Applications of Large Random Matrices and Free Probability Theory in Wireless Communications

Shahab Sanayei

Multimedia and Communications Laboratory

University of Texas at Dallas

August 2004

Overview(1)

- Motivations
- Modern Communications Techniques
  - Wireless Link
    - Multiple Antenna
  - Wireless Network
    - Multiuser Detection
    - CDMA systems

Overview (2)

- Large Random Matrices:
  - Asymptotic Analysis
    - Large Number of Users (CDMA Systems,Capacity of Ad hoc Networks)
    - Large Number of Antennas (MIMO Capacity)
  - System Design
    - Reduced Rank Multiuser Detection
- Open Problems
- Conclusion

Motivations(1)

- More demand for higher data rates and increased mobility (eg for multimedia communications)
- More data + More users = More Complexity
- Complexity does not only affect the hardware design it also affects analysis and system level design
- In a network with large number of users, macroscopic behavior of the system can be an index of how well the systems is operating
- Microscopic statiscal modeling can be used to predict macroscopic behavior and properties of the system

Motivations(2)

Inherent Features of Wireless Networks

- Large number of users (LAN$\approx 10 - 100$, WAN≈ 1000,..., Internet$\approx 10^8$!!)
- Randomness
  - Additive Noise
  - Multi-path Fading
  - Shadowing

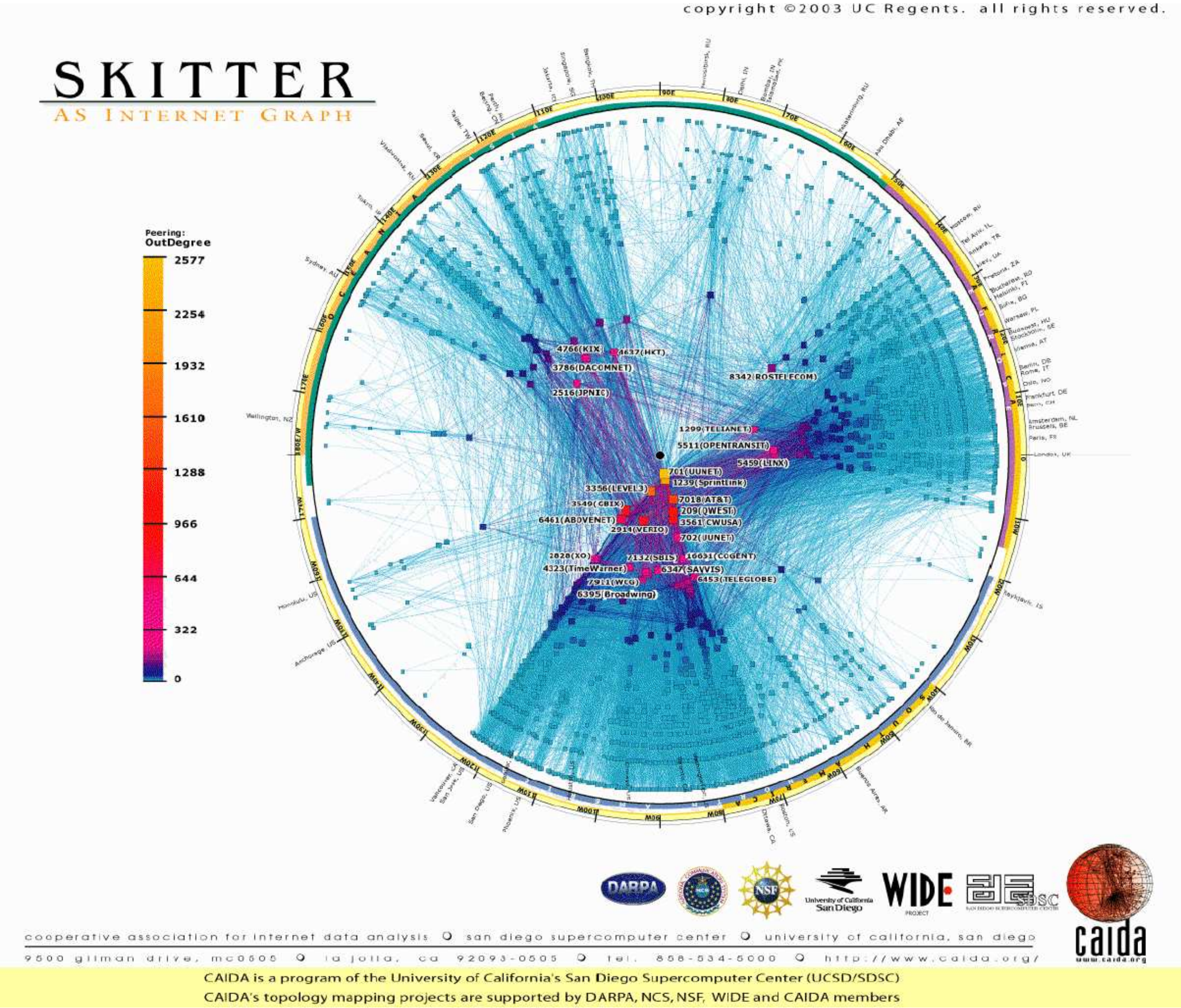
Communication networks are large random non-commutative objects!

