UNIVERSITY OF SOUTHAMPTON
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

Minimum Bit Error Ratio Beamforming

by

Nurul Nadia Ahmad
BEng.

A doctoral thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy at the University of Southampton

March 2005

SUPERVISORS:
Professor Lajos Hanzo
Dipl. Ing., MSc., PhD, Fellow IEEE
Chair of Telecommunications
School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, United Kingdom

Dr. Sheng Chen
BEng., MSc., PhD, SMIEEE
Reader
School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, United Kingdom

To my family in Malaysia.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

Minimum Bit Error Ratio Beamforming

by Nurul Nadia Ahmad

The main objective of employing smart antennas is that of combating the effects of multipath fading on the desired signal and suppressing interfering signals, thereby potentially increasing both the bit error ratio performance and the capacity of wireless systems. Specifically, in smart antenna aided systems multiple antennas may be invoked at the transmitter and/or the receiver, where the antennas may be arranged for achieving spatial diversity, directional beamforming or for attaining both. In smart antenna assisted systems the achievable performance improvements are usually a function of the antenna spacing and that of the algorithms invoked for processing the signals received by the antenna elements. This treatise is focused on adaptive beamforming employing the temporal reference technique (TRT). We explore a range of efficient adaptive algorithms designed for the sake of optimising the achievable beamforming performance.

Based on the traditional Minimum Mean Square Error (MMSE) criterion we commence our study by investigating three classic algorithms, namely the Recursive Least Square (RLS), the Direct Matrix Inversion (DMI) and the Recursive Sample Matrix Inversion (RSMI) algorithms in order to determine the optimal array weight values. We evaluate and compare their performance in terms of the attainable Signal-to-Interference Ratio (SIR), as a function of the received signal power and that of the reference sequence length used for Binary Phase Shift Keying (BPSK) transmissions over an Additive White Gaussian Noise (AWGN) channel. A three-element uniform linear array was then employed to observe the array's angular response at the defined angles, as the number of interfering signals is increased. The RLS algorithm was used for conducting these simulations over both AWGN and flat Rayleigh fading channels. For the adaptive algorithms discussed, we evaluate the associated complexity as a function of the number of antenna elements invoked.

Our further investigations were motivated by the rationale that the ultimate performance measure is the achievable bit error ratio. Therefore, instead of the MMSE criterion we continue our investigations by invoking the novel approach of Minimum Bit Error Ratio (MBER) beamforming, which is based on directly minimising the system's Bit Error Ratio (BER). We employed a simplified conjugate gradient algorithm for determining the array weights of this MBER beamforming solution. Several adaptive versions of the MBER algorithm were presented, which were categorised into two classes, namely the family of block-data based and the set of sample-by-sample adaptive stochastic gradient based algorithms.

In the block-data based adaptive algorithm category we studied the Block Adaptive Conjugate Gradient (BACG) algorithm, while in the stochastic gradient category we investigated both the Least Bit Error Rate (LBER) and the Approximate LBER (ALBER) algorithms. The ALBER algorithm consistently outperformed the LBER algorithm, despite having as low a complexity as the well-known LMS algorithm, provided that the related algorithmic parameters were appropriately chosen.

To circumvent the drawbacks of the gradient based adaptive MBER algorithms, namely those of the BACG, LBER and ALBER algorithms we invoked Genetic Algorithms (GAs) in conjunction with the MBER beamforming scheme. The convergence behaviour of the GA was studied by evaluating the probability density function (PDF) of the BER at the beamformer's output. It was shown that GA is capable of circumventing many of the problems encountered by the MBER beamforming scheme.

Acknowledgements

I would like to thank all of those people who helped me to complete this thesis.

First, I wish to thank both my supervisors, Prof. Lajos Hanzo and Dr. Sheng Chen for their willingness to share with me their wealth of intuition, patience and support as they guided me in this work. Prof. Hanzo has been an invaluable guide throughout this entire process. He has dedicated his time to helping me with every aspect of this work. I am truly grateful for his kindness and his generous financial assistance. Dr. Chen's enthusiasm and determination has always inspired me. Advice and words of encouragement from both of them kept me motivated. My gratitude for their assistance is simply immeasurable. Many thanks also to Dr. Lie-Liang Yang and to Dr. Tim O'Farrell for their suggestions on the improvement of the report.

I would also like to thank the Multimedia University for approving my study leave in pursuing this degree and Majlis Amanah Rakyat (MARA) for their financial support.

Many thanks to Dr. Soon Xin Ng and Dr. Bee Leong Yap for their assistance with the computer system as well as with their help in programming and in using the Linux operating system. My thanks also go to Dr. Byoung Jo Choi for his Latex tutorial, to Dr. Jin Yee Chung, Dr. Hafizal Mohamad and Noor Shamsiah Othman for support, helps and discussions in C programming, GLE and Latex, to Dr. Matthias Münster for his assistance in understanding the beamforming concept, to Dr. Jason Wong Pee Ng for introducing me to Genetic Algorithms (GAs), to Chandy Vitale and Gianluca Vesentini for their help and insightful discussions on GAs while doing their project in our group. I would like to continue by thanking Dr. Ahmad Kamsani Samingan for his tutorial, to Dr. Mohamad Yusoff Alias and Andreas Wolfgang for their generous help and discussions on several related topic. My thanks are also due to everyone in the Communications Research Group, both in the past and presently, thank you for the friendship, knowledge, discussions, advice and help throughout this project. A special thank you goes to Denise Harvey and Rebecca Earl for their help with administrative matters.

Outside the research group, I have also cherished the company of several friends throughout my stay at Southampton. Thanks also due to every single friend, both in the UK and Malaysia - too numerous to be mentioned here, for their wonderful friendship and for their support over the years.

To my family in Malaysia, thank you for their love, continual support and prayer. Especially to my parents, Ahmad Abdollah and Aminah Abdullah, I cannot thank them enough for the sacrifices they made in raising me. I am very grateful to have them as my parents.

Finally, I extend my greatest gratitude to Allah for giving me the opportunity to meet the wonderful people mentioned above and for the strength to complete my studies.

List of Publications

1. **S. Chen, L. Hanzo, N. N. Ahmad and A. Wolfgang**, "Adaptive Minimum Bit Error Rate Beamforming Assisted Receiver for QPSK Wireless Communications", *accepted in* Digital Signal Processing, Elsevier, 2005.

2. **S. Chen, N. N. Ahmad and L. Hanzo**, "Adaptive Minimum Bit Error Rate Beamforming", *to appear in* IEEE Transactions on Wireless Communications, March 2005.

3. **S. Chen, L. Hanzo, N. N. Ahmad and A. Wolfgang**, "Adaptive Minimum Bit Error Rate Beamforming Assisted QPSK Receiver", International Conference on Communications (ICC) 2004 - Wireless Communications Symposium, Paris, France, June 20-24, 2004, Vol. 6, pp. 3389-3393.

4. **A. Wolfgang, N. N. Ahmad, S. Chen and L. Hanzo**, "Genetic Algorithm Assisted Minimum Bit Error Rate Beamforming", IEEE Vehicular Technology Conference (VTC) 2004 Spring, Milan, Italy, May 17-19, 2004, Vol. 1, pp. 142-146.

5. **A. Wolfgang, N. N. Ahmad, S. Chen and L. Hanzo**, "Genetic Algorithm Assisted Error Probability Optimisation for Beamforming", IEE Electronics Letters, 4th of March, 2004, Vol. 40, No. 5, pp. 320-322.

6. **S. Chen, L. Hanzo and N. N. Ahmad**, "Adaptive Minimum Bit Error Rate Beamforming Assisted Receiver for Wireless Communications", in Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2003, Hong Kong, China, April 6-10, 2003, Vol. IV, pp. 640-643.

7. **S. Chen, N. N. Ahmad and L. Hanzo**, "Smart Beamforming for Wireless Communications: A Novel Minimum Bit Error Rate Approach", in Proceedings of 2nd. IMA International Conference on Mathematics in Communications, Lancaster, UK, December 16-18, 2002, 4 pages.

8. **M. Vitale, G. Vesentini, N. N. Ahmad and L. Hanzo**, "Genetic Algorithm Assisted Adaptive Beamforming", in Proceedings of VTC 2002 Fall, Vancouver, Canada, 2002, Vol. 1, pp. 601-605.

Contents

Chapter 1

Introduction

We commence our discourse by outlining the organisation and the novel contributions of the thesis in Section 1.1. This will be followed by a brief overview of the various sources of impairments in wireless communications and by an introduction to smart antennas as a means of mitigating these impairments in Sections 1.2.1 and 1.2.2, respectively. Finally, a summary of the chapter is provided in Section 1.3.

1.1 Organisation of the Thesis

The thesis is organised as follows:

Chapter 2: An overview of adaptive beamforming schemes will be presented, beginning with a brief historical perspective on its evolution. A rudimentary classification of various adaptive beamforming schemes will also be provided. We will mainly concentrate on adaptive beamforming techniques based on the temporal reference technique (TRT). Several classic algorithms, such as the Least Mean Square (LMS) algorithm, the Recursive Least Square (RLS) algorithm and the Sample Matrix Inversion (SMI) algorithm, which were all based on the Minimum Mean Square Error (MMSE) criterion will be outlined and compared in terms of their operating principle and computational complexity. We will also characterise both the RLS and SMI algorithms' performance in terms of the achievable Signal-to-Interference Ratio (SIR) as a function of the Signal-to-Noise Ratio (SNR) and also as a function of the reference sequence length used, when the users' signal power is fixed.

Chapter 3: This chapter introduces the novel concept of Minimum Bit Error Ratio (MBER) criterion based adaptive beamforming. We will use the simplified conjugate gradient method for directly minimising the Bit Error Ratio (BER). We will demonstrate the superiority of the MBER solution in comparison to the array weight solution offered by the closed-form MMSE equation in almost all scenarios investigated. The robustness of the MBER solution against array-overloading, i.e. against a scenario when the number of users is higher than the number of array elements will also be shown. In the context of adaptive MBER beamforming, we will characterise the performance of

several adaptive MBER beamforming algorithms, namely that of the block-data based Block Adaptive Conjugate Gradient (BACG) algorithm as well as that of two stochastic gradient based algorithms, namely the Least Bit Error Rate (LBER) and the approximate LBER (ALBER) algorithms.

Chapter 4: This chapter aims to overcome the limitations of the simplified conjugate gradient method used in Chapter 3 by employing a random guided search method based on Genetic Algorithms (GAs). We will demonstrate with the aid of BER histograms that the GA is capable of arriving at the exact MBER solution. The required complexity will be quantified in terms of the number of objective function evaluations, as the number of the antenna array elements L increases. We will also compare the BER performance attained by using the GA to that of the block-data based BACG algorithm.

Chapter 5: Finally, in this chapter we will summarise the findings of the thesis and offer our conclusions. Suggestions for future research will also be provided.

1.2 Motivation of the Thesis

1.2.1 Sources of Impairments in Wireless Communications

A radio signal transmitted in cellular systems is attenuated by three main factors, namely path-loss, fast- and slow fading [1–4]. Path-loss describes how the attenuation varies with distance between the transmitter and the receiver. Fast fading, which is sometimes referred to as short-term fading, is caused by the multipath effects and scattering, where the superposition of multiple paths generates a rapidly fluctuating fading envelope. If there is a dominant path or line of sight (LOS) path between the transmitter and the receiver, the fading channel is referred to as a Rician channel, while if there is no dominant LOS path, it is referred to as a Rayleigh channel. On the other hand, slow fading is caused by shadowing of the radio signal, owing mainly to building and terrain features, such as mountains and hills. It is also referred to as long-term fading. Fading however may also boost a signal instead of attenuating it, since it involves the combination of energy arriving from several signal paths. In general, the main scattering sources that affect the propagation environment are scatterers local to the mobile, scatterers local to the base station and remote scatterers, as briefly portrayed in Figure 1.1. A commonly used technique of combating signal fading in mobile radio system is constituted by transmit and receive diversity assisted by various diversity combining methods [3,5], which exploit the fact that the signals arriving from different locations fade at different rates.

Co-channel interference (CCI) is one of the primary channel impairments associated with digital radio communications. In radio applications, where the signals are assigned to different frequency bands, appropriate filter masks are imposed on the transmitted spectrum for the sake of limiting the adjacent channel interference. This interference is generated by transmitters assigned to adjacent

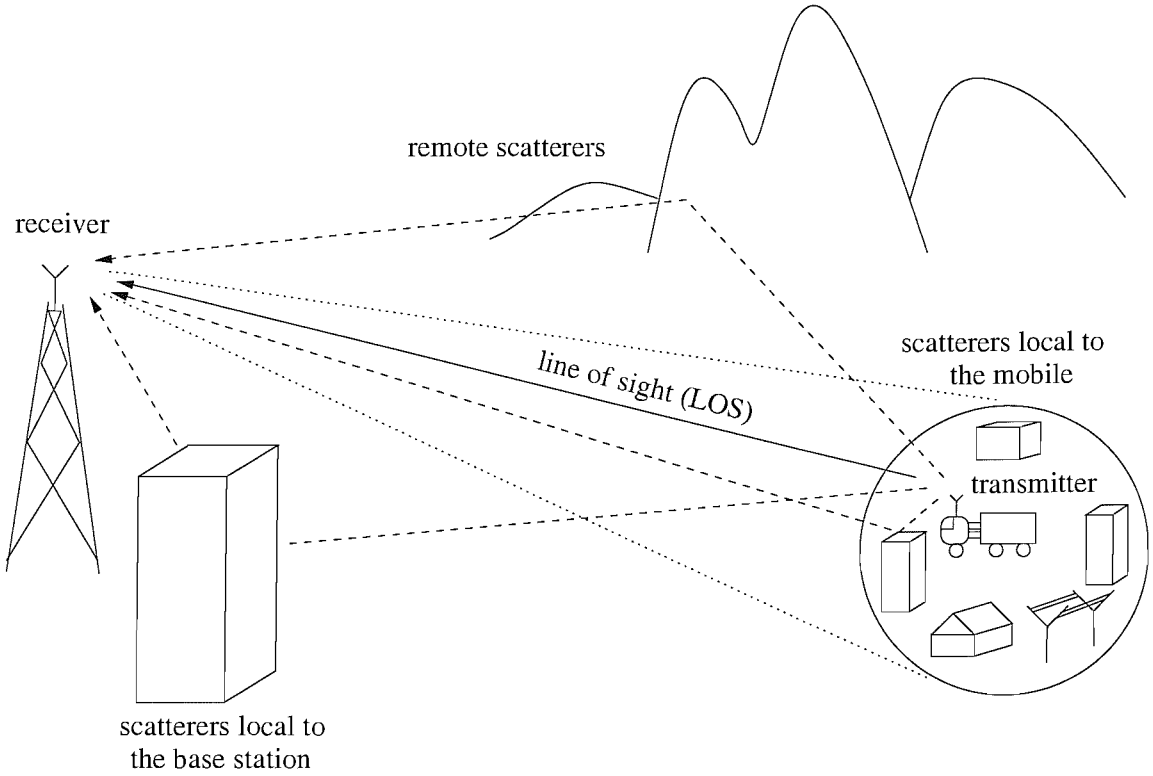


Figure 1.1: Main scattering sources that affect the propagation environment resulting in a 'multipath' propagation scenario.

frequency bands. In addition to this adjacent channel interference, co-channel interference (CCI) may be generated by other transmitters assigned to the same frequency band as the desired signal. The amount of CCI caused is dependent on numerous parameters, such as the geographical position of the users, the associated channel fading and the transmission power [4]. Since the primary objective of the mobile radio system is to conserve the available spectrum by reusing allocated frequency channels in areas that are geographically located as close to each other as possible, CCI imposes a limitation on the frequency reuse distance. On the other hand, the adjacent channel interference limits the spectral separation between adjacent channels and the assignment of frequency channels within a particular area.

Linear distortions imposed by dispersive multipath propagation result in intersymbol interference (ISI), which may inflict high error rates. In mobile cellular communications the multipath propagation of the transmitted signals may result in severe ISI. The solution to the ISI problem is to invoke a channel equaliser [2, 3, 6, 7].

1.2.2 Smart Antennas

The increasing demand for mobile communication services supported in a limited radio-frequency (RF) bandwidth motivates the design of novel communications techniques. There are numerous ways of

increasing the achievable capacity of cellular radio systems. Acquiring new frequency spectra is perhaps the easiest but not the most cost effective approach. Classic techniques that allow multiple users to share the available bandwidth include techniques, such as frequency, time, polarisation, code and spatial division multiple access [3]. A particular approach, which is receiving increasing attention and shows real promise of substantial capacity enhancement that would lead to a better quality of service (QoS) is constituted by spatial processing employing adaptive antenna arrays [4, 5, 8–11].

1.2.2.1 Advantages of Smart Antennas

Multiple antennas are capable of combating the effects of multipath fading with the aid of diversity combining techniques. For example, selection diversity selects the signal received from that particular antenna for processing, which has the highest probability of error-free detection. Equal gain combining (EGC) adjusts the phase of the desired signals and phase-coherently combines them after equal weighting. By contrast, maximal ratio combining (MRC) weights the contributions of the individual antennas in proportion to their Signal-to-Noise Ratio (SNR) and coherently combines the weighted signals in-phase. The employment of these classic techniques is common in communications systems [2–5, 7, 8, 12]. Apart from fading, the capacity of wireless communication systems, is however, gravely affected by another major impairment known as co-channel interference (CCI). Especially in Code Division Multiple Access (CDMA) systems, interference is imposed not only by the adjacent cells, but also by users communicating within the cell considered. An efficient way of overcoming this limitation is constituted by the employment of beamforming based smart antennas, provided that the desired and interfering signals originate from different spatial locations associated with different angles of arrival. By exploiting the spatial selectivity of an antenna array, CCI experienced in the uplink (i.e. mobile to base station link) and in the downlink (i.e. base station to mobile link) of wireless systems may be cancelled or mitigated, depending on the propagation environment. The reduced CCI can be traded for an increased system capacity. Apart from the increased system capacity, other operational advantages offered by deploying smart antennas in a mobile communication network include achieving an extended coverage, more efficient power control, reduction of the required transmit power, the support of smart handovers *etc.* [13]. These benefits are summarised in Figure 1.2. For detailed discussions on these and on several other advantages, readers are referred to [5, 8–10, 13, 14].

These benefits are basically due to the ability of smart antennas to form focused transmit and/or receive beams in certain directions and nulls in others. It is worth noting that the beam's limited angle of arrival results in a reduction of the number of multipath components, which leads to a reduced delay spread. Observed that this reduced delay spread may in fact become an impediment in conjunction with MRC, for example, since the reduced number of multipath components reduces the achievable receiver diversity gain, even if the MRC scheme has the ability to combine a high number of received signal components [15]. Nonetheless, beamforming schemes are capable of reducing the CCI,

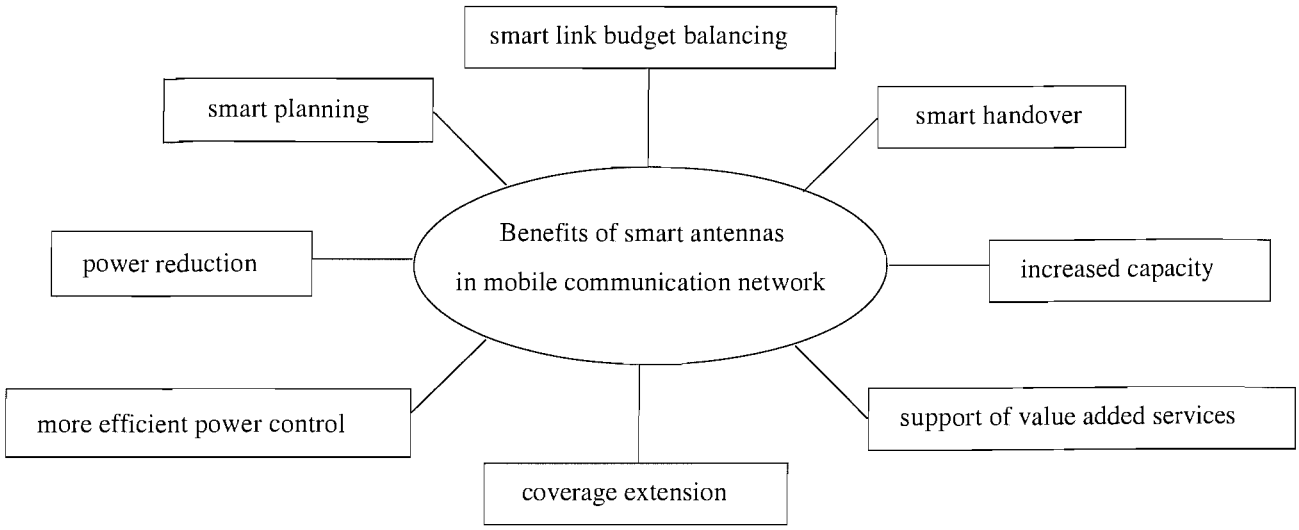


Figure 1.2: Operational benefits achieved with the aid of smart antennas [13].

which then may contribute towards the reduction in outage probability or decreases the Bit Error Ratio (BER) at a given SNR or vice versa, thus improving the achievable spectral efficiency. Smart antennas are also capable of achieving transmit diversity gain [16], which reduces fading of the radio signal and therefore ease the power control requirements. The required transmission power levels may also be reduced, which then relaxes the requirements on the batteries. The knowledge of the location and speed of a mobile user assists in smoothing the handover process. It also contributes to several value added services, as pointed out for example by Tsoulous in [13], some of which are listed as follows:

1. **On-demand location specific services:** Such services may include roadside assistance, real-time traffic updates, tourist information and electronic yellow pages, complemented by nearby petrol station, local entertainment and dining information.

2. **Ability to support user location for emergency calls:** Adaptive antennas are also capable of providing user location information in conjunction with direction finding in case of emergencies, such as accidents or abduction.

3. **Location of fraud perpetrators:** Since adaptive antennas are capable of providing user location information, they may be used for locating fraud perpetrators.

4. **Location sensitive billing:** Instead of the currently predominantly used billing principle of charging as a function of the time of day, the inclusion of location as the second measure may provide an operator with the ability to control its network by encouraging (or discouraging) any type of service access. The objective here is not so much that of simply reducing the cost of the calls, which would increase the network traffic, while producing the same revenue, but that of

Beamforming [4]	Typically $\lambda/2$-spaced antenna elements are used for the sake of creating a spatially selective transmitter/receiver beam. Smart antennas using beamforming have been employed for mitigating the effects of co-channel interfering signals and for providing beamforming gain.
Spatial Diversity [16] and Space-Time Spreading	In contrast to the $\lambda/2$-spaced phased array elements, in spatial diversity schemes, such as space-time block or trellis codes [16] the multiple antennas are positioned as far apart as possible, so that the transmitted signals of the different antennas experience independent fading, resulting in the maximum achievable diversity gain.
Space Division Multiple Access [12]	SDMA exploits the unique, user-specific 'spatial signature' of the individual users for differentiating amongst them. This allows the system to support multiple users within the same frequency band and/or time slot.
Multiple Input Multiple Output Systems [17]	MIMO systems also employ multiple antennas, but in contrast to SDMA arrangements, not for the sake of supporting multiple users. Instead, they aim for increasing the throughput of a wireless system in terms of the number of bits per symbol that can be transmitted by a given user in a given bandwidth at a given integrity.

Table 1.1: Applications of smart antennas in wireless communications [12].

improving the capacity of the system in the most efficient manner. This technique may enable tariff plans to be tailored to individual user needs, for example by providing low-cost zones when the user is in his preferred zone.

1.2.2.2 Classification of Smart Antennas [15]

In smart antenna assisted systems multiple antennas may be invoked at the transmitter and/or the receiver, where the antennas may be arranged for achieving spatial diversity, directional beamforming or for attaining both diversity and beamforming. In smart antenna systems the achievable performance improvements are usually a function of the antenna spacing and that of the algorithms invoked for processing the signals received by the antenna elements [15].

Terms commonly used today that embrace various aspects of smart antenna systems include intelligent antennas [18], adaptive antennas [5], [18], phased arrays [12, 15], Spatial Division Multiple Access (SDMA) [12,15,19,20], spatial processing, digital beamforming (DBF) [8,19,21] and others [18]. Smart antennas however, are mainly categorised based on their specific application in improving the wireless communications quality of service, as summarised in Table 1.1 [12,15].

In beamforming arrangements [4] typically $\lambda/2$-spaced antenna elements are used for the sake of cre-

ating a spatially selective transmitter/receiver beam. Smart antennas using beamforming have widely been employed for mitigating the effects of various interfering signals and for providing beamforming gain. Furthermore, the beamforming arrangement is capable of suppressing co-channel interference, which allows the system to support multiple users within the same bandwidth and/or same time-slot by separating them spatially. This spatial separation, however, becomes only feasible, if the corresponding users are separable in terms of the angle of arrival of their beams. An illustration of a system that supports multiple users within the same bandwidth is provided in Figure 1.3 a). Observed that all the co-channel users' main beam is in the direction of nulls of the central user. These beamforming schemes, which employ appropriately phased antenna array elements that are spaced at distances of $\lambda/2$ typically result in an improved Signal-to-Interference-plus-Noise Ratio (SINR) distribution and enhanced network capacity [4].

In contrast to the $\lambda/2$-spaced phased array elements, *in spatial diversity schemes* the multiple antennas are positioned as far apart as possible, using a typical spacing of 10λ [22], so that the transmitted signals of the different antennas experience independent fading, when they reach the receiver. This is because the maximum diversity gain can be achieved, when the signal replicas experience independent fading. Although spatial diversity can be achieved by employing multiple antennas at either the base station, mobile station, or both, it is more cost effective and practical to employ multiple antennas at the base station. A system having multiple receiver antennas has the potential of achieving receiver diversity, while that employing multiple transmit antennas exhibits transmit diversity. Recently, the family of transmit diversity schemes based on space-time coding [16], either space-time block codes or space-time trellis codes, has received wide attention and has been invoked in the third-generation systems [23]. The aim of using spatial diversity is to provide both transmit as well as receive diversity and hence enhance the system's integrity/robustness. This typically results in a better physical-layer performance and hence a better network-layer performance. Therefore, space-time codes are capable of indirectly increasing not only the transmission integrity, but also the achievable spectral efficiency [16].

A third application of smart antennas is often referred to as *Space Division Multiple Access* (SDMA) [12, 15], a technique which exploits the unique, user-specific 'spatial signature' of the individual users for differentiating amongst them. In simple conceptual terms one could argue that both a conventional CDMA spreading code and the Channel Impulse Response (CIR) affect the transmitted signal similarly - they are namely convolved with it. Hence, provided that the CIR is accurately estimated, it becomes known and certainly unique, although - as opposed to orthogonal Walsh-Hadamard spreading codes, for example - not orthogonal to the other CIRs. Nonetheless, it may be used for uniquely identifying users after channel estimation and hence for supporting several users within the same bandwidth. Provided that a powerful multiuser detector is available, one can support even more users than the number of antennas [15]. Hence this method enhances the achievable spectral efficiency directly. As seen in Figure 1.3 b), the spatially dispersed SDMA users have their own beam sector in

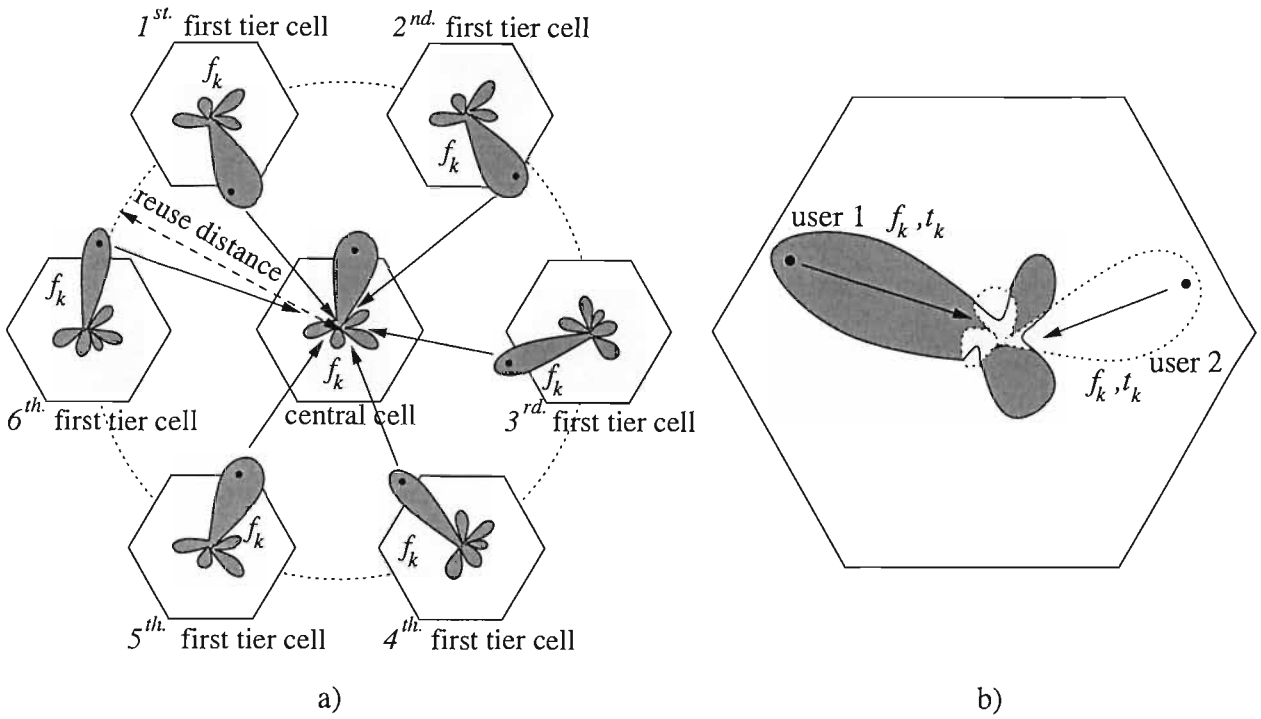


Figure 1.3: Concept of a) beamforming (Spatial Filtering for Interference Reduction (SFIR) [13]) and b) Spatial Division Multiple Access (SDMA) [13].

a shared cell. The shape of each beam sector is dynamically updated according to the specific propagation and interference environment encountered by a particular mobile user. The above-mentioned beamforming concept that aims for reducing the frequency reuse distance, while maintaining a low co-channel interference has also been referred to as Spatial Filtering for Interference Reduction (SFIR) by Tsoulous in [13], as justified by Figure 1.3 a). By contrast, the SDMA principle characterised in Figure 1.3 b) allows several users to communicate at the same time instant, t_k and frequency f_k within a given cell, again, as mentioned above [12], using the unique users-specific CIRs of the individual mobile users. This principle has the potential of substantially increasing the number of users supported.

Finally, Multiple Input Multiple Output (MIMO) systems [17] also employ multiple antennas, but in contrast to SDMA arrangements, not for the sake of supporting multiple users. Instead, they aim for increasing the throughput of a wireless system in terms of the number of bits per symbol that can be transmitted by a given user in a given bandwidth at a given integrity [15].

1.3 Summary

In this chapter, the organisation of the thesis was outlined. We then briefly characterised several sources of impairments encountered in wireless communications systems. The increasing number of communication services that have to be supported in a limited frequency spectrum requires efficient

techniques for the sake of exploiting the resources available. Smart antennas [4, 8–12] constitute an efficient technique of achieving these objectives. Their benefits were summarised in Figure 1.2, while their categorisation based on their specific application in improving the wireless communications quality of service (QoS) was summarised in Table 1.1. Spatial diversity and space-time spreading [24–33], SDMA [34–37] as well as MIMO schemes for achieving multiplexing gain [25, 38–47] are also timely research topics. In the next chapter we will discuss adaptive beamforming, i.e. the formation of angularly selective beams using multiple antennas separated by a distance of $\lambda/2$.

Chapter 2

Adaptive Beamforming

2.1 Introduction

Adaptive beamforming constitutes a process, which is capable of separating signals collocated in the frequency domain, but separated in the angular spatial domain [8]. It combines the signal received by a number of antenna elements with the aid of baseband signal processing for the sake of optimising its radiation or receiver beam pattern automatically in response to the interfering and desired signal environment. In other words, this smart antenna system has the capability of controlling the antenna beam patterns, in which the main lobe having an enhanced gain will be directed to the desired signal of a particular user and at the same time, a null or side lobes with minimal gain will be formed towards the interferers constituted by other nearby users.

In this chapter, we commence our discourse by providing a glimpse of the historical background of beamforming. We will then outline the basic concepts of beamforming in Section 2.2 and introduce the signal model that is used throughout this report. A detailed example is provided for the sake of illustrating the basic beamforming process. In Section 2.3 we will discuss and compare several design criteria that can be used in determining the optimal array weights. We will then focus our attention on the adaptive beamformer's operation portrayed in Figure 2.11 and in Section 2.4 we proceed with the portrayal of some of the commonly used adaptive array-weight update algorithms. Various issues of beamforming invoked in both the uplink (i.e. mobile to base station link) and the downlink (i.e. base station to mobile link) are highlighted in Section 2.5 and finally, some performance results are provided in Section 2.6.

2.1.1 Historic Background of Beamforming

Adaptive beamforming was conceived for employment in military sonar and radar systems [8]. It was primarily invoked for direction finding and for nulling enemy interferers. Generally, techniques

developed for counteracting intentional jamming have also proved suitable for communication through dispersive channels in cellular applications. Here, we highlight several milestones during the evolution of beamforming, commencing with the invention of the intermediate frequency (IF) sidelobe canceller (SLC) originally devised by Howells [48] in the late 1950s [6, 49]. This first bona-fide adaptive antenna system that incorporated the key capability of automatic interference nulling, however, was only partially adaptive. More specifically, the main beam antenna has a fixed pattern and the auxiliary array contains only a few controlled elements. Subsequent developments of this simple adaptive antenna lead to the emergence of the first fully adaptive array in 1965 [8]. In contrast to Howells' original SLC, a fully adaptive array exhibits a higher grade of flexibility in terms of its adaptive performance characteristics [49], albeit this is achieved at a significantly increased implementation cost.

In terms of publications, the first special issue on adaptive antennas [50] was published in March 1964, which was edited by Hansen. The collection of papers mainly concentrated on the family of retrodirective systems and self-steering or self-focusing arrays. These were essentially based upon phase-locked loop (PLL) and phase-conjugate schemes often invoked for employment in diversity combining [2,8]. The term 'adaptive antenna' was used for describing the self-phasing antenna system, which reradiates a signal in the direction from which it was received. This type of system is referred to as adaptive, since it operates without any prior knowledge of the direction in which it has to transmit. For clarity, such a system has been referred to as an adaptive transmitting array by Widrow *et al.* [51], while describing their proposed system as an adaptive receiving array in [51]. In Howells' special issue [50] however, there is no publication on adaptive interference nulling, although it had actually been under development for some years prior to that [52].

It was only in 1966, that Howells' coworker, Applebaum [53] published the first contribution concerning adaptive interference nulling, in which he derived the algorithm governing the operation of an adaptive array antenna, invoking a control loop for each element of the array. Applebaum [53] shows that the adaptive array processing technique is a generalisation of the SLC procedure of [48]. The algorithm, now commonly known as the Howells-Applebaum algorithm is based on the general problem of maximising the Signal-to-Noise Ratio (SNR) at the array output and thus it is sometimes referred to as the Maximum Signal-to-Noise ratio (MSN) algorithm [53, 54]. The noise term includes both the receiver noise and all other external noise components, which implies that essentially this algorithm is based on the maximum Signal-to-Interference-plus-Noise-Ratio (SINR). The MSN algorithm is typically used for signal detection by determining the array weight vector, which yields the maximum probability of error-free detection. Further seminal contributions are due to another independent research group lead by Widrow [51], who co-developed the well known Widrow-Hoff Least Mean Square (LMS) algorithm invoking the Minimum Mean Square Error (MMSE) criterion for automatic adjustment of the array weights. This algorithm overcomes the main difficulty encountered in applying the steepest descent algorithm for array weight adjustment, since the correlation matrix is not required for the weight calculation. The Widrow-Hoff LMS algorithm was first applied to adap-

tive RF antenna arrays in the open literature in 1967 [51]. Both the MSN and LMS techniques were invoked for iteratively cancelling the interferers in the presence of the signal of interest, both of which were analogue signals [55]. With the advent of digital signal processing, these techniques were then employed also to time- as well as amplitude-discretised signals. It was found later [54] that if an actual target signal is used in the Widrow-Hoff method [51], it leads to a biased solution, i.e. it does not converge to the MMSE filter. The proof of this is provided in the Appendix of [54]. Elimination of this problem has been proposed by Griffith in [54] for large antenna arrays, where all terms used by the algorithm are either known *a priori* or are extracted directly from the received signal.

In 1969, Capon [56] introduced a different beamforming technique, using a desired direction of arrival (DOA) constraint, which leads to an adaptive beamformer referred as the Minimum Variance Distortionless Response (MVDR) beamformer. He applied the theory of signal estimation in the context of seismometer arrays using a maximum likelihood criterion. His work was motivated by realising that the poor performance of the conventional (sometimes known as the delay-and-sum) beamformer is due to the fact that its response along the direction of interest depends not only on the power of the incoming target signal, but also on that of the undesirable contributions arriving from other sources of interference [6]. He then attempted to minimise the power contributed by both the noise and the interfering signals arriving from other directions, while maintaining a fixed gain in the direction of interest. Capon's technique is one of the earliest adaptive beamforming techniques that offers the ability to resolve signals that are separated by a fraction of an antenna beamwidth [8]. In the context of angle of arrival estimation, this algorithm is termed as the Maximum Likelihood Method (MLM) [57], since it also maximises the likelihood function of the input signal vector.

Despite the completely independent discovery and development of the MSN and LMS algorithms, they are fairly similar in that they are both based on the steepest descent gradient search technique. Furthermore, for stationary received signals, the algorithms converged to the optimum Wiener solution [8]. The popularity of the algorithms is due to their simplicity and robustness, as well as a consequence of requiring no data storage and their flexibility of implementation in either analogue or digital form. However, achieving a satisfactory performance can only be guaranteed under specific operational conditions. The LMS algorithm was then further developed following the introduction of various constraints by Frost [58] and Griffiths [59]. When these constraints are applied, the desired signal is never filtered out along with the unwanted signals. The optimisation process is still the same as before, but now the required antenna gain could be maintained in the desired direction. Frost's algorithm [58] has a unique advantage over other constrained optimisation algorithms in that the algorithm has a self-correcting feature, which makes it numerically stable despite the round-off and truncation errors of the array weights. The algorithm can be operated for an arbitrarily long time, i.e. large number of iterations, because error accumulation has been eliminated.

The MSN algorithm was further developed during the 1970s for example by Brennan and Reed [60–

62]. Their research team has made several significant contributions to the theory of adaptive radar systems, involving simultaneous adaptivity in both spatial and temporal domains [52]. In [61], Brennan, Pugh and Reed developed a theory of control-loop noise for adaptive arrays, which maximise the SNR. Their analysis shows that both the array weight noise and the array convergence rate are determined by the eigenvalues of the noise covariance matrix. In other words, the selection of optimum parameters requires striking a compromise between the achievable convergence rate and the array weight noise. This compromise-solution was improved later with the introduction of envelope limiting [60]. Generally, the Maximum Signal-to-Noise criterion constitutes an optimisation principle in the sense that the maximisation of the SNR is equivalent to the maximisation of the probability of error-free detection, provided that the noise and interference may be approximated by a Gaussian process. In 1974, Reed *et al.* [63] presented a technique referred to as the Sample Matrix Inversion (SMI) procedure, which exhibits a rapid convergence (that depends only on the number of adaptive array elements) owing to the direct computation of the adaptive weights. In addition to achieving a higher convergence rate, the performance of the SMI technique is almost independent of the eigenvalue spread encountered, which is a limitation in the context of the LMS and MSN algorithms. These advantages are achieved, however at the cost of an increased complexity. Further beamforming investigations were then focused on identifying appropriate search techniques combined with simple array performance monitoring [64], in an effort to contrive implementations, which are less complex or costly [52]. Widrow and McCool [64] compared three array weight adjustment algorithms using either the method of steepest descent or the method of random search. Their indepth discussions provide performance comparisons in terms of the achievable convergence rate, gradient estimation methods and array weight vectors quantisation noise effects. Furthermore, the term misadjustment was introduced, which is defined as the average excess Mean Square Error (MSE) divided by the Minimum Mean Square Error (MMSE) [64]. In other words, it can be referred as the interference rejection performance penalty arising from the imperfect statistical estimation process. It was shown that the LMS algorithm outperforms the Differential Steepest Descent (DSD) and the Linear Random Search (LRS) algorithms and therefore its employment is highly recommended, whenever circumstances permit. Note furthermore that the Linear Random Search (LRS) algorithm originates from the concept of random search by natural selection [65].

As for the SMI algorithm, the major drawback is the requirement of performing a matrix inversion every time, when a new set of array weight has to be determined. The matrix inversion complexity would increase proportionally to the cube of the number of array elements. This problem can be mitigated by using the recursive method introduced by Brennan *et al.* [66], in which the array weights can be computed without matrix inversion by directly updating the inverse of the received samples' covariance matrix. The performance of this recursive method was compared to that of the direct method and Applebaum's adaptive loops. The inverse update and the direct matrix inversion method showed a rapid convergence compared to that of the adaptive loop based method of [53]. Apart from the maximum SNR criterion [53], the LMS error criterion [51] and the maximum likehood criterion [56],

the constrained optimisation procedure of [58, 59] was also analysed, amongst other authors also by Applebaum and Chapman [67] as well as by Takao *et al.* [68]. The capabilities of adaptive arrays in practical communication systems were also studied by a research team at the Ohio State University, where Compton *et al.* [69] presented an overview of the work conducted, providing also a bibliography of the papers, reports and other documents published. Apart from discussions on several adaptive systems that were built and tested, this team also contributed several studies [69] that covered topics such as the choice of the reference signal, the improvement of the LMS algorithm [70] for the sake of obtaining fixed nulls at certain angular positions with the aid of an adaptive array, the effects of multipath propagation on the performance of an adaptive array applied in the context of a spread spectrum system [71] and on the stochastic properties of the array weights [72]. Partial adaptivity has also been introduced by Chapman [73] as a way of balancing the implementation cost imposed versus the achievable complexity, which becomes an important consideration in the context of large arrays. For further material focusing on the issues of adaptive interference nulling, motivated readers are referred to a journal special issue on adaptive antennas, which was published in 1976 [52].

Based on further substantial developments in the field of adaptive antennas a third special issue was published in 1986 [74]. The primary difference with respect to the previous special issue is the inclusion of 'super-resolution' spatial spectrum estimation, whose resolution is higher than the traditional estimation approaches. The word 'resolution' generally refers to the ability of the algorithm to distinguish closely spaced signal sources. The broad and diverse topics of the special issue were divided into several classes, namely, spatial spectrum estimation, adaptive look-direction constraints, adaptive algorithms or techniques and application oriented contributions. In the adaptive algorithms category, a novel algorithm and architecture designed for adaptive digital beamforming was contributed by Ward, Hargrave and McWhirter [75]. A parallel spatial processing structure proposed by Su *et al.* [76] was aimed at combating signal cancellation effects in a correlated jamming environment. Their approach results in a spatially smoothened maximum likelihood estimate, while maintaining the maximum achievable array gain in the direction of the desired signal. Godara and Cantioni [77] extended their seminal work on gradient estimation, in which they considered both the scenario, when all of the array element signals were accessible as well as when they are inaccessible. In the context of spectral estimation, two established algorithms are the MUltiple SIgnal Classification (MUSIC) algorithm and the Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) algorithm. MUSIC [78] was first proposed in 1979 by Schmidt and independently by Bienvenu and Kopp [79], while ESPRIT was proposed by Paulraj, Roy and Kailath [80]. Several refinements of the MUSIC and ESPRIT algorithms [81] include the min-norm MUSIC, the smoothed MUSIC, the root-MUSIC [82], the beam-space MUSIC, the total least-squares ESPRIT [83], the multi-invariance ESPRIT [1] and the higher order ESPRIT techniques.

As mentioned before, the early use of adaptive antennas in telecommunications was found in the context of military applications. It was during the 1990s that the interest in using antenna arrays

Year	Author	Contribution
'59	Howells [48]	Invented the IF sidelobe canceller (SLC). This first bona-fide adaptive antenna system that incorporated the capability of automatic interference nulling was partially adaptive.
'64	Special Issue on Active and Adaptive Antennas (I) [50] edited by Hansen	A collection of papers mainly concentrating on the family of self-steering or self-focusing arrays. The term 'adaptive antenna' was used for describing the self-phasing antenna system, which reradiates a signal in the direction from which it was received.
'65-'66	Applebaum [53]	The emergence of the first fully adaptive array [8]. Applebaum showed that the adaptive array processing technique is a generalisation of the SLC procedure. In comparison to the SLC, a fully adaptive array exhibits a higher grade of flexibility in terms of its adaptive performance characteristics [49], albeit this is achieved at a significantly increased implementation cost. Applebaum also derived the Howells-Applebaum algorithm, which is based on maximising the SNR at the array's output.
'67	Widrow, Mantey, Griffiths and Goode [51]	Developed the well-known Widrow-Hoff LMS algorithm based on the MMSE criterion. This simple LMS algorithm overcomes the main difficulty encountered in applying the steepest descent algorithm for array weight adjustment, since the calculation of the correlation matrix is not required for the weight calculation.
'69	Capon [56]	Introduced a beamforming technique using a desired DOA constraint referred to as the Minimum Variance Distortionless Response (MVDR) beamformer. Capon's technique belongs to the family of adaptive beamforming techniques that offers the ability to resolve signals that are separated by a fraction of an antenna beamwidth [8].
'69	Griffiths [54]	Introduced an algorithm for real-time processing in large antenna arrays.
'72	Frost [58]	Developed the constrained LMS algorithm, which has a self-correcting feature allowing it to operate for an arbitrarily long time, i.e. for a large number of iterations, because error accumulation has been eliminated.
'74	Reed, Mallet and Brennan [63, 66]	Presented a technique referred to as the Sample Matrix Inversion (SMI) procedure, which exhibits a rapid convergence (that depends only on the number of adaptive array elements) owing to the direct computation of the adaptive weights. Later the authors have overcome the major drawback of the SMI technique, namely its increased complexity due to performing matrix inversion, by introducing the recursive SMI algorithm [66].
'76	Special Issue on Adaptive Antennas (II) [52] edited by Gabriel	A collection of papers mainly concentrating on the issues of adaptive interference nulling. The antenna system is capable of responding to an unknown interference environment by automatically steering the nulls or reducing the sidelobes in the direction of the interferers, while maintaining the desired signal's beam.

Table 2.1: A summary of selected contributions on adaptive beamforming.

Year	Author	Contribution
'76	Compton, Huff, Swarner and Ksienski [69]	Assessing the capabilities of adaptive arrays in realistic wireless environments. The authors succesfully incorporated adaptive arrays in a single-channel spread spectrum communication system, in a TDMA satellite communication system and in a remote sensor communication system.
'76	Chapman [73]	Introduced partial adaptivity as a way of balancing the implementation cost imposed versus the achievable performance.
'78	Morgan [84]	Studied the selection of auxiliary elements in partially adaptive antenna arrays for the sake of optimising the performance in a multiple narrowband jammer environment. Morgan also derived an explicit solution for a two-jammer problem. Apart from the ULA configuration, a planar array has also been considered.
'79	Schmidt [78]	Introduced the MUSIC algorithm, which is based on spectral estimation.
'80	Gabriel [85]	Studied various nonlinear spectral analysis techniques in the context of RF adaptive antenna array systems.
'81	Mayhan, Simmons and Cummings [86]	Developed quantative estimates of how the performance of antenna arrays varies with the antenna and delay-line parameters and also developed some tools for the performance evaluation of frequency-dependent weighting, i.e. for wideband beamforming.
'84	Mucci [87]	Presented various concepts, which aid the efficient implementation of discrete-time beamformers, including discussions on the associated hardware considerations, particularly in terms of analog-to-digital conversion, input data storage and the beamformer's computational complexity.
'86	Paulraj, Roy and Kailath [80]	Proposed the ESPRIT algorithm for DOA estimation.
'88	Veen and Buckley [88]	Provides an overview of beamforming from a signal processing perspective. The authors also discussed the concepts of data independent, statistically optimum, adaptive and partially adaptive beamforming.
'90	Anderson, Millnert, Viberg and Wahlberg [89]	Suggested a scheme for combatting CCI in a mobile communication system by means of adaptive antenna techniques.
'92	Chang, Yang and Chan [90]	Proposed a neural network approach to the MVDR beamforming problem.
'92	Kim, Cha and Youn [91]	Derived an expression for the output power of the GSC in the presence of correlated interference.
'92	Li [92]	Described the combined approach of beamforming and DOA-estimation using the ESPRIT algorithm for two widely spaced (25^o) incident signals and two closely spaced (4^o) incident signals for improving the performance of the ESPRIT algorithm.

Table 2.2: A summary of selected contributions on adaptive beamforming.

Year	Author	Contribution
'94	Woerner, Reed and Rappaport [93]	Presented an overview of the key simulation issues in the context of evaluating the performance of wireless systems that employ adaptive antenna arrays and DSP-based interference rejection techniques.
'94	Barrett and Arnott [94]	Presented the findings of the Smart Communication Antenna Research Programme (SCARP), describing the application of adaptive arrays or smart antennas in commercial communications applications.
'96	Klouche-Djedid and Fujita [95]	Analysed DOA estimation using high-resolution algorithms followed by a signal estimation procedure using several beamforming methods for the enhancement of the channel capacity in cellular systems.
'96	Litva and Lo [8]	Introduced digital beamforming and its applications in both wireless and satellite communications.
'97	Paulraj and Papadias [1]	Focused mainly on high-mobility TDMA networks, developing a signal model incorporating a range of channel effects and discussing the relationship between the forward-link (transmit) and reverse-link (receive) channels. Single- and multiuser models are proposed for space-time processing problems and the underlying spatial and temporal structures are discussed. Several algorithmic approaches to reverse-link space-time processing using both blind and trained methods for single- and multiple user scenarios are covered. Applications of space-time processing techniques in the context of both CDMA and other current cellular systems were also brought to the reader's attention.
'97	Godara [9, 81]	Provided a comprehensive and detailed treatment of beamforming schemes and their adaptive algorithms designed for both weight adjustment and DOA estimation along with their performance comparisons. A rich list of references covering most aspects of array signal processing was also provided.
'98	Rappaport [96]	A compilation of several smart antenna-related papers, categorised into five sections, namely introductory reading, the algorithms, the architechure, hardware and applications, the channel models as well as the corresponding performance evaluations.
'98	Winters [22]	Discussed current and future antenna technology and the performance improvement that adaptive antenna arrays can provide. A range of further issues involved in incorporating these antennas in CDMA, GSM and IS-36 wireless systems operating in different environments were also described.

Table 2.3: A summary of selected contributions on adaptive beamforming.

Year	Author	Contribution
'99	Sheikh, Gesbert, Gore and Paulraj [97]	Provided an overview of smart antenna applications in fixed broadband wireless access (BWA) networks. Different smart antenna techniques were described, including recent advances such as spatial multiplexing that is capable of dramatically increasing the performance of BWA networks. The impact of smart antenna techniques on the capacity and throughput of BWA networks was also discussed.
'00	Dietrich, Stutzman, Kim and Dietze [5]	Provided a review of smart antenna research applied to both base stations and handsets, with particular attention dedicated to the research conducted by the Smart Antenna Group at Virginia Tech.
'00	Bhobe and Perini [21]	Discussed both current and future smart antenna technology, focusing on the aspects of interference reduction and on increasing the attainable capacity. Different types of smart antenna systems using both switched beam and adaptive antenna array techniques were discussed along with their application in different multiple access schemes, such as FDMA, TDMA and CDMA.
'00	Boukalov and Häggman [98]	Discussed several issues that are important in future radio network design in conjunction with smart antennas. Three main system aspects of smart antenna technology were considered, namely smart antenna aided receivers and their related algorithms, wireless network control and smart antenna assisted cellular planning. The existing experimental and commercially available smart antennas and their performance were also surveyed.
'02	Blogh and Hanzo [4]	The authors studied the network capacity gains that can be achieved with the advent of adaptive antenna arrays and adaptive modulation techniques in both FDMA/TDMA and CDMA based mobile cellular networks. The performance of various UTRA soft-handover techniques was compared and the network capacity performance attained with the aid of adaptive antenna arrays was evaluated.
'02	Bellofiore, Balanis, Foutz and Spanias [99, 100]	Part 1 of the paper provided an overview of smart antenna aided systems and presented a planar array as a design example. The potential of smart antennas with regard to providing increased capacity in wireless communication networks was also discussed. As a continuation of Part 1, Part 2 introduced the signal processing aspects of antenna arrays. Specifically, both DOA- and adaptive beamforming algorithms were described. A study of the achievable network throughput, of the antenna radiation patterns and the required length of the training sequence used by the beamforming algorithm was also provided.

Table 2.4: A summary of selected contributions on adaptive beamforming.

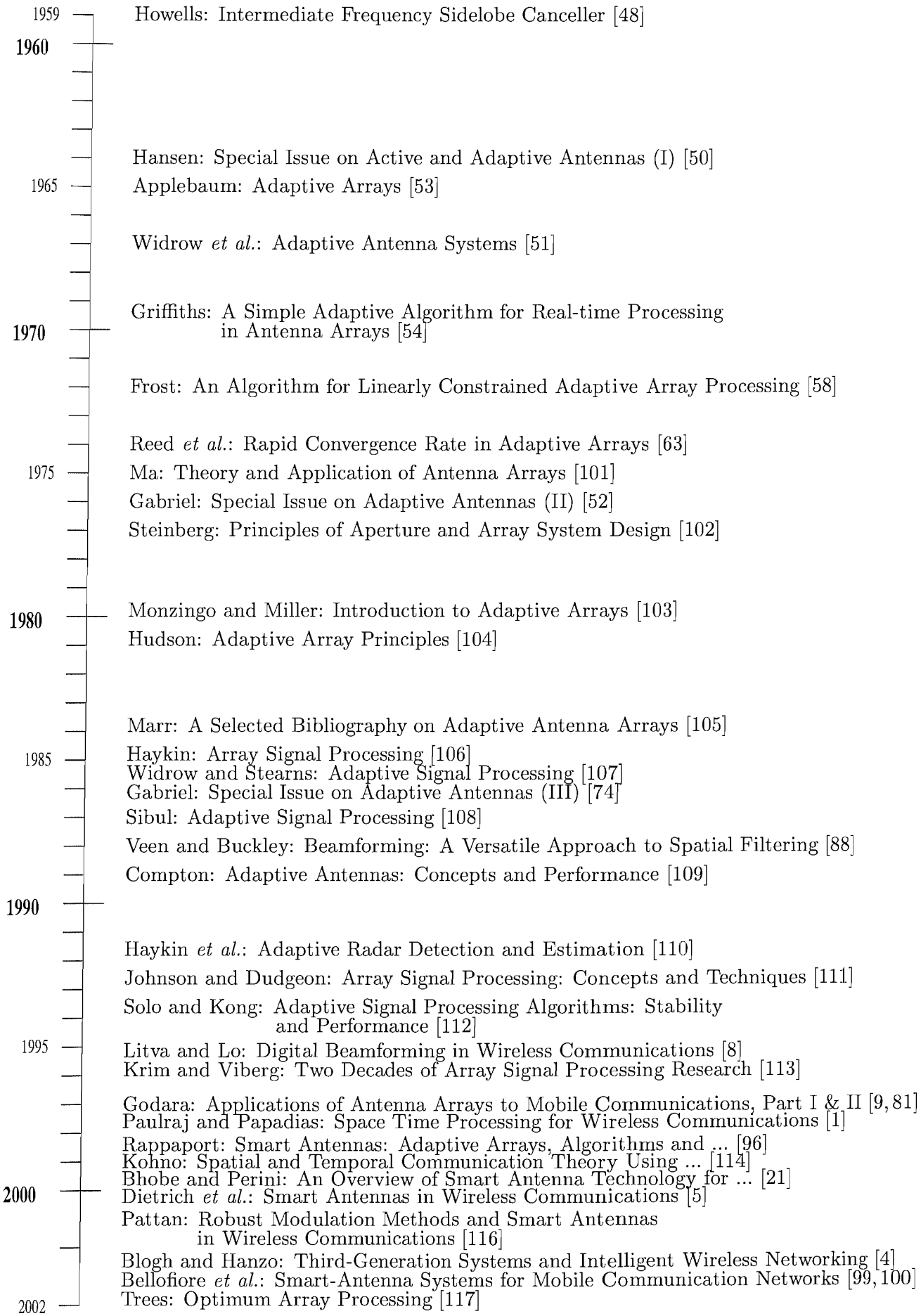


Figure 2.1: The evolution of research topics in adaptive beamforming.

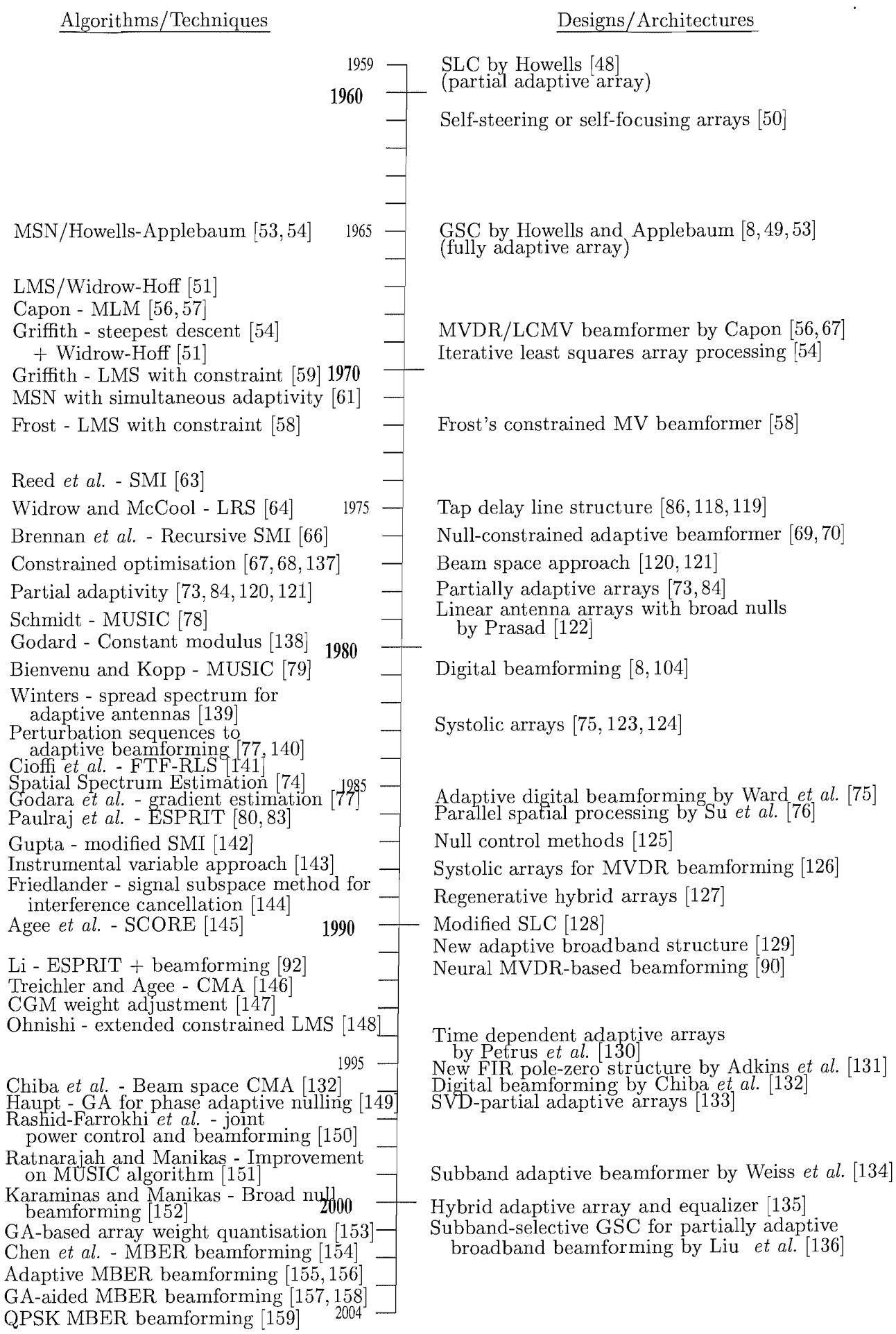


Figure 2.2: Evolution of adaptive beamforming algorithms and beamformer designs.

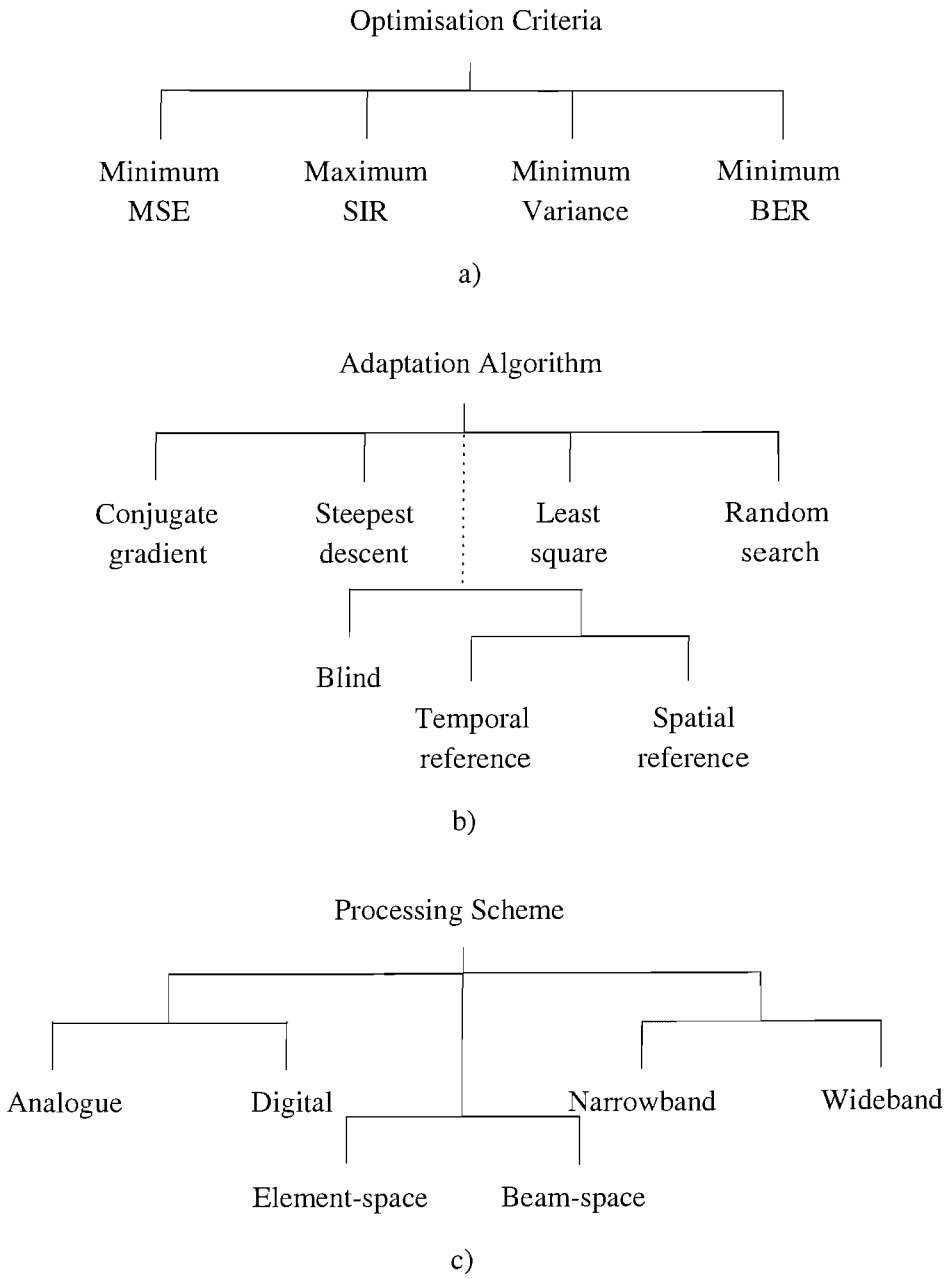
in the context of civilian cellular communications began to grow rapidly. This trend was also motivated by the advances in processor cost and speed. A large variety of new algorithms was derived as refinements of previously proposed algorithms, which were designed for different environments using Frequency Divison Multiple Access (FDMA), Time Divison Multiple Access (TDMA), Code Divison Multiple Access (CDMA) or Spatial Divison Multiple Access (SDMA) schemes. A summary of the above-mentioned contributions, including several selected publications on the evolution of adaptive beamforming was given in Tables 2.1, 2.2, 2.3 and 2.4. Specifically, Figure 2.1 summarised several important milestones in the context of monographs, collections of papers and text books, while Figure 2.2 outlined the evolutionary path of algorithms and array designs that emerged following the seminal contribution of Howells in 1959. Having reviewed the history and evolution of adaptive beamforming, we will now proceed by classifying the adaptive beamforming schemes known at the time of writing.

2.1.2 Classification of Adaptive Beamforming

The adaptive beamformer may be viewed as a learning and self-optimising system. It can be categorised either on the basis of the optimisation criterion employed or with respect to the algorithm used for adjusting the array weights [100]. Thirdly, there are several processing schemes available for beamforming, which include digital beamforming (DBF) [8], analogue beamforming [8], element-space beamforming [8, 81] and beam-space beamforming [8, 81]. Beamforming may also be categorised as narrowband [81, 88] or wideband beamforming [51, 54, 58, 81, 88, 97]. The breakdown of these categories is provided in Figure 2.3. In the first category, optimisation is governed by the performance criterion used. For example, when using the Mean Square Error (MSE) criterion, the output of the beamformer is optimised such that the MSE between the beamformer's output $y(n)$ and some desired response $r(n)$ is minimised. When using the Signal-to-Interference Ratio (SIR) criterion, the output is optimised such that the SIR is maximised, while in the case of the minimum variance beamformer the variance between the desired and actual output is minimised. In the context of the Minimum Bit Error Ratio (MBER) criterion the output of the beamformer is optimised such that the BER between the transmitted and the received bits is minimised.

As seen in Figure 2.3, several adaptive algorithms are available for adjusting the array weights, such as the steepest descent method, least squares method, conjugate gradient method and random search method. The choice of the adaptive algorithm to be used is perhaps the most important design decision to be made, since it determines for example the speed of convergence and also the hardware complexity required for implementing the algorithm, which will then contribute to the overall performance of the system. Haykin [6] and Rappaport [7] listed several factors that may affect the preferences of one algorithm over another in the context of adaptive filters, which can also be applied to adaptive beamforming.

1. **Rate of convergence:** This is defined as the number of array weight adjustment iterations



MSE : Mean Square Error

SIR : Signal-to-Interference Ratio

BER : Bit Error Ratio

Figure 2.3: Classification of adaptive beamforming based on a) optimisation criteria, b) weight adaptation algorithm and c) processing scheme employed.

required for the algorithm, in response to stationary inputs, to converge sufficiently close to the optimum Wiener solution in the sense of a defined criterion, such as the MSE. A fast rate of convergence allows the algorithm to adapt rapidly to a stationary environment of unknown statistics.

2. **Misadjustment:** For an algorithm of interest, this parameter provides a quantitative measure of the amount by which the final value of the MSE, averaged over an ensemble of adaptive filters, deviates from the optimal MMSE solution that is produced by the Wiener filter.

3. **Tracking:** In a non-stationary environment, the algorithm is required to track statistical variations in the environment. The tracking performance of an algorithm, however, is influenced by two contradictory features, namely by the rate of convergence and by the steady state fluctuation due to the array weights' quantisation noise.

4. **Robustness:** This measure characterises the ability of an algorithm to operate satisfactorily with ill-conditioned input data, which shows the stability of the algorithm. The data sequence is said to be ill-conditioned when producing large error on the solutions while having a small error. Robustness may also be used in the context of the algorithm's numerical behaviour (see item 6).

5. **Computational complexity:** This can be quantified in terms of the number of arithmetic operations required for a complete iteration of the algorithm and/or the size of the memory required to store both the data and the program.

6. **Numerical properties:** When an algorithm is implemented, inaccuracies are produced due to round-off noise and finite-precision errors. These errors influence the stability of an algorithm.

The factors listed above are also summarised in Figure 2.4. The choice of an algorithm and its corresponding convergence speed would also depend on the radio channel characteristics encountered. Note that an algorithm may have a very high rate of convergence, but may be suffering from a high computational complexity. In other words, there is always a trade-off between these two features.

Another significant parameter influencing the performance of an adaptive beamforming scheme is the choice of the reference signal transmitted in order to differentiate the desired signal from the interferers; if an explicit reference signal is not used or unavailable, the adaptive beamforming scheme is termed as a blind adaptive beamforming regime. Adaptive beamforming can be further divided into adaptive beamforming schemes using temporal reference or that involving spatial reference techniques. Further comments on these categories will be provided in the subsequent sections. In the next section, we will elaborate on our signal model that will be used throughout this chapter.

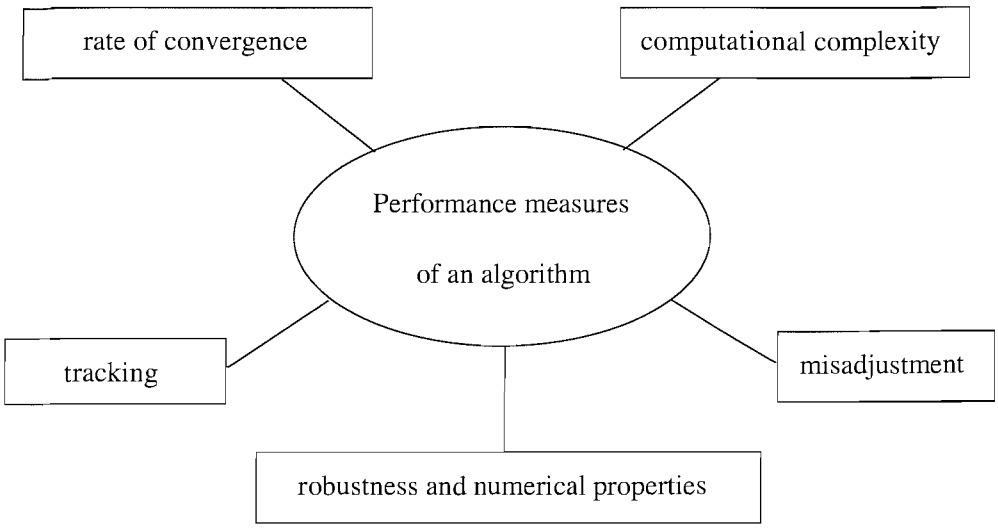


Figure 2.4: Several contradictory design factors that contribute to the overall performance of a beamforming algorithm.

2.2 Basic Concepts of Beamforming

2.2.1 Signal Model

The employment of antenna arrays in mobile radio systems for combating the effects of co-channel interference was first discussed in 1980 by Yeh and Reudink [160], demonstrating that with the aid of a moderate number of antenna elements, it was possible to substantially increase the achievable spectral efficiency [8]. The advantages of antenna arrays have been lavishly documented in the literature [4, 8–10, 13, 14, 22]. The antenna elements can be arranged according to different geometries, with the uniform linear [8, 113, 161], circular [8, 113, 161] and planar [8, 100, 161] arrays being the most popular design options. Here, we continue our discourse by presenting our signal model, which is based on the uniform linear array (ULA) configuration.

Consider an array of L omni-directional antenna elements separated by a distance d in the far-field of M uncorrelated point sources, as shown in Figure 2.5 a). In general, the M point sources represent M mobile stations, where one of them is the desired user, while $(M-1)$ users impose interference. Let us first concentrate on the scenario of having a single source, say source 1. If the far-field wavefront of source 1 impinges upon the array at an angle θ_1 with respect to the array normal, as shown in Figure 2.5 b), the wavefront will reach element l, before arriving at element $l-1$, for example it will reach element 2 before arriving at element 1. The extra distance the wave has to travel for reaching element 1 after arriving at element 2 is $d\sin\theta_1$. As seen in Figure 2.5, the extra distance the wave has to travel from the lth element to element 1 is given by:

$$d_l(\theta_1) = d(l-1)\sin\theta_1. \tag{2.1}$$

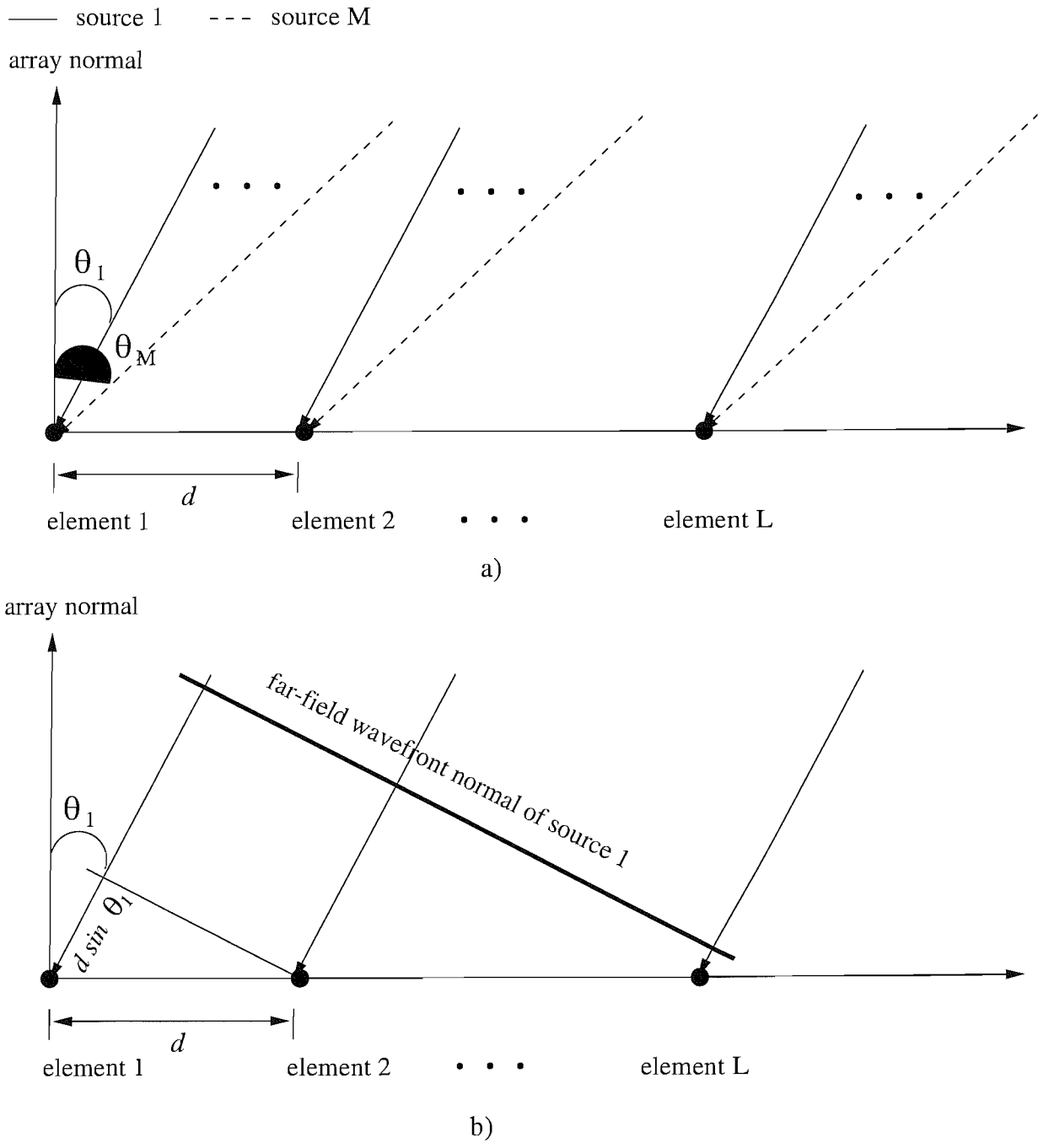


Figure 2.5: L-element linear antenna array in the far-field of a) M point sources of direction θ_1, θ_2, ..., θ_M, respectively b) a single point source of direction θ_1.

Hence, for an arbitrary source direction θ_m, $m = 1 \ldots M$, this relationship can be expressed as:

$$d_l(\theta_m) = d(l-1)\sin\theta_m. \tag{2.2}$$

Equivalently, in terms of time the delay measured from element l to element 1 is given by:

$$t_l(\theta_m) = \frac{d(l-1)}{c}\sin\theta_m, \tag{2.3}$$

where c is the speed of light, which is related to the frequency f and the wavelength λ by $c = f\lambda$. These equations are based on the assumption that the elements are located along the x-axis of the Cartesian plane with the reference element positioned at $(0,0)$. For the general case of an arbitrary array on the x-y Cartesian plane, the time delay encountered by an element at position (p,q) with respect to the origin is given by:

$$t_{(p,q)}(\theta_m) = \frac{|p|\sin\theta_m + |q|\cos\theta_m}{c}. \tag{2.4}$$

The signal induced in the lth array element due to the mth signal source, where $m = 1, 2, \ldots, M$ is normally expressed in complex form as:

$$\bar{x}_{l,m}(n) = \bar{m}_m(n)e^{j2\pi f t_l(\theta_m)}, \tag{2.5}$$

where we use a bar, ' $\bar{}$ ' over x (i.e. $\bar{x}$) to denote a pure noise-free signal and $\bar{m}_m(n)$ denotes the complex-valued modulating signal of user m. The structure of the modulating signal reflects the particular modulation scheme used in the communication system considered [81]. In our case, we assume a simple Binary Phase Shift Keying (BPSK) modulated signal in a non-dispersive narrowband channel, thus the modulating signal of user m can be expressed as:

$$\bar{m}_m(n) = a_m(n)b_m(n), \tag{2.6}$$

where the complex-valued coefficient $a_m(n)$ models the multiplication of the channel coefficient of user m with the transmitted signal power of user m and therefore $|a_m(n)|^2$ denotes the received signal power for user m, while $b_m(n) \in \{\pm 1\}$, for $m = 1, 2, \ldots, M$.

In vectorial representation, the total output signal of the array is given by:

$$\bar{\mathbf{x}}(n) = \begin{bmatrix} \bar{x}_1(n) \\ \vdots \\ \bar{x}_L(n) \end{bmatrix} = \begin{bmatrix} \bar{m}_1(n)e^{j2\pi f t_1(\theta_1)} + \ldots + \bar{m}_M(n)e^{j2\pi f t_1(\theta_M)} \\ \vdots \\ \bar{m}_1(n)e^{j2\pi f t_L(\theta_1)} + \ldots + \bar{m}_M(n)e^{j2\pi f t_L(\theta_M)} \end{bmatrix}, \tag{2.7}$$

where the so-called steering vector $\mathbf{s}_m$ associated with the mth signal source, $m = 1, ..., M$, namely with the signal arriving from direction θ_m, can be represented by an L-row complex vector:

$$\mathbf{s}_m = \begin{bmatrix} e^{j2\pi f t_1(\theta_m)} \\ e^{j2\pi f t_2(\theta_m)} \\ \vdots \\ e^{j2\pi f t_L(\theta_m)} \end{bmatrix}. \tag{2.8}$$

For an element spacing d of $d = \lambda/2$ the steering vector can be simplified to:

$$\mathbf{s}_m = \begin{bmatrix} 1 \\ e^{j\pi \sin\theta_m} \\ \vdots \\ e^{j\pi(L-1)\sin\theta_m} \end{bmatrix}. \tag{2.9}$$

Since in the context of BPSK we have a constant modulating signal magnitude, the symbol index n constitutes the only parameter that influences the modulating signal of $\bar{m}_m(n) = \pm 1$ in Equation (2.7). We could therefore express the total noise-free output signal of the antenna array in terms of an $(L \times M)$-dimensional matrix $\mathbf{G}$ acting as the system matrix of the beamformer, such that:

$$\bar{\mathbf{x}}(n) = \begin{bmatrix} a_1 e^{j2\pi f t_1(\theta_1)} & \dots & a_M e^{j2\pi f t_1(\theta_M)} \\ & \vdots & \\ a_1 e^{j2\pi f t_L(\theta_1)} & \dots & a_M e^{j2\pi f t_L(\theta_M)} \end{bmatrix} \begin{bmatrix} b_1(n) \\ \vdots \\ b_M(n) \end{bmatrix} \tag{2.10}$$

$$= \begin{bmatrix} a_1\mathbf{s}_1 & \dots & a_M\mathbf{s}_M \end{bmatrix} \mathbf{b}(n) \tag{2.11}$$

$$= \mathbf{SAb}(n) \tag{2.12}$$

$$= \mathbf{Gb}(n), \tag{2.13}$$

where $\mathbf{S}$ is the $(L \times M)$-dimensional steering matrix given by:

$$\mathbf{S} = \begin{bmatrix} e^{j2\pi f t_1(\theta_1)} & \dots & e^{j2\pi f t_1(\theta_M)} \\ & \vdots & \\ e^{j2\pi f t_L(\theta_1)} & \dots & e^{j2\pi f t_L(\theta_M)} \end{bmatrix} \tag{2.14}$$

and $\mathbf{A}$ is a diagonal matrix that contains the binary modulating signals a_m of the M point sources or users. Note that we use boldface lowercase characters to denote the vectors and boldface uppercase characters for matrices. In a more compact form, the total noise-contaminated array output vector $\mathbf{x}(n)$ can be expressed as:

$$\mathbf{x}(n) = \bar{\mathbf{x}}(n) + \mathbf{n}(n)$$

$$= \mathbf{Gb}(n) + \mathbf{n}(n) \tag{2.15}$$

$$= \left[\sum_{m=1}^{M} \bar{m}_m(n)\mathbf{s}_m(n)\right] + \mathbf{n}(n), \tag{2.16}$$

which consists of the superposition of the desired and interfering signals and random noise. The vector $\mathbf{n}(n)$ denotes the complex array noise vector of $\mathbf{n}(n) = [n_1(n), n_2(n), ..., n_L(n)]^T$, where $n_l(n)$ is the complex random noise component imposed on the lth array element and T represents the transpose of a vector or a matrix. The noise is assumed to be temporally white with zero-mean and variance of $E[|n_l(n)|^2] = 2\sigma_n^2$. Therefore, the desired user's Signal-to-Noise Ratio is defined as SNR $= |a_1|^2/2\sigma_n^2$, assuming that user 1 is the desired user. Similarly, the Interference-to-Noise Ratio of user m is given by $\text{INR}_m = |a_m|^2/2\sigma_n^2$ and the desired Signal-to-Interference Ratio with respect to user m is defined as $\text{SIR}_m = |a_1|^2/|a_m|^2$, where $m = 2, ..., M$.

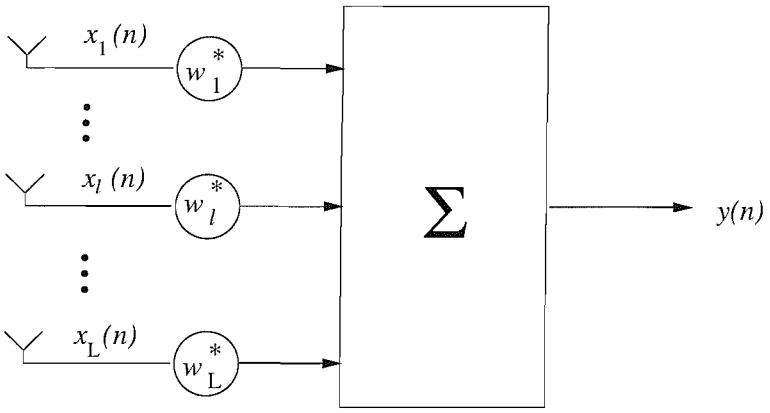


Figure 2.6: A narrowband beamformer structure.

Figure 2.6 shows the simple structure of a narrowband beamformer. It can be directly observed that the output $y(n)$ of the beamformer is constituted by the sum of L appropriately weighted array element signals. More explicitly, the signal received by each element, namely $x_l(n)$ is multiplied by a complex array weight w_l^* and summed, yielding the output $y(n)$ as:

$$y(n) = \sum_{l=1}^{L} w_l^*(n)x_l(n), \tag{2.17}$$

where '*' in $w_l^*(n)$ denotes the complex conjugate value of the weights $w_l(n)$. It is convenient to multiply the data by the conjugates of the weights for the sake of later notation simplification. In vectorial notation the weights of the beamformer are represented as:

$$\mathbf{w}(n) = \begin{bmatrix} w_1(n) \\ \vdots \\ w_L(n) \end{bmatrix} \tag{2.18}$$

and taking into account the signals induced in all elements as formulated in Equation (2.7), the output of the beamformer can be expressed as:

$$y(n) = \mathbf{w}^H(n)\mathbf{x}(n), \tag{2.19}$$

where the superscript H denotes the complex conjugate transpose of a vector or a matrix. Assuming that the components of the array output $\mathbf{x}(n)$ can be modeled as zero-mean stationary complex processes, the mean output power of the beamformer is given by:

$$\begin{aligned} P(\mathbf{w}) &= E\left[y(n)y^*(n)\right] \\ &= \mathbf{w}^H\mathbf{R}\mathbf{w}, \end{aligned} \tag{2.20}$$

where $E[\cdot]$ denotes the expectation operator and $\mathbf{R}$ is the $(L \times L)$-dimensional auto-correlation matrix

of the received signal samples at the array output, defined as:

$$\mathbf{R} = E\left[\mathbf{x}(n)\mathbf{x}^H(n)\right] = E\left[\begin{bmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_L(n) \end{bmatrix} \begin{bmatrix} x_1^*(n) & x_2^*(n) & \dots & x_L^*(n) \end{bmatrix}\right]. \tag{2.21}$$

Expanding Equation (2.21) and using the condition of wide sense stationarity, we obtain:

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} r(0) & r(1) & \dots & r(L-1) \\ r(-1) & r(0) & \dots & r(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-L+1) & r(-L+2) & \dots & r(0) \end{bmatrix} \\ &= \begin{bmatrix} r(0) & r(1) & \dots & r(L-1) \\ r^*(1) & r(0) & \dots & r(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(L-1) & r^*(L-2) & \dots & r(0) \end{bmatrix}. \end{aligned} \tag{2.22}$$

Elements of the auto-correlation matrix are the correlation coefficients between the signals received by the various elements. For example, $\mathbf{R}_{ij}$ denotes the correlation between the received signals of the ith and the jth element of the array. Observe furthermore that the structure of $\mathbf{R}$ is Hermitian, i.e. we have $\mathbf{R}^H = \mathbf{R}$. More explicitly, this correlation matrix has a Toeplitz structure, i.e. $r(0)$ is real and the off-diagonal values parallel to the main diagonal elements are all equal. Hence the correlation matrix of the received signal samples of the various array elements can be expressed in a compact form as:

$$\mathbf{R} = \left[\sum_{m=1}^{M} \sigma_m^2 \mathbf{s}_m \mathbf{s}_m^H\right] + 2\sigma_n^2 \mathbf{I}_L, \tag{2.23}$$

where σ_m^2 is the variance, which is also equal to the power of the associated zero-mean incoming signal of user m and $\mathbf{I}_L$ is an $(L \times L)$-dimensional identity matrix. Using matrix notation, Equation (2.23) can be expressed as:

$$\mathbf{R} = \mathbf{G}\mathbf{G}^H + 2\sigma_n^2 \mathbf{I}_L, \tag{2.24}$$

or as:

$$\mathbf{R} = \mathbf{S}\mathbf{P}\mathbf{S}^H + 2\sigma_n^2 \mathbf{I}_L, \tag{2.25}$$

where $\mathbf{G}$ is the system matrix as defined in Equation (2.13), while $\mathbf{S}$ is the $(L \times M)$-dimensional matrix of Equation (2.14) comprised of the steering elements defined in Equation (2.8) and $\mathbf{P}$ is given by $\mathbf{P} = \mathbf{A}\mathbf{A}^H = \mathbf{A}\mathbf{A}^* = diag[|a_m|^2], m = 1, 2, \dots, M$, with $\mathbf{A}$ defined in Equation (2.12). The elements of the $(M \times M)$-dimensional diagonal matrix $\mathbf{P}$ are given by the variance or power of the

incoming signals $|a_m|^2, m = 1, 2, \ldots, M$. It should also be noted that if the signals arriving from the M point sources are uncorrelated, we have $E[\bar{m}_i \bar{m}_j] = 0$ for $i \neq j$, hence the total correlation matrix $\mathbf{R}$ can be expressed as the superposition of the correlation matrices of the individual sources and noise, yielding $\mathbf{R} = \mathbf{R}_{m_i} + \mathbf{R}_{m_j} + 2\sigma_n^2 \mathbf{I}_L$.

For a linear antenna array constituted by L omni-directional elements and placed along the $x-$axis, i.e. based on Equation (2.4), as in Figure 2.5 the beam pattern is given by [8]:

$$F(\theta_m) = \sum_{l=1}^{L} w_l e^{-j2\pi f t_l(\theta_m)}, \tag{2.26}$$

which describes the response of the beamformer to the source signal arriving at an angle of θ_m. If we closely observe Equation (2.26), it is readily seen that it has a similar structure to that of the beamformer's output defined in Equation (2.17). More explicitly, while in Equation (2.17) the time-domain array output is formulated as the sum of the weighted array element output signals, in Equation (2.26) the response of the beamformer to the source signal arriving at an angle of θ_m is formulated as the sum of the kernel functions $e^{-j2\pi f t_l(\theta_m)}$, weighted by the array weights.

Having introduced the signal model and some elementary equations associated with beamforming, let us now take a closer look at two specific categories of beamforming, namely at element-space [8,81] and beam-space [8,81] processing schemes as well as at both narrowband and wideband beamforming [81,88], which will be defined in Sections 2.2.2.1 and 2.2.2.2, respectively.

2.2.2 Beamforming Processing Scheme

The process of beamforming may broadly be classified as either analogue or digital. It can also be divided into two significant categories known as element-space beamforming [8, 81] and beam-space beamforming [8,81], as well as into narrowband- [81,88] and wideband beamforming [51,54,58,81,88].

2.2.2.1 Element-Space Beamforming and Beam-Space Beamforming

In element-space beamforming, as discussed so far, the received signals of the array elements $(x_1, x_2, ..., x_L)$ are directly multiplied by a set of array weights $(w_1, w_2, ..., w_L)$ to form a transmit or receive beam in the desired direction θ, where L indicates the number of elements in the antenna array. The number of beams to be formed is however, not limited to one. For the sake of forming multiple beams receiving from I number of desired directions, the received signals $(x_1, x_2, ..., x_L)$ are multiplied by I number of different sets of weights, namely by $(w_1^1, w_2^1, ..., w_L^1)$,$(w_1^2, w_2^2, ..., w_L^2)$, ..., $(w_1^I, w_2^I, ..., w_L^I)$. Each set of weights yields a different output, representing a different array pattern or beam. Hence we have:

$$y(\theta_i) = \sum_{l=1}^{L} w_l^{i*} x_l \,, \qquad i = 1, 2, \ldots, I, \tag{2.27}$$

where $y(\theta_i)$ denotes the output of the beamformer, x_l is the lth array element's output signal and w_l^{i*} is the corresponding weight assisting in forming a beam at the desired angle θ_i, $i = 1, 2, \ldots, I$.

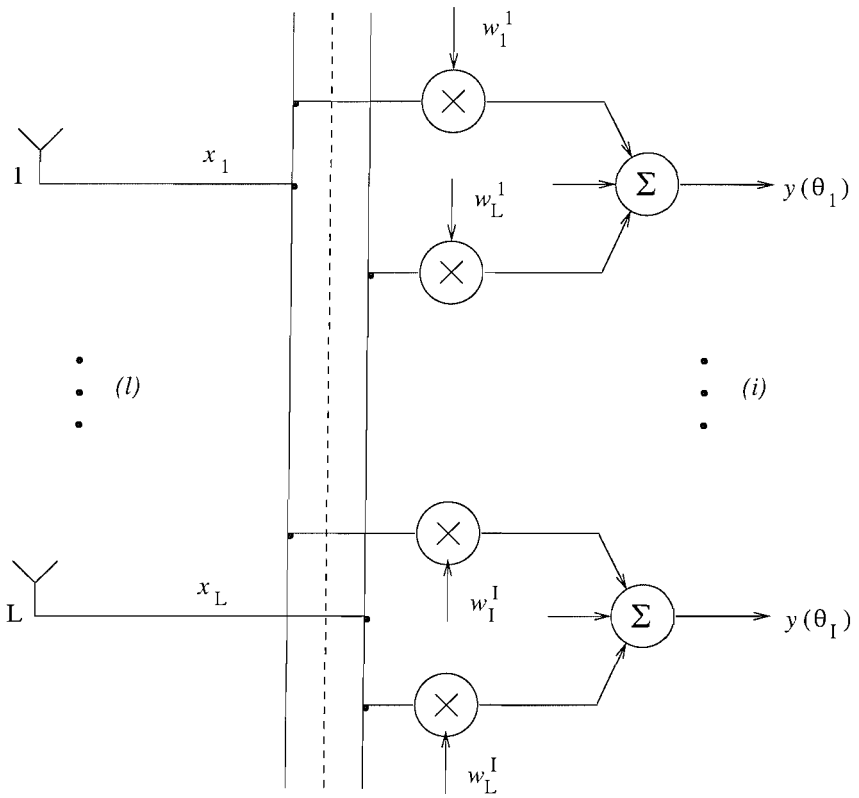


Figure 2.7: Element-space beamformer designed for simultaneously generating I number of receiver beams [8].

In contrast to the method of element-space beamforming, in which the output $y(\theta_i)$ of the beamformer is produced by summation of the multiplication between the array received signals x_l and array weights w_l, the beam-space beamforming approach first forms a number of fixed beams. This multiple fixed beams may be produced using a fixed beamforming network, in which commonly the weights used are fixed [8,11] and spatially orthogonal, i.e. the weight vectors' dot product is equal to zero [81]. The outputs of these beams are then adaptively weighted and combined in order to form the desired outputs. The simplified schematic of the element-space and beam-space processing arrangements designed for generating multiple receiver beams is provided in Figure 2.7 and 2.8, respectively.

With reference to the beam-space beamforming scheme of Figure 2.8, the Fast Fourier Transform (FFT) block in the diagram generates L orthogonal beams v_l. The 'beam select' block of Figure 2.8 will then choose a number of these orthogonal beams that are to be weighted for the sake of forming the desired output. For example, assume that the desired output y_i requires the weighted

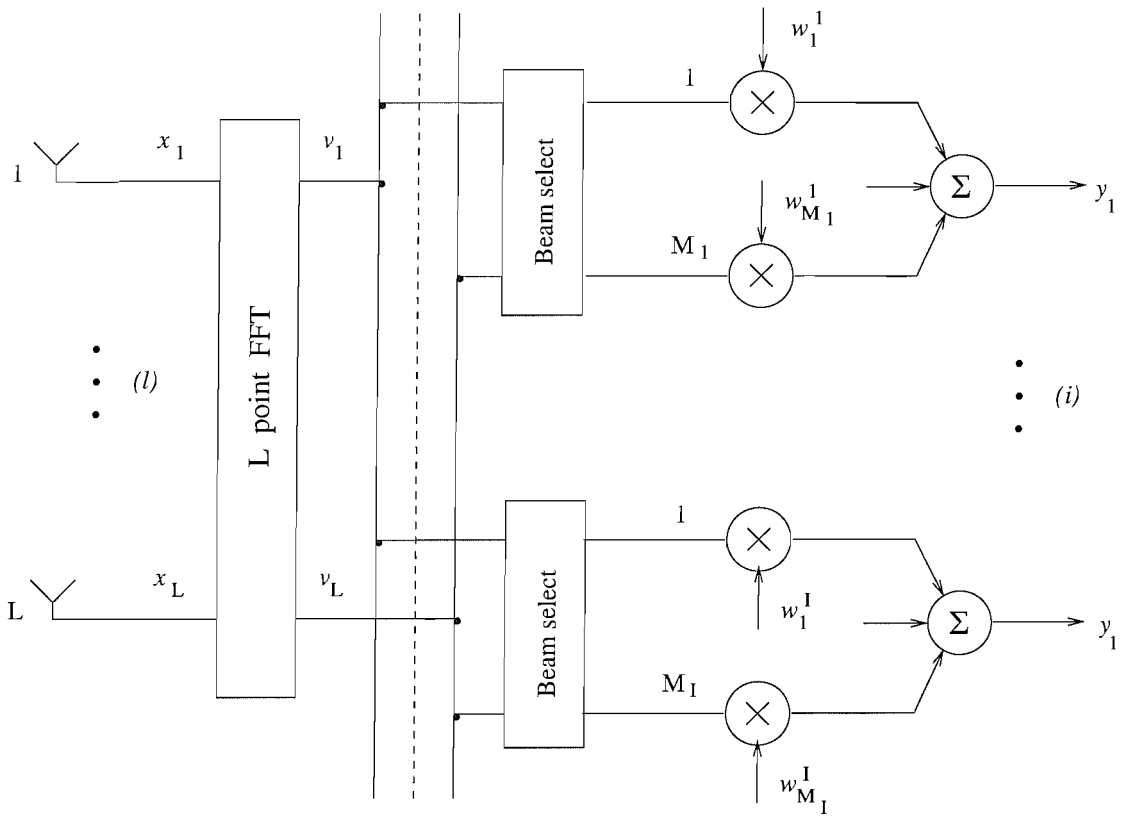


Figure 2.8: Beam-space beamformer designed for simultaneously generating I number of receiver beams [8].

sum of the lth and $(l+2)$nd beams, yielding:

$$y_i = w_1^i v_l + w_2^i v_{l+2} \tag{2.28}$$

$$= \sum_{m=1}^{M_i} w_m^i v_{i_m}, \tag{2.29}$$

where i_m is the selected beam's index (e.g. $i_1 = l$ and $i_2 = l+2$) and M_i is the number of orthogonal beams that are required for forming the ith desired beam. In this case we have $M_i = 2$.

Beam-space processors have been studied under numerous different names, including the terminology of Howells-Applebaum array [53,67], Generalised Sidelobe Canceller (GSC) [49,88], adaptive-adaptive array [81,162] and multiple-beam antennas [11,81,163]. Reference [81] provides a good source of further reading on beam-space processing.

2.2.2.2 Narrowband Beamforming and Wideband Beamforming

Beamforming may also be grouped into either narrowband [81,88] or wideband beamforming [51,54,58, 81,88,97], depending on the bandwidth of the signal environment. Narrowband beamforming assumed that the signal bandwidth is sufficiently narrow to be considered as a single frequency. Generally, if the signal bandwidth is a small fraction (say less than 1%) of the carrier frequency, this scenario may

be classified as narrowband, otherwise the signal is deemed to exhibit wideband characteristics [164]. However, the presence of wideband beamforming conditions also depends on the time interval over which the signal is observed [88]. Based on statistical arguments, it turns out that the observation time interval and bandwidth product is the fundamental parameter that determines, whether a signal can be considered as narrowband [88].

Therefore a signal may be considered to exhibit wideband characteristics, when the signal's bandwidth is wider than the channel's frequency response. Viewing matters in the time domain, since the CIR exhibits several delayed tap, the received signal includes multiple versions of the transmitted waveform, which are attenuated and delayed in time, thus linearly distorting the received signal. This results in intersymbol interference (ISI). In our investigations we will be concentrating on the non-dispersive AWGN channel, thus we will invoke spatial processing but no time-domain processing. The use of narrowband beamformers for detecting wideband signals will degrade the attainable performance of the beamformer. When receiving wideband signals, the received signal of each array element has to be fed into a time-domain equaliser, which combines the signal components having different delays, leading to the concept of temporal processing [1,114], [165]. The combination of both spatial and temporal processing involves simultaneous filtering in both the space- and time-domain and therefore it is referred to as space-time processing [1, 114], [165]. Narrowband beamforming is conceptually simpler than wideband beamforming, since we can ignore the temporal domain.

Having provided the necessary rudimentary foundations on beamforming, in the next section we will consider a simple worked example.

2.2.3 Beamforming Example

In reference to our signal model outlined in Section 2.2.1, let us consider having only two antenna elements separated by a distance d which equals to $\lambda/2$ (half a wavelength), as shown in Figure 2.9. We assume that there are only two signals transmitted at the same frequency f, one arriving from $\theta_1 = 0^o$ as the desired signal $\bar{m}_1(n)$ and another one from $\theta_2 = 30^o$ (equivalent to $\frac{\pi}{6}$ radians) as the interferer $\bar{m}_2(n)$.

