Security Level:

Scientific Challenges of 5G

Mérouane Debbah

Mathematical and Algorithmic Sciences Lab

Huawei

www.huawei.com

HUAWEI TECHNOLOGIES CO., LTD.

HUAWEI

5G-4G=1G

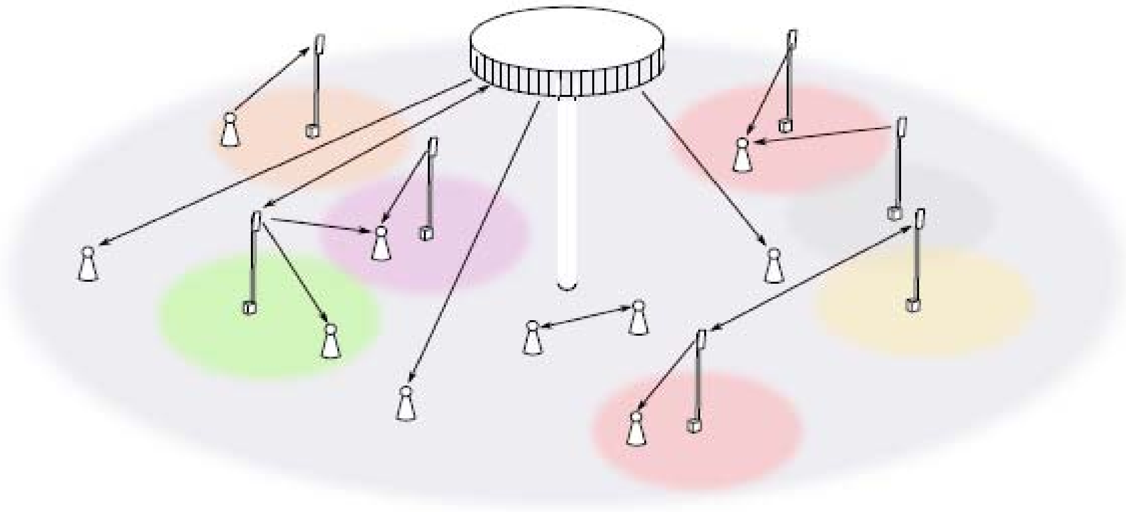
What is exactly 5G?

- **Tons of Plenary Talks and Overview Articles**
  - Fulfilling dream of ubiquitous wireless connectivity

- **Expectation: Many Metrics Should Be Improved in 5G**
  - Higher user data rates
  - Higher area throughput
  - Great scalability in number of connected devices
  - Higher reliability and lower latency
  - Better coverage with more uniform user rates
  - Improved energy efficiency

- **These are Conflicting Metrics!**

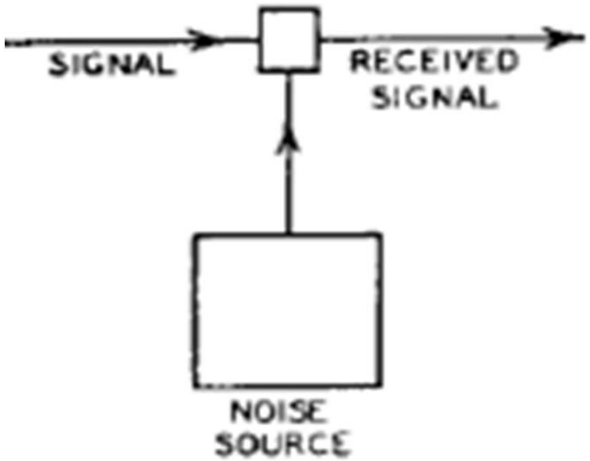
Some possible solutions



- More cells (small cells, multi-tier/heterogeneous networks)
- More antennas (MIMO techniques, massive MIMO)
- Cooperation and coordination (network MIMO, interference coordination, relays)
- 3D beamforming, antenna tilting, and smart antennas
- Device-to-device communication, distributed caching
- Cognitive radio (dynamic spectrum access)
- More spectrum (millimeter waves)
- Full-duplex transceivers
- New coding and modulation schemes

1948: Cybernetics and Theory of Communications

- ”A Mathematical Theory of Communication”, Bell System Technical Journal, 1948, C. E. Shannon
- ”Cybernetics, or Control and Communication in the Animal and the Machine”, Herman et Cie/The Technology Press, 1948, N. Wiener



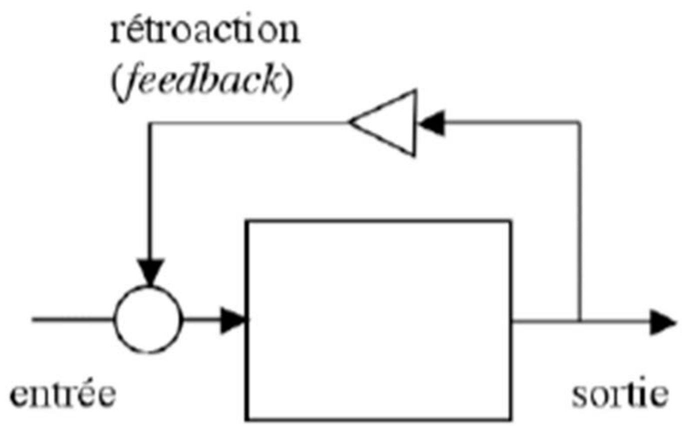
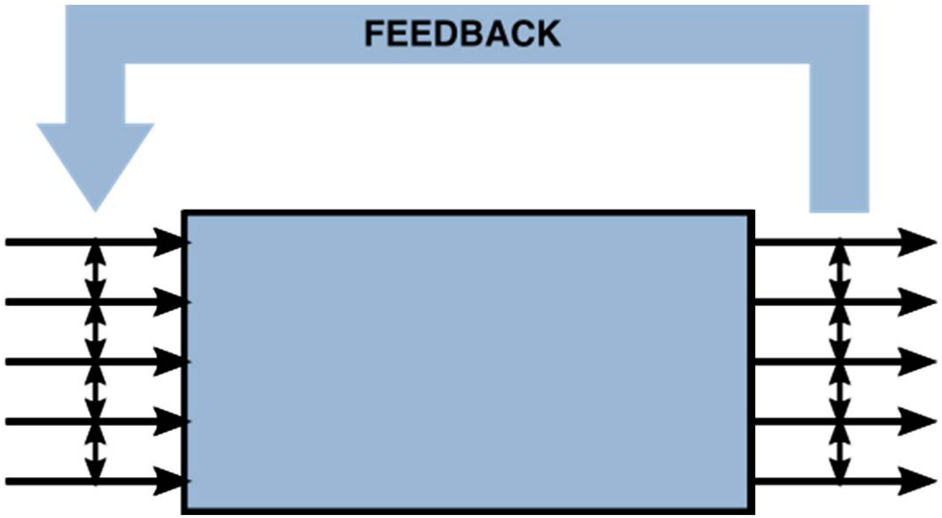


