Mathematics of Multi-Antenna Transmission in 5G networks

Ilya Sinilshchikov, Evgeny Bobrov
Sinill57@gmail.com
eugenbobrov@ya.ru

Agenda

Introduction to cellular communication systems

1. **Massive MIMO precoding in 5G**
    1. **Precoding basics**
    2. **Optimization approaches**
    3. **Quantization**
    4. **User selection**
2. **Link adaptation**
3. **Brief overview of other problems**

Cellular Networking

2020s: + cloud AR/VR/Gaming, smart cities, robots

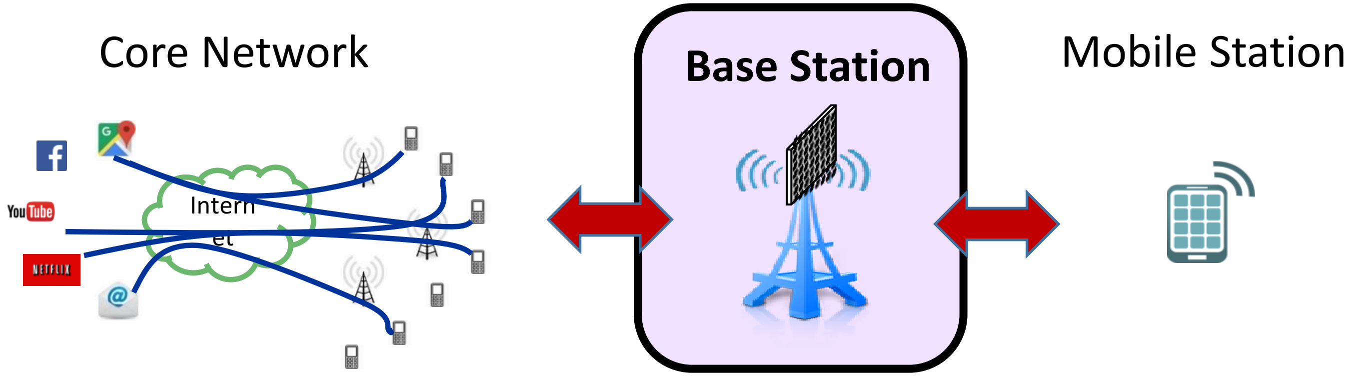
Evolution mainly due to **mathematics**

2010s: + video, live streaming, maps

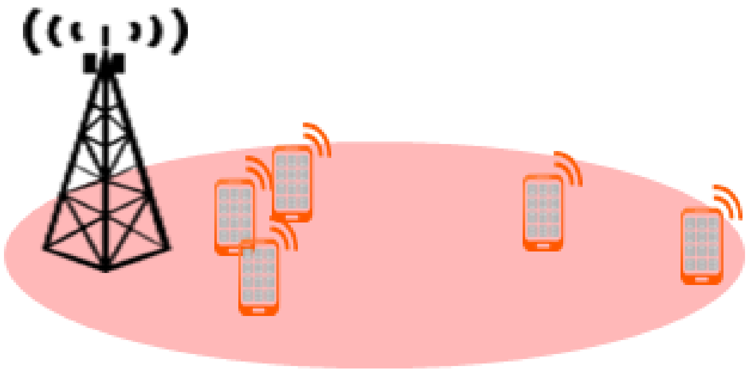
2000s: + web surfing, music

Evolution mainly due to **physics**

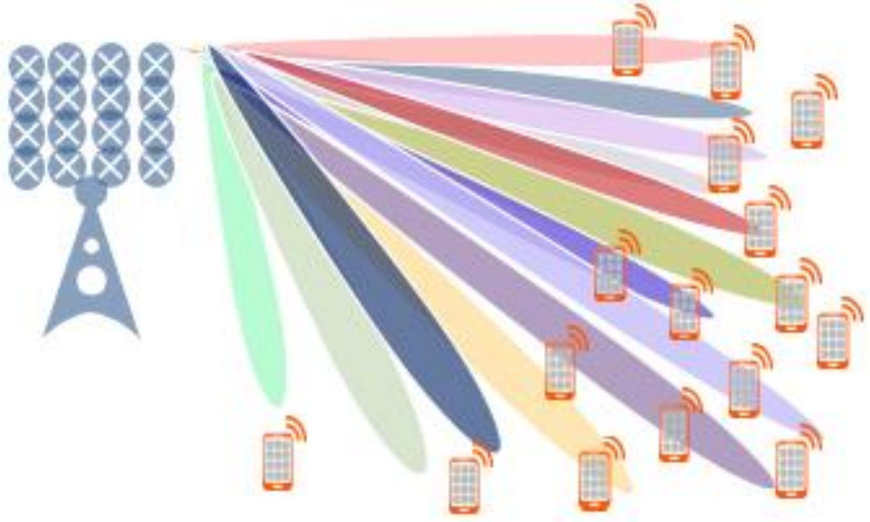
1990s: voice calls, text messaging, basic data service



Single antenna transmission



Massive MIMO transmission – key 5G technology

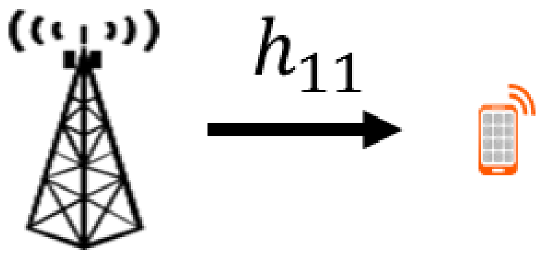


- Evolution due to efficient resource utilization
- **More sophisticated transmission requires cutting-edge mathematical methods**

Source of

- **Non-convex optimization problems**
- **Combinatorial optimization problems**
- **Stochastic optimization problems**
- **Dynamic control problems**

Single antenna transmission



After modulation data is represented by a complex number (symbol):

$$x_1 \in \mathbb{C}$$

$$|x_1|^2 = 1$$

Channel h_{ij}:

$h_{ij} \in \mathbb{C}, |h_{ij}| \leq 1$

$i \to j$

Transmitter antenna

Receiver antenna

Describes signal transformation during transmission from antenna i to antenna j

Transmission with power p

$$y_1 = h_{11} \cdot \sqrt{p} \cdot x_1 + noise$$

Received symbol

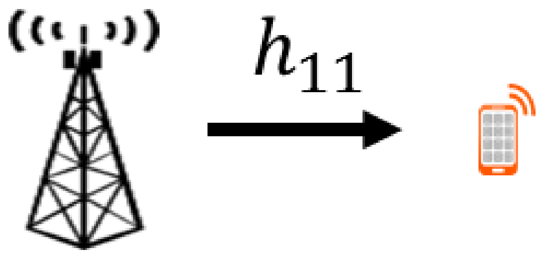
Transmitted symbol

Input-output relation of linear system:

$$y(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau$$

$$y_n = x_n \circledast h_n \Longleftrightarrow Y(k) = X(k)H(k)$$

Single antenna transmission



After modulation data is represented by a complex number (symbol):

$$x_1 \in \mathbb{C}$$

$$|x_1|^2 = 1$$

Channel h_{ij}:

$h_{ij} \in \mathbb{C}$, $|h_{ij}| \leq 1$

$i \rightarrow j$

Transmitter antenna

Receiver antenna

Describes signal transformation during transmission from antenna i to antenna j

Transmission with power p

$$y_1 = h_{11} \cdot \sqrt{p} \cdot x_1 + noise$$

Received symbol

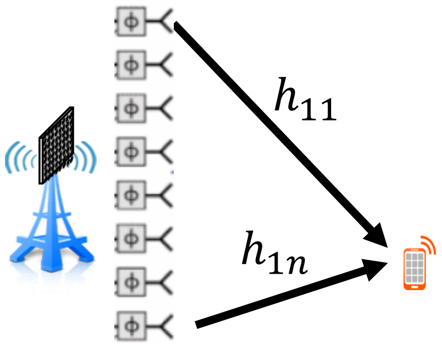
Transmitted symbol

$$SINR_1 = \frac{P(signal)}{P(noise + interference)} = \frac{p \cdot |h_{11}|^2}{\delta^2}$$

