Multiuser Receivers, Random Matrices and Free Probability *

David N.C. Tse
Department of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720, USA
dtse@eecs.berkeley.edu

There are often curious and deep connections between engineering and mathematics. We tell one such story here. It starts with the engineering problem of capacity analysis of direct-sequence spread spectrum wireless networks. Using results from convergence of spectra of large random matrices, we obtain simple answers for a basic model of this problem, in terms of notions of *effective bandwidths* and *effective interference*. In the process of understanding the structure of our solution better, we stumble on *free probability*, a theory for non-commutative random variables. We establish a striking connection between certain basic results in this theory and the concept of effective interference. Using free probability as a new tool, we are in turn able to analyze more complex models for the original wireless communications problem. Our story has come a full circle.

1 The Problem

With the introduction of the IS-95 Code-Division Multiple-Access (CDMA) standard, the use of spread-spectrum as a multiple-access technique in commercial wireless systems is growing rapidly in popularity. Unlike more traditional methods such as time-division multiple access (TDMA) or frequency-division multiple access (FDMA), spread-spectrum techniques are *broadband* in the sense that the entire transmission bandwidth is shared between all users at all times. This is done by the spreading of the users' signals onto a bandwidth much larger than an individual user's information rate. The advantages of spread-spectrum techniques include simpler statistical multiplexing without explicit scheduling of time or frequency slots, universal frequency reuse between cells, graceful degradation of quality near congestion, and exploitation of frequency-selective fading to avoid the harmful effects of deep fades that afflict narrowband systems.

*This research is supported by a National Science Foundation Early Faculty CAREER Award.

One form of spread-spectrum is *direct-sequence* CDMA. In this multiple access method, each user is assigned a *signature sequence* on which it modulates its data. A simple channel model for such a system is:

$$\mathbf{y} = \sum_{k=1}^{K} x_k \mathbf{s}_k + \mathbf{w} \in \Re^N, \tag{1}$$

where K is the number of users, x_k the transmitted data symbol of user k, $\mathbf{w} \sim N(0, \sigma^2 I_N)$ is the additive Gaussian noise in the channel, and $\mathbf{y}$ is the received vector. We assume that $E[x_k] = 0$, $E[x_k^2] = p_k$, the received power of user k. The parameter N, sometimes called the *processing gain*, is the ratio of the channel bandwidth and the data symbol rate and quantifies the amount of spreading in the system. The larger the channel bandwidth, the larger N can be. Geometrically, communication takes place in a signal space of dimension N, with each user occupying a signal direction defined by its signature sequence $\mathbf{s}_k$. The parameter N can also be thought of as the number of degrees of freedom in the system. The task of a receiver is to estimate the data symbol of each of the users from the received vector $\mathbf{y}$. We assume that the receiver has knowledge of the signature sequences and received powers of the users.

A natural class of receivers are the *linear receivers*. Focusing on user 1, the estimate $\hat{x}_1$ of the transmitted symbol x_1 is of the form $\hat{x}_1 = \mathbf{c}_1^t \mathbf{y}$, where $\mathbf{c}_1 \in \Re^N$ is the linear receiver for user 1. For example, choosing $\mathbf{c}_1 = \mathbf{s}_1$ yields the well-known *matched-filter* receiver; it projects the received signal onto the direction of the signature sequence of user 1. This is the receiver used in current-generation CDMA systems, and is simple in the sense that the receiver needs only keep track of the signature sequence of user 1 and not the others. However, one can expect that better performance can be attained by taking advantage of the knowledge of the signature sequences and received powers of other users as well.

