Supelec

Randomness in Wireless Networks: how to deal with it?

Mérouane Debbah
Alcatel-Lucent Chair on Flexible Radio
merouane.debbah@supelec.fr

The performance dilemma...

La théorie, c'est quand on sait tout et que rien ne fonctionne.

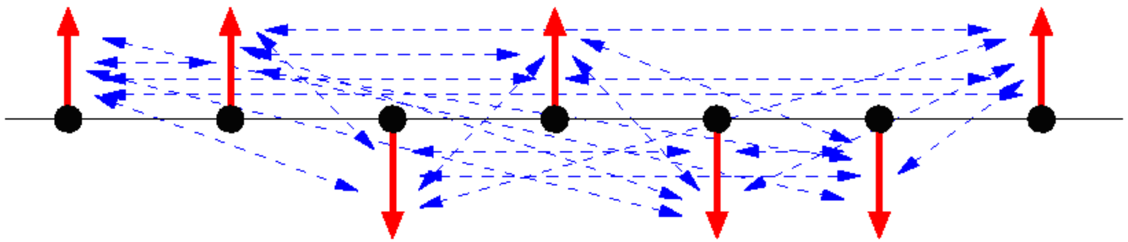
La pratique, c'est quand tout fonctionne et que personne ne sait pourquoi.

Ici, nous avons réuni théorie et pratique : rien ne fonctionne... et personne ne sait pourquoi!

Schrondinger's equation

$$H\mathbf{\Phi}_i = E_i\mathbf{\Phi}_i$$

$\mathbf{\Phi}_i$ 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

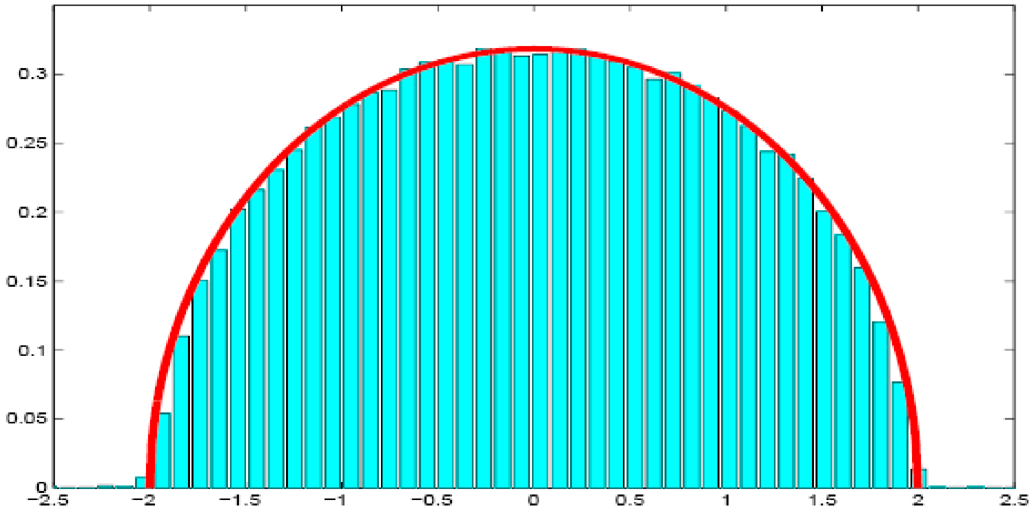


Figure 2: The semicircle law density function (4) compared with the histogram of the average of 100 empirical density functions for a Wigner matrix of size $n = 100$.

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.

Wigner Matrices: the semi-circle law

Calculus based on recursion:

We integrate by parts and get:

$$
\begin{aligned}
\alpha_{2k} &= \frac{1}{\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx \\
&= -\frac{1}{2\pi}\int_{-2}^{2} \frac{-x}{\sqrt{4-x^2}} x^{2k-1}(4-x^2)dx \\
&= \frac{1}{2\pi}\int_{-2}^{2} \sqrt{4-x^2}(x^{2k-1}(4-x^2))'dx \\
&= 4(2k-1)\alpha_{2k-2} - (2k+1)\alpha_{2k}
\end{aligned}
$$

In this way, the recursion is obtained:

$$\alpha_{2k} = \frac{2(2k-1)}{k+1}\alpha_{2k-2}$$

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,

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

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 $\mid x \mid \leq 2$

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

Remarks on asymptotics

Distribution Insensitivity: The asymptotic distribution does not depend on the distribution of the independent entries.

Ergodicity: The eigenvalue histogram of one realization converges almost surely to the asymptotic eigenvalue distribution.

Speed of Convergence: $8 = \infty$

Gaussian Case: Non-asymptotic joint distribution of the entries known.

