Frequency Assignment and Network Planning for Digital Terrestrial Broadcasting Systems

Frequency Assignment and Network Planning for Digital Terrestrial Broadcasting Systems

Roland Beutler

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 1-4020-7873-0
Print ISBN: 1-4020-7872-2

To Kai and Chris

Contents

Preface

Frequency management and network planning of terrestrial digital broadcasting systems are not counted among the exact natural sciences. Communication science and electronic engineering can be considered as being the basis. However, they do not fulfill all demands thereto. Typical frequency planning problems exhibit such a high level of complexity that very sophisticated mathematical methods have to be consulted. Last but not least, a profound knowledge in physics is required, too, since any transmission of information is governed by the Maxwell equations in any case. This is independent from the way the transmission is actually accomplished, i.e. if either terrestrial transmitter networks are used, cable connections or satellite links.

Further, a potential approach to frequency assignment or network planning which solely rests on rigorous scientific principles only partially covers the whole variety of different tasks. Every planning strategy must be embedded into a corresponding media political milieu. On the level of one individual sovereign state different rules apply than in an international environment. The larger the influence of politics on the planning process becomes the more solutions to frequency assignment problems are found by negotiations and not by the application of mathematical optimization strategies.

Frequency or network planner usually is not a job people are taking up by intention as they for example choose the focus of their academic studies. In most cases pure chance plays a major role. Hence, it comes as no surprise that in the field of frequency and network planning people from nearly all technically oriented faculties can be found. There are engineers from communication science to electronics accompanied by mathematicians and last but not least a lot of physicists. Very likely all of them share one common experience, namely the first steps on the new ground called frequency or network planning proves to be rather weary. Basically, this is connected to the fact that in particular network planning is an activity where a lead in knowledge gives a direct competitive edge. Consequently, successful planning approaches are published only to the extend that the proper ascendancy in the market of network providers is not affected by given away know-how. Such an attitude is very common as can be seen for example from the situation on the radio or TV receiver market. All manufactures need to build their products in accordance with well-known published standards. Nevertheless, there are good and bad receivers to be bought at nearly the same price.

Beside the lack of officially published documentation the highly practically oriented working method of the planners has to be considered a real obstacle for any novice. Many ideas will never be published in articles. Most of the work is carried out by project groups developing new concepts which are not documented very elaborately. Once the new ideas have been put into practice most of the presentations and manuscripts vanish again.

Clearly, the results of international frequency planning conferences are memorized very carefully and detailed. The documentation prepared under such circumstances must contain precise directions how to make use of the planning results. But at the same time they represent some kind of a very compressed condensate of information. For a newcomer this might lead to nearly the same level of forlornness as having no documentation at all.

This background and the experience of the author during his confrontation with frequency and network planning gave umbrage to the present book. It is meant to give insight into the problems and the thereby emerging strategies to find solutions for them. It is quite obvious in view of the manifold of different tasks under the umbrella of frequency and network planning that no exhaustive presentation of the whole field can be provided. Rather, it was intended to sketch the principle ideas. By means of several planning examples an idea about the characteristics of typical problems should be conveyed. Furthermore, possible ways to cope with these problems in order to arrive at useful results are addressed. The selection of documents referenced in this book has to be considered as a subjective subset of what might be relevant. Nevertheless, it gives first indications and hints for further reading.

The book is divided into three parts. The first part sets the frame for the problem of frequency and network planning for digital terrestrial broadcasting systems. The general context is discussed as well as the broadcasting systems under consideration. Furthermore, a very brief introduction into the principles of coverage prediction are given, too. The second part is concerned about frequency planning only. Different aspects are explained and discussed in detail by means of examples. Finally, in the third part network planning issues are addressed.

The subdivision of the book is not quite arbitrary, but is instead oriented at the so-called allotment planning approach which is the method of choice for digital terrestrial broadcasting systems nowadays. It will be presented later in full detail. For the time being, it is sufficient to mention that this means to assign a frequency to a given geographical area to be used there without specifying the details of a corresponding transmitter network implementation. Seen from that perspective, network planning always is the second step. Frequency assignment comes first.

This order should by no means suggest that both tasks are independent from each other. Rather, the converse statement is true. First of all this touches upon the definition of the coverage areas which are the basis for the frequency assignment process. Depending on the number and the shape of the coverage areas different strategies for solution are required. However, any frequency assignment for a particular area must take into account both the size of the envisaged coverage area and the topographical conditions encountered there. Large area coverage demands for frequencies that according to their corresponding wave propagation characteristics allow the implementation of an adequate transmitter network. This limits the maximum extension of a reasonable coverage area. A larger number of individual areas to be provided with frequencies gives rise to a higher spectrum demand. So, both frequency assignment and network planning are inseparably linked.

All examples presented here have their source in practice, even though especially in the case of network planning a purely hypothetical model has been used. Both the considered digital terrain model as well as the location of transmitter sites a pure fiction. Nevertheless, the qualitative properties of digital terrestrial broadcasting networks can be studied without problems on that basis. The consequences emerging from the results of the investigations have general validity since they shed light on the structure of problems rather than any particular implementation in terms of a model.

Most of the facts collected here trace back to the large amount of contributions for national and international working groups in the process of the preparation of the revision conference of the ITU frequency plan assembled in Stockholm in 1961. Therefore, this revision will be present everywhere in the book. It has to be emphasized, however, that the universality and applicability of the methods presented is thereby not brought into question. Rather, the preparation of the revision conference has to be considered as an inexhaustible source which can be exploited to develop and test new planning approaches. Future will tell to what extend the ideas discussed in this book will find their way into the forthcoming new frequency plans for digital terrestrial radio and television.

Chapter 1

Introduction

Resources, their availability as well as the unbarred access to them are driving forces for all civilizations. Food, clean water and clean air are essential, energy in all its aspects likewise. Usually, the term resource is automatically linked to material resources. In recent times, however, immaterial resources are getting more and more important. In the age of globalization states possessing only few or none natural resources are becoming increasingly dependent on the fact that their working population has sufficient technical or economical knowledge and know-how. This might allow to compensate for a lack of natural resources.

Services of very different kind take the place of pure processing of raw materials or the agricultural production of food. Of course, the generation of knowledge and its provision presumes a set of adequate communication possibilities. The so-called information age, as a consequence, promotes and pulls through a general digitization of information and all types of information carriers on every level in society. There is basically no part of our daily life that is not subjected to dramatic change. Without the help of computers, the synonym of digitization, probably nothing would work any more in all civilized societies around the globe in particular in those fields of activity which are to be considered essential. It seems, however, this development has just begun. Where it will lead to is far from clear. More and more tasks are currently deputed onto computers. The performance of state-of-art computer technology already today is breathtaking and every day new applications are presented showing new, so far not reached functionality.

The amazing changes taking place can nowhere better be seen as when looking at the incredible spread of the Internet. It took less than a decade to pull through a far reaching social transformation which compares to a phase transition. The networking of computers across the whole planet implies completely new economic and political structures whose profound consequences today are still to be discovered. By using the Internet people are able to communicate nearly without limitations in space and time. They can share any type of information by employing cable connections, satellite links or radio transmission. In most cases all three communication types are involved when establishing a point-to-point connection across a large distance.

A common feature of all these possibilities is that they make use of electromagnetic waves as the medium for the transmission of information. Electromagnetic spectrum is an immaterial resource. It is limited but in contrast to petrol the reserve will never run short. In addition, it is not subject to a reduction of quality like for example air which gets polluted by harmful emissions. Electromagnetic spectrum is available everywhere either on ground, under water or in outer space. Of course, not the all parts of the electromagnetic spectrum are equally well suited for the transmission of information. Figure 1.1 sketches the different regimes which according to their characteristics can be used for different applications.

One of the central problems the information society has to overcome in the future is to provide transmission systems which can cope with the permanently increasing demand of people for different types of communication. Clearly, the development and introduction of such systems is limited by the scarcity of appropriate electromagnetic spectrum. Now, if the amount of spectrum cannot be increased forthcoming generations of telecommunication systems must be designed to allow for a highly efficient usage of spectrum. Depending on the type of communication different approaches might be adapted.

Basically, there are three different communication levels to be distinguished. If one person tries to establish a communication link with a single other person this is simply called a point-to point connection. In addition, there exist point-to-multi-point and broadcasting scenarios. Point-to-multi-point is typically encountered in computer networks, for example when several people connect to the same webserver to download exactly the same content. The number of users thereby tends from one

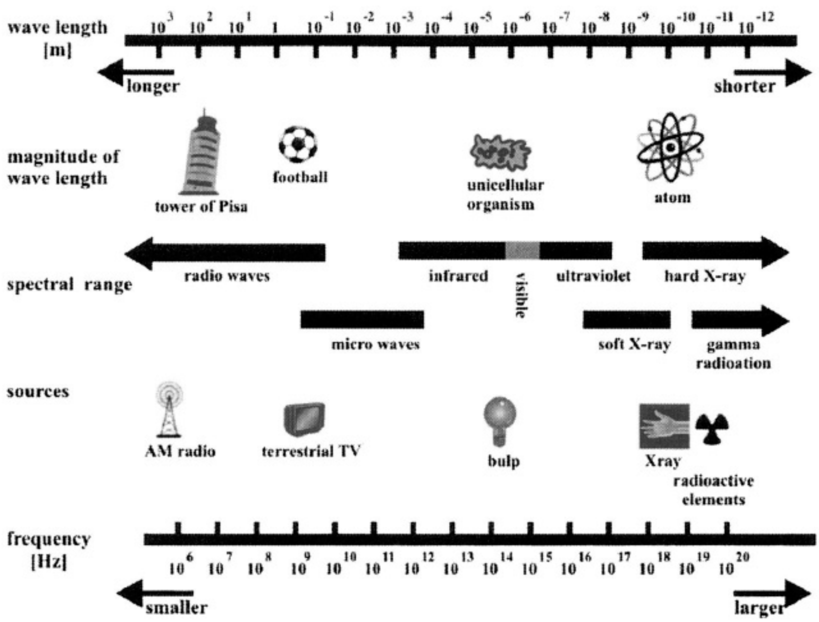

Figure 1.1: Survey of the electromagnetic spectrum.

to some hundred persons at the same time. Broadcasting, in contrast, means to distribute the same information to an in principle unlimited number of customers at the same time. It is evident that all of these communication types give rise to completely different requirements for the underlying system design. It should be noted, however, that this does not prohibit mutual compatibility and transparency. Indeed, it is crucial that interfaces have to be developed which allow for a seamless link between different communication platforms. Furthermore, it has become clear in recent years that in the future it will be mandatory for transmission systems to possess the ability of dynamic operation mode reconfiguration in order to allow for a very efficient usage of the available data capacities.

The described profound structural change started to seize both radio and television at the beginning of the nineties, too. On one side any production means were switched to digital. Digital cameras, digital cut but in first line digital archive systems are meanwhile taken as granted. On the other side, the delivery of programs to listeners and viewers, however, today (2004) is still in most cases accomplished in an analogue fashion. It is true that both on cable and satellite digital services are constantly gaining ground. But, so far the vast majority of customers still relies on analogue broadcasting technology.

Without doubt the situation will change in the future. On occasion of the Internationale Funkaustellung 2001 in Berlin German chancellor Gerhard Schröder confirmed in public during his opening speech that after 2010 a digital only provision of radio and television programs is to be realized in Germany. Analogue terrestrial TV will be switched off by then. FM radio is likely to exist for a somewhat longer period. However, the expectation of life for that type of service is to be considered ended around 2015, too.

Before the all digital terrestrial broadcasting scenario has become reality a lot of obstacles must be removed. A rigorous switch-over from analogue to digital transmission will probably not be possible. It seems that a more or less extended so-called simulcast period will be necessary for the introduction of digital terrestrial radio and television. Simulcast means the simultaneous transmission of one particular program both by analogue and digital means. The question how the costs naturally arising by that can be compensated for in an acceptable way cannot be answered yet. At least currently, the public opinion does not support the idea of longer lasting simulcast phases since nobody is willing to pay for that. On the other side, switching off all analogue transmitters at one chosen date would certainly create a lot of political and social concern and troubles. There seems to be no chance for any acceptance of such an idea as a general rule and consequently it could not be enforced.

Hence, strategies for the introduction of digital terrestrial broadcasting systems have to be developed which allow for a pampered and smooth transition. However, it is evident that this also will lead to a great variety of problems to be solved. In first place, spectrum scarcity has to be mentioned here. A brief look on analogue television in Germany might clarify this. Today analogue terrestrial television occupies the entire frequency range of band IV and V in UHF. Only in a very few numbers of occasions new TV channels can be found at a particular location if a gap in the existing coverage has to be closed. In most cases it is not possible to improve the provision with analogue TV programmes. Since digital terrestrial TV is to be implemented in the same frequency range it is unavoidable that its introduction must be accompanied by switch-off or rearrangement of analogue transmitter sites.

So far digital terrestrial radio and television have been mentioned without specifying any system characteristics. Nevertheless, it has been tacitly implied that in Europe the two standardized systems “Digital Audio Broadcast” (T-DAB) [ETS97a] and “Digital Video Broadcast” (DVB-T) [ETS97b] are to be introduced and hence treated here. Both systems have been designed for several different transmission environments. The letter ‘T’ in the abbreviations specifies the terrestrial variant of both systems. There exist standards for cable and satellite transmission for both systems, too. The position of the letter ‘T’ is purely historically based and has no deeper meaning.

In particular, in the United States of America there exist competing terrestrial radio and television systems which are also designed for terrestrial broadcasting purposes (see e.g. [ITU01b]). It seems as if at least in Europe and wider parts of Asia T-DAB and DVB-T are able to assert themselves to become the one and only digital terrestrial broadcasting standard. However, the race is not yet finished. Very recently in Japan a system has been put forward which combines radio, television and telecommunication under one single technical roof. Especially from the point of efficient usage of the spectrum such an approach might seem to be very attractive (see e.g. [ITU02]).

Efficient usage of the available spectrum is also one of the characteristic features of both T-DAB and DVB-T. A relatively large bandwidth occupied by the broadcast signal combined with sophisticated technical measures gives rise to very robust transmission systems. In particular, T-DAB has been designed and optimized for mobile reception. Portable or fixed reception are clearly possible without problems as well. In the DVB-T case the latter reception modes were the major design targets of the system. But even for mobile reception significant improvements could be achieved which are due to more efficient antennas and better design and performance of the electronic components of the receivers.

Analogue radio and television systems require a large amount of spectrum to provide a wide area coverage with one dedicated content. In order to serve an area in the order of 200 km $\times$ 200 km with only one FM radio or analogue television program several frequencies are needed. A typical FM high power transmitter has an output of about 100 kW. In case there is no other transmitter around using the same FM frequency an area of up to 100 km radius depending on the topography could be served.

In practice, such an isolated existence is never met anywhere. As a consequence, there are a extended areas where the signals from different co-frequency transmitters superimpose at the location of the receiver. Analogue radio or television receivers cannot cope very well with such a superposition of different contributions from different sources. Usually, the reception will then fail. Basically, this leads to a reduction of the theoretically achievable coverage area. Moreover, even if two transmitters using the same frequency would broadcast the same content a FM radio or an analogue television receiver would have problems decoding the information properly. This is the reason why in an extended area of the dimensions just mentioned 10 or more high power transmitters using different frequencies are utilized. Under such conditions they have a typical mean inter-transmitter distance of about 50 km. The important issue, however, is that each transmitter needs to use its own frequency well separated from all others.

Both T-DAB and DVB-T are based on techniques which differ completely from those employed in their analogue counterparts. First of all, several programs are bundled into so-called multiplexes and thus broadcast simultaneously. This possibility is a direct consequence of the system design. Both T-DAB and DVB-T employ a large number of individual high frequency carriers distributed across the spectrum range used by the systems to carry the information. This is the basis to compensate effectively for service degradations due to critical wave propagation conditions.

As just mentioned, the spectrum being at disposal for T-DAB or a DVB-T transmission is exploited to transmit several programs at the same time. Audio or video data together with additional information are combined to build a multiplex which with the help of modern modulation schemes is transmitted by the set of high frequency carriers. In the case of T-DAB six to eight audio programs in nearly CD quality per multiplex are the rule while for DVB-T two to four PAL equivalent programs can be broadcast within one multiplex. Clearly, if more programs than that are to be provided within a particular area additional frequency resources are needed in order to establish and broadcast additional multiplexes.

The most important consequence of the system design for frequency and network planning of T-DAB and DVB-T is the possibility to implement the systems in terms of single frequency networks (SFN). SFN configuration means that all transmitters constituting the network within a specified coverage area are employing the same spectrum range in order to broadcast the same content. Clearly, also with digital signals superposition of different signal contributions is encountered. This is just physics and is naturally not connected to the characteristics of the broadcast signals. However, even though there is definitely a negative impact for certain parts of the used spectrum range due to the superposition of several signals, it turns out that in average superposition will lead to better reception. It must be emphasized, nevertheless, that this statement is true only as long as the spread in the times of arrival of the signals at the receiver location is not too large. The details of this particular system feature will be in the focus of the discussion later in this book extensively.

Since the propagation of electromagnetic waves is governed by basic physical laws it does not naturally terminate at national or international borders. Therefore, the assignment of frequencies onto coverage areas is in principle an issue of international nature. It has to be dealt with in accordance with all affected foreign neighbors. There are several international organizations who organize and manage the usage of the entire electromagnetic spectrum. The most import body certainly is the International Telecommunications Union (ITU). For Europe the Conference Européenne des Administration des Postes et des Telecommunications (CEPT) plays a major role as well.

In 1995 a CEPT conference was held at Wiesbaden, Germany, in order to establish the access to spectrum for T-DAB and to set up a corresponding frequency plan. Both in the VHF range and in the L-band certain spectrum ranges have been allotted which gave the possibility to allow every country in Europe the implementation of two nation-wide coverages with T-DAB. It turned out that a vast number of different national and political constraints had to be taken into account explicitly. This led to a frequency plan which formally is full of incompatibilities. Fortunately, most of these problems could be solved by bi- or multilateral coordination. The reason for the incompatibilities lies in the very different spectrum usage across Europe. There were a lot of countries where the originally envisaged spectrum ranges could not be used because other telecommunication services were already occupying them. In view of these obstacles it is evident that the assembly of a detailed frequency plan was only possible by massive use of computer power and the exploitation of highly sophisticated mathematical methods.

The solution to the problem of finding a set of mutually compatible frequencies for well-defined coverage areas seems to be very easy and straightforward on a first glance. Indeed, a more closer look reveals the extremely high complexity of the underlying optimization problem. Already the simplest task, namely just to find a set of frequency assignments for some areas such that there are no compatibility problems and at the same time the use of spectrum is minimized falls in to the class of mathematical problems which are labelled "NP-hard". A typical feature of such types of problems is that if one would try to find the one and only optimal solution by checking every possible solution, this would lead to a computational effort which increases incredibly fast as the characteristic system size increases. For the case of an frequency assignment this means the computation time blows up faster than exponentially as a function of the number of treated areas. In the case of typical, realistic numbers of coverage areas to be included in the investigation, it proves that it is no longer technically feasible to find a solution by pure trail and error. This is true irrespective of any future performance boost of the then available computer technology.

What is left then as the only practical way forward is to change the search target from the one and only optimal solution to a nearly optimal or at least very good solution. During the last decades there has been tremendous progress providing adequate methods thereto. Nowadays there exists a vast number of different means to find satisfying solutions for frequency assignment problems in the sense just mentioned. A very elaborate list of references to relevant work can be found in [KZZ02]. It has to be noted that most of the articles listed there refer to mobile telecommunication and not to broadcasting. However, the principle ideas can be applied to the field of broadcasting as well and remain valid.

After the end of any frequency planning conference all participants hopefully are in possession of the rights to use particular frequencies within their defined coverage areas. This is the time when network providers can seriously start to think about the implementation of their networks. The question what type of coverage targets can be realized must be addressed. Typical examples would be wide area coverage or coverage according to the population density. Besides, the type of reception that is focused on determines the details of the network implementation as well. Very different efforts have to be made if fixed, portable or mobile reception is to be aimed at.

What ever the coverage intentions are, the characteristics of the T-DAB or DVB-T networks have to be specified in all details. This means answers to several important questions have to be found. How many transmitters are actually needed? At which locations should they be preferable be situated? What are the effective output powers that are necessary to reach a certain coverage goal? In order to be able to find an answer to these questions a model for the whole broadcasting network has to be developed. Clearly, wave propagation issues have to be included as well. A high power transmitter on a high mountain will certainly have a completely different impact on the coverage than a low power transmitter somewhere in the middle of a big city with large buildings or even skyscrapers nearby. The freedom to adjust the number of transmitters in the network and their individual characteristic properties seems to be unrestricted, at least in terms of a theoretical planning model. It is obvious that once again the planner is confronted with an optimization problem possessing an enormous number of potential solutions. And once again the task is to find a solution which compared to predefined requirements is (nearly) optimal.

The present book tries to give insight into the manifold of mathematical and practical problems of frequency assignment and network planning for digital terrestrial broadcasting systems. By employing several examples the principles and most important features will be visualized and explained. The book comes in three parts. In the first part both T-DAB and DVB-T will be presented. This is followed by those aspects which are relevant for frequency assignment problems and network planning. Topics like wave propagation and coverage prediction fall into that class. In part two of the book different typical and interesting problems connected to frequency management and assignment are discussed in detail. Finally, the last part is concerned about network planning of digital terrestrial broadcasting systems. Mathematical methods employed in the course of the investigation can be found in the annex.

Chapter 2

Digital Terrestrial Broadcasting Systems

The success story of radio and television was originally based on terrestrial transmission. At a suitable location a transmitter was set up which broadcast radio and TV signals using corresponding equipment and antennas. The listener or viewer was expected to maintain a certain reception effort which in the case of television was to implement an antenna on top of the roof of the house having a sufficiently high directivity. Most of the stationary radio receivers were fed by correspondingly adopted antenna systems.

At the end of the 20th century the supremacy of terrestrial broadcasting over other distribution paths was definitely lost. At least this is true for television. According to market studies in 2000 only 12 % - 15 % of German households still were relying on terrestrial broadcasting as their primary reception method for television. One of the reasons certainly is the lack of a sufficient high number of attractive programs that could be provided by terrestrial means. This, however, is directly linked to the scarcity of corresponding frequencies.

For radio the situation is different in several aspects. Terrestrial broadcasting still is the most important way to deliver audio programs to the listeners. Nevertheless, also in the case of radio an irreversible change is taking place. Since the introduction of CDs as carrier for music people started to get used to the high quality standards provided by that digital audio system. FM radio is not able to offer an equivalent service quality.

Another difference between radio and television is that the main reception environment for radio is not fixed reception but portable or mobile reception. FM radio has not been designed for mobile reception in a vehicle. Hence, more or less serious perturbations have to be accepted under these conditions.

Basically, FM radio and analogue television suffer common problems. As in the case of TV there is not enough spectrum to satisfy all demands public and private broadcasters would wish to comply. In recent years the former very strict rules applied for coordination of FM transmissions very gradually softened in order to find new frequencies at all. The current situation, however, is very tight. In most cases the benefit gained by bringing a new FM frequency on air in some area has to be paid by over proportionally high level of interference at other locations. The FM spectrum has been squeezed out to the maximum.

All these facts triggered the development of the new digital terrestrial broadcasting systems T-DAB and DVB-T. Primary objectives were a resource saving usage of radio frequencies, high transmission quality, a large enough data capacity to allow for a sufficient number of offered programs and in the case of radio the possibility for mobile reception even at high velocities. The target of the latest requirement were vehicles moving on highways and also high velocity trains.

T-DAB and DVB-T quite naturally exhibit differences since they are optimized with respect to the transmission of audio or video data, respectively. Nevertheless, they have several technical features in common. Both systems employ a multi-carrier technology and in both cases psycho-accoustic or psycho-visual insights are exploited to significantly reduce the amount of data to be transmitted. Furthermore, the transmission is protected against perturbation by the application of sophisticated error protection mechanisms.

In the sequel, both systems will be described very briefly. The focus will lie on those aspects of the systems which are essential in connection with questions arising in the context of frequency assignment problems and network planning. For a very detailed technical description it is referred to the standardization documents [ETS97a] and [ETS97b] as well as to the textbooks [Lau96] and [Rei01].

Besides T-DAB there exists another digital transmission system which is suitable for terrestrial radio broadcasting. It is called Digital Radio Mondiale (DRM) [ITU01b]. It has been designed to substitute the analogue systems in the short and medium wave regions. DRM is a multi-carrier system, too. But in contrast to T-DAB or DVB-T only a very small bandwidth is used which is nevertheless very well adapted to short and medium wave transmission purposes. The characteristics of the system which are relevant for planning tasks, however, are totally different from those of T-DAB and DVB-T. Therefore DRM will not be treated here.

Mainly in Japan a further digital terrestrial broadcasting system is being introduced. It is called Integrated Services Digital Broadcasting-Terrestrial (ISDB-T) and combines system characteristics of DVB-T and T-DAB (see e.g. [ITU02]). Due to the close technical resemblance to T-DAB/DVB-T the Japan system will not dealt with here, too.

2.1 COFDM Modulation

In contrast to cable or satellite based distribution of radio of television programs terrestrial transmission has to cope with the negative impacts of multi-path propagation conditions. Any transmitter in charge broadcasts the signals with a constant radiation power. Usually, the transmitting antenna gives rise to an angular radiation pattern which means that in some directions there will more power radiated than in others. On its way from the transmitter to the point of reception the broadcast signal can undergo reflection or diffraction caused by obstacles like mountains, hills or buildings. Thus, in general at the receiver site not only the directly impinging signal from the transmitter is received but also contributions caused by the mentioned physical effects. This additional signals arrive under different angles, with different amplitudes and at different times than the direct signal. Figure 2.1 sketches a typical transmission scenario in the case of mobile reception.

The linear superposition of the different individual signal contributions generates a characteristic interference pattern in the vicinity of the point of reception. Constructive interference leads to points where the resulting amplitude takes a maximum value whereas in other locations due to destructive interference the amplitude is minimal. The field pattern depends on all details of the impinging signals, i.e. the number of contributing waves, their individual amplitudes, their angle of incidence and

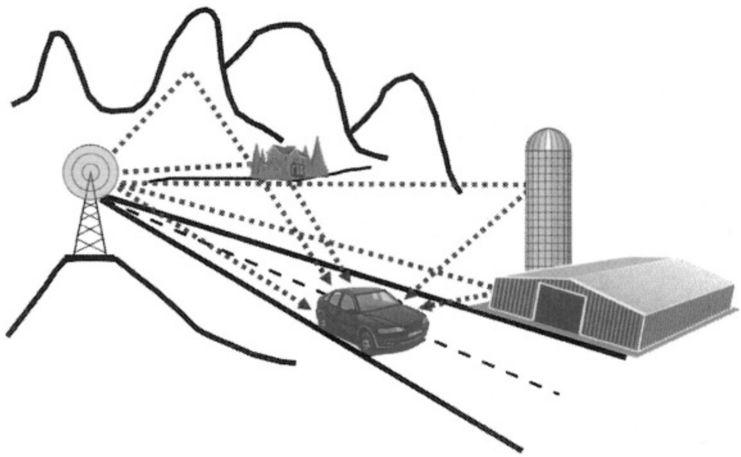

Figure 2.1: Typical multi-path reception scenario for terrestrial transmission.

last but not least the relative times of arrival. Clearly, another very import parameter is the frequency of the waves.

In case the reception quality for a particular transmission system is primarily determined by the field strength that is delivered at the point of reception it is possible to improve a deficient stationary reception quality simply by moving the receiver a distance in the order of the used wave length. A very well-known example for such a situation is mobile FM reception. A driver moving in a vehicle and listening to a dedicated FM program might experience a reduction of service quality when coming to stop for example in front of traffic lights. The position of the car might coincide with the location of a minimum of the field strength caused by the destructive interference of several signals. Moving the vehicle just about one meter might clarify the reception again. In the case of analogue television multi-path propagation typically leads to so-called ghost images. This corresponds to the appearance of laterally shifted weak pictures in addition to the proper TV picture.

In addition to the problems that arise from multi-path environments mobile reception imposes further critical requirements to be met. The receiver passes through the interference pattern of the electromagnetic waves. This is equivalent to a dynamic change of the usable field strength level. Furthermore, the motion gives rise to Doppler shifts of the arriving waves. Depending on the angle of incidence relative to the direction of motion the shifts are varying. As a consequence, however, the frequencies of the broadcast signal and the frequency the receiver is tuned to do not match anymore. The effect is frequency dependent in the sense that higher frequencies will suffer a larger Doppler shift. Also, increasing velocity gives rise to larger mismatch. Usually, broadcasting systems are able to cope Doppler shifts up to a critical value which depends on the characteristics of the system design.

Multi-path propagation without doubt affects the performance of broadcasting systems. The negative influence cannot not only be tracked in the time-domain but also when looking at the changes of the spectrum of the transmitted signals. Superposition of several waves leads to periodic cuts in the spectrum. The dimension of these degradations depends basically on the so-called delay spread Δt. This parameter describes the difference of the times of arrival between the first and the last incoming signal. In order to assess the degree of degradation it has to be compared to the bandwidth B the transmission system employs. The impact of multi-path environments is different for systems using a small or a large bandwidth. If B is kept fixed for the time being the region where

$$\Delta t >> \frac{1}{B} \tag{2.1}$$

holds will give rise to periodic reductions of the spectrum levels which are highly localized in comparison to the bandwidth and separated only by a small difference in frequency Δf. The smaller Δt becomes the more fractions of the spectrum occupied by the transmission system get affected by the perturbations until beyond a certain value of Δt the spectrum as a whole suffers a degradation.

Figure 2.2 is provided to give an graphic imagination of the impact of multi-path propagation on the spectrum of the received signal. Two cases for different values of the time spread Δt are shown. The pictures have been generated by using an arbitrary broadband signal $x(t)$ which has been manipulated according to the impulse response functions in upper part of figure 2.2 to derive a fictitious received signal $r(t)$, i.e.

$$r(t) = x(t) + x(t - \Delta t) \ . \tag{2.2}$$

Then, a simple FFT has been applied to calculate the spectrum of $r(t)$.

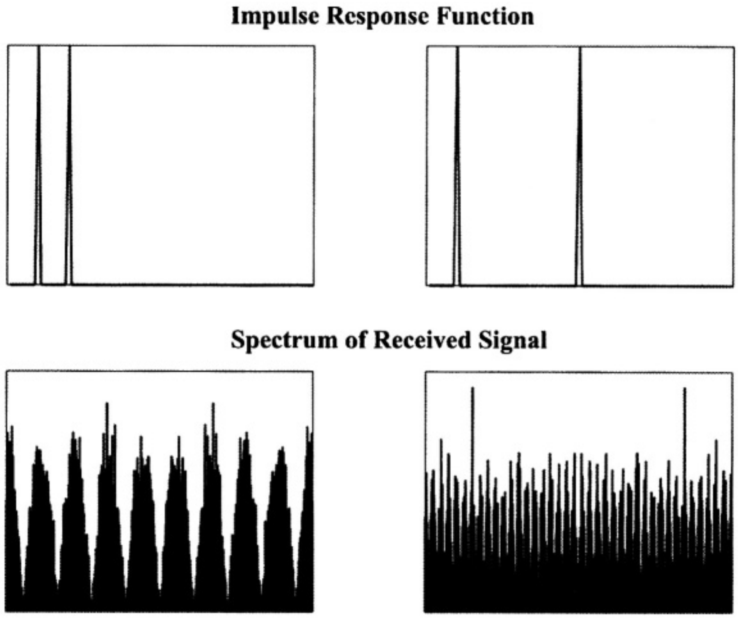

Figure 2.2: Impact of multi-path propagation on the spectrum.

If Δt is small, the as on the left part of figure 2.2, the spectrum is distorted throughout a larger frequency range. In the other case, the effect is not that pronounced. The distortion becomes more located.

The smaller the used bandwidth of the transmission system is, the more dramatic the consequences of multi-path transmission can be become. Total failure cannot not be excluded under some circumstances. In digital terrestrial broadcasting delay spreads in the order of 5×10^{-4} sec are quite common. So, in the case of two nearly equally strong signals every 2 kHz the spectrum shows the characteristic fading depicted in figure 2.2.

The modulation scheme "Coded Orthogonal Frequency Division Multiplexing " (COFDM) provides a solution to the problems caused by multi-path propagation effects (see e.g. [She95] or [Sto98]). COFDM is a multi-carrier system. In contrast to analogue transmission the temporal signal stream is subdivided into so-called symbols of duration T_S. Each symbol is the result of the superposition of a set of carriers whose amplitudes and phases can be fixed to certain allowed values only. After the time T_S has elapsed new amplitudes and phases might be chosen thus giving rise to another, different symbol which is then broadcast for a time T_S again.

The structure of COFDM already indicates that it is a modulation scheme which is applicable to digital signals only. There are, in fact, two levels of discretization involved. First, the input data stream is transmitted in terms of successive discrete symbols of duration T_S. Furthermore, the carriers can be adjusted only to defined discrete states. If M carriers are employed and each of them can take one of $2K$ allowed states then each symbol has a data capacity of $M * K$ bits.

But, when using such a technique how is it possible to overcome problems caused by multi-path environments? To understand this it is once again helpful to consider a simple scenario consisting of two signal contributions only. Depending on the relative temporal delay between the two signals the receiver will experience a more or less strong constructive or destructive interference at the point of reception. Figure 2.3 sketches some of the superposition possibilities.

In order to decode the transmitted information properly the receiver must temporally synchronize to the incoming stream of symbols. Then, a portion of length T_W of the received signal is sampled. By applying a Fast Fourier Transform the amplitudes and the phases are calculated. If the evaluation window T_W is adjusted to coincide completely with symbols of the first path, then nevertheless, it will inevitably take contributions from other symbols via the second path as well. In case, the delay between the two signals is larger than the symbol length T_S the second path gives rise to destructive interference only and hence a reduction of reception quality.

Interference which is caused by a retarded or advanced symbol falling into the evaluation window is called inter-symbol interference or selfinterference. In a situation where the time delay is smaller than the symbol duration only a fraction of an unwanted symbol falls into the evaluation window. In this case there is also a part of the wanted symbol arriving across the second path. Depending on the relative phases of the individual carriers of the two signals it might lead to constructive or destructive interference.

A straightforward way to overcome these problems is to introduce a symbol length which is larger than the utmost in practice appearing

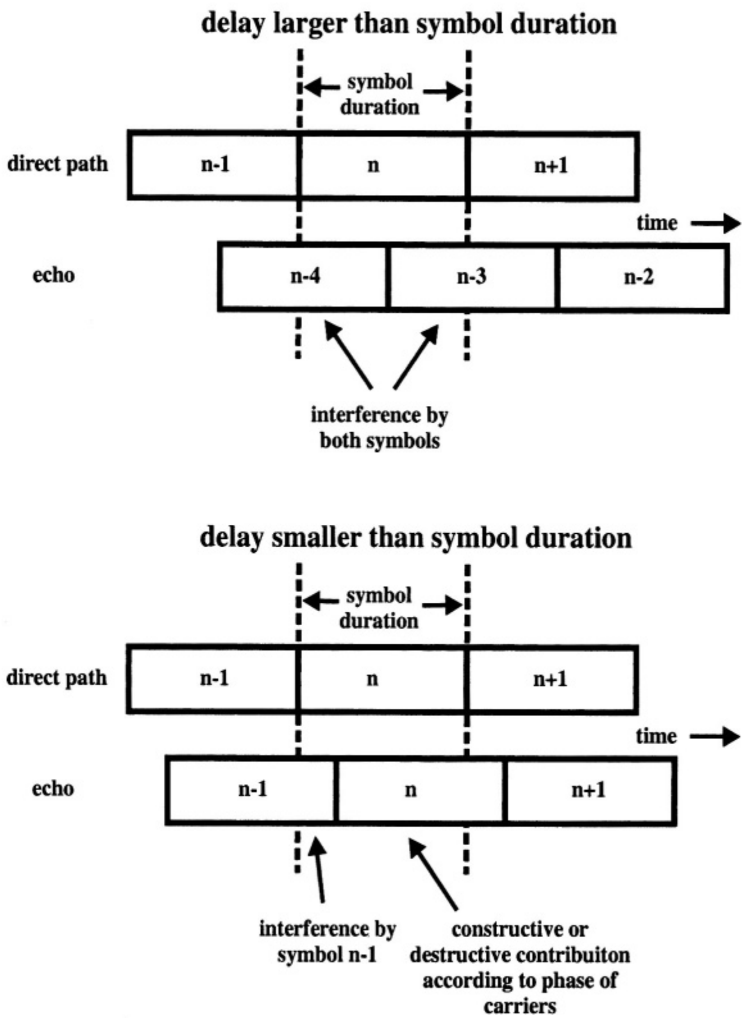

Figure 2.3: Inter-symbol interference caused by superposition of two signals which are delayed with respect to each other.

time delay. Thereby situations leading to symbols contributing as a total to selfinterference can be avoided. Increasing the symbol duration, unfortunately, reduces the data capacity which can be transmitted. In principle, this loss can be compensated for by employing a larger number of carriers. Of course, if the evaluation window T_W is equal to the symbol length T_S like in figure 2.3, then it is inevitable that contributions from different symbols will be present in the evaluation window. Hence, selfinterference would always be encountered.

The crucial move to avoid even these remaining problems is to use an evaluation window whose duration T_W is smaller than the symbol length T_S. The difference $T_G = T_S - T_W$ is called the guard interval. It is reasonable to choose a duration of the guard interval which should exceed the largest time delay to be confronted with in a practical situation as pointed out already above. But, concerning the data capacity the same constraints apply to a reasonable choice of the length of T_G. A typical value for T_G is for example $T_W/4$.

As already mentioned, the demodulation of the received symbols is accomplished by utilizing the Fast Fourier Transformation (FFT). A successful application of that technique presumes that the duration of the evaluation window T_W and the spectral separation between two carriers are mutually linked. Hence, it is reasonable to use carriers that are equally distributed across the bandwidth with a spectral distance of

$$\Delta f = \frac{1}{T_W} \tag{2.3}$$

separating them.

In mathematical terms a set of such defined carriers can be described as an orthogonal system. If the k-th carrier in baseband is written as

$$\Psi_k(t) = e^{ik\omega_W t} \tag{2.4}$$

with

$$\omega_W = \frac{2\pi}{T_W} \tag{2.5}$$

the orthogonality relation

$$\frac{1}{T_W} \int_t^{t+T_W} dt' \; \Psi_k(t') \Psi_l^*(t') = \delta_{k,l} \tag{2.6}$$

between two carriers holds. The asterix indicates complex conjugation while the parameter $\delta_{k,l}$ is the Kronecker symbol

$$\delta_{k,l} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases} . \tag{2.7}$$

Figure 2.4 outlines the position of the individual carriers. Usually an even number of carriers is employed. Therefore, the middle carrier f_C is in general not set.

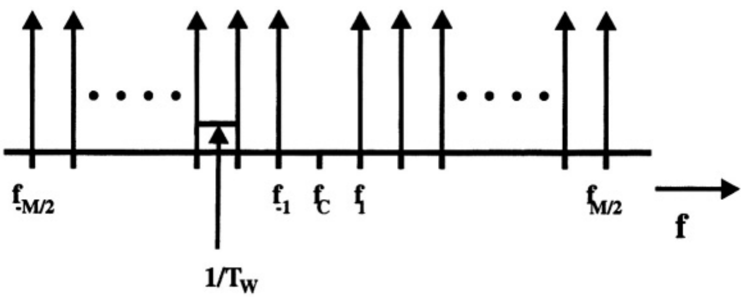

Figure 2.4: Schematic representation of the carriers of a COFDM signal consisting of M carriers.

For the demodulation the received signal is downshifted to baseband. Different kinds of filtering are applied as well. Then the signal is digitized on the basis of an appropriate sampling rate. From a set of samples corresponding to a time interval of T_W the Fourier coefficients are calculated by means of the FFT. The amplitudes and phases of the carriers contain the transmitted information. It has to be noted that the restriction of the evaluation to a finite time interval implies that the calculated spectrum corresponds to a convolution between the COFDM signal and the applied weighting function used for the FFT. The theoretical representation of figure 2.4 is replaced by a spectrum of the form given in figure 2.5. The underlying weighting function in that case is a simple rectangular window.

The spreading of the spectrum does not cause any problems as long as the receiver frequency tuning is correct. The proper choice of the COFDM parameters like width of the evaluation window T_W and the spectral carrier separation guarantee that at the location of the absolute

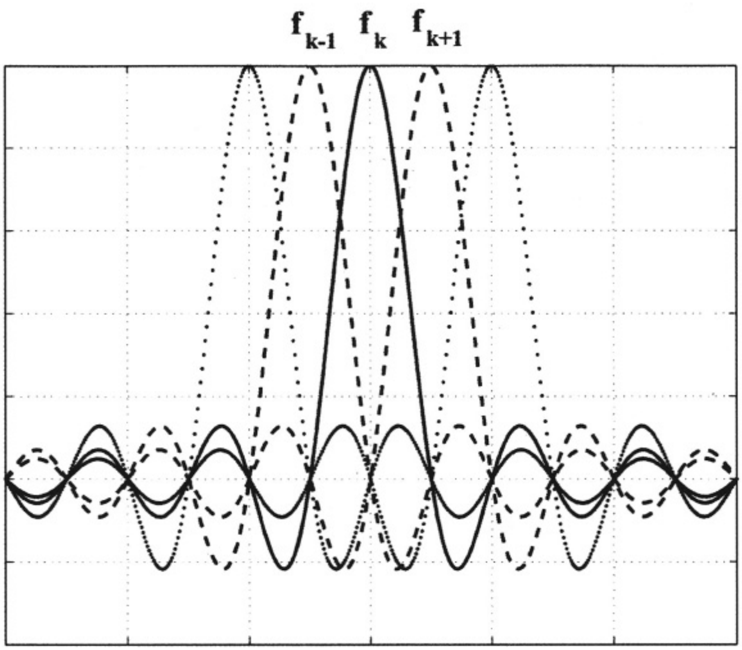

Figure 2.5: Spectrum of a COFDM signal convoluted with a rectangular window.

maximum of one particular carrier all other carriers have their nulls. The evaluation of the signals on the basis of FFT is thus not effected.

Clearly, this is no longer true if selfinterference comes into play. Then orthogonality is violated. Taken seriously, this holds for mobile reception, too. In that case Doppler shifts according to

$$\Delta f^{\text{Doppler}} = f\frac{v}{c} \tag{2.8}$$

appear. The magnitude of the shift depends on velocity component parallel to the direction of motion of the incoming wave and on the absolute value of the frequency. The quantity c represents the velocity of light. These Doppler shifts are the reason why the spectral location of the carriers and the position of the Fourier coefficient do no longer coincide. As has become clear by now, this is, however, a prerequisite for a successful demodulation.

In principle, COFDM systems are well suited to cope with perturbations caused by Doppler shifts during mobile reception. However, some care has to be taken when defining the COFDM parameters. A good system should work without problems as long as the Doppler shift does not exceed 5 % of the inter-carrier spacing, i.e.

$$\frac{\Delta f^{\text{Doppler}}}{\Delta f} < 0.05 \,. \tag{2.9}$$

With the help of relation (2.3) this can be cast into the form

$$\Delta f^{\text{Doppler}} \times T_W < 0.05 \,. \tag{2.10}$$

Equation (2.10) allows a rough estimate up to which velocity Doppler shifts can be compensated for.

The utilization of a large number of carriers and the introduction of a guard interval are the foundations of any COFDM system. Despite these features it cannot be avoided that depending on the transmission characteristics the decoding of the information is in general erroneous. If by destructive interference the amplitude of a certain carrier becomes too small then it is no longer possible to determine its phase correctly with a high probability. For other carriers within the same symbol the situation might be better at the same time. The bits derived from a carrier with a small amplitude thus tend to be wrong then. In case this affects bits which are important, for example because they carry information about the structure of the entire broadcast signal, the reception quality will degrade dramatically.

A very natural solution to this problem is to add some degree of redundancy to the original bit stream. Then, if a certain amount of bits cannot be decoded properly this will cause no harm to the reception. The receiver still gets all the information which is necessary to decode the signal completely. However, such a robustness of the system has to be paid for in terms of reduction of net data capacity.

Further protection mechanism can be applied as well. Time and frequency interleaving are very common approaches to make the transmission less susceptible to the impact of severe propagation conditions. Interleaving means to separate adjacent quantities from each other according to a defined scheme. The temporal sequence of bits can be reorganized as well as information on carriers can be mutually exchanged. The influence of localized perturbations both in the time and the frequency domain can be reduced thereby. More details concerning interleaving issues can be found for example in [Lau96] or [Sto98].

2.2 Digital Audio Broadcasting

The digital broadcasting system known as T-DAB has been drafted around 1990 in the framework of the European research and development program Eureka 147 [EUR96]. It has been standardized in 1997 [ETS97a] and rests on three basic pillars. The data rate of the source data is reduced with the help of the data compression method MPEG-3, layer II [ISO93]. Psycho-accoustic effects are exploited thereto. The second important issue is that the total data capacity of the system can be distributed between several independent data source streams. Both audio and other arbitrary data are allowed. Each source data stream can have its own data rate and can be protected individually. At the end, all data streams are combined to build a data multiplex. The final pillar is the application of the modulation scheme presented above, namely COFDM where a nominal bandwidth of 1.75 MHz is occupied by the T-DAB signals. Figure 2.6 gives an schematic overview of the T-DAB signal generation procedure.

A single audio program can be supplemented by more than one data service. The combination of audio program and data service(s) represents a program content. The data rate of the individual radio programs is reduced by employing the source coding mechanisms. As a next step each stream of information undergoes an appropriate channel coding scheme. In the case of T-DAB punctured convolutional codes are applied. This adds the required redundant information to the signal which provides the necessary basis for the protection against transmission losses. The error protection means include time interleaving patterns as well. If one audio program is to be supplemented by more than one data service the corresponding services are bundled into a proper data service multiplex. Channel coding is added, too (see e.g. [Lau96]).

The information how the T-DAB multiplex is structured is prepared independently from the program branches and is contained in the so-called Fast Information Channel (FIC). Together with program data the FIC is fed into the T-DAB multiplexer whose output goes to an COFDM modulator. Its task is to generate a corresponding baseband signal.

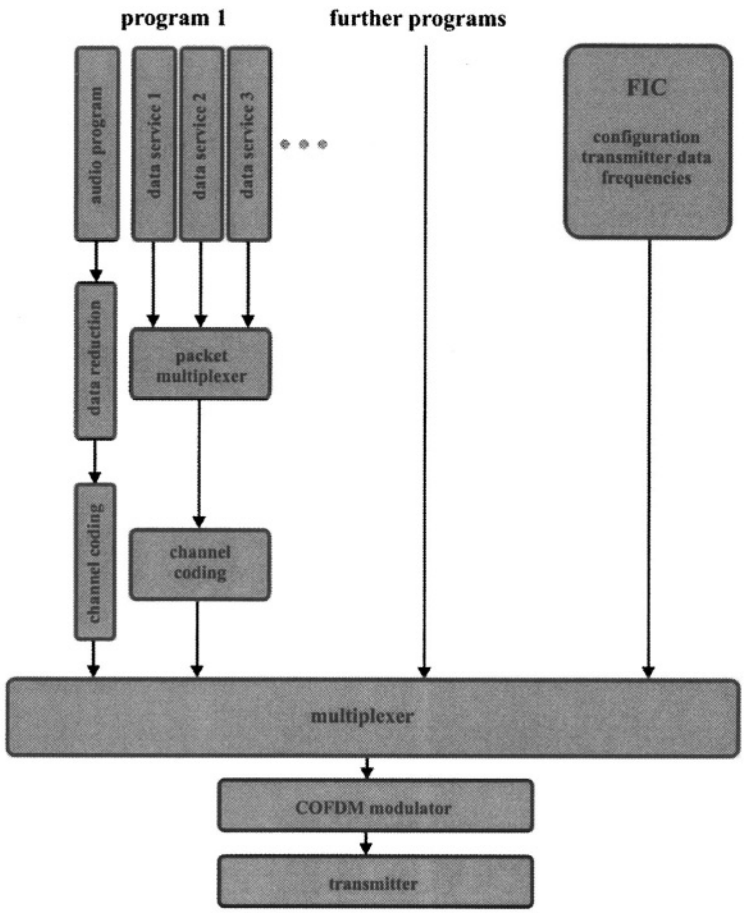

Figure 2.6: Procedure for the generation of a T-DAB multiplex. FIC stands for 'Fast Information Channel'.

This process rests primarily on the application of the FFT. Subsequently, a D/A conversion is carried out and the signal is shifted to the required radio frequency. Clearly, this generation is accompanied by appropriate filtering and amplifying.

The cycle to pass through during reception in order to detect and provide the requested audio signal or data service is sketched in figure 2.7. The T-DAB signal is received by the antenna and passes through an adequate process of filtering, downshifting to baseband and digitizing. The

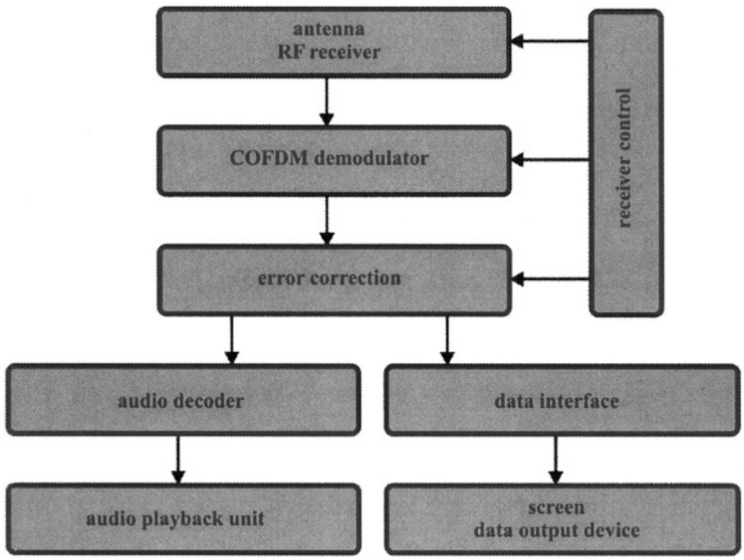

Figure 2.7: General demodulation procedure for the reception of a T-DAB signal.

COFDM demodulator retrieves the phases and amplitudes of the used carriers. This is accomplished by applying an inverse FFT. The FIC is analyzed which allows to restrict the data reconstruction on those parts of the entire bit stream that actually contain the requested audio program or data services. Errors that were caused by the transmission can be removed by making use of error correction algorithms unless the damage of the signal is not irreparable. Usually, soft-decision Viterbi methods are utilized for that purposes. The resulting bit stream is then passed to a playback unit or an adapted data output device.

Since T-DAB has been designed for mobile reception a very robust modulation scheme was mandatory. Therefore, the differential modulation DQPSK[1] has been chosen. The information that is to be transmitted is coded in terms of phase differences of identical carriers of successive T-DAB symbols. A defined number of symbols is grouped to build a so-called T-DAB frame which is preceded by the null symbol. This means simply that during the duration of a T-DAB symbol there is no power output of the transmitter at all. The null symbol is employed in

[1]DQPSK means **D**ifferential **Q**uadrature **P**hase **S**hift **K**eying.

order to establish a first rough synchronization of the receiver. The null symbol is followed by the phase reference symbol which within each T-DAB frame acts as a precisely defined starting point for the differential demodulation.

As already discussed in the previous section it is possible to adjust the COFDM parameters to adapt the broadcasting system to different coverage environments and targets. For T-DAB in total four different operation modes have been defined in [ETS97a] which can be employed under different conditions. Table 2.1 gives an outline of the four sets of allowed COFDM parameters.

	Mode I	**Mode II**	**Mode III**	**Mode IV**
number of carriers	1536	384	192	768
carriers spacing Δf (kHz)	1	4	8	2
symbol length T_S (μs)	1246.0	311.5	155.75	623.0
guard interval T_G (μs)	246.0	61.5	30.75	123.0

Table 2.1: The four possible operation modes for T-DAB.

Mode I assumes a transmission in the VHF range which in particular is suitable for the coverage of wider areas due to the corresponding wave propagation characteristics of VHF. In order to reduce degradations caused by violations of the guard intervals the network implementation should take into account an inter-transmitter distance that should not exceed 73 km significantly. This distance corresponds exactly to the lap electromagnetic waves can travel within the period of a guard interval of 246.0 μs.

Mode II has its focus on L-band application. According to the smaller guard interval the inter-transmitter distance in single frequency networks should be reduced as well. Therefore only smaller areas can be covered without problems. This goes hand in hand with the impaired wave propagation conditions in comparison with the VHF range. Nevertheless, for large city coverage this mode is very well suited.

The third mode has originally been defined for satellite broadcasting. In that case large Doppler shifts due to the high velocities of the satellites relative to the ground have to be accounted for. Therefore a carrier spacing of 8 kHz has been chosen. In contrast to terrestrial broadcasting echoes by reflections do not play a significant role for satellite distribution. Hence, the relatively small guard interval is justified. Finally, a fourth mode has been defined for coverage tasks which lie in between the wide area and the city only case.

2.3 Digital Video Broadcasting

Already with analogue terrestrial television a significantly larger technical effort was necessary than with sound broadcasting. This did not change in the digital age since the data capacities that are required to provide a satisfying service are larger by an order of magnitude. Furthermore, the acceptable error rates are smaller at the same time.

Terrestrial broadcasting has no longer the same importance for television as this is still the case for radio transmission. There certainly exist variations across Europe but the majority of people either uses cable or satellite links for television reception. In order to open the path for an successful introduction and market penetration of digital terrestrial television it is necessary that programs so far exclusively distributed via cable or satellite can be broadcast terrestrially without problems, too. Further design requirements were the possibility to transmit several different independent programs concurrently. The data rate for each of them should be adaptable to comply with predefined coverage targets. Clearly, the rates should be large enough to allow for an sufficient quality of service.

In Europe there are several channel rasters used in the spectrum ranges destined by the ITU for television broadcasting. Basically, the spectrum bands are subdivided into 8 MHz channels. This holds in particular for the UHF range. In VHF there are some countries like Germany for example which use a 7 MHz bandwidth. In other parts of the world there are also channel rasters based on a 6 MHz spacing. It is obvious that a new digital television system which is to be introduced around the globe must take these facts into consideration.

The former fixed antenna reception of audio and television programs is no longer a general coverage target. Portable and mobile reception have become more important in recent years due to a significant change of costumer's daily life circumstances and habits. It must be possible to carry the DVB-T receiver outside to be used during different spare time activities. This implies that DVB-T terrestrial reception has to be independent of a complicated roof antenna connection. Instead, a simple pole antenna must suffice.

All these requirements have been taken into account in [ETS97b] where Digital Video Broadcasting (DVB-T) has been standardized. The overall design approach is very similar to that of T-DAB. On the side of the transmission technology COFDM constitutes the nucleus of the system. It is possible to use one of the common bandwidths, namely 6, 7 or 8 MHz. The system has been designed initially for the 8 MHz case only. So, all DVB-T parameters mentioned explicitly somewhere in the text refer to that variant. In case another bandwidth is discussed then this is pointed out. Values for system parameters connected to a bandwidth of 6 MHz or 7 MHz can be derived from the 8 MHz values by a corresponding scaling of the underlying system clock by a factor 6/8 and 7/8, respectively.

Apart from the basic decision which bandwidth is to be utilized there are two fundamental system configurations that can be implemented. They differ by the number of carriers employed for the COFDM modulation. It is possible to use either 1705 or 6817 carriers. They are called 2k or 8k mode, respectively. Depending on the used bandwidth different durations of the evaluation window T_W and the carrier spacing Δf result. Table 2.2 summarizes the most important parameters.

For the two modes a total of six different guard intervals have been defined. The four values $T_G = 224$, 112, 56 and $28\mu s$ can be used in 8k mode while for the 2k mode $T_G = 56$, 28, 14 and $7\mu s$ are allowed. If two transmitters in DVB-T single frequency network are separated by more than a distance $\Delta r = c * T_G$ selfinterference can result. The quantity c denotes the velocity of light. As a consequence, the choice of the mode and the value of the guard interval is not a totally free one. It is to a large extend determined by the coverage target. For wide area coverage across in terms of an SFN structure there is, however, not much freedom left. The only promising DVB-T variant thereto should rest on the utilization of the 8k mode together with $T_G = 224\ \mu s$.

In contrast to T-DAB, it is possible to employ several different mod-

	2k mode			8k mode		
bandwidth of TV channel [MHz]	6	7	8	6	7	8
evaluation window T_W [μs]	298	256	224	1194	1024	896
carrier spacing Δf [Hz]	3348	3906	4464	837	977	1116
DVB-T bandwidth B [MHz]	5.72	6.66	7.612	5.71	6.66	7.609

Table 2.2: Potential COFDM parameters for DVB-T [ETS97b].

ulation schemes. Either QPSK, 16-QAM or 64-QAM can be applied[2]. The amount of data that can be transmitted increases along this ordering of modulations. On the other hand, the transmission becomes less rugged at the same time. In fact, from QPSK to 64-QAM an increasing protection ration between the useful and the unwanted signal contributions has to be taken into account. As a matter of fact, more elaborate network structures might be needed. Fortunately, DVB-T offers the possibility to adjust different error protection levels. This can be used to counterbalance the consequences of higher modulation schemes.

DVB-T does not use a differential modulation scheme. Therefore it is necessary to dedicate a fraction of the total data capacity for synchronization purposes. A subset of the total number is utilized as pilot carriers. They have precisely defined amplitudes and phases which are known to the receiver. There exist two types of pilots. The first type has fixed positions within the used bandwidth. Furthermore, there are pilots which change their position within the spectrum from one symbol to the next. The way they move is purely deterministic and also know to the receiver. This offers additional protection for the synchronization against degradation caused by narrow band fading as a consequence of multi-path propagation conditions.

The net data rate of DVB-T is independent of the chosen mode. Both 2k as well as 8k allow the transmission of the same amount of data per

[2] QPSK means **Q**uadrature **P**hase **S**hift **K**eying whereas QAM stands for **Q**uadrature **A**mplitude **M**odulation.

second. It is true, that the 8k mode employes four times more carriers than in the 2k case. But, at the same time the symbol length is four times as large for 8k variants, so that after all the data capacity remains the same. The crucial factors determining the data capacity are the modulation scheme applied, the error protection level, the duration of the guard interval and the bandwidth used. By variation of these parameters a huge variety of different operational system variants can be put into practice. Table 2.3 presents the most important possibilities. For a more profound discussion it is referred to [Rei01].

In the case of T-DAB as well as for DVB-T other planning parameters like minimum required field strength or protection ratio between wanted and unwanted signal contributions have to be taken into account as well. A full presentation of all these quantities would lie outside of the scope of this book. Very detailed tables containing all this information have been published both by EBU and ITU (see e.g. [ITU00], [EBU98], [EBU01], [EBU02]). They are constantly updated. Those parameters which are needed in the course of the investigations presented here are introduced and explained whenever it seems requisite.

		net bit rate (MBits / sec)			
modulation	error protection ratio	$\Delta/T_W =$ 1/4	$\Delta/T_W =$ 1/8	$\Delta/T_W =$ 1/16	$\Delta/T_W =$ 1/32
QPSK	1/2	4.98	5.53	5.85	6.03
QPSK	2/3	6.64	7.37	7.81	8.04
QPSK	3/4	7.46	8.29	8.78	9.05
QPSK	5/6	8.29	9.22	9.76	10.05
QPSK	7/8	8.71	9.68	10.25	10.56
16 QAM	1/2	9.95	11.06	11.71	12.06
16 QAM	2/3	13.27	14.75	15.61	16.09
16 QAM	3/4	14.93	16.59	17.56	18.10
16 QAM	5/6	16.59	18.43	19.52	20.11
16 QAM	7/8	17.42	19.35	20.49	21.11
64 QAM	1/2	14.93	16.59	17.56	18.10
64 QAM	2/3	19.91	22.12	23.42	24.13
64 QAM	3/4	22.39	24.88	26.35	27.14
64 QAM	5/6	24.88	27.65	29.27	30.16
64 QAM	7/8	26.13	29.03	30.74	31.67

Table 2.3: Net data rates for different DVB-T operation modes in the case of an 8 MHz TV channel [ETS97b].

Chapter 3

Coverage Prediction

The assignment of frequencies onto coverage areas is a prerequisite to offer broadcasting services. However, prior to getting involved in the struggle for the rights to use certain frequency resources at international planning conferences it is indispensable to have a very clear imagination which frequencies are favorable for the realization of different broadcasting intentions.

Once the frequency range for a particular purpose has been identified and the spectrum ranges are officially licensed the second step on the way to an operating digital terrestrial transmitter network can be addressed. The design and the implementation of transmitter networks require careful planning stages before the first transmitter is actually switched on. No network provider can afford to choose the locations of his transmitters by chance. The trial to simply bring a transmitter into operation and only then assess its effectiveness is prohibitive due to the enormous costs involved. Only by serious preparatory planning efforts a transmitter network can be successfully brought up nowadays.

Depending on the wave length very large areas can be covered in principle. High power AM transmitters using medium frequency waves can be received in most cases far beyond national boundaries. The data capacity that can be achieved with this kind of transmission is, however, not very large. This explains the very poor quality of AM radio. In the regime of short or very short wave lengths only regional or sometimes even only local areas can be served[1]. The advantage, in contrast, is that

[1]Actually, there is a difference between the terms "covered" and "served". Cover-

a large amount of data or TV pictures and the associated sound that can be transported.

It is evident that network planning cannot be tackled without a reliable prediction of the coverage that can be expected. Clearly, this requires to approximate the real world by reliable and funded models including the entire network structure and all associated aspects like wave propagation. It has to be emphasized that it is important to balance very carefully between the proximity of the models to reality and a reasonable effort when applying them. It does not help anything to create a perfect model that might give answers to any question with enormous precision if the time to get hold of this information tends to infinity or is impractically large due to unacceptably large computing times .