Shannon theorem for gaussian channel:
upper bound for information transmission capacity is

$$C = \max_{p(x):E|x|^2 \leq 1} I(X;Y) = \max_{p(x):E|x|^2 \leq 1} H(Y) - H(Y|X) = \log(1 + SINR)$$

Multi antenna transmission



n – number of transmitting antennas

w_k^1– "**weight**" of the symbol at antenna

Symbol x_1 is multiplied by w_k^1 and then transmitted from k-th antenna

$$w^1 = \begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix}$$ -- precoding vector

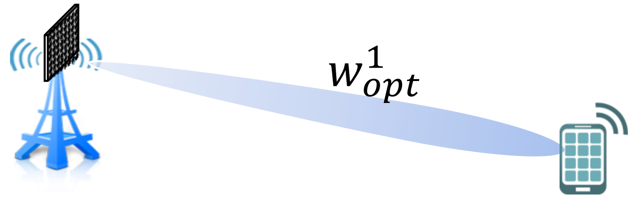
$w^1 \in \mathbb{C}^n$,
$||w^1||_{L^\infty} \le p$,

$$y_1 = (h_{11} \ldots h_{1n}) \cdot \begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix} \cdot x_1 + noise$$

$$SINR_1 = \frac{\left|\langle h_1, w^{1^*} \rangle\right|^2}{\delta^2}$$

Which **precoding vector** w^1 maximizes $SINR$ for user with channel h_1?

Answer: $\boldsymbol{w_{opt}^1 = c \cdot h_1^*}$



CR-calculus

$$f(z) = f(z, \bar{z}) = f(x, y) = u(x, y) + j\, v(x, y)$$

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

exists only for holomorphic functions

MSE is not holomorphic!

$$f(z) = |z|^2 = \bar{z}z = x^2 + y^2$$

Conjugate Coordinates: $c \triangleq (z, \bar{z})^T \in \mathbb{C} \times \mathbb{C}\,, \quad z = x + j\, y$ and $\bar{z} = x - j\, y$

$\underline{\mathbb{R}\text{-Derivative of } f(c)} \triangleq \left.\frac{\partial f(z, \bar{z})}{\partial z}\right|_{\bar{z} = \text{const.}}$ and $\underline{\text{Conjugate } \mathbb{R}\text{-Derivative of } f(c)} \triangleq \left.\frac{\partial f(z, \bar{z})}{\partial \bar{z}}\right|_{z = \text{const.}}$

Cauchy Riemann Condition: $\frac{\partial f}{\partial \bar{z}} = 0$

CR-calculus identities

$$f(z) = f(z, \bar{z}) = f(x, y) = u(x, y) + j\, v(x, y)$$

$$\boxed{\underline{\mathbb{R}\text{-Derivative of } f(c)} \triangleq \left.\frac{\partial f(z,\bar{z})}{\partial z}\right|_{\bar{z}=\text{ const.}} \quad \text{and} \quad \underline{\text{Conjugate } \mathbb{R}\text{-Derivative of } f(c)} \triangleq \left.\frac{\partial f(z,\bar{z})}{\partial \bar{z}}\right|_{z=\text{ const.}}}$$

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - j\frac{\partial f}{\partial y}\right) \quad \text{and} \quad \frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y}\right)$$

$$\begin{aligned} \frac{\partial \bar{f}}{\partial \bar{z}} &= \overline{\left(\frac{\partial f}{\partial z}\right)} \\ \frac{\partial \bar{f}}{\partial z} &= \overline{\left(\frac{\partial f}{\partial \bar{z}}\right)} \\ df &= \frac{\partial f}{\partial z}\, dz + \frac{\partial f}{\partial \bar{z}}\, d\bar{z} \end{aligned} \quad \Longrightarrow \quad f(z) \in \mathbb{R} \Rightarrow \overline{\left(\frac{\partial f}{\partial z}\right)} = \frac{\partial f}{\partial \bar{z}}$$

Application to gradient descent

$$L = L(s) \in \mathbb{R}; \qquad s = f(z) = f(x, y) \in \mathbb{C}$$

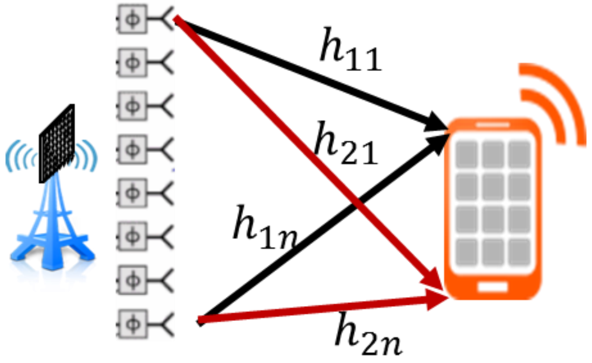
$$\begin{aligned} z_{n+1} &= x_n - (\alpha/2) * \frac{\partial L}{\partial x} + 1j * (y_n - (\alpha/2) * \frac{\partial L}{\partial y}) \\ &= z_n - \alpha * 1/2 * (\frac{\partial L}{\partial x} + j\frac{\partial L}{\partial y}) \\ &= z_n - \alpha * \frac{\partial L}{\partial z^*} \end{aligned}$$

$$\frac{\partial L}{\partial z^*} = \frac{\partial L}{\partial u} * \frac{\partial u}{\partial z^*} + \frac{\partial L}{\partial v} * \frac{\partial v}{\partial z^*}$$

$$\begin{aligned} \frac{\partial L}{\partial z^*} &= (\frac{\partial L}{\partial s} + \frac{\partial L}{\partial s^*}) * \frac{\partial u}{\partial z^*} - 1j * (\frac{\partial L}{\partial s} - \frac{\partial L}{\partial s^*}) * \frac{\partial v}{\partial z^*} \\ &= \frac{\partial L}{\partial s} * (\frac{\partial u}{\partial z^*} + \frac{\partial v}{\partial z^*} j) + \frac{\partial L}{\partial s^*} * (\frac{\partial u}{\partial z^*} - \frac{\partial v}{\partial z^*} j) \\ &= \frac{\partial L}{\partial s^*} * \frac{\partial (u + vj)}{\partial z^*} + \frac{\partial L}{\partial s} * \frac{\partial (u + vj)^*}{\partial z^*} \\ &= \frac{\partial L}{\partial s} * \frac{\partial s}{\partial z^*} + \frac{\partial L}{\partial s^*} * \frac{\partial s^*}{\partial z^*} \end{aligned}$$

$$\frac{\partial L}{\partial z^*} = (\frac{\partial L}{\partial s^*})^* * \frac{\partial s}{\partial z^*} + \frac{\partial L}{\partial s^*} * (\frac{\partial s}{\partial z})^*$$ - backpropagation rule

Multi-stream transmission



$$y_1 = (h_{11} \dots h_{1n}) \cdot \left(\begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix} \cdot x_1 + \begin{pmatrix} w_1^2 \\ \vdots \\ w_n^2 \end{pmatrix} \cdot x_2 \right) + noise_1$$

$$y_2 = (h_{21} \dots h_{2n}) \cdot \left(\begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix} \cdot x_1 + \begin{pmatrix} w_1^2 \\ \vdots \\ w_n^2 \end{pmatrix} \cdot x_2 \right) + noise_2$$

$$y = H \cdot W \cdot x + noise$$

Channel matrix **Precoding matrix**

Real-world channel matrix origin:

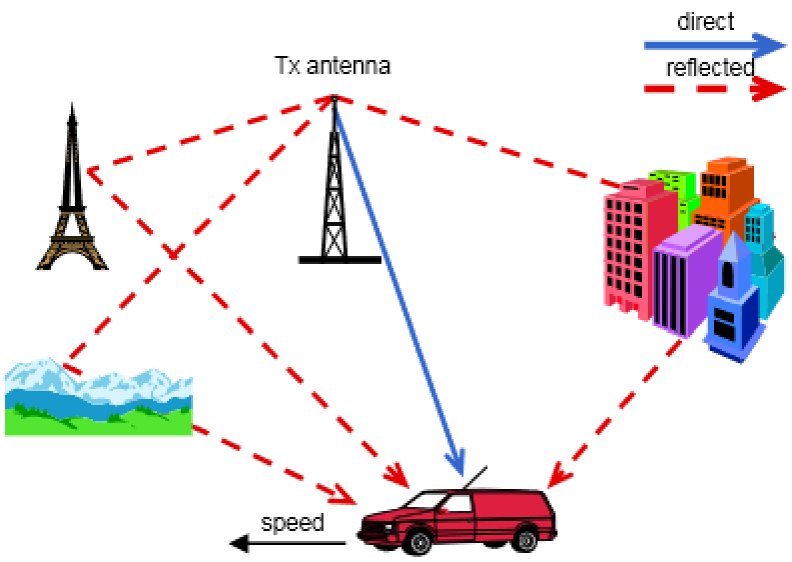
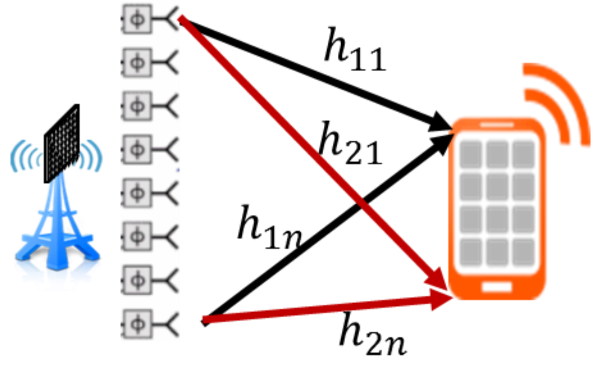


Fig. 1: Fading principle.

Multi-stream transmission



$$y_1 = (h_{11} \dots h_{1n}) \cdot \left(\begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix} \cdot x_1 + \begin{pmatrix} w_1^2 \\ \vdots \\ w_n^2 \end{pmatrix} \cdot x_2 \right) + noise_1$$

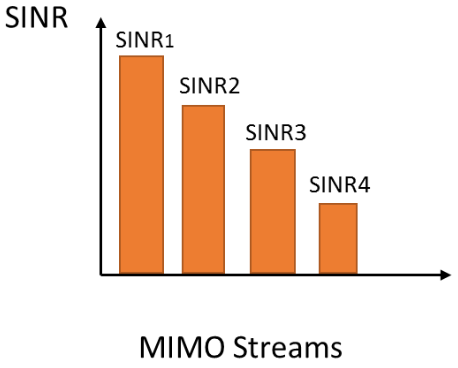
$$y_2 = (h_{21} \dots h_{2n}) \cdot \left(\begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix} \cdot x_1 + \begin{pmatrix} w_1^2 \\ \vdots \\ w_n^2 \end{pmatrix} \cdot x_2 \right) + noise_2$$

$$y = H \cdot W \cdot x + noise$$

Channel matrix **Precoding matrix**

Shannon formula:

$$Rate \leq \log_2 \det\left(I + \frac{P}{N \cdot \delta^2} HW(HW)^*\right)$$



Capacity optimization:
Ideal receiver is assumed

SVD-based Single-User precoding

W is selected as singular vectors of SVD decomposition of H:

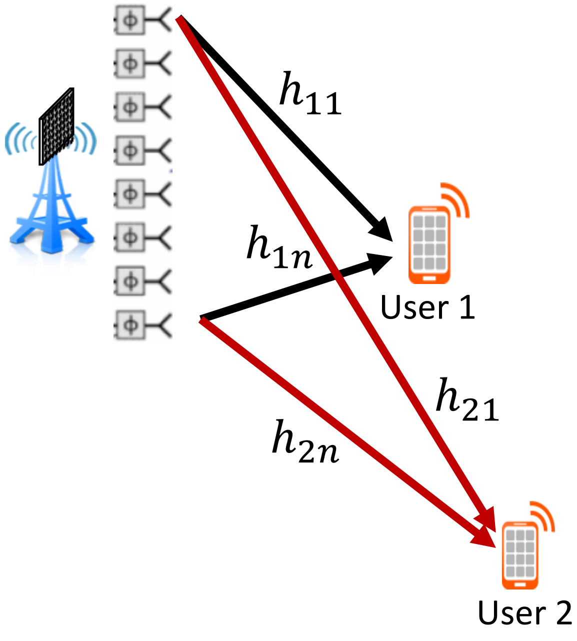
$$H = U \cdot \Lambda \cdot V^*$$

$$W = V$$

Physical meaning:

1. Choose the best transmitting direction
2. Choose the next best direction in the orthogonal complement to the first one
3. etc…

Multi-user transmission



Transmit two symbols to two different users simultaneously

$$y_1 = (h_{11} \dots h_{1n}) \cdot \left(\begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix} \cdot x_1 + \begin{pmatrix} w_1^2 \\ \vdots \\ w_n^2 \end{pmatrix} \cdot x_2 \right) + noise_1$$

$$y_2 = (h_{21} \dots h_{2n}) \cdot \left(\begin{pmatrix} w_1^1 \\ \vdots \\ w_n^1 \end{pmatrix} \cdot x_1 + \begin{pmatrix} w_1^2 \\ \vdots \\ w_n^2 \end{pmatrix} \cdot x_2 \right) + noise_2$$

$$w^k = \begin{pmatrix} w_1^k \\ \vdots \\ w_n^k \end{pmatrix}$$

$w^k \in \mathbb{C}^n$,

$\left\| \left| \sum_k w^k \right| \right\|_{L^\infty} \le p$,

$$y = H \cdot W \cdot x + noise$$

Channel matrix **Precoding matrix**

$$SINR_1(W) = \frac{\left|\langle h_1, w^{1^*} \rangle\right|^2}{\left|\langle h_1, w^{2^*} \rangle\right|^2 + \delta_1^2}$$

$$SINR_2(W) = \frac{\left|\langle h_2, w^{2^*} \rangle\right|^2}{\left|\langle h_2, w^{1^*} \rangle\right|^2 + \delta_2^2}$$

How to choose precoding matrix?

Maximizing weighted sum of spectral efficiency:

$$\sum_{k \in U} \alpha_k \cdot \log(1 + SINR_k) \to \max_W$$

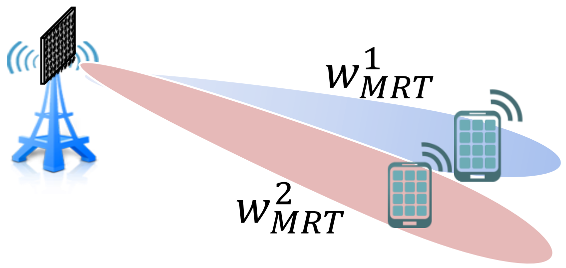
Multi-user beamforming

$$\sum_{u \in U} \alpha_u \cdot \log(1 + SINR_u(W)) \rightarrow \max_W$$

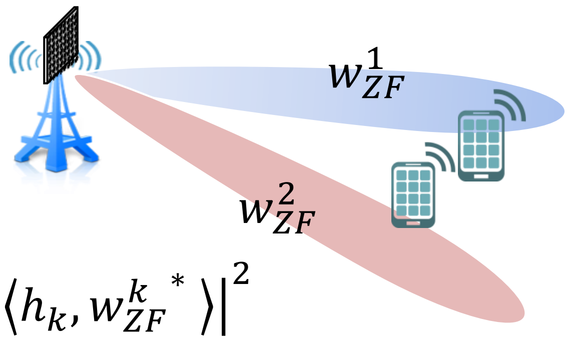
$$SINR_k(W) = \frac{\left|\langle h_k, w^{k^*} \rangle\right|^2}{\sum_{k \neq l}\left|\langle h_k, w^{l^*} \rangle\right|^2 + \delta_1^2}$$

Classic solution #1: Maximum-rate transmission

$$w_{MRT}^k = c_k \cdot h_k^*$$



Classic solution #2: Zero Forcing



$$SINR_k(W_{ZF}) = \frac{\left|\langle h_k, w_{ZF}^{k}{}^* \rangle\right|^2}{\delta_1^2}$$

