ANTENNA ARRAYS FOR MOBILE COMMUNICATIONS

Wireless cellular networks are growing rapidly around the world, and this trend is likely to continue for several years. The progress in radio technology enables new and improved services. Current wireless services include transmission of voice, fax, and low-speed data. More bandwidth-consuming interactive multimedia services such as video-on-demand and Internet access will be supported in the future. Wireless networks must provide these services in a wide range of environments, spanning dense urban, suburban, and rural areas. Varying mobility needs must also be addressed. Wireless local loop networks serve fixed subscribers. Microcellular networks serve pedestrians and other slow-moving users, and macrocellular networks serve high-speed vehicle-borne users. Several competing standards have been developed for terrestrial networks. AMPS (advanced mobile phone system) is an example of a first-generation frequency division multiple-access analog cellular system. Second-generation standards include GSM (global system for mobile) and IS-136, using time division multiple access (TDMA); and IS-95, using code division multiple access (CDMA). IMT-2000 is proposed to be the third-generation standard and will use mostly a wideband CDMA technology.

Increased services and lower costs have resulted in an increased air time usage and number of subscribers. Since the radio (spectral) resources are limited, system capacity is a primary challenge for current wireless network designers. Other major challenges include (1) an unfriendly transmission medium, with multipath transmission, noise, interference, and time variations, (2) the limited battery life of the user's handheld terminal, and (3) efficient radio resource management to offer high quality of service.

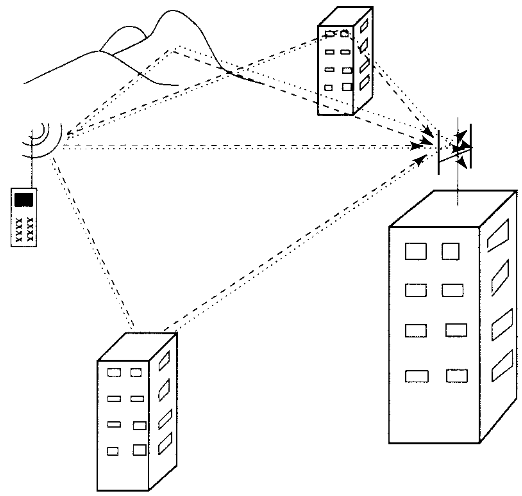

Figure 1. The user signal experiences multipath propagation and impinges on a two-element array on a building rooftop.

Current wireless modems use signal processing in the time dimension alone through advanced coding, modulation, and equalization techniques. The primary goal of smart antennas in wireless communications is to integrate and exploit efficiently the extra dimension offered by multiple antennas at the transceiver in order to enhance the overall performance of the network. Smart antenna systems use modems that combine the signals of multielement antennas in both space and time. Smart antennas can be used for both receive and transmit, both at the base station and at the user terminal. The use of smart antennas at the base alone is more typical, since practical constraints usually limit the use of multiple antennas at the terminal. See Fig. 1 for an illustration.

Space–time processing offers various advantages. The first is array gain: multiple antennas capture more signal energy, which can be combined to improve the signal-to-noise ratio (SNR). Next, spatial diversity obtained from multiple antennas can be used to combat channel fading. Finally, space–time processing can help mitigate intersymbol interference (ISI) and cochannel interference (CCI). These leverages can be traded for improvements in:

- Coverage: square miles per base station
- Quality: bit error rate (BER); outage probability
- Capacity: erlangs per hertz per base station
- Data rates: bits per second per hertz per base station

EARLY FORMS OF SPATIAL PROCESSING

Adaptive Antennas

The use of adaptive antennas dates back to the 1950s with their applications to radar and antijam problems. The primary goal of adaptive antennas is the automatic generation of beams (beamforming) that track a desired signal and possibly reject (or *null*) interfering sources through linear combining of the signals captured by the different antennas. An early contribution in the field of beamforming was made in 1956 by Altman and Sichak, who proposed a combining device based on a phase-locked loop. This work was later refined in order to incorporate the adjustment of antenna signals in both phase and gain, allowing improved performance of the receiver in the presence of strong jammers. Howells proposed the sidelobe canceler for adaptive nulling. Optimal combining schemes were also introduced in order to minimize different criteria at the beamformer output. These include the minimum mean squared error (MMSE) criterion, as in the LMS algorithm proposed by Widrow; the signal-to-interference-and-noise ratio (SINR) criterion proposed by Applebaum, and the minimum-variance beamformer distortionless response (MVDR) beamformer proposed by Capon. Further advances in the field were made by Frost, Griffiths, and Jim among several others. A list of references in beamforming can be found in Refs. 1 and 13.

Besides beamforming, another application of antenna arrays is direction-of-arrival (DOA) estimation for source or target localization purposes. The leading DOA estimation methods are the MUSIC and ESPRIT algorithms (2). In many of the beamforming techniques (for instance in Capon's method), the estimation of the source direction is an essential step. DOA estimation is still an area of active research.

Antenna arrays for beamforming and source localization are of course of great interest in military applications. However, their use in civilian cellular communication networks is now gaining increasing attention. By enabling the transmission and reception of signal energy from selected directions, beamformers play an important role in improving the performance of both the base-to-mobile (forward) and mobile-to-base (reverse) links.

Antenna Diversity

Antenna diversity can alleviate the effects of channel fading, and is used extensively in wireless networks. The basic idea of space diversity is as follows: if several replicas of the same information-carrying signal are received over multiple branches with comparable strengths and exhibit independent fading, then there is a high probability that at least one branch will not be in a fade at any given instant of time. When a receiver is equipped with two or more antennas that are sufficiently separated (typically several wavelengths), they offer useful diversity branches. Diversity branches tend to fade independently; therefore, a proper selection or combining of the branches increases link reliability. Without diversity, protection against deep channel fades requires higher transmit power to ensure the link margins. Therefore, diversity at the base can be traded for reduced power consumption and longer battery life at the user terminal. Also, lower transmit power decreases the amount of co-channel-user interference and increases the system capacity.

Independent fading across antennas is achievable when radio waves impinge on the antenna array with sufficient angle spread. Paths coming from different arriving directions will add differently (constructive or destructive manner) at each antenna. This requires the presence of significant scatterers in the propagation medium, such as in urban or hilly terrain.

Diversity also helps to combat large-scale fading effects caused by shadowing from large obstacles (e.g., buildings or terrain features). However, antennas located in the same base station experience the same shadowing. Instead, antennas from different base stations can be combined to offer a protection against such fading (*macro* diversity).

Antenna diversity can be complemented by other forms of diversity. Polarization, time, frequency, and path diversity are some examples. These are particularly useful when physical constraints prevent the use of multiple antennas (for instance at the hand-held terminal). See Ref. 3 for more details.

Combining the different diversity branches is an important issue. The main options used in current systems are briefly described below. In all cases, independent branch fading and equal mean branch powers are assumed. However, in nonideal situations, branch correlation and unequal powers will result in a loss of diversity gain. A correlation coefficient as high as 0.7 between instantaneous branch envelope levels is considered acceptable.

Selection Diversity. Selection diversity is one of the simplest form of diversity combining. Given several branches with varying carrier-to-noise ratios (C/N), selection diversity consists in choosing the branch having the highest instantaneous C/N. The performance improvement from selection diversity is evaluated as follows: Let us suppose that M branches experience independent fading but have the same mean C/N, denoted by Γ. Let us now denote by Γ_s the mean C/N of the selected branch. Then it can be shown that (4)

$$\Gamma_s = \Gamma \sum_{j=1}^{M} \frac{1}{j}$$

For instance, selection over two branches increases the mean C/N by a factor of 1.5. More importantly, the statistics of the instantaneous C/N is improved. Note that selection diversity requires a receiver behind each antenna.

Switching diversity is a variant of selection diversity. In this method, a selected branch is held until it falls below a threshold T, at which point the receiver switches to another branch, regardless of its level. The threshold can be fixed or adaptive. This strategy performs almost as well as the selection method described above, and it reduces the system cost, since only one receiver is required.

