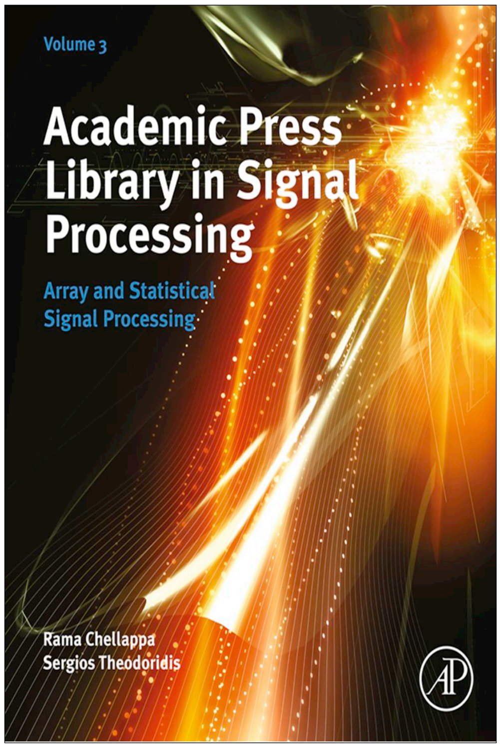


From A. Lee Swindlehurst et al., Applications of Array Signal Processing. In: Rama Chellappa and Sergios Theodoridis, editors, Academic Press Library in Signal Processing. Vol 3, Array and Statistical Signal Processing, Chennai: Academic Press, 2014, p. 859-953.
ISBN: 978-0-12-411597-2

CHAPTER

Applications of Array Signal Processing

20

A. Lee Swindlehurst[*], **Brian D. Jeffs**[†], **Gonzalo Seco-Granados**[‡], **and Jian Li**[§]

[*]*Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, USA*
[†]*Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT, USA*
[‡]*Department of Telecommunications and Systems Engineering, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*
[§]*Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA*

3.20.1 Introduction and background

The principles behind obtaining information from measuring an acoustic or electro-magnetic field at different points in space have been understood for many years. Techniques for long-baseline optical interferometry were known in the mid-19th century, where widely separated telescopes were proposed for high-resolution astronomical imaging. The idea that direction finding can be performed with two acoustic sensors has been around at least as long as the physiology of human hearing has been understood. The mathematical duality observed between sampling a signal either uniformly in time or uniformly in space is ultimately just an elegant expression of Einstein's theory of relativity. However, most of the technical advances in array signal processing have occurred in the last 30 years, with the development and proliferation of inexpensive and high-rate analog-to-digital (A/D) converters together with flexible and very powerful digital signal processors (DSPs). These devices have made the chore of collecting data from multiple sensors relatively easy, and helped give birth to the use of sensor arrays in many different areas.

Parallel to the advances in hardware that facilitated the construction of sensor array platforms were breakthroughs in the mathematical tools and models used to exploit sensor array data. Finite impulse response (FIR) filter design methods originally developed for time-domain applications were soon applied to uniform linear arrays in implementing digital beamformers. Powerful data-adaptive beamformers with constrained look directions were conceived and applied with great success in applications where the rejection of strong interference was required. Least-mean square (LMS) and recursive least-squares (RLS) time-adaptive techniques were developed for time-varying scenarios. So-called "blind" adaptive beamforming algorithms were devised that exploited known temporal properties of the desired signal rather than its direction-of-arrival (DOA).

For applications where a sensor array was to be used for locating a signal source, for example finding the source's DOA, one of the key theoretical developments was the parametric vector-space formulation introduced by Schmidt and others in the 1980s. They popularized a vector space signal model with a parameterized array manifold that helped connect problems in array signal processing to advanced estimation theoretic tools such as Maximum Likelihood (ML), Minimum Mean-Square Estimation (MMSE) the Likelihood Ratio Test (LRT) and the Cramér-Rao Bound (CRB). With these

Academic Press Library in Signal Processing. http://dx.doi.org/10.1016/B978-0-12-411597-2.00020-5

tools, one could rigorously define the meaning of the term "optimal" and performance could be compared against theoretical bounds. Trade-offs between computation and performance led to the development of efficient algorithms that exploited certain types of array geometries. Later, concerns about the fidelity of array manifold models motivated researchers to study more robust designs and to focus on models that exploited properties of the received signals themselves.

The driving applications for many of the advances in array signal processing mentioned above have come from military problems involving radar and sonar. For obvious reasons, the military has great interest in the ability of multi-sensor surveillance systems to locate and track multiple "sources of interest" with high resolution. Furthermore, the potential to null co-channel interference through beamforming (or perhaps more precisely, "null-steering") is a critical advantage gained by using multiple antennas for sensing and communication. The interference mitigation capabilities of antenna arrays and information theoretic analyses promising large capacity gains has given rise to a surge of applications for arrays in multi-input, multi-output (MIMO) wireless communications in the last 15 years. Essentially all current and planned cellular networks and wireless standards rely on the use of antenna arrays for extending range, minimizing transmit power, increasing throughput, and reducing interference. From peering to the edge of the universe with arrays of radio telescopes to probing the structure of the brain using electrode arrays for electroencephalography (EEG), many other applications have benefited from advances in array signal processing.

In this chapter, we explore some of the many applications in which array signal processing has proven to be useful. We place emphasis on the word "some" here, since our discussion will not be exhaustive. We will discuss several popular applications across a wide variety of disciplines to indicate the breadth of the field, rather than delve deeply into any one or try to list them all. Our emphasis will be on developing a data model for each application that falls within the common mathematical framework typically assumed in array processing problems. We will spend little time on algorithms, presuming that such material is covered elsewhere in this collection; algorithm issues will only be addressed when the model structure for a given application has unique implications on algorithm choice and implementation. Since radar and wireless communications problems are discussed in extensive detail elsewhere in the book, our discussion of these topics will be relatively brief.

3.20.2 Radar applications

We begin with the application area for which array signal processing has had the most long-lasting impact, dating back to at least World War II. Early radar surveillance systems, and even many still in use today, obtain high angular resolution by employing a radar dish that is mechanically steered in order to scan a region of interest. While such slow scanning speeds are suitable for weather or navigation purposes, they are less tolerable in military applications where split-second decisions must be made regarding targets (e.g., missiles) that may be moving at several thousand miles per hour. The advent of electronically scanned phased arrays addressed this problem, and ushered in the era of modern array signal processing.

Phased arrays are composed of from a few up to several thousand individual antennas laid out in a line, circle, rectangle or even randomly. Directionality is achieved by the process of *beamforming*: multiplying the output of each antenna by a complex weight with a properly designed phase (hence the term "phased" array), and then summing these weighted outputs together. The conventional "delay-and-sum"

FIGURE 20.1

A phased array radar enclosed in the nose of a fighter jet.

beamforming scheme involves choosing the weights to phase delay the individual antenna outputs such that signals from a chosen direction add constructively and those from other directions do not. Since the weights are applied electronically, they can be rapidly changed in order to focus the array in many different directions in a very short period of time. Modern phased arrays can scan an entire hemisphere of directions thousands of times per second. Figures 20.1 and 20.2 show examples of airborne and ground-based phased array radars.

For scanning phased arrays, a fixed set of beamforming weights is repeatedly applied to the antennas over and over again, in order to provide coverage of some area of interest. Techniques borrowed from time-domain filter design such as windowing or frequency sampling can be used to determine the beamformer weights, and the primary trade-off is beamwidth/resolution versus sidelobe levels. Adaptive weight design is required if interference or clutter must be mitigated. In principle, the phased array beamformer can be implemented with either analog or digital hardware, or a combination of both. For arrays with a very large number of antennas (e.g., the Patriot radar has in excess of 5000 elements), analog techniques are often employed due to the hardware and energy expense required in implementing a separate RF receive chain for each antenna. Hybrid implementations are also used in which analog beamforming over subsets of the array is used to create a smaller number of signal streams, which are then processed by a digital beamformer. This is a common approach, for example, in shipborne radar systems, where the targets of interest (e.g., low altitude cruise missiles) are typically located near the horizon. In such systems, analog beamforming with vertically-oriented strips of antennas are used to create a set of narrow azimuthal beams whose outputs can be flexibly combined using digital signal processing.

FIGURE 20.2

The phased array used for targeting the Patriot surface-to-air missile system, composed of over 5000 individual elements.

In this section, we will briefly discuss the two radar array applications that have received the most attention in the signal processing literature: space-time adaptive processing (STAP) and MIMO radar. Since these are discussed in detail elsewhere in the book, our discussion will not be comprehensive. While STAP and MIMO radar applications are typically used in active radar systems, arrays are also useful for passive radars, such as those employed in radio astronomy. We will devote a separate section to array signal processing for radio astronomy and discuss this application in much more detail, since it is not addressed elsewhere in the book.

3.20.2.1 Space-time adaptive processing

In many tactical military applications, airborne surveillance radars are tasked with providing location and tracking information about moving objects both on the ground and in the air. These radars typically

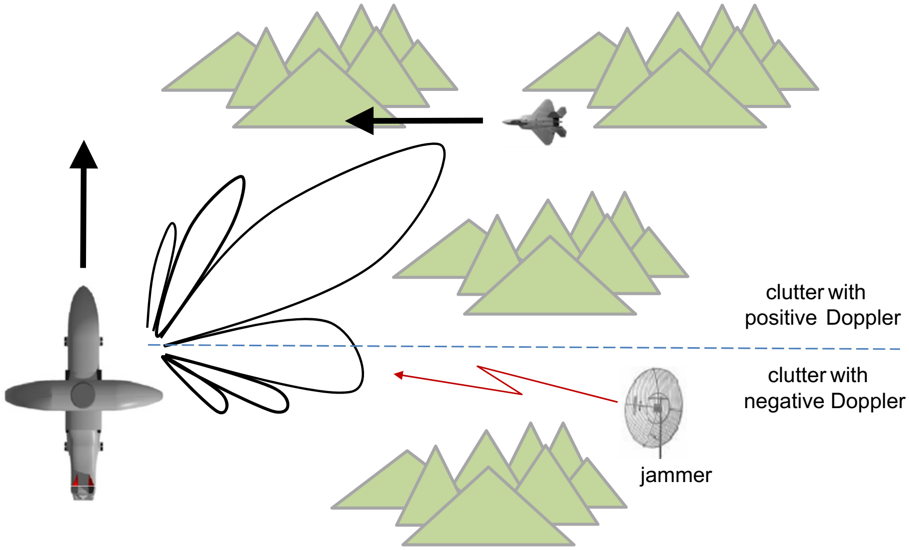


FIGURE 20.3

Airborne STAP scenario with clutter and jamming.

use pulse-Doppler techniques since measuring the velocity of the objects of interest (or "targets") is a key to accurately tracking them. As depicted in Figure 20.3, even when the targets are airborne, the transmit mainbeam and sidelobes will still illuminate the ground, especially when the radar look-direction is at a negative elevation angle (the targets may be below the radar platform). This means that the radar returns will contain significant energy from ground reflections, referred to as *clutter*. In addition, since pulse-Doppler techniques require an active radar, the frequency support of the radar signal is known, and an adversary can employ strong jamming to further mask the target returns. Often, the target signal is many tens of dB (e.g., 50 or more) weaker than the combination of jamming and clutter.

The difficulty of the situation is revealed by Figure 20.4, which shows the angle-Doppler power spectrum of data that contains a target together with clutter and jamming at a particular range. The jamming signal is due to a point source, so it is confined to a single arrival angle, but the jamming signal extends across the entire bandwidth of the data. The clutter energy lies on a ridge that cuts across the angle-Doppler space in a direction that is a function of the heading, altitude and velocity of the radar, and the current range bin of interest. Clutter in front of the radar will have a positive Doppler, and that behind it will be negative (as seen in Figure 20.3). Compared with the clutter and jamming, the target signal is weak and cannot be distinguished from the background due to the limited dynamic range of the receiver. Doppler filtering alone is not sufficient to reveal the target, since the jamming signal cuts across the entire bandwidth of the signal. On the other hand, using spatial filtering (beamforming) to null the jammer will still leave most of the clutter untouched. What is needed is a two-dimensional space-time filter. The process of designing and applying such a filter is referred to as space-time adaptive processing (STAP).

To better place STAP in the context of array signal processing problems, consider Figure 20.5 which depicts how data is organized in an M-antenna pulse-Doppler radar. The radar transmits a series of K

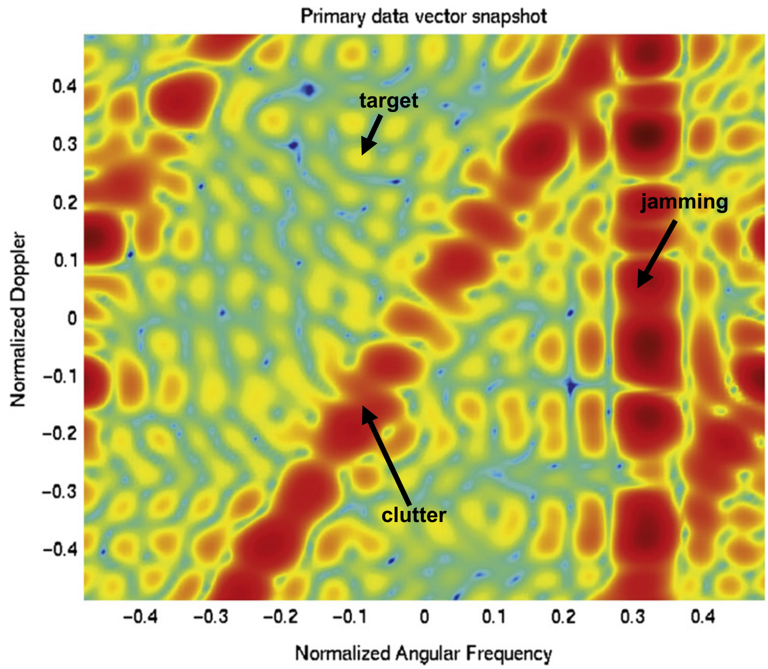


FIGURE 20.4

Angle-Doppler spectrum with weak target in the presence of clutter and jamming.

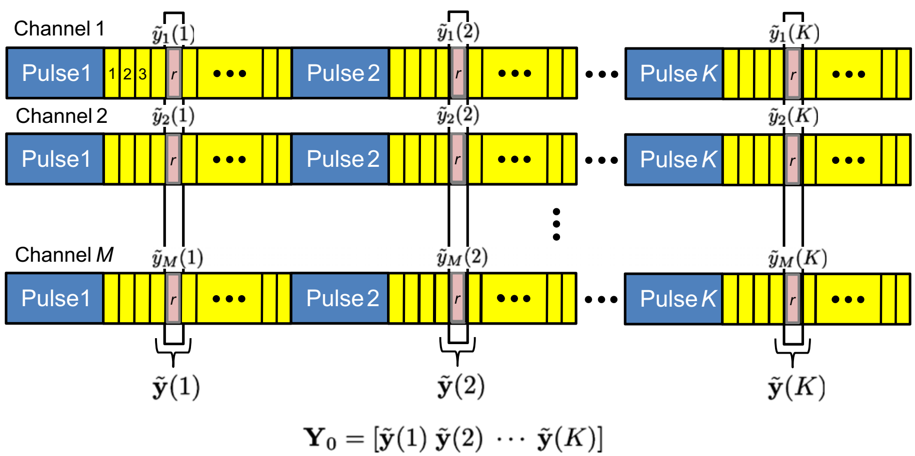


FIGURE 20.5

Organization of data for range bin r in STAP pulsed-Doppler radar.

pulses separated in time by a fixed pulse repetition interval (PRI). In order to focus sufficient energy to obtain a measurable return from a target, the transmitted pulse is typically a very spatially focused signal steered towards a particular azimuth and elevation angle or look direction. However, the mathematical description of the STAP process can be described independently of this assumption. In between the pulses, the radar collects the returns from each of the M antennas, which are sampled after the received data is passed through a pulse-compression matched filter. Each sample corresponds to the aggregate contribution of scatterers (clutter and targets, if such exist) at a particular range together with any noise, jamming or other interference that may be present. The range for a given sample is given by the speed of light multiplied by half the time interval between transmission of the pulse and the sampling instant. Suppose we are interested in a particular range bin r. As shown in the figure, we will let

$$\tilde{\mathbf{y}}(t) = \begin{bmatrix} \tilde{y}_1(t) \\ \vdots \\ \tilde{y}_M(t) \end{bmatrix}, \tag{20.1}$$

$$\mathbf{Y}_0 = \begin{bmatrix} \tilde{\mathbf{y}}(1) & \cdots & \tilde{\mathbf{y}}(K) \end{bmatrix} \tag{20.2}$$

represent the $M \times 1$ vector of returns from the array after pulse t and the $M \times K$ matrix of returns from all K pulses for range bin r, respectively.

Alternatively, as shown in Figure 20.6, the data can be viewed as forming a cube over M antennas, K pulses, and B total range bins. Each range bin corresponds to a different slice of the data cube. Data from adjacent range bins $\mathbf{Y}_k$ will be used to counter the effect of clutter and jamming in the range bin of interest, which we index with $k = 0$. The time required to collect the data cube for a given look direction is referred to as a coherent processing interval (CPI). If the radar employs multiple look directions, a separate CPI is required for each. Assuming the target, clutter and jamming are stationary over different CPIs, data from these CPIs can be combined to perform target detection and localization. However, in our discussion here we will assume that data from only a single CPI is available to determine the presence or absence of a target in range bin r.

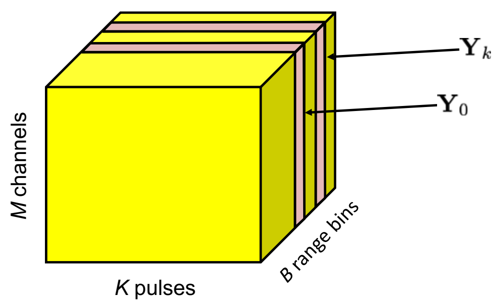


FIGURE 20.6

STAP data cube showing slices for range bin of interest ($\mathbf{Y}_0$) and secondary range bin ($\mathbf{Y}_k$).

If a target is present in the data set $\mathbf{Y}_0$, then the received signal can be modeled as

$$\tilde{\mathbf{y}}(t) = b_0\mathbf{a}(\theta_0, \phi_0)e^{j\omega_0 t} + \underbrace{\sum_{i=1}^{D_c} b_i \mathbf{a}(\theta_i, \phi_i)e^{j\omega_i t} + \sum_{j=1}^{D_j} \mathbf{a}(\theta_j^*, \phi_j^*)x_j(t) + \mathbf{n}(t)}_{\tilde{\mathbf{e}}(t)}, \tag{20.3}$$

where b_i is the amplitude of the return from the ith scatterer ($i = 0$ corresponds to the target), (θ_i, ϕ_i) are the azimuth and elevation angles of the ith scatterer, ω_i is the corresponding Doppler frequency, $\mathbf{a}(\theta, \phi)$ is the response of the M-element receive array to a signal from direction (θ_i, ϕ_i), $x_j(t)$ is the signal transmitted by the jth jammer, (θ_j^*, ϕ_j^*) denote the DOA of the jth jammer signal, D_c represents the number of distinct clutter sources, D_j the number of jammers, and $\mathbf{n}(t)$ corresponds to any remaining background noise and interference. We have also defined $\tilde{\mathbf{e}}(t)$ to contain all received signals except that of the target. Note that the above model assumes the relative velocity of the radar and all scatterers is constant over the CPI, so that the Doppler effect can be described as a complex sinusoid.

Technically, the amplitude and Doppler terms b_i and ω_i will also depend on the azimuth and elevation angles of the ith scatterer since the Doppler frequency is position-dependent and the strength of the return is a function of the transmit beampattern in addition to the intrinsic radar cross section (RCS) of the scatterer. This is clear from Figure 20.7, which shows the geometry of the airborne radar with respect a clutter patch on the ground at some range r. The Doppler frequency for the given clutter patch at azimuth θ and elevation ϕ can be determined from the following equations:

$$\sin\phi = \frac{H}{r} + \frac{r^2 - H^2}{2r(r_e + H)}, \tag{20.4}$$

$$\cos\alpha = \sin\theta\cos\phi, \tag{20.5}$$

$$\omega = \frac{4\pi V}{\lambda}\cos\alpha, \tag{20.6}$$

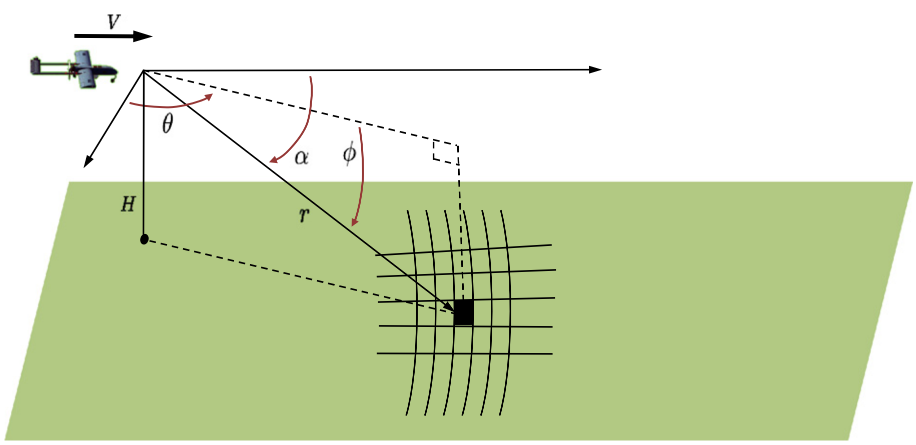


FIGURE 20.7

Geometry for determining the Doppler frequency due to a ground clutter patch at range r.

where r_e denotes the earth's radius, H is the altitude of the radar, and α is the angle between the velocity vector of the radar and the clutter patch. To simplify the notation, we have dropped the explicit dependence of b_i and ω_i on θ_i, ϕ_i. While the highest Doppler frequencies obviously occur for small α (forward- or rear-looking radar), the fact that $\cos\alpha$ changes relatively slowly for small α compared with α near 90° means that the Doppler spread of the clutter for a forward- or rear-looking radar will be smaller than that for the side-looking case.

Rather than working with the data matrix $\mathbf{Y}_0$, for STAP it is convenient to vectorize the data as follows:

$$\mathbf{y}_0 = \text{vec}(\mathbf{Y}_0) = \begin{bmatrix} \tilde{\mathbf{y}}(1) \\ \vdots \\ \tilde{\mathbf{y}}(K) \end{bmatrix} = b_0 \mathbf{s}(\theta_0, \phi_0, \omega_0) + \mathbf{e}_0, \tag{20.7}$$

where $\mathbf{e}_0$ is defined similarly to $\mathbf{y}_0$ for the clutter and jamming, and where

$$\mathbf{s}(\theta_0, \phi_0, \omega_0) = \text{vec}\left(\mathbf{a}(\theta_0, \phi_0)\left[e^{j\omega_0}\ e^{j2\omega_0}\ \cdots\ e^{jK\omega_0}\right]\right) \tag{20.8}$$

$$= \begin{bmatrix} e^{j\omega_0} \\ \vdots \\ e^{jK\omega_0} \end{bmatrix} \otimes \mathbf{a}(\theta_0, \phi_0). \tag{20.9}$$

The $MK \times 1$ vector $\mathbf{y}_0$ is the space-time snapshot associated with the given range bin (r) of interest. To detect whether or not a target signal was present in $\mathbf{y}_0$, one may be tempted to use a minimum-variance distortionless response (MVDR) space-time filter of the form

$$\mathbf{w}(\theta, \phi, \omega) = \frac{\mathbf{R}_{y_0}^{-1}\mathbf{s}(\theta, \phi, \omega)}{\mathbf{s}^H(\theta, \phi, \omega)\mathbf{R}_{y_0}^{-1}\mathbf{s}(\theta, \phi, \omega)}, \tag{20.10}$$

apply it to $\mathbf{y}_0$ for various choices of (θ, ϕ, ω), which then should lead to a peak in the filter output when (θ, ϕ, ω) corresponds to the parameters of the target. The problem with this approach is that we will not have enough data available to estimate the covariance $\mathbf{R}_{y_0}$; if the target signal is only present in this range bin, then with a single CPI we only have a single snapshot that possesses this covariance.

Fortunately, an alternative approach exists, since it can be shown via the matrix inversion lemma (MIL) that the optimal MVDR space-time filter is proportional to another vector that can be more readily estimated:

$$\mathbf{w}(\theta, \phi, \omega) \propto \mathbf{R}_{e_0}^{-1}\mathbf{s}(\theta, \phi, \omega), \tag{20.11}$$

which depends on the covariance $\mathbf{R}_{e_0}$ of the clutter and jamming. In particular, STAP relies on the assumption that the statistics of the clutter and jamming in range bins near the one in question are similar, and can be used to estimate $\mathbf{R}_{e_0}$. For example, let $S_0 = \{k_1, k_2, \ldots, k_{N_s}\}$ represent a set containing the indices of N_s target-free range bins near r (since the target signal may leak into range bins immediately adjacent to bin r, these are typically excluded), then a sample estimate of $\mathbf{R}_{e_0}$ may be formed as

$$\widehat{\mathbf{R}}_{e_0} = \sum_{k \in S_0} \mathbf{y}_k \mathbf{y}_k^H = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^H, \quad \boldsymbol{\Gamma} = \left[\mathbf{y}_{k_1} \ldots \mathbf{y}_{k_{N_s}}\right], \tag{20.12}$$

where $\mathbf{y}_k$ is the space-time snapshot from range bin k. The N_s samples that compose $\boldsymbol{\Gamma}$ are referred to as secondary data vectors.

Implementation of the space-time filter in (20.11) using a covariance estimate such as (20.12) is referred to as the "fully adaptive" STAP algorithm. The number N_s of secondary data vectors chosen to estimate $\mathbf{R}_{e_0}$ is a critical parameter. If it is too small, a poor estimate will be obtained; if it is too large, then the assumption of statistical similarity may be strained. Another critical parameter is the rank of $\mathbf{R}_{e_0}$. While in theory $\mathbf{R}_{e_0}$ may be full rank, in practice its effective rank ρ is typically much smaller than its dimension MK, since the clutter and jamming are usually orders of magnitude stronger than the background noise. According to Brennan's rule [1], the value of ρ for a uniform linear array is $M + (K - 1)\beta$, where β is a factor that depends on the speed of the array platform and the pulse repetition frequency (PRF), and is usually between 0.5 and 1.5. The rank of $\mathbf{R}_{e_0}$ for non-linear array geometries will be greater, although no concise formula exists in the general case. Factors influencing the rank of $\mathbf{R}_{e_0}$ include the beamwidth and sidelobes of the transmit pulse (narrower pulses and lower sidelobes mean smaller ρ), the presence of intrinsic clutter motion (e.g., leaves on trees in a forest) or clutter discretes (strong specular reflectors), and whether the radar is forward- or side-looking (the Doppler spread of the clutter and hence ρ is much smaller in the forward-looking case).

The rank of $\mathbf{R}_{e_0}$ is important in determining the minimum value for N_s required to form a sufficiently accurate sample estimate. A general rule of thumb is that the number of required samples is on the order of 2ρ–5ρ. Even when these many stationary secondary range bins are available, N_s may still be much smaller than MK, and $\widehat{\mathbf{R}}_{e_0}$ will not be invertible. In such situations, a common remedy is to employ a diagonal loading factor δ, and use the MIL to simplify calculation of the inverse:

$$\left(\widehat{\mathbf{R}}_{e_0} + \delta \mathbf{I}\right)^{-1} = \left(\boldsymbol{\Gamma}\boldsymbol{\Gamma}^H + \delta \mathbf{I}\right)^{-1} \tag{20.13}$$

$$= \frac{1}{\delta}\left(\mathbf{I} - \boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^H\boldsymbol{\Gamma} + \delta \mathbf{I}\right)^{-1}\boldsymbol{\Gamma}^H\right). \tag{20.14}$$

Another approach is to use a pseudo-inverse based on principal components.

Still, the computation involved in implementing the fully adaptive STAP algorithm is often prohibitive. The dimension MK of $\mathbf{R}_{e_0}$ is often in the hundreds, and computational costs add up quickly when one realizes the STAP filtering must be performed in multiple range bins for each look direction. Most of the STAP research in recent years has been aimed at reducing the computational load to more reasonable levels. Two main classes of approaches have been proposed: (1) partially adaptive STAP and (2) parametric modeling. In the partially adaptive approach, the dimensions of the space–time data slice are reduced by means of linear transformations in space or time or both:

$$\mathbf{Y}_0 \rightarrow \mathbf{T}_a \mathbf{Y}_0 \mathbf{T}_\omega^H. \tag{20.15}$$

Techniques for choosing the transformation matrices include beamspace methods, Doppler binning, PRI staggering, etc. The classical moving target indicator (MTI) approach can be thought of as falling in this class of algorithms for the special case where $\mathbf{T}_a$ is one-dimensional. The dimension reduction achieved by partially adaptive methods not only reduces the computational load, but it improves the numerical conditioning and decreases the required secondary sample support as well.

The parametric approach is based on the observation that in (20.14), as $\delta \to 0$, we have

$$\lim_{\delta \to 0} \left(\widehat{\mathbf{R}}_{e0} + \delta \mathbf{I}\right)^{-1} \propto \left(\mathbf{I} - \mathbf{\Gamma}\left(\mathbf{\Gamma}^H \mathbf{\Gamma}\right)^{-1} \mathbf{\Gamma}^H\right). \tag{20.16}$$

Thus, the effect of $\mathbf{R}_{e0}^{-1}$ is to approximately project the space-time signal vector onto the space orthogonal to the clutter and jamming. While $\mathbf{\Gamma}$ could be used to define this subspace, a more efficient approach has been proposed based on vector autoregressive (VAR) filtering. To see this, note from (20.3) and (20.7) that the clutter and jamming vector $\mathbf{e}_k$ for range bin k over the full CPI can be partitioned into samples for each individual pulse within the CPI:

$$\mathbf{e}_k = \begin{bmatrix} \mathbf{n}_k(1) \\ \mathbf{n}_k(2) \\ \vdots \\ \mathbf{n}_k(K) \end{bmatrix}. \tag{20.17}$$

The VAR approach assumes that the clutter and jamming obey the following model for each pulse t:

$$\mathbf{H}_0 \mathbf{n}_k(t) + \mathbf{H}_1 \mathbf{n}_k(t-1) + \cdots + \mathbf{H}_L \mathbf{n}_k(t-L+1) = 0, \tag{20.18}$$

where L is typically assumed to be small (e.g., less than 5–7) and each matrix $\mathbf{H}_i$ is $M' \times M$ for some chosen value of M'. The matrix coefficients of the VAR can be estimated for example by solving a standard least-squares problem of the form

$$\min_{\mathcal{H}} \sum_{k=1}^{N_s} \|\mathcal{H}\mathbf{e}_k\|^2 \quad \text{s.t.} \quad \mathcal{H}^H \mathcal{H} = \mathbf{I}, \tag{20.19}$$

where

$$\mathcal{H} = \begin{bmatrix} \mathbf{H}_L & \mathbf{H}_{L-1} & \cdots & \mathbf{H}_0 & & & \\ & \mathbf{H}_L & \mathbf{H}_{L-1} & \cdots & \mathbf{H}_0 & & \\ & & \ddots & & & \ddots & \\ & & & \mathbf{H}_L & \mathbf{H}_{L-1} & \cdots & \mathbf{H}_0 \end{bmatrix} \tag{20.20}$$

and the constraint $\mathcal{H}^H \mathcal{H} = \mathbf{I}$ is used to prevent a trivial solution. The matrix $\mathcal{H}^H$ will approximately span the subspace orthogonal to $\mathbf{\Gamma}$, and based on (20.16) a suitable space-time filter would be given by

$$\mathbf{w} = \mathbf{P}_{\mathcal{H}^H} \mathbf{s}(\theta, \phi, \omega), \tag{20.21}$$

where

$$\mathbf{P}_{\mathcal{H}^H} = \mathcal{H}^H \left(\mathcal{H}\mathcal{H}^H\right)^{-1} \mathcal{H}. \tag{20.22}$$

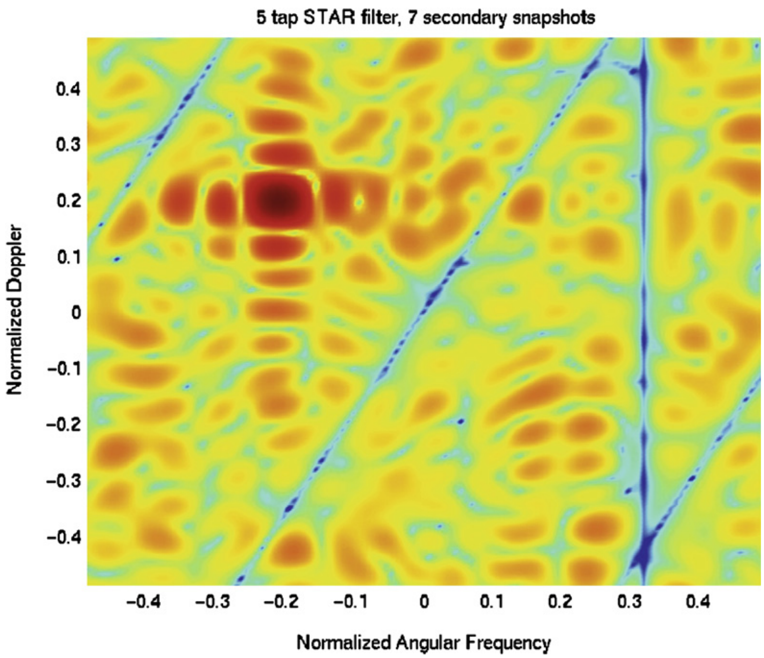


FIGURE 20.8

Angle-Doppler spectra after STAP filtering.

This approach is referred to as the space-time autoregressive (STAR) filter. An example of the performance of the STAR filter is given in Figure 20.8 for a case with $L = 4$ and $N_s = 7$. These results are for the same data set that generated the unfiltered angle-Doppler spectrum in Figure 20.4. Note that the clutter and jamming have been removed, and the target is plainly visible. Similar results were obtained in this case with the fully adaptive STAP method with diagonal loading, but required a value of N_s near 60.

3.20.2.2 MIMO radar

Multi-input multi-output (MIMO) radar is beginning to attract a significant amount of attention from researchers and practitioners alike due to its potential of advancing the state-of-the-art of modern radar. Unlike a standard phased-array radar, which transmits scaled versions of a single waveform, a MIMO radar system can transmit via its antennas multiple probing signals that may be chosen quite freely (see Figure 20.9). This waveform diversity enables superior capabilities compared with a standard phased-array radar. For example, the angular diversity offered by widely separated transmit/receive antenna elements can be exploited for enhanced target detection performance. For collocated transmit and receive antennas, the MIMO radar paradigm has been shown to offer many advantages including long virtual array aperture sizes and the ability to untangle multiple paths. Array signal processing plays critical roles in reaping the benefits afforded by the MIMO radar systems. In our discussion here, we focus on array signal processing for MIMO radar with collocated transmit and receive antennas.

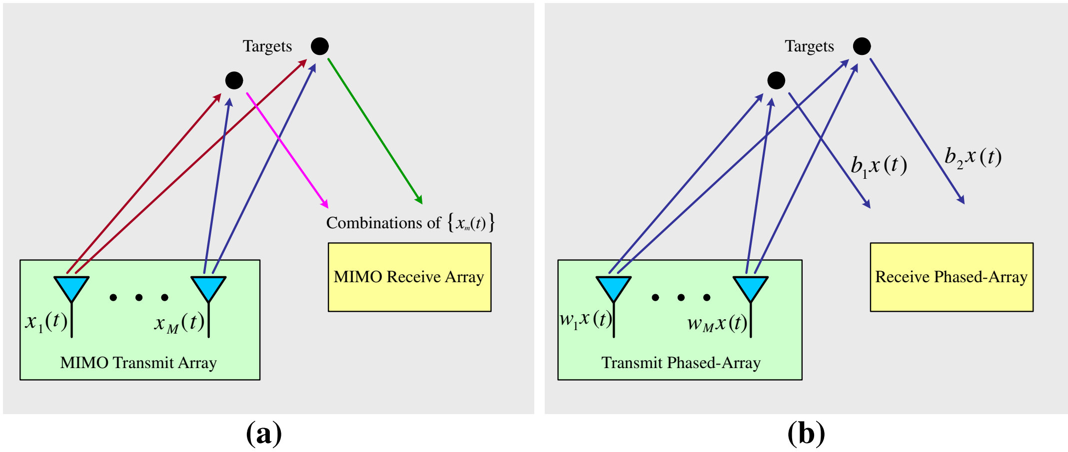


FIGURE 20.9

(a) MIMO radar and (b) phased-array radar.

