

Base Station Transmitter Antenna Arrays with Mobile to Base Feedback

Derek Gerlach, Arogyaswami Paulraj

Information Systems Lab
Stanford University
Stanford, California 94305

ABSTRACT

A central base station simultaneously communicates with several mobile users by allocating a separate time or frequency channel for each mobile to prevent undesired crosstalk. However, each time or frequency channel may be reused among several mobiles by means of an antenna array at the base station which points a separate beam at each user. The downlink beamformer would normally operate in an "open loop" mode, in which the base steers a mainlobe in the direction of each mobile. Such a system may operate effectively in a free space environment with no multipath. In the presence of scattering, open loop methods will not perform adequately. A new "closed loop" technique is presented in which each mobile user feeds back to the base estimates of the received signal amplitudes. Using feedback, the base station can achieve precision beamforming resulting in lower crosstalk and improved signal separation even in the presence of strong scattering environments.

1. INTRODUCTION

In cellular radio and other wireless communication systems, a central base station simultaneously communicates with several mobile users. To accommodate these users without crosstalk, the available bandwidth is partitioned into several independent channels. The methods of frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA) divide the resources of bandwidth and time among the users to create a collection of separate channels.

The scarcity of bandwidth creates a need to increase the number of users serviceable by a single base station using a fixed, allocated bandwidth. One way of increasing capacity is through frequency reuse, in which more than one mobile user simultaneously shares the same time or frequency channel, and crosstalk normally present in frequency reuse is eliminated via spatial division multiplexing (SDMA). SDMA recognizes that each user typically has a different angle of arrival to the base station. SDMA uses an antenna array at the base to create a separate beampattern for each user. Such beampatterns provide an additional means of multi-

plexing beyond time, frequency, or code, and makes possible the reuse of bandwidth [4]. We consider antenna arrays at the base, and the mobile is assumed to have a single omnidirectional antenna.

The aim of SDMA is slightly different in the uplink and the downlink. In receive beamforming, the object is to separate several cochannel signals which are simultaneously impinging on the array, by forming separate beams for each incoming signal. In this way, several mobiles may simultaneously talk to the base on the same frequency band, and the base hears each without mutual crosstalk by virtue of their angle difference. The beamformer and separated signals are *collocated at the base*; therefore, the beamformer can gauge the separated signals and modify itself to improve signal separation. Antenna arrays used in receive mode are easier to implement because there is a natural feedback structure in the receive array scenario [6] [1] [2].

In transmit beamforming, the object is to simultaneously transmit several signals so that each signal is received only by its intended mobile. However, the transmitter, located at the base, normally has no measure of the signal separation, which occurs at the mobile. Therefore, there is no convenient feedback structure in a transmitting array scenario.

In the absence of feedback, transmit beamforming is particularly difficult in scattering environments. A transmitting array will have no a priori knowledge of the location of scattering bodies; therefore, the array, by directing a mainlobe towards a mobile, could cause crosstalk through scattering bodies to several other mobiles. Unless the base can account for all of the scattering bodies in the environment, undesired crosstalk is liable to occur.

The proposed approach adds the missing element in the transmit problem by providing feedback to gauge the separation performance at the mobile, resulting in lower crosstalk even in the presence of scatterers. In this closed loop method, a probing tone is periodically sent by the base and measured by the mobiles. These measurements are fed back to the base by the mobile. Using the response to the probing tones, the base can send the information signals using the array so that each signal is received only by its intended receiver,

even with strong scattering.

2. DEFINITIONS AND ASSUMPTIONS

Let

d = the number of transmitted information signals (= number of receivers).

p = the number of transmitted probing signals (defined in III).

m = the number of antenna elements at the base.

Assume that there is one information signal intended for each receiver. Denote the d information signals as

$$s_1(t), \dots, s_d(t) \quad (1)$$

Each of these information signals sent by the base is multiplied by its own vector which gives its complex amplitude weighting on the antenna elements.