Searching for determinism in randomness

- Distribution of $\lambda(\mathbf{H})$.
- Distribution of $\lambda(\mathbf{H}^H\mathbf{H})$.
- Distribution of $\lambda_{\max}(\mathbf{H})$.
- Joint distribution of $\lambda_1(\mathbf{H}), \ldots, \lambda_N(\mathbf{H})$.
- Distribution of the spacings between adjacent eigenvalues (linked to the Riemann Hypothesis).
- Distribution of $\mathbf{H}^H\mathbf{H}$.
- Distribution of the matrix of eigenvectors of $\mathbf{H}^H\mathbf{H}$.

Randomness in 1967

$\mathbf{H}$ matrix $N \times K$ with i.i.d. elements, zero mean and variance $1/N$.

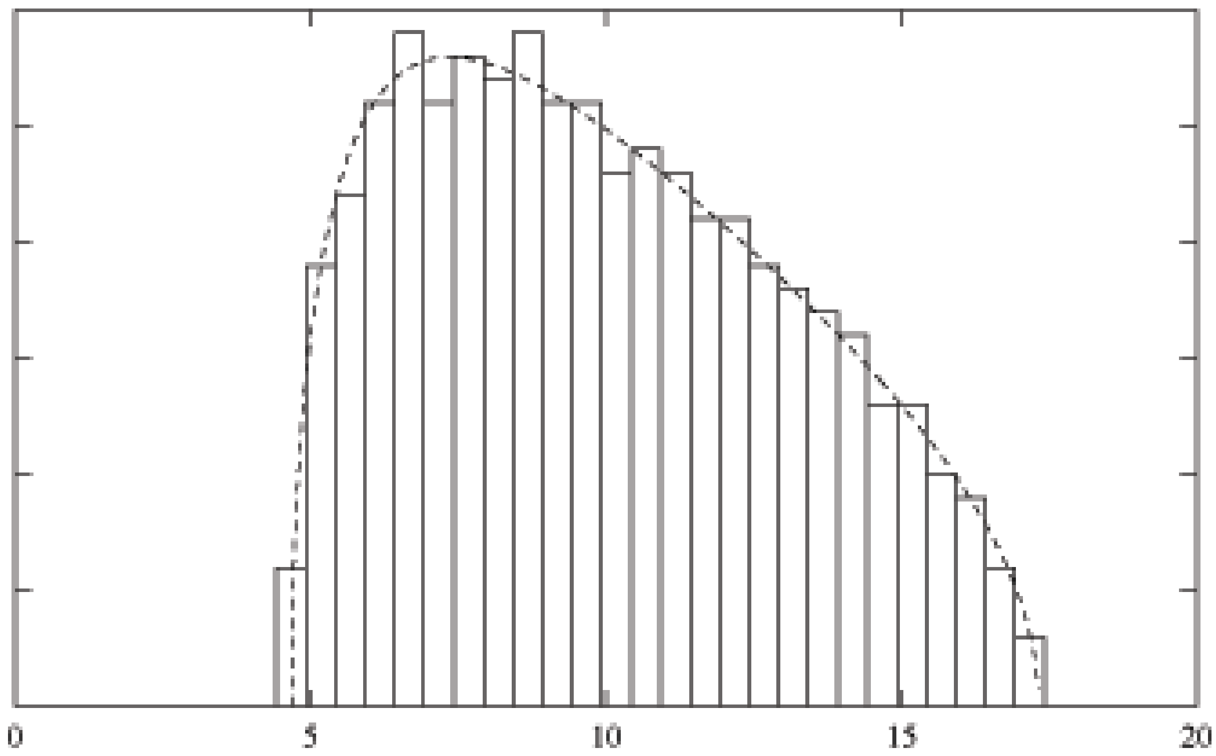
Eigenvalues of the matrix

$$N \left\{ \underbrace{\begin{bmatrix} & & \\ & \mathbf{H} & \\ & & \end{bmatrix}}_{K} \right. \begin{bmatrix} \\ \\ \mathbf{H}^H \\ \\ \\ \end{bmatrix}$$

when $N \to \infty, \quad K/N \to \alpha$ **IS NOT IDENTITY!**

Remark: If the entries are Gaussian, the matrix is called a Wishart matrix with K degrees of freedom. The **exact** distribution is known in the finite case.

Limiting eigenvalue distribution for $\alpha = 10$



Remark: Quite remarkably, the support is bounded even though the entries can take any values!

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

$\mathbf{H}$ is $N \times K$ i.i.d Gaussian with $\frac{K}{N} = \alpha$

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

The moments of this distribution are given by:

$$\begin{aligned}
m_1^N &= \frac{1}{N}\text{tr}\left(\mathbf{HH}^H\right) = \frac{1}{N}\sum_{i=1}^{N} \lambda_i \rightarrow 1 \\
m_2^N &= \frac{1}{N}\text{tr}\left(\mathbf{HH}^H\right)^2 = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^2 \rightarrow 1 + \alpha \\
m_3^N &= \frac{1}{N}\text{tr}\left(\mathbf{HH}^H\right)^3 = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^3 \rightarrow \alpha^2 + 3\alpha + 1
\end{aligned}$$

The Marchenko-Pastur Distribution Law

V. A. Marchenko and L. A. Pastur, ”Distributions of eigenvalues for some sets of random matrices,” Math USSR-Sbornik, vol.1 pp.457-483, 1967.