A commonly used measure to evaluate the performance of linear receivers is the output *signal-to-interference ratio*:

$$\mathrm{SIR}_1 := \frac{(\mathbf{c}_1^t \mathbf{s}_1)^2 p_1}{(\mathbf{c}_1^t \mathbf{c}_1)\sigma^2 + \sum_{i \neq 1} (\mathbf{c}_1^t \mathbf{s}_i)^2 p_i}$$

This is simply the ratio of the variance of user 1's signal to the variance of noise plus interference from other users, measured at the output of the linear receiver. It is not too difficult to show that the optimal receiver which maximizes the output SIR is the *minimum mean-square error* (MMSE) receiver which minimizes $E[(\hat{x}_1 - x_1)^2]$. This is the classic least squares problem and has a well-known solution. The MMSE receiver and its SIR performance is given by:

$$\begin{aligned} \mathbf{c}_1 &= (SDS^t + \sigma^2 I)^{-1} \mathbf{s}_1 \\ \mathrm{SIR}_1 &= p_1 \mathbf{s}_1^t (SDS^t + \sigma^2 I)^{-1} \mathbf{s}_1 \end{aligned} \tag{2}$$

where

$$S := [\mathbf{s}_2, \ldots, \mathbf{s}_K], \qquad D := \mathrm{diag}(p_2, \ldots, p_K). \tag{3}$$

The MMSE receiver is an example of a *multiuser receiver* which, in contrast to the conventional matched filter, exploits the information about the other users to mitigate their interference.

The capacity analysis problem is this: given users each with its own target SIR requirement, find the mix of users that can be admissible under the MMSE receiver. This problem is currently relevant in the design of future-generation wireless systems. The capability of multiuser receivers needs to be better understood to assess whether the performance gain over the conventional receiver warrants the additional complexity in implementation.

2 The Solution

Although (2) gives a formula for the SIR of the MMSE receiver, the complicated dependence on the signature sequences and received powers of the users makes it difficult to be used directly for capacity analysis. Moreover, the performance certainly depends on which specific sequences are used. We bypass these drawbacks by instead using a *random signature sequence model*. The entries of $\mathbf{s}_i$'s are modeled as i.i.d. zero-mean random variables, with variance normalized to be $1/N$, and are also independent across users. This model provides analytical tractability, and it is often a reasonable approximation in practice as the wireless multipath channel typically randomizes the sequences of the users. Moreover, many CDMA systems use pseudorandom spreading sequences. It should also be noted that the random sequence model is used only for performance analysis purpose; from the point of view of the receiver, the assumption of perfect knowledge of (the realization of) the signature sequences is still retained. In practice, this information is obtained through an adaptive tracking algorithm.

In this model, SIR_1 is a random variable, being a function of the random sequences. The following is the key result, showing that in a large system, SIR_1 converges to a deterministic constant. To simplify the notation, we set $p_1 = 1$ without loss of generality.

Theorem 1 *[1] Let $N, K \to \infty$ such that $\frac{K}{N} \to \alpha$, and assume that the empirical distribution of the users' powers converges to a limiting distribution F. Then SIR_1 converges in probability to β^*, where β^* is the unique solution to the fixed-point equation:*

$$\beta^* = \frac{1}{\sigma^2 + \alpha \int_0^\infty I(p, \beta^*) dF(p)}$$

where

$$I(p, x) \equiv \frac{p}{1 + px}$$

Here, α is the system loading in terms of number of users per degree of freedom. Heuristically, the result says that in a large system,

$$\mathrm{SIR}_1 \approx \frac{1}{\sigma^2 + \frac{1}{N} \sum_{k=2}^K I(p_k, \mathrm{SIR}_1)}$$

Although in general this fixed-point equation has no closed-form solution, it can be shown that the fixed point can be computed numerically by simply iterating from any initial point. More importantly, if user 1 has a target SIR requirement β, then it follows from the monotonicity of the fixed-point equation that its SIR requirement is satisfied asymptotically if and only if

$$\beta \geq \frac{1}{\sigma^2 + \frac{1}{N}\sum_{k=2}^{K} I(p_k, \beta)}. \tag{4}$$

This condition can be checked explicitly.

It is natural to define $I(p, \beta)$ as the *effective interference* of an interferer with received power p on a user with a target SIR β. This summarizes succinctly the effect of an interferer. It is interesting to contrast this to the corresponding result for the matched filter receiver [1]: the effective interference under the matched filter is simply $I_{\text{mf}}(p) = p$. Note that under the matched filter, the effect of an interferer is proportional to its received power. In contrast, under the MMSE receiver, the effect is highly nonlinear in the received power, and ceiling out at $\frac{1}{\beta}$. This is a testimony to the interference suppression capability of the MMSE receiver.