Every coverage prediction for a terrestrial transmission system, be it of analogue or digital nature, is based on two essential factors. These are field strength prediction and the incorporation of relevant system aspects. First, it must be possible to give information about the field strength that can be expected at a certain point of reception. Since a single frequency network does not consist of just one transmitter a set of very different transmitters must be handled concurrently. This step mainly relies on the application of wave propagation models of various types. Concepts must be developed how to combine the different signal contributions and to distinguish between wanted and unwanted signals.

Clearly, prediction of field strengths cannot be the final answer alone. Granted, without a high enough field strength level at the receiver site no terrestrial broadcasting system will ever work, but especially for digital transmission systems additional criteria have to be met to guarantee a satisfying reception quality. In principle, a link between the received field strength and the bit error ratio has to be established to be able to assess the coverage quality appropriately. This touches upon the incorporation of system features which quite naturally differ sometimes dramatically from system to system.

Furthermore, a coverage prediction for just one single point of reception has only very limited value. What network providers need to have is a

---

age area is usually understood as the area throughout which a service can be received physically whereas service area refers to an area in which a service is to be provided on the basis of media political defaults. Therefore in practice, a transmitter usually covers a larger area than what needs to be served. However, for the issues addressed in this book this distinction is not relevant and thus both terms are used synonymously.

statement about a particular area being served or not. In other words, it is necessary to provide a coverage prediction that is valid within a defined area. If it is possible to make a statement about the coverage status of one single point it is in principle no problem to simply increase the number of points throughout the considered area in order to collect the desired information. Such an approach remains, however, by definition a one-dimensional method since there is always a limitation in resolution. Mathematically, an infinite number of points would have to be taken into account to arrive at some conclusion valid in two dimensions. Clearly, this not feasible.

An area coverage prediction must perforce exploit means that are qualified to close the gaps between individual points for which a prediction is possible and the a whole area. As a consequence, statistical methods are employed. Pure deterministic wave propagation is meant to give a field strength level which is then assumed to represent the average field strength value within a small area of maybe 100m $\times$ 100 m. Different wave propagation conditions enter the statistical analysis as necessary since there is a significant difference between a reception scenario where direct line-of-sight connection between transmitter and receiver exists or not. The impact of multi-path environments on the coverage prediction must be appropriately reflected in the prediction method, too.

3.1 The Terrestrial Radio Channel

The primarily relevant frequency ranges for the terrestrial transmission of radio and television lie between 30 MHz and 1.5 GHz in Europe. Corresponding to such a large frequency range, the wave propagation conditions are varying heavily. On their path from the transmitter to the receiver the electromagnetic signals are affected by the topographical and morphographical nature of the earth's surface. At a first approximation, the wave propagation takes place quasi-optical. Surely, this assumption complies better for high frequencies than for the frequencies at about 100 MHz.

At the point of reception the resulting signal is very often generated by the superposition of several different contributions. If there is a line-of sight connection between the transmitter and the receiver location than the direct signal makes up the largest contribution. Since radio and television transmitters are usually set up at locations high above the sea level whereas the reception antenna is in general only a few meters above ground, so-called ground reflections from the immediate vicinity of the receiver are quite common. They make a large contribution as well. Furthermore, signals will be scattered or reflected by mountains, hills or buildings. This creates additional echoes arriving at the receiver with characteristic time delays. All this taken together is the already mentioned multi-path environment (see the discussion in chapter 2.1).

The details of the transmission characteristic mainly depend on the radio frequency of the carrier(s). The higher the frequency the smaller structures can in principle act as scattering centers for the electromagnetic waves. In the range of 1.5 GHz, where for example T-DAB is operated, the wave length is about only 20 cm. Each mast of a traffic light, each traffic sign or the metallic window frames in multistory buildings should in principle be included in the description of the transmission. It is quite obvious that this is not possible since this would require an enormous increase of data to be handled provided this detailed information is available at all.

To aggravate the situation, many potential scattering centers are not fixed. This includes cars but also in the case of indoor reception people moving around. As a matter of fact, the transmission channel is time variant by nature. The temporal fluctuations of the composition of the set of contributing waves at a particular point of reception and thus the resulting field strength follow stochastic or even chaotic dynamics.

Measurements of typical transmission scenarios fortunately show, however, that in practice in most cases the situation is quite stable. At a chosen point of reception usually only a couple of signal contributions significantly add to the resulting field. In first line, this comprises the direct signal and the thereby caused ground reflection. For analogue systems which typically use a smaller bandwidth than digital ones this sometimes creates severe problems because these two components are of comparable strength. Their relative time delay is very small and thus the degradations in the spectrum might be fatal (see chapter 2.1). As a consequence, the receiver no longer can demodulate the signals and retrieve the transmitted information without errors.

In addition to direct path and ground reflection there are also reflections caused by geographic obstacles. Nevertheless, the number of significantly contributing echoes is limited in practice. Figure 3.1 shows the result of one particular measurement taken in the country of Baden-Württemberg in Germany in the running T-DAB transmitter network. It is a T-DAB network which is operated in VHF using T-DAB block 12B.

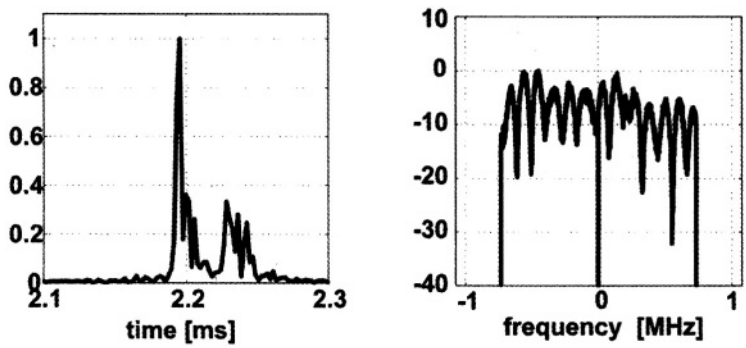

Figure 3.1: Measured impulse response (on left side) and associated transfer function (on right side) taken from the VHF T-DAB single frequency network in Baden-Württemberg, Germany.

At the location where the measurements were carried out two transmitters of the single frequency network could be received. Since the two transmitters were separated from the point of reception by different distances the corresponding signals had different times of arrivals. Indeed, there are two groups of signals to be seen on the left picture of figure 3.1 where the impulse response function of the radio channel is depicted. Both groups consist of a direct signal and a number of echoes which are caused by reflections somewhere on the way from the transmitter to the receiver. The right picture of figure 3.1 shows the corresponding transfer function which has been calculated from the impulse response function by applying the FFT.

What attracts the attention are the periodic fadings in the spectrum. They are a direct consequence of the presence of the two signal groups. If the measured data are used to determine the distance between two minima in the spectrum a value of $\Delta f \simeq 118$ kHz is found which according to the rules of the Fourier transformation leads to a temporal separation of two localized structures in the impulse response function of about 0.05 ms. This is approximately equal to the time difference between the two groups in the impulse response picture.

The description of a transmission channel by means of the phrases "impulse response" and "transfer function" implies a linear dependence between received signal $s(t)$ and the transmitted signal $\overline{s}(t)$. Nonlinearities that might be important for different propagation mechanism do not have any relevance here[2]. However, it has to be noted that linear not necessarily means proportional. Rather, the general link between transmitted and received signal is given as

$$s(t) = \int dt' \ R(t-t')\overline{s}(t') + n(t) \tag{3.1}$$

where $R(t)$ represents the impulse response function and $n(t)$ is an additive noise term. Equation (3.1) says that the received signal s at the time t depends on all values of the transmitted signal $\overline{s}$ for all times t'. Clearly, in reality the causality between transmitting and receiving must be kept. Only signals which have been broadcast before time t can have a physical influence on the received signal.

Impulse response $R(t)$ and transfer function $\Gamma(\omega)$ can be transformed into each other with the help of the Fourier transformation. In the spectral regime equation (3.1) thus is equivalent to

$$\xi(\omega) = \Gamma(\omega)\overline{\xi}(\omega) + \eta(\omega), \tag{3.2}$$

where $\eta(\omega)$ is the Fourier transform of the noise term $n(t)$. The quantities $\overline{\xi}(\omega)$ and $\xi(\omega)$ are the Fourier transforms of the transmitted and the received signal, respectively.

The impulse response function that can be seen in figure 3.1 is valid in that form only at one single point of reception. In the vicinity of this point the transmission characteristics might change dramatically. A very simple example will shed some light on this. If two equally powerful transmitters can be received under a relative angle of incidence of 30° between them a very typical interference pattern is generated by the superposition of the two signals within an area with a dimension of several wave lengths. Such a wave field is shown in figure 3.2.

[2]In fibre cable nonlinear propagation modes become more and more important. They can be excited by using very high power ultra-short lasers pulses. Then the nonlinear impact of the refraction index on the propagtion comes into play giving rise to nonlinear propagation modes called solitons.

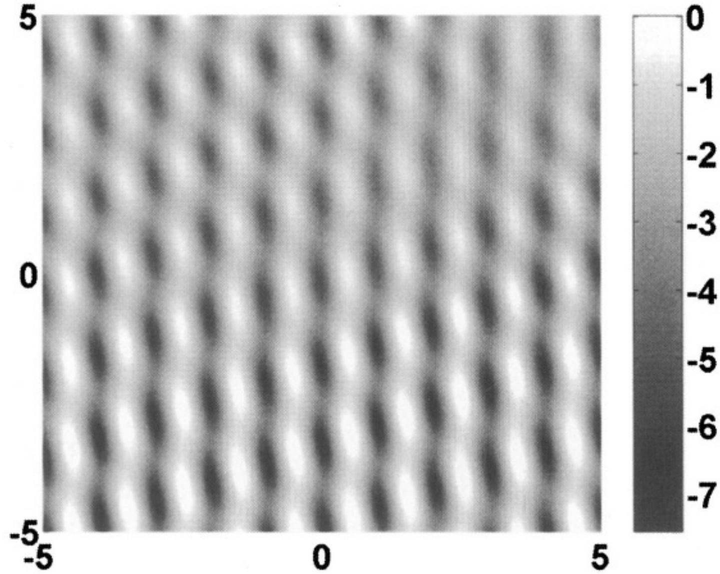

Figure 3.2: Wave field in the vicinity of a chosen point of reception which is assumed to be located in the center. The amplitudes are given in dB and normalized with respect to the maximum value of the field. The distances are multiples of the wave length.

Now, if the receiver starts to move from one point to another variations of the field strength level up to 20 dB are experienced. Sometimes at critical locations the changes can be in the order of 40 dB, too. This fluctuations of the field strength in an area of the order of several wave lengths are usually called ‘fast fading’. Its origin is of purely physical nature, namely the interference of a set of signals.

In reality the structure depicted in figure 3.2 is time variant, too. Even when remaining fixed at one particular location the wave field is changing in the course of time depending on how the composition of the set of arriving signals changes. There might be changes in the number of incoming waves, their amplitudes, their angles of incidence and their relative time delays. These changes cannot be described by deterministic means. Only in very few and very special situations this is possible and even reasonable.

In most cases the description of the terrestrial radio transmission channel is based on statistical approaches in order to model the spatial and temporal behavior of the transmission characteristics adequately. Against this background numerous channel models have been developed from which some crystallized as being more relevant for the planning of terrestrial transmission systems than others. The four most important models will be present here very briefly just to summarize their basic features. For a more profound discussion it is referred to standard text books of transmission technology like [Pro89] or [Kam92].

3.1.1 Gauss Channel

In the literature the Gauss channel is very often called 'AWGN channel'. The capital letters stand for '*A*dditive *W*hite *G*aussian *N*oise'. This refers to a transmission scenario where the signal at the point of reception is given as the superposition of the transmitted signal and a noise term according to

$$s(t) = \overline{s}(t) + n(t) \ . \tag{3.3}$$

The impulse response function consequently reads

$$R(t) = 1 \tag{3.4}$$

according to the general formulation (3.1). The quantity $n(t)$ is assumed to be Gaussian noise, i.e. $n(t)$ has to be considered as a statistical variable which follows a Gaussian probability distribution with zero mean and standard deviation σ_n. Hence $s(t)$ has to be treated as a statistical variable, too, with Gaussian distribution according to $n(t)$.

For terrestrial broadcasting the Gauss channel is only of subordinate value in practice. The fact that theoretical investigations again and again refer to the Gauss channel is connected to its mathematical simplicity. Only this very simple model allows the closed, analytical calculation of quantities like bit error ratios at tolerable costs. The advantage of closed form expressions is obvious. Systematic investigations of the impact of certain parameters on the transmission quality can be carried out easily with their help. In all other cases it is necessary to fall back on very time-consuming numerical simulations to come to reliable conclusions.

Another advantage of the simplicity of the Gauss channel has to be mentioned. More complex channel types normally require higher field strengths at the point of reception to guarantee a certain reception quality. If direct calculations on the basis of corresponding channel models cannot be afforded it might be possible, however, to give gross estimates for examples of the protection requirements of a complex transmission scenario by adding margin values to the results derived for the Gauss channel.

3.1.2 Stationary Multi-Path Channel

Every time not only one but several signals superimpose at a point of reception this constitutes a multi-path channel. In contrast to the analogue world, the wanted signal contributions in digital single frequency networks can come from different transmitters. The signals broadcast by each of the transmitters may undergo different propagation effects leading to a great variety of different signal contributions arriving at the point of reception.

The most simple way to illustrate such a scenario is by employing sinusoidal waves. The m-th transmitter is assumed to broadcast the signal

$$\overline{s}_m(t) = \exp\left[i\omega t + \varphi_{m0}\right] \tag{3.5}$$

where φ_{m0} is an arbitrary initial phase and ω the frequency. According to the channel characteristics each of the signals has to undergo, the received signal is to be described by the linear superposition

$$s(t) = \sum_m \sum_k A_{mk_m} \exp\left[i\omega\left(t - \tau_{mk_m}\right) + \varphi_{m0}\right] + n(t) \; . \tag{3.6}$$

Any noise effects are accounted for by adding the noise term $n(t)$. The parameter A_{mk_m} represents the attenuation factor the k_m-th path originated from transmitter m has suffered. Its time delay is given by τ_{mk_m}. Equation (3.6) can be cast into a form which is in accordance with relation (3.1). The impulse response function thus reads

$$R(t) = \sum_m \sum_k A_{mk_m} \delta\left(t - \tau_{mk_m}\right) \; . \tag{3.7}$$

In (3.7) $\delta\left(t\right)$ denotes the delta function. The associated transfer function is given by

$$\Gamma(\omega) = \sum_m \sum_k \widetilde{A}_{mk_m} \exp\left[-i\omega\tau_{mk_m}\right] + \eta(\omega) \; . \tag{3.8}$$

The constants $\widetilde{A}_{mk_m}$ and A_{mk_m} differ only by some numerical factors which depend on the employed definition of the Fourier transformation.

The radio channel defined in equation (3.7) is assumed to be not time-variant, i.e. all parameters involved do not show any time dependence. In practice, a Gaussian noise term $n(t)$ is used. Such types of multipath channels are typically applied for numerical simulations or also for measurements in the laboratory. In [Com88] four channel models for mobile telecommunication systems like GSM have been defined. These models were applied to investigations for T-DAB and DVB-T very often, too.

3.1.3 Rayleigh Channel

In both preceding sections is was presumed that the parameters of the radio channel are not time-variant. However, in practice this is the exceptional case. Normally the channel parameters vary in time. Both scattering as well as reflection of electromagnetic waves are affected by fluctuations. Furthermore, mobile reception as a typical reception situation naturally leads to time dependencies. Consequently, the definition of the impulse response (3.7) must be modified such to take into account time dependent channel parameters. This gives the representation

$$R(t) = \sum_m \sum_k A_{mk_m}(t)\ \delta(\ t - \tau_{mk_m}(t)\ )\ . \tag{3.9}$$

The time dependence of the the amplitudes A_{mk_m} and the time delays τ_{mk_m}, unfortunately, is not known. Furthermore, it is rather likely that also the number of signals is changing in time. The impulse response function thus might change its structure drastically.

Without any knowledge about the detailed dynamics of the channel parameters there remains no other possibility but to interpret these quantities as statistical variables. If the number of independent signal contributions which superimpose at the point of reception is sufficiently large then it is permissible to treat the resulting received signal as a Gaussian statistic variable. This approximation gets better the larger the number of contributing signals becomes. In the mathematical literature this is known as the central limit theorem.

The superposition of a large number of complex valued plane waves according to an expression like (3.6) therefore means that $s(t)$ is complex valued, too, i.e.

$$s(t) = s'(t) + i * s''(t) \ . \tag{3.10}$$

The Gaussian assumption hence implies that both the real part $s'(t)$ and the imaginary part $s''(t)$ have to be considered as uncorrelated statistic variables which both independently follow a corresponding Gauss probability distribution function. For the magnitude of $s(t)$,

$$|s(t)| = \sqrt{s'^2(t) + s''^2(t)} \ , \tag{3.11}$$

a so-called Rayleigh distribution can be derived. It can be cast into the form

$$p(x) = \begin{cases} \dfrac{2x}{\sigma^2} \exp\left[-\dfrac{x^2}{\sigma^2}\right] & , \qquad x \geq 0 \\ 0 & , \qquad \text{else} \ . \end{cases} \tag{3.12}$$

The parameter σ stands for the standard deviation while σ^2 gives a value for the received power. The mean value of the Rayleigh distribution is given by $\frac{\sigma}{2}\sqrt{\pi}$. The phase $\varphi(t)$ of the received signal is equally distributed across the interval $[-\pi, \pi]$.

Situations that can be satisfactorily characterized by a Rayleigh channel are encountered if there is no direct line-of-sight connection between transmitter and receiver. Furthermore, it is necessary that the direct path suffers so much attenuation by diffraction that it does not contribute significantly to the resulting received signal. The second group of signals which can be seen in the impulse response function in figure 3.1 could act as an example. However, it has to be noted that probably the assumption for the application of the Gauss channel is violated because there is only a very limited number of individual signals involved.

3.1.4 Rice Channel

Very often the direct signal does not suffer such an attenuation that it can be neglected. In most cases there is one strong contribution whose amplitude not even fluctuates much. The time of arrival also has a very constant value. In single frequency networks it might happen that several such contributions arising from different transmitters can be received. In addition to these more or less static signals there is large number of signals having pronounced smaller amplitudes. Under the assumption that both large peaks in the impulse response of figure 3.1 do not change significantly in time this would constitute a nice example for a Rice channel.

The power ratio between the direct and the scattered components is called Rice factor c. As in the case of the Rayleigh channel the magnitude of the received signal $|s(t)|$ is a statistical variable following the probability distribution function

$$p(x) = \begin{cases} \frac{2x}{\sigma^2} \exp\left[-\left(\frac{x^2}{\sigma^2} + c\right)\right] \; I_0\left(\frac{2x}{\sigma}\sqrt{c}\right) & , \quad x \geq 0 \\ 0 & , \quad \text{else} . \end{cases} \quad . \tag{3.13}$$

The quantity I_0 denotes the modified Bessel function of order zero. The function (3.13) is called Rice distribution.

The phase of the received signal is not equally distributed any more as in the case of the Rayleigh distribution. A shape of the distribution for the phase $\varphi(t)$ can be derived which at a first glance is quite similar to an ordinary Gauss function. The larger the Rice factor c becomes the more localized the phase distribution gets. This behavior comes as no surprise. An increasing Rice factor means that the echoes are getting smaller and smaller with respect to the amplitudes of the directly incoming signals. If the direct signals indeed have constant amplitude and phase than the impact of the additional contributions is vanishing and both the distribution for the magnitude and the phase of the received signal must converge to a pure delta function [Kam92].

3.2 Wave Propagation Models

The detailed knowledge about the physical characteristics of the terrestrial radio channel certainly is one essential information in order to assess the coverage situation within a particular area. Sure enough, this alone is not sufficient. The field strength to be expected at the point of reception has to be provided, too. To this end, partially very different wave propagation models are applied.

Seen from a physical perspective free space propagation, reflection, diffraction and scattering are the physical mechanisms of electromagnetic wave propagation that determine the terrestrial transmission of radio and television signals. The number of publications extensively dealing with this topic is nearly inexhaustible. As introductive literature the two text books [Hal96] and [Hes98] could be put forward. Both cover the problem of wave propagation in all details. However, this suggestion does not exclude other sources from being adequate as well.

The calculation of the field strength at the receiver location presumes sufficient knowledge about all relevant topographic and morphographic issues which might affect the propagation of the electromagnetic waves on their way from the transmitter to the receiver. Shorter wave lengths require an accordingly adapted higher effort when creating digital terrain data models and the corresponding morphographical data. Depending on the desired accuracy of the field strength calculations there are several alternatives applied in practice.

All methods have to struggle with the temporarily changing fields. In practice, the field strength level is fluctuating more or less heavily with respect to a mean value. As discussed in the previous sections, this fluctuations can only be described statistically. Hence, strictly speaking no deterministic field strength F is calculated. Rather, the result of any calculation has to be interpreted in terms of saying the field strength at the point of reception under consideration exceeds the value F in x % of the time. Typical values of x are 1 %, 10 % and 50 % (see e.g. [Hal96] or [Hes98]). Figure 3.3 demonstrates the interpretation of a calculated field strength value.

To avoid any misunderstandings, it has to be emphasized that the term field strength has a very distinct meaning in the context of wave propagation models. Physically, the impact of an electromagnetic wave at the location of the receiver is to deliver a certain power P which is linked to the amplitude E, i.e. the physical electromagnetic field strength of the wave according to $P \sim |E|^2$. However, it is common ground in the field of network planning to use the terms "field strength" or "field strength level" for the logarithm of the normalized power P of the wave, i.e.

$$F = 10 \log \frac{P}{P_0} = 10 \log \frac{E^2}{E_0^2} \tag{3.14}$$

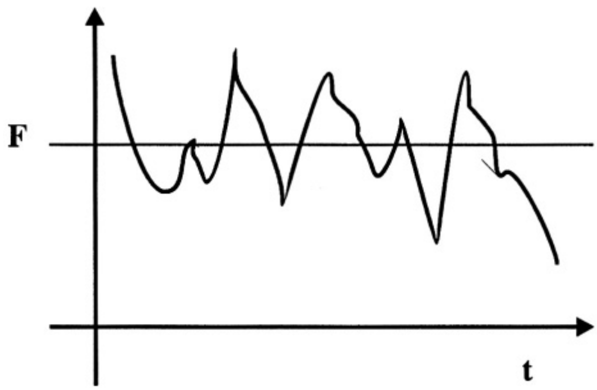

Figure 3.3: Interpretation of calculated field strength value F as the value that will be exceeded in x % of the time. In this case $x \approx 50$ %.

where E_0 corresponds to $1\ \mu\mathrm{V/m}$. All wave propagation models refer to the quantity F, i.e. they give field strengths in units of $\mathrm{dB}(\mu\mathrm{V/m})$.

3.2.1 Recommendation ITU-R P. 1546

If there is only little information about the topography of a region for which a field strength prediction is to be calculated, then the model of the ITU based on the recommendation ITU-R P.1546 [ITU01a] offers a way to do so. The recommendation 1546 replaced the formerly valid recommendation 370 which has been used for decades at the end of 2001. This wave propagation model is very common in the field of broadcasting because last not least all recent international frequency planning conferences were based on this model or some of its derivatives.

Topography is not taken into account explicitly. In contrast, the model is based on a vast number of field strength measurements taken over a period of many years. This information has been condensed into a number of propagation curves from which the field strength value at a chosen distance from a transmitter can be extracted. Clearly, the propagation curves are valid only for a predefined set of special transmitter characteristics, for example they assume a transmitter output power of 1 kW.

It has to be kept in mind that the calculated field strength values have to be interpreted in a very particular way. First of all, the values are to be understood as a value exceeded by the physical field a certain amount of time as described in the previous section. Second, due to the derivation of the curves from measurements it is very unlikely that a calculated value for a certain point of reception will be reproduced when making a measurement of the field strength there. The calculated value is valid only in a statistical way in the sense that if a lot of measurements would be carried out at the same distance from the considered transmitter then the mean of these measured values should coincide with the calculated figure.

On a first glance the method based on the ITU recommendation 1546 seems to be applicable only in very few special occasions. This is clearly not true since then the asset of the method would tend to zero for practical application. In principle, any transmitter and receiver site characteristics can be taken into consideration. This covers arbitrary output powers of the transmitter as well as for example different heights of the receiving antenna above ground.

In order to grasp the essential technical features both at the location of the transmitter and the receiver site which can influence the wave propagation, a set of parameters has to be provided for the application of the method. The most important parameters are the time percentage associated with the field strength level, the distance where the field has to be calculated, the height of the transmitting antenna and the frequency of the signal.

The field strength curves of the recommendation 1546 are given for the three time percentage values 1 %, 10 % and 50 %. With respect to the distance, 78 different values between 1 km and 1000 km are included. For the antenna height eight nominal values are taken into consideration reaching from 10 m to 1200 m. The frequency range covered by the recommendation reaches from 30 MHz up to 3000 MHz. However, only for 100 MHz, 600 MHz and 2000 MHz field strength curves are quoted.

Apart from these four essential parameters there is also a categorization of the general propagation type. Since the wave propagation conditions above land, cold sea and warm sea differ dramatically these three propagation regimes are distinguished. Cold sea refers to the Atlantic Ocean while warm sea is connected to the Mediterranean Sea.

All curves are provided both graphically and tabularly. The later form is important if the characteristics of a transmitter under consideration do not match with those values given in the recommendation. For that instance interpolation rules and formulas are given as well. In most cases it is necessary to make several interpolations. To this end, not only the interpolations but also rules about the sequence of their application are presented. Both the tables as well as the interpolation rules are innovations which have been introduced when migrating from recommendation 370 to 1546. The application of the 370 version very often led to problems during application in the past due to the lack of precise rules for interpolation.

The impact of topography is mainly covered by the height h_1 of the transmitter. Its definition is slightly broader in recommendation 1546 than in the former one. Depending on the distance in which the field strength is to be calculated different rules apply. In order to calculate the parameter h_1 first of all the height of the antenna center above ground h_a has to be known. Furthermore, the effective height h_{eff} of the transmitter with respect to the receiver location has to be known, too. This concept has been taken over from the former recommendation 370. It is defined as being the difference between h_a and the the average value of the ground level height calculated from an interval of 3 km and 15 km in the direction towards the receiver. Figure 3.4 visualizes the definition.

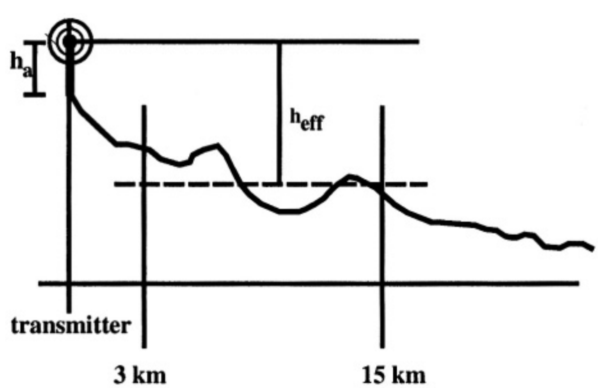

Figure 3.4: Definition of the effective height h_{eff} according to [ITU01a].

The parameter h_1 is defined differently depending on the distance between transmitter and receiver, namely

$$h_1 = \begin{cases} h_a & , \quad d < 3 \text{ km} \\ h_a + (h_{eff} - h_a)\,\dfrac{(d-3)}{12} & , \quad 3 \text{ km} \leq d < 15 \text{ km} \\ h_{eff} & , \quad 15 \text{ km} \leq d \end{cases} \quad . \tag{3.15}$$

The topographic and morphographic conditions in the vicinity of the receiver are captured by providing different heights R for the receiving antenna height h_2. In order to characterize an inner-city situation the value $R = 30$ m should be used, less populated cities are to be represented by $R = 20$ m and rural reception conditions are taken into account with the choice $R = 10$ m. If the actual receiving antenna height does not coincide with one of the reference values, i.e. if for example in a rural area an antenna height $h_2 = 20$ m must be used, then the field strength values have to be corrected following given procedures.

Since the real topography between transmitter and receiver is not taken into account explicitly, the calculated field strength values sometimes diverge strongly from the actually measured values. The invaluable asset of the ITU-1546 curves is that they represent some kind of widely accepted international minimal consensus. Nearly all frequency planning conferences in recent years would not have come to any result at all if the problems connected to wave propagation would not have been dealt with by an ITU recommendation or a modification thereof. The application of [ITU01a] creates a common basis for all partners participating in the international frequency planning business. When it comes to bilateral negotiations between countries, regions or even network providers, however, it is true that more refined models are employed to solve problems on a more detailed level.

3.2.2 Terrain Based Propagation Models

There exist numerous mathematical methods which provide the means for a field strength prediction within a particular area for both the VHF and the UHF frequency range. The work of [Lon68], [Oku68], [Cau82] and [Gro86] as well as the references found there offer a profound outline of the different approaches. In essence, all methods are based on the calculation of the diffraction attenuation caused by obstacles on the way from the transmitter to the receiver. Which geographical objects are included is determined by examining a corresponding ground level profile. An important features of all these approaches is that they only exploit obstacles which lie on a plane of the profile between transmitter and receiver. Hence, they are two-dimensional methods by definition.

The field strength at the point of reception is determined by reducing the free space propagation value by factors derived from the properties of the relevant topographic and morphographic structures. The models differ in the way these corrections are actually carried out and in which way morphographic data enter the process.

Seen from a physical perspective, diffraction means that the electromagnetic waves manage to get in the geometrical shadow zone of an obstacle. The basis for all calculations concerning diffraction effects is the artificial problem of diffraction at an infinitely extended half plane perpendicular to the direction of propagation. The analytical mathematical treatment of that configuration leads to solutions that can be expressed in terms of Fresnel integrals (see e.g. [Lue84]). In geometrical terms this means that in the plane of ground height level an geometrical region, the so-called Fresnel zone, is defined according to a precise construction rule. Any obstacle falling substantially into the Fresnel zone contributes to the attenuation.

In practice, transmitter and receiver are quite often separated by not only one but several obstacles. There exist various approaches to tackle that problem. In most cases, the obstacles are dealt with in terms of subsequently ordered half planes. An exact solution to that mathematical problem is feasible for a moderate number of diffraction edges, however, due to the high numerical effort involved they have virtually no meaning for practical applications.

A quite satisfying approximation is to treat the obstacles sequentially not according to their geographical proximity to the transmitter location but according to their impact on the diffraction attenuation [Mee83]. A general constraint for all methods trying to cope with multi-edge diffraction is that none of the attenuating humps between transmitter and receiver must be located near to the transmitter. Likewise, the distance between two obstacles should be larger than 2 km. The first case does not play an important role in broadcasting where typically the transmitter sites are chosen on high mountains or hills far enough away from any nearby heights.

Morphographic corrections also show up in very different ways across the various methods to determine the diffraction loss. Usually, morphographic data have an accuracy and resolution not as good as the topographic data. Therefore, very often additional global attenuations are added to compensate that loss of precision. On the other hand, there are models which combine the morphographic data with the digital terrain data in the sense that depending on the morphography the ground height level of a considered point is artificially raised. Typical values are 25 m for forest and 10 m for built-up areas.

Since there is no strict physical method for the calculation of the morphographic corrections enormous differences between the results of different approaches are encountered. This makes it very difficult to compare the methods. Sometimes it is even totally impossible.

3.2.3 3D-Models

The propagation of the electromagnetic waves from the transmitter to the receiver is three-dimensional phenomenon. In general, a multitude of signal contributions arrives at the point of reception from different directions (see figure 2.1). This multi-path environment might lead to a totally different field strength level as in the case with only the direct signal. The set of signals can superimpose constructively or destructively giving rise to either a higher or lower resulting field strength level.

By reducing the wave propagation model to a two-dimensional ground height profile calculated along the line between transmitter - receiver an error is introduced into the description of the wave propagation which cannot be compensated for by empirical margins anymore. Furthermore, the information how obstacles are oriented with respect to the ground height profile gets lost. It is obvious that there must be a difference if the diffraction edge is indeed perpendicular to the direction of propagation or not.

To overcome these shortcomings several 3D-models have been developed in the past. The publications [Gro95] and [Leb92] can be used as starting points to enter into that field. There are models which are entirely based on ray approximations. Both ray tracing and ray launching algorithms have been applied and adapted to individual needs. Heuristic semi-empirical methods are the alternative. They rest on matching a set of empirical factors by comparison with measurements.

In practice and in particular in the field of broadcasting where wide area predictions have high priority, there does not exist a 3D-model which achieved real relevance for everyday planning purposes. In first place, most of the models do not produce field strength values that are in line with measured values. Shortcomings in the model assumptions might be the reason for that. Another problem of some of the three-dimensional wave propagation models is connected to their implementation as computer programs. Still their time-consumption is not acceptable to allow the application on a daily basis. However, in view of the steadily increasing computer performance this might be a restriction to be overcome in the future.

3.3 Full Area Field Strength Prediction

Independent of the used wave propagation model any field strength prediction in first place makes a statement only for one single mathematical point. Since the radio channel in general is time variant this field strength value has to be interpreted as a limit which will be exceeded in a certain percentage of time. For wanted signals the percentage values range from 50 % to 99 % whereas for unwanted signals the percentages 1%, 5% or 10% are interesting (see e.g. [ITU95]).

Since as a matter of principle not for all points within a considered area a field strength value can be calculated it is quite common to use a regular grid of points. Their mutual separation is determined by the required accuracy and the availability of topographic and morphographic data with good enough resolution. Typical distances between grid points lie in the interval between 100 m and 400 m for broadcasting applications. At each of these points a field strength value is determined with the help of a properly chosen wave propagation model. As a consequence, each grid point is surrounded by an area of corresponding size, i.e. 100 m $\times$ 100 m or 400 m $\times$ 400 m. These small areas are called pixels.

It is well known that the field strength level significantly varies across the area of a pixel. To first approximation, the fluctuations of the level are made up by a slowly changing component which is superimposed by a fast changing component. The slow changes are connected to the variation of the angle of incidence of the incoming signals when moving within the pixel area from one location to another. On a scale in the order of the wave length fluctuations of the field strength level are found which cannot be described deterministically in full detail. Their physical origin is the interference of several individual signals. This has been discussed already in section 3.1. The slow changing component is usually called “slow fading” and the quickly changing phenomenon consequently “fast fading”.

In practice, both fluctuations are described in statistical terms. The statistics of the slowly changing field strength level, i.e. its variation across the pixel area, is governed by a Gauss distribution to a good approximation. According to relation (3.14) the power of the signals follows a log-normal probability distribution (see e.g. [Hal96]). The signal power connected to the quickly changing component follows to sufficient accuracy a Rayleigh distribution (see e.g. [Hal96] and the discussion in section 3.1.3). In practical applications, these dependencies are usually assumed and both mean and standard deviation are fixed on the basis of measurements.

Equipped with these assumptions the probability can be calculated that the field strength level exceeds a given value at a predefined percentage of locations within the pixel under consideration. To this end, the field strength value calculated by means of the wave propagation model is employed as the mean value for the distribution of the slow fading for the corresponding pixel. For the Rayleigh distribution of the fast fading component a zero mean value is always presumed. As standard deviations the values found in measurements are used.

3.4 Coverage Assessment

The calculation of the field strength inside a defined area is only the first step on the way to an profound assessment of the coverage situation. It means that the characteristic features of the transmitter have been treated. In first place this relates to the geographical location of the transmitter site, the broadcasting frequency and the output power. Furthermore, the consideration of the antenna diagram is essential. In general, it is not possible to adjust the output power of the transmitter freely. Indeed, both output power and antenna diagram have to be tuned to comply with constraints imposed by other adjacent transmitters. A further property of the transmitter site directly influencing the range of the wave propagation is the antenna height above ground.

Apart from the parameters characterizing the transmitter it is necessary to specify how large the reception effort a customer is expected to bear at the point of reception should be. It has to be defined what is the proper receiving antenna height. Is it 10 m which more or less would correspond to a roof antenna or is it 1.5 m above ground which typically is considered as being representative for mobile reception? Moreover, the selectivity and the sensitivity of the receiver taken into account as a common type of receiver during calculation must be specified.

From these data the minimum required field strength at the location of the receiver can be deduced. Minimum field strength is a synonym for the fact that the receiver needs a certain input power to be able to distinguish the transmitted information from the noise floor. Using a wave propagation model then allows to determine those point where the considered transmitter delivers a field strength level above the minimum level.

In practice, there are no isolated transmitters. Usually, there are other transmitters in operation using the same or adjacent radio frequencies. Depending on the distance from the transmitter under consideration these other transmitters might lead to more or less harmful perturbations. According to the employed type of transmission systems different protection ratios between the useful signal and the sum of all interfering contributions must be met. For digital terrestrial transmission systems the situation usually gets even more complicated due to the enormous number of different operation modes which all might demand for special individual protection ratios. The documents [CEP95], [CEP96] and [CEP97] offer a very detailed overview about the situation in connection with T-DAB and DVB-T.

If several different signals have to be considered at a certain point of reception at the same time the question arises how these signals can be combined to a resulting signal. Seen from a purely physical perspective, the solution to the problem is obvious. At the point of reception all contribution superimpose according to their amplitudes and phases. So, a full linear superposition of the electromagnetic waves is the result. It is quite evident that such an approach is not feasible on the basis of the available information provided by the wave propagation models at hand. Therefore other possibilities have to be utilized.

In order to derive a method for the combination of several signals it is helpful to reconsider what the wave propagation models actually provide and how this information is interpreted. Given a set of transmitters the result of the wave propagation models is a set of field strength values at the receiver site which are exceeded for a certain percentage of time. They are interpreted as being the mean values inside the area of a pixel. Thus, a deterministic combination of the quantities does not make any sense.

Indeed, a statistical treatment of the unification of the field strength values is mandatory. The problem of superimposing several electromagnetic waves is hence mapped to the problem to find the probability distribution function for the sum of two or more independent signals. The assumption about independence, however, is not quite correct in reality because different signals are sometimes completely correlated as in the case of a direct signal and a corresponding ground reflection. An adequate and settled treatment of correlation when deriving the distribution function of the sum of signals is still missing and under investigation.

As discussed in the previous section in practice it is assumed that the distribution functions for the field strengths are either log-normal or of Rayleigh type for single signals. The calculation of the distribution function for the sum is a time-consuming numerical process. Therefore solutions based on approximations have been developed.

Especially in single frequency networks it is very often not possible to carry out a clear distinction between wanted an unwanted signal. Sure enough, all signals arriving within the lap of a guard interval after the first used signal can contribute positively. On the contrary, signals arriving with a delay larger than a full symbol length will only produce negative interference. For those in between the situation is not that obvious a priori. Both wanted and unwanted impact might result (see the discussion in section 2.1 and figure 2.3).

When having distinguished wanted from unwanted contributions the resulting probability distribution functions for both sums have to be determined. Since both useful and interfering signals are dealt with in a statistical way this quite naturally carries over to their ratio. Once this is established all information is available to calculate the so-called coverage probability which allows to label a given pixel as served or not served.

The statement “a pixel is covered or served” can be given only in statistical terms, too. Two conditions have to be met thereto at the same time. First, it has to be checked if the minimum field strength exceeds the required value with a probability of p. Second, if the probability that the protection ratio exceeds the required value is larger than p as well, then the pixel is called covered or served. If the minimum field strength criterion is met but the protection ratio is not sufficient the pixel is classified as interfered. Finally, in case not even the minimum field is strong enough the pixel is simply not served.

It has to be emphasized that this is the traditional and for the most part employed definition. However, other definitions are conceivable as well. If for example it is implicitly assumed that high power transmitter networks are implemented in any case, the minimum field strength criterion could be omitted. Furthermore, instead of considering two independent probabilities a joint probability function might be considered more adequate.

As already indicated above, the crucial point, however, in the whole process is the statistical combination of several independent signal contribution. The most important methods at least for practical applications are discussed in some detail below.

3.4.1 Monte-Carlo Method

Apart from determining the joint probability density functions of several independent fields by means of numerical integration the well-known Monte-Carlo approach can be exploited. Sure enough, it is the most accurate of all methods based on some kind of approximation. If the mean values of all wanted and unwanted contributions as well as their standard deviations are known it is possible to calculate the shape of the sought after distribution function for the protection ratio between the total wanted and unwanted signal by simulation.

To this end, for every signal a set of random values is generated on the basis of its corresponding distribution function. The set should contain a significantly large number of entries. Then, a set of wanted and a set of unwanted signals are created simply by summing. From this a histogram of the protection ratio is derived which subsequently can be used to calculate the coverage probability numerically.

Such an investigation has to be carried out for each of the pixels of the considered coverage area. It is obvious that this requires an enormous numerical calculation effort to obtain a coverage assessment for a wide area. In practical applications this is prohibitive.

3.4.2 Power Sum Method

If only a very rough but nevertheless fast solution is asked for the so-called "power sum method" can be applied [EBU98]. The underlying idea is very radical in the sense that the statistical nature of the wanted and unwanted signals is neglected completely. Useful and interfering components are only summed in terms of powers and then their ratio is determined.

The link between the mean value of the field strength level $\overline{F}$ and the corresponding power $\overline{P}$ reads

$$\overline{F} = 10\ dB(\mu V/m) \cdot \log_{10}(\frac{\overline{P}}{P_0}) \tag{3.16}$$

with

$$P_0 = \frac{E_0^2}{Z_0} \quad , \quad E_0 = 1\ \mu V/m\ . \tag{3.17}$$

Summing all powers according to

$$\overline{P}_S = \sum_n \overline{P}_n \tag{3.18}$$

leads to the desired mean power sum $\overline{P}_S$ which by applying the inverse relation of (3.17) can be mapped to a corresponding field strength again.

Such a summation is carried out both for the useful and the interfering parts independently. The situation that both wanted and unwanted sums are equal is interpreted as a coverage probability of 50 %. In this particular case the results obtained by application of the power sum method are equal to first approximation to those of typical statistical methods. Other ratios lead to different coverage probability whose deviation from other approaches, however, increases and thus are no longer reliable.

3.4.3 Log-Normal Method

The log-normal method (LNM) provides a mean to calculate an approximation to the probability distribution function of a sum of log-normally distributed, not correlated signals. It is based on the assumption that the distribution function of the sum of two log-normally distributed signals is also a log-normal distribution. In practice, two implementations of this method won recognition. Only a brief summary of the methods will be given here. A detailed description of both can be found in the annex of [EBU98].

To clarify the difference between the two types of LNMs it must be recalled that the link between the field strength and the power (3.17) establishes also a link for the associated probability distribution functions of both quantities. It is widely accepted that the statistics of $\overline{F}$ is governed by a log-normal distribution. Relation (3.17) forces $\overline{P}$ to be normally distributed.

In principle this gives two possibilities to combine several signals statistically. With the help of (3.17) all individual signal field strengths are mapped to powers. The statistics of the sum of powers is straightforward due to the fact that the individual $\overline{P}$s follow a Gauss distribution, respectively. The only problem is to map back the normal distribution of the power sum to the corresponding log-normal distribution of the sum of field strengths. This process has to be carried out only once and thus can be easily accomplished by numerical means [Bru92].

The other possibility to statistically combine the set of signals is to operate in the regime of log-normal distributions. Even if the joint probability distribution function of two log-normal distributions is log-normal, too, it is in contrast to the Gaussian case not possible to calculate the mean value and the standard deviation of the resulting log-normal density function by analytical means. Instead, rather complicated numerical calculations have to be carried out. For an application to real world problems the corresponding time consumption is quite large. Some years ago this created real problems with respect to the existing computer performance. Meanwhile, it is no longer a problem. So an exact numerical solution should be always feasible. If performance is still an issue the exact numerical treatment can be substituted by employing an approximation based on bilinear interpolation [Phi95].

The latter improvement has facilitated the application of LNM methods to a large extend. Usually, the first method was utilized for extensive investigations due to the advantages with respect to computing time. However, it has been evident all the time that the method lacks accuracy. So, when compared to full blown Monte-Carlo simulations the results obtained by Gauss distribution based LNM were not very satisfying. This drawback does not exist with the second approach.

3.4.4 Other Methods

All methods presented so far are based on the assumption that by calculating field strengths and combining them appropriately it is possible to arrive at some conclusion about the coverage probability. However, strictly speaking already for analogue transmission systems where all these methods described above have their origin this idea is a very strong simplification of reality. For digital systems like T-DAB and DVB-T this is even more true. The ratio of the wanted and the unwanted signal levels is certainly an important criterion since without sufficient field strength at the point of reception no receiver will be able to demodulate the transmitted information. Unfortunately, the inverse conclusion that enough field strength will cure any reception problems is only conditionally valid.

The adequate quantity which allows to assess the reception quality in most cases is the bit error ratio. T-DAB and DVB-T are equipped with very efficient error protection mechanisms as described in sections 2.2 and 2.3. This means that also in areas where wave propagation models predict only a very low field strength an immaculate reception quality can be found. On the other hand, multi-path effects in particular when guard interval violations are encountered can lead to completely failed reception.

The correct way to derive a criterion which would allow the assessment of the transmission would try to establish a link between the radio channel characteristics and the bit error ratio. First steps in that direction have been made (see e.g. [Beu98a] and [Kue98]). A full description of the situation in all aspects is still to be expected.

Chapter 4

Management of the Electromagnetic Spectrum

Telecommunication is omnipresent today. Nearly every household in the western world owns several radios and most of the families have available at least one sometimes even more television sets. In some countries for examples in Scandinavia, already today (2003) the number of customers having exclusively a mobile phone contract exceeds the number of cable based subscriptions. Further wireless telecommunication systems are pushing into the market trying to gain ground and customers.

All these systems are utilizing a part of the electromagnetic spectrum. In order to guarantee interference free coexistence several international organizations are monitoring and controlling the spectrum usage. In first place, this task is taken over by the International Telecommunications Union (ITU) and with focus on Europe by the Conference Européenne des Administration des Postes et des Télécommunications (CEPT).

4.1 Spectrum Ranges for T-DAB and DVB-T

Currently, three spectral ranges have been identified for the usage by T-DAB and DVB-T. These are the VHF band III, i.e. the spectrum range from 174 MHz up to 230 MHz, the UHF range from 470 MHz to 890 MHz and parts of the L-band, namely the interval between 1452 MHz and 1479.5 MHz. At the moment, it is possible to operate T-DAB only in band III and the L-band while the primary frequency range for DVB-T is the band IV and V in UHF. However, DVB-T can also be implemented in VHF in band III. Therefore this latter spectrum range has to be shared by the two services T-DAB and DVB-T. Clearly, this enforces to establish special means to guarantee their spectral coexistence.

Radio and television services are not allowed to use an arbitrary frequency freely chosen from one of the mentioned spectral ranges. Throughout all bands there exist distinct frequency rasters which are to be respected both on a national and an international level. This means the entire spectrum bands are subdivided into channels of a particular width. In the UHF range there is a common 8 MHz channel raster across Europe in all countries. Thus the whole spectrum from 470 MHz up to 862 MHz allows to accommodate 49 channels who can be addressed by the corresponding channel number starting from 21 and ending with 69. Any TV transmitter must use one these channels.

According to the utilization by T-DAB the L-band is subdivided into intervals which are compatible with spectrum requirements of T-DAB. In order to broadcast a T-DAB multiplex a bandwidth of 1.75 MHz is occupied. These intervals are called T-DAB blocks. The nomenclature for these frequency blocks is LX where X is one of the characters from A to P.

The most difficult situation is encountered in the VHF range. There are quite a number of countries in Europe which decided to subdivide band III in terms of 7 MHz channels. In other countries a channel bandwidth of 8 MHz is used. In the first case eight TV channels result while the 8 MHz raster allows only seven channels. Also in France 8 MHz are used but the ordering of the channels is different. Thereby only six channels are disposable. A quite similar situation is found in Italy which utilizes an individual 7 MHz scheme resulting in seven usable TV channels. In the lower part of band III there are gaps between the channels which is a unique feature of the Italian raster compared to all other alignments. Figure 4.1 provides an overview.

The spectrum rasters in VHF are fairly well adapted to the individual needs of each country. With the introduction of T-DAB, however, the problem arose how to accommodate it spectrally in a television only frequency environment. It was decided to subdivide the entire band III continuously into T-DAB blocks starting at 174 MHz. Even though this

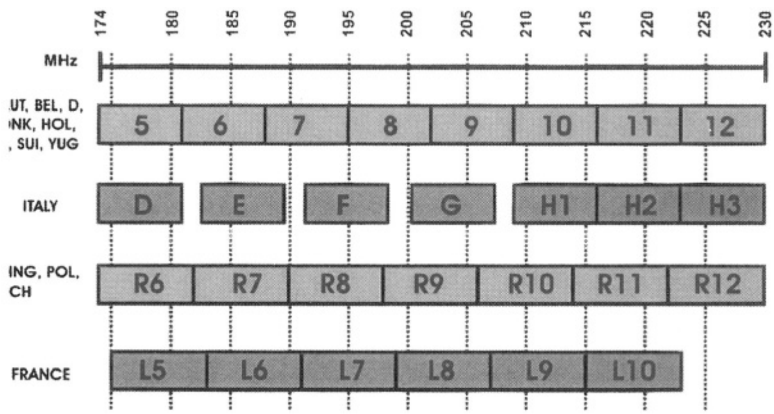

Figure 4.1: Some utilized spectrums rasters in the VHF range in Europe.

sounds very reasonable and straightforward it creates nevertheless some problems. While with respect to a 7 MHz raster everything is fine this is not the case with an 8 MHz scheme or one of the non-standard channel distributions. The consequence to bear is an spectral overlap between TV channels and T-DAB blocks. Figure 4.2 sketches the relative positions of the T-DAB blocks with respect to both a standard 7 MHz and an 8 MHz raster.

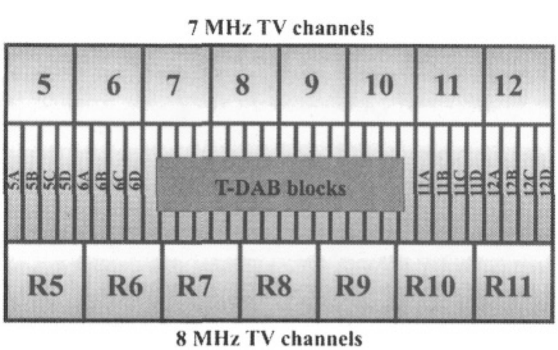

Figure 4.2: Relative position of the T-DAB blocks in VHF with respect to a 7 MHz (5-12) and a 8 MHz (R5-R11) channel raster.

In contrast to the nomenclature of the L-band in VHF the T-DAB blocks are identified by giving the channel number and a character A to D. A specifies the block having the smallest center frequency whereas D is the block with largest center frequency.

4.2 Frequency Planning and Radio Conferences

The details about the utilization of frequencies are formulated at international radio frequency planning conferences. The countries that sign a conference agreement oblige themselves to respect the rules for the utilization of the frequency assignments and procedures to modify entries in the agreed frequency plan. Each country is free to use or not use the frequency resources assigned to it and notified within the plan. Any implementation and operation of broadcasting infrastructure by employing any part of the electromagnetic spectrum, however, has to be internationally coordinated as soon as any other country or network provider affected by such activities. In any case, the neighboring state of a country have to be informed according to defined procedures.

There is an international stipulation that an administrative body like the ITU which is a branch of the United Nations should monitor and control the adherence to the frequency plan. Furthermore, it has to be logged who is operating which service in which area. Any changes of the utilization of the electromagnetic spectrum have to be notified to the ITU once all necessary coordination activities have been passed through. The ITU finally checks the compatibility of the notified frequency usage with the rules given in the corresponding agreements.

Even if it might suggest, the task of the ITU concerning broadcasting is far from being static or routine. New telecommunication systems are being developed and lay claim to appropriate electromagnetic spectrum. Thus, the actually existing frequency plans need to be revised from time to time. Sometimes, however, it is even necessary to reorganize the spectrum usage in parts of the electromagnetic spectrum from scratch. Then, it is the ITU to prepare and conduct the corresponding frequency planning conferences.

One of the obvious principles for the generation of any frequency plan is to avoid interference between services as much as possible. To this end, transparent and reliable methods to assess the interference have to be agreed upon and respected by all interested parties. On the level of plan generation it does not make any sense if during a conflict one country complains about interference produced by a neighbor and justifies its complains by results of some analysis carried out by methods the other side does not know of or does not accept as being appropriate. Even if refering to the procedures given in the agreed plan is the general rule, on a bilateral level, however, countries can agree to other procedures if necessary to solve a common problem.

Moreover, the possibility to adapt frequencies plans to future developments is another crucial factor which has to be taken into account when setting up a frequency plan. The technological development especially in the sector of telecommunications can not be foreseen with high reliability over extended periods of time. If the plan does not leave enough freedom to incorporate such progress then it is not compliant with today's demands. It must be possible to modify the plan to some extend and the procedures to do so must be defined.

For T-DAB and DVB-T there exist three international agreements which rule the frequency usage of these digital terrestrial systems so far. These are the final acts of the ITU conference of Stockholm 1961 [ITU61] and the agreements of the CEPT meetings in Wiesbaden 1995 [CEP95] and in Chester 1997 [CEP97].

All of these conferences had a different scope and validity. The Stockholm conference had to deal with the entire UHF range in the period from 26. May until 22. June 1961 which at that time was not already used by any legitimized telecommunication services. So, the huge spectral range from 470 MHz up to 960 MHz was at nearly free disposal for the planners. Altogether there were about 5300 requirements for analogue terrestrial television to be dealt with. Even though from a planners point of view the situation was ideal the political problems shaking Europe in that time were reflected by the accomplishment and the results of the Stockholm conference, too. In view of these problems it is amazing how well the plan worked out for analogue television during the following forty years.

The planning principle applied in Stockholm was to distribute the available TV channels according to a deterministic scheme. To this end, a nonuniform grid consisting of quadrilaterals of different size and shape was laid across a map of Europe. The dimensions of the quadrilaterals was chosen in accordance with the transmitter density covered by the quadrilaterals, i.e. the higher the density the smaller the area of the cell. However, the vertices of the quadrilaterals had to be separated far enough to allow the usage of the same channel there. Furthermore, it was assumed that all transmitters are identical having the same power, the same antenna heights and an omnidirectional antenna pattern. In addition, radio wave propagation should not depend on neither frequency nor direction of propagation. Then, frequencies were assigned to locations within the quadrilaterals according to a deterministic algorithm. The planning approach of Stockholm lead to a very efficient frequency plan in terms of efficient usage of channels.

Nonetheless, the plan had some inherent drawbacks. First of all, the transmitter locations used during the planning process did not coincide with the transmitter sites in reality. Therefore, coordination procedures had to be followed to bring a transmitter into operation because very often the assumed necessary separation between transmitters using the same channel was violated. Moreover, due to the political tension in 1961 and the existence of two political blocks different channel distribution schemes were applied in the western and the eastern world giving rise to additional problems that had to be solved in very lengthy and tedious negotiations (see e.g. [ERC98]). This burden could not be cured until today which is one of the reasons for the ITU planning conference scheduled for 2004 - 2006.

Any change of the plan after the conference had to be coordinated with all countries that were touched upon the consequences. Nevertheless, in Europe the Stockholm plan paved the way to allow each country the implementation of at least three nation-wide coverages with analogue television services. Regionally or locally it was possible to provide a lot more programmes in particular throughout the population centers. The crucial point about the Stockholm plan, however, is the fact that the utilization of frequencies is managed and controlled down to each single transmitter site. Each participant of the conference took home precise statements and specifications for the implementation of his analogue transmitter network.

This was completely different at the CEPT conference at Wiesbaden, Germany, in 1995 which was held to provide spectrum for the introduction of T-DAB. The target of Wiesbaden 1995 was to set up the basic conditions in Europe to allow every country the introduction of national nation-wide coverages with T-DAB. A total of 756 requirements had to be dealt with. In contrast to the Stockholm conference however, there was not enough spectrum available at that time. In band III the channels 11 and 12 were decided to build the core spectrum range which should accommodate T-DAB in the future. It turned out, however, that this amount of spectrum was just enough to establish one single coverage with T-DAB across Europe. But in order to do so, this implied already to accept a considerable level of interference in particular regions. Furthermore, there were a lot of countries where the utilization of the channels 11 and 12 by T-DAB was blocked since these channels were already occupied by other telecommunication services. This lead to the usage of T-DAB blocks located in other VHF channels, too.

The assignment of frequencies in the VHF range allows the transmission of one T-DAB multiplex across a wide area for example a total German Bundesland. The second coverage for which at Wiesbaden spectrum has been assigned is based on T-DAB blocks in the range of about 1.5 GHz in the L-band. Since the wave propagation conditions lead to significant reduction of range it is only possible to cover regional or local areas without problems.

Apart from the scarcity of spectrum the Wiesbaden meeting had to cope with, there is a fundamental difference in terms of planning principles between Stockholm and Wiesbaden. In contrast to Stockholm where all details of all transmitters have been specified, at Wiesbaden all frequency requirements were given only in the form of geographical areas throughout which T-DAB services should be provided. Implementation issues of transmitters did not lie in the scope of the conference in first place. Clearly, implementation has been an important point even during the meeting but it was not necessary to cope with that for the actual frequency assignment process.

The reason for that can be traced back to the fact that T-DAB as a digital terrestrial broadcasting system offers the possibility to be operated in terms of single frequency networks. Thus, the dimensions of coverage areas are not directly linked to the characteristic features of a single transmitter. Instead, it is possible to invert the process, i.e. first to define the coverage area and afterwards look for transmitters that operated together as a SFN actually guarantee the coverage. It is evident, however, that also this freedom has its limits in the sense that the outgoing interference produced by a network has to be controlled in order not to perturb an neighboring co-block network.

The third important conference in connection with T-DAB and DVB-T took place in 1997 at Chester, UK. In contrast to both conferences just mentioned the focus was different. The scope was not to assign frequencies but to find means to manage the transition from analogue TV to DVB-T that could be applied throughout Europe. Strong emphasis has been put thereby on the protection of existing analogue television stations. International coordination rules have been established at Chester in order to organize the anlogue-digital switch-over. This gave the possibility for each European country to initiate a step-by-step introduction of digital terrestrial television networks towards a full digital final scenario.

Seen from a frequency planning perspective, however, the final agreement of Chester does not stand for efficient usage of spectrum. The main reason for that is connected to the fact that the planning legacy from Stockholm is carried over to the digital world by the Chester final acts. Both protection of existing analogue stations and the shortcomings of the Stockholm plan with respect to a direct adoption to digital terrestrial broadcasting systems need to be reconsidered in the future.

Further conferences to regulate the spectrum usage of T-DAB and DVB-T have been already carried out or are currently being prepared. In June 2002 another CEPT conference took place at Maastricht, Netherlands, with the aim to assign seven additional T-DAB blocks from the upper part of the L-band. At present, the preparations for a revision of the Stockholm plan are under way. This conference will be organized in two parts, one in 2004 and the second part in 2005/6 or even later. Project groups have taken up their work to set the frame for a successful new frequency plan which hopefully will pave the way for T-DAB and DVB-T on large scale.

4.3 Assignments and Allotments

The main difference between the Stockholm and the Wiesbaden conference with respect to the planning principles applied is that according to common terminology Stockholm was based on assignment planning whereas at Wiesbaden allotment planning was preferred. The difference can be summarized by saying that at Stockholm requirements for spectrum were handed in by specifying all characteristic features of each transmitter. The frequency assignment was strictly linked to these data. Any modification of the transmitter like displacing the transmitter site, changing the output power or the antenna pattern and the antenna height have to be coordinated with all relevant neighbors. Finally, after all negotiations have settled and come to a result the new characteristics have to be notified to the ITU. After having checked the changes with respect to the principles of the Stockholm agreement, the data are archived.

In contrast, an allotment represents a geographical area inside of which a defined spectrum range can be utilized to provide a particular broadcasting service. Implementation issues in terms of one or more transmitters and the determination of the their characteristics is not touched upon by the allotment definition per se. It is only assumed that a future implementation of a transmitter network using the given frequency will be carried out in accordance with predefined coordination rules. When the network is actually implemented the allotment is formally converted into a set of assignments.

The big advantage of allotment planning as seen from the network providers point of view is that any terrestrial transmitter network can be modified at any time without coordination with neighboring countries as long as the total interfering field strength produced by the network outside its proper coverage area is not increased. This gives the network providers the possibilities to adapt their network structures in a more or less dynamic manner to economical changes or changes concerning the coverage target.

For the revision of the Stockholm plan it is necessary to incorporate both planning approaches. This is due to the fact that the revision conference has to deal with digital and analogue systems at the same time. So, both principles have to be adopted appropriately to cope with all potential input requirements.

Needless to say, this implies that the difference between assignment and allotment but also their common ground in terms of an unified planning approach is one of the central issues that the preparation activities of the upcoming revision of the Stockholm plan have to focus on. For the case of digital terrestrial transmission systems giving the possibility of single frequency operation of the networks it is necessary to define methods which formally map a requirement given in terms of an assignment into a corresponding allotment without problems.

One straightforward solution to this mapping task could be that for each assignment under consideration the associated analogue coverage area is calculated. This area could then be formally interpreted as an allotment area and subsequently dealt with accordingly. Certainly, the details of such a procedure still have to be elaborated in detail. But, since a formal mapping from an assignment to an allotment seems to be technically feasible this book is only concerned about allotment planning. All methods presented here, however, can be very easily applied to assignment planning as well.

4.4 Equitable Access to Spectrum

Independent of the spectrum ranges under consideration at a conference or the type of telecommunication system for which spectrum has to be provided all planning conferences are governed by some irrevocable principles. One of these pillars is the equitable access to spectrum. This means that all nations have the same rights to utilize any telecommunication service. Irrespective of political bias or economical power any country can claim equal demand for radio or television services. This is also independent of the geographical location or the size of a country. The same holds for the number of inhabitants or their geographic distribution across the area of the considered country.