The signals received by the elements are given by:

$$x_1 = \bar{m}_1(n)e^{j2\pi f t_1(\theta_1)} + \bar{m}_2(n)e^{j2\pi f t_1(\theta_2)} \quad (2.30)$$

$$x_2 = \bar{m}_1(n)e^{j2\pi f t_2(\theta_1)} + \bar{m}_2(n)e^{j2\pi f t_2(\theta_2)}, \quad (2.31)$$

respectively, where $t_l(\theta_m)$ is the time delay defined in Equation (2.3). This is equivalent to applying Equation (2.16) apart from the absence of the noise vector term. In vectorial representation, the array output is given by:

$$\bar{\mathbf{x}}(n) = \bar{m}_1(n)\mathbf{s}_1(n) + \bar{m}_2(n)\mathbf{s}_2(n), \quad (2.32)$$

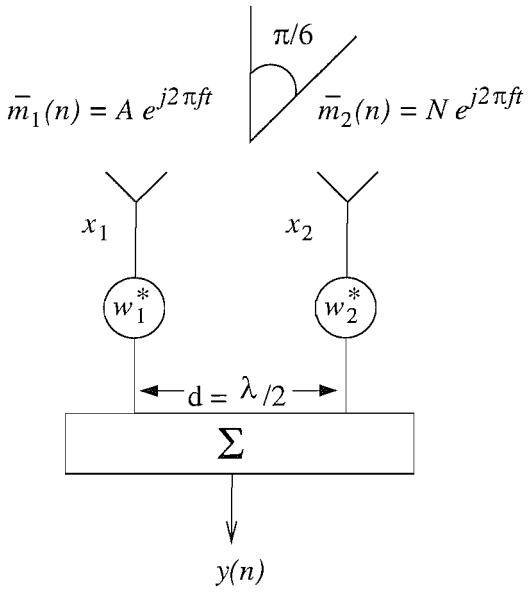


Figure 2.9: Beamforming receiver example with a wanted signal at 0^o and an interfering signal at 30^o for an array having an element spacing of $\lambda/2$.

where for the signals arriving from 0 radian and $\frac{\pi}{6}$ radians, we have the steering vectors of:

$$\begin{aligned} \mathbf{s}_1 &= [e^{j2\pi ft_1(\theta_1)}, e^{j2\pi ft_2(\theta_1)}]^T = [1,1]^T \\ \mathbf{s}_2 &= [e^{j2\pi ft_1(\theta_2)}, e^{j2\pi ft_2(\theta_2)}]^T = [1,j]^T \end{aligned} \tag{2.33}$$

for the desired signal and the interferer, respectively.

As illustrated in Figure 2.6, the signal emerging from each of the two elements is multiplied by a complex-valued array weight, and the weighted signals are then summed in order to form the beamformer's output. Therefore, the noise-free output of the beamformer is given by:

$$\begin{aligned} \bar{y}(n) &= \sum_{l=1}^{2} w_l^*(n) x_l(n) \\ &= w_1^*[\bar{m}_1(n) + \bar{m}_2(n)] + w_2^*[\bar{m}_1(n) + j\bar{m}_2(n)] \\ &= \bar{m}_1(n)(w_1^* + w_2^*) + \bar{m}_2(n)(w_1^* + jw_2^*). \end{aligned} \tag{2.34}$$

When steering the array's beam pattern in the direction of the desired signal, while simultaneously suppressing the interfering signals, the array's output due to the desired signal $\bar{y}_d(n)$ becomes:

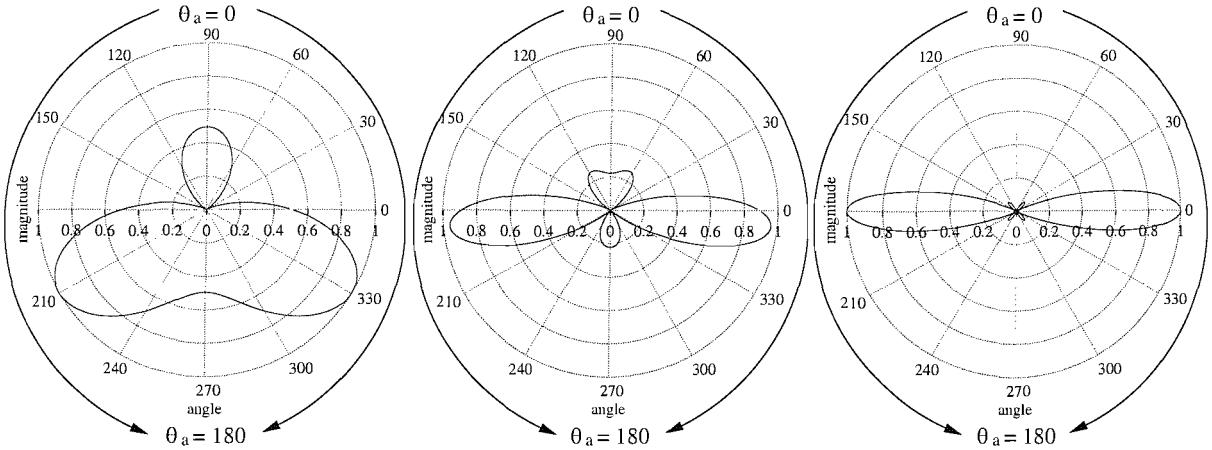
$$\bar{y}_d(n) = \bar{m}_1(n)(w_1^* + w_2^*) = Ae^{j2\pi ft}(w_1^* + w_2^*) = Ae^{j2\pi ft} \tag{2.35}$$

and the array output owing to the interfering signal $\bar{y}_I(n)$ becomes zero, yielding:

$$\bar{y}_I(n) = \bar{m}_2(n)(w_1^* + jw_2^*) = Ne^{j2\pi ft}(w_1^* + w_2^* e^{j\pi/2}) = w_1^* Ne^{j2\pi ft} + w_2^* Ne^{j(2\pi ft + \pi/2)} = 0. \tag{2.36}$$

Equations (2.35) and (2.36) lead to two sets of complex weight values:

$$\begin{aligned} \Re[w_1] + \Re[w_2] &= 1 \\ \Im[w_1] + \Im[w_2] &= 0 \end{aligned} \tag{2.37}$$



(a) $L = 2$ elements (b) $L = 3$ elements (c) $L = 4$ elements

Figure 2.10: Polar beam pattern for the beamforming receiver example of Figure 2.9 employing a (a) two-element, (b) three-element and (c) four-element antenna array with a wanted signal at 0^o and an interfering signal at 30^o.

and

$$\begin{aligned} \Re[w_1] - \Re[w_2] &= 0 \\ \Im[w_1] + \Im[w_2] &= 0, \end{aligned} \tag{2.38}$$

where $\Re$ and $\Im$ denote taking the real and imaginary values, respectively. Solving Equations (2.37) and (2.38) simultaneously yields:

$$\begin{aligned} w_1^* &= 0.5 - j0.5 \\ w_2^* &= 0.5 + j0.5. \end{aligned} \tag{2.39}$$

The set of values in Equation (2.39) are the weighting coefficients to be used by a two-element antenna array for acquiring the desired signal $\bar{m}_1(n)$ arriving from $\theta = 0^o$, while eliminating the interfering signal $\bar{m}_2(n)$ impinging from $\theta = 30^o$. Figure 2.10 (a) shows the beam pattern in polar form for this specific example. Observe that the interfering signal from the direction of $\theta = 30^o$ is perfectly nulled. The effects of adding more array elements can be seen from Figures 2.10 (b) and 2.10 (c). Apart from placing a null in the direction of the interfering user at $\theta = 30^o$, the extra degrees of freedom (DOF), i.e. DOF$= L-1$ has been used to increase/improve the antenna gain in the direction of the desired user's signal arriving from $\theta = 0^o$, while simultaneously reducing the sidelobes formed in other angular directions. Note that the gain of the maximum sidelobe and the width of the main beam jointly determine the effective interference rejection capability. Note also that all beam patterns of Figure 2.10 are symmetrical with respect to the vertical line extending from 90^o to 270^o, i.e. from the line where the antenna elements are placed. However, we normally consider only half of the plane, i.e. from $\theta_a = 0^o$ to $\theta_a = 180^o$ [117], where θ_a denotes the angle measured from the line connecting the antenna array elements, which is given by $\theta = 90^o - \theta_a$. Furthermore, the symmetry

of the radiation characteristic follows from Equation (2.26), which produces identical responses for angles of $\pm\theta_a$ for $0^o \leq \theta_a^o \leq 180^o$.

Naturally, in practical situations we have no prior knowledge concerning the direction of arrival and many other factors should be taken into consideration such as the contaminating effects of noise, the characteristics of the received signals and the nature of the propagation environment. Next, we elaborate on some of the criteria often used for calculating the optimal weights.

2.3 Criteria for Determining the Optimal Beamforming Weights

An antenna array's output is constituted by the weighted combination of signals induced in the array elements. The set of array weights used, when the best performance resulting in a beam having a sharp directivity in the desired direction and a null in the direction of the interferers is achieved is referred to as the optimal weights. Several performance measures, such as the maximum desired signal power, maximum SNR, maximum SIR and minimum interference power may be used, but essentially the objective is to optimise the beamformer's response with respect to a prescribed optimisation criterion, such as the Minimum Mean Square Error (MMSE) between the array's actual and desired output so that the output of the beamformer contains the lowest possible contribution from noise and interference.

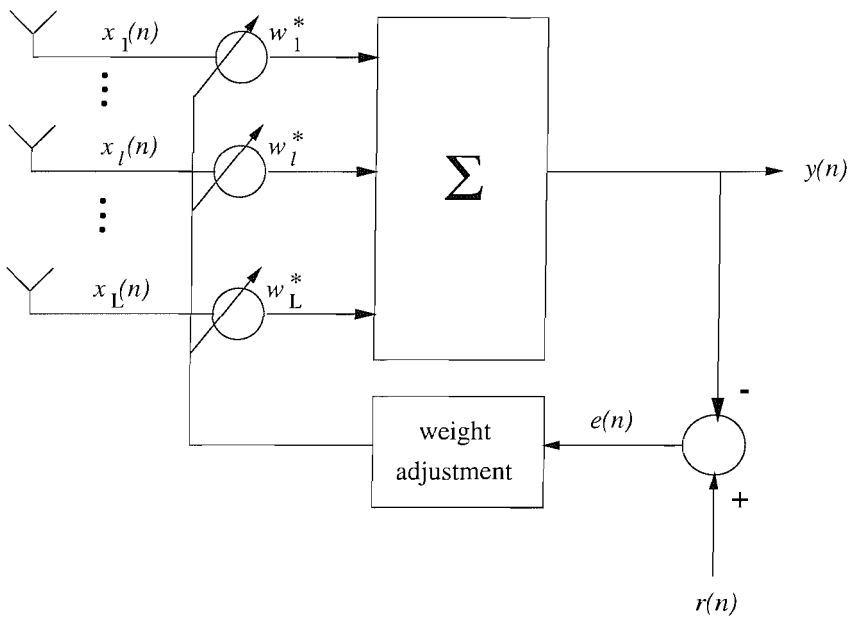


Figure 2.11: Structure of a beamformer using the reference signal $r(n)$.

The general optimisation problem to be solved here is centred around the schematic of Figure 2.11, where the array weights are optimised for the sake of minimising for example the variance of the error signal $e(n)$ between the array output $y(n)$ and the reference signal $r(n)$. The reference signal $r(n)$ is

typically a unique sequence, which unambiguously identifies the desired user and may be considered equivalent to the known channel sounding sequence used in the context of channel equalisation for optimising the equaliser coefficients, in order to minimise the effects of intersymbol interference [2].

In order to simplify our notations, in this section we denote the desired signal vector by $\mathbf{d}$ (assuming that the first signal $\bar{m}_1(n)$ of Equation (2.7) having a steering vector $\mathbf{s}_1$ described by Equation (2.8) is the desired user), the noise vector by $\mathbf{n}$ and the sum of all interfering signal vectors by $\mathbf{u}$ (assuming the presence of $M-1$ interfering signals $\bar{m}_m(n)$ formulated in Equation (2.7) with the corresponding steering vector $\mathbf{s}_m$ given by Equation (2.8), where we have $m = 2, 3, ..., M$).

2.3.1 Minimum Mean Square Error Criterion

As in Equation (2.16), the array output $\mathbf{x}(n)$ can be expressed as:

$$\mathbf{x}(n) = \bar{m}_1(n)\mathbf{s}_1 + \sum_{i=2}^{M} \bar{m}_i(n)\mathbf{s}_i + \mathbf{n}(n) = \mathbf{d}(n) + \mathbf{u}(n) + \mathbf{n}(n). \tag{2.40}$$

As seen in Figure 2.11, the Minimum Mean Square Error (MMSE) criterion aims for minimising the Mean Square Error (MSE) between the beamformer's output $y(n)$ and the desired response $r(n)$. Ideally, the reference signal, which is also referred to as the desired response $r(n)$ is assumed to be highly correlated with the desired signal $\bar{m}_1(n)$ and uncorrelated with all unwanted interfering signals $\bar{m}_m(n)$, resulting in $E[r(n)\bar{m}_m(n)] = 0$, where $m = 2, 3, \ldots, M$ is the index of an unwanted signal. Since the array output consists of the superposition of all incoming signals and random noise, the difference between the beamformer's output and the reference signal will create an error signal $e(n)$ expressed as:

$$e(n) = r(n) - y(n) = r(n) - \mathbf{w}^H\mathbf{x}(n), \tag{2.41}$$

which is used for adjusting the weights $\mathbf{w}$, such that the MSE between the array output $y(n)$ and the reference signal $r(n)$ is minimised. The MSE is obtained by squaring and expanding Equation (2.41) and then taking the ensemble average, resulting in:

$$\begin{aligned} \text{MSE} &= E\left[|e(n)|^2\right] \\ &= E\left[|r(n)|^2\right] - 2\mathbf{w}^H\mathbf{z} + \mathbf{w}^H\mathbf{R}\mathbf{w}, \end{aligned} \tag{2.42}$$

where $\mathbf{z} = E[r^*(n)\mathbf{x}(n)]$ is the cross-correlation between the array output and the complex conjugate of the reference signal, while $\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^H(n)]$ is the auto-correlation of the array output.

Observe in Equation (2.42) that the MSE expression defined is a quadratic function of $\mathbf{w}$ and hence it may be minimised by taking its derivative with respect to the array weights $\mathbf{w}$ and then setting it to zero, as follows:

$$\nabla_{\mathbf{w}} E\left[|e(n)|^2\right] = -2\mathbf{z} + 2\mathbf{R}\mathbf{w} = 0, \tag{2.43}$$

yielding the Wiener-Hopf or optimum Wiener solution of:

$$\mathbf{w}_{opt} = \mathbf{R}^{-1}\mathbf{z}. \tag{2.44}$$

Under ideal conditions, when the reference signal is equal to the desired signal, i.e. when we have $\bar{m}_1(n) = r(n)$, we may define the cross-correlation vector of $\mathbf{z} = E[|r(n)|^2]\mathbf{s}_1$ and the auto-correlation matrix of $\mathbf{R} = E[|r(n)|^2]\mathbf{s}_1\mathbf{s}_1^H + \mathbf{R}_u + 2\sigma_n^2\mathbf{I}$, where $\mathbf{R}_u = E[\mathbf{u}\mathbf{u}^H]$ and $\mathbf{I}$ denotes an identity matrix. According to Equation (2.44) it is necessary to invert the matrix $\mathbf{R}$ every time, when the set of optimum array weights has to be updated, which is associated with a complexity proportional to the cube of the number of array weights. Fortunately, however, it is possible to directly update $\mathbf{R}^{-1}$ without performing matrix inversion, by applying the Matrix Inversion Lemma [107, 166]:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{D}\mathbf{A}^{-1} \tag{2.45}$$

to $\mathbf{R}$, where $\mathbf{A} = \mathbf{R}_{u+n} = \mathbf{R}_u + 2\sigma_n^2\mathbf{I}$, $\mathbf{B} = \mathbf{s}_1$, $\mathbf{C} = E[|r(n)|^2]$ and $\mathbf{D} = \mathbf{s}_1^H$. It follows that the inverse of $\mathbf{R}$ can be regularly updated according to:

$$\mathbf{R}^{-1} = \left[\frac{1}{1 + E[|r(n)|^2]\mathbf{s}_1^H\mathbf{R}_{u+n}^{-1}\mathbf{s}_1}\right]\mathbf{R}_{u+n}^{-1}, \tag{2.46}$$

which leads to the generalised optimum Wiener solution expressed as [8]:

$$\mathbf{w}_{opt} = \beta\mathbf{R}_{u+n}^{-1}\mathbf{s}_1, \tag{2.47}$$

where the scalar coefficient β is defined as [8]:

$$\beta = \frac{E[|r(n)|^2]}{1 + E[|r(n)|^2]\mathbf{s}_1^H\mathbf{R}_{u+n}^{-1}\mathbf{s}_1}. \tag{2.48}$$

Note that the Matrix Inversian Lemma is also referred to in the literature as Woodbury's Identity [6,8]. Following the above brief discourse on the MMSE array weight optimisation criterion, let us now consider the maximisation of the SIR in the next section.

2.3.2 Maximum Signal-to-Interference Ratio Criterion

The array weights can also be chosen to directly maximise the Signal-to-Interference Ratio (SIR) at the beamformer's output. Assuming that the M arriving signals; $\bar{m}_1(n), \bar{m}_2(n), \bar{m}_3(n), ..., \bar{m}_M(n)$ are uncorrelated and the correlation matrices of the desired signal $\mathbf{R}_d = E[\mathbf{d}\mathbf{d}^H]$ as well as that of the undesired signals $\mathbf{R}_u = E[\mathbf{u}\mathbf{u}^H]$ are known, we may choose to maximise the ratio of the desired signal power σ_d^2 and the total interfering signal power σ_u^2. With the mean output power given by $\sigma^2 = E[y(n)y^*(n)]$ and the array output formulated by $y(n) = \mathbf{w}^H\mathbf{x}(n)$, where we have $\mathbf{x}(n) = \mathbf{d}(n) + \mathbf{u}(n) + \mathbf{n}(n)$, the desired signal power may be written as:

$$\begin{aligned}\sigma_d^2 &= E[|\mathbf{w}^H\mathbf{d}|^2] \\ &= \mathbf{w}^H\mathbf{R}_d\mathbf{w}\end{aligned} \tag{2.49}$$

and the interfering signals power as:

$$\begin{aligned} \sigma_u^2 &= E[|\mathbf{w}^H\mathbf{u}|^2] \\ &= \mathbf{w}^H\mathbf{R}_u\mathbf{w}. \end{aligned} \tag{2.50}$$

Therefore, the SIR is given by:

$$\begin{aligned} \text{SIR} &= \frac{\sigma_d^2}{\sigma_u^2} \\ &= \frac{\mathbf{w}^H\mathbf{R}_d\mathbf{w}}{\mathbf{w}^H\mathbf{R}_u\mathbf{w}}. \end{aligned} \tag{2.51}$$

Upon setting the derivative of Equation (2.51) with respect to $\mathbf{w}$ to zero, one arrives at:

$$\mathbf{R}_d\mathbf{w} = \frac{\mathbf{w}^H\mathbf{R}_d\mathbf{w}}{\mathbf{w}^H\mathbf{R}_u\mathbf{w}}\mathbf{R}_u\mathbf{w}, \tag{2.52}$$

which has to be solved for the weights $\mathbf{w}$. Having defined the desired signal vector $\mathbf{d}(n)$ as $\mathbf{d}(n) = \bar{m}_1(n)\mathbf{s}_1$ in Equation (2.40), we have $\mathbf{R}_d = E[|r(n)|^2]\mathbf{s}_1\mathbf{s}_1^H$, thus we obtain:

$$\mathbf{w}_{opt} = \beta\mathbf{R}_u^{-1}\mathbf{s}_1, \tag{2.53}$$

where in this case, the scalar coefficient β is expressed as [8]:

$$\beta = \frac{E[|r(n)|^2]}{\text{SIR}}\mathbf{s}_1^H\mathbf{w}_{opt}. \tag{2.54}$$

In order to directly maximise the SINR, rather than the SIR, we must include the noise power in the SINR calculation such that the expression of:

$$\text{SINR} = \frac{\mathbf{w}^H\mathbf{R}_d\mathbf{w}}{\mathbf{w}^H\mathbf{R}_{u+n}\mathbf{w}}. \tag{2.55}$$

replaces the SIR calculation of Equation (2.51). Having described the maximum SIR/SINR array weight optimisation criterion, we now proceed to another common criterion referred to as the minimum variance optimisation criterion.

2.3.3 Minimum Variance Criterion

The utilisation of the reference signal based approach to the optimisation of the array weights is prevented, when too little statistical knowledge is available concerning the desired signal $\bar{m}_1(n)$ [88]. The desired signal maybe of unknown power and may not always be present, thus preventing the estimation of the signal and correlation matrices required by the maximum SNR beamformer. An attractive way of ensuring a good signal reception may be arrived at by the application of linear constraints to the array weight vector and then minimising the output noise variance (i.e. average power). Recall that the beamformer's output in Figure 2.11 is given by:

$$\begin{aligned} y(n) &= \mathbf{w}^H\mathbf{x}(n) \\ &= \mathbf{w}^H\mathbf{d}(n) + \mathbf{w}^H\mathbf{u}(n) + \mathbf{w}^H\mathbf{n}(n). \end{aligned} \tag{2.56}$$

In order to ensure that the desired signal is received with a specific gain and phase, a constraint may be used so that the response of the beamformer to the desired signal becomes [8]:

$$\mathbf{w}^H\mathbf{s}_1 = g, \tag{2.57}$$

where $\mathbf{s}_1$ is the desired user's steering vector and g is a complex constant. The employment of this linear constraint constitutes a general approach that facilitates maintaining an extensive control over the response of the beamformer [88]. Multiple linear constraints may also be imposed for the sake of maintaining more control over the beam pattern. For example, if there is a fixed interference source at a known direction θ_u, then it is possible to enforce maintaining a zero gain in that direction in addition to maintaining the response g towards the desired signal arriving from direction θ_d. This is expressed as:

$$\begin{bmatrix} \mathbf{s}_1^H(\theta_d) \\ \mathbf{s}_1^H(\theta_u) \end{bmatrix} \mathbf{w} = \begin{bmatrix} g \\ 0 \end{bmatrix}. \tag{2.58}$$

In other words, multiple linear constraints can be imposed for the sake of handling any uncertainty concerning the desired signal's direction of arrival and to provide a degree of robustness to signal cancellation due to signal direction mismatch. In [68], multiple constraints were used for ensuring that the beamformer response complies with given specifications at certain spatial or temporal frequencies. The number of points at which the beamformer's response can be constrained is however limited to the number of array elements L, since each constraint exhausts one degree of freedom (DOF) in the weight vector [88]. Therefore, given N constraints such that $N < L$, there are only $(L - N)$ DOF left for minimising the error signal's variance. If L constraints are used, then there are no DOF available for power minimisation and therefore the beamformer's response will be fixed, i.e. remains independent of the array's received signal for all signal and interference scenario. This particular case is analogous to that of the data-independent beamformer [88] or switched beam approach [13, 20, 21].

Minimisation of the contributions of the interferers to the array output is achieved by choosing the array weights to minimise the output variance or power, which is formulated as:

$$\begin{aligned} \mathrm{Var}[y] &= \mathbf{w}^H\mathbf{R}\mathbf{w} \\ &= \mathbf{w}^H\mathbf{R}_d\mathbf{w} + \mathbf{w}^H\mathbf{R}_{u+n}\mathbf{w} \end{aligned} \tag{2.59}$$

subject to the constraint defined in Equation (2.57). This is equivalent to minimising the quantity $\mathbf{w}^H\mathbf{R}_{u+n}\mathbf{w}$, where $\mathbf{R}_{u+n} = E[\mathbf{u}\mathbf{u}^H] + 2\sigma_n^2\mathbf{I}$. Using the method of Lagrange, we have [8, 56, 88]:

$$\nabla_{\mathbf{w}}\left(\frac{1}{2}\mathbf{w}^H\mathbf{R}_{u+n}\mathbf{w} + \beta[1 - \mathbf{w}^H\mathbf{s}_1]\right) = \mathbf{R}_{u+n}\mathbf{w} - \beta\mathbf{s}_1, \tag{2.60}$$

and upon setting this expression to zero, we arrive at:

$$\mathbf{w}_{opt} = \beta\mathbf{R}_{u+n}^{-1}\mathbf{s}_1, \tag{2.61}$$

where the scalar coefficient β represents [8]:

$$\beta = \frac{g}{\mathbf{s}_1^H \mathbf{R}_{u+n}^{-1} \mathbf{s}_1}. \tag{2.62}$$

The beamformer employing this particular optimisation criterion is referred to as the Linearly Constrained Minimum Variance (LCMV) [58] scheme. For the specific case, when the complex constant in Equation (2.57) is set to $g = 1$, the response of the beamformer is often termed the Minimum Variance Distortionless Response (MVDR) [67], a beamforming scheme, which was pioneered by Capon [56]. Having discussed the three classic array weight optimisation criteria, in the next section we will compare and summarise them.

2.3.4 Comparison Between Beamforming Optimisation Criteria

When invoking the MMSE criterion, the direction of the desired signal does not have to be known. However, this technique requires the transmission of a known reference signal that is highly correlated with the desired signal and therefore the statistics of the desired signal are required for the design of the reference signal. This also applies to the maximum SIR criterion, where the correlation matrices of the desired and undesired signals, namely $\mathbf{R}_d$ and $\mathbf{R}_u$ have to be known for the sake of determining the array weights $\mathbf{w}$. As for the LCMV criterion of Section 2.3.3, it guarantees that the desired signal will not be cancelled with the advent of applying linear constraints to the array weight vector and therefore it requires the knowledge of the steering vector $\mathbf{s}_1$. Satisfying the above requirements will lead to optimum beamforming in the sense of the specific optimisation criterion concerned, provided that the required knowledge is accurate and that the signals impinging on the array are uncorrelated.

Having determined the optimum array weights from Equation (2.52) with $\mathbf{R}_{u+n} = \mathbf{R}_u + 2\sigma_n^2\mathbf{I}$ replacing the correlation matrix of the interfering signals $\mathbf{R}_u$, the SINR can also be determined from Equation (2.55), yielding:

$$\begin{aligned} \text{SINR} &= \frac{\mathbf{w}_{opt}^H \mathbf{R}_d \mathbf{w}_{opt}}{\mathbf{w}_{opt}^H \mathbf{R}_{u+n} \mathbf{w}_{opt}} \\ &= \frac{\beta^2 E[|r(n)|^2] \mathbf{s}_1^H \mathbf{R}_{u+n}^{-1} \mathbf{s}_1 \mathbf{s}_1^H \mathbf{R}_{u+n}^{-1} \mathbf{s}_1}{\beta^2 \mathbf{s}_1^H \mathbf{R}_{u+n}^{-1} \mathbf{s}_1} \\ &= E[|r(n)|^2] \mathbf{s}_1^H \mathbf{R}_{u+n}^{-1} \mathbf{s}_1, \end{aligned} \tag{2.63}$$

which is independent of β [8, 88]. It is interesting to note from Equations (2.47), (2.53) and (2.61), that the above three criteria used for calculating the optimal array weights resulted in an optimum weight expression of the same form [8]. Below we contrast the corresponding expressions of the scalar coefficient β in terms of the SINR for the above three optimisation criteria introduced, which are

derived from Equations (2.48), (2.54) and (2.62):

$$\beta_{\text{MSE}} = \frac{E[|r(n)|^2]}{1+\text{SINR}} \tag{2.64}$$

$$\beta_{\text{SINR}} = \frac{E[|r(n)|^2]}{\text{SINR}} \mathbf{s}_1^H \mathbf{w}_{opt} \tag{2.65}$$

$$\beta_{\text{MV}} = g\frac{E[|r(n)|^2]}{\text{SINR}} \tag{2.66}$$

$$= \frac{E[|r(n)|^2]}{\text{SINR}} \mathbf{w}_{opt}^H \mathbf{s}_1. \tag{2.67}$$

Note that, β_{MV} is identical to β_{SINR}, while for the MSE criterion, we infer from Equation (2.67) that the constant gain β_{MSE} can be expressed in terms of β_{MV} as:

$$\beta_{\text{MSE}} = \frac{E[|r(n)|^2]\beta_{\text{MV}}}{gE[|r(n)|^2]+\beta_{\text{MV}}}. \tag{2.68}$$

The optimum array weights of the three different optimisation criteria described are all given by the Wiener solution of Equation (2.44), showing the importance of the Wiener-Hopf equation in establishing the theoretical adaptive beamforming steady-state performance limits [8]. This also implies that the choice of the optimisation criterion is not as critical as that of an appropriate adaptive array weight update algorithm, which will be discussed in the following section. As a summary, Table 2.5 compares the three criteria discussed.

Criterion	Minimum MSE (Section 2.3.1)	Maximum SIR (Section 2.3.2)	Minimum Variance (Section 2.3.3)
Parameters	$\mathbf{x} = \mathbf{d} + \mathbf{u} + \mathbf{n}$ $\mathbf{x}$: array output r: reference signal $\mathbf{z} = E[r^*\mathbf{x}]$ $\mathbf{R} = E[\mathbf{x}\mathbf{x}^H]$ (2.21) output: $y = \mathbf{w}^H\mathbf{x}$ (2.41)	$\mathbf{x} = \mathbf{d} + \mathbf{u} + \mathbf{n}$ $\mathbf{x}$: array output $\mathbf{d}$: signal component $\mathbf{u}$: interference component $\mathbf{R}_d = E[\mathbf{d}\mathbf{d}^H]$ (2.49) $\mathbf{R}_u = E[\mathbf{u}\mathbf{u}^H]$ (2.50) output: $y = \mathbf{w}^H\mathbf{x}$	$\mathbf{x} = \mathbf{d} + \mathbf{u} + \mathbf{n}$ $\mathbf{x}$: array output $\mathbf{s}_1$: constraint vector g: response constant $\mathbf{R} = E[\mathbf{x}\mathbf{x}^H]$ (2.21) output: $y = \mathbf{w}^H\mathbf{x}$ (2.56)
Method	$\min_{\mathbf{w}} E\left[\lvert r-y\rvert^2\right]$ (2.42)	$\max_{\mathbf{w}} \frac{\mathbf{w}^H\mathbf{R}_d\mathbf{w}}{\mathbf{w}^H\mathbf{R}_u\mathbf{w}}$ (2.51)	$\min_{\mathbf{w}} \mathbf{w}^H\mathbf{R}\mathbf{w}$ (2.59) subject to $\mathbf{w}^H\mathbf{s}_1 = g$ (2.57)
Optimum Weights	$\mathbf{w} = \mathbf{R}^{-1}\mathbf{z}$ (2.44)	$\mathbf{R}_s\mathbf{w} = \frac{\mathbf{w}^H\mathbf{R}_d\mathbf{w}}{\mathbf{w}^H\mathbf{R}_u\mathbf{w}}\mathbf{R}_u\mathbf{w}$ (2.52)	$\mathbf{w} = \mathbf{R}_{u+n}^{-1}\mathbf{s}_1[\mathbf{s}_1^H\mathbf{R}_{u+n}^{-1}\mathbf{s}_1]^{-1}g$ (2.61)
Advantages	Direction of desired signals not required	True maximisation of SNR	Flexible and general constraints
Dis-advantages	Requires a reference signal	Requires the knowledge of $\mathbf{R}_d$ and $\mathbf{R}_u$	Computation of constrained weight vector is required

Table 2.5: Summary of the array weight optimisation criteria; MMSE, maximum SIR and minimum variance.

All the three criteria require a reference signal in their adaptive optimisation process. The choice

of a reference signal usually requires explicit *a priori* knowledge about the statistics of the signal of interest. The reference signal can be classified into two distinct categories, namely spatial and temporal reference classes [8]. The design option of using a spatial reference is often referred to as the employment of direction of arrival (DOA) information concerning the desired signal. By contrast, a temporal reference signal may be a pilot signal transmitted for example in each transmission burst that is correlated with the desired signal, a specific bit sequence embedded in a data packet or a known pseudo-noise (PN) code in a CDMA system. The form of reference signal used depends on the particular system, in which adaptive beamforming is to be implemented. If an explicit reference signal is available in a system, it should be used as efficiently as possible for the sake of achieving the lowest possible complexity, high accuracy and fast convergence.

In applications involving multipath propagation, it is not uncommon to experience multipath interference that is coherent with the signal of interest. The utilisation of an LCMV adaptive beamformer in conjunction with one or several constraints will not only suppress the interferers but also the desired signal. An example of preventing the cancellation of the desired signal in the presence of coherent interference has been proposed by Sivaradje *et al.* [167] using three methods known as spatial smoothing, parallel spatial processing and the so-called split-polarity transformation technique.

Having characterised three different array-weight optimisation criteria of beamforming, namely the MMSE, the maximum SIR and the minimum variance criteria, let us now move on to considering the most important aspects of adaptive beamforming, namely the choice of the adaptive array-weight update algorithms, which play a major role in determining the behaviour of the antenna array.

2.4 Adaptive Beamforming Algorithms

The array weight optimisation criteria described in the previous section fall into the statistically optimum beamforming category, where the weights are chosen based on the statistics of the signal received by the array. However, the signal statistics are often unknown and they may change over time, for example, as a consequence of the time-variant interferers [88]. Therefore, typically adaptive algorithms are used for obtaining a number of weights that will converge to the statistically optimum solution.

2.4.1 Temporal Reference Technique

Recall from our earlier discourse that Figure 2.11 shows an adaptive beamformer employing the temporal reference based array weight optimisation technique. The aim of the weight adjustment process is to find a set of weights that will allow the output $y(n)$ of the array at each time instant to be equal to or as close as possible to the desired response $r(n)$, i.e. the error signal should be rendered statistically as small as possible. The reference signal $r(n)$ is generated based on the available information related

to the statistical characteristics of the desired signal $\bar{m}_1(n)$. For example, in a CDMA system $r(n)$ may simply be the PN sequence of the intended user.

In adapting the weighting coefficients, this temporal reference technique may apply the MMSE criterion, which was derived in Section 2.3.1. Minimisation of the corresponding MSE has resulted in the well-known Wiener-Hopf equation, restated here for convenience as:

$$\mathbf{w}_{opt} = \mathbf{R}^{-1}\mathbf{z}. \tag{2.69}$$

The resultant MMSE of the beamformer constituted by the classic Wiener filter, using the optimum weights may be derived from Equation (2.42), which is given by:

$$\text{MMSE} = E\left[|r(n)|^2\right] - \mathbf{z}^H\mathbf{R}^{-1}\mathbf{z}. \tag{2.70}$$

The temporal reference model may invoke a variety of adaptive array weight update algorithms, such as the Least Mean Square (LMS), the Recursive Least Square (RLS) and the Sample Matrix Inversion (SMI) algorithms to be outlined in Sections 2.4.1.1, 2.4.1.2 and 2.4.1.3, respectively.

2.4.1.1 Least Mean Square Beamforming Algorithm

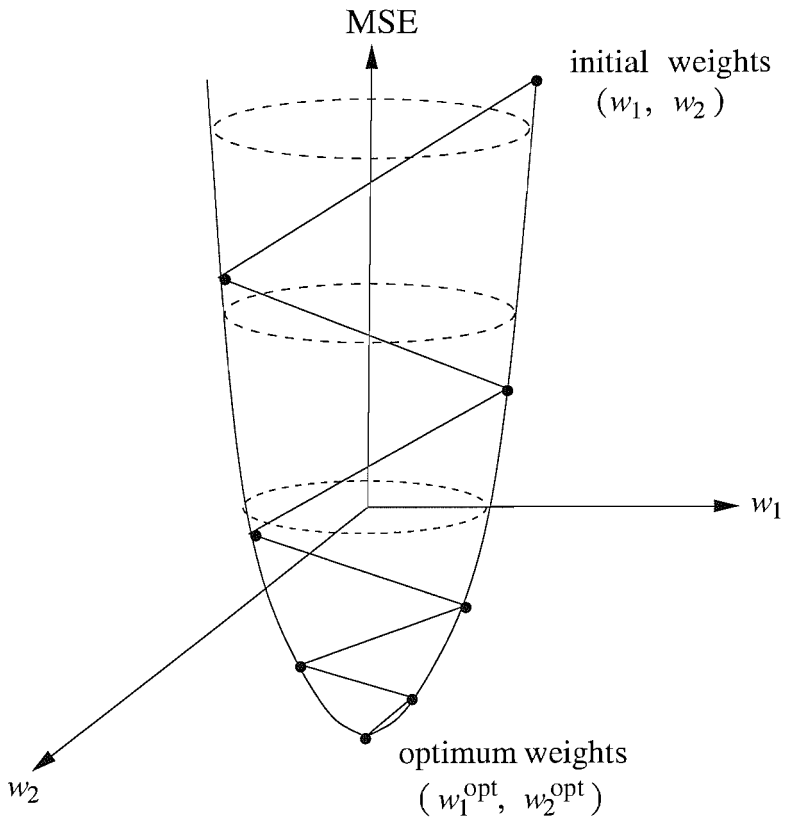


Figure 2.12: An example of the MSE surface and the weights following the negative direction of the gradient to minimise the MSE.

The well-known LMS algorithm falls into the family of stochastic gradient-based techniques, since it is based on the steepest descent method that operates by iteratively computing and updating the weight vector. The updated value of the weight vector at time instant $n+1$ based on the steepest descent method is computed using the relation of [4, 6, 8, 166]:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\left[-\nabla_{\mathbf{w}}\mathrm{MSE}(n)\right], \tag{2.71}$$

where n indicates the discrete time index, μ is a positive real-valued constant referred to as the step size and $\mathbf{w}(n)$ is the current weight vector at time index n. The simple and plausible philosophy of Equation (2.71) is that the array weights are regularly updated according to the step size μ in the direction of the negative gradient of the cost function $\nabla_{\mathbf{w}}\mathrm{MSE}(n)$, as defined in Equation (2.43), in order to yield an updated coefficient vector $\mathbf{w}(n+1)$. In other words, the algorithm follows the negative gradient of the MSE cost function step by step and this process will eventually lead to a unique global minimum, namely to a point where the weight vector assumes its optimum value [166], as shown in Figure 2.12. The initial weight values of $\mathbf{w}(0)$ is typically, arbitrarily chosen, but commonly it is set to be a null vector. Equation (2.71) thus describes the formulation of the steepest descent algorithm.

Practically, an exact measurement of the gradient vector is not possible [6, 54, 64], since no prior knowledge is available concerning the auto-correlation matrix $\mathbf{R}$ and the cross-correlation vector $\mathbf{z}$ before the commencement of communications. Therefore, the true gradient is replaced by a short-term estimate based on the previous samples of the input signal $\mathbf{x}(n)$ and the reference signal $r(n)$. The estimated auto-correlation $\hat{\mathbf{R}}$ and cross-correlation vector $\hat{\mathbf{z}}$ based on a single sample are defined, respectively, as:

$$\hat{\mathbf{R}}(n) = \mathbf{x}(n)\mathbf{x}^H(n) \tag{2.72}$$

$$\hat{\mathbf{z}}(n) = r^*(n)\mathbf{x}(n). \tag{2.73}$$

Note that we used a hat, '$\hat{}$' over the variables to indicate an estimated value, in order to distinguish them from their exact value. The estimated gradient is obtained by inserting these estimated statistics into Equation (2.43), which is then substituted into Equation (2.71) in order to yield the LMS array weight update equation in the form of:

$$\begin{aligned}\hat{\mathbf{w}}(n+1) &= \hat{\mathbf{w}}(n) + \mu[\hat{\mathbf{z}} - \hat{\mathbf{R}}\hat{\mathbf{w}}(n)] \\ &= \hat{\mathbf{w}}(n) + \mu\mathbf{x}(n)[r^*(n) - \mathbf{x}^H(n)\hat{\mathbf{w}}(n)] \\ &= \hat{\mathbf{w}}(n) + \mu\mathbf{x}(n)e^*(n),\end{aligned} \tag{2.74}$$

where, as noted previously, the constant μ is the step size that controls the convergence characteristics of the weight vector update algorithm. The procedure of updating the array weights with the aid of the LMS algorithm is summarised in Table 2.6. The LMS algorithm requires only $2L+1$ complex multiplications and $2L$ complex additions per array weight update iteration, where L is the number of elements in the antenna array.

1:	$y(n) = \hat{\mathbf{w}}^H(n)\mathbf{x}(n)$
2:	$e(n) = r(n) - y(n)$
3:	$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu\mathbf{x}(n)e^*(n)$

Table 2.6: Updating the array weight vector of the LMS algorithm [166].

In order for the estimated array weights to approach the optimal weights, the step size μ should be sufficiently small. However, this renders the convergence rate slow, which is the major disadvantage of the LMS algorithm. Increasing μ will facilitate a faster convergence, however, it will result in a higher residual MSE due to using non-optimal weights. For the sake of maintaining algorithmic stability, the step size μ must satisfy the following condition [6, 7, 168]:

$$0 < \mu < \frac{2}{\lambda_{max}} , \tag{2.75}$$

where λ_{max} is the largest eigenvalue of the correlation matrix $\mathbf{R}$. This constraint can be simplified by using the trace of $\mathbf{R}$, which is easier to determine than the value of λ_{max}, as the array's dimension becomes larger, resulting in [4, 6, 8]:

$$0 < \mu < \frac{2}{tr[\mathbf{R}]} , \tag{2.76}$$

where trace of $\mathbf{R}$, $tr[\mathbf{R}]$ is given by the sum of the diagonal elements of the correlation matrix $\mathbf{R}$ expressed as $tr[\mathbf{R}] = \sum_{i=1}^{L} \lambda_i$ for L antenna elements. Physically this quantity represents the sum of the input powers of the array elements.

The convergence rate of the algorithm is also controlled by the eigenvalue spread [6, 169, 170]:

$$\chi(\mathbf{R}) = \frac{\lambda_{max}}{\lambda_{min}} , \tag{2.77}$$

where $\chi(\mathbf{R}) \geq 1$. Rapid convergence will be attained only, when $|1 - \mu\lambda_i|$ is small ($\ll 1$), which is achieved, when either μ or $\lambda_i, i = 1, 2, \ldots, L$ or both are relatively high [6, 64, 168]. However, this desirable condition cannot be readily achieved, when the step size is close to the stability upper bound of $\simeq \frac{2}{\lambda_{max}}$, which is likely to occur, when $\chi(\mathbf{R}) = \frac{\lambda_{max}}{\lambda_{min}} \gg 1$. In other words, the convergence will now depend on the smallest eigenvalue λ_{min}. Therefore, the fastest convergence is achieved, when $\lambda_{max} = \lambda_{min}$ and as the eigenvalue spread increases, the convergence rate becomes slower. Experiments conducted by Haykin [6] have demonstrated that the LMS algorithm is highly sensitive to variations of the step size μ and to the eigenvalue spread $\chi(\mathbf{R})$ of the correlation matrix of the array output vectors. Briefly, the value of λ_{max} will determine the excess MSE, while that of λ_{min} is expected to limit the convergence of the algorithm, as the eigenvalue spread increases.

Blogh *et al.* [4] and Hanzo *et al.* [12] showed the performance dependence of the LMS algorithm upon the desired signal's power, when aiming for an adequate level of interference rejection. In conjunction with a step size μ as small as 0.0000005, the convergence rate remained low even with a

reference length of 1024 bits [4]. A faster convergence was observed, when the step size was increased to $\mu = 0.0005$ [4] although a sudden decrease of the achievable SIR level was observed, when the signal power exceeded 20dB. This phenomenon is likely to be related to the fact that the adaptation of the array weight vector according to the factor of $\mu\mathbf{x}(n)e^*(n)$ in Equation (2.74) is directly proportional to the array output vector $\mathbf{x}(n)$. Explicitly, if $\mathbf{x}(n)$ is large, the LMS algorithm is expected to experience a noise amplification problem, since $e^*(n)$ is multiplied by $\mathbf{x}(n)$, which prevents the array weight vector from reaching the vicinity of its optimum value. This observation motivated the design of one of the variants of the LMS algorithm, which is known as the Normalised LMS (NLMS) algorithm.

The NLMS algorithm uses a signal-dependent step size in each iteration [81], since it normalises the correction vector of $\mu\mathbf{x}(n)e^*(n)$ in Equation (2.74) with respect to the squared magnitude of the array output vector $\mathbf{x}(n)$, resulting in a step size expressed in the form of [4, 8]:

$$\mu(n) = \frac{\mu_o}{a + \mathbf{x}^H(n)\mathbf{x}(n)} = \frac{\mu_o}{a + \|\mathbf{x}(n)\|^2} , \tag{2.78}$$

where both μ_o and a is a positive constant. The algorithm converges in the mean squared sense, if we have $0 < \mu_o < 2$ [6, 168]. The parameter a is chosen to be higher than zero [4, 6, 168] for low values of the array output vector $\mathbf{x}(n)$ in order to avoid division by a small number. The NLMS algorithm is therefore less sensitive to the power fluctuations of the input signals. Hence the algorithm's stability is improved and its performance in terms of the achievable convergence rate is far better than that of the LMS algorithm, as shown in [4] and [12]. In terms of complexity, the NLMS algorithm requires a slightly higher number of complex-valued mathematical operations, than the LMS technique, namely $3L + 2$ complex multiplications and $3L$ complex additions.

Next, we will elaborate on another classic adaptive algorithm known as the Recursive Least Square (RLS) algorithm.

2.4.1.2 Recursive Least Square Beamforming Algorithm

The philosophy of the Recursive Least Square (RLS) algorithm is based on the least squares principle, directly optimising the weight vector by minimising the sum of squared errors of all previous samples up to that of the current time index n. Mathematically, the Least Squares Error (LSE) cost function is written as [166]:

$$\text{LSE} = \sum_{n=0}^{N} \alpha^n e(N-n)e^*(N-n) , \tag{2.79}$$

where the error e is the difference between the actual array output y and the desired response r, as defined in Equation (2.41), while α represents a weighting factor having a value spanning the range of $0 < \alpha \leq 1$. The weighting factor α is often referred to as the forgetting factor due to its function of de-emphasising past error contributions, because it determines how rapidly the previous data bits are 'forgotten'. Furthermore, the term $(1 - \alpha)$ is an approximate measure of the algorithm's memory,

thus the special case of $\alpha = 1$ corresponds to having an infinite memory. The choice of the forgetting factor α is strongly dependent upon the fading rate in the channel [7]. Commonly its value is chosen to be higher than 0.8 [7]. For a rapidly fading channel, α should be slightly less than unity, and the choice of $\alpha = 0.95$ was reported in [171] to be a reasonable value. For a stationary propagation environment the forgetting factor α is set to unity, implying that all previous data are taken into account for determining the beamforming algorithm's updated parameters.

Similar to the case of employing the LMS algorithm for minimising the MSE, the minimisation of the LSE criterion is also achieved by taking its derivative with respect to the array weights $\mathbf{w}$ and then setting it to zero, yielding the Wiener-Hopf solution of Equation (2.69). Based on the weighted sum, the estimated auto-correlation matrix $\hat{\mathbf{R}}$ and the estimated cross-correlation vector $\hat{\mathbf{z}}$ are defined, respectively, as:

$$\hat{\mathbf{R}}(n) = \sum_{n=0}^{N} \alpha^n \mathbf{x}(N-n)\mathbf{x}^H(N-n) \tag{2.80}$$

and

$$\hat{\mathbf{z}}(n) = \sum_{n=0}^{N} \alpha^n r^*(N-n)\mathbf{x}(N-n) \ , \tag{2.81}$$

where r is the reference signal and $\mathbf{x}$ is the array output.

The RLS algorithm aims for updating the weight vector $\mathbf{w}(n+1)$ based on the knowledge of the previous variables, namely on $\mathbf{w}(n)$, $\mathbf{R}(n-1)$ and $\mathbf{z}(n-1)$. This is based on recursively updating the estimates of $\mathbf{R}(n)$ and $\mathbf{z}(n)$ in Equations (2.80) and (2.81), respectively, by applying:

$$\hat{\mathbf{R}}(n) = \alpha\hat{\mathbf{R}}(n-1) + \mathbf{x}(n)\mathbf{x}^H(n) \tag{2.82}$$

$$\hat{\mathbf{z}}(n) = \alpha\hat{\mathbf{z}}(n-1) + r^*(n)\mathbf{x}(n) \tag{2.83}$$

and solving $\hat{\mathbf{R}}(n)\hat{\mathbf{w}}(n) = \hat{\mathbf{z}}(n)$ for each time index n. For the optimal Wiener-Hopf solution of Equation (2.69), we have to compute $\hat{\mathbf{R}}^{-1}(n)$. The inverse of $\hat{\mathbf{R}}(n)$ can be determined iteratively by applying the Matrix Inversion Lemma [166] of Equation (2.45) to Equation (2.82) using $\mathbf{A} = \alpha\hat{\mathbf{R}}(n-1)$, $\mathbf{B} = \mathbf{x}(n)$, $\mathbf{C} = 1$ and $\mathbf{D} = \mathbf{x}^H(n)$, yielding:

$$\hat{\mathbf{R}}^{-1}(n) = \frac{1}{\alpha}\left[\hat{\mathbf{R}}^{-1}(n-1) - \frac{\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)}{\alpha + \mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)}\right], \tag{2.84}$$

with the initial inverse auto-correlation matrix given by:

$$\hat{\mathbf{R}}^{-1}(0) = \tfrac{1}{e_o}\mathbf{I}_L; \qquad e_o > 0, \tag{2.85}$$

where $\mathbf{I}_L$ is the $(L \times L)$-dimensional identity matrix and e_o is the initial error. Upon defining a gain vector $\mathbf{g}(n)$ as [4, 8]:

$$\mathbf{g}(n) = \frac{\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)}{\alpha + \mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)} \ , \tag{2.86}$$

the RLS minimisation process results in the following weight update equation:

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mathbf{g}(n)e^*(n). \tag{2.87}$$

1:	$y(n) = \hat{\mathbf{w}}^H(n)\mathbf{x}(n)$
2:	$e(n) = r(n) - y(n)$
3:	$\mathbf{r} = \mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)$
4:	$\kappa = \alpha + \mathbf{r}\mathbf{x}(n)$
5:	$\mathbf{g}(n) = \hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)/\kappa$
6:	$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mathbf{g}(n)e^*(n)$
7:	$\hat{\mathbf{R}}^{-1}(n) = \frac{1}{\alpha}\left(\hat{\mathbf{R}}^{-1}(n-1) - \mathbf{g}(n)\mathbf{r}\right)$

Table 2.7: Updating the array weight vector of the RLS algorithm [166].

Observe that in Equation (2.86) the inversion of the auto-correlation matrix $\hat{\mathbf{R}}$ is replaced at each iteration by a simple scalar division. This reduces the implementational complexity imposed by an order L from that, when the matrix inversion is computed directly. The array weight update procedure of the RLS algorithm is summarised in Table 2.7. The value of the forgetting factor α has no influence on the rate of convergence, but does determine the channel variation tracking capability of the beamforming process [7]. The smaller the value of α, the better the tracking capability of the algorithm. Therefore, if the channel estimate is varying significantly, as a function of time, it is better to use a small forgetting factor in order to 'forget' the past channel estimate. However, if α is significantly less than unity, the algorithm may become unstable [7].

In general, the RLS algorithm converges faster than the LMS algorithm [6, 7, 81, 168]. It is also robust, since its rate of convergence is essentially insensitive to the eigenvalue spread problem. It has however some limitations, especially in terms of its increased computational complexity, since the number of complex multiplications and complex additions per iteration is $4L^2 + 4L$ and $3L^2 + L$, respectively, which is an order of magnitude higher than that of the LMS algorithm. Similar to the LMS technique, there is also a number of variants of the RLS algorithm that may be invoked for adapting the weighting coefficients, such as the square-root RLS (SQ-RLS) [6] and the fast RLS algorithms [6]. These algorithms have their own advantages and disadvantages in comparison to one another. The best design option depends on the communications environment considered and on the statistical knowledge available concerning the received signals. A more recent variant of the RLS algorithm is referred to as the variable forgetting factor nonlinear RLS (VFF-NRLS) adaptive algorithm, which was introduced by Leung and So in [172]. In the context of this algorithm, the auto-correlation of the error is modified and used for controlling the forgetting factor of the nonlinear RLS algorithm. This scheme equips the algorithm with a fast tracking capability and results in a small misadjustment MSE in case of a sub-optimum choice of the array weights, thus this technique outperforms other variable forgetting factor aided RLS algorithms.

Having discussed the RLS algorithm, in the following section we introduce the concept of Sample Matrix Inversion (SMI).

2.4.1.3 Sample Matrix Inversion Beamforming Algorithm

The Sample Matrix Inversion (SMI) algorithm is another technique, which is used to approximate the solution to the MMSE problem. It offers a faster convergence than the LMS method, since it performs direct inversion of the auto-correlation matrix $\mathbf{R}$, thus sometimes it is also referred to as the Direct Sample Covariance Matrix Inversion or Direct Matrix Inversion (DMI) procedure of solving the Wiener-Hopf equation. Unlike in the context of the LMS algorithm, its rate of convergence is independent of the eigenvalues of $\mathbf{R}$, i.e. from the power level of the array elements. Due to requiring no priori information concerning the desired and interfering signals, the DMI algorithm updates the array weight vector by replacing $\mathbf{R}$ with its estimate based on a finite number of signal samples [66]. Hence, this method is termed as block-adaptive. When using N samples of the array element signals $\mathbf{x}(n)$, where $n = 0, 1, 2, ..., N-1$, the estimate of $\mathbf{R}$ may be obtained using a simple averaging scheme, resulting in:

$$\hat{\mathbf{R}}(n) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{x}(n)\mathbf{x}^H(n) \tag{2.88}$$

and similarly, the sample cross-correlation vector $\hat{\mathbf{z}}$ is evaluated as:

$$\hat{\mathbf{z}}(n) = \frac{1}{N}\sum_{n=0}^{N-1}r^*(n)\mathbf{x}(n). \tag{2.89}$$

The number of samples N may be referred to as the size of the observation interval or the window width of the block of samples used. This block-adaptive approach can also be used in a time-varying environment, provided that the array weights are recomputed periodically [8]. The estimated array weight vector is given by:

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{R}}^{-1}\hat{\mathbf{z}} \tag{2.90}$$

and the error associated with the estimation of the cross-correlation vector on the basis of N samples may be evaluated as:

$$\mathbf{e} = \hat{\mathbf{R}}\mathbf{w}_{opt} - \mathbf{z}. \tag{2.91}$$

The solution given by Equation (2.90) may be viewed as the least squares formulation of the problem, since the error $\mathbf{e}$ of Equation (2.91) is defined to be the difference of the estimated sample cross-correlation vector $\hat{\mathbf{z}}$ and the actual vector $\mathbf{z}$, in which the weight vector is directly optimised based on minimising the sum of the squared errors of all N samples. In other words, the cross-correlation and auto-correlation matrices of Equations (2.88) and (2.89), respectively, are estimated using the $(N-1)$

previous samples up to the current time index n. Thus the weight vector generated is a least squares solution [8]. It has been shown in [63] that the adaptive system achieves an SIR performance, which is roughly 3dB worse than that of the optimum solution constituted by the true cross-correlation $\mathbf{z}$ and auto-correlation $\mathbf{R}$ matrices, when the number of samples N is twice the number of total degrees of freedom (DOF). Explicitly, since we have DOF $= L-1$, therefore approximately, $N = 2L$ are used. The estimated value of $\hat{\mathbf{R}}$ may be updated, when the set of N new samples arrives using [81]:

$$\begin{aligned}\hat{\mathbf{R}}(n+1) &= \frac{n\hat{\mathbf{R}}(n) + \mathbf{x}(n+1)\mathbf{x}^H(n+1)}{n+1} \\ &= \alpha\hat{\mathbf{R}}(n) + (1-\alpha)\mathbf{x}(n+1)\mathbf{x}^H(n+1); \qquad 0<\alpha<1, \end{aligned} \tag{2.92}$$

where α is a real scalar forgetting factor that is smaller than but close to unity, which is used for exponentially weighting the past data.

The philosophy of the DMI method is quite straightforward, but as in Equation (2.90), the expression used for calculating the array weights requires $\hat{\mathbf{R}}^{-1}$, and the inversion of $\hat{\mathbf{R}}$ for each new block of received array signal samples is computationally expensive, particularly if there are many elements in the antenna array. According to Equation (2.88), this algorithm requires an averaging for the computation of $\hat{\mathbf{R}}$. Explicitly, for each element of the matrix $\hat{\mathbf{R}}$, N complex multiplications and $N-1$ complex additions are needed. Due to the Hermitian nature of $\hat{\mathbf{R}}$, it is sufficient to execute these instructions $L(L+1)/2$ times, rather than L^2 times, if $\hat{\mathbf{R}}$ were non-Hermitian. This totals $NL(L+1)/2$ complex multiplications and $(N-1)L(L+1)/2$ complex additions, when forming the estimated auto-correlation matrix $\hat{\mathbf{R}}$. However, when using Multiply-and-ACcumulate (MAC) instructions, which are readily available in state-of-the-art processors, the number of mathematical operations imposed reduces to $NL(L+1)$. For the inversion of $\hat{\mathbf{R}}$, another $L^3/2 + L^2$ complex operations are required, when taking the Hermitian nature of the matrix $\hat{\mathbf{R}}$ into account. Otherwise this step requires L^3 operations [4]. To elaborate further, the computation of the cross-correlation vector $\hat{\mathbf{z}}$ would require $N(L+1)$ complex multiplications and $N(L-1)$ complex additions or, equivalently, $N(L+1)$ operations if the availability of the MAC instruction is assumed. Finally, according to Equation (2.90), the array weight vector is obtained with the aid of L^2 complex multiplications and $L(L-1)$ complex additions, or using L^2 complex operations, provided that the MAC instruction is available. When using $N = 2L$ samples, the complexity of the algorithm, again provided that the MAC instruction is available and that the Hermitian property of the matrix $\hat{\mathbf{R}}$ is taken into account, is at least $2.5L^3 + 5L^2 + L$, which is proportional to the order of L^3. This is clearly a major drawback of the DMI algorithm. In order to overcome this problem, again, we apply the Matrix Inversion Lemma of Equation (2.45) to $\hat{\mathbf{R}}$ as formulated in Equation (2.92). Then, it follows that [4, 12]:

$$\begin{aligned}\hat{\mathbf{R}}^{-1}(n) &= \alpha^{-1}\hat{\mathbf{R}}^{-1}(n-1) - \frac{(1-\alpha)\alpha^{-2}\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)}{1+(1-\alpha)\alpha^{-1}\mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)} \\ &= \frac{1}{\alpha}\left[\hat{\mathbf{R}}^{-1}(n-1) - \frac{\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)}{\frac{\alpha}{(1-\alpha)} + \mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)}\right], \end{aligned} \tag{2.93}$$

using the initialisation of:

$$\hat{\mathbf{R}}^{-1}(0) = \frac{1}{e_o}\mathbf{I}_L; \qquad e_o > 0, \tag{2.94}$$

where $\mathbf{I}_L$ is the $(L \times L)$-dimensional identity matrix and again L is the number of array elements. We refer this technique as the Recursive SMI (RSMI) algorithm, in order to differentiate it from the DMI technique, which is taking the inverse of $\mathbf{R}$ of Equation (2.88) for each array weight update.

1:	$\hat{\mathbf{z}}(n) = \frac{1}{n}[(n-1)\hat{\mathbf{z}}(n-1) + r^*(n)\mathbf{x}(n)]$
2:	$\mathbf{r} = \mathbf{x}^H(n)\hat{\mathbf{R}}^{-1}(n-1)$
3:	$\kappa = \frac{\alpha}{1-\alpha} + \mathbf{r}\mathbf{x}(n)$
4:	$\mathbf{g}(n) = \hat{\mathbf{R}}^{-1}(n-1)\mathbf{x}(n)/\kappa$
5:	$\hat{\mathbf{R}}^{-1}(n) = \frac{1}{\alpha}\left(\hat{\mathbf{R}}^{-1}(n-1) - \mathbf{g}(n)\mathbf{r}\right)$
6:	$\hat{\mathbf{w}}(n) = \hat{\mathbf{R}}^{-1}(n)\hat{\mathbf{z}}(n)$

Table 2.8: Updating the array weight vector of the RSMI algorithm.

Table 2.8 summarises the array weight update procedure of the RSMI technique. For stationary signals, the ratio $\frac{\alpha}{(1-\alpha)}$ in step 3 of Table 2.8 or in Equation (2.93) is replaced by the value of unity. This step is different from the LMS and RLS algorithms, which are based on continuous adaptation, where the array weights are adjusted on a sample-by-sample basis as the data is sampled. The RSMI array weight updating procedure is an open loop technique, since it uses a block-adaptive approach in that the statistics are estimated from a temporal block of array data and then used in the optimum weight update procedure outlined in Table 2.8. Following our updating steps summarised in Table 2.8, the procedure requires $5L^2 + 3L + 1$ complex multiplications and $4L^2 - L - 1$ complex additions per iteration. Therefore, by using the recursive formula of Equation (2.93), the complexity is reduced to an order of L^2, while the achievable performance remains similar to that of the DMI algorithm. It should be noted that as the number of samples increases ($n \to \infty$), the matrix update procedure approaches the true value, i.e. we have $\hat{\mathbf{R}}(n) \to \mathbf{R}$ and thus, the estimated array weights approach the optimal weights, corresponding to $\hat{\mathbf{w}}(n) \to \mathbf{w}_{opt}$. For further understanding on the SMI algorithm, we provide an example employing the DMI approach as follows.

An example: Using the same interference environment as in Section 2.2.3, here we provide an example of calculating the array weights optimised for acquiring the desired signal arriving from an angle of $\theta = 0$ radian, while eliminating the signal arriving from an angle of $\theta = \frac{\pi}{6}$ radians using the SMI algorithm in the form of the DMI. We use two four-bit orthogonal reference sequences, one for each signal source and maintain a correlation of zero between them. Specifically, the reference sequence chosen for the desired signal $\bar{m}_1(n)$ is $r_1(n) = [1, 1, 1, 1]$ and that for the interferer $\bar{m}_2(n)$ is $r_2(n) = [1, -1, 1, -1]$. Observe that $r_1(n)$ and $r_2(n)$ are orthogonal.

For an observation interval of $N = 4$, the sample correlation matrix $\hat{\mathbf{R}}(n)$ is calculated based on Equation (2.88) as follows:

$$\begin{aligned}
\hat{\mathbf{R}}(4) &= \frac{1}{4}\sum_{n=0}^{3}\mathbf{x}(n)\mathbf{x}^H(n) \\
&= \frac{1}{4}\left[\begin{bmatrix} 2\bar{m} \\ \bar{m}(1+j) \end{bmatrix}\begin{bmatrix} 2\bar{m} & \bar{m}(1-j) \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{m}(1-j) \end{bmatrix}\begin{bmatrix} 0 & \bar{m}(1+j) \end{bmatrix} + \right. \\
&\quad \left. \begin{bmatrix} 2\bar{m} \\ \bar{m}(1+j) \end{bmatrix}\begin{bmatrix} 2\bar{m} & \bar{m}(1-j) \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{m}(1-j) \end{bmatrix}\begin{bmatrix} 0 & \bar{m}(1+j) \end{bmatrix}\right] \\
&= \frac{1}{4}\left[\begin{bmatrix} 4\bar{m}^2 & 2\bar{m}^2(1-j) \\ 2\bar{m}^2(1+j) & 2\bar{m}^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 2\bar{m}^2 \end{bmatrix} + \right. \\
&\quad \left. \begin{bmatrix} 4\bar{m}^2 & 2\bar{m}^2(1-j) \\ 2\bar{m}^2(1+j) & 2\bar{m}^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 2\bar{m}^2 \end{bmatrix}\right] \\
&= \begin{bmatrix} 2\bar{m}^2 & \bar{m}^2(1-j) \\ \bar{m}^2(1+j) & 2\bar{m}^2 \end{bmatrix},
\end{aligned} \tag{2.95}$$

where we assumed that both signals have the same power of $\bar{m}^2$, since we have $|\bar{m}_1(n)| = |\bar{m}_2(n)| = \bar{m}$. Note that the structure of the correlation matrix is indeed Hermitian, having real values along its main diagonal. For the sample cross-correlation matrix of Equation (2.89), we have:

$$\begin{aligned}
\hat{\mathbf{z}}(4) &= \frac{1}{4}\sum_{n=0}^{3}\mathbf{x}(n)r^*(n) \\
&= \frac{1}{4}\left[\begin{bmatrix} 2\bar{m} \\ \bar{m}(1+j) \end{bmatrix}\bar{m} + \begin{bmatrix} 0 \\ \bar{m}(1-j) \end{bmatrix}\bar{m} + \begin{bmatrix} 2\bar{m} \\ \bar{m}(1+j) \end{bmatrix}\bar{m} + \begin{bmatrix} 0 \\ \bar{m}(1-j) \end{bmatrix}\bar{m}\right] \\
&= \frac{1}{4}\left[\begin{bmatrix} 2\bar{m}^2 \\ \bar{m}^2(1+j) \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{m}^2(1-j) \end{bmatrix} + \begin{bmatrix} 2\bar{m}^2 \\ \bar{m}^2(1+j) \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{m}^2(1-j) \end{bmatrix}\right] \\
&= \begin{bmatrix} \bar{m}^2 \\ \bar{m}^2 \end{bmatrix},
\end{aligned} \tag{2.96}$$

where ideally, $r^*(n) = \bar{m}_1^*(n) = \bar{m}^*(n)$ and for the real-valued case we have $\bar{m}^*(n) = \bar{m}(n)$. The weights of the beamformer are then calculated using the average values of $\hat{\mathbf{R}}(4)$ and $\hat{\mathbf{z}}(4)$ according to the optimum Wiener-Hopf equation of:

$$\hat{\mathbf{w}} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{z}}. \tag{2.97}$$

Since we have a correlation matrix of (2×2)-dimension, it is straightforward to compute the inverse of $\hat{\mathbf{R}}$. However, when a larger correlation matrix is required, the computations will become more elaborate and the recursive formula of Equation (2.93) should be used. For a (2×2)-dimensional matrix, let us assume an example of matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where the inverse $\mathbf{A}^{-1}$ can be calculated by the following procedure [169, 170]:

1. Interchange the diagonal elements a and d.
2. Change the sign of the other two elements, i.e. c and b.
3. Divide the resultant new matrix by the determinant of $\mathbf{A}$, i.e. $|\mathbf{A}| = ad - bc$. Note that, for a matrix to have an inverse the determinant $|\mathbf{A}|$ must not be zero, i.e. $|\mathbf{A}| = ad - bc \neq 0$.

This gives the inverse of matrix $\mathbf{A}$, such that:

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \tag{2.98}$$

and therefore the inverse of the correlation matrix, namely $\hat{\mathbf{R}}^{-1}$ is expressed as:

$$\begin{aligned} \hat{\mathbf{R}}^{-1}(4) &= \frac{1}{2\bar{m}^4} \begin{bmatrix} 2\bar{m}^2 & -\bar{m}^2(1-j) \\ -\bar{m}^2(1+j) & 2\bar{m}^2 \end{bmatrix} \\ &= \frac{1}{\bar{m}^2} \begin{bmatrix} 1 & -0.5(1-j) \\ -0.5(1+j) & 1 \end{bmatrix}, \end{aligned} \tag{2.99}$$

leading to the following array weights:

$$\begin{aligned} \hat{\mathbf{w}} &= \hat{\mathbf{R}}(4)\hat{\mathbf{z}}(4) \\ &= \frac{1}{\bar{m}^2} \left[\begin{bmatrix} 1 & -0.5(1-j) \\ -0.5(1+j) & 1 \end{bmatrix} \begin{bmatrix} \bar{m}^2 \\ \bar{m}^2 \end{bmatrix} \right] \\ &= \begin{bmatrix} 0.5 + j0.5 \\ 0.5 - j0.5 \end{bmatrix}. \end{aligned} \tag{2.100}$$

In the context of the temporal reference technique of Figure 2.11, we have to take the conjugate of the array weights $\hat{\mathbf{w}}$ in Equation (2.100) for the sake of obtaining the correct weight values, yielding $\mathbf{w}_1^* = 0.5 - j0.5$ and $\mathbf{w}_2^* = 0.5 + j0.5$. This is equivalent to the values calculated earlier in Section 2.2.3, namely in Equation (2.39). Note that in case of a single interfering user and a single desired user under perfect channel conditions, when the received signal is uncontaminated by channel effects and hence there is a perfect match between the reference and the desired signal, we are able to obtain the correct array weight values after receiving the corresponding two reference sequences.

Having discussed the algorithms employing the temporal reference technique, we shall next briefly introduce some algorithms using the spatial reference technique.

2.4.2 Spatial Reference Technique

Apart from the family of temporal reference techniques, there is also a large body of research based on direction of arrival (DOA) estimation [1, 81, 113], approaching the problem of desired and interfering users identification from the viewpoint of spectral analysis [173], [85, 174] in the space domain. The

information on the DOA is usually obtained by applying a particular DOA estimation technique for processing the received array signals [8].

At this stage it is necessary to physically interpret the philosophical differences between the temporal and spatial reference techniques. The temporal reference approach is conceptually similar to using a channel sounding or equaliser training pattern. More explicitly, in the latter case often a dirac-delta-like impulse is transmitted in each transmission burst for the sake of estimating the Channel's Impulse Response (CIR) and adjusting the equaliser's weights, in order to minimise the multipath interference imposed by the dispersive CIR's paths. By contrast, in the context of the temporal reference approach the reference signal is used for differentiating the interfering users, rather than the interfering paths. The interfering users' signals arrive from specific angles, and are associated with specific delays and similarly, the interfering paths of the CIR are also associated with specific delays. The corresponding training or reference signals are used for adjusting the equaliser or beamformer weights, respectively.

Viewing the same processes in the frequency domain now, the well-understood channel equaliser may use a 'spectrally white' signal for estimating the channel's frequency domain transfer function, instead of the CIR. Hence the dirac-delta-like channel sounding sequence could be transformed to its 'spectrally white' frequency-domain representation with the aid of the DFT and the effects of multipath interference induced spectral-domain linear distortions could be compensated with the aid of frequency-domain division of the received signal by the channel's transfer function.

In the spectral estimation special issue [175] published in 1982, the family of spectral estimation techniques has been grouped into two broad categories, namely the parametric and non-parametric classes. In the class of parametric spectral analysis methods a model is assumed in the formulation of the problem, and the requirement is to estimate the parameters of the model from observations of a given process encompassing a limited duration of time. The family of algorithms employing this approach include the Maximum Entropy Method (MEM) [81] and the Maximum Likelihood Estimation (MLE) technique [8, 113]. The parametric methods however, typically impose a high degree of computational complexity. On the other hand, in the class of non-parametric spectral analysis methods no specific model is assumed in formulating the estimation problem as seen, for example in the spectral-based approach of [113]. Using the spectral-based approach, some spectrum-like function of the parameters of interest, such as the DOA is formulated. Examples of algorithms using this approach include the MUSIC algorithm [78,79,81,82,113] and the approach of the Estimation of Signal Parameters by Rotational Invariance Techniques (ESPRIT) [80, 81, 83, 113].

The MUSIC algorithm is a technique devised for estimating the DOA that is based on decomposing the auto-correlation matrix $\mathbf{R}$ of the received array signal into diagonal signal eigenvalues Λ_d and diagonal noise eigenvalues Λ_n, as well as into the corresponding signal eigenvectors $\mathbf{E}_d$ and noise eigenvectors $\mathbf{E}_n$. Thus the auto-correlation matrix $\mathbf{R}$ of Equation (2.25) discussed earlier in Section 2.2.1

can be expressed as:

$$\mathbf{R} = \mathbf{S}\mathbf{P}\mathbf{S}^H + 2\sigma_n^2\mathbf{I}_L, \tag{2.101}$$

$$= \underbrace{\mathbf{E}_d\Lambda_d\mathbf{E}_d}_{\text{signal subspace}} + \underbrace{\mathbf{E}_n\Lambda_n\mathbf{E}_n}_{\text{noise subspace}}, \tag{2.102}$$

where $\mathbf{S}$ is the $(L\times M)$-dimensional steering matrix defined in Equation (2.14), $\mathbf{P} = diag[p_1, p_2, \ldots, p_M]$ is the $(M \times M)$-dimensional diagonal matrix, whose elements are constituted by the power of the incoming signals $p_m = |a_m|^2, m = 1, 2, \ldots, M$, given the modulating signal as defined in Equation (2.6), while $2\sigma_n^2$ is the complex-valued noise variance and $\mathbf{I}_L$ is an $(L \times L)$-dimensional identity matrix. The eigenvectors corresponding to the principal eigenvalues represent the signal subspace. The smallest eigenvalues and the corresponding eigenvectors represent the noise subspace, which is the basis of the MUSIC algorithm. This implies that the MUSIC algorithm requires an antenna array having L number of elements, where L is higher than the number of users M [113]. For a wave impinging on the antenna array at an angle θ_m, the pseudo-spectrum is given by [78, 81, 170]:

$$P_{\text{MUSIC}}(\theta_m) = \frac{1}{\mathbf{s}^H(\theta_m)\mathbf{E}_n\mathbf{E}_n^H\mathbf{s}(\theta_m)}, \tag{2.103}$$

where $\mathbf{s}(\theta_m)$ is the steering vector of the L antenna elements corresponding to direction θ_m, defined as in Equation (2.8), restated here as:

$$\mathbf{s}(\theta_m) = \left[e^{j2\pi f t_1(\theta_m)} \quad \ldots \quad e^{j2\pi f t_L(\theta_m)}\right]^T, \tag{2.104}$$

where $2\pi f t_l(\theta_m)$ relates the phase of the received signal at each element to the angle of arrival (AOA) θ_m and the delay $t_l(\theta_m)$ is defined in Equation (2.3). The peaks obtained in the MUSIC spectrum defined in Equation (2.103) correspond to the M estimated DOA.

Beamforming is ineffective, if an interfering user is located close to the desired user, such that the angular separation becomes lower than the beamwidth, since it becomes difficult to reproduce the desired user's signal at the receiver accurately. In fact, it was pointed out in [176] in a wireless multipath environment that pathological geometrics corresponding to any particular location of the users may be encountered, for which sufficiently accurate signal detection becomes impossible.

DOA estimation techniques are best combined with a beamforming technique, rather than operated consecutively, since the DOA of mobile users is very important in order to form an optimum radiation pattern for the antenna array. More explicitly, once the DOA of the various sources becomes available, the beamforming algorithms can be used for tracking the sources of interest as they roam and null the interfering sources by controlling the beam pattern of the antenna array in an adaptive manner. A design example that incorporates both DOA estimation and beamforming was contributed by Margarita *et al.* [177]. They used the MUSIC algorithm [78,81,113] for the DOA estimation and the General Data Independent Response Design (GDIRD) [88] beamforming technique for determining the weights of the array elements for application in the GSM and DECT systems. The GDIRD technique falls into the data independent, rather than in the statistically optimum beamformer category of

Section 2.4, where the weights do not depend on the array's received signal or on its statistics and are optimised for achieving a specified response for a selected number of desired signal and interference scenarios. Although the algorithm was unable to find the DOA, when the interferer was located at an angular separation of $\pm 5^o$ from the desired user, but nonetheless, generally speaking, satisfactory results have been obtained, demonstrating that the employment of SDMA techniques was capable of usefully increasing the system capacity of both GSM and DECT type systems.

The popularity of the MUSIC and ESPRIT algorithms is due to their high angular resolution. However, when compared to DFT-based spectral-domain arrays, their implementation complexity is higher owing to the need of calculating the eigenvalues of the correlation matrix. A lower-complexity method than the MUSIC and ESPRIT algorithms, that combines the spatial and temporal reference approaches has been proposed by Watanabe *et al.* [178]. The authors proposed and investigated an adaptive antenna array using a DFT-based LMS algorithm, aiming for maintaining the achievable advantages, while reducing the associated weaknesses, when using each algorithm separately. The authors also showed that the combined algorithm is capable of improving the attainable convergence speed in comparison to a system using temporal reference based techniques, such as for example the LMS algorithm alone, while also improving the accuracy of array weight values in comparison to using the DFT-based arrays of [178]. In short, the proposed algorithm compensated for the weaknesses of each individual algorithm achieving a rapid adaptation, accurate estimation and a low implementation complexity.

In conclusion, the family of spatial spectral estimation algorithms can be invoked for obtaining optimum or sub-optimum array weights by using spatial received signal samples at a given time instant and therefore, if the processing speed is sufficiently high for the sake of tracking the time variations of the channels encountered, these algorithms may become more attractive, than the class of temporal reference based array weight updating algorithms of Section 2.4.1 [11, 114].

2.4.3 Blind Adaptive Beamforming Algorithms

The family of blind adaptive algorithms does not require a training sequence and hence its representatives utilise the available bandwidth efficiently. More explicitly, the transmission of training sequences wastes valuable bandwidth and the associated overhead increases, if frequent re-training is required for avoiding their deficient convergence in high-Doppler time-varying propagation channels [179]. Hence, in the scenario, where an explicit reference signal is unavailable, blind adaptive beamforming is used. In this scenario the system will generate its own reference signal by exploiting the characteristics of the transmitted signal and therefore it may be able to dispense with training sequences. This class of beamforming techniques includes algorithms such as the Constant Modulus Algorithm (CMA) [8, 81, 146], the Spectral COherence REstoral algorithm (SCORE) [8, 145] and Decision Directed Algorithms (DDA) [8].

Important performance metrics characterising blind algorithms are the speed of convergence, the ability to reach the global optimum and their capability of tracking time-varying mobile channels [1]. The success of blind detection algorithms depends on their ability to exploit input signal features such as their spectral whiteness, or lack of it, the size of the received signal alphabet, their constant modulus and cyclostationarity or the shape of the probability density function (PDF) [1,8,179]. From the family of blind algorithms, the CMA is probably the most widely investigated blind algorithm and it is also most widely used in practice [6,81]. This is owing to the fact that this family of algorithms has also been recognised to exhibit robustness against beamforming array deficiencies and multipath environments, which preclude the employment of super-resolution beamforming techniques for DOA estimation in a multiuser communications context.

2.5 Uplink versus Downlink Beamforming

For years, significant research efforts have been devoted to the study of adaptive beamforming designed for the uplink (i.e. mobile to base station). This due to the fact that adaptive beamforming has been traditionally used for reception in remote sensing [180], radar [180, 181] and sonar systems [182]. The techniques and algorithms that have been developed for reception readily lend themselves to employment in wireless communications systems. Furthermore, the spatial channel information is more readily available on the uplink, where the base station (BS) may be conveniently fitted with multiple antenna elements, when invoking beamforming techniques and algorithms for mitigating the effects of multipath fading and co-channel interference [8]. As discussed before, the main objective of uplink receiver beamforming is to receive as much power as possible from the desired user and as little power as possible from any undesired interfering user.

In contrast to the uplink, research aimed at designing beamforming algorithms for the downlink or forward link (i.e. base station to mobile) has only been receiving research attention during the last few years. Similarly to the previously discussed uplink scenario, the aim of adaptive beamforming designed for the downlink is also to transmit as much power as possible to the desired mobile and as little power as possible to any mobile where this signal would inflict interference. In other words, this scenario is similar to that of the uplink case in terms of maximising the SINR for the desired user. The only difference is that we intend to maximise the received signal strength at the desired mobile, while minimising the interference inflicted upon other mobiles as well as upon the adjacent base stations, thus maximising the achievable downlink SINR.

One of the main issues associated with downlink beamforming using smart antennas is the impact of the desired user's signal on other subscribers [183], since beamforming conveying energy to a desired user affects the interference imposed on others. Therefore, downlink beamforming must be performed by jointly considering the entire mobile population roaming within the entire cell. Shim and Alam [183]

proposed a novel algorithm employing the Conjugate Gradient Method (CGM) for mitigating the effects of the interference. Their algorithm, explicitly referred to as Adaptive Interference Mitigation using the CGM (AIM-CGM) was evaluated using the signal format of FDD/WCDMA (frequency division duplex/wideband CDMA). It was shown that the achievable gain of the beamforming process depends on the power distribution of the users as determined by their power control and also on the DOA of the interferers. A typically better performance is achieved for scenarios having higher-power interferers, which are located closer to the desired user in terms of their angle of arrival separation, but the number of users that may be supported by the proposed algorithm is limited, since it cannot be higher than the number of antennas. However, in WCDMA systems [15] the number of users cannot be high, if the users require high-rate services and hence for this particular scenario the algorithm is capable of offering an improved performance.

In time division duplex (TDD) systems similar channel conditions prevail in the uplink and downlink, provided that the transmission frame duration is short compared to the reciprocal of the maximum Doppler frequency [8,183,184]. In other words, the assumption of having similar uplink/dowlink scenarios is valid, if the carrier frequency of the uplink and downlink is the same and when both the interference and the propagation environment remains similar during the time interval between reception and transmission. In this case the same beamforming weights may be used for both the uplink and downlink.

Unlike TDD systems, FDD arrangements encounter more complex beamforming problems as a consequence of the difference between the interference scenarios of the uplink and downlink [183]. Standard systems, such as IS-54 [185], IS-95 [7] and GSM [185–187] cannot readily apply the array weight reuse concept as a consequence of the separation between the downlink and uplink frequencies, since the corresponding channel transfer function as well as interference scenario is expected to be different to a large extent [8]. For example, the downlink bandwidth may be outside the coherence bandwidth of the uplink channel, hence the instantaneous fading on the two radio links may be uncorrelated.

One of the techniques that may be used for mitigating this challenging problem is the application of fixed-beam techniques both in the uplink and downlink directions at the base station [188,189]. The strongest beam identified in the uplink will be used for transmission in the downlink. This concept can be extended to the employment of steerable multiple beams [190]. For the uplink reception process, the base station determines the direction of the path in which the strongest component of the desired signal arrives at the base station, while for downlink transmission, the base station points a beam in the corresponding direction. Although this approach is not optimal, the achievable SINR level experienced at the mobile can be improved. Furthermore, at the base station power is not a critical resource and a relatively high-power beam can be used for boosting the SINR level, if necessary. This is in contrast to the uplink case, where a mobile typically transmits using an omni-directional antenna

at a limited power and therefore the desired signal components arrive at the base station via multiple paths.

The attainable performance of fixed-beam downlink beamforming applied to a WCDMA system was evaluated in [189]. More specifically, the performance of a fixed-beam antenna array having two beams was compared to that of a single-antenna aided base station. Simulations were performed over a single-path Rayleigh-fading channel and two dispersive channels similar to the Modified Pedestrian A channel and Modified Vehicular A channel of the ITU [191]. These investigations indicate that the fixed-beam approach may be expected to offer a substantial coverage and capacity extension, especially in macrocells having a relatively narrow angular spread encountered owing to the typically high base station-mobile station distance. However, having only two beams per sector is insufficient for obtaining an adequate performance in scenarios, when the desired user is between two adjacent beams. Furthermore, in environments having a large angular spread, the fixed-beam approach fails to provide a substantial performance improvement in comparison to that of the single antenna transmission.

However, if the frequency separation between the uplink and downlink bands is not too high, they are likely to be affected by the propagation environment in a similar fashion, although their SINR may be rather different. Nonetheless, downlink beamforming may be able to use the uplink weight vector.

Czylwik and Matsumoto [192] discussed the optimisation of downlink beamforming schemes designed for FDD systems. More specifically, they presented a novel approach, which exploits the knowledge of the propagation channel's statistics, which was evaluated amongst each base station and each mobile in order to estimate the achievable diversity gain of receivers, which combine the signals arriving over different paths. This approach can be applied in conjunction with RAKE receivers in the context of CDMA systems, as well as for TDMA receivers invoking adaptive channel equalisers. The simulation results provided show that when the transmitted power is essentially conveyed in the direction of the path having the minimum average attenuation or when the power is distributed over several propagation paths, yield a rather similar response. The advantage of concentrating the power to the main path is that the interference power imposed on other mobiles may be reduced, while in the case of distributing the transmit power to several paths has the advantage of achieving a higher path diversity. An SIR gain of about 8dB was obtained at low SIRs using smart antennas when compared to omni-directional antennas [192]. At low SIRs, the attainable performance was poor, when no diversity was invoked. More recent papers that address the issue of joint downlink beamforming and power control include [193, 194].

Another source of co-channel interference in the downlink is constituted by the adjacent base stations. Until recently, most research on the application of adaptive antenna arrays in mobile communications networks was focused on the base station site. This may be attributed to the associated hardware complexity requirements [195]. However, owing to the increasing demand for higher trans-

mission rates and with the advent of rapid developments in electronics, employing adaptive arrays at the mobile may become both necessary and feasible, provided that the achievable performance gain becomes sufficiently high to offset the additional cost and power consumption. In comparison to the base station, the mobile typically encounters a less complex co-channel interference pattern, where the number of interfering sources (i.e. the corresponding adjacent base stations) is comparatively low. Therefore, using adaptive arrays at the mobile has the potential of suppressing the co-channel interference, which in turn results in an increased SINR, thus increasing the interference-limited capacity of a CDMA system. Gao and Sousa [195] studied the efficiency of this approach and concluded that by using a three- or four-element antenna array the interference is indeed substantially reduced. However, owing to the limited physical size of the mobile station, the number of antenna elements used should be kept as low as possible. Furthermore, the number of interfering signals that may be cancelled cannot be more than three owing to the associated high computational power requirement.

2.6 Beamforming Simulation Results

2.6.1 Performance of a Two-Element Uniform Linear Array System

Against this background, in the next section we characterise the achievable performance of a range of beamforming arrangements. The aim of our simulations is to investigate the attainable performance of a range of adaptive algorithms using the temporal reference technique [8, 114] of Section 2.4.1 for narrowband beamforming, in terms of their interference rejection gain or equivalently, Signal-to-Interference Ratio (SIR) gain for transmission over both AWGN and flat Rayleigh fading channels. The fading channel considered assumes a carrier frequency f_c of 1.9GHz and a transmission rate R_b of 2.6×10^6 symbols/s (Baud) at a vehicular speed of $v = 30$mph. The corresponding Doppler frequency f_D is calculated to be about 85Hz.

The parameters used in the simulations are as follows. Two antenna elements having a spacing of half a wavelength were used, the desired signal of the wanted user arrived from a direction of 0 radian, while the interferer from $\frac{\pi}{6}$ radians. The reference signal was generated employing an orthogonal Walsh sequence and Binary Phase Shift Keying (BPSK) modulation was used. A total of 10000 iterations are used for the AWGN channel, but this number varies for the Rayleigh channel. For the fading channel, we ensure that all simulations experienced at least 200 fades or in other words, we used a minimum of 3×10^6 transmitted symbols. Due to the instability and poor convergence rate of the LMS channel estimation algorithm [4, 12], here we concentrates on the DMI and RSMI algorithms of Section 2.4.1.3 and RLS algorithm of Section 2.4.1.2.

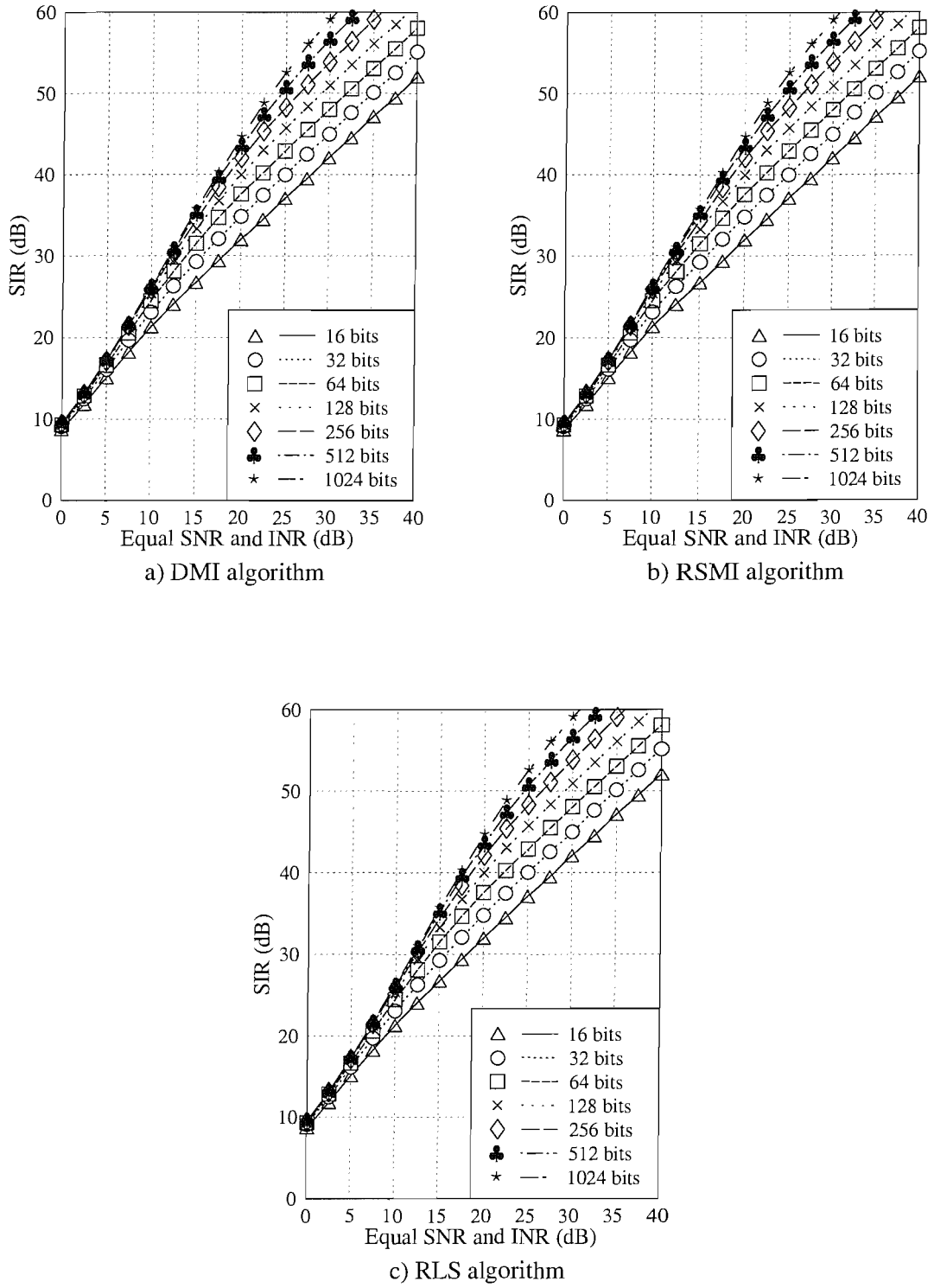


Figure 2.13: The Signal-to-Interference Ratio (dB) achieved with respect to equal desired and interfering signal power (i.e. for **SNR=INR**) using a) the **DMI** and b) the **RSMI** algorithms of Section 2.4.1.3 and c) the **RLS** algorithm of Section 2.4.1.2 employing reference signals of different length over the **AWGN channel**. The **initial error** e **was set to 0.01** and the **forgetting factor** α **was unity**.

2.6.1.1 AWGN Channel

Figure 2.13 shows the performance of the DMI and RSMI algorithms of Section 2.4.1.3 and RLS algorithm of Section 2.4.1.2 as the power of the incoming signals is increased. The initial estimate of the inverse correlation matrix $\hat{\mathbf{R}}^{-1}(0)$ of Equations (2.85) and (2.94) in the RLS and RSMI algorithms, respectively, is biased with an error value of 0.01. When the forgetting factor was set to unity, the DMI, RSMI and RLS algorithms exhibited an identical performance. This is expected since both techniques are similar in that the weight vectors are based on the least squares solution and in the context of the RLS and RSMI algorithms, their difference essentially lies in the formula used for calculating the correlation vector $\hat{\mathbf{z}}$. Recall from Equations (2.81) and (2.89) that as the number of samples is increased, the estimated array weight coefficients approach the optimum value and hence for a given signal power, a longer reference signal is required for achieving a better performance. For SNR and INR values of at least 15dB, a satisfactory performance of more than 25dB beamforming gain was achieved using a reference sequence length as low as 16 bits.

The plots seen in Figure 2.14 demonstrate the achievable interference rejection of the DMI and RSMI algorithms of Section 2.4.1.3 and RLS algorithm of Section 2.4.1.2 as a function of the reference signal length for an equal desired and interfering signal power (SNR = INR) of $6, 12, 18$ and 33dB. Higher received signal powers result in a reduced interference level, but actually only the power of the interference signal (INR) affects the associated performance curves [4,12]. For a fixed INR, varying the SNR values resulted in a similar performance curve. Observe that there is a slight difference between the RLS and SMI algorithm's performance at the initial point of the curve. This is likely to be due to the difference in the calculation of the cross-correlation matrix $\hat{\mathbf{z}}$ of Equations (2.83) and (2.89) for the RLS and RSMI algorithms, respectively. However, the associated difference is not critical, since the reference signal length is commonly higher than eight bits.

Variation of the value of the initial error e_o in Equations (2.85) and (2.94) does not result in a significant SIR difference for the RLS and RSMI schemes of Sections 2.4.1.2 and 2.4.1.3, respectively. Increasing the initial error from 0.01 to 0.9 resulted in a difference of about 0.01dB for the RSMI algorithm of Section 2.4.1.3 and even less for the RLS algorithm of Section 2.4.1.2. In terms of the forgetting factor α, the performance of the algorithm is optimum, when it is set to unity, where $\alpha = 1.0$ is the desired value for the AWGN channel. The effect of varying the forgetting factor α in conjunction with a fixed reference sequence length of 256 bits and assuming different signal powers is shown in Figure 2.15. For both algorithms, the performance is improved as the forgetting factor is increased towards unity, where the RSMI algorithm has a higher rate of improvement as opposed to the shallower slope of the RLS algorithm. In terms of the achievable interference rejection, the RLS algorithm performs better, which is evidenced by the observation that the final SIR value is always higher than that of the RSMI algorithm, except when the forgetting factor α equals to unity, where both algorithms have an equivalent performance. This may be due to the difference of the recursive

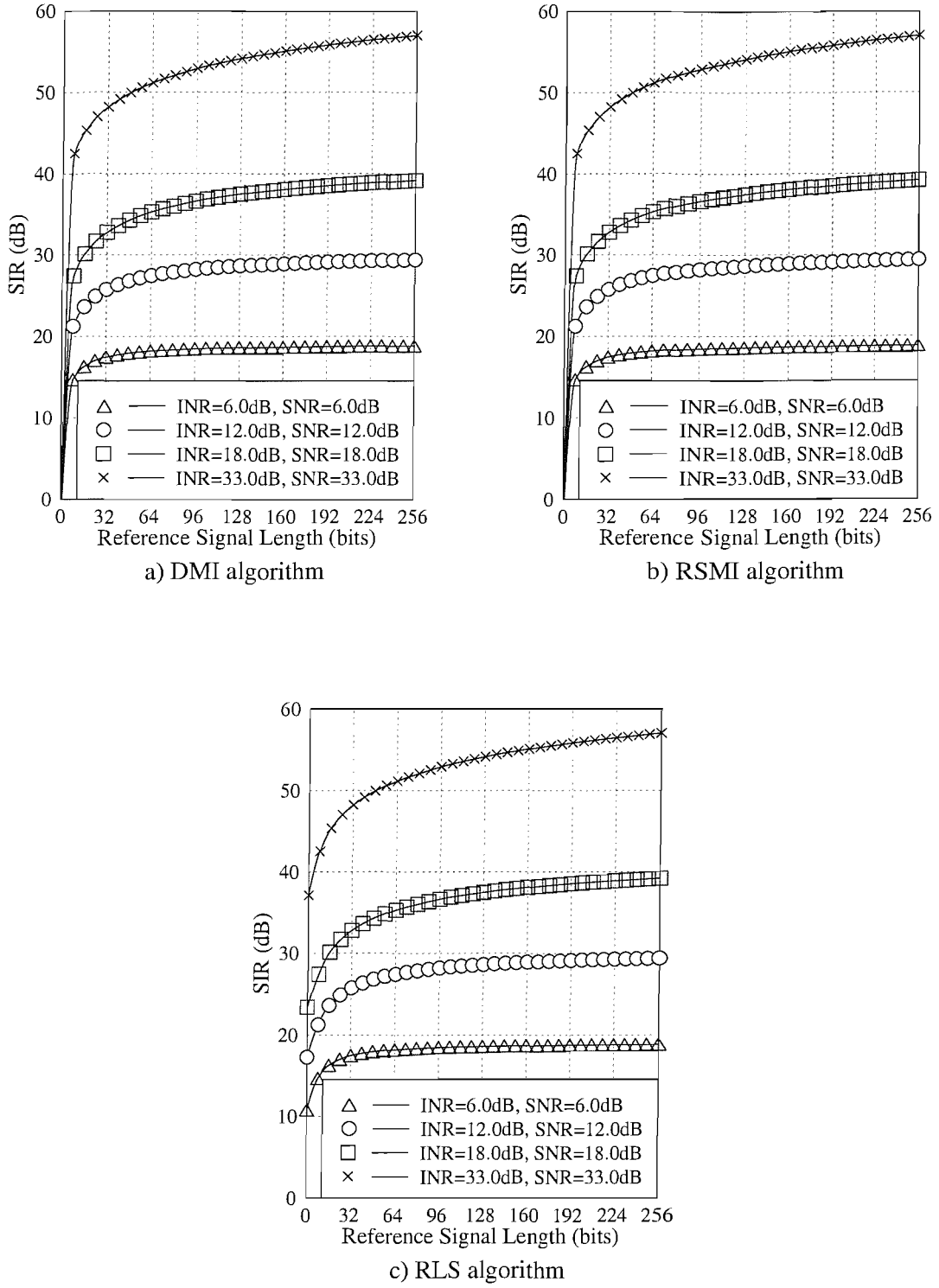
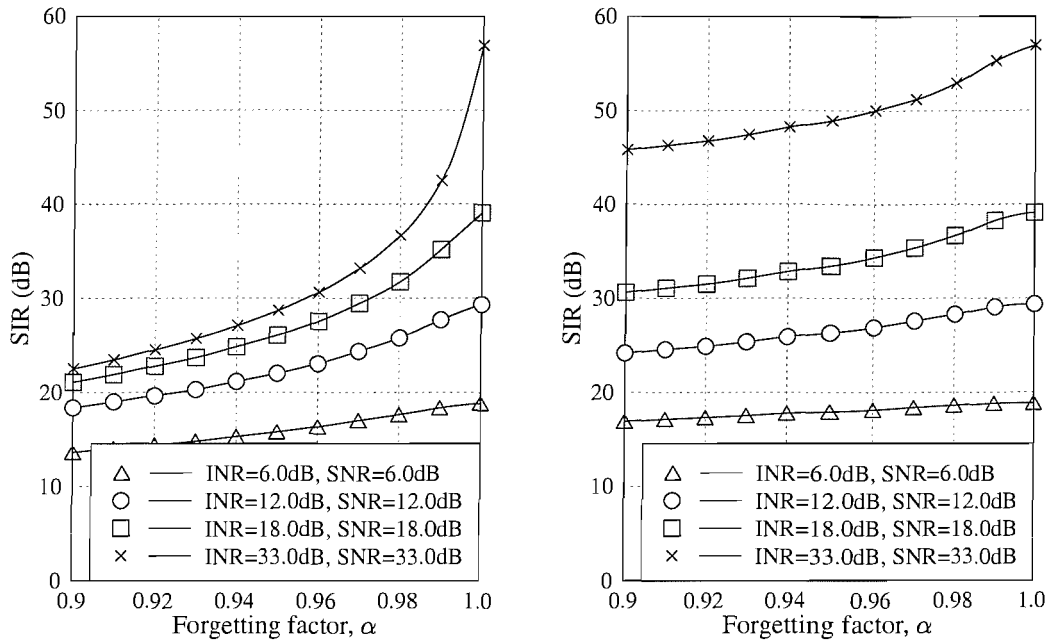


Figure 2.14: The Signal-to-Interference Ratio (dB) achieved with respect to reference signal length using a) the **DMI** and b) the **RSMI** algorithms of Section 2.4.1.3 and c) the **RLS** algorithm of Section 2.4.1.2 in conjunction with different sets of equal desired and interfering signal power (i.e. for **SNR=INR**) over the **AWGN channel**. The **initial error** e **was set to 0.01** and the **forgetting factor** α **was unity**.



a) RSMI algorithm

b) RLS algorithm

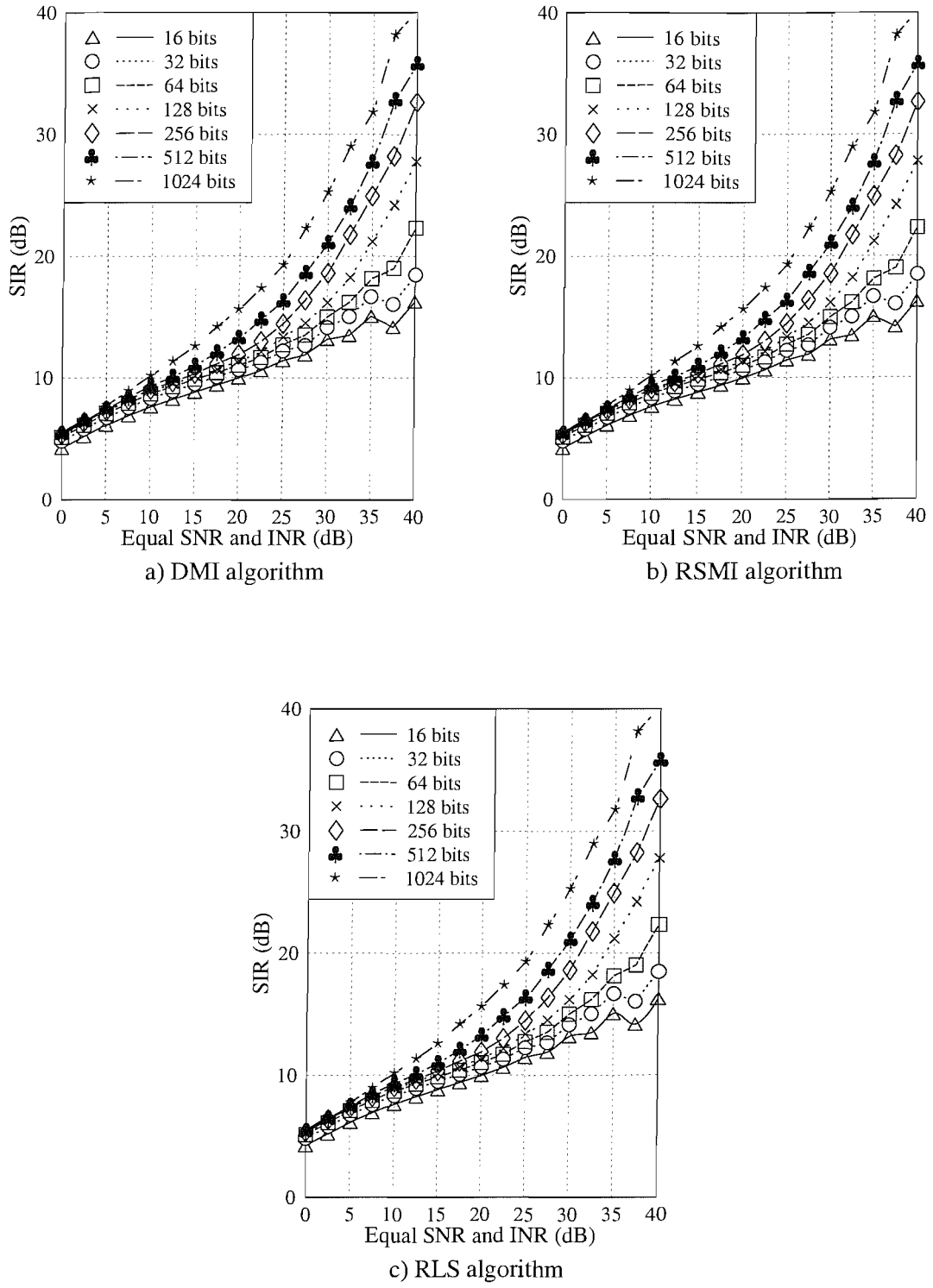
Figure 2.15: The Signal-to-Interference Ratio (dB) as a function of the **forgetting factor** α using a) the **RSMI** algorithm of Section 2.4.1.3 and b) the **RLS** algorithm of Section 2.4.1.2 in conjunction with different sets of equal desired and interfering signal powers (i.e. for **SNR=INR**) and a **reference length of 256 bits** over the **AWGN channel.**

formulation of the inverse auto-correlation matrix $\hat{\mathbf{R}}^{-1}$ of Equations (2.84) and (2.93) for the RLS and RSMI algorithms, respectively, for $\alpha < 1$. We, however, only considered the scenario of having a unity forgetting factor in the context of our simulations conducted over the AWGN channel.