Theorem. Consider an $N \times K$ matrix $\mathbf{H}$ whose entries are independent zero-mean complex (or real) random variables with variance $\frac{1}{N}$ and fourth moments of order $O(\frac{1}{N^2})$. As $K, N \to \infty$ with $\frac{K}{N} \to \alpha$, the empirical distribution of $\mathbf{H}^H\mathbf{H}$ converges almost surely to a nonrandom limiting distribution with density

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

where $a = (1 - \sqrt{\alpha})^2$ and $b = (1 + \sqrt{\alpha})^2$.

Randomness in 1980's

What can one say about the distribution of $\mathbf{C} = \mathbf{A} + \mathbf{B}$

- In general, we cannot find the eigenvalues of sums/products of independent random matrices from the eigenvalues of individual matrices.
- The exception is when the matrices have the same eigenvectors, as for diagonal matrices.
- If this is not the case, it is hard to combine the eigenvectors of A and B to find the eigenvectors of AB.

The empirical eigenvalue distribution of $\mathbf{C} = \mathbf{A} + \mathbf{B}$

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

In many cases, the asymptotic moments of $\mathbf{C}$ can be expressed only with the asymptotic moments of $\mathbf{A}$ and $\mathbf{B}$.

$$m_k^C \quad = \quad \lim_{N\to\infty} \frac{1}{N}\text{tr}\left(\mathbf{C}\right)^k = f(m_1^A, \ldots, m_k^A, m_1^B, \ldots, m_k^B)$$

In other words, the asymptotic empirical eigenvalue distribution of $\mathbf{C}$ depends only on the asymptotic empirical eigenvalue distribution of $\mathbf{A}$ and $\mathbf{B}$.

When this happens, we say that the matrices are free and the framework falls in the realm of free probability theory. The same holds for $\mathbf{C} = \mathbf{AB}$.

Voiculescu

- Free probability was developed as a probability theory for random variables which do not commute, like matrices.
- Many similarities with classical probability.
- Random variables are elements in what we will call a *noncommutative probability space*: A pair (A, ϕ), where A is a unital $*$-*algebra* with unit I, and ϕ is a normalized (i.e. $\phi(I) = 1$) linear functional on A.
- For matrices, ϕ will be the normalized trace tr_n, defined by

$$tr_n(a) = \frac{1}{n} \sum_{i=1}^{n} a_{ii}.$$

For random matrices, $\phi = \tau_n$ is defined by

$$\tau_n(a) = \frac{1}{n} \sum_{i=1}^{n} E(a_{ii}) = E(tr_n(a)).$$

The unit in these $*$-algebras is the $n \times n$ identity matrix I_n.

Applications of Random Matrix Theory

- Wigner (55) , Dyson (67) : Random matrix theory and the statistical theory of energy levels of nuclei.
- Potters (00), Bouchaud (00) : Random matrix theory and financial correlations.
- Voiculescu (91) , Biane (00), Hiai, Petz (00): Random matrix theory and Free probability Theory.
- Silverstein (89), Pastur (72), Girko (90), Edelman (89): Random matrix theory and Cauchy-Stieltjes transform.
- Speicher (92): Random matrix theory and Combinatorics.
- Tanaka (01), Moustakas (03), Sengupta (03) : Random matrix Theory and statistical mechanics approach.

Random matrix theory versus Information Theory

For a discrete finite random Gaussian vector $\mathbf{x}_i$ of size n, the differential entropy per dimension is given by:

$$
\begin{aligned}
H &= \log(\pi e) + \frac{1}{n} \log \det(\mathbf{R}) \\
&= \log(\pi e) + \frac{1}{n} \sum_{i=1}^{n} \log(\lambda_i)
\end{aligned}
$$

where $\mathbf{R} = \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^*)$ is the covariance and λ_i the associated eigenvalues.

The covariance carries therefore all the information of networks where Gaussian signaling are in use!

The finite size considerations of random networks

For a number of observations K of the vector $\mathbf{x}_i, i = 1, \ldots, K$, one can only extract:

$$\hat{\mathbf{R}} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{x}_i \mathbf{x}_i^H$$
$$= \mathbf{R}^{\frac{1}{2}} \mathbf{U}\mathbf{U}^H \mathbf{R}^{\frac{1}{2}}$$

Here, $\mathbf{U} = [\mathbf{u}_1, \ldots \mathbf{u}_p]$ is an $N \times K$ i.i.d zero mean Gaussian vector of variance $\frac{1}{K}$ (and $\mathbf{x}_1 = \mathbf{R}^{\frac{1}{2}}\mathbf{u}_1$).