Equitable access to spectrum refers to future, not yet realized usage of spectrum as well. Just in relation to the very different stages of economical development of the countries around the globe this item is very important. A frequency assignment following a "first come, first serve" principle in not acceptable on no account.

Even though demanding equitable access is an easy thing to do, giving a proper definition in the context of a radio conference is far from being trivial. Looking at past conferences it turns out that so far frequencies have been distributed mainly in terms of wide area coverage only. This refers to the idea to provide the frequency resources such that every single spot throughout the territory of a country could be served in principle. To do so, the entire national area is usually subdivided into smaller pieces of regional or even local character for which frequencies are required.

The traditional realization of equitable access to spectrum hence is to allow for the same number of nation-wide coverages in each country. Certainly, this represents an evident and working application of that principle. However, in the age of digital broadcasting systems and an ever increasing influence of market forces onto the implementation of transmitter networks it might turn out that different approaches are needed in the future. Equitable access could be determined by the total data capacity a country is going to use independent of the underlying character of the coverage area. Another possibility would be to link the assigned frequencies to the number of potential customers.

Any definition of the principle of equitable access to spectrum seems to be rather complicated and even politically delicate. Due to the enormous differences in terms of economical power for example between countries participating at radio planning conferences it is sometimes not possible to arrive at common view on that issue. In any case, the least common denominator always seems to be to say equitable access has been granted already by the right of any country to participate at a radio conference and thereby putting forward its requirements to be treated on an equal basis.

Chapter 5

Mathematical Basis of Frequency Assignment Problems

The assembly of the UHF frequency plan in Stockholm 1961 was accomplished almost entirely by hand. Computers as they are known and used today did not exist at that time. Granted, there were computers around already in the sixties but their performance was that of a contemporary pocket calculator.

At the Wiesbaden conference in 1995 the situation had changed completely. The making of the frequency plan was done with the help of computers without exception. Both the interference produced by one transmitter network on others as well as the actual derivation of the frequency plan would not have been possible without computers.

Even though the number of requirements at Wiesbaden was only about 700 in comparison to 5300 in Stockholm there were an awful number of very different constraints to be met. The initially derived plan was full of incompatibilities due to the lack of enough spectrum. To cure the shortcomings a lot of hand work was necessary. Incompatibilities had to be negotiated on a bi-or multi-lateral level. However, also that process was accompanied by adequate computing power. The reason for that was that the mathematical problems at the bottom of the frequency assignment process were so complex and involved that an educated guess did not give rise to reasonable solutions in most cases.

The common ground on which both conferences rested, however, was to map the actual real world frequency assignment problem to a mathematical model for which solutions could be found by mathematical means. Clearly, frequency assignment problems popping up in different areas of telecommunications have many similarities. Since the focus of this book is on digital terrestrial broadcasting only those methods will be discussed which are relevant in the context of T-DAB or DVB-T networks.

5.1 Mathematical Model for the Frequency Assignment Process

Any mathematical model of the frequency assignment process is based on the definition of the coverage areas, their mutual interaction or interference potential and mathematical algorithms for the actual assignment of frequencies, i.e. T-DAB blocks of TV channels onto coverage areas.

5.1.1 Definition of Coverage Areas

Before a radio frequency planning conference like the revision of the Stockholm agreement in 2004-2006, all affected administrations are called upon to identify their individual demands. This may vary from country to country and be subject to completely different media political constraints. Nevertheless, the resulting requirements are to be cast into a set of correspondingly defined coverage areas to be dealt with during the allotment planning process.

The geometry of real world coverage areas usually is very complex. In most cases they are not given in terms of simple geometrical shapes like circles or quadrilaterals. The line separating all served points from those not being served has a very complicated structure. A coverage area needs not necessarily be simply connected in a mathematical sense. Holes in the middle of the coverage area are the order of the day in regions with complicated topography. Therefore, the area that is to be understood as the coverage area for a particular service very often consists of several not connected and isolated pieces.

The standard way of describing such geometrical two-dimensional objects is to employ polygons as an approximation to the real shape. This can be implemented very well in software and leaves enough freedom to increase the accuracy of the approximation simply by increasing the number of vertices of the polygons. Figure 5.1shows a simple example from Germany.

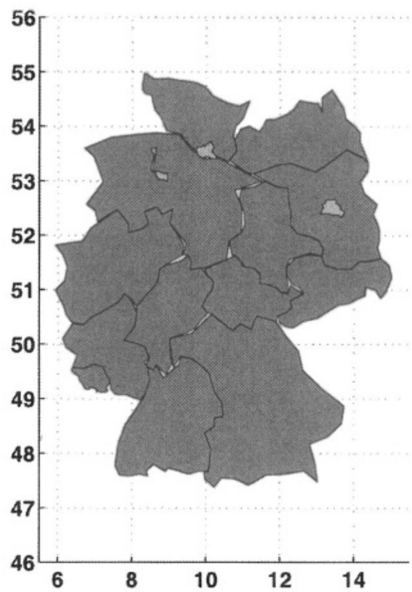

Figure 5.1: Example for a set of coverage areas which are given in terms of polygons.

One of the basic mathematical operations to be applied over and over again in relation to frequency assignment is to calculate the distance between two geographical points, like for example the transmitter and the receiver location. For calculations on an international scale the earth's curvature cannot be neglected thereto. In most cases, it is, however, sufficient to approximate the earth by a sphere. More elaborated analyses which assume the earth being an ellipsoid or a geoid are usually not necessary. The error which is made by relying on the spherical approximation is negligible in comparison with other error sources. Already the approximation of coverage areas by polygons for examples introduces variations with respect to the actual location of the allotment boundaries in the region of ten or twenty kilometers. The differences between distances calculated on the basis of spherical or ellipsoidal approximations are by far smaller.

When applying the spherical approximation of the earth, it is clear that the coverage polygons are to be considered as polygons on a sphere. Thus, every geometrical relation between two polygons has to be investigated on the basis of spherical geometry. As usual, the polygons are specified by their number of vertices and the coordinates of each vertex. Usually, a maximum allowed number of vertices for a single polygon is defined. For national borders 99 vertices are common whereas regional or local coverage areas in most cases have to be approximated with only 36 vertices.

5.1.2 Reference Networks

In order to avoid unacceptable perturbations by interference geographically adjacent transmitters broadcasting different programmes are obliged to use different frequencies which are spectrally separated far enough. The answer to the question what is the minimum allowed separation distance between transmitters using the same frequency or how much two frequencies need to be separated given the distance between them depends on a several issues. Sure enough, the characteristics of the transmitters and the topographic conditions along a line connecting them are very important. If the transmitter sites are separated by high mountains, then they are geographically decoupled. In this case each of the transmitters can be operated independently from the other.

Furthermore, the utilized frequency range determines the possibility to re-use the frequencies. The higher a frequency is the larger is usually its effective propagation attenuation. Generally speaking, electromagnetic waves in the VHF possess a larger propagation range than in the UHF regime under the assumption that the transmitters have the same power.

Finally, it is essential what types of broadcasting systems are under consideration. There is a big difference concerning the re-use of frequencies if two FM stations, two T-DAB transmitters or DVB-T transmitters are analyzed. The interaction between different types of broadcasting systems, like for example analogue TV against DVB-T, leads to further special mutual protection demands.

The assessment of the interference potential arising from a particular transmitter can be very easily accomplished. That is also true for the inverse problem, namely the investigation how much a transmitters is interfered by others with respect to its coverage area. This stems from the fact that all relevant details of the transmitter representing the assignment are given. The transmitter location is known, its output power and the antenna diagram as well. On the basis of an adequate wave propagation model the field strength can be predicted at an arbitrary point (see section 3). This gives a first indication about the coverage status of that pixel by comparing with the minimum required field strength. If all contributions from other transmitter at point are known, too, it can be checked if the required protection ratio between the useful signal and the sum of all interfering signals is exceeded or not.

Clearly, this requires that corresponding knowledge about the protection ratio is indeed available. Even though this might seem obvious the derivation of appropriate values is far from being trivial. In the case of T-DAB and DVB-T a vast number of both theoretical and experimental investigations were carried out to determine the necessary protection ratios for all possible system variants and combinations of interacting system variants. The final acts of the Wiesbaden planning conference [CEP95] as well as those of Chester [CEP97] contain all relevant data in detail.

The situation is completely different for the case of pure allotment planning. It is a characteristic feature of that planning principle that no transmitter specifications are made. However, it is evident that also here adjacent or overlapping areas cannot make use of the same frequencies to broadcast different content. Joining two adjacent or overlapping areas to build a larger common coverage area in which the same programs are to be delivered is a different story. Different content on the same frequency will definitely lead to not acceptable interference in parts of both affected areas. The basic and crucial question thus is how it is possible to assess the outgoing interference potential of an area in which a dedicated frequency can be used to provide a certain broadcasting service without the detailed specification of all transmitters employed for the network?

To answer that question the concept of the so-called reference networks has been developed [OLe98]. As the name already indicates, a reference network is a single frequency network consisting of a defined number of transmitters which generates a coverage area that is to resemble the size and the shape of a typical realistic allotment area.

The location of the transmitters is fixed, as well as their characteristic features like power and antenna patterns. Moreover, their effective heights have to be provided, too. The coverage status is calculated by means of the ITU-R recommendation P.1546 [ITU01a]. At Wiesbaden there were two such networks defined for T-DAB, one for the VHF range and one for the L-band. With respect to the Stockholm revision it is clear that the situation will be more complicated due to the large number of different operation modes of DVB-T.

Both reference networks from Wiesbaden consist of seven transmitters. They are laid out in form of a hexagon with one transmitter located in the very center. The other six have their position at the vertices of the hexagon. According to the different propagation conditions in VHF and L-band different network parameters were chosen. The distance between two transmitters in the VHF network is 60 km whereas for the L-band only 15 km were decided to be appropriate. In both cases it is assumed that all transmitters have an effective height of 150 m for all directions. The power of the central transmitter is smaller than that of the other transmitters. Its antenna is considered to be omnidirectional.

There is one significant difference between the reference networks in VHF and L-band. For VHF the transmitters at the periphery have directional antennas which have their maximum lobe directed towards the center. The forward-backward ratio of the antenna diagrams is 12 dB with an aperture angle of 120 °. Such a configuration is called a “closed” network. In contrast, the L-band reference network employs omnidirectional antennas at the periphery as well. This is called an “open” network. The covered area is more or less equivalent to the area of the underlying hexagon with the difference that in the case of an open network the coverage area extends beyond the hexagon boundaries. Figures 5.2 and 5.3 highlight the differences.

Apart from the covered area there is another very important parameter which can be deduced from the properties of the reference networks. The impact of a reference network on other networks can be calculated explicitly since all characteristic features of the involved transmitters are given. This information can be employed to derive the so-called re-use distance for co-spectrum usage. The re-use distance is defined as the minimum distance the borders of two coverage areas must be separated so that both can utilize the same T-DAB block or TV channel. Then they do not mutually cause inacceptable interference levels.

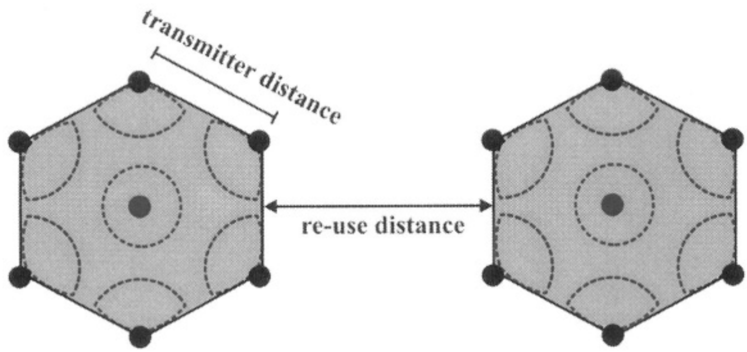

Figure 5.2: T-DAB reference network for the VHF range. The dashed lines sketch the antenna diagrams of the individual transmitters.

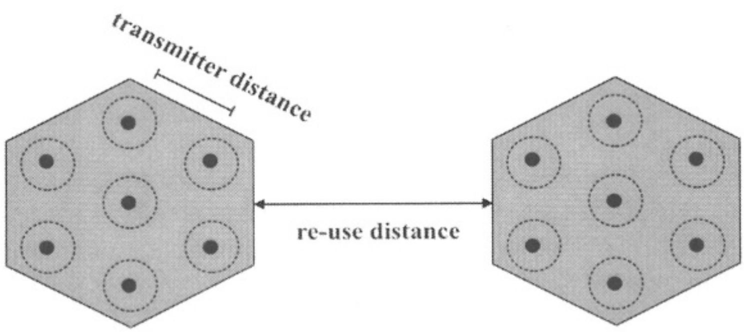

Figure 5.3: T-DAB reference network for the L-band. The dashed lines sketch the antenna diagrams of the individual transmitters.

In the case of T-DAB two re-use distances were calculated at Wiesbaden from the definitions of the two reference networks for VHF and L-band, respectively. According to [CEP95] for propagation above land in the VHF range the two networks must be separated by more than 81 km whereas in L-band the result is 61 km. If the two coverage areas are separated by cold or warm sea only, the critical distances increase to values of 142 km and 348 km for VHF and 173 km and 485 km in L-band. Cold sea refers to the Atlantic sea while warm sea stands for the Mediterranean sea. In case of mixed paths, i.e. paths consisting of both land and sea fractions appropriate interpolation schemes have to be applied [CEP95]. The relevant parameters of both reference networks are summarized in table 5.1 [CEP95].

	VHF	**L-band**
type of network	closed	open
inter-transmitter distance	60 km	15 km
effective heights	150 m	150 m
power of central transmitter	100 W	500 W
antenna diagram of central transmitter	omnidirectional	omnidirectional
power of transmitter at periphery	1 kW	1 kW
antenna diagram at periphery	directed towards center	omnidirectional
diameter of coverage area	<= 120 km	<= 60 km
re-use distance above land	81 km	61 km
re-use distance above cold sea	142 km	348 km
re-use distance above warm sea	173 km	485 km

Table 5.1: Parameters of the Wiesbaden reference networks for VHF and L-band.

When comparing figures 5.2 and 5.3 with the polygons depicted in figure 5.1 it becomes evident that coverage areas of reference networks in general do not coincide with the real shapes of allotment areas. This fact created and still is creating some confusion amongst people involved in the business of frequency management and network planning. In particular, after the conference of Wiesbaden the idea was widespread that inside an allotment area a network is to be implemented which corresponds to the geometry of the reference network, i.e. seven transmitters laid out as a hexagon. Clearly, this is not true. The reference networks serve only for the purpose to estimate the outgoing interference with the help of a hypothetical network whose size is adopted to the frequency range under consideration.

The figures that can be derived from the properties of the reference networks for the size of the coverage area and the re-use distance, however, have to be considered as very rough estimates for real world applications. This becomes clear when looking at the concept of reference network as a whole. Neither the network configuration nor the wave propagation model represents the real world with a high accuracy. Reference networks do not take into account any topography. So, they are applied in Netherlands and in the Alps without any modification. It is obvious that this cannot be completely correct. On the other hand, this shortcoming leaves enough freedom for negotiations both during the frequency assignment process as well as with respect to network implementation issues.

Subsequent to the Wiesbaden conference and during the preparations for the L-band conference which was held in Maastricht, Netherlands, in 2002 it became evident that further types of reference networks would be desirable. They could help to better do justice to all different demands of broadcasters and network providers in Europe. It had been shown that very often only very small coverage areas are required to be served by T-DAB in L-band. Unfortunately, these areas cannot be represented adequately by the former L-band reference network. Therefore both CEPT and EBU collaborated to produce two more types of smaller reference networks which were used for the first time during the Maastricht conference [CEP02].

The frequency assignment process both at Wiesbaden and Maastricht was based on the assumption that the allotment areas could be represented by one of the reference networks at least in relation to their outgoing interference potential. As mentioned already, any type of reference network gives rise to particular re-use distances. So, once an allotment area is associated with a certain reference network variant it is in principle straightforward to check if two allotments are allowed to use the same frequency or not. To this end, their mutual geographical separation simply has to be compared with the corresponding re-use distance.

However, for both conferences a slightly different approach was chosen. In order to assess the outgoing interference it is essential to know what field strength is produced by the set of transmitters of a reference network outside the coverage area as a function of the distance from the border. So, calculations were carried out to find these field strength curves. Then it is assumed that a real network implemented within the considered allotment area would produce an outgoing interference according to these derived curves.

The idea behind the application of these curves to arbitrarily shaped allotment areas is the following. In order to calculate the field strength that is produced by an allotment area on the border of another, a reference network is virtually aligned to the allotment such that both the border of the allotment and that of the reference network are congruent. Then the field strength curves are used to determine the field strength at the border of the second allotment border. That way the compatibility of two allotments is based on field strengths rather than on distances. Network planners usually are in favor of such an approach since they are used to dealing with field strengths and protection ratios.

However, there is no principle difference between the two possibilities of deciding if two areas can get the same frequency or not. In both cases violations of the constraints to be met can be dealt with very easily. Working with re-use distances then simply means to allow areas which are separated less than the critical distance to use the same spectrum. On the other hand, for a treatment based on field strengths this corresponds to a higher interfering field strength at the border of an allotment to be accepted.

5.1.3 Effective Distances and Adjaceny Relations

Before any allotment planning activities can start at a conference it is necessary that all requirements are known and given in terms of corresponding polygons approximating the required coverage areas. The first step on the way to a frequency plan is to investigate the mutual interference status of all pairs of polygons. To this end, one of methods described above has to be used. The re-use distance method seems to be less technical and can be understood intuitively. Therefore the discussion will proceed along that line only. However, this does not imply any restricted validity of the conclusion presented here.

Thus, checking the mutual interference potential between two areas means to calculate the minimal distance between them and compare it to the corresponding re-use distance. This has to be carried out then for all pairs of allotment areas. To make the discussion more transparent, a very simple example is used in order to explain the context. Five polygons are sufficient for that purpose. They are sketched in figure 5.4.

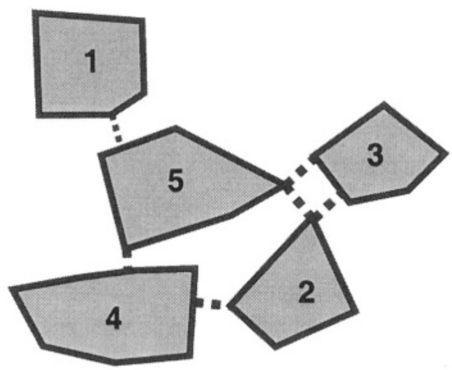

Figure 5.4: Five coverage areas approximated by simple polygons. The red, dashed lines indicate those minimal distances between two areas which are assumed to be smaller than the re-use distance.

All calculated distances between the polygons are stored in a symmetrical matrix. Sometimes it is helpful to derive an auxiliary matrix from this distance matrix. The former is usually called the adjacency matrix. Its elements are either one or zero depending on the decision if the minimal distance between the two considered areas is larger or smaller than the re-use distance. In figure 5.4 the re-use distance has been chosen such that an adjacency matrix of the form

$$\mathbf{N} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} \tag{5.1}$$

arises.

As long as only one type of propagation, i.e. land, cold sea or warm sea has to be accounted for nothing changes essentially. Care has to be taken only with respect to selecting the correct re-use distance. In this case the minimal distance between two areas must be one of two particular connecting lines between the polygons. Either it is given by a line connecting two vertices or a line starting at one vertex and ending at some location on an edge perpendicularly. Figure 5.5 illustrates the possibilities.

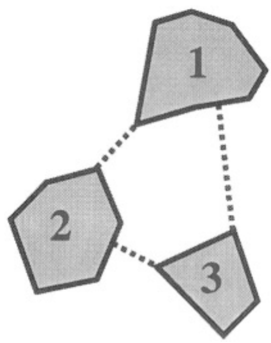

Figure 5.5: Minimal distance between polygons under the assumption that only one type of propagation is relevant. The minimum separation line connects either two vertices or one vertex and an edge to which it is perpendicular then.

If mixed paths come into play, then the analysis is getting more complicated. The simple concept of the geometrical distance can no longer be applied directly. The different propagation conditions add to result in one effective interference impact. Clearly, the influence of a particular propagation type will depend on its corresponding fraction of the total path length. This has an important consequence. Figure 5.6 depicts the new situation.

The two areas 1 and 2 are partially separated by water. Path A is the shortest pure land path. However, it is very likely that the critical path, i.e. the path whose effective impact is largest, includes a passage over water like e.g. path B. As can be seen from figure 5.6 such a path must no longer connect two vertices or one vertex and an edge perpendicularly.

If the interaction analysis for two areas is based on field strength curves derived from ITU-R recommendation P.1546 mixed paths can be dealt

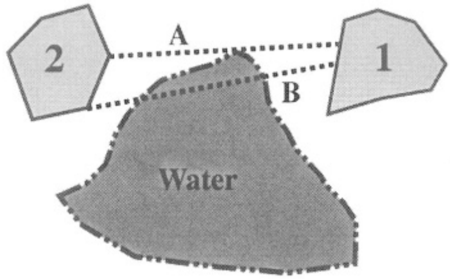

Figure 5.6: Different effective distances between areas for mixed path propagation.

with according to the interpolation schemes given therein. For the re-use distance approach an adequate formulation does not exist. However, a very simple and evident method which rests on appropriate scaling of distances gives satisfying results.

To explain the idea behind this approach the parameters for the VHF reference network of Wiesbaden are employed for the time being. According to Wiesbaden a warm sea path of 173 km has the same interference potential as a land path of 81 km. Now, if the distance between two polygons is composed by a fraction of x km above land, y km above cold sea and z km above warm sea then the combined interference potential is assumed to be equivalent to a situation when the two areas would be separated by a distance

$$d = x + y * \frac{81}{142} + z * \frac{81}{173} \tag{5.2}$$

above land only. The quantity d is called the effective distance between the two areas. Then, the decision if the two areas could use the same frequency or not is to be taken by comparing the effective distance d with the re-use distance for land propagation. Naturally, the same method can be applied for any spectral range and any type of reference network as long as the re-use distances for land, cold sea and warm sea are known.

It has to be noted that the concept of the re-use distances which rests on the wave propagation model according to the recommendation [ITU01a] does not take into account any topographic conditions between the two areas under consideration. Clearly, the adherence of the re-use distance criterion becomes less critical if both areas are geographically decoupled. To give an example, at Wiesbaden there were polygons on the northern and the southern side of the Alps that were separated only by a few kilometers. But, due to the high mountains in between they could be considered decoupled without problems. If more effort is put into the calculation of the effective distance than done with the application of equation (5.2) then such configurations could be detected from the very beginning.

5.2 Graph Theory Based Algorithms

Once the adjacency matrix has been set up all information is available which is necessary to assign frequencies onto coverage areas. Clearly, an important objective is to utilize as few frequencies as possible. Even if on a first glance this seems to be an easy task it turns out that in fact the underlying mathematical problem is extremely complex.

A short example might give some insight. For the time being it is assumed that the number of coverage areas to be dealt with is 20. For a single country this is a realistic figure. On a European scale, however, hundreds or thousands of polygons are the order of the day. Furthermore, it is supposed that one of 10 different frequencies can be assigned to each of the areas. This means all coverage areas are on an equal footing in terms of the absence of access restrictions for individual frequencies. Under these circumstances there are 10^{20} different possibilities to assign a frequency to each of the polygons.

Now, the target is to find an assignment of frequencies onto the areas which is compatible with the adjacency relations and makes use of a minimum number of frequencies. Usually, this task is accomplished by the help of computers. The actual assignment, the compatibility check and the determination of the number of utilized frequencies clearly depends on the available computing power. One microsecond might be a typical value. In order to find the one and only optimal solution to the frequency assignment problem one could simply check each of the possible solutions. If a new best solution is found temporarily it is saved and overwritten if a better one appears.

This method would require a time $t = 10^{20} * 10^{-6}\text{sec} = 10^{14}\text{sec}$ which corresponds to $2.78 * 10^{10}$ hours or $1.16 * 10^{9}$ days or $3.17 * 10^{6}$ years. In other words, the computer would need to run for more than three million years! Even if a supercomputer would be available which reduces the computing time of one check to just one nanosecond still the total time would exceed three thousand years. This simple result proves very clearly that any so-called brute force approach like the one described here is completely useless in practice.

Seen from a mathematical perspective frequency assignment falls into the class of NP-hard mathematical problems. This means that the computing time to find the optimal solution grows faster than exponentially with respect to the characteristic system size. For a very small number of coverage areas it is possible to calculate the optimal solution by a simple trial and error strategy. However, as can be seen from the example given above already for 20 areas this is no longer feasible. In reality, problems are much more complex, so different strategies must be developed.

5.2.1 Frequency Assignment and Graph Theory

Keeping in mind the difficulties in view of finding the optimal frequency plan the question must be asked if this is really always necessary for practical applications. Can it not be satisfactory to get hold of a nearly optimal solution that can be found, however, within a finite and reasonable interval of time? In order to rely on such a workaround it would be very helpful if some criteria would exist which would give some indication about how far the accepted nearly optimal and the one and only optimal solution are apart.

In principle, graph theory can deliver both, namely very good approximative solutions and lower bounds for the minimal number of required frequencies. A very good and clear introduction to the field of graph theory can be found in [Jun98]. For a very detailed description of the application of graph theory in particular to problems arising in frequency assignment for T-DAB it is referred to [Gra02]. A very extensive reference for a lot of other different aspects of graph theoretical issues is the website of the Konrad-Zuse-Zentrum für Informationstechnik in Berlin (ZIB). This website contains a section which also provides a very long list of scientific articles connected to graph theory and frequency assignment [ZIB01].

The connection between frequency assignment and graph theory can be established very easily. A situation as depicted in figure 5.4 can be mapped to a simple graph on the basis of the adjacency matrix (5.1). Figure 5.7 shows the graph corresponding to figure 5.4. Every polygon shown in figure 5.4 is represented by a vertex in figure 5.7. Two areas which according to the adjacency matrix (5.1) are not allowed to use the same frequencies are connected by an edge.

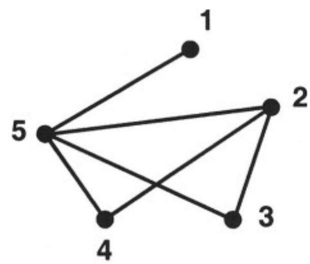

Figure 5.7: Simple graph corresponding to figure 5.4.

The original task to assign frequencies to areas thus is equivalent to the problem to color a graph. This means to assign each of the vertices of the graph a color such that vertices that are connected by an edge do not get the same color. In practice, the property "color" is represented simply by an integer. Clearly, the coloring has to be accomplished under the constraint to use as few colors as possible. The minimum number of required colors is called the chromatic number χ of the graph.

What would be the most natural and obvious way to color the graph? Clearly, the most simple thing to do would be to choose an arbitrary vertex and assign a color to it. Then a another vertex not yet being dealt with is chosen. It is given a color in accordance with its individual adjacency relations. This process is continued until all vertices have been assigned colors. Finally, the number of different colors used is determined. By definition, such a strategy will lead to an admissible solution of the graph coloring problem. However, its drawback is that the quality of the result can be arbitrarily poor in practice.

The quality of the solution can be improved considerably by taking into account the adjacency relations of the vertices in an appropriate way. When looking at the graph of figure 5.7 it becomes obvious that all vertices have different types of adjacency relations. Vertex 1 only has one single neighbor while vertex number 5 is connected with all other vertices. This differences can be quantified with the help of the so-called vertex degree κ. The degree of a vertex corresponds to the number of edges ending at the vertex under consideration. Coloring methods making explicit usage of the vertex degrees of the graph will be presented in the following sections in detail.

Clearly, those vertices having large degrees very likely require a large number of different colors to be used in order to satisfy all adjacency relations appropriately. Nevertheless, the knowledge of all vertex degrees of a graph is not sufficient to determine the chromatic number χ. It is not even sufficient in general to derive a satisfying lower bound for the minimum number of required colors.

This can be understood when looking at figure 5.4. Vertex 5 has degree $\kappa = 4$. It can be shown, however, that three colors are needed for this simple graph. So obviously, the maximum vertex degree of a graph is not necessarily linked to the chromatic number.

In order to find a reliable bound for the chromatic number it is necessary to consider subsets of vertices, so-called sub-graphs, instead of dealing with individual vertices alone. Moreover, it is crucial to identify complete sub-graphs. A complete graph is a graph whose vertices are all mutually connected by edges, i.e. all vertices have the same degree κ. The number of vertices ω of a complete graph and its vertex degrees κ are linked by

$$\omega = \kappa + 1 \ . \tag{5.3}$$

Sometimes a complete sub-graph is also called clique and its number of vertices ω is the clique number.

Each graph which does not consist only of isolated vertices contains complete sub-graphs. The reason is that already two connected vertices constitute a trivial complete graph. Figure 5.8 shows a simple example of a graph having several non-trivial complete sub-graphs. The vertices {4, 5, 6} or {2, 3, 4, 5} for example form complete sub-graphs.

An important parameter connected to the existence of complete sub-graphs is the maximum clique number $\omega_{\max}$ which is nothing but the number of vertices of the largest complete sub-graph of the considered graph. Clearly, the maximum clique number $\omega_{\max}$ determines the chromatic number to a large extent. Indeed, it can be considered as a lower

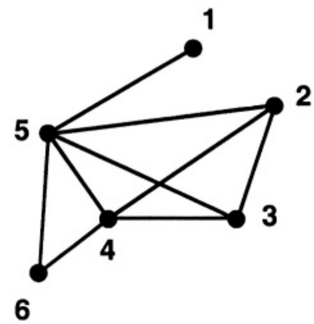

Figure 5.8: Graph containing non-trivial complete sub-graphs.

bound for χ, i.e.

$$\omega_{\max} = \kappa_{\max} + 1 \leq \chi \tag{5.4}$$

where $\kappa_{\max}$ is the corresponding degree of the vertices.

In most cases the maximum clique number $\omega_{\max}$ represents a very good lower bound in the sense that the chromatic number χ is not much larger. The problem, however, is that the determination of the clique number falls into the class of NP-hard problems, too, like the graph coloring itself. Thus, brute force approaches to find the chromatic number are usually of no help.

In figure 5.8 the set of vertices {2, 3, 4, 5} not only forms a complete sub-graph, but at the same time the vertices form the maximum clique as well. For this particular graph the chromatic number is 4 which is equal to the maximum clique number. In general, this need not be so.

For a graph consisting of six vertices it is no problem to determine the maximum clique number even by trial and error. For medium size graphs, i.e. a number of vertices in the order of 1000, it is possible to exploit an exact algorithm which provides the maximum clique number and the vertices forming the maximum clique within an acceptable period of time [Car90].

5.2.2 Sequential Graph Coloring

Vertices having large degrees might lead to a high demand for colors. However, this statement has to be interpreted with care as seen above.

Nevertheless, it is reasonable to start the coloring process with those vertices being linked to many others. This is guided by the hope that there is enough freedom for the rest of the vertices to assign a color once the difficult ones are dealt with without increasing the overall consumption of colors.

Indeed, it became evident that such strategies are likely to produce good results. But, it should be born in mind that even if algorithms are utilized which are based on strict and clear rules it cannot be guaranteed that the optimal solution will be found this way. Normally, this will not be the case. Hence, all deterministic algorithms which do not necessarily lead to the determination of the optimal solution like a straightforward trial-and-error approach are called heuristics.

The most simple way to color the apparently critical vertices at the beginning is to establish a sequence of the vertices according to their degrees. Typically, a descending ordering is applied. However, the inverse ordering from small to large degrees can give good results, too. Then, all vertices are sequentially given colors such that each assignment is in accordance with the corresponding adjacency relations.

In terms of mathematical algorithms the attribute "color" of a vertex is usually represented by an integer. So, applying a sequential coloring strategy means for example to assign the number zero to the first vertex of the sequence. The next one gets the smallest integer which is compatible with all its neighboring vertices that already were assigned a number. This is continued until each vertex has been dealt with.

Experience shows that a descending order is more advantageous than an ascending one. However, one should in any case check both possibilities because it cannot be excluded that another ordering might give a better result. Table 5.2 shows the assignment of colors for the graph depicted in figure 5.7 obtained by employing a vertex sequence of descending degree order.

vertex	5	2	4	3	1
color	0	1	2	2	1

Table 5.2: Result of the sequential coloring for the graph shown in figure 5.7

The arrangement of vertices according to descending or ascending vertex degrees obviously is not the only possibility. The so-called “smallest-last” ordering is another very famous and promising way to build a sequence [Mat83]. In contrast to the aforementioned schemes the generation of this sequence is more complex.

The first step is to set up a sequence with descending vertex degree. The vertex at the end of the series, i.e. the vertex having the smallest degree, is shifted to the last position of an auxiliary sequence. Then this vertex is removed from the original graph together with all the edges ending in that particular vertex. Thereby a new graph is created which has one vertex less than the original one.

For this new graph once again a vertex ordering based on descending degrees is established. The last vertex of this sequence is shifted to the last but one position of the auxiliary sequence. As before, a new graph is build by removing this vertex from the actual graph thus giving rise to the second new graph having two vertices less than the original one. This procedure is continued until all positions in the auxiliary sequence are finally filled. The subsequent coloring is carried out as in the case of table 5.2.

If several vertices have the same minimal degree, additional criteria have to be applied in order to decide which vertex is to be removed from the graph. To this end, there are different possibilities. One of the vertices can be chosen by chance or that vertex is removed which had the smallest degree of all vertices under consideration with respect to the original graph. Figure 5.9 sketches the principle approach.

Other systematic ordering strategies have been investigated in recent years. The ordering according to the maximum saturation is one of the most famous approaches [Bre79]. For further methods it is referred to [Wei98]. Unfortunately, there is no generally valid rule which suggests one or the other procedures in a given situation. Therefore, it is advised to always try several different sequences. This strategy has been followed at the Wiesbaden conference, too, where over 1000 different ordering schemes have been exploited [OLe98].

Furthermore, it has to be emphasized that a graph coloring which is carried out on the basis of one the methods described here implies that all constraints in terms of adjacency relations are strictly obeyed. In case two vertices are linked by an edge they are strictly assigned distinct colors. Speaking in terms of frequency assignment this means the frequency

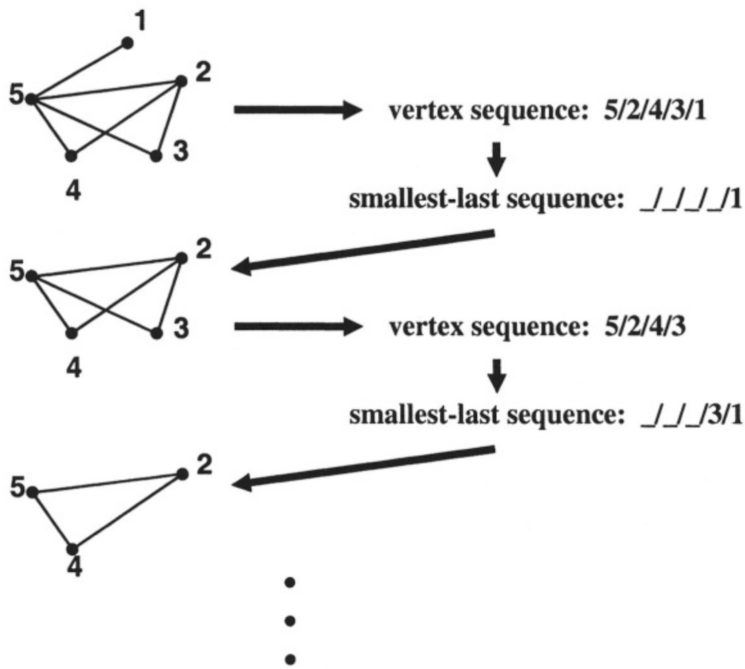

Figure 5.9: Smallest-last ordering of the graph 5.7.

plan derived thereby does not contain any additional interference beyond the accepted and unavoidable levels.

5.2.3 Weighted Graphs

The graphs shown in figure 5.7 or 5.8 represent very special frequency assignment problems. Two vertices can either get the same color or not. Basically, this is a clear and mutually exclusive decision. However, real frequency assignment problems are usually more complex and cannot always be mapped to a simple black-or-white decision.

A typical more involved example would be to incorporate constraints on how far the frequencies must separated which are assigned to two connected vertices. The problems given in terms of the figures 5.7 or 5.8 tacitly imply a separation distance of one channel or block or whatever frequency unit is actually addressed.

Another task could be to find a frequency assignment for a set of coverage areas where certain areas using the same frequency would give rise to a certain penalty to be borne, for example a level of interference between them. The task then would be to find a distribution of colors onto the graph such that the total interference produced is minimal.

One way to represent both problems in terms of graphs is to introduce the concept of weighted graphs. This means that each edge can be provided with a number which has to be interpreted accordingly. Figure 5.10 shows a corresponding extension of the graph 5.7.

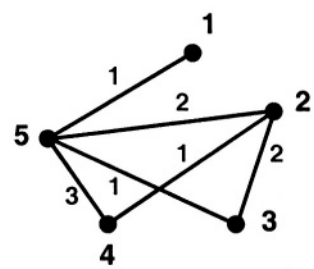

Figure 5.10: Simple weighted graph. The number at the edges might represent the number of channels the two frequency assignments must be separated.

The interaction between two vertices in the graph of figure 5.10 is symmetric. Sometimes it is necessary to introduce an asymmetric interaction and to operate on the basis of weighted and oriented graphs. Thereby, the complexity of the graph coloring problem is even more increased.

In principle, the same type of approaches to find a coloring as described above can be applied. However, it must be emphasized that the more constraints are to be met the less effective usually such simple and straightforward strategies become. Hence, a lot more effort has to be put into the development of specially adapted coloring methods. Unfortunately, this means they lose their general applicability at the same time.

5.3 Frequency Assignment by Stochastic Optimization

The more constraints have to be met during a graph coloring process the more complicated and complex the problem becomes. Deterministic coloring strategies tend to loose their efficiency under such conditions. Fortunately, there exist other methods which can be used instead. In order to solve highly complex frequency assignment problems stochastic optimization algorithms can be employed. Details concerning the mathematical background can be found in section A.

In principle, there are several possibilities for the application of stochastic optimization methods. A first and very straightforward utilization can be identified already in the case of the simple sequential coloring. As has been described above the ordering of the vertices is the crucial factor that determines the quality of the solution. The so-called Great Deluge algorithm described in detail in section A.1 could be used to find a sequence of vertices which leads to a minimum number of required colors. The search for such an optimal ordering is carried out in terms of an iteration.

To this end, an arbitrary sequence is chosen as the starting configuration and the resulting graph coloring is calculated by application of the sequential procedure. Then, two vertices are chosen arbitrarily and their positions within the sequence are exchanged. Again, the colors are assigned. The next step would be to decide if the second result should be kept as new starting point for the progression of the Great Deluge algorithms or not. By iterating this process a very good solution can be found.

It must be pointed out, however, that such an approach is only reasonable in practice if the number of vertices is large enough. For a small or moderate number of vertices the variance of the results is too small. Most of the sequences will lead to the same number of required colors. So, for the iteration process it is not obvious which of them is to be chosen as the new starting point.

If in such an undecided situation the old configuration is always kept, the algorithm carries out more or less a local search in the vicinity of the starting configuration. Since this sequence has been chosen by chance it can give rise to a very high color consumption. In the opposite case, when always the new configuration is chosen the algorithm proceeds along an arbitrary path in configuration space. Then the stochastic optimization process bears resemblance to an unrestricted Monto-Carlo simulation.

A further possible application of stochastic optimization to frequency assignment problems is connected to principle issues concerning the applicability of graph theory for that purpose. Situations are quite common where the available spectrum is so limited that none of the constraints can be satisfied exactly. Also, there might be completely different types of constraints to be obeyed at each vertex.

All these complications are not very easily modelled in terms of graph coloring problems. If a mapping to a corresponding graph coloring problem can be established, highly sophisticated strategies to solve it must be developed and applied. Thus, most of the graph theoretical algorithms are highly adapted to a very special set of constraints. Often they are optimized to cope with exactly these boundary condition as good as possible.

The derivation of a working frequency plan for a particular broadcasting service is a dynamic process. This means that constraints are very likely to change in the course of the generation of the plan. Hence, optimal adaption of an algorithm to a specified set of constraints clearly is a major obstacle when deriving a good frequency plan. A method which works very well for a certain type and number of constraints need not necessarily be as efficient if that basis changes.

Stochastic optimization algorithms have undoubtedly advantages in that respect. In contrast to dedicated graph coloring algorithms they have a more fundamental nature in the sense that they have to be considered as very flexible and generic principles to solve a problem rather than explicit instructions. Usually, they can be adapted very easily to different optimization scenarios. Clearly, the price to pay is that a flexible approach can never be as sufficient as a specially designed method.

In practice, one of the major problems frequency planning must face is the scarcity of available frequencies. As a consequence, high interference levels can usually not be avoided. A natural objective to set up a working frequency plan thus is to reduce interference as much as possible. All mutual perturbations to be expected between different coverage areas can be combined to generate a global quality criterion. Exploiting this criterion allows to distinguish between good and bad frequency plans. It is evident that such a quality can be employed as the target function of a particularly chosen stochastic optimization algorithm.

Chapter 6

Geometrical Allotment Planning

The starting-point of every allotment planning procedure is a set of specified geographical areas which are to be provided with TV channels as in the case of DVB-T or frequency blocks for T-DAB. In principle, there are several very different approaches that can be followed. One of them is the systematic assignment of frequency resources on the basis of periodic geometric structures. This can be traced back to the ITU planning conference of Stockholm in 1961. Originally, such strategies have been developed within the context of a pure assignment plan. However, they can be very easily applied to the case of allotment planning as well. Methods following completely different routes will be dealt with in the following sections in detail.

The application of systematic geometry based planning principles to allotment planning is straightforward. In case wide-area coverages are the planning target, this can be most easily achieved if simple geometric areas are used as basic elements to cover the territory. The whole planning region, i.e. for example Germany or Europe, is covered by a net consisting of these elements. The only constraint to be met thereto is to employ elements which allow for a gap free parquet of the plane. Rhombuses or hexagons are very well suited. Circles are usually not used since a gap free two-dimensional coverage of the plane can be generated only if overlapping areas are accepted. Figure 6.1 shows a rhombic net across Germany whereas in figure 6.2 a coverage in terms of hexagons is presented.

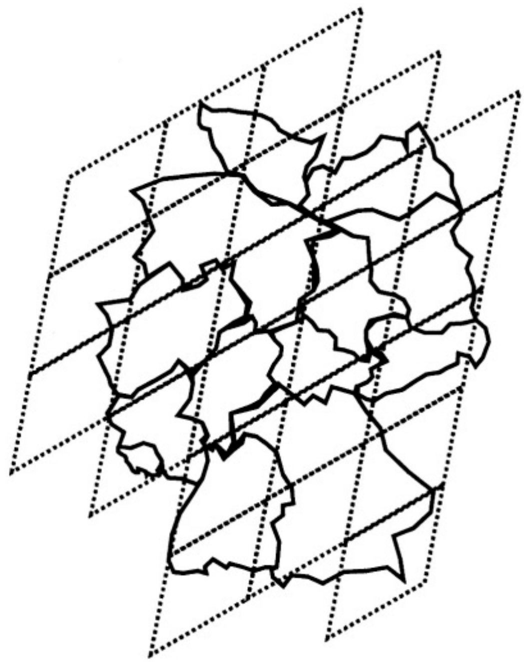

Figure 6.1: Allotment areas in terms of rhombuses across Germany.

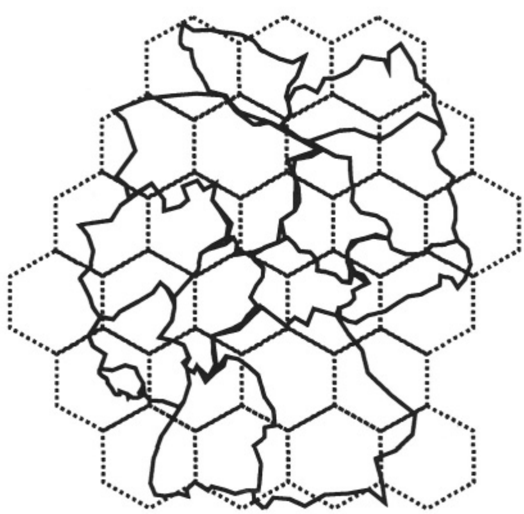

Figure 6.2: Allotment areas in terms of hexagons across Germany.

Frequencies are assigned to these symmetrical coverage areas by applying a deterministic scheme which respects the re-use distance criterion for co-channel or co-block usage. Depending on the ratio between the characteristic dimensions of the area elements and the re-use distance a different spectrum demand for an entire wide-area coverage results. Figure 6.3 illustrates the connection in the case of planning with the help of hexagons.

Even though this planning principle is very attractive from a frequency management point of view the question arises how the achieved distribution of frequencies is compatible with national and international media political requirements. A short look on the figures 6.1 and 6.2 makes clear that for example the boundaries of the German Bundesländer are nowhere respected by the presented layout of rhombuses or hexagons. In case such a principle would be applied at a radio conference for the derivation of the frequency plan, it would be obvious that the affected administrations and network providers would need to enter bi- or multilateral coordination negotiations after the conference to figure out the details of the frequency usage in neighboring countries.

Seen from a practical point of view the geometrical allotment planning does not possess a lot of relevance today. However, it has nevertheless great conceptional significance. First of all, geometric allotment planning represents some kind of optimal planning in the sense that usage of spectrum is optimal. Therefore, it can be employed as a reference in order to assess the efficiency of other planning results.

Moreover, a very simple but nevertheless important rule of thumb for the generation of allotment areas by administrations and the thereby emerging consequences for the spectrum demand can be derived from its results. The extension of allotment areas in relation to the re-use distance determines the number of required frequencies. Therefore, it has to be taken care when defining allotment areas that their minimal extension should preferably be larger than the relevant re-use distance.

This corresponds to the optimal situation. Every violation of this rule inevitably leads to an increase of the spectrum demand. To quantify the actual demand it is necessary to investigate the mutual interference potential between the allotment areas in detail. In view of the complexity of typical frequency assignment problems no generally valid tendencies can be given thereto in advance.

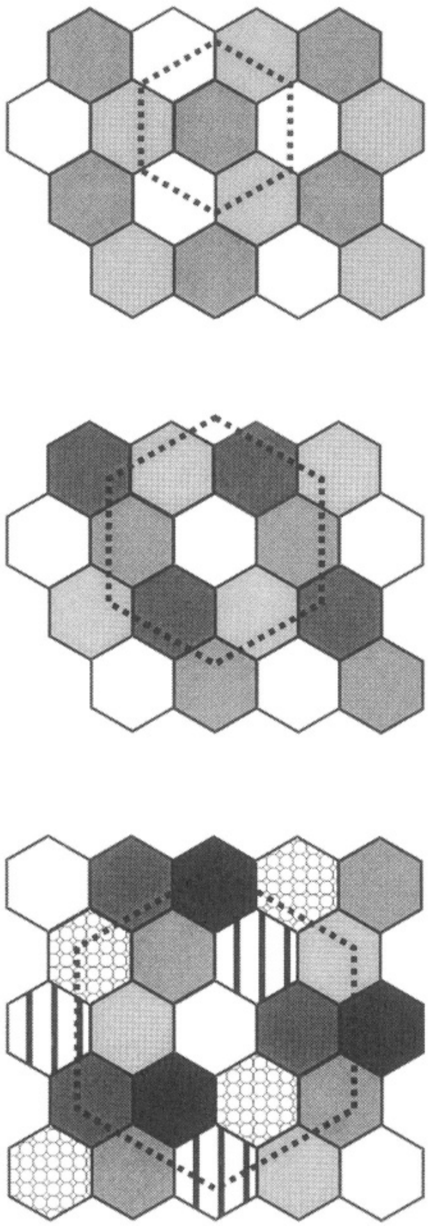

Figure 6.3: Three gap free parquets on the basis of hexagons. The re-use distance is indicated by the dashed hexagon surrounding the central hexagon. The spectrum demand is 3, 4 and 7 different frequencies, respectively.

Chapter 7

Spectrum Demand Estimation

The successful introduction of radio or telecommunication services has to meet several technical and non-technical constraints. The technical side deals with the question which transmission technique is utilized to provide which service. Non-technical issues are economic feasibility and efficiency which tend to become more and more important. This touches also upon the question about the coverage target. Which special services are to be provided in which areas and how much of these can be implemented in accordance with consumer demands and their will to spend some money for these new services?

In the case of radio and television the situation has an additional flavour which is not present for pure telecommunication services like mobile personal communication. The reason is due to the role public broadcasters play in the field. They are liable to national or regional governments and their corresponding administrations under rules defining precisely the task and scope of public broadcasting.

Political and economic demands thus stipulate the introduction of radio and telecommunication services. Once a road map for the introduction and the future development has been settled the question arises how much spectrum is necessary to bring the considered service into operation. Hence, spectrum demand estimation is a crucial issue to be investigated in detail. Mathematical tools to accomplish this estimation are appropriately chosen frequency assignment methods.

In contrast to situations where there is limited spectrum available, spectrum demand estimation rests on the fundamental assumption that an infinite number of channels or frequency blocks is on-hand. Frequency assignment problems under the restriction of spectrum scarcity will be dealt with in subsequent chapters. This distinction might seem rather arbitrary, however, experience shows that there are significant practical differences between both problems. Therefore this formal separation can be justified.

7.1 T-DAB Coverage for Europe

The frequency block distribution achieved at CEPT conference of Wiesbaden in 1995 for T-DAB was based on wide-area coverage. That approach respected the demand for equitable access to the transmission capacities for all European states and has been consequently put into practice.

At the conference the way was paved to allow for the implementation of two T-DAB multiplexes at each point throughout the planning area. As a rule, one frequency block located in the VHF range and one in the L-band could be utilized. There were countries which due to certain national restrictions could not use either of the two frequency ranges. They were assigned two coverages nonetheless accommodating both coverages then in the same available band.

According to the propagation conditions of electromagnetic waves the VHF range is more suited to serve large areas. In L-band the wave propagation conditions suggest a coverage of areas with limited size like urban areas. Before the Wiesbaden conference a large number of investigations were carried out to derive an order or magnitude for the amount of spectrum that would be required to satisfy probable planning scenarios.

In contrast to analogue transmission systems T-DAB offers the possibility to build single frequency networks, i.e. all transmitters which are to serve a particular coverage area are using the same frequency. This feature is fully taken into account by the Wiesbaden allotment planning approach. Nevertheless, it does not include any statements about actual network implementation issues. Neither the number or location of the transmitters of a network nor their characteristic features had to be specified in advance. This is entirely left to the network provider during implementation as long as adjacent T-DAB network are not affected.

Even though T-DAB allows the utilization of single frequency technology this does not mean that geographically adjacent networks which are broadcasting different content can make use of the same frequency. Moreover, depending of the topographic and morphographic conditions co-block areas have to be separated by a sufficiently large distance.

This important fact lead to many investigations already before the Wiesbaden conference in order to assess the mutual compatibility of different T-DAB networks. As a result, re-use distances for co-block usage both for band III and the L-band were derived. In the VHF range a value of 81 km has been agreed upon for propagation above land while in L-band 61 km were found to be adequate. Sea propagation conditions will increase the re-use distance dramatically [CEP95].

The basis for the determination of the re-use distances is the concept of reference networks. They can be used to estimate the interference a T-DAB single frequency network is likely to produce outside its proper coverage area (see section 5.1.2). This way it is possible to decide if the protection needs of other T-DAB networks are violated. Clearly, this investigation has to be carried out with respect to other services like DVB-T and non-broadcasting applications as well.

The starting-point of the frequency assignment of Wiesbaden was a set of coverage areas which were submitted by the participating European administrations. All areas had to be provided in terms of closed polygons. Figures 7.1 and 7.2 show a set of corresponding coverage areas which have been generated on the basis of the information given in [CEP95]. A total of 278 polygons for VHF coverage areas and 476 L-band areas were delivered as input to the conference.

The generation of the frequency plan at the Wiesbaden conference had to face severe restrictions with respect to the accessibility to spectrum in parts of the planning area. It was intended to use the TV-channels 11 and 12 in band III as the core frequency resource to be used for the first layer while the second layer was to be established with the help of L-band frequencies.

In some parts of Europe, like for example Scandinavia, channel 13 could be utilized, too. In other regions, for example along the German-Polish

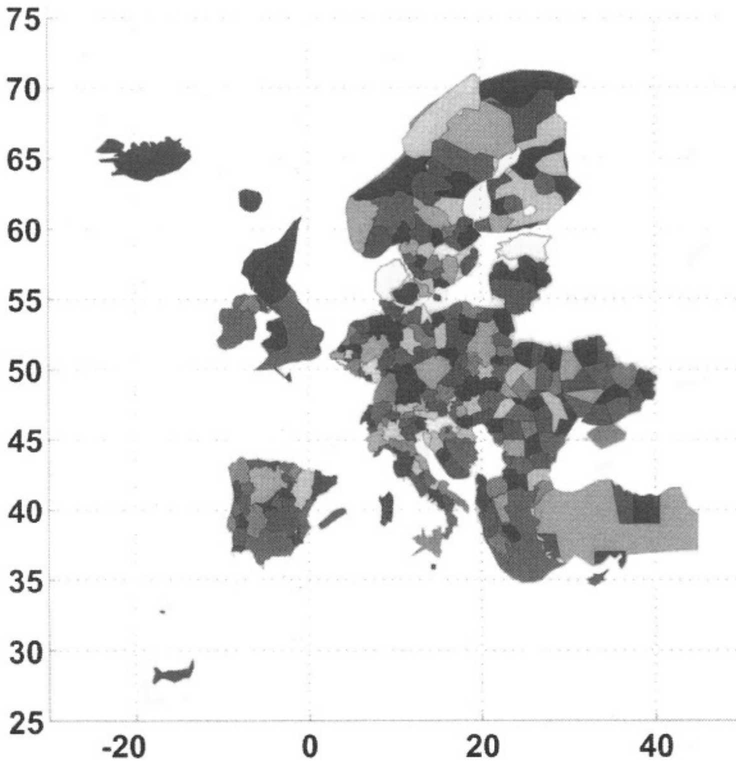

Figure 7.1: Result of a VHF spectrum demand estimation for a set of polygons adapted from [CEP95]. Identical gray shades of the areas indicate the assignment of the same T-DAB block.

border, even channel 12 could not used without creating problems. This particular case was connected to high power analogue TV stations in Poland. The Polish administration, however, was not willing to shut them down or use other channels in order to facilitate the introduction of T-DAB in Europe.

Apart from these ordinary restrictions there were also some extreme cases. In France VHF frequencies could not be used at all. Therefore both T-DAB layers were based on L-band blocks. Exactly the inverse situation was encountered in Great Britain and Scandinavia where L-band frequencies were not at disposal. Furthermore, other radio services had to be protected in several parts of Europe, too.

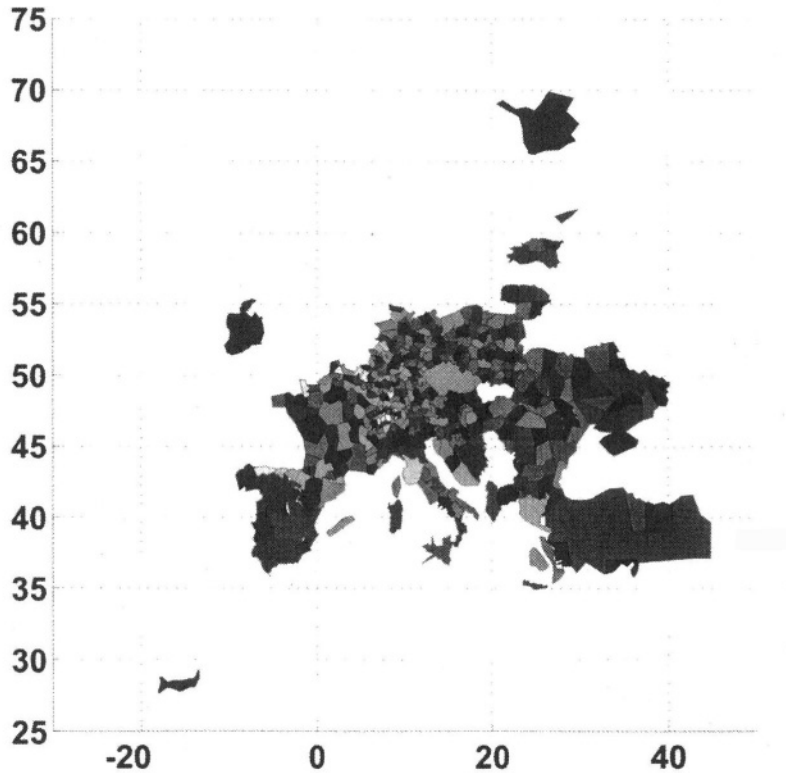

Figure 7.2: Result of an L-band spectrum demand estimation for a set of polygons adapted from [CEP95]. Identical gray shades of the areas indicate the assignment of the same T-DAB block.

In view of these severe and not uniform constraints it is obvious that the generation of the frequency plan was not an easy task to accomplish. According to the definition given at the beginning of this chapter it falls into the class of restricted frequency assignment problems to be dealt with later. However, it is worthwhile to carry out a spectrum demand estimation on the basis of the set of polygons depicted in figures 7.1 and 7.2 in order to classify the quality of the agreed Wiesbaden plan.

As described in section 5 the problem of estimating the T-DAB block demand can be mapped to a corresponding simple graph coloring problem. A solution can be found by applying standard graph coloring algorithms (see also [OLe98]). By employing the methods presented in section 5.1 the spectrum demand can be estimated both for a wide-area coverage in VHF and in L-band without problems.

It has to be emphasized that such an estimation follows very strict assumptions. The methods describe in section 5.1 rest on the constraint that any violation of the re-use distance between two coverage areas does not allow simultaneous co-block usage. Even the slightest transgression of the re-use distance criterion is not accepted. This is in blatant contrast to the Wiesbaden agreement where substantial violations of the re-use distance have been cleared off by bilateral negotiations between the affected administrations.

The first step of the spectrum demand estimation is to calculate the effective distances between all pairs of polygons shown in figures 7.1 and 7.2, respectively, according to the principles laid out in section 5.1.3. From this information the associated adjacency matrices are build. This is all the input required to apply an appropriate graph coloring algorithm. In the VHF range the result is that 9 T-DAB blocks are needed while in L-band a total of 17 blocks is necessary to comply with the adjacency constraints. In VHF the demand corresponds to a little bit more than two TV channels of 7 MHz bandwidth.

The question arises how to rate these results. To find an answer to that question it is helpful to calculate the maximum clique number ω_{max} of the associated graphs on the basis of the algorithm given in [Car90]. For the VHF problem a clique number is found to be $\omega_{\text{max}} = 9$ while in the L-band case the value for ω_{max} reads $\omega_{\text{max}} = 16$. The first result confirms that the graph coloring algorithm provides the optimal solution to that special problem. It is mathematically not possible to find a solution with less colors.

In the latter case the situation is different. In L-band only solutions can be found which require one block more than the maximum clique number. Nothing can be concluded about the question if a solution with only 16 colors exists. This would require further more sophisticated investigations. However, it is very likely that such a solution cannot be realized.

In both cases the calculated spectrum demand to establish a frequency plan which is assumed to be free of interference is significantly higher than the amount of spectrum available at Wiesbaden. Hence, it becomes clear that in order to arrive at an agreed frequency plan compromises had to be found. This lead to the aforementioned negotiations between administrations to find solutions for the critical cases.

The reason why most of the formal incompatibilities of the Wiesbaden plan could be cleared off is related to the fact that many areas whose effective distance was less than the re-use distance were topographically decoupled. Examples thereto can be found along the border between Austria and Germany where the high mountains of the Alps justified the decoupling assumption.

Since the creation of the frequency plan, i.e. the calculation of the mutual interference between allotment areas was based on the former ITU-R recommendation P.370 no topography was taken into account at all. This leaves an additional degree of freedom when the final decision has to be taken on whether two areas can use the same T-DAB block or not by checking their compatibility individually by employing terrain based wave propagation methods.

7.2 Mapping of FM to T-DAB

T-DAB is supposed to substitute FM radio as the standard sound broadcasting system in the intermediate future. This statement is valid at least for Europe according to the opinions expressed by network providers, receiver manufactures and regulators. The data transmission capacity that has been assigned to each participating country during the Wiesbaden conference in 1995 allows in principle to provide two T-DAB multiplexes at each point within the national boundaries [CEP95]. So, a total of 12 to 16 different programmes can be offered depending on the data rates chosen for each of them.

When having a look at a typical large city in Europe it becomes obvious that currently about 30 or even a lot more different FM stereo programmes can be received there without problems. If the customers are to be convinced that T-DAB offers a real benefit for them, then it seems that they have to be offered at least the same broad variety of programmes they are used to get from FM radio at the moment. A reception quality similar to a CD and the provision of additional non-audio services alone certainly will not make people to spent money on a new and so far still expensive technology. The existing FM radio reception quality still seems to be good enough to satisfy most customer's current needs.

The above mentioned 12 to 16 programmes the frequency assignment of the Wiesbaden plan allows for can be provided in principle at every single location throughout the defined allotment areas. This includes the very centers of big cities like London, Paris or Berlin as well as rural regions for example the Black Forest in Germany. Since in the future certainly economic issues are getting more and more important the question arises if the planning principles need to be refined.

Many broadcasters, in particular the private companies, will be very reluctant to generally accept the wide-area coverage target as the only possibility. Obviously, they will prefer to offer their services in economically interesting areas. This areas, however, will be precisely defined and certainly limited in size. Even more, if T-DAB will be an economic success demand for additional coverages in the same areas will arise.

