Advanced Communication Signal Processing (MIMO)

December 16, 2010

Contents

Chapter 1

Preliminary

1.1 Wireless channel review

(Ref. Ch.15 Digital Communication by B.Skalar)

- For multipath channels, the baseband equivalent channel coefficient consist of multipath components as

$$h = \sum_n (x_n + \jmath y_n) = x + \jmath y$$

- Due to the central limit theorem, both x & y have a Gaussian distribution. Also they are independent.

- Let $x_n = r_n \cos\theta_n$ & $y_n = r_n \sin\theta_n$

$$\begin{aligned} \mathrm{E}(xy) &= \mathrm{E}(\sum_n r_n \cos\theta_n \sum_m r_m \cos\theta_n) \\ &= \sum_n \sum_m \mathrm{E}(r_n r_m \cos\theta_n \sin\theta_m) \\ &= \sum_n \sum_m \mathrm{E}(r_n r_m)\mathrm{E}(\cos\theta_n \sin\theta_m) \end{aligned}$$

Since $\mathrm{E}(r_n r_m) = \sigma^2 \delta_{nm}$, we have

$$\sum_n \sigma^2 \mathrm{E}(\cos\theta_n \sin\theta_n) = \sum_n \sigma^2 \mathrm{E}(\frac{1}{2}\sin 2\theta_n) = 0$$

$\therefore$ x & y are uncorrelated $\rightarrow$ independent.

- The channel coefficient h becomes a "complex Gaussian" random variable and the baseband equivalent received signal in fading channels are often modeled as

$$y = hx + n,$$

where $h \sim \mathcal{N}(0,1)$, $n \sim \mathcal{N}(0,\sigma^2)$, $\mathrm{E}(x^2) = \mathrm{E}(y^2) = \frac{1}{2}$

- Defining the magnitude an the phase of h as $R = |h| = \sqrt{x^2+y^2}$ and $\Theta = \angle h = \tan^{-1}\frac{y}{x}$, it can be shown that R has a Rayleigh distribution, and Θ has a uniform distribution, respectively

$$\begin{aligned} f_R(r) &= \frac{r}{\sigma^2}\exp(-\frac{r^2}{2\theta^2}) \qquad \text{for} \quad r \geq 0 \\ f_\Theta(\theta) &= \frac{1}{2\pi} \qquad \text{for} \quad \theta \in [0, 2\pi] \end{aligned}$$

- Joint pdf of x & y is given by

$$f_{X,Y} = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$$

- Also R & Θ are independent, and the joint pdf becomes

$$f_{R,\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2}\exp(-\frac{r^2}{2\sigma^2})$$

- flat fading : $T_m < T_s$
  frequency selective fading fading : $T_m > T_s$
  T_m : delay spread, T_s : symbol period

1.2 Information theory

- Suppose we have a discrete random variable (RV) X with a distribution according to $p(x)$. Then, entropy is defined as

$$\begin{aligned} H(x) &\triangleq \mathrm{E}(\log_2 \frac{1}{p(x)}) \quad \text{(bit)} \\ &= \sum_n p_n \log_2 \frac{1}{p_n} \end{aligned}$$

Similarly, for a continuous RV X, we have

$$H(x) = \int p(x) \log \frac{1}{p(x)}$$

The entropy is a measure of the uncertainty
ex) dice $H(x) = \sum_1^6 \frac{1}{6} \log_2 6$ bits
coin $H(x) = \sum_1^2 \frac{1}{2} \log_2 2 = 1$ bits

Theorem 1 *$p(x)$ which yields the maximum entropy subject to the condition that the standard deviation of X is equal to σ is Gaussian*

Proof 1

$$\max_{p(x)} \int_{-\infty}^{\infty} p(x) \log_2 \frac{1}{p(x)} dx$$
$$s.t \quad \int_{-\infty}^{\infty} x^2 p(x) dx = \sigma^2 \quad \& \quad \int_{-\infty}^{\infty} p(x) dx = 1$$

Using Lagrange multipliers, the cost function becomes

$$C \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx + \lambda_1 \int_{-\infty}^{\infty} x^2 p(x) dx + \lambda_2 \int_{-\infty}^{\infty} p(x) dx$$

By differentiating the terms inside the integral with respect to $p(x)$, we have

$$-\log_2 p(x) - \frac{1}{\log_e 2} + \lambda_1 x^2 + \lambda_2 = 0$$
$$\Rightarrow \ln p(x) - \lambda_1' x^2 - \lambda_2' = 0 \qquad p(x) = e^{\lambda_1 x^2 + \lambda_2}$$

By plugging $p(x)$ into two constraints. it follows

$$\int_{-\infty}^{\infty} x^2 e^{\lambda_1 x^2+\lambda_2} dx = \sigma^2 \quad \& \quad \int_{-\infty}^{\infty} e^{\lambda_1 x^2+\lambda_2} dx = 1$$

After some manipulation, we obtain $\lambda_1 = -\frac{1}{2\sigma^2}$ *&* $\lambda_2 = \ln \frac{1}{\sqrt{2\pi\sigma^2}}$

$$\therefore p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

Then, the entropy is given by

$$\begin{aligned} H(x) &= -\int_{-\infty}^{\infty} p(x) \log_2 p(x) dx \\ &= \int_{-\infty}^{\infty} p(x) \log_2 \sqrt{2\pi\sigma_2} dx + \int_{-\infty}^{\infty} p(x) \frac{x^2}{2\sigma^2} \log_2 e dx \\ &= \log_2 \sqrt{2\pi\sigma^2} + \frac{1}{2} \log_2 e \\ &= \log_2 \sqrt{2\pi e \sigma^2} \end{aligned}$$

- Conditional entropy

$$H(Y|X) = \mathbb{E}\left[\log_2 \frac{1}{p(Y|X)}\right] = \int p(x) H(Y|X=x) dx$$

- For $Y = X + N$ assuming X & N are independant, $H(Y|X = x) = H(N|X = x)$ since Y is just a shifted version of N for a given x. Thus $H(Y|X)$ is given by $H(Y|X) = \int p(x)H(N|X = x)dx = \int p(x)H(N)dx = H(N)$.

- Mutual Information $I(X;Y)$

$I(X;Y)$ indicates the reduction in the uncertainty of X due to the knowledge of Y, and is defined by

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

- Channel capacity

consider a probabilistic channel model

$$X \quad \longrightarrow \quad \boxed{p(y|x)} \quad \longrightarrow \quad Y$$

The channel capacity is defined by the maximum of the mutual information between X & Y

$$C \triangleq \max_{p(x)} I(X;Y)$$

The capacity represents the maximum transmission rate of any communication systems

Theorem 2 *Consider a AWGN channel $Y = X + N$ where X has a power constraint and is independent of N. Then, the channel capacity equals*

$$C = \frac{1}{2}\log_2\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right),$$

where σ_X^2, σ_X^2 indicate variance of X & Y.

Proof 2

$$C = \max_{p(x)} (H(Y) - H(Y|X)) = \max_{p(x)} (H(Y) - H(N))$$

Since $H(Y) \leq \log_2 \sqrt{2\pi e(\sigma_X^2 + \sigma_N^2)}$ with equality when Y is Gaussian, we have

$$\begin{aligned} C &= \frac{1}{2}\log_2\left(2\pi e(\sigma_X^2 + \sigma_N^2)\right) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \\ &= \frac{1}{2}\log_2 \frac{\sigma_X^2 + \sigma_Y^2}{\sigma_N^2} = \frac{1}{2}\log_2\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \end{aligned}$$

C is achieved when X is Gaussian.

1.3 Space-Time (ST) configuration

- single-input, single-output (SISO)
  single-input, multiple-output (SIMO)
  multiple-input, single-output (MISO)
  multiple-input, multiple-output (MIMO)

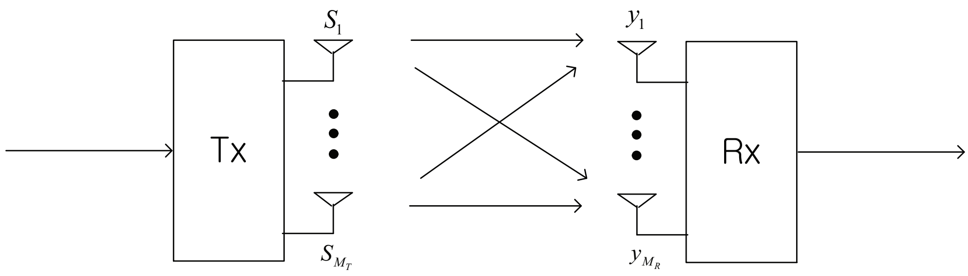

Figure 1.1: MIMO discrete-time channel model

- MIMO discrete-time channel model

Consider a flat-fading MIMO channel with M_T Tx antennas & M_R Rx antennas.

The received signal at the ith Rx antenna is given by

$$y_i(t) = \sqrt{E_s} \sum_{j=1}^{M_T} h_{i,j} S_j + n_i \qquad \text{for} \;\; i = 1, 2, \cdots, M_R$$

where $h_{i,j}$: fading coefficeint between the jth Tx and the ith Rx antenna
E_s : average symbol energy $E_s = \mathbb{E}|s|^2$

Here $h_{i,j}$is an independent identical distributed(i,i,d) Complex Gaussian with zero mean & unit variance. $(\mathbb{E}(h_{i,j}) = 0, \mathbb{E}(|h_{i,j}|^2) = 1, \mathbb{E}(h_{i,j}h^*_{m,n}) = 0)$

- In a matrix form, We have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{M_R} \end{bmatrix} = \sqrt{E_s} \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,M_T} \\ h_{2,1} & h_{22} & \cdots & h_{2M_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_R,1} & h_{M_R,2} & \cdots & h_{M_R,M_T} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{M_T} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{M_R} \end{bmatrix}$$

Equivalently, $\mathbf{y} = \sqrt{E_s}\mathbf{H}\mathbf{s} + \mathbf{n}$

where $\mathbf{H}$: $M_R \times M_T$ channel matrix, $\quad \mathbf{H} \in \mathbb{C}^{M_R \times M_T}$
$\mathbf{y}, \mathbf{n}$: column vector of length M_R, $\quad \mathbf{y}, \mathbf{n} \in \mathbb{C}^{M_R \times 1}$
$\mathbf{s}$: column vector of length M_T, $\quad \mathbf{s} \in \mathbb{C}^{M_T \times 1}$

- Normalization

Total Tx power should be the same regardless of M_T

$\Rightarrow$ Tx Symbol energy per Tx antenna is normalized to $\frac{E_s}{M_T}$

1.4 Spatial correlation

- In practice, the channel may have spatial correlation due to inadequate antenna spacing, insufficient scattering or LOS element

- Correlated channel model

$$\mathbf{H} = \mathbf{R}_r^{\frac{1}{2}} \mathbf{H}_w \mathbf{R}_t^{\frac{1}{2}}$$

where $\mathbf{R}_t$: Tx covariance matrix of size $M_T \times M_T$

$\mathbf{R}_r$: Rx covariance matrix of size $M_R \times M_R$

$\mathbf{H}_w$: spatially white channel

1.5 MIMO channel estimation

- The channel estimation is obtained by transmitting a known training sequence to Rx.(Sounding)

- Estimation complexity grows with delay spread, M_T, M_R Doppler, etc.

1.6 Channel state information (CSI)

- open loop : CSI is available only at Rx.

closed loop : CSI is available both Tx & Rx.

- CSI at Tx can improve the system performance by using precoder design, etc.

- In closed loop frequency division duplex(FDD) systems the CSI is first estimated at Rx, and then transmitted to Tx through reliable feedback channels. Feedback delay should be much smaller than the coherence time.

- Reciprocity principle
If time, frequency and antennas are the same, then the channels in the forward and backward links are identical.

- Estimation using reciprocity.

- In time division duplex(TDD), the reverse channel is directly obtained from the forward channel as long as the duplexing time delay is much smaller than coherence time.

- Adaptive modulation and coding(AMC) adjusts modulation levels and code rates depending on the channel conditions.