Note that the non-zero eigenvalues of $\hat{\mathbf{R}}$ are the same as $\mathbf{U}\mathbf{U}^H\mathbf{R}$

One needs ”finite asymptotic techniques” to compute the spectrum of of $\mathbf{R}$ with respect to the spectrum of $\hat{\mathbf{R}}$.

The spectral efficiency barrier: energy and bandwith

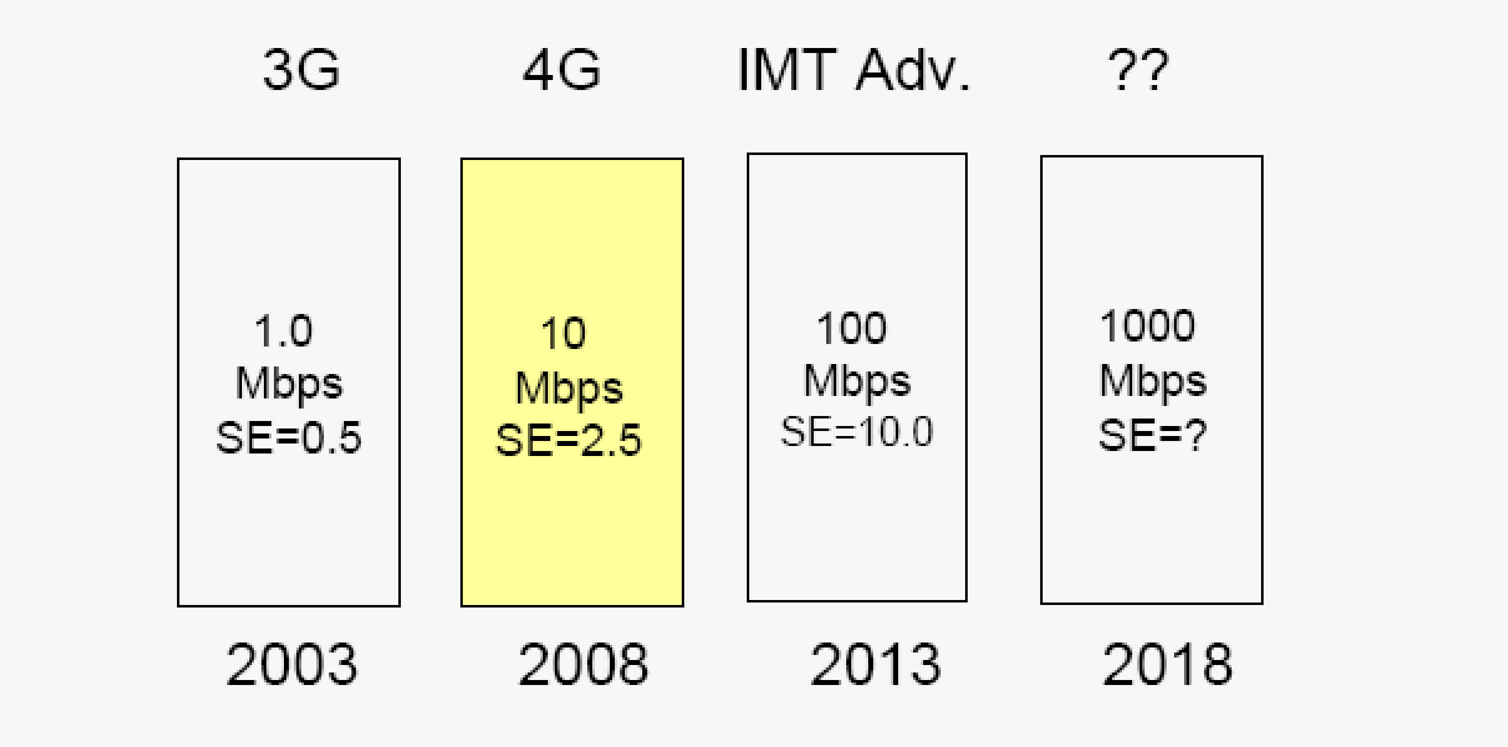
$$C = W \log(1 + \mathrm{SNR})$$

$$\mathrm{SNR} = C \frac{Eb}{N0}$$

$$\lim_{C \to 0} \frac{Eb}{N0} = \frac{2^C - 1}{C} = \log(2)$$

The two degrees of freedom (Energy and Bandwidth) are the barrier limit of communication.

Is there a limit to spectral efficiency?



”As the number of users in the network increases, interference becomes the bottleneck”...the resources must be shared...

Random Networks: from b/s/Hz to b/s/Hz/m^2

Random Networks (also known as Flexible networks) do not consider wireless resources as a ”cake to be shared” among users but take benefit of the high number of interacting devices to increase the spectral efficiency frontier.