We can apply this result to analyze the capacity of a power-controlled system under the MMSE receiver. Consider a set of users belonging to J classes, with target SIR requirement of β_j for users in class j. Moreover, suppose we can control the received powers of the individual users, but subject to a received power constraint of $\bar{p}_j$ for users in class j. Then it can be shown that α_j users per degree of freedom in class j can be simultaneously supportable under the MMSE receiver if and only if:

$$\sum_{j=1}^{J} \alpha_j \frac{\beta_j}{1+\beta_j} \leq \min_{1\leq j\leq J}\left[1 - \frac{\beta_j \sigma^2}{\bar{p}_j}\right].$$

The set of class mixes $(\alpha_1, \ldots, \alpha_J)$ of users that can be simultaneously supportable is the *capacity region* of the system. The above equation says that the capacity region admits a simple description via a single linear constraint. When there are no power constraints, the right hand side simplifies to 1.

It is now natural to define the *effective bandwidth* of a user with target SIR requirement of β:

$$e(\beta) = \frac{\beta}{1+\beta}.$$

This is the fraction of a degree of freedom occupied by the user. To compute the admissibility of a set of users, one only needs to add up the effective bandwidths of the individual users. The additivity of effective bandwidths follows from the decoupling of the aggregate interference effect into the effective interference of the individual interferers.

Two curious questions arise at this point:

- Why does the SIR converge to a deterministic limit independent of the random signature sequences?
- Why does the interfering effects of the users decouple in such a simple way?

We investigate these two questions in the following sections.

3 Random Matrices

Recall that the SIR of user 1 (with unit received power) is given by:

$$\text{SIR}_1 = \mathbf{s}_1^t (SDS^t + \sigma^2 I)^{-1} \mathbf{s}_1 \tag{5}$$

where S and D are defined in eqn. (3). Observe that S depends only on the signature sequences of the interferers and is hence independent of $\mathbf{s}_1$. Direct computation yields:

$$E[\text{SIR}_1 | S, D] = \frac{1}{N} \mathbf{Tr} \left(SDS^t + \sigma^2 \right)^{-1}.$$

Moreover, it can be shown that the conditional variance goes to zero like $1/N$. Hence,

$$\text{SIR}_1 - \frac{1}{N} \mathbf{Tr} \left(SDS^t + \sigma^2 \right)^{-1} \xrightarrow{\mathcal{P}} 0. \tag{6}$$

We can write:

$$\frac{1}{N} \mathbf{Tr} \left(SDS^t + \sigma^2 I \right)^{-1} = \int_0^\infty \frac{1}{\lambda + \sigma^2} dG_N(\lambda) \tag{7}$$

where λ has the (random) empirical eigenvalue distribution of SDS^t. Thus, the convergence of the SIR hinges on the convergence of the spectrum G_N. This problem has been solved by [2, 3], where they show that the limit exists and moreover does not depend on the distribution of the individual elements of S. The limiting spectrum is described by a functional fixed-point equation for its *Steltjes transform*. The Steltjes transform of a distribution G is defined to be:

$$m(z) := \int \frac{1}{\lambda - z} dG(\lambda)$$

The functional fixed-point equation for the Steltjes transform $m^*(z)$ of the limiting spectrum G^* of SDS^t is given by:

$$m(z) = \frac{1}{-z + \alpha \int \frac{\tau dF(\tau)}{1 + \tau m(z)}} \tag{8}$$

where F is the limiting spectrum of D.

This seems like a complicated characterization of the limiting spectrum. However, we observe from (6) and (7) that what we need to characterize the limiting SIR is precisely the Steltjes transform of the limiting spectrum of SDS^t evaluated at $z = -\sigma^2$. Theorem 1 now follows.

4 Free Probability

The effective interference interpretation follows directly from the random matrix result (8). Let us then take a deeper look at the structure of this equation from a different perspective.