1.7 MIMO channel estimation

- For a fading channel $y = hx + n$, the received power of y equals $\sigma_y^2 = |h|^2\sigma_x^2 + \sigma_n^2$

- The channel capacity is maximized when x is Gaussian

$$C = \max_{p(x)} I(X;Y) = \log_2(1 + |h|^2\rho) \quad \text{bps/Hz}$$

where $\rho = \sigma_x^2/\sigma_n^2$
$\rightarrow$ Capacity is also a $R.V$ depending on channel realization.

- Cutoff rate R_o
For $R < R_o$, the average probability of error approaches 0 as the code block length goes to infinity.

Chapter 2

MIMO Capacity

2.1 Ergodic capacity $C = \mathbb{E}_h(C)$

- Ergodic channel : For each channel use an independent realization of h is drawn from a given distribution.

- Assuming that the transmission time is long enough to reveal the long-term ergodic properties of the ergodic fading channels, the ergodic capacity is defined as the ensemble average of the information rate

$$\begin{aligned} C &= \mathbb{E}_h[\log_2(1+|h|^2\rho)] \\ &= \int_0^\infty \log_2(1+|h|^2\rho) f_{|h|^2}(\alpha) d\alpha \end{aligned}$$

- Ergodic capacity can be achieved by using an adaptive transmission policy where the power and data rate vary relative to CSI and transmitting a codeword over a very long number of independent fading block assuming infinite delay.$\rightarrow$ fast fading

- Using the Jensen's inequality $\mathbb{E}(f(x)) \leq f(\mathbb{E}(x))$ for a concave function $f(x)$, we have

$$C = \mathbb{E}_h[\log_2(1+|h|^2\rho)] \leq \log_2(1+\mathbb{E}|h|^2\rho)$$

This indicates that the ergodic capacity of an open loop fading channel is less than the AWGN channel capacity with the same average SNR.

2.2 Outage capacity (slow fading)

- For non-ergodic channels where h is chosen randomly at the begining of the transmission and is held fixed for all channel users ergodic capacity has no meaning.

- Outage probability : $P_{out} = P_r(C \leq C_{out}^{\epsilon})$
Probability that the transmission rate C_{out}^{ϵ} exceeds the channel capacity.

- Outage capacity : the maximum rate that can be supported by the channel with a given outage probability.

2.3 MIMO capacity

- MIMO signal model

$$\mathbf{y} = \sqrt{\frac{E_s}{M_t}}\mathbf{H}\mathbf{s} + \mathbf{n}$$

where $\mathbf{H}$: $M_R \times M_T$ channel matirx
$\mathbf{y}, \mathbf{n}$: column vector of length M_R
$\mathbf{s}$: column vector of length M_T

- The Covariance Matrix of $\mathbf{y}$ is given by

$$\begin{aligned}
\mathbf{R}_{\mathbf{yy}} \triangleq \mathbb{E}[\mathbf{y}\mathbf{y}^*] &= \mathbb{E}\left[\left(\sqrt{\frac{E_s}{M_T}}\mathbf{H}\mathbf{s} + \mathbf{n}\right)\left(\sqrt{\frac{E_s}{M_T}}\mathbf{H}\mathbf{s} + \mathbf{n}\right)^*\right] \\
&= \frac{E_s}{M_T}\mathbf{H}\mathbf{R}_{\mathbf{ss}}\mathbf{H}^* + \mathbf{R}_{\mathbf{n}}
\end{aligned}$$

where $\mathbf{R}_{\mathbf{ss}} \triangleq \mathbb{E}[\mathbf{s}\mathbf{s}^*]$: covariance matrix of $\mathbf{s}$
$\mathbf{R}_{\mathbf{n}} \triangleq \mathbb{E}[\mathbf{n}\mathbf{n}^*]$: covariance matrix of $\mathbf{n}$

- m-dimensional complex Gaussian distribution

$$p(\mathbf{x}) = \frac{1}{\pi^m |\mathbf{R}_{\mathbf{x}}|} \exp\left(-(\mathbf{x} - \mu)^* \mathbf{R}_{\mathbf{x}}^{-1} (\mathbf{x} - \mu)\right) \tag{2.1}$$

where μ : mean of $\mathbf{x}$
$\mathbf{R_X} = \mathbb{E}[\mathbf{xx}^*]$ covariance matrix of $\mathbf{x}$.
For AWGN with $\mathbf{R_n} = N_0\mathbf{I}$, $p(\mathbf{x}) = \frac{1}{(\pi N_0)^m}\exp\left(-\frac{|\mathbf{X}|^2}{N_0}\right)$

Theorem 3 *For complex Gaussian* $\mathbf{x}$*,* $H(\mathbf{x})$ *is given by*

$$H(\mathbf{x}) = \log_2 |\pi e \mathbf{R_X}|$$

Proof 3

$$\begin{aligned} H(\mathbf{x}) &= -\mathbb{E}[\log_2 p(\mathbf{x})] \\ &= \mathbb{E}[\log_2 \pi^m |\mathbf{R_X}| + \mathbf{x}^*\mathbf{R_X}^{-1}\mathbf{x}\log_2 e] \\ &= \log_2 |\pi\mathbf{R_X}| + \log_2 e\mathbb{E}[\mathbf{x}^*\mathbf{R_X}^{-1}\mathbf{x}] \end{aligned}$$

Here $\mathbb{E}[\mathbf{x}^*\mathbf{R_X}^{-1}\mathbf{x}]$ *is computed by*

$$\begin{aligned} \mathbb{E}[\mathbf{x}^*\mathbf{R_X}^{-1}\mathbf{x}] &= \mathbb{E}[\mathrm{tr}(\mathbf{x}^*\mathbf{R_X}^{-1}\mathbf{x})] = \mathbb{E}[\mathrm{tr}(\mathbf{R_X}^{-1}\mathbf{xx}^*)] \\ &= \mathrm{tr}(\mathbf{R_X}^{-1}\mathbb{E}[\mathbf{xx}^*]) = \mathrm{tr}(\mathbf{I}) = m \\ \therefore H(\mathbf{x}) &= \log_2 |\pi\mathbf{R_X}| + \log_2 e^m \\ &= H(\mathbf{x}) = \log_2 |\pi e\mathbf{R_X}| \end{aligned}$$

- Mutual information

$$I(\mathbf{s}, \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{s}) = H(\mathbf{y}) - H(\mathbf{n})$$

This is maximized when $\mathbf{y}$ is Gaussian

$$\therefore C = \max I(\mathbf{s}, \mathbf{y}) = \log_2 |\pi e\mathbf{R_{yy}}| - \log_2 |\pi e\mathbf{R_n}|$$

Using $\frac{|\mathbf{A}|}{|\mathbf{B}|} = |\mathbf{AB}^{-1}| = |\mathbf{B}^{-1}\mathbf{A}|$, we have

$$\begin{aligned} C &= \log_2 \frac{|\mathbf{R_{yy}}|}{|\mathbf{R_n}|} = \log_2 |\mathbf{R_n}^{-1}\mathbf{R_{yy}}| \\ &= \log_2 \left|\frac{E_s}{M_T}\mathbf{R_n}^{-1}\mathbf{H}\mathbf{R_{ss}}\mathbf{H}^* + \mathbf{I}\right| \qquad (\log - \det) \end{aligned}$$

- For AWGN ($\mathbf{R} = N_0\mathbf{I}$),

$$C = \log_2 \left|\mathbf{I} + \frac{E_s}{M_T}\mathbf{H}\mathbf{R_{ss}}\mathbf{H}^*\right|$$

2.4 Open loop capacity

- We assume that CSI is available only at Rx. Thus independent equal power signal $\mathbf{R}_{\mathbf{S}]sv} = \mathbf{I}$ achieve the open loop capacity

$$C = \log_2 |\mathbf{I} + \rho \mathbf{H}\mathbf{H}^*| \qquad \text{where} \quad \rho = \frac{E_s}{M_T N_0}$$

- Consider the eigenvalue decomposition of $\mathbf{H}\mathbf{H}^*$ as

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

where $\mathbf{Q}$ is a unitary matrix ($\mathbf{Q}\mathbf{Q}^* = \mathbf{Q}^*\mathbf{Q} = \mathbf{I}$)
$\mathbf{\Lambda} = \text{diag}\{\lambda_1, \cdots, \lambda_{M_R}\}$

- Using $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$, it follows

$$\begin{aligned} C &= \log_2 |\mathbf{I} + \rho \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^*| = \log_2 |\mathbf{I} + \rho\mathbf{\Lambda}| \\ &= \log_2 \left| \begin{bmatrix} 1+\rho\lambda_1 & & & \mathbf{0} \\ & 1+\rho\lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & 1+\rho\lambda_{M_R} \end{bmatrix} \right| \\ &= \log_2 \prod_{i=1}^{M_R} (1+\rho\lambda_i) \\ &= \sum_{i=1}^{r} \log_2 (1+\rho\lambda_i) \end{aligned}$$

where $r:$ rank of the channel ($r = \text{rank}(\mathbf{H}) = \text{rank}(\mathbf{H}\mathbf{H}^*)) \leq \min(M_T, M_R)$
$\Rightarrow$ MIMO capacity is equivalent to the sum of the SISO capacities, each having power gain $\sqrt{\lambda_i}$

- Power constraint

$$\begin{aligned} E_s &= \mathbb{E}|\mathbf{s}|^2 = \mathbb{E}[\mathbf{s}^*\mathbf{s}] = \mathbb{E}[\text{tr}(\mathbf{s}^*\mathbf{s})] = \mathbb{E}[\text{tr}(\mathbf{s}\mathbf{s}^*)] \\ &= \text{tr}(\mathbb{E}[\mathbf{s}\mathbf{s}^*]) = \text{tr}(\mathbf{R}_{\mathbf{SS}}) \end{aligned}$$

2.5 Closed loop capacity

[**ref**] : "Capacity of Multi-antenna Gaussian channels", E. Telatar, Euro. Trans. Telecom., 1999

- Assuming that CS is available both at Tx and Rx, the input covariance $\mathbf{R_{SS}}$ can be optimized based on two know ledge of the channel

- Consider singular value decomposition(SVD) of $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U\Sigma V}^*$$

where $\mathbf{U}$: $M_R \times M_R$ unitary matrix whose column are eigenvectors of $\mathbf{HH}^*$
$\mathbf{V}$: $M_T \times M_T$ unitary matrix whose column are eigenvectors of $\mathbf{H}^*\mathbf{H}$
$\mathbf{\Sigma}$: $M_R \times M_T$ real diagonal matrix with diagonal entries $\sigma_1, \cdots, \sigma_r$
Here, the singular value σ_i of $\mathbf{H}$are equal to the square root of the eigenvalue λ_i of $\mathbf{HH}^*$ or $\mathbf{H}^*\mathbf{H}$, since

$$\begin{aligned} \mathbf{HH}^* &= \mathbf{U\Sigma V^* V \Sigma^* U^*} = \mathbf{U\Sigma^2 U^*} \\ \mathbf{H^*H} &= \mathbf{V\Sigma U^* U \Sigma^* V^*} = \mathbf{V\Sigma^2 V^*} \qquad (\sigma_i = \sqrt{\lambda_i}) \end{aligned}$$

Then, the channel model becomes

$$\mathbf{y} = \sqrt{\frac{E_s}{M_T}}\mathbf{U\Sigma V^* s} + \mathbf{n}$$

- Defining $\tilde{\mathbf{s}} = \mathbf{V}^*\mathbf{s}$ as the precoded input, the received signal $\tilde{\mathbf{y}} = \mathbf{U}^*\mathbf{y}$ becomes

$$\tilde{\mathbf{y}} = \sqrt{\frac{E_s}{M_T}}\mathbf{\Sigma}\tilde{\mathbf{s}} + \tilde{\mathbf{n}}$$

where $\tilde{\mathbf{n}} = \mathbf{Un}$ with $\mathbb{E}[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^*] = \mathbb{E}[\mathbf{U^* nn^* U}] = \mathbf{U}^*\mathbb{E}[\mathbf{nn}^*]\mathbf{U} = N_0\mathbf{I}$

$$\begin{aligned} \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \vdots \\ \tilde{y}_{M_R} \end{bmatrix} &= \sqrt{\frac{E_s}{M_T}} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{M_T} \\ & \mathbf{0} & \end{bmatrix} \begin{bmatrix} \tilde{s}_1 \\ \vdots \\ \tilde{s}_{M_T} \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \vdots \\ \vdots \\ \tilde{n}_{M_R} \end{bmatrix} \\ \Rightarrow \tilde{y}_i &= \sqrt{\frac{E_s}{M_T}}\sigma_i \tilde{s}_i + \tilde{n}_i \qquad \text{for} \quad i = 1, \cdots, r \end{aligned}$$

Here $\tilde{y}_i$ for $i > r$ is independent of the transmit data and $\tilde{s}_i$ for $i > r$ does not play any role.