An example of a UAV equipped with a MIMO radar system is shown in Figure 20.10, where the transmit array is sparse and the receive array is a filled (half-wavelength inter-element spacing) uniform linear array. When the transmit antennas transmit orthogonal waveforms, *the virtual array of the radar system is a filled array with an aperture up to M times that of the receive array, where M is the number of transmit antennas*. Many advantages of MIMO radar with collocated antennas result directly from this significantly increased virtual aperture size. For example, for small aerial vehicles (with medium or short range applications), a conventional phased-array system could be problematic since it usually weighs too much, consumes too much power, takes up too much space, and is too expensive. In contrast, MIMO radar offers the advantages of reduced complexity, power consumption, weight and cost by obviating phase shifts and affording significantly increased virtual aperture size.

Some typical examples of array processing in MIMO radar include transmit beampattern synthesis, transmit and receive array design, and adaptive array processing for diverse MIMO radar applications. We briefly describe these array processing examples in MIMO radar.

3.20.2.2.1 *Flexible transmit beampattern synthesis*

The probing waveforms transmitted by a MIMO radar system can be designed to approximate a desired transmit beampattern and also to minimize the cross-correlation of the signals reflected from various targets of interest—an operation that would hardly be possible for a phased-array radar.

The recently proposed techniques for transmit narrowband beampattern design have focused on the optimization of the covariance matrix $\mathbf{R}$ of the waveforms. Instead of designing $\mathbf{R}$, we might think of directly designing the probing signals by optimizing a given performance measure with respect to the matrix $\mathbf{X}$ of the signal waveforms. However, compared with optimizing the same performance measure with respect to the covariance matrix $\mathbf{R}$ of the transmitted waveforms, optimizing directly with respect to

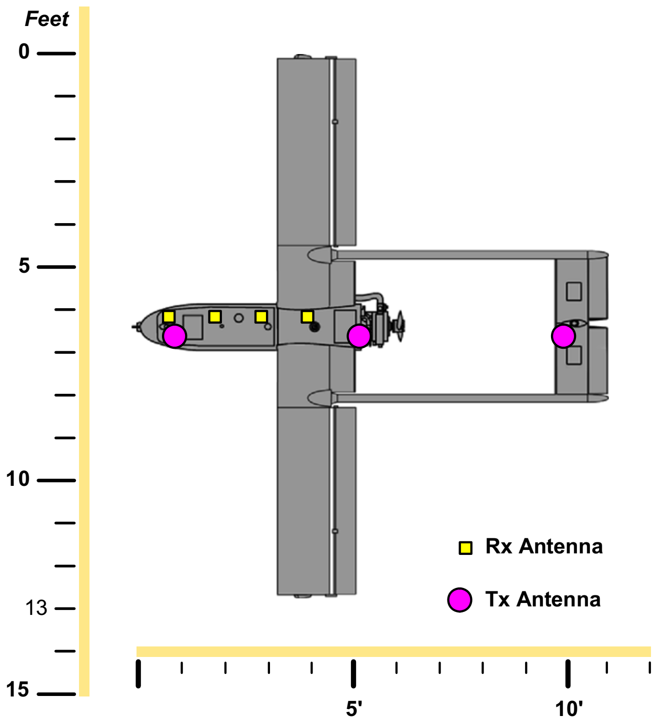


FIGURE 20.10

A UAV equipped with a MIMO radar.

X is a more complicated problem. This is so because **X** has more unknowns than **R** and the dependence of various performance measures on **X** is more intricate than the dependence on **R**.

There are several recent methods that can be used to efficiently compute an optimal covariance matrix **R**, with respect to several performance metrics. One of the metrics consists of choosing **R**, under a uniform elemental power constraint (i.e., under the constraint that the diagonal elements of **R** are equal), to achieve the following goals:

a. Maximize the total spatial power at a number of given target locations, or more generally, match a desired transmit beampattern.
b. Minimize the cross-correlation between the probing signals at a number of given target locations.

Another beampattern design problem is to choose **R**, under the uniform elemental power constraint, to achieve the following goals:

a. Minimize the sidelobe level in a prescribed region.
b. Achieve a predetermined 3 dB main-beam width.

It can be shown that both design problems can be efficiently solved in polynomial time as a semi-definite quadratic program (SQP).

We comment in passing on the conventional phased-array beampattern design problem in which only the array weight vector can be adjusted and therefore all antennas transmit the same differently-scaled waveform. We can readily modify the MIMO beampattern designs for the case of phased-arrays by adding the constraint that the rank of $\mathbf{R}$ is one. However, due to the rank-one constraint, both of these originally convex optimization problems become non-convex. The lack of convexity makes the rank-one constrained problems much harder to solve than the original convex optimization problems. Semi-definite relaxation (SDR) is often used to obtain approximate solutions to such rank-constrained optimization problems. The SDR is obtained by omitting the rank constraint. Hence, interestingly, *the MIMO beampattern design problems are the SDRs of the corresponding phased-array beampattern design problems*.

We now provide a numerical example below, where we have used a Newton-like algorithm to solve the rank-one constrained design problems for phased-arrays. This algorithm uses SDR to obtain an initial solution, which is the exact solution to the corresponding MIMO beampattern design problem. Although the convergence of the said Newton-like algorithm is not guaranteed, we did not encounter any apparent problem in our numerical simulations.

Consider the beampattern design problem with $M = 10$ transmit antennas. The main-beam is centered at $\theta_0 = 0°$, with a 3 dB width equal to 20° ($\theta_1 = -10°, \theta_2 = 10°$). The sidelobe region is $\Omega = [-90°, -20°] \cup [20°, 90°]$. The minimum-sidelobe beampattern design is shown in Figure 20.11a. Note that the peak sidelobe level achieved by the MIMO design is approximately 18 dB below the mainlobe peak level. Figure 20.11b shows the corresponding phased-array beampattern obtained by using the additional constraint $\text{rank}(\mathbf{R}) = 1$. The phased-array design fails to provide a proper mainlobe (it suffers from peak splitting) and its peak sidelobe level is much higher than that of its

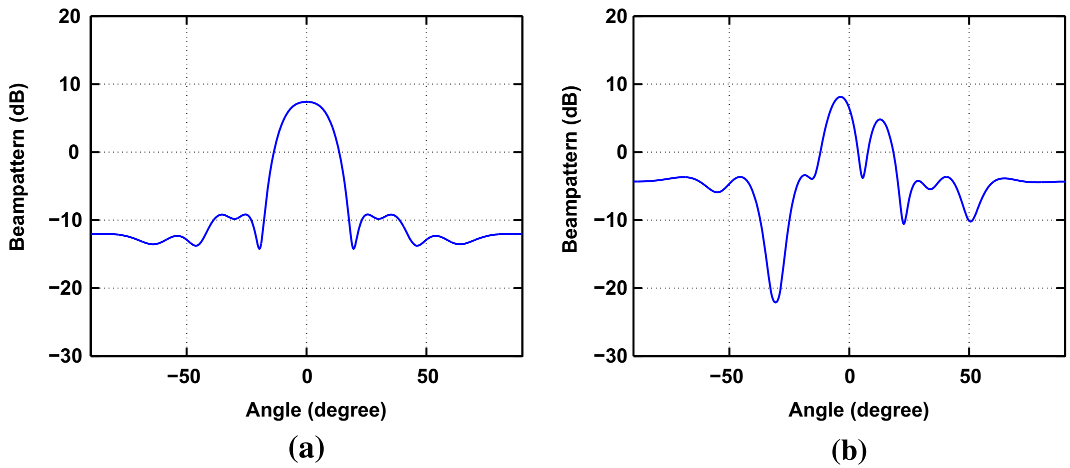


FIGURE 20.11

Minimum sidelobe beampattern designs, under the uniform elemental power constraint, when the 3 dB main-beam width is 20°. (a) MIMO and (b) phased-array.

MIMO counterpart. We note that, under the uniform elemental power constraint, the number of degrees of freedom (DOF) of the phased-array that can be used for beampattern design is equal to only $M - 1$; consequently, it is difficult for the phased-array to synthesize a proper beampattern. The MIMO design, on the other hand, can be used to achieve a much better beampattern due to its much larger number of DOF, viz. $M^2 - M$.

The radar waveforms are generally desired to possess constant modulus and excellent auto- and cross-correlation properties. Consequently, the probing waveforms can be synthesized in two stages: at the first stage, the covariance matrix $\mathbf{R}$ of the transmitted waveforms is optimized, and at the second stage, a signal waveform matrix $\mathbf{X}$ is determined whose covariance matrix is equal or close to the optimal $\mathbf{R}$, and which also satisfies some practically motivated constraints (such as constant modulus or low peak-to-average-power ratio (PAR) constraints). A cyclic algorithm for example, can be used for the synthesis of such an $\mathbf{X}$, where the synthesized waveforms are required to have good auto- and cross-correlation properties in time.

3.20.2.2.2 Array design

For a phased-array radar system, the transmission of coherent waveforms allows for a narrow mainbeam and, thus, a high signal-to-noise ratio (SNR) upon reception. When the locations of targets in a scene are unknown, phase shifts can be applied to the transmitting antennas to steer the focal beam across an angular region of interest. In contrast, MIMO radar systems, by transmitting different, possibly orthogonal waveforms, can be used to illuminate an extended angular region over a single processing interval, as we have demonstrated above.

Waveform diversity permits higher degrees of freedom, which enables the MIMO radar system to achieve increased flexibility for transmit beampattern design. The assumptions used in the discussions above are that the positions of the transmitting antennas, which also affect the shape of the beampattern, are fixed prior to the construction of $\mathbf{R}$ followed by the synthesis of $\mathbf{X}$. At the receiver, sparse, or thinned, array design has been the subject of an abundance of literature during the last 50 years. The purpose of sparse array design has been to reduce the number of antennas (and thus reduce the cost) needed to produce desirable spatial receiving beampatterns. The ideas behind sparse receive array methodologies can be extended to that of sparse, MIMO array design. For example, cyclic algorithms can be used to approximate desired transmit and receive beampatterns via the design of sparse antenna arrays. These algorithms can be seen as extensions to iterative receive beampattern designs.

3.20.2.2.3 Adaptive array processing at radar receivers

Adaptive array processing plays a vital role at radar receivers, including those of MIMO radar. Conventional data-independent algorithms, such as the delay-and-sum approach for array processing, suffer from poor resolution and high sidelobe level problems. Data-adaptive algorithms, such as MVDR (Capon) receivers, have been widely used in radar receivers. These adaptive signal processing algorithms offer much higher resolution and lower sidelobe levels than the data-independent approaches. However, these algorithms can be sensitive to steering vector errors and also require a substantial number of snapshots to determine the second-order statistics (covariance matrices). To mitigate these problems, diagonal loading has been used extensively in practical applications to make adaptive algorithms feasible. However, too much diagonal loading makes the adaptive algorithm degenerate into data-independent methods, and the diagonal loading level may be hard to determine in practice. Parametric methods tend to be sensitive to data model errors and are not as widely used as the aforementioned data-adaptive algorithms.

In MIMO radar, adaptive array processing is essential, especially because many of the simple tricks used to achieve the longer virtual arrays, such as randomized antenna switching (also called randomized time-division multiple access (R-TDMA)) and slow-time code-division multiple access (ST-CDMA), provide sparse random sampling. Because of such sampling, the high sidelobe level problem suffered by data-independent approaches are exacerbated. Moreover, most of the radar signal processing problems encountered in practice do not have multiple snapshots. In fact, in most practical applications, only a single data measurement snapshot is available for adaptive signal processing. For example, in synthetic aperture radar (SAR) imaging, just a single phase history matrix is available for SAR image formation. Moreover the phase history matrix may not be uniformly sampled. In MIMO radar applications, including MIMO-radar-based space-time adaptive processing (STAP), synergistic MIMO SAR imaging and ground moving target indication (GMTI), and untangling multiple paths for diverse radar operations such as those encountered by MIMO over-the-horizon radar (OTHR), we essentially have just a single snapshot available at the radar receiver, especially in a heterogeneous clutter environment.

Fortunately, the recent advent of iterative adaptive algorithms, such as the iterative adaptive approach (IAA) and sparse learning via iterative minimization (SLIM), obviate the need of multiple snapshots and the uniform sampling requirements but retain desirable properties, including high resolution, low sidelobe level, and robustness against data model errors, of the conventional adaptive array processing methods. Moreover, for uniformly sampled data, various fast implementation strategies of these algorithms have been devised to exploit the Toeplitz matrix structures. These iterative adaptive algorithms are particularly suited for signal processing at radar receivers. They can also be used in diverse other applications, such as in sonar, radio astronomy, and channel estimation for underwater acoustic communications.

3.20.3 Radio astronomy

Radio astronomy is the study of our universe by passive observation of extra-terrestrial radio frequency emissions. Sources of interest for astronomers include (among others) radio galaxies, pulsars, supernova remnants, synchrotron radiation from excited material in a star's magnetic field, ejection jets from black holes, narrowband emission and absorption lines from diffuse elemental or chemical compound matter that can be assayed by their characteristic spectral structure, and continuum thermal black body radiation emitted by objects ranging from stars to interstellar dust and gasses. The radio universe provides quite a different and complementary view to that which is visible to more familiar optical telescopes. Radio astronomy has enabled a much fuller understanding of the structure of our universe than would have been possible with visible light alone. With Doppler red shifting, the spectrum of interest ranges from as low as the shortwave regime near 10 MHz, to well over 100 GHz in the millimeter and submillimeter bands, and there are radio telescopes either in use or under development to cover much of this spectrum.

From the earliest days of radio astronomy, detecting faint deep space sources has pushed available technology to extreme performance limits. Early progress was driven by improvements in hardware with relatively straightforward signal processing and detection techniques. With the advent of large synthesis arrays, signal processing algorithms increased in sophistication. More recently, interest in phased array feeds (PAFs) has opened a new frontier for array signal processing algorithm development for radio astronomical observations.

Radio astronomy presents unique challenges as compared to typical applications in communications, radar, sonar, or remote sensing:

- *Low SNR:* Deep space signals are extremely faint. SNRs of −30 to −50 dB are routine.
- *Radiometric detection:* A basic observational mode in radio astronomy is "on-source minus off-source" radiometric detection where the source level is well below the noise floor and can only be seen by differencing with a noise only estimate. This requires stable power estimates of (i) system noise plus weak signal of interest (SOI) and (ii) noise power alone with the sensor steered off the SOI. The standard deviation of the noise power estimate determines the minimum detectable signal level, so that long integration times (minutes to hours) are required.
- *Low system temperatures:* With cryogenically cooled first stage low noise amplifiers, system noise temperatures can be as low as 15 K at L-band, including LNA noise, waveguide ohmic losses, downstream receiver noise, and spillover noise from warm ground observed beyond the rim of a dish reflector.
- *Stability:* System gain fluctuations increase the receiver output variance and place a limit on achievable sensitivity that cannot be overcome with increased integration time. High stability in gain, phase, noise, and beamshape response over hours is required to enable long term integrations to tease out detection of the weakest sources.
- *Bandwidth:* Some scientific observations require broad bandwidths of an octave or more. Digital processing over such large bandwidths poses serious computational burdens.
- *Radio frequency interference (RFI):* Observations in RFI environments outside protected frequency bands are common. Interference levels below the noise floor may be as problematic as strong interferers, since they are hard to identify and attenuate. Cancelation approaches also cause pattern rumble which limits sensitivity.

3.20.3.1 Synthesis imaging

Radio astronomical synthesis imaging uses interferometric techniques and some of the world's largest sensor arrays to form high resolution images of the distribution of radio sources in deep space. Figure 20.12 presents two examples of the beautiful high resolution detail revealed by synthesis imaging from the Very Large Array (VLA) in New Mexico, and Figure 20.13 shows the VLA with its antennas configured in a compact central core configuration. The key to this technology is coherent cross-correlation processing (i.e., interferometry) of RF signals seen by pairs of widely separated antennas (up to 10s of kilometers and more). Each such antenna typically consists of a high gain dish reflector of 12–45 m diameter which serves as a single element in the larger array. At lower frequencies, in order to avoid difficulties of physically steering the large aperture needed for high gain, array elements may themselves be built up as electronically steered beamforming aperture array "stations" using clusters of fixed bare antennas without a reflector (for example, the LOFAR array). Whether implemented with a collection of large dish telescopes, or with a beamforming array, these elements of the full imaging array provide a sparse spatial sampling of the wavefront that would have been observed by a much larger, imaginary "synthetic" encompassing dish. Though the array cannot match the collecting areas of the

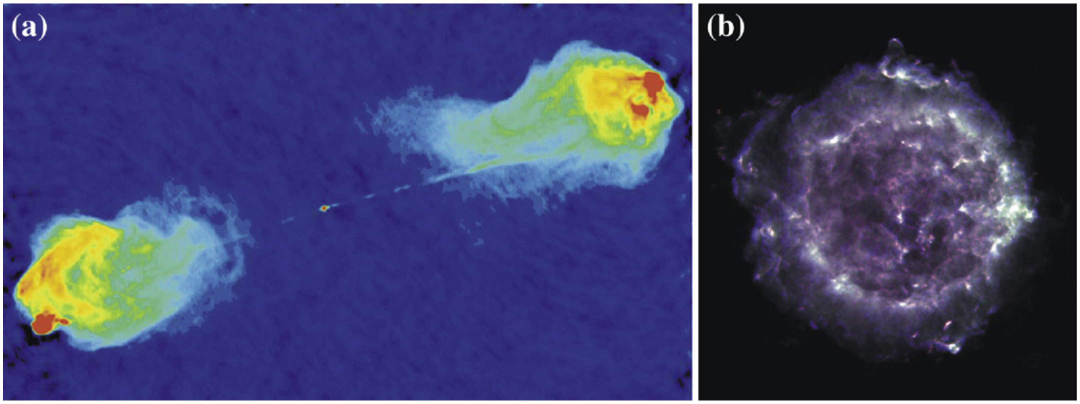


FIGURE 20.12

VLA images of radio sources not visible to optical astronomy. (a) An early image of the gas jet structures in Cygnus A (ejected from the spinning core of the radio galaxy in the constellation Cygnus) seen at 5.0 GHz 1983 by Perley, Carilli, and Dreher. (b) Supernova remnant Cassiopeia A, 1994 composite of 1.4, 4.0, and 8.4 GHz images, by Rudnick, Delaney, Keohane, Koralesky, and Rector.

Credits: National Radio Astronomy Observatory/Associated Universities, Inc./National Science Foundation.

FIGURE 20.13

The central core of the Very Large Array (VLA) in compact configuration.

Credit: Dave Finely, National Radio Astronomy Observatory/Associated Universities, Inc./National Science Foundation.

synthesized aperture, the long "baseline" distances between antennas yield spatial imaging resolution comparable to that of the encompassing dish aperture which inscribes the baseline vectors. Exploiting the earth's rotation over time relative to the distant celestial sky patch being observed fills in sampling gaps between sparse array elements.

There are a number of aspects of synthesis imaging arrays that are distinct from many other array signal processing applications. Due to wide separation there is no mutual coupling and noise is truly independent across the array. The large scale, long baselines, and critical dependence on phase relationships require very long coherent signal transport or precision time stamping of data sets using atomic clock references. Each array element is itself a high gain, highly directive antenna with a sizable aperture. Precision array calibration is required, but due to large scale hardware this cannot be done in a laboratory or on an antenna range. Self calibration methods are employed that use known point-source deep space objects in the field of view to properly phase the array. Array geometry is sparse with either random-like or log-scaled spacing. Extreme stability is required due to the need for coherent integration over hours, and bandwidths of interest can cover and octave or more.

3.20.3.1.1 *The imaging equation*

While the signals of interest are broadband, processing typically takes place in frequency subchannels so that narrowband models can typically be used. Further, since deep space sources are typically seen through line-of-sight propagation, multipath scattering is limited and occurs only locally as reflections off antenna support structures. Thus the propagation channel can be considered to be memoryless (zero delay spread). The synthesis imaging equations relate the observed cross correlation between pairs of array elements to the expected electromagnetic source intensity spatial distribution over a patch of the

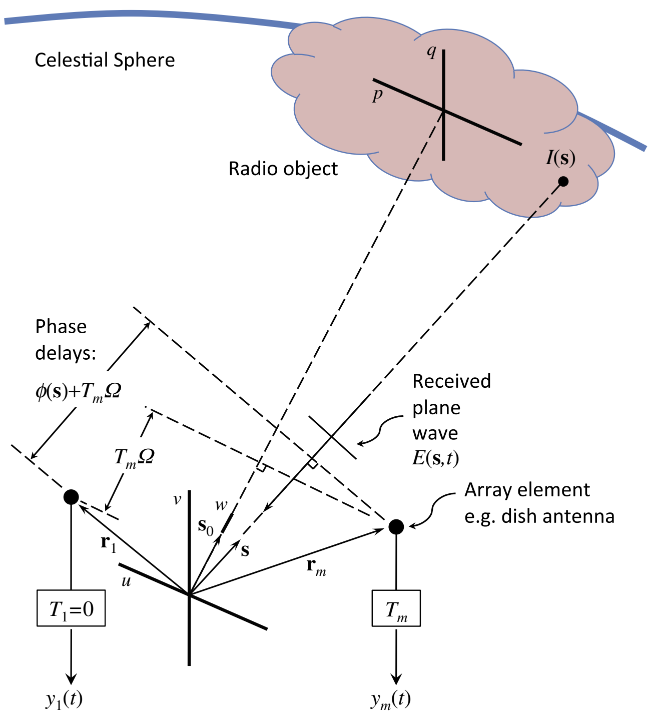


FIGURE 20.14

Geometry and signal definitions for the synthesis imaging equations.

celestial sphere. Figure 20.14 illustrates the geometry, signal definitions, and coordinate systems for one of the baseline pairs of antennas used to develop the imaging equations.

Consider the electric field $E(\mathbf{s}, t) = E(\mathbf{s})e^{j\Omega t}$ observed by the array at frequency Ω due to a narrowband plane wave signal arriving from the direction pointed to by the unit length 3-space vector $\mathbf{s}$. We consider only the quasi monochromatic case where a single radiation frequency Ω is observed by subband processing. To simplify discussion, polarization effects are not considered so $E(\mathbf{s})$ is treated as a scalar rather than vector quantity, though working synthesis arrays typically have dual polarized antennas and receiver systems to permit studying source polarization. Since distance is indeterminate to the array, in our model the observed $E(\mathbf{s})$ and its corresponding intensity distribution $I(\mathbf{s}) = E\left[|E(\mathbf{s})|^2\right]$ are projected without time or phase shifting onto a hypothetical far-field celestial sphere that is interior to the nearest observed object. The goal of synthesis imaging is to estimate $I(\mathbf{s})$ from observations of sensor array $\mathbf{y}(t)$.

Define the image coordinate axes (p, q) to be fixed on the celestial sphere and centered in the imaging field of view patch. Since $\mathbf{s}$ is unit length, we may use these coordinates to express it as $\mathbf{s} = \left(p, q, \sqrt{1 - p^2 - q^2}\right)$. Let $\mathbf{s}_0$ point to the $(p = 0, q = 0)$ origin, thus $\mathbf{s}_0 = (0, 0, 1)$. For small values of p and q, such as being contained within a field-of-view limited by the narrow beamwidth of array antennas, $\mathbf{s} \approx (p, q, 1)$. Time delays T_m are inserted in the signal paths for receiver outputs $y_m(t)$ to compensate for the differential propagation times of a plane wave originating from the (p, q) origin. The most distant antenna is arbitrarily designated as the $m = 1$st element, and $T_1 = 0$. Thus the array is co-phased for a signal propagating along $\mathbf{s}_0$.

Receiver output voltage signal $y_m(t)$, $1 \leq m \leq M$, is given by the superposition of scaled electric field contributions from across the full celestial sphere surface S, plus local sensor noise:

$$y_m(t) = \int_S A(\mathbf{s}) E(\mathbf{s}) e^{j(\Omega t + \phi_m(\mathbf{s}))} d\mathbf{s} + n_m(t), \tag{20.23}$$

where $A(\mathbf{s})$ represents the known antenna element directivity pattern and downstream receiver gain terms, $\phi_m(\mathbf{s})$ is the phase shift due to differential geometric propagation distances for a source from $\mathbf{s}$ relative to a co-phased source from $\mathbf{s}_0$ as shown in Figure 20.14, and $\mathbf{n}_m(t)$ is the noise seen in the mth array element. For simple imaging algorithms, it is assumed that all elements (e.g., dish antennas) have identical spatial response patterns and that each is steered mechanically or electronically to align its beam mainlobe with $\mathbf{s}_0$, so $A(\mathbf{s})$ does not depend on m and sources outside the elemental beams are strongly attenuated. The beamwidth defined by $A(\mathbf{s})$ determines the maximum imaging field of view, or patch size. Considering the full array, (20.23) can be expressed in vector form as:

$$\mathbf{y}(t) = \int_S A(\mathbf{s}) E(\mathbf{s}) e^{j(\Omega t + \boldsymbol{\phi}(\mathbf{s}))} d\mathbf{s} + \mathbf{n}(t) \tag{20.24}$$

where $\boldsymbol{\phi}(\mathbf{s}) = [\boldsymbol{\phi}_1(\mathbf{s}) \ \cdots \ \phi_M(\mathbf{s})]^T$.

Consider the vector distance between two array elements, $(\mathbf{r}_l - \mathbf{r}_m)$, $l \neq m$, where $\mathbf{r}_m$ is the location of the mth antenna. This is known as an interferometric "baseline," and it plays a critical role in synthesis imaging. Longer baselines yield higher resolution images by increasing the synthetic array aperture diameter, and using more antennas provides more distinct baseline vectors which will be shown to more fully sample the image in the angular spectrum domain. In the following all functions of element position depend only on such vector differences, so it is convenient to define a relative coordinate system (u, v, w) in the vicinity of the array to express the difference as $(\mathbf{r}_l - \mathbf{r}_m) = (u, v, w)$. Align (u, v) with (p, q), and w with $\mathbf{s}_0$. Scale these axes so distance is measured in wavelengths, i.e., so that a unit distance corresponds to one wavelength $\lambda = \frac{2\pi c}{\Omega}$, where c is the speed of light. In this coordinate system we have by simple geometry

$$\phi_m(\mathbf{s}) + T_m\Omega = -2\pi\mathbf{s}(\mathbf{r}_m - \mathbf{r}_1), \text{ and } T_m\Omega = -2\pi\,\mathbf{s}_0(\mathbf{r}_m - \mathbf{r}_1). \tag{20.25}$$

At array outputs $y_m(t)$, after the inserted delays T_m, the effective phase difference between two array elements is then

$$\phi_l(\mathbf{s}) - \phi_m(\mathbf{s}) = -2\pi(\mathbf{s} - \mathbf{s}_0)^T(\mathbf{r}_l - \mathbf{r}_m). \tag{20.26}$$

Using the signal models of (20.23) and (20.26), the cross correlation of two antenna signals as a function of their positions is given by:

$$R(\mathbf{r}_l, \mathbf{r}_m) = \mathrm{E}\left[y_l(t)y_m^*(t)\right] \quad \text{for } l \neq m \tag{20.27}$$

$$= \mathrm{E}\left[\left(\int_S A(\mathbf{s})E(\mathbf{s})e^{j(\Omega t + \phi_l(\mathbf{s}))}d\mathbf{s} + n_l(t)\right)\right.$$
$$\left.\times\left(\int_S A(\mathbf{s}')E(\mathbf{s}')e^{j(\Omega t + \phi_m(\mathbf{s}'))}d\mathbf{s}' + n_m(t)\right)^*\right] \tag{20.28}$$

$$= \int_S |A(\mathbf{s})|^2 I(\mathbf{s})e^{-j2\pi(\mathbf{s}-\mathbf{s}_0)^T(\mathbf{r}_l - \mathbf{r}_m)}d\mathbf{s}$$

$$= \int_S |A(\mathbf{s})|^2 I(\mathbf{s})e^{-j2\pi(p,q,a-1)^T(u,v,w)}d\mathbf{s} \tag{20.29}$$

$$= \iint_{-\infty}^{\infty} |A(p,q)|^2 \frac{1}{a} I(p,q)e^{-j2\pi(up+vq+w(a-1))}dp\,dq \tag{20.30}$$

$$\approx \iint_{-\infty}^{\infty} |A(p,q)|^2 I(p,q)e^{-j2\pi(up+vq)}dp\,dq = R(u,v) \quad \text{for } u, v \neq 0, \tag{20.31}$$

where $a = \sqrt{1 - p^2 - q^2}$. We have assumed zero mean spatially independent radiators for $E(\mathbf{s})$ and $n_m(t)$, a narrow field of view so $a \approx 1$, and that $\mathbf{s}_0 = (0, 0, 1)$. The quantity $R(u, v)$ is known by radio astronomers as a "visibility function" where arguments $\mathbf{r}_l$ and $\mathbf{r}_m$ are replaced by u and v since the final expression depends only on these terms. A cursory inspection of (20.31) reveals that it is precisely a 2-D Fourier transform relationship, so the inversion method to obtain $I(\mathbf{s})$ from visibilities $R(u, v)$ suggests itself:

$$I(p,q) = \frac{1}{|A(p,q)|^2}\iint_{-\infty}^{\infty} R(u,v)e^{j2\pi(up+vq)}du\,dv, \quad \forall\{(p,q)|A(p,q) \gg 0\} \tag{20.32}$$

$$= \frac{1}{|A(p,q)|^2} F^{-1}(R(u,v)), \tag{20.33}$$

where $F^{-1}(\cdot)$ is the inverse 2-D Fourier transform. This is the well known synthesis imaging equation. Since only cross correlations between distinct antennas are measured by this imaging interferometer, the self power terms $R(\mathbf{r}_l, \mathbf{r}_m)|_{\mathbf{r}_l=\mathbf{r}_m} = R(0,0)$ are not computed or used in the Fourier inverse. The d.c. level in the image which normally depends on these terms must rather be adjusted to provide a black, zero valued background.

3.20.3.1.2 Algorithms for solving the imaging equation

The geometry of the imaging problem described in (20.32) and illustrated in Figure 20.14 is continually changing due to Earth rotation. The fixed ground antenna positions $\mathbf{r}_m$ rotate relative to the (u, v) axis, which remains aligned to the (p, q) axis fixed on the celestial sphere. On one hand, this is a negative effect because it limits the integration time that can be used to estimate $R(u, v)$ under a stationarity assumption. On the other hand, rotation produces new baseline vectors $(\mathbf{r}_l - \mathbf{r}_m)$ with distinct orientations, filling in the Fourier space coverage for $R(u, v)$ and improving image quality. To exploit rotation, imaging observations are made over long time periods, up to 12 h, to form a single image.

Receiver outputs are sampled as $\mathbf{y}(i) \equiv \mathbf{y}(iT_s)$ at frequency $f_s = 1/T_s$, and sample covariance estimates of the visibility function (assuming zero mean signals) are obtained as

$$\widehat{\mathbf{R}}_k = \frac{1}{N} \sum_{i=kN}^{(k+1)N-1} \mathbf{y}(i)\mathbf{y}^H(i), \tag{20.34}$$

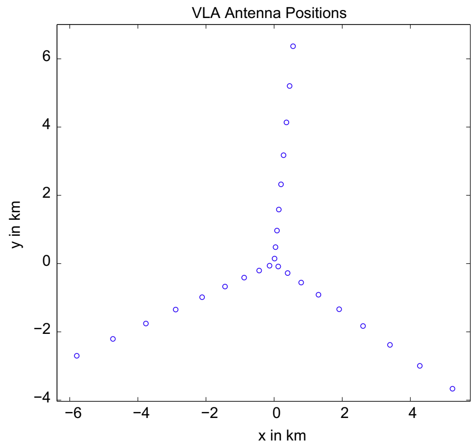
where N is the number of samples in the long term integration (LTI) window over which the imaging geometry and thus cross correlations may be assumed to be approximately stationary, and k is the LTI index. (We will later introduce a short term integration window length N_{sti} over which moving interference sources appear statistically stationary.)

Since covariance estimates are only available at discrete time intervals (one per LTI index k), and the antennas have fixed Earth positions, only samples of $R(u, v)$ are available with irregular spacing in the (u, v) plane, so (20.32) must be solved with discreet approximations. However, noting that due to Earth rotation, the corresponding antenna position vector orientations $\mathbf{r}_m$ depend on time through k, a new set of (u, v) samples with different locations is available at each LTI. Index k is thus added to the notation to distinguish distinct baseline vectors $(\mathbf{r}_{k,l} - \mathbf{r}_{k,m})$ for the same antenna pairs during different LTIs. So the (l, m)th element of $\widehat{\mathbf{R}}_k$ relates to the sampled visibility function as

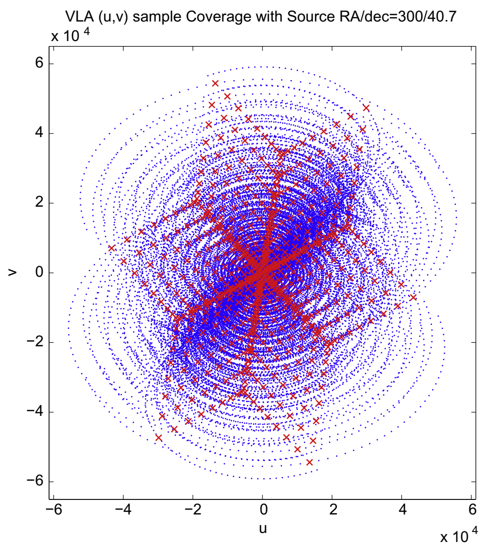
$$\{\widehat{\mathbf{R}}_k\}_{lm} = \widehat{R}_{k,lm} \approx R(u_{k,lm}, v_{k,lm}), \tag{20.35}$$

where $(u_{k,lm}, v_{k,lm}, w_{k,lm}) = (\mathbf{r}_{k,l} - \mathbf{r}_{k,m})$ and where as in (20.26) and (20.31), due to inserted time delays T_m we may take $w_{k,lm}$ to be zero. For simplicity we will use a single index κ to represent unique LTI-antenna index triples $\{k, lm\}$ to specify vector samples in the (u, v) plane, so $(u_{k,lm}, v_{k,lm}) = (u_\kappa, v_\kappa)$ and $\widehat{R}_{k,lm} = \widehat{R}_\kappa$. Thus elements of the sequence of matrices $\widehat{\mathbf{R}}_k$ provide a non-uniformly sampled representation of the visibility function, or frequency domain image. Consistent with the treatment of $R(0, 0)$ in (20.32), diagonal elements in $\widehat{\mathbf{R}}_\kappa$ are set to zero.

Figure 20.15a presents an example of a certain VLA geometry, and Figure 20.15b shows where the (u_κ, v_κ) samples would lie, with each point representing a unique sample κ. This plot includes 61 LTIs (i.e., $0 \le k \le 60$) over a 12 h VLA observation for the Cygnus A radio galaxy of Figure 20.12a. This sample pattern would change for sources with different positions on the celestial sphere (expressed by astronomers in right ascension and declination).



(a)



(b)

FIGURE 20.15

(a) An example VLA antenna element geometry with the repositionable 25 m dishes in a compact log spacing along the arms. Axis units are in kilometers. (b) Corresponding (u, v) sample grid for a 12 h observation of Cygnus A. Each point represents a (u_κ, v_κ) sample corresponding to a unique baseline vector where a visibility estimate $\widehat{R}_\kappa$ is available. Red crosses denote baselines from a single LTI midway through the observation, and blue points are additional samples available using Earth rotation, with a new $\widehat{\mathbf{R}}_k$ computed every 12 min. Observation is at 1.61 GHz and axis units are in wavelengths. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this book.)

With this frequency domain sampling and including noise effects (20.32) becomes

$$\widehat{I}_D(p,q) = \frac{1}{|A(p,q)|^2} \iint_{-\infty}^{\infty} \Psi(u,v) \left(R(u,v) + \widetilde{R}(u,v) \right) e^{j2\pi(up+vq)} du\, dv \qquad (20.36)$$

$$= \frac{1}{|A(p,q)|^2} \sum_{\kappa} \widehat{R}_\kappa \, e^{j2\pi(u_\kappa p + v_\kappa q)}, \qquad (20.37)$$

where $\widehat{I}_D(p,q)$ is known as the "dirty image," the sampling function $\Psi(u,v) = \sum_\kappa \delta(u - u_\kappa, v - v_\kappa)$, and $\widetilde{R}(u,v)$ represents sample estimation error in the covariance/visibility. Since the (u,v) plane is sparsely sampled, $\Psi(u,v)$ introduces a bias in the inverse which must be removed by deconvolution as described below. This also means that (20.37) is not a true inverse Fourier transform due to the limited set of basis functions used. It is referred to as the "direct Fourier inverse" solution.

There are two common approaches to solving (20.36) or (20.37) for $\widehat{I}_D(p,q)$ given a set of LTI covariances $\widehat{R}_\kappa$. The most straightforward though computationally intensive method is a brute force evaluation of (20.37) given knowledge of the (u_κ, v_κ) sample locations (e.g., as in Figure 20.15). Alternately, the efficiencies of a 2-D inverse FFT can be exploited if these samples and corresponding visibilities $\widehat{R}_k$ are re-sampled on a uniform rectilinear grid in the (u,v) plane. "Cell averaging" assigns the average of visibility samples contained in a local cell region to the new rectilinear grid point in the middle of the cell. Other re-gridding methods based on higher order 2-D interpolation have also been used successfully. When large fields of view are required, or array elements are not coplanar, then any of these approaches based on (20.31) will not work and a solution to the more complete expression of (20.30) must be found. Cornwell has developed the W-Projection method to address these conditions [27].