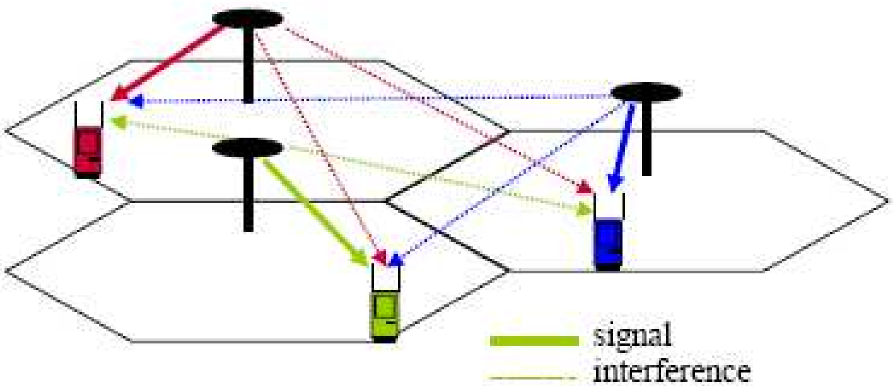
More devices represent more opportunities to schedule information which enhances the global throughput.

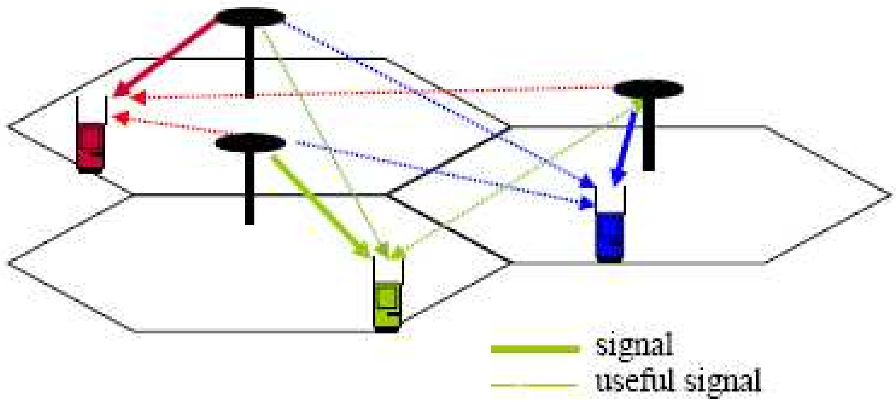
Interference is considered as an opportunity rather than a drawback by exploiting intelligently the degrees of freedom of wireless communications:

- Space (MIMO Network)
- Frequency/time (Cognitive Network)
- User (Opportunistic Network).