The i^{th} information signal transmitted by the elements is given by the m dimensional vector $\mathbf{w}_{Ii}s_i(t)$. The information signals are transmitted simultaneously, each according to its own complex information weight vector. The complex weights of the d signals across the m elements are summarized by the $m \times d$ matrix $\mathbf{W}_I$. The i^{th} column of $\mathbf{W}_I$ is the vector of weights according to which the i^{th} information signal is applied to the array. The *information weight matrix* is given by

$$\mathbf{W}_I = [\mathbf{w}_{I1} \cdots \mathbf{w}_{Id}] \quad (2)$$

This matrix is chosen by the base in order to transmit each signal only to its intended receiver.

Assume that the channel gain from each element to each receiver is flat over the bandwidth of the information signals, so that it is characterized by a single complex scalar. Thus, the channel from array elements to mobiles is characterized by the *channel matrix* $\mathbf{A}$ of size $m \times d$. The (i, j) entry contains the channel gain from the i^{th} element to the j^{th} receiver. The channel matrix $\mathbf{A}$ summarizes the entire propagation environment, which may include scattering and fading processes. It is normally unknown to the base, and since it is determined by the scattering and mobile locations, it is not under the control of the base.

The complex amplitude c_{Iij} of i^{th} information signal at the j^{th} receiver is given by the inner product of the signal's weight vector with the corresponding channel vector:

$$c_{Iij} = \mathbf{w}_{Ii}^H \mathbf{a}_j. \quad (3)$$

where vector $\mathbf{w}_{Ii}$ is the i^{th} column of matrix $\mathbf{W}_I$, vector $\mathbf{a}_j$ is the j^{th} column of $\mathbf{A}$, and H denotes the Hermitian conjugate. Equation (3) is a vector version of the familiar statement

$$\begin{pmatrix} \text{receiver} \\ \text{amplitude} \end{pmatrix} = \begin{pmatrix} \text{transmitter} \\ \text{amplitude} \end{pmatrix} \times (\text{channel gain}).$$

The information signals at each of the receivers can therefore be characterized by the *received amplitude matrix* $\mathbf{C}_I$, which is size $d \times d$. The (i, j) entry specifies the complex signal amplitude of the i^{th} signal at the j^{th} receiver. In other words, the off-diagonal elements express the crosstalk, which the diagonal elements express the desired signal amplitudes. The relationship between each transmitted signal amplitude and each received amplitude is given by:

$$\mathbf{W}_I^H \mathbf{A} = \mathbf{C}_I \quad (4)$$

Feedback is needed to give the transmitter a basis for choosing the information weights. The transmitter chooses the weights $\mathbf{W}_I$, so that each signal is only received by its intended user, i.e.,

$$\mathbf{C}_I = \mathbf{I}. \quad (5)$$

Beampatterns which induce this condition in eq (5) are "complementary null beampatterns." The base has no control over the scattering environment $\mathbf{A}$. The information signal levels at the receivers $\mathbf{C}_I$ are a result of the information weights interacting with the environment through eq (4), and eq (5) represents an implicit constraint on $\mathbf{W}_I$. Feedback gives the base an estimate of $\mathbf{A}$, so that it can induce eq (5) at the mobiles.

Since each element adds a degree of freedom, to achieve eq (5), we must have $m \geq d$.

3. INCORPORATION OF FEEDBACK

In the method described here, closing the loop provides the transmitter with feedback data characterizing the spatial channel. This knowledge enables the base to use to reflections present in the environment to form a condition of constructive interference at the desired mobile, and destructive interference at all the remaining mobiles.

