Security Level:

Advanced Mathematical Tools for Complex Network Engineering

Mérouane Debbah

Mathematical and Algorithmic Sciences Lab

Huawei

www.huawei.com

HUAWEI TECHNOLOGIES CO., LTD.

Where to start from?

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

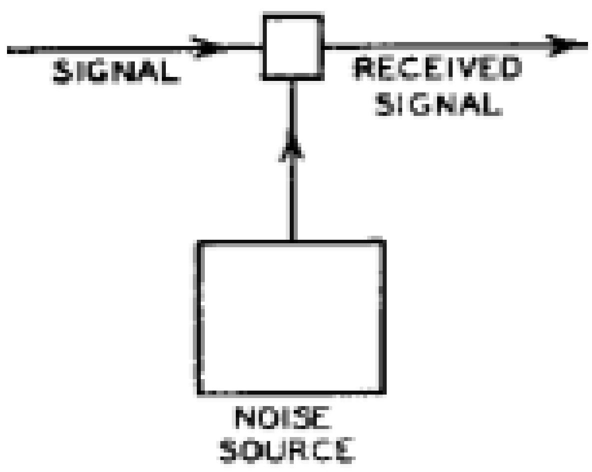
- Expectation: Many Metrics Should Be Improved in 5G
  - Higher user data rates
  - Higher area throughput
  - Great scalability in number of connected devices
  - Higher reliability and lower latency
  - Better coverage with more uniform user rates
  - Improved energy efficiency

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

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



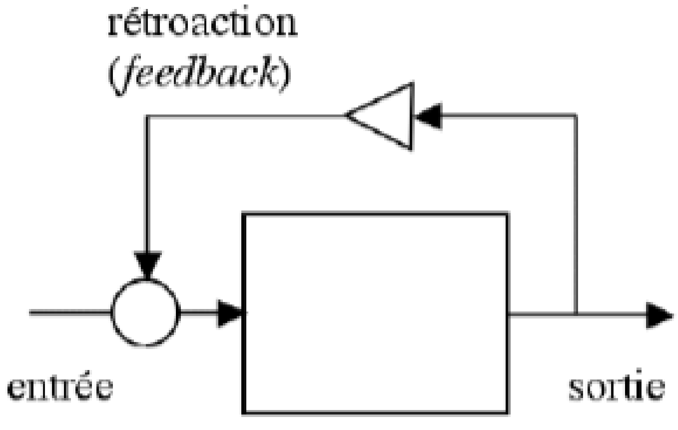
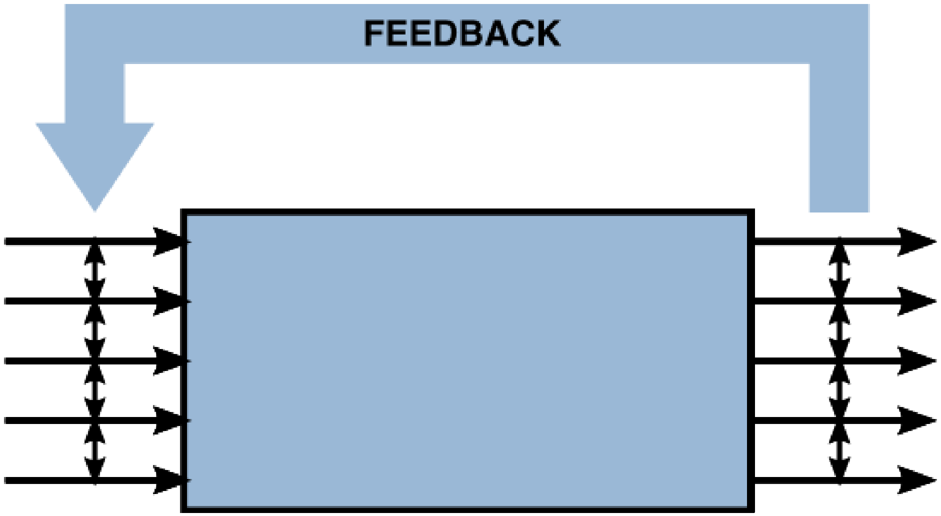


FIG. 1 – Boucle de rétroaction

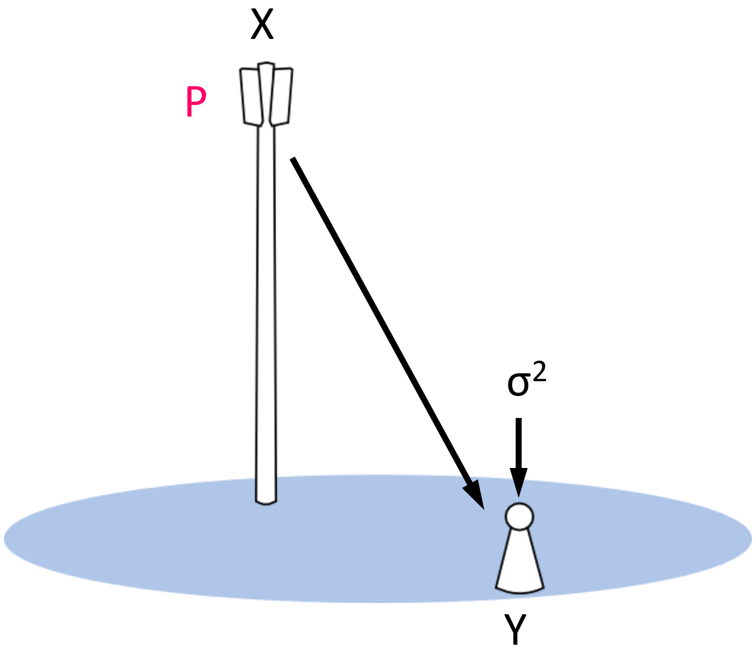
60 years later..the network paradigm…



We must learn and control the black box

- within a fraction of time
- with finite energy.

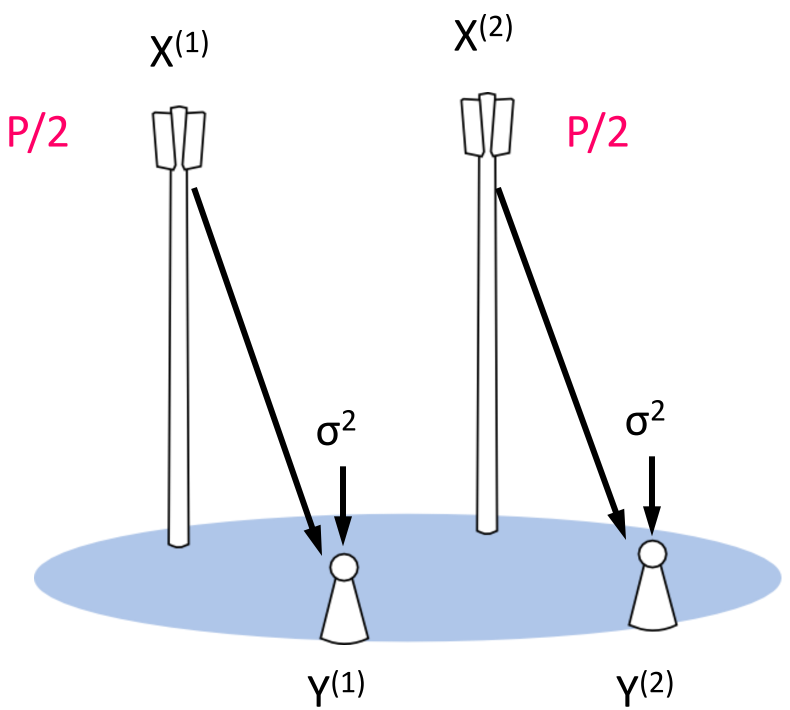
Basics



$C_1 = \log(1+P/\sigma^2)$

Basics

$X = X^{(1)} + X^{(2)}$



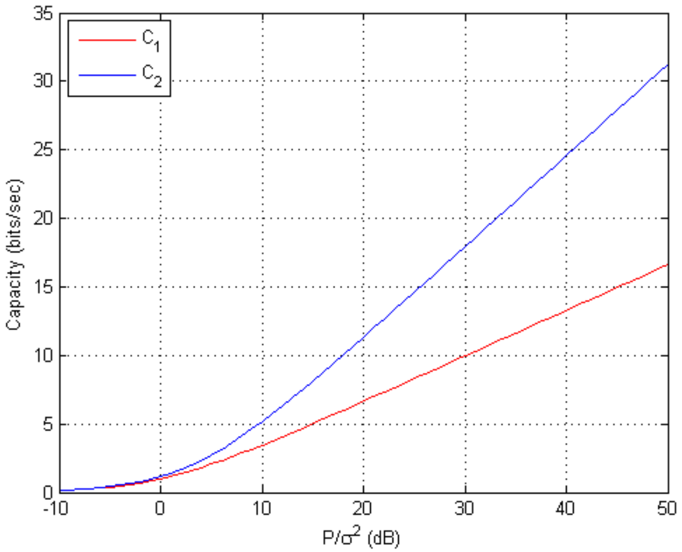
$C_2^{(1)} = \log(1+P/2\sigma^2)$

$C_2^{(2)} = \log(1+P/2\sigma^2)$

$C_2 = C_2^{(1)} + C_2^{(2)} = 2\log(1+P/2\sigma^2)$

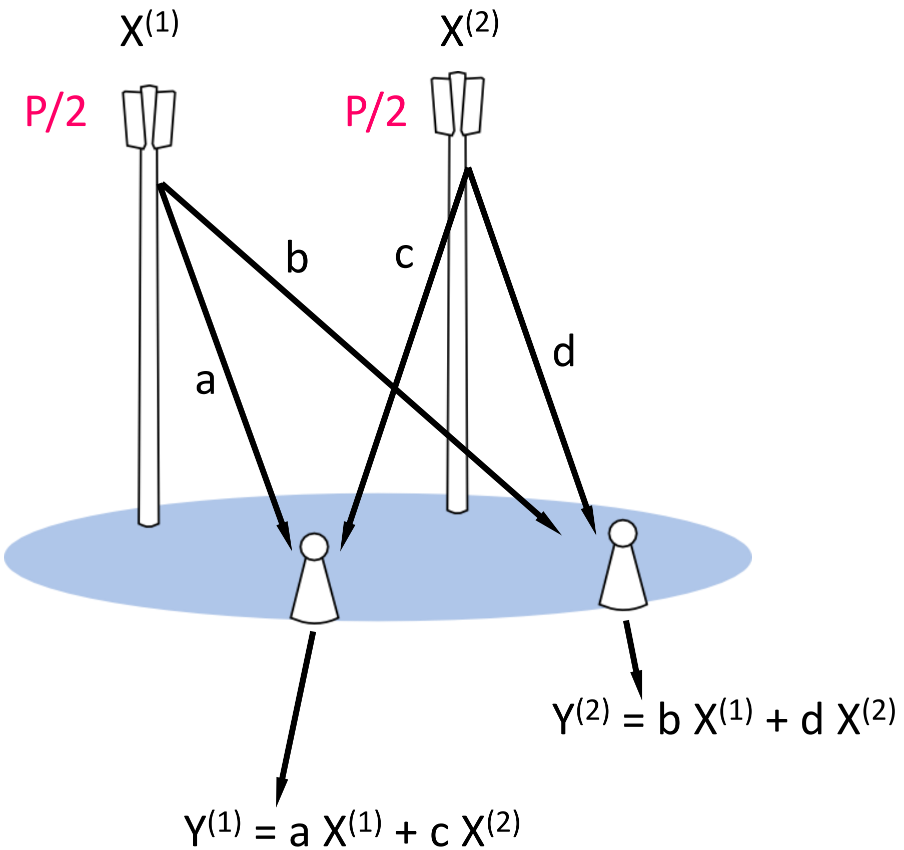
C_1 and C_2, which is better?

$C_1 = \log(1+P/\sigma^2)$ v.s. $C_2 = 2\log(1+P/2\sigma^2)$

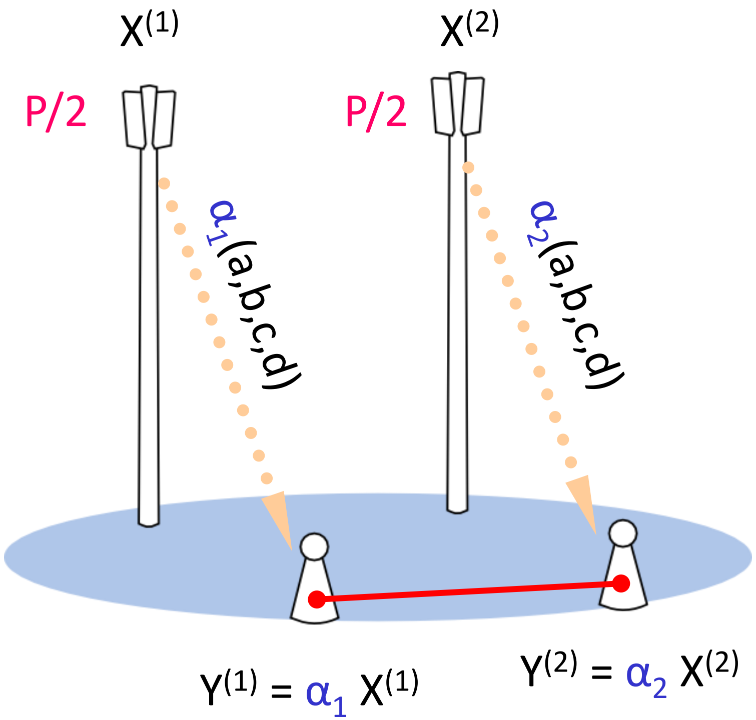


- **Lesson learned:**
  - **High SNR regime**
  - **No need for transmit and receive cooperation**

