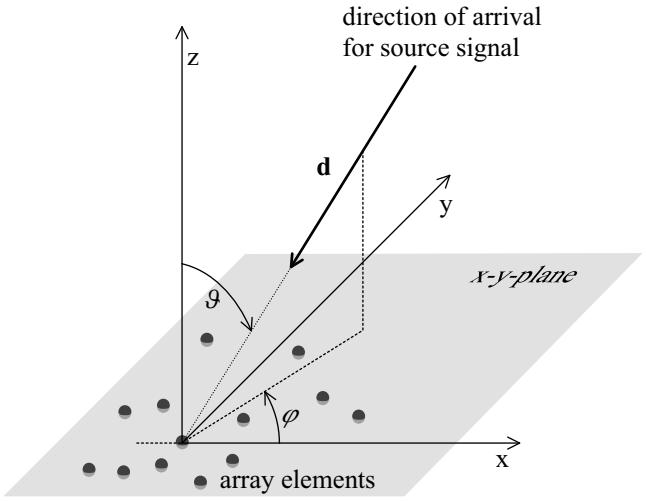
which define the coordinates of the phase centres of the L array elements.

Further it is assumed that a signal transmitted by a source impinges on the antenna array from the corresponding DOA defined by angles (ϑ, φ) (cf. Fig. 3.12(a)).

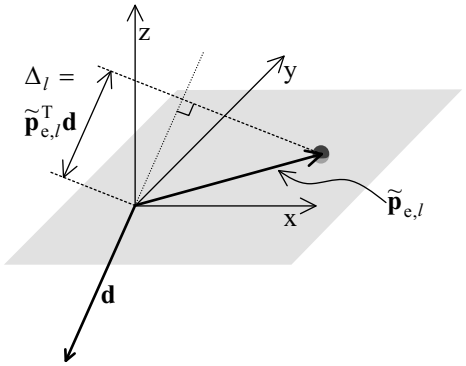
Due to the spatial distribution of the array elements the source signal arrives with different delays at the different array element phase centres, whereas the delays depend solely on the array geometry and on the DOA of the signal.

Hereby it is sufficient to consider delays relative to an arbitrary reference point[2] and we choose the origin of the coordinate system as reference. Then the delays τ_l with respect to the reference point

[2]Usually, the reference point is arbitrary and it even needs not to coincide with a point inside the array area, but there are exceptions. The ESPRIT direction finding algorithm, which relies on certain properties of the underlying array (centro-symmetry), requires selection of a *particular* reference point [Haa96].



(a) Definition of the DOA by angles (ϑ, φ). The planar array comprises L array elements, w.l.o.g. the elements are arranged in the x-y-plane.



(b) Path length difference Δ_l with respect to reference point for array element l is given by the scalar product $\tilde{\mathbf{p}}_{\mathrm{e},l}^{\mathrm{T}}\mathbf{d}$, where $\mathbf{d}$ denotes the incident direction vector defined by the DOA of the source signal, and $\tilde{\mathbf{p}}_{\mathrm{e},l}$ is the location of array element l.

Figure 3.12: Definition of direction of arrival (ϑ, φ) for the source signal (a). Further illustration of calculation the propagation path length difference Δ_l (b). The L array elements are all located in a plane.

are given by (note that the delays can be negative, cf. Fig. 3.12(b))

$$\tau_l = \frac{\Delta_l}{\mathrm{c}}, \tag{3.29}$$

where c denotes speed of light (we will always assume $c = c_0$, with c_0 being the vacuum speed of light) and Δ_l is the propagation path length difference with respect to the origin, which is defined as [God97a]

$$\Delta_l = \tilde{\mathbf{p}}_{\mathrm{e},l}^{\mathrm{T}}\mathbf{d} \quad \text{with} \quad \mathbf{d} = -\begin{pmatrix} \sin\vartheta\cos\varphi \\ \sin\vartheta\sin\varphi \\ \cos\vartheta \end{pmatrix}, \tag{3.30}$$

with $\tilde{\mathbf{p}}_{\mathrm{e},l}$ being the position of the l-th array element, and $\mathbf{d}$ being a unit length vector defined by the direction of arrival of the source signal.

To simplify notation in the following, the planar array element position vector

$$\mathbf{p}_{\mathrm{e},l} = \begin{pmatrix} p_{\mathrm{e},l}^{\mathrm{x}} \\ p_{\mathrm{e},l}^{\mathrm{y}} \end{pmatrix} \in \mathbb{R}^2, \qquad l = 1, 2, \ldots, L, \tag{3.31}$$

is introduced, which is obtained from $\tilde{\mathbf{p}}_{\mathrm{e},l}$ by simply dropping the zero z-coordinate.

The *spatial frequency* $\boldsymbol{\nu}$ of a signal arriving at the array from direction (ϑ, φ) is given by

$$\boldsymbol{\nu} = \frac{1}{\lambda}\begin{pmatrix} u \\ v \end{pmatrix}, \tag{3.32}$$

where u, v are the sine space coordinates introduced in (3.19).

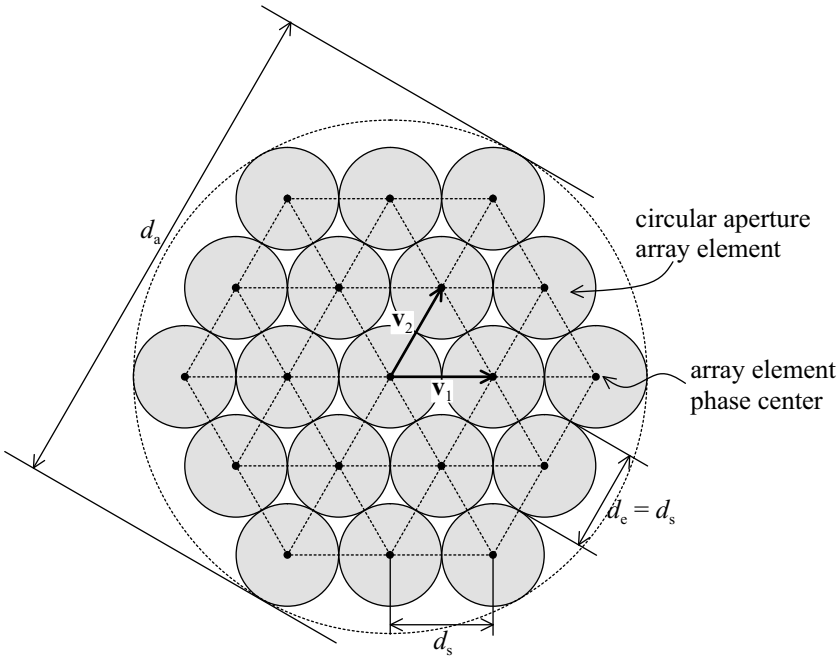


Figure 3.13: Hexagonal array geometry where the phase centres of the array elements are located on a regular triangular lattice (example shows $L = 19$ array elements). The diameter d_e of the circular aperture array elements is equal to phase centre spacing d_s. In this work only directly radiating hexagonal arrays comprising array elements with circular aperture are considered.

Finally, for an unmodulated carrier of frequency f_c, the delays τ_l translate with (3.28) – (3.32) to phase shifts [MM80]

$$a_l \triangleq \mathrm{e}^{-j2\pi f_\mathrm{c}\tau_l} = \mathrm{e}^{-j2\pi \mathbf{p}_{\mathrm{e},l}^\mathrm{T}\boldsymbol{\nu}}. \tag{3.33}$$

The vector perpendicular to the array plane is the array boresight and it is assumed that the boresight axes of the array elements are all aligned with the array boresight axis. The array has a hexagonal contour and the phase centres of the array elements are arranged on a regular triangular lattice as depicted in Fig. 3.13. The triangular lattice on which the phase centres of the array elements are located is defined by the basis $\mathbf{V}$ [DM84], which is given by

$$\mathbf{V} = d_\mathrm{s} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix}, \tag{3.34}$$

where the columns of $\mathbf{V}$, $\mathbf{v}_1$ and $\mathbf{v}_2$, are the basis vectors of the triangular lattice (cf. Fig. 3.13 and App. D), and d_s denotes the phase centre spacing. With (3.34) the array element positions can be given according to

$$\mathbf{p}_{\mathrm{e},l} = \mathbf{V}\mathbf{i}_l, \tag{3.35}$$

with $\mathbf{i}_l \in \mathbb{Z}^2$ being the integer position vector of element l on the array lattice.

The gain pattern of the antenna array depends on the gain pattern of the array elements $G_\mathrm{e}(\vartheta)$, on the array geometry (i.e. on the arrangement of the array elements) and on the beamforming. In particular, grating lobes may occur in real space, if the phase centre separation exceeds approx. $\lambda/2$, i.e. $d_\mathrm{s} > \lambda/2$ [Kra88, CL96]. This will be discussed in the following section.

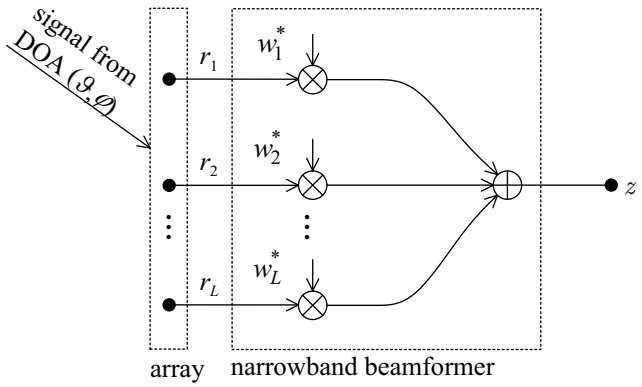


Figure 3.14: Schematic representation of narrowband beamforming for the source signal by multiplying the array element receive signals r_l with complex weights w_l^*.

3.4.3 Narrowband Beamforming

Up to this point we have introduced assumptions on array element patterns and array geometry which will hold throughout this work. Finally, we have to turn to the question how the array output is established from the receive signals from the L array elements.

As it was mentioned before, we initially consider an unmodulated signal and in this case the narrowband beamformer is applicable [MM80] (again, refer to Sec.3.5.2 where it will be discussed under which conditions the narrowband assumption is sufficiently fulfilled for modulated signals of non-zero bandwidth).

The general structure of the narrowband beamformer is depicted in Fig. 3.14. Note that no noise sources are included in the array model at this point, because we aim firstly at a characterisation of the antenna array in terms of directivity and gain, respectively, which are defined independent of any noise.

Furthermore, collecting the beamforming weights w_l in a vector, we define the *beamforming weight vector* $\mathbf{w} \in \mathbb{C}^L$, with

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{pmatrix}, \tag{3.36}$$

and, accordingly the *steering vector* $\mathbf{a} \in \mathbb{C}^L$ is obtained by collecting the phase factors a_l of (3.33) in a vector

$$\mathbf{a}(\vartheta, \varphi) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_L \end{pmatrix}. \tag{3.37}$$

It is well known (see standard antenna literature, e.g. [ST81]) that for an antenna array, comprising array elements having identical magnitude patterns $g_{\mathrm{e}}(\vartheta)$, the array field pattern $f_{\mathrm{ea}}(\vartheta, \varphi)$ is provided by the *pattern multiplication* theorem, yielding

$$f_{\mathrm{ea}}(\vartheta, \varphi) = g_{\mathrm{e}}(\vartheta) f_{\mathrm{a}}(\vartheta, \varphi), \tag{3.38}$$

where the factor

$$\begin{aligned} f_{\mathrm{a}}(\vartheta,\varphi) &= \sum_{l=1}^{L} w_l^* \mathrm{e}^{-j2\pi \mathbf{p}_{\mathrm{e},l}^{\mathrm{T}}\boldsymbol{\nu}} \\ &= \mathbf{w}^{\mathrm{H}}\mathbf{a}(\vartheta,\varphi), \end{aligned} \tag{3.39}$$

is termed the *array factor* of the array [LL96]. Based on (3.38) the array directivity is defined.

Array Directivity

Finally, with the array field pattern $f_{\mathrm{ea}}(\vartheta,\varphi)$ given by (3.38) the equation for directivity $D_{\mathrm{a}}(\vartheta,\varphi)$ of the antenna array is therefore readily provided by (cf. (3.11))

$$\begin{aligned} D_{\mathrm{a}}(\vartheta,\varphi) &= \frac{|f_{\mathrm{ea}}(\vartheta,\varphi)|^2}{\frac{1}{4\pi}\int_{\varphi'=0}^{2\pi}\int_{\vartheta'=0}^{\pi}|f_{\mathrm{ea}}(\vartheta',\varphi')|^2\sin(\vartheta')\,d\vartheta' d\varphi'} \\ &= \frac{D_{\mathrm{e}}(\vartheta)|\mathbf{w}^{\mathrm{H}}\mathbf{a}(\vartheta,\varphi)|^2}{\frac{1}{4\pi}\int_{\varphi'=0}^{2\pi}\int_{\vartheta'=0}^{\pi}D_{\mathrm{e}}(\vartheta')|\mathbf{w}^{\mathrm{H}}\mathbf{a}(\vartheta',\varphi')|^2\sin(\vartheta')\,d\vartheta' d\varphi'}. \end{aligned} \tag{3.40}$$

Accordingly, the gain pattern $G_{\mathrm{a}}(\vartheta,\varphi)$ of the array is given by (cf. (3.14))

$$G_{\mathrm{a}}(\vartheta,\varphi) = \eta_{\mathrm{r,a}} D_{\mathrm{a}}(\vartheta,\varphi), \tag{3.41}$$

where $\eta_{\mathrm{r,a}}$ is the overall radiation efficiency of the array, i.e. including the radiation efficiency of the array elements $\eta_{\mathrm{r,e}}$ (cf. (3.27)) as well as additional losses introduced in the beamforming.

As already mentioned in the introductory part of Sec. 3.4, the beam pattern can be controlled by choosing the weight vector $\mathbf{w}$. The beamforming can in this way be implemented either statically, e.g. to create a fixed main beam pointing towards a constant direction, or dynamically, then the weight vector is adapted to a changing signal scenario from time to time, e.g. to steer the main lobe of the array to a desired direction by choosing $\mathbf{w} = \mathbf{a}(\vartheta,\varphi)$. Further $\mathbf{w}$ can be chosen to control the sidelobe level of the spot beams to reduce the received power from interfering signal sources (spatial filtering for interference reduction) [MM80, God97b, Gay02].

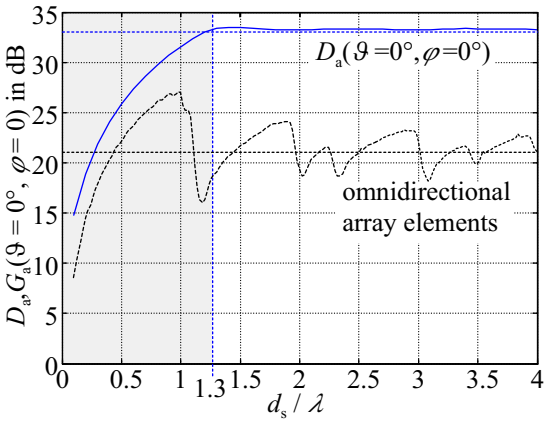
In general, directivity and gain, respectively, of an antenna array depend strongly on the array element locations $\mathbf{p}_{\mathrm{e},l}$ for a given total number L of array elements. Depending on the radiation characteristics of the array elements, the integral in the denominator of (3.40) and thus $D_{\mathrm{a}}(\vartheta,\varphi)$ possibly varies strongly with the array element phase centre separation d_{s} and also with $\mathbf{w}$. The same holds for the appearance of grating lobes.

Impact of Element Pattern and Separation on Array Directivity If for the array element phase centre separation it holds $d_{\mathrm{s}} \leq \lambda/2$, then typically directivity is independent of $\mathbf{w}$ and no grating lobes appear. In fact, it is often encountered in literature dealing with smart antennas in communications systems to choose this case, possibly to avoid difficulties coming with the case of $d_{\mathrm{s}} > \lambda/2$. Also for satellite scenarios the assumption of $\lambda/2$-spacing of the array elements is sometimes used in this connection, e.g. [GG95, LC00].

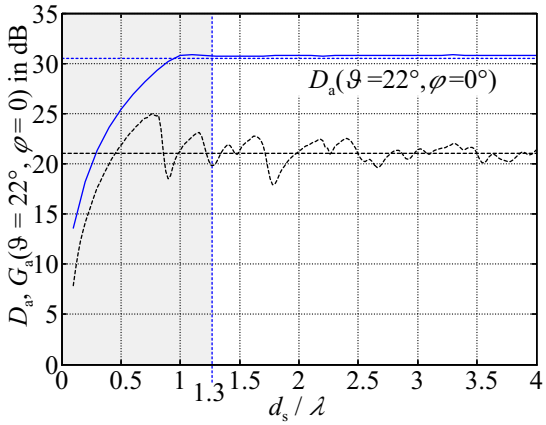
However, in particular for the satellite application it usually holds for the satellite array antenna that $d_s > \lambda/2$ (e.g. [IHLA98, CCAL01]).

Therefore, the impact of $d_s > \lambda/2$ on array directivity has to be discussed.

Especially for omnidirectional array elements the array directivity strongly varies with d_s as it is demonstrated in Fig. 3.15(a) for a hexagonal array of $L = 127$ omnidirectional elements, where the phase centre separation d_s is varied in a range $d_s/\lambda = 0 \ldots 4$.



(a) Main beam is steered towards antenna boresight $\vartheta = 0°$. Note that the directivity is virtually constant for $d_s \geq d_e = 1.3\lambda$ for the array of circular aperture elements, while it varies significantly for the array of omnidirectional elements.



(b) Main beam is steered towards ICO satellite edge of coverage $\vartheta = 22°$.

Figure 3.15: Dependency of array directivity $D_a(\vartheta, \varphi)$ on array element separation d_s for a hexagonal array comprising $L = 127$ elements. As d_s increases the directivity of the array of omnidirectional array elements approaches L, and directivity of the array of circular aperture elements approaches $LD_e(\vartheta, \varphi)$.

In contrast, for the antenna array with circular aperture array elements (diameter $d_e = 1.3\lambda$), directivity is virtually constant for $d_s \geq d_e$ (the range of $d_s < d_e$, where the apertures of adjacent array elements would overlap is highlighted grey in Fig. 3.15)). Fig. 3.15(b) shows array directivity for a weight vector $\mathbf{w}$ that is chosen such that the main beam is steered towards $\vartheta = 22°$ (this is the off-boresight angle for edge of coverage for an ICO satellite [MS98]).

A possible interpretation of the oscillations of directivity with respect to d_s in case of omnidirectional elements is that the effective apertures of neighbouring elements overlap[3], where additionally this overlap depends not only on the array element separation, but also on the considered angular direction (ϑ, φ). Due to overlap of effective apertures, the directivity $D_e(\vartheta, \varphi) = 1$ of an isolated omnidirectional array element is not applicable anymore for the array element located in the vicinity of other array elements.

Obviously, the directivity of the array of circular aperture elements reaches $LD_e(\vartheta, \varphi)$ for $d_s \geq d_e$, where then the directivity does not vary much, which is in contrast to the case of omnidirectional array elements. This can be interpreted in that the effective aperture of a circular array element

[3]Personal communication. Dr. Dreher, DLR e.V., Institute for High Frequency and Radar Systems, November 2003.

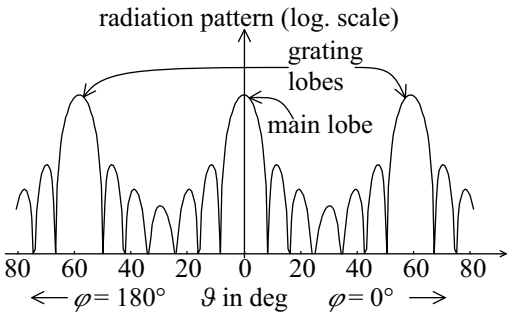


Figure 3.16: Schematic representation of grating lobes for the main lobe being steered toward $\vartheta = 0$.

is well approximated by its geometric aperture. Therefore, no overlap of the effective apertures can occur and the directivity $D_{\mathrm{e}}(\vartheta, \varphi)$ of the isolated array element is not altered by arranging the elements in an array.

From this we conclude that the different array element receive signals can be treated independently from each other, regardless of the particular value of $d_{\mathrm{s}} \geq d_{\mathrm{e}}$.

Grating Lobes

Assume now that the spatial frequency $\boldsymbol{\nu}_0$ (according to (3.32)) corresponds to the main lobe, i.e. the angular direction (ϑ_0, φ_0) for that array factor $f_{\mathrm{a}}(\vartheta, \varphi)$ according to (3.39) is maximised for given weight vector $\mathbf{w}$.

Due to the 2π-periodicity of the complex exponential function in (3.39) and depending on the array element spacing there may be angular directions for that the array factor shows side lobes of the same amplitude as for the main lobe. These extra main lobes are referred to as *grating lobes* (cf. Fig. 3.16) [MM80, ST81].

The spatial frequencies $(\boldsymbol{\nu}_0 + \boldsymbol{\nu}^*)$ for that grating lobes occur are obtained from (3.35) and (3.39) according to

$$2\pi \mathbf{i}_l^{\mathrm{T}} \mathbf{V}^{\mathrm{T}} (\boldsymbol{\nu}_0 + \boldsymbol{\nu}^*) = 2\pi \mathbf{i}_l^{\mathrm{T}} \mathbf{V}^{\mathrm{T}} \boldsymbol{\nu}_0 + k2\pi \Rightarrow \mathbf{i}_l^{\mathrm{T}} \mathbf{V}^{\mathrm{T}} \boldsymbol{\nu}^* = k, \qquad k \in \mathbb{Z}. \tag{3.42}$$

Therefore for $\boldsymbol{\nu}^*$ it must hold

$$\boldsymbol{\nu}^* = \left(\mathbf{V}^{\mathrm{T}}\right)^{-1} \mathbf{i}^*, \qquad \mathbf{i}^* \in \mathbb{Z}^2, \tag{3.43}$$

where

$$\mathbf{V}^* = \left(\mathbf{V}^{\mathrm{T}}\right)^{-1} = \frac{1}{d_{\mathrm{s}}} \begin{pmatrix} 1 & 0 \\ -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{3}} \end{pmatrix} \tag{3.44}$$

is the *periodicity matrix* in the spatial frequency domain (cf. App. D) [DM84] and $\mathbf{V}$ was provided in (3.34) for a triangular lattice.

Finally, the angles (ϑ, φ) where the grating lobes occur are obtained from $(\boldsymbol{\nu}_0 + \boldsymbol{\nu}^*)$ using (3.19) and (3.32):

$$\begin{aligned} \vartheta &= \arcsin \sqrt{(u_0 + u^*)^2 + (v_0 + v^*)^2} \\ \varphi &= \operatorname{arctan4}\left(v_0 + v^*, u_0 + u^*\right), \end{aligned} \tag{3.45}$$

with

$$\begin{pmatrix} u_0 + u^* \\ v_0 + v^* \end{pmatrix} = \lambda \left(\boldsymbol{\nu}_0 + \boldsymbol{\nu}^* \right) = \lambda \left(\boldsymbol{\nu}_0 + \left(\mathbf{V}^{\mathrm{T}} \right)^{-1} \mathbf{i}^* \right), \qquad \mathbf{i}^* \in \mathbb{Z}^2. \tag{3.46}$$

Note that there is an infinite number of grating lobes in the spatial frequency domain, but only those grating lobes are in real space for that $(u_0 + u^*)^2 + (v_0 + v^*)^2 \leq 1$ such that ϑ according to (3.45) is real [CL96].

The discussion of grating lobes will be revisited in Sec. 3.6.1 and 3.6.2 where two particular satellite systems and the respective satellite array antenna models are introduced.

3.5 Signal Model

Recalling the definition equations of antenna array directivity (3.40) and gain (3.41), respectively, it becomes obvious that an important factor in any communications system was not required for these definitions, namely noise.

As depicted in Fig. 3.17, a common antenna array model assumes that independent noise sources are present in each element receive path[4], and in consequence the noise power present at the output of the beamformer is actually a function of the beamforming weight vector $\mathbf{w}$, cf., e.g., [MM80, God97b, LL96].

The noise sources $n_l(t)$ are mutually uncorrelated and modelled by AWGN, where the noise power will be denoted with

$$\mathrm{E}\left\{ |n_l(t)|^2 \right\} = 2\sigma_{\mathrm{n}}^2, \qquad \forall l = 1, 2, \ldots, L \tag{3.47}$$

(we choose to write the noise power in this way to emphasise that the noise is modelled by a proper complex Gaussian random variable, i.e. with the real and imaginary part being independent, zero-mean, and both having equal variance σ_{n}^2 (cf. Sec. 4.3)).

In the following, an array signal model including the satellite channel, introduced in Sec. 2.2, will be presented. Finally, we have to consider the multi-user case, which follows in Sec. 3.5.3.

3.5.1 Single Source Signal Model

Signals will be treated in the complex base band [Pro95]. For a generic real RF signal $x_{\mathrm{RF}}(t)$ of bandwidth B_{s} (one-sided) around the centre frequency f_{c}, the complex base band representation is given by $x(t)$ with

$$x_{\mathrm{RF}}(t) = \mathrm{Re}\left\{ x^+(t) \right\}, \quad \text{with} \quad x^+(t) = x(t) \mathrm{e}^{j2\pi f_{\mathrm{c}} t}, \tag{3.48}$$

where $x^+(t)$ denotes the *analytical signal*, and is the complex baseband signal $x(t)$ shifted to centre frequency f_{c} (note that the spectrum of $x^+(t)$ is therefore only defined for positive frequencies).

[4] For this antenna array model the signal-to-noise-power-ratio (SNR) at the output of the antenna array may *not* be proportional to the antenna array directivity [SC01].

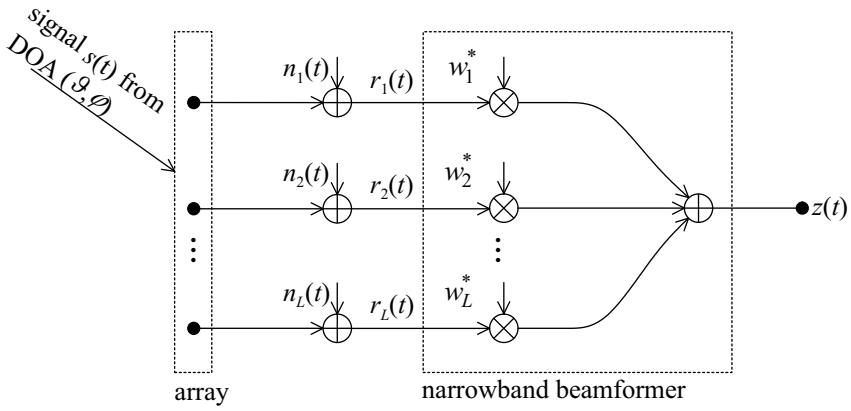


Figure 3.17: Schematic representation of narrowband beamforming for the source signal by multiplying the array element receive signals r_l with complex weights w_l^*. Noise sources $n_l, (l = 1, 2, \ldots, L)$ model white Gaussian receiver noise, which is present at each array element receive branch. Note the presence of noise sources in the receive branch of each element (cf. Fig. 3.14 where no noise sources are considered for the definition of array directivity).

Let a signal source m emit signal $s_m(t)$ (complex baseband signal). In the following the carrier frequency has to be taken into account therefore the analytical signal

$$s_m^+(t) = s_m(t)\mathrm{e}^{j2\pi f_c t} \tag{3.49}$$

will be used.

As it has been pointed out in Sec. 3.4.3, a source signal $s_m^+(t)$ arrives with different delays at the different array elements due to their spatial distribution. The delays $\tau_{l,m}$ have been provided in (3.29) with (3.30), and have been given to $\tau_{l,m} = \Delta_{l,m}/\mathrm{c}$.

With the above definitions and using (2.11), the receive signal $r_l^+(t)$ (cf. (3.48)) of array element l can be written according to

$$r_l^+(t) = s_m^+(t) * h_{l,m}'^+(t,\tau) + n_l^+(t), \tag{3.50}$$

where $h_{l,m}'^+(t,\tau)$ denotes the channel impulse response applying for the l-th array element and the m-th source, further $n_l^+(t)$ is the thermal noise introduced at the l-th array element.

Assuming that the array elements have all identical characteristics, i.e. all have identical amplitude pattern $g_\mathrm{e}(\vartheta)$, the channel impulse response $h_{l,m}'^+(t)$ can be written according to

$$h_{l,m}'^+(t,\tau) = h_m'^+(t,\tau) * \delta(t - \tau_{l,m}), \tag{3.51}$$

which means that the channel can be separated into the contribution of the transmission channel $h_m'^+(t,\tau)$ from source m to the array reference point and relative delays $\tau_{l,m}$ among the different array elements. Therefore, $h_m'^+(t,\tau)$ includes all satellite communications propagation channel effects for source signal m (i.e. free space loss, and, further, shadowing, delay spread, Doppler spread, Doppler shift[5], and also gains of the transmit and the receive antenna [Jah94, Jah99]),

[5]It will be assumed that the Doppler shift is tracked and (pre-)compensated in the transmitter and/or the receiver.

whereas $\delta(t-\tau_{l,m})$ takes the relative propagation delays among the various antenna array elements into account.

In the further, the signal model will be constrained in that we will assume that the propagation channel is a frequency non-selective fading channel as discussed in Sec. 2.2.1, i.e.

$$s_m^+(t) * h_m'^+(t,\tau) = s_m^+(t) h_m'^+(t). \tag{3.52}$$

Then (3.51) in (3.50) yields

$$r_l^+(t) = \underbrace{\tilde{r}_m(t-\tau_{l,m})\mathrm{e}^{j2\pi f_c(t-\tau_{l,m})}}_{=\tilde{r}_m^+(t-\tau_{l,m})} + n_l^+(t), \tag{3.53}$$

with the complex baseband signal (cf. (2.16) and (2.17))

$$\tilde{r}_m(t) = s_m(t)\frac{h_m(t)}{\sqrt{\mu_m}}, \tag{3.54}$$

which holds according to Sec. 2.2.1, where it was pointed out that the propagation channel can be described by a quasi-constant component $\sqrt{\mu_m}$ with (cf. (3.23))

$$\mu_m = \frac{L_{\mathrm{bf},m}}{G_\mathrm{T} G_\mathrm{e}(\vartheta_m)} \frac{c_\mathrm{R}}{1+c_\mathrm{R}}. \tag{3.55}$$

and a possibly rapidly varying fading factor $h_m(t)$ (normalised to unit power).

The $\tau_{l,m}$, as well as the signal DOAs (ϑ_m, φ_m) must, in general, be considered time-variant, as a mobile ground terminal, LEO and MEO satellites change their position over time. However, to simplify notation an explicit designation of time dependency of the DOA and the relative delays is omitted.

Finally, the complex baseband array element signal is obtained from (3.53) using (3.48), and given by (cf. Fig. 3.18)

$$r_l(t) = \underbrace{\tilde{r}_m(t-\tau_{l,m})\mathrm{e}^{-j2\pi f_c \tau_{l,m}}}_{\triangleq\, v_{l,m}(t)} + n_l(t), \tag{3.56}$$

where we have further introduced the array element receive signal without noise $v_{l,m}(t)$.

3.5.2 Narrowband Signal Model for Array Signal Processing

The narrowband signal model is a well known concept in array signal processing (cf. [MM80]), therefore we will omit a detailed derivation.

Firstly, the *array propagation time* is required, defined as

$$\Delta\tau = \max_{l_1,l_2,m} |\tau_{l_1,m} - \tau_{l_2,m}|, \tag{3.57}$$

which denotes the maximum delay difference that can occur between any two array elements.

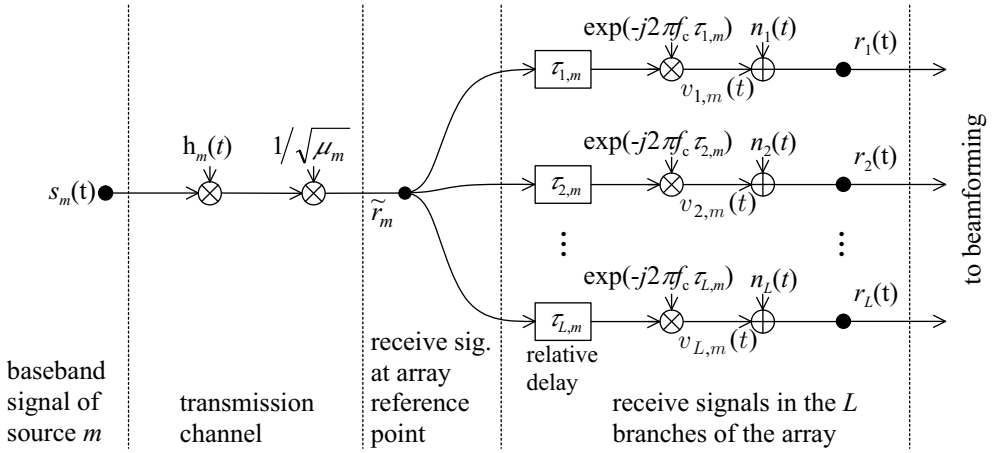


Figure 3.18: Schematic representation of the definition of the array element receive signals $r_l(t)$. Source m transmits signal $s_m(t)$ over the multipath channel characterised by the multiplicative factor $h_m(t)/\sqrt{\mu_m}$. The resulting signal $\tilde{r}_m$ is then received by the array elements with different delays $\tau_{l,m}$. The relative delays (w.r.t. array reference point) are negligible if the narrowband assumption holds, then $\tilde{r}_m(t-\tau_{l,m}) \approx \tilde{r}_m(t)$.

Then the narrowband assumption is fulfilled if it holds

$$\Delta\tau B_\mathrm{s} \ll 1, \tag{3.58}$$

where B_s is the one-sided bandwidth of the incident signal. The relation given by (3.58) is known as the *narrowband assumption for array signal processing*[6] [MM80, God97b] and says that the array propagation time $\Delta\tau$ is negligible.

If the narrowband assumption holds, the relative signal delays with respect to the reference point translate to phase shifts $\mathrm{e}^{-j2\pi f_\mathrm{c}\tau_{l,m}}$, i.e. the relative delays as depicted in Fig. 3.18 can be neglected. In consequence, $v_{l,m}(t)$ in (3.56) simplifies to

$$v_{l,m}(t) = a_{l,m}\tilde{r}_m(t), \tag{3.59}$$

where

$$a_{l,m} = \mathrm{e}^{-j2\pi f_\mathrm{c}\tau_{l,m}} = \mathrm{e}^{-j2\pi\frac{\Delta_{l,m}}{\lambda_\mathrm{c}}}. \tag{3.60}$$

Therefore, we can rewrite (3.56) as

$$\boxed{r_l(t) = \tilde{r}_m(t)a_{l,m} + n_l(t).} \tag{3.61}$$

Obviously, according to (3.57) validity of the narrowband assumption depends not only on signal bandwidth B_s, but also on the array size and the range of DOAs (ϑ_m, φ_m) which can occur, as these factors determine the maximum range of array propagation time $\Delta\tau$ in (3.57).

In the next section it will be verified that narrowband assumption for array signal processing indeed normally holds in satellite scenarios.

[6]The term *narrowband assumption* is in a way misleading, as by means of the channel transfer function the signal may well be broadband, i.e. the channel may be frequency selective.

Validity of Narrowband Assumption in Satellite Scenarios

Recalling the requirement for narrowband beamforming that the bandwidth-aperture-propagation-delay-product must be negligible, i.e. $B_\mathrm{s}\Delta\tau \ll 1$ (cf. (3.58)), we can write that

$$B_\mathrm{s}\Delta\tau \approx B_\mathrm{s}\frac{d_\mathrm{a}\sin\vartheta_\mathrm{max}}{\mathrm{c}} = \frac{B_\mathrm{s}}{f_\mathrm{c}}\frac{d_\mathrm{a}}{\lambda_\mathrm{c}}\sin\vartheta_\mathrm{max} \ll 1 \tag{3.62}$$

must hold for the applicability of narrowband beamforming (f_c is the centre frequency, λ_c is the wave length at f_c, and d_a is the antenna diameter).

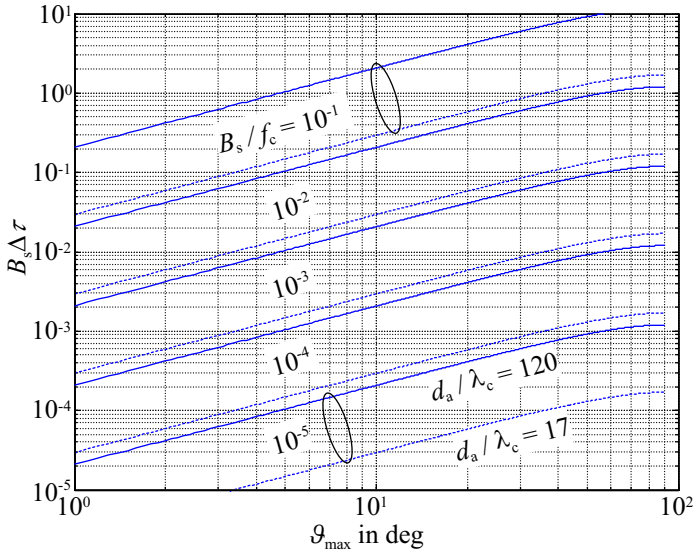


Figure 3.19: Bandwidth-aperture-propagation-delay-product $B_\mathrm{s}\Delta\tau$ versus maximal occurring off-boresight angle ϑ_max for various relative bandwidths $B_\mathrm{s}/f_\mathrm{c}$, and for $d_\mathrm{a}/\lambda_\mathrm{c} = 120$ (solid lines) and $d_\mathrm{a}/\lambda_\mathrm{c} = 17$ (dashed lines).

Fig. 3.19 shows the bandwidth-aperture-propagation-delay-product $B_\mathrm{s}\Delta\tau$ versus the maximal occurring off-boresight angle ϑ_max for various relative bandwidths $B_\mathrm{s}/f_\mathrm{c}$ and for two values of $d_\mathrm{a}/\lambda_\mathrm{c}$, namely $d_\mathrm{a}/\lambda_\mathrm{c} = 120$ (typical for GEO satellite) and $d_\mathrm{a}/\lambda_\mathrm{c} = 17$ (typical for MEO satellite).

Considering maximal off-boresight angles of $\vartheta_\mathrm{max} \approx 9°$ for GEO satellites and $\vartheta_\mathrm{max} \approx 20°$ for MEO satellites [LWJ00], it can be concluded that the narrowband assumption (3.58) is applicable for relative bandwidths up to an order of magnitude of $B_\mathrm{s}/f_\mathrm{c} \approx 10^{-3}$ for both a GEO satellite, as well as for a MEO satellite (assuming that $B_\mathrm{s}\Delta\tau \approx 10^{-2}$ sufficiently fulfills (3.58)).

3.5.3 M Source Signal Model

We will now extend the case of a single source treated above to the general M source signal case. To simplify notation, the dependency on time t will be omitted in the notation in the following.

In Fig. 3.20 the principle characterisation of the M source signal model is sketched, summarising the assumptions presented in Sec. 2.2 and 3.4.

Fig. 3.21 schematically depicts the signal model which will be discussed in detail below.

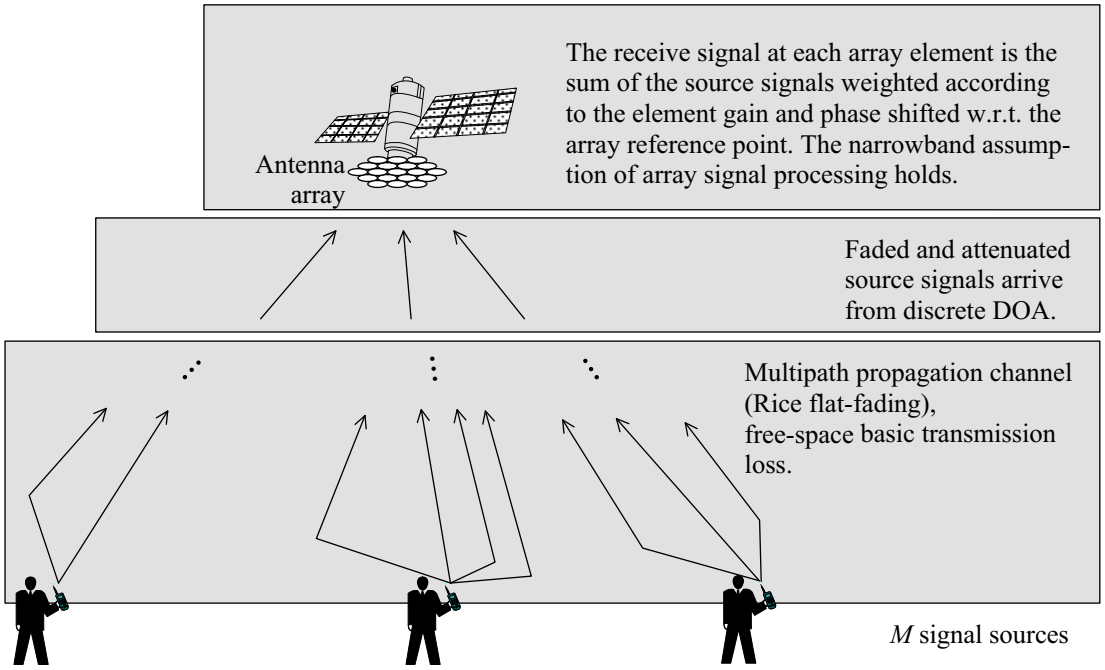


Figure 3.20: Summary of the assumptions regarding signal propagation. The propagation environment in the vicinity of the sources creates a multipath propagation channel resulting in fading. Due to geometry the signal of each source arrive approximately from a single discrete direction-of-arrival (DOA). Finally, the narrowband assumption of array signal processing holds.

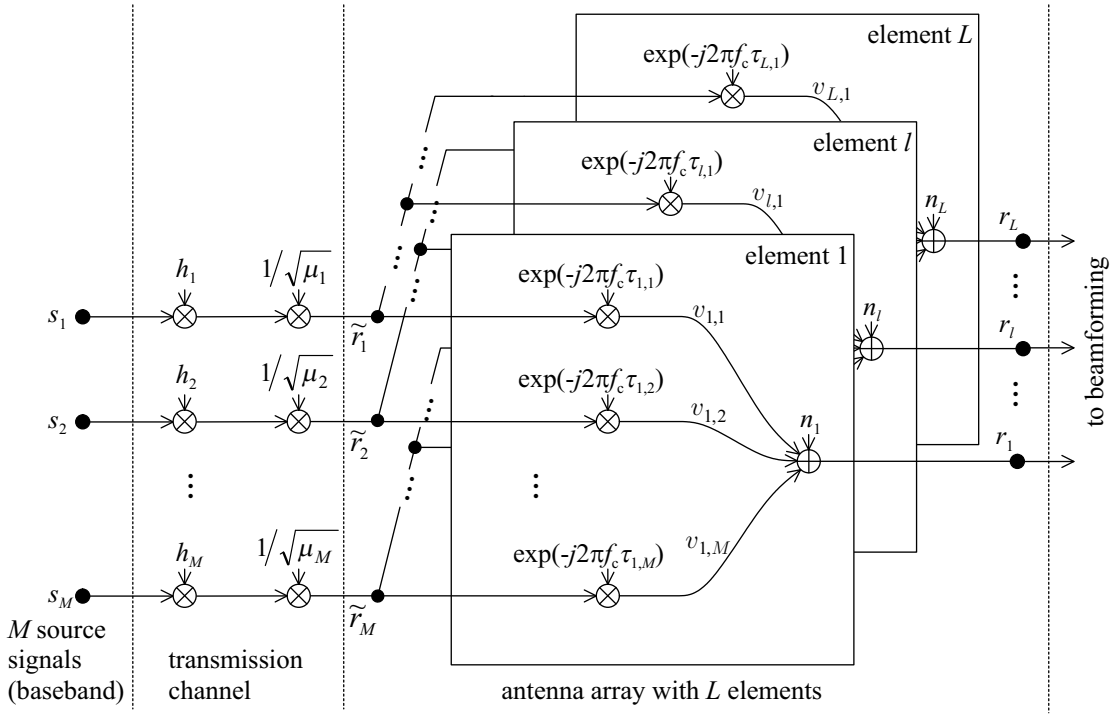


Figure 3.21: Schematic representation of the definition of the array element receive signals r_l for the narrowband case and M sources. (Note that the segmentation in transmission channel and antenna array is not strict, as the μ_m include the source antennas and array element gains.)

If M signals are received at the antenna array and assuming linear system properties, the element receive signal r_l at the output of array element l is given by summing the contributions of each source m (cf. (3.61)) resulting in

$$r_l = \sum_{m=1}^{M} v_{l,m} + n_l = \sum_{m=1}^{M} a_{l,m}\tilde{r}_m + n_l. \tag{3.63}$$

Further, the following definitions shall hold.

Collecting the array element receive signals in a vector yields the *array element receive signal vector*

$$\mathbf{r} = (r_1, r_2, \ldots, r_L)^{\mathrm{T}}, \tag{3.64}$$

further, $\mathbf{A} \in \mathbb{C}^{L\times M}$ denotes the *steering matrix* defined according to

$$\mathbf{A} = (\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_M), \tag{3.65}$$

with the M *steering vectors* $\mathbf{a}_m \in \mathbb{C}^L$,

$$\mathbf{a}_m = (a_{1,m}, a_{2,m}, \ldots, a_{L,m})^{\mathrm{T}} \tag{3.66}$$

and the *array input signal vector* $\tilde{\mathbf{r}} \in \mathbb{C}^M$,

$$\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_M)^{\mathrm{T}} \tag{3.67}$$

and, finally, the noise vector $\mathbf{n} \in \mathbb{C}^L$,

$$\mathbf{n} = (n_1, n_2, \ldots, n_L)^{\mathrm{T}}. \tag{3.68}$$

Applying the definition of the phase factors $a_{l,m}$ provided in (3.60) on (3.66) the steering vector of source m is given by

$$\mathbf{a}_m = \left(\mathrm{e}^{-j2\pi f_c \tau_{1,m}}, \mathrm{e}^{-j2\pi f_c \tau_{2,m}}, \ldots, \mathrm{e}^{-j2\pi f_c \tau_{L,m}}\right)^{\mathrm{T}}. \tag{3.69}$$

Using the above definitions and with the multiplicative fading assumption (3.61), (3.63) can be rewritten in matrix notation, yielding

$$\boxed{\mathbf{r} = \underbrace{\mathbf{AHs}}_{\triangleq\,\mathbf{v}} + \mathbf{n}.} \tag{3.70}$$

Here we have additionally introduced the diagonal channel matrix $\mathbf{H} \in \mathbb{C}^{M\times M}$, with

$$\mathbf{H} = \begin{pmatrix} \frac{h_1}{\sqrt{\mu_1}} & 0 & \cdots & 0 \\ 0 & \frac{h_2}{\sqrt{\mu_2}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{h_M}{\sqrt{\mu_M}} \end{pmatrix}, \tag{3.71}$$

and the *source signal vector* $\mathbf{s} = (s_1, s_2, \ldots, s_M)^{\mathrm{T}}$, and finally the *receive signal vector* $\mathbf{v} \in \mathbb{C}^M$.

Adopting the matrix notation, Fig. 3.21 can be redrawn according to Fig. 3.22, representing finally the signal model that shall be valid throughout this work.

For the remainder of this work the following assumptions shall hold.

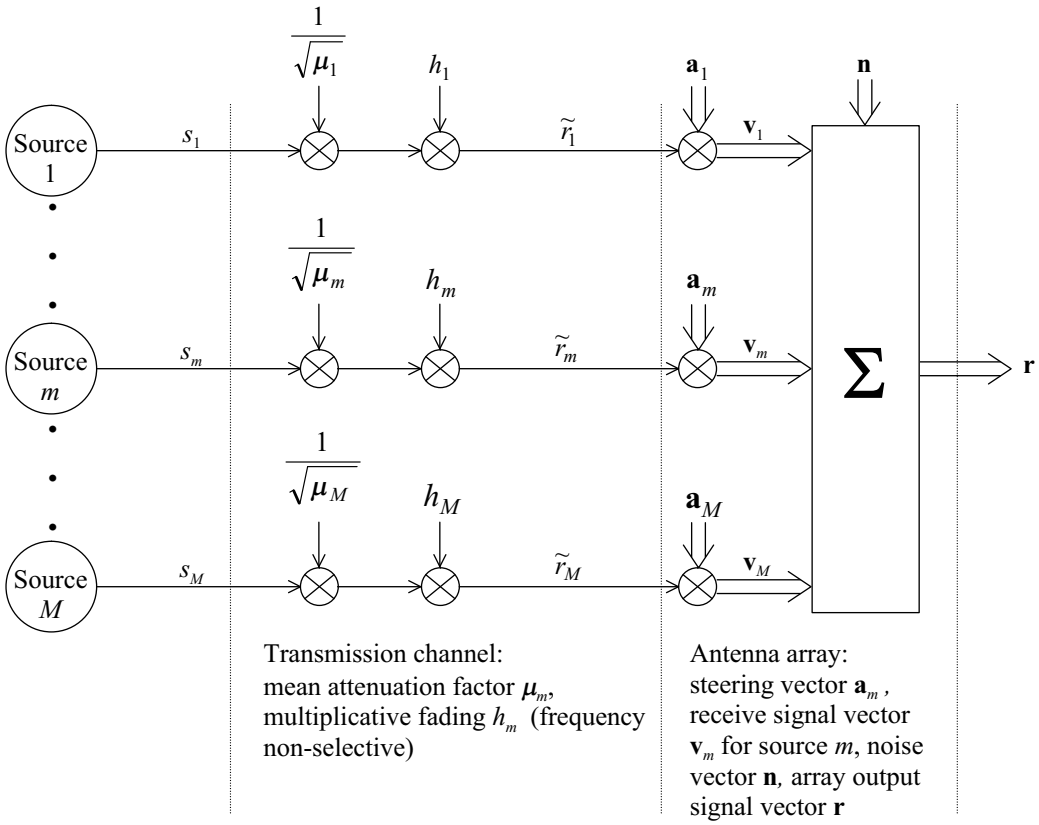


Figure 3.22: Block diagram of the vector multiple-access fading channel (double arrows indicate vectors), which can be separated into the influence of the transmission channel and the array. (Note that the gains of the antenna array elements are included in the transmission channel via the μ_m, cf. (3.55).) The link attenuation factors μ_m and the steering vectors $\mathbf{a}_m$ (L elements) are assumed fixed during transmission, while the fast-fading factors h_m are changing. The noise vector $\mathbf{n}$ takes care of both noise in the receiver, as well as noise received by the antenna.

Definitions of Required Covariance Matrices

Both the signals s_m, as well as the noise are ergodic processes, therefore time and group averages are the same.

The source signals are mutually uncorrelated, then

$$\mathrm{E}\left\{s_m^* s_i\right\} = \begin{cases} p_m & m = i \\ 0 & m \neq k \end{cases}, \quad m, i = 1, 2, \ldots, M, \tag{3.72}$$

where p_m is the transmit power of source m (hence, the p_m correspond to P_T introduced in Sec. 3.2.4).

Then the *source signal covariance matrix* $\mathbf{K_s} \in \mathbb{R}_+^{M \times M}$ is a diagonal matrix, with

$$\mathbf{K_s} = \mathrm{E}\left\{\mathbf{s}\mathbf{s}^\mathrm{H}\right\} = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & p_M \end{pmatrix} \quad \text{(uncorrelated source signals).} \tag{3.73}$$

The AWGN noise vector $\mathbf{n}$ is *spatially white* (i.e. the noise sources in the L receive branches of the array are uncorrelated, which is a common assumption, e.g. [LL96, God97b, SXLK98]), then

$$\mathrm{E}\left\{n_l^* n_k\right\} = \begin{cases} 2\sigma_\mathrm{n}^2 & l = k \\ 0 & l \neq k \end{cases}, \quad l, k = 1, 2, \ldots, L, \tag{3.74}$$

where $2\sigma_\mathrm{n}^2$ is the power of a single noise source.

Therefore, also the *noise covariance matrix* $\mathbf{K_n} \in \mathbb{C}^{L \times L}$ is diagonal with

$$\mathbf{K_n} = \mathrm{E}\left\{\mathbf{n}\mathbf{n}^\mathrm{H}\right\} = \mathbf{I}_L 2\sigma_\mathrm{n}^2 \qquad \text{(spatially white noise).} \tag{3.75}$$

However, in the following we will mostly use the more general notation $\mathbf{K_n}$ in order to provide the according equations for the general case of eventually correlated noise sources, and only if required in the argumentation we will use (3.75).

Further we define the *receive signal covariance matrix without noise* $\mathbf{K_v} \in \mathbb{C}^{L \times L}$ (Hermitian, positive semidefinite [Pap91]) according to (with $\mathbf{v}$ being defined in (3.70))

$$\mathbf{K_v} = \mathrm{E}\left\{\mathbf{v}\mathbf{v}^\mathrm{H}\right\} = \mathbf{A}\mathbf{H}\mathbf{K_s}\mathbf{H}^\mathrm{H}\mathbf{A}^\mathrm{H}. \tag{3.76}$$

Note that it holds

$$\mathbf{A}\mathbf{H}\mathbf{K_s}\mathbf{H}^\mathrm{H}\mathbf{A}^\mathrm{H} = \sum_{m=1}^{M} \mathbf{a}_m \mathbf{a}_m^\mathrm{H} \frac{|h_m|^2}{\mu_m} p_m, \tag{3.77}$$

which means that the covariance matrix $\mathbf{K_v}$ of the sum of the receive signals is the sum of the covariance matrices of each single receive signal, because the source signals are mutually uncorrelated.

Finally, with the above definitions, the *array output covariance matrix* $\mathbf{K_r} \in \mathbb{C}^{L \times L}$ (Hermitian, positive definite for $\sigma_\mathrm{n}^2 > 0$ [Pap91]), which is the covariance matrix of the receive signal and noise, is introduced and defined according to

$$\boxed{\mathbf{K_r} = \mathrm{E}\left\{\mathbf{r}\mathbf{r}^\mathrm{H}\right\} = \mathbf{K_v} + \mathbf{K_n} = \mathbf{A}\mathbf{H}\mathbf{K_s}\mathbf{H}^\mathrm{H}\mathbf{A}^\mathrm{H} + \mathbf{K_n}.} \tag{3.78}$$

The covariance matrix of signal vector $\mathbf{v}$ plus noise vector $\mathbf{n}$ is the sum of the covariance matrix of the signals $\mathbf{K_v}$ and the covariance matrix of the noise $\mathbf{K_n}$, because noise and receive signals are uncorrelated.

Additional Definitions Further, we define in the following the covariance matrices for some subset of the M sources. This does not add any new aspects to the definitions above, but it will support a clearer notation in the remaining chapters.

The set

$$\mathcal{M} = \{1, 2, \ldots, M\} \tag{3.79}$$

denotes the set containing all M source indices, further we define a subset $\mathcal{S}$ of $\mathcal{M}$,

$$\mathcal{S} \subseteq \mathcal{M}. \tag{3.80}$$

Then, let $\mathbf{A}(\mathcal{S}) \in \mathbb{C}^{L \times M_\mathcal{S}}$ denote the steering matrix that contains only the $M_\mathcal{S} = |\mathcal{S}|$ steering vectors $\mathbf{a}_i$ of the sources $i \in \mathcal{S}$.

The covariance matrix $\mathbf{K_s}(\mathcal{S}) \in \mathbb{R}_+^{M_\mathcal{S} \times M_\mathcal{S}}$ is the diagonal matrix of transmit powers p_i of sources $i \in \mathcal{S}$, and, equivalently, the channel matrix $\mathbf{H}(\mathcal{S}) \in \mathbb{C}^{M_\mathcal{S} \times M_\mathcal{S}}$ is a diagonal matrix of the channel factors $h_i/\sqrt{\mu_i}$ for sources $i \in \mathcal{S}$. Accordingly we define the receive signal covariance matrix $\mathbf{K_v}(\mathcal{S}) \in \mathbb{C}^{L \times L}$ for a subset of the M sources with

$$\mathbf{K_v}(\mathcal{S}) = \mathbf{A}(\mathcal{S})\mathbf{H}(\mathcal{S})\mathbf{K_s}(\mathcal{S})\mathbf{H}(\mathcal{S})^{\mathrm{H}}\mathbf{A}(\mathcal{S})^{\mathrm{H}}, \qquad \mathcal{S} \subseteq \mathcal{M}. \tag{3.81}$$

(Obviously it holds that $\mathbf{A}(\mathcal{M}) = \mathbf{A}$, $\mathbf{K_s}(\mathcal{M}) = \mathbf{K_s}$, $\mathbf{K_v}(\mathcal{M}) = \mathbf{K_v}$, and $\mathbf{H}(\mathcal{M}) = \mathbf{H}$.)

According to (3.77) we can therefore write for any subset $\mathcal{S}$ of $\mathcal{M}$:

$$\mathbf{K_v}(\mathcal{S}) = \mathbf{K_v}(\mathcal{T}) + \mathbf{K_v}(\bar{\mathcal{T}}), \qquad \mathcal{T} \subseteq \mathcal{S} \subseteq \mathcal{M}, \bar{\mathcal{T}} = \mathcal{S} \setminus \mathcal{T}. \tag{3.82}$$

Finally, the array output vector covariance matrix $\mathbf{K_r}(\mathcal{S})$ for a set $\mathcal{S} \subseteq \mathcal{M}$ is given according to

$$\mathbf{K_r}(\mathcal{S}) = \mathbf{K_v}(\mathcal{S}) + \mathbf{K_n}, \qquad \mathcal{S} \subseteq \mathcal{M}. \tag{3.83}$$

With these additional definitions, (3.78) can be split in the notation in three components, namely the contribution from a single source m (we will refer to this as the *wanted signal*), the contributions from the signals from sources in $\mathcal{I}_m = \mathcal{M} \setminus \{m\}$ interfering with source m, and finally the noise, i.e.

$$\mathbf{K_r} = \mathbf{K_v}(m) + \mathbf{K_v}(\mathcal{I}_m) + \mathbf{K_n}, \qquad \mathcal{I}_m = \mathcal{M} \setminus m, \tag{3.84}$$

where it holds according to (3.81) that

$$\mathbf{K_v}(m) = \mathbf{a}_m \mathbf{a}_m^{\mathrm{H}} \frac{|h_m|^2}{\mu_m} p_m \tag{3.85}$$

3.5.4 Output SINR of the Optimal Beamformer

Based on the narrowband beamformer structure (depicted in Fig. 3.14 for a single source signal) we will in the following provide the expression for the SINR at the output of an antenna array if M source signals are present at the antenna array aperture [MM80, LL96].

The output signal of the narrowband beamformer for source m, $z_m \in \mathbb{C}$, is readily given by

$$\begin{aligned} z_m &= \mathbf{w}_m^{\mathrm{H}}\mathbf{r} \\ &= \mathbf{w}_m^{\mathrm{H}}\mathbf{A}\mathbf{H}\mathbf{s} + \mathbf{w}_m^{\mathrm{H}}\mathbf{n} \\ &= \underbrace{\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_m \frac{h_m}{\sqrt{\mu_m}} s_m}_{\text{wanted signal}} + \underbrace{\sum_{\substack{i=1 \\ i\neq m}}^{M} \mathbf{w}_m^{\mathrm{H}}\mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}} s_i}_{\text{interference}} + \underbrace{\mathbf{w}_m^{\mathrm{H}}\mathbf{n}}_{\text{noise}}, \end{aligned} \tag{3.86}$$

where the complex weight vector $\mathbf{w}_m \in \mathbb{C}^L$, is defined as $\mathbf{w}_m = (w_{1,m}, w_{2,m}, \ldots, w_{L,m})^{\mathrm{T}}$. Further we have split the signal in (3.86) in wanted signal m, interference from sources $i \neq m$, and thermal noise.

The output signal z_m is further processed in receiver stages following the beamformer. This will be discussed in detail in Chap. 5 and 6.

Using (3.84) to determine the covariance of the beamformer output signal z_m for source m yields

$$\begin{aligned} \mathrm{E}\left\{z_m z_m^{\mathrm{H}}\right\} = \mathbf{w}_m^{\mathrm{H}}\mathrm{E}\left\{\mathbf{r}\mathbf{r}^{\mathrm{H}}\right\}\mathbf{w}_m = \mathbf{w}_m^{\mathrm{H}}\mathbf{K}_{\mathbf{r}}\mathbf{w}_m = \\ \underbrace{\mathbf{w}_m^{\mathrm{H}}\mathbf{K}_{\mathbf{v}}(m)\mathbf{w}_m}_{\text{wanted signal}} + \underbrace{\mathbf{w}_m^{\mathrm{H}}\mathbf{K}_{\mathbf{v}}(\mathcal{I}_m)\mathbf{w}_m}_{\text{interference}} + \underbrace{\mathbf{w}_m^{\mathrm{H}}\mathbf{K}_{\mathbf{n}}\mathbf{w}_m}_{\text{noise}}, \qquad \mathcal{I}_m = \mathcal{M}\setminus\{m\}, \end{aligned} \tag{3.87}$$

and allows to provide for source m the expression for the output SINR $\Gamma_m(\mathbf{h})$ after the beamformer for given joint fading state $\mathbf{h} = (h_1, h_2, \ldots, h_M)^{\mathrm{T}}$ according to

$$\Gamma_m(\mathbf{h}) = \frac{p_m\left|h_m\right|^2}{\mu_m} \frac{\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_m\mathbf{a}_m^{\mathrm{H}}\mathbf{w}_m}{\mathbf{w}_m^{\mathrm{H}}\left(\mathbf{K}_{\mathbf{v}}(\mathcal{I}_m) + \mathbf{K}_{\mathbf{n}}\right)\mathbf{w}_m}, \qquad \mathcal{I}_m = \mathcal{M}\setminus m \tag{3.88}$$

where further (3.85) was used (also cf. [SXLK98] where the SINR for an array is considered without fading).

The optimal weight vector $\mathbf{w}_{\mathrm{opt},m}(\mathbf{h})$, maximising $\Gamma_m(\mathbf{h})$ is given by [MM80, SXLK98]

$$\boxed{\mathbf{w}_{\mathrm{opt},m}(\mathbf{h}) = \beta\mathbf{K}_{\mathbf{r}}(\mathcal{I}_m)^{-1}\mathbf{a}_m = \beta\left(\mathbf{K}_{\mathbf{v}}(\mathcal{I}_m) + \mathbf{K}_{\mathbf{n}}\right)^{-1}\mathbf{a}_m, \qquad \mathcal{I}_m = \mathcal{M}\setminus m,} \tag{3.89}$$

where β is an arbitrary scalar that leaves $\Gamma_m(\mathbf{h})$ unchanged, as can be readily verified by plugging using (3.89) into (3.88).

Note that the optimal weight vector for source m depends also on the steering vectors $\mathbf{a}_i$ and on the instantaneous receive powers $\left|h_i\right|^2 p_i/\mu_i$ of all sources $i \in \mathcal{I}_m$ interfering with source m according to the definition of $\mathbf{K}_{\mathbf{v}}(\mathcal{S})$ (with $\mathcal{S} = \mathcal{I}_m$) in (3.81).

The maximal SINR $\Gamma_{\mathrm{opt},m}(\mathbf{h})$ achievable for source m with optimal weight vector $\mathbf{w}_{\mathrm{opt},m}(\mathbf{h})$ is obtained by using $\mathbf{w}_m = \mathbf{w}_{\mathrm{opt},m}(\mathbf{h})$ in (3.88) and is hence given by (cf. [MM80, SXLK98])

$$\boxed{\Gamma_{\mathrm{opt},m}(\mathbf{h}) = \frac{p_m\left|h_m\right|^2}{\mu_m 2\sigma_{\mathrm{n}}^2}\mathbf{a}_m^{\mathrm{H}}\left(\sum_{i\in\mathcal{I}_m}\mathbf{a}_i\mathbf{a}_i^{\mathrm{H}}\frac{p_i\left|h_i\right|^2}{\mu_i 2\sigma_{\mathrm{n}}^2} + \frac{1}{2\sigma_{\mathrm{n}}^2}\mathbf{K}_{\mathbf{n}}\right)^{-1}\mathbf{a}_m, \qquad \mathcal{I}_m = \mathcal{M}\setminus m,} \tag{3.90}$$

where we have also used (3.77).

To ease notation later on, we introduce the instantaneous element (or input) SNR $\gamma_m(h_m)$ at array element level before the beamformer according to

$$\gamma_m(h_m) = \frac{p_m \left|h_m\right|^2}{\mu_m 2\sigma_{\mathrm{n}}^2}, \tag{3.91}$$

and the mean input SNR γ_m given by

$$\boxed{\gamma_m = \mathrm{E}\left\{\gamma_m(h_m)\right\}_{h_m} = \frac{p_m}{2\sigma_{\mathrm{n}}^2 \mu_m} = p_m G_{\mathrm{T}} \frac{G_{\mathrm{e}}(\vartheta_m)}{2\sigma_{\mathrm{n}}^2 L_{\mathrm{bf},m}} \left(1 + \frac{1}{c_{\mathrm{R}}}\right).} \tag{3.92}$$

The product $p_m G_{\mathrm{T}}$ is the EIRP of source m (the EIRP is the power required for an isotropic radiator to produce the same power flux-density at the receiver as the antenna with gain G_{T}).

Because the optimal weight vector $\mathbf{w}_{\mathrm{opt},m}(\mathbf{h})$ depends on the joint fading state $\mathbf{h}$ it must be re-computed at the rate of change of the fading to maintain optimality, i.e. with approximately rate $1/T_{\mathrm{coh}}$. This raises the question of the loss in the SINR, if the weight vectors are computed based on *average* receive powers p_m/μ_m ($m \in \mathcal{M}$) to eventually reduce the update rate requirement. However, a more detailed discussion of this is presented in App. C (concepts and definitions are used there that are introduced not before Chap. 4).

We conclude this section with a simple example illustrating optimal beamforming.

Example 3.1: **Optimal Beamforming with 2 Sources** Here we want to shortly address how the optimal beamformer performs for the limiting cases that, on the one hand, thermal noise, and, on the other hand, interference is dominating the output SINR. The simple scenario of two sources is considered (source signal 1 at $0°$, signal 2 at $5°$), with a 10-element linear array of omnidirectional elements and $\lambda/2$ element spacing. Source 1 is considered the wanted-signal, while source 2 is the interferer.

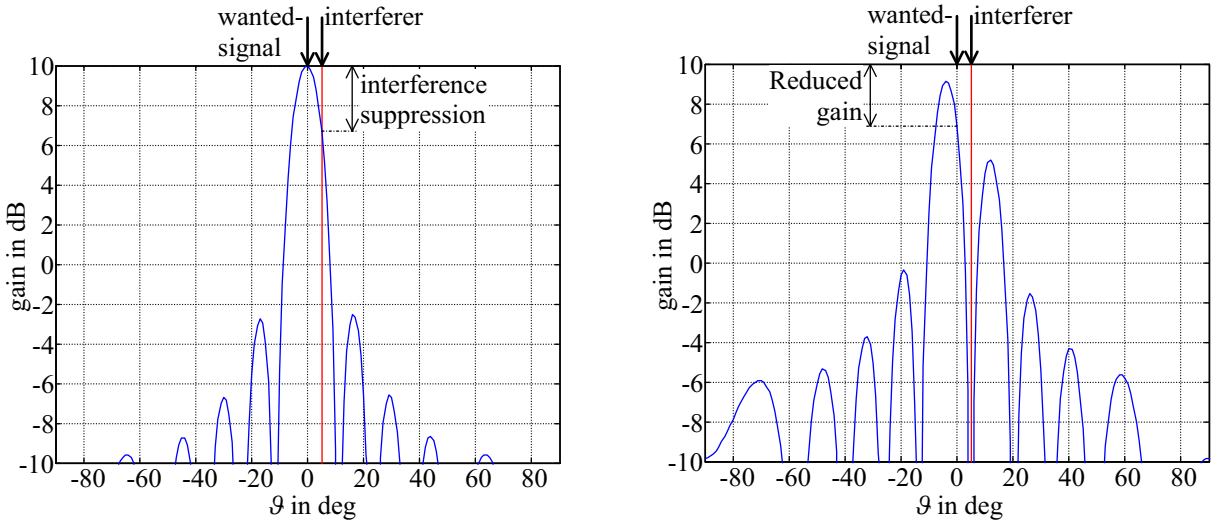
Firstly, considering that thermal noise is dominant ($p_2 \ll \sigma_{\mathrm{n}}^2$), the optimal beamformer aims at maximising the beamforming gain towards the wanted-signal, while interference is allowed to have only little suppression (Fig. 3.23(a)).

If interference is the decisive element ($p_2 \gg \sigma_{\mathrm{n}}^2$), then the optimal beamformer places a pattern null at the interfering signal, even at the cost of reducing significantly the gain for the wanted-signal (cf. Fig. 3.23(b)).

If the optimal beamformer places a pattern null at the interfering signal[7], it is obvious that the output SINR no longer depends on the interference power. Instead, the output SINR is given by the power of the wanted-signal and the gain for the wanted-signal, where this gain is a function of the array geometry and the source distribution.

■

[7] In the limiting case where interference tends to infinity, the optimal beamformer places a pattern null at the direction of interference, then the optimal beamformer performs identical to the null steering beamformer, on the other hand, if interference is zero, then the optimal performer corresponds to the conventional beamformer (see e.g. [God97b])



(a) Thermal noise dominant. The SINR is maximised mainly by pointing the peak gain towards the wanted signal, while interference is allowed to have a significant gain as well.

(b) Interference dominant. The SINR is maximised by spatial filtering of interference, while it is tolerated to reduce the gain towards the wanted signal.

Figure 3.23: Gain pattern for optimal beamforming, if either thermal noise or interference is dominant. Wanted-signal at $\vartheta = 0°$, interferer at $\vartheta = 5°$.

3.6 Satellite System Scenarios

Although we will aim to provide results of general validity, we have to restrict the investigations at first to particular scenarios, because we want to create realistic interference scenarios. Such interference scenarios have to consider realistic satellite antenna array parameters, but also a reasonable distribution of mutually interfering sources.

In the following two scenarios are described in detail that will be basis of the further investigations.

3.6.1 MEO Satellite Scenario: the ICO Satellite System

The ICO satellite system was designed in the mid-1990's to provide global voice and low-rate data services via a constellation of satellites in medium earth orbit (MEO). In the following important system parameters of ICO will be presented, representing the status of the ICO system design as published 1999 [GST99]).[8]

The ICO space segment comprises a constellation of 10 satellites in the MEO. The orbit altitude amounts to approx. $h_\mathrm{S} = 10390\ \mathrm{km}$, and accordingly the orbit time is $T_\mathrm{o} = 6\ \mathrm{h}$. The satellites are arranged in two planes of five satellites each (cf. Fig 3.24). However, in this work only a single satellite will be considered. Each satellite covers approximately 20%–30% of the earth's surface, where the coverage areas of neighbouring satellites overlap to provide path diversity (satellite diversity) via multiple satellites [LWJ00].

[8] ICO corporation planned originally to commence service in the year 2000, however, because of the commercially disastrous fate of the Iridium satellite system that provides voice services since 1998 [Vat91], ICO corporation is undertaking a redesign of the system with the date of start of service now being unknown.

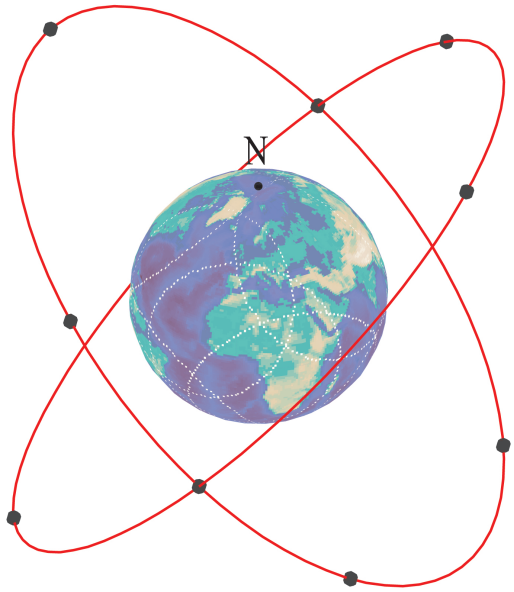
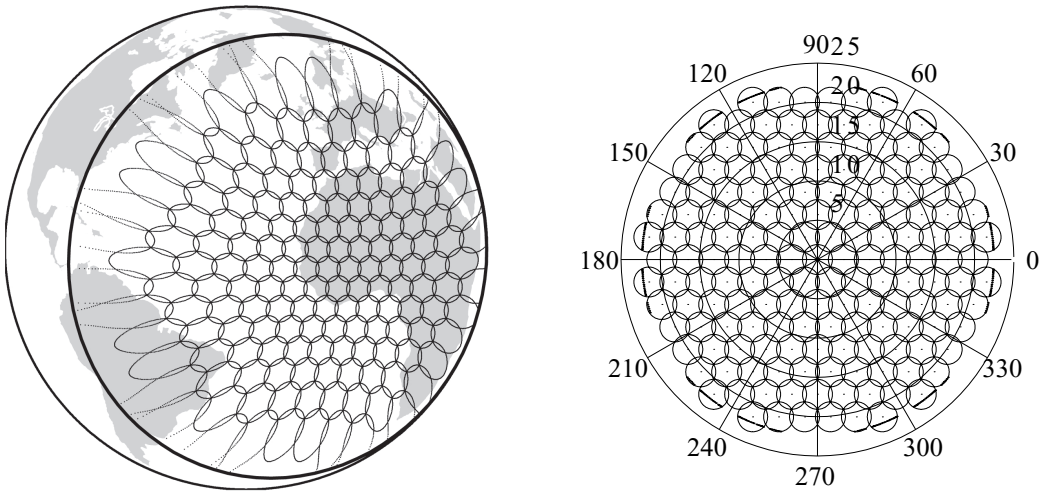


Figure 3.24: View of ICO satellite constellation comprising 10 MEO satellites in two planes (taken from [LWJ00]).

The service area of a satellite is completely covered with 163 cells (cf. Fig. 3.25), where the cell boundaries are defined according to the 3 dB-below-maximum spot beam gain contour. The cell pattern for uplink and downlink in the mobile link are congruent.



(a) Spot beam pattern shown in projection on earth surface.

(b) Spot beam pattern shown in antenna ϑ-φ-coordinates.

Figure 3.25: ICO spot beam pattern comprising 163 cells, 3 dB below maximum gain contours are shown (taken from [LWJ00]).

The 163 spot beams for the mobile link are realised with DRA antennas comprising 127 array elements, where two separate DRA are employed for uplink and downlink (cf. Fig. 3.26). Further, the 163 spot beams are created by digital fixed beamforming.

A frequency reuse with cluster size 4 is realised to maximise the spectral efficiency by optimal use of the mobile link spectrum (cf. Sec.3.3.3 and Fig. 3.27). The side lobe level of each spot

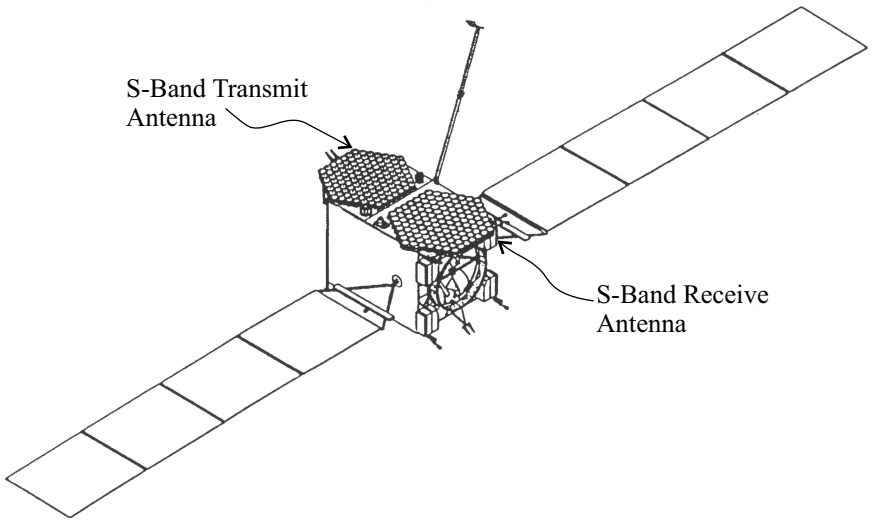


Figure 3.26: View of ICO satellite (taken from [MS98]). The S-band transmit and receive arrays for the mobile link comprise 127 array elements each.