In this closed loop approach, the base has two modes in which two types of signals are transmitted: *information signals* and *probing signals*. While transmitting information signals in the information mode using eq (6), the base station periodically pauses and sends a burst of probing signals. In the probing mode, each element is excited in turn by a single probing tone at the carrier frequency. Each mobile receives the probing tones and measures their relative amplitudes and phases. Then, the amplitude and phase data is fed back to the base. These complex responses measured at the mobiles form the entries of the *probing response matrix* $\mathbf{C}_P$. Instead of exciting each element in turn, it is also possible to excite linear combinations of elements. Therefore, in the probing mode, each probing tone is transmitted according to its probing vector $\mathbf{w}_P$. This vector is the i^{th} column of the *probing weight matrix* $\mathbf{W}_P$:

$$\mathbf{W}_P = [\mathbf{w}_{P1} \cdots \mathbf{w}_{Pp}] \quad (6)$$

In general, there will be p different excitations so that there are p separate probing vectors. Probing using various linear combinations instead of element by element can increase robustness to measurement noise. The probing weight matrix $\mathbf{W}_P$ is $m \times p$ and the probing response matrix $\mathbf{C}_P$ is $p \times d$. In this case, the relationship between the probing vectors and the measured response is analogous to the information signal case

$$\mathbf{C}_P = \mathbf{W}_P^H \mathbf{A} \quad (7)$$

The receivers measure the amplitudes $\mathbf{C}_P$ of the probing tones and send this data back to the transmitter. In each measurement, the receivers measure the probing signal in the presence of zero mean Gaussian noise using a matched filter, so the transmitter is fed back the matrix:

$$\hat{\mathbf{C}}_P = \mathbf{C}_P + \mathbf{E}_C \quad (8)$$

where the entries of $\mathbf{E}_C$ are zero mean iid Gaussian random variable. Knowing $\hat{\mathbf{C}}_P$, the transmitter forms an estimate of $\mathbf{A}$ according to

$$\hat{\mathbf{A}} = \mathbf{W}_P^{+H} \hat{\mathbf{C}}_P \quad (9)$$

where $+$ denotes the pseudoinverse operation. The pseudoinverse in the above equation computes the least squares estimate of the channel matrix from an overdetermined set of equations. The independent additive noise model is appropriate if the measurements from receiver to receiver and probing signal to probing signal are independent.

Regardless of what combinations of elements are used for probing ($\mathbf{W}_P$), once the estimate $\hat{\mathbf{A}}$ is known, the transmitter forms a $\mathbf{W}_I$ which causes

$$\mathbf{W}_I^H \hat{\mathbf{A}} = \mathbf{I} \quad (10)$$

A set of information transmission weights which achieve this is given by

$$\mathbf{W}_I = \hat{\mathbf{A}}^{+H} \quad (11)$$

Each additional probing vector provides another equation involving $\mathbf{A}$. Since $\mathbf{A}$ has m rows, the condition to completely determine $\mathbf{A}$ is $p \geq m$. Extra probing vectors beyond m will improve the accuracy of the least-squares estimate of $\mathbf{A}$.

Incorporation of feedback therefore consists of the following five steps:

1. Transmit the information signals to the mobiles according to an initial set of information weight vectors.
2. Monitor the level of crosstalk as the mobiles move in the propagation environment, and halt the information transmission when the crosstalk at mobiles exceeds a threshold of acceptability.

3. Enter probing mode:

- (a) Choose probing vectors $\mathbf{W}_P$.
- (b) Transmit the probing vectors $\mathbf{W}_P$ and measure the probing response matrix $\hat{\mathbf{C}}_P$ at the mobiles.
- (c) Feed back the probing response matrix $\hat{\mathbf{C}}_P$ to the base.
- (d) Estimate $\mathbf{A}$ via $\hat{\mathbf{A}} = \mathbf{W}_P^{+H} \hat{\mathbf{C}}_P$
- (e) Form information signal weights according to $\mathbf{W}_I = \hat{\mathbf{A}}^{+H}$

4. Resume information transmission with the new choice of information weights.

5. Repeat starting at step 2.

During probing, no information can be transmitted. Therefore, the duration for which the downlink channel is unavailable will be determined by the value of p , the value of m , and the time of flight to and from the mobile. However, the larger the value of p , the more accurate the least squares channel estimate (7) will be, resulting in an improved beam pattern.