2.6.1.2 Flat Rayleigh Fading Channel

Figures 2.16 and 2.17 show the performance of the adaptive beamforming algorithms of Sections 2.4.1.2 and 2.4.1.3 over flat fading channels, when using a forgetting factor α fixed to unity. Similar to the AWGN scenario, the DMI, RSMI and RLS algorithms exhibits a similar response, although they experience a significant degradation in terms of the achievable interference rejection, when compared to the results of Figures 2.13 and 2.14 as a function of the signal power variation and reference sequence length, respectively.

As the forgetting factor α of Equations (2.84) and (2.93) is reduced, the attainable performance of the RLS and RSMI algorithms, respectively, is degraded. The SIR versus SNR and INR characteristic

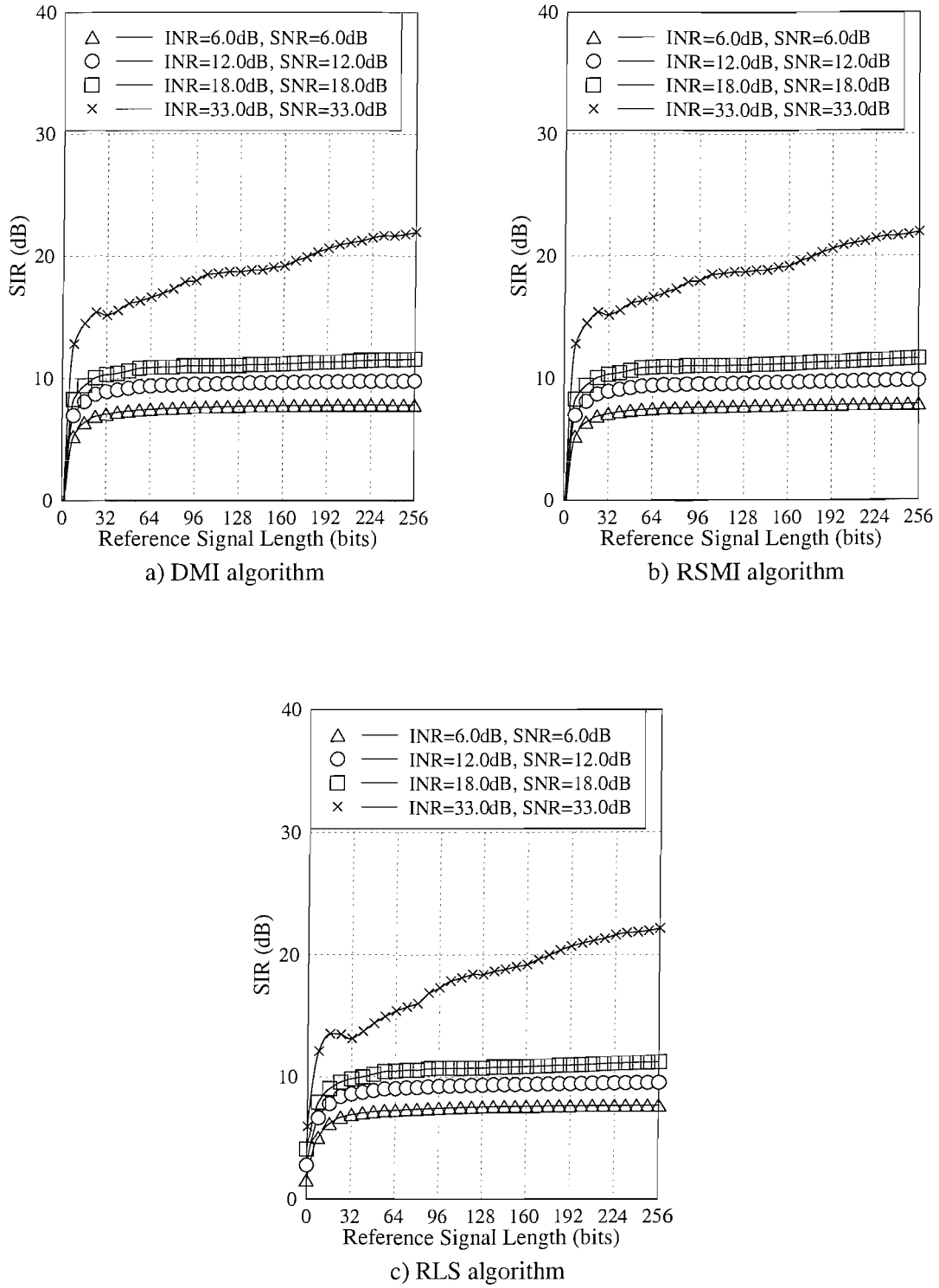


a) DMI algorithm

b) RSMI algorithm

c) RLS algorithm

Figure 2.16: The Signal-to-Interference Ratio (dB) achieved with respect to equal desired and interfering signal power (i.e. for **SNR=INR**) using a) the **DMI** and b) the **RSMI** algorithms of Section 2.4.1.3 and c) the **RLS** algorithm of Section 2.4.1.2 employing reference signals of different length over the **flat Rayleigh fading channel**. The **initial error** e **was set to 0.01** and the **forgetting factor** α **was unity**.



a) DMI algorithm

b) RSMI algorithm

c) RLS algorithm

Figure 2.17: The Signal-to-Interference Ratio (dB) achieved with respect to reference signal length using a) the **DMI** and b) the **RSMI** algorithms of Section 2.4.1.3 and c) the **RLS** algorithm of Section 2.4.1.2 in conjunction with different sets of equal desired and interfering signal power (i.e. for **SNR=INR**) over the **flat Rayleigh fading channel**. The **initial error** e **was set to 0.01** and the **forgetting factor** α **was unity**.

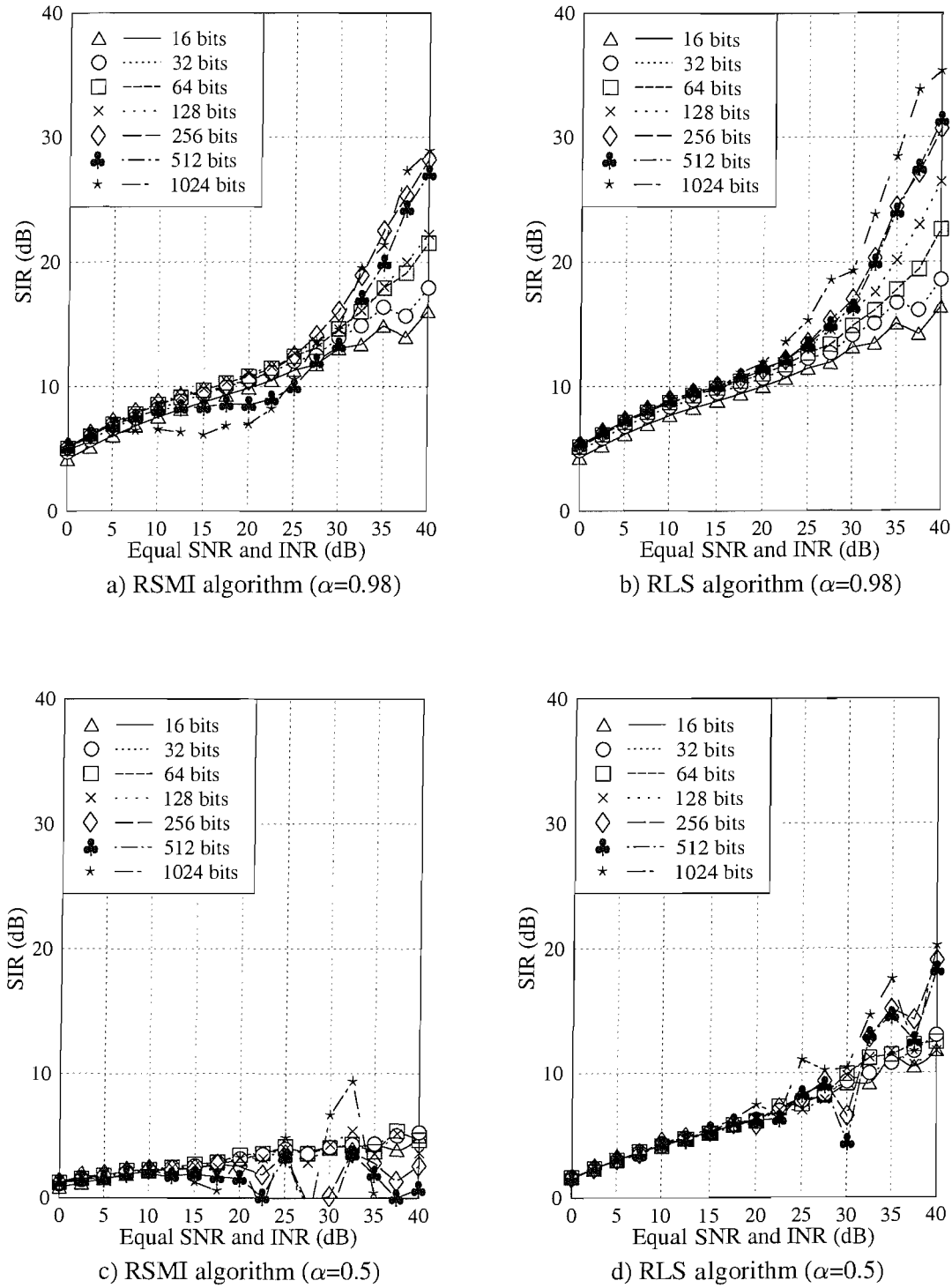


Figure 2.18: The Signal-to-Interference Ratio (dB) achieved with respect to equal desired and interfering signal power (i.e. for **SNR=INR**) using the **RSMI** algorithm of Section 2.4.1.3 for a) $\alpha = 0.98$ and c) $\alpha = 0.5$ and the **RLS** algorithm of Section 2.4.1.2 for b) $\alpha = 0.98$ and d) $\alpha = 0.5$ employing reference signals of different length over the **flat Rayleigh fading channel**.

of the algorithms expressed as a function of the incoming signals' power is portrayed in Figure 2.18. In conjunction with $\alpha = 0.98$, the highest SIR improvement was found for the RSMI algorithm of Section 2.4.1.3, which is about 25dB while for the RLS algorithm of Section 2.4.1.2 this was 30dB, across the 40dB range of the incoming signal power. The level of interference rejection found for the RLS algorithm in Figure 2.18 b) was proportional to the length of the reference sequence used, but for the RSMI algorithm of Section 2.4.1.3, the achievable SIR does not necessarily improve, when a longer reference sequence is used as seen in Figure 2.18 a). Instead, it is obvious that the attainable rejection of the RSMI algorithm of Section 2.4.1.3 is degraded, when a longer reference sequence is used, in particular for the 512-bit and 1024-bit sequences. It is also observed in Figures 2.18 a) and 2.18 b) that the SIR versus SNR and INR characteristic are similar, regardless of the reference length for signal powers up to 20dB, except for the 512-bit and 1024-bit reference sequence length based RSMI scenario. As the forgetting factor is further reduced to $\alpha = 0.5$ for example, the level of interference rejection was similar, regardless of the length of the reference sequence at low signal power levels, namely for SNR or INR values below 20dB. The achievable SIR performance of the algorithms fluctuates at higher signal powers, especially for the reference sequence lengths of 256-bit, 512-bit and 1024-bit, as seen in Figures 2.18 c) and 2.18 d).

In Figure 2.19, we plot the SIR versus reference sequence length characteristics of both the RSMI and RLS algorithms for forgetting factors of $\alpha = 0.98$, $\alpha = 0.95$ and $\alpha = 0.5$. Observe that the SIR curve of the RLS algorithm exhibits undulations for an INR value of 33dB. More explicitly, the range of fluctuation is about 3dB for $\alpha = 0.95$ and increases to about 5dB for the case of $\alpha = 0.5$, tending to increase as we decrease the forgetting factor further. Moreover, for lower signals powers, i.e. for low INR values the range of fluctuations and the associated reduced interference rejection capability is not as pronounced as in the INR = 33dB scenario. Comparing the two algorithms, the RSMI regime exhibits worse performance, where the range of SIR fluctuations is about 3dB for $\alpha = 0.98$. The SIR performance fluctuates even more dramatically, namely up to 10dB, for the case of $\alpha = 0.95$ and $\alpha = 0.5$ as seen in Figures 2.19 c) and 2.19 e), respectively. Although the associated SIR fluctuation of the RSMI algorithm seen in Figure 2.19 c) is in a similar range to that of Figure 2.19 e), it is observed that the SIR undulations are more rapid for lower values of α, such as for $\alpha = 0.5$. For the RSMI algorithm of Figure 2.19 e), we observed that the achievable SIR decreases as the reference length is increased. However, for signal power of $6, 12$ and 18dB, decreasing the forgetting factor from $\alpha = 1.0$ to $\alpha = 0.5$ does not affect the convergence rate of the algorithms, as seen by comparing Figures 2.17 and 2.19.

Figure 2.20 characterises the achievable performance of the RSMI and RLS algorithms of Sections 2.4.1.3 and 2.4.1.2, respectively, with respect to the variation of the forgetting factor ranging from $\alpha = 0.9$ to $\alpha = 1.0$ for a reference sequence length of 256 bits. Similar to our observations in the AWGN scenario, the slope of the RLS scheme's response is less steep and at each corresponding α value, the level of interference rejection achieved by the RLS algorithm is higher or at least identical

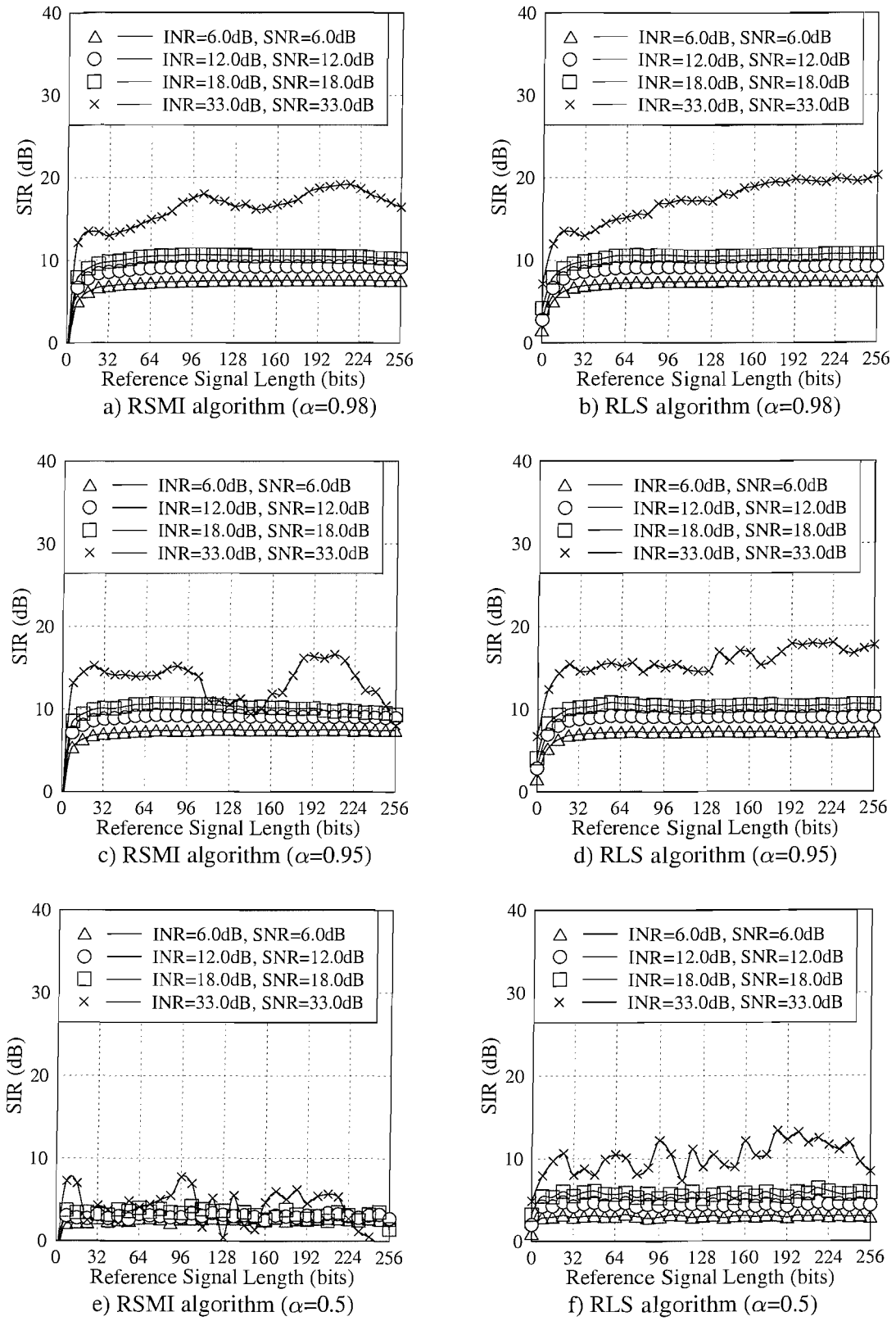
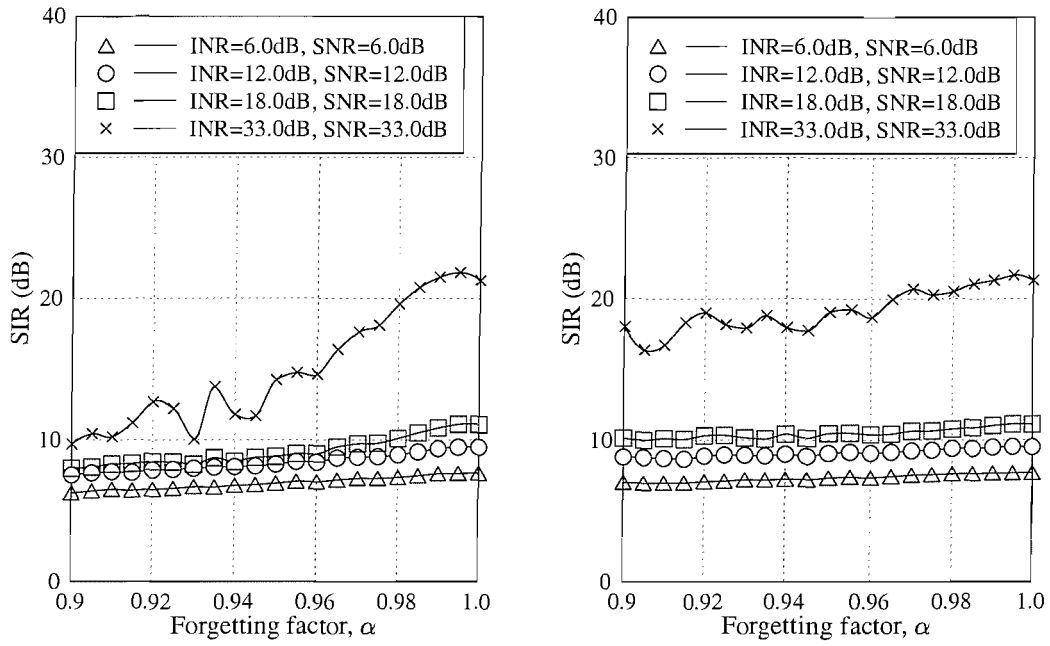


Figure 2.19: The Signal-to-Interference Ratio (dB) achieved with respect to reference signal length using the **RSMI** algorithm of Section 2.4.1.3 for a) $\alpha = 0.98$, c) $\alpha = 0.95$ and e) $\alpha = 0.5$ and the **RLS** algorithm of Section 2.4.1.2 for b) $\alpha = 0.98$, d) $\alpha = 0.95$ and f) $\alpha = 0.5$ in conjunction with different sets of equal desired and interfering signal power (i.e. for **SNR=INR**) over the **flat Rayleigh fading channel**.



a) RSMI algorithm

b) RLS algorithm

Figure 2.20: The Signal-to-Interference Ratio (dB) as a function of the **forgetting factor** α using a) the **RSMI** algorithm of Section 2.4.1.3 and b) the **RLS** algorithm of Section 2.4.1.2 in conjunction with different sets of equal desired and interfering signal powers (i.e. for **SNR=INR**) and a **reference length of 256 bits** over the **flat Rayleigh fading channel**.

to that of the RSMI algorithm. Based on our earlier observations of Figures 2.18 and 2.19, we may conclude that the RSMI algorithm is more sensitive to the choice of the forgetting factor α, particularly when a longer reference sequence is used at a high signal power. This may be related to the recursive calculation of the estimated inverse auto-correlation matrix $\hat{\mathbf{R}}^{-1}$ defined in Equation (2.93), where we have the variable scalar expression of $\frac{\alpha}{1-\alpha}$ in the denominator.

2.6.1.3 Complexity Analysis

Apart from the achievable level of interference rejection (SIR), we also compare the performance of the algorithms in terms of their computational complexity. Figures 2.21 a) and 2.21 b) show the relative complexity of the LMS, NLMS, RLS and RSMI algorithms for multiple antenna aided systems having up to 16 elements in terms of the number of complex additions per iteration and complex multiplications per iteration, respectively. The total number of complex-valued operations required is compared in Figure 2.22 a) and also in Figure 2.22 b) based on the assumption that the Multiply-and-ACcumulate (MAC) instruction is available in the system, also assuming that the Hermitian property

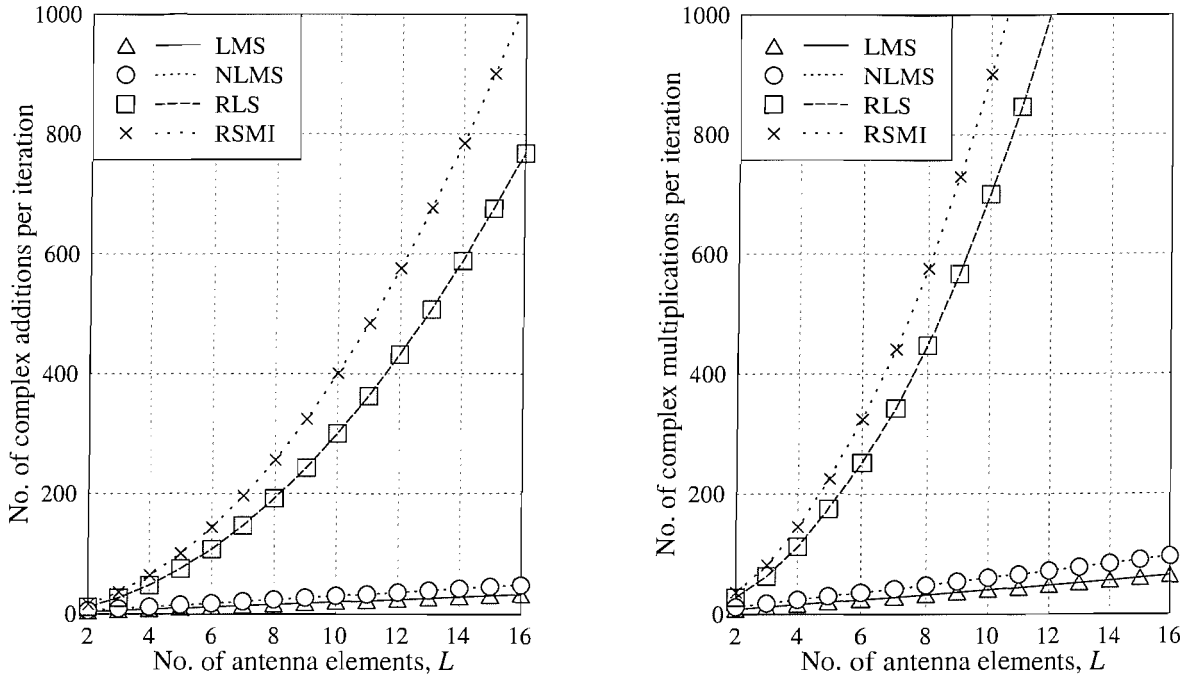


Figure 2.21: Estimated complexity of the LMS, NLMS, RLS and RSMI beamforming algorithms in terms of a) the number of complex additions per iteration and b) the number of complex multiplications per iteration.

of the auto-correlation matrix $\hat{\mathbf{R}}$ is exploited. The total number of mathematical operations plotted in Figure 2.22 is based on the assumption that a reference length of 8-bit was used for the LMS, NLMS algorithms of Section 2.4.1.1 and for the RLS and RSMI algorithms of Sections 2.4.1.2 and 2.4.1.3, respectively. A window length of $2L$ was employed for the DMI algorithm of Section 2.4.1.3.

From Figures 2.21 a) and 2.21 b) we conclude that the RSMI algorithm is the most complex in the set of algorithms investigated, since it requires the highest number of complex additions and complex multiplications per iteration, while the RLS technique is the second most complex, followed by the NLMS and finally the LMS method. In general, the RLS and RSMI algorithms typically converge faster than the LMS and NLMS algorithms of Section 2.4.1.1, since they are based on the least-squares principle, where the error estimate was based on the sequence of data observed at the time of the estimate. On the other hand, the LMS and NLMS algorithms of Section 2.4.1.1 were based on the mean-square approach, which require a statistical model for determining the auto-correlation $\mathbf{R}$ and cross-correlation $\mathbf{z}$ defined in Equations (2.72) and (2.73), respectively. This implies that

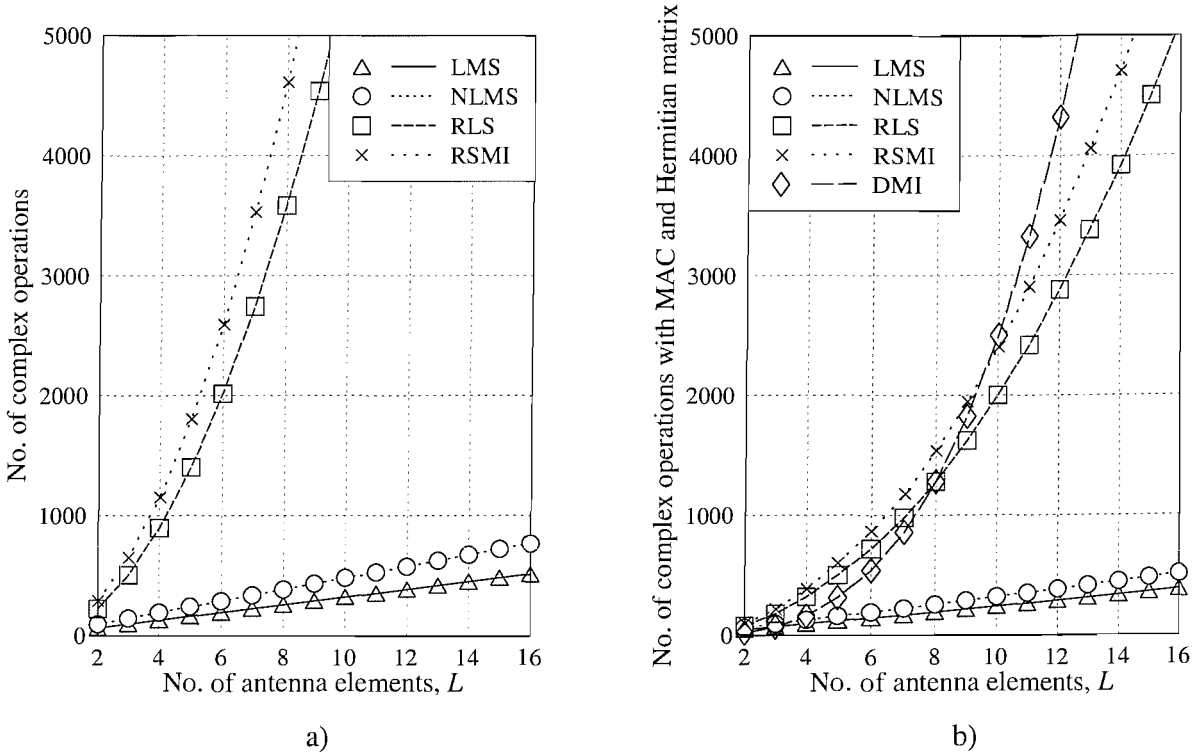


Figure 2.22: Estimated complexity of the various beamforming algorithms in terms of a) the total number of complex operations and b) the total number of complex operations assuming the availability of the MAC operation and the Hermitian property of the auto-correlation matrix $\hat{\mathbf{R}}$. We used 8-bit reference sequence for the LMS, NLMS, RLS and RSMI algorithms, and employed a window length of $N = 2L$ for the DMI algorithm.

the assignment of a similar length of eight bits to the reference sequence was slightly unfair, but the complexity curves seen in Figures 2.22 a) and 2.22 b) were simply used for the sake of general comparisons.

The plots seen in Figures 2.21 and 2.22 are based on the calculated complexity per iteration tabulated in Table 2.9, except for the DMI algorithm of Section 2.4.1.3, where the calculated complexity was based on assuming $N = 2L$ iterations. However, a reference sequence length of eight bits has been employed for the LMS, NLMS, RLS and RSMI algorithms of Figure 2.22 a) and 2.22 b). According to the plots seen in Figures 2.22 a) and 2.22 b), the complexity of the RSMI algorithm is always higher than that of the RLS algorithm. The RSMI and RLS algorithms' complexity was significantly reduced, when assuming the availability of the MAC instruction and upon exploiting that the auto-correlation matrix $\mathbf{R}$ is Hermitian. This can be observed by comparing the plots associated with the

RSMI and RLS algorithms seen in Figure 2.22 a) and Figure 2.22 b), or simply by comparing at the calculated complexity equations shown in Table 2.9. From Figure 2.22 b), we observed that for a low number of antenna elements, i.e. for $L < 8$ the DMI algorithm of Section 2.4.1.3 is less complex, than the RLS algorithm of Section 2.4.1.2, while the complexity is similar when both the RLS and DMI algorithms employed eight antenna elements. The RSMI algorithm of Section 2.4.1.3 however, has a higher complexity than the DMI algorithm, when the number of antenna elements is less than 10.

Complexity	LMS	NLMS	RLS	RSMI	DMI
No. of complex additions	$2L$	$3L$	$3L^2 + L$	$4L^2 - L - 1$	
No. of complex multiplications	$2L + 1$	$3L + 2$	$4L^2 + 4L$	$5L^2 + 3L + 1$	
Total complex operations	$4L + 1$	$6L + 2$	$7L^2 + 5L$	$9L^2 + 2L$	
Total complex operations with Hermitian matrix and MAC assumptions	$3L + 2$	$4L + 4$	$2.5L^2 + 7.5L + 2$	$3L^2 + 6L + 2$	$2.5L^3 + 5L^2 + 2L$

Table 2.9: Summary of the various algorithms' complexity in terms of the different mathematical operations involved per iteration (except for the DMI, in which case the complexity quoted is valid for $2L$ iterations).

2.6.2 Performance of a Three-Element Uniform Linear Array System

An L-element antenna array is said to have degrees of freedom (DOF) given by $L - 1$, where the DOF refers to the number of unconstrained or free weights that can be used to form a beam. This means that the antenna system is capable of forming a beam directed at the signal of interest (SOI) and simultaneously form a null in the direction of $L - 1$ interferers [10]. This section will investigate the interference rejection capabilities of three uniformly spaced antenna elements as the number of interfering sources is increased even beyond the number of elements, in other word beyond the DOF of the system. Our investigations are conducted over both the AWGN and flat Rayleigh fading channels.

In our simulations, the signal of interest (SOI) is assumed to arrive from the direction of 15^o, while the interferers from $-30^o, 60^o, 80^o$ and -70^o viewed from the array normal with reference to the centre element, as shown in Figure 2.23. The simulation parameters include a 256-bit orthogonal reference signal invoked by the RLS algorithm and BPSK modulation.

2.6.2.1 Spatial Selectivity over an AWGN Channel

Figure 2.24 shows several beam patterns of the system receiving two signals, namely the SOI and a single interfering signal. It can be observed that the interferer at any of the four angles is nulled successfully with an attenuation of 38dB, even when the Interference-to-Noise Ratio (INR) is as low

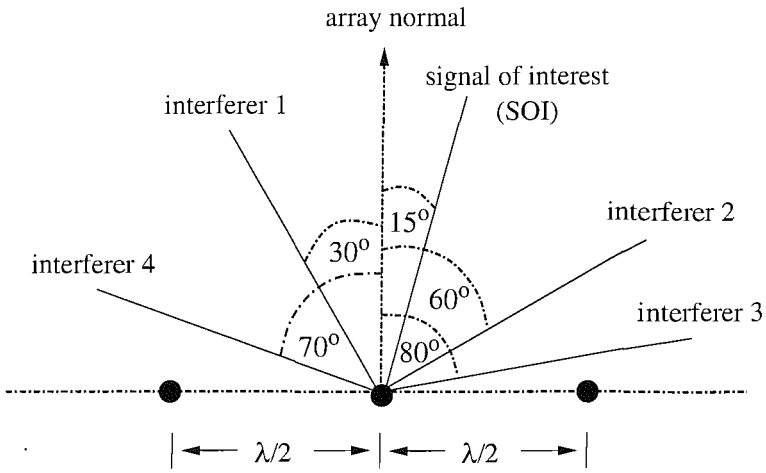


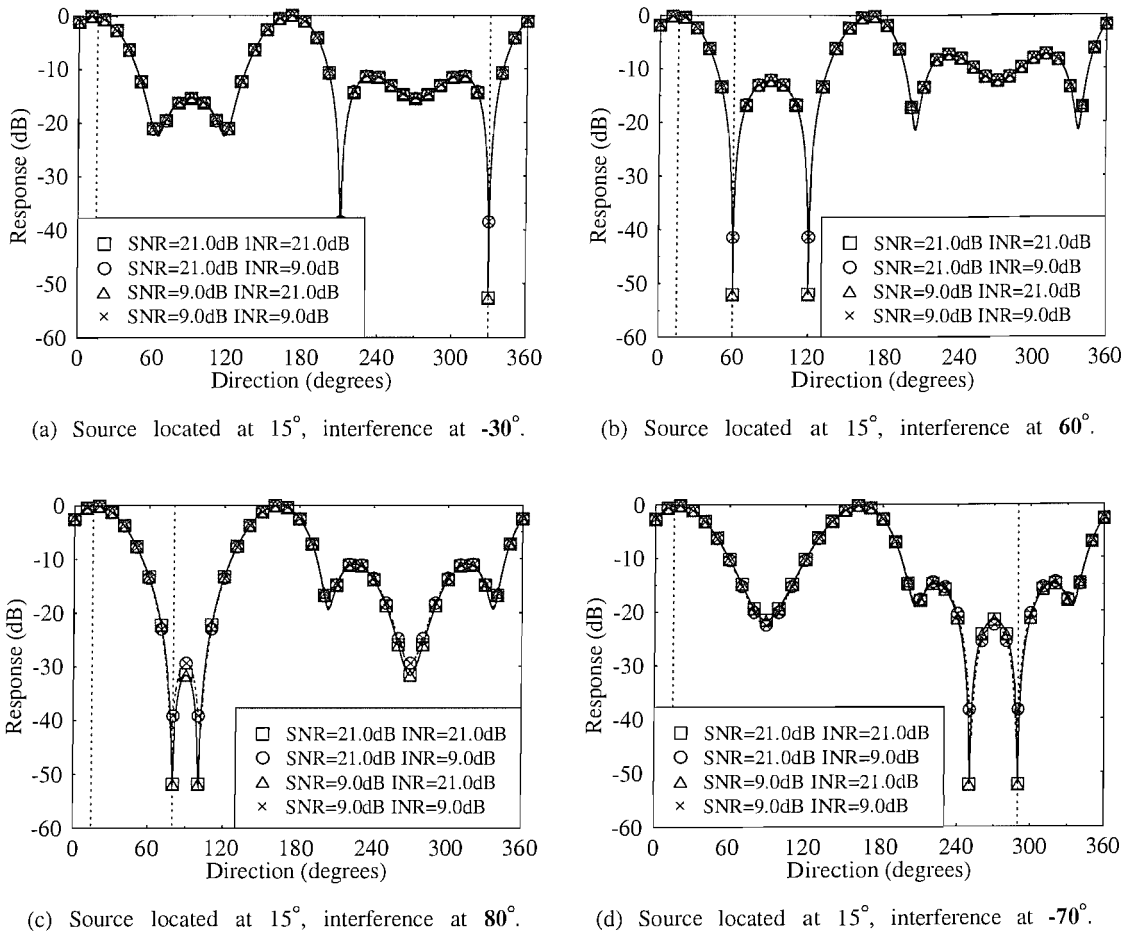
Figure 2.23: Three $\lambda/2-$spaced antenna elements, where the signal of interest arrives from 15^o, while the interfering sources from $-30^o, 60^o, 80^o$ and -70^o.

as 9dB. As mentioned in the previous section, the Signal-to-Noise Ratio (SNR) essentially does not affect the achievable performance curve, while the INR slightly does, since as its value becomes higher, the achievable interference rejection is increased.

When the number of signals impinging on the antenna is equal to the number of array elements, the associated degrees of freedom are exhausted. When we have SNR = INR = 21dB, the achievable interference rejection is in excess of 45dB in the worst-case scenario of having an interferer at 60^o and 80^o, as shown in Figure 2.25 (b), which is only 7 to 8dB lower than the best attenuation response of 53dB obtained in Figure 2.24. It is observed in the figure that as expected, the separation between the desired and interfering signals does affect the associated response. For a pair of interferers that are closely located, as for example in the case of the interferers arriving from 60^o and 80^o seen in the right-hand figure, a strong rejection in excess of 45dB is observed at 60^o and 80^o, but this is achieved at the cost of a slightly degraded response at the desired signal's angle. This phenomenon is also visible, when the interfering sources are better separated, as seen for example for the angles of 80^o and -70^o.

The array response obtained in Figure 2.25 assumes that the interferers are of equal power. For unequal INRs different attenuations are observed. The interferer associated with a higher INR will be better attenuated than a weaker one. This is explicitly shown in Figure 2.26.

Figure 2.27 shows the achievable performance, as four incoming signals impinge upon the 3-element antenna array. Except for the second scenario of Figure 2.27 (b), the attainable performance of the system is substantially reduced, exhibiting the lowest interference rejection of 17dB in the context of the second scenario of Figure 2.27 (a). Note that the DOF of the system in this scenario has been exceeded by one. In Figure 2.28 the same SNR was maintained as in Figure 2.27, while the INR of the interfering signals was reduced from 21dB to 9dB. As expected, some of the nulls recorded at the



(a) Source located at 15°, interference at **-30°**.

(b) Source located at 15°, interference at **60°**.

(c) Source located at 15°, interference at **80°**.

(d) Source located at 15°, interference at **-70°**.

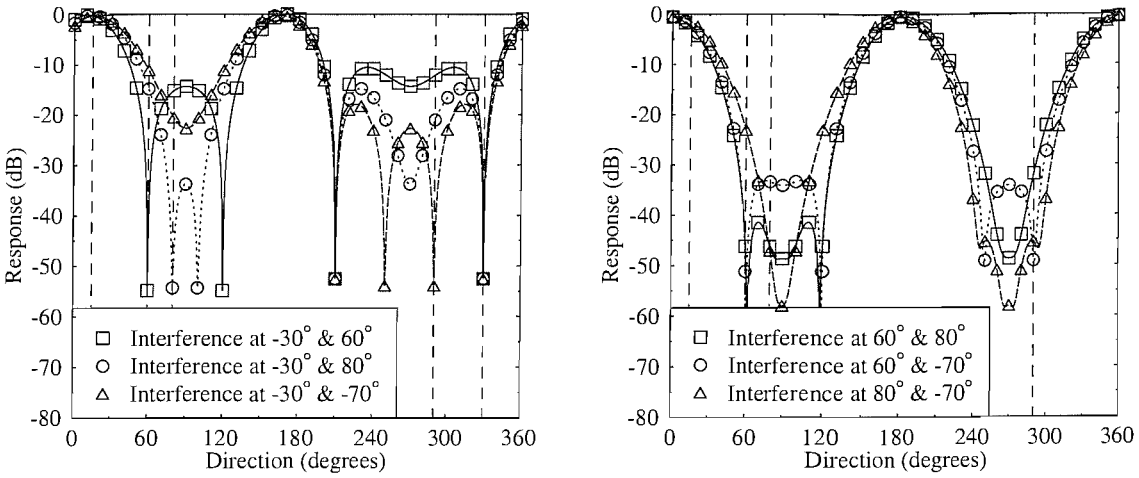
Figure 2.24: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **a single interfering signal** from different directions over the **AWGN channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits.

corresponding angles are not as deep, as those in Figure 2.27.

When the signals shown in Figure 2.23 are all simultaneously impinging upon the antenna array, the resultant beam pattern is shown in Figure 2.29. It is observed from Figure 2.29 (a) that the interfering signals are rejected by at least 17dB for both INR = 21dB and 9dB. If the INR of the interfering signal arriving at 80^o is reduced to 9dB, while maintaining it at 21dB for the rest of the interfering signals, the null at 80^o will become less deep, as shown in Figure 2.29 (b).

2.6.2.2 Spatial Selectivity over a Flat Rayleigh Fading Channel

In our investigations conducted in the context of a flat Rayleigh fading channel, we opted for $\alpha = 0.95$ as the forgetting factor. We aim for observing the array response and compare the corresponding



(a) Source located at 15°, interferers located at **-30°** and **60°**, or **-30°** and **80°**, or **- 30°** and **-70°**.

(b) Source located at 15°, interferers located at **60°** and **80°**, or **60°** and **-70°**, or **80 °** and **-70°**.

Figure 2.25: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **two interfering signals** from different directions over the **AWGN channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits. The **SNR and INRs were 21dB**.

performance to that recorded over the AWGN channel. All the beam patterns plotted in this section are the equivalents of the scenarios described in the AWGN propagation scenario of Section 2.6.2.1.

The array's response recorded for a single interferer is portrayed in Figure 2.30. The rejection of the interferers is substantial, although not quite as high as in the corresponding AWGN scenario of Figure 2.24. Specifically, the lowest interference rejection value was 32dB for an INR as low as 9dB. In comparison to the responses seen in Figure 2.24, a few minor differences occur at some angles other than the interferer's angles. For example, in Figure 2.30 (a) the sidelobe level achieved at $\theta^o = 90^o$ is approximately -20dB, which is 5dB lower than that of the corresponding AWGN scenario of Figure 2.24 (a).

Figure 2.31 and 2.32 portray the responses in conjunction with three signals impinging on the 3-element array. When we have SNR= 21dB, the rejection level is at least 38dB as demonstrated by Figure 2.31 (b). Although this is about 7dB lower than that of the AWGN scenario seen in Figure 2.25 (b), we may consider a rejection higher than 30dB as adequate. Observe that the beam patterns of Figure 2.31 are similar to those of Figure 2.25, except for the scenario when the interferers are arriving from 60^o and 80^o. However, the associated difference is not so dramatic and we are more interested in the achievable performance at the defined angles. For the SNR values portrayed in Figure 2.32, the interferers are nulled successfully.

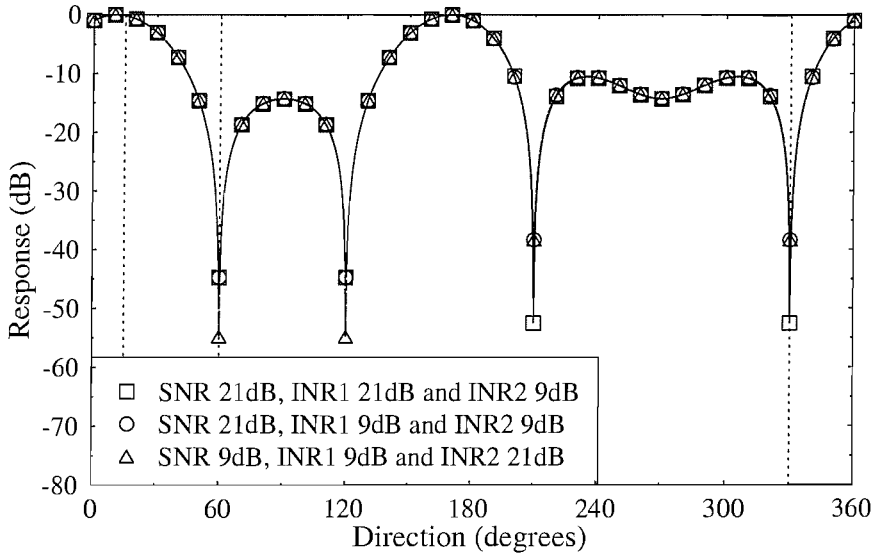
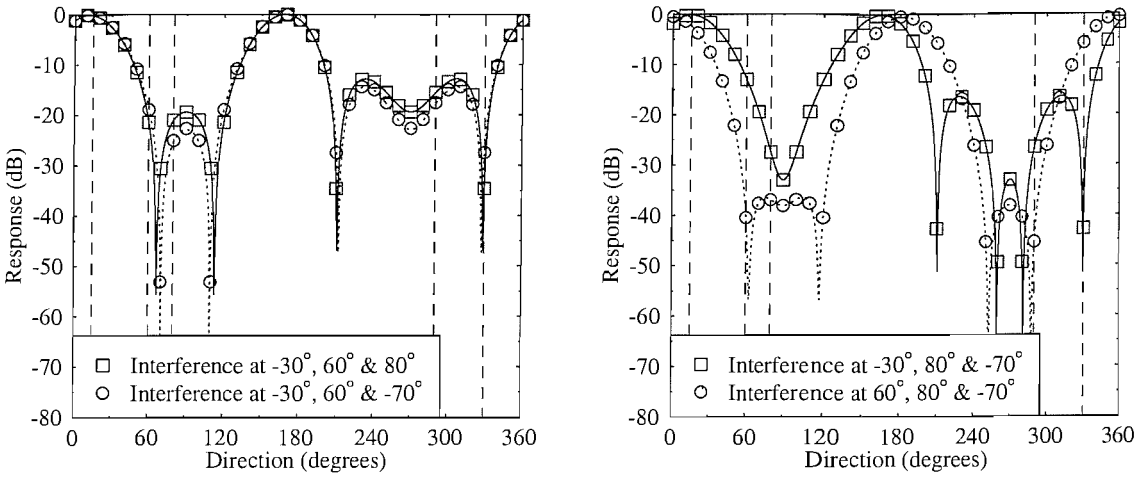


Figure 2.26: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and two interfering signals from -30^o **(interferer 1)** and 60^o **(interferer 2)** over the **AWGN channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits and **unequal desired and interfering signal power**.

As stated before, when the number of signals to be separated exceeded the number of elements in the array, the DOF is also exceeded. Therefore, the depth of the nulls will be reduced. As seen in Figures 2.33 and 2.34 in conjunction with SNRs of 21dB and 9dB, respectively, the lowest rejection was a modest 17dB. This rejection is similar to that achieved over the AWGN channel. Apart from attaining less deep nulls, the beam patterns of Figures 2.33 and 2.34 are similar to the corresponding scenarios in Figures 2.27 and 2.28. The most obvious difference is observed perhaps between Figure 2.28 (b) and Figure 2.34 (b) in conjunction with interferers arriving from 60^o, 80^o and -70^o.

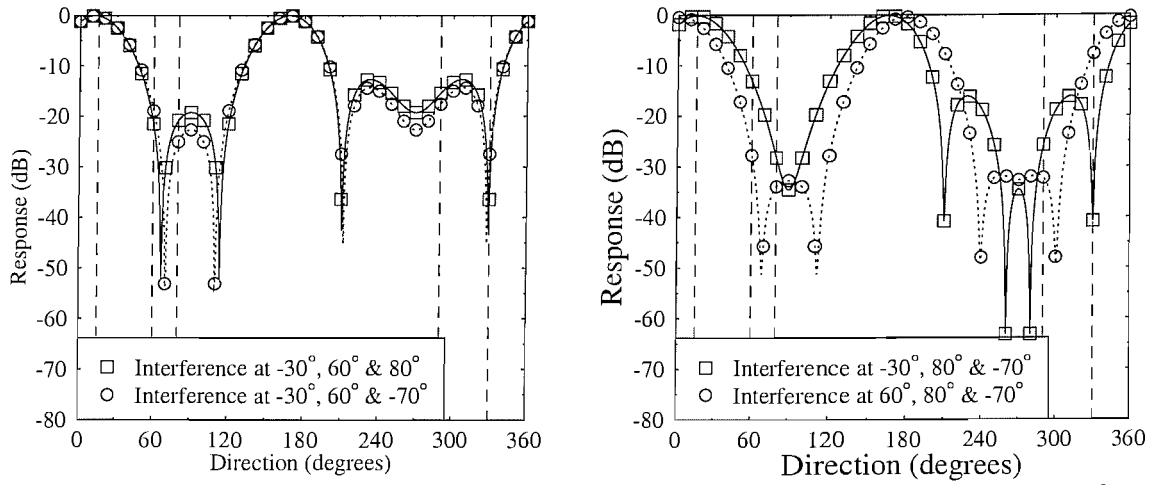
The array response in conjunction with five incoming signals is shown in Figure 2.35. Observe that the lowest rejection is 17dB and all other nulls seen at the defined angles are below -30dB. The beamforming patterns obtained in all scenarios are fairly similar to those obtained for transmission over the AWGN channel and reduced differences are observed, as the corresponding number of signals impinging on the antenna array becomes higher than the number of elements, i.e. when the corresponding DOF is exceeded.



(a) Source located at 15°, interferers located at **-30°**, **60°** and **80°** or **-30°**, **60°** and **-70°**.

(b) Source located at 15°, interferers located at **-30°**, **80°** and **-70°** or **60°**, **80°** and **-70°**.

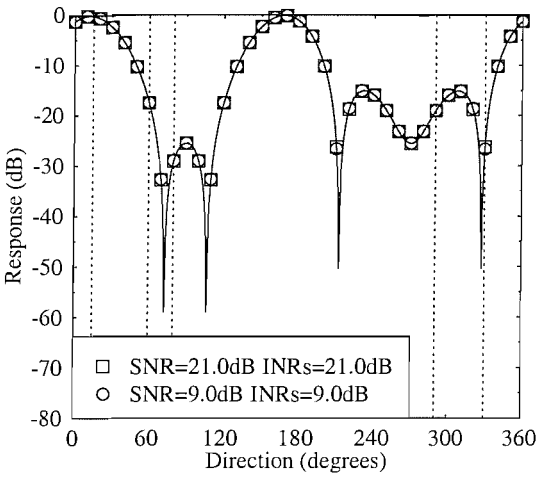
Figure 2.27: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **three interfering signals** from different directions over the **AWGN channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits. The **SNR and INRs were 21dB**.



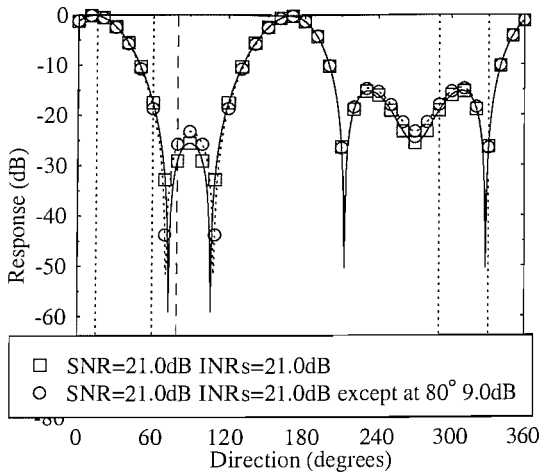
(a) Source located at 15°, interferers located at **-30°**, **60°** and **80°** or **-30°**, **60°** and **-70°**.

(b) Source located at 15°, interferers located at **-30°**, **80°** and **-70°** or **60°**, **80°** and **-70°**.

Figure 2.28: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **three interfering signals** from different directions over the **AWGN channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits. The **SNR was 21dB** while the **INRs were 9dB**.



(a) Equal SNR and INRs of **21.0dB** and **9.0dB**.



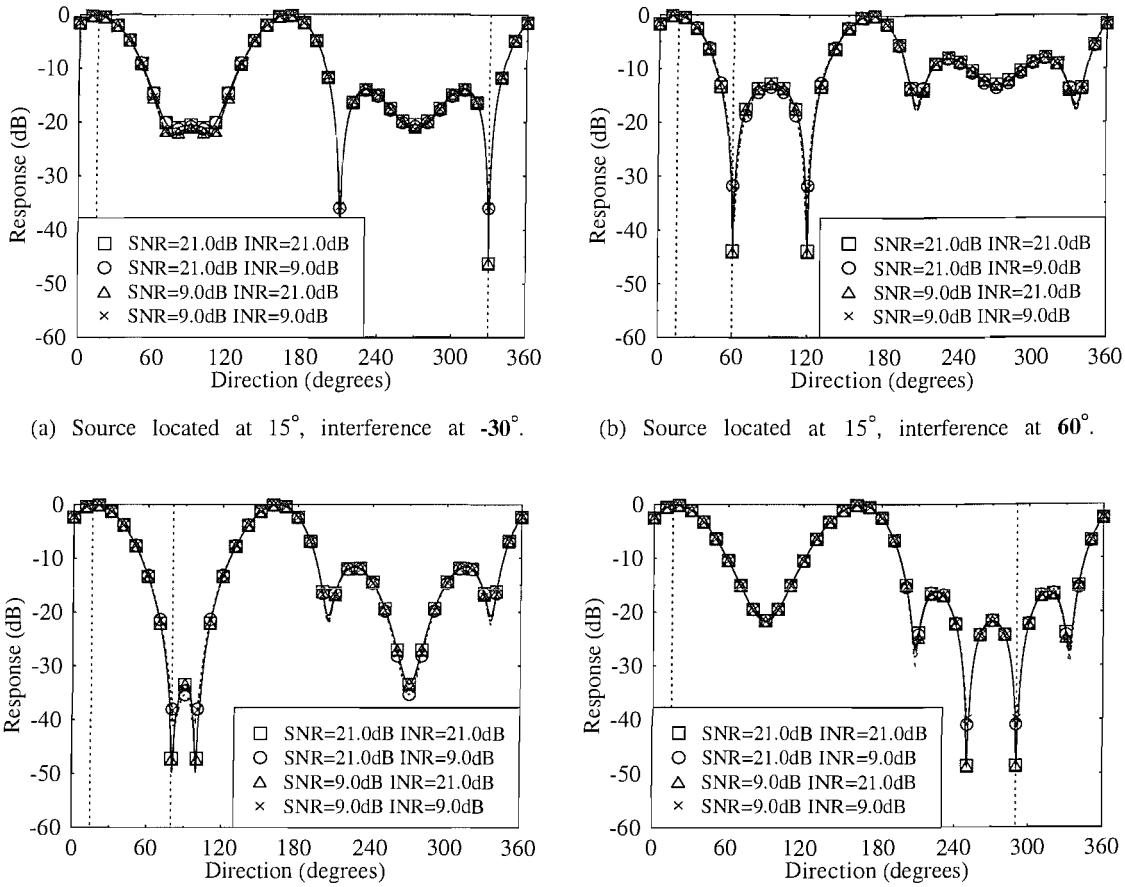
(b) Comparison between equal SNR and INRs of 21.0dB and that with one interferer at **80°** having INR of 9.0dB.

Figure 2.29: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **four interfering signals** from direction $-30^o, 60^o, 80^o$ and -70^o over the **AWGN channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits.

2.7 Conclusions

In Section 2.1.1, we presented the historic background of beamforming. A few preliminary concepts were also introduced, which include our signal model and a simple demonstrative example. For the family of statistically optimum beamforming techniques, it has been shown that regardless of the criterion chosen for optimising the array weights, we will arrive at the same array weight expression, namely the Wiener solution of $\mathbf{w}_{opt} = \mathbf{R}^{-1}\mathbf{z}$. Often the statistics of the array output signal are not completely known and vary over time, thus in Sections 2.4.1.1, 2.4.1.2 and 2.4.1.3 we introduced several algorithms that may result in a set of weights that will converge to the statistically optimum solution. In Section 2.4.1 we used the temporal reference technique, which minimises the error based on the MMSE criterion of Equation (2.70) and in Figures 2.13 to 2.20 we characterised the achievable interference rejection level expressed in terms of the SIR.

When communicating over the AWGN channel, it was observed that the DMI and RSMI algorithms of Section 2.4.1.3 and RLS algorithm of Section 2.4.1.2 give a similar performance. However, they differ in terms of their computational complexity. Specifically, the DMI algorithm imposes the highest order of complexity, i.e. $O(L^3)$, while the RLS and RSMI algorithms exhibit a complexity on the order of $O(L^2)$. This complexity is an order lower than that of the DMI algorithm, but an order higher than that of the LMS and NLMS algorithms of Section 2.4.1.1. For the flat fading channel, as expected, the achievable interference rejection level was significantly lower than that of the corresponding AWGN



(a) Source located at 15°, interference at **-30°**.

(b) Source located at 15°, interference at **60°**.

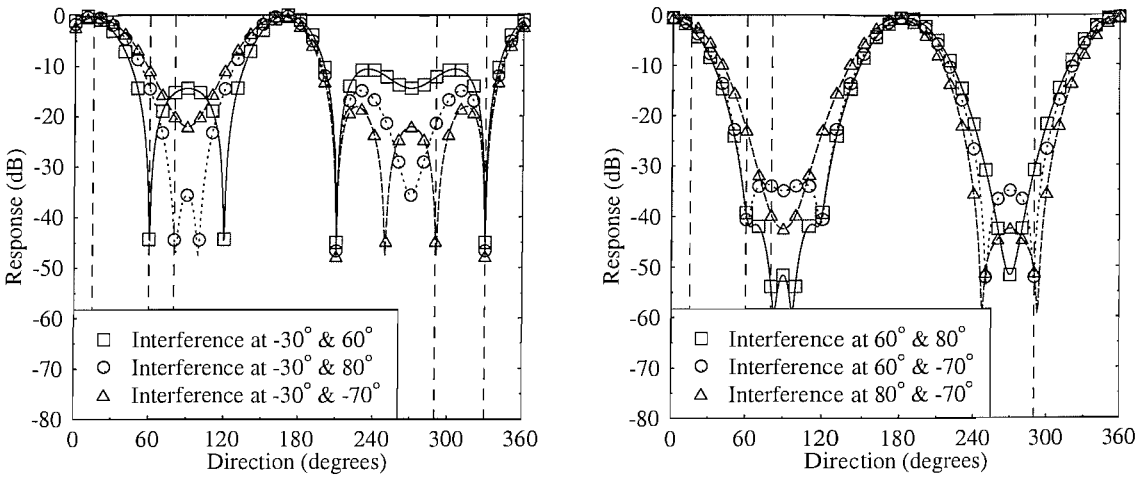
(c) Source located at 15°, interference at **80°**.

(d) Source located at 15°, interference at **-70°**.

Figure 2.30: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **a single interfering signal** from different directions over the **flat Rayleigh fading channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits.

channel. Focusing on the attainable performance of the RLS and RSMI algorithms, it is observed that the performance of the RLS technique of Section 2.4.1.2 is always similar or better than that of the RSMI algorithm of Section 2.4.1.3 for any arbitrary value of the forgetting factor α, i.e. the RSMI algorithm is more sensitive to the value of α. This may be due to the variable scalar $\frac{\alpha}{1-\alpha}$ in Equation (2.93), which is derived as a result of the estimated auto-correlation matrix defined in Equation (2.92).

The beam pattern of the three-element array having 2, 3, 4 and 5 arriving beams investigated, when communicating over the AWGN and the flat Rayleigh fading channel is generally adequate. When the interferers are located at similar angles, typically, a stronger rejection is observed. Interferers exhibiting higher INRs will be nulled more effectively than those arriving at lower INRs. The interference rejection of the system communicating over the AWGN channel is superior in comparison to that over the fading



(a) Source located at 15°, interferers located at **-30°** and **60°**, or **-30°** and **80°**, or **- 30°** and **-70°**. Forgetting factor, α= **0.95**.

(b) Source located at 15°, interferers located at **60°** and **80°**, or **60°** and **-70°**, or **80 °** and **-70°**. Forgetting factor, α= **0.95**.

Figure 2.31: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **two interfering signals** from different directions over the **flat Rayleigh fading channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits. The **SNR and INRs were 21dB**.

channel, although the associated difference is reduced as the array's DOF is exceeded. For the 5-user scenario of Figures 2.29 and 2.35, where the DOF has been exceeded by two, a minimum of 17dB rejection is observed for both the AWGN and flat fading scenarios.

In this chapter, we mainly focused our attention on beamforming schemes based on the MMSE criterion. In the following chapter, we will investigate a novel family of beamforming schemes based on the minimum bit error ratio criterion.

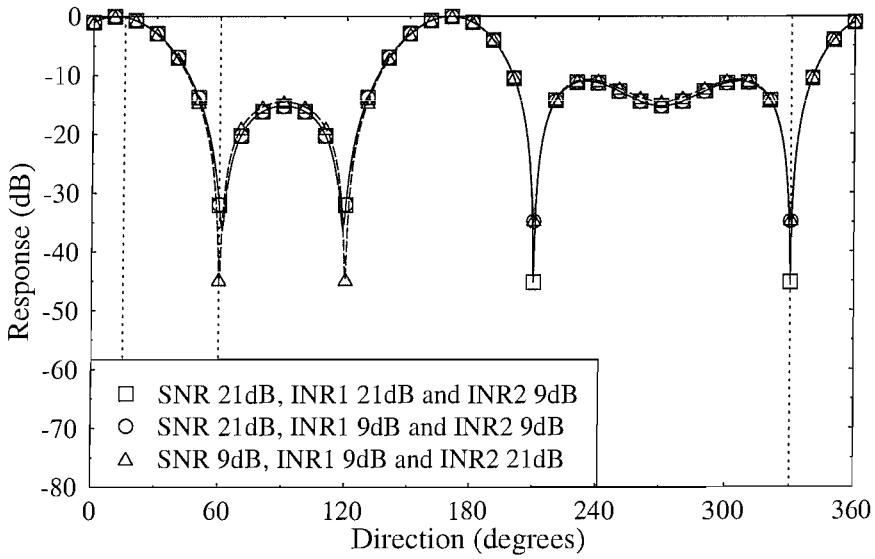
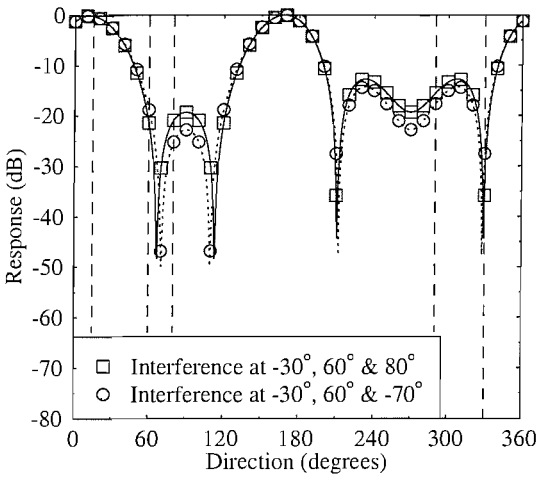
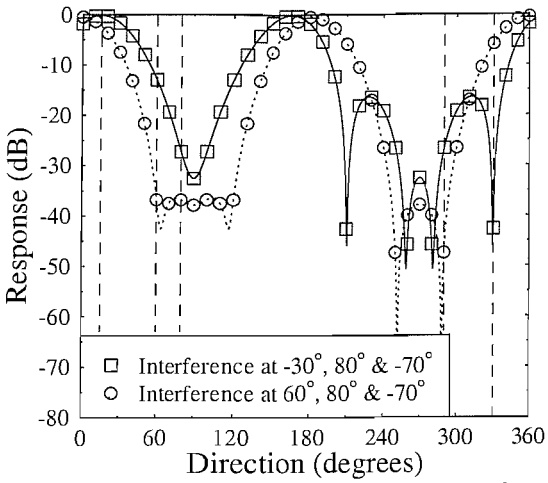


Figure 2.32: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and two interfering signals from -30^o **(interferer 1)** and 60^o **(interferer 2)** over the **flat Rayleigh fading channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits and **unequal incoming power levels**.

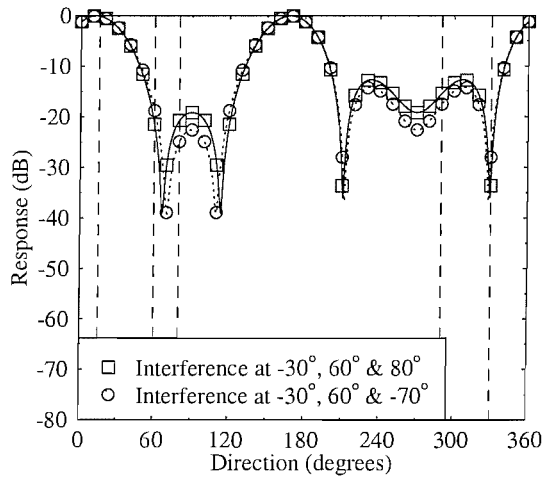


(a) Source located at 15°, interferers located at **-30°**, **60°** and **80°** or **-30°**, **60°** and **-70°**. Forgetting factor, α= **0.95**.

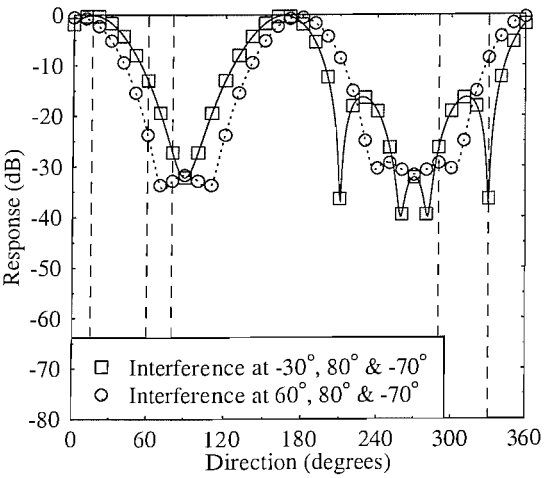


(b) Source located at 15°, interferers located at **-30°**, **80°** and **-70°** or **60°**, **80°** and **-70°**. Forgetting factor, α= **0.95**.

Figure 2.33: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **three interfering signals** from different directions over the **flat Rayleigh fading channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits. The **SNR and INRs were 21dB**.

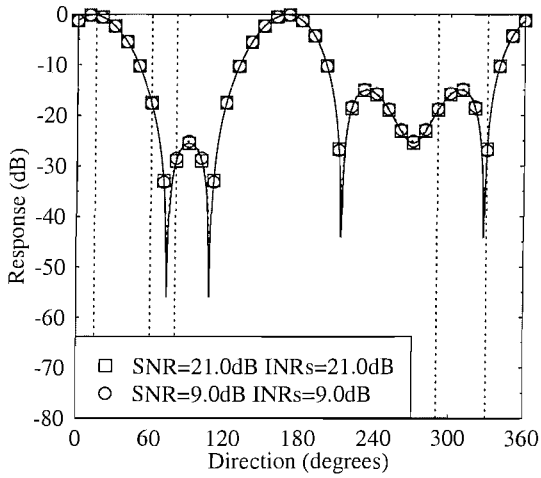


(a) Source located at 15°, interferers located at **-30°**, **60°** and **80°** or **-30°**, **60°** and **-70°**. Forgetting factor, α= **0.95**.

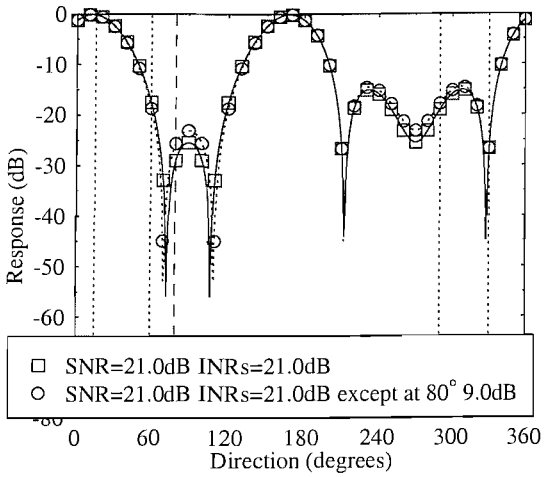


(b) Source located at 15°, interferers located at **-30°**, **80°** and **-70°** or **60°**, **80°** and **-70°**. Forgetting factor, α= **0.95**.

Figure 2.34: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **three interfering signals** from different directions over the **flat Rayleigh fading channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits. The **SNR was 21dB** while the **INRs were 9dB**.



(a) Equal SNR and INRs of **21.0dB** and **9.0dB** with forgetting factor, α= **0.95**.



(b) Comparison between equal SNR and INRs of 21.0dB and that with one interferer at **80°** having INR of 9.0dB. Forgetting factor, α= **0.95**.

Figure 2.35: Beam patterns for a three-element ULA system of $\lambda/2$ element-spacing receiving the **SOI from a direction of** 15^o and **four interfering signals** from direction $-30^o, 60^o, 80^o$ and -70^o over the **flat Rayleigh fading channel**. The **RLS** algorithm was employed in conjunction with a reference sequence length of 256 bits.

Chapter 3

Minimum Bit Error Ratio Beamforming

3.1 Introduction

Adaptive beamforming is capable of separating signals transmitted on the same carrier frequency provided that they originate from different angular directions. In Chapter 2 we have briefly described several classic algorithms, amongst others those applying the Minimum Mean Square Error (MMSE) criterion for deriving the optimal array weights leading to the Wiener-Hopf solution for the beamformer. Specifically, in Section 2.4.1.3 we focused our attention on the Sample Matrix Inversion (SMI) algorithm that adjusts the array weights on a block-by-block adaptation basis, while in Sections 2.4.1.1 and 2.4.1.2 the sample-by-sample adaptation approach of the Least Mean Square (LMS) and Recursive Least Square (RLS) algorithms, respectively, was invoked. Apart from the MMSE criterion, we also outlined in Section 2.3, the maximum Signal-to-Interference Ratio (SIR) and the minimum variance criteria. However, for a communication system, it is the achievable Bit Error Ratio (BER) that really matters. Ideally, the system design should be based on minimising the BER, rather than the Mean Square Error (MSE) [155], although in certain situations the MMSE solution produces a relatively good BER performance.

When applying the MMSE criterion, the achievable performance will be highly dependent on whether the beamformer's output $y(n)$ is exactly identical to the desired response $r(n)$ as seen in the schematic of Figure 2.11. The performance comparison of the MMSE and Minimum Bit Error Ratio (MBER) approaches has been carried out in the context of single-user channel equalisation [196, 197] and also in multiuser detection (MUD) [198, 199]. It was found that in certain situations the MMSE solution is distinctly inferior to the MBER solution. Motivated by these findings, in this chapter we focus our attention on adaptive beamforming techniques based on directly minimising the system's BER. Hence we refer to this approach as adaptive MBER beamforming. Several adaptive

implementations of the theoretical MBER solution have been studied in the literature [196–201]. Table 3.1 summarises the range of past contributions on the design of various MBER detectors [202].

To begin our investigation on MBER beamforming, we will first obtain the BER expression at the beamformer's output in Section 3.3 as a comparison to the MMSE beamforming solution derived in Section 3.2.1. A suitable algorithm in minimising the BER will also be introduced in Section 3.3. Having a theoretical solution, we then proceed with adaptive implementation of MBER beamforming in Section 3.4 using the block-based and sample-by-sample approaches. Finally, in Sections 3.5 and 3.6 we will provide simulation results for characterising the corresponding MBER algorithms. We will commence our discourse by briefly reviewing our signal model in the following section, noting the differences in comparison to the classic beamformer's model outlined in Section 2.2.1.

3.2 Signal Model

In this chapter we will be using the same signal model as in Chapter 2, but for the sake of convenience we revisit some of the equations. As before, the modulation scheme is assumed to be Binary Phase Shift Keying (BPSK), and the channel is assumed to be non-dispersive, which does not induce any intersymbol interference (ISI). We use the narrowband beamforming structure of Figure 2.11.

Let us consider a system, which supports M users, each transmitting on the same carrier frequency f. The baseband signal of user m is formulated as in Equation (2.6):

$$\bar{m}_m(n) = a_m(n)b_m(n), \tag{3.1}$$

where the modulating signal is given by $b_m(n) \in \{\pm 1\}$, the complex-valued channel coefficient $a_m(n)$ models the multiplicative effect of the channel of user m, where the transmitted signal power of user m is unity and therefore $|a_m(n)|^2$ denotes the received signal power of user m, $m = 1, 2, ..., M$, at instant n. The signals received by the lth element of the L-element uniformly spaced linear antenna array of Figure 2.5 are given by:

$$\begin{aligned} x_l(n) &= \sum_{m=1}^{M} \bar{m}_m(n) e^{j2\pi f t_l(\theta_m)} + n_l(n) && (3.2) \\ &= \bar{x}_l(n) + n_l(n), && (3.3) \end{aligned}$$

where $l = 1, 2, ..., L$, $t_l(\theta_m)$ is the relative time delay at array element l for a source direction of θ_m, and $n_l(n)$ is the complex-valued white Gaussian noise having a zero-mean and a variance of $E[|n_l(n)|^2] = 2\sigma_n^2$. Therefore, the desired user's signal to noise ratio is defined as SNR $= |a_1|^2/2\sigma_n^2$, assuming user 1 is the desired user. The interference to noise ratio of user m is given by $\text{INR}_m = |a_m|^2/2\sigma_n^2$ and the desired signal to interference ratio with respect to user m is defined as $\text{SIR}_m = |a_1|^2/|a_m|^2$, for $m = 2, ..., M$. Note that we use a bar '$\bar{}$' for indicating a noise-free signal component. In vectorial

Year	Author	Contribution
'66	Aaron and Tufts [203]	Establishing the interrelationship of intersymbol interference and error probability.
'74	Shamash and Yao [204]	Outlining the structure and characterising the performance of a linear decision feedback equaliser (DFE) based on the minimum error probability criterion.
'96	Chen, Chng, Mulgrew and Gibson [205]	Deriving an MBER solution for the DFE that employs a linear combination of the channel outputs and the past decisions.
'97	Yeh and Barry [206]	Proposing algorithms approaching the performance of the linear and decision-feedback MBER equalisers for binary signalling.
'97	Mandayam and Aazhang [207]	A non-adaptive linear MBER multiuser detector based on gradient optimisation for CDMA systems communicating over narrowband Gaussian channels which do not introduce ISI.
'98	Yeh and Barry [208]	Approximate MBER equalisation for both pulse-amplitude and quadrature amplitude modulation (QAM).
'98	Yeh, Lopes and Barry [209]	Approximate MBER multiuser detection (MUD).
'98	Chen, Mulgrew, Chng and Gibson [200]	Space translation properties and the MBER linear combiner DFE.
'99	Chen and Mulgrew [210]	The minimum-SER linear-combiner decision feedback equaliser.
'99	Psaromiligkos, Batalama and Pados [211]	Adaptive MBER receivers using linear filters in the context of DS-CDMA.
'00	Yeh and Barry [196]	Adaptive MBER equalisation for binary signalling.
'00	Mulgrew and Chen [212]	Stochastic gradient MBER DFEs.
'00	Wang, Lu and Antoniou [213]	Design of a constrained MBER MUD.
'01	Chen, Samingan, Mulgrew and Hanzo [198, 214]	Adaptive MBER linear MUD for DS-CDMA signals transmitted over multipath fading channels.
'01	Mulgrew and Chen [197]	Adaptive MBER DFEs designed for binary signalling.
'01	Samingan, Chen and Hanzo [215]	Adaptive MBER linear MUD for CDMA signals transmitted over multipath channels using 4-QAM.
'03	Chen, Mulgrew and Hanzo [216]	Least bit-error rate adaptive nonlinear equalisers for binary signalling.
'03	de Lamare and Sampaio-Neto [217]	Adaptive MBER decision feedback multiuser receivers for frequency selective fading channels.
'03	Chen, Hanzo and Ahmad [156]	Adaptive MBER beamforming assisted receiver for wireless communications.
'03	Gesbert [218]	A minimum error-rate approach for robust linear MIMO receivers.
'03	Alias, Samingan, Chen and Hanzo [219]	SDMA OFDM employing MBER MUD.

Table 3.1: Contributions on MBER receivers [202].

form, the array input or the received signal vector $\mathbf{x}(n)$ can be expressed as:

$$\begin{aligned}\mathbf{x}(n) &= [x_1(n)\ x_2(n)\cdots x_L(n)]^T && (3.4)\\ &= \bar{\mathbf{x}}(n)+\mathbf{n}(n) && (3.5)\\ &= \mathbf{SAb}(n)+\mathbf{n}(n) && (3.6)\\ &= \mathbf{Gb}(n)+\mathbf{n}(n), && (3.7)\end{aligned}$$

where $\mathbf{n}(n) = [n_1(n)\ n_2(n)\cdots n_L(n)]^T$ has a covariance matrix of $E[\mathbf{n}(n)\mathbf{n}^H(n)] = 2\sigma_n^2\mathbf{I}_L$ with $\mathbf{I}_L$ representing the $(L \times L)$-dimensional identity matrix. Furthermore, $\mathbf{G}$ is the $(L \times M)$-dimensional system matrix given by the product of the users' $(L\times M)$-dimensional steering matrix $\mathbf{S}$ and $(M\times M)$-dimensional diagonal amplitude matrix $\mathbf{A}$, as we have seen in Equation (2.13):

$$\mathbf{G} = [a_1\mathbf{s}_1\ a_2\mathbf{s}_2\cdots a_M\mathbf{s}_M], \tag{3.8}$$

which consists of steering vectors for M sources, each represented as in Equation (2.8), namely as:

$$\mathbf{s}_m = [e^{j\omega t_1(\theta_m)}\ \ e^{j\omega t_2(\theta_m)}\ \ \ldots\ \ e^{j\omega t_L(\theta_m)}]^T, \tag{3.9}$$

where $m = 1, 2, \ldots, M$ and $\mathbf{b}(n)$ is the transmitted bit vector expressed as $\mathbf{b}(n) = [b_1(n)\ b_2(n)\cdots b_M(n)]^T$.

The beamformer's output is given by:

$$\begin{aligned}y(n) &= \mathbf{w}^H\mathbf{x}(n) && (3.10)\\ &= \mathbf{w}^H\bar{\mathbf{x}}(n)+\mathbf{w}^H\mathbf{n}(n) && (3.11)\\ &= \mathbf{w}^H\mathbf{Gb}(n)+\mathbf{w}^H\mathbf{n}(n) && (3.12)\\ &= \bar{y}(n)+e(n), && (3.13)\end{aligned}$$

where $\mathbf{w} = [w_1\ w_2\cdots w_L]^T$ is the complex-valued beamformer weight vector and $e(n)$ is Gaussian distributed having zero-mean and a variance of $E[|e(n)|^2] = 2\sigma_n^2\mathbf{w}^H\mathbf{w}$. When dealing with complex-valued variables, such as the weight w_l, received signal x_l and noise samples n_l, the resultant output $y(n)$ will be also complex-valued. In other words, the output of the beamformer can be expanded as:

$$\begin{aligned}y(n) &= \mathbf{w}^H\mathbf{x}(n)\\ &= \mathbf{w}^H\bar{\mathbf{x}}(n)+\mathbf{w}^H\mathbf{n}(n)\\ &= (\mathbf{w}_R(n)-j\mathbf{w}_I(n))^T(\bar{\mathbf{x}}_R(n)+j\bar{\mathbf{x}}_I(n))+(\mathbf{w}_R(n)-j\mathbf{w}_I(n))^T(\mathbf{n}_R(n)+j\mathbf{n}_I(n)) && (3.14)\\ &= \mathbf{w}_R^T\bar{\mathbf{x}}_R(n)-j\mathbf{w}_I^T\bar{\mathbf{x}}_R(n)+j\mathbf{w}_R^T\bar{\mathbf{x}}_I(n)+\mathbf{w}_I^T\bar{\mathbf{x}}_I(n)+\\ &\quad \mathbf{w}_R^T\mathbf{n}_R(n)-j\mathbf{w}_I^T\mathbf{n}_R(n)+j\mathbf{w}_R^T\mathbf{n}_I(n)+\mathbf{w}_I^T\mathbf{n}_I(n) && (3.15)\\ &= \mathbf{w}_R^T\bar{\mathbf{x}}_R(n)+\mathbf{w}_I^T\bar{\mathbf{x}}_I(n)+\mathbf{w}_R^T\mathbf{n}_R(n)+\mathbf{w}_I^T\mathbf{n}_I(n)+\\ &\quad j(\mathbf{w}_R^T\bar{\mathbf{x}}_I(n)-\mathbf{w}_I^T\bar{\mathbf{x}}_R(n)+\mathbf{w}_R^T\mathbf{n}_I(n)-\mathbf{w}_I^T\mathbf{n}_R(n)) && (3.16)\\ &= \bar{y}_R(n)+e_R(n)+j(\bar{y}_I(n)+e_I(n)) && (3.17)\\ &= y_R(n)+jy_I(n), && (3.18)\end{aligned}$$

where we have $y_R(n) = \Re[y(n)]$, $y_I(n) = \Im[y(n)]$, $e_R(n) = \Re[e(n)]$ and $e_I(n) = \Im[e(n)]$. The subscripts R and I denote the real and imaginary parts, respectively. We estimate the desired user's transmitted bit $b_d(n)$ as follows:

$$\hat{b}_d(n) = \begin{cases} +1, & y_R(n) > 0, \\ -1, & y_R(n) \leq 0, \end{cases} \tag{3.19}$$

where we only take the real part of the beamformer's output as the decision variable. We are basically interested in the sign of the decision variable in order to decide, whether the estimated bit is +1 or -1, such that:

$$\hat{b}_d(n) = \text{sgn}(y_R(n)), \tag{3.20}$$

where $\text{sgn}(\cdot)$ is the signum function and $\hat{b}_d(n)$ is the estimated bit at signalling instant n.

3.2.1 MMSE Beamforming Solution

Previously, we determined the beamforming weight vector by minimising the MSE term of $E[|r(n) - y(n)|^2]$ between the desired response and the beamformer's output as seen in Equation (2.42). In the context of the MBER technique, we focused our attention on the transmitted bits $\mathbf{b}(n)$ and the system matrix $\mathbf{G}$, which we introduced in Equations (2.13) and (2.15). Commencing with Equation (2.42) employing the subscript d for the desired user and replacing the desired response $r(n)$ with the desired user's transmitted bit $b_d(n)$, we derived the MMSE solution for uncorrelated transmitted bits and noise samples as follows:

$$\begin{aligned} \text{MSE} &= E[|b_d(n) - y(n)|^2] && (3.21) \\ &= \sigma_d^2 - 2\mathbf{w}^H E[\mathbf{Gb} + \mathbf{n}]b_d^* + \mathbf{w}^H E[(\mathbf{Gb} + \mathbf{n})(\mathbf{Gb} + \mathbf{n})^H]\mathbf{w} && (3.22) \\ &= \sigma_d^2 - 2\mathbf{w}^H \mathbf{g}_d + \mathbf{w}^H(\sigma_d^2 \mathbf{G}\mathbf{G}^H + 2\sigma_n^2 \mathbf{I}_L)\mathbf{w}. && (3.23) \end{aligned}$$

This MSE expression was then minimised by taking its derivative with respect to the conjugate of the array weight vector element w^* [166] and setting it to zero:

$$\nabla_{\mathbf{w}}\text{MSE} = \frac{\partial}{\partial \mathbf{w}^*}\text{MSE} = -2\mathbf{g}_d + 2(\sigma_d^2 \mathbf{G}\mathbf{G}^H + 2\sigma_n^2 \mathbf{I}_L)\mathbf{w} = 0, \tag{3.24}$$

yielding a closed-form solution for the array weight vector $\mathbf{w}$ as:

$$\mathbf{w}_{\text{MMSE}} = \left(\sigma_d^2 \mathbf{G}\mathbf{G}^H + 2\sigma_n^2 \mathbf{I}_L\right)^{-1} \mathbf{g}_d, \tag{3.25}$$

where the vector $\mathbf{g}_d$ is the desired user's column in the system matrix $\mathbf{G}$, and where we may have $1 \leq d \leq M$. Since we deal with BPSK and always assume that the transmitted bit obeys $b_d(n) \in \{\pm 1\}$, the signal variance will always be equal to unity, i.e. we have $\sigma_d^2 = 1$ and therefore we can simplify Equation (3.25) to [155]:

$$\mathbf{w}_{\text{MMSE}} = \left(\mathbf{G}\mathbf{G}^H + 2\sigma_n^2 \mathbf{I}_L\right)^{-1} \mathbf{g}_d. \tag{3.26}$$

Alternatively, the derivation of the MMSE solution could also commence by directly differentiate the Mean Square Error (MSE) expression of Equation (3.21) and equating it to zero, such that we have:

$$\begin{aligned}
\nabla_{\mathbf{w}}\text{MSE} &= 0 \\
\nabla_{\mathbf{w}} E[|b_d(n) - y(n)|^2] &= 0 \qquad (3.27)\\
2E[(b_d(n) - \mathbf{w}^H(\mathbf{Gb}+\mathbf{n}))(\mathbf{Gb}+\mathbf{n})] &= 0 \qquad (3.28)\\
2E[(\mathbf{Gb}+\mathbf{n})b_d^*(n)] - 2E[(\mathbf{Gb}+\mathbf{n})(\mathbf{Gb}+\mathbf{n})^H\mathbf{w}] &= 0 \qquad (3.29)\\
2\mathbf{g}_d - 2(\sigma_d^2\mathbf{G}\mathbf{G}^H + 2\sigma_n^2\mathbf{I}_L)\mathbf{w} &= 0 \qquad (3.30)\\
\mathbf{w}_{\text{MMSE}} &= \left(\sigma_d^2\mathbf{G}\mathbf{G}^H + 2\sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{g}_d. \qquad (3.31)
\end{aligned}$$

In general, the system matrix $\mathbf{G}$ is unknown [155]. However, the MMSE solution can be readily realised for example, by using either the block-data based adaptive Sample Matrix Inversion (SMI) algorithm or the sample-by-sample adaptation based LMS algorithm, which is a derivative of the stochastic gradient approach, as explained in Chapter 2. Theoretically, the array weight solution of Equation (3.25) or (3.31) is the optimum array weight based on the MMSE criterion. This set of weight values will be used as our benchmarker, characterising the MMSE solution.

The array factor, which describes the response of the beamformer to the source arriving at angle θ_m is given by:

$$F(\theta_m) = \sum_{l=1}^{L} w_l e^{-j\omega t_l(\theta_m)}, \qquad (3.32)$$

where $\omega = 2\pi f$. This is similar to taking the Discrete Fourier Transform (DFT) of the beamformer weights. Previously in Chapter 2 we were concerned only with the magnitude of $F(\theta_m)$ in characterising the performance of a beamformer. However, it is desirable to characterise both the magnitude and phase of $F(\theta_m)$. When determining the probability of error, the ultimate factor is the probability density function (PDF) of the beamformer's output, which fully characterises the BER performance of the beamformer [154–156].

Having determined the closed-form expression of the MMSE beamforming solution, let us next concentrate our attention on the derivation of the MBER beamforming solution.

3.3 MBER Beamforming Solution

As opposed to the MMSE solution, we now focus our attention on the Minimum BER (MBER) criterion invoked for finding the optimum beamforming weight solution. The MBER criterion involves calculating the probability that the decision function considered is in error. In reference to Equation (3.19), we will register a transmission error every time we received a logical zero or a negative

real component of the output, i.e. when we encounter $y_R(n) \leq 0$ having transmitted a logical one bit ($b_d(n) = +1$) or simply when we have $\hat{b}_d(n) \neq b_d(n)$. The BER, which is interchangeably referred to as the probability of error of the signed decision function of:

$$y_s = \text{sgn}(b_d(n))y_R, \tag{3.33}$$

according to the decision regime defined in Equation (3.20) as a function of the weight vector $\mathbf{w}$ can be represented as:

$$P_E(\mathbf{w}) = P(\text{sgn}(b_d(n))y_R \leq 0). \tag{3.34}$$

Before proceeding further we note that in conjunction with M incoming signals and applying the BPSK modulation scheme, there will be $N_b = 2^M$ number of possible transmitted bit sequences of $\mathbf{b}(n)$, which we represent as $\mathbf{b}_q$, where $1 \leq q \leq N_b$. Assuming that we are interested in the first signal $\bar{m}_1$ the desired user's transmitted bit will always be the first bit in the sequence, hence we will refer to it as $b_{q,1}$. Distinct values of the array input signal $\bar{\mathbf{x}}(n)$ will be the signals defined by:

$$\bar{\mathbf{x}}_q = \mathbf{G}\mathbf{b}_q, \tag{3.35}$$

where again, we have $1 \leq q \leq N_b$. Having only two possible values of $b_{q,d}(n)$, the received signals can be separated into two categories, depending on the desired user's transmitted bit $b_{q,d}(n)$ such that:

$$\bar{\mathbf{x}}_q^{(+)} \quad \text{for} \quad b_{q,d}(n) \quad = \quad +1 \quad \text{and} \tag{3.36}$$

$$\bar{\mathbf{x}}_q^{(-)} \quad \text{for} \quad b_{q,d}(n) \quad = \quad -1. \tag{3.37}$$

When using a notation customary in conjunction with mathematical sets, these can be expressed as [155]:

$$\mathcal{X}^{(\pm)} \triangleq \{\bar{\mathbf{x}}_q^{(\pm)} \in \mathcal{X} : \ b_{q,d}(n) = \pm 1\}, \tag{3.38}$$

where we have $\mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{G}\mathbf{b}_q, \ 1 \leq q \leq N_b\}$. In a similar fashion, the noiseless output of the beamformer $\bar{y}(n)$ will be one of the N_b signals defined by:

$$\bar{y}_q = \mathbf{w}^H \bar{\mathbf{x}}_q, \tag{3.39}$$

whose real part we denote as $\bar{y}_{R,q}$, such that we have:

$$\bar{y}_{R,q} = \Re[\bar{y}_q], \tag{3.40}$$

that may belong to one of the two categories with reference to the desired user's transmitted bit $b_{q,d}(n)$ i.e. $\bar{y}_{R,q}^{(+)}$ for $b_{q,d}(n) = +1$ and $\bar{y}_{R,q}^{(-)}$ for $b_{q,d}(n) = -1$. Alternatively, the corresponding signals are those in the set defined by [155]:

$$\mathcal{Y}_R^{(\pm)} \triangleq \{\bar{y}_{R,q}^{(\pm)} \in \mathcal{Y}_R : \ b_{q,d}(n) = \pm 1\}, \tag{3.41}$$

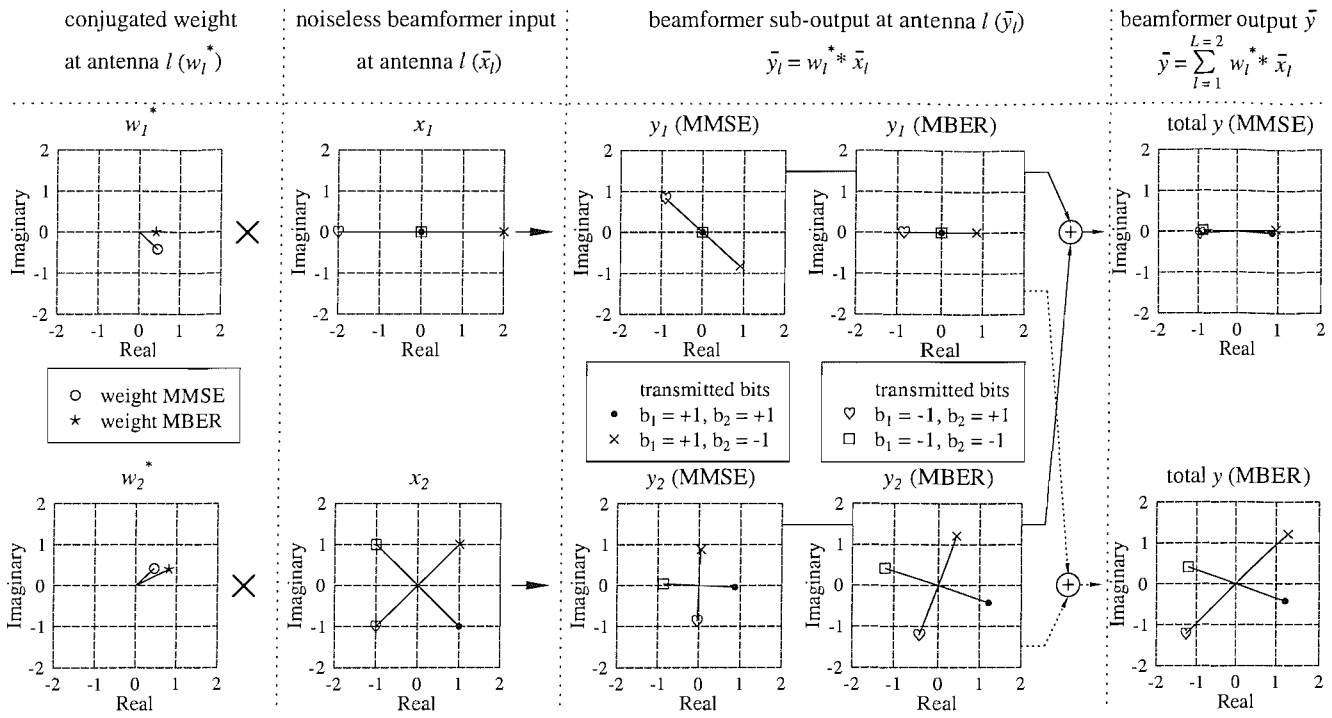


Figure 3.1: Composition of the noiseless beamformer output $\bar{y}$ defined in Equation (3.39) for both the MMSE and MBER beamformers at SNR = INR = 10dB. Both the noiseless input $\bar{x}_l$ and the sub-output $\bar{y}_l$ observed at the corresponding antenna element l are defined in Equations (2.7) and (2.17), respectively. The AOA of the desired and interfering user was 0^o and 30^o, respectively.

where $\mathcal{Y}_R \triangleq \{\bar{y}_{R,q} = \Re[\bar{y}_q],\ 1 \leq q \leq N_b\}$. Note that our definition of beamforming here refers to linear beamforming [155]. We assume that $\mathcal{X}^{(+)}$ and $\mathcal{X}^{(-)}$ are linearly separable, such that there exists an array weight vector $\mathbf{w}$, such that the two real output sets $\mathcal{Y}_R^{(+)}$ and $\mathcal{Y}_R^{(-)}$ are completely separable by a linear decision boundary, in our BPSK scenario we have $y_R = 0$.

Figure 3.1 shows the procedure of composing the noiseless beamformer output $\bar{y}_q$ of Equation (3.39). In the figure $\bar{y}_q$ is simply written as 'total y'. For $M = 2$ BPSK users and an $L = 2$-element antenna array, the total number of possible transmitted bit sequences $\mathbf{b}_q$ is given by $N_b = 2^M$, which results in $N_b = 4$, i.e. we have $1 \leq q \leq N_b = 4$ possible 2-bit combinations. For example, for the specific scenario, when the desired user always transmits a logical one, i.e. for $b_1 = +1$ assuming that user 1 is the desired user, the corresponding noiseless beamformer output $\bar{y}_q^{(+)}$ was denoted by the markers cross and dot in the final MMSE and MBER phasor plots, respectively, seen at the right of Figure 3.1. By contrast, for $b_1 = -1$, the corresponding noiseless beamformer output $\bar{y}_q^{(-)}$ was denoted by the markers square and heart, respectively. In Figure 3.1 both the MMSE and MBER beamformers' noiseless outputs were succesfully separated by the vertical division boundary at $y_R = 0$. However, it is observed from Figure 3.1 that the MMSE beamformer minimises the distance between the desired user's transmitted bit $b_1 = \pm 1$ and the received noiseless output y, hence the corresponding markers are close to ± 1. By contrast, the MBER beamformer searches for the specific weights $\mathbf{w}$ that will

produce the received noiseless outputs y, that are as far as possible from the decision boundary at $y_R = 0$, rather than being as close to $b_1 = \pm 1$ as possible. More explicitly, the MBER beamformer attempts to keep all the four legitimate constellation points as far from $y_R = 0$, as possible. In other words, the MBER beamformer concentrates on reducing the BER of the received noiseless output, and the appropriate weights derived for the sake of achieving this are typically different from the MMSE weight solution, as will be shown in the context of our simulation results in Sections 3.5 and 3.6.

The probability of error P_E in a single user scenario of $M = 1$, in conjunction with BPSK modulation is given by [220]:

$$P_E(y|x_1) = \int_{x_1}^{\infty} p(y|x_1)\, dy, \tag{3.42}$$

where x_1 is one of the two legitimate transmitted waveforms in the set of $x = \{x_1, x_2\}$, as plotted in Figure 5.1 (a) in [220]. Considering that the conditional probability of receiving any particular analogue sample y given that x_1 or x_2 was transmitted is quantified by the Gaussian PDFs seen in Figure 5.1 (a), which can be described by [220]:

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(\frac{-(y-x)^2}{2\sigma_n^2}\right), \tag{3.43}$$

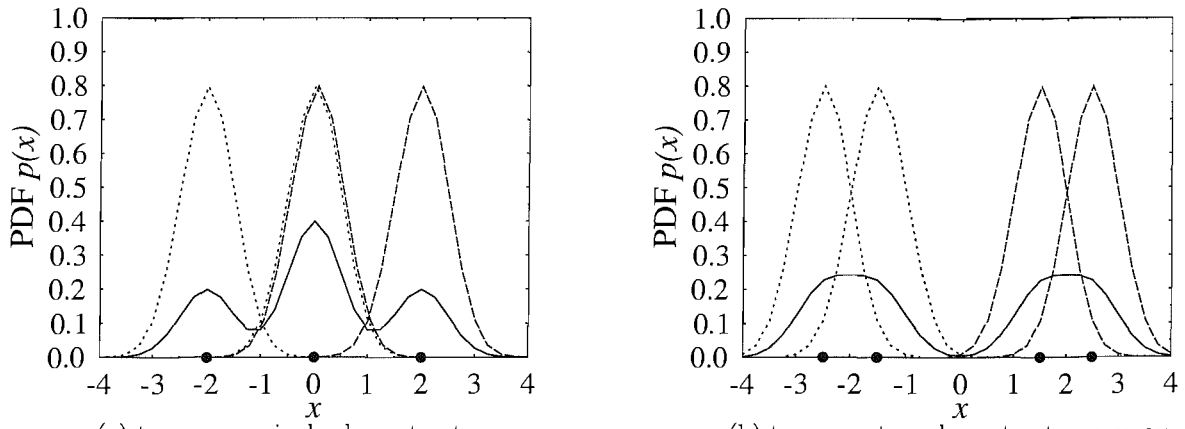
where $x = x_1$ or x_2 is the mean and σ_n^2 is the variance, Equation (3.42) corresponds to the Gaussian Q-function of $Q(x_1)$, which may be interpreted as the probability of a Gaussian noise sample exceeding the value x_1. In other words, it represents the complementary cumulative density function (CDF) of the Gaussian distribution [220]. Assuming that $x_1 = -x_2$, the probability that the noise can carry x_1 across a threshold of $x = 0$ is equal to that of x_2 being corrupted in the negative direction. Hence, assuming that we have $p(x_1) = p(x_2) = 0.5$, the overall error probability is given by [220]:

$$\begin{aligned} P_E &= p(x_1)Q(x_1) + p(x_2)Q(x_2) && (3.44)\\ &= \frac{1}{2}Q(x_1) + \frac{1}{2}Q(x_1) && (3.45)\\ &= Q(x_1). && (3.46) \end{aligned}$$

Let us now proceed by considering a scenario, when there are several users, i.e. when we have $M > 1$. For simplicity, we assume that there are $M = 2$ equal-power users transmitting over an AWGN channel to a single-antenna receiver. Then, there will be $2^M = 4$ possible transmitted BPSK sequences, which can be written in the following set:

$$\mathbf{b}_q = \left\{ \begin{bmatrix} +1 \\ +1 \end{bmatrix}, \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ +1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right\}, \tag{3.47}$$

which results in the noiseless transmitted waveforms of x in the set of $x = \{-2, 0, 0, 2\}$, when assuming a unity channel gain and noiseless transmitted waveforms. Similar to the previous single user case, the resultant PDF is the superposition of the appropriately positioned Gaussian PDFs, as portrayed in Figure 3.2 (a). The resultant mathematical expression is similar to that of Equation (3.42), except



(a) two-user, a single element antenna (b) two-user, two-element antenna array

Figure 3.2: PDF of the receiver output; the dotted and dashed lines represent the PDF of the individual noisy samples, while the solid line represents the resultant PDF constituted of all the legitimate points.

that x_1 now represents a condition associated with $b_{q,d} = +1$ and x_2 is associated with $b_{q,d} = -1$, where the subscript d denotes one of the users, i.e. we may have $1 < d < M$.