Maximum-Ratio Combining. Maximum-ratio combining (MRC) is an optimal combining approach to combat fading. The signals from M branches are first cophased to mutual coherence and then summed after weighting. The weights are chosen to be proportional to the signal level to maximize the combined C/N. It can be shown that the gain from MRC in mean C/N is directly proportional to the number of branches:

$$\Gamma_s = M\Gamma$$

Equal-Gain Combining. Although optimal, MRC is expensive to implement. Also, MRC requires accurate tracking of the complex fading, which is difficult to achieve in practice. A simpler alternative is given by equal-gain combining, which consists in summing the cophased signals using unit weights.

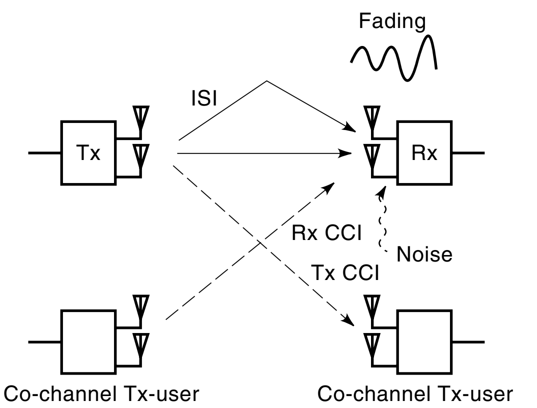

Figure 2. Smart antennas help mitigate the effects of cellular radio propagation.

The performance of equal-gain combining is found to be very close to that of MRC. The SNR of the combined signals using equal gain is only 1 dB below the SNR provided by MRC (4).

EMERGING APPLICATION OF SPACE–TIME PROCESSING

While the use of beamforming and space diversity proves useful in radio communication applications, an inherent limitation of these techniques lies in the fact that they exploit signal combining in the space dimension only. Directional beamforming, in particular, heavily relies on the exploitation of the spatial signatures of the incoming signals but does not consider their temporal structure. The techniques that combine the signals in both time and space can bring new advantages, and their importance in the area of mobile communications is now recognized (5).

The main reason for using space–time processing is that it can exploit the rich temporal structure of digital communication signals. In addition, multipath propagation environments introduce signal delay spread, making techniques that exploit the complete space–time structure more natural.

The typical structure of a space–time processing device consists of a bank of linear filters, each located behind a branch, followed by a summing network. The received space–time signals can also be processed using nonlinear schemes, for example, maximum-likelihood sequence detection. The space–time receivers can be optimized to maximize array and diversity gains, and to minimize:

1. Intersymbol interference (ISI), induced by the delay spread in the propagation channel. ISI can be suppressed by selecting a space–time filter that equalizes the channel or by using a maximum-likelihood sequence detector.
2. Co-channel-user interference (CCI), coming from neighboring cells operating at the same frequency. CCI is suppressed by using a space–time filter that is orthogonal to the interference's channel. The key point is that CCI that cannot be rejected by space-only filtering may be handled more effectively using space–time filtering.

Smart antennas can also be used at the transmitter to maximize array gain and/or diversity, and to mitigate ISI and CCI. In the transmit case, however, the efficiency of space–time processing schemes is usually limited by the lack of accurate channel information.

The major effects induced by radio propagation in a cellular environment are pictured in Fig. 2. The advantages offered by space–time processing for receive and transmit are summarized in Table 1.

In the following sections, we describe channel models and algorithms used in space–time processing. Both simple and advanced solutions are presented, and tradeoffs highlighted. Finally, we describe current applications of smart antennas.

CHANNEL MODELS

Channel models capture radio propagation effects and are useful for simulation studies and performance prediction. Channel models also help in motivating appropriate signal-processing algorithms. The effects of radio propagation on the transmitted signal can be broadly categorized into two main classes: fading and spreading.

Fading refers to the propagation losses experienced by the radio signal (on both the forward and reverse links). One type of fading, called *selective fading,* causes the received signal level to vary around the average level in some regions of space, frequency, or time. *Channel spreading* refers to the spreading of the information-carrying signal energy in space, and on the time or frequency axis. Selective fading and spreading are complementary phenomena.

Channel Fading

Mean Path Loss. The mean path loss describes the attenuation of a radio signal in free-space propagation, due to isotropic power spreading, and is given by the well-known inverse square law:

$$P_{\mathrm{r}} = P_{\mathrm{t}} \left(\frac{\lambda}{4\pi d}\right)^2 G_{\mathrm{t}} G_{\mathrm{r}}$$

where P_{r} and P_{t} are the received and transmitted powers, λ is the radio wavelength, d is the range, and G_{t}, G_{r} are the gains of the transmit and receive one-element antennas respectively. In cellular environments, the main path is often accompanied by a surface-reflected path which may interfere destructively with the primary path. Specific models have been developed that consider this effect. The path loss model becomes (4)

Table 1. Advantages of Space–Time Processing

For transmit (Tx)	Reduces Tx CCI Maximizes Tx diversity Reduces ISI Increases Tx EIRP
For receive (Rx)	Reduces Rx CCI Maximizes Rx diversity Eliminates ISI Increases C/N

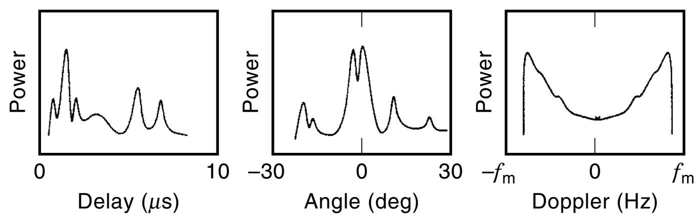

Figure 3. The radio channel induces spreading in several dimensions. These spreads strongly affect the design of the space–time receiver.

$$P_r = P_t \left(\frac{h_r h_r}{d^2}\right)^2 G_t G_r$$

where h_t, h_r are the effective heights of the transmit and receive antennas, respectively. Note that this particular path loss model follows an inverse fourth-power law. In fact, depending on the environment, the path loss exponent may vary from 2.5 to 5.

Slow Fading. Slow fading is caused by long-term shadowing effects of buildings or natural features in the terrain. It can also be described as the local mean of a fast fading signal (see below). The statistical distribution of the local mean has been studied experimentally and shown to be influenced by the antenna height, the operating frequency, and the type of environment. It is therefore difficult to predict. However, it has been observed that when all the above-mentioned parameters are fixed, then the received signal fluctuation approaches a normal distribution when plotted on a logarithmic scale (i.e., in decibels) (4). Such a distribution is called lognormal. A typical value for the standard deviation of shadowing distribution is 8 dB.

Fast Fading. The multipath propagation of the radio signal causes path signals to add up with random phases, constructively or destructively, at the receiver. These phases are determined by the path length and the carrier frequency, and can vary extremely rapidly along with the receiver location. This gives rise to *fast fading:* large, rapid fluctuations of the received signal level in space. If we assume that a large number of scattered wavefronts with random amplitudes and angles of arrival arrive at the receiver with phases uniformly distributed in $[0, 2\pi)$, then the in-phase and quadrature phase components of the vertical electrical field E_z can be shown to be Gaussian processes (4). In turn, the envelope of the signal can be well approximated by a Rayleigh process. If there is a direct path present, then it will no longer be a Rayleigh distribution but becomes a Rician distributed instead.

Channel Spreading

Propagation to or from a mobile user, in a multipath channel, causes the received signal energy to spread in the frequency, time, and space dimensions (see Fig. 3, and also Table 2 for typical values). The characteristics of the spreading [that is to say, the particular dimension(s) in which the signal is spread] affects the design of the receiver.

Doppler Spread. When the mobile user is in motion, the radio signal at the receiver experiences a shift in the frequency domain (called the Doppler shift), the amplitude of which depends on the path direction of arrival. In the presence of surrounding scatterers with multiple directions, a pure tone is *spread* over a finite spectral bandwidth. In this case, the Doppler power spectrum is defined as the Fourier transform of the time autocorrelation of the received signal, and the Doppler spread is the support of the Doppler power spectrum. Assuming scatterers uniformly distributed in angle, the Doppler power spectrum is given by the so-called classical spectrum:

$$S(f) = \frac{3\sigma^2}{2\pi f_m}\left[1 - \left(\frac{f - f_c}{f_m}\right)^2\right]^{-1/2}, \qquad f_c - f_m < f < f_c + f_m$$

where $f_m = v/\lambda$ is the maximum Doppler shift, v is the mobile velocity, f_c is the carrier frequency, and σ^2 is the signal variance. When there is a dominant source of energy coming from a particular direction (as in line-of-sight situations), the expression for the spectrum needs to be corrected according to the Doppler shift of the dominant path f_D:

$$S(f) + B\delta(f - f_D)$$

where B denotes the ratio of direct to scattered path energy.

