Unit 9. Introduction to MIMO

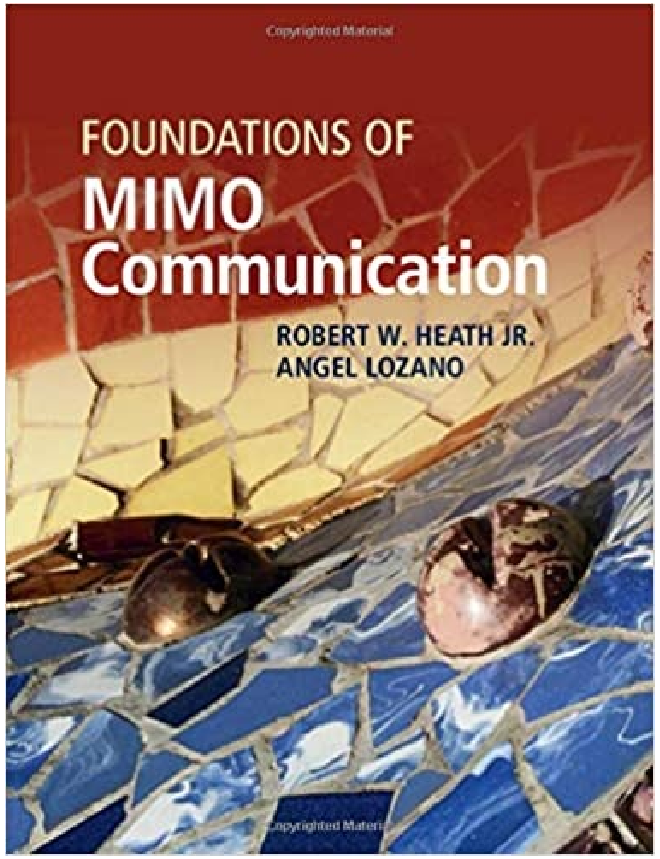
EL-GY 6023. WIRELESS COMMUNICATIONS

PROF. SUNDEEP RANGAN

Learning Objectives

- Assuming CSI-R and CSI-T, describe the diagonalization of a MIMO channel
  - Compute the virtual directions and their SNRs
- Compute the capacity for a MIMO channel using diagonalization
  - Narrowband and wideband
- Mathematically formulate the power allocation problem and find optimal power allocations
- Describe linear receivers, identify the main blocks and compute their capacity
  - Zero forcing and LMMSE
- Describe reference signals for MIMO channel estimation in 4G and 5G systems
- Compute optimal statistical pre-coders and compute the capacity

Excellent Text for This Section



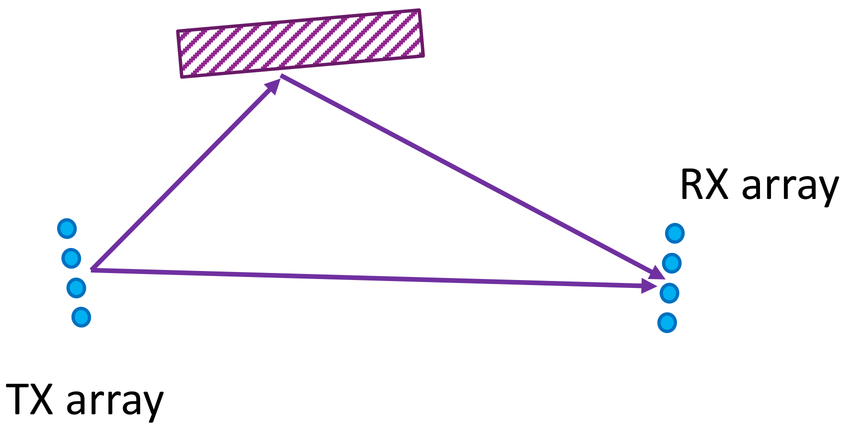
- Some material in this section is from this recent text
- Provides excellent:
  - Information theoretic background
  - Practical guidelines for implementation
  - Up-to-date examples with issues for mmWave
- We only cover a small section
  - Single user MIMO
  - Many derivations are left for the text

Outline

- Spatial Multiplexing with CSI-T and CSI-R
- Power Allocation and Rank Selection
- Spatial Multiplexing with CSI-R Only
- Channel Estimation and CSI-R
- CSI-T Feedback and Statistical Pre-Coding

Spatial Multiplexing

- Many environments have multiple spatial paths
  - LOS, reflections, diffraction, …
- Spatial multiplexing concept
  - Transmit separate information streams on different paths
- Increases degrees of freedom
- Requires:
  - Channel rank $r \geq K$ where K is the number of streams
  - In particular, $N_r, N_t \geq K$
  - Also need sufficient power for the K streams



MIMO History

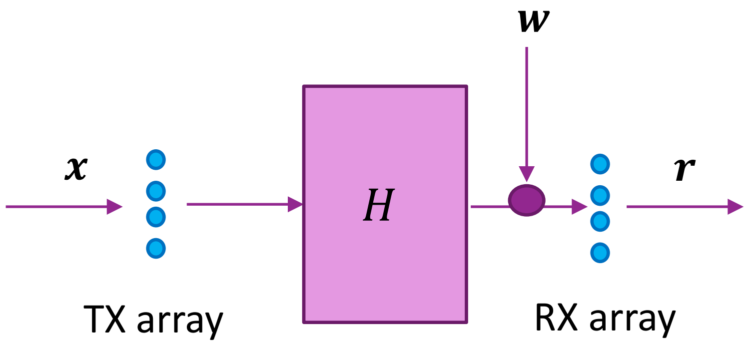
- Early research:
  - ArrayComm 1991
  - Paulraj and Kailath, initial patent on SDMA
  - Foschini and others, initial capacity estimates, 1996
  - Bell Labs prototype, 1998
- Commercialization in LANs
  - Began study in 2003.
  - First appeared in 802.11n 2009
- Commercialization in cellular:
  - 4G systems, approximately 2004
  - 5G systems: Integral component for support for mmWave

MIMO Narrowband Capacity

- Consider narrowband MIMO channel from previous lecture:

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \qquad \boldsymbol{w} \sim CN(0, N_0 \boldsymbol{I})$$

  - $\boldsymbol{H} \in \mathbb{C}^{N_r \times N_t}$ Channel matrix
  - $\boldsymbol{x} = \left(x_1, \ldots, x_{N_t}\right)^T$: signals to the TX antennas



- TX power constraint: $\|\boldsymbol{x}\|^2 \le E_x$
  - Total energy constraint on all antennas

- This section: make two critical assumption:
  - TX and RX knows $\boldsymbol{H}$ and N_0 exactly (called CSI-T and CSI-R)
  - We will relax these later

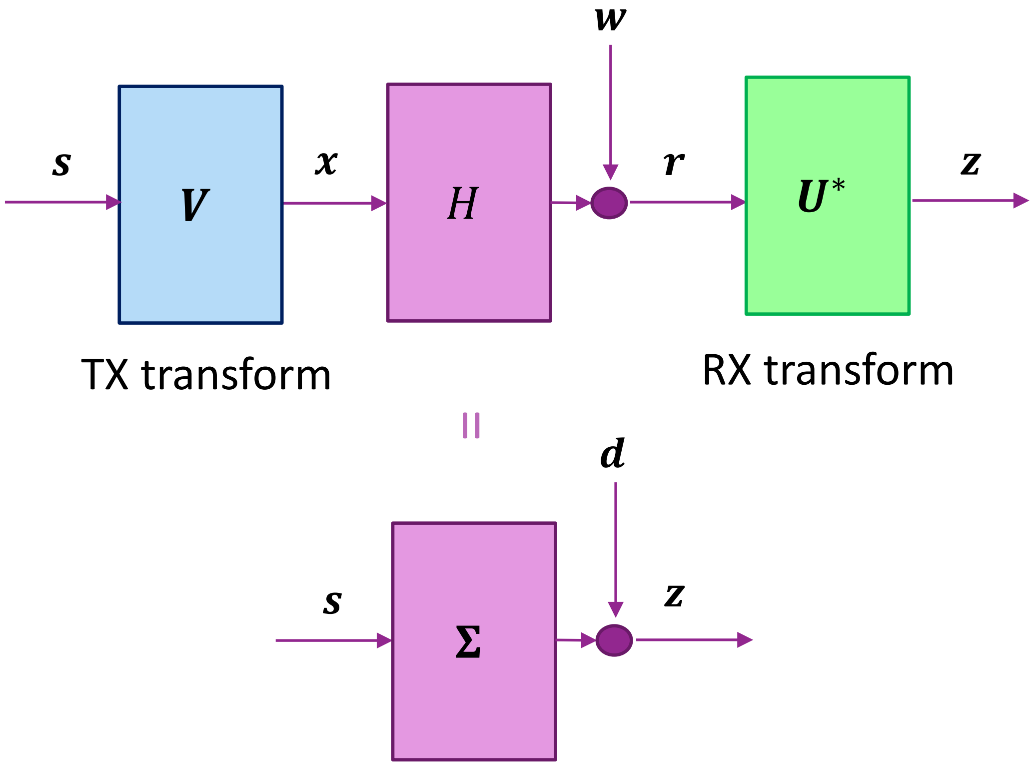
Applying Transforms with the SVD

- Take reduced SVD of the channel: $\boldsymbol{H} = \boldsymbol{U\Sigma V}^*$
  - $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$
  - $r = rank(\boldsymbol{H})$
- TX and RX apply transforms
  - TX transform: $\boldsymbol{x} = \boldsymbol{Vs}$ (also called a pre-coder)
  - RX transform: $\boldsymbol{z} = \boldsymbol{U}^*\boldsymbol{r}$
- Theorem: The channel from s to z is diagonal:

$$\boxed{\boldsymbol{z} = \boldsymbol{\Sigma s} + \boldsymbol{d}, \qquad \boldsymbol{d} \sim CN(0, N_0\boldsymbol{I}_r)}$$

- Creates r independent channels:

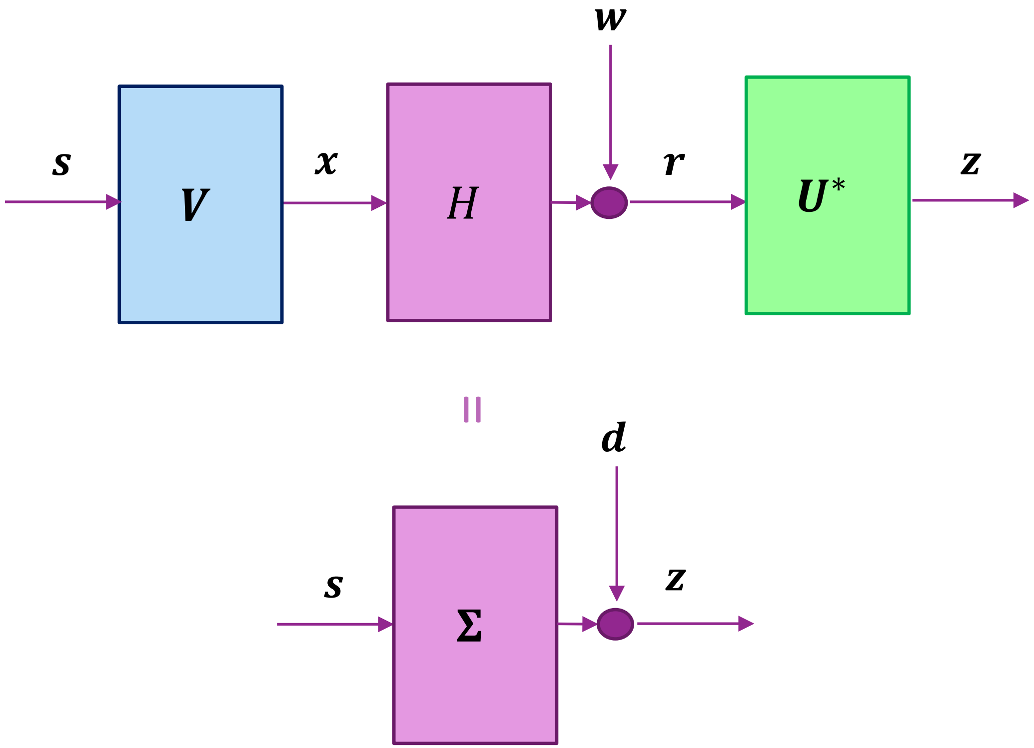
$$z_i = \sigma_i s_i + d_i, \qquad d_i \sim CN(0, N_0)$$



Proof of the Diagonalization

- Consider channel from $\boldsymbol{s}$ to $\boldsymbol{z}$:
  - $\boldsymbol{z} = \boldsymbol{U}^*\boldsymbol{r} = \boldsymbol{U}^*\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}\boldsymbol{x} + \boldsymbol{U}^*\boldsymbol{w} = \boldsymbol{\Sigma}\mathbf{s} + \mathbf{d}$
- Noise:
  - Since $\boldsymbol{w} \sim CN(0, N_0\boldsymbol{I})$ and $\boldsymbol{d} = \boldsymbol{U}^*\boldsymbol{w}$, $\boldsymbol{d}$ is also Gaussian
  - $E(\boldsymbol{d}) = \boldsymbol{U}^* E(\boldsymbol{w}) = \mathbf{0}$
  - $var(\boldsymbol{d}) = \boldsymbol{U}^* var(\mathbf{w})\boldsymbol{U} = N_0\boldsymbol{U}^*\boldsymbol{U} = N_0\boldsymbol{I}_r$
- Hence, transforms diagonalize the channel:

$$\boldsymbol{z} = \boldsymbol{\Sigma}\boldsymbol{s} + \boldsymbol{d}, \qquad \boldsymbol{d} \sim CN(0, N_0\boldsymbol{I}_r)$$