An alternate "parametric matrix" representation of (20.31) and (20.37) has been developed. This is particularly convenient because it models the imaging system in a familiar array signal processing form that lends itself readily to analysis, adaptive array processing and interference canceling, and opens up additional options for solving the synthesis imaging and image restoration problems. Returning to the indexing notation of (20.34), note that since $(\mathbf{r}_l - \mathbf{r}_m) = (\mathbf{r}_l - \mathbf{r}_1) - (\mathbf{r}_m - \mathbf{r}_1)$ one may express $(u_{k,lm}, v_{k,lm})$ as $(u_{k,l1} - u_{k,m1}, v_{k,l1} - v_{k,m1})$. Let $J(p,q) = |A(p,q)|^2 I(p,q)$ be the desired image as scaled (i.e., vignetted) by the antenna beam pattern, and sample it on a regular 2-D grid of pixels (p_d, q_d), $1 \le d \le D$. The conventional visibility Eq. (20.31) then becomes

$$R_{k,lm} = \sum_{d=1}^{D} J(p_d, q_d) e^{-j2\pi(u_{k,lm} p_d + v_{k,lm} q_d)} + \sigma_n^2 \delta(l-m) \qquad (20.38)$$

$$= \sum_{d=1}^{D} e^{-j2\pi(u_{k,l1} p_d + v_{k,l1} q_d)} J(p_d, q_d) e^{j2\pi(u_{k,m1} p_d + v_{k,m1} q_d)} + \sigma_n^2 \delta(l-m), \qquad (20.39)$$

which in matrix form is

$$\mathbf{R}_k = \mathbf{A}_k \mathbf{J} \mathbf{A}_k^H + \sigma_n^2 \mathbf{I}, \text{ where} \qquad (20.40)$$

$$\mathbf{A}_k = [\mathbf{a}_{k,1}, \ldots, \mathbf{a}_{k,D}], \tag{20.41}$$

$$\mathbf{a}_{k,d} = \left[e^{-j2\pi(u_{k,11}p_d + v_{k,11}q_d)}, \ldots, e^{-j2\pi(u_{k,M1}p_d + v_{k,M1}q_d)}\right]^T, \tag{20.42}$$

and where $\mathbf{J} = \mathrm{Diag}([J(p_1, q_D), \ldots, J(p_D, q_D)])$ is the diagonal image matrix representation of sampled $J(p,q)$, M is the total number of array elements, and though noise is independent across antennas, the self noise terms have been included to allow for the $l = m$ case that contributes to the diagonal of full matrix $\mathbf{R}_k$. The matrix discrete "direct Fourier inverse" relationship corresponding to (20.37) is

$$\widehat{\mathbf{J}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{A}_k^H \mathbf{R}_k \mathbf{A}_k, \tag{20.43}$$

where K is the number of available LTIs. Equations (20.40) and (20.43) are well suited to address synthesis imaging as an estimation problem, facilitating use of Maximum Likelihood, maximum a posteriori, constrained minimum variance, or robust beamforming techniques. Note that (20.43) is not a complete discrete inverse Fourier transform, indeed, often $D > MK$ so a one-to-one inverse relationship between $\mathbf{R}_k$ and $\mathbf{J}$ does not exist and $\widehat{\mathbf{J}}$ is significantly blurred.

By the Fourier convolution theorem, the effect of frequency sampling by $\Psi(u,v)$ in (20.36) is to convolve the desired image $I(p,q)$ with the "dirty beam response" $\psi_D(p,q) = F^{-1}(\Psi(u,v))$. Neglecting the effect of individual antenna directivity pattern $A(p,q)$, $\psi_D(p,q)$ can be interpreted as the point spread function, or synthetic beam pattern of the imaging array for the given observation scenario. Significant reduction of this blurring effect can be achieved by an image restoration/deconvolution step. The dirty image of (20.36) may be expressed as

$$\widehat{I}_D(p,q) = \psi_D(p,q) * (I(p,q) + \widetilde{I}(p,q)), \tag{20.44}$$

where $\widetilde{I}(p,q) = F^{-1}(\widetilde{R}(u,v))$ is due to sample estimation error in the visibilities. Since antenna locations in the rotating (u,v) plane are known precisely over the full observation, $\psi_D(p,q)$ is known to high accuracy, and with well calibrated dish antennas so is $A(p,q)$. Thus (20.44) may be solved as

$$\widehat{I}(p,q) = \widehat{I}_D(p,q) *^{-1} \psi_D(p,q), \tag{20.45}$$

where "$*^{-1}$" denotes deconvolution with respect to the right argument. Due to the spatial lowpass nature of dirty beam $\psi_D(p,q)$ this problem is ill conditioned and must be regularized by imposing some assumptions about the image. The most popular reconstruction methods impose a sparse source distribution model and use an iterative source subtraction approach related to the original CLEAN algorithm [32]. The sparse model is justifiable for point-source images of star fields, and works well even with more complex distributions of gas and nebular structures given that much of the field of view is expected to be dark. Several variants and extensions to CLEAN have been proposed, some applying source subtraction in the spatial (p,q) domain, and some in the frequency (u,v) domain. Typically these have performance tuning parameters which astronomers adjust for most pleasing results. Thus the effective regularization term or mathematical optimization expression is often not known precisely and the process is a bit ad hoc, but solutions with higher contrast and resolution, and with reduced noise and reconstruction artifacts are preferred. Maximum entropy reconstruction has also been used effectively.

3.20.3.2 Astronomical phased array feeds

A new application for array signal processing in radio astronomy is phased array feeds (PAFs) where the traditional single large horn antenna feed at the focus of large telescope dish is replaced with a closely spaced (order of 1/2 wavelength) 2-D planar array of small antennas located at the dish focal plane. The primary motivation for such a system, as shown in Figure 20.16 is to form multiple simultaneous beams steered to cover a grid pattern in a field of view that is many times larger than the single pixel horn fed dish. PAFs are ideal for wide-field and survey instruments where it is desired to cover large regions of the sky in the shortest possible time. They provide the ability to capture a small image over the field of view, with one pixel per simultaneously formed beam, using a single snapshot pointing of the dish. Such systems have been referred to as "radio cameras." Additional advantages of PAFs include sensitivity optimization with respect to the noise environment, and spatial interference cancelation capabilities (see Figure 20.16 and Section 3.20.3.3) albeit at the expense of increased hardware and processing complexity.

In some ways PAF processing is simply conventional beamforming for an array of microwave receiving antennas, but there are several unique aspects of the application that provide some challenges.

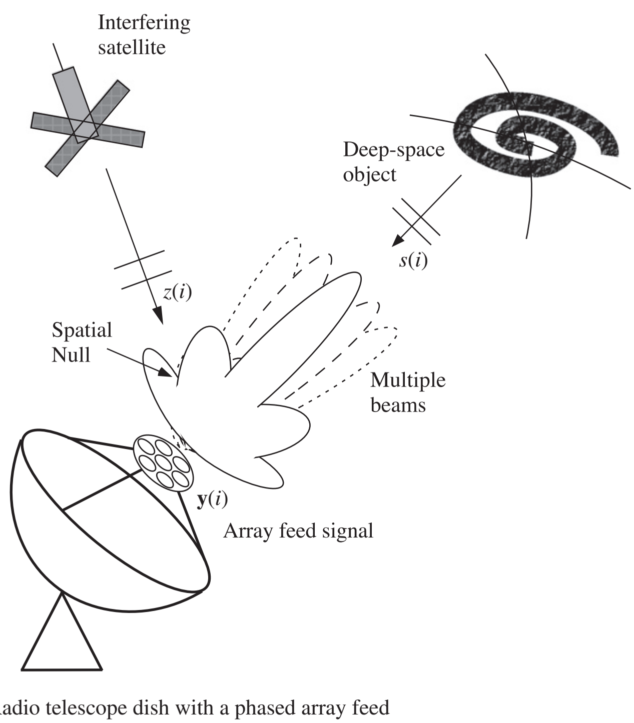


FIGURE 20.16

The primary advantage of FPA telescopes is increased field of view provided by multiple, simultaneously formed beams. Spatial cancelation of interfering signals is also possible, but very deep nulls are required.

The following technical hurdles are why PAFs have not been previously adopted in radio astronomy, but these issues have largely been resolved and working platforms have now been demonstrated.

First, the PAF is not a bare aperture array but operates in conjunction with a very large reflector which for an on-axis far field point source focusses a tight Airy pattern spot of energy at the array that spans little more than a single array element. For off-axis sources the spot moves across the array and undergoes coma shaped pattern distortion. So, though noise and interference are seen on all elements, only a few antennas see much of the SOI. The combined dish and PAF can be viewed as a dense array of small but high gain, highly directive elements, but not all of these have equal SNR. Elements outside the focal spot must however be used in beamforming to control the illumination pattern on the dish and thus reduce spillover noise from observing warm ground beyond the edge of the dish. The focal properties of the dish also limit the achievable field of view, even with electronic steering, since deviation from the boresight axis beyond a few beamwidths leads to defocusing and loss of gain, no matter how large the PAF is.

Second, array calibration is critical to achieve maximum sensitivity (gain over noise power) and due to the huge sizes of these instruments, must be performed in situ using known deep space objects as calibration sources of opportunity. Calibrations must be performed periodically (order of weeks) to account for electronic and structural drift, and must estimate array response vectors in every direction that a beam is to be steered or a response constraint is to be placed.

Third, beamformer weight calculation is non-trivial. Astronomers want maximum sensitivity and stable beampatterns on the sky, but these competing requirements are challenging. The variable correlated noise field environment of a radio telescope calls for an adaptive approach, but it is difficult to obtain low error array calibrations at enough points to control beam sidelobe structure. Also, due to complexity of the antenna structures, it is impossible to design usable beamformer weights from even a very detailed electromagnetic system simulation.

Fourth, as discussed in Section 3.20.3.3, many of the conventional adaptive canceling beamforming methods are not very effective for astronomical PAFs. This is because observations are frequently done when both the SOI and interference power levels are well below the noise floor. New approaches are required to form deeper spatial nulls in scenarios where it is difficult to estimate interference parameters.

Fifth, replacing a single horn feed channel with 38, or 200 array elements, as have been proposed for PAFs, has major implications on the back end processing. Processed bandwidths of 300 MHz or more per antenna are needed, so a real-time DSP processor with capacity to serve as digital receiver, multiple beamformer, and array correlator for calibrationm constitutes a major infrastructure investment.

And finally, in a field where cryogenically cooled antennas and LNAs are the norm to reduce receiver noise, cooling a large array is daunting. Most current development projects have opted for room temperature arrays and trade off the then necessary longer integration times with faster survey speeds possible with multiple beams.

3.20.3.2.1 Signal model

After analog frequency down conversion, sampling, and complex baseband bandshifting, the array signal time sample vector of Figure 20.16 is modeled as

$$\mathbf{y}(i) = \mathbf{a}s(i) + \sum_{d=1}^{D} \mathbf{v}_d(i)\, z_d(i) + \mathbf{n}(i), \tag{20.46}$$

where **a** is the array response vector for signal of interest (SOI) $s(i)$, $\mathbf{v}_d(i)$ is the time varying array response for the dth independent interfering source $z_d(i)$, and $\mathbf{n}(i)$ is the noise vector. Source response **a** is assumed to be constant, even for observation times on the order of an hour because the dish mechanically tracks a point in the sky. Even fixed ground interference sources must be modeled as moving (thus $\mathbf{v}_d(i)$ depends on i) due to this tracking motion of the dish. Approaches to address man–made interference are discussed in Section 3.20.3.3. This model for $z_d(i)$ can also include natural deep space sources which are bright enough to overwhelm the SOI even when seen in the beam sidelobe pattern. Their apparent rotational motion about the SOI is due to Earth rotation. When the corresponding $\mathbf{v}_d(i)$ is known accurately, these can be removed through a successive subtraction algorithm known as peeling. As with synthesis imaging, broadband processing for PAFs is accomplished by FFT based subband decomposition, often with thousands of frequency bins. So in the following we consider only a single frequency channel and adopt the standard narrowband array processing model.

Any array signal processing, including beamforming, must take into account the fact that, unlike synthesis imaging, the PAF noise vector $\mathbf{n}(i)$ is correlated across the array. Even with cryogenic cooling, first stage amplifier LNA noise is correlated due to electromagnetic mutual coupling at the elements. Another major component, spillover noise from warm ground black body radiation as seen by the feed array, is spatially correlated because it is not isotropic since it stops above the horizon and is blocked over a large solid angle by the dish.

In a practical PAF scenario the beams are steered in a rectangular or hexagonal grid pattern with crossover points at the −1 to −3 dB levels. The total number of beams, J, is limited by the maximum steering angle which is determined by the diameter of the array feed and the focal properties of the dish, by the acceptable limit for beamshape distortion, and by the available processing capacity for real-time simultaneous computation of multiple beams. As illustrated in Figure 20.17, the time series output for

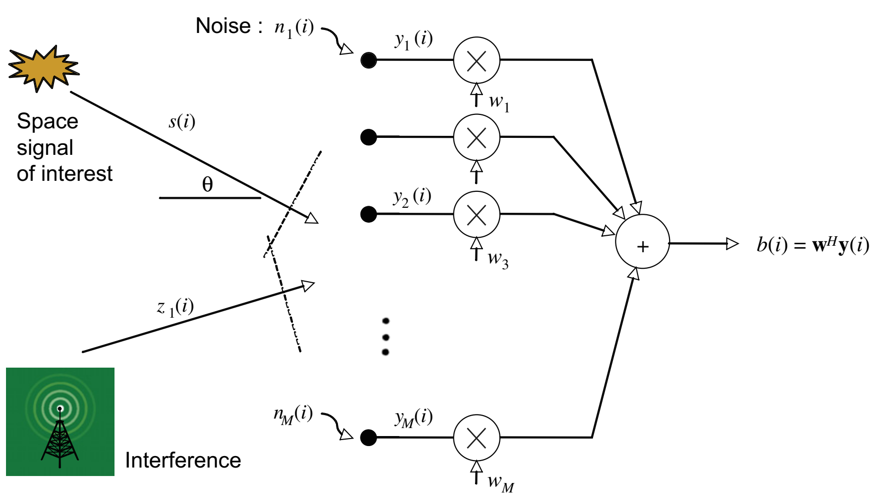


FIGURE 20.17

Beamformer architecture. Narrowband operation is assumed and for PAF beamforming, interaction with the large reflector dish is not shown.

a beam steered in the jth direction is given by

$$b_j(i) = \mathbf{w}_j^H \mathbf{y}(i), \tag{20.47}$$

where $\mathbf{w}_j$ is the vector of complex weights for beamformer j, $1 \leq j \leq J$. Weights are designed based on array calibration data and the desired response pattern constraints and optimization as described in the following two sections. Separate beamformers with their own sets of J distinct weight vectors are computed for each frequency channel, though we consider only a single channel in this discussion.

3.20.3.2.2 Calibration

Since multiple simultaneous beams are formed with a PAF as shown in Figure 20.16, a calibration for the signal array response vector $\mathbf{a}_j$ must be performed for each direction, $\mathbf{s}_j$, corresponding to each formed beam's boresight direction, and any additional directions where point constraints in the beam pattern response will be placed. Periodic re-calibration is necessary due to strict beam pattern stability requirements, to correct for differential electronic phase and gain drift, and to characterize changes in receiver noise temperatures. Calibration is based on sample array covariance estimates $\widehat{\mathbf{R}}$ as described in (20.34) while observing a dominant bright calibration point source in the sky. For example, in the northern hemisphere, Cassiopeia A and Cygnus A shown in Figure 20.12 are the brightest continuum (broadband) sources, and with a typical single dish telescope aperture they are unresolved and appear as point sources. Both have been used as calibrators.

3.20.3.2.3 Beamformer calculation

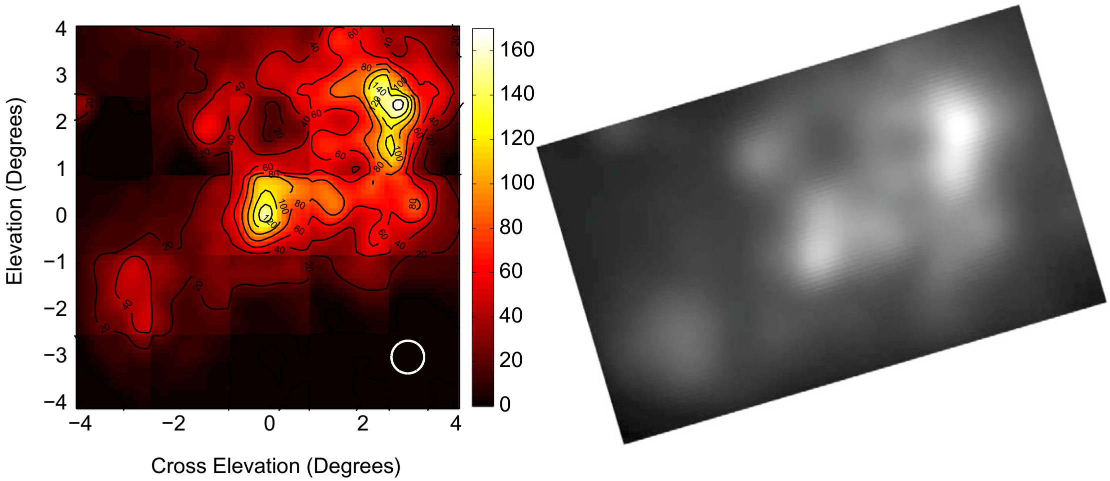
Since discovery of the weakest, most distant sources is a primary aim of radio astronomers, it is paramount to design a dish and feed combination to achieve high sensitivity, which has been derived for a phased array feed to be

$$\frac{A_\mathrm{e}}{T_\mathrm{sys}} = \frac{k_b B}{F_\mathrm{s}} \frac{\mathbf{w}^H \mathbf{R}_\mathrm{s} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_\mathbf{n} \mathbf{w}} \quad (\mathrm{m}^2/\mathrm{K}), \tag{20.48}$$

where A_e (in m^2) represents directivity in terms of the effective antenna aperture collecting area, T_sys is system noise power at the beamformer output expressed as a black body temperature, k_b is Boltzmann's constant, B is system bandwidth, F_s, (in Watts/m^2) is the signal flux density in one polarization, and $\mathbf{R}_\mathrm{s}$ and $\mathbf{R}_\mathbf{n}$ are the signal and noise components of $\mathbf{R}$ respectively. Here we have assumed $D = 0$, i.e., that there are no interferers. For a reflector antenna with a traditional horn feed, maximizing sensitivity involves a hardware-only tradeoff between aperture efficiency, which determines the received signal power, and spillover efficiency, which determines the spillover noise contribution. With a PAF, sensitivity is determined by the beamforming weights as well as the array and receivers. Adjusting $\mathbf{w}$ controls both the PAF illumination pattern on the dish which affects A_e, and the response to the noise field, which affects T_sys. Noting that all other right hand side terms in (20.48) are constant, sensitivity can be maximized with the well known maximum signal to noise ratio (SNR) beamformer

$$\mathbf{w}_\mathrm{snr} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^H \mathbf{R}_\mathrm{s} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_\mathrm{n} \mathbf{w}}. \tag{20.49}$$

To date all hardware demonstrated PAF telescopes have used this maximum sensitivity beamformer. However, a hybrid beamformer design method for PAFs that parametrically trades off sensitivity maximization with constraining mainlobe shape and sidelobe levels has been proposed.



(a) **(b)**

FIGURE 20.18

(a) Cygnus X region at 1600 MHz. 5×5 mosaic of images using the 19-element prototype PAF on the Green Bank 20-Meter Telescope. The circle indicates the half-power beamwidth. (b) Canadian Galactic Plane Survey image [38] convolved to the 20-m effective beamwidth. The center of the map is approximately $20^h 44^m$, $+42°$ (J2000) with north to the upper left.

Credit: Karl Warnick in [33].

*3.20.3.2.4 **Radio camera results***

In 2008, ASTRON and BYU/NRAO independently demonstrated the first radio camera images with a PAF fed dish. Figure 20.18 presents an example of the BYU work as a mosaic image of a complex source distribution in the Cygnus X region. As a comparison, the right image is from the Canadian Galactic Plane Survey image, but blurred by convolution with the equivalent beam pattern of the 20-Meter Telescope to match resolution scales. We expect that the image artifacts caused by discontinuities at mosaic tile boundaries could be eliminated with more sophisticated processing. The Cygnus X radio camera image contains approximately 3000 pixels. A more practical coarse grid spacing of about half the HPBW would require about 600 pixels. A single horn feed would require 600 pointings (one for each pixel) to form such an image, compared to 25 (one for each mosaic tile) for the radio camera. Thus for equal integration times per pixel, this radio camera provides an imaging speed up of 24 times.

3.20.3.3 **Interference mitigation for radio astronomy**

From a regulatory and spectrum management point of view, radio astronomy is a passive wireless service which must co-exist with many other licensed communications activities. Though international treaties have long been established to protect a few important frequency bands for astronomical use only (e.g., around 1420 MHz for emission lines of abundant deep space neutral Hydrogen) these precautions have become wholly inadequate. Astronomers' current scientific goals require observing emissions across the radio spectrum from molecules of more exotic gas compounds, from broad spectrum sources such as pulsars, and from highly red shifted objects nearing the edge of the observable universe where Doppler effects dramatically reduce the frequencies. Thus there is virtually no frequency band devoid of interesting sources to study. Astronomers cannot rely solely on protected bands and must develop methods to mitigate ubiquitous man-made radio transmission interference.

The problem is further exacerbated because one of the fundamental aims of radio astronomy is to discover the weakest of sources which are often at signal levels many tens of decibels below the noise floor. Successful detection usually requires long integration times on the order of hours to average out noise induced sample estimation error variance, combined with on-source minus off-source subtraction to find subtile differences in power levels between a noise-only background and noise plus SOI. Thus even very weak interference levels that would hardly hinder wireless communications can completely obscure an astronomical source of interest.

There is a long laundry list of troublesome RFI sources for radio astronomy. Examples of man-made signals encountered at radio observatories for which mitigation strategies have been demonstrated include: satellite downlink transmissions, radar systems, air navigation aids, wireless communications, and digital television broadcasts. Even locating instruments in undeveloped areas with regulatory protection for radio quiet zones does not avoid many man-made sources such as satellite downlinks. Low frequency synthesis arrays such as LOFAR, PAPER, LWA, and the Murchison Widefield Array operate in the heavily used VHF bands (30–300 MHz) to detect highly redshifted emissions, and as such must contend with very powerful commercial TV and FM radio broadcasts, as well as two-way mobile communications services.

There are a variety of RFI mitigation methods used in radio astronomy. The major approaches include *avoidance* (simply wait until the interference stops or observe in a different frequency band), *temporal excision* (blank out only the small percentage of data samples corrupted by impulsive interference), waveform subtraction (estimate parameters for known structured interference and subtract a synthetic copy of this signal from the data), *anti-coincidence* (remove local interference by retaining only signals common to two distant observing stations), and *spatial filtering* (adaptive array processing to place spatial nulls on interference). Since this present article emphasizes array signal processing, we will address spatial filtering in the following discussion.

Figures 20.16 and 20.19 illustrate interference scenarios for a phased array feed and synthesis imaging array respectively. For PAFs the closely packed antennas in the feed enable for the first time adaptive spatial filtering on single dish telescopes. This would also be possible with PAFs on the multiple dishes of a large imaging array, but even with just typical single horn feeds (as in Figure 20.19) the covariance matrix used to compute imaging visibilities as in (20.34) and (20.40) can also be used for interference canceling. Some proposed algorithms use only the main imaging array antennas, while others achieve improved performance with additional smaller auxiliary antennas trained on the interferers as shown in the figure. The various algorithm approaches will be discussed below. Most spatial filtering work to-date has been at frequencies in L-band (1–2 GHz) and below because this includes important astronomical sources and because of the abundance of man-made interference in these bands.

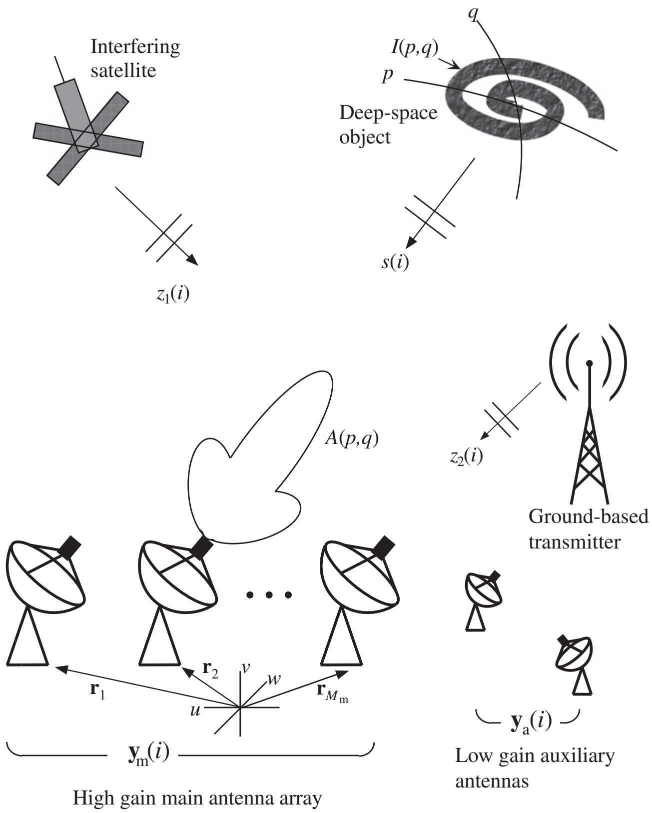


FIGURE 20.19

An RFI scenario at a synthesis imaging array. Two independent interference sources are illustrated: a satellite downlink and a ground-based broadcast transmitter. The main imaging array consists of typical single feed dishes (i.e., PAF feeds are not used here). In addition to the main array, a subarray of smaller auxiliary antennas is shown which can be used with some algorithms discussed below to improve cancelation performance. If tracking information is available, these auxiliaries are steered to the offending sources to provide a high INR copy of the interference.

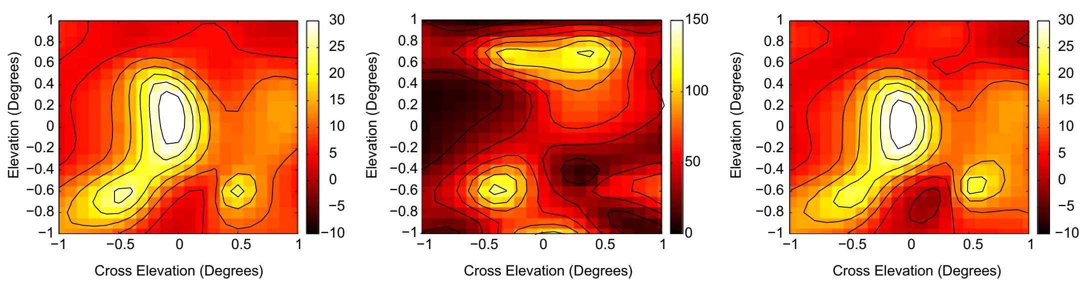
3.20.3.3.1 Challenges and solutions to radio astronomical spatial filtering

Many of the well-known adaptive beamforming algorithms appear at first glance to be promising candidates for interference mitigation in astronomical array processing, including maximum SNR, minimum variance distortionless response (MVDR or Capon), linearly constrained minimum variance (LCMV), generalized sidelobe canceler (GSC), Wiener filtering, and other algorithms. Robust canceling beamformers which are less sensitive to calibration error have also been considered for aperture arrays used as stations in large low frequency imaging arrays like LOFAR. However, due to several challenging characteristics of the radio astronomical RFI problem, most of these approaches are less successful here than they would be in typical radar, sonar, wireless communications, or signal intercept applications. These problems have made many astronomers reluctant to adopt the use of adaptive array processing methods for regular scientific observations. We note though that the intrinsic motivations to observe in RFI corrupted bands are becoming strong enough that rapid progress toward adoption is necessary and is anticipated by most practitioners. New algorithm adaptations are being introduced which are better suited for radio astronomical spatial filtering. We consider below some of the significant aspects of radio astronomy that complicate spatial filtering.

The typical astronomical SOI power level is 30 dB or more below the system noise over comparable bandwidth, even when cryogenically cooled LNAs are used with instruments located in radio quiet zones. Canceling nulls must therefore be deep enough to drive interference below the SOI level, i.e., below the on-source minus off-source detection limit, not just down to the system noise level. Most algorithms require a dominant interferer to form deep nulls because minimum variance methods (MVDR, LCMV, max SNR, Wiener Filtering, etc.) which balance noise variance with residual interference power cannot drive a weaker interferer far below the noise floor. The residual will remain well above the SOI level.

Another promising solution to limited null depth is a zero forcing beamformer like subspace projection (SP) where the null in the estimated vector subspace for interference is theoretically infinitely deep. A number of proposed radio astronomical RFI cancelers have adopted the SP approach and some experimental demonstration results have appeared. Figure 20.20 illustrates the first use of subspace projection RFI mitigation with a PAF as reported in [53]. Data were collected from a 19 element PAF mounted on the 20-Meter Telescope at the NRAO Green Bank, West Virginia observatory while observing the deep space Hydroxl Ion (OH) maser radiation source designated in star catalog as "W3OH." An FM-modulated RFI source overlapping the W3OH spectral line at 1665 MHz was created artificially using a signal generator. The RFI was removed using the subspace projection algorithm. Snapshot radio camera images (see Section 3.20.3.2) of the source with and without RFI mitigation are shown in Figure 20.20. The source which was completely obscured by interference is now clearly visible.

Typically interference subspace estimation is poor in SP and all other cancelers without a dominant RFI signal so null depth suffers at lower INR levels. Short integration times, needed to avoid subspace



(a) No RFI (b) With RFI (c) RFI Canceled

FIGURE 20.20

W3OH image with and without RFI. The color scale is equivalent antenna temperature (K).

smearing with moving interference, increase covariance sample estimation error which also limits null depth. To address these issues, an SP canceler using auxiliary antennas as in Figure 20.19 and a new parametric model-based SP approach for tracking low INR moving interferers have been proposed which significantly improves null depth [50].

Adaptive beamformers must distort the desired quiescent (interference free) beam pattern in order to place deep nulls on interferers. For astronomy, even modest beamshape distortions can be unacceptable. A small pointing shift in mainlobe peak response, or coma in the beam mainlobe can corrupt sensitive calibrated measurements of object brightness spatial distribution. Due to strict gain stability requirements it has been preferable to lose some observation time and frequency bands to interference rather than draw false scientific conclusions from corrupted on-sky beam patterns.

For PAF beamforming a potential solution is to use one of several classical constrained adaptive beamformers. Due to the inherent tendency for off-axis steered beams with a parabolic dish reflector to develop a mainlobe coma distortion, it would be necessary to employ several mainlobe point constraints to maintain a consistent symmetric beampattern. It has also been demonstrated that without multiple mainlobe constraints, RFI canceling nulls in the beampattern sidelobes can cause significant distortion in the mainlobe.

A more subtle undesirable effect for both PAF and synthesis imaging arrays is that variations in the effective sidelobe patterns due to moving RFI nulling can translate directly to an increase in the minimum detectable signal level for the radiometer. Weak astronomical sources can only be observed by integrating the received power for a long period to obtain separate low variance estimates of signal plus noise power (on source), and noise only (off source). Both signal and noise (including leakage from other deep space source through beam sidelobe patterns) must be stable to an extreme tolerance requirement over the full integration time. Even small variations in the sidelobe structure can significantly perturb background source and noise signal levels, causing intolerable time variation. This sidelobe pattern rumble due to adaptive cancelation increases the "confusion limit" to detection since unstable noise and background are not fully canceled in the on-source minus off-source subtraction. This occurs even if the beam pattern mainlobe is held stable using constrained or robust beamformer techniques.

3.20.4 Positioning and navigation

The Global Positioning System (GPS) is the most widely adopted positioning system in the world. It is a prominent example of what is known as Global Navigation Satellite Systems or GNSS, which represent any system that provides position information to users equipped with appropriate receivers at any time and anywhere around the globe based on signals transmitted from satellites. Currently there are two operating GNSS: GPS (developed by the USA) and Glonass (developed by the former USSR and now by Russia), while there are a number of systems under deployment, such as Galileo in Europe and Compass in China. Despite the differences in the satellite constellation, signal parameters, etc., all of these systems share the same operating principles and use similar types of signals. Therefore, while we will often refer to the case of GPS, all of what we discuss here is also applicable to the other systems as well.

The GPS constellation is formed by approximately 30 satellites orbiting at a distance of about 26,560 km from earth's center. Each satellite transmits several Direct-Sequence Spread-Spectrum

(DS-SS) signals, and the main task of a GPS receiver is to measure the distances to the satellites via the time delay of the signals. In applications requiring high-accuracy positions, the phase of the received signal is also used as a source of information about the propagation delay of the signal. Once the receiver has obtained these distances, it can compute its position by solving a geometrical problem. Apart from the satellites themselves, the core of a GNSS is the ground segment that consists of a set of ground stations monitoring the satellites and computing their positions.

Unlike communication receivers, where timing and phase synchronization are intermediary steps to recovering the transmitted information, for positioning receivers it is the synchronization that is the information. Significantly greater synchronization precision is required in a GNSS receiver than in a communications system. As discussed below, the positioning accuracy of GNSS is degraded by many effects. Multipath propagation and certain types of interference are very difficult to mitigate with single-antenna receivers. Spatial processing has proven to be the most effective approach to combat these sources of degradation, making it possible to obtain in some cases the same accuracy as in a multipath- and interference-free scenario. The next two sections describe the error sources in GNSS, with special emphasis on the multipath effects, and an appropriate signal model for spatial processing. They serve as a justification of why the use of antenna arrays in the context of GNSS has been receiving considerable attention since the mid-1990s. The rest of the sections discuss the advantages and limitations of different approaches for exploiting the spatial degrees of freedom or spatial diversity in satellite-based navigation systems.

3.20.4.1 Error sources and the benefits of antenna arrays in GNSS

The synchronization accuracy demanded by GPS receivers is very stringent, on the order of a few nanoseconds, and exceeds by far the levels usually required in communications receivers. The difficulties in achieving such ranging accuracy are due to the presence of different sources of error, which can be categorized in three groups: (i) the errors due to the ground segment and the satellites, (ii) propagation-induced errors, and (iii) local errors at the receiver. The first category includes the discrepancy between the estimates of the satellite positions and clocks, which are computed by the ground segment and broadcast by the satellites themselves, and the actual values. The second category corresponds to the changes in the propagation delay, phase and amplitude of the signals caused by the atmosphere. Finally, local errors refer to the effects of thermal noise, interference and multipath components.

The largest contributors to the total error budget are typically the ionospheric delay and local effects. The size of the errors in the first category is progressively decreasing as the ground segment and satellites are modernized. Moreover, one can also access alternative providers of more accurate satellite coordinates and clocks. Another option is to use differential methods, where the user receiver makes use of corrections computed by another receiver at a known position, or relative methods, where the position relative to that second receiver is computed. The use of differential or relative methods virtually eliminates the errors from the first category. These methods also help mitigate the propagation-induced errors. Alternatively, the ionospheric delay can be essentially canceled using measurements at two or more frequency bands. In short, the errors from the first two categories can typically be mitigated at the measurement or system levels, and hence the local errors remain as the limiting factor in the ultimate accuracy achievable with GNSS. This is the reason why it is of high interest to use signal processing techniques, and in particular antenna array-based methods, to combat multipath and interference effects in GNSS.

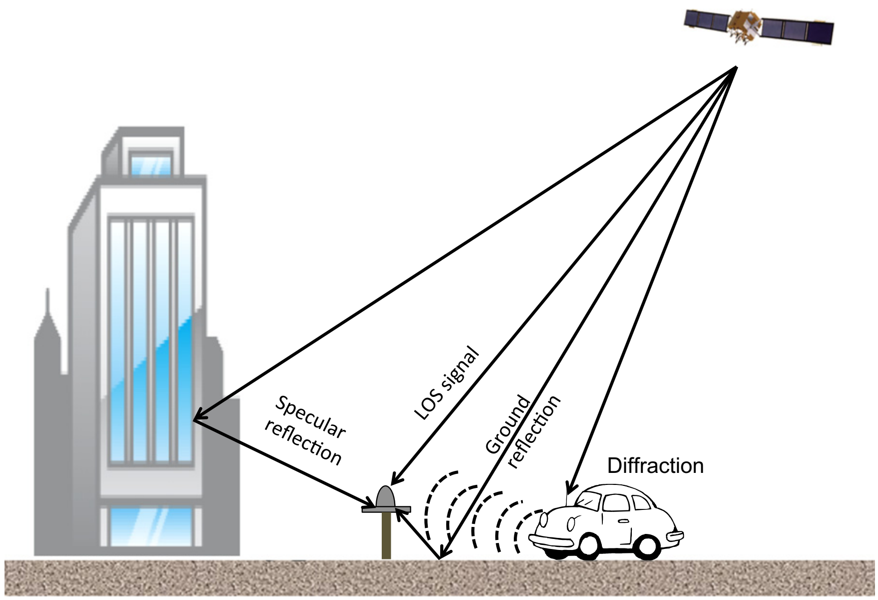


FIGURE 20.21

Environment with multipath propagation.

