MIMO Fundamentals and Signal Processing Course

Erik G. Larsson
Linköping University (LiU), Sweden
Dept. of Electrical Engineering (ISY)
Division of Communication Systems
www.commsys.isy.liu.se

slides version: September 25, 2009

Objectives and Intended Audience

- This short course will give an introduction to the basic principles and signal processing for MIMO wireless links.
- Intended audience are graduate students and industry researchers
- Prerequisites:
  General mathematical maturity.
  Solid knowledge of linear algebra and probability theory.
  Good understanding of digital and wireless communications.
  Some basic coding/information theory.

Organization

- 4 lectures:
  Le 1: MIMO fundamentals
  Le 2: Low-complexity MIMO
  Le 3: MIMO receivers
  Le 4: MIMO in 3G-LTE (by P. Frenger, Ericsson Research)

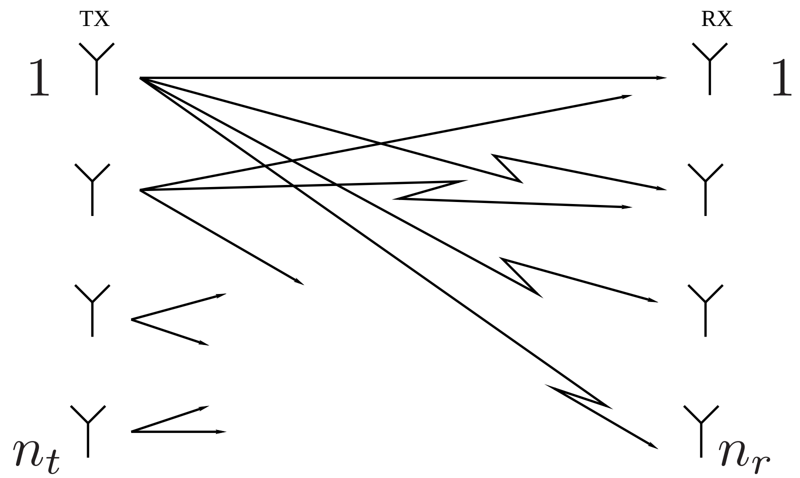
- Reading, in addition to course notes:
  - Le 1: D. Tse & P. Viswanath, Fundamentals of Wireless Communications, Cambridge Univ. Press 2005, Chs. 7–8
  - Le 2: E. G. Larsson & P. Stoica, Space-time block coding for wireless communications, Cambridge Univ. Press 2003, Chs. 2, 4–8
  - Le 3: E. G. Larsson, "MIMO detection methods: How they work", IEEE SP Magazine, pp. 91–95, May 2009

- Examination (for Ph.D. students): TBD

Le 1: MIMO fundamentals

Basic channel model

- Flat fading; linear, time-invariant channel



- Complex data $\{x_1, \ldots, x_{n_t}\}$ are transmitted via the n_t antennas

- Received data: $y_m = \sum_{n=1}^{n_t} h_{m,n} x_n + e_m$

Basic MIMO Input-Output Relation

- Transmission model (single time interval):

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_{n_r} \end{bmatrix}}_{\boldsymbol{y} \text{ (RX data)}} = \underbrace{\begin{bmatrix} h_{1,1} & \cdots & h_{1,n_t} \\ \vdots & & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix}}_{\boldsymbol{H} \text{ (channel)}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_{n_t} \end{bmatrix}}_{\boldsymbol{x} \text{ (TX data)}} + \underbrace{\begin{bmatrix} e_1 \\ \vdots \\ e_{n_r} \end{bmatrix}}_{\boldsymbol{e} \text{ (noise)}}$$

- Transmission model (N time intervals):

$$\underbrace{\begin{bmatrix} y_{1,1} & \cdots & y_{1,N} \\ \vdots & & \vdots \\ y_{n_r,1} & \cdots & y_{n_r,N} \end{bmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{bmatrix} h_{1,1} & \cdots & h_{1,n_t} \\ \vdots & & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix}}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} x_{1,1} & \cdots & x_{1,N} \\ \vdots & & \vdots \\ x_{n_t,1} & \cdots & x_{n_t,N} \end{bmatrix}}_{\substack{\boldsymbol{X} \in \mathcal{X} \\ \text{"code matrix"}}} + \underbrace{\begin{bmatrix} e_{1,1} & \cdots & e_{1,N} \\ \vdots & & \vdots \\ e_{n_r,1} & \cdots & e_{n_r,N} \end{bmatrix}}_{\boldsymbol{E}}$$

- AWGN: $e_{k,l} \sim N(0, N_0)$, i.i.d.

MIMO Design Space

- Fast fading: codeword spans ∞ number of channel realizations
  Channel can be time- or frequency-variant (e.g., MIMO-OFDM), or both
- Slow fading: codeword spans one channel realization
- For point-to-point MIMO, four basic cases (reality inbetween)

	Fast fading	Slow fading
Channel unknown at TX	V-BLAST optimal no coding across antennas	V-BLAST suboptimal coding across antennas D-BLAST optimal
Channel known at TX	waterfilling over space& time	waterfilling over space

Slow fading, full CSI at TX

- Deterministic channel, known at TX and RX

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e} \qquad \boldsymbol{H} \in \mathbb{C}^{n_r \times n_t}$$

- Power constraint: $E[||\boldsymbol{x}||^2] \leq P$
- $\boldsymbol{e}$ is white noise, $N(\mathbf{0}, N_0\boldsymbol{I})$
- SVD: $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H, \quad \boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}, \quad \boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{I}$
- Dimensions: $\boldsymbol{U} \in \mathbb{C}^{n_r \times n_r}, \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n_r \times n_t}, \quad \boldsymbol{V} \in \mathbb{C}^{n_t \times n_t}$

- Introduce transform: $\boldsymbol{y} = \underbrace{\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H}_{\boldsymbol{H}}\boldsymbol{x} + \boldsymbol{e}$

$$\tilde{\boldsymbol{y}} \triangleq \boldsymbol{U}^H\boldsymbol{y} = \boldsymbol{\Lambda}\underbrace{\boldsymbol{V}^H\boldsymbol{x}}_{\triangleq\tilde{\boldsymbol{x}}} + \underbrace{\boldsymbol{U}^H\boldsymbol{e}}_{\tilde{\boldsymbol{e}}}$$

- $\tilde{\boldsymbol{e}}$ is white noise, $N(\boldsymbol{0}, N_0\boldsymbol{I})$
- $\tilde{\boldsymbol{x}}$ is precoded TX data. Note: $E[||\tilde{\boldsymbol{x}}||^2] = E[||\boldsymbol{x}||^2]$
- Equivalent model with parallel channels: $(n = \text{rank}(\boldsymbol{H}))$

$$\tilde{y}_1 = \lambda_1\tilde{x}_1 + \tilde{e}_1$$

$$\ldots$$

$$\tilde{y}_n = \lambda_n\tilde{x}_n + \tilde{e}_n$$

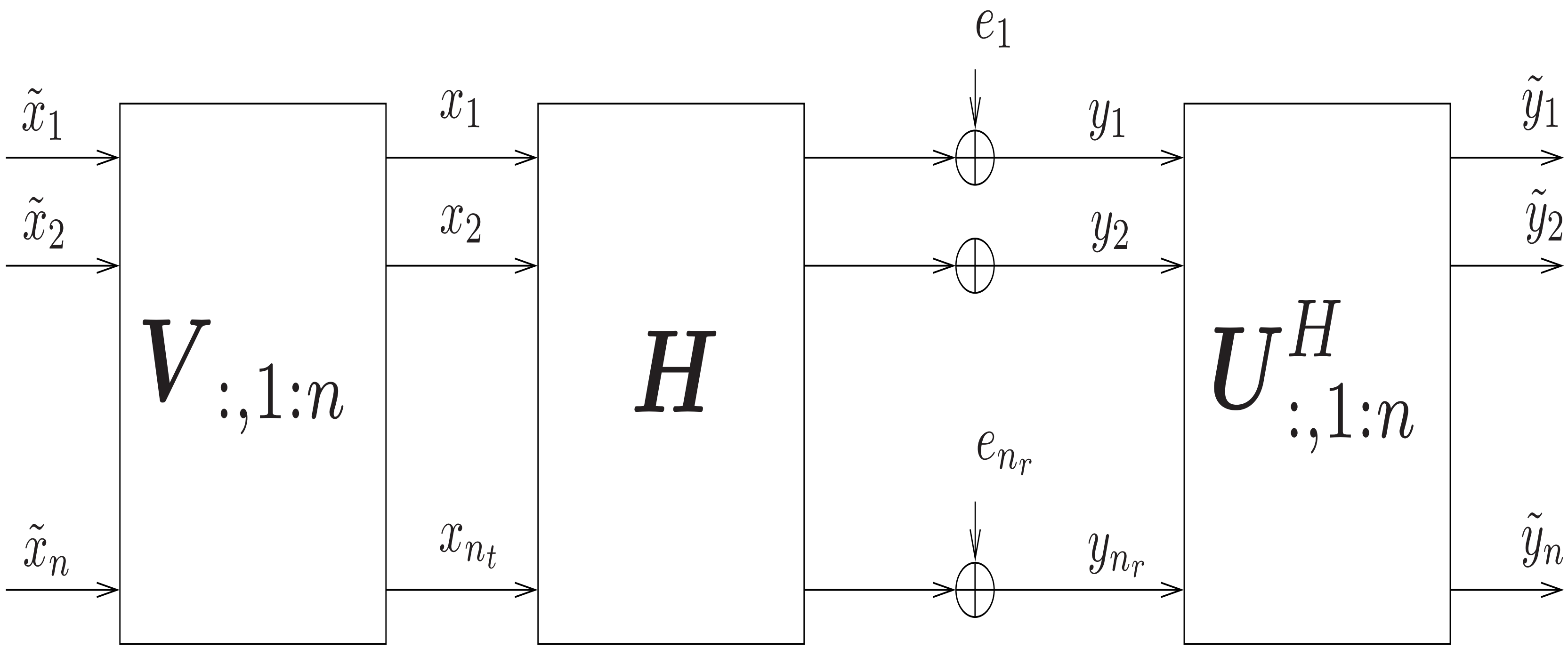
- No gain by coding across streams ➠ $\tilde{x}_k$ independent
- Operational meaning is multistream beamforming:

$$\boldsymbol{x} = \boldsymbol{V}\tilde{\boldsymbol{x}} = \sum_{k=1}^{n} \tilde{x}_k \boldsymbol{v}_k$$

  - ➠ n indep. streams $\{\tilde{x}_k\}$ are transmitted over orthogonal beams $\{\boldsymbol{v}_k\}$
  - ➠ We'll assume that each stream $\tilde{x}_k$ has power P_k
  - ➠ In the special case of $n = 1$ (rank one channel), then we have MRT:

$$\boldsymbol{H} = \lambda \boldsymbol{u}\boldsymbol{v}^H \quad \Rightarrow \quad \boldsymbol{x} = \tilde{x}\boldsymbol{v}$$

Architecture



- $\tilde{x}_k$ are independent streams with powers P_k and rates R_k

Optimal power allocation over n subchannels

- Each subchannel can offer the rate

$$C_k = \log_2\left(1 + \frac{P_k}{N_0}\lambda_k^2\right) \qquad \text{(bits/cu)}$$

- The power constraint is

$$E[||\tilde{\boldsymbol{x}}||^2] = E[||\boldsymbol{x}||^2] = \sum_{k=1}^{n} P_k \leq P$$

- Optimal power allocation $P_1^*, ..., P_n^*$

$$\max_{P_k, \sum_{k=1}^{n} P_k \leq P} \sum_{k=1}^{n} C_k \quad \Leftrightarrow \quad \boxed{\max_{P_k, \sum_{k=1}^{n} P_k \leq P} \sum_{k=1}^{n} \log_2\left(1 + \frac{P_k}{N_0}\lambda_k^2\right)}$$

Waterfilling solution

- Problem: $$\max_{P_k, \sum_{k=1}^{n} P_k \leq P} \sum_{k=1}^{n} \log_2\left(1 + \frac{P_k}{N_0}\lambda_k^2\right)$$

- Solution: (need solve for μ)

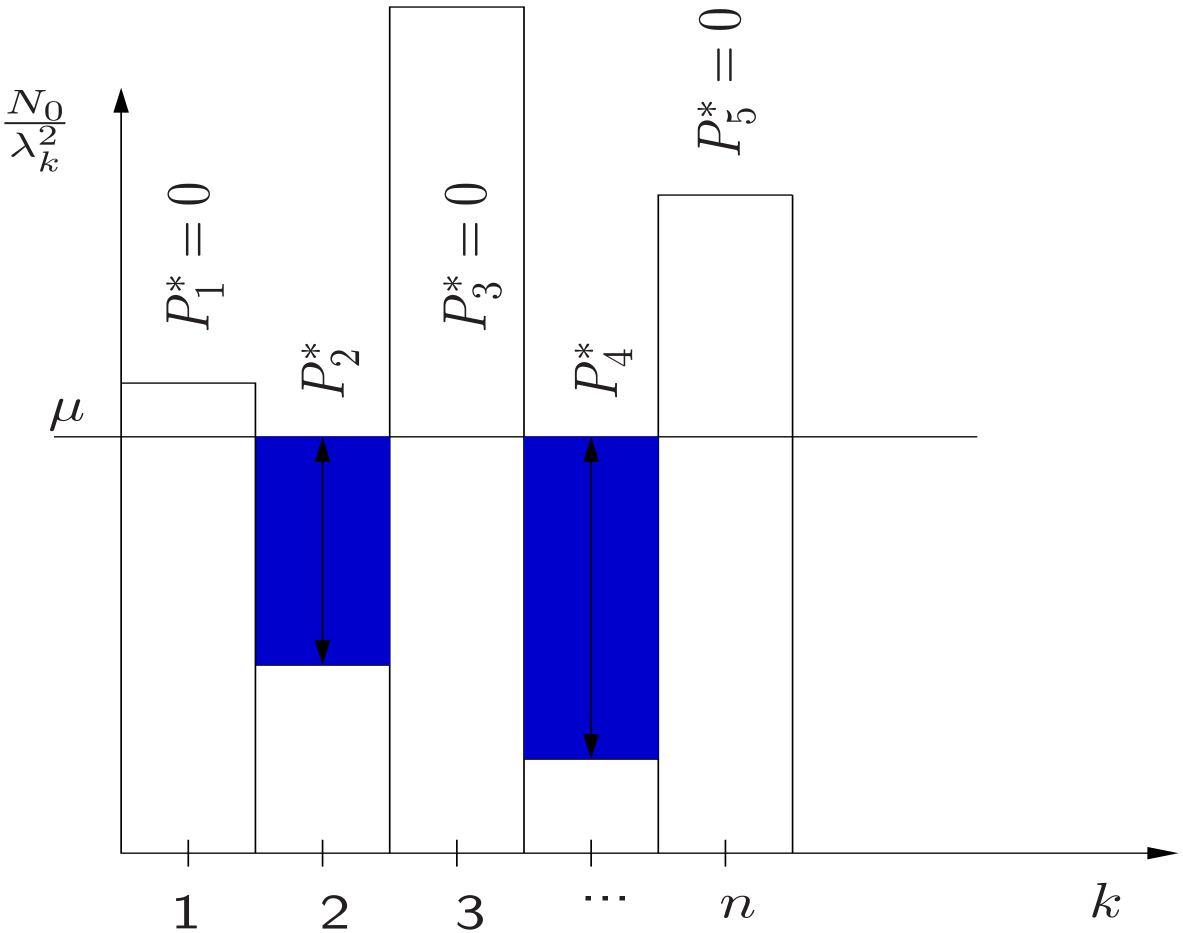
$$C = \sum_{k=1}^{n} \log_2\left(1 + \frac{P_k^* \lambda_k^2}{N_0}\right)$$

$$P_k^* = \left(\mu - \frac{N_0}{\lambda_k^2}\right)^+, \qquad \sum_{k=1}^{n} P_k^* = P$$

- Special case: $\boldsymbol{H} = \lambda \boldsymbol{u}\boldsymbol{v}^H$. Transmit one stream

$$C = \log_2\left(1 + \frac{P}{N_0}\lambda^2\right) = \log_2\left(1 + \frac{P}{N_0}\|\boldsymbol{H}\|^2\right)$$

Waterfilling solution



Waterfilling at high SNR

- At high SNR, the water is deep, so $P_k^* \approx \frac{P}{n}$ and

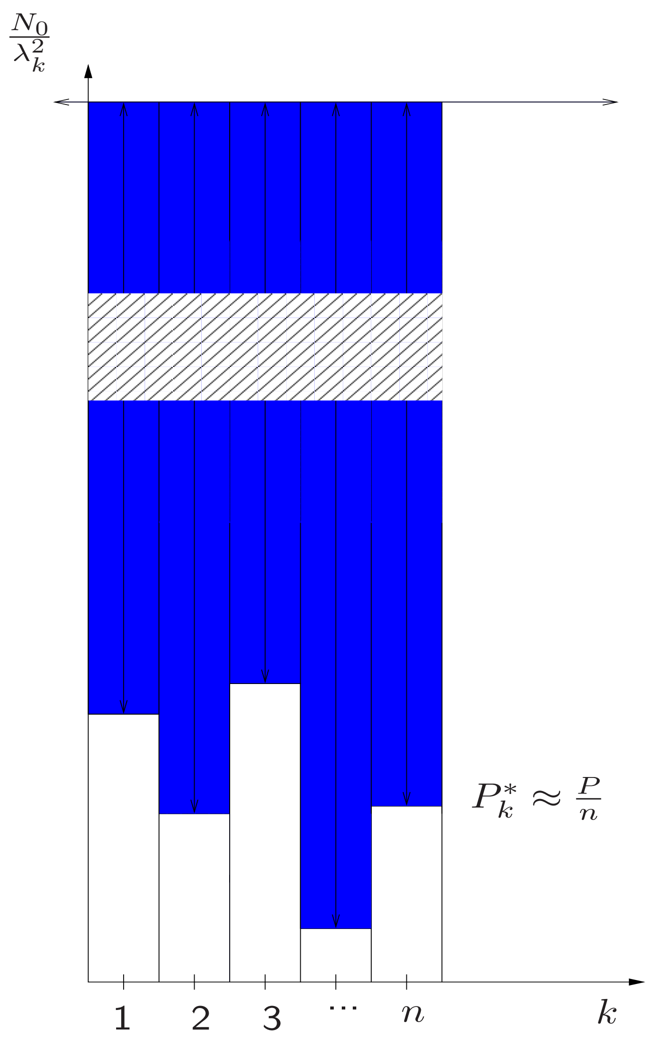
$$C = \sum_{k=1}^{n} C_k \approx \sum_{k=1}^{n} \log_2\left(1 + \frac{P}{N_0}\frac{\lambda_k^2}{n}\right) \approx n \cdot \log_2\left(\frac{P}{N_0}\right) + \sum_{k=1}^{n} \log_2\left(\frac{\lambda_k^2}{n}\right)$$

- With SNR $\triangleq P/N_0$, we have $C \sim n \log_2(\text{SNR})$. We say that

The channel offers $n = \text{rank}(\boldsymbol{H})$ **degrees of freedom** (DoF)

- Best capacity for well conditioned $\boldsymbol{H}$ (all λ_k's equal)
- Transmit n equipowered data streams spread on orthogonal beams
- Channel knowledge $\sim$unimportant
- No coding across streams

Waterfilling at high SNR



Waterfilling at low SNR

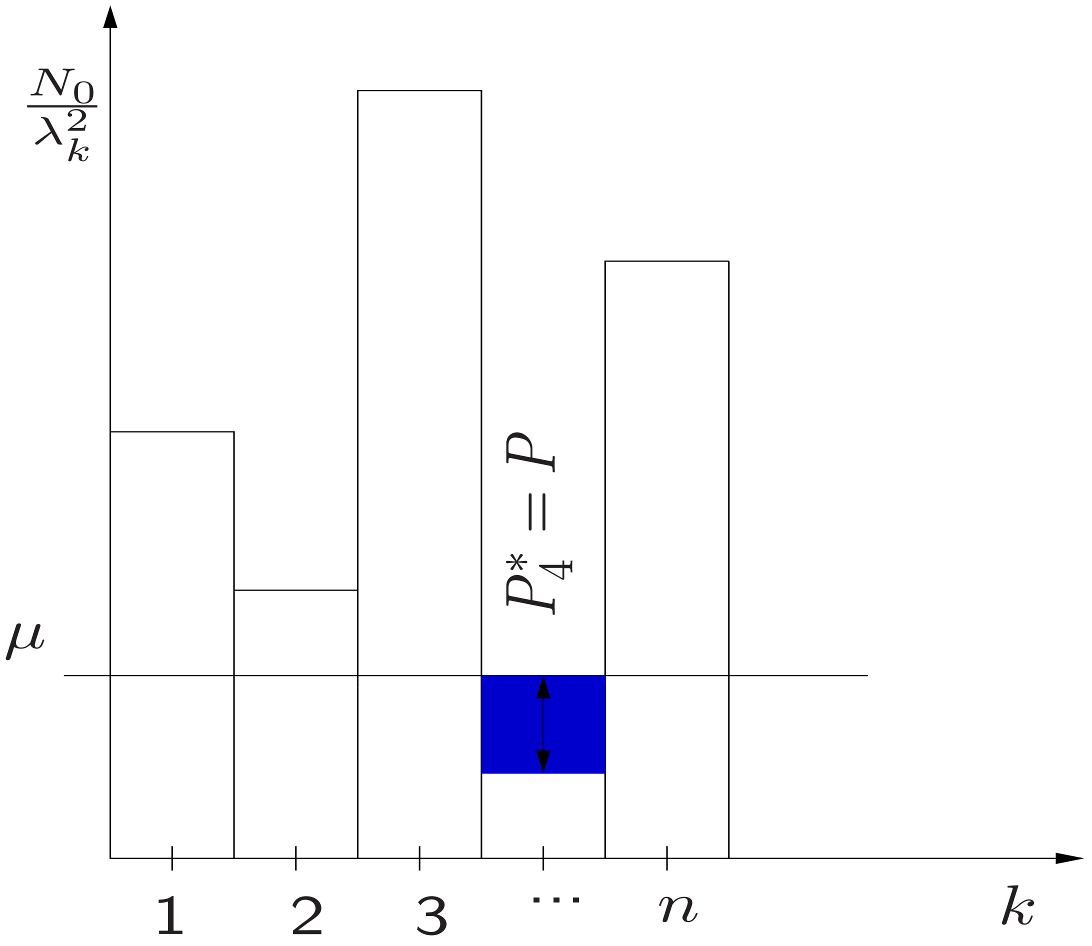
- At low SNR, the water is shallow. Then

$$P_k^* = \begin{cases} P, & k = \text{argmax} \ \lambda_k^2 \\ 0, & \text{else} \end{cases}$$

$$C = \log_2\left(1 + \frac{P}{N_0}\lambda_{\text{max}}^2\right) \approx \left(\frac{P}{N_0}\lambda_{\text{max}}^2\right) \cdot \log_2(e)$$

- MIMO provides an array gain (power gain of λ_{max}^2) but no DoF gains.
- Channel rank does not matter, only power matters.
- Transmit <u>one beam</u> in the direction associated with largest λ_k
- Knowing $\boldsymbol{H}$ is very important! (to select what beam to use)

Waterfilling at low SNR



In practice

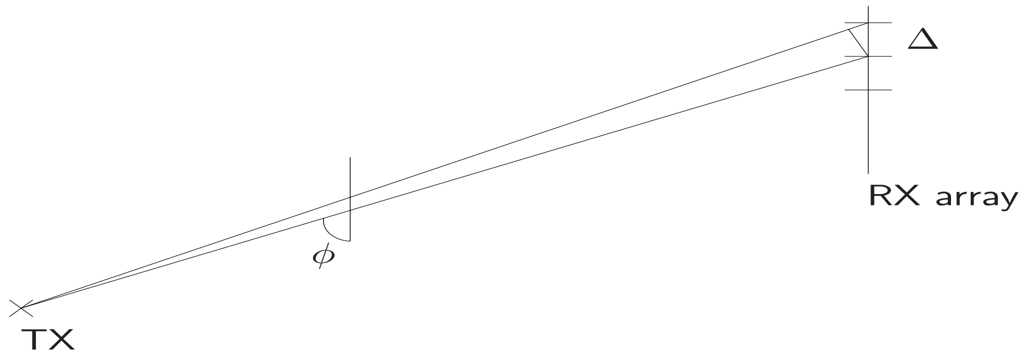
- Feedback of channel state information, requires quantization
- Potentially, by scheduling only "good" users, one may always operate at high SNR
- Selection of modulation scheme
  - — e.g., M-QAM per subchannel, different M
  - — better channel, larger constellation
  - — should be done with outer code in mind
- Imperfect CSI ➠ cross-talk!