Motivation (3)

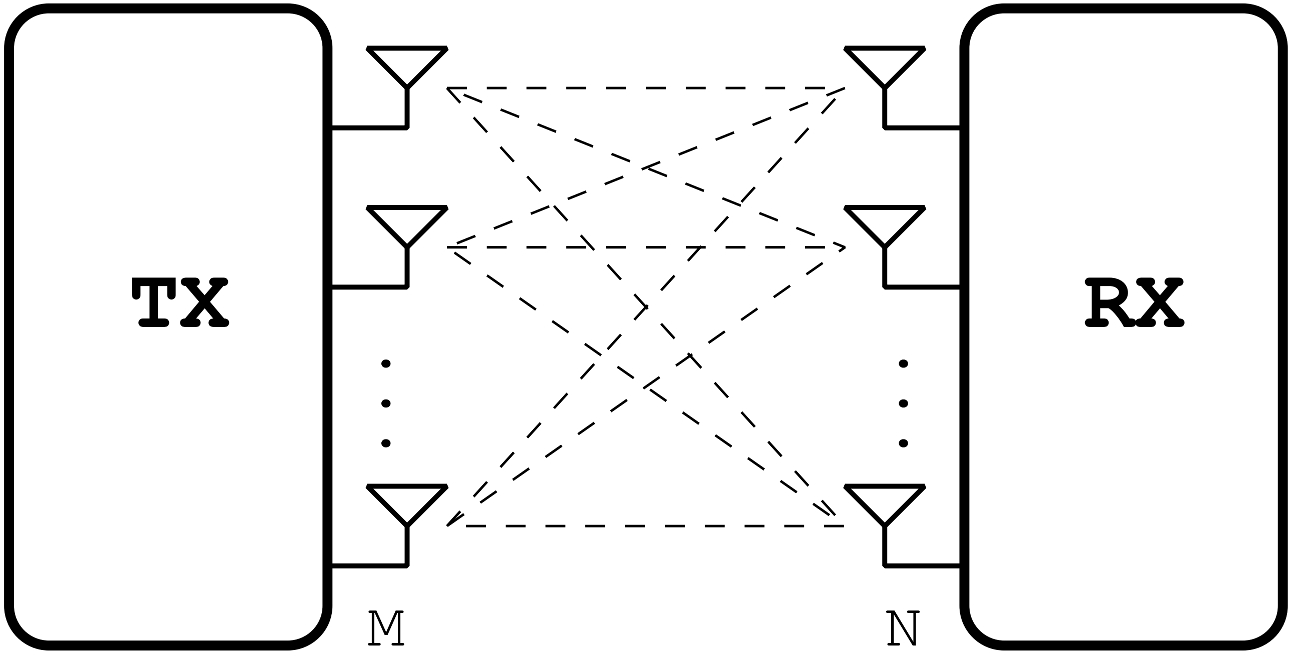
and above all ...

Spectral Efficiency =

A Global Map of Internet



MIMO System



Signal Model

$$y(t) = \sqrt{\frac{\rho}{M}} Hx(t) + n(t)$$

- $x \in \mathbb{C}^M$, $y \in \mathbb{C}^N$
- Rayleigh Fading Model: $H_{N \times M}$ with i.i.d. complex Gaussian elements
- $h_{ij} \sim \mathcal{CN}(0, 1)$
- $n(t)$ additive white Gaussian noise, $n_i(t) \sim \mathcal{CN}(0, 1)$
- H and $n(t)$ are independent
- ρ is the transmitted power and $\mathbb{E}[x^* x] \leq 1$

Capacity of MIMO (1)

$$\mathcal{I}(X;Y|H) = \log\det(I + \frac{\rho}{M}HH^*) = \sum_{i=1}^{m}\log(1+\frac{\rho}{M}\lambda_i)$$

where $m = \min(M, N)$ and λ_i's are the non-zero eigen values of HH^*.

- Ergodic Capacity of MIMO channel:

  $$C = \mathbb{E}_H[\mathcal{I}(X;Y|H)] = m\mathbb{E}_\lambda[\log(1+\frac{\rho}{M}\lambda)]$$

  where λ is an unordered eigen value of H.

- for H with iid Gaussian elements, the distribution of λ can be explicitly calculated in terms of Laguerre Polynomials.