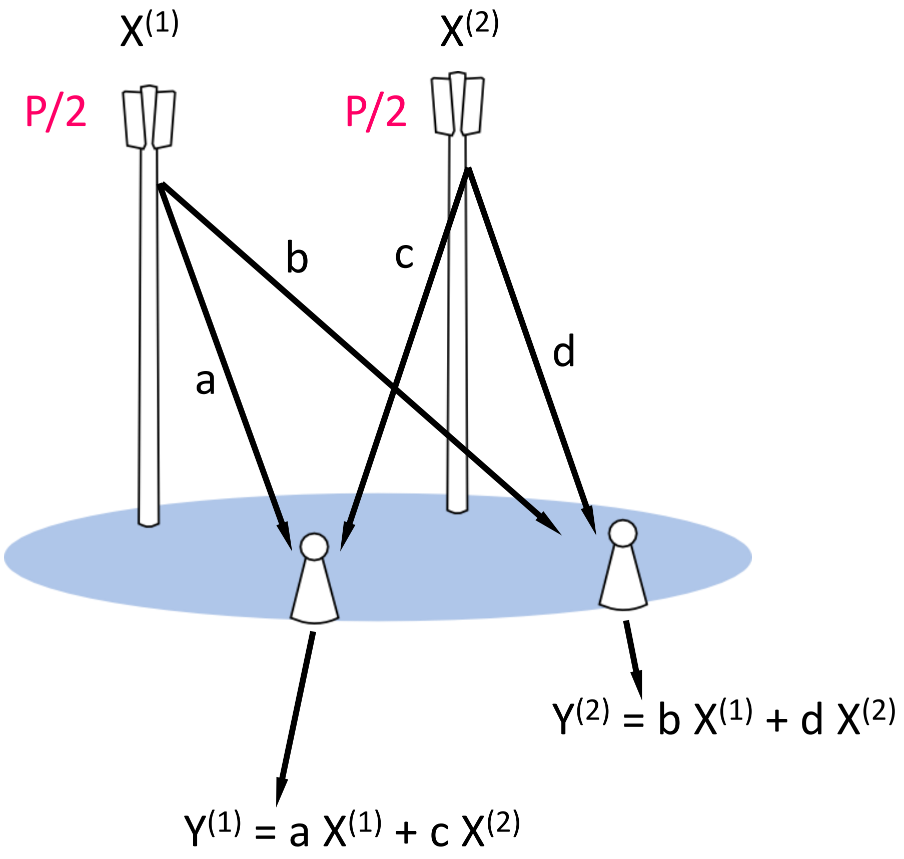
Nature



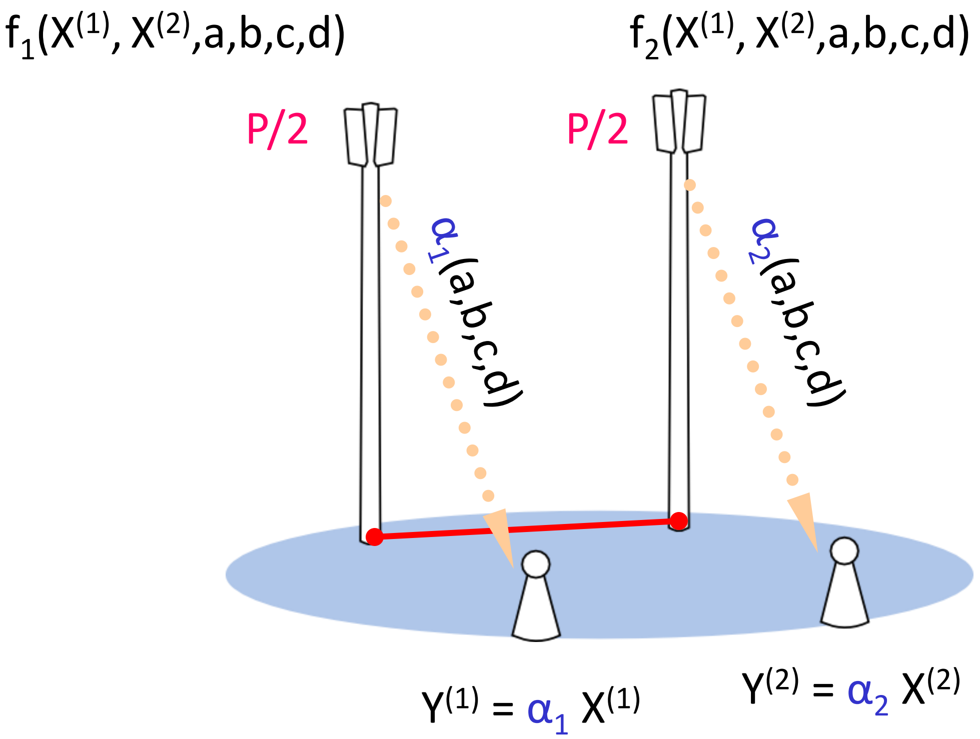
Receiver cooperation

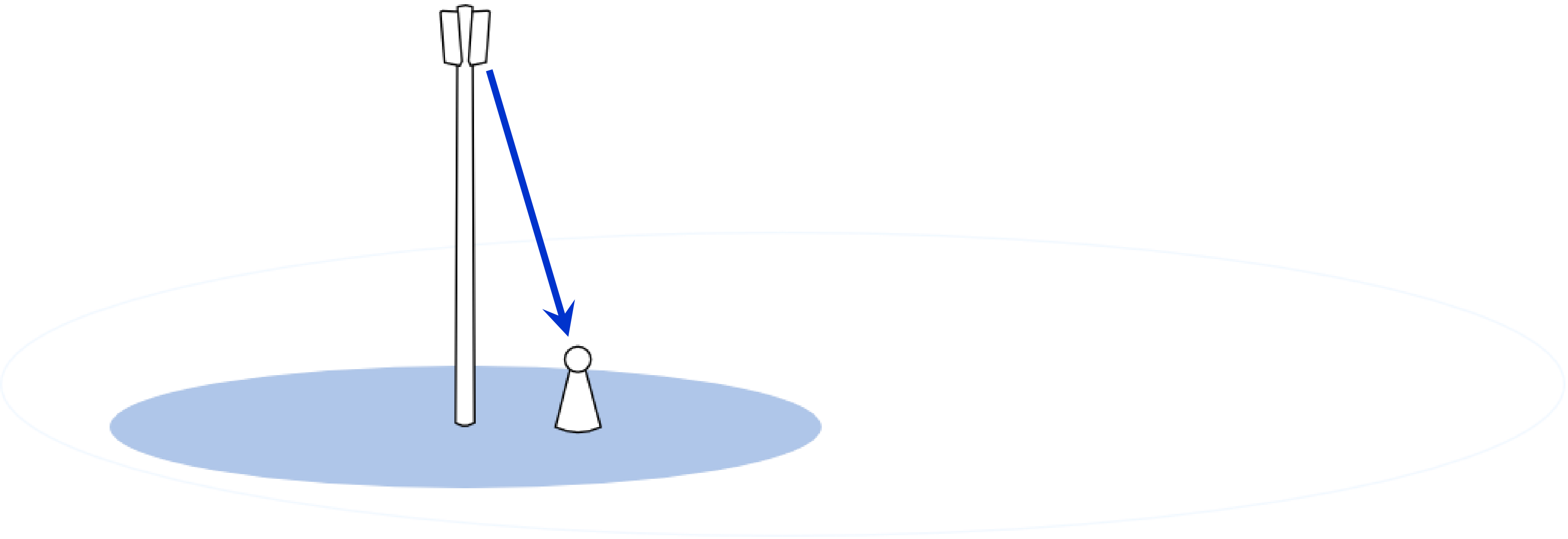


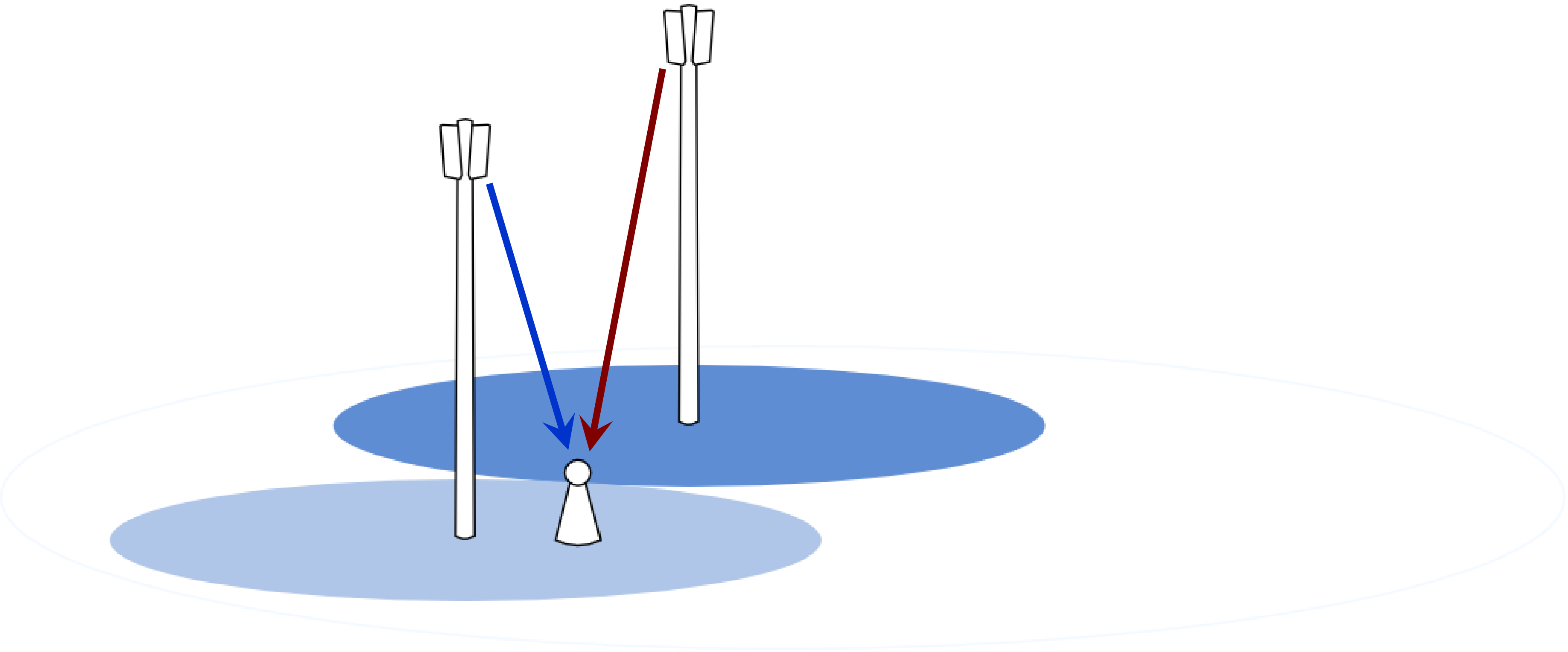
Transmitter cooperation

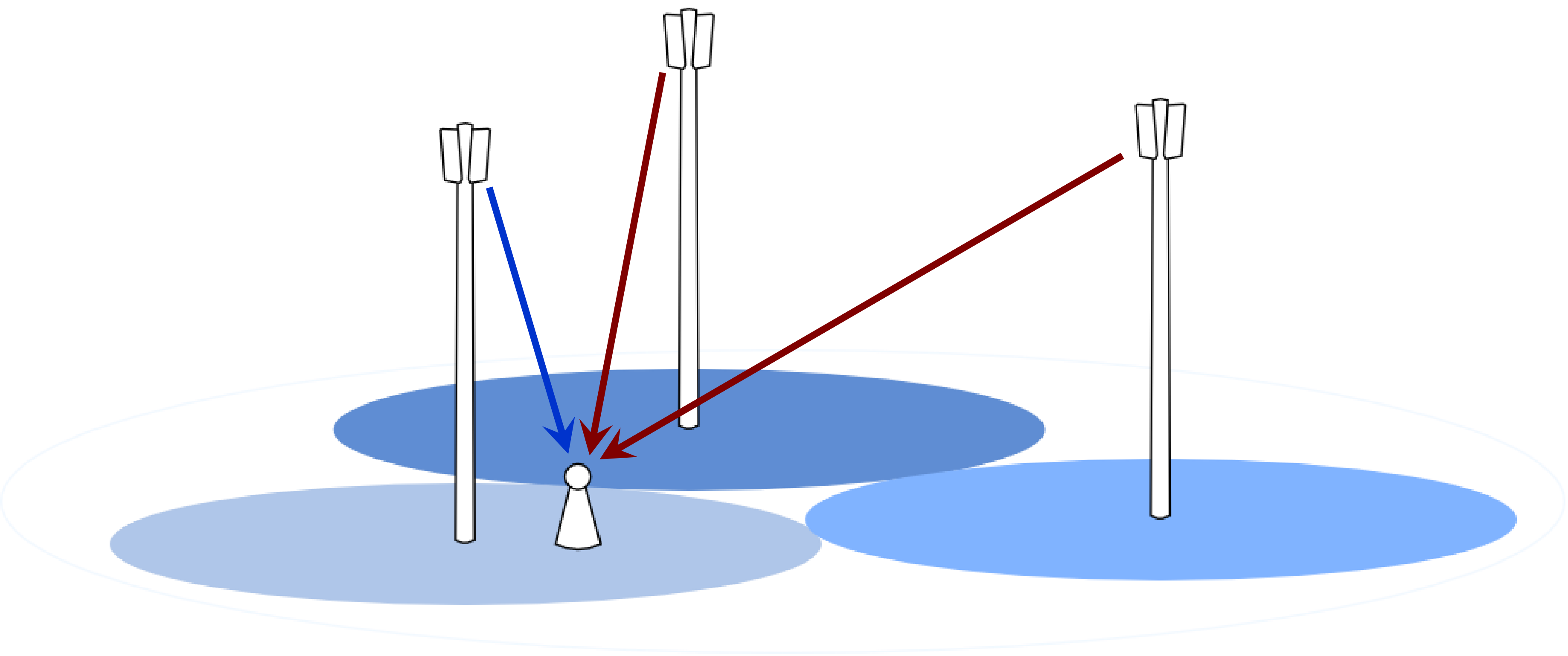


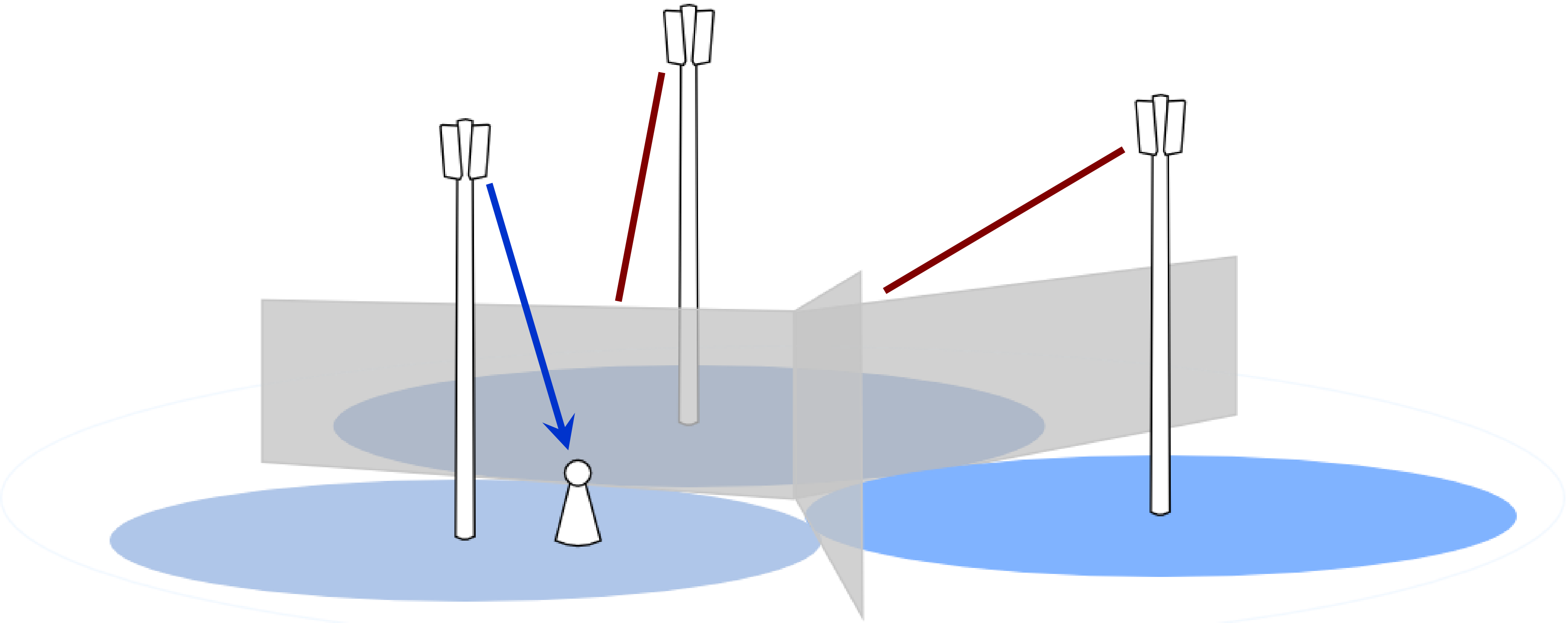
Transmitter cooperation

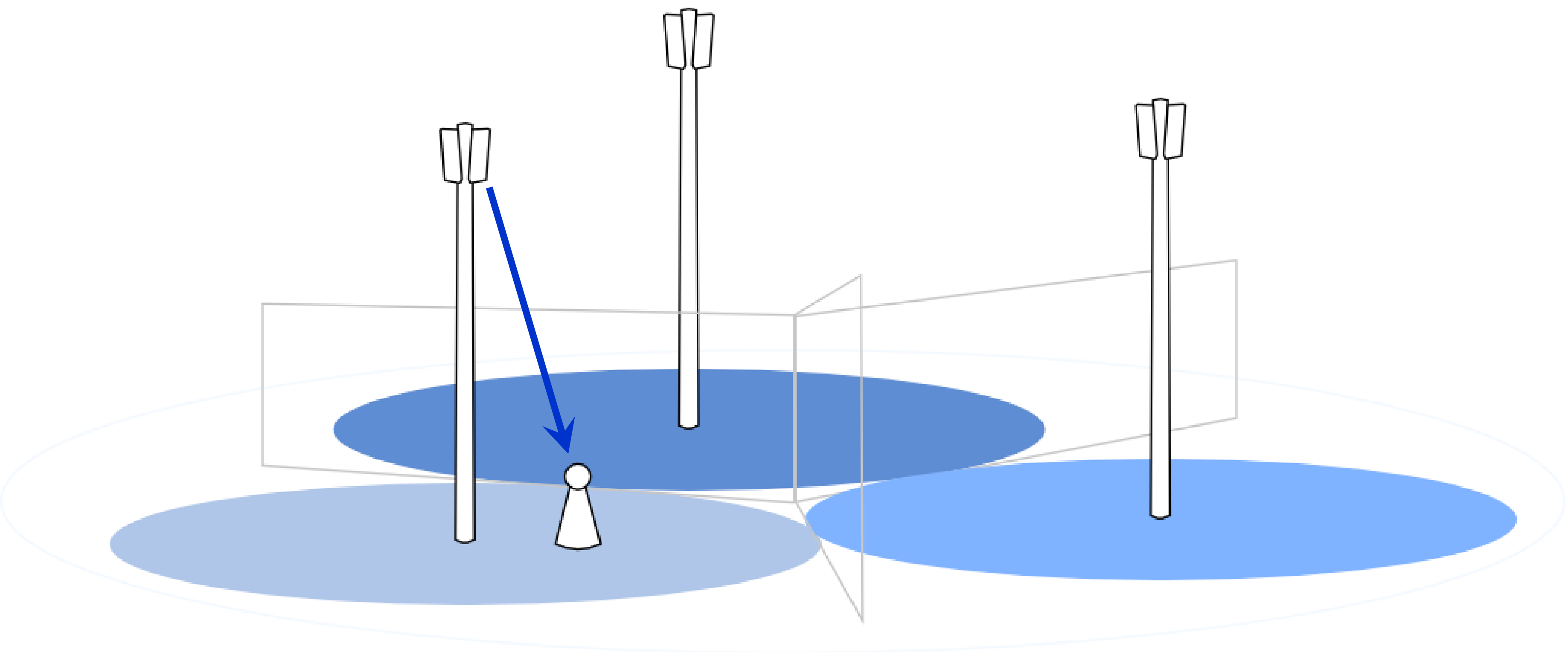




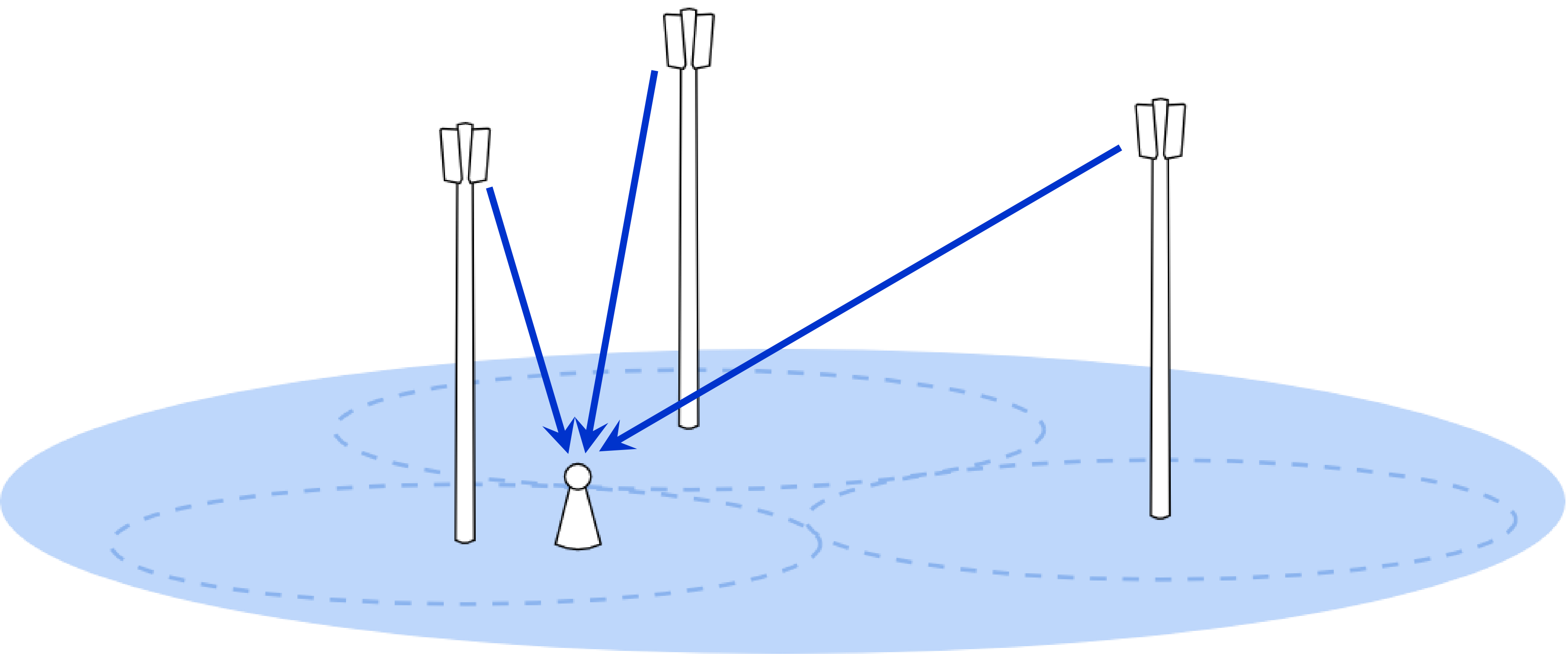


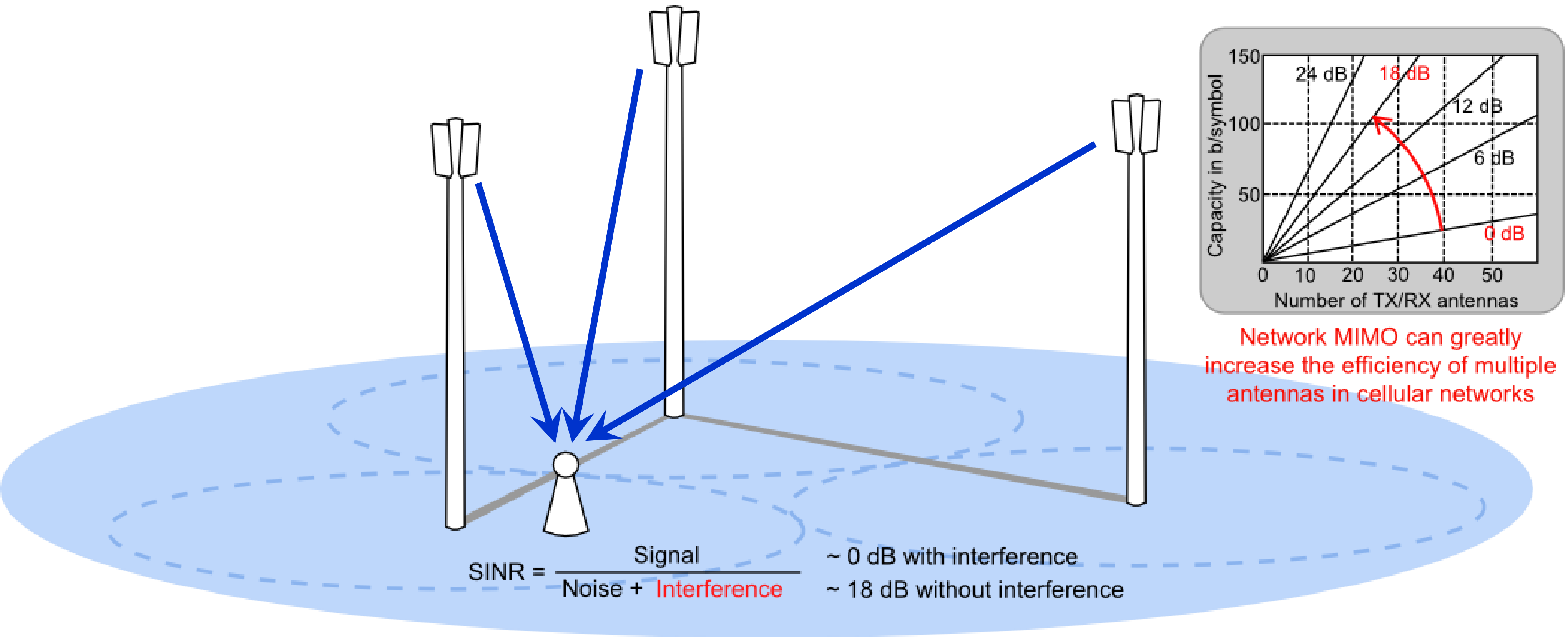


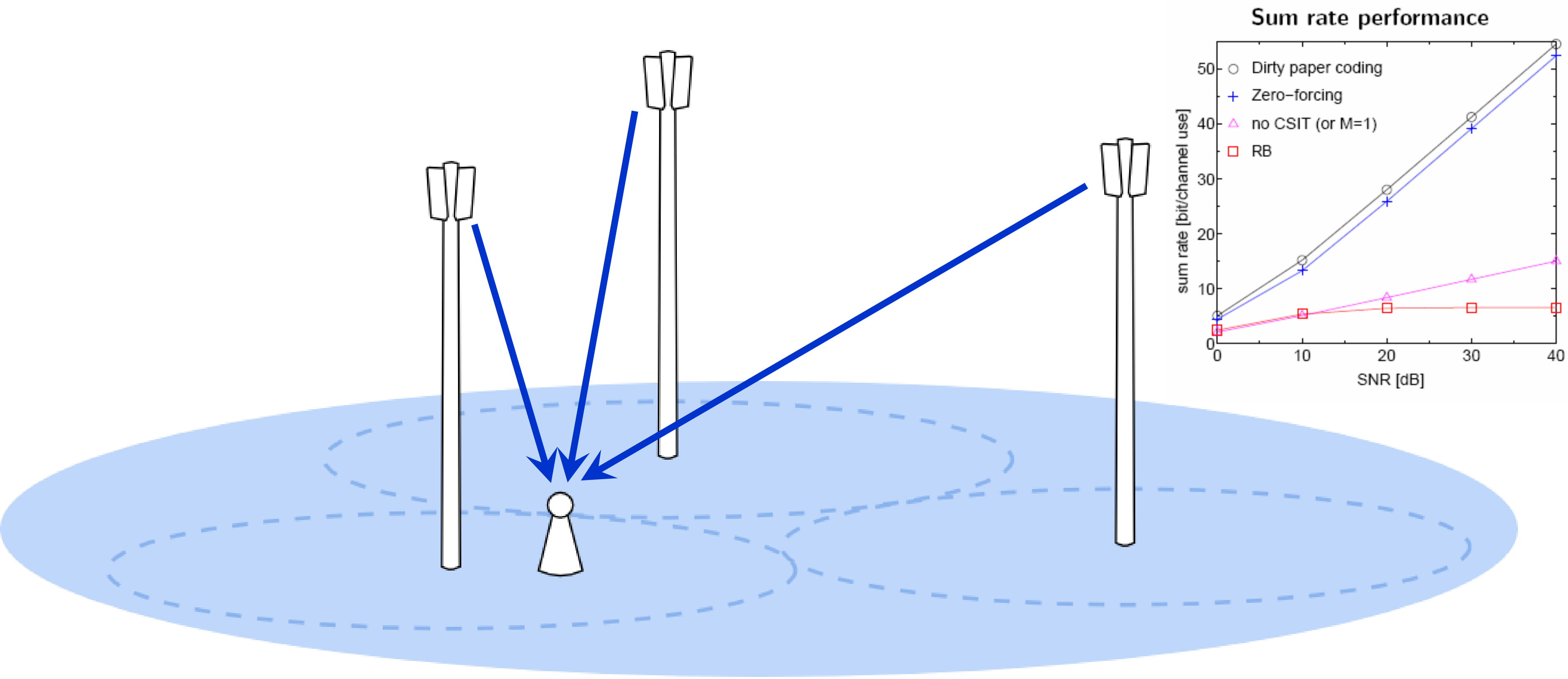




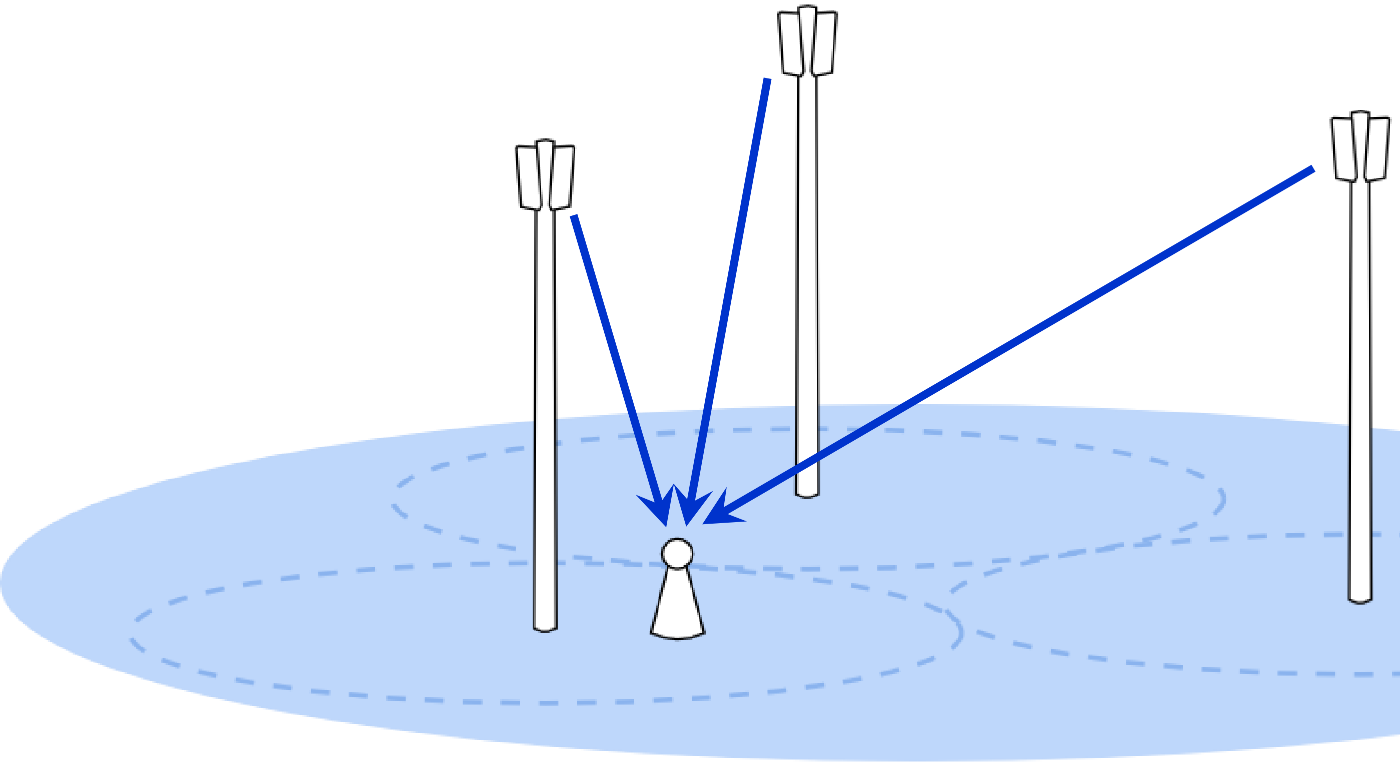
"We build too many walls and not enough bridges."
Isaac Newton

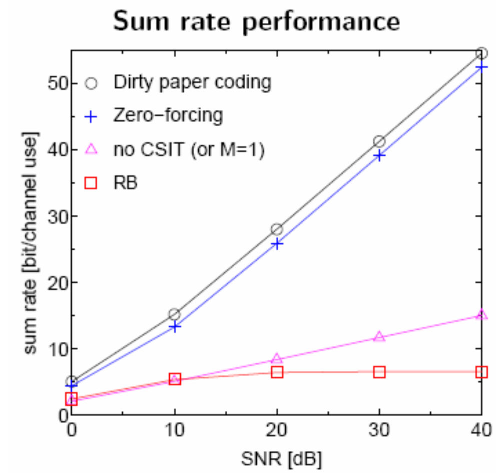



150
100
50
Capacity in b/symbol
24 dB
18 dB
12 dB
6 dB
0 dB
0 10 20 30 40 50
Number of TX/RX antennas
Network MIMO can greatly increase the efficiency of multiple antennas in cellular networks
SINR = Signal / Noise + Interference
~ 0 dB with interference
~ 18 dB without interference



- For large SNR, the sum rate scales as $R_{sum} \sim M \log SNR$ with perfect channel state information at Tx and Rx (CSIT/R)
- The capacity is achieved by a combination of MMSE beamforming and interference pre-cancelation encoding, **Dirty-paper coding**

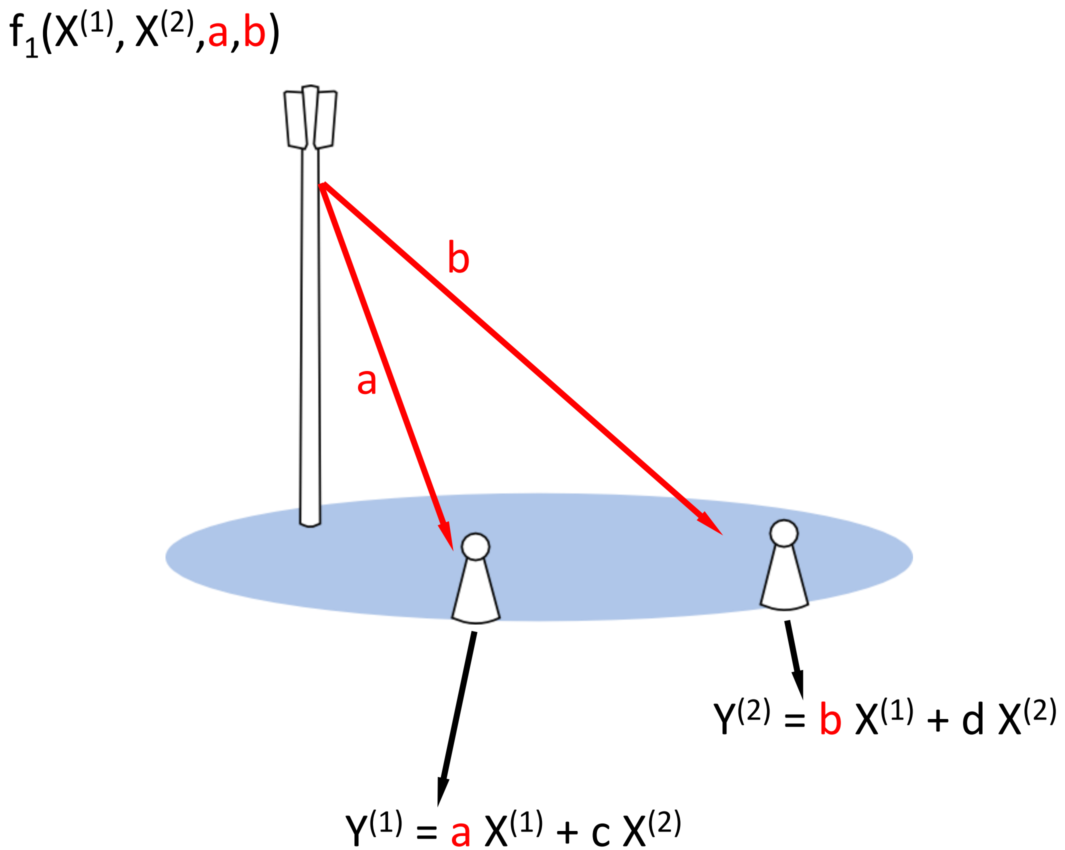




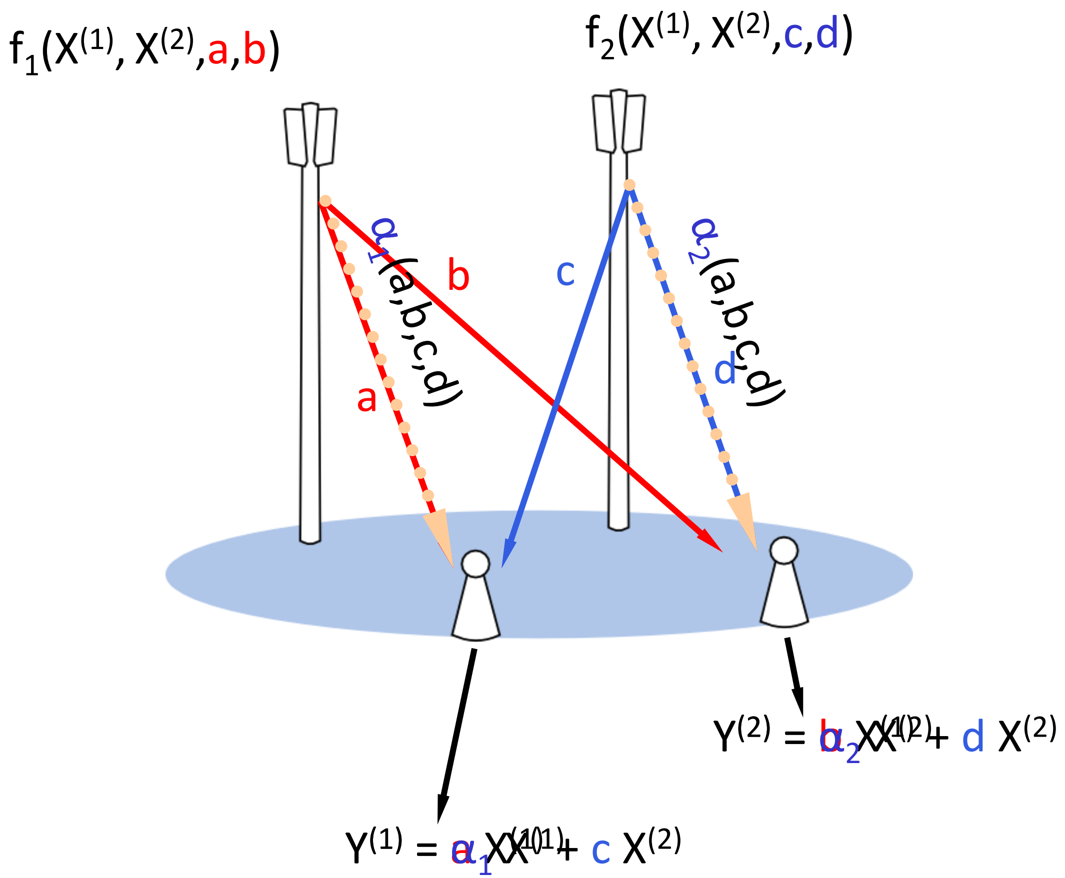
•Recent results show that a naive ZF with training and analog/digital feedback can indeed achieve the full multiplexing.

•The number of feedback bits should scale as log(SNR).

Our Goal: Self-Learning Base Station



Our Goal: Self-Learning Base Station

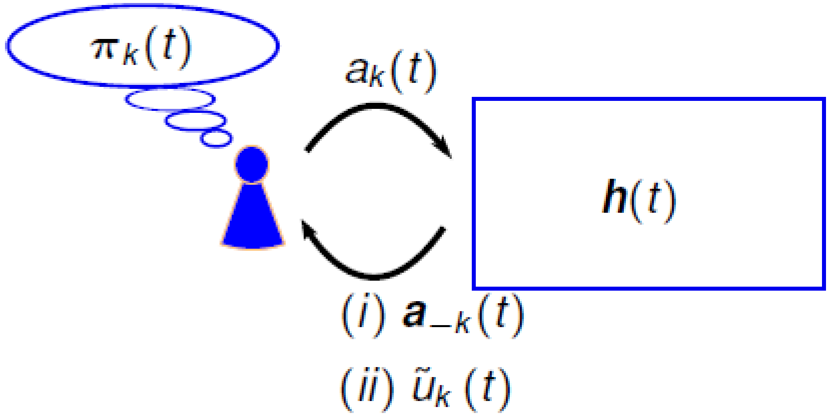



Fun
Mansion.com

Learning Equilibria in Cognitive Radio Networks

Learning Iterative Steps:

- **Choose** action $a_k(t) \sim \pi_k(t)$.
- **Observe** game outcome, e.g.,
  $\boldsymbol{a}_{-k}(t)$
  $u_k(a_k(t), \boldsymbol{a}_{-k}(t))$.
- **Improve** $\pi_k(t+1)$.



Thus, we can expect that: $\forall k \in \mathcal{K}$,

$$\pi_k(t) \xrightarrow{t\to\infty} \pi_k^* \tag{6}$$

$$\bar{u}_k(\pi_k(t), \pi_{-k}(t)) \xrightarrow{t\to\infty} \bar{u}_k(\pi_k^*, \pi_{-k}^*) \tag{7}$$

where, $\pi^* = (\pi_1^*, \ldots, \pi_K^*)$ is a NE strategy profile.

Shannon's Small Self-Learning Mouse

Computers and Automata*

CLAUDE E. SHANNON†, FELLOW, IRE

C. E. Shannon first became known for a paper in which he applied Boolean Algebra to relay switching circuits; this laid the foundation for the present extensive application of Boolean Algebra to computer design. Dr. Shannon, who is engaged in mathematical research at Bell Telephone Laboratories, is an authority on information theory. More recently he received wide notice for his ingenious maze-solving mechanical mouse, and he is well-known as one of the leading explorers into the exciting, but uncharted world of new ideas in the computer field.

The Editors asked Dr. Shannon to write a paper describing current experiments, and speculations concerning future developments in computer logic. Here is a real challenge for those in search of a field where creative ability, imagination, and curiosity will undoubtedly lead to major advances in human knowledge.—*The Editor*

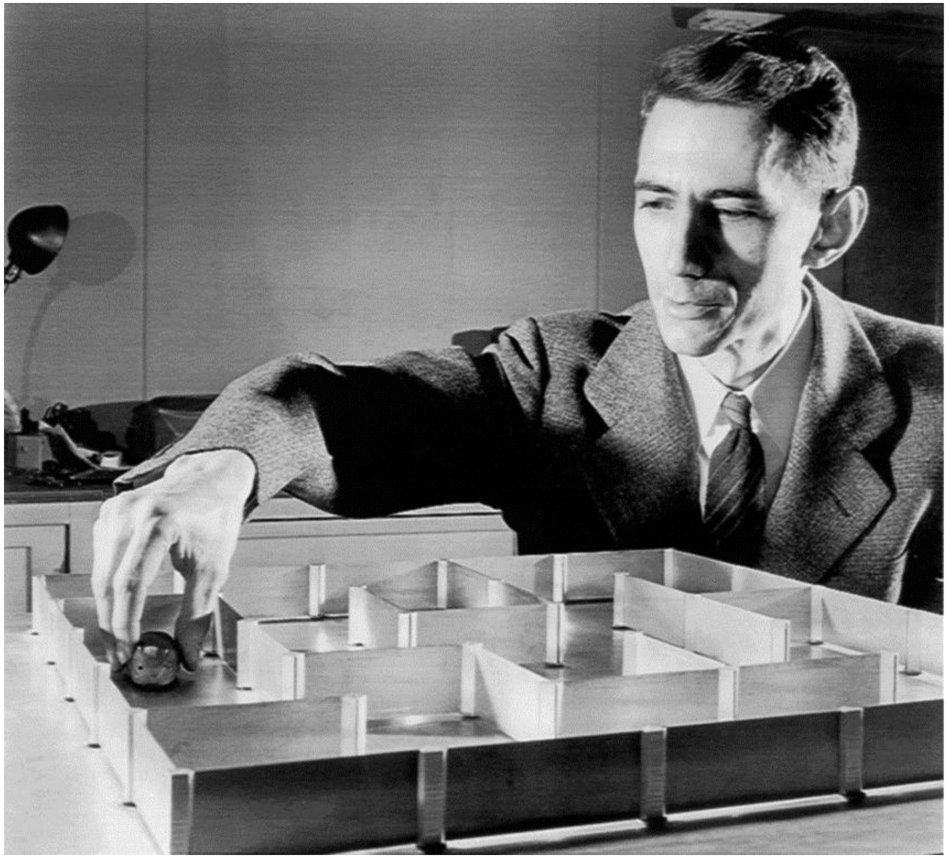
***Summary*—This paper reviews briefly some of the recent developments in the field of automata and nonnumerical computation. A number of typical machines are described, including logic machines, game-playing machines and learning machines. Some theoretical questions and developments are discussed, such as a comparison of computers and the brain, Turing's formulation of computing machines and von Neumann's models of self-reproducing machines.**

* Decimal classification: 621.385.2. Original manuscript received by the Institute, July 17, 1953.