MIMO channel models

- MIMO channel modeling is a rich research field, with both empirical (measurement) work and theoretical models.
- We will explore the main underlying physical phenomena of MIMO propagation and how they connect to the DoF concept.

Line-of-sight SIMO channel

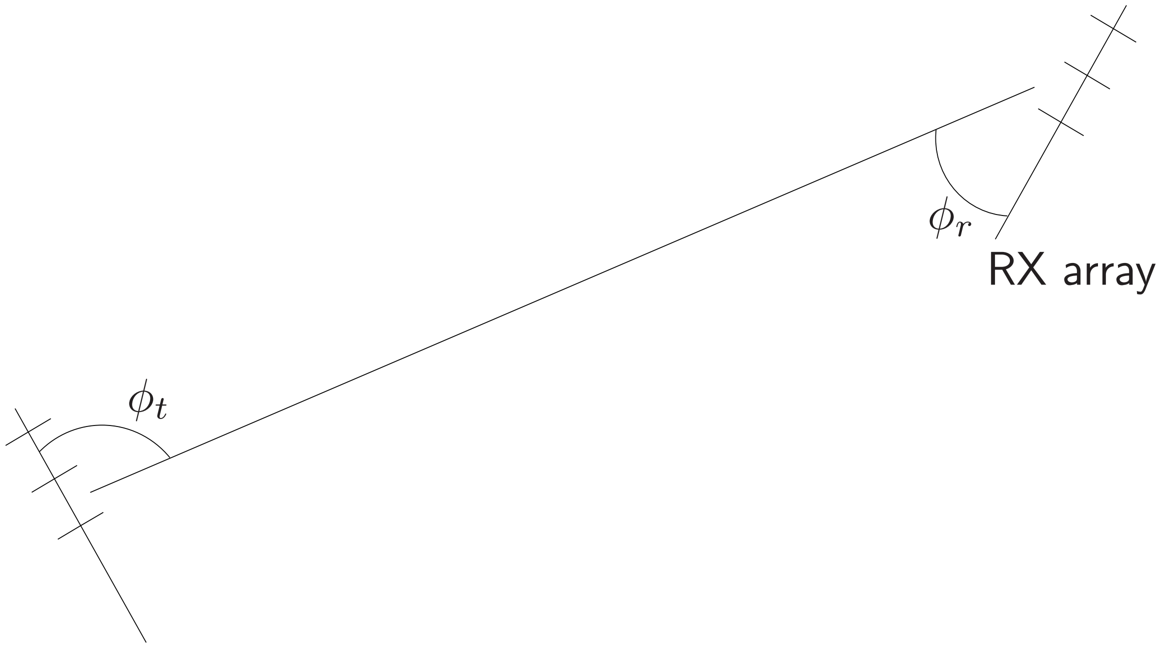
- Consider m-ULA at TX and RX, wavelength $\lambda = c/f_c$, ant. spacing Δ



- Let $\boldsymbol{u}(\phi) \triangleq \begin{bmatrix} 1 \\ e^{-j\frac{2\pi}{\lambda}\Delta\cos\phi} \\ \vdots \\ e^{-j(m-1)\frac{2\pi}{\lambda}\Delta\cos\phi} \end{bmatrix}$

- Signal from point source impinging on RX array (large TX-RX distance):
$\boldsymbol{y} = \alpha\boldsymbol{u}(\phi)\cdot s + \boldsymbol{e}, \qquad (\alpha \in \mathbb{C}, \text{ dep. on distance})$

Line-of-sight MIMO channel



- MIMO channel: $\boldsymbol{y} = \alpha \cdot \underbrace{\underbrace{\boldsymbol{u}(\phi_r)}_{n_r \times 1} \cdot \underbrace{\boldsymbol{u}(\phi_t)}_{n_t \times 1}^{H}}_{\boldsymbol{H}} \cdot \boldsymbol{x} + \boldsymbol{e}, \quad n = \mathrm{rank}(\boldsymbol{H}) = 1$

- The LoS-MIMO channel has rank one, so no DoF gain!

Lobes and resolvability

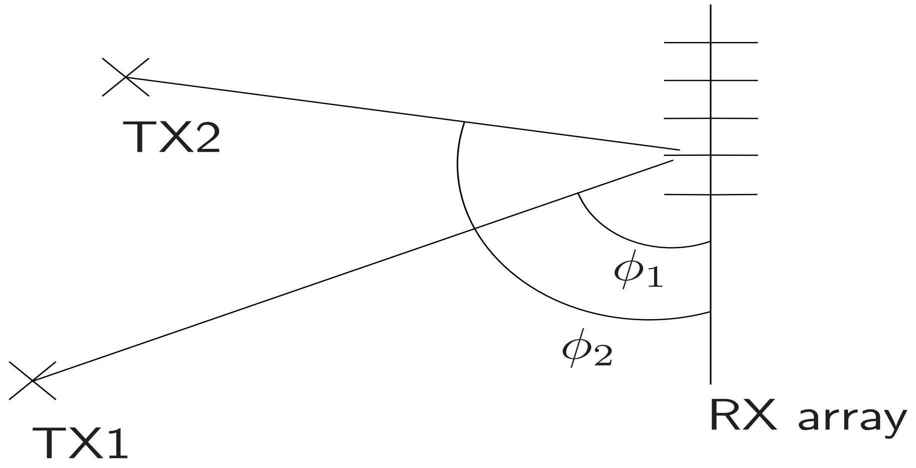
- Consider unit-power point sources at ϕ_1, ϕ_2 with sign. $\boldsymbol{u}(\phi_1)$, $\boldsymbol{u}(\phi_2)$. How similar do these signatures look?

$$\frac{1}{m}\|s_1\boldsymbol{u}(\phi_1) - s_2\boldsymbol{u}(\phi_2)\|^2 = 2 - 2\mathrm{Re}\left(s_1^* s_2 \cdot \underbrace{\frac{1}{m}\boldsymbol{u}^H(\phi_1)\boldsymbol{u}(\phi_2)}_{|\cdot|=f(\cdot)} \right)$$

  where the lobe pattern $f(\cos(\phi_1) - \cos(\phi_2)) \triangleq \frac{1}{m}|\boldsymbol{u}^H(\phi_1)\boldsymbol{u}(\phi_2)|$.

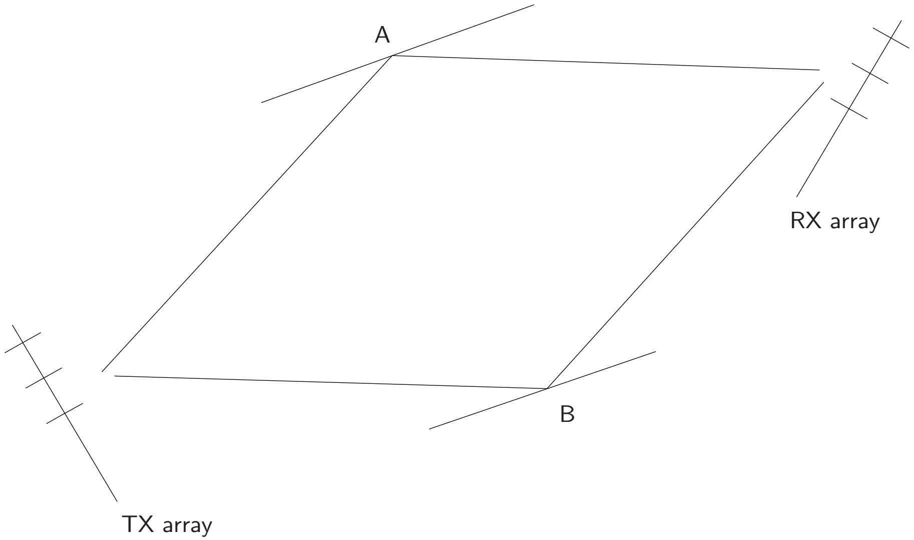
- If $f(\cdot) < 1$, then ϕ_1, ϕ_2 resolvable.
- Resolvability criterion: $|\cos\phi_1 - \cos\phi_2| \geq \frac{2\pi}{A}, \quad A \triangleq (m-1)\Delta$
- Grating lobes avoided if $\Delta \leq \dfrac{\lambda}{2} \Rightarrow A \leq (m-1)\dfrac{\lambda}{2}$.

Two separated point sources and an m-receive-array



- Define $\boldsymbol{H} = [\boldsymbol{h}_1 \quad \boldsymbol{h}_2]$, $\qquad \boldsymbol{h}_i = \alpha_i \boldsymbol{u}(\phi_i)$
- Condition number $\kappa(\boldsymbol{H}) = \dfrac{\lambda_{\text{max}}(\boldsymbol{H})}{\lambda_{\text{min}}(\boldsymbol{H})} = \sqrt{\dfrac{1 + f(\cos(\phi_1) - \cos(\phi_2))}{1 - f(\cos(\phi_1) - \cos(\phi_2))}}$
- $\kappa(\boldsymbol{H})$ is small if $f(\cdots) \neq 1 \Leftrightarrow \phi_1, \phi_2$ resolvable $\Leftrightarrow |\cos\phi_1 - \cos\phi_2| > \frac{2\pi}{A}$

MIMO with two plane scatters



- Here, $\boldsymbol{H}_{\text{TX-RX}} = \boldsymbol{H}_{\text{AB-RX}} \cdot \boldsymbol{H}_{\text{TX-AB}}$
- We have $\text{rank}(\boldsymbol{H}_{\text{TX-RX}}) = 2$
  only if $\text{rank}(\boldsymbol{H}_{\text{AB-RX}}) = 2$ and $\text{rank}(\boldsymbol{H}_{\text{TX-AB}}) = 2$
- For $\boldsymbol{H}$ to offer 2 DoF, A and B must be sufficiently separated in angle, as seen both from TX and RX

Angular decomposition of MIMO channel

- For $\phi_1, \phi_2, ..., \phi_m$ define

$$\boldsymbol{U} \triangleq \frac{1}{\sqrt{m}}[\boldsymbol{u}(\phi_1) \ \cdots \ \boldsymbol{u}(\phi_m)]$$

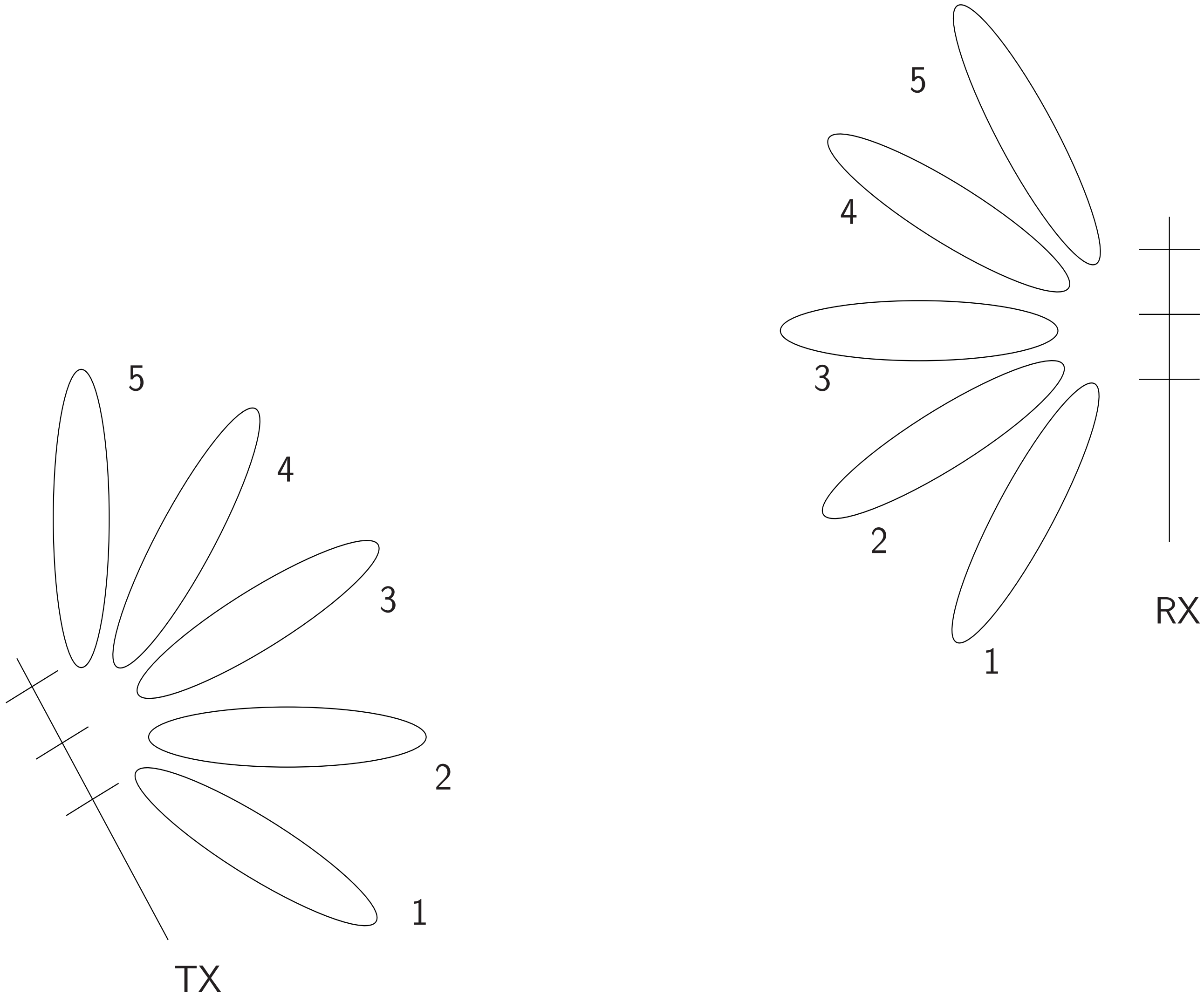
- Can show: With $\cos(\phi_k) = k/m$, $\{\boldsymbol{u}(\phi_k)\}$ forms ON-basis. Then $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}$.

- Let $\boldsymbol{U}_r$ and $\boldsymbol{U}_t$ be the $\boldsymbol{U}$ matrices associated with the TX and RX arrays. Note that $\boldsymbol{U}_r^H\boldsymbol{U}_r = \boldsymbol{I}$ and $\boldsymbol{U}_t^H\boldsymbol{U}_t = \boldsymbol{I}$

- If $\Delta = \lambda/2$, then $\boldsymbol{u}(\phi_i)$ correspond to simple, perfectly resolvable beams, with a single mainlobe.

- We assume $\Delta = \lambda/2$ from now on. The case of $\Delta \neq \lambda/2$ is more involved.

Angular decomposition, cont.



Angular decomposition, cont.

- Now define

$$\boldsymbol{H}_a \triangleq \boldsymbol{U}_r^H \boldsymbol{H} \boldsymbol{U}_t$$

$$\Rightarrow H_{a,(k,l)} = \boldsymbol{u}^H(\phi_k^r)\boldsymbol{H}\boldsymbol{u}(\phi_l^t)$$

$$\boldsymbol{u}^H(\phi_k^r) \underbrace{\left[\sum_i \alpha_i \boldsymbol{u}(\phi_i'^r)\boldsymbol{u}^H(\phi_i'^t)\right]}_{\text{physical model}} \boldsymbol{u}(\phi_l^t)$$

$$= \sum_i \alpha_i \left(\underbrace{\boldsymbol{u}^H(\phi_k^r)\boldsymbol{u}(\phi_i'^r)}_{=0 \text{ unless } \phi_i'^r \text{ falls in lobe } \phi_k^r} \right) \cdot \left(\underbrace{\boldsymbol{u}^H(\phi_i'^t)\boldsymbol{u}(\phi_l^t)}_{=0 \text{ unless } \phi_i'^t \text{ falls in lobe } \phi_l^t} \right)$$

- Elements of $\boldsymbol{H}_a$ correspond to different propagation paths

- $H_{a,(k,l)}$=gain of ray going out in TX lobe l and arriving in RX lobe k

Angular decomposition, key points

- "Rich scattering" if all angular bins filled ($\boldsymbol{H}_a$ has "no zeros")
- "Diversity order"
  = measure of error resilience
  = number of propagation paths
  = number of nonzero elements in $\boldsymbol{H}_a$
- Number of DoF
  = rank($\boldsymbol{H}$)
  = rank($\boldsymbol{H}_a$)
- If $\boldsymbol{H}_{a,(k,l)}$ are i.i.d. then $H_{k,l}$ are i.i.d.
- With i.i.d. $\boldsymbol{H}_a$ and many terms in $\sum$, then we get i.i.d. Rayleigh fading.

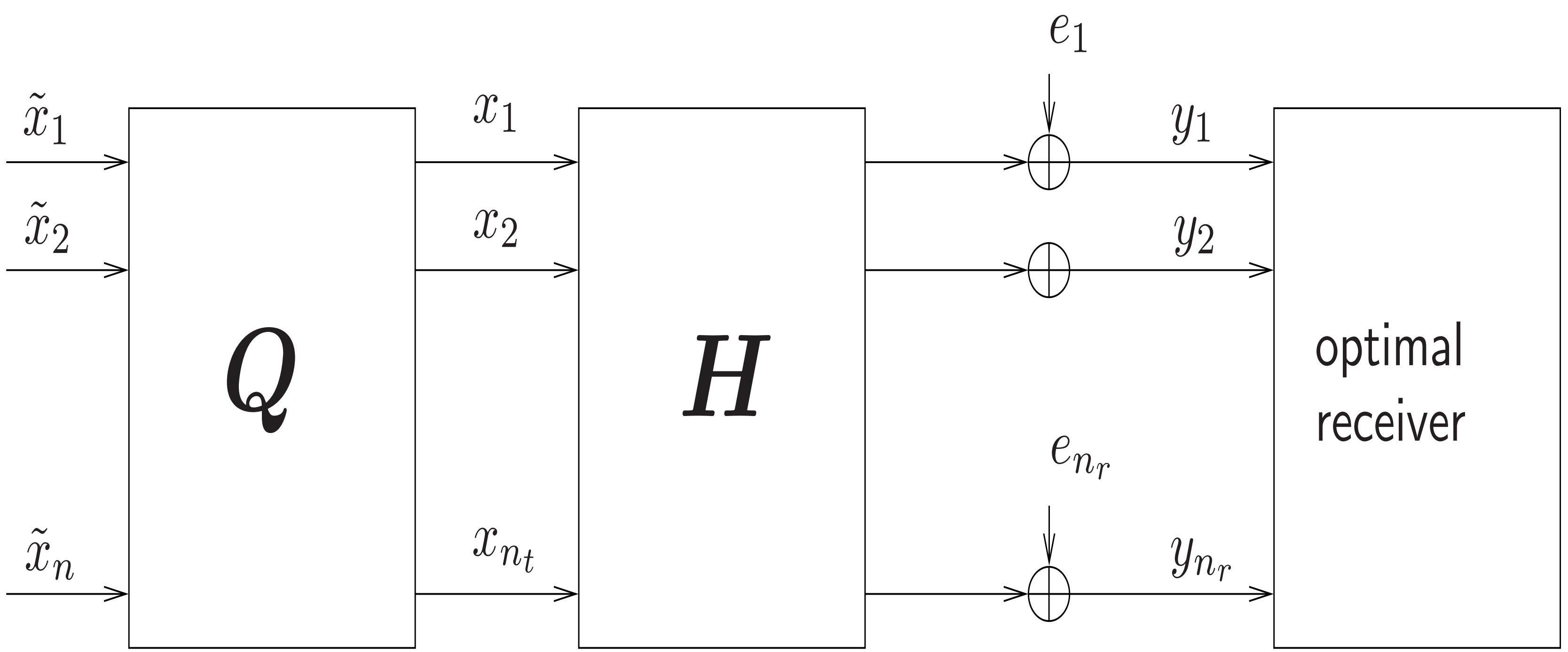
Fast fading, no CSI at TX

- Each codeword spans ∞ number of $\boldsymbol{H}$
- The V-BLAST architecture is optimal here
  Note: Reminiscent of architecture for slow fading and full CSI@TX
- Transmit vectors

$$\boldsymbol{x} = \boldsymbol{Q}\tilde{\boldsymbol{x}}$$

  where $\tilde{x}_1, ..., \tilde{x}_n$ are independent streams with powers P_k and rates R_k

V-BLAST architecture



- Transmit covariance: $\boldsymbol{K}_x \triangleq \text{cov}(\boldsymbol{x}) = \boldsymbol{Q} \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_n \end{bmatrix} \boldsymbol{Q}^H$

- Achievable rate, for fixed $\boldsymbol{H}$:

$$R = \log_2 \left| \boldsymbol{I} + \frac{1}{N_0} \boldsymbol{H} \boldsymbol{K}_x \boldsymbol{H}^H \right|$$

- Intuition: Volume of noise ball is $|N_0 \boldsymbol{I}|^N$.
  Volume of signal ball is $|\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}^H + N_o \boldsymbol{I}|^N$.

