INTRODUCTION

1.1 SOME REFLECTIONS ON CURRENT THOUGHTS

The fundamental bottleneck in mobile communication is that many users want to access the base station simultaneously and thereby establish the first link in the communication chain. The way the scarce resources of the base station are distributed to mobile users is through sharing. This is a technical definition of the term *multiple access*. Therefore, multiple accesses are implemented by sharing one or more of the four resources of the base station by the various mobile users randomly located in space and time. By *time* we imply that different users may start using the system at different times. This sharing can take place in any of the following four ways [1, 2]:

1.) *Bandwidth* (*Frequency Division Multiple Access* or in short, FDMA). Here, the frequency spectrum or the entire bandwidth is portioned off to different users and allocated for that communication duration. Hence each user communicates with the base station over an allocated narrow frequency band for the entire duration of the communication.

2.) *Time* (*Time Division Multiple Access* or in short, TDMA). Here, each mobile has the entire frequency resource of the base station for a short duration of the time (i.e., each user accesses the entire spectrum of the base station for a finite duration in an ordered sequence). With the advent of digital technology it is possible to have an intermittent connection for each mobile with the base station for a short period of time, and in this way the valuable frequency resource of the base station is shared.

3.) *Code* (*Code Division Multiple Access* or in short, CDMA). In this case, each user is assigned a unique code. In this way the user is allowed to access all the bandwidth, as in TDMA, and for the complete duration of the call, as in FDMA. All the users have access simultaneously to the entire spectrum for all the time. They are interfering with each other, and that is why this methodology was originally conceived as a covert mode
of communication. There are two main types of CDMA. One is called
*direct sequence spread spectrum multiple access*, and the other is called
*frequency hopped spread spectrum multiple access*. In the first case, two-way
communication is accomplished through spread spectrum modulation
where each user's digital waveform is spread over the entire frequency
spectrum that is allocated to that base station. Typically, on transmit, the
actual signal is coded and spread over the entire spectrum, where on
receive, the intended user first detects the signal by convolving the
received signal with his/her unique code and then demodulates the
convolved signal. In the second case the transmit carrier frequency
changes as a function of time in an ordered fashion so that the receiver can
decode each narrowband transmission. At first glance it appears that
CDMA is more complex than TDMA or FDMA, but with the advent of
novel digital chip design, it is easy to implement CDMA in hardware.

4.) *Space (Space Division Multiple Access)* or in short, SDMA. If a base
station has to cover a large geographical area, the region is split into cells
where the same carrier frequency can be reused in each cell. Therefore, for
a large number of cells there is a high level of frequency reuse, which
increases the capacity. In this primitive form the transmitted power of the
base station limits the number of cells that may be associated with a base
station since the level of interference at a base station is determined by the
spatial separation between cells, as the mobiles are using the same
frequency. This is one of the reasons that microcells and picocells have
been proposed for personal communication systems. However, it was soon
realized that the capacity of the base station could be increased further by
spatially focusing the transmitted energy along the direction of the
intended users. In this way, transmission can be achieved at the same
carrier frequency simultaneously with different users. This can be
accomplished by using an array of antennas at the base station and either a
switched beam array or a tracking beam array can be used to direct the
electromagnetic beam to the intended users.

In current times it appears that further enhancement in the capacity of a
communication system can be achieved primarily in the implementation of
SDMA. This is generally carried out using an adaptive process where we have a
collection of antennas called *phased arrays*. One now dynamically combines the
output from each antenna element using different weights. The weights modify
the amplitude and phase of the voltages received at each antenna element.
Through an appropriate combination of the voltages that are induced in them by
the incident electromagnetic fields, one forms an antenna beam. This antenna
beam can either be steered continuously or the beam can be switched along
certain prefixed directions by selecting a set of *a priori* weights. This can be
achieved in either of two ways.