In the case of beamforming we use a multi-element antenna, i.e. $L > 1$. The simplest case is that of considering $L = 2$. As described previously in Section 3.2, the beamformer's output due to the noise-contaminated array input signals $\mathbf{x}$ is given by Equation (3.10), while the real part $\bar{y}_{R,q}$ of the noiseless output of the beamformer belongs to the set described by Equation (3.41). In a similar fashion to the single-element antenna case, the conditional probability of receiving any particular beamformer output y associated with our decision variable y_s, based on Equation (3.33), given that one of the bit sequences $\mathbf{b}_q$ was transmitted is quantified by one of the Gaussian PDFs seen in Figure 3.2 (b), which can be described by:

$$p(\mathrm{sgn}(b_d(n))y_R|\mathbf{b}_q) = \frac{1}{\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\exp\left(-\frac{(\mathrm{sgn}(b_d(n))y_R - \mathrm{sgn}(b_{q,d})\bar{y}_{R,q})^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right), \quad (3.48)$$

where $\mathrm{sgn}(b_{q,d})\bar{y}_{R,q}$ is the mean and $\sigma_n^2\mathbf{w}^H\mathbf{w}$ is the variance, which follows the variance of the real part of the error e_R, based on the following calculation:

$$\begin{aligned}
\sigma_{e_R}^2 &= E[|\Re[\mathbf{w}^H\mathbf{n}]|^2] && (3.49)\\
&= E[|\mathbf{w}_R^T\mathbf{n}_R(n) + \mathbf{w}_I^T\mathbf{n}_I(n)|^2] && (3.50)\\
&= E[(\mathbf{w}_R^T\mathbf{n}_R(n) + \mathbf{w}_I^T\mathbf{n}_I(n))^T(\mathbf{w}_R^T\mathbf{n}_R(n) + \mathbf{w}_I^T\mathbf{n}_I(n))] && (3.51)\\
&= E[(\mathbf{n}_R^T\mathbf{w}_R(n) + \mathbf{n}_I^T\mathbf{w}_I(n))(\mathbf{w}_R^T\mathbf{n}_R(n) + \mathbf{w}_I^T\mathbf{n}_I(n))] && (3.52)\\
&= E[\mathbf{n}_R^T\mathbf{w}_R\mathbf{w}_R^T\mathbf{n}_R(n) + \mathbf{n}_I^T\mathbf{w}_I\mathbf{w}_R^T\mathbf{n}_R(n) + \mathbf{n}_R^T\mathbf{w}_R\mathbf{w}_I^T\mathbf{n}_I(n) + \mathbf{n}_I^T\mathbf{w}_I\mathbf{w}_I^T\mathbf{n}_I(n)] && (3.53)\\
&= E[\mathbf{n}_R^T\mathbf{n}_R\mathbf{w}_R^T\mathbf{w}_R(n) + \mathbf{n}_I^T\mathbf{n}_I\mathbf{w}_I^T\mathbf{w}_I(n)] && (3.54)\\
&= \sigma_n^2\mathbf{w}_R^T\mathbf{w}_R + \sigma_n^2\mathbf{w}_I^T\mathbf{w}_I && (3.55)\\
&= \sigma_n^2\mathbf{w}^H\mathbf{w}, && (3.56)
\end{aligned}$$

where the real and imaginary components are assumed to be uncorrelated, i.e. we have $\mathbf{n}_I^T\mathbf{n}_R(n) = \mathbf{n}_R^T\mathbf{n}_I(n) = 0$ and are of equal variance expressed as $\mathbf{n}_R^T\mathbf{n}_R(n) = \mathbf{n}_I^T\mathbf{n}_I(n) = \sigma_n^2$. Based on the

assumption that the N_b number of vectors $\mathbf{b}$ are equiprobable, the overall probability of $\text{sgn}(b_d(n))y_R$ is given by [198]:

$$p(\text{sgn}(b_d(n))y_R) = \frac{1}{N_b}\sum_{q=1}^{N_b} p(\text{sgn}(b_d(n))y_R|\mathbf{b}_q) \tag{3.57}$$

$$= \frac{1}{N_b\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{q=1}^{N_b}\exp\left(-\frac{(\text{sgn}(b_d(n))y_R - \text{sgn}(b_{q,d})\bar{y}_{R,q})^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right). \tag{3.58}$$

Assuming that both logical values of the transmitted bits are equally likely, the conditional PDF of $y_R(n)$ associated with $b_d(n) = +1$ is given by:

$$p(y_R|b_d(n) = +1) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{q=1}^{N_{sb}}\exp\left(-\frac{\left(y_R - \text{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right), \tag{3.59}$$

where $\bar{y}_{R,q}^{(+)} \in \mathcal{Y}_R^{(+)}$ and N_{sb} is the number of distinct sets of bits transmitted with $b_{q,d}$ fixed to a specific value (in this case $b_{q,d} = +1$), which should be half of N_b, i.e. we have $N_{sb} = N_b/2 = 2^{M-1}$, where M is the number of users, since $b_{q,d}$ is either $+1$ or -1.

The probability of error $P_E(\mathbf{w})$ of our decision function defined in Equation (3.33) can be represented by the area under the PDF within the interval of $(-\infty, 0)$, which is given by [169]:

$$P_E(\mathbf{w}) = \int_{-\infty}^{0} p(y_R| + 1)\, dy_R. \tag{3.60}$$

Substituting Equation (3.59) into Equation (3.60), we arrive at:

$$P_E(\mathbf{w}) = \int_{-\infty}^{0}\frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{q=1}^{N_{sb}}\exp\left(-\frac{\left(y_R - \text{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right)dy_R \tag{3.61}$$

$$= \sum_{q=1}^{N_{sb}}\int_{-\infty}^{0}\frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\exp\left(-\frac{\left(y_R - \text{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right)dy_R. \tag{3.62}$$

Upon exploiting that the Gaussian Q-function is given by [7, 221]:

$$Q(z) = \frac{1}{\sqrt{2\pi}}\int_z^{\infty}\exp\left(-\frac{v^2}{2}\right)dv, \tag{3.63}$$

we can simplify the above expression of $P_E(\mathbf{w})$ in terms of the standard Q-function $Q(z)$, having a zero-mean and a standard deviation of unity, which quantifies the probability of $v > z$. Upon representing v as:

$$v = \frac{y_R - \text{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}, \tag{3.64}$$

and its derivative as:

$$dv = \frac{dy_R}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}, \tag{3.65}$$

Equation (3.61) may be expressed as:

$$P_E(\mathbf{w}) = \sum_{q=1}^{N_{sb}} \int_{-\infty}^{-\frac{\mathrm{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}} \frac{1}{N_{sb}\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv \tag{3.66}$$

$$= \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} \frac{1}{\sqrt{2\pi}} \int_{\frac{\mathrm{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv. \tag{3.67}$$

Thus, taking into account Equations (3.63) and (3.64) the BER of the beamformer associated with the weight vector $\mathbf{w}$ is given by:

$$P_E(\mathbf{w}) = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q\left(c_{q,+}(\mathbf{w})\right), \tag{3.68}$$

where we used the shorthand of:

$$c_{q,+}(\mathbf{w}) = \frac{\mathrm{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} = \frac{\mathrm{sgn}(b_{q,d})\Re[\mathbf{w}^H\bar{\mathbf{x}}_q^{(+)}]}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}. \tag{3.69}$$

Note that the BER is invariant to a positive scaling of $\mathbf{w}$, which implies that if we scale the array weight vector by a positive integer α, Equation (3.69) will result in the same value of $c_{q,+}(\mathbf{w})$, which is expressed as:

$$c_{q,+}(\alpha\mathbf{w}) = c_{q,+}(\mathbf{w}). \tag{3.70}$$

This indicates that the BER expression in Equation (3.68) is independent of the positive scaling of the array weight vector $\mathbf{w}$, which results in an infinite number of optimum solutions for $\mathbf{w}_{\mathrm{MBER}}$. Following a similar approach, the BER can also be calculated using $\mathcal{Y}_R^{(-)}$.

The MBER beamforming solution is therefore defined as:

$$\mathbf{w}_{\mathrm{MBER}} = \arg\min_{\mathbf{w}} P_E(\mathbf{w}). \tag{3.71}$$

Unlike for the MMSE approach of Equation (3.25), due to the complex nature of the BER cost function [199], no closed-form solution is available for the MBER approach. In [199], it was pointed out that an attractive practical approach is to solve the MBER solution by employing an iterative strategy based on the steepest descent gradient method [6]. The algorithm, referred to as the steepest descent gradient algorithm operates by updating the array weight vector $\mathbf{w}$ iteratively in the direction opposite to the gradient of the BER cost function. Mathematically, this is formulated as follows:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{d}(n) \tag{3.72}$$

$$= \mathbf{w}(n) - \mu\nabla_{\mathbf{w}}P_E(\mathbf{w}(n)), \tag{3.73}$$

where $\mathrm{d}(n) = -\nabla_{\mathbf{w}}P_E(\mathbf{w}(n))$ represents the negative direction of the gradient vector $P_E(\mathbf{w}(n))$ weighted by the step size μ. We calculate the gradient of $P_E(\mathbf{w})$ using the following identity [169]:

$$\frac{\partial}{\partial t}\int_{a(t)}^{c(t)} f(y)dy = f(c(t))\frac{\partial c(t)}{\partial t} - f(a(t))\frac{\partial a(t)}{\partial t}. \tag{3.74}$$

It can be shown that the gradient of $P_E(\mathbf{w})$ of y_R with respect to $\mathbf{w}$ may be expressed as:

$$\nabla_{\mathbf{w}} P_E(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} P_E(\mathbf{w}) \tag{3.75}$$

$$= \frac{\partial}{\partial \mathbf{w}} \left(\frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} \frac{1}{\sqrt{2\pi}} \int_{c_{q,+}(\mathbf{w})}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv \right) \tag{3.76}$$

$$= \frac{\partial}{\partial \mathbf{w}} \left(\frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-c_{q,+}(\mathbf{w})} \exp\left(-\frac{v^2}{2}\right) dv \right) \tag{3.77}$$

$$= \frac{1}{N_{sb}\sqrt{2\pi}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}} \right) \frac{\partial}{\partial \mathbf{w}} \left(-\frac{\mathrm{sgn}(b_{q,d}) \bar{y}_{R,q}^{(+)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \right) \tag{3.78}$$

$$= -\frac{1}{N_{sb}\sqrt{2\pi}\sigma_n} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}} \right) \mathrm{sgn}(b_{q,d}) \frac{\partial}{\partial \mathbf{w}} \left(\frac{\bar{y}_{R,q}^{(+)}}{\sqrt{\mathbf{w}^H \mathbf{w}}} \right), \tag{3.79}$$

where we have $\bar{y}_{R,q}^{(+)} = \Re[\mathbf{w}^H \bar{\mathbf{x}}_q^{(+)}]$ and $\mathbf{w}$ is a complex-valued vector. Differentiation with respect to a complex-valued variable of $w = w_R + jw_I$ is somewhat more involved than that with respect to a real-valued one. More specifically, the corresponding derivatives are given by Wirtinger calculus [166] as follows:

$$\frac{\partial f(w)}{\partial w} = \frac{1}{2} \left(\frac{\partial f(w)}{\partial w_R} - j \frac{\partial f(w)}{\partial w_I} \right) \tag{3.80}$$

$$\frac{\partial f(w)}{\partial w^*} = \frac{1}{2} \left(\frac{\partial f(w)}{\partial w_R} + j \frac{\partial f(w)}{\partial w_I} \right), \tag{3.81}$$

where we have

$$\frac{\partial w}{\partial w} = 1 \quad \text{and} \quad \frac{\partial w*}{\partial w} = 0. \tag{3.82}$$

Applying the Wirtinger derivative with respect to the conjugate of the array weight element w^* obeying the definition in Equation (2.17) and by using the product rule expressed as $\frac{\partial}{\partial t} c(t)a(t) = \frac{\partial c(t)}{\partial t} a(t) + \frac{\partial a(t)}{\partial t} c(t)$ in the context of $\frac{\bar{y}_{R,q}^{(+)}}{\sqrt{\mathbf{w}^H \mathbf{w}}} = \Re[\mathbf{w}^H \bar{\mathbf{x}}_q^{(+)}](\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}}$, results in:

$$\frac{\partial}{\partial \mathbf{w}} \left(\frac{\bar{y}_{R,q}^{(+)}}{\sqrt{\mathbf{w}^H \mathbf{w}}} \right) = \frac{\partial}{\partial \mathbf{w}^*} \left(\Re[\mathbf{w}^H \bar{\mathbf{x}}_q^{(+)}] \right) (\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} + \frac{\partial}{\partial \mathbf{w}^*} \left((\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} \right) \left(\bar{y}_{R,q}^{(+)} \right) \tag{3.83}$$

$$= \frac{\partial}{\partial \mathbf{w}^*} \left(\Re[(\mathbf{w}_R - j\mathbf{w}_I)^T (\bar{\mathbf{x}}_{R,q}^{(+)} + j\bar{\mathbf{x}}_{I,q}^{(+)})] \right) (\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} + \frac{\partial}{\partial \mathbf{w}^*} \left((\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} \right) \left(\bar{y}_{R,q}^{(+)} \right)$$

$$= \frac{\partial}{\partial \mathbf{w}^*} \left(\mathbf{w}_R^T \bar{\mathbf{x}}_{R,q}^{(+)} + \mathbf{w}_I^T \bar{\mathbf{x}}_{I,q}^{(+)} \right) (\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} + \frac{\partial}{\partial \mathbf{w}^*} \left((\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} \right) \left(\bar{y}_{R,q}^{(+)} \right) \tag{3.84}$$

$$= \frac{1}{2} \left(\bar{\mathbf{x}}_{R,q}^{(+)} + j\bar{\mathbf{x}}_{I,q}^{(+)} \right) (\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} + \left(-\frac{1}{2} \right) (\mathbf{w}^H \mathbf{w})^{-\frac{3}{2}} \left(\left(\frac{\partial}{\partial \mathbf{w}^*} \mathbf{w}^H \right) \mathbf{w} + \left(\frac{\partial}{\partial \mathbf{w}^*} \mathbf{w}^T \right) \mathbf{w}^* \right) \left(\bar{y}_{R,q}^{(+)} \right) \tag{3.85}$$

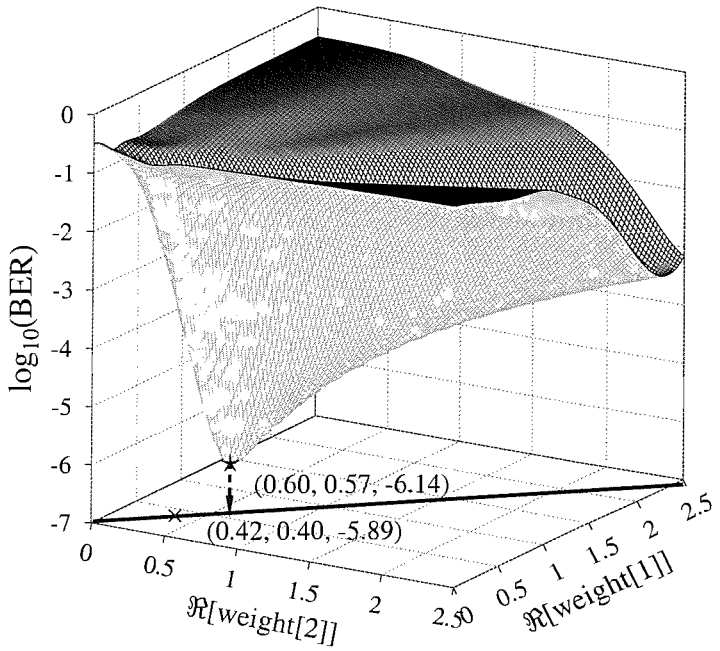
$$= \frac{1}{2} \bar{\mathbf{x}}_q^{(+)} (\mathbf{w}^H \mathbf{w})^{-\frac{1}{2}} - \frac{1}{2} (\mathbf{w}^H \mathbf{w})^{-\frac{3}{2}} \mathbf{w} \left(\bar{y}_{R,q}^{(+)} \right) \tag{3.86}$$

$$= \frac{1}{2} \left(\frac{\bar{\mathbf{x}}_q^{(+)}}{\sqrt{\mathbf{w}^H \mathbf{w}}} - \frac{\bar{y}_{R,q}^{(+)} \mathbf{w}}{\left(\sqrt{\mathbf{w}^H \mathbf{w}}\right)^3} \right). \tag{3.87}$$

Therefore the gradient of $P_E(\mathbf{w})$ in Equation (3.68) may be expressed as:

$$\nabla_{\mathbf{w}} P_E(\mathbf{w}) = -\frac{1}{N_{sb}\sqrt{2\pi}\sigma_n} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^H\mathbf{w}}\right) \mathrm{sgn}(b_{q,d}) \frac{1}{2} \left(\frac{\bar{\mathbf{x}}_q^{(+)}}{\sqrt{\mathbf{w}^H\mathbf{w}}} - \frac{\bar{y}_{R,q}^{(+)}\mathbf{w}}{\left(\sqrt{\mathbf{w}^H\mathbf{w}}\right)^3}\right) \quad (3.88)$$

$$= \frac{1}{2N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^H\mathbf{w}}\right) \mathrm{sgn}(b_{q,d}) \left(\frac{\bar{y}_{R,q}^{(+)}\mathbf{w}}{\mathbf{w}^H\mathbf{w}} - \bar{\mathbf{x}}_q^{(+)}\right). \quad (3.89)$$



α= 1.0

Figure 3.3: Error surface of the normalised MBER solution for $\alpha = 1$ and SNR = 10dB as well as INR_i = 10dB for $i = 2, 3, 4, 5$, when communicating over an AWGN channel. The associated interference scenario was plotted in Figure 3.18. The thick line on the horizontal plane is where the MBER solutions reside, corresponding to BER $= 10^{-6.14} \approx 7.24 \cdot 10^{-7}$. The MMSE solution is denoted by a cross, which should be located at $(0.42, 0.40)$ having a BER of $10^{-5.89} \approx 1.29 \cdot 10^{-6}$. The corresponding array weight's imaginary part was fixed at $(0.55, 1.01 \cdot 10^{-25})$ for the MBER solution and at $(0.39, 0.17)$ for the MMSE solution. The BER was evaluated based on Equation (3.68).

Unlike the paraboloid-shaped quadratic MSE cost function shown in Figure 2.12 of Section 2.4.1.1, the MBER cost function is not well-behaved. Using the MMSE approach, there is a unique global minimum, while for the MBER approach it is clear from Equation (3.70) that there is an infinite number of possible MBER solutions, since the equation is invariant to a positive scaling of $\mathbf{w}$. This is indicated by a straight line seen on the horizontal plane of Figure 3.3, an issue, which will be revisited in more depth in Section 3.5. Nonetheless, we may arrive at a single global minimum by normalising the array weight vector $\mathbf{w}$ to a unit length after every iteration, such that the normalised weight

denoted as $\ddot{\mathbf{w}}$ is given by:

$$\ddot{\mathbf{w}} = \frac{\mathbf{w}}{\|\mathbf{w}\|} = \frac{\mathbf{w}}{\sqrt{\mathbf{w}^H\mathbf{w}}} . \tag{3.90}$$

Hence the gradient of $P_E(\mathbf{w})$ expressed in Equation (3.89) can be simplified to:

$$\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}) = \frac{1}{2N_{sb}\sqrt{2\pi}\sigma_n} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2} \right) \text{sgn}(b_{q,d}) \left(\bar{y}_{R,q}^{(+)} \ddot{\mathbf{w}} - \bar{\mathbf{x}}_q^{(+)} \right) . \tag{3.91}$$

Apart from giving a single optimum solution, Equation (3.91) also offers a reduction in terms of the computational complexity imposed. An example of the BER surface using the normalised MBER beamforming weight vector $\ddot{\mathbf{w}}$, associated with $\alpha = 1.0$ in Equation (3.70) is shown in Figure 3.3. Note that the MMSE solution, which is also plotted in Figure 3.3 gives a BER performance of BER $= 10^{-5.89} \approx 1.29 \cdot 10^{-6}$, i.e. a higher BER. It is a coincidence that the MMSE solution lies on the MBER solution's line. We will explain this coincidence in more depth in Section 3.5.2. In the context of multiuser detection (MUD), a BER surface showing an infinite number of solutions has been graphically portrayed by Samingan in [199].

However, the convergence of the steepest descent algorithm is relatively slow [199]. In order to overcome this impediment, the Gauss-Newton algorithm [222] based on the nonlinear least square optimisation criterion can be applied. The algorithm updates the weight vector $\mathbf{w}$ in the direction given by [222]:

$$\mathbf{d}(n+1) = -[\nabla_{\mathbf{w}}^2 P_E(\mathbf{w}(n))]^{-1} \nabla_{\mathbf{w}} P_E(\mathbf{w}(n)), \tag{3.92}$$

where $\nabla_{\mathbf{w}}^2 P_E(\mathbf{w}(n))$ is the second order derivative of $P_E(\mathbf{w}(n))$ with respect to $\mathbf{w}$ at the nth iteration. Owing to the employment of the second derivative, the algorithm becomes computationally more complex than the previously mentioned steepest descent algorithm of Equation (3.73).

Method/ Parameter	steepest descent	Gauss-Newton	conjugate gradient	simplified conjugate gradient
$\mathbf{d}(n)$	$-\nabla_{\mathbf{w}} P_E(\mathbf{w}(n))$	$-[\nabla_{\mathbf{w}}^2 P_E(\mathbf{w}(n))]^{-1} \nabla_{\mathbf{w}} P_E(\mathbf{w}(n))$	$\phi_{n-1}\mathbf{d}(n-1) - \nabla_{\mathbf{w}} P_E(\mathbf{w}(n))$, where $\phi_{n-1} = \frac{\|\nabla_{\mathbf{w}} P_E(\mathbf{w}(n))\|^2}{\|\nabla_{\mathbf{w}} P_E(\mathbf{w}(n-1))\|^2}$	
μ	constant	constant	variable	constant
Remarks	simple but slow convergence	high computational complexity	requires optimal step size in each update	uses a fixed step size

Table 3.2: Comparison of the steepest descent, Gauss-Newton, conjugate gradient and simplified conjugate gradient methods in terms of the array weight vector direction $\mathbf{d}(n)$ and step size μ of array weight vector update of $\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{d}(n)$.

As a compromise between the speed of convergence and complexity imposed, we solve the optimisation problem of Equation (3.71) by using the conjugate gradient approach [223]. The conjugate

gradient method updates its current array weight vector direction by using the conjugate of the previous one. This approach is expected to be efficient, as it avoids using the same update direction twice. The algorithm is completed after L (the dimension of the array weight vector $\mathbf{w}$) iterative update steps using the optimal step size in each iteration. Comparison of the array weight vector direction $\mathbf{d}(n)$ update of $\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{d}(n)$ for the above-mentioned steepest-descent, Gauss-Newton and conjugate gradient methods is summarised in Table 3.2. For the sake of finding the MBER solution, we adopt a simplified version of the conjugate gradient approach [155, 198] that uses a constant step size for all iterations. The algorithm operates as follows [155, 198]:

Initialisation: Choose a step size of $\mu > 0$ and a termination scalar of $\beta > 0$.

Given $\ddot{\mathbf{w}}(1)$ and $\mathbf{d}(1) = -\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(1))$; set the iteration index to $n = 1$.

Loop: If $\|\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n))\| = \sqrt{(\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n)))^H \nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n))} < \beta$: goto *Stop*. Else,

$$\mathbf{w}(n+1) = \ddot{\mathbf{w}}(n) + \mu\mathbf{d}(n) \tag{3.93}$$

$$\ddot{\mathbf{w}}(n+1) = \frac{\mathbf{w}(n+1)}{\|\mathbf{w}(n+1)\|} \tag{3.94}$$

$$\phi_n = \frac{\|\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n+1))\|^2}{\|\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n))\|^2} \tag{3.95}$$

$$\mathbf{d}(n+1) = \phi_n \mathbf{d}(n) - \nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n+1)) \tag{3.96}$$

$n = n + 1$, goto *Loop*.

Stop: $\ddot{\mathbf{w}}(n)$ is the solution.

The optimum MBER solution is arrived at, when the gradient is zero, i.e. when we have $\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n)) = 0$. However, arriving at this condition would require numerous iterative loops. We used a small value of 10^{-15} for the scalar β as the termination constant in the above algorithm, so that a near-MBER solution can be found using a lower number of iterations. Note that we normalise the array weight vector in Equation (3.94), for the sake of arriving at a single optimum array weight solution. The vectorial search direction of the algorithm $\mathbf{d}(n)$ is updated by taking into account some contributions of the previous search direction $\mathbf{d}(n-1)$, where the value of this previous search direction is weighted by ϕ_n, i.e. by the square of the magnitude ratio of the current to the previous gradient of P_E. This results in introducing some randomness in the process converging towards the solution, which is typically beneficial in the context of nonlinear optimisation problems. As the BER surface is claimed to be highly irregular [198], as also evidenced by Figure 3.3, it is advisable to periodically reset the search direction to the negative gradient, according to $\mathbf{d}(n+1) = -\nabla_{\mathbf{w}} P_E(\ddot{\mathbf{w}}(n+1))$, especially when we have $\phi_n \geq 1$, which indicates that the magnitude of the current BER gradient is higher than or equal to the previous one. This is similar to the philosophy of the steepest descent algorithm, where we have $\phi_n = 0$ in Equation (3.96). Due to the irregular shape of the BER surface $P_E(\mathbf{w})$ shown in Figure 3.3, we should be aware that the convergence of this algorithm is not always guaranteed. Furthermore, it

is not obvious, how to adjust the initial array weight values of the algorithm. This may result in slow convergence, if the initial array weight vector is in the region of small gradient values.

In this section we have formulated the basic MBER beamforming approach. However, according to our previous approach the knowledge of the parameters involved, including the system matrix defined in Equation (3.8) is assumed. Therefore, the expressions derived in this section can only be used as a benchmarker or theoretical upper bound of the achievable MBER beamforming performance, which can readily be compared to the MMSE beamforming solution of Equation (3.25). Let us now consider a variety of algorithms, which allow us to approximate the MBER theoretical solution, while dispensing with the knowledge of unavailable system parameters.

3.4 Adaptive Minimum Bit Error Ratio Beamforming

The direct application of the algorithms mentioned in the previous section is unrealistic, since we assumed that all parameters required for the computation of the gradient of $P_E(\mathbf{w})$ expressed in Equation (3.68) are perfectly known. Owing to the time-variant nature of the mobile radio environment, an adaptive implementation should be considered and this would require the approximation of some of the parameters involved in Equation (3.68), including the PDF $p(y_R)$ of the decision variable y_R. In supporting the adaptive implementation of the MBER beamforming algorithm, we will adopt the temporal reference technique described in Chapter 2. We are interested in constructing the array weight adaptation procedure using both the block-data based SMI algorithm of Section 2.4.1.3 and the sample-by-sample adaptation aided LMS algorithm of Section 2.4.1.1.

Having $N_b = 2^M$ possible combinations of the M simultaneous users' transmitted bits $\mathbf{b}_q$, where we have $1 \leq q \leq N_b = 2^M$ and assuming that all $N_b = 2^M$ vectors are equiprobable, the true PDF $p(y_s)$ of $y_s(n)$ defined in Equation (3.33) can be shown to be given by [169,170]:

$$p(y_s) = \frac{1}{N_b\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} \sum_{q=1}^{N_b} \exp\left(-\frac{(y_s - \mathrm{sgn}(b_{q,d})\bar{y}_{R,q})^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right) \tag{3.97}$$

and as in Equation (3.68) of Section 3.3, the BER can alternatively be expressed as:

$$P_E(\mathbf{w}) = \frac{1}{N_b}\sum_{q=1}^{N_b} Q\left(c_q(\mathbf{w})\right), \tag{3.98}$$

where

$$c_q(\mathbf{w}) = \frac{\mathrm{sgn}(b_{q,d})\bar{y}_{R,q}}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} \tag{3.99}$$

and $\bar{y}_{R,q} \in \mathcal{Y}_R$. In reality, the true PDF $p(y_s)$ of $y_s(n)$ is unknown and its approximation is commonly based on the so-called kernel density or Parzen window based estimate [224]. We will provide a rudimentary introduction to this kernel density estimation procedure in the context of our MBER

beamforming scheme in Section 3.4.1. In Sections 3.4.2, 3.4.3.1 and 3.4.3.2 we present three algorithms, namely the Block Adaptive Conjugate Gradient (BACG) algorithm, the Least Bit Error Rate (LBER) algorithm and the Approximate LBER (ALBER) algorithm, respectively, that will adopt this density estimation approach. Simulation results characterising the BER performance of both the theoretical and adaptive MBER beamforming schemes will be provided in Sections 3.5 and 3.6, respectively.

3.4.1 Kernel Density Estimation

The kernel density estimator [224,225] was developed by Parzen [224] in 1962, which is defined by [225]:

$$\hat{p}(x) = \frac{1}{Ih} \sum_{i=1}^{I} \mathcal{K}\left(\frac{x - X_i}{h}\right), \tag{3.100}$$

where h is the window width that is sometimes referred as the smoothening parameter or bandwidth. The kernel density estimator of Equation (3.100) may be viewed as the superposition of I number kernel functions $\mathcal{K}\left(\frac{x-X_i}{h}\right)$, each positioned at the observed data X_i, where the shape of the kernels is determined by the specific choice of the kernel function $\mathcal{K}(x)$, while their widths is determined by the window width h. In our case, we consider the employment of the Gaussian kernel function, which is expressed as:

$$\mathcal{K}_{\text{Gaussian}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{3.101}$$

The Gaussian kernel function $\mathcal{K}_{\text{Gaussian}}(x)$ given in Equation (3.101) satisfies both properties to be fulfilled by the PDF $p(x)$ of a random variable x, such that we have [221]:

1. $p(x) \geq 0$, implying that $p(x)$ is a non-negative function of x; and that
2. $\int_{-\infty}^{\infty} p(x)\, dx = 1$, i.e. the area under the PDF curve $p(x)$ is unity.

Therefore the Gaussian kernel function is itself a PDF and this property is required in order for the estimated PDF produced by the kernel density estimator defined in Equation (3.100) to be a valid PDF. An example of the PDF estimates of variable x generated with the aid of $I = 10$ observed data values is shown in Figure 3.4 (a). The respective kernel function positioned at the observed data x in the set $X_i = \{-3.5, -2.0, -1.5, -1.3, 0, 0.3, 0.7, 1.0, 3.0, 6.0\}$ for $i = 1, ..., 10$, is represented by the dashed lines, while the resultant unnormalised PDF estimate is shown by the solid line. A window width of $h = 1.6$ was assumed for Figure 3.4 (a).

Let us now vary the window width h to study its effect on the kernel density. The resultant PDF estimate generated in conjunction with a window width smaller than $h = 1.6$, i.e. for example for $h = 0.5$ and a window width larger than 1.6, such as for example $h = 3.0$ can be observed from Figure 3.4 (b) and Figure 3.4 (c), respectively. All PDF estimate curves in Figure 3.4 are constructed using the same data samples of $X_i = \{-3.5, -2.0, -1.5, -1.3, 0, 0.3, 0.7, 1.0, 3.0, 6.0\}$, for $i = 1, ..., 10$,

which is marked by the dots on the x-axis. It can be observed from the plots that the window width h determines the width of the kernel function $\mathcal{K}(x)$ positioned at each data sample X_i. A small value of h results in a thin kernel, which smoothens the data very little and this may produce estimates that are irregular or under-smoothed, hence exhibiting spurious peaks in the estimates. By contrast, a large window width h widens the kernel, which may excessively smoothen the data, thus eliminating some of the details in the kernel estimates. This smoothening role played by the window width h may be the particular reason why it is often referred to as the smoothening parameter.

Having seen the effect of varying the window width or the smoothening parameter h, the issue now is how to determine the optimum window width h_{opt} in order to strike a good balance, i.e. to avoid under- and over-smoothening the kernel. Several techniques were proposed in [199] for obtaining the optimal window width for a particular problem. One of them is Silverman's rule of thumb [225], which states that the optimal window width for the Gaussian kernel can be calculated from the following expression [225]:

$$h_{opt} = \left(\frac{4\hat{\sigma}^5}{3I}\right)^{1/5} \approx 1.06\hat{\sigma}I^{-1/5} , \tag{3.102}$$

where $\hat{\sigma}$ is the standard deviation of the I number of data samples in the set $X_i, i = 1, ..., I$. The standard deviation $\hat{\sigma}$ may be calculated as [225]:

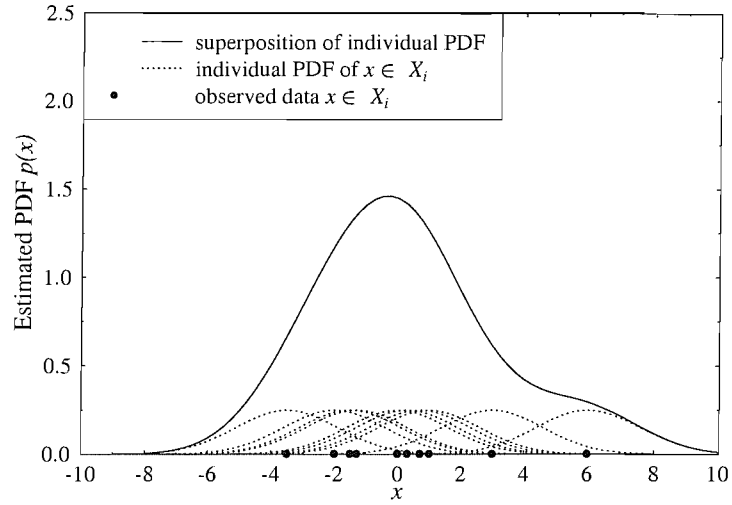
$$\hat{\sigma} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(X_i - \hat{\mu})^2} , \tag{3.103}$$

where $\hat{\mu}$ represents the mean of the observed data X_i. For the $I = 10$ data samples we used in our example of Figure 3.4, whose average was $\hat{\sigma} = 0.27$, the optimal width h_{opt} calculated according to Equation (3.102) can be shown to be $h_{opt} \approx 1.6$.

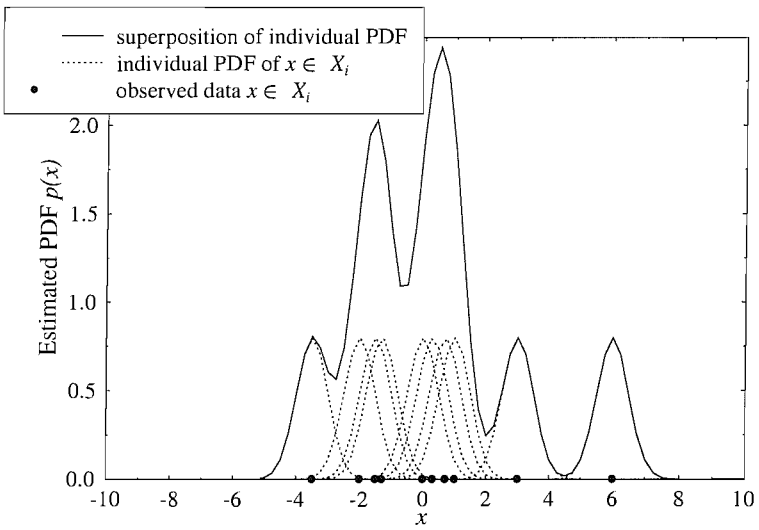
In our specific case of MBER beamforming, the PDF $p(y_s)$ of the signed decision function y_s defined in Equation (3.33) has been given in Equations (3.58) and (3.97). For convenience, again we repeat the PDF expression $p(y_s)$ here:

$$p(y_s) = \frac{1}{N_b\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{q=1}^{N_b}\exp\left(-\frac{(y_s - \mathrm{sgn}(b_{q,d})\bar{y}_{R,q})^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right) , \tag{3.104}$$

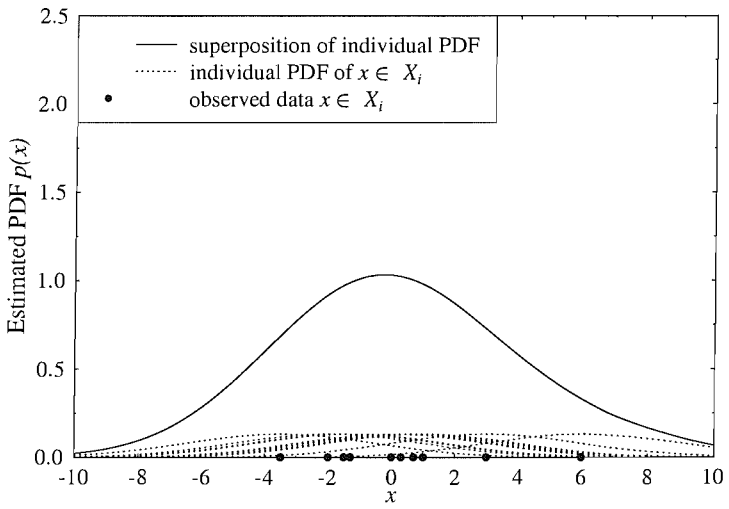
where $N_b = 2^M$ is the number of possible combinations of the M simultaneous user's transmitted bit $\mathbf{b}_q$, where we have $1 \leq q \leq N_b = 2^M$, while $b_{q,d} = \pm 1$ is the desired user's transmitted bit within one of the $N_b = 2^M$ number of possible M-bit combinations. In order to arrive at Equation (3.104) we assumed that all of the $N_b = 2^M$ number of vectors are equiprobable, as discussed in the context of Equation (3.57). The parameter $\bar{y}_{R,q}$ is the real part of the noiseless desired output $\bar{y}_q$, as defined in Equations (3.40) and (3.39), respectively. The kernel density estimate of the corresponding formula in Equation (3.104) may be obtained by using the Gaussian kernel function defined in Equation (3.101) and substituting the expression $\mathrm{sgn}(b_{q,d})\bar{y}_{R,q}$ of Equation (3.104) into the parameter X_i of the kernel



(a) window width $h = 1.6$



(b) window width $h = 0.5$



(c) window width $h = 3.0$

Figure 3.4: PDF estimate of variable x with the aid of 10 observed data at $X_i = \{-3.5, -2.0, -1.5, -1.3, 0, 0.3, 0.7, 1.0, 3.0, 6.0\}, i = 1, ..., 10$. Window width of (a) $h = 1.6$, (b) $h = 0.5$, (c) $h = 3.0$ and the Gaussian kernel function of Equation (3.101) were used for the estimation.

density estimator of Equation (3.100), such that we arrive at:

$$\hat{p}(y_s) = \frac{1}{Ih}\sum_{i=1}^{I}\mathcal{K}_{\text{Gaussian}}\left(\frac{y_s - X_i}{h}\right) \tag{3.105}$$

$$\hat{p}(y_s) = \frac{1}{Ih\sqrt{2\pi}}\sum_{i=1}^{I}\exp\left[\frac{-\left(\frac{y_s - X_i}{h}\right)^2}{2}\right] \tag{3.106}$$

$$= \frac{1}{Ih\sqrt{2\pi}}\sum_{i=1}^{I}\exp\left(-\frac{(y_s - X_i)^2}{2h^2}\right) \tag{3.107}$$

$$= \frac{1}{N_b h\sqrt{2\pi}}\sum_{q=1}^{N_b}\exp\left(-\frac{(y_s - \text{sgn}(b_{q,d})\bar{y}_{R,q})^2}{2h^2}\right), \tag{3.108}$$

where we assumed that $I = N_b$ number of samples are used in the PDF estimation. The signed decision function y_s was defined in Equation (3.33).

Having defined the estimated PDF for the signed decision function y_s for our BPSK based MBER beamforming scheme, we now present the three previously mentioned adaptive algorithms, namely the Block Adaptive Conjugate Gradient (BACG) algorithm, the Least Bit Error Rate (LBER) algorithm and the Approximate LBER (ALBER) algorithm, respectively, each of which will employ this estimated PDF operating on the basis of either block-data based or sample-by-sample based array-weight adaptation.

3.4.2 Block-Data Based Gradient Adaptive MBER Algorithm

Given $N_b = 2^M$ number of samples, a kernel density estimate of the true PDF of Equation (3.97) is readily derived with the aid of Equation (3.105). Similarly, for a block of T training samples constituted by the desired user's transmitted bits $b_d(t)$, $1 < t < T$, the estimated PDF of Equation (3.97) at time instant n can be expressed as [169, 170, 224, 225]:

$$\hat{p}(y_s(n)) = \frac{1}{T\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}}\sum_{t=1}^{T}\exp\left(-\frac{(y_s(n) - \text{sgn}(b_d(t))y_R(t))^2}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right), \tag{3.109}$$

where ρ_n is weighted by the magnitude of the array weights $\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$, i.e. $\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$ is the window width h represented in the kernel density estimator of Equation (3.100). The window width $\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$ is preferably referred here as the kernel width, in which ρ_n is a radius related to the standard deviation σ_n of the noise. The weighting of the kernel width $\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$ by the weight-vector related factor $\sqrt{\mathbf{w}^H\mathbf{w}}$ was due to the $\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}$ expression in the true PDF $p(y_s)$ of Equation (3.97). Note that a different notation, namely $y_R(t)$ is used for representing the real part of the noise-contaminated array output associated with the tth transmitted bit $b_d(t)$ of the desired user's within the set of T bits in a training block. Similar to the derivation seen in the context of Equation (3.59) of Section 3.3, the estimated conditional PDF of the signed decision function $y_s(n)$

at time instant n defined in Equation (3.33), given that we have $b_d(t) = +1$, for $1 < t < T$ of a block of T training samples can be expressed as:

$$\hat{p}(y_s(n)|b_d(t) = +1) = \hat{p}(y_R(n)) = \frac{1}{T\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}} \sum_{t=1}^{T} \exp\left(-\frac{(y_R(n) - \text{sgn}(b_d(t))y_R(t))^2}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right). \tag{3.110}$$

Taking into account that the transmitted bit of the desired user is $b_d(t) = +1$, according to Equation (3.19) the probability of error $\hat{P}_E(\mathbf{w})$ can be represented by the area under the estimated conditional PDF of Equation (3.110) within the interval of $(-\infty, 0)$. The specified area under the estimated PDF can be evaluated following the procedure outlined in the context of Equations (3.61) to (3.67), which can be expressed as [169]:

$$\hat{P}_E(\mathbf{w}(n)) = \int_{-\infty}^{0} \hat{p}(y_R(n))\, dy_R \tag{3.111}$$

$$= \frac{1}{T}\sum_{t=1}^{T} Q\left(\hat{c}_t(\mathbf{w}(n))\right), \tag{3.112}$$

where the standard Q-function $Q(z)$ is defined in Equation (3.63) and $\hat{c}_t(\mathbf{w}(n))$ is given by:

$$\hat{c}_t(\mathbf{w}(n)) = \frac{\text{sgn}(b_d(t))y_R(t)}{\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}}. \tag{3.113}$$

The expression given in Equation (3.112) is referred to as block-based estimate of the BER since a training block of T samples is used for the BER's estimation. According to Equation (3.71), we have to evaluate the gradient of this block-estimated BER of Equation (3.112) for the sake of obtaining a solution close to the MBER solution.

Using the identity of Equation (3.74), the gradient of the estimated BER $\hat{P}_E(\mathbf{w}(n))$ can be calculated following a similar procedure to that outlined in the context of Equations (3.75) to (3.89), resulting in:

$$\nabla_{\mathbf{w}}\hat{P}_E(\mathbf{w}(n)) = \frac{1}{T\sqrt{2\pi}}\sum_{t=1}^{T} \exp\left(-\frac{y_R^2(t)}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right)\frac{\partial}{\partial\mathbf{w}}\left(-\hat{c}_t(\mathbf{w}(n))\right) \tag{3.114}$$

$$= -\frac{1}{2T\sqrt{2\pi}\rho_n} \sum_{t=1}^{T} \exp\left(-\frac{y_R^2(t)}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right)\text{sgn}(b_d(t))\left(\frac{\mathbf{x}(t)}{\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}} - \frac{\mathbf{w}(n)^H\mathbf{x}(t)\mathbf{w}(n)}{\left(\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}\right)^3}\right) \tag{3.115}$$

$$= \frac{1}{2T\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}} \sum_{t=1}^{T} \exp\left(-\frac{y_R^2(t)}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right)\text{sgn}(b_d(t))\left(\frac{y_R(t)\mathbf{w}(n)}{\mathbf{w}(n)^H\mathbf{w}(n)} - \mathbf{x}(t)\right). \tag{3.116}$$

Since we are dealing with complex weight values, for a two-element array there are a total of four array weights, that have to be considered in producing the BER surface. However, for reasons of practical

data visualisation we were limited to portraying the 3-dimensional BER surface seen in Figure 3.3 by fixing the imaginary parts of both array weights w_1 and w_2. For the sake of showing the infinite number of optimum array weight values' location, we drew a straight line on the horizontal plane of Figure 3.3 as a result of varying the values of α. However, it is desirable to arrive at a single optimum array weight solution, and hence we normalised the weight vector to unit magnitude, as in Equation (3.90). Since the array weight normalisation ensures that we have $\mathbf{w}(n)^H\mathbf{w}(n) = 1$, the expression of Equation (3.116) may be simplified to:

$$\nabla_{\mathbf{w}}\hat{P}_E(\ddot{\mathbf{w}}(n)) = \frac{1}{2T\sqrt{2\pi}\rho_n}\sum_{t=1}^{T}\exp\left(-\frac{y_R^2(t)}{2\rho_n^2}\right)\mathrm{sgn}(b_d(t))\left(y_R(t)\ddot{\mathbf{w}}(n) - \mathbf{x}(t)\right). \tag{3.117}$$

Upon substituting $\nabla_{\mathbf{w}}P_E(\ddot{\mathbf{w}}(n))$ by $\nabla_{\mathbf{w}}\hat{P}_E(\ddot{\mathbf{w}}(n))$ in the simplified conjugate gradient algorithm of Equation (3.96), a block-data based adaptive algorithm is obtained, which we refer to as the Block Adaptive Conjugate Gradient (BACG) algorithm. Using this BACG algorithm the array weight vector is basically updated based on the block of T training samples in each subsequent iteration. As seen in Equation (3.93), apart from the step size μ, the kernel width radius ρ_n is another parameter for this algorithm, which is used in determining $\nabla_{\mathbf{w}}\hat{P}_E(\ddot{\mathbf{w}}(n))$ of Equation (3.117). The achievable performance of this technique will be characterised in Section 3.6.

Having introduced the BACG algorithm, which is a block-data based updating algorithm relying on T training samples, in the next section we will describe two different Adaptive MBER algorithms, which operate on a sample-by-sample basis, namely the Least BER (LBER) and the Approximate LBER (ALBER) algorithms.

3.4.3 Stochastic Gradient Based Adaptive MBER Algorithms

In the context of the stochastic approach the adaptive MBER algorithm updates the array weight vector $\mathbf{w}$ instantaneously, such that a different training sample is used in each array weight vector update iteration. Unlike the block adaptive approach of Section 3.4.2, which uses a block of T training samples, in the stochastic approach we use only a single training sample.

3.4.3.1 Least Bit Error Rate Algorithm

With regards to the estimated PDF given in Equation (3.109) associated with a block of T training samples, the estimated PDF used in the context of the sample-by-sample stochastic adaptive algorithm may be evaluated via the same procedure, with the exception of using an arbitrary single sample in each array weight update iteration. When utilising Parzen's kernel density estimator of Equation (3.100) for algorithms based on the stochastic adaptive approach the corresponding estimated PDF based on

the Gaussian kernel function of Equation (3.101) can be expressed as [197]:

$$\hat{p}(y_s(n)) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}} \exp\left(-\frac{(y_s(n) - \text{sgn}(b_d(n))y_R(n))^2}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right), \tag{3.118}$$

where $\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$ is the kernel width having a radius of ρ_n weighted by $\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$. Note that a different notation of $y_R(n)$ is used for representing the real part of the noise-contaminated array output associated with the transmitted desired user's bit $b_d(n)$ at time instant n. Similar to the approach derived for the BACG algorithm of Section 3.4.2, which is based on the derivation seen in Equation (3.59) of Section 3.3, the estimated conditional PDF of the signed decision function y_s defined in Equation (3.33) can be expressed as:

$$\hat{p}(y_s|b_d(n) = +1) = \hat{p}(y_R) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}} \exp\left(-\frac{(y_R - \text{sgn}(b_d(n))y_R(n))^2}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right), \tag{3.119}$$

given that the transmitted desired user's bit was $b_d(n) = +1$ at time instant n. The probability of error $P_E(\mathbf{w}(n))$, which may be associated with the area under the estimated PDF can be evaluated similar to the BACG algorithm of Section 3.4.2 following the procedure of Equations (3.61) to (3.67), such that it can be expressed as [169]:

$$\hat{P}_E(\mathbf{w}(n)) = \int_{-\infty}^{0} \hat{p}(y_R)\, dy_R \tag{3.120}$$

$$= Q\left(\hat{c}_t(\mathbf{w}(n))\right), \tag{3.121}$$

where the standard Q-function $Q(z)$ is defined in Equation (3.63) and $\hat{c}_t(\mathbf{w}(n))$ is given by:

$$\hat{c}_t(\mathbf{w}(n)) = \frac{\text{sgn}(b_d(n))y_R(n)}{\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}}. \tag{3.122}$$

The instantaneous gradient estimate $\nabla_{\mathbf{w}}\hat{P}_E(\mathbf{w}(n))$ based on a single training sample may be expressed with the aid of Equation (3.74), yielding:

$$\nabla_{\mathbf{w}}\hat{P}_E(\mathbf{w}(n)) = \frac{\text{sgn}(b_d(n))}{2\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}} \exp\left(-\frac{y_R^2(n)}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right)\left(\frac{y_R(n)\mathbf{w}(n)}{\mathbf{w}(n)^H\mathbf{w}(n)} - \mathbf{x}(n)\right), \tag{3.123}$$

which gives rise to a stochastic gradient adaptive algorithm, which we referred to as the Least Bit Error Rate (LBER) algorithm, whose array weights are updated based on the following expression [6]:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\nabla_{\mathbf{w}}\hat{P}_E(\mathbf{w}(n)). \tag{3.124}$$

This formula represents the well-known steepest descent gradient algorithm, where the array weight vector is updated in a direction opposite to the BER-gradient $\nabla_{\mathbf{w}}\hat{P}_E$, weighted by the step size μ.

Using Equation (3.123), this expression may be formulated more explicitly as follows:

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) - \\ &\mu\frac{\mathrm{sgn}(b_d(n))}{2\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}}\exp\left(-\frac{y_R^2(n)}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right)\left(\frac{y_R(n)\mathbf{w}(n)}{\mathbf{w}(n)^H\mathbf{w}(n)}-\mathbf{x}(n)\right) \quad (3.125)\\ &= \mathbf{w}(n) + \\ &\mu\frac{\mathrm{sgn}(b_d(n))}{2\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}}\exp\left(-\frac{y_R^2(n)}{2\rho_n^2\mathbf{w}(n)^H\mathbf{w}(n)}\right)\left(\mathbf{x}(n)-\frac{y_R(n)\mathbf{w}(n)}{\mathbf{w}(n)^H\mathbf{w}(n)}\right), \quad (3.126)\end{aligned}$$

where the step size μ and the kernel width ρ_n are the two algorithmic parameters that have to be set appropriately for the sake of ensuring convergence towards the true MBER solution. Again, the weight vector is normalised to a unit length after each update for the sake of simplifying the LBER algorithm's expression seen in Equation (3.126) to the form:

$$\mathbf{w}(n+1) = \ddot{\mathbf{w}}(n) + \mu\frac{\mathrm{sgn}(b_d(n))}{2\sqrt{2\pi}\rho_n}\exp\left(-\frac{y_R^2(n)}{2\rho_n^2}\right)(\mathbf{x}(n) - y_R(n)\ddot{\mathbf{w}}(n)). \quad (3.127)$$

Note in Equation (3.127) that the array weight vector is denoted as $\ddot{\mathbf{w}}(n)$, which implies that this LBER algorithm requires the array weight $\mathbf{w}(n)$ to be normalised after each updating instant n.

3.4.3.2 Approximate Least Bit Error Rate Algorithm

In the kernel density estimate of the sample PDF seen in Equation (3.119), a variable kernel width of $\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$ is used. This is because the true standard deviation of $y_R(n)$ is given by $\sigma_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}$, which is related to the magnitude of the beamformer's weight vector $\mathbf{w}(n)$. If this kernel-width expression is approximated by a constant width of ρ_n, the associated computational complexity can be considerably reduced [155], since the computation of $\mathbf{w}(n)^H\mathbf{w}(n)$ or the normalisation of the weight vector $\mathbf{w}(n)$ is no longer necessary.

Formally, this leads to using the kernel density estimate of:

$$\tilde{p}(y_R) = \frac{1}{\sqrt{2\pi}\rho_n}\exp\left(-\frac{(y_R - \mathrm{sgn}(b_d(n))y_R(n))^2}{2\rho_n^2}\right), \quad (3.128)$$

as an approximation of the true kernel density given by Equation (3.97) following the kernel density estimator of Equation (3.100). By the same procedure of Equations (3.119) and (3.120) for the LBER algorithm, we may arrive at:

$$\tilde{P}_E(\mathbf{w}(n)) = Q\left(\tilde{c}_n(\mathbf{w}(n))\right) \quad (3.129)$$

in conjunction with

$$\tilde{c}_n(\mathbf{w}(n)) = \frac{\mathrm{sgn}(b_d(n))y_R(n)}{\rho_n} \quad (3.130)$$

as the BER estimate. This approximation does not result in any dramatic performance degradation, provided that the constant width ρ_n is chosen appropriately [155] as it will be shown in Section 3.6. Therefore, the gradient of $\tilde{P}_E(\mathbf{w}(n))$ has a simpler form expressed as:

$$\nabla_{\mathbf{w}}\tilde{P}_E(\mathbf{w}(n)) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y_R^2(n)}{2\rho_n^2}\right)\frac{\partial}{\partial \mathbf{w}}\left(-\tilde{c}_n(\mathbf{w}(n))\right) \tag{3.131}$$

$$= -\frac{\text{sgn}(b_d(n))}{2\sqrt{2\pi}\rho_n}\exp\left(-\frac{y_R^2(n)}{2\rho_n^2}\right)\mathbf{x}(n). \tag{3.132}$$

Note that no weight normalisation is necessary for simplifying the gradient expression seen in Equation (3.132). Adopting this approach leads to the definition at another stochastic gradient algorithm referred to here as the Approximate LBER (ALBER) technique, which is formulated as:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\nabla_{\mathbf{w}}\tilde{P}_E(\mathbf{w}(n)) \tag{3.133}$$

$$= \mathbf{w}(n) + \mu\frac{\text{sgn}(b_d(n))}{2\sqrt{2\pi}\rho_n}\exp\left(-\frac{y_R^2(n)}{2\rho_n^2}\right)\mathbf{x}(n). \tag{3.134}$$

Apart from its similar philosophy, the ALBER algorithm has a similar computational complexity order $O(L)$ to that of the low-complexity LMS algorithm [6, 166]. Specifically, the LMS algorithm discussed in Section 2.4.1.1 requires $2L+1$ complex multiplications and $2L$ complex additions per array weight update iteration, while the ALBER algorithm requires $2L+7$ complex multiplications and $2L-1$ complex additions per array weight update iteration, where L is the number of elements in the antenna array.

Having outlined three different adaptive MBER algorithms, in Section 3.6 we will characterise their performance relative to the theoretical MBER performance. Prior to that, however, it is beneficial for us to first compare the theoretical performance of MBER beamforming to that of MMSE beamforming in Section 3.5.

3.5 Simulation Results for Exact MBER Beamforming

The performance of the beamformer based on the MBER weight optimisation criterion is studied with the aid of computer simulations, where the BER is calculated based on Equation (3.68) assuming the employment of BPSK modulation, where the Gaussian Q-function is approximated using the formula given by Ziemer and Tranter [221, 226]:

$$Q(x) \approx \frac{\exp^{-x^2/2}}{\sqrt{2\pi}}\left(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5\right), \qquad x \geq 0 \tag{3.135}$$

where we have $t = \frac{1}{(1+\rho x)}$, $\rho = 0.2316419$, and

$$b_1 = 0.31981530 \quad (3.136)$$

$$b_2 = -0.356563782 \quad (3.137)$$

$$b_3 = 1.781477937 \quad (3.138)$$

$$b_4 = -1.821255978 \quad (3.139)$$

$$b_5 = 1.330274429. \quad (3.140)$$

x	$Q(x)$	x	$Q(x)$	x	$Q(x)$	x	$Q(x)$
0.0	0.50000						
0.1	0.46017	1.1	0.13567	2.1	0.01786	3.1	0.00097
0.2	0.42074	1.2	0.11507	2.2	0.01390	3.2	0.00069
0.3	0.38209	1.3	0.09680	2.3	0.01072	3.3	0.00048
0.4	0.34458	1.4	0.08076	2.4	0.00820	3.4	0.00034
0.5	0.30854	1.5	0.06681	2.5	0.00621	3.5	0.00023
0.6	0.27425	1.6	0.05480	2.6	0.00466	3.6	0.00016
0.7	0.24196	1.7	0.04457	2.7	0.00347	3.7	0.00011
0.8	0.21186	1.8	0.03593	2.8	0.00256	3.8	0.00007
0.9	0.18406	1.9	0.02872	2.9	0.00187	3.9	0.00005
1.0	0.15866	2.0	0.02275	3.0	0.00135	4.0	0.00003

Table 3.3: Q-function values $Q(x)$ for the independent variable values x ranging from $x = 0$ to $x = 4$, calculated according to Equation (3.135) [221].

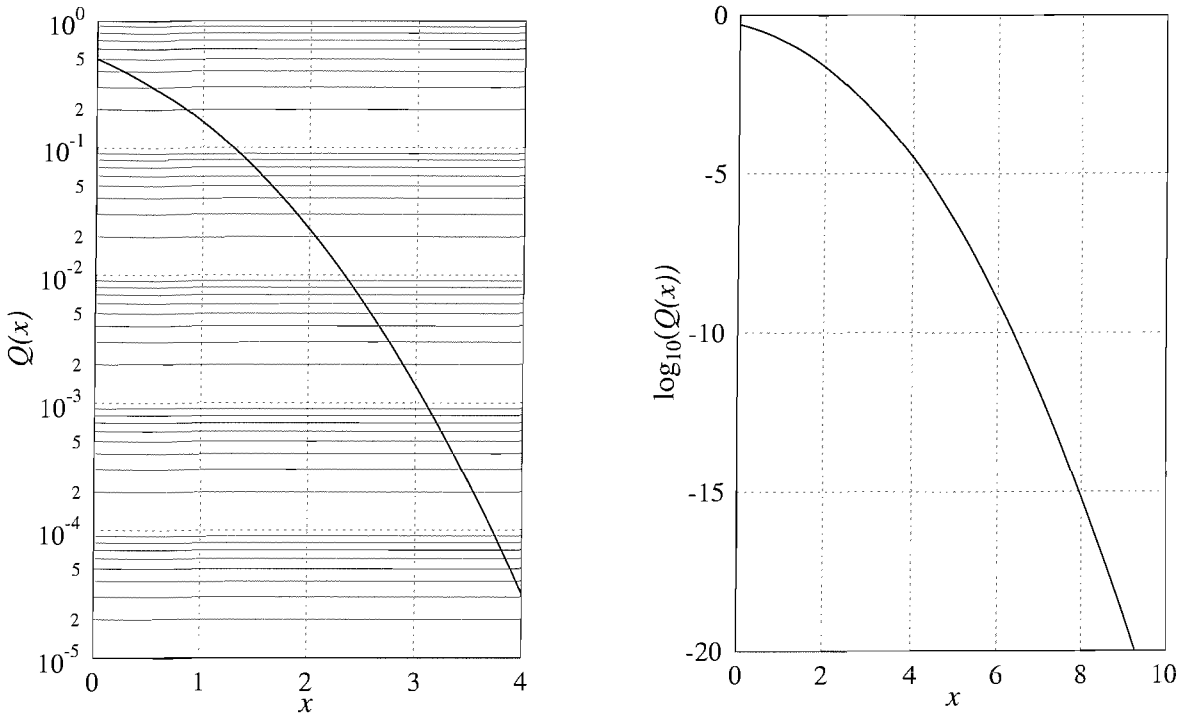
A plot of the Q-function seen in Equation (3.135) and evaluated for the abscissa values of Table 3.3 [221] is shown in Figure 3.5.

The simulations carried out considered several scenarios when communicating over a non-dispersive Additive White Gaussian Noise (AWGN) channel that differ in terms of both the number of users supported and the angle of arrivals (AOAs) encountered.

3.5.1 Supporting Two Users Employing a Two-Element Uniform Linear Array

We commence our simulation experiments with the example of a two-user scenario, where the signal of the desired user arrives from 0^o, while that of the interfering user from 30^o.

Figure 3.6 shows the attainable BER performance as a function of the users' signal power, in conjunction with two-, four- and eight-element uniform linear arrays. Let us first concentrate our attention on the two-element array, whose MMSE and MBER design based BER curves are represented by the continuous line marked by circles and the dotted line associated with stars, respectively. In this scenario it is observed that the MBER solution performs consistently better, exhibiting an SNR advantage of about 2dB and requiring about 1dB higher SNR than the upper-bound performance



(a) $Q(x)$ (b) $\log_{10}(Q(x))$

Figure 3.5: (a) The Gaussian Q-function $Q(x)$ based on the approximation seen in Equation (3.135) as proposed by Ziemer and Tranter [226] for the abscissa values shown in Table 3.3. Subfigure (b) shows $\log_{10}Q(x)$ for $0 \leq x < 10$.

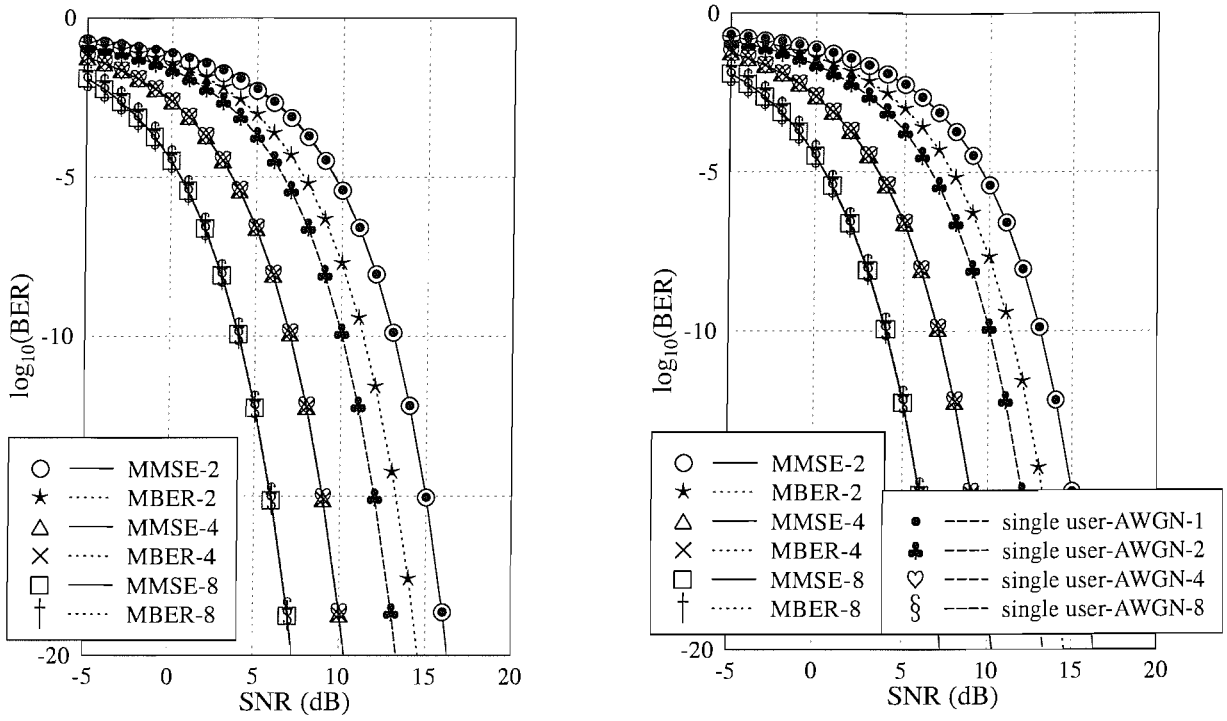
of the single-user, two-element array scenario. By contrast, the MMSE solution performed similarly to the single-user, single-element case. For the sake of interpreting these performance trends let us consider the achievable BER at SNR = 10dB. Specifically, the calculation of the probability of error $P_E(\mathbf{w})$ for the MBER beamforming is using the formulation given by Equation (3.68), while for the single-user, single-element antenna scenario $P_{E,\text{BPSK}}$ can be expressed as [168, 221]:

$$P_{E,\text{BPSK}} = Q\left(\sqrt{\frac{a^2}{\sigma_n^2}}\right), \tag{3.141}$$

where a^2 is the received signal power and $\sigma_n^2 = N_0/2$, where N_0 is the white Gaussian noise power. For a two-user scenario, Equation (3.68) may be simplified to:

$$P_E(\mathbf{w}) = \frac{1}{2}\sum_{q=1}^{2} Q\left(\frac{\bar{y}_{R,q}}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\right), \tag{3.142}$$

where $\bar{y}_{R,q}$ is the real part of the noiseless desired output $\bar{y}_q$, as defined in Equations (3.40) and (3.39), respectively. The values of the parameter $\bar{y}_{R,q}$ were plotted in the top right-most subfigure of Figure 3.1 for the MMSE beamformer and also in Figure 3.13 a) for SNR = INR = 10dB. Note that the beamformer's output values in Figure 3.13 a) were $\bar{y}_{R,q} = 0.87$ and $\bar{y}_{R,q} = 0.95$ in conjunction with $\sqrt{\mathbf{w}_{\text{MMSE}}^H\mathbf{w}_{\text{MMSE}}} = 0.872357$, which results in abscissa values of $x = 0.997$ and $x = 1.089$, respectively, for the Q-function of Equation (3.142). Assuming a unity amplitude of



(a) SNR = INR

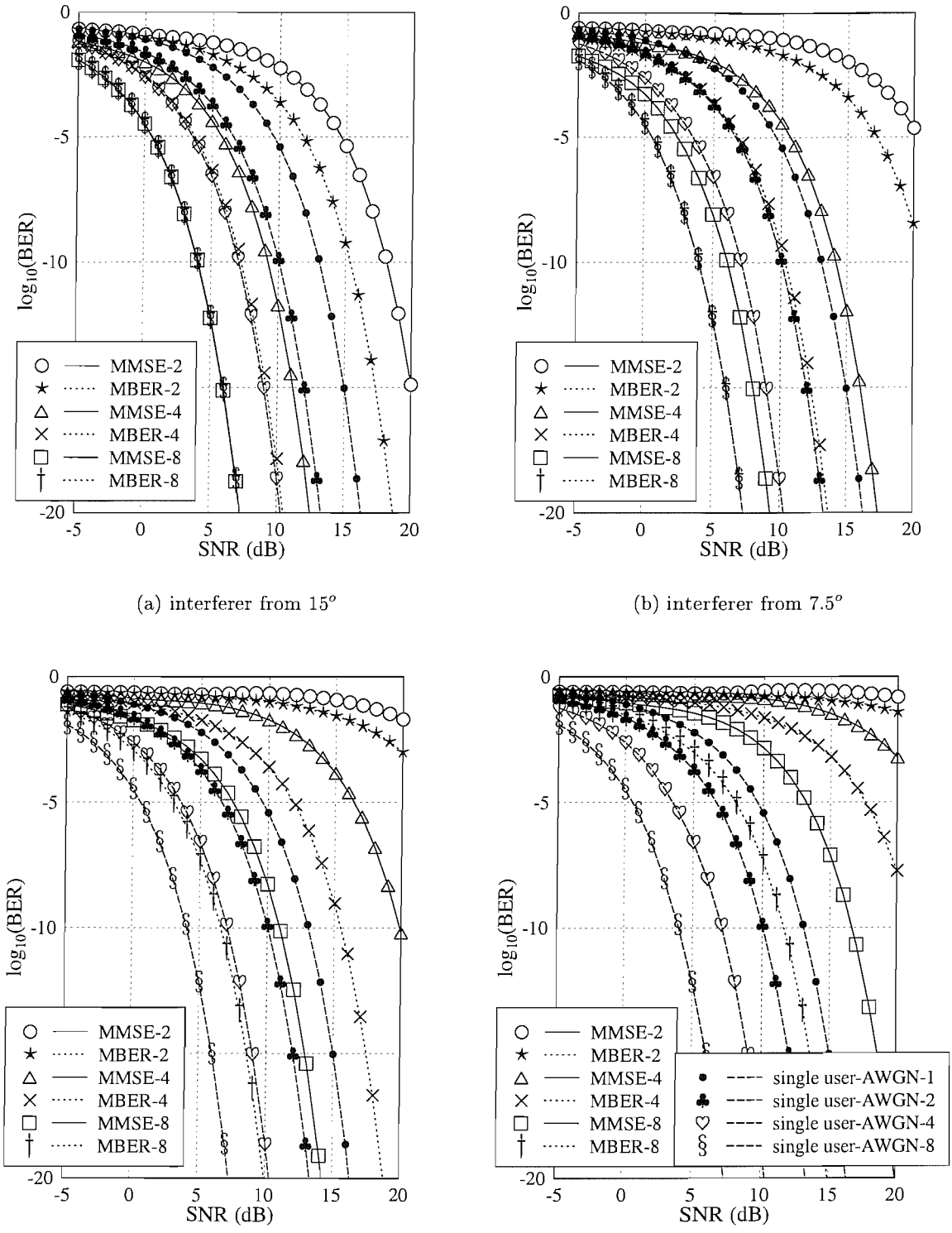
(b) INR = SNR +6dB

Figure 3.6: Comparison of the BER performance of the MMSE and the MBER beamformers, based on Equation (3.68), for two different conditions; (a) SNR = INR, and (b) INR = SNR +6dB, in conjunction with two-, four- and eight-element uniform linear array system in an AWGN channel[1]. The AOA of the desired and interfering user was 0^o and 30^o, respectively.

$a = 1$ in Equation (3.141), the BER calculation using Equations (3.141) and (3.142) results in $P_E(\mathbf{w}_{\text{MMSE}}) = \frac{1}{2}(Q(4.46) + Q(4.87)) = 2.33 \cdot 10^{-6}$ at SNR =10dB, i.e. for $\sigma_n = \sqrt{0.05}$, while the corresponding simulated value was $P_{E,\text{BPSK}} = Q(4.47) = 3.88 \cdot 10^{-6}$. As the SNR is increased, the associated variance of σ_n^2 becomes lower. However, the calculation of the MMSE weights according to Equation (3.26) also involves the variance σ_n^2, thus balancing the numerator and denominator of the Q-function's abscissa value in Equation (3.142).

It is also observed from Figures 3.6 (a) and 3.6 (b) that the performance of both the MMSE and MBER solutions recorded for the various number of array elements is fairly independent of the SIR in the investigated scenario. The BER curve of the three-element array is not plotted in Figure 3.6. For this specific two-user scenario, associated with the AOA of 0^o and 30^o, adding further antenna elements will produce no significant performance difference between the MMSE and MBER solutions. Furthermore, for a minimum of four array elements the BER performance of both the MMSE and MBER solutions achieved the optimum single-user performance associated with the same number of antenna elements. Let us now move the interfering user closer to the desired user by reducing their angle of arrival difference. As shown in Figures 3.7 (a), 3.7 (b), 3.7 (c) and 3.7 (d), we halved the

[1]The single-user, multi-element upper-bound curves were plotted based on the single-user, single-element curve, with a 3dB shift to the left for each doubling on the number of antenna elements.



(a) interferer from 15^o

(b) interferer from 7.5^o

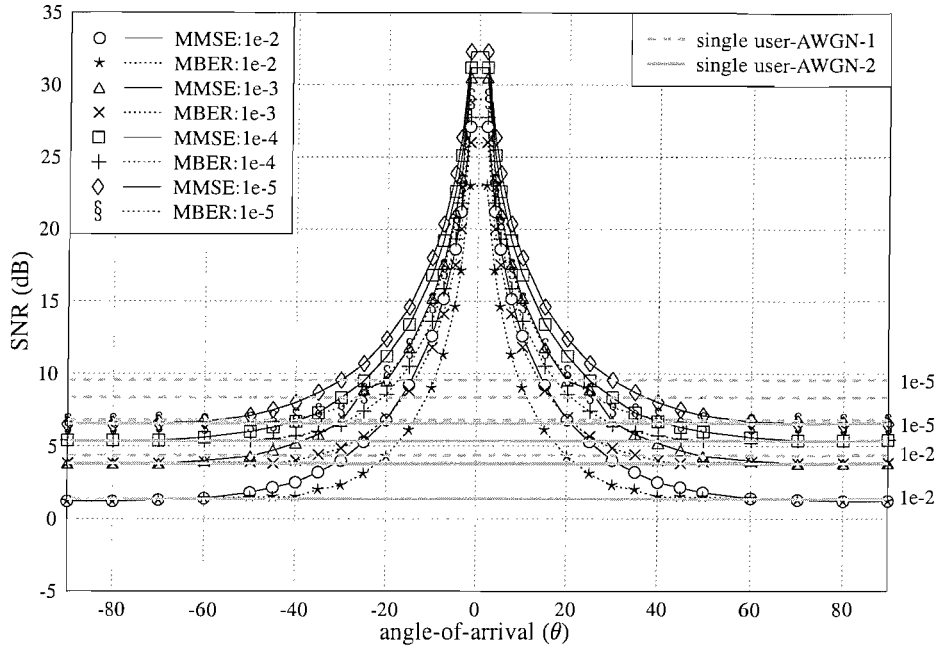
(c) interferer from 3.75^o

(d) interferer from 1.875^o

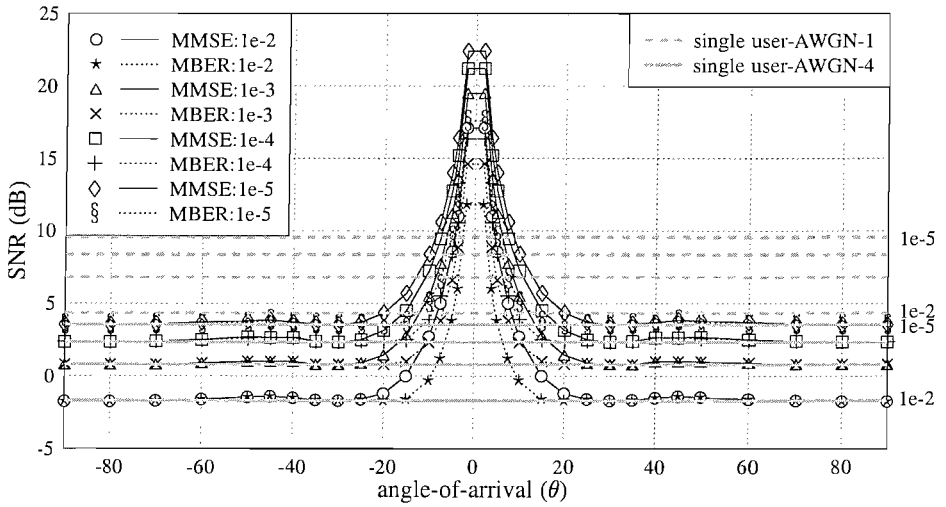
Figure 3.7: Comparison of the BER performance of the MMSE and the MBER beamformers, based on Equation (3.68), for two equal-power users, i.e. for SNR = INR, in conjunction with two-, four- and eight-element uniform linear arrays communicating over an AWGN channel. The AOA of the desired and interfering user was 0^o and (a) 15^o, (b) 7.5^o, (c) 3.75^o, (d) 1.875^o, respectively.

angular distance between the two users repeteadly, i.e. the interfering user's signal is now assumed to arrive from $15^o, 7.5^o, 3.75^o$ and 1.875^o, respectively, with respect to the desired user's signal arriving from 0^o. Most plots in Figure 3.7 show a degraded BER performance when compared to the plots seen in Figure 3.6, with a few exceptions noted for the MBER curves. Explicitly, the respective MBER curves of the eight-element array seen in Figures 3.7 (a) and 3.7 (b) and in the four-element scenario of Figure 3.7 (a) are not degraded, despite the reduced AOA difference. By contrast, the MMSE beamformer was capable of maintaining a similar BER performance to that of Figure 3.6 for the eight-element array of Figure 3.7 (a), when the AOA difference was reduced from 30^o to 15^o. However, the BER performance of the two-user scenario was significantly degraded, where even the originally superior MBER curve of Figure 3.7 (a) is facing an SNR degradation of approximately 2.5dB in comparison to the single-user, single-element BER curve.

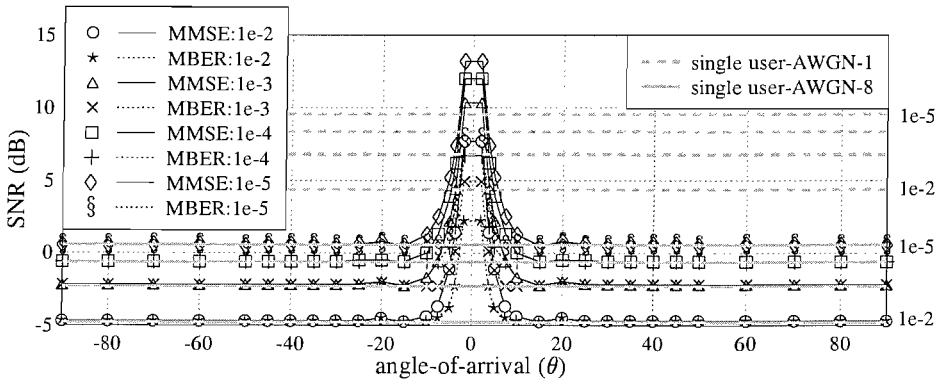
For a more clear visualisation, Figures 3.8 and 3.9 show the performance of both the MMSE and MBER beamformers as the interfering user's AOA θ is varied from -90^o to 90^o, i.e. for $-90^o \leq \theta \neq 0^o \leq 90^o$. Note that the interfering user's signal cannot arrive from 0^o, since this is assumed to be the AOA for the desired user, thus having an interfering user in the same direction would render their separation infeasible. In Figure 3.8 it is shown that the signal power required for the users in order to achieve a specific BER performance increases, as the interfering user moves closer in angular terms to the desired user located at 0^o. When aiming for BER $= 10^{-2}$ in the context of the two-element array, it is clear from Figure 3.8 (a) that the MBER beamformer would require a lower desired user signal power than that required by the MMSE beamformer. The difference in terms of the required signal power becomes, however, negligible, when the interfering user's AOA is higher than 60^o, i.e. it is $|\theta| \geq 60^o$ away from the desired user. This is also true, when the number of array elements is increased. Note also the reduced angular difference requirement between the two users for the four- and eight-element case, respectively, which were found to be $|\theta| \geq 25^o$ and $|\theta| \geq 10^o$ in the corresponding Figures 3.8 (b) and 3.8 (c). Upon comparing Figures 3.8 (a), 3.8 (b) and 3.8 (c), it may be observed that as the number of antenna array elements is doubled, the signal power required by the two users is reduced by 3dB. This observation is only valid, however, in the angular interval of $|\theta| \geq 60^o$. It is noted from Figure 3.8 that the best performance the beamformer was capable of achieving as the interfering user was angularly further separated from the SOI, was the optimum single-user performance of the corresponding number of antenna elements. Figure 3.8 also shows that even for an extremely small angular separation of 1.875^o between the desired and interfering user the MBER beamformer remains capable of attaining a BER of 10^{-5}, provided that the desired user's signal power is in excess of 29dB, 17.5dB and 7.5dB for the two-, four- and eight-element antenna arrays, respectively. In other words, doubling the number of antenna elements to four or eight allows us to decrease the reference user's signal power by 11.5dB and 10dB, respectively. By contrast, the MMSE beamformer would require a desired user signal power of 32.5dB, 22.5dB and 13dB for the two-, four- and eight-element antenna array, respectively, for the sake of maintaining a BER of 10^{-5}, resulting in a signal power reduction of



(a) two-element uniform linear array system

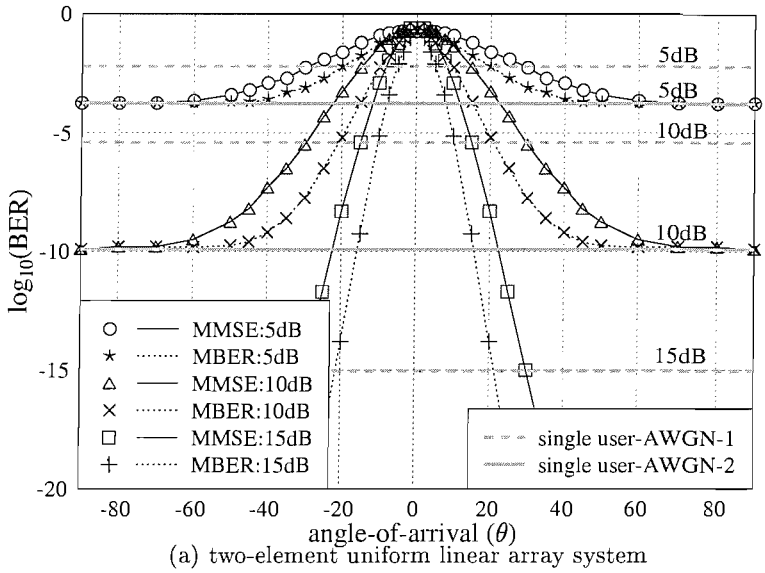


(b) four-element uniform linear array system

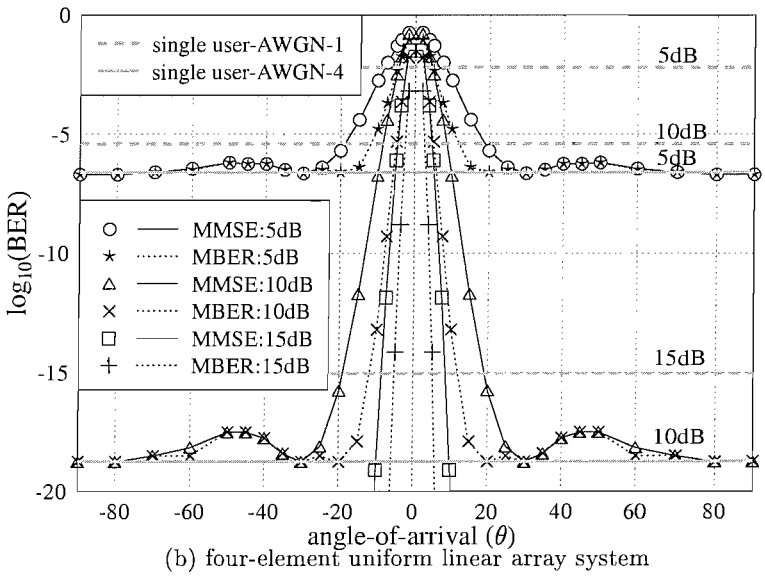


(c) eight-element uniform linear array system

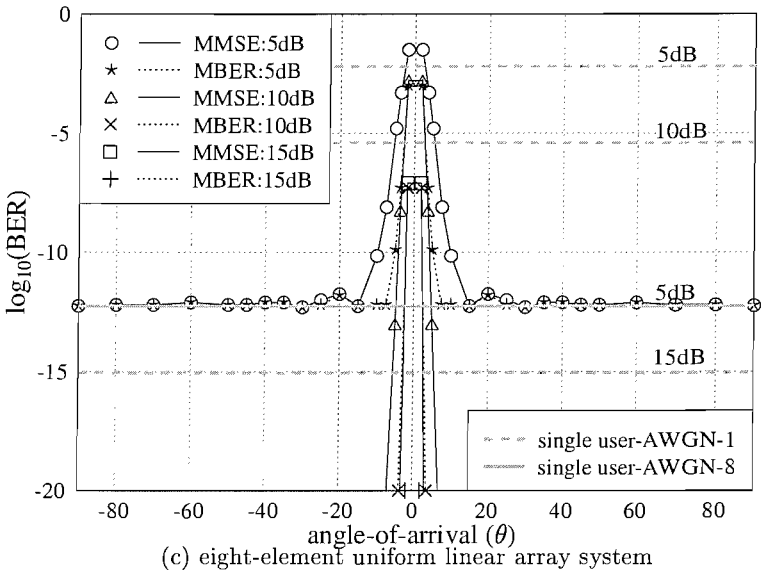
Figure 3.8: Comparison of the signal power required for a BER of $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, evaluated based on Equation (3.68), when using the MMSE and the MBER beamformers for two equal-power users, i.e. at SNR = INR, for (a) two-, (b) four- and (c) eight-element uniform linear arrays in an AWGN channel.



(a) two-element uniform linear array system



(b) four-element uniform linear array system

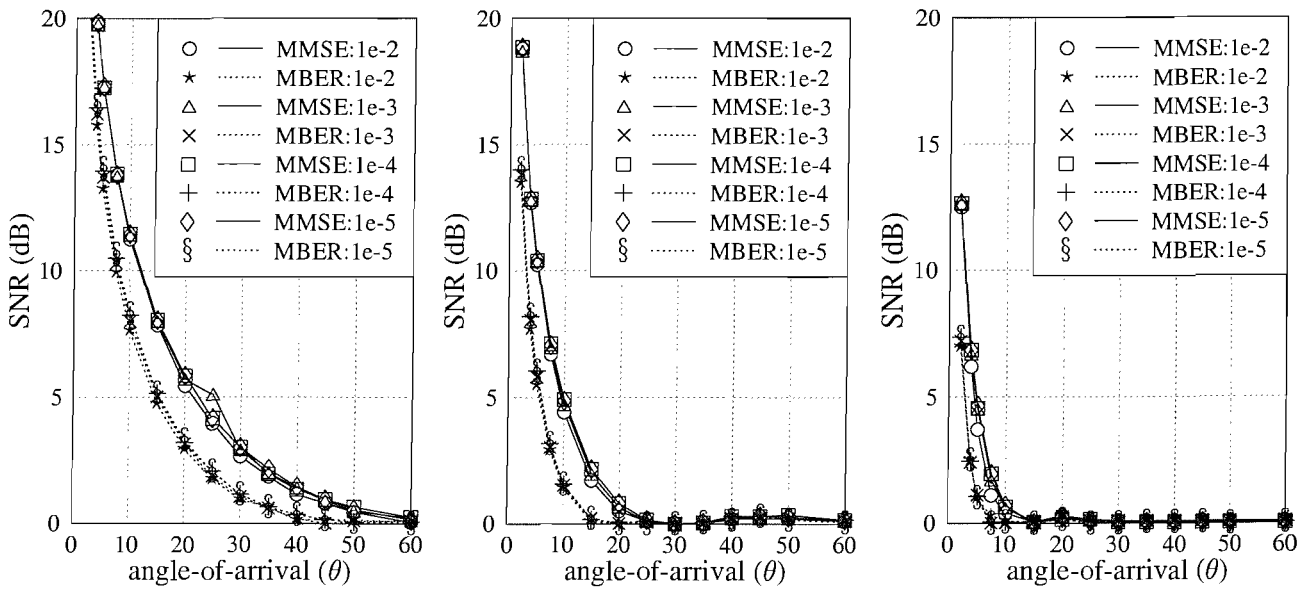


(c) eight-element uniform linear array system

Figure 3.9: Comparison of the achievable BER of the MMSE and the MBER beamformers, based on Equation (3.68), supporting fixed equal-power users, both experiencing SNR = INR of 5dB, 10dB, 15dB and 20dB for (a) two-, (b) four- and (c) eight-element uniform linear arrays in an AWGN channel.

10dB and 9.5dB, for the four- and eight-element arrays. This indicates that the MBER beamformer was capable of acquiring a substantially higher gain than the MMSE beamformer, as the number of array elements was doubled.

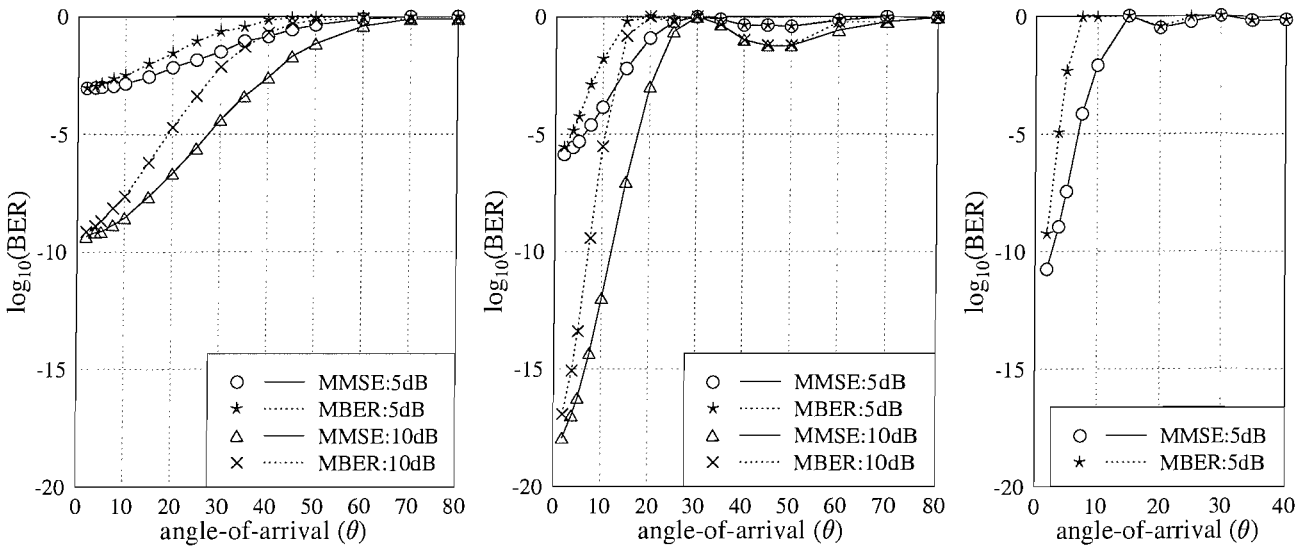
Figure 3.9 shows the achievable BER performance of both the MMSE and MBER beamformers from a different perspective, for a fixed desired user signal power, as the interfering signal's AOA is varied from -90^o to 90^o, i.e. for $-90^o \leq \theta \neq 0^o \leq 90^o$. For the two-element array of Figure 3.9 (a), only 10dB signal power is required by the MBER solution for achieving a BER performance of 10^{-5}, when the interfering user's signal is arriving from an angle of $|\theta| \geq 20^o$. By contrast, the MMSE solution requires an extra $2-3$dB higher signal power for achieving a BER of 10^{-5}. When the number of antenna array elements is increased, a lower signal power will be required for achieving a similar BER performance, as shown in Figures 3.9 (b) and 3.9 (c) for the four- and eight-element array, respectively. The signal power and BER differences between the MMSE and MBER solutions and the optimum single-user performance of the two-, four- and eight-element arrays are shown in Figures 3.10 and 3.11, respectively. We note that in Figures 3.8 to 3.11, the desired and interfering user had the same power, i.e. we assumed SNR = INR.



(a) $L = 2$ elements (b) $L = 4$ elements (c) $L = 8$ elements

Figure 3.10: The SNR difference between the MMSE and MBER beamformers and the optimum single user performance, i.e. $\text{SNR}_{\mathbf{w}} - \text{SNR}_{\text{single-user},L}$ as a function of the interfering user's AOA of $0^o < \theta \leq 60^o$ for BER of $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, evaluated based on Equation (3.68), for (a) two-, (b) four- and (c) eight-element antenna arrays in an AWGN channel. The desired user's signal is arriving from 0^o with respect to the normal of the antenna array.

Figure 3.12 portrays the near-far performance of both the MMSE and MBER solutions for a desired



(a) $L = 2$ elements (b) $L = 4$ elements (c) $L = 8$ elements

Figure 3.11: The BER difference of the MMSE and MBER beamformers from the optimum single user performance, i.e. $\log_{10}(\text{BER})_{\text{single-user},L} - \log_{10}(\text{BER})_{\mathbf{w}}$, based on Equation (3.68), as a function of the interfering user's AOA $0^o < \theta \leq 80^o$ or 40^o for SNR of 5dB and 10dB for (a) two- and (b) four-element antenna arrays as well as for SNR = 5dB for (c) eight-element antenna array in an AWGN channel. The desired user's signal is arriving from 0^o with respect to the normal of the antenna array.

signal power of 10dB and for a variable interferer power spanning the range of 36dB to -18dB. The initial array weight of the MBER solution at SNR = 0dB was using the MMSE solution, while at the subsequent SNRs the MBER solution's array weights derived for the previous SNR were used for initialisation. Figure 3.12 shows that the MBER solution is more near-far resistant than the MMSE solution, since the probability of error dependence of the former is about half of that recorded for the MMSE solution.

As seen in Equation (3.60) the probability density function (PDF) $p(y_R)$ of the decision variable y_R plays an important role in calculating the probability of error P_E, therefore let us now compare the PDF of both solutions. The PDF $p(y_R)$ conditioned on $b_1(n) = +1$ when the incoming signals are of equal power is shown in Figure 3.13, while that associated with INR = SNR +6dB is shown in Figure 3.14. It is observed that the shape of the PDF of both the MMSE and MBER beamformers is similar to each other and the respective PDF shape is Gaussian-like. The legends 'circle' and 'star' correspond to the real part of the noiseless output $\bar{y}_{R,q}$ of the beamformer in the set $\mathcal{Y}_R^{(+)}$, given $b_d(n) = +1$, as defined in Equation (3.41). In our simulations we always assume that the first user associated with $d = 1$ is the desired user. In both Figures 3.13 and 3.14, the noiseless state represented by the circles at positions $y_R = 0.87$ and $y_R = 0.95$ of Figure 3.13 a) and at positions $y_R = 1.92$ and $y_R = 1.95$ of Figure 3.14 a) correspond to the MMSE solution's real part of the noiseless output $\bar{y}_{R,q}$,

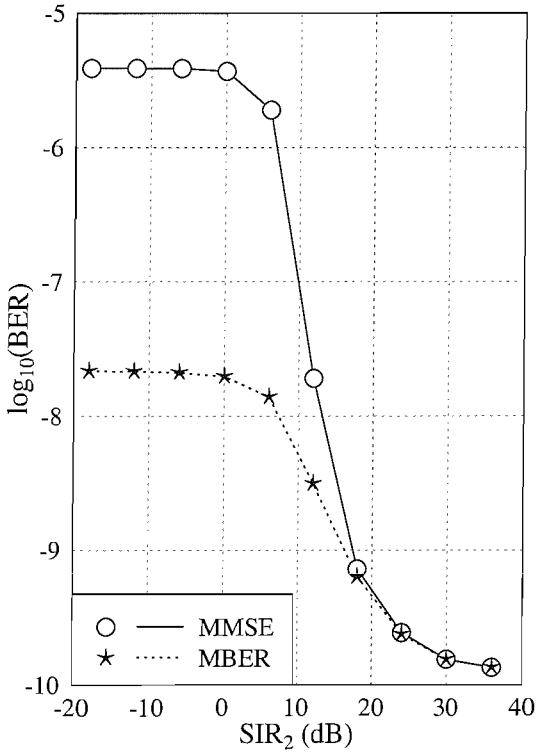
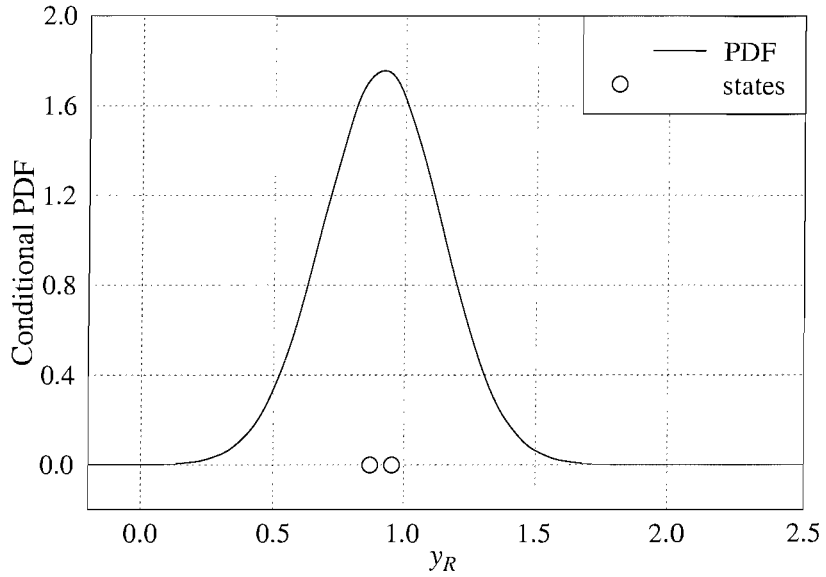


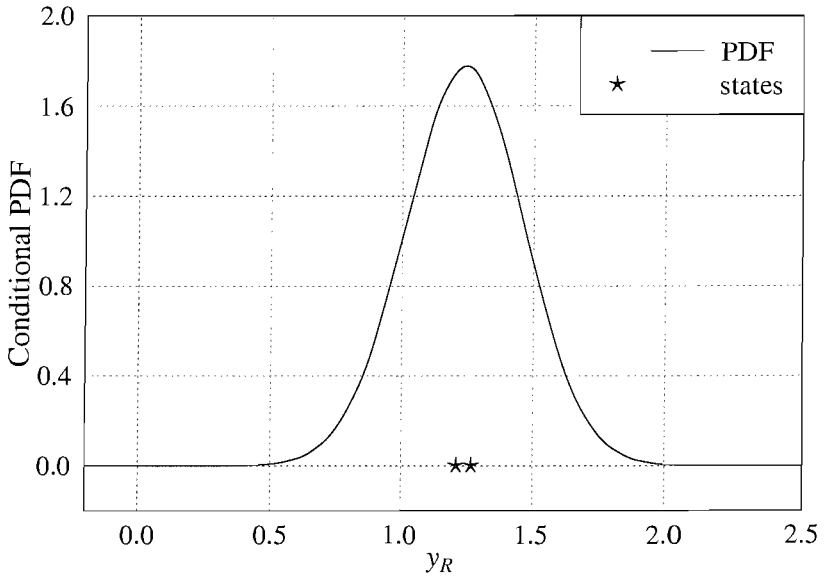
Figure 3.12: Influence of the near-far effect on the BER performance of the MMSE and MBER beamformers, evaluated based on Equation (3.68), for SNR = 10dB and SIR_2 in the range of $-18 \leq SIR_2 \leq 36$, where the desired user's signal is arriving from 0^o and that of the interferer from 30^o in an AWGN channel.

while the noiseless states represented by stars at positions $y_R = 1.21$ and $y_R = 1.28$ of Figure 3.13 b) and at positions $y_R = 2.43$ and $y_R = 2.49$ of Figure 3.14 b) denote the MBER solution's real part of the noiseless output $\bar{y}_{R,q}$. In Figure 3.13 we used the original data values, while in Figure 3.14 we have normalised the values by a factor of 2 for the sake of plotting the conditional PDF on the same Cartesian grid. Note that the respective MMSE and MBER noiseless states of Figures 3.13 a) and 3.13 b) are constituted by the real part of the noiseless beamformer output represented by the cross and the dot legends of the 'total y' plot seen at the right of Figure 3.1.