- Fast fading, coding over ∞ number of $\boldsymbol{H}$ matrices gives ergodic capacity

$$C = E\left[\log_2\left|\boldsymbol{I} + \frac{1}{N_0}\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}^H\right|\right]$$

- Choose $\boldsymbol{Q}$ and P_k to

$$\max_{\boldsymbol{K}_x, \mathrm{Tr}(\boldsymbol{K}_x)\leq P} E\left[\log_2\left|\boldsymbol{I} + \frac{1}{N_0}\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}^H\right|\right]$$

- Optimal $\boldsymbol{K}_x$ depends on the statistics of $\boldsymbol{H}$

- In i.i.d. Rayleigh fading, $\boldsymbol{K}_x^* = \frac{P}{n_t}\boldsymbol{I}$ (i.i.d. streams) and

$$C = E\left[\log_2\left|\boldsymbol{I} + \frac{P}{N_0}\frac{1}{n_t}\boldsymbol{H}\boldsymbol{H}^H\right|\right] = \sum_{k=1}^{n} E\left[\log_2\left(1 + \frac{\mathsf{SNR}}{n_t}\lambda_k^2\right)\right]$$

where

$$n = \mathsf{rank}(\boldsymbol{H}) = \min(n_r, n_t)$$

$$\mathsf{SNR} \triangleq \frac{P}{N_0}$$

$\{\lambda_k\}$ are the singular values of $\boldsymbol{H}$

  - Antennas then transmit separate streams.
  - Coding across antennas is unimportant.

Some special cases

- SISO: $n_t = n_r = 1$

$$C = E\left[\log_2(1 + \text{SNR}|h|^2)\right]$$

At high SNR, the loss is -0.83 bpcu relative to AWGN channel

- SIMO: $n_t = 1$ (power gain relative to SISO)

$$C = E\left[\log_2\left(1 + \text{SNR}\sum_{k=1}^{n_r}|h_k|^2\right)\right]$$

- MISO: $n_r = 1$ (no power gain relative to SISO)

$$C = E\left[\log_2\left(1 + \frac{\text{SNR}}{n_t}\sum_{k=1}^{n_t}|h_k|^2\right)\right]$$

Large arrays (infinite apertures)

- Large MISO (n_t TX, 1 RX) becomes AWGN channel:

$$C = E\left[\log_2\left(1 + \frac{\mathsf{SNR}}{n_t}\sum_{k=1}^{n_t}|h_k|^2\right)\right] \rightarrow \log_2(1+\mathsf{SNR})$$

- Large SIMO (1 TX, n_r RX)

$$C = E\left[\log_2\left(1 + \mathsf{SNR}\sum_{k=1}^{n_r}|h_k|^2\right)\right] \approx \log_2(n_r\mathsf{SNR}) = \log_2(n_r)+\log_2(\mathsf{SNR})$$

- Large square MIMO (n_t TX, n_r RX, $n_r = n_t = n$): Linear incr. with n:

$$C \approx n \cdot \left(\frac{1}{\pi}\int_0^4 \left(\log_2(1+t\cdot\mathsf{SNR})\sqrt{\frac{1}{t}-\frac{1}{4}}\right)dt\right)$$

Fast fading, no CSI at TX, high SNR

- Here

$$C = \sum_{k=1}^{n} E\left[\log_2\left(1 + \frac{\text{SNR}}{n_t}\lambda_k^2\right)\right] \approx n \log_2(\text{SNR}) + \text{const}$$

- Both n_r and n_t must be large to provide DoF gain

- "Capacity grows as $\min(n_r, n_t)$"

Fast fading, no CSI at TX, low SNR

- Here

$$C = \sum_{k=1}^{n} E\left[\log_2\left(1 + \frac{\mathsf{SNR}}{n_t}\lambda_k^2\right)\right] \approx \log_2(e) \cdot \frac{\mathsf{SNR}}{n_t} \cdot \sum_{k=1}^{n} E[\lambda_k^2]$$

$$= \log_2(e) \cdot \frac{\mathsf{SNR}}{n_t} \cdot \underbrace{E[||\boldsymbol{H}||^2]}_{=n_r n_t} = \log_2(e) \cdot n_r \cdot \mathsf{SNR}$$

- Capacity independent of n_t!
- No DoF gain. All what matters here is **power**
- Relative to SISO, a power gain of n_r (array/beamforming gain)
- Multiple TX antennas do not help here
  (but with CSI at TX, things are very different)

V-BLAST in practice

- Transmitter architecture "simple" but the **receiver** must separate the streams ➠ major challenge
- Problems are conceptually similar to uplink MUD in CDMA and to equalization for ISI channels
- Stream-by-stream receivers: Successive-interference-cancellation
  - ➠ MMSE-SIC is theoretically optimal but suffers from error propagation
  - ➠ Rate allocation necessary
- Iterative architectures
  - ➠ Iteration between outer code and demodulator
  - ➠ Demodulator design is major problem
- Receivers for MIMO to be discussed more in lecture 3

Fast fading, full CSI at TX

- The transmitter can do waterfilling over both space and time
- Parallel channels: $\tilde{y}_k[m] = \lambda_k[m]\tilde{x}_k[m] + \tilde{e}_k[m]$
  - Waterfilling over space (k) and time (m).
  - Optimal powers $P_k^*[m]$
  - Capacity

$$C = \sum_{k=1}^{n} E\left[\log_2\left(1 + \frac{P^*(\lambda_k)\lambda_k^2}{N_0}\right)\right]$$

- High SNR: $P^*(\lambda_k) \approx \frac{P}{n}$ (equal power)

$$C \approx \sum_{k=1}^{n} E\left[\log_2\left(1 + \frac{\mathsf{SNR}}{n}\lambda_k^2\right)\right], \qquad n \text{ D.o.F.}$$

An SNR gain (compared to no CSI) of

$$\frac{n_t}{n} = \frac{n_t}{\min(n_t, n_r)} = \frac{n_t}{n_r}, \quad \text{if } n_t \geq n_r$$

- Low SNR: Even larger gain, so here multiple antennas do help!

Slow fading, no CSI at TX

- Reliable communication for fixed $\boldsymbol{H}$ if $\log_2\left|\boldsymbol{I}+\frac{1}{N_0}\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}^H\right| > R$
- Outage probability, for fixed R: $P_{out} = P\left[\log_2\left|\boldsymbol{I}+\frac{1}{N_0}\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}^H\right| < R\right]$
- Optimal $\boldsymbol{K}_x$ as function of $\boldsymbol{H}$'s statistics:

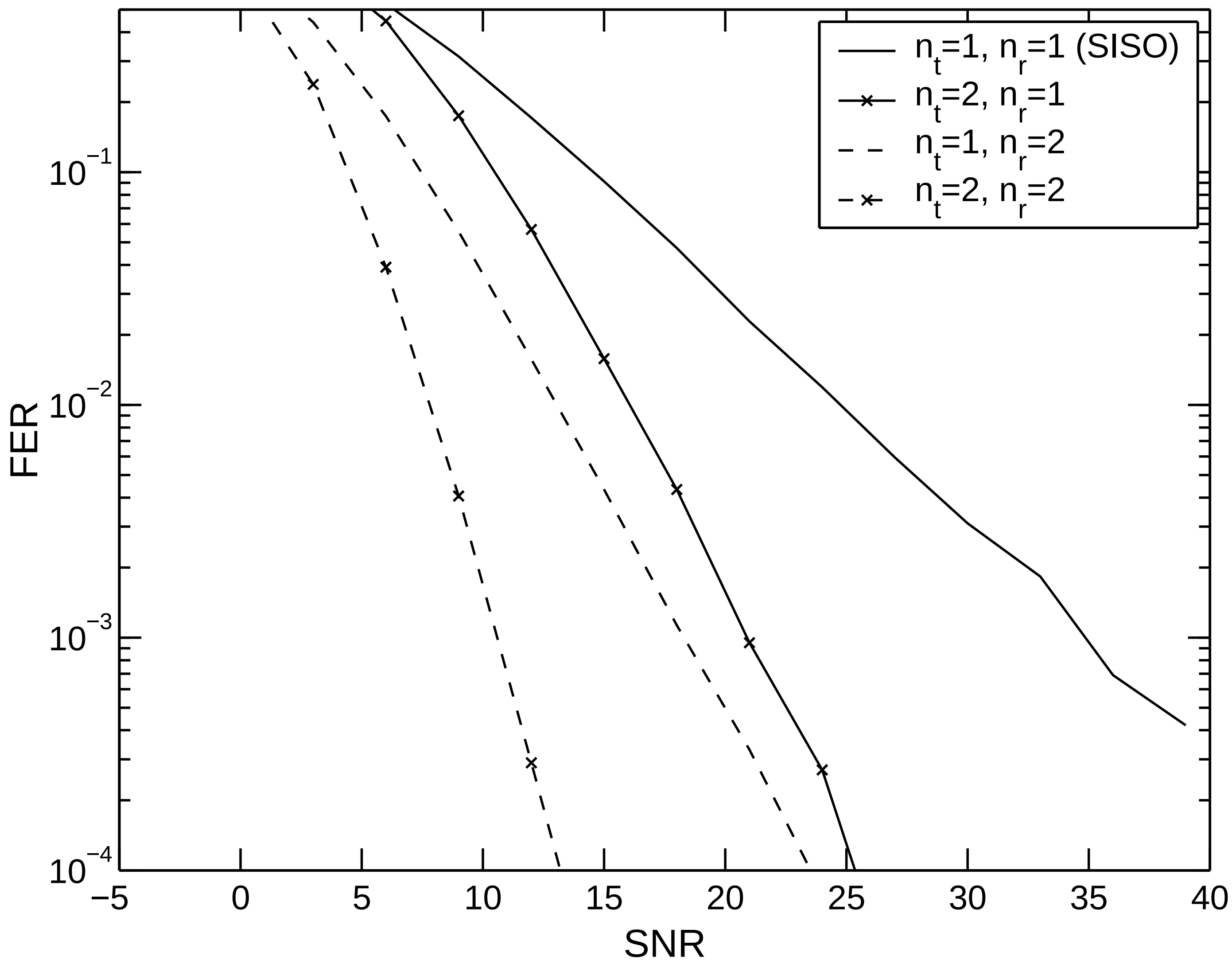
$$\boldsymbol{K}_x^* = \underset{\boldsymbol{K}_x, \mathrm{Tr}\,\boldsymbol{K}_x \le P}{\operatorname{argmin}} P\left[\log_2\left|\boldsymbol{I}+\frac{1}{N_0}\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}^H\right| < R\right]$$

For $\boldsymbol{H}$ i.i.d. Rayleigh fading:

- $\boldsymbol{K}_x^* = \frac{P}{n_t}\boldsymbol{I}$ optimal at large SNR
- $\boldsymbol{K}_x^* = \frac{P}{n'}\mathrm{diag}\{1, ..., 1, 0, ..., 0\}$ at low SNR ($n' < n_t$)

- Notion of **diversity**: P_{out} behaves as SNR^{-d} where d=diversity order
- Maximal diversity: $d = n_r n_t$
- To achieve diversity, we need **coding across streams**
- V-BLAST does not work here.
  Each stream has diversity at most n_r, while the channel offers $n_r n_t$
- Architectures for slow fading:
  - Theoretically, D-BLAST is optimal
  - Pragmatic approaches include STBC combined with FEC

Example: Outage probability at rate $R = 2$ bpcu



D-BLAST architecture

	$\boldsymbol{x}_B(1)$	$\boldsymbol{x}_B(2)$		
$\boldsymbol{x}_A(1)$	$\boldsymbol{x}_A(2)$	$\boldsymbol{x}_A(3)$		

- Decoding in steps:
  1. Decode $\boldsymbol{x}_A(1)$
  2. Decode $\boldsymbol{x}_B(1)$, suppressing $\boldsymbol{x}_A(2)$ via MMSE
  3. Strip off $\boldsymbol{x}_B(1)$, and decode $\boldsymbol{x}_A(2)$
  4. Decode $\boldsymbol{x}_B(2)$, suppressing $\boldsymbol{x}_A(3)$ via MMSE
- One codeword: $\boldsymbol{x}(i) = [\boldsymbol{x}_A(i) \;\; \boldsymbol{x}_B(i)]$
- Requires appropriate rate allocation among $\boldsymbol{x}_A(i), \boldsymbol{x}_B(i)$
- In practice, error propagation and rate loss due to initialization

Le 2: Low-complexity MIMO

Antenna diversity basics

- Recall transmission model (single time interval):

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_{n_r} \end{bmatrix}}_{\boldsymbol{y}\ \text{(RX data)}} = \underbrace{\begin{bmatrix} h_{1,1} & \cdots & h_{1,n_t} \\ \vdots & & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix}}_{\boldsymbol{H}\ \text{(channel)}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_{n_t} \end{bmatrix}}_{\boldsymbol{x}\ \text{(TX data)}} + \underbrace{\begin{bmatrix} e_1 \\ \vdots \\ e_{n_r} \end{bmatrix}}_{\boldsymbol{e}\ \text{(noise)}}$$

- Transmission model (N time intervals):

$$\underbrace{\begin{bmatrix} y_{1,1} & \cdots & y_{1,N} \\ \vdots & & \vdots \\ y_{n_r,1} & \cdots & y_{n_r,N} \end{bmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{bmatrix} h_{1,1} & \cdots & h_{1,n_t} \\ \vdots & & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix}}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} x_{1,1} & \cdots & x_{1,N} \\ \vdots & & \vdots \\ x_{n_t,1} & \cdots & x_{n_t,N} \end{bmatrix}}_{\substack{\boldsymbol{X}\in\mathcal{X} \\ \text{“code matrix”}}} + \underbrace{\begin{bmatrix} e_{1,1} & \cdots & e_{1,N} \\ \vdots & & \vdots \\ e_{n_r,1} & \cdots & e_{n_r,N} \end{bmatrix}}_{\boldsymbol{E}}$$

Introduction and preliminaries

- Transmitting with low error probability at fixed rate requires N large.
- For practical systems, it is often of interest to design short space-time blocks (small N) with good error probability performance. Outer FEC can then be used over these blocks.
- Throughout, we will assume Gaussian noise,

$$\boldsymbol{e} \sim N(0, N_0 \boldsymbol{I})$$

Usually, we assume i.i.d. Rayleigh fading,

$$H_{i,j} \qquad \text{i.i.d. } N(0,1)$$

Sometimes, for SIMO/MISO, we take

$$\boldsymbol{h} \sim N(\boldsymbol{0}, \boldsymbol{Q})$$

Receive diversity ($n_t = 1$)

- Suppose s transmitted, and $\boldsymbol{h}$ known at RX.
- Receive: $\boldsymbol{y} = \boldsymbol{h}s + \boldsymbol{e}$
- Detection of s via maximum-likelihood (in AWGN):

$$\|\boldsymbol{y} - \boldsymbol{h}s\|^2 = ... = \|\boldsymbol{h}\|^2 \cdot \left|s - \hat{s}\right|^2 + \text{const.}, \quad \text{where} \quad \hat{s} \triangleq \frac{\boldsymbol{h}^H \boldsymbol{y}}{\|\boldsymbol{h}\|^2}$$

- MRC+scalar detection problem!
- Distribution of $\hat{s}$ determines performance:

$$\hat{s} \triangleq \frac{\boldsymbol{h}^H \boldsymbol{y}}{\|\boldsymbol{h}\|^2} \sim N\left(s, \frac{N_0}{\|\boldsymbol{h}\|^2}\right), \qquad \text{SNR}|_{\boldsymbol{h}} = \frac{\|\boldsymbol{h}\|^2}{N_0} \cdot \underbrace{E\left[|s|^2\right]}_{=P} = \|\boldsymbol{h}\|^2 \cdot \underbrace{\frac{P}{N_0}}_{\text{SNR}}$$

Diversity order ($n \times 1$ fading vector $\boldsymbol{h}$)

- $P(e|\boldsymbol{h}) = Q\left(\sqrt{\mathsf{SNR} \cdot \|\boldsymbol{h}\|^2}\right)$ and $\boldsymbol{h} \sim N(\boldsymbol{0}, \boldsymbol{Q})$ (SNR up to a constant)

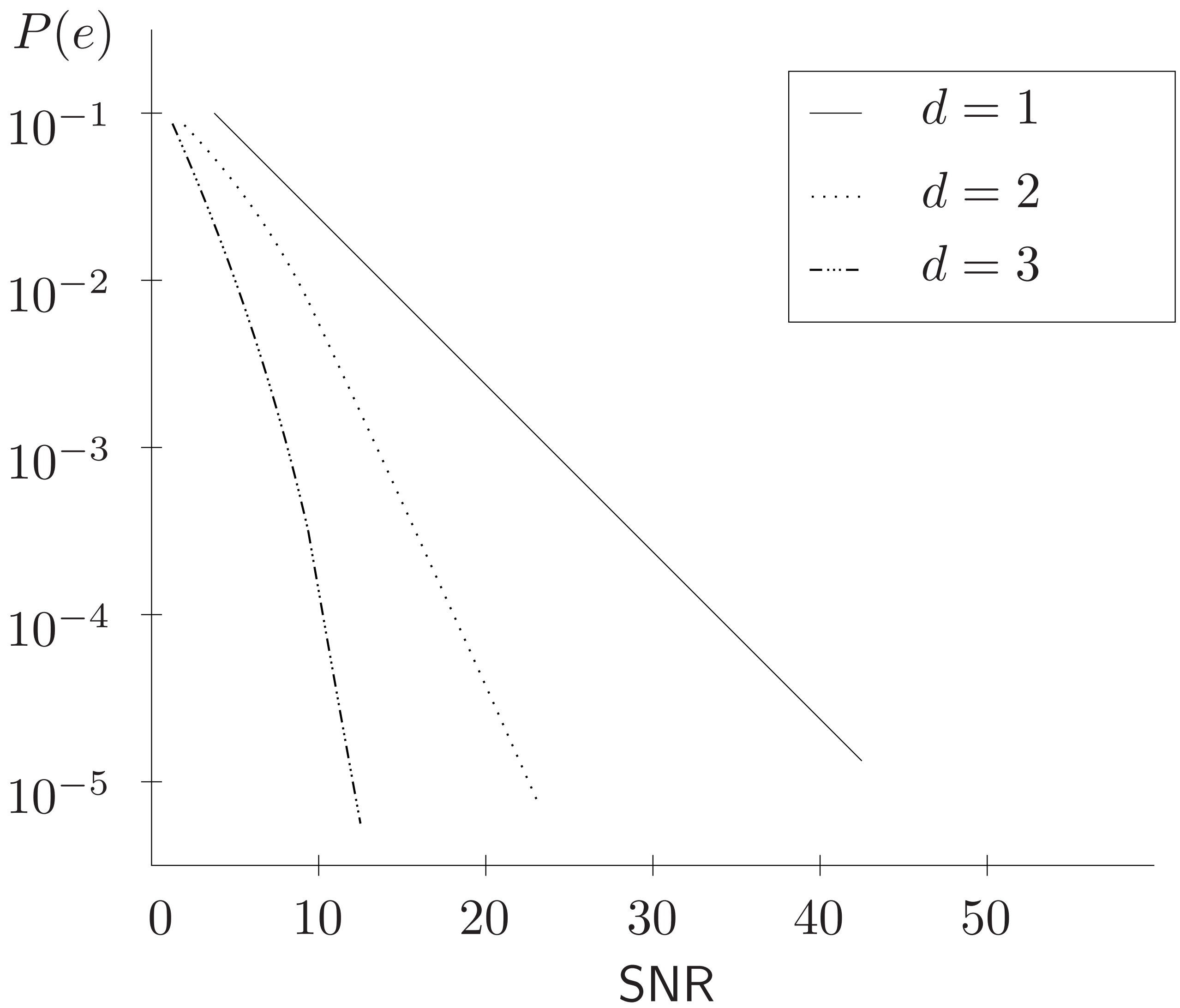
- Then $P(e) = E[P(e|\boldsymbol{h})] \leq \left|\boldsymbol{I} + \frac{\mathsf{SNR}}{2}\boldsymbol{Q}\right|^{-1} = \prod_{k=1}^{n}\left(1 + \frac{\mathsf{SNR}}{2}\lambda_k(\boldsymbol{Q})\right)^{-1}$

- As SNR $\to \infty$, $P(e) \leq \left(\frac{\mathsf{SNR}}{2}\right)^{-\operatorname{rank}(\boldsymbol{Q})} \cdot \frac{1}{\prod_{k=1}^{\operatorname{rank}(\boldsymbol{Q})} \lambda_k(\boldsymbol{Q})}$

- Diversity order $d \triangleq -\frac{\log P(e)}{\log \mathsf{SNR}} = \operatorname{rank}(\boldsymbol{Q})$

- Note that $\left(\prod_{k=1}^{n} \lambda_k\right)^{1/n} \leq \frac{1}{n}\sum_{k=1}^{n} \lambda_k = \frac{1}{n}\mathrm{Tr}\{\boldsymbol{Q}\} = \frac{1}{n}E[\|\boldsymbol{h}\|^2]$
  with eq. if $\lambda_1 = \cdots = \lambda_n$ so $\boldsymbol{Q} \propto \boldsymbol{I}$ minimizes the bound on $P(e)$