ZF beam orthogonal to all other users channel vectors:

$$w_{ZF}^k \in < h_1, \dots, h_{k-1,} h_{k+1}, \dots, h_n >^{\perp}$$

w_{ZF}^k maximizes $\left|\langle h_k, w_{ZF}^{k}{}^* \rangle\right|^2$ in this subspace

$\mathrm{W_{ZF}}$ is a pseudo-inverse matrix to H:

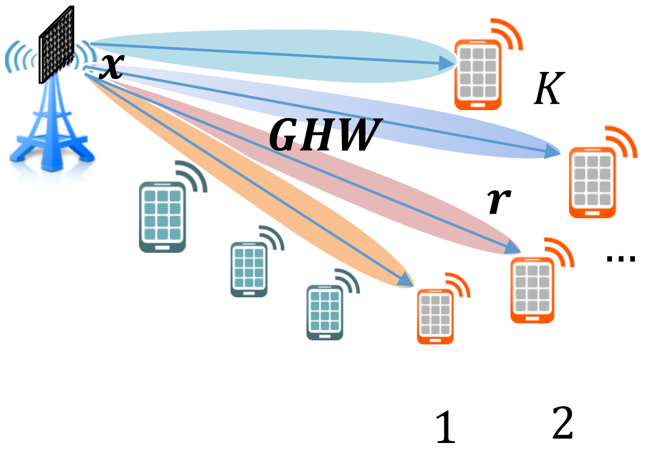
$$\mathrm{W_{ZF}} = H^* \cdot (HH^*)^{-1}$$

Eigen Zero Forcing Theory

Combination of SVD decompositions

User also performs beamformig:

$$\boldsymbol{r} = \boldsymbol{G}(\boldsymbol{H}\boldsymbol{W}\boldsymbol{x} + \boldsymbol{n}) \in \mathbb{C}^{L}$$

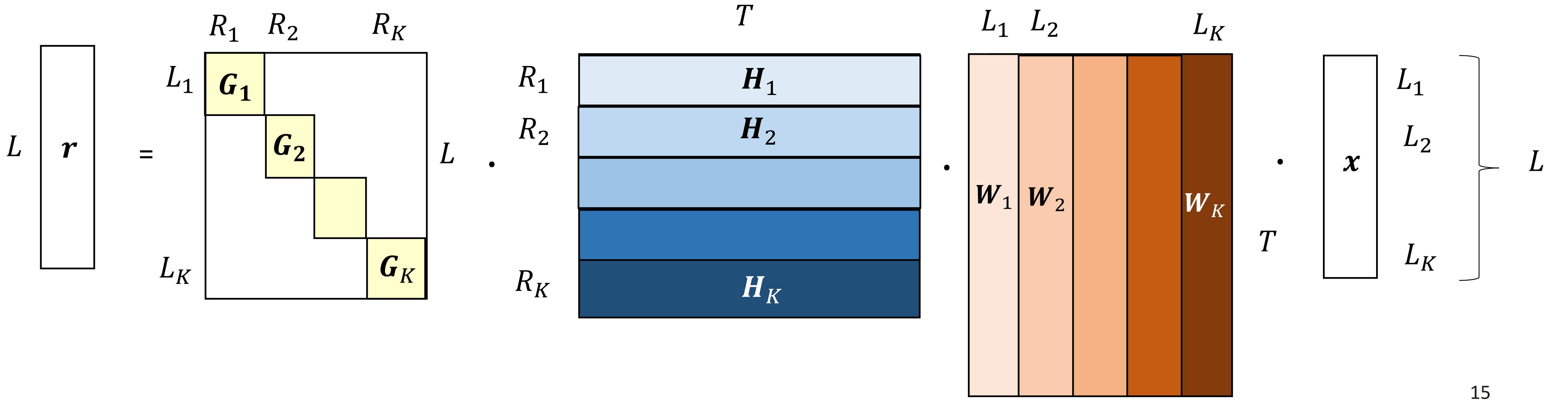


$$\boldsymbol{H} = [\boldsymbol{H}_1 \dots \boldsymbol{H}_K] = \left[\boldsymbol{U}_1^H \boldsymbol{S}_1 \boldsymbol{V}_1 \dots \boldsymbol{U}_K^H \boldsymbol{S}_K \boldsymbol{V}_K\right] =: \boldsymbol{U}^H \boldsymbol{S}\boldsymbol{V}$$

$$\begin{cases} \boldsymbol{H} = \boldsymbol{U}^H \boldsymbol{S}\boldsymbol{V} \\ \boldsymbol{G} = \boldsymbol{S}^{-1}\boldsymbol{U} \end{cases} \rightarrow \boldsymbol{G}\boldsymbol{H}\boldsymbol{W} = \boldsymbol{S}^{-1}\boldsymbol{U}\boldsymbol{U}^H\boldsymbol{S}\boldsymbol{V}\boldsymbol{W} = \boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{V}\boldsymbol{W} = \boxed{\boldsymbol{V}\boldsymbol{W} = \boldsymbol{I}}$$

Theorem. The same holds for $L_i < R_i$

Bobrov, Evgeny, et al. "Adaptive Regularized Zero-Forcing Beamforming in Massive MIMO with Multi-Antenna Users." *arXiv preprint arXiv:2107.00853* (2021).



Gradient-based optimization

Assuming special form of the matrix $\boldsymbol{G}$ we have come to the following approximated function of the $SINR$:

$$SINR_l^C(\boldsymbol{W}, \widetilde{\boldsymbol{v}}_l, s_l, \sigma^2, P) = \frac{|\widetilde{\boldsymbol{v}}_l \boldsymbol{w}_l|^2}{\sum_{i \neq l}^{L} |\widetilde{\boldsymbol{v}}_l \boldsymbol{w}_i|^2 + s_l^{-2} \frac{\sigma^2}{P}}$$

Spectral Efficiency function may be simplified in the following way:

$$SE^C(\boldsymbol{W}, \widetilde{\boldsymbol{V}}, \boldsymbol{S}, \sigma, P) = \sum_{l=1}^{L} \log_2(1 + SINR_l^C(\boldsymbol{W}, \widetilde{\boldsymbol{v}}_l, s_l, \sigma, P)) =$$

$$\sum_{l=1}^{L} \log_2 \left(\sum_{i=1}^{L} |\widetilde{\boldsymbol{v}}_l \boldsymbol{w}_i|^2 + s_l^{-2} \frac{\sigma^2}{P} \right) - \sum_{l=1}^{L} \log_2 \left(\sum_{i \neq l}^{L} |\widetilde{\boldsymbol{v}}_l \boldsymbol{w}_i|^2 + s_l^{-2} \frac{\sigma^2}{P} \right)$$

1. $\widetilde{\boldsymbol{v}}_l \in \mathbb{C}^T$ is the singular vector of the l-th symbol;
2. $\boldsymbol{s}_l \in \mathbb{R}$ is the singular value of the l-th symbol.

Gradient-based optimization

This method explicitly constrains antenna rows using projection:

$$\underset{\boldsymbol{W} \in \mathbb{C}^{T \times L}}{\text{maximize}} \quad SE^C(\text{proj}_{P,T}(\boldsymbol{W}), \widetilde{\boldsymbol{V}}, \boldsymbol{S}, \sigma, P)$$

$$\text{proj}_{P,T}(\boldsymbol{W}) = \begin{cases} \boldsymbol{w}^m, & \|\boldsymbol{w}^m\|^2 \leqslant \frac{P}{T} \\ \frac{\boldsymbol{w}^m}{\|\boldsymbol{w}^m\|}\sqrt{\frac{P}{T}}, & \|\boldsymbol{w}^m\|^2 > \frac{P}{T}, \quad \forall m = 1 \dots T \end{cases}$$

Algorithm 1: On the optimal precoding matrix

Input: Initial precoding matrix $\boldsymbol{W}$, channel singular vectors $\widetilde{\boldsymbol{V}}$, channel singular values $\boldsymbol{S}$, station power P, noise σ^2, iterations N

for $t = 1$ **to** N **do**

  Calculate the gradient: $\nabla SE^C(\text{proj}_{P,T}(\boldsymbol{W}), \widetilde{\boldsymbol{V}}, \boldsymbol{S}, \sigma, P)$;

  Find the optimal direction recursively: $\boldsymbol{D} = \boldsymbol{D}(\nabla SE^C)$;