- To maximize the mutual information, we need to choose $\tilde{s}_i$ having independent Gaussian. Then capacity becomes

$$C = \sum_{i=1}^{r} \log_2(1 + \rho\gamma_i\sigma_i^2)$$

where $\gamma_i = \mathbb{E}|\tilde{s}_i|^2$ denotes the power for the i^{th} eigenmode with power constraint $\sum_i \gamma_i \leq P$

- Now we need to find the optimal energy allocation for γ_i to maximize the capacity using Lagrange multiplier.

$$J = \sum_{i=1}^{r} \log_2(1 + \rho\gamma_i\sigma^2) + \lambda \sum_{i=1}^{r} \gamma_i$$

By differentiating J with respect to γ_j, we have

$$\begin{aligned} \frac{\partial J}{\partial \gamma_j} = \frac{\log_2 e\rho\sigma_i^2}{1 + \rho\gamma_i\sigma_i^2} + \lambda = \quad 0 \qquad &\rightarrow \qquad \frac{1}{\gamma_i + \frac{1}{\rho\sigma_i^2}} + \lambda' = 0 \\ \gamma_i + \frac{1}{\rho\sigma_i^2} = \lambda'' \qquad &\text{or} \qquad \gamma_i = \lambda'' - \frac{1}{\rho\sigma_i^2} \end{aligned}$$

Since $\gamma_i \leq 0$, γ_i is expressed as

$$\gamma_i = \left(\lambda'' - \frac{1}{\mu\sigma_i^2}\right)^+ \qquad ; \quad \text{water} - \text{filling} \quad \text{solution}$$

where $(a)^+ = \max(a, 0)$ and μ is chosen to meet the power constraint $\sum_i \gamma_i \leq P$

- The SVD operation combined with the water pouring achieves the closed loop capacity. The power is allocated to dominant eigenvectors of $\mathbf{H}$.

- Since the nonzero eigenvalues of $\mathbf{H}^*\mathbf{H}$ are the same as those of $\mathbf{H}\mathbf{H}^*$, the capacities of channels corresponding to $\mathbf{H}$ and $\mathbf{H}^*$ are the same.(reciprocity)

2.6 SIMO capacity ($M_T = 1$)

- For open loop case, the capacity becomes

$$\begin{aligned} C &= \log_2 \left| \ \mathbf{I} + \tfrac{E_s}{M_T N_0} \mathbf{H}\mathbf{H}^* \ \right| \\ &= \log_2 \left| \ \mathbf{I} + \tfrac{E_s}{N_0} \mathbf{h}\mathbf{h}^* \ \right| \end{aligned}$$

where $\quad \mathbf{h} = [h_1 \cdots h_{M_R}]^T$
Using $|\mathbf{I} = \mathbf{AB}| = |\mathbf{I} = \mathbf{BA}|$

$$\begin{aligned} C &= \log_2 \left(1 + \frac{E_s}{N_0} \|\mathbf{h}\|^2 \right) \\ &= \log_2 \left(1 + \frac{E_s}{N_0} \sum_{j=1}^{M_R} |h_j|^2 \right) \end{aligned}$$

- The SIMO capacity is achieved by maximum ratio combining(MRC). Additional Rx antenna yields only a logarithmic increase in capacity.

- CSI at Tx provides no gain in capacity.

2.7 MISO capacity ($M_R = 1$)

$$\begin{aligned} C &= \log_2 \left| \ \mathbf{I} + \tfrac{E_s}{M_T N_0} \mathbf{h}\mathbf{h}^* \ \right| \\ &= \log_2 \left(1 + \frac{E_s}{M_T N_0} \sum_{i=1}^{M_T} |h_i|^2 \right) \end{aligned}$$

where $\quad \mathbf{h} = [h_1 \cdots h_{M_T}]$

- A capacity gain due to additional Tx antenna is smaller than SIMO. The MISO capacity is attainable with Tx beamforming(Tx MRC).

- For high SNR values, MISO has a penalty of $\log_2 M_T$ over SIMO.

2.8 Ergodic capacity of MIMO

- The ergodic capacity $\bar{C}$ of a MIMO channel is the ensemble average of the information rate over the distribution of the elements of $\mathbf{H}$.

- When the channel is unknown to Tx, $\bar{C}$ is given by

$$\begin{aligned}\bar{C} &= \mathbb{E}_H \left(\log_2 \left| \ \mathbf{I} + \rho \mathbf{H}\mathbf{H}^* \ \right|\right) \\ &= \mathbb{E}_H \left(\sum_{i=1}^{r} \log_2 (1 + \rho\lambda_i) \right)\end{aligned}$$

- For high SNR, $\bar{C}$ becomes

$$\begin{aligned}\bar{C} &\simeq \sum_{i=1}^{r} \mathbb{E}_H(\log_2 (\rho\lambda_i)) \\ &= \sum_{i=1}^{r} (\log_2 \rho + \mathbb{E}_H(\log_2 \lambda_i)) \\ &= r \log_2 \rho + \mathbb{E}_H \left(\sum_{i=1}^{r} \log_2 \lambda_i \right)\end{aligned}$$

Since the second term is finite, $\bar{C}$ grows linearly with r.

- For the large system limit of $M_R = M_T = M \to \infty$, from the strong law of large numbers, we have

$$\begin{aligned} \frac{1}{M}\mathbf{H}\mathbf{H}^* &= \mathbf{I}_n \\ \bar{C} &= \mathbb{E}_H\left(\log_2 \left|\ \mathbf{I} + \tfrac{E_s}{M_T N_o}\mathbf{H}\mathbf{H}^*\ \right|\right) \\ &\rightarrow \mathbb{E}_H\left(\log_2 \left|\ \mathbf{I} + \tfrac{E_s}{N_o}\mathbf{I}\ \right|\right) \\ &= M \log_2 (1 + \frac{E_s}{N_o}) \end{aligned}$$

$\therefore$ The MIMO ergodic capacity increases by M times compared with the SISO capacity.

- Exact computation of the ergodic capacity was first derived in Telatar's paper.

- Ergodic capacity grows linearly with $M_T = M_R$(Fig 4,5)

- The capacity gain when CSI is available at Tx reduces at high SNR. The water pouring solution approaches the uniform power distribution at high SNR.(Fig 4.6)

Chapter 3

MIMO Systems

• Space-time coding(STC) : multiple Tx antennas are exploited to achieve a diversity gain for higher link performance.
• Spatial Multiplexing(SM) : independent data streams are transmitted through each Tx antenna to increase the throughput.

3.1 Diversity techniques

- Variation is signal levels can result is low SNR and poor link performance. Diversity methods are used to mitigate the fading effects.

- Diversity approaches
1) Time diversity : Channel coding with inter leaver.
2) Frequency diversity : The signal is transmitted through subcarriers with different channel gains.(OFDM)
3) Antenna(space) diversity
   Rx diversity : MRC
   Tx diversity : STC

3.1.1 MRC (SIMO)

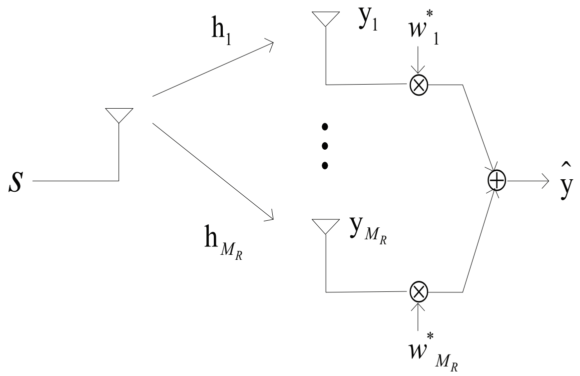

Figure 3.1: MRC

- The received signal at the ith antenna is given by

$$y_i = \sqrt{E_s} h_i s + n_i \qquad \text{for} \quad i = 1, \cdots, M_R$$

where n_i is the AWGN with variance N_o.
Then, $\mathbf{y} = [y_1, y_2, \cdots, y_{M_R}]^T$ is equal to

$$\mathbf{y} = \sqrt{E_s}\mathbf{h}s + \mathbf{n}$$

where $\mathbf{h} = [h_1, h_2, \cdots, h_{M_R}]^T$
$\mathbf{n} = [n_1, n_2, \cdots, n_{M_R}]^T$

- The combined signal $\tilde{y}$ is expressed by

$$\tilde{y} = \sum_{i=1}^{M_R} w_i^* y_i = \mathbf{w}^*\mathbf{y} = \sqrt{E_s}\mathbf{w}^*\mathbf{h}s + \tilde{n}$$

where $w = [w_1, w_2, \cdots, w_{M_R}]^T$: Rx combining vector.
$\tilde{n} = \mathbf{w}^*\mathbf{n}$

Theorem 4 *The output SNR is maximized with $w_i = h_i$*

Proof 4 *The output SNR is defined as*

$$\text{SNR} = \frac{\mathbb{E}(E_S|\mathbf{w}^*\mathbf{h}s|^2)}{\mathbb{E}|\tilde{n}|^2} = \frac{E_s|\mathbf{w}^*\mathbf{h}|^2}{N_o\|\mathbf{w}\|^2}$$

using the Schwarz inequality $|\mathbf{a}^*\mathbf{b}|^2 \le \|\mathbf{a}\|^2\|\mathbf{b}\|^2$ *where equality holds when* $\mathbf{b} = k\mathbf{a}$ *with a constant* k*, we have*

$$\text{SNR} \le \frac{E_s}{N_o}\|\mathbf{h}\|^2 \quad \text{with equality when } w_i = h_i$$

- Denoting α as $\alpha = \sum_{i=1}^{M_R} |h_i|^2 = \|\mathbf{h}\|^2$, the probability of error with BPSK equals $Q(\sqrt{\frac{\alpha E_s}{N_o}})$.

Using the Chernoff bound ($Q(x) \le \exp\left(-\frac{x^2}{2}\right)$), it follows

$$P_{e|h} \le \exp\left(-\alpha\frac{E_s}{2N_0}\right)$$

where α has a Chi-square distribution with $2M_R$ degree of freedom as

$$f_\alpha(x) = \frac{x^{M_R-1}}{(M_R-1)!}e^{-x} \qquad , \quad x \ge 0$$

- Averaging $P_{e|h}$ over the fading distribution yields

$$\begin{aligned} P_e &= \int_0^\infty f_\alpha(x)P_{e|h}dx \\ &\le \int_0^\infty \frac{x^{M_R-1}}{(M_R-1)!}\exp(-x)\exp\left(-\alpha\frac{E_s}{2N_0}\right)dx \\ &= \prod_{i=1}^{M_R}\frac{1}{1+\frac{E_s}{2N_0}} \\ &= \left(1+\frac{E_s}{2N_0}\right)^{-M_R} \end{aligned}$$

$\therefore$ For high SNR,

$$P_e \le \left(\frac{E_s}{2N_0}\right)^{-M_R}$$

- Diversity order

$$D = -\lim_{\text{SNR}\to\infty}\frac{P_e(\text{SNR})}{\log(\text{SNR})}$$

the slope of BER/FER curves in the log-log scale

- Diversity gain determines the slope of BER curves while a coding specifies the horizontal shift of the curve.