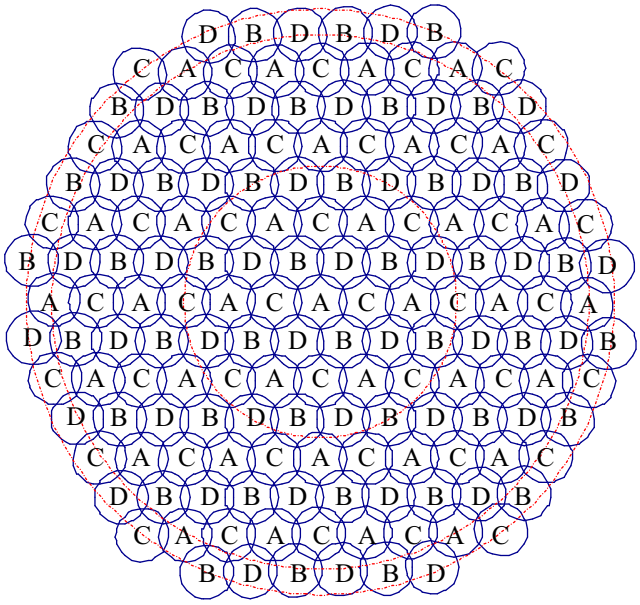


Figure 3.27: ICO cell pattern with cluster size 4 frequency reuse pattern.

beam is controlled to ensure a tolerable level of intra-system interference from co-channel cells. However, an even higher spectral efficiency employing cluster size 3 is not possible with the fixed cell pattern, because this would result in too high a mutual interference.

The frequencies used for the ICO system are located in the S-band (2483.5–2500 MHz), where in particular for the uplink in the mobile link (Earth-to-space link, cf. Fig. 2.3) right-hand circular polarised signals in the range 1985–2015 MHz are used. The bandwidth allocated for the mobile link suffices to provide 4500 voice channels per satellite.

The signal format used on the air interface in the mobile link is a 6-slot TDMA with 40 ms frame duration, where the TDMA rate is 36 kbps. With QPSK modulation with raised-cosine pulse-

shaping with roll-off factor $\beta_{\mathrm{ro}} = 0.4$ the bandwidth requirement amounts to

$$B = \frac{1 + \beta_{\mathrm{ro}}}{T_{\mathrm{sym}}} = 25.2 \text{ kHz}, \tag{3.93}$$

with symbol duration $T_{\mathrm{sym}} = 1/18$ ms. The bit error rate (BER) requirement for voice services is 4%, leading to a $E_{\mathrm{b}}/N_0 = 1.9$ dB.

For data services rate $1/2$ convolutional coding provides a bit rate of 2.4 kbps per TDMA slot at a BER of 10^{-5}, where higher bit rates are realised by using multiple TDMA slot.

Various types (5 in total) of user terminals are proposed for the ICO system with different values of antenna gain, figure-of-merit G/T, and EIRP.

Tab. 3.1 describes three terminal types, providing representative values for the considered parameters. Generally, a limitation for maximum antenna gain for mobile terminals results from the requirement that no active pointing is needed because of an eventually narrow beamwidth of the mobile terminal antenna, this holds in particular for hand-held terminals. Further, the maximum EIRP may be limited to save battery power or due to health concerns, again in particular for hand-held terminals.

Terminal Type	Antenna Gain (dBi)	Peak transmit power P_{T} (W/dBW)	Peak EIRP (dBW)
Hand-held	2	3/4.8	6.8
Private vehicle	3.5	7.1/8.5	12
Semi-fixed	10.5	1.4/1.5	12

Table 3.1: Parameters of user terminals for ICO system.

Of particular importance are the parameters of the DRA antenna for the mobile uplink, as we will focus in Chap. 5 on receiver structures using adaptive array processing, exploiting the spatial distribution of the user terminal signals. In the following section we will therefore provide the array model for an ICO satellite.

ICO Satellite DRA Antenna Model

An ICO satellite employs two DRA (one as transmit antenna for the forward link, the other as receive antenna in the return link) of similar design as introduced in Sec. 3.4.2 (cf. in particular Fig. 3.13) for the mobile uplink and downlink [GST99] (cf. Fig. 3.26). Both DRA of an ICO satellite comprise $L = 127$ array elements each. In references [IHLA98, MS98] a detailed description of the antenna system and its performance is provided, and this will be used in the following to demonstrate the validity of the array model we will use later.

The patterns of the 163 spot beams of an ICO satellite are static, i.e. the spot beam patterns are not adapted to the time-varying signal environment. However, the beam patterns are optimised to minimise the distance between co-channel cells (*reuse distance*) and, consequently, increase spectral efficiency as discussed in Sec. 3.3.3. This is achieved by suppressing intra-system interference more efficiently by minimisation of the sidelobe level for the respective spatial directions where interference is expected to come from, i.e. the co-channel cells. In this way a 4-cell frequency reuse

(cluster size 4) is realised by controlling for each beam the sidelobe level in the corresponding co-channel cells.

For the uplink antenna, the diameter of a single array element is assumed to be $d_\mathrm{e} = 0.195\ \mathrm{m}$, therefore it holds $d_\mathrm{e}/\lambda = 1.29 \ldots 1.31$ for the according uplink frequencies. Accordingly, the gain of a single array element for $f_\mathrm{c} = 2\ \mathrm{GHz}$ ($d_\mathrm{e}/\lambda = 1.3$) at edge-of-coverage for $\vartheta = 22°$ is with (3.27) $2.68\ \mathrm{dB}$ below the peak element gain at boresight for $\vartheta = 0°$.

The array element phase centre separation is given by $d_\mathrm{s} = d_\mathrm{e}$ (cf. Fig. 3.13).

The diameter of the receive array is $d_\mathrm{a} = 2.54\ \mathrm{m}$, and the narrowband assumption for array signal processing (cf. (3.62)) is fulfilled ($f_\mathrm{c} = 2\ \mathrm{GHz}$, $\lambda = 0.15\ \mathrm{m}$):

$$\frac{B_\mathrm{s}}{f_\mathrm{c}} \frac{d_\mathrm{a}}{\lambda} \sin\vartheta_\mathrm{max} = \frac{25.2\ \mathrm{kHz}}{2\ \mathrm{GHz}} \frac{2.54\ \mathrm{m}}{0.15\ \mathrm{m}} \sin 22° \approx 8 \cdot 10^{-5} \ll 1. \tag{3.94}$$

Because $d_\mathrm{s}/\lambda \approx 1.3$, the phase centre separation is larger than $\lambda/2$, and therefore grating lobes appear in the beam pattern.

However, grating lobes are reduced by the gain characteristics of the array elements, and, further, grating lobes do not enter the coverage area even if the beam is steered towards edge-of-coverage with $\vartheta = 22°$: the three grating lobes located in real space are in sine space at (using (3.45)) $(u, v) = (-0.47, 0.64), (-0.47, 0.27), (0.32, 0.72)$ and, equivalently, in angular directions $(\vartheta, \varphi) = (52.5°, 125.9°), (32.5°, -150°), (52.5°, -65.9°)$. Fig. 3.28 shows the the array factor $f_\mathrm{a}(\vartheta, \varphi)$ according to (3.39) for a beam steered towards edge-of-coverage ($\vartheta = 22°$, $\varphi = 30°$), also validating the calculated locations of grating lobes.

Validation of the Array Model In [IHLA98] the measured radiation pattern of the nadir beam[9] is provided. With the element gain pattern $G_\mathrm{e}(\vartheta)$ according to (3.27) and with $d_\mathrm{s} = 0.195\ \mathrm{m}$, the array field pattern can be computed using equation (3.38) and compared with the measured beam pattern.

For a complete comparison we have to address in the following also a method to reduce the sidelobe level of the spot beams in the respective co-channel cells. Reduction of the sidelobe level for antenna arrays with fixed beamforming can be achieved by applying algorithms from adaptive array signal processing. The algorithm used here was proposed in [OC90] and we set aside a detailed discussion.

To compute the beamforming weight vectors producing low sidelobe level for co-channel cells, a high number of interfering sources is placed in the corresponding co-channel cells. Fig. 3.29(a) shows the 163 cell pattern of an ICO satellite in the projection on earth's surface (the satellite is assumed to be located at longitude $0°$, latitude $0°$), further, Fig. 3.29(b) shows the cell pattern in antenna u-v-coordinates with the interfering sources placed in the co-channel cells of the centre cell.

Firstly, for comparison the weight vector for the nadir beam is computed without reducing the sidelobe level in co-channel cells. This is shown in Fig. 3.30(a), whereas the array field pattern is depicted for constant $\varphi = 30°$ (in Figs. 3.30(a) and 3.30(b) this cut is indicated by bold lines). Comparison with the measured beam pattern shows that additional effort is required to reduce the sidelobe level to that of the measured pattern.

[9] The nadir beam is that spot beam which main lobe is pointing towards the satellites nadir point.

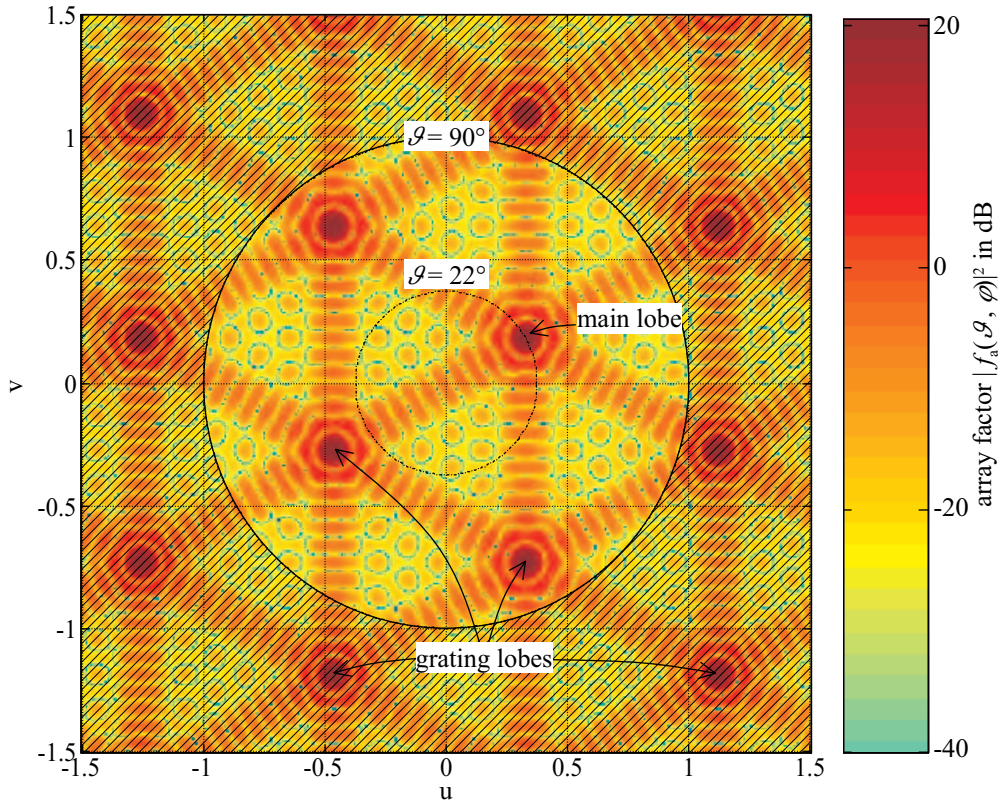
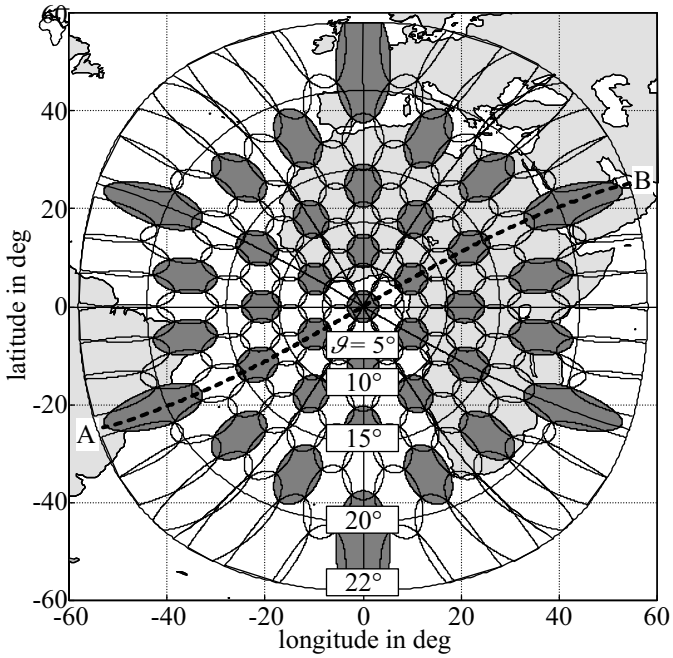
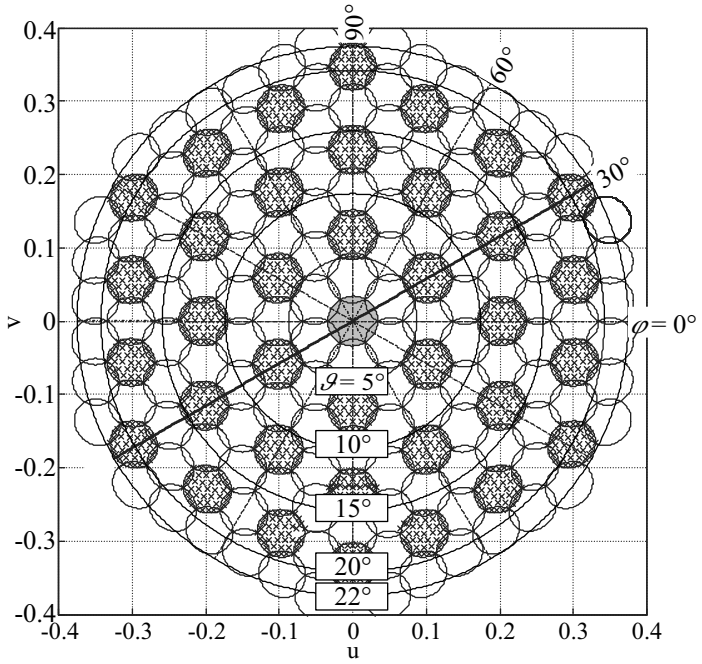


Figure 3.28: Array factor $|f_{\mathrm{a}}(\vartheta, \varphi)|^2$ (logarithmic) for ICO array with the main lobe steered towards edge-of-coverage ($\vartheta = 22°$, $\varphi = 30°$). Only three grating lobes are in real space (imaginary space is indicated as hatched area for $u^2 + v^2 > 1$).

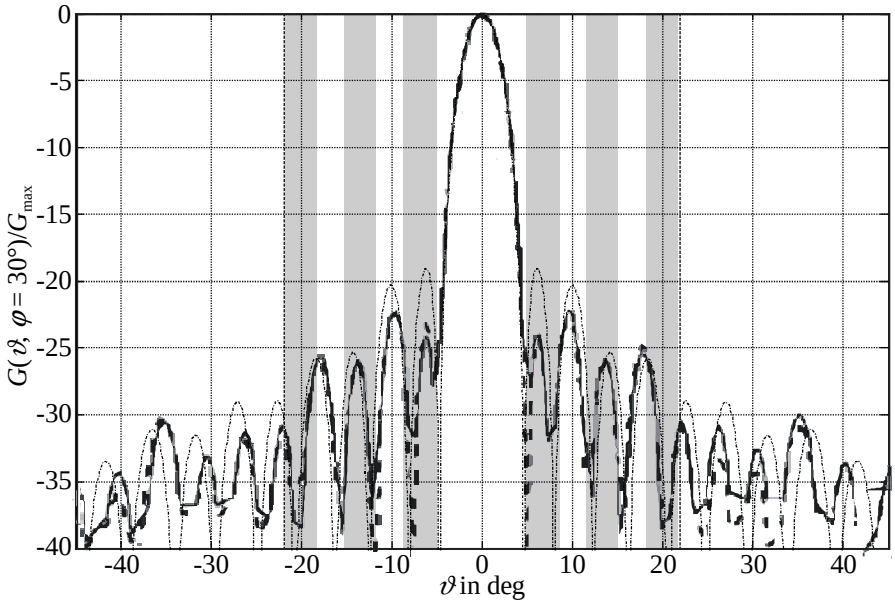


(a) Projection of ICO cell pattern on earth's surface (satellite is located at latitude 0°, longitude 0°). Filled cells (dark grey) indicate the nadir cell and the corresponding co-channel cells.

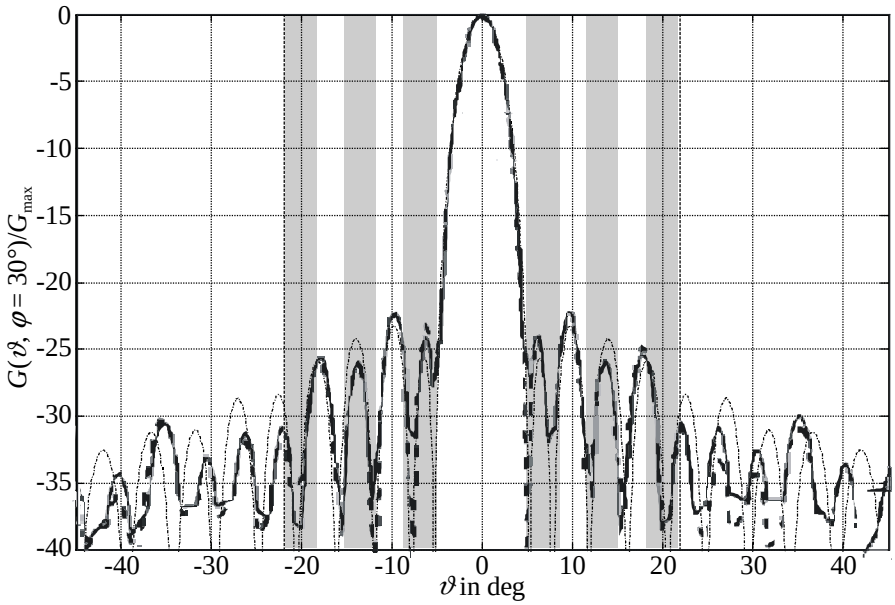


(b) ICO cell pattern in antenna u-v-coordinates. Locations of interferers in 36 co-channel cells (out of 37; cluster size 4) used in optimisation procedure to reduce sidelobe level of the nadir spot beam in the co-channel spot beams. In total, 1540 interferers are considered (43 interferers in each cell, except for the two outermost cells where 8 interferers would be located outside the coverage area).

Figure 3.29: ICO spot beam pattern projected on earth's surface and in antenna u-v-coordinates.



(a) Simulated array field pattern without reduction of sidelobe level in co-channel cells (bold: measurement).



(b) Simulated array field pattern with reduction of sidelobe level in co-channel cells. The sidelobe level of the simulated array pattern agrees very well with the measurements (bold: measurement).

Figure 3.30: Comparison of simulated nadir spot beam pattern based in the array model used in this work and actual measured pattern of the ICO antenna from [IHLA98]. The ranges of ϑ corresponding to co-channel cells are highlighted grey.

The beam pattern obtained from the pattern optimisation procedure, with the sidelobe level being reduced in the co-channel cells, is shown in Fig. 3.30(b). Reducing the sidelobe level in the co-channel cells requires a tapered, non-uniform aperture illumination, which results in an aperture efficiency reduced to $\eta_{\mathrm{A,a}} \approx 0.95$.

It can be concluded that the obtained sidelobe level agrees well with the measurements in the

important range of off-boresight angles $0 \leq \vartheta \leq 22°$.

Directivity and Gain From the definition of the field pattern we know that it provides only *relative* radiation intensities, and in Fig. 3.30 we have indeed only compared the beam patterns normalised to the maximum value.

Finally, a radiation efficiency $\eta_{\mathrm{r,a}} = 0.63$ must be assumed to obtain an array gain $G_{\mathrm{a}}(\vartheta, \varphi) = \eta_{\mathrm{r,a}} D_{\mathrm{a}}(\vartheta, \varphi)$ that agrees sufficiently well with the published ICO satellite antenna measurements (30.4 dB at nadir and 28.1 dB peak gain for a beam at edge-of-coverage) [MS98, IHLA98].

The array model introduced in this section reproduces the field pattern and the absolute gains achieved by the real ICO DRA antenna with sufficient accuracy.

Concluding this section, Tab. 3.2 summarises the satellite DRA antenna parameters for the ICO scenario.

Parameter	Value
Carrier frequency f_{c}	2 GHz
Wave length λ	0.15 m
Max. off-boresight angle $\vartheta_{\max}$	$22°$
Effective Noise Temperature T_{e}	290 K
Bandwidth B_{s}	25.2 kHz
Element	
Diameter d_{e}/λ	1.3
Gain G_{e}	10 dB ($\vartheta = 0$),
	7.5 dB ($\vartheta = \vartheta_{\max}$)
Array	
Number of elements	127
Phase centre separation d_{s}	1.3
Diameter d_{a}	2.54 m
Aperture efficiency $\eta_{\mathrm{A,a}}$	0.95
Radiation efficiency $\eta_{\mathrm{r,a}}$	0.63
Gain G_{a}	30.8 dB (beam 1, peak gain)
	28.3 dB (beam 163, peak gain)
One-sided 3 dB-beamwidth $\vartheta_{3\mathrm{dB}}$	$2°$

Table 3.2: Important system and DRA antenna parameters assumed for the ICO satellite scenario.

3.6.2 GEO Satellite Scenario: the EuroSkyWay System

In particular for LEO and MEO satellites the direct radiating array is considered an attractive option over reflector antennas [Bjö93], providing possibility of large off-boresight scan angles [Bjö93]. Due to the comparably low free-space basic transmission loss, even in the S-band the required antenna gain can be provided with moderate, technologically manageable (with respect to size, weight, and number of array elements) aperture area and number of array elements as it was discussed in more detail in the preceding section. Therefore, LEO and MEO satellite systems usually employ DRA antennas [MS98, SUM+99].

In contrast, for GEO satellites the larger free-space basic transmission loss has to be compensated, and, further, to obtain the possibility of frequency reuse, the cell area (accordingly the spot beam width) must be significantly smaller than the coverage area (cf. Fig. 3.8). Depending on the frequency band used to provide the satellite services, both large reflector antennas (e.g. the already mentioned Thuraya satellites with $12\ \mathrm{m}$-diameter reflector antennas for operation in L-band), as well as DRA are possible choices.

For GEO satellites, the DRA antenna is a technologically feasible option only for frequencies at K/Ka-band and higher, where for regional coverage (e.g. Europe) an array aperture of approx. $1\ \mathrm{m}$ diameter with about 200 array elements can be assumed a realistic order of magnitude [CCAL01].

Adaptive beamforming is also possible with reflector antennas [LL96, KB96, And99], but we will restrict in this work to DRA, in particular also for the GEO satellite scenario.

The system parameters we will assume are taken from the EuroSkyWay system description [ESW97]. Again, we will concentrate on the uplink in the mobile link.

EuroSkyWay System Outline

Development of the EuroSkyWay satellite system was initiated in 1994 and is still on-going. Therefore many parameters of EuroSkyWay are subject to change, and we will restrict to present shortly the most important system parameters only (the description of EuroSkyWay presented below mirrors the design status as published 1999 and 2001, respectively [ESW97, LMM01]).

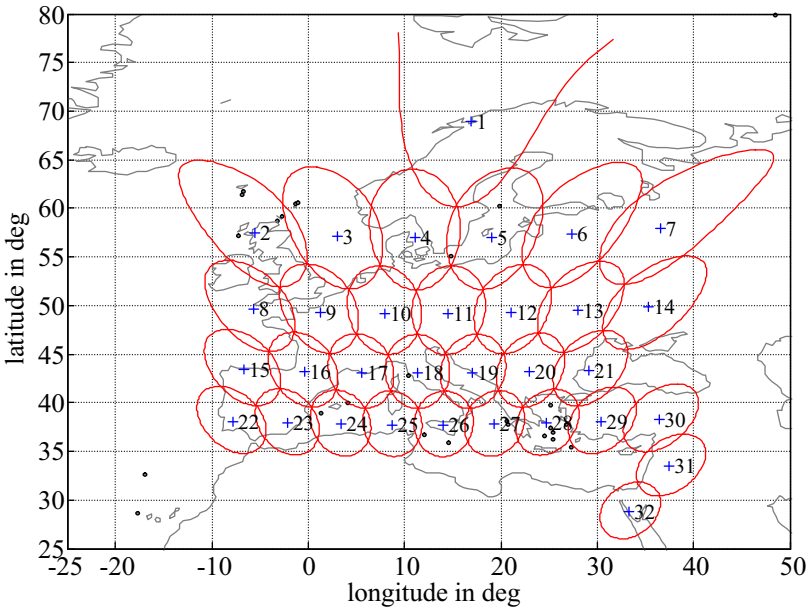
EuroSkyWay is a satellite communication system that aims at provision of high-speed, broadband connectivity services to service providers.

It is planned that a GEO satellite in an orbital position at latitude $0°$, longitude $12°\mathrm{E}$ provides coverage of Europe and part of the Middle East, Mediterranean Africa, and part of the CIS countries. The service area is covered by 32 cells, where again a 4-frequency reuse pattern is realised (cf. Fig. 3.31(b)). The cell boundaries are defined as the $4.3\ \mathrm{dB}$-below-maximum spot beam gain contour.

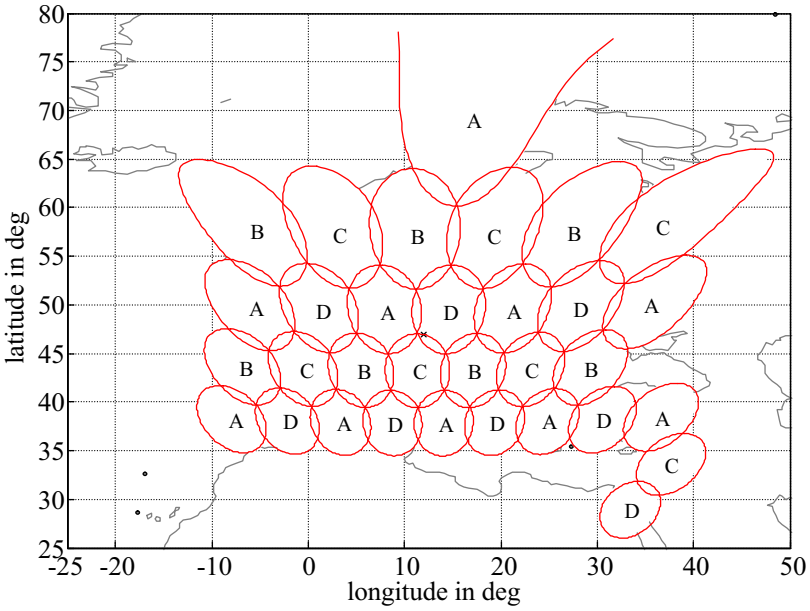
The frequencies allocated for EuroSkyWay are in the Ka-band ($20\ \mathrm{GHz}/30\ \mathrm{GHz}$).

High usable data rates between approx. $193.2\ \mathrm{kbps}$ and $2473\ \mathrm{kbps}$ are supported by different user terminals, and with code rate of approx. 0.8 this results in coded data rates between $242.3\ \mathrm{kbps}$ and $3091.3\ \mathrm{kbps}$.

Parameters of two types of user terminals are shown in Tab. 3.3. Further QPSK modulation is employed, and a bandwidth requirement between approx. $190\ \mathrm{kHz}$ and $2.5\ \mathrm{MHz}$ is stated in [ESW97].



(a) Cell pattern (cell boundary defined by 4.3 dB-below-maximum spot beam contour). The "+" indicate the cell centers (maximum spot beam gain).



(b) 4-frequency reuse scheme in the uplink.

Figure 3.31: EuroSkyWay service area, 32 cells cover Western Europe and the Middle East.

Terminal Type	Antenna Gain (dBi), diameter	Peak EIRP (dBW)
SaT-A	33.5, 0.19 m	37.8
SaT-C	36.6, 0.28 m	49.9

Table 3.3: Parameters of user terminals for EuroSkyWay system. The terminal antenna efficiency is 0.6.

GEO Satellite DRA Antenna Model

The original system design of EuroSkyWay does not foresee a DRA antenna, therefore we cannot provide a similarly detailed comparison between DRA model and measurements as for the ICO

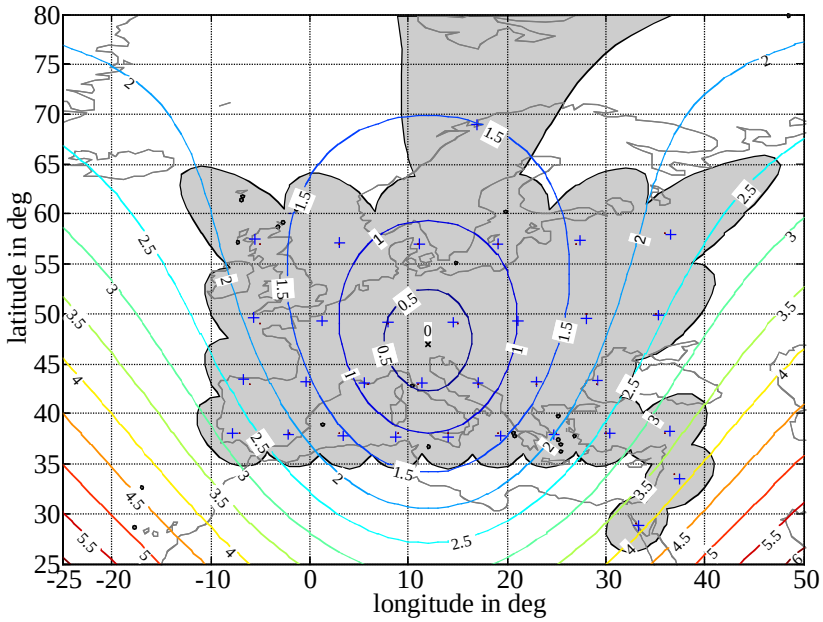


Figure 3.32: Off-boresight angles for antenna boresight pointed towards longitude 12°E, latitude 47°N. The maximum off-boresight angle amounts to approx. $\vartheta_{\max} = 4.3°$. Area shaded grey indicates the EuroSkyWay service area.

satellite antenna in Sec. 3.6.1. Instead, we will use parameters from [CCAL01] for comparison, where requirements for a DRA antenna for a similar GEO satellite scenario are defined.

To optimise the DRA antenna gain characteristics inside the service area, the boresight of the satellite antenna is pointed towards longitude 12°E, latitude 47°N, such that the maximum off-boresight angle amounts to approx. $\vartheta_{\max} = 4.3°$ (cf. Fig. 3.32).

The DRA consists of $L = 169$ identical circular elements, again arranged on a triangular lattice. Using (3.27) the array element diameter is chosen to maximise the element gain at the edge-of-coverage $G_{\mathrm{e}}(\vartheta_{\max})$, which results in $d_{\mathrm{e}}/\lambda = 8$ for $f_{\mathrm{c}} = 30$ GHz (i.e. $d_{\mathrm{e}} = 0.08$ m). The element gain for $\vartheta = \vartheta_{\max}$ at edge-of-coverage is then approx. 4.3 dB below the element gain at boresight for $\vartheta = 0$ [CCAL01].

The phase centre separation is $d_{\mathrm{s}} = d_{\mathrm{e}}$, and consequently the array diameter amounts to $d_{\mathrm{a}} = 1.21$ m (cf. Fig. 3.13). The narrowband assumption for array signal processing (cf. (3.62)) is fulfilled ($f_{\mathrm{c}} = 30$ GHz, $\lambda = 0.01$ m):

$$\frac{B_{\mathrm{s}}}{f_{\mathrm{c}}} \frac{d_{\mathrm{a}}}{\lambda} \sin\vartheta_{\max} = \frac{2.5 \text{ MHz}}{30 \text{ GHz}} \frac{1.2 \text{ m}}{0.01 \text{ m}} \sin 4.3° \approx 7.8 \cdot 10^{-4} \ll 1. \tag{3.95}$$

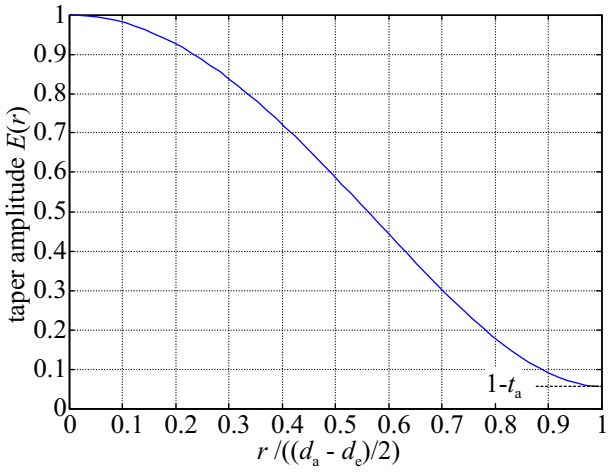
To obtain sufficiently low sidelobe level an amplitude taper is introduced according to

$$E(r) = 1 - t_{\mathrm{a}} + t_{\mathrm{a}} \left(1 - \frac{4r^2}{(d_{\mathrm{a}} - d_{\mathrm{e}})^2}\right)^p, \qquad 0 \le r \le (d_{\mathrm{a}} - d_{\mathrm{e}})/2, \tag{3.96}$$

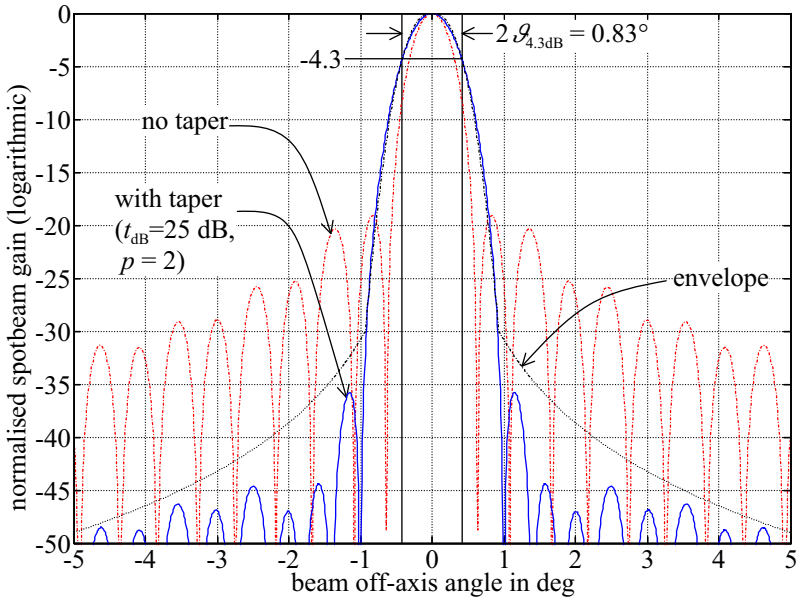
where $r = |\mathbf{p}_{\mathrm{e},l}|$ ($l = 1, 2, \ldots, L$, cf. (3.31)) is the distance of the array element phase centre from the array centre, further, t_{a} is the taper amplitude at the edge of the aperture, and $p = 0, 1, 2, \ldots$ determines the rate at which $E(r)$ drops from $E(r = 0) = 1$ to $E(r = (d_{\mathrm{a}} - d_{\mathrm{e}})/2) = 1 - t_{\mathrm{a}}$

[CCF+92]. For higher values of $t_{\mathrm{dB}} = -20\log_{10}(1 - t_{\mathrm{a}})$, the sidelobe level decreases as desired, whereas, as an undesired consequence, the beamwidth increases and, accordingly, the peak gain decreases.

The requirements for the spot beam patterns for EuroSkyWay (found in [ESW97]) are fulfilled using a taper with $t_{\mathrm{dB}} = 25$ dB and $p = 2$ (cf. Fig. 3.33), which results in an aperture efficiency of approx. 0.6.



(a) Aperture taper function $E(r)$ with $t_{\mathrm{dB}} = 25$ dB, $p = 2$. The resulting aperture efficiency is $\eta_{\mathrm{A,a}} = 0.66$



(b) With taper the sidelobe level requirement (envelope) is fulfilled.

Figure 3.33: Comparison of the beam pattern obtained with and without aperture taper.

Fig. 3.34 shows the cell pattern obtained with the DRA model described above with the cell pattern as proposed for the EuroSkyWay system.

Again, grating lobes are present as the array element phase centre separation is larger than $\lambda/2$. From Fig. 3.35 it can be seen that no grating lobe is located in the service area and, further, that the level of grating lobes is significantly reduced due to the element gain characteristics $G_{\mathrm{e}}(\vartheta)$.

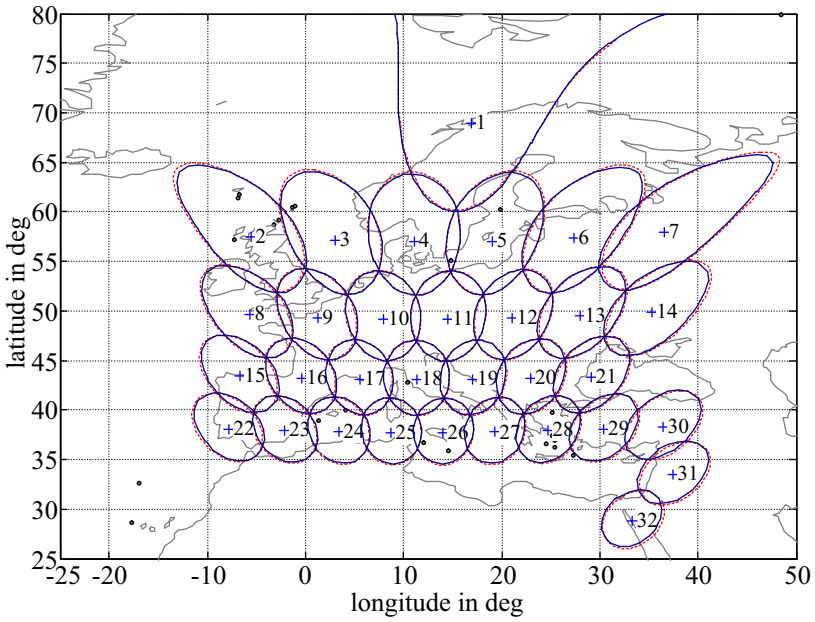
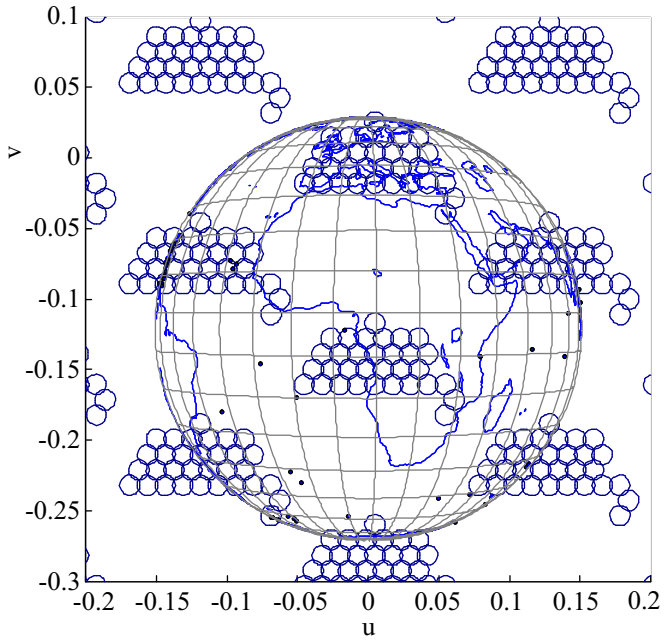


Figure 3.34: Comparison of the cell pattern as proposed in EuroSkyWay system design (dashed) with cell pattern obtained with the described array model (solid). Small deviations from the idealised cell pattern occur at the edge-of-coverage.

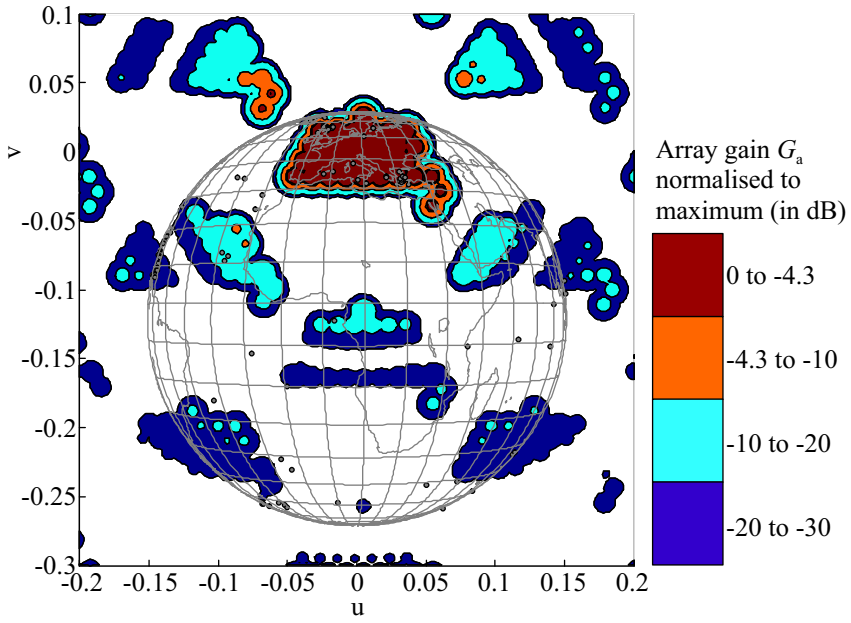
Concluding, Tab. 3.4 summarises the satellite DRA antenna parameters for the EuroSkyWay scenario.

Parameter	Value
Carrier frequency f_c	30 GHz
Wave length λ	0.01 m
Max. off-boresight angle ϑ_max	4.3°
Effective Noise Temperature T_e	444 K
Bandwidth B_s	190 kHz–2.5 MHz
Element	
Diameter d_e/λ	8
Gain G_e	26.9 dB ($\vartheta = 0$),
	22.6 dB ($\vartheta = \vartheta_\mathrm{max}$)
Array	
Number of elements	169
Phase centre separation d_s/λ	8
Diameter d_a	1.2 m
Aperture taper	$t_\mathrm{dB} = 25$ dB, $p = 2$
Aperture efficiency $\eta_\mathrm{A,a}$	0.66
Radiation efficiency $\eta_\mathrm{r,a}$	0.75
Spot beam gain G_a	47.3 dB (beam 18, peak gain)
	43.5 dB (beam 32, peak gain)
One-sided 4.3 dB-beamwidth $\vartheta_\mathrm{4.3dB}$	0.415°

Table 3.4: Important system and DRA antenna parameters assumed for the EuroSkyWay satellite scenario.



(a) 4.3 dB-contours of the array factor $f_{\mathrm{a}}(\vartheta, \varphi)$ (i.e. element gain G_{e} is not taken into account). Note that a significant number of grating lobes is present on earth's surface.



(b) Filled contour plot of the array gain G_{a} including element gain G_{e}, normalised to maximum. The grating lobes level is significantly reduced due to element gain characteristics, the strongest sidelobes are pointing towards space.

Figure 3.35: Positions of grating lobes in u-v-coordinates.

Chapter 4

Resource Allocation for the Fading Vector Multiple-Access Channel

In the last chapter it was mentioned that it is necessary to cover the service area with a number of spot beams if a high satellite antenna gain is required for the link budget. The usage of spot beams introduces also the advantage of frequency reuse in different cells, depending on the multiple access scheme used. This improves the spectral efficiency of the system, where the minimal cluster size is determined by the tolerable level of co-channel interference, which arises between sources in cells using the same frequency. Satellite antennas are therefore designed to have low sidelobes to reduce interference from co-channel cells.

However, the sidelobe level cannot be reduced arbitrarily because of the related decrease of aperture efficiency and an undesired broadening of the main lobe. Finally the beamwidth of the main lobe is the limiting factor regarding the minimal distance of co-channel sources (cf. Fig. 3.30).

Of course, interference can be counteracted to a certain extent by simply increasing transmit powers. Yet increasing the transmit power of one source to compensate interference from other sources means in turn a higher interference level for the other co-channel sources, and this mutual interference may build up to excessively high transmit powers.

To achieve lower transmit powers and better spectral efficiency more elaborate techniques are required to suppress interference more efficiently. Generally there are two approaches to solve this task.

Firstly, interference can be suppressed better at the beamforming stage of the receiver by employing adaptive beamforming instead of fixed beamforming. This is discussed for the satellite application in a number of publications, e.g. [CLW92, Bjö93, GG95, LL96, Yu96, Lüc98, LC00, Lüc00, Gay02], just to name a few. Adaptive beamforming does not only reduce interference, but also compensates partially or even completely the edge-of-cell loss that comes with fixed beams.

Secondly, interference can be reduced by joint decoding, where we will in particular deal with successive decoding. Here, the source signals are decoded sequentially, where the contribution of each source signal to the received signal can be removed after it is decoded, such that interference to the signals decoded subsequently vanishes [CT91]. Also this approach is discussed for satellite scenarios in various articles, e.g. [Ern99]. Finally, the optimal receiver employs both (optimal) adaptive beamforming and successive decoding.

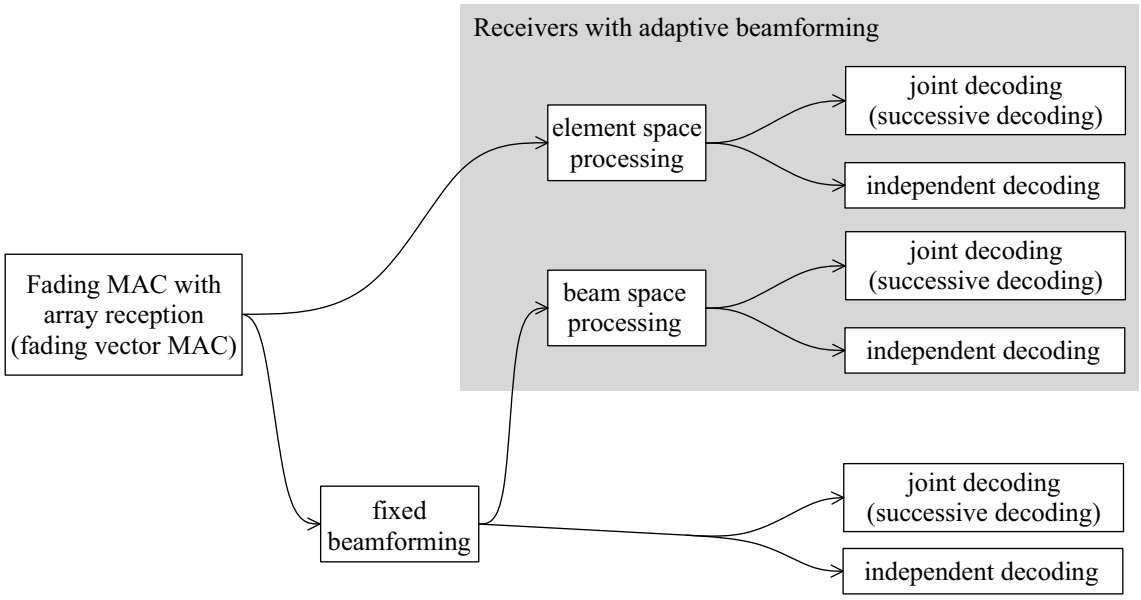


Figure 4.1: Overview of the receiver/decoder options.

What is missing in literature so far is a detailed comparison of the two said approaches, i.e. an analysis of the advantages and disadvantages with respect to performance and complexity under particular consideration of the conditions encountered in satellite communications scenarios.

Fig. 4.1 depicts the options for the receiver structures that shall be investigated, which can be divided in receivers employing fixed or adaptive beamforming (element or beam space), and independent or joint (successive) decoding.

Fixed beamforming with independent decoding represents the receiver scheme being implemented in present-day satellite systems.

The various receiver options depicted in Fig. 4.1 can all be traced back to the *fading vector MAC* with independent or joint decoding. Therefore the fading vector MAC will be discussed in detail in this chapter to lay foundation to a comparison of the receiver options in a unified way.

We will then present the actual comparison in Chap. 5, based on the results presented in this chapter.

As a central point, we will present in the following the expressions for the region of achievable rates (capacity region) and the power region, and discuss their properties. In particular we will investigate the fading vector MAC in the context of *polymatroids*, which have proven to be an important concept for resource allocation problems[1] in MAC scenarios [TH98a, TH98b].

In fact, scanning the literature shows that, so far, the vector MAC is not looked into in connection with this resource allocation problem and polymatroids.

We will in the following introduce terms related to channel capacity and discuss properties of polymatroids. Further, we will continue in this chapter with the well known classical MAC with fading as an introduction to the matter before finally dealing with the vector MAC.

[1] As *resource allocation* we understand here the process of (i) allocating transmit powers to the sources to achieve a certain transmission rate R, (ii) determining the transmission rate R that is achievable when the maximum transmit powers of the sources are limited.

4.1 Polymatroids

Let f be a *set function*, with $f : 2^{\mathcal{M}} \to \mathbb{R}_+$, where $\mathcal{M} = \{1, 2, \ldots, M\}$ ($2^{\mathcal{M}}$ denotes the *power set* of $\mathcal{M}$, which is defined as the set of all subsets of $\mathcal{M}$. The cardinality is $|2^{\mathcal{M}}| = 2^{|\mathcal{M}|}$, including the empty set $\emptyset$[2]; the following expressions are equivalent: $\mathcal{S} \subseteq \mathcal{M}$, $\mathcal{S} \in 2^{\mathcal{M}}$).

Set function f is called a *rank function* if it has the following properties

$$f(\emptyset) = 0 \qquad \text{(normalised)}, \tag{4.1}$$
$$f(\mathcal{S}) \leq f(\mathcal{T}), \mathcal{S} \subseteq \mathcal{T} \qquad \text{(nondecreasing)}, \tag{4.2}$$
$$f(\mathcal{S}) + f(\mathcal{T}) \geq f(\mathcal{S} \cup \mathcal{T}) + f(\mathcal{S} \cap \mathcal{T}) \, \forall \mathcal{S}, \mathcal{T} \subseteq \mathcal{M} \qquad \text{(submodular)}. \tag{4.3}$$

Then the *polytope*[3] in M-dimensional space [Wel76]

$$\mathcal{B}(f) = \left\{ (x_1, x_2, \ldots, x_M) : \sum_{m \in \mathcal{S}} x_m \leq f(\mathcal{S}) \qquad \forall \mathcal{S} \subseteq \mathcal{M}, x_m \geq 0 \, \forall m \right\} \tag{4.4}$$

is a *polymatroid*.

Note that $\mathcal{B}(f)$ is defined over $2^M - 1$ inequalities, not counting the inequalities $x_m \geq 0$, and the trivial case $\mathcal{S} = \emptyset$. Further, the polytope $\mathcal{B}(f)$ has $M!$ vertices (a vertex is a point on the boundary of $\mathcal{B}(f)$ where M out of the $2^M - 1$ defining inequalities become strict equalities).

Similarly a polytope $\mathcal{G}(f)$ with

$$\mathcal{G}(f) = \left\{ (x_1, x_2, \ldots, x_M) : \sum_{m \in \mathcal{S}} x_m \geq f(\mathcal{S}) \qquad \forall \mathcal{S} \subseteq \mathcal{M} \right\} \tag{4.5}$$

is a *contra-polymatroid* if the rank function f is normalised, nondecreasing, and further satisfies

$$f(\mathcal{S}) + f(\mathcal{T}) \leq f(\mathcal{S} \cup \mathcal{T}) + f(\mathcal{S} \cap \mathcal{T}) \qquad \text{(supermodular)}. \tag{4.6}$$

Finally, a rank function f of a polymatroid $\mathcal{B}(f)$ is said to be *generalised symmetric* if there exists a vector $\mathbf{y} \in \mathbb{R}_+^M$ and a nondecreasing concave function h such that

$$f(\mathcal{S}) = h\left(\sum_{m \in \mathcal{S}} y_m \right), \mathcal{S} \subseteq \mathcal{M}. \tag{4.7}$$

Then, for all vectors $\mathbf{x} \in \mathbb{R}_+^M$, the set $\{\mathbf{y} : \mathbf{x} \in \mathcal{B}(h(\sum_{m \in \mathcal{S}} y_m))\}$ is a contra-polymatroid.

Key properties of polymatroids are shortly reviewed below [TH98a, Edm69].

[2] E.g., if $\mathcal{M} = \{1, 2, 3\}$, we obtain $2^{\mathcal{M}} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$, the cardinality is $2^{|\mathcal{M}|} = 8$.

[3] A *polytope* is defined as the bounded and nonempty intersection of a finite set of halfspaces (a hyperplane defines two halfspaces). From the convexity of halfspaces it follows directly that a polytope is a convex set. An M-dimensional polytope is defined as the set of solutions to the system of n linear inequalities $\mathbf{Mx} \leq \mathbf{b}$, $\mathbf{M} \in \mathbb{R}^{n \times M}, \mathbf{x} \in \mathbb{R}^M, \mathbf{b} \in \mathbb{R}^n$. Each row of $\mathbf{M}$ is a normal vector of a hyperplane, where the corresponding element of $\mathbf{b}$ is the distance between hyperplane and origin. (A polyhedron is a polytope in 3-dimensional space [PS98].)

4.1.1 Characterisation of the Vertices of a (Contra-)Polymatroid

Let

$$\boldsymbol{\pi}_v = (\pi_v(1), \pi_v(2), \ldots, \pi_v(M)), \qquad v = 1, 2, \ldots, M!, \tag{4.8}$$

denote the $M!$ permutations on the set $\mathcal{M}$.

If $\mathcal{B}(f)$ is a polymatroid (the following holds equally for contra-polymatroid $\mathcal{G}(f)$) then the $M!$ vertices $\boldsymbol{\omega}(\boldsymbol{\pi}_v) = (w_1, w_2, \ldots, w_M)$ of $\mathcal{B}(f)$ ($\mathcal{G}(f)$) are provided by the following equations, where we use v to index the vertices:

$$\begin{aligned} \omega_{\pi_v(1)} &= f(\{\pi_v(1)\}) \\ \omega_{\pi_v(m)} &= f(\{\pi_v(1), \pi_v(2), \ldots, \pi_v(m)\}) - f(\{\pi_v(1), \pi_v(2), \ldots, \pi_v(m-1)\}), \\ & \quad m = 2, 3, \ldots, M, \qquad v = 1, 2, \ldots, M!, \end{aligned} \tag{4.9}$$

(obviously, M inequalities out of $2^M - 1$ become tight at a vertex [RU96]).

Conversely, if f is a set function and $\boldsymbol{\omega}(\boldsymbol{\pi}_v) \in \mathcal{B}(f)$ holds for all permutations $\boldsymbol{\pi}_v$ then $\mathcal{B}(f)$ is a polymatroid. The same holds for contra-polymatroids.

4.1.2 Polymatroids and Linear Programming

A *linear program* is defined as (this is the so-called standard form)

$$\min_{\mathbf{x}} \{\boldsymbol{\lambda} \cdot \mathbf{x} : \mathbf{C}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}, \tag{4.10}$$

where matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$, vectors $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ are given, and $\mathbf{x} \in \mathbb{R}^n$ is the vector of variables to be solved for. By minimising $-\boldsymbol{\lambda} \cdot \mathbf{x}$ we can maximise $\boldsymbol{\lambda} \cdot \mathbf{x}$.

Further, $\boldsymbol{\lambda} \cdot \mathbf{x}$ is called the *objective function* and $\mathbf{C}\mathbf{x} = \mathbf{b}$ are the *constraints*. The constraints $\mathbf{C}\mathbf{x} = \mathbf{b}$ define the *feasible region* (or *set*) for $\mathbf{x}$.

For the linear program, the feasible region is a polytope and it is well known that, if a solution to (4.10) exists, it is obtained at a vertex of this polytope [PS98].

That the optimal solution must be found at a vertex of the feasible set is evident, because the objective function $\boldsymbol{\lambda} \cdot \mathbf{x}$ defines a set of hyperplanes $\boldsymbol{\lambda} \cdot \mathbf{x} = d$, having $\boldsymbol{\lambda}$ as normal vector and with d being the distance of the hyperplanes to the origin of the coordinate system; since polytopes are convex there must be a unique d^* for that the according hyperplane $\boldsymbol{\lambda} \cdot \mathbf{x} = d^*$ touches the feasible set in a single vertex or on a face of the polytope. This is illustrated in Fig. 4.2.

There is a number of optimisation algorithms to find the optimal vertex in an iterative way, like the simplex method, the ellipsoid method, or interior point methods [PS98].

Solving (4.10) is simplified enormously if the feasible set is a (contra-)polymatroid, in fact the solution to the linear programs

$$\min_{\mathbf{x}} \{\boldsymbol{\lambda} \cdot \mathbf{x} : \mathbf{x} \in \mathcal{G}(f)\} \tag{4.11}$$

$$\max_{\mathbf{x}} \{\boldsymbol{\lambda} \cdot \mathbf{x} : \mathbf{x} \in \mathcal{B}(f)\}, \tag{4.12}$$

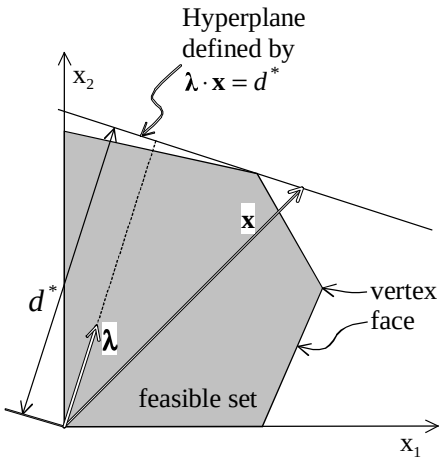


Figure 4.2: The optimal solution to (4.10) is always found at a vertex of the polytope that represents the feasible set ($\mathbf{x} = (x_1, x_2)$). It is possible that the hyperplane $\boldsymbol{\lambda} \cdot \mathbf{x}$ runs through a face of the polytope, then all points of the face including the according two vertices similarly fulfill the optimality criterion, i.e. maximise $\boldsymbol{\lambda} \cdot \mathbf{x}$.

for a given vector $\boldsymbol{\lambda} \in \mathbb{R}_+^M$, is then in both cases attained at a vertex $\mathbf{x} = \mathbf{w}(\boldsymbol{\pi}_{v^*})$ as defined by (4.9), where $\boldsymbol{\pi}_{v^*}$ is a permutation such that

$$\lambda_{\pi_{v^*}(1)} \geq \lambda_{\pi_{v^*}(2)} \geq \cdots \geq \lambda_{\pi_{v^*}(M)}. \tag{4.13}$$

This means that the solution to (4.10) can be calculated directly (in time $\mathcal{O}(M \log_2 M)$ [TH98a], which is actually the minimum complexity for sorting a list of M elements), without the need to employ an iterative optimisation algorithm.

Concluding it can be stated that the properties of polymatroids allow an efficient (so called *greedy*[4]) algorithm to solve (4.10).

The relation between polymatroids and the problem of optimal resource allocation related to the MAC will become evident in the following sections.

4.2 Definitions and Assumptions

As already mentioned in the introduction of this chapter we want to compare different receiver options in terms of the maximum rates that are achievable on the underlying fading vector MAC. When addressing capacity in connection with fading channels we have to specify a number of assumptions concerning the particular definition of capacity, the availability of channel state information (CSI), and power control.

[4]Greedy Algorithm: An algorithm that always takes the best *local* solution by moving in the direction of steepest ascent while staying inside the feasible region defined by the constraints. Greedy algorithms find the *globally* optimal solution for some optimisation problems (but not generally), in particular if the underlying problem has a matroidal structure [TH98a].

4.2.1 Information Capacity

The rate with that a source can transmit information over a channel is bounded by the (information) capacity of the underlying channel [CT91]. Depending on the variability of the fading process during the transmitted block, different definitions of capacity may be applied, namely *ergodic capacity* (this is also termed the *Shannon capacity* or *throughput capacity* [TH98a]), *distribution of capacity* (closely related to outage-capacity), and *delay-limited capacity* [EB98].

When the data is transmitted in blocks of a length that is sufficient to capture the fading statistics of the channel, *ergodic capacity* applies, and this will be the approach that will be followed below. The ergodic capacity of a channel defines its theoretical upper bound for the maximum rate of data transmission at a BER tending to zero, and without any delay or complexity constraints [AG99].

The capacity is measured in *bits per transmission* (also: *bits per channel use*, or *bits per use*).

Further, in App. B a review of some of the most important concepts of information theory is presented, which should aid understanding of the investigations presented in the following sections.

4.2.2 Channel State Information and Power Control

The amount of information that can be transmitted over the fading channel depends on the availability of channel state information (CSI). The term CSI refers to knowledge of the channel state that is possibly available at the transmitting and receiving side, which is obtained by some channel estimation technique, e.g. by means of regularly transmitted training symbols known to the receiver, or pilot tones [EB98].

The channel state according to the channel model introduced in Sec. 2.2.1 is characterised by the link attenuation factors $\sqrt{\mu_m}$ and the fading factors h_m, and we will assume that the receiver (the satellite) has ideal knowledge of these factors to coherently demodulate and decode the received source signals.

In contrast, the transmit power of the sources is only adapted to the slowly varying link attenuation factors $\sqrt{\mu_m}$, but not to the rapidly changing fading factors h_m. This assumption is justified by the difficulties to implement a reliable channel estimation at the transmitters side in a satellite scenario, mainly due to the delay on the link source–satellite [Jah99].

The impact of partial or noisy CSI on achievable rates is discussed in [Mec02b, Mec02a].

4.3 The Classical Fading Multiple-Access Channel

The classical fading MAC is described by the following model (flat-fading, discrete-time, synchronous, complex-valued) (cf. Fig. 4.3) [Mec00b]

$$r_k = \sum_{m=1}^{M} s_{m,k} \frac{h_{m,k}}{\sqrt{\mu_{m,k}}} + n_k, \tag{4.14}$$

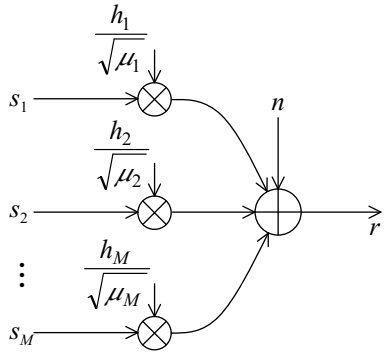


Figure 4.3: The classical MAC with fading. Assumptions are: multiplicative (flat) fading, discrete-time, synchronous, complex-valued.

where k denotes the discrete-time index. This corresponds to the receive signal at a single array element according to (3.63) (considering a single array element only, the phase factors $a_{1,m}$ are irrelevant).

As pointed out in App. B, the source symbols s_m and the noise n are realisations of proper complex Gaussian random variables (cf., e.g., [CT91, NM93]).

Further, the s_m are the source signals with power constraint $\mathrm{E}\{|s_m|^2\} = p_m$, n is a white Gaussian random process with variance $\mathrm{E}\{|n|^2\} = 2\sigma_{\mathrm{n}}^2$ and r is the signal available at the receiver for decoding, further, $\sqrt{\mu_m}$ models the (amplitude) link loss and h_m are the fast-fading factors (cf. Sec. 3.5 and 2.2.1).

The mean received powers are defined as

$$P_m = \mathrm{E}\left\{\left|\frac{s_m h_m}{\sqrt{\mu_m}}\right|^2\right\} = \frac{p_m}{\mu_m}, \tag{4.15}$$

since we have defined $\mathrm{E}\{|h_m|^2\} = 1$ in Sec. 2.2.1 (also cf. (2.20), where P_m and p_m correspond to P_{R} and P_{T}, respectively).

The central property of the MAC is that the M signal sources, communicating with a single receiver via a common channel, cause mutual interference. Of course, the goal of the receiver is to reliably decode all M source signals contained in the receive signal r despite the mutual interference.

For retrieving the different source signals s_m from the receive signal r, there are generally two decoder options [CT91, SXLK98]:

- *Single-user decoding* (also referred to as *independent decoding* or *non-cooperative decoding*), and
- *multi-user decoding* (also referred to as *joint decoding* or *cooperative decoding*).

For joint decoding it is exploited that each source signal, though random in nature, is known at the receiver after decoding. Theoretically joint decoding can achieve rates equal to capacity of the underlying fading MAC and is in this sense considered as optimal decoding.

For independent decoding each signal is decoded at the receiver separately considering the interference simply as unknown noise, which is of course suboptimal.

4.3.1 Independent Decoding

For independent decoding of the m-th user signal (this will be also referred to as the *wanted signal*) the signals of the $(M-1)$ co-channel signals are considered unknown, i.e. treated as interference. For given transmit power constraints, the achievable rates in a multi-user environment are significantly lower than those rates achievable with joint decoding.

On the other hand, the advantage of independent decoding clearly is the simplicity of the approach and that single-user coding is very well understood. Single-user codes working close to capacity are known and can be decoded with low complexity (e.g. parallel concatenated convolutional codes ("Turbo") codes [BGT93]).

For independent decoding, and assuming that CSI is available only at the receiver (according to Sec. 4.2.2), the achievable rate R_m for each source m is bounded by [EB98]

$$R_m \leq \mathrm{E}\left\{\log_2\left(1+\frac{|H_m|^2 P_m}{2\sigma_\mathrm{n}^2+\sum\limits_{i\in\mathcal{I}_m=\mathcal{M}\setminus m}|H_i|^2 P_i}\right)\right\}_{\boldsymbol{H}}, \qquad m\in\mathcal{M}=\{1,2,\ldots,M\}, \quad (4.16)$$

where $\mathcal{I}_m=\mathcal{M}\setminus m$ denotes obviously the set of sources interfering with source m.

Recalling the assumption that the input symbols s_m are chosen from a proper Gaussian random process, the denominator in (4.16) obviously denotes the variance of the Gaussian random process created by summation of the interfering source signals and the noise, where all random processes are uncorrelated (signals add in power).

Power Allocation for Independent Decoding

If all M sources have to achieve equal rate $R_m=R$, then, assuming the same fading statistics for all sources, all receive powers P_m have to be equal as well, i.e. $P_m=P$.

Further, there is a maximal rate $R_\mathrm{max}^\mathrm{CMAC}$ that can be achieved at most due to the mutual interference for $P\to\infty$, which is given by (cf. [Mec00a])

$$R_\mathrm{max}^\mathrm{CMAC}=\lim_{P\to\infty}\mathrm{E}\left\{\log_2\left(1+\frac{|H_m|^2 P}{2\sigma_\mathrm{n}^2+\sum\limits_{i\in\mathcal{M}\setminus m}|H_i|^2 P}\right)\right\}_{\boldsymbol{H}}=\mathrm{E}\left\{\log_2\left(1+\frac{|H_m|^2}{\sum\limits_{i\in\mathcal{M}\setminus m}|H_i|^2}\right)\right\}_{\boldsymbol{H}}, \quad (4.17)$$

where the result must be independent of m. Without fading it evidently holds that

$$R_\mathrm{max}^\mathrm{CMAC}=\log_2\left(1+\frac{1}{M-1}\right). \quad (4.18)$$

4.3.2 Joint Decoding

To characterise the achievable rates for the classical fading MAC, the notion of *capacity region* will be introduced, further the closely related *power region*. These regions define then the feasible sets of the related resource allocation problems formulated as linear programs further below.

Achievable Rates, Capacity Region and Power Region

The achievable rates $R_m \in \mathbb{R}_+$ in the ergodic case are constrained by [EB98]

$$\sum_{m\in S} R_m \leq \mathrm{E}\left\{\log_2\left(1+\frac{1}{2\sigma_{\mathrm{n}}^2}\sum_{m\in\mathcal{S}}|H_m|^2 P_m\right)\right\}_{\boldsymbol{H}}, \quad \forall \mathcal{S}\subseteq\mathcal{M}. \tag{4.19}$$

Note that (4.19) defines a region (or set) of rates that is bounded by (2^M-1) hyperplanes of dimension $(M-1)$ in $\mathbb{R}_+^M$.

Based on (4.19) the definitions for the *capacity region* and the *power region* for the classical fading MAC are provided in the following.

Capacity Region The ergodic capacity region is the set of rate tuples[5] $\mathbf{R}=(R_1,R_2,\ldots,R_M)$ that satisfy (4.19), i.e. are achievable if a power constraint tuple $\mathbf{P}=(P_1,P_2,\ldots,P_M)$ is given.

This region of rates for the classical fading MAC is given by [EB98, TH98a]

$$\mathcal{R}_{\mathrm{CMAC}}=\left\{\mathbf{R}\in\mathbb{R}_+^M : \sum_{m\in\mathcal{S}} R_m \leq \mathrm{E}\left\{\log_2\left(1+\frac{1}{2\sigma_{\mathrm{n}}^2}\sum_{m\in\mathcal{S}}|H_m|^2 P_m\right)\right\}_{\boldsymbol{H}}, \forall\mathcal{S}\subseteq\mathcal{M}\right\}, \tag{4.20}$$

where the mean receive powers P_m are considered as given and fixed.

A set of rates $(R_1,R_2,\ldots,R_M)$ is said to be achievable, if all rates R_m, $m=1,2,\ldots,M$, lie inside the capacity region $\mathcal{R}_{\mathrm{CMAC}}$ defined by (4.20).

The *dominant face* $\mathcal{D}_{\mathcal{R}_{\mathrm{CMAC}}}$ of the region of achievable rates $\mathcal{R}_{\mathrm{CMAC}}$ is defined as the set of points that fulfill (4.20) with equality for $\mathcal{S}=\mathcal{M}$. Points on the dominant face $\mathcal{D}_{\mathcal{R}_{\mathrm{CMAC}}}$ are characterised by the property that it is not possible to increase the rate of one source without decreasing the rate of another source in order to stay inside the region of achievable rates [RU96, TH98a].

In [TH98a] it is pointed out that the capacity region $\mathcal{R}_{\mathrm{CMAC}}$ is a polymatroid, and this can be readily proved by verifying that (4.19) (which is used in (4.20) to define $\mathcal{R}_{\mathrm{CMAC}}$) is a rank function according to (4.1) – (4.3)

Power Region Equivalently to the above, the power region is defined as the set of power tuples $\mathbf{P}$ that satisfy (4.19) for given rate tuple $\mathbf{R}$.

If we consider fixed rate tuple $\mathbf{R}$, then the required mean receive powers $\mathbf{P}$ can be calculated such that the corresponding region $\mathcal{R}_{\mathrm{CMAC}}$ contains the rate tuple $\mathbf{R}$. Thus, the set of powers $\mathbf{P}$ that achieve a given set of rates $\mathbf{R}$ is implicitly defined by

$$\mathcal{P}_{\mathrm{CMAC}}=\left\{\mathbf{P}\in\mathbb{R}_+^M : \sum_{m\in\mathcal{S}} R_m \leq \mathrm{E}\left\{\log_2\left(1+\frac{1}{2\sigma_{\mathrm{n}}^2}\sum_{m\in\mathcal{S}}|H_m|^2 P_m\right)\right\}_{\boldsymbol{H}}, \forall\mathcal{S}\subseteq\mathcal{M}\right\}, \tag{4.21}$$

where now the R_m are considered fixed and given.

[5] equivalently the terms *point* or *vector* will be used

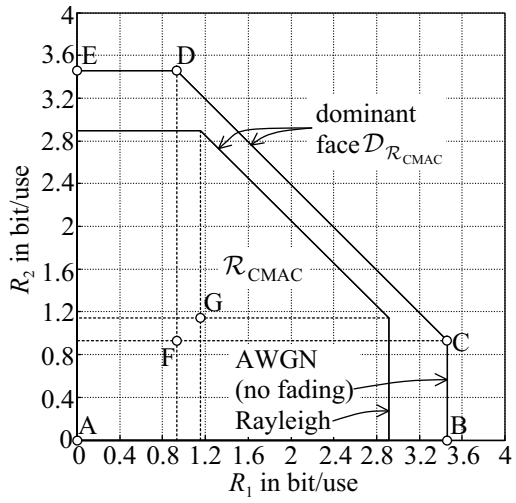
The dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{CMAC}}}$ of the power region $\mathcal{P}_{\mathrm{CMAC}}$ is equivalently defined with $\mathcal{S} = \mathcal{M}$ in (4.21), with the difference, of course, that points on the dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{CMAC}}}$ are characterised by the property that it is not possible to decrease the power of a single source without increasing the power of another source in order to stay inside the power region.

The power tuples lying on the boundary of $\mathcal{P}_{\mathrm{CMAC}}$ correspond to a rate tuple $\mathbf{R}$ lying on the boundary of $\mathcal{R}_{\mathrm{CMAC}}$, the same holds in particular for the vertices. Furthermore, it can be readily deduced from the above definition of the dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{CMAC}}}$ that the optimal power tuple must lie on $\mathcal{D}_{\mathcal{P}_{\mathrm{CMAC}}}$.

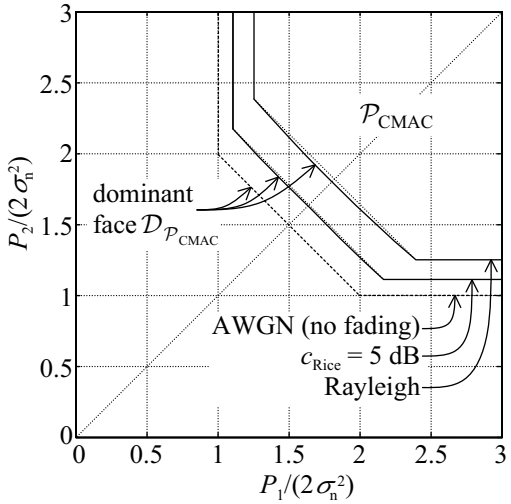
In contrast to $\mathcal{R}_{\mathrm{CMAC}}$, the power region $\mathcal{P}_{\mathrm{CMAC}}$ is not a polytope which can be proved by verifying that the dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{CMAC}}}$ of $\mathcal{P}_{\mathrm{CMAC}}$ is a *concave* set [Mec00a] (also cf. Fig. 4.4(b)). (Faces of convex polytopes are again convex polytopes, therefore it follows if a face of $\mathcal{P}_{\mathrm{CMAC}}$ is not convex, it cannot be a polytope [PS98].)

Hence, $\mathcal{P}_{\mathrm{CMAC}}$ is *not* a contra-polymatroid for the fading channel, except for constant fading factors h_m, i.e. for the non-fading AWGN MAC, which follows from the fact that, for constant h_m, the right hand side of (4.19) is a generalised symmetric rank function according to (4.7).

Example 4.1: **Capacity and Power Region for $M = 2$ Sources** For clarification of the definitions (4.20) and (4.21), Fig. 4.4 shows examples for $\mathcal{R}_{\mathrm{CMAC}}$ and $\mathcal{P}_{\mathrm{CMAC}}$ for $M = 2$ sources (cf. [CT91]).