For an M-user BPSK-modulated beamforming scenario, there will be $N_b = 2^M$ number of possible transmitted bit sequences $\mathbf{b}$, which can be divided into two categories according to the desired user's transmitted bit b_d, where we have $N_{sb} = 2^{M-1}$ number of possible bit sequences, given $b_d = +1$, and another $N_{sb} = 2^{M-1}$ number of possible bit sequences, given $b_d = -1$. Specifically, for the two-user scenario we will have $N_b = 4$ and $N_{sb} = 2$ possible bit sequences corresponding to the two noiseless states $\bar{y}_{R,q}$ in all plots of Figures 3.13 and 3.14. The corresponding states of the MMSE and MBER solutions are different due to the different array weight values $\mathbf{w}$ of the MMSE and MBER solutions substituted into Equation (3.39). Having normalised the MBER array weight vector $\mathbf{w}$ to a unit-length, the BER or synonymously the probability of error $P_E(\mathbf{w})$ expressed in Equation (3.68) associated with

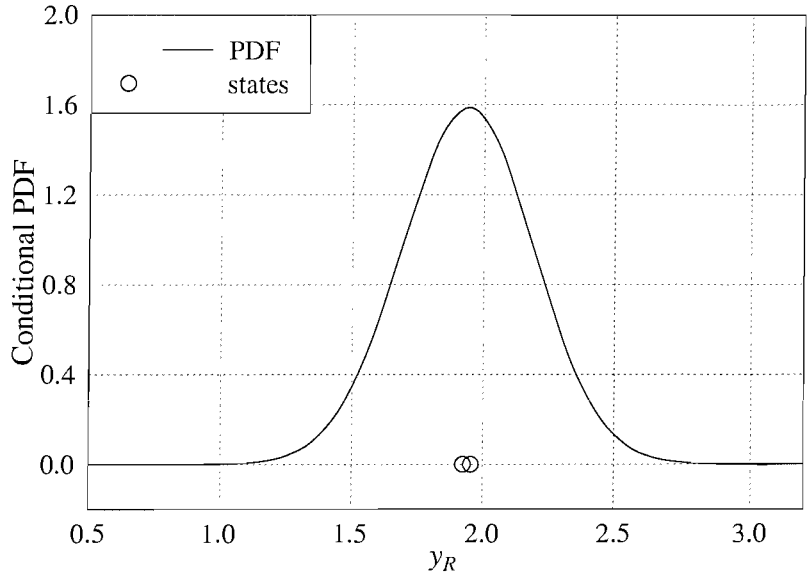


a) MMSE

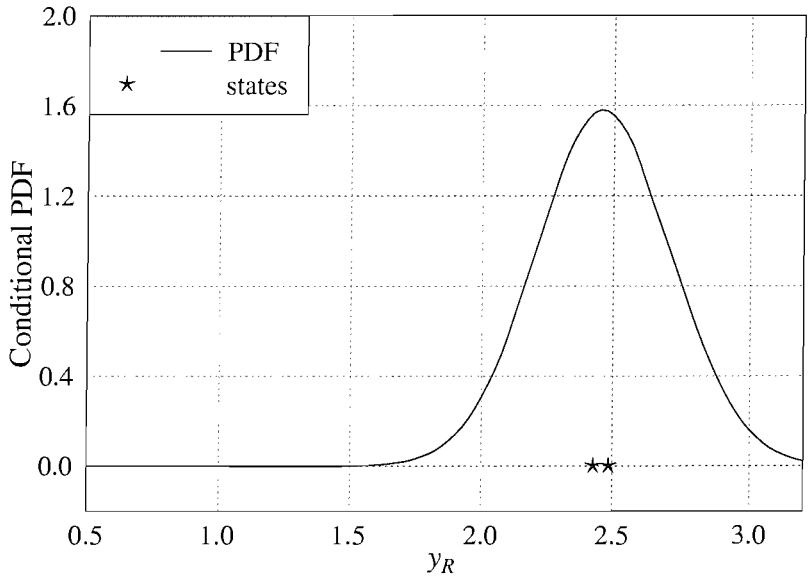


b) MBER

Figure 3.13: Conditional PDF of a) the MMSE and b) the MBER beamformers, given $b_1(n) = +1$ and the subset $\mathcal{Y}_R^{(+)}$ defined in Equation (3.41) of Section 3.3 for SNR = 10dB and INR = 10dB in an AWGN channel. The AOA of the desired and interfering user was 0^o and 30^o, respectively.



a) MMSE



b) MBER

Figure 3.14: Conditional PDF of a) the MMSE and b) the MBER beamformers, given $b_1(n) = +1$ and the subset $\mathcal{Y}_R^{(+)}$ defined in Equation (3.41) of Section 3.3 for SNR = 15dB and INR = SNR +6dB in an AWGN channel. The AOA of the desired and interfering user was 0^o and 30^o, respectively.

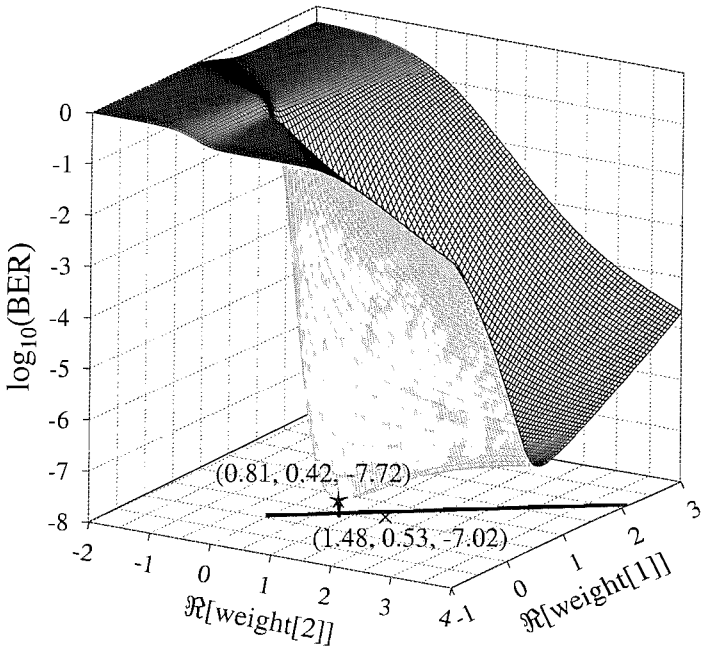


Figure 3.15: Error surface of the normalised MBER solution for $\alpha = 1$ and SNR = 10dB as well as INR = 10dB in an AWGN channel, where the AOA of the desired and interfering user was 0^o and 30^o, respectively. The thick line on the horizontal plane is where the MBER solutions reside, corresponding to BER = $10^{-7.72} \approx 1.91 \cdot 10^{-8}$. The MMSE solution is denoted by a cross, which should be located at $(1.48, 0.53)$ having a BER of $10^{-7.02} \approx 9.55 \cdot 10^{-8}$. The corresponding array weight's imaginary part is plotted in Figure 3.17 (b), such that it was located at $(-0.39, 3.01 \cdot 10^{-23})$ for the MBER solution and at $(-0.48, 0.48)$ for the MMSE solution. The BER was calculated based on Equation (3.68).

Equation (3.69) is predominantly determined by the minimum distance of the subset $\mathcal{Y}_R^{(+)}$ from the decision threshold of $y_R = 0$, namely by the smallest value of the noiseless beamformer output $\bar{y}_{R,q}$. Again, these noiseless beamformer outputs correspond to the left hand-side states denoted by the legends circle or star in Figures 3.13 and 3.14. In our case, although it was observed in Figures 3.13 and 3.14 that for each scenario the plots show only a slight difference between the states corresponding to the MMSE and the MBER solutions, the error probability determined by the Q-function [7, 221] $Q(z)$ of Equation (3.63) obeys a negative exponential function, which is a strongly non-linear function and hence may result in a significant BER difference at low values of z. As can be seen in Figures 3.13 a) and 3.13 b), the left circle in Figure 3.13 a) is at position $y_R = 0.87$, while in Figure 3.13 b) the left star is at position $y_R = 1.21$. Explicitly, $y_R = 0.87$ is closer to the decision boundary of $y_R = 0$, therefore it will result in a higher bit error probability calculated with the aid of Equation (3.68) as compared to that of the MBER solution having $y_R = 1.21$. This observation explains, why the MBER beamformer has a slightly lower BER than that of the MMSE beamformer.

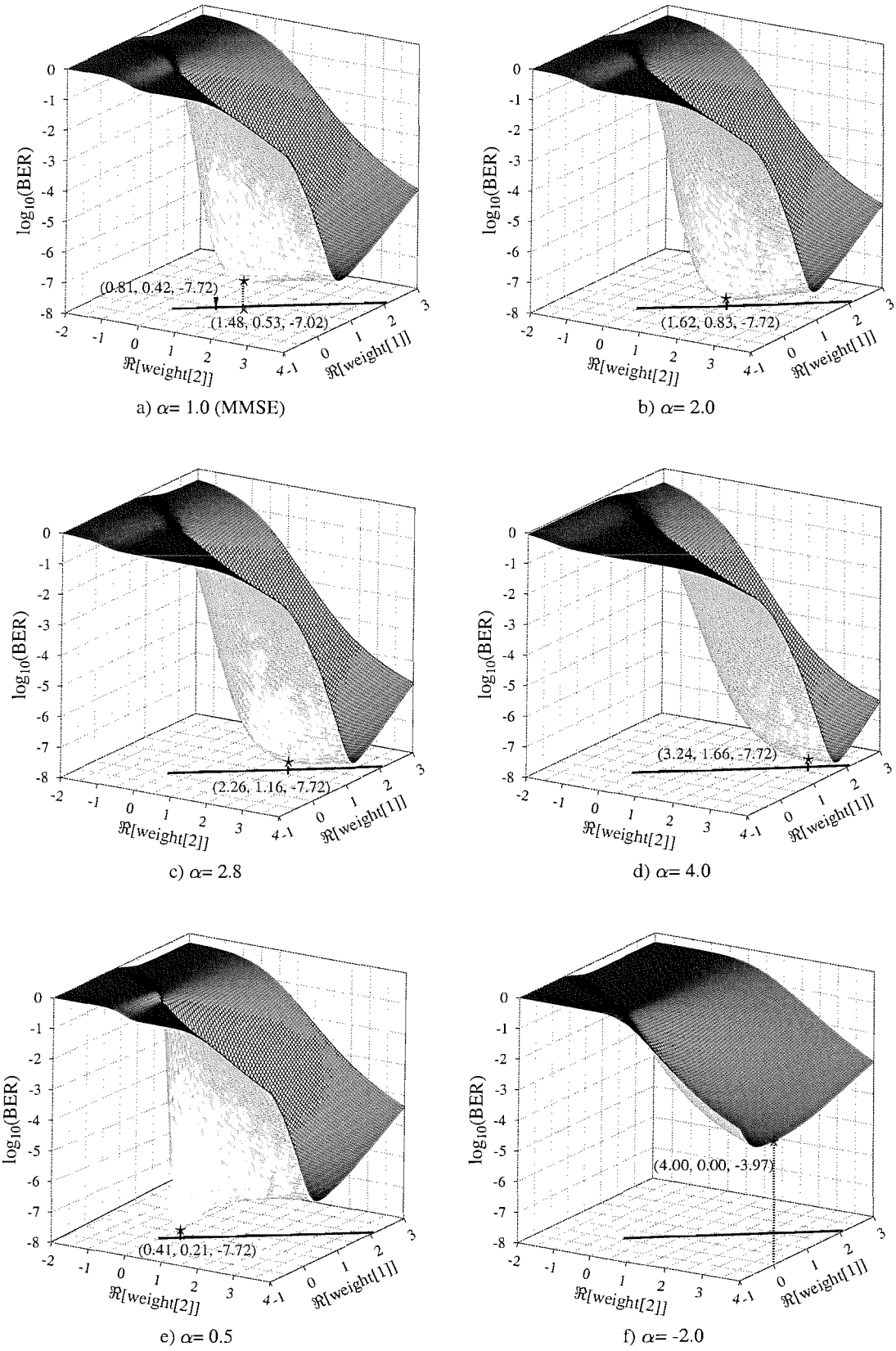
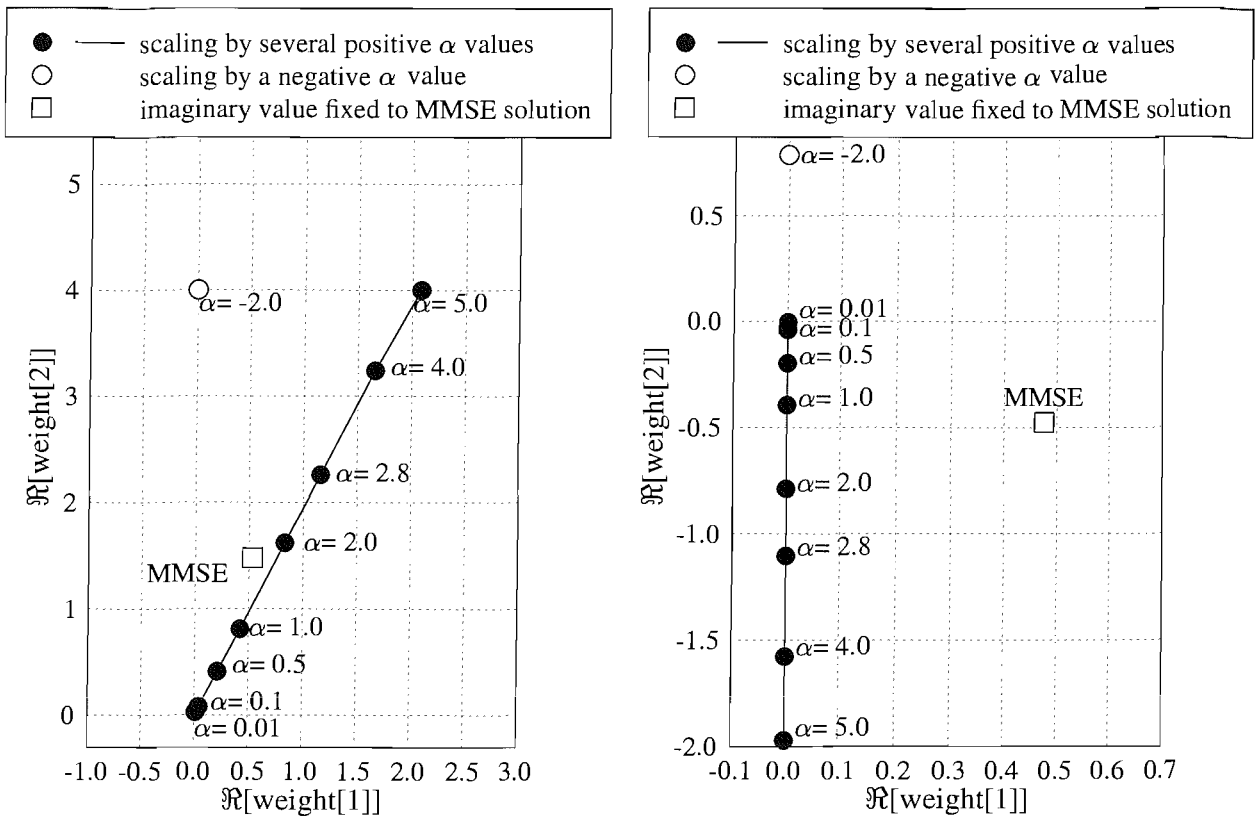


Figure 3.16: Error surfaces of the MMSE and MBER solutions plotted for different values of α; a) MMSE b) $\alpha = 2.0$ c) $\alpha = 2.8$ d) $\alpha = 4.0$ e) $\alpha = 0.5$ f) $\alpha = -2.0$ for SNR = 10dB and INR = 10dB in an AWGN channel, where the AOA of the desired and interfering user was 0^o and 30^o, respectively. The plot was varying the array weight's real part and the corresponding fixed array weight's imaginary part associated with the α value was shown in Figure 3.17 (b). The arrow in subfigure (a) points to the MBER solution associated with the $\alpha = 1$. The BER was evaluated based on Equation (3.68).

Let us now consider the BER surface of both the MMSE and MBER solutions. In a two-element system, each weight vector solution will consists of two complex weight values formulated as $\mathbf{w} = [w_1.r + jw_1.i \quad w_2.r + jw_2.i]^T$. Hence, there will be four parameters, the real and imaginary parts of each weight. Since we will be plotting the BER surface in a 3-dimensional space, we have to fix two of the variables, while varying the other two. We opted for fixing the imaginary values of both weights of both solutions to their optimum value for the sake of illustration. Figure 3.15 shows the BER surface of the normalised MBER solution, i.e. when we have $\alpha = 1$ in Equation (3.70), where the probability of error is $10^{-7.72} \approx 1.91 \cdot 10^{-8}$. On the other hand, the BER surface of the MMSE solution characterised in Figure 3.16 a) resulted in a BER of $10^{-7.02} \approx 9.55 \cdot 10^{-8}$, which is higher than that of the MBER solution. Although it was stated before in the context of Equation (3.70) that the BER surface exhibits an infinite number of MBER solutions, the plots portrayed in Figures 3.15 and 3.16 exhibit a single solution as a result of fixing the imaginary parts of both weights. For the sake of demonstrating that multiple solutions do exist in the context of the MBER optimisation criterion according to Equation (3.70), which implies that we have $c_{q,+}(\alpha\mathbf{w}) = c_{q,+}(\mathbf{w})$, we will multiply the MBER solution of $\ddot{\mathbf{w}}_{\text{MBER}}$ derived for $\alpha = 1$ by various positive values of α and verify that the resultant minimum BER value is the same as before, i.e. BER $\approx 1.91 \cdot 10^{-8}$. The BER surfaces associated with different values of α as well as different real parts of the weights and for fixed imaginary parts of αw_1 and αw_2 are plotted in Figures 3.16 b) to 3.16 f). Observe that all error surfaces of Figure 3.16, except for the MMSE solution of Figure 3.16 a) and the MBER solution of Figure 3.16 f) associated with $\alpha = -2$ (where α is negative) give the same minimum probability of error value as the one seen in Figure 3.15. For positive values of α, the value of the real and imaginary parts at the minimum BER point is exactly a factor α times higher than the normalised weight $\ddot{\mathbf{w}}_{\text{MBER}}$, which is the weight in the case of $\alpha = 1$.

The real part $\Re[\mathbf{w}]$ of the array weights of all solutions is plotted in Figure 3.17 (a), where we observe that all the optimum MBER solutions are on a straight line. This line is where the infinite number of MBER solutions associated with the scenario considered resides, i.e. for any point on the line we arrive at the minimum BER, where we have BER $\approx 1.91 \cdot 10^{-8}$, provided that the imaginary parts of the weights $\Im[\mathbf{w}]$ are scaled accordingly, i.e. with respect to that shown in Figure 3.17 (b). A complete set of weight values is determined by referring to the associated α value in Figures 3.17 (a) and 3.17 (b), for the real and imaginary parts, respectively. For example, when $\alpha = 1$, the associated weight locations, arranged in the form of (weight[1], weight[2]), are at $(0.42, 0.81)$ and $(3.00 \cdot 10^{-23}, -0.39)$ in Figures 3.17 (a) and 3.17 (b), respectively, resulting to beamforming weight vector of $\mathbf{w}_{\text{MBER},\alpha=1} = [0.42 + j0 \quad 0.81 - j0.39]^T$. In a similar way, the MMSE beamforming weight vector will be $\mathbf{w}_{\text{MMSE}} = [0.53 + j0.48 \quad 1.48 - j0.48]^T$. The MBER solutions associated with the line seen in Figure 3.17 (a) were also shown in Figure 3.15. Apart from varying the real part of the weights, it is also possible to accordingly vary their imaginary part as shown in Figure 3.17 (b) or both parts of one of the weights, but all these scenarios will yield the same weight values and thus the same minimum BER value.



(a) real-part of weight location

(b) imaginary-part of weight location

Figure 3.17: Linear relationship of (a) the two MBER array weights' real part and (b) the two MBER array weights' imaginary part for SNR = 10dB and INR = 10dB in an AWGN channel. The AOA of the desired and interfering user was 0^o and 30^o, respectively.

3.5.2 Supporting Five Users Employing a Two-Element Uniform Linear Array

To elaborate a little further a system employing a two-element antenna array for supporting five users can be viewed in Figure 3.18, where the signal of interest (SOI) of the desired user arrives from a direction of 15° while the interfering signals arrive from $-30°, 60°, 80°, -70°$ relative to the normal of the array.

Similar to the example of Section 3.5.1, the BER performance of the MBER solution is compared to that of the MMSE solution under four different conditions; (a) the desired user and all the four interfering sources have an equal power; (b) all the interfering sources have 6dB higher power than the desired user; (c) the desired user and the interfering sources $3, 4, 5$ have an equal power, while the interfering source 2 has 6dB higher power than the desired user; and (d) the desired user and the interfering sources $2, 3, 4$ have an equal power, but the interfering source 5 has 6dB higher power than the desired user.

The results shown in Figure 3.19, in particular, the system employing a two-element antenna array, demonstrate the superior performance of the MBER beamforming technique in comparison to the classic MMSE approach. The results also show that the MBER solution is more robust to the

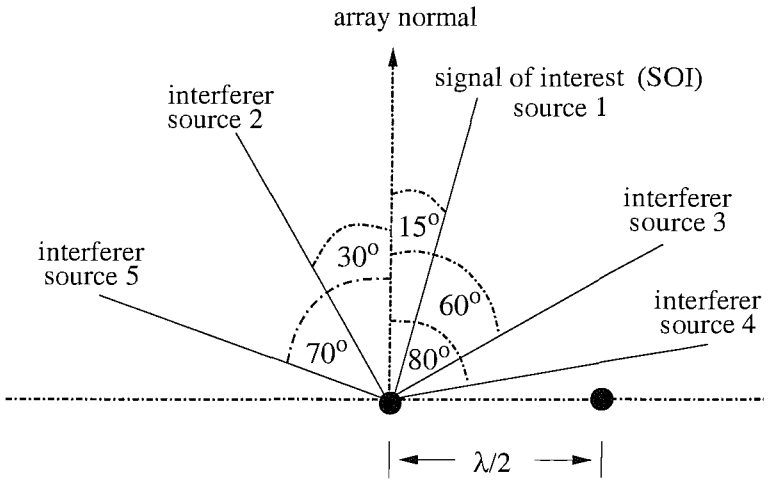


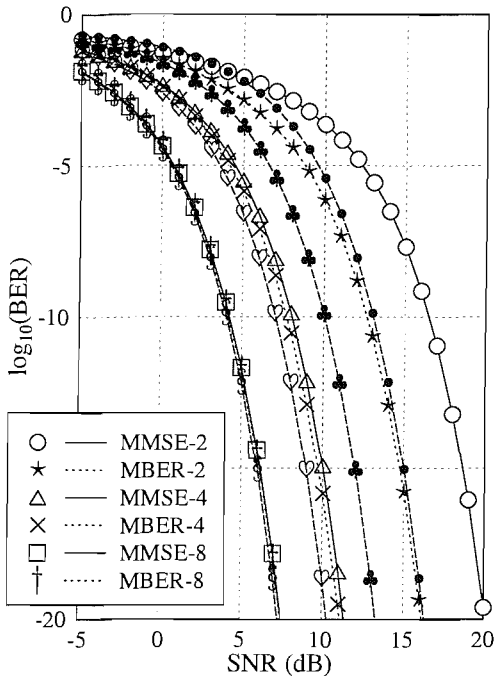
Figure 3.18: Two-element antenna array having an element spacing of $\lambda/2$ and receiving signals from **15°** (SOI), $-$**30°**, **60°**, **80°** and $-$**70°** (the interfering signals).

near-far effect, i.e. the same BER performance is maintained, regardless of the fact that the users' transmit powers are different, even if the desired user's signal arrives from an angle close to that of the interfering users. This is clearly demonstrated, if we compare the results seen in Figures 3.19 (a) and Figure 3.19 (c), where the AOAs are $15°, -30°, 60°, 80°$ and $-70°$ and when the user whose signal is arriving from an angle of $-30°$ has 6dB higher power than the desired user's signal arriving from a direction of $15°$. From Figures 3.19 (c) and 3.19 (d), we also observe that the BER performance of the MBER solution improves, when the stronger interfering signal arrives from an angle closer to the desired signal. Figure 3.19 (b) shows that the achievable BER performance is degraded, when the power of all the interferers is higher than that of the desired user, but the degradation suffered by the MBER solution is not dramatic, when compared to that of the MMSE beamformer, which exhibits a high BER floor. We will return to this phenomenon at a later stage in the context of Figures 3.25 and 3.26.

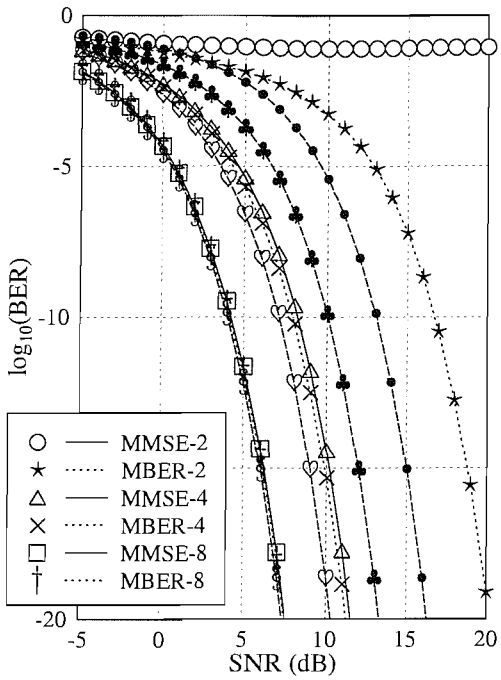
AOA	Figure 3.18	Figure 3.20 (a)	Figure 3.20 (b)	Figure 3.20 (c)	Figure 3.20 (d)
θ_1	15^o	15^o	15^o	15^o	15^o
θ_2	-30^o	-7.5^o	3.75^o	9.375^o	12.1875^o
θ_3	60^o	37.5^o	26.25^o	20.625^o	17.8125^o
θ_4	80^o	80^o	80^o	80^o	80^o
θ_5	-70^o	-70^o	-70^o	-70^o	-70^o

Table 3.4: AOA of the users for the beamforming simulation of Figures 3.18 and 3.20. Note that the AOAs of user 2 and 3, i.e. θ_2 and θ_3, respectively, were varied by repeatedly halving their angular separation with respect to the desired user located at $\theta_1 = 15^o$.

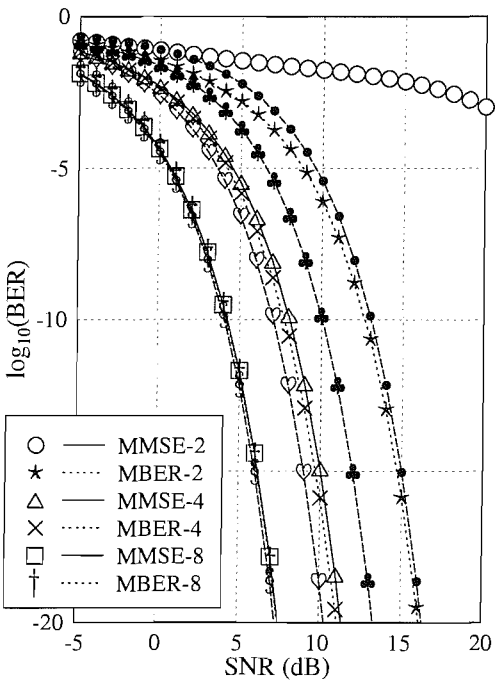
Let us investigate the BER performance of both the MMSE and MBER solution as the interfering sources 2 and 3 portrayed in Figure 3.18 are gradually moved towards the SOI located at 15^o with



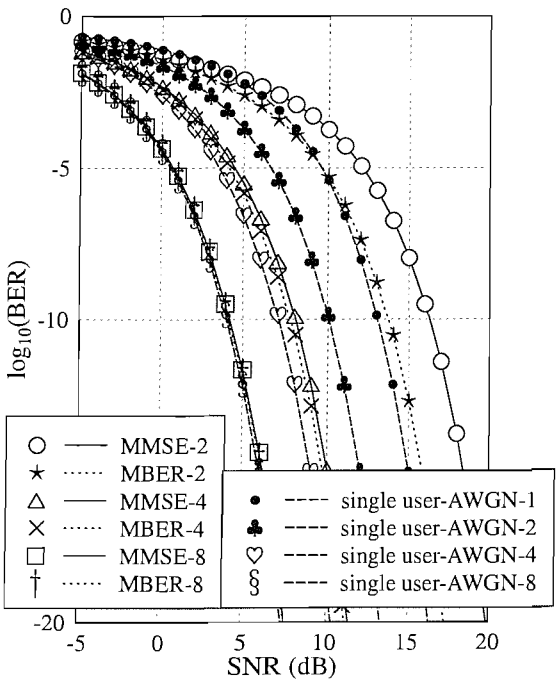
(a) SNR = INR_i for $i = 2, 3, 4, 5$



(b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$

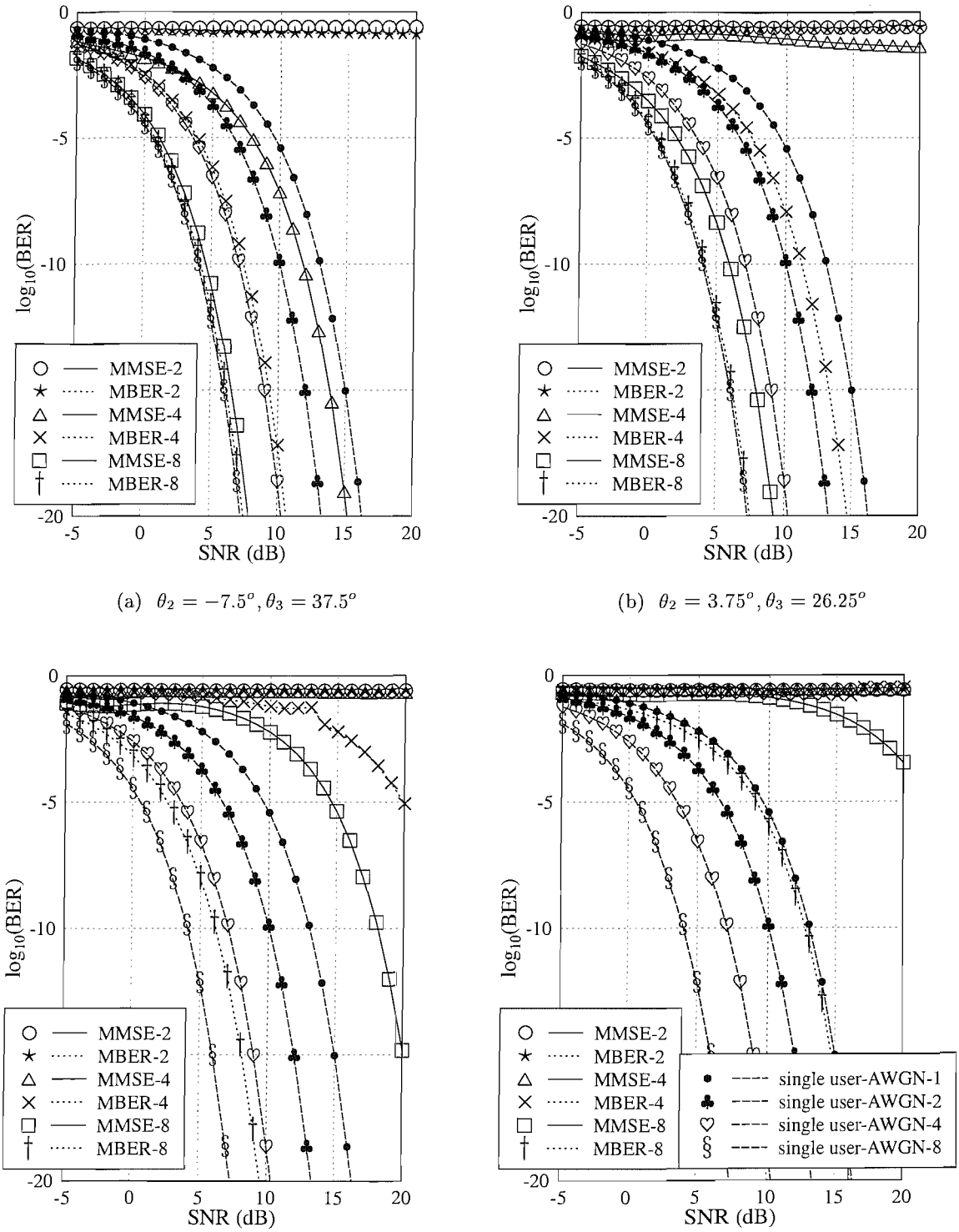


(c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB



(d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB

Figure 3.19: Comparison of the BER performance of the MMSE and the MBER beamformers, based on Equation (3.68), under four different conditions; (a) SNR = INR_i for $i = 2, 3, 4, 5$, (b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$, (c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB for two-, four- and eight-element uniform linear arrays. The associated interference scenario was plotted in Figure 3.18. These results were recorded when communicating over an AWGN channel.



(a) $\theta_2 = -7.5^o, \theta_3 = 37.5^o$

(b) $\theta_2 = 3.75^o, \theta_3 = 26.25^o$

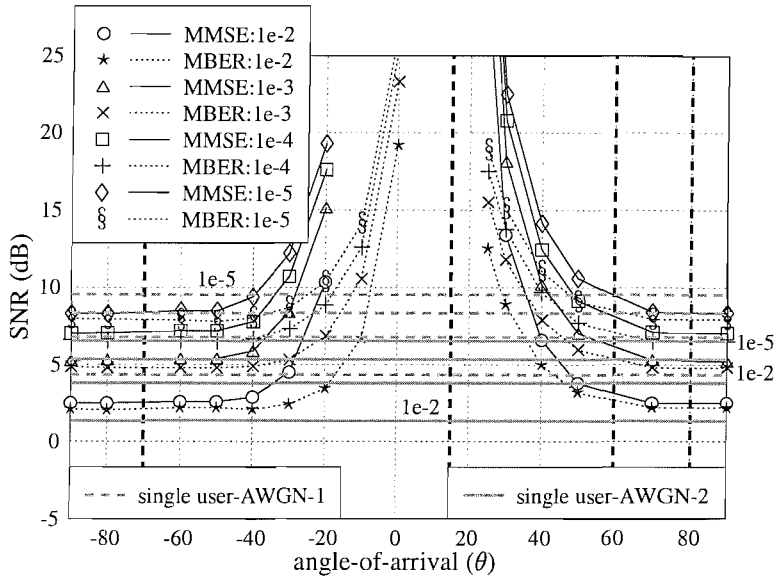
(c) $\theta_2 = 9.375^o, \theta_3 = 20.625^o$

(d) $\theta_2 = 12.1875^o, \theta_3 = 17.8125^o$

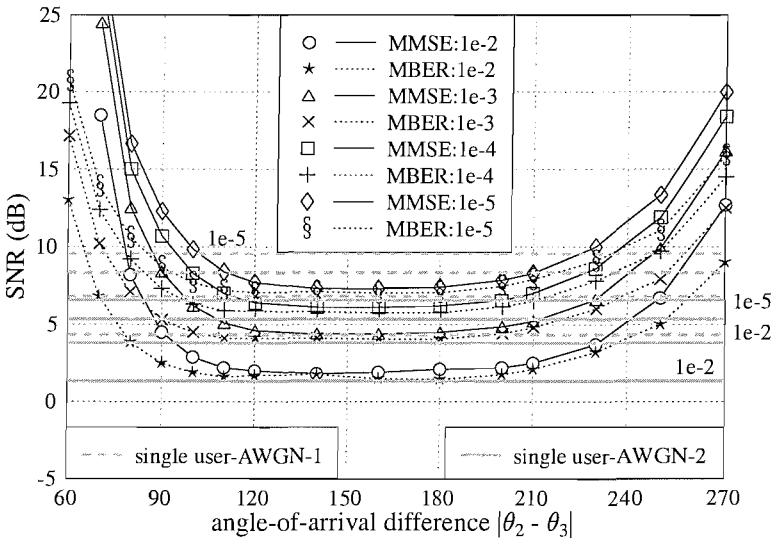
Figure 3.20: Comparison of the BER performance of the MMSE and the MBER beamformers, based on Equation (3.68), for five equal-power users, i.e. SNR = INR, in conjunction with two-, four- and eight-element uniform linear arrays in an AWGN channel. The AOA of the desired and interfering users was plotted in Figure 3.18, but those AOA of the interfering source 2 and 3 were varied according to Table 3.4.

respect to the array normal. As observed in Figure 3.20 (a), the BER performance of both the MMSE and MBER solution is severely degraded in the context of a two-element array, when the angular separation of the two nearest interfering sources with respect to the SOI, i.e. source 2 and 3 was halved in comparison to their original location, i.e. they are now at -7.5^o and 37.5^o, respectively. Despite increasing the dimensionality of the antenna array to four elements, the BER versus SNR performance of the MMSE solution was degraded by about 4dB, when compared to that observed in Figure 3.19 (a), while the MBER solution managed to maintain a similar performance to the scenario having twice the angular separation for the interfering sources 2 and 3 with the advent of doubling the number of array elements to four. For example, the MBER solution still achieves a BER of 10^{-5} for a desired user signal power of 4dB, while the MMSE solution would require a desired user signal power of about 8dB for achieving the same BER performance. Observe also in Figure 3.20 (d) that for the eight-element antenna array operating at SNR = 9dB the MBER solution maintains a BER of 10^{-5}, but the MMSE solution would require more than 20dB desired user signal power to achieve a BER of 10^{-5}. More BER performance characteristics generated by repeteadly halving the angular separation between the interfering sources 2 and 3 in the context of two-, four- and eight-element antenna arrays can be seen in Figure 3.20 (a), 3.20 (b), 3.20 (c) and 3.20 (d) with the angular separation between user 2 and 3 set to 45^o, 22.5^o, 11.25^o and 5.625^o, respectively. Note that the angular separation is halved under the condition that the SOI is always at the centre between user 2 and 3. Explicitly, recall that the angular positions of the five-user scenario were shown in Table 3.4.

Returning to the user scenario of Figure 3.18, Figure 3.21 (a) shows the performance of both the MMSE and MBER beamformers as the AOA of the interfering source 2 is varied from -90^o to 90^o, i.e. for $-90^o \leq \theta_2 \neq 15^o, 60^o, 80^o, -70^o \leq 90^o$. Note that the interfering signals must not arrive from $15^o, 60^o, 80^o$ and -70^o, since these are the arriving angles of the other four remaining users, i.e. those of the SOI and the other three interfering signals, otherwise the corresponding signals could not be separated. Therefore we indicate these AOA values by thick vertical dashed lines in Figure 3.21 (a), implying that the corresponding SNR values do not exist. It is observed from Figure 3.21 (a) that the signal power required for the five users in order to achieve a specific BER performance of $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$ increases as the interfering user 2 moves closer to the desired user located at 15^o. It is also clear from Figure 3.21 (a) that the MBER solution always requires a lower desired user signal power than that required by the MMSE solution. The range of the requirement of different signal powers for the MMSE and MBER beamformers reduces, when the interfering user 2 arrives from an angular direction that obeys $-40 \geq \theta_2 \geq 50^o$ with respect to the array normal. It is also observed that the two-element system is particularly more robust in the context of the MBER solution, since it is capable of separating interfering user 2 whose angular separation is as low as 10^o or 15^o, requiring a desired user signal power of less than 27dB. By contrast, for the MMSE solution, the 'just-tolerable' AOA of user 2 is 30^o or -20^o, which results in a minimum AOA separation of 15^o or 35^o with respect to the desired user located at 15^o.



(a) AOA of interferer 2(θ_2)



(b) AOA difference of interferer 2 and 3 ($|\theta_2 - \theta_3|$)

Figure 3.21: Comparison of the SNR required for achieving a BER of $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, evaluated based on Equation (3.68), between the MMSE and the MBER beamformers for two equal-power users, i.e. for SNR = INR, as a function of (a) the AOA of interferer 2 and (b) the AOA difference of interferer 2 and 3, i.e. $|\theta_2 - \theta_3|$. The thick vertical dashed lines in subfigure (a) mark the corresponding AOA of the remaining four users. The associated interference scenario was plotted in Figure 3.18. These results were recorded when communicating over an AWGN channel.

In Figure 3.21 (b), we also consider a scenario of simultaneously moving both the interfering sources 2 and 3 in equal angular separations from the SOI located at 15^o. For simplicity, we plot the desired user's signal power as a function of the angular separation between the interfering sources 2 and 3, i.e. versus $|\theta_2 - \theta_3|$, in conjunction with target BERs of $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$. It can be seen from Figure 3.21 (b) that the MMSE solution would require a minimum angular separation of 70^o between users 2 and 3 at a target BER performance of 10^{-2} as well as 10^{-3}, since otherwise an excessive SNR would be required for meeting these target BERs. This is indicated in Figure 3.21 (b) by the fact that the corresponding approximately 18dB and 25dB points are the lowest-SNR points for the target BERs of 10^{-2} and 10^{-3}, respectively. On the other hand, the MBER solution managed to achieve the BERs of 10^{-2} and 10^{-3} for signal powers of 7dB and 10dB, respectively, at the same angular separation of 70^o. It may also be observed in Figure 3.21 (b) that the associated signal power requirement is reduced, as users 2 and 3 are further separated, and it becomes approximately constant in the angular separation region of $120^o \leq |\theta_2 - \theta_3| \leq 200^o$. For angular differences higher than 200^o, the associated signal power requirement increases again. This is a consequence of the circular symmetry with respect to 360^o, since the maximum angular separation from a location is limited to 180^o. Therefore, any angular separation θ_i greater than 180^o will produce the same effect, as if the separation was $360 - \theta_i$. For equal angular separations with respect to the SOI, the variation of the angular position of users 2 and 3 in the range of $-150^o \leq \theta_2 \leq 10^o$ and $20^o \leq \theta_3 \leq 180^o$ with respect to the array normal, respectively, results in angular differences of $10^o \leq |\theta_2 - \theta_3| \leq 330^o$. For desired user signal powers spanning the range of -5dB to 25dB, the acceptable angular separation of interfering users 2 and 3 for the MMSE and MBER beamformers in conjunction with the user scenario of Figure 3.18 are $70^o \leq |\theta_2 - \theta_3| \leq 270^o$ and $60^o \leq |\theta_2 - \theta_3| \leq 270^o$, respectively. Again, it becomes explicit also from Figure 3.21 (b) that the MBER solution performed better than the MMSE solution, i.e. it required a lower signal power, particularly when the angular separation was in the ranges of $60^o \leq |\theta_2 - \theta_3| \leq 100^o$ and $250^o \leq |\theta_2 - \theta_3| \leq 270^o$. Note that in all cases studied in this section we assumed that all five users are of equal signal power.

The system's robustness to the near-far effect is further confirmed by Figure 3.22, where the performance of a system is characterised in conjunction with variable values of SIR_2, spanning the range of -18dB to 36dB. The SNR was fixed to 10dB and we had $\text{SIR}_i = 24$ dB for $i = 3, 4, 5$. The initial array weight of the MBER beamformer used for each SNR was the MBER solution generated at the previous SNR except at 0dB, where the MMSE solution was used as the initial weight value.

A comparison of the beam-patterns of the MBER and MMSE beamformer was provided in Figure 3.23. Having equal-power users of each transmitting at SNR = 10dB, at first sight it may appear that the MMSE beamformer exhibits a better amplitude response than the MBER beamformer. More explicitly, the MMSE beamformer exhibits more attenuated magnitude responses at the angles of three of the four interfering sources, namely for the users transmitting from directions of $-70°, 60°, 80°$. By contrast, the MBER beamformer appears to have a better amplitude response for the interfering user

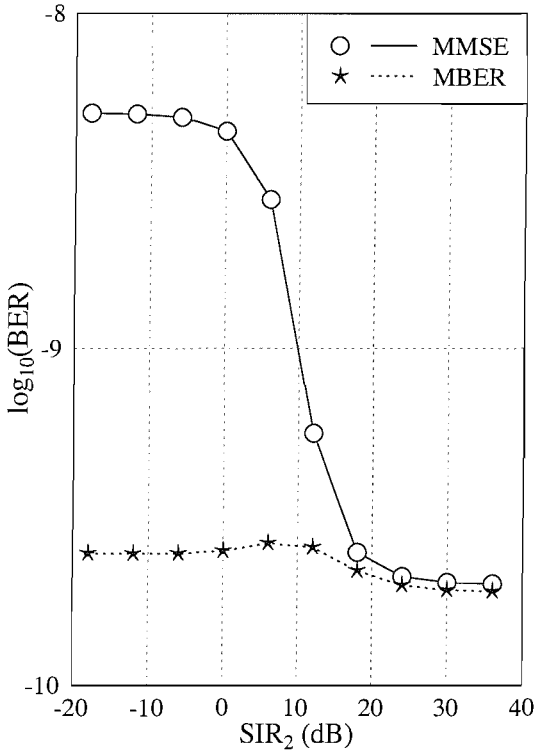


Figure 3.22: Influence of the near-far effect on the BER performance of the MMSE and MBER beamformers, evaluated based on Equation (3.68), for SNR = 10dB and SIR_i = 24dB for $i = 3, 4, 5$, when communicating over an AWGN channel. The associated interference scenario was plotted in Figure 3.18.

at an angle of $-30°$, i.e. for the user at the lowest angular separation. Therefore, if the BER performance was determined on the basis of the amplitude response alone, it would be more beneficial to use the MMSE beamformer. However, considering the magnitude response alone and neglecting the phase response is misleading, since both the magnitude and phase of $F(\theta)$ should be used together for characterising the beamformer.

The absolute value of the real and imaginary parts of $[F(\theta)]$, i.e. $|\Re[F(\theta)]|$ and $|\Im[F(\theta)]|$, respectively, corresponding to the scenario of Figures 3.23 a) and 3.23 b) can be viewed in Figures 3.23 c) and 3.23 d). In our case, we are only interested in the real response of $|\Re[F(\theta)]|$ due to our decision function defined in Equation (3.19), which is valid for BPSK transmissions. In terms of the response $|\Im[F(\theta)]|$, the MMSE solution appears to perform better, particularly at the 15°- angle of the desired user's signal, where it exhibits a deep null. This is crucial for the sake of minimising the function $\text{MSE} = E[|b_d(n) - y(n)|^2]$, since $b_d(n)$ is real-valued, although this function is irrelevant as regards to the system's BER performance. By contrast, the MBER beamformer exploits the systems' 'resources' more intelligently. This may be visualised from the total beamformer output y plot of Figure 3.34.

As mentioned earlier, the best indicator of the expected BER performance is constituted by the PDF of the decision variable, as given in Equation (3.59). Figure 3.24 b) depicts the PDF of the

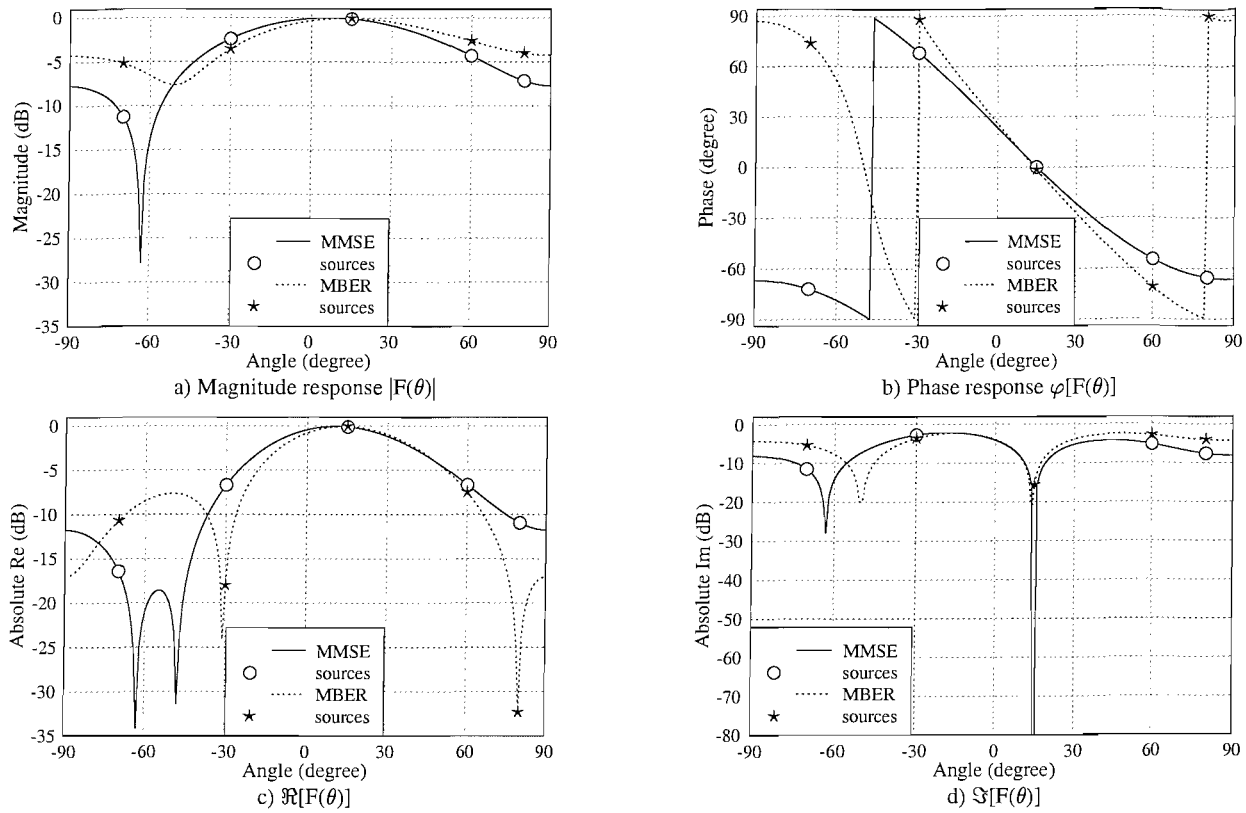
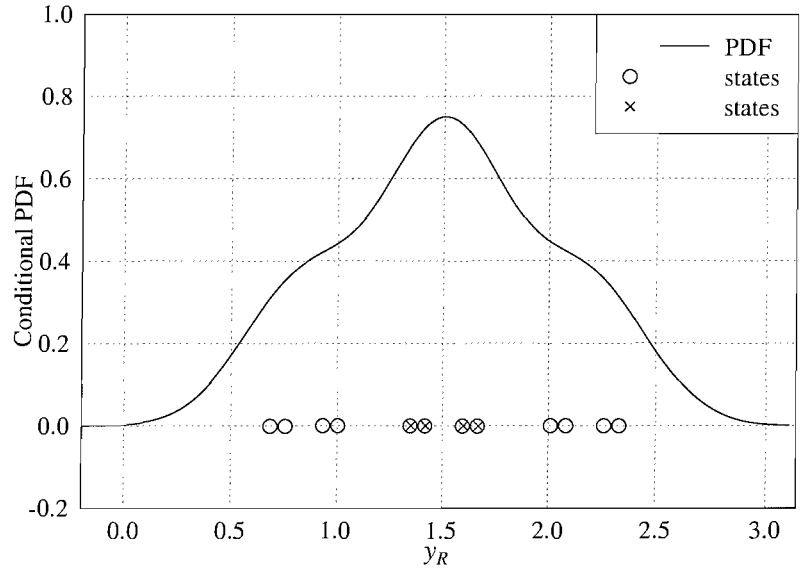
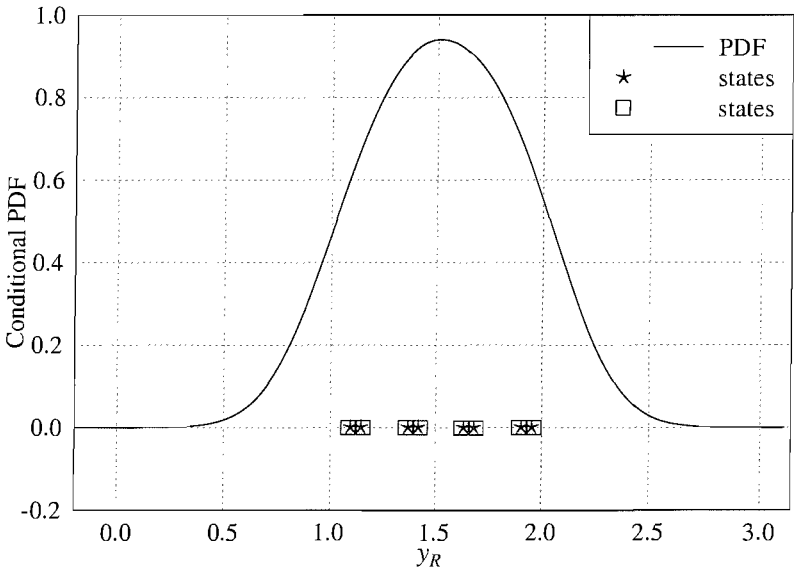


Figure 3.23: Comparison of the beam-patterns of both the MMSE and MBER beamformers for $\mathbf{SNR} = 10\mathbf{dB}$ and $\mathbf{INR}_i = 10\mathbf{dB}$ for $i = 2, 3, 4, 5$, when communicating over an AWGN channel. The associated interference scenario was plotted in Figure 3.18.

MBER solution conditioned on $b_1(n) = +1$, while that of the MMSE is shown in Figure 3.24 a), both using the interference scenario of Figure 3.18, as in Figure 3.23. Having normalised the MBER beamformer's weight vector, the associated BER expression of Equations (3.61) and (3.69), which is based on the Q-function defined in Equation (3.63) is dominated by the specific decision variable value in the subset $\mathcal{Y}_R^{(+)}$, which is closest to the decision boundary $y_R = 0$, namely by the circle and square at the left-most positions in Figures 3.24 a) and 3.24 b) for the MMSE and MBER solutions, respectively. For the $M = 5$-user scenario, there are $2^{M-1} = 16$ points in both the MMSE and MBER subset associated with $b_1(n) = +1$, as seen in Figures 3.24 a) and 3.24 b). For the sake of showing all the states, i.e. the real part of the noiseless output $\bar{y}_{R,q}$ of the beamformer as defined in Equation (3.40), which are sometimes quite similar hence may overlap each other, we employ circles and crosses as markers for the real part of the MMSE beamformer's noiseless output states, while star and square markers for the real part of the MBER beamformer's noiseless output states. Table 3.5 lists the 16 possible values of the noiseless output $\bar{y}_{R,q}$ for the real component of both the MMSE and MBER beamforming solutions. Observe that the nearest point of the MBER solution to the decision boundary $y_R = 0$ is at $y_R = 1.08$, which corresponds to nearly twice the distance from the decision boundary in comparison to the nearest MMSE decision variable at $y_R = 0.68$. The Q-function



a) MMSE



b) MBER

Figure 3.24: Conditional PDF of a) the MMSE and b) the MBER beamformers given $b_1(n) = +1$ and the subset $\mathcal{Y}_R^{(+)}$ defined in Equation (3.41) of Section 3.3 for SNR = 10dB and $\text{INR}_i = 10\text{dB}$ for $i = 2, 3, 4, 5$, when communicating over an AWGN channel. The associated interference scenario was plotted in Figure 3.18.

State q	MMSE	MBER
1:	0.68	1.37
2:	1.34	1.43
3:	1.35	1.90
4:	2.00	1.96
5:	0.93	1.36
6:	1.59	1.42
7:	1.60	1.89
8:	2.26	1.95
9:	0.75	1.09
10:	1.42	1.16
11:	1.42	1.63
12:	2.08	1.69
13:	1.00	1.08
14:	1.66	1.14
15:	1.67	1.62
16:	2.33	1.67

Table 3.5: The real part of the noiseless output $\bar{y}_{R,q}$ of the beamformer of Equation (3.41) for the MMSE and MBER solution given that $b_d(n) = +1$, as inferred from Figures 3.24 a) and 3.24 b), respectively.

$Q(\bar{y}_{R,q})$ results in a high probability of error $P_E(\mathbf{w})$ according to Equation (3.68) when the noiseless beamformer's ouput value is nearer to the decision boundary of $y_R = 0$. This explains, why the MBER beamformer has a lower BER compared to that of the MMSE beamformer.

Returning to Figure 3.19 (b) for a moment, it is intriguing to speculate why the MMSE beamformer is overwhelmed by interference, thus exhibiting a high BER floor. To investigate this phenomenon, let us first consider its beam pattern. Given an SNR of 15dB and $\text{INR}_i = \text{SNR} + 6\text{dB}$ for $i = 2, 3, 4, 5$, the corresponding beam patterns are presented in Figure 3.25. Generally, the beam patterns have not changed dramatically in comparison to Figure 3.23 for the equal-power users experiencing SNR = 10dB. Since no explicit conclusion could be drawn from these patterns, we proceed to plotting the PDFs of the decision variables, which are shown in Figure 3.26 for the unequal desired and interfering user power scenario. From the location of the states, i.e. from the real component of the noiseless beamformer's output $\bar{y}_{R,q}$ located on the y_R-axis it becomes more obvious, why the MMSE beamformer has a high BER floor. Explicitly, it was found for the MMSE solution that $\mathcal{Y}_R^{(-)}$ and $\mathcal{Y}_R^{(+)}$ are linearly nonseparable, i.e. given a transmitted bit of $b_d(n) = +1$, the resultant noiseless beamformer output $\bar{y}_{R,q}$ should be in the positive subset of $\mathcal{Y}_R^{(+)}$ as defined in Equation (3.41), but one of the total of 16 possible states has violated this rule. In other words, the corresponding state falls in the wrong subset $\mathcal{Y}_R^{(-)}$. As seen in Figure 3.26 a) as well as listed in the second column of Table 3.6, from the 16 points in the subset $\mathcal{Y}_R^{(+)}$, one is at the wrong side of the decision boundary of $y_R = 0$, yielding $y_R = -0.11$ and another one is nearly on the decision boundary of $y_R = 0$, namely at $y_R = 0.01$. As

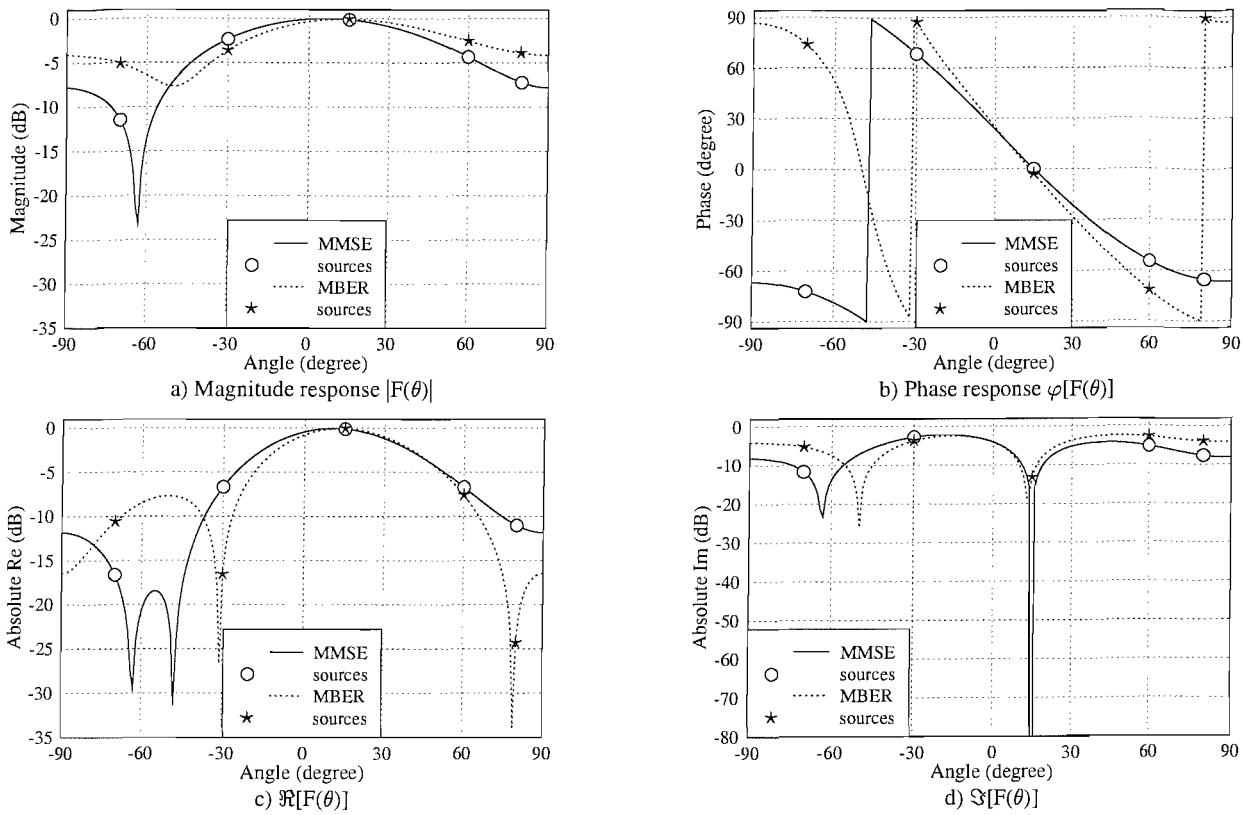
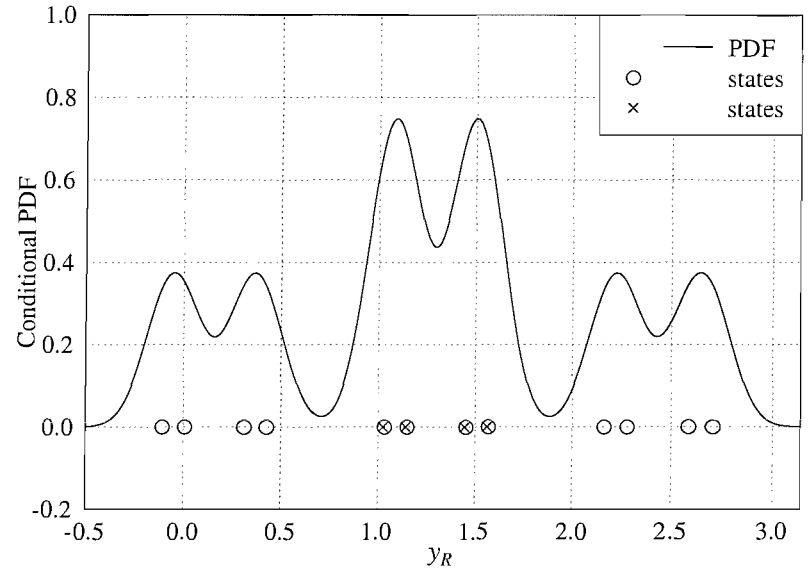


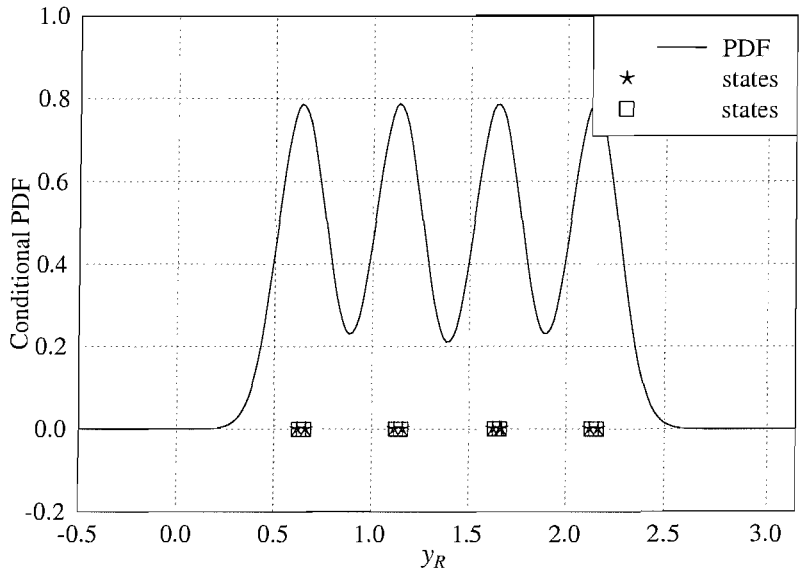
Figure 3.25: Comparison of the beam-patterns of both the MMSE and MBER beamformers for $\mathbf{SNR} = 15\mathbf{dB}$ and $\mathbf{INR}_i = \mathbf{SNR} + 6\mathbf{dB}$ for $i = 2, 3, 4, 5$, when communicating over an AWGN channel. The associated interference scenario was plotted in Figure 3.18.

opposed to the MMSE beamformer, the MBER solution characterised in Figure 3.26 b) successfully separates $\mathcal{Y}_R^{(-)}$ and $\mathcal{Y}_R^{(+)}$.

For a more clear explanation, let us return to some of the basic formulae of Equations (3.36) to (3.39) in Section 3.3. Specifically, Figure 3.27 shows the procedure of composing the noiseless beamformer output $\bar{y}_q$ of Equation (3.39), given that the desired user's transmitted bit b_1 is always $+1$. For an $M = 5$-user scenario supported by an $L = 2$-element antenna array, the total number of possible transmitted bit sequences $\mathbf{b}_q$ is given by $N_b = 2^M = 32$, i.e. $1 \leq q \leq N_b = 32$. For the condition that the desired user always transmits $b_1 = +1$, the number of possible bit sequences now reduces to $N_{sb} = 2^{M-1} = 16$. For the sake of showing all the unnormalised noiseless output states $\bar{y}$, we use different ranges on the Cartesian axes. There should be $N_{sb} = 16$ dots in each input and output plots, but some states especially those for antenna element 1 were having similar values that they overlaps each other. Similar to the noiseless output states previously shown in Figure 3.26 a), one of the noiseless output states of the MMSE beamformer was in the wrong decision region, i.e. at the negative side of the decision boundary of $y_R = 0$, corresponding to a negative noiseless output value y_R, despite the fact that the desired user's transmitted bit was always $b_1 = +1$. This implies that



a) MMSE



b) MBER

Figure 3.26: Conditional PDF of a) the MMSE and b) the MBER beamformers, given $b_1(n) = +1$ and the subset $\mathcal{Y}_R^{(+)}$ defined in Equation (3.41) of Section 3.3 for SNR = 15dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$, when communicating over an AWGN channel. The associated interference scenario was plotted in Figure 3.18.

State q	MMSE-normalised (Figure 3.26 a))	MMSE (Figure 3.27)	MBER (Figures 3.26 b) and 3.27)
1:	−0.11	−0.026	1.12
2:	1.03	0.250	1.16
3:	1.02	0.248	2.13
4:	2.16	0.525	2.16
5:	0.31	0.075	1.11
6:	1.45	0.352	1.15
7:	1.44	0.350	2.12
8:	2.58	0.627	2.15
9:	0.01	0.002	0.63
10:	1.15	0.279	0.66
11:	1.14	0.276	1.63
12:	2.28	0.554	1.67
13:	0.43	0.103	0.62
14:	1.57	0.381	0.65
15:	1.56	0.378	1.62
16:	2.70	0.655	1.65

Table 3.6: The real part of the noiseless output $\bar{y}_{R,q}$ of the beamformer of Equation (3.41) for the MMSE and MBER solution given that $b_d(n) = +1$, as inferred from Figures 3.26 a), 3.27, and Figures 3.26 b) and 3.27, respectively.

the noiseless output states $\bar{y}$ for the MMSE beamformer are linearly nonseparable, which results in the high BER floor seen in Figure 3.19 (b). Note that the corresponding noiseless MMSE and MBER beamformer outputs y_R shown in Figure 3.26 are identical to the real part of the respective noiseless outputs y shown in Figure 3.27, except that the values of the MMSE beamformer in Figure 3.26 a) were normalised approximately by a factor of 4.12, for the sake of portraying the associated conditional PDF within the same ranges as the MBER beamforming of Figure 3.26 b). The imaginary-part of the noiseless beamformer output was ignored in Figure 3.26, since in the case of BPSK transmissions we are only interested in the real-part of the output y_R.

In the context of the shape of the PDF curve it is transparent that both the shape of the MMSE and MBER PDF curves is non-Gaussian. In addition to our observation drawn from Figures 3.13 and 3.14 of Section 3.5.1, we may conclude that the MMSE beamforming solution performs best, when its PDF shape is more-Gaussian, while for the MBER its performance is significantly better than that of the MMSE, when having a non-Gaussian PDF. Therefore the performance difference between the MMSE and MBER solutions would be minimal if both have similar Gaussian-like PDF, with the latter always perform better, as shown in Figures 3.13 and 3.14.

Let us now plot the BER surface of both the MMSE and MBER solutions for this five-user scenario. Similar to the two-user case, we fixed the imaginary part of the weights, while varying

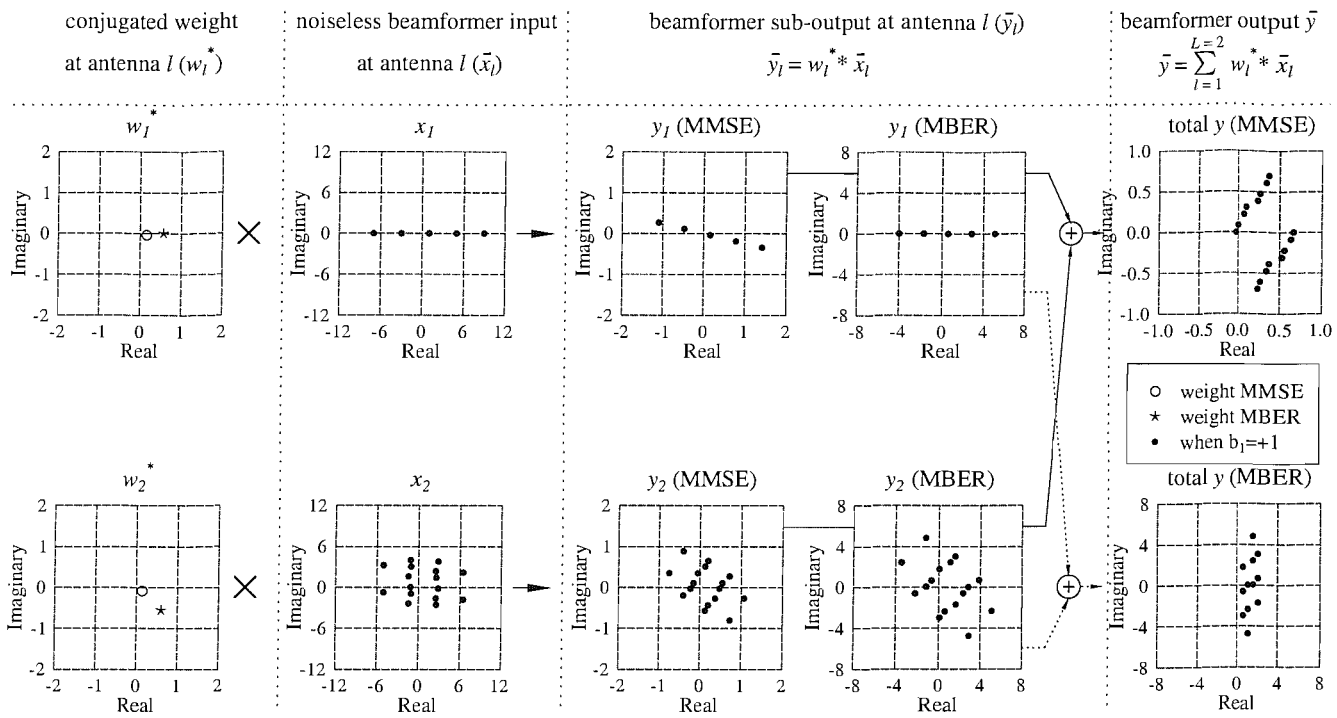


Figure 3.27: Composition of the noiseless beamformer output $\bar{y}$ defined in Equation (3.39) for both the MMSE and MBER beamformers at SNR = 15dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$. Both noiseless input $\bar{x}_l$ and sub-output $\bar{y}_l$ at the corresponding antenna l is defined in Equation (2.7) and (2.17), respectively. The associated interference scenario was plotted in Figure 3.18. Recall that a simpler two-user scenario was plotted in Figure 3.1. Whilst the MMSE solution attempts to adjust the weights for positioning the output phasors as close to ± 1 as possible, the MBER beamformer maximises their joint distance from $y_R = 0$.

the real part. Figure 3.3 shows the corresponding BER surface of the normalised MBER solution $\ddot{\mathbf{w}}_{\text{MBER}}$, where the probability of error is $\text{BER} = 10^{-6.14} \approx 7.24 \cdot 10^{-7}$. In contrast to the MBER beamformer, the error surface of the MMSE solution was shown in Figure 3.28 a) which resulted in $\text{BER} = 10^{-5.89} \approx 1.29 \cdot 10^{-6}$, corresponding to a higher BER. Several error surfaces associated with various positive values of α, which were used for multiplying the fixed imaginary parts of the MBER solution of Figure 3.3 can be observed in Figures 3.28 b) to 3.28 f). As expected, all solutions give the same minimum BER value, except for the one seen in Figure 3.28 f) in conjunction with $\alpha = 0.1$, which produces a BER of $10^{-6.03} \approx 9.33 \cdot 10^{-7}$, thus exhibits a small BER difference of $2.09 \cdot 10^{-7}$. This is likely to be a consequence of round-off errors, when multiplying a small value of $\alpha \ll 1$, such as $\alpha = 0.1$ by $\ddot{\mathbf{w}}_{\text{MBER}}$. But again, in general the value of the real and imaginary parts at the minimum corresponds to $\alpha \ddot{\mathbf{w}}_{\text{MBER}}$.

The relationship of the real parts of the MBER beamformer weights is plotted in Figure 3.29 (a) and as expected, all the optimum solutions including the one due to the MMSE solution are on a straight line. This line is, where the infinite number of MBER solutions for this specific five-user scenario resides. Hence, for any position on the line, we will attain the minimum BER of $10^{-6.14} \approx 7.24 \cdot 10^{-7}$.

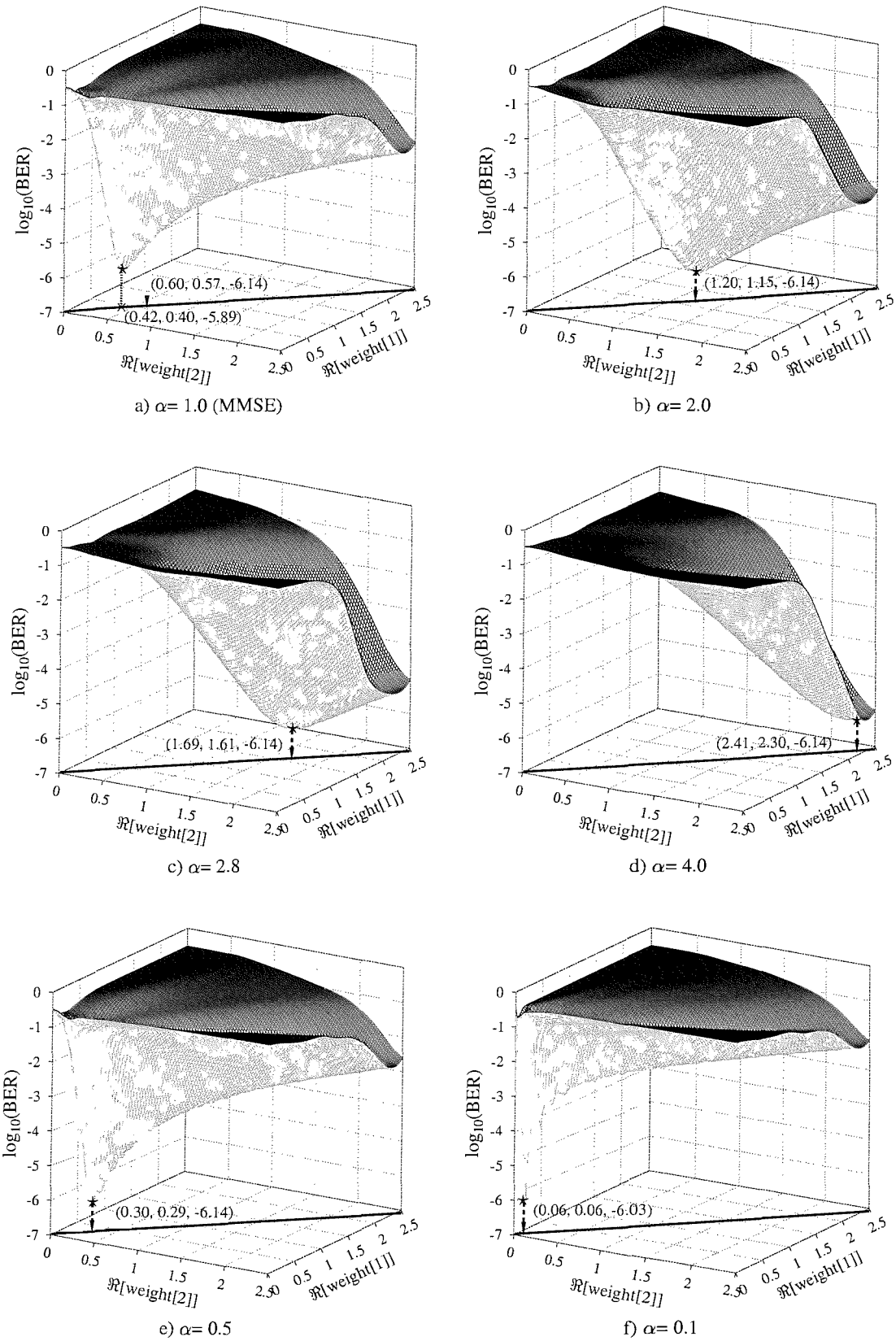
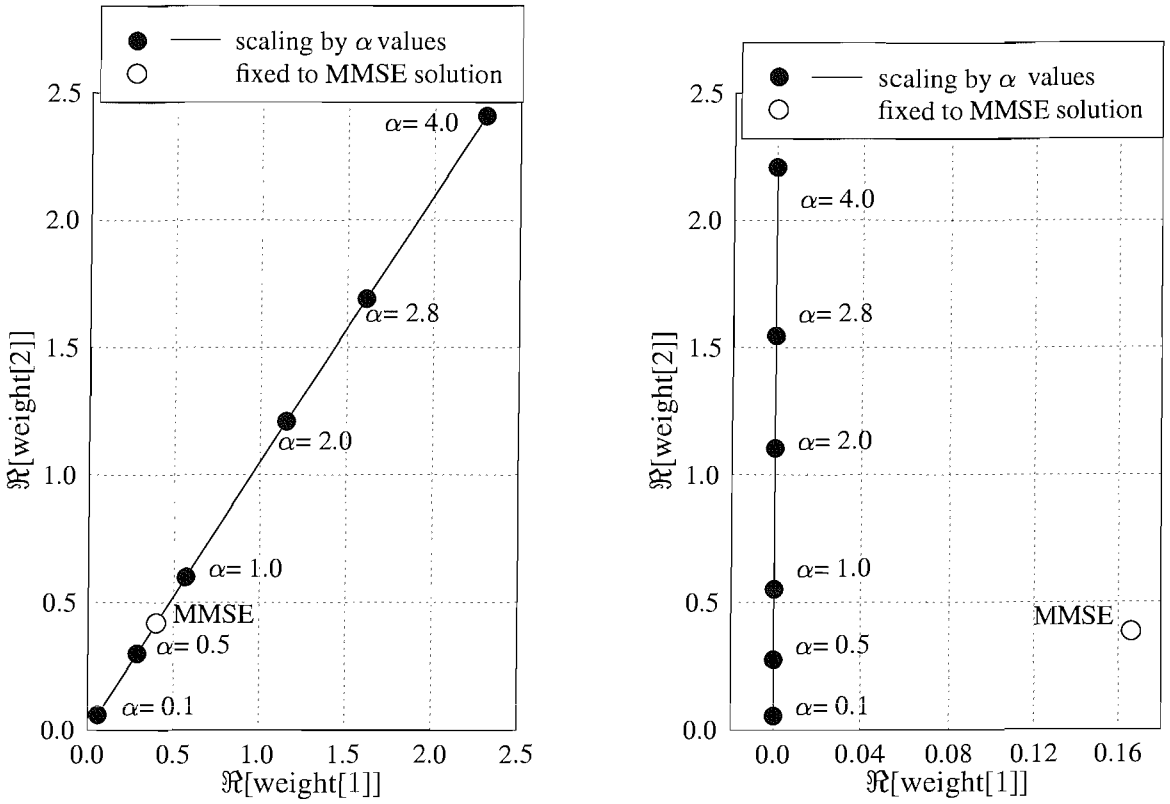


Figure 3.28: Error surfaces of the MMSE and MBER solutions plotted for different values of α; a) MMSE b) $\alpha = 2.0$ c) $\alpha = 2.8$ d) $\alpha = 4.0$ e) $\alpha = 0.5$ f) $\alpha = 0.1$ at SNR = 10dB and $\text{INR}_i = 10\text{dB}$ for $i = 2, 3, 4, 5$, when communicating over an AWGN channel. The plot was varying the array weight's real part and the corresponding fixed array weight's imaginary part associated with the α value was shown in Figure 3.29 (b). The arrow points to the MBER solution associated with the specific α values considered. The associated interference scenario was plotted in Figure 3.18. The BER was evaluated based on Equation (3.68).



(a) real-part of weight location

(b) imaginary-part of weight location

Figure 3.29: Linear relationship of (a) the two MBER array weights' real part and (b) the two MBER array weights' imaginary part for SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$, when communicating over the AWGN channel. The associated interference scenario was plotted in Figure 3.18.

We also plotted this MBER solution line in Figure 3.3. Let us next consider, why the MMSE solution also lies on the optimum line of Figure 3.29 (a), yet failing to achieve the same optimum value of BER $\approx 7.24 \cdot 10^{-7}$. The MBER weight solution extracted from our simulations was:

$$\ddot{\mathbf{w}}_{\text{MBER}} = \begin{bmatrix} 0.57 & + & j1.00608 \cdot 10^{-25} \\ 0.60 & + & j0.552012 \end{bmatrix}, \tag{3.143}$$

while the MMSE weights were:

$$\mathbf{w}_{\text{MMSE}} = \begin{bmatrix} 0.40 & + & j0.166005 \\ 0.42 & + & j0.385217 \end{bmatrix}, \tag{3.144}$$

where the real components were varied using a step-size of 0.01. It can be shown that the real part values of the MMSE solution are given by 0.7 times the MBER solution, i.e. we have $\alpha = 0.7$. However, the imaginary parts of the MMSE solution do not satisfy this condition, as seen in Figure 3.29 (b). More explicitly, we have $0.7 \cdot 0.552012 = 0.3864084$, which is not dramatically different from 0.385217. However, $0.7 \cdot 1.00608 \cdot 10^{-25} = 0.704256 \cdot 10^{-25} \approx 0$ is significantly smaller than 0.166005 and the point does not lie on the MBER solution line of Figure 3.29 (b). This is the reason, why the BER of the MMSE solution is not as low as that of the MBER solution, although the real parts lie on the MBER solution line of Figure 3.29 (a).

3.5.3 Supporting Several Users by a Four-Element Uniform Linear Array

Our previous examples provided in Sections 3.5.1 and 3.5.2 investigated user scenarios supported by a two-element uniform linear array (ULA). In this section, we will provide performance results considering a four-element ULA supporting three to nine users, which may be viewed in Figure 3.30. Explicitly, the SOI of the desired user arrives from a direction of 15^o, while the interfering signals may arrive from $-30°, 60°, -10°, 40°, 0°, 30°, -50°, 80°$ relative to the array normal.

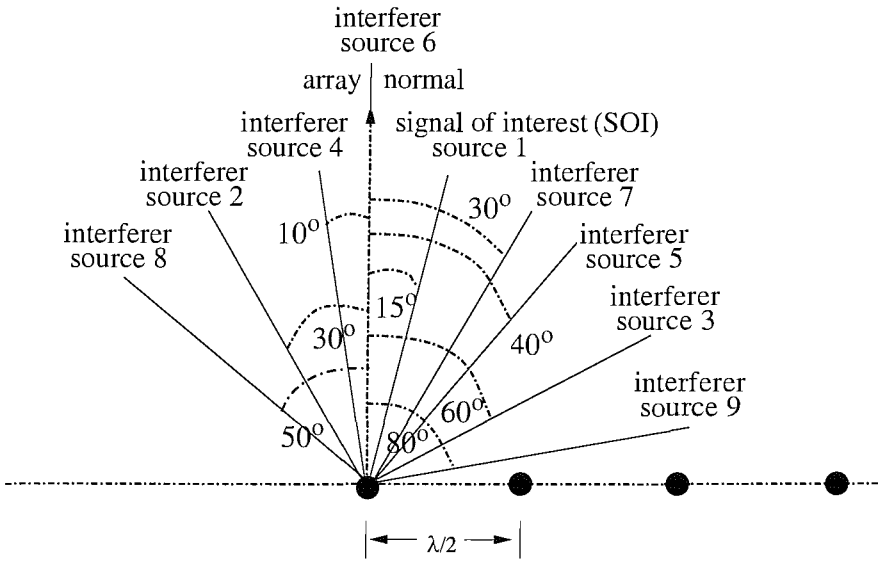
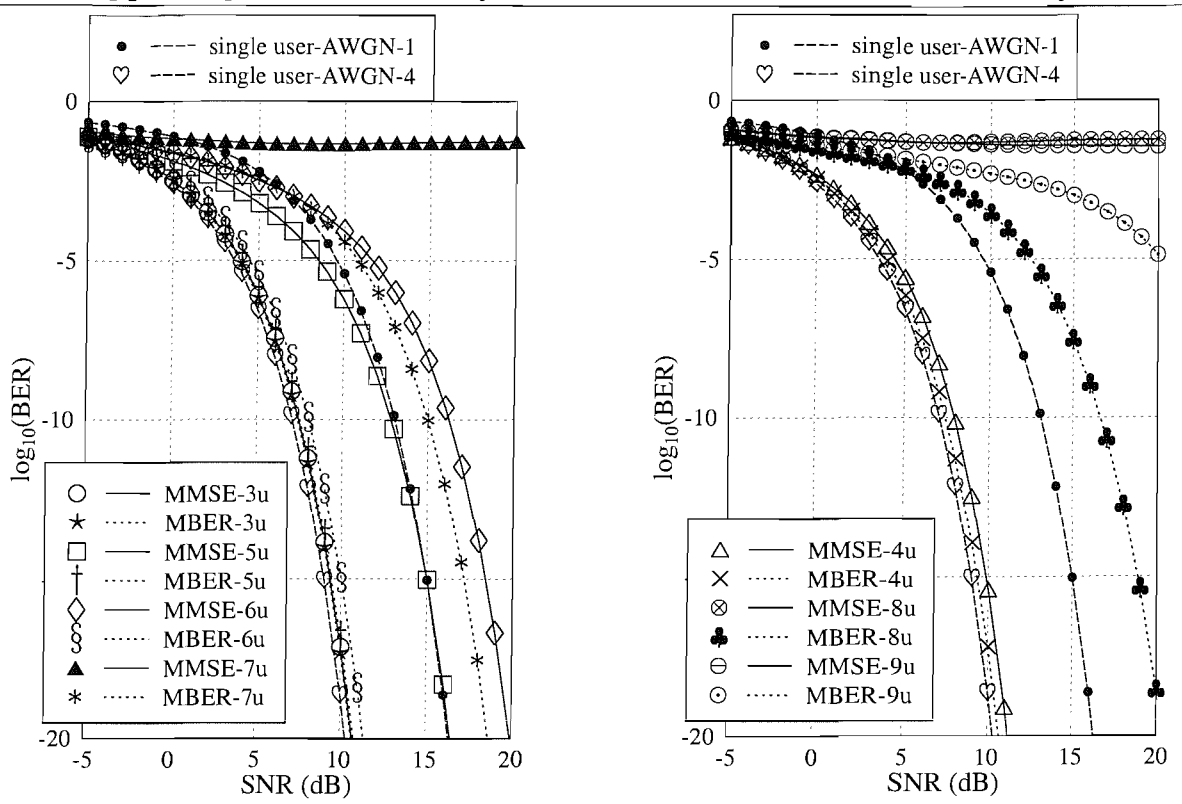


Figure 3.30: Four-element antenna array having an element spacing of $\lambda/2$ and receiving signals from **15°** (SOI), $\mathbf{-30°, 60°, -10°, 40°, 0°, 30°, -50°}$ and **80°** (the interfering signals).

Figure 3.31 shows the attainable BER performance as a function of the users' signal power, in conjunction with three to nine supported users employing a four-element uniform linear array. To avoid severe overlapping of the curves, we separate the plots into two subfigures. Figure 3.31 (a) portrays the BER curves of both the MMSE and MBER solutions, denoted by the continuous line and the dotted line, respectively, supporting three, five, six and seven users, while Figure 3.31 (b) shows the BER curves supporting four, eight and nine users. In all the seven different user-scenarios, it is observed that the MBER solution performs progressively better than the MMSE solution, as the number of user is increased. It is also observed from Figures 3.31 (a) and 3.31 (b) that when the number of users supported in the four-element system increases to seven, eight and nine, the achievable BER performance of the MMSE beamformer is severely degraded and hence it exhibits a high BER floor. Although the MBER beamformer also suffered from a significant BER performance degradation, when the system supported in excess of seven users, it remained capable of maintaining a good BER performance, even when supporting all the nine users portrayed in Figure 3.30, as evidenced by Figure 3.31 (b). Explicitly, a BER of 10^{-5} was attainable, when all the nine users had an SNR of 20dB. In addition to the BER performance results of Figure 3.19 discussed in Section 3.5.2,



(a) four-element array receiving 3, 5, 6 and 7 signals

(b) four-element array receiving 4, 8 and 9 signals

Figure 3.31: Comparison of the BER performance of the MMSE and the MBER beamformers, based on Equation (3.68), for equal-power users, i.e. for SNR = INR, in conjunction with a four-element uniform linear array in an AWGN channel. The associated interference scenario was plotted in Figure 3.30.

Figure 3.31 further confirms that MBER beamforming tolerates 'user-overloading', i.e. a scenario when the number of users is higher than the number of antenna elements, especially in comparison to the MMSE beamformer. The BER differences of the multi-user scenario with respect to the optimum performance achieved by the single-user system supported by a four-element antenna array is shown in Figure 3.32 as a function of the number of users supported for the AOAs plotted in Figure 3.30.

Figures 3.33 (a) and 3.33 (b) show the performance of both the MMSE and MBER solutions as the AOA of the interfering sources 6 and 7 of Figure 3.30 is varied from -90^o to 90^o, more specifically for $-90^o \leq \theta_6 \neq 15^o, -30^o, 60^o, -10^o, 40^o \leq 90^o$ and $-90^o \leq \theta_7 \neq 15^o, -30^o, 60^o, -10^o, 40^o, 0^o \leq 90^o$, respectively. As mentioned earlier in the context of Figure 3.21 (a), the varied interfering source 6 must not arrive from the same direction as the other five users, since this would render their separation infeasible. Therefore we drew thick vertical dashed lines in Figures 3.33 (a) and 3.33 (b), for the sake of marking the corresponding AOA of other users in the system, where no legitimate SNR value may be inferred. Similar to our findings shown in Figure 3.21 (a), the signal power required for all users in order to achieve a specific BER increases, as the interfering users 6 and 7 moved closer to the desired user located at 15^o. It is also clear from Figure 3.33 that the MBER solution always requires a lower desired user signal power than that required by the MMSE solution. More importantly, Figure 3.33 (a) shows that the interfering user 6 may get as close as 10^o or 20^o, maintaining a mere 5^o separation

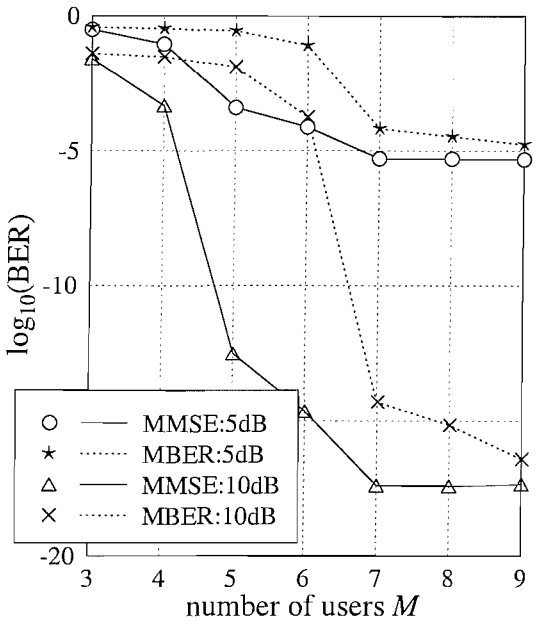
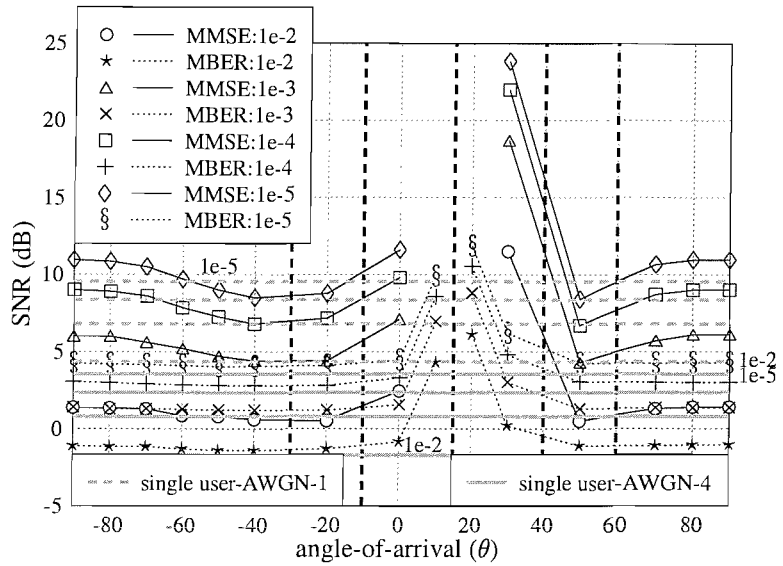


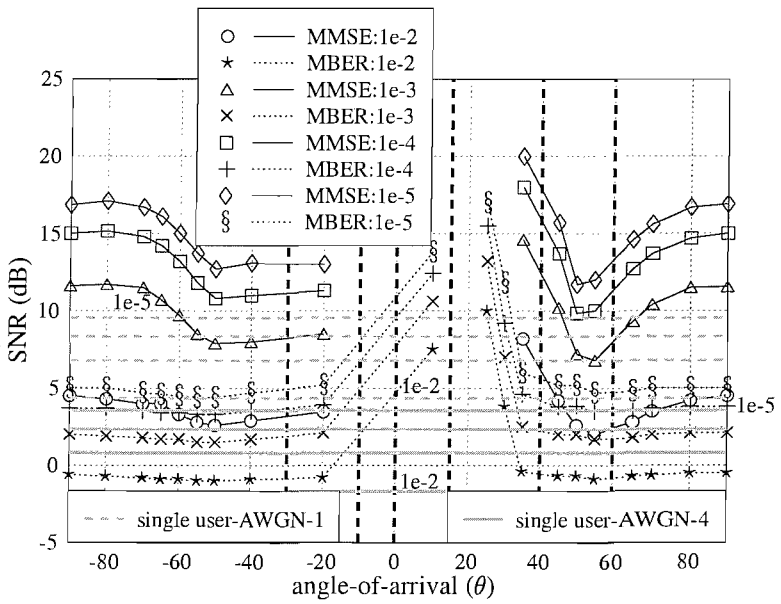
Figure 3.32: The BER difference of the MMSE and MBER beamformers from the optimum single user performance, i.e. $\log_{10}(\text{BER})_{\text{single-user,four-element}} - \log_{10}(\text{BER})_{\mathbf{w}}$, where BER was evaluated based on Equation (3.68), as a function of the number of users supported for $3 \leq M \leq 9$ for SNR of 5dB and 10dB for a four-element antenna array in an AWGN channel. The associated interference scenario was plotted in Figure 3.30.

from the SOI for the MBER solution. For the MMSE solution, the closest AOA user 6 may arrive is 0^o or 30^o, that is 15^o away from the SOI. Relative to the SNR versus AOA plot of Figure 3.21 (a) the performance of the six-user scenario of Figure 3.33 (a) is considerably better, since the minimum required angular separation between user 6 and the SOI is smaller, namely 5^o, even when all the other interfering signals are much closer to the SOI than in the scenario of Figure 3.18. Furthermore, for the MBER beamformer of Figure 3.33 (a), the desired user's signal power is always required to be significantly lower than that in Figure 3.21 (a). This consistent trend, is however, not observed for the MMSE beamformer. Figure 3.33 (b) shows that upon adding the extra interfering user at an angle of 0^o, a substantially increased SNR was required not only for the MMSE beamformer, but even for the MBER beamformer at angles of $\theta = 10^o, 25^o, 30^o$. By contrast, at higher angular separations the required SNR was near-constant for the MBER beamformer, as evidenced by Figure 3.33 (a).

Figure 3.34 shows the procedure of composing the noiseless beamformer output $\bar{y}_q$ of Equation (3.39). For an $L = 4$-element antenna array supporting $M = 7$ users, the total number of possible transmitted bit sequences is $N_{sb} = 2^{M-1} = 64$, given that the desired user's transmitted bit is always $b_1 = +1$. As explicitly seen for the beamformer's output $\bar{y}$ in Figure 3.34, some of the noiseless output phasors y of the MMSE beamformer were in the negative half-plane, i.e. on the left side of the vertical decision boundary at $y_R = 0$, while some phasors lie directly on the origin itself. Hence transmission errors would always occur, even in the absence of noise, which is the reason for



(a) AOA of interferer 6 (θ_6)



(b) AOA of interferer 7 (θ_7)

Figure 3.33: Comparison of the SNR required for maintaining a BER of $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, evaluated based on Equation (3.68), by both the MMSE and the MBER beamformers for (a) six equal-power users, i.e. for SNR = INR, as a function of the AOA of interferer 6 and (b) seven equal-power users, as a function of the AOA of interferer 7. The thick vertical lines in the subfigures (a) and (b) mark the corresponding AOA of the remaining users. The associated interference scenario was plotted in Figure 3.30. An AWGN channel was used.

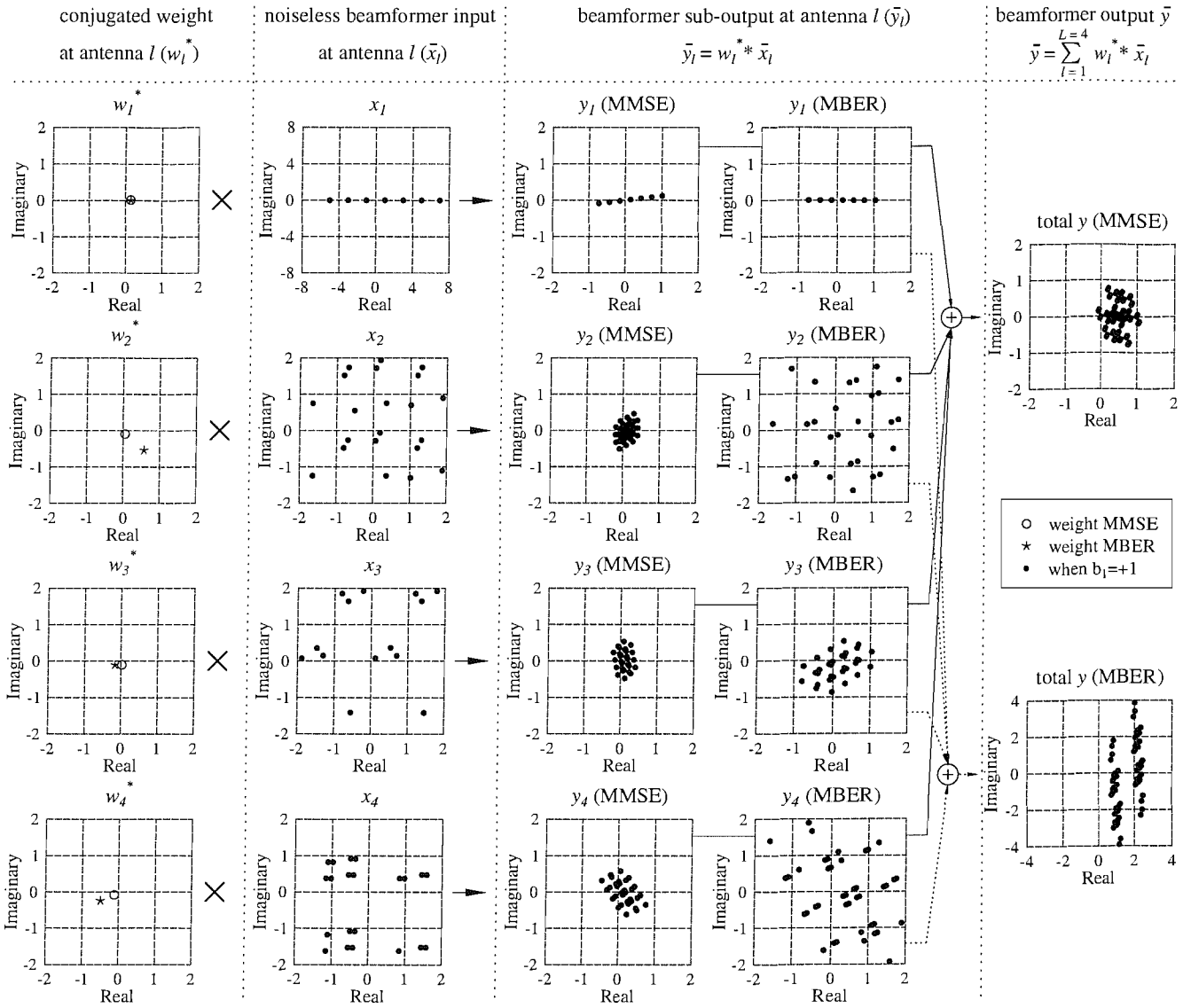


Figure 3.34: Composition of the noiseless beamformer output $\bar{y}$ defined in Equation (3.39) for both the MMSE and MBER beamformers supporting seven equal-power users at SNR = INR_i = 10dB for $i = 2, 3, ..., 7$. Both the noiseless input $\bar{x}_l$ and the sub-output $\bar{y}_l$ recorded at the corresponding antenna l were defined in Equations (2.7) and (2.17), respectively. The associated interference scenario was plotted in Figure 3.30. Recall that a simpler two-user scenario was portrayed in Figure 3.1. Whilst the MMSE solution attempts to adjust the weights for the sake of positioning the output phasors as close to ± 1 as possible, the MBER beamformer maximises their joint distance from $y_R = 0$.

the irreducible BER observed in Figure 3.31 (a). Since not all of the noiseless total output phasors y of the MBER beamformer are associated with real and/or imaginary values confined within the interval $|2|$, we extend the plotted range of the Cartesian axes to $|4|$ for the sake of showing all the 64 noiseless output phasors y. However, it is also possible to have several noiseless outputs y associated with similar values, which result in overlapping points in Figure 3.34. The noiseless output y plotted in Figure 3.34 also shows the difference between the optimisation philosophy of the MMSE and MBER beamformers. Explicitly, Figure 3.34 demonstrates that the MBER beamformer constructs the system more intelligently, in that it arranges the real part of the beamformer output y_R to be as far as possible from the decision boundary at $y_R = 0$ for the sake of minimising the BER.