MIMO Network

”We build to many walls and not enough bridges”, Isaac Newton

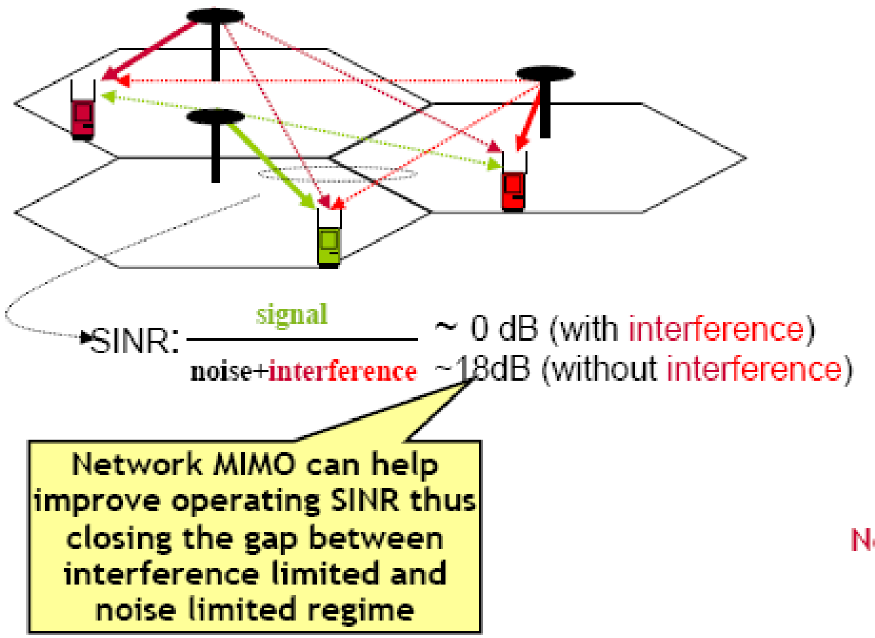


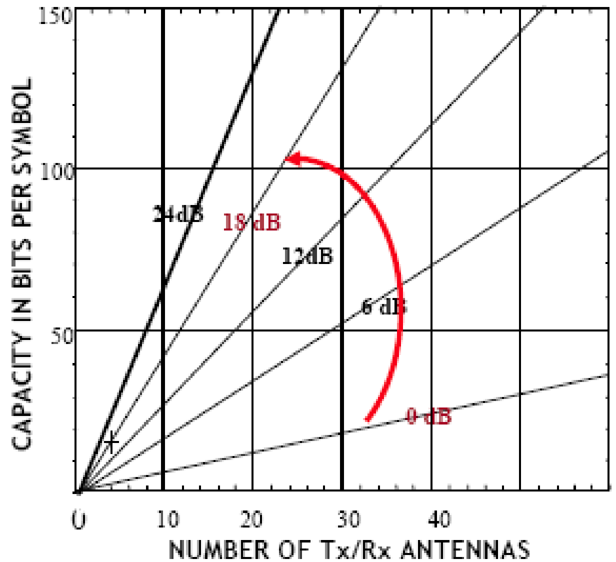


Combat interference through frequency/space reuse or power control.

Exploit interference through coordination and cooperation.

MIMO Network





Network MIMO can greatly improve the efficiency of multiple antennas in cellular networks.

In theory, the only limiting factor for increasing the spectral efficiency is the number of base stations.

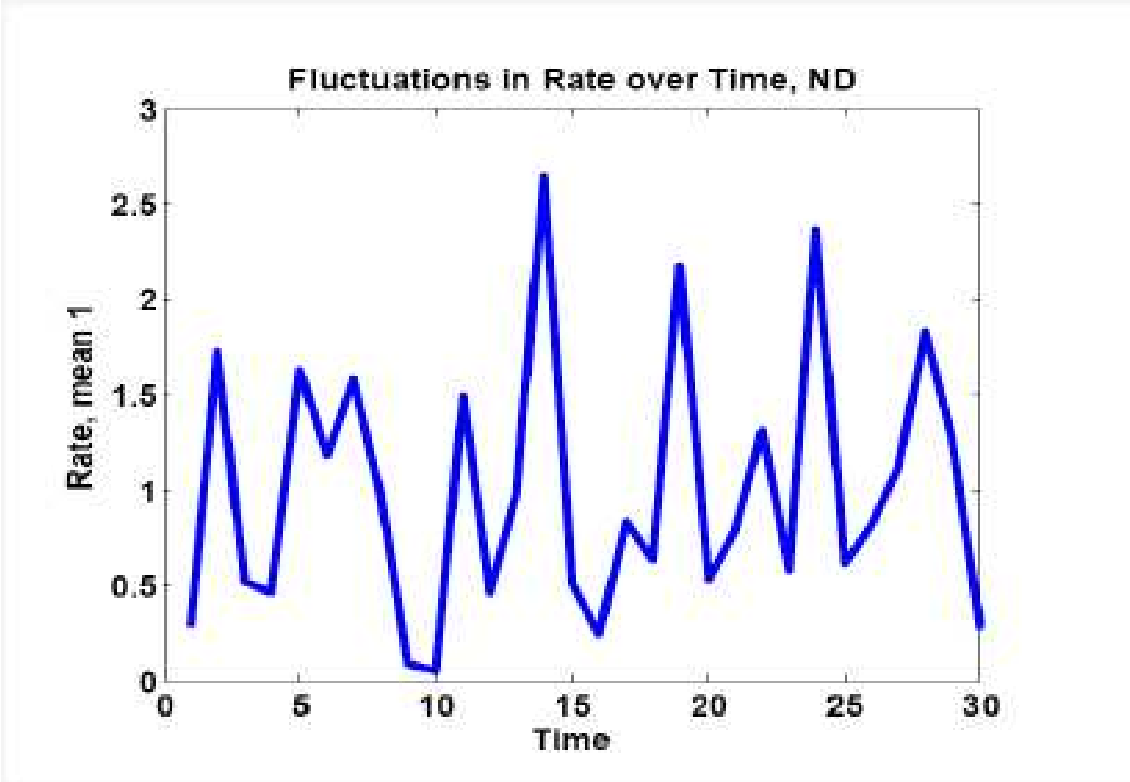
Cognitive Networks

Cognitive networks coordinate the transmission over various bands of frequencies/technologies by exploiting the vacant bands/technologies in idle periods.