† Bell Telephone Laboratories, Murray Hill, N. J.

Introduction

SAMUEL BUTLER, in 1871, completed the manuscript of a most engaging social satire, *Erewhon*. Three chapters of *Erewhon*, originally appearing under the title "Darwin Among the Machines," are a witty parody of *The Origin of Species*. In the topsy-turvy logic of satirical writing, Butler sees machines as gradually evolving into higher forms. He considers the classification of machines into genera, species and vari-



Computers and Automata*

CLAUDE E. SHANNON†, FELLOW, IRE

C. E. Shannon first became known for a paper in which he applied Boolean Algebra to relay switching circuits; this laid the foundation for the present extensive application of Boolean Algebra to computer design. Dr. Shannon, who is engaged in mathematical research at Bell Telephone Laboratories, is an authority on information theory. More recently he received wide notice for his ingenious maze-solving mechanical mouse, and he is well-known as one of the leading explorers into the exciting, but uncharted world of new ideas in the computer field.

The Editors asked Dr. Shannon to write a paper describing current experiments, and speculations concerning future developments in computer logic. Here is a real challenge for those in search of a field where creative ability, imagination, and curiosity will undoubtedly lead to major advances in human knowledge.—*The Editor*

Summary—**This paper reviews briefly some of the recent developments in the field of automata and nonnumerical computation. A number of typical machines are described, including logic machines, game-playing machines and learning machines. Some theoretical questions and developments are discussed, such as a comparison of computers and the brain, Turing's formulation of computing machines and von Neumann's models of self-reproducing machines.**

* Decimal classification: 621.385.2. Original manuscript received by the Institute, July 17, 1953.

† Bell Telephone Laboratories, Murray Hill, N. J.

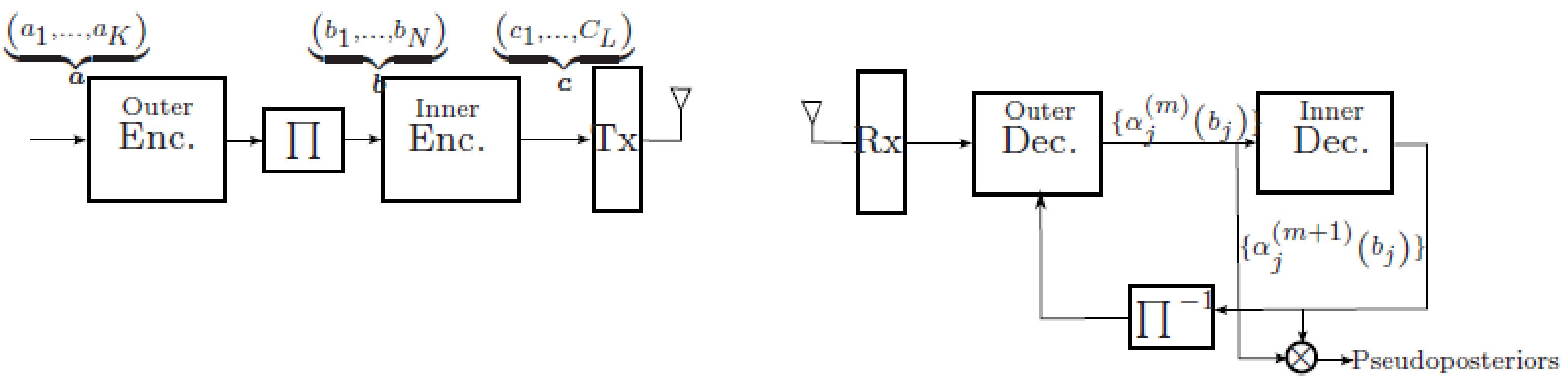
Introduction

SAMUEL BUTLER, in 1871, completed the manuscript of a most engaging social satire, *Erewhon*. Three chapters of *Erewhon*, originally appearing under the title "Darwin Among the Machines," are a witty parody of *The Origin of Species*. In the topsy-turvy logic of satirical writing, Butler sees machines as gradually evolving into higher forms. He considers the classification of machines into genera, species and vari-

Motivation: How To Tackle the Challenge

Basic Turbo Deconding Concept:

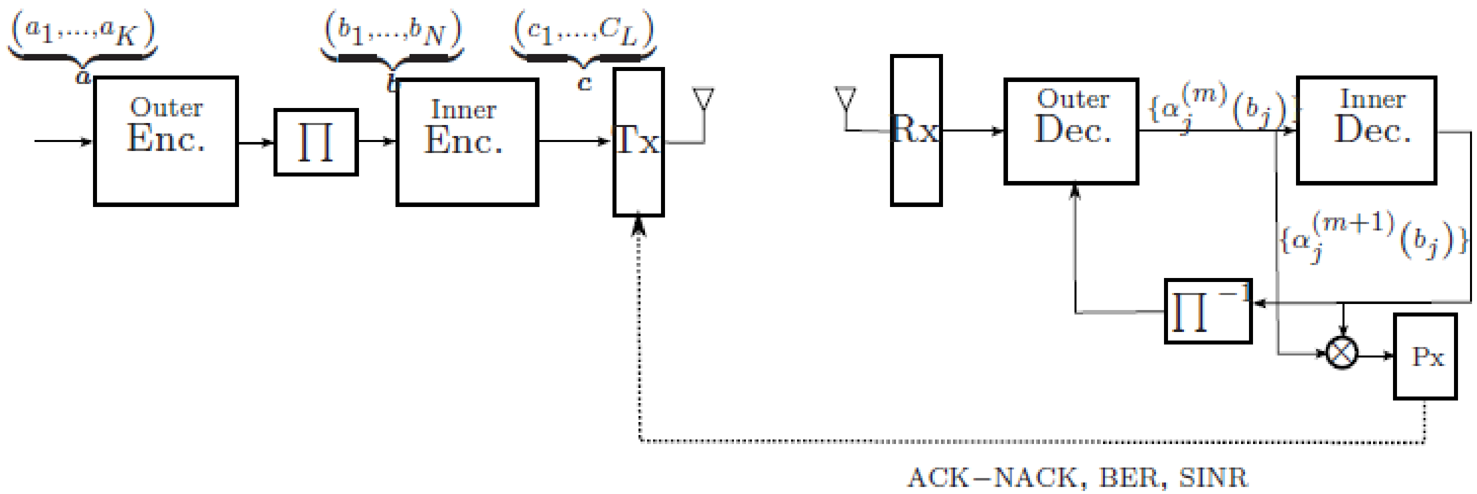
- Two decoders iterate beliefs (probability measures) about the received sequence.
- The iteration process produces a countinous improvement of the beliefs
- The process converges when both decoders produce the same beliefs.



Motivation: How To Tackle the Challenge

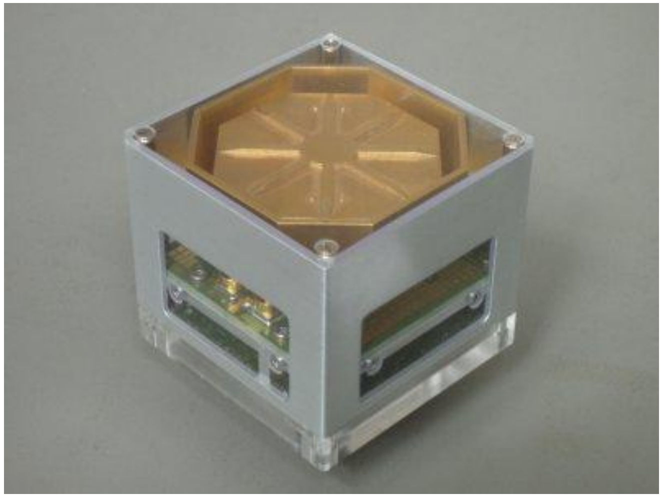
What about iterating with the transmitter?

e.g., To use ACK/NACK, BER or SINR feedback to iteratively improve a belief of the optimal transmission configuration, given the existence of other CRs.

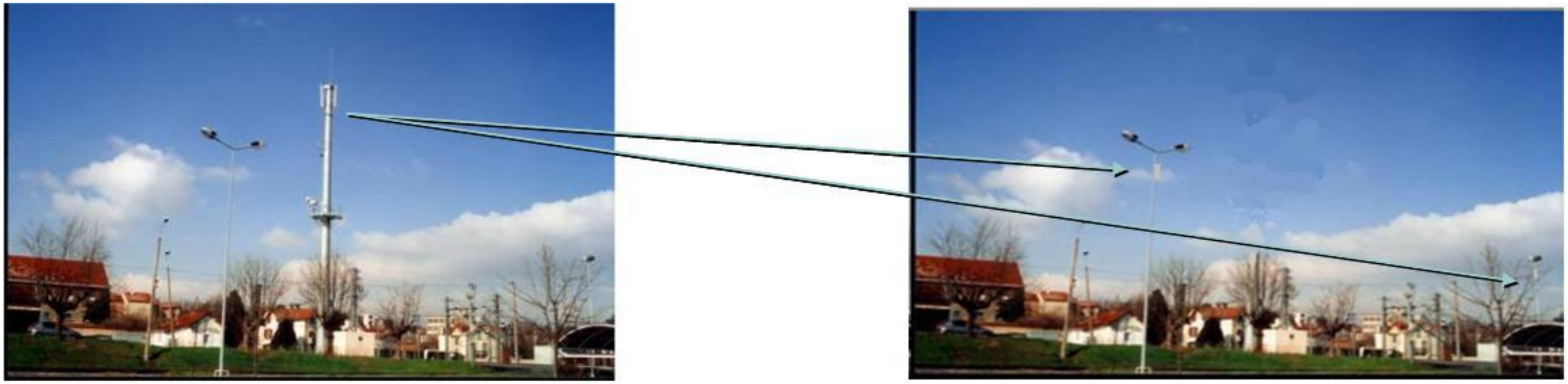


Vision

- 1Gbps/Km^2 for 10 MHz
- Environment constraints = <1W EIRP
- Constraint: ~10 W power consumption

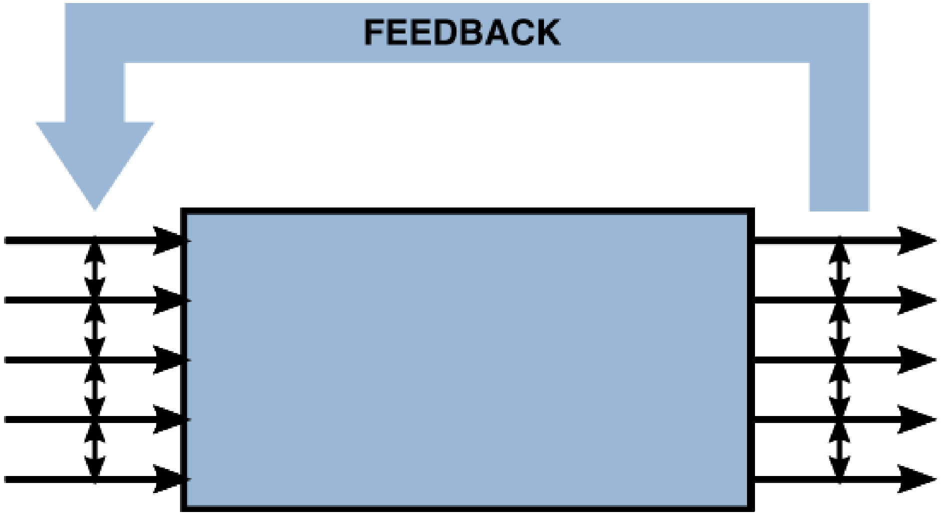


Bell Labs lightradio antenna module – the next generation small cell
(picture from www.washingtonpost.com)




DIAS
TRAFALGAR S°

Let us go back to the MIMO paradigm…

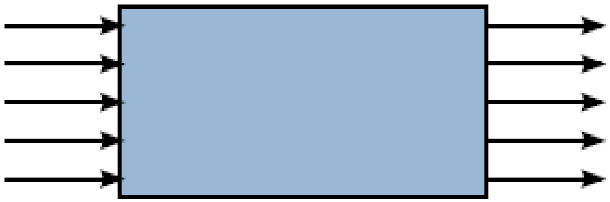


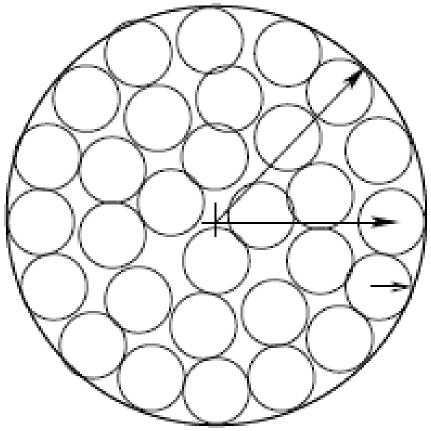
We must learn and control the black box

- within a fraction of time
- with finite energy.

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:

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

Channel Modelling Perspective

Where do we stand on Channel Modelling

Google search: "MIMO Wireless Channel Modelling"

- Over 15 000 publications on channel modelling
- At a rate of 10 papers per day, 1 500 days (nearly 4 years)!
- The models are different and many validated by measurements!

Three conflicting schools

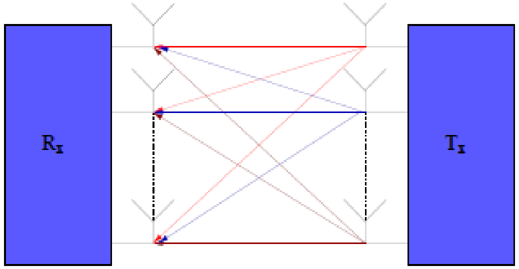
- Geometry based channel models.
- Stochastic channel models based on channel statistics
- Do not model, use test measurements

Not even within each school, all experts agree on fundamental issues.

How to increase YOUR number of publications

or how to increase the length and shrink the content...

MIMO System Model



The channel is linear, noise is additive

$$\mathrm{y}(u) = \sqrt{\frac{\mathrm{SNR}}{n_t}} \int \mathrm{H}_{n_r \times n_t}(\tau)\mathrm{x}(u-\tau)d\tau + \mathrm{n}(u)$$

$$\mathrm{Y}(f,u) = \sqrt{\frac{\mathrm{SNR}}{n_t}} \mathrm{H}_{n_r \times n_t}(f,u)\mathrm{X}(f) + \mathrm{N}(f)$$

u, f, SNR are respectively time, frequency and the SNR per receive antenna.

My first experience in channel modelling

Suppose that you read a paper and find a "good model" (which complies with capacity measurements for example) M represented by the structured matrix H.

Create the new model M_1 represented by the structured matrix $H_1 = \Theta H$ (where Θ is a unitary transform, Fourier matrix for example so that Θ represents some directions of arrival).

Call it the double scattering model (and pretend that the Θ matrix represents the information coming from scatterers to the receiving antennas)

You have also a very good model since:

$$\begin{aligned} \log \det(I + \frac{\text{SNR}}{n_t} H_1 {H_1}^H) &= \log \det(I + \frac{\text{SNR}}{n_t} \Theta H H^H \Theta^H) \\ &= \log \det(I + \frac{\text{SNR}}{n_t} H H^H) \end{aligned}$$

The two models can predict very well capacity measurements. Which one of the models should we choose?

Occam's razor

The razor states that:

- Entities are not to be multiplied without necessity.
- The simplest answer is usually the correct answer.

William Ockham, 1295-1349

General idea: When multiple competing theories have equal predictive powers, the principle recommends selecting those that introduce the fewest assumptions and postulate the fewest hypothetical entities **Note:** Occam's razor (due to the franciscan friar Ockham) is also known as the principle of parsimony.

Information theory and channel modeling

Still, occam's razor does not provide us the answer and has no theoretical background.

$$\Theta H \rightarrow H \rightarrow H_{\text{ideal}}$$

Questions information theory should answer:

- Why is Occam's razor true?
- If is true then what is the most adequate (parsimonious or less redundant or less parametrized) representation of the channel?

Model Comparison

How to compare models?

What is the right model, in other words how to choose between the set $\{M_0, M_1, \ldots, M_K\}$ of K models?

Different methods based on different metrics:

- Parameter estimation methods
- Mutual information
- Bayesian methods

Is there a universal metric?

Model Comparison

Parameter Estimation methods

- The data is cut in two parts.
- The first part is used for estimation.
- The second part is used to test the model with respect to the mean square error.

Deficiencies:

- How does one cut the set of data?
- By adding more and more parameters, one can always fit the model to the data!

Model Comparison

Mutual information compliance

- Explicit expressions of the mutual information are derived.
- Find the optimal parameters of the problem such as the model has the same mutual information as measurements.

Deficiencies:

- Mutual information is a symmetric measure:

$$I(\mathrm{x};\mathrm{y}) = I(\mathrm{y};\mathrm{x}). \tag{1}$$

- We can exchange transmitter and receiver without any change of mutual information (although the model is different)!

Bayesian Viewpoint

- Ranking of the models based on computing their likelihood.
- The method punishes models with too many parameters over those with fewer by the fact that they are less likely to be true.
- It can be proved that there is an optimum number of parameters when representing information which fulfills Occam's razor.

Model Comparison

What is the right model?

Bayes rule gives the posterior probability for the i^{th} model according to:

$$P(M_i \mid Y, I) = P(M_i \mid I) \frac{P(Y \mid M_i, I)}{P(Y \mid I)}$$

For comparing two models M and M_1, one has to compute the ratio:

$$\frac{P(M_1 \mid Y, I)}{P(M \mid Y, I)} = \frac{P(M_1 \mid I)}{P(M \mid I)} \frac{P(Y \mid M_1, I)}{P(Y \mid M, I)}$$

If π is the set of parameters on which is based the model, then

$$\begin{aligned} P(Y \mid M, I) &= \int P(Y, \pi \mid M, I) d\pi \\ &= \int P(Y \mid \pi, M, I) P(\pi \mid M, I) d\pi \end{aligned}$$

What is the right model?

Comparing with traditional methods

In traditional "advanced methods", one maximizes the likelihood $P(Y \mid \hat{\pi}, M, I)$.

But what if both models have the same likelihood?

let us expand $\log P(Y \mid \hat{\pi}, M, I)$ around the maximum likelihood point
$\hat{\pi} = \pi^1{}_{\max}, \ldots, \pi^m{}_{\max}\}$

$$\log P(Y \mid \hat{\pi}, M, I) = \log P(Y \mid \hat{\pi}_{\max}, M, I) + \frac{1}{2} \sum_{i,j=1}^{m} \frac{d^2 \log(P)}{d\pi^i d\pi^j} (\pi^i - \pi^i{}_{\max})(\pi^j - \pi^j{}_{\max})$$

then near the peak a good approximation is a multivariate gaussian such as:

$$P(Y \mid \hat{\pi}, M, I) = P(Y \mid \hat{\pi}_{\max}, M, I) e^{-\frac{1}{2}(\hat{\pi} - \hat{\pi}_{\max})^T \Delta^{-1} (\hat{\pi} - \hat{\pi}_{\max})}$$

with the inverse covariance matrix defined as:

$$\Delta^{-1}{}_{ij} = \left(\frac{d^2 \log(P)}{d\pi^i d\pi^j} \right)_{\pi = \pi_{\max}}$$

What is the right model?

A common misconception

Occam's razor: Nature prefers simplicity!

$$\begin{aligned} P(Y \mid M, I) &= P(Y \mid \hat{\pi}_{\text{max}}, M, I) \int e^{-\frac{1}{2}(\pi - \hat{\pi}_{\text{max}})^T \Delta^{-1} (\hat{\pi} - \hat{\pi}_{\text{max}})} P(\pi \mid M, I) d\pi \\ &= P(Y \mid \hat{\pi}_{\text{max}}, M, I) G(M, I) \end{aligned}$$

Usually, the more model has parameters, the more $P(\pi \mid M, I)$ is spread out and $G(M, I)$ decreases.

Indeed:

$$\begin{aligned} P(\pi, \alpha \mid M, I) &= P(\pi \mid M, I, \alpha) P(\alpha \mid M, I) \\ \leq P(\pi \mid M, I, \alpha) \end{aligned}$$

What is the right model?

There are three terms to compare....

$$\frac{P(M \mid Y, I)}{P(M_1 \mid Y, I)} = \frac{P(M \mid I)}{P(M_1 \mid I)} \frac{P(Y \mid M, I)}{P(Y \mid M_1, I)}$$
$$= \frac{P(M \mid I)}{P(M_1 \mid I)} \frac{P(Y \mid \hat{\pi}_{\max}, M, I)}{P(Y \mid \hat{\pi}_{1\max}, M_1, I)} \frac{G(M, I)}{G(M_1, I)}$$

Remark: The term $\frac{P(M_1|I)}{P(M|I)}$ can be seen as the revenge of the measurement field scientist over the mathematician. It shows that modelling is both an experimental and theoretical science and that the experience of the field scientist (which attributes the values of the prior probabilities) does matter.

Clearing up mysteries

Model validation is a question of trading adequacy for complexity.

Usual validation methods validate functionals (mutual information..) of the model and not the model itself (transforming a multi-dimensional problem into a one dimensional problem).

This explains why so many models comply with measurements!

Models have to be constructed in a consistent manner with the functionals of the model of interest i.e marginalizing over the functionals should lead to consistent solutions.

A distribution probability consistent model (as we did) is for the case where we do not know which functions we are interested in.

Information and Complexity: One can know less but understand more

- Should we take into account all the information provided, in other words, what information is useful?
- From a bayesian point of view, we should consider all the available information BUT there is a compromise to be made in terms of model complexity.
- Each information added will not have the same effect on the channel model and might as well more complicate the model for nothing rather than bring useful insight.
- To assume further information by putting some additional structure would not lead to incorrect predictions:
  - However, if the predictions achieved with or without the details are equivalent
  - Then this means that the details may exist but are irrelevant for the understanding of our model

Let us start...

Model Construction

The i.i.d Gaussian model

The modeler would like to attribute a joint probability distribution to:

$$\mathbf{H}(f) = \begin{pmatrix} h_{11}(f) & \dots & \dots & h_{1n_t}(f) \\ \vdots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ h_{n_r1}(f) & \dots & \dots & h_{n_rn_t}(f) \end{pmatrix} \quad (2)$$

Assumption 1: The modeler has no knowledge where the transmission took place (the frequency, the bandwidth, the type of room, the nature of the antennas...)

Assumption 2: The only things the modeler knows:
For all $\{i, j\}$,

$$\mathbb{E}(\sum_{i,j} \mid h_{ij} \mid^2) = n_r n_t E$$

What distribution $P(\mathbf{H})$ should the modeler assign to the channel based only on that specific knowledge?

The i.i.d Gaussian model

Principle of maximum entropy

Maximize the following expression:

$$-\int d\mathbf{H}P(\mathbf{H})\log P(\mathbf{H})+\gamma[n_r n_t E-\int d\mathbf{H}\sum_{i=1}^{n_r}\sum_{j=1}^{n_t}\mid h_{ij}\mid^2 P(\mathbf{H})]$$
$$+\beta\left[1-\int d\mathbf{H}P(\mathbf{H})\right]$$

Solution:

$$P(\mathbf{H})=\frac{1}{(\pi E)^{n_r n_t}}\exp\{-\sum_{i=1}^{n_r}\sum_{j=1}^{n_t}\frac{\mid h_{ij}\mid^2}{E}\}$$

Contrary to past belief, the i.i.d Gaussian model is not an assumption but the result of finite energy knowledge.

This method can be extensively used whenever additional information is provided in terms of expected values.

Knowledge of the covariance structure

In the general case, under the constraint that:

$$\int_{\mathbb{C}^N} h_i h_j^* P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}) d\mathbf{H} = q_{i,j}$$

for $(i,j) \in [1,\ldots,N]^2$ $(N = n_r n_t)$. Then using Lagrangian multipliers,

$$\begin{aligned} L(P_{\mathbf{H}|\mathbf{Q}}) &= \int_{\mathbb{C}^N} -\log(P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}))\mathrm{P}_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} \\ &+ \beta\left[1 - \int_{\mathbb{C}^N} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H}\right] \\ &+ \sum \alpha_{i,j}\left[\int_{\mathbb{C}^N} h_i h_j^* P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} - \mathrm{q_{i,j}}\right]. \end{aligned}$$

we obtain:

$$P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}) = \frac{1}{\det(\pi\mathbf{Q})}\exp\left(-(\mathrm{vec}(\mathbf{H})^H\mathbf{Q}^{-1}\mathrm{vec}(\mathbf{H}))\right).$$

Existence of Correlation

Question

What to do if we know the existence of correlation but not its exact value?

Answer

$$P(\mathbf{H}) = \int P(\mathbf{H}, \mathrm{Q})d\mathrm{Q} = \int P(\mathbf{H} \mid \mathrm{Q})P(\mathrm{Q})d\mathrm{Q}$$

1- Determine the a priori distribution of the covariance matrix based on limited information at hand
2- Marginalize with respect to the a priori distribution

Construction of the a priori

Let us determine the a priori distribution of the covariance

Suppose that we only know that $\mathbb{E}(\mathrm{Trace}(\mathbf{Q})) = n_r n_t E$ (The covariance is not fixed but varies due to mobility for example)

Result. The maximum entropy distribution for a covariance matrix $\mathbf{Q}$ under the constraint $\mathbb{E}(\mathrm{Trace}(\mathbf{Q})) = n_r n_t E$ is such as:

$$\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$$

where:

- $\mathbf{U}$ is Haar unitary distributed matrix.
- $\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_{n_r n_t}\right)$ is diagonal matrix with independent Laplacian distributions.

$$P(\mathbf{Q})d\mathbf{Q} = \frac{1}{E^{n_r n_t}} \Pi_{n=0}^{n_r n_r - 1}(n!(n+1)!) e^{-\frac{\mathrm{Trace}(\mathbf{Q})}{E}} \Pi_{i>j}(\lambda_i - \lambda_j)^2 d\mathbf{U} d\mathbf{\Lambda}$$

Note that $\mathbf{Q}$ is nothing else than a Wishart matrix with $n_r n_t$ degrees of freedom.

MIMO Channel distribution with correlation

What do we need to do?

$$
\begin{aligned}
P(\mathbf{H}) &= \int P(\mathbf{H} \mid \mathrm{Q}) P(\mathrm{Q}) d\mathrm{Q} \\
&= \int \int \frac{1}{\pi \prod_{i=1}^{n_r n_t} \lambda_i} e^{-\mathrm{Trace}(\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H \mathbf{U}^H \mathbf{\Lambda}^{-1} \mathbf{U})} \\
&\quad \frac{1}{E^{n_r n_t}} \Pi_{n=0}^{n_r n_r - 1} (n!(n+1)!) e^{-\frac{\sum_i \lambda_i}{E}} \Pi_{i>j} (\lambda_i - \lambda_j)^2 d\mathbf{U} d\mathbf{\Lambda}
\end{aligned}
$$

We need to integrate over $\mathbf{U}$ and $\mathbf{\Lambda}$!

Difficult problem...but well known in statistical physics!

MIMO Channel distribution with correlation

Harish-Chandra, "Differential Operator on a Semi-Simple Lie Algebra", Amer. J. Math. 79 87-120 (1957)

Harish-Chandra, 1923-1983

Harish-Chandra integral

$$\int_{\mathbf{U}\in U(m)} e^{-m\mathrm{Trace}(\mathbf{\Sigma}^{-1}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1})} d\mathbf{U} = \frac{\det(e^{-\sigma_j^{-1}\lambda_k})}{\Delta(\Sigma^{-1})\Delta(\Lambda)}$$

MIMO Channel distribution with correlation

"Maximum Entropy Analytical MIMO Channel Models", M. Guillaud, M. Debbah and A. Moustakas, submitted to IEEE transactions on Information Theory, 2007.

Solution. $P(\mathbf{H})$ is given by

$$P(\mathbf{H}) = \sum_{n=1}^{n_r n_t} 2\left(\frac{\mathrm{Trace}(\mathbf{H}\mathbf{H}^H)}{E}\right)^{\frac{n+n_r n_t-2}{2}} K_{n+n_r n_t-2}\left(2\sqrt{\frac{\mathrm{Trace}(\mathbf{H}\mathbf{H}^H)}{E}}\right). \frac{(-1)^{n n_r n_t}}{[(n-1)!]^2 (n_r n_t - n)!}$$

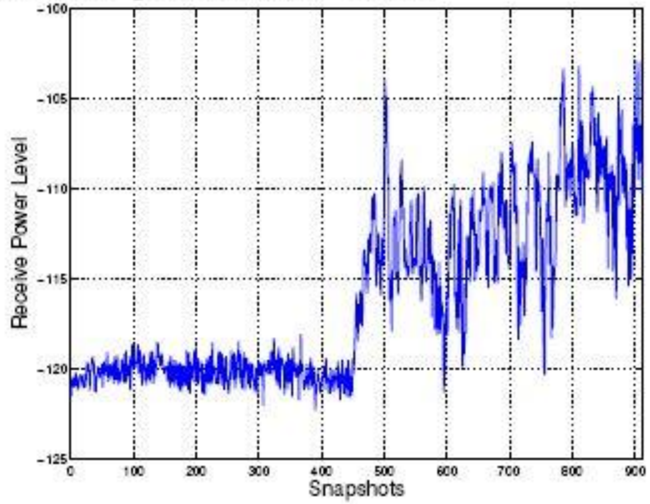
$K_n(x)$ are bessel functions of order n.

We have therefore an explicit form that can be used for design when correlation exists in the MIMO model but we are not aware of the explicit value of the correlation!

Figure 1.4: Mutual information of a 2x2 (dual polarized at the BS) MIMO system in an urban area for a SNR of 10 dB and constant transmit power.