The Doppler spread causes the channel characteristics to change rapidly in time, giving rise to the so-called *time selectivity.* The coherence time, during which the fading channel can be considered as constant, is inversely proportional to the Doppler spread. A typical value of the Doppler spread in a macrocell environment is about 200 Hz at 30 m/s (65 mi/h) in the 1900 MHz band. A large Doppler spread makes good channel tracking an essential feature of the receiver design.

Delay Spread. Multipath propagation is often characterized by several versions of the transmitted signal arriving at the receiver with different attenuation factors and delays. The spreading in the time domain is called *delay spread* and is responsible for the selectivity of the channel in the frequency domain (different spectral components of the signal carry different powers). The coherence bandwidth, which is the maximum range of frequencies over which the channel response can be viewed as constant, is inversely proportional to the delay spread. Significant delay spread may cause strong intersymbol interference, which makes necessary the use of a channel equalizer.

Angle Spread. Angle spread at the receiver refers to the spread of directions of arrival of the incoming paths. Like-

Table 2. Typical Delay, Angle, and Doppler Spreads in Cellular Radio Systems

Environment	Delay Spread (μs)	Angle Spread (deg)	Doppler Spread (Hz)
Flat rural (macro)	0.5	1	190
Urban (macro)	5	20	120
Hilly (macro)	20	30	190
Micro cell (mall)	0.3	120	10
Pico cell (indoors)	0.1	360	5

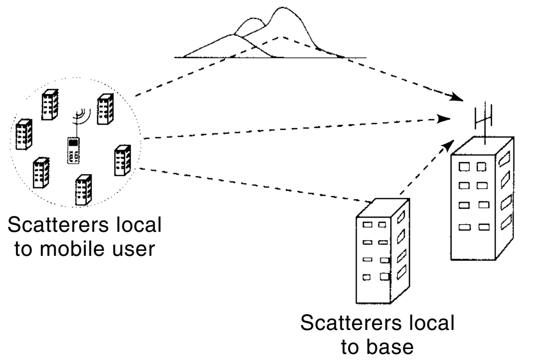

Figure 4. Each type of scatterer introduces specific channel spreading characteristics.

wise, angle spread at the transmitter refers to the spread of departure angles of the paths. As mentioned earlier, a large angle spread will cause the paths to add up in a random manner at the receiver as the location of the receive antenna varies; hence it will be source of *space*-selective fading. The range of space for which the fading remains constant is called the coherence distance and is inversely related to the angle spread. As a result, two antennas spaced by more than the coherence distance tend to experience uncorrelated fading. When the angle spread is large, which is usually the case in dense urban environments, a significant gain can be obtained from space diversity. Note that this usually conflicts with the possibility of using directional beamforming, which typically requires well-defined and dominant signal directions, that is, a low angle spread.

Multipath Propagation

Macro Cells. A macro cell is characterized by a large cell radius (up to a few tens of kilometers) and a base station located above the rooftops. In macro-cell environments, the signal energy received at the base station comes from three main scattering sources: scatterers local to the mobile, remote dominant scatterers, and scatterers local to the base (see Fig. 4 for an illustration). The following description refers to the reverse link but applies to the forward link as well.

The scatterers local to the mobile user are those located a few tens of meters from the hand-held terminal. When the terminal is in motion, these scatterers give rise to a Doppler spread, which causes time-selective fading. Because of the small scattering radius, the paths that emerge from the vicinity of the mobile user and reach the base station show a small delay spread and a small angle spread.

Of the paths emerging from the local-to-mobile scatterers, some reach remote dominant scatterers, such as hills or high-rise buildings, before eventually traveling to the base station. These paths will typically reach the base with medium to large angle and delay spreads (depending of course on the number and locations of these remote scatterers).

Once these multiple wavefronts reach the vicinity of the base station, they usually are further scattered by local structures such as buildings or other structures that are close to the base. These scatterers local to the base can cause large angle spread; therefore they can cause severe space-selective fading.

Micro Cells. Micro cells are characterized by highly dense built-up areas, and by the user's terminal and base being relatively close (a few hundred meters). The base antenna has a low elevation and is typically below the rooftops, causing significant scattering in the vicinity of the base. Micro-cell situations make the propagation difficult to analyze, and the macro-cell model described earlier no longer can be expected to hold. Very high angle spreads along with small delay spreads are likely to occur in this situation. The Doppler spread can be as high as in macro cells, although the mobility of the user is expected to be limited, due to the presence of mobile scatterers.

Parametric Channel Model

A complete and accurate understanding of propagation effects in the radio channel requires a detailed description of the physical environment. The *specular model,* to be presented below, only provides a simplified description of the physical reality. However, it is useful, as it describes the main channel effects and it provides the means for a simple and efficient mathematical treatment. In this model, the multiple elementary paths are grouped according to a (typically small) number L of main path clusters, each of which contains paths that have roughly the same mean angle and delay. Since the paths in these clusters originate from different scatterers, the clusters typically have near-independent fading. Based on this model, the continuous-time channel response from a single transmit antenna to the ith antenna of the receiver can be written as

$$f_i(t) = \sum_{l=1}^{L} a_i(\theta_l)\alpha_l^{\mathrm{R}}(t)\delta(t-\tau_l) \tag{1}$$

where $\alpha_l^{\mathrm{R}}(t)$, θ_l, and τ_l are respectively the fading (including mean path loss and slow and fast fading), the angle, and the delay of the lth receive path cluster. Note that this model also includes the response of the ith antenna to a path from direction θ_l, denoted by $a_i(\theta_l)$. In the following we make use of the specular model to describe the structure of the signals in space and time. Note that in the situation where the path cluster assumption is not acceptable, other channel models, called *diffuse* channel models, are more appropriate (6).

DATA MODELS

This section focuses on developing signal models for space–time processing algorithms. The transmitted information signal is assumed to be linearly modulated. In the case of a nonlinear modulation scheme, such as the Gaussian minimum shift keying (GMSK) used in the GSM system, linear approximations are assumed to hold. The baseband equivalent of the transmitted signal can be written (7)

$$u(t) = \sum_k g(t-kT)s(k) + n(t) \tag{2}$$

where $s(k)$ is the symbol stream, with rate $1/T$, $g(t)$ is the pulse-shaping filter, and $n(t)$ is an additive thermal noise. Four configurations for the received signal (two for the reverse link and two for the forward link) are described below.

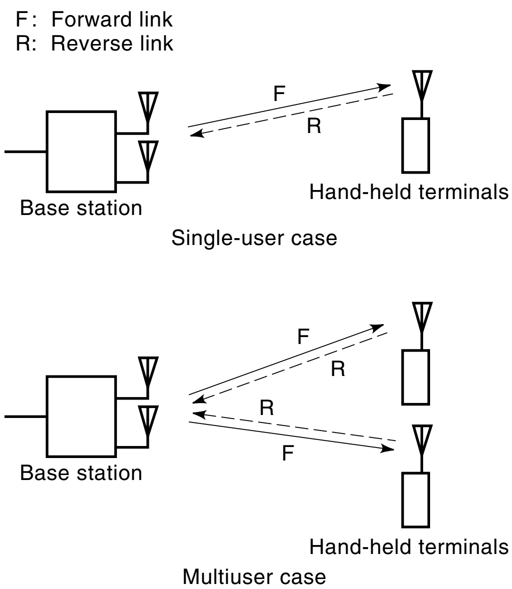

Figure 5. Several configurations are possible for antenna arrays, in transmit (T) and in receive (R).

These are also depicted in Fig. 5. In each case, one assumes $M > 1$ antennas at the base station and a single antenna at the mobile user.