Consider the scenario when there are two groups of interferers $\mathcal{C}_1$ and $\mathcal{C}_2$, one in which the interferers have common received power p and one in which the interferers have common received power q respectively. Suppose there are K_1 users in group $\mathcal{C}_1$ and K_2 users in group $\mathcal{C}_2$, with $K_1/N = \alpha_1$ and $K_2/N = \alpha_2$. The key random matrix SDS^t can be written as an outer sum:

$$SDS^t = p \sum_{i \in \mathcal{C}_1} \mathbf{s}_i \mathbf{s}_i^t + q \sum_{i \in \mathcal{C}_2} \mathbf{s}_i \mathbf{s}_i^t \equiv U_1 + U_2.$$

The asymptotic interfering effect of each group $\mathcal{C}_i$, when present in isolation, depends only on the spectrum of the random matrix U_i. How the overall interfering effect of the two groups can be decoupled into the effects of the individual groups is then a question of how the spectrum of the matrix $U_1 + U_2$ depends on the individual spectra of the matrices U_1 and U_2. This leads to a more general question: if we are given the spectra of two random matrices A and B, what can we say about the spectrum of the sum $A + B$?

For *deterministic* matrices A and B, one cannot in general determine the eigenvalues of $A + B$ from those of A and of B alone, as they depend on the eigenvectors of A and B as well. However, it turns out that for large random matrices A and B satisfying a property called *freeness*, the limiting spectrum of $A+B$ can indeed be determined from the individual spectra of A and B. This is a central result in *free probability theory*, which we very briefly introduce now. For more details, please consult [4] or [5].

Classical probability theory is concerned with commutative random variables. Free probability, on the other hand, deals with non-commutative ones. More formally, a (non-commutative) probability space $(\mathcal{A}, \varphi)$ is an algebra $\mathcal{A}$ over $\mathcal{C}$ with an unit element 1 and endowed with a linear functional, called the *trace*, $\varphi : \mathcal{A} \to \mathcal{C}$, $\varphi(1) = 1$. Elements of $\mathcal{A}$ are called (non-commutative) random variables. Classical probability theory is obtained when the algebra $\mathcal{A}$ consists of scalar random variables and the functional φ is the standard expectation operator. The focus of the theory however is on random variables which do not commute.

The distribution of a random variable $X \in \mathcal{A}$ is specified by the moments $\varphi(X^k)$, for $k \geq 1$. The distribution defined in this form allow us to compute the expectation of any function of the random variable X that can be approximated by polynomials. Similarly, the *joint distribution* of a collection of random variables $X_1, \ldots, X_m \in \mathcal{A}$ is specified by all the joint moments $\varphi(X_{i_1} \ldots X_{i_p})$, $p \geq 1$.

The central notion in classical probability theory is *independence*. In the notations introduced above, two scalar random variables X, Y are independent if $\varphi(X^k Y^l) = \varphi(X^k)\varphi(Y^l)$ for all k, l. The analogous notion in free probability theory is *freeness.*

Definition 2 *Let X and Y be two random variables. Let $\mathcal{A}_1$ be the sub-algebra generated by X and the unit 1 and $\mathcal{A}_2$ be the sub-algebra generated by Y and the unit 1 respectively, i.e. they consist of polynomials of X and of Y respectively. The random variables X and Y are* free *if $\varphi(Z_1 \ldots Z_m) = 0$ whenever $\varphi(Z_k) = 0$ for all $k = 1, \ldots, m$ and $Z_k \in \mathcal{A}_{i(k)}$ where consecutive indices $i(k) \neq i(k+1)$ are distinct.*

For independent random variables, the joint distribution can be specified completely by the marginal distributions. For free random variables, the same result can be proved, directly from definition. In particular, if X and Y are free, then the moments $\varphi((X+Y)^n)$ of $X+Y$ can be completely specified by the moments of X and the moments of Y. The distribution is naturally called the *free convolution* of the two marginal distributions. Classical convolution can be computed via transforms: the log moment generating function of the distribution of $X+Y$ is the sum of the log moment generating functions of the individual distributions. For free convolution, the appropriate transform is called the R-transform. This is defined via the Steltjes transform.