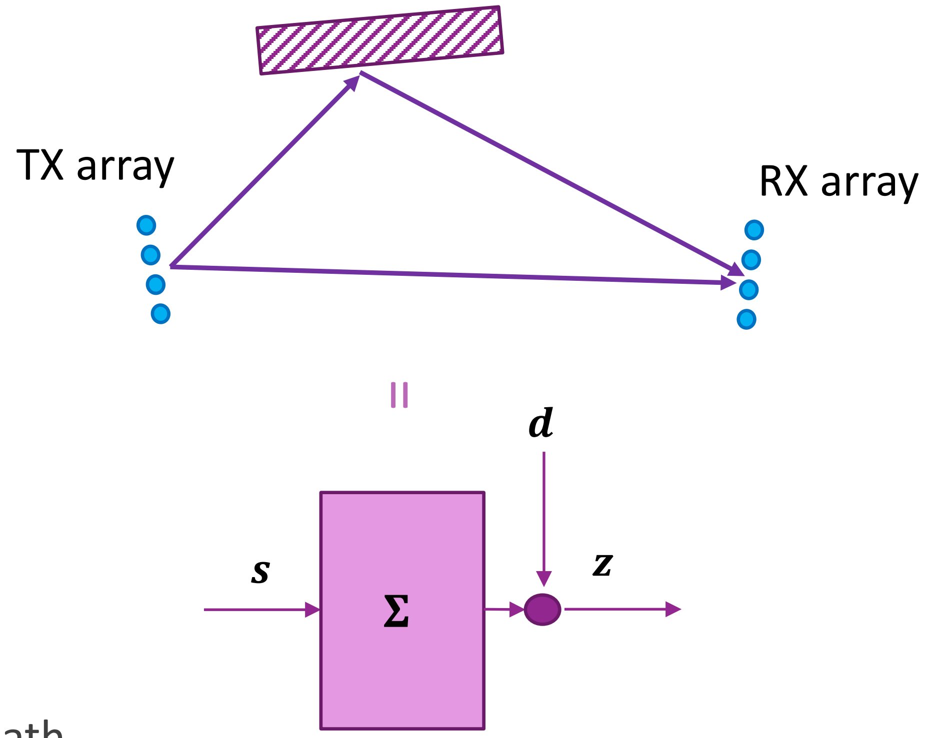


Virtual Channels

- Diagonalizing the channel creates r virtual channels

$$z_i = \sigma_i s_i + d_i, \qquad d_i \sim CN(0, N_0)$$

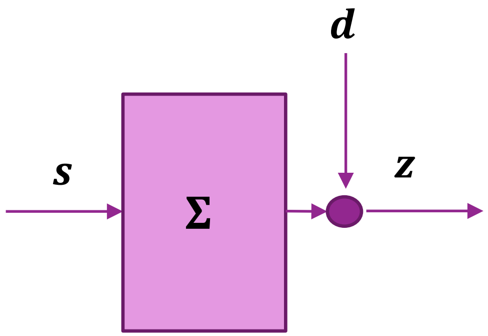
- Number of virtual channels = $rank(H)$
  - = Number of orthogonal paths in the environments
- Correspond loosely to the physical paths
  - Suppose spatial signature of each physical path is orthogonal
  - In this case, directions of virtual channel = direction of physical path



Shannon Capacity

- With diagonalizing transform:

$$z_i = \sigma_i s_i + d_i, \qquad d_i \sim CN(0, N_0)$$



- Assume TX allocates power uniformly across all r virtual channels
  - Each channel gets $E|s_i|^2 = \frac{E_x}{r}$ energy per symbol
  - This is not optimal. We will look at improved allocations laer
- Total capacity (bits per degree of freedom)

$$C = \sum_{i=1}^{r} \log_2\left(1 + \sigma_i^2 \frac{E_x}{rN_0}\right) = \sum_{i=1}^{r} \log_2\left(1 + \frac{\gamma_i}{r}\right)$$

  - $\lambda_i = \sigma_i^2 =$ eigenvalues of $\boldsymbol{H}^*\boldsymbol{H} =$ eigenvalues of $\boldsymbol{H}^*\boldsymbol{H}$
  - $\gamma_i = \frac{\lambda_i E_x}{N_0} =$ SNR on virtual path i

Log-Det Form of the Shannon Capacity

- The Shannon capacity is commonly written in an alternate form:

$$C = \log_2 \det\left(\boldsymbol{I} + \frac{E_x}{rN_0} \boldsymbol{H}^* \boldsymbol{H} \right)$$

- Proof:
  - Take eigenvalue decomposition: $\boldsymbol{H}^*\boldsymbol{H} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^*$, $\boldsymbol{D} = diag(\lambda_1, \dots, \lambda_{N_t})$
  - Let $\alpha = \frac{E_x}{rN_0}$
  - $det(\boldsymbol{I} + \alpha \boldsymbol{H}^*\boldsymbol{H}) = \det\left[\boldsymbol{V} diag\left(1 + \alpha\lambda_1, \dots, 1 + \alpha\lambda_{N_t}\right)\boldsymbol{V}^*\right] = \det diag\left(1 + \alpha\lambda_1, \dots, 1 + \alpha\lambda_{N_t}\right)$ $= \prod(1 + \alpha\lambda_i)$
  - Hence $\log_2 \det(\boldsymbol{I} + \alpha \boldsymbol{H}^*\boldsymbol{H}) = \sum \log(1 + \alpha\lambda_i)$
  - But this is the capacity from the previous slide

SNR Per Antenna

- To understand benefit of spatial multiplexing, compare to a SISO system
- Channel from TX j to RX i has SNR $\frac{|H_{ij}|^2 E_x}{N_0}$
- Definition: The SNR per antenna is the average single antenna SNRs:

$$\boxed{\gamma_0 = \frac{1}{N_r N_t} \sum_{ij} |H_{ij}|^2 \frac{E_x}{N_0} = \frac{1}{N_r N_t} \|\boldsymbol{H}\|_F^2 \frac{E_x}{N_0}}$$

  - $\|\boldsymbol{H}\|_F^2 = \sum_{ij} |H_{ij}|^2$ =“Frobenius” norm of H

Frobenius Norm=Sum of Eigenvalues

- Key property: The Frobenius norm $\|\boldsymbol{H}\|_F^2 = \sum_i \lambda_i$ where λ_i = eigenvalues of $\boldsymbol{Q} = \boldsymbol{H}^*\boldsymbol{H}$
- Proof:
  - Diagonal entries of $\boldsymbol{Q}$: $Q_{ii} = \sum_j \left|H_{ij}\right|^2$
  - Hence, $\|\boldsymbol{H}\|_F^2 = \sum_i Q_{ii} = Tr(\boldsymbol{Q})$ = "Trace" = sum of diagonals
  - Property of trace: $Tr(\boldsymbol{AB}) = Tr(\boldsymbol{BA})$
  - Take digonalization: $\boldsymbol{Q} = \boldsymbol{VDV}^*$, $\boldsymbol{D} = diag(\lambda_1, \ldots, \lambda_{N_t})$
  - Therefore: $\|\boldsymbol{H}\|_F^2 = Tr(\boldsymbol{Q}) = Tr(\boldsymbol{VDV}^*) = Tr(\boldsymbol{V}^*\boldsymbol{VD}) = Tr(\boldsymbol{D}) = \sum \lambda_i$
- Hence, SNR per antenna is sum of SNR per virtual path divided by $N_r N_t$

$$\gamma_0 = \frac{1}{N_r N_t} \|\boldsymbol{H}\|_F^2 \frac{E_x}{N_0} = \frac{1}{N_r N_t} \sum_{i=1}^{r} \lambda_i \frac{E_x}{N_0} = \frac{1}{N_r N_t} \sum_{i=1}^{r} \gamma_i$$

SNR Per Antenna and Path Gains

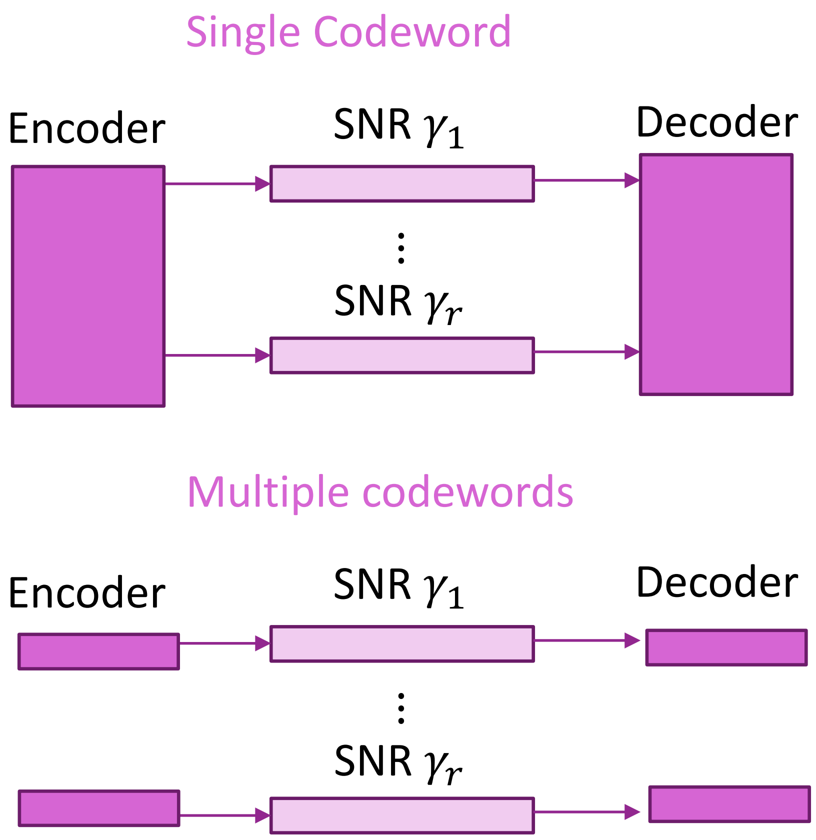
- Consider channel with L paths $\boldsymbol{H} = \sum_{\ell=1}^{L} g_\ell e^{i\theta_\ell} \, \boldsymbol{u}_r(\Omega_\ell^r)\boldsymbol{u}_t^T(\Omega_\ell^t)$
- Assume:
  - Phases θ_ℓ are uniform in $[0,2\pi]$ since they vary with time and frequency
  - $\|\boldsymbol{u}_r(\Omega_\ell^r)\|^2 = N_r$ and $\left\|\boldsymbol{u}_t\left(\Omega_\ell^t\right)\right\|^2 = N_t$ (spatial signatures only include phase rotations)
  - Element gains are included in the complex gains
- Theorem: Taking average over phases

$$E(\gamma_0) = \frac{E_x}{N_0}\frac{1}{N_r N_t} E\|\boldsymbol{H}\|_F^2 = \frac{E_x}{N_0}\sum_{\ell=1}^{L} |g_\ell|^2$$

- Conclusion: SNR per antenna = sum of SNRs of each path without beamforming gain

Coding Architectures

- Spatial multiplexing creates r virtual channels
  - Each channel has SNR $\gamma_i = \frac{E_x}{N_0 r}\lambda_i$
- Two possible transmission methods
- Single codeword:
  - Encode bits for rN symbols into one codeword
  - Codewords sees varying SNR across symbols
  - Adjust MCS for ergodic capacity
- Multiple codewords:
  - Divide bits into r streams
  - In each stream, encode bits for N symbols into a codeword
  - Each codewords sees a constant SNR
  - Set MCS for each codeword to match SNR

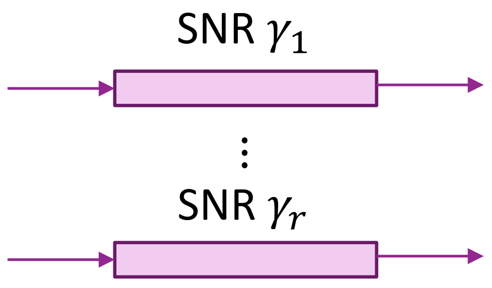


Capacity with Practical Codes

- Shannon capacity is $C = \sum_{i=1}^{r} \log_2(1 + \frac{\gamma_i}{r})$
  - Bits per channel use
  - Can be achieved with optimal single or multiple codeword method
- To account for practical codes, usually assume a model:

$$C = \sum_{i=1}^{r} R\left(\frac{\gamma_i}{r}\right) = \sum_{i=1}^{r} \min\{\rho_{max}, \alpha \log_2\left(1 + \beta \frac{\gamma_i}{r}\right)\}$$

  - ρ_{max} = Max spectral efficiency, based on modulation
  - α = bandwidth loss
  - β = SNR loss
- Typical values for cellular systems: $\rho_{max} = 4.5, \alpha = 0.6, \beta = 1$



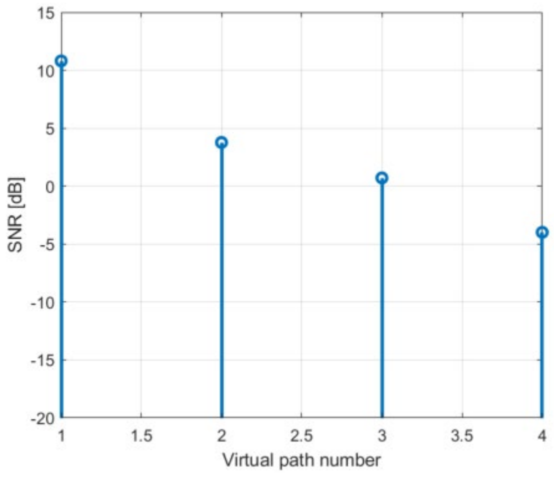
Examples

- We will illustrate the calculations in this unit in two cases

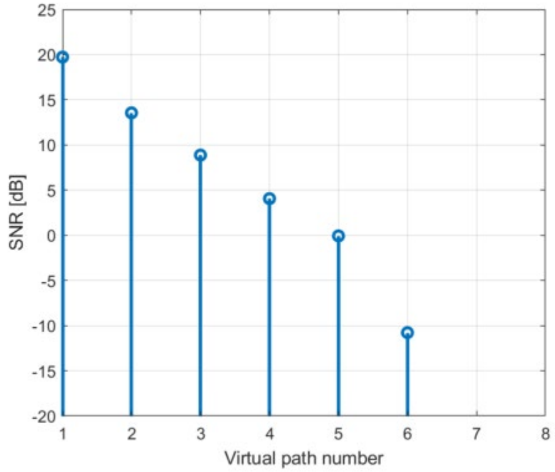
Parameter	High-Dim Array	Low-Dim Array	Remark
Carrier f_c	28 GHz	2.3 GHz	
TX Array	4x4 URA	4x1 ULA	Typical gNB
RX Array	8x1 ULA	4x1 ULA	Typical UE
Num paths	20		
Relative path gains	Exponential, Mean = 10 dB		
AoA Az	Unif[-60,60]	Unif[-180,180]	Low-dim case has rich scattering
AoD Az	Unif[-30,30]	Unif[-180,180]	
AoA and AoD El	Unif[-20,20]	Unif[-20,20]	