Against this background, it can be expected that sooner or later the necessity to abandon or at least soften the principle of wide-area coverage target will become evident. Any planning approaches therefore must be able to cope with inhomogeneous distributions of allotment areas and point out their consequences for the correspondingly derived frequency plans.

One first step in that direction is an investigation under what conditions it would be possible to map the entire actual FM radio programme landscape to T-DAB. It has to noted that this means not only to estimate the spectrum demand to broadcast the same programmes on T-DAB but also to map the individual coverage area of each program.

Furthermore, it is very interesting with respect to the future perspectives of T-DAB to study the spectrum demand arising from the introduction of additional new T-DAB services on top of the mapped FM programmes. If there is no future development potential for T-DAB then certainly its introduction will fail.

7.2.1 Model for the Mapping of FM to T-DAB

The development of an appropriate model for the mapping of the actual FM radio landscape to T-DAB needs further detailed specification. Currently, the number of programmes that can be received in stereo quality varies across a country to a large extend. In typical dense urban areas the number of receivable programmes is significantly higher than in most rural regions.

Along the border lines between two countries usually the programmes of both states can be received. The same situation can be found between two independent national areas like the German Bundesländer where different media political concepts are followed. Thus, in practice it is very common that a programme is listened to or watched even beyond its generic coverage area. Such an overspill seems to be very welcome by the customers since it increases quite naturally the selection of available programmes. Clearly, any realistic spectrum demand estimation which claims to take into account the customers habits should incorporate overspill coverage in a reasonable way.

At the Wiesbaden conference in 1995 it has been assumed that within each of the coverage polygons the transmission capacity corresponding to one T-DAB multiplex is to be provided. To this end, one T-DAB frequency block must be employed. The corresponding data capacity allows to broadcast between six and eight stereo audio programmes at the same time. Thus by definition, these six programmes can be received within the same coverage area.

Is is evident that such a structure is not found in FM broadcasting. In order to broadcast one single programme a bandwidth of about 150 kHz is needed. The basic problem to be solved when trying to map the FM scenario to T-DAB thus is the different partitioning of the spectrum. The naive idea to map one FM programme to one T-DAB multiplex is not reasonable. In fact, such a correlation would imply an increase of the number of receivable programmes by a factor of six or more. Even if such a demand for programmes would be formulated in the future it clearly does not reflect the actual FM landscape.

Therefore, several individual FM programmes have to be combined to form the content of a T-DAB multiplex. To this end, the coverage area of each of the FM programmes has to be calculated and approximated by an appropriate polygon. In the very unlikely case that the coverage areas of all FM programmes in one multiplex are identical the coverage area of the multiplex quite naturally results to be the same as well. Since such a situation will certainly be the exception from the rule it has to be specified how the coverage areas of the individual FM programmes have to be combined.

An evident option is to define the union of the coverage areas of all programmes of a multiplex as the multiplex coverage area. It should be noted that this choice implies significant overspill for most of the involved FM programs. The question if this is a reasonable approach clearly has to be addressed and discussed as well.

When trying to figure out the spectrum demand for the mapping of FM radio to T-DAB by combining individual programmes to multiplexes it has to be decided according to what principles programmes can be put into the same multiplex. A first criterion could be to create multiplexes hosting only programmes of public or private broadcasters. Public broadcasters in general offer several programmes which are meant to be broadcast within large areas trying to serve all locations therein. In most cases there are enough public programmes to fill a T-DAB multiplex.

The situation is quite different with respect to private broadcasters. Usually, there are very few broadcasting companies that would like to provide several programmes within exactly the same region. Just the small radio stations are serving distinct proper coverage areas. There is typically a lot of overlap between coverage areas of different programmes, but they are seldom identical. In principle, the transition from analogue to digital radio could be seen as a good chance to re-design the coverage areas of all private broadcasters. If they all would be happy with that is in question.

In view of the necessity to analyze the future spectrum demand for T-DAB a detailed study was carried on behalf of WorldDAB[1] and presented at the CEPT radio conference in Lisbon in 2000. By means of concentrating on several representative regions in Europe a method has been proposed which allows for the accomplishment of realistic spectrum demand analyses [Wor00] (see also [Beu01a]). It followed the above mentioned ideas of combining FM programmes to fit into multiplexes. In detail, the environments of Paris, London, Rome, Stuttgart and Malmoe have been considered. The dimensions of the geographic regions to be investigated were different in all cases whereas the German region was the largest.

[1]WorldDAB is an international organisation that tries to promote T-DAB around the globe. Its activities include investigations concerning regulartory, technical and economical aspects relating to T-DAB.

The selection of examples should cover as much as possible different aspects that are important for the spectrum demand investigation. In France so far only L-band T-DAB blocks can be used while in Great Britain and Sweden only the VHF range can be employed for T-DAB. In addition, the southern part of Sweden is surrounded by sea so that effects due to wave propagation above water need to be taken care of.

The purpose of the study was not only to estimate the amount of spectrum which would be required for a complete mapping of the actual FM radio landscape to T-DAB. Indeed, a forecast with respect of the future spectrum demand should be attempted, too. Clearly, this implied to make comprehensible assumptions with regard to the potential progress of the T-DAB system during the next ten years.

It proved to be very difficult to find any reliable information thereto on which a reasonable forecast could be based. There are still too many open questions concerning the future market penetration of T-DAB which last but not least is determined also by the available spectrum. Therefore, all estimates presented here have to be considered as first vague indications only. However, the general assumption was that T-DAB is going to be a success thereby leading to an increasing demand for transmission capacity on the part of the broadcasters.

As described above, a set of polygons has been generated each of them representing the coverage area of one single FM radio program. In the first step, it has been assumed that six FM programmes should be combined to build one T-DAB multiplex. This decision is obvious because the total data capacity of a T-DAB multiplex allows the transmission of six audio signals of 192 kBits/sec. This corresponds to stereo programmes in near CD quality. Clearly, the number of programmes within a multiplex could be varied, too. A figure between four and eight programmes seems to be reasonable. It has to be emphasized, however, that accepting this additional degree of freedom will further increase the complexity of the problem.

The investigation started by bundling six programmes of the present set of polygons to represent T-DAB multiplexes, respectively. The associated coverage area of the multiplex has been assumed to be given by the union of the individual programme coverage areas. Once this has been done the distances between all multiplex coverage areas have to be determined. In order to reduce the computational time the mutual distance between the coverage areas of all FM programmes are calculated before the simulation has been started. This way the distance calculation for multiplexes can be reduced to checking the involved programme area distances.

So far, T-DAB can be implemented by employing frequencies from two distinct spectral ranges, namely band III in the VHF range and the L-band. According to the wave propagation conditions between 176 MHz and 230 MHz the VHF range is very well suited to be used for the coverage of wide areas. Extension of 200 km $\times$ 200 km are quite common, sometimes even more can be realized. The L-band is to be utilized in first place to provide services on a regional or local basis.

For the mapping of FM audio programmes to T-DAB there is no a priori predefinition which programmes are to be broadcast in VHF or L-band. An obvious strategy would be to accommodate those multiplexes whose associated coverage area exceeds a maximum extension in the VHF range. All “smaller” multiplexes should use an T-DAB block located in L-band. The critical extension has to be chosen properly. Against that background, the maximal extension of a multiplex is an important parameter which has to be calculated as well.

After having determined all mutual multiplex distances the corresponding adjacency matrix can be filled. Clearly, VHF and L-band multiplexes do not interact at all. Therefore, they need not obey a minimal separation distance. The elements of the adjacency matrix are either zero or one according to if two multiplexes can use the same frequency block or not. This way, all input information has been provided to allow the application of graph theoretical methods to derive the spectrum demand for the current set of multiplexes (see section 5.2). Figure 7.3 illuminates the process.

Without doubt, an arbitrary distribution of programmes onto multiplexes can give rise to an extremely high spectrum demand. It is necessary to determine an optimal constitution of the set of multiplexes. This task can be achieved by embedding the determination of the block demand based on graph coloring into a stochastic optimization framework. The recourse to stochastic optimization is needful since once again a very complex combinatorial problem has to be faced. The brute force approach of checking all possible permutations of distributing the programmes onto the multiplexes cannot be accomplished in practice within

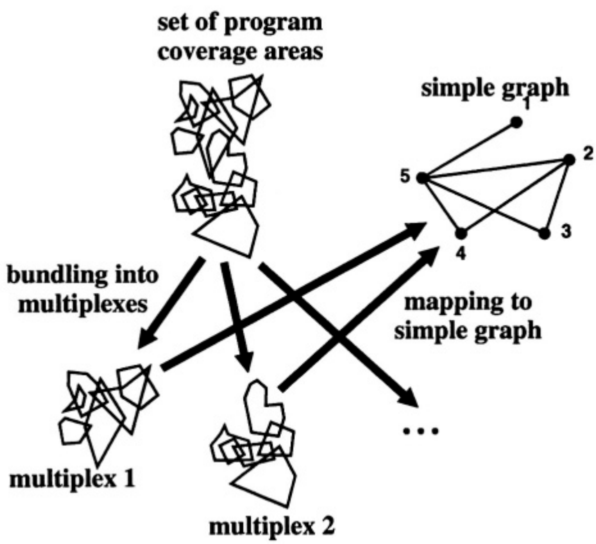

Figure 7.3: Bundling of programmes into multiplexes.

a finite period of time.

If for the time being 120 FM programmes have to be dealt with which are to be distributed onto multiplexes in order to find the minimum number of required T-DAB blocks, then there are $120! \approx 7 \times 10^{198}$ different permutations which is an astronomically large number. It is evident that any brute force method trying to check all these possibilities must fail. This becomes even more clear in view of the number of programmes typically dealt with which is by far larger than 120.

Every complete distribution of the set of programmes onto multiplexes is called a configuration in the context of stochastic optimization. The quality of a configuration corresponds to the minimal number of T-DAB blocks it requires. According to the discussion in section A.1 the application of an adequate optimization algorithm presumes the existence of a defined neighborhood relation between configurations. Two configurations are neighbors if one can be converted into the other by changing its characteristic parameters only slightly.

In the present case such a relation can be established by exchanging only one arbitrarily chosen programme of one multiplex with another also arbitrarily chosen programme of another multiplex. Such a modification does not affect the global properties of the configuration, i.e. the set of multiplexes as a whole. Furthermore, the associated spectrum demand of the configuration will not change significantly, if it changes at all.

However, the latter fact can lead to problems for the stochastic optimization simulation. Especially in those cases where no change in quality is experienced secondary criteria have to be applied as well. The number of required blocks alone might not be sufficient to distinguish between good and bad configurations..

For the present task it is desirable to arrive at as compact multiplex coverage areas as possible. However, following the construction principles this needs not be the case. It is possible that multiplexes are generated giving rise to a small spectrum demand which might consist of six independent non-overlapping programme coverage areas. Seen from a purely practical point of view concerning the implementation of T-DAB networks such solutions have to be sorted out. This suggests to classify those configurations consisting of a large number of multiplexes having compact coverage areas as better configurations.

The question is, however, how this constraint can be cast into a reasonable secondary criterion. One way would be to calculate the mean maximum extension of all multiplexes being part of a particular configuration. In case the spectrum demand required by two sets of multiplexes is identical that configuration whose average extension is smaller is considered to possess a higher quality. Such a secondary criterion quite naturally leads to configurations being built from compact multiplexes. Figure 7.4 shows a flow chart indicating the algorithm that has been utilized for the mapping study.

7.2.2 T-DAB Block Demand in South-West-Germany

The federal structure of Germany is dominating the German radio landscape. This is one of the German peculiarities. With respect to T-DAB planning each of the German Bundesländer has to be considered as completely independent having proper demands and rights. For this reason the boundary between for example the two Bundesländer Baden-Württemberg and Bayern has to be dealt with on the same level as the boundary between France and Baden-Württemberg. Therefore, any spectrum demand estimation for Baden-Württemberg does only make

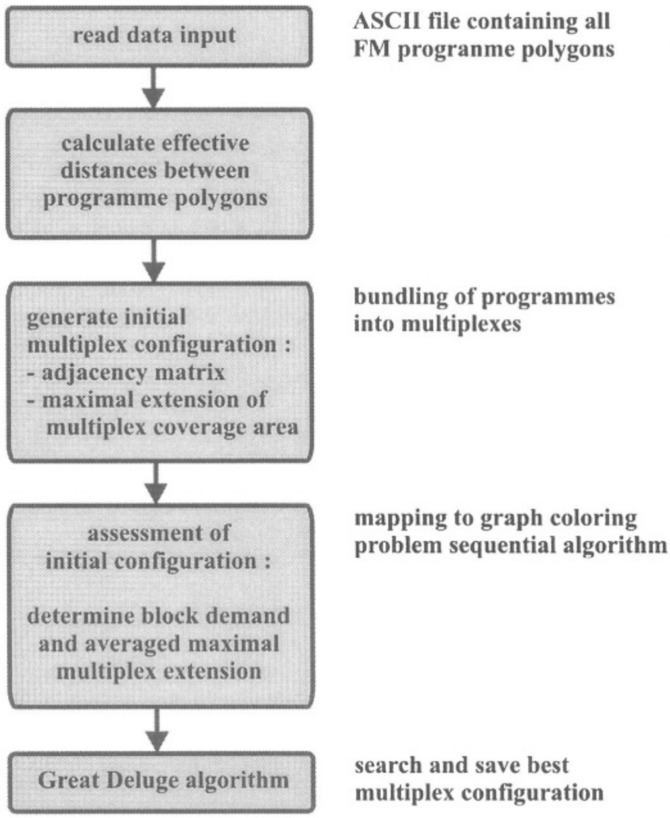

Figure 7.4: Schematic representation of the solution strategy used for the WorldDAB study.

sense if the adjacent Bundesländer Bayern, Hessen, Rheinland-Pfalz and Saarland are taken into account as well. Clearly, coverage areas in the neighboring countries France, Switzerland and Austria have to be considered, too. Due to the lack of reliable information, however, non-german programmes have not been considered for the study.

The investigation covered an area in Germany which extended from the most south-eastern part of Bavaria up to Cologne. Its maximal extension thus reached 650 km. Thereby, it comprises several German Bundesländer. Figure 7.5 shows a polygon drawn into a map of Germany characterizing the boundaries of the planning area.

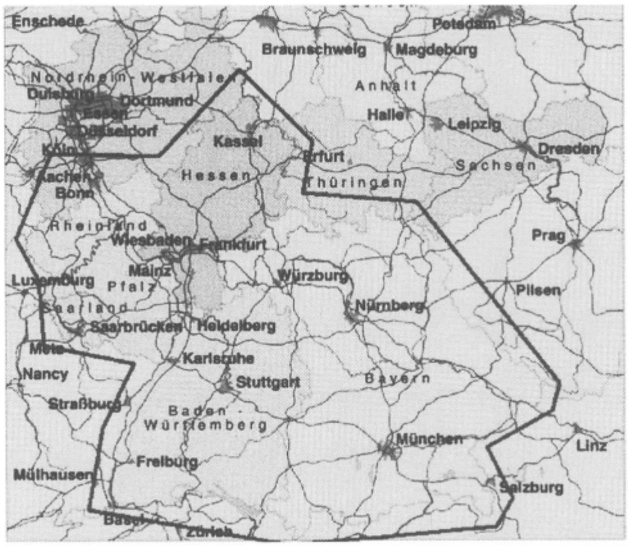

Figure 7.5: Investigated region in south-west Germany.

As mentioned, the objective of the FM to T-DAB mapping study was twofold. First, the spectrum demand resulting from the mapping of the existing radio landscape to T-DAB was to be estimated. This included first of all a very detailed stocktaking of the actual FM radio broadcasting situation throughout the considered planning area. The second objective was to make a reasonable forecast for the future T-DAB demand and subsequently determine the associated demand for T-DAB blocks.

Existing Analogue FM Services

In each of the Bundesländer falling into the planning area depicted in figure 7.5 there exists a public broadcasting company whose task is to provide a wide-area coverage with a variety of different radio programmes. Some of the programmes are being broadcast only in parts of the Bundesländer on a regional basis which is often determined by the occurrences of cultural or administrative communities. Furthermore, there are also local broadcasting windows in the sense that a special service for several urban areas is offered during a limited period of time during daytime.

Moreover, there are two German public broadcasters which are operating on a national basis. i.e. they are not restricted to respect the boundaries of the German Bundesländer. They offer programmes which are not adapted to any special regional conditions. These programmes can be received across the total planning area of figure 7.5.

Besides, there are several private broadcasting companies which are supposed to cover parts of the Bundesländer. The variety of FM radio programmes encompasses also a lot of local stations that offer very specialized programmes primarily focused onto local content. The radio landscape is rounded off by the non-commercial radio programmes for example in Stuttgart and Freiburg and in addition the stations of the foreign military services. This leads to a reception situation in large cities like Stuttgart, Frankfurt or München where up to 35 stereo programmes can be received in rather good quality.

The total number which was derived when counting all existing programmes is 139. Clearly, this can only be considered as a rough estimate since it was not possible to obtain the latest up to date information from all affected Bundesländer. In order to carry out the intended mapping of FM services to DAB for all these programmes their individual coverage area has been approximated by a polygon.

In order to bundle 139 FM programmes into multiplexes a total of 24 multiplexes is created. Under the assumption that each multiplex is to contain six programmes there are five free programme slots since $24 \times 6 = 144$. These free capacity can be used later for the introduction of additional programmes.

Estimated demands for services

The forecast for future programmes proved to be very complicated. The reason was that basically no reliable information could be obtained on which a reasonable prediction could have been based. There were neither any official statements of politicians in charge of media political issues nor did any profound market surveys exist which could be used to assess the potential commercial success of T-DAB.

Therefore, simple assumptions had to be made like for example that the demand for T-DAB services would be steadily increasing. Clearly, such ideas can be completely wrong. However, for the introduction of T-DAB it is crucial to know what problems are to be expected in terms of spectrum demand if certain expansion scenarios are to be satisfied in the future. If it would turn out that the spectrum demand for a reasonable market penetration goes far beyond the available spectrum the question certainly would arise if T-DAB really is such a promising radio system capable to cope with the future developments.

In order to model the future demand for T-DAB services new coverage areas were defined in addition to the already existing programmes. Since it can be assumed that in the future economic factors will play an even more important role for the success of a new system the new coverage areas were supposed to cover primarily urban areas and traffic paths. Thus, the creation of polygons for these new services was guided by an examination of the population density within the planning area.

Once a potential new coverage area had been identified and approximated by a polygon it was checked how many people are living within the boundaries of that area. According to the result copies of this polygon were added to the set of programme coverage areas. If more than one million people were counted a total of six identical polygons was taken into account. In case the number of inhabitants was between 500000 and 1000000 three copies were used, for the interval from 300000 to 500000 it was two copies. All polygons encompassing less than 500000 people were taken into account only once. This procedure finally lead to a total of 300 radio programmes for the planning area depicted in figure 7.5.

Required frequency blocks to satisfy the service demands

Two sets of programme coverage areas have been used as input to answer the question about the spectrum demand for a mapping of FM programmes to T-DAB. As a secondary criterion it has been assumed that if a multiplex has a coverage area which is less than 100 km it will automatically be assigned a L-band frequency.

For the 24 multiplexes which represent the mapping of the actual FM landscape to T-DAB a demand of 14 blocks in VHF has been calculated. No multiplex was generated in the context of the optimal solution which had an extension that would force to assign an L-band block. The spectrum demand thus is about 3.5 TV channels having a bandwidth of 7 MHz.

The forecast scenario being based on 300 programmes, i.e. 50 multiplexes gave rise to a total demand of 22 T-DAB blocks. These consisted of 20 VHF T-DAB blocks and only two from L-band. Twenty T-DAB blocks correspond to five 7MHz TV channels in the VHF range. This is significantly more than the two channels that are currently considered to represent the core frequency range which is to be dedicated to satisfy the T-DAB demands.

Apart from the spectrum demand figures the investigations showed that the details of the polygon shapes crucially influence the results. Any changes of the polygons might lead to different solutions. This should be considered as an additional degree of freedom in terms of reducing the number of required T-DAB blocks derived for a particular scenario.

The central question when looking at the calculated spectrum demand, however, is how these results must be interpreted. The block figures alone do not give a clear indication since they have to be seen in connection with different coverage targets. Therefore, it is essential to derive some kind of measure which allows to assess the efficiency of the spectrum usage. Unfortunately, the definition of “efficient spectrum usage” is far from being obvious.

In order to get some idea about the quality of the calculated frequency plans a comparison between the two mapping scenarios and existing situations like the actual FM landscape and the Wiesbaden plan might help. Table 7.1 summarizes all relevant information. The values of the presented parameters all refer to the planning area depicted in figure 7.5. The 24.5 MHz for the Wiesbaden case include all T-DAB blocks that are used somewhere within the planning area of figure 7.5, i.e. all blocks in channel 12, the nine L-band blocks and block 6C used in Saarland.

Table 7.1 sets out three different possible definitions of efficient spectrum usage. They lead to different assessments of the four planning approaches. The first difference between the planning examples is that obviously they consume different amounts of spectrum. It is nevertheless remarkable that the Wiesbaden plan and the mapping of the actual FM landscape occupy the same bandwidth. However, the forecast utilizes nearly double the amount of spectrum than does the existing FM landscape.

	FM	Wiesbaden	direct mapping	forecast
occupied bandwidth B [MHz]	20.5	24.5	24.5	38.5
total number Z_{ges} of receivable programmes	139	76	144	300
Z_{ges}/B [1/MHz]	6.78	3.10	5.88	7.79
average number $\overline{Z}$ of receivable programmes	19.5	12	22.2	27.5
$\overline{Z}/B$ [1/MHz]	0.95	0.49	0.90	0.71
average number $\overline{Z}_{UKW}$ of receivable FM equivalents	19.5	16	29.6	36.6
$\overline{Z}_{UKW}/B$ [1/MHz]	0.95	0.65	1.21	0.95

Table 7.1: Comparision of different T-DAB coverage scenarios. Direct mapping refers to the actual FM situation and forecast covers the extension to 300 programmes.

Sure enough, this figure alone is misleading as can be easily seen from the results of the three different definitions of efficiency. If just the total number of receivable programmes is related to the bandwidth then the forecast case is the best. This changes for the second definition. The forecast falls back beyond the FM landscape which seems to be optimal then.

However, also this conclusion has to be taken with care because the absolute figures put it into perspective again. Both direct mapping and forecast provide a larger mean number of programmes and in particular in the forecast case there is more than twice the number of receivable programmes than for the FM landscape. Locally, i.e. in urban areas the differences are even larger.

The figures clearly show that it is difficult to find a common basis on which all planning scenarios can be assessed under the same premises. This becomes even more obvious when considering the third definition of efficient spectrum usage. Actually, the direct comparison between FM radio and T-DAB as done by the first two definitions is not really admissible. If a T-DAB multiplex contains six programmes only, the quality of services is far better than anything FM radio can offer. In order to compare the quality of T-DAB to FM services the term "FM equivalent" has been coined. This means that if eight programmes are broadcast within one T-DAB multiplex the corresponding audio quality is about the same as in the FM radio case due to smaller data capacity available for each programme.

A comparison based on the FM equivalent approach leads to a completely new picture. The total number of receivable programmes goes up to 400 in the forecast scenario. Correspondingly, the average number of receivable programmes, i.e. the FM equivalents goes up. Amazingly, the direct mapping beats the FM landscape scenario in any aspect. For the forecast the situation becomes more favorable for T-DAB as well.

Furthermore, there is one principal intrinsic drawback of the presented mapping scenarios. The starting point of the investigations was the actual FM landscape in all details. This covers both the number of programmes as well as the shape of their coverage areas. As is well known, this situation is far from being optimal. Thus, any expansion thereof must cope with all the immanent problems. Finally, the Wiesbaden plan seems to be suboptimal in any respect. This, however, comes as no surprise in view of the very limited amount of available spectrum.

7.3 Spectrum Demand for DVB-T in Germany

Since 1961 the spectrum utilization in the VHF and the UHF range by broadcasting is governed by the Stockholm agreement [ITU61]. In face of the developments in the course of the transition to the digital era it comes as no surprise that this plan can no longer cope with the current and future requirements efficient spectrum management must meet. Before the international radio conference 2000 in Berlin ("Internationale Funkaustellung (IFA)") the initiative digital radio (IDR) on behalf of the German government formulated intermediate and long term expansion scenarios for T-DAB and DVB-T in Germany. Since then, these statements have been commonly considered as reasonable and feasible guidelines for the German transition to digital terrestrial broadcasting[IDR00].

Amongst other points they stipulate the far term target to provide Germany with six DVB-T multiplexes on a nation-wide coverage basis. First investigations with respect to the corresponding channel demand have been carried out by the Institut für Rundfunktechnik in München in 1999 [IRT99]. They showed very clearly that the different operation modi of DVB-T would require to obey very different re-use distances for co-channel usage. As point of reference a mean value of 88 km for the UHF range has been chosen. Several scenarios which consisted of nation-wide, regional and local coverages were considered then.

All investigations were based on a fixed set of coverage areas whose shapes have not been changed during the study. The results obtained by similar investigations for T-DAB [Wor00] lead to the conclusion that the crucial factor for an efficient usage of the available spectrum is the explicit definition of the coverage area shapes. Consequently, in [Beu00] coverage scenarios of different constitution have been dealt with in detail. The re-use distance has been chosen to coincide with the 88 km introduced in the IRT study [IRT99]. The determination of the channel distributions and the thereby resulting spectrum demand have been carried out with the help of the deterministic graph coloring algorithms presented in section 5.2. Some of the results will be discussed below.

7.3.1 Planning Scenarios

The first example to be presented corresponds to the DVB-T scenario 1 from the IRT study from 1999. Since the exact coordinates of the vertices of the polygons have not been known at the time being they have been extracted by hand from figure 10.11 of [IRT99]. Only areas in Germany have been taken into consideration. Figure 7.6 gives an overview about the polygons.

This first example deals with 48 areas which have been designed by unification of several L-band polygons from the Wiesbaden agreement. The second set of polygons is derived from the first one simply by deleting four very small areas, namely Hamburg, Bremen, Berlin and Helgoland. This leaves the set which can be seen in figure 7.7. The third and final scenario has been generated by merging several polygons of the second example 7.7 in order to obtain a smaller number of nevertheless larger areas. This situation is shown in figure 7.8.

For each set of allotment areas three different scenarios have been investigated. By making copies of the polygons cases with two and six

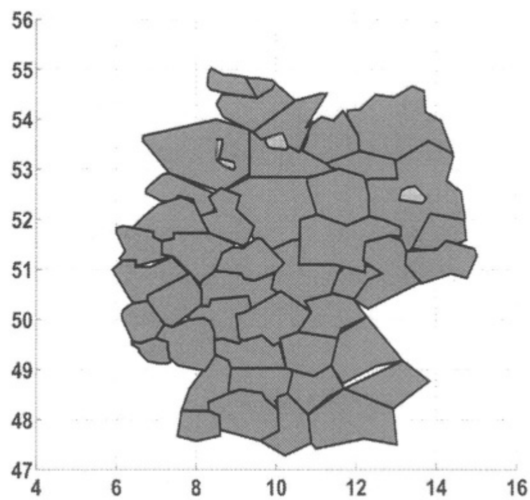

Figure 7.6: Polygons for the first DVB-T scenario.

identical coverage layers have been created. Thus, a total of nine different frequency assignment problems has been prepared. For each of them the spectrum demand has been calculated by employing graph coloring methods. The results are summarized in table 7.2.

The standard approach for frequency assignment problems containing several identical layers assumes that an allowed distribution of channels onto the polygons and thus an associated spectrum demand can be obtained by multiplying the result of one layer by the number of layers. This certainly gives a result that is in full accordance with all constraints expressed in terms of the adjacency matrix. Implicitly, this means that the different coverage layers dealt with are completely independent from each other.

coverages	1	2	6
scenario 1	8	15	45
scenario 2	7	14	42
scenario 3	6	12	36

Table 7.2: Number of required channels for the different DVB-T planning scenarios.

It is well known, however, that better results can be achieved if all layers

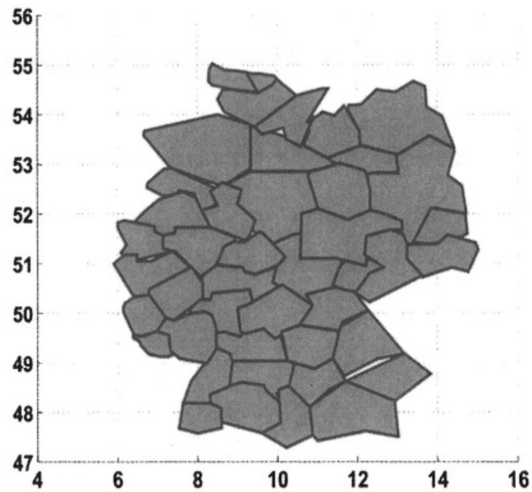

Figure 7.7: Polygons for the second DVB-T spectrum demand scenario. In contrast to figure 7.6 the polygons for Hamburg, Bremen, Berlin and Helgoland have not been taken into account.

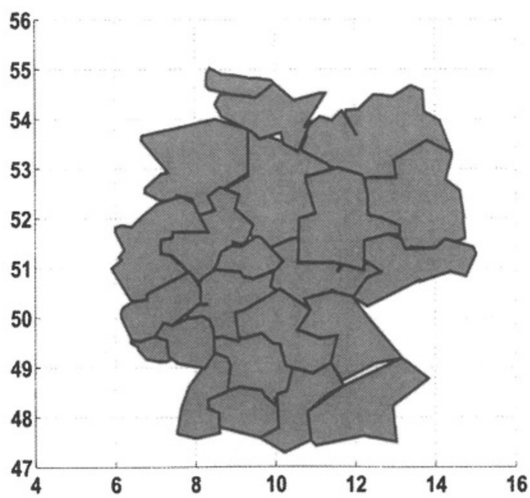

Figure 7.8: Polygons for the third spectrum demand scenario. This set of polygons has been derived by merging several polygons of figure 7.7.

coverages	1	2	6
scenario 1	7	14	42
scenario 2	7	14	42
scenario 3	6	12	36

Table 7.3: Clique numbers for the different DVB-T planning scenarios of table 7.2.

are treated at once. In particular, the figures for scenario 1 in table 7.2 support this claim. When going from one to two layers one channel less is needed and in the case of six layers even three channels are saved. In the following section this interesting phenomenon will be discussed further.

The question arises also here if it is possible to assess the results given in table 7.2 with regard to their quality. As has been discussed in section 5.2 the application of sequential graph coloring algorithms usually leads to very good results. A priori, however, it is not clear how large the quality of the calculated solutions differs from the quality of the optimal solution.

For that reason it is interesting to determine the maximum clique numbers associated [Car90]. This way, at least a lower bound for the number of required channels is found which can be used as a reference. All clique numbers for the scenarios of table 7.2 are shown in table 7.3. Obviously, the quality of the results derived on the basis of deterministic algorithms proves to be very good.

7.3.2 Multi-Layer Gain

The fact that a joint coloring of several layers might lead to better results is known as “multi layer gain”. It can be explained with the help of a very simple example without diving too deep into the mathematical basics of graph coloring.

Figure 7.9 sketches the example. It consists of two simple graphs one having four and the other five vertices. The vertices are labelled by characters. Each vertex has exactly two neighbors which consequently are not allowed to use the same channel. For one single layer two colors are needed in the case of four vertices whereas for the graph with five vertices three colors must be used. The assigned colors or channels are indicated by the attached numbers.

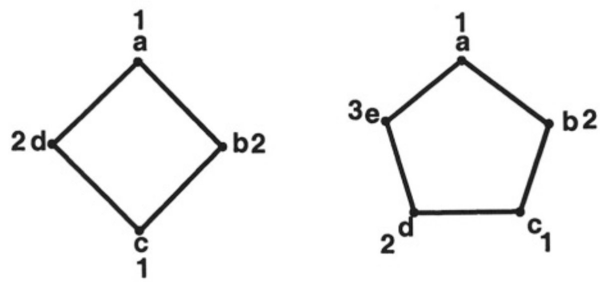

Figure 7.9: Two simple graphs representing a single coverage layer consisting of four and five allotment areas, respectively. The characters denote the vertex while the numbers represent the assigned colors.

Figure 7.10 shows a solution for the corresponding two layer scenario. Speaking in terms of allotment areas this would correspond to a duplication of the already existing allotments. Thus, a valid solution for this graph coloring problem is a duplication of the number of required colors. In the case of four vertices there is no other possibility. However, the structure of the graph having five vertices allows for a solution that needs five instead of six colors as obtained from the simple duplication scheme.

Compared to real world problems in the field of frequency assignment and spectrum demand estimation these examples indeed have to be considered as being trivial. However, if already under such simple and transparent conditions multi-layer gain effects can be observed it seems to be quite obvious that in the case of more complex problems such a degree of freedom can be expected just waiting to be exploited.

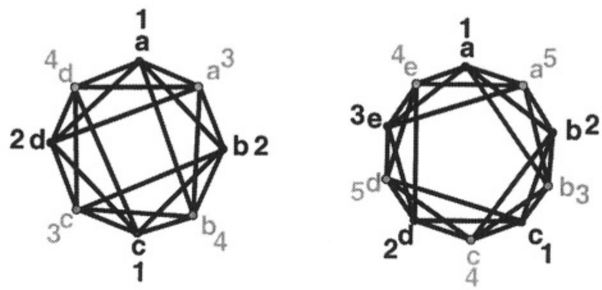

Figure 7.10: Two simple graphs representing two coverage layers. The characters denote the vertex while the numbers represent the assigned colors. The two layers are differentiated by black and gray.

7.4 Spectrum Demand for DVB-T from a European Perspective

The investigations presented so far did not take into consideration the central geographic location of Germany within Europe. The conclusion is evident that any analysis that is concerned about a geographically isolated Germany only will very likely underestimate the actual spectrum demand because further constraints are imposed by the presence of other countries. Therefore, a scenario is to be discussed now which explicitly takes this into account.

None of the DVB-T planning examples in the preceding sections made any reference to a specific spectrum range. However, it is clear that such a large number of channels is only disposable in UHF if at all. On the other hand, in particular public broadcasters in Germany are very much in favor of at least one nation-wide coverage with DVB-T in the VHF range. The reason is that they have been operating extended transmitter networks in band III for decades. So, they can refer to an existing technical infrastructure.

Since the wave propagation conditions in VHF are much more favorable than in UHF it is possible to establish large coverage areas. For that reason, the Wiesbaden polygons for T-DAB in VHF have been employed for the studies presented in the sequel as examples for DVB-T allotment areas. Small areas like the allotments for Bremen, Helgoland, Hamburg and Berlin have not been included. All coverage areas dealt with is to be provided with one TV channel in order to allow the broadcasting of one DVB-T multiplex.

During the Wiesbaden conference there were no French allotments in VHF to be taken into account. The usage of band III for T-DAB was blocked then by existing other services in that frequency range. In order to round off the investigation here, polygons for France have been generated by merging an adequate number of French L-band polygons from Wiesbaden to represent one new allotment area.

The union of L-band areas was necessary to arrive at a reasonable size of the new allotments. It has to be emphasized, however, that the actual design of an area has not been governed by any French media political constraints. The only criterion to guide the generation process was to obtain some areas of reasonable dimensions. The figures 7.11 - 7.13 show the used allotment areas for Europe and in detail for Germany and France.

At this stage, it is necessary to stress the important fact that all allotment areas employed here for any DVB-T planning example have not been confirmed by national administrations nor have they been agreed upon yet. This is in particular valid for the British areas which certainly will not coincide with actual DVB-T requirements. They have been chosen entirely by the idea to define a representative set of polygons which taken as allotments cover all interesting planning aspects.

The re-use distance very strongly depends on the operation mode the DVB-T network [IRT99]. The interesting question to be answered is what is the influence of this parameter on the spectrum demand. Therefore, a sequence of calculations has been carried out by varying the re-use distance between 40 km and 150 km. The set of polygons has been the same for all the distances.

For each value of the re-use distance a channel distribution has been derived using the graph coloring algorithms discussed earlier. Three different ordering schemes for the graph coloring have been utilized thereto. Subsequently, the number of required channels has been determined. Apart from the total number of TV channels also the number of channels used in Germany only has been checked. Furthermore, as before the maximum clique number has been calculated as well. All results are summarized in table 7.4.

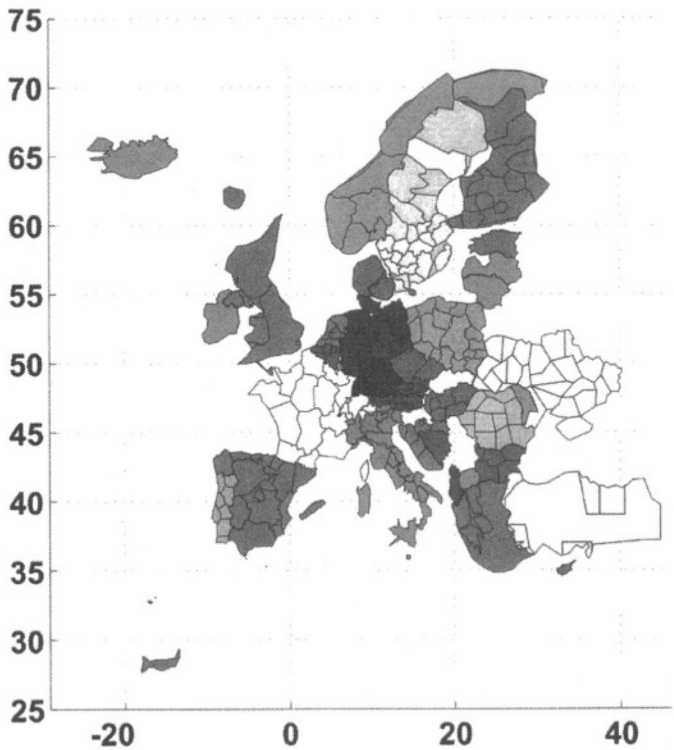

Figure 7.11: Employed DVB-T coverage areas for Europe in VHF.

A general conclusion from table 7.4 is that the channel demand increases with increasing re-use distance. Moreover, the results obtained seem to be very good as can be concluded from the calculated maximum clique numbers. The three different ways to order the vertex sequences lead to results that partially differ significantly. Method III seems to give the best results, however, this is not true all the time. As a consequence of the variation of the results, it is always necessary to make use of different algorithms in order to obtain reliable results.

There is another peculiarity in table 7.4 which should be noted here. All results have been derived by using the entire set of polygons at the same time. The results correspond to the minimum number of channels which are necessary to comply with the adjacency constraints across the whole of Europe. The figures for Germany in table 7.4 are determined by simply checking which of these channels are actually used somewhere

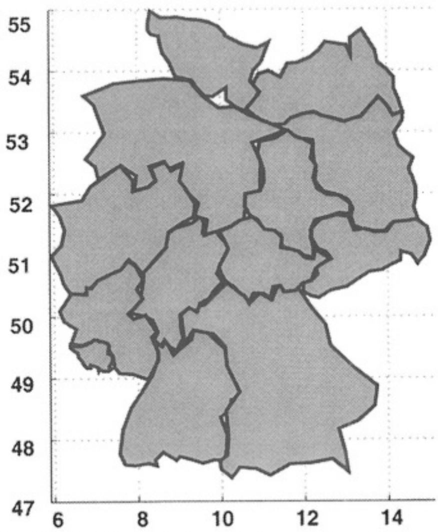

Figure 7.12: Employed DVB-T coverage areas for Germany in VHF.

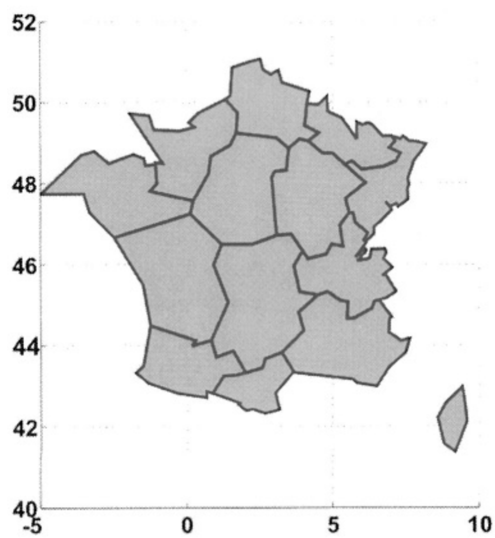

Figure 7.13: Emloyed DVB-T coverage areas for France in VHF.

in Germany.

As can be seen very easily, the German demand is always smaller than the European demand. Obviously, the total demand is not dominated by the German polygons. There must be areas in Europe which are in a sense more complex than the German situation. A closer look at the members of the maximum cliques reveals that the Baltic sea is the most problematic region. This is due to the wave propagation above cold sea which links areas which under land propagation conditions would be separated far enough.

Another interesting feature of the result is the fact that the channels assigned to areas in Germany are not necessarily contiguous, i.e. they do not form a sequence like 1, 2, 3, 4, 5, etc.. This means for example that if across Europe the channels 1 - 11 have been used it might turn out that in Germany only channels 1,2,3,4,6,7,9 and 10 are needed. For such a special solution TV channels 5 and 8 are not assigned to any area in Germany. However, it is very likely that even if they seem to be disposable they cannot be used due to usage in other countries near the German border. A situation like that is encountered in the case of a 85 km re-use distance. Virtually, this corresponds to an increased spectrum demand because the channels not used should be counted as well. In table 7.4 these type of solutions are indicated by a star.

The starting point for all the investigations in this chapter has been the question how many channels or blocks are needed for a more or less realistic DVB-T coverage of Europe taking into account different wave propagation mechanisms. Even if the results produced are strictly valid only for the special choice of allotment areas, the spectrum demand figures give a first indication what has to be expected. Different shapes and number of polygons quite certainly will lead to different results in detail. However, several independent studies have shown that the global amount of spectrum that is required in order to provide one DVB-T layer across Europe does not vary too much. For realistic re-use distance values a spectrum demand of six to seven TV channels seems to be necessary to allow for a nation-wide DVB-T coverage in Germany only.

re-use distance in km	Europe			clique number	Germany		
	I	II	III		I	II	III
40	7	7	7	7	5	5	5
45	7	7	7	7	6	6	5
50	9	8	8	8	6	6	5
55	10	9	9	9	6	6	5
60	10	9	9	9	6	5	5
65	10	9	9	9	7	6	5
70	10	9	9	9	7	5	7
75	10	9	9	9	7	7*	6
80	10	10	10	10	7	6	6
85	11	10	10	10	7	6	8*
90	12	11	11	11	8	7	7
95	12	11	11	11	8	7	7
100	12	11	11	11	8	7*	7
105	12	11	11	11	9	8	7*
110	13	11	11	11	8	8*	8*
115	13	12	12	12	8*	8*	6
120	13	12	12	12	8*	8*	6
125	13	12	12	12	9*	8*	8*
130	13	12	12	12	10	9*	7
135	13	12	12	12	10*	9*	8
140	13	12	12	12	10*	9*	8
145	13	13	13	13	9*	9*	9
150	13	13	13	13	10*	10*	9*

Table 7.4: Number of channels to establish a plan without any violation of the adjacency relations between allotment areas as a function of the re-use distance. Three different graph coloring algorithms have been employed (I,II,III). Non contiguous sets of channels are labelled by an asterix.

Chapter 8

Constrained Frequency Assignment Problems

The estimation of the spectrum demand is a very important issue with respect to the introduction of a new broadcasting service. Knowledge of the spectrum demand is a prerequisite in order to decide if the envisaged planning targets are realistic or need to be revised.

In practice, usually another problem has to be solved more often. It concerns the question how the different needs of broadcasters can be satisfied under the constraint of having only a limited portion of spectrum at disposal. A very prominent example is the sharing of the VHF range by T-DAB and DVB-T. It is intended for example by the German public broadcasters to provide nation-wide coverages both with T-DAB and DVB-T in band III. The near future target to establish two layers of T-DAB and one layer of DVB-T has been put forward in [IDR00]. According to the commercial success of both systems more coverages are very likely needed in the future.

The compliance of the demands formulated in [IDR00] for Germany, however, cannot be tackled without taking into account the constraints imposed by other European countries. Unfortunately, there is no possibility to introduce new systems on a broad basis on top of the existing services since the actual spectrum usage both in VHF and UHF definitely impedes such an intention. The entire spectrum is already used for analogue terrestrial television and also for other non-broadcasting service.

Therefore, the ITU decided to revise the frequency plan laid down in the Stockholm agreement from 1961 in order to pave the way for the all-digital terrestrial broadcasting future. This revision will be held mainly at the instigation of several European organizations like EBU and CEPT. The preparations are currently underway. In 2004 a first conference will be held whose task is to collect and settle all technical planning parameters as well as the planning principles.

Two years later a second conference is to be held during which the new frequency plan for digital terrestrial broadcasting in the bands III, IV and V has to be generated on the basis of the principles defined by the first session. At the beginning, the planning area encompassed only the European Broadcasting Area (EBA). It turned out, however, that there was also a need for a digital plan in other regions. Therefore, the planning area has been extended and now extends from the very North of Europe down to South Africa. The eastern border line is drawn 200 km before Wladiwostok in Russia. Including all of the Russian area would have led to the inclusion of the American continent, too. Since the Americans at least currently are not interested neither in T-DAB nor in DVB-T it did not make any sense to also include America.

The task that has to be accomplished by the ITU conference is far from being easy. On one hand, the new plan must take into consideration both T-DAB and DVB-T along with their different individual requirements in terms of frequency management and network implementation issues. This inevitably touches upon the spectral coexistence of both systems in band III. On the other hand, it is absolutely crucial that the new frequency plan leaves sufficient freedom and is flexible enough to allow the incorporation of future developments which cannot be foreseen today.

Just the latter requirement is a very important issue to be taken care of by the conference. The Stockholm plan from 1961 has been build under the technical conditions that existed at that time. It served by now more than forty years, a period of time which has seen dramatic changes in terms of development of new broadcasting and telecommunication systems. In the sixties there was only analogue technique. Nobody could have foreseen a digital revolution as we face it today.

However, the important point is that the Stockholm plan for some time allowed to accommodate a certain amount of new digital terrestrial broadcasting systems. The Wiesbaden agreement for T-DAB is an example [CEP95]. Also, the Chester agreement [CEP97] opened the way to introduce DVB-T into the existing analogue environment. It has to be emphasized that all that happened under the umbrella of the Stockholm plan from 1961.

Even if most of the European countries are going to introduce DVB-T they will certainly not be able to use the same frequency resources. Some will accommodate DVB-T only in UHF, whereas others will make use of VHF, too. In Germany, it seems that both options will be employed. Thus it follows that T-DAB and DVB-T will have to share the VHF range in some regions of the planning area while others do not have to face that problem.

Depending on the different countries there is large probability that very different constraints and boundary conditions in terms of spectrum availability will have to be accounted for. Furthermore, there are a lot of different radio services which have nothing to do with broadcasting. However, they make extensive use of spectrum both in VHF and UHF. They require an adequate protection. Examples thereto can be found in military communication systems but also radio astronomy is one of the services to be protected by all means.

In order to be able to cope with all these factors very sophisticated and flexible planning tools will have to be developed and applied. Along the line of discussion presented in the preceding sections it seems to be obvious that both graph coloring algorithms as well as stochastic optimization routines will certainly play an important role in that context.

As has been already pointed out in section 4.3 there are two principle planning options, namely assignment or allotment planning. Currently (2004) it seems that both approaches have to be dealt with during the revision conference on an equal basis. How this can be accomplished in detail has so far not yet been elaborated. From a mathematical point of view there is no fundamental difference between them.

Assignment planning assumes that all characteristics of a transmitter are known. As a consequence, its coverage area is then fixed, too. This can be approximated by a corresponding polygon and subsequently treated as an ordinary allotment in an allotment planning process. The details of this mapping of an assignment into an allotment, above all appropriate protection mechanisms, have to be developed in the near future. Once the frequency plan has been established, the original transmitter could be used for the implementation. Furthermore, there is also the possibility to implement a single frequency network according to the principles of the allotment plan. Due to the obvious connection between assignments and allotments only the latter will be dealt with here.

8.1 The Need for Flexible Planning Approaches

The starting point of every realistic frequency assignment based on allotments as well as any estimation of the spectrum demand is a set of coverage areas. In principle, they can have arbitrary shapes. They are approximated by correspondingly chosen polygons. In order to keep the administrative effort as small as possible the number of vertices that can be used to describe a coverage area is usually limited. At Wiesbaden 36 vertices could be used for regional areas whereas a national territory could be approximated by up to 99 vertices.

The specification of the number and the location of the vertices defines the geographical shape of the coverage area. Usually, geographical coordinates are used to represent the vertices. According to the requirements they can be easily converted into other coordinate systems which might be better adapted to cope with regional or even local conditions.

For practical reasons, the coverage areas are supposed to posses a simple geometrical structure. In mathematical term this means they should preferably be simply connected geometrical objects. Holes or loops are to be avoided. It might happen, however, that a coverage area across which one single programme content should be broadcast is broken down into two or more separate areas. In that case they should be treated as independent areas because otherwise if the same frequency is assigned to them they have to be separated from each other more than the corresponding re-use distance to avoid too high interference levels. If there is a reason that they must have the same frequency then this can be cast into a corresponding set of constraints for the frequency planning process.

Such cases can be considered as being pathological and will therefore not be pursued here. Rather, it will be assumed that disjoint coverage areas are to be considered as being independent. Furthermore, there are no other restrictions for the definition of the coverage areas. They are even allowed to partially or totally overlap.

Apart from the geographical description of a coverage area it must be specified whether the area under consideration is supposed to represent a T-DAB or a DVB-T coverage area. This is essential in connection with the assessment of violations of the re-use distance or the determination of the mutual interference potential of two allotment areas. Two T-DAB areas using the same T-DAB block need to be separated by a different distance as two DVB-T allotments which utilize the same TV channel. In the case of DVB-T there is in addition a large variation of the re-use distance with regard to the different operation modes that can be used.

The definition of a coverage area is concluded by specifying a list of all frequencies that can be used within that area. During the frequency assignment process the allotment is assigned only one of those channels or blocks that are contained in its list. That way it is assured that any assignment is valid at least as far as the usage of the frequency resources is concerned. Any frequency plan derived therefore does not include any cases where a channel or block cannot be used. It has to be noted, however, that this is not linked in no way to the assessment of the plan which is a separate issue.

The propagation of electromagnetic waves is not affected by the existence of any national or regional boundaries. As already seen in the previous sections, a realistic frequency plan has to be derived on the basis of a large geographical planning area. Too small planning regions will lead to unreliable results since the boundary conditions are not taken into account properly.

Considering large geographical areas requires to take into account the curvature of the earth's surface when calculating distances between two points. Usually, it is sufficient for frequency planning tasks to employ a spherical approximation of the earth. This facilitates the distance calculations significantly since the geographical coordinates can used directly together with spherical geometry.

What cannot be neglected, however, is the impact of the different wave propagation conditions above land or sea. All calculations carried out in order to estimate the interaction between two allotment areas must take into account these differences. One way to cope with these conditions is to introduce the concept of effective distances between points as described in detail in section 5.1.3.

The crucial parameter to which in principle the whole frequency assignment problem can be mapped is the re-use distance for the usage of the same frequency. In the case of DVB-T this means the same TV channel while for T-DAB the same block is meant by “the same frequency”. The determination of the re-use distance rests on the concept of reference networks as discussed in section 5.1.2.

In order to calculate the re-use distance it is necessary to fix the characteristics of the reference networks. Apart from the obvious question how many transmitters are to build the network and where are they supposed to be located, it has to be decided what should be the coverage target for the reference network. This determines the details of the network configuration like power of the transmitters, type of antenna diagrams and so on.

Furthermore, a decision has to be taken what type of operation mode is to be considered. For T-DAB this is a relatively easy choice since there are only four different operation modes. They are, moreover, intended to be used in different spectrum ranges. That means that for a T-DAB reference network in VHF essentially there is no choice for the operation mode because mode I has been designed for VHF and mode II for L-band.

The situation is completely different for DVB-T. There is great variety of different operation modes that can be used when implementing the transmitter networks. Theoretical investigations have shown that the resulting re-use distances can differ by a factor of three or more [IRT99].

Furthermore, in the VHF range the spectrum sharing scenario between T-DAB and DVB-T requires to consider also the mixed interaction case. This means that if in one area a T-DAB network is to be set up and in another area there exists a DVB-T network it must be possible to make statements about co-channel usage and the distance the two coverage areas need to be separated in order not to produce inacceptable mutual interference levels.

It can be assumed that within the territory of a country there will be in general several coverage layers with T-DAB or DVB-T at the same time. Most administrations strive for a multitude of services to be receivable at each point throughout their area of responsibility. Clearly, the coverage areas in which for example two different DVB-T multiplexes are broadcast need not be identical. They probably will only partially overlap.

Normally, two adjacent areas which do not overlap can be assigned adjacent TV channels or T-DAB blocks. This does not give rise to any problems for the implementation of the corresponding networks. The situation changes if the two areas partially or totally overlap. Investigations in particular in Great Britain have shown that severe problems may be encountered if different network operators are running their networks in overlapping areas using their own transmitter sites on adjacent T-DAB blocks. In the vicinity of a transmitter of one network the service provided by the other network is considerably perturbed. If both networks providers can agree on broadcasting their services from the same sites using the same antennas there are no or only few problems to be expected [Rad01].

Overlapping areas need to be marked when applying a frequency assignment method. Depending on the underlying model of the frequency assignment it can be important to quantify the overlap of areas. This is relevant in the context of deciding if the overlap is to be neglected because its origin might be a failed coordination of the two affected administrations. On the other hand, there might be cases where different rules for the frequency assignment apply depending if the overlap is 50 % or 100 %.

In order to avoid problems in overlapping areas the frequencies assigned to them should have a certain minimal spectral distance. In that context it is important that the services which are to be provided in the overlapping areas are specified. The reason is that T-DAB and DVB-T occupy a different bandwidth. Thus, the critical spectral distance to be obeyed depends on the type of service under consideration.

Once a set of allotment area has been specified along with all necessary planning parameters, then every area can be assigned a frequency taken from its proper pool of allowed frequencies. Clearly, an arbitrary assignment of frequencies will in general lead to a large number of incompatibilities in terms of violations of the corresponding re-use distances. Different assignment will give rise to different sets of incompatibilities.

Therefore, it is mandatory to establish a quality measure for frequency plans in order to distinguish good from bad plans. The proper definition of this quality measure is crucial for a successful generation of a frequency plan. All items discussed above have to be taken into account appropriately.

The definition of the quality measure for the plan is directly linked to the question what the global planning target governing the creation of a frequency plan should be. Usually, there are two main targets that can be pursued. Either the number of used is frequencies is to be reduced at the cost of accepting incompatibilities or the avoidance of these incompatibilities has highest priority, however, at the cost of using more frequencies. Both targets usually cannot be obtained at the same time. Therefore, a trade-off between them has to be envisaged. The better this balance is cast into mathematical terms the better, clearly, the resulting frequency plans will be in the end.

8.2 Key Elements of Flexible Planning

A large number of individual constraints usually prohibits the application of highly sophisticated specially designed algorithms in order to calculate a frequency plan. This is in particular valid since very often the constraints are not fixed for all times. Just the preparation work for the revision of the Stockholm plan demonstrates that there is a need to very quickly react to modifications of the constraints. Especially during a planning period in which the planning parameters are not finally agreed it is necessary to have the possibility to test different variants. Hence, the only approach which seems to be practical rests on the application of stochastic optimization algorithms (see section A).

In the subsequent sections one particular strategy will be discussed as a feasible method in the light of several constrained frequency assignment scenarios. It is based on the Great Deluge algorithm presented in detail in section A.1. The same investigations can be carried out with the help of Tabu Search, Simulated Annealing or other methods. The application of one or the other algorithms mainly seems to be a matter of taste.

All of these stochastic optimization methods have their advantages and disadvantages. There is no algorithm which could be generally speaking considered as being superior to others in terms of performance. The investigation of the problems arising in the context of the central topic to be discussed in the sequel, namely the sharing of VHF by T-DAB and DVB-T is not affected by the application of a particular strategy.

8.2.1 Re-Use Distances for T-DAB and DVB-T

There are several factors which influence the values of the re-use distances for T-DAB and DVB-T. It is evident that the frequency range under consideration will affect them to a large extend. This is connected to the different wave propagation conditions in VHF, UHF and L-band.

Apart from the frequency re-use distances are determined mainly by the protection ratios for T-DAB and DVB-T with respect to their different operation modes as defined in the final agreement documents of Wiesbaden [CEP95] and Chester [CEP97]. They indicate the robustness of the systems under the influence of any perturbations by other transmitters or networks.

These values are explicitly used in order to calculate the re-use distances on the basis of the reference networks (see for example 5.1.2). It is obvious that for different interaction scenarios different re-use values are derived. Two co-block T-DAB allotments must certainly obey another re-use distance as two co-channel DVB-T areas. The same is true for the mixed case of T-DAB against DVB-T.

All protection ratios have been determined on the basis of theoretical studies and laboratory experiments. Depending on the envisaged coverage targets like fixed, portable or mobile reception and the in each case required minimum field strength different protection ratios are found. Furthermore, the choice of the modulation scheme affects the protection ratio as well.

In particular, the mixed interference scenario is very complex. A 7MHz TV channel accommodates four T-DAB blocks. Depending on the number of T-DAB blocks interfering the DVB-T signal different protection ratios are necessary. All four T-DAB blocks on air corresponds to the worst-case situation which, however, seems to be rather likely if several layers of T-DAB are realized on top of one DVB-T layer in the VHF range.

The great variety of different DVB-T and T-DAB system variants gives rise to a very large number of possible re-use distances. Whether all these values should be used for planning purposes is controversial. Obviously, if allotment planning is carried out on the basis of system variants requiring large re-use distances it is no problem to subsequently use the assigned frequencies also for the implementation of variants which are compatible with smaller re-use distances. The only drawback is a possible inefficient use of spectrum. Thus, it might be advantageous to identify a small subset of re-use distances which are compliant with a reasonable number of different system variants.

Such an approach is followed here. Based on the available information for the protection ratios a set of re-use distances has been derived which was used for the investigations presented here. Table 8.1 shows the values.

T-DAB-T-DAB	DVB-T-DVB-T	T-DAB-DVB-T
81	120	150

Table 8.1: Re-use distances for T-DAB and DVB-T in km for the VHF range employed in the simulations.

The value for T-DAB against T-DAB has already been defined at Wiesbaden. So, there is little doubt that it will be used also in the future. The values for DVB-T instead still have to be considered preliminary. The reason is that no reference networks for DVB-T have been agreed upon yet. Nevertheless, for the DVB-T against DVB-T case a value of 120 km seems to be good compromise. This refers to a 64QAM system using a code rate of 2/3. For the implementation of the frequency plan this means that all other system variants which are more robust than this variant could be put into operation as well. Most of the interesting operation modi of DVB-T fall into this class. However, the 150 km for the mixed scenario indeed need to be proved by further studies.

The values of table 8.1 refer to propagation above land only. If two coverage areas are separated by sea then the re-use distances increase considerably. At the Wiesbaden conference in the case of T-DAB in VHF the values 142 km and 173 km have been applied for propagation above cold sea and warm sea, respectively [CEP95].

Since so far no agreed values for DVB-T are available a simple but obvious approach can be chosen to derive the corresponding sea path values from the Wiesbaden figures. The propagation of electromagnetic waves certainly is not linked to any network configuration. It is governed purely by the laws of physics. Therefore, the ratio between the re-use distances for cold sea and land as well as warm sea and land should be constant for all re-use distances regardless to which type of reference network of interaction scenario they refer. The re-use distance ratios are

$$\frac{\text{re-use distance above cold sea}}{\text{re-use distance above land}} = 0.570 \tag{8.1}$$

and

$$\frac{\text{re-use distance above warm sea}}{\text{re-use distance above land}} = 0.468\ . \tag{8.2}$$

Applying this relations to the values of table 8.1 the following matrix of re-use distances can be calculated.

	T-DAB-T-DAB	DVB-T-DVB-T	T-DAB-DVB-T
land	81	120	150
cold sea	142	210	263
warm sea	173	256	320

Table 8.2: Employed re-use distances for T-DAB and DVB-T in km for the VHF range for propagation above land, cold sea and warm sea.

8.2.2 Adjacent Spectrum Usage in Overlapping Areas

Overlapping coverage areas require special rules for the frequency assignment. This fact just recently became obvious. So far this has been no topic in most European countries since there was no need for T-DAB coverages beyond what Wiesbaden provided in VHF and L-band. These two spectral ranges, however, are separated far enough that no interaction will take place. But since in some regions in Europe the demand for T-DAB coverages exceeded the capacities assigned at the Wiesbaden conference new allotments have been created and corresponding transmitters networks have been established.

This happened especially in Great Britain. Due to the lack of available spectrum the administrations were forced to use T-DAB blocks which were adjacent to the blocks already in use in overlapping areas [Rad01]. Unfortunately, the T-DAB networks were run by different operators. Each of them wanted to use his own transmitter sites. Partially drastic perturbations of the service quality were encountered in both the new and the old networks. In the vicinity of each transmitter site the coverage of the other network was affected to the extend that no reception was possible any longer.

Clearly, if only a very limited amount of spectrum is available such consequences cannot be avoided without putting a very large effort on the upgrade of the network structures. Adding more transmitters to the existing single frequency network might cure the problem. This however, is directly linked to dramatically increasing costs for the network operation. The only possibility to use adjacent frequencies in overlapping areas is to employ co-sited transmission of both multiplexes. If this is always a practical solution is another question.

On the occasion of the generation of a new plan as is envisaged with the revision of the Stockholm plan such problems could be relaxed from the very beginning by handling them appropriately. A simple way to deal with adjacent frequency problems is sketched in figure 8.1.