As in other systems, interference obviously affects the quality of time delay and phase estimates in GNSS. On the other hand, the study of multipath effects requires a different treatment to the one that is typically employed in wireless communications. While multipath components can be useful in communications systems as a source of diversity or to increase the total received signal power, they are always a source of error in navigation systems, and can lead to positioning inaccuracies reaching up to many tens of meters. For the case of a satellite-based transmission, multipath is produced by objects that are close to the receiver, as depicted in Figure 20.21. The only signal of interest in a navigation receiver is the line-of-sight (LOS) signal, since it conveys information about the transmitter-receiver distance through its time delay and phase information. While the multipath in a frequency-flat channel with zero delay-spread theoretically arrives at the same time as the LOS, the resulting fading can lead to signal drop-outs and poor localization performance. A second antenna (i.e., forming a small array) can be used to overcome this difficulty. More challenging are multipath signals that arrive with non-zero delay relative to the LOS, but still within 1–1.5 chip periods of the LOS (for civilian GPS, the chip period is 1 μs, corresponding to about 300 m). Such signals are commonly referred to as *coherent* multipath, and cause biases in the LOS signal time delay and carrier phase estimates. Signal replicas with delays greater than about 1.5 chip periods can essentially be eliminated via the despreading process.

Narrowband or pulsed interference can be canceled in single antenna receivers using excision filters or pulse blanking. Wideband non-pulsed interference cannot be combatted with time-domain processing, but it is in principle an easy target for array processing. Harmful interference usually stands out clearly above the noise, and this makes its identification and subsequent nulling with a spatial filter relatively easy. On the other hand, multipath mitigation is an extremely difficult task in single-antenna receivers and also a difficult problem when using antenna arrays. In the single-antenna case where time-domain methods must be used, the problem is ill-conditioned since one is attempting to estimate the parameters of signal replicas that are very similar to each other. If a reflection and the LOS signal differ by a very small delay (compared to the inverse of the signal bandwidth), they are almost identical and it is very difficult to accurately measure the exact LOS signal delay. On the other hand, the spatial selectivity offered by antenna arrays can be used to differentiate the LOS signal from multipath, since the multipath will arrive from directions different from the LOS (it is very unlikely to have reflectors close to the direct propagation path). The application of spatial processing for multipath mitigation is not without difficulties. The main problem is that the LOS signal and the coherent multipath are strongly correlated, which causes problems for many array processing techniques.

3.20.4.2 Signal model for positioning applications

The signal received by the antenna array can be written as

$$\mathbf{y}(t) = \sum_{k=0}^{D} \alpha_k \mathbf{a}_k x(t-\tau_k) e^{j2\pi f_k t} + \mathbf{n}(t). \tag{20.50}$$

In particular, in our problem the sources are not different signals, but delayed replicas of a single signal. Each replica is shifted by a different Doppler frequency f_k, and its complex amplitude is α_k. The subindex 0 is reserved for the LOS signal, and this implies that $\tau_i > \tau_0, \forall i$. The term $\mathbf{n}(t)$ includes the thermal noise and any (possibly directional) interference. The key parameters of interest for positioning applications are τ_0 and possibly the argument of α_0 (i.e., $\angle\alpha_0$, which is the carrier phase of the LOS signal).

According to the discussion above, we assume that the delays of the replicas are in the range $[\tau_0, \tau_0 + 1.5T_c]$, where T_c is the chip duration. Each replica may represent a single reflection or a cluster of reflections with very similar delays. This leads to different possible parameterizations for the vectors $\mathbf{a}_k$, as listed below:

1. an unstructured spatial signature (i.e., each $\mathbf{a}_k$ is an arbitrary complex vector). In this case, there is an inherent ambiguity between the definition of α_k and $\mathbf{a}_k$, which can be simply avoided by defining $\alpha_k \mathbf{a}_k$ as the overall spatial signature. One element of the spatial signature is identified as α_k, and hence the carrier phase of the LOS signal is given by the argument of that element of the spatial signature.
2. a steering vector (or also referred to as structured spatial signature), which is a function of the DOA.
3. a weighted sum of steering vectors: $\mathbf{a}_k = \sum_{l=1}^{D_k} \alpha_{k,l} \mathbf{a}_{k,l}(\theta_{k,l}, \phi_{k,l})$, where each term corresponds to the amplitude and the steering vector of one of the reflections of the cluster. In this case, the ambiguity between α_k and $\mathbf{a}_k$ can be handled in the same way as in the first model.

The signal $x(t)$ may represent the GNSS signal itself or the signal after some processing. The most common case of processing in our context is the despreading operation, which consists in cross-correlating the received signal with a local replica of the pseudorandom or pseudonoise (PN) sequence.

In this case, the variable t in the signals may be interpreted as the correlation lag. Unlike communications receivers, a single correlation lag is not sufficient. A single correlation lag may be appropriate for data detection but in a GNSS receiver, where the timing of the PN sequence has to be measured, several correlation lags are required. The correlation of the incoming signal with the local sequence is usually computed as a multiply-integrate-and-dump operation, which is carried out for each lag. However, the despread signal, depicted in Figure 20.22, can also be interpreted as a portion of the output of a matched filter. Figure 20.23 shows how the reception of multiple replicas affects the shape of the despread signal, and it is clear from there that identifying the components that form the signal is a very complicated task.

The choice of whether to base the computation of beamformers or other estimation methods on the pre-despreading (pre-correlation) or post-despreading (post-correlation) signal has a crucial impact on the performance and limitations of the array processing algorithms. GNSS signals typically have a Carrier-Power-to-Noise-Spectral-Density (C/N_0) of about 45 dB Hz. The chip rate and hence the bandwidth is greater than 1 MHz, so this results in an SNR on the order of -15 dB or less. This means that the GNSS signals and also their reflections are buried in the background noise. If one computes the spatial correlation matrix $\mathbf{R}_{yy} = E\{\mathbf{y}(t)\mathbf{y}^H(t)\}$ in a pre-correlation scheme, only the noise and interference have a noticeable contribution to the matrix, so in practical terms the "total" correlation matrix $\mathbf{R}_{yy}$ really only represents the noise-plus-interference correlation matrix.

The situation is completely different in the post-correlation scheme. The SNR of the correlation maximum is equal to C/N_0 times the duration of the local reference. The duration of PN sequences in GNSS is several milliseconds, so the SNR of the maximum is typically on the order of several tens of dBs. The average SNR of the signal depends on the length of the portion of the correlation around the maximum that is taken as the observation window. This length is normally not too large, usually only a few chips, so the average SNR stays at the level of tens of dBs. In this case, the post-correlation matrix $\mathbf{R}_{yy}$ includes noticeable contributions from the LOS and reflected signals besides the noise and

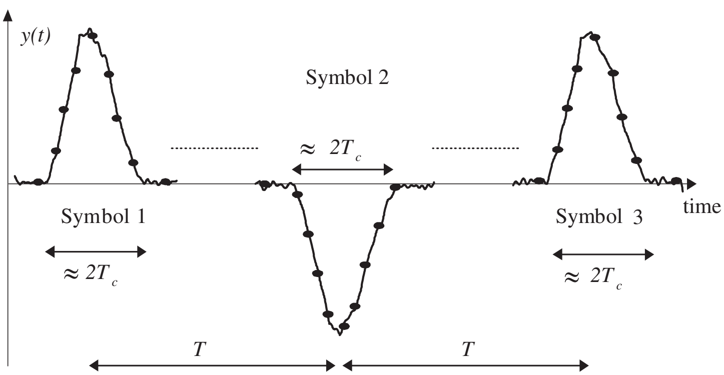


FIGURE 20.22

Qualitative example of the signal at one antenna after the despreading (parameter T is the symbol period: 20 ms in GPS C/A, and T_c is the chip duration: $\sim$1 μs in GPS C/A.)

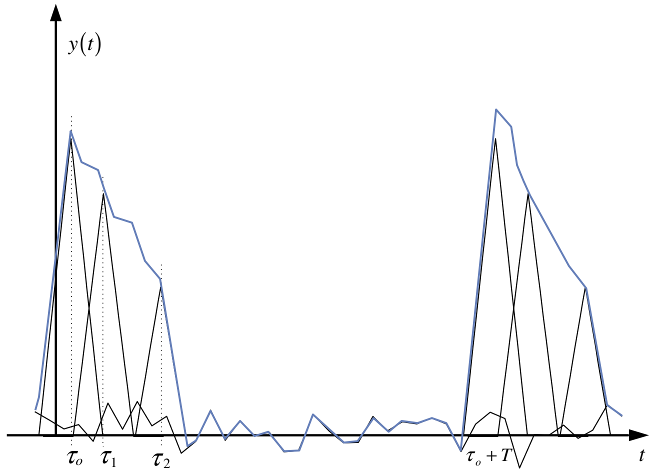


FIGURE 20.23

Qualitative example of the despreaded signal composed of the LOS component and two reflections. These reflections are considered as coherent multipath because their contributions overlap with that of the LOS component.

interference. To conclude, in order to make multipath *visible* in the spatial correlation matrix, one has to work with the post-despreading correlation matrix. If one wants to hide multipath from the spatial correlation matrix, the pre-despreading correlation matrix has to be used.

The location of the beamformer (if any) with respect to the despreader has an impact on computational complexity, but it does not have an effect on performance since only its position within a set of linear operations is changed. Note that we are referring here to the placement of the beamformer in the receive chain, and not to the input data used for its computation, which is a totally different aspect as explained above. Some examples of the placement of the beamformer as well as the input data used for its computation are shown in Figures 20.24 and 20.25. Note that all combinations are in principle possible, although some cases, such as pre-despreading beamforming with weights computed using the post-despread signals, do not have a clear justification. As an example of a typical approach, the beamforming vector is computed using the pre-despread correlation matrix as $\mathbf{w} = \mathbf{R}_{yy}^{-1}\mathbf{a}_0$, and applied to the despread signal to obtain $z(t) = \mathbf{w}^H\mathbf{y}(t)$. In the first formula, the symbol $\mathbf{y}$ refers to signals before despreading, whereas in the second formula it refers to the despread signals.

3.20.4.3 Beamforming

The objective is to synthesize an array pattern that attenuates the reflections and interference. In the context of GNSS, antenna-array beamformers are customarily referred to as CRPAs (Controlled Reception Pattern Antennas). Adaptive (or data-dependent) beamforming is appropriate for situations where little *a priori* information about the scenario is available, or when the scenario is likely to change with time.

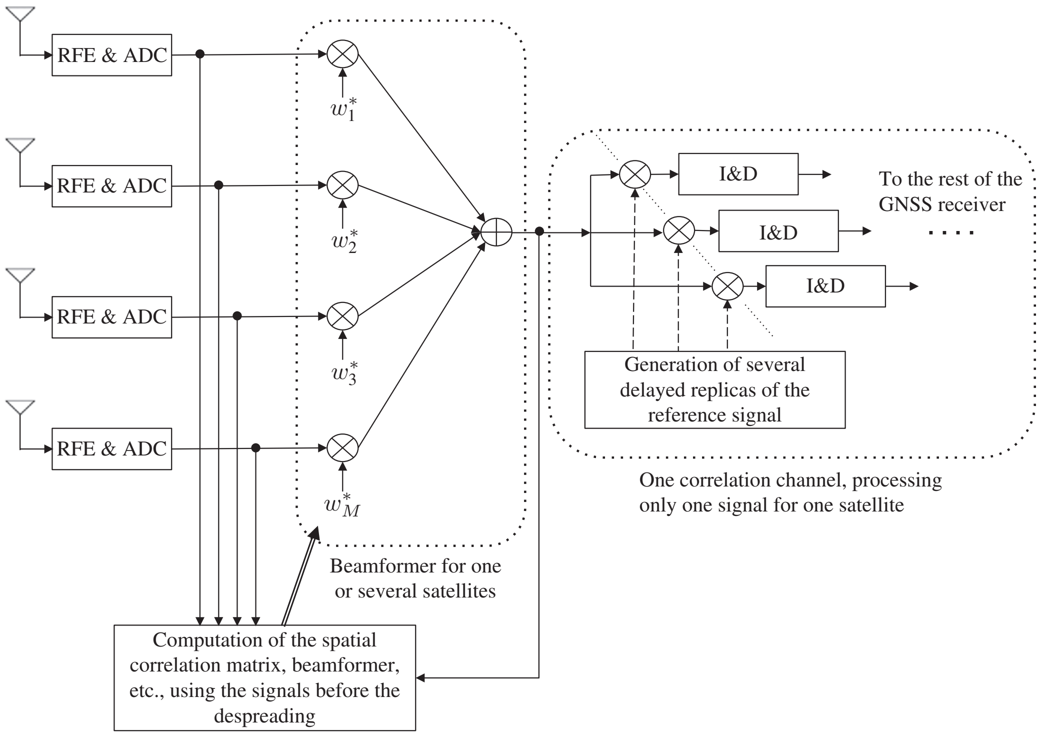


FIGURE 20.24

Example of a GNSS receiver using an antenna array where the beamformer is applied before despreading and it is computed using the pre-despread signals. The output of the beamformer is processed by a conventional GNSS receiver channel, as if it was the signal coming from a single antenna. Either option is possible: the beamformer can be the same for all satellites, or different beamformers for different satellites can be used. The complexity bottleneck is due to the fact that the beamformer weights are applied to high-rate samples coming from the RF front-end.

This is the typical situation for user receivers. On the other hand, deterministic (or data-independent) beamforming is more suitable for static and relatively controlled scenarios. This is typically the case for ground reference stations. These reference stations refer to both the receivers that form part of the ground segment of the GNSS (i.e., those receivers providing the measurements used to compute the position of the satellites) and the user receivers that are static and typically used as references in differential or relative positioning.

3.20.4.3.1 ***Adaptive beamforming***

As outlined below, several different types of adaptive beamforming algorithms have been proposed for GNSS. Some of these are adaptations of standard algorithms, others have been designed specifically for conditions specific to positioning applications.

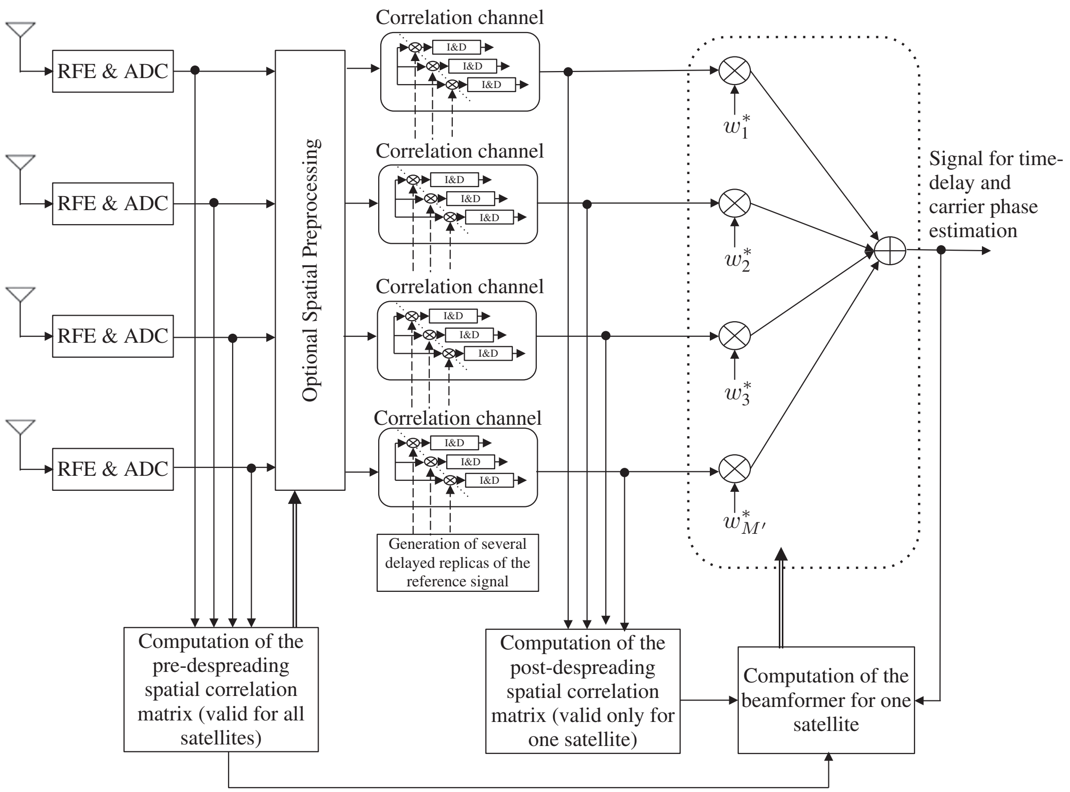


FIGURE 20.25

Example of a GNSS receiver using an antenna array where the beamformer is applied after despreading. The beamformer vector is calculated using the pre-despreading or the post-despreading spatial correlation matrix. An optional spatial preprocessing block is included, which can be used to cancel some spatial sectors. The number of outputs of the preprocessing block, M', is equal to or smaller than the number of antennas, M. In this configuration, the application of the beamformer weights do not entail a significant computational load because the correlation channels generate samples at a very low rate. Hence the fact that a different beamformer is applied for each satellite is not a problem. Here the computational bottleneck comes from the need to use a correlation channel at each antenna or at each output of the preprocessing block.

Algorithms employing a spatial reference: These approaches are based on knowledge of the steering vector of the LOS signal, $\mathbf{a}_0$. Assuming this *a priori* information is reasonable in some GNSS applications since the satellite position can be known thanks to the navigation message (transmitted by the satellite itself) or to assistance from ground stations, and a rough estimate of the receiver position may be available from previous position fixes or from the application of a basic positioning algorithm (e.g., using only one antenna and not exploiting the antenna array). Moreover, the accuracy of the satellite and receiver positions is not important in determining the DOA of the signal; errors of several hundreds of meters can be tolerated without affecting the satellite DOA estimate, given that the satellite-receiver distance is more than 20,000 km. However, the assumption of a known $\mathbf{a}_0$ relies on the availability of array calibration and especially on the knowledge of the receiver orientation (also known as attitude in the GNSS literature). Errors or uncertainty in the array response correspond to the standard calibration problem found in many applications of antenna arrays, and robust methods developed for generic applications are also applicable here. On the other hand, the need to know the receiver orientation is a feature more specific to GNSS receivers. Assuming that $\mathbf{a}_0$ is known, the use of the MVDR beamformer (and variants) is possible, and it is most appropriate to apply them in a pre-despreading scheme. If these beamformers are computed with the post-despreading correlation matrix and multipath components are present, they will suffer from the cancellation of the desired signal.

Algorithms employing a temporal reference: These methods are based on knowledge of the GNSS signal waveform. Knowledge of the waveform can be exploited to design a beamformer that minimizes the difference between its output and the reference signal (e.g., as in a Wiener filter). In practice, the situation is not that straightforward because even though the shape of the signal is known, the delay and frequency shift are not, so the beamformer weights and the signal parameters have to be computed jointly or iteratively. The expression for the beamformer is

$$\mathbf{w}_T = \mathbf{R}_{yy}^{-1}\mathbf{r}_{yx}\left(\hat{\tau}_0, \hat{f}_0\right), \tag{20.51}$$

where $\mathbf{r}_{yx}(\hat{\tau}_0, \hat{f}_0)$ is the cross-correlation between the array output and a local replica of the LOS signal generated using estimates of its delay and frequency shift, $\hat{\tau}_0$ and $\hat{f}_0$, respectively. In this case, it only makes sense to work with correlations computed after the despreading, otherwise the contribution of the GNSS signals is hardly present in the correlations. This beamformer is able to cancel interference, but its performance in the presence of multipath is not satisfactory, although not as bad as with spatial-reference beamformers. The temporal reference beamformer combines the multipath and the LOS signal in a constructive manner, so as to increase the SNR. This is useful behavior in communications but not in navigation systems, since the increase in SNR comes at the price of a bias in the estimation of the delay and phase due to the presence of strong multipath at the beamformer output.

Hybrid beamformers: The opposite behavior of the spatial-reference and temporal-reference beamformers suggests that their combination may have good properties. Both of them provide the LOS signal at the output, but the former changes the phase of the multipath so that it is roughly in counter-phase with the LOS signal, whereas the later modifies the multipath phase to align it with that of the LOS signal. Therefore, if the output of both beamformers is added together, the multipath will tend to cancel. This observation has led to the proposal of a hybrid beamformer that can be expressed as

$$\mathbf{w}_H = \beta\mathbf{w}_T + \gamma\mathbf{w}_S, \tag{20.52}$$

where $\mathbf{w}_S$ is a spatial-reference beamformer, and β and γ are two scalars weighting the contribution of each beamformer. When $\mathbf{w}_S$ is chosen as the MVDR beamformer, it can be shown that the optimal weights are

$$\begin{aligned} \beta &= \alpha_0, \\ \gamma &= 1 - \alpha_0\mathbf{a}_0^H\mathbf{R}_{yy}^{-1}\mathbf{r}_{yx}(\tau_0, f_0). \end{aligned} \tag{20.53}$$

Since the optimal weights depend on the unknown parameters to be estimated, a practical way to proceed is to use an iterative algorithm where the calculation of the beamformer according to (20.52) is done using the previous estimates of $\{\alpha_0, \tau_0, f_0\}$, and next these estimates are updated using the output of the just computed beamformer.

Blind algorithms: This class of methods refers to techniques that do not exploit *a priori* knowledge of the exact signal or the steering vector, and hence are more robust to errors in these assumptions. Examples of such methods include those based on the constant modulus (CM) assumption, cyclostationarity and the power inversion approach. The civil GPS signal in current use, referred to as the C/A signal, has constant modulus because it is formed by almost rectangular chips. Most other GNSS signals also satisfy the CM property. However, this property cannot be exploited before despreading since the array cannot provide enough SNR gain to bring the signal above the noise. Therefore, the CM beamformer has to be applied after despreading, but in order to do so the despread samples corresponding to the LOS signal have to be CM. This happens when only one sample per integration period is used. However, the presence of multipath does not alter the constant-modulus property of the signal, so the CM beamformer is not useful in combating multipath.

GNSS signals are obviously cyclostationary since the repeated use of the PN spreading sequence introduces periodicity into the statistics of the signal. The fact that several repetitions of the PN code are present during a bit time (a property sometimes referred to as self-coherence) can also be exploited, as depicted in Figure 20.26. Interference will not have in general this structure, so it is possible to design the beamformer by imposing that its output should be as similar as possible to a version of itself delayed by a time equal to the PN code duration. Because of the same reasons as in the case of the CM beamformer, this technique should be applied to the despread signals and it will only be effective against interference and not multipath.

A very simple but rather effective approach is the power inversion beamformer. The weights are obtained as the beamformer vector that minimizes the total output power subject to a simple constraint to avoid the null solution. The constraint is chosen without using any information about the signal, typically forcing a given beamformer coefficient to be equal to one. This method has to be applied to the signals before despreading and, since the response is independent of the GNSS signals, it may happen that some nulls of the reception pattern are near to the DOAs of some of the GNSS signals. However, this situation can often be accepted since it is assumed that many GNSS satellites will be

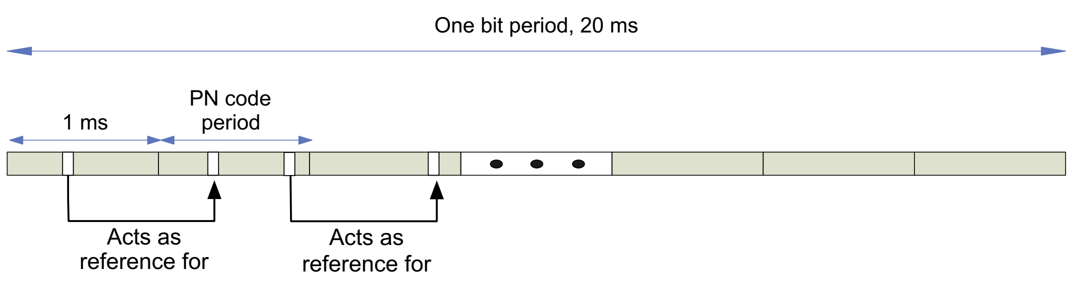


FIGURE 20.26

Structure of the GPS signal that allows to implement self-coherence restoration beamforming methods.

visible (around 10 satellites), and if a few of them are lost due to the coincidence of the pattern nulls with their DOAs, there will still be a sufficient number of satellite signals available (i.e., four or more) to compute the position. In this method, all satellites are received through the same beamformer, so it offers the possibility of being deployed as an add-onto existing single-antenna receivers. This is an important advantage of this method; more sophisticated beamformers that require information provided the receiver (e.g., the DOA or the delay of the LOS signal) or that generate one beam per satellite cannot be coupled with existing single-antenna receivers and require the development of a completely new receiver. The number of antennas used with power inversion should be large enough to cancel the existing interference sources, but not much larger in order not to increase the number of nulls in the pattern and thus the probability that a GNSS signal is canceled. The power inversion approach is popular in military systems where jamming from highly maneuverable sources (fighter jets) is a pivotal concern. The fast maneuvers of these vehicles makes the use of spatial-reference beamformers virtually impossible, and therefore a simple and robust method like power inversion, that does not need any reference or calibration procedure, is an excellent option.

3.20.4.3.2 *Deterministic beamforming*

Although it is recognized that data-dependent beamformers are more powerful in general than deterministic versions, there are some situations where the latter may be advantageous. Deterministic beamformers are clearly more robust against calibration errors and other uncertainties in the signal parameters. Moreover, if the desired and non-desired signals are known to be confined to distinct spatial regions, the deterministic design may offer an adequate solution since the problem reduces to designing a spatial filter with given pass and stop bands. This *a priori* spatial separability occurs in several circumstances in GNSS, particularly in GNSS ground stations. In this case, the interference is normally ground-based, and the multipath normally arises from ground-based scatterers, so both interference and multipath impinge on the receiver from relatively low elevation angles. This is contrasted with the satellite signals, which originate from the entire upper hemisphere. Thus, as illustrated in Figure 20.27, a fixed beamformer can be designed to minimize reception of signals from these low elevations. The complicating factor here is that an upwards-facing array typically cannot provide a sharp stop-band to pass-band transition for directions near end-fire. Another advantage of deterministic beamformers is that they allow an easier control of the trade-off between array gain (understood here as the increase of the ratio between the desired signal power and the white noise power) and interference cancellation. In adaptive beamformers, these two characteristics are tightly coupled. For instance, with MVDR, the presence of a strong interference gives rise to a deep null in the pattern, and this null necessarily increases the beampattern in other directions, thus degrading the array gain.

3.20.4.4 DOA estimation

DOA estimation algorithms can be used as a processing stage prior to beamforming. If the DOAs of the LOS and reflected signals and interferences can be determined, then it is possible to design a beamformer that, for instance, attenuates the reflections and interferences while maximizing the SNR of the LOS signal. Given the identifiability limitations of DOA estimation methods and their sensitivity to highly correlated signals, such a method would likely only be suitable in situations where there were a small number of multipath and interference arrivals in addition to the LOS signal. The DOAs of the

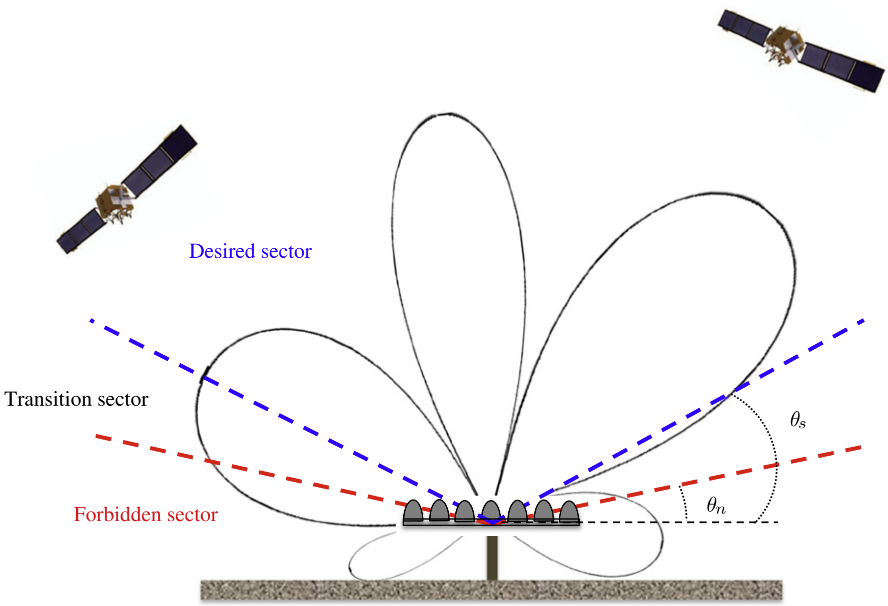


FIGURE 20.27

Desired and forbidden regions for the design of deterministic beamformers.

non-desired signals can typically be obtained in two stages; interference sources can be localized prior to despreading, while the DOAs for multipath sources would have to be found after despreading. Even when it is not possible to use estimation methods to determine the DOAs of all signals, such methods can still be useful for detecting scenarios where the LOS signal is obstructed. Such information is critical in tracking applications, since highly erroneous estimates due to non-LOS measurements can be eliminated.

3.20.4.5 Array-based parameter estimators

Probably the most rigorous approach to the use of antenna arrays for multipath and interference mitigation consists not in focusing on the use of the array to synthesize a beam that attenuates those unwanted signals, but in formulating the measurement of the time delay and carrier phase of the LOS signal as an estimation problem. The Maximum Likelihood (ML) approach used in many other areas of array processing can also be used here. There is a large variety of models and assumptions that have been used to derive ML estimators, as we briefly outline below. Recall the expression of $\mathbf{y}(t)$ in (20.50) and assume that K samples or snapshots are taken from the array, which form the columns of a matrix $\mathbf{Y}$. This matrix can be expressed as

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\Gamma}(\mathbf{X}(\boldsymbol{\tau}) \odot \mathbf{D}(\mathbf{f})) + \mathbf{N}, \tag{20.54}$$

where $\mathbf{A} = [\mathbf{a}_0, \ldots, \mathbf{a}_D]$, $\boldsymbol{\Gamma} = \text{diag}\{\boldsymbol{\alpha}\} = \text{diag}\{\alpha_0, \ldots, \alpha_D\}$, $\boldsymbol{\tau} = [\tau_0, \ldots, \tau_D]$, and $\mathbf{f} = [f_0, \ldots, f_D]$. The (k,n)th components of $\mathbf{X}$ and $\mathbf{D}$ are: $[\mathbf{X}(\boldsymbol{\tau})]_{k,n} = x(t_n - \tau_k)$ and $[\mathbf{D}(\mathbf{f})]_{k,n} = e^{j2\pi f_k t_n}$. Matrix $\mathbf{N}$ contains the snapshots of $\mathbf{n}(t)$, which includes all disturbances present in the received signal except for the multipath components. In particular, as it can include directional interference, it is logical to assume that $\mathbf{N}$ is spatially colored. The spatial correlation matrix is denoted as $\mathbf{Q}$ and it is in general assumed to be unknown. If we further assume for simplicity that $\mathbf{N}$ is temporally white, zero-mean and circularly-symmetric complex Gaussian distributed, the negative log-likelihood function can be expressed as

$$
\begin{aligned}
L(\mathbf{Y}, \mathbf{A}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \mathbf{f}) = {} & M \ln \det(\mathbf{Q}) \\
& + \operatorname{tr}\left(\mathbf{Q}^{-1}(\mathbf{Y} - \mathbf{A}\boldsymbol{\Gamma}(\mathbf{X}(\boldsymbol{\tau}) \odot \mathbf{D}(\mathbf{f})))(\mathbf{Y} - \mathbf{A}\boldsymbol{\Gamma}(\mathbf{X}(\boldsymbol{\tau}) \odot \mathbf{D}(\mathbf{f})))^H\right),
\end{aligned} \tag{20.55}
$$

where $\mathbf{A}$ can be replaced with the alternative parameterization discussed in Section 3.20.4.2. The ML estimates of the different parameters can be obtained as the arguments that minimize (20.55). This optimization problem, in its most general form, cannot be easily tackled because it has a large number of variables and it is highly non-linear (and non-convex). As discussed below, simplifications are possible depending on various modeling assumptions and how the problem is parameterized.

Unstructured spatial signatures and spatially white noise: The assumption of spatially white noise allows determinant in (20.55) to be eliminated, and the ML problem turns into a least squares problem. Since the resulting problem has the same structure as the estimation of the DOAs of unknown deterministic signals, most DOA estimations algorithms can be adapted to the estimation of time delays and frequencies in this new setup. A number of algorithms have been developed based on this parallelism between conventional DOA estimation and time delay estimation with unstructured spatial signatures. However, all techniques derived under the assumption of spatial whiteness suffer from a lack of interference mitigation capability. In addition, the use of unstructured spatial signatures causes the variance of the estimates to grow when the difference in delay and frequency shift of the replicas becomes smaller (see Figure 20.28).

Structured spatial signatures and spatially white noise: In this approach, the steering vectors $\mathbf{a}_k$ are parameterized by the corresponding DOAs instead of being arbitrary complex vectors. This change makes the estimation problem more non-linear and hence more complex to solve, but on the other hand it provides in general more accurate estimates because the model parsimony is improved. The increased accuracy is largely observed when the signals are very close to each other in the delay and frequency dimension.

Unstructured spatial signatures and unknown spatial correlation: The ML estimator for this model involves the determinant of the correlation matrix of the fitting residuals, and hence it does not correspond to a least squares problem like the techniques derived under the assumptions of unstructured spatial signatures and spatial white noise. Consequently, it is not possible to establish a clear parallelism with DOA estimation algorithms, but techniques have been developed that are robust to directional interference. In addition, asymptotically equivalent algorithms have been proposed that admit a simple solution based on polynomial rooting.

Structured spatial signatures and unknown spatial correlation: This constitutes the most detailed model for the problem at hand, and also the one that leads to the best performance as long as there are no severe model mismatches with respect to reality. A direct optimization for this model requires a highly non-linear search in the DOA, time delay and frequency spaces, which cannot be implemented easily in an efficient manner. This limitation has been recently overcome by applying the Extended Invariance

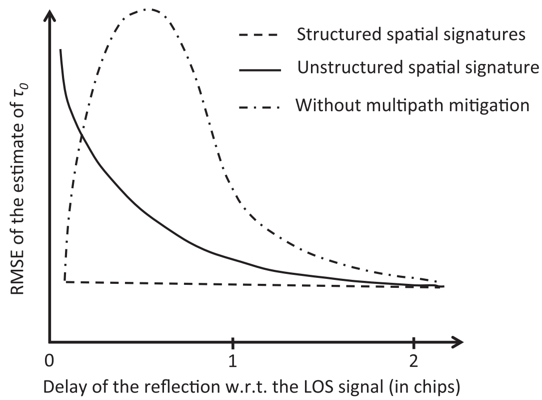


FIGURE 20.28

Qualitative representation of the behavior achieved with the estimators derived under different models. It is assumed that the LOS signal and one reflection are received. The solid line corresponds to the model with unstructured spatial signatures. The dashed line corresponds to the model with structured spatial signatures; the line may be not totally constant, but in any case it shows a much smaller dependence with the delay than the solid line. When the reflection is not mitigated, the errors have the shape depicted by the dash-dotted line, where the increase in the RMSE is normally not due to an increase of the variance as in the other two cases, but to the existence of large bias.

Principle (EXIP). The EXIP technique begins with ML estimates corresponding to the model with unstructured spatial signatures and unknown spatial correlation described above, which can be obtained with relatively low complexity. Then, these estimates are refined by means of a weighted least-squares fit, resulting in improved estimates that have the same asymptotic accuracy as the exact ML estimates directly derived from the model with structured spatial signatures and unknown spatial correlation. The refinement approach boils down to a DOA estimation problem, and if the antenna array response has a Vandermonde structure, a polynomial-rooting based DOA estimator can be used. Thus, in the most difficult case involving DOAs and time delays of several replicas received in noise of unknown spatial correlation, estimates asymptotically equivalent to the ML ones can be obtained simply rooting two polynomials.

GNSS-specific signal models: Although the methods described above rigorously follow the logic of model-based estimation, they may present limitations in some practical conditions. This is exemplified by these two cases:

- The model in (20.54) assumes that the received signal is formed by several replicas of the transmitted GNSS signal. The real received signal may not be constituted by a few clearly defined reflections, but instead may consist of a large number or even a continuous distribution of components. In principle, the accurate modeling of this reality would require the use of a very large value for D in the model, and this would prohibitively increase the number of parameters to estimate and hence the complexity. One can argue that a reasonable model can be obtained using only a few replicas that capture most of the contribution of the actual multipath environment. But even if this is true, the estimation of the appropriate value of D (large enough to represent well the received signal, but not too large to avoid overfitting of the model and excessive complexity) is an issue that needs to be addressed in any of the model-based estimators presented above.