- The maximum diversity order of $M_T \times M_R$ MIMO systems is

$$D = M_T \cdot M_R \quad \Rightarrow \quad \text{full diversity}$$

- For SISO systems, $P_e \propto \frac{E_s}{N_0}^{-1} \rightarrow D = 1$
$\Rightarrow$ BER decreases linearly with SNR
cf) AWGN : $P_e = Q(\sqrt{\frac{E_s}{N_0}}) \leq \exp\left(-\frac{1}{2}\frac{E_s}{N_0}\right)$

- The diversity order of MRC is M_R. As M_R grows infinity, the MRC performance approaches the AWGN.

- Array Gain : the average SNR increase at Rx that arrises form the coherent combining effect of multiple Rx antennas.

- The MRC benefits from the array gain $10 \log M_R$dB over a SISO link.

3.2 Space-time Coding (STC)

- No channel information knowledge is assumed at Tx

- Consider the channel output at time k as

$$\mathbf{y}_k = \sqrt{\frac{E_s}{M_T}} \bar{\mathbf{H}} \mathbf{s}_k + \mathbf{n}_k, \quad k = 1, \cdots, T \text{ (MIMO)}$$

where T : code block length

- The channel matrix $\bar{\mathbf{H}} \in \mathbb{C}^{M_R \times M_T}$ is assumed to be quasi-static flat fading duration T slots.

- The By collection T signals together, we have

$$\mathbf{Y} = \sqrt{\frac{E_s}{M_T}} \bar{\mathbf{H}} \mathbf{S} + \mathbf{N}$$

where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_T] \in \mathbb{C}^{M_R \times T}$
$\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_T] \in \mathbb{C}^{M_R \times T}$
$\mathbf{N} = [\mathbf{n}_1, \cdots, \mathbf{n}_T] \in \mathbb{C}^{M_R \times T}$

-ML decoding

$$\begin{aligned}\hat{\mathbf{S}} &= \arg \min_{\mathbf{S}} ||\mathbf{Y} - \sqrt{\frac{E_s}{M_t}}\mathbf{HS}||_F^2 \\ &= \arg \min_{\mathbf{S}} \sum_{k=1}^{T} ||\mathbf{y}_k - \sqrt{\frac{E_s}{M_t}}\mathbf{H}\mathbf{s}_k||^2\end{aligned}$$

where $||\cdot||_F^2$ indicate Frobenius norm
$(||\mathbf{A}||_F^2 = \mathrm{tr}(\mathbf{AA}^*))$

3.2.1 Alamouti code

$(M_T = T = 2)$

- In the first time slot, s_1 & s_2 are transmitted at two Tx antennas, and in the second time slot, $-s_2^*$ & s_1^* are transmitted.

- Channel is assumed to be constant over two time slot

- Let y_i be the received signal at time i $(i = 1, 2)$

$$\begin{aligned}y_1 &= \sqrt{\frac{E_s}{2}}(h_1 s_1 + h_2 s_2) + n_1 \\ y_2 &= \sqrt{\frac{E_s}{2}}(-h_1 s_2^* + h_2 s_1^*) + n_2\end{aligned}$$

where n_i is the complex Gaussian noise with variance N_0

- In a matrix form, we have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \sqrt{\frac{E_s}{2}} \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

- By applying the conjugate of y_2, it follows

$$\begin{aligned} y_1 &= \sqrt{\frac{E_s}{2}}(h_1 s_1 + h_2 s_2) + n_1 \\ y_2^* &= \sqrt{\frac{E_s}{2}}(h_2^* s_1 - h_1^* s_2) + n_2^* \end{aligned}$$

- Denoting $\mathbf{y}$as $\mathbf{y} = [\ y_1 \quad y_2^*\ ]^T$, we have

$$\begin{aligned} \mathbf{y} &= \sqrt{\frac{E_s}{2}} \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix} \\ &= \sqrt{\frac{E_s}{2}} \mathbf{Hs} + \mathbf{n} \end{aligned}$$

where $\mathbf{H}$ is orthogonal $\mathbf{H}^*\mathbf{H} = \begin{bmatrix} |h_1|^2 + |h_2|^2 & 0 \\ 0 & |h_1|^2 + |h_2|^2 \end{bmatrix} = \alpha \mathbf{I}$
and $\mathbb{E}\{\mathbf{nn}^*\} = N_0 \mathbf{I}$

- By multiplying $\mathbf{H}^*$ to $\mathbf{y}$ (matched filtering), it follows

$$\begin{aligned} \mathbf{z} = \mathbf{H}^*\mathbf{y} &= \sqrt{\frac{E_s}{2}} \mathbf{H}^*\mathbf{Hs} + \mathbf{H}^*\mathbf{n} \\ &= \sqrt{\frac{E_s}{2}} \alpha \mathbf{s} + \tilde{\mathbf{n}} \qquad (\tilde{\mathbf{n}} = \mathbf{H}^*\mathbf{n}) \end{aligned}$$

- Then, the matched filter output becomes

$$z_i = \sqrt{\frac{E_s}{2}} \alpha s_i + \tilde{n}_i \qquad i = 1, 2$$

where $\mathbb{E}[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^*] = \mathbb{E}[\mathbf{H}^*\mathbf{nn}^*\mathbf{H}] = \mathbf{H}^*\mathbb{E}[\mathbf{nn}^*] = N_0\mathbf{H}^*\mathbf{H} = \alpha N_0 \mathbf{I}$

- The received SNR is expressed as

$$\text{SNR} = \frac{\frac{E_s}{2}\alpha^2}{\alpha N_0} = \frac{\alpha}{2}\frac{E_s}{N_0} = \frac{1}{2}(|h_1|^2 + |h_2|^2)\frac{E_s}{N_0}$$

$\Rightarrow$ The diversity performance is equivalent to MRC ($D = 2$)

- Matched filtering decouples two streams and allows independent ML decoding on each stream (Very low complexity)

- Due to power normalization factor, the Alamouti code has 3dB loss compared to the MRC (No array gain)

- Diversity stabilizes the fading fluctuations (channel hardening)

- code rate (multiplexing gain) r_x : the ratio of the number of transmitted symbols to the transmission time slot. $r_s = 1$ indicates full rate.

- Alamouti code achieves full rate ($r_s = \frac{2}{2} = 1$) and full diversity ($D = 2$)

- For complex constellations, and orthogonal design with full rate and full diversity does not exist when $M_T > 2$

- For system with $M_T > 2$, a penalty incurs in either orthogonality (decoder complexity), code rate (rate loss), or diversity (performance loss)

- From the original signal nodel with $M_R = 1$

$$\mathbf{Y} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}, \quad \tilde{\mathbf{H}} = \begin{bmatrix} h_1 & h_2 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix}$$

- orthogonal STC for $M_t = 4$ (Tirkkoen)

$$\mathbf{C}(s) = \begin{bmatrix} s_1 & s_2 & s_3 & 0 \\ -s_2^* & s_1^* & 0 & -s_3 \\ -s_3^* & 0 & s_1^* & s_2 \\ 0 & s_3^* & -s_2* & s_1 \end{bmatrix} \quad \rightarrow \quad \begin{array}{c} \text{orthogonal} \\ r_s = \frac{3}{4} \\ D = 4 \end{array}$$

Higher modulation level should be employed to compensate rate loss.

- In next generation systems. the role of STC is getting smaller

i) A diversity gain from STC may become marginal for channels with frequency diversity & time diversity

ii) Closed loop systems become more popular and outperform open loop systems such as STC

3.2.2 Pairwise error probability (PEP) for STC

$$\mathbf{Y} = \sqrt{\frac{E_s}{M_T}}\mathbf{H}\mathbf{s} + \mathbf{N}$$

- PEP represents the probability that ML mistakes a codeword $\mathbf{s}$ for an errorous codework $\hat{\mathbf{s}}$.

- Given $\mathbf{H}$, the conditional PEP is given by

$$\begin{aligned} P_{e|H} &= P(\mathbf{s} \rightarrow \hat{\mathbf{s}} \mid \mathbf{H}) \\ &= P(||\mathbf{Y} - \sqrt{\frac{E_s}{M_T}}\mathbf{H}\mathbf{s}||_F^2 > ||\mathbf{Y} - \sqrt{\frac{E_s}{M_T}}\mathbf{H}\hat{\mathbf{s}}||_F^2 | \ \mathbf{H}) \\ &= P(||\mathbf{N}||_F^2 > ||\mathbf{N} - \sqrt{\frac{E_s}{M_T}}\mathbf{H}\mathbf{E}||_F^2 | \ \mathbf{H}) \qquad \text{where} \ \ \mathbf{E} = \hat{\mathbf{s}} - \mathbf{s} \end{aligned}$$

Since $||\mathbf{A}+\mathbf{B}||_F^2 = ||\mathbf{A}||_F^2 + ||\mathbf{B}||_F^2 + \text{trace}(\mathbf{A}\mathbf{B}^* + \mathbf{B}\mathbf{A}^*)$, we have

$$\begin{aligned} P_{e|H} &= P(0 > \frac{E_s}{M_T}||\mathbf{H}\mathbf{E}||_F^2 - \sqrt{\frac{E_s}{M_T}}\text{trace}(2 \times \Re\{\mathbf{H}\mathbf{E}\mathbf{N}^*\})| \ \mathbf{H}) \\ &= P(\text{trace}(\Re\{\mathbf{H}\mathbf{E}\mathbf{N}^*\}) > \frac{1}{2}\sqrt{\frac{E_s}{M_T}}||\mathbf{H}\mathbf{E}||_F^2 | \ \mathbf{H}) \end{aligned}$$

Since elements in $\mathbf{N}$ are white Gausssian with variance N_0

$$\text{var}(\text{trace}(\Re\{\mathbf{H}\mathbf{E}\mathbf{N}^*\})) = \tfrac{N_0}{2}||\mathbf{H}\mathbf{E}||_F^2$$

$$\therefore \ \ P_{e|H} = Q\left(\tfrac{1}{2}\sqrt{\tfrac{E_s}{M_T}}\frac{||\mathbf{H}\mathbf{E}||_F^2}{\sqrt{\frac{N_0}{2}||\mathbf{H}\mathbf{E}||_F^2}}\right) = Q\left(\sqrt{\tfrac{E_s}{2M_T N_0}||\mathbf{H}\mathbf{E}||_F^2}\right)$$

- Using the chernoff bound $Q(x) \leq \exp(-\frac{x^2}{2})$, we have

$$P_{e|H} \leq \exp\left(-\frac{E_s}{4M_T N_0}||\mathbf{H}\mathbf{E}||_F^2\right)$$

Denoting $\mathbf{h}^*$ as the i^{th} row of $\mathbf{H}$, we have

$$
\begin{aligned}
||\mathbf{HE}||_F^2 &= \mathrm{trace}(\mathbf{HEE}^*\mathbf{H}^*) \\
&= \mathrm{trace}\left(\begin{bmatrix} \mathbf{h}_1^* \\ \vdots \\ \mathbf{h}_{M_R}^* \end{bmatrix} \mathbf{EE}^* \begin{bmatrix} \mathbf{h}_1 & \cdots & \mathbf{h}_{M_R} \end{bmatrix}\right) \\
&= \mathrm{trace}\left(\begin{bmatrix} \mathbf{h}_1^* \\ \vdots \\ \mathbf{h}_{M_R}^* \end{bmatrix} \begin{bmatrix} \mathbf{EE}^*\mathbf{h}_1 & \cdots & \mathbf{EE}^*\mathbf{h}_{M_R} \end{bmatrix}\right) \\
&= \mathbf{h}_1^*\mathbf{EE}^*\mathbf{h}_1 + \cdots + \mathbf{h}_{M_R}^*\mathbf{EE}^*\mathbf{h}_{M_R} \\
&= \sum_{i=1}^{M_R} \mathbf{h}_i^*\mathbf{EE}^*\mathbf{h}_i
\end{aligned}
$$