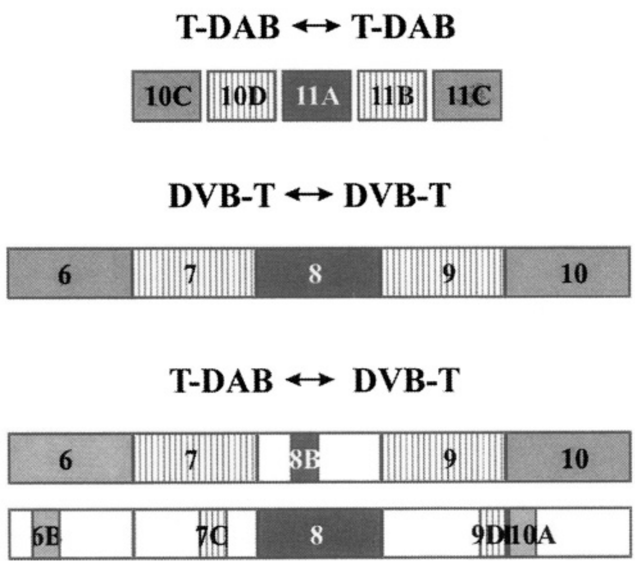

Figure 8.1: Adjacency relations between channels and blocks for overlapping areas. The dark gray indicates the spectral interval under consideration, dashed frequencies are to be avoided in the case of overlap and light gray is acceptable.

Figure 8.1 has to be interpreted such that there is an area which currently uses one of the dark gray channels or blocks in the middle of figure 8.1. Another area is overlapping the first one. As usual there are three possible interaction scenarios, i.e. T-DAB against T-DAB, DVB-T against DVB-T or T-DAB against DVB-T. It is obvious that the usage of a dark gray frequency in the overlapping area is prohibited since none of the two networks would work properly.

If the overlapping area makes use of a dashed channel or block then mutual perturbations would result if the transmission were not co-sited. In general, such assignments such be avoided. The light gray channels or blocks could be used without any problems.

In general, it is assumed that two overlapping T-DAB areas should use blocks that are spectrally separated at least by one block. In the case of two DVB-T coverage areas a buffer of an entire channel should be obeyed. According to figure 8.1 the same spectral distance should apply in the case of a mixed interaction. This conditions seems to be obvious, however, it is also very restrictive.

That way, the position of a T-DAB block inside a TV channel is not important. This implies that throughout the overlap area for example the block 12D has the same negative impact on the a DVB-T service in channel 11 as a block 12A. This is certainly not true. Applying the rules of figure 8.1 strictly will lead to a frequency plan advising the usage of a large number of frequencies. It has to be decided in each case individually if these severe restrictions cannot be relaxed. Figure 8.2 shows a less restrictive alternative.

For the application of the model of figure 8.2 it is assumed that a receiver can demodulate the wanted signal if other signals obey a certain minimal spectral distance Δf_c. This critical spectral distance is different for T-DAB and DVB-T receivers. Furthermore, it is very likely that receivers from different manufactures have different capabilities to suppress adjacent channels or blocks.

Seen from a planning perspective it is necessary to define one single value for Δf_c which can be used for all system variants. In principle, also in that case the weakest system needs to be protected for a common planning approach. However, in contrast to the discussion about the different protection ratios for different systems variants of DVB-T the definition of Δf_c is not that difficult. A value for the critical spectral distance of $\Delta f_c = 0.5$ MHz has proved to be a reasonable choice.

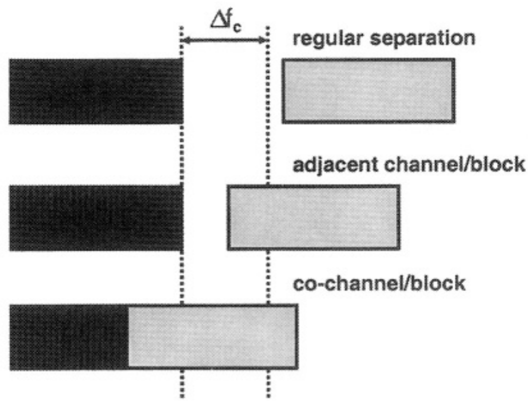

Figure 8.2: Less restrictive model for the adjecency relations between channels and blocks in the case of overlapping areas.

8.2.3 Assessment of a Frequency Plan

To generate a frequency plan means to assign to each of the coverage areas under consideration one of its permissible frequencies. An arbitrary assignment will in most cases lead to a very inefficient frequency plan full of incompatibilities and thus a high level of interference. There will be both violations of the re-use distance and adjacent or even co-channel assignments in overlapping areas.

Any other arbitrarily chosen frequency plan might differ in detail but the probability that it suffers from the same problems is very high. Therefore, two questions arise. First, how can the quality of a frequency plan be assessed, i.e. how can "good" plans distinguished from "bad" plans? The second question coming up immediately then is how can a good frequency plan be derived or calculated?

The latter question is more easy to answer. The derivation of a good frequency plan corresponds to a very complex combinatorial problem from a mathematical point of view. Due to the large number of individual constraints to be taken into account stochastic optimization algorithms can be applied to find an optimal solution.

The former question concerning the quality needs further specification. First of all, it must be possible to define an attribute like "good" in the context of a frequency plan anyway. This means a precisely defined quality function must exist. A first requirement thereto is the uniqueness of this function. A particular frequency plan can only have one single quality.

Moreover, the mathematical form of the quality function depends on several factors. Clearly, the number of incompatibilities will certainly be important. This alone, however, will not be sufficient. Rather, it is necessary to establish a quantitative measure for the interference in the plan. The negative mutual impact of two areas using the same frequency is increasing the smaller their separation distance is with respect to the re-use distance. So, in general there are four categories of incompatibilities which are sketched in figure 8.3.

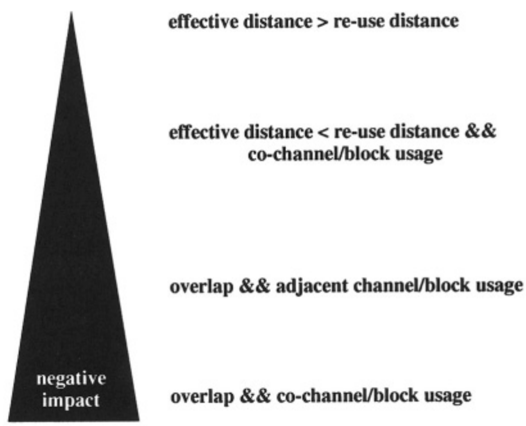

Figure 8.3: Negative impact of different imcompatibilities in a frequency plan.

A first and obvious approach to quantify the incompatibilities between two areas i and j in a frequency plan is to define their incompatibility Δ_{ij} as the difference between the corresponding re-use distance and the geographical distance of the two areas. Since wave propagation conditions above land and sea have to be taken into account geographical distance has to be interpreted in term of the effective distance introduced in section 5.1.3. The idea behind this concept is that areas separated by land and sea affect each other as if they would be virtually closer but separated by land only. Effective distance thus is nothing but a properly scaled geographic distance. Therefore, the effective distance has to be compared to the re-use distance above land. In mathematical terms this can written as

$$\Delta_{ij} = RU \; - ED(i,j) \tag{8.3}$$

where RU denotes the re-use distance above land and $ED(i,j)$ the effective distance between areas i and j.

In the case of overlapping areas the effective distance clearly is equal to zero. However, it might be necessary to distinguish incompatibilities according to the degree of overlap. A quite natural measure is to calculate the fraction ψ of the overlap with respect to the total area of the two coverage areas, i.e. for example

$$\psi = \frac{2 \times \text{size of overlapping area}}{\text{area}(i) + \text{area}(j)} \; . \tag{8.4}$$

Then, the definition of the incompatibility Δ_{ij} has to be extended for example in the form

$$\Delta_{ij} = \begin{cases} RU \; - ED(i,j) & , \quad ED(i,j) > 0 \\ \\ RU + 100 \times \psi & , \quad ED(i,j) \leq 0 \end{cases} \; . \tag{8.5}$$

A straightforward idea to generate a global measure for the quality of a frequency plan is then to sum all Δ_{ij}. The quality function to be exploited by the stochastic optimization algorithms thus reads

$$\Delta = \sum_{i,j} \Delta_{ij} \; . \tag{8.6}$$

The definition (8.6), (8.5) has to be considered only as one special way to handle this. In practice, it might turn out that a different formulation has to be found because the difference of the quality of two frequency plans does not allow to objectively decide which one is better. Quite often it is necessary to declare the case three and four of figure 8.3 as forbidden to suppress the occurrence of such incompatibilities with high probability. The quality "forbidden" can be realized in terms of a predefined very large value for Δ_{ij} so that the penalty for accepting the corresponding assignment in the plan will be extremely high. Figure 8.4 sketches a decision tree in order to assess the incompatibility between two areas for such an approach.

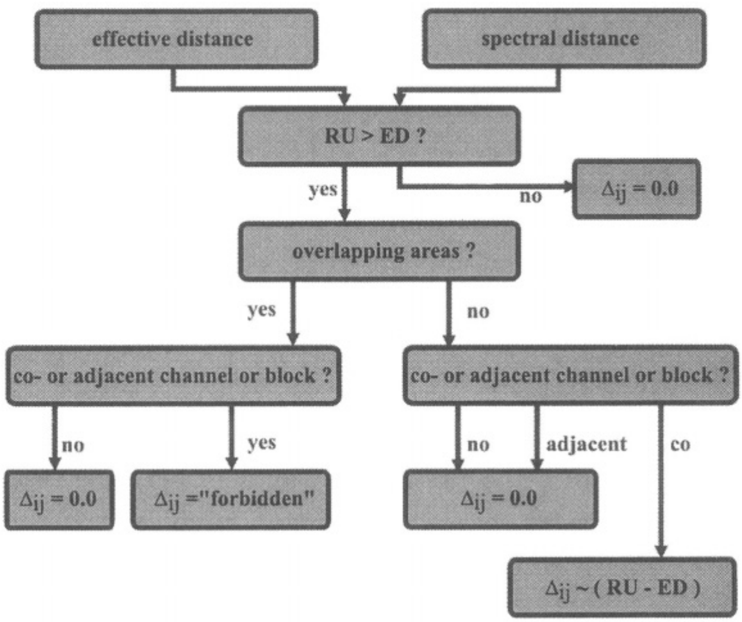

Figure 8.4: Decision tree in order to assess the incompatibility between two coverage areas.

The decision process starts with the calculation of both the effective distance and the spectral distance between the two areas under consideration. Then the incompatibility Δ_{ij} is calculated along the lines of the decision tree of figure 8.4. As before, a global measure in order to assess the total set of assignments, i.e. the frequency plan, could be to sum all Δ_{ij}. If necessary appropriately chosen weights can be used for the summation to put more emphasis on particular sections of the frequency plan.

8.2.4 Implementation Issues

There are certainly various way to take into account of all relevant factors discussed above when designing some software tools for the generation of new frequency plans. To highlight the special problems that might be encountered a very special implementation of the aforementioned planning principles will be presented in the sequel. It has been developed during the preparation of the revision of the Stockholm plan. Numerous studies have been carried out with help of this particular software implementation. Different aspects of the whole planning complex were covered (see for example [Beu02]).

It has to be emphasized, however, that this method does not claim to be the only feasible approach. Very different strategies will certainly lead to similar results. Nevertheless, the chosen approach proved to be very flexible, in particular with respect to changes of the planning constraints. During the preparations of the revision conference it turned out that once a planning result was available, immediately the demand for a new planning exercise emerged which usually differed simply by modifying some of the constraints.

Furthermore, the method described below allows to illustrate the essential features that any implementation of a planning strategy must take care of in a transparent way. The core of the frequency assignment method roots on the Great Deluge algorithm [Due93] discussed in detail in section A.1. This algorithm proved to be a very robust and reliable, nevertheless very easily implementable mathematical machinery.

To keep the simulation as simple as possible all input data should preferably be given in terms of pure ASCII files. This way it is possible to introduce changes very quickly and independent of any special computer configurations like the employed operating system. Tables 8.3 and 8.4 show a reasonable data structure and a short section of a typical input data file as an example.

Basically, there are two possible ways to specify the frequencies that can be used during the simulation. In tables 8.3 and 8.4 each frequency is coded in terms of frequency band, TV channel and frequency block. No explicit frequencies have been used. Such a scheme can only be applied if one common channel raster is utilized across the considered planning area. This happens to be the case in the UHF range in the European

file entry	explanation
#########	separating line between entries
service	DAB, DVB
actual frequency	band (0: VHF, 1: UHF, 2: L-band), channel (VHF: 5-13, UHF: 21-69, L: 100-103), block (DVB: 0, DAB: 1-6)|
name of service	string
number of vertices	integer
geographical coordinates of vertices	longitude,latitude|longitude,latitude|...
number of allowed channels or blocks	Integer
allowed frequencies	band,channel,block|band,channel,block| ...

Table 8.3: Structure of an input data file.

```
#####################
DAB
0,12,1|
BADEN-WUERTTEMBERG-DAB
36
+8.42,+49.58|+8.58,+49.52|+8.65,+49.62| ...
8
0,11,1|0,11,2|0,11,3|0,11,4|0,12,1|0,12,2|0,12,3|0,12,4|
#####################
DVB
0,12,0|
BADEN-WUERTTEMBERG-DVB
36
+8.42,+49.58|+8.58,+49.52|+8.65,+49.62| ...
8
0,5,0|0,6,0|0,7,0|0,8,0|0,9,0|0,10,0|0,11,0|0,12,0|
...
```

Table 8.4: Input data example for a T-DAB and a DVB-T allotment area.

countries which all use TV channels of 8 MHz bandwidth with defined center frequencies.

However, when looking at the VHF range the situation is different. First of all, there is a set of distinct channel rasters used in different countries as shown in figure 4.1. For example in Germany TV channels are defined to use a bandwidth of 7 MHz. Each of these channels can accommodate exactly four T-DAB blocks. Other countries have TV channels of 8 MHz in band III. In that case the T-DAB blocks are not perfectly aligned with the channel raster.

Furthermore, there are special cases like France and Italy which use also 8 MHz or 7 MHz, respectively, but in connection with different center frequencies (see also 4.1). All this leads to a very complicated situation in particular along boundaries between countries employing different raster schemes. Clearly, any investigation that tries to find a frequency plan under such circumstances must explicitly make use of precisely defined center frequencies and bandwidths.

The first step of any calculation of a frequency plan is to provide a list of all allotment areas for example in the form shown in table 8.4 or a modification thereof to cope with different channel rasters. At the beginning of the simulation the effective distances between all pairs of areas have to be determined. This includes to account for all relevant wave propagation conditions that need to be considered such as propagation above land or sea. Moreover, the wave propagation depends essentially on the frequency. These effects have to incorporated as well.

Overlapping areas must be marked, either in a simple way indicating only overlap or not or, if necessary, the degree of overlap has to be calculated. Then, an arbitrary initial configuration, i.e. an initial frequency plan, is chosen. This is accomplished by choosing for each of the allotments one if its allowed frequencies by chance. Subsequently, this initial plan is assessed by passing through the decision tree of figure 8.4 for each pair of polygons. Subsequently, the optimization cycle can be started.

During the simulation a new configuration is generated from the current one by arbitrarily choosing one allotment and assigning another of its allowed frequencies also by chance. Clearly, it is also possible to change more than one allotment during one iteration step. It is reasonable to try different strategies in order to obtain the best results. Such freedom is typical for stochastic optimization algorithms. There are no precise rules how to proceed in a given situation. On that level, the solution process is governed to some extend by trial and error. However, this does normally not pose any problems in practical applications.

The simulation will be carried out until a predefined stop criterion is met. In most cases a maximum number of iteration steps is given. At the beginning of the stochastic search usually a lot of new configurations are found having a quality which is currently the best. As the simulation proceeds it takes more and more time to find a new best solution. The algorithm that way converges to the final solution.

If the simulation does not produce a new best solution after a reasonable period of time there are several options how to stimulate the search again. The most straightforward way is to re-start the whole process. To this end, a new initial configuration is chosen and the algorithm started again. In principle, this can be done as often as desired.

Sometimes, however, it is better not to start from scratch again. The search might have come to an end because there is not much freedom left to generate new configurations from the current one. Thus, relaxing this a little bit and allowing for more freedom might help to escape from the dead-end situation again. Like before, such a relaxation can be applied as often as it seems necessary.

8.3 Planning Examples

The planning examples presented here refer to the shared usage of band III by T-DAB and DVB-T. The major part of the examples is concerned about Germany. The neighboring countries of Germany have not been taken into account to reduce the complexity of the problems and make the examples more transparent. At the end, an example referring to central Europe is discussed to put some emphasis on those issues which are directly linked to the different ways of using the spectrum in different countries.

It is by no means intended to give a thorough presentation of the entire sharing problem in the VHF range. Neither an optimal solution to that problem is proposed. Currently the latter cannot be achieved primarily because neither the planning targets nor the planning principles for the revision of the Stockholm plan are known.

Rather, the following examples should give insight into the manifold of open questions and problems which certainly will arise before, during or even after the conference. This great variety of different planning tasks strikingly gives evidence for the need of a flexible planning method like for example described in the previous sections.

The starting point of the investigations were modifications to the T-DAB polygons used at the Wiesbaden conference. Originally, at Wiesbaden all German Bundesländer were supposed to be treated as indivisible units. Due to the lack of spectrum the eastern Bundesländer in Germany had to be split into two parts in order to avoid incompatibilities with analogue television assignments in channel 12 in the neighboring countries in eastern Europe.

This unnatural division which furthermore was not desired from a media political point of view in Germany has been neglected for the planning examples here. Each Bundesland is represented by exactly one single polygon. Moreover, any overlap of allotment areas which was present with the original Wiesbaden polygons has been polished, too. Obviously however, the allotment areas of the small Bundesländer like Berlin, Hamburg or Bremen are embedded into larger areas. Therefore, they completely overlap. But since this makes part of the German media landscape this has not been touched upon. Figure 8.5 shows the basic set of polygons.

The examples below can be subdivided into two main categories. The first covers planning scenarios with regard to T-DAB only. Starting from the results contained in the final agreement of the Wiesbaden conference some extensions and possible further developments are discussed. The second category is explicitly concerned about the way both systems T-DAB and DVB-T can be accommodated in VHF and what are the consequences that have to be expected.

8.3.1 T-DAB after Wiesbaden 1995

Revision of the VHF-Plan of Wiesbaden

In order to have a reference for further planning examples the first study presented here concerns the original Wiesbaden allotments in band III in Germany. As already mentioned above, the result of the Wiesbaden

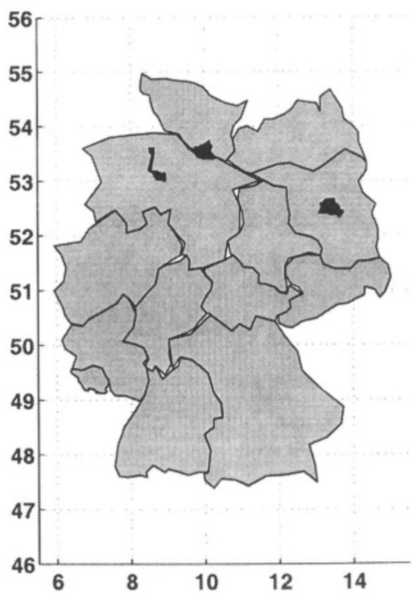

Figure 8.5: Basic set of polygons for Germany.

conference has been a sub-optimal plan due to constraints imposed by analogue television assignments in channel 12 in eastern Europe and the presence of other services which blocked the usage of T-DAB blocks in channel 12 in some areas, too.

To have a proper basis, it has been assumed that these restrictions do no longer apply. Thus, in all German VHF coverage areas it should be possible to use any of the T-DAB blocks from channel 11 and 12. Since the small Bundesländer like Hamburg, Bremen and Berlin have been fully integrated adjacent channel restrictions had to be taken care of explicitly.

By applying the frequency assignment machinery described in the previous section in detail it is possible to obtain a distribution of T-DAB blocks onto the polygons at hand which first of all does not contain any violation of the re-use distance. Furthermore, not all allowed blocks are being used. Instead, only six out of eight T-DAB block are necessary. The embedded areas for Hamburg, Bremen and Berlin are assigned blocks such that no problems due to adjacent frequency usage would result. Figure 8.6 illustrates the findings.

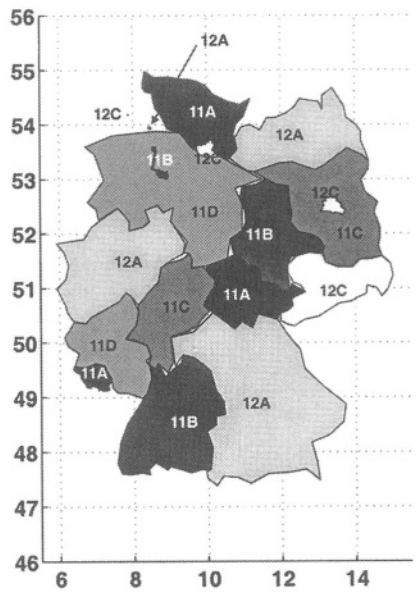

Figure 8.6: Re-distribution of T-DAB blocks onto the German band III polygons that have been designed for the Wiesbaden conference. Different blocks are indicated by different gray shadings. The labels denote the blocks explicitly by giving both channel (number) and block (character).

Relocation of Wiesbaden Allotments below Channel 12

The TV channels 11 and 12 in band III are considered to form the spectral core interval for T-DAB. Nevertheless, during the Wiesbaden conference it was not possible to assign one of these core blocks to all the submitted allotment areas. In Germany for example, it was agreed to shift all analogue television services in channel 12 to other channels in band III. Therefore, in principle channel 12 could have been used exclusively by T-DAB throughout Germany.

However, in Poland analogue transmission in channel 12 existed at the time of the Wiesbaden conference. Consequently, at the border between

Germany and Poland the German allotments could not make use of any channel 12 T-DAB block. Moreover, allotment areas had to be split into two parts with the eastern part being assigned a block from any other channel except channel 12.

It has been agreed that the analogue television transmission should come to end in the near future. Therefore, the question arises under what conditions a relocation of T-DAB blocks not accommodated within the core interval into channels 11 and 12 is possible. Clearly, the already existing channel 12 assignments from the Wiesbaden agreement are to be kept fixed.

This defines another type of planning task which can be tackled by the methods described earlier without problems. Those polygons that should be shifted to channel 11 or 12 are supplied with a corresponding list of allowed frequencies. On the other hand, the allotments that should not be touched will get only one single frequency entry, namely just the one they already use. Figure 8.7 shows the result of the investigation.

The best solution found for that problem is not free of interference, i.e. there are some violations of the re-use distance criterion. However, these are incompatibilities already present in the Wiesbaden plan. Then, it has been agreed to accept violations of the re-use distance between the allotment pairs Schleswig-Holstein/Brandenburg and Hessen/Sachsen-Anhalt. Since these assignments have not been changed during the simulation, it is evident that the associated incompatibilities are preserved. In contrast, the relocation of T-DAB blocks into channels 11 and 12 does not lead to additional problems.

Additional Coverage in VHF on top of Wiesbaden

It took some time after the Wiesbaden conference but since about the turn of the century T-DAB left the status of pilot projects in Germany and arrived at the operation as regular service. Since in Germany broadcasting falls under the sovereignty of the Bundesländer the level of operation is not homogenous across Germany. There are several network provider companies trying to push T-DAB to obtain a satisfying market penetration.

So far, the basis from a frequency management point of view is still the Wiesbaden agreement and the transmission capacities provided there.

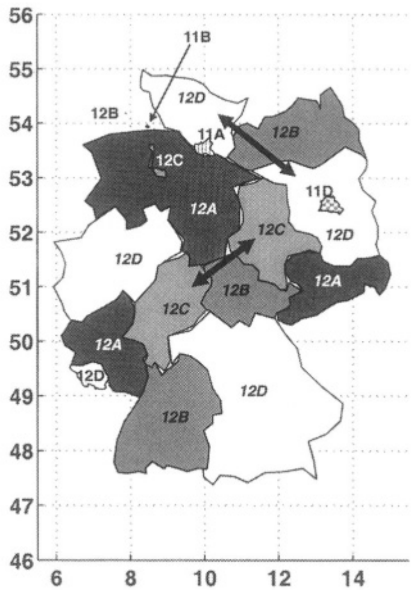

Figure 8.7: Relocation of all T-DAB blocks into band III channels 11 and 12 for the German Wiesbaden allotments. The slanted labels denote those assignments that have not been changed. The black arrows mark incompatibilities.

This means that in Germany currently there is the possibility to offer two T-DAB multiplexes at each single point throughout the territory. In order to establish an attractive programme offer to the listeners more spectrum is obviously needed as the estimations presented in section 7.2 have shown. The prospects formulated in [IDR00] with respect to the development of T-DAB tend into the same direction.

Therefore, it seems reasonable to have a closer look at the question if an additional T-DAB layer in VHF on top of the Wiesbaden allotments based solely on using the core channels 11 and 12 is feasible at all. To this end, it is legitimate as a first step to neglect all European boundary conditions and concentrate on a Germany only scenario. If already such an attempt fails or can be achieved only when accepting severe interference problems any extension to an European scale is superfluous.

A first analysis immediately shows that the two TV channels 11 and

12 are not sufficient to keep the interference levels within acceptable limits and at the same time avoid the usage of adjacent T-DAB blocks in overlapping areas. This comes as no surprise in view of the result of section 8.3.1 where a total of six T-DAB blocks were found to be required to establish one layer without re-use distance violations.

If, however, channel 13 would be available the situation would relax significantly. Channel 13 in band III is so far mainly occupied by military services in Europe. In contrast to other TV channels in band III it has a bandwidth of 10 MHz which could accommodate not only four but six T-DAB blocks. The claim to use channel 13 for civil services on a permanent basis is politically a very delicate issue.

Depending on the global political constellations and confrontations militaries are more or less inclined to get involved in discussions about a cession of channel 13 in favor of other users. Currently, there are no signs that digital terrestrial broadcasting can get hold of this additional resource within a conceivable period of time. However, this naturally does not prohibit any investigations that try to illuminate the consequences to be expected if channel 13 would be at broadcaster's disposal.

Figure 8.8 shows the result in case a second identical layer would be introduced. Channel 13 has been assumed to be available everywhere. Like in the previous section those allotment areas that already at Wiesbaden were given a channel 12 block were kept fixed. The allotments of the second layer were created by duplicating all polygons of the first layer. This includes also the small areas for Hamburg, Bremen, Helgoland, Neuwerk and Berlin.

It is possible to realize two T-DAB layers in Germany using channels 11 - 13 without causing additional interference problems. Also, there are no conflicts due to adjacent block usage in overlapping allotment areas.

8.3.2 Shared Usage of the VHF Range by T-DAB and DVB-T

Even if the conclusions of IDR [IDR00] with respect to the introduction of T-DAB and DVB-T are not fully shared by all involved parties in Germany they nevertheless form some kind of basis which could be used as a starting point for further developments. Following [IDR00] the first

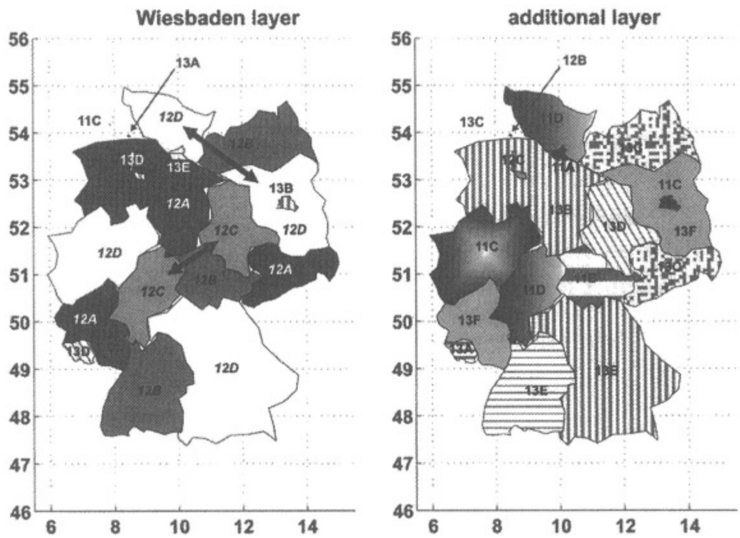

Figure 8.8: Additional T-DAB coverage layer on top of the Wiesbaden plan in VHF. All T-DAB blocks in channels 11 - 13 have been used. The slanted labels denote those assignments that have not been changed. The arrows mark incompatibilities.

near future target to be achieved is a scenario consisting of two T-DAB layers and one DVB-T layer in VHF throughout Germany on a wide area coverage basis.

The demand for DVB-T in channels 5 - 12 originates basically from the public broadcasters point of view in Germany. Traditionally, they are broadcasting the first terrestrial analogue television programme in band III using their own transmitter networks. These comprise an extended network infrastructure public broadcasters would like to exploit also in the digital era. The same argument applies in the case of T-DAB which should preferably be operated in VHF. Hence, a clash of interests results in view of the spectrum scarcity which then triggered the search for reliable and flexible planning methods.

In principle, there are three different basic possibilities to share the band III spectrum between T-DAB and DVB-T. The first option consequently pursues the idea of core channels for T-DAB initiated at Wiesbaden.

This means that channels 11 and 12 should be exclusively used by T-DAB while any DVB-T demand is to be satisfied with the help of channels 5 - 10. As long as the planning area is small like the German example of the previous sections or the number of requirements is very limited such an approach might work. If a wider planning area is envisaged and thus more severe constraints have to be taken into account or the number of T-DAB and DVB-T allotments increases significantly it seems to be rather doubtful that the core channel idea will be successful.

The second basic sharing scenario is to consequently relax the core channel approach such that both T-DAB and DVB-T allotments can be assigned any frequency in band III without any restriction. It is evident that this can lead to a more efficient usage of the spectrum due to the increased degree of freedom for planning. Since both approaches are quite similar their results can be compared very easily.

The third variant differs significantly from the two others. The idea is to make a plan for DVB-T only. Any submitted allotment area thus is assigned a full TV channel. However, since four T-DAB blocks can be exactly accommodated within the bandwidth of a 7 MHz channel in band III a network provider could decide afterwards to use that TV channel to broadcast up to four T-DAB multiplexes instead. The decision what type of service is finally offered thus could be governed to some extend by market forces.

Moreover, the very attractive advantage of that sharing strategy is that in principle there would exist the possibility to change the type of service at some point in time without the need for a new frequency assignment. If the assigned TV channel is for example initially used to broadcast DVB-T and it would turn out that no satisfying market penetration can be achieved this could be stopped and four T-DAB services could be brought on air instead. Clearly, this would require some coordination effort and also changes to the transmitter networks, but the underlying frequency plan would not be touched upon by this step. Some examples thereto will be discussed here as well.

Spectral Partitioning

The first sharing example to be discussed shall answer the question what consequences are to be expected if T-DAB is restricted to employ only blocks from a defined and limited number of channels. It concerns two T-DAB layers and one DVB-T layer. All sets of polygons constituting a single layer are identical to the Wiesbaden T-DAB polygons for band III. However, the small areas for Hamburg, Bremen, Helgoland, Neuwerk and Berlin have been omitted.

To be in line with the investigation in section 8.3.1 those allotments that already were assigned a T-DAB block from channel 12 should be kept unchanged. Furthermore, it has been assumed that all other T-DAB allotments can employ any block from channels 11 and 12. For the case of the DVB-T areas the assumption was that they could use any channel from 5 to 12.

A first alternative thereto is put forward by removing the core channel restriction for T-DAB. Thus a T-DAB allotment could be assigned any T-DAB block from any TV channel throughout band III. For both scenarios frequency plans have been calculated on the basis of the re-use distances given in section 8.2.1. The results are shown in figures 8.9 and 8.10.

In the case of the core channel restriction for T-DAB the derived frequency plan has four incompatibilities. In fact, the best solution found gives rise to spectral partitioning automatically without having imposed that a priori. The DVB-T allotments are assigned only channels 5-10 even though they were allowed to use 11 and 12 as well. On the other hand, when relaxing the core channel restriction by allowing T-DAB to use all of band III then a set of frequency assignments is found containing only one single incompatibility. Also, the segmentation of the VHF range into a part for T-DAB and another part for DVB-T is not realized.

In order to compare the planning strategy which gives the most freedom to the network provider in terms of implementing different types of services with the other two approaches only two layers have to be considered now. Actually, these are two DVB-T layers which can be used optionally either for the implementation of T-DAB or DVB-T networks. Several possibilities exist. First of all, two independent DVB-T coverage layers could be established. Another configuration could include one DVB-T layer and make use of the second TV channel for T-DAB only. Finally, both TV channels could be dedicated to T-DAB thus giving rise to a maximum of eight T-DAB multiplexes on air simultaneously.

The maximum of eight T-DAB layers probably would be possible only if all were operated by the same network provider using the same trans-

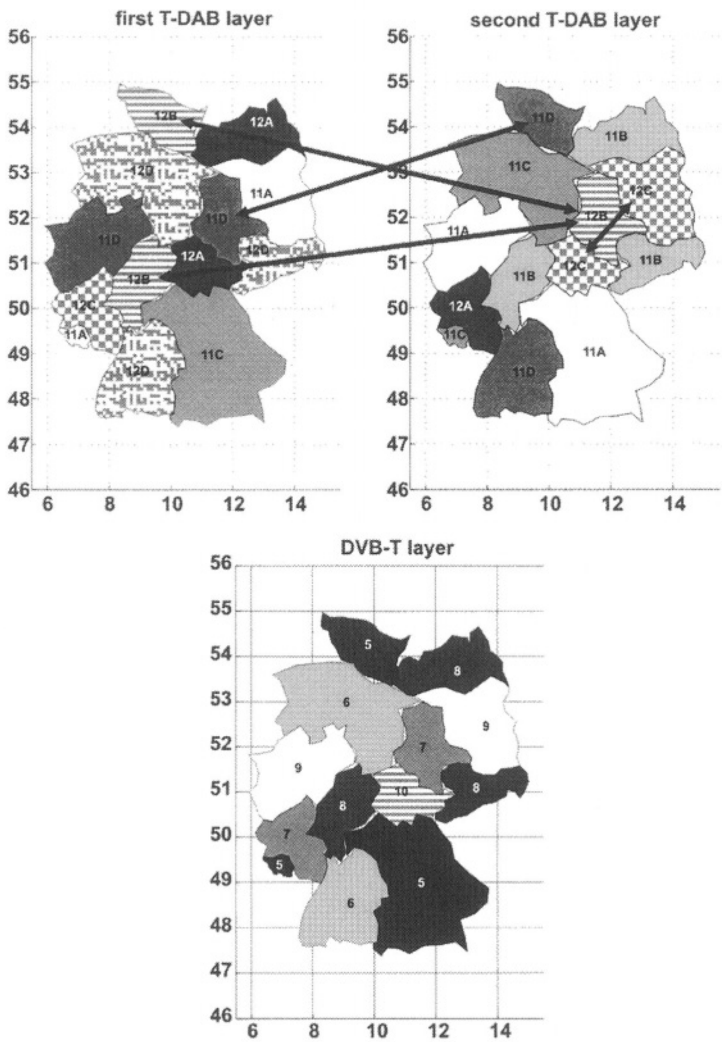

Figure 8.9: Shared usage of band III by T-DAB and DVB-T. T-DAB is restricted to use only blocks from channels 11 and 12. For DVB-T all channels from 5 to 12 are possible. The arrows indicate incompatibilities.

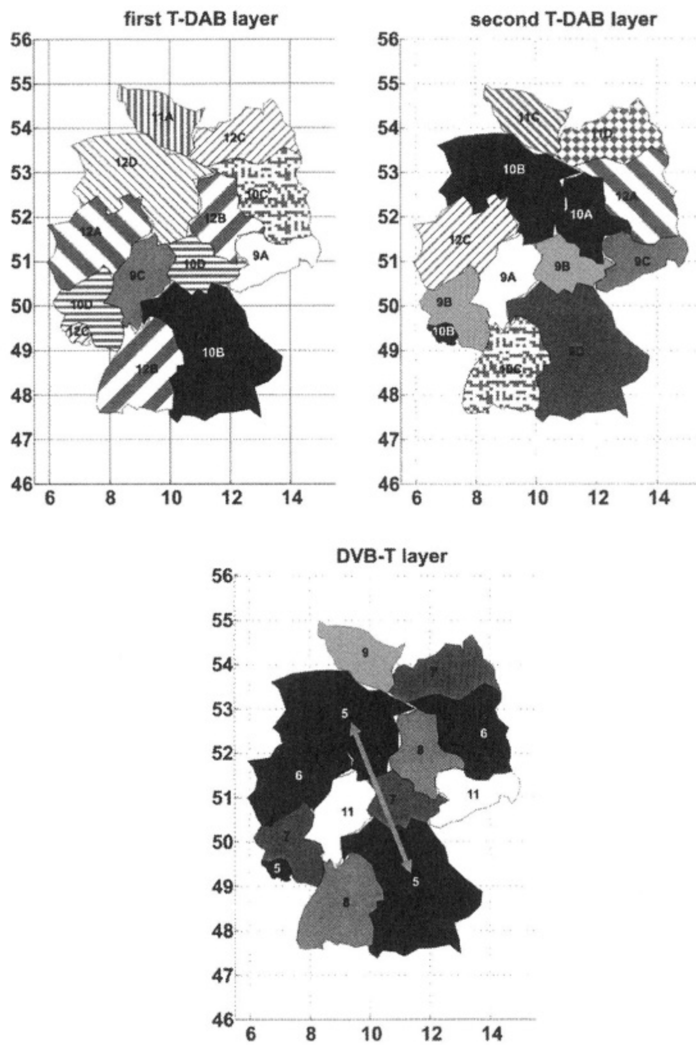

Figure 8.10: Shared usage of band III by T-DAB and DVB-T. Both services are allowed to use any appropriate frequency resources in band III. The arrows indicate incompatibilities.

mitter sites for all T-DAB transmissions. In case different providers are equipped with one of the T-DAB blocks and each of them would use his own network infrastructure problems due to adjacent block usage in overlapping areas would result. Therefore, it seems that under realistic general assumptions a mixed scenario of one DVB-T layer and two T-DAB layers not using adjacent blocks could be realized without problems.

If, however, the decision which system is to be implemented is not taken, this freedom automatically requires to protect the weakest system or better most sensitive type of interaction during the planning process. According to table 8.1 this is the DAB-DVB case which results in the largest re-use distance to be obeyed. As a consequence, the frequency assignment for the two layers of allotment areas discussed here has to be based on the critical distance of 150 km between co-channel areas. Figure 8.11 illustrates the results of a corresponding calculation.

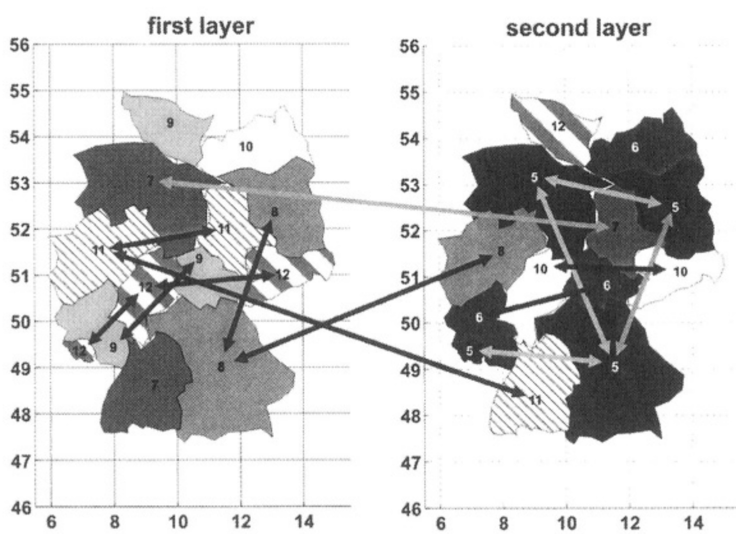

Figure 8.11: Service independent frequency assignment in band III. The employed re-use distance is 150 km. The arrows indicate incompatibilities.

The consequence of applying such a large re-use distance is a very large number of incompatibilities. In some cases there are also adjacent areas which were assigned the same channel. This is clearly prohibitive. Furthermore, in parts the violations of the re-use distance are so severe that they even cannot be compensated for by exploiting sophisticated and hence expensive network implementation strategies.

On the other hand, there are some incompatibilities shown in figure 8.11 that can be removed by fixing the services to be offered in some areas. This concerns for example Nordrhein-Westfalen and Baden-Württemberg which both are meant to use channel 11. The effective distance between these two Bundesländer is about 120 km. If channel 11 would be used exclusively for T-DAB in both areas then according to the re-use distance already employed at Wiesbaden of 81 km there would be no problem at all. It could even turn out that DVB-T networks could be operated in both allotments in channel 11 at the same time without too much mutual interference.

However, a coordinated operation of networks in both Bundesländer would be the prerequisite to do so. This obviously contradicts with the charming advantage of the underlying planning philosophy, namely to separate the process of frequency assignment from the subsequent usage in terms of implementing a particular service independently from any neighboring allotments.

Impact of Different Channel Rasters in VHF

The different spectrum rasters in the VHF range across Europe and the associated implications is one of the central issues that has to be dealt with by the forthcoming revision of the Stockholm plan. A look at figure 4.1 which sketches the most prominent rasters used in Europe underlines that in particular along borders between different countries problems are very likely to be expected. The borderland between France and Germany as well as between the Czech Republic and Germany may serve as examples thereto.

In principle, there are two different options this issue could be dealt with when generating a new frequency plan. Either it is agreed that a common raster is to be used in the future or the status quo is kept. In the case of a common raster it has to be decided whether it should be a 7 MHz or 8 MHz raster in all countries of the planning area.

All possibilities have their own advantages or disadvantages. The 7 MHz variant would allow for a total of eight TV channels in band III while in the case of an 8 MHz raster only seven channels would result.

One channel more would certainly give more freedom for frequency planning purposes whereas a bandwidth of 8 MHz provides a larger data capacity and, in addition, in VHF the same bandwidth would be used as in the UHF range. This could be interesting in the context of wide-area coverage like a coverage of a national territory by one single programme content.

Since a country like Germany probably cannot be served by one single frequency network the area has to be subdivided into smaller regions. According to the principles of the frequency assignment each of these regions would need to use its own frequency to broadcast the same content. Clearly, some of the areas could be assigned VHF frequencies while others are given UHF channels.

A multiplex which is to be broadcast within a whole country by digital terrestrial networks nevertheless will be produced at one single location before the signal is conveyed to the transmitter sites for transmission. However, when generating the digital terrestrial signal the bandwidth determines the details of the modulation process. Therefore, a 7 MHz channel requires a different processing than a 8 MHz channel, i.e. two different types of multiplexes would have to be created. Hence, having a 8 MHz raster both in VHF and UHF would allow in principle to use frequencies of both spectrum ranges without the need for additional signal processing.

On the other hand, if a change-over to one unique raster is decided this would cause a large technical expenditure and financial outlay for a many countries. So, leaving the channel rasters as they are would certainly facilitate the introduction of T-DAB and DVB-T on a national level. The price to pay for that would be the just mentioned problems of overlapping channels along national boundaries.

Since this issue is very complex and cannot be assessed by intuition detailed investigations of all aspects related to the usage of different spectrum rasters are needed. In order to highlight the consequences resulting from a retention of the existing rasters it is sufficient to consider a planning example consisting of one single coverage layer across Europe. Figure 8.12 shows a set of polygons representing a scenario that could be used for DVB-T in Europe.

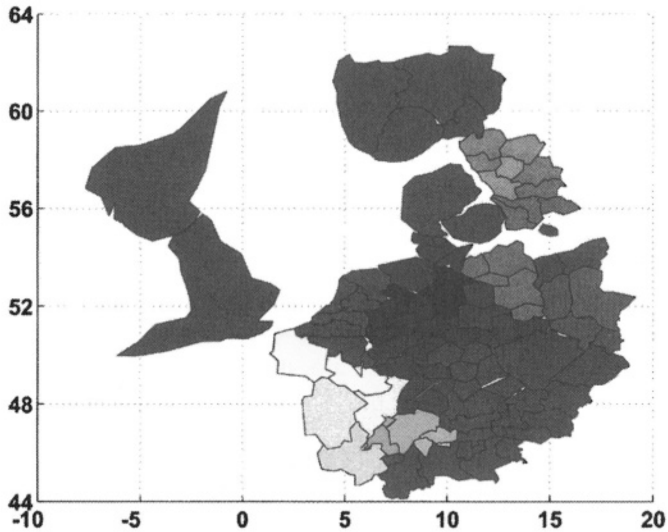

Figure 8.12: Potential DVB-T polygons for central Europe.

Basically, the set of polygons has been taken from the T-DAB allotments of Wiesbaden except the areas for France and Germany. Since at Wiesbaden there were no French VHF requirements the corresponding French allotments in figure 8.12 have been constructed from L-band polygons. In order to build a VHF polygon an appropriate number of L-band polygons have been joined. In contrast, the German coverage areas have been designed from scratch and chosen in line with obvious media political constraints.

However, as in the case of the simple spectrum demand estimation for DVB-T in Europe discussed in section 7.4 the set of polygons used here must not be considered as the official set of DVB-T requirements. They are chosen in order to mimic a particular scenario highlighting the problems arising from different spectrum rasters. This is evident when looking at the British areas which certainly will not coincide with actual DVB-T requirements. They have been chosen that way and included primarily to include the effect of cold sea wave propagation.

For the calculation of a frequency plan the re-use distances from table 8.1 have been taken into account. The set of polygons comprises 103 areas, i.e. there are 5253 different pairs of polygons. A total of 759 thereof are separated by an effective distance which is less than 120 km. It has been assumed that all allotment areas are allowed to use any channel of band III, i.e. channels 5 - 12.

First attempts to establish a working frequency plan which were based on the scheme 8.4 in order to assess the mutual interference potential between two areas did not give satisfying results. This is independent of any assumption about the spectrum rasters used in Europe.

As quality function to quantify an arbitrary distribution of channels onto the allotments, a weighted sum of all interference contributions combined with the explicit number of re-use distance violations has been used. It turned out that in general the number of incompatibilities was relatively small compared to the 759 critical cases. However, the best plan found contained also cases in which adjacent allotment areas were assigned the same channel. Clearly, this is not acceptable.

The reason for that is connected to the principle problem that a frequency plan which on a global scale seems to be rather good might nevertheless exhibit severe problems locally. Such phenomena emerge the more frequently the more complex the assignment problem is. A large number of polygons can create such situations or also very difficult constraints. In these cases the quality criterion employed for the generation of the plan needs to be revised.

Applied to the present example the interpretation is clear. The available VHF spectrum already is overused and consequently the severe violations of the re-use distance have to be accepted. Since this cannot be avoided a revision of the quality function offers the only possibility to find a more acceptable plan.

One idea is to relax the very strict violation criterion. This means to define a limit up to which violations of the re-use distance are considered as being somehow acceptable. They will not create too much problems. Clearly, such a limit cannot be fixed a priori but must be found more or less by trial and error.

An implementation of such a relaxed approach would increase the penalty for any violation of the re-use distance which is larger than the limit drastically. If for example a limit value of 50 % would have been found to be acceptable this means that violations which are smaller than 50 % of the re-use distance would be accounted for in the standard way. As soon as the limits is exceeded the interference value goes up to an arbitrarily large but fixed number. Practically, for the plan generation process the underlying frequency assignments will be considered as being forbidden. It can be expected that as a result the total number of incompatibilities in the plan will increase at the benefit of hopefully avoiding any forbidden assignments.

According to such a strategy a new frequency plan for the set of allotment areas of figure 8.12 has been calculated. As limit separating just acceptable assignments from totally unacceptable ones a value of 50 % has been chosen. Like before, all allotments were assumed to be allowed to make use of any channel in band III. Under these conditions simulations have been carried out for all three options concerning the European spectrum rasters, i.e. 7 MHz or 8 MHz everywhere or a mixture of rasters in Europe.

Due to large number of re-use distance violation still present in the best solutions found it is obvious that a representation of the results as in the case of figures 8.9 - 8.11 is not reasonable. Therefore the results were summarized in terms of histograms as in figure 8.13.

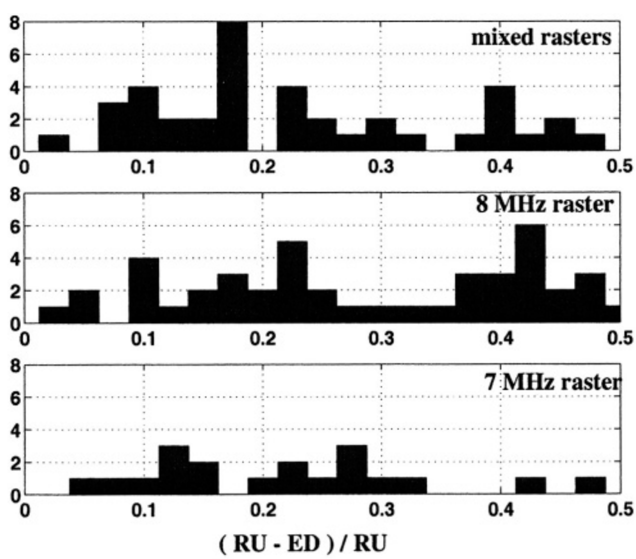

Figure 8.13: Normalized re-use distance violations for the planning scenario of figure 8.12.

In the case of mixture of rasters in Europe the best solution found still has 39 incompatibilities. If 8 MHz channels are used everywhere then 44 co-channel cases with effective distance less than 120 km are left while for the 7 MHz situation there are only 19 collisions.

Related to the total number of 759 polygons pairs having an effective distance less than 120 km the results do not seem to be too bad. The ratios read 5.1 %, 5.8 % and 2.5 %. This figures are quite good compared to typical results obtained for example during the Wiesbaden conference for T-DAB in 1995 when 20 % was a quite common figure.

The histograms 8.13 show however, that there are some cases with re-use distance violations up to 50 %. This means, instead of 120 km the two co-channel allotments are separated by an effective distance of 60 km only.

For a planning conference such a result could be acceptable if the incompatibilities could be removed by bilateral coordination agreements between the administrations of the corresponding countries. However, the question should be addressed what are the reasons for this results and is there a possibility to modify the input requirements or the quality function in order to obtain better results?

To this end, it is helpful to have a look at the adjacency relations between the allotment areas in detail. An important parameter in that respect is certainly the number of areas that have an effective distance from a considered allotment which is less the corresponding re-use distance. From a mathematical point of view this is nothing but the vertex degree discussed in section 5.2.

Calculating the vertex degree for each polygon of the set of figure 8.12 leads to an interesting result. The average degree over all polygons is 15.7. If only the German areas are averaged the value is 18.6 while the mean degree for all non-German polygons reads 13.6. Obviously, the German allotments are surrounded by an above average number of areas. However, as pointed out already in section 5.2.1 the vertex degree is not sufficient to judge the complexity of the frequency assignment problem.

More insight can be achieved by having a closer look on the details behind the incompatibilities. For the mixture of rasters it is observed that 28 of the 39 collision cases involve German polygons whereby 14 interferences are Germany only cases. For the two other scenarios the findings are similar. In the case of the 8 MHz raster even 36 of 44 incompatibilities have German participation. A total of 21 collision is between German polygons only. For the 7 MHz case the figures read 14 German of 19 and 8 German only incompatibilities.

The striking fact of the results is that the involvement of German polygons in collisions is above average. Their central geographical location in the middle of the planning area might be one reason for that. In addition, there it could be that both size and shape of the German allotment have not been chosen in favorable way. Germany is covered by a large number of relatively small allotments compared to the re-use distance of 120 km.

Therefore a new set of polygons has been generated. The German allotment areas have been substituted by the T-DAB allotments from Wiesbaden 1995. Figure 8.14 shows the new set of areas. Thus, the total number of polygons for this example is reduced to 72.

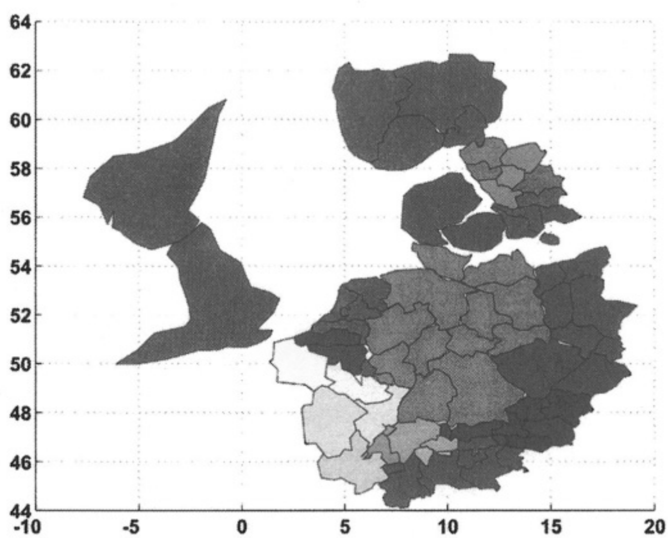

Figure 8.14: New set of DVB-T polygons for central Europe.

The geographical arrangement of the 72 allotment pairs gives rise to 410 pairs of areas having an effective distance that is smaller than the re-use distance of 120 km. A similar calculation as for the layout 8.12 then gives the results depicted in figure 8.15. Most important, the number of incompatibilities decreases to 14, 14 and 4 for the mixed, the 8MHz and the 7 MHz case, respectively. There are no German allotment only collision any more in none of the three cases. The German collision involvement decreases to 4, 6 and 1 cases.

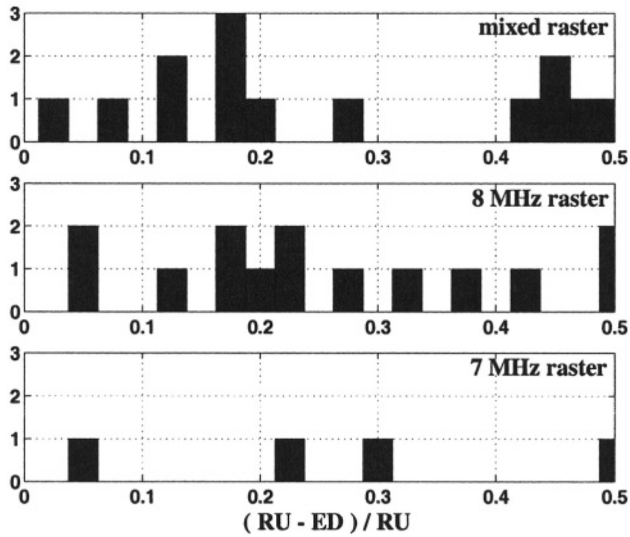

Figure 8.15: Normalized re-use distance violation for the planning scenario of figure 8.14.

Certainly two conclusions can be drawn from the examples above. First of all, it is definitely worth the effort to try to find reasonably designed polygons with respect to their size and the shape. This is a key issue for the derivation of a successful frequency plan. Furthermore, seen from a frequency management point of view it seems to be desirable to go for a unique 7 MHz raster across the planning area in the VHF regime. However, the drawbacks of such an approach in terms of data capacities of network structure have to be discussed very seriously before taking any decisions.

Chapter 9

From Frequency Assignment to Network Planning

Frequency assignment and network planning can be considered as the two sides of the same medal. They are independent from each other on one hand. On the other hand they are closely linked. Usually, frequency assignment is the first process to be completed before any explicit network planning tasks can be taken up. A look at the T-DAB plan from Wiesbaden might illuminate the interrelation between the two aspects.

The CEPT planning conference for T-DAB at Wiesbaden in 1995 finalized a frequency plan which was entirely based on allotment planning. Two coverages for T-DAB could be achieved, one employing band III frequencies while the other has been accommodated in L-band.

At the beginning, it was intended to use primarily channel 12 for the coverage in the VHF range. As it turned out however, this was not possible due to the protection needs of analogue TV transmission in that channel or other non-broadcasting services employing the same frequencies. Therefore, also T-DAB blocks from other channels had to be utilized.

All European administrations taking part in the conference submitted their requirements in terms of polygons which were to represent the coverage areas. With the help of several runs of a specially designed planning software the allotments have been assigned T-DAB blocks in such way to minimize the number of incompatibilities [OLe98].

Since the available spectrum in particular in the VHF range was far from being sufficient to satisfy all requirements a significant number of allotments remained without a frequency assignment. Time-consuming bi- and multi-lateral coordination negotiations had to be carried out to find solutions to these cases all parties could live with.

After having signed the final acts of the conference every administration and every network provider knew which frequencies could be used in which geographical areas in order to build and run the corresponding transmitter networks. Network implementation issues, however, have been dealt with at the Wiesbaden conference only as far as they were connected to the concept of reference networks.

It has to be emphasized explicitly that reference networks should not be considered as examples for real world network implementations. Their one and only purpose is to estimate the interference potential a typical single frequency network is going to produce at a chosen point beyond its proper coverage area. This is the basis on which the re-use distance for co-block usage can be derived.

The final acts of Wiesbaden and the subsequent conference at Bonn which was mainly concerned about coordination procedures in connection with Wiesbaden precisely define the conditions under which the conversion of an allotment into a set of transmitter assignments has to be accomplished. It has to be noted, however, that these guidelines govern only the external impact of the implementation of a single frequency network. Any issues connected to the coverage of the network are not of interest for the coordination procedures.

Apart from the re-use distance reference networks are exploited to define a set of test points surrounding the allotment area. All transmitters inside a particular allotment area are only allowed to produce a limited total field strength at these test points. For the network provider this means that the network configuration can be changed at will as long as the field strength limits at the test points are not exceeded. This concerns the number of utilized transmitters as well as their geographical locations.

Clearly, a closer look will reveal that the freedom how to realize a network is limited, too. In order to comply with the test point criteria it is necessary to choose the transmitter powers and antenna diagrams appropriately. If the resulting network really can be put into operation, i.e. if it is economically feasible is a question that has be answered by the network provider alone.

Every network planning for a terrestrial radio or television system rests on coverage predictions as described in section 3. To this end, an adequate wave propagation model is needed which allows the calculation of the electromagnetic field strength at a specified point of reception. According to the accuracy of the prediction digital terrain data can be exploited or not.

Since as a matter of principle it is not possible to predict the field strength for each mathematical point within a given area statistical methods are employed. Thereby it is possible to describe the variation of the field strength across an extended area on the basis of a set of distinct points of reception inside this area.

But this is not the end of the story. The properties of the receivers have to be taken into account explicitly. For analogue broadcasting systems like FM radio of analogue TV the specification of the field strength is sufficient to allow an assessment of the reception quality. The received signal must be strong enough to be distinguished by the receiver against a noise background produced by several sources including the receiver itself.

In general, there is not one single isolated signal arriving at the location of the receiver. Usually, there are other contributions originating from other transmitters using the same or adjacent frequencies. Under such conditions the ratio between the wanted signal and the sum of all interfering signals must exceed a so-called protection ratio to guarantee satisfying reception quality [IRT82].

Strictly speaking, the situation is not that simple in the case of digital transmission systems. The details of the radio channel directly influence the reception quality. This is connected to the technical characteristics of digital broadcasting systems like for example very efficient error correction mechanisms. There is no simple link between the provided field strength and the reception quality. The magnitude of the received field strength certainly is a first indicator since if it is too small then no error free demodulation of the received signals will be possible anyway.

Even more, relying only on the field strength can lead to completely wrong assessments of the reception quality as measurements have clearly shown. There are cases in which the predicted field strength is large enough to allow for good reception. Nevertheless, unacceptable quality degradations are experienced due to self interference arising quite naturally in multi-path environments.

On the other hand, the reverse situation may be encountered as well. The wave propagation model predicts a field strength which is considered as being not large enough for good reception. Nonetheless, there is good reception. Obviously, there are additional constructive contributions to the signal at the receiver site due to reflections.

Therefore, the assessment of the reception quality can no longer be based on the field strength alone as quality measure. A new appropriate quantity like the bit error rate before the application of the error correction algorithms [Beu98a] should be employed instead.

In order to incorporate the bit error rate future prediction methods have to take into account all details of the transmission channel. In the case of single frequency networks this means the number of impinging signals, their amplitudes, angle of incidence and relative time of arrival need to be known. Clearly, this implies the demand for better wave propagation models with particular emphasis on three-dimensional models.

Apart from the need to modify the quality measure for the coverage predictions of digital terrestrial broadcasting systems there are further fundamental differences between analogue and digital systems in terms of network planning. Generally, FM radio and analogue TV have to be planned as multi frequency networks. This means that in order to offer one particular programme throughout a large area like a whole national territory a multitude of different frequencies has to be utilized.

Even if in band II, the frequency range where FM radio is accommodated, where the wave propagation conditions are fair enough to cover larger areas, it is necessary to make use of several transmitters. In particular, in regions with high mountains and deep valleys quite a large number of transmitters has to be put into operation.

In those reception areas where two or more transmitters produce sufficiently high field strengths simultaneously the corresponding signals interfere. If they are spectrally separated far enough the receiver is able to tune to one of them and discriminate the rest. The quality of the corresponding filters distinguishes good from bad receivers. For planning purposes assumptions have to be made about the features of a typical receiver. This leads to certain protection ratios to be obeyed by wanted and unwanted signals in order to guarantee good reception (see for example [IRT82]).

If all analogue transmitters would broadcast a programme using the same frequency the superposition of several signals could not be broken down into separate parts. Since in general each signal will arrive at the receiver site at a different time relative to the others, no demodulation of the signals would be possible any more. This leads to the usage of different frequencies for different transmitters.

To give an example the German Bundesland Baden-Württemberg can be considered. It occupies an area of about 200 km × 250 km. In order to provide one single FM programme throughout this topographically difficult region a total of more than ten high power transmitters are needed using different FM frequencies.

In contrast, T-DAB and DVB-T offer the possibility to implement single frequency networks. In order to broadcast one T-DAB or DVB-T multiplex within Baden-Württemberg the same amount of spectrum is employed everywhere, either one T-DAB block or one TV channel.