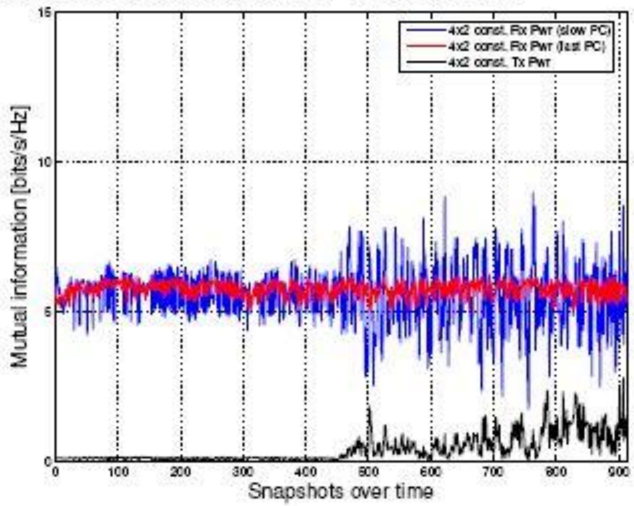
Figure 1.10: Mutual information of a 1x1 SISO system in an urban area for a SNR of 10 dB and constant transmit power.

Receive power over time:

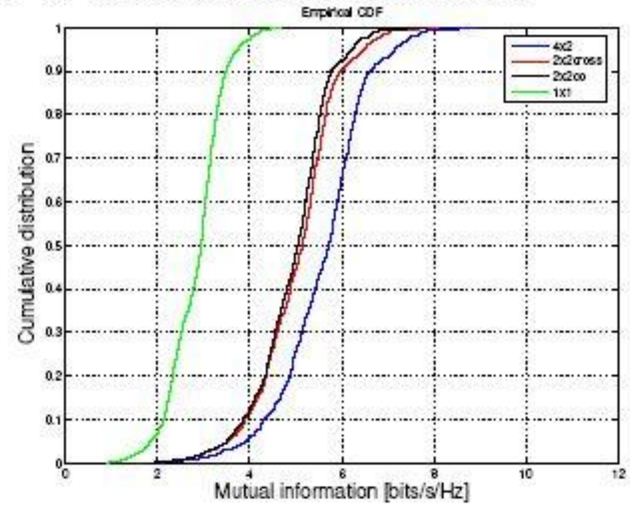


Map of the measurement run:

Mutual information over time:

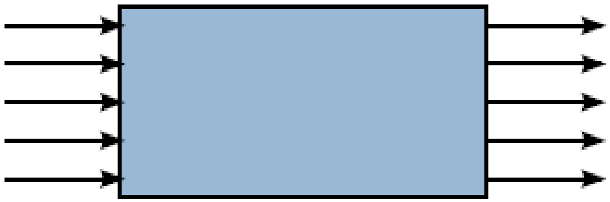


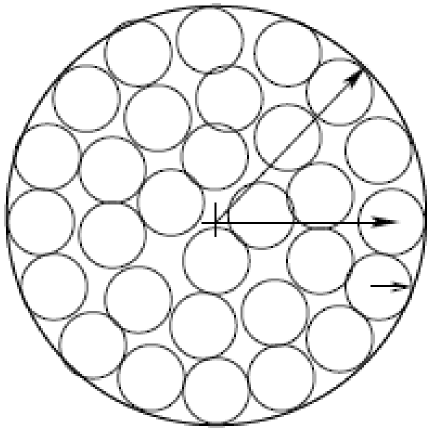
CDF of the Mutual information:



Information transfer in MIMO flexible networks

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





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

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

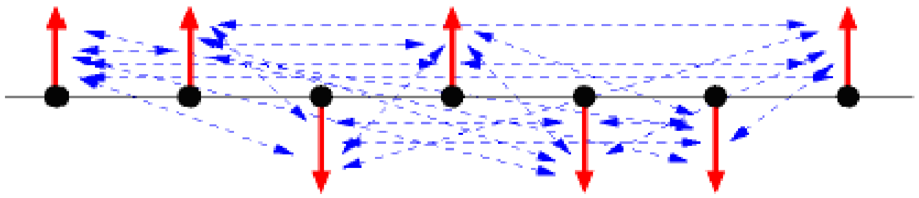
The rate is:

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

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

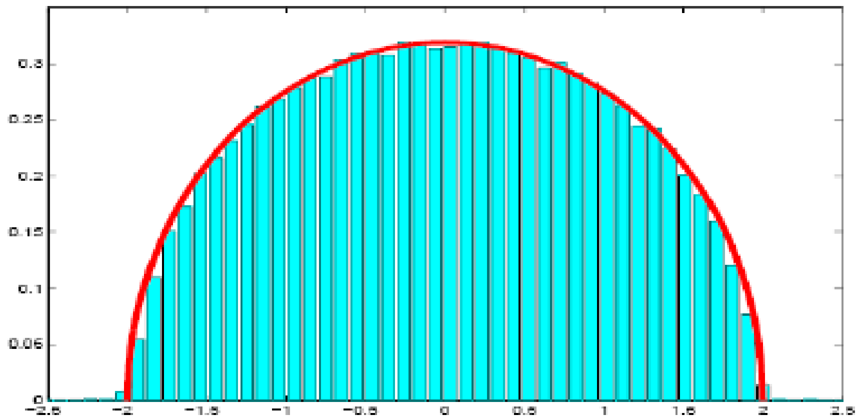


Figure 2: The semicircle law density function (4) compared with the histogram of the average of 100 empirical density functions for a Wigner matrix of size $n = 100$.

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:

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

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

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

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,

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

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.

Square Matrix of i.i.d coefficients

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

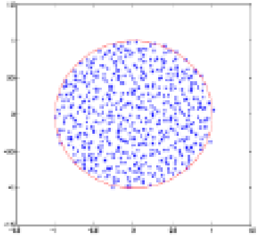


Figure 3: The full-circle law and the eigenvalues of a realization of a 500×500 matrix

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.

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 $b = (1 + \sqrt{\alpha})^2$.

The Marchenko-Pastur Distribution Law

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

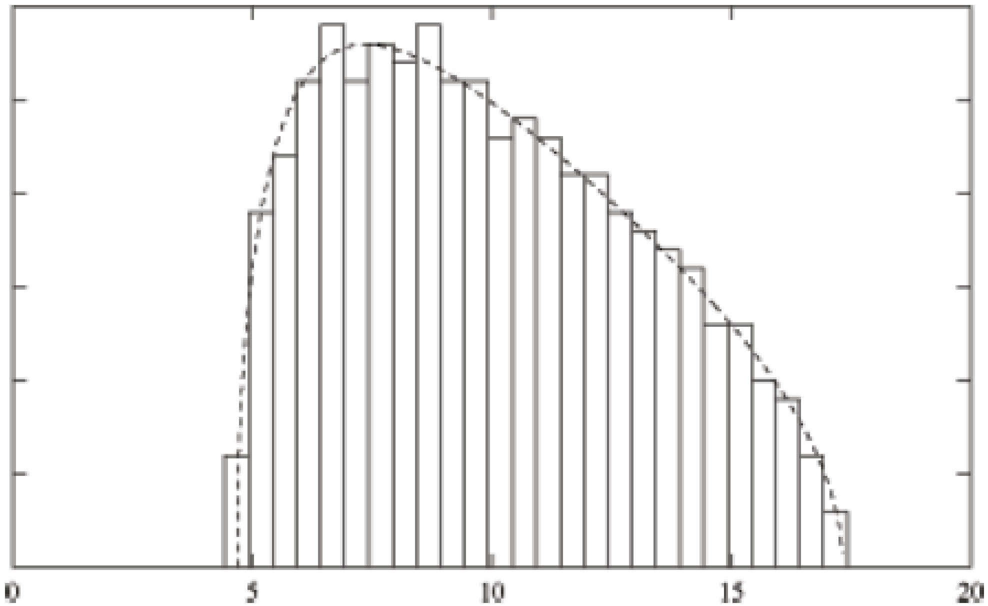
Eigenvalues of the matrix

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

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!

Empirical Eigenvalue Distribution

Denote the eigenvalues of $\mathbf{W}_N$ as $\lambda_1, \ldots, \lambda_N$.

Definition. The empirical eigenvalue distribution function of $\mathbf{W}_N$ is defined as the function that, for each x, gives the proportion of eigenvalues lower than or equal to x, i.e

$$F^N(x) = \frac{1}{N}\sharp\{n \leq N, \lambda_n \leq x\}$$

where $\sharp\{\}$ stands for the cardinality of the set.

Remark The eigenvalues of $\mathbf{W}_N$ are random so is their empirical distribution function $F^N(x)$.

Cauchy-Stieltjes Transform

N. I. Akhiezer, "the classical moment problem", Oliveir & Boyd, 1965.

The Cauchy-Stieltjes transform $G_\nu(z)$ associated to the probability measure ν is

$$G_\nu(z) = \int \frac{1}{t-z}\nu(dt)$$

Analytic in the half complex plan $\mathrm{Im}(z) > 0$.

A one to one mapping between ν and $G_\nu(z)$.

$$\int_a^b d\nu = \frac{1}{\pi} \lim_{y \to 0^+} \int_a^b \mathrm{Im} G_\nu(x+iy)$$

with a, b two continuity point of ν.

Cauchy-Stieltjes Transform

Some intuitions. Let ν be compactly supported. Then $G_\nu(z)$ is analytic in a neighborhood of ∞. Since $(z-t)^{-1} = \sum_{k=0}^{\infty} t^k z^{-k-1}$, then $G_\nu(z)$ has the following expansion at $z = \infty$:

$$-G_\nu(z) = z^{-1} + \sum_{k=1}^{\infty} m_k(\nu) z^{-k-1}.$$

where $m_k(\nu) = \int t^k d\nu(t) (k \in \mathbb{Z}^+)$

Cauchy-Stieltjes Transform and Matrices

Definition. The resolvent matrix of a hermitian matrix $\mathbf{W}_N$ is defined as:

$$\mathbf{R}_w(z) = (\mathbf{W}_N - z\mathbf{I})^{-1}$$

and therefore:

$$\int \frac{dF^N(x)}{x-z} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} = \frac{1}{N}\mathrm{Trace}(\mathbf{R}_w(z))$$

since $dF^N(x) = \frac{1}{N}\sum_{i=1}^{N}\delta(x-\lambda_i)$.

Interestingly, finding the Stieltjes transform can be related to a problem of computation of the trace of the inverse of a matrix.

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 $K/N \to \alpha$,

$$C_N \to \int \log \left(1 + \frac{1}{\sigma^2} t \right) \mu(dt) \text{ a.s.}$$

$$\frac{dC_N}{d\frac{1}{\sigma^2}} \to \sigma^2 - \sigma^4 G_\mu(-\sigma^2) \text{ a.s.}$$

The **capacity** is strongly related to the **Cauchy-Stieltjes transform**.

Information transfer in MIMO flexible networks

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

Consider the random variable

$$R_N = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{W}\mathbf{W}^H\right) = \int \log\left(1 + \frac{1}{\sigma^2}\lambda\right) 1]N \sum_i \delta(\lambda - \lambda_i) d\lambda$$

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

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

with

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

MMSE Receiver

Model example :

$$\begin{aligned}\mathbf{y} &= \mathbf{W}\mathbf{s} + \mathbf{n} \\ &= \mathbf{u}s_1 + \mathbf{U}\mathbf{x} + \mathbf{n} \\ &= \mathbf{u}s_1 + \mathbf{n}'\end{aligned}$$

$$\mathbb{E}(\mathbf{n}'\mathbf{n}'^H) = (\mathbf{U}\mathbf{U}^H + \sigma^2\mathbf{I}) = \mathbf{Q}\Lambda\mathbf{Q}^H$$

Whitening filter:

$$\begin{aligned}\tilde{\mathbf{y}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H\mathbf{y} &= \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H\mathbf{u}s_1 + \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H\mathbf{n}' \\ &= \mathbf{g}s_1 + \mathbf{b}\end{aligned}$$

$\mathbf{b}$ is a white Gaussian noise.

MMSE Receiver

$$\tilde{\mathbf{y}} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H\mathbf{u}s_1 + \mathbf{b}$$

Define $\mathbf{g} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H\mathbf{u}$

The output SINR is maximized with:

$$\mathbf{g}^H\tilde{\mathbf{y}} = \mathbf{g}^H\mathbf{g}s_1 + \mathbf{g}^H\mathbf{b}$$

As a consequence, the receiver is:

$$\mathbf{g}^H\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H = \mathbf{u}^H\left(\mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^H\right) = \mathbf{u}^H\left(\mathbf{U}\mathbf{U}^H + \sigma^2\mathbf{I}_N\right)^{-1}$$

Remark: The usual MMSE receiver is the unbiased one:

$$\mathbf{u}^H\left(\mathbf{W}\mathbf{W}^H + \sigma^2\mathbf{I}_N\right)^{-1} = \frac{1}{1 + \mathbf{u}^H\left(\mathbf{U}\mathbf{W}\mathbf{U}^H + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{u}}\mathbf{u}^H\left(\mathbf{U}\mathbf{U}^H + \sigma^2\mathbf{I}_N\right)^{-1}$$

MMSE Receiver

After MMSE filtering, we obtain:

$$\mathbf{g}^H \tilde{\mathbf{y}} = \mathbf{g}^H \mathbf{g} s_1 + \mathbf{g}^H \mathbf{b}$$

with $\mathbf{g} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Q}^H \mathbf{u}$

Signal to Interference plus Noise Ratio (SINR):

$$\beta_N = \frac{(\mathbf{g}^H \mathbf{g})^2 \mathbb{E}(\mid s_1 \mid^2)}{\mathbf{g}^H \mathbf{g}} = \mathbf{g}^H \mathbf{g} = \mathbf{u}^H \left(\mathbf{U}\mathbf{U}^H + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{u}$$

Depends strongly on the choice of $\mathbf{U}$.

Example: the i.i.d. model case

$$\beta_N = \mathbf{u}^H \left(\mathbf{U}\mathbf{U}^H + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{u}$$

Suppose the matrix $\mathbf{W} = [W_{ij}]$ has i.i.d. elements, $E[W_{ij}] = 0$, $E\left[W_{ij}^2\right] = 1/N$.
Example : IS95.

You remember that important lemma?
$\mathbf{u}$ vector $N \times 1$ with i.i.d elements. Each element : zero mean and variance $1/N$.
$\mathbf{A}$ matrix $N \times N$ independent of $\mathbf{u}$. Then, under some assumptions,

$$\mathbf{u}^H \mathbf{A} \mathbf{u} - \frac{1}{N} \mathrm{trace}\left(\mathbf{A}\right) \rightarrow 0 \text{ a.s.}$$

when $N \rightarrow \infty$.

Application : $\mathbf{u}$ et $\mathbf{U}$ independent, so

$$\beta_N - \frac{1}{N} \mathrm{trace}\left(\left(\mathbf{U}\mathbf{U}^H + \sigma^2 \mathbf{I}_N \right)^{-1} \right) \rightarrow 0 \text{ a.s.}$$

Example: the i.i.d. model case

$$\frac{1}{N}\mathrm{trace}\left(\left(\mathbf{U}\mathbf{U}^H + \sigma^2\mathbf{I}_N\right)^{-1}\right) = \frac{1}{N}\mathrm{trace}\left(f\left(\mathbf{U}\mathbf{U}^H\right)\right)$$

where $f(t) = 1/(t+\sigma^2)$.

Since $\mathbf{U}\mathbf{U}^H$ has a limiting Marchenko-Pastur distribution law μ, we have

$$\beta_N \to \beta = \int \frac{1}{t+\sigma^2}\mu(dt) = G_\mu(-\sigma^2) \text{ a.s.}$$

Solution :

$$\beta = \frac{1-\alpha}{2\sigma^2} - \frac{1}{2} + \sqrt{\frac{(1-\alpha)^2}{4\sigma^4} + \frac{(1+\alpha)}{2\sigma^2} + \frac{1}{4}}$$

The **SINR** at the output of the MMSE receiver is exactly the **Cauchy-Stieltjes transform**.

MMSE and Capacity

$$\mathrm{SINR}_{\mathrm{MMSE}} = G_\mu(-\sigma^2)$$
$$\frac{dC}{d\frac{1}{\sigma^2}} = \sigma^2 - \sigma^4 G_\mu(-\sigma^2)$$

$$\frac{dC}{\sigma^2} = \mathrm{SINR}_{\mathrm{MMSE}} - \frac{1}{\sigma^2}$$

The derivative of the capacity is strongly related to the performance of the MMSE receiver!

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 $\mathbf{A}$ and $\mathbf{B}$ to find the eigenvectors of $\mathbf{A} + \mathbf{B}$.

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 $\mathbf{B}$.

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

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 $\mathbf{C} = \mathbf{A}\mathbf{B}$.

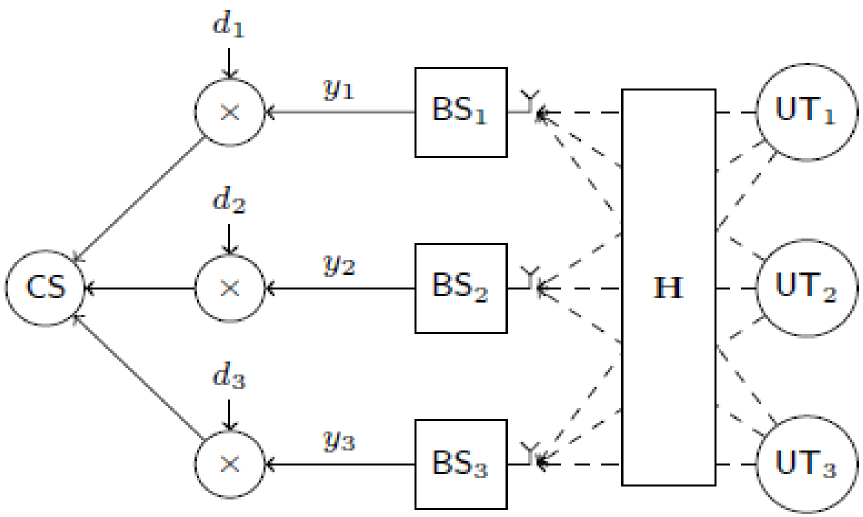

RMT
inside

Applications

Optimal channel training

Optimal Number of Base Stations

Network MIMO Uplink Channel



- B BSs and B UTs, each equipped with a single antenna.
- The BSs forward their received signals to the CS via high capacity backhaul links.
- The backhaul links fail randomly with probability ϵ (modeled by d_i).
- The CS detects which links are defective, estimates the channel $\mathbf{H}$ and *jointly* processes all received signals.

Motivation

The potential gain of network MIMO increases with B.

$$\Updownarrow$$

In practice, the CS needs to estimate the uplink channels during a fixed coherence time T. This overhead becomes quickly paramount when B grows.

1. What is the optimal fraction of the coherence time T used for channel training?

2. How many BSs should cooperate to maximize the per-cell rate?

Related work

- Practical limitations of network MIMO arise from many factors: limited backhaul capacity, local connectivity, processing complexity and delay, imperfect/local CSI [Shamai et al.'07, Shamai et al.'08, Gesbert et al.'10].
- Information theoretic treatment of a source communicating via randomly failing relays to a destination [Simeone et al.'09].
- Optimal 'training-data'-tradeoff has been studied for point-to-point MIMO [Hassibi-Hochwald'03] and multi-user downlink [Kobayashi-Jindal-Caire'09].

System Model

$$\boxed{\mathbf{y} = \mathbf{D}\underbrace{(\mathbf{H}\mathbf{x} + \mathbf{n}_{\mathsf{BS}})}_{\text{forwarded from BSs}} + \mathbf{n}_{\mathsf{CS}}}$$

where

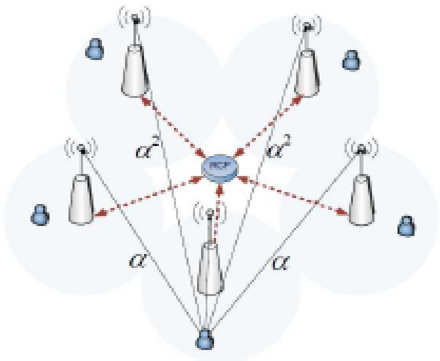
$$\mathbf{D} = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_B \end{pmatrix} \quad , \quad d_i \sim Bernoulli(1-\epsilon)$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_B \end{pmatrix} \quad , \quad \mathbb{E}\left[|x_i|^2\right] \le P$$

$$\mathbf{n}_{\mathsf{BS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2_{\mathsf{BS}}\mathbf{I}_B)$$

$$\mathbf{n}_{\mathsf{CS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2_{\mathsf{CS}}\mathbf{I}_B) \quad , \quad \sigma^2_{\mathsf{BS}} + \sigma^2_{\mathsf{CS}} = 1.$$

Path loss: Extended Wyner Model



$$\mathbf{V} = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^2 & \alpha \\ \alpha & 1 & \alpha & \alpha^2 & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha^2 & \alpha & 1 & \alpha \\ \alpha & \alpha^2 & \alpha^2 & \alpha & 1 \end{pmatrix}, \quad \alpha \in [0,1]$$

- The entries of the channel matrix $\mathbf{H}$ are modeled as independent, complex Gaussians $h_{bk} \sim \mathcal{N}_{\mathbb{C}}(0, v_{bk})$.
- The variance v_{bk} captures the path loss between UT k and BS b.
- Due to the circular symmetry, the sums of the entries of every row and every column of $\mathbf{V}$ are identical, i.e.,

$$\mathcal{K} \triangleq \sum_{i=1}^{B} v_{ij} = \sum_{j=1}^{B} v_{ij}, \qquad \forall i, j.$$

Channel Estimation

coherence time T

training τ	data transmission

- The CS estimates the channel from UT k to BS b based on

$$r_{bk} = \sqrt{\tau P} h_{bk} + n_{bk}, \qquad \text{if } d_b = 1$$

where $n_{bk} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

- Taking the MMSE estimate yields

$$h_{bk} = \hat{h}_{bk} + \tilde{h}_{bk}$$

with variances

$$\hat{v}_{bk}(\tau) \triangleq \mathbb{E}\left[|\hat{h}_{bk}|^2\right] = \frac{\tau P (v_{bk})^2}{\tau P v_{bk} + 1}$$
$$\tilde{v}_{bk}(\tau) \triangleq \mathbb{E}\left[|\tilde{h}_{bk}|^2\right] = \frac{v_{bk}}{\tau P v_{bk} + 1}.$$

Net Ergodic Achievable Rate

- The CS decomposes the received signal vector into

$$\mathbf{y} = \mathbf{D}\hat{\mathbf{H}}\mathbf{x} + \underbrace{\mathbf{D}\left(\tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}_{\mathsf{BS}}\right) + \mathbf{n}_{\mathsf{CS}}}_{\text{effective noise}}$$

- The *ergodic achievable rate* per cell is given by

$$\boxed{R(\tau) \triangleq \frac{1}{B}\mathbb{E}_{\hat{\mathbf{H}},\mathbf{D}}\left[\log\left|\mathbf{I}_B + \frac{\hat{\rho}(\tau)}{B}\mathbf{D}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}\right|\right]}$$

where

$$\hat{\rho}(\tau) \triangleq \frac{PB}{\tilde{\mathcal{K}}(\tau)P + 1}, \qquad \tilde{\mathcal{K}}(\tau) \triangleq \sum_{j=1}^{B} \tilde{v}_{1j}(\tau).$$

- Our goal is to maximize the *net ergodic achievable rate*

$$\boxed{R_{\mathsf{net}}(\tau) \triangleq \left(1 - \frac{\tau}{T}\right) R(\tau)}$$

Deterministic Equivalent

Theorem 1 (Hachem-Loubaton-Najim'07, Theorem 4) *For any fixed SNR ρ, fixed diagonal matrix $\mathbf{D}$ and variance profile $\hat{\mathbf{V}}$, the ergodic achievable rate converges almost surely for $B \to \infty$ to*

$$\overline{R}(\rho, \mathbf{D}) \triangleq \frac{1}{B} \sum_{i=1}^{B} \log \left(\frac{\rho^2}{\Psi_i(\rho)\Upsilon_i(\rho)} \right) - \frac{\mathbf{\Psi}(\rho)^\mathsf{T} \mathbf{D} \hat{\mathbf{V}} \mathbf{\Upsilon}(\rho)}{B^2 \rho}$$

where the vectors $\mathbf{\Psi}(\rho)$ and $\mathbf{\Upsilon}(\rho)$ are the unique solutions to the set of $2B$ fixed point equations

$$\Psi_i(\rho) = \frac{\rho}{1 + \frac{1}{B} \sum_{j=1}^{B} d_i \hat{v}_{ij} \Upsilon_j(\rho)}, \quad 1 \leq i \leq B$$

$$\Upsilon_j(\rho) = \frac{\rho}{1 + \frac{1}{B} \sum_{i=1}^{B} d_i \hat{v}_{ij} \Psi_i(\rho)}, \quad 1 \leq j \leq B.$$

Deterministic Equivalent (II)

Corollary 1 *For any fixed SNR ρ, variance profile $\hat{\mathbf{V}}$ and backhaul link failure probability ϵ, the ergodic achievable rate converges almost surely for $M \to \infty$ to*

$$R^{\text{det}}(\rho) \triangleq \mathbb{E}_{\mathbf{D}}\left[\overline{R}(\rho, \mathbf{D})\right] = \sum_{n=1}^{B} \frac{f(n,\epsilon)}{N_n} \sum_{i=1}^{N_n} \overline{R}(\rho, \mathbf{D}_i^n)$$

where $N_n = \binom{B}{n}$ and $\mathcal{D}^n = \{\mathbf{D}_1^n, \ldots, \mathbf{D}_{N_n}^n\}$ is the set of all matrices $\mathbf{D}$ such that the number of '1s' on the diagonal is equal to n and

$$f(n,\epsilon) = \binom{B}{n}(1-\epsilon)^n \epsilon^{B-n}, \qquad 1 \leq n \leq B.$$

The computation of the expectation becomes quickly prohibitive for large B since one needs to consider all 2^B possible realizations of $\mathbf{D}$.