(a) $P_1/(2\sigma_{\mathrm{n}}^2) = P_2/(2\sigma_{\mathrm{n}}^2) = 10$. The pentagon A-B-C-D-E circumscribes $\mathcal{R}_{\mathrm{CMAC}}$ for the non-fading channel. The inner pentagon indicates $\mathcal{R}_{\mathrm{CMAC}}$ for the Rayleigh fading channel. Note that the region of achievable rates is reduced for the Rayleigh fading channel, however, the rates achieved by independent decoding are increased (points F,G).



(b) $\mathcal{P}_{\mathrm{CMAC}}$ for $R_1 = R_2 = 1$. The dominant face of the power regions for fading channels is concave. (Note that the power regions extend towards $P_1/(2\sigma_{\mathrm{n}}^2), P_2/(2\sigma_{\mathrm{n}}^2) \to \infty$.)

Figure 4.4: Region of achievable rates $\mathcal{R}_{\mathrm{CMAC}}$ and power region $\mathcal{P}_{\mathrm{CMAC}}$ for the non-fading, Rice, and Rayleigh channel. The fading coefficients are normalised to unit power, $\mathrm{E}\left\{|H_1|^2\right\} = \mathrm{E}\left\{|H_2|^2\right\} = 1$.

■

Optimal Resource Allocation: Rate and Power Allocation

The definitions of the region of achievable rates $\mathcal{R}_{\mathrm{CMAC}}$ and power region $\mathcal{P}_{\mathrm{CMAC}}$ allow to investigate in detail the following optimisation problems [TH98a]:

- Given a power tuple $\mathbf{P}$, the optimal rate allocation is sought.
- Given rate tuple $\mathbf{R}$ the optimal power allocation is sought.

Optimality for rate allocation is defined as maximising the weighted sum of rates $\boldsymbol{\lambda}_{\mathrm{R}} \cdot \mathbf{R}$ for given power tuple $\mathbf{P}$, where $\boldsymbol{\lambda}_{\mathrm{R}} \in \mathbb{R}_{+}^{M}$ is a vector of rate rewards, which can be used to possibly prioritise some sources over others [TH98a].

Likewise, optimality for power allocation will be defined as minimising the weighted sum of mean receive powers $\boldsymbol{\lambda}_{\mathrm{P}} \cdot \mathbf{P}$ for given rate vector $\mathbf{R}$, here $\boldsymbol{\lambda}_{\mathrm{P}} \in \mathbb{R}_{+}^{M}$ is a vector of power costs.

Optimal Rate Allocation Firstly, we turn to the problem of maximising the weighted rate sum $\boldsymbol{\lambda}_{\mathrm{R}} \cdot \mathbf{R}$, if received powers P_m are given.

Because the capacity region $\mathcal{R}_{\mathrm{CMAC}}$ is a polymatroid, the optimisation problem

$$\max_{\mathbf{R}} \{\boldsymbol{\lambda}_{\mathrm{R}} \cdot \mathbf{R} : \mathbf{R} \in \mathcal{R}_{\mathrm{CMAC}}\} \tag{4.22}$$

is a *linear program* for all receive power vectors $\mathbf{P}$ and fading statistics [TH98a] (according to Sec. 4.1.2, $\boldsymbol{\lambda}_{\mathrm{R}} \cdot \mathbf{R}$ is the *objective function*, and $\mathcal{R}_{\mathrm{CMAC}}$ is the *feasible set* for $\mathbf{R}$), and can be solved most efficiently by exploiting the polymatroidal structure of the feasible set [TH98a].

Because $\mathcal{R}_{\mathrm{CMAC}}$ is a polymatroid, and according to the characterisation of the vertices of a polymatroid provided by (4.9), the rates at any vertex v of $\mathcal{R}_{\mathrm{CMAC}}$ are given by (we define that $\sum_{i=a}^{b} x_i = 0$ for $b < a$)

$$\begin{aligned} R_{\pi_v(m)} =& \mathrm{E}\left\{ \log_2\left(1 + \frac{1}{2\sigma_{\mathrm{n}}^2} \sum_{i=1}^{m} \left|H_{\pi_v(i)}\right|^2 P_{\pi_v(i)}\right)\right\}_{\boldsymbol{H}} - \\ & \mathrm{E}\left\{ \log_2\left(1 + \frac{1}{2\sigma_{\mathrm{n}}^2} \sum_{i=1}^{m-1} \left|H_{\pi_v(i)}\right|^2 P_{\pi_v(i)}\right)\right\}_{\boldsymbol{H}}, \\ & m = 1, 2, \ldots, M, \end{aligned} \tag{4.23}$$

where the optimal permutation $\boldsymbol{\pi}_{v^*}$, which solves the posed optimisation problem (4.22), is given by (cf. Sec. 4.1.1 and 4.1.2)

$$\lambda_{\mathrm{R},\pi_{v^*}(1)} \geq \lambda_{\mathrm{R},\pi_{v^*}(2)} \geq \cdots \geq \lambda_{\mathrm{R},\pi_{v^*}(M)}, \tag{4.24}$$

with $\boldsymbol{\lambda}_{\mathrm{R}} = (\lambda_{\mathrm{R},1}, \lambda_{\mathrm{R},2}, \cdots, \lambda_{\mathrm{R},M})$.

Note that the rates at a vertex according to (4.23) are also obtained from the *chain rule of mutual information* [CT91, RU96], and, therefore, any capacity region is a polymatroid according to the characterisation of the vertices of a polymatroid (Sec. 4.1.1).

The structure of the decoder that achieves the rates given by (4.23) is discussed in Sec. 4.3.3.

Optimal Power Allocation Because $\mathcal{P}_{\text{CMAC}}$ is no polytope it directly follows that

$$\min_{\mathbf{P}} \left\{ \boldsymbol{\lambda}_{\text{P}} \cdot \mathbf{P} : \mathbf{P} \in \mathcal{P}_{\text{CMAC}} \right\}, \tag{4.25}$$

is **not** a linear program, but it can be readily shown that both the objective function $\boldsymbol{\lambda}_{\text{P}} \cdot \mathbf{P}$, as well as the feasible set $\mathcal{P}_{\text{CMAC}}$ are convex and, hence, (4.25) formulates a *convex program* [PS98]. Therefore, the optimum power allocation can be computed via standard optimisation techniques [Mec00a]. However, note that the feasible set $\mathcal{P}_{\text{CMAC}}$ is defined by an exponential number $(2^M - 1)$ of constraints, such that, although it is guaranteed that the global optimum to (4.25) can by found in a limited number of steps, computation of the solution to (4.25) can take a prohibitive amount of time.

4.3.3 Successive Decoding

The capacity and power regions provide the ultimate limits of either rates that are achievable with given power constraints $\mathbf{P}$, or, vice versa, the powers that are required to achieve given target rates $\mathbf{R}$.

It is known that certain points of the capacity region are achievable with an implementation complexity significantly less than a general point, where the complexity is defined by the required effort in encoding the information at the M sources and decoding it at the receiver [RU96].[6]

Generally, to achieve points on the dominant face is of particular interest, because, firstly considering the capacity region, for any point inside the capacity region, at least one component of the rate vector $\mathbf{R}$ can be increased until the dominant face $\mathcal{D}_{\mathcal{R}_{\text{CMAC}}}$ is reached, while the other elements of $\mathbf{R}$ remain fixed; it is said that any point in a capacity region is dominated by some point on the dominant face [RU96]. Equivalently, given a point inside the power region, at least one component can be reduced with the other powers being fixed, until the dominant face $\mathcal{D}_{\mathcal{P}_{\text{CMAC}}}$ is reached.

The $M!$ vertices of $\mathcal{R}_{\text{CMAC}}$ and $\mathcal{P}_{\text{CMAC}}$, respectively, are on the according dominant face and are of particular interest, because, as we will see in the following, rate tuples at the vertex of the capacity region are achieved with an implementation complexity significantly less than that required to achieve a general point in $\mathcal{R}_{\text{CMAC}}$ [CT91, RU96].

In the following it will be described how rates at the vertices of the capacity region can be achieved at reduced receiver complexity.

Using basic algebraic manipulations, (4.23) can be rewritten according to

$$R_{\pi_v(m)} = \text{E}\left\{ \log_2 \left(1 + \frac{\left|H_{\pi_v(m)}\right|^2 P_{\pi_v(m)}}{2\sigma_{\text{n}}^2 + \sum_{i=1}^{m-1} \left|H_{\pi_v(i)}\right|^2 P_{\pi_v(i)}} \right) \right\}_{\boldsymbol{H}}, \qquad m = 1, 2, \ldots, M. \tag{4.26}$$

Comparison with the rate achievable by independent decoding (4.16), reveals that (4.26) is evidently the rate achievable by independent decoding of source $\pi_v(m)$ while all sources in the set

$$\mathcal{S}_{v,m-1} = \{\pi_v(1), \pi_v(2), \ldots, \pi_v(m-1)\} \tag{4.27}$$

[6] E.g., joint encoding/decoding of all users has a decoding complexity of 2^{nMR} if M sources transmit at rate R and block length n.

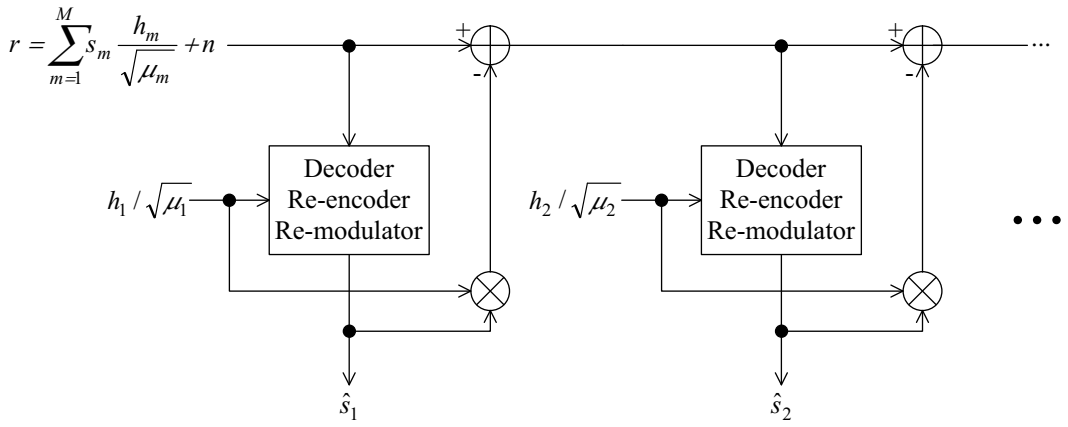


Figure 4.5: Successive decoding for the classical fading MAC. (Decoding order is assumed as $1, 2, \ldots$ to simplify notation.)

have to be considered as interferers [RU96]. Hence, when decoding source $\pi_v(m)$, only $(m-1)$ sources out of $(M-1)$ potential interferers actually contribute to interference.

This means that rates at the vertices of the capacity region, solving the optimisation problem (4.22), can be achieved by successive decoding [CT91], that is a decoder that removes the contributions of already decoded source signals from the received signal r given by (4.3) (cf. Fig. 4.5).

The decoding order is $(\pi_v(M), \pi_v(M-1), \ldots, \pi_v(1))$. So, starting with $m = M$, where initially $(M-1)$ interferers are relevant, source $\pi_v(M)$ is decoded. Subsequently, re-encoding and re-modulating allows to subtract the contribution of source $\pi_v(M)$ from the received symbol r, such that source $\pi_v(M)$ does not contribute to the interference as seen by the sources still to be decoded. All sources are decoded, re-encoded and re-modulated in that way, and, finally, for $m = 1$ only thermal noise deteriorates the signal (cf. (4.23)) [Wyn74, BC74].

This decoding procedure is known variously as *onion peeling*, *stripping*, *successive cancellation*, *successive decoding*, *interference cancellation*, and *superposition coding* [RU96].

Power Allocation for Successive Decoding

The result of the above considerations is that it says that important rate tuples on the dominant face of the capacity region (where the sum-rate is maximised for a given receive power allocation $\mathbf{P}$) can be achieved with M single-user decoders of comparably low complexity and successive decoding.

Vice versa, if we demand a rate tuple $\mathbf{R}$, then the powers to be allocated can in general be significantly reduced compared to independent decoding without successive decoding as will be demonstrated later.

While the optimal solution to the rate allocation problem is always found at a vertex of the capacity region, we have seen in the last section that this is possibly not so for the power region, because the optimum solution to the minimum sum power allocation problem (4.25) is not necessarily a vertex of $\mathcal{P}_{\text{CMAC}}$, since the feasible set $\mathcal{P}_{\text{CMAC}}$ is no polytope. This means that the optimal power allocation does, in general, **not** lead to successive decoding and vice versa.

However, optimality with respect to (4.25) may be sacrificed in favour of a comparably simple decoder structure that still achieves rate tuples on the dominant face of $\mathcal{R}_{\text{CMAC}}$, providing significant advantage over independent decoding (i.e. higher achievable rate, lower transmit powers).

Optimal Decoding Order for Successive Decoding Generally, the powers are allocated in the order $\pi_v(1), \pi_v(2), \ldots, \pi_v(M)$. Given a permutation $\boldsymbol{\pi}_v$ and target rates $\mathbf{R}$ the required transmit power $p_{\pi_v(1)}$ can directly be calculated using (4.23), without the need to account for the other sources. Then, the interference relevant for the source $\pi_v(2)$ being second-last in the decoding order is known, because only source $\pi_v(1)$ has to be considered as interference. Hence, with the order of power allocation being reversed with respect to the decoding order, unique transmit powers can be allocated to all users.

Restricting in the further to successive decoding and demanding equal rates for all sources, i.e. $R_m = R$, it must hold

$$R_{\pi_v(m)} = R = \mathrm{E}\left\{\log_2\left(1 + \frac{\left|H_{\pi_v(m)}\right|^2 P_{\pi_v(m)}}{2\sigma_\mathrm{n}^2 + \sum_{i=1}^{m-1}\left|H_{\pi_v(i)}\right|^2 P_{\pi_v(i)}}\right)\right\}_{\boldsymbol{H}}, \qquad m = 1, 2, \ldots, M. \quad (4.28)$$

This provides the power allocation procedure for given π_v, but it remains to determine that permutation $\boldsymbol{\pi}_{v^*}$ for that

$$\min_v \left\{\boldsymbol{\lambda}_\mathrm{P} \cdot \mathbf{P} : R = \mathrm{E}\left\{\log_2\left(1 + \frac{\left|H_{\pi_v(m)}\right|^2 P_{\pi_v(m)}}{2\sigma_\mathrm{n}^2 + \sum_{i=1}^{m-1}\left|H_{\pi_v(i)}\right|^2 P_{\pi_v(i)}}\right)\right\}_{\boldsymbol{H}}, \forall m \in \mathcal{M}\right\} \quad (4.29)$$

is solved.

Eqn. (4.29) formulates a *combinatorial* optimisation problem. In this class of optimisation problems the optimal solution is attained for an object from a finite (or countably infinite) set, such as an integer, set, permutation, or graph [PS98].

The brute force approach for solving (4.29) is enumeration of all $M!$ permutations[7] $\boldsymbol{\pi}_v$. Of course, depending on the structure of the underlying optimisation problem, there are more elaborate approaches to solve a combinatorial optimisation problem, but for many problems of this class no solution algorithms except enumeration are known which provable provide the optimal solution (e.g. for the knapsack problem) (see, e.g. [PS98]).

This becomes clearer if we represent the problem to determine the optimal permutation as a tree graph (see Fig. 4.6). Each branch of the tree is corresponding to a particular decoding order. At the m-th stage of the tree there are $(M - m)$ edges leaving each vertex, corresponding to the possible choices in the decoding order; each edge connecting a vertex $\pi_v(m-1)$ in level $(m-1)$ with a vertex $\pi_v(m)$ in level m is associated with the corresponding cost $\lambda_{\mathrm{P},\pi_v(m)} P_{\pi_v(m)}$.

Aiming at the minimisation of sum transmit power $\sum_{m\in\mathcal{M}} p_m$, we choose $\lambda_{\mathrm{P},m} = \mu_m$ [Mec00a], because then it holds with (4.15)

$$\sum_{m\in\mathcal{M}} \mu_m P_m = \sum_{m\in\mathcal{M}} p_m, \quad (4.30)$$

and, therefore, (4.25) and (4.29), respectively, formulate the minimisation of the sum-*transmit* power.

[7] Because of the rapid growth of the factorial $x!$, the term *Combinatorial Explosion* was coined.

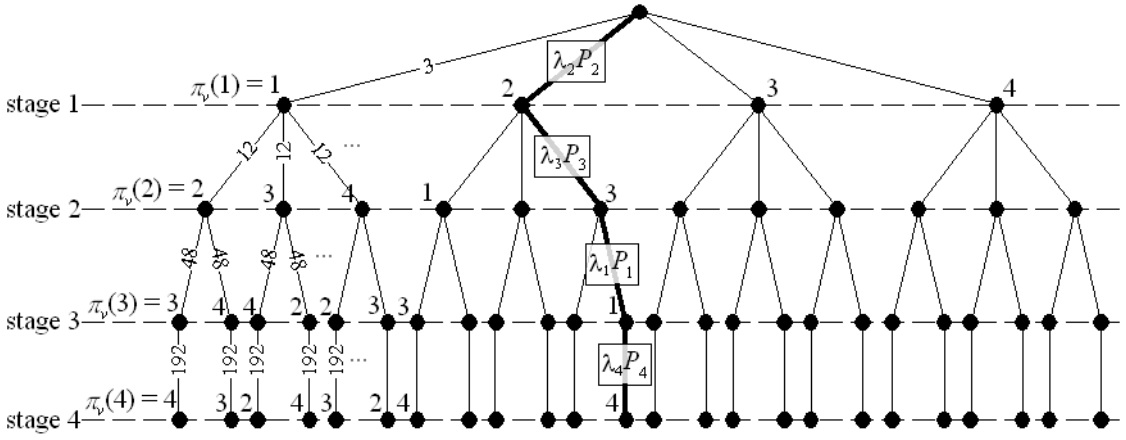


Figure 4.6: Tree graph representing the possible decoding orders for successive decoding for $M = 4$. The branch associated with $\boldsymbol{\pi}_v = (2, 3, 1, 4)$ is highlighted (bold lines), the according decoding order is $(4, 1, 3, 2)$. If for all sources equal rates $R_m = R$ are required, the necessary receive power at stage m for source $\pi_v(m)$ are independent of the decoding order itself, i.e. $P_{\pi_{v1}(m)} = P_{\pi_{v2}(m)}$ for any two arbitrary permutations $\boldsymbol{\pi}_{v1}, \boldsymbol{\pi}_{v2}$.

It is quite a trivial but none the less essential observation that, if there is no known structure in the edge weights that can be exploited, there cannot be an algorithm other than enumeration of all paths that will find the path with minimum sum of weights. However, for the classical fading MAC this is indeed not required.

If we demand equal rates $R_m = R$ for all sources and assuming further equal statistics for the fading factors h_m, then it is evident from (4.28) that the vector of received powers, reordered according to $\boldsymbol{\pi}_v$, i.e. $(P_{\pi_v(1)}, P_{\pi_v(2)}, \ldots, P_{\pi_v(M)})$, is independent of the chosen permutation, i.e.

$$P_{\pi_{v1}(m)} = P_{\pi_{v2}(m)} \qquad m = 1, 2, \ldots M, \tag{4.31}$$

for any two arbitrary permutations $\boldsymbol{\pi}_{v1}, \boldsymbol{\pi}_{v2}$. Note that if we choose two permutations with $\pi_{v1}(m) = \pi_{v2}(m)$, then, because of $P_{\pi_{v1}(m)} = P_{\pi_{v2}(m)}$, it follows directly that the power allocated at stage m is independent of the decoding order of the sources decoded after (i.e. stages 1 to $(m-1)$) and of that decoded before (i.e. stages $(m+1)$ to M).

From (4.28) we can also reason that $P_{\pi_v(m)} > P_{\pi_v(n)}$ for $m > n$ and for a given permutation $\boldsymbol{\pi}_v$. Therefore, if $P_m > P_n$ (for any $m, n \in \mathcal{M}$, $m \neq n$), and assuming that $\mu_m > \mu_n$, then the sum transmit power is reduced by $(P_m - P_n)(\mu_m - \mu_n)$ if the positions of source m and n in the decoding order are exchanged.

Finally, the optimal permutation $\boldsymbol{\pi}_{v^*}$ solving (4.29) is as is known provided by choosing that permutation for that [Mec00a]

$$\mu_{\pi_{v^*(1)}} \geq \mu_{\pi_{v^*(2)}} \geq \cdots \geq \mu_{\pi_{v^*(M)}}. \tag{4.32}$$

Therefore it is the optimal approach for an arbitrary number of sources to decode the sources in the order of increasing link loss, i.e. to decode that source first that has the least link loss, while decoding last that source which shows the highest link loss.

Optimal power allocation for successive decoding has a greedy solution, allowing to solve this optimisation problem very efficiently. However, bear in mind that this solution is not necessarily a

global optimum for the power allocation problem, as the power region is no polytope as explained before!

Example 4.2: **Optimal Power Allocation and Power Allocation for Successive Decoding for 2 Sources** To illustrate in more detail the difference between the minimum sum-power allocation and the powers required for successive decoding to achieve rates $R_1, R_2 \geq R$ we again consult the simple example of 2 sources. Fig. 4.7(a) shows the power region $\mathcal{P}_{\mathrm{CMAC}}$ obtained from (4.21) for 2 sources for a fading MAC (Rice, $c_{\mathrm{R}} = 0$ dB). Two cases are considered regarding the link loss, namely $\mu_1 = \mu_2$ and $\mu_1 = 2\mu_2$. Further, it will be assumed $R = 1$. In Fig. 4.7(a) the power regions for required rates $R = 1$ are shown, further the objective functions for $\boldsymbol{\lambda}_{\mathrm{P}} = (\mu_1, \mu_2)$ (cf. (4.25) and (4.30)) with $\mu_1 = \mu_2$ and $\mu_1 = 2\mu_2$. Depending on the concavity of the dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{CMAC}}}$ and on the particular values of μ_1, μ_2, the optimal receive power tuple (P_1, P_2) solving (4.25) may be obtained at a vertex of $\mathcal{P}_{\mathrm{CMAC}}$. Then successive decoding is optimal in the sense of minimum sum transmit power.

■

Above we have understood optimality of $\mathbf{P}$ in the context of minimising the sum transmit power

$$\min_{\mathbf{p} \in \mathcal{P}_{\mathrm{CMAC}}} \sum_{m \in \mathcal{M}} p_m. \tag{4.33}$$

In particular for the case of independent sources, that is considered in this work, minimisation of the maximum transmit power may seem more appropriate:

$$\min_{\mathbf{p} \in \mathcal{P}_{\mathrm{CMAC}}} \max_{m \in \mathcal{M}} p_m. \tag{4.34}$$

Power allocation under this optimality constraint is described in [TH98a], where availability of perfect CSI at the transmitters is considered. However, the solution is rather involved and, therefore, we will restrict to optimality in terms of minimal sum transmit power as defined by (4.33).

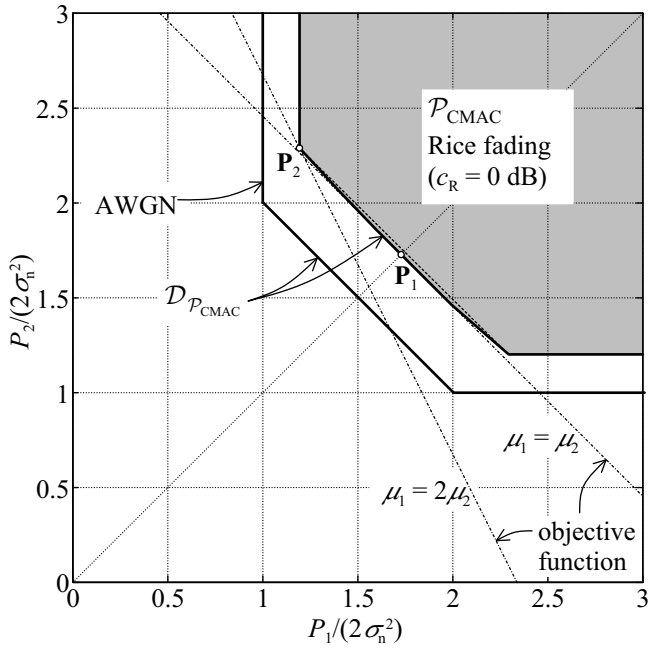
4.3.4 Comparison Independent and Successive Decoding

Fig. 4.8 shows a comparison of the required sum-transmit power for $M = 20$ sources, where all μ_m are equal with $\mu_m = 1$, and, further, the AWGN case without fading is considered (cf. [Mec00a]).

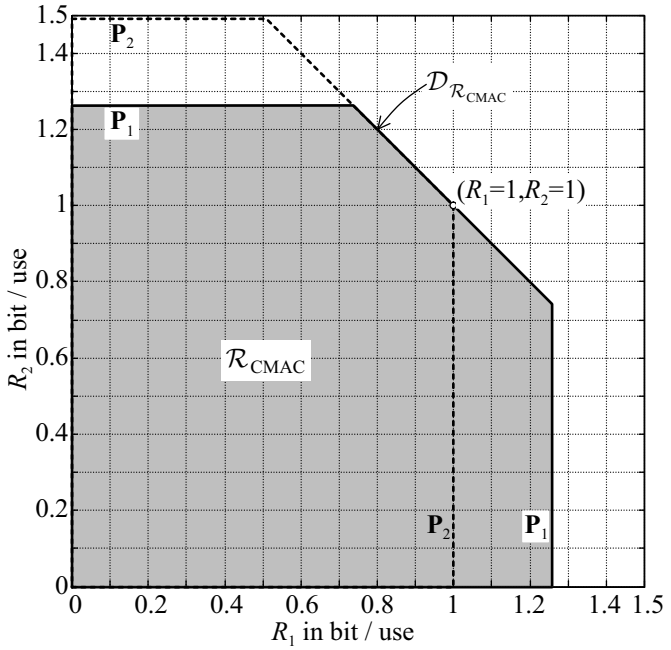
The sum transmit power required for successive decoding is obtained from (4.26), where it is sufficient to consider a single arbitrary permutation, because all permutations produce the same sum transmit power for the classical MAC as all $\mu_m = 1$ by assumption (this follows from (4.32)).

While the maximal achievable rate for independent decoding is according to (4.18) given by $R_{\max}^{\mathrm{CMAC}} = \log_2(1 + 1/19) = 0.074$ bit/channel use, no rate limit exists for successive decoding.

This finally concludes the review of the characterisation of the classical fading MAC. With the understanding of the terms and concepts related to the classical fading MAC presented above, we can now turn to the power allocation for successive decoding for the fading vector MAC.



(a) Power region $\mathcal{P}_{\mathrm{CMAC}}$ (also the power region for AWGN is shown for comparison). In general, because the dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{CMAC}}}$ is not convex for fading, the optimal power tuple differs from that for successive decoding.



(b) Capacity regions for power tuple $\mathbf{P}_1$ according to (a) (optimal for $\mu_1 = \mu_2$) and for successive decoding corresponding to power tuple $\mathbf{P}_2$ (here, optimal for $\mu_1 = 2\mu_2$). Power allocation according to $\mathbf{P}_1$ does **not** allow to achieve required rates via successive decoding.

Figure 4.7: Power region $\mathcal{P}_{\mathrm{CMAC}}$ for 2 sources to achieve rates $R_1, R_2 \geq R$ with $R = 1$. Further, the capacity regions for equal receive powers and for successive decoding with decoding order $2 \rightarrow 1$ are shown. Rice fading with $c_{\mathrm{R}} = 0$ dB is assumed.

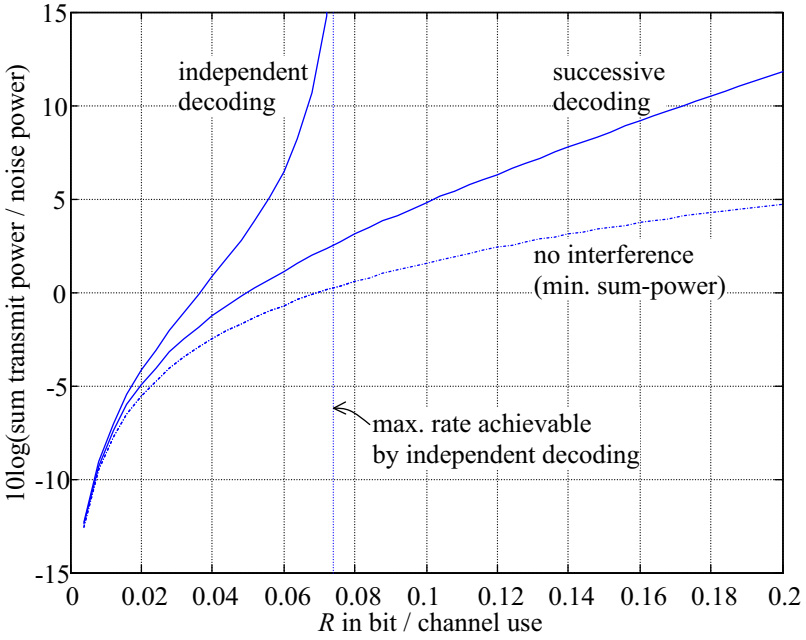


Figure 4.8: Comparison of the required sum transmit power for independent and successive decoding for $M = 20$ sources with equal steering vectors (resulting in the classical MAC), without fading (AWGN). Also shown is the sum transmit power for the interference free case, which is providing the lower bound for sum transmit power.

4.4 Independent Decoding for the Fading Vector Multiple-Access Channel

Due to antenna array reception there is the opportunity to efficiently separate user signals based on the spatial dimension (mirrored in the steering vectors $\mathbf{a}_m$), which is in contrast to the classical MAC [SXLK98].

In fact, spatial separation could, in the best case, create independent transmission channels for the respective sources, transforming the MAC into independent channels that are not impaired by interference but only thermal noise, therefore allowing to achieve channel capacity with a single-user decoder.

Since the user signals are received by an antenna array, an essential part of the possible receiver structures will be a beamforming stage to collect the signals received by the array elements.

This gives rise to the question how close the performance of single-user decoding with a preceding optimal beamforming stage can get to optimal joint decoding considering realistic satellite scenarios.

As already pointed out before, the receive signal vector $\mathbf{r}$ and its covariance matrix $\mathbf{K}_\mathbf{r}$ can be split into the contribution of the wanted signal from source m, the interference from other sources, and the noise (cf. (3.87)). We will use the according definitions introduced in Sec. 3.5.3 to provide in the following the formulations for the mutual information and achievable rates for independent decoding.

Mutual information is the amount of information that one random variable contains about another random variable; we will adopt the notation of [CT91], where capital italics (X) are used to indicate a random variable, while minuscule italics (x) indicate the concrete realisations of a random variable.

Considering independent decoding for the vector MAC, the two random variables that we look at in connection with mutual information are on the one side the source signal S_m, and on the other side the receive signal vector $\boldsymbol{R}$ (with s_m and $\mathbf{r}$, respectively, being the concrete realisations).

The mutual information between source signal s_m and receive signal vector $\mathbf{r}$ is given by the entropy of the receive signal vector $\mathcal{H}(\boldsymbol{R})$ minus the entropy of the receive signal vector conditioned on the transmit signal of source m, $\mathcal{H}(\boldsymbol{R}|\, S_m)$ [CT91, SXLK98]

$$\begin{aligned} \mathcal{I}(S_m; \boldsymbol{R}) =& \mathcal{H}(\boldsymbol{R}) - \mathcal{H}(\boldsymbol{R}|S_m), \\ & \mathcal{I}_m = \mathcal{M} \setminus m. \end{aligned} \tag{4.35}$$

The rates that are achievable with independent decoding, if the joint fading state $\mathbf{h} = (h_1, h_2, \ldots, h_M)$ is fixed, are those of the AWGN vector MAC, where the signal powers are scaled by $|h_m|^2 / \mu_m$.

In consequence, using the definitions of Sec. 3.5.3, the region of achievable rates is obtained from

(4.35) according to (cf. [SXLK98])

$$\begin{aligned} R_m(\mathbf{h}) \;\leq\; & \max_{p(s_1)p(s_2)\ldots p(s_M):\mathrm{E}\{|s_j|^2\}\leq p_j} \mathcal{I}(S_m;\boldsymbol{R}) \\ & = \log_2\left(\frac{\det\left(\mathbf{K}_\mathbf{r}(\mathcal{I}_m)+\frac{p_m|h_m|^2}{\mu_m}\mathbf{a}_m\mathbf{a}_m^{\mathrm{H}}\right)}{\det\left(\mathbf{K}_\mathbf{r}(\mathcal{I}_m)\right)}\right) \\ & = \log_2\left(\det\left(\mathbf{I}_L+\frac{p_m\,|h_m|^2}{\mu_m}\mathbf{a}_m\mathbf{a}_m^{\mathrm{H}}\mathbf{K}_\mathbf{r}(\mathcal{I}_m)^{-1}\right)\right) \\ & = \log_2\left(1+\frac{p_m\,|h_m|^2}{\mu_m}\mathbf{a}_m^{\mathrm{H}}\mathbf{K}_\mathbf{r}(\mathcal{I}_m)^{-1}\mathbf{a}_m\right), \qquad \mathcal{I}_m=\mathcal{M}\setminus m, \end{aligned} \tag{4.36}$$

where the array output covariance matrix $\mathbf{K}_\mathbf{r}$ was introduced in (3.78), and $\mathbf{K}_\mathbf{r}(\mathcal{I}_m)$ associated with the sources interfering with source m was defined in (3.83); in the last step we have used the identity (A.2). The maximisation in (4.36) is performed over the product distribution of the source symbols under the transmit power constraint $\mathrm{E}\{|s_j|^2\}\leq p_j$ ($\forall j\in\mathcal{M}$), where it is well known that the maximum is achieved for normal distributed source symbols [CT91].

Obviously, the rate $R_m(\mathbf{h})$ that is achievable for source m with fixed joint fading state $\mathbf{h}$, if all other sources are considered as interference, depends only on the optimal SINR

$$\Gamma_{\mathrm{opt},m}(\mathbf{h})=\frac{p_m\,|h_m|^2}{\mu_m}\mathbf{a}_m^{\mathrm{H}}\mathbf{K}_\mathbf{r}(\mathcal{I}_\mathbf{m})^{-1}\mathbf{a}_m, \qquad \mathcal{I}_m=\mathcal{M}\setminus m, \tag{4.37}$$

which we have already come across in (3.90) in Sec. 3.5.4 and which is achievable with optimal beamforming [SXLK98].

It is interesting to note that in the calculation of the achievable rate for independent decoding no explicit assumptions on using beamforming were used. However, the equation for achievable rate $R_m(\mathbf{h})$ (4.36) together with (4.37) says that a receiver employing optimal beamforming can in principle reach $R_m(\mathbf{h})$ [SXLK98].

To consider now the effect of the time-dependent fading on achievable rates, the ergodic rates are considered by averaging $R_m(\mathbf{h})$ in (4.36) over the joint fading state $\mathbf{h}$, where it is further assumed that the receiver has perfect CSI available.

No CSI is available at the transmitters preventing to compensate the fading h_m by adaptively controlling the transmit power p_m (cf. Sec. 4.2.2). Therefore, averaging over the joint fading state $\mathbf{h}$ is performed without allowing the transmitters to adapt the p_m according to $\mathbf{h}$.

Then the achievable rates are under these assumptions given by (cf. [EB98])

$$\boxed{R_m\leq\mathrm{E}\left\{\log_2\left(1+\frac{p_m\,|H_m|^2}{\mu_m}\mathbf{a}_m^{\mathrm{H}}\mathbf{K}_\mathbf{r}(\mathcal{I}_\mathbf{m})^{-1}\mathbf{a}_m\right)\right\}_{\boldsymbol{H}},} \tag{4.38}$$

where the expectation operation is performed with respect to the joint fading state $\mathbf{h}$.

The receiver structure that can achieve the rates R_m given by (4.38) is depicted schematically in Fig. 4.9. Recall that the optimal beamforming vector $\mathbf{w}_{\mathrm{opt},m}$ of source m, defined in (3.89),

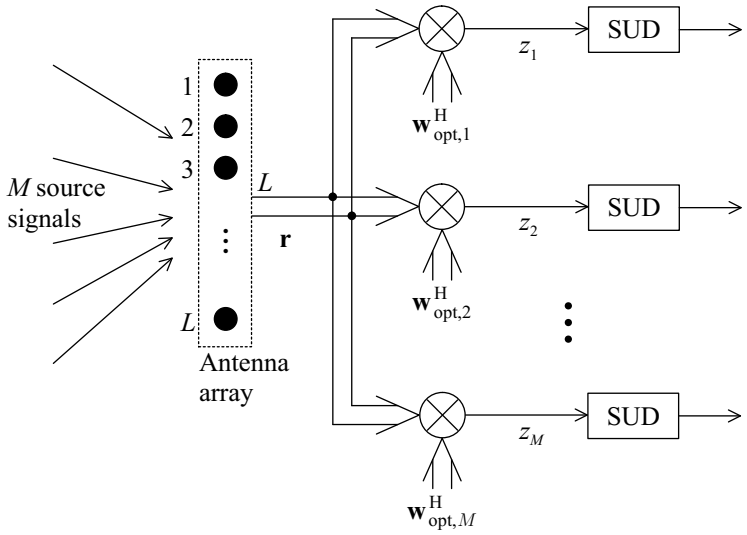


Figure 4.9: Block diagram of the receiver for the fading vector MAC employing optimal beamforming and independent (single-user) decoding for each source signal.

depends on the instantaneous receive powers $p_i \left|h_i\right|^2 / \mu_i$, $i \in \mathcal{I}_m$ of the sources interfering with source m (cf. Sec. 3.5.4).

In contrast to the classical MAC, there is in general **no rate limit** for the vector MAC with optimal beamforming and subsequent independent decoding, i.e. as the powers tend to infinity also the achievable rates tend to infinity, as long as the steering vectors are linear independent. This holds because interference can always be perfectly spatially filtered as long as the steering matrix $\mathbf{A} \in \mathbb{C}^{L \times M}$ has rank M, at the cost of possibly large, but finite noise amplification (e.g. [God97b]).

In the following an iterative power allocation algorithm will be described that guarantees demanded rates R_m according to (4.38) for all users while maintaining minimal transmit powers, if such a power allocation exists.

4.4.1 Power Allocation for Independent Decoding

Assume that for the sources support of a certain rate vector $\mathbf{R}$ is required. Then the power allocation is sought which achieves the required rates R_m for all sources, while maintaining minimal transmit powers $\mathbf{p}$, where we define $\mathbf{p} = (p_1, p_2, \ldots, p_m)$.

In [Mec00a] the case of the classical MAC with fading without array reception is considered and it is pointed out that by assuming equal rates for the transmitters, the receive powers $P_m = p_m / \mu_m$ must be equal as well. This is clearly not the case here, since in general an asymmetry is introduced by the array reception if more than 2 sources are considered.

However, the unique power vector $\mathbf{p}^*$ achieving rates $\mathbf{R}$ minimising transmit powers can be found by a simple iterative algorithm according to

$$p_m(n+1) = \mathrm{I}_m(\mathbf{p}(n)), \tag{4.39}$$

where $\mathrm{I}_m(\mathbf{p})$ is defined as the inverse function of (4.38) with respect to p_m, associating a unique p_m to a demanded R_m for given $\mathbf{p}$ and, further, $p_m(n)$ is the power allocated to user m at iteration step n.

If $\mathbf{p}^*$ exists then it is a fixed point of (4.39), such that

$$p_m^* = \mathrm{I}_m(\mathbf{p}^*), \qquad m \in \mathcal{M}, \tag{4.40}$$

and the p_m will converge to p_m^* for all initial power vectors $\mathbf{p}(0)$ and for all m. (A proof of the above allegations follows arguments presented in [Yat95].)

Note that the iteration rule can be modified such that update of powers by calculating $p_m(n+1)$ is based on $\mathbf{p}(n)$ and on the $p_l(n+1)$ $(l = 1, 2, \ldots, m-1)$ already calculated (it is assumed that the update of powers p_m is done in increasing order with respect to m, i.e. p_m is updated before p_{m+1}). Simulations show that convergence to $\mathbf{p}^*$ is faster compared to (4.39). However, we will not elaborate on this, because, at least in this work, we are not interested in the efficiency of algorithms that find the p_m^* for demanded rates R_m, rather we are only interested in the resulting powers to allow comparison of the different receiver structures.

Finally, an illustrative example derived from simulation and comparison with bounds derived from theory will be presented.

Example 4.3: **Power allocation for Independent Decoding for the Vector MAC** Equal rates $R_m = R$ are demanded for all $M = 20$ sources. Further, the simplified case is considered that all signal sources have the same steering vectors, i.e.

$$\mathbf{a}_m^{\mathrm{H}} \mathbf{a}_i = L, \qquad m, i \in \mathcal{M}. \tag{4.41}$$

Note that this basically resembles the example scenario treated in Sec. 4.3.4 for the classical MAC, and, therefore, a rate limit at $R = 0.074$ exists.

Further, no fading will be considered. Because of the requirement of equal rates, all receive signal powers after beamforming p_m/μ_m must be equal for all m, i.e. $p_m/\mu_m \triangleq p$ is required [Mec00a]. Then (4.36) simplifies to

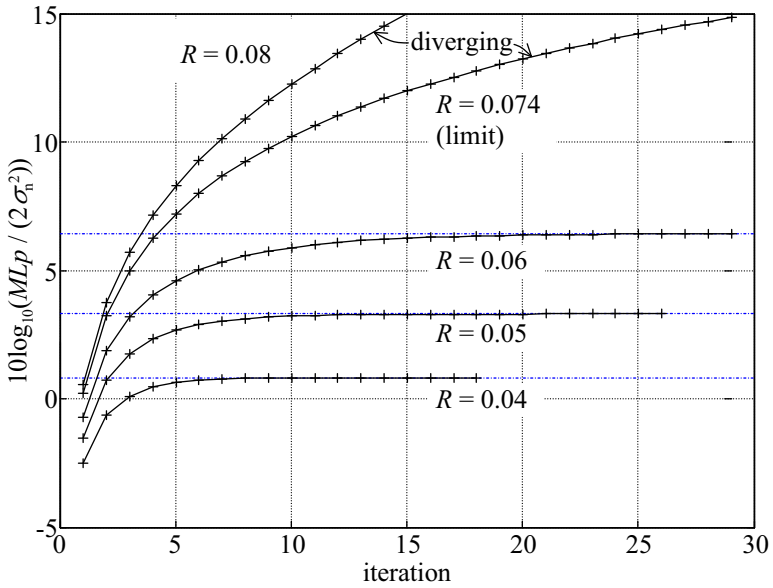
$$R_m = R = \log_2\left(1 + \frac{Lp}{2\sigma_{\mathrm{n}}^2 + (M-1)Lp}\right), \qquad m \in \mathcal{M}. \tag{4.42}$$

Fig. 4.10 shows the evolution of the source powers for $M = 20$ in the course of the iterative power allocation procedure described above.

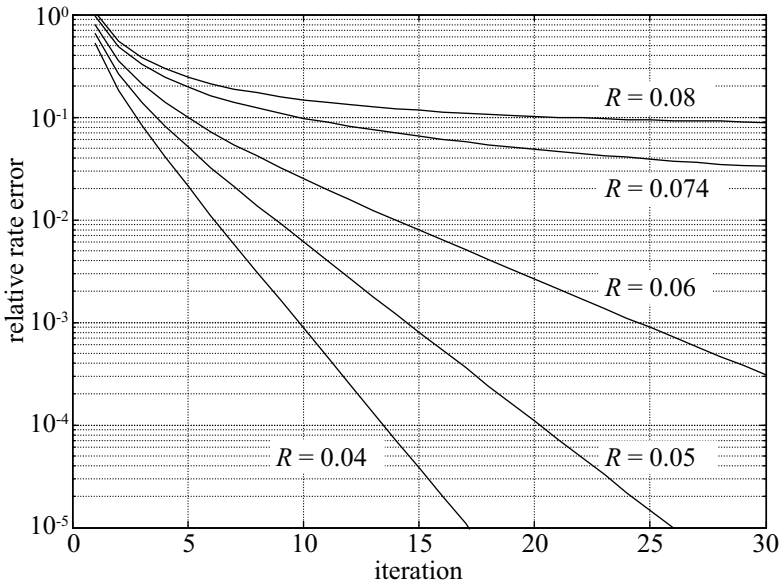
■

4.5 Joint Decoding for the Fading Vector Multiple-Access Channel

As for the classical MAC, we will firstly derive the capacity and power region, respectively. With these definitions at hand, it is possible to undertake a detailed analysis of the optimal resource



(a) Normalised sum power $MLp/(2\sigma_{\mathrm{n}}^2)$ (logarithmic) vs. iterations. The limiting rate $R = 0.074$ is obtained for $p \to \infty$, where for rates above this limit the powers tend to infinity and the power control algorithm cannot converge.



(b) Relative rate error vs. iterations. Note that convergence rate slows down as rate R approaches the limit $R = 0.074$.

Figure 4.10: Evolution of relative rate error and sum power vs. iterations. The $M = 20$ sources have equal steering vectors (no fading). The initial power vector is $\mathbf{p} = \mathbf{0}$.

allocation problem for the fading vector MAC in terms of linear programming and polymatroids, and further convex programming.

To put the investigations presented in the remainder of this chapter into the context of available publications and the investigations therein, we will shortly review the literature dealing either with resource allocation in the fading MAC or with the fading vector MAC.

Suard, Xu et al. derived the region of achievable rates for independent and optimal joint decoding

for the AWGN vector MAC [SXLK98]. There, the simple case of AWGN without fading is considered and it is pointed out that optimal beamforming does not change achievable rates. Further it is stated that independent decoding can in fact achieve the rates of optimal joint decoding if the steering vectors of the sources are orthogonal. However, the problem of resource allocation is not addressed.

Rapajic derived the capacity considering the steering vectors as random variables, neglecting the influence of a channel allocation entity that would avoid to allocate the same physical channel to users having similar steering vectors [Rap99]. Also restrictions on the range of possible steering vectors as a result from a limited service area was not taken into account. Therefore, it seems inevitable, if realistic constraints on the steering vectors should be introduced, particular satellite scenarios have to be considered, mainly regarding the assumed distribution of co-channel users.

On the other hand, the standard publications on the capacity of the fading MAC under various assumptions regarding availability of ideal or partial CSI to the transmitters do not consider the case of array reception.

Goldsmith and Varaiya treated the single-source fading scenario assuming ideal CSI being available at transmitter and receiver (the resulting optimal power allocation strategy is water-filling[8] in time.) [GV93].

Knopp and Humblet treated the classical fading MAC with ideal CSI being available at transmitters and receiver, and symmetric rate and power constraints (leading to water-filling in time with a TDMA-flavour) [KH95].

Tse and Hanly characterised the region of achievable rates for the classical fading MAC providing solutions to most relevant rate and power allocation problems [TH98a, TH98b] (a review of this was presented in the previous section).

It will be shown in this chapter that array reception adds some considerably new aspects to the investigations presented in the said publications, mainly with respect to the problem of power allocation for the fading vector MAC.

4.5.1 Capacity Region and Power Region

A characterisation of the resource allocation problem in the frame of linear and convex programming will be presented and we will turn to the question whether there is again a simple solution to the power allocation problem for successive decoding, as there is one for the classical (fading) MAC (cf. Sec. 4.3.3).

Anticipatory, it shall be said that a solution to this power allocation problem is unfortunately significantly harder to obtain from the computational point of view than it is for the classical MAC as demonstrated in Sec. 4.3.3.

The starting point for the definition of the capacity and power region of the fading vector MAC is the equation for the mutual information between source signals from a set $\mathcal{S} \subseteq \mathcal{M}$ and the receive signal vector $\boldsymbol{R}$, while source signals from $\bar{\mathcal{S}}$ (the complement of $\mathcal{S}$ with respect to $\mathcal{M}$) are considered known.

[8] This power allocation strategy is also referred to as water-pouring

The mutual information is provided by [SXLK98, CT91]

$$\begin{aligned}\mathcal{I}(S_m, m\in\mathcal{S};\boldsymbol{R}|S_j, j\in\bar{\mathcal{S}}) =& \mathcal{H}(\boldsymbol{R}|S_j, j\in\bar{\mathcal{S}}) - \mathcal{H}(\boldsymbol{R}|\boldsymbol{S})\\ =& \mathcal{H}(\boldsymbol{R}|S_j, j\in\bar{\mathcal{S}}) - \mathcal{H}(\boldsymbol{N}),\\ & \mathcal{S}\subseteq\mathcal{M}, \bar{\mathcal{S}} = \mathcal{M}\setminus\mathcal{S}.\end{aligned} \tag{4.43}$$

Then, for given receive power vector $\mathbf{P} = (P_1, P_2, \ldots, P_M)$ and for fixed joint fading state $\mathbf{h}$, the inequalities defining the region of achievable rates for the fading MAC are [SXLK98] (this can be readily verified using the definition of conditional mutual information (B.9) and the entropy of a proper complex Gaussian random variable with given covariance matrix (B.11))

$$\begin{aligned}\sum_{m\in\mathcal{S}} R_m(h_m) \quad\leq\quad & \max \mathcal{I}(S_m, m\in\mathcal{S};\boldsymbol{R}|S_j, j\in\bar{\mathcal{S}})\\ & = \log_2\left((\pi\mathrm{e})^L \det\left(\mathbf{K}_\mathbf{r}(\mathcal{S})\right)\right) - \log_2\left((\pi\mathrm{e})^L\det\left(\mathbf{K}_\mathbf{n}\right)\right)\\ & = \log_2\left(\det\left(\mathbf{I}_L + \mathbf{K}_\mathbf{n}^{-1}\sum_{m\in\mathcal{S}}\mathbf{a}_m\mathbf{a}_m^{\mathrm{H}} P_m\left|h_m\right|^2\right)\right),\\ & \forall\mathcal{S}\subseteq\mathcal{M}, \bar{\mathcal{S}} = \mathcal{M}\setminus\mathcal{S}.\end{aligned} \tag{4.44}$$

where we further have used as before the definition of the mean received powers $P_m = p_m/\mu_m$ to simplify notation.

Capacity Region

With the definition in (4.44) and allowing the h_m to be random variables (we will again adopt capital italics H_m to indicate this) known to the receiver only, the ergodic capacity region is provided by (cf. Sec. 4.3.2)

$$\boxed{\mathcal{R}_{\mathrm{VMAC}} = \left\{\mathbf{R}\in\mathbb{R}_+^M : \sum_{m\in\mathcal{S}} R_m \leq \mathrm{E}\left\{\log_2\left(\det\left(\mathbf{I}_L + \mathbf{K}_\mathbf{n}^{-1}\sum_{m\in\mathcal{S}}\mathbf{a}_m\mathbf{a}_m^{\mathrm{H}}\left|H_m\right|^2 P_m\right)\right)\right\}_{\boldsymbol{H}}, \forall\mathcal{S}\subseteq\mathcal{M}\right\}.} \tag{4.45}$$

Power Region

Further, as power region for given $\mathbf{R}$ and for the ergodic capacity case, we obtain the implicit definition

$$\boxed{\mathcal{P}_{\mathrm{VMAC}} = \left\{\mathbf{P}\in\mathbb{R}_+^M : \sum_{m\in\mathcal{S}} R_m \leq \mathrm{E}\left\{\log_2\left(\det\left(\mathbf{I}_L + \mathbf{K}_\mathbf{n}^{-1}\sum_{m\in\mathcal{S}}\mathbf{a}_m\mathbf{a}_m^{\mathrm{H}}\left|H_m\right|^2 P_m\right)\right)\right\}_{\boldsymbol{H}}, \forall\mathcal{S}\subseteq\mathcal{M}\right\}.} \tag{4.46}$$

Every power tuple $\mathbf{P} = (P_1, P_2, \ldots, P_M)\in\mathcal{P}_{\mathrm{VMAC}}$ corresponds to a rate region $\mathcal{R}_{\mathrm{VMAC}}$ containing a demanded rate tuple $\mathbf{R}$.

As for the classical MAC, the capacity region and the power region given by (4.45) and (4.46) are both defined by $(2^M - 1)$ constraints (cf. the according investigations for the classical MAC presented in Sec. 4.3.2).

Next we have to discuss the properties of the capacity region $\mathcal{R}_{\mathrm{VMAC}}$ and power region $\mathcal{P}_{\mathrm{VMAC}}$, respectively, in the context of resource allocation, i.e. the optimisation problems of maximising the sum-rate for given powers $\mathbf{P}$ and minimising the sum-power for given target rates $\mathbf{R}$.

Therefore we now turn to a characterisation of $\mathcal{R}_{\mathrm{VMAC}}$ and $\mathcal{P}_{\mathrm{VMAC}}$ to find out whether a possible polymatroidal structure can be exploited to efficiently solve the resource allocation problems (4.47) and (4.54).

4.5.2 Optimal Resource Allocation: Rate and Power Allocation

Optimal Rate Allocation

In accordance to the optimisation problem formulated for the classical MAC in (4.22), here maximisation for given power constraints $\mathbf{p}$ according to

$$\max_{\mathbf{R}} \{\boldsymbol{\lambda}_{\mathrm{R}} \cdot \mathbf{R} : \mathbf{R} \in \mathcal{R}_{\mathrm{VMAC}}\} \tag{4.47}$$

is sought.

In Sec. 4.3.2 it was pointed out that any capacity region regardless of the underlying channel must be a polymatroid, because the chain rule of mutual information directly provides the necessary and sufficient property of the vertices of a polymatroid given by (4.9).

Therefore, $\mathcal{R}_{\mathrm{VMAC}}$ must be a polymatroid and the optimisation problem (4.47) can be easily solved using this property.

Then, analogous to the classical MAC (cf. (4.23)), the rate allocation that maximises (4.47) for given power constraints is given by the rates R_m at vertex v according to (4.9), i.e.

$$\begin{aligned} R_{\pi_v(m)} \;=\; & \mathrm{E}\left\{\log_2\left(\det\left(\mathbf{K_n} + \sum_{i=1}^{m} \mathbf{a}_{\pi_v(i)}\mathbf{a}_{\pi_v(i)}^{\mathrm{H}} \left|H_{\pi_v(i)}\right|^2 P_{\pi_v(i)}\right)\right)\right\}_{\boldsymbol{H}} - \\ & \mathrm{E}\left\{\log_2\left(\det\left(\mathbf{K_n} + \sum_{i=1}^{m-1} \mathbf{a}_{\pi_v(i)}\mathbf{a}_{\pi_v(i)}^{\mathrm{H}} \left|H_{\pi_v(i)}\right|^2 P_{\pi_v(i)}\right)\right)\right\}_{\boldsymbol{H}}, \\ & m = 1, 2, \ldots, M, \end{aligned} \tag{4.48}$$

where π_v ($v = 1, 2, \ldots, M!$) denotes again the $M!$ possible permutations of $\mathcal{M}$.

The optimal permutation $\boldsymbol{\pi}_{v^*}$, which solves the posed optimisation problem (4.47), is given by (4.24):

$$\lambda_{\mathrm{R},\pi_{v^*}(1)} \geq \lambda_{\mathrm{R},\pi_{v^*}(2)} \geq \cdots \geq \lambda_{\mathrm{R},\pi_{v^*}(M)},$$

with $\boldsymbol{\lambda}_{\mathrm{R}} = (\lambda_{\mathrm{R},1}, \lambda_{\mathrm{R},2}, \cdots, \lambda_{\mathrm{R},M})$.

In the following, an example for 2 sources will demonstrate the dependency of the capacity region $\mathcal{R}_{\mathrm{VMAC}}$ on the steering vectors.

Example 4.4: **Capacity Region for 2 Sources for the VMAC** For simplicity it will be assumed in the following that $|h_m|^2/\mu_m = 1, \forall m$, and further that the L noise sources are uncorrelated (spatially white noise, cf. Sec. 3.5.3) such that according to (3.75) it holds

$$\mathbf{K_n} = 2\sigma_{\mathrm{n}}^2 \mathbf{I}_L,$$

where $2\sigma_{\mathrm{n}}^2$ is the power of the Gaussian noise.

For the case of 2 transmitters, i.e. $\mathcal{M} = \{1, 2\}$, (4.44) provides $\left(2^2 - 1 = 3\right)$ inequalities, which are

$$R_1 \leq \log_2\left(\det\left(\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\frac{p_1}{2\sigma_{\mathrm{n}}^2} + \mathbf{I}_L\right)\right) = \log_2\left(1 + \frac{Lp_1}{2\sigma_{\mathrm{n}}^2}\right) \tag{4.49}$$

$$R_2 \leq \log_2\left(\det\left(\mathbf{a}_2\mathbf{a}_2^{\mathrm{H}}\frac{p_2}{2\sigma_{\mathrm{n}}^2} + \mathbf{I}_L\right)\right) = \log_2\left(1 + \frac{Lp_2}{2\sigma_{\mathrm{n}}^2}\right) \tag{4.50}$$

$$\begin{aligned} R_1 + R_2 &\leq \log_2\left(\det\left(\mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}\frac{p_1}{2\sigma_{\mathrm{n}}^2} + \mathbf{a}_2\mathbf{a}_2^{\mathrm{H}}\frac{p_2}{2\sigma_{\mathrm{n}}^2} + \mathbf{I}_L\right)\right) = \\ &= \log_2\left(\left(1 + \frac{Lp_1}{2\sigma_{\mathrm{n}}^2}\right)\left(1 + \frac{Lp_2}{2\sigma_{\mathrm{n}}^2}\right) - \left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2 \frac{p_1p_2}{4\sigma_{\mathrm{n}}^4}\right), \end{aligned} \tag{4.51}$$

where we have used that $\det\left(\mathbf{I}_L + \mathbf{a}_m\mathbf{a}_m^{\mathrm{H}}p_m/(2\sigma_{\mathrm{n}}^2)\right) = \det\left(1 + \mathbf{a}_m^{\mathrm{H}}\mathbf{a}_m p_m/(2\sigma_{\mathrm{n}}^2)\right)$ (according to (A.2)) and $\mathbf{a}_m^{\mathrm{H}}\mathbf{a}_m = L$ being the array gain.

The term $\left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2$ in (4.51) indicates that the achievable sum-rate strongly depends on the relative user positions as the steering vectors $\mathbf{a}_m$ only depend on the array geometry and the incident angles of the source signals on the satellite antenna array.

If $\mathbf{a}_1$ and $\mathbf{a}_2$ are orthogonal, i.e. the two signals can be perfectly spatially separated, then the sum-rate becomes maximal according to

$$\begin{aligned} R_1 + R_2 &\leq \log_2\left(\left(1 + \frac{Lp_1}{2\sigma_{\mathrm{n}}^2}\right)\left(1 + \frac{Lp_2}{2\sigma_{\mathrm{n}}^2}\right)\right) = \\ &= \log_2\left(1 + \frac{Lp_1}{2\sigma_{\mathrm{n}}^2}\right) + \log_2\left(1 + \frac{Lp_2}{2\sigma_{\mathrm{n}}^2}\right), \qquad \text{for } \mathbf{a}_1 \perp \mathbf{a}_2, \end{aligned} \tag{4.52}$$

which is, as expected, simply the sum of rates of two independent sources. On the other hand, if for the worst-case $\mathbf{a}_1 = \mathbf{a}_2$ (which provides now the classical Gaussian MAC) the sum-rate is limited to that of a single source transmitting with a power $(p_1 + p_2)$:

$$R_1 + R_2 \leq \log_2\left(1 + \frac{L(p_1 + p_2)}{2\sigma_{\mathrm{n}}^2}\right), \qquad \text{for } \mathbf{a}_1 = \mathbf{a}_2. \tag{4.53}$$

To illustrate these results, Fig. 4.11 shows the capacity region $\mathcal{R}_{\mathrm{VMAC}}$ for 2 sources for $p_1/(2\sigma_{\mathrm{n}}^2) = p_2/(2\sigma_{\mathrm{n}}^2)$ (rates are normalised with respect to the maximum) and for various values of $\left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2$ (cf. [SXLK98]). The capacity region has $2!$ vertices in the positive orthant and the corresponding points $(R_1^*, 1)$ and $(1, R_2^*)$ are achievable via successive decoding and coincide for $\mathbf{a}_1 \perp \mathbf{a}_2$; in this

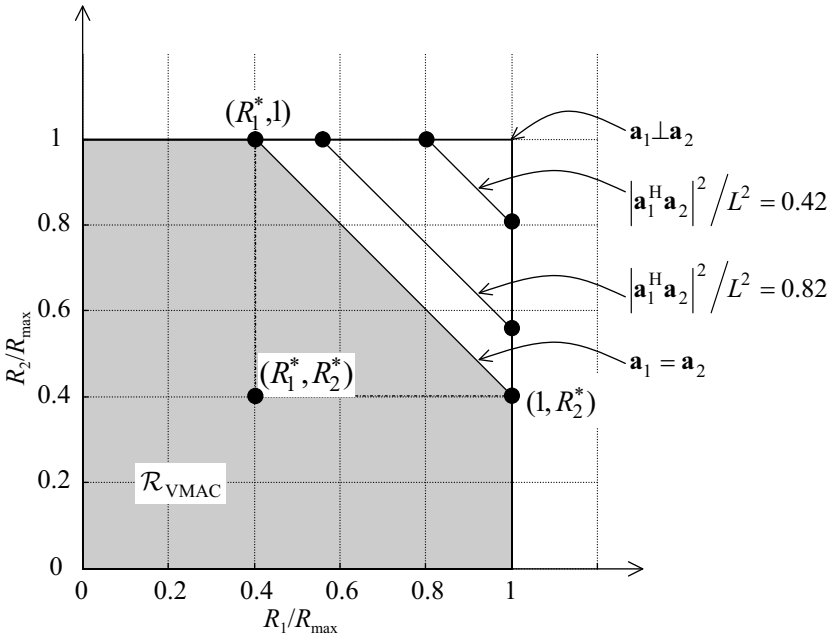


Figure 4.11: Capacity region $\mathcal{R}_{\mathrm{VMAC}}$ for two sources for $Lp_1/(2\sigma_{\mathrm{n}}^2) = Lp_2/(2\sigma_{\mathrm{n}}^2) = 1$, maximal rate $R_{\max} = 1$. The capacity regions for various values of $\left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2$ are pentagons with vertices $(0,0), (1,0), (1,R_2^*), (R_1^*,1), (0,1)$. For $\mathbf{a}_1 \perp \mathbf{a}_2$ it holds for the normalised rates $R_1^* = R_2^* = 1$.

case successive decoding is cannot perform better than independent decoding as interference is already removed completely by spatial filtering.

Obviously, according to (4.36), the points (R_1^*, R_2^*) (again, cf. Fig. 4.11) are those achievable when independent decoding is employed without successive decoding.

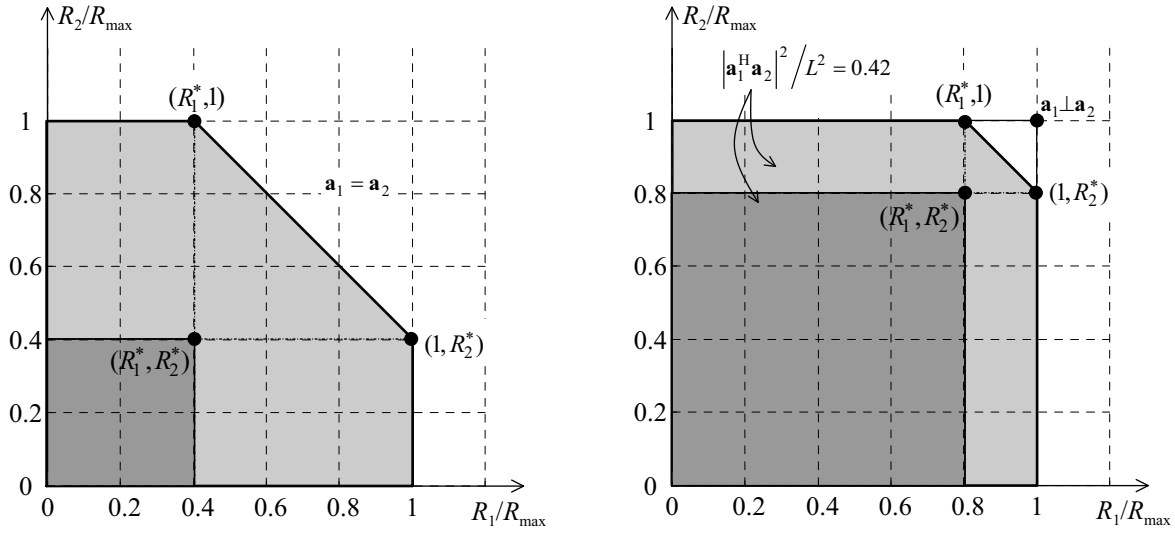
In Fig. 4.12, the capacity regions for joint decoding are compared with the regions of achievable rates for independent decoding, from which it becomes obvious that performance of independent decoding approaches that of optimal joint decoding as $\mathbf{a}_1$ and $\mathbf{a}_2$ become increasingly orthogonal.

Furthermore, as stated above there are $2!$ permutations on $\mathcal{M}$, namely, $\{1,2\}$ and $\{2,1\}$. According to the corresponding decoding orders, the rate tuple $(R_1^*, 1)$ is achievable by firstly decoding source 1, subsequently subtracting the re-encoded and re-modulated signal from the receive signal vector, and, finally, decoding source 2, which is only impaired by thermal noise. Analogously, rate tuple $(1, R_2^*)$ is achieved (cf. Sec. 4.6).

■

It was mentioned before that there is a symmetry in the classical MAC with respect to the amount of interference which is added by a particular source, i.e. the interference power introduced by a source i is the same for all other users $m \in \mathcal{M} \setminus i$. This holds also for array reception for 2 sources, as the factor $\left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2$ is symmetric with respect to the source indices.

A different situation is encountered when we are dealing with $M > 2$ sources. This will be illustrated for $M = 3$ sources in the following.



(a) $\mathbf{a}_1 = \mathbf{a}_2$ **(b)** $\left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2/L^2 = 0.42$, and $\mathbf{a}_1 \perp \mathbf{a}_2$

Figure 4.12: Comparison of capacity region (without fading) for joint decoding (light gray area) and region of achievable rates for independent decoding (dark gray area) for 2 sources with equal power constraint. As $\mathbf{a}_1$ and $\mathbf{a}_2$ become increasingly orthogonal independent decoding approaches optimal joint decoding.

Example 4.5: **Capacity region for 3 users for the VMAC** Considering the positive quadrant again, the capacity region is bounded by $(2^3 - 1 = 7)$ faces and there are $3! = 6$ vertices.

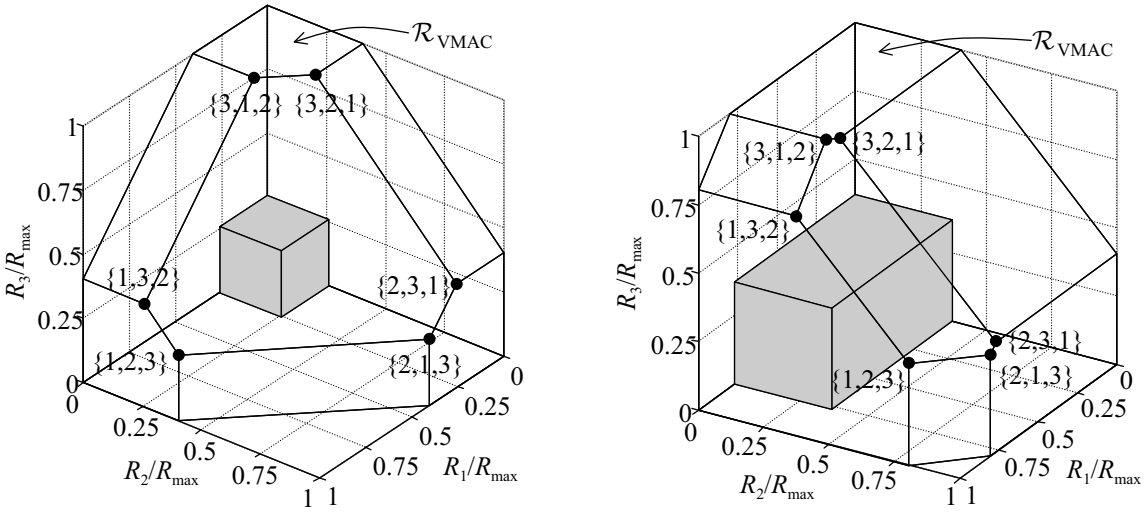
In Fig. 4.13 the capacity regions for joint decoding and the regions of achievable rates for independent decoding are shown for the two cases that $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3$, and $\mathbf{a}_2 = \mathbf{a}_3$, $\left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2/L^2 = \left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_3\right|^2/L^2 = 0.42$. All sources are subject to the same power constraint, i.e. $p_1 = p_2 = p_3$.

Clearly, although using the same power, with independent decoding source 1 can achieve a higher rate than source 2 and 3, because of the difference in the steering vectors and the resulting spatial separability. Further, the rates of source 2 and 3 are also slightly higher than for the classical MAC, because interference from source 1 is reduced.

Also with joint decoding a rate gain due to spatial separability that comes with array reception is obvious.

It was mentioned above that for the classical MAC the important equal rate point (here: $R_1 = R_2 = R_3$) is achieved with equal powers (cf. Fig. 4.13(a)), whereas for the general case of unequal steering vectors this point requires unequal power allocations to minimise transmit powers (cf. Fig. 4.13(b)).

■



(a) $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3$ **(b)** $\left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2/L^2 = \left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_3\right|^2/L^2 = 0.42$, $\mathbf{a}_2 = \mathbf{a}_3$

Figure 4.13: Comparison of capacity region (normalised, no fading) for joint decoding (wireframe polyhedron) with region of achievable rates for independent decoding (grey cuboid) for 3 sources with equal power constraint. Also shown is next to each vertex the corresponding permutation π_v of $\mathcal{M}$ indicating the power allocation and (reverse) decoding order required to achieve the rates at the vertex. Note that for successive decoding the rate of the source being decoded first is limited to the rate achievable with independent decoding as indicated (grey cuboid).

Optimal Power Allocation

Optimal power allocation is for the vector MAC defined by the optimisation problem (absolutely analogous to the classical MAC, (4.25))

$$\min_{\mathbf{P}} \left\{ \boldsymbol{\lambda}_{\mathrm{P}} \cdot \mathbf{P} : \mathbf{P} \in \mathcal{P}_{\text{VMAC}} \right\}. \tag{4.54}$$

What is of interest for optimal power allocation, as formulated by (4.54), is the characterisation of the power region $\mathcal{P}_{\text{VMAC}}$.

The following extends results for the classical fading MAC presented in [TH98a] and [Mec00a] (a review was presented in Sec. 4.3.2), respectively, to the fading vector MAC.

Firstly, it shall be proved that $\mathcal{P}_{\text{VMAC}}$ is a convex set also for the vector MAC. Convexity of the feasible set $\mathcal{P}_{\text{VMAC}}$ is of paramount importance in order to guarantee convergence of an optimisation algorithm to the global optimum, solving the power allocation problem (4.54).

Proof. To show convexity of $\mathcal{P}_{\text{VMAC}}$ it is used that $\log\det(\mathbf{I}+\mathbf{S})$ is a concave function over the set of positive semidefinite (also: nonnegative definite) matrices $\mathbf{S}$. This follows from strict concavity of $\log\det(\mathbf{D})$ over the set of positive definite matrices $\mathbf{D}$ (for a proof see, e.g., [CT91, HJ99]).

With concavity of $\log\det(\mathbf{I}+\mathbf{S})$ it follows analogously to concave scalar functions that

$$\log\det(\mathbf{I}+\lambda\mathbf{S}_1+(1-\lambda)\mathbf{S}_2) \geq \lambda\log\det(\mathbf{I}+\mathbf{S}_1)+(1-\lambda)\log\det(\mathbf{I}+\mathbf{S}_2), \tag{4.55}$$

for $0 \leq \lambda \leq 1$ and for any positive semidefinite matrices $\mathbf{S}_1, \mathbf{S}_2$.

Further, because $\mathbf{K}_{\mathbf{n}}^{-1}$ is hermitian and

$$\sum_{m \in \mathcal{S}} \mathbf{a}_m \mathbf{a}_m^{\mathrm{H}} \left|H_m\right|^2 P_m$$

is positive semidefinite, it must hold that the term

$$\mathbf{K}_{\mathbf{n}}^{-1} \sum_{m \in \mathcal{S}} \mathbf{a}_m \mathbf{a}_m^{\mathrm{H}} \left|H_m\right|^2 P_m$$

in (4.44) is positive semidefinite [HJ99], which we need in the following in connection with (4.55).

Given two power allocations $\mathbf{P}^{(1)}, \mathbf{P}^{(2)} \in \mathcal{P}_{\mathrm{VMAC}}$ it follows from (4.55) that it must hold for $0 \leq \lambda \leq 1$ and $\mathcal{S} \subseteq \mathcal{M}$

$$\begin{aligned}
&\mathrm{E}\left\{\log_2\left(\det\left(\mathbf{I}_L + \mathbf{K}_{\mathbf{n}}^{-1} \sum_{m \in \mathcal{S}} \mathbf{a}_m \mathbf{a}_m^{\mathrm{H}} \left|H_m\right|^2 \left(\lambda P_m^{(1)} + (1-\lambda) P_m^{(2)}\right)\right)\right)\right\}_{\boldsymbol{H}} \\
&\quad \overset{(a)}{\geq} \lambda \mathrm{E}\left\{\log_2\left(\det\left(\mathbf{I}_L + \mathbf{K}_{\mathbf{n}}^{-1} \sum_{m \in \mathcal{S}} \mathbf{a}_m \mathbf{a}_m^{\mathrm{H}} \left|H_m\right|^2 P_m^{(1)}\right)\right)\right\}_{\boldsymbol{H}} + \\
&\qquad (1-\lambda) \mathrm{E}\left\{\log_2\left(\det\left(\mathbf{I}_L + \mathbf{K}_{\mathbf{n}}^{-1} \sum_{m \in \mathcal{S}} \mathbf{a}_m \mathbf{a}_m^{\mathrm{H}} \left|H_m\right|^2 P_m^{(2)}\right)\right)\right\}_{\boldsymbol{H}} \\
&\quad \overset{(b)}{\geq} \lambda \sum_{m \in \mathcal{S}} R_m + (1-\lambda) \sum_{m \in \mathcal{S}} R_m = \sum_{m \in \mathcal{S}} R_m,
\end{aligned} \tag{4.56}$$

where (a) follows from (4.55) (concavity) and (b) from $\mathbf{P}^{(1)}, \mathbf{P}^{(2)} \in \mathcal{P}_{\mathrm{VMAC}}$. Note that equality in (a) is only given if and only if $\lambda = 0$, $\lambda = 1$, or $\mathbf{P}^{(1)} = \mathbf{P}^{(2)}$.

Rewriting (4.56) without the intermediate step in (a) yields

$$\sum_{m \in \mathcal{S}} R_m \leq \mathrm{E}\left\{\log_2\left(\det\left(\mathbf{I}_L + \mathbf{K}_{\mathbf{n}}^{-1} \sum_{m \in \mathcal{S}} \mathbf{a}_m \mathbf{a}_m^{\mathrm{H}} \left|H_m\right|^2 \left(\lambda P_m^{(1)} + (1-\lambda) P_m^{(2)}\right)\right)\right)\right\}_{\boldsymbol{H}}, \tag{4.57}$$

which simply means that the convex combination $\left(\lambda \mathbf{P}^{(1)} + (1-\lambda)\mathbf{P}^{(2)}\right)$ must lie in $\mathcal{P}_{\mathrm{VMAC}}$. This proves convexity of $\mathcal{P}_{\mathrm{VMAC}}$. □

Knowing now that the power region for the vector MAC $\mathcal{P}_{\mathrm{VMAC}}$ is convex just as for the classical MAC, we turn again towards the question under what conditions $\mathcal{P}_{\mathrm{VMAC}}$ may be a contra-polymatroid.