Diversity order



Transmit diversity, $\boldsymbol{H}$ known at transmitter

- Try transmit $\boldsymbol{w} \cdot s$ where $\boldsymbol{w}$ is function of $\boldsymbol{H}$! (as we did in Le 1)
- RX data is $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{w}s + \boldsymbol{e}$ and optimal decision minimizes the ML metric:

$$\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{w}s\|^2 = \ldots = \|\boldsymbol{H}\boldsymbol{w}\|^2 \cdot |s - \hat{s}|^2 + \text{const.}$$

where $\hat{s} \triangleq \dfrac{\boldsymbol{w}^H \boldsymbol{H}^H \boldsymbol{y}}{\|\boldsymbol{H}\boldsymbol{w}\|^2} \sim N\left(s, \dfrac{N_0}{\|\boldsymbol{H}\boldsymbol{w}\|^2}\right)$

- The SNR$|_{\boldsymbol{H}}$ in $\hat{s}$ is max for $\boldsymbol{w}$=normalized dominant RSV of $\boldsymbol{H}$
- Resulting SNR$|_{\boldsymbol{H}} = \frac{1}{N_0} \underbrace{\lambda_{\text{max}}(\boldsymbol{H}^H\boldsymbol{H})}_{\geq \|\boldsymbol{H}\|^2/n_t} \cdot E\left[|s|^2\right] \geq \frac{1}{N_0}\frac{\|\boldsymbol{H}\|^2}{n_t} \cdot E\left[|s|^2\right]$.
- Diversity order: $d = n_r n_t$
- For $n_r = 1$ take $\boldsymbol{w}_{\text{opt}} = \frac{\boldsymbol{h}^*}{\|\boldsymbol{h}\|}$ so SNR$|_{\boldsymbol{h}} = \frac{\|\boldsymbol{h}\|^2}{N_0} E\left[|s|^2\right]$ (same as for RX-d)

Transmit diversity, $\boldsymbol{H}$ unknown at transmitter

- From now on, TX does **not** know $\boldsymbol{H}$!
- Consider $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E}$. Optimal receiver in AWGN ($\boldsymbol{H}$ known at RX):

$$\max_{\boldsymbol{X}} P(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{H}) \Leftrightarrow \min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|^2$$

- Pairwise error probability $P(\boldsymbol{X}_0 \to \boldsymbol{X}|\boldsymbol{H}) = Q\left(\sqrt{\frac{\|\boldsymbol{H}(\boldsymbol{X}_0 - \boldsymbol{X})\|^2}{2N_0}}\right)$
- Consider $P(\boldsymbol{X}_0 \to \boldsymbol{X}) = E[P(\boldsymbol{X}_0 \to \boldsymbol{X}|\boldsymbol{H})]$. For i.i.d. Rayleigh fading,

$$P(\boldsymbol{X}_0 \to \boldsymbol{X}) \leq \underbrace{\left|\boldsymbol{I} + \frac{1}{4N_0}(\boldsymbol{X}_0 - \boldsymbol{X})(\boldsymbol{X}_0 - \boldsymbol{X})^H\right|^{-n_r}}_{\sim\left(\frac{1}{N_0}\right)^{-d} \sim \text{SNR}^{-d}}$$

d= "diversity order". Note: $d \leq n_r n_t$ and $d = n_r n_t$ if $\boldsymbol{X}_0 - \boldsymbol{X}$ full rank

Linear space-time block codes (STBC)

- STBC maps n_s complex symbols onto $n_t \times N$ matrix $\boldsymbol{X}$:

$$\{s_1, \ldots, s_{n_s}\} \rightarrow \boldsymbol{X}$$

- Linear STBC:

$$\boldsymbol{X} = \sum_{n=1}^{n_s} (\bar{s}_n \boldsymbol{A}_n + i\tilde{s}_n \boldsymbol{B}_n)$$

where $\{\boldsymbol{A}_n, \boldsymbol{B}_n\}$ are *fixed* matrices

- Typically N small. Need $N \geq n_t$ for max diversity (why?)
- Rate: $R \triangleq \frac{N}{n_s}$ bits/channel use

STBC with a single symbol

- Transmit one symbol s during N time intervals, weighted by $\boldsymbol{W}$:

$$\boldsymbol{X} = \boldsymbol{W} \cdot s, \qquad \boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E} = \boldsymbol{H}\boldsymbol{W}s + \boldsymbol{E}$$

- Average error probability in Rayleigh fading:

$$P(s_0 \to s) \leq |\boldsymbol{W}\boldsymbol{W}^H|^{-n_r} |s - s_0|^{-2n_r n_t} \left(\frac{1}{4N_0}\right)^{-n_r n_t}$$

- What is the optimum $\boldsymbol{W}$? Try to maximize:

$$\max_{\boldsymbol{W}} \quad |\boldsymbol{W}\boldsymbol{W}^H|$$

$$\text{s.t.} \quad \text{Tr}\left\{\boldsymbol{W}\boldsymbol{W}^H\right\} = \|\boldsymbol{W}\|^2 \leq 1 \qquad \text{(power constraint)}$$

- Solution: $\boldsymbol{W}\boldsymbol{W}^H = \frac{1}{n_t}\boldsymbol{I}$, (antenna cycling). Diversity but rate $1/N$!

Alamouti scheme for $n_t = 2$

- $\boldsymbol{X} = \frac{1}{\sqrt{2}}\begin{bmatrix} s_1 & s_2^* \\ s_2 & -s_1^* \end{bmatrix}$. That is:

	Time 1	Time 2
Ant 1	$s_1/\sqrt{2}$	$s_2^*/\sqrt{2}$
Ant 2	$s_2\sqrt{2}$	$-s_1^*/\sqrt{2}$

- RX data: $\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{h}_1 s_1 + \boldsymbol{h}_2 s_2 \\ \boldsymbol{h}_1 s_2^* - \boldsymbol{h}_2 s_1^* \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{bmatrix}$

- Consider $\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2^* \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{h}_1 s_1 + \boldsymbol{h}_2 s_2 \\ \boldsymbol{h}_1^* s_2 - \boldsymbol{h}_2^* s_1 \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2^* \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 \\ -\boldsymbol{h}_2^* & \boldsymbol{h}_1^* \end{bmatrix}\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2^* \end{bmatrix}$

- ML detector

$$\min_{s_1, s_2} \left\| \underbrace{\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2^* \end{bmatrix}}_{\boldsymbol{y}} - \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 \\ -\boldsymbol{h}_2^* & \boldsymbol{h}_1^* \end{bmatrix}}_{\triangleq \boldsymbol{G}} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{\boldsymbol{s}} \right\|^2$$

- Observation: $\boldsymbol{G}^H\boldsymbol{G} = \frac{1}{2}\begin{bmatrix} \boldsymbol{h}_1^H & -\boldsymbol{h}_2^T \\ \boldsymbol{h}_2^H & \boldsymbol{h}_1^T \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 \\ -\boldsymbol{h}_2^* & \boldsymbol{h}_1^* \end{bmatrix} = \frac{\|\boldsymbol{h}_1\|^2+\|\boldsymbol{h}_2\|^2}{2}\boldsymbol{I}$

- Hence $\min \|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s}\|^2 \Leftrightarrow \min \|\hat{\boldsymbol{s}} - \boldsymbol{s}\|^2$, $\quad \hat{\boldsymbol{s}} = 2\frac{\boldsymbol{G}^H\boldsymbol{y}}{\|\boldsymbol{h}_1\|^2 + \|\boldsymbol{h}_2\|^2}$

- Distribution of $\hat{\boldsymbol{s}}$:

$$\hat{\boldsymbol{s}} = 2\frac{\boldsymbol{G}^H\boldsymbol{y}}{\|\boldsymbol{h}_1\|^2 + \|\boldsymbol{h}_2\|^2} = 2\frac{\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{s}+\boldsymbol{e})}{\|\boldsymbol{h}_1\|^2 + \|\boldsymbol{h}_2\|^2} \sim N\left(\boldsymbol{s}, \frac{2N_0}{\|\boldsymbol{H}\|^2}\boldsymbol{I}\right)$$

- $\text{SNR}|_{\boldsymbol{H}} = \frac{\|\boldsymbol{H}\|^2}{2N_0}$. For 2×1 system, 3 dB less than 1×2 with MRC

- Diversity order: $2n_r$

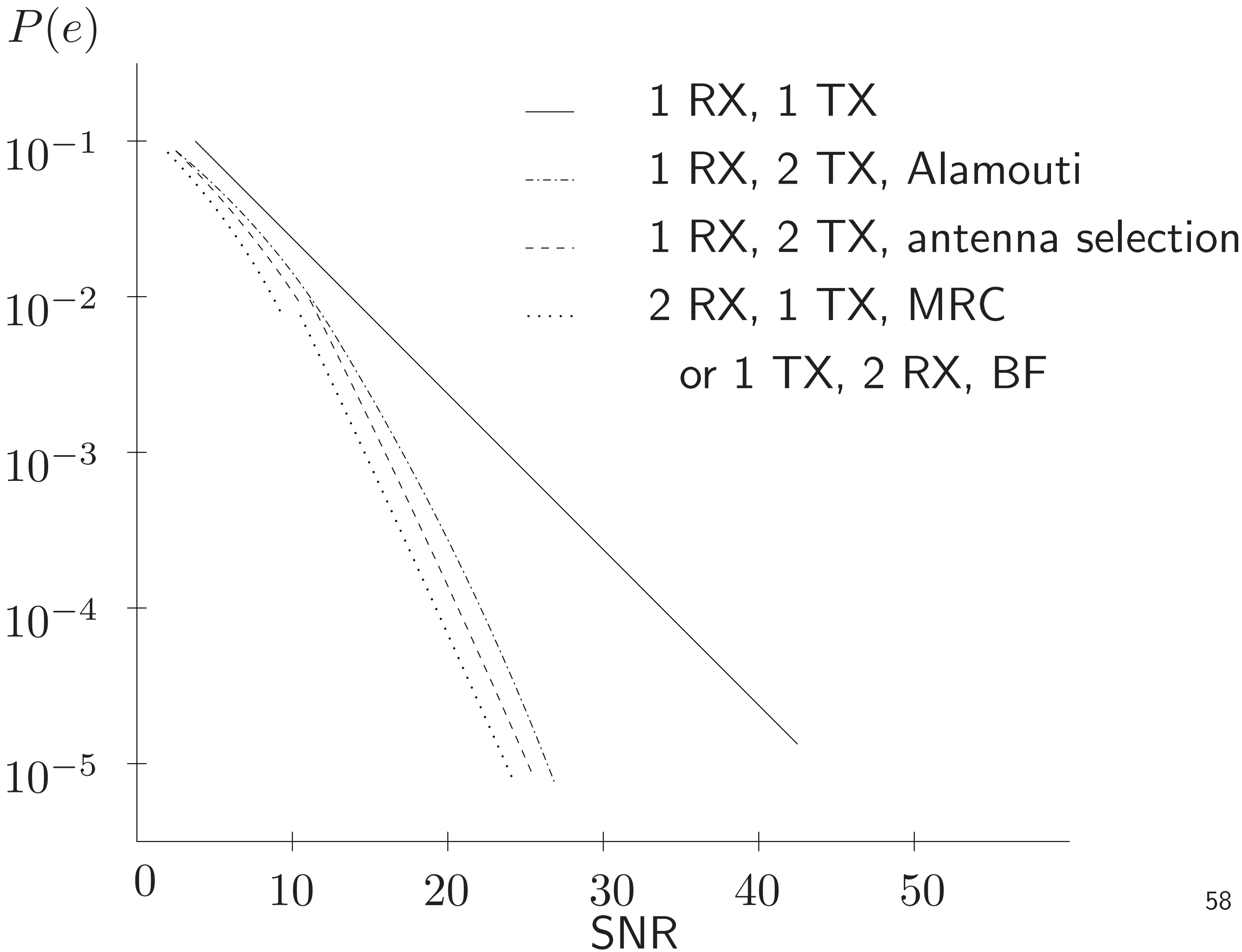
Overview of 2-antenna systems

Method	SNR	rate	TX knows h_1, h_2
1 TX, 2 RX, MRC	$\frac{\lvert h_1\rvert^2 + \lvert h_2\rvert^2}{N_0}$	1	no
2 TX, 1 RX, BF	$\frac{\lvert h_1\rvert^2 + \lvert h_2\rvert^2}{N_0}$	1	yes
2 TX, 1 RX, ant. cycl.	$\frac{\lvert h_1\rvert^2 + \lvert h_2\rvert^2}{2N_0}$	1/2	no
2 TX, 1 RX, Alamouti	$\frac{\lvert h_1\rvert^2 + \lvert h_2\rvert^2}{2N_0}$	1	no
2 TX, 1 RX, Ant. sel.	$\geq \frac{\lvert h_1\rvert^2 + \lvert h_2\rvert^2}{2N_0}$	1	partly

- For antenna selection, note that

$$\frac{\max |h_n|^2}{N_0} \geq \frac{1}{2}\frac{|h_1|^2 + |h_2|^2}{N_0}$$

2-antenna systems, cont



Orthogonal STBC (OSTBC)

- Important special case of linear STBC:

$$\boldsymbol{X} = \sum_{n=1}^{n_s} (\bar{s}_n \boldsymbol{A}_n + i\tilde{s}_n \boldsymbol{B}_n) \quad \text{for which} \quad \boxed{\boldsymbol{X}\boldsymbol{X}^H = \sum_{n=1}^{n_s} |s_n|^2 \cdot \boldsymbol{I} = \|\boldsymbol{s}\|^2 \cdot \boldsymbol{I}}$$

Notation: $\bar{(\cdot)}$=real part, $\tilde{(\cdot)}$=imaginary part

- This is equivalent to requiring for $n = 1, \ldots, n_s, \quad p = 1, \ldots, n_s$

$$\boldsymbol{A}_n \boldsymbol{A}_n^H = \boldsymbol{I}, \boldsymbol{B}_n \boldsymbol{B}_n^H = \boldsymbol{I}$$

$$\boldsymbol{A}_n \boldsymbol{A}_p^H = -\boldsymbol{A}_p \boldsymbol{A}_n^H, \quad \boldsymbol{B}_n \boldsymbol{B}_p^H = -\boldsymbol{B}_p \boldsymbol{B}_n^H, \quad n \neq p$$

$$\boldsymbol{A}_n \boldsymbol{B}_p^H = \boldsymbol{B}_p \boldsymbol{A}_n^H$$

Proof

- To prove $\Rightarrow$), expand:

$$\begin{aligned} \boldsymbol{X}\boldsymbol{X}^H &= \sum_{n=1}^{n_s}\sum_{p=1}^{n_s}(\bar{s}_n\boldsymbol{A}_n + i\tilde{s}_n\boldsymbol{B}_n)(\bar{s}_p\boldsymbol{A}_p + i\tilde{s}_p\boldsymbol{B}_p)^H \\ &= \sum_{n=1}^{n_s}(\bar{s}_n^2\boldsymbol{A}_n\boldsymbol{A}_n^H + \tilde{s}_n^2\boldsymbol{B}_n\boldsymbol{B}_n^H) \\ &+ \sum_{n=1}^{n_s}\sum_{p=1,p>n}^{n_s}\Big(\bar{s}_n\bar{s}_p(\boldsymbol{A}_n\boldsymbol{A}_p^H + \boldsymbol{A}_p\boldsymbol{A}_n^H) + \tilde{s}_n\tilde{s}_p(\boldsymbol{B}_n\boldsymbol{B}_p^H + \boldsymbol{B}_p\boldsymbol{B}_n^H)\Big) \\ &+ i\sum_{n=1}^{n_s}\sum_{p=1}^{n_s}\tilde{s}_n\bar{s}_p(\boldsymbol{B}_n\boldsymbol{A}_p^H - \boldsymbol{A}_p\boldsymbol{B}_n^H) \end{aligned}$$

- Proof of $\Leftarrow$), see e.g., EL&PS book.

Some properties of OSTBC

- Manifests the intuition that unitary matrices are good
- Alamouti code is an OSTBC (up to $1/\sqrt{2}$ normalization)
- Pros
  - Diversity of order $n_r n_t$
  - Detection of $\{s_n\}$ is *decoupled*
  - Converts space-time channel into n_s AWGN channels
  - Combination with outer coding is straightforward
- Cons
  - Rate loss for $n_t > 2$, i.e., $n_t > 2 \Rightarrow R = \dfrac{n_s}{N} < 1$
  - Information loss except for when $n_t = 2, n_r = 1$

Diversity order of OSTBC

- Suppose $\{s_n^0\}_{n=1}^{n_s}$ are true symbols and $\{s_n\}$ are any other symbols. Then

$$X - X_0 = \sum_{n=1}^{n_s} \Big((\bar{s}_n - \bar{s}_n^0) A_n + i(\tilde{s}_n - \tilde{s}_n^0) B_n \Big)$$

$$\Rightarrow \quad (X - X_0)(X - X_0)^H = \sum_{n=1}^{n_s} |s_n - s_n^0|^2 \cdot I$$

- Full rank ➠ full diversity for i.i.d. Gaussian channel

Derivation of decoupled detection

- Write the ML metric as

$$
\begin{aligned}
&\|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}\|^2 \\
=&\|\boldsymbol{Y}\|^2 - 2\mathrm{ReTr}\{\boldsymbol{Y}^H\boldsymbol{H}\boldsymbol{X}\} + \|\boldsymbol{H}\boldsymbol{X}\|^2 \\
=&\|\boldsymbol{Y}\|^2 - 2\sum_{n=1}^{n_s}\mathrm{ReTr}\{\boldsymbol{Y}^H\boldsymbol{H}\boldsymbol{A}_n\}\bar{s}_n + 2\sum_{n=1}^{n_s}\mathrm{Im}\,\mathrm{Tr}\{\boldsymbol{Y}^H\boldsymbol{H}\boldsymbol{B}_n\}\tilde{s}_n \\
&+\|\boldsymbol{H}\|^2\cdot\|\boldsymbol{s}\|^2 \\
=&\sum_{n=1}^{n_s}\Big(-2\mathrm{ReTr}\{\boldsymbol{Y}^H\boldsymbol{H}\boldsymbol{A}_n\}\bar{s}_n + 2\,\mathrm{Im}\,\mathrm{Tr}\{\boldsymbol{Y}^H\boldsymbol{H}\boldsymbol{B}_n\}\tilde{s}_n + |s_n|^2\|\boldsymbol{H}\|^2\Big) \\
&\quad + \text{const.} \\
=&\|\boldsymbol{H}\|^2\cdot\sum_{n=1}^{n_s}\left|s_n - \frac{\mathrm{ReTr}\{\boldsymbol{Y}^H\boldsymbol{H}\boldsymbol{A}_n\} - i\,\mathrm{Im}\,\mathrm{Tr}\{\boldsymbol{Y}^H\boldsymbol{H}\boldsymbol{B}_n\}}{\|\boldsymbol{H}\|^2}\right|^2 + \text{const.}
\end{aligned}
$$

Decoupled detection, again

- Linearity: $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E} \quad \Leftrightarrow \quad \boldsymbol{y} = \boldsymbol{F}\boldsymbol{s}' + \boldsymbol{e}, \qquad \boldsymbol{s}' \triangleq [\bar{\boldsymbol{s}}^T \ \ \tilde{\boldsymbol{s}}^T]^T$

- Theorem: $\boldsymbol{X}$ is an OSTBC if and only if

$$\text{Re}\{\boldsymbol{F}^H\boldsymbol{F}\} = \|\boldsymbol{H}\|^2 \cdot \boldsymbol{I} \qquad \forall \quad \boldsymbol{H}$$

- ML metric:

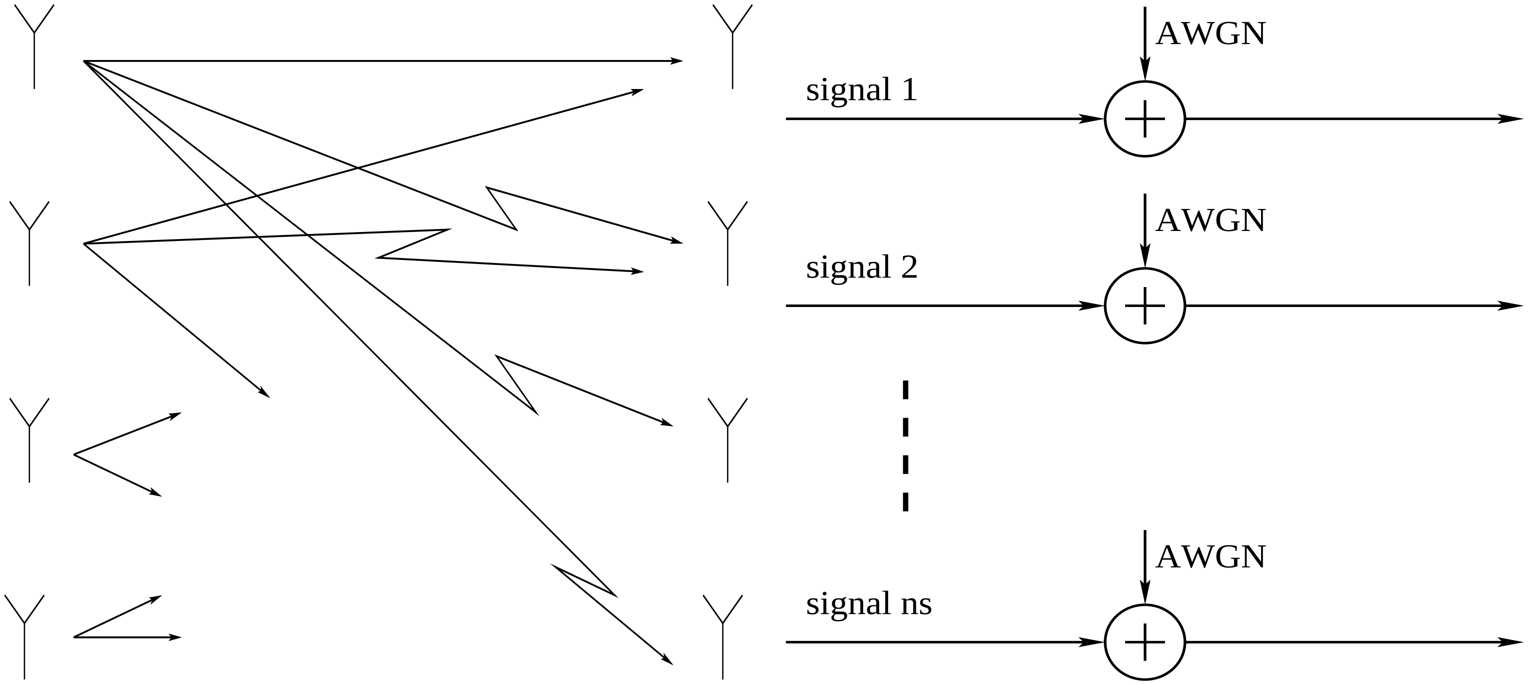
$$\begin{aligned}\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{s}'\|^2 &= \|\boldsymbol{y}\|^2 - 2\text{Re}\{\boldsymbol{y}^H\boldsymbol{F}\boldsymbol{s}'\} + \text{Re}\{\boldsymbol{s}'^T\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{s}'\} \\ &= \|\boldsymbol{y}\|^2 - 2\text{Re}\{\boldsymbol{y}^H\boldsymbol{F}\boldsymbol{s}'\} + \|\boldsymbol{H}\|^2 \cdot \|\boldsymbol{s}'\|^2 \\ &= \|\boldsymbol{H}\|^2 \cdot \|\boldsymbol{s}' - \hat{\boldsymbol{s}}'\|^2 + \text{const.}\end{aligned}$$

where $\hat{\boldsymbol{s}}' \triangleq \begin{bmatrix} \hat{\bar{\boldsymbol{s}}} \\ \hat{\tilde{\boldsymbol{s}}} \end{bmatrix} = \dfrac{\text{Re}\{\boldsymbol{F}^H\boldsymbol{y}\}}{\|\boldsymbol{H}\|^2} \sim N\left(\begin{bmatrix} \bar{s} \\ \tilde{s} \end{bmatrix}, \dfrac{N_0/2}{\|\boldsymbol{H}\|^2}\boldsymbol{I}\right)$

- $\boldsymbol{F}$ is a “Spatial/temporal (code) matched filter”

Interpretation of decoupled detection

- Space-time channel decouples into n_s AWGN channels



(a) $n_t \times n_r$ space-time channel (b) n_s independent AWGN channels

- SNR per subchannel: $\text{SNR}|_{\boldsymbol{H}} = \frac{N}{n_s} \cdot \frac{\|\boldsymbol{H}\|^2}{n_t} \cdot \frac{P}{N_0}$

Example: Alamouti's code is an OSTBC

- Consider the Alamouti code (re-normalized):

$$\boldsymbol{X} = \begin{bmatrix} s_1 & s_2^* \\ s_2 & -s_1^* \end{bmatrix}, \quad \boldsymbol{X}\boldsymbol{X}^H = (|s_1|^2 + |s_2|^2)\boldsymbol{I}$$

- Identification of $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ gives

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \boldsymbol{A}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\boldsymbol{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{B}_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Examples of OSTBC

- Best known OSTBC for $n_t = 3$, $N = 4$, $n_s = 3$:

$$\boldsymbol{X} = \begin{bmatrix} s_1 & 0 & s_2 & -s_3 \\ 0 & s_1 & s_3^* & s_2^* \\ -s_2^* & -s_3 & s_1^* & 0 \end{bmatrix}$$

Code rate: 3/4 bpcu

- For $n_t = 4$, $N = 4$, $n_s = 3$:

$$\boldsymbol{X} = \begin{bmatrix} s_1 & 0 & s_2 & -s_3 \\ 0 & s_1 & s_3^* & s_2^* \\ -s_2^* & -s_3 & s_1^* & 0 \\ s_3^* & -s_2 & 0 & s_1^* \end{bmatrix}$$

Rate is $3/4$ bpcu.