Heuristic Approximation

Theorem 2 (Tulino-Verdú'05, Theorem 4) *For any fixed* SNR ρ, *backhaul link failure probability* $\epsilon = 0$ *and variance profile* $\hat{\mathbf{V}}$, *the ergodic achievable rate converges almost surely to*

$$\overline{R}_0\left(\frac{\hat{\mathcal{K}}}{B}\rho, 1\right)$$

where

$$\overline{R}_0(\rho, /\beta) \triangleq \log(1+\rho\beta - \mathcal{F}(\rho,\beta)) + \beta\log(1+\rho-\mathcal{F}(\rho,\beta)) - \mathcal{F}(\rho,\beta)/\rho$$

$$\hat{\mathcal{K}} \triangleq \sum_{j=1}^{B} \hat{v}_{1j}, \quad \mathcal{F}(x,y) \triangleq \frac{1}{4}\left(\sqrt{1+x(1+\sqrt{y})^2} - \sqrt{1+x(1-\sqrt{y})^2}\right)^2.$$

For $\epsilon \neq 0$, the variance profile is not symmetric any more. By the LLN, $(1-\epsilon)$ elements of $\mathbf{D}$ are equal to one. This motivates the heuristic approximation

$$\boxed{R^{\text{heur}}(\rho) \triangleq \overline{R}_0\left(\frac{\hat{\mathcal{K}}}{B}\rho, 1-\epsilon\right)}$$

Optimization of the Training Length

$$\begin{aligned} \text{maximize} \quad & \left(1 - \frac{\tau}{T}\right) R(\tau) \\ \text{subject to} \quad & B \leq \tau \leq T \end{aligned}$$

- Assuming that the objective function is concave in τ, we look for τ^* satisfying

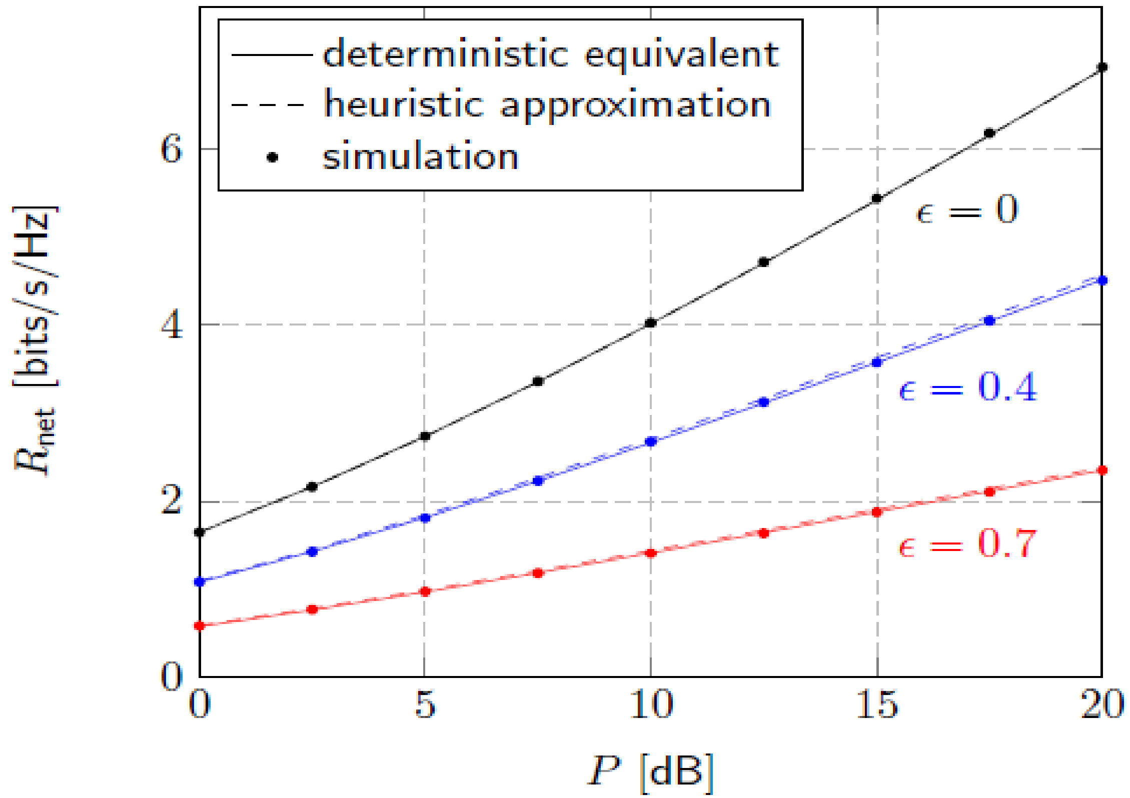
$$\left(1 - \frac{\tau^*}{T}\right) R'(\tau^*) - \frac{1}{T} R(\tau^*) = 0$$

- Using the following approximations

$$R(\tau) \approx \begin{cases} R^{\text{det}}((\hat{\rho}(\tau)) \\ R^{\text{heur}}((\hat{\rho}(\tau)) \end{cases} , \quad R'(\tau) \approx \begin{cases} R^{\text{det}\prime}((\hat{\rho}(\tau)) \\ R^{\text{heur}\prime}((\hat{\rho}(\tau)) \end{cases}$$

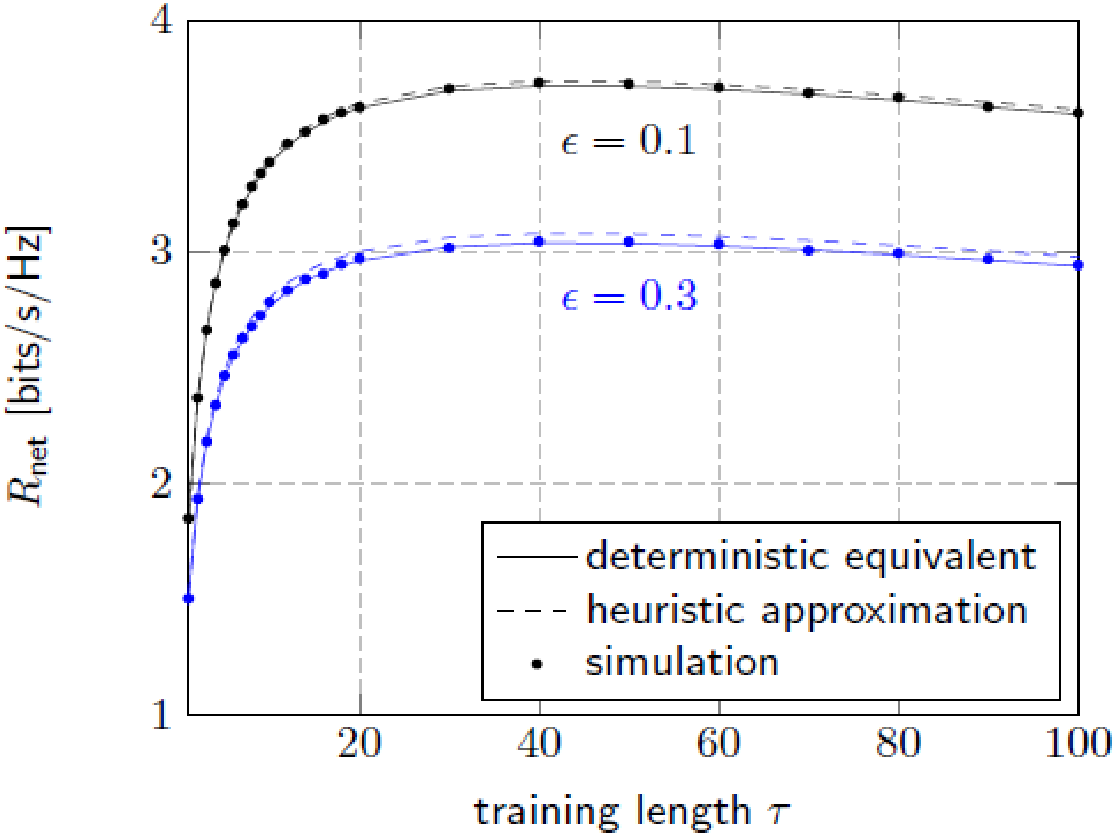
τ^* can be found by a simple line search.

Numerical Results (I): Tightness of Approximations



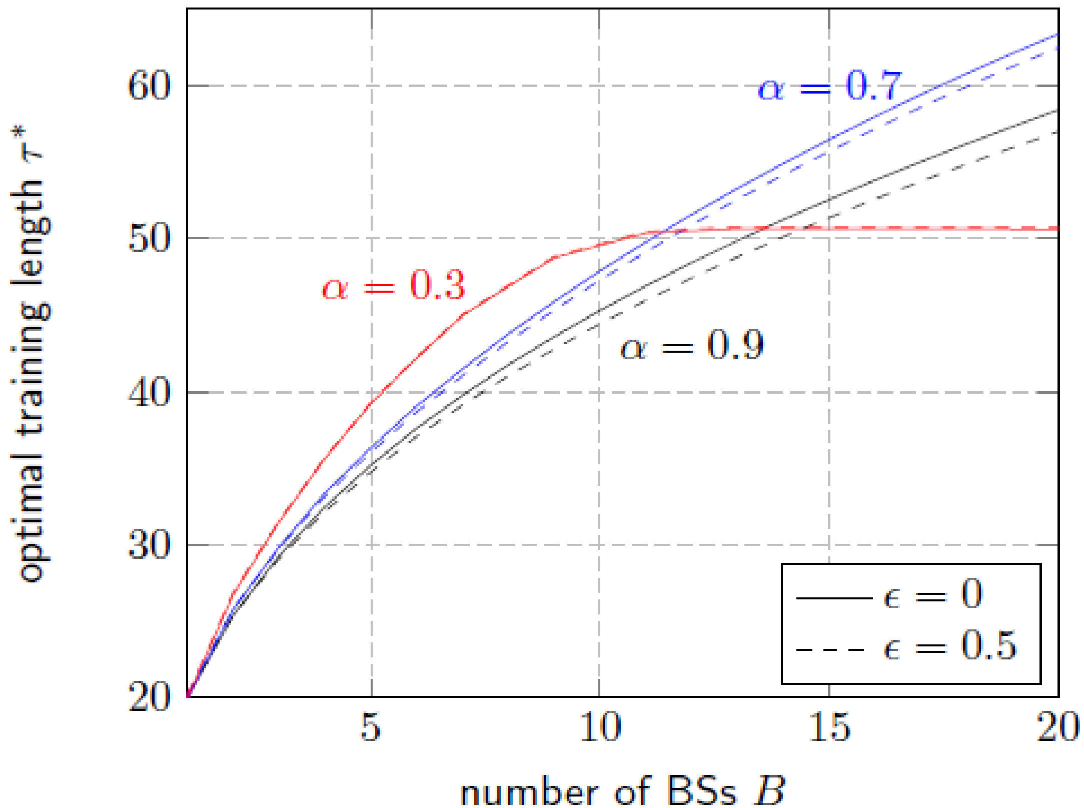
$$T = 1000, \quad \tau = 40, \quad M = 8, \quad \alpha = 0.7$$

Numerical Results (II): Concavity of the Objective Function



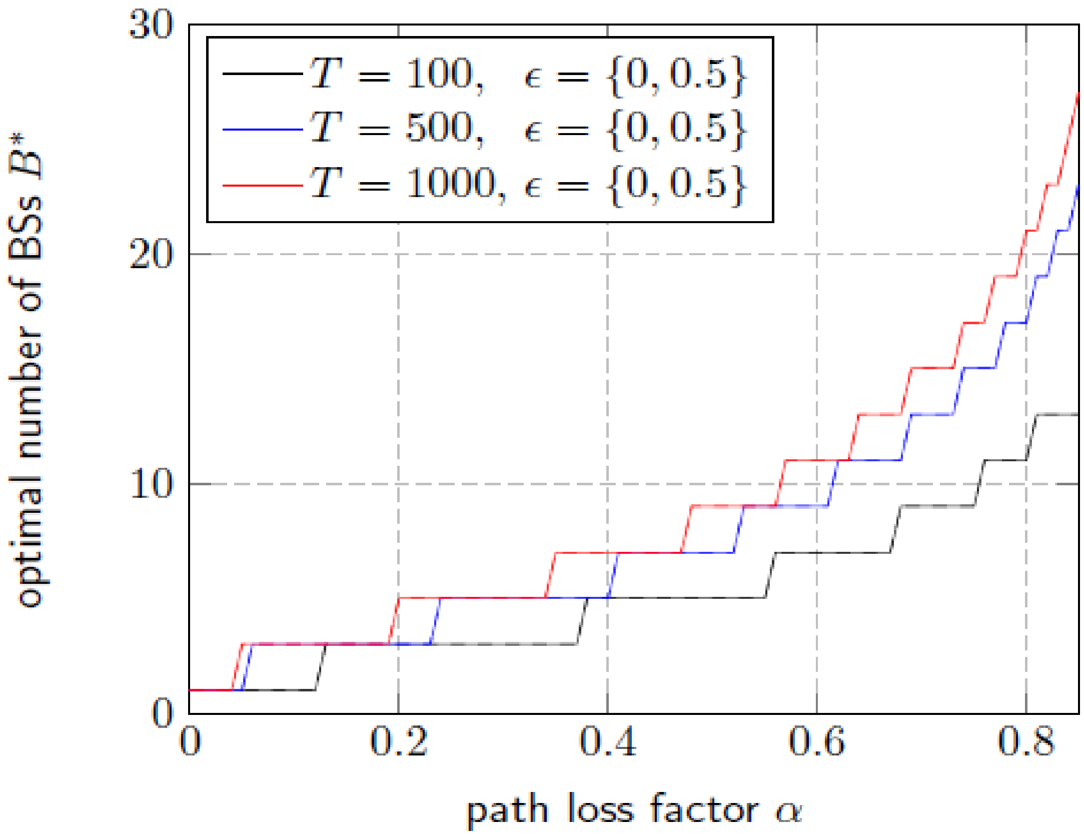
$$T = 1000, \quad P = 10\,\text{dB}, \quad M = 8, \quad \alpha = 0.7$$

Numerical Results (III): Optimal Training Length



$T = 1000, \quad P = 10\,\mathrm{dB}$

Numerical Results (IV): Optimal Number of Cooperative BSs



$\boxed{P = 10\,\text{dB}}$

Answers?

1. What is the optimal fraction of the coherence time T used for channel training?
   In general, non-trivial dependence of the optimum training length on the path loss, number of BSs, transmit powers, etc. For reasonably small networks:

$$\begin{array}{rcl} \#\text{BS } B \uparrow & \Longrightarrow & \tau^* \uparrow \\ \text{path loss factor } \alpha \downarrow & \Longrightarrow & \tau^* \uparrow \\ \text{link failure probability } \epsilon \uparrow & \Longrightarrow & \tau^* \approx \text{ const.} \end{array}$$

2. How many BSs should cooperate to maximize the per-cell rate?

$$\begin{array}{lcl} T = 1000, \quad \alpha = 0.9 & \Longrightarrow & B^* \approx 30 \\ T = 1000, \quad \alpha = 0.1 & \Longrightarrow & B^* \approx 3 \end{array}$$

Polynomial expansion detectors

Problem setting and motivation

Consider the $N \times K$ MIMO channel

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \sigma\mathbf{n} = \sum_{k=1}^{K} \mathbf{h}_k x_k + \sigma\mathbf{n}$$

where $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{H} \in \mathbb{C}^{N \times K}$ random but known to the receiver.

Several linear detectors require the matrix inversion ($\alpha \geq 0$):

$$\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \alpha\mathbf{I}_N\right)^{-1}$$

- This operation is expensive (complexity and energy) : $\mathcal{O}(N^2)$ [40].
- In particular, for large (distributed) antenna arrays for multi-user communications.
- $N \approx 100$, $K \approx 50$ (or even more)

$\boxed{\text{Asymptotic moments can be used calculate approximations of the matrix inverse.}}$

A note on matrix inversion

Let $\mathbf{A}$ be a non-singular $N \times N$ matrix with characteristic polynomial:

$$\det(z\mathbf{I}_N - \mathbf{A}) = \prod_{i=1}^{N}(z - \lambda_i) = \sum_{i=0}^{N}\alpha_i z^i$$

where λ_i are the eigenvalues of $\mathbf{A}$ and α_i depend on the eigenvalues of $\mathbf{A}$.

Caley-Hamilton Theorem ("Every matrix satisfies its own characteristic polynomial"):

$$\sum_{i=0}^{N}\alpha_i \mathbf{A}^i = \mathbf{0} \qquad \Longleftrightarrow \qquad \mathbf{A}^{-1} = \sum_{l=0}^{N-1} -\frac{\alpha_{l+1}}{\alpha_0}\mathbf{A}^l.$$

Polynomial expansion detectors

Assume we want to estimate $\mathbf{x}$ with a linear detector:

$$\hat{\mathbf{x}} = \mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \alpha \mathbf{I}_N\right)^{-1}\mathbf{y} \approx \mathbf{H}^{\mathrm{H}}\sum_{l=0}^{L-1} w_l \left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{l}\mathbf{y}, \quad L < \mathrm{rank}(\mathbf{H}).$$

The weights $\mathbf{w} = [w_0, \ldots, w_{L-1}]^{\mathrm{T}}$ can be chosen to minimize:

$$\mathbf{w} = \arg\min_{\mathbf{w}} \mathbb{E}\left[\left\|\mathbf{x} - \mathbf{H}^{\mathrm{H}}\left(\sum_{l=0}^{L-1} w_l \left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{l}\right)\mathbf{y}\right\|_2^2\right] = \mathbf{\Phi}^{-1}\varphi$$

where $\mathbf{\Phi} \in \mathbb{R}_+^{L\times L}$ and $\varphi \in \mathbb{R}_+^{L}$ are defined as

$$[\mathbf{\Phi}]_{ij} = \frac{1}{N}\mathrm{tr}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{i+j} + \sigma^2\frac{1}{N}\mathrm{tr}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{i+j-1}$$

$$[\varphi]_i = \frac{1}{N}\mathrm{tr}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{i}.$$

Idea: Replace the moments $\frac{1}{N}\mathrm{tr}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^k$ by their deterministic approximations $\overline{M}_k$.

Related works and existing moment results

Asymptotic moment results:

[12] : $\mathbf{H}$ as i.i.d. entries with zero mean and variance $1/K$.

[41] : $\mathbf{H}_{ij}$ have zero mean and variance $\frac{1}{K}\mathbf{V}_{ij}$.

[41] : $\mathbf{H} = [\mathbf{A}_1\mathbf{h}_1, \ldots, \mathbf{A}_K\mathbf{h}_K]\,\mathbf{B}$, where $\mathbf{h}_j$ are vectors of i.i.d. entries with zero mean and variance $1/K$, $\mathbf{B}$ is diagonal and $\mathbf{A}_j$ are absolutely summable Toeplitz matrices.

[42] : $\mathbf{H} = \mathbf{T}\mathbf{W}\mathbf{P}^{\frac{1}{2}}$, where $\mathbf{T}$ is Toeplitz, $\mathbf{P}$ is diagonal and $\mathbf{W}$ has either i.i.d. elements with zero mean and variance $1/K$ or is created by taking $K \leq N$ columns from a random Haar (unitary) matrix.

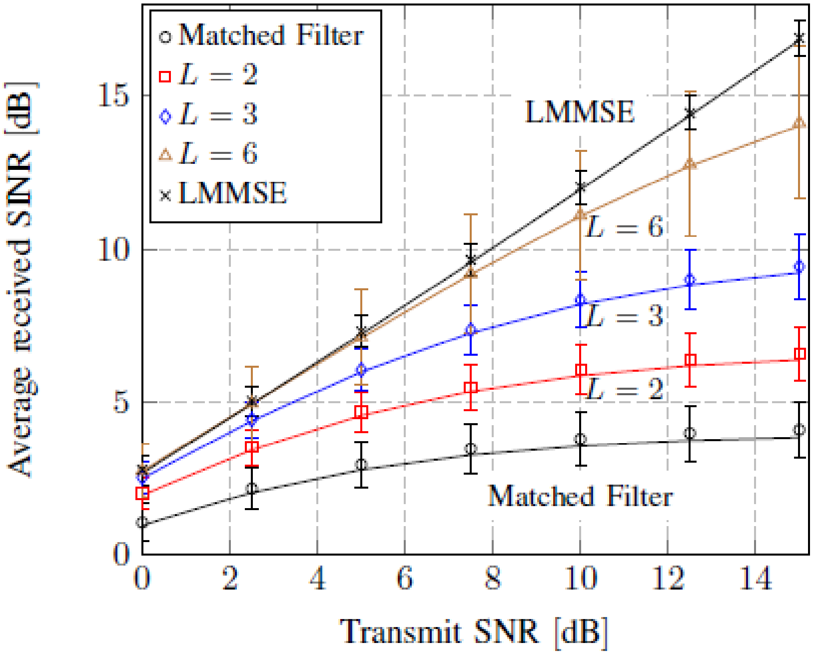
[43] : $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$ where $\mathbf{h}_k = \mathbf{R}_k^{\frac{1}{2}}\mathbf{w}_k$ and $\mathbf{w}_k$ are vectors of i.i.d. with zero mean and variance $1/K$.

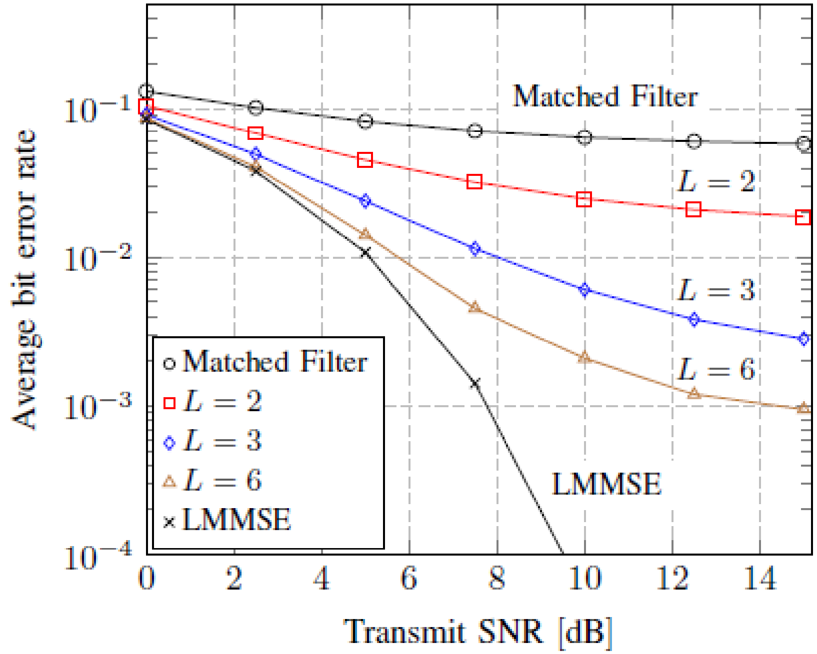
[44] : $\mathbf{H}$ is a random Vandermonde matrix.

Related works on polynomial expansion receivers:

[45, 46, 47, 48, 40, 49, 50, 43, 51]

Polynomial expansion receivers : Numerical results





- $N = 100$, $K = 40$
- Channel with a generalized variance profile: Correlation matrices $\mathbf{R}_k$ randomly created (extended version of Jake's model) (see [43] for details)
- Uncoded BER $= \mathbb{E}\left[Q\left(\sqrt{\mathsf{SINR}}\right)\right]$
- One can also derive a deterministic equivalent of the SINR for a given approximation order L, as shown by solid lines in the left plot (see, e.g., [40, 43]).

Some remarks

- Polynomial expansion detectors are a low complexity receiver architecture for large MIMO channels (also for CDMA).
- Polynomial expansion detectors allow to trade-off complexity against performance with a very fine granularity ($L = 1$: matched filter, $L = \mathrm{rank}(\mathbf{H})$: MMSE detector).
- If the channel coherence time is large and the channel statistics are known, the asymptotic moments can be precomputed and only the matrices $(\mathbf{H}\mathbf{H}^{\mathrm{H}})^l$, $l = 1, \ldots, L$, must be calculated.
- The asymptotic moments for a wide range of channel models are known and could be used.
- Practical implementations?

Massive MIMO

Where to start from?

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

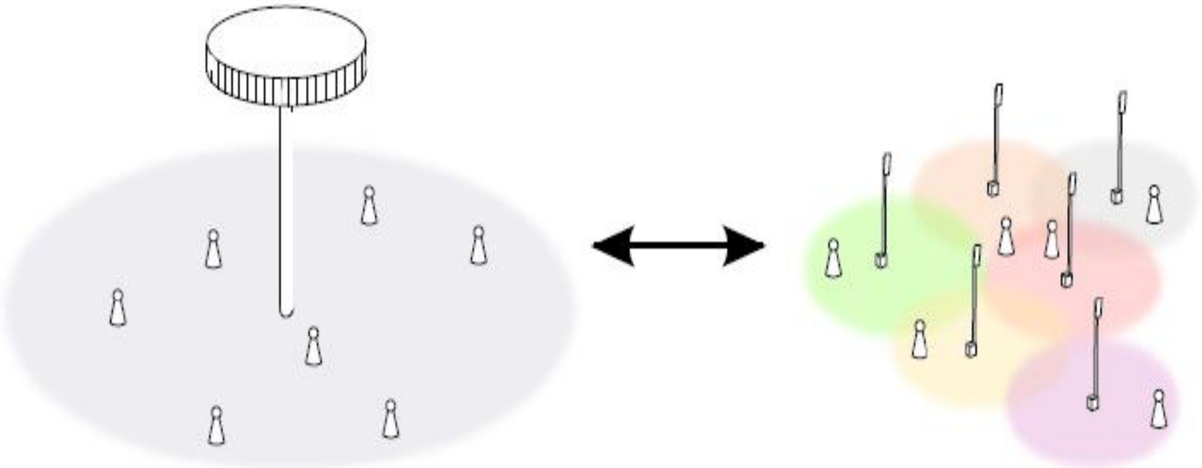
- Expectation: Many Metrics Should Be Improved in 5G
  - Higher user data rates
  - Higher area throughput
  - Great scalability in number of connected devices
  - Higher reliability and lower latency
  - Better coverage with more uniform user rates
  - Improved energy efficiency

- These are Conflicting Metrics!

- Higher user data rate

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

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:

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

- $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$: 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:

- $\mathcal{K} = \frac{\lambda_{\text{UT}}}{\lambda_{\text{BS}}}$ 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:

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

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
- $S = \left|\mathbf{h}_0^{\mathrm{H}} \mathbf{w}_{0,1}\right|^2 \sim \Gamma(N-K+1, 1)$, $\quad g_i = \sum_{k=1}^{K} \left|\mathbf{h}_i^{\mathrm{H}} \mathbf{w}_{i,k}\right|^2 \sim \Gamma(K, 1)$
- 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)

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

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

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

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

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

Fully distributing the antennas gives highest throughput gains!

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

- Distributed network densification is preferable over massive MIMO if the average throughput per UT should be increased.
- More antennas increase the coverage probability, but more BSs lead to a linear increase in area spectral efficiency (with constant total transmit power).
- If we use other metrics such as coverage probability or goodput, the picture might change.

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

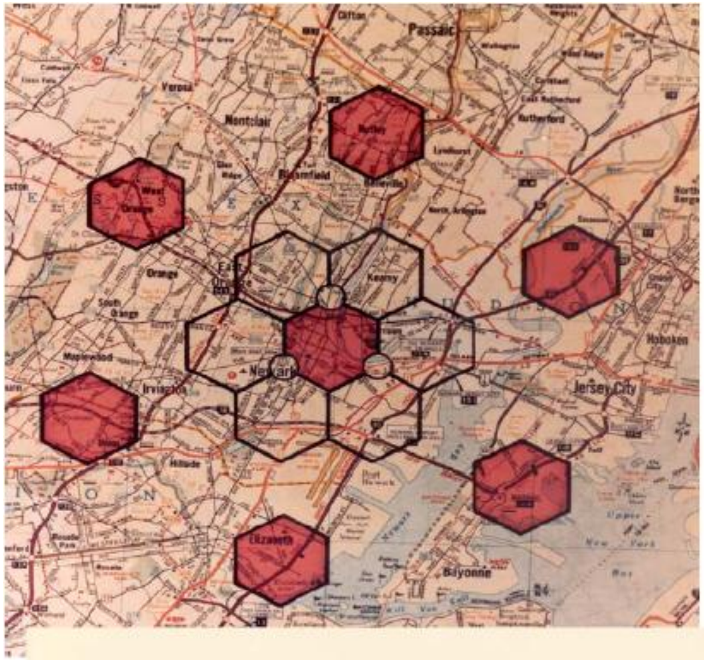
What if we are only interested in the average throughput per UT?

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

Bi-Bop (commercial launch in 1991-1997)

Let us know focus on two metrics…

- 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!
  - Difficult to maximize theoretically all metrics simultaneously
  - **Our goal: High energy efficiency (EE) with uniform user rates**

How to Measure Energy-Efficiency?

- Energy-Efficiency (EE) in bit/Joule

$$EE = \frac{\text{Average Sum Rate [bit/s/cell]}}{\text{Power Consumption [Joule/s/cell]}}$$

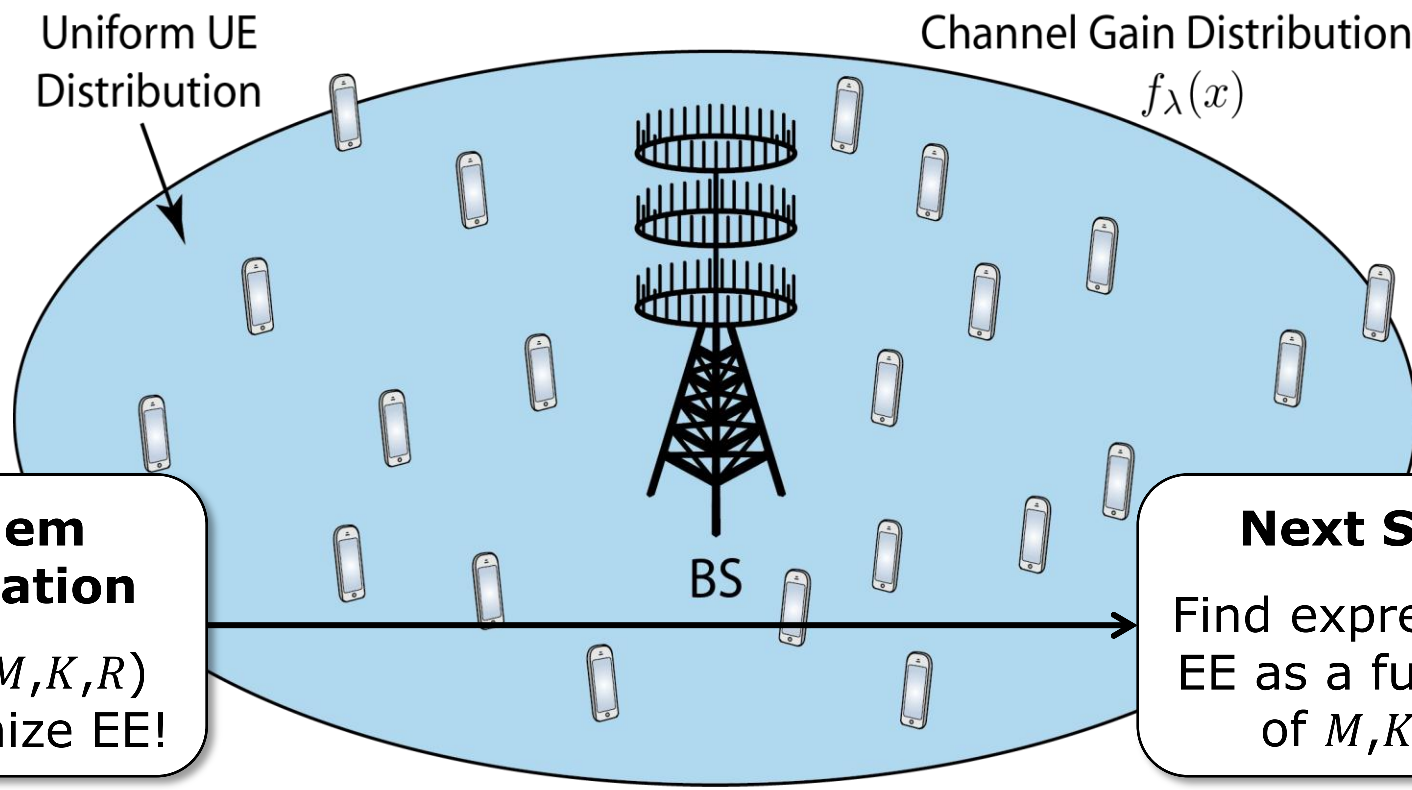
- Conventional Academic Approaches:
  - Maximize rates with fixed power
  - Minimize transmit power for fixed rates

New Problem: Balance rates and power consumption

Important to account for overhead signaling and circuit power!

Single-Cell: Optimizing for Energy-Efficiency

- Clean Slate Design
  - Single Cell: One base station (BS) with M antennas
  - Geometry: Random distribution for user locations and pathlosses
  - Multiple users: Pick K users randomly and serve with some rate R



Problem Formulation

Select (M,K,R) to maximize EE!