Reverse Link

We consider the signal received at the base station. Since the receiver is equipped with M antennas, the received signal can be written as a vector $\boldsymbol{x}(t)$ with M entries.

Single-User Case. Let us assume a single user transmitting towards the base (no CCI). Using the specular channel model in Eq. (1), the received signal can be written as follows:

$$\boldsymbol{x}(t) = \sum_{l=1}^{L} \boldsymbol{a}(\theta_l)\alpha_l^{\mathrm{R}}(t)u(t-\tau_l) + \boldsymbol{n}(t) \tag{3}$$

where $\boldsymbol{a}(\theta_l) = (a_1(\theta_l), \ldots, a_M(\theta_l))^{\mathrm{T}}$ is the vector array response to a path of direction θ_l, and where $^{\mathrm{T}}$ refers to the transposition operator.

Multiuser Case. We now have Q users transmitting towards the base. The received signal is the following sum of contributions from the Q users, each of them carries a different set of fading, delays, and angles:

$$\boldsymbol{x}(t) = \sum_{q=1}^{Q}\sum_{l=1}^{L_q} \boldsymbol{a}(\theta_{lq})\alpha_{lq}^{\mathrm{R}}(t)u_q(t-\tau_{lq}) + \boldsymbol{n}(t) \tag{4}$$

where the subscript q refers to the user index.

Forward Link

Single-User Case. In this case, the base station uses a transmitter equipped with M antennas to send an information signal to a unique user. Therefore, space–time processing must be performed *before* the signal is launched into the channel. As will be emphasized later, this is a challenging situation, as the transmitter typically lacks reliable information on the channel.

For the sake of simplicity, we will assume here that a space-only beamforming weight vector $\boldsymbol{w}$ is used, as the extension to space–time beamforming is straightforward. The baseband signal received at the mobile station is scalar and is given by

$$x(t) = \sum_{l=1}^{L} \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\theta_l)\alpha_l^{\mathrm{F}}(t)u(t-\tau_l) + n(t) \tag{5}$$

where $\alpha_l^{\mathrm{F}}(t)$ is the fading coefficient of the lth transmit path in the forward link. Superscript $^{\mathrm{H}}$ denotes the transpose-conjugation operator. Note that path angles and delays remain theoretically unchanged in the forward and reverse links. This is in contrast with the fading coefficients, which depend on the carrier frequency. Frequency division duplex (FDD) systems use different carriers for the forward and reverse links, which result in $\alpha_l^{\mathrm{F}}(t)$ and $\alpha_l^{\mathrm{R}}(t)$ being nearly uncorrelated. In contrast, time division duplex (TDD) systems will experience almost identical forward and reverse fading coefficients in the forward and reverse links. Assuming however that the transmitter knows the forward fading and delay parameters, transmit beamforming can offer array gain, ISI suppression, and CCI suppression.

Multiuser Case. In the multiuser case, the base station wishes to communicate with Q users, simultaneously and in the same frequency band. This can be done by superposing, on each of the transmit antennas, the signals given by Q beamformers $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_Q$. At the mth user, the received signal waveform contains the signal sent to that user, plus an interference from signals intended for all other users. This gives

$$x_m(t) = \sum_{q=1}^{Q}\sum_{l=1}^{L_q} \boldsymbol{w}_q^{\mathrm{H}}\boldsymbol{a}(\theta_{lm})\alpha_{lm}^{\mathrm{F}}(t)u_q(t-\tau_{lm}) + n_m(t) \tag{6}$$

Note that each information signal $u_q(t)$ couples into the L_m paths of the mth user through the corresponding weight vector $\boldsymbol{w}_q$, for all q.

A Nonparametric Model

The data models above build on the parametric channel model developed earlier. However, there is also interest in considering the end-to-end channel impulse response of the system to a transmitted symbol rather than the physical path parameters. The channel impulse response includes the pulse-shaping filter response, the propagation phenomena, and the antenna response as well. One advantage of looking at the impulse response is that the effects of ISI and CCI can be described in a better and more compact way. A second advantage is that the nonparametric channel only relies on the channel linearity assumption.

We look at the reverse-link and single-user case only. Since a single scalar signal is transmitted and received over several branches, this corresponds to a single-input multiple-output

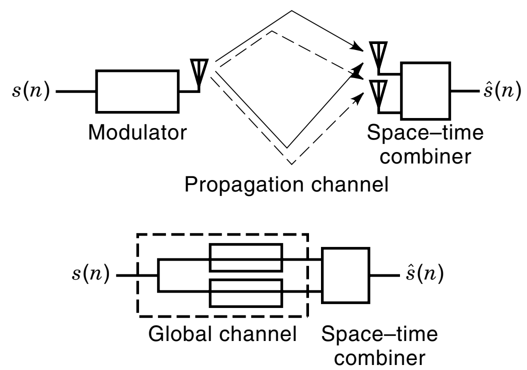

Figure 6. The source signal $s(n)$ can be seen as driving a single-input multiple-output filter with M outputs, where M is the number of receive antennas.

(SIMO) system, depicted in Fig. 6. The model below is also easily extended to multiuser channels. Let $\boldsymbol{h}(t)$ denote the $M \times 1$ global channel impulse response. The received vector signal is given by the result of a (noisy) convolution operation:

$$\boldsymbol{x}(t) = \sum_k \boldsymbol{h}(t - kT)s(k) + \boldsymbol{n}(t) \tag{7}$$

From Eqs. (2) and (3), the channel response may also be expressed in terms of the specular model parameters through

$$\boldsymbol{h}(t) = \sum_{l=1}^{L} \boldsymbol{a}(\theta_l)\alpha_l^{\mathrm{R}}(t)g(t - \tau_l) \tag{8}$$

Signal Sampling. Consider sampling the received signal at the baud (symbol) rate, that is, at $t_k = t_0 + kT$, where t_0 is an arbitrary phase. Let N be the maximum length of the channel response in symbol periods. Assuming that the channel is invariant for some finite period of time [i.e., $\alpha_l^{\mathrm{R}}(t) = \alpha_l^{\mathrm{R}}$], the received vector sample at time t_k can be written as

$$\boldsymbol{x}(k) = \boldsymbol{H}\boldsymbol{s}(k) + \boldsymbol{n}(k) \tag{9}$$

where $\boldsymbol{H}$ is the sampled channel matrix, with size $M \times N$, whose (i, j) term is given by

$$[\boldsymbol{H}]_{ij} = \sum_{l=1}^{L} a_i(\theta_l)\alpha_l^{\mathrm{R}} g(t_0 + jT - \tau_l)$$

and where $\boldsymbol{s}(k)$ is the vector of N ISI symbols at the time of the measurement:

$$\boldsymbol{s}(k) = (s(k), s(k-1), \ldots, s(k - N + 1))^{\mathrm{T}}$$

To allow for the presence of CCI, Eq. (9) can be generalized to

$$\boldsymbol{x}(k) = \sum_{q=1}^{Q} \boldsymbol{H}_q \boldsymbol{s}_q(k) + \boldsymbol{n}(k) \tag{10}$$

where Q denotes the number of users and q the user index. Most digital modems use sampling of the signal at a rate higher than the symbol rate (typically up to four times). Oversampling only increases the number of scalar observations per transmitted symbol, which can be regarded mathematically as increasing the number of channel components, in a way similar to increasing the number of antennas. Hence the model above also holds true when sampling at $T/2$, $T/3$, ... However, though mathematically equivalent, spatial oversampling and temporal oversampling lead to different signal properties.

Structure of the Linear Space–Time Beamformer

Space combining is now considered at the receive antenna array. Let $\boldsymbol{w}$ be a $M \times 1$ space-only weight vector (a single complex weight is assigned to each antenna). The output of the combiner, denoted by $y(k)$ is as follows:

$$y(k) = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{x}(k)$$

The resulting beamforming operation is depicted in Fig. 7. The generalization to space–time combining is straightforward: Let the combiner have m time taps. Each tap, denoted by $\boldsymbol{w}(i)$, $i = 0, \ldots, m - 1$, is an $M \times 1$ space weight vector defined as above. The output of the space–time beamformer is now written as

$$y(k) = \sum_{i=0}^{m-1} \boldsymbol{w}(i)^{\mathrm{H}}\boldsymbol{x}(k - i) \tag{11}$$

which can be reformulated as

$$y(k) = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{X}(k) \tag{12}$$

where $\boldsymbol{W} = (\boldsymbol{w}(0)^{\mathrm{H}}, \ldots, \boldsymbol{w}(m - 1)^{\mathrm{H}})^{\mathrm{H}}$ and $\boldsymbol{X}(k)$ is the data vector compactly defined as $\boldsymbol{X}(k) = (\boldsymbol{x}(k)^{\mathrm{H}}, \ldots, \boldsymbol{x}(k - m + 1)^{\mathrm{H}})^{\mathrm{H}}$.