Capacity of MIMO (2)

- But in general, the distribution of λ is may not be easy to calculate
- If $\frac{M}{N} \rightarrow \beta$ then distribution of the normalized eigen value of HH^* obeys the *Marchenko-Pastur* law:

  $$f_\beta(x) = (1-\beta)^+\delta(x) + \frac{\sqrt{(x-a)^+(x-b)^+}}{2\pi\beta x}$$

  where, $a = (1-\sqrt{\beta})^2$, $b = (1+\sqrt{\beta})^2$ and $(x)^+ = \max(x,0)$
- Thus:

  $$C \longrightarrow m\int_a^b \log(1+\frac{\rho}{\beta}x)dF_\beta(x)$$

  as $M, N \rightarrow \infty$ with $\frac{M}{N} \rightarrow \beta$.

Capacity of MIMO (3)

- The Marchenko-Pastur law is valid even if the distribution of the elements of H are not Gaussian, they only need to have identical mean and variance and bounded forth moment (even a weaker Lindeberg type condition is enough).
- The rate of convergence is fast so even for not so large M and N, this still gives an acceptable approximation.

Capacity of Correlated MIMO Channel (1)

- Correlation model:
  - Row correlation:
    $$H = R^{1/2}\widetilde{H}$$
    where $R = \mathbb{E}[HH^*]$
  - Column correlation:
    $$H = \widetilde{H}R^{1/2}$$
    where $R = \mathbb{E}[H^*H]$
- $\widetilde{H}$ is a matrix whose elements are i.i.d.

Capacity of Correlated MIMO Channel (2)

- Ergodic capacity for receiver correlation (Mestre et al 03):

$$\mathbb{E}[\log\det(I + \frac{\rho}{M} H H^T)] = m\mathbb{E}[\log(1 + \rho\frac{N}{M}\lambda)]$$

  where λ is an unordered nonzero eigen value of $R\widetilde{H}\widetilde{H}^*$

- For large number of antennas, we need to calculated the spectral measure of $R\widetilde{H}\widetilde{H}^*$
- We use free probability theory!

$\mathcal{S}$ Transform (1)

- $\mathcal{S}$-transform:
$$\mathcal{S}(z) = \frac{1+z}{z}\psi^{-1}(z)$$
where ψ^{-1} is the inverse of
$$\psi(z) = z^{-1}m(z^{-1}) - 1$$
- When R and $\frac{1}{N}HH^*$ are free we have:
$$\mathcal{S}_{\frac{1}{N}HH^*}(z) = \mathcal{S}_R(z)\mathcal{S}_{\frac{1}{N}\widetilde{H}\widetilde{H}^*}(z)$$
- we know $\frac{1}{N}\widetilde{H}\widetilde{H}^*$ obeys the Marchenko-Pastur law for which
$$\mathcal{S}_{\frac{1}{N}\widetilde{H}\widetilde{H}^*}(z) = \frac{1}{\beta+z} \quad , \quad \frac{M}{N} \to \beta$$

$\mathcal{S}$ Transform(2)

- Tilted semi-circle law:

$$f_R(x) = \frac{1}{\pi \mu x^2} \sqrt{\left(\frac{x}{\sigma_1} - 1\right)\left(1 - \frac{x}{\sigma_2}\right)}\mathbf{1}_{[\sigma_1,\sigma_2]}(x)$$

where, $\mu = \frac{(\sqrt{\sigma_2} - \sqrt{\sigma_1})^2}{4\sigma_1\sigma_2}$

- This is a good approximation for practical purposes
- σ_1 and σ_2 are determined via data fitting
- $\mathcal{S}$-Transform (under normalization assumption $\sigma_1\sigma_2 = 1$)

$$\mathcal{S}_R(z) = 1 - \mu z \quad \Rightarrow \quad \mathcal{S}_{\frac{1}{N}HH^*} = \frac{1 - \mu z}{\beta + z}$$

$\mathcal{S}$ Transform(3)

- Stieltjes transform:

$$m(z) = \frac{z + 2z\mu + 1 - \beta + \sqrt{[z - (1+\beta)]^2 - 4\beta(1+\mu z)}}{2z(1+\mu z)}$$