Summary of OSTBC Relations

$$XX^H = \sum_{n=1}^{n_s} |s_n|^2 \cdot I = \|s\|^2 \cdot I$$

$$\Updownarrow$$

$$\begin{aligned} A_n A_n^H &= I \quad , \quad B_n B_n^H = I \\ A_n A_p^H &= -A_p A_n^H \quad , \quad B_n B_p^H = -B_p B_n^H, \qquad n \neq p \\ A_n B_p^H &= B_p A_n^H \end{aligned}$$

$$\Updownarrow$$

$$\text{Re}\{F^H F\} = \|H\|^2 \cdot I \quad \text{where } F \text{ is such that}$$

$$\text{vec}(Y) = F \cdot \begin{bmatrix} \bar{s} \\ \tilde{s} \end{bmatrix} + \text{vec}(E)$$

Mutual Information Properties of OSTBC

- Average transmitted energy per antenna and time interval $= 1/n_t$
- Channel mutual information, with i.i.d. streams of power $1/n_t$:

$$C_{\mathsf{MIMO}}(\boldsymbol{H}) = \log \left| \boldsymbol{I} + \frac{1}{n_t} \frac{\boldsymbol{H}\boldsymbol{H}^H}{N_0} \right|$$

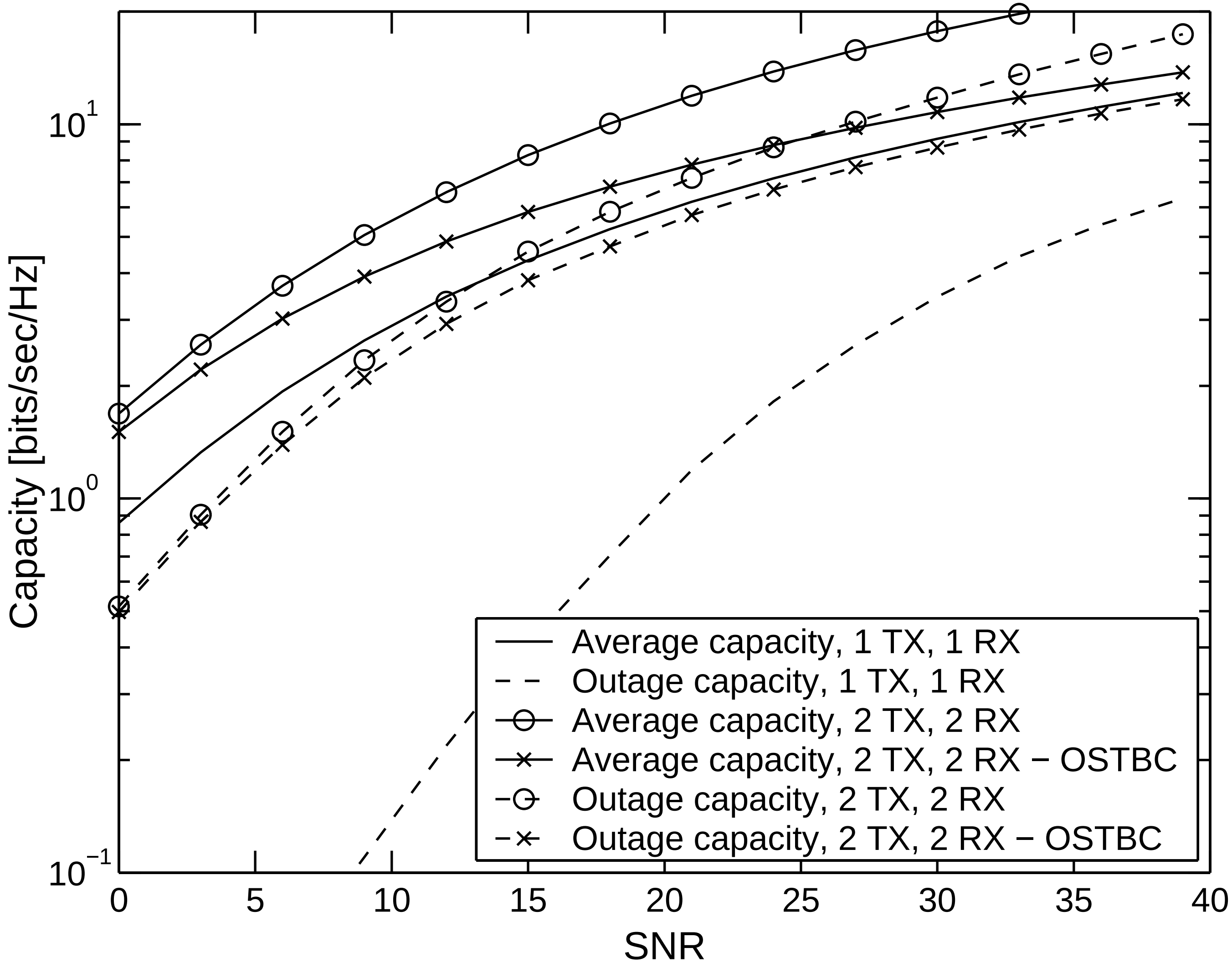
- Mutual information of OSTBC coded system:

$$C_{\mathsf{OSTBC}}(\boldsymbol{H}) = \frac{n_s}{N} \log \left(1 + \frac{N}{n_s} \frac{\|\boldsymbol{H}\|^2}{n_t N_0} \right)$$

- Theorem:

$$C_{\mathsf{MIMO}} \geq C_{\mathsf{OSTBC}}, \quad \text{equality only for } n_t = 2, n_r = 1$$

Capacity comparison (1% outage)



Non-orthogonal linear STBC

- Also called linear dispersion codes
- Different approaches:
  - Optimization of mutual information between the TX & RX:

$$\max \frac{1}{2} E_{\boldsymbol{H}} \left[\log_2 \left| \boldsymbol{I} + \frac{2}{N_0} \mathrm{Re}\{ \boldsymbol{F}^H \boldsymbol{F} \} \right| \right]$$

(no explicit guarantee for full diversity here)

  - Quasi-orthogonal codes
  - Codes based on linear constellation (complex-field) precoding

$$\boldsymbol{s}' = \boldsymbol{\Phi} \boldsymbol{s}$$

Example: A non-OSTBC

- Consider the following *diagonal code*, where $|s_n| = 1$:

$$X = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}$$

- Then

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

- ML metric for symbol detection:

$$\begin{aligned} \|Y - HX\|^2 =& \|Y\|^2 - 2\mathrm{ReTr}\{X^H H^H Y\} + \|H\|^2 \\ =& -2\mathrm{Re}\{[H^H Y]_{1,1} \cdot s_1\} - 2\mathrm{Re}\{[H^H Y]_{2,2} \cdot s_2\} + \text{const.} \end{aligned}$$

- Decoupled detection, but *not* OSTBC, and not full diversity

More examples of linear but not orthogonal STBC

- Alamouti code with forgotten conjugates

$$\boldsymbol{X} = \begin{bmatrix} s_1 & s_2 \\ s_2 & -s_1 \end{bmatrix}$$

- "Spatial multiplexing" ($R = n_t$, $N = 1$, $n_s = n_t$, $d = n_r$).
  For $n_t = 2$:

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$\boldsymbol{B}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{B}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Linearly precoded STBC

- Transmit $\boldsymbol{W}\boldsymbol{X}$ where $\boldsymbol{W} \in \{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_K\}$. Data model:

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{W}\boldsymbol{X} + \boldsymbol{E}$$

- Fact: If $\mathrm{rank}\{\boldsymbol{W}_k\} = n_t$ then same diversity order as but without $\boldsymbol{W}$
- Consider correlated fading: $\boldsymbol{R} = E[\boldsymbol{h}\boldsymbol{h}^H] = \boldsymbol{R}_t^T \otimes \boldsymbol{R}_r, \quad \boldsymbol{h} = \mathrm{vec}(\boldsymbol{H})$
- Error probability:

$$E_{\boldsymbol{H}}[P(\boldsymbol{X}_0 \to \boldsymbol{X})] \leq \text{const.} \cdot \left| \boldsymbol{I} + \frac{1}{N_0}(\boldsymbol{X}_0 - \boldsymbol{X})(\boldsymbol{X}_0 - \boldsymbol{X})^H \cdot \boldsymbol{W}^H \boldsymbol{R}_t \boldsymbol{W} \right|^{-n_r}$$

- For OSTBC, $(\boldsymbol{X}_0 - \boldsymbol{X})(\boldsymbol{X}_0 - \boldsymbol{X})^H \propto \boldsymbol{I}$. Hence, $\min_{\boldsymbol{W}} \left| \boldsymbol{I} + \frac{n_s}{N_0} \boldsymbol{W}^H \boldsymbol{R}_t \boldsymbol{W} \right|$

Ex. OSTBC with One-Bit Feedback for $n_t = 2$

- One bit used to choose between

$$\underbrace{\boldsymbol{W}_1 = \begin{bmatrix} |a| & 0 \\ 0 & \sqrt{1-|a|^2} \end{bmatrix}}_{\text{if } \|\boldsymbol{h}_1\| > \|\boldsymbol{h}_2\|}, \quad \underbrace{\boldsymbol{W}_2 = \begin{bmatrix} \sqrt{1-|a|^2} & 0 \\ 0 & |a| \end{bmatrix}}_{\text{if } \|\boldsymbol{h}_2\| > \|\boldsymbol{h}_1\|}$$

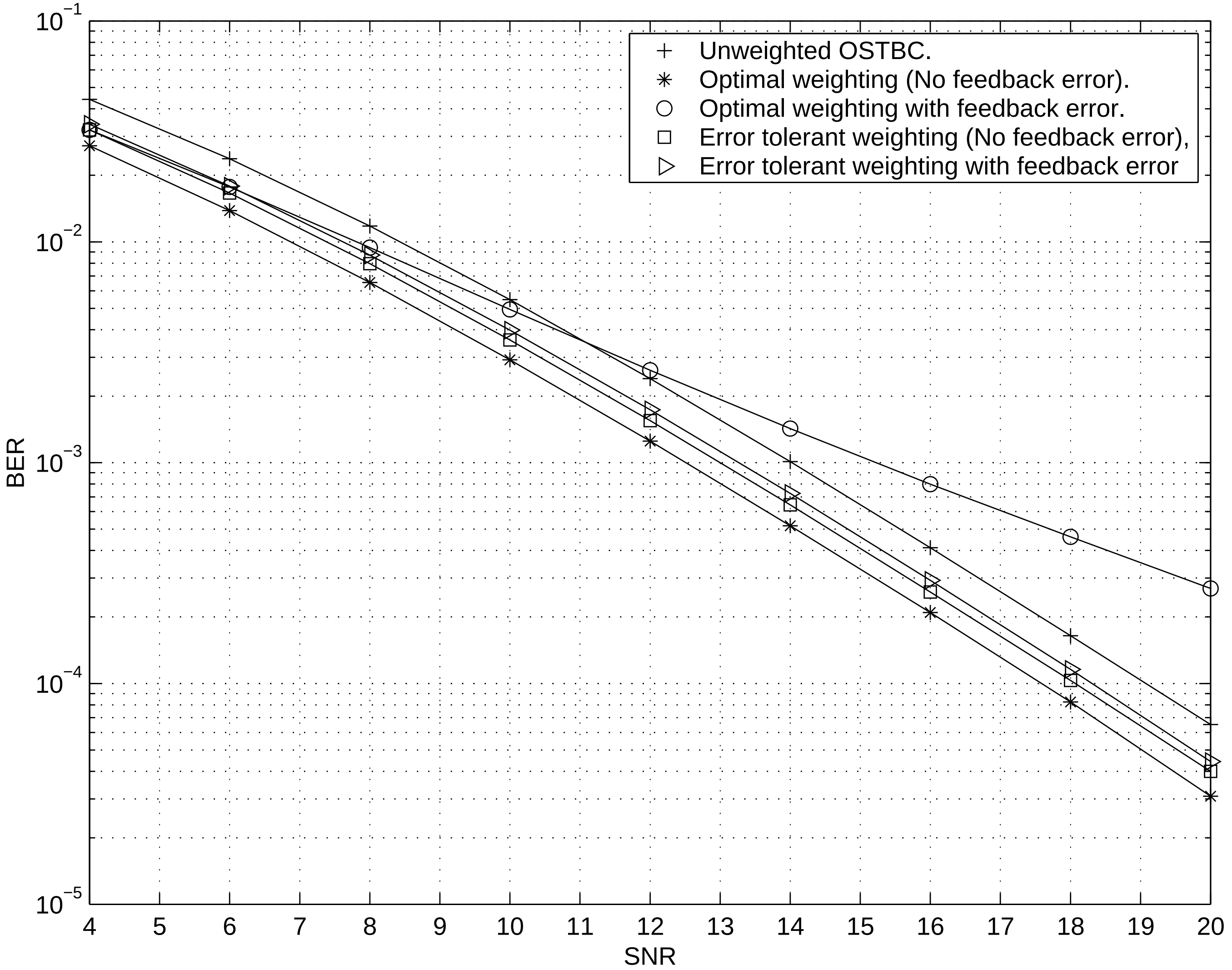
- Let P_c be the probability that the feedback bit is correct

- For $P_c = 1$ (reliable feedback), $a = 1$ is optimal ➠ antenna selection

  $\boldsymbol{W}$ may be multiplied with fixed unitary matrix ➠ grid of beams

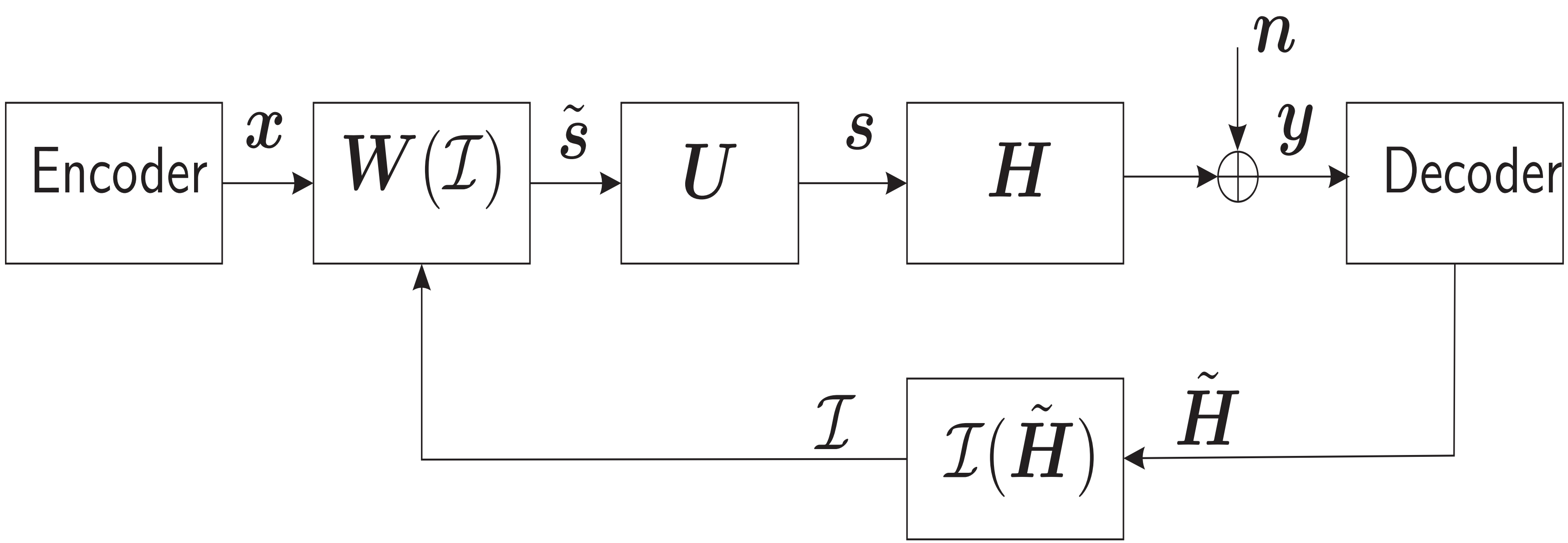
- For $P_c < 1$ (erroneous feedback),

$$E_{\boldsymbol{H}}[P(\boldsymbol{X}_0 \to \boldsymbol{X})] \leq \frac{2}{\mathsf{SNR}^2} \cdot \left(\frac{P_c}{|a|^2} + \frac{1-P_c}{1-|a|^2} \right) \quad \text{(for } n_r = 1\text{)}$$



Unweighted OSTBC.
Optimal weighting (No feedback error).
Optimal weighting with feedback error.
Error tolerant weighting (No feedback error),
Error tolerant weighting with feedback error
BER
SNR
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

MIMO with feedback - optimized transmission



- Here
  - U depends on long-term feedback
  - $\mathcal{I}$ depends on short-term (few bits) feedback
- State-of-the art designs rely on vector quantization techniques

Frequency-selective channels

- Maximum diversity order (with ML detection) will be $n_r n_t L$ where L=length of CIR
- Variety of techniques to achieve maximum diversity
- Most widely used transmission technique is MIMO-OFDM
  - coding across multiple OFDM symbols
  - coding across subcarriers within one OFDM symbol
- Basic model per subcarrier is

$$\boldsymbol{Y}_n = \boldsymbol{H}_n \boldsymbol{X}_n + \boldsymbol{E}_n$$

Le 3: MIMO receivers

Summary of MIMO receivers

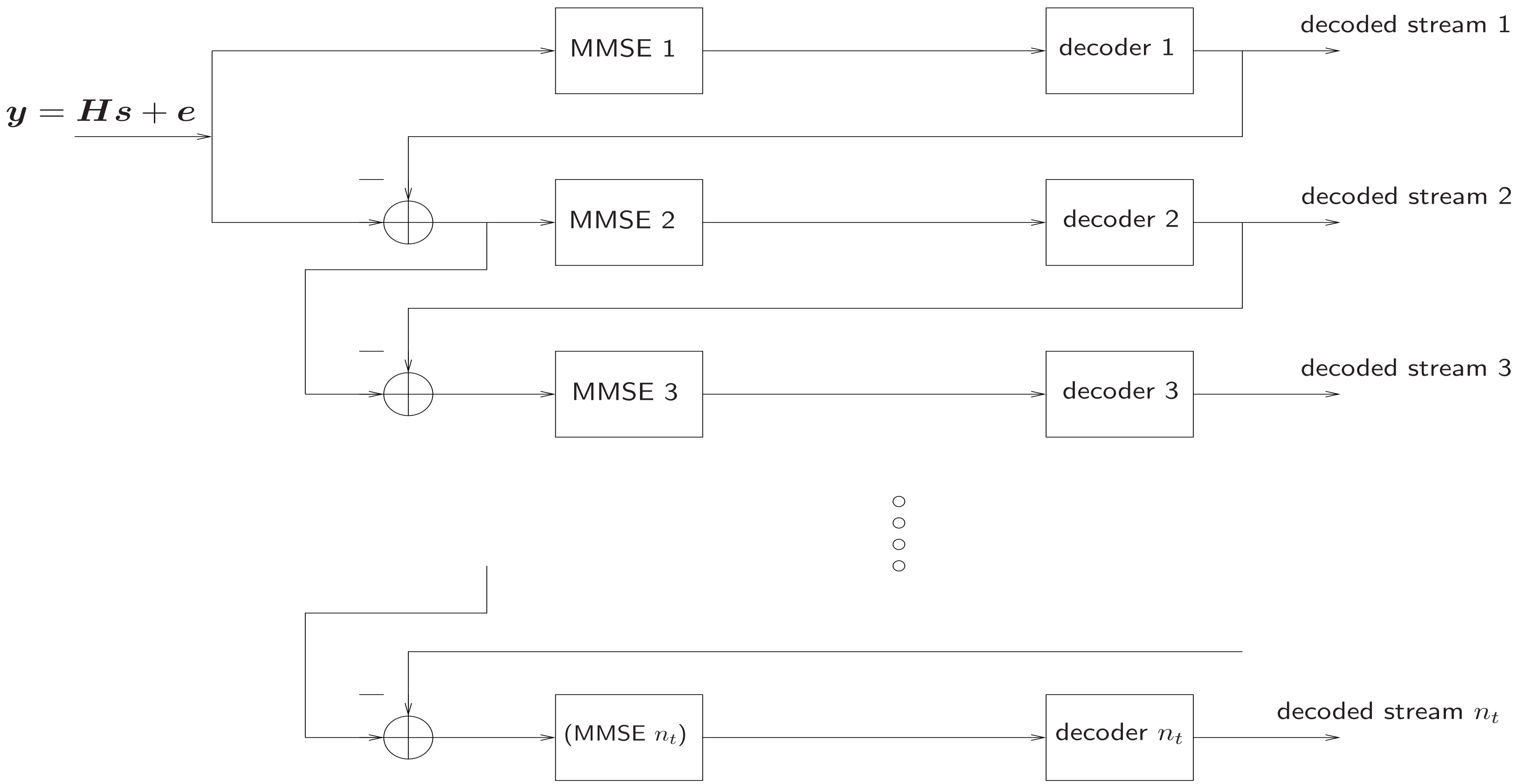
- Optimal architectures (from Le 1):
  - CSI@TX (any fading): linear processing, $\tilde{\boldsymbol{y}} = \boldsymbol{U}^H \boldsymbol{y}$, separates streams
  - no CSI@TX, fast fading: V-BLAST, optimal receiver is more involved
    - linear receiver (channel inversion) is grossly suboptimal
    - successive interference cancellation (SIC)
    - using soft MIMO demodulator + decoder, possibly iterative
  - no CSI@TX, slow fading: D-BLAST

- Architectures with STBC+outer FEC (from Le 2)
  - With OSTBC, decoupled detection and thing are simple:

$$\min \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\| \sim \min \left(|s_1 - \hat{s}_1|^2 + |s_2 - \hat{s}_2|^2\right)$$

  - With non-OSTBC, $\min \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|$ does not decouple
    - problem similar to for V-BLAST

Theoretically optimal V-BLAST receiver based on SIC



- Optimality only for fast fading. Requires rate allocation on streams.
- Major drawback: Requires long codewords. Prone to error propagation.