Eigenvalue Distribution

- Plot: SNR per virtual path γ_i
  - Channel matrix normalized to SNR per antenna = 0 dB
  - Max gain in any one path is $N_r N_t$
- Low-dim case
  - Up to 4 paths
  - Eigenvalues ≈ evenly spread due to rich scattering
- High-dim case
  - Up to 8 paths
  - Eigenvalues concentrated in a few dominant paths



Low-dim array

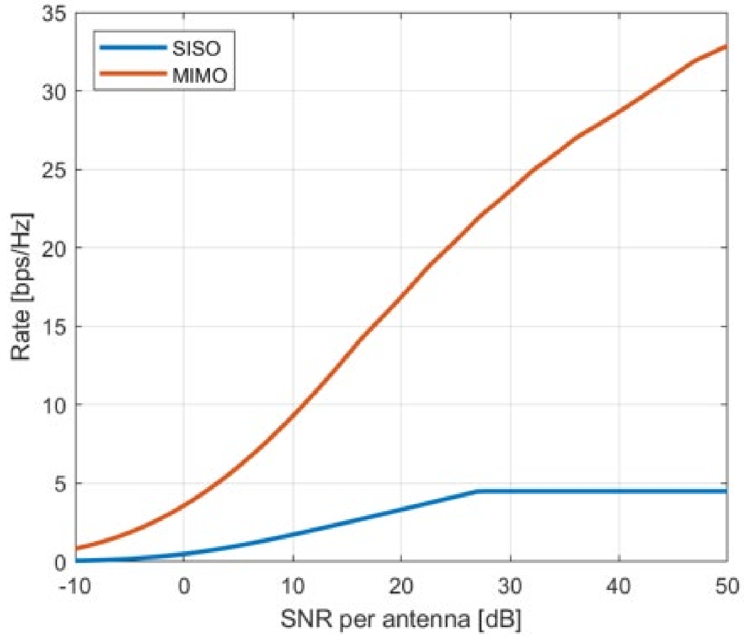


High-dim array

Rate vs SNR

- For each SNR per antenna:
  - Rescale channel matrix
  - Compute rate assuming uniform distribution across all streams
  - Assume rate per stream: $\mathrm{R(s)} = \min\{\rho_{max}, \alpha \log_2(1+\beta s)\}$
  - $\rho_{max} = 4.5, \alpha = 0.6, \beta = 1$
- Also plotted:
  - SISO rate with SNR per antenna
- See significant possible gain
  - But not a fair comparison
  - Should compare against beamforming

High-dim array

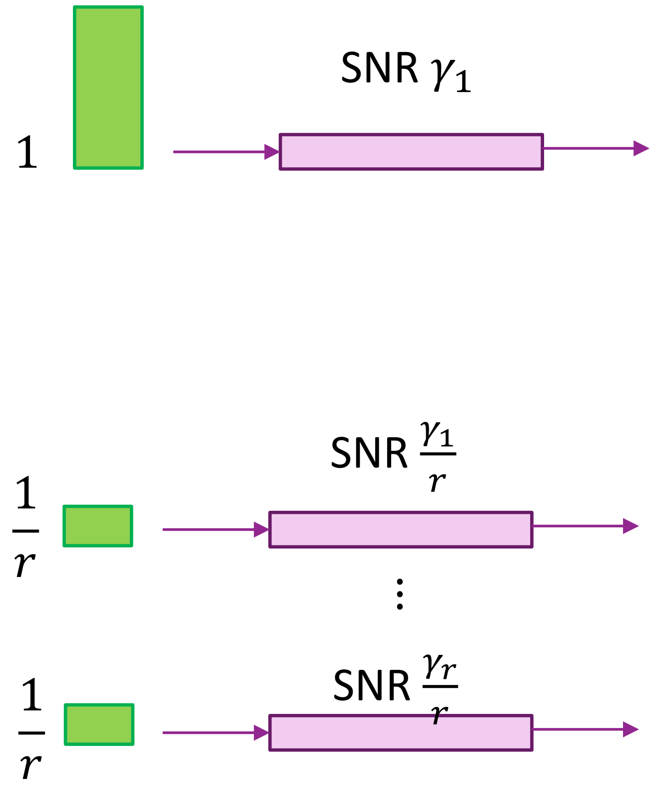


Outline

- ❑Spatial Multiplexing with CSI-T and CSI-R
- ❑Power Allocation and Rank Selection
- ❑Spatial Multiplexing with CSI-R Only
- ❑Channel Estimation and CSI-R
- ❑CSI-T Feedback and Statistical Pre-Coding

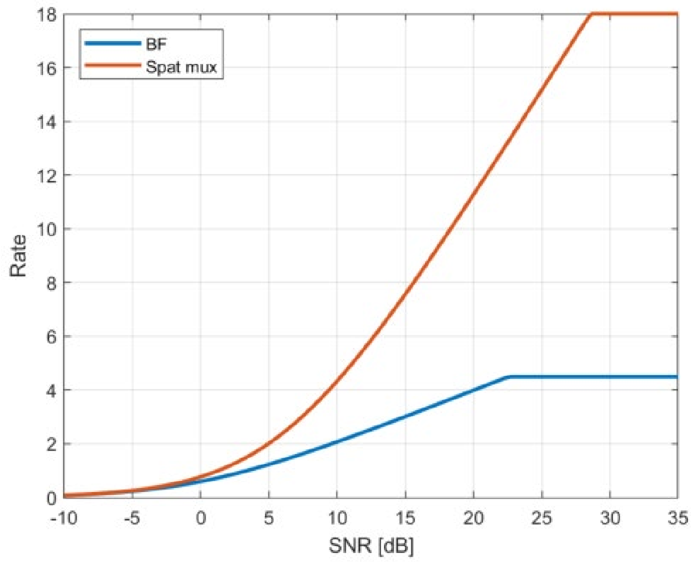
Spatial Multiplexing vs Beamforming

- Beamforming:
  - Places all energy on virtual path with strongest SNR
  - Achieves rate $R_{BF} = R(\gamma_1)$
  - If codes achieve capacity, $R_{BF} = \log_2(1 + \gamma_1)$
- Spatial multiplexing:
  - Transmit energy evenly on r virtual paths
  - $R_{SM} = \sum_i R(\frac{\gamma_i}{r}) = \sum_i \log_2(1 + \frac{\gamma_i}{r})$
- To compare, consider two extreme cases:
  - Equal SNRs $\gamma_i = \gamma$ for all i
  - Single dominant SNR γ_1



Spatial Multiplexing vs Beamforming
Equal SNR streams

- First, suppose all virtual directions have same power
  - $\gamma_i = \gamma, \ i = 1, \dots, r$
  - BF rate: $R_{BF} = R(\gamma)$
  - Spatial mux rate $R_{SM} = rR\left(\frac{\gamma}{r}\right)$
- Low SNR regime (power limited)
  - $R_{SM} \approx R_{BF}$. No gain
- High SNR regime (bandwidth limited)
  - $R_{SM} \approx rR_{BF}$. Gain of r.
  - Spatial multiplexing adds degrees of freedom
- Spatial multiplexing is like adding bandwidth



4×4.5 bps/Hz

4.5 bps/Hz

Simulation:

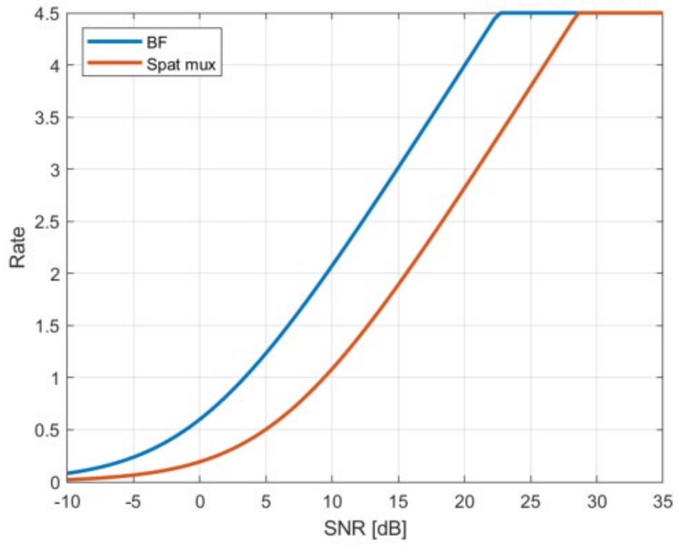
$r = 4$ streams

$R(\gamma) = \min\{4.5, 0.5 \log_2(1 + \gamma)\}$

NYU TANDON SCHOOL OF ENGINEERING

Spatial Multiplexing vs Beamforming
Single Dominant Stream

- Now suppose that there is a single dominant virtual direction
  - $\gamma_i = 0$ for $i = 2, \ldots, r$
- BF rate: $R_{BF} = R(\gamma_1)$
- Spatial mux rate $R_{SM} = R\left(\frac{\gamma_1}{r}\right)$
- Spatial multiplexing with uniform power allocation is worse!
  - Waste energy in directions with no SNR



Power Optimization and Water-Filling

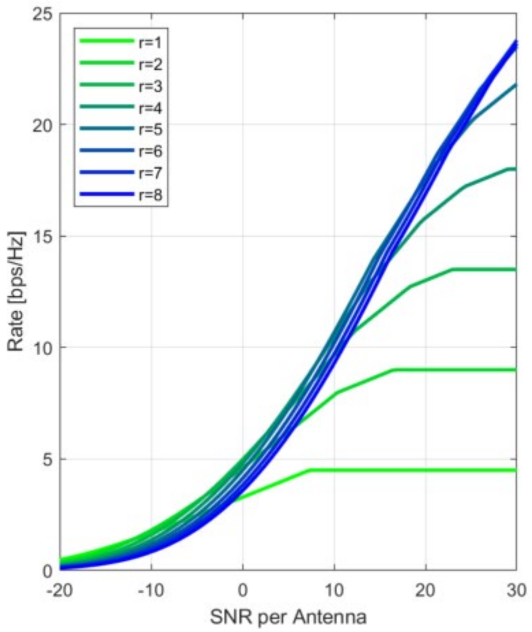
- We can allocate the power optimally
- Let x_i = fraction of power allocated to stream i
- Optimize the rate:

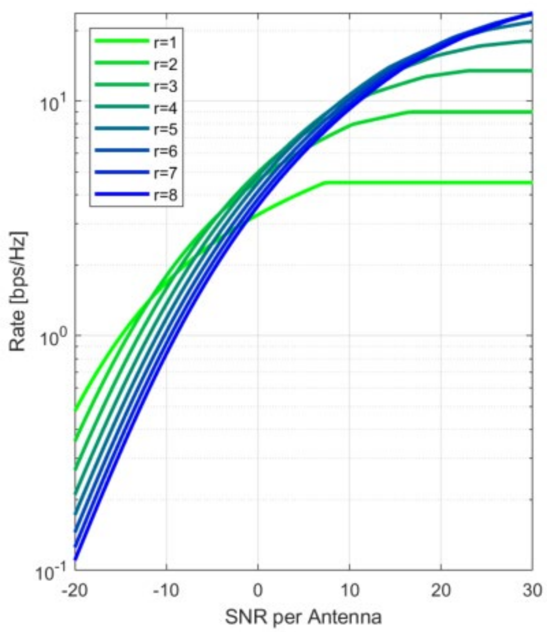
$$\max \sum_{i=1}^{r} R(\gamma_i x_i) \;\; \text{s.t.} \quad x_i \geq 0, \qquad \sum_{i=1}^{r} x_i \leq 1$$

- Some special cases:
  - Beamforming: Place all power in best stream: $x_1 = 1,\ x_i = 0$ for $i > 1$
  - Uniform power allocation: $x_i = \frac{1}{r}$
- When $R(\gamma) = \log_2(\gamma)$, optimal solution is given by "water-filling" (see text)
  - $x_i = \max\{ c - \frac{1}{\gamma_i}, 0\}$
  - Constant c set to satisfy constraint that $\sum_{i=1}^{r} x_i \leq 1$
  - Allocate energy inversely proportional to SNR
  - Some streams are allocated zero energy

Power Allocation

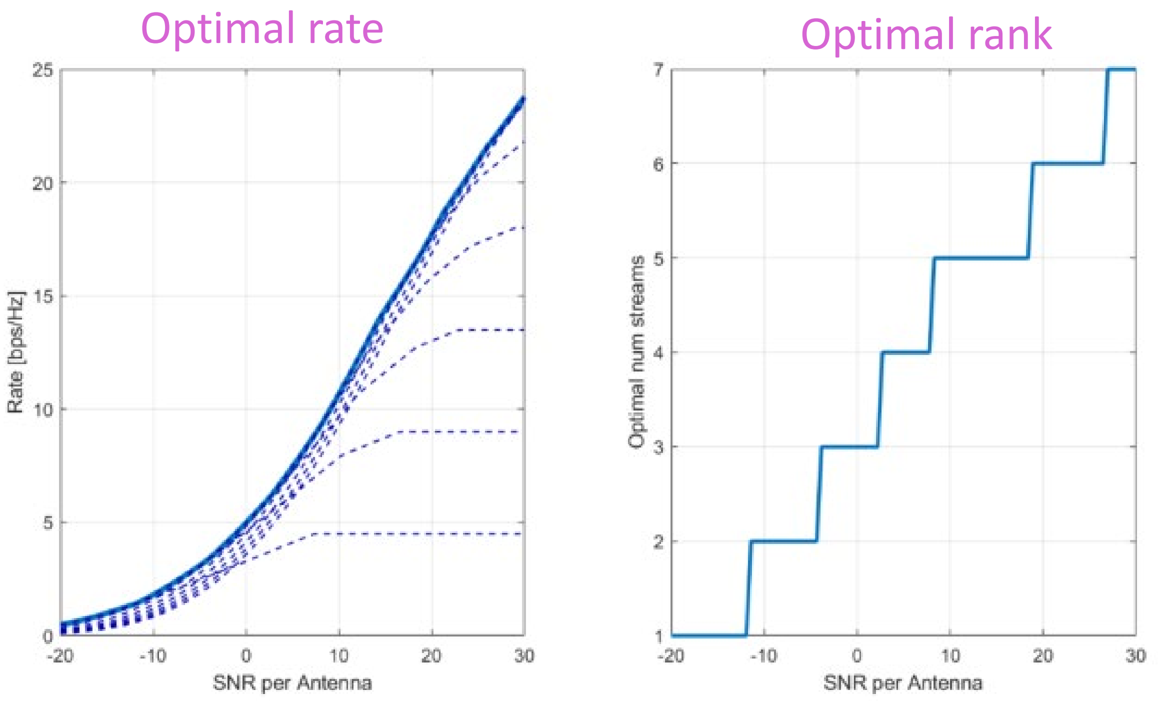
- Consider sub-optimal strategy
  - Uniformly allocate to top r streams
- Plot to the right:
  - Synthetic channel from before
  - We see that $r = 1$ (BF) optimal at lower SNRs
  - At higher SNRs, multi-streams become useful