The frequency with which it will be necessary to enter the probing mode will be determined by the rate at which matrix $\mathbf{A}$ changes, which is in turn determined by the mobiles' speeds. Matrix $\mathbf{A}$ will also change rapidly in complex scattering environments.

4. GENERALIZED PROBING METHODS

In the example of the last section, each element, or each combination of elements, was excited in turn by a probing tone. The difference in time of transmission of each probing vector allowed each corresponding response to be measured without interference from the other probing vectors. Instead of separation by time of transmission, each probing vector and associated probing tone can be sent on a slightly different frequency, and the responses at the mobiles measured simultaneously. The shifts would be small enough so that all the tones still lie in or nearby the data signal bandwidth. In fact, any form of orthogonality among the probing signals will suffice. Denote the probing signals by

$$p_1(t), \dots, p_p(t) \text{ for } t \in [0, T_p] \quad (12)$$

The interval $[0, T_p]$ is the period over which the probing signals are transmitted. Most importantly, the probing signals must lie in or very close to the same frequency band as the information signals, because their purpose is to probe the spatial channel over which the information signals are transmitted. With each probing signal is associated a probing vector $\mathbf{w}_{P_i}$ which gives the probing signal amplitude at each element. To insure that each probing vector excitation is measurable without

interference from the other probing signals, each probing signal must be assigned to its own time-bandwidth channel, ie, the probing signals are orthogonal:

$$\int_0^{T_p} p_i(t)p_j^*(t)dt = \delta_{(i,j)} \quad (13)$$

This equation states that any combination of T/F/CDMA may be used to separate each probing signal. Implementation complexity determines which is most appropriate. Exciting each element in turn with a single frequency pilot tone seems easiest, but other combinations are possible.

5. CHOOSING THE INFORMATION WEIGHTS

Each mobile user gives rise to an additional constraint on the information weight vectors: the desired signal at the mobile should have unity amplitude, and the remaining signals should have zero amplitude. Eq (5) places d constraints on the information weight matrix, but since there are m degrees of freedom, the weight matrix satisfying eq (5) is underdetermined. The information weight matrix based on the "A-estimation" method given in eq (11) is therefore a special case of a general class of information weight matrices which achieve condition eq (5) when $\mathbf{E}_C = 0$. Eq (11) is the the "Moore-Penrose" pseudoinverse which achieves eq (11) with minimum norm of the information weights [5].

The A-estimation choice eq (11) is very important not only because it achieves eq (5) with minimum energy, but also because it generates an explicit estimate of the matrix $\mathbf{A}$. This allows various manifold and/or subspace constraints on $\mathbf{A}$ to be incorporated into both the probing weight vector and information weight vector choices. Although these incorporations are very significant in making transmission beamforming practical in scattering environments, they will not be addressed in this introductory paper.

If there is no error in $\hat{\mathbf{C}}_P$, then all information weight matrices achieving eq (11) will be equally suitable, except that some will have larger norms than others, and will therefore require more transmitter power to send the signals reliably to the mobiles. However, in the presence of errors in $\mathbf{C}_P$, some information weight matrices will give rise to beampatterns which deviate less than others from a complementary null beampattern. Which beampattern will deviate least? Let us use the following performance measure for comparison:

$$J = E \|\mathbf{W}_I^H \mathbf{A} - \mathbf{I}\|_F^2 \quad (14)$$

This "Frobenius norm" is the square root of the sum of the squares of the entries of the difference. As such, it gauges how far the information signals' amplitudes at

each mobile deviate on average from the complementary null beampattern.