Given a random variable X, let

$$m_X(z) = E\left[\frac{1}{X-z}\right]$$

be the Stieltjes transform of its distribution. Let m_X^{-1} be the inverse of m_X. The R-transform is defined as:

$$R(\beta) := \frac{1}{\beta} + m_X^{-1}(\beta).$$

Theorem 3 *If X and Y are two free random variables, then $R_{A+B} = R_A + R_B$.*

A good example of non-commutative random variables are random matrices. Let $\mathcal{A} = M_N$ be the algebra of complex N by N random matrices whose entries are scalar random variables defined on some underlying common probability space. The trace is defined as:

$$\varphi_N(X) := \frac{1}{N} E\left[\mathbf{Tr} X\right].$$

For any N by N random Hermitian matrix $X \in M_N$ with random eigenvalues $\lambda_1, \ldots \lambda_N$, the rth moment of X in the non-commutative probability space (M_N, φ_N) is given by

$$\varphi_N(X^r) = \frac{1}{N} E[\mathbf{Tr} X^r] = \frac{1}{N} E\left[\sum_{i=1}^{N} \lambda_i^r\right]$$

If we let $F_X(\cdot)$ be the expected empirical distribution of the eigenvalues of X:

$$F_X(\lambda) := \frac{1}{N} E\left[|\{i : \lambda_i \leq \lambda\}|\right]$$

then the moments of the distribution F_X are precisely the moments of X as a non-commutative random variable.

The deep connection between random matrices and free probability is first made by Voiculescu [6], who showed that in a lot of cases of interest, large random matrices become asymptotically free. This answers the question we posed earlier: the limiting spectra of the sum of two large random matrices is the free convolution of the limiting spectra of the individual matrices, if they become asymptotically free.

Returning to our specific problem, it can be shown as a consequence of Voiculescu's results that the matrices $U_1 = p\sum_{i\in\mathcal{C}_i}\mathbf{s}_i\mathbf{s}_i^t$ and $U_2 = \sum_{i\in\mathcal{C}_2}\mathbf{s}_i\mathbf{s}_i^t$ are in fact asymptotically free as $N, K \to \infty, K/N \to \alpha$. . The R-transforms can be computed explicitly for this problem:

$$R_{U_1}(\beta) = \alpha_1\frac{p}{1+p\beta}, \qquad R_{U_2}(\beta) = \alpha_2\frac{q}{1+q\beta},$$

and

$$R_{U_1+U_2}(\beta) = R_{U_1}(\beta) + R_{U_2}(\beta).$$

What is the connection of this to the notion of effective interference? Observe that the R-transform satisfies:

$$m(z) = \frac{1}{-z + R[m(z)]}.$$

Putting in $z = -\sigma^2$ and noting that $m(-\sigma^2)$ is the limiting SIR, we conclude that at target SIR requirement of β, $R_{U_i}(\beta), i = 1,2$, is the effective interference of group $\mathcal{C}_i$ when considered in isolation, and $R_{U_1+U_2}(\beta)$ is the limiting aggregate interference of the two groups when both are present. **The decoupling of effective interference is nothing but the additivity of the R-transforms of free random matrices.**

5 Back to Multiuser Receivers

The previous section re-interprets the structure of the fixed-point equation in Theorem 1 in terms of notions from free probability. The proof of Theorem 1 itself does not require free probability, as the structure of the relevant random matrix SDS^t is a classical one for which a limit theorem already exists. However, the analysis of more sophisticated CDMA models [7, 8] led to non-classical random matrix structure, for which new limit theorems had to be proved. Free probability is a valuable tool for solving these problems. We discuss one such problem here [8].

One detrimental effect of a wireless link is the random *fading* of the channel due to constructive and destructive interference between multiple signal paths from the transmitter to the receiver. A countermeasure to alleviate this problem is the use of multiple antennas at the receiver. Provided that the antennas are placed sufficiently far apart, the received signals at the antennas fade independently and the probability is high that at least one of the antennas receives a strong copy of the transmitted signal. The following is a model of a direct sequence CDMA system with 2 receive antennas:

$$\mathbf{y}(l) = \sum_{k=1}^{K}\gamma_k(l)x_k\mathbf{s}_k + \mathbf{w}(l) \in \mathcal{C}^N, \qquad l = 1,2.$$

where as before, $x_k, \mathbf{s}_k$ are the transmitted symbol and signature sequence of the kth user respectively, $\mathbf{y}(l)$ is the received signal at antenna l, and $\gamma_k(l)$ is the random channel gain from user k to an antenna l.