- Applying eigen-decomposition to $\mathbf{EE}^*$ as $\mathbf{EE}^* = \mathbf{P}^*\mathbf{\Lambda P}$

$$
\begin{aligned}
||\mathbf{HE}||_F^2 &= \sum_{i=1}^{M_R} \mathbf{h}_i^*\mathbf{P}^*\mathbf{\Lambda P}\mathbf{h}_i \\
&= \sum_{i=1}^{M_R} \tilde{\mathbf{h}}_i^*\mathbf{\Lambda}\tilde{\mathbf{h}}_i
\end{aligned}
$$

where $\mathbf{\Lambda}$: $\mathrm{diag}\{\lambda_1, \cdots, \lambda_{M_T}\}$, $\mathbf{P}$: unitary, $\tilde{\mathbf{h}}_i = \mathbf{Ph}_i = [\tilde{h}_{i1} \cdots \tilde{h}_{1M_T}]^T$

$$
\begin{aligned}
\tilde{\mathbf{h}}_i^*\mathbf{\Lambda}\tilde{\mathbf{h}}_i &= \begin{bmatrix} \tilde{h}_{i1}^* & \cdots & \tilde{h}_{iM_T}^* \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{M_T} \end{bmatrix} \begin{bmatrix} \tilde{h}_{i1} \\ \vdots \\ \tilde{h}_{iM_T} \end{bmatrix} \\
&= \sum_{j=1}^{M_T} \lambda_j |\tilde{h}_{ij}|^2
\end{aligned}
$$

$\therefore ||\mathbf{HE}||_F^2 = \sum_{i=1}^{M_R}\sum_{j=1}^{M_T} \lambda_j |\tilde{h}_{ij}|^2$

Thus, $P_{e|H}$ becomes

$$\begin{aligned} P_{e|H} &\leq \exp\left(-\frac{E_s}{4M_T N_0}\sum_{i=1}^{M_R}\sum_{j=1}^{M_T}\lambda_j|\tilde{h}_{ij}|^2\right) \\ &= \prod_i\prod_j \exp(-\rho\lambda_j|\tilde{h}_{ij}|^2) \end{aligned}$$

where $\rho = \dfrac{E_s}{4M_T N_0}$

Since $\mathbf{P}$ is unitary , $|\tilde{h}_{ij}|$ is also Rayleigh distributed with $\mathrm{E}|\tilde{h}_{ij}|^2 = 1$. Averaging $P_{e|H}$ over a Rayleigh distribution , we have

$$\begin{aligned} P_e &= \int P_{e|H} f_H(\mathbf{x}) d\mathbf{x} \\ &\leq \prod_i\prod_j \int_0^\infty 2x_{ij}\exp(-(1+\rho\lambda_j)x_{ij}^2)dx_{ij} \\ &= \prod_i\prod_j -\frac{1}{(1+\rho\lambda_j)}\exp(-(1+\rho\lambda_j)x_{ij}^2)\Big|_0^\infty \\ &= \prod_i^{M_R}\prod_j^{M_T}\frac{1}{(1+\rho\lambda_j)} \\ &= \prod_j^{M_T}\left(\frac{1}{1+\rho\lambda_j}\right)^{M_R} \\ &= \prod_j^{M_T}(1+\rho\lambda_j)^{-M_R} \end{aligned}$$

- Let the rank of $\mathbf{EE}^*$ be $r = \mathrm{rank}(\mathbf{EE}^*) = \mathrm{rank}(\mathbf{E})$. Then $M_T - r$ eigenvalues are zero such that

$$P_e \leq \prod_j^{r}(1+\rho\lambda_j)^{-M_R}$$

- For high SNR, the PEP becomes

$$\begin{aligned} P_e &\leq \prod_{j}^{r}(\rho\lambda_j)^{-M_R} \\ &= \rho^{-rM_R}\prod_{j=1}^{r}\lambda^{-M_R} \\ &= (\frac{E_s}{4M_TN_0})^{-rM_R}\prod_{j=1}^{r}\lambda^{-M_R} \end{aligned}$$

- Since $r \leq M_T$, the maximum diversity order equals $M_T \times M_R$ (full diversity).

3.2.3 STC criteria

(1) rank criteria
To achieve full diversity ($r = M_T$), $\mathbf{E}$ should be full rank for any pair of $\mathbf{s}$ & $\hat{\mathbf{s}}$.
(2) determinant criteria
With full rank $\mathbf{E}$, the coding gain $\prod_{j=1}^{r}\lambda_j$ becomes $\det(\mathbf{E}\mathbf{E}^*)$, $\rightarrow$The minimum of $\det(\mathbf{E}\mathbf{E}^*)$ over all possible pairs of $\mathbf{s}$ & $\hat{\mathbf{s}}$ should be maximized.

ex) Alamouti scheme (T=2)

$$\mathbf{C}(s) = \begin{bmatrix} s_1 & -s_2^* \\ -s_2 & s_1^* \end{bmatrix} \qquad \mathbf{C}(\hat{s}) = \begin{bmatrix} \hat{s}_1 & -\hat{s}_2^* \\ \hat{s}_2 & \hat{s}_1^* \end{bmatrix}$$

$$\mathbf{E} = \hat{\mathbf{s}} - \mathbf{s} = \begin{bmatrix} e_1 & -e_2^* \\ e_2 & e_1^* \end{bmatrix}$$

where $e_i = \hat{s}_i - s_i$

$$\text{For } (e_1 = 0, e_2 \neq 0), \mathbf{E} = \begin{bmatrix} 0 & -e_2^* \\ e_2 & 0 \end{bmatrix}$$

$$\text{For } (e_1 \neq 0, e_2 = 0), \mathbf{E} = \begin{bmatrix} e_1 & 0 \\ 0 & e_1^* \end{bmatrix}$$

$\rightarrow$full rank(r=2)

3.3 Spatial multiplexing (SM)

- Each transmit antenna transmits independent information streams to maximize the throughtput.

- Higher code rate ($r_s = M_T$) is achieved at the expense of the reduced diversity gain.

- Normally SM is employed where higher throughput should be obtained , while STC is applied where a better link performance is desirable.

- receiver structure
i) Maximum Likelihood (ML) detection
ii) Serial Interference cancelation (SIC).
iii) Linear equalization : Zero forcing or MMSE

3.3.1 ML detection (MLD)

- Consider $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$. Then, a ML solution is obtained as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2$$

- MLD compares all possible input vectors $\mathbf{x}$ and selects the one which minimizes the Euclidean distance.

- Total number of search candidates is N^{M_T} where N is the modulation level. Thus, Computational complexity increases exponentially with M_T and the number of bits per symbol.

- Near optimal detectors with reduced complexity have been proposed including sphere decoder & QRD-M.

§ Performance analysis of MLD

- Let $\mathbf{x}$ & $\hat{\mathbf{x}}$ be two possible Tx input vectors of length M_T. Denoting $\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x}$, we have

$$\begin{aligned} P(\mathbf{x} \rightarrow \hat{\mathbf{x}}|\mathbf{H}) &\leq \exp(-\frac{E_s}{4N_0}||\mathbf{He}||^2) \\ &= \exp(-\frac{E_s}{4N_0}\mathbf{z}^*\mathbf{z}) \end{aligned}$$

where $\mathbf{z} = \mathbf{He}$

- Since the elements of $\mathbf{H}$ are i.i.d complex Gaussian, $\mathbf{z}$ has a multi-variate complex Gaussian distribution.

- Averaging over all channel realization yields

$$P(\mathbf{x} \rightarrow \hat{\mathbf{x}}) \leq \int \exp(-\frac{E_s}{4N_0}\mathbf{z}^*\mathbf{z})\frac{1}{|\pi\mathbf{R_Z}|}\exp(-\mathbf{z}^*\mathbf{R_Z}^{-1}\mathbf{z})\ d\mathbf{z}$$

where $R_{\mathbf{Z}}$ is a convariance matrix of $\mathbf{z}$ as $R_{\mathbf{Z}} = \mathbb{E}(\mathbf{zz}^*) = \mathbb{E}(\mathbf{Hee}^*\mathbf{H}^*)$

- Then PEP is bounded by

$$P(\mathbf{x} \rightarrow \hat{\mathbf{x}}) \leq \frac{1}{|\pi\mathbf{R_Z}|}\int \exp(-\mathbf{z}^*\mathbf{R_{Z'}}^{-1}\mathbf{z})\ d\mathbf{z}$$

where $R_{\mathbf{Z}'} = \left(\frac{E_s}{4N_0}\mathbf{I} + R_{\mathbf{Z}}^{-1}\right)^{-1}$
Finally we have

$$\begin{aligned} P(\mathbf{x} \rightarrow \hat{\mathbf{x}}) &\leq \frac{|\pi\mathbf{R_{Z'}}|}{|\pi\mathbf{R_Z}|}\frac{1}{|\pi\mathbf{R_{Z'}}|}\int \exp(-\mathbf{z}^*\mathbf{R_{Z'}}^{-1}\mathbf{z})\ d\mathbf{z} \\ &= \frac{|\pi\mathbf{R_{Z'}}|}{|\pi\mathbf{R_Z}|} \\ &= \frac{1}{|\mathbf{R_Z}\mathbf{R_{Z'}}^{-1}|} \\ &= \frac{1}{|\frac{E_s}{4N_0}\mathbf{R_Z} + \mathbf{I}|} \end{aligned}$$

- By applying eigen decomposition to $\mathbf{ee}^*$, we have $\mathbf{ee}^* = \mathbf{U}^*\mathbf{\Lambda}\mathbf{U}$

where $\mathbf{U}$ is a unitary matrix and $\mathbf{\Lambda}$ denotes a diagonal matrix with eigenvalues.

- Since $\mathbf{ee}^*$ is a a rank one matrix , $\mathbf{\Lambda}$ has only one eigenvalue λ. Thus $\mathbf{R_Z}$ becomes

$$\begin{aligned} R_{\mathbf{Z}} &= \mathbb{E}(\mathbf{HU}^*\mathbf{\Lambda}\mathbf{UH}^*) \\ &= \mathbb{E}\left(\mathbf{H}'\begin{bmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & 0 \end{bmatrix}\mathbf{H}'^*\right) \end{aligned}$$

where $\mathbf{H}' = \mathbf{HU}^*$

- Denoting $\mathbf{h}'$ as the first column of $\mathbf{H}'$, we have $\mathbb{E}(\mathbf{h}'\mathbf{h}'^*) = \mathbf{I}$ since $\mathbf{U}$ is unitary.
$\Rightarrow \mathbf{R_Z} = \mathbb{E}(\lambda\mathbf{h}'\mathbf{h}'^*) = \lambda\mathbf{I}_{M_R}$

- Finaly, the PEP becomes

$$P(\mathbf{x} \rightarrow \ \hat{\mathbf{x}}) \ \leq \ \left(\lambda\frac{E_s}{4N_0} + 1\right)^{-M_R}$$

$\therefore$ Diversity order of MLD $= M_R$

3.3.2 Zero forcing (ZF) Receiver

- The ZF receiver $\mathbf{W}$ is obtained as an $M_T \times M_R$ pseudo-inverse matrix of $\mathbf{H}$

$$\mathbf{W} = \mathbf{H}^+ = (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*$$

- In order for the pseudo-inverse to exist, M_T should be loss than or equal to M_R ($M_T \leq M_R$), as opposed to MLD which does not have any constraint on M_T & M_R

- The estimate of $\mathbf{x}$ is computed by

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{W}\mathbf{y} = (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{H}\mathbf{x} + (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{n} \\ &= \mathbf{x} + \tilde{\mathbf{n}}\end{aligned}$$

where $\tilde{\mathbf{n}} = \mathbf{W}\mathbf{n}$
$\rightarrow$ ZF receivers suffer from noise enhancement

§ Performance analysis of ZF receiver

- Let us define the estimation error $\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x} = \tilde{\mathbf{n}}$

Assume that $\mathbf{n}$ is i.i.d. with $\mathbb{E}[\mathbf{n}\mathbf{n}^*] = \sigma_n^2\mathbf{I}$, the covariance matrix of $\mathbf{e}$ is given by

$$\begin{aligned}\mathbf{R_e} &= E[\mathbf{e}\mathbf{e}^*] \\ &= E\left[(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{n}\mathbf{n}^*\mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}\right] \\ &= \sigma_n^2(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1} = \sigma_n^2(\mathbf{H}^*\mathbf{H})^{-1}\end{aligned}$$

- The mean squared error (MSE) for the i^{th} symbol x_i $(1 \le i \le M_T)$ is equal to $\sigma_i^2 = \mathbf{R}_{\mathbf{e},ii} = \sigma_n^2(\mathbf{H}^*\mathbf{H})_{ii}^{-1}$ where $\mathbf{A}_{ii}$ indicates the i^{th} diagonal element of $\mathbf{A}$ where $\mathbf{A}_{ii}$ indicates the i^{th} diagonal element of $\mathbf{A}$.