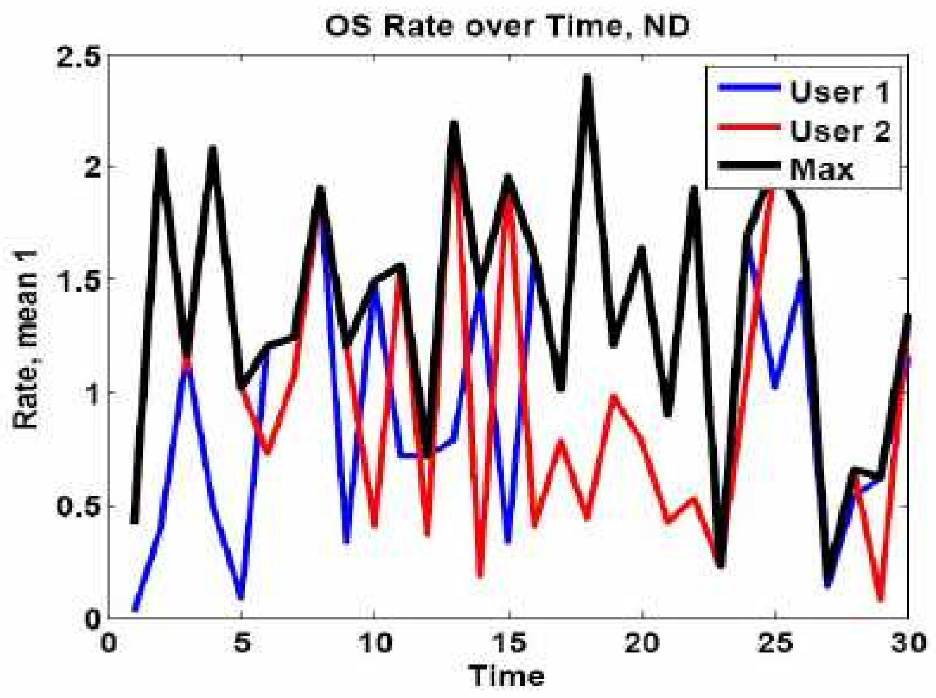
It requires base stations able to work in a large range of bandwidth for sensing the different signals in the network and reconfigure smartly.

Example: They will sense the different technologies, the energy consumption of the terminals and reconfigure (changing from a GSM to UMTS base station if UMTS terminals are present) to adapt to the standards or services to be delivered at a given time.

Opportunistic Networks

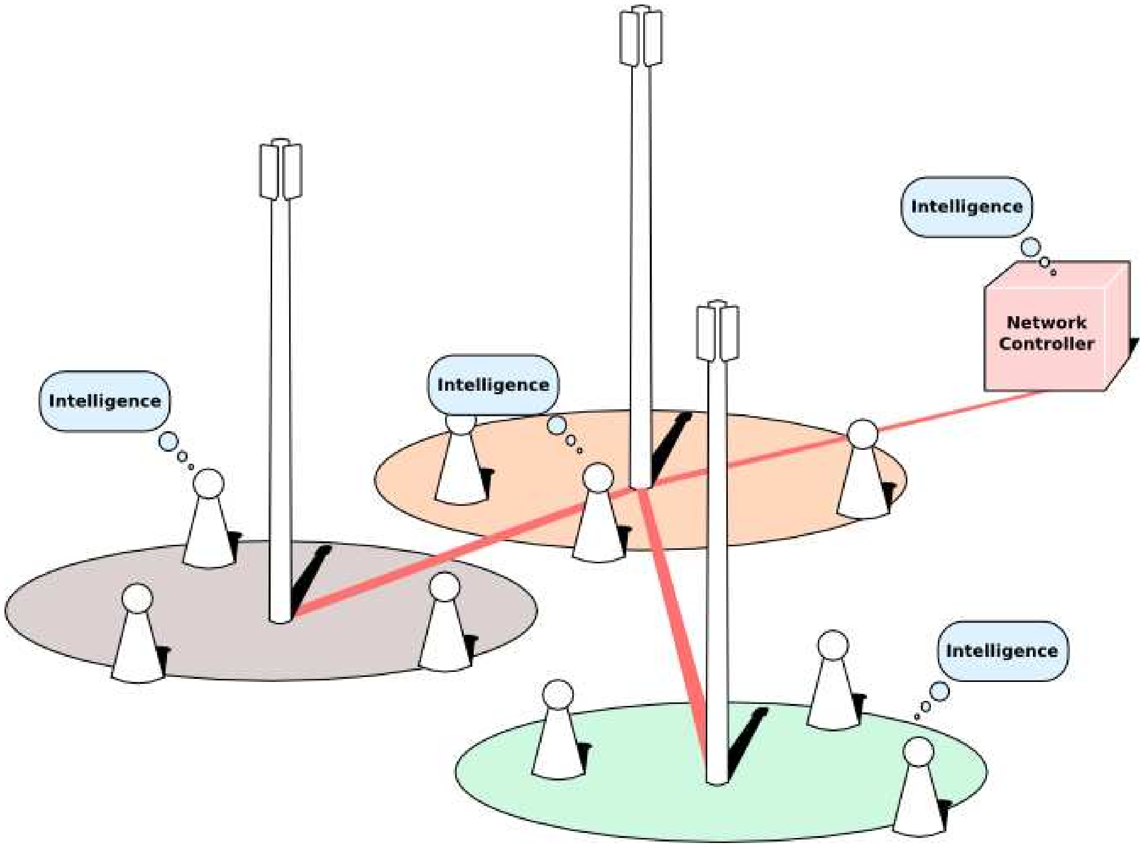


Opportunistic Networks



As the number of users increase, the spectral efficiency of opportunistic network increases, as the probability of having a user with a good channel increases with the number of users.