Clearly, also in that scenario there are geographical regions where the coverage areas of individual transmitters overlap. There the signals from different transmitters are superimposed as well. However, due to the features of the digital broadcasting systems this superpositions can be exploited in a constructive way (see section 2.1). This capability is limited by the magnitude of the guard interval. Signals which arrive at a time interval significantly larger than the guard interval give rise to interference, too.

Just in topographically difficult regions the concept of single frequency network offers an important advantage. If the direct path between transmitter and receiver is blocked by some obstacle like a mountain then it is still possible that a satisfying reception quality can be achieved with the help of reflections. In the case of analogue systems such a multi-path environment will lead to severe service quality degradations.

FM radio and in particular analogue TV have been planned for stationary reception using a directional antenna. The customer was expected to accept some technical and financial expenditure to have a good reception quality. Therefore, all planning activities assumed a receiving antenna height of 10 m above ground and an antenna gain of 6 dB.

At the time FM radio was introduced it could not be foreseen that mobile reception in vehicles has become as important and common as well as portable indoor reception. Since the receiver manufactures made large progress in terms of providing steadily increased receiver quality both mobile and portable indoor reception are considered as being satisfying today.

However, this is not true for terrestrial analogue television. Still a directional antenna is the prerequisite for acceptable reception quality. In the vicinity of a television transmitter it is possible to use small portable television receivers equipped with a short rod antenna. However, mobile reception in vehicles usually does not work. Neither for FM radio nor for analogue TV there are different operation modes adapted to different reception conditions as is the case for their digital counterparts.

The situation is completely different for T-DAB and DVB-T. Digital radio has been designed for mobile reception from the very beginning. To this end, part of the potential data capacity has been sacrificed in favor of more robustness against perturbations during the transmission. In principle, also DVB-T can be used for mobile reception even though significantly higher requirements with respect to minimum field strength and protection ratios have to be met [EBU02].

As has been shown in sections 2.2 and 2.3 there are different operation modes for T-DAB and DVB-T. In particular in the case of DVB-T there is a great variety of possibilities that require different minimum field strengths and protection ratios, respectively. Conceptually, they differ by the potential data rate and the error protection required thereto.

Apart from mobile reception other planning targets are feasible as well. Both portable and stationary reception are being considered independently. In the case of portable reception it is furthermore distinguished between indoor and outdoor reception. In the documents [CEP95], [CEP96], [CEP97] and [ITU02] all required minimum field strengths and protection ratios for all possible interactions between system variants are given in detail.

However, most of them still have to be considered as being preliminary. They are mainly based on simulations rather than on measurements.

There is some experimental evidence that in particular before the revision of the Stockholm plan these planning parameters have to be critically reviewed. Especially in relation with current receiver developments it has to be clarified if these values can be used as basis for the new digital frequency plan or not.

Chapter 10

Planning Principles for Single Frequency Networks

Seen from a technical point of view, T-DAB and DVB-T rest on the same principles like for example COFDM. Nevertheless, there are significant differences between the two systems as has been discussed in detail already. They differ with respect to the occupied bandwidth and the different possible operation modes.

This has direct consequences for the frequency assignment process. As discussed, each operation mode requires strictly speaking its own re-use distances. Furthermore, the interaction between different system variants has to be accounted for explicitly. Hence, the great diversity of planning parameters imposes very complex constraints on any frequency assignment process.

In contrast thereto, there are no principle distinctions between the two systems with regard to the implementation of single frequency networks. It is true that different minimum field strengths and protection ratios are to be accounted for. However, the basics of the treatment of single frequency networks for T-DAB and DVB-T are the same. Therefore, it is possible to explain network planning approaches almost without referring to the characteristics of one of the systems. If in the sequel special properties of T-DAB or DVB-T are needed this is highlighted in each case explicitly.

Each network planning is based on precisely defined coverage targets. Before starting any network planning effort it has to be decided if fixed, portable or mobile reception is to be considered. Furthermore, it has to be specified which geographical areas are to be covered. Exactly this issue has had and still has outstanding significance. It is obvious that this point differentiates the views of public and private broadcasters.

During the introduction of a digital terrestrial broadcasting system it is evident that the focus is on densely populated urban areas. In case mobile reception is a relevant issue then quite clearly all major traffic paths possess high priority, too. Thus, all T-DAB pilot projects in Germany at the end of the last century were put into operation correspondingly.

Such coverage targets fit very well with the interests of private broadcasters. Usually, they are not trying to reach wide-area coverage. Most people live in urban areas and consequently this is where the addresses of advertisement are found by which these companies are financed. Public broadcasters on the other hand must fulfil the task to provide a wide-area coverage for everybody.

However, from a network planner's point of view both objectives have one thing in common, namely that in any case the definition of an envisaged coverage area is the starting point of the network planning process. In general, the intended coverage area cannot be served by one single transmitter alone. This is even more valid the more complicated in terms of topology the planning area proves to be. Typical examples thereto can be found in southern Germany or the alpine regions.

The first and essential question that needs to be answered in relation to the network implementation concerns the search and choice of adequate transmitter sites. For every network provider it would be desirable if as few as possible transmitters would be required to establish a satisfying coverage. The more effort has to be put into the network infrastructure the more complicated and economically unalluring the setting up of the network and its subsequent operation will be.

Therefore it would be optimal if during the network planning process the choice of the transmitter sites would be a free choice. There is no principle problem to determine the geographical location of the transmitters with the help of appropriate optimization strategies as for example described in section A. Actually, this is just a question of how much effort is acceptable when setting up a mathematical model of the entire problem.

Seen from a practical perspective such an approach has to be considered as being completely hypothetical. The opening up of a high power transmitter site is a very cost-intensive issue. Acquisition costs for an appropriate premises or the costs for rent if it is not possible to buy it, construction costs or expenses for maintenance and many more beyond the pure telecommunication sector have to be taken into account. Therefore, such an investment requires sincere consideration.

Furthermore, electromagnetic compatibility issues gain more and more importance when trying to find new transmitter sites. Currently, this still affects mainly network providers for mobile communication. However, public broadcasters are beginning to feel the impacts of the ongoing discussion on the business as well. Moreover, since quite some time there is also a discussion about a drastic reduction of the threshold values for electromagnetic exposition. If this is really going to be finalized the terrestrial distribution of radio and television would be existentially affected.

It follows quite naturally form all these facts that a network provider will be using existing network infrastructure in first place to provide the wanted coverage. Only after detailed investigations showing that more transmitters are required new investments will be considered. If the license for the operation of the transmitter network allows to cover only a smaller part this could be an option, too. However, this degree of freedom is not always given. In the case of digital broadcasting networks the regulatory framework that controls the licensing policy usually prescribes strict time schedules for the expansion of the networks. Defined percentages of area coverage are usually to be achieved within certain given periods of time.

In most cases the implementation of a digital terrestrial network starts with the collection of all potentially useful transmitter sites from which a subset is chosen according to economic and coverage related issues. If the number of available transmitter sites is large, the selection of the subset might be very complicated.

Seen from a mathematical point of view this is again a combinatorial optimization problem of a type that in connection with frequency assignment problems has been encountered several times so far. The problem of choosing the optimal set of transmitters will be dealt with in the sequel.

By selecting the transmitters the planning process for a single frequency network is not finished. Even if the geographical locations of all transmitters are fixed there are still transmitting power and antenna diagram which need to be adjusted in each case. These two technical parameters determine the coverage of a network to a large extend.

However, for digital terrestrial single frequency networks typically utilized by T-DAB and DVB-T there is in addition another degree of freedom which can be employed for the design of the networks. Both, T-DAB and DVB-T are very well adapted to cope with multi-path environments. As long as the individual signal contributions from different transmitters arrive within a time interval not exceeding the guard interval, all of them can give rise to constructive superposition.

Thus, since the system design of T-DAB and DVB-T already offers the means to cope with several subsequently arriving signals, it is natural to exploit this feature in order to compensate for potential perturbations by selfinterference. To this end, the temporal synchronization of the transmitters in the single frequency network can be adjusted to reduce the temporal spread of the times of arrival to values smaller than the guard interval. Thereby it is possible to avoid selfinterference at points of reception where otherwise problems would occur. Compared to analogue transmitter networks this is a completely new degree of freedom that can be utilized to optimize the network design.

So far designing a transmitter network has been discussed from the perspective of reaching a certain coverage only. However, this is only one side of the medal. Any network that serves a particular area inevitably creates interference at points outside its proper coverage area. This aspect has to be taken into account during the implementation phase of the networks as well.

As has been described in the previous sections, reference networks can be used to estimate the outgoing interference of a typical network designed to fulfil a defined coverage target. To this end, every allotment is accompanied by a set of test points outside the actual allotment area. When putting into operation one or more transmitters inside the allotment area it has to be checked that the total field strength produced by the network at these test points does not exceed fixed limits. This way it is guaranteed that the implicit assumptions about outgoing interference of a typical network under which the frequency assignment has been carried out are not violated.

The process of configuring a network with respect to its test points is usually called coordination. Subsequent to the CEPT conference of Wiesbaden another meeting was held at Bonn, Germany, in 1996 which covered the main coordination issues in relation to T-DAB [CEP96]. Complementary, the final acts of the Chester CEPT meeting in 1997 [CEP97] contain precise coordination instructions how manage the transition from analogue to digital television on the basis of the Stockholm plan from 1961.

Coordination problems are not covered here. First of all, this would require a significantly larger effort in terms of modelling an appropriate example. Each network planning example would need to be supplemented by a corresponding external environment in order grasp and study effects which are solely coordination based. This would hamper a transparent presentation of the planning methods and approaches without adding too much to the understanding.

In the case of network planning the consideration of coordination constraints follows the same mechanisms as an isolated network planning example. Only the number of constraints to be taken into account is increased. As a consequence, a significantly higher numerical effort would result. The strategy itself is thereby not touched upon.

10.1 Model of a Single Frequency Network

In contrast to frequency management, network planning is a task that each network provider accomplishes reclusively. Apart from the need to meet the coordination constraints at the location of the test points, the internal configuration of single frequency networks can be dealt with independent from adjacent countries or other networks.

Similar to the receiver manufactures which all exploit the same technical standards like T-DAB or DVB-T yet offering different products, reasonable network planning constitutes the core competence of a successful network provider. Thus, it is evident that on one hand there is only a limited number of openly available publications. On the other hand, the technical details of a network are classified.

Relevant literature can be found for example on the website [ZIB01]. There, however, the focus lies on mobile communication and not on broadcasting. The dissertation [Que93] deals with network planning by employing a strict graph theoretical approach. It is mainly concerned about FM radio networks while in [Beu95] and [Beu98b] planning of digital single frequency networks is treated. The discussion below follows the outline given in the latter two documents.

Each model of a transmitter network rests on four pillars. First of all, an adequate wave propagation model is needed which allows to calculate the field strength a transmitter produces at a given point of reception. Furthermore, the characteristics of the receivers under multi-path propagation conditions must be known at least qualitatively. Then, the target coverage area is to be defined appropriately. Finally, the set of transmitters forming the basis of the network have to be specified by fixing all relevant parameters.

10.1.1 Wave Propagation

In particular single frequency networks benefit from the fact that at a given point of reception not only one single signal is received but a number of signals. These originate either from different transmitters or are caused by reflections. The superposition of all contributions can lead to better coverage.

This requires to take into account two important aspects with respect to the prediction of the field strength. One point is related to the occurrence of reflections or echoes and the other concerns the selection of reception points for which the field strength is predicted.

Reflections are a consequence of the electromagnetic wave propagation under typical topographical conditions where mountains or hills exist (see section 3). If echoes are important for the received signal strength the three-dimensional wave propagation models should preferably be employed.

The development of such models has been subject to scientific research for a long time already. In the meantime, there are several models available which differ in particular with respect to the run times of the corresponding software (see for example [Leb92] and [Gro95]). Due to the mostly unacceptable large computation times these algorithms cannot be used for wide-area predictions on daily basis.

Therefore, mainly two-dimensional wave propagation models are employed thereto. They evaluate only the topography along the line connecting the location of the transmitter and the receiver (see [Oku68], [Cau82], [Mee83], [Gro86] or [Lon68]). Heuristic models like [ITU01a] are only of little help in that context.

The prediction of field strengths in the case of single frequency networks requires the utilized wave propagation models to be so flexible that they can be applied for any arbitrary point of reception. This demand becomes essential in relation to the selection of the set of reception points that should represent the coverage area.

The background is that all existing wave propagation models can make statements only for single points and not for areas. Thus in practice, an area has to be approximated by a large number of individual points. The more points are used, i.e. the smaller the gaps are between them the more accurate the results usually are.

However, is must be possible to change that resolution. Quite often there are situations in which for a small part of the original area a higher resolution is required, for example if an urban area is included. Wave propagation models should be able to offer such modifications without problems.

Another important issue in relation to the calculation of field strengths concerns the geographical location of the considered points of reception that represent the explored coverage area. In most cases a set of points is used which form a regular grid. For investigations that have to deal with area coverages where each point is to be treated on an equal footing this is a common and reliable approach.

If, however, other coverage targets are considered like for example the coverage of densely populated areas or other specially chosen parts of a national territory then it is advantageous to base the calculations on irregular distributions of reception points. In section 10.1.3 this aspect is deepened.

Current wave propagation models provide information about the field strength value produced by the network at a chosen point of reception. In view of the technical characteristics of COFDM transmission systems it would be desirable if the output of the wave propagation model would comprise the details of the radio channel as a function of time of arrival and angle of incidence of the signals. This would certainly allow for a more accurate determination of the field strength to be expected (see section 3.4.4).

10.1.2 Receiver Models

The second important item that has to be incorporated into a model of a single frequency network touches upon the question how the receiver responds to a multi-path environment and how the received signal is evaluated.

In order to demodulate and decode the transmitted information the receiver has to synchronize to the received signal, i.e. to the temporal succession of T-DAB or DVB-T symbols. The signal is evaluated by periodically recording a set of samples within a time interval of duration T_W.

If there is only one signal contribution then the synchronization strategy of the receiver is obvious. The evaluation window has to be aligned with the arriving signal. However, in a multi-path environment the way this synchronization process is carried out needs some care. In that case there is a variety of different signal contributions arriving at the point of reception at different times.

If the time interval between the first and the last signal is smaller than the guard interval a system like T-DAB or DVB-T can benefit from the superposition of several signals. In case the first and the latest arriving signals are separated by a time larger than the guard interval both constructive and destructive effects will result.

Seen from a physical point of view, the set of different signals are linearly superimposed at the point of reception according to their amplitudes and phases. Both quantities are subject to the details of the wave propagation from the transmitters to the receiver.

The correct way to deal with a multi-path environment in terms of a wave propagation model would be to add all signals coherently, i.e. fully taking into account their amplitudes and phases. Then, the resulting signal would have to be assessed. However, there is no wave propagation model so far that is able to provide the necessary information thereto.

First of all, the exact determination of the phases is not feasible. To make that statement clear it is reminded to the rule of thumb that any object in the order of the wave length has to be considered as a potential scatterer for a propagating wave. In the VHF range wave lengths of one to two meters are found.

As a consequence, a topographic data base that is supposed to fully cover this would need to possess a resolution of less than one meter. There are certainly data bases of that type available today, but it is completely impossible to rest wave propagation calculations for a distance of 100 km on such a data base. Computation times would definitely explode. Furthermore, the scattering process of the electromagnetic wave would need to be captured on all scales exactly. This is clearly not possible and therefore the principle inaccuracy of typical wave propagation models does not allow to calculate the phases of the waves.

Since the physically correct way to deal with a multi-path environment is actually not possible alternatives are required. The standard approach usually adopted operates on the level of signal powers and assumes the arriving signals can be virtually subdivided into two parts, a wanted or useful power C_k and a corresponding unwanted or interfering power I_k. The values of C_k and I_k depend on the time of arrival of the k-th signal relative to the position of the evaluation window. Figure 10.1 sketches the situation.

From looking at figure 10.1 it becomes clear that all black parts can be counted constructively. In the case of the third signal only a fraction can be added to the wanted signal. The rest increases the unwanted part. The different contributions to the wanted and unwanted signals are taken into account according to appropriate weighting factors. These weighting functions are different for T-DAB and DVB-T. The reason is that the processing of the signals by the corresponding receivers is slightly different. More details can be found in [EBU98] and [ERC98].

In order to apply the weighting functions it is necessary to specify the synchronization strategy of the receiver. This means how the receiver positions its evaluation window of duration T_W relative to the temporal layout of the incoming signals. In figure 10.1 the receiver synchronizes to the first arriving signal. Without restriction the time t_1 can be put equal

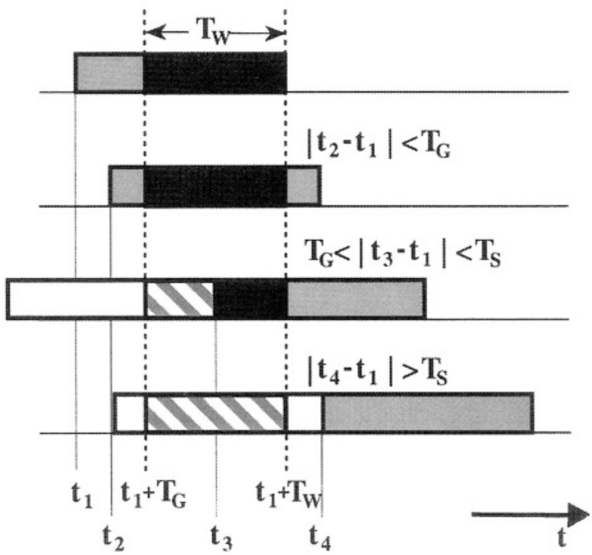

Figure 10.1: Constructive and destructive contribution of different signals depend on their relative time of arrival with respect to the position of the receiver evaluation window. Black areas contribute constructively while the dashed parts give rise to selfinterference.

to zero. Then for T-DAB the weighting factors f_N and f_S for wanted and unwanted as a function of the relative time of arrival read

$$f^{DAB}_{wanted} = \begin{cases} 0 & , \quad t_k < -T_W \\ \left(\dfrac{T_S + t_k - T_G}{T_W}\right)^2 & , \quad -T_W \le t_k < 0 \\ 1 & , \quad 0 \le t_k < T_G \\ \left(\dfrac{T_S - t_k}{T_W}\right)^2 & , \quad T_G \le t_k < T_S \\ 0 & , \quad T_S \le t_k \end{cases} \tag{10.1}$$

and

$$f_{unwanted}^{DAB} = \begin{cases} 1 & , \quad t_k < -T_W \\ 1 - \left(\dfrac{T_S + t_k - T_G}{T_W}\right)^2 & , \quad -T_W \le t_k < 0 \\ 0 & , \quad 0 \le t_k < T_G \\ 1 - \left(\dfrac{T_S - t_k}{T_W}\right)^2 & , \quad T_G \le t_k < T_S \\ 1 & , \quad T_S \le t_k \end{cases} \tag{10.2}$$

Figure 10.2 shows the curves corresponding to the definitions (10.1) and (10.2).

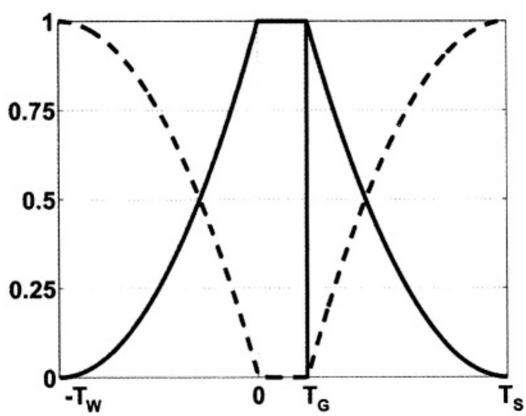

Figure 10.2: T-DAB weighting factors for wanted (solid) and unwanted (dashed) contributions of delayed signals.

Clearly, other receiver synchronization strategies are possible, too. They are already implemented in receivers on the market. The positioning of the evaluation window by aligning to the first arriving signal can be a very bad choice indeed.

This might happen if for example the direct path between receiver and the nearest transmitter is blocked by an obstacle. Even if there is still some power arriving along that way the attenuation suffered by the presence of the obstacle might be so large that the main contribution comes from a different transmitter site. If its corresponding time delay is larger than the guard interval severe degradation of the reception quality would result up to a total failure.

An obvious approach to exploit the arriving signal in a more efficient way is to position the evaluation window according to the strongest contribution. However, if this happens to be a single signal that is separated by a lap of time larger than the guard interval from a group of other signals then once again selfinterference might occur.

The optimal strategy with respect to improving the reception quality as much as possible certainly is to adjust the evaluation window such that the ratio between useful and interfering contributions is maximal. The drawback of this method is the need for more complex receivers and corresponding sophisticated algorithms to find this position of the evaluation window. More details on receiver synchronization strategies can be found in [EBU03].

Whatever synchronization strategy is employed by the receiver it is evident that this is one of the features which differentiates good and bad receivers. This is exactly the issues where the know-how and the experience of the manufacturer becomes relevant both in terms of performance as well as costs of production of the receivers. Hence, it come as no surprise that there is nearly no information available about the actually implemented synchronization strategies.

Simulations and hardware tests of DVB-T systems have shown that the receiver model appropriate for T-DAB according to (10.1) and (10.2) cannot be applied to DVB-T as well. The main reason lies in the fact that both systems use different modulation techniques. T-DAB employes a differential QPSK modulation scheme while DVB-T rests on coherent QPSK or QAM modulations (see section 2).

The coherent demodulation in the case of DVB-T requires an estimation of the channel response function. The pilot carriers are utilized thereto. Strictly speaking, the impulse response function then is only known for those frequencies the pilot carriers are located at. Therefore, a spectral interpolation is carried out which includes filtering. The corresponding filter bandwidth leads to a characteristic time interval T_F that influences the receiver properties in the presence of interfering components

[ERC98]. Echoes arriving later than T_F thus introduce degradations in the determination of the impulse response and thus contribute only in a destructive manner.

Like in the case of T-DAB, signals having a spread of their times of arrival smaller than the guard interval are considered to contribute constructively. Signals arriving later but before T_F are also split into wanted and unwanted parts taken into account adequate weighting factors. All contributions with a time of arrival beyond the filter time T_F are considered to be interfering only. In mathematical terms the weighting functions can be written as

$$f_{wanted}^{DVB} = \begin{cases} 0 & , \quad t_k < -T_F \\ \left(\dfrac{T_S + t_k - T_G}{T_W}\right)^2 & , \quad -T_F + T_G \leq t_k < 0 \\ 1 & , \quad 0 \leq t_k < T_G \\ \left(\dfrac{T_S - t_k}{T_W}\right)^2 & , \quad T_G \leq t_k < T_F \\ 0 & , \quad T_F \leq t_k \end{cases} \tag{10.3}$$

and

$$f_{unwanted}^{DVB} = \begin{cases} 1 & , \quad t_k < -T_F \\ 1 - \left(\dfrac{T_S + t_k - T_G}{T_W}\right)^2 & , \quad -T_F + T_G \leq t_k < 0 \\ 0 & , \quad 0 \leq t_k < T_G \\ 1 - \left(\dfrac{T_S - t_k}{T_W}\right)^2 & , \quad T_G \leq t_k < T_F \\ 1 & , \quad T_F \leq t_k \end{cases} \tag{10.4}$$

Figure 10.3 shows the curves corresponding to the definitions (10.3) and (10.4).

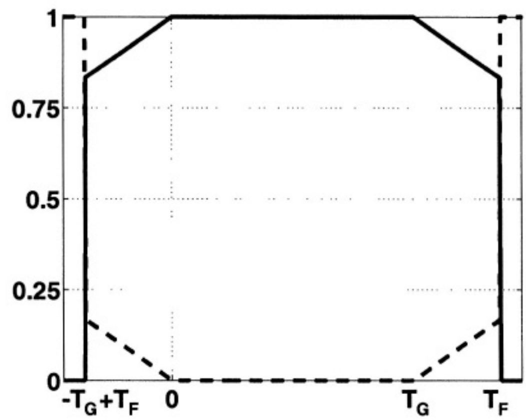

Figure 10.3: DVB-T weighting factors for wanted (solid) and unwanted (dashed) contributions of delayed signals.

10.1.3 Envisaged Coverage Area

The main target for most network planning activities in broadcasting is to provide service for a defined area. As discussed before, it is not possible to investigate the coverage situation at every mathematical point inside a considered area. Therefore, the problem has to be reduced to include only a limited set of so-called test points. Their number and geographical location is determined by the principle coverage target of the network provider. Public broadcasters are requested to cover the whole area while private broadcasters can focus on specially chosen parts or regions.

In order to deal with a wide-area coverage like an entire national territory an equidistant grid comes forward as basis for the test point distribution. The location of the test points then coincides with the grid points. Obviously, the number of test points that fall into a particular coverage area depends on the horizontal and vertical distances of the grid points.

The choice of these two lattice constants is critical with respect to the computational effort when evaluating the coverage situation throughout the considered area. Too small lattice constants will lead to unacceptably high computer run times whereas too large values will restrict the worthiness of the coverage prediction. Thus, a reasonable trade-off between computational time and accuracy must be achieved. Figure 10.4 shows an example of a realistic grid of test points for the German Bundesland Baden-Württemberg.

In this particular case both in horizontal and vertical direction 50 points have been chosen. This gives a total of 2500 points to represent the coverage area. They have been aligned equidistantly with regard to their geographical coordinates. The area is limited by 7.5° and 10.5° in longitudinal direction whereas the latitudes vary between 47.5° and 50.0°. This leads to different absolute distances between grid points in east-west direction and in north-south direction.

Clearly, another definition of the lattice could be employed as well. First of all the, it is not necessary to use the geographical coordinates to generate the grid. With the same right the geometrical distance between points could be fixed to for example 1 km. Furthermore, those test points lying outside the boundaries of Baden-Württemberg could be neglected. However, a rectangular structure as seen in figure 10.4 allows to produce faster computational software due to simpler access to the set of test points.

However, selecting the test points according to a grid neglects completely that different points within an area can be of totally different importance for the network provider. Even a public broadcaster whose primary obligation is to provide wide-area coverage might distinguish between test points being located in urban areas or in the middle of a forest. The proximity of a test point to main traffic routes might be relevant, too.

Moreover, some times certain test points have to be taken into consideration because the topography in the vicinity has special characteristics that seem to be important. Also, it might be necessary to take into account certain test points because they have economic or political relevance.

Apart from the major traffic routes the population distribution certainly is one of the most important factors that can determine the selection of a set of test points. Employing a grid surely is very well adapted to wide-area coverage problems. This is not valid for population density oriented planning activities.

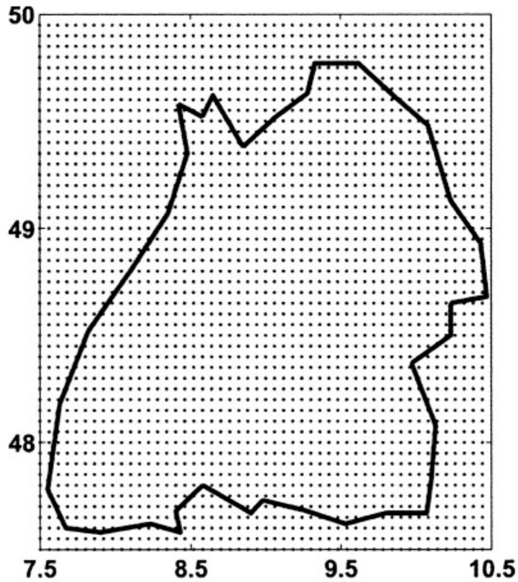

Figure 10.4: Example for a set of test points based on a rectangular grid as basis for network planning in Baden-Württemberg, Germany.

Each test point represents an area of approximately 200 m × 200 m. This is connected to the philosophy of commonly used coverage prediction methods (see section 3). Thus, each test point can be endowed by the number of people that happen to live inside its associated area. To this end, population data bases with a large enough resolution are needed.

It is evident that in the case of a grid based test point generation it might turn out that a large number of test points are taken into account which are located in areas where only very few or even no customers live. Obviously, any network planning based on such data will not be optimal.

Especially private broadcasters are interested in providing coverage only in those areas with high population densities like urban areas. In order to configure the network to suit optimal to this target it is reasonable to base the coverage prediction on a set of test points that maps the coverage target.

One way to generate such a set of points could be to define a threshold and compare the number of inhabitants associated to each point with this value. If the actual figure is larger than the threshold than the corresponding point is included in the set of test points, if not it is neglected.

Applied to the example of figure 10.4 a threshold value of 30 results in a more or less uniform distribution across the area of Baden-Württemberg. A corresponding example is shown in figure 10.5. If the threshold is chosen significantly larger like for example 300, then only points belonging to the highly populated urban areas in Baden-Württemberg are left. This can be seen in figure 10.6.

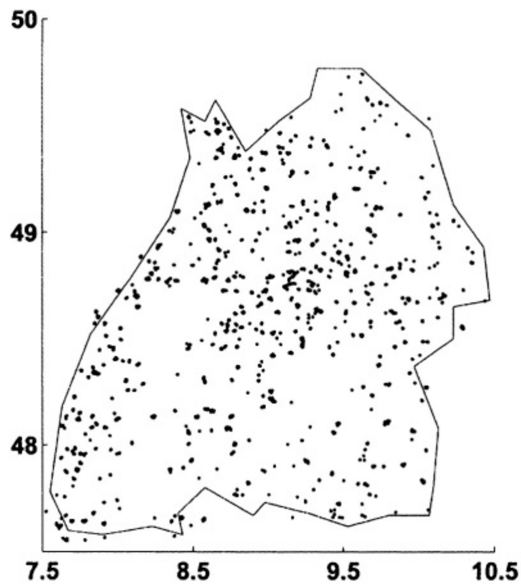

Figure 10.5: Example for a set of test points based on population data. Only points with a moderate number of inhabitants are accepted.

10.1.4 Network Parameter

Apart from the wave propagation model, a selection of test points and the specification the receiver properties the characteristics of the transmitters of the network have to be defined as well. A single frequency

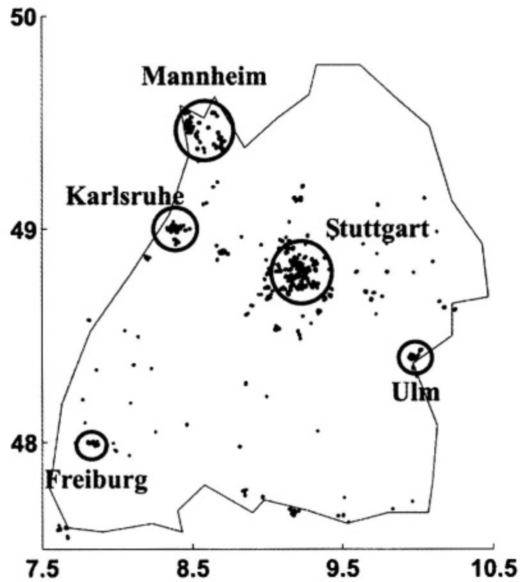

Figure 10.6: Example for a set of test points based on population data. Only points with a large number of inhabitants are accepted.

network is built from a set of N transmitters. They are located at the positions $\mathbf{r}_k$ $(k = 1,..,N)$.

Each of the transmitters is operated at an individually adjustable radiation power. The antenna is fed with the input power P_I which is then transformed into the radiation power P_T according to the efficiency $\eta = P_T/P_I$. The quantity η accounts also for cable losses. The transmission generates a power density $S(\mathbf{r})$ at a distant point $\mathbf{r}$ whose value depends on the distance to the antenna and as discussed before on the topographic conditions.

Another important factor determines the field strength in practice, too. Every antenna employed gives rise to a characteristic there-dimensional angular antenna pattern. The most simple antenna pattern is that of a three-dimensional isotropic radiator. In that case, the radiated power is released uniformly in each direction.

However, spherical symmetry is not desirable for broadcasting purposes.

The transmission should preferably take place in horizontal direction. A slight tilt towards the earth's surface is even better. Furthermore, most transmissions are subject to certain restricting constraints in the sense that the output power towards specified directions must not exceed limits imposed by the presence of other radio services.

With the help of proper antenna design it is possible to bundle the available power towards defined directions at the expense of others (see for example [Bal82]). This non-isotropic transmission is described in terms of the so-called antenna gain. It is defined as the ratio between the power density S of the considered antenna into the boresight direction and the power density of an ideal isotropic radiator S_i under the condition that both are virtually fed by the same input power P_I.

Radio or television antennas are usually based on dipoles, a set or dipoles or related designs. Therefore, it is reasonable to relate the antenna gain not to the isotropic radiator but to the power density S_d of the main lobe of an ideal half wave dipole [Kat89]. The definition of the antenna gain used in broadcasting thus reads

$$G_d = \frac{S}{S_d} \; . \tag{10.5}$$

The antenna gain is a global characterization of an antenna. Basically, it is a first indication for the directivity of the antenna. To provide the full information about the angular behavior it is, however, necessary to make reference to the complete antenna pattern or diagram. Usually, two types of antenna patterns are given, one in the horizontal plane and one in the plane perpendicular thereto. The values of the antenna diagrams represent a measure for the power emitted into a particular direction. They are normalized to the boresight direction. Thus, in logarithmic representation only negative values appear.

To establish a unique characterization of the transmission power of different transmitters the half wave dipole is utilized thereto as reference as in the case of the antenna gain. The so-called effective radiated power (ERP) is just the product between the input power and the antenna gain:

$$\mathrm{ERP} = P_I * G_d \; . \tag{10.6}$$

Usually, logarithmic representation is used for the ERP as well,

$$\mathrm{ERP}_{\log} = 10 * \log(\frac{\mathrm{ERP}}{P_{\mathrm{ref}}})$$

where for the reference power P_{ref} both 1 kW as well as 1 W are used in practice. The statement of 3 dBkW thus corresponds to an ERP of 2 kW.

The isotropic radiator mentioned above cannot be realized as transmitting antenna. Even an antenna whose pattern shows a circular geometry in the horizontal plane cannot be put into practice without tremendous effort. Significant deviations from the circular shape have to be accepted usually. Also the implementation of an exact half wave dipole remains an exception.

Real antennas consists of dipoles and Yagi antennas which are combined appropriately to approximate the desired theoretical antenna pattern as good as possible. Exact antenna diagrams are usually not known. As a rule, numerical simulations are carried out for a particular antenna configuration.

Sometimes, if the effort is worthwhile for example in the case of antennas of high power stations, measurements with the help of helicopters are pursued in order to improve the knowledge about the actual transmission characteristics. Experience showed that unfortunately in most cases there is a large discrepancy between antenna patterns employed for theoretical planning and those actually employed in practice. Therefore, it has to be borne in mind that theoretical antenna diagrams have to be considered as rough approximations of the real patterns only.

Usually, for network planning purposes only horizontal antenna diagrams are taken into account. Vertical patterns are considered only for special case studies. The angular dependence of the horizontal radiation patterns are approximated by specifying thirty-six parameters A_k attached to angles from zero degrees to 350° separated by an interval of 10°. The north direction coincides with the angle 0° increasing with mathematically negative orientation. For angles between that are not equal to a multitude of 10° a linear interpolation is applied. Figure 10.7 shows a typical diagram that is used for planning.

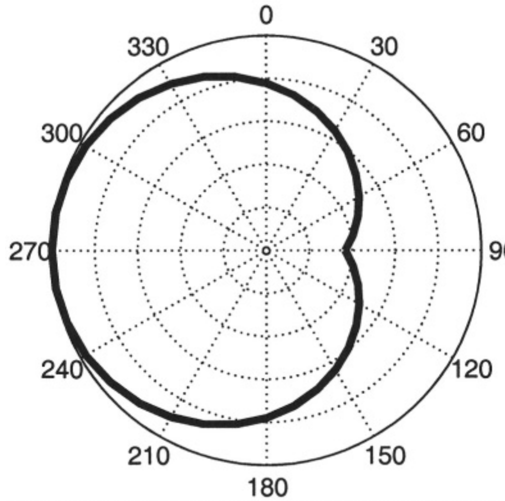

Figure 10.7: Example for an antenna diagram to be used for planning. The scale is 2 dB, so the maximum reduction is about 6 dB towards east.

There are two basic constraints to be taken into account when generating theoretical antenna diagrams. In practice, reductions beyond 20 dB cannot be realized without very large effort. When bringing into operation any transmitter its compatibility with already existing transmitters has to be checked. If during the corresponding coordination procedures reductions of that order of magnitude result, the only solution is to reduce the total ERP of the considered transmitter.

The second constraint to be met when designing theoretical diagrams is to take care that the reduction as a function of the angle does not exhibit too strong changes, i.e. the corresponding gradient needs to be sufficiently small. Cast into practical terms, this means that a difference between adjacent antenna diagram values must not exceed 5 dB.

In contrast to analogue transmitter networks there exists another important parameter in the case of T-DAB and DVB-T single frequency networks which primarily determines the reception quality. This crucial feature is connected to the way COFDM systems can cope with the special characteristics of multi-path transmission channels. Typically, under such conditions more than one signal contributions arrive at the point of reception (see section 3.1).

In the case of FM radio two signals caused by reflection and thus arriving at different times at the receiver will lead to service degradations due to their interference which cannot be overcome by the FM receiver. Only in the very unlikely situation that they arrive exactly at the same time they add constructively. Just a very small deviation from the contemporaneous arrival creates perturbations. The larger the difference between their times of arrival becomes the more serious the problems gets.

For digital terrestrial broadcasting systems based on COFDM this problem does not exist up to certain limits beyond which also problems appear. However, the systems design was carried out exactly such as to compensate for typically occurring time of arrival delay spreads. The introduction of a guard interval allows so to speak the "constructive collection" of different signal contributions.

If the time of arrival of different signals at the location of the receiver is no longer a critical issue then conversely it is no longer necessary to exactly synchronize in time all transmitters participating in a single frequency network. The actual transmission instant is allowed to vary within certain limits. But, this gives rise to the possibility to exploit this degree of freedom by configuring the network correspondingly. Thereby, it is possible to eliminate the occurrence of selfinterference problems.

Such situations might be encountered if there is a bounded part of the total coverage area in which due to special topographic conditions a distant transmitter still can be received. If its time of arrival is such that the difference with respect to the times of the others is larger than the guard interval, selfinterference would result. However, by advancing the transmission instant of that particular transmitter a delay spread inside the critical area could be configured which would be harmless.

In principle, such a type of configuration could be applied to any transmitter making part of the single frequency network. Consequently, in order to characterize a transmitter completely it is necessary to describe it not only in terms of power and antenna diagram but to include an individual transmission delay τ_k, too.

Apart from pure technical parameters realistic network planning necessarily must deal with economic aspects as well. The erection and the operation of a transmitter is a very expensive enterprise. Several factors have to be considered. First of all, an adequate property to build a tower is needed. Maintenance facilities are to be provided as well. Furthermore, an expensive technical infrastructure must be set up. This covers antennas, cable connections, amplifier, transmitter unit and much more.

Expenses related to local infrastructure is one thing, another issue is the way in which the signal that is to be broadcast is delivered to the transmitter location. According to individual local conditions data communication lines, satellite links or radio-link systems can be utilized thereto. Each of them require different technical equipment and can be employed under different conditions.

All these factors have to be included into a proper calculation of the costs for setting up and maintaining a transmitter site. Therefore, quite naturally in practice it happens that although a particular coverage target has been defined and should be reached this is not realized at any costs. Applied to the problem of network planning this means that it should be possible to include the costs K_k a transmitter causes into the planning process from the very beginning.

Certainly there are further aspects that could be considered in order to model the characteristics of the transmitters. However, the principle ideas about network planning are not affected by reducing the network parameters to the set of quantities discussed so far. Table 10.1 summarizes those parameters that are used for the planning examples here.

$\mathbf{r}_k$:	location vector of the k-th transmitter site
P_k :	ERP of the k-th transmitter
A_{km} :	antenna factor of the k-th transmitter towards direction m
τ_k :	time delay of the k-th transmitter
K_k :	costs of the k-th transmitter

Table 10.1: A set of network parameters employed for planning purposes.

To conclude this section, one point not mentioned so far needs to be emphasized. In addition to the set of parameters given in table 10.1, usually the description of a transmitter site is supplemented by the specification of thirty-six effective height values (see section 3.2.1). They are to be interpreted the same way as it is done for the antenna factors, i.e. they represent the values for the directions 0°, 10°, 20°, etc. As with the quantities A_{km} a linear interpolation is applied for angles in between.

It is evident that this information is essential for the application of the wave propagation model according to the ITU recommendation 1546 [ITU01a]. However, this parameters are fixed as long as the location of the transmitter is fixed. If the position of a transmitter is changed these parameters have to be calculated again. Their values are linked to the topographic conditions and thus cannot be adjusted freely. Therefore, they cannot be considered free parameters in the literal sense and hence they are neglected here.

10.2 Practicable Network Planning Approach

Any transmitter network is build from a set of *N* transmitters that are described by defining *N* sets of network parameters as listed in table 10.1. In order to assess the quality of this network it is necessary to specify a wave propagation model, a receiver model and to define the coverage area envisaged.

A typical planning scenario for a digital terrestrial single frequency network consists of about 10 - 30 transmitter. However, a significantly larger number might be necessary in some cases as well. Planning the network means to adjust the set of network parameters in line with some predefined quality criterion.

Even though mathematically speaking ERP, time delay and antenna factors are real parameters it makes sense for practical reasons to introduce a set of distinct values from which each of them has to be chosen, respectively. If each of the network parameters has only ten different values to choose from then already for one single transmitter there are $10*10*10*10^{36} = 10^{39}$ different configuration possibilities. Considering a small network of only ten transmitters therefore gives rise to the astronomically large number $\left(10^{39}\right)^{10} = 10^{390}$ of different network configurations! Practically, this is equivalent to infinity.

As in the case of frequency assignment problems obviously also here a NP-hard combinatorial optimization problem is encountered. Any brute-force approach based on pure trial-and-error must necessarily fail. Once again, it seems that more sophisticated methods like for example stochastic optimization algorithms need to be adapted in order to find a satisfying configuration of the network.

However, it turns out that employing the network parameters of table 10.1 leads to combinatorial problems that are very, very complex. All stochastic optimization algorithms evaluate only a very small fraction of the total number of possible configurations. The determination of a satisfying solution is mainly connected to the application of more or less sophisticated strategies to decide which configurations are tested and which not.

The large number of network parameters is connected to the commonly used description of antenna patterns in terms of thirty-six different values. Such a parameterization does not make too much sense when see from the optimization algorithm's point of view. It seems to be reasonable to introduce a representation of an antenna diagram which is based on a significantly smaller number of parameters.

Basically, there are two promising possibilities. Looking at real transmitters indicates that both the shape as well as the direction of an antenna diagram is usually subject to restrictions. Therefore, it would be natural to define three to five different antenna diagrams that can be employed. Furthermore, it could be assumed that these diagrams could be directed only towards thirty-six directions, namely from 0° to 350° with a step-size of 10°. Instead of 10^{36} there would be only $5*36=180$ possible antenna configurations for one transmitter.

Another possibility would be to define antenna diagrams in terms of mathematical formulas depending on a small number of free parameters only. A very simple diagram can be described by two parameters for example. One of them is the angle Φ_k of the direction of the main lobe. The other specifies the width κ_k of the antenna diagram.

The vector

$$\mathbf{s}_k = \begin{pmatrix} \cos\Phi_k \\ \sin\Phi_k \end{pmatrix} \tag{10.7}$$

is oriented along the direction of the main lobe while

$$\mathbf{d}_k = \begin{pmatrix} x - x_k \\ y - y_k \end{pmatrix} \tag{10.8}$$

represents the line connecting the location of transmitter k and the point of reception. Calculating the scalar product between $\mathbf{s}_k$ and $\mathbf{d}_k$ allows to determine the angle ϕ between the two vectors in (10.7) and (10.8) :

$$\phi = \arccos\left\{\frac{(x-x_k)\cos\Phi_k + (y-y_k)\sin\Phi_k}{\left[(x-x_k)^2+(y-y_k)^2\right]^{1/2}}\right\} . \tag{10.9}$$

It has to be noted that the definition (10.14) implies the angle zero to coincide with the east direction which is in contrast to the common geographical definition where the north direction is usually considered as representing zero degrees. Furthermore, the angle increases by counter-clockwise rotation.

With the help of (10.14) the antenna diagram is then given as

$$A(\phi, \Phi_k, \kappa_k) = \exp\left[-\kappa_k\phi^2\right] . \tag{10.10}$$

This means the ERP has to be multiplied by $A(\phi, \Phi_k, \kappa_k)$. In logarithmic representation the multiplication corresponds to adding the term

$$10\log_{10}\left[A(\phi, \Phi_k, \kappa_k)\right] = -10\ \kappa_k\phi^2 \log e \tag{10.11}$$

[Beu95]. If both for the parameter κ_k as well as for Φ_k ten different values are allowed, a total of 10*10=100 configurations is possible.

The function defined in equation (10.10) is monotonous and hence has only one local maximum. The corresponding antenna diagram thus shows exactly one main lobe. It has to be checked on a case-by-case basis if such an approach is reasonable or not. By superimposing several equivalent terms corresponding to different values κ_k and Φ_k it is in principle possible to create a a diagram having more significant lobes. The price to pay, however, is a corresponding increase of the number of network parameters.

10.2.1 Assessment of a Network Configuration

Fixing the set of network parameters to certain values corresponds to establishing a particular network configuration. Clearly, an arbitrary choice can give rise to very poor coverage. Therefore, it is necessary to select from the vast number of possible configurations one that gives satisfying coverage results.

However, the terms “good” or “bad” in relation with a network configuration is necessarily a relative assessment of the network. It crucially depends on the way the quality of a network is defined. What represents an optimal choice with respect to a particular quality function might be considered bad in the context of other quality criteria.

Furthermore, coverage obviously is a two-dimensional concept. Therefore, it might turn out that globally a certain network configuration is being considered good. At the same time, however, there might be very unsatisfying coverage conditions on a local level in special regions of the coverage area. Clearly, the inverse situation might be encountered as well. This phenomenon has to be taken into account properly when defining the quality function by which the network is to be assessed.

If wide-area coverage is the primary coverage target, it is obvious to select a sufficiently large number of test points spread out across the envisaged coverage area and use them as a representative base for an appropriately chosen quality measure. In general, there are three types of test points that are relevant for an assessment of the network configuration. There are served points, points where the minimum field strength is not reached and points where no satisfying service can be provided due to selfinterference problems.

One way to cover a larger area is to, quite generally speaking, simply increase the emitted power in the network. However, more power output entails significantly higher costs. In addition, for digital single frequency networks it is the effect that increasing the power of individual stations in the network might lead to an increase of selfinterference in first place.

A natural starting point to find a good network configuration is to adjust the network parameters such that the number of served test points Z_S is maximized. If this does not lead to satisfying results it is worthwhile to try to minimize the number of points with selfinterference Z_{SI} and the number of points where the minimum field strength Z_{Min} is not reached. Very often a combination like

$$Q = \frac{Z_S}{(1 + Z_{SI})(1 + Z_{Min})} \tag{10.12}$$

of all three figures gives the best results. In order to apply the quality criterion (10.12) the total number of test points,

$$Z_T = Z_S + Z_{SI} + Z_{Min} \ , \tag{10.13}$$

must be kept fixed.

Equation (10.12) can be used as quality function for example when employing a stochastic optimization algorithms. Then, the task is to find that set of network parameters which corresponds to a maximum value of the quality Q. In equation (10.12) the denominator has been chosen such to avoid the occurrence of singularities. Other combinations of the three figures Z_S, Z_{SI} and Z_{Min} can be utilized as well.

By using nonlinearities like exponential functions it is possible to introduce different weights. There is actually no limit with regard to inventing new quality functions. Sometimes it is even unavoidable to put some effort into the creation of a reasonable quality function because it might turn out that a straightforward definition like (10.12) is not sensitive enough.

Apart from the different types of test points it might be necessary to take into account the number of served inhabitants, too. Each test point is associated with a number of people. Hence, the summation of the number of served test points can be supplemented by a summation of the corresponding numbers of inhabitants E_m. In the same way, the operating costs K_k for the transmitters can be taken in account.

If the coverage target is to maximize the number of served test points and the number of served inhabitants along with a simultaneous minimization of the number of interfered or unserved pixels and the network costs, then a quality function like

$$Q = \frac{A\ Z_V + B \sum_{m=1}^{Z_S} E_m}{C\ \ (1 + Z_{ESI})\,(1 + Z_{Min}) + D \sum_{k=1}^{N} K_k} \tag{10.14}$$

could be employed. In definition (10.14) the four scaling factors *A, B, C* and *D* have been introduced. By appropriately adjusting the their values the optimization criteria gain or loose weight with respect to the application of the quality (10.14). If such an approach will give satisfying results must be decided on a case by case basis.

Chapter 11

Fictitious Network Planning Scenario

The planning of transmitter networks is a highly sensitive issue. Obviously, the manner in which a network is actually configured determines the quality of service provided. In particular, in those geographical regions where several network providers are competing, solid network planning is the key to economical success. This is certainly true in the field of mobile communication. So far broadcasting was affected by such economical competition only to a minor extend, at least as far as terrestrial broadcasting is concerned.

In Germany there are two network providers maintaining terrestrial transmitter networks for the distribution of radio and TV programmes. These are the public broadcasters and the Deutsche Telekom AG. Since there are no other competitors these two organizations share the market between them. The situation is different with respect to cable and satellite distribution. Several national and international companies offer transmission capacity. The German cable networks have been recently privatized with the result that also foreign cable network providers entered the market.

The start of digital transmission has changed the situation with respect to terrestrial broadcasting as well. The German telecommunication act in its present form demands for digital terrestrial broadcasting the separation between network and content providers. Therefore, the introduction of T-DAB has been put into practice by founding network provider companies. Both public and private broadcasters and Deutsche Telekom AG hold shares of these companies.

In most cases, however, these network provider companies could not develop beyond the status of pure marketing agencies. A proper operation of the networks in the sense of possessing and maintaining the transmitter sites does not exist. Thus, the joint creation of network provider companies shared by all interested organizations is currently considered a very critical solution. Hence, it is not very likely that this concept will be applied again for the DVB-T transmitter networks in the future.

Not only the framework given by the updated telecommunication act but also the changed utilization of radio and television systems in the digital era will affect the network providership. Amongst the transmission of pure audio and video data T-DAB and DVB-T in principle offer the possibility to broadcast arbitrary digital data. Both systems could be used also as broadband data delivery systems. Combined with GSM or UMTS services an asymmetric communication system could be established providing an enormous amount of data capacity for the downstream to the customer. Especially these new ideas could be the starting point for other so far not active network providers to try to get hold of licenses to operate T-DAB or DVB-T networks.

Up to now the German broadcasters are either funded by public fees or have to rely on advertising revenues. Pure Pay-TV has had only a small market share. At the beginning of 2002 Pay-TV in Germany ran into severe problems. This confirmed that a business model which is entirely based on the provision of radio or television content does not seem to be successful enough to survive on the market. At least this is true as long as there is such a large share of free-to-air programmes as is the case in Germany at the moment.

If, however, the data capacity a system like DVB-T offers is exploited to transmit other data as well, it seems other business models than pure audio and video broadcasting might become attractive. Under such conditions it becomes even more crucial for network providers not to make public too much details of their network planning and implementation activities.

These considerations have a significant impact for the presentation of any planning examples. It is evident that in contrast to the frequency planning part of this book in this case no real world scenarios can be discussed here in detail. Therefore, a completely fictitious scenario is presented. Nevertheless, it is detailed enough to discuss all interesting and relevant aspects in relation to the planning of digital single frequency networks.

11.1 Model for the Fictitious Transmitter Network

The model presented here rests on a purely artificial landscape. A set of transmitters is arranged throughout the considered area. All of them can be included to make part of a single frequency network. The values of the employed network parameters are in line with a typical T-DAB network. It has to be emphasized, however, that this does not constitute a drawback for the general conclusions to be drawn from the following investigations. The reason is that the planning principles are to a large extend independent from the explicit type of terrestrial COFDM system under consideration. Clearly, the values of parameters like the guard interval, the minimum field strength or the protection ratios need to be changed. But, this does not affect the fundamental strategies. In any case, it is possible to obtain a qualitative picture of the characteristics of network planning and optimization activities.

11.1.1 Digital Terrain Model

The typical extension of the coverage area in the case of T-DAB in VHF lies in the range between 150 km and 250 km. In order to have enough freedom for the investigations a rectangular area has been defined here having edges of 200 km in east-west direction and 250 km north-south direction. This corresponds more or less to the size of the German Bundesland Baden-Württemberg.

As resolution of the digital terrain model a value of 1 km has been used. Thus each test point represents an area of 1 km $\times$ 1 km. Real world planning activities are usually based on a 200 m grid. However, for studies on a qualitative level a more coarse resolution is sufficient, especially in view of a reduction of the computational time to carry out the calculations. Furthermore, the characteristic features of a hilly terrain like that in the south-western part of Germany are still very well captured by a lower resolution.

With the help of a random number generator based on a two-dimensional uniform probability distribution a set of points within the envisaged rectangular planning area having coordinates $(\overline{x}_k, \overline{y}_k)$ has been chosen. Each of these points was assumed to be the centre of a hill whose height as a function of location is given by

$$H_k(x,y) = \alpha_k \exp\left[-w_k \left\{(x-\overline{x}_k)^2 + (y-\overline{y}_k)^2\right\}\right] . \tag{11.1}$$

Both the amplitudes α_k as well as the widths w_k were assumed to be uniformly distributed statistical variables taken from a pre-defined interval. The final digital terrain has been generated by superimposing all individual height profiles, i.e. the summation of *N* functions $H_k(x,y)$ gives the terrain

$$H(x,y) = \sum_{k=1}^{N} H_k(x,y) = \sum_{k=1}^{N} \alpha_k \exp\left[-w_k \left\{(x-\overline{x}_k)^2 + (y-\overline{y}_k)^2\right\}\right] . \tag{11.2}$$

All parameters have been chosen such that realistic height values in the order of some hundred meters resulted. The figures 11.1 and 11.2 show the utilized landscape seen from two different perspectives. The presented digital terrain features two mountain ridges both in the north and the south which are connected by several smaller hills in the west. Towards east the height falls off. As a result a deep lying valley shaped like a vessel is created in the centre of the planning area.

The same method has been adopted for the generation of a reasonable population distribution. It is evident that such a distribution cannot be independent from the underlying topography since in reality there are areas which are more likely going to be populated than others. Sea regions can be considered as being unpopulated for example, and very high mountain tops usually have a very low population density. Thus, it has been assumed that there exists a critical height above which no population is found any more. This is certainly a reasonable assumption as can be confirmed by looking at a corresponding real world map.

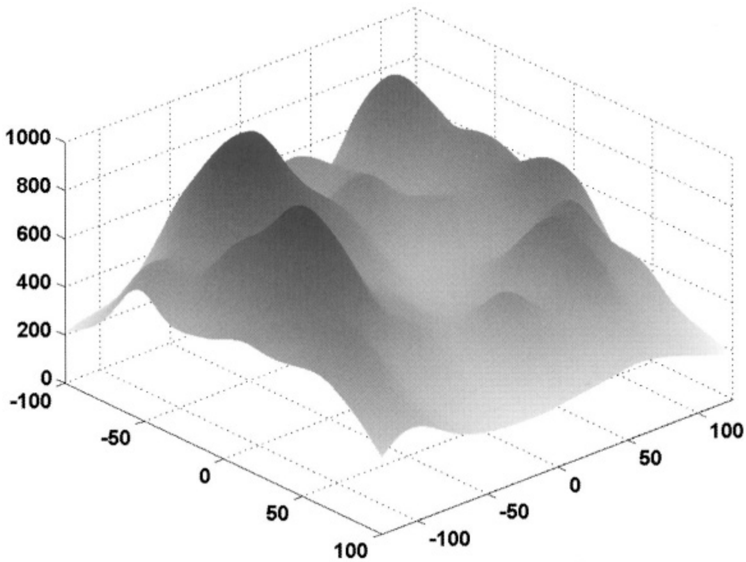

Figure 11.1: Utilized ficticious landscape seen from the south-east.

Apart from that, the same creation mechanism has been employed as in the case of the terrain height. Starting with a two-dimensional uniform distribution a set of points lying inside the rectangular planning area has been chosen. However, only those points have been accepted whose height was below an appropriately fixed critical value. These points were interpreted as the centres of local population distributions P_k of the form

$$P_k(x,y) = \beta_k \exp\left[-v_k\left\{(x-\overline{x}_k)^2 + (y-\overline{y}_k)^2\right\}\right] . \qquad (11.3)$$

The amplitudes and widths of the exponential functions in (11.3) were considered as statistical variable within certain upper and lower bounds so that the total number of inhabitants at the point (x, y) was defined as the summation of all individual contributions $P_k(x, y)$ as

$$P(x,y) = \sum_{k=1}^{N} P_k(x,y) = \sum_{k=1}^{N} \beta_k \exp\left[-v_k\left\{(x-\overline{x}_k)^2 + (y-\overline{y}_k)^2\right\}\right] . \qquad (11.4)$$

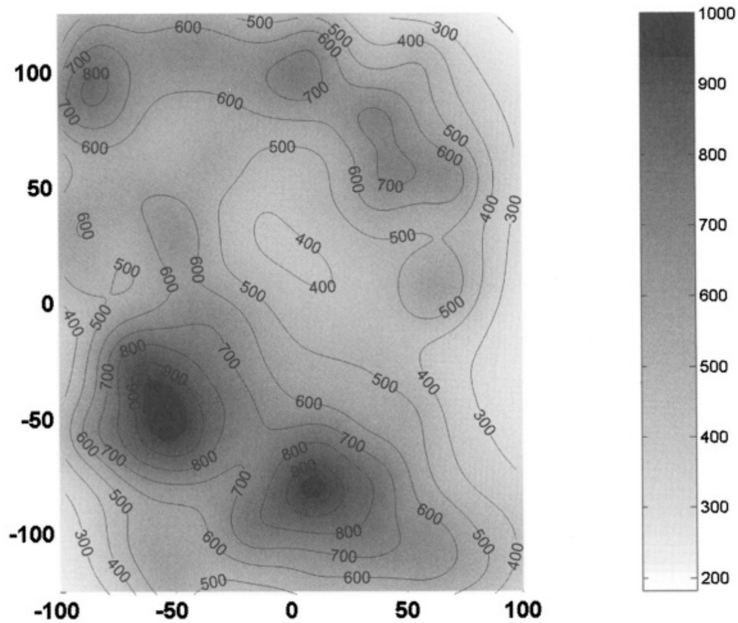

Figure 11.2: Bird eye's view of the utilized ficticious landscape.

Figures 11.3 and 11.4 illustrate the final population distribution seen from the same perspectives as for the terrain height above.

In particular, figure 11.4 very clearly indicates that a population distribution has been generated which has indeed a very realistic structure. The areas which are highly populated are located where the terrain height is small. There are high density areas, one in the centre of the planning area exactly at the ground of the valley. This population centre extends to the east near the entrance to the central lowlands region. The other highly populated area is found in the north-eastern part of the rectangular area, beyond the northern mountain ridges.

11.1.2 Simplified Wave Propagation Model

After an artificial topography as well as a correspondingly adapted population density have been created, it is now possible to identify reasonable transmitter sites. The planning area has a maximum extension of 200

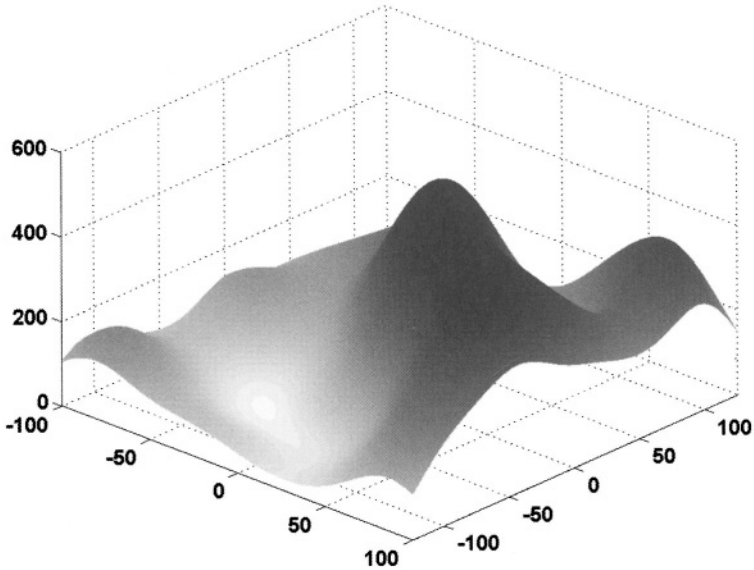

Figure 11.3: Utilized population distribution seen from south-east. The figures on the z-axis represent the number of inhabitants living in an area of 1 km^2.

km $\times$ 250 km. This corresponds more or less to the area of Baden-Württemberg, Germany, where a T-DAB single frequency network has been put into operation just some time ago. This transmitter network comprises about twenty transmitters located at carefully chosen sites.

To keep the discussion as close as possible to reality and furthermore possess enough degrees of freedom for the various network planning scenarios discussed later a transmitter pool of twenty-six transmitters has been set up here. The are spread out across the planning area as can be seen in figure 11.5.