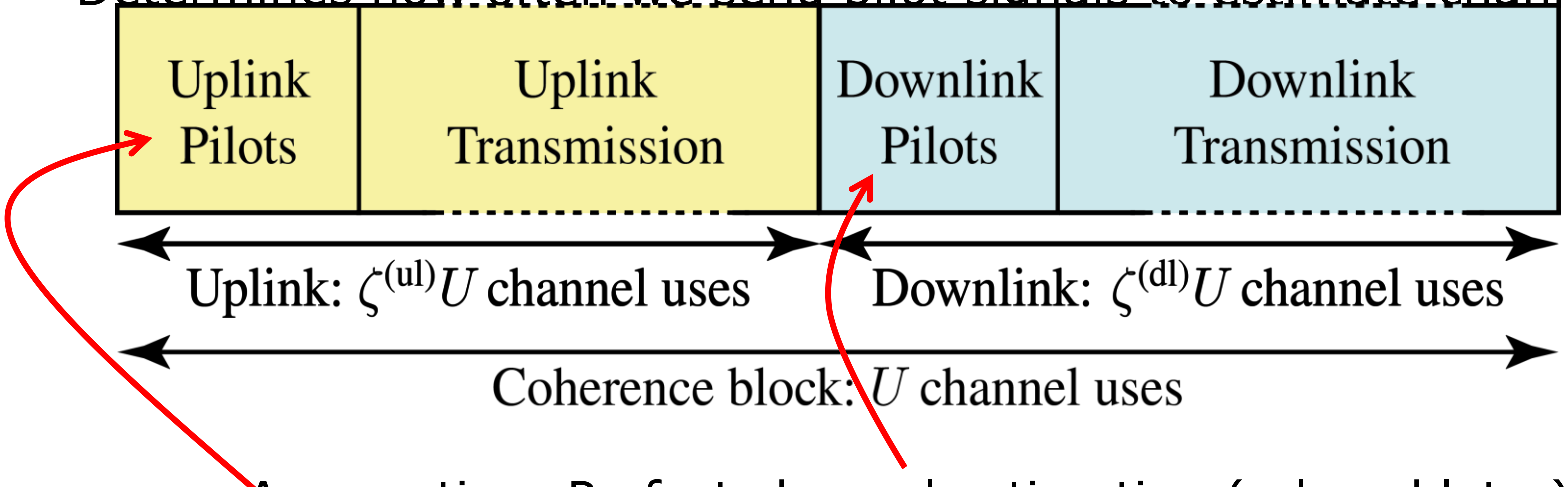
Next Step

Find expression: EE as a function of M,K,R.

System Model: Protocol

- Time-Division Duplex (TDD) Protocol
  - Uplink and downlink separated in time
  - Uplink fraction $\zeta^{(\mathrm{ul})}$ and downlink fraction $\zeta^{(\mathrm{dl})}$

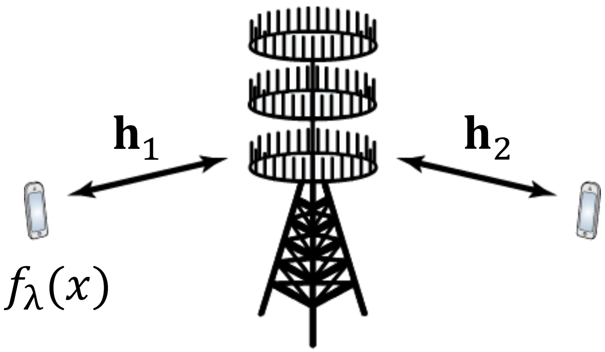
- Coherence Block
  - B Hz bandwidth = B "channel uses" per second (symbol time $1/B$)
  - Channel stays fixed for U channel uses (symbols) = Coherence block
  - Determines how often we send pilot signals to estimate channels



Assumption: Perfect channel estimation (relaxed later)

System Model: Channels

- Flat-Fading Channels
  - Channel between BS and User k: $\mathbf{h}_k \in \mathbb{C}^M$
  - Rayleigh fading: $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \lambda_k \mathbf{I})$
  - Channel variances λ_k: Random variables, pdf $f_\lambda(x)$



- Uplink Transmission
  - User k transmits signal s_k with power $\mathbb{E}\{|s_k|^2\} = \mathrm{p}_k^{(\mathrm{ul})}$ [Joule/channel use]
  - Received signal at BS:

Signals from other users (interference)

$$\boldsymbol{y} = \mathbf{h}_k s_k + \sum_{i=1, i \neq k}^{K} \mathbf{h}_i s_i + \mathbf{n}$$

Signal of User k

Noise $\sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$

  - Recover s_k by receive beamforming $\mathbf{g}_k$ as $\mathbf{g}_k^H \boldsymbol{y}$:

$$\mathrm{SINR}_k^{(\mathrm{ul})} = \frac{\mathbb{E}\left\{|s_k|^2 \left|\mathbf{g}_k^H \mathbf{h}_k\right|^2\right\}}{\sum_{i \neq k} \mathbb{E}\left\{|s_i|^2 \left|\mathbf{g}_k^H \mathbf{h}_i\right|^2\right\} + \mathbb{E}\left\{\left|\mathbf{g}_k^H \mathbf{n}\right|^2\right\}} = \frac{\mathrm{p}_k^{(\mathrm{ul})} |\mathbf{g}_k^H \mathbf{h}_k|^2}{\sum_{i \neq k} \mathrm{p}_i^{(\mathrm{ul})} |\mathbf{g}_k^H \mathbf{h}_i|^2 + \sigma^2 \|\mathbf{g}_k\|^2}$$

System Model: Channels (2)

- Flat-Fading Channels
  - Channel between BS and User k: $\mathbf{h}_k \in \mathbb{C}^M$
  - Rayleigh fading: $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \lambda_k \mathbf{I})$
  - Channel variances λ_k: Random variables, pdf $f_\lambda(x)$

- Downlink Transmission
  - BS transmits d_k to User k with power $\mathbb{E}\{|d_k|^2\} = \mathrm{p}_k^{(\mathrm{dl})}$ [Joule/channel use]
  - Spatial directivity by beamforming vector $\mathbf{v}_k$
  - Received signal at User k:

Signals from other users (interference)

$$y_k = \mathbf{h}_k^H \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|} d_k + \sum_{i=1,\, i\neq k}^{K} \mathbf{h}_k^H \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} d_i + n_k$$

Signal to User k

Noise $\sim \mathcal{CN}(0, \sigma^2)$

  - Recover d_k at User k:

$$\mathrm{SINR}_k^{(\mathrm{dl})} = \frac{\mathrm{p}_k^{(\mathrm{dl})} |\mathbf{h}_k^H \mathbf{v}_k|^2 / \|\mathbf{v}_k\|^2}{\sum_{i\neq k} \mathrm{p}_i^{(\mathrm{dl})} |\mathbf{h}_k^H \mathbf{v}_i|^2 / \|\mathbf{v}_i\|^2 + \sigma^2}$$

System Model: How Much Transmit Power?

- Design Parameter: *Gross* rate R
  - Make sure that $R = \begin{cases} B\log_2(1+\mathrm{SINR}_k^{(\mathrm{ul})}) & \text{for all } k \text{ in uplink} \\ B\log_2(1+\mathrm{SINR}_k^{(\mathrm{dl})}) & \text{for all } k \text{ in downlink} \end{cases}$
  - Select beamforming $\mathbf{g}_k$ and $\mathbf{v}_k$, adapt transmit power $\mathrm{p}_k^{(\mathrm{ul})}$ and $\mathrm{p}_k^{(\mathrm{dl})}$

  - Gives K Equations:

$$\begin{cases} \mathrm{p}_k^{(\mathrm{ul})}|\mathbf{g}_k^H\mathbf{h}_k|^2 = (2^{R/B}-1)\left(\sum_{i\neq k}\mathrm{p}_i^{(\mathrm{ul})}\left|\mathbf{g}_k^H\mathbf{h}_i\right|^2 + \sigma^2\|\mathbf{g}_k\|^2\right) & \text{for } k = 1,\dots,K \\ \mathrm{p}_k^{(\mathrm{dl})}\frac{\left|\mathbf{h}_k^H\mathbf{v}_k\right|^2}{\|\mathbf{v}_k\|^2} = (2^{R/B}-1)\left(\sum_{i\neq k}\mathrm{p}_i^{(\mathrm{dl})}\frac{\left|\mathbf{h}_k^H\mathbf{v}_i\right|^2}{\|\mathbf{v}_i\|^2} + \sigma^2\right) & \text{for } k = 1,\dots,K \end{cases}$$

  - Linear equations in transmit powers → Solve by Gaussian elimination!

Total Transmit Power [Joule/s] for $\mathbf{g}_k = \mathbf{v}_k$

Uplink energy/symbol: $\sigma^2\mathbf{D}^{-H}\mathbf{1}$

Downlink energy/symbol: $\sigma^2\mathbf{D}^{-1}\mathbf{1}$

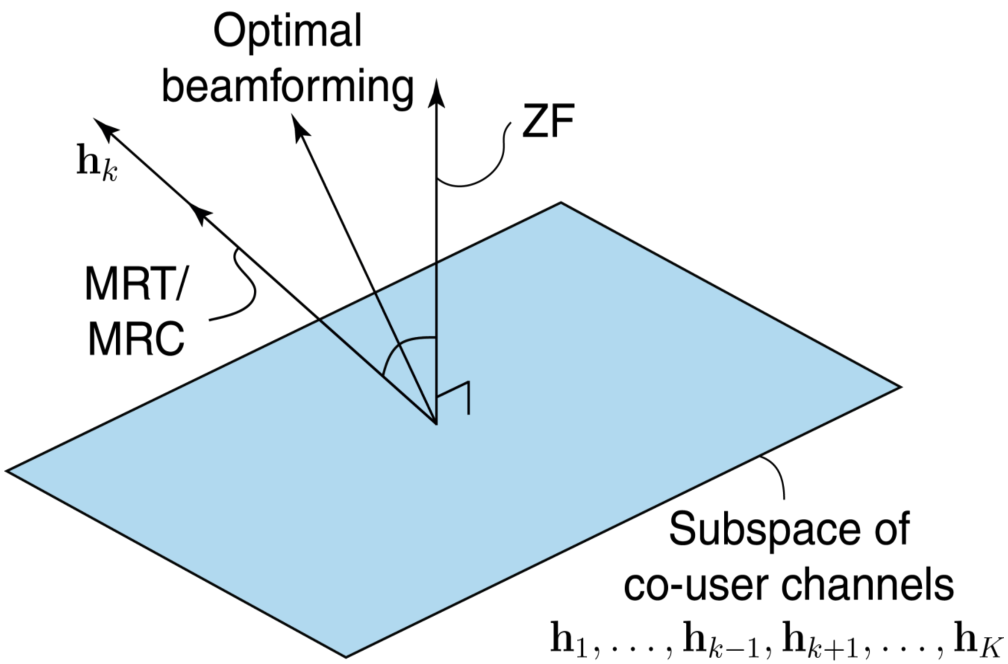
Same total power: $P_{\mathrm{trans}} = B\mathbb{E}\{\sigma^2\mathbf{1}^H\mathbf{D}^{-1}\mathbf{1}\} = B\mathbb{E}\{\sigma^2\mathbf{1}^H\mathbf{D}^{-H}\mathbf{1}\}$

where $[\mathbf{D}]_{k,l} = \begin{cases} \frac{\left|\mathbf{h}_k^H\mathbf{v}_k\right|^2}{(2^{R/B}-1)\|\mathbf{v}_k\|^2} & \text{for } k = l \\ -\frac{\left|\mathbf{h}_k^H\mathbf{v}_l\right|^2}{\|\mathbf{v}_l\|^2} & \text{for } k \neq l \end{cases}$

System Model: How Much Transmit Power? (2)

- What did we Derive?
  - Optimal power allocation for fixed beamforming vectors

- Different Beamforming
  - Notation: $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_K]$
    $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_K]$,
    $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]$,
    $\mathbf{P}^{(\mathrm{ul})} = \mathrm{diag}(\mathrm{p}_1^{(\mathrm{ul})}, \dots, \mathrm{p}_K^{(\mathrm{ul})})$



Minimize interference

Maximize signal

  - Maximum ratio trans./reception (MRT/MRC): $\mathbf{G} = \mathbf{V} = \mathbf{H}$
  - Zero-forcing (ZF) beamforming: $\mathbf{G} = \mathbf{V} = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$
  - Optimal beamforming: $\mathbf{G} = \mathbf{V} = \left(\sigma^2\mathbf{I} + \mathbf{H}\,\mathbf{P}^{(\mathrm{ul})}\mathbf{H}^H\right)^{-1}\mathbf{H}$

Balance signal and interference (iteratively!)

System Model: How Much Transmit Power? (3)

- Simplified Expressions for ZF ($M \geq K+1$)
  - Main property: $\mathbf{H}^H\mathbf{V} = \mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1} = \mathbf{I}$
  - Hence: $[\mathbf{D}]_{k,l} = \begin{cases} \frac{|\mathbf{h}_k^H\mathbf{v}_k|^2}{(2^{R/B}-1)\|\mathbf{v}_k\|^2} & \text{for } k = l \\ -\frac{|\mathbf{h}_k^H\mathbf{v}_l|^2}{\|\mathbf{v}_l\|^2} & \text{for } k \neq l \end{cases} = \begin{cases} \frac{1}{(2^{R/B}-1)\|\mathbf{v}_k\|^2} & \text{for } k = l \\ 0 & \text{for } k \neq l \end{cases}$

Property of Wishart matrices

  - Total transmit power:

$$P_{\text{trans}} = \mathbb{E}\{B\sigma^2\mathbf{1}^H\mathbf{D}^{-1}\mathbf{1}\} = B\sigma^2(2^{R/B}-1)\underbrace{\sum_k \mathbb{E}\{\|\mathbf{v}_k\|^2\}}_{\mathbb{E}\{\mathrm{tr}((\mathbf{H}^H\mathbf{H})^{-1})\}}$$

$$= B\sigma^2(2^{R/B}-1)\frac{K}{M-K}\underbrace{\left(\frac{1}{\lambda}\right)}_{\text{Call this } \mathcal{S}_\lambda}$$

Call this $\mathcal{S}_\lambda$ (depends on cell)

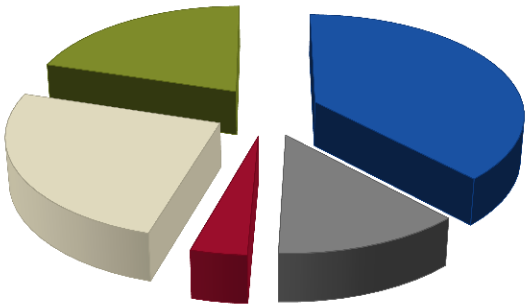
Summary: Transmit Power with ZF

Parameterize gross rate as $R = B\log_2(1+\alpha(M-K))$ for some α

Total transmit power: $P_{\text{trans}} = \alpha B\sigma^2\mathcal{S}_\lambda K$ [Joule/s]

Detailed Power Consumption Model

- What Consumes Power?
  - Not only radiated transmission power
  - Circuits, signal processing, backhaul, etc.
  - Must be specified as functions of M, K, R



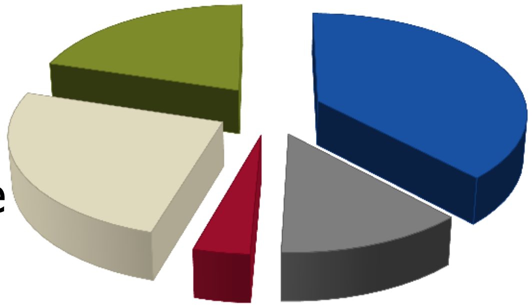
- Power Amplifiers
  - Amplifier efficiencies: $\eta^{(\mathrm{ul})}, \eta^{(\mathrm{dl})} \in (0,1]$
  - Average inefficiency: $\frac{\zeta^{(\mathrm{ul})}}{\eta^{(\mathrm{ul})}} + \frac{\zeta^{(\mathrm{dl})}}{\eta^{(\mathrm{dl})}} = \frac{1}{\eta}$

Summary: $\frac{P_{\mathrm{trans}}}{\eta}$

- Active Transceiver Chains
  - P_{FIX} = Fixed power (control signals, oscillator at BS, standby, etc.)
  - P_{BS} = Circuit power / BS antenna (converters, mixers, filters)
  - P_{UE} = Circuit power / user (oscillator, converters, mixer, filters)

Summary: $P_{\mathrm{FIX}} + M \cdot P_{\mathrm{BS}} + K \cdot P_{\mathrm{UE}}$

Detailed Power Consumption Model (2)



- Signal Processing
  - Channel estimation and beamforming
  - Efficiency: $L_{\mathrm{BS}}, L_{\mathrm{UE}}$ arithmetic operations / Joule

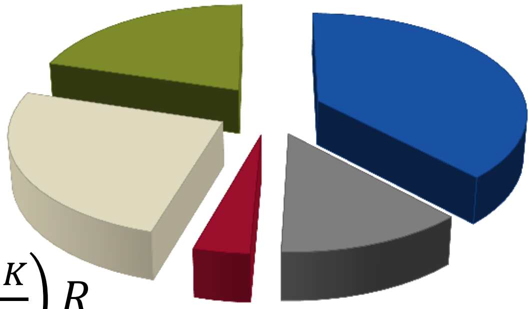
- Channel Estimation: $\frac{B}{U}\left(\frac{2\tau^{(\mathrm{ul})}MK^2}{L_{\mathrm{BS}}} + \frac{4\tau^{(\mathrm{dl})}K^2}{L_{\mathrm{UE}}}\right)$
  - Once in uplink/downlink per coherence block
  - Pilot signal lengths: $\tau^{(\mathrm{ul})}K, \tau^{(\mathrm{dl})}K$ for some $\tau^{(\mathrm{ul})}, \tau^{(\mathrm{dl})} \geq 1$

- Linear Processing (for $\mathbf{G} = \mathbf{V}$): $\frac{B}{U}\frac{C_{\mathrm{beamforming}}}{L_{\mathrm{BS}}} + B\left(1 - \frac{(\tau^{(\mathrm{ul})}+\tau^{(\mathrm{ul})})K}{U}\right)\frac{2MK}{L_{\mathrm{BS}}}$
  - Compute beamforming vector once per coherence block
  - Use beamforming for all $B\left(1 - \left(\tau^{(\mathrm{ul})} + \tau^{(\mathrm{ul})}\right)K/U\right)$ symbols
  - Types of beamforming:

$$C_{\mathrm{beamforming}} = \begin{cases} 3MK & \text{for MRT/MRC} \\ 3MK^2 + MK + \frac{1}{3}K^3 & \text{for ZF} \\ Q\left(3MK^2 + MK + \frac{1}{3}K^3\right) & \text{for Optimal} \end{cases}$$

Number of iterations

Detailed Power Consumption Model (3)



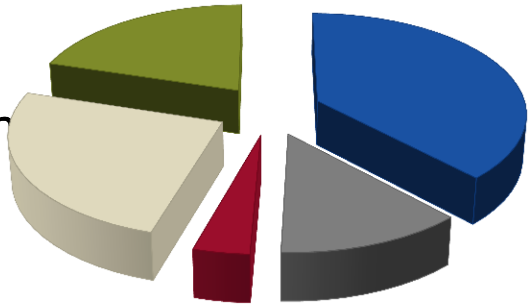
- Coding and Decoding: $R_{\mathrm{sum}}(P_{\mathrm{COD}} + P_{\mathrm{DEC}})$
  - P_{COD} = Energy for coding data / bit
  - P_{DEC} = Energy for decoding data / bit
  - Sum rate: $R_{\mathrm{sum}} = K\left(\zeta^{(\mathrm{ul})} - \frac{\tau^{(\mathrm{ul})}K}{U}\right)R + K\left(\zeta^{(\mathrm{dl})} - \frac{\tau^{(\mathrm{dl})}K}{U}\right)R$

$$= K\left(1 - \frac{(\tau^{(\mathrm{ul})} + \tau^{(\mathrm{dl})})K}{U}\right)R$$

- Backhaul Signaling: $P_{\mathrm{BH}} + R_{\mathrm{sum}}P_{\mathrm{BT}}$
  - P_{BH} = Load-independent backhaul power
  - P_{BT} = Energy for sending data over backhaul / bit

Detailed Power Consumption Model: Summary

- Many Things Consume Power
  - Parameter values (e.g., $P_{\mathrm{BS}}, P_{\mathrm{UE}}$) change over time
  - Structure is important for analysis



Generic Power Model

Fixed power

$$\frac{P_{\mathrm{trans}}}{\eta} + \underbrace{C_{0,0} + C_{0,1}M}_{\text{Circuit power per transceiver chain}} + \underbrace{C_{1,0}K + C_{1,1}MK + C_{2,0}K^2 + C_{3,0}K^3 + C_{2,1}MK^2}_{\text{Cost of signal processing}} + \underbrace{AK\left(1 - \frac{(\tau^{(\mathrm{ul})} + \tau^{(\mathrm{dl})})K}{U}\right)\bar{R}}_{\text{Coding/decoding/ backhaul}}$$

Transmit with amplifiers

for some parameters $C_{l,m}$ and A

- Observations
  - Polynomial in M and K → Increases faster than linear with K
  - Depends on cell geometry only through P_{trans}

Finally: Problem Formulation

- Maximize Energy-Efficiency:

Average Sum Rate [bit/s/cell]

$$\underset{M,K,R}{\text{maximize}} \quad \frac{K\left(1-\frac{(\tau^{(\mathrm{ul})}+\tau^{(\mathrm{dl})})K}{U}\right)R}{\underbrace{\frac{P_{\mathrm{trans}}}{\eta}+\sum_{i=0}^{3} C_{i,0}K^{i}+\sum_{i=0}^{2} C_{i,1}MK^{i}+AK\left(1-\frac{(\tau^{(\mathrm{ul})}+\tau^{(\mathrm{dl})})K}{U}\right)R}_{\text{Power Consumption [Joule/s/cell]}}}$$

Closed Form Expressions with ZF

Recall: $R = B\log_2(1+\alpha(M-K))$ for some α and $P_{\mathrm{trans}} = \alpha B\sigma^2\mathcal{S}_\lambda K$

Define: $\tau = \tau^{(\mathrm{ul})}+\tau^{(\mathrm{dl})}$

$$\underset{M,K,\alpha}{\text{maximize}} \quad \frac{K\left(1-\frac{\tau K}{U}\right)B\log_2(1+\alpha(M-K))}{\frac{\alpha B\sigma^2\mathcal{S}_\lambda K}{\eta}+\sum_{i=0}^{3} C_{i,0}K^{i}+\sum_{i=0}^{2} C_{i,1}MK^{i}+AK\left(1-\frac{\tau K}{U}\right)B\log_2(1+\alpha(M-K))}$$

Simple ZF expression: Used for analysis, other beamforming by simulation

Why Such a Detailed/Complicated Model?

- Simplified Model → Unreliable Optimization Results
  - Two examples based on ZF
  - Beware: Both has appeared in the literature!

- Example 1: Fixed circuit power and no coding/decoding/backhaul

$$\underset{M,K,\alpha}{\text{maximize}} \quad \frac{K\left(1-\frac{\tau K}{U}\right)B\log_2(1+\alpha(M-K))}{\frac{\alpha B\sigma^2\mathcal{S}_\lambda K}{\eta}+C_{0,0}}$$

  - If $M \to \infty$, then $\log_2(1+\alpha(M-K)) \to \infty$ and thus $\mathrm{EE} \to \infty$!

- Example 2: Ignore pilot overhead and signal processing

$$\underset{M,K,\alpha}{\text{maximize}} \quad \frac{KB\log_2(1+\alpha(M-K))}{\frac{\alpha B\sigma^2\mathcal{S}_\lambda K}{\eta}+C_{0,0}+C_{1,0}K+C_{0,1}M} = \frac{B\log_2(1+\alpha K(\frac{M}{K}-1))}{\frac{\alpha B\sigma^2\mathcal{S}_\lambda}{\eta}+\frac{C_{0,0}}{K}+C_{1,0}+C_{0,1}\frac{M}{K}}$$

  - If $M, K \to \infty$ with $\frac{M}{K} = \text{constant} > 1$, then $\log_2(1+\alpha K(\frac{M}{K}-1)) \to \infty$ and $\mathrm{EE} \to \infty$!

Optimization of Energy-Efficiency

Preliminaries

- Our Goal
  - Optimize number of antennas M
  - Optimize number of active users K
  - Optimize the (normalized) transmit power α

  For ZF processing

- Outline
  - Optimize each variable separately
  - Devise an alternating optimization algorithm

Definition (Lambert W function)

- Lambert W function, $W(x)$, solves equation $W(x)e^{W(x)} = x$
- The function is increasing and satisfies $W(0) = 0$
- $e^{W(x)}$ behaves as a linear function (i.e., $e^{W(x)} \approx x$):

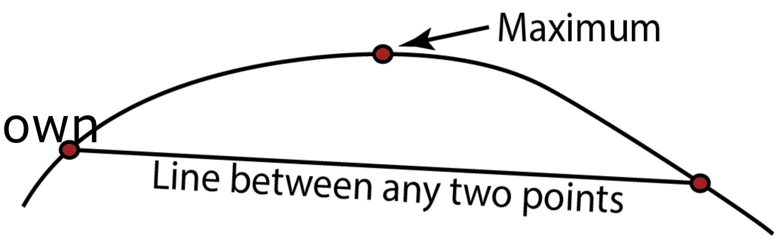
$$\frac{x\,e}{\log_e(x)} \leq e^{W(x)+1} \leq \frac{x}{\log_e(x)}(1+e) \quad \text{for} \quad x \geq e.$$

Solving Optimization Problems

- How to Solve an Optimization Problem?
  - Simple if the function is "nice":

Quasi-Concave Function

For any two points on the graph of the function,
the line between the points is below the graph



Property: Goes up and then down

Examples: $-x^2, \log(x)$

- Maximization of a Quasi-Concave Function $\varphi(x)$:
  1. Compute the first derivative $\frac{d}{dx}\varphi(x)$
  2. Find switching point by setting $\frac{d}{dx}\varphi(x) = 0$
  3. Only one solution → It is the unique maximum!

Optimal Number of BS Antennas

- Find M that maximizes EE with ZF:

$$\underset{M \geq K+1}{\text{maximize}} \quad \frac{K\left(1-\frac{\tau K}{U}\right) B \log_2(1+\alpha(M-K))}{\frac{\alpha B \sigma^2 \mathcal{S}_\lambda K}{\eta} + \sum_{i=0}^{3} C_{i,0} K^i + \sum_{i=0}^{2} C_{i,1} M K^i + AK\left(1-\frac{\tau K}{U}\right) B \log_2(1+\alpha(M-K))}$$

Theorem 1 (Optimal M)

EE is quasi-concave w.r.t. M and maximized by

$$M^* = \frac{e^{W\left(\frac{\alpha(B\sigma^2 \mathcal{S}_\lambda K/\eta + \sum_{i=0}^{3} C_{i,0} K^i)}{e \sum_{i=0}^{2} C_{i,1} K^i} + \frac{\alpha K - 1}{e}\right)+1} + \alpha K - 1}{\alpha}$$

- Observations
  - Increases with circuit coefficients independent of M (e.g., $P_{\mathrm{FIX}}, P_{\mathrm{UE}}$)
  - Decreases with circuit coefficients multiplied with M (e.g., $P_{\mathrm{BS}}, 1/L_{BS}$)
  - Independent of cost of coding/decoding/backhaul
  - Increases with power α approx. as $\frac{\alpha}{\log \alpha}$ (almost linear)

Optimal Transmit Power

- Find α that maximizes EE with ZF:

$$\underset{\alpha \geq 0}{\text{maximize}} \quad \frac{K\left(1-\frac{\tau K}{U}\right) B \log_2(1+\alpha(M-K))}{\frac{\alpha B \sigma^2 \mathcal{S}_\lambda K}{\eta} + \sum_{i=0}^{3} C_{i,0} K^i + \sum_{i=0}^{2} C_{i,1} M K^i + AK\left(1-\frac{\tau K}{U}\right) B \log_2(1+\alpha(M-K))}$$

Theorem 2 (Optimal α)

EE is quasi-concave w.r.t. α and maximized by

$$\alpha^* = \frac{e^{W\left(\frac{\eta}{B\sigma^2 \mathcal{S}_\lambda} \frac{(M-K)\left(\sum_{i=0}^{3} C_{i,0} K^i + \sum_{i=0}^{2} C_{i,1} M K^i\right)}{e} - \frac{1}{e}\right)+1} - 1}{M-K}$$

- Observations
  - Increases with all circuit coefficients (e.g., $P_{\text{FIX}}, P_{\text{BS}}, P_{\text{UE}}, 1/L_{BS}$)
  - Independent of cost of coding/decoding/backhaul
  - Increases with M approx. as $\frac{M}{\log M}$ (almost linear)

More circuit power →
More transmit power

Optimal Number of Users

- Find K that maximizes EE with ZF:

$$\underset{K \geq 0}{\text{maximize}} \quad \frac{K\left(1-\frac{\tau K}{U}\right) B \log_2(1+\bar{\alpha}(\bar{\beta}-1))}{\frac{\bar{\alpha} B \sigma^2 \mathcal{S}_\lambda}{\eta} + \sum_{i=0}^{3} C_{i,0} K^i + \sum_{i=0}^{2} C_{i,1} \bar{\beta} K^{i+1} + AK\left(1-\frac{\tau K}{U}\right) B \log_2(1+\bar{\alpha}(\bar{\beta}-1))}$$

where $\bar{\alpha} = \alpha K$ and $\bar{\beta} = \frac{M}{K}$ are fixed

Theorem 3 (Optimal K)

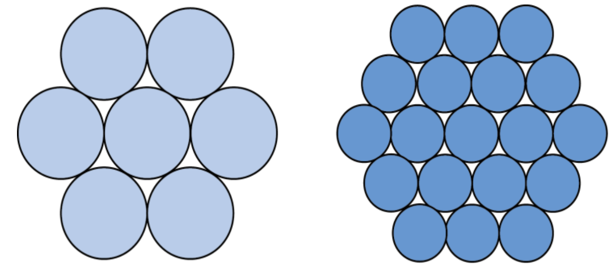
EE is quasi-concave w.r.t. K

Maximized by the root of a quartic polynomial:
Closed form for K^* but very "large" expressions

- Observations
  - Increases with fixed circuit power (e.g., P_{FIX})
  - Decreases with circuit coefficients multiplied with M or K ($P_{\text{BS}}, P_{\text{UE}}, 1/L_{BS}$)

Impact of Cell Size

- Are Smaller Cells More Energy Efficient?
  - Recall: $\mathcal{S}_\lambda = \mathbb{E}\left\{\frac{1}{\lambda}\right\}$
  - Smaller cells → λ is larger → $\mathcal{S}_\lambda$ is smaller



- For any given parameters M, α, K
  - Smaller $\mathcal{S}_\lambda$ → smaller transmit power $\alpha B \sigma^2 \mathcal{S}_\lambda K$
  - Higher EE!