As in the basic model, we can consider the MMSE estimator for transmitted symbol x_1, based now on $\mathbf{y}(1)$ and $\mathbf{y}(2)$, the received signals at both of the antennas. In place of SDS^t for the basic model, the key random matrix for this problem is $\bar{S}\bar{S}^\dagger$, where

$$\bar{S} = \begin{bmatrix} SB_1 \\ SB_2 \end{bmatrix}$$

and $S = [s_2 \dots s_k]$ and $B_l = \text{diag}(\gamma_2(l), \dots, \gamma_K(l))$.

There is strong dependency between entries in the random matrix $\bar{S}$, due to the repetition of the signature sequences at the two receive antennas. Existing results on spectral convergence do not apply. However, we observe that $\bar{S}\bar{S}^\dagger$ and $\bar{S}^\dagger\bar{S}$ have the same nonzero eigenvalues, and the latter can be written as a sum of two matrices:

$$\bar{S}^\dagger\bar{S} = B_1^\dagger S^t S B_1 + B_2^\dagger S^t S B_2$$

We have the following theorem.

Theorem 4 *[1] If B_1, B_2 are independent and zero mean, then the matrices $B_1^\dagger S^t S B_1$ and $B_2^\dagger S^t S B_2$ are asymptotically free as $N, K \to \infty, K/N = \alpha$.*

Moreover, it can be shown that if S_1 and S_2 are two independent matrices, each having the same distribution as S, then $B_1^\dagger S_1^t S_1 B_1$ and $B_2^\dagger S_2^t S_2 B_2$ are also asymptotically free. The implication is that the limiting spectrum of $\bar{S}^t\bar{S}$ is the same as if independent copies of the signature sequences were received at the different antennas. For this latter "independent sequence" model, existing random matrix results can be applied and the performance of the MMSE receiver can be analyzed. More details can be found in [9] in these proceedings. The main point to be made here is that the important notion is freeness and not independence: although there is strong dependence between the entries of the matrix $\bar{S}$, the randomness in D_1 and D_2 is enough to make the components free. This result translates into the engineering conclusion that there is asymptotically no loss of degrees of freedom for communication even though the signature sequences are repeated at each of the receive antennas.

6 Conclusions

In this paper, we develop interesting connections between the engineering problem of capacity analysis of multiuser receivers on the one hand, and random matrices and free probability theory on the other. The connections are unexpected and deep. Is there anything fundamental relating the least square estimation problem and free probability, or is the connection purely coincidental? The reader is encouraged to look further.

References

[1] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, Mar. 1999.

[2] V. A. Marcenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sb.*, vol. 1, pp. 457–483, 1967.

[3] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[4] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables.* CRM Monogragh Series, Volume 1, Providence, Rhode Island, USA: American Mathematical Society, 1992.

[5] D. Voiculescu, "Lectures on free probability theory," 1998. Notes for a Course at the Saint-Flour Summer School on Probability Theory (unpublished).

[6] D. Voiculescu, "Limit laws for random matrices and free products," *Inventiones mathematicae*, vol. 104, pp. 201–220, 1991.

[7] J. Evans and D. Tse, "Linear multiuser receivers for multipath fading channels," *Submitted to IEEE Trans. Inform. Theory*, 1999.

[8] S. Hanly and D. Tse, "Resource pooling and effective bandwidths in CDMA systems with multiuser receivers and spatial diversity," *Submitted to IEEE Trans. Inform. Theory*, 1999.

[9] S. Hanly and D. Tse, "Resource pooling and effective bandwidths for CDMA antenna arrays," in *Proc. Allerton Conference on Communications, Computing and Control, Monticello, Illinois*, Sept. 1999.