Fig. 1 – Boucle de rétroaction

60 years later... MIMO Flexible Networks



We must learn and control the black box

- within a fraction of time
- with finite energy.

In many cases, the number of inputs/outputs (the dimensionality of the system) is of the same order as the time scale changes of the box.

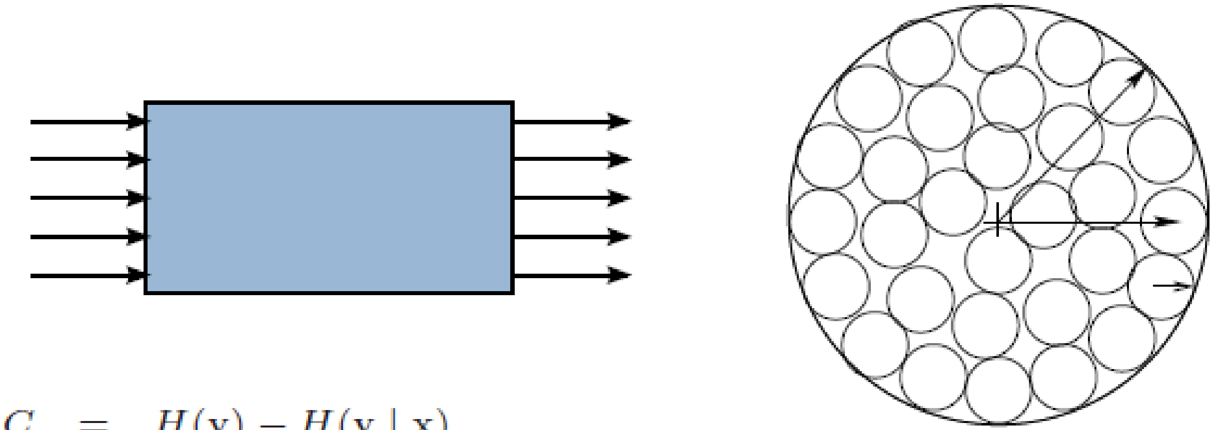
Huawei Mathematical and Algorithmic Sciences Lab:

20 Advanced Mathematical Tools for Engineering

- Discipline of Random Matrix Theory
- Discipline of Free Probability Theory
- Discipline of Stochastic Geometry
- Discipline of Discrete Mathematics
- Discipline of Statistics
- Discipline of Game Theory
- Discipline of Mean Field Theory
- Discipline of Information Theory
- Discipline of Signal Processing
- Discipline of Queuing Theory
- Discipline of Estimation Theory
- Discipline of Decision theory
- Discipline of Probability Theory
- Discipline of Optimization Theory
- Discipline of Statistical Mechanics
- Discipline of Factor Graphs
- Discipline of Control Theory
- Discipline of Learning theory
- Discipline on Partial Differential Equations Theory
- Discipline of Optimal Transport Theory

Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$



$$\begin{aligned} C &= H(\mathbf{y}) - H(\mathbf{y} \mid \mathbf{x}) \\ &= \log \det(\pi e \mathbf{R}_y) - \log \det(\pi e \mathbf{R}_n) \end{aligned}$$

$$\text{Rate} = \log\left(\frac{\det(\mathbf{R}_y)}{\det(\mathbf{R}_n)}\right)$$

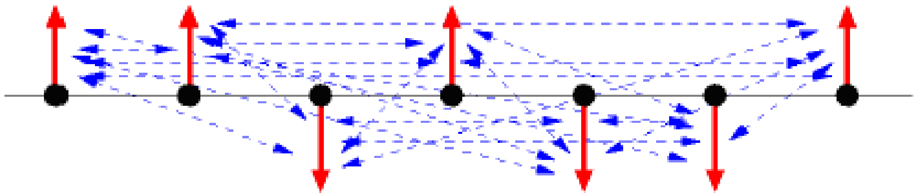
The rate is:

$$\begin{aligned} C_N &= \log_2 \det(\pi e(\sigma^2 \mathbf{I}_N + \mathbf{W}\mathbf{W}^H)) - \log_2 \det(\pi e \sigma^2 \mathbf{I}_N) \\ &= \log_2 \det(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{W}\mathbf{W}^H) \end{aligned}$$

Schrondinger's equation

$$H\Phi_i = E_i\Phi_i$$

Φ_i is the wave function
E_i is the energy level
H is the hamiltonian



Magnetic interactions between the spins of electrons

The Birth of Asymptotic Random Matrix Theory

Eugene Paul Wigner, 1902-1995

Randomness in 1955

E. Wigner. "Characteristic Vectors of bordered matrices with infinite dimensions", The annal of mathematics, vol. 62, pp.546-564, 1955.

$$\frac{1}{\sqrt{n}}\begin{bmatrix} 0 & +1 & +1 & +1 & -1 & -1 \\ +1 & 0 & -1 & +1 & +1 & +1 \\ +1 & -1 & 0 & +1 & +1 & +1 \\ +1 & +1 & +1 & 0 & +1 & +1 \\ -1 & +1 & +1 & +1 & 0 & -1 \\ -1 & +1 & +1 & +1 & -1 & 0 \end{bmatrix}$$

As the matrix dimension increases, what can we say about the eigenvalues (energy levels)?