Optimal Rank Selection

- Water Filling is difficult for a general rate function
- Consider sub-optimal strategy
  - Uniformly allocate to top r streams
- Left plot:
  - Best rate among all possible ranks
- Right plot:
  - Best rank
- We see rank increases with SNR



Water-Filling

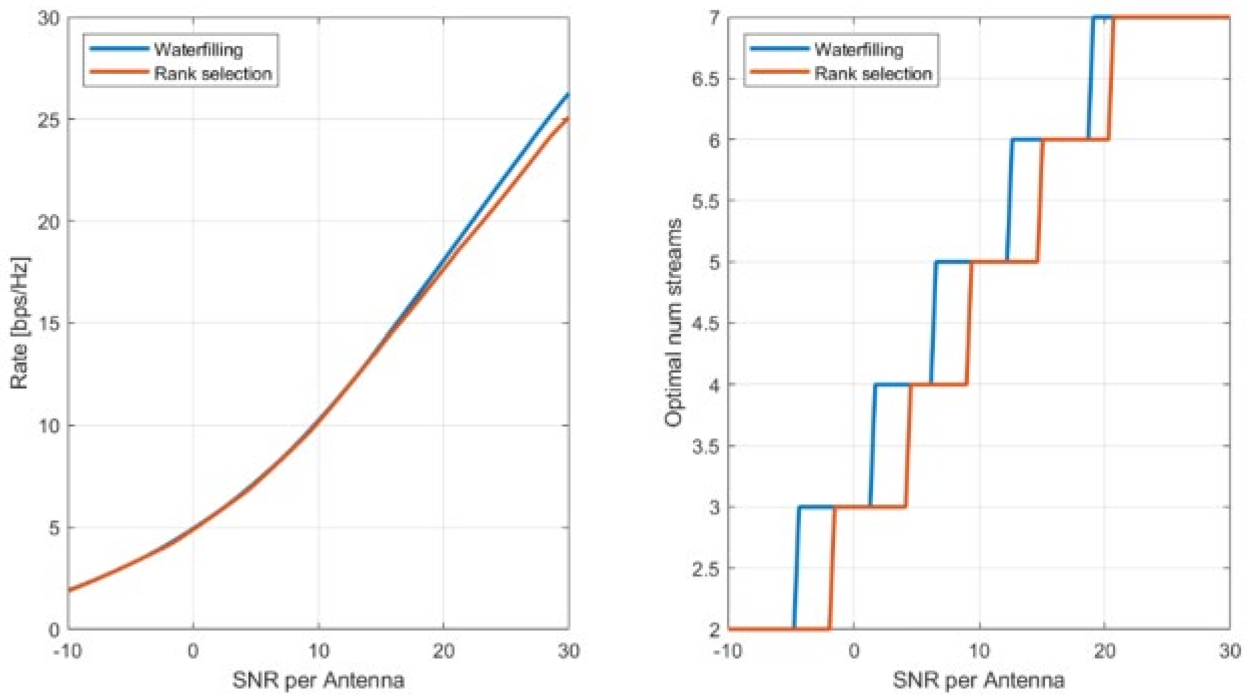
- Assume the rate per stream is given by Shannon's formula:

$$r = \log_2(1 + \frac{s_i^2}{N_0} E_i)$$

- Maximize $R = \frac{1}{\log(2)} \sum_{i=1}^{r} \log(1 + \gamma_i x_i)$ s.t. $\sum_{i=1}^{r} x_i = 1$
  - $\gamma_i = \frac{s_i^2 E_x}{N_0}$ = SNR on virtual stream i if all power is allocated to stream
- Take Lagrangian: $L = \sum_{i=1}^{r} [\log(1 + \gamma_i x_i) + \lambda x_i - \mu_i\, x_i] - \lambda$
  - $\mu_i \geq 0$
- Take derivative: $\frac{1}{1+\gamma_i x_i} = \lambda - \mu_i$
- For positive streams: $1 + \gamma_i x_i = c \Rightarrow x_i = \max(0, c - \frac{1}{\gamma_i})$

Example: Water-Filling vs Rank Selection

- Compare two strategies:
- Water-filling: Optimal allocation
- Rank selection:
  - Uniform power among best r directions
- Comparison:
  - Gain from waterfilling is small
  - Only helps at very high SNR

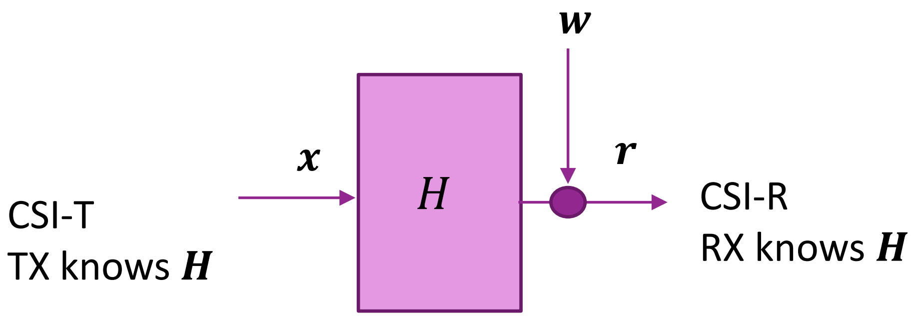


Outline

- Spatial Multiplexing with CSI-T and CSI-R
- Power Allocation and Rank Selection
- Spatial Multiplexing with CSI-R Only
- Channel Estimation and CSI-R
- CSI-T Feedback and Statistical Pre-Coding

Challenges in Obtaining CSI

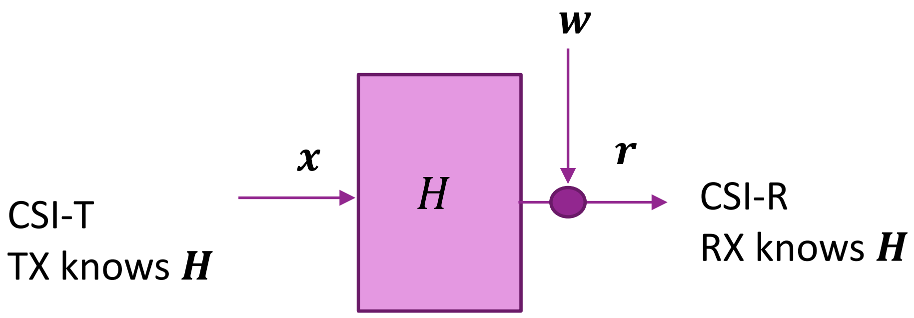
- Above analysis assumes TX and RX have CSI
- CSI-R: RX knows $\boldsymbol{H}$
  - Generally possible with sufficient reference signals
  - Reference signals add overhead
  - Overhead is reasonable if number of streams $\ll$ coherence time



- CSI-T: TX knows $\boldsymbol{H}$
  - Requires feedback from RX or measurements in reverse direction
  - With small-scale fading, $\boldsymbol{H}$ changes rapidly
  - Instantaneous $\boldsymbol{H}$ is generally difficult
- Later, we will discuss how to obtain both
- This section: Understand effect of lacking CSI-T

Problems in Lacking CSI-T

- We will see that lacking CSI-T causes two problems
- TX power mis-allocation
  - TX does not know virtual path directions
  - TX may send data into directions with poor SNR
  - Wastes TX energy



- Inter-stream interference
  - Streams end up “mixed” at RX (Each r_i depends on multiple x_j)
  - RX must separate the streams
  - Can be mitigated with equalization or joint decoding
  - But equalization or joint decoding adds complexity and may incur a penalty loss

Information Theoretic Formulation

- Consider transmission over multiple symbols: $\boldsymbol{r}_n = \boldsymbol{H}\boldsymbol{x}_n + \boldsymbol{w}_n, \quad \boldsymbol{w}_n \sim CN(0, N_0\boldsymbol{I})$
  - $n = 1, \dots, N$, N =Block length $\rightarrow \infty$
- TX encodes $M = 2^{RN}$ messages, R = bits per symbol
  - Denote messages $\boldsymbol{x}_n^{(m)}, m = 1, \dots, M$
- Transmission energy assumption: $\mathrm{E}(\boldsymbol{x}_n \boldsymbol{x}_n^*) \leq \frac{E_x}{N_t}\boldsymbol{I}$
  - Assumes TX energy is uniform over all N_t antennas
  - We assume this since the TX does not know the optimal direction
  - We will see this loss can be significant
- RX can perform any decoding. No computational limits
  - For example, optimal decoding $\hat{m} = \arg\min \sum_{n=1}^{N} \left\| \boldsymbol{r}_n - \boldsymbol{H}\boldsymbol{x}_n^{(m)} \right\|^2$

Information Theoretic Capacity

- Information theoretic model: $\boldsymbol{r}_n = \boldsymbol{H}\boldsymbol{x}_n + \boldsymbol{w}_n\,, \quad \boldsymbol{w}_n \sim CN(0, N_0\boldsymbol{I})$
  - TX power constraint: $\mathrm{E}(\boldsymbol{x}_n\boldsymbol{x}_n^*) \le \frac{E_x}{N_t}\boldsymbol{I}$

- Theorem [Teletar,~2003]:

Under optimal decoding with the block length $N \to \infty$ the capacity is:

$$C = \log\det(\boldsymbol{I} + \alpha\boldsymbol{H}^*\boldsymbol{H}), \qquad \alpha = \frac{E_x}{N_0 N_t}$$

Capacity of Multi-antenna Gaussian Channels*

EMRE TELATAR
Lucent Technologies Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974, USA
telatar@lucent.com

Proof of the Teletar’s Result

- Channel model: $\boldsymbol{r} = \boldsymbol{Hx} + \boldsymbol{w}, \ \boldsymbol{w} \sim CN(0, N_0\boldsymbol{I})$
- From Shannon’s Theorem, capacity is $C = \max I(\boldsymbol{r};\boldsymbol{x}) = \max[H(\boldsymbol{r}) - H(\boldsymbol{r}|\boldsymbol{x})]$
  - Max is over distributions on $\boldsymbol{x}$
- Fact 1: For any Gaussian $\boldsymbol{z} \sim CN(\boldsymbol{\mu}, \boldsymbol{Q})$ entropy is $H(\boldsymbol{z}) = \log\det(\pi e \boldsymbol{Q})$
- Fact 2: For $\boldsymbol{z}$ with $E(\boldsymbol{z}\boldsymbol{z}^*) \le \boldsymbol{Q}$ entropy is bounded $H(\boldsymbol{z}) \le \log\det(\pi e \boldsymbol{Q})$
- Given $\boldsymbol{x}, \ \boldsymbol{r} \sim CN(\boldsymbol{Hx}, N_0\boldsymbol{I}) \Rightarrow H(\boldsymbol{r}|\boldsymbol{x}) = \log\det(\pi e N_0 \boldsymbol{I}) = N_t \log(\pi e N_0)$
- If $E(\boldsymbol{x}\boldsymbol{x}^*) \le \frac{E_x}{N_t}\boldsymbol{I}$ then $E(\boldsymbol{r}\boldsymbol{r}^*) \le N_0\boldsymbol{I} + \frac{E_x}{N_t}\boldsymbol{H}\boldsymbol{H}^* = N_0(\boldsymbol{I} + \alpha\boldsymbol{H}\boldsymbol{H}^*), \quad \alpha = \frac{E_x}{N_0 N_t}$
- Hence $H(\boldsymbol{r}) \le N_t \log(\pi e N_0) + \log\det(\boldsymbol{I} + \alpha\boldsymbol{H}\boldsymbol{H}^*)$
- Capacity is $C \le \log\det(\boldsymbol{I} + \alpha\boldsymbol{H}\boldsymbol{H}^*)$
- Get $C = \log\det(\boldsymbol{I} + \alpha\boldsymbol{H}\boldsymbol{H}^*)$ by using Gaussian distribution

Comparison to CSI-T Case

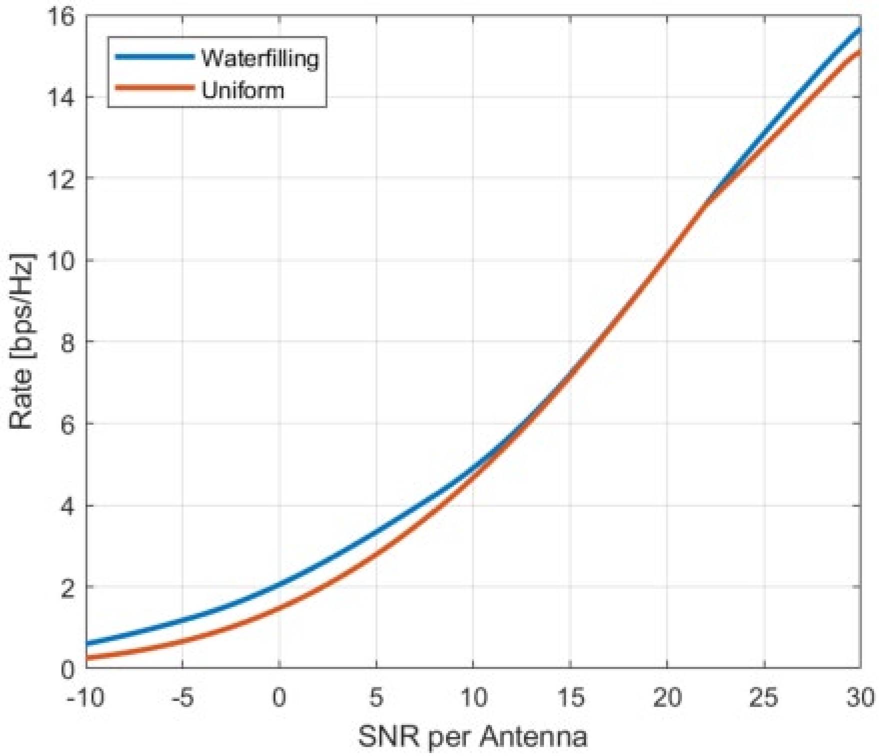
- We saw that if TX allocates energy TX power uniformly on all r virtual directions, capacity is

$$C = \log\det\left(\boldsymbol{I} + \frac{\gamma_x}{r}\boldsymbol{H}^*\boldsymbol{H}\right), \qquad \gamma_x = \frac{E_x}{N_0}$$