Is $\mathcal{P}_{\mathrm{VMAC}}$ a Contra-Polymatroid? Recall that every face of a convex polytope is again a convex polytope [PS98]. Therefore, it suffices to proof that at least one face is not convex to verify that the feasible set is no polytope and hence cannot be a polymatroid [Mec00a].

To test whether $\mathcal{P}_{\mathrm{VMAC}}$ is a polymatroid, we test the dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{VMAC}}}$ of $\mathcal{P}_{\mathrm{VMAC}}$ for convexity. The dominant face $\mathcal{D}_{\mathcal{P}_{\mathrm{VMAC}}}$ of $\mathcal{P}_{\mathrm{VMAC}}$ is the set of power tuples $\mathbf{P}$ for that no component can be reduced without increasing other powers to stay in $\mathcal{P}_{\mathrm{VMAC}}$ (cf. Sec. 4.3.2) [TH98a].

We will see that $\mathcal{D}_{\mathcal{P}_{\mathrm{VMAC}}}$ is not convex and, therefore, $\mathcal{P}_{\mathrm{VMAC}}$ cannot be a polymatroid.

Proof. For this we choose two power allocations $\mathbf{P}^{(1)}, \mathbf{P}^{(2)} \in \mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ in (4.56) and $\mathcal{S} = \mathcal{M}$. Under these assumption the inequality in (b) in (4.56) becomes a strict equality. If the convex combination $\left(\lambda \mathbf{P}^{(1)} + (1-\lambda)\mathbf{P}^{(2)}\right)$ has to be inside $\mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$, then equality is also required in (a) in (4.56) for all λ, but this is only given if and only if $\lambda = 0$, $\lambda = 1$, or $\mathbf{P}^{(1)} = \mathbf{P}^{(2)}$ as was already stated before. Therefore the line connecting any $\mathbf{P}^{(1)}, \mathbf{P}^{(2)} \in \mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ is not in $\mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ except the endpoints, and hence $\mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ is concave.

For the AWGN case it is assumed that $|H_m|^2 = 1, \forall m$. Then equality in (a) in (4.56) would also be achieved if $\sum_{m\in\mathcal{M}} \mathbf{a}_m \mathbf{a}_m^{\text{H}} P_m^{(1)} = \sum_{m\in\mathcal{M}} \mathbf{a}_m \mathbf{a}_m^{\text{H}} P_m^{(2)}$, but this requires $\mathbf{P}^{(1)} = \mathbf{P}^{(2)}$ or that all $\mathbf{a}_m$ are equal, again leading to the classical MAC where $\sum_{m\in\mathcal{M}} P_m^{(1)} = \sum_{m\in\mathcal{M}} P_m^{(2)}$ guarantees convexity of the dominant face. Therefore, the dominant face $\mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ is concave also for the non-fading case.

By proving that the dominant face of $\mathcal{P}_{\text{VMAC}}$ is a concave set it follows that, in general, the power region cannot not be a polytope.

Hence, we conclude with the statement that $\mathcal{P}_{\text{VMAC}}$ is **not** a contra-polymatroid, unless no fading is present and all steering vectors are equal (resulting in the classical AWGN MAC), or orthogonal (resulting in M mutually independent channels, where the dominant face becomes a point). □

Finally, a 2-source example shall illustrate the dependency of the power region $\mathcal{P}_{\text{VMAC}}$ on the steering vectors.

Example 4.6: **Power Region $\mathcal{P}_{\text{VMAC}}$ for 2 Sources** Above it was proven that the power region $\mathcal{P}_{\text{VMAC}}$ cannot be a contra-polymatroid, because the dominant face $\mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ is concave, even for the non-fading case (unless the steering vector of all sources are mutually orthogonal). This is illustrated in Fig. 4.14 for 2 sources and without fading. There $\mathcal{P}_{\text{VMAC}}$ is shown for different values of $|\mathbf{a}_1^{\text{H}}\mathbf{a}_2|^2/L^2$ according to varying spatial separation of the 2 source signals. It can be seen clearly that the dominant face $\mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ is concave for $0 < |\mathbf{a}_1^{\text{H}}\mathbf{a}_2|^2/L^2 < 1$ (for better visibility of concavity the graph obtained from (4.46) for $\mathcal{S} = \mathcal{M}$ defining $\mathcal{D}_{\mathcal{P}_{\text{VMAC}}}$ is drawn also outside the relevant range).

■

4.6 Successive Decoding for the Fading Vector Multiple-Access Channel

We have shown in the last section that the power region of the vector MAC is **not** a contra-polymatroid, and there are two consequences of this. Firstly, the optimisation problem of minimising sum-transmit powers for given target rates (4.54) cannot be easily solved exploiting the property of the vertices of a polymatroid. Secondly, because the optimal solution may not be obtained at a vertex of the capacity region, successive decoding is in general not leading to the minimal sum transmit power required to achieve a given rate tuple.

However, just as discussed for the classical MAC in Sec. 4.3.3, successive decoding is still attractive, because its ability to achieve maximal sum-rates with reduced complexity.

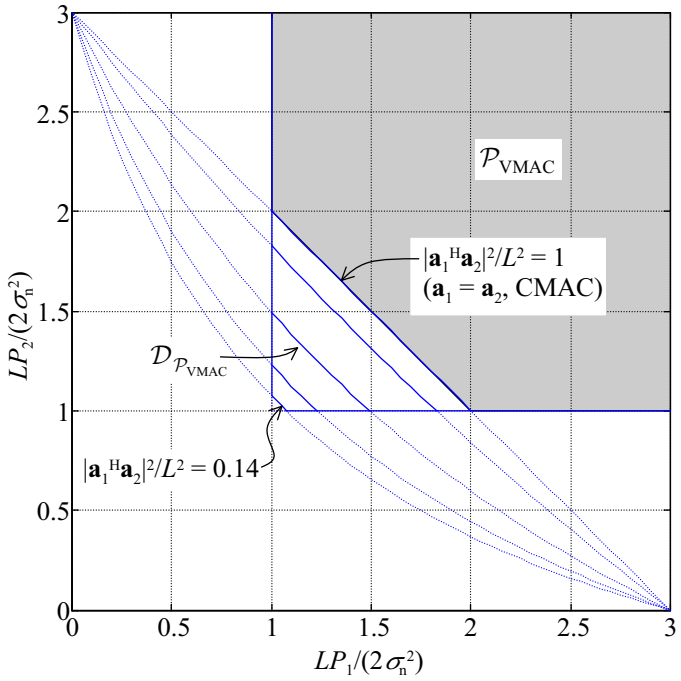


Figure 4.14: Power region $\mathcal{P}_{\text{VMAC}}$ for 2 sources for various values of $|\mathbf{a}_1^{\text{H}}\mathbf{a}_2|^2/L^2$ (AWGN, $R_1^* = R_2^* = 1$).

Therefore, we want to restrict to successive decoding for the vector MAC due to its attractiveness with respect to performance and decoder complexity.

In particular we have to clarify whether there is a simple optimal power allocation procedure for successive decoding as there is for the classical fading MAC as pointed out in Sec. 4.3.3.

The rates that are achievable by successive decoding at a vertex v of the capacity region are given by (4.48). Rewriting this equation, we obtain

$$\begin{aligned} R_{\pi_v(m)} =& \mathrm{E}\left\{ \log_2 \left(\frac{\det\left(\mathbf{K_n} + \sum_{i=1}^{m} \mathbf{a}_{\pi_v(i)}\mathbf{a}_{\pi_v(i)}^{\mathrm{H}} P_{\pi_v(i)} \left|H_{\pi_v(i)}\right|^2\right)}{\det\left(\mathbf{K_n} + \sum_{i=1}^{m-1} \mathbf{a}_{\pi_v(i)}\mathbf{a}_{\pi_v(i)}^{\mathrm{H}} P_{\pi_v(i)} \left|H_{\pi_v(i)}\right|^2\right)} \right) \right\}_{\boldsymbol{H}} \\ =& \mathrm{E}\left\{ \log_2 \left(1 + P_{\pi_v(m)} \left|H_{\pi_v(m)}\right|^2 \mathbf{a}_{\pi_v(m)}^{\mathrm{H}} \left(\mathbf{K_n} + \mathbf{K_v}(\mathcal{S}_{v,m-1})\right)^{-1} \mathbf{a}_{\pi_v(m)} \right) \right\}_{\boldsymbol{H}}, \\ & m = 1, 2, \ldots, M, \end{aligned} \tag{4.58}$$

where we have defined $\mathcal{S}_{v,m-1}$ in (4.27) according to

$$\mathcal{S}_{v,m-1} = \{\pi_v(1), \pi_v(2), \ldots, \pi_v(m-1)\},$$

and, further, $\mathbf{K_v}(\mathcal{S}_{v,m-1})$, as provided by the definition in (3.82), is the receive signal covariance matrix considering only source signals contained in the set $\mathcal{S}_{v,m-1}$. (Recall that for a given permutation π_v the decoding order is $(\pi_v(M), \pi_v(M-1), \ldots, \pi_v(1))$.)

Compare (4.58) with the rates achievable by optimum beamforming in conjunction with independent decoding given by (4.38), respectively. Then it is easy to see that (4.58) is the rate achieved

by optimal beamforming with subsequent independent decoding of source $\pi_v(m)$, where source signals from the set $\mathcal{S}_{v,m-1}$ are considered as interference and signals from the remaining sources, i.e. from set $\bar{\mathcal{S}}_{v,m-1} = \mathcal{M} \setminus \mathcal{S}_{v,m-1}$, are considered known and, thus, not contributing to the interference for source $\pi_v(m)$.

In the following section the description of the receiver structure achieving rates given by (4.58) is discussed.

4.6.1 Receiver Structure

To simplify notation it is assumed that the source indices are sorted according to the decoding order, i.e. source 1 is decoded first, followed by source 2, and so on, with source M being decoded last. Therefore, it holds in the following for the set of sources $\mathcal{J}_m$ interfering with source m

$$\mathcal{J}_m = \{m+1, m+2, \ldots, M\}\,. \tag{4.59}$$

The receiver structure for successive decoding for the fading vector MAC is depicted in Fig. 4.15.

The optimal beamforming weight vector for source m is according to (3.89) given by

$$\mathbf{w}_{\mathrm{opt},m} = \beta \left(\mathbf{K}_{\mathbf{n}} + \mathbf{K}_{\mathbf{v}}(\mathcal{J}_m)\right)^{-1} \mathbf{a}_m, \tag{4.60}$$

where only interference from sources $m+1, m+2, \ldots, M$ not yet decoded is considered.

The input z_m to the m-th single-user decoder is obtained by

$$z_m = \mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}} \left(\mathbf{A}\tilde{\mathbf{r}} + \mathbf{n}\right) - \sum_{i=1}^{m-1} B_{mi}\hat{r}_i = \underbrace{\mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}}\mathbf{a}_m\tilde{r}_m + \overbrace{\sum_{\substack{i=1\\ i\neq m}}^{M} \mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}}\mathbf{a}_i\tilde{r}_i}^{\text{Interference}} + \mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}}\mathbf{n}}_{\triangleq\, \tilde{z}_m} - \sum_{i=1}^{m-1} B_{mi}\hat{r}_i, \tag{4.61}$$

with

$$B_{mi} = \mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}}\mathbf{a}_i, \tag{4.62}$$

further, the elements of the array input signal vector $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_M)^{\mathrm{T}}$ were introduced in (3.54) according to

$$\tilde{r}_m(t) = \frac{h_m(t)}{\sqrt{\mu_m}} s_m(t).$$

In (4.61) $\tilde{z}_m$ denotes the output of the m-th beamformer, which contains interference from sources $1, 2, \ldots, m-1$ already decoded, as well as from sources $m+1, m+2, \ldots, M$ not yet decoded (cf. Fig. 4.16).

The terms $B_{mi}\hat{r}_i$ in (4.61) are estimates of the interference for source m caused by sources $i = 1, 2, \ldots, m-1$, i.e. caused by sources already decoded before source m. The required estimates $\hat{r}_i$ of the receive signal $\tilde{r}_i$ are provided by the i-th single-user decoder.

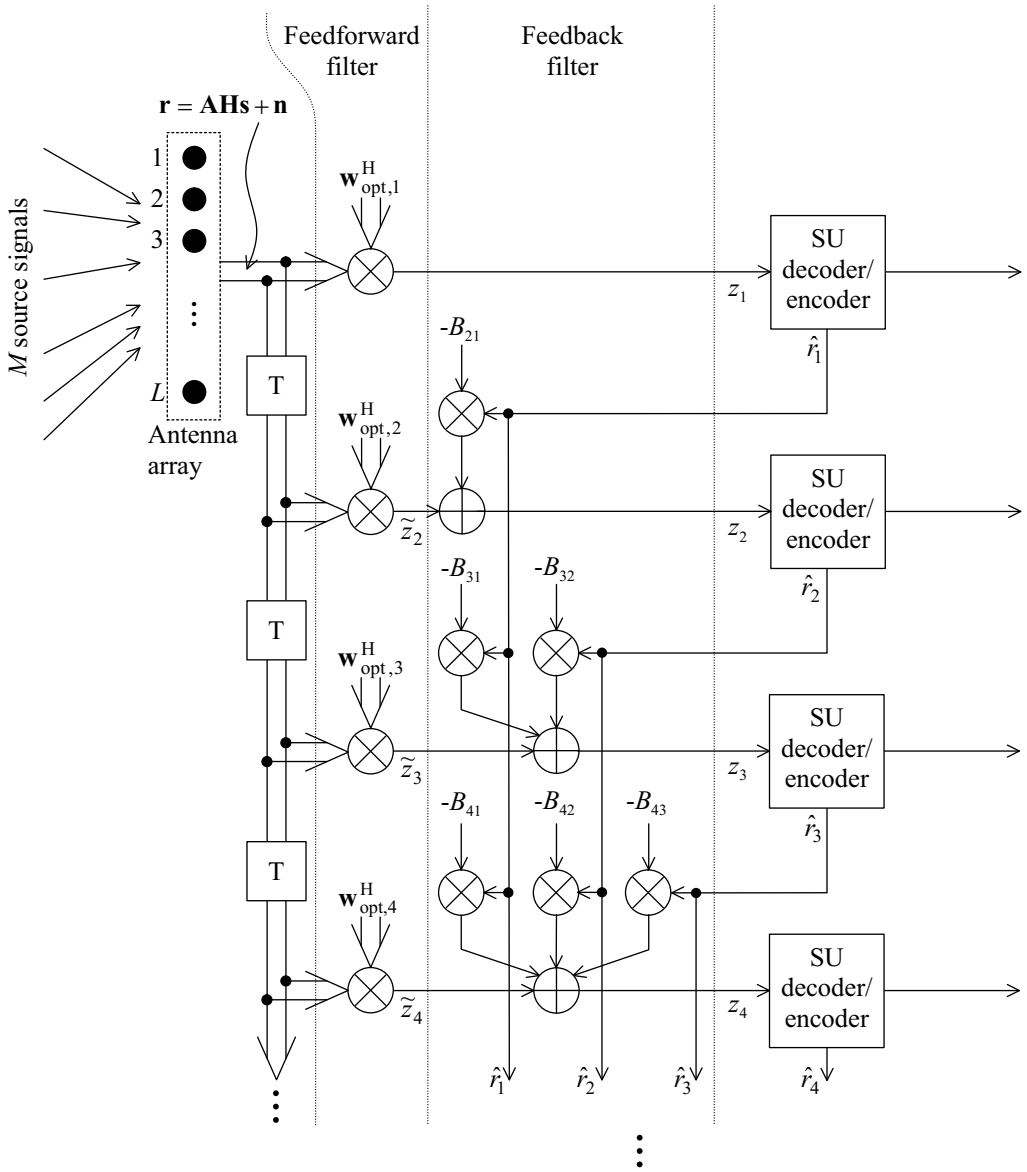


Figure 4.15: Block diagram of the receiver implementing successive decoding. Delays T are introduced to compensate the decoding delay in each decoder stage. (For simpler notation the source indices are sorted according to the decoding order.)

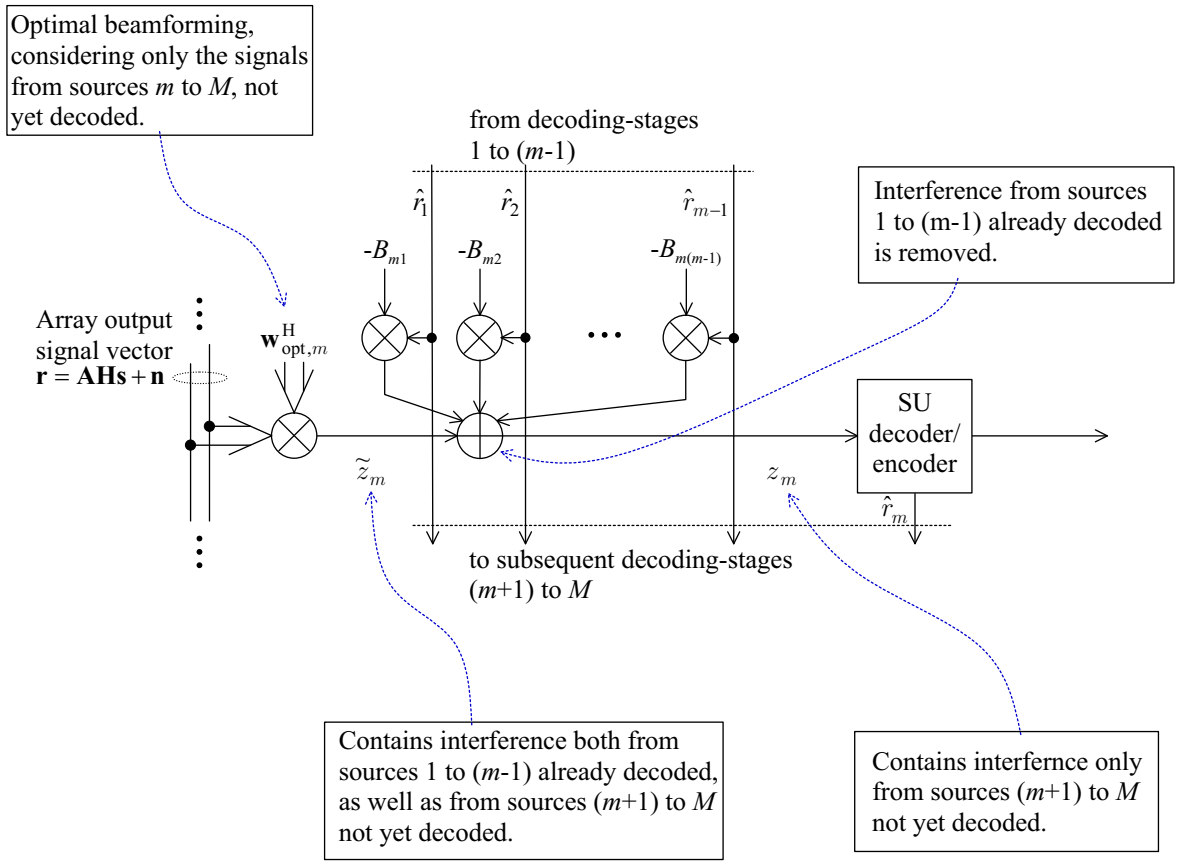


Figure 4.16: Detail of the receiver structure implementing successive decoding. Shown is the section where source signal m is processed.

Often, the effect of erroneous decoding and resulting error propagation by feeding back erroneous estimates $\hat{r}_i \neq \tilde{r}_i$ is neglected, such that $\hat{r}_i = \tilde{r}_i$. This assumption is further justified because we consider capacities, i.e. transmission is error free by definition [VG97].[9]

Therefore it is assumed that interference from sources $1, 2, \ldots, m-1$ can be perfectly removed from $\tilde{z}_m$, such that the resulting z_m contains only interference from source $m+1, m+2, \ldots, M$ not yet decoded.

Then, with $\hat{r}_i = \tilde{r}_i$, (4.61) can be rewritten according to

$$z_m = \mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}} \mathbf{a}_m \tilde{r}_m + \sum_{i=m+1}^{M} \mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}} \mathbf{a}_i \tilde{r}_i + \mathbf{w}_{\mathrm{opt},m}^{\mathrm{H}} \mathbf{n}. \tag{4.63}$$

This, in turn, can be rewritten more compact in matrix notation according to

$$\mathbf{z} = \mathbf{F}\mathbf{r} - \mathbf{B}\tilde{\mathbf{r}}, \tag{4.64}$$

where $\mathbf{F} \in \mathbb{C}^{M \times L}$ is the *linear feedforward filter*, and $\mathbf{B} \in \mathbb{C}^{M \times M}$ is the *linear causal feedback filter* [VG97], further $\mathbf{r} \in \mathbb{C}^{L}$ was introduced as the array output signal vector in (3.64).

The feedforward filter matrix $\mathbf{F}$ is defined by the weight vectors $\mathbf{w}_{\mathrm{opt},m}$ according to

$$\mathbf{F} = \begin{pmatrix} \mathbf{w}_{\mathrm{opt},1}^{\mathrm{H}} \\ \mathbf{w}_{\mathrm{opt},2}^{\mathrm{H}} \\ \vdots \\ \mathbf{w}_{\mathrm{opt},M}^{\mathrm{H}} \end{pmatrix}, \tag{4.65}$$

where the function of the feedforward filter is to minimise the impact of multiple-access interference from the yet undecoded sources $m+1, m+2, \ldots, M$.

The feedback filter matrix $\mathbf{B}$ is the strictly lower part of the matrix product $\mathbf{FA}$, hence

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \mathbf{w}_{\mathrm{opt},2}^{\mathrm{H}} \mathbf{a}_1 & 0 & 0 & \cdots & 0 \\ \mathbf{w}_{\mathrm{opt},3}^{\mathrm{H}} \mathbf{a}_1 & \mathbf{w}_{\mathrm{opt},3}^{\mathrm{H}} \mathbf{a}_2 & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{w}_{\mathrm{opt},M}^{\mathrm{H}} \mathbf{a}_1 & \mathbf{w}_{\mathrm{opt},M}^{\mathrm{H}} \mathbf{a}_2 & \cdots & \mathbf{w}_{\mathrm{opt},M}^{\mathrm{H}} \mathbf{a}_{M-1} & 0 \end{pmatrix}, \tag{4.66}$$

where the function of the feedback filter is to mitigate interference caused by sources already decoded. The element of $\mathbf{B}$ in the m-th row and i-th column, denoted as B_{mi}, was already defined in (4.62).

As for the classical MAC the central question for successive decoding is again how to find the optimal decoding order minimising the sum transmit power for given target rates. This will be dealt with in the next section.

[9] However, there are some publications dealing with error-prone successive decoding, e.g. [Mec01]. There it is stated that error-prone stripping may require only a small power overhead to achieve close to optimum performance.

4.6.2 Power Allocation for Successive Decoding

We have seen from the consideration presented above that we cannot rely on a polymatroidal structure of $\mathcal{P}_{\mathrm{VMAC}}$ when looking for the optimal power allocation, which would naturally also lead to successive decoding.

In Sec. 4.3.3, where we have dealt with the optimal decoding order for successive decoding for the classical fading MAC, it was pointed out that there is a greedy algorithm to identify that decoding order for that the sum transmit power is minimised among all decoding orders possible, although the power region $\mathcal{P}_{\mathrm{CMAC}}$ is not a contra-polymatroid.

We will see in the following that this simple solution is not applicable for the vector MAC, unfortunately.

Optimal Decoding Order for Successive Decoding for the VMAC

For successive decoding, the receive powers P_m (or equivalently the element SNR γ_m) which are required to achieve demanded rates R_m with successive decoding are obtained from (4.58) by solving for $P_{\pi_v(m)}$ for given $R_{\pi_v(m)}$:

$$P_{\pi_v(m)} = \left\{ P : R_{\pi_v(m)} = \mathrm{E}\left\{ \log_2\left(1 + P\left|H_{\pi_v(m)}\right|^2 \mathbf{a}_{\pi_v(m)}^{\mathrm{H}} \left(\mathbf{K_n} + \mathbf{K_v}(\mathcal{S}_{v,m-1})\right)^{-1} \mathbf{a}_{\pi_v(m)}\right)\right\}_{\boldsymbol{H}}\right\},$$
$$m = 1, 2, \ldots, M, \qquad \mathcal{S}_{v,m-1} = \{\pi_v(1), \pi_v(2), \ldots, \pi_v(m-1)\}. \tag{4.67}$$

(Note that (4.67) can be readily solved for $P_{\pi_v(m)}$ unambiguously employing standard numerical methods since the right hand side is concave in P.)

When looking for the optimal decoding order π_v^* for that the sum transmit power obtained with (4.67) and (4.15) (i.e. $P_m = p_m/\mu_m$) is minimised, the central point is again that we are facing a combinatorial optimisation problem, as it was pointed out in Sec. 4.3.3. There we have seen that for the classical fading MAC this combinatorial optimisation problem is reduced to the very simple problem of list sorting due to the particular structure of the edge weights of the tree-graph depicted in Fig. 4.6 (recall that the set of mean received powers was independent of the particular decoding order).

Referring to Fig. 4.6 where the combinatorial optimisation problem is shown as a tree graph, we have to ask whether there is again a structure in the edge weights that could be exploited to find the optimal branch without enumeration of all solutions.

The simple solution to the combinatorial optimisation problem for the classical MAC shown in Sec. 4.3.3 is established on the fact that the amount of interference "seen" by source $\boldsymbol{\pi}_v(m)$ depends only on the *number* $(m-1)$ of interfering sources not yet decoded and *not* on the particular permutation $\boldsymbol{\pi}_v$. This is related to the fact that the amount of interference that an individual source is causing is the same for all sources. Therefore it holds for the classical fading MAC $P_{\pi_{v1}(m)} = P_{\pi_{v2}(m)}$, $m = 1, 2, \ldots, M$, for any two arbitrary permutations $\boldsymbol{\pi}_{v1}, \boldsymbol{\pi}_{v2}$.

This is generally **not** the case for the vector MAC, because the amount of interference for a source m caused by a source i depends on their steering vectors $\mathbf{a}_m$ and $\mathbf{a}_i$.

An exception are the two special cases of orthogonal steering vectors, and equal steering vectors (resulting in the classical MAC). In the first case the decoding order is irrelevant as there is no interference, and in the latter case the decoding order is easily obtained by sorting according to the link loss as pointed out in Sec. 4.3.3 for the classical MAC.

For the general case of the vector MAC (i.e. the steering vectors are neither mutual orthogonal, nor are they all equal) it holds for some $m = 1, 2, \ldots, M$:

$$\begin{aligned} P_{\pi_{v1}(n)} &= P_{\pi_{v2}(n)}, \text{ if } \pi_{v1}(n) = \pi_{v2}(n), \forall n = 1, 2, \ldots, m \\ P_{\pi_{v1}(n)} &\neq P_{\pi_{v2}(n)}, \text{ if } \pi_{v1}(n) \neq \pi_{v2}(n), \forall n = m+1, m+2, \ldots M. \end{aligned} \tag{4.68}$$

Thus, between the two extremes of equal and orthogonal steering vectors, respectively, where all powers $P_{\pi_{v1}(m)}$ are equal, we expect that there are source distributions for that the transmit sum-power is more sensitive to the particular choice of decoding order than for other source distributions.

Eqn. (4.68) allows to represent power allocation for successive decoding for the vector MAC in a tree-graph similar to the classical MAC (cf. Fig. 4.6).

With a simple example with $M = 4$ sources we now want to shed more light on the problem of finding the optimal decoding order for successive decoding for the vector MAC.

Example 4.7: **Decoding Order for Onion Peeling for the Vector MAC with 4 Sources** The sources are located on a line where the off-boresight angles are given by

$$\vartheta_1 = 0^\circ, \vartheta_2 = 2^\circ, \vartheta_3 = 4^\circ, \vartheta_4 = 6^\circ,$$

such that the angular separation between any two adjacent sources is 2° (cf. Fig. 4.17). The angular separation is chosen such that we can expect a distinct dependency of the transmit sum-power on the particular decoding order (recall that the transmit sum-power does not depend on the decoding order if the steering vectors are identical for all sources, or if they are mutually orthogonal). Further, in the following we will neglect the slight differences in free-space basic transmission loss and array element gains, respectively, for the different sources, i.e. we assume $\mu_m = \mu_j$, for $m, j = 1, 2, 3, 4$.

Further, it is assumed that the satellite array antenna has a hexagonal contour, and consists of $L = 127$ circular array elements, arranged on a triangular grid (this array model was presented in detail in Sec. 3.6.1). Finally, the required rate for all sources is $R = 2$ and we assume an AWGN channel (i.e. no fading).

There are $M! = 24$ possible decoding orders for successive decoding, and the receive powers P_m, required to achieve $R = 2$, are computed for all 24 permutations $\boldsymbol{\pi}_v, v = 1, 2, \ldots, 24$ with (4.58) (because an AWGN channel is assumed, averaging over fading is omitted):

$$R_{\pi_v(1)} = R = \log_2\left(1 + \frac{P_{\pi_v(1)} L}{2\sigma_{\mathrm{n}}^2}\right) \Rightarrow \frac{P_{\pi_v(1)}}{2\sigma_{\mathrm{n}}^2} = \frac{1}{L}\left(2^R - 1\right) = 0.0236, \tag{4.69}$$

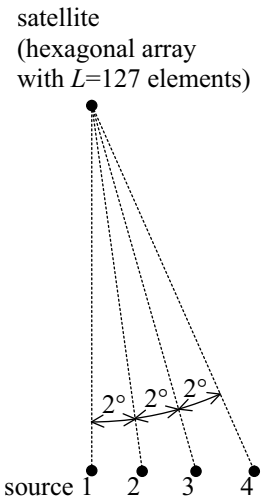


Figure 4.17: Distribution of the $M = 4$ sources.

where it is obvious that this is independent of the particular decoding order, and for the remaining sources $m = 2, 3, 4$ the receive powers are obtained from

$$\begin{aligned} R_{\pi_v(m)} = R &= \log_2\left(1 + \frac{P_{\pi_v(m)}}{2\sigma_{\mathrm{n}}^2}\mathbf{a}_{\pi_v(m)}^{\mathrm{H}}\left(\sum_{i=1}^{m-1}\frac{P_{\pi_v(i)}}{2\sigma_{\mathrm{n}}^2}\mathbf{a}_{\pi_v(i)}\mathbf{a}_{\pi_v(i)}^{\mathrm{H}} + \mathbf{I}_L\right)^{-1}\mathbf{a}_{\pi_v(m)}\right) \Rightarrow \\ \frac{P_{\pi_v(m)}}{2\sigma_{\mathrm{n}}^2} &= \frac{2^R - 1}{\mathbf{a}_{\pi_v(m)}^{\mathrm{H}}\left(\sum_{i=1}^{m-1}\frac{P_{\pi_v(i)}}{2\sigma_{\mathrm{n}}^2}\mathbf{a}_{\pi_v(i)}\mathbf{a}_{\pi_v(i)}^{\mathrm{H}} + \mathbf{I}_L\right)^{-1}\mathbf{a}_{\pi_v(m)}}, \; m = 2, 3, 4. \end{aligned} \tag{4.70}$$

As already pointed out before (cf. (4.68)), it is evident from (4.70) that the receive power $P_{\pi_v(m)}$ required to achieve the rate tuple $\mathbf{R}$ depends on the particular permutation, because of the dependence on the steering vectors $\mathbf{a}_{\pi_v(i)}$ of the sources $i = 1, 2, \ldots, m-1$ interfering with source m.

For the example discussed here, the resulting normalised receive powers $P_m/(2\sigma_{\mathrm{n}}^2)$ are shown for all permutations in a tree graph in Fig. 4.18. As expected, the required receive powers depend on the particular decoding order (cf. (4.68)). Further, the resulting receive powers illustrate that the greedy algorithm for power allocation that is optimal for the classical MAC, is not applicable for the vector MAC: the greedy algorithm cannot decide in the first stage of power allocation for the permutations that will lead to a minimum sum-power solution, i.e. $\pi_v(1) = 2$ and $\pi_v(1) = 3$, respectively.

Fig. 4.19 depicts that permutations that lead to minimum sum-powers and shows also worst-case permutations resulting in maximum sum-powers. (Note that we can generally expect that there is a unique optimal permutation, however, because of the particular source distribution chosen for this example there are four permutation resulting in similar minimum and maximum, respectively, sum-powers.)

Obviously, the permutations resulting in similar minimal sum-power have in common that the two outer sources 1 and 4, respectively, are decoded first, followed from the two inner sources 2 and 3. Further, the worst-case, resulting in maximum sum-power, is encountered when the two inner

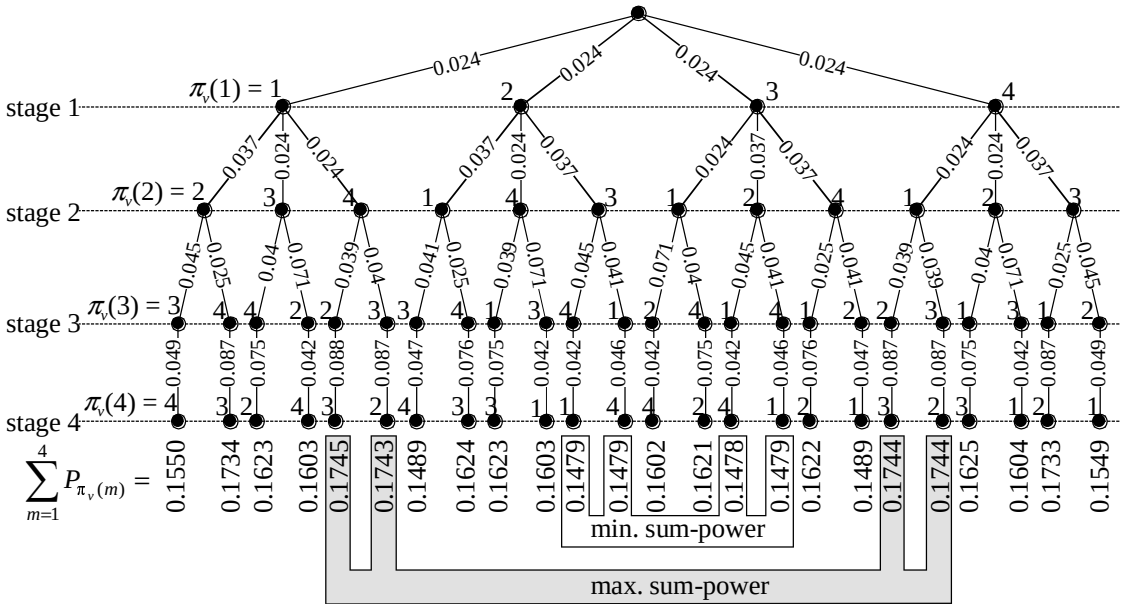


Figure 4.18: Tree graph representing the $M!$ possible decoding orders for $M = 4$. The required rate is $R = 2$, such that for the source that is decoded last the required normalised receive power is $P_{\pi_v(1)}/(2\sigma_{\mathrm{n}}^2) = (2^R - 1)/L = 0.0236$, where $L = 127$ is the number of array elements. Each vertex of the tree is labelled with the normalised required receive power $P_{\pi_v(m)}/(2\sigma_{\mathrm{n}}^2)$. (Note that it holds $\sum_{m=1}^{4} p_m = \mu \sum_{m=1}^{4} P_m$, because it is assumed $\mu_m = \mu_j = \mu$, $m, j = 1, 2, 3, 4$. Therefore minimising $\sum_{m=1}^{4} P_m$ minimises also $\sum_{m=1}^{4} p_m$.)

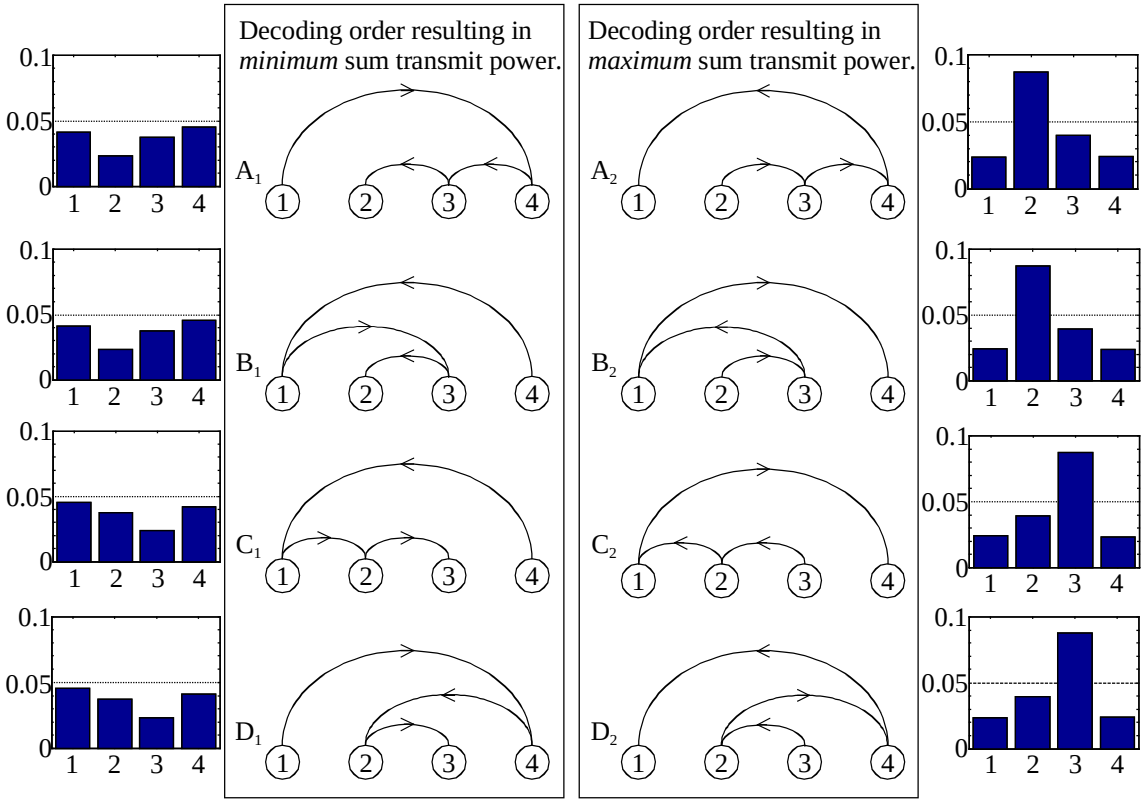


Figure 4.19: Due to symmetry of the source locations, decoding orders A_1 is equivalent to C_1, and B_1 to D_1, the same holds for the worst-case decoding orders.

sources 2 and 3 are decoded first. The explanation for this observation follows below.

■

The source being decoded first has to cope with the maximum number of interferers ($M-1$ sources), therefore it is advantageous to begin successive decoding with a source that is, owing to spatial filtering, less sensitive to the interference from the $M-1$ sources. For this reason the minimum sum-power decoding orders start in the example above with either source 1, or with source 4, because these two sources have only a single direct neighbour, whereas source 2 and 3 have two direct neighbours. Equivalently, we can state that power allocation should start with that source that has the most close neighbours (in the example: sources 2 and 3).

4.7 Fixed Beamforming

In Sec. 4.4–4.6 we have seen that optimal beamforming in element space[10] is maximising the SINR for a given source m and a set of interfering sources. In fact, a receiver employing optimal element space beamforming and successive decoding already achieves rates on the boundary of the capacity region of the vector MAC.

So, if we expect that we cannot do any better than with optimal beamforming in element space, why do we want to deal with fixed beamforming at all?

In Sec. 3.6 we have introduced two exemplary satellite scenarios, which employ both fixed beams. Fixed beamforming in combination with independent decoding is the usual approach in nowadays satellite systems [LWJ00].

Also successive decoding is suggested in combination with fixed beamforming [Ern99, Ern01], and we want to include the rates achievable with these approaches in a comparison with what is achievable with optimal adaptive beamforming with independent and successive decoding in Chap. 5.

Further, a fixed beamforming stage can be combined with subsequent adaptive *beam space* beamforming, i.e. the adaptive beamforming is applied to the output of the fixed beamforming stage. In this case especially *partial adaptive* beam space beamforming is an attractive option (cf. Sec. 5.2.2) [LL96].

Therefore in this section we will discuss the impact of fixed beamforming on the achievable rates in comparison with optimal element space processing.

4.7.1 Signal Model for Fixed Beamforming

Fig. 4.20 shows the block diagram of a receiver employing a fixed beamforming network (BFN) located behind the antenna array. The fixed BFN creates L_{B} beams from the L-element antenna

[10] The term *element space* refers to the fact that optimal adaptive beamforming is applied to the element receive signal vector $\mathbf{r}$.

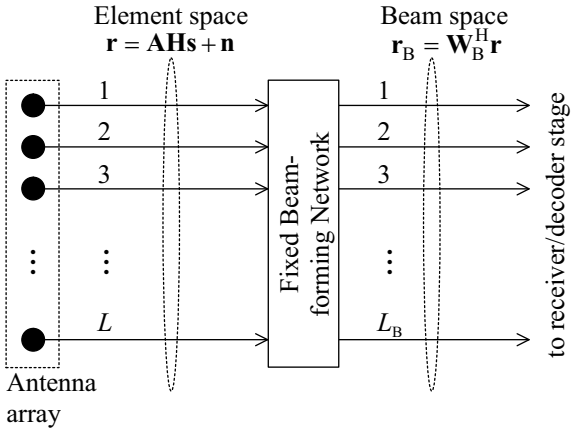


Figure 4.20: Block diagram of receiver employing a generic beamforming network. The L branches of the antenna array represent the input for the beamforming network (i.e. the input vector to the BFN is $\mathbf{r}$). The beamforming operation of the BFN is described by the weight matrix $\mathbf{W}_\mathrm{B}$. The output vector of the BFN is $\mathbf{r}_\mathrm{B}$, having L_B elements.

array, such that the *beamforming network output vector* $\mathbf{r}_\mathrm{B}$ can be written according to

$$\mathbf{r}_\mathrm{B} = \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{r}, \tag{4.71}$$

where the L_B columns of $\mathbf{W}_\mathrm{B} \in \mathbb{C}^{L\times L_B}$ are the weight vectors defining the L_B fixed beams.

Because $\mathbf{r}$ is a proper, zero-mean multivariate Gaussian random vector, also $\mathbf{r}_\mathrm{B}$ is proper and zero-mean (a random vector obtained from an affine transformation of a proper random vector is again proper [NM93], also cf. App. B).

Then the PDF of $\mathbf{r}_\mathrm{B}$ is defined by (B.3) with the covariance matrix $\mathbf{K}_{\mathbf{r}_\mathrm{B}} \in \mathbb{C}^{L_B\times L_B}$ of $\mathbf{r}_\mathrm{B}$. Using (4.71) the covariance matrix of the beamforming network output vector $\mathbf{K}_{\mathbf{r}_\mathrm{B}}$ is given by

$$\begin{aligned}\mathbf{K}_{\mathbf{r}_\mathrm{B}} &= \mathrm{E}\left\{\mathbf{r}_\mathrm{B}\mathbf{r}_\mathrm{B}^\mathrm{H}\right\} = \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_\mathbf{r}\mathbf{W}_\mathrm{B} \\ &= \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_\mathbf{v}\mathbf{W}_\mathrm{B} + \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_\mathbf{n}\mathbf{W}_\mathrm{B} \\ &= \sum_{m=1}^{M}\mathbf{a}_{\mathrm{B},m}\mathbf{a}_{\mathrm{B},m}^\mathrm{H}\frac{p_m\left|h_m\right|^2}{\mu_m} + 2\sigma_\mathrm{n}^2\mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{W}_\mathrm{B},\end{aligned} \tag{4.72}$$

where the *beam space steering vector* $\mathbf{a}_{\mathrm{B},m}$ of source m is defined as

$$\mathbf{a}_{\mathrm{B},m} = \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{a}_m, \tag{4.73}$$

and

$$\mathbf{K}_{\mathbf{n}_\mathrm{B}} = \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_\mathbf{n}\mathbf{W}_\mathrm{B} = 2\sigma_\mathrm{n}^2\mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{W}_\mathrm{B} \tag{4.74}$$

is the *beam space noise covariance matrix*. Note that $\mathbf{K}_{\mathbf{n}_\mathrm{B}}$ may not be a diagonal matrix depending on $\mathbf{W}_\mathrm{B}$, such that the noise in different beams may be correlated.

Recall the definition of the covariance matrix for the array output vector $\mathbf{K}_\mathbf{r}$ ((3.78) with (3.76)), which is repeated here for the sake of clarity

$$\mathbf{K}_\mathbf{r} = \sum_{m=1}^{M}\mathbf{a}_m\mathbf{a}_m^\mathrm{H}\frac{p_m\left|h_m\right|^2}{\mu_m} + \mathbf{K}_\mathbf{n},$$

where the more general case of possibly correlated noise with covariance matrix $\mathbf{K}_{\mathbf{n}}$ is considered. Comparing the definition of $\mathbf{K}_{\mathbf{r}}$ with the equation for the fixed BFN output vector covariance matrix $\mathbf{K}_{\mathbf{r}_\mathrm{B}}$, given by (4.72), reveals the general equivalence of the descriptions of the receiver without and with fixed beamforming.

Therefore, Sec. 4.4–4.6 apply also to the receiver employing fixed beamforming. This holds in particular for the power allocation for independent decoding, the definitions of capacity and power region and, related to this, optimal resource allocation procedures.

For convenience of notation we introduce similar to the definitions in Sec. 3.5.3 the covariance matrix $\mathbf{K}_{\mathbf{r}_\mathrm{B}}(\mathcal{S})$ for a subset of sources $\mathcal{S} \subseteq \mathcal{M}$, with

$$\begin{aligned}\mathbf{K}_{\mathbf{r}_\mathrm{B}}(\mathcal{S}) &= \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_{\mathbf{r}}(\mathcal{S})\mathbf{W}_\mathrm{B} = \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_{\mathbf{v}}(\mathcal{S})\mathbf{W}_\mathrm{B} + \mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_{\mathbf{n}}\mathbf{W}_\mathrm{B} \\ &= \sum_{m\in\mathcal{S}} \mathbf{a}_{\mathrm{B},m}\mathbf{a}_{\mathrm{B},m}^\mathrm{H}\frac{p_m\left|h_m\right|^2}{\mu_m} + 2\sigma_\mathrm{n}^2\mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{W}_\mathrm{B}. \end{aligned} \tag{4.75}$$

With the definition of the covariance matrix of the fixed BFN output vector $\mathbf{r}_\mathrm{B}$ at hand, it is easy to provide the equation for maximum mutual information $\mathcal{I}(\boldsymbol{S};\boldsymbol{R}_\mathrm{B})$:

$$\begin{aligned}\max\mathcal{I}(\boldsymbol{S};\boldsymbol{R}_\mathrm{B}) &\overset{(a)}{=} \log_2\left((\pi\mathrm{e})^{L_\mathrm{B}}\det\left(\mathbf{K}_{\mathbf{r}_\mathrm{B}}\right)\right) - \log_2\left((\pi\mathrm{e})^{L_\mathrm{B}}\det\left(2\sigma_\mathrm{n}^2\mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{W}_\mathrm{B}\right)\right) \\ &= \log_2\left(\det\left(\frac{1}{2\sigma_\mathrm{n}^2}\mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{K}_{\mathbf{v}}\mathbf{W}_\mathrm{B}\left(\mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{W}_\mathrm{B}\right)^{-1} + \mathbf{I}_{L_\mathrm{B}}\right)\right), \end{aligned} \tag{4.76}$$

where equality in (a) follows from the definition of mutual information using differential entropy of a proper multivariate normal random variable with covariance matrix $\mathbf{K}_{\mathbf{r}_\mathrm{B}}$ given by (4.72) (also cf. (B.3) and (4.43) with $\mathcal{S} = \mathcal{M}$).

The fixed beamformer is fully specified by the fixed beamforming matrix $\mathbf{W}_\mathrm{B}$, which in general may be implement by an analog beamforming network or via digital signal processing. However, analog beamforming always introduces an increasing signal degradation with increasing number of beams, while with digital beamforming the number of beams is only limited by the processing power of the involved digital signal processors.

Some of the nowadays and upcoming satellite systems employ multi-beam antennas creating around 200 spot beams, making digital beamforming the only viable implementation method. For example, the ICO system with 163 (cf. Sec. 3.6.1), and Thuraya with 246 fixed spot beams [SDR+02], both employ digital beamforming. Therefore, we can assume that also fixed beamforming that does not implement any possibly elaborate adaptive beamforming algorithm is implemented via digital signal processors.

Besides the number of beams that are created by the fixed beamformer, an even more important characteristics is of course the concrete realisation of the beamforming matrix $\mathbf{W}_\mathrm{B}$. With $\mathbf{W}_\mathrm{B}$ not only the number of beams has to be defined, but also the shape of the beams (e.g. sidelobe level) and the angular directions which the spot beams are pointing to.

We have seen in Sec. 3.6, where the ICO and EuroSkyWay array antenna models have been discussed, that the beamforming matrix $\mathbf{W}_\mathrm{B}$ may be designed to control the sidelobe level to reduce interference from co-channel cells. Further, the spot beams may be arranged to completely cover the service area of the respective satellite with a regular cell pattern, where a cell may be defined by various spot beam gain contours (typical values are $3\ \mathrm{dB}$, $4.3\ \mathrm{dB}$).

In the further, we also want to address *beam space beamforming*, where subsequent to the fixed beamforming stage adaptive beamforming is performed.

In this context, in the literature it is often stated that *orthogonal* fixed beams are created via a Butler matrix beamformer in the analog domain or via discrete fourier transformation (DFT) for digital beamforming (see e.g. [LL96]). Orthogonality refers to the property of the beamforming matrix $\mathbf{W}_{\mathrm{B}}$ that the column and row vectors are mutually orthogonal, such that the beam space noise covariance matrix in (4.74) becomes a diagonal matrix, which in turn means that the noise in different orthogonal fixed beams is uncorrelated. In contrast, we will see that the fixed beams created for a complete coverage of the service area with overlapping cells which are often defined as the $3\ \mathrm{dB}$ of $4.3\ \mathrm{dB}$ spot beam gain contour (cf. Sec. 3.6), are typically non-orthogonal.

Finally, all this raises the question what principle limits are posed by the fixed beamforming on the achievable rates, both orthogonal, as well as non-orthogonal. This will be discussed below.

4.7.2 Impact of Fixed Beamforming on Achievable Rates

From a general result of information theory, namely the *data processing inequality* [CT91], it follows that no processing of the array output vector $\mathbf{r}$ can increase the information that $\mathbf{r}$ contains about the source signals $\mathbf{s}$.

Because $\mathbf{r}_{\mathrm{B}}$ is a function of $\mathbf{r}$ according to (4.71), the corresponding random variables $\boldsymbol{S}$, $\boldsymbol{R}$ and $\boldsymbol{R}_{\mathrm{B}}$ form a Markov chain in the order $\boldsymbol{S} \rightarrow \boldsymbol{R} \rightarrow \boldsymbol{R}_{\mathrm{B}}$ [CT91].

(Random variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ form a Markov chain in that order, denoted by $\boldsymbol{X} \rightarrow \boldsymbol{Y} \rightarrow \boldsymbol{Z}$, if their joint PDF can be written as $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{z}|\boldsymbol{y})$, i.e. the conditional distribution of $\boldsymbol{Z}$ is independent of $\boldsymbol{X}$ such that $p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{z}|\boldsymbol{y})$. In particular, if we can write $\boldsymbol{Z} = f(\boldsymbol{Y})$, i.e. $\boldsymbol{Z}$ is a function of $\boldsymbol{Y}$, then $\boldsymbol{X} \rightarrow \boldsymbol{Y} \rightarrow \boldsymbol{Z}$.)

Then the data processing inequality says that

$$\mathcal{I}(\boldsymbol{S}; \boldsymbol{R}_{\mathrm{B}}) \leq \mathcal{I}(\boldsymbol{S}; \boldsymbol{R}), \qquad \text{if } \boldsymbol{S} \rightarrow \boldsymbol{R} \rightarrow \boldsymbol{R}_{\mathrm{B}}, \tag{4.77}$$

i.e., if the random variables $\boldsymbol{S}$, $\boldsymbol{R}$, and $\boldsymbol{R}_{\mathrm{B}}$ form a Markov chain in that order, then the information in $\boldsymbol{r}_{\mathrm{B}}$ about the source signal vector $\mathbf{s}$ can only be equal or less than the information available in $\boldsymbol{r}$.

Therefore, as mentioned previously, introducing fixed beamforming only can leave the information about $\mathbf{s}$ available at the receiver unchanged or it even reduces it.

Of course, it is of particular interest under what conditions the available information about the source signals $\mathbf{s}$ remains unchanged by the fixed BFN (then equality is obtained in (4.77)), and that regardless of the particular spatial distribution of the sources (we will say in this case that the fixed beamforming is lossless) and when it is possibly reduced.

If the fixed beamforming is **not** lossless, then in general equality does not hold in (4.77), and the degree of reduction of mutual information strongly depends on the particular source distribution.

Generally, equality in (4.77) is obtained if and only if the mutual information between $\boldsymbol{S}$ and $\boldsymbol{R}$ is zero given $\boldsymbol{R}_{\mathrm{B}}$, i.e. $\mathcal{I}(\boldsymbol{S}; \boldsymbol{R}|\boldsymbol{R}_{\mathrm{B}}) = 0$. This arises if the conditional PDF of $\boldsymbol{R}$ depends only on $\boldsymbol{R}_{\mathrm{B}}$, i.e. $p(\mathbf{r}|\mathbf{s}, \mathbf{r}_{\mathrm{B}}) = p(\mathbf{r}|\mathbf{r}_{\mathrm{B}})$ and, therefore, $\boldsymbol{S} \rightarrow \boldsymbol{R}_{\mathrm{B}} \rightarrow \boldsymbol{R}$ (this is readily verified using the definition of conditional mutual information in (B.15)) [CT91]. This holds in particular if $\boldsymbol{R}$ is a

function of $\boldsymbol{R}_{\mathrm{B}}$, i.e. $\boldsymbol{R} = g(\boldsymbol{R}_{\mathrm{B}})$, and, finally, this function is defined by solving (4.71) for $\mathbf{r}$, when $\mathbf{W}_{\mathrm{B}}$ and $\mathbf{r}_{\mathrm{B}}$ are given.

In other words, we obtain the intuitive result that, if it is possible to reconstruct the array output vector $\mathbf{r}$ from observation of the fixed BFN output vector $\mathbf{r}_{\mathrm{B}}$ with knowledge of $\mathbf{W}_{\mathrm{B}}$, this must mean that $\mathbf{r}_{\mathrm{B}}$ contains all information about $\mathbf{s}$ also contained in $\mathbf{r}$. Then equality in (4.71) is obtained, regardless of the particular source distribution.

With $\mathbf{W}_{\mathrm{B}}$ being a $L \times L_{\mathrm{B}}$ matrix, we discuss in the following the three cases $L = L_{\mathrm{B}}$, $L < L_{\mathrm{B}}$, and $L > L_{\mathrm{B}}$. The rank of the fixed steering matrix is generally bounded by [GL96]

$$\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) \leq \min\left\{L, L_{\mathrm{B}}\right\}, \tag{4.78}$$

where it is reasonable to assume in the following that the fixed beamforming matrix $\mathbf{W}_{\mathrm{B}}$ is full rank, such that equality holds in (4.78).

Equal number of fixed beams and array elements: $L_{\mathrm{B}} = L$ If $L = L_B$, then $\mathbf{W}_{\mathrm{B}}$ is a square matrix with $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$, and therefore invertible. Then the transformation from element to beam space is generally lossless, as the array output vector $\mathbf{r}$ can be derived from the beam space receive vector $\mathbf{r}_{\mathrm{B}}$ by inversion of the fixed beamforming matrix $\mathbf{W}_{\mathrm{B}}$ according to

$$\mathbf{r} = \left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\right)^{-1} \mathbf{r}_{\mathrm{B}} \qquad (\mathbf{W}_{\mathrm{B}} \in \mathbb{C}^{L \times L}, \mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L), \tag{4.79}$$

and equality in (4.77) holds.

In this case, equality in (4.77) also follows simply if we explicitly write the mutual information $\mathcal{I}(\boldsymbol{S}; \boldsymbol{R}_{\mathrm{B}})$, which is given by

$$\begin{aligned} \mathcal{I}(\boldsymbol{S}; \boldsymbol{R}_{\mathrm{B}}) &\overset{(a)}{=} \log_2\left(\det\left(\frac{1}{2\sigma_{\mathrm{n}}^2}\mathbf{K}_{\mathbf{v}}\mathbf{W}_{\mathrm{B}}\left(\mathbf{W}_{\mathrm{B}}\right)^{-1}\left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\right)^{-1}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}} + \mathbf{I}_L\right)\right) \\ &= \log_2\left(\det\left(\frac{1}{2\sigma_{\mathrm{n}}^2}\mathbf{K}_{\mathbf{v}} + \mathbf{I}_L\right)\right) \\ &= \mathcal{I}(\boldsymbol{S}; \boldsymbol{R}), \end{aligned} \tag{4.80}$$

where equality in (a) follows from (4.76) because the matrix identity (A.2).

Number of fixed beams lesser than number of array elements: $L_{\mathrm{B}} < L$ Next, the case is considered that $L_{\mathrm{B}} < L$, $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L_{\mathrm{B}}$. Under this assumption, the array output vector $\mathbf{r}$ cannot be uniquely determined from the fixed BFN output vector $\mathbf{r}_{\mathrm{B}}$, because then (4.71) is an *underdetermined* system of equations with an infinite number of solutions for $\mathbf{r}$ for given $\mathbf{W}_{\mathrm{B}}$ and $\mathbf{r}_{\mathrm{B}}$ [GL96]. Although in this case there is in general a loss introduced by the fixed BFN with respect to the information available at the receiver about the source signals $\mathbf{s}$, this loss depends strongly on the distribution of the sources and may be small. This will be discussed in adequate detail in Sec. 5.2.2.

Number of fixed beams greater than number of array elements: $L_{\mathrm{B}} > L$ Finally, the case $L_{\mathrm{B}} > L$ is considered, as it is encountered, e.g., for the ICO satellite system ($L_{\mathrm{B}} = 163$ fixed beams are created with an antenna array of $L = 127$ elements, cf. Sec. 3.6.1). There can be at most L linearly independent columns (which are the beamforming weight vectors for the fixed beams) in $\mathbf{W}_{\mathrm{B}}$. Because it is reasonable to assume that the fixed beams are looking in different directions, we assume that there are indeed L linearly independent columns, such that $\mathbf{W}_{\mathrm{B}}$ is full rank with $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$. In this case we can write by multiplying both sides of (4.71) with $\mathbf{W}_{\mathrm{B}}$ from the left

$$\begin{aligned} \mathbf{W}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{r} &= \mathbf{W}_{\mathrm{B}}\mathbf{r}_{\mathrm{B}} \Rightarrow \\ \mathbf{r} &= \left(\mathbf{W}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\right)^{-1}\mathbf{W}_{\mathrm{B}}\mathbf{r}_{\mathrm{B}} \\ & \quad \left(\mathbf{W}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \in \mathbb{C}^{L\times L}, \mathrm{rank}\left(\mathbf{W}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\right) = L\right), \end{aligned} \tag{4.81}$$

where we have used that it holds for any matrix $\mathbf{X} \in \mathbb{C}^{m\times n}$ that $\mathrm{rank}\left(\mathbf{X}\mathbf{X}^{\mathrm{H}}\right) = \mathrm{rank}(\mathbf{X})$ [HJ99] and, therefore, $(\mathbf{W}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}})^{-1} \in \mathbb{C}^{L\times L}$ must exist because $\mathbf{W}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}$ has full rank since $\mathrm{rank}\left(\mathbf{W}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\right) = \mathrm{rank}(\mathbf{W}_{\mathrm{B}})$ and $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$ by assumption.[11] As stated before, because the array output vector $\mathbf{r}$ can be retrieved from fixed BFN output vector $\mathbf{r}_{\mathrm{B}}$, the fixed BFN with $L_{\mathrm{B}} > L$, $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$, is lossless.

(Note that, because $\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{B}} \in \mathbb{C}^{L_{\mathrm{B}}\times L_{\mathrm{B}}}$ and $\mathrm{rank}\left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{B}}\right) = \mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$, it must hold $\det\left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{B}}\right) = 0$. Therefore, the mutual information $\mathcal{I}(\boldsymbol{S}, \boldsymbol{R}_{\mathrm{B}})$ is not defined for $L_{\mathrm{B}} > L$ according to the definition of $\mathcal{I}(\boldsymbol{S}, \boldsymbol{R}_{\mathrm{B}})$ in (4.76). However, because it is assumed $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$, it is possible to obtain an invertible covariance matrix by choosing L elements in $\mathbf{r}_{\mathrm{B}}$.)

As expected, at least from the information theoretic point of view, it does not make much sense to employ a fixed BFN with $L_{\mathrm{B}} > L$, as the information available in $\mathbf{r}_{\mathrm{B}}$ about the source signals $\mathbf{s}$ is the same as for $L_{\mathrm{B}} = L$ (assuming in both cases $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$).

4.7.3 Orthogonal Beams

It is well known that orthogonal beams can be created in a digital beamformer by means of the discrete fourier transformation (DFT)[12] [LL96, CL96, Haa96, TT98].

Although the DFT approach for creating orthogonal beams is definitely present in the antenna literature, it is either treated without the detail required for implementation [LL96, TT98] or it is discussed only for the most simple cases of the uniform linear array (ULA)[13] or the uniform rectangular array (URA)[14] [Haa96]. In [TT98], which deals with DOA estimation using hexagonal arrays, it is mentioned that it is possible to obtain a (orthogonal) transformation to beam space also for arrays where the elements are located on a hexagonal lattice, but beyond this, no details or indications to relevant literature are provided there. Also in standard literature dealing with phased

[11] Note that (4.71) is an *overdetermined*, full rank system of equations for $L_{\mathrm{B}} > L$ and $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L$. Generally, there is no unique solution for $\mathbf{r}$ if $\mathbf{W}_{\mathrm{B}}$ and $\mathbf{r}_{\mathrm{B}}$ are known, because $\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}$ defines a subspace of $\mathbb{R}^L$ such that, in principle, $\mathbf{r}_{\mathrm{B}}$ could lie outside this subspace [GL96]. Of course, this cannot be the case here, since the known fixed beamforming output vector $\mathbf{r}_{\mathrm{B}}$ is obtained from a corresponding array output vector $\mathbf{r}$ according to (4.71), such that $\mathbf{r}_{\mathrm{B}}$ must be an element of the subspace defined by $\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}$.

[12] For analog beamforming the also well known Butler matrix, being the analog equivalent of the DFT, is used.

[13] In the ULA, all array elements are located on a straight line with a constant inter-element spacing.

[14] In the URA, the array elements are located in a square plane and are positioned on a regular rectangular lattice.

array antennas the concept of creating orthogonal beams is of course present. However, in [Mai94] the focus is put on the implications of creating orthogonal beams by analog beamforming networks, further, detailed discussion is mostly provided only for the linear array. Similar limitations apply to [Bro91].

In a URA, the array elements are located on a rectangular lattice and the DFT is readily calculated by ordering the array elements in rows and columns, and performing 1-D DFT separately across the rows and columns [Haa96].

In the general case the array elements are located on an arbitrary lattice (the most general case, allowing also irregular array element locations, like it is encountered for sparse (thinned) arrays, will not be addressed here).

We will see below that when 2-D arrays based on arbitrary lattices are treated, the theory of 2-D DFT gets more complicated, such that the calculation of appropriate beamformer weights using the DFT to create orthogonal beams is surely not obvious.

We take this lack of a detailed derivation in the literature dealing with antenna arrays as a motivation to shortly present the generation of orthogonal beams for 2-D arrays with a hexagonal lattice geometry by employing the 2-D DFT (the extension to general lattices is straight forward). Part of the following is based on [BM94] and [Dub85] where the main ideas of the theory of 2-D DFT are presented in short (for a more detailed discussion see Appendix D).

Without loss of generality it is assumed that the array elements are arranged in the x-y-plane. The array elements are further located on a lattice Λ (this will be called the *spatial sampling lattice*, indicating that the array antenna can be understood as a sensor that samples the spatially continuous wavefront of the incident signals at the discrete array element locations), which is defined by a basis $\mathbf{V} \in \mathbb{R}^{2\times 2}$.

For a hexagonal array basis $\mathbf{V}$ is provided in (3.34) according to [TT98]

$$\mathbf{V} = d_{\mathrm{s}} \begin{pmatrix} 1 & 0.5 \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix}.$$

The position $\mathbf{p}_{\mathrm{e},l}$ of element l in the x-y-plane is provided by (3.31) according to

$$\mathbf{p}_{\mathrm{e},l} = \mathbf{V}\mathbf{i}_l, \quad l = 1, 2, \ldots, L.$$

Further, we define a matrix $\mathbf{N}$ according to

$$\mathbf{N} = \begin{pmatrix} 2q-1 & q-1 \\ 1-q & q \end{pmatrix}, \tag{4.82}$$

where q is linked with L via

$$L = |\det(\mathbf{N})| = 3q^2 - 3q + 1, \quad q = 1, 2, 3, \ldots. \tag{4.83}$$

Finally, with the above definitions the L DFT beamforming weight vectors $\mathbf{w}_l^{\mathrm{DFT}}$ are given by [DM84]

$$(\mathbf{w}_l^{\mathrm{DFT}})^{\mathrm{H}} = \left(\mathrm{e}^{-j2\pi \mathbf{k}_l^{\mathrm{T}} \mathbf{N}^{-1} \mathbf{i}_1}, \mathrm{e}^{-j2\pi \mathbf{k}_l^{\mathrm{T}} \mathbf{N}^{-1} \mathbf{i}_2}, \ldots, \mathrm{e}^{-j2\pi \mathbf{k}_l^{\mathrm{T}} \mathbf{N}^{-1} \mathbf{i}_L} \right), \quad l = 1, 2, \ldots, L, \tag{4.84}$$

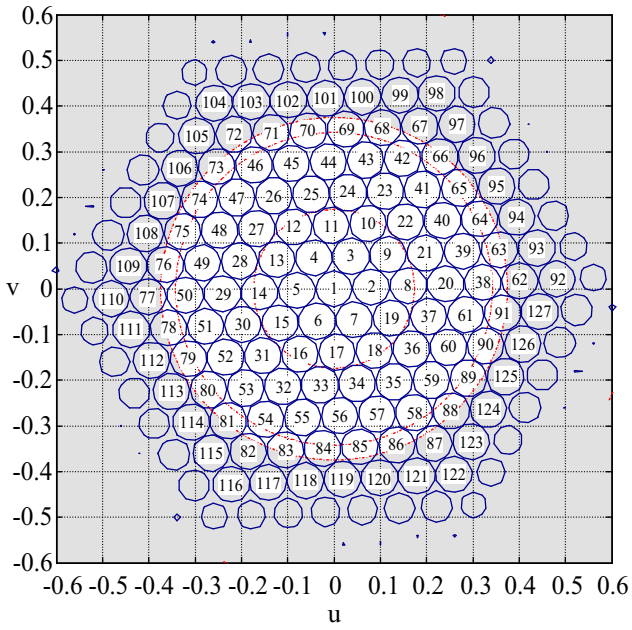


Figure 4.21: Spot beam contours (4.3 dB) and numbering of the 127 orthogonal DFT spot beams.

with $\mathbf{k}_l$ being a two-element index vector associated with the l-th beam given by

$$\mathbf{k}_l = \mathbf{N}\mathbf{i}_l. \tag{4.85}$$

Fig. 4.21 shows the 4.3 dB-contours of the spot beams resulting from the DFT beamforming for the ICO antenna array comprising 127 array elements (cf. Sec. 3.6.1).

Grouping the L DFT beamforming vector in a matrix we define the DFT beamforming matrix

$$\mathbf{W}_{\mathrm{DFT}} = \left(\mathbf{w}_1^{\mathrm{DFT}}, \mathbf{w}_2^{\mathrm{DFT}}, \ldots, \mathbf{w}_L^{\mathrm{DFT}}\right). \tag{4.86}$$

The DFT beamforming vectors $\mathbf{w}_l^{\mathrm{DFT}}$ are mutually orthogonal (we say, the beams are orthogonal), therefore in this case the beam space noise covariance matrix $\mathbf{K}_{\mathbf{n}_{\mathrm{B}}}$ (cf. (4.74)) is diagonal:

$$\mathbf{K}_{\mathbf{n}_{\mathrm{B}}} = 2\sigma_{\mathrm{n}}^2 \mathbf{W}_{\mathrm{DFT}}^{\mathrm{H}} \mathbf{W}_{\mathrm{DFT}} = 2\sigma_{\mathrm{n}}^2 L \mathbf{I}_L, \tag{4.87}$$

i.e. the noise in any two DFT spot beams is uncorrelated.

4.7.4 Non-Orthogonal Beams

We have seen in Sec. 3.6.1 that an ICO satellite creates 163 fixed spot beams, and that the resulting 163 cells are defined by the corresponding 3 dB spot beam contours (cf. Fig. 4.22). Clearly,

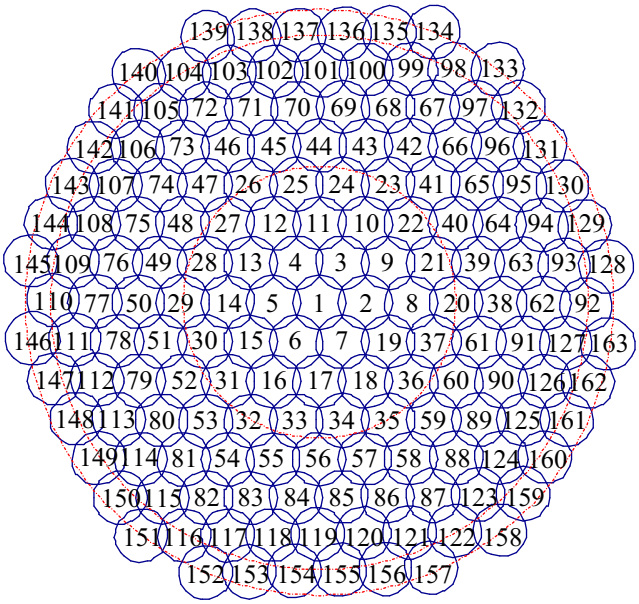


Figure 4.22: Spot beam contours (3 dB) and numbering of the 163 ICO spot beams.

the beams cannot be orthogonal (since $L_{\mathrm{B}} > L$), such that the noise in different beams must be correlated.

Fig. 4.23 shows the absolute value of the elements of the matrix product $\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{B}}$, which determines the beam space noise covariance matrix $\mathbf{K}_{\mathbf{n}_{\mathrm{B}}}$ according to (4.74).

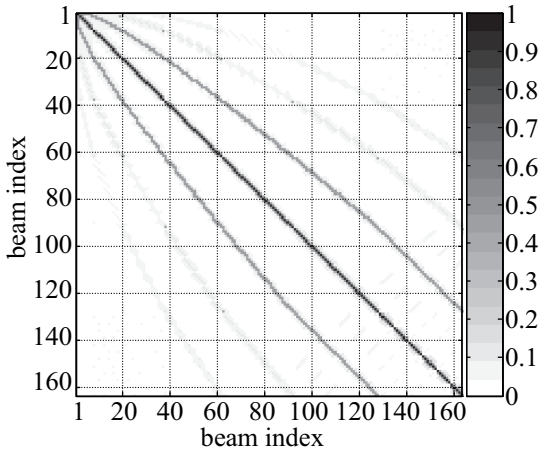


Figure 4.23: Covariance (normalised to maximum) of the noise in different beams for fixed beamforming according to the ICO scenario.

The correlation of the noise is strongest for adjacent beams and is significantly lower for all other beams.

4.8 Summary

We have presented in this section a thorough discussion of the capacity and power regions for the fading vector MAC. In particular, it was investigated whether a possibly polymatroidal structure would help in the optimal resource allocation problem, as it is the case for the classical fading MAC [TH98a, TH98b].

It was pointed out that the capacity region $\mathcal{R}_{\mathrm{VMAC}}$ is again a polymatroid (as any capacity region must be due to the chain rule of mutual information), further that the power region $\mathcal{P}_{\mathrm{VMAC}}$ is no contra-polymatroid, indicating that there may not be a greedy algorithm available for optimal power allocation.

This extended results known from the classical fading MAC presented in [TH98a, TH98b] and, further, added the discussion of optimal rate resource allocation for the fading vector MAC to the investigations presented in [SXLK98].

Chapter 5

Receiver Structures for the Fading Vector Multiple-Access Channel in Satellite Scenarios

At the beginning of the last chapter different receiver structures were presented in Fig. 4.1. The different options can be reduced to the fading vector MAC with independent or joint decoding of the source signals. Therefore, we have then turned to a thorough discussion of optimal resource allocation for the fading vector MAC.

We will now return to the said different receiver options, and present a comparison, using the results from the last chapter. The comparison will relate to minimum input SNR γ_m and EIRP $p_m G_{\mathrm{T}}$, respectively, that are required to allow all sources to transmit information at a desired rate R, employing the respective receiver structure.

Recall that if the input SNR γ_m are given, the according EIRP follow uniquely from (3.92). Therefore, in a single-source scenario the input SNR and the EIRP would be equivalent measures of receiver performance. However, in the multiple-source scenario where also successive decoding is investigated, it is indeed required to consider the EIRP or the transmit powers, being proportional to the EIRP. This is due to the fact that the optimal decoding order for successive decoding is defined on basis of the transmit sum-powers.[1] Further, independent of the particular receiver variant, the required EIRP to achieve desired information rates for the sources can be compared with the EIRP that is technologically feasible with the envisaged terminal type (e.g. hand-held). This allows to identify a region of rates that is achievable considering this technological constraint.

For the comparison of the different receiver structures we will restrict to particular scenarios, namely the two satellite system scenarios that were described in Sec. 3.6.

Comparison based on both the input SNR as well as on the EIRP is in some sense unsatisfactory, because it involves a particular source distribution instead of providing a single figure of merit related to the receiver performance in a more comprehensive way. This is inevitable, because the performance of the receivers is determined by the different capabilities to cope with interference from co-channel sources, and interference in turn depends on the particular source distribution.

[1] In fact, we have seen in particular for the classical MAC that all decoding orders yield the same set of required receive powers (equivalent to the input SNR for the vector MAC), but assuming different attenuation factors for the different sources, there is a unique decoding order minimising transmit sum-power (cf. Sec. 4.3.3).

However, for a considered satellite scenario with given fixed cell pattern, the problem that the sources can in principle be distributed arbitrarily, possibly producing required input SNR and EIRP in a wide range, is avoided by assuming only the respective worst-case source distribution to yield the maximum mutual interference. Looking at various frequency reuse schemes (e.g. cluster size 4 or 3, or other schemes), the maximum spectrum efficiency that can be achieved by the different receiver structures can be assessed (cf. Sec. 3.3.3).

5.1 Fully Adaptive Element Space Processing

We have seen in the last chapter that there is no simple method to obtain the optimal decoding order, but it is also unclear how significant the performance difference is between the optimal decoding order and the worst-case one.

Therefore, in this section we will firstly discuss the relevance of choosing the decoding order for successive decoding.

Secondly, what is of prominent interest is the comparison of successive decoding and independent decoding, where in this section it is assumed that fully adaptive (optimal) beamforming (element or beam space) is employed as discussed in Sec. 4.4 and 4.5.

Further, the parameters of the ICO satellite system are assumed, i.e. a MEO satellite employing an antenna array made up of 127 array elements is considered (cf. Sec. 3.6.1).