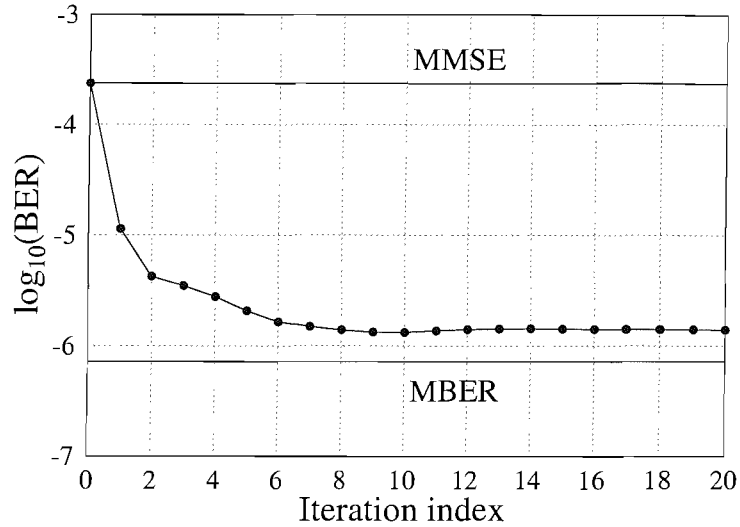
3.6 Simulation Results for Adaptive MBER Beamforming

Let us now consider the performance of the adaptive MBER beamforming solution. First, we focus our attention on the block-data based gradient adaptive MBER algorithm of Section 3.4.2, which we refer to as the Block Adaptive Conjugate Gradient (BACG) algorithm. We will then investigate the attainable performance of two different sample-by-sample based adaptive MBER algorithms, namely that of the Least BER (LBER) and the Approximate LBER (ALBER) algorithms of Section 3.4.3.1 and 3.4.3.2, respectively.

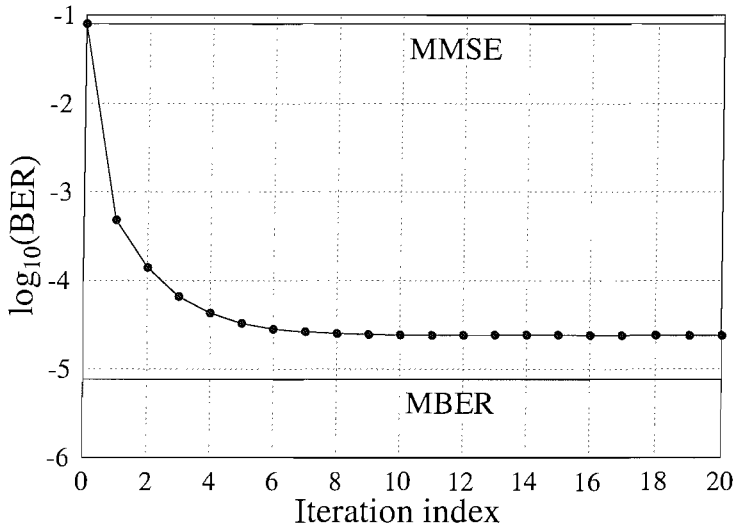
3.6.1 Block Adaptive Conjugate Gradient Algorithm

Figure 3.35 illustrates the rapid convergence of the BACG algorithm employing a block of $T = 200$ training samples, under two different conditions, namely when (a) all users have an equal SNR of 10dB and (b) when all interfering users benefit from a 6dB higher SNR than the desired user having an SNR of 13dB. More explicitly, Figure 3.35 demonstrates that six to eight iterations are sufficient for the algorithm's convergence to a BER near the MBER solution. The initial array weights have been set to the MMSE solution of $\mathbf{w}_{\text{MMSE}}$ in both scenarios of Figure 3.35, but our experiments suggest that this is not necessarily the best option for weight-initialisation. More specifically, as observed in Figures 3.36 (a) and 3.36 (b), we may attain more rapid convergence and possibly a better steady state BER performance than that seen in Figure 3.35 (a), when the array weights of the algorithms were initialised to W1 $= [0.2+j0.05 \quad 0.05+j0.2]^T$ and W2 $= [0.2+j0.05 \quad 0.15+j0.15]^T$, respectively, i.e. when the weights are different from the MMSE solution of $\mathbf{w}_{\text{MMSE}} = [0.31+j0.08 \quad 0.27+j0.18]^T$ determined for this specific scenario. It was also shown in [199] that the algorithm may become trapped in a local minimum, when it is initialised to the MMSE solution. The issue of initialisation hence has to be given special attention, but for simplicity's sake we will initialise the beamformer weights to the MMSE solution, $\mathbf{w}_{\text{MMSE}}$, in all our simulations.

Figures 3.37 (a) and 3.37 (b) show the effects of using different training block sizes on the achievable

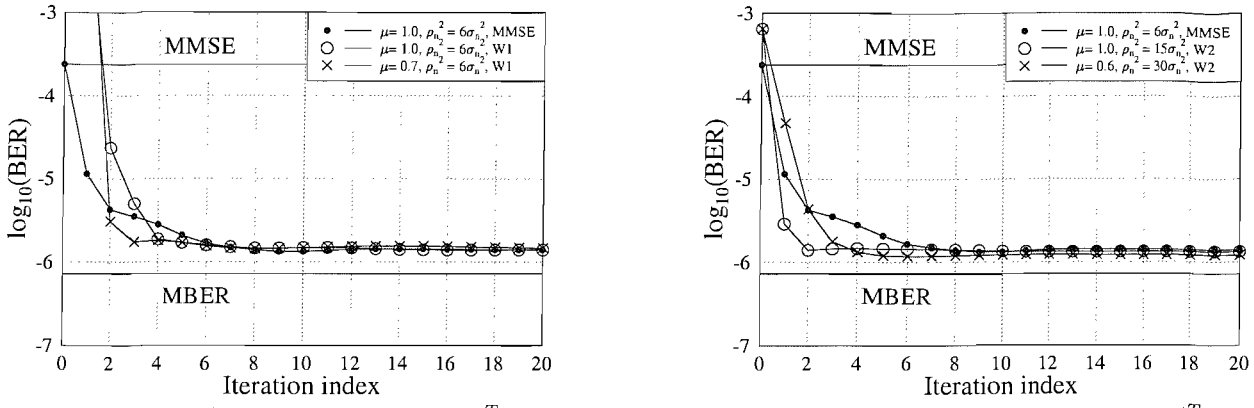


(a) SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$, $\mu = 1.0$, $\rho_n^2 = 6\sigma_n^2 = 0.3$



(b) SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$, $\mu = 0.5$, $\rho_n^2 = 6\sigma_n^2 = 0.15$

Figure 3.35: Convergence behaviour of the block-data based BACG MBER algorithm, based on Equation (3.112), for a block size of $T = 200$ samples under two different conditions; (a) SNR = 10dB and INR_i = 10dB for users $i = 2, 3, 4, 5$, when employing a step size of $\mu = 1.0$ and a kernel width ρ_n of $\rho_n^2 = 6\sigma_n^2 = 0.3$, where $\sigma_n^2 = 0.05$ and (b) SNR = 13dB and INR_i = SNR +6dB for users $i = 2, 3, 4, 5$, involving a step size of $\mu = 0.5$ and a kernel width ρ_n of $\rho_n^2 = 6\sigma_n^2 = 0.15$, where $\sigma_n^2 = 0.025$. The related AOA scenarios were depicted in Figure 3.18 and BPSK transmissions were used over an AWGN channel.

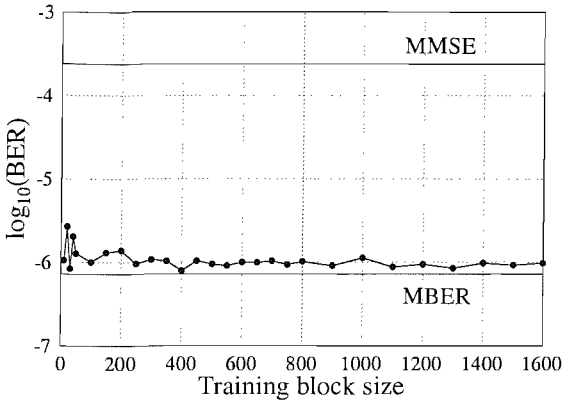


(a) W1 = $|0.2 + j0.05 \quad 0.05 + j0.2|^T$ (b) W2 = $|0.2 + j0.05 \quad 0.15 + j0.15|^T$

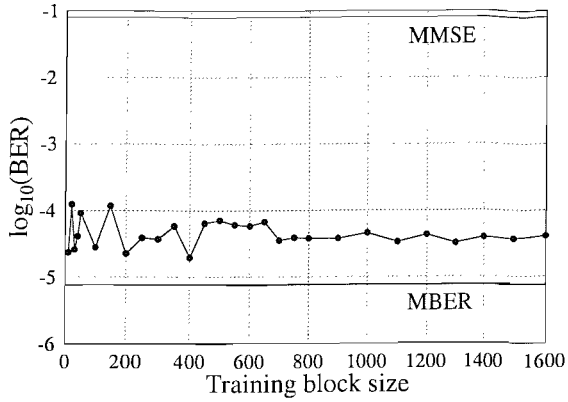
Figure 3.36: Convergence behaviour of the block-data based BACG MBER algorithm, based on Equation (3.112), in conjunction with the SNR = 10dB equal-power user scenario, employing a training block size of $T = 200$ samples under two different non-MMSE weight initialisation conditions; (a) weight vector initialised to W1 = $[0.2 + j0.05 \quad 0.05 + j0.2]^T$ for step sizes of $\mu = 1.0$ and $\mu = 0.7$ and a kernel width ρ_n of $\rho_n^2 = 6\sigma_n^2 = 0.3$ and (b) weight initialised to W2 = $[0.2 + j0.05 \quad 0.15 + j0.15]^T$ for step sizes of $\mu = 1.0$ and $\mu = 0.6$ and a kernel width ρ_n of $\rho_n^2 = 15\sigma_n^2 = 0.75$ and $\rho_n^2 = 30\sigma_n^2 = 1.5$, respectively. The corresponding AOA scenarios were depicted in Figure 3.18 and BPSK transmissions were used over an AWGN channel.

convergence rate of the BACG algorithm, when the adaptive parameters μ and ρ_n are fixed, employing values similar to those of Figures 3.35 (a) and 3.35 (b), respectively. Specifically, the step size was set to $\mu = 1.0$ and the kernel width to $\rho_n = \sqrt{6}\sigma_n = 0.5477$ in Figure 3.37 (a), when considering equal-power users having an SNR of 10dB, while Figure 3.37 (b) employs $\mu = 0.5$ and $\rho_n = \sqrt{6}\sigma_n = 0.3878$ for the unequal-power user scenario of SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$. It was observed in both scenarios that the BACG algorithm is not overly sensitive to the specific training block size used. The steady state BER performance seen from both plots of Figure 3.37 are BER = $10^{-6.04} \approx 9.12 \cdot 10^{-7}$ and BER = $10^{-4.42} \approx 3.8 \cdot 10^{-5}$, respectively. The fluctuation of the BER values observed for the equal-power user scenario of Figure 3.37 (a) is not as significant as that observed in Figure 3.37 (b). It is readily seen from Figure 3.37 (a) that for a training block size higher than 50 bits, the BER values' average deviation from the mean is approximately BER $\approx \pm 0.3 \cdot 10^{-6}$, while in Figure 3.37 (b), the BER values' average deviation from the mean for training block sizes in excess of 200 bits is approximately BER $\approx 2.37 \cdot 10^{-5}$. The maximum deviation from the respective steady state BER value for both scenario is approximately BER $\approx 1.76 \cdot 10^{-6}$ and BER $\approx 8.57 \cdot 10^{-5}$, respectively. Furthermore, the BACG algorithm operating under the equal-power user scenario shows a near-constant BER performance for training block sizes higher than 400 bits, while for the unequal-power user scenario of Figure 3.37 (b) this trend prevails, when the training block size employed is in excess of 700 bits. Using a short training block length of 100 bits may also achieve a similar steady state BER value to that of the longer block size scenarios. As the specific choice of training block size is not critical, we will consider a block size of $T = 256$ in our forthcoming BACG algorithmic simulations.

Figures 3.36 (a) and 3.36 (b) also show that the BER performance of the BACG algorithm is



(a) SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$, $\mu = 1.0$, $\rho_n^2 = 6\sigma_n^2 = 0.3$

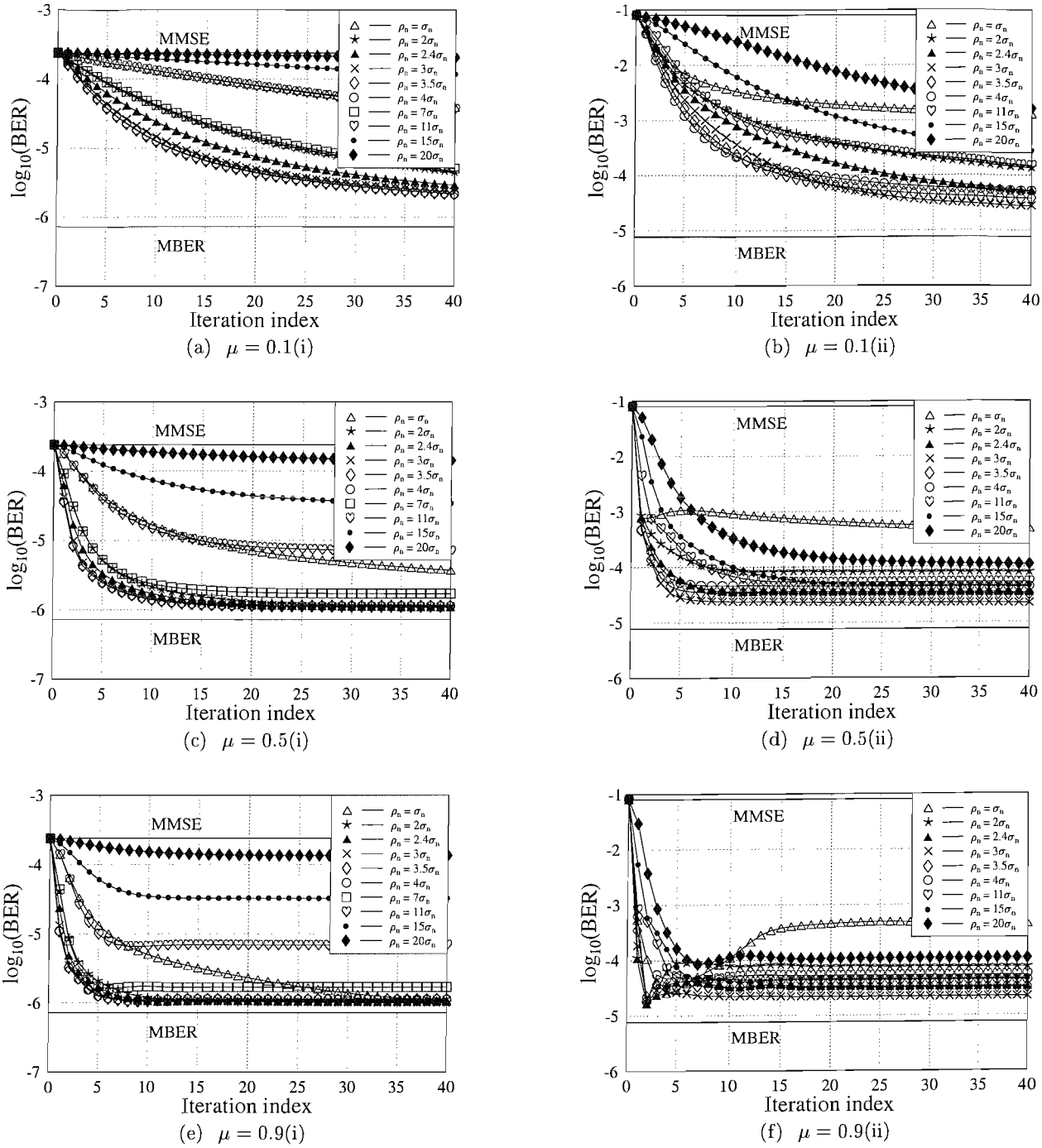


(b) SNR = 13dB and INR_i = SNR +6dB for users $i = 2, 3, 4, 5$, $\mu = 0.5$, $\rho_n^2 = 6\sigma_n^2 = 0.15$

Figure 3.37: Effect of training block size on the achievable performance of the block-data based BACG MBER algorithm, based on Equation (3.112), under two different conditions; (a) SNR = 10dB and INR_i = 10dB for users $i = 2, 3, 4, 5$, employing a step size of $\mu = 1.0$ and a kernel width ρ_n of $\rho_n^2 = 6\sigma_n^2 = 0.3$ and (b) SNR = 13dB and INR_i = SNR +6dB for users $i = 2, 3, 4, 5$, invoking a step size of $\mu = 0.5$ and a kernel width ρ_n of $\rho_n^2 = 6\sigma_n^2 = 0.15$. The related AOA scenarios were depicted in Figure 3.18 and BPSK transmissions were used over an AWGN channel.

affected, when the step size μ and the kernel width ρ_n, obeying $\rho_n^2 = \rho_r^2\sigma_n^2$ is varied. It has been stated in Section 3.4.2 that the appropriate values of the step size μ and kernel width ρ_n should be used for the sake of arriving at the optimum performance of the adaptive algorithm. Therefore it is beneficial to observe the effect of varying these parameters on the performance of the BACG algorithm. Figure 3.38 shows the effect of varying the kernel width values ρ_n in conjunction with several step sizes, namely $\mu = 0.1, 0.5$ and 0.9 in a scenario experiencing an identical SNR of 10dB and also for SNR = 13dB, INR_i = SNR +6dB for $i = 2, 3, 4, 5$. For a particular step size, it is observed that as ρ_n is increased to $\rho_n = 3\sigma_n$, the achievable convergence rate also increases. Its further increase to the values of $\rho_n > 7\sigma_n$ and $\rho_n > 3\sigma_n$ selected for the equal-power and unequal-power user scenarios, respectively, will result in a reduced convergence rate for the BACG algorithm as evidenced by Figure 3.38. For the equal-power user scenario of Figure 3.38 (a), 3.38 (c) and 3.38 (e), it is observed that a simultaneous improvement of both the convergence rate and the steady state BER performance may only be achieved for $\sigma_n < \rho_n \leq 4\sigma_n$. Finally, as suggested by Figures 3.38 (c) and 3.38 (e), when the kernel width ρ_n increases, such that we have $4\sigma_n < \rho_n \leq 7\sigma_n$, the convergence rate maintains or continues to improve, although this is achieved at the expense of a higher steady state BER.

For the sake of a more explicit visualisation, in Figure 3.39 we also characterise the effect of the step size μ for several kernel width values ρ_n. It is observed from Figure 3.39 (b) in conjunction with the kernel width of $\rho_n = 2\sigma_n$ invoked for the unequal-power user scenario that more iterations may be required for achieving the steady state BER, when employing a large step size of $\mu > 0.7$. This phenomenon may be avoided, provided that a suitable kernel width is used and a higher number of iterations is affordable in complexity terms. Based on the plots of both Figures 3.38 and 3.39, we may conclude that the convergence rate and steady-state BER performance of the BACG algorithm

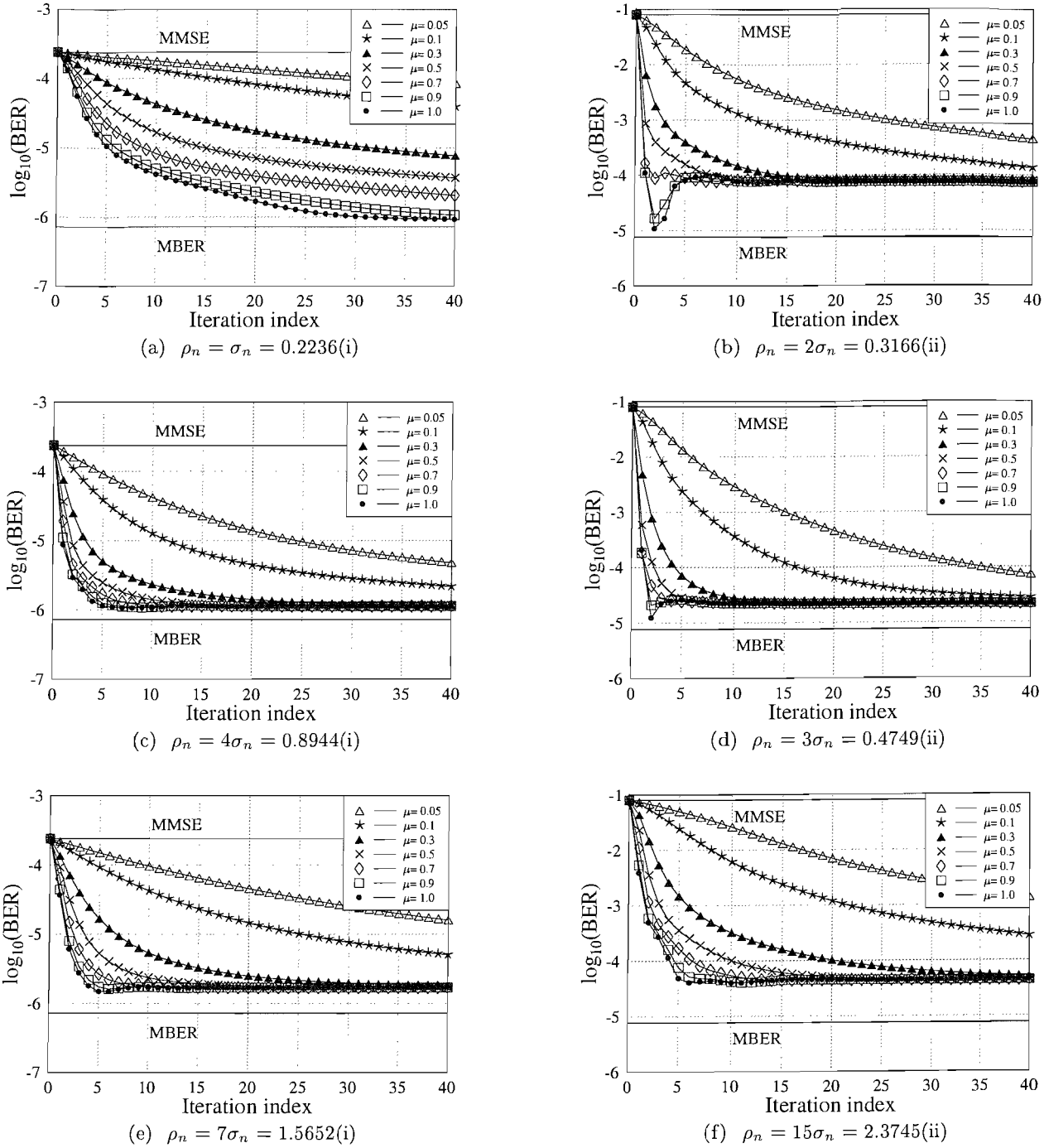


(a) $\mu = 0.1$(i) (b) $\mu = 0.1$(ii)

(c) $\mu = 0.5$(i) (d) $\mu = 0.5$(ii)

(e) $\mu = 0.9$(i) (f) $\mu = 0.9$(ii)

Figure 3.38: Convergence behaviour of the BACG algorithm, based on Equation (3.112), for kernel width values of $\rho_n = \sigma_n, 2\sigma_n, 2.4\sigma_n, 3\sigma_n, 3.5\sigma_n, 4\sigma_n, 11\sigma_n, 15\sigma_n, 20\sigma_n$ and for the step sizes of $\mu = 0.1, 0.5, 0.9$ in conjunction with a training block size of $T = 256$ under two different conditions; (i) SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$ and $\sigma_n = \sqrt{0.05} = 0.2236$ in subfigures (a), (c) and (e), and (ii) SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$ and $\sigma_n = \sqrt{0.025} = 0.1583$ in subfigures (b), (d) and (f). The corresponding AOA scenario was depicted in Figure 3.18 and an AWGN channel was used.



(a) $\rho_n = \sigma_n = 0.2236$(i)

(b) $\rho_n = 2\sigma_n = 0.3166$(ii)

(c) $\rho_n = 4\sigma_n = 0.8944$(i)

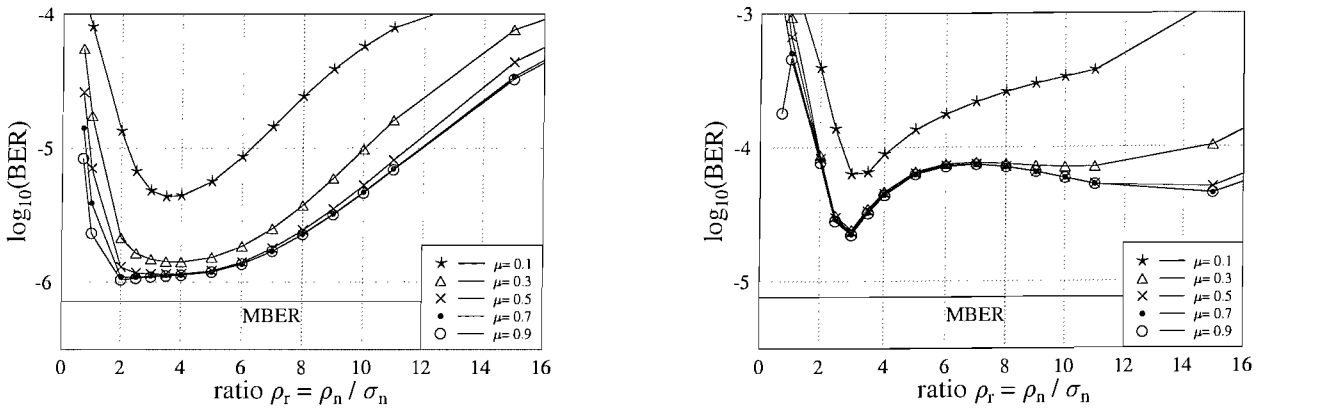
(d) $\rho_n = 3\sigma_n = 0.4749$(ii)

(e) $\rho_n = 7\sigma_n = 1.5652$(i)

(f) $\rho_n = 15\sigma_n = 2.3745$(ii)

Figure 3.39: Convergence behaviour of the BACG algorithm, based on Equation (3.112), for step sizes of $\mu = 0.05, 0.1, 0.3, 0.5, 0.9, 1.0$ and for the kernel width values of (a) $\rho_n = \sigma_n$, (c) $\rho_n = 4\sigma_n$, (e) $\rho_n = 7\sigma_n$ under the condition (i) SNR = 10dB and $\text{INR}_i = 10$dB for $i = 2, 3, 4, 5$ and $\sigma_n = \sqrt{0.05} = 0.2236$, and kernel width values of (b) $\rho_n = 2\sigma_n$, (d) $\rho_n = 3\sigma_n$, (f) $\rho_n = 15\sigma_n$ under the condition (ii) SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$ and $\sigma_n = \sqrt{0.025} = 0.1583$, in conjunction with a training block size of $T = 256$. The corresponding AOA scenario was depicted in Figure 3.18 and an AWGN channel was used.

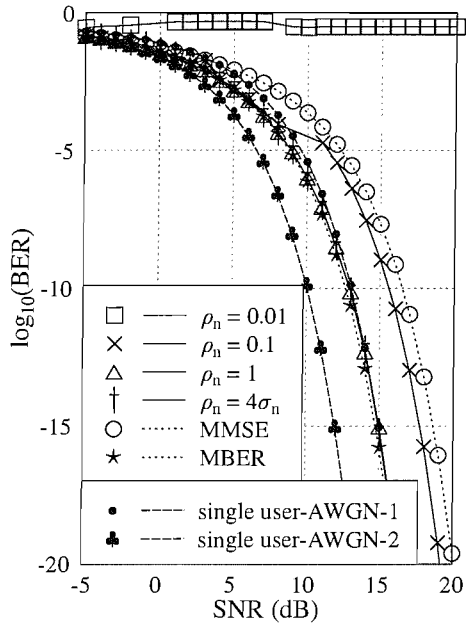
is affected by both the step size μ and the kernel width ρ_n. An appropriate balance of these two parameters has to be maintained for the sake of achieving the lowest steady state BER, while attaining a rapid convergence. Using a small step size will reduce the rate of convergence and although this may be partially compensated for by having a better steady state BER value, it will be at the cost of a high complexity, expressed in terms of the number of operations invoked during the iterations. It is therefore advisable to use a sufficiently high step size of $\mu \geq 0.5$ in conjunction with a moderate value of the kernel width, which is within the range of $2\sigma_n < \rho_n < 7\sigma_n$ and $2.5\sigma_n \leq \rho_n \leq 3.5\sigma_n$ for an equal-power and unequal-power user scenario, respectively. This however is solely based on the observation of Figure 3.40, where we plotted the achievable steady state BER of the BACG algorithm versus the kernel width ratio of $\rho_r = \rho_n/\sigma_n$ at iteration index 20, as the kernel width ρ_n was varied, along with several step size values of $\mu = 0.1, 0.3, 0.5, 0.7, 0.9$. Figure 3.40 also suggests that the specific choice of the kernel width ρ_n is more crucial than that of the step size μ. For a particular scenario, it is beneficial to know the range of the acceptable kernel width values for the sake of attaining a rapid convergence and a near-optimum BER performance.



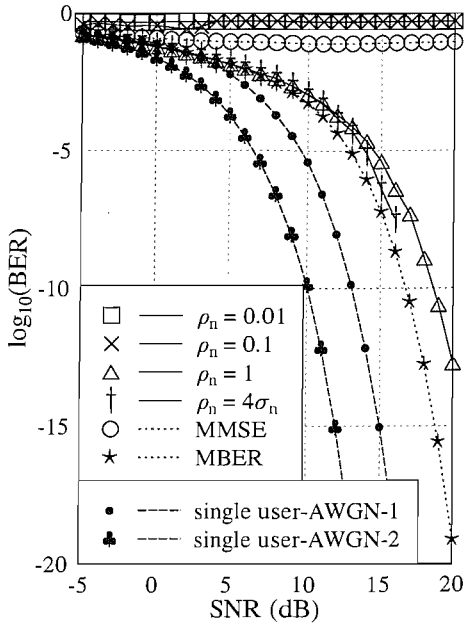
(a) SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$ (b) SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$

Figure 3.40: The achievable steady state BER performance of the BACG algorithm, based on Equation (3.112), at iteration index 20, when the kernel width ratio of $\rho_r = \rho_n/\sigma_n$ is varied for step sizes of $\mu = 0.1, 0.3, 0.5, 0.7, 0.9$ using a training block size of $T = 256$ in conjunction with two different user power scenarios of (a) SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$ and (b) SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$. The corresponding AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.

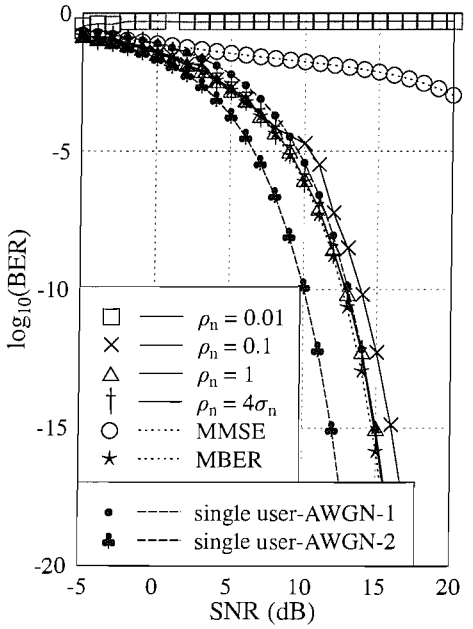
As a comparison to the performance of the exact MBER for a system employing the two-element antenna array portrayed in Figure 3.19, we plotted the BACG BER performance in Figures 3.41 and 3.42 under the same conditions; (a) the desired user and all the four interfering sources have an equal power; (b) all the interfering sources have 6dB higher power than the desired user; (c) the desired user and the interfering sources $3, 4, 5$ have an equal power, while the interfering source 2 has 6dB higher power than the desired user; and (d) the desired user and the interfering sources $2, 3, 4$ have an equal power, but the interfering source 5 has 6dB higher power than the desired user. In



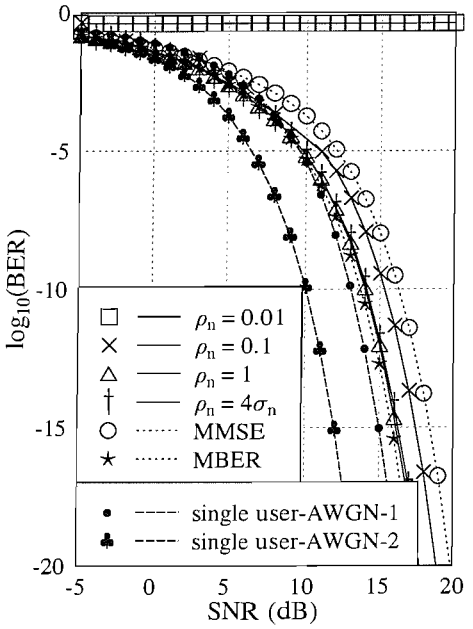
(a) SNR = INR_i for i = 2, 3, 4, 5



(b) INR_i = SNR +6dB for i = 2, 3, 4, 5



(c) SNR = INR_i for i = 3, 4, 5 and INR_2 = SNR +6dB



(d) SNR = INR_i for i = 2, 3, 4 and INR_5 = SNR +6dB

Figure 3.41: BER performance of the BACG algorithm, based on Equation (3.112), under four different conditions; (a) SNR = INR_i for $i = 2, 3, 4, 5$, (b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$, (c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB for two-element uniform linear array. The BACG algorithm was using training samples of length $T = 256$, kernel width values of $\rho_n = 0.01, 0.1, 1, 4\sigma_n$ and a step size of $\mu = 0.9$. These results were recorded for transmission over the AWGN channel for five users and the corresponding AOA scenario was plotted in Figure 3.18.

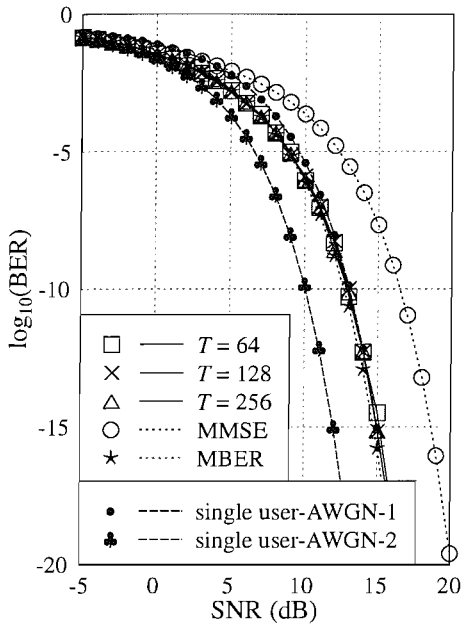
Figure 3.41 we used a step size of $\mu = 0.9$ and kernel width of $\rho_n = 0.01, 0.1, 1, 4\sigma_n$. It is observed in Figure 3.41 that the BACG algorithm is unable to tolerate an inappropriate choice of the kernel width value, particularly a value of $\rho_n = 0.01$. By contrast, a near-optimum BER performance may be observed in Figure 3.41, when using the kernel width values of $\rho_n = 1$ and $\rho_n = 4\sigma_n$. However, for a particular scenario of having all the interfering sources' power at a level of 6dB higher than the desired user, a significant BER performance degradation has been observed in Figure 3.41 (b). It is seen in Figure 3.41 (b) that the BACG algorithm becomes deficient, when using small constant kernel width values of $\rho_n = 0.01, 0.1$. For a variable kernel width value of $\rho_n = 4\sigma_n$, i.e. for $5.0297 < \rho_n < 0.2828$ and for SNR values spanning the range of $-5 < \text{SNR(dB)} < 20$, the BER performance of the BACG significantly deviates from the exact MBER performance, as the users signal power is increased. Since the kernel width value plays an important role in attaining a BER performance that is near to the exact MBER solution, we opt for employing the optimal kernel width radius suggested by Silverman's rule of thumb, which was given in Equation (3.102). The corresponding BER performance is shown in Figure 3.42 in conjunction with training block of lengths of $T = 64, 128, 256$. The BER performance shown in Figure 3.42 is in good agreement with the exact MBER solution. It is observed that the BER performance of the BACG algorithm is nearly identical to that of the exact MBER solution for all scenarios considered, except for the particular case of Figure 3.42 (b). It is also seen in Figure 3.42 that the BER performance improves, as a longer training block is used, especially for SNRs higher than 10dB.

Having characterised the achievable BER performance of the block-data based BACG algorithm, let us next study the attainable performance of the stochastic gradient based adaptive MBER algorithms of Section 3.4.3, namely that of the Least Bit Error Rate (LBER) and Approximate (ALBER) algorithms.

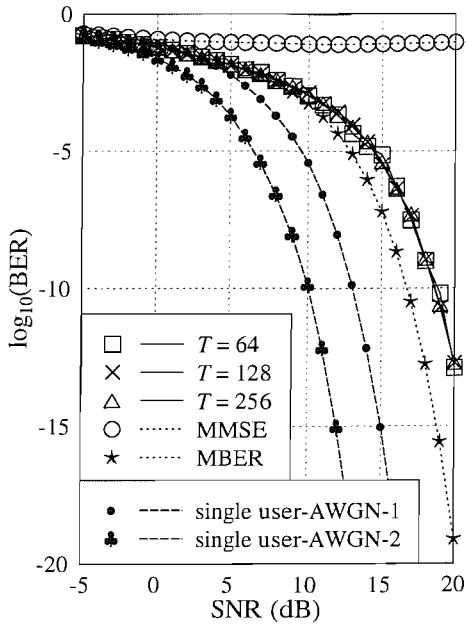
3.6.2 Stochastic Gradient Based Adaptive MBER Algorithms

Conditioned on whether the kernel width ρ_n chosen for the kernel density estimation process is dependent or independent of the magnitude of the beamformer's weight vector, we may define two different types of sample-by-sample adaptive MBER algorithms, which we referred to as the LBER algorithm and ALBER algorithm, respectively. Figures 3.43 and 3.44 present the associated learning curves of the sample-by-sample adaptive algorithms of Sections 3.4.3.1 and 3.4.3.2, i.e. those of the LBER algorithm and the ALBER algorithm, respectively, under two different conditions. In the first scenario all users experience an equal SNR of 10dB, while in the second the SNR of all the interfering users is increased by another 6dB with respect to the previously employed SNR = 13dB of the desired user. As previously, the initial array weight vector was set to the MMSE solution of $\mathbf{w}_{\text{MMSE}}$.

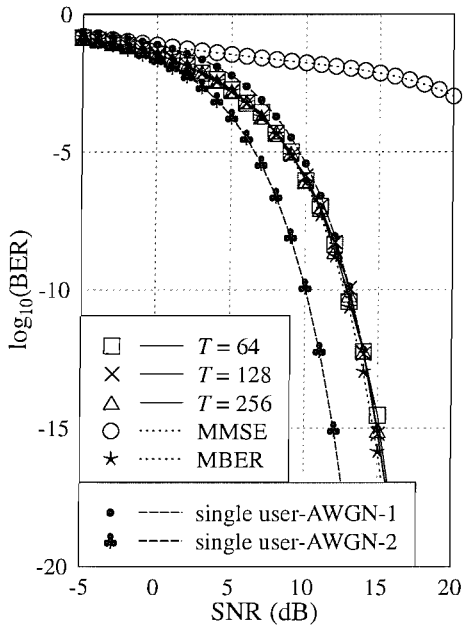
For the equal-SNR scenario characterised in Figure 3.43 (a), it is observed that the ALBER solution has a better convergence rate and settles at a lower steady-state BER value than the LBER



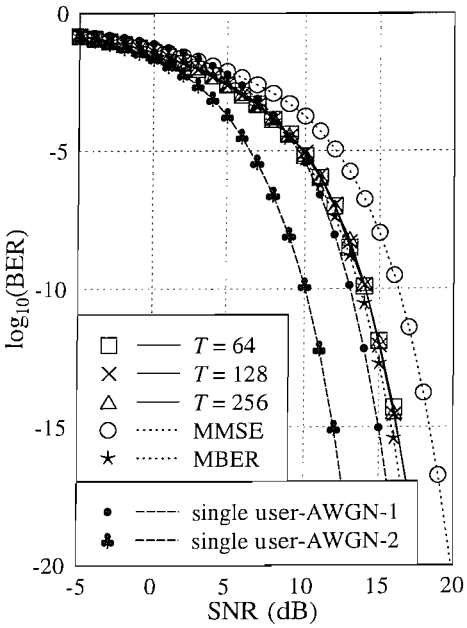
(a) SNR = INR_i for i = 2, 3, 4, 5



(b) INR_i = SNR +6dB for i = 2, 3, 4, 5

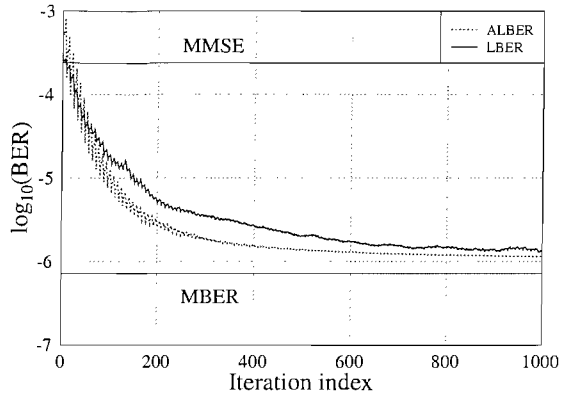


(c) SNR = INR_i for i = 3, 4, 5 and INR_2 = SNR +6dB

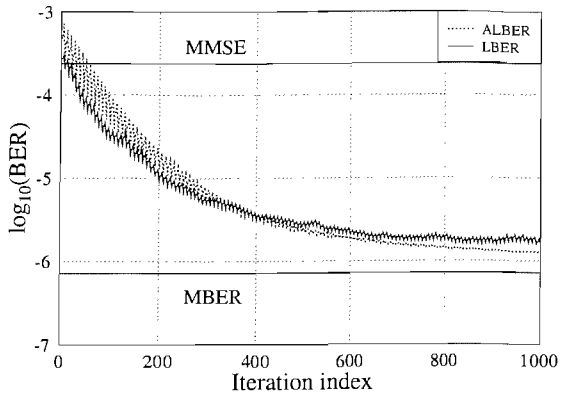


(d) SNR = INR_i for i = 2, 3, 4 and INR_5 = SNR +6dB

Figure 3.42: BER performance of the BACG algorithm, based on Equation (3.112), under four different conditions; (a) SNR = INR_i for $i = 2,3,4,5$, (b) INR_i = SNR +6dB for $i = 2,3,4,5$, (c) SNR = INR_i for $i = 3,4,5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2,3,4$ and INR_5 = SNR +6dB for two-element uniform linear array. The BACG algorithm was using training samples of length $T = 64, 128, 256$, optimum kernel width radius ρ_n according to Silverman's rule of thumb of Equation (3.102) and a step size of $\mu = 0.9$. These results were recorded for transmission over the AWGN channel for five users and the corresponding AOA scenario was plotted in Figure 3.18.

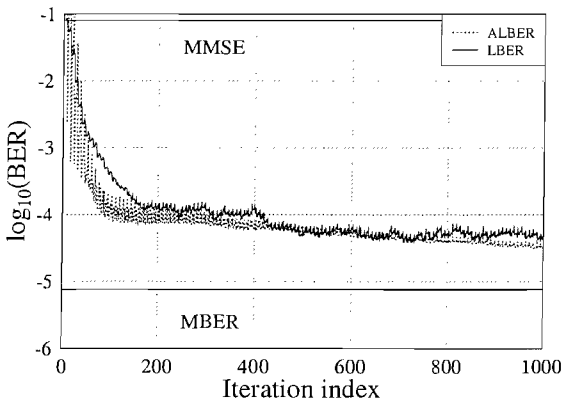


(a) $\mu = 0.01, \rho_n = 4\sigma_n = 0.8944$

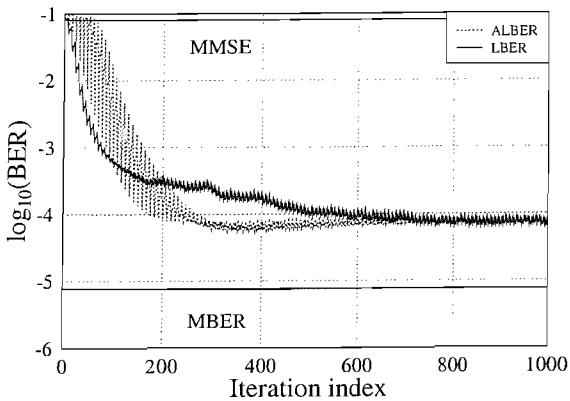


(b) $\mu = 0.01, \rho_n = 6\sigma_n = 1.3416$

Figure 3.43: Convergence behaviour of two stochastic gradient adaptive MBER algorithms averaged over 100 independent simulation runs having SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$ and an initial weight vector of $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$. For both the LBER and ALBER algorithms we used a step size of $\mu = 0.01$ and a kernel width of (a) $\rho_n = 4\sigma_n = 0.8944$ and (b) $\rho_n = 6\sigma_n = 1.3416$, where $\sigma_n^2 = 0.05$. The BER of the LBER and ALBER algorithms was evaluated based on Equations (3.121) and (3.129), respectively. The corresponding AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.



(a) $\mu = 0.01, \rho_n = 4\sigma_n = 0.6332$



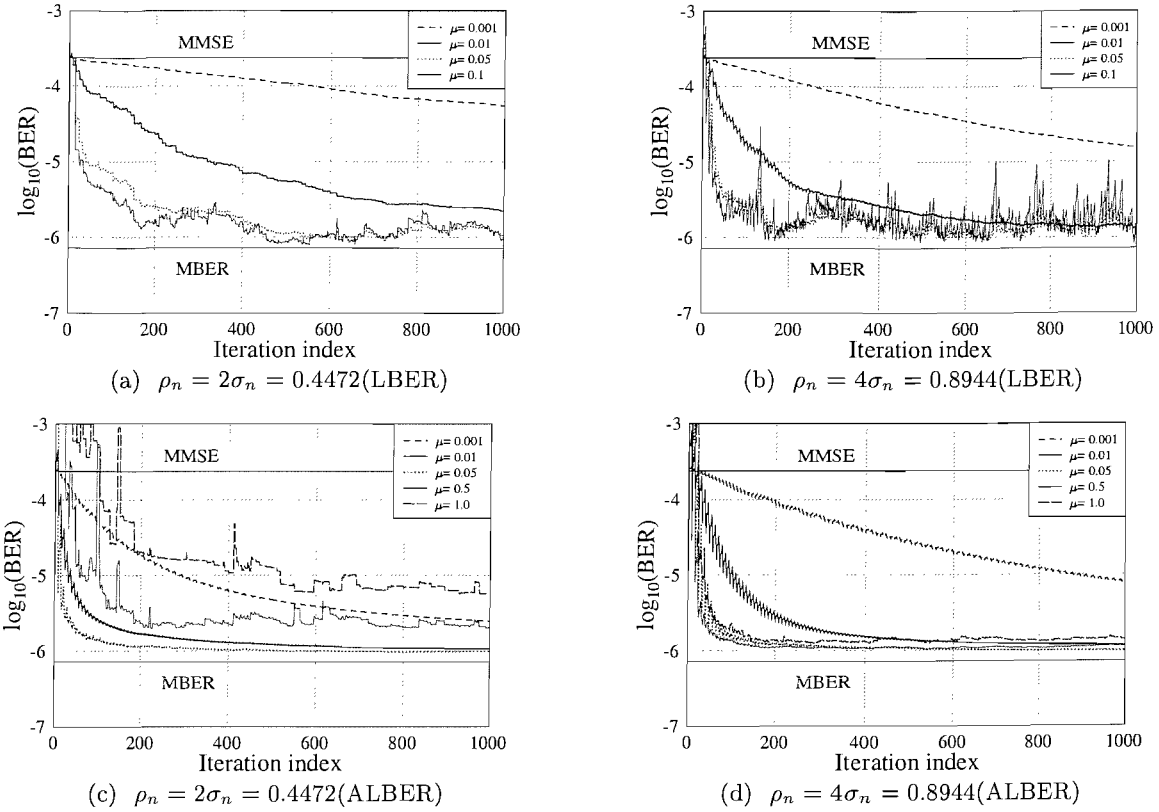
(b) $\mu = 0.01, \rho_n = 8\sigma_n = 1.2664$

Figure 3.44: Convergence behaviour of two stochastic gradient adaptive MBER algorithms averaged over 100 independent simulation runs having SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$ and an initial weight vector of $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$. For both the LBER and ALBER algorithms we used a step size $\mu = 0.01$ and a kernel width of (a) $\rho_n = 4\sigma_n = 0.6332$ and (b) $\rho_n = 8\sigma_n = 1.2664$, where $\sigma_n^2 = 0.025$. The BER of the LBER and ALBER algorithms was evaluated based on Equations (3.121) and (3.129), respectively. The corresponding AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.

algorithm, despite the fact that the former has a lower computational complexity. As the kernel width is increased to $\rho_n = 6\sigma_n$, as seen in Figure 3.43 (b), the convergence rate of both algorithms becomes slower, especially for the ALBER algorithm. It is observed in Figure 3.43 (b) that the LBER algorithm converges faster than the ALBER algorithm, but it settles at a higher steady-state BER value. The BER versus iteration index curves of the algorithms cross each other around the iteration index of 400. Note that both the LBER and ALBER curves of Figure 3.43 (b) fluctuate somewhat more widely than the corresponding curves of Figure 3.43 (a). The BER fluctuations of the ALBER algorithm are more visible during the transient period, while for the LBER algorithm the BER versus iteration index curve slightly fluctuates across the entire observation interval, although upon approaching the steady state BER the fluctuations become more moderate. This phenomenon is more obvious for the unequal-power user scenario of Figure 3.44, where it may be observed that the BER of the ALBER algorithm continues to slightly undulate after approaching the steady state BER. It is observed from Figure 3.44 (b) that although the ALBER algorithm converges earlier, namely after approximately 300 iterations, it finally settles at a similar BER value of BER $= 10^{-4.2} \approx 6.31 \cdot 10^{-5}$ to that of the LBER algorithm. The performance of both the LBER and ALBER algorithms characterised in Figure 3.44 is less impressive than one might expect, since in Figure 3.44 (a) both curves exhibit a slow convergence, while in Figure 3.44 (b) the algorithm's achievable BER is still substantially higher than that of the exact MBER solution.

As in the context of the BACG algorithm of Section 3.4.2, the specific choice of the parameters μ and ρ_n is also important in the context of the LBER and ALBER algorithms of Sections 3.4.3.1 and 3.4.3.2, respectively. Figures 3.45, 3.46, 3.47 and 3.48 demonstrate the effects of varying these two parameters.

For the LBER algorithm of Section 3.4.3.1 invoked in an equal-power user scenario, it can be seen from Figures 3.45 (a) and 3.45 (b) that using a small step size of $0.05 < \mu < 0.1$ is necessary for the sake of attaining a rapid convergence, provided that the associated kernel width ρ_n is not too small, i.e. when we have $2\sigma_n < \rho_n < 4\sigma_n$. A smaller step size of $\mu \leq 0.01$ results in a slow convergence, although a good steady state BER value may be achieved, when using a kernel width of $4\sigma_n$, as shown in Figure 3.45 (b). Figure 3.46 (a) portrays the convergence behaviour of the LBER algorithm for a step size of $\mu = 0.05$, employing kernel widths of $\rho_n = 0.7\sigma_n, 2\sigma_n, 4\sigma_n$. Note in Figure 3.45 (b) that when using a kernel width of $\rho_n = 4\sigma_n$ and a step size of $\mu = 0.05$, the LBER algorithm converges after about 400 iterations, but it exhibits BER fluctuations. This may be engendered by the fact that the LBER algorithm fails to approach the exact minimum BER value and also a consequence of its sensitivity to the choice of the algorithmic parameters μ and ρ_n. Observe in Figures 3.45 (a) and 3.45 (b) that for step size values of $\mu = 0.05$ and $\mu = 0.1$, the fluctuations of the LBER algorithm's BER curve become more substantial as the kernel width ρ_n is increased from $\rho_n = 2\sigma_n$ to $\rho_n = 4\sigma_n$. Increasing the step size beyond $\mu = 0.1$ will produce more dramatic fluctuations, as shown in Figure 3.45 (b). For a smaller step size of $\mu = 0.01$ it is observed in Figure 3.46 (b) that the range of kernel width



(a) $\rho_n = 2\sigma_n = 0.4472$(LBER) (b) $\rho_n = 4\sigma_n = 0.8944$(LBER)

(c) $\rho_n = 2\sigma_n = 0.4472$(ALBER) (d) $\rho_n = 4\sigma_n = 0.8944$(ALBER)

Figure 3.45: Convergence behaviour of two stochastic gradient adaptive MBER algorithms averaged over 100 independent simulation runs having SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$ and an initial weight vector of $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$. For the LBER algorithm we used step sizes of $\mu = 0.001, 0.01, 0.05, 0.1$ and a kernel width of (a) $\rho_n = 2\sigma_n = 0.4472$ and (b) $\rho_n = 4\sigma_n = 0.8944$. For the ALBER algorithm we used step sizes of $\mu = 0.001, 0.01, 0.05, 0.5, 1.0$ and a kernel width of (c) $\rho_n = 2\sigma_n = 0.4472$ and (d) $\rho_n = 4\sigma_n = 0.8944$. The BER of the LBER and ALBER algorithms was evaluated based on Equations (3.121) and (3.129), respectively. The associated AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.

values ρ_n has now been changed to $4\sigma_n < \rho_n < 6\sigma_n$. This new kernel width range however imposed a slower convergence rate on the LBER algorithm's BER curve, which is associated with a slightly higher steady-state BER value, than that associated with the parameters' optimum range of $0.05 < \mu < 0.1$ and $2\sigma_n < \rho_n < 4\sigma_n$.

On the other hand, Figures 3.47 (a), 3.47 (b), 3.48 (a) and 3.48 (b) show the convergence behaviour of the LBER algorithm for the unequal signal power scenario, where we have SNR = 13dB and all four interfering users have a 6dB higher signal power, i.e. we have $\text{INR}_i = \text{SNR} + 6\text{dB}$ for $i = 2, 3, 4, 5$. In comparison to the equal-power user scenario of Figure 3.45, the BER curves of Figure 3.47 fluctuate more dramatically. It is explicit in both Figures 3.47 (a) and 3.47 (b) that the LBER algorithm requires a step size of $\mu = 0.01$ for the sake of attaining both a rapid convergence and a less erratically fluctuating BER curve. It is also observed from Figure 3.48 (a) that for this particular step size of $\mu = 0.01$, the LBER algorithm's sensitivity to the specific choice of the kernel width ρ_n has been reduced, i.e. we have $2\sigma_n < \rho_n < 10\sigma_n$. More smooth BER curves may be obtained by using a step

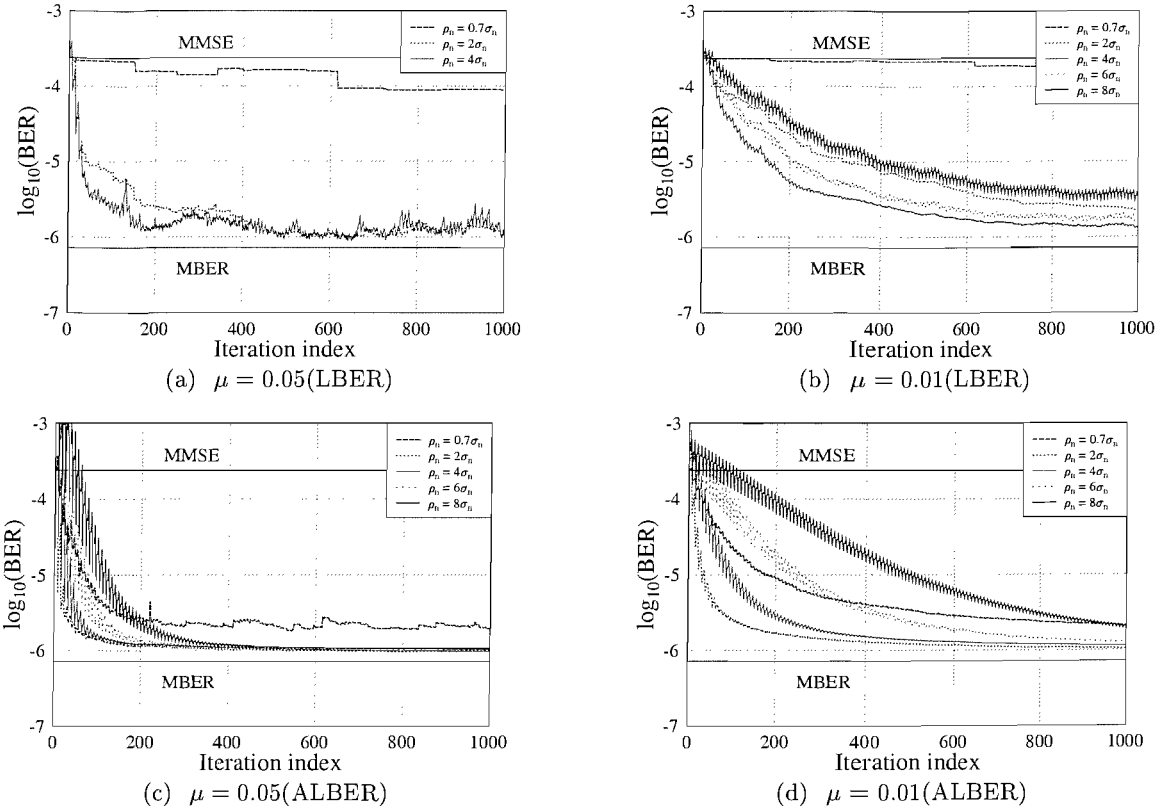
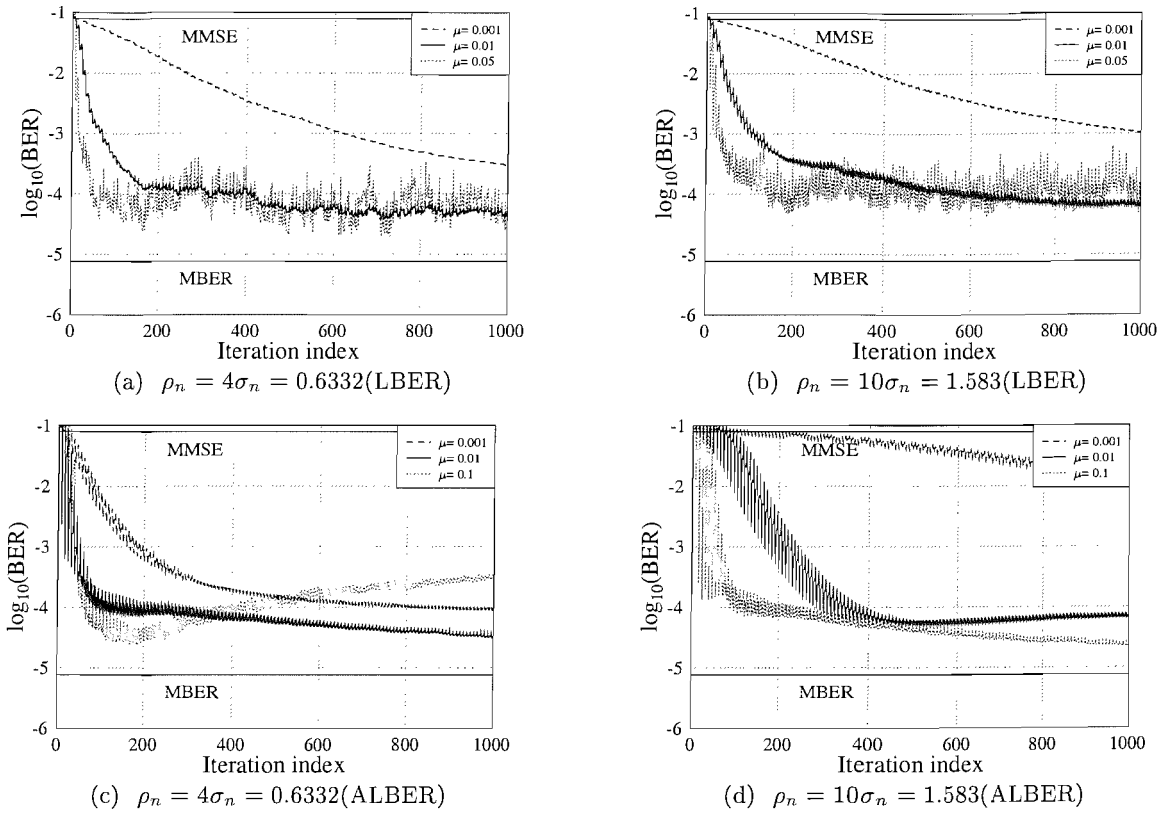


Figure 3.46: Convergence behaviour of two stochastic gradient adaptive MBER algorithms averaged over 100 independent simulation runs having SNR = 10dB and INR_i = 10dB for $i = 2, 3, 4, 5$ and an initial weight vector of $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$. For the LBER algorithm we used kernel widths of $\rho_n = 0.7\sigma_n, 2\sigma_n, 4\sigma_n$ with a step size of $\mu = 0.05$ in subfigure (a) and kernel widths of $\rho_n = 0.7\sigma_n, 2\sigma_n, 4\sigma_n, 6\sigma_n, 8\sigma_n$ with a step size of $\mu = 0.01$ in subfigure (b). For the ALBER algorithm we used kernel widths of $\rho_n = 0.7\sigma_n, 2\sigma_n, 4\sigma_n, 6\sigma_n, 8\sigma_n$ for a step size of (c) $\mu = 0.05$ and (d) $\mu = 0.01$. The BER of the LBER and ALBER algorithms was evaluated based on Equations (3.121) and (3.129), respectively. The associated AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.

size of $\mu = 0.001$, although the attainable BER performance verges on the unacceptable, since the associated convergence rate is too slow.

Let us now focus our attention on the Approximate LBER (ALBER) algorithm characterised in Figures 3.45 and 3.46, as well as in Figures 3.47 and 3.48 for equal-power users and unequal-power users, respectively. We observe in Figures 3.46 (c) and 3.46 (d) that the BER curve of the ALBER algorithm generated for the equal-power users tends to fluctuate during the transient period. After approaching the steady state BER value, we observe that the BER fluctuations had disappeared, resulting in a fairly smooth BER curve. Observe in Figures 3.46 (c) and 3.46 (d) that for scenarios having a kernel width of $\rho_n = 2\sigma_n, 4\sigma_n, 6\sigma_n$ as well as $\rho_n = 2\sigma_n, 4\sigma_n$ and for the particular step sizes of $\mu = 0.05$ as well as $\mu = 0.01$, respectively, the achievable BER is lower than that of the corresponding LBER algorithm, in addition to the benefit of attaining a convergence at approximately 200 and 400 iterations. More explicitly, these convergence periods are shorter than those needed by



(a) $\rho_n = 4\sigma_n = 0.6332$(LBER) (b) $\rho_n = 10\sigma_n = 1.583$(LBER)

(c) $\rho_n = 4\sigma_n = 0.6332$(ALBER) (d) $\rho_n = 10\sigma_n = 1.583$(ALBER)

Figure 3.47: Convergence behaviour of two stochastic gradient adaptive MBER algorithms averaged over 100 independent simulation runs having SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$ and an initial weight vector of $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$. For the LBER algorithm we used step sizes of $\mu = 0.001, 0.01, 0.05$ and a kernel width of (a) $\rho_n = 4\sigma_n = 0.6332$ and (b) $\rho_n = 10\sigma_n = 1.583$. For the ALBER algorithm we used step sizes of $\mu = 0.001, 0.01, 0.1$ and a kernel width of (c) $\rho_n = 4\sigma_n = 0.6332$ and (d) $\rho_n = 10\sigma_n = 1.583$. The BER of the LBER and ALBER algorithms was evaluated based on Equations (3.121) and (3.129), respectively. The corresponding AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.

the corresponding LBER algorithm, as shown in Figures 3.46 (a) and 3.46 (b), respectively. For the particular kernel width of $\rho_n = 4\sigma_n$ invoked in Figure 3.45 (d) the ALBER algorithm was found to be less sensitive to the variation of the step size μ, provided that the step size is in the range of $0.01 < \mu < 1.0$. However, for the sake of achieving a rapid convergence within less than 200 iterations, the employment of a step size range of $0.01 \ll \mu < 1.0$ is recommended. Having too large a step size for the ALBER algorithm may result in the algorithm settling at a higher steady state BER value, as suggested by the curves of Figure 3.45 (c) associated with $\mu = 0.5$ and $\mu = 1.0$, especially if small kernel width, such as $\rho_n \leq 2\sigma_n$ is used.

Figures 3.47 and 3.48 also show the achievable performance of the ALBER algorithm for the unequal-power user scenario, i.e. when all interfering users' signals arrive at 6dB higher power than that of the desired user, who has an SNR of 13dB. We observed in Figure 3.47 (c) that the ALBER algorithm is becoming more sensitive to the choice of the step size μ. Too small a step size, such as $\mu = 0.001$ results in a slow convergence, while a larger step size of $\mu = 0.1$ may result in instability

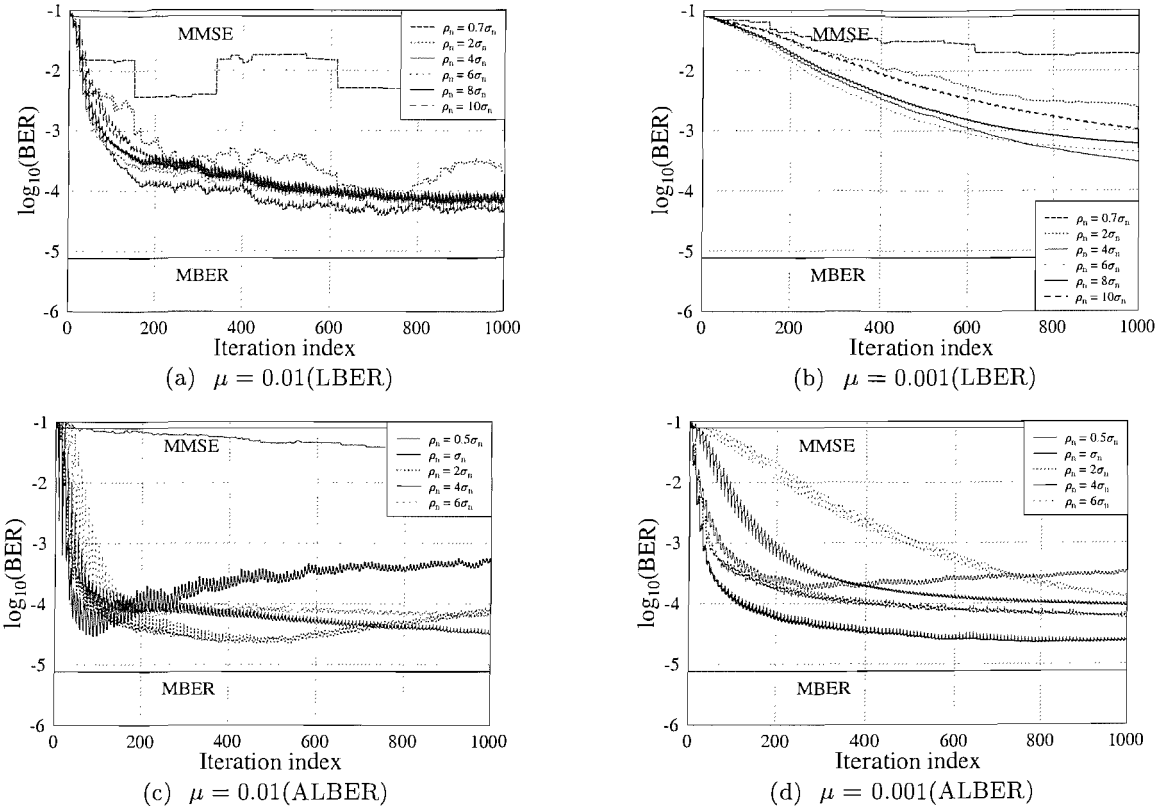


Figure 3.48: Convergence behaviour of two stochastic gradient adaptive MBER algorithms averaged over 100 independent simulation runs having SNR = 13dB and INR_i = SNR +6dB for $i = 2, 3, 4, 5$ and an initial weight vector of $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$. For the LBER algorithm we used kernel widths of $\rho_n = 0.7\sigma_n, 2\sigma_n, 4\sigma_n, 6\sigma_n, 8\sigma_n, 10\sigma_n$ and a step size of (a) $\mu = 0.01$ and (b) $\mu = 0.001$. For the ALBER algorithm we used kernel widths of $\rho_n = 0.5\sigma_n, \sigma_n, 2\sigma_n, 4\sigma_n, 6\sigma_n$ and a step size of (c) $\mu = 0.01$ and (d) $\mu = 0.001$. The BER of the LBER and ALBER algorithms was evaluated based on Equations (3.121) and (3.129), respectively. The corresponding AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.

of the ALBER algorithm, potentially perturbing its ability to converge to a steady state BER value. The best option is to use the step size of $\mu = 0.01$, since it allows its convergence towards a lower BER value, but it may require an excessive number of iterations to arrive at a steady state. For larger kernel width values, such as $\rho_n = 10\sigma_n$, a slightly larger step size than $\mu = 0.01$ is required for achieving a better steady state BER value, as exemplified by Figure 3.47 (d). These observations are further evidenced by the corresponding BER curves of the ALBER algorithm portrayed in Figure 3.48. Figure 3.48 (d) shows that for a step size of $\mu = 0.001$ the kernel width value range of $\sigma_n \leq \rho_n < 4\sigma_n$ is acceptable, while for the step size of $\mu = 0.01$ used in Figure 3.48 (c) the acceptable range was $4\sigma_n < \rho_n < 6\sigma_n$.

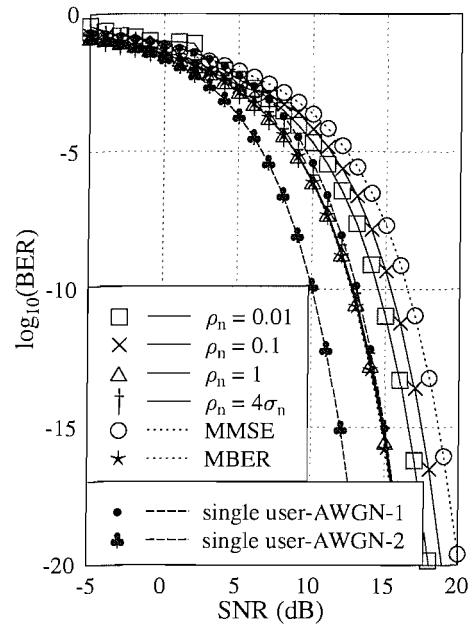
We observe in all scenarios investigated in conjunction with the equal-power users of Figures 3.43, 3.45 and 3.46, that the ALBER algorithm generally performs better or at least as well as the LBER algorithm in terms of both the attainable convergence rate and the achievable steady state BER value, despite its low complexity. We infer based on all plots seen in Figures 3.45, 3.46, 3.47 and

Condition	SNR = 10dB and INR_i = 10dB ($\sigma_n = \sqrt{0.05}$)		SNR = 13dB and INR_i = SNR +6dB ($\sigma_n = \sqrt{0.025}$)	
Adaptive algorithm	step size μ	kernel width ρ_n	step size μ	kernel width ρ_n
LBER	$0.05 < \mu < 0.1$	$2\sigma_n < \rho_n < 4\sigma_n$	$\mu = 0.01$	$2\sigma_n < \rho_n < 10\sigma_n$
ALBER	$0.01 \ll \mu < 1.0$	$2\sigma_n < \rho_n < 6\sigma_n$	$0.001 < \mu \ll 0.01$	$\sigma_n \leq \rho_n < 4\sigma_n$
BACG	$\mu \geq 0.5$	$2\sigma_n < \rho_n \leq 7\sigma_n$	$\mu \geq 0.5$	$2.5\sigma_n \leq \rho_n \leq 3.5\sigma_n$

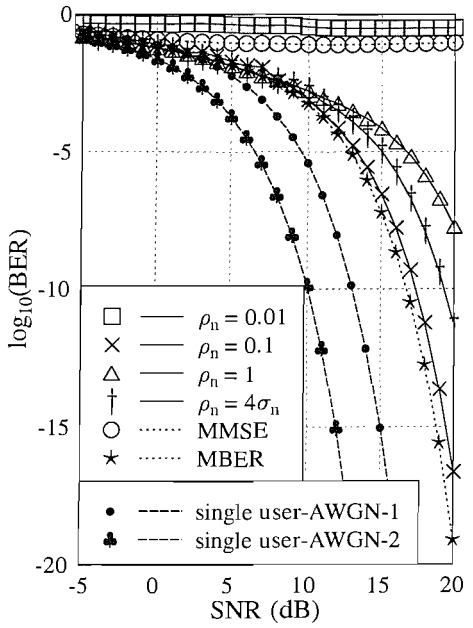
Table 3.7: Acceptable range of values for step size μ and kernel width radius ρ_n of LBER, ALBER and BACG algorithms of Sections 3.4.3.1, 3.4.3.2 and 3.4.2, respectively, for equal-power users of SNR = 10dB (i.e. $\sigma_n = \sqrt{0.05}$) and INR_i = 10dB and unequal-power users of SNR = 13dB (i.e. $\sigma_n = \sqrt{0.025}$) and INR_i = SNR +6dB for $i = 2, 3, 4, 5$. The corresponding AOA scenario was plotted in Figure 3.18 and an AWGN channel was used.

3.48 that regardless of the steady state BER value the choice of a particular step size mainly affects the convergence rate of the algorithms. The LBER algorithm appears to perform best, when using a step size from the range of $0.05 < \mu < 0.1$ and $\mu = 0.01$ for the equal-power and unequal-power user scenarios, respectively, provided that it is accompanied by a suitable kernel width, perhaps values of $2\sigma_n < \rho_n < 4\sigma_n$ for the equal-power user scenario and $2\sigma_n < \rho_n < 10\sigma_n$ for the unequal-power user scenario. By contrast, for the ALBER algorithm employed for detecting equal-power users, a pair of the adaptive parameters μ and ρ_n residing in the range of $0.01 \ll \mu < 1.0$ and $2\sigma_n < \rho_n < 6\sigma_n$ is recommended. For the associated unequal-power users, we suggest to use a small step size of $0.001 < \mu \ll 0.01$ and a kernel width value within the range of $\sigma_n \leq \rho_n < 4\sigma_n$, for the sake of attaining the best possible performance, both in terms of the achievable convergence rate and the steady state BER value. Table 3.7 summarises the acceptable ranges of the adaptive parameters μ and ρ_n of the LBER and ALBER algorithms, including also the block-data based BACG algorithm.

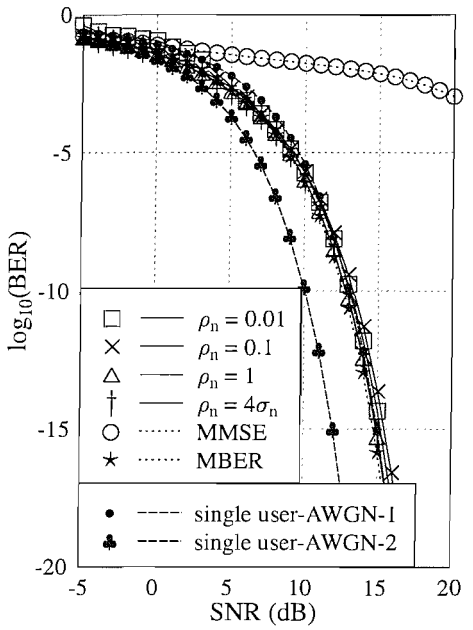
For the sake of comparing the BER performance of the stochastic gradient based adaptive MBER algorithms to that of the block-data based BACG algorithm, we plotted the ALBER algorithm's BER performance in Figure 3.49 as a function of the users' signal power spanning the range of $-5 \leq$ SNR(dB) ≤ 20, under the same conditions as those used for Figure 3.41. However, for the ALBER algorithm of Figure 3.49 we use a step size of $\mu = 0.05$ instead. Similar to the BACG algorithm's performance portrayed in Figure 3.41, it is observed in Figures 3.49 (b) and 3.49 (d) that the ALBER algorithm is also deficient in conjunction with a small kernel width value of $\rho_n = 0.01$. However, for the other conditions considered, the ALBER algorithm generally performs better than the BACG algorithm of Figure 3.41. Note that increasing the kernel width values does not necessarily increase the associated BER performance of the ALBER algorithm. Specifically, kernel width values of $\rho_n = 1, 4\sigma_n$ perform best for equal-power users and for the INR_2 = SNR +6dB conditions of Figures 3.49 (a) and 3.49 (c), respectively. By contrast, the kernel width value of $\rho_n = 0.1$ performs best for unequal-power users associated with INR_i = SNR +6dB for $i = 2, 3, 4, 5$ as well as INR_5 = SNR +6dB, as seen in Figures 3.49 (b) and 3.49 (d), respectively.



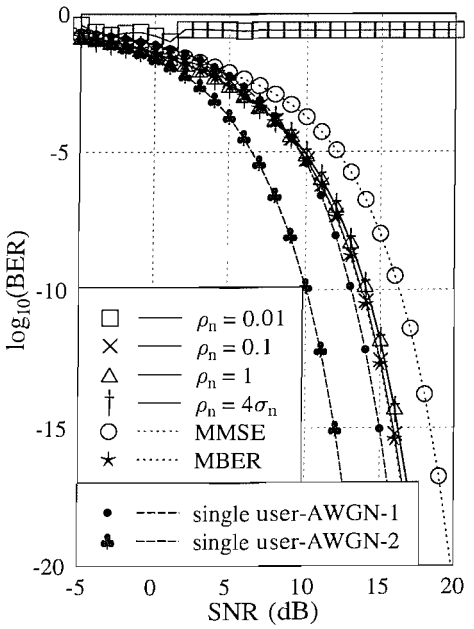
(a) SNR = INR_i for i = 2, 3, 4, 5



(b) INR_i = SNR +6dB for i = 2, 3, 4, 5



(c) SNR = INR_i for i = 3, 4, 5 and INR_2 = SNR +6dB



(d) SNR = INR_i for i = 2, 3, 4 and INR_5 = SNR +6dB

Figure 3.49: BER performance of the ALBER algorithm, based on Equation (3.129), under four different conditions; (a) SNR = INR_i for $i = 2, 3, 4, 5$, (b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$, (c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB for two-element uniform linear array. The ALBER algorithm was using kernel width values of $\rho_n = 0.01, 0.1, 1, 4\sigma_n$ and a step size of $\mu = 0.05$. These results were recorded for transmission over the AWGN channel for five users and the corresponding AOA scenario was plotted in Figure 3.18.

3.7 Conclusions

An adaptive MBER beamforming technique has been developed. As opposed to the closed-form solution of the MMSE approach characterised in Equation (3.25), the beamformer's weight values based on the MBER criterion may be obtained using the simplified conjugate gradient algorithm, as outlined in Equations (3.93) to (3.96) in conjunction with the BER gradient of $\nabla_{\mathbf{w}} P_E(\mathbf{w})$ expressed in Equation (3.89). Assuming a perfect knowledge of all system parameters for the sake of obtaining this optimum solution is impractical in realistic situations. Therefore a number of adaptive MBER beamforming algorithms were introduced. In obtaining the adaptive MBER algorithms, a PDF estimate using Parzen's [224] kernel density estimator of Equation (3.100), which was outlined in Section 3.4.1 was derived.

In addition to the theoretical MBER beamforming solution, the performance of the various adaptive versions of the MBER-based beamforming algorithms was characterised. Specifically, the block-data based adaptive algorithm known as the Block Adaptive Conjugate Gradient (BACG) algorithm of Section 3.4.2 and the stochastic gradient based adaptive algorithms, which include the Least Bit Error Rate (LBER) and the Approximate LBER (ALBER) algorithms of Sections 3.4.3.1 and 3.4.3.2, respectively, were characterised in Section 3.6, in terms of their achievable convergence rate, while also quantifying their ability to converge to a BER close to that of the ideal MBER solution.

It has been shown in Figures 3.7 to 3.11 that the MBER beamforming solution has an advantage over the MMSE beamforming solution. It was observed in Figures 3.19 and 3.20 that the MBER optimisation criterion is superior to the more commonly used MMSE optimisation criterion. Figure 3.21 has shown that the MBER solution always requires a lower desired user signal power than that necessitated by the MMSE solution for the sake of achieving a particular BER, especially when the angular separation between the desired user and the nearest interfering source is small. Explicitly, as shown in Figure 3.21 (b), this happens when the angular separation is in the ranges of $60^o \leq |\theta_2 - \theta_3| \leq 100^o$ and $250^o \leq |\theta_2 - \theta_3| \leq 270^o$. A low angular separation of 5^o between the desired user and the closest interfering user may be observed in Figure 3.33 (a), when six users are supported by a four-element antenna array. Furthermore, Figure 3.22 shows that the MBER solution has a better near-far resistance than the MMSE solution. We also observed in Figures 3.31 (a) and 3.31 (b) that the MBER solution is more robust to 'array-overloading', when the number of users supported is higher than the number of array elements. The MMSE beamformer exhibited a high BER floor, when more than six users were supported. On the other hand, the MBER beamformer was capable of maintaining a BER as low as 10^{-5}, even when supporting nine users, each having a SNR of 20dB. The philosophical differences between the MMSE and MBER approach in optimising the beamformer's weights may be visualised with the aid of Figures 3.1, 3.27 and 3.34. Given that the desired user always transmits $b_1 = +1$, it becomes explicit from Figure 3.34 that whilst the MMSE solution attempts to adjust the weights for the sake of positioning the output phasors as close to ± 1 as possible, the MBER beamformer

maximises the distance of all phasor points from $y_R = 0$, concentrating on minimising the BER.

The conditional probability density function (PDF) $p(y_R)$ of Equation (3.59) plays an important role in analysing the performance of the MBER algorithm, since the probability of error $P_E(\mathbf{w}(n))$ expressed in Equation (3.61) is associated with the specified area under the PDF curve. Furthermore, the shape of the PDF provides an indication of the expected BER performance gap between the MMSE and MBER beamforming solutions. Explicitly, the MMSE solution would perform best, when the PDF of the beamformer's output is Gaussian-like. By contrast, the MBER algorithm performs best, when the PDF of the beamformer's output is non-Gaussian. For the Gaussian-like PDF shown for the two-user scenario of Figures 3.13 and 3.14, the associated BER performance difference would be minimal, but nonetheless the MBER beamformer always outperform the MMSE scheme.

In the context of the MBER beamformer's adaptive implementations, the BACG algorithm of Section 3.4.2 has been shown to converge rapidly, while requiring a reasonably small training block size for accurately approximating the theoretical MBER solution. An appropriate choice of the adaptive parameters is necessary, especially for the kernel width values, which has a significant effect on the attainable performance of the BACG algorithm. The appropriate kernel width however lies within a particular range of values, whose impact on the BACG algorithm's performance would be minimal. Observations inferred from Figures 3.38, 3.39 and 3.40 suggest that a sufficiently larger step size of $\mu \geq 0.5$ invoked in conjunction with a moderate kernel width value spanning the range of $2\sigma_n < \rho_n < 7\sigma_n$ and $2.5\sigma_n \leq \rho_n \leq 3.5\sigma_n$ for the equal-power and unequal-power user scenario, respectively, for the sake of obtaining a near-optimum steady state BER value results in rapid convergence. In comparison to the exact MBER performance of Figure 3.19 for a two-element antenna array, Figure 3.41 shows further that the BACG algorithm's performance is dependent on the kernel width values. For the sake of improving the BACG performance, the Silverman's rule of thumb given in Equation (3.102) may be employed for estimating an optimal kernel width value, provided that a suitable step size is used. The corresponding BER performance of BACG algorithm portrayed in Figure 3.42 shows a nearly identical BER performance to that of the exact MBER solution for all scenarios, except for the most difficult situation of receiving all interfering users at a level of 6dB higher signal power than the desired user. As the training block length is increased, small improvement on the BER performance has been observed.

The convergence behaviour of the sample-by-sample based LBER and ALBER algorithms was portrayed in Figures 3.43 to 3.48. From our observations based on the equal-power user scenario of the LBER algorithm characterised in Figures 3.43, 3.45 and 3.46, we would suggest that a good choice of the adaptive parameters μ and ρ_n resides within the range of $0.05 < \mu < 0.1$ and $2\sigma_n < \rho_n < 4\sigma_n$, respectively. On the other hand, the ALBER algorithm requires a different range, where the step size μ is within the range of $0.01 \ll \mu < 1.0$, while it tolerates a wider span of kernel width values chosen from the range of $2\sigma_n < \rho_n < 6\sigma_n$. Comparing Figure 3.45 (d) characterising the ALBER algorithm

to Figure 3.45 (b) related to the LBER algorithm, it is seen that the ALBER algorithm is generally insensitive to the specific choice of the step size μ, while for the LBER algorithm it is not possible to tolerate a larger step size of $\mu \geq 0.1$, as this will produce more frequent and larger BER spikes than those shown in Figure 3.45 (b). For the unequal-power user scenario the achievable performance shown in Figures 3.44, 3.47 and 3.48, both the LBER and ALBER algorithms were found to be sensitive to the choice of the step size μ, especially when its value was high. Although the LBER algorithm may tolerate a wider range of kernel width values of $2\sigma_n < \rho_n < 10\sigma_n$, a good balance between the attainable convergence rate and the steady state BER value is only achievable, when the step size of $\mu = 0.01$ is used. It was also observed in Figure 3.47 (c) that the ALBER algorithm requires a much smaller step size of $0.001 < \mu < 0.01$ in the unequal-power scenario, which is in contrast to the equal-power user scenario of Figure 3.45 (d) that requires step size of $\mu \gg 0.01$. In a number of scenarios, the steady state BER value of the ALBER algorithm was found to be better than that of the LBER scheme, as suggested by Figures 3.43, 3.44 (a), 3.45 and 3.46. We may conclude from our observations that the ALBER algorithm of Section 3.4.3.2, whose complexity is of similar order to that of the simple LMS algorithm of Section 2.4.1.1 is capable of performing well in approximating the ideal MBER solution in conjunction with appropriate values of the step size μ and kernel width ρ_n, may perform better than the LBER algorithm. Table 3.7 summarised the acceptable range of the step size μ and kernel width ρ_n values for the equal-power users as well as for the unequal-power users scenario. Similar to Figure 3.41 of the BACG algorithm the ALBER algorithm also shows that its performance depends on the kernel width value used, as observed from Figure 3.49. However, both BACG and ALBER algorithms have their own acceptable range of kernel width values ρ_n and their performance is always worst for condition (b) of Figures 3.41 and 3.49, respectively.

In addition to finding the suitable range of the adaptive parameters μ and ρ_n, we are also concerned with expediting the convergence of this stochastic gradient algorithm, while achieving a BER as near the exact MBER solution as possible. An attractive approach is to employ a genetic algorithm in assisting our search for a good balance between the step size μ and kernel width ρ_n and also for good initial weights for the weight adaptation, as it will be presented in Chapter 4.

Chapter 4

Genetic Algorithm-Assisted MBER Beamforming[2]

4.1 Introduction

In contrast to the multi-dimensional paraboloid-shaped error surface of the Minimum Mean Square Error (MMSE) beamformer, the function describing the Bit Error Ratio (BER) surface of a beamformer as a function of the array-weights is highly nonlinear and has numerous local minima, as shown earlier in Figure 3.3. More explicitly, the BER surface is unlikely to be that of a unimodal function, where we will have only a single minimum point, instead it contains an unknown number of local minima. In Chapter 3 we solved this complex optimisation problem by employing the stochastic gradient based algorithm [155]. This algorithm however requires an appropriate set of initial array weight values and the adjustment of several adaptation parameters, which have to be carefully chosen in order to approach the optimum performance instead of a sub-optimum performance due to getting trapped in a local minimum.

In order to counteract the problems imposed by both the irregular BER surface and the initialisation-dependent final solution when using the conjugate gradient algorithm, in this chapter we invoke Genetic Algorithms (GAs) as a means of obtaining the array weights. The family of GAs has also been shown to be successful in the context of multiuser detection (MUD) in [227–229]. As pointed out in [230–232], GAs are capable of circumventing the above-mentioned potential local minima problems and can be randomly initialised. Thus here we embark on investigating, whether they are capable of overcoming the above-mentioned weaknesses of the gradient based algorithms. Since GAs were detailed in [15,230], in this chapter we will directly invoke them for finding the array weight values that minimise the BER at the beamformer's output, resulting in the MBER solution. Main contributions in the field of using GAs for solving multiuser detection problem is summarised on Table 4.1 [202].

[2]**This chapter is based on collaboration with the coauthors of [157].**

Year	Author	Contribution
'75	Holland [233]	Originally proposed GAs for studying the adaptive process of natural evolution in an artificial system software.
'89	Goldberg [230]	Goldberg further developed GAs in the context of optimisation and machine learning.
'96	Mitchell [232]	A further advancement of GAs in machine learning.
'97	Juntti, Schlosser and Lilleberg [234]	First known study of the application of GAs in MUDs.
'98	Wang, Lu and Antoniou [235]	Proposed a detector for multiuser communications, which is based on the maximum-likelihood decision rule and employs a GA for detecting the user bits sequentially.
'00	Ergun and Hacioglu [236]	Suggested a hybrid approach that combines GAs with a multistage multiuser detector (MSD) in the context of a CDMA system.
'01	Yen and Hanzo [228]	Employed a novel CDMA multiuser receiver based on GAs, for jointly estimating the transmitted symbols and the fading channel coefficients of all users.
'01	Abedi and Tafazolli [237]	Proposed and characterised a genetic implementation of the optimal MUD.
'02	Ng, Yen and Hanzo [238]	Advocated a turbo trellis coded modulation assisted GA-aided reduced complexity MUD (TTCM-GA-MUD) that is capable of providing a considerable coding gain without any bandwidth expansion, while maintaining a low complexity compared to the optimum MUD.
'02	Chiang and Chang [239]	Improved the GA and MSD using eugenic[3] population.
'03	Shayesteh, Menhaj and Nobary [240]	Proposed a modified genetic algorithm for multiuser detection in DS/CDMA systems, which attains a performance comparable to that of the optimum detector with a lower complexity.
'03	Yen and Hanzo [241]	Advocated a spatial diversity reception assisted CDMA multiuser detector based on GAs.
'03	Du and Chan [242]	Invoked a GA for sub-optimal detection in space-time block coding (STBC) aided multiuser detection systems.
'04	Wolfgang, Ahmad, Chen and Hanzo [157]	Introduced a novel GA assisted Minimum Bit Error Ratio (MBER) beamforming technique.
'04	Alias, Chen and Hanzo [243]	Employing GA to solve for the MUD's weight of an SDMA OFDM system based on the MBER criterion.

Table 4.1: Contributions towards the development of GA-aided MUDs [202].

We commence in Section 4.2 by introducing the fundamental concepts of GAs and by defining some of the terminology or procedures involved in a typical GA cycle. In Section 4.3 we will present the specific GA configuration, that will be used in our investigations. This will then be followed in Section 4.4 by our simulation results using both the true and the estimated bit error probability of Equations (4.11) and (4.15), respectively. Our conclusions are provided in Section 4.5.

4.2 Fundamental Concepts and Basic Procedures of Genetic Algorithms

Genetic algorithms constitute an optimisation technique based on the 'survival of the fittest' paradigm found in nature [230]. They commence their search from a large population of randomly generated potential solutions, which are simulated and ranked based on their 'fitness', i.e. how well they perform a given task. It is not guaranteed that a perfect solution for the particular application concerned will be produced, but statistically speaking, some will perform better than others. These 'high performers' will have higher fitness values than their peers and will therefore have more chances of being selected for the production of future solutions. The main philosophy is that by exchanging information between two good solutions, a better solution may be produced. This evolutionary process is carried out for many generations, until a solution that meets the requirement of a particular application or termination criterion is achieved [244].

The flowchart of Figure 4.1 provides an overview of the sequence of operations encountered in a typical genetic algorithm. In order to solve a specific problem, we first have to decide on several parameters and variables that will control the algorithm. Firstly, in the context of our beamforming problem a quantised version of all possible array weight solutions must be generated. Furthermore, a termination criterion has to be chosen, which must be related to the accuracy of the final array weight vector solution. Having initialised the algorithm, the GA's search operations commence with the generation of an initial **population** of P solutions, which are referred to in GA parlance as individuals. Given this initial population, the generation index $g = 1$ denotes the first of the total of G number of generations. We refer to P as the population size. This initial population may be generated randomly, although if some knowledge concerning the optimum solution is known *a priori*, it can be exploited in this initial population for expediting the search. The product of P and G determines the complexity of the optimisation procedure and given a fixed product, different P and G combination may result in dramatically different solutions.

Instead of keeping the population size to a constant P throughout the consecutive generations, it

[3]Unlike the elitist selection which ensures only the fittest member of the parents population is kept in the next generation, the eugenic GA combines both offspring and old population into union population. Then, the fitness values of all the elements in the union population are evaluated and the better chromosomes are selected to produce the eugenic population.

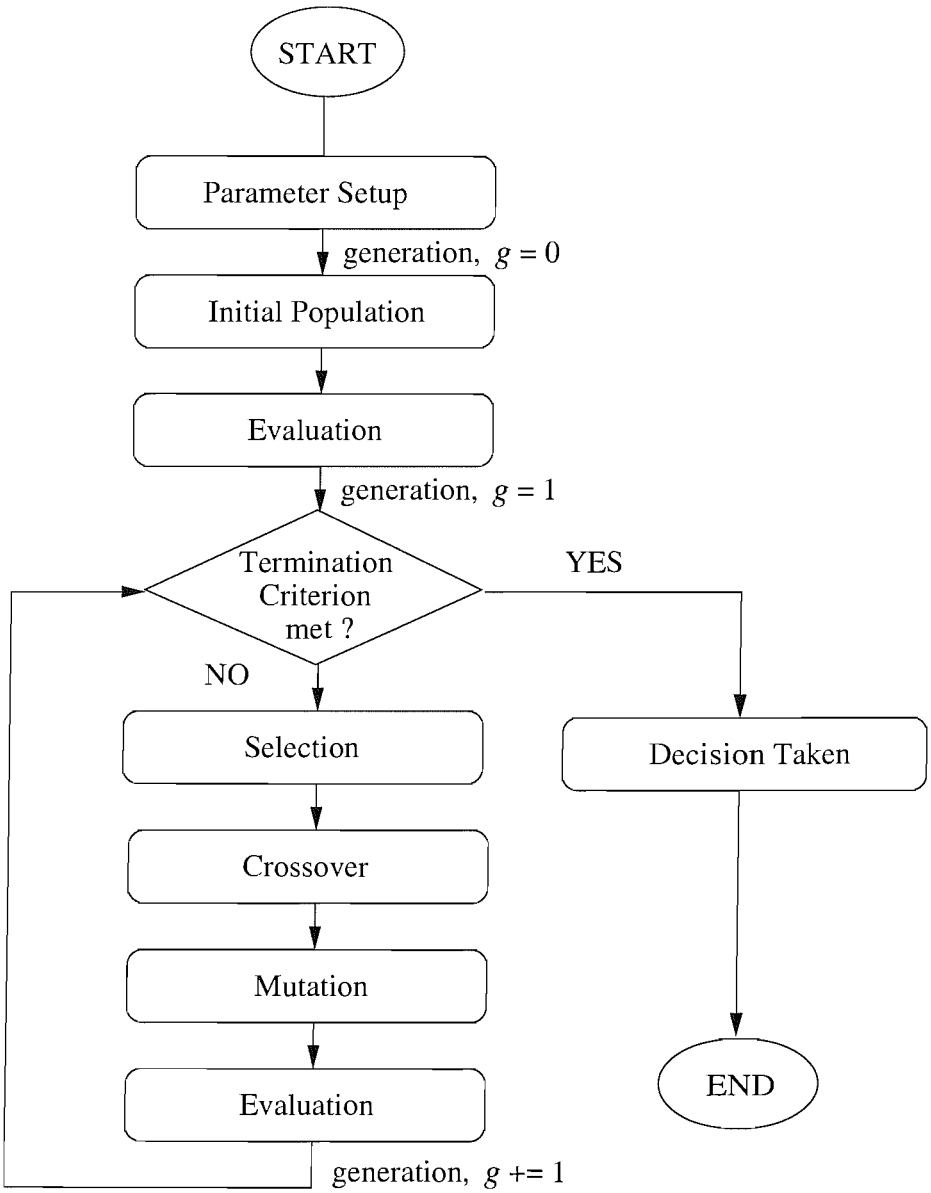


Figure 4.1: Flowchart of a genetic algorithm cycle.

may be reduced to $T < P$, as the generation index g is increasing. Fundamentally, there is no particular reason to keep the population size constant, except for reasons of simplicity. In the GA's population, each individual can be regarded as a string, such as the B-bit vector portrayed in Figure 4.2, whose elements are the decision variables to be optimised. Commonly the vector is in the form of binary bits, thus if the decision variables are not binary in nature, they have to be discretised and encoded to a bit vector. Upon completing the search the optimum B-bit vector will be decoded to its original form in order to obtain the final optimised solution. However, it is also possible to directly optimised real-valued variables, which is more convenient in some applications. Furthermore, real-valued GAs exhibit an increased complexity [230].

One of the most important aspects in the design of GA-based optimisation is the choice of the

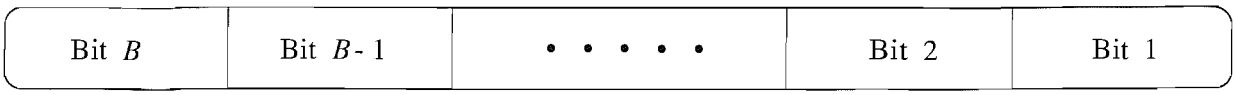


Figure 4.2: A B-bit vector representing an individual. For population of size P, there will be P number of this vector.

fitness measure or the GA cost function, which we always referred to as the fitness function. The merits of each individual will be evaluated and a corresponding fitness value is assigned that reflects how meritorious this individual is in terms of the corresponding beamforming objective function value, such as the Minimum Bit Error Ratio (MBER) in our case. The corresponding fitness is compared to that of the others in the population. In each generation, a new population will be generated by selecting several individuals based on their merit or fitness value. It is the fitness function that will be optimised during the process of finding the best solution. The fitness value is evaluated by substituting the candidate solution represented by the individual under consideration into the objective function. It is a common practice to normalise the fitness value such that it assumes values spanning the range of zero to unity. The GA's cycle proceeds by comparing for example the fitness of all individuals to a defined termination criterion. If the termination condition is met, be it satisfying the fitness threshold or the maximum number of generations allowed, a solution is considered found and the search will be concluded. Otherwise, the GA-based reproduction of individuals continues using one of the GA-based selection methods available and this is then followed by the so-called crossover and mutation processes, before the fitness evaluation procedure is applied again.