- There are some particular aspects of the GNSS application that are not adequately exploited. For instance, the above methods provide estimates of all parameters in the model, but this is overkill in GNSS, where only the parameters of the LOS signal are of interest for positioning. Moreover, there is some side information that is not employed in the models, such as the *a priori* knowledge of the DOA of the LOS signal, and the fact that reflections always arrive later than the LOS signal and usually with smaller amplitudes.

As a consequence, a different way of proceeding consists in abandoning very detailed models attempting to provide a very precise description of the received signal (and maybe not achieving it because the signal includes other effects not accounted for in the model), in favor of simpler models that focus on particular aspects related to the GNSS application, even if they do not necessarily provide a comprehensive representation reality. For example, when LOS DOA can be assumed to be known, the vector $\mathbf{a}_0$ can act as a spatial reference to the LOS signal, what makes it possible to approximately model the reflections as part of the noise term with unknown spatial correlation and, hence, the value of D is simply taken as zero. It can be shown that there is an equivalence between the estimator resulting from this simplified model and the hybrid beamformer presented in Section 3.20.4.3.1, while a model with $D > 0$ would provide a more accurate representation of reality (at the expense of increased complexity), results have shown that there is typically not a large penalty in assuming $D = 0$. Problems will arise in situations when the delay of the reflections are close to that of the LOS signal, in which case the time delay and carrier phase estimates are biased. However, the degradation in such cases remains bounded.

3.20.5 Wireless communications

The use of antenna arrays in wireless communications provides one or more of the following types of advantages: diversity gain, array (or beamforming) gain and multiplexing gain. Given that in practice these gains lead to increases in capacity and spectral efficiency as well as improved robustness against fading, multiple antenna (or MIMO) techniques have been included in recent wireless communication standards. While the signal models and algorithms for multi-antenna wireless communications have been studied in detail in other chapters, here we focus on specific ways in which multiple antennas are exploited in current wireless standards.

3.20.5.1 Multiple antennas techniques in LTE

MIMO constitutes an essential element of LTE in order to achieve the highly demanding requirements for transmission rate and spectral efficiency. LTE exploits multiple antennas for both diversity and multiplexing [78,79, Ch.11], and also for both the downlink and uplink portions of the network.

3.20.5.1.1 Diversity schemes

Various sources of diversity are available to average out channel variations due to fading. This includes time and frequency diversity, as well as transmit and receive diversity. Receive diversity is mandatory for user handsets, usually referred to as UE's (User Equipment). It is the baseline receiver functionality for which performance requirements are defined. The typical method consists in performing maximum ratio combining (MRC) of the signals received at several antennas. We will focus however on transmit diversity since many schemes were analyzed in detail during the standardization phase of LTE. Some of the characteristics sought for the final selection of techniques were:

- Absence of puncturing in the presence of correlated channels. This eliminated the use of Cyclic Delay Diversity (CDD) and Precoding Vector Switching (PVS) in favor of block-code-based schemes.
- Low decoding complexity, which eliminated the option of non-orthogonal block codes.
- Power efficiency. Each antenna has an instantaneous power limitation, so the power that is not employed during one OFDM symbol (referred to as a "resource element" in LTE) cannot be shifted to the following ones. On the other hand, power can be adaptively allocated in the frequency domain; power that is not used in some subcarriers can be employed in others. The objective of using the maximum available power makes it advisable to select schemes where all antennas transmit at all times, though not necessarily in all subcarriers. This objective takes precedence over achieving a uniform power distribution in the frequency domain.
- Robustness to channel estimation errors. Orthogonal block codes lose the orthogonality property due to channel estimation errors at the receiver. As a certain level of error is unavoidable, it has to be checked that these errors do not cause large interfering terms. Estimation errors are not the only source of loss of orthogonality; variations of the channel that violate some design assumptions (such as the channel is constant across a certain group of subcarriers or during some symbols) may also create self-interference. It is desired to use techniques that do not make stringent assumptions about the evolution of the channel in time or frequency. Moreover, the quality of the channel estimation is not necessarily the same at all antennas. This means that all antennas are not statistically equivalent on average and a proper balancing of the symbols among them is needed.
- Good adaptation to the structure of the signals. The signals are mapped to two-dimensional resource blocks formed by a certain number of symbols and subcarriers. Some codes must be applied over a number of symbols or subcarriers that is a multiple of a given value (typically two or four). It may be easier to achieve the structure required by the code in one of the two dimensions. Since there are many more subcarriers than symbols forming a resource block, it is usually simpler to apply the code in the frequency domain because selecting a certain number of subcarriers is more manageable than changing the number of symbols in a block. Furthermore, in LTE the number of available OFDM symbols in a resource block is often odd.
- Reduced inter-cell interference. One must consider the impact of diversity techniques on the interference produced in neighboring cells.

For two-transmit-antenna diversity, the well-known Alamouti code is applied in the frequency domain, constituting a Space-Frequency Block Code (SFBC). If $y^{(p)}(k)$ denotes the symbols transmitted from the pth antenna on the kth subcarrier, at a given OFDM symbol period, the transmission strategy of the eNodeB (i.e., the base station in LTE terminology) can be represented as follows:

$$\begin{bmatrix} y^{(0)}(k) & y^{(0)}(k+1) \\ y^{(1)}(k) & y^{(1)}(k+1) \end{bmatrix} = \begin{bmatrix} x(n) & x(n+1) \\ -x^*(n+1) & x^*(n) \end{bmatrix}, \tag{20.56}$$

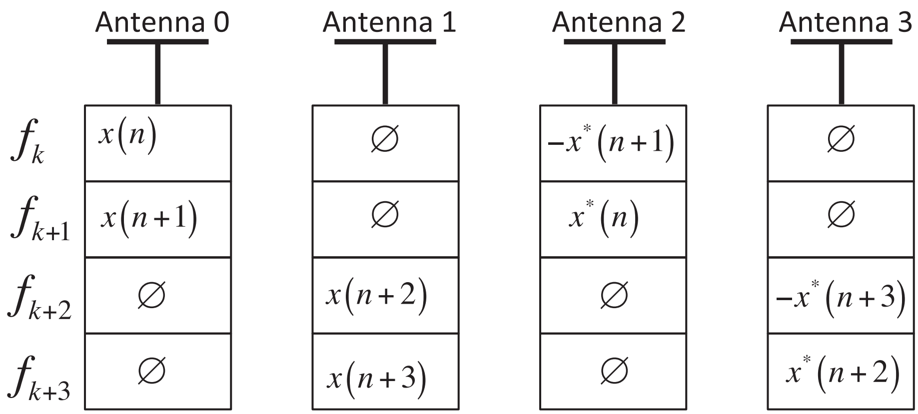


FIGURE 20.29

Space-Frequency Block Code for four antennas used in LTE.

where $x(n)$ represents the stream of symbols to be transmitted. In the case of four transmit antennas, the previous code is applied to each pair of antennas. Each pair of antennas uses a different set of frequencies, and hence the scheme is referred to as SFBC-FSTD, where FSTD stands for Frequency Shift Transmit Diversity (also known as Frequency Switched Transmit Diversity). This is depicted in Figure 20.29, and can be expressed as

$$\begin{bmatrix} y^{(0)}(k) & y^{(0)}(k+1) & y^{(0)}(k+2) & y^{(0)}(k+3) \\ y^{(1)}(k) & y^{(1)}(k+1) & y^{(1)}(k+2) & y^{(1)}(k+3) \\ y^{(2)}(k) & y^{(2)}(k+1) & y^{(2)}(k+2) & y^{(2)}(k+3) \\ y^{(3)}(k) & y^{(3)}(k+1) & y^{(3)}(k+2) & y^{(3)}(k+3) \end{bmatrix} = \begin{bmatrix} x(n) & x(n+1) & 0 & 0 \\ 0 & 0 & x(n+2) & x(n+3) \\ -x^*(n+1) & x^*(n) & 0 & 0 \\ 0 & 0 & -x^*(n+3) & x^*(n+2) \end{bmatrix}. \tag{20.57}$$

This mapping is a full-rate orthogonal code with diversity order equal to two, which is smaller than the possible maximum of four since full-rate full-diversity orthogonal codes do not exist for four antennas and complex symbols. Note also that each pair of symbols uses antennas $\{0, 2\}$ and $\{1, 3\}$. This is because the channel estimates are better in antennas 0 and 1 since more pilot symbols are employed for these antennas than for antennas 2 and 3. Thus, each pair of symbols makes use of one of the antennas for which the receiver can obtain better channel estimates and another antenna for which the estimates are worse.

3.20.5.1.2 *Multiplexing schemes*

LTE supports closed-loop and open-loop MIMO transmission in the downlink using $P = 2$ or 4 antennas and a number of multiplexing layers equal to $\upsilon = 1, 2, 3,$ or 4. A layer is a term used in LTE to denote the different data streams to be transmitted simultaneously using spatial multiplexing. As a consequence, the number of layers represents the multiplexing gain and cannot exceed the number of transmit antennas; thus, $\upsilon \leq P$. The number of layers is also referred to as the rank of the transmission. The mapping of between codewords (i.e., an independently encoded data block) and layers is also specified in LTE. For a transmission rank greater than 1, up to two codewords can be transmitted. In this case, each codeword is assigned to each layer if $\upsilon = 2$, one codeword is assigned to one layer and the other codeword is split between the other two layers if $\upsilon = 3$, and each codeword is mapped to a different pair of layers if $\upsilon = 4$. Multi-codeword transmission allows for the use of the computationally simpler MMSE-SIC (Minimum Mean Square Error-Successive Interference Cancellation) detector, providing comparable performance to the more complex ML detector applied to the single-codeword case, which on the other hand enjoys an advantage in terms of ARQ ACK/NACK signaling.

The relation between the symbols at the antenna ports, $y^{(p)}(n)$, and the symbols in layer l, $x^{(l)}(n)$, is

$$\begin{bmatrix} y^{(0)}(n) \\ \vdots \\ y^{(P-1)}(n) \end{bmatrix} = \mathbf{W}(n) \begin{bmatrix} x^{(0)}(n) \\ \vdots \\ x^{(\upsilon-1)}(n) \end{bmatrix}, \quad P = 1, 2, \text{ or } 4, \quad P \geq \upsilon = 1, 2, 3, \text{ or } 4, \tag{20.58}$$

where $\mathbf{W}(n)$ is a $P \times \upsilon$ precoding matrix. Next we describe the closed- and open-loop approaches to forming this matrix.

Closed-loop multiplexing schemes: The precoding matrices belong to a codebook. The receiver selects the best precoding matrix based on its current channel estimates and feeds back an index to the transmitter. For the case of rank-1 transmission with 2 antennas, the precoders are

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ j \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -j \end{bmatrix}. \tag{20.59}$$

The elements of the precoders are limited to the QPSK alphabet $\{\pm 1, \pm j\}$ to reduce computational complexity at the UE by avoiding the use of complex multiplications. Moreover, there are no amplitude differences between antennas because it is desired to use the maximum available power at each antenna. These two properties are also valid for the other cases, with the caveat that with four antennas the elements of the matrices can also belong to the 8-PSK alphabet: $\{\pm 1, \pm j, (\pm 1 \pm j)/\sqrt{2}\}$.

For the case of rank-2 transmission with 2 antennas, the precoders are

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix}. \tag{20.60}$$

The codebook for four antenna ports is formed by 16 matrices, which are obtained from 16 generating vectors, $\mathbf{v}_k$, whose components belong to the 8-PSK alphabet by applying the Houselholder matrix definition: $\mathbf{I} - 2\mathbf{v}_k\mathbf{v}_k^H$. The precoders for ranks lower than four are obtained by a selected subset of the columns of each matrix. This makes it straightforward to fulfill the nested property, whereby columns of lower rank precoders are subsets of the columns of higher rank precoders, which considerably facilitates the precoder evaluation at the UE. LTE admits both frequency-selective precoding, in which precoding weights are selected independently for different sub-bands of bandwidth ranging from 360 kHz to 1.44 MHz, and also wideband precoding, where a single set of single precoding weights are applied to the entire transmission band.

Note that rank-1 transmission amounts to beamforming. Besides the beamforming case, LTE also allows for UE-specific beamforming, which is not based on the feedback of precoding-related information, but on channel state information obtained by the eNodeB using for instance DOAs measured from the uplink signals or exploiting reciprocity in TDD scenarios.

Open-loop multiplexing schemes: The same diversity schemes as described in Section 3.20.5.1.1 are used for rank-1 open-loop communication. For higher ranks, the general approach is to employ layer cycling together with precoder cycling. Layer cycling is implemented by means of CDD (Cyclic Delay Diversity), and the net effect is to circularly change the order of the columns of the precoding matrix. Specifically, this type of CDD is called long-delay CDD in LTE terminology. This means each layer is transmitted using a different column of the precoding matrix at successive OFDM symbols. The precoder cycling consists simply of changing the precoding matrix after each set of υ resource elements, that is to say, when a complete circular shift of the current matrix has been done. The logic behind this approach is that precoder cycling provides a new realization of SINRs across the layers every time the precoding matrix is changed, and layer cycling makes each codeword experience an SINR that is the average of the SINRs of the layers because each codeword ends up using all columns of the precoding matrix.

In order to put the description above in formulas, we can consider for example the case of $P = 4$ antennas and $\upsilon = 3$ layers. The description is also valid for two or four layers with obvious modifications. The same relation as in (20.58) is valid with the replacement of $\mathbf{W}(n)$ with

$$\mathbf{W}\left(\left\lfloor \frac{n}{\upsilon} \right\rfloor \bmod 4\right)\mathbf{D}(n)\mathbf{U}, \tag{20.61}$$

where $\mathbf{W}(n)$ represents one of the 16 matrices previously mentioned, matrix $\mathbf{D}(n)$ applies CDD in the frequency domain and $\mathbf{U}$ is the $\upsilon \times \upsilon$ DFT matrix:

$$\mathbf{U} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-j2\pi/3} & e^{-j4\pi/3} \\ 1 & e^{-j4\pi/3} & e^{-j8\pi/3} \end{bmatrix}. \tag{20.62}$$

The cyclic delay applied to the lth layer is equal to a fraction l/υ of the symbol duration, and then the CDD matrix is

$$\mathbf{D}(n) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-j2\pi n/3} & 0 \\ 0 & 0 & e^{-j4\pi n/3} \end{bmatrix}. \tag{20.63}$$

The set of possible precoding matrices contains four elements (for any number of layers), and the index $(\lfloor \frac{n}{\upsilon} \rfloor \bmod 4)$ selects another matrix every time a given matrix has been used for υ symbols. The important fact is that the combined effect of the CDD and the DFT matrices is

$$\mathbf{D}(3m)\mathbf{U} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-j2\pi/3} & e^{-j4\pi/3} \\ 1 & e^{-j4\pi/3} & e^{-j8\pi/3} \end{bmatrix},$$
$$\mathbf{D}(3m+1)\mathbf{U} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ e^{-j2\pi/3} & e^{-j4\pi/3} & 1 \\ e^{-j4\pi/3} & e^{-j8\pi/3} & 1 \end{bmatrix}, \tag{20.64}$$
$$\mathbf{D}(3m+2)\mathbf{U} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ e^{-j4\pi/3} & 1 & e^{-j2\pi/3} \\ e^{-j8\pi/3} & 1 & e^{-j4\pi/3} \end{bmatrix},$$

which means that columns of the resulting matrix are shifted for successive symbols.

The case for two antennas is simpler because the precoding matrix $\mathbf{W}(n)$ is always the identity matrix (so no precoder cycling is applied) and

$$\mathbf{D}(n) = \begin{bmatrix} 1 & 0 \\ 0 & (-1)^n \end{bmatrix}. \tag{20.65}$$

This implies that

$$\mathbf{D}(2m)\mathbf{U} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$
$$\mathbf{D}(2m+1)\mathbf{U} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \tag{20.66}$$

which simply represents a swap of the columns between the two layers for consecutive symbols.

3.20.5.1.3 Multiple user MIMO (MU-MIMO)

The previous description is based on Release 8 of the LTE standard and has considered only Single-User MIMO (SU-MIMO). That release includes a rather minimal MU-MIMO transmission scheme. It is based on codebook feedback and uses the same codebooks as SU-MIMO. Actually, only the rank-1 precoders are employed because only one layer is utilized by each UE. The performance of this MU-MIMO scheme is limited by the coarse codebook quantization and the lack of support for cross-talk suppression at the UE. As a consequence, MU-MIMO only offers marginal gain with respect to SU-MIMO. The shortcomings of this simple MU-MIMO approach are fixed in the subsequent releases [80–82]. The set of new features included in Release 10 of the standard has made it possible to reach spectral efficiencies of 30 bits/s/Hz in the downlink and 15 bits/s/Hz in the uplink [83, Section 7.3].

Release 9 allows for beamforming for up to four UEs. The beamformers are constructed by exploiting channel reciprocity. It also includes the option of rank-2 transmissions to two UEs. Release 10 (also known as LTE-A or LTE-Advanced) supports configurations with up to 8 × 8 MIMO with eight transmission layers, and as a consequence the set of precoding codebooks has also been extended using the dual-codebook approach. That is, the precoding matrix is obtained as the multiplication of two matrices, $\mathbf{W}_1$ and $\mathbf{W}_2$, where $\mathbf{W}_1$ is a block diagonal matrix matching the spatial covariance matrix of the dual-polarized antenna setup, and $\mathbf{W}_2$ is the antenna selection and cophasing matrix. The LTE-A UEs have to provide feedback information for both $\mathbf{W}_1$ and $\mathbf{W}_2$. When only two or four antennas are used at the eNodeB, $\mathbf{W}_1$ is the identity matrix and backwards compatibility with Releases 8 and 9 is achieved. For the 8-transmit antenna configuration, $\mathbf{W}_1$ is obtained from the coefficients of the DFT.

An important contribution in LTE-A is the inclusion of Coordinated Multipoint transmission (CoMP), whereby multiple eNodeBs can cooperate to determine the scheduling, transmission parameters, and transmit antenna weights for a specific UE [84,85]. The objective is to reduce interference at the UEs, making universal frequency reuse possible and, hence improving cell-edge throughput as well as average sector throughput with little complexity increase at the receiver. Two major types of CoMP transmission are identified for the downlink (DL) of LTE-A:

- *Coordinated beamforming/coordinated scheduling (CB/CS)* refers to techniques that do not required data sharing between cells. However, CSI may be shared among cells. This family of techniques includes coordinated beamforming/scheduling, adaptive fractional frequency reuse, interference alignment, PMI (Precoding Matrix Indicators) coordinations, etc.
- *Joint processing* is characterized by the fact that data are shared, and it includes techniques such as dynamic cell selection and joint transmission (network MIMO).

The CoMP concept can also be employed in the uplink by coordinating multiple cells to perform joint reception of the transmitted signal at multiple receiving eNodeBs and/or by taking coordinated scheduling decisions. Nevertheless, CoMP transmission/reception is an active area of research and further studies are needed to reliably evaluate the gains of CoMP in LTE.

3.20.5.1.4 *Uplink MIMO*

In Release 8, only one antenna of the UE can be used for transmission, so it is possible to achieve transmit diversity using an antenna selection mechanism, but single-user spatial multiplexing is not feasible. However, the uplink (UL) can support MU-MIMO transparently with 2–6 UEs (although in practice only two UEs are considered in order to limit receiver complexity). The number of UEs that can share a resource block is determined by the number of orthogonal reference signals that can be assigned to the UEs. The different reference signals are used by the eNodeB to estimate the channels of each UE, from which a multiuser detector (e.g., using the MMSE criterion) is derived. In Release 10, spatial multiplexing with 1, 2, or 4 transmit antennas at the UE and up to four layers is introduced. Open-loop and closed-loop spatial multiplexing as well as transmit diversity are supported. Closed-loop multiplexing relies on codebook-based precoding, and the codebooks are optimized to maintain a low PAPR.

To sum up, the evolution of MIMO techniques in the different releases of the standard is summarized in Table 20.1.

3.20.5.2 Multiple antennas techniques in WiMAX

The IEEE 802.16m standard is the core technology for WiMAX Release 2 (WiMAX-2 in short), and it contains the addition of several MIMO technologies to the ones included in IEEE 802.16e (which was the basis of WiMAX Release 1) [86]. MIMO plays an essential role in WiMAX-2, as well as in LTE-A, in order to meet the IMT-Advanced 4G requirements. Although terminologies in the IEEE 802.16 and 3GPP LTE standards differ and the comparison may be confusing, the MIMO techniques used in both WiMAX-2 and LTE-A, while different in various details, share in general the same fundamental approaches.

Table 20.1 Evolution of the Support of MIMO Techniques in LTE

	LTE (Rel-8)	LTE (Rel-9)	LTE-A (Rel-10)
Downlink	• Codebook-based SU- & MU-MIMO • Transmit diversity • Dedicated reference signal-based beamforming	• Dual-stream beamforming	• Non-codebook-based precoding for eight layers • Enhanced MU-MIMO • Inclusion of CoMP
Uplink	• MU-MIMO • Antenna selection		• Spatial multiplexing with codebook-based precoding • Transmit diversity

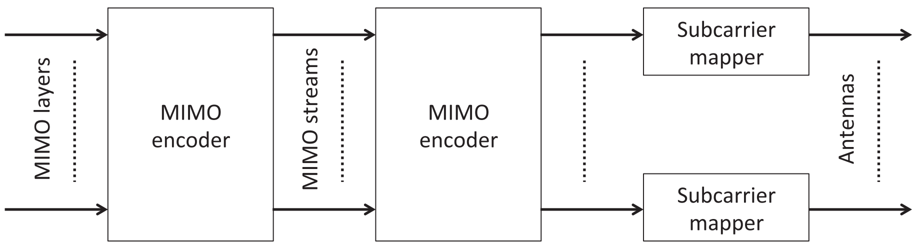


FIGURE 20.30

WiMAX DL MIMO architecture (as shown in [89]).

Therefore, rather than describing the details of MIMO techniques in WiMAX-2, for which an excellent review can be found in [87, Ch.10], we will focus on the similarities between LTE-A and WiMAX (a compared overview can be found in [88]), and comment on some specifics aspects of the latter.

Both 802.16m and LTE-A support MIMO implementations with the same sets of antennas: 2, 4, or 8 transmit antennas and a minimum of 2 receive antennas in the DL; 1, 2, or 4 transmit antennas and a minimum of 2 receive antennas in the UL. The two systems also specify schemes for: open-loop transmit diversity, open- and closed-loop spatial multiplexing, and MU-MIMO both in the UL and DL. The WiMAX downlink architecture is represented in Figure 20.30. For open-loop transmit diversity, WiMAX employs SFBC encoding combined with precoder cycling, whereas LTE employs either SFBC or SFBC-FSTD. As far as open-loop spatial multiplexing is concerned, both systems propose precoder cycling, but WiMAX does not include layer permutation with CDD. Closed-loop spatial multiplexing relies on codebook-based precoding, and WiMAX has three feedback mechanisms: base mode, transformation mode, and differential mode.

Codebook adaptation is defined in 802.16m, and it consists in changing the codeword distribution according to long-term channel statistics. Each vector codeword of the rank-1 base codebook is linearly transformed and normalized to create a codeword in the new codebook. As a result, more codewords are steered towards the ideal beamformer vectors and the codebook quantization error is reduced. Moreover, codebook adaptation is also useful in achieving robustness against calibration errors in the antenna array and transceiver chains. The derivation of the adaptive precoding matrix is specific to the implementation and is not included in the standard. The case where the columns of the precoding matrix are orthogonal to each other is called unitary precoding. Otherwise, it is defined as non-unitary precoding. Non-unitary precoding is only allowed with closed-loop MU-MIMO. Advanced beamforming is also enabled by this precoding mechanism. Besides the closed-loop MU-MIMO scheme, which is also present in LTE, WiMAX-2 allows for open-loop MU-MIMO, where each terminal selects the preferred column from a unitary matrix that has been preset for each frequency-domain resource. Each terminal reports the channel quality indicator (not the spatial correlation matrix, which is the reason why the scheme is considered to be open-loop), and the technique shows good performance with limited feedback in uncorrelated and semi-correlated channels typically corresponding to urban areas with high user density and no line-of-sight. A summary of the MIMO modes proposed for the downlink and uplink of WiMAX-2 are summarized in Tables 20.2 and 20.3.

3.20.5.3 Multiple Antenna Techniques in IEEE 802.11

MIMO techniques play an essential role in the significant increase (54–600 Mbits) in the maximum data rate provided by the IEEE 802.11n amendment to the IEEE 802.11-2007 standard. The single largest contributor to the rate increment comes from the use of multiple antennas, which has the effect of a fourfold increase. A factor of 2 can be attributed to the widening of the channels from 20 MHz to

Table 20.2 Downlink MIMO Modes

Mode index	Description	MIMO Encoding Format	MIMO Precoding
0	Open-loop single-user (transmit diversity)	Alamouti encoding in space-frequency	Non-adaptive
1	Open-loop single-user (spatial multiplexing)	Transparent encoding	Non-adaptive
2	Closed-loop single-user (spatial multiplexing)	Transparent encoding	Adaptive
3	Open-loop multiple-user (spatial multiplexing)	Multi-layer encoding	Non-adaptive
4	Closed-loop multiple-user (spatial multiplexing)	Multi-layer encoding	Adaptive
5	Open-loop single-user (transmit diversity)	Conjugate data repetition	Non-adaptive

Table 20.3 Uplink MIMO Modes

Mode index	Description	MIMO Encoding Format	MIMO Precoding
0	Open-loop single-user (transmit diversity)	Alamouti encoding in space-frequency	Non-adaptive
1	Open-loop single-user (spatial multiplexing)	Transparent encoding	Non-adaptive
2	Closed-loop single-user (spatial multiplexing)	Transparent encoding	Adaptive
3	Open-loop multiple-user (collaborative spatial multiplexing)	Transparent encoding	Non-adaptive
4	Closed-loop multiple-user (collaborative spatial multiplexing)	Transparent encoding	Adaptive

40 MHz; and the rest of the improvement (roughly about 40%) to reducing the overhead in the signal [90]. IEEE 802.11a/g allowed only for a very basic exploitation of multiple antennas. Its method for obtaining diversity was simple antenna selection, while IEEE 802.11n allows for the use of Space-Time Block Codes (STBC), spatial multiplexing and transmit beamforming. Any number of transmit and receive antennas with a maximum number of 4 at each side is permitted, and up to four data streams can be multiplexed.

Two processing blocks are sequentially applied to the spatial streams to obtain the data streams to be transmitted from each antenna [91,92]:

- *STBC encoder:* Spreads constellation points from N_{SS} spatial streams into N_{STS} space-time streams using a space-time block code. The STBC encoder is used only when $N_{SS} < N_{STS}$, otherwise it is a transparent block. If $N_{SS} = 1$ and $N_{STS} = 2$, the Alamouti code is employed; if $N_{SS} = 2$ and $N_{STS} = 3$, one spatial stream is encoded by the Alamouti approach and the other stream is directly mapped to the third space-time stream; if $N_{SS} = 2$ and $N_{STS} = 4$, two disjoint pairs of space-time streams are obtained by applying the Alamouti code to each spatial stream; and finally if $N_{SS} = 3$ and $N_{STS} = 4$, one spatial stream is coded with the Alamouti code and the other two streams are directly mapped to the output. The cases for a single spatial stream $N_{SS} = 1$ with three or four antennas are handled through the use of spatial expansion, which is mentioned below.
- *Spatial mapper:* Maps the N_{STS} space-time streams to the N_{TX} antennas (where $N_{TX} \geq N_{STS}$) by multiplying the space-time streams by a matrix, which is then passed along to each transmit chain. Different matrices can be used for different subcarriers. Some examples of spatial mapping are presented below, but other alternatives are possible and the standard does not restricts the implementation to these instances.
  - *Direct mapping:* Each space-time stream is directly assigned to each antenna (only possible when $N_{TX} = N_{STS}$), possibly after multiplication by a complex exponential in order to implement CDD.

- *Indirect mapping:* The two sets of streams are related by a square unitary matrix such as the Hadamard matrix or the Fourier matrix.
- *Spatial expansion:* The standard proposes several binary-valued (ones and zeros) matrices covering the different combinations of the values of N_{TX} and N_{STS}. The effect of these matrices is simply to translate each of the N_{STS} streams to one or several antennas. For instance, if $N_{\mathrm{STS}} = 1$ and $N_{\mathrm{TX}} = 3$, the following matrix (vector, in this case) is proposed: $\mathbf{D} = 1/\sqrt{3}[1\ 1\ 1]^T$, which implies that the same symbols are transmitted simultaneously from the three antennas.
- *Beamforming matrix:* Represents any matrix that improves the reception based on some knowledge of the channel between the transmitter and the receiver. Two mechanisms are considered in the standard to obtain CSI at the transmit side. The first is called *implicit feedback*, which relies on reciprocity in the TDD operation mode to estimate the channel based on a reference signal transmitted by the device that will act as receiver in the subsequent communication. In the second mechanism, denoted as *explicit feedback*, the receiver sends to the transmitter either the measured channel response or a beamforming matrix that it has computed based on the measured channel. In the latter case, there are two possibilities, namely, to simply transmit the coefficients of the beamforming matrix (called noncompressed beamforming feedback matrix) or a set of angles and phases that parameterize that matrix (called compressed beamforming feedback matrix).

As a final remark, it is worth mentioning that a cyclic shift can also be applied to the signal in each antenna to prevent unintentional beamforming. The shift can be inserted either in the frequency or in the time domain (i.e., before or after the IDFT).

3.20.6 Biomedical

There is an immense interest in processing the electrical, magnetic and acoustic signals that originate from physiological processes, and extracting information that is useful for diagnosis and treatment. Examples of such signals include those obtained via electrocardiography, measurements of the electrical behavior of the heart; electroencephalography and magnetoencephalography, which measure the electrical and magnetic activity of the brain; electromyography, observations of electrical signals in muscle tissue, and so on. In addition, active measurement approaches exist that collect the response of the body to magnetic or acoustic stimulation, as in magnetic resonance or ultrasonic imaging systems. Arrays of sensors are used in many of these applications, primarily for localizing the source of the signals in passive measurement systems, or for non-invasively imaging the internal structure of the body in active systems.

In this section, we will briefly discuss three biomedical applications of array signal processing that are widely employed in both clinical or research settings. These are by no means exhaustive; a notable omission is magnetic resonance imaging (MRI), which uses a large array of coils to detect the precession of molecules in response to applied external magnetic fields. However, these examples serve to illustrate the important role array signal processing has in biomedicine.

3.20.6.1 Ultrasonic imaging

Ultrasonic arrays for biomedical imaging are in widespread clinical use today, most commonly for monitoring fetal development and for real-time imaging of heart valve operation and related blood flow. Ultrasound imaging is relatively inexpensive compared with other imaging modalities, and the array and associated equipment is relatively compact and portable. Ultrasonic images can achieve sub-millimeter resolution, but the imaging process is more susceptible to noise and unpredictable propagation effects than, say, MRI.

Ultrasound imaging is based on pulse-echo signal processing, much like an active radar. An array of from many tens to a few hundred piezoelectric transducers transmits baseband signals with bandwidths up to tens of MHz and then receives the resulting echoes. Broadband signals are usually employed for imaging human tissue, while narrowband CW signals are employed for Doppler measurements of blood flow velocities. An ultrasonic array is relatively compact, with an aperture of 50 mm or less, and can be condensed to fit in a handheld wand that is manually placed on the body and oriented in some direction of interest. The speed of sound in human tissue is approximately 1500 m/s, so the resulting wavelength is typically much less than 1 mm. Consequently, near-field modeling of the acoustic wavefronts is often necessary. The array typically has a slight inward curve to create a larger "fan-beam" image.

Traditionally, ultrasonic imagers have employed delay-and-sum beamforming to focus both the transmit and receive signals, although with improvements in computational power, systems are now being designed with adaptive (e.g., MVDR) beamforming to improve resolution and eliminate artifacts due to interference entering through sidelobes. An important difference compared to radar is the severe range-dependent attenuation the ultrasonic signal undergoes—with signal intensity decreasing by a factor of two at 5 MHz for approximately every cm of distance the signal travels. This necessitates the use of gain compensation on receive and leads to low SNRs at longer ranges. As focus moves towards higher frequencies for better resolution, the attenuation problem increases.

3.20.6.2 EEG and MEG signal processing

Electroencephalography (EEG) and magnetoencephalography (MEG) are widely used in both clinical practice and research since they provide direct measurement of cerebral activity with much higher temporal resolution than other non-invasive methods such as functional MRI (fMRI). The analysis of EEG/MEG signals is used for detecting and diagnosing neurological disorders such as epileptic seizures, monitoring brain activity during sleep or anesthesia, analyzing the extent of brain damage due to stroke or traumatic injury, etc. Such signals are also currently being investigated as a tool for brain-computer interface applications that would allow individuals with sensory-motor impairments (e.g., a paraplegic) to control a wheelchair, prosthetic limb or a computer input device via focused cognitive activity. Although EEG/MEG techniques have lower spatial resolution than, for example, fMRI, relatively high-resolution techniques for locating sources of cerebral activity have been proposed to cope with this issue.

To obtain high spatial precision, EEG/MEG localization requires a large array of sensors or electrodes (an example of a typical EEG array is shown in Figure 20.31), which leads to a high-dimensional inverse problem that in general does not have a unique solution. Thus, in practice, a "forward" propagation model for the brain, skull and scalp is adopted, and one attempts to estimate the parameters of the model corresponding to the source activity. A common approach is to model the signal source in a

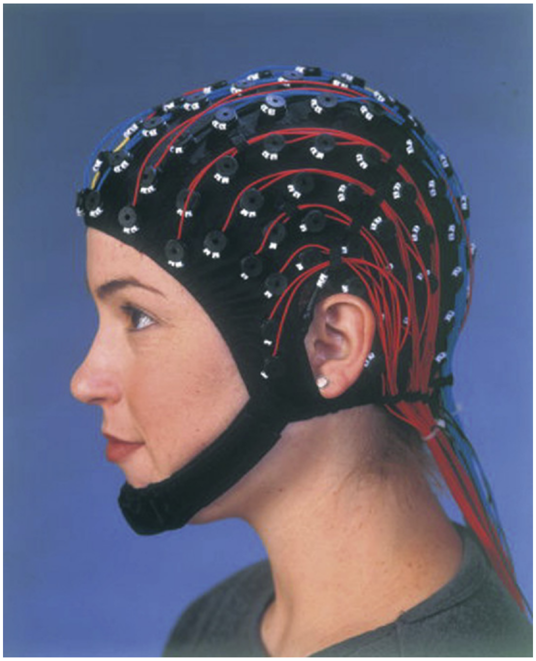


FIGURE 20.31

Cap with electrodes for collecting EEG data.

small region of the brain as originating from an equivalent current dipole, treating the dipole location, the orientation and magnitude of the dipole moment as dipole parameters to be estimated. Figure 20.32 depicts the equivalent dipole model, where the charge difference along a neuron (or neuron cluster) causes a flow of current whose resulting electric or magnetic field can be measured by a sensor. While EEG caps like the one in Figure 20.31 are used to place the sensor as close to the brain as possible, there is increasing interest in and use of intracranial EEG measurements, or electrocorticography (ECoG), where the electrode is placed beneath the skull immediately adjacent to the neural area of interest. ECoG signals avoid the attenuation of the skull and scalp and provide a much higher SNR, but at the expensive of an invasive implantation. In the next section, we describe the mathematical model that results from the equivalent dipole assumption to illustrate its connection with other array signal processing applications.

3.20.6.2.1 ***Unified dipole model for EEG/MEG/ECoG measurements***

Assume a current dipole located at position $\mathbf{r} \in \mathbb{R}^{3\times 1}$ and an electric or magnetic sensor located at $\mathbf{s} \in \mathbb{R}^{3\times 1}$. In the following, we derive expressions for the electric or magnetic field $y(t)$ at $\mathbf{s}$ due to the dipole at $\mathbf{r}$ for the three measurement modalities EEG, MEG, and ECoG. We will see that all three lead to expressions with a similar structure that allows us to formulate a unifying model for the output of arrays of such sensors.

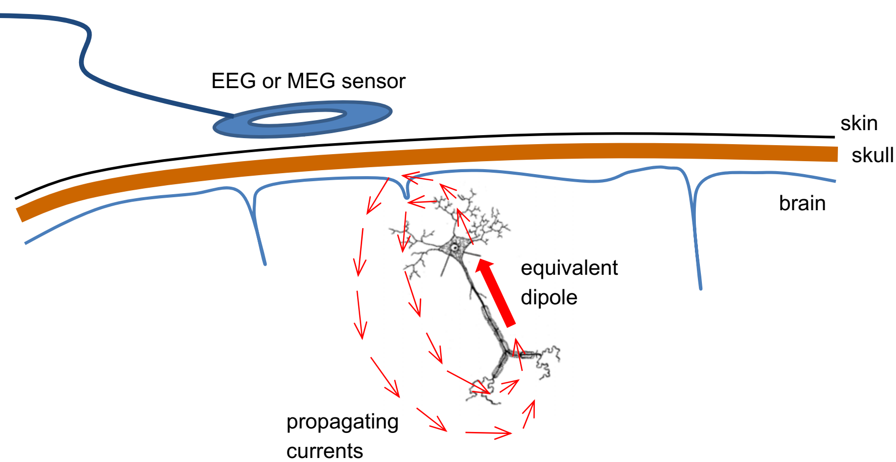


FIGURE 20.32

Equivalent current dipole model for EEG measurements.