Theoretically optimal D-BLAST receiver based on SIC

	$\boldsymbol{x}_B(1)$	$\boldsymbol{x}_B(2)$		
$\boldsymbol{x}_A(1)$	$\boldsymbol{x}_A(2)$	$\boldsymbol{x}_A(3)$		

- One codeword split as $\boldsymbol{x}(i) = [\boldsymbol{x}_A(i) \;\; \boldsymbol{x}_B(i)]$, with rate allocation
- Decoding in steps:
  1. Decode $\boldsymbol{x}_A(1)$
  2. Decode $\boldsymbol{x}_B(1)$, suppressing $\boldsymbol{x}_A(2)$ via MMSE
  3. Strip off $\boldsymbol{x}_B(1)$, and decode $\boldsymbol{x}_A(2)$
  4. Decode $\boldsymbol{x}_B(2)$, suppressing $\boldsymbol{x}_A(3)$ via MMSE
- Drawbacks:
  - error propagation
  - rate loss due to initialization
  - requires long codewords

Receivers for linear STBC architectures

	Training	Data 1	Data 2			

- Received block: $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E}$.

- $\boldsymbol{X}$ linear in $\{s_1, ..., s_{n_s}\}$ so with appropriate $\boldsymbol{F}$, the ML metric is

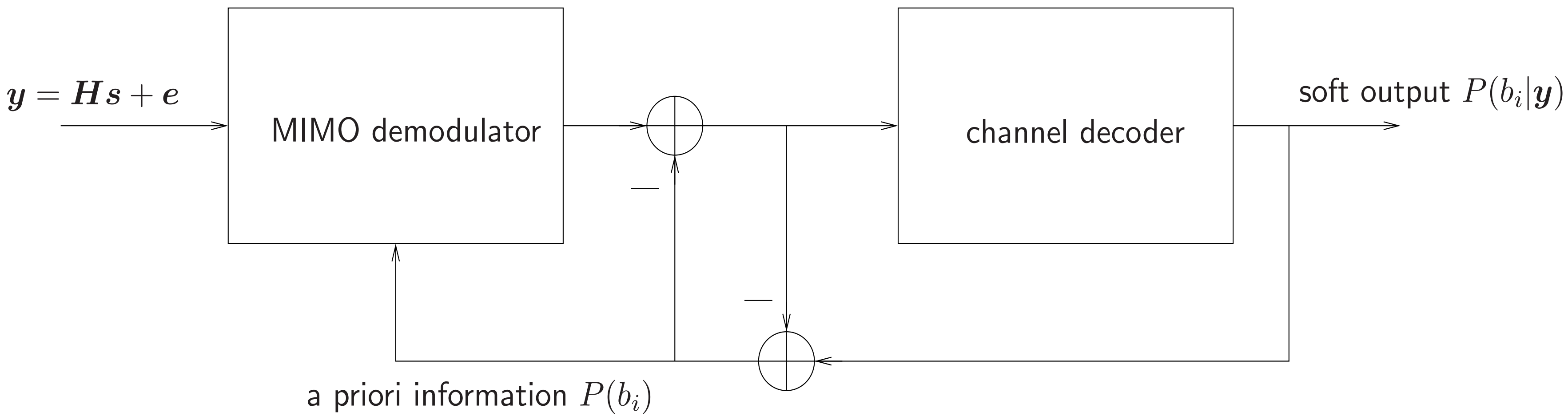
$$\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|^2 = \left\| \begin{bmatrix} \mathrm{vec}(\bar{\boldsymbol{Y}}) \\ \mathrm{vec}(\tilde{\boldsymbol{Y}}) \end{bmatrix} - \boldsymbol{F} \begin{bmatrix} \bar{\boldsymbol{s}} \\ \tilde{\boldsymbol{s}} \end{bmatrix} \right\|^2$$

- For OSTBC, $\boldsymbol{F}^T\boldsymbol{F} = \|\boldsymbol{H}\|^2\boldsymbol{I}$ so detection decouples

- Spatial multiplexing (V-BLAST) can be seen a degenerated special case of this architecture, with

$$\boldsymbol{X} = \begin{bmatrix} s_1 \\ \vdots \\ s_{n_t} \end{bmatrix}$$

Demodulator+decoder architectures



- Demodulator computes $P(b_i|\boldsymbol{y})$ given a priori information $P(b_i)$
- Decoder adds knowledge of what codewords are valid
- Added knowledge in decoder is fed back to demodulator as a priori
- Iteration until convergence (a few iterations, normally)

MIMO demodulation (hard)

- General transmission model, with $G_{i,j} \in \mathbb{R}$

$$\underbrace{\boldsymbol{y}}_{m\times 1} = \underbrace{\boldsymbol{G}}_{m\times n} \cdot \underbrace{\boldsymbol{s}}_{n\times 1} + \underbrace{\boldsymbol{e}}_{m\times 1}, \quad s_k \in \mathcal{S}$$

- Models V-BLAST architectures, and (non-O)STBC architectures
- Other applications: multiuser detection, ISI, crosstalk in cables, ...
- Typically, $m \geq n$ and $\boldsymbol{G}$ is full rank and has no structure.

The problem

- If $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma \boldsymbol{I})$ then the problem is to detect $\boldsymbol{s}$ from $\boldsymbol{y}$

$$\min_{\boldsymbol{s} \in \mathcal{S}^n} \|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s}\|^2, \qquad \boldsymbol{y} \in \mathbb{R}^m, \quad \boldsymbol{G} \in \mathbb{R}^{m \times n}$$

- Let $\boldsymbol{G} = \boldsymbol{Q}\boldsymbol{L}$ where $\begin{cases} \boldsymbol{Q} \in \mathbb{R}^{m \times n} & \text{is orthonormal } (\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}) \\ \boldsymbol{L} \in \mathbb{R}^{n \times n} & \text{is lower triangular} \end{cases}$

Then $\|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s}\|^2 = \left\|\boldsymbol{Q}\boldsymbol{Q}^T(\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s})\right\|^2 + \left\|(\boldsymbol{I} - \boldsymbol{Q}\boldsymbol{Q}^T)(\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s})\right\|^2$

$$= \left\|\boldsymbol{Q}^T\boldsymbol{y} - \boldsymbol{L}\boldsymbol{s}\right\|^2 + \left\|(\boldsymbol{I} - \boldsymbol{Q}\boldsymbol{Q}^T)\boldsymbol{y}\right\|^2$$

so $\min_{\boldsymbol{s} \in \mathcal{S}^n} \|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s}\|^2 \quad \Leftrightarrow \quad \min_{\boldsymbol{s} \in \mathcal{S}^n} \|\tilde{\boldsymbol{y}} - \boldsymbol{L}\boldsymbol{s}\|$ where $\tilde{\boldsymbol{y}} \triangleq \boldsymbol{Q}^T\boldsymbol{y}$

Some remarks

- Integer-constrained least-squares problem, known to be NP hard
- Brute force complexity $O(|\mathcal{S}^n|)$
- Typical dimension of problem: $n \sim 8-16$, so $|\mathcal{S}| \sim 2$–8, $|\mathcal{S}^n| \sim 256$–10^{14}
- Needs be solved
  - *in real time*
  - *once per received vector* $\boldsymbol{y}$
  - *in power-efficient hardware* (beware of heavy matrix algebra)
  - possibly *fixed-point arithmetics*
  - preferably, in a parallel architecture
- In communications, we can accept a suboptimal algorithm that finds the correct solution quickly, with high probability

Some remarks, cont.

- For $\boldsymbol{G} \propto$ orthogonal (OSTBC), the problem is trivial.
- Our focus is on unstructured $\boldsymbol{G}$
- If $\boldsymbol{G}$ has structure (e.g., Toeplitz) then use algorithm that exploits this
- Generally, slow fading (no time diversity) is the hard case

Zero-Forcing

- Let

$$\tilde{\boldsymbol{s}} \triangleq \arg \min_{\boldsymbol{s} \in \mathbb{R}^n} \|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s}\| = \arg \min_{\boldsymbol{s} \in \mathbb{R}^n} \|\tilde{\boldsymbol{y}} - \boldsymbol{L}\boldsymbol{s}\| = \boldsymbol{L}^{-1}\tilde{\boldsymbol{y}}$$

E.g., Gaussian elimination: $\tilde{s}_1 = \tilde{y}_1 / L_{1,1}$

$$\tilde{s}_2 = (\tilde{y}_2 - \tilde{s}_1 L_{2,1}) / L_{2,2}$$

$$\vdots$$

- Then project onto $\mathcal{S}$: $\hat{s}_k = [\tilde{s}_k] \triangleq \arg \min_{s_k \in \mathcal{S}} |s_k - \tilde{s}_k|$

- This works very poorly. Why? Note that

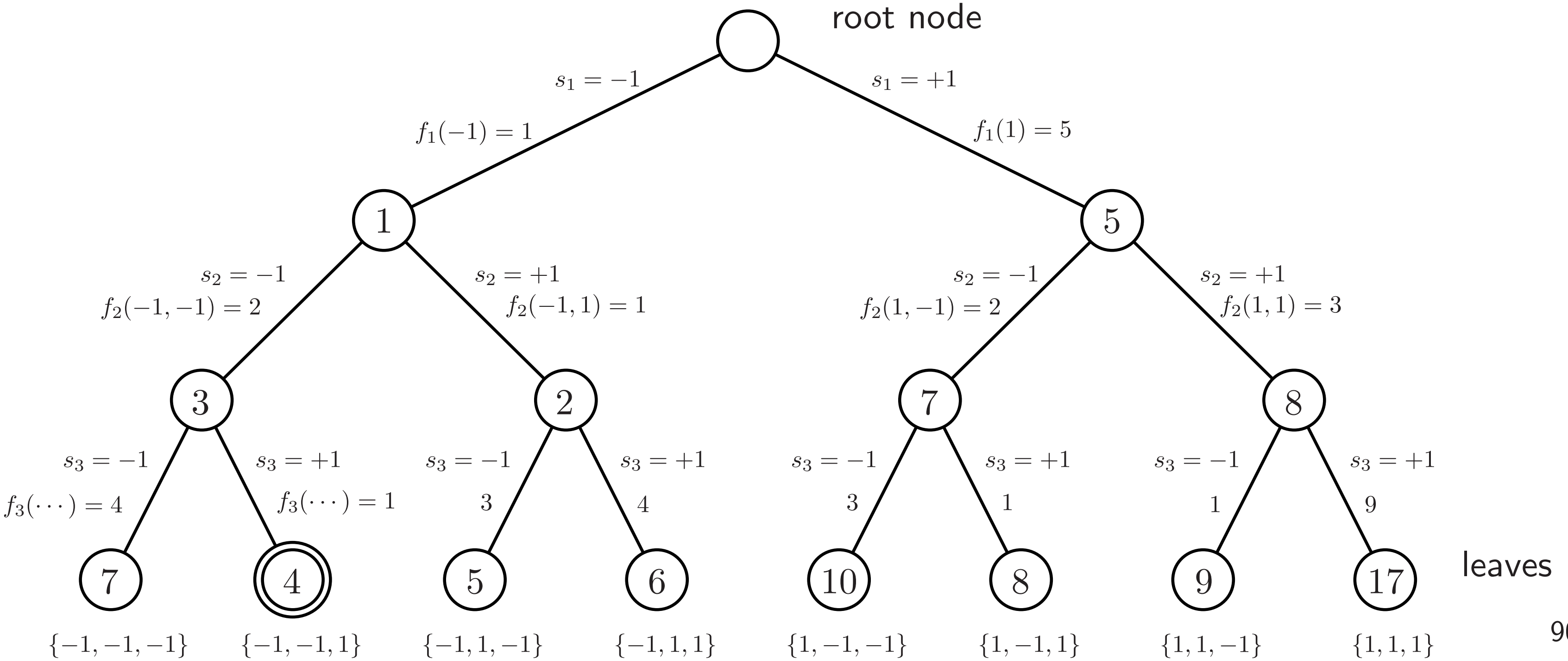
$$\tilde{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{L}^{-1}\boldsymbol{Q}^T\boldsymbol{e} = \boldsymbol{s} + \tilde{\boldsymbol{e}}, \qquad \text{where } \mathrm{cov}(\tilde{\boldsymbol{e}}) = \sigma \cdot (\boldsymbol{L}^T\boldsymbol{L})^{-1}$$

ZF neglects the correlation between the elements of $\tilde{\boldsymbol{e}}$

Decision tree view

$$\min_{\substack{\{s_1,\ldots,s_n\} \\ s_k\in\mathcal{S}}} \{f_1(s_1)+f_2(s_1,s_2)+\cdots+f_n(s_1,\ldots,s_n)\}$$

where $$f_k(s_1,...,s_k) \triangleq \left(\tilde{y}_k - \sum_{l=1}^{k} L_{k,l}s_l\right)^2$$



Zero-Forcing with Decision Feedback (ZF-DF)

- Consider the following improvement

  i) Detect s_1 via: $\hat{s}_1 = \left[\frac{\tilde{y}_1}{L_{1,1}}\right] = \arg\min_{s_1 \in \mathcal{S}} f_1(s_1)$

  ii) Consider s_1 known and set $\hat{s}_2 = \left[\frac{\tilde{y}_2 - \hat{s}_1 L_{2,1}}{L_{2,2}}\right] = \arg\min_{s_2 \in \mathcal{S}} f_2(\hat{s}_1, s_2)$

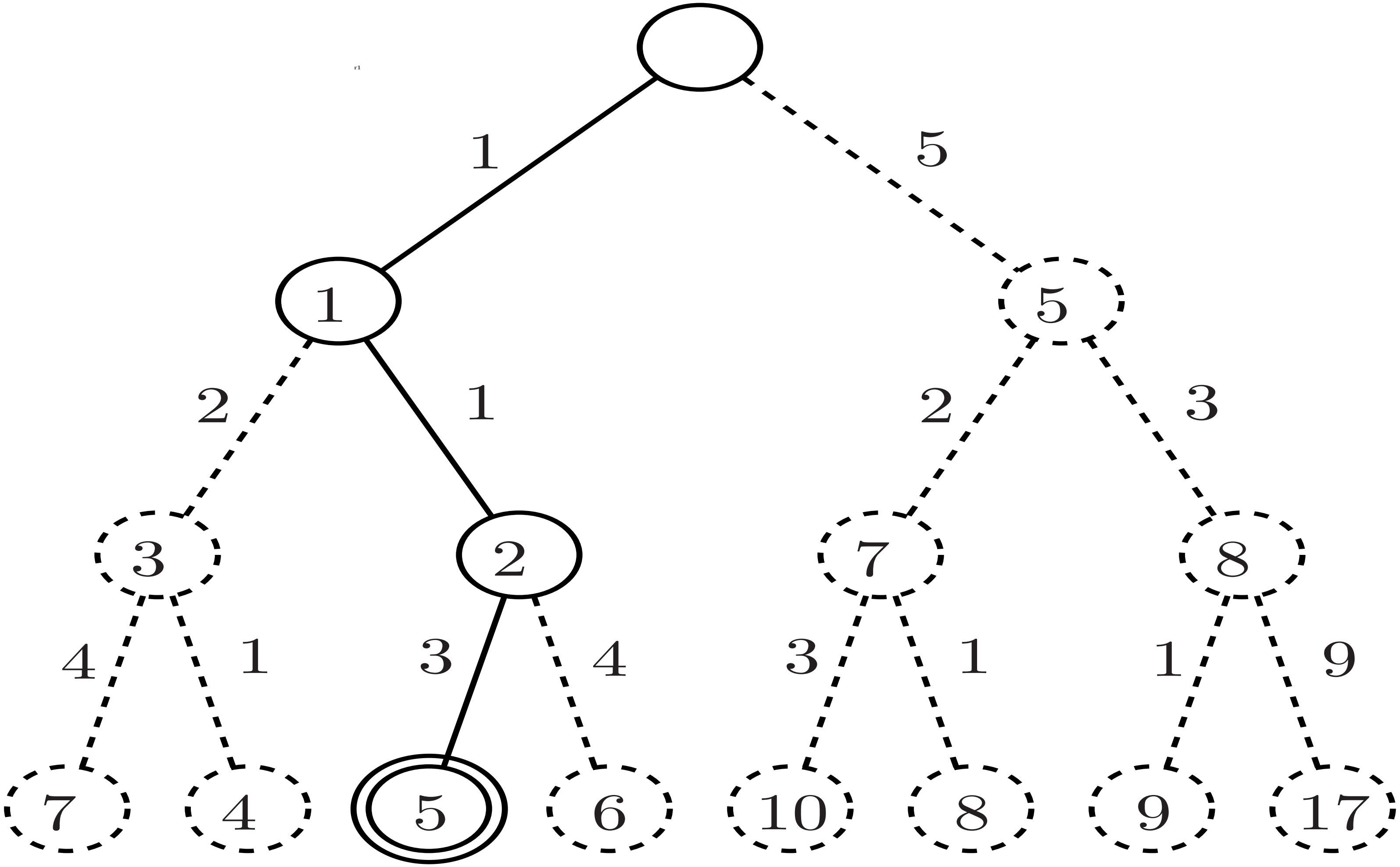
  iii) Continue for $k = 3, ..., n$:

$$\hat{s}_k = \left[\frac{\tilde{y}_k - \sum_{l=1}^{k-1} L_{k,l}\hat{s}_l}{L_{k,k}}\right] = \arg\min_{s_k \in \mathcal{S}} f_k(\hat{s}_1, ..., \hat{s}_{k-1}, s_k)$$

- This also works poorly. Why? Error propagation.
  Incorrect decision on s_i ➠ most of the following s_k wrong as well.

- Optimized detection order (start with the best) does not help much.