5.1.1 Impact of Decoding Order on Transmit Powers

As already mentioned, if we want to compare the performance of independent decoding with that of successive decoding with optimal power allocation, the investigations are severely complicated by the fact that the greedy power allocation algorithm applicable for the classical MAC does not work for the vector MAC (cf. Sec. 4.3.3 and 4.6).

It was pointed out in the last section that the transmit sum-powers for all $M!$ possible decoding orders have to be explicitly computed, in order to identify the optimal decoding order, minimising transmit sum-powers.

This poses eventually a severe computational problem depending on the number of sources M, and, therefore, we restrict in the investigations to a moderate number of sources with $M = 7$, resulting in $M! = 5040$ possible decoding orders[2].

The 7 sources are distributed as shown in Fig. 5.1, where source 1 is located at the nadir and the remaining 6 sources are located at the corners of a hexagon. For this geometry, a single parameter $\Delta\vartheta_s$ characterises the angular separation between the sources.

In Sec. 4.6.2 it was shown that the two special cases of the vector MAC, where either all steering vectors are mutually orthogonal, or all are equal, lead to simple solutions for the problem of finding the optimal decoding order for successive decoding (in the first case any decoding order is optimal

[2] In publications dealing with interference analysis for satellite scenarios, the authors often restrict to $M = 7$, cf. e.g. [LWJ00]

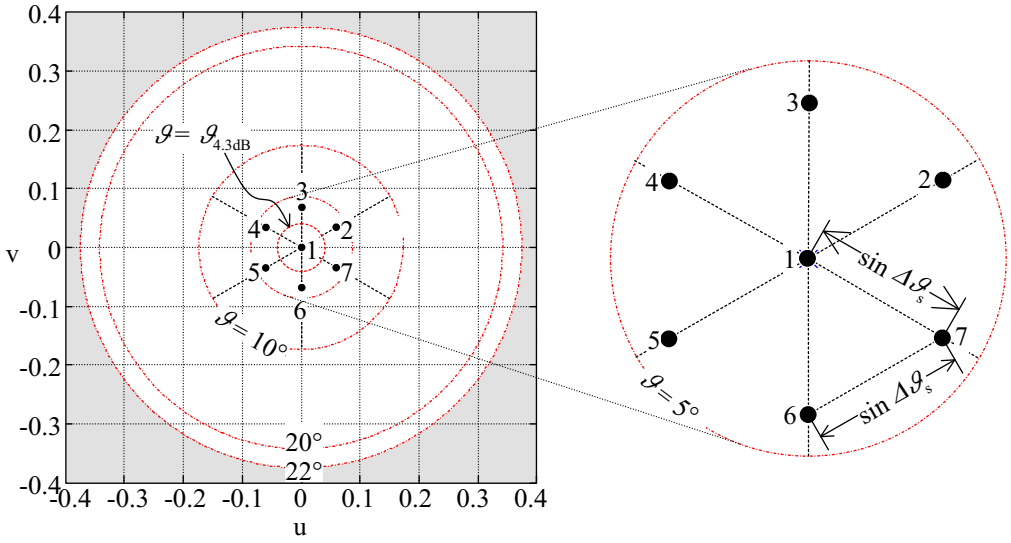


Figure 5.1: Positions of the $M = 7$ sources (in u-v-coordinates) to evaluate and compare the performance of independent and successive decoding depending on angular separation of the sources. Due to the assumed regular geometry of the source distribution, angular separation is defined by a single parameter $\Delta\vartheta_\mathrm{s}$. Concentric circles indicate off-boresight angles $\vartheta = 2.3°, 5°, 10°, 20°, 22°$, where the one-sided 4.3 dB-beamwidth for the assumed array is $\vartheta_{4.3\mathrm{dB}} = 2.3°$.

because no mutual interference is produced, in the latter case the greedy algorithm for the classical MAC is applicable). The mentioned special cases arise for some sufficiently large $\Delta\vartheta_\mathrm{s}$ to obtain (nearly) orthogonal steering vectors, and, further, for $\Delta\vartheta_\mathrm{s} = 0$ to obtain equal steering vectors, resulting in the classical MAC.

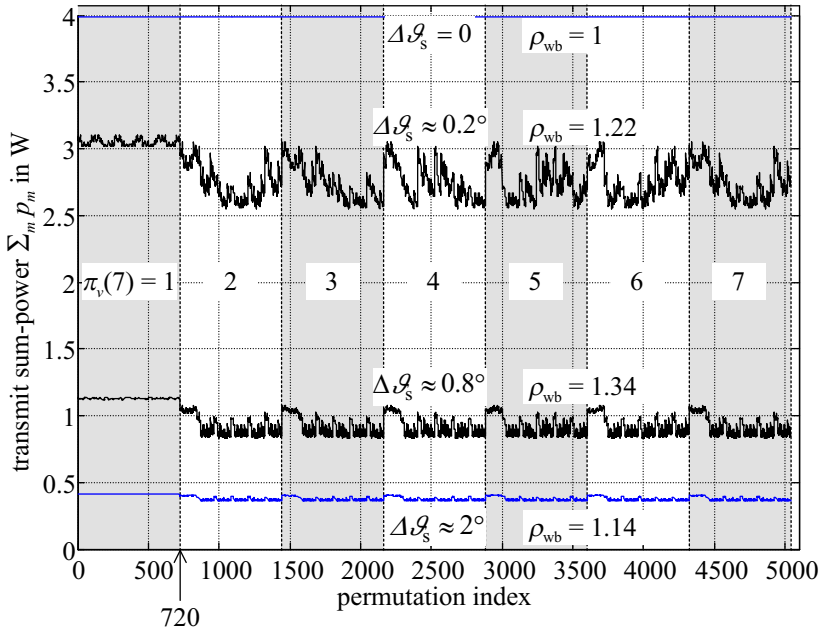
The transmit powers p_m required to achieve rate $R = 1$ for all sources are computed using (4.67), (4.15) and Fig. 5.2 shows the transmit sum-power for some values of $\Delta\vartheta_\mathrm{s}$ and for all 5040 possible decoding orders.

As expected, the ratio ρ_wb, defined as the quotient of maximum and minimum sum-power over all permutations for a given $\Delta\vartheta_\mathrm{s}$, approaches 1 as $\Delta\vartheta_\mathrm{s}$ approaches 0, or for sufficiently large $\Delta\vartheta_\mathrm{s}$, such that the steering vectors become approximately orthogonal.

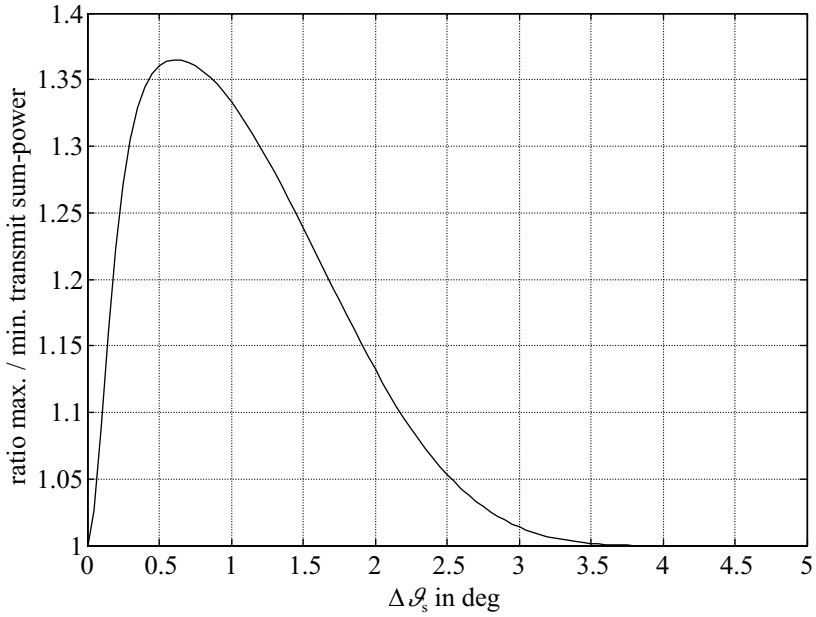
From the calculation of transmit sum-power for all permutations, those permutations can be identified that yield the optimal and the worst-case sum-powers in the most relevant range of values for $\Delta\vartheta_\mathrm{s}$, where the transmit sum-power distinctly depends on the chosen decoding order, i.e. approx. $0.1° \leq \Delta\vartheta_\mathrm{s} \leq 2.5°$ (cf. Fig. 5.2(b)).

Note that there is no unique optimal (or worst-case) permutation due to symmetry of the source distribution. However, it suffices to restrict to one particular decoding order each for minimal and maximal transmit sum-powers and Fig. 5.3 schematically depicts the decoding orders used for some further investigations (in the following we will refer to the decoding order $(4, 2, 6, 7, 5, 3, 1)$ as the *optimal* decoding order, and the decoding order $(1, 5, 4, 6, 3, 7, 2)$ as the *worst-case* decoding order).

Fig. 5.4 shows the EIRP of the $M = 7$ sources that are required to achieve $R = 1$ for the two selected decoding orders depicted in Fig. 5.3.



(a) Transmit sum-power for selected values of $\Delta\vartheta_s$ for all 5040 possible decoding orders. Note that the worst-case (maximum) sum-powers are obtained for $\pi_v(7) = 1$, i.e. if source 1 is decoded first. ρ_{wb} indicates the ratio of maximum to minimum sum-power over all permutations for a given $\Delta\vartheta_s$.



(b) Ratio of maximum to minimum sum-power ρ_{wb} over all permutations for given $\Delta\vartheta_s$. As expected, the ratio ρ_{wb} approaches 1 as $\Delta\vartheta_s$ approaches 0, or for sufficiently large $\Delta\vartheta_s$, such that the steering vectors become orthogonal.

Figure 5.2: Dependency of the transmit sum-power on the decoding order and ratio ρ_{wb} of the maximum (worst-case) sum-power and the minimum (best) sum-power for all permutations, depending on angular separation $\Delta\vartheta_s$. Target rate is $R = 1$.

Again we observe for identical or for increasingly orthogonal steering vectors that the required EIRP are independent of the chosen decoding order, as the required EIRP for the optimal and the

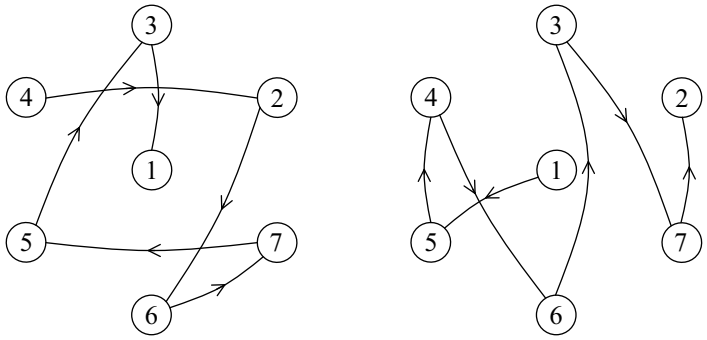


Figure 5.3: Left: Decoding order yielding minimal transmit sum-power. Right: Decoding order yielding maximal transmit sum-power.

worst-case permutation converge for $\Delta\vartheta_\mathrm{s} = 0$ and approx. $\Delta\vartheta_\mathrm{s} > 4.5°$.

The difference is maximum at approx. $\Delta\vartheta = 1°$ (cf. Fig. 5.5 for $R = 1$), where the maximal EIRP for the worst-case decoding order is approx. 2.7 times (equals 4 dB) higher than for the optimal decoding order.

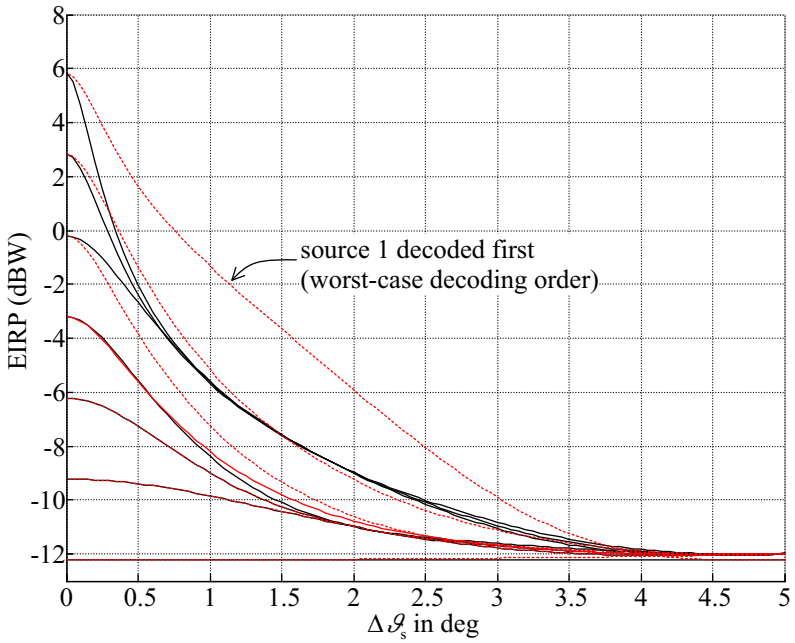


Figure 5.4: EIRP of the $M = 7$ sources for the decoding orders as depicted in Fig. 5.3 to achieve target rate $R = 1$ (Solid (black): optimal dec. order $(4, 2, 6, 7, 5, 3, 1)$. Dashed (red): worst-case dec. order $(1, 5, 4, 6, 3, 7, 2)$). Note the significant increase of transmit power for source 1, if it is decoded first.

Summarising the above results, it can be stated that selection of the decoding order for the vector MAC can have a significant impact on the required EIRP. However, the relevance of the decoding order depends on the particular distribution of the sources and, further, on the target rate R. The higher the target rate R the higher the transmit powers are and, in consequence, interference becomes the dominating factor with regard to thermal noise, which in turn is dominating for lower rates. Therefore, the impact of the decoding order on the maximum transmit powers is reduced for lower rates while it is more pronounced the higher the target rate R (again cf. Fig. 5.5).

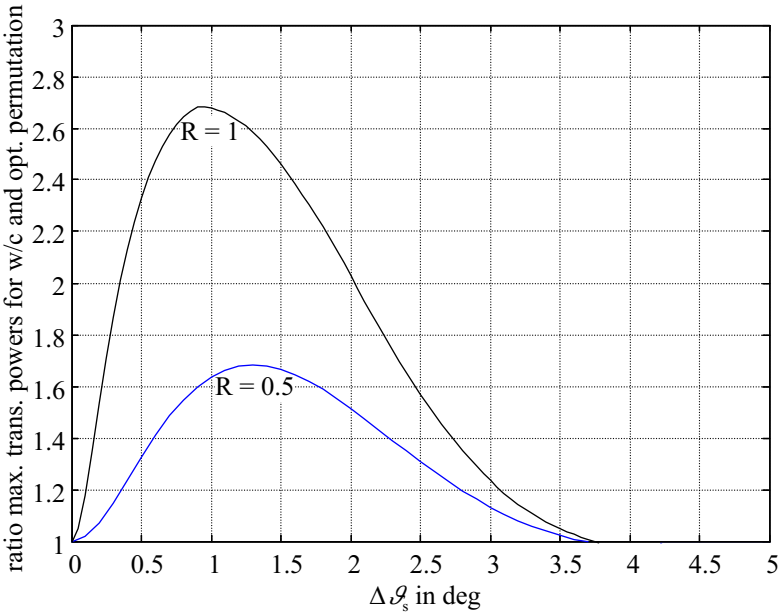


Figure 5.5: Ratio (linear) of the maximal transmit power occurring for optimal and worst-case decoding order as depicted in Fig. 5.3. For $R = 1$ the maximal transmit power for the worst-case decoding order is approx. 2.7 times (equivalent to 4 dB) higher than for the optimal decoding order. For $R = 0.5$ this factor is only approx. 1.7 at maximum.

In the following we will restrict for successive decoding to the optimal decoding order $(4, 2, 6, 7, 5, 3, 1)$.

5.1.2 Comparison Independent and Successive Decoding

We will turn now to a comparison of independent and successive decoding.

Fig. 5.6 shows the element (input) SNR that are required to achieve rate $R = 1$ for all sources by independent and successive decoding, respectively, in the AWGN channel (no fading). The transmit powers for independent decoding are computed as described in Sec. 4.4.1.

As expected, the required γ_m for independent and successive decoding are the same if the angular separation of the source signals is sufficiently large, because then independent channels are obtained. This holds for approximately $\Delta\vartheta_s \geq 2\vartheta_{4.3\text{dB}} = 4.6°$.

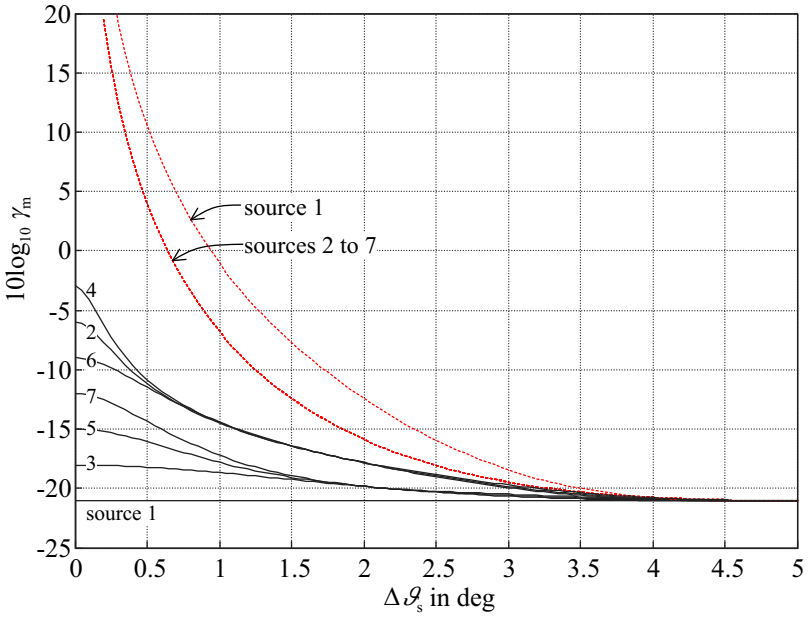


Figure 5.6: Required element SNR γ_m for independent (dashed) and successive decoding (solid). Target rate is $R = 1$, AWGN channel.

For independent decoding, the γ_m tend to infinity for $\Delta\vartheta_s = 0$ (when the classical MAC is obtained), because the rate limit that cannot be achieved with finite γ_m amounts to $R_{\max}^{\text{CMAC}} = 0.22$ bit/channel use, according to (4.18) with $M = 7$.

The according EIRP show, of course, generally the same dependency on angular separation $\Delta\vartheta_s$ (cf. Fig. 5.7), but the EIRP of the outer 6 sources grow slightly with increasing $\Delta\vartheta_s$ as at the same time the slant range, and, therefore, the free-space basic transmission loss increase as well. Considering the maximal EIRP 6.8 dB and 12 dB, respectively, that are specified for the ICO terminals (cf. Tab. 3.1), the minimal separation between the sources that can be achieved for independent decoding is approx. $\Delta\vartheta_s = 1°$ and $0.75°$, respectively. For successive decoding, the required EIRP are always below the possible maximum values.

This changes when the target rate is $R = 2$ (cf. Fig. 5.8). Then also for successive decoding there is a minimum admissible angular separation of approx. $\Delta\vartheta_s = 0.5°$ and $0.25°$ in order to obtain EIRP below the maximum of 6 dB and 12 dB, respectively. For independent decoding the minimum separation is approx. $\Delta\vartheta_s = 1.75°$ and $1.3°$.

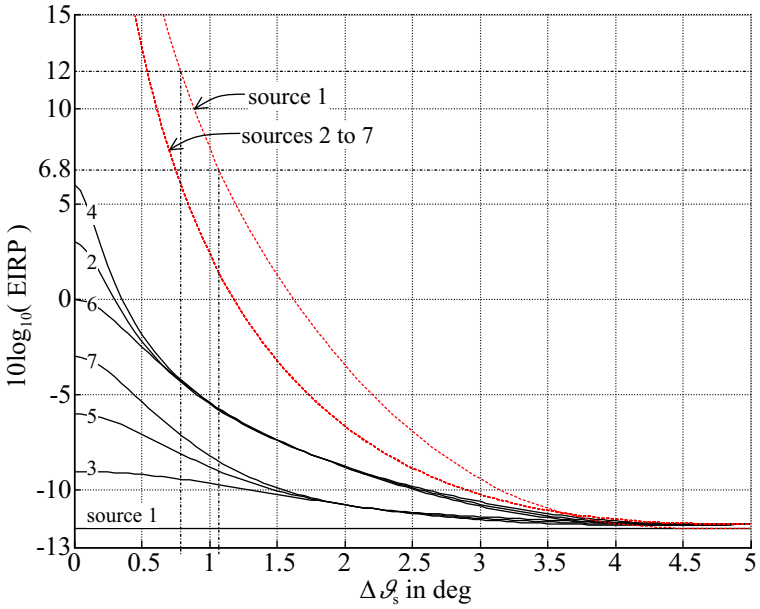


Figure 5.7: Required EIRP for independent (dashed) and successive decoding (solid). Target rate is $R = 1$, AWGN channel. Further, the maximal EIRP 6.8 dB and 12 dB, respectively, for the ICO scenario are plotted, indicating that the minimal separation for independent dec. is approx. $\Delta\vartheta_{\mathrm{s}} = 1°$ and $0.75°$, respectively. For successive dec. the EIRP are always below the maximal EIRP.

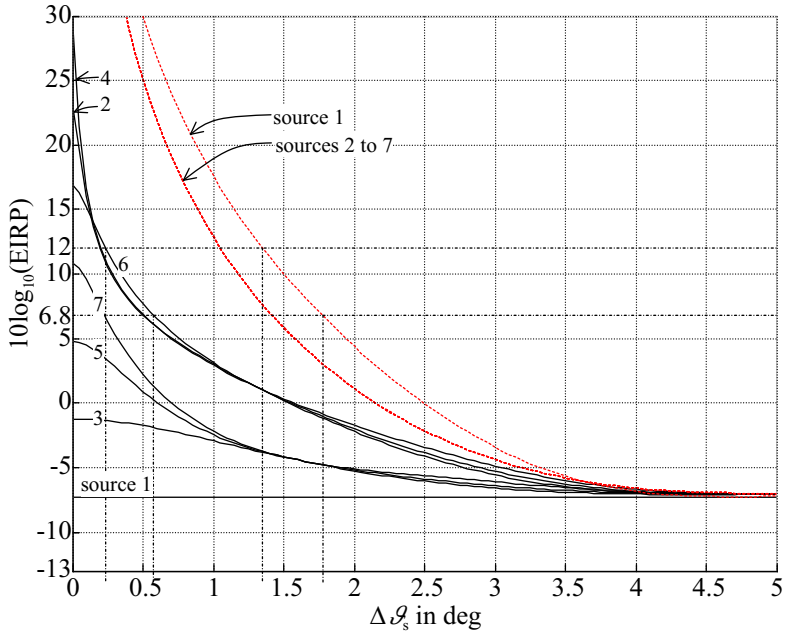


Figure 5.8: Required EIRP for independent (dashed) and successive decoding (solid). Target rate is $R = 2$, AWGN channel. Further, the maximal EIRP 6.8 dB and 12 dB, respectively, for the ICO scenario are plotted, indicating that the minimal separation for independent dec. is approx. $\Delta\vartheta_{\mathrm{s}} = 1.75°$ and $1.3°$, respectively, and for successive dec. $\Delta\vartheta_{\mathrm{s}} = 0.5°$ and $0.25°$, respectively.

Tab. 5.1 shows the minimal angular separation for values of $R = 1, 2, 3, 4, 5$, from which it can be seen that the advantage of successive decoding vanishes as the target rate and, in consequence,

	Minimum separation $\Delta\vartheta_s$			
	Independent Dec.		Successive Dec.	
Target rate R	max. EIRP 6.8 dB	max. EIRP 12 dB	max. EIRP 6.8 dB	max. EIRP 12 dB
1	$1°$	$0.75°$	$0°$	$0°$
2	$1.75°$	$1.3°$	$0.5°$	$0.25°$
3	$2.2°$	$1.7°$	$1.4°$	$0.75°$
4	$2.6°$	$2°$	$2.3°$	$1.5°$
5	$3.1°$	$2.4°$	$2.9°$	$2.1°$

Table 5.1: Minimal angular separation $\Delta\vartheta_s$.

the transmit powers increase. Clearly, this is due to the increasing interference deteriorating the sources being decoded first. Significantly smaller angular separation (approx. 0.5 to 0.3 times smaller) can be achieved with successive decoding for rates $R < 3$.

Finally, Rice fading with $c_R = 5$ dB is considered (cf. Fig. 5.9), which is a reasonable value for the low gain terminal antennas employed in the ICO system (cf. Sec. 2.2.1).

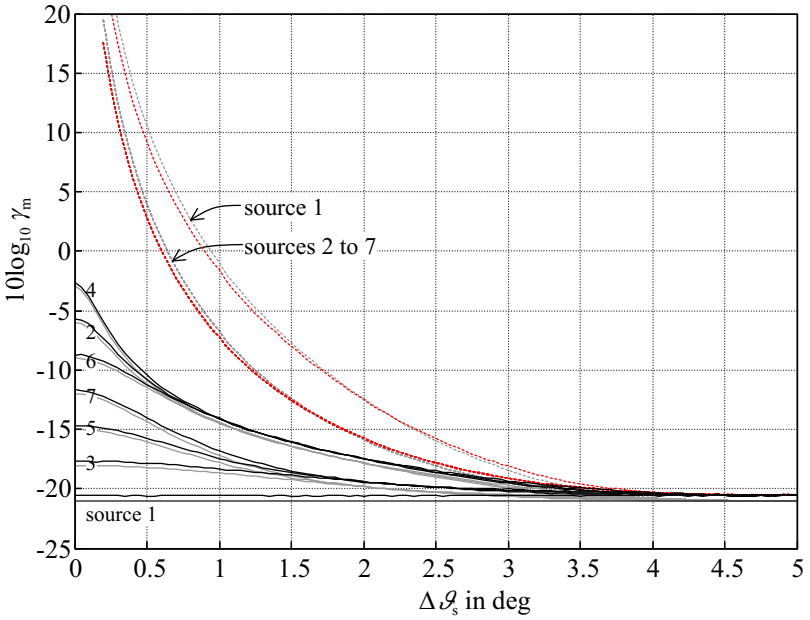


Figure 5.9: Required element SNR γ_m for independent (dashed) and successive decoding (solid). Target rate is $R = 1$, Rice fading channel with $c_R = 5$ dB. Also shown are for comparison the required element SNR for the AWGN channel (grey).

It can be stated that the results for Rice fading with $c_R = 5$ dB are very similar to the AWGN case. However, as expected the transmit power for independent decoding are slightly reduced in the fading channel (cf. Sec. 4.3.2) for angular separation $\Delta\vartheta_s < 2°$, but this effect is very small and negligible.

5.2 Receiver Structures with Fixed Beamforming and with Adaptive Beam Space Processing

In the preceding section we have considered a receiver which employs element space beamforming. In Chap. 2 the concept of spot beams that are created with fixed beamforming was introduced, and finally it was pointed out in Sec. 4.7 that adaptive beam space beamforming subsequent to fixed beamforming can in principle achieve the same rates as it is possible with element space.

In Sec. 4.4 and 4.5, where we have dealt with the receiver for the fading vector multiple-access channel without fixed beamforming, it was pointed out that both for successive decoding, as well as for independent decoding, optimum beamforming to maximise the individual SINR is required to achieve the respective maximal rates.

Also for the receiver with fixed beamforming, optimum beam space processing prior to decoding (successive or independent) is required to exploit the spatial separability of the source signals in the optimal way, maximising the SINR.

The fixed BFN is designed such that the spot beams and the resulting cells, respectively, cover completely the service area of the satellite. Without additional adaptive beam space processing a single source usually cannot utilise the maximum array gain, the worst case being encountered when the source is located at the edge of the corresponding fixed cell. Further, spatial filtering of interference from co-channel users is, of course, less efficient without adaptive beam space processing. However, the complexity of adaptive beam space processing can eventually be saved at the cost of both a worse spatial separation of the sources and a usually less than optimal array gain (edge of cell loss), on the condition that the rates achievable with fixed beamforming alone are sufficient.

In fact, fixed beamforming without adaptive beam space processing together with subsequent independent decoding is the approach followed in present satellite communication systems (cf., e.g., [LWJ00, Gay02]).

Especially in the satellite environment the combination of a fixed beamformer with subsequent adaptive beam space beamforming is particularly attractive, due to the following reasons.

Firstly, one can imagine a communications satellite that covers the service area with a cell pattern employing fixed beamforming as described in Chap. 2 to serve basic traffic load, and adaptive processing is only employed to deal more efficiently with increased interference caused by the higher number of sources during peak traffic hours [Jah99]. In this way adaptive beamforming is used to trade the higher bandwidth requirement during peak traffic hours for the higher computational complexity that comes with adaptive beamforming.

Secondly, it was already indicated in Sec. 4.7.2 that beam space processing is lossless if the same number of beams as there are array elements is used. However, it was also mentioned that, if fewer beams are used for adaptive beam space processing, the losses may be negligible, depending on the particular source distribution. Therefore, fixed beamforming allows *partial adaptive beam space processing* [MM80, LL96], i.e. in the adaptive beamforming stage it may be sufficient to use the signals of only a few fixed beams for adaptive processing. This may reduce significantly the computational complexity of the adaptive beamforming stage with respect to fully adaptive element or beam space beamforming.

We will in the following investigate the performance of the receiver employing fixed beamforming, where in particular the following scenarios are considered (we introduce here some abbreviations to address the various receiver types in a simple way, also cf. Fig. 5.10):

Fixed beamforming, followed by

- adaptive beamforming (i.e. beam space beamforming using the output of the fixed beamformer) followed by independent decoding ("AB+ID", p. 131 f).
- optimal joint decoding, i.e. optimal beamforming with the output of the fixed beamforming networks with subsequent successive decoding ("AB+SD", p. 132 f).
- independent decoding ("FB+ID", p. 141 f). This is today's usual approach.
- successive decoding, without an optimal beamforming stage ("FB+SD", p. 145 f).

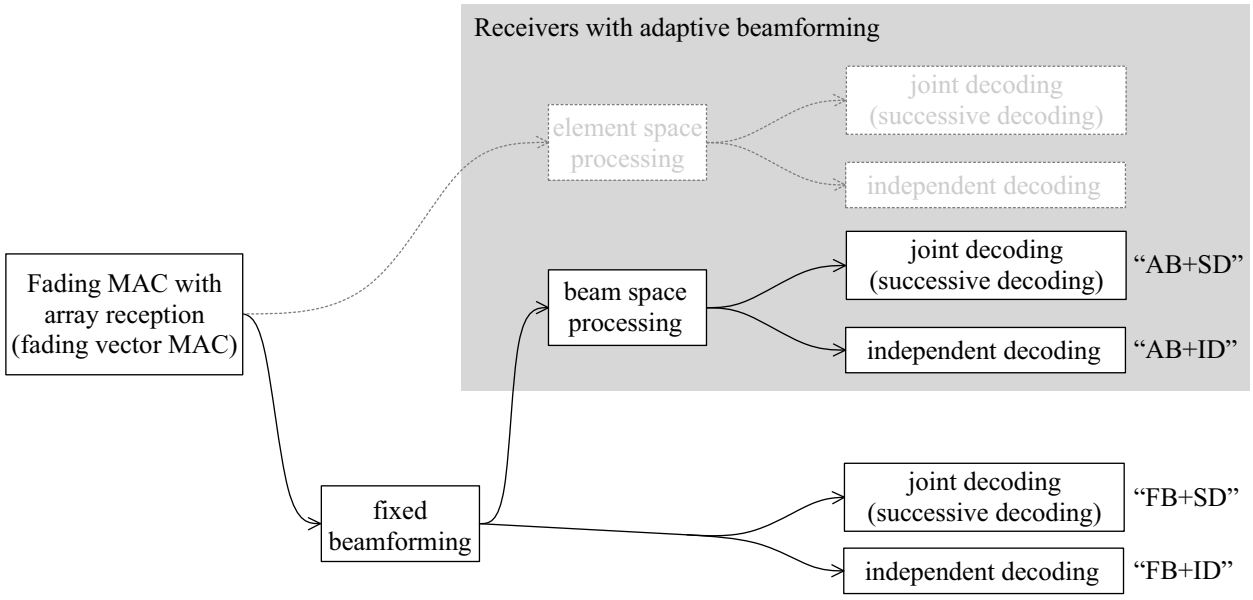


Figure 5.10: Outline of the receiver/decoder options that will be considered in the remainder of this chapter. AB: adaptive beamforming, FB: only fixed beamforming, SD: successive decoding, ID: independent decoding.

Of particular interest is the comparison of *adaptive beamforming with independent decoding* (AB+ID) and *fixed beamforming with successive decoding* (FB+SD) with the optimal receiver employing both adaptive beamforming and successive decoding (AB+SD), because both AB+ID and FB+SD options rely purely on one or the other of the two interference mitigation techniques, i.e. the implementation complexity is reduced with respect to the optimal receiver AB+SD.

5.2.1 Receivers with Optimal Beam Space Processing: "AB+ID" and "AB+SD"

In Sec. 4.7.2 it was pointed out that the covariance matrix $\mathbf{K}_{\mathbf{r}_\mathrm{B}}$ of the fixed BFN output vector $\mathbf{r}_\mathrm{B}$ has in principle the same structure as for the receiver without fixed beamforming (cf. (4.72) and

(4.73)). Consequently, the description of the receiver with fixed beamforming in terms of region of achievable rates, power region, and successive decoding is simply obtained from the equations for element space processing presented in Sec. 4.4 and 4.5 by replacing the element steering vectors $\mathbf{a}_m$ with the beam space steering vectors $\mathbf{a}_{\mathrm{B},m}$ as defined in (4.73) and, further, using the beam space noise covariance matrix from (4.74):

$$\mathbf{K}_{\mathbf{n}_B} = 2\sigma_{\mathrm{n}}^2 \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{W}_{\mathrm{B}}.$$

Therefore, we can avoid in the following the lengthy derivation of the required equations.

Finally, in the following it will always be assumed that the covariance matrix $\mathbf{K}_{\mathbf{r}_{\mathrm{B}}}$ of $\mathbf{r}_{\mathrm{B}}$ is invertible, i.e. $\det\left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{B}}\right) \neq 0$, which is equivalent to $\mathrm{rank}(\mathbf{W}_{\mathrm{B}}) = L_{\mathrm{B}} \leq L$. Then the L_{B}-dimensional PDF of the fixed BFN output vector $\mathbf{r}_{\mathrm{B}}$ exists and is defined by (B.3), in consequence also mutual information is defined.

Independent Decoding ("AB+ID")

After fixed beamforming and optimal beam space processing for each of the M source signals, each source signal is decoded by single-user decoding. The receiver structure for this case is schematically depicted in Fig. 5.11.

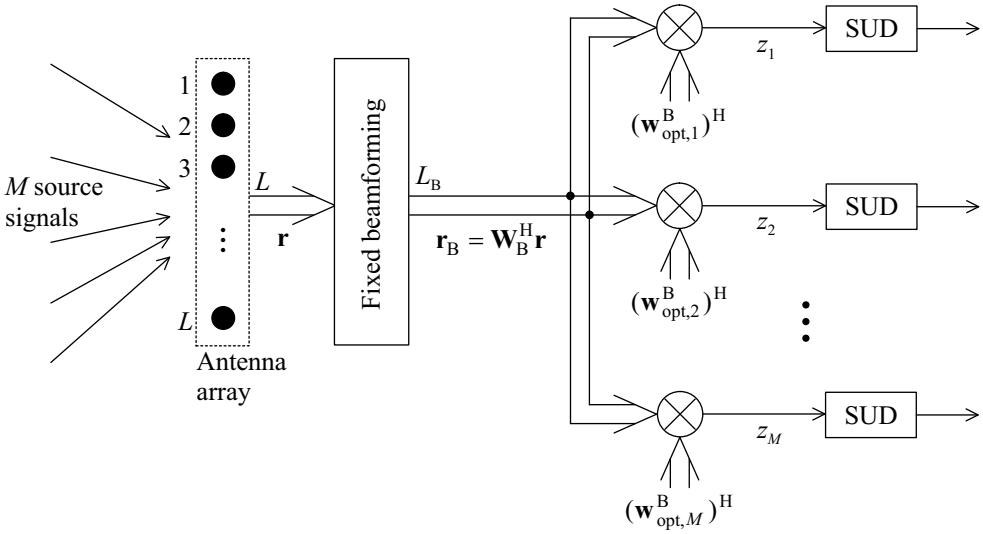


Figure 5.11: Block diagram of the receiver with fixed beamforming, optimal beam space processing, and subsequent single-user decoding (AB+ID). With $\mathbf{W}_{\mathrm{B}}$ being a $L \times L_{\mathrm{B}}$ matrix, L_{B} fixed beams are created from the L array element output signals. Subsequently, optimal beam space processing, maximising the respective SINR, and single-user decoding for each of the M source signals is performed. (Also cf. Fig. 4.9, showing the receiver employing optimal element space beamforming with independent decoding.)

In this case the expression for mutual information between source signal s_m and the fixed BFN output vector $\mathbf{r}_{\mathrm{B}}$ reads

$$\mathcal{I}(S_m; \boldsymbol{R}_{\mathrm{B}}) = \mathcal{H}(\boldsymbol{R}_{\mathrm{B}}) - \mathcal{H}(\boldsymbol{R}_{\mathrm{B}}|S_m), \tag{5.1}$$

and the rates achievable by independent decoding are provided by (cf. (4.38))

$$\begin{aligned} R_m &\leq \max \mathcal{I}(S_m; \boldsymbol{R}_{\mathrm{B}}) \\ &= \mathrm{E}\left\{ \log_2 \left(1 + P_m \left| H_m \right|^2 \mathbf{a}_m^{\mathrm{H}} \mathbf{W}_{\mathrm{B}} \left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{K}_{\mathbf{v}}(\mathcal{I}_m) \mathbf{W}_{\mathrm{B}} + 2\sigma_{\mathrm{n}}^2 \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{W}_{\mathrm{B}} \right)^{-1} \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{a}_m \right) \right\}_{\boldsymbol{H}}, \\ \mathcal{I}_m &= \mathcal{M} \setminus m, \end{aligned} \tag{5.2}$$

with $\mathcal{I}_m$ being the set of source indices interfering with source m.

By the analogy of element space and beam space beamforming demonstrated in Sec. 4.7.1, it is evident that the term

$$\Gamma_{\mathrm{opt},m}^{\mathrm{B}}(\mathbf{h}) = \frac{P_m \left| h_m \right|^2}{2\sigma_{\mathrm{n}}^2} \mathbf{a}_m^{\mathrm{H}} \mathbf{W}_{\mathrm{B}} \left(\frac{1}{2\sigma_{\mathrm{n}}^2} \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{K}_{\mathbf{v}}(\mathcal{I}_m) \mathbf{W}_{\mathrm{B}} + \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{W}_{\mathrm{B}} \right)^{-1} \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{a}_m \tag{5.3}$$

in (5.2) denotes the maximal SINR achievable by optimal beam space processing (compare (5.2) with the equation for the achievable rates for independent decoding without fixed beamforming (4.16)).

The corresponding optimal beam space weight vector $\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}$ for source m is given by

$$\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}(\mathbf{h}) = \beta \left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{K}_{\mathbf{v}}(\mathcal{I}_m) \mathbf{W}_{\mathrm{B}} + 2\sigma_{\mathrm{n}}^2 \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{W}_{\mathrm{B}} \right)^{-1} \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{a}_m, \tag{5.4}$$

where β is an arbitrary scalar (cf. (3.89)).

Rates R_m can be easily determined for given powers P_m based on the relation between powers P_m and achievable rates R_m given by (5.2). For required rates R_m, power allocation is analogous to Sec. 4.4.1, i.e. the transmit powers $p_m = \mu_m P_m$ are computed by an iterative algorithm.

Successive Decoding ("AB+SD")

The receiver structure for this case is shown in Fig. 5.12, where it is assumed to simplify notation that the source indices are rearranged to yield decoding order $(1, 2, 3, \ldots, M)$.

The mutual information between source signal vector $\mathbf{s}$ and fixed BFN output vector $\mathbf{r}_{\mathrm{B}}$ for a subset of sources $\mathcal{S} \subseteq \mathcal{M}$ is given by (cf. the corresponding equation for element space processing in Sec. 4.6)

$$\begin{aligned} \mathcal{I}(S_m, m \in \mathcal{S}; \boldsymbol{R}_{\mathrm{B}} | S_j, j \in \bar{\mathcal{S}}) &= \mathcal{H}(\boldsymbol{R}_{\mathrm{B}} | S_j, j \in \bar{\mathcal{S}}) - \mathcal{H}(\boldsymbol{R}_{\mathrm{B}} | \boldsymbol{S}) \\ &= \mathcal{H}(\boldsymbol{R}_{\mathrm{B}} | S_j, j \in \bar{\mathcal{S}}) - \mathcal{H}(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \boldsymbol{N}), \\ & \mathcal{S} \subseteq \mathcal{M}, \bar{\mathcal{S}} = \mathcal{M} \setminus \mathcal{S}. \end{aligned} \tag{5.5}$$

With the equation for the covariance matrix of $\mathbf{r}_{\mathrm{B}}$ for a subset $\mathcal{S}$ according to (4.75), the region of achievable rates for the receiver employing a fixed BFN and joint decoding is readily given by

$$\begin{aligned} \sum_{m \in \mathcal{S}} R_m &\leq \max \mathcal{I}(S_m, m \in \mathcal{S}; \boldsymbol{R}_{\mathrm{B}} | S_j, j \in \bar{\mathcal{S}}) \\ &= \mathrm{E}\left\{ \log_2 \left(\det \left(\mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{K}_{\mathbf{v}}(\mathcal{S}) \mathbf{W}_{\mathrm{B}} \left(2\sigma_{\mathrm{n}}^2 \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{W}_{\mathrm{B}} \right)^{-1} + \mathbf{I}_{L_{\mathrm{B}}} \right) \right) \right\}_{\boldsymbol{H}}, \\ \mathcal{S} &\subseteq \mathcal{M}, \end{aligned} \tag{5.6}$$

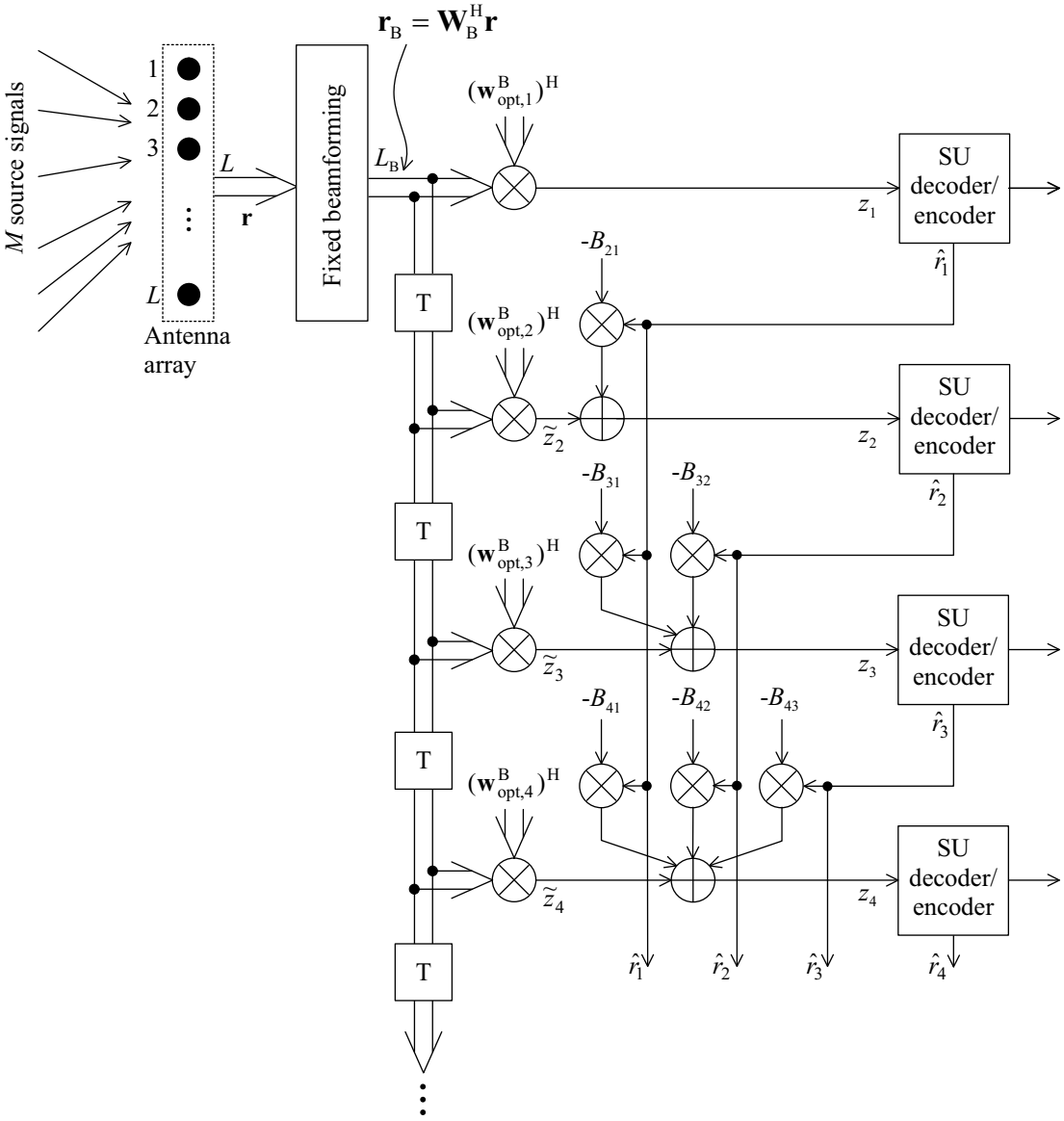


Figure 5.12: Block diagram of a receiver with fixed beamforming, optimal beam space processing, and subsequent successive decoding (AB+SD; to simplify notation and without loss of generality it is assumed that the source indices correspond to the decoding order). The delays T are equal to the decoding delay in one stage of the successive decoder and are introduced to synchronise the symbol streams in the different decoder stages.

which is similar to (4.44) representing the achievable rates for the receiver without fixed beamforming, where the only difference between (5.6) and (4.44) is the covariance matrix of the receive signal vectors $\mathbf{r}_B$ (cf. (4.75)) and $\mathbf{r}$ (cf. (3.78)), respectively.

Using either the chain rule of mutual information or, equivalently, the definition of the vertices of a polymatroid (recall that the region of achievable rates must be a polymatroid, cf. Sec. 4.5.1) the rates achievable by successive decoding for a permutation $\boldsymbol{\pi}_v$ on the set $\mathcal{M} = \{1, 2, \ldots, M\}$ are

for $m = 1, 2, \ldots, M$ given by

$$R_{\pi_v(m)} = \mathrm{E}\left\{\log_2\left(1 + P_{\pi_v(m)}\left|H_{\pi_v(m)}\right|^2 \mathbf{a}^{\mathrm{H}}_{\pi_v(m)}\mathbf{W}_{\mathrm{B}} \ldots \right.\right. \\ \left.\left. \ldots \left(\mathbf{W}^{\mathrm{H}}_{\mathrm{B}}\mathbf{K}_{\mathbf{v}}(\mathcal{S}_{v,m-1})\mathbf{W}_{\mathrm{B}} + 2\sigma^2_{\mathrm{n}}\mathbf{W}^{\mathrm{H}}_{\mathrm{B}}\mathbf{W}_{\mathrm{B}}\right)^{-1}\mathbf{W}^{\mathrm{H}}_{\mathrm{B}}\mathbf{a}_{\pi_v(m)}\right)\right\}_{\boldsymbol{H}}, \tag{5.7}$$

with the set $\mathcal{S}_{v,m} = \{\pi_v(1), \pi_v(2), \ldots, \pi_v(m)\}$, such that $\mathcal{S}_{v,m-1}$ contains the indices of the sources not yet decoded and, hence, interfering with source $\pi_v(m)$ (remember that the decoding order is $(\pi_v(M), \pi_v(M-1), \ldots, \pi_v(1))$, cf. Sec. 4.3.3).

Again, the rate $R_{\pi_v(m)}$ that is achievable for source $\pi_v(m)$ is that of a receiver performing optimal beam space processing, where only sources in the set $\mathcal{S}_{v,m-1}$ contribute to interference (cf. the achievable rates with optimal beam space processing and independent decoding in (5.2), also cf. Sec. 4.6 where the rates achievable with optimal element space processing have been provided).

It is straight forward to derive the feedforward and feedback filter matrices according to Sec. 4.6.1, and we therefore set aside to explicitly write them here.

5.2.2 Partially Adaptive Beam Space Processing

In Sec. 4.7.2 it was pointed out that fixed beamforming with subsequent optimal adaptive beam space processing leaves the information at the receiver about the transmitted source symbols unchanged, if the transformation from element space to beam space, realised by the fixed BFN, is reversible. It was demonstrated that the transformation from element space to beam space is reversible regardless of the particular source distribution, only if at least L beams with linearly independent weight vectors are created from the L-element array output vector.

Further, the computational complexity of the adaptive processing depends, besides the chosen algorithm, strongly on the dimension of the beamforming input vector, which is L for adaptive element space beamforming and $L_{\mathrm{B}} \leq L$ for beam space beamforming, respectively [LL96]. Hence, it seems that there is no advantage in introducing a fixed BFN, as long as the same number of beams as there are array elements are used for adaptive beam space processing.

Therefore, the case being of particular interest for further investigations for the considered satellite scenarios, is encountered if only a subset of the L_{B} beams is used for adaptive beam space processing.

This is referred to as *partial adaptivity* [MM80, LL96], because not all L_{B} available degrees of freedom are used for adaptive processing. It was already pointed out in Sec. 4.7.2 that in this case equality of mutual information $\mathcal{I}(\boldsymbol{S}; \boldsymbol{R})$ and $\mathcal{I}(\boldsymbol{S}; \boldsymbol{R}_{\mathrm{B}})$ is in general not given (cf. (4.77)), and the loss of information depends strongly on the particular source distribution for a given subset of beams.

Partial adaptivity offers the possibility of a trade-off between reduction of computational complexity in the adaptive beam space processing at the receivers side on the one hand, and, on the other hand, the performance degradation due to partially adaptive beam space processing of the source signals with respect to full adaptivity, which results in less efficient spatial filtering of interfering source signals and increased required transmit powers for the sources.

The following definitions will be used in this section to describe partially adaptive beamforming. Let

$$\mathcal{B}_m \subseteq \{1, 2, \ldots, L_{\mathrm{B}}\} \tag{5.8}$$

denote the set that contains the indices of the beams that are used for source m for partially adaptive beam space beamforming and, further, let

$$L_{\mathrm{B},m} = |\mathcal{B}_m| \leq L_{\mathrm{B}} \tag{5.9}$$

denote the according number of beams. Selecting the beams according to the set $\mathcal{B}_m$ can be described by multiplying the fixed beamforming weight matrix $\mathbf{W}_{\mathrm{B}}$ with a selection matrix $\mathbf{S}_m$ that chooses the columns of $\mathbf{W}_{\mathrm{B}}$ according to the beam indices contained in $\mathcal{B}_m$:

$$\mathbf{W}_{\mathrm{B},m} = \mathbf{W}_{\mathrm{B}}\mathbf{S}_m, \tag{5.10}$$

where the selection matrix $\mathbf{S}_m$ ($L_{\mathrm{B}} \times L_{\mathrm{B},m}$) is obtained from the $L_{\mathrm{B}} \times L_{\mathrm{B}}$ unit matrix $\mathbf{I}_{L_{\mathrm{B}}}$ by deleting the $(L_{\mathrm{B}} - L_{\mathrm{B},m})$ columns according to the beam indices not contained in $\mathcal{B}_m$.

From the definition of the output SINR for optimal beam space beamforming (5.3), it is easy to provide the expression for the output SINR for source m for partially adaptive beam space beamforming by replacing $\mathbf{W}_{\mathrm{B}}$ with $\mathbf{W}_{\mathrm{B},m}$ given by (5.10).

This yields for the output SINR $\Gamma^{\mathrm{B}}_{\mathrm{PA},m}(\mathbf{h})$ that is achievable at maximum if only those beams contained in $\mathcal{B}_m$ are employed

$$\Gamma^{\mathrm{B}}_{\mathrm{PA},m}(\mathbf{h}) = \gamma_m \left|h_m\right|^2 \mathbf{a}_m^{\mathrm{H}}\mathbf{W}_{\mathrm{B}}\mathbf{S}_m \left(\frac{1}{2\sigma_{\mathrm{n}}^2}\mathbf{S}_m^{\mathrm{T}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{K}_{\mathbf{v}}(\mathcal{I}_m)\mathbf{W}_{\mathrm{B}}\mathbf{S}_m + \mathbf{S}_m^{\mathrm{T}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{B}}\mathbf{S}_m\right)^{-1} \mathbf{S}_m^{\mathrm{T}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{a}_m, \tag{5.11}$$

where

$$\mathbf{K}_{\mathbf{n}_{\mathrm{B}}}(\mathcal{B}_m) = 2\sigma_{\mathrm{n}}^2\mathbf{S}_m^{\mathrm{T}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{B}}\mathbf{S}_m \tag{5.12}$$

is the beam space noise covariance matrix, considering only the beams contained in $\mathcal{B}_m$. The term $W_{\mathrm{B}}^{\mathrm{H}}\mathbf{a}_m$ was introduced in Sec. 4.7.1 as the beam space steering vector of source m.

The output SINR according to (5.11) depends strongly on the beam selection matrix $\mathbf{S}_m$. If the number $L_{\mathrm{B},m}$ of beams employed for partially adaptive beamforming is given, then, of course, those $L_{\mathrm{B},m}$ out of L_{B} beams should be included in $\mathcal{B}_m$ for that the output SINR is maximised. Vice versa, if a tolerable degradation due to partial adaptivity with respect to full adaptive beamforming is given, we want to know the minimum required number of beams $L_{\mathrm{B},m}$. It seems not trivial to provide the $\mathcal{B}_m$ for given $L_{\mathrm{B},m}$ for that optimality in that or the other sense can be strictly proven. However, it can be expected that those L_{B} beams should be employed that show the highest spot beam gains with respect to the sources.

A central point is the eventual difference in the number of beams required for partially adaptive beamforming to achieve a certain percentage of the optimal output SINR between partially adaptive beamforming with orthogonal beams and non-orthogonal beams, respectively. We would expect at a first glance that fewer beams suffice for non-orthogonal beams as adjacent beams overlap more than it is the case for orthogonal beams (cf. Fig. 3.25 and Fig. 4.21). On the other hand, noise in different beams is correlated for non-orthogonal beams (this holds in particular for adjacent beams)

and we are therefore interested in the influence of the correlated noise on the required number of beams.

Therefore, we have to address two questions. Firstly, how many beams are required for partially adaptive beamforming to yield a performance close enough to that of the full adaptive beamformer, and, secondly, whether there is a significant difference between partially adaptive beamforming using orthogonal beams and using non-orthogonal beams. These questions shall be answered in the following by numerical evaluation of (5.11) both for orthogonal beams as well as for non-orthogonal beams.

To simplify the discussion here, we will only consider the single-source case.

Orthogonal Beams

For orthogonal beams it holds (cf. (4.87))

$$\mathbf{K}_{\mathbf{n}_\mathrm{B}} = 2\sigma_\mathrm{n}^2 \mathbf{S}_m^\mathrm{T} \mathbf{W}_\mathrm{B}^\mathrm{H} \mathbf{W}_\mathrm{B} \mathbf{S}_m = 2\sigma_\mathrm{n}^2 L \mathbf{I}_{L_{\mathrm{B},m}}. \tag{5.13}$$

Then, for orthogonal beams and without interference, the expression for the output SINR (5.11) yields

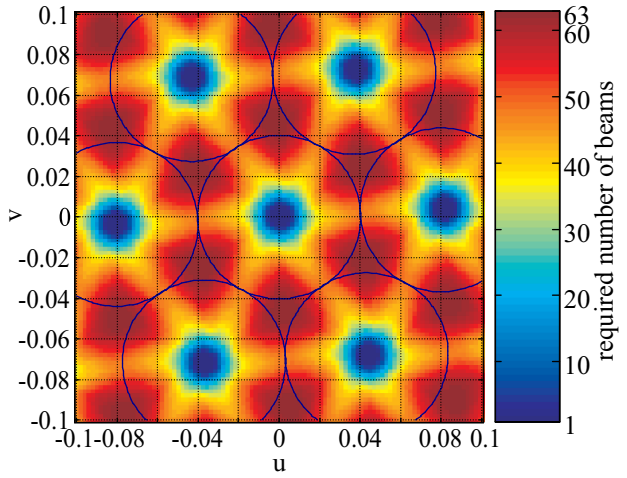
$$\Gamma_{\mathrm{PA},m}^\mathrm{B}(h_m) = \gamma_m \left|h_m\right|^2 \frac{1}{L} \mathbf{a}_m^\mathrm{H} \mathbf{W}_\mathrm{B} \mathbf{S}_m \mathbf{S}_m^\mathrm{T} \mathbf{W}_\mathrm{B}^\mathrm{H} \mathbf{a}_m = \gamma_m \left|h_m\right|^2 \frac{1}{L} \sum_{b \in \mathcal{B}_m} \left|\mathbf{a}_m^\mathrm{H} \mathbf{w}_{\mathrm{B},b}\right|^2, \tag{5.14}$$

and for given number of beams $L_{\mathrm{B},m}$ the SNR $\Gamma_{\mathrm{PA},m}^\mathrm{B}(h_m)$ is maximised if it holds, possibly after reordering, that

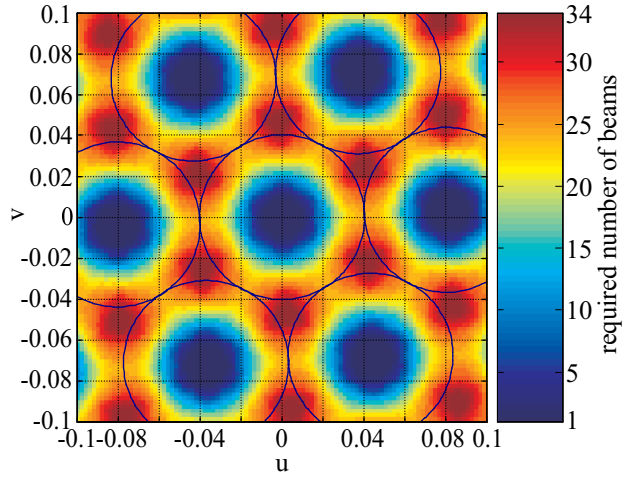
$$\left|\mathbf{a}_m^\mathrm{H} \mathbf{w}_{\mathrm{B},1}\right|^2 \geq \left|\mathbf{a}_m^\mathrm{H} \mathbf{w}_{\mathrm{B},2}\right|^2 \geq \cdots \geq \left|\mathbf{a}_m^\mathrm{H} \mathbf{w}_{\mathrm{B},L_\mathrm{B}}\right|^2 \tag{5.15}$$

Fig. 5.13 shows for the ICO scenario the number of beams required to achieve a $\Gamma_{\mathrm{PA},m}^\mathrm{B}(\mathbf{h})$ for a single source according to (5.14) that amounts to 99%, 95%, and 90%, respectively, of the optimal $\Gamma_{\mathrm{opt},m}^\mathrm{B}(\mathbf{h})$ achievable when all 127 DFT beams are used.

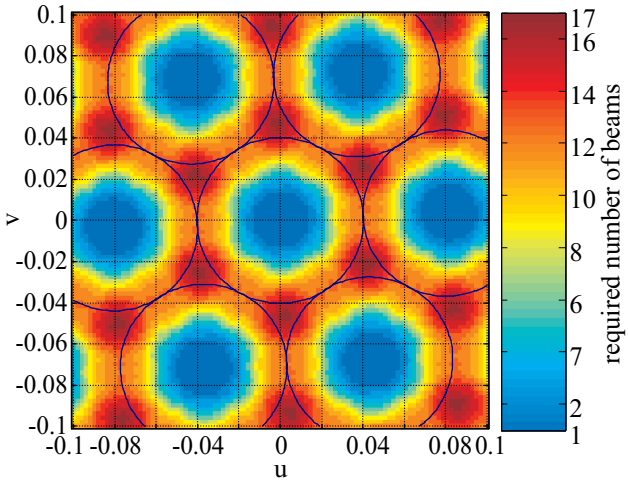
To achieve 99% of the optimal SINR 63 beams have to be employed for partial adaptivity at maximum, while for 95% and 90% 34 beams and 17, respectively, are sufficient.



(a) 99%



(b) 95%



(c) 90%

Figure 5.13: Number of DFT beams required for a single source located at (u, v) to achieve the indicated percentage of the optimal output SNR achieved with optimal beam space beamforming using all 127 orthogonal DFT beams.

Non-Orthogonal Beams

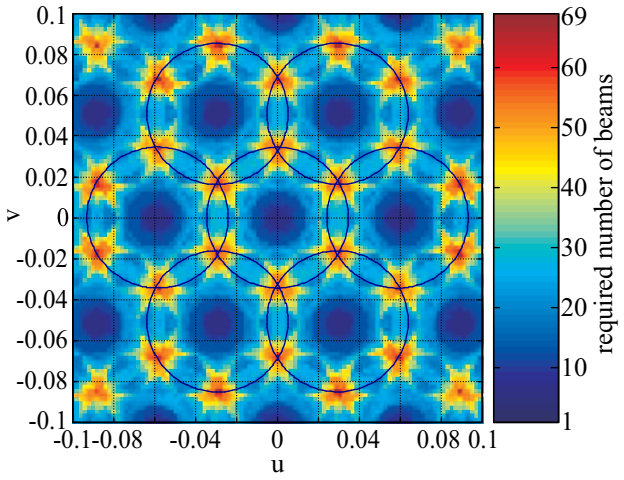
As pointed out before one could expect that fewer beams are required for partial adaptivity employing non-orthogonal beams, because of the larger overlap of adjacent beams (cf. Fig. 3.25 and Fig. 4.21, respectively), than it is required for orthogonal beams in order to achieve the same performance.

On the other hand, we know that the covariance matrix $\mathbf{K}_{\mathbf{n}_\mathrm{B}}(\mathcal{B}_m)$ is not diagonal for non-orthogonal beams and, in consequence, the noise in adjacent beams is correlated, such that the gain achieved through combining the receive signals of different beams is reduced. Indeed numerical evaluation of (5.11) for non-orthogonal beams (the spot beam pattern of the ICO scenario is assumed) reveals that a similar number of beams for partial adaptivity is required in the worst case to obtain the same performance as with orthogonal beams (cf. Fig. 5.14): to achieve 99% of the optimal SINR 69 beams have to be employed at maximum, whereas for 95% and 90% 53 beams and 6, respectively, are sufficient.

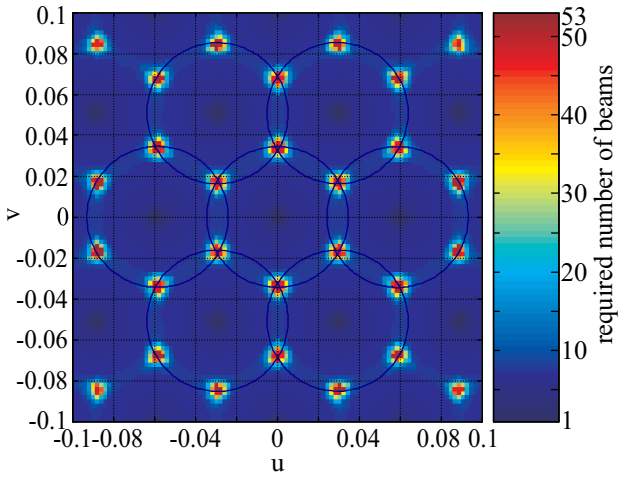
Finally, if the considered source is randomly placed in the coverage area (for the sake of simplicity we assume here a uniform distribution in the u-v-plane) then the required number of beams becomes a random number as well. Fig. 5.15 shows the resulting probabilities that more than a certain number of beams if required to achieve 99%, 95%, and 90% of the optimal SNR. This indicates that with non-orthogonal beams fewer beams are required in the average than for orthogonal beams.

We conclude this section with the following observations:

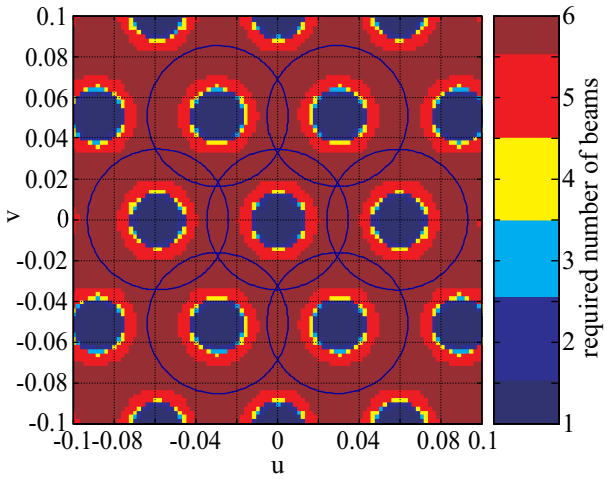
- In the worst-case, partial adaptivity requires approximately the same number of beams for both orthogonal as well as for non-orthogonal beams to achieve similar performance.
- With non-orthogonal beams fewer beams are required in the average.



(a) 99%

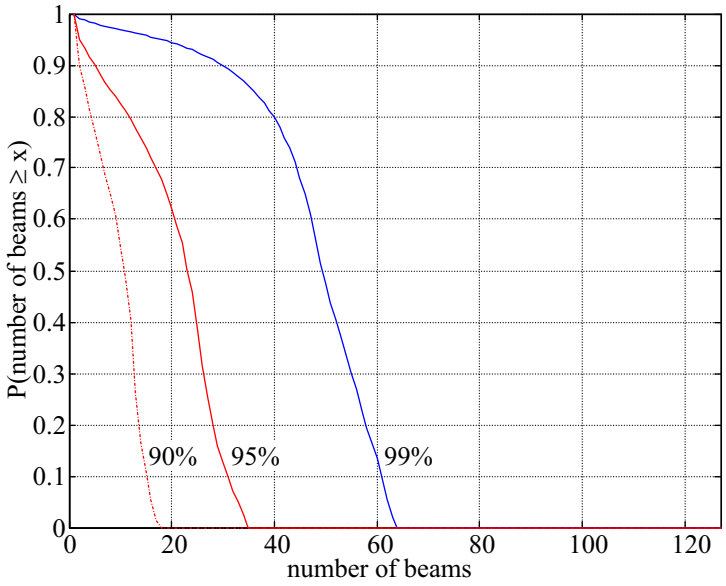


(b) 95%

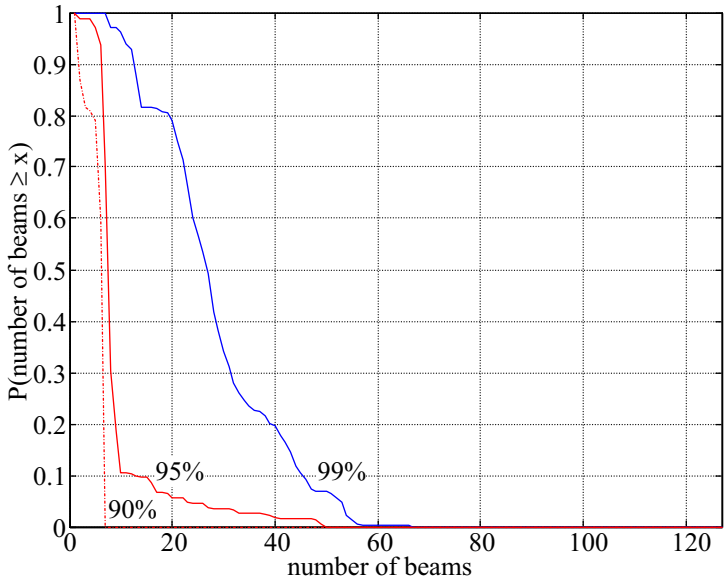


(c) 90%

Figure 5.14: Number of beams required for a single source located at (u, v) to achieve the indicated percentage of the optimal output SNR achieved with optimal beam space beamforming using 127 out of 163 non-orthogonal beams (ICO scenario).



(a) Orthogonal beams.



(b) Non-orthogonal beams (ICO spot beam pattern).

Figure 5.15: Probability that a given number of beams required for partial adaptivity to achieve a given percentage (99%, 95%, and 90%) of the optimal output SNR is exceeded.

5.3 Receiver Structures without Adaptive Beam Space Processing

In this section we will present the equations for achievable rates for a receiver that does not employ any adaptive beam space processing. Like before we will consider independent and successive decoding of the source signals.

Without beam space processing it is required to decode the source signal employing only the receive signal of a single beam. For a source m, we will denote the index of the associated beam with b_m (cf. Fig. 5.16). The output port of the beam-source assignment block for source m contains

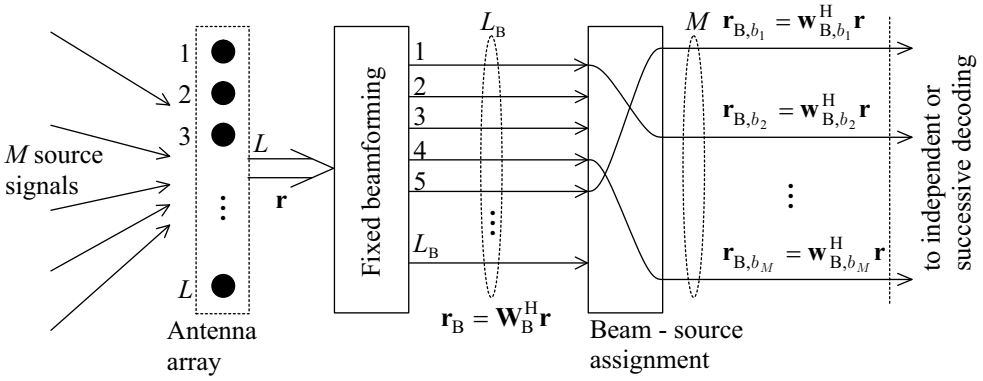


Figure 5.16: Beam-source assignment for fixed beamforming without beam space processing. As an example, the beams assigned to the sources are $b_1 = 5$, $b_2 = 1$, and $b_M = 4$.

only the signal of beam b_m, i.e.

$$r_{\mathrm{B},b_m} = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{r}, \tag{5.16}$$

since the fixed beam with index b_m is associated to column $\mathbf{w}_{\mathrm{B},b_m}$ of the fixed beamforming matrix $\mathbf{W}_{\mathrm{B}}$.

5.3.1 Independent Decoding ("FB+ID")

To decode the signal of a particular source m, the output signal of the respective associated beam port b_m is used as input to the single-user decoder, cf. Fig. 5.17. Obviously, the received symbols z_m at the input to the coherent single-user demodulator/decoder for source m are equal to the signal received in beam b_m, such that

$$z_m = r_{\mathrm{B},b_m} = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{r} = \underbrace{\mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_m \frac{h_m}{\sqrt{\mu_m}} s_m}_{\text{wanted signal}} + \underbrace{\sum_{\substack{i=1 \\ i \neq m}}^{M} \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}} s_i}_{\text{interference}} + \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{n}. \tag{5.17}$$

Such a receiver employing fixed beamforming only, together with single-user decoding is the standard in nowadays satellite systems [LWJ00].

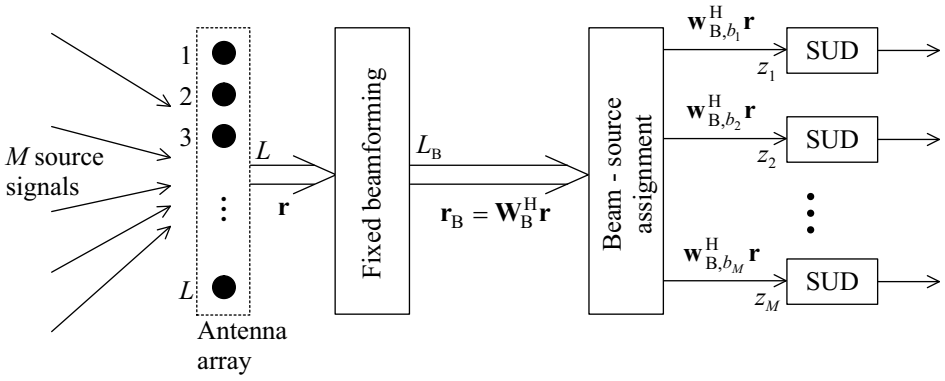


Figure 5.17: Block diagram of a receiver with fixed beamforming and subsequent independent decoding (FB+ID). The beam-source assignment interconnects the output of beam b_m with the single-user decoder of source m. The input to the m-th single-user decoder z_m is the output signal of beam b_m, given by $r_{\mathrm{B},b_m} = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}}\mathbf{r}$, i.e. $z_m = r_{\mathrm{B},b_m}$.

Because no beam space processing is employed, the decoding of a particular source signal m has to rely on the information available at a single beam port b_m, and therefore the expression for mutual information reads in this case

$$\mathcal{I}(S_m; R_{\mathrm{B},b_m}) = \mathcal{H}(R_{\mathrm{B},b_m}) - \mathcal{H}(R_{\mathrm{B},b_m}|S_m), \tag{5.18}$$

(compare this with the expression for mutual information for optimal beam space processing with independent decoding in (5.1), where the fixed BFN output vector $\mathbf{r}_{\mathrm{B}}$ is considered available to retrieve a source signal s_m at the receiver).

Then the achievable rates are bounded by

$$R_m \leq \mathrm{E}\left\{\log_2\left(1 + \frac{P_m\left|H_m\right|^2\left|\mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}}\mathbf{a}_m\right|^2}{\mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}}\mathbf{K}_{\mathbf{v}}(\mathcal{I}_m)\mathbf{w}_{\mathrm{B},b_m} + 2\sigma_{\mathrm{n}}^2\left|\mathbf{w}_{\mathrm{B},b_m}\right|^2}\right)\right\}_{\boldsymbol{H}}, \qquad \mathcal{I}_m = \mathcal{M}\setminus m. \tag{5.19}$$

If the rates R_m are given, then the required powers P_m are calculated from (5.19) using the iterative power allocation algorithm specified in Sec. 4.4.1.

Maximal Achievable Rates

We have seen in Sec. 4.3.4 that for the classical MAC there is a rate limit (cf. (4.18)) beyond which no rate can be achieved likewise for all sources with finite element SNR γ_m.

Also with fixed beamforming there exists such a rate limit, and calculation of this limit for the vector MAC with fixed beamforming will be demonstrated in the following (fading is not considered).

According to (5.19) the output SINR for source m is for the AWGN case given by (using (3.76) and (3.77))

$$\Gamma_{\mathrm{FB},m} = \frac{\gamma_m G_{mm}}{1 + \sum\limits_{i\in\mathcal{I}_m}\gamma_i G_{mi}}, \quad m = 1, 2, \ldots, M, \tag{5.20}$$

with

$$G_{kl} = \frac{\left|\mathbf{w}_{\mathrm{B},b_k}^{\mathrm{H}} \mathbf{a}_l\right|^2}{\left|\mathbf{w}_{\mathrm{B},b_k}\right|^2}, \tag{5.21}$$

being the spot beam array gain (without considering the array element gain) for source signal l with respect to beam b_k.