EEG: The electric potential at $\mathbf{s}$ caused by a current density $\mathbf{J}_S$ at location $\mathbf{r}$ under a quasistatic assumption (i.e., setting all time derivatives in Maxwell's equations equal to zero) can be expressed as

$$y(t) = -\frac{1}{4\pi\sigma}\int_V \frac{\nabla \mathbf{J}_S(\mathbf{r}, t)}{\|\mathbf{s}-\mathbf{r}\|}\, d\mathbf{r}, \tag{20.67}$$

where ∇ is the divergence operator and V is the volume of interest. A current dipole can be idealized as a source and sink with equal magnitude, denoted by $I_0(t)$, and separated by a very small distance d, which leads to

$$\nabla \mathbf{J}_S = -I_0(t)[\delta(\mathbf{s}-\mathbf{r}_+) - \delta(\mathbf{s}-\mathbf{r}_-)], \tag{20.68}$$

where δ is the Dirac delta function, and $\mathbf{r}_+$ ($\mathbf{r}_-$) is the source (sink) location. The dipole location $\mathbf{r}$ is assumed to be at the midpoint between $\mathbf{r}_+$ and $\mathbf{r}_-$, and the orientation of the dipole is in the direction of $\mathbf{d} = \mathbf{r}_+ - \mathbf{r}_- = \frac{1}{2}(\mathbf{r}_+ - \mathbf{r}) = \frac{1}{2}(\mathbf{r} - \mathbf{r}_-)$. Substituting (20.68) into (20.67), the potential generated by the ideal dipole source becomes

$$y(t) = \frac{I_0(t)}{4\pi\sigma\|\mathbf{s}-\mathbf{r}_+\|} - \frac{I_0(t)}{4\pi\sigma\|\mathbf{s}-\mathbf{r}_-\|}. \tag{20.69}$$

Assuming that $\|\mathbf{s}-\mathbf{r}_+\| \gg d$, and similarly for $\mathbf{r}_-$, then

$$\frac{1}{\|\mathbf{s}-\mathbf{r}_+\|} \approx \frac{1}{\|\mathbf{s}-\mathbf{r}\|} + \frac{(\mathbf{s}-\mathbf{r})^T\mathbf{d}}{2\|\mathbf{s}-\mathbf{r}\|^3},$$
$$\frac{1}{\|\mathbf{s}-\mathbf{r}_-\|} \approx \frac{1}{\|\mathbf{s}-\mathbf{r}\|} - \frac{(\mathbf{s}-\mathbf{r})^T\mathbf{d}}{2\|\mathbf{s}-\mathbf{r}\|^3}.$$

Substituting this approximation into (20.69), the potential received at $\mathbf{r}$ becomes

$$y(t) = \frac{I_0(t)}{4\pi\sigma}\left[\frac{1}{\|\mathbf{s}-\mathbf{r}\|} + \frac{(\mathbf{s}-\mathbf{r})^T\mathbf{d}}{2\|\mathbf{s}-\mathbf{r}\|^3} - \left(\frac{1}{\|\mathbf{s}-\mathbf{r}\|} - \frac{(\mathbf{s}-\mathbf{r})^T\mathbf{d}}{2\|\mathbf{s}-\mathbf{r}\|^3}\right)\right] = \frac{1}{4\pi\sigma}\frac{(\mathbf{s}-\mathbf{r})^T}{\|\mathbf{s}-\mathbf{r}\|^3}\mathbf{m}(t), \quad (20.70)$$

where the dipole moment is defined as $\mathbf{m}(t) = \mathbf{d}I_0(t)$. In the sequel, we will write $\mathbf{m}(t) = \boldsymbol{\phi}s(t)$, where $\boldsymbol{\phi} = \mathbf{d}/\|\mathbf{d}\|$ is the unit-magnitude dipole orientation, and $s(t) = I_0(t)\|\mathbf{d}\|$ is the moment magnitude.

MEG: Extracranial magnetic fields produced by neuronal activity within the brain can be calculated using Biot-Savart's law. A dipole source at $\mathbf{r}$ with dipole moment $\mathbf{m}(t)$ will generate a magnetic field $y(t)$ at sensor location $\mathbf{s}$ given by

$$y(t) = \frac{\mu_0(\mathbf{m}(t)\times(\mathbf{s}-\mathbf{r}))^T}{4\pi\|\mathbf{s}-\mathbf{r}\|^3}\mathbf{t} = \frac{\mu_0((\mathbf{s}-\mathbf{r})\times\mathbf{t})^T}{4\pi\|\mathbf{s}-\mathbf{r}\|^3}\mathbf{m}(t), \quad (20.71)$$

where $\times$ denotes the vector cross product and $\mathbf{t}$ is a unit vector defining the orientation of the sensor. As in the case of EEG, we will write $\mathbf{m}(t) = \boldsymbol{\phi}s(t)$ with $\boldsymbol{\phi}$ defining the orientation of the dipole moment. Note that a dipole inside a spherically symmetric conductor with $\boldsymbol{\phi}$ aligned with the sphere's radius will produce no external magnetic field. Consequently, for MEG applications, the orientation $\boldsymbol{\phi}$ (and hence the moment $\mathbf{m}(t)$) is often expressed using only two rather than three coordinates.

ECoG: More involved models have been developed for ECoG settings due to the presence of local currents and higher SNR. For a sensor inside the skull on the surface of the brain at position $\mathbf{s}$, the measured potential $y(t)$ due to a dipole source at $\mathbf{r}$ with moment $\mathbf{m}(t)$ in a homogeneous conducting sphere is given by

$$y(t) = \frac{1}{4\pi\sigma}\left(2\frac{\mathbf{r}-\mathbf{s}}{r_{d^3}} + \frac{1}{\|\mathbf{r}\|^2 r_d}\left[\mathbf{r} + \frac{\mathbf{r}\|\mathbf{s}\|\cos\theta - \|\mathbf{r}\|\mathbf{s}}{\|\mathbf{r}\| + r_d - \|\mathbf{s}\|\cos\theta}\right]\right)^T \mathbf{m}(t), \quad (20.72)$$

where σ is the conductivity value for the brain, θ denotes the angle between $\mathbf{r}$ and $\mathbf{s}$ and $r_d = \|\mathbf{r}-\mathbf{s}\|$.

General multi-source multi-sensor model: In all three cases discussed above, the equation for the electric or magnetic field has the same general form. In particular, for an array of M sensors at positions $\mathbf{s}_i$, $i = 1, \ldots, M$, the field can be represented as

$$y_i(t) = \mathbf{g}(\mathbf{s}_i, \mathbf{r})^T\mathbf{m}(t), \quad (20.73)$$

where the gain vector depends on which measurement system is employed:

$$\text{EEG:}\quad \mathbf{g}(\mathbf{s}_i, \mathbf{r}) = \frac{1}{4\pi\sigma}\frac{\mathbf{s}_i-\mathbf{r}}{\|\mathbf{s}_i-\mathbf{r}\|^3}, \quad (20.74)$$

$$\text{MEG:}\quad \mathbf{g}(\mathbf{s}_i, \mathbf{r}) = \frac{\mu_0(\mathbf{s}_i-\mathbf{r})\times\mathbf{t}_i}{4\pi\|\mathbf{s}_i-\mathbf{r}\|^3}, \quad (20.75)$$

$$\text{ECoG:}\quad \mathbf{g}(\mathbf{s}_i, \mathbf{r}) = \frac{1}{4\pi\sigma}\left(2\frac{\mathbf{r}-\mathbf{s_i}}{r_{d,i}^3} + \frac{1}{\|\mathbf{r}\|^2 r_{d,i}}\left[\mathbf{r} + \frac{\mathbf{r}\|\mathbf{s}_i\|\cos\theta_i - \|\mathbf{r}\|\mathbf{s_i}}{\|\mathbf{r}\| + r_{d,i} - \|\mathbf{s}_i\|\cos\theta_i}\right]\right), \quad (20.76)$$

where all variables are as defined above, with the subscript i referencing sensor i. Thus, for all three models, stacking the outputs of the M sensors together in the vector $\mathbf{y}(t)$ yields the same general equation:

$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{g}^T(\mathbf{s}_1, \mathbf{r}) \\ \vdots \\ \mathbf{g}^T(\mathbf{s}_M, \mathbf{r}) \end{bmatrix} \mathbf{m}(t) = \mathbf{G}(\mathbf{r})\mathbf{m}(t) = \mathbf{a}(\mathbf{r}, \boldsymbol{\phi})s(t), \tag{20.77}$$

where $\mathbf{G}(\mathbf{r})$ is $M \times 3$ (or possibly $M \times 2$ in the case of MEG data), and where the steering vector for the source depends on its location and dipole orientation:

$$\mathbf{a}(\mathbf{r}, \boldsymbol{\phi}) = \mathbf{G}(\mathbf{r})\boldsymbol{\phi}. \tag{20.78}$$

The steering vector model in (20.78) has the same form as in RF applications with diversely polarized signals. Finally, augmenting the model with the superposition of N sources as well as background interference $\mathbf{n}(t)$, we end up with the standard array processing equation:

$$\mathbf{y}(t) = \left[\mathbf{a}(\mathbf{r}_1, \boldsymbol{\phi}_1) \ \cdots \ \mathbf{a}(\mathbf{r}_N, \boldsymbol{\phi}_N)\right] \begin{bmatrix} s_1(t) \\ \vdots \\ s_N(t) \end{bmatrix} + \mathbf{n}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \tag{20.79}$$

where the vector $\boldsymbol{\theta}$ contains the source location and dipole orientation parameters.

The fact that the steering or array manifold vectors depend linearly on the dipole orientations, and that these vectors are assumed to satisfy $\boldsymbol{\phi}_k^T\boldsymbol{\phi}_k = 1$, lead to special types of solutions when estimating the parameters. For example, a direct implementation of the MUSIC algorithm leads to

$$\hat{\mathbf{r}}, \hat{\boldsymbol{\phi}} = \arg\min_{\mathbf{r},\boldsymbol{\phi}} \frac{\boldsymbol{\phi}^T\mathbf{G}^T(\mathbf{r})\mathbf{E}_n\mathbf{E}_n^{\mathbf{T}}\mathbf{G}(\mathbf{r})\boldsymbol{\phi}}{\boldsymbol{\phi}^T\mathbf{G}^{\mathbf{T}}(\mathbf{r})\mathbf{G}(\mathbf{r})\boldsymbol{\phi}} \quad \text{s.t.} \quad \boldsymbol{\phi}^T\boldsymbol{\phi} = 1, \tag{20.80}$$

where $\mathbf{E}_n$ are the noise subspace eigenvectors of the covariance of $\mathbf{y}(t)$. It is straightforward to show that minimizing the MUSIC criterion is equivalent to solving the following generalized eigenvalue problem as a function of $\mathbf{r}$:

$$\hat{\mathbf{r}} = \arg\min_{\mathbf{r}} \lambda_{\min}(\mathbf{r}), \tag{20.81}$$

$$\mathbf{G}^T(\hat{\mathbf{r}})\mathbf{E}_n\mathbf{E}_n^T\mathbf{G}(\hat{\mathbf{r}})\hat{\boldsymbol{\phi}} = \lambda_{\min}(\hat{\mathbf{r}})\mathbf{G}^T(\hat{\mathbf{r}})\mathbf{G}(\hat{\mathbf{r}})\hat{\boldsymbol{\phi}}, \tag{20.82}$$

where $\lambda_{\min}(\mathbf{r})$ is the smallest generalized eigenvalue for a given $\mathbf{r}$. The position estimates $\hat{\mathbf{r}}$ are found by searching for the value of $\mathbf{r}$ for which $\lambda_{\min}(\mathbf{r})$ is minimized, and the dipole orientation estimate is then given by the generalized eigenvector associated with $\lambda_{\min}(\hat{\mathbf{r}})$.

3.20.6.2.2 Interference mitigation

For EEG and MEG measurements, where the sensors are separated from the brain by the skull and scalp, the signals of interest are very weak, and embedded in strong, spatially correlated noise and interference due primarily to background brain activity not related to the stimulus of interest. If standard source localization algorithms are applied without some attempt at mitigating this interference, the results are typically very poor. A common strategy in such situations is to design experiments with dual conditions, one (control state) prior to application of the stimulus and one (activity state) after the stimulus has been applied. In principle, the control state data will contain only background interference and sensor noise, while the activity state data will contain statistically similar noise and interference as well as the event-related signals. Prewhitening approaches are typically applied in dual-condition experiments like these. In these approaches, the control state data are first used to estimate the spatial covariance matrix of the interference plus noise using, for example, the following sample average:

$$\widehat{\mathbf{R}}_C = \frac{1}{n_C}\sum_{t=1}^{n_C}\mathbf{y}_C(t)\mathbf{y}_C^T(t), \tag{20.83}$$

where n_C is the number of control state samples. The activity state data, $\mathbf{y}_A(t)$, is then prewhitened in an attempt to eliminate the influence of the interference and noise as follows:

$$\mathbf{y}'_A(t) = \widehat{\mathbf{R}}_C^{-1/2}\mathbf{y}_A(t). \tag{20.84}$$

A drawback to the use of prewhitening is that it requires that the spatial and temporal statistics of the interference and noise during the control state be identical to those during the activity state. If the assumption of stationarity between these two states is violated, then methods based on prewhitening can suffer a significant performance degradation. An alternative is to use projection-based methods that estimate a spatial-only subspace in which the bulk of the interference energy lies during the control state, and then project away this subspace in the activity state data. This method eliminates the need for temporal stationarity, and relies only on the assumption that the locations of the interference in the control and activity states remain unchanged. This is a reasonable assumption since any "new" source that appears during the activity state is considered to be related to the stimulus, and is thus a source of interest.

As an example, we present here the results of an experimental study with real EEG data. Experiments with an auditory stimulus applied to the left ear were conducted with a single human subject to elicit auditory-evoked potentials. MUSIC and LCMV were applied to the resulting data using both prewhitening (PW) and the projection (NP) technique and assuming the number of sources was one. Figures 20.33 and 20.34 show the spatial spectra of the four algorithm combinations. The projection-based methods provide an activity map that closely corresponds to what one would expect, with energy confined to the auditory cortex. On the other hand, the prewhitening-based methods contain a number of apparently unrelated artifacts, and the PW-LCMV method does not even show any energy near the auditory cortex.

3.20.6.3 Multi-sensor extracellular probes

Direct measurement of neural action potentials (APs) using electrodes inserted directly into biological tissue has become an important neurological research and diagnostic tool. The goal is to record the APs of individual neurons, often referred to as "single-unit activity," in order to obtain a more precise view of the underlying neurophysiology. Information from these recordings are potentially useful in the development of artificial prostheses and in the diagnosis and treatment of paralysis and brain disorders such as epilepsy and memory loss. Even though the electrodes are small and can be inserted with high accuracy to target a specific location, they will typically record the superposition of the activity from

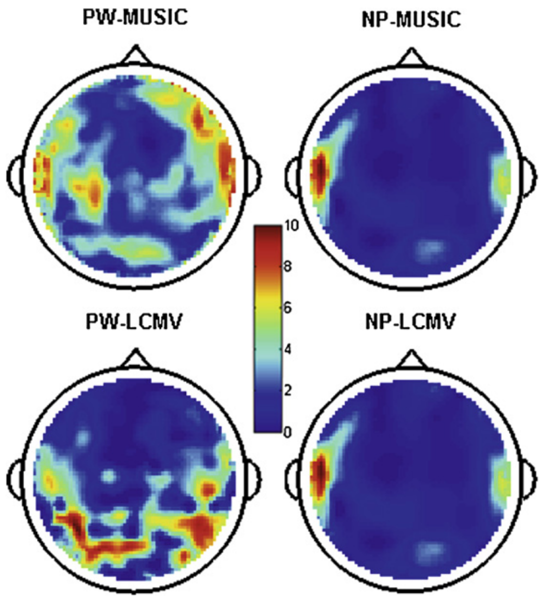


FIGURE 20.33

Spatial spectra of four algorithm combinations using experimental data (top view).

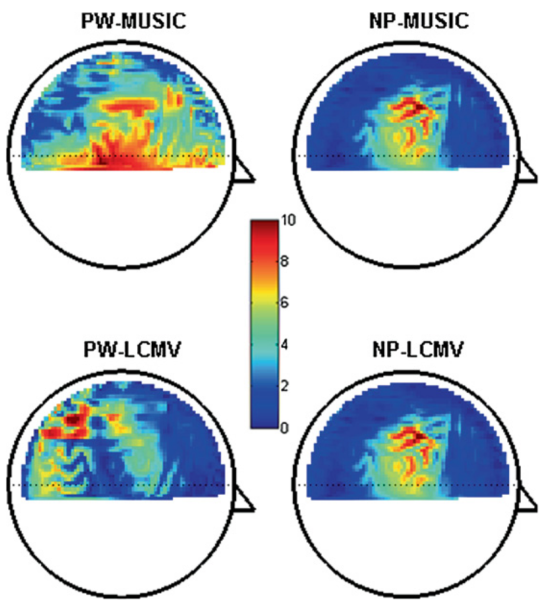


FIGURE 20.34

Same as previous figure (side view).

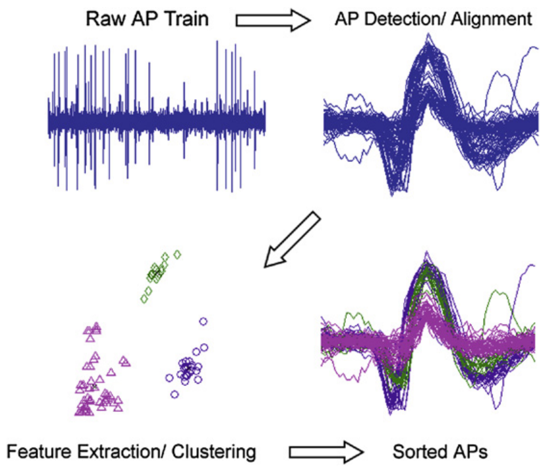


FIGURE 20.35

Automated procedures for AP sorting.

several neurons. The process of separating out the single-unit activity of individual neurons from the multi-unit activity in the noisy electrode measurements is often referred to as AP or "spike" sorting.

In practice, manual sorting of APs in large volumes of experimental data is prohibitively time-consuming, and automated procedures for AP sorting have become essential. As depicted in Figure 20.35, an automated AP sorting algorithm can be divided into three main steps: (1) *AP detection and time alignment*: determining the locations of the APs in the electrode time series and arranging the isolated AP waveforms so that they "line up" in time, (2) *feature extraction*: extracting a low-dimensional set of parameters for each detected AP that can be used to discriminate between different sources, and (3) *clustering*: grouping the extracted features into clusters in order to associate them with individual neurons. The feature extraction step is crucial since it reduces the effect of noise and removes redundant information in the input data so that clustering algorithms can work efficiently. The three most common feature categories discussed in the literature are: (1) AP shape-related features, such as AP height, width, peak-to-peak amplitude, inter-AP interval, and first-order derivative, (2) wavelet coefficients, and (3) principal components (PCs). One common characteristic of these features is that they only capture "temporal" information since they are obtained by processing single-sensor measurements. However, AP sorting based only on temporal features is challenging since neurons with similar geometries located at roughly equal distances to the electrode can generate very similar AP waveforms and therefore similar features.

To overcome this problem, multi-sensor extracellular probes (e.g., tetrodes) that record a time-aligned multi-channel data set have been suggested. The simplest way to use the data from multi-sensor probes is to apply standard feature extraction techniques to all of the channels individually, and then combine all the extracted features as inputs for clustering. Other approaches use the availability of spatially distinct channel measurements to obtain neuron location estimates or independent components as feature vectors for clustering. Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents. While ICA can potentially resolve overlapping spikes, it requires strong assumptions regarding the non-Gaussianity and independence of the APs, and a separate feature extraction step is still required to identify the source of the recovered AP waveform. Matched subspace (MS) techniques attempt to detect the presence of a signal that lies in an *a priori* unknown low-dimensional subspace of the data. Unlike multi-sensor principal component analysis and algorithms based on location estimates, which respectively allow only temporal or spatial information to be extracted, the MS approach provides a joint spatio-temporal feature vector that is more effective for differentiating between individual neurons. Furthermore, the spatial information obtained by MS techniques is achieved without the need for a forward propagation model as required by location-based methods.

3.20.6.3.1 Data model

Assume that the APs have been accurately detected in a previous step using existing approaches. A block of samples around each detected AP peak is isolated, and it is time-aligned with data blocks obtained for other detected APs. Assuming M electrodes and N samples per block, the data for the ith detected AP will consist of an $M \times N$ matrix $\mathbf{Y}_i$, which is referred to as an AP "bundle." Assuming that each bundle consists of an AP from a single neuron, and assuming that the AP signal results in an instantaneous mixture at the electrode array (i.e., a rank-one signal component), an appropriate mathematical model for $\mathbf{Y}_i$ is:

$$\mathbf{Y}_i = \mathbf{S}_i + \mathbf{W}_i = \mathbf{a}_i \mathbf{v}_i^T + \mathbf{W}_i, \tag{20.85}$$

where $\mathbf{S}_i \in \mathbb{R}^{M \times N}$ represents the noise-free multi-sensor signal corresponding to the AP, $\mathbf{W}_i$ is composed of zero-mean background neural and sensor noise, $\mathbf{a}_i \in \mathbb{R}^{M \times 1}$ is the spatial signature of the target neuron, and $\mathbf{v}_i \in \mathbb{R}^{N \times 1}$ corresponds to the sampled AP waveform.

By vectorizing the data matrix, we obtain

$$\begin{aligned}
\mathbf{y}_i &= \mathbf{s}_i + \mathbf{w}_i = \mathbf{v}_i \otimes \mathbf{a}_i + \mathbf{w}_i \\
&= \mathbf{\Phi}\mathbf{c}_i \otimes \mathbf{a}_i + \mathbf{w}_i && (20.86)\\
&= (\mathbf{\Phi} \otimes \mathbf{a}_i)\mathbf{c}_i + \mathbf{w}_i, && (20.87)
\end{aligned}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{y}_i$, $\mathbf{s}_i$, and $\mathbf{w}_i$ are $MN \times 1$ vectors formed from $\mathbf{Y}_i$, $\mathbf{S}_i$, and $\mathbf{W}_i$, respectively. The term $\mathbf{v}_i = \mathbf{\Phi}\mathbf{c}_i$ models the AP in the absence of any specific information about the waveform, with matrix $\mathbf{\Phi} \in \mathbb{R}^{N \times p}$ $(p \leq N)$ representing a chosen orthonormal basis and $\mathbf{c} \in \mathbb{R}^{p \times 1}$ representing the corresponding coefficient vector. Modeling the AP signal in this way not only provides the possibility of a compact representation for the AP but also eliminates the need for AP templates, which enables unsupervised spike sorting. Although $\mathbf{\Phi}$ can be any orthonormal basis, one with a compact support such as the wavelet basis is preferred in general since APs tend to be pulse-like.

3.20.6.3.2 Multi-sensor feature extraction

In the discussion that follows, we briefly describe several popular algorithms for feature extraction from a given $MN \times 1$ AP bundle $\mathbf{y}_i$. In some cases, the methods either assume a single sensor ($M = 1$) or they operate on each sensor independently. We will use the notation $\mathbf{y}_i^{(k)} = \mathbf{y}_i(k : M : M(N-1) + k)$ to represent the data from the kth sensor, where the indexing $k : M : M(N-1) + k$ indicates we select every Mth sample from $\mathbf{y}_i$ starting with sample k. The variable p will be used to denote the dimension of the extracted feature vector.

Discrete wavelet transform: The wavelet transform is a popular choice for feature extraction in the spike sorting application since it offers simultaneous interpretation of the signal in both time and scale (frequency), which allows local, transient or intermittent components to be elucidated. It has advantages over the traditional Fourier transform in analyzing physical signals since it can provide a compact signal representation in both the time and scale domains. The discrete wavelet transform (DWT) decomposes the data from a single sensor as follows:

$$\mathbf{y}_i^{(k)} = \mathbf{\Phi}_w \mathbf{c}_i^{(k)}, \tag{20.88}$$

where $\mathbf{\Phi}_w \in \mathbb{R}^{N \times N}$ is a basis matrix that defines the DWT, and $\mathbf{c}_i^{(k)} \in \mathbb{R}^{N \times 1}$ represents the DWT coefficient vector. The DWT basis is typically assumed to be orthonormal, so the coefficient vector is found by simply computing $\mathbf{c}_i^{(k)} = \mathbf{\Phi}_w^T \mathbf{y}_i^{(k)}$. The feature vector $\hat{\mathbf{c}}_i^{(k)} \in \mathbb{R}^{p \times 1}$ is determined by choosing a subset of p of the coefficients in the full DWT vector $\mathbf{c}_i^{(k)}$. The choice of which p coefficients to use can in principle be different for each sensor k, but must be the same for each AP bundle. For example, feature reduction for the DWT can be achieved by selecting the p coefficients that have the largest average magnitudes. Once a reduced-dimension set of features is chosen for each sensor, the complete feature vector is formed by stacking them all together:

$$\hat{\mathbf{c}}_w = \mathrm{vec}\left(\left[\hat{\mathbf{c}}_w^{(1)} \ \cdots \ \hat{\mathbf{c}}_w^{(m)}\right]\right). \tag{20.89}$$

Principal component analysis: A difficulty associated with the DWT approach is there is no systematic way to choose the wavelet basis so that it is somehow optimized for the signals at hand. Principal component analysis (PCA) addresses this issue by calculating a data-dependent basis that corresponds to the principle subspace where most of the signal energy resides. This is most commonly achieved by performing the singular value decomposition (SVD) on a subset of $n > p$ of the AP bundles

$$\mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^T = \left[\mathbf{y}_{i_1}^{(k)} \mathbf{y}_{i_2}^{(k)} \ \ldots \ \mathbf{y}_{i_n}^{(k)}\right], \tag{20.90}$$

where $i_1, i_2, \ldots, i_n$ are the indices corresponding to the n AP bundles chosen for the analysis. The PCA basis $\mathbf{\Phi}_p^{(k)} \in \mathbb{R}^{N \times p}$ is then taken to be the first p columns of $\mathbf{U}_k$, and the PCA feature vector (sometimes referred to as the "score" vector) is calculated as $\hat{\mathbf{c}}_i^{(k)} = \mathbf{\Phi}_p^{(k)T} \mathbf{y}_i^{(k)}$. In most applications of PCA to this problem, p is chosen to be between two to three. Alternatively, a single basis for all k can be found by including AP bundles from all sensors in the SVD of (20.90). As in the DWT approach, once features are extracted for each k, the complete feature vector is found by stacking them together as in (20.89).

Matched Subspace Detector: The Matched Subspace Detector (MSD) can be thought of as a generalization of the well-known matched filter from signal processing, where a noisy signal $\mathbf{y}_i$ is correlated with a parameterized version of the signal of interest $\mathbf{s}$ to produce the output $\mathbf{s}^T \mathbf{y}_i$. The parameters are chosen as those that maximize the resulting correlation. The single-sensor versions of the DWT and PCA algorithms described in the previous section, where $\mathbf{s} = \mathbf{\Phi c}$, can be thought of as implementing a simple matched filter:

$$\begin{aligned} \hat{\mathbf{c}}_i &= \arg\max_{\mathbf{c}} \|\mathbf{s}^T \mathbf{y}_i\|^2 = \arg\max_{\mathbf{c}} \|\mathbf{c}^T \mathbf{\Phi}^T \mathbf{y}_i\|^2 \\ &\text{s.t. } \|\mathbf{c}\| = \|\mathbf{y}_i\| = \alpha_i, \end{aligned} \tag{20.91}$$

where the constraint on $\mathbf{c}$ is used to maintain energy equivalence. The solution to (20.91) is the same as that given earlier for DWT and PCA: $\hat{\mathbf{c}}_i = \mathbf{\Phi}^T \mathbf{y}_i$.

The general MSD approach can be viewed as a natural multi-sensor extension of the single-sensor DWT or PCA approaches. Instead of the single-sensor parameterization $\mathbf{s} = \mathbf{\Phi c}$, the multi-sensor parameterization in (20.87) is used. In particular, MSD solves the following generalized version of (20.91):

$$\hat{\mathbf{a}}_i, \hat{\mathbf{c}}_i = \arg\max_{\mathbf{a},\mathbf{c}} \left\| \mathbf{s}^T \mathbf{y}_i \right\|^2 = \arg\max_{\mathbf{a},\mathbf{c}} \left\| (\mathbf{\Phi c} \otimes \mathbf{a})^T \mathbf{y}_i \right\|^2$$
$$\text{s.t. } \|\mathbf{a}\| = 1, \; \|\mathbf{c}\| = \|\mathbf{y}_i\| = \alpha_i, \tag{20.92}$$

where the constraints match those used in the model to ensure identifiability. It is straightforward to find a closed form solution for both $\hat{\mathbf{a}}_i, \hat{\mathbf{c}}_i$.

The MSD algorithm can be used in conjunction with either the DWT or PCA, or any other choice of the temporal basis matrix $\mathbf{\Phi}$. The number of features produced by the MSD algorithm will be the M spatial features from the elements of $\hat{\mathbf{a}}_i$, plus however many temporal features are provided by $\hat{\mathbf{c}}_i$, which in turn depends on the dimension of $\mathbf{\Phi}$. For the case of PCA, where $\mathbf{\Phi} \in \mathbb{R}^{N \times p}$ and typically $p \ll N$, the temporal feature vector $\hat{\mathbf{c}}_i$ will have p elements. For the DWT, where $\mathbf{\Phi}$ is $N \times N$, $\hat{\mathbf{c}}_i$ will have N elements. Note that in this case, $\hat{\mathbf{a}}_i$ can be found from the SVD of $\mathbf{X}'_i$ rather than $\mathbf{X}'_i\mathbf{\Phi}$, since both matrices have the same set of left singular vectors. Whether the total number of space-time features obtained by MSD is $M + N$ or $M + P$, it is often desirable to reduce the number of features to a more manageable number.

3.20.7 Sonar

In the context of naval warfare, sonar is used to detect, locate, track and identify surface and submerged vehicles. This is one of the classical early applications of digital array signal processing. Since required bandwidths are often small and operating frequencies are low (10s of Hz to about 30 kHz) due to propagation limitations at higher frequencies in the ocean environment, corresponding sample and data rates are low enough to have been accommodated by ADCs and signal processing computers available in the 1970s and 1980s. The well funded military applications spurred rapid development in that era, and the transition from analog to digital systems enabled significant capability enhancements and increased processing complexity. Many of the classical array processing and statistically optimal beamforming algorithms were first demonstrated in sonar applications.

We will address two major classes of sonar: active and passive. Active sonar has much in common with radar systems in that a pulse, or sequence of pulses, is transmitted and the return echo signal is analyzed to detect vehicle range, direction, and range rate (radial velocity). This can be viewed as an non-cooperative digital wireless communications problem where the transmitted pulse corresponds to communications symbol, and two-way propagation effects including reflection from the target correspond to the communications channel.

Passive sonar is a “listen only” mode used when stealthy operation is important so as not to reveal one’s own position with a transmitted pulse. Acoustic radiation is detected from the target’s turning propeller, internal machinery, occupants, or flow turbulence as it moves through the water. Passive systems can typically estimate target direction, and can classify the source as to speed, vehicle type, or even specific hull number by comparing the signal spectrum to previously obtained acoustic signature data bases.

The primary use of sensor and transmitting arrays in sonar is to exploit spatial information in the channel, including estimating directions of arrival, improving gain, and mitigating against noise and other interfering sources. We will also discuss in Section 3.20.7.3 how array processing combined with good acoustic propagation models can be used in passive sonar to estimate range and depth at a distance, without access to two-way propagation time-of-flight information.

3.20.7.1 Sonar arrays

The sensors used in sonar arrays are hydrophones that act as underwater microphones and acoustic drivers. Most hydrophones are constructed of ceramic piezoelectric transducer material, which operates effectively over the range of a few Hz to tens of kHz. In active sonar the same hydrophone elements are used for both transmit and receive. The total instantaneous array output power for long range sonar systems can be many tens of kilowatts. At very low frequencies some systems use electromagnetic linear motors (like speaker driver coils) or hydraulic actuators. Infrasonic pulses have also been generated using explosive charges.

Depending on the intended application and the supporting platform, sonar arrays are found in a variety of physical forms. Spherical or cylindrical arrays as seen in Figure 20.36 are housed in the bulbous bow protrusions below the water line of many military surface ships, and encased in the streamlined bow of submarines. These typically operate in the 1–6 kHz range and are capable of steering pencil beams in both azimuth and depression angle. An example of a spherical array is the US Navy AN/BSY-2 sonar on the Sea Wolf submarine.

FIGURE 20.36

An illustration of the spherical array in the Virginia III class of submarines.

Credit: Defense Industry Daily.

Conformal arrays use a thinly layered grid of hydrophones mounted on the nose or sides of vessels so as to blend smoothly with the contours of the hull design. Though this may be a less than ideal geometry for acoustic beamforming, it has the benefit of maintaining a streamlined structure for reduced drag and flow noise turbulence while allowing a larger aperture than is practical with a spherical array.

Long tubular towed array lines are pulled behind surface ships, submarines, and barges. Many hydrophones are spaced regularly inside a garden-hose-like tube that can be thousands of feet long. Depth is controlled either by adjusting the payout of the tow cable, or with an actively controlled tow body at the end of the array or at the tow cable attachment point. This enables steering to, and maintaining a desired depth (see Figure 20.37). Because of their length, towed arrays offer very large apertures for increased bearing resolution, narrow beams, high sensitivity due to many sensors and separation from ship self noise, and lower frequency operation as compared to hull mounted arrays. One drawback with the towed array is its one-dimensional linear geometry which leads to annularly symmetric (donut shaped) formed beampatterns. This yields no directivity in the vertical dimension, and a left-right ambiguity that requires the support ship to make turn maneuvers to resolve. The low frequency, long range, barge-towed US Navy SURTASS system is an example of a towed array sonar.

When mobility is essential or when submarine detection is needed at the far perimeter of the sonar reach from a naval battle group, then helicopter-borne dipping sonar is highly effective. A sonar array is reeled down to great depth from a hovering helicopter. Figure 20.38 shows a 1980s era system that is still in service, the US Navy AN/AQS-13 sonar. More recent developments like the US Navy AN/AQS-22 ALFS dipping sonar include extendable hydrophone support arms which increase aperture and permit lower frequency operation.

Modern torpedoes like the US Navy Mk 48 ADCAP shown in Figure 20.39 are quite autonomous. They are able to search out, detect, track, and target surface ships and submarines without the necessity

FIGURE 20.37

A French type F70 frigate (the Motte-Picquet) fitted with VDS (Variable Depth Sonar) type DUBV43 or DUBV43C towed array sonars. The array reeling mechanism and tow depth control body can be seen.

Credit: Used by permission, NetMarine. Photographer: Jean-Michel Roche.

FIGURE 20.38

A US Navy 1980s era Sikorsky SH-3H Sea King helicopter lowers its AN/AQS-13 dipping sonar.

Credit: US DefenseImagery (www.defenseimagery.mil), PH1 R.O. Overholt, USN.

FIGURE 20.39

Maintenance on an early development model of the US Navy Mk 48 ADCAP torpedo. The sonar hydrophone array lies behind the rubber shielded flat front nose plate.

Credit: US DefenseImagery (www.defenseimagery.mil).

of guidance and control from the launching boat. These tasks are performed using a nose-mounted planar array and on-board signal processing. The Mk 48 array is a nose-mounted grid of piezoelectric hydrophones which steer pencil beams for detection and direction finding.

Other sonar arrays are not mobile, but are permanently moored to the ocean floor. We will discuss in Section 3.20.7.3 how a fixed vertical line array can be used in matched field processing to estimate source range and depth using only passive observations. There are a number of very large and widely dispersed bottom-affixed passive sensor arrays used for surveillance in strategic ocean regions, including the US Navy's SOSUS network.

3.20.7.2 The undersea acoustic channel

Effective sonar signal processing requires an understanding of the challenging characteristics of sound propagation in an the ocean environment. In many ways sonar propagation is more complex and variable than the radio frequency channel encountered in wireless communications, radio astronomy, or radar. Fortunately though, propagation in the deep ocean is well understood, can be modeled accurately, and coherent processing across a large sensor array is possible even for distant sources.