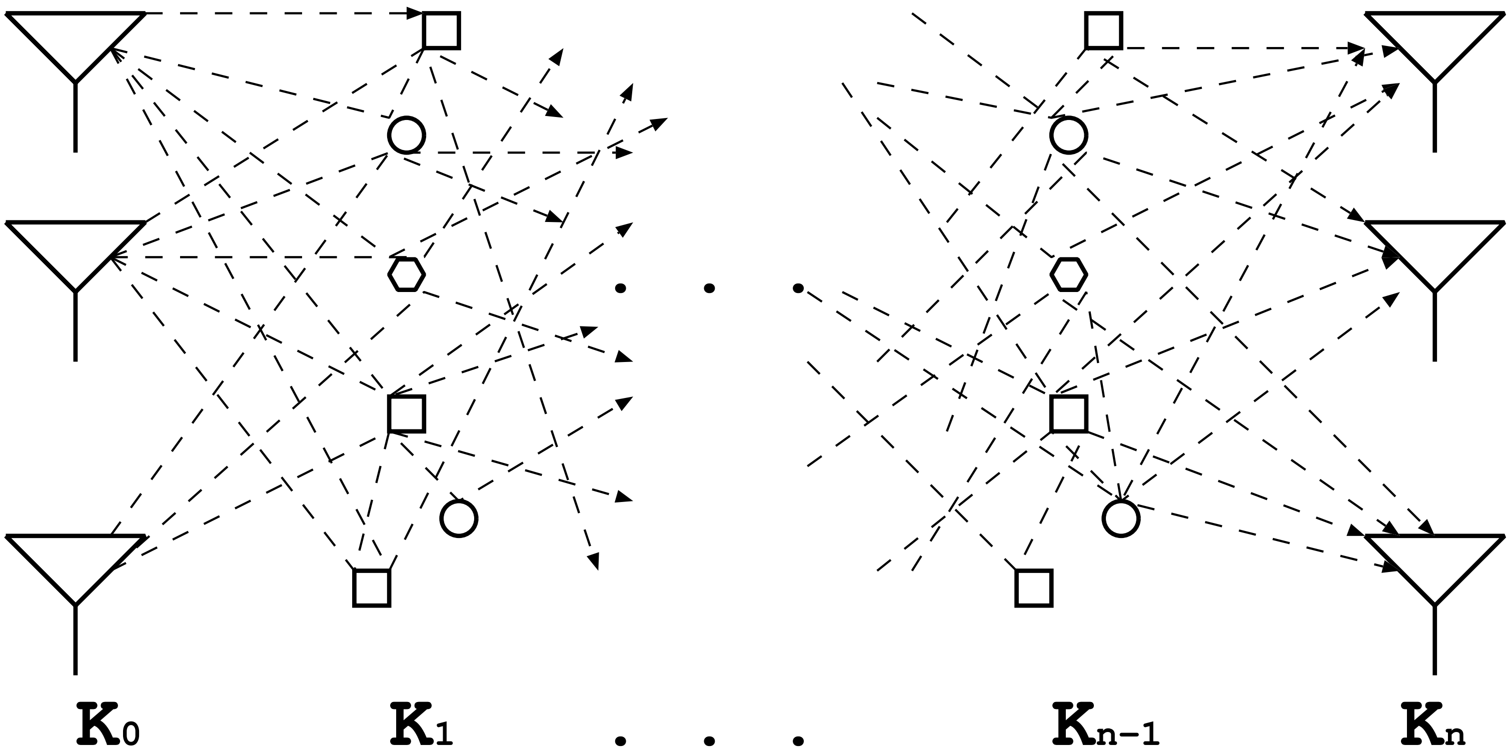
- Using inverse Stieltjes:

$$f(x) = \max\{0, 1-\beta\}\delta_0(x) + \frac{\sqrt{(x-\xi_1)(\xi_2-x)}}{2\pi x(1+\mu x)}\mathbf{1}_{[\xi_1,\xi_2]}(x)$$

$$\xi_1 = 1 + \beta + \frac{\mu}{\beta} - 2\sqrt{\beta}\sqrt{(1+\mu)(1+\mu\beta)}$$
$$\xi_2 = 1 + \beta + \mu\beta + 2\sqrt{\beta}\sqrt{(1+\mu)(1+\mu\beta)}$$

MIMO channel Modeling

Cluster of Scatterers:



Multiplicative MIMO Model (1)

- Multiplicative model (Mueller 01):

$$\mathbf{H}_N = \mathbf{M}_N \mathbf{M}_{N-1} \ldots \mathbf{M}_2 \mathbf{M}_1 = \prod_{n=1}^{N} \mathbf{M}_n$$

  where $\mathbf{M}_i$ is a $K_{i-1} \times K_i$ matrix with i.i.d. elements

- We assume that the family $\{\{\mathbf{M}_1^* \mathbf{M}_1\}, \{\mathbf{M}_2^* \mathbf{M}_2\}, \ldots, \{\mathbf{M}_N^* \mathbf{M}_N\}\}$ is asymptotically free and for all n, $K_n \to \infty$, but $\beta_n \overset{\triangle}{=} \frac{K_{n-1}}{K_n}$ is constant.