Having briefly introduced the philosophy of a genetic algorithmic cycle, in the next section we will further detail the basic processes involved. As shown in the GA cycle of Figure 4.1, the process of selection, crossover and mutation will be detailed in Sections 4.2.1, 4.2.2 and 4.2.3, respectively.

4.2.1 Selection

The creation of a new generation of the GA commences with the process of reproduction or selection. Selection constitutes a particular way of creating the next generation of the population, i.e. a new set of individuals. In the selection process, the fitness of each individual plays an important role, since the GA chooses T number of individuals having the highest fitness value for the next generation, where T may range from 2 to P.

A commonly used selection method is referred to as fitness-proportionate selection [230], where the probability p_p of selecting the pth individual for creating the next generation is proportional to the ratio of its fitness value f_p to the total fitness value of all the T number of selected individual,

Individual p	Fitness f_p	Probability of selection p_p
1	$f_1 = 6.25$	$p_1 = 0.0625$
2	$f_2 = 50$	$p_2 = 0.5$
3	$f_3 = 31.25$	$p_3 = 0.3125$
4	$f_4 = 12.5$	$p_4 = 0.125$
$P = 4$	$\sum_{p=1}^{T=4} f_p = 100$	$\sum_{p=1}^{T=4} p_p = 1$

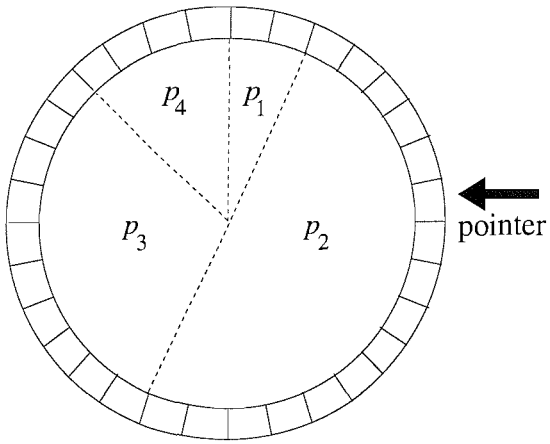


Figure 4.3: An example of fitness-proportionate or roulette-wheel based selection in a GA having a population of $P = 4$ individuals.

which is expressed as:

$$p_p = \frac{f_p}{\sum_p^T f_p} . \tag{4.1}$$

This selection approach may be implemented using the so-called roulette-wheel sampling [230], whereby each individual is allocated a slice of a circular roulette-wheel, proportional in area to the individual's probability of selection [245], i.e. its fitness. For a population size of $T = P = 4$ this procedure is exemplified in Figure 4.3, where we selected four individuals for the next generation by spinning the weighted roulette-wheel four times. As expected, a highly fit individual will have a higher probability of being selected, since it was assigned a higher fraction of the wheel. The second individual involved in this example has the highest fitness ratio, which covers 50% of the roulette-wheel, therefore each spin returns that particular individual ($p = 2$) with a probability of 0.5. Each time we require another offspring of the previous generation for creating the new generation, a simple spin of the weighted roulette-wheel yields the selected candidate. In this way, the more fit individuals have a higher number of offspring in the succeeding generation. Once an individual was selected, a copy of its string represented by the B-bit vector of Figure 4.2 is entered into a so-called mating pool constituted by a tentative new population for further genetic operations.

Apart from being the simplest and hence the most straightforward to implement, roulette-wheel based selection favours the highly fit individuals, which allows the GA to converge faster to a near-optimum solution [246]. However, a specific disadvantage of fast convergence is that it might discourage the exploration of the entire search space. We refer to this phenomenon as 'premature convergence', where the highly fit individuals dominate the population. In order to prevent the predominance of this premature convergence, in certain applications roulette-wheel based selection is only used at a later stage of the optimisation process.

Apart from roulette-wheel based selection, there are numerous other ways of selecting individuals for creating a new generation. For example, the stochastic remainder sampling with replacement [230, 232] or the tournament selection [230, 232] schemes may be invoked for biasing the population towards the best individuals. The principle of the stochastic remainder sampling combined with replacement [230] is that each individual is reproduced using the integer part of the expected number of individuals e_p, calculated according to:

$$e_p = \frac{f_p}{\sum_p^T f_p} \cdot T = p_p \cdot T, \tag{4.2}$$

where T is the total number of individuals to be selected for the next generation and p_p is the probability of selection, similar to that expressed in Equation (4.1). The fraction part of e_p will be used as the weight, i.e. as the direct probability of selection of an individual p on the roulette-wheel. The number of spins made will depend on the balance of individuals needed to complete the T-individual population. This technique combines the deterministic and stochastic approaches, while the earlier fitness-proportionate selection is entirely based on using the stochastic approach. On the other hand, the tournament selection scheme operates by randomly choosing N individuals from the population and copying the best individual of the selected N-individual group into the new population. This process is repeated T times. An illustration of the tournament selection procedure is given in Figure 4.4.

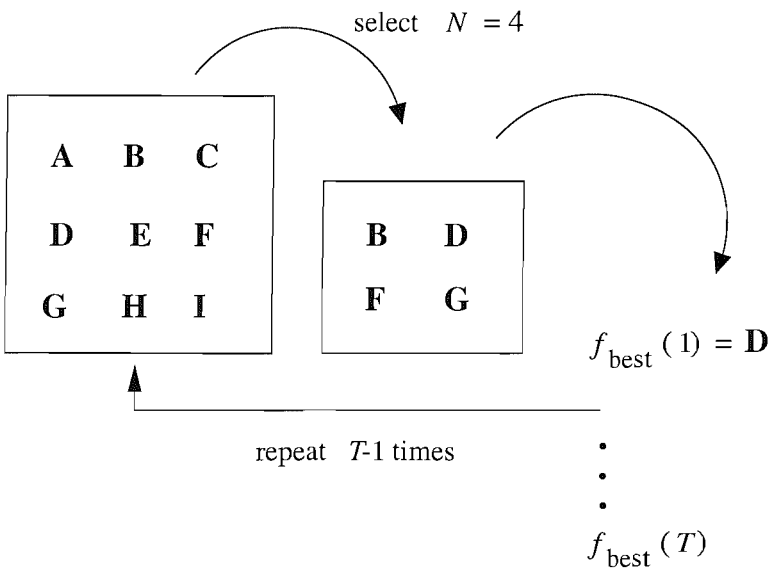


Figure 4.4: An example of tournament selection in a GA having a tournament size of $N = 4$ individuals.

All the selection schemes mentioned can be considered as methods of defining the individual's merit, i.e. the probability of producing offspring. The selection process however does not change any features of the individuals. Let us now proceed by considering the next procedure in the genetic algorithm's cycle seen in Figure 4.1, which is termed as 'crossover'.

4.2.2 Crossover

Reproduction or selection alone does not promote the exploration of new regions of the search space, since it only copies the selected individuals' structure to the so-called mating pool. This is where crossover between a pair of individuals takes place.

In contrast to the selection process, crossover is a process that combines individuals of the current generation based on genetic operations for the sake of generating the individuals of the next generation. It can also be viewed as a structured yet randomised information exchange between strings or individuals of the GA. In its simplest form crossover exchanges certain bits of the parent individuals for the sake of creating an offspring, which is performed with a probability of $p_c < 1$, implying that some of the selected parents might not be swapping any bits. Those parent vectors that are not crossed are simply transferred to the new population unchanged.

In this simple GA operation, the T selected parents will be randomly paired for the sake of exchanging genetic information, hence creating two new offspring, as shown in Figure 4.5 a). This process aims for combining the bit vectors of the parents for the sake of producing a higher fitness pair of offspring for the next generation. Some crossovers might produce less fit offspring than the parents, but nonetheless the average fitness of the offspring's generation is expected to be higher than that of their parents' generation. When considering a B-bit array weight vector as that of Figure 4.2, the crossover point of a single-point scheme may be chosen randomly, for example between the bth and the $(b+1)$st bit position. Apart from the **single-point crossover** scheme shown in Figure 4.5 a), it is also possible to opt for more than one crossover points for the sake of speeding up the search-space exploration process, instead of favouring the convergence to highly fit individuals early in the search. Having an increased number of crossover points however renders the population more random and therefore may delay convergence to the optimum solution. Figure 4.5 b) shows a **double-point crossover** operation, which uses two randomly chosen crossover points. The array weight vector bits that fall between these crossover points are then exchanged between the parents. Another form of crossover can be seen in Figure 4.5 c), which is referred to as the **uniform crossover**, where a 'crossover mask' is invoked instead of one or two crossover points. The crossover mask is a vector that consists of randomly generated binary 1s and 0s, having a length of B. In Figure 4.5 c), offspring 1 is produced by inheriting the array weight vector bit from parent 1 if the corresponding mask 1 bit is a binary 1 or the bit from parent 2 if the corresponding mask 1 bit is a binary 0. Offspring 2 is created in a similar way using the inverse of mask 1, mask 2.

The GA's operations considered so far were quite simple, involving nothing more than random number generations and probabilistic string exchanges. However, the approach of GAs is not entirely random, hence we refer to it as a guided random algorithm. As shown in Figures 4.2 and 4.5, each individual is represented by a string of bits, in that substrings contain hints as to what is important, relevant or optimum to the task concerned [230]. Interpreting GAs from this perspective,

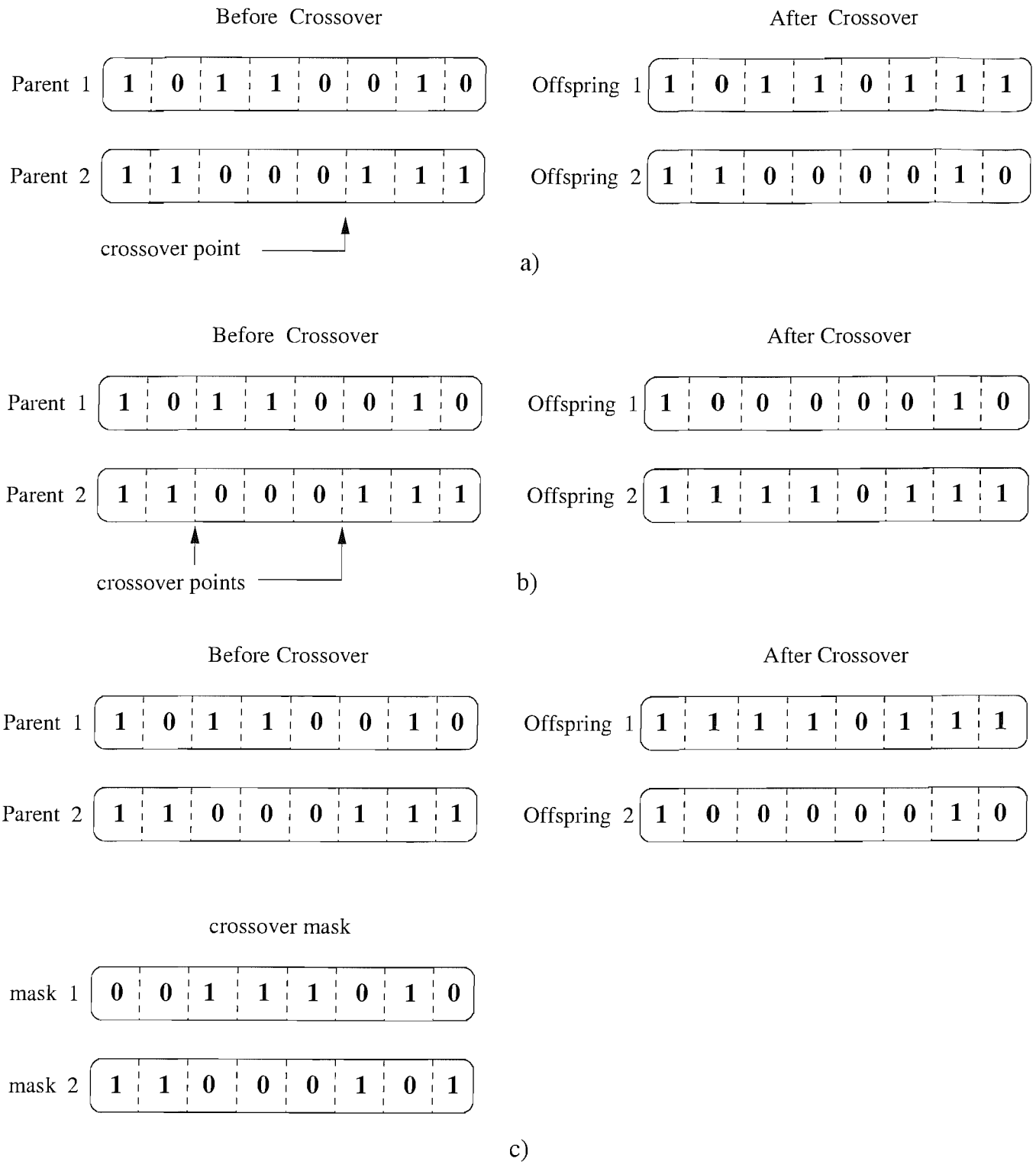


Figure 4.5: An illustration of a) single-point crossover, b) double-point crossover and c) uniform crossover operations.

the population contains not a random set of P individuals, but rather P individuals associated with a fitness-based ranking, which is indicative of the probability of each of them being the optimum solution. Let us now consider the 'mutation' operation of Figure 4.1.

4.2.3 Mutation

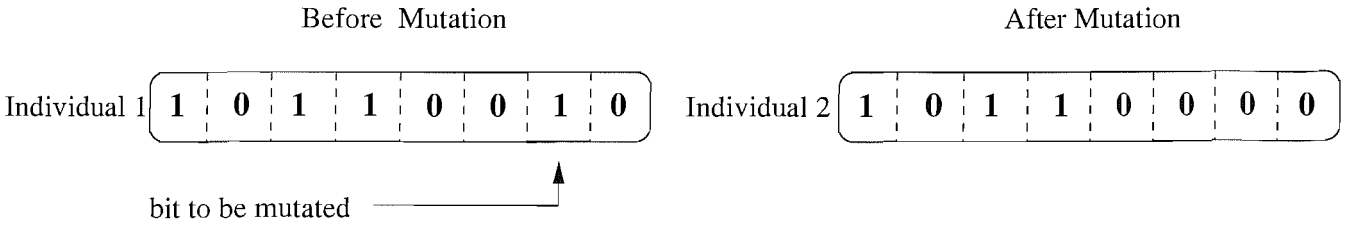


Figure 4.6: An illustration of the mutation process.

When using crossovers, the individuals of population are altered but nonetheless the population may not be sufficiently diverse, converging to a fitness maximum that is not certain to be the global maximum [246]. In order to increase and maintain the diversity of the population and hence force the algorithm to extend the search to previously unexplored areas of the search space, another operation referred to as mutation is introduced. In other words, mutation can be viewed as an operator invoked for the sake of avoiding premature convergence, which might trap the GA in local minima. Similarly to the crossover operation, mutation also permits the creation of new individuals from the existing individuals, but it is performed on a bit-by-bit basis. It can be defined as the random alteration of the bth decision variable with a probability of p_m, where we have $b = 1, 2, \ldots, B$.

The probability of mutation p_m is typically set to a lower value than the crossover probability p_c. Often it is set to the reciprocal value of the population size P or chromosome length B, i.e. we set $p_m = 1/P$ or $p_m = 1/B$, respectively. Alternatively, it may be expressed as a function of the crossover probability p_c to obey $p_m = 1 - p_c$, or may be fixed to a constant value, such as $p_m = 10^{-3}$. The probability p_m must not be set high, since it might mutate the fit individuals, decreasing the fitness of the individuals, which renders the convergence to the global solution slower. Assigning a mutation probability p_m as high as 0.5 resembles a random search, regardless of the crossover probability p_c and the population size P. In general, the role of mutation is to prevent premature convergence, which might lead to a local optimum without exploring the entire search space.

For **binary coded** individuals, mutation simply implies toggling the value of '0' to '1' or vice versa. Assuming a mutation probability p_m of 0.05, for example, with 40 (i.e. four 10-bit individuals) available bit positions we would expect $40 \times 0.05 = 2$ bits to undergo mutation in a given generation, i.e. two out of 40 bits will be changed from 0 to 1 or vice versa. For **real-valued** individuals the process of mutation is more elaborate, since each decision variable may assume an infinite number of

possible values. As outlined in [245], when a decision variable b is picked for mutation, the direction of mutation is chosen randomly with equal probability, before a real-valued random mutation size $\triangle$ is generated. The value of $\triangle$ is in the range of $[0, \triangle_{max}]$ with $\triangle_{max}$ denoting the maximum range of change. The value of $\triangle_{max}$ has to be determined in advance in order to ensure that the mutated decision variable $b \pm \triangle$ will not exceed the minimum b_{min} and the maximum b_{max} limits specified by the problem, i.e. we have $b_{min} \leq b \pm \triangle \leq b_{max}$.

Indeed, mutation is important, since it permits the incorporation of new individuals and hence avoids the generation of similar individuals, a condition which was referred to above as 'premature convergence'. An example of the binary mutation process is illustrated in Figure 4.6. Having discussed the three fundamental procedures, namely selection, crossover and mutation, it is now necessary to evaluate the fitness of the individuals. In the next section we will describe how the evaluation of the individuals is handled in GAs in order to find a solution to an optimisation problem.

4.2.4 Evaluation: Encoding of Individuals

Following reproduction or selection, crossover and mutation, the individuals of the new population are ready to be evaluated. This is carried out by decoding the new strings from their binary representation and calculating their fitness value.

The simplest encoding method is to use binary coded strings represented by binary 0s and 1s. A binary string C can be decoded as an unsigned integer u. For example, the binary string $C = c_B c_{B-1} \ldots c_2 c_1$ can be decoded to a parameter value of:

$$u = \sum_{b=1}^{B} c_b \cdot 2^{b-1}. \tag{4.3}$$

Unfortunately, the binary encoding of integers results into a limited range of values confined to $[0, 2^B - 1]$, which may not provide the required variety of values for solving a specific optimisation problem. In other words, this unsigned fixed-point integer encoding procedure is straightforward, however, an excessive number of encoding bits may be required for covering the parameter range hosting the optimum value. There are several options for mapping the finite-length string C to the floating point real-valued parameter w of a specific optimisation problem. One option is to map the decoded unsigned integer u linearly from $|0, 2^B - 1|$ to a specified interval $|D_{min}, D_{max}|$ of w. In this way, we can carefully control the range and encoding precision of the decision variables. The precision η of this mapped encoding procedure may be calculated as:

$$\eta = \frac{D_{max} - D_{min}}{2^B}, \tag{4.4}$$

where B is the length of a chromosome or the vector representing an individual, while D_{max} and D_{min} are the maximum and minimum real values of the parameter w, respectively.

In order to construct a multi-parameter encoding scheme, we can simply concatenate the binary encoded representation of as many single parameters as we require. For example, the three parameters $C(1), C(2)$ and $C(3)$ may be concatenated, such that we have:

$$C = C(1)C(2)C(3) \tag{4.5}$$

$$= c(1)_i c(1)_{i-1} \dots c(1)_2 c(1)_1 c(2)_k c(2)_{k-1} \dots c(2)_2 c(2)_1 c(3)_l c(3)_{l-1} \dots c(3)_2 c(3)_1. \tag{4.6}$$

In this multi-parameter encoding scheme each parameter may have its own binary length, i.e. i, k and l does not have to be equal, but for the sake of simplicity we may consider them to be of equal length, corresponding to $i = k = l$. Furthermore, the parameters $C(1), C(2)$ and $C(3)$ may have their own D_{max} and D_{min} values. Having decoded the B-bit string into the desired real-valued representations $w(1), w(2)$ and $w(3)$, in which B is now the sum of the parameters' length expressed as $B = i + k + l$, we can calculate the fitness value of the string C.

With reference to Figure 4.1, we will repeat the consecutive cycles of selection, crossover, mutation and evaluation, until a specific termination criterion is met. In the following section, we will briefly highlight the design options available as termination criterion.

4.2.5 Termination Criterion

Each cycle of selection, crossover, mutation and fitness calculation is deemed to be one generation in the execution of a GA. This cycle will continue until the defined termination condition is met, as shown in Figure 4.1.

Common termination criteria are constituted by the maximum number of generations permitted or attaining a specific cost function value. Generally, if the GA's cycle is executed for many generations, the population will eventually converge to a set of individuals, in which the individual that corresponds to the highest fitness value is deemed to be the optimum or near-optimum solution. Therefore, the relevant parameters of the GA have to be carefully configured for the sake of avoiding premature termination or unnecessary further cycles after reaching a satisfactory solution, i.e. after attaining convergence.

Having introduced the basic procedures of GAs, as outlined in Figure 4.1, in the next section we will focus our attention on the specific GA configuration designed for assisting the operation of MBER beamforming.

4.3 Genetic Algorithm-Asisted MBER Beamforming and Array Weight Quantisation

In this section the GA configuration that will be used for calculating the MBER weights of an antenna array is presented. The general procedure of a GA, which involves creating an initial population, performing selection, crossover, mutation and the individual's evaluation process is portrayed in Figure 4.1. The basic GA operations of selection, crossover and mutation are similar to those described in Sections 4.2.1, 4.2.2 and 4.2.3, respectively. Therefore, here we only highlight the associated differences of the GA procedures described in Section 4.2 and introduce a few additional operations.

4.3.1 Representation

In the context of GAs, the first step of our optimisation process is to encode the parameter to be optimised as a finite-length string, which will then be decoded and evaluated in the context of the specific objective function used, before being assigned an appropriate fitness value f.

Several individuals constitute a population, in which each of them represents a complete solution of the problem to be solved, as shown in Figure 4.7. In a simple GA, an individual typically has a single chromosome, which always represents a positive integer. In adaptive beamforming, we however deal with vectors and floating point complex-valued variables, such that the array weight vector $\mathbf{w}$ is constituted by:

$$\mathbf{w} = [w_1 \;\; w_2 \;\; \ldots \;\; w_L]^T \tag{4.7}$$

$$= [w_{\Re,1} + jw_{\Im,1} \;\; w_{\Re,2} + jw_{\Im,2} \;\; \ldots \;\; w_{\Re,L} + jw_{\Im,L}]^T, \tag{4.8}$$

where $\Re$ and $\Im$ denote the real and imaginary parts of the weights, respectively. We therefore have to find a method for converting our chromosomes into signed floating point values for both the real and imaginary parts of the L array weights. This implies that our chromosome is represented by a complex number having two real-valued parameters, i.e. the real part $\Re$ and the imaginary part $\Im$. The entire representation now depends on the number of antenna elements L used in the antenna array, where we have $L \geq 2$, giving a chromosome length B, which is constituted by at least $2L = 4$ concatenated real values having at least $2LJ = 4J$ bits, where J is the length of each parameter's bit-based representation. In other words, we visualise a chromosome as a vector, which consists of several weight values corresponding to L. Due to having complex-valued array weights, we separate the real and imaginary parts and therefore, the total number of vector elements in a chromosome is twice the number of antenna elements. Using this approach, the length of the chromosome B will also depend on the number of bits J we used for representing a real-valued number. Hence the chromosome length is given by $B = 2LJ$ bits. The detailed structure of an individual is illustrated in Figure 4.7 b).

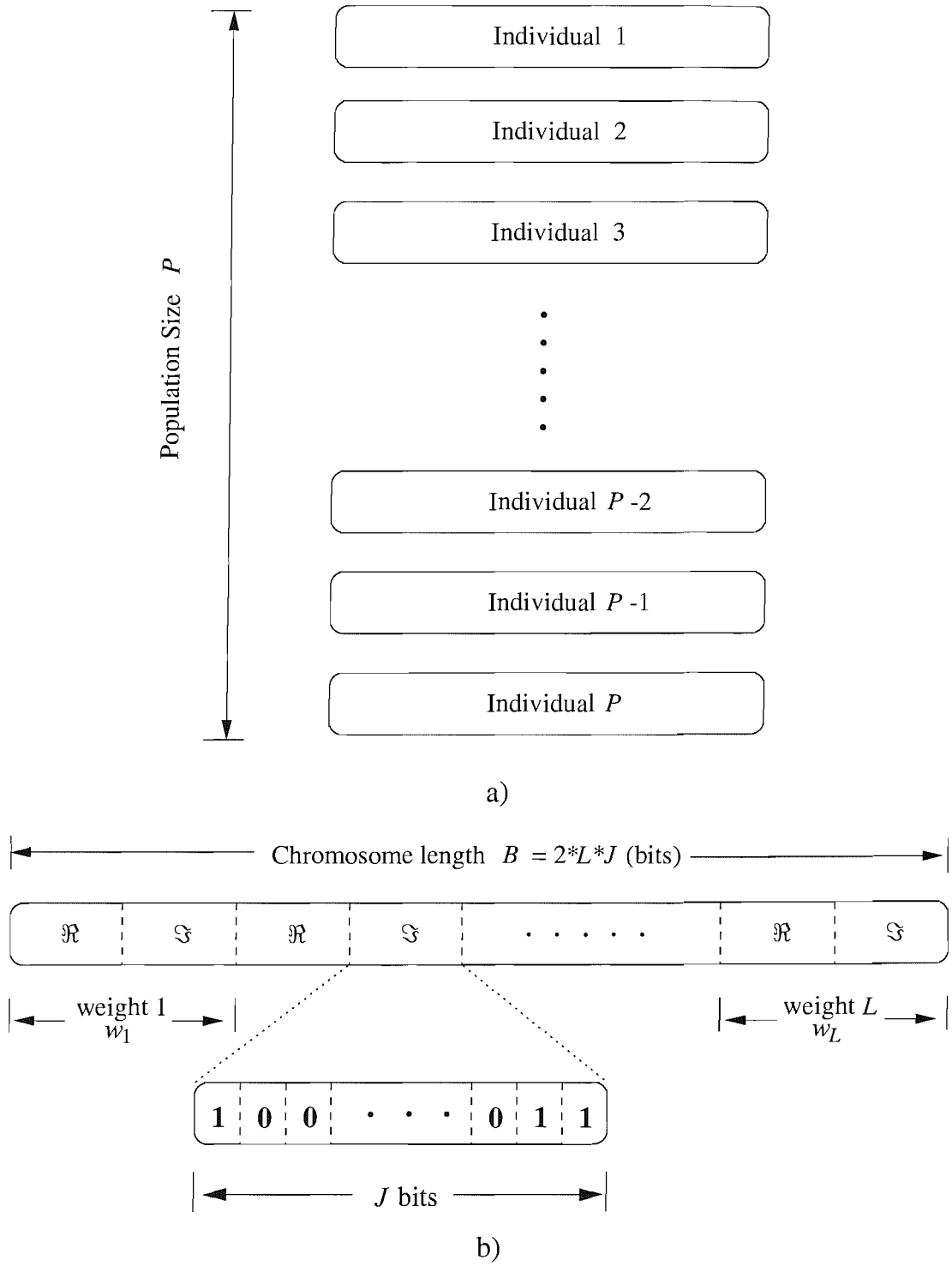


Figure 4.7: An illustration of a) a population of P individuals and b) an individual (or chromosome), where $\Re$ and $\Im$ denote the real and imaginary parts of the corresponding array weight values $w_1, w_2, \ldots, w_L$, respectively.

4.3.2 Encoding of Individuals

We used binary encoding for our GA-assisted MBER beamformer. First we perform the binary to unsigned integer conversion seen in Equation (4.3). We then proceed to convert the unsigned integers u to the signed real numbers w. There are a number of ways to perform this conversion. One method is to reserve a bit (usually the first bit) as the sign bit, i.e. a '1' indicates a negative and a '0' indicates a positive weight. Another way of quantising the array weights is by representing the entire dynamic range of the array weights accommodating both the positive and negative values by J bits. For example, having J bits we can represent 2^J integers ranging from 0 to $2^J - 1$. The signed real value w is obtained by:

$$w = \frac{\sum_{j=1}^{J} c(1)_j 2^{j-1}}{2^J}(D_{max} - D_{min}) + D_{min}, \tag{4.9}$$

where $c(1)_j$ with $j = 1, \ldots, J$ representing the bits in the substring $C(1)$ of a string (an individual) C, such that we have $C = C(2L)C(2L-1)\ldots C(2)C(1)$ and $C(l) = c(l)_J c(l)_{J-1} \ldots c(l)_2 c(l)_1$ with $l = 1, \ldots, 2L$. Assuming that all substrings $C(2L), \ldots, C(1)$ have a J-bit representation, we have $j = 1, \ldots, J = B/2L$. The parameters D_{max} and D_{min} represent the upper and lower boundary of the signed real-valued variable w, respectively.

With reference to Figure 4.1, the operation of a GA commences with an initial set of potential array weights, which may be chosen randomly within a given search-space. The array weight values of the MBER beamforming scheme may be normalised to be within the search-space of $[-1, 1]$, based on Equation (3.70), which states that the BER is invariant to a positive scaling of $\mathbf{w}$. This fixed dynamic range results in a coefficient-resolution that depends only on the number of bits J used for representing a real-valued variable, w. The resolution of the GA's individuals can now be calculated as:

$$\eta = \frac{D_{max} - D_{min}}{2^J} = \frac{1 - (-1)}{2^J} = 2^{1-J}, \tag{4.10}$$

where J is the number of bits representing each concatenated substring describing an individual given by a signed real-valued variable w, while $D_{min} = -1$ and $D_{max} = +1$ are the minimum and maximum legitimate value of each variable w, respectively.

Using $J = 8$ bits for each signed real-valued variable w, the resolution is fixed to $2^{-7} = 1/128 = 0.0078125$. A higher accuracy may be obtained by increasing the number of bits J used for representing the variable w. Having described the GA's individual encoding requirements, in the next section we will consider the MBER beamforming scheme's objective function, which has to be evaluated by the GA.

4.3.3 Fitness Evaluation

It has been shown in the context of Equations (3.68) and (3.69) in Chapter 3 that the probability of error P_E for a given beamformer weight vector $\mathbf{w}$ can be calculated as:

$$P_E(\mathbf{w}) = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q\left(c_{q,+}(\mathbf{w})\right), \tag{4.11}$$

with

$$c_{q,+}(\mathbf{w}) = \frac{\text{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} = \frac{\text{sgn}(b_{q,d})\Re[\mathbf{w}^H\bar{\mathbf{x}}_q^{(+)}]}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}, \tag{4.12}$$

where N_{sb} is the total number of distinct sets of bits transmitted by the M number of simultaneous users with $b_{q,d}$ fixed to a specific value (in this case, $b_{q,d} = +1$), such that for BPSK modulation we have $N_{sb} = 2^{M-1}$, where M is the number of users. Still considering Equation (4.12) $\mathbf{w}$ is the array weight vector, $\mathbf{x}$ is the received signal vector at the antenna array or equivalently the input signal vector of the beamformer defined in Equation (3.4) and σ_n^2 is the variance of the noise vector $\mathbf{n}$. The probability of error P_E quantified in terms of Equations (4.11) and (4.12) was arrived at by calculating the area under the conditional PDF of $y_R(n)$ associated with $b_d(n) = \pm 1$, such that we have:

$$P_E(\mathbf{w}) = \int_{-\infty}^{0} p(y_R| + 1)\, dy_R, \tag{4.13}$$

where the conditional PDF $p(y_R| + 1)$ was given by Equation (3.59) and repeated here as:

$$p(y_R|b_d(n) = +1) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(y_R - \text{sgn}(b_{q,d})\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right). \tag{4.14}$$

The detailed derivation may be obtained with the aid of Equations (3.61) to (3.67). We refer to Equation (4.11) as the true probability of error P_E, which will be used in the GA's fitness function expression for calculating the fitness value of the GA's individuals. For the sake of investigating also a more realistic receiver structure, we replaced the true probability of error P_E by the estimated probability of error $\hat{P}_E$, which can be obtained using Parzen's kernel density estimation [224] outlined in Section 3.4.1 and expressed as:

$$\hat{P}_E(\mathbf{w}(n)) = \frac{1}{T} \sum_{t=1}^{T} Q\left(\hat{c}_t(\mathbf{w}(n))\right), \tag{4.15}$$

where the employment of T training samples has been assumed. The standard Q-function $Q(z)$ was defined in Equation (3.63), while the shorthand of $\hat{c}_t(\mathbf{w}(n))$ is given by:

$$\hat{c}_t(\mathbf{w}(n)) = \frac{\text{sgn}(b_d(t))y_R(t)}{\rho_n\sqrt{\mathbf{w}(n)^H\mathbf{w}(n)}}, \tag{4.16}$$

where $b_d(t)$ is the desired user's tth transmitted bit, $y_R(t)$ is the real part of the beamformer's noise-contaminated output and ρ_n is the radius of the kernel width or smoothening parameter [224, 225] of the density estimation process, as discussed in Section 3.4.1.

Our task is now to create an expression for the GA's fitness function that will correspond to the specific complex-valued array weight vector $\mathbf{w}$ that will minimise the true error probability P_E or the estimated probability of error $\hat{P}_E$ according to Equation (4.11) or (4.15), respectively. In our initial investigations of the achievable BER versus the GA complexity required we simply expressed our fitness function as:

$$f_p = 1 - P_E(\mathbf{w}_p), \tag{4.17}$$

where f_p is the fitness value of the pth individual representing the array weight vector $\mathbf{w}_p$. We observed, however that as the error probability decreases, the expression in Equation (4.17) becomes less efficient, because fairly similar fitness values were obtained for all the individuals resulting in a bit error ratio of $P_E \geq 10^{-3}$. For example, two individuals $P1$ and $P2$ having a probability of error of $P_{E,P1} = 10^{-3}$ and $P_{E,P2} = 10^{-7}$, respectively, are highly likely to be of the same priority, since the fitness value difference between $f_{P1} = 0.999$ and $f_{P2} = 0.9999999$ is rather small, namely $|f_{P2} - f_{P1}| = 0.0009999 \approx 0.001$.

Having briefly characterised the behaviour of the algorithm using the fitness function of Equation (4.17), we obtained an improved fitness function, which involves taking the logarithm of the probability of error P_E for the sake of expanding the fitness differences, as the error probability reduces. The new fitness function f_p of an individual can be expressed as [157]:

$$f_p = 1 - \frac{1}{1 - \log_{10} P_E(\mathbf{w}_p)} \,. \tag{4.18}$$

Both the fitness functions of Equations (4.17) and (4.18) approach unity as P_E decreases, but using the fitness function expressed in Equation (4.18) produces a more distinct fitness value difference between the individuals $P1$ and $P2$, namely $f_{P1} = 0.75$ and $f_{P2} = 0.875$. Equation (4.18) also outperforms other linear functions quantifying the relationship between the fitness f_p of an individual and the BER, particularly at low BERs [157].

In the context of vast search spaces and complex optimisation problems premature convergence has to be particularly avoided. In the following sections we will introduce both elitism and fitness scaling for the sake of avoiding this problem.

4.3.4 Elitism

Elitism is applied during the reproduction process. It refers to the preservation of the fittest individual or a group of the fittest individuals in an effort to maximise each generation's fitness. In a simple GA it is possible for the best individual of generation g to have a lower fitness value than the best individual of the previous generation $g-1$. Therefore in our case, we replace the least fit individual in a generation with the most fit individual of the previous generation. There are also other options of implementing elitism, one of which is based on replacing the fittest individual of the current generation $p_{best}(g)$ with

that of the previous generation $p_{best}(g-1)$, if it was more fit, i.e. if we have $f_{p_{best}}(g-1) > f_{p_{best}}(g)$, where f denotes the fitness value and g is the generation index. These indicate that elitism can be used for insuring that statistically speaking there is a monotonic increase in the best fitness f_{best} in the population as a function of the generation index.

4.3.5 Fitness Scaling

The procedure of selecting individuals to undergo crossover for the sake of creating new individuals is of paramount importance. Typically, the selection procedure is probabilistically biased towards individuals associated with high fitness values. This simple principle may however exhibit some undesirable properties. For a given optimisation function that is characterised by a relatively flat surface, the fitness of all individuals will be similar. However, regardless of the shape of the error surface, the specific choice of individuals becomes fairly random after the GA approached the optimum solution, i.e. at a high generation index g, when the population converged, and hence has numerous similar individuals. This is undesirable since many individuals associated with average fitness will have a similar chance to the best individuals for being selected for crossover. On the other hand, convergence to a local rather than global optimum may also occur, if we have a 'super-individual' exhibiting a very high fitness in a given population, particularly if this occurs at an early generation g.

For the sake of overcoming these problems, several fitness scaling techniques, such as windowing-based [247], exponential [247], linear [230–232], [247], sigma [230–232] and power scaling [230] have been introduced in [230–232], [247]. Generally speaking, a scaling procedure involves the alteration of the fitness value f_p, with the aim of promoting competition among the individuals. In the GA assisted MBER beamforming scheme, we incorporated sigma scaling [230–232] and also a new fitness scaling technique referred to as span scaling [157, 248], which will be described in the following subsections.

4.3.5.1 Sigma Scaling

Sigma scaling uses the populations' standard deviation σ and the populations' mean fitness $\bar{f}$ in modifying the individual's fitness, such that:

$$f'_p = f_p - (\bar{f} - c\sigma), \tag{4.19}$$

where f'_p is the modified fitness value and the coefficient c is usually chosen between 1 and 3. For a population having a high variance σ^2, i.e. for widely spread fitness values, the new fitness of the average individuals will be set to a large value for the sake of promoting their participation in the reproduction. On the other hand, the average individuals in a population having a low fitness variance, which may be encountered towards the end of a genetic run, are assigned small modified fitness values f'_p for the sake of differentiating them from the better-than-average individuals, so that the latter individuals may have higher chances of being selected for recombination.

Individual p	1	2	3	4
Fitness f_p	6.25	50	31.25	12.5
(%)	6.25%	50%	31.25%	12.5%
Sigma f'_p $(c = 2)$	15.48	59.23	40.49	21.73
(%)	11.31%	43.26%	29.57%	15.87%

Table 4.2: Example of scaling effects using sigma scaling [230–232] technique having a coefficient of $c = 2$.

Table 4.2 shows the effect of sigma scaling having a coefficient of $c = 2$. We use the same candidates of population size $P = 4$, when describing the roulette-wheel selection based method of Section 4.2.1. As a result of the sigma scaling applied to each fitness value according to Equation (4.19), it is observed in Table 4.2 that the influence of the strongest individual has been reduced. In case of encountering negative scaled fitness values, sigma scaling [230] will automatically assign them to zero.

4.3.5.2 Span Scaling

The new fitness scaling technique referred to as span scaling [157, 248] is expected to perform better in a particular configuration than the classic sigma scaling [230] as the generation index g increases. More specifically, span scaling maps the raw fitness f_p^g of an individual p of generation g according to [248]:

$$f_p'^g = \frac{f_p^g - \bar{f}^g}{\sigma_g} + C_g, \tag{4.20}$$

where we have [157]

$$C_g = 1 - \frac{1}{1 + \exp(-(\frac{10g}{G} - 4))}, \tag{4.21}$$

where $\bar{f}^g$ and σ_g are the average fitness and the standard deviation of the fitness of generation g, respectively. The function C_g in Equation (4.21) is a modified version of the one used in [248], which is characterised graphically in Figure 4.8. For a given generation index g the individuals having an average fitness value of $f_p = \bar{f}^g$ will be assigned a new fitness $f_p'^g$ value equal to the value of $C(g)$, according to $f_p'^g = C(g)$, as shown in Figure 4.8. Clearly, it is observed that during the early generations the average individuals are assigned relatively high fitness values, so that they have similar chances to the fit individuals to undergo crossover. However, as the population becomes saturated with similar-fitness individuals at higher generation indices, the average individuals are assigned relatively low new fitness values, so that they do not prevent the above-average individuals' competition for being selected for crossover.

In comparison to other scaling techniques, the adaptive span scaling regime of Equation (4.20) has the advantage of forcing the GA to converge to a solution after a fixed number of generations by

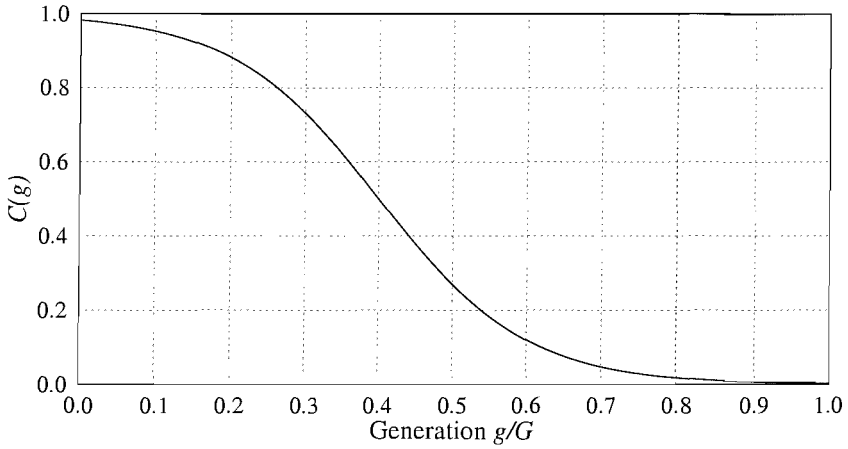


Figure 4.8: The span-scaling function $C(g)$ versus the generation index g.

increasing the selection 'pressure', as the generation index g increases. Let us next outline how we determine the convergence of our GA-assisted MBER beamforming scheme and its complexity.

4.3.6 Convergence and Complexity

We opted for using the maximum number of generations as the termination criterion of the algorithm. This has the advantage of maintaining a constant complexity. The statistical behaviour of the algorithm will be characterised in terms of the estimated probability density function (PDF) of the BER at the beamformer's output, rather than simply portraying the average BER. More specifically, this PDF is approximated with the aid of the BER histogram, which is constructed from 1000 independent GA runs.

The complexity of the GA is characterised by the number of objective function evaluations encountered, which is proportional to the product of the population size P and the maximum number of generations G, expressed as:

$$\text{Complexity} \sim P \cdot G. \tag{4.22}$$

For algorithms having the same total complexity, generally the one having a higher population size is preferable, since in a particular generation, the P number of objective function evaluations can be processed in parallel. We shall next characterise the achievable performance of the GA-assisted MBER beamforming scheme.

Population size	10 to 50
Generation size	1 to 50
Bits per complex array weight	16
Selection type	roulette-wheel
Single-point crossover probability	0.7, 0.9
Bit-inversion mutation probability	0.001, 0.01, 0.1
Additional operations	sigma scaling span scaling elitism
Fitness function f_p	$F1: f_p = 1 - P_E(\mathbf{w}_p)$ (4.17) $F2: f_p = 1 - \frac{1}{1-\log_{10} P_E(\mathbf{w}_p)}$ (4.18)

Table 4.3: Summary of the GA configurations and the range of values considered.

4.4 Simulation Results

In order to generate attractive GA configurations, we varied the associated GA parameters, such as the crossover probability p_c, mutation probablity p_m, population size P and the number of generations G, according to the values given in Table 4.3. The additional operations of elitism, sigma- and span scaling have been described in Sections 4.3.4, 4.3.5.1 and 4.3.5.2, respectively. For the sigma scaling technique, which uses the populations' standard deviation σ and the populations' mean fitness $\bar{f}$ for modifying an individual's fitness according to $f'_p = f_p - (\bar{f} - c\sigma)$, we fixed the coefficient c to unity. The search space of the GA was limited to the region of $|\Re\{w_l\}| \leq 1$ and $|\Im\{w_l\}| \leq 1$, where $1 \leq l \leq L$ is the number of antenna array elements, since the MBER solution found for BPSK modulated users is invariant to a positive linear scaling and thus can be confined to this search space. The initial population was randomly chosen in conjunction with a uniform probability within the given search space. In our simulations we limited the number of bits used for representing each signed real-valued array coefficient $\Re\{w_l\}$ and $\Im\{w_l\}$ to $J = 8$ bits, resulting in a maximum of $2^{J \cdot 2L} = 2^{16L}$ possible individuals. For $L = 2$, there will be approximately $4.29 \cdot 10^9$ legitimate individuals, although naturally, only a fraction of this solution space will be searched.

As in Chapter 3, we used the beamforming scenario of Figure 3.18, which consists of five users supported by a two-element antenna array having an element spacing of $\lambda/2$. The desired user is at an angular direction of 15^o and the four interfering users are at directions $-30^o, 60^o, 80^o$ and -70^o with respect to the array normal, which are assumed to have an equal power of 10dB, i.e. we have SNR = INR_i = 10dB for $i = 2, 3, 4, 5$, unless otherwise stated. In some cases we varied the direction of the nearest interferer, namely that of user 2 and user 3 for the sake of further characterising the GA-assisted MBER beamforming performance, when the desired user is very close in terms of its angular separation to the interferers. We list the scenarios considered in Table 4.4.

In the first stage of our simulations we used Equation (4.17) as the fitness function, which we

User	1	2	3	4	5
Scenario 1	15^o	-30^o	60^o	80^o	-70^o
Scenario 2	15^o	-7.5^o	37.5^o	80^o	-70^o
Scenario 3	15^o	-3.75^o	26.25^o	80^o	-70^o
Scenario 4	15^o	-9.375^o	20.625^o	80^o	-70^o
Scenario 5	15^o	-12.1875^o	17.8125^o	80^o	-70^o

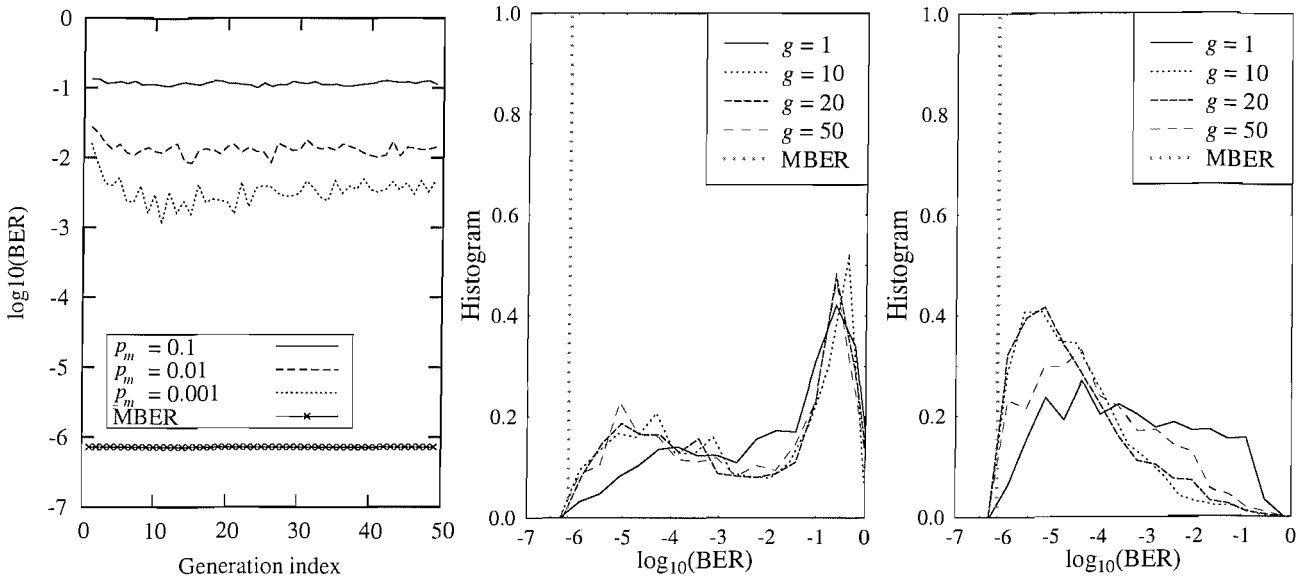
Table 4.4: Several angular direction scenarios for the five users considered.

refer to as fitness function $F1$. We characterised the attainable performance of the GA both with the aid of the BER histogram and the average value of the achievable BER at the beamformer's output. Again, the histogram was constructed from 1000 randomly initialised GA runs. The fitness function of Equation (4.18) is referred to as fitness function $F2$. The results generated will also be compared to the theoretical bounds of the MBER solution obtained by using the simplified conjugate gradient algorithm of Section 3.5.2. Instead of using the perfect MBER assumption according to Equation (4.11) for Section 4.4.1, in Section 4.4.2 we will proceed by determining the best achievable BER performance of the GA using the objective function based on the estimated BER of Equation (4.15).

4.4.1 GA-Assisted MBER Beamforming Upper-Bound Performance

Figure 4.9 (a) shows the average BER achieved with the aid of the GA having a population size of $P = 30$, as the generation index g increases from 1 to 50. As expected, it is observed that the average BER decreases, as we reduce the mutation size p_m from 0.1 to 0.001, but none of these scenarios was capable of approaching the theoretical MBER solution of BER $= 10^{-6.14} \approx 7.24 \cdot 10^{-7}$. Let us consider the BER histogram associated with the mutation probabilities of $p_m = 0.1$ and $p_m = 0.001$ shown in Figures 4.9 (b) and 4.9 (c), respectively. For the GA configured for a mutation probability of $p_m = 0.1$ in Figure 4.9 (b), most of the recorded BERs were of high values, centred approximately around BER $= 10^{-0.8} \approx 1.58 \cdot 10^{-1}$, regardless of the number of generations. By contrast, when a significantly lower mutation probability of $p_m = 0.001$ is used, some improvement of the BER histogram may be observed. As seen in Figure 4.9 (c), we have a more attractive BER histogram, exhibiting a peak shifted towards lower BER values around BER $= 10^{-5}$ at a generation index of $g = 10$ and $g = 20$, but nonetheless, the average BER is still relatively high, around BER $= 10^{-2.5} \approx 3.16 \cdot 10^{-3}$. It is also observed that as the generation index g increases beyond $g > 10$, there is no further improvement of the BER histogram. In fact, for $G = g = 50$ number of generations, the histogram becomes flatter, concentrating towards higher BERs again, as shown by the plot associated with $g = 50$ in Figure 4.9 (c). This phenomenon shows signs of premature convergence, which may be due to having 'super-individuals' in the population in the low generation index region of $g \leq 10$.

The BER histogram recorded for the mutation probability of $p_m = 0.01$ may be observed in Figure 4.10 (b). As in the previous scenario of $p_m = 0.001$ characterised in Figure 4.9, this choice of



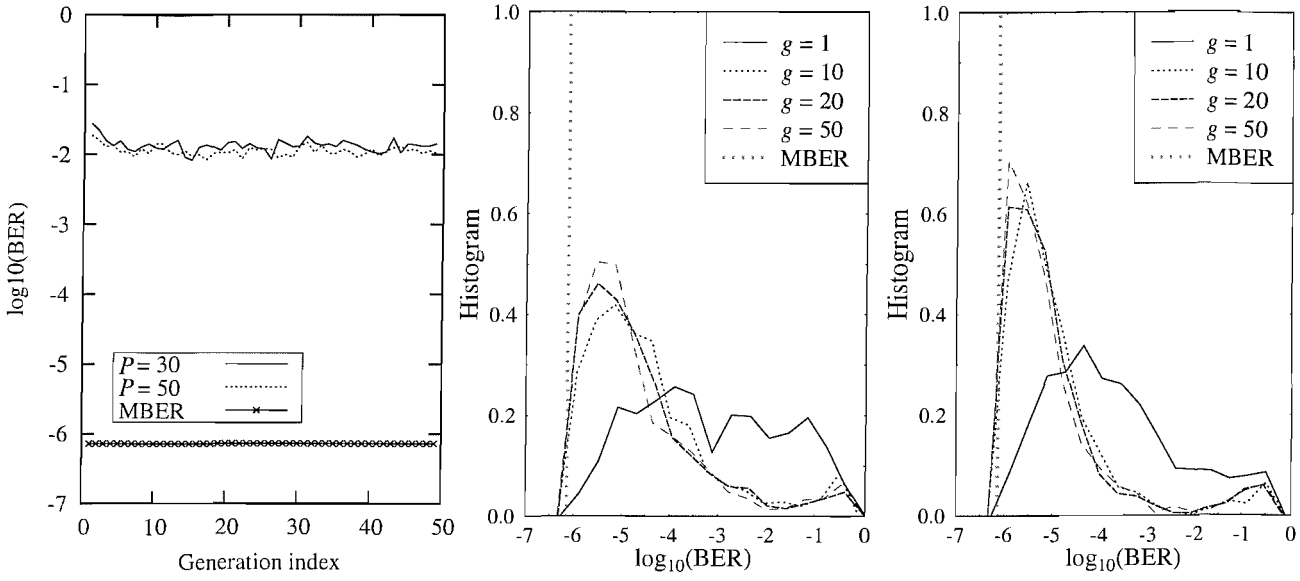
(a) $P = 30$ (b) $p_m = 0.1$ (c) $p_m = 0.001$

Figure 4.9: (a) The average BER of the GA-aided beamformer, based on Equation (4.11), versus the generation index g, for $1 \leq g < 50$ and for the mutation probability values of $p_m = 0.1, 0.01, 0.001$, in conjunction with a GA using the fitness function $F1$[4], roulette-wheel selection having a population size of $P = 30$ and a crossover probability of $p_c = 0.9$. Also shown is the associated BER histogram of the GA used in (a) employing the mutation probabilities of (b) $p_m = 0.1$ and (c) $p_m = 0.001$. The BER values were averaged over 1000 randomly initialised GA runs for transmission over the AWGN channel, for five equal-power users, each having SNR = 10dB, and the angular directions specified by Scenario 1 of Table 4.4.

p_m does not dramatically affect the shape of the BER histogram in conjunction with different number of generations. Increasing the population size to $P = 50$ does not substantially improve the average BER either, as seen from Figure 4.10 (a). We may surmise that the limited performance of the system is a consequence of the deficient choice of the fitness function $F1$ of Equation (4.17). As noted earlier in Section 4.3.3, it may be inferred that the fitness function $F1$ formulated as $f_p = 1 - P_E(\mathbf{w}_p)$ becomes ineffective for BER values below BER $= 10^{-3}$. Before we consider the more promising fitness function $F2$, let us briefly characterise the effect of incorporating scaling or/and elitism, as shown in Figures 4.11 and 4.12.

Figures 4.11 (b) and 4.11 (c) show the BER histogram of the GA-assisted MBER beamforming scheme using the sigma scaling algorithm of Section 4.3.5.1 in conjunction with $c = 1$ and elitism, respectively. No significant average BER improvement is observed, when employing the sigma scaling technique, although the associated histogram of Figure 4.11 (b) indicates some improvement in com-

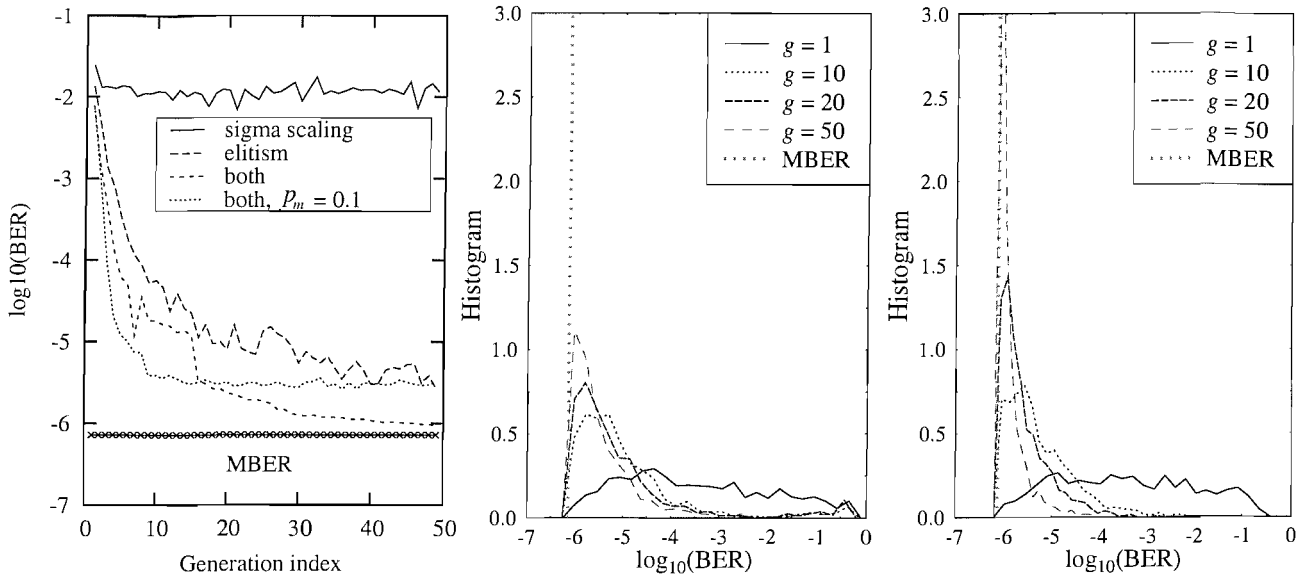
[4]The exact BER calculated based on Equation (4.11), was used in the objective function $F1$ of the GA-assisted MBER approach described by Equation (4.17).



(a) $p_m = 0.01$ (b) $P = 30$ (c) $P = 50$

Figure 4.10: (a) The average BER of the GA-aided beamformer, based on Equation (4.11), versus the generation index g, for $1 \leq g < 50$ and for population sizes of $P = 30, 50$, in conjunction with a GA using the fitness function $F1$, roulette-wheel selection and having a mutation probability of $p_m = 0.01$ and a crossover probability of $p_c = 0.9$. Also shown is the associated BER histogram PDF of the GA used in (a) having population sizes of (b) $P = 30$ and (c) $P = 50$. The BER values were averaged over 1000 randomly initialised GA runs for transmission over the AWGN channel, for five equal-power users, each having SNR = 10dB, and the angular directions specified by Scenario 1 of Table 4.4.

parison to the basic GA characterised in Figures 4.9 and 4.10. This may be a consequence of having a relatively high mutation probability of $p_m = 0.01$, which is capable of mutating the meritorious individuals, thus preventing them from surviving to the next generation. By contrast, performing elitism shows more significant improvements in comparison to all earlier measures, since elitism guarantees that the most fit invidual of a generation survives to the next generation. In this way elitism has the potential of preserving the high-fitness individuals for the sake of producing even better individuals. As seen from Figure 4.11 (a), the average BER values become slightly less than BER $= 10^{-5.5} \approx 3.16 \cdot 10^{-6}$, as we increase the number of generations to $G = g = 50$. The convergence of the GA was further improved, when we combined elitism with sigma scaling. The average BER is seen to approach the theoretical MBER solution at BER $= 10^{-6.14} \approx 7.24 \cdot 10^{-7}$. Explicitly, as shown in Figure 4.11 (a), we observed that the average BER associated with a mutation probability of $p_m = 0.1$ converges faster, namely as early as at the generation index of $g = 10$, but it settles at a higher average BER value of BER $= 10^{-5.5} \approx 3.16 \cdot 10^{-6}$ than that when having a lower mutation probability of $p_m = 0.01$, which settles near BER $= 10^{-6}$ after 50 generations. Figures 4.12 (a), 4.12 (b) and 4.12 (c) demonstrate the sequence of PDF improvements, as we increase the population size from $P = 20$ to $P = 30$ and

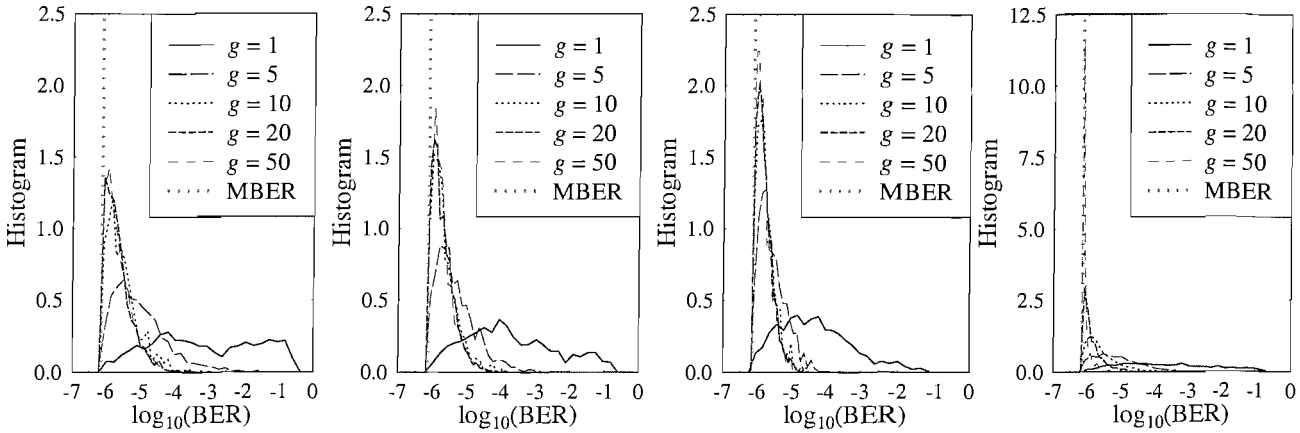


(a) $p_m = 0.01$, $P = 30$ (b) sigma scaling (c) elitism

Figure 4.11: (a) The average BER of the GA-aided beamformer, based on Equation (4.11), versus the generation index g, for $1 \leq g < 50$ and for a mutation probability of $p_m = 0.01$, in conjunction with a GA using the fitness function $F1$, roulette-wheel selection as well as a population size of $P = 30$, a crossover probability of $p_c = 0.9$ and incorporating sigma scaling associated with $c = 1$ or/and performing elitism. Also shown is the associated BER histogram of the basic GA used in (a) employing a mutation probability of $p_m = 0.01$; (b) fitness value f_p scaled according to sigma scaling using $c = 1$ and (c) elitism. The BER values were averaged over 1000 randomly initialised GA runs for transmission over the AWGN channel, for five equal-power users, each having SNR = 10dB and the angular directions specified by Scenario 1 of Table 4.4.

$P = 50$, respectively. Comparing Figures 4.12 (c) and 4.12 (d), it is observed that Figure 4.12 (d) has a spike-like PDF at a generation index of $g = 50$, but the histogram peak at lower generation indices of $g = 5$ and $g = 10$ remains lower than that of Figure 4.12 (c), which is likely to be the reason why we observed early convergence of the $p_m = 0.1$ curve in Figure 4.11 (a).

Let us now consider the achievable BER performance, when employing the fitness function $F2$ of Equation (4.18). The performance attained with the aid of the fitness function $F2$ using the GA configuration of Figure 4.12 is portrayed in Figure 4.13. In all plots of Figures 4.13 (a) to 4.13 (d), the BER histogram difference becomes explicit in comparison to the respective plots of Figures 4.12 (a) to 4.12 (d), in terms of both of its height and width. At a particular generation index in the range of $g \geq 5$, the width of the PDF is reduced to approximately half of that associated with the previous fitness function $F1$, while the histogram becomes higher, at least twice higher than in the corresponding plot of Figure 4.12. A significantly lower complexity is required for the sake of attaining a similar PDF shape, when using the fitness function $F2$. For example, it is seen that in Figure 4.13 (a) the

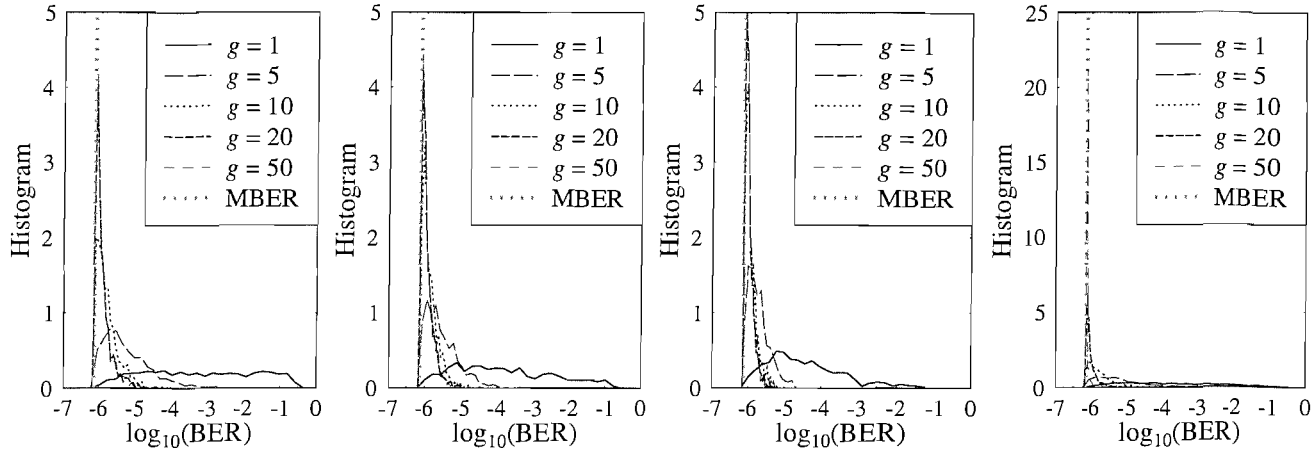


(a) $p_m = 0.1$, $P = 20$ (b) $p_m = 0.1$, $P = 30$ (c) $p_m = 0.1$, $P = 50$ (d) $p_m = 0.01$, $P = 30$

Figure 4.12: The BER histogram of the GA-aided beamformer, evaluated based on Equation (4.11), using the fitness function $F1$, roulette-wheel selection, elitism, sigma scaling employing $c = 1$, having a crossover probability of $p_c = 0.9$ and (a) a population size of $P = 20$ as well as a mutation probability of $p_m = 0.1$; (b) a population size of $P = 30$ and mutation probability of $p_m = 0.1$; (c) a population size of $P = 50$ and mutation probability of $p_m = 0.1$ as well as (d) a population size of $P = 30$ and a mutation probability of $p_m = 0.01$. The BER histogram for each of the $G = g$ number of generations was constructed from 1000 randomly initialised GA runs for transmission over the AWGN channel, for five equal-power users, each having SNR = 10dB and the angular directions specified by Scenario 1 of Table 4.4.

PDF of a population size of $P = 20$ having $G = 10$ generations is similar to that of the population size of $P = 50$ having $G = 20$ generations in Figure 4.12 (c), despite reducing the complexity defined in Equation (4.22) from 1000 to 200 objective function evaluations.

Figure 4.14 (a) portrays the gradual improvement of the average BER at the beamformer's output, as we incorporate several additional parameters in the GA's configuration such as sigma scaling and elitism, when using the fitness function $F2$, for a population size of $P = 30$ employing a mutation probability of $p_m = 0.01$ and a crossover probability $p_c = 0.9$. We also included the BER curve for the population size of $P = 50$ for the sake of showing the effect of using a larger population size. It is observed in Figure 4.14 (a) that the elitism operation has a beneficial effect on the GA's performance. Exploiting the configurations considered in Figure 4.14 (a), in Figure 4.14 (b) we also plotted the BER variance as a function of the affordable complexity, defined as the product of the population size P and the number of generations G given in Equation (4.22). Apart from the basic GA configuration, which employed neither scaling nor elitism and with the exception of the GA using sigma scaling associated with $c = 1$, all the BER variance values seen in Figure 4.14 (b) substantially decrease, as the complexity is increased. It is observed that the GA employing elitism and using the fitness function $F1$ is capable of converging to the MBER solution, although at a slower rate, hence requiring a higher complexity. For our particular case of supporting five users as specified in Scenario 1 of

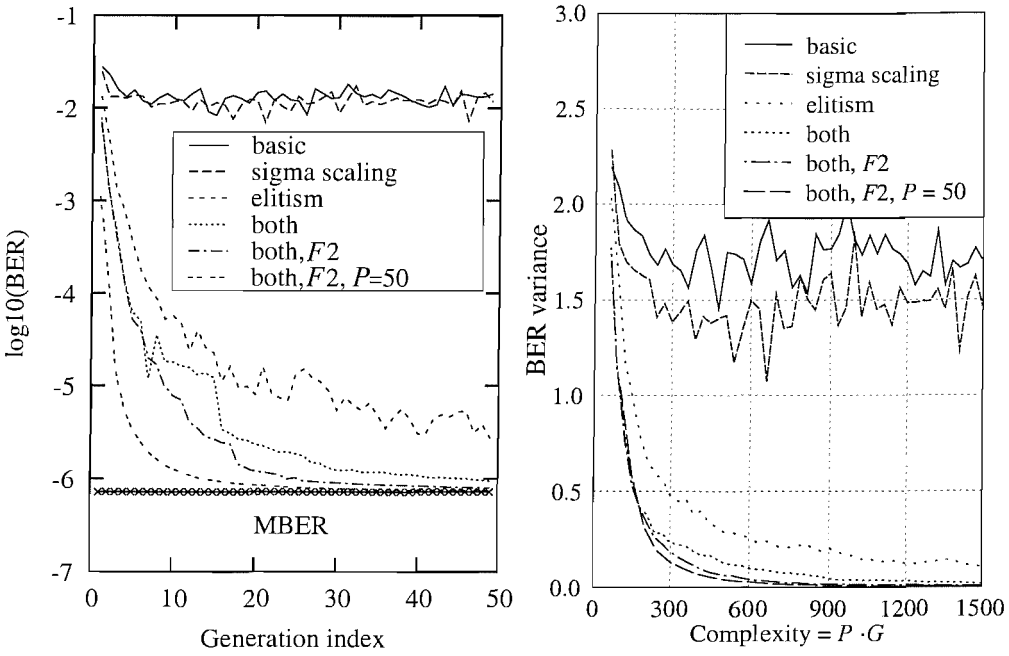


(a) $p_m = 0.1$, $P = 20$ (b) $p_m = 0.1$, $P = 30$ (c) $p_m = 0.1$, $P = 50$ (d) $p_m = 0.01$, $P = 30$

Figure 4.13: The BER histogram of the GA-aided beamformer, evaluated based on Equation (4.11), using the fitness function $F2$, roulette-wheel selection, elitism, sigma scaling employing $c = 1$, having a crossover probability of $p_c = 0.9$ and (a) a population size of $P = 20$ as well as a mutation probability of $p_m = 0.1$; (b) a population size of $P = 30$ and a mutation probability of $p_m = 0.1$; (c) a population size of $P = 50$ and a mutation probability of $p_m = 0.1$ as well as (d) a population size of $P = 30$ and mutation probability of $p_m = 0.01$. The BER histogram for each of the $G = g$ number of generations was constructed from 1000 randomly initialised GA runs for transmission over the AWGN channel, for five equal-power users, each having SNR = 10dB and the angular directions specified by Scenario 1 of Table 4.4.

Table 4.4 having $G = 50$ generations, a small population size of $P \approx 30$ is sufficient for achieving both rapid convergence and a reasonably good steady-state BER performance, when using the fitness function $F2$. Let us now consider, how we can further improve the GA for the sake of having a lower complexity without sacrificing the attainable BER performance.

According to the complexity measure defined in Equation (4.22), the GA's complexity will increase as the number of generations G increases. In order to achieve a performance advantage as a consolation for the increased complexity, an attractive approach is to invoke a so-called scaling technique [230–232, 248] in such a way that the scaling function involves the generation index g. The span scaling [157,248] technique introduced in Section 4.3.5.2 exhibits this desirable characteristic, which has the potential of forcing the GA to converge to a solution after a fixed number of generations by increasing the GA's selection pressure, as the generation index g increases. More explicitly, the selection pressure refers to scaling the individuals' fitness values f_p according to the population's variance at a particular generation. As mentioned earlier in Section 4.3.5, we prefer that the average individuals compete with the fit individuals during the early generations, but we would like to separate them later, as the generation index g approaches the total affordable number of generations G specified. Figures 4.15 (a) and 4.15 (b) show the BER variance of the GA-aided beamformer configured for using the mutation probabilities of $p_m = 0.01$ and $p_m = 0.1$, respectively. The GA-assisted beamformer's performance



(a) $p_m = 0.01, P = 30$ (b) $p_m = 0.01, P = 30$

Figure 4.14: (a) The average BER of the GA-aided beamformer, evaluated based on Equation (4.11), versus the generation index g, for $1 \leq g < 50$ and for a mutation probability of $p_m = 0.01$, in conjunction with a GA using the fitness function $F1$, roulette-wheel selection having a population size of $P = 30$ and a crossover probability of $p_c = 0.9$. Also shown in (b) is the associated BER variance of the respective GA used in (a), as a function of the associated complexity. The BER values were averaged over 1000 randomly initialised GA runs for transmission over the AWGN channel, for five equal-power users, each having SNR = 10dB and the angular directions specified by Scenario 1 of Table 4.4.

is evaluated according to the fitness function $F2$, in conjunction with the population sizes of $P = 20, 30, 50$, a crossover probability of $p_c = 0.9$ and employing elitism in conjunction with either sigma- [230–232] or span scaling [157, 248]. Apart from the population size of $P = 30$ characterised in Figure 4.15 (a), all the other plots in Figures 4.15 (a) and 4.15 (b) show that the BER variance associated with the GA-aided beamformer employing span scaling is reduced faster or at least as fast as that of the GA employing sigma scaling. Although the span scaling based plot associated with the population size of $P = 30$ in Figure 4.15 (a) converges at a lower rate initially, the curve crosses the respective plot associated with sigma scaling at a complexity of approximately 450 objective function evaluations, beyond which it settles at a lower BER variance value of 0.01 after 800 objective function evaluations. Upon comparing Figures 4.15 (a) and 4.15 (b), it may be observed that the GA using span scaling performs best, when employing a higher mutation probability of $p_m = 0.1$, rather than $p_m = 0.01$, while the reverse is true for the GA using sigma scaling. Therefore, in Figure 4.15 (c) we plotted the average BER value for the respective GA using either span- or sigma

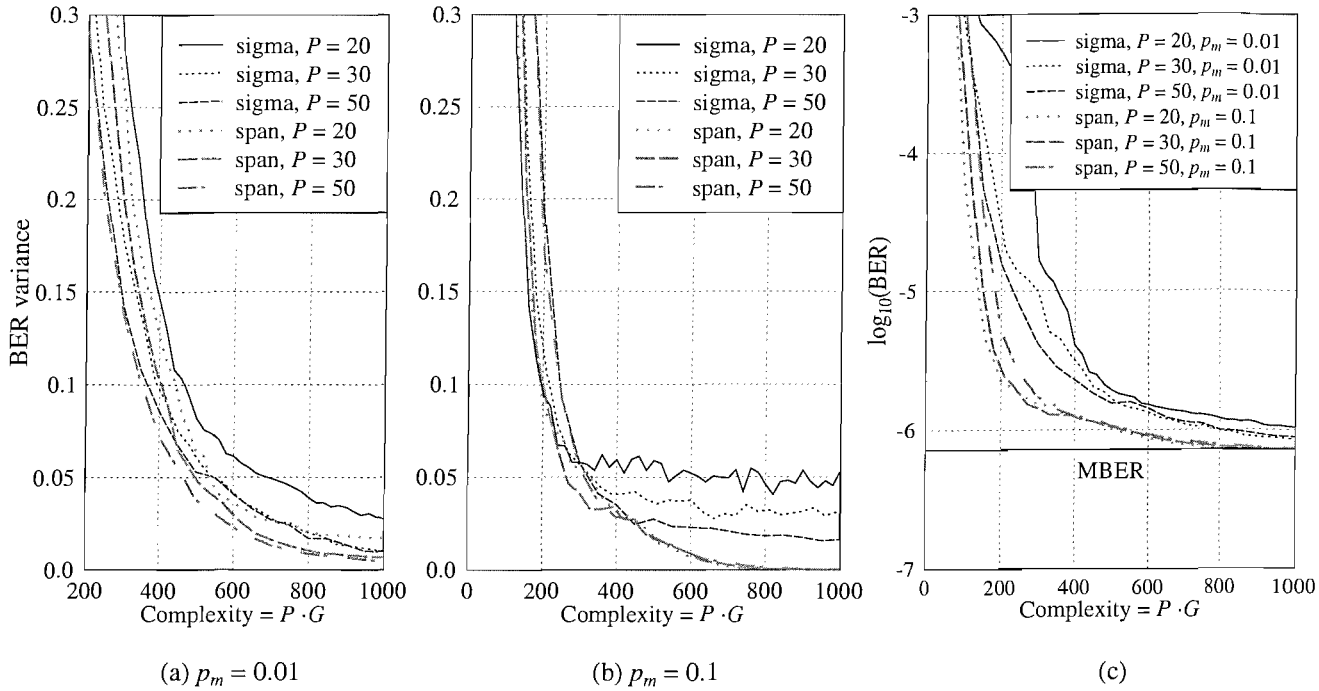
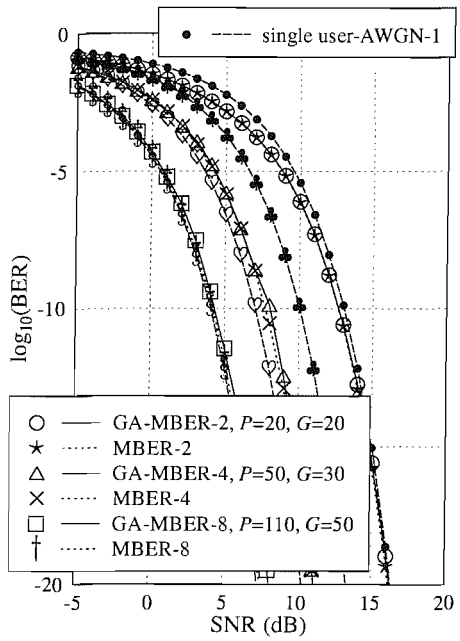


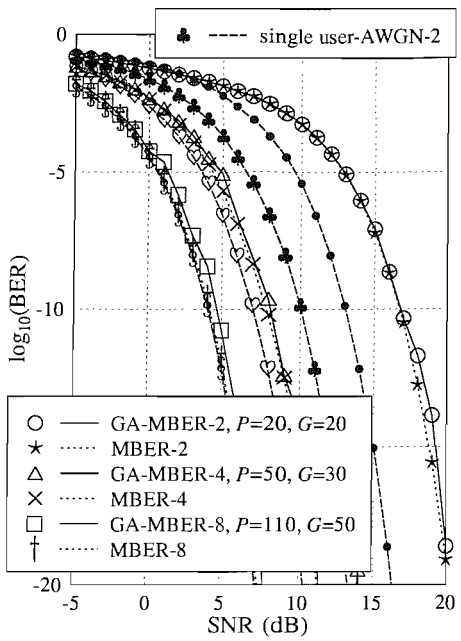
Figure 4.15: The BER variance of the GA-aided beamformer, evaluated based on Equation (4.11), for a mutation probability of (a) $p_m = 0.01$ and (b) $p_m = 0.1$, in conjunction with a GA using the fitness function $F2$[5], roulette-wheel selection having a population size of $P = 20, 30, 50$, a crossover probability of $p_c = 0.9$ and employing elitism as well as either sigma- or span scaling. Also shown in (c) is the associated average BER of the respective GA using the sigma scaling associated with $c = 1$ having $p_m = 0.01$ in (a) and that using the span scaling having $p_m = 0.1$ in (b), as a function of the associated complexity. The BER values were accumulated over 1000 randomly initialised GA runs for transmissions over the AWGN channel, for five equal-power users, each having SNR = 10dB, and the angular directions specified by Scenario 1 of Table 4.4.

scaling, employing their optimum mutation probability of $p_m = 0.1$ and $p_m = 0.01$, respectively, as a function of the affordable complexity. It is observed that in the best-case scenario considered the GA-aided beamformer employing span scaling outperforms that using sigma scaling at a lower complexity. All the BER curves associated with span scaling are below the corresponding BER curves associated with sigma scaling beyond the complexity value of 100 objective function evaluations. Figure 4.15 (c) also shows that all GA configurations employing span scaling are capable of approaching the exact theoretical MBER bound of BER $= 10^{-6.14} \approx 7.24 \cdot 10^{-7}$ at a maximum complexity of just above 800 objective function evaluations. Observe also in Figure 4.15 (c) that all the span scaling related curves produce a similar average BER performance beyond a complexity of approximately 380 objective function evaluations.

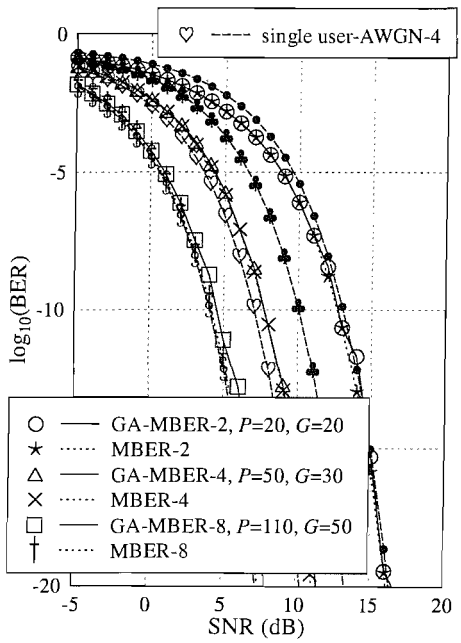
[5]The exact BER calculated based on Equation (4.11), was used in the objective function $F2$ of the GA-assisted MBER approach described by Equation (4.18).



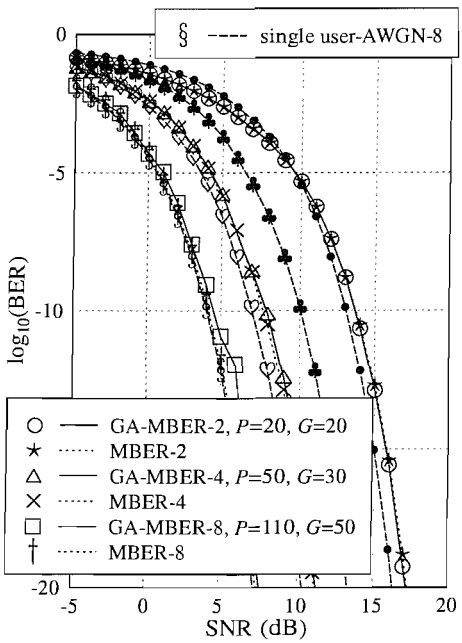
(a) SNR = INR_i for $i = 2, 3, 4, 5$



(b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$



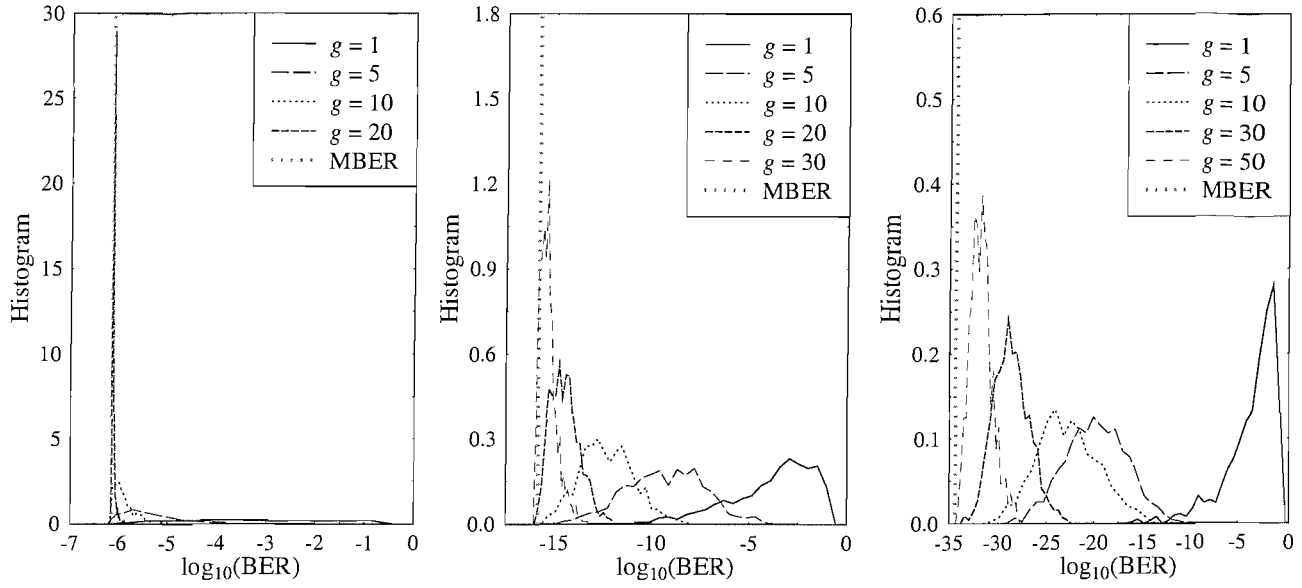
(c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB



(d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB

Figure 4.16: BER performance of the GA-assisted MBER beamformer, based on Equation (4.11), under four different conditions; (a) SNR = INR_i for $i = 2, 3, 4, 5$, (b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$, (c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB for two-, four- and eight-element uniform linear arrays. The GA employed was using the roulette-wheel selection, elitism, span scaling, the fitness function $F2$ and having a mutation as well as a crossover probability of $p_m = 0.1$ and $p_c = 0.9$, respectively. The associated two-, four- and eight-element antenna arrays were configured to have a complexity of $P \cdot G \sim 400, 1500$ and 5500, respectively. These results were recorded for transmission over the AWGN channel for five users and for the angular directions of 15^o (SOI), $-30^o, 60^o, 80^o, -70^o$, as also specified according to Scenario 1 in Table 4.4.

Let us next observe the BER performance of the GA-assisted MBER beamformer as the SNR of the users increases. Figure 4.16 shows the achievable BER performance at the beamformer's output for the GA-assisted MBER solution in comparison to that of the exact closed-form-solution based simplified conjugate gradient MBER technique of Section 3.2.1, i.e. in comparison to the MBER theoretical bound, under four different conditions for Scenario 1 of Table 4.4; (a) the desired user and all the four interfering sources have an equal power; (b) all the interfering sources have 6dB higher power than the desired user; (c) the desired user and the interfering sources $3, 4, 5$ have an equal power, while the interfering source 2 has 6dB higher power than the desired user; and (d) the desired user and the interfering sources $2, 3, 4$ have an equal power, but the interfering source 5 has 6dB higher power than the desired user. It can be seen in Figure 4.16 that the BER performance attained by the GA-assisted two-element antenna array is identical to the MBER theoretical bound for all the scenarios considered in Figures 4.16 (a) to 4.16 (d) for at least up to SNR = 11dB. For a four-element antenna array, the possible number of individuals, i.e. the search space expands quadratically from that of the two-element array having $2^{J \cdot 2L} = 2^{16L} = 4.29 \cdot 10^9$ legitimate solutions to $1.84 \cdot 10^{19}$. The increased search space indicates that the complexity of the GA required for the sake of arriving at the global optimum solution will definitely be higher. It is seen for the two-element, equal-power user scenario of Figure 4.16 (a), that the GA having a total complexity of 400 is capable of approaching the theoretical MBER bound up to SNR = 13dB, achieving a low average BER of approximately BER = 10^{-11}. Focussing our attention on the SNR range up to SNR = 10dB, it is also possible for the GA to approach a similar BER performance to that seen in Figure 4.16 (a), given a slightly lower complexity than 400 objective function evaluations. Hence, instead of imposing a GA complexity of $4 \cdot 400 = 1600$ objective function evaluations for the four-element antenna array, we opt for a slightly lower complexity of 1500 objective function evaluations, constituted by the population size of $P = 50$ and $G = 30$ generations. It is observed for the equal-power scenario of Figure 4.16 (a), that identical solutions are achievable up to a maximum of SNR = 7dB. Although for SNRs higher than 7dB the BER performance shows a slight degradation, it is still close to the MBER theoretical bound, where we have BER = 10^{-10}. At certain SNR values seen in Figures 4.16 (b) to 4.16 (d) there are deviations from the theoretical MBER bound, but this artifact may be eliminated by increasing the affordable complexity, i.e. the population size P and/or the number of generations G. However, bearing in mind that the beamformer weight values optimised by the GA are encoded or quantised values, thus they have a limited precision. The beamformer weights are encoded according to the specific number of bits used for representing an individual, where the array-weight precision was calculated in Section 4.3.2 to be $2^{J-1} = 2^{-7} = 1/128 = 0.0078125$, since we assign $J = 8$ bits for each real-valued beamformer weight. In all plots of Figures 4.16 (a) to 4.16 (d), we used the same population size P and the same number of generations G in conjunction with a given size of the antenna array for the sake of showing that by using the same GA complexity as in the equal-power scenario of Figure 4.16 (a) the GA is still capable of producing meritorious MBER solutions, regardless of the power differences of the users



(a) 2-element, $P = 20$, $G = 20$ (b) 4-element, $P = 50$, $G = 30$ (c) 8-element, $P = 110$, $G = 50$

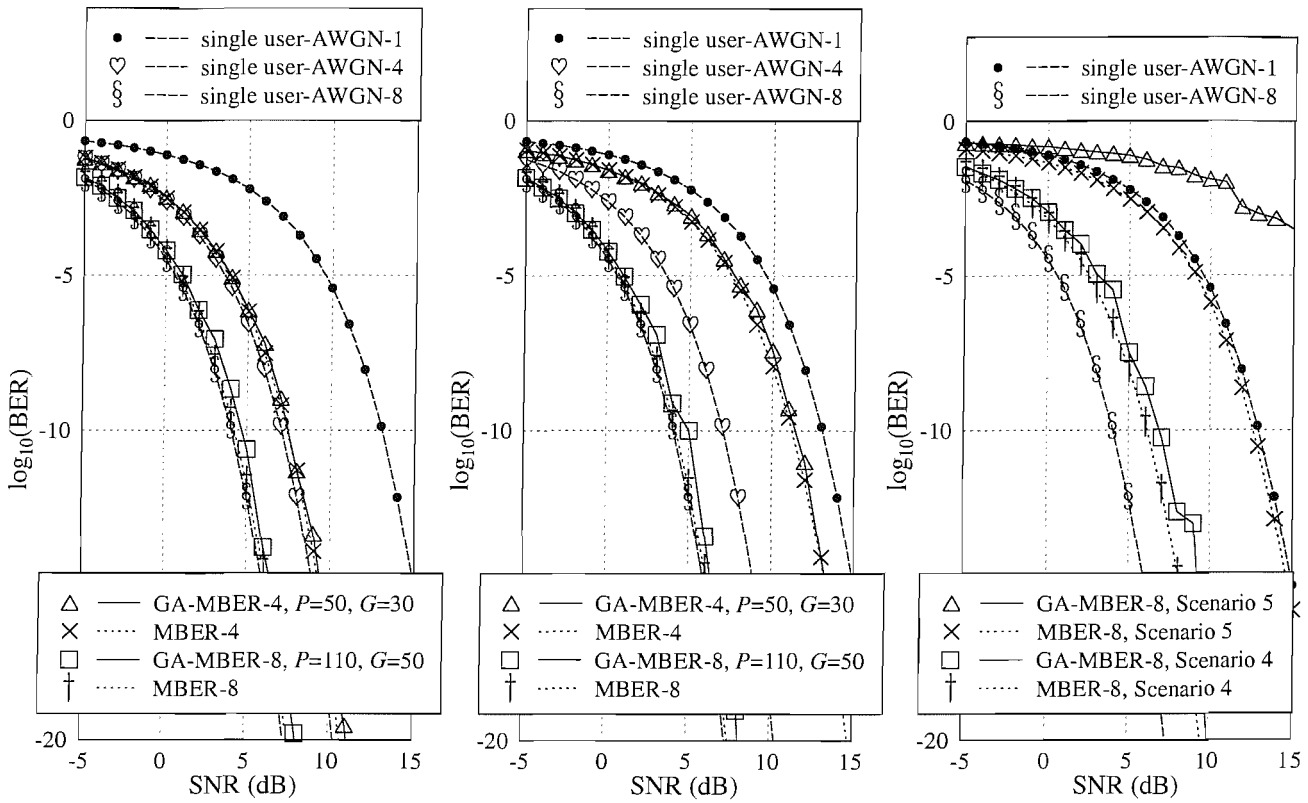
Figure 4.17: The BER histogram of the GA-aided beamformer, evaluated based on Equation (4.11), using the fitness function $F2$, roulette-wheel selection, elitism, span scaling and having a crossover as well as a mutation probability of $p_c = 0.9$ and $p_m = 0.1$, respectively, for (a) a population size of $P = 20$ and $G = 20$ generations for a two-element antenna array; (b) a population size of $P = 50$ and $G = 30$ generations for a four-element antenna array and (c) a population size of $P = 110$ and $G = 50$ generations for an eight-element antenna array. The BER histogram recorded for each of the generations $G = g$ was constructed from 1000 randomly initialised GA runs for transmission over the AWGN channel, for five equal-power users, each having SNR = 10dB, and for the angular directions specified by Scenario 1 of Table 4.4.

supported. Apart from the single-user, single-element BER curve, the single-user curves for the two-, four- and eight-element antenna arrays are also shown in Figures 4.16 (a) to 4.16 (d), which represent the best possible BER performance.

Figures 4.17 (a), 4.17 (b) and 4.17 (c) show the BER histogram of the GA-assisted MBER beamformer characterised in Figure 4.16 (a), when all users operate at SNR = 10dB, for two-, four- and eight-element antenna arrays, respectively, for the sake of characterising the convergence behaviour of the GA-aided solutions at generation indices of $g = 1, 5, 10, 20$ and for the final generations at $g = 30$ and $g = 50$, for the four- and eight-element case, respectively. Figure 4.17 demonstrated that the GA approaches the theoretical MBER bound, as the generation index g increases. For the two-element antenna array of Figure 4.17 (a), the BER histogram is particularly beneficial, when the generation index g reaches $G = 20$. Comparing the BER histogram of other antenna array configurations in Figures 4.17 (b) and 4.17 (c) to their corresponding counterparts in Figure 4.16 (a), we concluded that we may slightly reduce the complexity imposed for this two-element array to 320 or 360 objective function

evaluations, i.e. we may employ $G = 16$ or $G = 18$, while still attaining a similar BER performance, respectively. However, this may affect the MBER performance observed at higher SNRs in the signal power scenarios of Figures 4.16 (a), 4.16 (b), 4.16 (c) and 4.16 (d). Note also for the eight-element antenna array that instead of using $4 \cdot 1500 = 6000$, we employed a total of $110 \cdot 50 = 5500$ objective function evaluations, since at this complexity no significant BER performance degradation has been observed. It is also seen from Figure 4.17, that the BER histogram of the four- and eight-element scenario is not as meritorious as that of the two-element case seen in Figure 4.17 (a). Nonetheless, the GA is still capable of approaching the theoretical MBER performance bound. Apart from the height of the BER histogram, its width also is taken into account in characterising the achievable BER. For the eight-element array of Figure 4.17 (c), the BER histogram is not as attractive as that of the four- or two-element array, but nonetheless, the BER may approach BER $= 10^{-30}$. Naturally, the resultant average BER value is higher than the MBER theoretical bound, which is approximately BER $= 10^{-35}$ at SNR $=$ 10dB. Observation of Figure 4.16 suggests that the GA is capable of reaching a reasonable BER, even if it does not approach the ideal MBER solution, as shown by the BER histograms of Figure 4.17, especially for the four- and eight-element antenna array scenarios. We also observed that the complexity of the GA required for approaching the MBER solution increases approximately quadratically with the array size L.

Having varied the users' signal power, next we investigate the GA-assisted MBER beamformer's performance having the same complexity as shown in Figure 4.16, in conjunction with varying the angular separation between the desired user located at 15^o with respect to the array normal and the nearest interfering users, namely user 2 and user 3. In these investigations, user 2 and user 3 are gradually moved towards the desired user in equal angular steps, namely by repeatedly halving their angular separation, as tabulated in the columns corresponding to user 2 and user 3 of Table 4.4. Based on the results shown in Figure 3.20 of Section 3.5.2, we leave out the scenarios, where the MBER beamformer was deficient, particularly the two-element array. As seen in Figures 4.18 (a) and 4.18 (b), the GA-assisted BER performance of the four-element array may be considered attractive, although it is not identical to the theoretical MBER solution using the simplified conjugate gradient algorithm of Chapter 3. In Figure 4.18 (a), we are able to observe identical solutions for the GA and for the simplified conjugate gradient algorithm up to SNR $=$ 5dB under Scenario 2, while for Scenario 3 of Figure 4.18 (b) the BER gap between both approaches starts to widen, although remains relatively insignificant at low SNR values satisfying SNR $<$ 8dB. On the other hand, the eight-element array has a significantly wider BER performance gap in comparison to the theoretical MBER bound. The BER performance difference also increases, as the beamforming scenario becomes more challenging, i.e. when the interfering user 2 and user 3 get closer in angular terms to the desired user. In particular, after the fourth halving, corresponding to Scenario 5 outlined in Table 4.4, the GA becomes unable to produce a weight-solution near the optimum MBER solution at a complexity of 5500 objective function evaluations. Increasing the complexity to 6500 objective function evaluations however becomes capable



(a) Scenario 2 (b) Scenario 3 (c) $P = 110, G = 50$

Figure 4.18: BER performance of the GA-assisted MBER beamformer, based on Equation (4.11), under four different five-user scenarios; (a) Scenario 1 of 15^o (SOI), $-\mathbf{7.5^o}, \mathbf{37.5^o}, 80^o, -70^o$ and (b) Scenario 2 of 15^o (SOI), $\mathbf{3.75^o}, \mathbf{26.25^o}, 80^o, -70^o$, for four- and eight-element uniform linear arrays, having a population size and number of generations given by $P = 50, G = 30$ and $P = 110, G = 50$, respectively, and (c) Scenario 4 of 15^o (SOI), $\mathbf{9.375^o}, \mathbf{20.625^o}, 80^o, -70^o$ and Scenario 5 of 15^o (SOI), $\mathbf{12.1875^o}, \mathbf{17.8125^o}, 80^o, -70^o$ for an eight-element uniform linear array. The GA employed was using the roulette-wheel selection, elitism, span scaling, the fitness function $F2$ having a mutation and a crossover probability of $p_m = 0.1$ and $p_c = 0.9$, respectively. The associated four- and eight-element antenna arrays had an affordable complexity of $P \cdot G \sim 1500$ and 5500, respectively. These results were recorded for transmission over the AWGN channel, for five equal-power users and for the AOAs specified by Scenarios 2, 3, 4 and 5 of Table 4.4.

of producing a similar curve. It is also plausible that as the beamforming scenario becomes more complex, for example due to having interfering users at extremely close angles to the desired user, the number of bits assigned to the quantised real-valued weight solutions must be increased. In our case the bit-representation should be more accurate than $J = 8$ bits. For a very complex BER surface encountered in case of interferers close to the desired user, such as that in Scenario 5 of Figure 4.18 (c), where two interfering users are separated by less than 3^o from the desired user due to the inadequate resolution of the 16 different quantised weights, where the number of quantised weights is twice the number of array elements given by $L = 8$, we will have to tolerate a high BER. For a particularly complex BER surface increasing the number of elements may not improve the BER performance further, since the ultimate effects of quantisation errors will be determined by the number of array weights to be quantised. Apart from Scenario 5, it has been observed that the GA is capable of converging to the theoretical MBER solution with a high probability.

4.4.2 GA-Assisted MBER Beamforming Performance Using Kernel Estimates

Our motivation for the employment of a GA in the context of MBER beamforming was to circumvent the weaknesses of the stochastic gradient based MBER algorithms investigated in Chapter 3. More explicitly, the LBER algorithm [155, 198] for example requires appropriate initial array weight values and the optimisation of several adaptation parameters, which have to be carefully chosen in order to attain the optimum performance, instead of a sub-optimum performance due to encountering local minima. In this chapter we have assumed the knowledge of the exact BER of Equation (4.11). We will now use the estimated BER of Equation (4.15), which is obtained using Parzen's kernel density estimation [224] discussed in Section 3.4.1. According to Equations (4.15) and (4.16) the estimated BER is based on a block of training samples of length T and on the kernel width or smoothening parameter ρ_n, respectively.