Demanding equal rates for all sources requires equal output SINR $\Gamma \leq \Gamma_{\mathrm{FB},m}$, and (5.20) can be written in matrix notation:

$$\underbrace{\begin{pmatrix} G_{11} & -\Gamma G_{12} & \cdots & -\Gamma G_{1M} \\ -\Gamma G_{21} & G_{22} & & \vdots \\ \vdots & & \ddots & \\ -\Gamma G_{M1} & \cdots & & G_{MM} \end{pmatrix}}_{\triangleq \mathbf{J}(\Gamma) \in \mathbb{R}^{M \times M}} \boldsymbol{\gamma} \geq \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \Gamma, \tag{5.22}$$

with $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_M)^{\mathrm{T}}$.

The rate limit, denoted with $R_{\max}^{\mathrm{FB}}$ in the following, that defines the rate corresponding to the maximal output SINR $\Gamma_{\max}^{\mathrm{FB}}$ that can be achieved at most for all sources at the same time as the γ_m tend to infinity. Finally, $\Gamma_{\max}^{\mathrm{FB}}$ is obtained for that Γ for that $\mathbf{J}(\Gamma)$, defined in (5.22), is singular and, in consequence

$$\boxed{\det\left(\mathbf{J}(\Gamma_{\max}^{\mathrm{FB}})\right) = 0.} \tag{5.23}$$

($\det(\mathbf{J}(\Gamma))$ is a polynomial in Γ of M-th order and has therefore M roots (zeros). $\Gamma_{\max}^{\mathrm{FB}}$ is then the smallest positive real zero of $\det(\mathbf{J}(\Gamma))$.)

Eqn. (5.23) is then evaluated numerically to obtain $\Gamma_{\max}^{\mathrm{FB}}$ and, finally, the according rate limit for fixed beamforming and subsequent independent decoding is given by

$$\boxed{R_{\max}^{\mathrm{FB}} = \log_2(1 + \Gamma_{\max}^{\mathrm{FB}}).} \tag{5.24}$$

Note that the rate limit depends only on the relative suppression G_{mi}/G_{mm} of interference i with respect to the wanted signal m, because scaling a column or row of a matrix scales the determinant. Therefore, we can normalise each row m of $\mathbf{J}(\Gamma)$ to G_{mm} without altering the Γ where $\det(\mathbf{J}(\Gamma)) = 0$.

Example 5.1: **Achievable SINR for 2 sources** In this case it holds

$$\Gamma_{\mathrm{FB},1} = \frac{\gamma_1 G_{11}}{1 + \gamma_2 G_{12}}, \quad \Gamma_{\mathrm{FB},2} = \frac{\gamma_2 G_{22}}{1 + \gamma_1 G_{21}}. \tag{5.25}$$

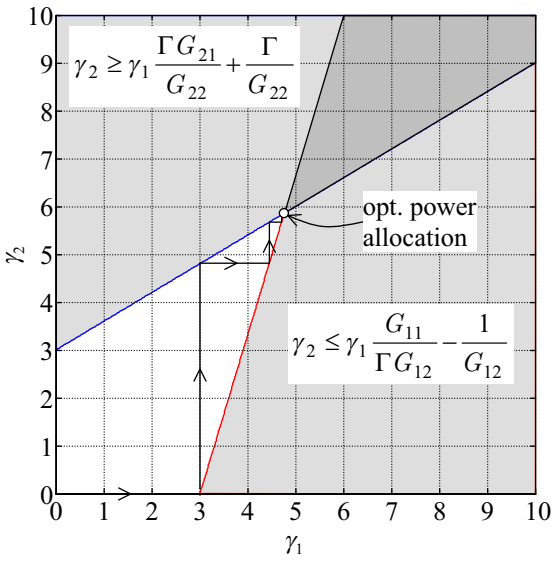
Demanding a minimum SINR $\Gamma \leq \Gamma_{\mathrm{FB},1}, \Gamma_{\mathrm{FB},2}$ and solving both above equations for γ_2 yields

$$\gamma_1 \frac{\Gamma G_{21}}{G_{22}} + \frac{\Gamma}{G_{22}} \leq \gamma_2 \leq \gamma_1 \frac{G_{11}}{\Gamma G_{12}} - \frac{1}{G_{12}}. \tag{5.26}$$

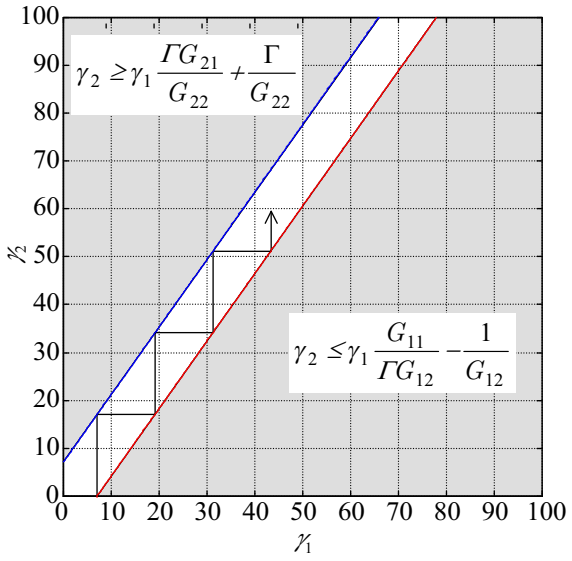
We assume, as an arbitrary example, $G_{11} = G_{22} = 1$, $G_{12} = 0.1$, $G_{21} = 0.2$.

Then the SINR limit according to (5.23) is $\Gamma_{\max}^{\mathrm{FB}} \approx 7.07$, i.e. only SINR $\Gamma < \Gamma_{\max}^{\mathrm{FB}} \approx 7.07$ can be achieved with finite input SNR γ_1, γ_2.

Fig. 5.18(a) shows the left and right hand side of (5.26) for $\Gamma = 3$. Obviously there is a region in the γ_1-γ_2-plane where both inequalities are fulfilled in (5.26), and the optimal power allocation is achieved at the point where strict equality holds in (5.26).



(a) $\Gamma = 3$. The arrows indicate the iterations of the power allocation procedure for independent decoding (starting at the origin), which obviously converges to the optimal point. Dark grey area indicates region where (5.26) is fulfilled.



(b) $\Gamma = \Gamma_{\max}^{\mathrm{FB}} \approx 7.07$. There is no region in the γ_1-γ_2-plane where (5.26) is fulfilled. Of course, the iterative power allocation procedure cannot converge.

Figure 5.18: Illustration of maximal achievable rates for 2 sources, $G_{11} = G_{22} = 1$, $G_{12} = 0.1$, $G_{21} = 0.2$.

However, if the required SINR is $\Gamma = \Gamma_{\max}^{\mathrm{FB}} \approx 7.07$ (cf. Fig. 5.18(b)), then there is no point in the γ_1-γ_2-plane for that (5.26) is fulfilled, because the graphs defined by the linear equations on the left and right hand side of (5.26) do not intersect in the positive quadrant for $\Gamma \geq \Gamma_{\max}^{\mathrm{FB}}$.

We have also indicated in Fig. 5.18 the iterations for the power allocation procedure for independent decoding (cf. Sec. 4.4.1) if the initial power allocation is $(0, 0)$. Clearly, the algorithm can converge only if the graphs defined by the linear equations on the left and right hand side of (5.26) intersect in the positive quadrant.

■

5.3.2 Successive Decoding ("FB+SD")

The receiver structure for fixed beamforming and successive decoding without a preceding adaptive beam space beamforming stage is depicted in Fig. 5.19. In this case the feedforward filter is

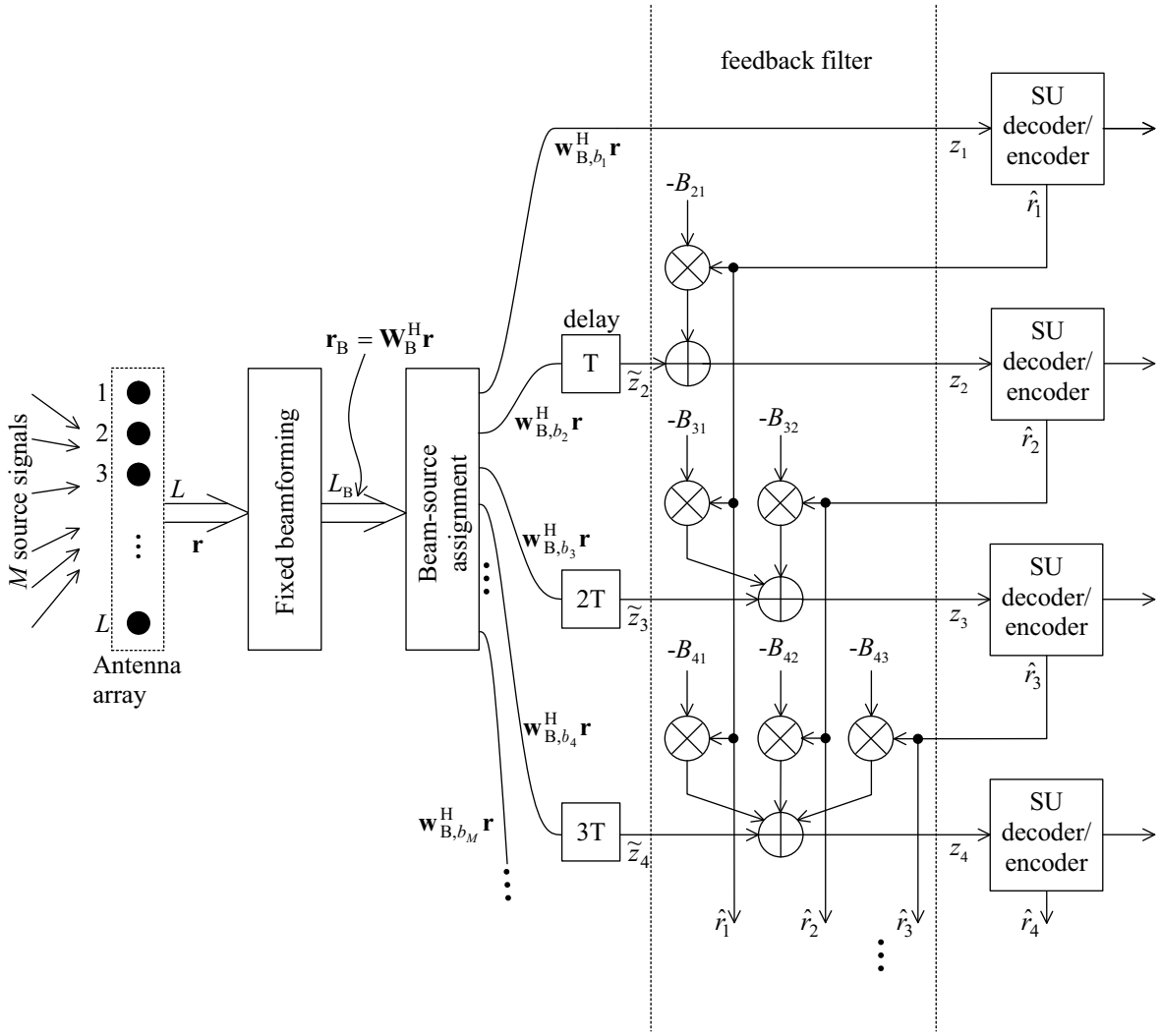


Figure 5.19: Block diagram of the receiver with fixed beamforming and subsequent successive decoding (FB+SD; without an adaptive beam space beamforming stage). To simplify representation it is assumed that the source indices are reordered such that decoding order is $(1, 2, \ldots, M)$. Delays of multiples of the decoding delay T of a single-user decoder are introduced to synchronise the signal flow in different stages of the successive decoder.

constituted by the beam-source assignment block. Furthermore, with the multiplicative factors of the linear feedback filter B_{mi} being defined as

$$B_{mi} = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_i, \tag{5.27}$$

and again assuming that the estimate of the received source symbol $\hat{r}_m$ is error-free such that it holds $\hat{r}_m = \tilde{r}_m$, the input z_m to the m-th single-user decoder is obtained by (cf. (4.61))

$$z_m = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \left(\mathbf{A}\tilde{\mathbf{r}} + \mathbf{n}\right) - \sum_{i=1}^{m-1} B_{mi}\hat{r}_i = \underbrace{\mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}}\mathbf{a}_m\tilde{r}_m + \overbrace{\sum_{\substack{i=1\\ i\neq m}}^{M} \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}}\mathbf{a}_i\tilde{r}_i}^{\text{Interference}} + \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}}\mathbf{n}}_{\triangleq\, \tilde{z}_m} - \sum_{i=1}^{m-1} B_{mi}\tilde{r}_i, \tag{5.28}$$

where finally the elements of the array input signal vector $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_M)^{\mathrm{T}}$ were introduced in (3.54) according to

$$\tilde{r}_m = \frac{h_m}{\sqrt{\mu_m}} s_m.$$

Therefore the symbols fed back in the process of successive decoding to remove interference of sources already decoded are given by

$$B_{mi}\frac{h_i}{\sqrt{\mu_i}} s_i = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}}\mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}} s_i. \tag{5.29}$$

The rates achievable by successive decoding as shown in Fig. 5.19 are readily provided by

$$\boxed{\begin{aligned} R_{\pi_v(m)} =& \mathrm{E}\left\{ \log_2\left(1 + \frac{P_{\pi_v(m)}\left|H_{\pi_v(m)}\right|^2 \left|\mathbf{w}_{\mathrm{B},b_{\pi_v(m)}}^{\mathrm{H}}\mathbf{a}_{\pi_v(m)}\right|^2}{\mathbf{w}_{\mathrm{B},b_{\pi_v(m)}}^{\mathrm{H}}\mathbf{K}_{\mathbf{v}}(\mathcal{S}_{v,m-1})\mathbf{w}_{\mathrm{B},b_{\pi_v(m)}} + 2\sigma_{\mathrm{n}}^2\left|\mathbf{w}_{\mathrm{B},b_{\pi_v(m)}}\right|^2}\right)\right\}_{\boldsymbol{H}} \\ & m = 1, 2, \ldots, M, \end{aligned}} \tag{5.30}$$

with the set $\mathcal{S}_{v,m} = \{\pi_v(1), \pi_v(2), \ldots, \pi_v(m)\}$ and, again, the sources are decoded in the order $(\pi_v(M), \pi_v(M-1), \ldots, \pi_v(1))$ for a given permutation $\boldsymbol{\pi}_v$.

Note that there is no rate limit as it was derived for the FB+ID receiver in the last section. A simple proof is based on the matrix notation of the relation between the vector of input SNR $\boldsymbol{\gamma}$ and the output SINR Γ provided in (5.22).

Proof. Without loss of generality we assume that the source indices are reordered such that the decoding order is $1, 2, 3, \ldots, M$ (also cf. Fig. 5.19). In this case, source 1 is affected by interfering sources $2, 3, \ldots, M$, source 2 by sources $3, 4, \ldots, M$ and so on. This can be written in matrix notation similar to (5.22), but yielding an upper triangular matrix:

$$\begin{pmatrix} G_{11} & -\Gamma G_{12} & -\Gamma G_{13} & \cdots & -\Gamma G_{1M} \\ 0 & G_{22} & -\Gamma G_{23} & \cdots & -\Gamma G_{2M} \\ 0 & 0 & G_{33} & \cdots & -\Gamma G_{3M} \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & G_{MM} \end{pmatrix} \boldsymbol{\gamma} \geq \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \Gamma, \tag{5.31}$$

The triangular matrix on the left hand side of (5.31) is non-singular, because the determinant (for a triangular matrix being the product of the diagonal elements [GL96]) is always non-zero for the reasonable case that $G_{mm} \neq 0$, $m = 1, 2, \ldots, M$. Thus there is always a solution vector of input SNR $\boldsymbol{\gamma}$ for all output SINR Γ. □

Next, a description of the scenarios considered here for evaluation and comparison of the performance of independent and successive decoding with fixed beamforming has to be provided.

5.3.3 Comparison Independent and Successive Decoding

For comparison of independent and successive decoding for the case of fixed beamforming with adaptive beam space beamforming we will consider two different interference scenarios, characterised by the considered satellite system (namely the ICO and the EuroSkyWay (ESW) system specified in Sec. 3.6) and the particular distribution of the sources.

ICO System

Fig. 5.20 shows again the cell pattern of an ICO satellite, created by the 163 fixed spot beams (cf. Sec. 3.6.1). In contrast to the original ICO system design where cluster size 4 is implemented [MS98, GST99], here cluster size $K_s = 3$ is assumed (cf. Sec. 3.3.3). According to (3.26), spectral efficiency is in this way increased by the factor $4/3$ (i.e. spectrum efficiency is increased by approx. 33%).

With cluster size 3 interference from co-channel sources is stronger, such that the advantages of both adaptive beamforming as well as successive decoding by more efficiently coping with interference than fixed beamforming or independent decoding, are more pronounced.

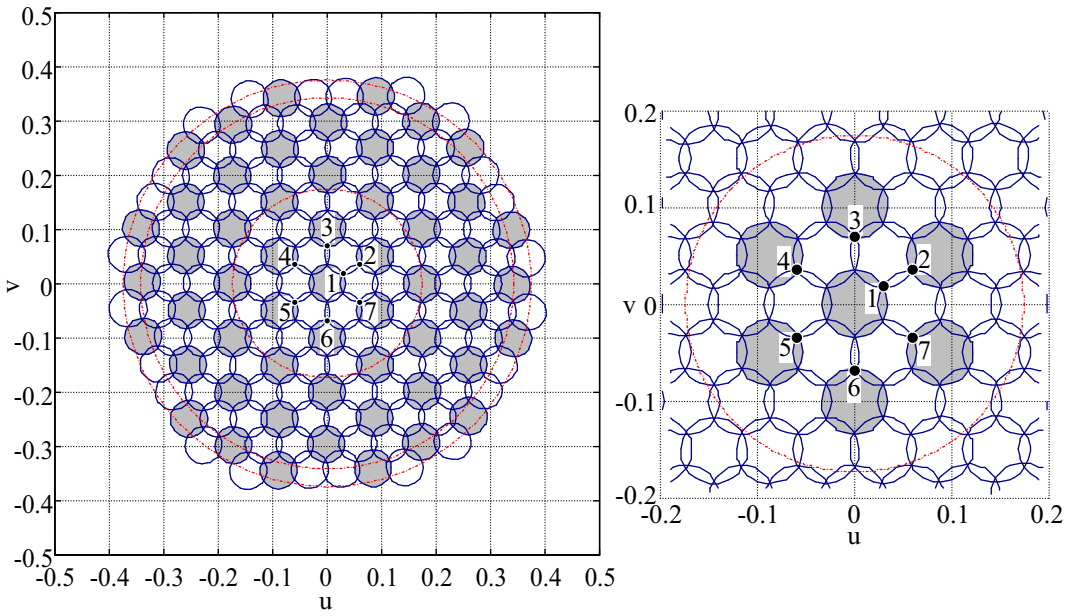


Figure 5.20: Left: ICO cell pattern resulting from 163 spot beams (cell contours defined by 3 dB spot beam contour) in u-v-coordinates. Further, the locations of the 7 sources are shown.

Again 7 sources are considered, and a worst-case distribution is considered for that both interference for source 1 caused by the other 6 sources is maximised, as well as the signal of source 1 suffers from the 3 dB edge-of-cell loss (cf. Fig. 5.20) [LWJ00].

For a better understanding of the results presented further below, Tab. 5.2 shows the spot beam array gains G_{kl} as defined by (5.21) for the assumed source distribution. As already mentioned, source 1 suffers the strongest interference, indeed the spot beam gains of interferers are only

Spot beam array gain G_{kl} (dB)							
$k \downarrow \backslash l \rightarrow$	1	2	3	4	5	6	7
1	18.16	8.39	8.39	8.39	8.39	8.39	8.39
2	8.72	18.26	-19.34	-12.38	-2.65	-12.45	-20.09
3	-18.69	-19.47	18.26	-20.38	-12.44	-2.65	-12.35
4	-3.7	-12.61	-20.44	18.24	-20.3	-12.59	-2.63
5	-26.79	-2.65	-12.45	-20.09	18.26	-19.34	-12.38
6	-3.64	-12.44	-2.65	-12.35	-19.47	18.26	-20.37
7	-19.6	-20.3	-12.59	-2.63	-12.61	-20.44	18.24

Table 5.2: Spot beam array gains G_{kl} (cf. (5.21)) for the spot beam associated with the k-th source with respect to the l-th source (for ICO).

10 dB below the gain for source 1 (contrast this with the minimal interference suppression of approx. 22 dB obtained for the original ICO design, cf. Fig. 3.30(b)). Further, source 2 faces similar interference from source 1, whereas interference from other sources is suppressed by approx. 20 dB–40 dB. For sources 3 to 7 all interferers are suppressed by approx. 20 dB–40 dB.

With (5.23) we have provided the equation which yields analytically the rate limit, which lies at $R_{\max}^{\mathrm{FB}} = 3.23$ bit/channel use for the assumed source distribution (the respective maximal output SINR is $\Gamma_{\max}^{\mathrm{FB}} = 9.2$ dB), which agrees perfectly with the numerical results shown in Fig. 5.21.

AWGN Using (5.19) the element SNR γ_m required to achieve a certain rate R with fixed beamforming and independent decoding (FB+ID receiver) are calculated for the AWGN case and shown in Fig. 5.21.

For low rates (approx. $R < 0.5$) the required γ_m to achieve rate R for all sources are determined by thermal noise alone and are in consequence very close to the minimum element SNR that would suffice to achieve R if only thermal noise would be present. The noise limit is almost the same for all sources, because the spot beam gains are almost identical as well, as indicated in Tab. 5.2 for $k = l$. However, in the general case the noise limit is different for each source due to the potentially different locations inside the respective spot beams and the resulting spot beam gain variations.

For higher rates interference becomes rapidly more and more dominant, and the distance between the actually required element γ_m and the noise limit increases with increasing R, up to the rate limit where the γ_m finally tend to infinity. In particular, interference is strongest for source 1 (caused by sources 2 to 7) and for source 2 (mainly caused by source 1), and therefore source 1 and 2 require the highest element SNR.

Contrast the results for the FB+ID receiver with the required element SNR for the FB+SD receiver, employing successive decoding as shown in Fig. 5.22.

The decoding order is $(4, 6, 2, 3, 7, 5, 1)$[3], where source 2 requires the highest element SNR due to interference caused by source 1, while for all other sources element SNR close to the noise limit suffice.

Obviously, mutual interference is efficiently suppressed by removing interfering signals in the process of successive decoding in the FB+SD receiver. Further, no rate limit exists, extending the region of achievable rates with respect to the FB+ID receiver, whereas the noise limit is the same as for FB+ID, since the sources suffer the same edge-of-cell loss in both cases.

[3]The optimal decoding order was found by simply trying all possible decoding orders.

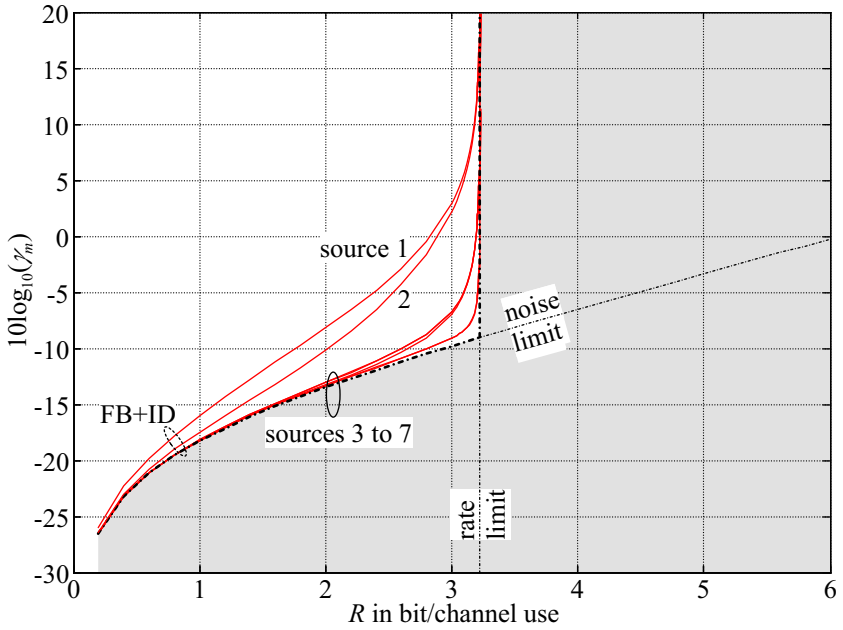


Figure 5.21: Element SNR γ_m required to achieve rate R for fixed beamforming with independent decoding (FB+ID). Also shown is the limit for minimal γ_m that is approached if only thermal noise is present (noise limit), further the limit of maximal achievable rate due to mutual interference $R_{\max}^{\mathrm{FB}} = 3.23$ is indicated (denoted as rate limit, can be computed with (5.23)). The area where no γ-R-pair can be achieved due to one or the other limit is highlighted grey.

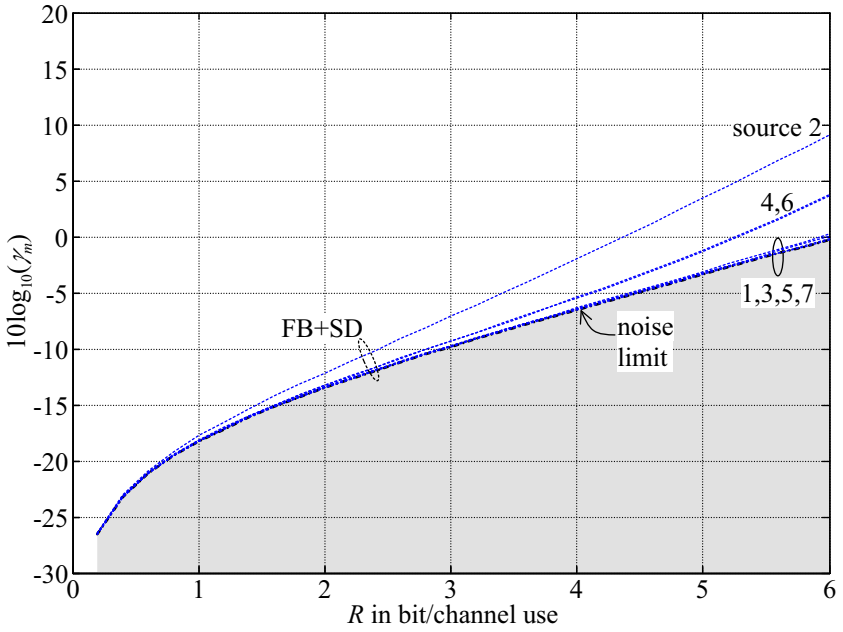


Figure 5.22: Fixed beamforming with successive decoding (FB+SD) with decoding order $(4, 6, 2, 3, 7, 5, 1)$. The greyed area indicates the region where no γ-R-pair can be achieved due to thermal noise.

Further, results for independent (FB+ID) and successive decoding (FB+SD) are compared directly in Fig. 5.23.

It can be seen that the worst-case element SNR (i.e. that of source 1 for FB+ID compared to that of source 2 for FB+SD) required for FB+SD are significantly reduced with respect to FB+ID, even for quite low target rate R.

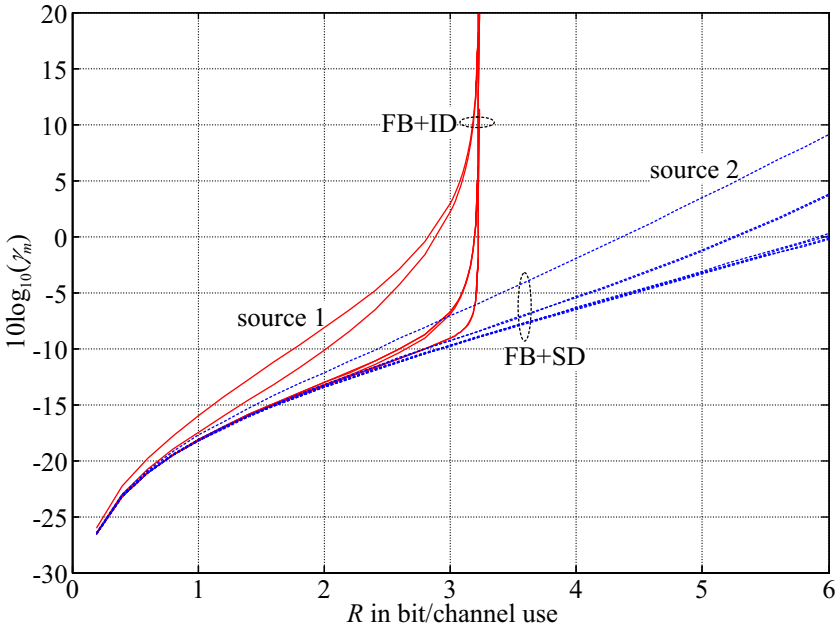


Figure 5.23: Fixed beamforming only, comparison of required element SNR γ_m for FB+ID (red, solid) and FB+SD (blue, dashed) with decoding order $(4, 6, 2, 3, 7, 5, 1)$.

Finally, we turn to the question what element SNR and, in consequence, what rates can actually be achieved, when the constraints regarding the terminal EIRP and the thermal noise power in the satellite receiving equipment are considered (Tab. 3.1 and Tab. 3.2).

The EIRP is obtained from given element SNR γ_m using (3.92) according to

$$p_m G_{\mathrm{T}} = \gamma_m \frac{2\sigma_{\mathrm{n}}^2 L_{\mathrm{bf},m}}{G_{\mathrm{e}}(\vartheta_m)} \frac{c_{\mathrm{R}}}{1 + c_{\mathrm{R}}}. \tag{5.32}$$

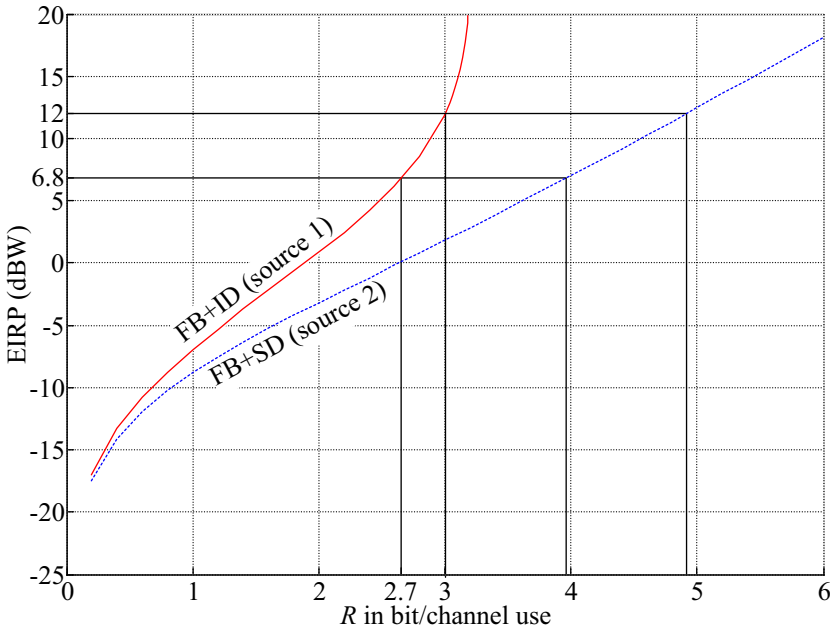
The noise power $2\sigma_{\mathrm{n}}^2$ is given by [LWJ00]

$$2\sigma_{\mathrm{n}}^2 = \mathrm{k} B T_{\mathrm{e}}, \tag{5.33}$$

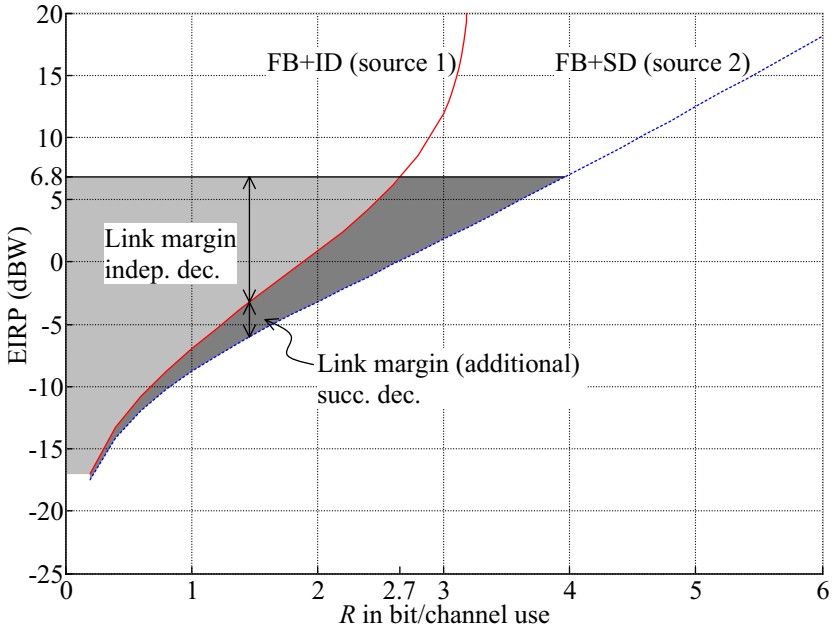
with k being the Boltzmann constant, noise equivalent bandwidth B, and effective noise temperature T_{e}.

Plugging the values for bandwidth B and effective noise temperature T_{e} from Tab. 3.2 into (5.33) together with $L_{\mathrm{bf},m} \approx 178.8$ dB and $G_{\mathrm{e}}(\vartheta_m) \approx 9.5$ dB we obtain from (5.32) the required EIRP, which is shown in Fig. 5.24(a) for the worst-case sources for independent and successive decoding. Taking the peak EIRP of the ICO terminals into account (6.8 dBW and 12 dBW, cf. Tab. 3.1), it can be stated that FB+SD theoretically allows rates that are by a factor of approx. 1.5 higher than that achievable with FB+ID.

Usually the system design does not envisage that the user terminals will permanently transmit at their peak EIRP, instead the EIRP required to achieve the desired transmission rate is several dB



(a) Maximal achievable rates, limited by peak EIRP.



(b) Link margin with respect to 6.8 dBW peak EIRP. With successive decoding an additional link margin is achieved.

Figure 5.24: Fixed beamforming only, comparison of rates achievable with peak EIRP and link margin for FB+ID (red, solid) and FB+SD (blue, dashed) with decoding order $(4, 6, 2, 3, 7, 5, 1)$.

below the peak value. In this way a link margin is obtained which allows to compensate signal shadowing to a certain extent [LWJ00].

Fig. 5.24(b) indicates that a significant additional link margin between 1.7 dB and approx. 10 dB can be achieved in the considered range of rates ($0 \leq R \leq 3$ bit/channel use) with FB+SD compared to FB+ID (cf. Tab. 5.3).

Peak EIRP (dBW)		Rate $R = 1$	$R = 2$	$R = 3$
6.8	Indep. dec.	13.7	5.9	/
	Succ. dec.	15.4	9.8	4.78
12	Indep. dec.	18.9	11.1	0.1
	Succ. dec.	20.6	15	10
	Additional link margin	1.7	3.9	9.9

Table 5.3: For required rate R the minimum link margin (in dBW) with respect to peak EIRP 6.8 dBW and 12 dBW, respectively, is shown for independent decoding and for successive decoding. The minimum link margin is obtained for source 1 for independent decoding and for source 2 for successive decoding (cf. Fig. 5.24(b)).

Fading Further, for independent decoding for the classical MAC it was observed in Sec. 4.3.2 (in particular Fig. 4.4(a)) that, in comparison with the AWGN case, fading reduces the achievable rates for a single source, while it increases the achievable rates in an interference scenario due to a statistical multiplexing effect. Vice versa, if the rates are fixed, then fading increases the required power to achieve the desired rate for a single source, while the required powers are reduced in an interference scenario. This effect is most pronounced for the classical MAC, but it is still visible, though eventually to a lesser extent, for the vector fading MAC where interference is spatially filtered to a certain degree.

Whether fading increases or decreases the rates achievable with independent decoding in a vector MAC depends on the spatial separability of the interfering sources (cf. Fig. C.1(a)). For the interference scenario considered in this section it can be seen from Fig. 5.25 that the required element SNR for independent decoding are increased for the most part if fading is present. Only as the element SNR rapidly grow for rates close to the rate limit at approx. $R = 3.2$, the statistical multiplexing effect becomes effective, and slightly, but actually negligibly higher rates can be achieved in the fading channel.

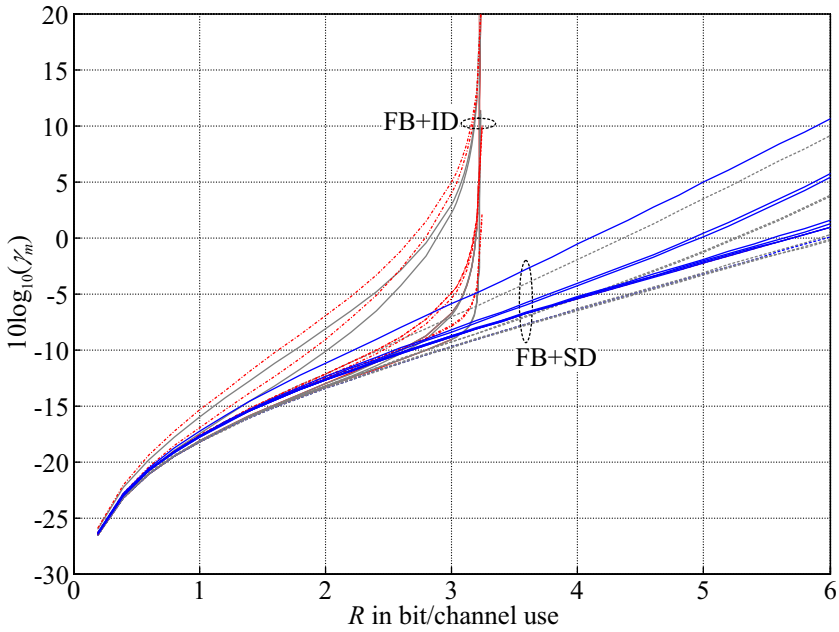


Figure 5.25: Fixed beamforming only. Comparison of required element SNR γ_m to achieve rate R in a fading channel (Rice, $c_{\mathrm{R}} = 5$ dB) and AWGN channel (grey), for FB+ID with fading (red, dot-and-dashed) and FB+SD with fading (blue, solid) with decoding order $(4, 6, 2, 3, 7, 5, 1)$.

EuroSkyWay System

Fig. 5.26 shows the EuroSkyWay (ESW) fixed cell pattern (cf. Sec. 3.6.2), together with the source distribution assumed in the further.

Again the sources are arranged according to cluster size 3, which is in deviation from the original system design foreseeing a 4-frequency reuse (cf. Fig. 3.31(b)).

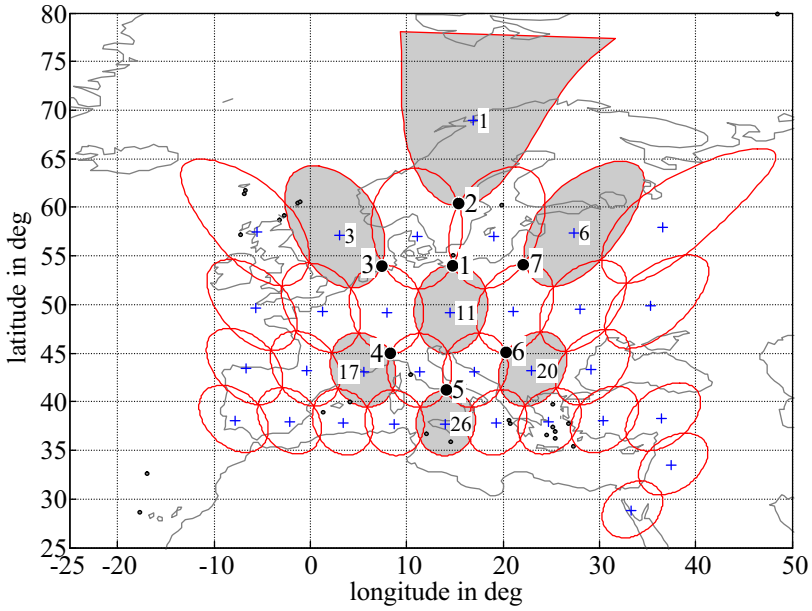


Figure 5.26: EuroSkyWay cell pattern resulting from 32 spot beams (cell contours defined by 4.3 dB spot beam contour). Further, the locations of the 7 sources are shown.

Tab. 5.4 shows the resulting spot beam array gains G_{kl}, from which the rate limit for fixed beamforming with independent decoding (FB+ID) is, again, obtained from (5.23) as $R_{\max}^{\mathrm{FB}} = 5.79$ bit/channel use (the maximal achievable output SINR as the γ_m tend to infinity is $\Gamma_{\max}^{\mathrm{FB}} = 17.3$ dB).

In comparison with the ICO scenario (cf. Fig. 5.20) a higher rate limit is achieved, because of a better suppression of interference: for source 1 the worst-case interference is suppressed by approx. 15 dB compared to only 10 dB for the ICO scenario (cf. Tab. 5.2).

We will consider in the further only the AWGN case, as we have seen in the last section that the fading case does not add any significantly new aspects.

AWGN As for the ICO scenario, the element SNR required to achieve rate R employing the FB+ID and the FB+SD receiver, respectively, are calculated for the AWGN channel using (5.19), and Fig. 5.27 shows the resulting element SNR for the EuroSkyWay scenario in comparison to the ICO scenario.

The noise limit for the ICO scenario lies approx. 2 dB below that for the EuroSkyWay scenario (compare the G_{kl} for $k = l$ in Tab. 5.2 with that in Tab. 5.4)[4].

[4]The edge-of-cell array gain is for the ICO system given by $10\log_{10}(127\eta_{\mathrm{A,a}}) - 3$ dB ≈ 18 dB (with aperture efficiency $\eta_{\mathrm{A,a}} = 0.95$, Tab. 3.2), and for the EuroSkyWay system it holds $10\log_{10}(169\eta_{\mathrm{A,a}}) - 4.3$ dB ≈ 16.2 dB (with aperture efficiency $\eta_{\mathrm{A,a}} = 0.66$, Tab. 3.4). From this we finally obtain with 18 dB $- 16.2$ dB $= 1.8$ dB the approximate 2 dB difference in the noise limit observed in Fig. 5.27

Spot beam array gain G_{kl} (dB)							
$k \downarrow \backslash l \rightarrow$	1	2	3	4	5	6	7
1	16.3	-2.58	-0.37	-0.08	-0.72	0.23	1.05
2	-1.88	16.48	-11.95	-24.48	-26.7	-24.36	-12.18
3	-12.85	-12.8	16.04	-12.02	-23.76	-26.14	-23.81
4	-41.5	-24.09	-12.49	16.02	-12.83	-23.87	-25.87
5	-24.67	-27.2	-23.94	-12.46	16.18	-12.41	-24.1
6	-39.12	-24.19	-26.08	-23.66	-12.4	16.17	-12.24
7	-12.7	-12.74	-24.03	-26.03	-24.04	-12.26	15.8

Table 5.4: Spot beam array gains G_{kl} (cf. (5.21)) for the spot beam associated with the k-th source with respect to the l-th source (for EuroSkyWay).

Despite this and the different maximal achievable rates (rate limit), the plots of required element SNR show otherwise a similar behaviour in both the considered scenarios.

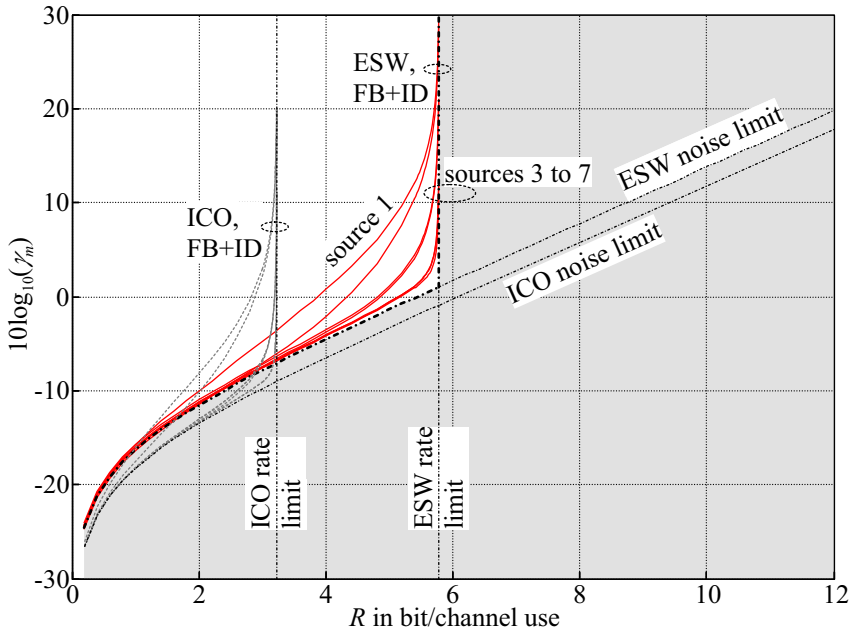


Figure 5.27: Element SNR γ_m required to achieve rate R in an AWGN channel for fixed beamforming with independent decoding (FB+ID). Also shown is the limit for minimal γ_m that is approached if only thermal noise is present (noise limit), further the limit for maximal achievable rates due to mutual interference (rate limit, can be computed with (5.23)). The area where no γ-R-pair can be achieved due to one or the other limit is highlighted grey. For comparison the results for the ICO system are included (grey).

Lower input SNR are again obtained when successive decoding is employed (cf. Fig. 5.28).

The optimal decoding order was found to be $(2, 5, 7, 3, 6, 4, 1)$, i.e. as expected the centre source with index 1 is decoded last, because it faces the strongest interference being surrounded by the 6 interfering sources.

Comparing the input SNR required for independent (FB+ID) and successive decoding (FB+SD) (cf. Fig. 5.29) clearly shows that also in the EuroSkyWay scenario for the considered source loca-

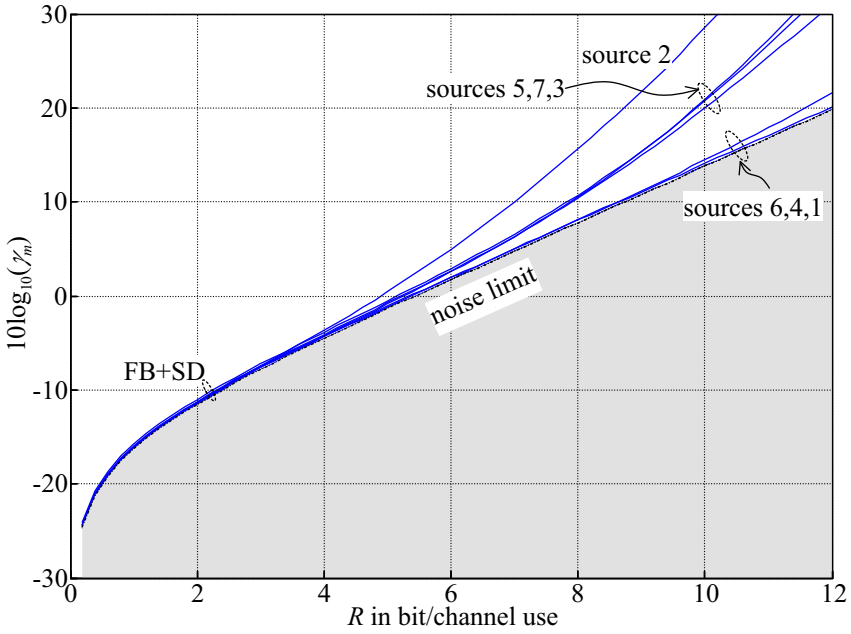


Figure 5.28: Fixed beamforming with successive decoding (FB+SD) with decoding order $(2, 5, 7, 3, 6, 4, 1)$. The grey area indicates the region where no γ-R-pair can be achieved due to thermal noise.

tions, substantial savings in transmit power can be achieved through successive decoding compared to independent decoding.

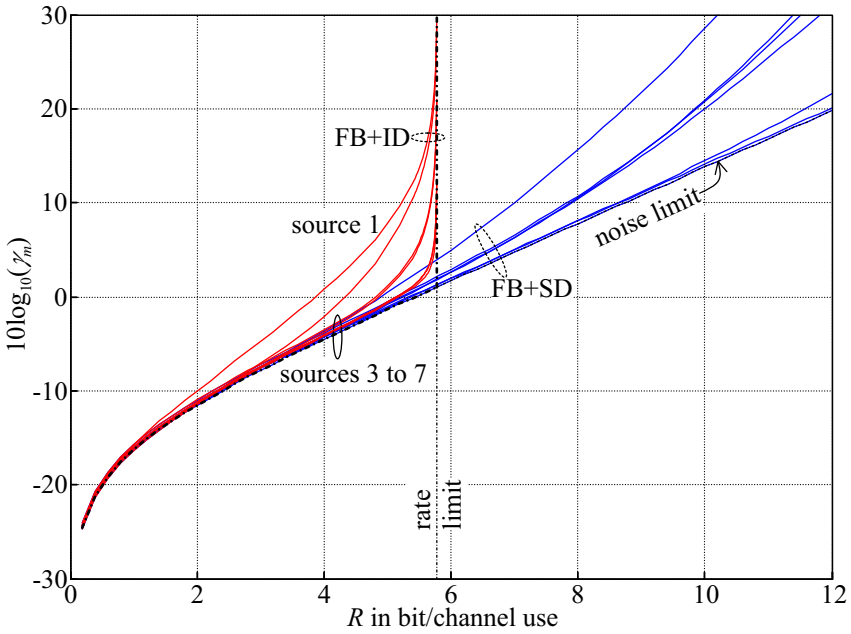


Figure 5.29: Fixed beamforming only, comparison of required element SNR γ_m for FB+ID (red, solid) and FB+SD (blue, dashed) with decoding order $(2, 5, 7, 3, 6, 4, 1)$.

A concluding comparison, considering the peak EIRP that can be reached by the satellite terminals specified for the EuroSkyWay system (37.8 dB and 49.9 dB, respectively, cf. Tab. 3.3) is shown in Fig. 5.30.

The EIRP $p_m G_{\mathrm{T}}$ is again calculated for each source m from the required input SNR γ_m using (5.32) and (5.33).

Here, the equivalent effective noise temperature is $T_{\mathrm{e}} = 444$ K (cf. Tab. 3.4).

Further, it was pointed out in the EuroSkyWay system outline presented in Sec. 3.6.2, that variable data rates requiring different bandwidth are supported. Therefore, we consider typical bandwidths $B = 200$ kHz and $B = 2$ MHz.

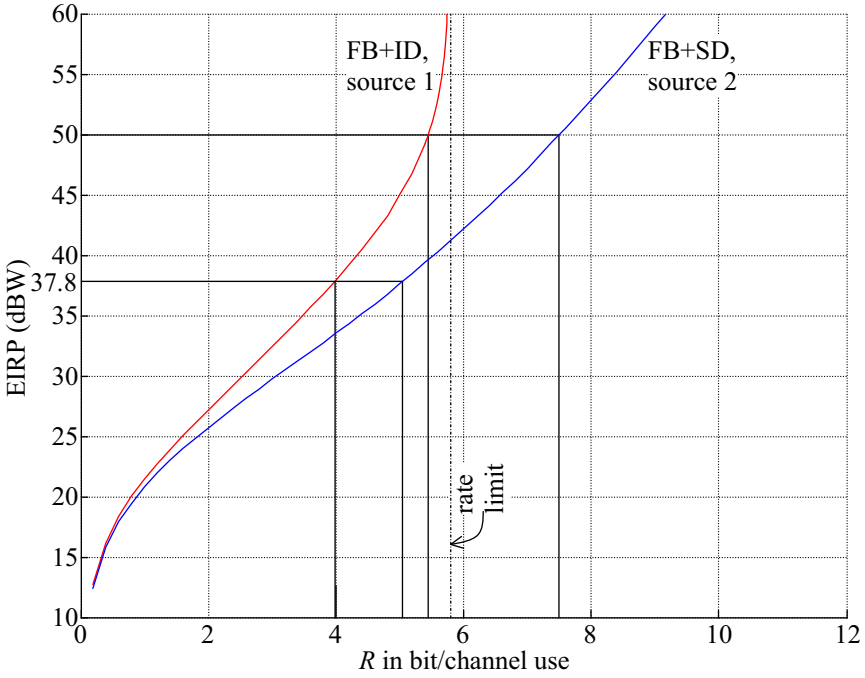
Finally, the free-space basic transmission loss is for all sources approx. $L_{\mathrm{bf},m} \approx 213$ dB, and the satellite antenna array element gain amounts to approx. 27 dB.

Considering the peak EIRP, successive decoding (FB+SD) can support roughly 1.3 times higher rates (bit/channel use) than independent decoding for both $B = 200$ kHz, as well as for $B = 2$ MHz. This is slightly less gain than what was observed for the ICO system (cf. Fig. 5.24(a)).

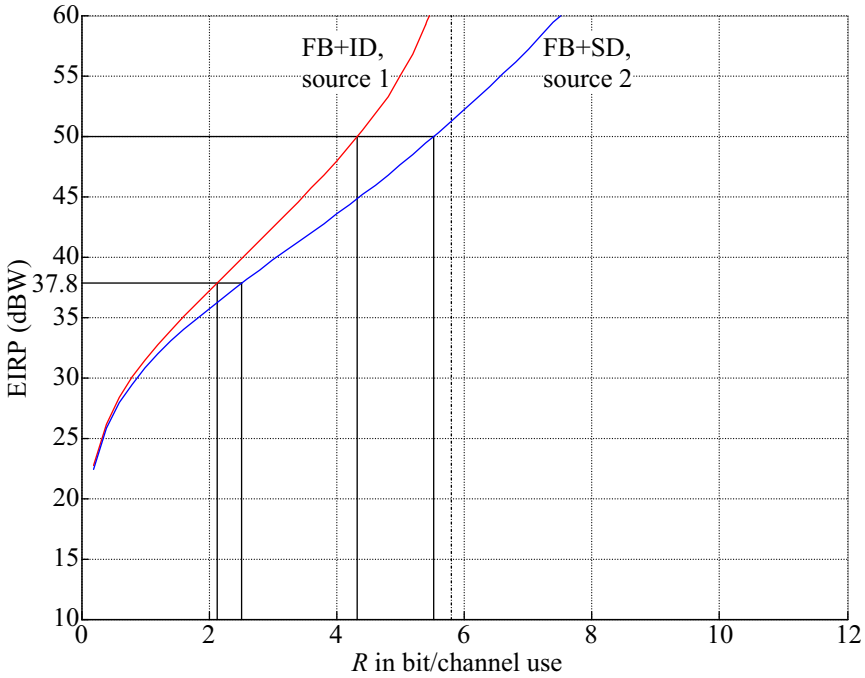
The link margin that is achieved with independent and successive decoding is for various rates shown in Tab. 5.5.

Band-width B	Peak EIRP (dBW)		Rate					
			R=1	R=2	R=3	R=4	R=5	R=6
200 kHz	37.8	FB+ID	16.3	10.8	5.3	/	/	/
		FB+SD	16.8	12.3	8.3	4.3	0.2	/
	49.9	FB+ID	28.4	22.9	17.4	11.9	5	/
		FB+SD	28.9	24.4	20.4	16.4	12.3	7.8
		Add. link margin	0.5	1.5	3	4.5	7.3	/
2 MHz	37.8	FB+ID	6.3	0.6	/	/	/	/
		FB+SD	6.8	2.1	/	/	/	/
	49.9	FB+ID	18.4	12.9	7.4	1.9	/	
		FB+SD	18.9	14.4	10.4	6.4	2.3	/

Table 5.5: For required rate R the minimum link margin (in dBW) with respect to peak EIRP 37.8 dBW and 49.9 dBW, respectively, is shown for independent decoding (FB+ID) and for successive decoding (FB+SD). The minimum link margin is obtained for source 1 for independent decoding and for source 2 for successive decoding.



(a) Maximal achievable rates, limited by peak EIRP for bandwidth $B = 200$ kHz.



(b) Maximal achievable rates, limited by peak EIRP for bandwidth $B = 2$ MHz.

Figure 5.30: Fixed beamforming only, comparison of rates achievable with peak EIRP for bandwidths $B = 200\,\text{kHz}, 2\,\text{MHz}$ for independent (solid) and successive decoding (dashed) with decoding order $(2, 5, 7, 3, 6, 4, 1)$.

5.4 Comparison Receivers with and without Adaptive Beamforming

In this section we finally turn towards the performance of a receiver employing adaptive beam space processing (cf. Sec. 5.2.1).

It was already indicated that it is of particular interest to compare performance of fixed beamforming with successive decoding, as discussed in the preceding section, and adaptive beamforming with independent decoding.

5.4.1 ICO System

Using (5.2) and the iterative power allocation algorithm from Sec. 4.4.1, the element SNR required to achieve rate R in the AWGN case are calculated for the AB+ID receiver (i.e. employing fixed beamforming with subsequent optimal beam space beamforming and independent decoding, cf. Fig. 5.31).

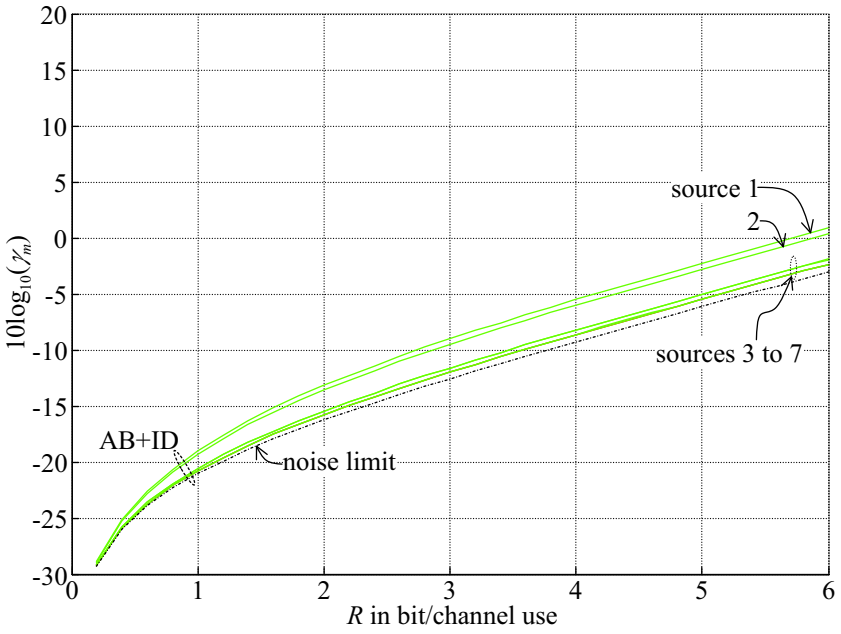


Figure 5.31: Required element SNR γ_m for fixed beamforming and subsequent optimal adaptive beam space beamforming (AB+ID).

For fixed beamforming it was observed that the slope of the element SNR graphs increases as the target rate R increases (this holds both for FB+ID, as well as for the FB+SD case, cf. Fig. 5.23), because of the mutual interference becoming stronger.

In contrast, we can observe here that the slope decreases, and finally reaches a constant value[5], as the target rate R increases. This results from the optimal beamformer that suppresses interference

[5]It can be readily verified (e.g. with (4.16), using that interference is perfectly removed) that the slope approaches $10\log_{10} 2$ in the logarithmic scale, i.e. for each increase in rate R by 1 bit/channel use, the element SNR is increased by 3 dB.

to the point where a pattern null completely removes any interference (cf. Fig. 3.23), such that the output SINR becomes independent of the interferers signal power.

Fig. 5.32 shows now the required element SNR for fully adaptive beamforming with independent decoding (AB+ID) in comparison with the results for successive decoding with fixed beamforming only (FB+SD).

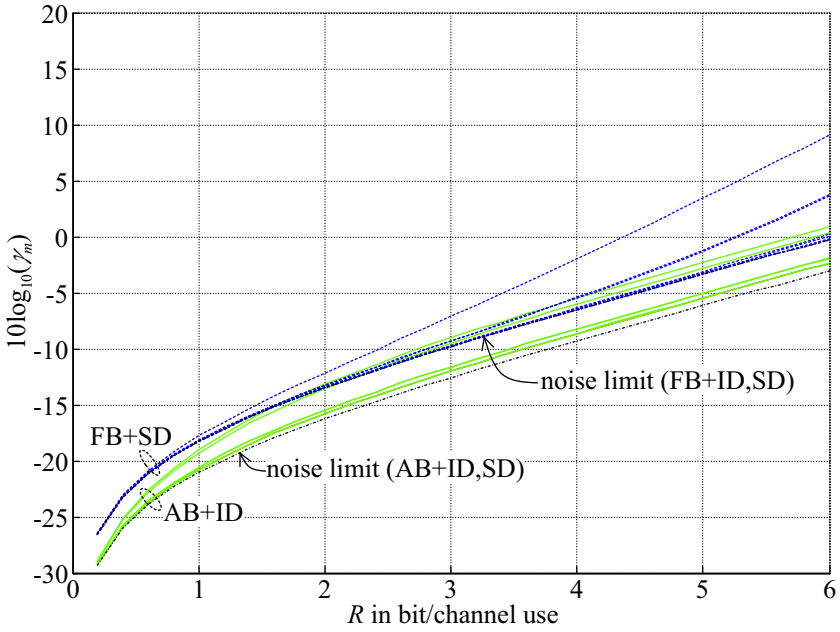


Figure 5.32: Comparison of required element SNR γ_m for fixed beamforming only with successive decoding (FB+SD), and optimal adaptive beamforming with independent decoding (AB+ID) (blue, dashed). The noise limit for receivers with fixed beamforming (FB) (green, solid) lies 3 dB above that for receivers employing adaptive beamforming (AB), because the 3 dB-edge-of-cell-loss is compensated in the latter case.

Firstly, one notices that for low rates (below approx. $R = 1.5$) the SNR of the receiver with fixed beamforming (FB+SD) are 1 dB to 3 dB higher than for the receiver with optimal adaptive beamforming (AB+ID), because the edge-of-cell loss is partially or fully (depending in interference) compensated employing adaptive beamforming. If two mutually interfering sources are located close together then it may not be possible to fully recover edge-of-cell, which is the case for sources 1 and 2, for the source distribution considered here (as depicted in Fig. 5.20).

The gap between fixed beamforming with successive decoding (FB+SD) and optimal adaptive beamforming with independent decoding (AB+ID) widens as for increasing rates (approx. $R > 1.5$) interference becomes more and more the dominating factor. Then the advantage of adaptive beamforming in optimally suppressing interference becomes decisive, and for the considered scenario adaptive beamforming with subsequent independent decoding (AB+ID) is for all target rates R superior to fixed beamforming with successive decoding (FB+SD).

In terms of maximal required input SNR, the optimal successive decoder employing both optimal beamforming together with successive decoding (AB+SD) performs only little better than the receiver with optimal beamforming and independent decoding (AB+ID) (cf. Fig. 5.33, where the maximum input SNR is required for source 1 in both cases).

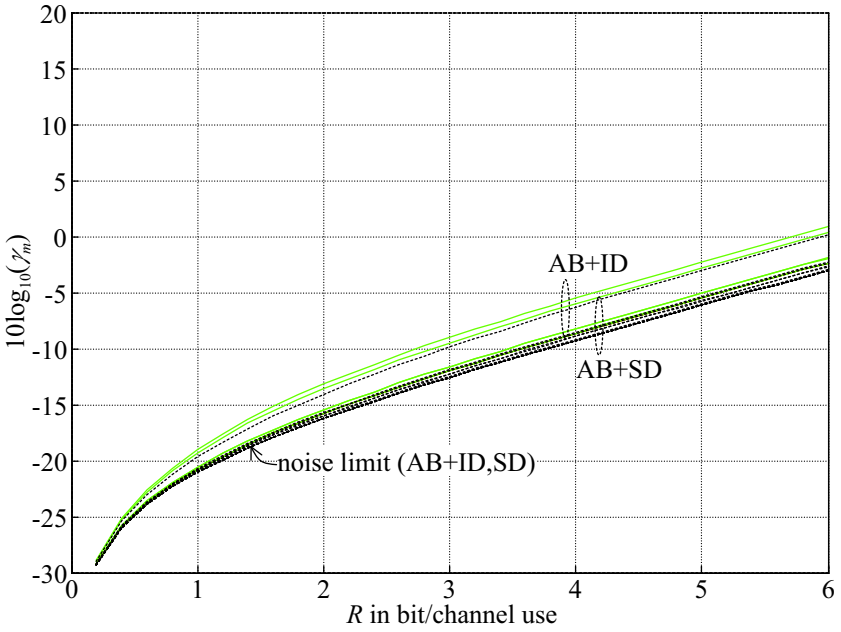


Figure 5.33: Comparison of required element SNR γ_m for receivers employing optimal adaptive beamforming, independent (AB+ID) (green, solid) and successive decoding (AB+SD) (black, dashed).

To conclude comparison of the different receiver options, Fig. 5.34 shows the maximum EIRP for fixed beamforming with independent (FB+ID) and successive decoding (FB+SD), further for the receiver with optimal adaptive beam space processing with subsequent independent decoding (AB+ID).

It can be stated that with AB+ID and the maximal EIRP being 6.8 dBW, rates can be achieved that are 1.3 times higher than that achievable with FB+SD, and 1.9 times higher than for FB+ID. With the EIRP being 12 dBW, the factors are 1.3 and 2.2.

Finally, it can be concluded for the scenario considered that the receiver employing optimal adaptive beam space processing with subsequent independent decoding (AB+ID) performs superior to the receiver employing fixed beamforming with successive decoding (FB+SD).

However, we will next address an interference scenario considered in [Ern01], where in this scenario it shows that, depending on the target rate, either fixed beamforming with successive decoding outperforms optimal adaptive beamforming with independent decoding or vice versa.

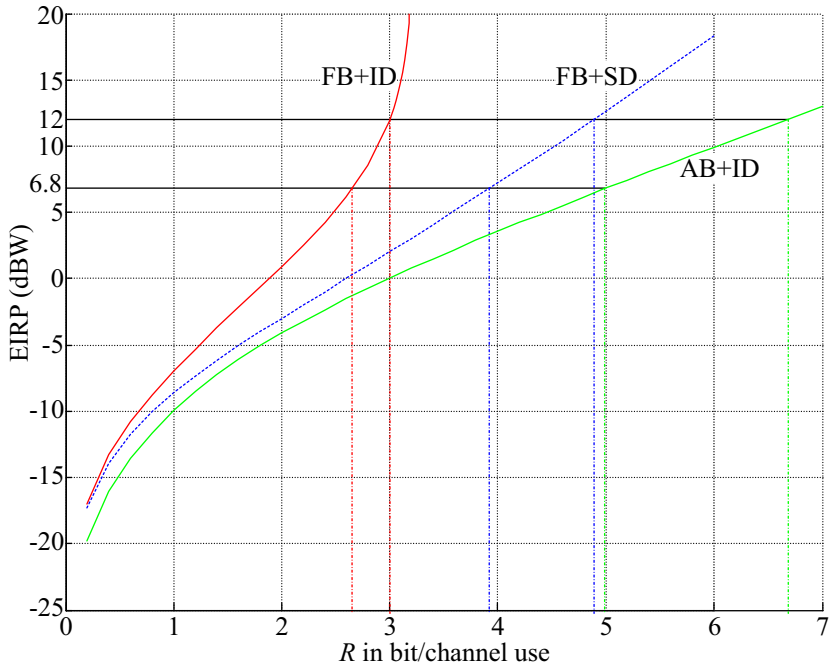


Figure 5.34: Comparison of rates achievable with peak EIRP and link margin for fixed beamforming with independent (FB+ID) and successive decoding (FB+SD) with decoding order $(4, 6, 2, 3, 7, 5, 1)$, and with optimal adaptive beam space processing with independent decoding (AB+ID).

Adjacent Cell Frequency Reuse

In this scenario, the area of each cell is divided into two subareas of equal size (grey circular area and the hexagon inscribed in each cell in Fig. 5.35). Further it is assumed that in the circular areas located around the cell centres, the same frequency band is reused in every cell, while otherwise, as discussed before, a frequency reuse according to cluster size 3 has to be adopted to avoid excessive interference [Ern01].

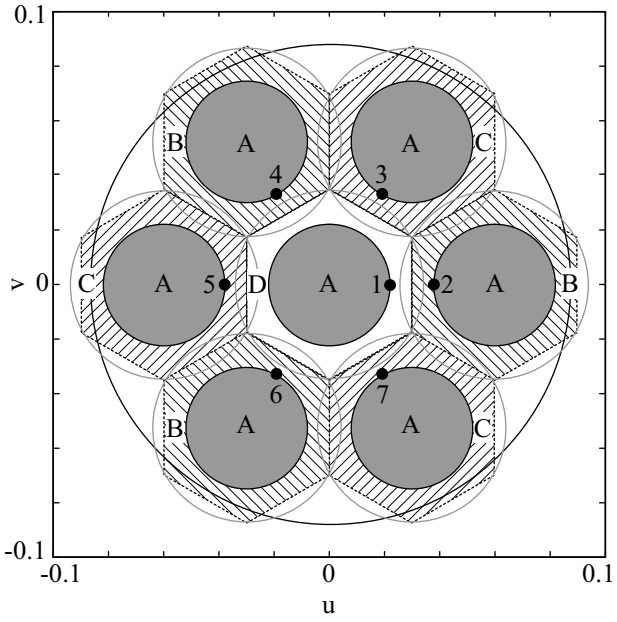


Figure 5.35: Frequency reuse scheme (frequency A is reused in every cell, while otherwise cluster size 3 with frequencies B–D is assumed) and worst-case source distribution for the scenario considered in [Ern99]. The 7 sources are located in adjacent cells, where mutual interference is limited by restricting the possible source positions to the areas highlighted grey.

With cluster size $K_\mathrm{s} = 3$ the bandwidth requirement per cluster is $B_\mathrm{sat} = 3B_\mathrm{c}$ (cf. (3.25)). The requirement is only $(3/2B_\mathrm{c} + 1/2B_\mathrm{c}) = 2B_\mathrm{c}$ for the modified reuse scheme, because of the equal size of the cell subareas and assuming homogeneous traffic. Therefore spectrum efficiency is increased by a factor of $3/2$ with respect to cluster size 3 and even by a factor of 2 with respect to cluster size 4 (cf. (3.26)).

Tab. 5.6 shows the spot beam array gains G_{kl}, indicating that for the worst-case source distribution, minimal attenuation of interference is only approx. 2.2 dB considering the centre source (recall that for the ICO cluster size 3 scenario, Fig. 5.20, a minimum relative interference suppression of approx. 10 dB was observed).

Again, using (5.23), the maximal rate that is achievable with fixed beamforming and independent decoding can be calculated, yielding $R_\mathrm{max}^\mathrm{FB} = 0.87$ bit/channel use. This agrees, as expected, with the results obtained from numerical evaluation (cf. Fig. 5.36).

Coming now to the most interesting point what concerns the considered scenario, we observe that for some interval of target rates R, fixed beamforming with successive decoding (FB+SD) outperforms optimal adaptive beam space processing with independent decoding (AB+ID), and

Spot beam array gain G_{kl} (dB)							
$k \downarrow \backslash l \rightarrow$	1	2	3	4	5	6	7
1	19.7	17.5	17.5	17.5	17.5	17.5	17.5
2	17.8	19.8	14.3	-3.9	-7.5	-4	14.4
3	14.4	14.3	19.8	14.3	-4	-7.5	-4
4	5.8	-4	4.4	19.9	14.4	-4	-7.5
5	-1.1	-7.5	-4	14.4	19.9	14.4	-4
6	5.8	-4	-7.5	-4	14.4	19.9	14.4
7	14.4	14.4	-4	-7.5	-4	14.4	19.9

Table 5.6: Spot beam array gains G_{kl} (cf. (5.21)) for the spot beam associated with the k-th source with respect to the l-th source (for ICO system, interference scenario according to Fig. 5.35).

vice versa (cf. Fig. 5.36). This is in contrast to the cluster size 3 scenario, where the AB+ID receiver structure performed always better than the FB+SD receiver, considering the maximal required input SNR (cf. Fig. 5.34).

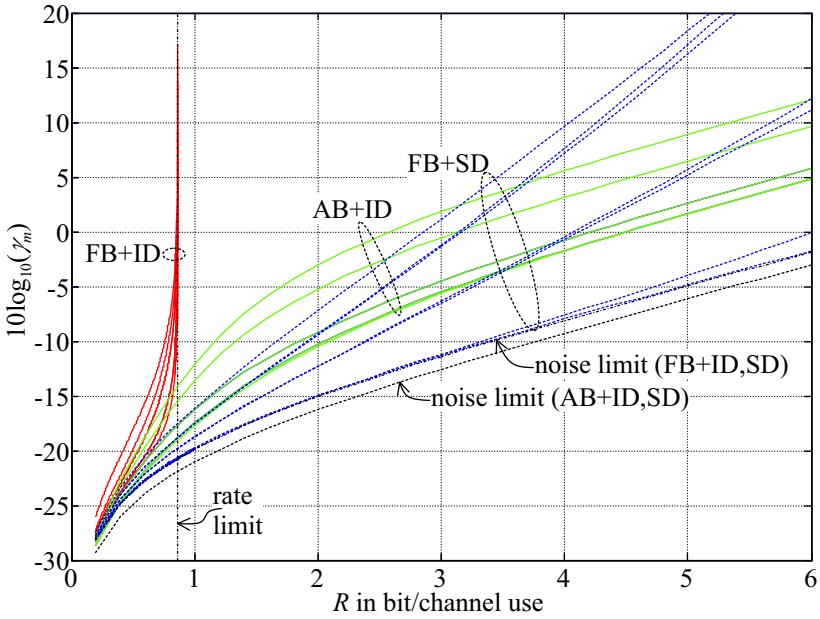


Figure 5.36: Comparison of required element SNR for fixed beam forming with independent decoding (FB+ID) (red), successive decoding (FB+SD) (blue), and optimal adaptive beam space beamforming with independent decoding (AB+ID) (green).

The explanation for the results observed here is as follows. Firstly, for target rates R below approx. 3 bit/channel use, the AB+ID receiver performs worse, because interference cannot be suppressed without significantly reducing at the same time also the gain for the wanted-signal (cf. Fig. 3.23), and, basically, thermal noise limits the interference reduction. However, as with increasing rates interference becomes dominant over thermal noise, the ability of adaptive beamforming to spatially filter interfering signals comes into effect.

Note that, as the rate R and therefore interference, grow larger, the receiver employing adaptive beamforming with independent decoding (AB+ID) will **always perform better** than the receiver

employing fixed beamforming with successive decoding (FB+SD). This holds, because as rates and input SNR tend to infinity, the AB+ID receiver places pattern nulls in the directions of interfering source signals, such that the required input SNR become independent of the interference power. This is confirmed by the plots shown in Fig. 5.37.

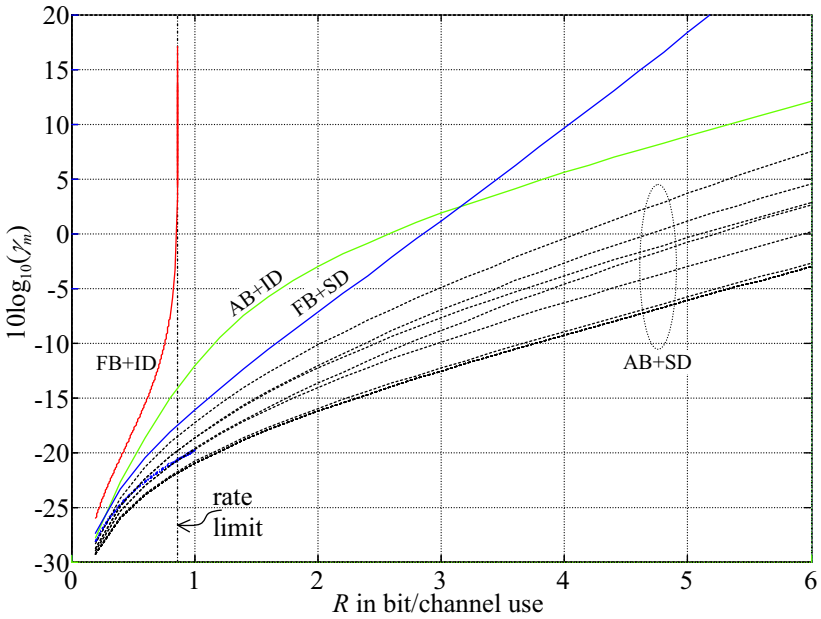


Figure 5.37: Required element SNR for optimal adaptive beam space beamforming with subsequent successive decoding (AB+SD). Also shown for comparison are the maximal element SNR for fixed beamforming with independent (FB+ID), successive decoding (FB+SD), and optimal adaptive beamforming with independent decoding (AB+ID).