- $P_{e|H}$ for ZF receiver is bounded by $P_{e|H} \le \exp\left(-\frac{E_s}{4\sigma_i^2}\right) = \exp\left(-\frac{\alpha_0}{4}E_s\right)$ where $\alpha_0 = \frac{1}{\sigma_i^2}$.

- It can be shown that when $\mathbf{H}$ is Rayleigh distributed, α_0 has a Chi-square distribution with $2(M_R - M_T + 1)$ degrees of freedom.

- It follows that the PEP is bounded by

$$\begin{aligned}P_e &\le \left(\frac{E_s}{4N_0}\right)^{-(M_R - M_T + 1)} \\ \Rightarrow\ & D = M_R - M_T + 1\end{aligned}$$

$\therefore$ The diversity performance of $M_T \times M_R$ MIMO ZF Receiver is equal to the MRC with $M_R - M_T + 1$ Rx antennas. For $M_R = M_T$ MIMO ZF receiver have no diversity gain over SISO.

3.3.3 Minimum mean squared error (MMSE) receiver

- Defining the error vector as $\mathbf{e} = \mathbf{x} - \mathbf{W}\mathbf{y}$, the MSE is given by $\sigma_e^2 = \mathbb{E}\left(||\mathbf{e}||^2\right) = \mathbb{E}\left[\left(\mathbf{x} - \mathbf{W}\mathbf{y}\right)^* \left(\mathbf{x} - \mathbf{W}\mathbf{y}\right)\right]$

- Assume that $\mathbf{x}$ and $\mathbf{n}$ are uncorrelated and $\mathbf{x}$ is i.i.d $\mathbb{E}\left(\mathbf{x}\mathbf{x}^*\right) = E_s\mathbf{I}$, In order to minimize the MSE σ_n^2, the orthogonal principle should be satisfied.

$$\begin{aligned} \mathbb{E}\left[\mathbf{e}\mathbf{y}^*\right] &= \mathbb{E}\left[\left(\mathbf{x} - \mathbf{W}\mathbf{y}\right)\mathbf{y}^*\right] = 0 \\ &\Rightarrow \mathbb{E}\left[\mathbf{x}\mathbf{y}^*\right] = \mathbf{W}\mathbb{E}\left[\mathbf{y}\mathbf{y}^*\right] \end{aligned}$$

where $\mathbb{E}\left[\mathbf{x}\mathbf{y}^*\right] = \mathbb{E}\left[\mathbf{x}\left(\mathbf{H}\mathbf{x} + \mathbf{n}\right)^*\right] = E_s\mathbf{H}^*$

$$\begin{aligned} \mathbb{E}\left[\mathbf{y}\mathbf{y}^*\right] &= E\left[\left(\mathbf{H}\mathbf{x} + \mathbf{n}\right)\left(\mathbf{H}\mathbf{x} + \mathbf{n}\right)^*\right] \\ &= \mathbb{E}\left[\mathbf{H}\mathbf{x}\mathbf{x}^*\mathbf{H}^*\right] + \mathbb{E}\left[\mathbf{n}\mathbf{n}^*\right] = E_s\mathbf{H}\mathbf{H}^* + \sigma_n^2\mathbf{I} \end{aligned}$$

$$\begin{aligned} \therefore\ \mathbf{W} &= E_s\mathbf{H}^*\left(E_s\mathbf{H}\mathbf{H}^* + \sigma_n^2\mathbf{I}\right)^{-1} \\ &= \mathbf{H}^*\left(\mathbf{H}\mathbf{H}^* + \alpha\mathbf{I}\right)^{-1} \qquad \text{where} \ \ \alpha = \tfrac{\sigma_n^2}{E_s} \end{aligned}$$

- Using the matrix inversion lemma,

$$\left(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{D}\mathbf{A}^{-1},$$

We have

$$\begin{aligned} \mathbf{W} &= \mathbf{H}^*\left(\mathbf{H}\mathbf{H}^* + \alpha\mathbf{I}\right)^{-1} \\ &= \mathbf{H}^*\left(\frac{1}{\alpha}\mathbf{I} - \frac{1}{\alpha}\mathbf{H}\left(\mathbf{I} + \mathbf{H}^*\mathbf{H}\frac{1}{\alpha}\right)^{-1}\mathbf{H}^*\frac{1}{\alpha}\right) \\ &= \frac{1}{\alpha}\mathbf{H}^*\left(\mathbf{I} - \mathbf{H}\left(\alpha\mathbf{I} + \mathbf{H}^*\mathbf{H}\right)^{-1}\mathbf{H}^*\right) \\ &= \frac{1}{\alpha}\left(\mathbf{H}^* - \mathbf{H}^*\mathbf{H}\left(\alpha\mathbf{I} + \mathbf{H}^*\mathbf{H}\right)^{-1}\mathbf{H}^*\right) \\ &= \frac{1}{\alpha}\left(\mathbf{H}^* - \left(\alpha\mathbf{I} + \mathbf{H}^*\mathbf{H} - \alpha\mathbf{I}\right)\left(\alpha\mathbf{I} + \mathbf{H}^*\mathbf{H}\right)^{-1}\mathbf{H}^*\right) \\ &= \frac{1}{\alpha}\left(\mathbf{H}^* - \mathbf{I} + \alpha\left(\alpha\mathbf{I} + \mathbf{H}^*\mathbf{H}\right)^{-1}\mathbf{H}^*\right) \\ &= \frac{1}{\alpha}\left(\mathbf{H}^* - \left(\mathbf{I} - \alpha\left(\alpha\mathbf{I} + \mathbf{H}^*\mathbf{H}\right)^{-1}\mathbf{H}^*\right)\right) \\ &= \frac{1}{\alpha}\left(\mathbf{I} - \mathbf{I} + \alpha\left(\alpha\mathbf{I} + \mathbf{H}^*\mathbf{H}\right)^{-1}\right)\mathbf{H}^* \\ &= \left(\mathbf{H}^*\mathbf{H} + \alpha\mathbf{I}\right)^{-1}\mathbf{H}^*. \end{aligned}$$

$\Rightarrow$ For high SNR ($\alpha \to 0$), the ZF solution becomes equivalent to the MMSE solution.

- The error covariance matrix is given as

$$\begin{aligned}
\mathbf{R_e} &= \mathbb{E}\left[\mathbf{e}\mathbf{e}^*\right] = \mathbb{E}\left[\mathbf{e}\left(\mathbf{x} - \mathbf{W}\mathbf{y}\right)^*\right] \\
&= \mathbb{E}\left[\mathbf{e}\mathbf{x}^*\right] - \mathbb{E}\left[\mathbf{e}\mathbf{y}^*\right]\mathbf{W}^* \\
&= \mathbb{E}\left[\left(\mathbf{x} - \mathbf{W}\mathbf{y}\right)\mathbf{x}^*\right] \\
&= \mathbb{E}\left[\mathbf{x}\mathbf{x}^*\right] - \mathbf{W}\mathbb{E}\left[\mathbf{y}\mathbf{x}^*\right] \\
&= E_s\mathbf{I} - \left(\mathbf{H}^*\mathbf{H} + \alpha\mathbf{I}\right)^{-1}\mathbf{H}^*\mathbf{H}E_s \\
&= E_s\left(\mathbf{I} - \left(\mathbf{H}^*\mathbf{H} + \alpha\mathbf{I}\right)^{-1}\mathbf{H}^*\mathbf{H}\right) \\
&= E_s\left(\mathbf{I} - \left(\mathbf{H}^*\mathbf{H} + \alpha\mathbf{I}\right)^{-1}\left(\mathbf{H}^*\mathbf{H} + \alpha\mathbf{I} - \alpha\mathbf{I}\right)\right) \\
&= E_s\left(\mathbf{I} - \mathbf{I} + \alpha\left(\mathbf{H}^*\mathbf{H} + \alpha\mathbf{I}\right)^{-1}\right) \\
&= \sigma_n^2\left(\mathbf{H}^*\mathbf{H} + \alpha\mathbf{I}\right)^{-1}
\end{aligned}$$

- The i^{th} diagonal element corresponds to the MSE of the i^{th} symbol estimate $\hat{x}_i$ ($1 \le i \le M_T$).

3.3.4 Bell-lab layered space-time (BLAST)

- Ordered Successive Interference Cancellation (OSIC)

As opposed to linear receiver where all M_T estimates of $\mathbf{x}$ are determined at once, BLAST selects only one estimate whose accuracy is the highest.

- Assuming a correct decision, interference is cancelled using the decision, then the detection process is repeated. Thus, the orders in which the decisions are made become important.

cf) Decision feedback equalizer (DFE)

§ Nulling and Cancelling

- MIMO channel output $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ can be written as

$$\begin{aligned}
\mathbf{y} &= \begin{bmatrix} \mathbf{h}_1 & \cdots & \mathbf{h}_{M_T} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_3 \end{bmatrix} + \mathbf{n} \\
&= \mathbf{h}_1 x_1 + \cdots + \mathbf{h}_{M_T} x_{M_T} + \mathbf{n}
\end{aligned}$$

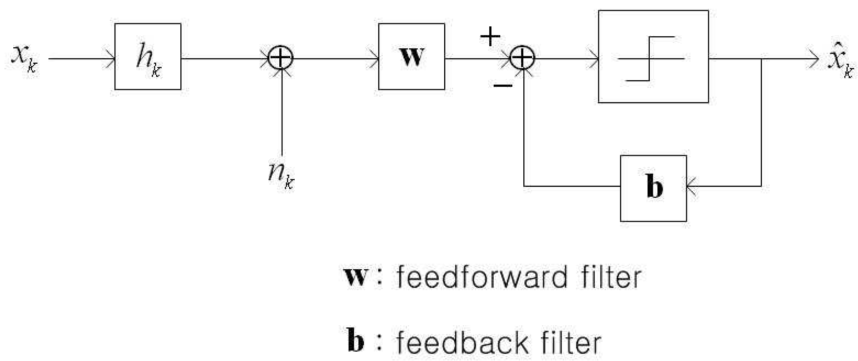

Figure 3.2: Decision Feedback Equalization

where $\mathbf{h}_i$: i^{th} column of $\mathbf{H}$.

- When detecting x_1, $\mathbf{h}_2 x_2 + \cdots + \mathbf{h}_{M_T} x_{M_T}$ is considered as interference from other Tx antennas.

- First by an equalization filtering, $\hat{x}_p$ is determined whose corresponding MSE is the smallest. (nulling)

- Then, assuming that $\hat{x}_p$ is correct, the interference term corresponding to $\hat{x}_p$ can be cancelled (cancellation) as

$$\mathbf{y}' = \mathbf{y} - \mathbf{n}_p x_p = \mathbf{h}_1 x_1 + \cdots + \mathbf{h}_{p-1} x_{p-1} + \mathbf{h}_{p+1} x_{p+1} + \mathbf{h}_{M_T} x_{M_T} + \mathbf{n}$$

- With the reduced number of the interference terms, the nulling & cancellation process is repeated.