Zero-Forcing with Decision Feedback (ZF-DF)

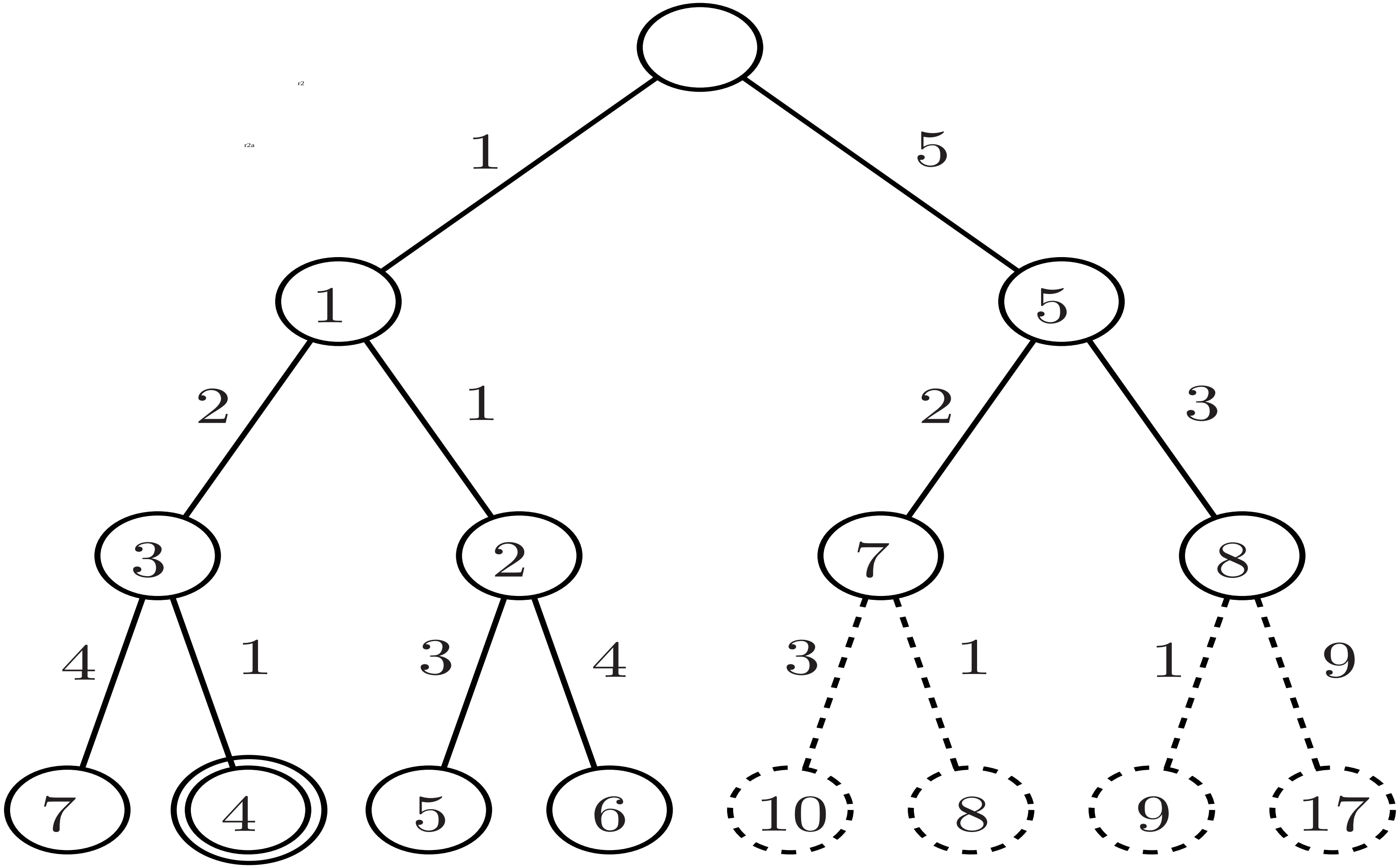


Sphere decoding (SD)

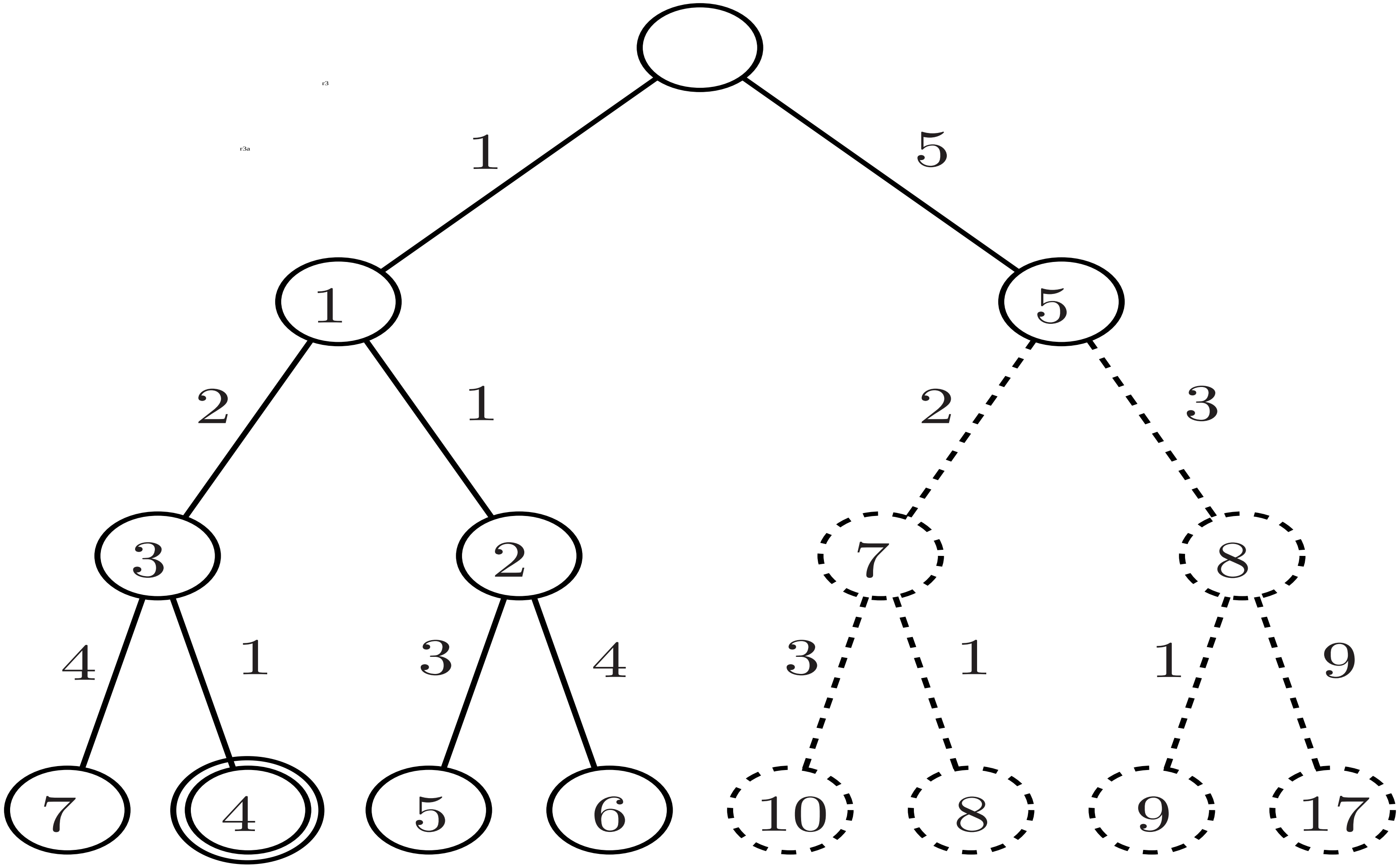
- Select a sphere radius, R. Then traverse the tree, but once encountering a node with cumulative metric $> R$, do not follow it down
- Enumerates all leaf nodes which lie inside the sphere $\|\tilde{\boldsymbol{y}} - \boldsymbol{L}\boldsymbol{s}\|^2 \leq R$
- Improvements:
  - Pruning: At each leaf, update R according to $R := \min(R, M)$
  - Improvements: optimal ordering of s_k
  - Branch enumeration

  (e.g., $s_k = \{-5, -3, -1, -1, 3, 5\}$ vs. $s_k = \{-1, 1, -3, 3, -5, 5\}$)
- Known facts:
  - The algorithm solves the problem, if allowed to finish
  - Runtime is random and algorithm cannot be parallelized
  - Under relevant circumstances, average runtime is $O(2^{\alpha n})$ for $\alpha > 0$

SD, without pruning, $R = 6$



SD, with pruning, $R = \infty$



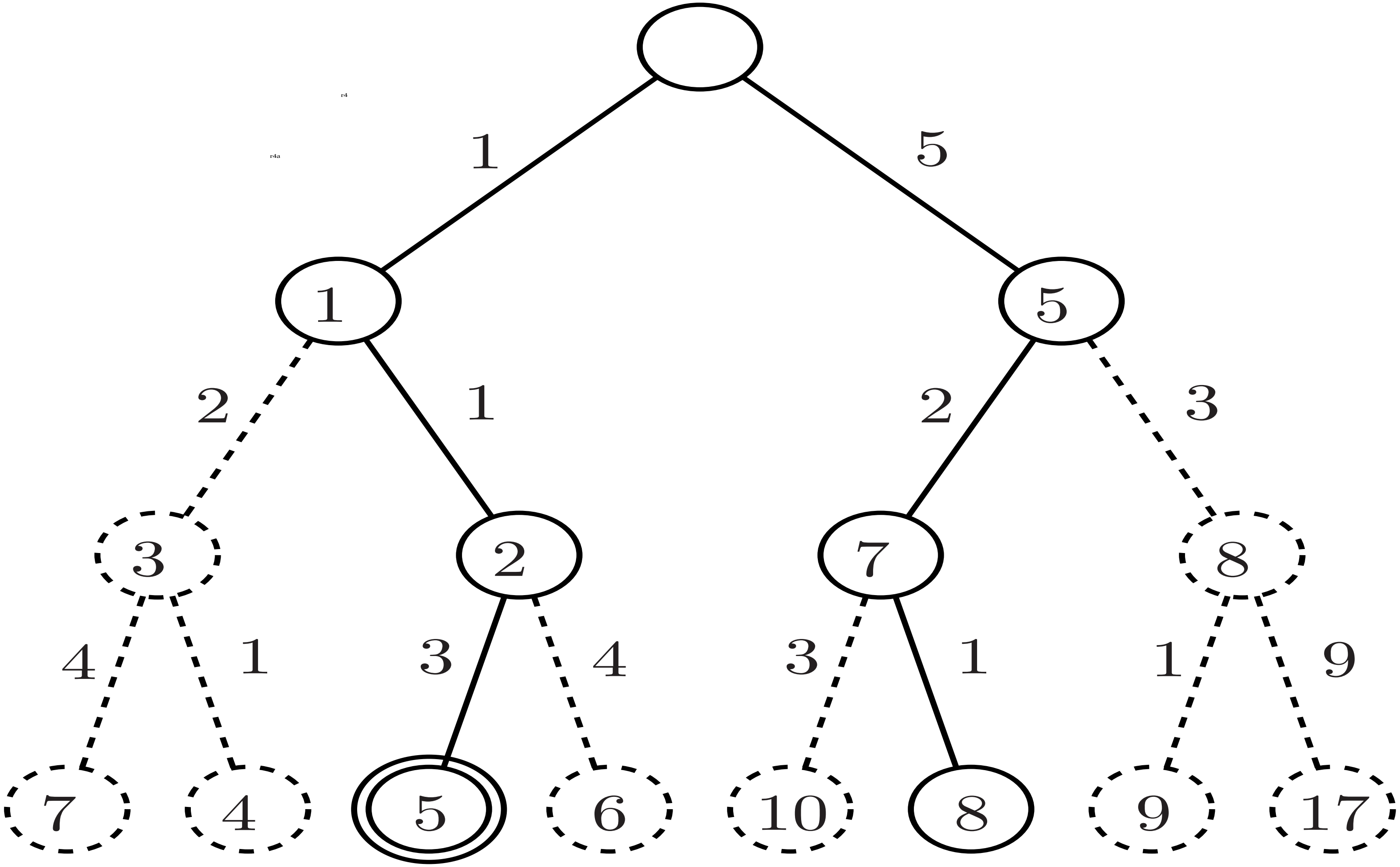
“Fixed complexity” sphere decoding (FCSD)

- Select a user parameter r, $0 \leq r \leq n$
- For each node on layer r, consider $\{s_1, ..., s_r\}$ fixed and solve

$$(*) \qquad \min_{\substack{\{s_{r+1},...,s_n\} \\ s_k \in \mathcal{S}}} \{f_{r+1}(s_1, ..., s_{r+1}) + \cdots + f_n(s_1, ..., s_n)\}$$

- Subproblem (*) solved using $|\mathcal{S}|^r$ times
- Low-complexity approximation (e.g. ZF-DF) can be used. Why? (*) is overdetermined (equivalent $\boldsymbol{G}$ is tall)
- Order can be optimized: start with the “worst”
- Fixed runtime, fully parallel structure

FCSD, $r = 1$



Semidefinite relaxation (for $s_k \in \{\pm 1\}$)

- Let $\bar{\boldsymbol{s}} \triangleq \begin{bmatrix} \boldsymbol{s} \\ 1 \end{bmatrix}$, $\quad \boldsymbol{S} \triangleq \bar{\boldsymbol{s}}\bar{\boldsymbol{s}}^T = \begin{bmatrix} \boldsymbol{s} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{s}^T & 1 \end{bmatrix}$, $\quad \boldsymbol{\Psi} \triangleq \begin{bmatrix} \boldsymbol{L}^T\boldsymbol{L} & -\boldsymbol{L}^T\tilde{\boldsymbol{y}} \\ -\tilde{\boldsymbol{y}}^T\boldsymbol{L} & 0 \end{bmatrix}$

  Then
  $$\|\tilde{\boldsymbol{y}} - \boldsymbol{L}\boldsymbol{s}\|^2 = \bar{\boldsymbol{s}}^T\boldsymbol{\Psi}\bar{\boldsymbol{s}} + \|\tilde{\boldsymbol{y}}\|^2 = \text{Trace}\{\boldsymbol{\Psi}\boldsymbol{S}\} + \|\tilde{\boldsymbol{y}}\|^2$$
  so the problem is to
  $$\min_{\substack{\text{diag}\{\boldsymbol{S}\}=\{1,\ldots,1\} \\ \text{rank}\{\boldsymbol{S}\}=1 \\ \bar{s}_{n+1}=1}} \text{Trace}\{\boldsymbol{\Psi}\boldsymbol{S}\}$$

- SDR proceeds by relaxing $\text{rank}\{\boldsymbol{S}\} = 1$ to $\boldsymbol{S}$ positive semidefinite

- Interior point methods used to find $\boldsymbol{S}$

- $\boldsymbol{s}$ recovered, e.g., by taking dominant eigenvector and project onto $\mathcal{S}^n$

Lattice reduction

- Extend $\mathcal{S}^n$ to lattice. For example, if $\mathcal{S} = \{-3, -1, 1, 3\}$, then $\bar{\mathcal{S}}^n = \{\ldots, -3, -1, 1, 3, \ldots\} \times \cdots \times \{\ldots, -3, -1, 1, 3, \ldots\}$.
- Decide on orthogonal *integer* matrix $\boldsymbol{T} \in \mathbb{R}^{n \times n}$ that maps $\bar{\mathcal{S}}^n$ onto itself:

$$T_{k,l} \in \mathbb{Z}, \quad |\boldsymbol{T}| = 1, \qquad \text{and} \qquad \boldsymbol{T}\boldsymbol{s} \in \bar{\mathcal{S}}^n \quad \forall \boldsymbol{s} \in \bar{\mathcal{S}}^n$$

- Find one such $\boldsymbol{T}$ for which $\boldsymbol{L}\boldsymbol{T} \propto \boldsymbol{I}$
- Then solve $\hat{\boldsymbol{s}}' \triangleq \arg \min_{\boldsymbol{s}' \in \bar{\mathcal{S}}^n} \|\tilde{\boldsymbol{y}} - (\boldsymbol{L}\boldsymbol{T})\boldsymbol{s}'\|^2$, and set $\hat{\boldsymbol{s}} = \boldsymbol{T}^{-1}\hat{\boldsymbol{s}}'$
- Critical steps:
  - Find suitable $\boldsymbol{T}$ (computationally costly, but amortize over many $\boldsymbol{y}$)
  - $\hat{\boldsymbol{s}} \in \bar{\mathcal{S}}^n$, but $\hat{\boldsymbol{s}} \notin \mathcal{S}^n$ in general, so clipping is necessary

MIMO demodulation (soft)

- Data model

$$\underbrace{\boldsymbol{y}}_{m\times 1} = \underbrace{\boldsymbol{G}}_{m\times n} \cdot \underbrace{\boldsymbol{s}}_{n\times 1} + \underbrace{\boldsymbol{e}}_{m\times 1}, \quad s_k = \mathcal{S}(b_1, ..., b_p) \in \mathcal{S}, \quad |\mathcal{S}| = 2^p$$

- Bits b_i a priori indep. with

$$L(b_i) = \log\left(\frac{P(b_i = 1)}{P(b_i = 0)}\right), \qquad i = 1, ..., np$$

- Objective: Determine

$$L(b_i|\boldsymbol{y}) = \log\left(\frac{P(b_i = 1|\boldsymbol{y})}{P(b_i = 0|\boldsymbol{y})}\right)$$

Posterior bit probabilities

$$L(b_i|\boldsymbol{y}) = \log\left(\frac{P(b_i=1|\boldsymbol{y})}{P(b_i=0|\boldsymbol{y})}\right) \overset{\text{(a)}}{=} \log\left(\frac{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=1} P(\boldsymbol{s}|\boldsymbol{y})}{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=0} P(\boldsymbol{s}|\boldsymbol{y})}\right) \overset{\text{(b)}}{=} \log\left(\frac{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=1} p(\boldsymbol{y}|\boldsymbol{s})P(\boldsymbol{s})}{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=0} p(\boldsymbol{y}|\boldsymbol{s})P(\boldsymbol{s})}\right)$$

$$\overset{\text{(c)}}{=} \log\left(\frac{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=1} p(\boldsymbol{y}|\boldsymbol{s})\left(\prod_{i'=1}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)}{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=0} p(\boldsymbol{y}|\boldsymbol{s})\left(\prod_{i'=1}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)}\right)$$

$$= \log\left(\frac{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=1} p(\boldsymbol{y}|\boldsymbol{s})\left(\prod_{i'=1,i'\neq i}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)\cdot P(b_i=1)}{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=0} p(\boldsymbol{y}|\boldsymbol{s})\left(\prod_{i'=1,i'\neq i}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)\cdot P(b_i=0)}\right)$$

$$= \log\left(\frac{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=1} p(\boldsymbol{y}|\boldsymbol{s})\left(\prod_{i'=1,i'\neq i}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)}{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=0} p(\boldsymbol{y}|\boldsymbol{s})\left(\prod_{i'=1,i'\neq i}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)}\right) + L(b_i)$$

In Gaussian noise $p(\boldsymbol{y}|\boldsymbol{s}) = \frac{1}{(2\pi\sigma)^{m/2}}\exp\left(-\frac{1}{2\sigma}\|\boldsymbol{y}-\boldsymbol{G}\boldsymbol{s}\|^2\right)$ so

$$L(b_i|\boldsymbol{y}) = \log\left(\frac{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=1} \exp\left(-\frac{1}{2\sigma}\|\boldsymbol{y}-\boldsymbol{G}\boldsymbol{s}\|^2\right)\left(\prod_{i'=1,i'\neq i}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)}{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=0} \exp\left(-\frac{1}{2\sigma}\|\boldsymbol{y}-\boldsymbol{G}\boldsymbol{s}\|^2\right)\left(\prod_{i'=1,i'\neq i}^{np} P(b_{i'}=b_{i'}(\boldsymbol{s}))\right)}\right) + L(b_i)$$

- With a priori equiprobable bits

$$L(b_i|\boldsymbol{y}) = \log\left(\frac{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=1} \exp\left(-\frac{1}{2\sigma}\|\boldsymbol{y}-\boldsymbol{G}\boldsymbol{s}\|^2\right)}{\sum_{\boldsymbol{s}:b_i(\boldsymbol{s})=0} \exp\left(-\frac{1}{2\sigma}\|\boldsymbol{y}-\boldsymbol{G}\boldsymbol{s}\|^2\right)}\right)$$

- $\sum$ can be relatively well approximated by its largest term

  That gives problems of the type

$$\min_{\boldsymbol{s}\in\mathcal{S}^n, b_i(\boldsymbol{s})=\beta} \|\boldsymbol{y}-\boldsymbol{G}\boldsymbol{s}\|^2$$

  This is called "max-log" approximation

- If many candidates $\boldsymbol{s}$ are examined, then use these terms in $\sum$
  - List-decoding algorithm

Incorporating a priori probabilities (BPSK/dim case)

- Consider $s_k \in \{\pm 1\}$, and let

$$s_k = 2b_k - 1$$

$$\gamma_k \triangleq \frac{1}{2}\log\left(P(s_k=-1)P(s_k=1)\right) = \frac{1}{2}\log\left(P(b_k=0)P(b_k=1)\right)$$

$$\lambda_k \triangleq \log\left(\frac{P(s_k=1)}{P(s_k=-1)}\right) = \log\left(\frac{P(b_k=1)}{P(b_k=0)}\right) = L(b_k)$$

- The prior is linear in s_k:

$$\log(P(s_k=s)) = \frac{1}{2}[(1+s)\log(P(s_k=1)) + (1-s)\log(P(s_k=-1))]$$

$$= \frac{1}{2}\gamma_k + \frac{1}{2}\lambda_k s_k$$

- Write

$$L(s_k|\boldsymbol{y}) = \log \left(\frac{\sum_{\boldsymbol{s}:s_k=1} \exp\left(-\frac{1}{\sigma}\|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s}\|^2 + \frac{1}{2}\sum_{i=1,i\neq k}^{n}(\gamma_i + \lambda_i s_i)\right)}{\sum_{\boldsymbol{s}:s_k=0} \exp\left(-\frac{1}{\sigma}\|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{s}\|^2 + \frac{1}{2}\sum_{i=1,i\neq k}^{n}(\gamma_i + \lambda_i s_i)\right)} \right) + \lambda_k$$

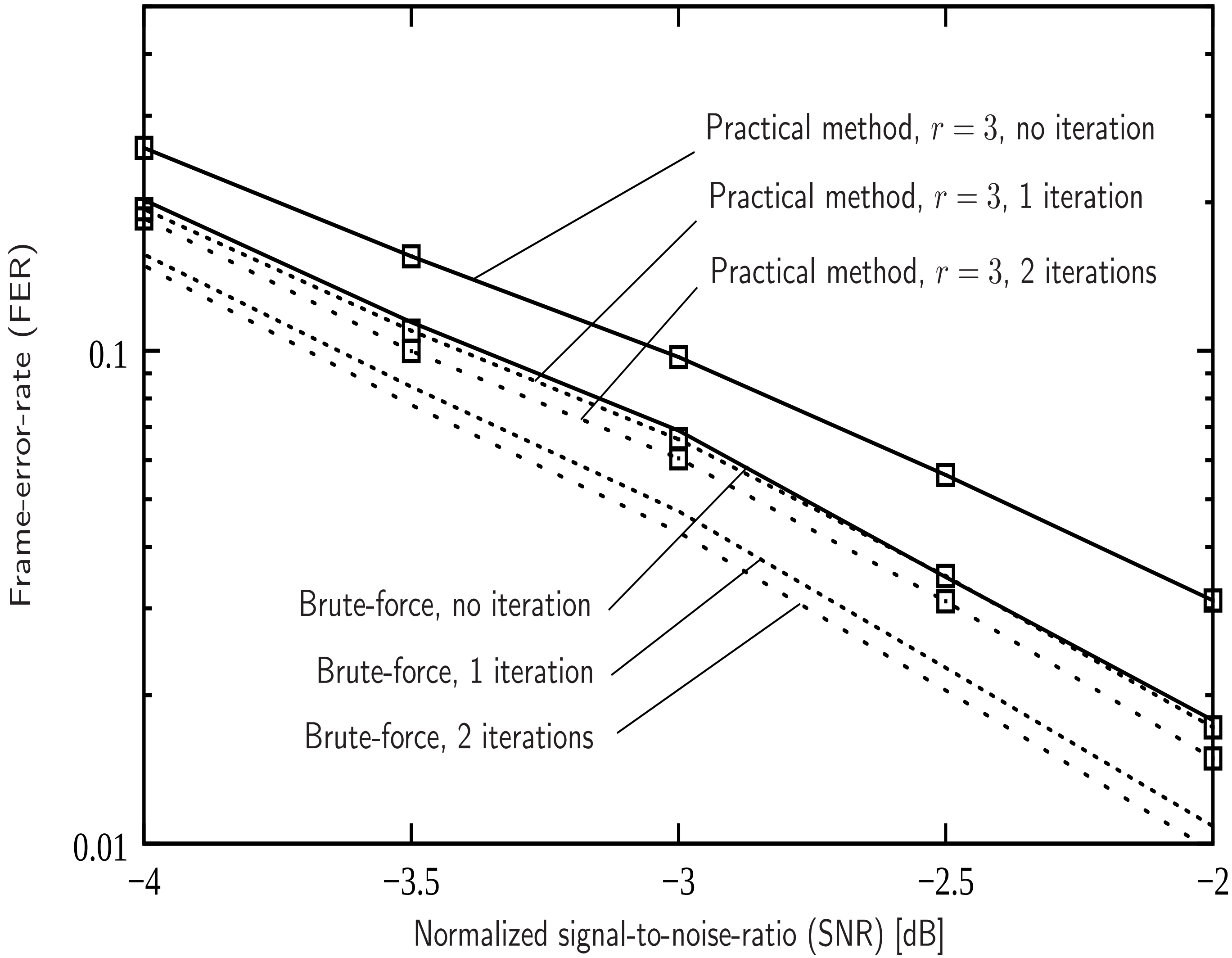
- Define $\tilde{\boldsymbol{y}} \triangleq [\boldsymbol{y}^T \ 1 \ \cdots \ 1]^T$ and $\tilde{\boldsymbol{G}} \triangleq \begin{bmatrix} \boldsymbol{G} \\ \boldsymbol{\Lambda}_k \end{bmatrix}$ where $\boldsymbol{\Lambda}_k \triangleq$ diag $\left\{\frac{\sigma}{4}\lambda_1, \cdots, \frac{\sigma}{4}\lambda_{k-1}, \frac{\sigma}{4}\lambda_{k+1}, \ldots, \frac{\sigma}{4}\lambda_n\right\}$

- Then

$$L(s_k|\boldsymbol{y}) = \log \left(\frac{\sum_{\boldsymbol{s}:s_k=1} \exp\left(-\frac{1}{\sigma}\|\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{G}}\boldsymbol{s}\|^2 + \sum_{i=1,i\neq k}^{n}\left(\frac{\sigma\lambda_i^2}{16} + \frac{\gamma_i}{2}\right)\right)}{\sum_{\boldsymbol{s}:s_k=0} \exp\left(-\frac{1}{\sigma}\|\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{G}}\boldsymbol{s}\|^2 + \sum_{i=1,i\neq k}^{n}\left(\frac{\sigma\lambda_i^2}{16} + \frac{\gamma_i}{2}\right)\right)} \right) + \lambda_k$$

- A priori information on s_k ➠ "virtual antennas"

Example w. iter. decod. 4×4, $r = 1/2$-LDPC, 1000 bits



Channel ($\boldsymbol{H}$) estimation and associated receivers

- Very often pilots are used to form a channel estimate.