ISI and CCI Suppression

The formulation above gives insight into the algebraic structure of the space–time received data vector. Also it allows us to identify the conditions under which the suppression of ISI and/or CCI is possible. Recalling the signal model in Eq. (9), the space–time vector $\boldsymbol{X}(k)$ can be in turn written as

$$\boldsymbol{X}(k) = \mathscr{H}\boldsymbol{S}(k) + \boldsymbol{N}(k) \tag{13}$$

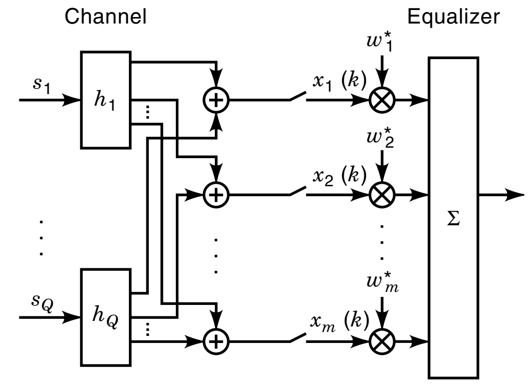

Figure 7. Structure of the spatial beamformer. The space–time beamformer is a direct generalization that combines in time the outputs of several spatial beamformers.

where $\boldsymbol{S}(k) = (s(k), s(k-1), \ldots, s(k-m+N+2))^{\mathrm{T}}$ and where

$$\mathscr{H} = \begin{pmatrix} \boxed{\boldsymbol{H}} & 0 & \cdots & 0 \\ 0 & \boxed{\boldsymbol{H}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boxed{\boldsymbol{H}} \end{pmatrix} \tag{14}$$

is a $mM \times (m + N - 1)$ channel matrix. The block Toeplitz structure in $\mathscr{H}$ stems from the linear time-invariant convolution operation with the symbol sequence.

Let us temporarily assume a noise-free scenario. Then the output of a linear space–time combiner can be described by the following equation:

$$y(k) = \boldsymbol{W}^{\mathrm{H}}\mathscr{H}\boldsymbol{S}(k) \tag{15}$$

In the presence of Q users transmitting towards the base station, the output of the space–time receiver is generalized to

$$y(k) = \sum_{q=1}^{Q} \boldsymbol{W}^{\mathrm{H}}\mathscr{H}_q\boldsymbol{S}_q(k) \tag{16}$$

ISI Suppression. The purpose of equalization is to compensate for the effects of ISI induced by the user's channel in the absence of CCI. Tutorial information on equalization can be found in Refs. 7, and 8. In general, a linear filter $\boldsymbol{W}_q$ is an equalizer for the channel of the qth user if the convolution product between $\boldsymbol{W}_q$ and the channel responses yields a Dirac function, that is, if $\boldsymbol{W}_q$ satisfies the following so-called zero-forcing condition:

$$\boldsymbol{W}_q^{\mathrm{H}}\mathscr{H}_q = (0, \ldots, 0, 1, 0, \ldots, 0) \tag{17}$$

Here, the location of the 1 element represents the delay of the combined channel–equalizer impulse response. Note that from an algebraic point of view, the channel matrix $\mathscr{H}_q$ should have more rows than columns for such solutions to exist: $mM \geq m + N - 1$. Therefore, it is essential to have enough degrees of freedom (number of taps in the filter) to allow for ISI suppression. Note that zero-forcing solutions can be obtained using temporal oversampling at the receive antenna only, since oversampling by a factor of M provides us theoretically with M baud-rate branches. However, having multiple antennas at the receiver plays a important role in improving the conditioning of the matrix $\mathscr{H}$, which in turn will improve the robustness of the resulting equalizer in the presence of noise. It can be shown that the condition number of the matrixe $\mathscr{H}_q$ is related to some measure of the correlation between the entries of $\boldsymbol{h}(t)$. Hence, a significant antenna spacing is required to provide the receiver with sufficiently decorrelated branches.

CCI Suppression. The purpose of CCI suppression in a multiple access network is to isolate the contribution of one desired user by rejecting that of others. One way to achieve this goal is to enforce *orthogonality* between the response of the space–time beamformer and the response of the channel of the users to be rejected. In other words, in order to isolate the signal of user q using $\boldsymbol{W}_q$, the following conditions must be satisfied (possibly approximately):

$$\boldsymbol{W}_q^{\mathrm{H}}\mathscr{H}_f = 0 \qquad \text{for all} \quad f \neq q \in [1, \ldots, Q] \tag{18}$$

If we assume that all the channels have the same maximum order N, Eq. (18) provides as many as $(Q-1)(m + N - 1)$ scalar equations. The number of unknowns is again given by mM (the size of $\boldsymbol{W}_q$). Hence a receiver equipped with multiple antennas is able to provide the number of degrees of freedom necessary for signal separation. This requires $mM \geq (Q-1)(m + N - 1)$. At the same time, it is desirable that the receiver capture a significant amount of energy from the desired user; hence an extra condition on $\boldsymbol{W}_q$ should be $\boldsymbol{W}_q^{\mathrm{H}}\mathscr{H}_q \neq 0$. From an algebraic perspective, this last condition requires that $\mathscr{H}_q$ and $\{\mathscr{H}_f\}_{f\neq q}$ should not have the same column subspaces. The required subspace misalignment between the desired user and the interferers in space–time processing is a generalization of the condition that signal and interference should not have the same direction, needed for interference nulling using beamforming.

Joint ISI and CCI Suppression. The complete recovery of the signal transmitted by one desired user in the presence of ISI and CCI requires both channel equalization and separation. A space–time beamformer is an exact solution to this problem if it satisfies both Eq. (17) and Eq. (18), which can be further written as

$$\boldsymbol{W}_q^{\mathrm{H}}(\mathscr{H}_1, \ldots, \mathscr{H}_{q-1}, \mathscr{H}_q, \mathscr{H}_{q+1}, \ldots, \mathscr{H}_Q) = (0, \ldots, 0, 1, 0, \ldots, 0) \tag{19}$$

where the location of the 1 element designates both the index of the user of reference and the reconstruction delay. The existence of solutions to this problem requires the multiuser channel matrix $\mathscr{H}_* \stackrel{\text{def}}{=} (\mathscr{H}_1, \ldots, \mathscr{H}_Q)$ to have more rows than columns: $mM \leq Q(m + N - 1)$. Here again, smart antennas play a critical role in offering a sufficient number of degrees of freedom. If, in addition, the global channel matrix $\mathscr{H}_*$ has full column rank, then we are able to recover any particular user using space–time beamforming. In practice, though, the performance of an ISI–CCI reduction scheme is limited by the SNR and the condition number of $\mathscr{H}_*$.

SPACE–TIME ALGORITHMS FOR THE REVERSE LINK

General Principles of Receive Space–Time Processing

Space processing offers several important opportunities to enhance the performance of the radio link. First, smart antennas offer more resistance to channel fast fading through maximization of *space diversity*. Then, space combining increases the received SNR through array gain, and allows for the suppression of interference when the user of reference and the co-channel-users have different DOAs.

Time processing addresses two important goals. First, it exploits the gain offered by path diversity in delay-spread channels. As the channel time taps generally carry independent fading, the receiver can resolve channel taps and combine them to maximize the signal level. Second, time processing can combat the effects of ISI through equalization. Linear zero-forcing equalizers address the ISI problem but do not fully exploit path diversity. Hence, for these equalizers, ISI suppression and diversity maximization may be conflicting goals. This not the case, however, for maximum likelihood sequence detectors.