3.20.7.2.1 Propagation models

Sound velocity in salt water is nominally 5000 ft per second, but this varies significantly with depth, sea temperature, and local salinity. Figure 20.40 illustrates a representative depth-dependent sound velocity profile $c(z)$ and the resulting ray propagation characteristics. The increasing velocity near the sea floor is due to greatly increased pressure in the deep isothermal layer below about 3000 ft. Velocity also increases as depth decreases between the deep sound channel and the surface duct due to rising temperature as depth decreases in the main thermocline layer. The cross-over between these two effects leads to a velocity minimum that focuses acoustic energy in the stable deep sound channel which can propagate great distances and maintain coherency across the ray paths.

Propagation effects near the surface are much more variable and depend on diurnal heating and cooling, surface mixing due to wind and wave action, latitude, and formation of seasonal thermocline layers. A relatively shallow surface layer duct often forms which traps energy near the surface, allowing sonar propagation and detection with shallow arrays. This may be the only possibility if the source also lies in the duct. However, due to sea roughness and losses at the sea-air interface, transmission loss is greater in this layer and rays die out more rapidly than in the deep sound channel.

Another important propagation effect not illustrated in Figure 20.40 is the convergence zone. Rays of higher angular incidence (at the sensor array or source) will periodically extend beyond the deep sound channel and intersect the surface. This forms a ring on the surface at a fixed range from the sonar array of convergent ray paths that enable surface ship detection at great distances. The radial separation between the successive convergence zones is typically on the order or 20 miles.

Optimal placement of sensor arrays (in depth), identifying convergence zones, and estimating range require specific knowledge of $c(z)$ to enable numerical ray path modeling of the sound channel. This information is obtained to great depths by bathythermograph and velocimeter sounders which are dropped overboard from surface ships or non-retrievably deployed from helicopters or other naval support fixed wing aircraft (e.g., the PC-3 Orion). When there are multiple vessels in the operational theater it is possible to collect these environmental data periodically over a large area. In situ measurements are supplemented with historical data, seasonal and weather models, and ocean bottom topography data to provide quite accurate sound velocity profile results. These enable useful acoustic channel ray trace modeling.

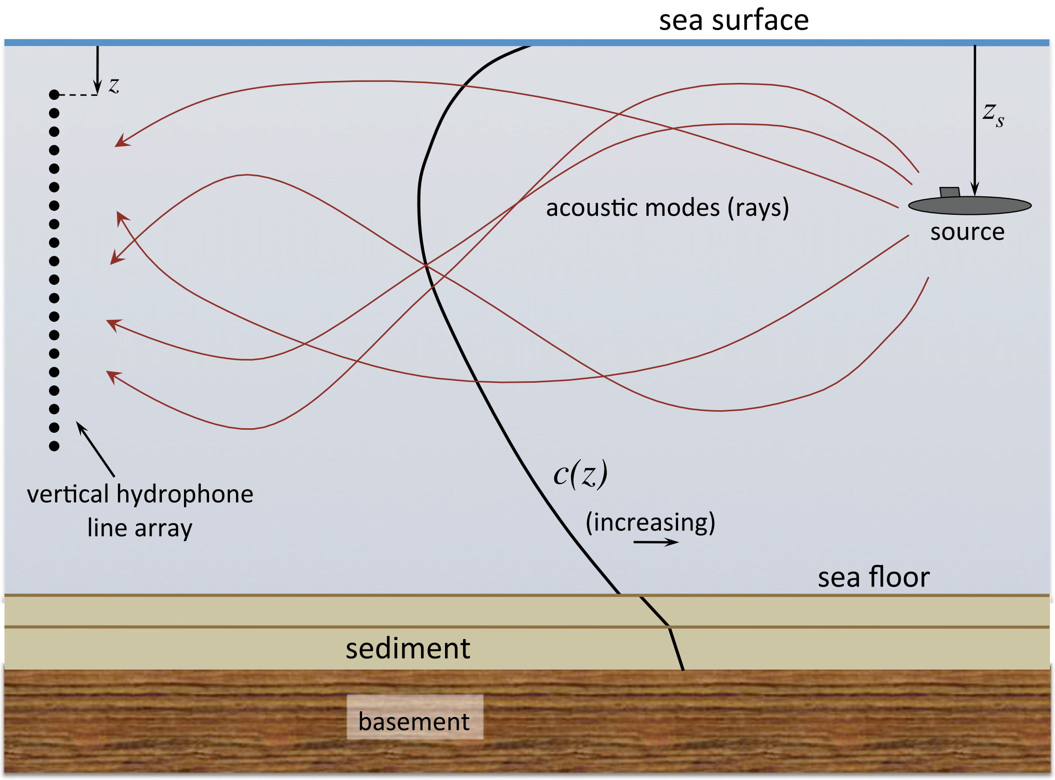


FIGURE 20.40

Acoustic propagation model for a horizontally stratified ocean. The convex-to-the left sound velocity function $c(z)$ forms long distance propagating modes (or acoustic ray paths) in the deep sound channel. The complex wavefront seen at the hydrophone array depends on source depth z and range, enabling estimation of these source parameters using MFP.

3.20.7.2.2 *Transmission loss*

Signal loss in the undersea acoustic channel is due to physical wavefront spreading, volume absorption, and leakage and scattering at the surface and bottom. One would expect near-field propagation to follow a spherical spreading law with loss proportional to $1/r^2$, and long distance propagation confined by the ocean surface and sea bottom to a planar disc to have cylindrical spreading loss proportional to $1/r$, where r is range to the source. However extensive field measurements suggest that due to scattering and leakage, spherical spreading with $1/r^2$ loss is a better match over a wide range of conditions. A commonly used model for sonar transmission loss in dB is

$$\mathrm{TL} \approx (20 \log_{10} r) + \alpha r \times 10^{-3}, \tag{20.93}$$

$$\alpha = \frac{16\pi^2}{3\rho c^3}\left(\mu_s + \frac{3}{4}\mu_v\right) f^2, \tag{20.94}$$

where α is defined in units of decibels loss per thousand yards, r is (following sonar convention) in yards, and f is frequency in Hz. The first term in (20.93) is due to spherical spreading. Other parameters for pure distilled water are density $\rho \approx 1$ gm/cm^3, sound velocity $c \approx 1.5 \times 10^5$ cm/s, shear viscosity $\mu_s \approx$ 0.01 poises, and volume viscosity $\mu_v \approx 0.0281$ poises. Below about 100 Hz the effective α in sea water increases (as compared to distilled water) by a factor of 30 due primarily to dissolved magnesium sulfate.

The fact that attenuation in dB is proportional to f^2 suggests that for long range detection it will be highly advantageous to use low frequencies. This is born out in practice where short range and targeting sonars with small arrays typically operate at tens of kHz, medium range hull mounted arrays and helicopter dipping sonars are at 1–6 kHz, and long range towed array sonars cover 10 Hz to a few 100 Hz. Even at low frequencies, spreading losses are significant at long ranges so in order to put sufficient energy into the water, active sonar systems typically use very long transmit pulses on the order of several seconds. Fortunately the sound channel is stable over such long pulse periods and coherent matched filter detection processing of echoes is possible.

3.20.7.2.3 *Noise and reverberation*

Sonar systems must detect weak signals in an inherently very noisy environment. Noise sources are many and varied, but we will mention approximate average levels for some significant sources in deep water conditions.

- Between 1 Hz and 10 Hz there are a variety of sources that contribute to an average level of approximately 105 dB rel 1 µPa at 1 Hz, which declines with a slope of −30 dB per decade of frequency increase.
- Between about 10 Hz and 150 Hz, the dominant source is mechanical and turbulence noise from distant surface shipping. Acoustic levels range from 60 to 85 dB rel 1 µPa for light to heavy shipping traffic conditions, respectively.
- Between about 100 Hz and 100 kHz The dominant source is surface noise from wind and wave action. Levels decline with increased frequency with a slope of about −20 dB per decade. At 1 kHz surface noise levels range from 44 to 70 dB rel 1 µPa for sea state 0–6, respectively.

Additional external noise sources include biologics such as shrimp (one of the loudest) and marine mammals. Self-generated noise from the platform vehicle is of course very local and potentially strong. It includes flow noise due to turbulence across the hydrophone surfaces for a moving platform, and propulsion, machinery and other noise associated with the support vehicle.

Flow noise is very local to each hydrophone and is thus well modeled as statistically independent per sensor, and often i.i.d. Surface sea state noise is typically quite widespread and can often be approximately modeled as isotropic within a horizontal plane containing the array. Shipping noise can be directional (spatially colored) since it is concentrated in well traveled shipping lanes. Biologic sources include very direction-dependent and distant marine mammals and more local swarms of shrimp-like noise makers. Beamforming algorithms which place nulls on the nearby directional noise sources can be very effective in improving SNR in this environment.

Detection processing in active sonar must deal with significant reverberation. There are three main sources: volume reverberation, surface reflections, and bottom backscatter from rough clutter topgraphic features. Volume reflection is due to widely distributed particulate matter and marine life in the path of the transmit beam. It is strongest during the early portion of the pulse period. The initial surface reflection arrives from directly above the array, with additional backscatter occurring as ray paths intersect the surface during rough sea-air interface conditions due to higher sea state. Multiple bottom-surface reflections lead to a nearly periodic structure for reverberation peaks within an exponentially decaying envelope. Reverberation can be reduced by extending the vertical array aperture to narrow the transmit and receive beampatterns in the vertical dimension. A time-dependent automatic gain control is also used in the receiver to avoid signal clipping during strong reverberation early in the pulse period. When the target of interest is moving, Doppler gating can also be used to reject reverberation from stationary clutter and to highlight the frequency-shifted target echo return.

3.20.7.3 Matched field processing

Matched field processing (MFP) is an interesting passive sonar application where modeled ray propagation is compared with the signal spatial structure at the receive array to estimate parameters of interest. The classical MFP application is to estimate source (target) range and depth at great distances, though it is theoretically possible to use the technique to identify environmental parameters such as ocean floor geography, propagation medium inhomogeneities, and even global undersea tomography. Unlike active sonar where two-way time of flight is used to directly measure range, and depth is not usually observable, in passive MFP it is possible to infer these parameters from the spatial phase and amplitude structure of the wavefront at the sensor array. Source localization can be viewed as an acoustic channel inversion problem exploiting sufficient complexity in the spatial signal distribution across a sensor array and a well-modeled channel transfer function between any candidate source position and each array sensor. This is not simple wavefront curvature estimation which can only be used effectively at shorter ranges within the Fresnel limits of the array. Success depends on the ability to accurately model the ducted, wave-guided acoustic propagation from source to sensor in a planar channel constrained by the ocean surface above, sea floor below, and a thermal-gradient-induced refractive deep-sound channel which directs (bends) acoustic rays with shallow incidence angles back toward the channel center. As illustrated in Figure 20.40, signals arrive at the array as a finite set of multipath rays, or acoustic modes, which are the discrete solutions to the frequency domain wave Eq. (20.95) and whose angular and depth distribution, or spatial spectrum, depends on the channel structure and source and sensor element positions.

To solve the channel inversion problem, source position parameters are varied within a propagation model for an optimization search to find the best "match" between the predicted and observed acoustic "field" at the sensor array. MFP relies on accurate full wave parametric modeling of acoustic waveguide propagation between the source and the array of hydrophone sensors, typically constructed as a vertical line array. Model mismatch of course impairs localization performance, but at low frequencies (10–100s of Hz) the sea channel maintains remarkable phase coherency across propagation modes and ray paths and existing models are sufficiently accurate. The MFP approach has been successfully demonstrated over ranges of hundreds of kilometers. Externally provided environmental parameters needed in the model include the sound velocity profile, bottom composition, and bottom topography. Contemporary measurements are obtained within a few hours of the MFP observations by deploying sounding instruments overboard to record sound velocity to great depths.

In this section we will follow in part the development found in [131]. Assuming waveguided propagation and a distant source, surface and bottom scattered signal components are attenuated to the point where they are negligible relative to modal components. These modes represent the multiple ray paths in the deep sound channel between the source and individual array elements, and are the discrete solutions to the temporal frequency domain wave equation

$$\left[\nabla^2 + K^2(\mathbf{v})\right] g(\mathbf{v}, \mathbf{v}_s) = -\delta(\mathbf{v} - \mathbf{v}_s), \tag{20.95}$$

where $\mathbf{v} = [x, y, z]^T$ is a position vector for an arbitrary point in the ocean channel, $g(\mathbf{v})$is the normalized (assuming a unit amplitude source) velocity potential or pressure, ∇^2 is the spatial Laplacian operator, $K(\mathbf{v}) = \frac{\Omega}{c(\mathbf{v})}$ is the position dependent medium wave number, Ω is the radian frequency of the narrowband acoustic source, $c(\mathbf{v})$ is the local sound velocity, $\mathbf{v}_s$ is the source position, and $\delta(\cdot)$ is the 3D delta function. Since the source is modeled as a fixed point radiator, solutions $g(\mathbf{v})$ are interpreted as the Green's function for the propagation channel between $\mathbf{v}_s$ and $\mathbf{v}$.

It is often possible to model the acoustic channel with a horizontally stratified ocean as shown in Figure 20.40, where sound speed $c(\mathbf{v}) = c(z)$ is a function of depth only and surface and bottom act as partially reflecting parallel plates. At greater distances this model is more accurate since only rays confined in the deep sound channel have survived surface and bottom scattering and attenuation. In this case we may separate out dependence on z in (20.95) and simplify by centering the (x, y) coordinate system on the source so $g(\mathbf{v}, \mathbf{v}_s)$ may be re-parameterized as $g(\bar{\mathbf{v}}, z, z_s)$ where $\bar{\mathbf{v}} = [x - x_s, y - y_s]^T$. The 2-D inverse spatial Fourier transform relationship with respect to only x and y is then

$$g(\bar{\mathbf{v}}, z, z_s) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} G(\mathbf{k}, z, z_s) e^{j\mathbf{k}^T \bar{\mathbf{v}}}\, d\mathbf{k}, \tag{20.96}$$

where $\mathbf{k} = [k_x, k_y]^T$ is the 2-D horizontal wavenumber vector. Note that the medium wavenumber is given by $K^2 = k_x^2 + k_y^2 + k_z^2$. Acoustic pressure $g(\bar{\mathbf{v}}, z, z_s)$ is interpreted as that seen by a hydrophone at depth z and 2D range $\bar{\mathbf{v}}$ relative to the source, which is at depth z_s. The wave equation is then expressed in the 2-D spatial frequency domain by substituting (20.96) into (20.95)

$$\frac{1}{4\pi^2} \int_{-\infty}^{\infty} \left[\nabla^2 + K^2(z)\right] G(\mathbf{k}, z, z_s) e^{j\mathbf{k}^T \bar{\mathbf{v}}}\, d\mathbf{k} = -\frac{1}{4\pi^2} \int_{-\infty}^{\infty} \delta(z - z_s) e^{j\mathbf{k}^T \bar{\mathbf{v}}}\, d\mathbf{k}, \tag{20.97}$$

$$\left[\frac{\partial^2}{\partial z^2} + K^2(z) - |\mathbf{k}|^2\right] G(\mathbf{k}, z, z_s) = -\delta(z - z_s), \tag{20.98}$$

where in (20.97) we have used $\delta(\mathbf{v} - \mathbf{v}_s) = \delta(\bar{\mathbf{v}})\delta(z - z_s)$ and $-\frac{1}{4\pi^2} \int_{-\infty}^{\infty} e^{j\mathbf{k}^T \bar{\mathbf{v}}}\, d\mathbf{k} = \delta(\bar{\mathbf{v}})$. Equation (20.98) follows by matching the Fourier transform arguments, noting that $G(\mathbf{k}, z, z_s)$ does not depend on x or y, and that the vertical wave number is given by $k_z^2 = K^2 - |\mathbf{k}|^2$. Solutions $G(\mathbf{k}, z, z_s)$ to (20.98) are known as the depth dependent Green's functions.

In general these solutions span a continuous range of $\mathbf{k}$ corresponding to the different directions of arrival for the wave fronts at z. But in the horizontally stratified, waveguided case we are considering, only discrete values of $\mathbf{k}$ for distinct ray paths represent propagation modes with significant energy, and these directions of arrival at $(\bar{\mathbf{v}}, z)$ are confined to the vertical plane containing the source and sensor. Thus $G(\mathbf{k}, z, z_s)$ typically consists of a series of complex-amplitude-scaled delta functions at discrete wavenumbers $\mathbf{k}_n$ corresponding to the nth propagation ray. The inverse transform of (20.96) then takes the form

$$g(\bar{\mathbf{v}}, z, z_s) = \frac{1}{4\pi^2} \sum_n G(\mathbf{k}_n, z, z_s) e^{j\mathbf{k}_n^T \bar{\mathbf{v}}}. \tag{20.99}$$

A central component of every MFP algorithm is a numerical simulation for the forward propagation model. Given a sound velocity profile $c(z)$, the simulation solves (20.98) for the discrete rays and uses (20.99) to compute $g(\bar{\mathbf{v}}, z, z_s)$ as a function of all $\bar{\mathbf{v}}$ and z_s values in the search range and for every z corresponding to a sensor array element depth. The fundamental MFP strategy is to solve the inverse propagation problem with an exhaustive search for the best match with respect to $\bar{\mathbf{v}}$ and z_s between the forward modeled responses and the measured sensor array response structure seen in output samples $\mathbf{y}(i)$.

Assuming a vertical line array of M elements and an MFP search in range and depth, the sampled data vector at the array is

$$\mathbf{y}(i) = \mathbf{a}(\bar{\mathbf{v}}, z_s)s(i) + \mathbf{n}(i), \tag{20.100}$$

where the parametric array spatial response vector is given by

$$\mathbf{a}(\bar{\mathbf{v}}, z_s) = \begin{bmatrix} g(\bar{\mathbf{v}}, z_1, z_s) \\ \vdots \\ g(\bar{\mathbf{v}}, z_M, z_s) \end{bmatrix} \tag{20.101}$$

and where z_m, $1 \leq m \leq M$, is the depth of the mth array hydrophone sensor, $s(i)$ is a zero-mean Gaussian random source process with variance σ_s^2 and $\mathbf{n}(i)$ is the noise sample vector including flow noise, surface winds and shipping, biologics, etc. Assuming wide sense stationarity, the covariance matrix is

$$\mathbf{R} = E[\mathbf{y}(i)\mathbf{y}^H(i)] = \sigma_s^2 \mathbf{a}(\bar{\mathbf{v}}, z_s)\mathbf{a}^H(\bar{\mathbf{v}}, z_s) + \mathbf{R}_{\mathrm{n}} = \mathbf{R}_s(\bar{\mathbf{v}}, z_s) + \mathbf{R}_{\mathrm{n}}. \tag{20.102}$$

As an MPF performance metric one can quantify how unique the array spatial response is for distinct values of the parameters $\bar{\mathbf{v}}$ and z_s, since this is related to the invertibility of the channel. To this end, a very useful measure is the ambiguity function defined as

$$\phi(\bar{\mathbf{v}}_1, z_1; \bar{\mathbf{v}}_2, z_2) = \left| \frac{\mathbf{a}^H(\bar{\mathbf{v}}_1, z_1)}{\|\mathbf{a}(\bar{\mathbf{v}}_1, z_1)\|} \cdot \frac{\mathbf{a}(\bar{\mathbf{v}}_2, z_2)}{\|\mathbf{a}(\bar{\mathbf{v}}_2, z_2)\|} \right|^2. \tag{20.103}$$

If $\phi(\bar{\mathbf{v}}_1, z_1; \bar{\mathbf{v}}_2, z_2)$ has multiple equally large peaks, then the MFP solution is ambiguous. Ideally it would have a single narrow peak at $(\bar{\mathbf{v}}_1 = \bar{\mathbf{v}}_2; z_1 = z_2)$ for all $\bar{\mathbf{v}}_1, z_1$, which would yield high resolution unique solutions, but significant sidelobe patterns are common. Array length and depth and the sound velocity profile $c(z)$ affect the shape of the ambiguity function and thus channel invertibility.

3.20.7.4 Acoustic vector sensors

As we transition to microphone arrays for aeroacoustical applications, we briefly mention here a relatively new type of acoustical *vector* sensor (AVS), which essentially amounts to an "array on a sensor." An AVS can measure the vector-valued acoustic particle velocity in addition to the scalar-valued sound pressure, and such sensors have been manufactured for acoustic measurements in both air and water. An example of an aeroacoustic vector sensor is shown in Figure 20.41. The output of a general AVS in free space can be represented as

$$\mathbf{y}(t) = \begin{bmatrix} 1 \\ \mathbf{u}(\theta, \phi) \end{bmatrix} x(t) + \mathbf{n}(t), \tag{20.104}$$

FIGURE 20.41

A vector acoustic sensor manufactured by Microflown Technologies, The Netherlands.

where $x(t)$ represents the sound pressure, $\mathbf{n}(t)$ noise and interference, and

$$\mathbf{u}^T(\theta, \phi) = [\cos(\theta)\cos(\phi) \quad \sin(\theta)\cos(\phi) \quad \sin(\phi)]^T$$

is a unit vector at the sensor pointing towards the source at azimuth angle θ and elevation angle ϕ. If the sensor is located near a reflecting surface (e.g., a wall or the ocean floor), then a reflection term is added to (20.104) to account for the source image.

The key distinguishing feature of an AVS is the fact that it produces a four-dimensional measurement at essentially a single point in space. A single AVS can be used to localize two separate sources, and additional resolving power can be obtained by an array of AVS within a relatively small aperture. AVS provide an interesting alternative to standard hydrophones or microphones in acoustic source localization and signal recovery.

3.20.8 Microphone arrays

The processing of acoustic signals in the air using arrays of microphones has received significant attention, although considerably less than for underwater acoustics due to the ubiquitous use of sonar in naval operations. While the use of microphone arrays has also been proposed for military applications, such as localization or identification of vehicles, helicopters, sniper fire, etc., such arrays have perhaps enjoyed more success in commercial settings, particulary those related to speech recovery or enhancement. For example, the ability of microphone arrays to locate an acoustic source such as a speaker, extract an acoustic signal in a noisy and reverberant environment, and synthesize arbitrary sound fields has led to

FIGURE 20.42

4×4 microphone array manufactured by iSEMcon GmbH, Germany.

their use in advanced video-conferencing systems, "hands-free" communication systems, surveillance of criminal activity and simulation of concert hall acoustics in high-end audio systems. Figure 20.42 shows a 4×4 array used for sound field mapping and source localization. A simple array typical of those used for teleconferencing applications is shown in Figure 20.43.

Similar to underwater acoustics, signal processing with microphone arrays relies on wideband data models, where the propagation time across the array is usually much greater than the inverse bandwidth. For example, it takes about 6 ms for sound to travel one half meter (a typical array aperture), while the inverse bandwidth of a speech signal is around 0.2–0.5 ms. On the other hand, since the speed of sound in air is over four times slower than in water, and since the frequencies of interest for aero-acoustics are usually higher than in sonar, microphone arrays can be much more compact. At 1 kHz, the wavelength of sound is about 30 cm, so arrays with apertures under a few meters are common. Consequently, plane-wave propagation models are typically assumed, at least locally, in the vicinity of the array. Propagation in outdoor environments is complicated by wind and temperature gradients that make precise localization difficult over long ranges. Even in situations where straight-line propagation can be assumed, random fluctuations in the air and temperature will cause a transmitted and received acoustic signal to lose temporal coherence if the signal travels a large distance. Indoors, the main obstacle to overcome is reverberation due to reflections of the sound from floors, walls, ceilings, furniture, etc.

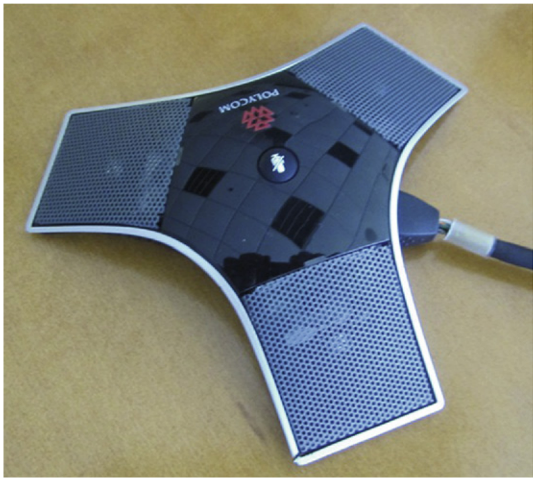


FIGURE 20.43

Microphone array used in video-conferencing applications. Manufactured by Polycom, Inc., San Jose, CA.

Consequently, the focus of most microphone array applications in outdoor settings is source localization, while indoors the most common problem is reconstruction of a desired acoustic source in the presence of noise and multipath. We briefly discuss aspects of these two problems below.

3.20.8.1 Aeroacoustic source localization

The term "acoustic camera" is often used to refer to microphone arrays that are used to characterize sound fields and locate sources of acoustic energy. Since the aeroacoustic signals used for localization are typically wideband, models for the problem tend to be formulated in the frequency domain. Let $\mathbf{y}(t) = [y_1(t) \ \cdots \ y_M(t)]^T$ denote the output of an M-microphone array. Assuming zero-mean wide-sense stationary signals, the array output is characterized by its cross-correlation matrix

$$\mathbf{R}_y(\tau) = \mathcal{E}\{\mathbf{y}(t+\tau)\mathbf{y}(t)^T\} \tag{20.105}$$

and the corresponding cross spectral density (CSD) matrix $\mathbf{G}_y(\omega)$ whose i, jth element is defined as

$$G_{y,ij}(\omega) = \int_{-\infty}^{\infty} R_{y,ij}(\tau)e^{-j\omega\tau}d\tau, \tag{20.106}$$

where $R_{y,ij}(\tau)$ is element i, j of $\mathbf{R}_y(\tau)$.

In general, the elements of the CSD may be expressed as

$$G_{y,ij}(\omega) = e^{-j\omega\tau_{ij}(\mathbf{p})}G_{s,ij}(\omega) + \sigma^2\delta_{ij}(\omega), \tag{20.107}$$

where $\sigma^2(\omega)$ is the CSD of the noise (assumed to be uncorrelated at each microphone), $\tau_{ij}(\mathbf{p})$ is the propagation delay between the two microphones, which is a function of the location of the source $\mathbf{p}$, and

$$G_{s,ij}(\omega) = \gamma_{s,ij}(\omega)[G_{s,i}(\omega)G_{s,j}(\omega)]^{1/2}, \tag{20.108}$$

where $\gamma_{s,ij}(\omega)$ is the spectral coherence function for the two sensors satisfying $0 \leq |\gamma_{s,ij}(\omega)| \leq 1$, and $G_{s,i}(\omega)$ represents the CSD of the source at microphone i. In general, $G_{s,i}(\omega) \neq G_{s,j}(\omega)$ when $i \neq j$ due to propagation inhomogeneities that occur as the signal travels between the two microphones. If microphones i and j are close enough together such that one can assume spatially coherent planewave propagation between them, then $\gamma_{s,ij}(\omega) = 1$ and $G_{s,i}(\omega) = G_{s,j}(\omega)$.

A convenient and very general approach is to assume an array-of-arrays situation, where several (say, K) arrays of closely-spaced microphones with locally coherent propagation are distributed over a larger area and separated by distances over which coherent propagation cannot generally be assumed. This model subsumes the two cases discussed above. If the vector outputs of each array $\mathbf{y}_k(t)$ are stacked on top of each other to form the super-vector $\mathbf{y}(t) = [\mathbf{y}_1(t)^T \cdots \mathbf{y}_K(t)^T]^T$, then the $MK \times MK$ CSD matrix will be given by

$$\mathbf{G}_y(\omega, \mathbf{p}) = \begin{bmatrix} \mathbf{a}_1(\omega, \mathbf{p})\mathbf{a}_1^H(\omega, \mathbf{p})G_{s,1}(\omega) & \cdots & \mathbf{a}_1(\omega, \mathbf{p})\mathbf{a}_K^H(\omega, \mathbf{p})e^{-j\omega\tau_{1K}(\mathbf{p})}G_{s,1K}(\omega) \\ \cdots & \ddots & \cdots \\ \mathbf{a}_K(\omega, \mathbf{p})\mathbf{a}_1^H(\omega, \mathbf{p})e^{-j\omega\tau_{K1}(\mathbf{p})}G_{s,K1}(\omega) & \cdots & \mathbf{a}_K(\omega, \mathbf{p})\mathbf{a}_K^H(\omega, \mathbf{p})G_{s,K}(\omega) \end{bmatrix} + \sigma^2(\omega)\mathbf{I}, \tag{20.109}$$

where $\mathbf{I}$ is an $MK \times MK$ identity matrix (assuming for simplicity that the noise CSD is the same at each array),

$$\mathbf{a}_k(\omega, \mathbf{p}) = \begin{bmatrix} e^{-j\omega\tau_{k,11}(\mathbf{p})} \\ \vdots \\ e^{-j\omega\tau_{k,1M_k}(\mathbf{p})} \end{bmatrix}, \tag{20.110}$$

and where $\tau_{k,ij}(\mathbf{p})$ represents the propagation delay between microphones i and j for array k with M_k elements.

If one has access to the outputs of all K of the arrays and the various source CSDs $G_{s,ij}(\omega)$ are known, a procedure for estimating the source location $\mathbf{p}$ based on samples of $\mathbf{G}_y(\omega, \mathbf{p})$ at different frequencies can easily be formulated. Such an approach would require the arrays to share all their data with a fusion center, which incurs a large communication overhead. In addition, knowledge of $G_{s,i}(\omega)$ implies that the arrays can somehow obtain time-aligned measurements of the source CSD, which is problematic without knowledge of the source location. The latter issue can be resolved by absorbing the time-delay terms $e^{-j\omega\tau_{kl}(\mathbf{p})}$ between arrays k and l into $G_{s,kl}(\omega)$, and basing the estimate of $\mathbf{p}$ on just the intra-array phase shifts. An alternative approach is to estimate the direction-of-arrival (DOA) of the source signal at each array using only the locally calculated CSD matrix $\mathbf{a}_k(\omega, \mathbf{p})\mathbf{a}_k^H(\omega, \mathbf{p})G_{s,k}(\omega)$ (the location of the source is not identifiable at each array individually, only the source DOA). Each array would then forward only its estimated DOA to the fusion center, which would then estimate $\mathbf{p}$ via triangulation. Various studies of the Cramér-Rao Bound have been conducted to determine the difference in achievable performance for these approaches.

3.20.8.2 Wideband adaptive beamforming

As mentioned above, in many microphone array applications, locating an acoustic source is less important than extracting its waveform in a reverberant and noisy environment. In relatively short-range indoor settings where factors that influence acoustic propagation (temperature, pressure, wind, etc.) are uniform, Doppler and dispersion effects can be ignored, and to a very good approximation the array will simply receive scaled and delayed versions of the source via a (potentially large) number of reverberant paths. In particular, at microphone m, the received acoustic signal can be represented as

$$y_m(t) = \sum_{i=1}^{N} \alpha_{i,m} s(t - \tau_{i,m}) + n_m(t), \tag{20.111}$$

where $s(t)$ is the desired source, N denotes the number of multipath echoes from the source to the microphone, $\{\alpha_{i,m}, \tau_{i,m}\}$ are the amplitude and the delay corresponding to path i at microphone m, and $n_m(t)$ is due to all other background noise and interference.

The most common approach to extracting $s(t)$ from the M-element microphone array output is via a wideband beamformer:

$$\hat{s}(t - t_0) = \sum_{m=1}^{M} \sum_{l=0}^{L} w_{ml} y_m(t - lT_s), \tag{20.112}$$

where T_s is the sampling period of the array, w_{ml} is the beamformer weight for microphone m at sample l, and t_0 is an arbitrary delay. This essentially amounts to a space-time equalizer similar to what might be employed in a frequency-selective wireless RF channel. The difference in the microphone array application is that one typically does not have access to periodic "training" data from the source to facilitate updates of the beamformer/equalizer weights, either in time via the LMS or RLS algorithms, or using a data-adaptive approach like MVDR beamforming. Instead, other factors must be exploited to adapt the weights. For example, one may know or be able to estimate the approximate location or DOA of the source, as in automobile voice-enhancement or video conferencing systems where the speaker(s) are confined to certain positions. Likewise, knowledge of the location of strong sources of acoustic interference (e.g., TVs, air conditioners, windows, etc.) can also be taken advantage of to help the filter focus on the source of interest. Adaptive noise canceling approaches are possible if reference waveforms are available for the interference, obtained for example by placing a microphone near the interfering source. One can also exploit situations where the source or interference is known to have strong components at certain frequencies, although in this case it is advantageous to implement the filter in the frequency domain:

$$\widehat{S}(\omega_k) = \sum_{m=1}^{M} W_m(\omega_k) Y_m(\omega_k). \tag{20.113}$$

Important factors to consider when implementing a wideband beamformer in microphone array applications are the sampling period T_s and the length L of the equalizer, which in many situations can be quite large for the required value of T_s. For example, to reconstruct a speech signal with a 3 kHz bandwidth in a room where the path lengths of the echos may vary by 5 m could require a value of L on the order of 300–400. For this reason, in computationally constrained scenarios, one may be forced to settle for a space-only beamformer followed by an adaptive echo canceler.

3.20.9 Chemical sensor arrays

In recent years, the use of model-based signal processing with chemical sensor arrays has received significant interest. Driving this interest has been important applications such as enironmental monitoring of air and water quality, chemical spills, detection and localization of air- or waterborne chemical weapons and even landmines. The ability to quickly discover and accurately locate sources of toxic chemicals is obviously a critical factor in mitigating their negative impact. The term "model-based" is used here to contrast against classical methods that simply use arrays of sensors to improve coverage or increase the probability of detecting a chemical event. While these are clearly important, our focus below will be on approaches that employ parametric models to describe chemical flow across the sensors, and as such can be used to locate and quantify other properties of the source(s) in addition to simply detecting their presence.

The key differentiating feature of applications involving chemical sensor arrays compared with others considered in this chapter is the fact that the signals of interest propagate according to diffusion rather than wave equations. Additional complications such as imprecisely known wind/currents, turbulence, eddys, vortices and boundary effects make it difficult to obtain an accurate mathematical model except in fairly simple circumstances. Nevertheless, results obtained from simplified models of the environment can serve as valuable approximations that provide useful information. Furthermore, they can be used to focus the local implementation of more complicated numerical operations that would be too involved to perform globally.

To illustrate the application of sensor arrays in localizing a diffusing chemical source, we will consider a simple example involving a point source in an open environment (all surfaces and other boundaries are far enough removed from the source and array so that their effects can be neglected) with homogeneous diffusivity in all directions. Assume the source is located at position $\mathbf{r}_0 = [x_0, y_0, z_0]^T$ and at time t_0 begins emitting the chemical substance at a constant rate of μ kg/s. Assume also that a wind/current is present with constant velocity vector $\mathbf{v}$. For this case, the diffusion equation that governs the concentration $c(\mathbf{r}, t)$ of the substance at some position $\mathbf{r}$ at time $t > t_0$ is given by

$$\frac{\partial c(\mathbf{r}, t)}{\partial t} = \kappa \nabla^2 c(\mathbf{r}, t) - \nabla c \cdot \mathbf{v}, \tag{20.114}$$

where κ measured in m^2/s is the diffusivity of the medium, which in the above expression is assumed to be incompressible. The solution to (20.114) is given by $c(\mathbf{r}, t) = \mu a(\mathbf{r}, t)$, where

$$\begin{aligned} a(\mathbf{r}, t) = &\frac{1}{8\pi\kappa \|\mathbf{r} - \mathbf{r}_0\|} \exp\left\{\frac{(\mathbf{r} - \mathbf{r}_0)^T \mathbf{v}}{2\kappa}\right\} \\ &\times \left[\exp\left\{\frac{\|\mathbf{r} - \mathbf{r}_0\| \|\mathbf{v}\|}{2\kappa}\right\} \operatorname{erfc}\left(\frac{\|\mathbf{r} - \mathbf{r}_0\|}{2\sqrt{\kappa(t - t_0)}} + \|\mathbf{v}\| \sqrt{\frac{t - t_0}{4\kappa}}\right)\right. \qquad (20.115) \\ &\left. \times \exp\left\{\frac{-\|\mathbf{r} - \mathbf{r}_0\| \|\mathbf{v}\|}{2\kappa}\right\} \operatorname{erfc}\left(\frac{\|\mathbf{r} - \mathbf{r}_0\|}{2\sqrt{\kappa(t - t_0)}} - \|\mathbf{v}\| \sqrt{\frac{t - t_0}{4\kappa}}\right)\right], \qquad (20.116) \end{aligned}$$

where $\text{erfc}(x) = \frac{2}{\pi}\int_x^\infty e^{-y^2}\,dy$ is the complementary error function. While technically the above model is appropriate for molecular diffusion, it can be applied to larger scale scenarios involving convective diffusion by adjusting the value of the diffusivity κ.