Wigner Matrices: the semi-circle law

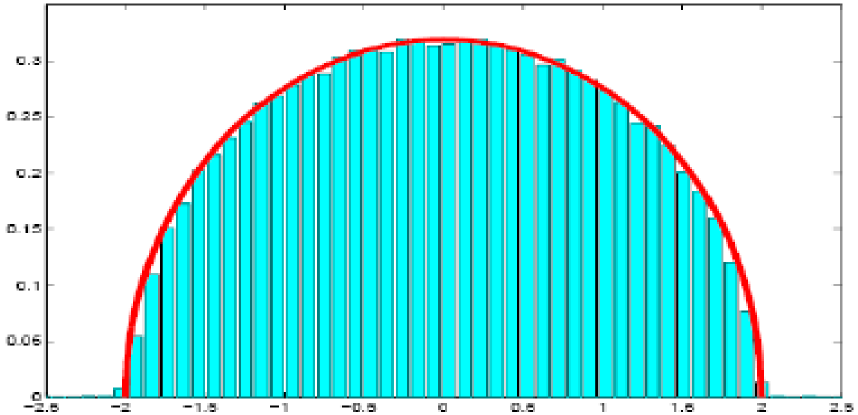


Figure 2: The semicircle law density function (4) compared with the histogram of the average of 100 empirical density functions for a Wigner matrix of size $n = 100$.

The empirical eigenvalue distribution of $\mathbf{H}$

$\mathbf{H}$ is Hermitian

$$dF_N(\lambda) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(\lambda - \lambda_1\right)$$

The moments of this distribution are given by:

$$\begin{aligned}
m_1^N &= \frac{1}{N}\text{tr}\left(\mathbf{H}\right) = \frac{1}{N}\sum_{i=1}^{N}\lambda_i = \int \lambda dF_N(\lambda) \\
m_2^N &= \frac{1}{N}\text{tr}\left(\mathbf{H}\right)^2 = \int \lambda^2 dF_N(\lambda) \\
\ldots &= \ldots \\
m_k^N &= \frac{1}{N}\text{tr}\left(\mathbf{H}\right)^k = \int \lambda^k dF_N(\lambda)
\end{aligned}$$

In many cases, all the moments converge. This is exactly the type of results needed to understand the network.

Wigner Matrices: the semi-circle law

Wigner's proof of the convergence to the semi-circle law:

The empirical moment $\frac{1}{N}\mathrm{Trace}(\mathbf{H}^{2k}) \longrightarrow$ The Catalan numbers

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\mathrm{Trace}(\mathbf{H}^{2k}) &= \int_{-2}^{2} x^{2k} f(x)dx \\ &= \frac{1}{k+1}C_k^{2k}\end{aligned}$$

Since the semi-circle law is symmetric, the odd moments vanish.

Catalan Numbers

Eugène Charles Catalan, 1814-1894

Wigner Matrices: the semi-circle law

E. Wigner. "On the Distribution of Roots of certain symmetric matrices", The Annals of Mathematics, vol. 67, pp.325-327, 1958.

Theorem2. Consider a $N \times N$ standard Wigner matrix $\mathbf{W}$ such that, for some constant κ and sufficiently large N,

$$\max_{i,j} \mathbb{E}(| w_{ij} |^4) \leq \frac{\kappa}{N^2}$$

Then the empirical distribution of $\mathbf{W}$ converges almost surely to the semi-circle law whose density is:

$$f(x) = \frac{1}{2\pi}\sqrt{4 - x^2}$$

with $| x | \leq 2$

The semi-circle law is also known as the non-commutative analog of the Gaussian distribution.

Square Matrix of i.i.d coefficients

$$\frac{1}{\sqrt{n}}\begin{bmatrix} -1 & +1 & +1 & -1 & -1 & +1 \\ -1 & +1 & -1 & -1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 & +1 \\ -1 & -1 & +1 & -1 & -1 & -1 \\ -1 & +1 & +1 & +1 & +1 & -1 \end{bmatrix}$$

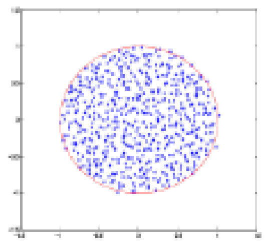


Figure 3: The full-circle law and the eigenvalues of a realization of a 500×500 matrix

On the empirical spectral distribution of large random matrices

Remember that $\mathbf{h}_j \sim \mathcal{CN}\left(0, \frac{1}{K}\mathbf{I}_N\right)$ i.i.d., for $j = 1, \ldots, K$.

What we expect from the strong law of large numbers:

For $K \to \infty$ and while $N =$ const., we have

$$\mathbf{H}\mathbf{H}^{\mathsf{H}} = \sum_{j=1}^{K} \mathbf{h}_j \mathbf{h}_j^{\mathsf{H}} = \frac{1}{K} \sum_{j=1}^{K} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^{\mathsf{H}} \xrightarrow{\text{a.s.}} \mathbb{E}\left[\tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^{\mathsf{H}}\right] = \mathbf{I}_N, \qquad \tilde{\mathbf{h}} \sim \mathcal{CN}(0, \mathbf{I}_N).$$

But what happens if also $N \to \infty$, while $N/K \to c \in (0, \infty)$?