- Expressions for M^*, α^*, K^*
  - M^* and K^* increases with $\mathcal{S}_\lambda$
  - α^* decreases with $\mathcal{S}_\lambda$

Smaller cells:
Less hardware and fewer users per cell
Use shorter distances to reduce power

Dependence on Other Parameters

Many other observations can be made
Example: Impact of bandwidth B, coherence block length U, etc.

Alternating Optimization Algorithm

- Joint EE Optimization
  - EE is a function of M, α, and K
  - Theorems 1-3 optimize one parameter, when the other two are fixed
  - Can we optimize all of them?

Algorithm: Alternating Optimization

1. Assume that an initial set (M, α, K) is given
2. Update number of users K (and implicitly M and α) using Theorem 3
3. Update number of antennas M using Theorem 1
4. Update transmit power (α) using Theorem 2
5. Repeat 2.-5. until convergence

Theorem 4

The algorithm convergences to a local optimum to the joint EE optimization problem

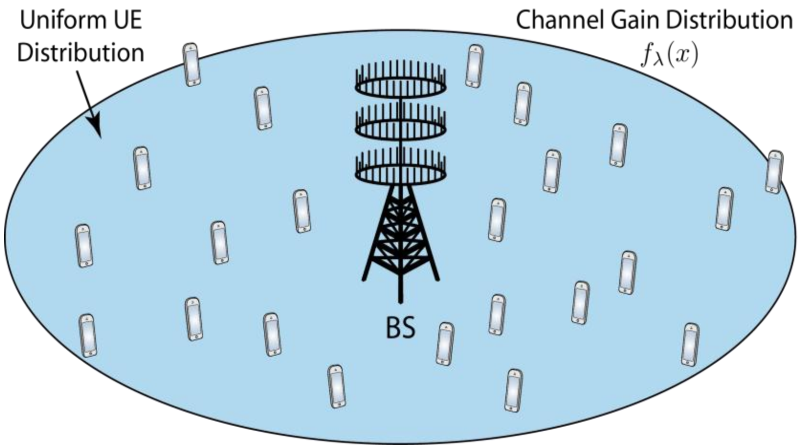
Disclaimer

M and K should be integers

Theorems 1 and 3 give real numbers → Take one of the 2 closest integers

Single-Cell Simulation Scenario

- Main Characteristics
  - Circular cell with radius 250 m
  - Uniform user distribution
  - Uncorrelated Rayleigh fading
  - Typical 3GPP pathloss model



- Many Parameters in the System Model
  - We found numbers from ≈ 2012 in the literature:

Parameter	Value	Parameter	Value
Cell radius (single-cell): $d_{\max}$	250 m	Fraction of downlink transmission: $\zeta^{(\mathrm{dl})}$	0.6
Minimum distance: $d_{\min}$	35 m	Fraction of uplink transmission: $\zeta^{(\mathrm{ul})}$	0.4
Large-scale fading model: $l(\mathbf{x})$	$10^{-3.53}/\|\mathbf{x}\|^{3.76}$	PA efficiency at the BSs: $\eta^{(\mathrm{dl})}$	0.39
Transmission bandwidth: B	20 MHz	PA efficiency at the UEs: $\eta^{(\mathrm{ul})}$	0.3
Channel coherence bandwidth: B_{C}	180 kHz	Fixed power consumption (control signals, backhaul, etc.): P_{FIX}	18 W
Channel coherence time: T_{C}	10 ms	Power consumed by local oscillator at BSs: P_{SYN}	2 W
Coherence block (channel uses): U	1800	Power required to run the circuit components at a BS: P_{BS}	1 W
Total noise power: $B\sigma^2$	-96 dBm	Power required to run the circuit components at a UE: P_{UE}	0.1 W
Relative pilot lengths: $\tau^{(\mathrm{ul})}, \tau^{(\mathrm{dl})}$	1	Power required for coding of data signals: P_{COD}	0.1 W/(Gbit/s)
Computational efficiency at BSs: L_{BS}	12.8 Gflops/W	Power required for decoding of data signals: P_{DEC}	0.8 W/(Gbit/s)
Computational efficiency at UEs: L_{UE}	5 Gflops/W	Power required for backhaul traffic: P_{BT}	0.25 W/(Gbit/s)

Optimal Single-Cell System Design: ZF Beamforming

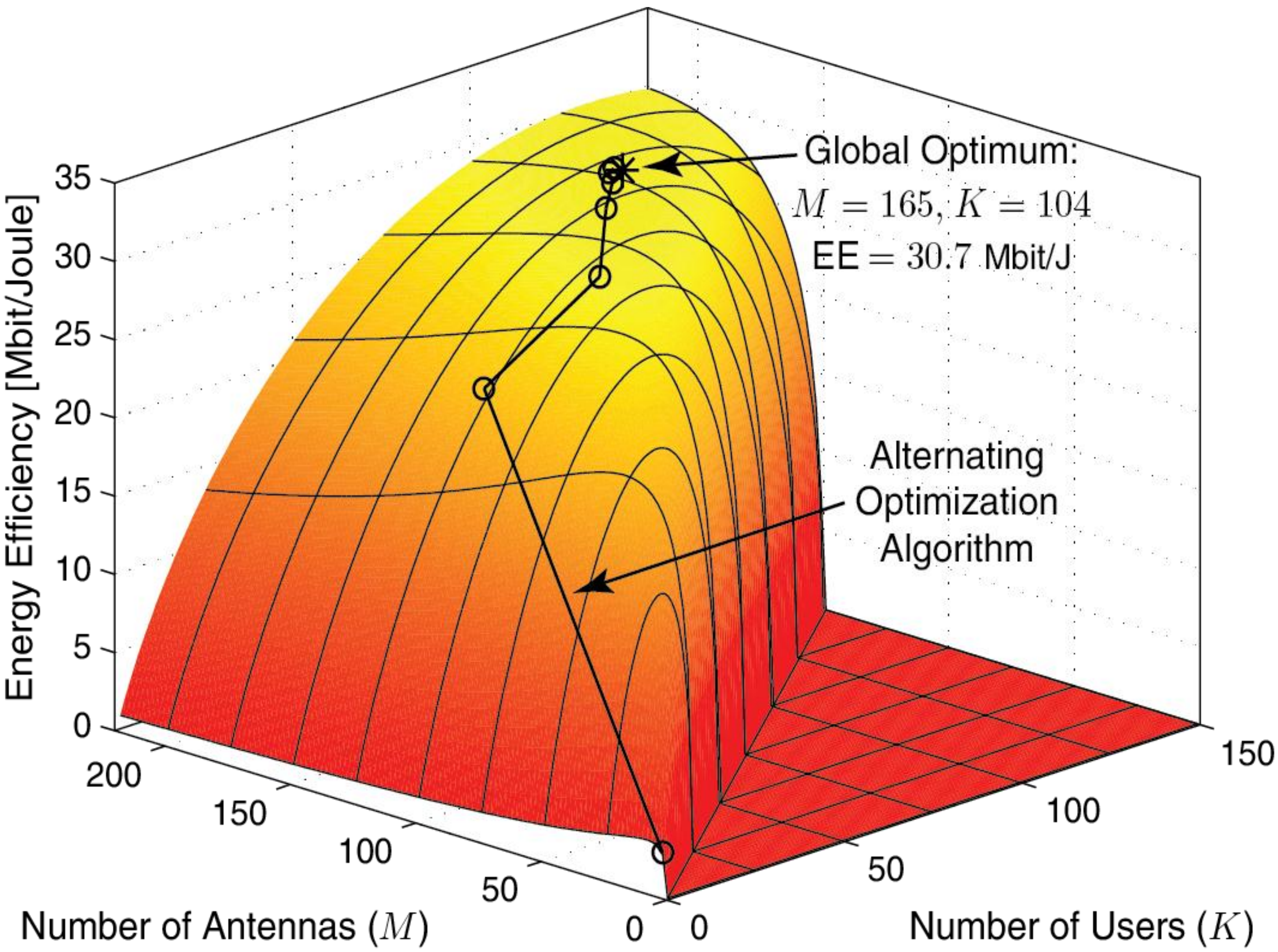
Optimum

$M = 165$
$K = 104$
$\alpha = 0.87$

User rates:
≈64-QAM

Massive MIMO!

Name for multi-user MIMO with very many antennas



Optimal Single-Cell System Design: "Optimal" Beamforming

Optimum

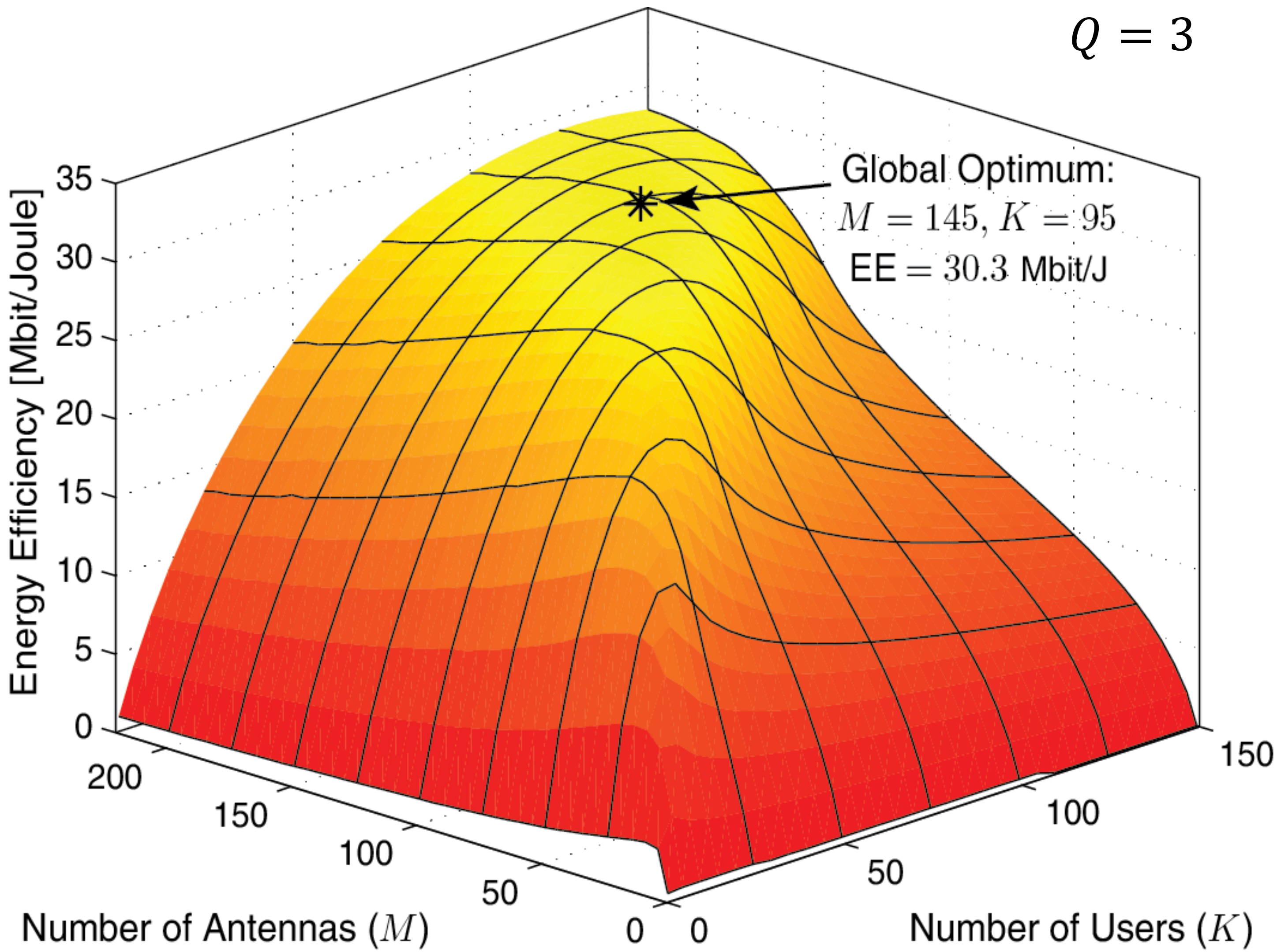
$M = 145$
$K = 95$
$\alpha = 0.91$

User rates: ≈64-QAM

Not optimal!

Gives optimal beamforming but computations are too costly

$Q = 3$



Optimal Single-Cell System Design: MRT/MRC Beamforming

Optimum

$M = 81$
$K = 77$
$\alpha = 0.24$

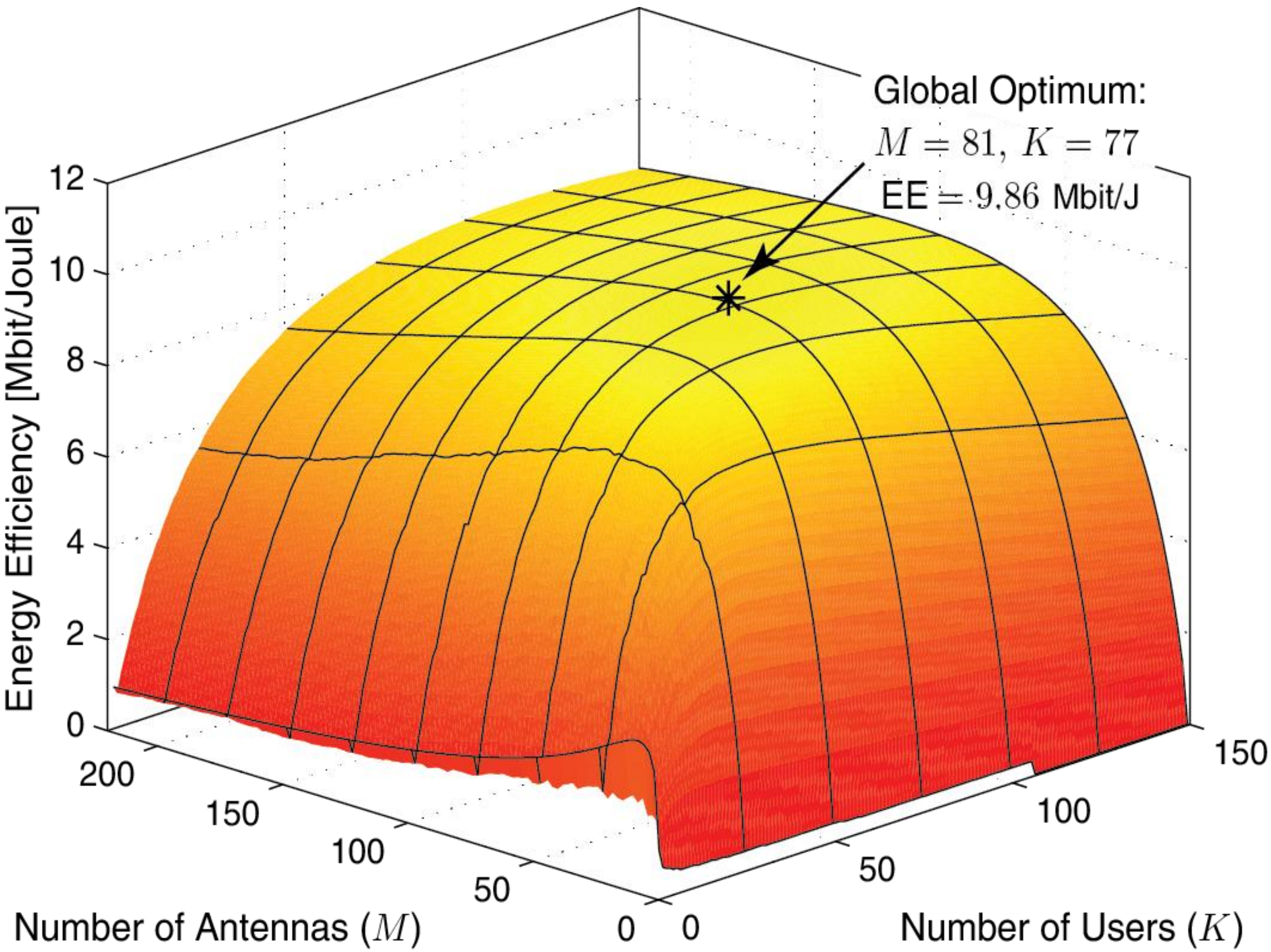
User rates:
≈2-PSK

Observation

Lower EE than with ZF

Also Massive MIMO setup

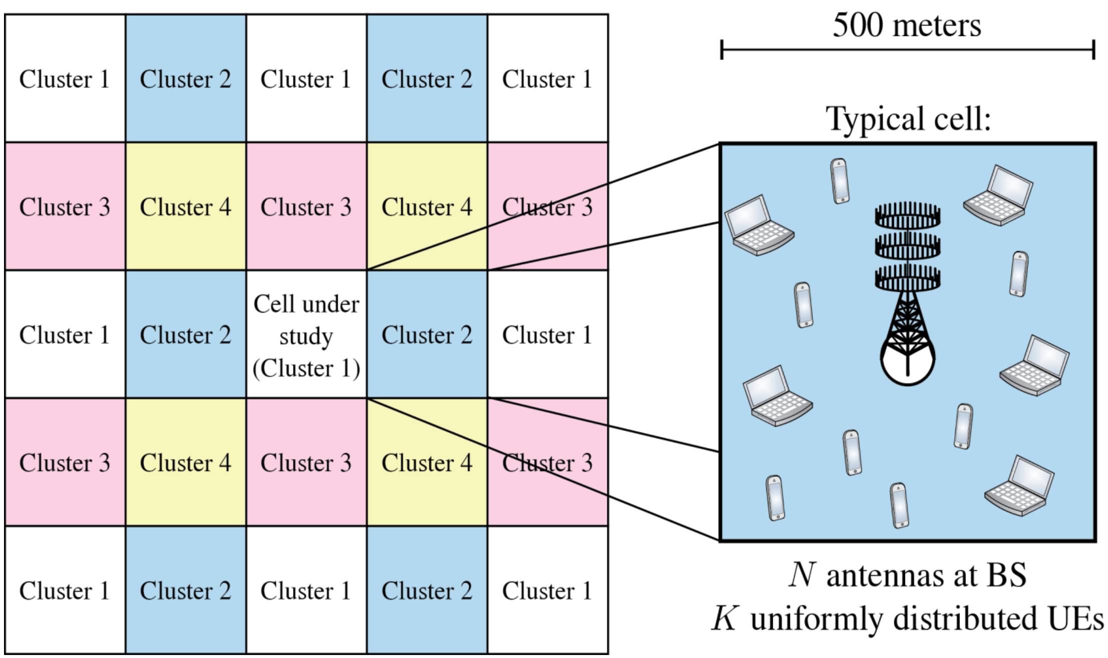
Low rates



Multi-Cell Scenarios and Imperfect Channel Knowledge

- Limitations in Previous Analysis
  - Perfect channel knowledge
  - No interference from other cells

- Consider a Symmetric Multi-Cell Scenario:



Assumptions

All cells look the same → Jointly optimized

All cells transmit in parallel

Fractional pilot reuse: Divide cells into clusters

Uplink pilot length $\tau^{(\mathrm{ul})}K$ for $\tau^{(\mathrm{ul})} \in \{1,2,4\}$

Multi-Cell Scenarios and Imperfect Channel Knowledge (2)

- Inter-Cell Interference
  - λ_{jl} = Channel attenuation between a random user in cell l and BS j
  - $\mathcal{I} = \sum_{l \neq j} \mathbb{E}\left\{\frac{\lambda_{jl}}{\lambda_{jj}}\right\}$ is relative severity of inter-cell interference

Lemma (Achievable Rate)

Consider same transmit power as before: $P_{\text{trans}} = \alpha B \sigma^2 \mathcal{S}_\lambda K$

Achievable rate under ZF and pilot-based channel estimation:

$$R = B \log_2\left(1 + \frac{\alpha(M-K)}{\alpha(M-K)\mathcal{I}_{\text{PC}} + \left(1 + \mathcal{I}_{\text{PC}} + \frac{1}{\alpha K \tau^{(\text{ul})}}\right)(1 + \alpha K \mathcal{I}) - \alpha K (1 + \mathcal{I}_{\text{PC}^2})}\right)$$

where $\mathcal{I}_{\text{PC}} = \sum_{l \neq j \text{ only in cluster}} \mathbb{E}\left\{\frac{\lambda_{jl}}{\lambda_{jj}}\right\}$ and $\mathcal{I}_{\text{PC}^2} = \sum_{l \neq j \text{ only in cluster}} \mathbb{E}\left\{\left(\frac{\lambda_{jl}}{\lambda_{jj}}\right)^2\right\}$

Pilot contamination (PC)
(Strong interference)

Intra/Inter-cell interference
(Weaker)

Multi-Cell Scenarios and Imperfect Channel Knowledge (3)

- Multi-Cell Rate Expression not Amenable for Analysis
  - No closed-form optimization in multi-cell case
  - Numerical analysis still possible

- Similarities and Differences
  - Power consumption is exactly the same
  - Rates are smaller: Upper limited by pilot contamination:

$$R = B\log_2\left(1 + \frac{\alpha(M-K)}{\alpha(M-K)\mathcal{I}_{\mathrm{PC}} + \left(1 + \mathcal{I}_{\mathrm{PC}} + \frac{1}{\alpha K \tau^{(\mathrm{ul})}}\right)(1+\alpha K \mathcal{I}) - \alpha K (1 + \mathcal{I}_{\mathrm{PC}^2})}\right) \leq B\log_2\left(1 + \frac{1}{\mathcal{I}_{\mathrm{PC}}}\right)$$

  - Overly high rates not possible (but we didn’t get that…)
  - Clustering (fractional pilot reuse) might be good to reduce interference

Optimal Multi-Cell System Design: ZF Beamforming

Optimum

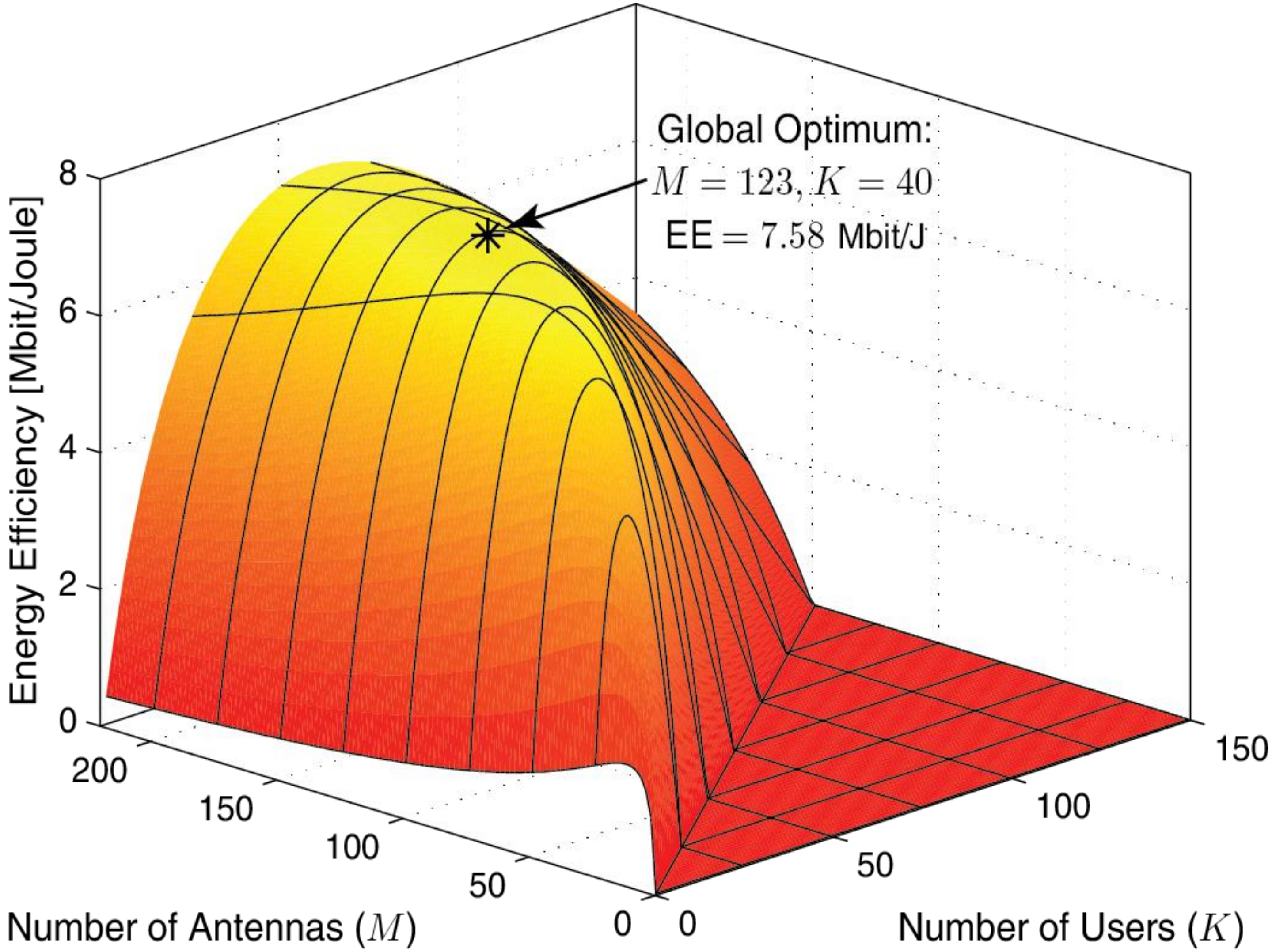
$M = 123$
$K = 40$
$\alpha = 0.28$
$\tau^{(\mathrm{ul})} = 4$

User rates:
≈4-QAM

Massive MIMO!

Many BS antennas

Note that
$M/K \approx 3$



Different Pilot Reuse Factors

Higher Pilot Reuse

Higher EE *and* rates!

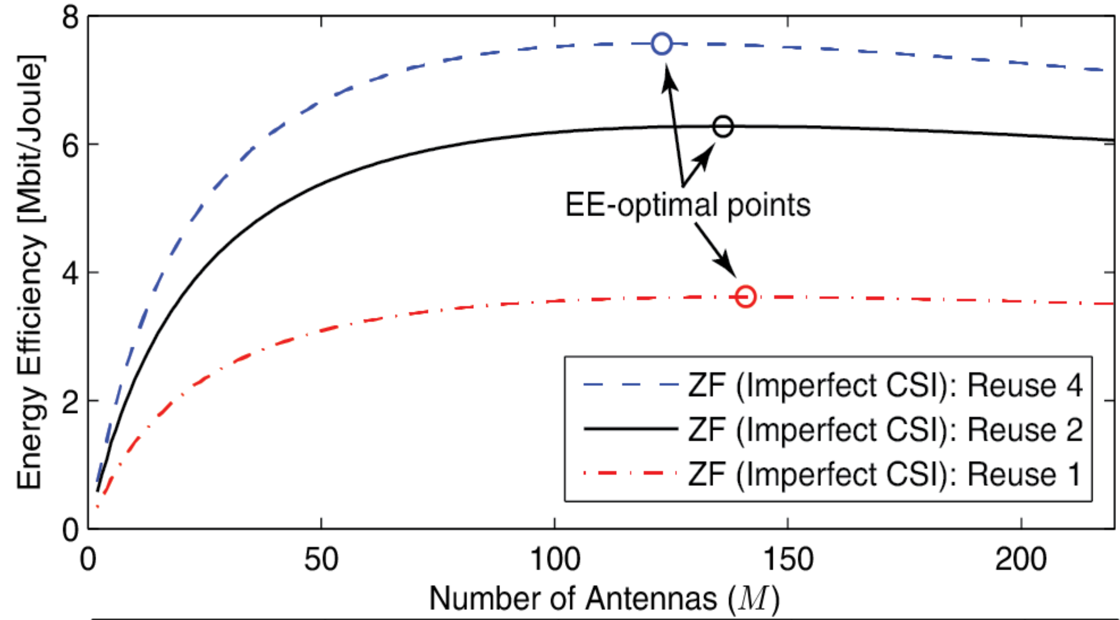
Controlling inter-cell interference is very important!