  Find the optimal step length $\alpha = \arg\max_{\alpha} SE^C(\text{proj}_{P,T}(\boldsymbol{W} + \alpha \boldsymbol{D}))$;

  Make the optimization step: $\boldsymbol{W} \leftarrow \boldsymbol{W} + \alpha \boldsymbol{D}$;

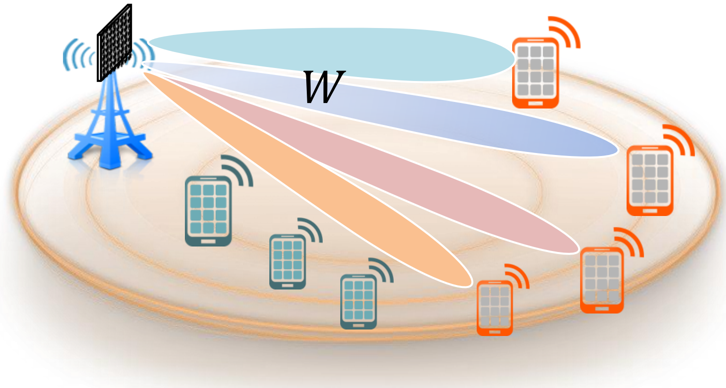
end

return $proj_{P,T}(\boldsymbol{W})$

Quantization justification

Math Modeling Lab

- Increase in antennas
- Increase in layers
- Stricter requirements for service quality



Even linear methods for uplink equalization and downlink precoding appear to be too expensive as dimension grows

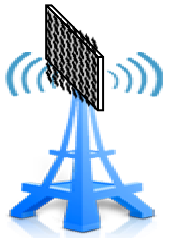
$$F = H^H \left(H \cdot H^H + R_{nn}\right)^{-1} \leftarrow \text{Coarse-grained (per-RB) computations} \rightarrow W = H^H \left(H \cdot H^H + \sigma I\right)^{-1}$$

$$\hat{s} = G \cdot y \leftarrow \text{Fine-grained (per-RE) computations} \rightarrow x = W \cdot s$$

DL MU Precoding

Precoding Matrix Computation in High-Precision values: $V \cdot W = I$	~10% of total complexity
Application of High Precision Precoding Matrix to low-bit vector of symbols: $x = Ws$	~90% of total complexity

Key direction
Reduce complexity of the RE-based computations



Quantization basics

PyTorch → PyTorch

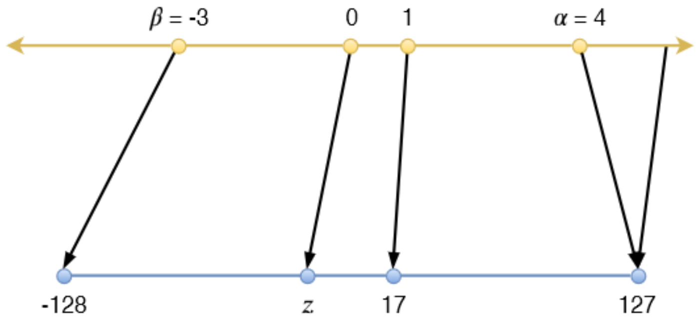
float32 int8

$$x_q = \text{quantize}(x, b, s, z) = \text{clip}(\text{round}(s \cdot x + z), -2^{b-1}, 2^{b-1} - 1)$$

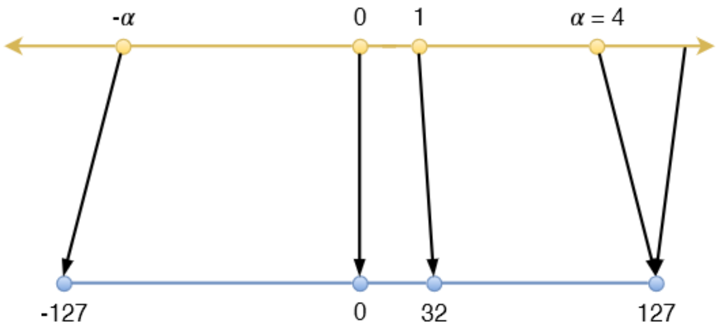
$$\hat{x} = \text{dequantize}(x_q, s, z) = \frac{1}{s}(x_q - z)$$

$$y_{ij} \approx \sum_{k=1}^{p} \frac{1}{s_x}(x_{q,ik} - z_x)\frac{1}{s_{w,j}}(w_{q,kj} - z_{w,j})$$

$$= \frac{1}{s_x s_{w,j}}\left(\underbrace{\sum_{k=1}^{p} x_{q,ik} w_{q,kj}}_{(1)} - \underbrace{\sum_{k=1}^{p}(w_{q,kj} z_x + z_x z_{w,j})}_{(2)} - \underbrace{\sum_{k=1}^{p} x_{q,ik} z_{w,j}}_{(3)}\right)$$



(a) Affine quantization



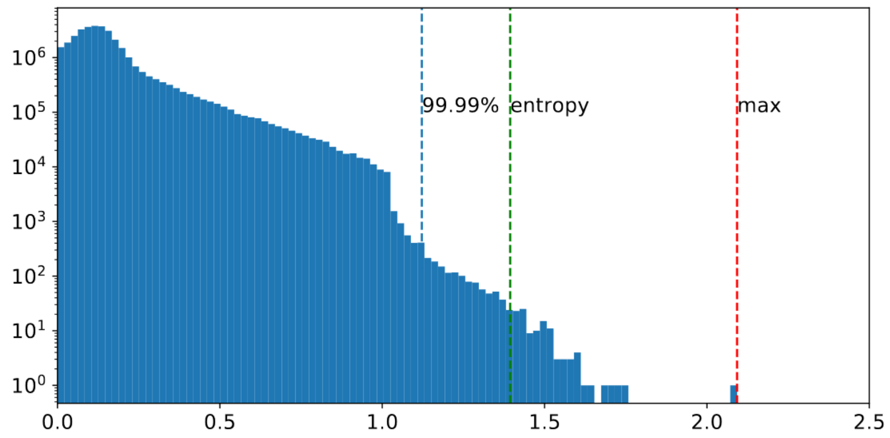
(b) Scale quantization

Quantization approaches

Post-training quantization

$$\sum_{k \in U} L_k \cdot \log(1 + SINR_k^{eff}(W)) \rightarrow \underset{W}{argmax} = W^*$$

$$\left\| W_q - W^* \right\| = \| diag(\beta) \cdot F - W^* \| \rightarrow \min_{\beta, F}$$

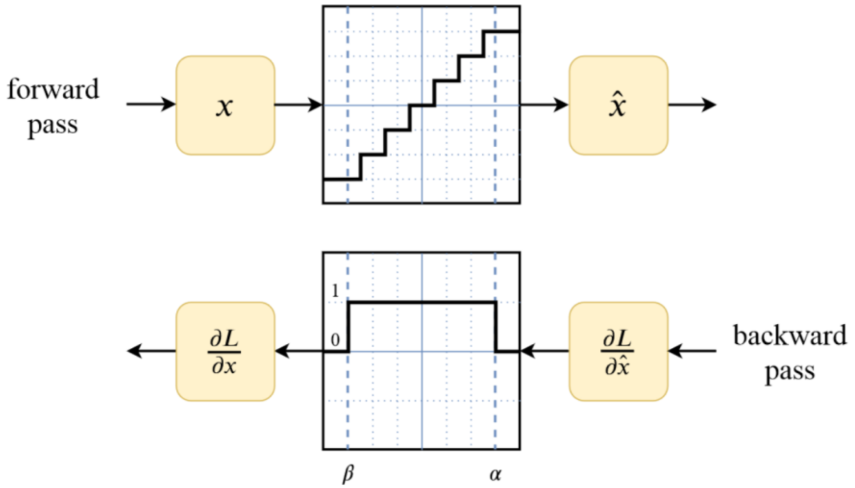


Histogram of input activations to layer 3 in ResNet50 and calibrated ranges

Quantization-aware training

$$\sum_{k \in U} L_k \cdot \log(1 + SINR_k^{eff}(fq(W))) \rightarrow \max_{W \in C^{T \times L}}$$

$$fq(W) = dequantize\big(quantize(W)\big) =$$
$$= \beta \cdot round(clip(\frac{W}{\beta}, -2^p, 2^p - 1)$$