- We can still say that $\left[\mathbf{H}\mathbf{H}^{\mathsf{H}}\right]_{i,i} \xrightarrow{\text{a.s.}} 1$ and $\left[\mathbf{H}\mathbf{H}^{\mathsf{H}}\right]_{i,j} \xrightarrow{\text{a.s.}} 0$ for $j \neq i$.
- However, it is not true that $\mathbf{H}\mathbf{H}^{\mathsf{H}} - \mathbf{I}_N \xrightarrow{\text{a.s.}} \mathbf{0}$!

What happens to the eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$?

Remark (Wishart matrices)

The matrix $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ is called a "Wishart matrix" with K degrees of freedom. For finite N, K, its exact eigenvalue distribution is known, e.g., [7].

Empirical and limiting spectral distribution

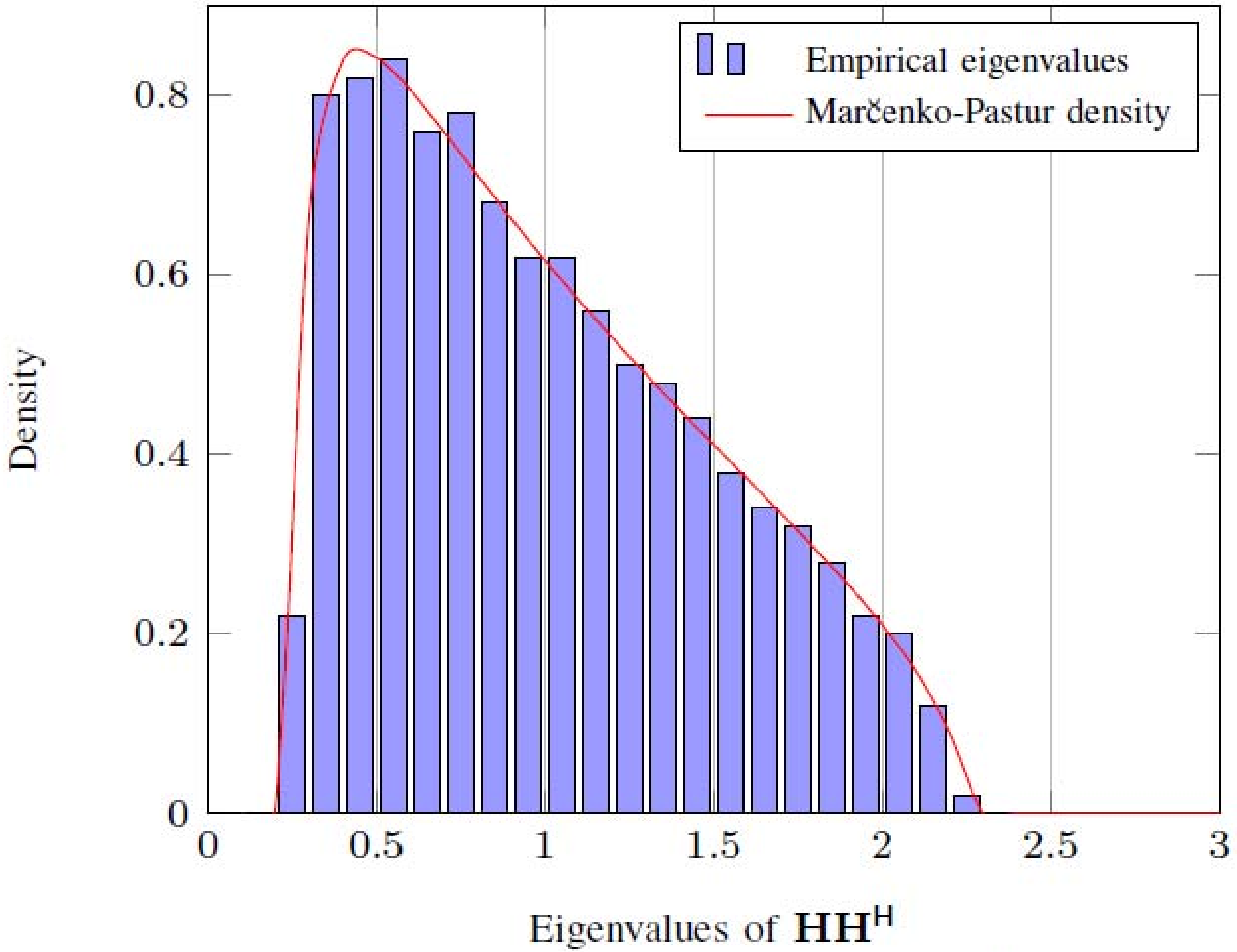


Figure: Histogram of the eigenvalues of a single realization of $\mathbf{HH}^{\mathsf{H}}$ for $N = 500$, $K = 2000$.

Asymptotic mutual information

$F^{\mathbf{HH}^\mathsf{H}} \Rightarrow F_c$, almost surely, implies that for any **bounded continuous** function g:

$$\int g(\lambda) dF^{\mathbf{HH}^\mathsf{H}}(\lambda) \xrightarrow{\text{a.s.}} \int g(\lambda) dF_c(\lambda).$$

The normalized mutual information between $\mathbf{y}$ and $\mathbf{x}$ is not bounded, since

$$I(\sigma^2) \triangleq \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{HH}^\mathsf{H}\right) = \int \log\left(1 + \frac{\lambda}{\sigma^2}\right) dF^{\mathbf{HH}^\mathsf{H}}(\lambda).$$

Theorem (No eigenvalues outside the support [9, Theorem 1.1], [10, Theorem 3])

Denote by λ_{max} and λ_{min} the largest and smallest eigenvalue of $\mathbf{HH}^\mathsf{H}$, respectively. Then,

$$(i)\ \lambda_{max} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2, \qquad (ii)\ \lambda_{min} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2 \mathbb{1}\{c \le 1\}.$$

Thus, we have (see, e.g., [11])

Theorem (Asymptotic mutual information)

$$I(\sigma^2) \xrightarrow{a.s.} \int g(\lambda) dF_c(\lambda) \triangleq \bar{I}(\sigma^2) = \frac{1}{c}\log(1+cm_c) + \log\left(1 + \frac{1}{\sigma^2}\frac{1}{1+cm_c}\right) - \frac{m_c}{1+cm_c}$$

where $m_c = m_c(-\sigma^2)$.