§ BLAST Algorithm

1) Compute $\mathbf{R_e} = \sigma_n^2 \left(\mathbf{H}^* \mathbf{H} + \alpha \mathbf{I}\right)^{-1}$. Then $\mathbf{W}$ is obtained from

$$\mathbf{W} = \left(\mathbf{H}^* \mathbf{H} + \alpha \mathbf{I}\right)^{-1} \mathbf{H}^* = \frac{1}{\sigma_n^2} R_{\mathbf{e}} \mathbf{H}^*$$

2) Determine p as

$$p = \arg\min_i \mathbf{R}_{\mathbf{e},ii}$$

where $\mathbf{w}_p$ is the p th row of $\mathbf{W}$.

3) Obtain a hard decision of $\hat{x}_p$ as

$$x_p = Q(\mathbf{w}_p\mathbf{y})$$

where $\mathbf{w}_p$: p^{th} row of $\mathbf{W}$

4) perform interference cancellation as

$$\mathbf{y} \leftarrow \mathbf{y} - \mathbf{h}_p x_p$$

- In each layer, decisions made in the previous layer are assumed to be correct, and are used to cancel interference. When those decisions are incorrect, subsequent interference cancellation process becomes very unreliable.
$\rightarrow$ error propagation

- Thus, the accuracy of decisions made in early layers determines the overall performance. $\Rightarrow$ The ordering is quite important.

- The diversity performance of BLAST is between ZF receiver ($M_R - M_T + 1$) and MLD (M_R).

3.3.5 SM architecture with channel coding

(1) Horizontal encoding

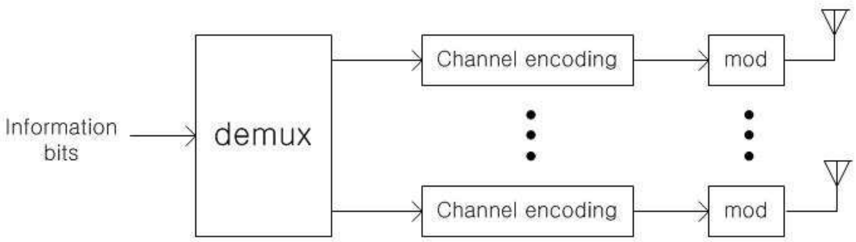

Figure 3.3: Horizontal encoding structure

- Channel coding is performed individually at each Tx antenna.
- Since any given symbol is transmitted through only one Tx antenna. Capacity may be lower. The weakest links limits the performance.

- The detection performance may be improved when cancelation using decoded decisions(CDD) is employed.

(2) Vertical encoding

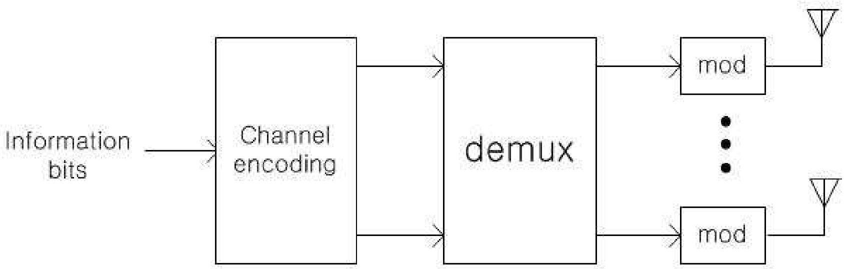

Figure 3.4: Vertical encoding structure

- A channel encoding is applied over all Tx antennas. Since the information symbols are encoded by a single channel encoder. The performance may be better than the horizontal encoding.

3.3.6 Closed loop SM systems

- Tx structure

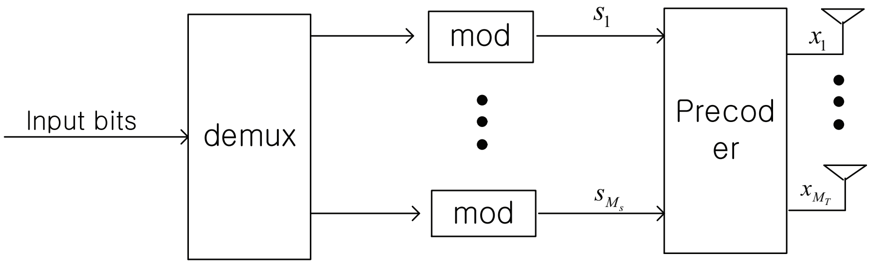

Figure 3.5: Tx structure

- For MIMO channels $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, the transmitted vector $\mathbf{x}$ of closed loop systems is precoded as $\mathbf{x} = \mathbf{P}\mathbf{s}$

where $\mathbf{s} = [s_1, s_2, \cdots, s_{M_s}]^T$

$\mathbf{P}$: $M_T \times M_s$ precoding matrix

M_s : number of streams

- M_s cannot be greater than the minimum of M_T or M_R ($M_s \leq \min(M_T, M_s)$)

$M_s = 1$: beamforming

$M_s > 1$: SM

- The precoded symbol is represented by

$$\mathbf{x} = \mathbf{P}\mathbf{s} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_{M_s}] \begin{bmatrix} s_1 \\ \vdots \\ s_{M_s} \end{bmatrix} = \sum_{i=1}^{M_s} \mathbf{p}_i s_i$$

where $\mathbf{p}_i$: i^{th} column of $\mathbf{P}$

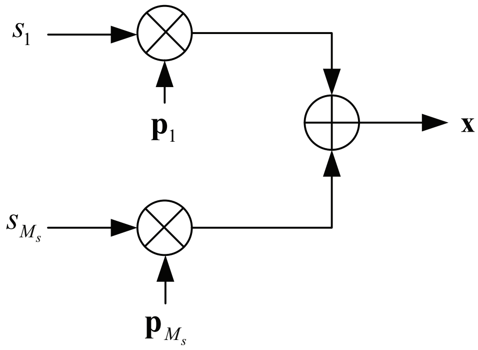

Figure 3.6: Precoder

- Power constraint

$$
\begin{aligned}
\mathbb{E}(\| \mathbf{x} \|^2) &= \mathbb{E}(\mathbf{x}^*\mathbf{x}) = \mathbb{E}(\mathrm{tr}(\mathbf{x}^*\mathbf{x})) = \mathbb{E}(\mathrm{tr}(\mathbf{x}\mathbf{x}^*)) \\
&= \mathbb{E}(\mathrm{tr}(\mathbf{P}\mathbf{s}\mathbf{s}^*\mathbf{P}^*)) = \mathrm{tr}(\mathbb{E}(\mathbf{P}\mathbf{s}\mathbf{s}^*\mathbf{P}^*)) = \mathrm{tr}(\mathbf{P}\mathbf{P}^*) \\
&= \| \mathbf{P} \|_F^2 = \sum_{i=1}^{M_s} \| \mathbf{p}_i \|^2 \le P_T
\end{aligned}
$$

- The channel output is written as

$$\mathbf{y} = \mathbf{H}\mathbf{P}\mathbf{s} + \mathbf{n} = \mathbf{H}_e\mathbf{s} + \mathbf{n}$$

where $\mathbf{H}_e = \mathbf{H}\mathbf{P}$ indicates the effective channel of size $M_R \times M_s$

- Rx structure

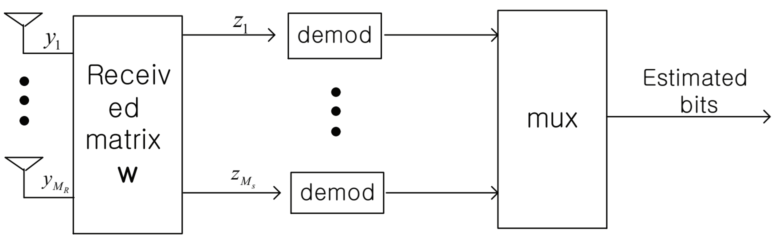

Figure 3.7: Rx structure

The channel output $\mathbf{y}$ is processed by the $M_R \times M_s$ received matrix $\mathbf{W}$ as $\mathbf{z} = \mathbf{W}^*\mathbf{y}$ where $\mathbf{z} = [z_1, z_2, \cdots, z_{M_s}]^T$

§ Beamforming(M_s=1)

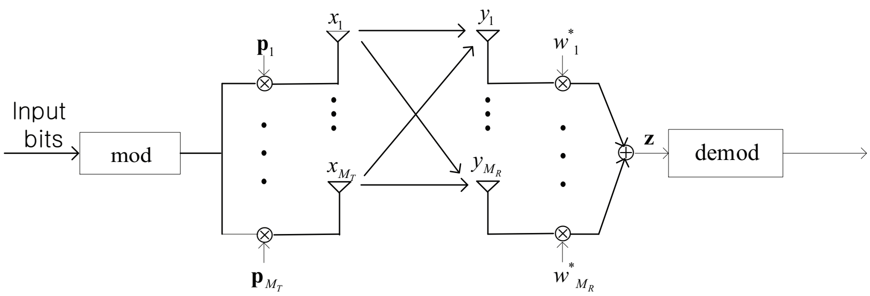

Figure 3.8: Beamforming structure

- The channel output of the beamforming scheme is given as

$$\mathbf{y} = \mathbf{Hp}s + \mathbf{n}$$

where $\mathbf{p} = [p_1, p_2, \cdots, p_{M_T}]^T$ represents the beamforming vector with $\| \mathbf{P} \|$=1

- By applying the receive filter $\mathbf{w} = [w_1, w_2, \cdots, w_{M_R}]^T$, the received filter output becomes

$$\mathbf{z} = \mathbf{w}^*\mathbf{y} = \mathbf{wHp}s + \mathbf{w}^*\mathbf{n}$$

- For given $\mathbf{P}$ from the MRC solution, the receive vector $\mathbf{w}$ is set to $\mathbf{w} = \mathbf{HP}$(matched filter).
Then we have

$$\mathbf{z} = \mathbf{p}^*\mathbf{H}^*\mathbf{Hp}s + \mathbf{p}^*\mathbf{H}^*\mathbf{n} = \| \mathbf{w}^* \| s + \mathbf{w}^*\mathbf{n}$$

- Assuming that $\mathbf{n}$ is a White Gaussian noise vector with variance N_o ($\mathbb{E}(\mathbf{nn}^*) = N_o\mathbf{I}$), the output SNR is given by

$$\text{SNR} = \frac{\mathbb{E}\, |\| \mathbf{w} \|^2 s\,|^2}{\mathbb{E}(\mathbf{w}^*\mathbf{nn}^*\mathbf{w})} = \frac{\| \mathbf{w} \|^4 \mathbb{E}(|\, s\,|^2)}{N_o \| \mathbf{w} \|^2} = \| \mathbf{w} \|^2 \frac{E_s}{N_o}$$

- To find $\mathbf{p}$ which maximize $\mathbf{p}^*\mathbf{H}^*\mathbf{H}\mathbf{p}$ with the constraint $\| \mathbf{p} \| = 1$, we set a Lagrange multiplier as

$$C = \mathbf{p}^*\mathbf{A}\mathbf{p} + \lambda(\mathbf{p}^*\mathbf{p} - 1) \qquad \text{where } \mathbf{A} = \mathbf{H}^*\mathbf{H}$$

Differentiating C with respect to $\mathbf{p}$ yields

$$\frac{\partial C}{\partial \mathbf{p}} = 2\mathbf{A}\mathbf{p} + \lambda 2\mathbf{p} = 0 \qquad \rightarrow \mathbf{A}\mathbf{p} = \lambda' \mathbf{p}$$

Thus, the solution is the eigenvector of $\mathbf{H}^*\mathbf{H}$ plugging this, it follows $\mathbf{p}^*\mathbf{H}^*\mathbf{H}\mathbf{p} = \lambda' \| \mathbf{p} \|^2 = \lambda'$

$\therefore$ The output SNR is maximized with $\mathbf{p}$ as the eigenvector of $\mathbf{H}^*\mathbf{H}$ corresponding to maximum eigenvalue λ_{max}, and is equal to $\lambda_{max}\frac{E_s}{N_o}$

- Alternatively, denoting the SVD of $\mathbf{H}$ as $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$, $\mathbf{H}\mathbf{H}^*$ becomes $\mathbf{V}\Sigma^*\mathbf{U}^*\mathbf{U}\Sigma\mathbf{V}^* = \mathbf{V}\Sigma^2\mathbf{V}^*$. Thus, the solution $\mathbf{p}$ & $\mathbf{w}$ equal the right singular vector and the left sigular vector of $\mathbf{H}$ corresponding to the largest singular value of $\mathbf{H}$, respectively.