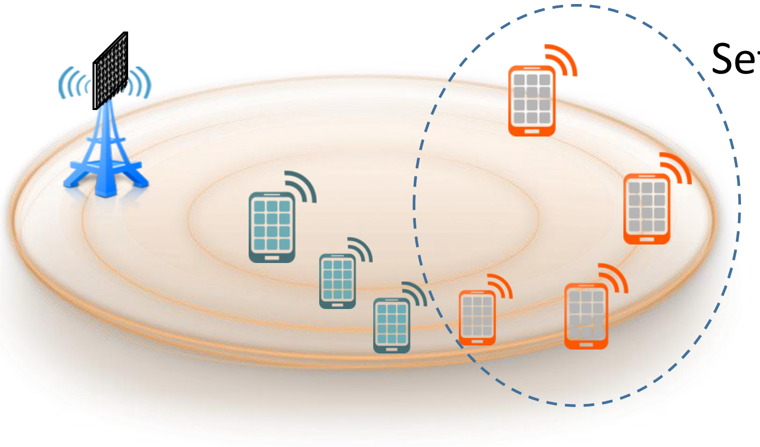


Straight-through gradient estimation

Multi-user pairing

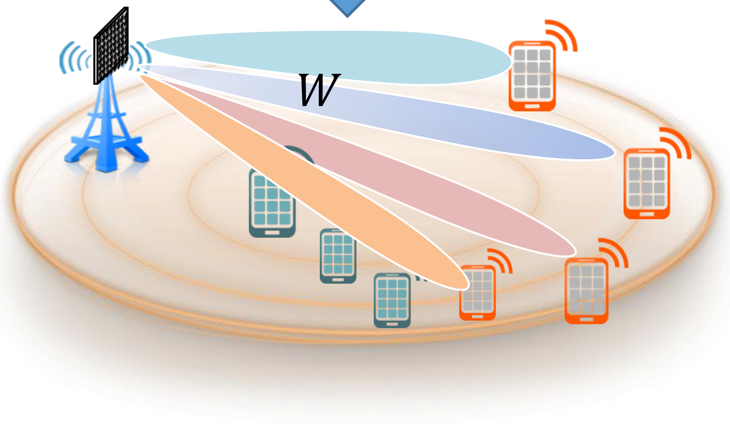
Not necessary to transmit to all active users A

$$\sum_{k \in U} \alpha_k \cdot \log(1 + SINR_k) \rightarrow \max_{U \subseteq A}$$



Set $U \subseteq A$

$$\sum_{k \in U} \alpha_k \cdot \log(1 + SINR_k) \rightarrow \max_{W}$$



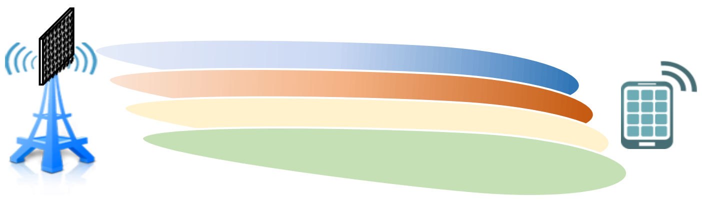
How to select optimal subset $U \subseteq A$ of users for transmission?

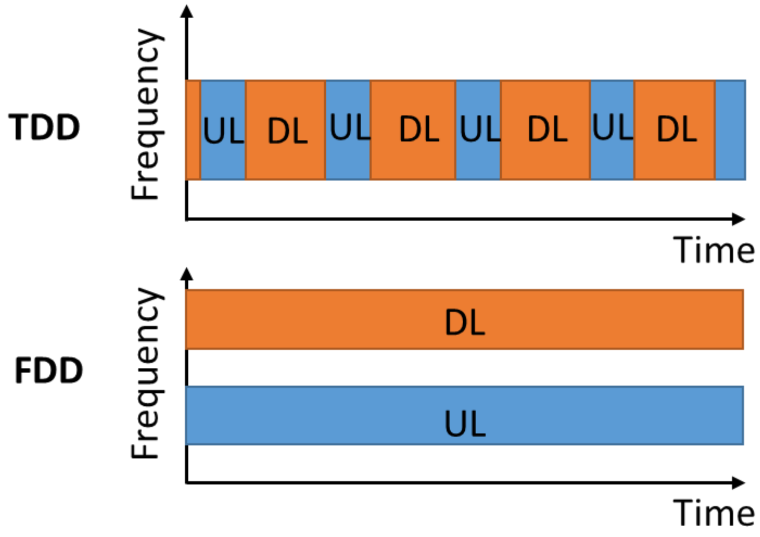
$$\sum_{k \in U} \alpha_k \cdot \log(1 + SINR_k(W)) \rightarrow \max_{\substack{U \subseteq A \\ W}}$$

Joint user selection and beamforming?

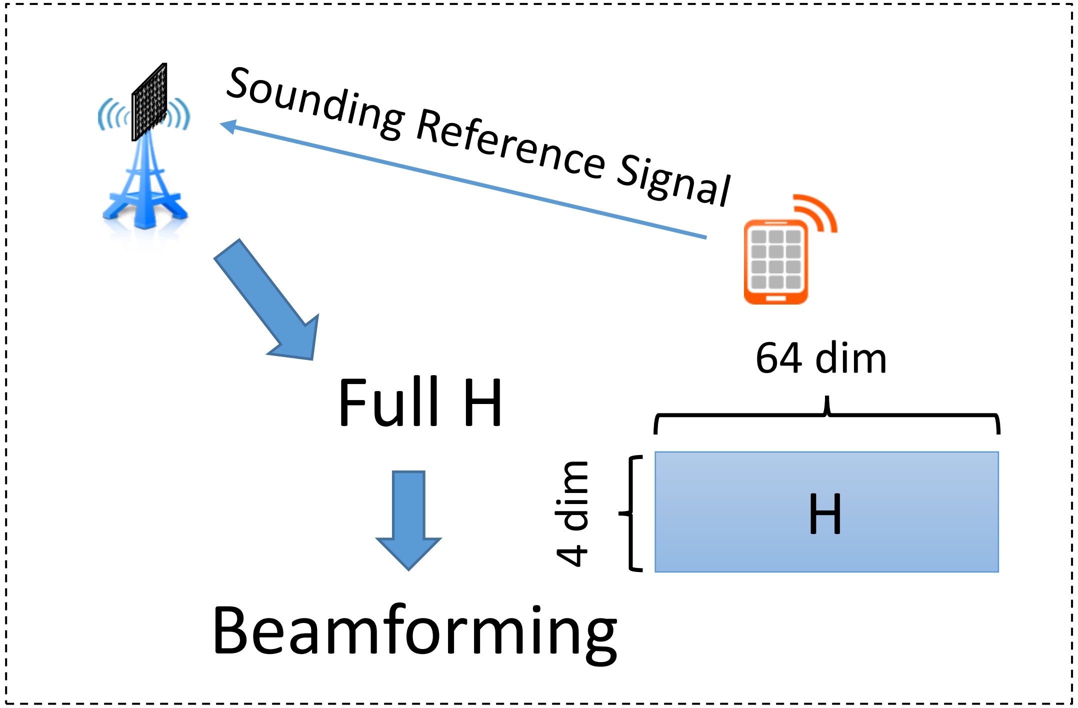
Channel Reconstruction

If we know channel matrix, we may speed up transmission several times:

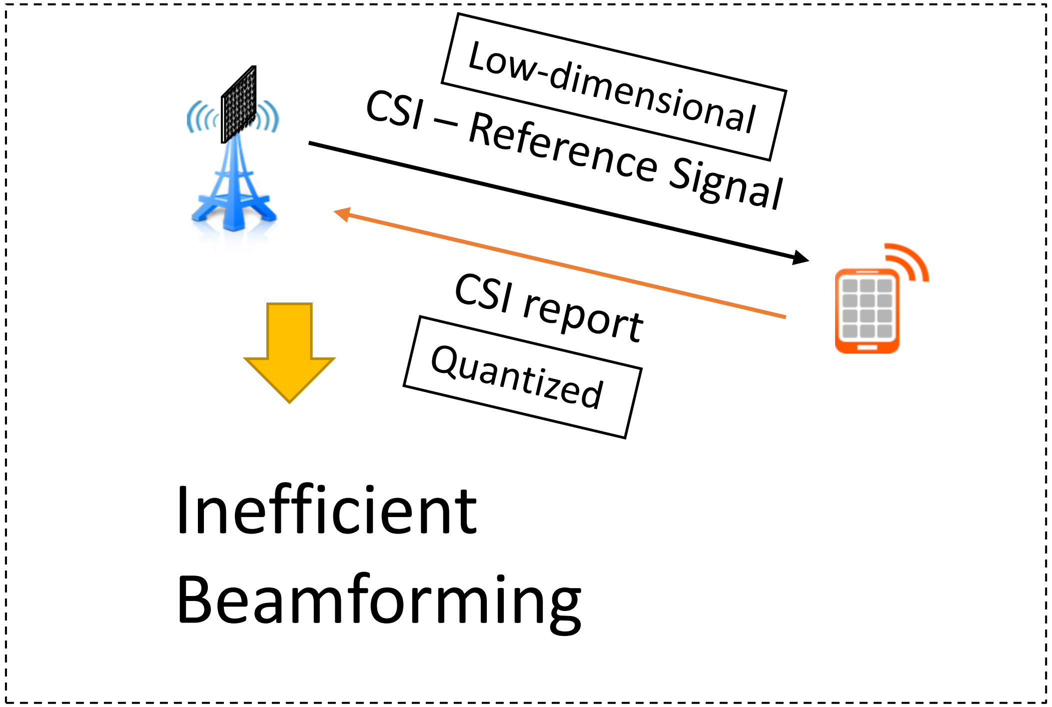




TDD: Uplink Measurement for Downlink Beamforming



FDD: Downlink Measurement and/or Channel Reconstruction



Link adaption problem

MCS – Modulation and Coding Scheme

29 options for MCS

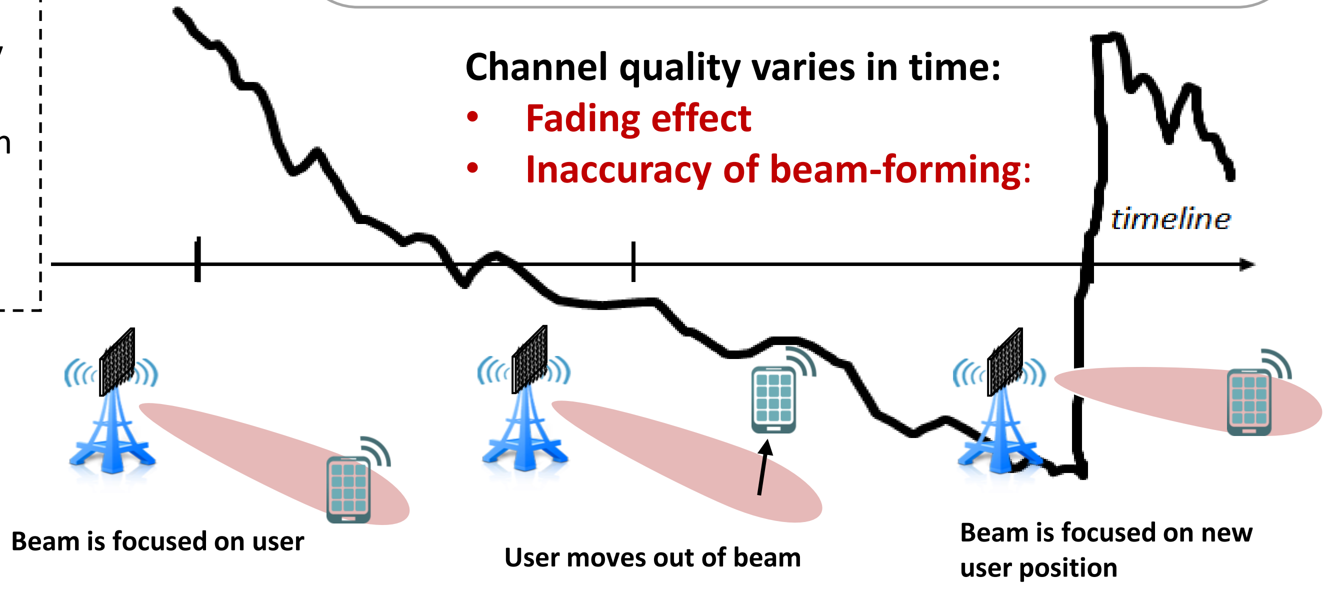
MCS	Modulation	Useful Data 100 PRBs	Total Data 100 PRBs	Code Rate
0	QPSK	2,792	30,000	0.094
1	QPSK	3,624	30,000	0.122
2	QPSK	4,584	30,000	0.154
3	QPSK	5,736	30,000	0.192
4	QPSK	7,224	30,000	0.242
5	QPSK	8,761	30,000	0.293
6	QPSK	10,296	30,000	0.344
7	QPSK	12,216	30,000	0.408
8	QPSK	14,112	30,000	0.471
9	QPSK	15,840	30,000	0.529
10	16 QAM	15,840	60,000	0.264
11	16 QAM	17,658	60,000	0.295
12	16 QAM	19,848	60,000	0.331
13	16 QAM	22,920	60,000	0.382
14	16 QAM	25,456	60,000	0.425
15	16 QAM	28,336	60,000	0.473
16	16 QAM	30,576	60,000	0.510
17	64QAM	30,576	90,000	0.340
18	64QAM	32,856	90,000	0.365
19	64QAM	36,696	90,000	0.408
20	64QAM	39,232	90,000	0.436
21	64QAM	43,816	90,000	0.487
22	64QAM	46,888	90,000	0.521
23	64QAM	51,024	90,000	0.567
24	64QAM	55,056	90,000	0.612
25	64QAM	57,336	90,000	0.637
26	64QAM	61,664	90,000	0.685
27	64QAM	63,776	90,000	0.709
28	64QAM	75,326	90,000	0.837

Modulation: how many bits are encoded into a complex number (4 / 16 / 64)

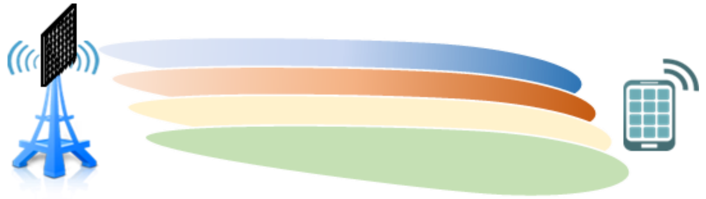
Coding: how many parity bits are used in an error correcting code

MCS should correspond to channel quality.
Higher MCS values
- **allow more information to be transmitted**
- **increase risk of unsuccessful transmission! Retransmissions are painful.**

Channel quality varies in time:
- **Fading effect**
- **Inaccuracy of beam-forming:**



Which number of layers is optimal for transmission?