Asymptotic mutual information: Numerical results

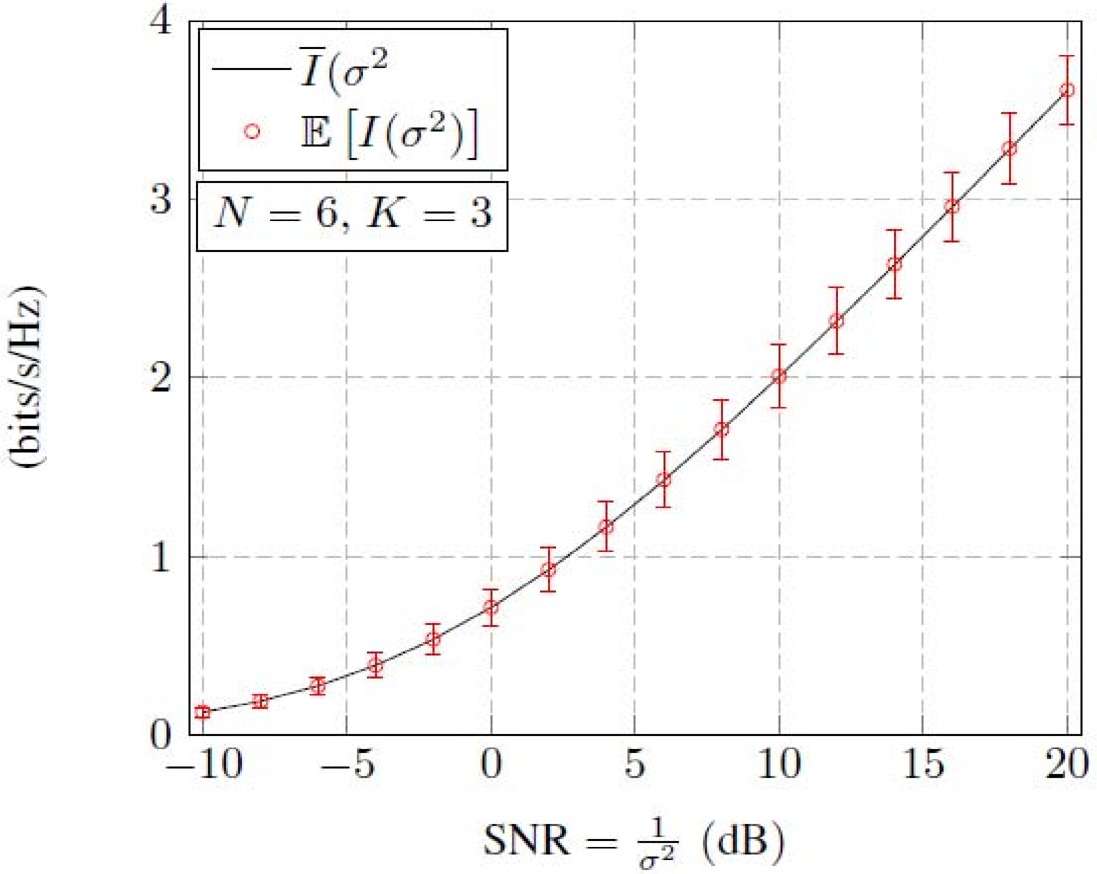
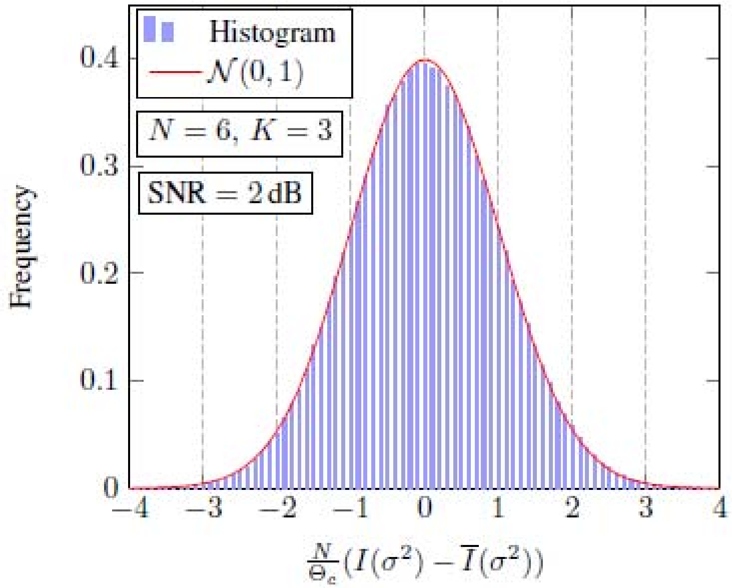


Figure: Average normalized mutual information $\mathbb{E}\left[I(\sigma^2)\right]$ and its asymptotic approximation $\bar{I}(\sigma^2)$ vs. SNR for $N = 6$ and $K = 3$. Errorbars correspond to one standard deviation in each direction.