- Capacity with TX uniformly on all N_t TX antennas is: $C = \log\det\left(\boldsymbol{I} + \frac{\gamma_x}{N_t}\boldsymbol{H}^*\boldsymbol{H}\right)$
- Conclusion: With optimal decoding:
  - Can obtain capacity identical to TX uniformly on N_t directions
- There is a TX power mis-allocation loss:
  - Loss of $\frac{r}{N_t}$ in SNR when $N_t > r$. This is especially large when $N_t > N_r$
  - No water-filling. Hence, within the rank r virtual streams, power is not allocated optimally

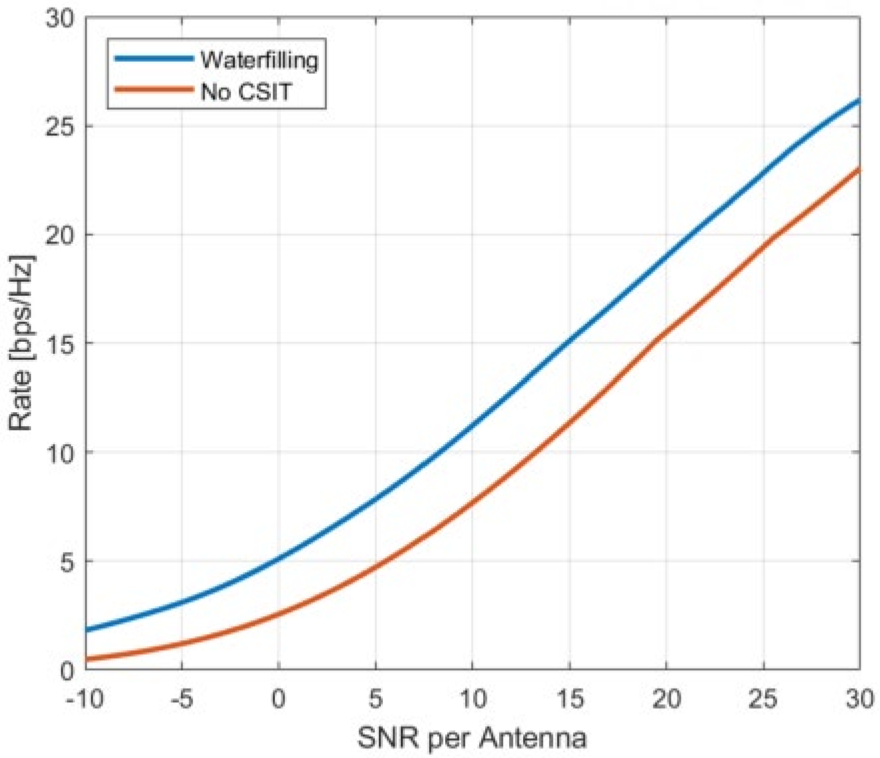
TX Mis-Allocation Loss Example
Low Dimensional Array Case

- Simulation:
  - $N_t = 4, N_r = 4$
  - Random channel with rich scattering
  - AoA and AoD uniform on $[0,2\pi]$
- TX mis-allocation loss is minimal
- With rich scattering and small num antennas:
  - All directions have good SNR
  - Uniform power allocation is optimal



TX Mis-Allocation Loss Example
High Dimensional Array Case

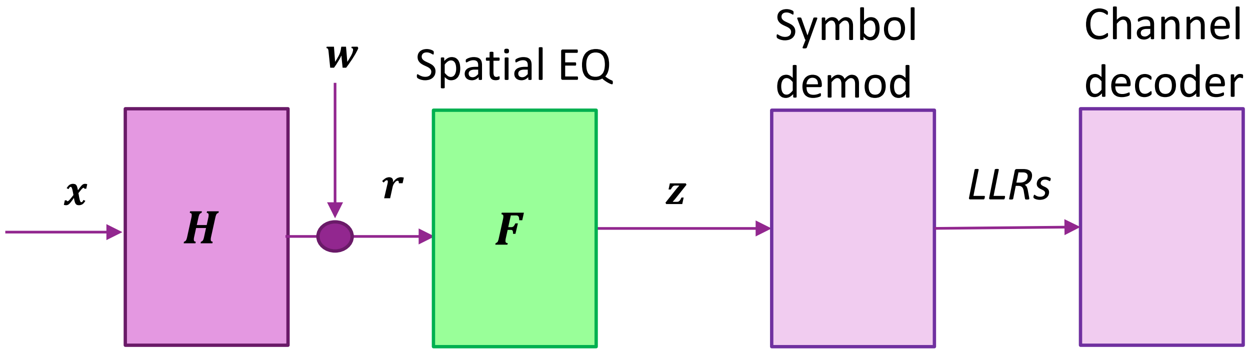
- Simulation from before:
  - $N_t = 16, N_r = 8$
  - Random channel
- Max rank is $r = \min(N_r, N_t) = 8$
- At high SNR, the loss is $\frac{r}{N_t} = \frac{1}{2}$
- At lower SNRs, the loss is larger
  - Wastes significant energy on poor rank directions



Loss from Inter-Stream Interference

- Prior capacity result required optimal decoding:
  - Search over all possible codewords $\hat{m} = \arg\min \sum_{n=1}^{N} \left\| \boldsymbol{r}_n - \boldsymbol{H}\boldsymbol{x}_n^{(m)} \right\|^2$
- Computationally impossible for even moderate block size
- We consider a simple linear equalization scheme:
  - Linear equalization followed by symbol demodulation and channel decoding
- This method introduces a second loss due to inter-stream interference

Linear Equalization Concept

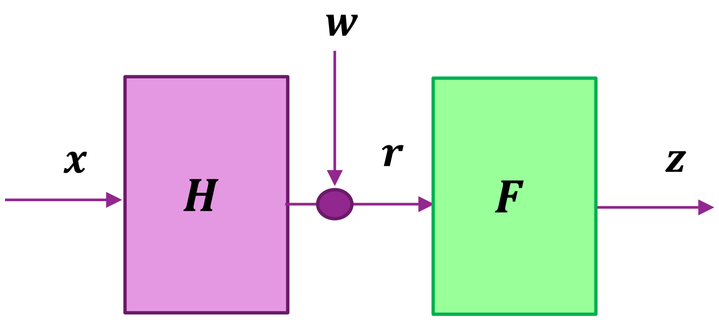


- Most practical systems use linear spatial equalization
  - Perform linear transform $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{r}$ to approximately invert $\boldsymbol{H}$ and recover $\boldsymbol{x}$
  - Followed by symbol demodulation on the symbols $\boldsymbol{z}$ to create LLRs
  - LLRs then used by the channel decoder
- In contrast, we call the optimal decoder the joint decoding
  - Since it jointly performs the symbol demodulation and channel decoding

Linear Zero-Forcing Equalizer

- Estimate $\boldsymbol{x}$ via a least-squares optimization

$$z = \arg\min_{x} \|\boldsymbol{r} - \boldsymbol{H}\boldsymbol{x}\|^2$$



- Solution is called the zero-forcing equalizer:

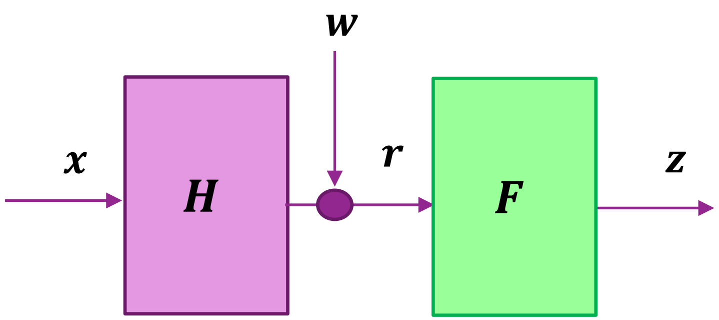
$$\boxed{\boldsymbol{z} = \boldsymbol{F}_{ZF}\boldsymbol{r}, \qquad \boldsymbol{F}_{ZF} = (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*}$$

  - The reason for the name will be clear later

- Note for the inverse to exist we need $N_r \geq N_t$
  - More specifically, we need $rank(\boldsymbol{H}) \geq N_t$

Zero-Forcing Equalizer Analysis

- Suppose we use ZF equalizer $\boldsymbol{z} = \boldsymbol{F}_{ZF}\boldsymbol{r}, \ \boldsymbol{F}_{ZF} = (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*$



- Then channel from $\boldsymbol{x}$ to $\boldsymbol{z}$ is
  - $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{r} = (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}) = \boldsymbol{x} + \boldsymbol{d}, \ \ \boldsymbol{d} = (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{w}$
- Creates N_t parallel channels
  - $z_i = x_i + d_i, i = 1, \ldots, N_t$
- Noise covariance matrix is:
  - $var(d) = (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^* var(\boldsymbol{w})\boldsymbol{H}(\boldsymbol{H}^*\boldsymbol{H})^{-1} = N_0(\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{H}(\boldsymbol{H}^*\boldsymbol{H})^{-1} = N_0(\boldsymbol{H}^*\boldsymbol{H})^{-1}$
- SNR on each channel: Since $E|x_i|^2 = \frac{E_x}{N_t}$

$$\boxed{\gamma_i^{ZF} = \frac{E_x}{N_t N_0}\frac{1}{Q_{ii}}, \qquad \boldsymbol{Q} = (\boldsymbol{H}^*\boldsymbol{H})^{-1}}$$

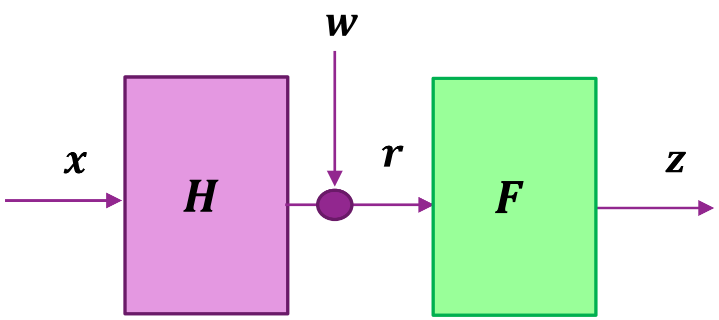
Problems with Zero Forcing

- SNR on each stream with zero forcing is: $\gamma_i^{ZF} = \frac{E_x}{N_t N_0} \frac{1}{Q_{ii}}, \quad \boldsymbol{Q} = (\boldsymbol{H}^* \boldsymbol{H})^{-1}$
- For inverse to exist, requires that $N_r \geq N_t$
- Also, when eigenvalues of $\boldsymbol{H}^* \boldsymbol{H}$ are small, $\boldsymbol{Q}$ will blow up
- What is the optimal linear transform?

Linear MMSE Equalization

- Narrowband channel $\boldsymbol{r} = \boldsymbol{Hx} + \boldsymbol{w}, \quad \boldsymbol{w} \sim CN(0, N_0 \boldsymbol{I})$,
- Assume
  - Assume $E(\boldsymbol{x}\boldsymbol{x}^*) = \frac{E_x}{N_t}\boldsymbol{I}$ and $E(\boldsymbol{x}) = \boldsymbol{0}$
  - TX energy is $\frac{E_x}{N_t}$ in each antenna
  - Select F to minimize average error:

$$E\|\boldsymbol{x} - \boldsymbol{Fr}\|^2 = E\|\boldsymbol{x} - \boldsymbol{F}(\boldsymbol{Hx} + \boldsymbol{w})\|^2$$

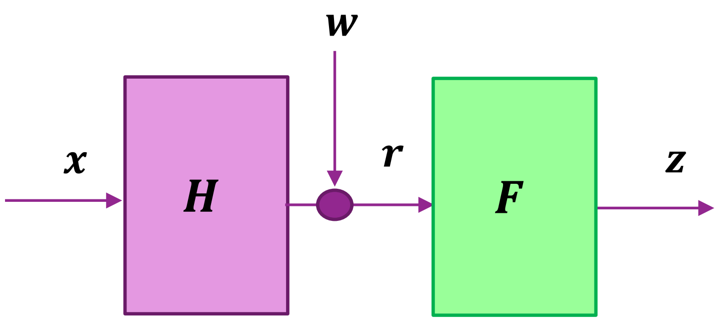


- Optimal linear estimator from probability theory is

$$\boldsymbol{z} = \boldsymbol{Fr}, \qquad \boldsymbol{F}_{LMMSE} = \alpha(\alpha \boldsymbol{H}^*\boldsymbol{H} + \boldsymbol{I})^{-1}\boldsymbol{H}^*, \qquad \alpha = \frac{E_x}{N_t N_0}$$

Linear Equalizer: Interference + Noise

- Narrowband channel $\boldsymbol{r} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}, \quad \boldsymbol{w} \sim CN(0, N_0\boldsymbol{I})$,
- Equalized symbols: $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{r}$
  - For now, consider general linear equalizer $\boldsymbol{F}$
- Hence: $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{w}$ and hence, per component



$$z_i = [FH]_{ii}x_i + \sum_{j \neq i} [FH]_{ij}x_j + [Fw]_i$$

Desired signal

Inter-stream interference

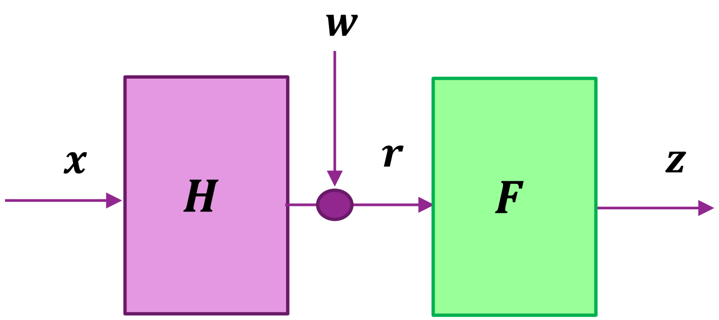
Noise

- LMMSE receiver: Decode x_i from z_i treating inter-stream interference as noise