Multiplicative MIMO Model (2)

- Random covariance matrix:

$$\mathbf{C}_i \overset{\triangle}{=} \mathbf{H}_i^* \mathbf{H}_i = \left(\prod_{n=1}^{i} \mathbf{M}_n \right) \left(\prod_{n=1}^{i} \mathbf{M}_n \right)^*$$

- To calculate the capacity we need to find the asymptotic spectral distribution of $\mathbf{C}_N$ which is possible by recursive application of the $\mathcal{S}$-transform
- let

$$\widetilde{\mathbf{C}}_i = \left(\prod_{n=1}^{i-1} \mathbf{M}_n \right) \left(\prod_{n=1}^{i-1} \mathbf{M}_n \right)^* \mathbf{M}_i^* \mathbf{M}_i$$

Multiplicative MIMO Model (3)

- The spectral measure of $\mathbf{C}_i$ and $\widetilde{\mathbf{C}}_i$ are the same
- Recursion: $\widetilde{\mathbf{C}}_{i+1} = \mathbf{C}_i \mathbf{M}_{i+1}^* \mathbf{M}_{i+1}$
- $\mathcal{S}$ transform:
  - $\mathcal{S}_{\mathbf{M}_i \mathbf{M}_i^*}(z) = \frac{1}{z+\beta_i}$
  - $\mathcal{S}_{\mathbf{M}_i^* \mathbf{M}_i}(z) = \frac{1}{1+z\beta_i}$
- Also:

$$\mathcal{S}_{\mathbf{C}_i}(z) = \frac{z+1}{z+\beta_i} \mathcal{S}_{\widetilde{\mathbf{C}}_i}\left(\frac{z}{\beta_i}\right)$$

Multiplicative MIMO Model (4)

- It can be shown (Meuller 01)

$$\mathcal{S}_{\mathbf{C}_n}(z) = \prod_{n=1}^{N} \frac{\rho_n}{z + \rho_{n-1}}$$

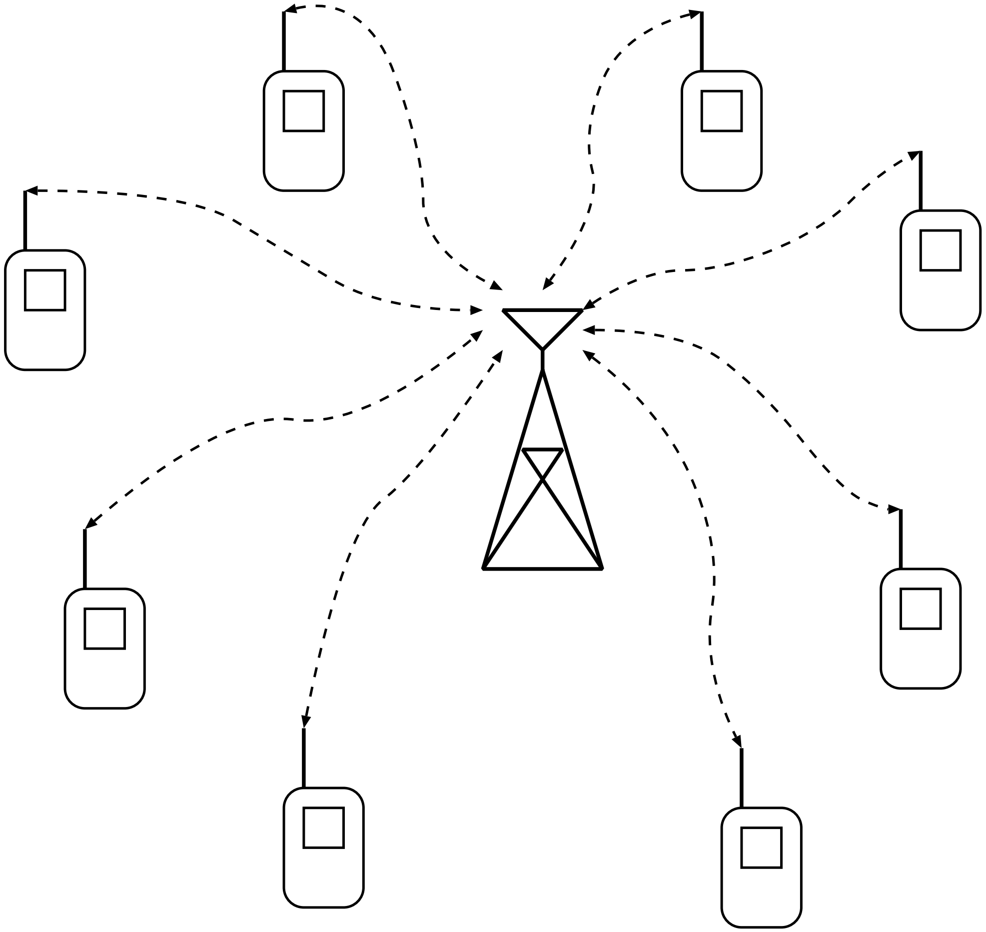
where $\rho_n = \frac{K_n}{K_N}$

- For the Stieltjes transform we have:

$$m_{\mathbf{C}_N}(z) \prod_{n=1}^{N} \frac{\rho_{n-1} - 1 - z m_{\mathbf{C}_N}(z)}{\rho_n} - z\, m_{\mathbf{C}_N}(z) = 1$$

- This can not be solved explicitly but can be used to fine the moments of asymptotic eigen value distribution.