When fixing the transmitter parameters like radiated power, antenna diagram, etc., an appropriate wave propagation model based on the given digital terrain model can be employed to calculate the field strength each transmitter is going to produce at each point of the entire planning area. Each of the practically relevant models discussed in section 3.2 could be utilized. With regard to a handling as easy as possible a heavily simplified wave propagation model has been used for the following planning

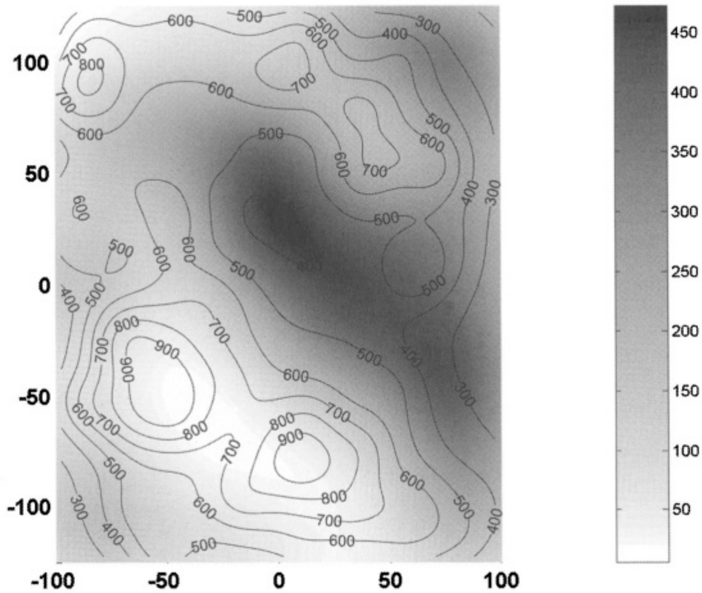

Figure 11.4: Bird eye's view of the utilized population distribution. The black curves represent the contours of constant height according to the digital terrain model shown in figures 11.1 and 11.2.

examples. It does not provide a good basis for real world planning scenarios. However, it is sophisticated enough for demonstration purposes.

Thus, it has been assumed that only at those pixels having a line-of-sight connection to the transmitter a significant field strength will be provided. Pixels not falling into this class will be assigned a very small dummy field strength value. Furthermore, all transmitters were supposed to possess antennas being mounted 100 m above ground. Clearly, this height offset has to be taken into account for the line-of-sight criterion.

The decay of the field strength with respect to an increase of distance between transmitter and receiving point was assumed to obey a simple inverse-square decay law. This means that the field strength after a distance d at the point (x, y) is given by

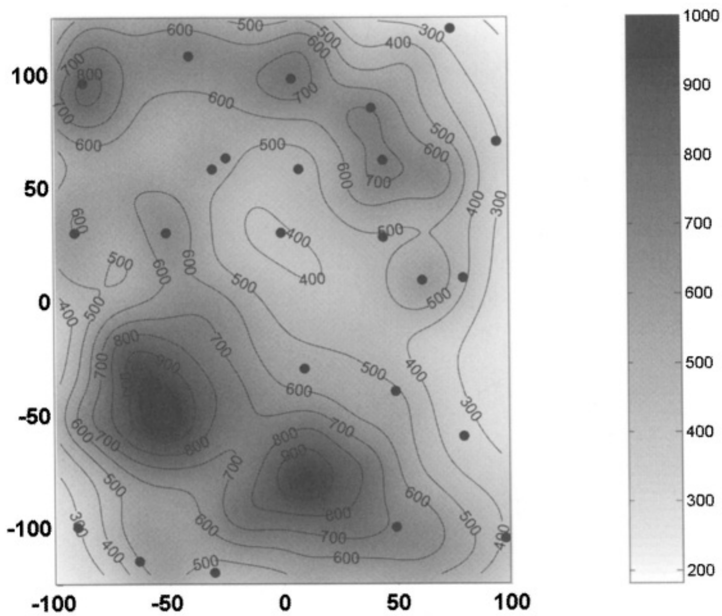

Figure 11.5: Location of the transmitter sites.

$$F(x,y) = \frac{F_0}{d^2} \tag{11.5}$$

with

$$d = \sqrt{(x - x_k)^2 + (y - y_k)^2} \tag{11.6}$$

where the location of the k-th transmitter site is described by the co-ordinates (x_k, y_k). F_0 corresponds to the field strength produced by the ERP in the close vicinity of the transmitter site.

Figure 11.6 shows the field strength distributions for four arbitrarily chosen transmitters of the set depicted in figure 11.5. It has to be emphasized, that these field strength values are valid strictly speaking only at the grid point separated by 1 km.

In practice, each of these values is considered as the mean value of a given field strength distribution across a pixel area. Adopting adequately defined distribution functions thus allows to make statements about the coverage probability (see section 3).

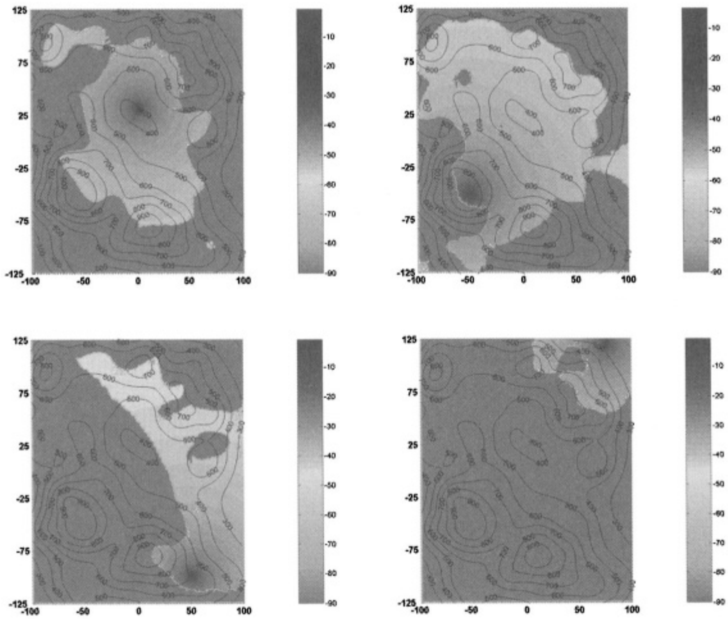

Figure 11.6: Field strength distribution of four arbitrarily chosen transmitters. All have an ERP of 1 kW and employ omnidirectional antennas. The field strength values are given in logarithmic scale normalized to the maximum value appearing. The contours of constant height are given as well.

In single frequency networks usually several transmitters are contributing to the received signal. This multi-path environment has to be fully taken into account when employing the statistical approach. In section 3 some practical methods are discussed.

For the model applied here this statistical treatment has been omitted. The following demonstration of certain important aspects of network planning of single frequency networks is entirely based on the power sum method described in section 3.4.2. This means that only a deterministic summation of the set of signal powers arriving at a particular point is carried out.

Basically, there are three different relevant power sums to be considered. First of all, all signals are summed independent of their actual relative

served	$F_{\text{total}} > F_{\text{min}}$	$C/I > F_P$
interfered	$F_{\text{total}} > F_{\text{min}}$	$C/I <= F_P$
unserved	$F_{\text{total}} <= F_{\text{min}}$	C/I arbitrary

Table 11.1: Utilized coverage criteria.

times of arrival. The result is the total received field strength F_{total}. Second, according to the synchronization strategy of the receiver the arriving signals are discriminated as wanted and unwanted parts. Both types of contributions are summed independently to give the total wanted signal F_{wanted} and the total unwanted signal F_{unwanted}. The ratio between F_{wanted} and F_{unwanted} is called the signal-to-interference ratio

$$C/I = \frac{F_{\text{wanted}}}{F_{\text{unwanted}}} . \tag{11.7}$$

A test point is called served if both the total received field strength F_{total} exceeds a certain threshold value, the so-called minimum field strength F_{min}, and the signal-to-interference ration is larger than the protection ration F_P. If *C/I* is smaller than required but at the same time the minimum field strength criterion is met, then the corresponding pixels is considered being interfered. In case there is not even enough field strength that the minimum field strength is reached, the point is classified as unserved. Table 11.1 summarizes the utilized coverage criteria.

For the examples presented here a minimum field strength value F_{min} of 58 dB has been used, while for F_P 20 dB have been assumed. Guard interval and duration of the evaluation window correspond to the values of T-DAB, mode I in the VHF range, i.e. $T_G = 246\mu s$ and $T_W = 1000\mu s$.

11.2 Network Planning Examples

It is crucial for any network planing process that the planning targets have been properly defined beforehand. To give an example, it has to be decided if wide-area coverage is to be aimed at or if only dense urban areas are to be covered. This is an issue where public and private broadcasters very often have diverging demands. For private broadcasters a population density oriented coverage usually is preferred. In order to meet such requirements it is necessary to possess reliable population density data.

However, the specification of the coverage area is just the first step. It is obvious that any attempts to reach a given coverage target have to incorporate economic issues as well. Even though economics certainly has higher priority for private than for public broadcasters also the latter need to be more and more aware of economical factors. The question how much money has to be spent on the implementation of a transmitter network is of fundamental importance for any network provider.

Apart from costs connected to the provision of a network also the operating costs need to be evaluated carefully. A primary cost factor certainly is the number of transmitters forming the network. The location of the transmitter sites influences directly the costs in terms of rents and expenses for maintenance. Furthermore, the broadcast signals have to be transported to the transmitters sites. If this possible via a satellite link it is certainly less expensive than compared to the engagement of cable links.

Beside the pure infrastructure, the radiated power of the transmitters influences the total costs of a network, too. The more power is to be radiated the more effort has to be taken in terms of technical equipment. More electric power is then consumed and at the same time stronger cooling is required.

All these factors play a key role when selecting the transmitters for a network. Together with the pure planning parameters discussed in the previous sections these economic parameters build the set of parameters that is to be adjusted in an optimal manner by the planner. Clearly, a simple trial-and-error scheme does not provide a good means to cope with such complex problems in a satisfying way. Once again more sophisticated methods like stochastic optimization algorithms can be employed instead.

The planning examples presented here are meant to illustrate interesting aspects of the entire problem. In order not to overload the examples some basic assumptions have to be made. One of the major problems of network planning concerns the selection of transmitters for the network. Usually, a network provider owns a pool of transmitters from which a subset can be selected in order to achieve a defined coverage target.

Thus, the 26 transmitters that are shown in figure 11.5 are supposed to build this pool.

In general, it is not possible to equip any transmitter site with technical devices such that any power could be radiated. In reality it might turn out that at a particular location only an ERP up to a given upper limit can be realized. Apart from technical restrictions there might be also restrictions due to general limits for electromagnetic emission. However, in order to simplify the examples it has been assumed that at a given transmitter site an arbitrary ERP can be realized.

The same situation is encountered with respect to the antenna diagrams. In reality, a transmitter tower is not only used for T-DAB or DVB-T. Usually, a tower carries antennas for all types of relevant broadcasting services from FM radio to DVB-T. Clearly, this limits some time the freedom to build antennas which are in accordance with the planning results. For example, there might not be enough space left on the antenna mast to mount new antenna elements which, however, would be needed in order to reduce the power output towards certain geographical directions. If this is not possible, the only measure that can be taken is to decrease the total ERP in order to meet such an requirement. This might pose very complicated problems in reality which, however, do not add anything fundamental to the planning aspect of antenna design. Therefore, as in the case of the ERPs it has been assumed that at any transmitter location any antenna diagram can be realized.

The planning examples presented here are generated such that they resemble a typical realistic planning scenario. The first step concerns the determination of the required output powers. Then, the antenna diagrams and the time delays are adjusted in order to optimize the coverage. Furthermore, planning examples are dealt with that take into account population densities and finally the explicit incorporation of the network costs into the planning process is considered.

11.2.1 Optimization of Transmitter Powers

According to the final acts of the Wiesbaden agreement [CEP95] real-world planning scenarios for T-DAB single frequency networks need to take into account an upper limit for the ERP of 1 kW. In the case of DVB-T currently ERPs between 20 kW and 50 kW are considered reasonable values.

More power output can be realized under special circumstances already today. However, it seems to be doubtful that this can be achieved on a general basis. The reason is that extreme requirements concerning the linearity of amplifiers have to be met in order to reach ERP above 50 kW when employing typical antenna configurations. If the antenna gain could be increased significantly, this would certainly boost the overall performance of DVB-T transmitters.

Despite that practical restriction an upper ERP limit of 100 kW has been assumed for the planning examples here. That way there is more freedom in terms of adjusting the transmitter power in an optimal manner. It should not be forgotten, however, that the planning examples only qualitatively render real planning scenarios. Thus the ERP values should not be taken literally with respect to their coverage potential. In reality, this is very strongly linked to the physical wave propagation conditions.

Figure 11.7 illustrates the covered area for the case of all 26 transmitters being used and radiating an ERP of 100 kW. All antenna diagrams are omnidirectional and the time delays of all transmitters have been put equal to zero.

The entire coverage area encompasses $201 \times 251 = 50451$ pixels. Those pixels where the minimum field strength is not reached are marked dark gray which here is a total of 6693 out of 50451 or 13.3 %. Light gray indicates served pixels (30567 or 60.6%) while throughout the medium gray areas selfinterference occurs. This corresponds to 13191 pixels or 26.1%.

The example of figure 11.7 is remarkable in many respects. Obviously, it is not possible to serve the whole area despite the massive use of ERP. There are two reasons for that. First, the employed wave propagation model is very restrictive in the sense that a line-of-sight relation between a transmitter location and the point of reception is required in order to provide any field strength. As soon as there is a simple obstacle in between the ERP could be infinite, however, there will be produced no field strength at the receiver location.

It is evident that this is not very realistic. Electromagnetic waves can

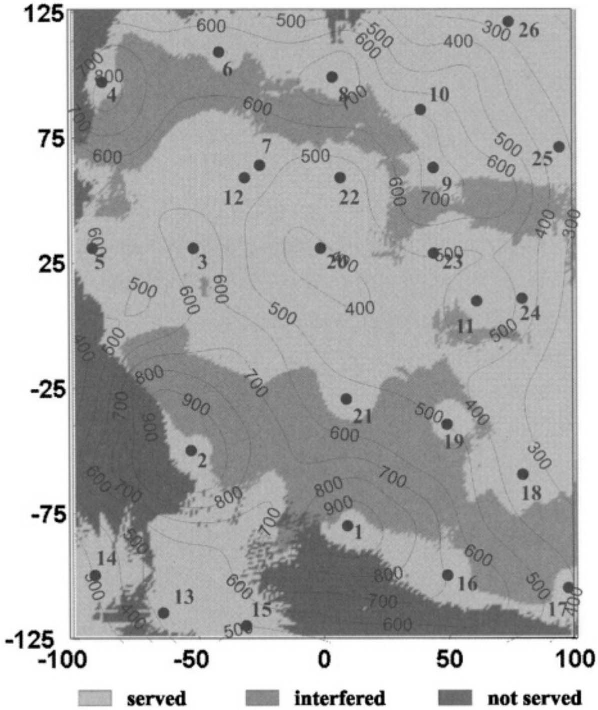

Figure 11.7: Covered area in the case of 26 identical high power transmitters.

seep into the shadowed region behind any obstacles by means of diffraction. Clearly, the thereby produced field strength levels will be attenuated significantly compared to any directly arriving signals. However, quite often the power thus delivered is sufficient to still allow for acceptable reception quality. Sometimes, even multiple diffractions can be borne.

The consequences arising from the line-of-sight demand can be quite well understood when looking at the three transmitters in the lower left part of figure 11.7. These three transmitters are supposed to radiate a power of 100 kW and at the same time are located very close to each other. Nevertheless, the area they cover is remarkably small.

The second important conclusion which can be drawn by looking at figure 11.7 is that provision of a sufficient amount of ERP obviously is not enough to achieve a reasonable coverage in the case of a digital terrestrial single frequency network. There are extended areas where the coverage is degraded by the occurrence of selfinterference. In the case of figure 11.7 these areas are fortunately located along the mountain chains. What is still noteworthy is the fact that even in the immediate vicinity of transmitters (like for example transmitters 1 or 2) selfinterference might be encountered.

This is mainly due to the difficult topographic layout of the planning area. The two mountain chains in the north and the south of the virtual landscape have direct line-of-sight but at the same time are separated such that their mutual distance is larger than the critical distance connected to the guard interval. If all transmitters are operated with zero time delay, the signals of the transmitters of one mountain ridge arrive at the other with a time delay larger than the guard interval. The question how this selfinterference can be reduced or even eliminated will be discussed later.

If the possible ERP is limited by the assumed 100 kW, then the coverage situation of figure 11.7 represents a limiting configuration with respect to the utilization of transmitter power. This means in particular that the number of pixels where the minimum field strength is not reached cannot be reduced further. Every means leading to a direct or indirect reduction of the total radiated power in the network can only maintain the status quo with respect to the number of pixels being provided not enough field strength. In general, the coverage situation will become worse.

On the other hand, it is clear that probably part of the power in the network is not needed to reach the minimum field strength threshold at all. It does very likely only contribute to the creation of selfinterference. Therefore, a first quite natural optimization step would be to reduce the ERP of individual transmitters without increasing the total number of pixels not meeting the minimum field strength criterion. Figure 11.8 shows the result of a corresponding simulation on the basis of the Great Deluge algorithm.

The starting point of the optimization run was the 100 kW configuration of figure 11.7. During the simulation the objective was to adjust the ERPs of the 26 transmitters such that the number of interfered pixels would be minimized. This should be accomplished, however, under the additional constraint that the number of unserved pixels would be kept

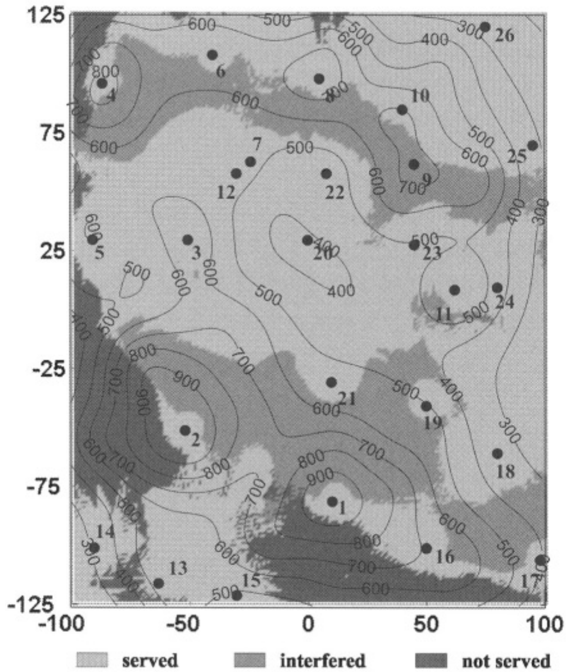

Figure 11.8: Covered area after optimization of the power output of all 26 transmitters.

fixed at the minimal value.

Throughout the studies presented here, ERPs were measured in dBkW. To simplify the calculations, only integer values have been allowed, i.e. 1dBkW or 3 dBkW. A power of 100 kW corresponds to 20 dBkW and 1 kW is equal to 0 dBkW. Table 11.2 gives an overview about the resulting transmitter powers.

Indeed, the improvement that can be obtained is not very large. The best network configuration found corresponds to 31271 served (62%), 12487 interfered (24.7%) and 6693 unserved pixels (13.3%). In comparison to figure 11.7 this leads to an increase of the served area of about 1.5%. The difference can be seen near the points having the coordinates (-90,60) and (75,-75), respectively. It has to be emphasized, however, that the reduction of the total power in the network does not lead to an increase of the unserved area. Only the number of interfered points has

transmitter	1	2	3	4	5	6	7	8	9
power (dBkW)	20	20	20	20	20	20	20	20	1
transmitter	10	11	12	13	14	15	16	17	18
power (dBkW)	19	20	20	20	19	20	19	18	20
transmitter	19	20	21	22	23	24	25	26	
power (dBkW)	20	20	20	20	20	20	14	20	

Table 11.2: Result of the pure power optimization.

be reduced.

Looking at the power figures of the individual transmitters proves, however, that nevertheless such a reconfiguration could be worthwhile for the network provider. The ERP of some transmitters could be reduced by 1-2 dB. For others like transmitter 9 the savings are more significant. This transmitter needs to be operated with 1.2 kW only. Also for transmitter 25 a power reduction by a factor 4 leading to a new ERP of 25 kW is achieved. These reductions would mean significant cost-savings for the network provider.

11.2.2 Reduction of Selfinterference

In the previous section the transmitter powers have been varied under the condition that the number of unserved pixels should not change. As a consequence, the number of interfered pixels can be reduced only marginally. However, apart from the transmitter ERPs it is possible to adjust both the antenna diagrams as well as the time delays of all transmitters in order to obtain further coverage improvements. In particular, the temporal synchronization of a single frequency network provides a degree of freedom in terms of network tuning that has no counterpart in the analogue world.

Under multi-path propagation conditions several signals arrive at the point of reception at different times even if metaphorically speaking each transmitter broadcasts the same “bit” at exactly the same time. This is simply due to the different distances between the transmitters and the receiver. As long as the relative time delays of the signals are not larger than the duration of the guard interval there are no problems for the receiver.

If it would happen by chance that one of the transmitters' broadcasting time would be slightly advanced or retarded with respect to the others, its corresponding signal would arrive at the receiver site at a correspondingly shifted time interval. Again, as long as the delay spread would still be smaller than the guard interval, no problem would arise. Clearly, in the reverted situation when a signal arrives too late the broadcasting time of the associated transmitter could deliberately be adjusted to reduce its relative signal delay below the duration of the guard interval.

However, care has to be taken. Any adjustment of the synchronization affects the relative time delay spread of the signals only with regard to the considered point of reception exactly in the desired way. The situation might be completely different with respect to other receiver locations. It might even happen that by tuning the broadcasting times of the transmitters selfinterference can be removed in one region at the price to generate an even larger interference area somewhere else. The optimal solution would be without doubt to create selfinterference in those areas where either no coverage is needed or the minimum field strength criterion is not met anyway.

Another possibility to suppress an interfering transmitter is to adjust its antenna diagram such that less power is broadcast into the direction under consideration. Generally speaking, imposing restrictions of that kind leads to a reduction of the total ERP in the network. This in turn might increase the number of pixels where the minimum field strength is not reached. As above, if this happens in uninteresting areas it does not do any harm. However, this has to be checked in detail.

A natural and quite straightforward strategy to reduce the amount of selfinterference in the network is to fix the antenna diagrams and time delays of the transmitters according to a systematic scheme. For the antenna diagrams it could be conceivable to equip all transmitters located in the center of envisaged coverage area with omnidirectional antennas while the transmitters at the periphery should get antennas pointing outwards from the core coverage area.

This resembles more or less the network configuration of the reference networks employed in allotment planning (see for example table 5.1). The difference lies in the fact that in the case of reference networks the transmitters at the periphery are directed towards the centre of the coverage area and not, as proposed here, away from the centre.

As it turns out, such a configuration is not suitable in practical applications. In general, transmitters at the periphery of a coverage area need to contribute to the core coverage as well. This is in particular true in regions with challenging topography like the south-western part of Germany for example.

Similar thoughts can be made with regard to the adjustment of the time delays of the transmitters. At first glance, it seems to be promising to implement some kind of "wave motion" throughout the entire planning area. This means first of all to identify a subset of transmitters located in the centre of the envisaged coverage area which all have a mutual distance that is smaller than the critical distance connected to the guard interval. Their time delays would be set equal to zero. Thus, they would not interfere with each other at no reception point.

Then, the center of gravity for the transmitters belonging to the subset is determined. This point is interpreted as the origin of an outgoing virtual electromagnetic wave. The propagation time from the origin to all so far not considered transmitters is then calculated. All these transmitters are advanced by exactly their proper time intervals. As a consequence, the signals they emit are synchronized in the sense that they will arrive at the centre of gravity at the same time. Since the propagation time from the core transmitters to the centre is by construction less than the guard interval all transmitters are radially synchronized with respect to an interference free central coverage area.

If such a strategy is applied to the network configuration of table 11.2, a coverage situation is produced as shown in figure 11.9. In that case the five transmitters being located a little bit northwest of the very centre of the planning area, i.e. transmitters 3, 7, 12 20 and 22, have been used as the core transmitter subset.

A total of 20019 served (39.7%), 23739 interfered (47.0%) and 6693 unserved pixels (13.3%) results. The number of unserved pixels is identical to the starting configuration of figure 11.8 because tuning the time delays does not affect the total amount of ERP of the single frequency network. However, a dramatic increase of the interfered area is found. Thus, the discussed wave synchronization strategy is completely useless.

The reason for this lies in the fact that a wave structure can only work properly in one direction. All transmitter have been tuned such that they do not cause any problems inside the core coverage area. Exactly this

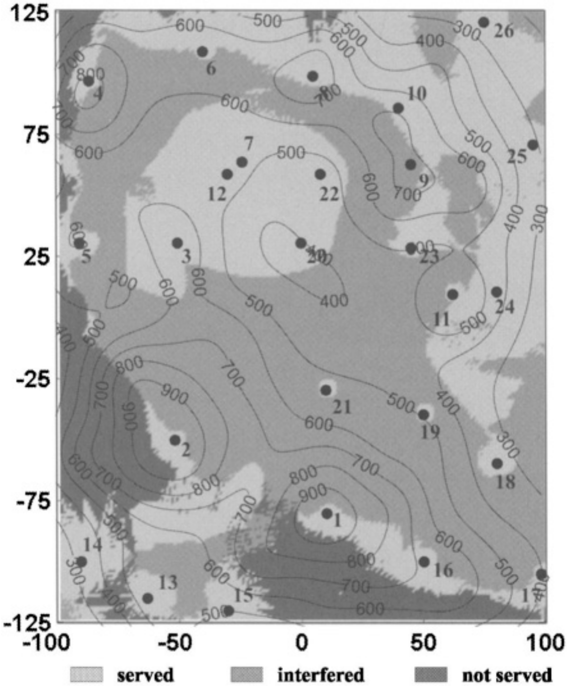

Figure 11.9: Served area generated by a radial wave synchronisation.

effect can be seen in figure 11.9. On the other hand, such a configuration necessarily leads to further divergence of the relative time of arrivals at other points outside the core area. Inevitably, the coverage becomes worse.

Despite this comprehensible qualitative explanation of the result it is still very amazing that an apparently advantageous choice of the transmitter time delays has so disastrous consequences. However, seen from a mathematical point of view the determination of optimal time delays is a highly overdetermined geometrical problem. It is always possible to find an optimal set of time delays for a limited number of reception points in order to eliminate selfinterference. But, as soon as the number of points exceeds the number of transmitters the problem is strictly speaking no longer exactly solvable. In the simple case of only two transmitters this can understood quite easily, but if there is a larger number of transmitters involved then the situation becomes very complex.

Two simple, entirely geometrical examples can help to get an idea of the complexity of the problem. The first case covers two transmitter sites while the second deals with three. On the basis of a guard interval $T_G = 2.46 * 10^{-4}$ sec all points inside a rectangle are classified in terms of the difference between the relative times of arrivals of the different arriving signals.

If only two transmitters are considered, then the areas in which the times of arrival differ by more than the guard interval can be easily described by analytical means. For the mathematical treatment a Cartesian co-ordinate system is chosen with the transmitters being located at the points $\mathbf{r}_1$ and $\mathbf{r}_2$. The times of arrival at an arbitrary point in the plane $\mathbf{r} = (x, y)$ read

$$T_1 = \frac{1}{c}\,|\mathbf{r} - \mathbf{r}_1| + \tau_1 \qquad , \qquad T_2 = \frac{1}{c}\,|\mathbf{r} - \mathbf{r}_2| + \tau_2 \,. \tag{11.8}$$

The time delays $\tau_{1/2}$ have to be taken into account explicitly. All points for which

$$|T_2 - T_1| = T_G \tag{11.9}$$

holds, are located on the branches of a hyperbola whose focal points are identical with the transmitters sites $\mathbf{r}_1$ and $\mathbf{r}_2$ [Bro85]. Depending on the distance $|\mathbf{r}_2 - \mathbf{r}_1|$ of the two transmitters and the choice of the two time delays various different patterns for the areas are found where the difference of the times of arrivals is smaller or larger than the guard interval. Figure 11.10 illustrates the possibilities.

The distance between the two transmitters is 120 km. For a guard interval $T_G = 2.46 * 10^{-4}$ sec this inevitably leads to regions where the times of arrival of the two signals differ by more than T_G. By varying the time delay of one of the transmitters first of all the symmetry of the layout is broken. However, what is more important is the fact that by increasing the time delay the not interfered region becomes smaller until beyond a critical value of the time delay there are left only regions where the guard interval is violated.

The situation becomes even more dramatic if a third transmitter is added. Even if the third transmitter is arranged in a symmetric manner

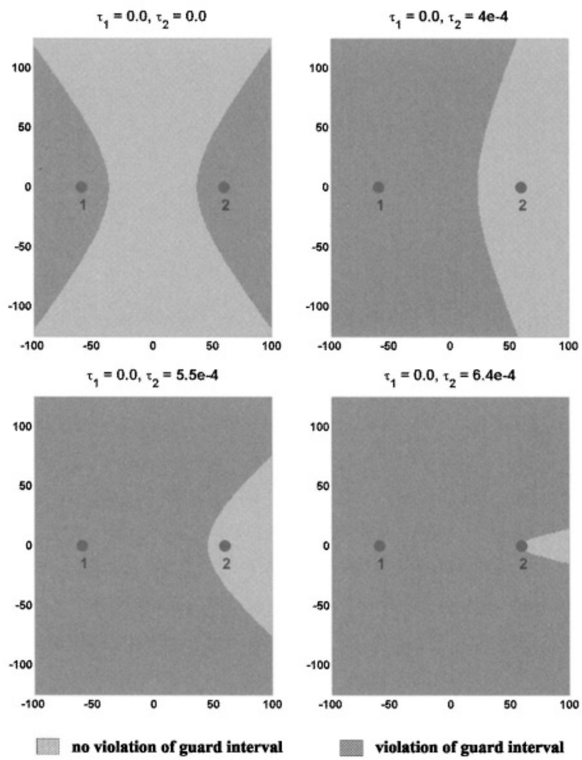

Figure 11.10: Areas with and without guard interval violations in the case of a two-transmitter scenario. The transmitter sites are indicated by the dots. The time delays are given in seconds.

with respect to the others, very complex patterns occur.This time, however, there are not only two categories, i.e. guard interval violated or not violated. In fact, there are four different cases to be distinguished. In order to classify a point in the plane all mutual differences between two transmitters have to be investigated.

In the optimal case all signals arrive within a time interval smaller than the guard interval, i.e. all mutual differences are smaller than T_G. But, any other relation between the times of arrival is found as well. For example three signals might be mutually compatible and only one arrives beyond the guard interval. However, there are also points where the temporal separation between the four signals is such that any pair of signals is delayed with respect to each other more than T_G. Figure 11.11 shows some examples.

As can be seen, already such a relatively simple configuration results in very complex geometrical structures. In a more realistic case like that of figure 11.9 the geometrical circumstances are amended by topography and the characteristics of wave propagation. Both aspects play an important role when determining at a certain point of reception if selfinterference is to be expected or not.

The effect of wave propagation in terms of provided field strength is non-linear[1] and topography is fractal by nature. Therefore, it seems to be obvious that a reduction of selfinterference by adjusting time delays of transmitters on the basis of simple deterministic strategies is not feasible. It does not provide reasonable results. Thus, the only solution seems to be the application of stochastic optimization routines in order to find a set of time delays for the entire network that finally leads to better coverage.

Quite similar as for the time delays one could try to reduce the selfinterference in the network by adept manual manipulation of the antenna diagrams without making any use of some algorithmic optimization strategy. In this case, there is certainly the price to be paid that the number of points where the minimum field strength is not reached will increase.

In figure 11.8 there are two extended interfered areas which are facing each other in north-south direction at a distance of about 100 km. On both mountain ridges there are transmitters whose signals can freely propagate to the other regions, respectively. Since the time of propagation is larger than the guard interval they are primarily generating selfinterference on the other side. One starting point for a manual antenna diagram adjustment to improve the situation thus could be to stipulate power output reductions into the critical directions. Thus, the transmitters in the south should emit less power towards north and vice versa for the northern transmitters. Figure 11.12 shows the coverage resulting from such a configuration.

[1] Clearly, the physics underlying the propagation of electro-magnetic waves is goverend by the Maxwell equations which are linear by definition. In the case of propagation through media showing special electromagnetic-properties non-linear wave propagation might nevertheless be desribed by the Maxwell equations as well.

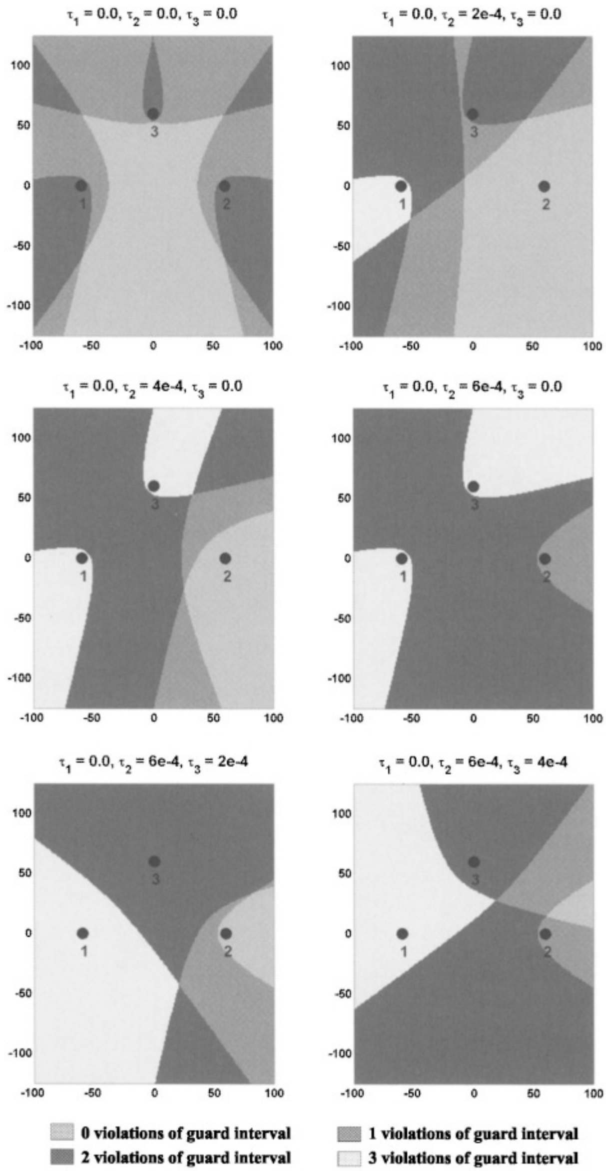

Figure 11.11: Areas with and without guard interval violations in the case of a three-transmitter scenario. The transmitter sites are indicated by the dots. The time delays are given in seconds.

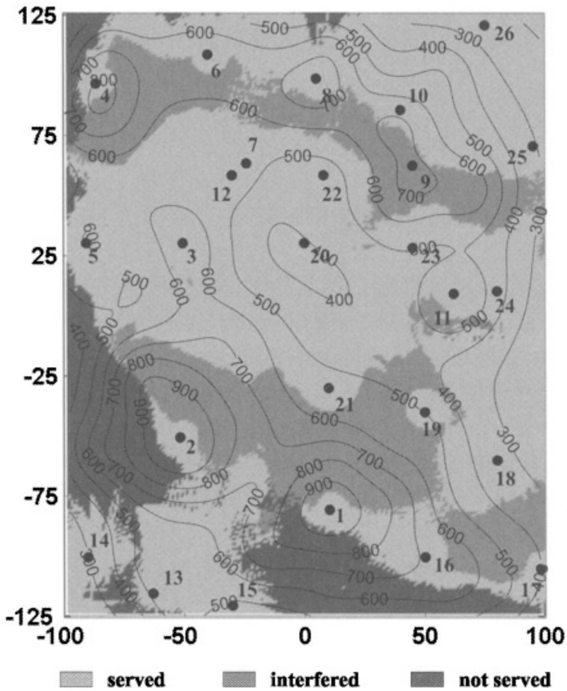

Figure 11.12: Covered area for the case of antenna diagrams directed towards north and south.

The starting configuration to derive the result of figure 11.12 once again has been the power optimized scenarios according to table 11.2. The transmitter 3, 7, 12, 20 and 22 are using omnidirectional antennas. All other transmitters in the south, i.e. 1, 2, 13, 14, 15, 16, 17, 18, 19 and 21, have been equipped with an antenna diagram following the definition (10.10). The parameter κ has been put equal to 0.2 which corresponds to a forward-backward ratio of about 10 dB. The main lobes of the diagrams of the southern transmitters were oriented exactly towards south. The transmitters in the norther part of the planning area were treated the same way with the only difference that their antenna diagrams were oriented towards north.

Altogether, a total of 31835 (63.1%) served pixels, 11767 (23.3%) interfered pixels and 6849 (13.6%) unserved points were found. In comparison with the configuration of table 11.2 an increase of the number of served pixel of about 1.1% can be observed. At the same time the selfinterference recedes by about 1.4% while as expected the number of unserved pixels increases slightly.

At this stage it seems that tuning the antenna diagrams systematically seems to be more efficient in terms of reducing the selfinterference than adjusting the time delays of the transmitters according to an outgoing wave. However, both approaches are very specific and it is not possible to draw any general conclusions. For example, in the antenna design case it is absolutely not clear how the coverage changes if the diagrams are modified slightly. This could be investigated by playing around with the parameters only. But then the approach starts to slide into the standard trial-and-error strategy again.

Carrying out a full-blown stochastic optimization aiming to adjust the time delays proves that the statement "it is better to optimize antenna diagram than time delays" has to be taken with care. Figure 11.13 gives the result of a corresponding calculation.

Also here the starting point has been the power optimized transmitter configuration of table 11.2. This time there are 34186 (67.8%) served, 9572 (18.9%) interfered and 6693 (13.3%) unserved pixels. Basically, this corresponds to an improvement of the coverage by nearly 6%. Table 11.3 summarizes the set of optimal time delays.

transmitter	1	2	3	4	5	6	7
time delay (μs)	24	106.6	32.1	0	36.1	0.5	30.8
transmitter	8	9	10	11	12	13	14
time delay (μs)	0	49.8	0	48.7	20	4.1	39.7
transmitter	15	16	17	18	19	20	21
time delay (μs)	27.3	49.2	0	41.8	83.3	47.6	67.5
transmitter	22	23	24	25	26		
time delay (μs)	43.6	51.9	43.8	133.8	0		

Table 11.3: Result of a time delay optimization.

A full-blown optimization can be carried out for the antenna diagrams without problems as well. However, in order to be able to compare the result with the delay optimization it is necessary to impose the additional constraint that the number of pixels where the minimum field strength is not reached should be kept fixed. This condition is satisfied in the

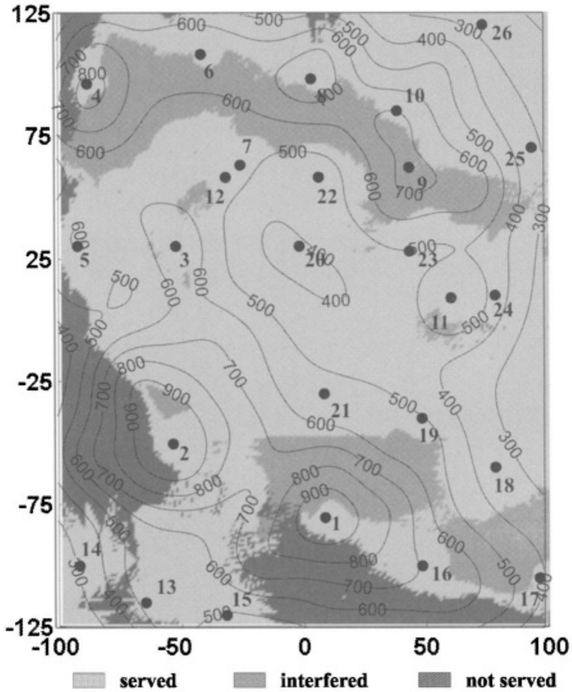

Figure 11.13: Covered area in the case of a full-blown optimization of the time delays of all transmitters.

former case by definition since adjusting the time delays has no impact on the minimum field strength at all.

The simulation has been based on the mathematical description of the antenna diagrams given in equation (10.10). Both the direction of the main lobe and the width of the beam were treated as free parameters to be fixed by the calculation. Figure 11.14 gives an impression of the result.

The coverage figures this time read 31470 (62.4%) served , 12288 (24.3%) interfered and 6693 (13.3%) unserved pixels. The constraint to keep the number of unserved pixels fixed has significant impact on the result. All transmitters are configured with omnidirectional antennas. Only transmitters 10, 14, 15, 16 and 24 have directional antennas. However, their directivity is only weak.

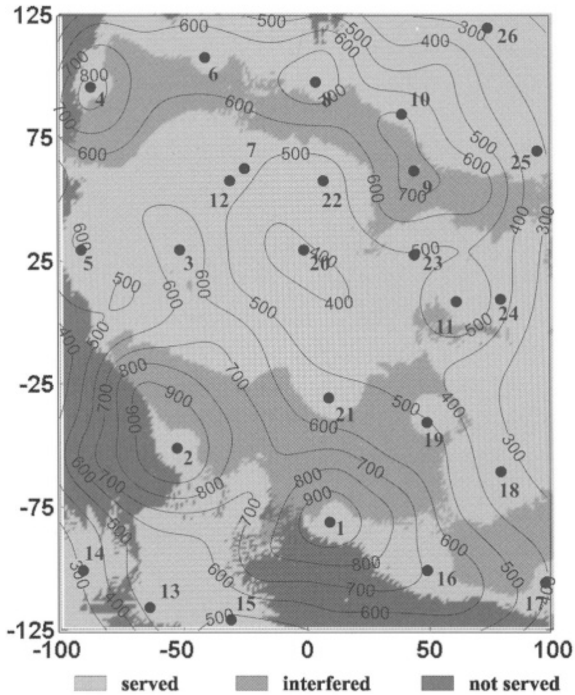

Figure 11.14: Covered area in the case of a full-blown optimization of the antenna diagrams.

Such a result is completely in line with the fact that as starting point a configuration has been chosen which already corresponds to a situation with a minimum of power being put into the network. Any application of antenna diagrams in first line further reduces the radiated power and thus would lead to an increase of the number of unserved pixels. But, exactly this has been prohibited for sake of comparison between a time delay and an antenna optimization. Thus, no significant gain in coverage can be expected by adjusting the antenna diagrams.

As the examples have shown the task to reduce the selfinterference in digital single frequency networks proves to be a very complex and very difficult problem. In particular, in the case of wide-area networks it is therefore mandatory to very carefully select the transmitters sites. In section 12 this aspect will be deepened further.

11.2.3 Planning Based on Population Density

So far the primary objective of all planing examples has been the achievement of wide-area coverage. Thus, all network parameters have been adjusted in such a way to maximize the served area, i.e. the to maximize the number of served pixels. This is a planning target which is typically aspired by public broadcasters. Usually, private broadcasters have different demands. For them it is less important to cover an area as large as possible. It is more interesting and economically crucial to reach a number of clients as large as possible at a minimal effort.

Such objectives can be modelled for the network planning process in two different ways. The first possibility is to define a special set of test points by selecting only pixels to which a number of inhabitants is assigned that is larger than a critical value (see section 10.1.3). It is clear that this corresponds to an a priori exclusion of test points which bears the problem that the selection of pixels has to be carried out very careful in order to produce reliable results. Furthermore, it is not guaranteed that a selection of test points appropriate for a particular planning target is also a reasonable choice for another problem.

In most cases it proves to be more reliable to base the planning process on a set of points uniformly distributed throughout the planning area with each of them being weighted by its associated number of inhabitants. Such an approach is independent from the actual planning objective and in addition offers a common basis in order to compare the results of different planning attempts. Figure 11.15 shows the result of a simulation based on the latter approach.

The set of test points utilized was the same as in the simulation above, i.e. a grid of $201 \times 251 = 50451$ pixels. Each of them has been associated a with a number of inhabitants according to the population density shown in figures 11.3 and 11.4. The initial network configuration as input to the simulation has been the situation where all 26 transmitters radiate 100 kW using an omnidirectional antenna. Both the time delays and the antenna diagrams of the transmitters have been kept fixed while the transmitter powers were adjusted such to maximize the number of served inhabitants. Table 11.4 summarizes the resulting ERPs.

Table 11.5 contains the calculated coverage figures for both pixels and inhabitants. In order to allow an easy comparison between this planning

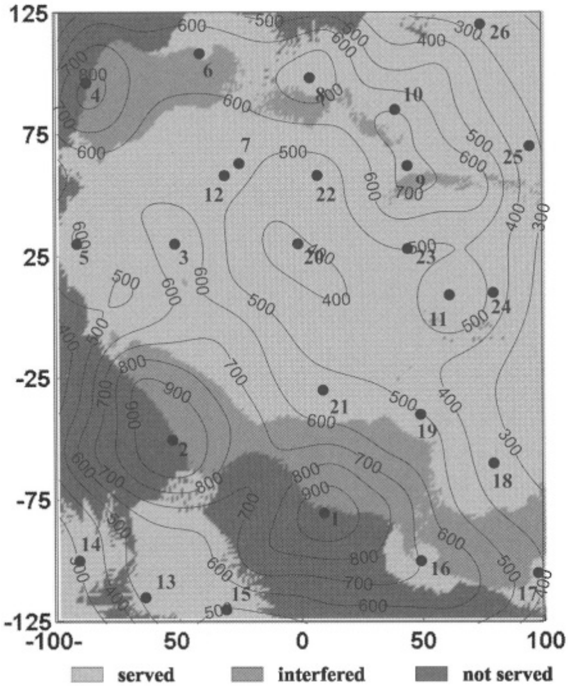

Figure 11.15: Covered area for an optimzation of the transmitter powers under the constraint to maximize the number of served inhabitants.

transmitter	1	2	3	4	5	6	7	8	9
power (dBkW)	-16	-16	18	-13	20	12	20	17	12
transmitter	10	11	12	13	14	15	16	17	18
power (dBkW)	16	19	20	20	19	20	13	14	20
transmitter	19	20	21	22	23	24	25	26	
power (dBkW)	20	20	20	19	17	20	16	20	

Table 11.4: Result of a power optimization aiming at a maximal number of served inhabitants.

type of simulation	number of served inhabitants	all transmitters 100 kW (no optimization)	power optimization
served pixels	32960 (65.3%)	30567 (60.6%)	31271 (62.0%)
served inhabitants	8337839 (83.9%)	7235901 (72.8%)	7372244 (74.2%)
interfered pixels	7881 (15.6%)	13191 (26.1%)	12457 (24.7%)
interfered inhabitants	989912 (10.0%)	2219678 (22.3%)	2083335 (20.9%)
unserved pixels	9610 (19.1%)	6693 (13.3%)	6693 (13.3%)
unserved inhabitants	607603 (6.1%)	479775 (4.9%)	479775 (4.9%)

Table 11.5: Comparision of different optimization results.

example and the former ones the results corresponding to figures 11.7 and 11.8 have been included as well. The case of figure 11.7 constitutes a network where all transmitters emit 100 kW using omnidirectional antennas while having identical time delays. Figure 11.8 represents the result of a power optimization whose aim has been a maximal extension of the coverage area under the constraint that the number of unserved pixels is kept the same as in figure 11.7.

Amazingly, the maximization of the number of served inhabitants leads to the largest number of served pixels as well. However, the price to pay thereto is a significant increase of the number of pixels where the minimum field strength criterion is not met. When looking at the corresponding power figures given in table 11.4, this comes as no surprise. In contrast to both former examples the maximization of the number of served inhabitants is achieved by significantly smaller ERPs. Thus, it is quite obvious that the number of unserved pixels increases.

On the other hand, it has to be concluded that the manifest demand to keep the number of unserved pixels fixed at the minimal possible value is heavily counterproductive for a transmitter power optimization in the context of a maximal wide-area coverage. Comparing the figures 11.8 and 11.15 directly clearly backs up the coverage improvement.

Once again, it is found that plausible and simple approaches for highly complex optimization problems do not lead to satisfying results. The better solution in any case seems to be the willingness to try out different formalized strategies like stochastic optimization algorithms.

11.2.4 Consideration of Network Costs

Economic efficiency becomes more and more the central issue of every network providership. During the days when there was only public broadcasting this was no big deal but nowadays every contender struggling for a broadcasting license which empowers the operation of transmitter network, has to prove the economic feasibility of the intended network implementation.

For private broadcasters cost-effective action is nothing new, it is self-evident. However, in the light of an increasing competition for listeners and viewers also public broadcasters are getting more and more confronted with the need to realize transmitter networks as cost-effective as possible.

Apart from license fees network costs primarily consist of costs for the supply of technical equipment, the operation of transmitter stations and their maintenance. It has to be distinguished between costs incurring only once and operating costs. The former cover mainly the acquisition costs of an appropriate realty to host the technical equipment as well as the antenna tower or mast. Clearly, the erection costs have to be included as well.

The operating costs include for example consumption of electricity for transmitter units and cooling systems. Basically, all costs being generated by the operation of the transmitter stations fall into this category. Furthermore, the broadcasting signals have to be transferred to the transmitters. If there is the possibility to do so via a satellite link the costs might be rather moderate. However, in case transfer via cable is needed the costs might rise dramatically.

All planning examples presented so far were based on the set of network parameters that have been introduced in section 10.1.4. By adjusting these parameters given coverage objectives like wide-area coverage or coverage according to population densities have been realized. If network cost are to be included, then appropriate cost functions have to be defined (see section 10.2.1 and [Lig99] for an alternative approach to cost optimization).

In principle, there are no restrictions in terms of applicability of stochastic optimization algorithms also to these type of network parameters. According to the requirements of the optimization task it is just necessary to model the cost specific features up to an adequate accuracy. When extending the quality function by which a network configuration is to be assessed later, care has to be taken with respect to the uniqueness of the quality function in terms of the network parameters linked to the network costs. This simply means that a given set of parameters must be associated with exactly one single value of the quality function. Clearly, different sets might lead to the same quality.

It has to be pointed out, however, that defining a suitable quality function is a task that is not trivially accomplished. Even if all relevant cost factors have been identified, still the question remains to be answered how they should be combined in a reasonable manner to build one unique quality function.

For a simple model it could be justified to link the cost a particular transmitter site produces to the power radiated from that site. It could be assumed to quantify the costs by defining them as proportional to the ERP. Unfortunately, this is no good approximation in reality. By virtue of the different technical infrastructure at two different transmitter sites it might be possible that the operation of a 10 kW transmitter does not produce the same costs everywhere. As a consequence, the costs for each transmitter site should be listed as a function of all other network parameters described in section 10.1.4.

In order not to overload the presentation the examples given here are based on generalized cost factors which in turn allow a straightforward definition of the quality function. Thereby the essential features and the impact of cost on the planning results can be demonstrated without problems.

For the first example costs proportional to the radiated power are to be taken into account only. The larger the power the larger the costs thereby incurred will be. In addition, fixed costs can be considered for each transmitter. too. They are meant to describe in general terms the costs connected to the installation and the operation of the transmitter site per se. Thus, an obvious mathematical form of the quality function would be

$$Q = \frac{V}{\sum_{k=1}^{N} \left[\alpha_k + \beta_k \{P_k\}^{\gamma_k}\right]} \tag{11.10}$$

or

$$Q = \frac{B}{\sum_{k=1}^{N} \left[\alpha_k + \beta_k \{P_k\}^{\gamma_k}\right]} \tag{11.11}$$

where the quantities V and B represent the number of served pixels or inhabitants, respectively. The summation runs over all transmitters. The parameter α_k accounts for the fixed costs of the k-th transmitter while by adjusting the parameters β_k and γ_k the impact of the ERP can be tuned.

Figure 11.16 shows two simulation results for the case of a quality function according to the definition (11.10). Both optimizations were based on the choice $\alpha_k = \gamma_k = 1$ for all k. The parameters β_k have been put equal to 0.1 or 1.0 for all transmitters. The objective of the simulation was in all cases to adjust the transmitter powers such that the function (11.10) takes a maximum value. Furthermore, all time delays were put equal to zero and all transmitters have been equipped with omnidirectional antennas. The results of the calculations for the case of quality function (11.10) are summarized in tables 11.6 and 11.7.

Substantially, the definition (11.10) means that the served area is to be maximized under the condition that the incurring costs should be as small as possible. Figure 11.16 very clearly shows the impact the expenses for transmitting power have. The transition from $\beta_k = 0.1$ to $\beta_k = 1.0$ corresponds in general terms to an increase of costs for ERP. Consequently, the total power output is reduced from 345 kW to 190 kW with a simultaneous decrease of the number of served pixels from about 45 % to 29 %.

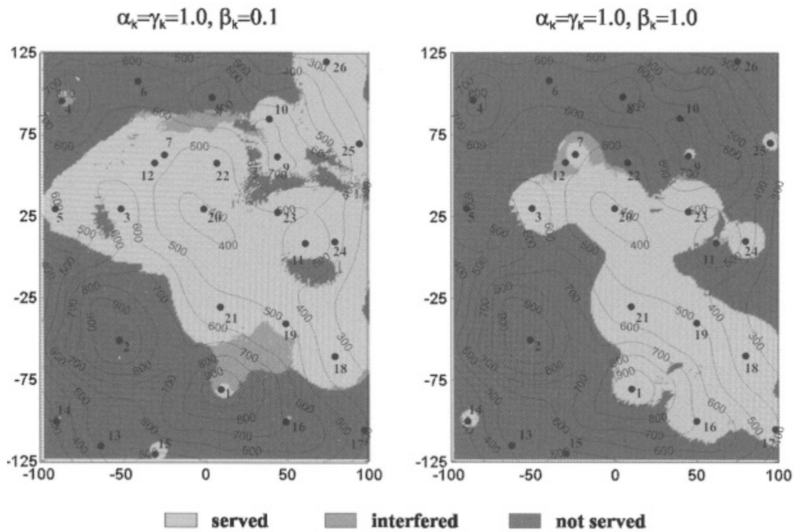

Figure 11.16: Covered area for an optimization based on the quality function (11.10) for two different sets of parameters.

transmitter	1	2	3	4	5	6	7	8	9
power (dBkW)	6	-20	18	2	10	-20	13	-20	11
transmitter	10	11	12	13	14	15	16	17	18
power (dBkW)	10	17	10	-20	-11	4	-6	-20	14
transmitter	19	20	21	22	23	24	25	26	
power (dBkW)	13	-20	15	11	-4	13	13	15	

Table 11.6: Result of an optimization of transmitter power under the constraints to maximize the served area and at the same time minimize the costs as given by equation (11.10). The parameters of the quality function $\alpha_k = \gamma_k = 1.0$ und $\beta_k = 0.1$ have been used.

transmitter	1	2	3	4	5	6	7	8	9
power (dBkW)	11	-12	12	-20	-9	-20	9	-16	0
transmitter	10	11	12	13	14	15	16	17	18
power (dBkW)	-11	-8	-9	-20	4	-11	13	10	14
transmitter	19	20	21	22	23	24	25	26	
power (dBkW)	11	13	15	-10	13	9	4	-11	

Table 11.7: Result of an optimization of transmitter power under the constraints to maximize the served area and at the same time minimize the costs as given by equation (11.10). The parameters of the quality function $\alpha_k = \gamma_k = 1.0$ und $\beta_k = 1.0$ have been used.

If the costs for transmission power goes up all transmitters will be fixed to smaller ERPs and primarily those are utilized which guarantee a large number of served pixels. Therefore, a coverage area is created which aligns along the valley regions of the planing area. There the wave propagation is not hindered by topographical obstacles.

A completely different result is obtained as expected if instead of the number of served pixels the number of served inhabitants is used in the quality function. Figure 11.17 shows the two results which have been derived on the basis of the quality function (11.11). The two pictures correspond to the same set of parameters as in the case of figure 11.16. As before, only the transmitting powers are optimized. All time delays are put to zero again and only omnidirectional antennas have been used.

Also in that case, there exists a significant difference between the low cost and the high cost situation. This time the total power radiated by the network changes from 312 kW to 203 kW. In contrast to the investigations of figure 11.16 now those transmitters are utilized in first place which are able to serve a large number of inhabitants even at moderate ERPs. Therefore, mainly those areas are covered in figure 11.17 which according to the representation of the population density 11.4 house the most people. Tables 11.8 and 11.9 provide the calculated ERPs of the transmitters in detail.

As already mentioned, the network costs which relate to the transport of the broadcasting signals from the central multiplexer to the transmitter sites cannot be disregarded in real network structures. If no transfer

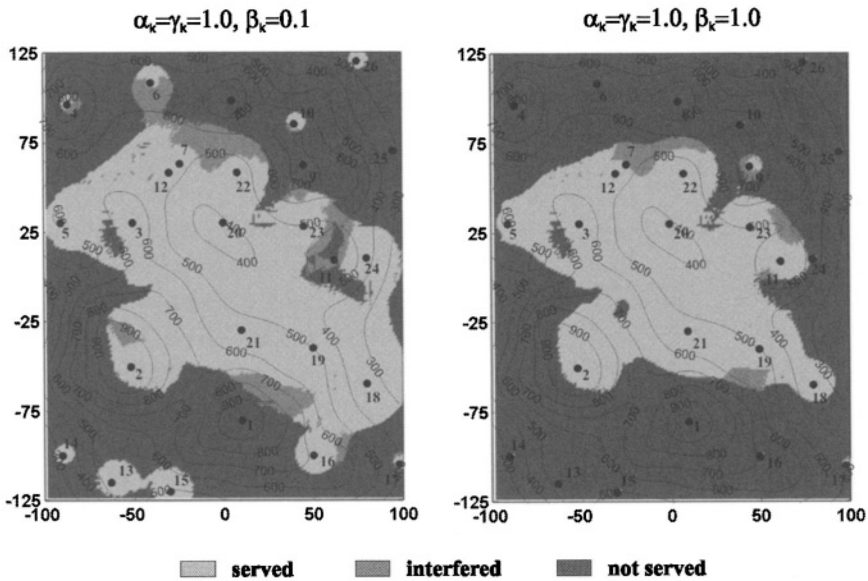

Figure 11.17: Covered area for an optimization based on the quality function (11.11) for two different sets of parameters.

transmitter	1	2	3	4	5	6	7	8	9
power (dBkW)	-20	15	17	3	10	8	12	-20	-20
transmitter	10	11	12	13	14	15	16	17	18
power (dBkW)	4	-20	10	9	4	10	10	2	13
transmitter	19	20	21	22	23	24	25	26	
power (dBkW)	18	2	12	11	12	13	-11	4	

Table 11.8: Result of an optimization of transmitter power under the constraints to maximize the number of served inhabitants and at the same time minimize the costs as given by equation (11.11). The parameters of the quality function $\alpha_k = \gamma_k = 1.0$ und $\beta_k = 0.1$ have been used.

transmitter	1	2	3	4	5	6	7	8	9
power (dBkW)	-20	14	18	-20	7	-20	-20	-20	4
transmitter	10	11	12	13	14	15	16	17	18
power (dBkW)	-11	11	8	-20	-11	-11	-12	0	8
transmitter	19	20	21	22	23	24	25	26	
power (dBkW)	14	3	14	12	11	-14	-11	-11	

Table 11.9: Result of an optimization of transmitter power under the constraints to maximize the number of served inhabitants and at the same time minimize the costs as given by equation (11.11). The parameters of the quality function $\alpha_k = \gamma_k = 1.0$ und $\beta_k = 1.0$ have been used.

via satellite link is possible this must be achieved by renting cable links. The price to pay thereto depends to a large extend on the data capacity required. However, the overall length of the cable link might be relevant, too. If transmitter sites are located in regions far outside the urban centres there might also be a shortage of appropriate cable infrastructure which might inflate prices, too.

A very simple model to map these conditions onto a mathematical quality function a least to a first approximation could be to assume a transmitter sites causes more expenses if its geographic height is larger. This traces back to the last sentence of the previous paragraph, namely that the higher a transmitter is situated the greater the probability becomes that there does not exist not sufficient cheap technical infrastructure. In mathematical terms a corresponding quality function could read

$$Q = \frac{V}{\sum_{k=1}^{N} \left[\alpha_k + \beta_k \left\{P_k\right\}^{\gamma_k} + \delta_k \left\{h_k\left(\mathbf{r}_k\right)\right\}^{\epsilon_k}\right]} \tag{11.12}$$

and

$$Q = \frac{B}{\sum_{k=1}^{N} \left[\alpha_k + \beta_k \left\{P_k\right\}^{\gamma_k} + \delta_k \left\{h_k\left(\mathbf{r}_k\right)\right\}^{\epsilon_k}\right]} . \tag{11.13}$$

The quantity $h_k(\mathbf{r}_k)$ denotes the geographical height above sea level of the k-th transmitter site while the additional parameters δ_k and ϵ_k can be used to tune the quality function as this has been done with β_k and γ_k already.

Depending on which values are chosen for the coefficients β_k and δ_k either the impact of the costs for power or costs for signal transport to the transmitter sites dominates. The upper transmitter power limit has been put to 100 kW here while according to the landscape introduced in section 11.1.1 the maximum height is 1000 m. If the coefficients β_k and δ_k in equations (11.12) or (11.13) are chosen such that their ratio is 1:10, it would be possible to explicitly study the competition between the power and transport costs.

Even if the models for network cost presented here certainly need to be refined in order to have any practical relevance, it seems to be evident that the method of stochastic optimization without doubt can be of great use in relation to planning digital single frequency networks under economic constraints. There is no principle obstacle to these algorithms, it is mainly a practical issue in terms of setting up reasonable models.

Chapter 12

Interference Free Allotment Areas

The features of digital terrestrial broadcasting systems show to advantage if they are operated as single frequency networks. As discussed already in detail, a very efficient spectrum usage can be achieved thereby which has an important impact on economic profitability, too. However, successful operation presupposes that the two aspects "frequency management" and "network planning" are not considered independently from each other.

International radio conferences are the forum where frequency resources are assigned to submitted requirements according to agreed principles. Those, who are asking for frequencies in order to provide services like T-DAB or DVB-T, have to very carefully define their requirements such that apart from any media political or economical constraints a reasonable network operation is feasible from a technical point of view.

Obviously, defining allotment areas as large as possible favors efficient usage of spectrum. As has been shown in the preceding sections this is, however, limited by the occurrence of selfinterference in wide-area networks.