Fluctuations of the mutual information



Example

Approximation of the outage probability:

$$P_{\text{out}}(r) = \Pr\left(NI(\sigma^2) < r\right)$$
$$\approx 1 - Q\left(\frac{r - N\bar{I}(\sigma^2)}{\Theta_c}\right).$$

Theorem (Central limit theorem of the mutual information [4, Theorem 3.18])

$$\frac{N}{\Theta_c}\left(I(\sigma^2) - \bar{I}(\sigma^2)\right) \Rightarrow \mathcal{N}(0,1)$$

where the asymptotic variance is given as $\Theta_c^2 = -\log\left(1 - \frac{c m_c(-\sigma^2)^2}{(1+c m_c(-\sigma^2))^2}\right).$

Beyond LTE: The 400-Antenna Base Station

Thomas L. Marzetta

Bell Laboratories

Alcatel-Lucent

28 May, 2010

Alcatel-Lucent

Infinitely Many Antennas: Forward-Link Capacity For 20 MHz Bandwidth, 42 Terminals per Cell, 500 μsec Slot

Interference-limited: energy-per-bit can be made arbitrarily small!

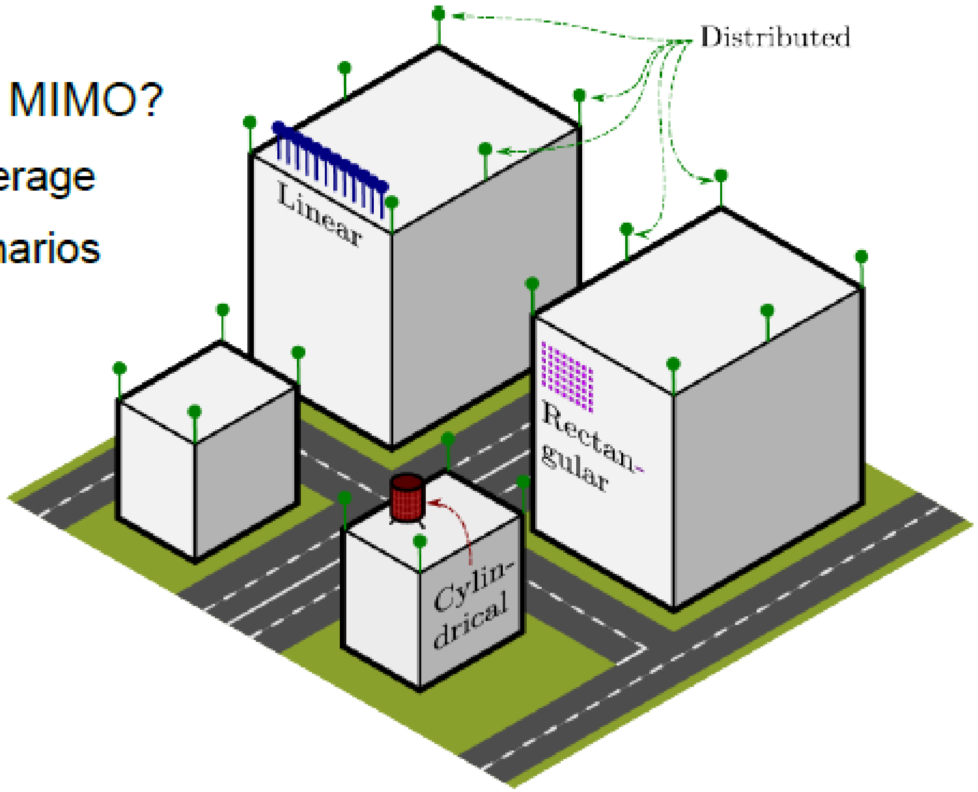
Frequency Reuse	.95-Likely SIR (dB)	.95-Likely Capacity per Terminal (Mbits/s)	Mean Capacity per Terminal (Mbits/s)	Mean Capacity per Cell (Mbits/s)
1	-29	.016	44	1800
3	-5.8	.89	28	1200
7	8.9	3.6	17	730

				Mean Capacity per Cell (Mbits/s)
LTE Advanced (>= Release 10)				74

Alcatel-Lucent

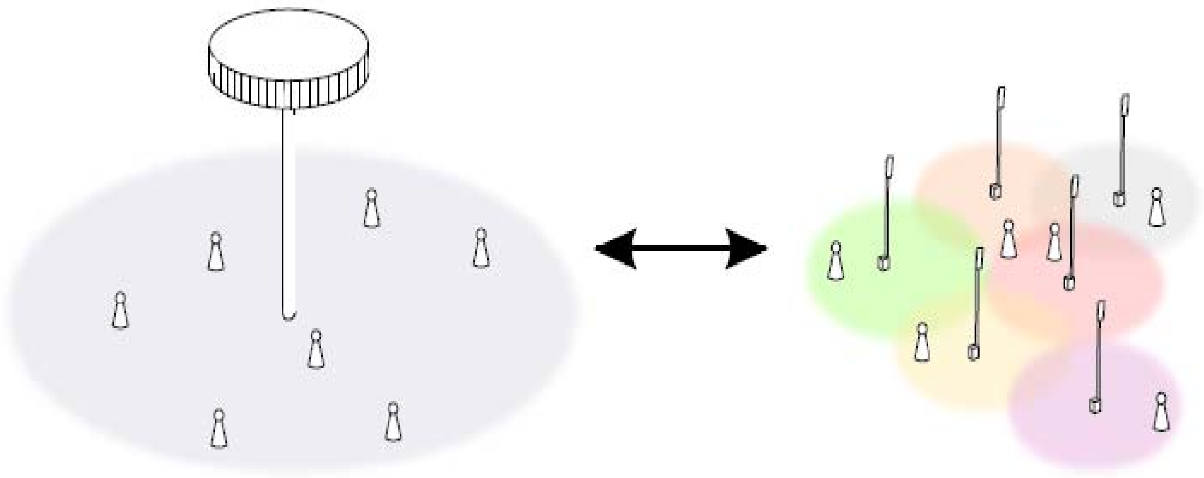
Massive MIMO Deployment

- When to Deploy Massive MIMO?
  - Improved wide-area coverage
  - Special superdense scenarios
- Co-located Deployment
  - 1D, 2D, or 3D arrays
  - One or multiple sectors
- Distributed Deployment
  - Remote radio heads
  - Cloud RAN



The clean slate approach

"David vs Goliath" or "Small Cells vs Massive MIMO"



How to densify: *"More antennas or more BSs?"*

What if we are only interested in the average throughput per UT?