Multiuser Systems



DS-CDMA Multiuser System

- Direct Sequence Code Division Spread Spectrum Multiple Access (DS-CDMA): For each user, every transmitted symbol is multiplied by a predetermined vector, called *the signature*
- In a CDMA network, all user transmit at the same time, hence the signal received at the central receiver is:

$$y = \sum_{i=1}^{K} x_k s_k + w \in \mathbb{C}^N$$

  where K is the number of active users in the network and N is the length of spreading sequence. $w \sim \mathcal{CN}(\mathbf{0}, \sigma^2 I_N)$.

- Assumption $\mathbb{E}[x_k] = 0$, $\mathbb{E}[|x_k|^2] = p_k$, p_k is the received power of user k
- N is called *processing gain*.

MMSE Multiuser Receiver (1)

- Objective: For user k, estimate x_k, with minimum estimation error
- We restrict ourselves to linear receivers, i.e. $\hat{x}_k = c_k^* y$
- $c_k = s_k$ is called the *matched filter* receiver.
- Matched is simple to implement, but it's not the best we can do with a linear receiver
- Minimizing the mean square error $\mathbb{E}[|\hat{x}_k - x_k|^2]$ leads to

  $$c_k = (SPS^* + \sigma^2 I)^{-1} s_k$$

  this is called the *MMSE receiver* and is the solution to a regularized least squares problem. $S = [s_1 \ldots s_{k-1}\ s_{k+1} \ldots s_K]$ and $P = \mathrm{diag}(p_1, \ldots, p_{k-1}, p_{k+1}, \ldots, s_K)$.

MMSE Multiuser Receiver (2)

- In multiuser system *signal-to-interference ratio* (SIR) for the k^{th} user is defined as:

$$SIR_k = \frac{p_k |c_k^* s_k|^2}{\sigma^2 c_k^* c_k + \sum_{i \neq k} p_i |c_i^* s_i|^2}$$

- for the MMSE receiver we have:

$$SIR_k = p_k s_k^* (SPS^* + \sigma^2 I)^{-1} s_k$$

- Unlike the matched filter receiver, the MMSE receiver exploits the information about the other users to mitigate their interference in an optimal way.

A Framework for Analysis of MMSE Receiver

- The exact analysis of MMSE receiver is difficult, due to dependence on the signature sequences and received powers of users.
- We assume a random signature sequence model. The elements of s_k are modeled as iid zero-mean r.v.'s with variance normalized to be $\frac{1}{N}$ and also dependent across different users.
- Definition: *system loading* (number of users per degree of freedom) is defined as: $\alpha = \frac{K}{N}$

Asymptotic Analysis of MMSE receiver

- **Theorem** (Tse 99): If $N, K \to \infty$ with $\frac{K}{N} \to \alpha$, also if the empirical distribution of the users' powers converges to a limiting distribution F, then $SIR_k \xrightarrow{i.p.} \rho$ where ρ is the solution to the following fixed-point equation:

$$\rho = \frac{1}{\sigma^2 + \alpha \int_0^\infty I(x,\rho) dF(x)}$$

where $I(x,\rho) = \frac{x}{1+\rho x}$.

- Large system approximation:

$$SIR_k \approx \frac{1}{\sigma^2 + \frac{1}{N} \sum_{i \neq k}^{K} I(p_i, SIR_i)}$$

Two Questions

- Why does SIR converges to a deterministic limit regardless of the distribution of the signature matrix?
- Why does the interfering effect of the users decouple in such a simple way?
- Answer: Concentration of the Spectral Measure for Large Random Matrices

Concentration of the Spectral Measure

- We have, $\mathbb{E}[SIR_k|S,P] = \frac{1}{N}\mathsf{Tr}(SPS^* + \sigma^2 I_N)^{-1}$
- It can be shown that: $\mathsf{var}(SIR_k|S,P) \to 0$
- Hence $SIR_k - \frac{1}{N}\mathsf{Tr}(SPS^* + \sigma^2 I_N)^{-1} \xrightarrow{i.p.} 0$
- But $\frac{1}{N}\mathsf{Tr}(SPS^* + \sigma^2 I_N)^{-1} = \int_0^\infty \frac{dG_N(\lambda)}{\lambda+\sigma^2}$ where $G_N(\cdot)$ is the empirical distribution of the matrix SPS^*
- as $N \to \infty$, $G_N \xrightarrow{D} G$
- The limiting measure, G, can be calculated using the Stieltjes transform.