Figure 4.19 shows the BER performance of the GA-assisted MBER beamformer for a population size of $P = 20$ and $G = 20$ generations, using the training segment length of $T = 512$ samples in conjunction with different smoothening parameters ρ_n of $0.01, 0.1$ and 1. Despite employing a long training segment, except perhaps for the smoothening parameter of $\rho_n = 0.1$, the BER performance observed is not very encouraging, especially for $\rho_n = 1$. The BER curve associated with the smoothening parameter of $\rho_n = 1$ became erratic, frequently exhibiting a worse BER performance than that of the MMSE solution. The erratic behaviour of Figure 4.19 may be explained by the BER histograms portrayed in Figure 4.17 associated with the GA-assisted MBER solutions of Figure 4.16 (a). It is observed from Figure 4.17 (c) for the eight-element case that the BER histogram, which is not as meritorious as that of the two-element case of Figure 4.17 (a), has resulted in a gap between the GA-aided MBER performance and the theoretical MBER bound. Similarly, for $\rho_n = 0.01$ the BER performance became worse than that of the MMSE solution, especially at higher SNR values. Hence,

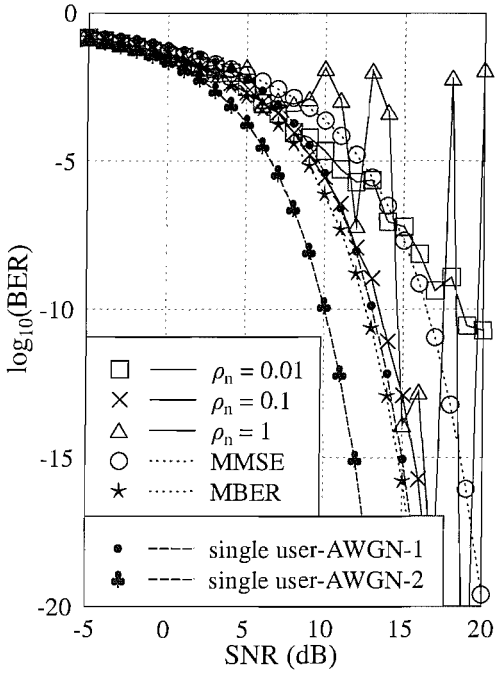
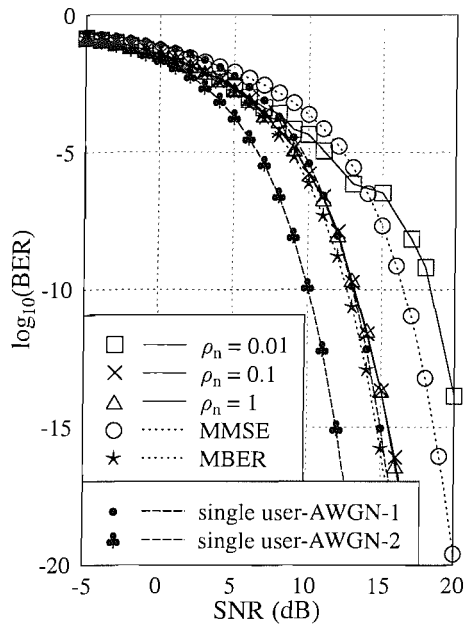


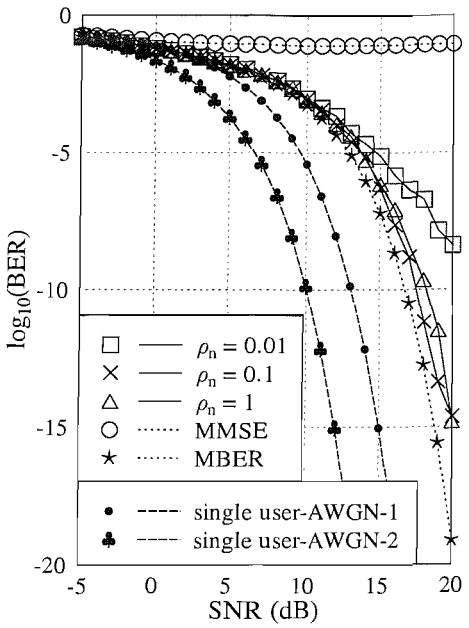
Figure 4.19: BER performance of the GA-assisted MBER beamformer, based on Equation (4.15), for five equal-power users and for the angular directions specified by Scenario 1 of 15^o (SOI), $-30^o, 60^o, 80^o, -70^o$ seen in Table 4.4 for a two-element uniform linear array. The GA employed was using roulette-wheel selection, elitism, span scaling, the fitness function $F2$ in conjunction with the estimated objective function of Equation (4.15)[6], using a training segment length of $T = 512$ samples and the smoothening parameters of $\rho_n = 0.01, 0.1, 1$. The population size and the number of generations was $P = 20$ and $G = 20$, respectively, while the mutation and crossover probability were $p_m = 0.1$ and $p_c = 0.9$, respectively. These results were recorded for transmission over the AWGN channel.

we opted for increasing the complexity, while using a shorter training segment length of $T = 256$ samples to compensate for the increased complexity. The results obtained using a population size of $P = 50$ and $G = 30$ generations, while using the probability of error expression corresponding to Equation (4.15), where the BER was estimated using the training segment length of $T = 256$ samples is plotted in Figure 4.20 under four different conditions; (a) SNR = INR_i for $i = 2, 3, 4, 5$, (b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$, (c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB. It is observed in Figure 4.20 (a) that in comparison to Figure 4.19 the BER performance seen in Figure 4.20 (a) improves, especially for the case of using the smoothening parameter of $\rho_n = 1$. This suggested that the GA is capable of improving the

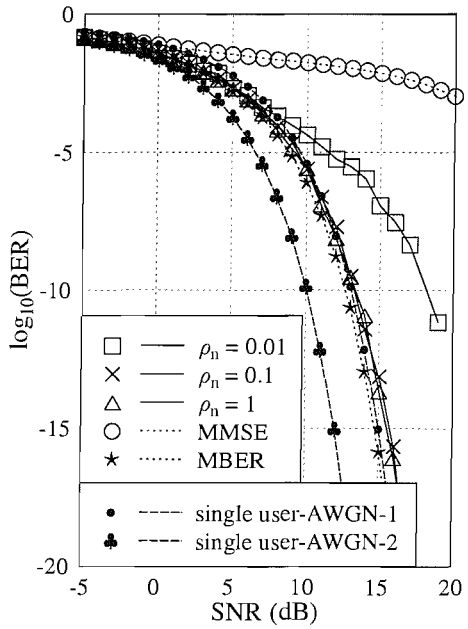
[6]The estimated BER calculated based on Equation (4.15), was used in the objective function $F2$ of the GA-assisted MBER approach described by Equation (4.18).



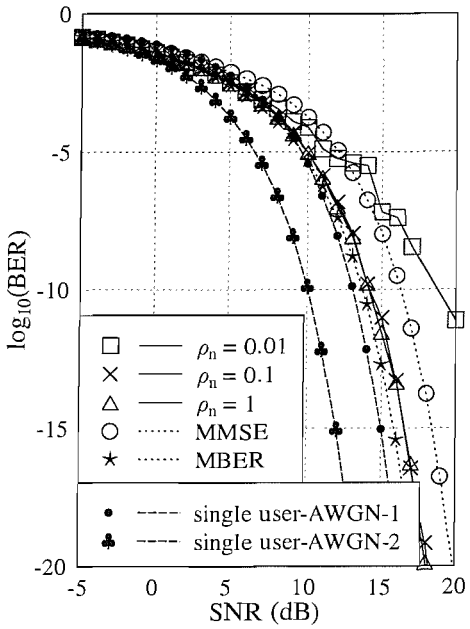
(a) SNR = INR_i for $i = 2, 3, 4, 5$



(b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$



(c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB



(d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB

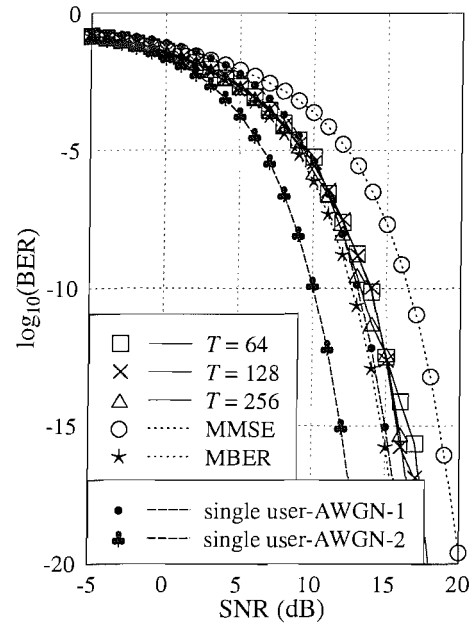
Figure 4.20: BER performance of the GA-assisted MBER beamformer, based on Equation (4.15), under four different conditions; (a) SNR = INR_i for $i = 2, 3, 4, 5$, (b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$, (c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB for a two-element uniform linear array. The GA employed was using the roulette-wheel selection, elitism, span scaling, the fitness function $F2$ in conjunction with the estimated objective function of Equation (4.15), using a training segment length of $T = 256$ samples and the smoothening parameter values of $\rho_n = 0.01, 0.1, 1$. The population size and the number of generations was $P = 50$ and $G = 30$, respectively, while the mutation and crossover probability were $p_m = 0.1$ and $p_c = 0.9$, respectively. These results were recorded for transmission over the AWGN channel for five users and the angular directions specified by Scenario 1 of Table 4.4.

estimated MBER beamforming performance in conjunction with a shorter training segment length, provided that we increase the affordable GA complexity. The remaining plots of Figures 4.20 (b), 4.20 (c) and 4.20 (d) portray the BER performance obtained, as we vary the interfering users' power with respect to the desired user's signal power. Note that there is always a BER performance gap between the GA solution and the theoretical MBER bound, particularly for the smallest smoothening parameter value of $\rho_n = 0.01$. This is unavoidable, since we are using the estimated probability of error $\hat{P}_E$ and the smoothening parameter of $\rho_n = 0.01$ which may be inadequate or inappropriate for this specific case, preventing the GA from converging to the MBER solution. Even for the exact error probability of Equation (4.11), at certain SNR values, we occasionally experienced a BER performance gap, although this was rarely the case. Furthermore, the kernel width has to be adjusted according to the variance of the training samples used and thus it is dependent on both the SNR and the INR [157].

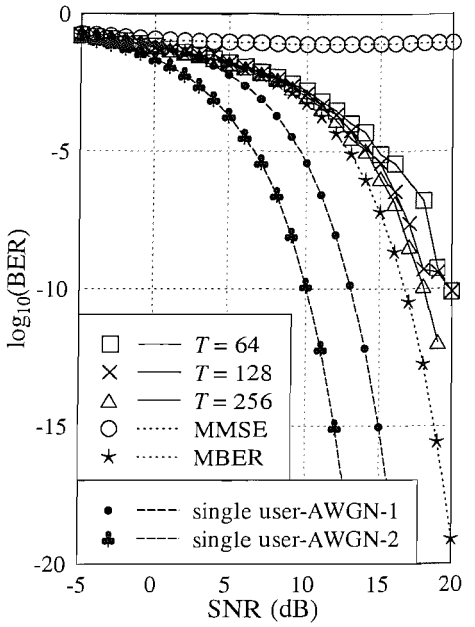
One of the reasons for invoking GAs for finding the MBER solution is because they are naturally guided by the objective function, such as the MBER objective function of this chapter. The BER performance shown in Figure 4.20 was unfortunately dependent on the specific choice of the smoothening parameter ρ_n used for satisfying Equation (4.16). One option of calculating a meritorious smoothening parameter ρ_n is based on Silverman's rule of thumb [225], where ρ_n is automatically determined by the block of T training samples and its standard deviation $\hat{\sigma}$. The associated expression for Silverman's smoothening parameter ρ_n has been expressed in Equation (3.102), which is restated here for convenience as:

$$\rho_n = \left(\frac{4\hat{\sigma}^5}{3T}\right)^{1/5} \approx 1.06\hat{\sigma}T^{-1/5} . \tag{4.23}$$

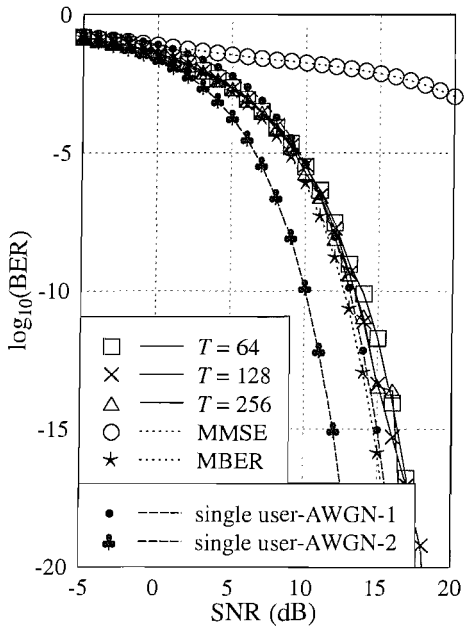
Figure 4.21 shows the achievable BER performance as a function of the users' signal power under the same four conditions as those stipulated in Figures 4.20 (a) to 4.20 (d). In comparison to the plots of Figure 4.20 the curves of Figure 4.21 exhibit a better BER performance, particularly in the SNR region up to SNR = 15dB. All plots of Figure 4.21 illustrate the superiority of the MBER approach over the MMSE approach in terms of counteracting the near-far effects, while demonstrating that the GA is capable of automatically adapting to a new BER surface without being reconfigured, while imposing the same complexity of $50 \cdot 30 = 1500$ objective function evaluations. It is also observed in Figure 4.21 that the GA's performance is not significantly affected at higher SNR values by using shorter training segment lengths of $T = 128$ and $T = 64$ samples, especially for SNR $\leq$ 10dB, apart from some BER curve glitches. The comparatively large BER performance gap seen in Figure 4.21 (b) at higher SNR values may relate to the insufficiently high complexity of the GA, given the more complex BER surface owing to having all interferers at a 6dB higher signal power than the desired user. Furthermore, Silverman's rule of thumb provides an adequate, but by no means optimum smoothening parameter value, therefore it does not guarantee the perfect smoothening of the probability of error estimates, thus it may also result in a high BER value. Another option for improving the attainable performance without increasing the GA's complexity is to increase the



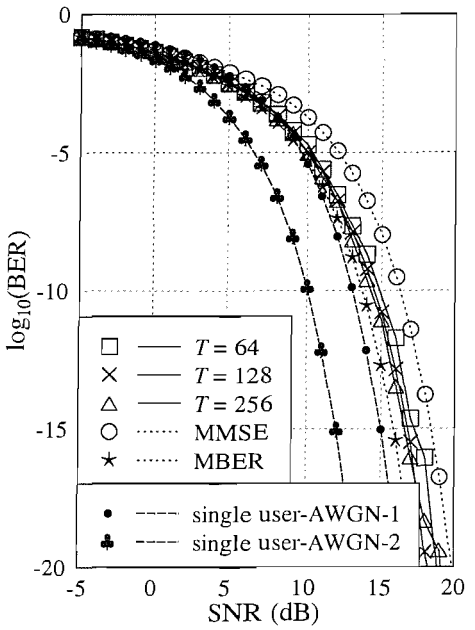
(a) SNR = INR_i for i = 2, 3, 4, 5



(b) INR_i = SNR +6dB for i = 2, 3, 4, 5



(c) SNR = INR_i for i = 3, 4, 5 and INR_2 = SNR +6dB



(d) SNR = INR_i for i = 2, 3, 4 and INR_5 = SNR +6dB

Figure 4.21: BER performance of the GA-assisted MBER beamformer, based on Equation (4.15), under four different conditions; (a) SNR = INR_i for $i = 2,3,4,5$, (b) INR_i = SNR +6dB for $i = 2,3,4,5$, (c) SNR = INR_i for $i = 3,4,5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2,3,4$ and INR_5 = SNR +6dB for a two-element uniform linear array. The GA employed was using the roulette-wheel selection, elitism, span scaling, the fitness function $F2$ in conjunction with the estimated objective function of Equation (4.15), using a training segment length of $T = 64, 128, 256$ samples and a smoothening parameter ρ_n obeying Silverman's rule of thumb outlined in Equation (4.23). The population size and the number of generations was $P = 50$ and $G = 30$, respectively, while the mutation and crossover probability was $p_m = 0.1$ and $p_c = 0.9$, respectively. These results were recorded for transmission over the AWGN channel for five users and the angular directions specified by Scenario 1 of Table 4.4.

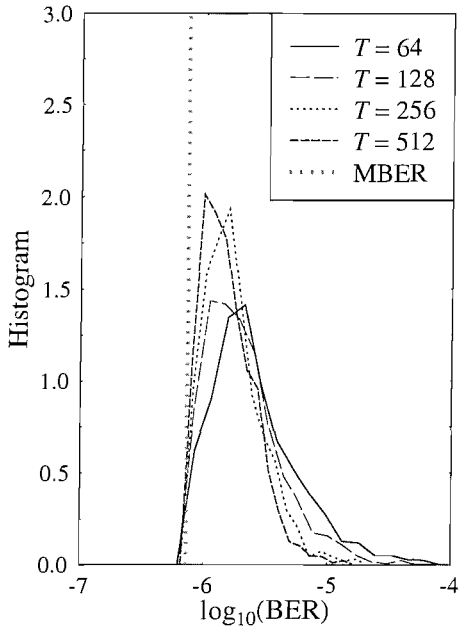
number of training samples T. This statement is based on the observation of Figure 4.22, since all the histograms of Figures 4.22 (a), 4.22 (b), 4.22 (c) and 4.22 (d), associated with user power scenarios similar to those of the corresponding Figures 4.21 (a) to 4.21 (d) demonstrated that the convergence of the GA improves, when a longer training segment is used. The resultant BER histogram becomes narrower, as the training segment length T increases.

4.5 Conclusions

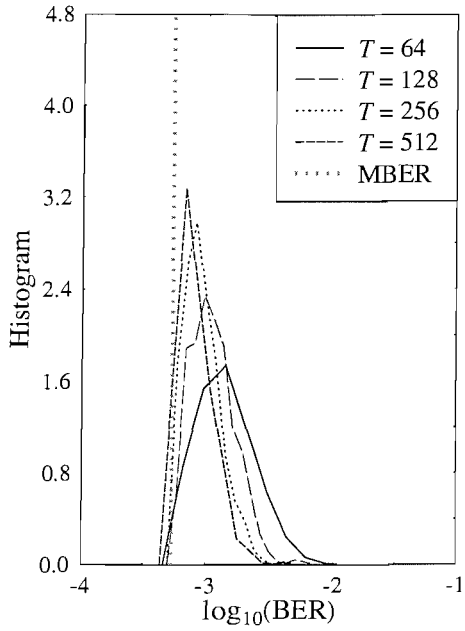
A novel GA-assisted MBER beamformer has been developed. The family of GAs operates on the basis of finding a solution from a rich database of candidate solutions by searching through a number of generations, rather than evolving from a single point to another single point. If the fitness function is not carefully chosen, it may lead the search to local minima. For our simulations we initially use the fitness function $F1$ of Equation (4.17), which becomes less efficient for lower BER values in the range of BER $\leq 10^{-3}$, since the fitness value difference between two different BERs satisfying BER $\leq 10^{-3}$ is rather small, even though their relative BER difference may be as high as BER $= 10^{-3}$ and BER $= 10^{-7}$. For the sake of overcoming this problem, we used a more effective fitness function, namely that expressed in Equation (4.18), which we referred to as the fitness function $F2$. In addition to the improved average BER performance, using fitness function $F2$ has reduced the GA complexity required for attaining a BER approaching the exact MBER solution.

Figure 4.13 demonstrated that the GA's BER performance improved drastically when using elitism combined with scaling and the fitness function $F2$. The BER performance seen in Figure 4.12, when using the less efficient fitness function $F1$ has also shown a tremendous improvement in terms of the attainable average BER, but nonetheless, the performance remained lower than that when using the fitness function $F2$, as demonstrated by the plots seen in Figure 4.14. This indirectly shows the robustness of GAs, provided that the GA has the option of offering fitness scaling combined with elitism, for the sake of improving the achievable BER performance. When fixing the complexity of the GA to the product of the population size P and the number of generations G, we have the advantage of knowing the GA's complexity in advanced. It is also advantageous to invoke a specific fitness scaling technique that increases the chances of selecting high-fitness individuals, when the generation index g becomes high, as span scaling, which was introduced in Section 4.3.5.2. More explicitly, in comparison to the more conventional sigma scaling, span scaling has the advantage of forcing the GA to converge to a solution after a fixed number of generations by increasing the selection pressure, as the generation index g increases [157]. It has been shown in Figure 4.15 that span scaling performs at least similarly, if not better than sigma scaling associated with the coefficient $c = 1$.

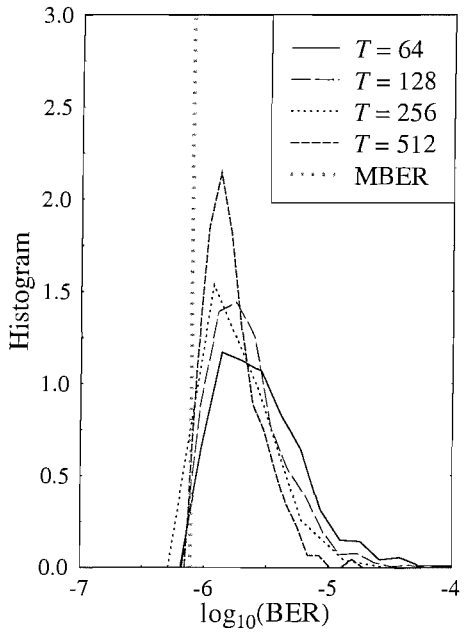
We used the BER histogram accumulated over 1000 randomly initialised GA runs for the sake of characterising the achievable performance of the MBER beamformer, as the number of generations



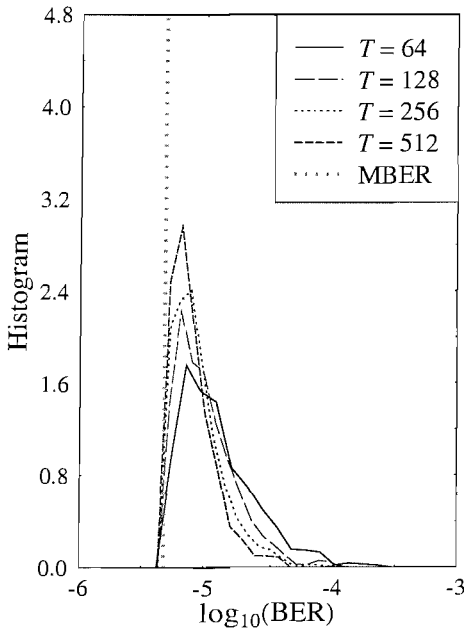
(a) SNR = INR_i = 10 dB for i= 2, 3, 4, 5



(b) SNR = 10 dB and INR_i = SNR +6dB for i = 2, 3, 4, 5



(c) SNR = INR_i = 10 dB for i = 3, 4, 5 and INR_2 = SNR +6dB



(d) SNR = INR_i = 10 dB for i = 2, 3, 4 and INR_5 = SNR +6dB

Figure 4.22: The BER histogram of the GA-aided beamformer used in Figure 4.21, evaluated based on Equation (4.15), in conjunction with a population size of $P = 50$ and $G = 30$ generations, at SNR = 10dB, under four different users signal power conditions; (a) SNR = INR_i for $i = 2, 3, 4, 5$, (b) INR_i = SNR +6dB for $i = 2, 3, 4, 5$, (c) SNR = INR_i for $i = 3, 4, 5$ and INR_2 = SNR +6dB and (d) SNR = INR_i for $i = 2, 3, 4$ and INR_5 = SNR +6dB for a two-element uniform linear array. The associated BER histogram for each block of $T = 64, 128, 256, 512$ training samples was constructed from 1000 randomly initialised GA runs, recorded when communicating over the AWGN channel for five users and the angular directions specified by Scenario 1 of Table 4.4.

G increases, as it was shown in Figure 4.12. From a different perspective, we also characterised the convergence rate of the algorithm in terms of the variance of the BER as a function of the affordable GA complexity, as shown in Figures 4.14 (b). In terms of the required complexity, the observation of Figure 4.16 suggested that the complexity of the GA increases quadratically, as the number of the antenna array elements L increases. The GA was also shown to be near-far resistant, approaching the theoretical MBER bound even in a large search space associated with a high number of users and numerous beamformer elements. For the particular case of Scenario 5 characterised in Figure 4.18, a large BER performance gap has been observed with respect to the exact MBER solution. Increasing the GA's complexity may not improve the situation further because the achievable performance is likely to be limited by the number of bits used for representing the quantised beamformer weights.

Instead of using the true BER formula of Equation (4.11) for guiding the GA's search, Figures 4.19 to 4.22 portray the GA's performance attained using the estimated BER of Equation (4.15). A good BER performance was observed even in conjunction with short training segments, such as $T = 64$ samples, although the spread of the BER histogram did improve, as a longer training segments were used, as evidenced by the BER histograms of Figure 4.22.

Indeed, GAs are suitable for MBER beamforming. Their ability of achieving a similar BER performance to the theoretical MBER solution has also been demonstrated. We have observed that this GA-aided approach exhibits a superior performance in terms of convergence, while requiring a modest complexity for arriving at the global solution and in this respect the GA-assisted MBER beamformer outperforms the conjugate gradient algorithm based solutions, such as the LBER or the approximate LBER algorithms investigated in Chapter 3, which often require the reoptimisation of the step size μ and the smoothening parameter ρ_n.

Chapter 5

Conclusions and Future Research

In this treatise we investigated the performance of several adaptive algorithms in the context of temporal reference based beamforming. The performance of the algorithms was analysed with the aid of computer simulations using BPSK modulated transmissions over an AWGN channel. The topics covered will be summarised in Section 5.1, while in Section 5.2 we will outline a range of possible further research directions.

5.1 Summary and Conclusions

In **Chapter 2** we presented a historical perspective on the evolution of adaptive beamforming. An overview of various beamforming algorithms was also provided, including the classification of adaptive beamforming techniques. The basic concept of beamforming was exemplified in Section 2.2.3.

For the sake of determining the optimal beamforming weights, several optimisation criteria were presented in Section 2.3, including the Minimum Mean Square Error (MMSE), the Maximum Signal-to-Interference Ratio (MSIR) and the Minimum Variance criterion. Table 2.5 summarised the features of these three criteria, along with their advantages and disadvantages. However, the beamformers based on these three criteria were all dependent on the time-variant statistics of the signal received by the antenna array of the beamformer, therefore necessitating the employment of adaptive algorithms. We characterised three classic algorithms, namely the Least Mean Square (LMS) algorithm of Section 2.4.1.1, the Recursive Least Square (RLS) algorithm of Section 2.4.1.2 and the Sample Matrix Inversion (SMI) algorithm of Section 2.4.1.3. For Binary Phase Shift Keying (BPSK) transmission over an Additive White Gaussian Noise (AWGN) channel the Recursive Least Square (RLS), the Direct Matrix Inversion (DMI), i.e. a direct form of the SMI algorithm and the recursive form of the SMI (RSMI) exhibit a similar performance. However, the complexity imposed by the DMI algorithm is of the order of $O(L^3)$, as opposed to the order of $O(L^2)$ of the RLS and RSMI algorithms, where L is the number of antenna elements in the array. A brief complexity comparison between the LMS,

Normalised LMS (NLMS), RLS, RSMI and DMI algorithms in the context of the number of complex multiplications and additions was provided in Section 2.6.1.3. In the context of flat Rayleigh fading channels, the performance of the DMI, RSMI and RLS algorithms was degraded and the associated degradation was dependent on the forgetting factor α used. The RSMI algorithm was observed to be more sensitive to the choice of the forgetting factor α, than the RLS algorithm.

The beam patterns of a three-element uniform linear antenna array (ULA) system, receiving signals from a number of combinations of the users' angle of arrival (AOA) were portrayed in Section 2.6.2. Specifically, from the beam patterns shown in Section 2.6.2.1 and 2.6.2.2 for both an AWGN and flat Rayleigh fading channels, respectively, a maximum of 8dB difference was observed as a function of the AOA. The achievable performance expressed in terms of interference rejection was improved, whenever the interferer closest to the desired user was allocated more power. Increasing the desired user's power does not result in significant improvements. As the number of users was increased, thus decreasing the degrees of freedom (DOF) exhibited by the system, the interference rejection became less effective. For example, the worst-case rejection in a five-user, three-element antenna array system communicating over both the AWGN and flat Rayleigh fading channels considered was reduced to 17dB as compared to 38dB and 32dB for a two-user scenario.

All simulations performed in Chapter 2 were based on the MMSE criterion. However, for a communication system, it is the achievable Bit Error Ratio (BER) that really matters. Ideally, the system design should be based on minimising the BER, rather than the MSE [155], although in certain situations even the MMSE solution produces a relatively good BER performance.

In **Chapter 3** we characterised various adaptive beamforming schemes, which were based on directly minimising the system's Bit Error Ratio (BER), thus we referred to these schemes MBER beamforming arrangements. Section 3.2 revised the signal model previously used in Chapter 2, streamlining some of the equations used in the context of MBER beamforming. Having derived the closed-form MMSE solution of Equation (3.25) in Section 3.2.1, we formulated a non-adaptive algorithm that directly minimises the system's BER based on the conjugate gradient technique, more precisely on the simplified conjugate gradient algorithm. The procedure of finding the MBER beamforming solution based on this simplified conjugate gradient algorithm was outlined in Equations (3.93) to (3.96).

The BER performance of the MBER beamforming scheme of Section 3.3 was analysed in Section 3.5 in comparison to that of the MMSE beamforming arrangement of Section 3.2.1. It can be concluded from the simulation results derived for a two-element antenna array system, which were recorded for BPSK transmissions over an AWGN channel, that the performance of MBER beamforming is superior to that of MMSE beamforming. Furthermore, for the four-element antenna arrays of Section 3.5.3, it is observed in Figures 3.31 (a) and 3.31 (b) that the MBER solution is more robust to the phenomenon of array-overloading, when the number of users is higher than the number of array element. In all related plots, we also included the single-user curves for the sake of benchmarking the achievable BER

performance of a beamformer. The performance gap between the MMSE and MBER beamforming solution with respect to the upper-bound performance the optimum has been portrayed in Figures 3.10 and 3.11 as a function of the interfering user's AOA for the two-user, multi-element scenarios with the SOI arriving from the direction of 0^o with respect to the normal of the antenna array. On the other hand, Figure 3.32 shows the signal power requirement difference with respect to the single-user performance, when supporting multiple users employing a four-element array in the AOA scenario plotted in Figure 3.30. We observe from Figures 3.10, 3.11 and 3.32 that the MBER solution always exhibits a performance closer to the optimum single-user case than the MMSE solution. For example, as shown in Figure 3.10, the SNR loss of the MBER solution is about 3 to 5dB lower than that of the MMSE solution in comparison to the optimum single-user case. The interpretation of the differences between the MMSE and MBER approach in optimising the beamformer's weights may be visualised with the aid of Figures 3.1, 3.27 and 3.34. It was observed in Section 3.5.2 that as the shape of the probability density function (PDF) became increasingly non-Gaussian, the performance difference between the MMSE and MBER beamforming schemes became increasingly more significant. Furthermore, MBER beamforming is more robust to the near-far effects often encountered in wireless systems owing to power-control errors.

In Section 3.4 we investigated an adaptive implementation of the MBER beamforming scheme advocated. Specifically, we classified these adaptive MBER beamforming algorithms into two categories, namely the family of block adaptive and the class of sample-by-sample adaptive stochastic gradient approaches, which were outlined in Sections 3.4.2 and 3.4.3, respectively. In the block-based implementation context we studied the Block Adaptive Conjugate Gradient (BACG) algorithm of Section 3.4.2, which used a block of T training samples constituted by the desired user's bits. On the other hand, the stochastic approach employed a sample-by-sample adaptive procedure for determining the optimum array weight values. For the Least Bit Error Rate (LBER) algorithm of Section 3.4.3.1, the kernel width radius ρ_n was shown to be dependent on the magnitude of the beamformer's weight vector $\sqrt{\mathbf{w}^H\mathbf{w}}$, while for the approximate LBER (ALBER) algorithm of Section 3.4.3.2 we assumed an independent ρ_n as the kernel width, for the sake of reducing the algorithm's complexity. Section 3.6 portrayed the performance of all the three adaptive MBER beamforming algorithms. It can be seen from both Figures 3.35 (a) and 3.35 (b) that the BACG algorithm has converged rapidly, approaching the steady BER state value after eight iterations, but resulted in a slightly higher BER than the theoretical MBER performance of the non-adaptive MBER beamforming algorithm obtained in Section 3.5.2. When using the BACG algorithm, the parameter ρ_n in Equation (3.117) has to be adjusted according to the step size value μ during the initialisation step of the simplified conjugate gradient algorithm outlined in Section 3.3, for the sake of attaining a rapid convergence towards the theoretical MBER performance. We then considered the option of using a single-sample based estimate of the PDF in the context of the LBER and ALBER algorithms of Sections 3.4.3.1 and 3.4.3.2, respectively.

According to Equations (3.116) and (3.123) the calculations required for determining the gradient of

the BER for both the BACG and LBER algorithms, respectively, have to use a variable kernel width of $\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}$ in the kernel density estimation process of Section 3.4.1, which is related to the beamformer's weight vector. The ALBER algorithm on the other hand eliminates this procedure and uses a constant width ρ_n for approximating the term $\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}$. Therefore the calculation of the gradient $\nabla_{\mathbf{w}}\hat{P}_E$ in the ALBER algorithm will only depend on the parameter ρ_n, not on $\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}$, as can be seen in Equation (3.132). In Figures 3.43 (a) and 3.44 recorded for the equal- and unequal-power user scenario, respectively, for the specific AOA considered in Figure 3.18 both the convergence rate and the steady state BER value of the ALBER algorithm is better than that of the LBER algorithm. According to Figure 3.43 (b) characterising the equal-power user scenario, although the LBER algorithm initially appears to have a faster convergence, near the iteration index 400 the performance curves of the algorithms crossed each other and beyond this point the ALBER technique outperformed the LBER algorithm, arriving at a BER lower than that of the LBER approach. Despite its reduced complexity, the ALBER algorithm exhibited an outstanding performance, as shown in Figures 3.46 and 3.48 for both the equal-power and unequal-power user scenarios, respectively.

The performance of all the adaptive MBER beamforming algorithms was however dependent on the step size μ and on the radius ρ_n related to the kernel width. Specifically, the BER performance of a two-element system evaluated in conjunction with the AOA scenarios represented by Figure 3.18 has been portrayed in Figures 3.41 and 3.42 as well as in Figure 3.49 for the BACG and ALBER algorithms, respectively. An appropriate choice of these parameters' value is necessary for the sake of obtaining an attractive adaptive MBER beamforming performance. Table 3.7 summarises the acceptable range of the step size μ and kernel width values ρ_n for our specific equal-power users and for the unequal-power users of the corresponding AOA scenario plotted in Figure 3.18.

In **Chapter 4** a novel GA-assisted MBER beamforming technique was introduced. We employed GAs for assisting the operation of the MBER beamforming scheme investigated in Chapter 3 using the BACG, LBER and ALBER algorithms. More specifically, GAs were invoked for overcoming certain drawbacks of the MBER beamforming scheme, which includes the irregular shape of the BER surface and also the adaptive MBER algorithms' performance dependence on the choice of the initial weights, on the kernel width radius ρ_n and on the step size μ.

The performance of the proposed GA was characterised both by the BER histogram and by the achievable average BER values at the beamformer's output. Having found a good GA configuration, we compared the BER performance results generated as a function of the SNR to the theoretical MBER bound, produced by the simplified conjugate gradient algorithm of Chapter 3. Apart from performing elitism for the sake of preserving the meritorious genes, in Section 4.4 we also stimulated the level of competition between individuals in the population for the sake of avoiding premature convergence, by incorporating fitness scaling techniques. In addition to the more conventional sigma scaling, we introduced a novel scaling technique that has some relation to the generation index g,

referred to as span scaling. It was observed in Figure 4.15 that the GA employing the span scaling technique performed better or at least similarly to sigma scaling having a coefficient of $c = 1$. In terms of the required complexity, the observation of Figure 4.16 suggested that the complexity of the GA increased quadratically, as the number of the antenna array elements L was increased.

It was demonstrated in Figures 4.16 to 4.22 of Section 4.4 that the GA is suitable for aiding the operation of MBER beamforming. In comparison to Figure 3.41 charactersing the stand-alone BACG algorithm, the BER performance of the GA-assisted MBER technique of Figure 4.20 was shown to be generally more robust to the algorithmic parameters' variations, although the best BER performance curve of Figure 4.20 is not as good as the best BER performance curve of Figure 3.41. Note that for the BACG algorithm the performance portrayed in Figure 3.41 was generated in conjunction with a step size of $\mu = 0.9$. On the other hand, Figure 4.21 shows the BER performance of the GA-aided approach as a result of using Silverman's rule of thumb given in Equations (3.102) or (4.23). Similar to the BACG algorithm's BER performance of Figure 3.42, the BER performance of the GA-aided system of Figure 4.21 was also improved, when a longer training block length T was used. It was observed in Figures 4.21 and 4.22 that as we increased the training block length T, the BER performance improvement attained by the GA became more significant. This may be attributed to the fact that the GA's approach is based on a random search method. However, owing to this particular reason, the employment of GAs is only beneficial, when using the block-data based approach. Based on the observation of Figure 4.22, using a sample-by-sample approach associated with $T = 1$, as used in the LBER and ALBER algorithms may result in a significantly higher BER than that using $T = 64$, as shown in Figure 4.21, in addition to its dependency on the specific choice of the smoothening parameter ρ_n used for satisfying Equation (4.16).

From all the above-mentioned observations stated in Chapters 3 and 4, particularly when comparing Figures 3.41, 3.49 and 4.20 as well as Figures 3.42 and 4.21, we concluded that the ALBER algorithm is capable of offering solutions nearly identical to the exact MBER performance, despite having the lowest computational complexity. More specifically, the complexity was similar to that of the LMS algorithm of Section 2.4.1.1, namely of the order of $O(L)$, where L is the number of antenna elements in the array. On the other hand, the GA is capable of attaining a solution, which is in good agreement with the exact MBER solution at the cost of a moderate complexity, whilst remaining robust to the variation of the adaptive parameters, namely to the step size μ and kernel width ρ_n, unlike the stochastic gradient based algorithms presented in Chapter 3, such as the BACG, LBER and ALBER algorithms.

Having concluded our findings, we are aware that several important areas still warrant further research, as suggested in the next section.

5.2 Suggestions for Future Research

In this section we suggest a number of further research items and open problems in the context of adaptive MBER beamforming.

1. Our investigations of adaptive MBER beamforming were concentrated on the simplest possible BPSK modulation scheme conducted over the non-dispersive AWGN channel. Further investigations may incorporate higher order modulation schemes, such as 4-QAM and 16-QAM. Furthermore, the channel considered may include hostile fading channels and dispersive wideband channels that induce ISI. An extension to wideband beamforming would also be interesting and beneficial in terms of fulfilling the requirements of higher data rate services.

2. It is noted from Figure 3.20 that the separation between the desired signal and the interfering signals expressed in terms of the AOA affect the system's BER performance. For a specific antenna array system, when the AOA separation is below a certain threshold, the modem's phasor constellation becomes linearly inseparable, which renders classic 'linear' beamforming, i.e. the linear combination of the received signals at the output of the antenna array, inadequate. This statement has been evidenced in [249–251] in the context of single-user channel equalisation. Therefore, it is interesting to consider nonlinear beamforming schemes, such as those referred to as Bayesian beamformers [252], which may still perform adequately in scenarios, when linear beamforming fails. Even for the linearly separable case, a nonlinear processing technique may be expected to provide a better performance than a linear one, although this is achieved at the cost of an increased complexity.

3. It would also be beneficial to compare the BER performance gain attained by the GA-aided MBER beamforming scheme to that of the GA-aided MMSE beamformer, as a function of the number of users [253] or the number of antenna elements. However, in contrast to MBER scheme, whose array weight values may be within a fixed range of $[D_{min}, D_{max}] = [1, -1]$, when considering Equation (3.70), the range of the MMSE solution's array weight values would be dependent on the scenarios considered. Therefore, a study of the D_{min} and D_{max} values, as a function of the number of users and/or number of antenna array elements used should be conducted, since this would affect the precision of the quantised array weights used, as suggested by Equation (4.10), which in turn would limit the attainable BER performance.

4. Simulations performed for the GA-aided MBER beamformer were based on the specific scenario of having the AOAs portrayed in Figure 3.18. Further research is required for generalising the results obtained to more complex beamformers and propagation scenarios. The GA procedures used, such as elitism and fitness scaling have significantly influenced the achievable BER performance of the investigated GA-aided MBER beamformer. More additional GA procedures,

such as weighted mutation, post selection and double search [153] may be included for the sake of improving the attainable BER performance versus the required complexity. The weighted mutation principle is portrayed in Figure 5.1, illustrating the mutation probability distribution of the various bits of an individual containing for example three concatenated parameters, namely $C(1)C(2)C(3)$. Specifically, we render the mutation probability to be dependent on the bit's position and on the number of bits used for a specific parameter, which may be formulated as:

$$p_{m,j} = \frac{kp_m}{\sum_{j=1}^{k} j}(k - (j-1)), \tag{5.1}$$

where $p_{m,j}$ is the weighted mutation probability for bit j, while k is the total number of bits in a parameter, say in $C(2)$. This method protects the Most Significant Bits (MSBs) from mutation and concentrates the search within the local region, once the GA is considered reasonably close to convergence.

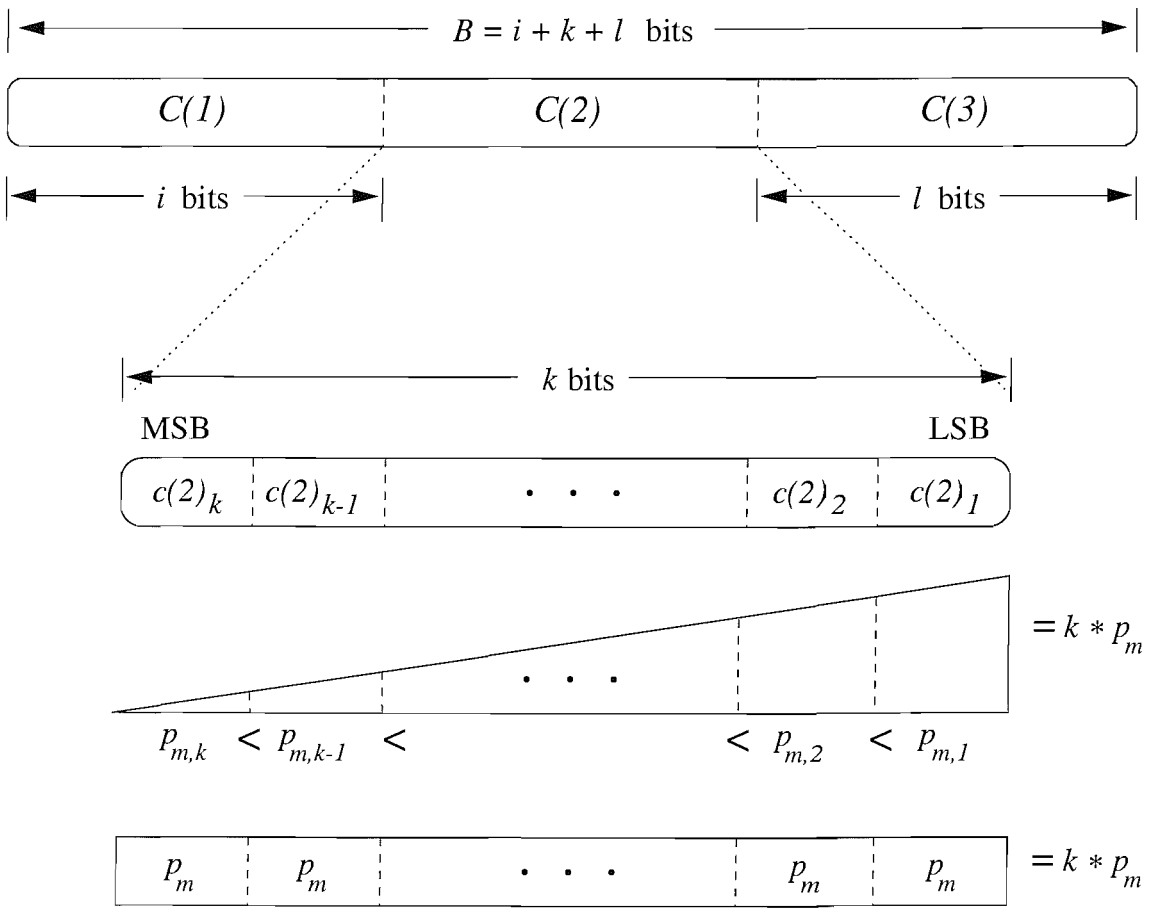


Figure 5.1: An illustration of the weighted mutation process.

List of Symbols

General Notations

Relations and Operators

$(\cdot)^*$	complex conjugate
$(\cdot)_{avg}$	average
$(\cdot)^H$	Hermitian transpose (complex conjugate transpose)
$(\cdot)_{max}$	maximum value
$(\cdot)_{min}$	minimum value
$(\cdot)^T$	transpose
$\nabla_a(\cdot)$	gradient with respect to a
$\exp^{(\cdot)}$	exponential
$E[\cdot]$	expectation operator
$\Im[\cdot]$	imaginary part
$p(\cdot)$	probability
$\Re[\cdot]$	real part
$\text{sgn}(\cdot)$	signum function
$tr[\cdot]$	trace of a matrix

Symbols and Variables

α	positive integer or forgetting factor
β	scalar coefficient
$\chi(\mathbf{R})$	eigenvalue spread of $\mathbf{R}$
Δ	real-valued mutation size
η	precision or efficiency
λ	the wavelength
λ_{max}	maximum eigenvalue

$\lambda_{\min}$	minimum eigenvalue
μ	positive real-valued constant or step-size
μ_o	a positive constant
θ_m	angle of arrival with respect to the normal of the array of point source/mobile user m
ρ_n	kernel width radius of the estimated PDF
$\hat{\sigma}$	estimated standard deviation
σ	standard deviation
σ_m^2	variance of transmitted signal of user m
ω	(angular) frequency
a	a positive constant or scalar coefficient
a_m	signal amplitude of mobile user m
A	$(M \times M)$-dimensional diagonal amplitude matrix
$b_m(n)$	pseudo-random binary sequence of mobile user m at time instant n; for BPSK $b_m \in \pm 1$
b	a scalar constant
b_d	user d transmitted bit
B	vector of an individual or chromosome length
$\tilde{c}$	estimate of approximated scalar constant
$\hat{c}$	estimated scalar constant
c	speed of light $3 \times 10^8 ms^{-1}$
$c_{q,+}$	a scalar constant as input variable to the Q-function
C	a string of bits (0s and 1s)
d	distance between two antenna elements of a linearly spaced array
$\mathbf{d}$	desired signal vector
$\mathbf{d}$	direction vector for updating gradient algorithm
D	a scalar constant
$e(n)$	error signal
e_o	initial error
e_p	expected number of individual for crossover
f_p'	scaled fitness of individual p
f	frequency
$\bar{f}^g$	average fitness of individuals in generation g
f_p^g	raw fitness of an individual p in generation g
f_p	fitness function of individual p
$F(\theta_m)$	array factor in response to source/mobile user m arriving from angle θ_m
g	complex constant/constraint
G	maximum number of generations

$\mathbf{G}$	$(L \times M)$-dimensional system matrix of the beamformer
$\mathbf{I}_L$	$(L \times L)$-dimensional identity matrix
J	number of bits representing a variable
k	a scalar constant
l	an omnidirectional antenna element of an array
L	total number of omnidirectional antenna elements of an array
$m_m(n)$	complex-valued modulating signal of an uncorrelated point source/mobile user m at time instant n
M	total number of uncorrelated point sources/mobile users
n	iteration/discrete time index
$n_l(n)$	complex random white noise signal on the l-th element of the antenna array at time instant n
$\mathbf{n}(n)$	complex array noise vector
N_b	number of distinct combination of bits transmitted
N_{sb}	number of distinct combination of bits transmitted with desired user's bit $b_{q,s}$ fixed to a value, which is half of N_b
O	complexity order
(p, q)	location of a point on the x-y Cartesian plane
$\tilde{p}$	estimate of approximated PDF
$\hat{p}$	estimated PDF
$p(a\|\mathbf{b}_q)$	probability of decision function a given bit vector $\mathbf{b}_q$ is transmitted
p_c	probability of crossover
p_m	probability of mutation
p_p	probability of individual p being selected
P	mean output power of beamformer
P	population size
$\mathbf{P}$	$(M \times M)$-dimensional users' power matrix
$\tilde{P}_E$	estimate of approximated probability of error
$\hat{P}_E$	estimated probability of error
P_E	probability of error
$Q(z)$	Q-function
$r(n)$	beamformer's reference signal/desired response at time instant n
$\mathbf{R}_\mathbf{x}$	$(L \times L)$-dimensional auto-correlation matrix of the input signal vector to beamformer or array output $\mathbf{x}$
$\mathbf{R}_{ij}$	a component of auto-correlation matrix between the i-th and the j-th element of an array
$\mathbf{s}_m$	steering vector of user m
$\mathbf{S}$	$(L \times M)$-dimensional steering matrix

t_k	time slot
$t_{(p,q)}(\theta_m)$	time delay of an element at position (p, q) with respect to the origin for point source/mobile user m having arriving angle of θ_m
$t_l(\theta_m)$	time delay measured from element l to element 1 for point source/mobile user m having arriving angle of θ_m
T	block of training samples
T	number of selected individuals
u	unsigned integer
$\mathbf{u}$	unwanted interfering signals vector
w_l	complex weight of element l of the antenna array of a beamformer
$\ddot{\mathbf{w}}$	normalised weight vector
$\mathbf{w}(n)$	complex weight vector of a beamformer at time instant n
$\mathbf{w}_{\text{opt}}$	optimum weight coefficient vector (Wiener filter)
$\mathcal{X}^{(\pm)}$	set consists of possible noise-free input vector to beamformer i.e. noise-free received signal vector at the antenna array, given the desired user's transmitted bit $\mathbf{b}_s = \pm 1$
$\bar{x}_l(n)$	noise-free input signal to beamformer i.e. noise-free received signal at element l of the antenna array for all users at time instant n
$\bar{x}_{l,m}(n)$	noise-free input signal to beamformer i.e. noise-free received signal at element l of the antenna array for user m at time instant n
$\bar{x}_q$	possible noise-free input vector to beamformer i.e. noise-free received signal vector at the antenna array
$\bar{x}_q^{(+)}$	possible noise-free input vector to beamformer i.e. noise-free received signal vector at the antenna array, given the desired user's transmitted bit vector $\mathbf{b}_s(n) = +1$
$\bar{\mathbf{x}}(n)$	noise-free input vector to beamformer i.e. noise-free received vector at the antenna array at time instant n
$\mathbf{x}(n)$	noise-contaminated input vector to beamformer i.e. noise-contaminated received signal at the antenna array at time instant n
$\mathcal{Y}^{(+)}$	set consists of real possible noise-free output vector of beamformer given the desired user's transmitted bit $b_s(n) = +1$
$\mathcal{Y}^{(\pm)}$	set consists of possible noise-free output of beamformer given the desired user's transmitted bit $b_s = \pm 1$
$y(n)$	received output of beamformer at time instant n
y_s	signed decision function
$\bar{y}_{R,q}$	real part of noise-free output of beamformer due to one of N_b possible input vector $\mathbf{x}_q$, where $1 \leq q \leq N_b$

$\bar{y}_{R,q}^{(+)}$	real part of noise-free output of beamformer due to one of N_b possible input vector $\mathbf{x}_q$, where $1 \leq q \leq N_b$, given the desired user's transmitted bit vector $\mathbf{b}_s = +1$
$\bar{y}_q$	noise-free output of beamformer due to one of N_b possible input vector, where $1 \leq q \leq N_b$
$\mathbf{z}$	cross-correlation vector of the array output $\mathbf{x}(n)$ and the conjugate of the reference signal $r^*(n)$

Glossary

AIM-CGM	Adaptive Interference Mitigation using the conjugate gradient method
ALBER	Approximate Least Bit Error Rate algorithm
AOA	angle of arrival
AWGN	Additive White Gaussian Noise
BACG	Block Adaptive Conjugate Gradient algorithm
BER	Bit Error Ratio
BPSK	Binary Phase Shift Keying
BS	base station
BWA	broadband wireless access
CCI	co-channel interference
CDMA	Code Division Multiple Access
CGM	Conjugate Gradient Method
CIR	Channel Impulse Response
CMA	Constant Modulus Algorithm
DBF	digital beamforming
DDA	Decision Directed Algorithm
DECT	Digital European Cordless Telephone
DFE	decision feedback equaliser
DFT	Discrete Fourier Transform
DMI	Direct Matrix Inversion algorithm
DOA	direction of arrival
DOF	degrees of freedom
DS/CDMA	Direct-Sequence/Code Division Multiple Access
DSD	Differential Steepest Descent algorithm
EGC	equal gain combining
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Technique algorithm
FDMA	Frequency Division Multiple Access
FDD	frequency division duplex

FFT	Fast Fourier Transform
GA	Genetic Algorithm
GDIRD	General Data Independent Response Design
GSC	Generalised Sidelobe Canceller
GSM	global system for mobile communication
IEEE	Institute of Electrical and Electronics Engineers
IF	intermediate frequency
INR	Interference-to-Noise Ratio
IS-36	Interim Standard-36
IS-54	Interim Standard-54
IS-95	Interim Standard-95
ISI	intersymbol interference
ITU	International Telecommunication Union
LBER	Least Bit Error Rate algorithm
LCMV	Linearly Constrained Minimum Variance
LMS	Least Mean Square algorithm
LOS	line of sight
LRS	Linear Random Search algorithm
LSE	Least Squares Error
MAC	Multiply-and-ACcumulates
MBER	Minimum Bit Error Ratio
MEM	Maximum Entropy Method
MIMO	Multiple Input Multiple Output
MLM	Maximum Likelihood Method
MLE	Maximum Likelihood Estimation
MMSE	Minimum Mean Square Error
MRC	maximal ratio combining
MSB	most significant bit
MSD	multistage multiuser detector
MSE	Mean Square Error
MSN	Maximum Signal-to-Noise ratio
MUD	multiuser detection
MUSIC	MUltiple SIgnal Classification algorithm
MV	minimum variance
MVDR	Minimum Variance Distortionless Response
NLMS	Normalised Least Mean Square algorithm
OFDM	Orthogonal Frequency Division Multiplexing

PDF	probability density function
PLL	phase-locked loop
PN	pseudo-noise
QAM	Quadrature Amplitude Modulation
QoS	quality of service
QPSK	Quadrature Phase Shift Keying
RF	radio-frequency
RLS	Recursive Least Square algorithm
RSMI	Recursive Sample Matrix Inversion algorithm
SCARP	Smart Communication Antenna Research Programme
SCORE	Spectral COherence REstoral algorithm
SDMA	Spatial Division Multiple Access
SFIR	Spatial Filtering for Interference Reduction
SIR	Signal-to-Interference Ratio
SINR	Signal-to-Interference-plus-Noise Ratio
SLC	sidelobe canceller
SMI	Sample Matrix Inversion algorithm
SNR	Signal-to-Noise Ratio
SOI	signal of interest
SQ-RLS	square-root RLS algorithm
TDD	time division duplex
TDMA	Time Division Multiple Access
TRT	temporal reference technique
TTCM	Turbo-Trellis-Coded Modulation
ULA	uniform linear array
UTRA	Universal Terrestrial Radio Access
VFF-NRLS	variable forgetting factor nonlinear RLS algorithm
WCDMA	wideband CDMA

Bibliography

[1] A. J. Paulraj and C. B. Papadias, "Space Time Processing for Wireless Communications," *IEEE Personal Communications*, vol. 14, pp. 49–83, November 1997.

[2] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-carrier Quadrature Amplitude Modulation.* New York, USA: John Wiley and Sons, 2000.

[3] W. C. Y. Lee, *Mobile Communications Engineering: Theory and Applications.* New York, USA: McGraw Hill, 1997.

[4] J. S. Blogh and L. Hanzo, *Third-Generation Systems and Intelligent Wireless Networking.* New York, USA: John Wiley and Sons, 2002.

[5] C. B. Dietrich Jr., W. L. Stutzman, B.-K. Kim, and K. Dietze, "Smart Antennas in Wireless Communications: Base-Station Diversity and Handset Beamforming," *IEEE Antennas and Propagation Magazine*, vol. 42, no. 5, pp. 142–151, October 2000.

[6] S. Haykin, *Adaptive Filter Theory.* New Jersey, USA: Prentice Hall, 1996.

[7] T. S. Rappaport, *Wireless Communications: Principles and Practice.* New Jersey, USA: Prentice Hall, 1996.

[8] J. Litva and T. K.-Y. Lo, *Digital Beamforming in Wireless Communications.* Boston, USA: Artech House Publishers,Inc., 1996.

[9] L. C. Godara, "Applications of Antenna Arrays to Mobile Communications, Part I: Performance Improvement, Feasibility, and System Considerations," in *Proceedings of the IEEE*, vol. 85, pp. 1031–1060, July 1997.

[10] J. H. Winters, J. Salz, and R. D. Gitlin, "The Impact of Antenna Diversity on the Capacity of Wireless Communications System," *IEEE Transactions on Communications*, vol. 42, pp. 1740–1751, February/March/April 1994.

[11] M. Chryssomallis, "Smart Antennas," *IEEE Antennas and Propagation Magazine*, vol. 42, pp. 129–136, June 2000.

[12] L. Hanzo, M. Münster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting.* New York, USA: John Wiley and IEEE Press, 2003.

[13] G. V. Tsoulos, "Smart Antennas for Mobile Communication Systems: Benefits and Challenges," *IEE Electronics and Communication Engineering Journal*, vol. 11, pp. 84–94, April 1999.

[14] G. V. Tsoulos, M. A. Beach, and S. C. Swales, "Adaptive Antennas for Third Generation DS-CDMA Cellular Systems," *IEEE Vehicular Technology Conference*, vol. 45, pp. 45–49, 1995.

[15] L. Hanzo, L.-L. Yang, E.-L. Kuan, and K. Yen, *Single- and Multi-Carrier DS-CDMA: Multi-User Detection, Space-Time Spreading, Synchronisation, Standards and Networking.* New York, USA: John Wiley and IEEE Press, 2003.

[16] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space-Time Coding for Transmission over Fading Channels.* New York, USA: John Wiley and IEEE Press, 2002.

[17] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multi-Element Antennas," *Bell Labs Technical Journal*, pp. 41–59, Autumn 1996.

[18] Beam Reach Networks, *Smart Antenna Systems.* http://www.iec.org/online/tutorials/smart_ant/: World Wide Web, 1998.

[19] G. V. Tsoulos, M. A. Beach, and J. P. McGeehan, "Wireless Personal Communications for the 21st Century: European Technological Advances in Adaptive Antennas," *IEEE Communications Magazine*, vol. 35, pp. 102–109, September 1997.

[20] R. H. Roy, "An Overview of Smart Antenna Technology: The Next Wave in Wireless Communications," in *IEEE Proceedings of the Aerospace Conference*, vol. 3, pp. 339–345, 21-28 March 1998.

[21] A. U. Bhobe and P. L. Perini, "An Overview of Smart Antenna Technology for Wireless Communication," in *IEEE Proceedings of the Aerospace Conference*, vol. 2, pp. 2/875–2/883, 10-17 March 2001.

[22] J. H. Winters, "Smart Antennas for Wireless Systems," *IEEE Personal Communications*, vol. 1, pp. 23–27, February 1998.

[23] R. T. Derryberry, S. D. Gray, D. M. Ionescu, G. Mandyam, and B. Raghothaman, "Transmit Diversity in 3G CDMA Systems," *IEEE Communications Magazine*, vol. 40, pp. 68–75, April 2002.

[24] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On Beamforming with Finite Rate Feedback in Multiple-Antenna Systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2579, October 2003.

[25] S. D. Blostein and H. Leib, "Multiple Antenna Systems: Their Role and Impact in Future Wireless Access," *IEEE Communications Magazine*, vol. 41, no. 7, pp. 94–101, July 2003.

[26] S. Zhou and G. B. Giannakis, "Optimal Transmitter Eigen-Beamforming and Space-Time Block Coding Based on Channel Correlations," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1673–1690, July 2003.

[27] D. Bartolome and A. I. Perez-Neira, "MMSE Techniques for Space Diversity Receivers in OFDM-Based Wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 151–160, February 2003.

[28] J. Li, K. B. Letaief, and Z. Cao, "Co-channel Interference Cancellation for Space-Time Coded OFDM Systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 1, pp. 41–49, January 2003.

[29] R. Negi, A. M. Tehrani, and J. M. Cioffi, "Adaptive Antennas for Space-Time Codes in Outdoor Channels," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 1918–1925, December 2002.

[30] K.-H. Li and M. A. Ingram, "Space-Time Block-Coded OFDM Systems with RF Beamformers for High-Speed Indoor Wireless Communications," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 1899–1901, December 2002.

[31] S. Zhou and G. B. Giannakis, "Optimal Transmitter Eigen-Beamforming and Space-Time Block Coding Based on Channel Mean Feedback," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2599–2613, October 2002.

[32] M. G. M. Hussain, "Principles of Space-Time Array Processing for Ultrawide-band Impulse Radar and Radio Communications," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 3, pp. 393–403, March 2002.

[33] G. Jongren, M. Skoglund, and B. Ottersten, "Combining Beamforming and Orthogonal Space-Time Block Coding," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 611–627, March 2002.

[34] A. T. Alastalo and M. Kahola, "Smart-Antenna Operation for Indoor Wireless Local-Area Networks Using OFDM," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 392–399, March 2003.

[35] S. Thoen, L. Deneire, L. V. der Perre, M. Engels, and H. D. Man, "Constrained Least Squares Detector for OFDM/SDMA-Based Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 2, no. 1, pp. 129–140, Jan 2003.

[36] S. Thoen, L. V. der Perre, M. Engels, and H. D. Man, "Adaptive Loading for OFDM/SDMA-Based Wireless Networks," *IEEE Transactions on Communications*, vol. 50, no. 11, pp. 1798–1810, November 2002.

[37] J.-H. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint Transmitter Receiver Diversity for Efficient Space Division Multiaccess," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 16–27, January 2002.

[38] Y. Sigen and R. S. Blum, "Optimized Signalling for MIMO Interference Systems with Feedback," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2839–2848, November 2003.

[39] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, October 2003.

[40] M. C. Bromberg, "Optimizing MIMO Multipoint Wireless Networks Assuming Gaussian Other-User Interference," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2352–2362, October 2003.

[41] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, October 2003.

[42] D. J. Love and R. W. Heath Jr., "Equal Gain Transmission in Multiple-Input Multiple-Output Wireless Systems," *IEEE Transactions on Communications*, vol. 51, no. 7, pp. 1102–1110, July 2003.

[43] D. P. Palomar and M. A. Lagunas, "Joint Transmit-Receive Space-Time Equalization in Spatially Correlated MIMO Channels: A Beamforming Approach," *IEEE Transactions on Selected Areas in Communications*, vol. 21, no. 5, pp. 740–743, June 2003.

[44] J. Chung, C.-S. Hwang, K. Kim, and Y. K. Kim, "A Random Beamforming Technique in MIMO Systems Exploiting Multiuser Diversity," *IEEE Transactions on Selected Areas in Communications*, vol. 21, no. 5, pp. 848–855, June 2003.

[45] M. Kang and M.-S. Alouini, "Largest Eigenvalue of Complex Wishart Matrices and Performance Analysis of MIMO MRC Systems," *IEEE Transactions on Selected Areas in Communications*, vol. 21, no. 3, pp. 418–426, April 2003.

[46] S. H. Simon and A. L. Moustakas, "Optimizing MIMO Antenna Systems with Channel Covariance Feedback," *IEEE Transactions on Selected Areas in Communications*, vol. 21, no. 3, pp. 406–417, April 2003.

[47] R. Knopp and G. Caire, "Power Control and Beamforming for Systems with Multiple Transmit and Receive Antennas," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 638–648, October 2002.

[48] P. W. Howells, "Intermediate Frequency Sidelobe Canceller," tech. rep., U. S. Patent 3202990, May 1959.

[49] W. F. Gabriel, "Adaptive Processing Array Systems," in *Proceedings of the IEEE*, vol. 80, pp. 152–162, January 1992.

[50] Special Issue on Active and Adaptive Antennas edited by R. C. Hansen, *IEEE Transactions on Antennas and Propagation*, vol. AP-12. March 1964.

[51] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive Antenna Systems," in *Proceedings of the IEEE*, vol. 55, pp. 2143–2159, December 1967.

[52] Special Issue on Adaptive Antennas edited by W. F. Gabriel, *IEEE Transactions on Antennas and Propagation*, vol. AP-24. September 1976.

[53] S. P. Applebaum, "Adaptive Arrays," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 585–598, September 1976.

[54] L. J. Griffiths, "A Simple Adaptive Algorithm for Real-Time Processing in Antenna Arrays," in *Proceedings of the IEEE*, vol. 57, pp. 1696–1704, October 1969.

[55] T. K. Sarkar, J. Koh, R. Adve, R. A. Schneible, M. C. Wicks, S. Choi, and M. S. Palma, "A Pragmatic Approach to Adaptive Antennas," *IEEE Antennas and Propagation Magazine*, vol. AP-42, pp. 39–55, April 2000.

[56] J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," in *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, August 1969.

[57] D. H. Johnson, "The Application of Spectral Estimation Methods to Bearing Estimation Problem," in *Proceedings of the IEEE*, vol. 70, pp. 1018–1028, September 1982.

[58] O. L. Frost III, "An Algorithm for Linearly Constrained Adaptive Array Processing," in *Proceedings of the IEEE*, vol. 60, pp. 926–935, August 1972.

[59] L. J. Griffiths, "Signal Extraction Using Real-Time Adaptation of a Linear Multichannel Filter," tech. rep., Stanford Electronic Laboratory, January 1968.

[60] L. E. Brennan and I. S. Reed, "Theory of Adaptive Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-9, pp. 237–252, March 1973.

[61] L. E. Brennan, E. L. Pugh, and I. S. Reed, "Control Loop Noise in Adaptive Array Antennas," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-7, pp. 254–262, March 1971.

[62] L. E. Brennan and I. S. Reed, "Effect of Envelope Limiting in Adaptive Array Control Loops," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-7, pp. 698–700, July 1971.

[63] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Transactions on Aerospace Electronic System*, vol. AES-10, pp. 853–863, November 1974.

[64] B. Widrow and J. M. McCool, "A Comparison of Adaptive Algorithms Based on the Method of Steepest Descent and Random Search," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 615–637, September 1976.

[65] C. Darwin, *On the Origin of Species*. London, UK: John Murray, 1859.

[66] L. E. Brennan, J. Mallet, and I. S. Reed, "Adaptive Arrays in Airborne MTI Radar," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 607–615, September 1976.

[67] S. P. Applebaum and D. J. Chapman, "Adaptive Arrays with Main Beam Constraints," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 650–662, September 1976.

[68] K. Takao, M. Fujita, and T. Nishi, "An Adaptive Array Under Directional Constraint," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 662–669, September 1976.

[69] R. T. Compton Jr., R. J. Huff, W. G. Swarner, and A. A. Ksienski, "Adaptive Arrays for Communication Systems: An Overview of Research at the Ohio State University," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 599–607, September 1976.

[70] R. T. Compton Jr., "Adaptive Arrays with Constraint Nulls," tech. rep., Report, The Ohio State University ElectroScience Laboratory, September 1972.

[71] K. L. Reinhard, "Adaptive Array Performance with Specular Multipath," tech. rep., Report, The Ohio State University ElectroScience Laboratory, February 1974.

[72] G. E. Kolezar, "The Stochastic Properties of the Weights in an Adaptive Antenna Array," tech. rep., Report, The Ohio State University ElectroScience Laboratory, August 1975.

[73] D. J. Chapman, "Partial Adaptivity for the Large Array," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 685–696, September 1976.

[74] Special Issue on Adaptive Antennas edited by W. F. Gabriel, *IEEE Transactions on Antennas and Propagation*, vol. AP-34. March 1986.

[75] C. R. Ward, P. J. Hargrave, and J. G. McWhirter, "A Novel Algorithm and Architecture for Adaptive Digital Beamforming," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 338–346, March 1986.

[76] Y. L. Su, T. J. Shan, and B. Widrow, "Parallel Spatial Processing: A Cure for Signal Cancellation in Adaptive Arrays," *IEEE Transactions on Antennas Propagation*, vol. AP-34, pp. 347–355, March 1986.

[77] L. C. Godara and A. Cantoni, "Analysis of Constrained LMS Algorithm with Application to Adaptive Beamforming Using Perturbation Sequences," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 368–379, March 1986.

[78] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas Propagation*, vol. AP-34, pp. 276–280, July 1986.

[79] G. Bienvenu and L. Kopp, "Adaptivity to Background Noise Spatial Coherence for High Resolution Passive Methods," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. AP-5, pp. 307–310, April 1980.

[80] A. Paulraj, R. Roy, and T. Kailath, "A Subspace Rotation Approach to Signal Parameter Estimation," in *Proceedings of the IEEE*, vol. 74, pp. 1044–1046, July 1986.

[81] L. C. Godara, "Application of Antenna Arrays to Mobile Communications, Part II: Beam-Forming and Direction-of-Arrival Considerations," in *Proceedings of the IEEE*, vol. 85, pp. 1195–1245, August 1997.

[82] P. Charge, Y. Wang, and J. Saillard, "A root-MUSIC Algorithm for Noncircular Sources," in *IEEE Acoustic, Speech and Signal Processing Proceedings*, vol. 5, pp. 2985–2988, 2001.

[83] R. Roy and T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-37, pp. 984–995, July 1989.

[84] D. R. Morgan, "Partially Adaptive Array Techniques," *IEEE Transactions on Antennas and Propagation*, vol. AP-26, pp. 823–833, November 1978.

[85] W. F. Gabriel, "Spectral Analysis and Adaptive Array Superresolution Techniques," in *Proceedings of the IEEE*, vol. 68, pp. 654–666, June 1980.

[86] J. T. Mayhan, A. J. Simmons, and W. C. Cummings, "Wide-Band Adaptive Antenna Nulling Using Tapped Delay Lines," *IEEE Transactions on Aerospace Electronic System*, vol. AES-29, pp. 923–936, November 1981.

[87] R. A. Mucci, "A Comparison of Efficient Beamforming Algorithms," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-32, no. 3, pp. 548–558, June 1984.

[88] B. D. V. Veen and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE Acoustic, Speech and Signal Processing Magazine*, pp. 4–24, April 1988.

[89] S. Anderson, M. Millnert, M. Viberg, and B. Wahlberg, "An Adaptive Array for Mobile Communication Systems," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 230–236, February 1991.

[90] P.-R. Chang, W.-H. Yang, and K.-K. Chan, "A Neural Network Approach to MVDR Beamforming Problem," *IEEE Transactions on Antennas and Propagation*, vol. AP-40, pp. 313–322, March 1992.

[91] K. M. Kim, I. W. Cha, and D. H. Youn, "On the Performance of the Generalized Sidelobe Canceller in Coherent Situations," *IEEE Transactions on Antennas and Propagation*, vol. AP-40, pp. 465–468, April 1992.

[92] J. Li, "Improving ESPRIT via Beamforming," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 2, pp. 520–528, April 1992.

[93] B. D. Woerner, J. H. Reed, and T. S. Rappaport, "Simulation Issues for Future Wireless Modems," *IEEE Communications Magazine*, vol. 32, pp. 42–53, July 1994.

[94] M. Barrett and R. Arnott, "Adaptive Antennas for Mobile Communications," *IEE Electronics & Communication Engineering Journal*, vol. 6, pp. 203–214, August 1994.

[95] A. Klouche-Djedid and M. Fujita, "Adaptive Array Sensor Processing Applications for Mobile Telephone Communications," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 405–416, August 1996.

[96] T. S. Rappaport, *Smart Antennas: Adaptive Arrays, Algorithms and Wireless Position Location.* New Jersey, USA: IEEE Press, 1998.

[97] K. Sheikh, D. Gesbert, D. Gore, and A. Paulraj, "Smart Antennas for Broadband Wireless Access Networks," *IEEE Communications Magazine*, vol. 37, no. 11, pp. 100–105, November 1999.

[98] A. O. Boukalov and S.-G. Häggman, "System Aspects of Smart-Antenna Technology in Cellular Wireless Communications-An Overview," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, no. 6, pp. 919–929, June 2000.

[99] S. Bellofiore, C. A. Balanis, J. Foutz, and A. S. Spanias, "Smart-Antenna Systems for Mobile Communication Networks, Part 1: Overview and Antenna Design," *IEEE Antennas and Propagation Magazine*, vol. 44, no. 3, pp. 145–154, June 2002.

[100] S. Bellofiore, J. Foutz, C. A. Balanis, and A. S. Spanias, "Smart-Antenna Systems for Mobile Communication Networks, Part 2: Beamforming and Network Throughput," *IEEE Antennas and Propagation Magazine*, vol. 44, no. 4, pp. 106–114, August 2002.

[101] M. T. Ma, *Theory and Application of Antenna Arrays.* New York, USA: John Wiley and Sons, 1974.

[102] B. D. Steinberg, *Principles of Aperture and Array System Design: Including Random and Adaptive Arrays.* New York, USA: John Wiley and Sons, 1976.

[103] R. Monzingo and T. Miller, *Introduction to Adaptive Arrays.* New York, USA: John Wiley and Sons, 1980.

[104] J. E. Hudson, *Adaptive Array Principles.* New York, USA: Peter Peregrinus, 1981.

[105] J. D. Marr, "A Selected Bibliography on Adaptive Antenna Arrays," *IEEE Transactions on Aerospace Electronic System*, vol. AES-22, pp. 781–788, 1986.

[106] S. Haykin, *Array Signal Processing.* New Jersey, USA: Prentice Hall, 1985.

[107] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* New Jersey, USA: Prentice Hall, 1985.

[108] L. H. Sibul, *Adaptive Signal Processing.* New Jersey, USA: IEEE Press, 1987.

[109] R. T. Compton Jr., *Adaptive Antennas: Concepts and Performance.* New Jersey, USA: Prentice Hall, 1988.

[110] S. S. Haykin, A. Steinhardt, and S. Haykin, *Adaptive Radar Detection and Estimation.* New York, USA: John Wiley and Sons, 1992.

[111] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques.* New Jersey, USA: Prentice Hall, 1993.

[112] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance.* New Jersey, USA: Prentice Hall, 1994.

[113] H. Krim and M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Processing Magazine*, pp. 67–94, July 1996.

[114] R. Kohno, "Spatial and Temporal Communication Theory Using Adaptive Antenna Array," *IEEE Personal Communications*, vol. 1, pp. 28–35, February 1998.

[115] G. V. Tsoulos, *Adaptive Antennas for Wireless Communications.* New Jersey, USA: IEEE Press, 2001.

[116] B. Pattan, *Robust Modulation Methods and Smart Antennas in Wireless Communications.* New Jersey, USA: Prentice Hall, 2000.

[117] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory.* New York, USA: John Wiley and Sons, 2002.

[118] W. E. Rodgers and R. T. Compton Jr., "Adaptive Array Bandwidth with Tapped Delay-Line Processing," tech. rep., Report, The Ohio State University ElectroScience Laboratory, May 1975.

[119] W. E. Rodgers and R. T. Compton Jr., "Adaptive Array Bandwidth with Tapped Delay Line Processing," *IEEE Transactions on Aerospace Electronic System*, vol. AES-15, pp. 21–28, 1979.

[120] R. N. Adams, L. L. Horowitz, and K. D. Senne, "Adaptive Main-Beam Nulling for Narrow-Beam Antenna Arrays," *IEEE Transactions on Aerospace Electronic System*, vol. AES-16, pp. 509–516, July 1980.

[121] W. F. Gabriel, "Using Spectral Estimation Techniques in Adaptive Processing Antenna Systems," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 291–300, March 1986.

[122] S. Prasad, "Linear Antenna Arrays with Broad Nulls with Applications to Adaptive Arrays," *IEEE Transactions on Antennas and Propagation*, vol. AP-27, pp. 185–190, March 1979.

[123] J. G. McWhirter, "Recursive Least-Squares Minimization Using a Systolic Array," in *Proceedings of the SPIE*, pp. 105–112, 1983.

[124] B. D. V. Veen, "Systolic Arrays for Partially Adaptive Beamforming," in *Proceedings of the 21st Asilomar Conference on Signals, Systems and Computers*, pp. 584–588, November 1987.

[125] R. S. H. Steyskal and R. Haupt, "Methods for Null Control and Their Effects on the Radiation Pattern," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 404–409, March 1986.

[126] J. G. McWhirter and T. J. Shepherd, "Systolic Array Processor for MVDR Beamforming," in *Proceedings of the IEE Radar and Signal Processing*, vol. 136, pp. 75–80, April 1989.

[127] C.-C. Yeh, W.-D. Wang, T.-H. S. Chao, and J. Mar, "Least-Squares Regenerative Hybrid Array for Adaptive Beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 38, pp. 489–497, April 1990.

[128] J. Ward and R. T. Compton Jr., "Sidelobe Level Performance of Adaptive Sidelobe Canceller Arrays with Element Reuse," *IEEE Transactions on Antennas and Propagation*, vol. AP-38, pp. 1684–1693, October 1990.

[129] Y.-H. Chen and C.-H. Chen, "A New Structure for Adaptive Broad-Band Beamforming," *IEEE Transactions on Antennas and Propagation*, vol. AP-39, pp. 551–555, April 1991.

[130] P. Petrus and J. H. Reed, "Time Dependent Adaptive Arrays," *IEEE Signal Processing Letters*, vol. 2, pp. 219–222, December 1995.

[131] C. N. Adkins and J. J. Tutora, "A Broadband Beam Former with Pole-Zero Unconstrained Jammer Rejection in Linear Arrays," *IEEE Transactions on Signal Processing*, vol. 44, pp. 438–441, December 1996.

[132] I. Chiba, R. Miura, T. Tanaka, and Y. Karasawa, "Digital Beam Forming (DBF) Antenna System for Mobile Communications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 12, pp. 31–41, September 1997.

[133] H. Yang and M. A. Ingram, "Design of Partially Adaptive Arrays Using the Singular-Value Decomposition," *IEEE Transactions on Antennas and Propagation*, vol. AP-45, pp. 843–850, May 1997.

[134] S. Weiss, R. W. Stewart, M. Schabert, I. K. Proudler, and M. W. Hoffman, "An Efficient Scheme for Broadband Adaptive Beamforming," in *Proceedings of the 33rd. Asilomar Conference on Signals, Systems & Computers*, vol. 1, (Pacific Grove), pp. 496–500, 24-27 October 1999.

[135] M.-L. Leou, C.-C. Yeh, and H.-J. Li, "A Novel Hybrid of Adaptive Array and Equalizer for Mobile Communications," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 1–10, January 2000.

[136] W. Liu, S. Weiss, and L. Hanzo, "A Novel Method for Partially Adaptive Broadband Beamforming," in *IEEE Proceedings of the Workshop on Signal Processing Systems (SIPS)*, (Antwerp, Belgium), pp. 361–372, 26-28 September 2001.

[137] H. H. Al-Khatib and R. T. Compton Jr., "A Gain Optimizing Algorithm for Adaptive Arrays," *IEEE Transactions on Antennas and Propagation*, vol. AP-26, pp. 228–235, March 1978.

[138] D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Transactions on Communications*, vol. 11, pp. 1867–1876, November 1980.

[139] J. Winters, "Spread Spectrum in a Four-Phase Communication System Employing Adaptive Antennas," *IEEE Transactions on Communications*, vol. 30, pp. 929–936, May 1982.

[140] L. C. Godara and A. Cantoni, "Analysis of the Performance of Adaptive Beam Forming Using Perturbation Sequences," *IEEE Transactions on Antennas and Propagation*, vol. AP-31, pp. 268–279, March 1983.

[141] J. M. Cioffi and T. Kailath, "Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-32, pp. 304–337, April 1984.

[142] I. Gupta, "SMI Adaptive Antenna Arrays for Weak Interfering Signals," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 1237–1242, October 1986.

[143] R. L. Moses and A. A. Beex, "Instrumental Variable Adaptive Array Processing," *IEEE Transactions on Aerospace Electronic System*, vol. 24, pp. 192–202, March 1988.

[144] B. Friedlander, "A Signal Subspace Method for Adaptive Interference Cancellation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1835–1845, December 1988.

[145] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral Self-coherence Restoral: A New Approach to Blind Adaptive Signal Extraction Using Antenna Arrays," in *Proceedings of the IEEE*, vol. 78, pp. 753–767, April 1990.

[146] J. R. Treichler and B. Agee, "A New Approach to Multipath Correlation of Constant Modulus," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-31, pp. 459–472, April 1993.

[147] S. Choi and D. H. Kim, "Adaptive Antenna Array Utilizing the Conjugate Gradient Method for Compensation of Multipath Fading in a Land Mobile Communication," in *IEEE Proceedings of the Vehicular Technology Conference*, vol. 1, pp. 33–36, 10-13 May 1992.

[148] K. Ohnishi and R. T. Milton, "A New Optimization Technique for Adaptive Antenna Arrays," *IEEE Transactions on Antennas and Propagation*, vol. AP-41, pp. 525–533, May 1993.

[149] R. L. Haupt, "Phase-Only Adaptive Nulling with a Genetic Algorithm," *IEEE Transactions on Antennas and Propagation*, vol. AP-45, pp. 1009–1015, June 1997.

[150] F. Rashid-Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint Optimal Power Control and Beamforming in Wireless Networks Using Antenna Arrays," *IEEE Transactions on Communications*, vol. 46, pp. 1313–1324, October 1998.

[151] T. Ratnarajah and A. Manikas, "An H^{∞} Approach to Mitigate the Effects of Array Uncertainties on the MuSIC Algorithm," *IEEE Signal Processing Letters*, vol. 5, pp. 185–188, July 1998.

[152] P. D. Karaminas and A. Manikas, "Super-Resolution Broad Null Beamforming for Cochannel Interference Cancellation in Mobile Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 689–697, May 2000.

[153] M. Vitale, G. Vesentini, N. N. Ahmad, and L. Hanzo, "Genetic Algorithm Assisted Adaptive Beamforming," *IEEE Vehicular Technology Conference*, vol. 1, pp. 601–605, Fall 1 2002.

[154] S. Chen, N. N. Ahmad, and L. Hanzo, "Smart Beamforming for Wireless Communications: A Novel Minimum Bit Error Rate Approach," in *Proceedings of 2nd. IMA International Conference on Mathematics in Communications*, (Lancaster, UK), 16-18 December 2002.

[155] S. Chen, N. N. Ahmad, and L. Hanzo, "Adaptive Minimum Bit Error Rate Beamforming," *to appear in IEEE Transactions on Wireless Communications*, 2004.

[156] S. Chen, L. Hanzo, and N. N. Ahmad, "Adaptive Minimum Bit Error Rate Beamforming Assisted Receiver for Wireless Communications," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. IV, pp. 640–643, 6-10 April 2003.

[157] A. Wolfgang, N. N. Ahmad, S. Chen, and L. Hanzo, "Genetic Algorithm Assisted Minimum Bit Error Rate Beamforming," in *IEEE Vehicular Technology Conference*, (Milan, Italy), 11-14 May 2004.

[158] A. Wolfgang, N. N. Ahmad, S. Chen, and L. Hanzo, "Genetic Algorithm Assisted Error Probability Optimisation for Beamforming," *IEE Electronics Letters*, vol. 40, pp. 320–322, 4 March 2004.

[159] S. Chen, L. Hanzo, N. N. Ahmad, and A. Wolfgang, "Adaptive Minimum Bit Error Rate Beamforming Assisted QPSK Receiver," in *Proceedings of International Conference on Communications (ICC)*, vol. 6, (Paris, France), pp. 3389–3393, 20-24 June 2004.

[160] Y. S. Yeh and D. O. Reudink, "Efficient Spectrum Utilization for Mobile Radio Systems Using Space Diversity," *IEEE Radio Spectrum Conservation Techniques Conference*, pp. 12–16, 1980.

[161] C. Balanis, *Antenna Theory: Analysis and Design.* New York, USA: John Wiley and Sons, 1982.

[162] E. Brookner and J. M. Howell, "Adaptive-adaptive Array Processing," in *Proceedings of the IEEE*, vol. 74, pp. 602–604, 1986.

[163] J. Gobert, "Adaptive Beam Weighting," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 744–749, September 1976.

[164] M. Ghavami, "Wideband Smart Antenna Theory Using Rectangular Array Structures," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2143–2151, September 2002.

[165] SYMENA Software & Consulting GmbH, *Smart Antennas - A Technical Introduction.* http://www.symena.com/capesso_orange.html: World Wide Web, 2001.

[166] S. Weiss and R. W. Stewart, *On Adaptive Filtering in Oversampled Subbands.* Germany: Shaker Verlag, 1998.

[167] G. Sivaradje, P. Dananjayan, and P. R. M. Kumar, "Prevention of Signal Cancellation for Coherent Signal and Interference in Adaptive Beamforming," *International Conference on Telecommunications*, May 2000.

[168] J. G. Proakis, *Digital Communications.* Singapore: McGraw Hill, 1995.

[169] K. Weltner, J. Grosjean, P. Schuster, and W. Weber, *Mathematics for Engineers and Scientists.* Cheltenham, UK: Stanley Thornes, 1994.

[170] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing.* New Jersey, USA: Prentice Hall, 2000.

[171] J. Fernandez, I. R. Corden, and M. Barrett, "Adaptive Array Algorithms for Optimal Combining in Digital Mobile Communications Systems," *IEE 8th International Conference on Antennas and Propagation*, pp. 983–986, 1993.

[172] S. H. Leung and C. F. So, "Nonlinear RLS Algorithm Using Variable Forgetting Factor in Mixture Noise," *International Conference on Acoustics, Speech and Signal Processing*, vol. 6, pp. 9777–9780, May 2001.

[173] Special Issue on Spectral Estimation edited by S. Haykin and J. A. Cadzow, *Proceedings of the IEEE*, vol. 70. September 1982.

[174] E. A. Robinson, "A Historical Perspective of Spectrum Estimation," *Proceedings of the IEEE*, vol. 70, pp. 885–907, September 1982.

[175] S. Haykin and J. A. Cadzow, "The Special Issue on Spectrum Estimation," *Proceedings of the IEEE*, vol. 70, pp. 883–884, September 1982.

[176] J. H. Winters, "Optimum Combining in Digital Mobile Radio with Cochannel Interference," *IEEE Transactions on Vehicular Technology*, vol. 33, pp. 144–155, August 1984.

[177] A. Margarita, S. J. Flores, L. Rubio, V. Almenar, and J. L. Corral, "Application of MUSIC for Spatial Reference Beamforming for SDMA in a Smart Antenna for GSM and DECT," *IEEE Vehicular Technology Conference*, vol. 1, pp. 123–126, May 2001.

[178] K. Watanabe, I. Yoshii, and R. Kohno, "Hybrid Adaptive Algorithm Based on Temporal Update and Spatial Spectrum Estimation for Adaptive Array Antenna," *IEEE Vehicular Technology Conference*, pp. 175–180, September 2000.

[179] T. Chen, "Highlights of Statistical Signal and Array Processing," *IEEE Signal Processing Magazine*, vol. 15, pp. 21–64, September 1998.

[180] M. I. Skolnik, *Radar Applications.* New York, USA: IEEE Press, 1988.

[181] M. I. Skolnik, *Introduction to Radar Systems.* New York, USA: McGraw Hill, 1962.

[182] V. V. Olśhevskii, *Statistical Method in Sonar.* New York, USA: Consultant Bureau, 1978.

[183] D. Shim and F. Alam, "A New Adaptive Downlink Beamforming Method for WCDMA System," *IEEE Vehicular Technology Conference*, vol. 1, pp. 157–161, May 2001.

[184] K. Hugl, J. Laurila, and E. Bonek, "Downlink Beamforming for Frequency Division Duplex Systems," *IEEE Global Communications (GLOBECOM) Conference*, pp. 2097–2101, 1999.

[185] W. C. Y. Lee, *Mobile Cellular Telecommunications: Analog and Digital Systems.* New York, USA: McGraw Hill, 1995.

[186] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third Generation Cellular and WATM Systems.* New York, USA: John Wiley and Sons and IEEE Press, 1999.

[187] S. Red, M. Oliphant, and M. Weber, *An Introduction to GSM.* London, UK: Artech House, 1995.

[188] R. Bernhardt, "The Use of Multiple Beam Directional Antenna in Wireless Messaging Systems," in *IEEE Vehicular Technology Conference*, pp. 858–861, August 1995.

[189] E. Tiirola and J. Ylitalo, "Performance Evaluation of Fixed-Beam Beamforming in WCDMA Downlink," *IEEE Vehicular Technology Conference*, vol. 2, pp. 700–704, May 2000.

[190] S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan, "The Performance Enhancement of Multibeam Adaptive Base-Station Antennas for Cellular Land Mobile Radio Systems ," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 56–67, February 1990.

[191] ITU Corporate Communication Unit, *International Telecommunication Union.* http://www.itu.int/aboutitu/: World Wide Web, 2002.

[192] A. Czylwik and T. Matsumoto, "Downlink Beamforming for Frequency Duplex Systems in Frequency Selective Fading," *IEEE Vehicular Technology Conference*, vol. 2, pp. 695–699, May 2000.

[193] M. Schubert and H. Boche, "An Efficient Algorithm for Optimum Joint Downlink Beamforming and Power Control," in *IEEE Vehicular Technology Conference*, May 2002.

[194] Y.-K. Choi, J.-K. Han, D.-H. Kim, C.-Y. Ahn, and D. K. Kim, "Performance Evaluation of Joint Beamforming and Power Control CDMA System with RAKE Combining," in *IEEE Vehicular Technology Conference*, May 2002.

[195] C. Gao and E. Sousa, "Forward Link Interference Suppression Using Multiple Element Adaptive Array at Mobile Terminals," *IEEE Vehicular Technology Conference*, vol. 2, pp. 922–925, May 2000.

[196] C.-C. Yeh and J. R. Barry, "Adaptive Minimum Bit-Error Rate Equalization for Binary Signalling," *IEEE Transactions on Communications*, vol. 48, pp. 1226–1235, July 2000.

[197] B. Mulgrew and S. Chen, "Adaptive Minimum-BER Decision Feedback Equalizers for Binary Signalling," *IEEE Transactions on Signal Processing*, vol. 81, pp. 1479–1489, 2001.

[198] S. Chen, A. K. Samingan, B. Mulgrew, and L. Hanzo, "Adaptive Minimum-BER Linear Multiuser Detection for DS-CDMA Signals in Multipath Channels," *IEEE Transactions on Signal Processing*, vol. 49, pp. 1240–1247, June 2001.

[199] A. K. Samingan, "Minimum Bit Error Rate Multiuser Detection Techniques for DS-CDMA," tech. rep., University of Southampton, United Kingdom, September 2003.

[200] S. Chen, B. Mulgrew, E. S. Chng, and G. Gibson, "Space Translation Properties and the Minimum-BER Linear-Combiner DFE," *IEE Proceeding on Communications*, vol. 145, pp. 316–322, October 1998.

[201] I. N. Psaromiligkos, S. N. Batalama, and D. A. Pados, "On Adaptive Minimum Probability of Error Linear Filter Receivers for DS-CDMA Channel," *IEEE Transactions on Communications*, vol. 47, pp. 1092–1102, July 1999.

[202] M. Y. Alias, *Minimum Bit Error Rate Multiuser Detection for Multiple Antenna Aided Uplink OFDM.* University of Southampton, United Kingdom: PhD thesis, School of Electronics and Computer Science, 2004.

[203] M. R. Aaron and D. W. Tufts, "Intersymbol Interference and Error Probability," *IEEE Transactions on Information Theory*, vol. IT-12, pp. 26–34, January 1966.

[204] E. Shamash and K. Yao, "On the Structure and Performance of a Linear Decision Feedback Equalizer Based on the Minimum Error Probability Criterion," in *Proceedings of the International Conference on Communications (ICC)*, (Minneapolis, Minnesota, USA), pp. 25F1–25F5, 17-19 June 1974.

[205] S. Chen, E. Chng, B. Mulgrew, and G. Gibson, "Minimum-BER Linear-Combiner DFE," in *Proceedings of International Conference on Communications (ICC)*, (Dallas, Texas, USA), pp. 1173–1177, 23-27 June 1996.

[206] C.-C. Yeh and J. R. Barry, "Approximate Minimum Bit-Error Rate Equalization for Binary Signaling," in *Proceedings of International Conference on Communications (ICC)*, vol. 1, (Montreal, Canada), pp. 1095–1099, 8-12 June 1997.

[207] N. B. Mandayam and B. Aazhang, "Gradient Estimation for Sensitivity Analysis and Adaptive Multiuser Interference Rejection in Code-Division Multiple-Access Systems," *IEEE Transactions on Communications*, vol. 45, no. 6, pp. 848–858, July 1997.

[208] C.-C. Yeh and J. R. Barry, "Approximate Minimum Bit-Error Rate Equalization for Pulse-Amplitude and Quadrature Amplitude Modulation," in *Proceedings of International Conference on Communications (ICC)*, vol. 1, (Atlanta, Georgia, USA), pp. 16–20, 7-11 June 1998.

[209] C.-C. Yeh, R. R. Lopes, and J. R. Barry, "Approximate Minimum Bit-Error Rate Multiuser Detection," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 1, (Sydney, Australia), pp. 3590–3595, 8-12 November 1998.

[210] S. Chen and B. Mulgrew, "The Minimum-SER Linear-Combiner Decision Feedback Equalizer," *IEE Proceedings on Communications*, vol. 146, no. 6, pp. 347–353, December 1999.

[211] I. N. Psaromiligkos, S. N. Batalama, and D. A. Pados, "On Adaptive Minimum Probability of Error Linear Filter Receivers for DS-CDMA Channels," *IEEE Transactions on Communications*, vol. 47, no. 7, pp. 1092–1102, July 1999.

[212] B. Mulgrew and S. Chen, "Stochastic Gradient Minimum-BER Decision Feedback Equalisers," in *Proceedings of IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control*, (Lake Louise, Alberta, Canada), pp. 93–98, 1-4 October 2000.

[213] X. Wang, W.-S. Lu, and A. Antoniou, "Constrained Minimum-BER Multiuser Detection," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2903–2909, October 2000.

[214] S. Chen, A. K. Samingan, B. Mulgrew, and L. Hanzo, "Adaptive Minimum-BER Linear Multiuser Detection," in *Proceedings of IEEE International Conference of Acoustics, Speech and Signal Processing (ICASSP)*, (Salt-Lake City, Utah, USA), pp. 2253–2256, 7-11 May 2001.

[215] A. K. Samingan, S. Chen, and L. Hanzo, "Adaptive Minimum-BER Linear Multiuser Detection for CDMA Signals in Multipath Channels with 4-QAM Constellation," *IEE Electronics Letters*, vol. 37, no. 11, pp. 721–723, 24 May 2001.

[216] S. Chen, B. Mulgrew, and L. Hanzo, "Least Bit Error Rate Adaptive Nonlinear Equalizers for Binary Signalling," *IEE Proceedings Communications*, vol. 150, no. 1, pp. 29–36, February 2003.

[217] R. C. de Lamare and R. Sampaio-Neto, "Adaptive MBER Decision Feedback Multiuser Receivers in Frequency Selective Fading Channels," *IEEE Communications Letters*, vol. 7, no. 2, pp. 73–75, February 2003.

[218] D. Gesbert, "Robust Linear MIMO Receivers: A Minimum Error-Rate Approach," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2863–2871, November 2003.

[219] M. Y. Alias, A. K. Samingan, S. Chen, and L. Hanzo, "Multiple Antenna Aided OFDM Employing Minimum Bit Error Rate Multiuser Detection," *IEE Electronics Letters*, vol. 39, no. 24, pp. 1769–1770, 27 November 2003.

[220] L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. West Sussex, England: John Wiley and IEEE Press, 2004.

[221] L. W. Couch II, *Digital and Analog Communication Systems.* New Jersey, USA: Prentice Hall, 1997.

[222] P. Guillaume and R. Pintelon, "A Gauss-Newton-Like Optimization Algorithm for Weighted Nonlinear Least Squares Problems," *IEEE Transactions on Signal Processing*, vol. 44, pp. 2222–2228, September 1996.

[223] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Non-linear Programming: Theory and Algorithm.* New York, USA: John Wiley and Sons, 1993.

[224] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, pp. 1066–1076, 1962.

[225] B. W. Silvermann, *Density Estimation for Statistics and Data Analysis.* New York, USA: Chapman and Hall, 1986.

[226] R. E. Ziemer and W. H. Tranter, *Principles of Communications.* Boston: Houghton Mifflin Company, 1990.

[227] S. Abedi and R. Tafazolli, "Genetically Modified Multiuser Detection for CDMA Systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 463–473, Feb 2002.

[228] K. Yen and L. Hanzo, "Genetic Algorithm Assisted Joint Multiuser Symbol Detection and Fading Channel Estimation for Synchronous CDMA Systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 985–998, June 2001.

[229] K. Yen and L. Hanzo, "Hybrid Genetic Algorithm Based Detection Schemes for Synchronous CDMA Systems," in *IEEE Vehicular Technology Conference*, vol. 2, pp. 1400–1404, May 2000.

[230] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning.* Reading, Massachusetts: Addison Wesley, 1989.

[231] D. A. Coley, *An Introduction to Genetic Algorithms for Scientists and Engineers.* River Edge, New Jersey, USA: World Scientific Publishing Co., Inc., 1999.

[232] M. Mitchell, *An Introduction to Genetic Algorithms.* Cambridge, Massachusetts: MIT Press, 1996.

[233] J. Holland, *Adaptation in Natural and Artificial Systems.* Ann Arbor, Michigan: University of Michigan Press, 1975.

[234] M. J. Juntti, T. Schlosser, and J. O. Lilleberg, "Genetic Algorithms for Multiuser Detection in Synchronous CDMA," in *Proceedings of IEEE Symposium on Information Theory*, (Ulm, Germany), p. 492, 29 June-4 July 1997.

[235] X. F. Wang and W.-S. Lu and A. Antoniou, "Genetic-algorithm-based Multiuser Detector for Multiple-access Communications," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 4, (Monterey, CA, USA), pp. 534–537, 31 May-3 June 1998.

[236] C. Ergun and K. Hacioglu, "Multiuser Detection Using a Genetic Algorithm in CDMA Communications Systems," *IEEE Transactions on Communications*, vol. 48, no. 8, pp. 1374–1383, August 2000.

[237] S. Abedi and R. Tafazolli, "A New CDMA Multiuser Detection Technique Using an Evolutionary Algorithm," *IEE Proceedings: Communications*, vol. 148, no. 6, pp. 393–399, December 2001.

[238] S. X. Ng, K. Yen, and L. Hanzo, "TTCM Assisted Genetic-algorithm Aided Reduced-Complexity Multiuser Detection," *IEE Electronics Letters*, vol. 38, no. 14, pp. 722–724, 4 July 2002.

[239] C.-T. Chiang and C.-Y. Chang, "An Improved Genetic Algorithm Based on Eugenic Population for Multiuser Detection in DS-CDMA Systems," in *Proceedings of IEEE Region 10 Annual International Conference (TENCON)*, vol. 2, (Beijing, China), pp. 984–987, 28-31 October 2002.

[240] M. G. Shayesteh, M. B. Menhaj, and B. G. Nobary, "A Modified Genetic Algorithm for Multiuser Detection in DS/CDMA," *IEICE Transactions on Communications*, vol. E86-B, no. 8, pp. 2377–2388, August 2003.

[241] K. Yen and L. Hanzo, "Antenna-diversity-assisted Genetic-algorithm-based Multiuser Detection Schemes for Synchronous CDMA Systems," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 366–370, March 2003.

[242] Y. Du and K. T. Chan, "Feasibility of Applying Genetic Algorithms in Space-time Block Coding Multiuser Detection Systems," in *Proceedings of the IASTED International Conference on Wireless and Optical Communications*, vol. 3, (Banff, Canada), pp. 469–473, 2-4 July 2003.

[243] M. Y. Alias, S. Chen, and L. Hanzo, "Genetic Algorithm Assisted Minimum Bit Error Rate Multiuser Detection in Multiple Antenna Aided OFDM," in *Proceedings of IEEE Vehicular Technology Conference (VTC-Fall)*, (Los Angeles, California, USA), 26-29 September 2004.

[244] E. A. Jones and W. T. Joines, "Genetic Design of Linear Antenna Arrays," *IEEE Antennas and Propagation Magazine*, vol. 42, pp. 92–100, June 2000.

[245] K. Yen, *Genetic Algorithm Assisted CDMA Multiuser Detection.* University of Southampton, United Kingdom: PhD thesis, Department of Electronics and Computer Science, 2000.

[246] P. Karamalis, A. Marousis, A. Kanatas, and P. Constantinou, "Direction of Arrival Estimation Using Genetic Algorithms," in *Proceedings of IEEE Vehicular Technology Conference (VTC-Spring)*, vol. 1, pp. 162–166, May 2001.

[247] J. P. Rennard, *Introduction to Genetic Algorithm.* http://www.rennard.org/alife/english/gavintrgb.htm World Wide Web, 2000.

[248] F. Leclerc and J.-Y. Potvin, "A Fitness Scaling Method Based on a Span Measure," in *IEEE International Conference on Evolutionary Computation*, vol. 2, pp. 561–565, 1995.

[249] S. Chen, B. Mulgrew, and P. M. Grant, "A Clustering Technique for Digital Communications Channel Equalisation Using Radial Basis Function Networks," *IEEE Transactions on Neural Networks*, vol. 4, pp. 570–579, 1993.

[250] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian Equaliser with Decision Feedback," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2918–2927, 1993.

[251] S. Chen, S. R. Gunn, and C. J. Harris, "The Relevance Vector Machine Technique for Channel Equalisation Application," *IEEE Transactions on Neural Networks*, vol. 12, pp. 1529–1532, Nov 2001.

[252] S. Chen, L. Hanzo, and A. Wolfgang, "Kernel-Based Nonlinear Beamforming Construction Using Orthogonal Forward Selection with the Fisher Ratio Class Separability," *IEEE Signal Processing Letters*, vol. 11, pp. 478–481, May 2004.

[253] M. Jiang and L. Hanzo, "Improved Hybrid MMSE Detection for Turbo-Trellis-Coded Modulation-Assisted Multiuser OFDM Systems," *IEE Electronics Letters*, vol. 40, pp. 1002–1003, August 2004.

Subject Index

Author Index

C

I

J

K

L

M

N

O

P

R

S

T

Z