The first way is to design an antenna with a narrow main beam. This is
generally implemented by using a physically large antenna, as the width of the
main lobe of the antenna is inversely proportional to the physical dimensions.
Hence, an electrically large antenna structure will have a very narrow beam and may also possess very low sidelobe levels. Creation of very low sidelobe levels may require extremely high tolerances in the variability of the actual physical dimensions of the radiating structures. This requires accurate design of the antenna elements in the phased array. Now one can mechanically steer this high-gain antenna to scan the entire geographical region of interest [3, 4]. This is actually done in developing the rotating antenna arrays in AWACS (Airborne Early Warning and Control System) radars [5]. Such a design makes the cost of the antennas very high. The other alternative is to use simple antenna elements such as dipoles, and then form the antenna beam by combining the received signals from a number of them by using a signal-processing methodology. This usually requires a receiver with an analog-to-digital converter (ADC) at the output of every antenna element, which also increases the cost. The signals from the antenna elements are now downconverted and sampled using an ADC, and then a digital beam-forming algorithm is used to form the main beam along the desired direction and to place nulls in the sidelobe regions along the direction of the interferers. The advantage of digital beam forming is that one can form any arbitrary low level of sidelobes with any width of the main lobe along the look direction [3, 4].

Historically, analog beam forming has been going on for a long time. Also, application of the Butler matrix to combine the outputs of the antenna elements is similar in principle to application of the fast Fourier transform (FFT) to the output voltages available at the antenna elements to form a beam [3, 4, 6]. This is because the far field is simply the Fourier transform of the induced current distribution on the radiating structure. Even though there is a one-to-one correspondence between the Butler matrix and the fast Fourier transform, there is an important fundamental difference. The Butler matrix processes the signals in the analog domain, whereas an FFT carries out similar processing in the digital domain. By processing signals in the analog domain, one is limited by the Rayleigh resolution criterion, which states that in order to resolve two closely spaced signals in space (i.e., their directions of arrival at the antenna array are very close to each other), one needs an antenna whose physical size is inversely proportional to the difference in the spatial angles of arrival at the array. Therefore, the closer two signals are located in space, the greater should be the physical size of the antenna in order to separate the two incoming signals. Therefore, the physical length of the array determines the angular resolution of a phased array performing analog processing. On the other hand, digital beam forming allows us to go beyond the curse of the Rayleigh limit if there is adequate signal strength and enough effective bits in the measured voltages (dynamic range) at each of the antenna elements to carry out beam forming [7].

Typically, adaptive beam forming is supposed to be synonymous with digital beam forming and smart antennas [1]. The term "smart antennas" implies that the antenna array can operate in any environment and has the capability to extract the signal of interest in the presence of interference and clutter and thus to adapt to the signal environment. However, a very important factor has been overlooked in the design process of adaptive systems. For example, if one observes a typical
cellular phone, the chip and the signal processors that have been used in the system were probably developed within the last year, but the key ingredient (i.e., the antenna) currently used in many systems was developed about 100 years ago by Hertz, as it is a modification of a simple dipole. Nowadays, the dipole is being replaced by some form of helix (bifilar or quadrifilar), which had been used in AM radios for almost 75 years. The same disparity in technology can also be observed in television sets. Even though a modern television set may have advanced components both for video displays and for processing the video and audio signals, the very high frequency (VHF) antenna is still the “rabbit ear”—a dipole, and the ultra high frequency (UHF) antenna is a loop which was developed in the early nineteen hundreds. The principle behind such wide disparities in component technologies of modern communication systems lies primarily in the assumption that an antenna captures a spatial-temporal signal propagating through space and transforms it into a pure temporal signal without any distortion. This assumes that the antennas are essentially isotropic omnidirectional point radiators. That is the reason why in contemporary literature the antenna is often referred to as a sensor of a temporal channel. In electromagnetics, the smallest source is an infinitesimally small dipole and it does not have an isotropic pattern, even though it is omnidirectional in certain planes.