Intelligent networks



It is the randomness in the network that makes the gains so important...we should exploit it and not combat!

The big dilemma

One of the most challenging problems in the development of this technology is to manage complexity. The key is to develop:

- The right abstractions to reason about the spatial and temporal dynamics of complex systems
- Understand how information can be processed, stored and transferred in the system with bounded delay.

Understanding what is the available ”information” in the network in a given window of observation is fundamental.

General Multiple Input Multiple Output Model

$\mathbf{y}$	$=$	$\mathbf{W}$	$\mathbf{s}$	$+$	$\mathbf{n}$
Received signal		MIMO matrix	emitted signal		AWGN
$N \times 1$		$N \times K$	$K \times 1$		$\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$

Shannon Capacity

Mutual information M between input and output:

$$
\begin{aligned}
M(\mathbf{s};(\mathbf{y},\mathbf{W})) &= M(\mathbf{s};\mathbf{W}) + M(\mathbf{s};\mathbf{y} \mid \mathbf{W}) \\
&= M(\mathbf{s};\mathbf{y} \mid \mathbf{W}) \\
&= H(\mathbf{y} \mid \mathbf{W}) - H(\mathbf{y} \mid \mathbf{s},\mathbf{W}) \\
&= H(\mathbf{y} \mid \mathbf{W}) - H(\mathbf{n})
\end{aligned}
$$

The differential entropy of a complex Gaussian vector $\mathbf{x}$ with covariance $\mathbf{R}$ is given by $\log_2 \det(\pi e \mathbf{R})$.

Shannon Capacity

In the case of Gaussian independent entries, since

$$\mathbb{E}(\mathbf{y}\mathbf{y}^H) = \sigma^2 \mathbf{I}_N + \mathbf{W}\mathbf{W}^H$$
$$\mathbb{E}(\mathbf{n}\mathbf{n}^H) = \sigma^2 \mathbf{I}_N$$

The mutual information per dimension is:

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

Shannon Capacity

Consider the random variable

$$C_N = \frac{1}{N} \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{W}\mathbf{W}^H \right) = \frac{1}{N} \sum_{k=1}^{N} \log \left(1 + \frac{1}{\sigma^2} \lambda_k \left(\mathbf{W}\mathbf{W}^H \right) \right)$$

When $N \to \infty$ and $K/N \to \gamma$ ($\frac{1}{\sigma^2} = \rho$),

$$\begin{aligned} C_N &= \int_0^\infty \log(1 + \rho\lambda) dF(\lambda) \\ &= \gamma \log(1 + \rho - \rho\alpha) + \ln(1 + \rho\gamma - \rho\alpha) - \alpha \end{aligned}$$

with

$$\alpha = \frac{1}{2}[1 + \gamma + \frac{1}{\rho} - \sqrt{(1 + \gamma + \frac{1}{\rho})^2 - 4\gamma}]$$

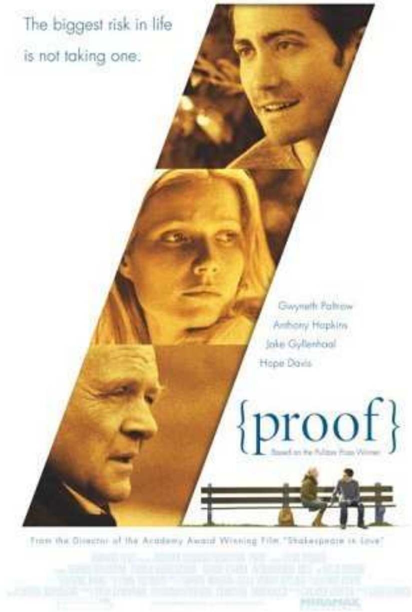
Measurement Validation

On-line course material

http://www.supelec.fr/d2ri/flexibleradio/cours.en.html

Course 1: **Overview and Historical development.**
Course 2: **Probability and convergence measures review.**
Course 3: **Basic Results on Random Matrix Theory**
Course 4: **What about deterministic matrices?**
Course 5: **Stieltjes Transform Method.**
Course 6: **Results on Unitary Random Matrix Theory**
Course 7: **The role of the Cauchy-Stieltjes transform in communications**
Course 8: **Free probability theory and random matrices**
Course 9: **Free deconvolution for signal processing applications**
Course 10 **MIMO Channel Modelling and random matrices**
Course 11: **Asymptotic analysis of (MC)-CDMA systems**
Course 12: **Asymptotic Analysis of MIMO systems**
Course 13: **Asymptotic design of receivers**
Course 14: **Decoding order in receivers**
Course 15: **Game theory and Random matrix theory**
Course 16: **Asymptotic Analysis of Multi-hop relay channels**

Random Matrix Theory



Last Slide

THANK YOU!