Each information weight matrix which achieves eq (5) can be written as the minimum norm solution with respect to a "weighting matrix" $\mathbf{M}$. This class of information weight matrices is given by

$$\mathbf{W}_I^H = (\hat{\mathbf{C}}_P^H \mathbf{M} \hat{\mathbf{C}}_P)^{-1} \hat{\mathbf{C}}_P^H \mathbf{M} \mathbf{W}_P^H \quad (15)$$

where matrix $\mathbf{M}$ must be $p \times p$ symmetric, and at least rank d . If $\mathbf{M}$ is of lower rank, then the inverse factor will not exist. For every choice of matrix $\mathbf{M}$, there is a corresponding $\mathbf{W}_I$. It can be shown that choosing $\mathbf{M} = \mathbf{W}_P^+ \mathbf{W}_P^{H+}$ corresponds to the A-estimation choice of information weight vectors.

The choice of information weights which minimizes eq (14) is denoted the "direct inversion" choice, and it occurs when $\mathbf{M} = \mathbf{I}$. In the "direct inversion" choice, the information weights are computed without forming an explicit estimate of $\mathbf{A}$, and are given by

$$\mathbf{W}_I = \mathbf{W}_P \hat{\mathbf{C}}_P^{H+} \quad (16)$$

To show optimality, plug eq (15) in eq (14), and the measure becomes

$$J = E \left\| (\hat{\mathbf{C}}_P^H \mathbf{M} \hat{\mathbf{C}}_P)^{-1} \hat{\mathbf{C}}_P^H \mathbf{M} \mathbf{C}_P - \mathbf{I} \right\|_F^2 \quad (17)$$

$$= E \left\| (\hat{\mathbf{C}}_P^H \mathbf{M} \hat{\mathbf{C}}_P)^{-1} \hat{\mathbf{C}}_P^H \mathbf{M} \mathbf{E}_C \right\|_F^2 \quad (18)$$

If the errors $\mathbf{E}_C$ are small, this measure becomes

$$J \doteq E \left\| (\mathbf{C}_P^H \mathbf{M} \mathbf{C}_P)^{-1} \mathbf{C}_P^H \mathbf{M} \mathbf{E}_C \right\|_F^2 \quad (19)$$

$$= \text{Tr} \left\{ E \left[(\mathbf{C}_P^H \mathbf{M} \mathbf{C}_P)^{-1} \cdot \mathbf{C}_P^H \mathbf{M} \mathbf{E}_C \mathbf{E}_C^H \mathbf{M}^H \mathbf{C}_P (\mathbf{C}_P^H \mathbf{M} \mathbf{C}_P)^{-1} \right] \right\} \quad (20)$$

$$= r\sigma_m^2 \text{Tr} \left\{ (\mathbf{C}_P^H \mathbf{M} \mathbf{C}_P)^{-1} \cdot \mathbf{C}_P^H \mathbf{M} \mathbf{M}^H \mathbf{C}_P (\mathbf{C}_P^H \mathbf{M} \mathbf{C}_P)^{-1} \right\} \quad (21)$$

Lemma II.2 in [3] uses a "Shur Complement" argument to show that the choice of $\mathbf{M}$ which minimizes J is $\mathbf{M} = \mathbf{I}$. This corresponds to direct inversion discussed above. Therefore, in the presence of measurement noise, the direct inversion gives the information weights with the lowest average squared distance from the complementary null beampattern.

While the direct method achieves the lowest average deviation from eq (5), and the A-estimation achieves the lowest average information signal energy. These two methods coincide if either

1. Matrix $\mathbf{W}_P$ is unitary, or

2. $p = d$

Consider the expression for the information weights in the A-estimation method.

$$\mathbf{W}_I^H = (\hat{\mathbf{C}}_P^H \mathbf{W}_P^+ \mathbf{W}_P^{+H} \hat{\mathbf{C}}_P)^{-1} \hat{\mathbf{C}}_P^H \mathbf{W}_P^+ \quad (22)$$

If $\mathbf{W}_P$ is unitary, then $\mathbf{W}_P^+ \mathbf{W}_P^{+H} = \mathbf{I}$ and $\mathbf{W}_P^+ = \mathbf{W}_P^H$, and the expression for $\mathbf{W}_I$ becomes that of the direct method. If $p = d$, then $\mathbf{C}_P$ is square, and $(\mathbf{W}_P^{+H} \hat{\mathbf{C}}_P)^+ = \hat{\mathbf{C}}_P^+ \mathbf{W}_P^H = \mathbf{W}_I$ for the direct method.