Space-time processing allows us to exploit the advantage of both the time and space dimensions. Space-time (linear) filters allow us to maximize space and path diversity. Also, space-time filters can be used for better ISI and CCI reduction. However, as was mentioned above, these goals may still conflict. In contrast, space–time maximum likelihood sequence detectors (see below) can handle harmoniously both diversity maximization and interference minimization.

Channel Estimation

Channel estimation forms an essential part of most wireless digital modems. Channel estimation in the reverse link extracts the information that is necessary for a proper design of the receiver, including linear (space–time beamformer) and nonlinear (decision feedback or maximum likelihood detection) receivers. For this task, most existing systems rely on the periodic transmission of training sequences, which are known both to the transmitter and receiver and used to identify the channel characteristics. The estimation of $\mathcal{H}$ is usually performed using the nonparametric FIR model in Eq. (9), in a least-squares manner or by correlating the observed signals against decorrelated training sequences (as in GSM). A different strategy consists in addressing the estimation of the physical parameters (path angle and delays) of the channel using the model developed in Eq. (8). This strategy proves useful when the number of significant paths is much smaller than the number of channel coefficients.

Channel tracking is also an important issue, necessary whenever the propagation characteristics vary (significantly) within the user slot. Several approaches can be used to update the channel estimate. The *decision-directed method* uses symbol decisions as training symbols to update the channel response estimate. Joint data–channel techniques constitute another alternative, in which symbol estimates and channel information are recursively updated according to the minimization of a likelihood metric: $\|\boldsymbol{X} - \mathcal{H}\boldsymbol{S}\|^2$.

Signal Estimation

Maximum Likelihood Sequence Detection (Single User). Maximum likelihood sequence detection (MLSD) is a popular nonlinear detection scheme that, given the received signal, seeks the sequence of symbols of one particular user that is most likely to have been transmitted. Assuming temporally and spatially white Gaussian noise, maximizing the likelihood reduces to finding the vector $\boldsymbol{S}$ of symbols in a given alphabet that minimizes the following metric:

$$\min_{\boldsymbol{S}} \|\boldsymbol{X} - \mathcal{H}\boldsymbol{S}\|^2 \tag{20}$$

where the channel matrix $\mathcal{H}$ has been previously estimated. Here, $\|\cdot\|$ denotes the conventional euclidean norm. Since $\boldsymbol{X}$ contains measurements in time and space, the criterion above can be considered as a direct extension of the conventional ML sequence detector, which is implemented recursively using the well-known Viterbi algorithm (7).

MLSD offers the lowest BER in a Gaussian noise environment, but is no longer optimal in the presence of co-channel users. In the presence of CCI, a solution to the MLSD problem consists in incorporating in the likelihood metric the information on the statistics of the interferers. This however assumes that the interferers do not undergo significant delay spread. In general though, the optimal solution is given by a multiuser MLSD detection scheme (see below).

Maximum Likelihood Sequence Detection (Multiuser). The multiuser MLSD scheme has been proposed for symbol detection in CCI-dominated channels. The idea consists in treating CCI as other desired users and detecting all signals simultaneously. This time, the Q symbol sequences $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_Q$ are found as the solutions to the following problem:

$$\min_{\{\boldsymbol{S}_q\}} \left\| \boldsymbol{X} - \sum_{q=1}^{Q} \mathcal{H}_q \boldsymbol{S}_q \right\|^2 \tag{21}$$

where again all symbols should belong to the modulation alphabet. The resolution of this problem can be carried out theoretically by a multiuser Viterbi algorithm. However, the complexity of such a scheme grows exponentially with the numbers of users and the channel length, which limits its applicability. Also, the channels of all the users are assumed to be accurately known. In current systems, such information is very difficult to obtain. In addition, the complexity of the multiuser MLSD detector falls beyond current implementation limits. Suboptimal solutions are therefore necessary. One possible strategy, known as *onion peeling,* consists in first decoding the user having the largest power and then subtracting it out from the received data. The procedure is repeated on the residual signal, until all users are decoded. Linear receivers, described below, constitute another form of suboptimal but simple approach to signal detection. Minimum mean square error detection is described below.

Minimum Mean Squared Error Detection. The space–time minimum mean squared error (STMMSE) beamformer is a space–time linear filter whose weights are chosen to minimize the error between the transmitted symbols of a user of reference and the output of the beamformer defined as $y(k) = \boldsymbol{W}_q^{\mathrm{H}}\boldsymbol{X}(k)$. Consider a situation with Q users. Let q be the index of the user of reference. $\boldsymbol{W}_q$ is found by:

$$\min_{\boldsymbol{W}_q} E|y(k) - s_q(k-d)|^2 \tag{22}$$

where d is the chosen reconstruction delay. E here denotes the expectation operator. The solution to this problem follows from the classical normal equations:

$$\boldsymbol{W}_q = E[\boldsymbol{X}(k)\boldsymbol{X}(k)^{\mathrm{H}}]^{-1}E[\boldsymbol{X}(k)s_q(k-d)^*] \tag{23}$$

The solution to this equation can be tracked in various manners, for instance using pilot symbols. Also, it can be shown that the intercorrelation term in the right-hand side of Eq. (23) corresponds to the vector of channel coefficients of the decoded user, when the symbols are uncorrelated. Hence Eq. (23) can also be solved using a channel estimate.

STMMSE combines the strengths of time-only and space-only combining, hence is able to suppress both ISI and CCI. In the noise-free case, when the number of branches is large enough, $\boldsymbol{W}_q$ is found to be a solution of Eq. (19). In the presence of additive noise, the MMSE solution provides a useful tradeoff between the so-called zero-forcing solution of Eq. (19) and the maximum-SNR solution. Finally the computational load of the MMSE is well below than that of the MLSD. However MLSD outperforms the MMSE solution when ISI is the dominant source of interference.

Combined MMSE–MLSD. The purpose of the combined MMSE–MLSD space–time receiver is to be able to deal with both ISI and CCI using a reasonable amount of computation. The idea is to use a STMMSE in a first stage to combat CCI. This leaves us with a signal that is dominated by ISI. After channel estimation, a single-user MLSD algorithm is applied to detect the symbols of the user of interest. Note that the channel seen by the MLSD receiver corresponds to the convolution of the original SIMO channel with the equalizer response.

Space–Time Decision Feedback Equalization. The *decision feedback equalizer* is a nonlinear structure that consists of a space–time linear feedforward filter (FFF) followed by a nonlinear feedback filter. The FFF is used for precursor ISI and CCI suppression. The nonlinear part contains a decision device which produces symbol estimates. An approximation of the postcursor ISI is formed using these estimates and is subtracted from the FFF output to produce new symbol estimates. This technique avoids the noise enhancement problem of the pure linear receiver and has a much lower computational cost than MLSD techniques.

Blind Space–Time Processing Methods

The goal of blind space–time processing methods is to recover the signal transmitted by one or more users, given only the observation of the channel output and minimal information on the symbol statistics and/or the channel structure. Basic available information may include the type of modulation alphabet used by the system. Also, the fact that channel is quasiinvariant in time (during a given data frame) is an essential assumption. Blind methods do not, by definition, resort to the transmission of training sequences. This advantage can be directly traded for an increased information bit rate. It also helps to cope with the situations where the length of the training sequence is not sufficient to acquire an accurate channel estimate. Tutorial information on blind estimation can be found in Ref. 9.

Blind methods in digital communications have been the subject of active research over the last twenty years. It was only recently recognized, however, that blind techniques can benefit from the utilization of the spatial dimension. The main reason is that oversampling the signal in space using multiple antennas, together with the exploitation of the signal–channel structure, allows for efficient channel and beamformer estimation techniques.

Blind Channel Estimation. A significant amount of research work has been focused lately on identifying blindly the impulse response of the transmission channel. The resulting techniques can be broadly categorized into three main classes: higher-order statistics (HOS) methods, second-order statistics (SOS) methods, and maximum-likelihood (ML) methods.

HOS methods look at third- and fourth-order moments of the received data and exploit simple relationships between those moments and the channel coefficients (assuming the knowledge of the input moments) in order to identify the channel. In contrast, the SOS of the output of a scalar (single input-single output) channel do not convey sufficient information for channel estimation, since the second-order moments are phase-blind.