Stieltjes Transform

- Stieltjes Transform:

$$m(z) = \int_0^\infty \frac{dG(\lambda)}{\lambda - z}$$

- Functional fixed point equation for the Stieltjes transform:

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

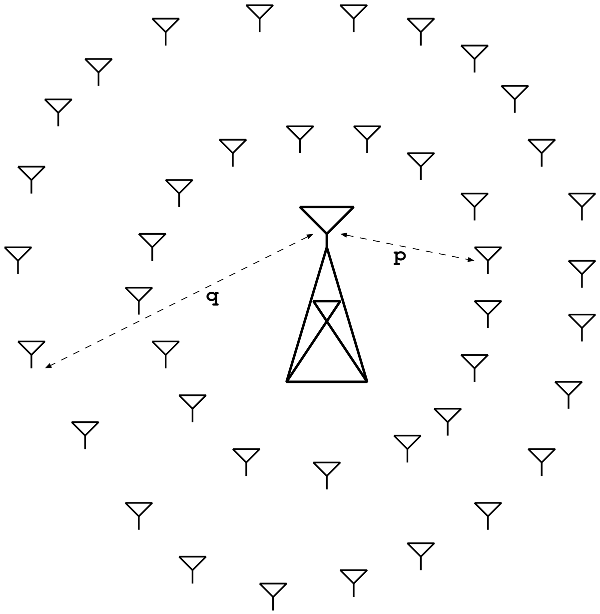
where F is the limiting distribution of P_k.

- Finally:

$$SIR_k \to m(-\sigma^2)$$

Free Probability Theory (1)

- Now let's assume there are two groups of users:
  - $\mathcal{C}_1$: Those who are close to base-station with a common received power of p
  - $\mathcal{C}_2$: Those who are far from the base-station with a common received power of q $(q < p)$.



Free Probability Theory (2)

- Also: $\beta_1 = \frac{K_1}{N}$ and $\beta_2 = \frac{K_2}{N}$
- The key random matrix SPS^* can be written as:

$$SPS^* = p \sum_{i \in \mathcal{C}_1} s_i s_i^* + q \sum_{i \in \mathcal{C}_2} s_i s_i^* = U_1 + U_2$$

- to calculate SIR_k we need to find the spectral density of $U_1 + U_2$
- Use the $\mathcal{R}$-Transform!

$\mathcal{R}$ Transform (1)

- The $\mathcal{R}$ Transform:

$$\mathcal{R}(s) = s^{-1} + m^{-1}(s)$$

- We know from free probability theory that if A and B are free, then

$$\mathcal{R}_{A+B} = \mathcal{R}_A + \mathcal{R}_B$$

$\mathcal{R}$ Transform (2)

- Also $U_1 = p \sum_{i \in \mathcal{C}_1} s_i s_i^*$ and $U_2 = q \sum_{i \in \mathcal{C}_2} s_i s_i^*$ are asymptotically free as $N, K_1, K_2 \to \infty$ with $\frac{K_1}{N} \to \beta_1$ and $\frac{K_2}{N} \to \beta_2$.
- The $\mathcal{R}$ transform can be calculated as:

$$\mathcal{R}_{U_1}(s) = \frac{\beta_1 p}{1 + ps} \quad , \quad \mathcal{R}_{U_2}(s) = \frac{\beta_2 q}{1 + qs}$$

- Thus:

$$\mathcal{R}_{U_1+U_2}(s) = \mathcal{R}_{U_1}(s) + \mathcal{R}_{U_2}(s)$$

Decoupling of Interference

- Effective Interference:

$$m(-z) = \frac{1}{z + \mathcal{R}(m(-z))}$$

- Hence $\mathcal{R}(m(-\sigma^2))$ is the effective interference!
- This implies that by adding two groups of users $\mathcal{C}_1$ and $\mathcal{C}_2$ the equivalent interference adds! in other words there is no coupling effect between two groups of users!
- The **decoupling** of effective interference is nothing but the **additivity** of the $\mathcal{R}$-transforms of free random matrices.