To characterize the chemical concentration at any point in space or time, one would need to know μ, the "strength" of the source, as well as the parameters in the vector $\boldsymbol{\theta} = [\mathbf{r}_0^T\ \kappa\ t_0]^T$, which include the location of the source and the time it became active. The diffusivity κ is also treated as an unknown constant, since it will depend on environmental factors (temperature, humidity, etc.) in a complicated way. To determine these unknowns, an array of sensors that measure the concentration of the chemical can be deployed. For example, a given sensor located at position $\mathbf{r}_i$ would observe the following concentration at some specific time t_k:

$$y_i(t_k) = a(\mathbf{r}_i, \boldsymbol{\theta}, t_k)\mu + n_i(t_k),$$

where $n_i(t)$ represents noise or modeling errors, and a is written as an explicit function of $\boldsymbol{\theta}$ to emphasize its dependence on the parameters of interest. If the observations from M sensors taken at K distinct time samples are stacked into a single observation vector $\mathbf{y}$, where element p of $\mathbf{y}$ is indexed according to $p = M(k-1)+i$ for $k = 1, \ldots, K$ and $i = 1, \ldots, M$, the standard array processing model is obtained:

$$\mathbf{y} = \begin{bmatrix} a(\mathbf{r}_1, \boldsymbol{\theta}, t_1) \\ \vdots \\ a(\mathbf{r}_M, \boldsymbol{\theta}, t_1) \\ a(\mathbf{r}_1, \boldsymbol{\theta}, t_2) \\ \vdots \\ a(\mathbf{r}_M, \boldsymbol{\theta}, t_K) \end{bmatrix} \mu + \mathbf{n} = \mathbf{a}(\boldsymbol{\theta})\mu + \mathbf{n}, \tag{20.117}$$

where the vector of noise samples $\mathbf{n}$ is organized like $\mathbf{y}$, and element $p = M(k-1)+i$ of the "steering" vector $\mathbf{a}(\boldsymbol{\theta})$ is given by $a(\mathbf{r}_i, \boldsymbol{\theta}, t_k)$. With the model of (20.117) in hand, one can apply standard array processing techniques to estimate μ and $\boldsymbol{\theta}$, provided that (20.117) is identifiable. In principle, unique identification of the three location parameters in $\boldsymbol{\theta}$, namely $\mathbf{r}_0$, requires that $M \geq 4$, and of course we require that the total number of observations MK exceed the number of free parameters (six in this model). In practice, of course, MK will likely need to be much larger than six in order to combat the effects of noise.

While the discussion above was for the simple case of an infinite open environment, a similar approach can be taken for more complicated scenarios provided that the diffusion equation can be solved. Cases of particular interest that have been addressed include a semi-infinite medium (e.g., a source on the ocean floor) and a large room of known dimensions. Boundary conditions play an important role in such cases, and different results are obtained depending on whether or not the boundaries are permeable to the chemical of interest. Source models different from the step function model assumed above can also be employed, such as impulse or pulsed waveforms. In settings involving very complicated geometries (e.g., urban canyons, buildings with offices and hallways, etc.), moving sources or sensors, or when more realistic propagation effects are taken into account (e.g., turbulence, eddys, inhomogeneous diffusivity, etc.), numerical methods are required to evaluate the response of the array to the chemical source. Details for these different modeling assumptions can be found in the references at the end of the chapter.

3.20.10 Conclusion

As we have seen above, the applications of array signal processing stretch from locating the sources of electrical energy from tiny neurons in the brain to astronomical objects millions of light-years away to submarines deep below the surface of the ocean. Remarkably, all of these applications share a very consistent underlying mathematical model that often allows techniques developed for one problem to apply to others in different fields. For this reason, we see similar methods appearing in journals related to radar, sonar, neurophysiology, acoustics, radio astronomy, medical imaging, seismology, and navigation, although explained in many cases with different terminology or emphases. Superficial differences in language aside, while the methods in the literature of these different areas are similar, they are not identical; each application has its own peculiarities that warrant special attention. Thus, in addition to showing what is common among the problems considered, our goal has also been to highlight the unique features of each application, and hence to provide motivation for the particular methodologies researchers and practitioners have adopted for these applications. Clearly, our discussion has only scratched the surface, and many details have been glossed over. It is our hope that we have piqued the reader's interest enough to pursue some of these details in the reference list (which is itself a small subset of what is available).

Relevant Theory: Signal Processing Theory, Machine Learning, and Statistical Signal Processing

See Vol. 1, Chapter 2 Continuous-Time Signals and Systems
See Vol. 1, Chapter 3 Discrete-Time Signals and Systems
See Vol. 1, Chapter 4 Random Signals and Stochastic Processes
See Vol. 1, Chapter 5 Sampling and Quantization
See Vol. 1, Chapter 6 Digital Filter Structures and Their Implementation
See Vol. 1, Chapter 7 Multirate Signal Processing for Software Radio Architectures
See Vol. 1, Chapter 8 Modern Transform Design for Practical Audio/Image/Video Coding Applications
See Vol. 1, Chapter 9 Discrete Multi-Scale Transforms in Signal Processing
See Vol. 1, Chapter 10 Frames in Signal Processing
See Vol. 1, Chapter 11 Parametric Estimation
See Vol. 1, Chapter 12 Adaptive Filters
See Vol. 1, Chapter 20 Clustering
See Vol. 1, Chapter 21 Unsupervised Learning Algorithms
See Vol. 1, Chapter 25 A Tutorial on Model Selection
See this Volume, Chapter 2 Model Order Selection
See this Volume, Chapter 7 Geolocation—Maps, Measurements, Models, and Methods
See this Volume, Chapter 8 Performance Analysis and Bounds

References and Further Reading

Space-time adaptive processing

[1] L. Brennan, F. Staudaher, Subclutter Visibility Demonstration, Adaptive Sensors, Inc., Technical Report RL-TR-92-21, 1992.

[2] I.S. Reed, J.D. Mallett, L.E. Brennan, Rapid convergence rate in adaptive arrays, IEEE Trans. Aerosp. Electron. Syst. AES-10 (1974) 853–862.
[3] B.D. Carlson, Covariance matrix estimation errors and diagonal loading in adaptive arrays, IEEE Trans. Aerosp. Electron. Syst. 24 (3) (1988) 397–401.
[4] I. Kirsteins, D. Tufts, Adaptive detection using low rank approximation to a data matrix, IEEE Trans. Aerosp. Electron. Syst. AES-30 (1) (1994) 55–67.
[5] P. Parker, A. Swindlehurst, space-time autoregressive filtering for matched subspace STAP, IEEE Trans. Aerosp. Electron. Syst. AES-39 (2) (2003) 510–520.
[6] J. Guerci, J. Goldstein, I. Reed, Optimal and adaptive reduced-rank STAP, IEEE Trans. Aerosp. Electron. Syst. 36 (2) (2000) 647–663.
[7] J. Guerci, Space-Time Adaptive Processing for Radar, Artech House, 2003.
[8] R. Klemm, Space-Time Adaptive Processing: Principles and Applications, IEE Press, 1998.
[9] J. Ward, Space-Time Adaptive Processing for Airborne Radar, MIT Lincoln Labs, Technical Report TR-1015, 1994.

MIMO radar

[10] J. Li, P. Stoica (Eds.), MIMO Radar Signal Processing, John Wiley & Sons, Inc., Hoboken, NJ, 2009.
[11] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, R. Valenzuela, MIMO radar: an idea whose time has come, in: Proceedings of the IEEE Radar Conference, April 2004, pp. 71–78.
[12] E. Fishler, A. Haimovich, R. Blum, L. Cimini, D. Chizhik, R. Valenzuela, Spatial diversity in radars—models and detection performance, IEEE Trans. Signal Process. 54 (3) (2006) 823–838.
[13] I. Bekkerman, J. Tabrikian, Target detection and localization using MIMO radars and sonars, IEEE Trans. Signal Process. 54 (2006) 3873–3883.
[14] P. Stoica, J. Li, Y. Xie, On probing signal design for MIMO radar, IEEE Trans. Signal Process. 55 (8) (2007) 4151–4161.
[15] J. Li, P. Stoica, MIMO radar with colocated antennas: review of some recent work, IEEE Signal Process. Mag. 24 (5) (2007) 106–114.
[16] L. Xu, J. Li, P. Stoica, Target detection and parameter estimation for MIMO radar systems, IEEE Trans. Aerosp. Electron. Syst. 44 (3) (2008) 927–939.
[17] A.H. Haimovich, R.S. Blum, L.J. Cimini, MIMO radar with widely separated antennas, IEEE Signal Process. Mag. 25 (1) (2008) 116–129.
[18] H. He, P. Stoica, J. Li, Designing unimodular sequence sets with good correlations—including an application to MIMO radar, IEEE Trans. Signal Process. 57 (11) (2009) 4391–4405.
[19] H. He, J. Li, P. Stoica, Waveform Design for Active Sensing Systems—A Computational Approach, Cambridge University Press, 2012.
[20] W. Roberts, P. Stoica, J. Li, T. Yardibi, F.A. Sadjadi, Iterative adaptive approaches to MIMO radar imaging, IEEE J. Sel. Top. Signal Process. 4 (1) (2010) 5–20.
[21] X. Tan, W. Roberts, J. Li, P. Stoica, Sparse learning via iterative minimization with application to MIMO radar imaging, IEEE Trans. Signal Process. (2011) 1088–1101.

Radio astronomy

[22] D. John, Kraus, Radio Astronomy, second ed., Cygnus-Quasar Books, Powell, Ohio, 1986.
[23] A.R. Thompson, J.M. Moran, G.W. Swenson Jr., Interferometry and Synthesis in Radio Astronomy, second ed., Wiley-Interscience, New York, 2001.

[24] B.D. Jeffs, K.F. Warnick, J. Landon, J. Waldron, J.R. Fisher D. Jones, R.D. Norrod, Signal processing for phased array feeds in radio astronomical telescopes, IEEE J. Sel. Top. Signal Process. 2 (5) (2008) 635–646.

[25] P.J. Napier, A.R. Thompson, R.D. Ekers, The very large array: design and performance of a modern synthesis radio telescope, Proc. IEEE 71 (1983) 1295–1320.

[26] R. Levanda, A. Leshem, Synthetic aperture radio telescopes, IEEE Signal Process. Mag. 27 (1) (2010) 14–29.

[27] R.J. Cornwell, K. Golap, S. Bhatnagear, The noncoplanar baselines effect in radio interferometry: the W-projection algorithm, IEEE J. Sel. Top. Signal Process. 2 (5) (2008) 647–657.

[28] A. Leshem, A.-J. van der Veen, A.-J. Boonstra, Multichannel interference mitigation techniques in radio astronomy, Astrophys. J. Suppl. 131 (1) (2000) 355–374.

[29] A. Leshem, A.-J. van der Veen, Radio-astronomical imaging in the presence of strong radio interference, IEEE Trans. Inform. Theory 46 (5) (2000) 1730–1747.

[30] J.A. Högbom, Aperture synthesis with a nonregular distribution of interferometer baselines, Astron. Astrophys. Suppl. 15 (1974) 417–426.

[31] F.R. Schwab, Relaxing the isoplanatism assumption in self-calibration: application to low-frequency radio interferometry, Astron. J. 131 (6) (1984) 646–659 (pt. F).

[32] R.J. Cornwell, Multiscale CLEAN deconvolutions of radio synthesis images, IEEE J. Sel. Top. Signal Process. 2 (5) (2008) 793–801.

[33] J. Landon, M. Elmer, D. Jones, A. Stemmons, B.D. Jeffs, K.F. Warnick, J.R. Fisher, R.D. Norrod, Phased array feed calibration, beamforming, and imaging, Astron. J. 139 (3) (2010) 1154–1167.

[34] K.F. Warnick, M.A. Jensen, Effects of mutual coupling on interference mitigation with a focal plane array, IEEE Trans. Antennas Propag. 53 (8) (2005) 2490–2498.

[35] M. Elmer, B.D. Jeffs, K.F. Warnick, J.R. Fisher, R. Norrod, Beamformer design methods for radio astronomical phased array feeds, IEEE Trans. Antennas Propag. 60 (2) (2012) 903–914.

[36] S. van der Tol, B.D. Jeffs, A.-J. van der Veen, Self calibration for the LOFAR radio astronomical array, IEEE Trans. Signal Process. 55 (9) (2007) 4497–4510.

[37] K.F. Warnick, B.D. Jeffs, Gain and aperture efficiency for a reflector antenna with an array feed, IEEE Antennas Propag. Lett. 5 (2006) 499–502.

[38] A.R. Taylor, S.J. Gibson, M. Peracaula, P.G. Martin, T.L. Landecker, C.M. Brunt, P.E. Dewdney, S.M. Dougherty, A.D. Gray, L.A. Higgs, C.R. Derton, L.B.G. Knee, R. Kothes, C.R. Purton, B. Uyaniker, B.J. Wallace, A.G. Willis, D. Durand, The Canadian galactic plane survey, Astron. J. 125 (2003) 3145–3164.

[39] Y. Bhattacharjee, Radio astronomers take arms against a sea of signals, Science 330 (6003) (2010) 444–445.

[40] J.F. Bell, S.W. Ellingson, J. Bunton, Removal of the GLONASS C/A signal from OH spectral line observations using a parametric modeling technique, Astrophys. J. Suppl. 135 (2001) 87–93.

[41] A.J. Poulsen, B.D. Jeffs, K.F. Warnick, J.R. Fisher, Programmable real-time cancellation of GLONASS interference with the Green Bank telescope, Astron. J. 130 (6) (2005) 2916–2927.

[42] W. Dong, B.D. Jeffs, J.R. Fisher, Radar interference blanking in radio astronomy using a Kalman tracker, Radio Sci. 40 (5) (2005).

[43] B.D. Jeffs, W. Lazarte, J.R. Fisher, Bayesian detection of radar interference in radio astronomy, Radio Sci. 41 (2006).

[44] S.W. Ellingson, G.A. Hampson, Mitigation of radar interfernece in L-band radio astronomy, Astrophys. J. Suppl. 147 (2003) 167–176.

[45] Q. Zhang, Y. Zheng, S.G. Wilson, J.R. Fisher, R. Bradley, Combating pulsed radar interference in radio astronomy, Astron. J. 126 (2003) 1588–1594.

[46] Q. Zhang, Y. Zheng, S.G. Wilson, J.R. Fisher, R. Bradley, Excision of distance measuring equipment interference from radio astronomy signals, Astron. J. 129 (6) (2005) 2933–2939.

[47] J.R. Fisher, Q. Zhang, S.G. Wilson, Y. Zheng, R. Bradley, Mitigation of pulsed interference to redshifted HI and OH observations between 960 and 1215 megahertz, Astron. J. 129 (6) (2005) 2940–2949.

[48] P.A. Fridman, W.A. Baan, RFI mitigation methods in radio astronomy, Astron. Astrophys. 378 (2001) 327–344.

[49] C. Barnbaum, R.F. Bradley, A new approach to interference excision in radio astronomy: real-time adaptive cancellation, Astron. J. 116 (1998) 2598–2614.

[50] B.D. Jeffs, L. Li, K.F. Wanick, Auxiliary antenna assisted interference mitigation for radio astronomy arrays, IEEE Trans. Signal Process. 53 (2) (2005) 439–451.

[51] S.W. Ellingson, G.A. Hampson, A subspace-tracking approach to interference nulling for phased array-based radio telescopes, IEEE Trans. Antennas Propag. 50 (1) (2002) 25–30.

[52] C.K. Hansen, K.F. Warnick, B.D. Jeffs, R. Bradley, Interferene mitigation using a focal plane array, Radio Sci. 40 (2005).

[53] J.R. Nagel, K.F. Warnick, B.D. Jeffs, J.R. Fisher, R. Bradley, Experimental verification of radio frequency interference mitigation with a focal plane array feed, Radio Sci. 42 (2007).

[54] J. Raza, A.-J. Boonstra, A.-J. van der Veen, Spatial filtering of RF interference in radio astronomy, IEEE Signal Process. Lett. 9 (2) (2002) 64–67.

Positioning and navigation

[55] E.D. Kaplan, C. Hegarty (Eds.), Understanding GPS: Principles and Applications, second ed., Artech House, 2005.

[56] Y.-H. Chen, J.-C.J.D.S.D. Lorenzo, J. Seo, S. Lo, P. Enge, D.M. Akos, Real-time software receiver for GPS controlled reception pattern antenna array processing, in: ION GNSS Conference, 2010.

[57] R.G. Lorenz, S.P. Boyd, Robust beamforming in GPS arrays, in: ION National Technical Meeting, 2002.

[58] S. Backén, On dynamic array processing for GNSS software receivers, Lulea University of Technology, Ph.D. Dissertation, 2011.

[59] G. Seco-Granados, J. Fernandez-Rubio, C. Fernandez-Prades, ML estimator and hybrid beamformer for multipath and interference mitigation in GNSS receivers, IEEE Trans. Signal Process. 53 (3) (2005) 1194–1208.

[60] M. Amin, W. Sun, A novel interference suppression scheme for global navigation satellite systems using antenna array, IEEE J. Sel. Areas Commun. 23 (5) (2005) 999–1012.

[61] D. Lu, Q. Feng, R. Wu, Survey on interference mitigation via adaptive array processing in GPS, PIERS Online 2 (4) (2006) 357–362.

[62] R. Fante, J. Vaccaro, Wideband cancellation of interference in a GPS receive array, IEEE Trans. Aerosp. Electron. Syst. 36 (2) (2000) 549–564.

[63] S.-J. Kim, R. Iltis, STAP for GPS receiver synchronization, IEEE Trans. Aerosp. Electron. Syst. 40 (1) (2004) 132–144.

[64] M. Amin, L. Zhao, A. Lindsey, Subspace array processing for the suppression of FM jamming in GPS receivers, IEEE Trans. Aerosp. Electron. Syst. 40 (1) (2004) 80–92.

[65] M.T. Brenneman, Y.T. Morton, Q. Zhou, GPS multipath detection with ANOVA for adaptive arrays, IEEE Trans. Aerosp. Electron. Syst. 46 (3) (2010) 1171–1184.

[66] J. Soubielle, I. Frijalkow, P. Duvaut, A. Bibaut, GPS positioning in a multipath environment, IEEE Trans. Signal Process. 50 (1) (2002) 141–150.

[67] A. Swindlehurst, Time delay and spatial signature estimation using known asynchronous signals, IEEE Trans. Signal Process. 46 (2) (1998) 449–462.

[68] A. Jakobsson, A. Swindlehurst, P. Stoica, subspace-based estimation of time delays and Doppler shifts, IEEE Trans. Signal Process. 46 (9) (1998) 2472–2483.

[69] M. Wax, A. Leshem, Joint estimation of time delays and directions of arrival of multiple reflections of a known signal, IEEE Trans. Signal Process. 45 (10) (1997) 2477–2484.

[70] H. Amindavar, A. Reza, A new simultaneous estimation of directions of arrival and channel parameters in a multipath environment, IEEE Trans. Signal Process. 53 (2) (2005) 471–483.
[71] M. Vanderveen, A.-J. van der Veen, A. Paulraj, Estimation of multipath parameters in wireless communications, IEEE Trans. Signal Process. 46 (3) (1998) 682–690.
[72] F. Antreich, J. Nossek, W. Utschick, Maximum likelihood delay estimation in a navigation receiver for aeronautical applications, Aero. Sci. Technol. 12 (3) (2008) 256–267 (online). <http://www.sciencedirect.com/science/article/pii/S1270963807000843>.
[73] G. Seco, A.L. Swindlehurst, D. Astély, Exploiting antenna arrays for synchronization, in: G.B. Giannakis, Y. Hua, P. Stoica, L. Tong (Eds.), Signal Processing Advances in Wireless Communications, Trends in Single- and Multi-User Systems, vol. II, Prentice-Hall, 2000, pp. 403–430 (Chapter 10).
[74] F. Antreich, J.A. Nossek, G. Seco-Granados, A.L. Swindlehurst, The extended invariance principle for signal parameter estimation in an unknown spatial field, IEEE Trans. Signal Process. 59 (7) (2011) 3213–3225.
[75] J.-C. Juang, G.-S. Huang, Development of GPS-based attitude determination algorithms, IEEE Trans. Aerosp. Electron. Syst. 33 (3) (1997) 968–976.
[76] J.K. Ray, M.E. Cannon, P.C. Fenton, Mitigation of static carrier-phase multipath effects using multiple closely spaced antennas, Navigation: J. Inst. Navigation (ION) 46 (3) (1999) 193–201.
[77] J. Ray, M. Cannon, P. Fenton, GPS code and carrier multipath mitigation using a multiantenna system, IEEE Trans. Aerosp. Electron. Syst. 37 (1) (2001) 183–195.

Wireless communications

[78] F. Khan, LTE for 4G Mobile Broadband—Air Interface Technologies and Performance, Cambridge University Press, 2009.
[79] S. Sesia, I. Toufik, M. Baker (Eds.), LTE—The UMTS Long Term Evolution: From Theory to Practice, Wiley, 2009.
[80] J. Lee, J.-K. Han, J.C. Zhang, MIMO technologies in 3GPP LTE and LTE-advanced, EURASIP J. Wireless Commun. Network. (2009) (online). <http://dx.doi.org/10.1155/2009/302092>.
[81] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, T. Thomas, LTE-advanced: next-generation wireless broadband technology (invited paper), IEEE Wireless Commun. 17 (3) (2010) 10–22.
[82] J. Duplicy, B. Badic, R. Balraj, P.H. Rizwan Ghaffar, F. Kaltenberger, R. Knopp, I.Z. Kovács, H.T. Nguyen, D. Tandur, G.Vivier, MU-MIMO in LTE systems, EURASIP J. Wireless Commun. Network. (2011).
[83] E. Dahlman, S. Parkvall, J. Skold, 4G: LTE/LTE-Advanced for Mobile Broadband, Academic Press, 2011.
[84] H. Taoka, S. Nagata, K. Takeda, Y. Kakishima, X. She, K. Kusume, MIMO and CoMP in LTE-advanced, NTT DOCOMO Techn. J. 12 (2) (2010) 20–28.
[85] E. Hossain, D.I. Kim, V.K. Bhargava, Cooperative Cellular Wireless Networks, Cambridge University Press, 2011.
[86] Q. Li, X. Lin, J. Zhang, W. Roh, Advancement of MIMO technology in WiMAX: from IEEE 802.16d/e/j to 802.16m, IEEE Commun. Mag. 47 (6) (2009) 100–107.
[87] S. Ahmadi, Mobile WiMAX: A Systems Approach to Understanding IEEE 802.16m Radio Access Technology, Academic Press, 2010.
[88] Q. Li, G. Li, W. Lee, M. il Lee, D. Mazzarese, B. Clerckx, Z. Li, MIMO techniques in WiMAX and LTE: a feature overview, IEEE Commun. Mag. 48 (5) (2010) 86–92.
[89] IEEE Std 802.16, IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 3: Advanced Air Interface, IEEE Std 802.16m-2011(Amendment to IEEE Std 802.16-2009), 2011.
[90] D. Halperin, W. Hu, A. Sheth, D. Wetherall, 802.11 with multiple antennas for dummies, SIGCOMM Comput. Commun. Rev. 40 (2010) 19–25 (online). <http://doi.acm.org/10.1145/1672308.1672313>.

[91] E. Perahia, R. Stacey, Next Generation Wireless LANs: Throughput, Robustness, and Reliability in 802.11n, Cambridge University Press, 2008.
[92] IEEE802.11n, IEEE Standard for Information technology—Telecommunications information exchange between systems—Local and metropolitan area networks—Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput, IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009), 2009.

Biomedical applications

[93] T. Szabo, Diagnostic Ultrasound Imaging: Inside Out, Elsevier Academic Press, 2004.
[94] B. Steinberg, Digital beamforming in ultrasound, IEEE Trans. Ultrasonics, Ferr. Freq. Cont. 39 (6) (1992) 716–721.
[95] J. Lu, H. Zou, J. Greenleaf, Biomedical ultrasound beamforming, Ultrasound Med. Biol. 20 (5) (1994) 403–428.
[96] J. Quistgaard, Signal acquisition and processing in medical diagnostic ultrasound, IEEE Signal Process. Mag. 14 (1) (1997) 67–74.
[97] Z. Wang, J. Li, R. Wu, Time-delay- and time-reversal-based robust Capon beamformers for ultrasound imaging, IEEE Trans. Med. Imag. 24 (2005) 1308–1322.
[98] S. Sanei, J. Chambers, EEG Signal Processing, Wiley & Sons, Ltd., West Sussex, England, 2007.
[99] P. Nunez, R. Srinivasan, Electric Fields of the Brain: The Neurophysis of EEG, Oxford University Press, 2006.
[100] T. Handy (Ed.), Brain Signal Analysis, Advances in Neuroelectric and Neuromagnetic Methods, MIT Press, 2009.
[101] Y. Salu, L.G. Cohen, D. Rose, S. Sato, C. Kufta, M. Hallett, An improved method for localizing electric brain dipoles, IEEE Trans. Biomed. Eng. 37 (7) (1990) 699–705.
[102] J.C. de Munck, The estimation of time varying dipoles on the basis of evoked potentials, Electroencephalogr. Clin. Neurophysiol. 77 (1990) 156–160.
[103] J. Mosher, P. Lewis, R. Leahy, Multiple dipole modeling and localization from spatio-temporal MEG data, IEEE Trans. Biomed. Eng. 39 (6) (1992) 541–557.
[104] J.W. Phillips, R.M. Leahy, J.C. Mosher, MEG-based imaging of focal neuronal current sources, IEEE Trans. Med. Imag. 16 (3) (1997) 338–348.
[105] B. van Veen, W. van Drongelen, M. Yuchtman, A. Suzuki, Localization of brain electrical activity via linearly constrained minimum variance spatial filtering, IEEE Trans. Biomed. Eng. 44 (9) (1997) 867–880.
[106] B. Lutkenhoner, Dipole source localization by means of maximum likelihood estimation: theory and simulations, Electroencephalogr. Clin. Neurophysiol. 106 (1998) 314–321.
[107] B. Cuffin, EEG dipole source localization, IEEE Eng. Med. Bio. Mag. 17 (5) (1998) 118–122.
[108] Z. Koles, Trends in EEG source localizatoin, Electroencephalogr. Clin. Neurophysiol. 106 (1998) 127–137.
[109] L. Zhukov, D. Weinstein, C. Johnson, Independent component analysis for EEG source localization, IEEE Eng. Med. Bio. Mag. 19 (3) (2000) 87–96.
[110] D. Yao, Electric potential produced by a dipole in a homogeneous conducting sphere, IEEE Trans. Biomed. Eng. 47 (7) (2000) 964–966.
[111] S. Baillet, J.C. Mosher, R.M. Leahy, Electromagnetic brain mapping, IEEE Signal Process. Mag. 18 (6) (2001) 14–30.

[112] J.C. de Munck, H.M. Huizenga, L.J. Waldorp, R.A. Heethaar, Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise, IEEE Trans. Signal Process. 50 (7) (2002) 1565–1572.
[113] C. Michela, M. Murraya, G. Lantza, S. Gonzaleza, L. Spinellib, R.G. de Peralta, EEG source imaging, Clin. Neurophysiol. 115 (2004) 2195–2222.
[114] K. Sekihara, K. Hild, S. Nagarajan, A novel adaptive beamformer for MEG source reconstruction effective when large background brain activities exist, IEEE Trans. Biomed. Eng. 53 (9) (2006) 1755–1764.
[115] T. Ferree, P. Nunez, Primer on electroencephalography for functional connectivity, in: V. Jirsa, A. McIntosh (Eds.), Handbook of Brain Connectivity, Springer-Verlag, 2007, pp. 169–200.
[116] K. Sekihara, K. Hild, S.S. Dalal, S. Nagarajan, Performance of prewhitening beamforming in MEG dual experimental conditions, IEEE Trans. Biomed. Eng. 55 (3) (2008) 1112–1121.
[117] S.C. Wu, A.L. Swindlehurst, P.T. Wang, Z. Nenadic, Projection vs. prewhitening for EEG interference suppression, IEEE Trans. Biomed. Eng. 59 (5) (2012) 1329–1338.
[118] S.C. Wu, A.L. Swindlehurst, P.T. Wang, Z. Nenadic, Efficient dipole parameter estimation in EEG systems with near-ML performance, IEEE Trans. Biomed. Eng. 59 (5) (2012) 1339–1348.
[119] S. Gibson, J.W. Judy, D. Markovic, Spike sorting: the first step in decoding the brain, IEEE Signal Process. Mag. 29 (1) (2012) 124–143.
[120] C.M. Gray, P.E. Maldonado, M. Wilson, B. McNaughton, Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex, J. Neurosci. Methods 63 (1995) 43–54.
[121] S. Takahashi, Y. Sakurai, M. Tsukada, Y. Anzai, Classifcation of neuronal activities from tetrode recordings using independent component analysis, Neurocomputing 49 (2002) 289–298.
[122] S. Takahashi, Y. Anzai, Y. Sakurai, A new approach to spike sorting for multi-neuronal activities recorded with a tetrode: how ICA can be practical, Neurosci. Res. 46 (2003) 265–272.
[123] M.I. Chelaru, M.S. Jog, Spike source localization with tetrodes, J. Neurosci. Meth. 142 (2005) 305–315.
[124] S. Micera, L. Citi, J. Rigosa, J. Carpaneto, S. Raspopovic, G.D. Pino, L. Rossini, K. Yoshida, L. Denaro, P. Dario, P.M. Rossini, Decoding information from neural signals recorded using intraneural electrodes: toward the development of a neurocontrolled hand prosthesis, Proc. IEEE 98 (3) (2010) 407–417.

Sonar

[125] W.C. Knight, R.G. Pridham, S.M. Kay, Digital signal processing for sonar, IEEE Proc. 69 (11) (1981) 1451–1506.
[126] R.J. Urick, Principles of Underwater Sound, third ed., John Wiley & Sons, West Sussex, England, 1983.
[127] R.G. Fizell, S.C. Wales, Source localization in range and depth in an arctic environment, J. Acoust. Soc. Am. Suppl. 78 (1985) p. S57.
[128] N.L. Owsley, Array Signal Processing, Prentice Hall, 1985.
[129] L. Brekhovskikh, Y. Lysanov, Fundamentals of Ocean Acoustics, Springer-Verlag, New York, 1991.
[130] Special issue on detection and estimation in matched—field processing, IEEE J. Ocean. Eng. 18 (3) (1993) 156–270.
[131] A.B. Baggeroer, W.A. Kuperman, P.N. Mikhalevsky, An overview of matched field methods in ocean acoustics, IEEE J. Ocean. Eng. 18 (4) (1993) 401–424.
[132] A. Tolstoy, Matched Field Processing for Underwater Acoustics, World Scientific, Singapore, 1993.
[133] A. Nehorai, E. Paldi, Acoustic vector-sensor array processing, IEEE Trans. Signal Process. 42 (9) (1994) 2481–2491.
[134] D.W. Tufts, J.P. Ianniello, I. Lourtie, J.C. Preisig, J.M.F. Moura, The past, present, and future of underwater acoustic signal processing, IEEE Signal Process. Mag. 15 (4) (1998) 21–51.

[135] M. Hawkes, A. Nehorai, Effects of sensor placement on acoustic vector-sensor array performance, IEEE J. Oceanic Eng. 24 (1999) 33–40.
[136] A.D. Waite, Sonar for Practising Engineers, third ed., West Sussex, England, 2002.
[137] M. Hawkes, A. Nehorai, Wideband source localization using a distributed acoustic vector-sensor array, IEEE Trans. Signal Process. 51 (6) (2003) 1479–1491.
[138] W. Xu, A.B. Baggeroer, C.D. Richmond, Bayesian bounds for matched—field parameter estimation, IEEE Trans. Signal Process. 52 (12) (2004) 3293–3305.

Microphone arrays

[139] J. Flanagan, J. Johnston, R. Zahn, G. Elko, Computer-steered microphone arrays for sound transduction in large rooms, J. Acoust. Soc. Am. 78 (5) (1985) 1508–1518.
[140] Y. Kaneda, J. Ohga, Adaptive microphone-array system for noise reduction, IEEE Trans. Acoust. Speech Signal Process. ASSP-34 (6) (1986) 1391–1400.
[141] Y. Grenier, A microphone array for car environments, Speech Commun. 12 (1993) 25–39.
[142] B. Ferguson, B. Quinn, Application of the short-time Fourier transform and the Wigner-Ville distribution to the acoustic localization of aircraft, J. Acoust. Soc. Am. 96 (1994) 821–827.
[143] M. Hoffman, K. Buckley, Robust time-domain processing of broadband microphone array data, IEEE Trans. Speech Audio Process. 3 (3) (1995) 193–203.
[144] S. Fischer, K. Simmer, Beamforming microphone arrays for speech acquisition in noisy environments, Speech Commun. 20 (1996) 215–227.
[145] G. Elko, Microphone array systems for hands-free telecommunication, Speech Commun. 20 (1996) 229–240.
[146] S. Affes, Y. Grenier, A signal subspace tracking algorithm for microphone array processing of speech, IEEE Trans. Speech Audio Process. 5 (5) (1997) 425–437.
[147] B. Ferguson, Time-delay estimation techniques applied to the acoustic detection of jet aircraft transits, J. Acoust. Soc. Am. 106 (1) (1999) 255–264.
[148] M. Dahl, I. Claesson, Acoustic noise and echo canceling with microphone array, IEEE Trans. Veh. Tech. 48 (5) (1999) 1518–1526.
[149] J. Benesty, Adaptive eigenvalue decomposition algorithm for passive acoustic source localization, J. Acoust. Soc. Am. 107 (1) (2000) 384–391.
[150] Y. Huang, J. Benesty, G. Elko, M. Mersereau, Real-time passive source localization: a practical linear-correction least-squares approach, IEEE Trans. Speech Audio Process. 9 (8) (2001) 943–956.
[151] J. Chen, L. Yip, J. Elson, H. Wang, D. Maniezzo, R. Hudson, K. Yao, D. Estrin, Coherent acoustic array processing and localization on wireless sensor networks, Proc. IEEE 91 (8) (2003) 1154–1162.
[152] T. Gustafsson, B. Rao, M. Trivedi, Source localization in reverberant environments: modeling and statistical analysis, IEEE Trans. Speech Audio Process. 11 (6) (2003) 791–803.
[153] R. Kozick, B. Sadler, Source localization with distributed sensor arrays and partial spatial coherence, IEEE Trans. Signal Process. 52 (3) (2004) 601–616.
[154] Z. Li, R. Duraiswami, Flexible and optimal design of spherical microphone arrays for beamforming, IEEE Trans. Audio Speech Lang. Process. 15 (2) (2007) 702–714.
[155] J. Benesty, J. Chen, Y. Huang, J. Dmochowski, On microphone-array beamforming from a MIMO acoustic signal processing perspective, IEEE Trans. Audio Speech Lang. Process. 15 (3) (2007) 1053–1065.
[156] X. Zhao, Z. Ou, Closely coupled array processing and model-based compensation for microphone array speech recognition, IEEE Trans. Audio Speech Lang. Process. 15 (3) (2007) 1114–1122.
[157] J. Benesty, J, Chen, Y. Huang, Microphone Array Signal Processing, Springer Verlag, 2008.
[158] M. Brandstein, D. Ward (Eds.), Microphone Arrays—Signal Processing Techniques and Applications, Springer Verlag, 2010.

Chemical sensor arrays

[159] A. Nehorai, B. Porat, E. Paldi, Detection and localization of vapor-emitting sources, IEEE Trans. Signal Process. 43 (1) (1995) 243–253.

[160] A. Gershman, V. Turchin, Nonwave field processing using sensor array approach, Signal Process. 44 (1995) 197–210.

[161] B. Porat, A. Nehorai, Localizing vapor-emitting sources by moving sensors, IEEE Trans. Signal Process. 44 (4) (1996) 1018–1021.

[162] A. Jerémic, A. Nehorai, Design of chemical sensor arrays for monitoring disposal sites on the ocean floor, IEEE J. Ocean. Eng. 23 (4) (1998) 334–343.

[163] Y. Nievergelt, Solution to an inverse problem in diffusion, SIAM Rev. 40 (1) (1998) 74–80.

[164] A. Jerémic, A. Nehorai, Landmine detection and localization using chemical sensor array processing, IEEE Trans. Signal Process. 48 (5) (2000) 1295–1305.

[165] J. Matthes, L. Gröll, H. Keller, Source localization based on pointwise concentration measurements, Sensor. Actuat. A: Phys. 115 (2004) 32–37.

[166] J. Matthes, L. Gröll, H. Keller, Source localization by spatially distributed electronic noses for advection and diffusion, IEEE Trans. Signal Process. 53 (5) (2005) 1711–1719.

[167] T. Zhao, A. Nehorai, Detecting and estimating biochemical dispersion of a moving source in a semi-infinite medium, IEEE Trans. Signal Process. 54 (6) (2006) 2213–2225.

[168] S. Vijayakumaran, Y. Levinbook, T. Wong, Maximum likelihood localization of a diffusive point source using binary observations, IEEE Trans. Signal Process. 55 (2) (2007) 665–676.

[169] M. Ortner, A. Nehorai, A. Jerémic, Biochemical transport modeling and Bayesian source estimation in realistic environments, IEEE Trans. Signal Process. 55 (6) (2007) 2520–2532.