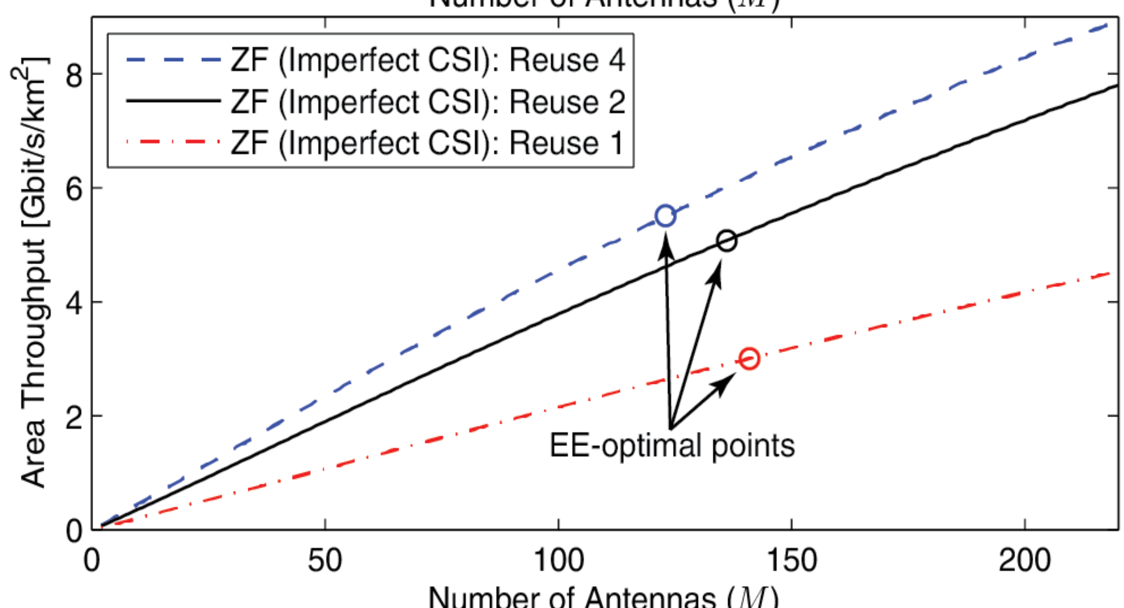
Area Throughput

We only optimized EE

Achieved $6\ \mathrm{Gbit/s/km^2}$ over 20 MHz bandwidth

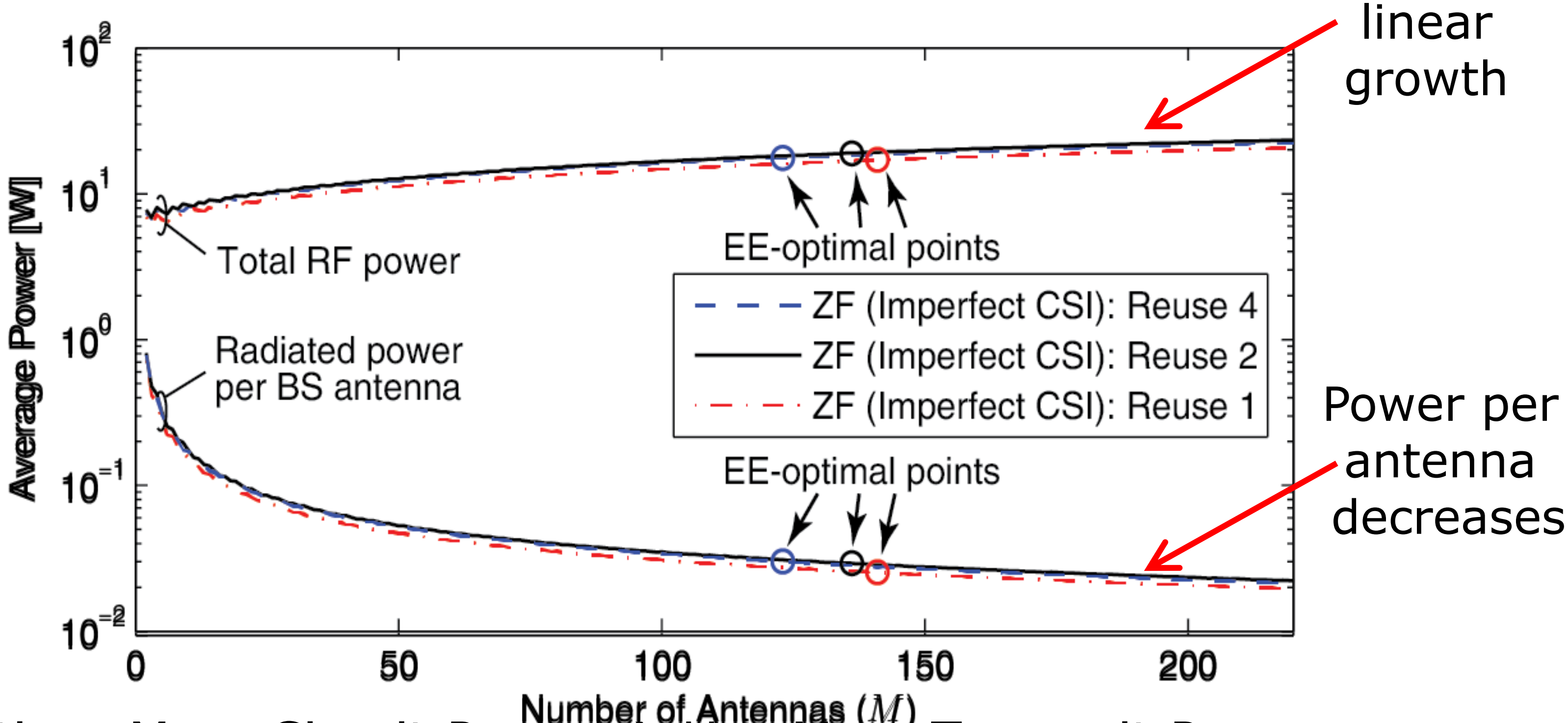
METIS project mentions $100\ \mathrm{Gbit/s/km^2}$ as 5G goal
→ Need higher bandwidth!





Energy Efficient to Use More Transmit Power?

- Recall from Theorem 2: Transmit power increases M
  - Figure shows EE-maximizing power for different M



- Intuition: More Circuit Power → Use More Transmit Power
  - Different from $1/\sqrt{M}$ scaling laws in recent massive MIMO literature
  - Power per antennas decreases, but only logarithmically

Summary

- Optimization Results
  - EE is a quasi-concave function of (M, K, α)
  - Closed-form optimal M, K, or α for single-cell
  - Alternating optimization algorithm

Simulations

Depends on parameters

Download Matlab code to try other values!

Reveals how variables are connected

	Increases with	Decreases with
Antennas M	Power α, coverage area $\mathcal{S}_\lambda$, and M-independent circuit power	M-related circuit power
Users K	Fixed circuit power $\mathcal{C}_{0,0}$ and coverage area $\mathcal{S}_\lambda$	K-related circuit power
Transmit power $\alpha B \sigma^2 \mathcal{S}_\lambda K$	Circuit power, coverage area $\mathcal{S}_\lambda$, antennas M, and users K	-

Large Cell

More antennas, users, RF power

Massive MIMO Appears Naturally

Fractional pilot reuse important!

More Circuit Power

Use more transmit power

Limits of M, K

Circuit power that scales with M,K

Optimize more than Energy-Efficiency

- Recall: Many Metrics in 5G Discussions
  - Average rate (Mbit/s/active user)
  - Average area rate (Mbit/s/km^2)
  - Energy-efficiency (Mbit/Joule)
  - Active devices (per km^2)
  - Delay constraints (ms)

- So Far: Only cared about EE
  - Ignored all other metrics

Optimize Multiple Metrics

We want efficient operation w.r.t. all objectives

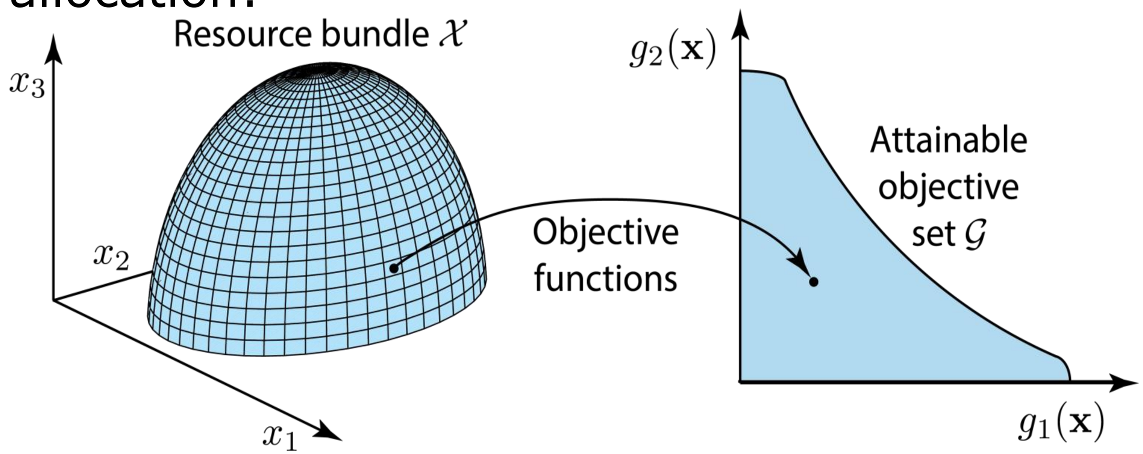
Is this possible?
For all at the same time?

Multi-Objective Network Optimization

Basic Assumptions: Multi-Objective Optimization

- Consider N Performance Metrics
  - Objectives to be maximized
  - Notation: $g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_N(\mathbf{x})$
  - Example: individual user rates, area rates, energy-efficiency

- Optimization Resources
  - Resource bundle: $\mathcal{X}$
  - Example: power, resource blocks, network architecture, antennas, users
  - Feasible allocation: $\mathbf{x} \in \mathcal{X}$



Single or Multiple Performance Metrics

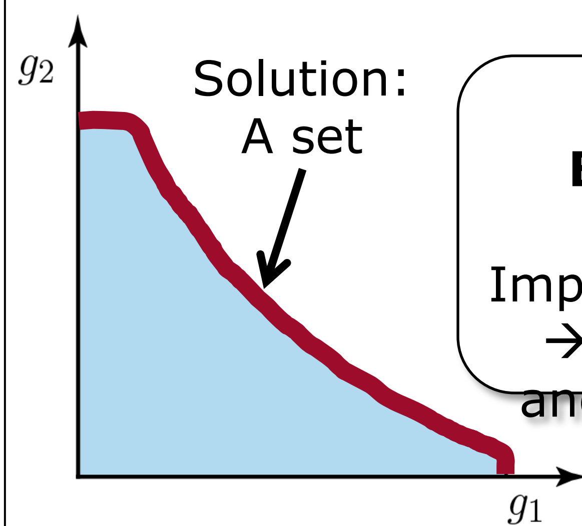
- Conventional Optimization
  - Pick one prime metric: $g_1(\mathbf{x})$
  - Turn $g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_N(\mathbf{x})$ into constraints
  - Optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & g_1(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \\ & g_2(\mathbf{x}) \geq C_2, \ldots, g_N(\mathbf{x}) \geq C_N. \end{aligned}$$

  - Solution: A scalar number
  - Cons: Is there a prime metric?
    How to select constraints?

- Multi-Objective Optimization
  - Consider all N metrics
  - No order or preconceptions!
  - Optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & [g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_N(\mathbf{x})] \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}. \end{aligned}$$



Pareto Boundary

Improve a metric → Degrading another metric

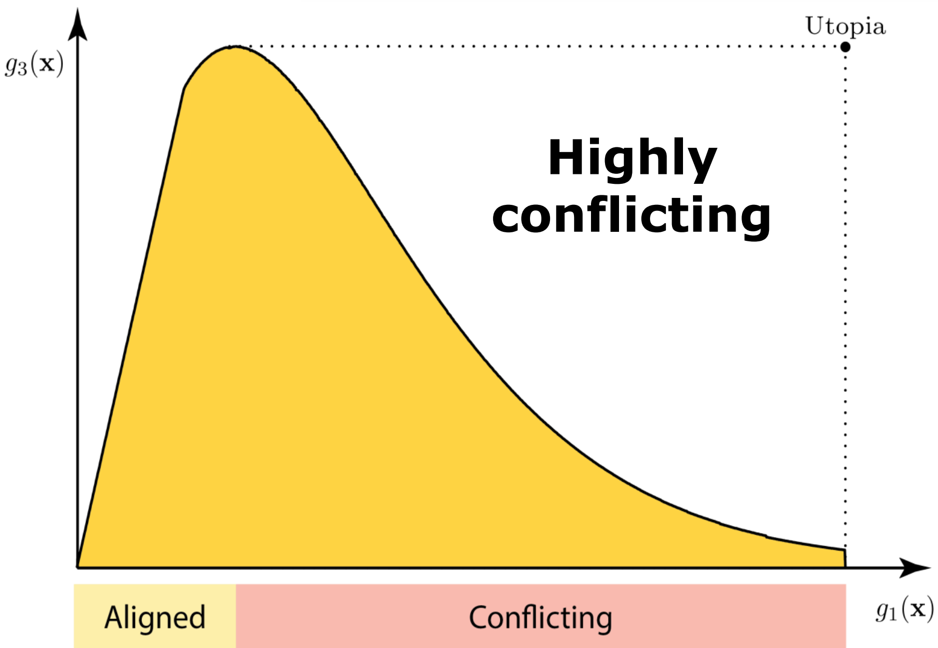
Why Multi-Objective Optimization?

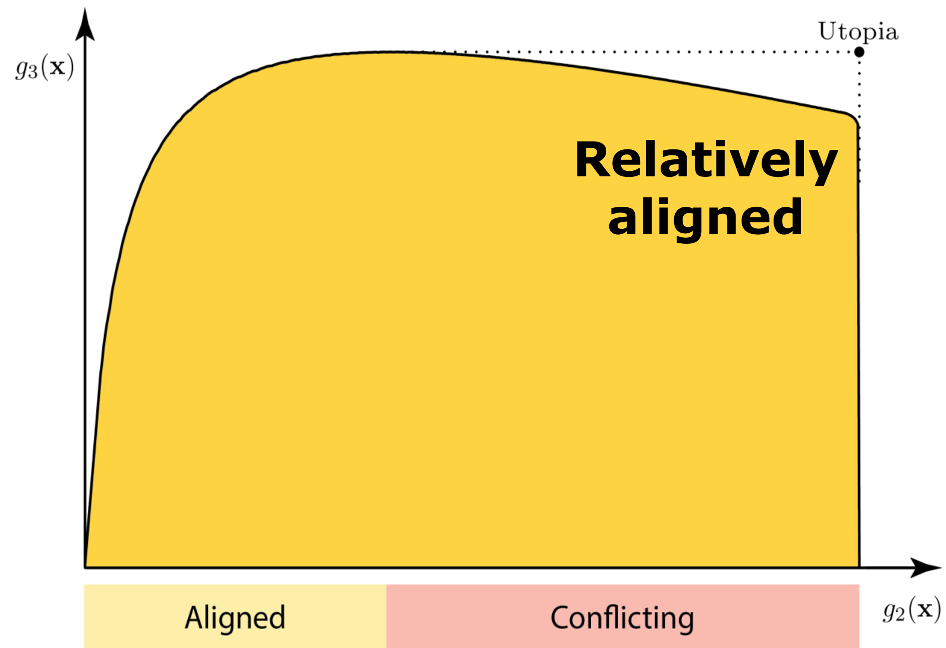
- Study Tradeoffs Between Metrics
  - When are metrics aligned or conflicting?
  - Common in engineering and economics – new in communication theory

***A Posteriori* Approach**

Generate region (computationally demanding!)

Look at region and select operating point





A Priori Approach

- No Objectively Optimal Solution
  - Utopia point outside of region → Only subjectively "good" solutions exist

- System Designer Selects Utility Function $f : \mathbb{R}^N \rightarrow \mathbb{R}$
  - Describes subjective preference (larger is better)

- Examples:
  - Sum performance: $f(\mathbf{g}) = \sum_k g_k$
  - Proportional fairness: $f(\mathbf{g}) = \prod_k g_k$
  - Harmonic mean: $f(\mathbf{g}) = K_r \left(\sum_k g_k^{-1}\right)^{-1}$
  - Max-min fairness: $f(\mathbf{g}) = \min_k g_k$

Aggregate metric

Fairness of metrics

We obtain a simplified problem:

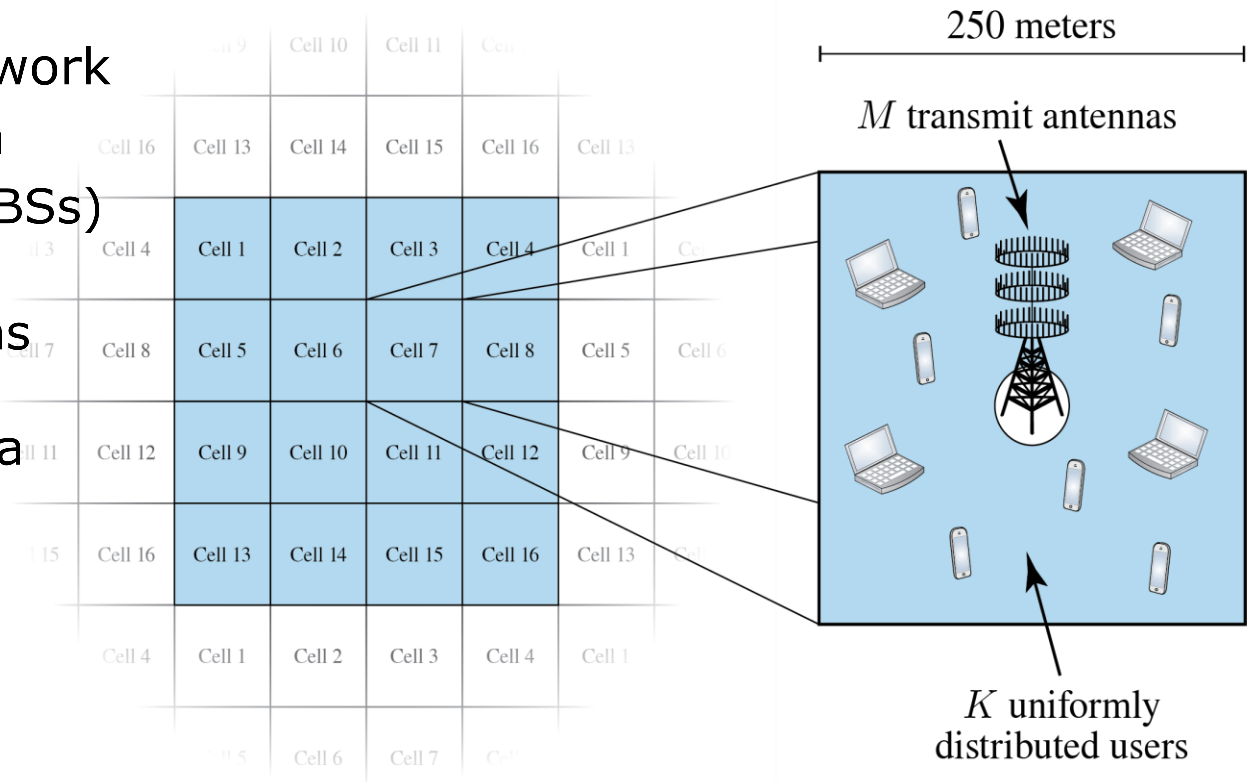
$$\underset{\mathbf{x}}{\text{maximize}} \quad f(g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_N(\mathbf{x}))$$
$$\text{subject to} \quad \mathbf{x} \in \mathcal{X}$$

- Solution: A scalar number
  (Gives one Pareto optimal point)

- Takes all metrics into

Example: Optimization of 5G Networks

- Design Cellular Network
  - Symmetric system
  - 16 base stations (BSs)
  - Select:
    M = # BS antennas
    K = # users
    P = power/antenna



- Resource bundle:

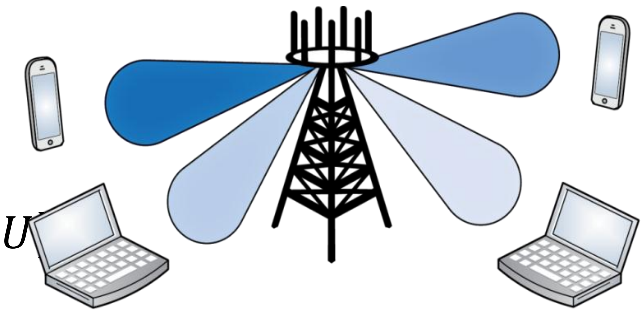
$$\mathcal{X} = \left\{ [K \ M \ P]^T : \begin{array}{c} 1 \leq K \leq \frac{M}{2}, \\ 2 \leq M \leq M_{\max}, \\ 0 \leq P \leq M P_{\max} \end{array} \right\}$$

500

20 W

Example: Optimization of 5G Networks (2)

- Downlink Multi-Cell Transmission
  - Each BS serves only its own K users
  - Coherence block length: U
  - BS knows channels within the cell (cost: K/U)
  - ZF beamforming: no intra-cell interference
  - Interference leaks between cells



- Average User Rate

Power/user

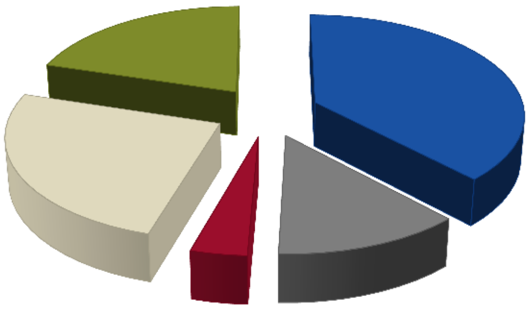
$$R_{\text{average}} = B \underbrace{\left(1-\frac{K}{U}\right)}_{\text{CSI estimation overhead } (U=1000)} \log_2\left(1+\frac{\frac{P}{K}\overbrace{(M-K)}^{\text{Array gain}}}{\mathcal{S}_\lambda \sigma^2 + \mathcal{I}}\right)$$

Bandwidth (10 MHz)

Noise / pathloss ($1.72 \cdot 10^{-4}$)

Relative inter-cell interference (0.54)

Example: Optimization of 5G Networks (3)



- What Consumes Power?
  - Transmit power (+ losses in amplifiers)
  - Circuits attached to each antenna
  - Baseband signal processing
  - Fixed load-independent power

- Total Power Consumption

$$P_{\text{total}} = \frac{P_{\text{trans}}}{\eta} + C_{0,0} + C_{1,0}K + C_{0,1}M + \frac{B\,C_{\text{beamforming}}}{U\,L_{\text{BS}}}$$

Amplifier efficiency (0.31)

Fixed power (10 W)

Circuit power per user (0.3 W)

Circuit power per antenna (1 W)

Computing ZF beamforming $(2.3 \cdot 10^{-6} \cdot MK^2)$

Example: Results

3 Objectives

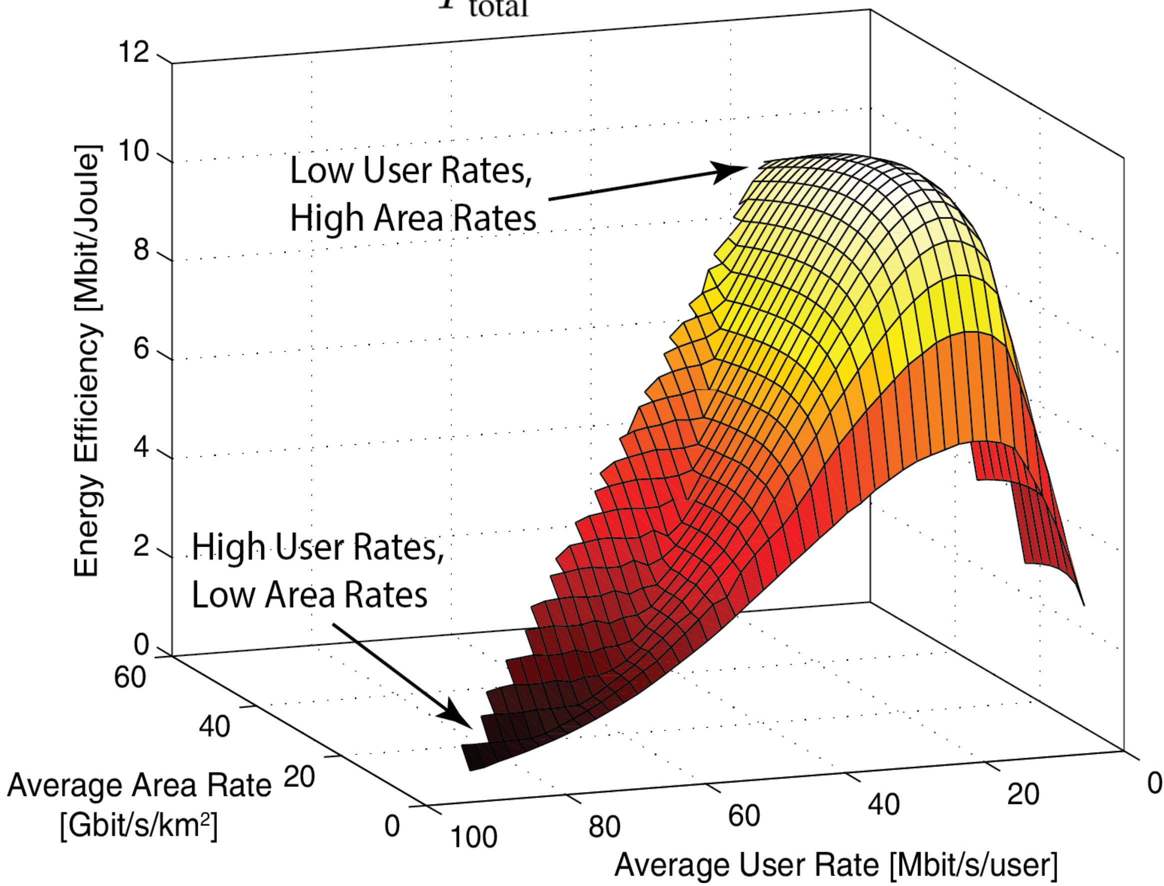
1. Average user rate $g_1(\mathbf{x}) = R_{\text{average}}$ [bit/s/user]
2. Total area rate $g_2(\mathbf{x}) = \frac{K}{A} R_{\text{average}}$ [bit/s/km^2]
3. Energy-efficiency $g_3(\mathbf{x}) = \frac{K R_{\text{average}}}{P_{\text{total}}}$ [bit/J]

Observations

Area and user rates are conflicting objectives

Only energy efficient at high area rates

Different number of users



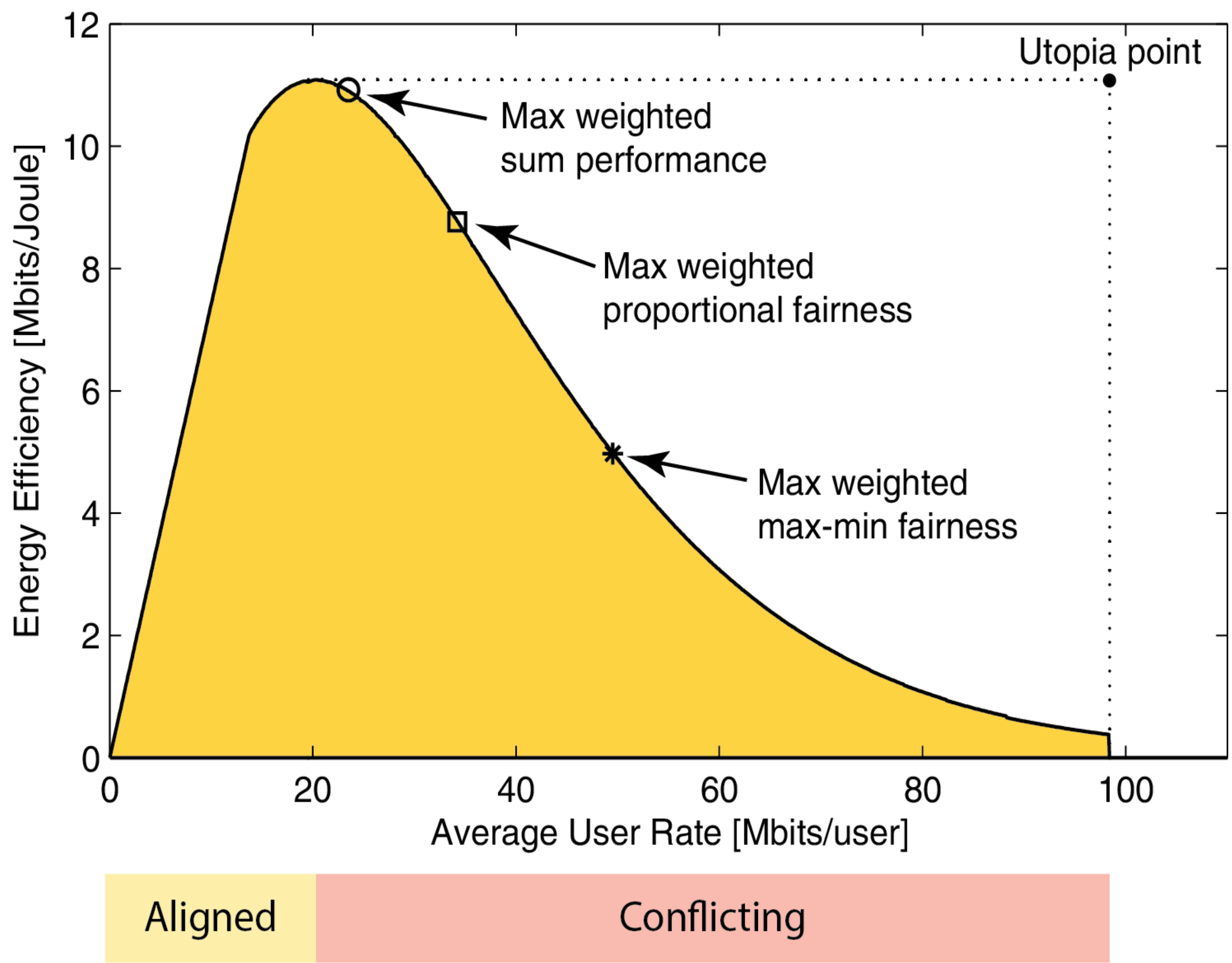
Example: Results (2)

- Energy-Efficiency vs. User Rates
  - Utility functions normalized by utopia point

Observations

Aligned for small user rates

Conflicting for high user rates



References

- L.Sanguinetti, A. L. Moustakas, E. Björnson and M. Debbah, ”**Large System Analysis of the Energy Consumption Distribution in Multi-User MIMO Systems with Mobility**”, submitted to IEEE Transactions on Wireless Communications, 2014
- E. Björnson, J. Hoydis, M. Kountouris and M. Debbah ”**Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits**”, accepted for publication, IEEE Transactions on Information Theory, 2014
- E. Björnson, M. Kountouris, and M. Debbah, “**Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination**,” in Proc. Int. Conf. Telecommun. (ICT), 2013.
- E. Björnson, L. Sanguinetti, J. Hoydis, M. Debbah, “**Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer?,”** IEEE Transactions on Wireless Communications, Submitted for publication, 2014
- E. Björnson, E. Jorswieck, M. Debbah, B. Ottersten, “**Multi-Objective Signal Processing Optimization: The Way to Balance Conflicting Metrics in 5G Systems**,” To Appear in IEEE Signal Processing Magazine, Special Issue on Signal Processing for the 5G Revolution.