Reduced Rank Multiuser Design (1)

- Multiuser detection involved inverting large matrices in real-time
- This might not be affordable for hand-held devices with limited power
- Approximate Solution: Reduced Rank Design

Reduced Rank Multiuser Design (2)

- Due to Cayley-Hamilton theorem, we have:

$$A^{-1} = \sum_{i=0}^{K-1} w_i A^i$$

- MMSE detection:

$$\hat{x} = s_k^*(A + \sigma^2 I)^{-1} y = s_k^* \sum_{i=0}^{K-1} w_i A^i y$$

  where $A = SPS^*$

- Approximate reduced rank solution: Projection onto the Krylov subspace = span$\{y \; Ay \; A^2y \; \ldots \; A^L y\}$ with $L \ll K$

Reduced Rank Multiuser Design (3)

- The optimal weights w_i can be obtained from a system of Yule-Walker equations:

$$\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_L \end{bmatrix} = \begin{bmatrix} m_2 & m_3 & \dots & m_{L+1} \\ m_3 & m_4 & \dots & m_{L+2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{L+1} & m_{L+2} & \dots & m_{2L} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{L-1} \end{bmatrix}$$

where $m_k = \frac{1}{K}\sum_{i=1}^{K} \lambda_i^k(A) = \frac{1}{K}\mathsf{Tr}(A^k)$.

Reduced Rank Multiuser Design (4)

- The weights can also be calculated adaptively by interpreting the reduced rank detector as a multi-stage Weiner filter.
- An easer solution: For large scale systems, m_k can be calculated from the asymptotic distribution of the eigen values of A which does not require solving of Yule-Walker equation for every A.

Other Applications

- Channel Estimation (Tse 00)
- Capacity of Ad hoc Networks (Telatar 03)
- Learning and Neural Networks (Solla 99)
- Queuing Analysis (Baryshnikov 00)
- Principle Component Analysis (Hoyle 04)
- Spectral Properties of Random Graphs (Chung 03)
- Computational Biology (Zee 01)

Some Open Problems

- Spectral properties of structured matrices (Banded, (block) Hankel, (block) Toeplitz, ...)
- Capacity of Frequency Selective MIMO (channel with memory) $\Rightarrow$ asymptotic spectral density of block Toeplitz matrices
- PAR (peak to average ratio) in OFDM systems $\Rightarrow$ maximum singular value of (block) Hankel matrices
- Capacity of doubly selective channel $\Rightarrow$ asymptotic spectral density of banded matrices

Some Open Problems (2)

- H is a structured matrix, quantities of interest:
  - freeness?
  - $F_\lambda(HH^*) \rightarrow\ ?$
  - $\lambda_{max}(HH^*) \rightarrow\ ?\ ,\ \ \lambda_{max}(HH^*) = \mathcal{O}(?)$
  - $\lambda_{min}(HH^*) \rightarrow\ ?\ ,\ \ \lambda_{min}(HH^*) = \mathcal{O}(?)$
  - $\log\det(HH^*) \rightarrow\ ?\ ,\ \ \log\det(HH^*) = \mathcal{O}(?)$

Banded Random Matrices

- The channel matrix $\mathcal{H}$ has the following banded structure:

$$\begin{bmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,\nu} & 0 & \ldots & & 0 \\ 0 & h_{2,1} & h_{2,2} & \ldots & h_{2,\nu} & 0 & \ldots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & & 0 \\ 0 & 0 & \ldots & 0 & h_{N,1} & h_{N,2} & \ldots & h_{N,\nu} \end{bmatrix}_{N \times N+\nu}$$

- All non-zero elements of $\mathcal{H}$ are iid and $\sim \mathcal{N}(0,1)$
- Fix ν (eg $\nu = 2$) $\Rightarrow$ Asymptotic eigen value distribution of $\frac{1}{N}\mathcal{H}\mathcal{H}^*$ **?**
- $L_N = \log\det(\mathcal{H}\mathcal{H}^*) \implies \frac{L_N - a_N}{b_N} \longrightarrow \mathcal{N}(0,1)$
  then, $a_N = \mathbf{?}$, $b_N = \mathbf{?}$

A Conjecture

- $B = A + XTX^*$
- Asymptotic eigen distribution (Silverstein)
  - General A, but diagonal T
  - $A = 0$, $T \succ 0$
- What if A is general and T non-diagonal?
  - If matrices are free ⇒ free probability
  - Conjecture (Mueller 04) : even if the matrices are not free the same law still holds! ⇒ proof?

Relation to Least Squares (Tse 00)

- We previously noticed that $\mathcal{R}$-transform relates the amount of uncertainty in the observations to the mean square error.
- Is that just a coincidence? is there any deeper connection between freeness and least squares technique?

Some Statistics

- More than 100 papers since 1998 with exponential growth
- "Free Probability" ⇒ Google Hits= 3560
- "Free Probability" AND "Communication"
  ⇒ Google Hits= 369
- about 10% and increasing...

Conclusion

- Communications Systems
- Many applications: Capacity analysis, CDMA analysis and design, estimation, etc...
- Open problems emerging from the engineering applications
- Large random matrices and free probability are important analytical tools for better understanding and more efficient design of the next generation of communication systems

Thank You!