A thought experiment

Consider an infinite large network of randomly uniformly distributed base stations and user terminals.

What would be better?

A $2 \times$ more base stations

B $2 \times$ more antennas per base station

How to optimally deploy your antennas?

A thought experiment

Consider an infinite large network of randomly uniformly distributed base stations and user terminals.

What would be better?

- A $2\times$ more base stations
- B $2\times$ more antennas per base station

Stochastic geometry can provide an answer.

What if we are only interested in the average throughput per UT?

System model: Downlink

Received signal at a tagged UT at the origin:

$$y = \underbrace{\frac{1}{r_0^{\alpha/2}}\mathbf{h}_0^{\mathsf{H}}\mathbf{x}_0}_{\text{desired signal}} + \underbrace{\sum_{i=1}^{\infty}\frac{1}{r_i^{\alpha/2}}\mathbf{h}_i^{\mathsf{H}}\mathbf{x}_i}_{\text{interference}} + n$$

- $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$: fast fading channel vectors
- r_i: distance to ith closest BS
- $P = \mathbb{E}\left[\mathbf{x}_i^{\mathsf{H}}\mathbf{x}_i\right]$: average transmit power constraint per BS

Assumptions:

- infinitely large network of uniformly randomly distributed BSs and UTs with densities λ_{BS} and λ_{UT}, respectively
- single-antenna UTs, N antennas per BS
- each UT is served by its *closest* BS
- distance-based path loss model with path loss exponent $\alpha > 2$
- total bandwidth W, re-used in each cell

What if we are only interested in the average throughput per UT?

Transmission strategy: Zero-forcing

Assumptions:

- $\mathcal{K} = \frac{\lambda_{\mathrm{UT}}}{\lambda_{\mathrm{BS}}}$ UTs need to be served by each BS on average
- total bandwidth W divided into $L \geq 1$ sub-bands
- $K = \mathcal{K}/L \leq N$ UTs are simultaneously served on each sub-band

Transmit vector of BS i:

$$\mathbf{x}_i = \sqrt{\frac{P}{K}} \sum_{k=1}^{K} \mathbf{w}_{i,k} s_{i,k}$$

- $s_{i,k} \sim \mathcal{CN}(0,1)$: message determined for UT k from BS i
- $\mathbf{w}_{i,k} \in \mathbb{C}^{N \times 1}$: ZF-beamforming vectors

What if we are only interested in the average throughput per UT?

Performance metric: Average throughput

Received SINR at tagged UT:

$$\gamma = \frac{r_0^{-\alpha} \left|\mathbf{h}_0^{\mathsf{H}} \mathbf{w}_{0,1}\right|^2}{\sum_{i=1}^{\infty} r_i^{-\alpha} \sum_{k=1}^{K} \left|\mathbf{h}_i^{\mathsf{H}} \mathbf{w}_{i,k}\right|^2 + \frac{K}{P}} = \frac{r_0^{-\alpha} S}{\sum_{i=1}^{\infty} r_i^{-\alpha} g_i + \frac{K}{P}}$$

Coverage probability:

$$P_{\mathrm{cov}}(T) = \mathbb{P}(\gamma \geq T)$$

Average throughput per UT:

$$\boxed{C = \frac{W}{L} \times \mathbb{E}\left[\log(1+\gamma)\right] = \frac{W}{L} \times \int_0^{\infty} P_{\mathrm{cov}}\left(e^z - 1\right) dz}$$

Remarks:

- expectation with respect to fading *and* BSs locations
- $S = \left|\mathbf{h}_0^{\mathsf{H}} \mathbf{w}_{0,1}\right|^2 \sim \Gamma(N-K+1, 1)$, $\quad g_i = \sum_{k=1}^{K} \left|\mathbf{h}_i^{\mathsf{H}} \mathbf{w}_{i,k}\right|^2 \sim \Gamma(K, 1)$
- K impacts the interference distribution, N impacts the desired signal
- for $P \to \infty$, the SINR becomes independent of λ_{BS}

What if we are only interested in the average throughput per UT?

A closed-form result

Theorem (Combination of Baccelli'09, Andrews'10)

$$P_{cov}(T) = \int_{r_0>0}\int_{-\infty}^{\infty} \mathcal{L}_{I_{r_0}}\left(i2\pi r_0^{\alpha} Ts\right) \exp\left(-\frac{i2\pi r_0^{\alpha} TK}{P} s\right) \frac{\mathcal{L}_S(-i2\pi s) - 1}{i2\pi s} f_{r_0}(r_0) ds dr_0$$

where

$$\mathcal{L}_{I_{r_0}}(s) = \exp\left(-2\pi\lambda_{BS}\int_{r_0}^{\infty}\left(1 - \frac{1}{(1+sv^{-\alpha})^K}\right) vdv\right)$$

$$\mathcal{L}_S(s) = \left(\frac{1}{1+s}\right)^{N-K+1}$$

$$f_{r_0}(r_0) = 2\pi\lambda_{BS} r_0 e^{-\lambda_{BS}\pi r_0^2}$$

The computation of $P_{cov}(T)$ requires in general three numerical integrals.

J. G. Andrews, F. Baccelli, R. K. Ganti, "A Tractable Approach to Coverage and Rate in Cellular Networks" IEEE Trans. Wireless Commun., submitted 2010.

F. Baccelli, B. Błaszczyszyn, P. Mühlethaler, "Stochastic Analysis of Spatial and Opportunistic Aloha" Journal on Selected Areas in Communications, 2009

What if we are only interested in the average throughput per UT?