  Consider

$$\boldsymbol{Y}_t = \boldsymbol{H}\boldsymbol{X}_t + \boldsymbol{E}_t$$

- Estimate $\boldsymbol{H}$ via training:
  - Maximum likelihood (in Gaussian noise):

$$\hat{\boldsymbol{H}} = \underset{\boldsymbol{H}}{\operatorname{argmin}} \|\boldsymbol{Y}_t - \boldsymbol{H}\boldsymbol{X}_t\|^2 = \boldsymbol{Y}_t\boldsymbol{X}_t^H(\boldsymbol{X}_t\boldsymbol{X}_t^H)^{-1}$$

  - Can show, estimate is the same also in colored noise (e.g. co-channel interference in multiuser system)
  - Can be somewhat improved by using MMSE estimation

- Estimate noise (co)variance:

$$\hat{\mathbf{\Lambda}} = \frac{1}{N_t}\boldsymbol{Y}_t \Pi^{\perp}_{\boldsymbol{X}_t^H} \boldsymbol{Y}_t^H, \quad \text{for colored noise}$$
$$\hat{N}_0 = \frac{1}{N_t n_r}\text{Tr}\left\{\boldsymbol{Y}_t \Pi^{\perp}_{\boldsymbol{X}_t^H} \boldsymbol{Y}_t^H\right\}, \quad \text{for white noise}$$

- This estimate can be used to prewhiten the received signal to suppress co-channel interference. E.g.

$$\tilde{\boldsymbol{Y}} = \mathbf{\Lambda}^{-1/2}\boldsymbol{Y}$$

- At most $n_r - 1$ rank-1 interferers can be suppressed.

The receive array has n_r degrees of freedom.

At high SNR, $\mathbf{\Lambda}^{-1}$ becomes a projection matrix that projects out interference.

More on training

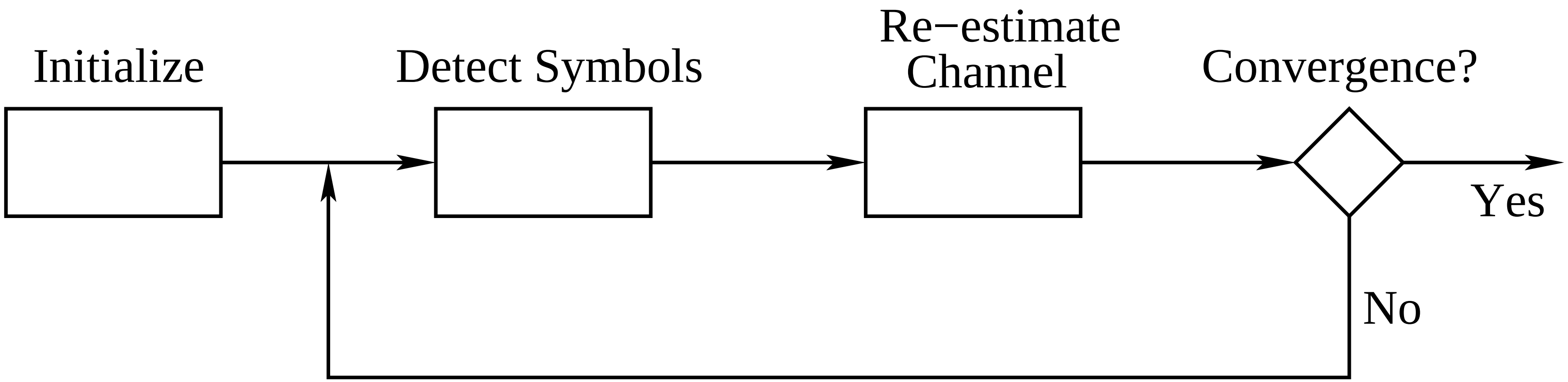
- Training-based detector uses $\hat{\boldsymbol{H}}$, $\hat{N}_0$ and $\hat{\boldsymbol{\Lambda}}$ in the coherent detector
- Can be improved, e.g., cyclic detection
- How should training be designed?

  Optimum pilots (in many respects) satisfy $\boldsymbol{X}_t \boldsymbol{X}_t^H \propto \boldsymbol{I}$
- Inserting pilot-based estimate in LF is *not* optimal

  Cf. the use of
  $p(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{H})$ (coherent)
  $p(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{H})|_{\boldsymbol{H}:=\hat{\boldsymbol{H}}}$ (training-based)
  $p(\boldsymbol{X}|\boldsymbol{Y}, \hat{\boldsymbol{H}})$ (best possible given $\hat{\boldsymbol{H}}$)
  $p(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{Y}_t)$ (best possible given training data)

  Later, we will give $p(\boldsymbol{X}|\boldsymbol{Y}, \hat{\boldsymbol{H}})$ on explicit form

Example, joint detection and estimation schemes



- Re-estimation step may make use of soft decoder output

Optimal training

- Recall: ML channel estimate $\hat{\boldsymbol{H}} = \boldsymbol{Y}_t \boldsymbol{X}_t^H (\boldsymbol{X}_t \boldsymbol{X}_t^H)^{-1}$
- Let $\boldsymbol{h} = \text{vec}(\boldsymbol{H})$, $\hat{\boldsymbol{h}} = \text{vec}(\hat{\boldsymbol{H}})$. Can show:

$$E[\hat{\boldsymbol{h}}] = \boldsymbol{h}$$

$$\boldsymbol{\Sigma} \triangleq E[(\hat{\boldsymbol{h}} - \boldsymbol{h})(\hat{\boldsymbol{h}} - \boldsymbol{h})^H] = \ldots = N_0\Big((\boldsymbol{X}_t \boldsymbol{X}_t^H)^{-T} \otimes \boldsymbol{I}\Big)$$

$$\text{Tr}\{\boldsymbol{\Sigma}\} = n_r \text{Tr}\big\{(\boldsymbol{X}_t \boldsymbol{X}_t^H)^{-1}\big\} N_0$$

- Lemma: Suppose $\text{Tr}\big\{\boldsymbol{X}\boldsymbol{X}^H\big\} \leq n_t$. Then

$$\text{Tr}\big\{(\boldsymbol{X}\boldsymbol{X}^H)^{-1}\big\} \geq n_t \quad \text{with equality if and only if } \boldsymbol{X}\boldsymbol{X}^H = \boldsymbol{I}$$

- Application of lemma ➠ optimal training block is (semi-)unitary:

$$\boldsymbol{X}_t \boldsymbol{X}_t^H \propto \boldsymbol{I}$$

Channel estimation for frequency-selective channels

- MIMO channel as matrix-valued FIR filter:

$$\boldsymbol{H}(z^{-1}) = \sum_{l=0}^{L} \boldsymbol{H}_l z^{-l}$$

- L is the length of the channel, $L = 0$ for ISI-free channel
- Transfer function:

$$\boldsymbol{H}(\omega) = \sum_{l=0}^{L} \boldsymbol{H}_l e^{-i\omega l}$$

- Transmission methods:
  - Single-carrier (block-based)
  - Multicarrier (e.g. OFDM)

Training for frequency selective channels

- Two basic approaches:
  - Frequency-domain estimation (estimate $\boldsymbol{H}(\omega)$)
  - Time-domain estimation (estimate $\boldsymbol{H}_l$, then compute $\boldsymbol{H}(\omega)$ via FT)
- Frequency-domain estimation of $\boldsymbol{H}(\omega)$ is straightforward:
  - Estimate in the frequency domain (via ML):

$$\hat{\boldsymbol{H}}(\omega) = \boldsymbol{Y}_t(\omega)\boldsymbol{X}_t^H(\omega)(\boldsymbol{X}_t(\omega)\boldsymbol{X}_t^H(\omega))^{-1}$$

  - Problem: suboptimal because parameterizations is not parsimonious. It does not exploit structure that

$$\boldsymbol{H}(\omega) = \sum_{l=0}^{L} \boldsymbol{H}_l e^{-i\omega l}$$

Time-domain estimation of $\boldsymbol{H}(\omega)$

- Let $\boldsymbol{x}(n)$ be time-domain training, $\boldsymbol{y}(n)$ received (t-d) training and define

$$\boldsymbol{X}_t = \begin{bmatrix} \boldsymbol{x}_t^T(0) & & \cdots & \cdots & \cdots & \boldsymbol{x}_t^T(-L) \\ \boldsymbol{x}_t^T(1) & \ddots & & & & \boldsymbol{x}_t^T(1-L) \\ \vdots & & \ddots & & & \vdots \\ \boldsymbol{x}_t^T(N-1) & \cdots & \cdots & \boldsymbol{x}_t^T(0) & \cdots & \boldsymbol{x}_t^T(N-1-L) \end{bmatrix}$$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_0^T \\ \vdots \\ \boldsymbol{H}_L^T \end{bmatrix}, \quad \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{y}_t^T(0) \\ \vdots \\ \boldsymbol{y}_t^T(N-1) \end{bmatrix}$$

- Then the ML estimate of $\boldsymbol{H}(\omega)$ is:

$$\hat{\boldsymbol{H}} = (\boldsymbol{X}_t^H \boldsymbol{X}_t)^{-1} \boldsymbol{X}_t^H \boldsymbol{Y}_t, \quad \hat{\boldsymbol{H}}(\omega) = \sum_{l=0}^{L} \hat{\boldsymbol{H}}_l e^{-i\omega l}$$

- Exploits structure (L unknowns but N equations)

Metrics with imperfect CSI (complex $\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{G}, \boldsymbol{e}$)

- $\boldsymbol{G}$ not known perfectly ➠ replacing $\boldsymbol{G}$ with $\hat{\boldsymbol{G}}$ in $p(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{G})$ **not** optimal!
- Instead, need to work with $p(\boldsymbol{y}|\boldsymbol{s}, \hat{\boldsymbol{G}})$. Write

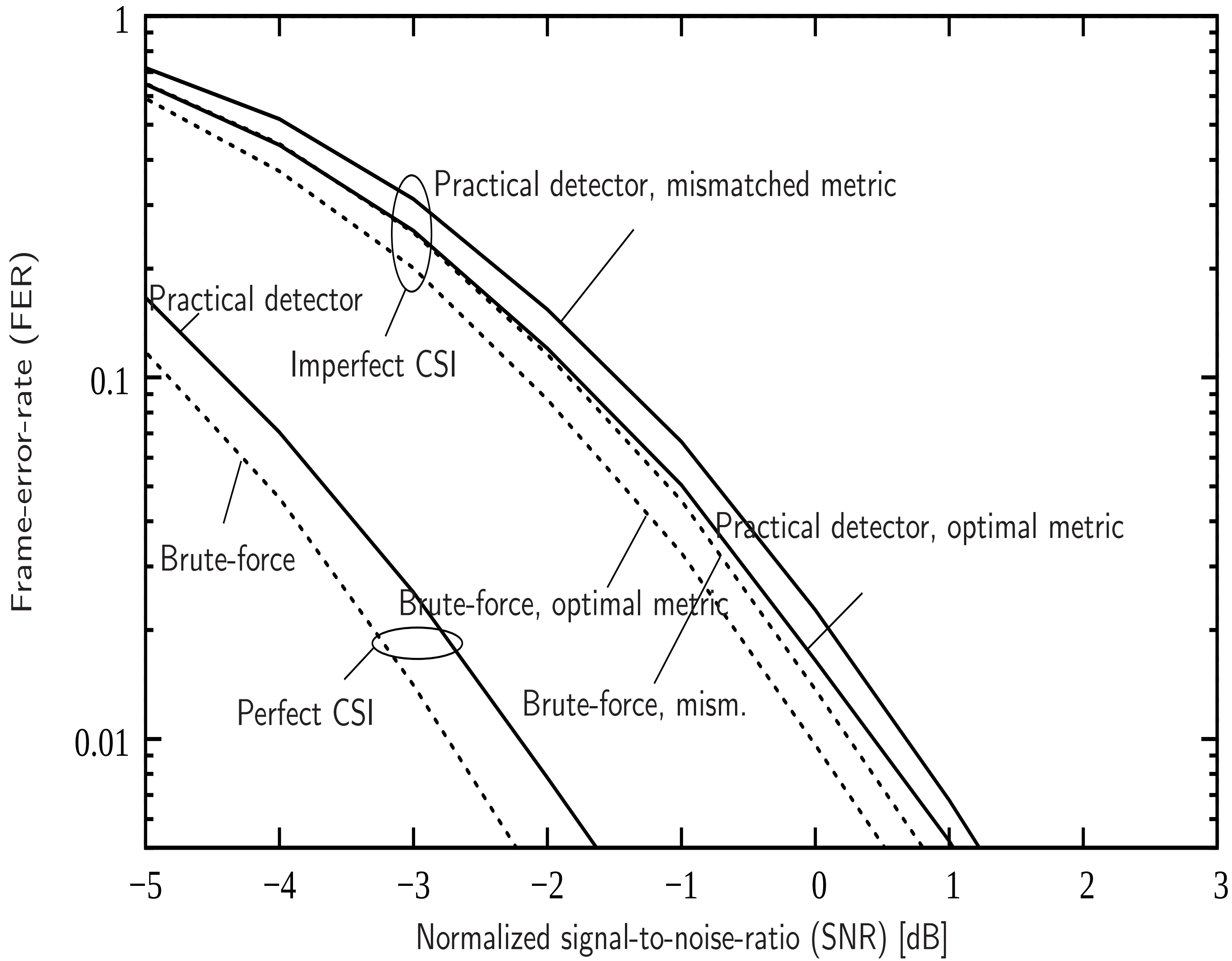
$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{s} + \boldsymbol{e} \quad \Leftrightarrow \quad \boldsymbol{y} = (\boldsymbol{s}^T \otimes \boldsymbol{I})\boldsymbol{h} + \boldsymbol{e}, \quad \boldsymbol{h} = \text{vec}(\boldsymbol{G}), \ \boldsymbol{e} = \text{vec}(\boldsymbol{E})$$

Suppose $\|\boldsymbol{s}\|^2 = n$ and $\begin{cases} \boldsymbol{h} \sim N(\boldsymbol{0}, \rho \boldsymbol{I}), & \boldsymbol{e} \sim N(\boldsymbol{0}, \sigma \boldsymbol{I}) \\ \hat{\boldsymbol{h}} = \boldsymbol{h} + \boldsymbol{\delta}, & \boldsymbol{\delta} \sim N(\boldsymbol{0}, \epsilon \boldsymbol{I}) \end{cases}$

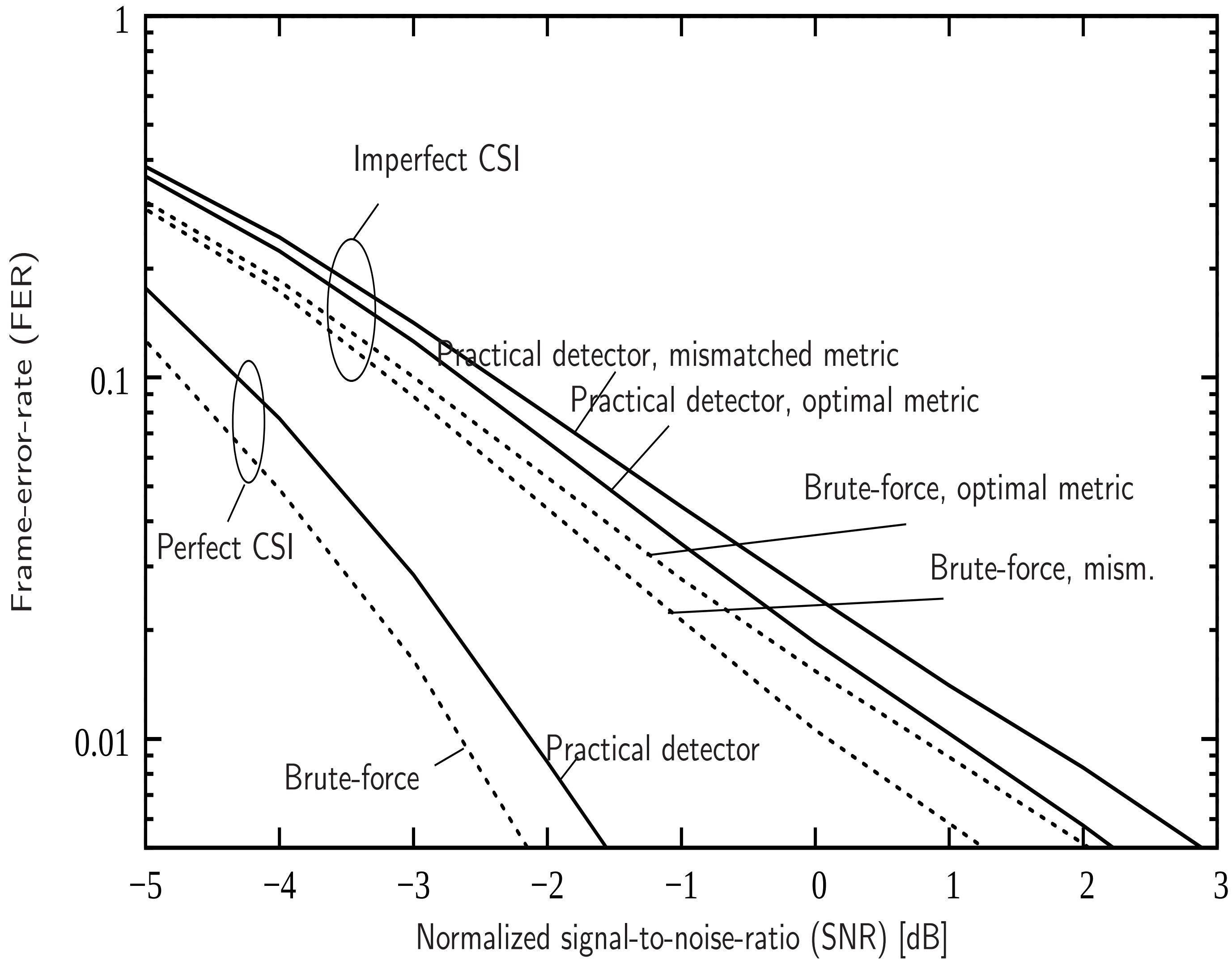
- Then $\begin{bmatrix} \boldsymbol{y} \\ \hat{\boldsymbol{h}} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} (n\rho + \sigma)\boldsymbol{I} & \rho(\boldsymbol{s}^T \otimes \boldsymbol{I}) \\ \rho(\boldsymbol{s}^* \otimes \boldsymbol{I}) & (\rho + \epsilon)\boldsymbol{I} \end{bmatrix}\right)$ so

$$p(\boldsymbol{y}|\hat{\boldsymbol{h}}, \boldsymbol{s}) = \frac{1}{\pi^n} \frac{1}{\frac{n\epsilon}{1+\epsilon/\rho} + \sigma} \exp\left(-\frac{1}{\frac{n\epsilon}{1+\epsilon/\rho} + \sigma} \left\| \boldsymbol{y} - \left(\frac{\rho}{\rho + \epsilon}\right) \hat{\boldsymbol{G}}\boldsymbol{s} \right\|^2 \right)$$

Example: 4×4 slow Rayl. fading MIMO, QPSK, est. G



Example: 4×4 slow Rayl. fad., QPSK, outdat. G



Last slide