LMMSE: SINR

- From previous slide:

$$z_i = [FH]_{ii} x_i + \sum_{j \neq i} [FH]_{ij} x_j + [Fw]_i$$



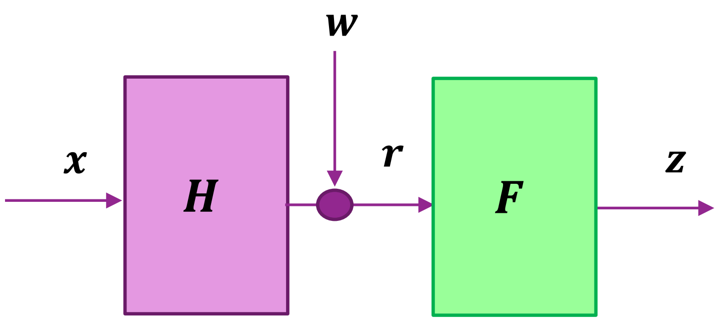
- On channel i:
  - Signal energy $E_{sig} = |[FH]_{ii}|^2 \ \frac{E_x}{N_t}$
  - Interference energy: $E_{int} = \frac{E_x}{N_t} \sum_{j \neq i} |[FH]_{ij}|^2$
  - Noise energy: $E_{noise} = N_0 \ [F^* F]_{ii}$
- SINR on channel i:
  - $\gamma_i^{LMMSE} = \frac{E_{sig}}{E_{noise} + E_{int}}$

LMMSE SINR

- Channel $\boldsymbol{r} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}, \quad \boldsymbol{w} \sim CN(0, N_0 \boldsymbol{I}), \quad E(\boldsymbol{x}\boldsymbol{x}^*) = \frac{E_x}{N_t}\boldsymbol{I}$
- Consider linear equalizer $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{r}$
- Theorem:
  - With the LMMSE equalizer, $\boldsymbol{F} = \boldsymbol{F}_{LMMSE}$, SINR per channel is:

$$\boxed{\gamma_i^{LMMSE} = \frac{1}{Q_{ii}} - 1, \qquad \boldsymbol{Q} = (\alpha \boldsymbol{H}^* \boldsymbol{H} + \boldsymbol{I})^{-1}, \qquad \alpha = \frac{E_x}{N_0 N_t}}$$

  - For any other linear transform $\boldsymbol{F}$, $\gamma_i \le \frac{1}{Q_{ii}} - 1$
- Proof: Follows from long linear algebra.
  - See text



LMMSE vs. ZF

- At high SNR, LMMSE → ZF:
  - Recall $\boldsymbol{F}_{LMMSE} = \alpha(\alpha\boldsymbol{H}^*\boldsymbol{H} + \boldsymbol{I})^{-1}\boldsymbol{H}^*, \alpha = \frac{E_x}{N_0 N_t}$
  - If $\boldsymbol{H}^*\boldsymbol{H}$ is invertible, as $\alpha \to \infty$, $\boldsymbol{F}_{LMMSE} = \alpha(\alpha\boldsymbol{H}^*\boldsymbol{H} + \boldsymbol{I})^{-1}\boldsymbol{H}^* \to (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^* = \boldsymbol{F}_{ZF}$

- For ZF, there is no inter-stream interference:
  - Recall $z_i = [FH]_{ii}x_i + \sum_{j\neq i}[FH]_{ij}x_j + [Fw]_i$
  - With $\boldsymbol{F}_{ZF}\boldsymbol{H} = (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{H} = \boldsymbol{I}$
  - Thus, $[FH]_{ij} = 0$ for $i \neq j$
  - Hence, ZF forces the inter-stream interference to zero. Hence, the name
  - But ZF amplifies the noise term $\boldsymbol{Fw}$
  - Linear MMSE optimally trades off noise and inter-stream interference

LMMSE vs. Joint Decoding

- We saw that with optimal joint decoding capacity is:

$$C_{joint} = \log_2 \det(\boldsymbol{I} + \alpha \boldsymbol{H}^* \boldsymbol{H}) = -\log_2 \det(\boldsymbol{Q}), \qquad \boldsymbol{Q} = (\alpha \boldsymbol{H}^* \boldsymbol{H} + \boldsymbol{I})^{-1}$$

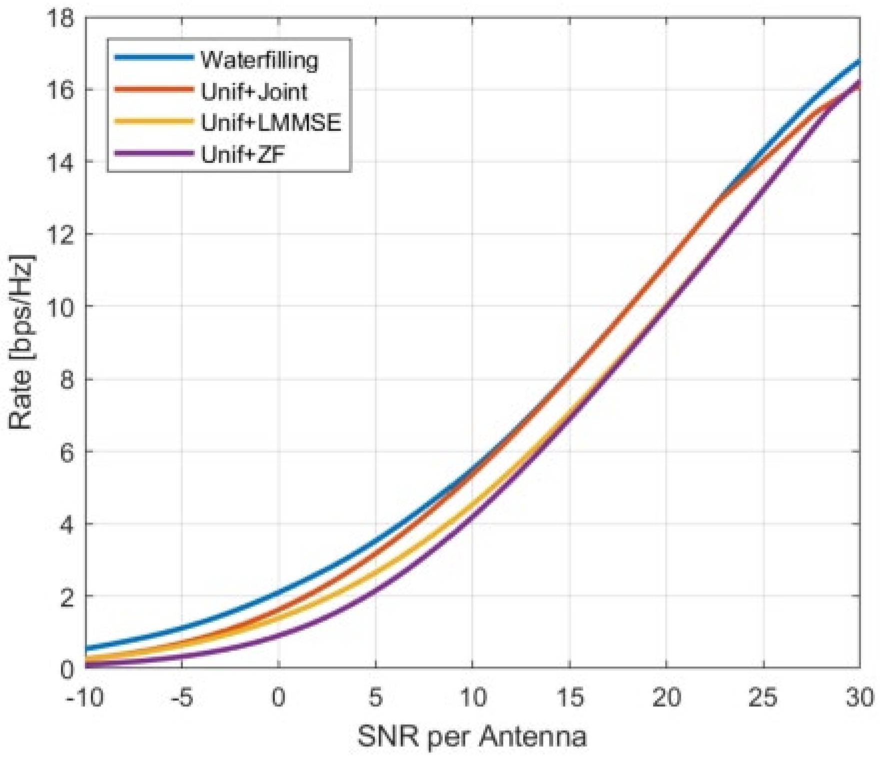
- With LMMSE, capacity is:

$$C_{LMMSE} = \sum_{i=1}^{N_t} \log_2(1 + \gamma_i^{LMMSE}) = \sum_{i=1}^{N_t} \log_2\left(\frac{1}{Q_{ii}}\right) = -\sum_{i=1}^{N_t} \log_2(Q_{ii})$$

- Fact: $C_{LMMSE} \le C_{joint}$
- Proof:
  - Linear algebra fact: for any matrix $\boldsymbol{Q} \ge 0$, $\det(\boldsymbol{Q}) \le \prod_i Q_{ii}$
  - This fact follows from a Cholesky factorization $Q = LL^*$
  - Therefore, $C_{joint} = -\log_2 \det(\boldsymbol{Q}) \ge -\log_2 \prod_i Q_{ii} = -\sum_{i=1}^{N_t} \log_2(Q_{ii}) = C_{LMMSE}$

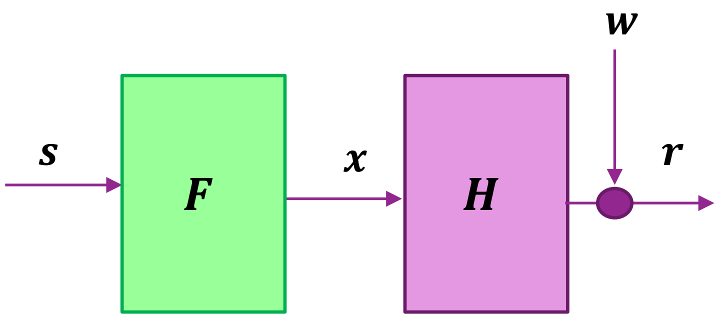
LMMSE Loss: Low-Dim Case

- Consider the low-dimensional example
  - 4 x 4 MIMO with rich scattering
  - See parameters above
- At low SNRs:
  - LMMSE performs close to Joint decoding
  - LMMSE performs much better than ZF
- At high SNRs:
  - Joint decoding provides a small advantage
  - Gain of ~2 dB
  - ZF performs close to LMMSE



LMMSE Loss: High-Dim Case

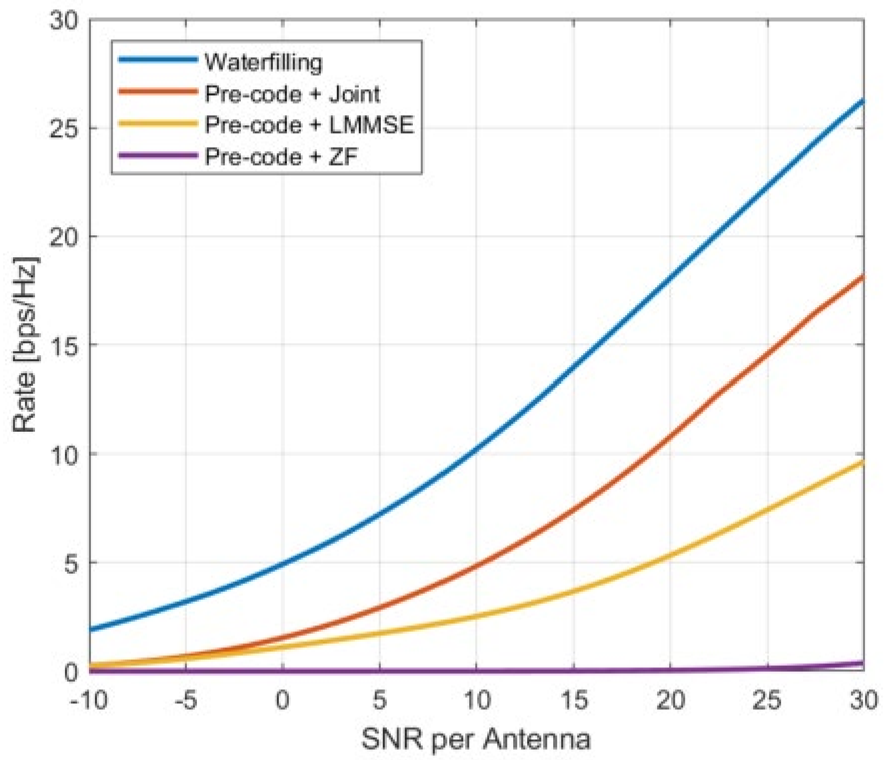
- Next consider the low-dimensional example
  - $N_t = 16$ and $N_r = 8$
  - See parameters above
- Linear equalization requires that $N_t > \mathrm{r}$
  - $r =$ channel rank = number of virtual directions with significant SNR
  - Otherwise, cannot recover $\boldsymbol{x}$
- When $N_t > r$, we need to perform pre-coding:
  - Estimate the number, r, of TX streams to use (somehow)
  - Pre-code $x = Fs$ where s is r −dimensional
- At the receiver perform decoding on pre-coded channel $\boldsymbol{G} = \boldsymbol{HF}$



Pre-coder

LMMSE Loss: High-Dim Case

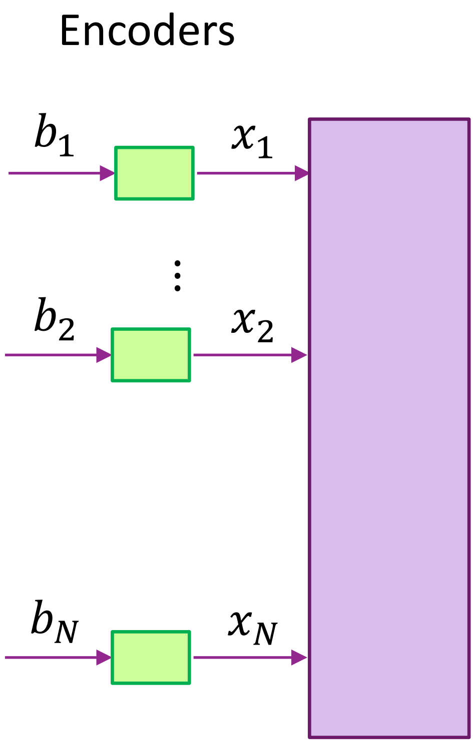
- We have $N_t = 16$ and $N_r = 8$
- Hence, some pre-coding is needed to use LMMSE
- As an example, suppose we use random pre-coding
  - Select $r = 6$ streams (num of significant evals)
  - Take a r= random orthogonal methods
- After random pre-coding:
  - There is a significant loss.
  - Even with optimal joint decoding
  - LMMSE and ZF have further losses
- For high-dim arrays, some CSI-T is needed
  - This is particularly important in mmWave
  - Need to intelligently select directions