In SIMO systems, SOS does provide the necessary phase information. Hence, one important advantage of multiantenna systems lies in the fact that they can be identified using second-order moments of the observations only. From an algebraic point of view, the use of antenna arrays creates a low-rank model for the vector signal given by the channel output. Specifically, the channel matrix $\mathscr{H}$ in Eq. (14) can be made tall and full-column-rank under mild assumption on the channels. The low-rank property allows one to identify the column span of $\mathscr{H}$ from the observed data. Along with the Toeplitz structure of $\mathscr{H}$, this information can be exploited to identify the channel.

Direct Estimation. Direct methods bypass the channel estimation stage and concentrate on the estimation of the space-time filter. The use of antenna arrays (or oversampling in time or space in general) offers important advantages in this context too. The most important one is perhaps the fact that, as was shown in Eq. (17), the SIMO system can be inverted exactly using a space–time filter with finite time taps, in contrast with the single-output case.

HOS methods for direct receiver estimation are typically designed to optimize a nonlinear cost function of the receiver output. Possible cost functions include Bussgang cost functions [Sato, decision-directed, and constant modulus (CM) algorithms] and kurtosis-based cost functions. The most popular criterion is perhaps the CM criterion, in which the coefficients of the beamformer $\boldsymbol{W}$ are updated according to the minimization (though gradient-descent algorithms) of

$$J(\boldsymbol{W}) = E[|y(k)|^2 - 1]^2$$

where $y(k)$ is the beamformer output.

SOS techniques (sometimes also referred to as "algebraic techniques") look at the problem of factorizing, at least implicitly, the received data matrix $\boldsymbol{X}$ into the product of a block-Toeplitz channel matrix $\mathscr{H}$ and a Hankel symbol matrix $\boldsymbol{S}$

$$\boldsymbol{X} \approx \mathscr{H}\boldsymbol{S} \tag{24}$$

A possible strategy is as follows: Based on the fact that $\mathscr{H}$ is a tall matrix, the row span of $\boldsymbol{S}$ coincides with the row span of $\boldsymbol{X}$. Along with the Hankel structure of $\boldsymbol{S}$, the row span of $\boldsymbol{S}$ can be exploited to uniquely identify $\boldsymbol{S}$.

MULTIUSER RECEIVER

The extension of blind estimation methods to a multiuser scenario poses important theoretical and practical challenges. These challenges include an increased number of unknown parameters, more ambiguities caused by the problem of user mixing, and a higher complexity. Furthermore, situations where the users are not fully synchronized may result in an abruptly time-varying environment which makes the tracking of the channel or receiver coefficients difficult.

As in the nonblind context, multiuser reception can be regarded as a two-stage signal equalization plus separation. Blind equalization of multiuser signals can be addressed using extensions of the aforementioned single-user techniques (HOS, CM, SOS, or subspace techniques). Blind separation of the multiuser signals needs new approaches, since subspace methods alone are not sufficient to solve the separation problem. In CDMA systems, the use of different user spreading codes makes this possible. In TDMA systems, a possible approach to signal separation consists in exploiting side information such as the finite-alphabet property of the modulated signals. The factorization Eq. (24) can then be carried out using alternate projections (see Ref. 10 for a survey). Other schemes include adjusting a space–time filter in order to restore the CM property of the signals.

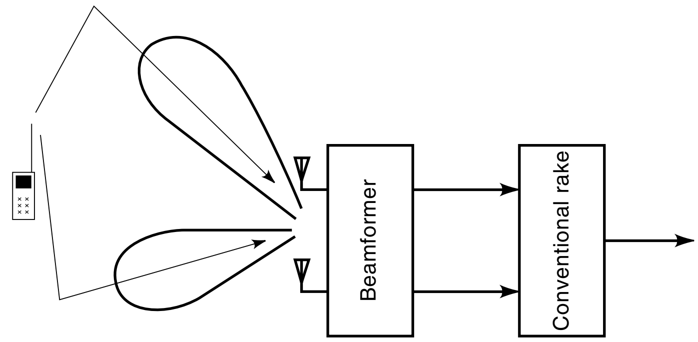

Figure 8. The space–time rake receiver for CDMA uses a beamformer to spatially separate the signals, followed by a conventional rake.

Space–Time Processing for Direct-Sequence Code Division Multiple Access

Direct-sequence CDMA (DSCDMA) systems are expected to gain a significant share of the cellular market. In CDMA, the symbol stream is spread by a *unique* spreading code before transmission. The codes are designed to be orthogonal or quasiorthogonal to each other, making it possible for the users to be separated at the receiver. See Ref. 11 for details. As in TDMA, the use of smart antennas in CDMA system improves the network performance.

We first introduce the DSCDMA model; then we briefly describe space–time CDMA signal processing.

Signal Model. Assume $M > 1$ antennas. The received signal is a vector with M components and can be written as

$$\boldsymbol{x}(t) = \sum_{q=1}^{Q} \sum_{k=-\infty}^{\infty} s_q(k)\boldsymbol{p}_q(t-kT) + \boldsymbol{n}(t)$$

where $s_q(k)$ is the information bit stream for user q, and $\boldsymbol{p}_q(t)$ is the composite channel for user q that embeds both the physical channel $\boldsymbol{h}_q(t)$ (defined as in the TDMA case) and the spreading code $c_q(p)$ of length P:

$$\boldsymbol{p}_q(t) = \sum_{p=0}^{P-1} c_q(p)\boldsymbol{h}_q\left(t - \frac{pT}{P}\right)$$

Space–Time Receiver Design. A popular single-user CDMA receiver is the *RAKE* combiner. The rake receiver exploits the (quasi)orthogonal codes to resolve and coherently combine the paths. It uses one correlator for each path and then combines the outputs to maximize the SNR. The weights of the combiner are selected using diversity combining principles. The rake receiver is a matched filter to the spreading code plus multipath channel.

The space–time rake is an extension of the above. It consists of a beamformer for each path followed by a rake combiner (see Fig. 8). The beamformer reduces the CCI at the rake input and thus improves the system capacity.

SPACE–TIME ALGORITHMS FOR THE FORWARD LINK

General Principles of Transmit Space–Time Processing

In transmit space–time processing, the signal to be transmitted is combined in time and space before it is radiated by the antennas to encounter the channel. The goal of this operation is to enhance the signal received by the desired user, while minimizing the energy sent towards co-channel-users. Space–time processing makes use of the spatial and temporal signature of the users to differentiate them. It may also be used to preequalize the channel, that is, to reduce ISI in the received signal. Multiple antennas can also be used to offer transmit diversity against channel fading.

Channel Estimation

The major challenge in transmit space–time processing is the estimation of the forward link channel. Further, for CCI suppression, we need to estimate the channels of all co-channel-users.

Time-Division Duplex Systems. TDD systems use the same frequency for the forward link and the reverse link. Given the reciprocity principle, the forward and reverse link channels should be identical. However, transmit and receive take place in different time slots; hence the channels may differ, depending on the ping–pong period (time duration between receive and transmit phases) and the coherence time of the channel.

Frequency-Division Duplex Systems. In frequency division duplex (FDD) systems, reverse and forward links operate on different frequencies. In multipath environment, this can cause the reverse and forward link channels to differ significantly.

Essentially, in a specular channel, the forward and reverse DOAs and times of arrival (TOAs) are the same, but not the path complex amplitudes. A typical strategy consists in identifying the DOAs of the dominant incoming path, then using spatial beamforming in transmit in order to focus energy in these directions while reducing the radiated power in other directions. Adaptive nulls may also be formed in the directions of interfering users. However, this requires the DOAs for the co-channel-users to be known.

A direct approach for transmit channel estimation is based on feedback. This approach involves the user estimating the channel from the downlink signal and sending this information back to the transmitter. In the sequel, we assume that the forward channel information is available at the transmitter.