For a concluding comparison we consider the respective maximum EIRP for the various receivers in relation with the peak EIRP that can be provided by the ICO terminals (cf. Fig. 5.38).

Clearly, the interval of rates where FB+SD outperforms AB+ID (approx. $0 \leq R \leq 3$ bit/channel use), is quite exactly the relevant interval defined by the peak EIRP of 6.8 dBW and 12 dBW, respectively.

The differences in required EIRP to achieve a desired rate R with the various receiver options are read off the graphs in Fig. 5.38 and compared in Tab. 5.7.

5.5 Summary

Based on Chap. 4, the aim of this chapter was a comparison of various receiver structures in exemplary communications satellite scenarios, employing fixed only or (optimal) adaptive beamforming, and independent or successive decoding.

Firstly, we have addressed the case of fully adaptive element space beamforming. After pointing out that beam space beamforming can in principle perform identical to element space beamforming, we have then focused on beam space beamforming, also because it comes with the possibility of partial adaptivity.

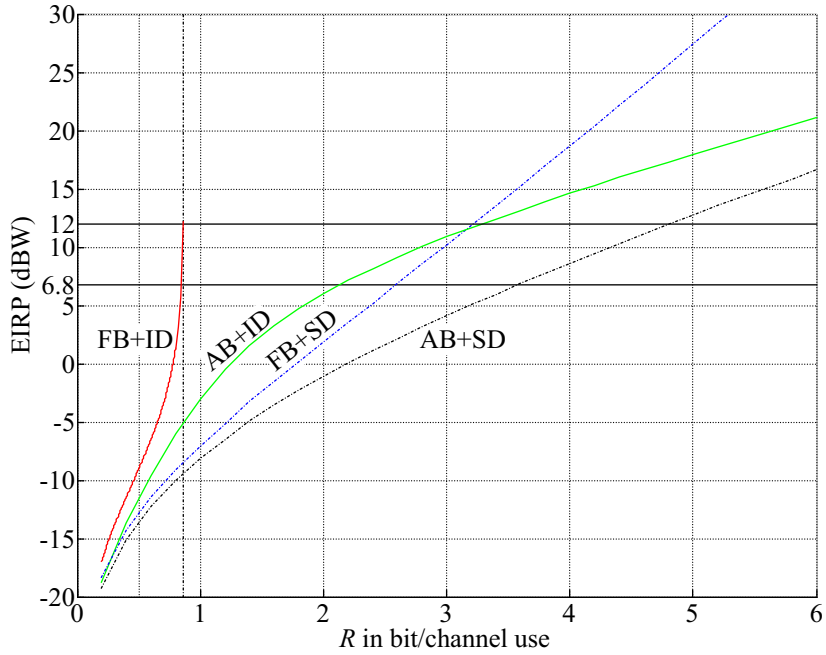


Figure 5.38: Comparison of the maximum required EIRP for optimal adaptive beam space beamforming with subsequent successive decoding (AB+SD), for fixed beamforming with independent (FB+ID), successive decoding (FB+SD), and optimal adaptive beamforming with independent decoding (AB+ID).

	Receiver type			
R bit/channel use	FB+ID	FB+SD	AB+ID	AB+SD
0.5	15.7	19.6	18.3	20.4
1	/	13.8	9.8	14.8
2	/	4.9	0.8	7.9
3	/	/	/	2.6

(a) Comparison of the link margin (in dB) with respect to lower EIRP limit 6.8 dBW.

	Receiver type			
R	FB+ID	FB+SD	AB+ID	AB+SD
0.5	4.7	0.8	2.1	0
1	∞	1	5	0
2	∞	3	6.3	0

(b) Increase (in dB) of required EIRP to achieve rate R for receivers FB+ID, FB+SD, and AB+ID with respect to receiver AB+SD.

Table 5.7: Comparison of optimal adaptive beam space beamforming with subsequent successive decoding (AB+SD), fixed beamforming with independent (FB+ID), successive decoding (FB+SD), and optimal adaptive beam space beamforming with independent decoding (AB+ID).

The performance comparison of the receiver employing fixed beamforming with independent decoding (FB+ID), common today, with the more complex receivers with adaptive beamforming or successive decoding, or both (AB+ID, FB+SD, and AB+SD) verified that significant gains can be

realised in terms of achievable rates, power and spectrum efficiency.

The comparison of AB+ID and FB+SD was considered of particular interest, because both receiver options are implemented with less complexity than the optimal AB+SD receiver, but possibly achieving similar performance. Further, these options are discussed for the satellite application independently in various publications (e.g. [Ern99, Ern01, Bjö93, LL96]), but there was no publication that would deal with a comparison of both receiver options AB+ID and FB+SD together.

This comparison was presented for the first time in this chapter, based on the common framework of resource allocation in the fading vector MAC presented in Chap. 4.

We have observed that under the assumption of sufficient source separation, AB+ID would outperform FB+SD, because,

- firstly, with AB+ID the edge-of-cell loss suffered with fixed beams (FB) can be recovered (yielding a noise limit 3 dB-4.3 dB lower, depending on the edge-of-cell gain), and,
- secondly, interferers are possibly ideally suppressed (then independent decoding (ID) performs equally good as more complex successive decoding (SD)).

On the other hand, if the angular source separation is smaller than the (one-sided) beamwidth, adaptive beamforming cannot efficiently suppress interference without compromising at the same time the gain for the wanted signal. Then, depending on the aimed at target rate R, it may be the case that AB+ID is inferior to FB+SD, which removes the interference by successive decoding without reducing the gain for the wanted signal.

Chapter 6

Implementation Considerations

In the previous section the different receiver options as sketched in Fig. 4.1 have been analysed and compared in an information theoretical sense. This is a very general approach that is mostly independent of particular implementations, and is based on assumptions regarding the channel model, availability of channel state information at the transmitters and receivers side, and the exploitation of available information at the receiver (is *all* available information used, or do we choose to discard some information in favour of a simpler receiver structure?). In that sense, all the different receiver structures are fully characterised by the respective expression for the achievable rates, providing strict limitations to maximum transmission rates for given transmit power constraints and, equivalently, minimum transmit powers required to achieve given rates. Therefore, the rate and power regions are considered as a reasonable basis for comparison of the potential performance of the different receiver structures.

However, it is evident that for a significant comparison of the different receiver structures, not only performance should be considered, but also complexity has to be included.

In the following section a more detailed, implementation-oriented characterisation is presented which will allow a basic comparison of complexity for the considered receiver structures.

6.1 Implementation Aspects Outline

Four separate signal processing stages in the receiver can be identified [Bjö93]:

- The analog front-end, consisting of the antenna array, amplifiers, filters, and down converters.
- The digital front-end including analog-to-digital conversion, comprising digital down-conversion, sample rate conversion, and channelisation.
- The digital beamforming (both fixed, as well as adaptive, and combinations). Computation of adaptive beamforming weights has to be considered, further computation of the scalar product of weight vector and signal vector (i.e. array output vector and fixed beamforming network output vector, respectively).

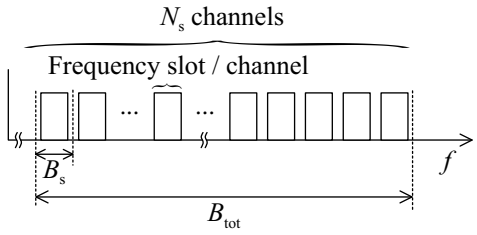


Figure 6.1: Definition of frequency slot/channel bandwidth B_s, total bandwidth B_tot, and total number of channels/slots N_s.

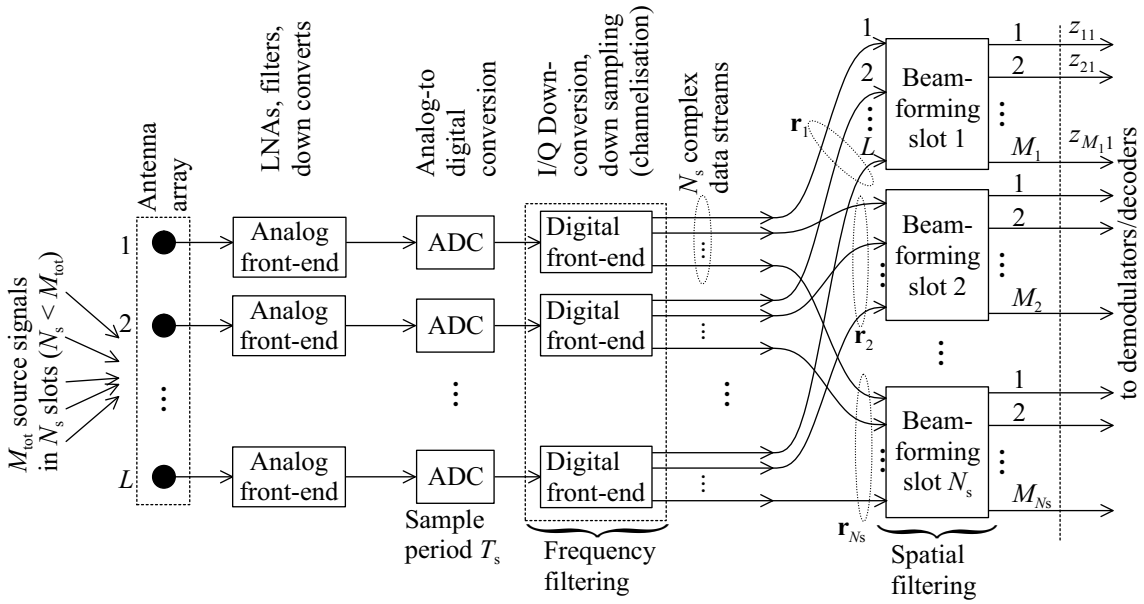


Figure 6.2: Overview of receiver, including analog front-end, digital front-end, and beamforming. The input data stream to a single digital front-end has sample rate $2B_\mathrm{tot}$, the N_s output data streams have each sample rate $2B_\mathrm{s}$. Generally, the digital signal processing including the beamforming is transparent to the subsequent demodulation and decoding. The signal vectors $\mathbf{r}_n \in \mathbb{C}^L$, $n = 1, 2, \ldots, N_\mathrm{s}$, are equivalent to signal vector $\mathbf{r}$ defined in (3.70). Each $\mathbf{r}_n$ contains the source signals of the M_n sources sharing frequency slot n, i.e. $\sum_{n=1}^{N_\mathrm{s}} M_n = M_\mathrm{tot}$. Each of the M_n output data streams of the n-th beamformer, $z_{1n}, z_{2n}, \ldots, z_{M_n n}$, still contains M_n source signals, but with the interference reduced by spatial filtering in the beamformers.

- The demodulator/decoder stage (digital).

In general, channelisation (also: demultiplexing) is understood as the process of extracting single channels from an input signal containing several channels, for further processing at baseband, thus mainly involving downconversion, and frequency filtering [HHF99].

For the following some definitions are introduced (cf. Fig. 6.1). Let B_s be the bandwidth of a single frequency slot (we will refer to this also simply as a channel) including guard bands and N_s the total number of frequency slots. Then $B_\mathrm{tot} = N_\mathrm{s} B_\mathrm{s}$ is the total bandwidth.

Fig. 6.2 essentially shows the analog receiver section, the digital front-end, and, further, the digital beamforming stage. The receive signal of any antenna array element is a superposition of all M_tot source signals (cf. (3.70)). Each source signal occupies one out of N_s frequency slots, and

to increase spectrum efficiency, several source signals may share a common frequency slot. The number of sources sharing the same slot n, $n = 1, 2, \ldots, N_\mathrm{s}$, is denoted with M_n.

After the receive signal of an antenna array element is amplified, frequency filtered, and, possibly, downconverted to an intermediate frequency (IF) in the analog-front end, it is converted to the digital domain by an analog-to-digital converter (ADC). Sampling in the ADC according to Nyquist rate guarantees that the receive signal is completely captured by the digital output data stream of the ADC (if oversampling is required at a later stage of the receiver, e.g. for matched filtering, this can be realised by upsampling the digital signal as required [HHF99, Gar93, HR01]; other notations in the literature used synonymously for *upsampling* are *interpolation*, *sampling-rate conversion*). Eventually the sampling rate can be significantly reduced by employing bandpass sampling [VSW91, HHF99].

This data stream is then further processed in the digital front-end, which performs channelisation of the N_s frequency slots contained in the receive signal of every antenna array element. For clarification Fig. 6.3 shows schematically the process of channelisation in the digital front-end. There it is depicted that a signal of bandwidth B_tot is the input to the digital-front end, and that this signal is channelised in N_s complex output data streams each of bandwidth B_s (sample rate is $2B_\mathrm{s}$).

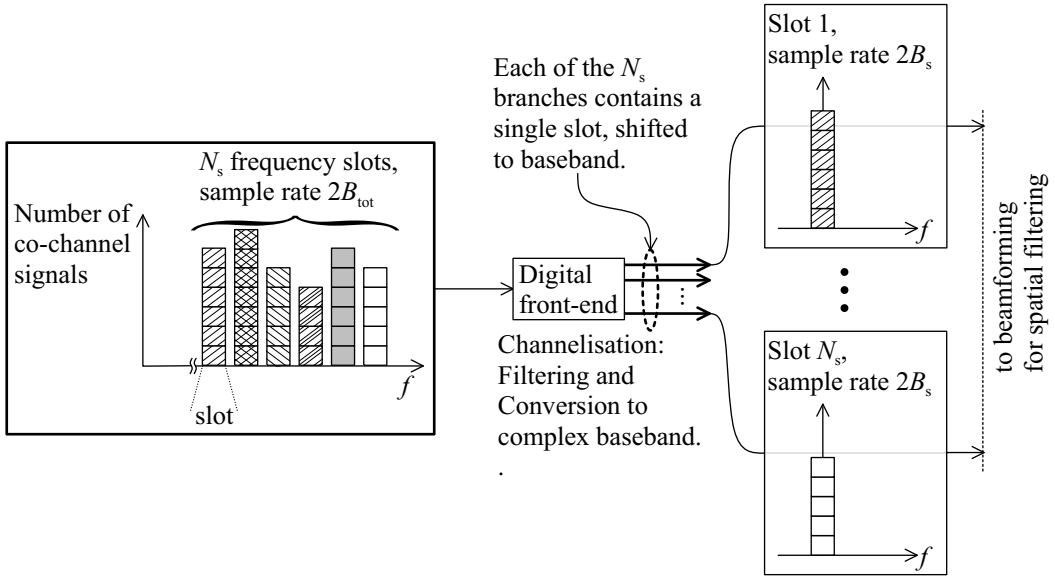


Figure 6.3: Schematic representation of channelisation in the digital front-end (cf. Fig. 6.2). At the input of the digital front-end there are N_s frequency slots, each occupied by one or more source signals. N_s output data streams are generated by channelisation, whereas each data stream contains a single frequency slots shifted to baseband. Channelisation is performed by frequency filtering and downconversion to baseband, independent of the number of co-channel signals eventually sharing a slot. Mutual interference from source signals sharing the same frequency slot is reduced by the spatial filtering in subsequent beamforming stages.

From Fig. 6.2 it is evident that the analog front-end is identical for all L array elements, and for all considered receiver structures as presented in Chap. 5. Therefore the analog front-end is irrelevant for comparison between the different receiver structures, and we will not go into further detail here

(for a detailed description of a feasible[1] approach for the analog front-end to be combined with subsequent digital signal processing see [HHF99]).

Further it suffices to consider the complexity of processing the signals of a single frequency slot n (again cf. Fig. 6.2), because the digital processing is the same for every slot n and mutual independent.

However, evidently a difference between the receiver structures arises from differences in the beamforming. Further, subsequent to the beamforming, demodulation and decoding is performed, which will be dealt with in the required detail at a later point in this work.

According to the general receiver structure just introduced the following complexity drivers can be identified:

- Digital beamforming, including computation of beamformer weights for adaptive beamforming.
- Decoder complexity (independent and successive decoding).

Therefore, the complexity for the considered receiver structures will be defined based on the complexity required for digital beamforming, and for demodulation/decoding.

Note that in [Bjö93] it is stated that channelisation is in fact the actual complexity driver in a digital signal processing architecture as depicted in Fig. 6.2, involving an order of magnitude more processing load than the beamforming. However, because channelisation down to single frequency slots is in any case required before demodulation and decoding, complexity of channelisation needs not to be included in the receiver structure comparison.

6.2 Receiver Complexity Outline

This section aims to provide a general outline of the different building blocks that are required in particular for the different receiver variants. This allows a coarse comparison of the receiver variants, however not concluding in detailed quantitative measures of receiver complexity (e.g. using floating point operations per second).

Fig. 6.4 shows the schematic block diagrams for the receiver options FB+ID, AB+ID, FB+SD, and AB+SD, introduced in Chap. 5, side by side to better point up the similarities and differences.

All receiver options require the same fixed beamforming stage, as well as single-user decoding (SUD) for each source. Also required for all receiver variants is the channel estimation that needs to be implemented for coherent demodulation and decoding of the source signals in the M single-user decoders, as stated in Sec. 4.2.2.

For successive decoding, re-encoding and re-modulation of the decoded bitstream is required to produce the estimates of the code symbols s_m. However, usually the re-encoded bitstream can

[1] It is desirable to place the ADC as close as possible to the antenna (eventually after amplification and anti-aliasing filtering, but without any analog downconversion), but this approach seems, at least today, not viable due to limitations of available ADCs

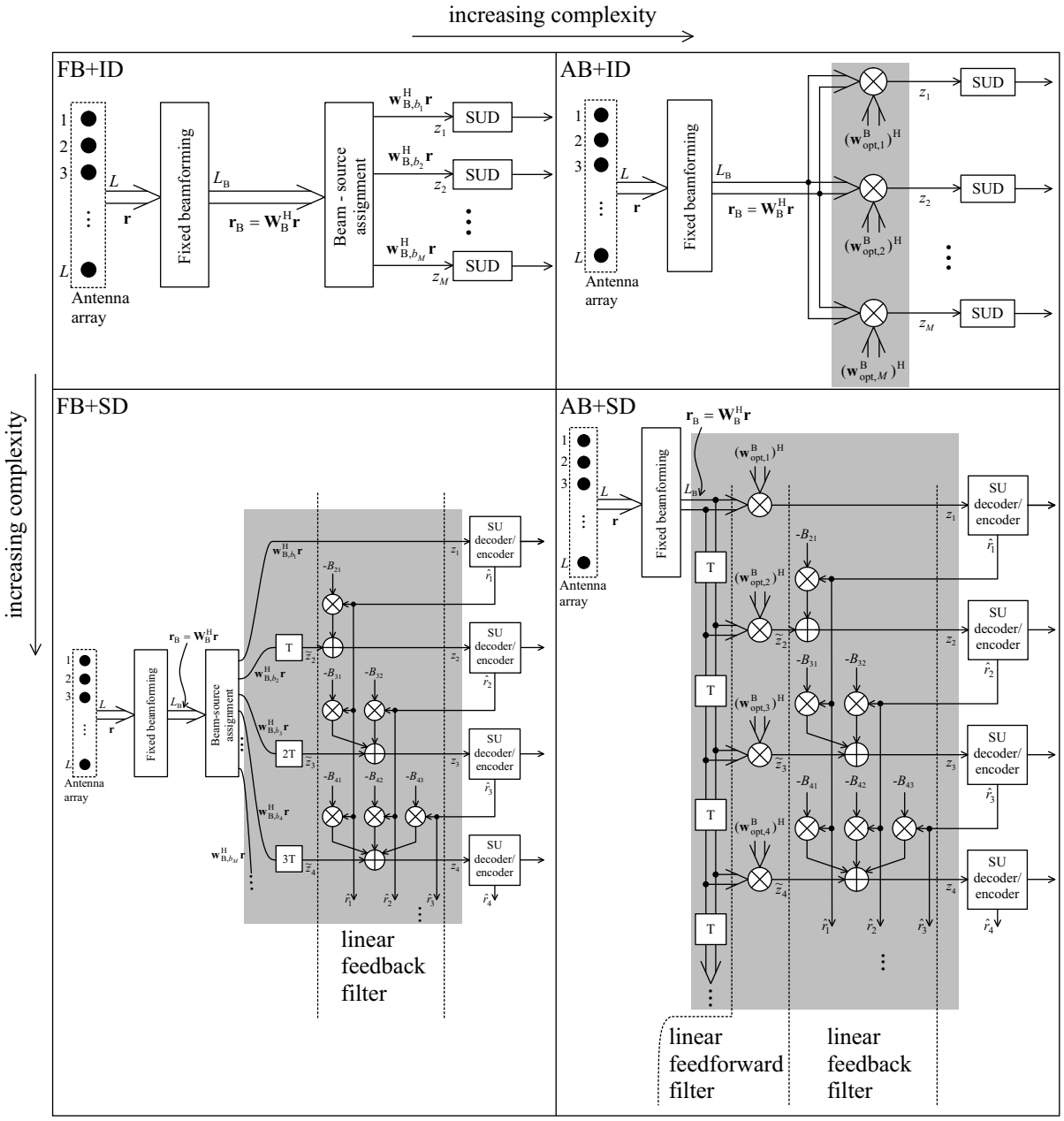


Figure 6.4: Receiver structures for fixed beamforming with independent decoding (FB+ID, Sec. 5.3.1), (optimal) adaptive beamforming with independent decoding (AB+ID, Sec. 5.2.1), fixed beamforming with successive decoding (FB+SD, Sec. 5.3.2), and adaptive beamforming with successive decoding (AB+SD, Sec. 5.2.1). Highlighted grey are the central differences with respect to FB+ID. The delay elements in the receiver structures employing successive decoding (SD) take care of the decoding delay introduced by the single-user decoders (SUD).

be made available in the decoding process as a byproduct (e.g. for the Viterbi decoder [Pro95]) without significant additional effort, and also complexity of mapping the re-encoded bits to code symbols s_m is considered negligible.

Further, the beam-source assignment can be considered of negligible complexity as its task is simply to choose 1 out of the L_{B} data streams at the output of the fixed beamforming stage.

Therefore, as pointed out in the last section, the building blocks of the receivers adding to complexity with respect to FB+ID are adaptive beamforming and successive decoding (highlighted grey in Fig. 6.4).

FB+ID This receiver employs fixed beamforming only, without any additional adaptive beam space processing, in combination with single-user decoding.

The receive symbols z_m that are available to the single-user decoder for source m are according to (5.17) given by

$$z_m = \underbrace{\mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_m \frac{h_m}{\sqrt{\mu_m}} s_m}_{\text{wanted signal}} + \underbrace{\sum_{\substack{i=1 \\ i \neq m}}^{M} \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}} s_i}_{\text{interference}} + \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{n}.$$

For coherent demodulation and decoding of the code symbols s_m (cf. Sec. 4.2.2), an estimate of the channel factor (including beamforming in addition to the transmission channel)

$$\alpha_m = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_m \frac{h_m}{\sqrt{\mu_m}} \quad \text{(fixed beamforming)} \tag{6.1}$$

must be provided by means of channel estimation. This allows to accordingly scale the receive symbol z_m by α_m to compensate for amplitude variation and phase shift that alter the source symbol s_m along the transmission path. If the time-variance of α_m is sufficiently slow, regular transmission of training sequences known to the receiver can be employed to yield the required estimates of α_m.

The amount of interference introduced for source m from all other code symbols s_i, $i \in \mathcal{M} \setminus m$, is determined by the interference channel factors α_{mi}, which are according to (5.17) given by

$$\alpha_{mi} = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}} \quad \text{(fixed beamforming)}, \tag{6.2}$$

however, these need not to be estimated for independent decoding.

AB+ID This receiver variant comprises a building block for adaptive beamforming, which requires computation of the (optimal) beamforming weights $\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}$ and beamforming itself (i.e. computation of the scalar product $\left(\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}\right)^{\mathrm{H}} \mathbf{r}_{\mathrm{B}}$). This has to be performed for each of the M sources.

As for the FB+ID receiver estimation of the M channel factors α_m is required here, where the α_m are here given by

$$\alpha_m = \left(\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}\right)^{\mathrm{H}} \mathbf{r}_{\mathrm{B}} \frac{h_m}{\sqrt{\mu_m}} \quad \text{(adaptive beamforming)}. \tag{6.3}$$

Partially adaptive beamforming may help to reduce the related complexity as discussed in Sec. 5.2.2.

In a time-variant signal environment the beamforming weight vectors (and also the coefficients of the feedback filter) have to be updated according to the rate of change of the signal

scenario. Three factors determine the required update rate for the weight vectors, if a tolerable performance degradation due to a delayed update of the beamforming weights is given. These factors are:

- Rate of change of the fading factors h_m.
- Rate of change of the steering vectors $\mathbf{a}_m$.
- Changes in the interference scenario: sources may start and stop using a frequency slot for communications.

FB+SD As this receiver variant does not implement adaptive beamforming, i.e no feedforward filter is implemented, the successive decoding requires a linear feedback filter only.

The input symbols z_m to the m-th single-user decoder, after interference from sources already decoded is removed, are according to (5.28) given by (without loss of generality, the decoding order is $1, 2, \ldots, M$)

$$z_m = \underbrace{\alpha_m s_m + \sum_{\substack{i=1 \\ i \neq m}}^{M} \alpha_{mi} s_i + \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{n}}_{= \tilde{z}_m} - \sum_{i=1}^{m-1} \underbrace{B_{mi} \frac{h_i}{\sqrt{\mu_i}}}_{= \alpha_{mi}} s_i, \tag{6.4}$$

where the channel factor α_m and the interference channel factors α_{mi} were defined in (6.1) and (6.2), respectively, and, further, the coefficients B_{mi} of the feedback filter are according to (5.27) given by

$$B_{mi} = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_i.$$

The code symbols s_i ($i = 1, 2, \ldots, m-1$) are known from the decoder stages of the successive decoder preceding that of source m, but additional effort is required to provide an estimate of the channel factors α_{mi} (cf. (6.4)) which determine the interference contribution of sources $i = 1, 2, \ldots, m-1$ for the received symbol z_m.

It may be significantly easier to provide an estimate of the α_m and α_{mi}, respectively, than it is to provide estimates for the B_{mi} and $h_m/\sqrt{\mu_m}$ separately, because channel estimation based on a training sequence cannot distinguish between the influence of the channel ($h_m/\sqrt{\mu_m}$) and the beamforming (B_{mi}).

According to the structure of the feedback filter it is required to estimate $M(M-1)/2$ interference channel factors α_{mi}. The required effort for estimation of the factors α_{mi} with sufficient accuracy based on a training sequence known to the receiver, depends obviously on the respective SINR for source i in beam b_m (where effort means required transmit power and training sequence length).

Finally, the input symbols z_m to the single-user decoders have to be computed from the symbols $\tilde{z}_m$ according to (cf. (6.4))

$$z_m = \tilde{z}_m - \sum_{i=1}^{m-1} \alpha_{mi} s_i, \tag{6.5}$$

which requires $M(M-1)/2$ complex multiplications and additions.

The delay elements that are introduced for successive decoding (i.e. also for the AB+SB receiver) are required to compensate for the decoding delays T.

AB+SD Here, both adaptive beamforming (linear feedforward filter) as well as successive decoding (linear feedback filter) are employed.

Computation of the optimal adaptive weight vectors $\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}$ and the scalar product

$$\tilde{z}_m = \left(\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}\right)^{\mathrm{H}} \mathbf{r}_{\mathrm{B}}$$

has to be performed for all M sources to implement the feedforward filter part of the AB+SD receiver variant.

As for the FB+SD receiver the feedback filter part requires computation of $M(M-1)/2$ factors α_{mi}, which are in this case defined as

$$\alpha_{mi} = \left(\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}\right)^{\mathrm{H}} \mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}} \quad \text{(adaptive beamforming)}. \tag{6.6}$$

Again $M(M-1)/2$ complex multiplications and additions are required to obtain the symbols z_m from the $\tilde{z}_m$.

Summarising, the required additional implementation effort for the receiver variants AB+ID, FB+SD, and AB+SD in comparison with FB+ID is given as follows:

AB+ID

- Estimation of M adaptive beamforming weights $\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}$.
- Computation of M scalar products $\left(\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}\right)^{\mathrm{H}} \mathbf{r}_{\mathrm{B}}$: M times L_{B} (number of beams) complex multiplications and complex additions.

FB+SD

- Estimation of $M(M-1)/2$ interference channel factors $\alpha_{mi} = \mathbf{w}_{\mathrm{B},b_m}^{\mathrm{H}} \mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}}$.
- Computation of $z_m = \tilde{z}_m - \sum_{i=1}^{m-1} \alpha_{mi} s_i$: $M(M-1)/2$ complex multiplications and complex additions.

AB+SD

- Estimation of the M optimal weight vectors $\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}$.
- Computation of M scalar products $\left(\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}\right)^{\mathrm{H}} \mathbf{r}_{\mathrm{B}}$: M times L_{B} complex multiplications and complex additions.
- Computation of $M(M-1)/2$ interference channel factors $\alpha_{mi} = \left(\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}\right)^{\mathrm{H}} \mathbf{a}_i \frac{h_i}{\sqrt{\mu_i}}$.
- Computation of $z_m = \tilde{z}_m - \sum_{i=1}^{m-1} \alpha_{mi} s_i$: $M(M-1)/2$ complex multiplications and complex additions.

The above summary allows a qualitative comparison of the expected implementation complexity and presents the starting point for comprehensive investigations into receiver complexity.

In particular, no detailed assumptions are made here regarding the algorithms for computation of the adaptive beamforming weights and the interference channel factors for successive decoding. A concluding in-depth comparison would require the identification of particular algorithms for weight computation for adaptive beamforming and channel estimation that are suitable for the envisaged satellite scenario. However, this would go beyond the scope of this work and such a detailed complexity analysis is opening future research.

Chapter 7

Conclusions

In this chapter, the main goals and most important results of this thesis are outlined. (A summary in German language can be found towards the end of this chapter.)

Summary

In this thesis, the uplink of the user link in a satellite communications system is considered, where at the satellite a direct radiating array (DRA) is employed. This scenario is described by the fading vector multiple access channel (MAC). In particular, the subject of investigation are various receiver options for the fading vector MAC.

The various receiver variants are encountered depending on whether fixed or (optimal) adaptive beamforming is performed with the antenna array receive signals, and, further, whether independent or successive decoding is implemented.

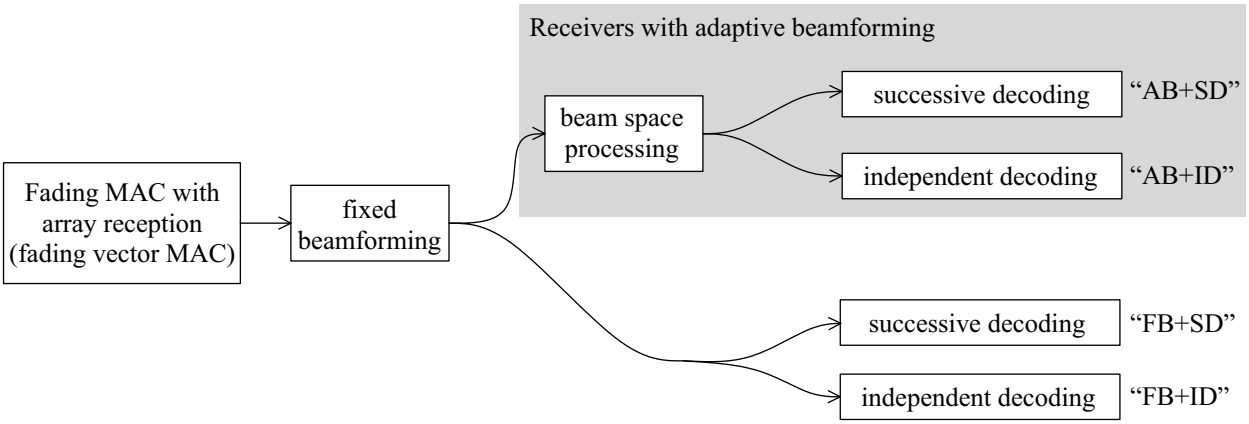


Figure 7.1: Outline of the considered receiver/decoder options. AB: adaptive beamforming, FB: only fixed beamforming, SD: successive decoding, ID: independent decoding.

The main objective of this thesis is a performance comparison of the resulting four receiver variants, which are (cf. Fig. 7.1)

- fixed beamforming with independent decoding, FB+ID,
- fixed beamforming with successive decoding, FB+SD,

- adaptive beamforming with independent decoding, AB+ID, und
- adaptive beamforming with successive decoding, AB+SD.

The areas of investigation of this thesis are:

- A unified description of the investigated receiver variants.
- The resource allocation problem of allocating transmit powers to the sources in order to achieve required information rates for all sources.
- Performance of the receivers employing adaptive beamforming and/or successive decoding (FB+SD, AB+ID, AB+SD) in comparison with the receiver being standard today, relying on fixed beamforming and independent decoding (FB+ID).
- Comparison of the performance of a receiver employing both adaptive beamforming and subsequent successive decoding (AB+SD) with a receiver employing only either adaptive beamforming or successive decoding (FB+SD, AB+ID) for better suppression of interference.
- Comparison of the performance of a receiver employing fixed beamforming with successive decoding (FB+SD) with that employing adaptive beamforming with independent decoding (AB+ID).
- Impact of the respective interference scenario on the performance of the considered receiver variants.

In the following, the main areas of investigation are outlined in more detail, further the main results are summarised.

Chapter 2 Among other things, the spatial characteristics of the satellite communications channel are addressed as this is in literature only available for the terrestrial case. It is pointed out that it can be assumed that a source signal emitted by a terrestrial terminal arrives at the satellite from a single discrete direction-of-arrival (DOA) in a plane wave front. Knowledge of the spatial characteristics is required, because the antenna array multi-user signal model developed in Chap. 3 has to include the spatial characteristics of the channel.

Chapter 3 This chapter is dedicated to the provision of antenna fundamentals (directivity, gain, radiation pattern etc.), and further of the antenna array multi-user signal model, which is required for the investigation of the receiver variants.

Two exemplary satellite systems are introduced in Sec. 3.6.1 and 3.6.2 that are used in Chap. 5 to compare the performance of the various receiver options. For the ICO satellite scenario (a medium earth orbit (MEO) system) detailed measurement data of the employed array antenna are available from publications and are used to successfully verify the satellite antenna array model developed in Sec. 3.4. Further, for EuroSkyWay (a geostationary orbit (GEO) system) an optimal antenna array model is developed, maximising the array gain at the edge of coverage.

Chapter 4 Here foundations are laid to a unified treatment of the said receiver options in an information theoretic framework. For this the definition of Shannon capacity of the fading vector MAC is provided, under the assumption that transmitters (sources) can adapt transmit powers only to the slowly varying, quasi-static attenuation factor, but not to the possibly fast-changing fading factors. In contrast, the common receiver of the source signal has ideal channel state information (CSI), i.e. knowledge of both the attenuation and the fading factors of all sources is available for coherent demodulation and decoding of the source signals.

The main part of this chapter is dedicated to an analysis of the resource allocation problems for the cases of, firstly, optimal rate allocation maximising sum-rates for given maximum transmit power constraints and, secondly, optimal transmit power allocation minimising transmit sum-power for required rates.

The thesis expands available literature considerably by discussing optimal resource allocation for the fading vector MAC in the framework of linear and non-linear convex programs. Basis for this are the definitions of the rate region (region of achievable rates) and power region for the fading vector MAC. The rate region and the power region, respectively, are the feasible sets for the said resource allocation problems.[1]

A task of considerable complexity is the power allocation problem, i.e. when required rates to be provided to the sources are given and the optimal transmit powers are sought.

For the first time ever this is analysed in this thesis in great detail for the fading vector MAC (cf. Sec. 4.5 and 4.6).

Related to this it is shown that the power region of the fading vector MAC is a polymatroid only for the case that the fading factors are constant (resulting in an AWGN channel) and only if all source signal steering vectors are identical (resulting in the classical MAC), allowing a greedy solution to the power allocation problem. The other case yielding a simple solution to the power allocation problem is encountered when all steering vectors are orthogonal, resulting in independent channels without interference, such that independent decoding is optimal.

The main results of the analysis of the resource allocation problem in the fading vector MAC are as follows:

In the general case of fading and unequal steering vectors the power region is no polytope. However, the power region is convex, such that optimal power allocation can be represented as a non-linear convex program, with the optimal solution being obtained using standard optimisation algorithms. The optimal power allocation may not be obtained at a vertex, such that successive decoding is possibly suboptimal. Nevertheless, due to the low-complexity implementation of successive decoding compared to joint decoding, while achieving rates at the boundary of the rate region, successive decoding is still attractive. Therefore the problem of finding the optimal vertex (i.e. the optimal decoding order) for successive decoding in the fading vector MAC is investigated.

As a central result of this work, it is shown that there cannot be a greedy algorithm to identify the optimal decoding order for successive decoding for the fading vector MAC. This is

[1] In an optimisation problem that can be expressed as a linear program, the feasible set is a polytope (the feasible set defines the region that the optimal solution must be element of). The optimal solution is always obtained at a vertex of the feasible set and if the feasible set is a polymatroid (a polytope with certain properties), then the optimal vertex can be obtained from a low-complexity greedy algorithm.

explained by using a tree graph representation of the procedure of successive decoding and pointing out that finding the optimal minimum sum-weight path (corresponding to the minimal transmit sum-power) through the tree graph can only be obtained by exhaustive search, i.e. by evaluating all possible paths (= decoding orders) through the tree. It is concluded that for the fading vector MAC, the procedure to identify the optimal decoding order for successive decoding is in general complete enumeration.

Further, fixed beamforming and its impact on achievable rates is discussed. It is pointed out that fixed beamforming is always lossless (i.e. that achievable rates with fixed beamforming are identical to achievable rates employing the antenna array without fixed beamforming), if the number of fixed beams (orthogonal or non-orthogonal) is equal or greater than the number of array elements, provided that the fixed beamforming weight vectors are all linearly independent.

Chapter 5 In Sec. 5.1.1 the impact of the decoding order for successive decoding on the transmit powers is discussed. For this a receiver employing optimal adaptive beamforming with successive decoding and an exemplary source distribution with 7 sources is assumed, with 6 sources arranged on the corners of a hexagon and a single source in the centre. The required transmit powers are computed for different values of angular source separation. It is observed that the required transmit powers can vary strongly with the particular decoding order, where the degree of variation depends on the targeted information rate and on the angular separation of the interfering sources. For the considered source distribution and for an exemplary target information rate of $R = 1$ bit/channel use, the optimal decoding order can yield maximum transmit powers being up to approx. 4 dB lower than the maximum transmit power for the worst-case decoding order. For higher targeted information rates the savings in transmit power increase, as they decrease for lower information rates (e.g. approx. 2 dB for $R = 0.5$ bit/channel use). It is concluded that the decoding order can have a significant impact on the maximal transmit powers. However, depending on the target information rate and on the source separation, the increase in transmit powers due to a non-optimal decoding order may be small.

Assuming the same source distribution with variable angular separation, performance of receivers employing optimal adaptive beamforming with independent and successive decoding is compared based on the element signal-to-noise-power-ratio (SNR) and the source equivalent isotropically radiated power (EIRP), required to achieve a certain target information rate. It is observed that with successive decoding the sources can be located significantly closer to each other than with independent decoding, while equal maximal element SNR and EIRP, respectively, are maintained. However, it is also observed that the advantage of successive decoding in terms of reduced maximum transmit powers vanishes for increasing target rates, because the adaptive beamforming stage has to suppress interference more and more efficiently in order to maximise the SINR, such that there is no additional benefit obtained with successive decoding.

Because the receivers AB+ID and AB+SD implement adaptive beam space beamforming, partial adaptivity is discussed in Sec. 5.2.2, where only a subset of the fixed beams is employed in the adaptive beamforming stage. In particular, it is investigated what differences arise from partial adaptive beam space beamforming based on orthogonal and non-orthogonal fixed beams. It is shown with the ICO satellite scenario that partial adaptivity using non-orthogonal beams is advantageous over using orthogonal beams, because in the

average a smaller number of fixed non-orthogonal beams is required to achieve the same level of performance.

Further, it is known for the classical MAC that with independent decoding there is a rate limit, which marks a maximum rate that can be achieved at most for all sources if the transmit powers tend to infinity (i.e. as the transmit powers tend to infinity, the achievable rate tends to a finite value, namely the rate limit). In Sec. 5.3.1 it is pointed out that also for the FB+ID receiver there is such a maximal rate limit, although interference is suppressed by the gain pattern of the fixed beams.Finally, a method is specified to compute this rate limit. There is no such rate limit for the other receiver options, i.e. as the transmit powers tend to infinity, so does the information rate achievable for the sources.

The main outcome of this chapter is a comparative investigation of the aforementioned different receiver options in satellite communications scenarios, namely fixed beamforming with independent decoding (FB+ID), fixed beamforming with successive decoding (FB+SD), fixed beamforming with subsequent adaptive beam space beamforming with independent decoding (AB+ID), and adaptive beamforming (again subsequent to fixed beamforming) with successive decoding (AB+SD).

In fact, this thesis provides for the first time a systematic comparison of the said receiver options in satellite communications scenarios, based on the resource allocation methodology for the fading vector MAC as presented in Chap. 4.

Finally, the aimed at comparison of the receiver options requires the definition of specific interference scenarios, because the mutual interference together with the target information rate is the decisive factor determining the performance differences of the various receiver configurations. The mutual interference is determined by the particular assumptions concerning the gain pattern of the fixed beams that cover the service area of the considered satellite system, and the source distribution. Therefore, various interference scenarios are defined, based on the two satellite systems introduced in Chap. 3, considering worst-case source distributions according to the considered frequency reuse scheme and cluster size.

The main results of the described investigations are as follows.

Generally it holds that FB+ID performs worst, while AB+SD performs best considering power efficiency, i.e. for a given target information rate, FB+ID requires highest element SNR[2], while AB+SD requires lowest element SNR. This is not surprising as FB+ID implements no adaptive interference mitigation technique at all, whereas AB+SD incorporates both. Of particular interest is a comparison of FB+SD and AB+ID as these receiver options implement only one of the two interference mitigation techniques, hence being implemented with less complexity compared to the AB+SD receiver. These receiver variants are suggested in various publications concerned with a better suppression of interference and the related potential for higher bandwidth efficiency in satellite scenarios, but have been investigated independently so far. Further, it is up to now unknown under what circumstances an additional advantage concerning interference mitigation can be gained with the more complex AB+SD receiver in comparison to the FB+SD and AB+ID variants.

Firstly, on basis of the considered interference scenarios it is observed that for FB+SD and AB+ID there is no general superiority over one or the other receiver option.

[2]The element SNR is the SNR at a single array element.

If the source separation is sufficiently large, then the AB+ID receiver is superior to the FB+SD receiver. This holds, because, in contrast to fixed beamforming, with adaptive beamforming, firstly, the edge-of-cell loss can be recovered, secondly, interference can be suppressed efficiently. In this case it holds also that the performance of the AB+ID receiver is very close to that of the more complex AB+SD receiver employing also successive decoding, because interference is already suppressed for the most part, such that successive decoding cannot achieve a significant additional advantage.

However, if the interfering sources are very closely spaced (in the order of the beamwidth of the fixed beams), the performance of the FB+SD receiver may exceed that of the AB+ID receiver for rates below a certain limit. Yet, if the target rate is successively increased, then there is always a certain rate limit beyond which AB+ID will eventually perform better than FB+SD what concerns the maximum element SNR. The explanation for these observations is as follows. Firstly, for target rates below a certain limit, the AB+ID receiver performs worse, because interference cannot be suppressed without significantly reducing at the same time also the gain for the wanted-signal, and, basically, thermal noise limits the interference reduction. However, as with increasing rates transmit powers are increasing as well, interference becomes dominant over thermal noise. Then the ability of adaptive beamforming to spatially filter interfering signals comes into effect. This holds, because the adaptive beamformer of the AB+ID receiver places pattern nulls in the directions of interfering source signals, such that the required element SNR become independent of the interference power.

Chapter 6 Aims to provide an outline of the implementation complexity of the different receiver variants.

Concluding, the main result of this thesis are summarised as follows:

- A method for the systematic comparison of the said receiver variants for satellite communications scenarios is presented, which is based on the discussion of the resource allocation problem for the fading vector multiple access channel (MAC).
- An analysis of the resource allocation problems of optimal rate allocation maximising sum-rates for given maximum transmit power constraints and optimal transmit power allocation minimising transmit sum-power for required rates is presented for the fading vector MAC. It is shown that there is in general no greedy algorithm to obtain the optimal decoding order for successive decoding for the vector fading MAC.
- Receivers implementing adaptive beamforming or successive decoding or both (AB+ID, FB+SD, and AB+SD) provide a significant advantage over the receiver employing fixed beamforming with independent decoding only (FB+ID), being standard in nowadays satellite communications systems. The performance differences between AB+ID, FB+SD, and AB+SD strongly depend on the considered interference scenario.
- The receiver variant AB+ID performs very close to the more complex AB+SD and outperforming the FB+SD receiver, if the interfering sources are moderately closely spaced (e.g. for cluster size 3).
- The receiver variant FB+SD may outperform AB+ID, if the interfering sources are very closely space (in the order of the beamwidth of the fixed beams). However, whether FB+SD

achieves a target rate with less worst-case transmit power depends on the target rate itself. Eventually, with successively increasing target rate, the AB+ID receiver variant performs always better than FB+SD in terms of maximal transmit powers.

Zusammenfassung

In dieser Arbeit wird die Aufwärtsstrecke (*uplink*) der Verbindung zwischen Benutzerterminal und Satellit (*user link*) betrachtet, wobei am Satelliten eine direkt-strahlende Gruppenantenne (*direct radiating array (DRA)*) eingesetzt wird. Dieses Szenario wird durch den Vektor-Mehrfachzugriffskanal mit Signalschwund (*fading vector multiple access channel (MAC)*) beschrieben. Gegenstand der Untersuchung sind insbesondere verschiedene Empfängervarianten für den Vektor-Mehrfachzugriffskanal mit Signalschwund.

Diese verschiedenen Empfängervarianten ergeben sich je nachdem, ob mit den Empfangssignalen der einzelnen Gruppenantennen-Elemente entweder nur feste oder zusätzlich adaptive Strahlformung (*fixed*, *adaptive beamforming*) durchgeführt wird, bzw. ob die Empfangssignale unabhängig voneinander oder sukzessiv dekodiert werden (*independent*, *successive decoding*).

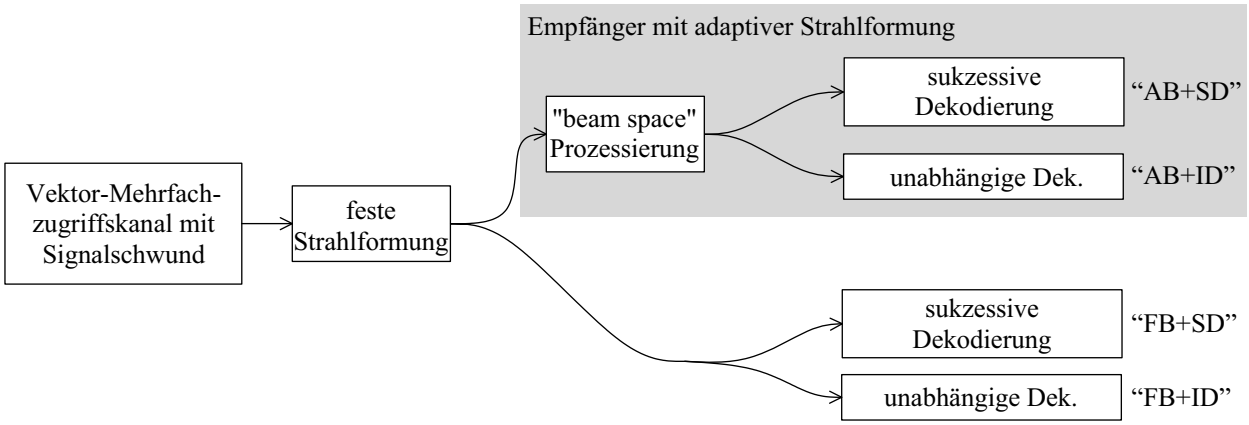


Figure 7.2: Übersicht über die betrachteten Empfänger/Dekodervarianten. AB: adaptive beamforming (adaptive Strahlformung), FB: fixed beamforming (feste Strahlformung), SD: successive decoding (sukzessive Dekodierung), ID: independent decoding (unabhängige Dekodierung). Adaptive Strahlformung erfolgt im “beam space”, d.h. im Anschluss an feste Strahlformung.

Das Hauptziel dieser Arbeit ist ein Vergleich der Leistungsfähigkeit der resultierenden vier Empfänger-Varianten. Diese sind (vgl. Abb. 7.2):

- feste Strahlformung mit unabhängiger Dekodierung (*fixed beamforming with independent decoding*, FB+ID),
- feste Strahlformung mit sukzessiver Dekodierung (*fixed beamforming with successive decoding*, FB+SD),
- adaptive Strahlformung mit unabhängiger Dekodierung (*adaptive beamforming with independent decoding*, AB+ID), und
- adaptive Strahlformung mit sukzessiver Dekodierung (*adaptive beamforming with successive decoding*, AB+SD).

Die Untersuchungsgebiete dieser Arbeit umfassen:

- Entwicklung einer allgemein gültigen Methodik zur einheitliche Beschreibung der untersuchten Empfängervarianten.
- Die Ressourcenzuteilung insbesondere im Zusammenhang mit dem Problem der Leistungszuteilung zu den Quellen (Sender), mit dem Ziel, geforderte Informationsraten für alle Quellen zu realisieren.
- Die Leistungsfähigkeit der Empfänger, die adaptive Strahlformung und/oder sukzessive Dekodierung einsetzen (FB+SD, AB+ID, AB+SD), im Vergleich mit der heutigen Standard-Empfängervariante, die auf fester Strahlformung mit unabhängiger Dekodierung basiert (FB+ID).
- Die Leistungsfähigkeit eines Empfängers, der sowohl adaptive Strahlformung, als auch anschließende sukzessive Dekodierung anwendet (AB+SD) im Vergleich mit einem Empfänger, der entweder nur adaptive Strahlformung (AB+ID) oder nur sukzessive Dekodierung (FB+SD) zur besseren Interferenzunterdrückung einsetzt.
- Die Leistungsfähigkeit eines Empfängers mit fester Strahlformung und sukzessiver Dekodierung (FB+SD) im Vergleich mit einem Empfänger mit adaptiver Strahlformung und unabhängiger Dekodierung (AB+ID).
- Einfluss des jeweiligen Interferenzszenarios auf die Leistungsfähigkeit der betrachteten Empfängervarianten.
- Betrachtung der Komplexität der verschiedenen Empfängervarianten.

Im folgenden werden die zentralen Fragestellungen in größerer Detailtiefe umrissen und die Hauptergebnisse zusammengefasst.

Kapitel 2 Hier wird unter Anderem die räumliche Charakteristik des Kommunikationskanals für Satellitenszenarien beleuchtet, weil dieser Aspekt in der Literatur nur für terrestrische Szenarien diskutiert wird. Es wird gezeigt, dass ein von einer terrestrischen Quelle ausgesendetes Signal am Satelliten aus einer einzigen diskreten Richtung (d.h. mit ebener Wellenfront) eintrifft. Das Wissen um die räumliche Charakteristik wird für das in Kapitel 3 entwickelte Mehrbenutzer-Signalmodel unter Einbeziehung der Gruppenantenne benötigt.

Kapitel 3 Dieses Kapitel dient dazu, Grundlagen der Antennentheorie (Direktivität, Gewinn, Richtcharakteristik etc.) zu vermitteln und das für die Untersuchung der Empfängervarianten benötigte Gruppenantennen-Mehrbenutzer-Signalmodel einzuführen.

Zwei exemplarische Satellitensysteme werden in Abschnitt 3.6.1 und 3.6.2 beschrieben, die in Kapitel 5 dazu dienen, die Leistungsfähigkeit der verschiedenen Empfängervarianten zu vergleichen. Für das ICO Satellitensystem (ein medium earth orbit (MEO) System) stehen aus Veröffentlichungen detaillierte Messdaten für die verwendete Satelliten-Gruppenantenne zur Verfügung. Diese werden herangezogen, um das in Abschnitt 3.4 entwickelte Satelliten-Gruppenantennen-Modell erfolgreich zu verifizieren. Zusätzlich wird für EuroSkyWay (ein geostationary orbit (GEO) System) ein optimales Gruppenantennen-Modell entwickelt, für das der Gruppenantennen-Gewinn am Rand des Versorgungsgebiets maximiert wird.

Kapitel 4 Hier wird das Fundament für eine einheitliche Behandlung der oben genannten Empfängervarianten im Rahmen der Informationstheorie bereitet. Dazu wird die Shannon-Kapazität des Vektor-Mehrfachzugriffskanal mit Signalschwund definiert, wobei die Annahme gilt, dass die Quellen ihre Sendeleistung nur an die langsam veränderlichen, quasi-statischen Dämpfungsfaktoren (*attenuation factors*) adaptieren können, jedoch nicht an die möglicherweise zeitlich schnell variierenden Schwundfaktoren (*fading factors*). Im Gegensatz dazu verfügt der gemeinsame Empfänger der Quellen-Signale über die Kenntnis der Dämpfungs- und Schwundfaktoren aller Quellen, um eine kohärente Demodulation and Dekodierung der Quellen-Signale zu ermöglichen.

Der Hauptteil dieses Kapitels widmet sich der Analyse des Problems der Ressourcenzuteilung für die Fälle, dass, erstens, die optimale Zuteilung der Informationsraten zu den Quellen gesucht wird, welche die Summen-Rate maximiert, während gleichzeitig eine Beschränkung der maximalen Sendeleistungen für die Quellen einzuhalten ist, und, zweitens, die optimal Zuteilung der Sendeleistungen zu den Quellen, welche die Summen-Sendeleistung minimiert, während geforderte Informationsraten für alle Quellen erreicht werden.

Die vorliegende Arbeit geht erheblich über die verfügbare Literatur hinaus, indem die optimale Ressourcenzuteilung für den Vektor-Mehrfachzugriffskanal mit Signalschwund im Zusammenhang mit linearen und nicht-linearen konvexen Programmen diskutiert wird. Die Basis dafür bildet die Definition der Ratenregion (Region erreichbarer Raten, *region of achievable rates*) und der Leistungsregion (*power region*) des Vektor-Mehrfachzugriffskanals mit Signalschwund. Die Ratenregion, bzw. die Leistungsregion ist die Menge zulässiger Lösungen für das entsprechende der oben genannten Probleme der Ressourcenzuteilung.[3]

Als schwierig erweist sich das Problem der optimalen Leistungszuteilung im Vektor-Mehrfachzugriffskanal mit Signalschwund, d.h. wenn für die Quellen geforderte Informationsraten realisiert werden sollen und die entsprechenden optimalen Sendeleistungen gesucht werden.

Zum ersten Mal überhaupt wird dies in der vorliegenden Arbeit im Detail für den Vektor-Mehrfachzugriffskanal mit Signalschwund untersucht (Abschnitt 4.5 und 4.6).

In diesem Zusammenhang wird gezeigt, dass die Leistungsregion des Vektor-Mehrfachzugriffskanals mit Signalschwund nur dann ein Polymatroid ist (was die Lösung des Problems der Leistungszuteilung mittels eines Greedy-Algorithmus gestatten würde), wenn die Schwundfaktoren konstant sind (resultierend in einem additive white gaussian noise (AWGN) Kanal) und die Steuervektoren (*steering vectors*) aller Quellen identisch sind (womit sich der klassische Mehrfachzugriffskanal ergibt). Eine einfache Lösung für die Leistungszuteilung ergibt sich auch, wenn alle Steuervektoren zueinander orthogonal sind, so dass sich voneinander unabhängige Kanäle ohne Interferenz ergeben und damit unabhängige Dekodierung optimal ist.

[3] In einem Optimierungsproblem, das als lineares Programm formuliert werden kann, ist die Menge zulässiger Lösungen (*feasible set*) durch ein Polytop gegeben. Die optimale Lösung wird immer an einem Eckpunkt (*vertex*) des Polytops erhalten. Wenn die Menge zulässiger Lösungen ein Polymatroid ist (ein Polytop mit bestimmten Eigenschaften), dann kann der optimale Eckpunkt mit geringem Aufwand mittels eines Greedy-Algorithmus (*greedy algorithm*) bestimmt werden.

Die Hauptergebnisse der Analyse des Ressourcenzuteilung-Problems für den Vektor-Mehrfachzugriffskanal mit Signalschwund sind wie folgt:

Für den allgemeinen Fall mit Signalschwund und ungleichen Steuervektoren ist die Leistungsregion kein Polytop. Die Leistungsregion ist jedoch konvex, so dass die optimale Leistungszuteilung sich als nicht-lineares konvexes Programm darstellt, wobei die optimale Lösung durch Standard-Optimierungsverfahren gefunden werden kann. Die optimale Leistungszuteilung liegt im Allgemeinen nicht auf einem Eckpunkt der Leistungsregion, so dass sukzessive Dekodierung suboptimal wäre. Trotzdem ist sukzessive Dekodierung attraktiv, weil diese sich im Vergleich zur Verbunddekodierung (*joint decoding*) mit geringerer Komplexität realisieren lässt und gleichzeitig Informationsraten am Rande der Ratenregion erzielt. Aus diesem Grund wird das Problem untersucht den optimalen Eckpunkt der Leistungsregion und somit die optimale Dekodierreihenfolge für sukzessive Dekodierung zu finden.

Als ein zentrales Ergebnis dieser Arbeit wird gezeigt, dass es für den Vektor-Mehrfachzugriffskanal mit Signalschwund im Allgemeinen keinen Greedy-Algorithmus geben kann, der die optimale Dekodierreihenfolge bestimmt. Dies wird erklärt, indem die verschiedenen Dekodierreihenfolgen für die sukzessive Dekodierung in einem Baumgraph (*tree graph*) dargestellt werden. Zusätzlich wird gezeigt, dass nur dann der minimale Summen-Pfad (entsprechend der minimalen Summen-Sendeleistung) durch den Baumgraph gefunden werden kann, wenn alle Pfade durch den Baumgraph durchsucht werden. Die Schlussfolgerung daraus ist, dass die Prozedur zur Bestimmung der optimalen Dekodierreihenfolge für sukzessive Dekodierung im Vektor-Mehrfachzugriffskanal mit Signalschwund im Allgemeinen das Ausprobieren aller Möglichkeiten ist (*complete enumeration*).

Schließlich wird die feste Strahlformung und ihr Einfluss auf die erreichbaren Raten diskutiert. Es wird gezeigt, dass feste Strahlformung immer verlustfrei ist (d.h. die mit fester Strahlformung erreichbaren Raten sind identisch mit den Raten, die ohne feste Strahlformung erzielt werden können), wenn die Zahl der festen (orthogonalen oder nicht-orthogonalen) Antennenkeulen (*fixed beams*) gleich der oder größer als die Zahl der Antennenelemente ist. Voraussetzung dabei ist, dass die Gewichtsvektoren zur festen Strahlformung linear unabhängig sind.

Kapitel 5 In Abschnitt 5.1.1 wird der Einfluss der Dekodierreihenfolge für sukzessive Dekodierung auf die Sendeleistungen der Quellen diskutiert. Dazu wird ein Empfänger mit optimaler adaptiver Strahlformung angenommen und eine exemplarische Quellen-Verteilung mit 7 Quellen, wobei 6 Quellen auf den Ecken eines Hexagons angeordnet sind und mit einer einzelnen Quelle im Zentrum. Die benötigten Sendeleistungen werden für unterschiedliche Winkeltrennungen zwischen den Quellen berechnet und es wird beobachtet, dass die benötigten Sendeleistungen mit der jeweiligen Dekodierreihenfolge stark variieren können. Der Grad der Variabilität hängt dabei von der anvisierten Informationsrate und der Winkeltrennung der interferierenden Signale ab. Für die betrachtete Quellen-Verteilung und für eine angenommene Informationsrate von $R = 1$ bit/channel use zeigt sich, dass die optimale Dekodierreihenfolge maximale Sendeleistungen liefert, die ca. 4 dB niedriger ausfallen, als es für die schlechteste Dekodierreihenfolge der Fall ist. Für höhere Informationsraten wachsen die Einsparungen weiter an, während sie für niedrigere Informationsraten geringer ausfallen (z.B. ca. 2 dB für $R = 0.5$ bit/channel use). Abschließend wird festgestellt, dass die Dekodierreihenfolge einen deutlichen Einfluss auf die maximalen Sende-

leistungen haben kann. Jedoch kann die Erhöhung der Sendeleistung aufgrund einer nicht optimal gewählten Dekodierreihenfolge gering ausfallen, abhängig von der gewünschten Informationsrate und der Winkeltrennung der Quellen.

Unter der Annahme derselben Quellen-Verteilung mit variabler Winkeltrennung wird die Leistungsfähigkeit von Empfängern mit optimaler Strahlformung und unabhängiger bzw. sukzessiver Dekodierung verglichen. Der Vergleich basiert auf dem Element-Signal-Rausch-Leistungs-Verhältnis (signal-to-noise-power-ratio (SNR))[4] und die von den Quellen aufzubringende Sendeleistung, um eine gewünschte Informationsrate zu erzielen. Es wird beobachtet, dass bei gleichem maximalem Element-SNR, bzw. gleicher Sendeleistung, die Quellen deutlich näher zueinander platziert werden können, falls sukzessive anstelle von unabhängiger Dekodierung eingesetzt wird. Es wird jedoch auch festgestellt, dass dieser Vorteil der sukzessiven Dekodierung bezüglich der Verminderung der maximalen Sendeleistung für wachsende Informationsraten verschwindet, weil dann die adaptive Strahlformung die Interferenz bereits so effizient unterdrücken muss, um das SINR zu maximieren, so dass durch sukzessive Dekodierung kaum mehr ein weiterer Vorteil erzielt werden kann.

Weil die Empfängervarianten AB+ID und AB+SD adaptive Strahlformung im sog. *beam space* realisieren (d.h. adaptive Strahlformung wird im Anschluss an feste Strahlformung implementiert), wird in Abschnitt 5.2.2 teil-adaptive Strahlformung (*partially adaptive beamforming*) diskutiert, wobei nur eine Untermenge der festen Antennenkeulen für die adaptive Strahlformung verwendet wird. Insbesondere wird untersucht, welche Unterschiede sich für teil-adaptive Strahlformung im *beam space* mit orthogonalen und nicht-orthogonalen festen Antennenkeulen ergeben. Anhand des ICO-Szenarios wird gezeigt, dass teil-adaptive Strahlformung mit nicht-orthogonalen Antennenkeulen vorteilhaft gegenüber der Verwendung von orthogonalen Antennenkeulen ist, weil im Mittel eine kleinere Anzahl von festen nicht-orthogonalen Antennenkeulen benötigt wird, um ein gleiches Maß an Leistungsfähigkeit zu erzielen.

Desweiteren ist es für den klassischen Mehrfachzugriffskanal bekannt, dass mit unabhängiger Dekodierung ein Ratenlimit existiert, welches eine maximale Rate angibt, die höchstens erzielt werden kann, selbst wenn die Sendeleistungen gegen Unendlich gehen (d.h. selbst wenn die Sendeleistungen gegen Unendlich gehen, streben die erzielten Informationsraten gegen einen endlichen Wert, nämlich dem Ratenlimit). In Abschnitt 5.3.1 wird gezeigt, dass auch für die Empfängervariante FB+ID eine solche maximale Rate existiert, obwohl die Interferenz durch die Richtcharakteristik (*gain pattern*) der festen Antennenkeulen teilweise unterdrückt wird. Schließlich wird eine Methode aufgezeigt, mit der dieses Ratenlimit für FB+ID berechnet werden kann. Ein solches Ratenlimit existiert nicht für die anderen Empfängervarianten, d.h. für gegen Unendlich strebende Sendeleistungen gehen auch die für die Quellen erreichbaren Informationsraten gegen Unendlich.

Der Kern dieses Kapitels ist eine vergleichende Untersuchung der zuvor erwähnten vier Empfängervarianten in Szenarien der Satellitenkommunikation.

In der Tat liefert die vorliegende Arbeit erstmals einen systematischen Vergleich der besagten Empfängervarianten in Szenarien der Satellitenkommunikation, basierend auf der in Kapitel 4 beschriebenen Methodik zur Ressourcenzuteilung für den Vektor-Mehrfachzugriffskanal mit Signalschwund.

[4] Das Element-SNR ist das SNR für ein einzelnes Gruppenantennen-Element.

Der anvisierte Vergleich der Empfängervarianten benötigt die Definition von bestimmten Interferenzszenarien, weil die gegenseitige Interferenz zusammen mit der gewünschten Informationsrate den entscheidenden Faktor darstellt, der die Unterschiede in der Leistungsfähigkeit der verschiedenen Empfängerkonfigurationen bestimmt. Die gegenseitige Interferenz wird festgelegt durch bestimmte Annahmen bezüglich der Richtcharakteristik der festen Antennenkeulen die das Versorgungsgebiet (*service area*) des betrachteten Satellitensystems bedecken und der Verteilung der Quellen selbst. Deshalb werden basierend auf den beiden in Kapitel 3 vorgestellten Satellitenszenarien verschiedene Interferenzszenarien definiert, wobei gemäß den angenommenen Frequenzwiederverwendungsmustern (*frequency reuse scheme, cluster size*) ungünstigste (*worst-case*) Verteilungen der Quellen angenommen werden.

Die Hauptergebnisse der beschriebenen Untersuchungen werden im folgenden präsentiert.

Generell gilt, dass FB+ID am schlechtesten abschneidet, während AB+SD die beste Leistungseffizienz zeigt, d.h. für eine gegebene gewünschte Informationsrate benötigt AB+SD das geringste Element-SNR, während FB+ID das höchste Element-SNR benötigt. Das ist weiter nicht überraschend, weil FB+ID keine adaptiven Verfahren zur besseren Unterdrückung von Interferenz beinhaltet, hingegen AB+SD beide umfasst. Von besonderem Interesse ist desweiteren ein Vergleich von FB+SD und AB+ID, weil diese Empfängervarianten nur eine von beiden Interferenz-reduzierenden Verfahren realisiert, weshalb der Implementierungsaufwand geringer ausfällt als für den Empfänger AB+SD. Diese Empfängervarianten werden für die Satellitenkommunikation in verschiedenen Veröffentlichungen vorgeschlagen, um eine bessere Unterdrückung von Interferenz zur erzielen und wegen der damit verbundenen besseren Bandbreiteneffizienz, jedoch wurden sie bisher unabhängig voneinander untersucht. Desweiteren war bisher unbekannt, unter welchen Umständen ein zusätzlicher Vorteil bzgl. Interferenzunterdrückung erzielt werden kann, wenn die im Vergleich zu FB+SD und AB+ID komplexere Empfängervariante AB+SD eingesetzt wird.

Zunächst kann man anhand der betrachteten Interferenzszenarien beobachten, dass für FB+SD und AB+ID weder die eine noch die andere Empfängervariante generell überlegen ist.

Wenn die Winkeltrennung der Quellen ausreichend groß ist, ist der AB+ID-Empfänger der FB+SD-Variante überlegen. Dies erklärt sich damit, dass erstens adaptive Strahlformung es ermöglicht, den Verlust am Rand einer festen Zelle (*edge-of-cell loss*) wiederherzustellen, und zweitens Interferenz effizient unterdrückt werden kann. In diesem Fall ist die Leistungsfähigkeit des AB+ID-Empfängers sehr nahe an der des komplexeren AB+SD-Empfängers, der zusätzlich sukzessive Dekodierung einsetzt. Interferenz wird dann zum größten Teil bereits durch die adaptive Strahlformung vermieden, so dass sukzessive Dekodierung kaum mehr einen signifikanten Vorteil bringen kann.

Wenn die interferierenden Quellen jedoch sehr dicht zusammen liegen (in der Größenordnung der Keulenbreite der festen Antennenkeulen), dann übertrifft die Leistungsfähigkeit des FB+SD-Empfängers unter Umständen die des AB+ID-Empfängers, zumindest für Informationsraten unterhalb einer bestimmten Grenze. Wenn die Informationsrate aber stetig erhöht wird, dann gelangt man zu einem Ratenlimit, jenseits dessen AB+ID stets besser abschneidet als FB+SD, was das maximal benötigte Element-SNR angeht. Die Erklärung für dieses Verhalten ist wie folgt. Zunächst schneidet der AB+ID-

Empfänger unterhalb einer bestimmten Informationsrate schlechter ab, weil die Interferenz durch adaptive Strahlformung nicht unterdrückt werden kann, ohne gleichzeitig den Gewinn für das Wunschsignal deutlich zu vermindern und damit das thermische Rauschen zum limitierenden Faktor zu machen. Wenn jedoch mit steigender Rate auch die Sendeleistungen ansteigen, dominiert der Einfluss der Interferenz über das thermische Rauschen. Die Fähigkeit von adaptiver Strahlformung, Interferenz räumlich zu filtern, kommt dann zum Tragen, weil der adaptive Strahlformer des AB+ID-Empfängers Nullstellen in der Richtcharakteristik plaziert für die Richtungen aus denen die interferierenden Quellensignale einfallen, so dass das benötigte Element-SNR gänzlich unabhängig von der Interferenz-Leistung wird.

Kapitel 6 Hier soll einen Einblick in die Implementierungskomplexität der verschiedenen Empfängervarianten verschafft werden.

Abschließend werden entsprechend den Hauptzielen, die Hauptresultate dieser Arbeit wie folgt zusammengefasst:

- Eine Methodik für den systematische Vergleich der besagten Empfängervarianten wird für Szenarien der Satellitenkommunikation präsentiert. Diese basiert auf der Diskussion des Problems der Ressourcenzuteilung im Vektor-Mehrfachzugriffskanal mit Signalschwund.

- Für den Vektor-Mehrfachzugriffskanal mit Signalschwund wird das Problem der Ressourcenzuteilung analysiert, d.h. die optimale Ratenzuteilung, die die Summen-Rate bei gegebener Beschränkung der Sendeleistungen maximiert und die optimale Zuteilung der Sendeleistungen, die die Summen-Sendeleistung minimiert, während geforderte Raten erzielt werden. Es wird gezeigt, dass es für den Vektor-Mehrfachzugriffskanal mit Signalschwund im Allgemeinen keinen Greedy-Algorithmus gibt, um die optimale Dekodierreihenfolge für sukzessive Dekodierung zu bestimmen.

- Es wird gezeigt, dass Empfänger, die adaptive Strahlformung oder sukzessive Dekodierung oder beides implementieren (Varianten AB+ID, FB+SD, und AB+SD), einen deutlichen Vorteil gegenüber der Variante bieten, die heutzutage den Standard in der Satellitenkommunikation darstellt und nur feste Strahlformung mit unabhängiger Dekodierung realisiert (FB+ID). Die Leistungsfähigkeit der Empfänger AB+ID, FB+SD, und AB+SD hängt dabei stark vom Interferenzszenario ab, wie ein Vergleich der Empfängervarianten in typischen Interferenzszenarien zeigt.

- Die Leistung der Empfängervariante AB+ID liegt sehr nahe an der des komplexeren AB+SD-Empfängers und übertrifft die des FB+SD-Empfängers, wenn die interferierenden Quellen mäßig dicht zusammen liegen (wie es z.B. bei Cluster-Größe 3 der Fall ist).

- Die Empfänger-Variante FB+SD kann die Variante AB+ID übertreffen, wenn die interferierenden Quellen sehr dicht zusammen liegen (in der Größenordnung der Keulenbreite). Ob mit FB+SD tatsächlich eine gewünschte Informationsrate mit einer niedrigeren maximalen Sendeleistung erzielt werden kann, hängt jedoch von der Informationsrate selbst ab. Erhöht man nach und nach die gewünschte Informationsrate, so ergibt sich immer eine Rate, jenseits der der AB+ID-Empfänger stets eine geringere maximale Sendeleistung als der FB+SD-Empfänger benötigt.

Bibliography

[AG99] M.-S. Alouini and A. J. Goldsmith. Capacity of rayleigh fading channels under different adaptive transmission and diversity-combining techniques. *IEEE Trans. on Vehicular Technology*, 48(4):1165–1181, July 1999.

[And99] B. R. Andersen. A survey of digital signal processing methods for advanced communication satellite. In *Proceedings Norsk Symposium I Signalbenhandling (NORSIG'99)*, Asker, Norway, 1999.

[BBJ95] J. Blanz, P. W. Baier, and P. Jung. A flexible configurable statistical channel model for mobile radio systems with directional diversity. In *Proceedings ITG Workshop on Mobile Communications*, Neu-Ulm, Germany, 1995.

[BC74] P. P. Bergmans and T. M. Cover. Cooperative broadcasting. *IEEE Trans. on Information Theory*, 20:317–324, May 1974.

[Ber77] L. A. Berry. Spectrum metrics and spectrum efficiency: Proposed definitions. *IEEE Trans. on Electromagnetic Compatibility*, 19(3):254–260, August 1977.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding: Turbo codes. In *IEEE International Conference on Communications (ICC'98)*, pages 1064–1070, Geneva, Switzerland, May 1993.

[Bjö93] G. Björnström. Digital payloads: Enhanced performance through signal processing. *ESA Journal*, 17:1–29, 1993.

[BM94] R. Bernardini and R. Manduchi. On the reduction of multidimensional DFT to separable DFT by smith normal form theorem. In *European Trans. on Communications*, volume 5, pages 377–381, May 1994.

[Bro91] E. Brookner, editor. *Practical Phased-Array Antenna Systems*. Artech House, Boston, USA, 1991.

[CCAL01] Y. Cailloce, G. Caille, I. Albert, and J. M. Lopez. Ka-band antennas providing multiple beams for a multimedia via satellite mission. In *Proceedings of the 19th International Communication Satellite Systems Conference and Exhibit*, Toulouse, France, April 2001. AIAA. 19th International Communication Satellite Systems Conference & Exhibit.

[CCF+92] C. Caini, G. E. Corazza, G. Falciasecca, M. Ruggieri, and F. Vatalaro. A spectrum- and power-efficient EHF mobile satellite system to be integrated with terrestrial cellular systems. *IEEE Journal on Selected Areas in Communication*, 10:1315–1325, 1992.

[CL96] P. A. Chiavacci and J. W. Locke. Planar phased array antenna relationships for FFT & butler matrix generated beamlets with graphical display of results. In *IEEE International Symposium on Phased Array Systems and Technology*, pages 196–202, Boston, USA, October 1996.

[CLW92] A. D. Craig, C. K. Leong, and A. Wishart. Digital signal processing in communications satellite payloads. *Electronics & Communication Engineering Journal*, pages 107–114, June 1992.

[CT91] Th. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[Dav97] F. Davarian. Channel behavior and power control in handheld mobile satellite links. In *Proceedings 5th International Mobile Satellite Conference (IMSC'97)*, pages 219–227, 1997.