Improving over LMMSE

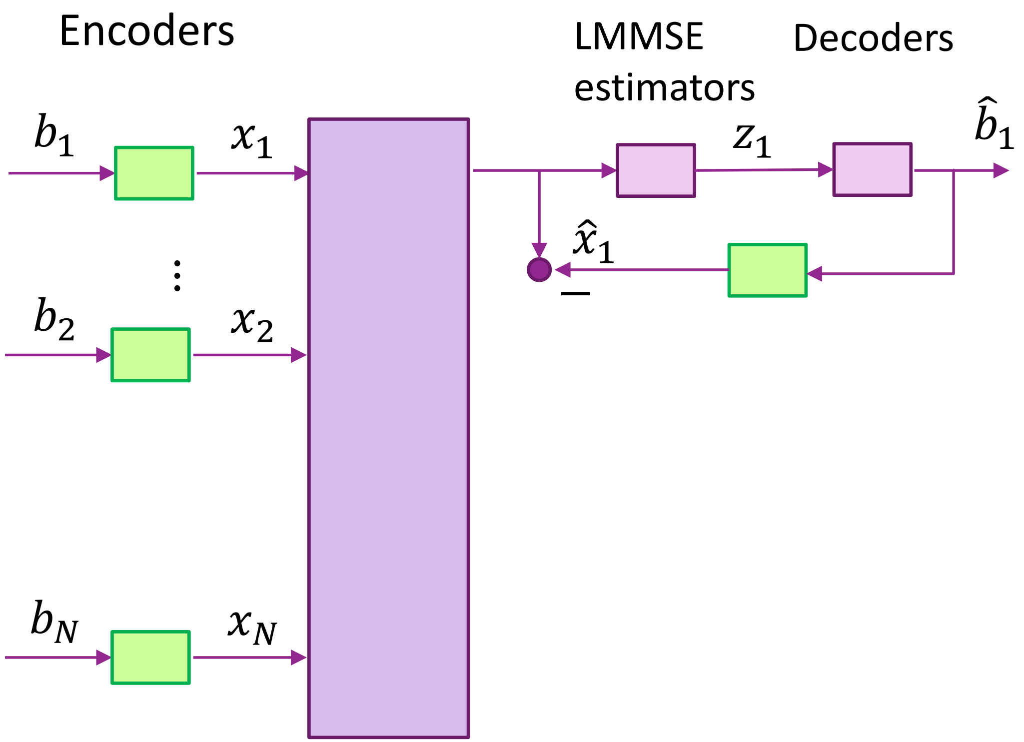
- We see that LMMSE decoding is not optimal
- Incurs a penalty due to inter-stream interference
- Several possible advanced receivers can be used to reduce inter-stream interference:
- LMMSE + SIC:
  - Successively decode each stream and cancel it out
  - We describe this next
- Turbo / joint decoding methods:
  - Iteratively perform iterations of the decoder with the LMMSE
  - Take the soft information of decoder to improve the LMMSE
  - See text

LMMSE + SIC: Transmitter



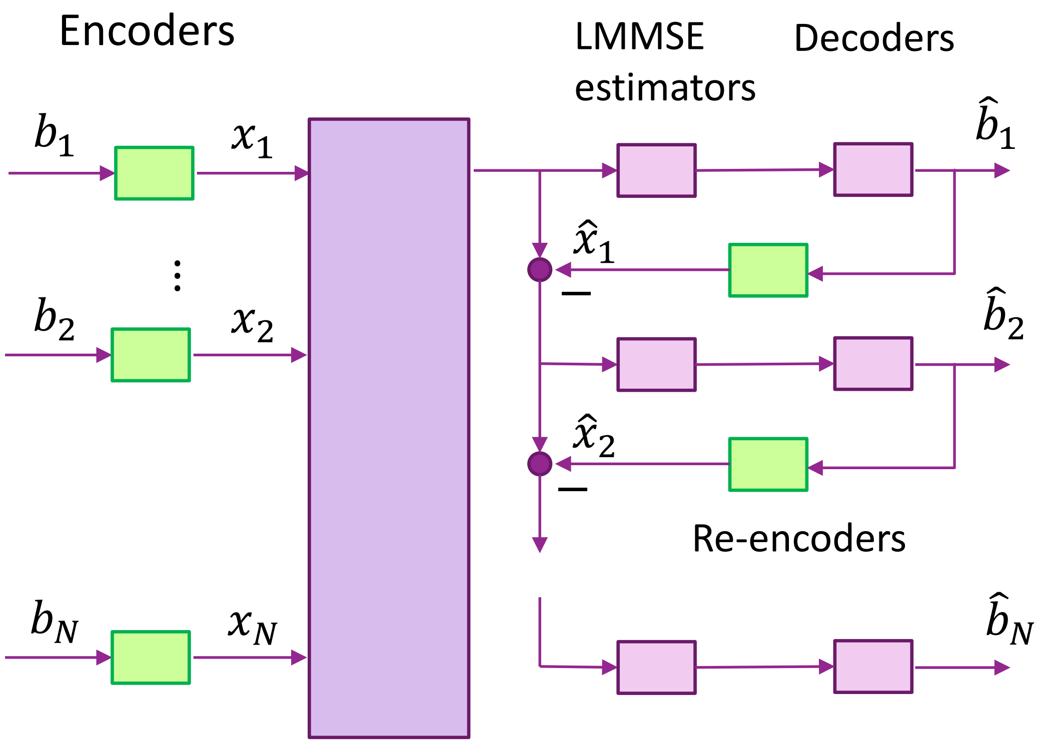
- TX:
  - Divide data into $N = N_t$ streams
  - Get information bits $b_1, \ldots, b_N$
  - Encode N codewords
  - Modulate to create N symbols $x_1, \ldots, x_N$
  - TX x_i on TX antenna i
- Channel: $\boldsymbol{r} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}$
  - Can be written: $\boldsymbol{r} = \boldsymbol{h}_1 x_1 + \cdots + \boldsymbol{h}_N x_N + \boldsymbol{w}$

LMMSE + SIC: First Stream



- Decode stream 1:
  - Perform LMMSE estimate of x_1 from $\boldsymbol{r}$ to get z_1
  - Signal z_1 will have interference from $x_2, \ldots, x_N$
  - Treat signals from $x_2, \ldots, x_N$ as noise
  - Decode bits b_1 from LMMSE signal z_1 (symbol demod + decoder)
- Cancellation phase
  - Re-encode estimated bits $\hat{b}_1$
  - Subtract out $\hat{x}_1$ to give $\boldsymbol{r} \leftarrow \boldsymbol{r} - \boldsymbol{h}_1 \hat{x}_1$
  - If $x_1 = \hat{x}_1$, $\boldsymbol{r} = \boldsymbol{h}_2 x_2 + \cdots + \boldsymbol{h}_N x_N + \boldsymbol{w}$
  - Residual signal r has no contribution from x_1!

LMMSE + SIC: Subsequent Streams



- For stream 2:
  - As stream 1, get LMMSE estimate of x_2
  - LMMSE uses residual
  - Residual r has no contribution from x_1
  - Interference is only $x_3, \ldots, x_N$
  - Treat signals from $x_3, \ldots, x_N$ as noise
  - Decode bits b_2, re-encode and subtract
- Continue for all N streams

LMMSE + SIC Performance

- Theorem: The capacity of the LMMSE SIC scheme is equal to optimal joint decoding
- Proof: Use linear algebra to estimate SINR in each stream.
  - With some linear algebra manipulations, you can show you end up at the same capacity
- Conclusions: LMMSE + SIC is a practical method to get optimal joint decoding
- Computational issues:
  - Error propagation: If one stream is in error, you cannot subtract
  - Difficult to merge with H-ARQ.
  - Need large buffer for symbols. This buffer is the main bottleneck in practical systems
- Commercial systems:
  - Have generally only used LMMSE+SIC on small numbers of antennas
  - LMMSE without SIC is overwhelmingly dominant implementation method

Outline

- Spatial Multiplexing with CSI-T and CSI-R
- Power Allocation and Rank Selection
- Spatial Multiplexing with CSI-R Only
- Channel Estimation and CSI-R
- CSI-T Feedback and Statistical Pre-Coding

Obtaining CSI-R

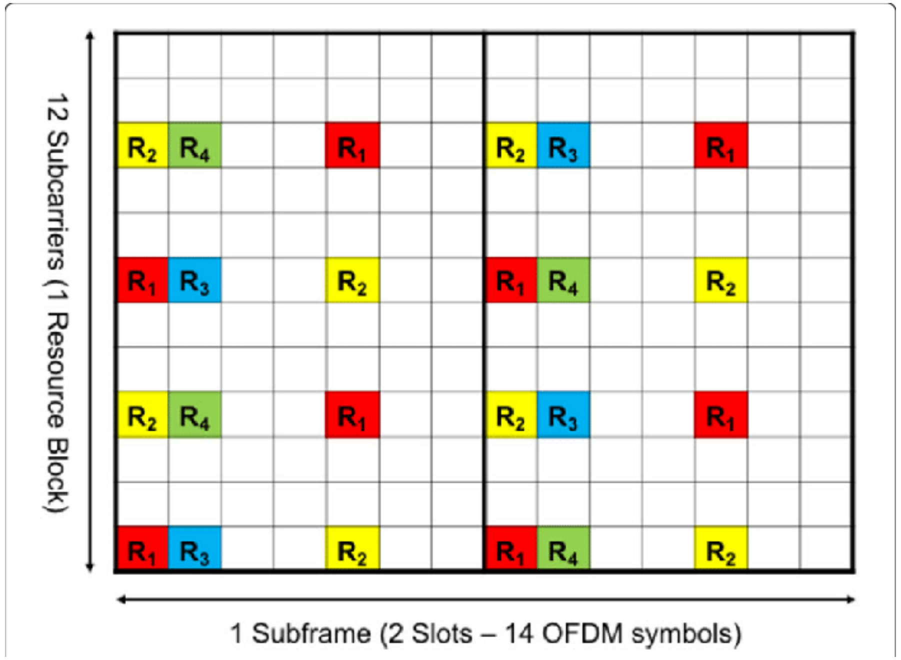
- Assume TX has selected a pre-coder $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{s}$
- RX sees channel $\boldsymbol{r}[n] = \boldsymbol{H}\boldsymbol{x}[n] + \boldsymbol{w}[n] = \boldsymbol{H}\boldsymbol{V}\boldsymbol{s}[n] + \boldsymbol{w}[n] = \boldsymbol{G}\boldsymbol{s}[n] + \boldsymbol{w}[n]$
- Results in a channel matrix $\boldsymbol{G} \in \mathbb{C}^{N_r \times N_s}$
  - N_s= number of streams
  - N_r= number of RX antennas
- For CSI-R, RX needs to estimate $\boldsymbol{G}$
  - RX does not need to know the pre-coder or the true channel!

Reference Signals

- ❑ Most systems use some form of reference signals
- ❑ One set of reference signals for each TX stream
  - Typically allocated on orthogonal resources
- ❑ Example to right: One sub-frame in LTE
  - Configuration for 4 TX "ports"
  - Resource elements R_1 to R_4 are the RS for each port
  - Each port has 2 to 4 REs per resource block
- ❑ In a RS for stream k, we get a measurement:

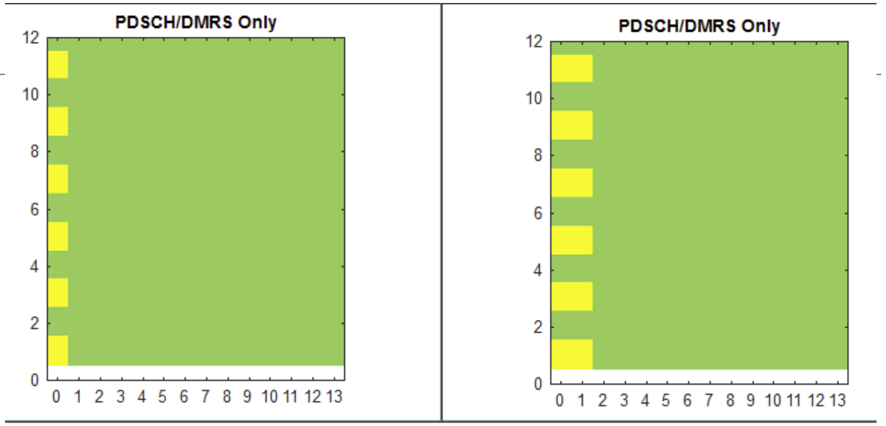
$$\boldsymbol{r} = \boldsymbol{g}_k x_k + \boldsymbol{w}$$

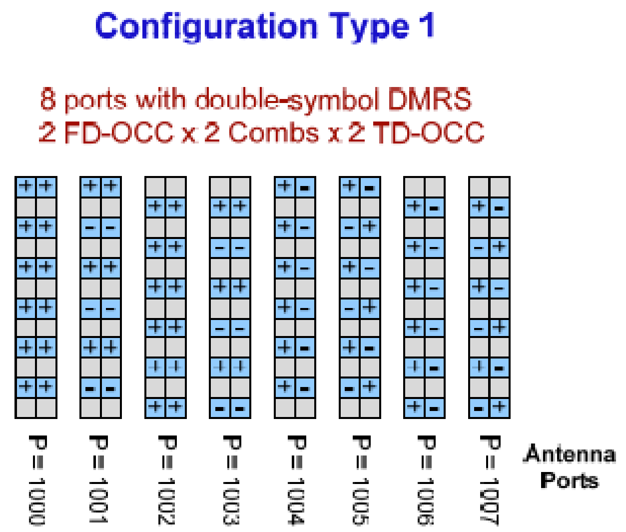
  - Estimate $\boldsymbol{g}_k = k$ –th column of $\boldsymbol{G}$



Example: 5G NR DM-RS

- DM-RS: Demodulation reference signals
  - Reference signals contained in downlink data
  - Shown in yellow squares
- Multiple layers in 5G:
  - Each spatial layer is modulated to a “port”
  - One set of reference signal for each port
  - Reference signals are different ports are orthogonal
  - Bottom right:
    RS for 8 port transmission
  - Each RS allocated on 12 REs
  - Each RE shared with 4 other ports
  - Uses an orthogonal covering code (OCC)





Overhead Issues

- Suppose channel is constant over $L = W_{coh}T_{coh}$ symbols
  - W_{coh} = coherence bandwidth, T_{coh} =coherence bandwidth
- We need at least one reference symbol per transmitted stream in each coherence block
- Overhead is $\frac{N_{RS}}{L}$, N_{RS} = number of RS $\geq N_S$ = number of streams
- Rate will be: $R = (1 - \frac{N_{RS}}{L}) \sum_{i=1}^{N_s} \log_2(1 + \frac{\gamma_i}{N_S})$
  - Training loss: $1 - \frac{N_{RS}}{L}$
  - Power loss per stream: $\frac{\gamma_i}{N_S}$
  - Bandwidth increase: Sum over i
- In general, there is a tradeoff

Information Theoretic Calculation

- Hochwald and Hassibi, 2003
- Simplified block fading model
  - Channel is constant over T uses
  - Allocates RS in each block
- With known channel
  - Increasing num TX streams helps
- Without non-coherent channel
  - Eventually hurts
  - Spend more energy on training
  - Less energy on data