An antenna to a spatial signal is equivalent to what an ADC is to a temporal signal. The purpose of an ADC is to produce high-fidelity temporal samples through the sample-and-hold mechanism of a temporal signal. For an ADC to be of good quality, it is essential that the sample time be much smaller than the hold period so that the sampled values provide a true representation of the analog signal. However, the quality of the ADC becomes questionable if the hold time is comparable to the sample time. In that case the temporal sample obtained from the ADC is not going to be representative of the true signal, as the ADC averages the output over the sample period, during which the signal of interest may have wide variations in amplitude. Under this scenario, where the hold time is comparable to the sample period, unless the effects of the ADC are removed through deconvolution, additional signal and data processing may not produce meaningful results.

This same problem arises in the practical application of antennas. An antenna is a spatial sampler of the electromagnetic fields propagating through space. A receiving antenna generally samples the electric field over its length and produces a voltage at the antenna terminals by integrating vectorially the electromagnetic fields incident upon it. When dealing with narrowband electromagnetic signals, a high-quality receiving antenna is often composed of an array of half-wavelength dipoles, typically spaced a half-wavelength apart [3–5]. So in an adaptive antenna environment, we are assuming the integrated value of the electrical field over a half wavelength to be equal to the actual value of the electric field at a point in space which corresponds to the feed point of the antenna. In other words, we are replacing the value of the incident electromagnetic field at the feed point of the antenna by a quantity that is the integral of the electromagnetic fields over a half wavelength in space. Thus by
comparing the performance of a finite-sized antenna in spatial sampling of electromagnetic fields to that of temporal sampling of a signal by an ADC, it is quite clear that unless the effects of the antenna are removed from the measured data, signal and data processing may not result in the desired output. This is due to the basic premise that the spatial integral value of the electric field along the half-wavelength antenna is representative of the actual value of the electric field at the antenna feed point. This is not correct. Hence, one of the objectives of this book is to merge the electromagnetic analysis with the signal processing [7–11]. Now one can implement adaptive processing using realistic antenna elements operating in close proximity and incorporate mutual coupling effects. Moreover, there may be coupling between the antenna elements and the platform on which it is mounted. In addition, there may be near-field scatterers, including other antennas, buildings, trees, and so on, near the array that may again distort the beam. In this book we present and illustrate methods for adaptive processing incorporating near-field electromagnetic effects.

When dealing with broadband signals, we often assume that the omni-directional isotropic point radiators have no effects on the signal. Such a simplistic assumption is seriously flawed. An antenna is not only a spatial sampler of the propagating electromagnetic field, it has a temporal response as well. It has a unique transfer function. For example, the far-field response from even an electrically small antenna is the result of a temporal differentiation of the driving time domain waveform. In addition, the radiated waveforms will have different signal shapes along different spatial directions. Moreover, an antenna of finite size will not mimic an omnidirectional point radiator in performance. On receive, an antenna vectorially integrates the spatial-temporal waveform that is incident on the structure. Therefore, unless the transfer function of the antenna is removed from the measured data, carrying out additional signal processing may not lead one to the correct solution to the problem at hand. This is not a simple problem, as the impulse response of both a transmitting and a receiving antenna of finite size is dependent simultaneously on both azimuth and elevation look directions. In a practical situation it is difficult to characterize the impulse response of either a transmitting or a receiving antenna, as it is difficult to know a priori at what azimuth and elevation angles the coupling is taking place. In broadband applications, the antenna responses must be accounted for. The easy way out in most practical systems in theory and in practice is to deal only with narrowband signals. One of the objectives of this book is to initiate a dialogue so that adaptive processing of the data collected through an antenna is performed in the correct fashion. Thus by combining electromagnetic analysis with signal processing, one can build toward a much more effective solution to the problem at hand.