[DM84] D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice-Hall, Englewood-Cliffs, 1984.

[DP02] E. Del Re and L. Pierucci, editors. *Satellite Personal Communications for Future-Generation Systems - Final Report: COST 252 Action*. Springer-Verlag London, 2002.

[Dub85] E. Dubois. The sampling and reconstruction of time-varying imagery with application in video systems. *Proceedings of the IEEE*, 73(4):502–522, April 1985.

[EB98] Sh. Shamai E. Biglieri, J. Proakis. Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Information Theory*, 44:2619–2692, 1998.

[ECS+98] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed. Overview of spatial channel models for antenna array communication systems. *IEEE Personal Communications*, 5(1):10–22, February 1998.

[Edm69] J. Edmonds. Submodular functions, matroids, and certain polyhedra. In *Proc. Calgary Int. Conf. Combinatorial Structures and Applications*, pages 69–87, Calgary, Canada, June 1969.

[Ega99] Sh. Egami. A power-sharing multiple-beam mobile satellite in ka band. *IEEE Journal on Selected Areas in Communications*, 17(2):145–152, February 1999.

[Ern99] H. Ernst. Interference cancellation for TDMA-satellite systems. In *Proceedings 6th International Mobile Satellite Conference (IMSC'99)*, pages 94–99, Ottawa, Canada, June 1999.

[Ern01] H. Ernst. Satellite UMTS and interference cancellation. In *Proceedings Conference on 3G Infrastructure and Services*, pages 58–62, Athens, Greece, July 2001.

[ESW97] Aeronautical link design report. In *SECOMS/ABATE Project, Deliverable D35*, August 1997.

[Far97] Ch. Farsakh. *Raummultiplex mit Intelligenten Antennen in zellularen Mobilfunksystemen*. Hieronymus, München, 1997.

[Gar93] F. M. Gardner. Interpolation in digital modems - part I: Fundamentals. *IEEE Trans. on Communications*, 41(3):501–507, March 1993.

[Gay02] J.-D. Gayrard. Evolution of telecommunication payloads: The necessity of new technologies. In *Proceedings of the 17th AIAA International Communications Satellite Systems Conference and Exhibit*, Montréal, Canada, May 2002. AIAA.

[GG95] Th. Gebauer and H. G. Göckler. Channel-individual adaptive beamforming for mobile satellite communications. *IEEE J. Selected Areas in Communications*, 13:439–448, 1995.

[GHLW03] D. Giggenbach, M. Holzbock, O. Lücke, and M. Werner. Optical-SDMA for broadband aeronautical communication. In *21st AIAA International Communications Satellite Systems Conference (ICSSC2003)*, Yokohama, Japan, April 2003.

[GL96] G. Golub and C. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3 edition, 1996.

[God97a] L. C. Godara. Applications of antenna arrays to mobile communications and part I: Performance improvement and feasibility and and system considerations. *IEEE Proceedings*, 85(7):1031–1060, July 1997.

[God97b] L. C. Godara. Applications of antenna arrays to mobile communications and part I: Performance improvement and feasibility and and system considerations. *IEEE Proceedings*, 85(8):1195–1245, August 1997.

[GST99] L. Ghedia, K. Smith, and G. Titzer. Satellite PCN - the ICO system. *Int. Journal of Satellite Communications*, 17:273–289, 1999.

[GV93] A. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. In *Proceedings IEEE International Conference on Communications (ICC'93)*, pages 600–604, Geneva, Switzerland, June 1993.

[Haa96] M. Haardt. *Efficient One-, Two-, and Multidimensional High-Resolution Array Signal Processing*. Shaker, Aachen, Germany, 1996.

[HHF99] T. Hentschel, M. Henker, and G. Fettweis. The digital front-end of software radio terminals. *IEEE Personal Communications*, 6(4):40–46, August 1999.

[HJ99] R. A. Horn and Ch. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1999.

[HR01] F. J. Harris and M. Rice. Multirate digital filters for symbol timing synchronization in software defined radios. *IEEE Journal on Selected Areas in Communications*, 19(12):2346–2357, December 2001.

[IHLA98] P. A. Ilot, R. Hladek, Ch. Liu, and B. Arnold. ICO S-band antennas test program. In *20th Antenna Measurement Techniques Association Conference*, Montréal, Canada, October 1998.

[ITU90] *Radiocommunication Vocabulary. ITU-R Recommendation 573-3*. International Telecommunication Union (ITU), Genf, 1990.

[ITU95] *The Concept of Transmission Loss for Radio Links. ITU-R Recommendation P.341-4*. International Telecommunication Union (ITU), Genf, 1995.

[Jah94] A. Jahn. Propagation data and channel model for LMS systems. In *Final Report, Inmarsat purchase Order P004001*, DLR, Institut für Nachrichtentechnik, 1994.

[Jah99] A. Jahn. *Ressourcenverwaltung in Kommunikationsnetzen mit niedrigfliegenden Satelliten*. PhD thesis, FernUniversität Hagen, Germany, 1999.

[KB96] J. Karimi and S. D. Blostein. Parabolic reflector array signal processing for improved rural area coverage in personal satellite communications. In *International Conference on Universal Personal Communications*, 1996.

[KH95] R. Knopp and P. A. Humblet. Information capacity and power control in single-cell multiuser communications. *IEEE International Conference on Communications (ICC)*, June 1995.

[Kra88] J. D. Kraus. *Antennas*. McGraw-Hill, New York, 1988.

[LC00] J. L. Lee and J. Chun. Adaptive interference nulling technique for a LEO cellular satellite system. In *A Collection of the 18th AIAA International Communications Satellite Systems Conference and Exhibit Technical Papers (AIAA2000)*, pages 14–18, Oakland, April 10-14, 2000.

[LL96] J. Litva and T. Kwok-Yeung Lo. *Digital Beamforming in Wireless Communications*. Artech House, Boston, London, 1996.

[LMM01] G. Losquadro, R. Mura, and D. Mignolo. The ESW home gateway supporting IP and MPEG services for residential users. In *19th AIAA International Communications Satellite systems Conference*, Tolosa, April 2001.

[Lüc98] O. Lücke. Intelligente antennen in der satellitenkommunikation. In *Systeme mit intelligenten Antennen für UMTS*, Karlsruhe, Germany, June 1998.

[Lüc00] O. Lücke. Capacity increase with an adaptive antenna for a GEO satellite scenario. In *Proceedings 4th European Mobile Personal Satellite Communications Workshop (EMPS 2000)*, pages 26–33, London, Sept. 2000.

[LWJ00] E. Lutz, M. Werner, and A. Jahn. *Satellite Systems for Personal and Broadband Communications*. Springer-Verlag, Berlin, 2000.

[Mai94] R. Mailloux. *Phased Array Antenna Handbook*. Artech House, Boston, USA, 1994.

[Mat00] R. Matzner. *Der Kapazitätsgewinn durch die ganzheitliche Beschreibung von Kabelbündeln und seine technische und ökonomische Nutzbarkeit*. PhD thesis, Universität der Bundeswehr München, 2000.

[MB98] G. Maral and M. Bousquet. *Satellite Communications Systems*. John Wiley & Sons, Chichester, 3rd edition, 1998.

[Mec00a] M. Mecking. Minimizing transmit power for fading multiple-access channels. In *IEEE International Symposium on Information Theory (ISIT 2000)*, Sorrento, Italy, June 2000.

[Mec00b] M. Mecking. Power control and accessing strategies for fading multiple-access channels. In *3rd ITG Conference Source and Channel Coding*, pages 231–236, München, January 2000.

[Mec01] M. Mecking. Multiple-access with spripping receivers. In *Proc. 4th European Mobile Communication Conference (EPMCC)*, Vienna, Austria, February 2001.

[Mec02a] M. Mecking. *Fading Multiple-Access with Channel State Information*. PhD thesis, Technische Universität München, 2002.

[Mec02b] M. Mecking. Resource allocation for fading multiple-access channels with partial channel state information. In *IEEE International Conference on Communications (ICC'02)*, New York, USA, April 2002.

[MG86] H. Meinke and F. W. Gundlach. *Taschenbuch der Hochfrequenztechnik*. Springer Verlag, Heidelberg, 1986.

[MM80] R. A. Monzingo and Th. W. Miller. *Introduction to Adaptive Arrays*. John Wiley & Sons, Chichester, 1980.

[Moh00] M. L. Moher. Multiuser decoding for multibeam systems. *IEEE Trans. on Vehicular Technology*, 49(4):1226–1234, July 2000.

[MPPL+98] M. Marinelli, V. Podda, R. Pérez-Leal, M. Luglio, L.S. Zhang, A. Jahn, and A. Franchi. SECOMS and ABATE: System architectures and relevant space segment. In *ACTS Mobile Communication Summit*, Rhodes, Greece, June 1998.

[MS98] F. Makita and K. Smith. Design and implementation of ICO system. In *Proceedings of the 17th AIAA International Communications Satellite Systems Conference and Exhibit*, pages 57–65, 1998.

[NM93] F. D. Neeser and J. L. Massey. Proper complex random processes with applications to information theory. *IEEE Trans. on Information Theory*, 39(4):1293–1302, July 1993.

[NYZ+01] I. Nakajima, Y. Yagi, Y. Zhao, N. Hamamoto, and A. Nakajima. Cost comparison between HEO satellites and 3G wireless mobile phones for video transmission from ambulances. In *Proceedings 4th International Symposium on Wireless Personal Multimedia Communications (WPMC 2001)*, Aalborg, Denmark, September 2001.

[OC90] C. A. Olen and R. T. Compton. A numerical pattern synthesis algorithm for arrays. *IEEE Trans. on Antennas and Propagation*, 38(10):1666–1676, October 1990.

[Pap91] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Singapore, 3 edition, 1991.

[Par92] D. Parson. *The Mobile Radio Propagation Channel*. Pentech Press, London, 1992.

[Pro95] J. G. Proakis. *Digital Communications*. McGraw-Hill, Singapore, 1995.

[PS98] Ch. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization - Algorithms and Complexity*. Dover Publications, Inc., Mineaola, New York, 1998.

[Rap99] P. B. Rapajic. Information capacity of space division multiple access mobile communiation systems. *Wireless Personal Communications*, 11:131–159, 1999.

[RU96] B. Rimoldi and R. Urbanke. A rate-splitting approach to the gaussian multiple-access channel. *IEEE Trans. on Information Theory*, 42(2):364–375, March 1996.

[SC01] D. P. Scholnik and J. O. Coleman. Superdirectivity and SNR constraints in wideband array-pattern design. In *2001 IEEE International Radar Conference (RADAR 2001)*, Atlanta, GA, May 2001.

[SDR+02] D. A. Sunderland, G. L. Duncan, B. J. Rasmussen, H. E. Nichols, D. T. Kain, L. C. Lee, B. A. Clebowicz, R. W. Hollis, L. Wissel, and T. Wilder. Megagate ASICs for the thuraya satellite digital signal processor. In *Proc. IEEE International Symposium on Quality Electronic Design*, page 479, San Jose, California, March 2002.

[ST81] W. L. Stutzmann and G. A. Thiele. *Antenna Theory and Design*. John Wiley & Sons, New York, 1981.

[SUM+99] J. J. Schuss, J. Upton, B. Myers, T. Sikina, A. Rohwer, P. Makridakas, R. Francois, L. Wardle, and R. Smith. The IRIDIUM main mission antenna concept. *IEEE Trans. on Antennas and Propagation*, 47(3):416–424, March 1999.

[SXLK98] B. Suard, G. Xu, H. Liu, and Th. Kailath. Uplink channel capacity of space-division-multiple-access schemes. *IEEE Trans. on Information Theory*, 44:1468–1469, July 1998.

[TH98a] D. N. C. Tse and S. V. Hanly. Multiaccess fading channels-part I: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Trans. on Information Theory*, 44(7):2796–2815, November 1998.

[TH98b] D. N. C. Tse and S. V. Hanly. Multiaccess fading channels-part II: Delay limited capacities. *IEEE Trans. on Information Theory*, 44(7):2816–2831, November 1998.

[TT98] Z. Tian and H. L. Van Trees. DOA estimation with hexagonal arrays. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP98)*, volume 4, pages 2053–2056, Seattle, USA, May 1998.

[Ung81] G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Trans. on Information Theory*, 28(1):55–67, January 1981.

[Vat91] G. Vatt. The Iridium system. In *Virginia Tech. 1st Symp. on Wireless Personal Communications*, pages 4.1–4.9, 1991.

[VG97] M. K. Varanasi and T. Guess. Optimum decision feedback equalization with successive decoding achieves the total capacity of the gaussian multiple-access channel. In *Proceedings of Asilomar Conference on Signal, Systems and Computing*, pages 1405–1409, Monterey, USA, November 1997.

[VSW91] R. G. Vaughan, N. L. Scott, and D. R. White. The theory of bandpass sampling. *IEEE Trans. on Signal Processing*, 39(9):1973–1984, September 1991.

[Wel76] D. Welsh. *Matroid Theory*. Academic Press, London, 1976.

[Wyn74] A. D. Wyner. Recent results in the shannon theory. *IEEE Trans. on Information Theory*, 20:2–10, January 1974.

[Yat95] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal On Selected Areas In Communications*, 13(7):1341–1347, September 1995.

[Yu96] K.-B. Yu. Adaptive beamforming for satellite communication with selective earth coverage and jammer nulling capability. *IEEE Trans. Signal Processing*, 44:3162–3166, 1996.

List of Symbols

General notation

$\mathbf{x}$	Column vector
x	Variable, element of a set
$x()$	Function
X	Random variable (exceptions possible)
$\boldsymbol{X}$	Vector or matrix of random variables
$\mathbf{X}$	Matrix
$\mathcal{X}$	Set
x^*	Complex conjugate of x

Matrix and vector operators

$\mathbf{X}^{-1}$	Inverse
$\mathbf{X}^{\mathrm{T}}, \mathbf{x}^{\mathrm{T}}$	Matrix/vector transpose
$\mathbf{X}^{\mathrm{H}}, \mathbf{x}^{\mathrm{H}}$	Transposed complex conjugate matrix/vector
$\det(\mathbf{X})$	Determinant
$\mathrm{rank}(\mathbf{X})$	Rank of the matrix
$\mathbf{x} \cdot \mathbf{y}$	Dot product

Symbols

	Description	Page
$\mathbf{A}$	Steering matrix	48
A_{eff}	Effective antenna aperture	25
A_{geo}	Geometric aperture area	25
A_{s}	Geometric area of coverage area	31

$\mathbf{B}$	Basis of spatial periodicity lattice	216
	Linear causal feedback filter matrix	105
$\mathcal{B}_m$	Set of beam indices used for partially adaptive beam space beamforming for source m	135
$\mathbf{B}^*$	Basis of spatial frequency sampling lattice	217
B_c	Coherence bandwidth	15
	Bandwidth allocated to cell	29
B_{f_D}	Doppler spread	15
B_{mi}	Linear feedback filter coefficients	102
B_s	Signal bandwidth	15
B_sat	Bandwidth allocated to satellite	29
B_tot	Total bandwidth	169
$\mathbb{C}$	Set of complex numbers	
C_s	Satellite traffic capacity	31
D, $D(\vartheta, \varphi)$	Directivity	24,23
$D_\mathrm{a}(\vartheta, \varphi)$	Array directivity	39
$\mathcal{D}_{\mathcal{P}_\mathrm{CMAC}}$	Dominant face of the power region for the classical fading MAC	78
$\mathcal{D}_{\mathcal{P}_\mathrm{VMAC}}$	Dominant face of the power region for the classical fading MAC	99
$D_\mathrm{e}(\vartheta)$	Directivity of an array element for angle ϑ	35
$\mathcal{D}_{\mathcal{R}_\mathrm{CMAC}}$	Dominant face of the capacity region for the classical fading MAC	77
$\mathbf{E}(\vartheta, \varphi, r, t)$	Electric field vector	20
$E_\vartheta, E_\varphi, E_r$	Components of electric field vector in spherical coordinates	20
$\mathbf{F}$	Linear feedforward filter matrix	105
$F(\vartheta, \varphi)$	Normalised field (also: magnitude) pattern	23
$\lvert F(\vartheta, \varphi)\rvert^2$	Power pattern	23
G, $G(\vartheta, \varphi)$	Gain	25, 25
$G_\mathrm{a}(\vartheta, \varphi)$	Array gain	39
$G_\mathrm{e}(\vartheta)$, $G_\mathrm{e}(\vartheta, \varphi)$	Gain of an array element	34
G_{kl}	Spot beam array gain	143
$\mathbf{H}(\vartheta, \varphi, r, t)$	Time-variant magnetic field vector	20
$\mathbf{H}$	Channel matrix	48
$H_\vartheta, H_\varphi, H_r$	Components of magnetic field vector in spherical coordinates	20
H_m	Fading factor of source m, random variable	76
$\mathbf{I}_m$	$m \times m$ identity matrix	
$\mathcal{I}_m$	Subset of $\mathcal{M}$	51

$\mathbf{K_n}$	Noise covariance matrix	50
$\mathbf{K}_{\mathbf{n}_\mathrm{B}}$	Beam space noise covariance matrix	111
$\mathbf{K_r}$	Array output covariance matrix	50
$\mathbf{K}_{\mathbf{r}_\mathrm{B}}$	Covariance matrix of the beamforming network output vector	111
K_s	Cluster size	29
$\mathbf{K_s}$	Source signal covariance matrix	50
$\mathbf{K_v}$	Receive signal covariance matrix without noise	50
L	Number of array elements	34
L_B	Number of fixed beams	111
$L_{\mathrm{B},m}$	Number of beams used for partially adaptive beam space beamforming for source m	135
L_bf	Free-space basic transmission loss	26
M	Number of sources	46
$\mathcal{M}$	Set of source indices	51
M_E	Earth mass, $5.974 \cdot 10^{24}$ kg	9
M_tot	Total number of source signals	169
N_0	Noise power spectral density	13
N_S	Minimum number of satellites for a complete coverage of Earth	11
N_s	Number of channels/slots	169
$\mathbf{P}$	Vector/tuple of receive powers or power constraints	77
$\mathcal{P}_\mathrm{CMAC}$	Power region for the classical fading MAC	77
P_LOS	Receive signal power in the line-of-sight path	17
P_m	Mean receive power for source m	75
P_mp	Multi-path component of received signal power	17
P_rad	Radiated power	23
P_R	Receive power	16
P_T	Transmit power	16
$\mathcal{P}_\mathrm{VMAC}$	Power region for the fading vector MAC	93
$\mathbf{R}$	Vector/tuple of rates	77
$\mathbb{R}$	Set of real numbers	
$\mathcal{R}_\mathrm{CMAC}$	Capacity region for the classical fading MAC	77
R_E	Earth radius, 6378.144 km	9
R_m	Information rate of source m	76
$R_\mathrm{max}^\mathrm{FB}$	Rate limit for fixed beamforming with independent decoding	143
$\mathcal{R}_\mathrm{VMAC}$	Capacity region for the fading vector MAC	93
$\mathcal{S}$	Subset of $\mathcal{M}$	51
$S(\vartheta, \varphi, r, t)$	Time-variant power flux-density	20

$\mathbf{S}(\vartheta, \varphi, r, t)$	Time-variant Poynting vector	20
$S_{\mathrm{i}}(r)$	Power flux-density of the isotropic radiator	23
$S_{\max}(r)$	Maximal power flux-density	23
$\mathbf{S}_m$	Selection matrix	135
$T(t, f)$	Time-variant transfer function	14
T_{A}	Antenna noise temperature	13
T_{c}	Coherence time	15
T_{e}	Effective noise temperature of receiver plus antenna	13
T_{o}	Satellite orbit period	9
T_{R}	Effective noise temperature of receiver without antenna	13
T_{s}	Symbol duration	15
$\mathcal{U}, \mathcal{U}^*$	Unit cell	215
$\mathbf{V}$	Basis of triangular array lattice	37
$\mathbf{V}^*$	Periodicity matrix for triangular array lattice	41
$\mathbf{W}_{\mathrm{B}}$	Matrix of fixed beamforming weights	111
$\mathbf{W}_{\mathrm{B},m}$	Matrix of fixed beamforming weights for partially adaptive beam space beamforming for source m	135
$\mathbf{W}_{\mathrm{DFT}}$	DFT beamforming matrix	117
$\mathbb{Z}$	Set of integers	
Z_{I}	Intrinsic Impedance	21
$\mathbf{a}$, $\mathbf{a}_m$	Steering vector	38, 48
a_l, $a_{l,m}$	Direction-of-arrival dependent phase shift for element l relative to reference point	37, 45
$\mathbf{a}_{\mathrm{B},m}$	Beam space steering vector	111
c	Speed of light in an arbitrary medium	
c_0	Speed of light in vacuum, $2.998 \cdot 10^8 \ \frac{\mathrm{m}}{\mathrm{s}}$	
c_{R}	Rice factor	17
d	Slant range (distance terminal–satellite)	10
$\mathbf{d}$	Direction-of-arrival vector	36
d_0	Maxima antenna aperture dimension	21
d_{a}	Satellite antenna diameter	27
d_{e}	Circular array element diameter	34
d_{s}	Hexagonal-array element phase centre spacing	37
d_{scat}	Maximal distance between source and scatterer	18

f	Frequency	
$f_{\mathrm{a}}(\vartheta, \varphi)$	Array factor	39
f_{c}	Carrier frequency, centre frequency	34
$f_{\mathrm{ea}}(\vartheta, \varphi)$	Array field pattern	38
$g(\vartheta, \varphi)$	Complex magnitude pattern	26
$g_{\mathrm{e}}(\vartheta)$, $g_{\mathrm{e}}(\vartheta, \varphi)$	Complex magnitude pattern of an array element	34
$h(t)$	Normalised multiplicative time-variant channel impulse response	16
$h'(t)$	Multiplicative time-variant channel impulse response	15
$h'(t, \tau)$	Linear time-variant channel impulse response	14
h_{S}	Satellite height over ground	9
$\mathbf{i}_l \in \mathbb{Z}^2$	Integer position vector of element l	37
k	Boltzmann constant, $1.38 \cdot 10^{-23}\ \frac{\mathrm{J}}{\mathrm{K}} = -228.6\ \mathrm{dBWs/K}$	13
$n_l(t)$	Noise at element l	42
$\mathbf{n}$	Noise vector	48
p	Rate of drop-off for aperture taper	65
$\tilde{\mathbf{p}}_{\mathrm{e},l}$	Array element position vector in $\mathbb{R}^3$	35
$\mathbf{p}_{\mathrm{e},l}$	Planar array element position vector in $\mathbb{R}^2$	36
$p^{\mathrm{x}}_{\mathrm{e},l}$, $p^{\mathrm{y}}_{\mathrm{e},l}$	x-y-coordinate of the phase centre of the l-th array element	35
p_m	Transmit power of source m	50
r	Distance from origin in spherical coordinates	21
$\mathbf{r}$	Array element receive signal vector	48
$\tilde{\mathbf{r}}$	Array input signal vector	48
$\tilde{r}(t)$	Receive signal at satellite	14
$\mathbf{r}_{\mathrm{B}}$	Beamforming network output vector	111
$\tilde{r}_m(t)$	Receive signal at array reference point for source m	44
$r_l(t)$	Receive signal of the l-th array element	44
r_{S}	Distance Earth centre to satellite	10
r_{T}	Distance Earth centre to terminal/source	10
$s(t)$	Transmit signal of a terminal/source	14
$\mathbf{s}$	Source signal vector	48
$s_m(t)$, s_m	Transmit signal of source m in complex baseband	43
t	Time	
t_{a}	Taper amplitude at edge of aperture	65

t_{dB}	Measure of taper amplitude at edge of aperture in dB	66
u	$= \sin\vartheta\cos\varphi$, Sine space coordinate	26
v	$= \sin\vartheta\sin\varphi$, Sine space coordinate	26
	Vertex-index of a (contra)-polymatroid	72
$\mathbf{v}$	Array element receive signal vector without noise	48
$\mathbf{v}_1$, $\mathbf{v}_2$	Basis vectors of triangular lattice	37
$v_{l,m}(t)$	Array element receive signal without noise	44
$\mathbf{w}$, $\mathbf{w}_m$	Beamforming weight vector	38, 52
$\mathbf{w}_l^{\mathrm{DFT}}$	DFT beamforming weight vector	116
$\mathbf{w}_{\mathrm{opt},m}(\mathbf{h})$	Optimal weight vector for given joint fading state $\mathbf{h}$	52
$\mathbf{w}_{\mathrm{opt},m}^{\mathrm{B}}$	Optimal beam space weight vector	132
$x_{\mathrm{RF}}(t)$	Real RF signal	42
$x^+(t)$	Analytical signal	42
z_m	Output symbol of m-th beamformer	52
	Input symbol to m-th single-user decoder	102
$\tilde{z}_m$	Output symbol of m-th beamformer in successive decoding	102
Γ	Spatial periodicity lattice	216
Γ^*	Spatial frequency sampling lattice	216
Γ_m	Generic output SINR	52
$\Gamma_{\mathrm{opt},m}$	Maximal output SINR	52
$\Gamma_{\mathrm{opt},m}^{\mathrm{B}}$	Maximal SINR achievable by optimal beam space processing	132
$\Gamma_{\mathrm{PA},m}^{\mathrm{B}}$	Generic output SINR for partially adaptive beam space beamforming for source m	135
Δ_l, $\Delta_{l,m}$	Relative path length difference for array element l	36, 43
$\Delta\vartheta_{\mathrm{s}}$	Angular separation between the sources	121
Δ_τ	Delay spread	15
Λ	Spatial sampling lattice	215
Λ^*	Periodicity lattice in spatial frequency domain	215
Ψ	Earth centre angle	10
$\Psi_{\max}$	Maximal Earth centre angle	10
Ω_{A}	Beam solid angle	24

$\alpha(\vartheta, \varphi)$	Phase shift	26
α_m	Channel factor	173
α_{mi}	Interference channel factor	173
γ_m	Mean element (also: input) SNR	53
$\gamma_m(h_m)$	Instantaneous element (also: input) SNR	53
γ_0	Gravitational constant, $6.672 \cdot 10^{-11}\ \frac{\mathrm{m}^3}{\mathrm{kg\,s}^2}$	9
$\delta(t)$	Dirac function	
δ_S	Latitude of satellite	10
δ_T	Latitude of terminal/source	10
ε	Elevation angle	10
	Electric permittivity	21
ε_0	Permittivity of the vacuum, $4\pi \cdot 10^{-7} \frac{\mathrm{H}}{\mathrm{m}} = 1.257 \frac{\mathrm{H}}{\mathrm{m}}$	21
ε_min	Minimal elevation angle	10
ε_r	Relative permittivity	10
η_a	Antenna efficiency	25
η_A	Aperture efficiency	25
η_r	Radiation efficiency	25
$\eta_\mathrm{r,e}$	Radiation efficiency of an array element	35
$\eta_\mathrm{r,a}$	Radiation efficiency of the array	39
η_s	Spectrum efficiency	31
ϑ	Nadir angle, antenna off-boresight angle	10
	Polar angle for spherical coordinates	21
ϑ_3dB	Onesided 3 dB beamwidth	27
λ	Wavelength	
$\boldsymbol{\lambda}$	Vector of costs or rewards	72
$\boldsymbol{\lambda}_\mathrm{P}$	Vector of power costs	79
$\boldsymbol{\lambda}_\mathrm{R}$	Vector of rate rewards	79
λ_S	Longitude of satellite	10
λ_T	Longitude of terminal/source	10
λ_c	Carrier wavelength	34
μ	Link (power) attenuation factor	16,27
	Magnetic permeability	21
μ_0	Permeability of the vacuum, $4\pi \cdot 10^{-7} \frac{\mathrm{H}}{\mathrm{m}} = 1.257 \frac{\mathrm{H}}{\mathrm{m}}$	21
μ_r	Relative permeability	21

$\boldsymbol{\nu}$	Spatial frequency	36
$\boldsymbol{\nu}_0$	Spatial frequency corresponding to the main lobe	41
$\boldsymbol{\pi}_v$	Permutation	72
ρ_{wb}	Quotient of maximum and minimum sum-power over all permutations	122
σ_{n}^2	Variance of a real Gaussian noise process	42
τ	Delay	14
τ_l	Relative delay for array element l	36
$\tau_{l,m}$	Relative delay of source signal m for array element l	43
φ	Azimuthal angle for spherical coordinates	21

List of Acronyms

ADC	analog-to-digital converter
AWGN	additive white gaussian noise
BER	bit error rate
BFN	beamforming network
BPSK	binary phase shift keying
CCI	co-channel interference
CDMA	code division multiple access
CEPT	European Conference of Postal and Telecommunications Administrations
CSI	channel state information
DRA	direct radiating array
DFT	discrete fourier transformation
DOA	direction-of-arrival
EIRP	equivalent isotropically radiated power
EOC	edge-of-cell
FDMA	frequency division multiple access
FLOPS	floating point operations per second
GEO	geostationary orbit
GSO	geosynchronous orbit
HEO	highly elliptical orbit
ICO	intermediate circular orbit
IF	intermediate frequency
ITU	International Telecommunication Union

LEO low earth orbit

LHCP left hand circularly polarised

LNA low-noise amplifier

LOS line-of-sight

MAC multiple access channel

MBA multiple-beam antenna

MEO medium earth orbit

nLOS non-line-of-sight

PDF probability density function

QPSK quaternary phase shift keying

RHCP right hand circularly polarised

SDMA space division multiple access

SINR signal-to-interference-and-noise-power-ratio

SNR signal-to-noise-power-ratio

TDMA time division multiple access

ULA uniform linear array

URA uniform rectangular array

WSSUS wide-sense stationary uncorrelated scattering

w.l.o.g. without loss of generality

Appendix A

Matrix Identities

Inverse Let $\mathbf{X} \in \mathbb{C}^{m \times m}$, $\mathbf{Y} \in \mathbb{C}^{n \times n}$ (both non-singular), further $\mathbf{U} \in \mathbb{C}^{m \times n}$, $\mathbf{V} \in \mathbb{C}^{m \times n}$. Then the matrix inversion lemma (also known as the Sherman-Morrison-Woodbury formula) says [GL96]

$$\left(\mathbf{X} + \mathbf{U}\mathbf{Y}\mathbf{V}^{\mathrm{H}}\right)^{-1} = \mathbf{X}^{-1} - \mathbf{X}^{-1}\mathbf{U}\left(\mathbf{Y}^{-1} + \mathbf{V}^{\mathrm{H}}\mathbf{X}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^{\mathrm{H}}\mathbf{X}^{-1}. \tag{A.1}$$

Determinant For any matrices $\mathbf{X} \in \mathbb{C}^{m \times n}$, $\mathbf{Y} \in \mathbb{C}^{n \times m}$ it holds (cf. [SXLK98])

$$\det\left(\mathbf{I}_m + \mathbf{X}\mathbf{Y}\right) = \det\left(\mathbf{I}_n + \mathbf{Y}\mathbf{X}\right). \tag{A.2}$$

Appendix B

Fundamental Concepts of Information Theory

In this section we will shortly review the fundamental concepts of information theory, namely *complex multivariate Gaussian (normal) random variables*, *differential entropy*, *mutual information* and *channel capacity* (the following can be found in [CT91]).

Regarding notation, upper case letters (e.g. X) designate random variables, while lower case letters (e.g. x) denote their values; if required, bold italic upper case letters (e.g. $\boldsymbol{X}$) indicate vectors of random variables, and bold lower case letters (e.g. $\mathbf{x}$) are vectors of their values. Further, the statistical properties of a complex random variable $Z = Z^{(\mathrm{r})} + jZ^{(\mathrm{i})}$ are determined by the joint PDF $p_{Z^{(\mathrm{r})}Z^{(\mathrm{i})}}(Z^{(\mathrm{r})}, Z^{(\mathrm{i})})$ of real and imaginary part, and for convenience of notation we define $p_Z(Z) \triangleq p_{Z^{(\mathrm{r})}Z^{(\mathrm{i})}}(Z^{(\mathrm{r})}, Z^{(\mathrm{i})})$ [NM93].

Complex multivariate Gaussian random variables Let $Z_1, Z_2, \ldots, Z_n$ denote complex jointly Gaussian random variables. Then the complex Gaussian random vector $\mathbf{z} \in \mathbb{C}^n$, given by ($\mathbf{z}^{(\mathrm{r})}, \mathbf{z}^{(\mathrm{i})} \in \mathbb{R}^n$)

$$\mathbf{z} = (z_1, z_2, \ldots z_n)^{\mathrm{T}} = \mathbf{z}^{(\mathrm{r})} + j\mathbf{z}^{(\mathrm{i})} = \left(z_1^{(\mathrm{r})} + jz_1^{(\mathrm{i})}, z_2^{(\mathrm{r})} + jz_2^{(\mathrm{i})}, \ldots z_n^{(\mathrm{r})} + jz_n^{(\mathrm{i})}\right)^{\mathrm{T}}, \quad \text{(B.1)}$$

is a realisation of the complex random variables $Z_1, Z_2, \ldots, Z_n$ having a multivariate normal distribution in real and imaginary part. We will assume that the random processes for real and imaginary part are uncorrelated. Further, it will be assumed that $\mathbf{z}$ has mean

$$\mathrm{E}\{\mathbf{z}\} = \boldsymbol{\mu}_{\mathbf{z}} = \mathrm{E}\left\{\mathbf{z}^{(\mathrm{r})}\right\} + j\mathrm{E}\left\{\mathbf{z}^{(\mathrm{i})}\right\} = \boldsymbol{\mu}_{\mathbf{z}^{(\mathrm{r})}} + j\boldsymbol{\mu}_{\mathbf{z}^{(\mathrm{i})}}.$$

With zero mean random variables

$$\mathbf{z}_0 = \mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}, \qquad \mathbf{z}_0^{(\mathrm{r})} = \mathbf{z}^{(\mathrm{r})} - \boldsymbol{\mu}_{\mathbf{z}^{(\mathrm{r})}}, \qquad \mathbf{z}_0^{(\mathrm{i})} = \mathbf{z}^{(\mathrm{i})} - \boldsymbol{\mu}_{\mathbf{z}^{(\mathrm{i})}},$$

the covariance matrix $\mathbf{K}_{\mathbf{zz}}$ of the complex vector $\mathbf{z}$ is defined as [Pap91]

$$\begin{aligned}
\mathbf{K}_{\mathbf{zz}} &= \mathrm{E}\left\{\mathbf{z}_0\mathbf{z}_0^{\mathrm{H}}\right\} \\
&= \mathrm{E}\left\{\mathbf{z}_0^{(\mathrm{r})}\left(\mathbf{z}_0^{(\mathrm{r})}\right)^{\mathrm{T}}\right\} + \mathrm{E}\left\{\mathbf{z}_0^{(\mathrm{i})}\left(\mathbf{z}_0^{(\mathrm{i})}\right)^{\mathrm{T}}\right\} - \\
&\quad j\left(\mathrm{E}\left\{\mathbf{z}_0^{(\mathrm{r})}\left(\mathbf{z}_0^{(\mathrm{i})}\right)^{\mathrm{T}}\right\} - \mathrm{E}\left\{\mathbf{z}_0^{(\mathrm{i})}\left(\mathbf{z}_0^{(\mathrm{r})}\right)^{\mathrm{T}}\right\}\right) \\
&= \mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}} + \mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{i})}} - j\left(\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{i})}} - \mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{r})}}\right). \qquad \text{(B.2)}
\end{aligned}$$

If it holds that $\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}} = \mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{i})}}$ and $\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{i})}} = -\mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{r})}}$, then $\mathbf{z}$ is called a *proper* Gaussian random vector, and it holds $\mathbf{K}_{\mathbf{zz}} = 2\left(\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}} - j\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{i})}}\right)$ (see [NM93] for a detailed discussion of proper complex random processes). In particular, if real part $\mathbf{z}^{(\mathrm{r})}$ and imaginary part $\mathbf{z}^{(\mathrm{i})}$ of $\mathbf{z}$ are realisations of uncorrelated normal random processes, we obtain

$$\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{i})}} = -\mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{r})}} = \mathbf{0} \Rightarrow \mathbf{K}_{\mathbf{zz}} = 2\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}} = 2\mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{i})}},$$

and in the following we will always assume that this condition holds, if not otherwise stated.

Then, for $\mathbf{K}_{\mathbf{zz}} = 2\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}} = 2\mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{i})}}$, the joint PDF of $Z_1, Z_2, \ldots, Z_n$ is given by[1] [Pap91]

$$p(z_1, z_2, \ldots, z_n) = \frac{1}{\pi^n \det\left(\mathbf{K}_{\mathbf{zz}}\right)} \mathrm{e}^{-(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})^{\mathrm{H}} \mathbf{K}_{\mathbf{zz}}^{-1} (\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})}. \tag{B.3}$$

Real part $\mathbf{z}^{(\mathrm{r})}$ is normal distributed according to (remember: real and imaginary part are assumed uncorrelated)

$$p(z_1^{(\mathrm{r})}, z_2^{(\mathrm{r})}, \ldots, z_n^{(\mathrm{r})}) = \frac{1}{\sqrt{(2\pi)^n \det\left(\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}}\right)}} \mathrm{e}^{-\frac{1}{2}(\mathbf{z}^{(\mathrm{r})}-\boldsymbol{\mu}_{\mathbf{z}^{(\mathrm{r})}})^{\mathrm{T}} \mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}}^{-1} (\mathbf{z}^{(\mathrm{r})}-\boldsymbol{\mu}_{\mathbf{z}^{(\mathrm{r})}})}, \tag{B.4}$$

and the PDF of imaginary part $\mathbf{y}$ is defined accordingly.

We will denote the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K}$ with $\mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$, where it will become obvious from the context whether complex normal distribution (B.3) or real normal distribution (B.4) has to be applied.

Any affine transform of a proper Gaussian random vector $\mathbf{z}$

$$\mathbf{w} = \mathbf{A}\mathbf{z} + \mathbf{b}, \quad \mathbf{A} \in \mathbb{C}^{n \times m}, \mathbf{b} \in \mathbb{C}^n, \tag{B.5}$$

is again proper Gaussian with mean

$$\mathrm{E}\{\mathbf{w}\} = \mathrm{E}\{\mathbf{A}\boldsymbol{\mu}_{\mathbf{z}}\} + \mathbf{b} \tag{B.6}$$

and covariance matrix

$$\mathbf{K}_{\mathbf{ww}} = \mathbf{A}\mathbf{K}_{\mathbf{zz}}\mathbf{A}^{\mathrm{H}}. \tag{B.7}$$

Differential Entropy The differential entropy is a measure of uncertainty of a continuous[2] random variable. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$ be a vector of continuous random variables with the continuous joint PDF $p(x_1, x_2, \ldots, x_n)$ (we will shortly write $p(\mathbf{x})$ in the following). Then, the differential entropy of $\boldsymbol{X}$, denoted as $\mathcal{H}(\boldsymbol{X})$, is given by[3]

$$\mathcal{H}(\boldsymbol{X}) = -\int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} p(\mathbf{x}) \log p(\mathbf{x}) dx_n \ldots dx_1, \tag{B.8}$$

[1] Actually, the correct notation for the PDF would be $p_{Z_1 Z_2,\ldots}(z_1, z_2, \ldots, z_n)$, however, for convenience we rather write $p(z_1, z_2, \ldots, z_n)$. Hence, $p(z_1, z_2, \ldots, z_n)$ and $p(y_1, y_2, \ldots, y_n)$ refer to a different set of random variables, and are in fact different PDF, $p_{Z_1 Z_2,\ldots}(Z_1, Z_2, \ldots, Z_n)$ and $p_{Y_1 Y_2,\ldots}(y_1, y_2, \ldots, y_n)$.

[2] For discrete random variables solely the term *entropy* is used.

[3] Note that the integral in (B.8) is w.r.t. complex variables, which is defined as the integrals over real and imaginary part, i.e. $\int_z \cdot dz \triangleq \int_{z^{(\mathrm{r})}} \int_{z^{(\mathrm{i})}} \cdot dz^{(\mathrm{i})} dz^{(\mathrm{r})}$.

where the logarithm $\log(\cdot)$ can be chosen to base e or 2 . In the latter case, differential entropy is expressed in *bits* and base 2 will be the assumption throughout the remainder of this work (for base e differential entropy is expressed in *nats*).

The probabilistic relation between two vectors of random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)^{\mathrm{T}}$ is given by the joint PDF $p(\mathbf{x}, \mathbf{y})$. Then, the *conditional differential entropy* of the vector of random variables $\boldsymbol{X}$, given the vector $\boldsymbol{Y}$, is denoted as $\mathcal{H}(\boldsymbol{X}|\boldsymbol{Y})$ and defined by

$$\begin{aligned}\mathcal{H}(\boldsymbol{X}|\boldsymbol{Y}) &= \\ &= -\int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \int_{y_1=-\infty}^{\infty} \cdots \int_{y_m=-\infty}^{\infty} p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} dy_m \ldots dy_1 dx_n \ldots dx_1 \qquad \text{(B.9)} \\ &\overset{\text{(a)}}{=} \mathcal{H}(\boldsymbol{X}, \boldsymbol{Y}) - \mathcal{H}(\boldsymbol{Y}), \qquad \text{(B.10)}\end{aligned}$$

where equality in (a) follows from (B.8).

The differential entropy of a set of complex normal distributed random variables $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)^{\mathrm{T}} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\mathbf{zz}})$ is given by (follows from (B.3) in (B.8))

$$\mathcal{H}(\boldsymbol{Z}) = \log\left((\pi \mathrm{e})^n \det\left(\mathbf{K}_{\mathbf{zz}}\right)\right) \qquad \text{(B.11)}$$

For $\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}} = \mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{i})}}$, and $\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{i})}} = \mathbf{K}_{\mathbf{z}^{(\mathrm{i})}\mathbf{z}^{(\mathrm{r})}} = \mathbf{0}$, the differential entropy of real and imaginary part are equal and given by (follows from (B.4) in (B.8))

$$\mathcal{H}(\boldsymbol{Z}^{(\mathrm{r})}) = \frac{1}{2} \log\left((2\pi \mathrm{e})^n \det\left(\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}}\right)\right). \qquad \text{(B.12)}$$

Note that, since $\det\left(\mathbf{K}_{\mathbf{zz}}\right) = 2^n \det\left(\mathbf{K}_{\mathbf{z}^{(\mathrm{r})}\mathbf{z}^{(\mathrm{r})}}\right)$, we can write

$$\mathcal{H}(\boldsymbol{Z}) = \mathcal{H}(\boldsymbol{Z}^{(\mathrm{re})}) + \mathcal{H}(\boldsymbol{Z}^{(\mathrm{im})}), \qquad \text{(B.13)}$$

i.e., if $Z_1, Z_2, \ldots, Z_n$ are proper complex random processes, with uncorrelated real and imaginary parts, the differential entropy of the complex random variables is simply the sum of the differential entropy of real and imaginary part random variables.

Mutual Information The mutual information $\mathcal{I}(X;Y)$ is the reduction in the uncertainty of X due to the knowledge of Y. The *mutual information* $\mathcal{I}(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_m)$ between two sets of random variables (complex or real) with joint PDF $p(\mathbf{x}, \mathbf{y})$ is given by

$$\begin{aligned}&\mathcal{I}(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_m) = \mathrm{E}\left\{\log \frac{p(\boldsymbol{X}, \boldsymbol{Y})}{p(\boldsymbol{X})p(\boldsymbol{Y})}\right\}_{\boldsymbol{X},\boldsymbol{Y}} = \\ &= \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \int_{y_1=-\infty}^{\infty} \cdots \int_{y_m=-\infty}^{\infty} p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} dy_m \ldots dy_1 dx_n \ldots dx_1 \\ &\overset{\text{(a)}}{=} \mathcal{H}(X_1, X_2, \ldots, X_n) - \mathcal{H}(X_1, X_2, \ldots, X_n | Y_1, Y_2, \ldots, Y_m) \\ &= \mathcal{H}(Y_1, Y_2, \ldots, Y_m) - \mathcal{H}(Y_1, Y_2, \ldots, Y_m | X_1, X_2, \ldots, X_n), \qquad \text{(B.14)}\end{aligned}$$

where equality in (a) follows directly from the definitions in (B.8) and (B.9).

Conditional mutual information $\mathcal{I}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{H})$ is the reduction in the uncertainty of $\boldsymbol{X}$ due to knowledge of $\boldsymbol{Y}$ when $\boldsymbol{H}$ is given, and can be written as (cf. (B.14))

$$\mathcal{I}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{H}) = \mathcal{H}(\boldsymbol{X}|\boldsymbol{H}) - \mathcal{H}(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{H}) = \mathrm{E}\left\{\log\frac{p(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{H})}{p(\boldsymbol{X}|\boldsymbol{H})p(\boldsymbol{Y}|\boldsymbol{H})}\right\}_{\boldsymbol{X},\boldsymbol{Y}}. \tag{B.15}$$

Information Capacity A communication channel can be viewed as a system in which complex symbols S at the channel input are received at the output as symbols R, whereas the channel is characterised by the probabilistic relation between symbols S and R, given by the conditional PDF $p(r|s)$. For the discrete-time single-source AWGN channel it holds $r_k = s_k + z_k$ (k is the discrete-time index), where the real and imaginary part of the noise z_k are normal distributed each with variance σ_n^2, uncorrelated (proper complex Gaussian random process), and zero mean; r_k, s_k, z_k are values of random variables R, S and Z. It holds $Z \sim \mathcal{N}(0, 2\sigma_\mathrm{n}^2)$. The capacity of the AWGN channel $\mathcal{C}_\mathrm{AWGN}$ is the maximum rate at which information can be sent over the channel with arbitrary low probability of error, and is given as

$$\mathcal{C}_\mathrm{AWGN} = \max_{p(s):\mathrm{E}\{S^2\}\leq P} \mathcal{I}(S;R), \tag{B.16}$$

where maximisation is over all input distributions $p(s)$, and P is a power constraint on the input symbols S. The maximum in (B.16) is obtained for $S \sim \mathcal{N}(0, P)$, and the capacity is given by the well known equation

$$\begin{aligned}\mathcal{C}_\mathrm{AWGN} &= \mathcal{H}(R) - \mathcal{H}(R|S) = \mathcal{H}(R) - \mathcal{H}(Z) = \\ &= \log\left((\pi\mathrm{e})^n\left(P + 2\sigma_\mathrm{n}^2\right)\right) - \log\left((\pi\mathrm{e})^L 2\sigma_\mathrm{n}^2\right) = \log\left(1 + \frac{P}{2\sigma_\mathrm{n}^2}\right) \textit{ bits per channel use}.\end{aligned} \tag{B.17}$$

Note that by assuming that S is normal distributed random variable implicates a continuous input alphabet, which is in contrast to the usual modulation techniques that use a discrete input alphabets, such as quadrature amplitude modulation (QAM). And indeed, the rates achieved with discrete input alphabets are strictly smaller than the channel capacity achieved by continuous Gaussian input alphabet [Ung81].

Another point to address is the definition of channel capacity for fading channels. Various definitions are possible, depending on assumptions regarding availability of information of the channel fading state (CSI) at transmitter and receiver side (e.g. [EB98]). Fig. B.1 shows the single-source fading channel model, including the random processes that model the fading itself H, the CSI available at the transmitter U and at the receiver V. Consider now the very simple discrete-time, single-source flat-fading channel model [CT91, EB98]

$$r_k = h_k s_k + n_k, \tag{B.18}$$

where it is assumed that the fading process is ergodic. Further assuming that no CSI is available at the transmitter (i.e. U is independent of H), and perfect CSI at the receiver (i.e. $V = H$), then the *ergodic* capacity of the fading channel is given by [EB98]

$$\begin{aligned}\mathcal{C}_\mathrm{RCSI} &= \max_{p(s):\mathrm{E}\{|s|^2\}\leq P} \mathcal{I}(S;R|H) = \mathcal{H}(R|H) - \mathcal{H}(R|S,H) = \mathcal{H}(R|H) - \mathcal{H}(Z) = \\ &= \mathrm{E}\left\{\log\left(1 + \frac{|H|^2 P}{2\sigma_\mathrm{n}^2}\right)\right\}_H,\end{aligned} \tag{B.19}$$

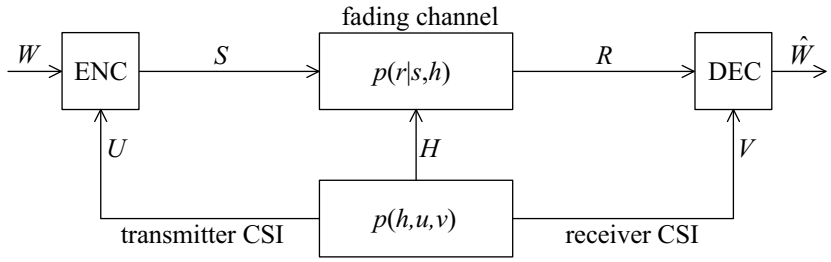


Figure B.1: Block diagram of the fading channel with time-varying channel state H, transmitter CSI U and receiver CSI V.

which directly follows (with (B.9) and $p(x,y) = p(x|y)p(y)$) from

$$\mathcal{H}(R|H) = \int_h p(h) \underbrace{\int_r p(r|h) \log\left(p(r|h)\right) dr}_{=\log(\pi \mathrm{e}(|h|^2 P + 2\sigma_{\mathrm{n}}^2))} dh. \tag{B.20}$$

Eqn. B.19 is the definition of ergodic capacity (also referred to as the *Shannon*, and *throughput* capacity of the fading channel), which assumes that the codeword length is chosen long enough to average over the fading. This introduces a delay depending on how fast the fading varies. If there is a constraint on the maximum tolerable delay, introduced by large codeword lengths, to be considered, then *delay-limited capacity* and *outage capacity* prove more appropriate [TH98a, TH98b, EB98]. However, in this work we will restrict to ergodic capacities.

Appendix C

Optimal Adaptive Beamforming in a Fading Channel

We have noted in the last section that the achievable rate for independent decoding depends solely on the SINR given by (4.37), which is obtained by source individual optimal beamforming [SXLK98]. Therefore, as it is well known, optimum beamforming does not reduce the maximal achievable rate of the corresponding source.

However, from (3.89) it is evident that the optimal beamforming weight has to account for the joint fading state $\mathbf{h}$ and, bearing in mind a concrete implementation of independent decoding, this means that all M optimal weight vectors have to be computed according to the instantaneous joint fading state. This raises the question what will be the related rate losses if we consider only *average* receive powers in the calculation of the beamformer weight vectors, which would bring a considerable saving in complexity of weight computation, because the steering vectors and the average receive powers vary much slower than the fast fading.

Let $\mathbf{w}_{\overline{\mathrm{H}},m}$ be the weight vector for source m computed based on average receive powers $\mathrm{E}\left\{P_i H_i\right\}_{H_i} = P_i$ of the interferers (i.e. $\forall i \in \mathcal{M} \setminus m$), and defined according to

$$\mathbf{w}_{\overline{\mathrm{H}},m} = \beta \left(2\sigma_{\mathrm{n}}^2 \mathbf{I}_L + \sum_{i \in \mathcal{M} \setminus m} \mathbf{a}_i \mathbf{a}_i^{\mathrm{H}} P_i \right)^{-1} \mathbf{a}_m. \tag{C.1}$$

Then the achievable rate $R_{\overline{\mathrm{H}},m}$ for source m is given by

$$R_{\overline{\mathrm{H}},m} = \mathrm{E}\left\{ \log_2 \left(1 + \frac{\left| \mathbf{w}_{\overline{\mathrm{H}},m}^{\mathrm{H}} \mathbf{a}_m \right|^2 P_m |H_m|^2}{\mathbf{w}_{\overline{\mathrm{H}},m}^{\mathrm{H}} \left(2\sigma_{\mathrm{n}}^2 \mathbf{I}_L + \sum_{i \in \mathcal{M} \setminus m} \mathbf{a}_i \mathbf{a}_i^{\mathrm{H}} P_i |H_i|^2 \right) \mathbf{w}_{\overline{\mathrm{H}},m}} \right) \right\}_{\boldsymbol{H}}. \tag{C.2}$$

It is evident that the rate loss from R_m to $R_{\overline{\mathrm{H}},m}$ depends both on the fading statistics, as well as on the steering vectors, and it seems hardly possible to obtain an exact value for the rate loss by means of arithmetic manipulations of (C.2) and, respectively, (4.37). However, it can be easily verified that the optimum beamforming weights become independent of the fast-fading factors for

the special cases that all steering vectors are either identical or all are mutually orthogonal. The equation for the optimum beamforming weight according to (3.89) can be rewritten as

$$\mathbf{w}_{\mathrm{opt},m}(\mathbf{h}) = \beta\left(\mathbf{I}_L + \frac{1}{2\sigma_{\mathrm{n}}^2}\mathbf{A}(\mathcal{I}_m)\mathbf{H}(\mathcal{I}_m)\mathbf{K}_{\mathbf{s}}(\mathcal{I}_m)\mathbf{H}(\mathcal{I}_m)^{\mathrm{H}}\mathbf{A}(\mathcal{I}_m)^{\mathrm{H}}\right)^{-1}\mathbf{a}_m \tag{C.3}$$

$$= \beta\left(\mathbf{I}_L - \frac{1}{2\sigma_{\mathrm{n}}^2}\mathbf{A}(\mathcal{I}_m)\left(\left(\mathbf{H}(\mathcal{I}_m)\mathbf{K}_{\mathbf{s}}(\mathcal{I}_m)\mathbf{H}(\mathcal{I}_m)^{\mathrm{H}}\right)^{-1} + \mathbf{A}(\mathcal{I}_m)^{\mathrm{H}}\mathbf{A}(\mathcal{I}_m)\right)^{-1}\mathbf{A}(\mathcal{I}_m)^{\mathrm{H}}\right)\mathbf{a}_m, \tag{C.4}$$

$$\mathcal{I}_m = \mathcal{M} \setminus m.$$

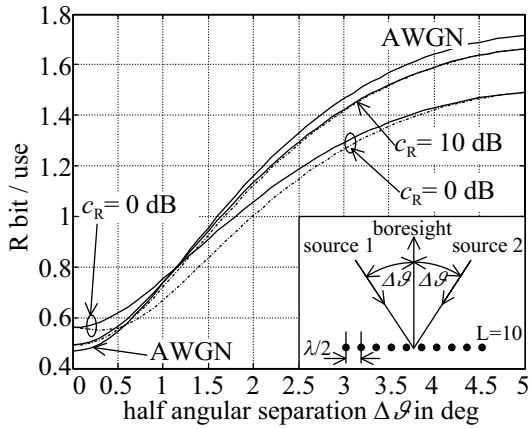
where we have used the matrix inversion lemma (A.1).

Furthermore, if all steering vectors are equal, i.e. $\mathbf{a}_m = \mathbf{a}_i = \mathbf{a}$, $m, i \in \mathcal{M}$, it can be shown that $\mathbf{a}$ is an Eigenvector of the inverse in (C.3), such that the optimal weight vector must be a scaled version of $\mathbf{a}$, and, finally, for orthogonal steering vector it holds $\mathbf{A}(\mathcal{I}_m)^{\mathrm{H}}\mathbf{a}_m = (0, 0, \ldots, 0)^{\mathrm{T}}$. Therefore it readily follows with (C.4) that the optimal weight vector is independent of the fading factors for equal or orthogonal steering vectors, such that

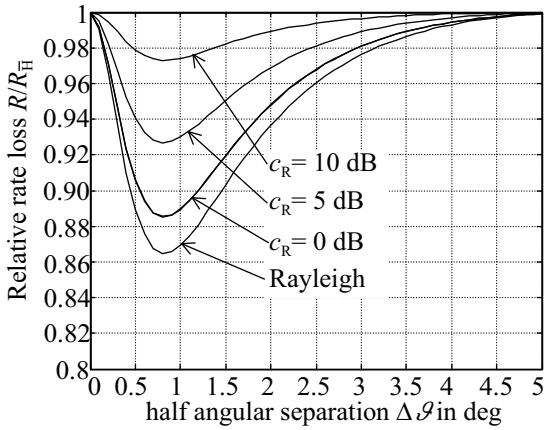
$$\mathbf{a}_m = \mathbf{a}_i \quad \text{or} \quad \mathbf{a}_i^{\mathrm{H}}\mathbf{a}_m = 0 \Rightarrow \mathbf{w}_{\mathrm{opt},m} = \beta_1\mathbf{a}_m, \mathbf{w}_{\overline{\mathrm{H}},m} = \beta_2\mathbf{a}_m. \tag{C.5}$$

For arbitrary steering vectors $R_m = R_{\overline{\mathrm{H}},m}$ holds only if $H_m = H_i = \text{const.}$, $\forall i, m \in \mathcal{M}$, thus otherwise $R_m > R_{\overline{\mathrm{H}},m}$ must hold.

Fig. C.1 shows as an example the rate loss due to beamforming based on the average receive powers for two sources and a linear array of $L = 10$ elements (omnidirectional). Clearly the relative rate loss $R_m/R_{\overline{\mathrm{H}},m}$ shown in Fig. C.1(b) depends on the angular separation $2\Delta\vartheta$ (cf. Fig. C.1(a)) of the two source signal, further on the fading statistics parameterised by the Rice factor c_{R}.



(a) Comparison of R (solid) and $R_{\overline{\mathrm{H}}}$ (dotted) as a function of source signal separation $2\Delta\vartheta$ and the Rice factor c_{R}.



(b) Relative rate loss $R/R_{\overline{\mathrm{H}}}$

Figure C.1: Comparison of rate R_m achieved with optimal beamforming and rate $R_{\overline{\mathrm{H}},m}$ achieved with beamforming based on average receive powers for two sources and depending on angular separation $2\Delta\vartheta$ (symmetric, $R_1 = R_2 = R$ and $R_{\overline{\mathrm{H}},1} = R_{\overline{\mathrm{H}},2} = R_{\overline{\mathrm{H}}}$). Further parameters: ULA with $L = 10$ omnidirectional elements, fading normalised to unit power $\mathrm{E}\left\{|H_1|^2\right\} = \mathrm{E}\left\{|H_2|^2\right\} = 1$, input $\gamma_1 = \gamma_2 = 1$.

Appendix D

Creating Orthogonal Beams via 2-D DFT

We have to start this section with some definitions [Dub85]. A lattice Λ in $\mathbb{R}^D$ is the set of all linear combinations with integer coefficients of linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_D \in \mathbb{R}^D$, such that

$$\Lambda = \{ i_1\mathbf{v}_1 + i_2\mathbf{v}_2 + \cdots + i_D\mathbf{v}_D |\, i_d \in \mathbb{Z}, d = 1, 2, \ldots, D \}\,. \tag{D.1}$$

A *unit cell* of a lattice Λ is defined as a set $\mathcal{U} \subset \mathbb{R}^D$ such that the disjoint union of copies of $\mathcal{U}$, centred on each lattice point, constitutes again $\mathbb{R}^D$:

$$\begin{aligned} &(\mathcal{U} + \mathbf{x}) \cap (\mathcal{U} + \mathbf{y}) = 0 \quad \text{(copies do not overlap)}, \\ &\bigcup_{\mathbf{x}} (\mathcal{U} + \mathbf{x}) = \mathbb{R}^D, \quad \mathbf{x}, \mathbf{y} \in \Lambda, \mathbf{x} \neq \mathbf{y} \end{aligned} \tag{D.2}$$

There are many definitions of the unit cell possible for a given lattice, here we will define the unit cell as the Voronoi cell. The Voronoi cell is defined as the set of points in $\mathbb{R}^D$ being closer to the origin $\mathbf{0}$ than to any other point on the lattice.

In the further only 2-D lattices are considered (i.e. $D = 2$). The spatial sampling lattice Λ is thus given by a basis $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2) \in \mathbb{R}^{2\times 2}$. Fig. D.1(a) shows a section of the triangular sampling lattice Λ, where next to some lattice points the corresponding indices are indicated.

The signals incident on the antenna array are spatially sampled in the x-y-plane at element positions given by $\mathbf{p}_l \in \Lambda$:

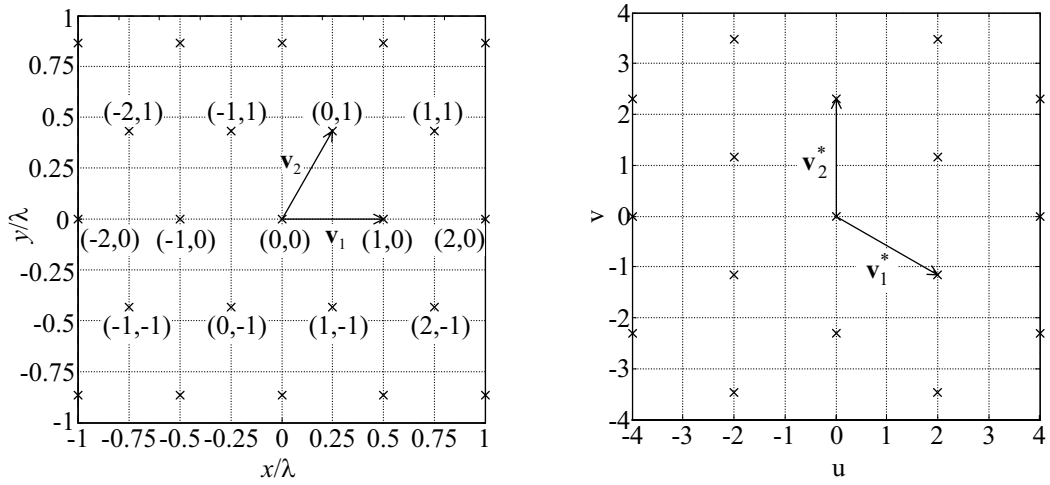
$$\mathbf{p}_l = \mathbf{V}\mathbf{i}_l, \quad l = 1, 2, \ldots, L, \tag{D.3}$$

where the vectors $\mathbf{i}_l \in \mathbb{Z}^2$ are chosen according to the indices indicated in Fig. D.1(a). Further we define a subset $\mathcal{E} \subset \Lambda$, being the points of Λ associated with the array elements:

$$\mathcal{E} = \{ \mathbf{x} |\, \mathbf{x} = \mathbf{V}\mathbf{i}_l, l = 1, 2, \ldots, L \}\,. \tag{D.4}$$

Just as in the 1-D case, where the Fourier transform of a discrete signal is periodic, the 2-D Fourier transform of a 2-D signal sampled on lattice Λ is periodic in the spatial frequency domain. The periodicity is characterised by the *periodicity lattice* Λ^*, which has the basis $\mathbf{V}^* = (\mathbf{V}^{-1})^{\mathrm{T}}$. Λ^* is also termed the *reciprocal lattice* of Λ. For $\mathbf{V}$ given according to (3.34) it holds

$$\mathbf{V}^* = \frac{1}{d_\mathrm{s}} \begin{pmatrix} 1 & 0 \\ -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{3}} \end{pmatrix}. \tag{D.5}$$



(a) Spatial sampling lattice Λ with basis $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2)$.

(b) Periodicity lattice Λ^* in the spatial frequency domain with basis $\mathbf{V}^* = (\mathbf{v}_1^*, \mathbf{v}_2^*)$.

Figure D.1: Spatial sampling lattice Λ and periodicity lattice in the spatial frequency domain Λ^*.

A section of the periodicity lattice Λ^* is shown in Fig. D.1(b).

The spatially discrete sampled signal is given by $\mathbf{r} = (r_1, r_2, \ldots, r_L)^{\mathrm{T}}$ according to (3.64). Then the 2-D spatial Fourier transform of $\mathbf{r}$ is given by

$$R(\boldsymbol{\nu}) = \sum_{l=1}^{L} r_l \mathrm{e}^{-j2\pi\boldsymbol{\nu}^{\mathrm{T}}\mathbf{V}\mathbf{n}_l}, \tag{D.6}$$

where $\boldsymbol{\nu} = 1/\lambda(u, v)^{\mathrm{T}} \in \mathbb{R}^2$ is the spatial frequency vector.

As already mentioned $R(\boldsymbol{\nu})$ is periodic with respect to Λ^*, such that

$$R(\boldsymbol{\nu}) = R(\boldsymbol{\nu} + \boldsymbol{\nu}^*), \quad \boldsymbol{\nu}^* \in \Lambda^*. \tag{D.7}$$

The periodicity of the Fourier transform corresponds to eventual occurrence of grating lobes when considering beamforming.

Eqn. (D.6) provides a spatial spectrum continuous in $\boldsymbol{\nu}$, and, because we are aiming at the DFT, $R(\boldsymbol{\nu})$ has to be discretised by sampling on a spatial frequency sampling lattice Γ^*. The discretisation of $R(\boldsymbol{\nu})$ introduces a periodisation of the sampled signal in the spatial domain.

The spatial periodicity lattice $\Gamma \subseteq \Lambda$ is defined by a basis $\mathbf{B}$, which can be described by

$$\mathbf{B} = \mathbf{V}\mathbf{N}, \tag{D.8}$$

where $\mathbf{N} \in \mathbb{Z}^{2\times 2}$ is an integer full rank matrix. Here it is required that $\mathbf{N}$ is chosen such that with (D.4) it holds

$$\begin{aligned} &(\mathcal{E} + \mathbf{x}) \cap (\mathcal{E} + \mathbf{y}) = 0, \\ &\bigcup_{\mathbf{x}} (\mathcal{E} + \mathbf{x}) = \Lambda, \quad \mathbf{x}, \mathbf{y} \in \Gamma. \end{aligned} \tag{D.9}$$

We will call $(\mathcal{E}+\mathbf{x}), \mathbf{x} \in \Gamma$, a Γ-period of Λ. Eqn. (4.82) provides $\mathbf{N}$ for hexagonal arrays that suffices (D.9).

Now that the spatial periodicity lattice Γ is defined, the sampling lattice Γ^* in the spatial frequency domain, being the reciprocal lattice of Γ, is given as well. A basis $\mathbf{B}^*$ of the spatial frequency sampling lattice Γ^* is given by

$$\mathbf{B}^* = \left(\mathbf{B}^{-1}\right)^{\mathrm{T}} \tag{D.10}$$

and the discretised version of $R(\boldsymbol{\nu})$ is obtained by setting

$$\boldsymbol{\nu} = \mathbf{B}^*\mathbf{k}, \qquad \mathbf{k} \in \mathbb{Z}^2. \tag{D.11}$$

Then (D.6) can be rewritten according to

$$R(\mathbf{B}^*\mathbf{k}) = \sum_{l=1}^{L} r_l \mathrm{e}^{-j2\pi\mathbf{k}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{V}\mathbf{i}_l} = \sum_{l=1}^{L} r_l \mathrm{e}^{-j2\pi\mathbf{k}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{i}_l}, \quad \mathbf{k} \in \mathbb{Z}^2, \tag{D.12}$$

where it is used that according to (D.8) it must hold that $\mathbf{B}^{-1}\mathbf{V} = \mathbf{N}^{-1}$.

In contrast to (D.12), where $\mathbf{k} \in \mathbb{Z}^2$ ($R(\mathbf{B}^*\mathbf{k})$ contains all periods), it suffices to restrict to a single Λ^*-period of $R(\mathbf{B}^*\mathbf{k})$. Those L indices belonging to a single period of $R(\mathbf{B}^*\mathbf{k})$ are denoted with $\mathbf{k}_l, L = 1, 2, \ldots, L$. Hence, we have to finally find the proper L indices $\mathbf{k}_l$ that belong to one Λ^*-period in the spatial frequency domain, i.e.

$$\mathbf{k}_l = \left\{\mathbf{x} | \mathbf{x} \in (\mathcal{U}^* \cap \Gamma^*)\right\}, \tag{D.13}$$

where $\mathcal{U}^*$ is a unit cell of lattice Λ^*. With this last definition (D.13) the DFT beamforming weight vectors can be written as

$$(\mathbf{w}_l^{\mathrm{DFT}})^{\mathrm{H}} = \left(\mathrm{e}^{-j2\pi\mathbf{k}_l^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{i}_1}, \mathrm{e}^{-j2\pi\mathbf{k}_l^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{i}_2}, \ldots, \mathrm{e}^{-j2\pi\mathbf{k}_l^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{i}_L}\right), \quad l = 1, 2, \ldots, L, \tag{D.14}$$

It is *not* required that the $\mathbf{k}_l$ are chosen from the same Λ^*-period and can be easily obtained via the transformation

$$\mathbf{k}_l = \mathbf{N}\mathbf{i}_l. \tag{D.15}$$

The question in what spatial directions the L orthogonal beams are steered to is answered by setting the DOA dependent phase at array element l equal to the corresponding phase of the DFT beamforming weight for element l. Thus it holds

$$2\pi \begin{pmatrix} u_l \\ v_l \end{pmatrix}^{\mathrm{T}} \mathbf{V}\mathbf{i}_l = 2\pi\mathbf{k}_l^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{V}\mathbf{i}_l \tag{D.16}$$

Eqn. (D.16) holds in particular if

$$\begin{pmatrix} \cos\varphi_l \sin\vartheta_l \\ \sin\varphi_l \sin\vartheta_l \end{pmatrix} = \left(\mathbf{B}^{-1}\right)^{\mathrm{T}} \mathbf{k}_l, \tag{D.17}$$

with $u_l = \cos\varphi_l \sin\vartheta_l$ and $v_l = \sin\varphi_l \sin\vartheta_l$ according to the definitions in (3.19). Thus, if we define $(\upsilon_l, \nu_l)^{\mathrm{T}} = \left(\mathbf{B}^{-1}\right)^{\mathrm{T}} \mathbf{k}_l$, it holds (cf. (3.20))

$$\varphi_l = \arctan\frac{\nu_l}{\upsilon_l} \tag{D.18}$$

$$\vartheta_l = \arcsin\sqrt{\upsilon_l^2 + \nu_l^2}. \tag{D.19}$$

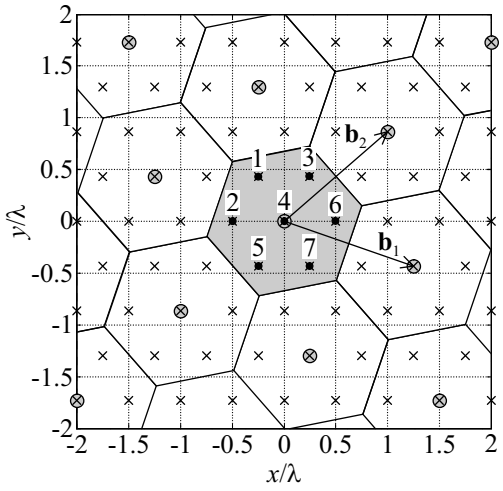
Of course, (D.19) provides real angles ϑ_l only if $(\upsilon_l^2 + \nu_l^2) \leq 1$, i.e. the vector $(\upsilon_l, \nu_l)^{\mathrm{T}}$ has a norm not greater than unity.

Again, to allow better understanding of the concepts just introduced, a simple example is presented in the following.

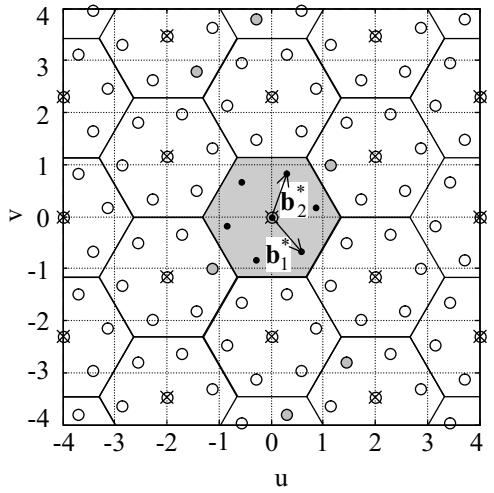
For a hexagonal array with $L = 7$ elements, the $\mathbf{i}_l$ can be chosen according to (not unique)

$$\begin{aligned} \mathbf{i}_1 &= \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \mathbf{i}_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \mathbf{i}_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mathbf{i}_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\ \mathbf{i}_5 &= \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \mathbf{i}_6 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{i}_7 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \end{aligned} \tag{D.20}$$

Besides the according indexing of the $L = 7$ array elements, Fig. D.2(a) shows the spatial periodicity lattice Γ for the hexagonal array. Further, Fig. D.2(b) shows sampling lattice in the spatial frequency domain Γ^*. Also the Λ^*-periods of Γ^* are shown there, indicating the proper indices $\mathbf{k}_l$ required for the calculation of the DFT beamforming weight vectors according to (D.14).



(a) Periodicity lattice Γ (filled grey circles) with basis $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2)$. The Γ-periods of Λ are indicated by hexagons. The unit cell of Γ contains those point of Λ corresponding with the array elements (filled black circles, numbered 1 to 7).



(b) Sampling lattice in spatial frequency domain Γ^* (circles) with basis $\mathbf{B}^* = (\mathbf{b}_1^*, \mathbf{b}_2^*)$. The Λ^*-periods are indicated by hexagons. The unit cell of Λ^* contains those points of Γ^* (filled circles) providing the indices $\mathbf{k}_l$ according to (D.13). Those indices $\mathbf{k}_l$ derived from (D.15) are shown as grey filled circles.

Figure D.2: Spatial periodicity lattice Γ and sampling lattice in the spatial frequency domain Γ^* for hexagonal array with $L = 7$ array elements. The unit cells in the space and spatial frequency domain, respectively, is the corresponding Voronoi cell. For the DFT it suffices to restrict to samples lying in one Γ- and Λ^*-period, respectively, e.g. in the unit cell centered on the origin of the x-y- and u-v-plane (grey hexagons).

Choosing the indices of the Λ^*-period contained in the unit cell (highlighted as a grey hexagon in Fig. D.2(b)) we obtain for the $\mathbf{k}_l$ (ordering arbitrary)

$$\begin{aligned} \mathbf{k}_1 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{k}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{k}_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{k}_4 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ \mathbf{k}_5 &= \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \mathbf{k}_6 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \mathbf{k}_7 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \end{aligned} \tag{D.21}$$

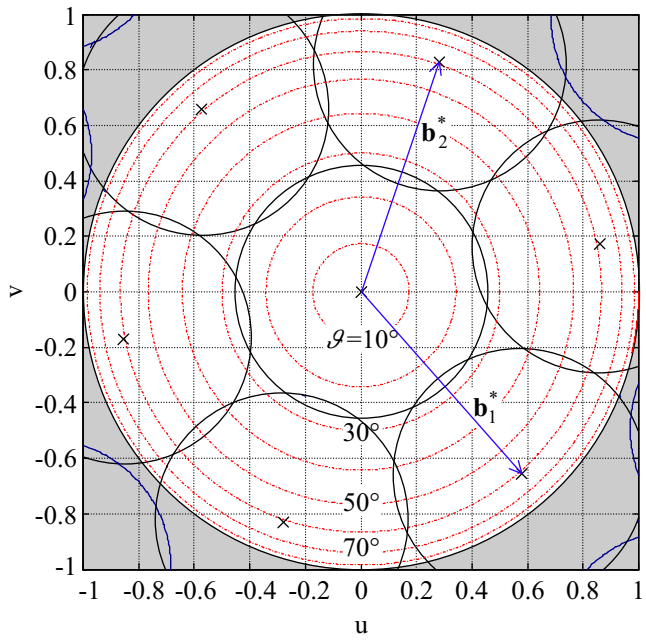


Figure D.3: 4.3 dB beam contours of the $L = 7$ orthogonal DFT beams for an hexagonal array of $L = 7$ omnidirectional elements (lattice spacing is $d = \lambda/2$) in the direction sine u-v-plane ($u = \cos\varphi \sin\vartheta$, $v = \sin\varphi \sin\vartheta$). Shown is the hemisphere for $0 \leq \vartheta \leq \pi/2$, i.e. $u^2 + v^2 \leq 1$. The area outside the unit circle (grey) indicates imaginary space, i.e. the beam contours shown in this region are not in real space.

Finally, Fig. D.3 shows the 4.3 dB-below-maximum spot beam contours of the 7 DFT-beams created by the 7-element array.