Single-User Minimum Mean Squared Error

The goal of space–time processing in transmit is to maximize the signal level received by the desired user from the base station, while minimizing the ISI and CCI to other users. The space–time beamformer $\boldsymbol{W}$ is chosen so as to minimize the following MMSE expression:

$$\min_{\boldsymbol{W}} \left(E \left\| \boldsymbol{W}^{\mathrm{H}} \mathscr{H}_q^{\mathrm{F}} \boldsymbol{S}_q(k) - s_q(k-d) \right\|_2^2 + \alpha \sum_{k=1, k \neq q}^{Q} \boldsymbol{W}^{\mathrm{H}} \mathscr{H}_k^{\mathrm{F}} \mathscr{H}_k^{\mathrm{FH}} \boldsymbol{W} \right) \tag{25}$$

where α is a parameter that balances the ISI reduction at the reference mobile and the CCI reduction at other mobiles. d is the chosen reconstruction delay. $\mathscr{H}_q^{\mathrm{F}}$ is the block-Toeplitz matrix [defined as in Eq. (14)] containing the coefficients of the forward link channel for the desired user. $\mathscr{H}_k^{\mathrm{F}}$, $k \neq q$, denotes the forward channel matrix for the other users.

Multiuser Minimum Mean Squared Error

Assume that Q co-channel-users, operating within a given cell, communicate with the same base station. The multiuser MMSE problem involves adjusting Q space–time beamformers so as to maximize the signal level and minimize the ISI and CCI at each mobile. Note that CCI that originates from other cells is ignored here. The base communicates with user q through a beamformer $\boldsymbol{W}_q$. All beamformers $\boldsymbol{W}_q$, $q = 1, \ldots, Q$, are jointly estimated by the optimization of the following cost function:

$$\min_{\boldsymbol{W}_q, q=1,\ldots,Q} \sum_{q=1}^{Q} \left(E \left\| \boldsymbol{W}_q^{\mathrm{H}} \mathscr{H}_q^{\mathrm{F}} \boldsymbol{S}_q(k) - s_q(k-d) \right\|_2^2 + \alpha \sum_{k=1, k \neq q}^{Q} \boldsymbol{W}_k^{\mathrm{H}} \mathscr{H}_q^{\mathrm{F}} \mathscr{H}_q^{\mathrm{FH}} \boldsymbol{W}_k \right) \tag{26}$$

It turns out that the problem above decouples into Q independent quadratic problem, each having the form shown in Eq. (25). The multiuser MMSE problem can therefore be solved without difficulty.

Space–Time Coding

When the forward channel is unknown or only partially known (in FDD systems), transmit diversity cannot be implemented directly as in TDD systems, even if we have multiple transmit antennas that exhibit low fade correlation. There is an emerging class of techniques that offer transmit diversity in FDD systems by using space–time channel coding. The diversity gain can then be translated into significant improvements in data rates or BER performance.

The basic approach in space–time coding is to split the encoded data into multiple data streams, each of which is modulated and simultaneously transmitted from a different antenna. Different choices of data-to-antenna mapping can be used. All antennas can use the same modulation and carrier frequency. Alternatively different modulation (symbol waveforms) or symbol delays can be used. Other approaches include use of different carriers (multicarrier techniques) or spreading codes. The received signal is a superposition of the multiple transmitted signals. Channel decoding can be used to recover the data sequence. Since the encoded data arrive over uncorrelated faded branches, diversity gain can be realized.

APPLICATIONS OF SPACE–TIME PROCESSING

We now briefly review existing and emerging applications of space–time processing that are currently deployed in base stations of cellular networks.

Switched-Beam Systems

Switched beam systems (SBSs) are nonadaptive beamforming systems that involve the use of four to eight antennas per sector at the base station. Here the system is presented for receive beamforming, but a similar concept can be used for transmit. The cell usually consists of three sectors that cover a 120° angle each. In each sector, the outputs of the antennas are combined to form a number of beams with predesigned patterns. These fixed beams are obtained through the use of a Butler matrix. In most current cellular standards (including analog FDMA and digital FDMA–TDMA), a sector and a channel–time-slot pair are assigned to one user only. In order to enhance the communication with this user, the base station examines, through an electronic *sniffer,* the best beam output and switches to it. In some systems, two beams may be picked up and their outputs forwarded to a selection diversity device. Since the base also receives signals from mobile users in surrounding cells, the sniffer should be able to detect the desired signal in the presence of interferers. To minimize the probability of incorrect beam selection, the beam output is validated by a color code that identifies the user. In digital systems, beam selection is performed at baseband, after channel equalization and synchronization.

SBSs provide array gain, which can be traded for an extended cell coverage. The gain brought by SBS is given by 10 log m, where m is the number of antennas. SBSs also help combat CCI. However, since the beams have a fixed width, interference suppression can occur only when the desired signal and the interferer fall into different beams. As a result, the performance of such a system is highly dependent on propagation environments and cell loading conditions. The SBS also experiences several losses, such as cusping losses (since there is a 2 to 3 dB cusp between beams), beam selection loss, mismatch loss in the presence of nonplanar wavefronts, and loss of path diversity.

Reuse within Cell

Since cellular communication systems are (increasingly) interference-limited, the gain in CCI reduction brought by the use of smart antennas can be traded for an increase in the number of users supported by the network for a given quality of service. In current TDMA standards, this capacity improvement can be obtained through the use of a smaller frequency reuse factor. Hence, the available frequency band is reused more often, and consequently a larger number of carriers are available in each cell.

Assuming a more drastic modification of the system design, the network will support several users in a given frequency channel in the same cell. This is called reuse within cell (RWC). RWC assumes these users have sufficiently different space–time signatures so that the receiver can achieve sufficient signal separation. When the users become too closely aligned in their signatures, space–time processing can no longer achieve signal recovery, and the users should be handed off to different frequencies or time slots. As another limitation of RWC, the space–time signatures (channel coefficients) of each user needs to be acquired with good accuracy. This can be a difficult task when the powers of the different users are not well balanced. Also, the propagation environment plays a major role in determining the complexity of the channel structure. Finally, angle spread, delay spread, and Doppler spread strongly affect the quality of channel estimation. As an additional difficulty, the channel estimation required in forward link space-time processing is made difficult in FDD systems.

SUMMARY

Smart antennas constitute a promising but still emerging technology. Space–time processing algorithms provide powerful tools to enhance the overall performance of wireless cellular networks. Improvements, typically by a factor of two in cell coverage or capacity, are shown to be possible according to results from field deployments using simple beamforming. Greater improvements can be obtained from some of the more advanced space–time processing solutions described in this paper. The successful integration of space–time processing techniques will however also require a substantial evolution of the current air interfaces. Also, the design of space–time algorithms must also be application- and environment-specific.

BIBLIOGRAPHY

1. R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays,* New York: Wiley, 1980.
2. D. Johnson and D. Dudgeon, *Array Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1993.
3. J. D. Gibson (ed.), *The Mobile Communications Handbook,* Boca Raton, FL: CRC Press, 1996.
4. W. C. Jakes, *Microwave Mobile Communications,* New York: Wiley, 1974.
5. *Proc. 4th Workshop on Smart Antennas in Wireless Mobile Communications,* Sanford, CA: Center for Telecommunications and Information Systems Laboratory, Stanford University, 1997.
6. R. S. Kennedy, *Fading Dispersive Communication Channels,* New York: Wiley, 1969.
7. J. G. Proakis, *Digital Communications,* New York: McGraw-Hill, 1989.
8. S. U. H. Qureshi, Adaptive equalization, *Proc. IEEE,* **53**: 1349–1387, 1985.
9. S. Haykin (ed.), *Blind Deconvolution,* Englewood Cliffs, NJ: Prentice-Hall, 1994.
10. Special issue on blind identification and estimation, *Proc. IEEE,* to be published 1998.
11. M. K. Simon et al., *Spread Spectrum Communications Handbook,* New York: McGraw-Hill, 1994.
12. A. Paulray and C. B. Papadias, Space-time processing for wireless communications, *IEEE Signal Processing Magazine,* **14** (6): 49–83, Nov. 1997.
13. S. U. Pillai, *Array Signal Processing,* New York: Springer-Verlag, 1989.

AROGYASWAMI PAULRAJ
DAVID GESBERT
Stanford University
CONSTANTINOS PAPADIAS
Lucent Technologies

ANTENNA NOISE. See RADIO NOISE.