A related problem in adaptive processing is that one often uses antenna elements that are very close to omnidirectional in nature. Even a dipole may have some directivity in elevation, but in azimuth it is still omnidirectional. This may not be an intelligent choice for cellular telephony, since a mobile will radiate most of the power in azimuth directions away from the intended user. The efficiency of a mobile communications system can be improved by using
INTRODUCTION

directive elements in a phased array. However, the problem with using directive elements is that it is not clear how to apply classical adaptive processing. Beamspace solutions offer one answer, but there may be others. One of the objectives of this book is to suggest adaptive systems with directive antenna elements and to illustrate how the measured outputs from directive elements can be combined when the directive element patterns of the antenna elements are properly oriented. Equivalently, the antenna elements in an array can be distributed nonuniformly to cover the physical structure and thereby further increase the radiation/receive efficiency. This is particularly useful in mobile systems, where the electromagnetic environment is not predictable nor may it be characterized in an accurate fashion. We address these issues and more in the following chapters.

1.2 ROADMAP OF THE BOOK

The book is organized as follows:

Chapter 2 provides a historical overview of Maxwell’s equations and presents some simple formulas to calculate the impulse response of selected canonical antennas. Our purpose is to demonstrate that even an infinitesimally small point source radiating a broadband signal in free space has a nontrivial impulse response and that their effects must be included in channel characterization. Measured results are also provided to illustrate impulse responses of some typical antennas and physical platforms over which they may be mounted.

In Chapter 3 we describe the anatomy of an adaptive process and present classical historical developments (i.e., statistically based methodologies where one needs to have an aggregate of the voltages at the antenna elements over some spatial-temporal duration). Appendix A further delineates the differences between a deterministic approach and a stochastic approach and illustrates the strengths and weaknesses of each.

In Chapter 4 we describe a direct data domain least squares (D³LS) approach, which operates, on a single snapshot of data. The advantages of using a D³LS over conventional stochastic methodologies are explained in Appendix A. There are various compounding factors such as nonstationarity of the data and real-time signal processing issues that are aided by a deterministic model, as it is well suited to applications in a highly dynamic environment where processing data on a snapshot-by-snapshot basis is appropriate. In addition, there is no need to develop a stochastic model for clutter, which in a direct data domain approach is treated as an undesired signal just like interferers and thermal noise. For a conventional adaptive system a snapshot is defined as the set of voltages measured at the terminals of the antennas. Both the interference and clutter in this algorithm are treated as undesired electromagnetic signals impinging on the array. Since no covariance matrix is formed in this method, a least squares method operating on a single snapshot of the data can be implemented in real time using a modern digital signal-processing device. To this end a survey of the
various forms of the adaptive conjugate gradient (CG) algorithm are presented in Appendix B. Their suitability when dealing with adaptive problems is also illustrated. In addition, several variants of this direct data domain least squares method can be implemented in parallel so that independent estimates of the same solution can be obtained. In reality, where the actual solution is unknown, different independent estimates can increase the level of confidence of the computed solution.

In Chapter 5 we illustrate how electromagnetic analysis can be utilized to correct for mutual coupling in an adaptive algorithm. Computed results with and without mutual coupling effects between the antenna elements are presented to illustrate the point that in a real system the finite size of the antenna must be accounted for. The efficiency of using a single snapshot of data for real-time processing is also discussed. It is shown that use of a direct data domain approach such as the Matrix Pencil along with an electromagnetic compensation technique leads to an accurate determination of the directions of arrival in a CDMA environment. The Matrix Pencil technique for both the one- and two-dimensional cases are described in Appendix C.

Chapter 6 demonstrates a two-step process implemented to extract the signal of interest in the presence of interferers and clutter when the antenna elements in the receiving array are nonuniformly spaced and operating in the presence of other near-field scatterers. The placement of the antenna elements in the array need not be coplanar. In addition, we illustrate how a conformal nonuniformly spaced microstrip patch array located on the side of an aircraft can be used for both direction finding and adaptive processing. Even though significant amounts of scattered energy are incident on the array from the wing and the fuselage, performance is not degraded and processing is carried out on a snapshot-by-snapshot basis. For these classes of radar-related problems, we generally assume that the direction of arrival of the signal of interest is known a priori. Additional examples are presented to illustrate how direction-of-arrival angle estimation can be carried out using conformal arrays on hemispherical and cylindrical surfaces having directive antenna elements with polarization diversity.