Although there are many methods corresponding to various choices of $\mathbf{M}$ which give rise to different $\mathbf{W}_I$, the A-estimation and the direct method are the only cases of importance.

6. SIMULATIONS

This feedback method was used to establish a separate channel to each of two mobiles simultaneously. The propagation environment was free space. Using an eight element circular array, a beam pattern for each signal was created according to eq (9) and eq (11). The base station periodically alternated between probing and information mode, and the mobiles moved around the base in a circular path. Because the channel changed as the mobiles moved, the interference was least immediately after each probe, and became worst immediately before the next probe. Figure 1 shows the worst case signal to noise plus interference ratio (SINR) vs. probing rate. The probing rate is given in terms of degrees moved between each probe. The mobiles were always separated by a fixed angle as they moved around the base. Matrix $\mathbf{W}_P$ was an 8×8 identity matrix. The measurement noise in $\mathbf{C}_P$ was additive zero-mean Gaussian noise 40 dB below the average values of the entries of $\mathbf{A}$. The plot shows that as the mobiles were spaced closer and closer, faster probing rates were needed to achieve the same SINR. This is because a more precise beam pattern was needed to control signal leakage with closer spacings.

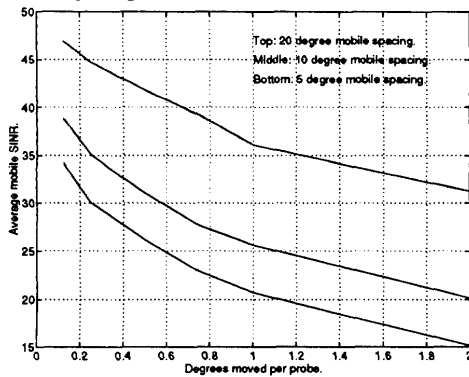


Fig.1 Average Mobile SINR vs Mobile Speed.

7. CONCLUSIONS

In a TDMA or FDMA access scheme, a closed loop feedback approach could permit two or more users to share the same time or frequency slot in the same cell because the precision null beam pattern keeps crosstalk at low levels. The number of array elements needed to achieve this is at least the number of mobiles sharing a channel, not the total number of mobiles in a cell. That is, for a x2 capacity increase, at least two antennas are needed at the base. Extra antennas add degrees of freedom and increase the robustness of the process so that it achieves the desired beam pattern with lower total transmitted power. Future work will focus on performance in more realistic scattering environments, and quantifying the probing overhead needed to achieve sufficient signal separation.

In contrast to TDMA or FDMA, using feedback in CDMA to achieve such precision nulling is not appropriate or necessary. In CDMA, increasing the number of users beyond nominal capacity does not entail the reuse of the existing codes for the new users. Rather, new users are assigned new codes and still retain a degree of code processing separation. This obviates the need for precision nulls for strict crosstalk suppression. Thus feedback control of precision nulling beam patterns is only appropriate in TDMA and FDMA access schemes.

8. REFERENCES

- [1] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, 1991.
- [2] J. E. Hudson. *Adaptive Array Principles*. Peter Peregrinus Ltd., Stevenage, UK, 1989.
- [3] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, 1987.
- [4] Ayman Naguib and Arogyaswami Paulraj. Performance of CDMA Cellular Networks with Base-Station Antenna Arrays. submitted to the International Zurich Seminar on Digital Communications, March 1994.
- [5] Gilbert Strang. *Linear Algebra and Its Applications*. Harcourt Brace Jovanovich, San Diego, 1988.
- [6] B. Widrow and S. Stern. *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, 1985.