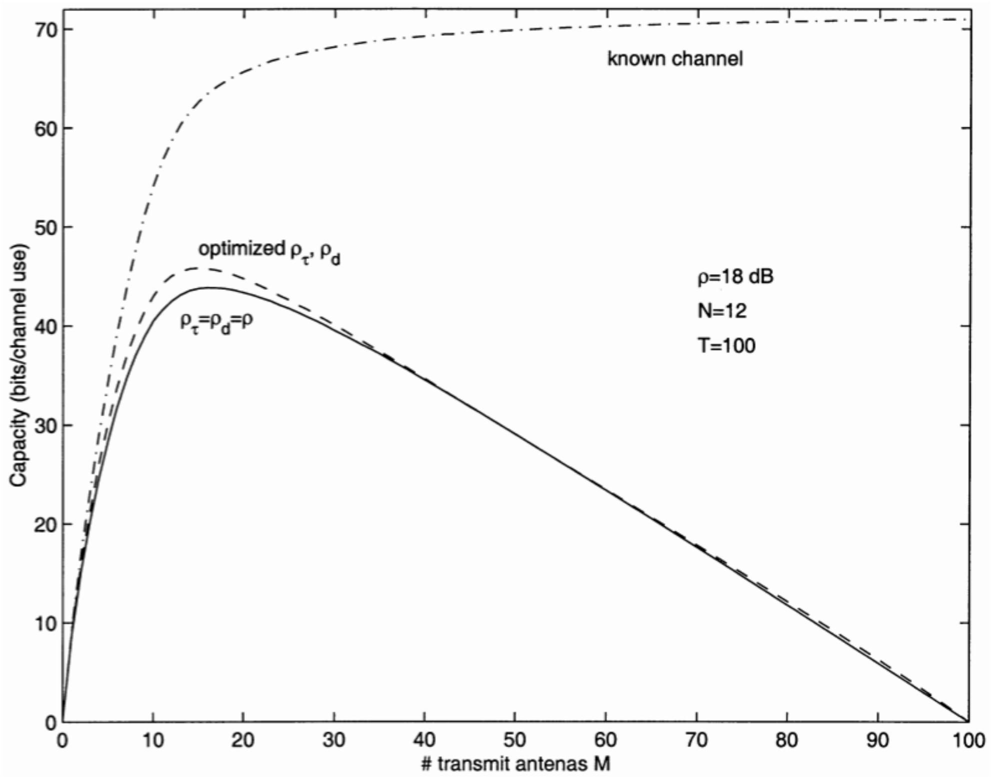


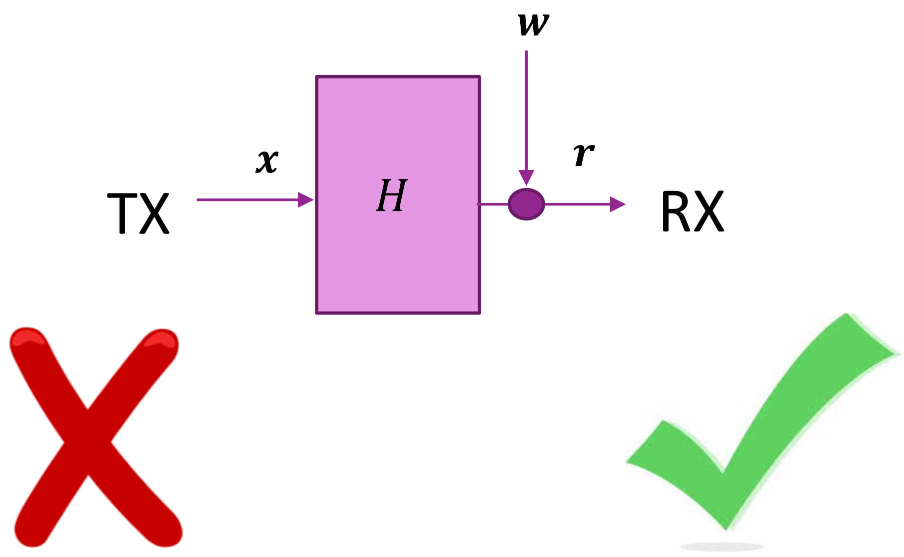
Fig. 5. Capacity as a function of number of transmit antennas M with $\rho = 18$ dB and $N = 12$ receive antennas. The solid line is optimized over T_τ for $\rho_\tau = \rho_d = \rho$ (see (40)), and the dashed line is optimized over the power allocation with $T_\tau = M$ (Theorem 3). The dash-dotted line is the capacity when the receiver knows the channel perfectly. The maximum throughput is attained at $M \approx 15$.

Outline

- ❑Spatial Multiplexing with CSI-T and CSI-R
- ❑Power Allocation and Rank Selection
- ❑Spatial Multiplexing with CSI-R Only
- ❑Channel Estimation and CSI-R
- CSI-T Feedback and Statistical Pre-Coding

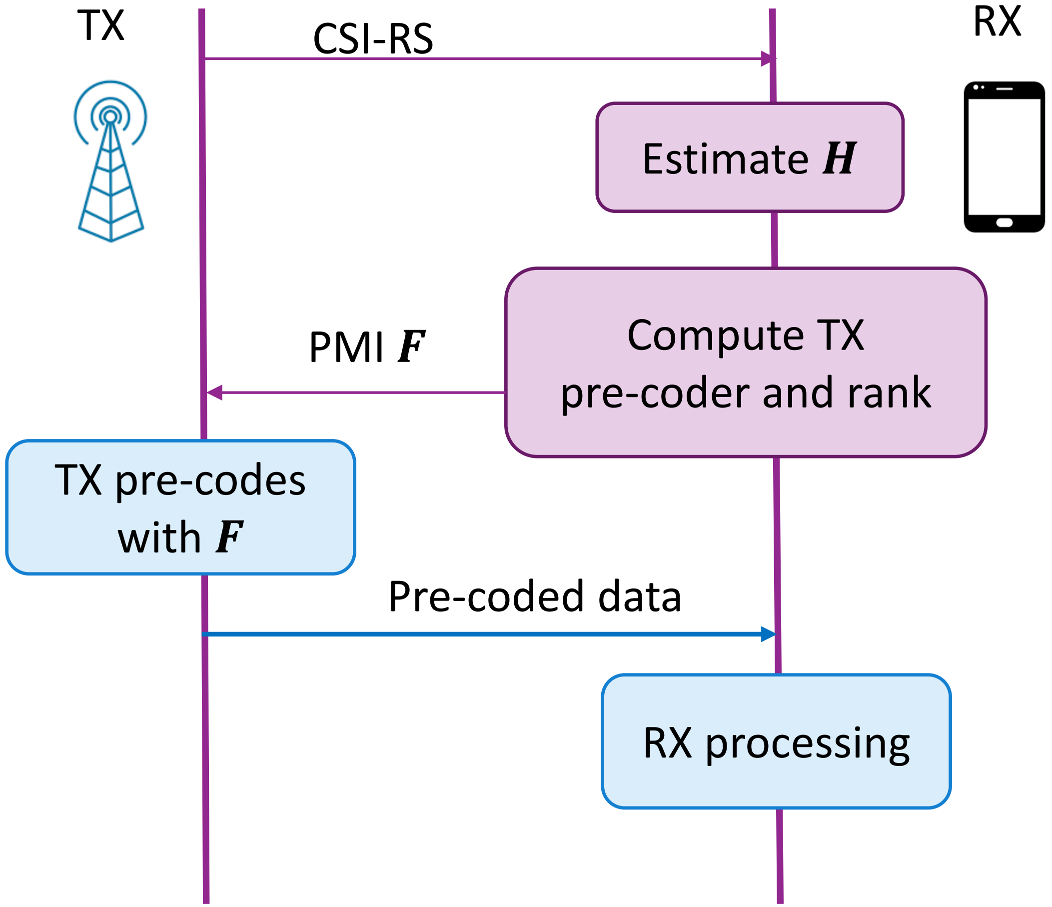
Problems in Obtaining CSI-T

- Channel state information is asymmetric
- Receiver:
  - Can directly measure the channel
- Transmitter:
  - No direct measurement
- We discuss two possible methods:
  - Precoding matrix feedback
  - Reverse link reference signal with reciprocity



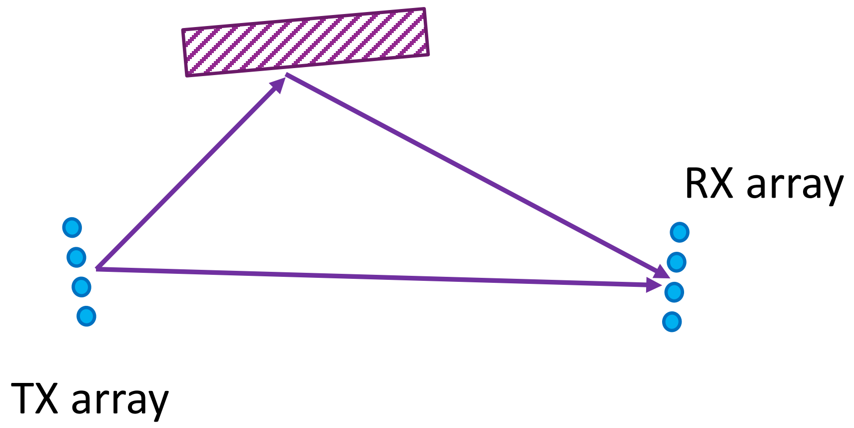
Pre-coding Feedback

- Feedback method
  - TX sends CSI reference signals from each antenna
  - RX measures complex channel matrix
  - RX computes optimal TX pre-coding matrix F
  - Also determines optimal rank
  - RX sends TX pre-coder back to TX
  - Pre-coder Matrix Indicator (PMI)
  - TX uses pre-coder in transmission
- Problem:
  - Must feedback within time coherence
  - Many need to send freq-dependent pre-coders
  - Potentially high overhead for fast varying channels



Statistical Pre-Coding

- Consider channel: $\boldsymbol{H} = \sum_{\ell=1}^{L} g_\ell e^{i\theta_\ell}\, \boldsymbol{u}_r(\Omega_\ell^r)\boldsymbol{u}_t^T(\Omega_\ell^t)$
- Parameters vary in two different time scales:
  - Small-scale variations: θ_ℓ vary with time and frequency. Difficult to track
  - Large-scale variations: $|g_\ell|$ and Ω_ℓ^r and Ω_ℓ^t vary with path gains. Slower varying. Easier to track
- Statistical pre-coding concept:
  - Measure H on many different time and frequencies
  - Assume θ varies but large-scale parameters are constant
  - Based TX pre-coding on statistics of H
  - Also called long-term pre-coding



TX and RX Spatial Covariance Matrices

- Statistical pre-coding is typically based on the spatial covariance matrices
- Consider channel: $\boldsymbol{H}(\boldsymbol{\theta}) = \sum_{\ell=1}^{L} g_\ell e^{i\theta_\ell}\, \boldsymbol{u}_r(\Omega_\ell^r)\boldsymbol{u}_t^T(\Omega_\ell^t)$
- Define TX and RX spatial covariance matrices

$$\boldsymbol{Q}_{tx} = E[\boldsymbol{H}(\theta)^*\boldsymbol{H}(\theta)] \in \mathbb{C}^{N_t \times N_t}, \qquad \boldsymbol{Q}_{rx} = E[\boldsymbol{H}(\theta)\boldsymbol{H}(\theta)^*] \in \mathbb{C}^{N_r \times N_r}$$

  - Average is over small-scale parameters

- If phases are i.i.d. $\theta_\ell \sim \mathrm{Unif}[0,2\pi]$:

$$\boldsymbol{Q}_{tx} = N_r \sum_{\ell=1}^{L} |g_\ell|^2 \boldsymbol{u}_t(\Omega_\ell^t)\boldsymbol{u}_t^*(\Omega_\ell^t), \qquad \boldsymbol{Q}_{rx} = N_t \sum_{\ell=1}^{L} |g_\ell|^2 \boldsymbol{u}_r(\Omega_\ell^r)\boldsymbol{u}_r^*(\Omega_\ell^r)$$

Statistical Pre-Coding with $\boldsymbol{Q}_{tx}$

- Estimate $\boldsymbol{Q}_{tx} = E[\boldsymbol{H}(\theta)^*\boldsymbol{H}(\theta)]$
  - Measured over many time and frequency instances
- Take eigenvalue decomposition: $\boldsymbol{Q}_{tx} = \boldsymbol{V\Lambda V}^*$, $\boldsymbol{\Lambda} = diag(\lambda_1, \dots, \lambda_{N_t})$
  - Assume sorted $\lambda_1 \geq \cdots \geq \lambda_{N_t}$
- To transmit on r streams:
  - Take TX pre-coder: $\boldsymbol{F}_{stat} = \boldsymbol{V}[:, 1:r]$ corresponding to r largest eigenvalues
  - Use pre-coding $\boldsymbol{x} = \boldsymbol{V}[:, 1:r]\boldsymbol{s}$

Instantaneous vs. Statistical Pre-Coding

- Statistical pre-coding:
  - Select one $\boldsymbol{F}_{stat} \in \mathbb{C}^{N_t \times r}$: Maps r streams to N_t antennas
  - On each channel realization, see channel matrix: $\boldsymbol{H}(\theta)\boldsymbol{F}_{stat}$
  - Obtain ergodic capacity: $C = E_\theta\left[\log_2 \det(I + \frac{E_x}{N_0}\boldsymbol{F}_{stat}^*\boldsymbol{H}^*(\theta)\boldsymbol{H}(\theta)\boldsymbol{F}_{stat})\right]$
- Instantaneous pre-coding:
  - Can select pre-coder $\boldsymbol{F}(\theta)$ for each channel realization
  - Get ergodic capacity: $C = E_\theta\left[\log_2 \det(I + \frac{E_x}{N_0}\boldsymbol{F}^*(\theta)\boldsymbol{H}^*(\theta)\boldsymbol{H}(\theta)\boldsymbol{F}(\theta)\right]$

Example

- High-dim array case: $N_t = 16, N_r = 8$
  - Transmission rank $r = 4$ and 8
  - Instantaneous selects best r directions in each realization. Statistical select r directions from $\boldsymbol{Q}_{tx}$
  - Uniform power allocation across all r