Furthermore, it should be noted that not the absolute extension of allotment areas is the relevant parameter for efficient spectrum usage, but rather the ratio between the minimal extension of areas and the re-use distance determines the spectrum demand. If this ratio is larger than one only the directly flanking areas of a particular allotment have to be considered when looking for a frequency that is not in conflict with others. In that case, next but one neighbor areas are not relevant. This can be confirmed by looking at the results of section 6. If the ratio between typical minimal extension of an allotment and re-use distance drops below one additional spectrum is required.

In any case, it is crucial for network providers to carry out extensive studies concerning the size of allotment areas before they submit their requirements. Usually, such investigations are based on existing transmitter sites. Topographical conditions have to be taken into account in order to come to a realistic assessment of the coverage situation.

Nevertheless, theoretical investigations are helpful as well. They should be based on properly defined reference networks. These studies constitute important indicators with respect to the question what is technically feasible for a particular digital transmission systems and what cannot be realized. Both types of investigations will be covered exemplarily in the sequel.

12.1 Interference Free Reference Networks

The reference networks described in section 5.1 are usually employed in the context of allotment planning in order to assess the interference a single frequency network is producing outside its proper coverage area. This provides a basis for the determination of the re-use distances. If the distance between two allotment areas is larger than the re-use distance, then both allotments are allowed to use the same frequency without mutually causing any problems as long as the network implementation guidelines are respected. Clearly, the total field strength produced by the network does not vanish beyond the re-use distance. However, the total field strength has then fallen to a level which is not critical anymore.

Obviously, the network parameters of the reference network cannot be chosen in an arbitrary manner. The fixing of the parameters of the theoretical networks is mainly carried out according to features of real-world networks. Normally, the existing transmitter network infrastructure that will be used for the implementation is analyzed with the objective of deriving typical parameters like distances between transmitter sites, effective heights of transmitters as well as antenna heights above ground.

In reality, these parameters will show a great variation depending on the geographical region where such data is collected. Thus, a decision has to be taken which parameters are to be used when defining a reference network. By definition, a reference network is to mimic the behavior of a real network in a reasonable way. Therefore, parameters should be chosen which in a sense cover a great variety of real situations and at the same time do not refer to an extreme case.

Even though reference networks show systematic drawbacks with respect to real-world scenarios like for example the missing incorporation of topographic data, they possess the valuable advantage that by systematically varying the network parameters over a certain span of values the interrelations between the covered area and the set of network parameters can be studied in detail. These findings constitute an important hint in which way single frequency networks should be implemented in reality if a certain coverage target is to be achieved.

The more extended the coverage area of single frequency networks becomes the more efficient the spectrum can be is utilized. However, extended areas also means either a large set of low power stations or a large separation distance between the high power transmitter sites. The former case very likely leads to network costs nobody is able to bear. In the latter case the times of arrivals of signal impinging from different transmitters at particular points of reception might differ by more than the guard interval (see for example section 10.1.2). This can give rise to selfinterference. Clearly, the geometrical conditions are just a prerequisite thereto. The final decision about the occurrence of selfinterference rests on the ratio between wanted and unwanted signal contributions as well as on the receiver properties.

In order to visualize different situations that can appear in reference networks a very simple example for a reference network is used here. It consists of seven transmitters arranged in hexagonal shape which is a standard layout. There is no topography involved, i.e. all transmitter are assumed to be located on a plane. All transmitters radiate a power of 100 kW and possess omnidirectional antenna diagrams. The antennas are mounted at a height of 100 m above ground and all transmitters are fully synchronized, i.e. they all have time delays equal to zero.

In order to decide if a point in the plane is served or not the same wave propagation model as in section 11 is employed. This means a

served	$F_{\text{total}} > F_{\text{min}}$	$C/I > F_P$
interfered	$F_{\text{total}} > F_{\text{min}}$	$C/I <= F_P$
unserved	$F_{\text{total}} <= F_{\text{min}}$	C/I arbitrary

Table 12.1: Utilized coverage criteria. F_{total} denotes the total field strength level at the point under consideration while C refers to the wanted and I to the unwanted signal. The mimimum field strength is indicated by F_{min} and the protection ratio is F_P.

transmitter can provide a non-vanishing field strength at the receiver location if there is a line-of-sight connection between the antenna mount point and the receiver. Since in that case there are no mountains or valleys this imposes no restriction or complication. The decay of the field strength as a function of distance is the same as in section 11, too, namely a inverse-square dependence on the field strength.

As before, the arriving signals will be classified according to their times of arrival into wanted and unwanted contributions. The total wanted and unwanted signals are calculated by summing the powers of the corresponding signals, respectively. The receiver properties according to the functions (10.1) and. (10.2) from section 5.1 are taken fully into account thereto. Moreover, the same criteria to assess the coverage status of a pixel as in section 5.1 are applied. As a remainder, they are repeated here again in table 12.1.

Also the other parameters like minimum field strength F_{min} are chosen as in the calculations of the previous sections. This means that for F_{min} a value of $F_{\text{min}} = 58$ dB has been used while for the protection ratio F_P it was 20 dB. For the guard interval and the duration of the evaluation window the T-DAB figures of mode I in the VHF range, i.e. $T_G = 246\ \mu s$ and $T_W = 1000\ \mu s$ were employed. Thus, the critical distance between two transmitters beyond which selfinterference might occur is given by $d_{\text{crit}} = 73.7$ km.

The hexagonal structure used for the reference networks is highly symmetric. It is invariant with respect to rotations by an angle of 60° and multiples thereof. The symmetry axis is perpendicular to the plane of the hexagon and passes through location of the central transmitter. All pairs of transmitters have the same mutual distance d. Only the length

d of the edge has been varied in the examples given below. Thus, opposite transmitters are separated by a distance $2d$. If $d/2$ exceeds the re-use distance *R*, then formally the geometrical layout would allow for selfinterference.

However, even in the case of this network structure combined with the simplified wave propagation model described above, a slight overstepping of the critical transmitter distance does not immediately lead to selfinterference. The reason for that is the nonlinear decay of the field strength as a function of distance and hence the nonlinear way wanted and unwanted signals contributions are built. The following pictures in figure 12.1 are to give an impression how the coverage area varies if the separation distance between the transmitters is changed.

If the edges of the hexagon have a length of 60 km two opposite transmitters are separated by 120 km. This is already significantly more than the critical distance d_{crit}. Nevertheless, there is no selfinterference inside the hexagon area which is considered the primary coverage area of the reference network. Outside of this region the network does not provide service anymore. Any coverage is destroyed by selfinterference.

Since the distance between two transmitters is smaller than d_{crit} in any case, there are at least two transmitters that completely contribute to the wanted signal. This is valid for any point of the plane. The interfering contribution of the remaining transmitters obviously does not reach a critical level at least inside the proper coverage area and in the close-up range outside the hexagon. Moving further away from the outer edges of the hexagon the situation changes. Due to the characteristics of the wave propagation the interfering components begin to dominate.

If the separation distances between the transmitters increase, the interfered region moves closer to the hexagonal target coverage area. Between a transmitter distance of 70 km and 80 km perturbations inside the hexagon are found as well. These gaps could be closed only by introducing additional transmitters.

These examples very clearly demonstrate that the system parameters on one hand and the network parameters on the other hand need to be harmonized in order to obtain a uniform coverage free of interference. The cases with a transmitter separation of 60 km and 70 km, respectively, comply with this demand while the results for the other transmitters distances are not acceptable.

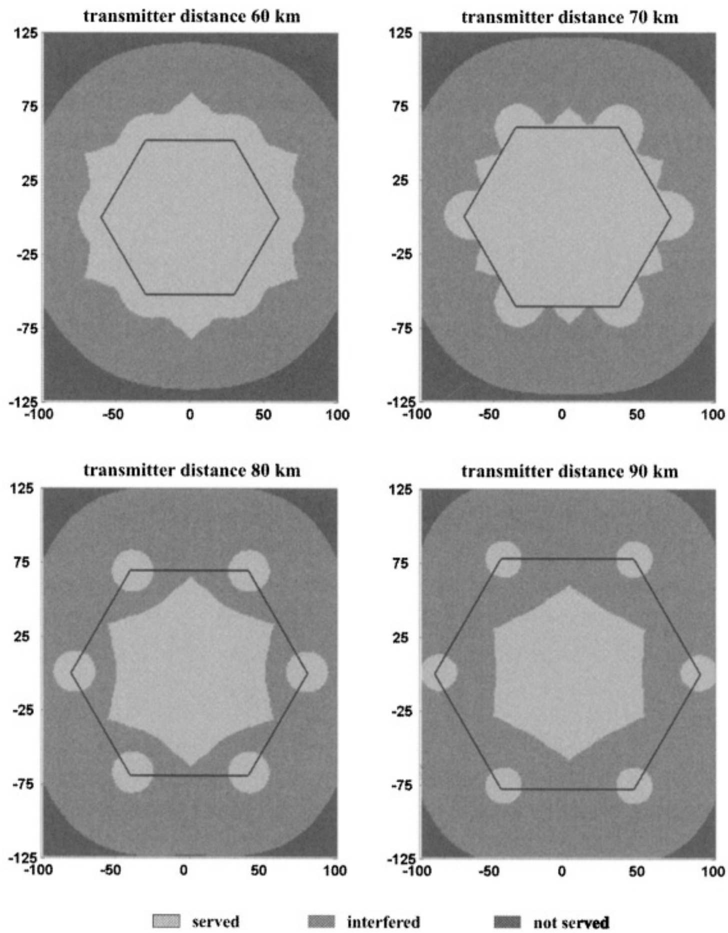

Figure 12.1: Covered area of four different reference networks. The transmitter sites are located at the vertices of the hexagon. One transmitter is situated in the centre.

An extension of the network structure by adding additional transmitters in accordance with the hexagonal symmetry in order to augment the coverage area is not possible in general. In the case of transmitter separations of 60 km or 70 km there are still only two fully constructively contributing transmitters while at the same time the number of interferers is growing. In particular for even larger inter-transmitter distances this can lead to a dramatic degradation of the coverage of the single frequency network. Figure 12.2 illustrates the situation.

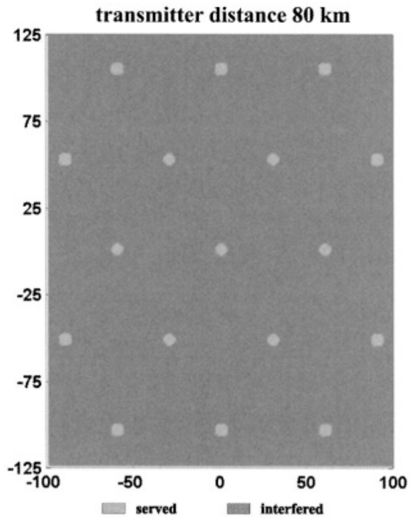

Figure 12.2: Covered area for an extended reference network. Service can only be provided in the close vicinity of the transmitters.

Only in the regions directly adjacent to the transmitters sites service can be provided. In fact, the minimum field strength criterion is met everywhere in the entire plane. However, the required protection ratio is not reached in most cases. All transmitters are radiating a power of 20 kW in this particular example. Changing the powers of the transmitters uniformly would not lead to an improvement of the coverage since the ratios between wanted and unwanted contributions would not be affected by such a change. The only consequence of a global reduction of the output power would be that areas would show up where the minimum field strength is no longer reached.

It has to be noted, however, that the failure of the extension of the reference network does not exclude the possibility to implement single frequency networks in reality having a larger extension than the one discussed here. The reason is that first of all the decay of the field strength is faster in reality than assumed here. This means that destructive impacts are not that far reaching. Second, the existence of mountains and valleys in real-world applications is an advantage when it comes to the implementation of wide-area single frequency networks.

12.2 Realistic Transmitter Networks

Theoretical reference networks are reflecting reality only inadequately. There are no real transmitter networks having a hexagonal symmetry in terms of the location of the transmitter sites. Also neither the antenna height above ground nor the antenna diagrams are the same for all transmitters. Quite the contrary, in most cases each transmitter has its own very special characteristic features not shared with any other transmitter in the network.

In principle, there is no problem to incorporate all these features into the concept of reference networks. However, when doing so the idea of reference networks looses its generality. Reference networks have been invented to qualitatively grasp the essential features of single frequency networks in a very general way. Furthermore, they are applied as some kind of placeholders in order to assess the compatibility of allotment areas which, until an explicit implementation of transmitters is carried out, do not have any substance. Basically, until the actual operation of transmitter takes place they are virtual objects.

Moreover, there is an additional fundamental difference between the usage of reference networks and networks in real environments. The former deliberately do not take into account any topographic data in order to establish a concept that is applicable in general terms. In contrast, the treatment of real networks should usually be as accurate as possible. Therefore, a lot of effort is put into the development of reliable wave propagation models on which the coverage predictions for real networks are based.

Difficult topographic conditions usually enforce more expenses in terms of network implementation. This concerns both the number of used transmitters as well as the radiated power. The examples of section 11 nevertheless show that sometimes even massive use of transmitting power does not guarantee complete coverage in topographically challenging regions.

On the other hand, topography often is the guarantor for the implementation of wide-area networks whose dimensions extend far beyond the critical distance derived from the underlying guard interval. Nature sometimes has mercy on planners and network providers in that respect. Mountains and valleys make sure that sometimes geographical regions are decoupled which without topography would otherwise perturb each other.

Figure 12.3 gives a very impressive example thereto. Two special coverage calculations are shown which are based on the fictitious network planning scenario that has been extensively discussed in section 11. The left picture in figure 12.3 is identical to figure 11.7 which corresponds to a situation where all 26 transmitters radiate 100 kW via omnidirectional antennas. In addition, all time delays are put equal to zero. The right picture illustrates the coverage that is obtained from the same network configuration with the only difference that topography has been virtually switched off. The wave propagation model is the same for both cases, too.

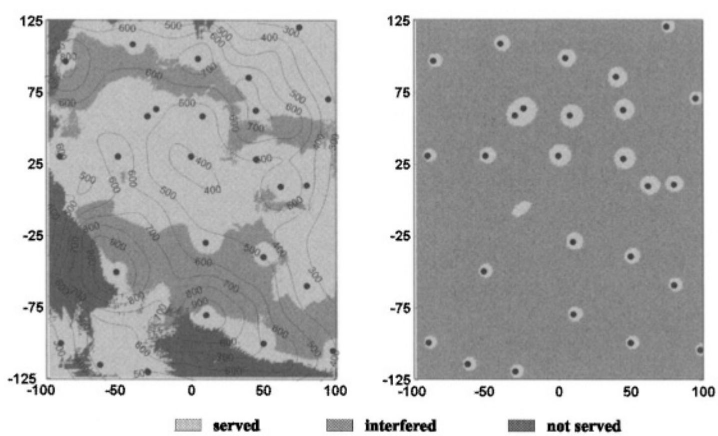

Figure 12.3: Comparision of the coverage with and without topography. The left picture corresponds to figure 11.7. The right picture has been generated by keeping all characteristics of the transmitters constant but vitually switching off topography.

The difference between the two scenarios is amazing. Without topography the coverage situation is similar to the extended reference network case above in figure 12.2. If a more sophisticated wave propagation model would have been used the differences certainly would not be that pronounced, qualitatively the situation would be the same.

In order to illustrate the connection between efficient network planning on one hand and a efficient usage of spectrum on the other, it is assumed for the following examples that the rectangular planning area shown in the pictures of figure 12.3 would represent the national boundaries of a fictitious state. The radio and telecommunications administration of this state would intend to set out the framework to provide a digital service like T-DAB or DVB-T throughout its entire territory. This intention would need to be cast into the definition of a set of allotment areas which could be submitted to an international radio conference as requirements.

If the radio conference would be able to elaborate a frequency plan in the sense that all requirements could be satisfied by assigning appropriate frequency resources, then the administration of the considered rectangular state would have the right to utilize several T-DAB blocks or TV channels under prescribed conditions. Therefore, the frequency technical basis would be given to open up a tender for network providers which are interested in setting up and running digital terrestrial broadcasting networks. The success of the implementation of single frequency networks with respect to the desired coverage targets would primarily depend on the fact if the allotment areas have been properly defined.

In view of the coverage situation in the left picture of figure 12.3 it would certainly not be very reasonable to define the entire national territory as one single allotment. Those regions which are affected by selfinterference have a dimension too large to be acceptable. The whole territory needs to be divided into at least two smaller parts.

Very often the definition of allotment areas is influenced and finally determined by political constraints in terms of mapping regional political or cultural structures to corresponding coverage areas. If an efficient usage of spectrum can be achieved, quite often has then secondary priority. Especially in the case when a multitude of very small areas is to be served individually and independently efficient use of frequency resources cannot be obtained. As has been already discussed in sections 7 and 8, allotment areas whose extension is smaller than the relevant re-use distances lead to an increased demand for spectrum.

If media political constraints need not be respected, the definition of appropriate allotment areas corresponds to the subdivision of a large region into smaller fractions such that as few as possible areas are established. The necessary number of areas is determined by the maximum extension of an allotment area that can be realized without generating selfinterference.

Usually, the starting point for this subdivision process is a set of all potentially usable transmitters scattered throughout the planning areas and endowed with realistic transmitter characteristics. An allotment can be created by selecting an appropriate subset of transmitters in such way that the coverage area created by them is free of interference. Depending on whether selfinterference is encountered or not transmitters have to be removed from the network or additional ones can be added. Then, the final coverage area can be approximated by a polygon which determines the allotment area. Clearly, this process of establishing an allotment represents a combinatorial optimization problem. Some methods to cope with these kind of problems have been repeatedly discussed in this book.

As alternative approach there is always the possibility to employ heuristic trial-and-error strategies. In contrast to other problems in the field of frequency assignment the chances to find a satisfying solution are pretty good for the case at hand. In practice, certainly geographically adjacent transmitters are being combined to form a single frequency network. Thus, it is promising to start with a core set of transmitters and successively add other transmitters to the network. That way it is possible in most cases to very quickly design an allotment. Figure 12.4 shows a decomposition of the original planning area into five allotments.

For the example given here, the generation of the allotment areas was started in the upper left corner of the planning area and then subsequently extended over the entire planning region. The total area which is served by any of the five networks implemented inside the corresponding allotment areas is shown in the lower right picture of figure 12.4. The black lines indicate the boundaries of the allotments as they are derived by the application of the above mentioned trial-and-error approach.

Two issues have to be borne in mind when looking at the result of figure 12.4. First, there is a significant overlap in terms of coverage from one allotment into the adjacent ones. In practice, this could avoided only by employing appropriately designed antenna diagrams. Furthermore,

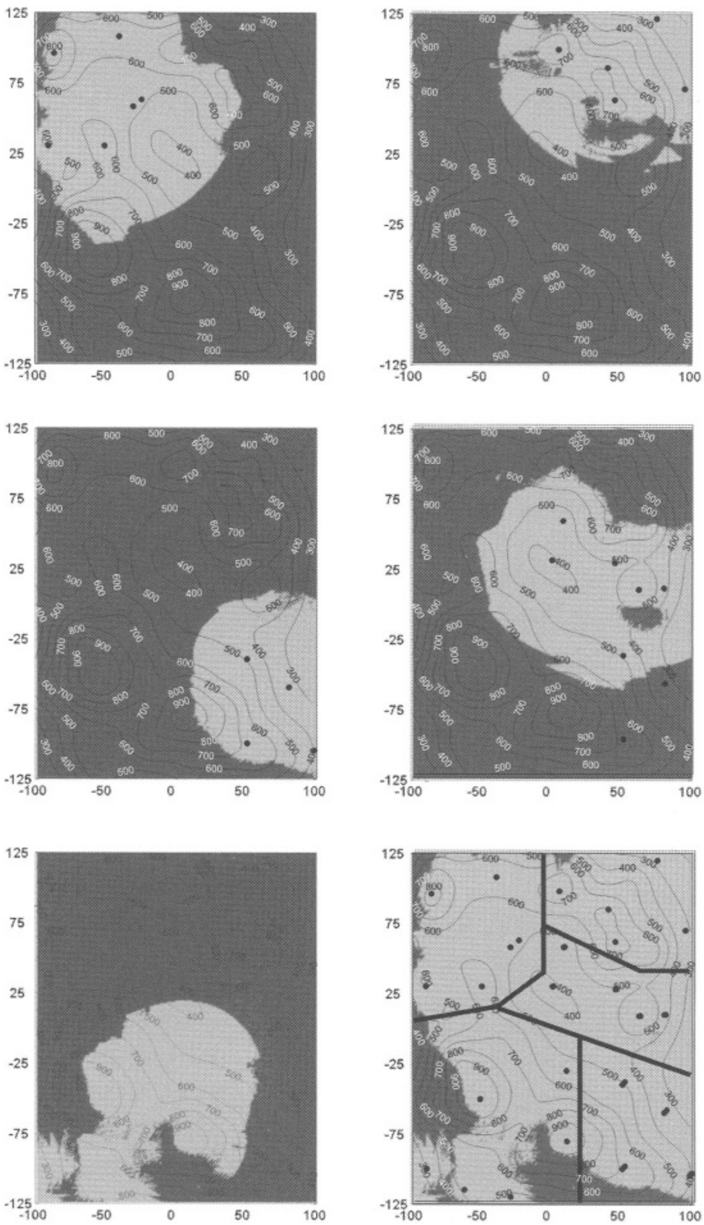

Figure 12.4: Decomposition of the original planning area into five allotments. The picure in the lower right corner shows the union of all pixels served by the five allotments. The black lines indicate the allotment boundaries.

inside some of the allotments there are still regions which are not covered because the minimum field strength is not reached. These gaps cannot be filled on the basis of the available transmitters under the restriction to respect an upper ERP limit of 100 kW which has been respected for all examples.

In a real network, the only way out of this problem would be to include further transmitter sites. Further problems with respect to the occurrence of additional selfinterference would very likely not be an issue. The reason is that the dimensions of the allotment areas have been chosen properly. All allotments have a maximal extension which is in the range of the dimensions of those reference networks from figure 12.1 that exhibit no selfinterference.

The decomposition of the whole planning area given in figure 12.4 is certainly a possible choice. However, when comparing the allotment areas with the population density shown in figure 11.4, it becomes clear that the allotments are not compatible with the population distribution. In practice, no administration would allow an allotment boundary to cross a large city or a connected urban region as long as there are no real technical restrictions that enforce such a solution.

Seen from a media political point of view it would be favorable to start with the central urban region and try to establish an allotment comprising this population centre. Figure 12.5 depicts a different decomposition of the planning area in also five allotments reflecting this idea.

In the southern part a very large allotment is created which, nevertheless, does not exhibit any signs of selfinterference. A brief examination proves that the single frequency network there is indeed implemented by employing transmitters which are separated by more than the critical guard interval distance. However, they are geographically decoupled. On the other hand, the allotment under consideration shows extended unserved regions. In order to close these coverage gaps additional transmitters would be needed. Probably, it would be even reasonable to subdivide this allotment into two parts.

Many other decompositions of the entire planning area into allotments are conceivable. In practice, allotment areas are designed by media political guidance. However, the aspect efficient usage of the precious resource spectrum has to be considered seriously. Otherwise all participants of the frequency assignment and the subsequent network implementation

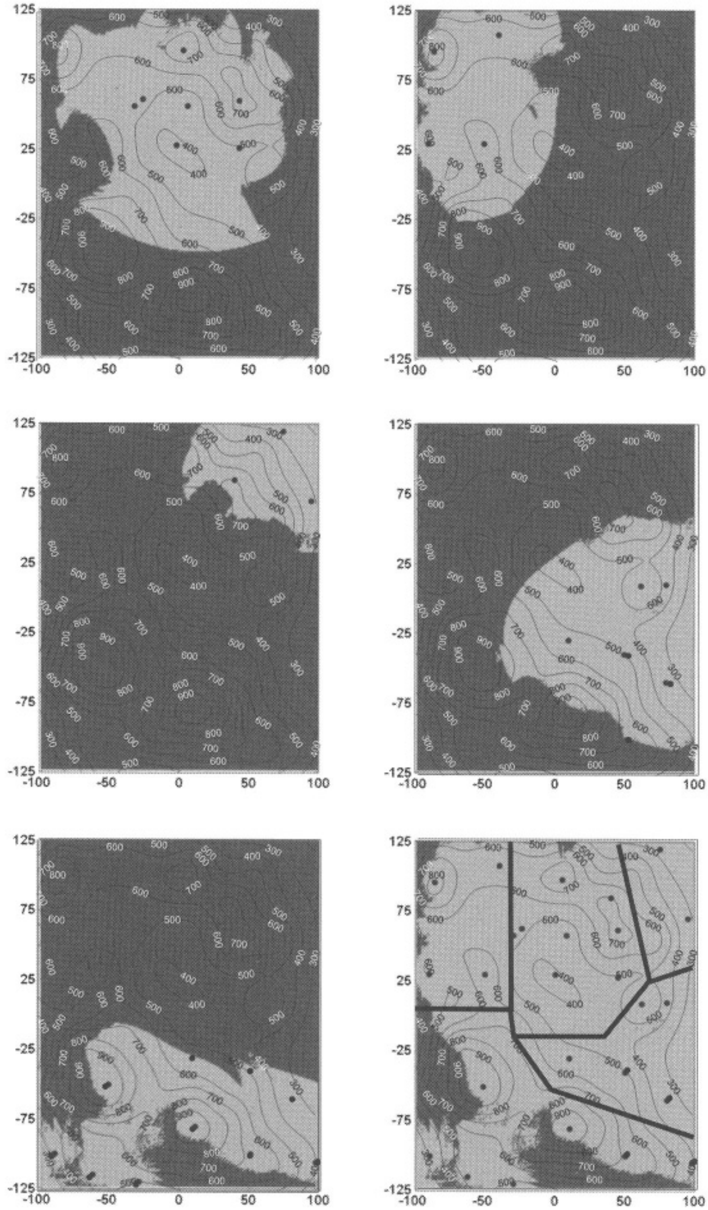

Figure 12.5: Second example for a decomposition of the original planning area into five allotments. The picure in the lower right corner shows the union of all pixels served by the five allotments. The black lines indicate the allotment boundaries.

are running the risk that the existing potential digital terrestrial broadcasting without doubt endues is exploited only in suboptimal manner and transmission capacity is wasted.

Appendix A

Stochastic Optimization

Both frequency assignment as well as network planning fall into a class of problems which on a first glance seem to be trivial but prove to be highly complex and involved on second consideration. The crucial point in both cases is the fact that there exists a gigantic number of potential solutions. In order to assess a single solution, usually, a number of sometimes complicated calculations has to be carried out. Thus, it is prohibitive even in the simplest cases to simply try to assess all existing solution sequentially in order to find the optimum.

A simple example taken from the field of frequency assignment is to illustrate this. It is assumed that there is a set of M geographically distinct areas which are to be given a TV channel each. However, this cannot be done freely. Instead, there is the constraint to be obeyed that geographically adjacent areas are not allowed to use the same channel. The total number of available channels should be N. Each of them should be usable in any area.

Clearly, a channel assignment onto the areas should be established which employs as few channels as possible. Thus, there is a variety of N^M different channel distributions. The only information available on which an assessment of a frequency assignment could be carried out are the adjacency relations between areas. This, however, is not enough to exclude a priori certain solutions. Hence, in principle all N^M would need to be checked in order to find the best solution.

Now, if N and M are fixed to $N = 20$ and $M = 10$ for example, then there are 20^{10} different configurations. If the assessment of a single solution takes only 10^{-6} sec, the exhaustive search for the best solution would require 2845 hours or 118 days! Since it is well known that typical frequency assignment problems dealt with at international radio conferences must cope with several hundreds of coverage areas to be assigned frequencies to, it becomes clear that even with the help of the fastest today existing computers such an approach must fail.

A very similar situation is encountered in the case of network planning. Typical wide-area T-DAB or DVB-T networks encompass a set of 10 - 50 individual transmitters that are operated as a single frequency network. Each of them is characterized by fixing the set of network parameters such as the radiated power, its antenna diagram and the individual broadcasting time delay which might be used to synchronize the whole network and to reduce selfinterference.

Usually, antenna diagrams are defined in terms of 36 numbers which correspond to a reduction of the effective radiated power towards a given radial direction. The standard of 36 values simply stems from the fact to use a 10 degree step between two directions. Therefore, every transmitter in that simple model is described by a set of 1 + 36 + 1 = 38 parameters.

In principle, these parameters can take any real value. However, practically only distinct values taken from a given set are used for planning purposes. For example, both ERP and antenna diagrams are usually given in terms of multiples of 1dB within certain limits while for the time delays the basic unit is 1 μsec. Now, assuming for all parameters a 10-member set each transmitter can be configured in 10^{38} different ways. If the single frequency network is build by a medium number of transmitters, let's say 30, there are $30 \times 10^{38} = 3 \times 10^{39}$ possible configurations of the network as a whole. As before, any brute force approach to find the best solution must necessarily fail.

From a mathematical point of view, both planning problems just described fall into the class of NP-hard problems. This characterizes a task whose systematic solution by testing every existing possibility leads to computing times which grow exponentially or even faster with respect to the relevant system size, i.e. the number of individual objects or parameters. In the case of typical system sizes encountered in real-world problems very quickly orders of magnitude are reached that can no longer be tackled that way.

In view of these well-known difficulties there has been put forward a tremendous effort during the last three decades in order to develop powerful methods which allow at least for a very efficient approximative treatment of the underlying high-dimensional optimization problems. They have been successfully applied to frequency assignment or network planning problems. These methods are the so-called stochastic optimization algorithms.

Stochastic optimization algorithms come in several different flavours. The core feature of all different approaches is to carry out a stochastic search in configuration space which in contrast to pure Monte-Carlo methods is accompanied by mechanisms that more or less effectively guide the search towards the desired solution. What is very important, however, is the fact that in any case, the basic idea consists of not seeking the one and only global optimal solution, but to accept a "nearly optimal" solution as well. The great advantage is that such a solution can be found within a reasonable period of time.

Meanwhile there is a vast number of different stochastic optimization algorithms. Some of them show only marginal differences. Therefore, only two approaches are described here in detail which have been exploited in the field of frequency and network planning. For further examples it is referred to the literature.

A.1 Great Deluge Algorithm

A special representative of stochastic optimization algorithms is the so-called Great-Deluge algorithm (GDA) [Due93]. Even though it is amazingly simple it produces very good and robust results. Its applicability is restricted by very few conditions only. Typical fields of application are problems in which a system is described in terms of a large set of parameters which determine the quality of the system in a very complicated way. Examples can be found in physics or economics when the degree of efficiency of a technical system or some logistic scheme has to be optimized.

The most important condition that has to be met in order to apply the GDA to a particular optimization problem is that the status of the system under consideration can be fully described by fixing a set of parameters. Further, it must be possible to quantify the status of the system in a non-ambiguous way. This means a quality function must exist which maps the set parameters to a reasonable quality measure. Finally, it is required that good configurations can be distinguished from bad ones and that the difference can be quantified numerically as well.

There are on the other hand not many restrictions on the explicit form of the quality function neither on the definition of the parameters. The domain of the parameters can be the set of integers or real numbers but in principle complex number are conceivable as well. The parameters might be restricted to certain intervals or even to a discrete set of values. The same is true for the quality function. It can be a continuos real-valued function or an integer function. Only complex-valued quality functions would create problems because in that case there is no clear mathematical distinction between good and bad.

Seen from a mathematical point of view the set of parameters spans a high-dimensional configuration space. The quality function can be visualized in the sense that it constitutes a very complex mountainous region above that configuration space as a function of its parameters. Depending on the domain of the parameters the quality function need not be a continuos more-dimensional surface. If the set of parameters contains both real and discrete values, then the quality function necessarily has discrete or continuos character in different directions of the configuration space.

The mathematical task underlying the optimization is to find a set of parameters that corresponds to a state or configuration of the system having optimal quality. In the case of the Great Deluge algorithm this is accomplished in an iterative manner based on a stochastic search in configuration space. Decisive for the success of the method is that the accessibility of certain regions in configuration space is dynamically reduced, i.e. more and more sectors become inaccessible as the simulation proceeds.

The course of the GDA can be summarized as follows. Starting point of the optimization process is an arbitrarily chosen configuration whose quality is calculated first. Then a new configuration is generated and its quality calculated as well. According to a given scheme the new configuration is either accepted as the new starting point or not. This procedure is continued until a particular stop criterion for the simulation is met.

Basically, this whole process is governed by two factors. First, the newly generated configurations should be in the "vicinity" of the old ones. This presumes that a vicinity of configurations can be defined at all. In most cases, an adjacent configuration can be generated by slightly changing the values of the parameters of the configuration at hand. Then, the quality of the new configuration will change only slightly, too. In the case of real parameters and smooth quality functions this is obvious. However, in general this might be far from being trivial. Depending on the structure and the complexity of the quality function it might be not possible at all.

Second, the decision which configurations are accepted as new starting points is taken dynamically during the simulation. The decision criterion applied is if the quality of the newly generated configuration is better than a dynamically increasing lower bound or not, the so-called water level WL. If the new quality is better than WL, then the new configuration is accepted as new starting point even if its quality is less than the old quality! Thus, degradations are temporarily accepted. Only if the new quality is smaller than WL the corresponding configuration is rejected and a new one has to be tried out. Depending on the change in quality when going from one configuration to another the water level is increased after each step. This way the dynamic blocking of certain parts of configuration space is realized.

The name "water level" stems from the way in which the optimization process can be visualized and which finally has been the origin of the name "Great Deluge" algorithm. The situation is similar to a blind man trying to find his way up to highest mountain peak or at least to a very high one. His situation is aggravated by the fact that on his way upwards it starts to rain so heavily that the valleys start to fill. Great Deluge has finally started. Furthermore, he is barefoot which certainly should be avoided in high mountains.

Since the man does not see anything he can only make one small step after the other. So, from where he actually stands he makes one single move into an arbitrary direction. If his feet become wet he immediately goes back to his former position and tries another step. If ending up on try land the man accepts his new position. While the water level is permanently rising the man is pushed up the hill towards a high peak. For the man's sake the rain should stop once he arrives at a peak from which there is no escape anymore due to being encircled by water. In mathematical terms, this corresponds to the stop criterion. Figure A.1 illustrates the process in terms of a one-dimensional example.

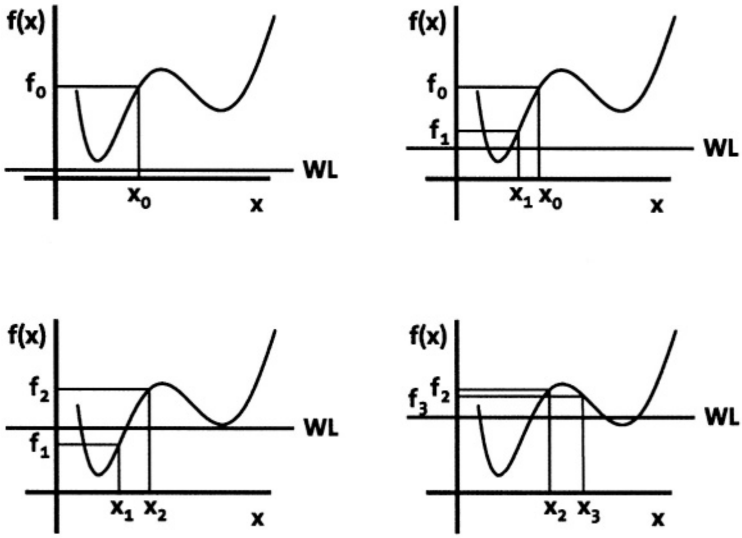

Figure A.1: One-dimensional example for the Great Deluge algorithm.

At the beginning of the simulation the initial configuration x_0 is chosen and the corresponding quality f_0 is calculated. Furthermore, the water level WL is fixed. The first iteration step consists of slightly changing the value x_0 and thus position x_1 is reached. Its quality reads f_1 which in example given here is smaller than that of x_0 but larger than WL. Thus, it is accepted as the new starting point. Next, the water level is raised.

During the second step the location x_2 is reached. If during the attempt to leave point x_1 a value x_2 is chosen which leads to quality smaller than WL, then this it would be necessary to reject this x_2 and try to find a better one. Repeating this procedures a sequence of points x_k is tested and finally a good solution can be found. The crucial point, however, is that by accepting temporarily reduced qualities it is possible to escape from the local maxima as is illustrated by the third step in figure A.1. Thus the path is open to reach the global maximum in this case. Figure A.2 shows the course of the GDA in term of a flow chart.

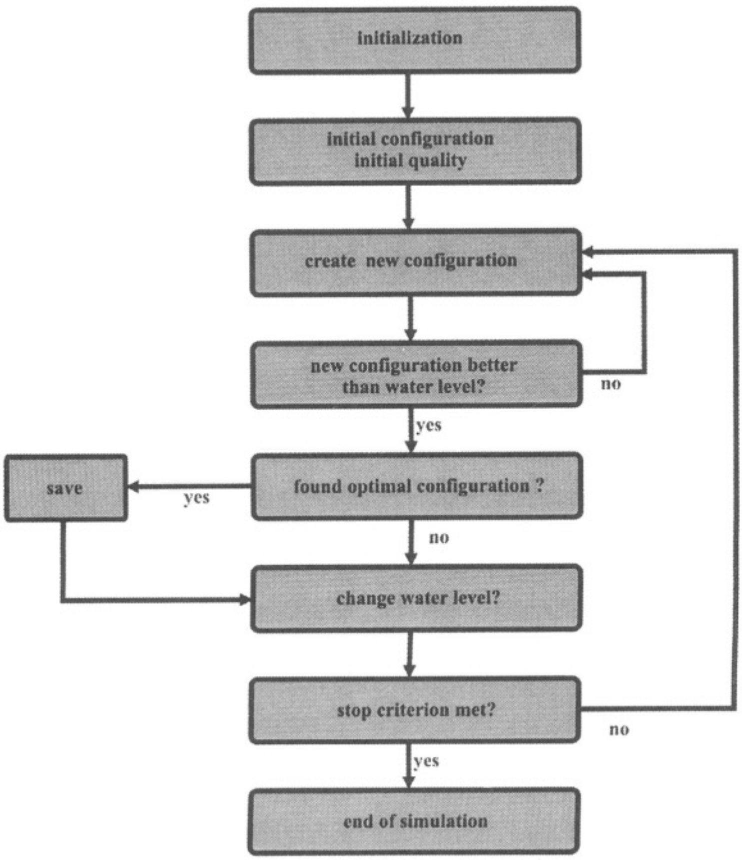

Figure A.2: Flow chart of the Great Deluge algorithm.

A.2 Evolutionary Strategies

Besides all algorithms that have their origin in different mathematical considerations there is one omnipresent, very flexible and overall very successful strategy to cope with high-dimensional optimization problems. Nature itself applies this strategy which is evolution. On the basis of recombination, mutation and selection it has been possible to create an almost unbelievably huge biological diversity on earth. Creatures were able to adopt to sometimes extreme living conditions.

The application of the underlying principles to the solution of high-dimensional optimization problems is rather young discipline. At about the end of the eighties the first attempts into that direction have been made. Since that time there has been great progress both with respect to the development of new algorithms and the application to concrete practical problems. A vast number of articles has been produced and published. In order to get a first impression of the field it is referred to [Abl87] and [Hei94].

In contrast to the Great Deluge algorithm, evolutionary strategies (ES) do not consider only one current configuration. Instead, an entire set, a so-called population, is dealt with at the same time, i.e. during one simulation step. A single configuration, called an individual, is characterized by its "genes" which is nothing but a set of parameters quantifying its features.

According to the principles of evolution, an simulation step consists of three parts. First, a number of offspring have to be generated by starting from an actually existing population. Then, each of the new individuals has to undergo several mutations by which the values of its parameters are affected. Finally, a subset of the entire set of new individuals is selected which is to build the new population.

The mechanisms of each step have to be specified explicitly. In nature, parents pass theirs genes to their children. In general, a child gets a full set of genes. Some of them derive from the mother and some from the father. The explicit apportionment is left to chance. In the context of an evolutionary strategy this means that parents, i.e. individuals of the current population, pass their genes, i.e. their parameter values to a child, i.e. a new individual. As in nature, who is passing which parameter value is a question of chance as well.

Clearly, for a computer simulation there is no need to literally adopt the biological principles put into practice by nature. Hence, it is possible that one child has more than two parents. This is one of the degrees of freedom that can be used to adjust the algorithm. Also, it has to be specified how many children are generated during one iteration. This number could be fixed or there is no principle objection to let it float according to a particular scheme.

The next step, namely mutation, needs to be defined properly as well. The number of parameters to be changed and also the extent of change must be fixed. It is possible to apply a modification to all parameters or to only change some of them, however, in a drastic way.

Finally, the population has to be set up. Also this step allows for a large amount of freedom. The new population could be recruited exclusively from the set of children. It is nevertheless conceivable to build a new population on the basis of both parents and children. Furthermore, what is the criterion for an individual to be included into the new population? Is it only the corresponding quality, are the individuals selected completely by chance or is it a mixture of both?

From this discussion it becomes clear that in comparison to the GDA evolutionary strategies bear a lot more degrees of freedom. This can be helpful in some cases in order to adapt the algorithm to a particular problem in a perfect way. However, on the other hand it is evident that a lot more effort has to be put into the application of the ES as with the rather simple GDA. Figure A.3 illustrates the basic ingredients of an evolutionary strategy in terms of a flow chart.

From a conceptional point of view, evolutionary strategies constitute some kind of generalization of the simple type of algorithms like Great Deluge. Not only one configuration is treated at one time but a whole set of system states are dealt with and assessed during one iteration step. This demands a significantly larger administrative effort. Correspondingly, the requirements in relation to computer resources like RAM are getting more severe, too. In addition, the computing times are increasing because not only one but several configurations have to be dealt with. On the other hand, the evolutionary algorithms are usually converging much faster towards an acceptable final solution.

The great difference between the ES and GDA lies in the selection and the acceptance of new configurations, respectively. Basically, in the case of the GDA there is a global criterion according to which it is decided if a new configuration is accepted or not. If the quality is larger than the water level the new configuration is taken independently from the difference in quality between the old and new configuration.

In the case of an evolutionary strategy a relative criterion is employed. If for example the twenty best configurations are supposed to build the new population then the qualities certainly show some variation. A moderate configuration might be rejected in the context of one particular

Figure A.3: Flow chart of an evolutionary strategy.

population but could have a completely different significance with respect to another population.

Moreover, the modifications that are allowed in ESs usually lead much more drastic changes as is typically accepted in GDA-type approaches. Using part of the values of the whole parameter set from one parent and another part from another individual quite naturally gives rise to discontinuities in terms of quality.

A.3 Parallel Algorithms

Apart from their efficiency and their robustness stochastic algorithms like the ones just presented offer an additional feature which makes them very attractive from a computational point of view. Nearly all methods can be parallelized in different ways without problems. All algorithms posses parts which not necessarily need to be processed sequentially. Thus, quite naturally a performance gain can be expected if these subtasks are processed in parallel.

There are two basic ways to carry out parallel computations. The traditional approach is to utilize several connected workstations. The computational tasks are then distributed across the local computer network onto the individual machines. Nowadays, combining computers across the internet is also an option. The distributed computation usually is established on the basis of remote procedure calls whereupon processes are started on remote machines in order to carry out some work.

If a particular computer is equipped with more than one processing units, then this fact can be exploited for parallel computing purposes as well. Multi-thread programming techniques are the right tool to make optimal use of this potential. For a detailed presentation of multi-thread programming it is referred to [Lew96] and [Kle96].

It is clear that a multi-tasking operating system is a prerequisite for any multi-thread programming activities. Multi-tasking operating systems can do more than one job at given time by running more than one process concurrently. In principle, a process can do the same thing by starting more than one thread at the same time. Each thread is a distinct program flow which carries out individual programming statements. One thread for example can open a window for displaying some graphic on screen while a second is occupied with I/O tasks and the third is making some numerical calculations.

What is the fundamental difference between a process and a thread? A process is a unit belonging to the operating system. It is characterized by a virtual memory map, file descriptors, user-IDs, source code, and so on. The only possibility of a program to access the data of the process or to enforce some change of the status of the process is via a system call which is a very expensive action where the phrase expensive refers to computer resources and computational time.

In contrast, a thread is a unit on user level which can be manipulated by ordinary user functions. Thus, a process associated to particular executable code can generate independent entities, namely threads, in order to carry out dedicated tasks independent of any other concurrent operations. This structure allows the generation of threads on different CPUs of a single machine without the need for special hardware.

If a program consisting of just one single thread requires some task to be carried out by the operating system the whole program must wait until the operating system has finished that task. Then the next program command can be executed. In contrast, a multi-thread program can execute other commands while waiting. Thereby, the total performance is increased and computing resources are utilized in a more efficient way. Performance is enhanced also by the fact that both the creation of threads as well as the inter-thread communication are much faster then the corresponding process operations. For a more detailed description of multi-thread programming techniques it is referred to [Lew96].

A closer look on figure A.3 reveals a quite obvious way to parallelize the previously discussed ES algorithm. During each iteration step a prescribed number of children is generated and assessed. Usually, this is carried out sequentially. Hence, it would be possible to exploit the multi-thread programming technique in order to start a set of threads each of them having the task to generate and assess exactly one child. Once this has been done the program proceeds again along the traditional line and decides which individuals are to be included into the population.

When assessing a configuration it is very often possible to subdivide this task into distinct pieces which could dealt with independently at the same time. Such a situation is found for example when calculating a coverage prediction for a single frequency network (see section 3). For every geographically point representing the coverage area the same set of calculations has to be carried out. Without problems the whole set of pixels could be subdivided into a number of smaller sets each of them being dealt with by an individual thread independent from the others. At the end the results of all threats are combined to give the total coverage.

Also for the Great Deluge algorithm parallelization is a conceivable way to improve the computational performance. The flow chart from figure A.2 could be modified by creating not only one new configuration per iteration step but allow the creation of a multitude of them. Both creation and assessment of these configurations could be carried out concurrently in terms of independent threads. Then, the selection mechanism is adapted in the sense that either the best or the worst configuration just above the water level is chosen as the new starting point.

The question if parallelization is to be exploited or not and in which form, i.e. via multi-thread programming techniques or via remote procedure calls to other machines, cannot be answered generally. This depends very heavily on the details on the considered optimization problem. In some cases parallelization is a good idea in others it is of no help because for example the administrative overload exceeds the potential performance gain.

In connection with the planning of digital single frequency networks investigations have been carried out to study the impact of multi-thread programming techniques on computational performance [Beu98b]. To this end, the coverage has been calculated for a given network within a prescribed planning area with the help of threads the way described above on a multi-processor machine. It has been expected that a significant performance gain could be achieved. However, the results seem to dampen the enthusiasm a little bit. Unfortunately, the simple equation N CPUs = N threads = N times faster is not valid. Increasing the number of threads inevitably increases the administrative effort as well which obviously limits the performance gain. Nevertheless, an improvement of about $N/3$ times faster, where N is the number of CPUs, can be realized without problems.

Appendix B

Abbreviations and Symbols

The following abbreviations are used throughout the book:

COFDM	:	Coded Orthogonal Frequency Division Multiplex
CEPT	:	Conférence Européenne des Administrations des Postes et des Télécommunication
DAB, T-DAB	:	Digital Audio Broadcasting
DQPSK	:	Differential Quadrature Phase Shift Keying
DRM	:	Digitale Radio Mondiale
DVB, DVB-T	:	Digital Video Broadcasting
EBU	:	European Broadcasting Union
ERC	:	European Radiocommunications Committee
ERP	:	Effective Radiated Power
FFT	:	Fast Fourier Transform
FIC	:	Fast Information Channel
ISDB-T	:	Integrated Services Digital Broadcasting-Terrestrial
ITU	:	International Telecommunication Union
LNM	:	Log-Normal Method
QAM	:	Quadrature Amplitude Modulation
QPSK	:	Quadrature Phase Shift Keying
UHF	:	Ultra High Frequency
VHF	:	Very High Frequency

The following table contains the explanations of the most important mathematical parameters that are used throughout this book:

A_{km}	:	antenna diagram factor of the k-th transmitter for the m-th direction
E_l	:	number of inhabitants corresponding to pixel l
F_{min}	:	minimum field strength
F_{total}	:	total field strength
F_{wanted}	:	wanted field strength
F_{unwanted}	:	unwanted field strength
F_P	:	protection ratio
G_d	:	antenna gain
K_k	:	costs of k-th transmitter
P_k	:	ERP of k-th transmitter
Φ_k	:	direction of main lobe of k-th antenna diagram
R	:	re-use distance
T_G	:	guard interval
T_S	:	symbol duration of a T-DAB or DVB-T symbol
T_W	:	duration of FFT evaluation window of the receiver
Q	:	quality function
c	:	velocity of light
κ_k	:	width of k-th antenna diagram
$\mathbf{r}_k$	:	location of k-th transmitter site
τ_k	:	time delay of k-th transmitter

Bibliography

[Abl87] Ablay, P., *Optimieren mit Evolutionsstrategien,* Spektrum der Wissenschaft, **7** (1987) 104

[Bal82] Balanis, C. A., *Antenna Theory - Analysis and Design,* John Wiley & Sons, New York, 1982

[Beu95] Beutler, R., *Optimization of Digital Single Frequency Networks,* Frequenz **49** (1995) 245

[Beu98a] Beutler, R., *A-Priori- und A-Posteriori-Dibitfehlerwahrscheinlichkeit DQPSK-modulierter Übertragung über Fading-Kanäle,* Forschungsverbund Medientechnik Südwest, Tagungsband des Projekttages “Messung und Modellierung von Funkkanälen”, Beutler, R., Prosch, Th. (Hrg.), Südwestrundfunk, Stuttgart, 1998

[Beu98b] Beutler, R., *Digital Single Frequency Networks : Improving Optimization Strategies by Parallel Computing,* Frequenz **52** (1998) 90

[Beu00] Beutler, R., *Channel Assignment for DVB-T - Spectrum Demand Estimation,* European Broadcasting Union, B/PPM Doc. 035, Geneva, 2000

[Beu01a] Beutler, R., *Abschätzung des Spektrumsbedarfs für DAB,* telekom praxis **78** (2001) 26

[Beu02] Beutler, R., *T-DAB and DVB-T in VHF - A Unified and Flexible Planning Approach,* Proceedings of the Open Symposium on Propagation and Remote Sensing, URSI Commission F, Garmisch-Partenkirchen, 2002

[Bre79] Brélaz, D., *New Methods to Color the Vertices of a Graph,* Communications of the ACM **22** (1979) 251

[Bro85] Bronstein, I.N., Semendjajew, K.A., *Taschenbuch der Mathematik,* Verlag Harri Deutsch, Frankfurt (Main) 1985

[Bru92] Brugger, R., *Die exakte Bestimmung der nutzbaren Feldstärke bei der Behandlung von Mehrfachstörungen und ein Vergleich älterer und neuer Näherungsmethoden,* Rundfunktechnische Mitteilungen **38** (1992) 23

[Car90] Carraghan, R., Pardalos, P.M., *An Exact Algorithm for the Maximum Clique Problem,* Operations Research Letters **9** (1990) 375

[Cau82] Causebrook, J. H., et.al., *Computer Prediction of Field Strength - A Manual of Methods developed by the BBC for the LF, MF, VHF and UHF Bands,* Engineering Division, BBC, 1982

[CEP95] CEPT, *Final Acts - CEPT T-DAB Planning Meeting Wiesbaden 1995,* European Conference of Postal and Telecommunications Administrations, Wiesbaden, Germany, 1995

[CEP96] CEPT, *Final Acts of the T-DAB Planning Meeing (2) Bonn, 7-8 November 1996,* European Conference of Postal and Telecommunications Administrations, Bonn, Germany, 1996

[CEP97] CEPT, *The Chester 1997 Multilateral Coordination Agreement relating to Technical Criteria, Coordination Principles and Procedures for the Introduction of Terrestrial Digital Video Broadcasting (DVB-T),* European Conference of Postal and Telecommunications Administrations, Chester, United Kingdom, 1997

[CEP02] CEPT, *Final Acts of the T-DAB Planning Meeing (2) Maastricht 2002,* European Conference of Postal and Telecommunications Administrations, Maastricht, The Netherlands, 2002

[Com88] Commission of the European Communities, *COST 207. Digital Land Mobile Radio Communication - Final Report,* Brussels, Belgium, 1988

[Due93] Dueck, G., *The Great Deluge Algorithm and the Record-to-Record Travel,* J. Comp. Physics. **104** (1993) 86

[EBU98] EBU, *Technical Bases for T-DAB Services Network Planning and Compatibility with Existing Broadcasting Services,* European Broadcasting Union, BPN Doc. 003-Rev.1, Geneva, Switzerland, 1998

[EBU01] EBU, *Terrestrial Digital Television Planning and Implementation Considerations - Edition 3,* European Broadcasting Union, BPN Doc. 005, Geneva, 2001

[EBU02] EBU, *Planning Criteria for Mobile DVB-T,* European Broadcasting Union, BPN Doc. 047, Geneva, Switzerland, 2002

[EBU03] EBU, *Impact on Coverage of Inter-Symbol Interference and FFT Window Positioning in OFDM Receivers,* European Broadcasting Union, BPN Doc. 059, Geneva, Switzerland, 2003

[ERC98] European Radiocommunications Committee, *ERC/EBU Report on Planning and Introduction of terrestrial Digital Television (DVB-T) in Europe,* Izmir, Turkey, 1997

[ERC01] European Radiocommunications Committee, *T-DAB Questionnaire for 3rd Priority in the 1.5 GHz Band,* Interlaken, Switzerland 2001

[ETS97a] European Telecommunications Standards Institute ETS 300 401, *Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers,* 1997

[ETS97b] European Telecommunications Standards Institute ETS 300 744, *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television (DVB-T)*, 1997

[EUR96] Eureka 147 Projekt, *Digital Audio Broadcasting,* 1996 [available from von Dr. Thomas Lauterbach c/o Robert Bosch GmbH, Abt. FV/SLM, Robert-Bosch-Straße 200, 31139 Hildesheim]

[Gra02] Gräf, A., McKenney, T., *Ensemble planning for digital audio broadcasting,* in "Handbook of Wireless Networks and Mobile

Computing”, Stojmenovic, J. (Hrg.), John Wiley and Sons, New York, USA, 2002

[Gro86] Großkopf, R., *Feldstärkevorhersgae im VHF-Bereich mit Hilfe topographischer Daten,* Rundfunktechnische Mitteilungen **30** (1986) 176

[Gro95] Großkopf, R., *Vergleich verschiedener Feldstärkeprognoseverfahren mit Messungen im DAB-Gleichwellennetz in Bayern,* Rundfunktechnische Mitteilungen **39** (1995) 102

[Hal96] Hall, M. P. M., Barcley, L.W., Hewitt, M. T. (Hrgs.), *Propagation of Radio Waves,* The Institution of Electrical Engineers, London, UK, 1996

[Hei94] Heistermann, J., *Genetische Algorithmen,* Teubner Texte zur Informatik 9, B.G. Teubner Verlagsgesellschaft, Stuttgart, Germany, 1994

[Hes98] Hess, G. C., *Handbook of Land-Mobile Radio System Coverage,* Artech House Books, London, UK, 1998

[IDR00] Bundesministerium für Wirtschaft und Technologie, Initiative “Digitaler Rundfunk”, *Einführung des digitalen Rundfunks in Deutschland - Startszenario 2000,* Berlin, Germany, 2000

[IRT82] Institut für Rundfunktechnik, *Richtlinie für die Beurteilung der UKW-Hörfunkversorgung (Mono und Stereo) bei ARD und DBP,* München, Germany, 1982

[IRT99] Institut für Rundfunktechnik, *Frequenzbedarf für den terrestrischen digitalen Rundfunk,* München, Germany, 1999

[IRT02] Institut für Rundfunktechnik, *Measurements of Protection Ratios for Various DVB-T Signals,* München, Germany, 2002

[ITU61] ITU, *Final Acts of the European VHF/UHF Broadcasting Conference,* Stockholm, Sweden, 1961

[ITU95] ITU-R Recommendation P. 370-7 (10/95), *VHF and UHF Propagation Curves for the Frequency Range from 30 MHz to 1000 MHz, Broadcasting Services,* Geneva, Switzerland 1995

[ITU00] ITU-R Recommendation BT. 1368-2, *Planning Criteria for Digital Terrestrial Television Services in the VHF/UHF Bands,* Geneva, Switzerland, 2000

[ITU01a] ITU-R Recommendation P. 1546, *Method for point-to-area predictions for terrestrial services in the frequency range 30 to 3 000 MHz,* Geneva, Switzerland, 2001

[ITU01b] ITU-R Recommendation BS. 1514, *System for Digital Sound Broadcasting in the Broadcasting Bands Below 30 MHz,* Geneva, Switzerland, 2001

[ITU02] ITU-R Recommendation BT. 1368-3, *Planning Criteria for Digital Terrestrial Television Services in the VHF/UHF Bands,* Geneva, Switzerland, 2002

[ISO93] International Organisation for Standardization, *Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio,* 1993

[Jun98] Jungnickel, D., *Graphs, Networks and Algorithms,* Springer Verlag, Berlin, Germany, 1998

[Kam92] Kammeyer, K. D., *Nachrichtenübertragung,* Teubner-Verlag, Stuttgart, Germany, 1992

[Kat89] Kathrein Werke KG, *FM- und TV-Sendesysteme,* Rosenheim, Germany, 1989

[Kle96] Kleiman, S., Shah, D., Smaalders, B., *Programming with Threads,* SunSoft Press, Mountain View, California, USA, 1996

[Kue98] Küchen, F., Haaß, U., Wiesbeck, W., *Vergleichende Beurteilung der Versorgungskriterien für DAB und DVB-T,* Rundfunktechnische Mitteilungen, Institut für Rundfunktechnik GmbH **42** (1998)

[KZZ02] Konrad-Zuse-Zentrum für Informationstechnik Berlin, *FAP web - A website about Frequency Assignment Problems,* http://fap.zib.de

[Lau96] Lauterbach, T., *Digital Audio Broadcasting - Grundlagen, Anwendungen und Einführung von DAB,* Franzis-Verlag, Feldkirchen, Germany, 1996

[Leb92] Lebherz, M., Wiesbeck, W., Krank, W., *A Versatile Wave Propagation Model for the VHF/UHF Range Considering Three Dimensional Terrain,* IEEE Trans. Antennas and Propagation **40** (1992) 1121

[Lew96] Lewis, B., Berg, D. J., *Threads Primer - A Guide to Multithread Programming,* SunSoft Press, Mountain View, California, USA, 1996

[Lig99] Ligeti, A., Zander, J., *Minimal Cost Coverage Planning for Single Frequency Networks,* http://citeseer.nj.nec.com/ligeti99minimal.html, 1999

[Lon68] Longley, A. G., Rice, P. L., *Prediction of Tropospheric Radio Transmission Loss over Irregular Terrain, a Computer Method,* ESSA Tech. Report ERL-79-ITS-67, Institute of Telecommunications Sciences, Boulder, USA, 1968

[Lue84] Luebbers, R. J., *Finite Conductivity Uniform GTD versus Knife Edge Diffraction Loss in Prediction of Propagation Path Loss,* IEEE Trans. Antennas and Propagation **32** (1984) 70

[Mat83] Matula, D. W., Beck, L. L., *Smallest-Last Ordering and Clustering and Graph Coloring Algorithms,* J. Association of the Computing Machinery **30** (1983) 417

[Mee83] Meeks, M. L., *VHF Propagation over Hilly Forested Terrain,* IEEE Trans. Antennas and Propagation **14** (1983) 480

[Oku68] Okumura, Y., Ohmori, E., Kawano, T., Fukuda, K., *Field Strenght and its Variability in VHF and UHF Land-Mobile Radio Service,* Rev. of the Electr. Comm. Labaratory **16** (1968) 825

[OLe98] O'Leary, T., *Wiesbaden '95 Revisited - T-DAB Planning Parameters, Reference Networks and Frequency Planning Algorithms,* EBU Technical Review, **278** (1998), Geneva, Switzerland

[Phi95] Philipp, J., *Einfache und genaue Approximation der Summenfeldstärke log-normal verteilter Einzelfelder durch sukzessive Verknüpfung der Komponenten,* Kleinheubacher Berichte **38** (1995) 195

[Pro89] Proakis, J. G., *Digital Communications,* McGraw-Hill, New York, USA, 1989

[Que93] Quellmalz, A., *Graphenorientierte Planung von Sendernetzen,* Nomos Verlagsgesellschaft, Baden-Baden, Germany, 1993

[Rad01] Radio Authority, *Adjacent-Frequency Working in Overlapping T-DAB Coverage Areas in the United Kingdom,* United Kingdom, London, 2001

[Rei01] Reimers, U., *Digital Video Broadcasting - The International Standard for Digital Television,* Springer-Verlag, Berlin, Germany, 2001

[She95] Shelswell, P., *The COFDM Modulation System : The Heart of Digital Audio Broadcasting,* Electronic & Communication Engineering Journal, (1995) 127

[Sto98] Stott, J., *The how and why of COFDM,* EBU Technical Review, **278** (1998), Geneva, Switzerland

[Wor00] WorldDAB, *Can We Satisfy the Future DAB Spectrum Demand,* SLC Report on DAB Spectrum Demand, London, UK, 2000

[Wei98] Weißenfels, G., *DAB-Blockzuweisung - Lösungsansätze,* Messung und Modellierung von Funkanälen, Projekttag des Forschungsverbundes Medientechnik Südwest, Beutler, R., Prosch. Th. (Hrgs.), Stuttgart, Germany, 1998

[ZIB01] Konrad-Zuse-Zentrum für Informationstechnik, *FAP web - A website about Frequency Assignment Problems,* Berlin, http://fap.zib.de, 2001

The list of references given here has to be considered as a subjective subset of those documents that might be relevant. It is by no means exaustive. Rather, it reflects the author's knowledge and usage of literature. It might give first indications and hints for further reading. Additional references can be found with the help of any Internet search engine without problems.

Some of the references given point to documents that are not officially available. This concerns for example documents submitted to EBU project teams. However, in most cases these documents are not classified and can be obtained by addressing the EBU or the authors directly.

List of Figures

List of Tables

Index