Example

- Density of UTs: $\lambda_{\mathrm{UT}} = 16$
- Constant transmit power density: $P \times \lambda_{\mathrm{BS}} = 10$
- Number of BS-antennas: $N = \lambda_{\mathrm{UT}} / \lambda_{\mathrm{BS}}$
- Path loss exponent: $\alpha = 4$
- UT simultaneously served on each band: $K = \lambda_{UT} / (\lambda_{\mathrm{BS}} \times L)$

$\Rightarrow$ Only two parameters: λ_{BS} and L

Table: Average spectral efficiency C/W in (bits/s/Hz)

sub-bands L	$\lambda_{\mathrm{BS}} = 1$	$\lambda_{\mathrm{BS}} = 2$	$\lambda_{\mathrm{BS}} = 4$	$\lambda_{\mathrm{BS}} = 8$	$\lambda_{\mathrm{BS}} = 16$
1	0.6209	0.8188	1.1964	1.5215	2.1456
2	1.1723	1.2414	1.3404	1.5068	×
4	0.8882	0.8973	1.1964	×	×
8	0.5689	0.5952		×	×
16	0.3532	×	×	×	×

Fully distributing the antennas gives highest throughput gains!

What if we are only interested in the average throughput per UT?

Cellular Dreams and Cordless Nightmares

Life at Bell Laboratories in Interesting Times

Richard H. Frenkiel

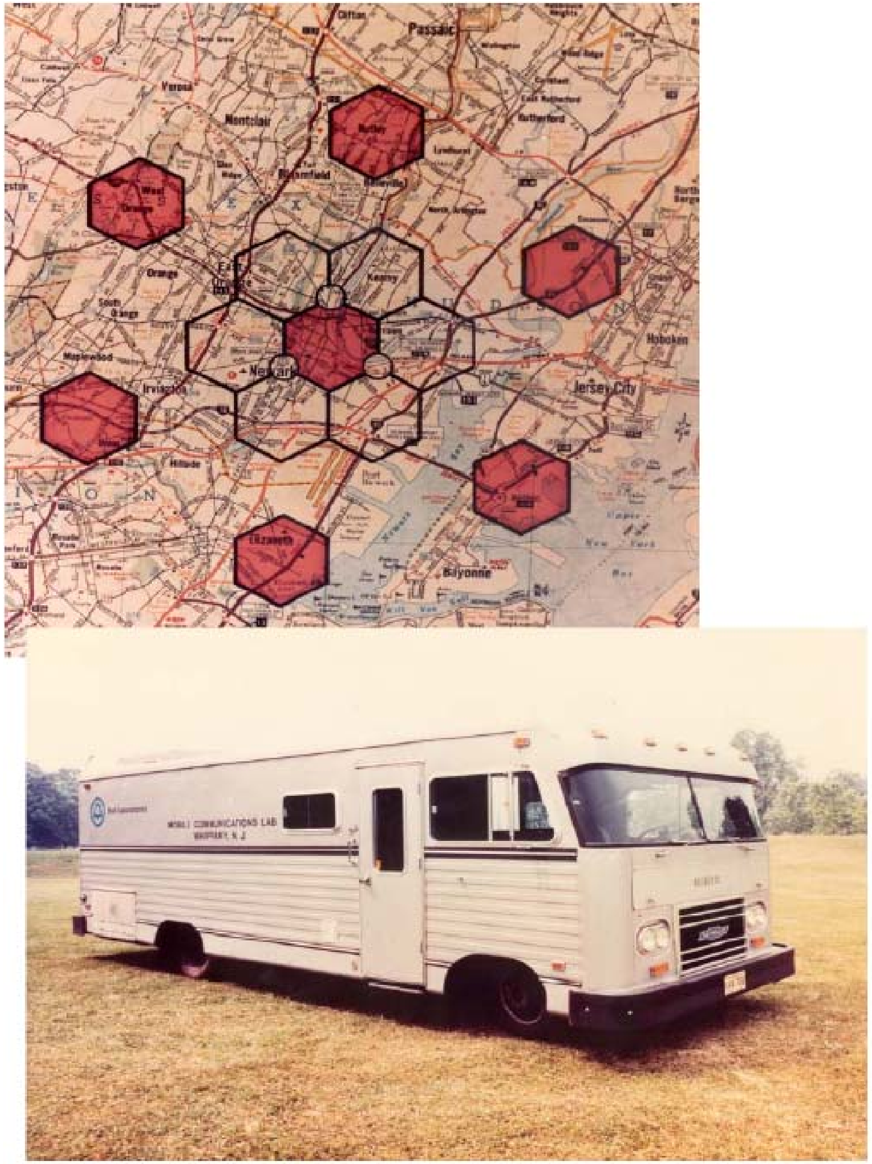
What if we are only interested in the average throughput per UT?

Trials and Tribulations

By 1976, the time had come to prove that our many claims could be turned into a practical system. Small cell coverage over a large service area would require hundreds of cells and cost hundreds of millions of dollars, so we applied for permission to conduct two separate trials. A large-cell Market Trial in Chicago would provide realistic service to more than 2000 customers, while a small-cell "Test Bed" in Newark, New Jersey, would demonstrate that the smallest cells could provide good service in the presence of nearby interference. In combination, these trials would provide a complete demonstration of our system.

Motorola objected to our proposal as inadequate, since neither the trial in Chicago nor the Test Bed in Newark demonstrated a fully developed small-cell system. Chicago, they argued, used very large cells, while Newark was only a partial grid of small cells. Since a demonstration of small cells over a large area was clearly impractical, we were confident that the FCC would see Motorola's objections for what they were—another smoke screen intended to delay progress. As it turned out, our faith was misplaced. The FCC ruled that our proposed trials were inadequate, using virtually the same arguments that Motorola had presented, and summarily denied our application.

What if we are only interested in the average throughput per UT?



The partial small-cell grid in Newark and the Test Van