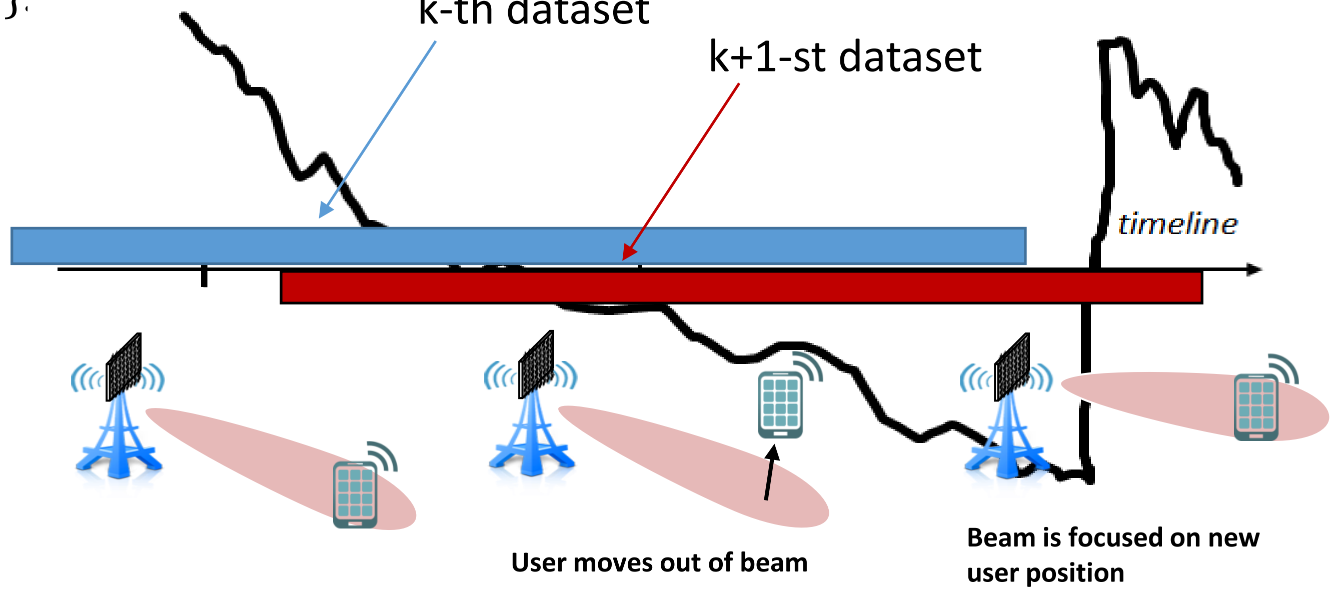
Which MCS is optimal for transmission?

Neural network MCS estimation

$$\widehat{mcs}(state) = \arg\max_{mcs} \{p_w(ack|mcs, state)SE(mcs)\}.$$

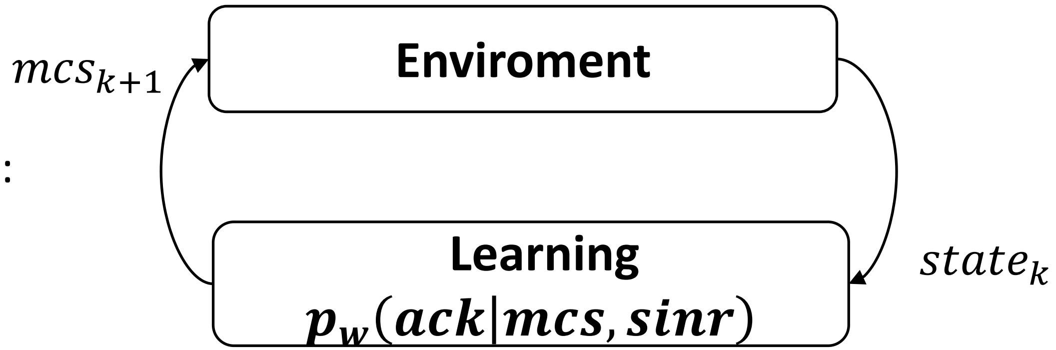
$$\hat{p}_{ack}(state;\theta) \approx p_w(ack|mcs, state)$$

state: SINR, UE CSI, etc.

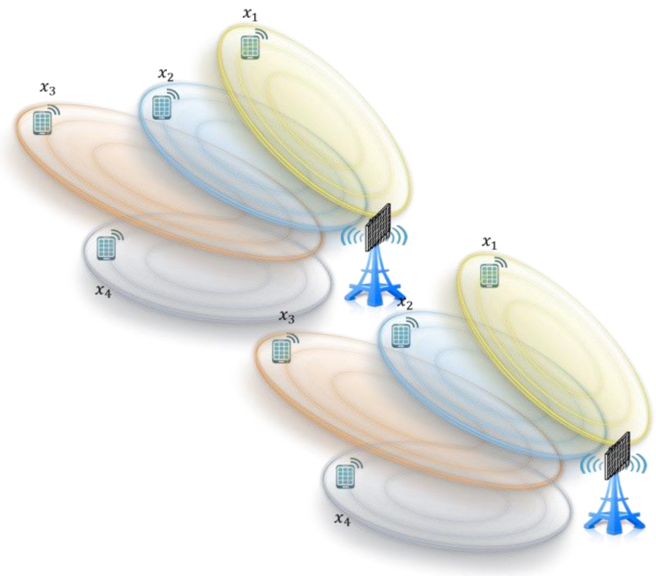


Online learning is essential:

But it is not RL!



Uplink Equalization Problem



$$y = H \cdot x + r$$

r **– random noise and interference**
H – channel matrix
x – transmitted signal (dimension = number of users)
y – received signal (dimension = number of base station antennas)

$$\hat{x} = G \cdot H \cdot x + G \cdot r$$

MMSE formulation

$$\boldsymbol{E}_{r,x}\left(||G \cdot y - x||_2^2\right) \to \min_G$$

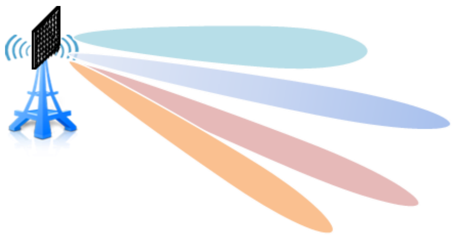
$$G_{MMSE} = H^* \cdot (H \cdot H^* + R_r)^{-1}$$

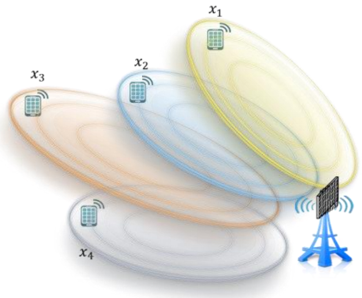
Noise + Interference Covariance matrix

Difference with precoding:

- Applied after interference and noise
- Can be nonlinear
- Can be combined with demodulation/decoding

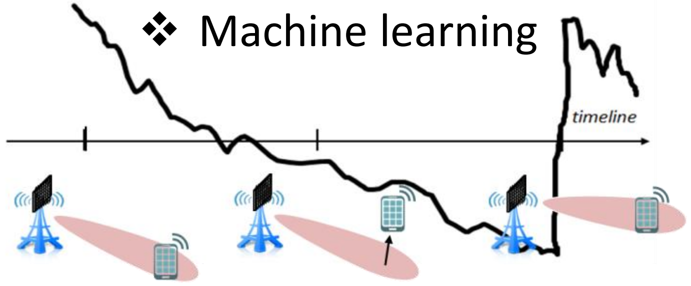
Beamforming problem
- Non-convex optimization
- Complex analysis
- Probability theory

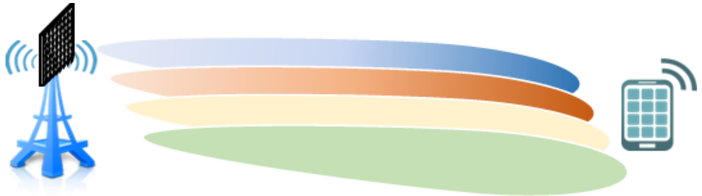




Equalization problem
- Statistical estimations
- Probability theory
- Statistical physics

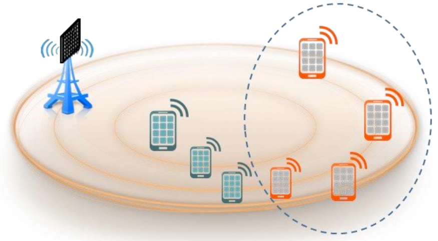
Link adaption problem
- Information theory
- Dynamic control theory
- Machine learning

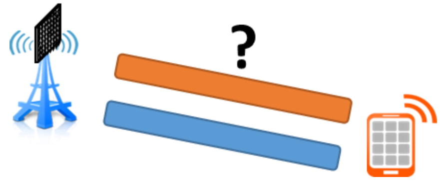




Multi-stream transmission problem
- Information theory
- Computational linear algebra
- Probability theory

User pairing problem
- Combinatorial optimization
- Submodular optimization
- Graph theory





Channel Reconstruction problem
- Non-convex optimization
- Complex analysis
- Stochastic process

Thank you!