However, in a mobile communication environment we do not know a priori the direction of arrival of the signal of interest, and hence a different type of a priori knowledge about the signal is required. Chapter 7 describes the concept of cyclostationarity to illustrate how the D^3LS method can be applied to a set of received voltages at an antenna array which need not be coplanar and operates in the presence of mutual coupling and near-field scatterers. These techniques are called blind methods, as no training signals are necessary. We still carry out processing on a snapshot-by-snapshot basis, so this adaptive procedure is highly suitable in a dynamic environment.

Chapter 8 provides a survey of the various propagation models currently used in characterizing mobile communication channels. It includes both stochastic and numerical models.

Chapter 9 describes methods for optimizing the location of base stations for indoor wireless communications subject to a certain quality of service in a given
environment. A survey of the various optimization techniques is presented to illustrate what class of methods is well suited for these types of problems.

In Chapter 10 we present a frequency diversity technique that can identify and eliminate various multipath components without spatial diversity.

In Chapter 11 we describe a methodology for directing the signal from a base station to a specified mobile user while simultaneously placing nulls along the direction of the other mobiles utilizing the principles of reciprocity. The advantage of this technique is that directing the electromagnetic signals to an intended user is possible without any knowledge of the physical location of the antennas or the electromagnetic multipath environment in which the system is operating. It is not even necessary to know the spatial coordinates of the transmitter or receiver.

Chapter 12 illustrates the extension of the D³LS method to space-time adaptive processing (STAP). In this section the single snapshot-based direct data domain methodology is applied to the data collected by a side-looking airborne radar to detect a Saberliner aircraft in the presence of terrain and sea clutter. Several variations of the direct data domain methods are presented. Here also we use a single snapshot and model clutter in a deterministic fashion as unwanted electromagnetic signals. The voltages received at each antenna element in space and sampled in time corresponding to a single range cell characterize a single space-time snapshot corresponding to a specific range cell. Hence, in this approach we process the data on a range cell-by-range cell basis. Comparisons are also made with conventional statistical methods to illustrate the quality of the solution that can be obtained by applying this method to measured radar data. Another important factor is that direct data domain procedures require far fewer computational resources than a conventional stochastic covariance-based methodology. The direct method is further extended to carry out STAP using data from a circular array. Next, a hybrid STAP technique is described which utilizes the good points of both a direct method and a stochastic method. Finally, a knowledge-based STAP is described which is capable of automatically selecting the most appropriate method for a given data set.

The unique features of this book are:

1.) Electromagnetic analysis and signal-processing techniques are combined to analyze and design adaptive systems. Thus the presence of mutual coupling between antenna elements and the presence of near-field scatterers can be incorporated in the analysis.

2.) A direct data domain least squares algorithm is developed which processes the data on a snapshot-by-snapshot basis. Thus it is quite suitable for real-time implementation the use of a priori information either through the direction of arrival or through use of the concept of cyclostationarity and processing a single snapshot of data.

3.) The principle of reciprocity is exploited to direct a signal to a mobile user while simultaneously placing nulls along other directions without any spatial information about the base station or the mobile user or exact characterization of the electromagnetic environment in which they are operating.
4.) The direct data domain approach is extended to include space-time adaptive processing for dealing with side-looking radars to carry out filtering in space and time to detect weak signals in the presence of terrain and sea clutter using either linear or circular arrays. A knowledge-based STAP approach is described which is quite suitable for making use of a variety of algorithms, depending on the given data set. One advantage of this methodology is that it is transparent to a user.

REFERENCES