- Special case of $M_R = 1$ (MISO)
The channel output is written as $y = \mathbf{h}\mathbf{p}s + n$ where $\mathbf{h} = [h_1, h_2, \cdots, h_{M_T}]$.

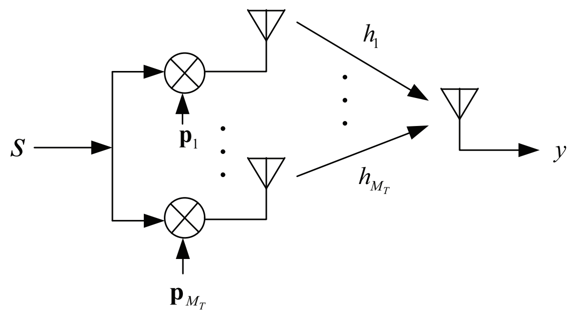

Figure 3.9: Tx MRC

Then the optimum weighting vector $\mathbf{p}$ equals $\mathbf{p} = \frac{\mathbf{h}^*}{\|\mathbf{h}\|}$ ($\mathbf{V}^* = \frac{\mathbf{h}}{\|\mathbf{h}\|} \quad \rightarrow \quad \mathbf{x} = \mathbf{V}\mathbf{s} = \frac{\mathbf{h}^*}{\|\mathbf{h}\|}\mathbf{s}$)

§ SM with multiple streams($M_s > 1$)

- Consider a SVD of $\mathbf{H}$ as $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$ where $\mathbf{U} : M_R \times M_R, \Sigma : M_R \times M_T, \mathbf{V} : M_T \times M_T$
Let us define $\mathbf{U}_{M_s}$ & $\mathbf{V}_{M_s}$ as matrices constructed from the first M_s column of $\mathbf{U}$ & $\mathbf{V}$, respectively.

Setting the precoder matrix $\mathbf{P}$ as $\mathbf{P} = \mathbf{V}_{M_s}$, the effective channel becomes

$$\mathbf{H}_e = \mathbf{HP} = \mathbf{U}\Sigma\mathbf{V}^*_{M_s} = \mathbf{U}\Sigma \begin{bmatrix} \mathbf{I}_{M_s} \\ 0 \end{bmatrix}$$

- Employing $\mathbf{U}_{M_s}$ as the receive matrix, the receive filter output is

$$\begin{aligned} \mathbf{z} = \mathbf{U}^*_{M_s}\mathbf{y} &= \mathbf{U}^*_{M_s}(\mathbf{H}_e\mathbf{s} + \mathbf{n}) \\ &= \mathbf{U}^*_{M_s}\mathbf{U}\Sigma \begin{bmatrix} \mathbf{I}_{M_s} \\ 0 \end{bmatrix} \mathbf{s} + \mathbf{U}^*_{M_s}\mathbf{n} \\ &= \begin{bmatrix} \mathbf{I}_{M_s} & 0 \end{bmatrix} \Sigma \begin{bmatrix} \mathbf{I}_{M_s} \\ 0 \end{bmatrix} \mathbf{s} + \mathbf{U}^*_{M_s}\mathbf{n} \\ &= \Sigma_{M_s}\mathbf{s} + \tilde{\mathbf{n}} \end{aligned}$$

$$\therefore z_i = \sigma_i s_i + \tilde{n}_i \quad \text{for} \quad i = 1, 2, \cdots, M_s$$

where $\Sigma_{M_s} = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_{M_s}) : M_s \times M_s$ upper-left submatrix of Σ
$\tilde{\mathbf{n}} = \mathbf{U}^*_{M_s}\mathbf{n}$

$\mathbf{z} = \sum_{M_s} \mathbf{s} + \tilde{\mathbf{n}} \qquad 1 \leq M_s \leq \min(M_T, M_R)$
where $\sum_{M_s} = \mathrm{diag}\{\sigma_1, \cdots, \sigma_{M_s}\}$; $M_s \times \ M_s$ diagonal matrix

$$\begin{bmatrix} z_1 \\ \vdots \\ z_{M_s} \end{bmatrix} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{M_s} \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_{M_s} \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \vdots \\ \tilde{n}_{M_s} \end{bmatrix}$$

$\therefore z_i = \sigma_i s_i + \tilde{n}_i$ for $i = 1, \cdots, M_s$ $(\sigma_1 \geq \cdots \geq \sigma_{M_s})$

- As a result, Ms parallel eigenmode channels are created with the channel gain $\{\sigma_i\}$.

- As the eigenmodes with small channel gains dominate the overall performance, the SM performance is improved by discarding $\min(M_T, M_R) - M_s$

eigenmodes with small channel gains.

- When comparing with the full SM case of $M_s = \min(M_T, M_R)$ with the same spectral efficiency, normally systems with small M_s exhibit better performance due to higher diversity order.

- The closed loop capacity is obtained with $M_s = \min(M_T, M_R)$ by applying the water-filling power allocation for all eigenmodes.

3.3.7 Diversity analysis of MIMO SVD systems

- It can be shown that the diversity order of coded SVD systems with code rate R_c is given by

$$D_{SVD} = (M_T - \lceil R_c M_s \rceil + 1)(M_R - \lceil R_c M_s \rceil + 1).$$

where $\lceil x \rceil$ equals the smallest integer more than x

- For uncoded systems ($R_c = 1$), D reduces to

$$D_{\text{uncoded}} = (M_T - M_s + 1)(M_R - M_s + 1).$$

The beamforming $M_s = 1$ achieves full diversity of $D = M_T \cdot M_R$ with the multiplexing gain of $r_s = M_s = 1$.
$\rightarrow$ diversity scheme

- For full SM case ($M_s = M_T = M_R$), $D = 1$ for uncoded system
$\rightarrow$ no diversity gain
cf) open loop SM systems with MLD

- With powerful code of $R_c \leq \frac{1}{M_s}$, the full diversity can be achieved for the full SM case.

3.4 Antenna selection

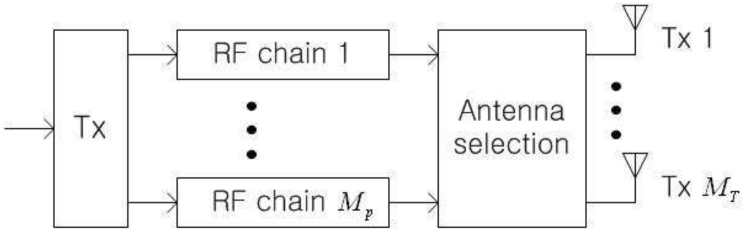

Figure 3.10: Transmit antenna selection

- Consider a system with M_{Ts} Tx RF chains and M_{Rs} Rx chains ($M_{Ts} \leq M_T$ & $M_{Rs} \leq M_R$).

- The number of all possible antenna subset selection is equal to $\binom{M_T}{M_{Ts}}$ or $\binom{M_R}{M_{Rs}}$.

- Various criteria for choosing the antenna subset to reduce the selection criteria are possible

3.5 MIMO-OFDM

- OFDM structure

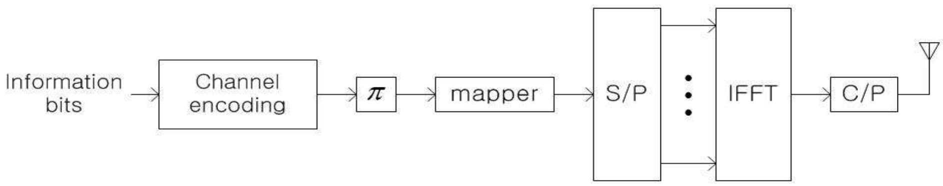

Figure 3.11: OFDM transmitter structure

- With proper cyclic prefix, IFFT/FFT operations convert frequency selective channels into M_c flat fading subchannels.

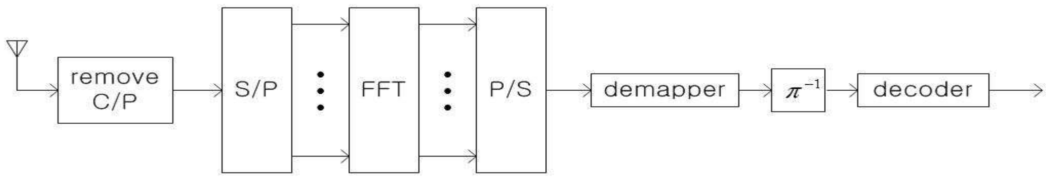

Figure 3.12: OFDM receiver structure

- MIMO techniques developed for flat fadings can be applied to each subcarrier.

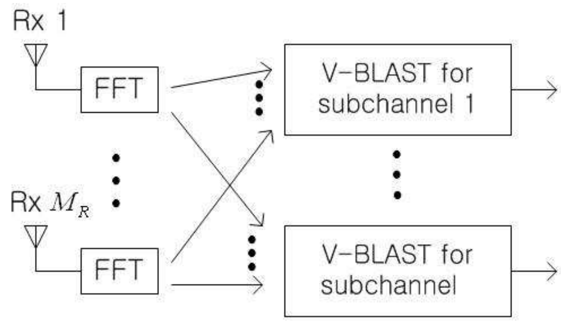

Figure 3.13: VBLAST processor

- With K tap equal power delaying profile, it can be shown that the maximum diversity order of MIMO-OFDM is given as $K \cdot M_T \cdot M_R$.

- For frequency selective channels, a diversity gain from MIMO is not as crucial as in flat fading cases.

Chapter 4

Limited Feedback MIMO Systems

"Grassmanian beamforming for MIMO wireless systems" D.love et'al, IEEE Info. Theory, 2003

- In ideal closed loop systems, it is assumed that perfect CSI is available at the Tx, which can be realized in TDD with channel reciprocity

- In FDD systems, a reliable feed back channel with no delay is required. However, the feedback overhead may become prohibitive especially for time varying channels.

- One solution which minimizes the required feedback overhead is to share a codebook of precoding matrices among Tx & Rx and to send back the index of the codeword to the Tx.

1) Grassmanian beamforming ($M_s = 1$)

. Suppose that both Tx & Rx know the codebook which contains N vectors of length M_T. Then Rx sends back $b = \log_2 N$bits to indicate the codeword index

. For beamforming systems $\mathbf{y} = \mathbf{H}\mathbf{p}s + \mathbf{n}$, the optimum solution $\mathbf{p}$ maximizes $||\mathbf{H}\mathbf{p}||^2$

. Denoting $\mathbf{\Omega} = [\mathbf{p}_1, \cdots, \mathbf{p}_N]$ as the codebook set, the Rx chooses the feedback index as

$$i^* = \arg\max_{\mathbf{p}_i \in \mathbf{\Omega}} ||\mathbf{H}\mathbf{p}||^2.$$

Then, the Rx employs the MRC filter $\mathbf{w} = \mathbf{H}\mathbf{p}_{i^*}$.

▪ The distance among vectors in $\mathbf{\Omega}$ is defined as

$$d(\mathbf{p}_1, \mathbf{p}_2) = \sqrt{1 - |\mathbf{p}_1^* \mathbf{p}_2|^2}$$

▪ The best codebook is designed such that the minimum distance among all possible vectors is maximized.

▪ For $N \leq M_T$, a codebook consisting of column vectors of unitary matrix results in a unit distance for all vectors, and thus becomes optimum.

▪ In contrast, for $N > M_T$, finding on optimal codebook in a closed form is complicated.

▪ Grassmanian line packing is the problem of optimally packing one-dimensional subspace, which finds the set of N lines in $\mathbb{C}^{M_T}$

▪ Using the results of Grassmanian line packing theory code book for arbitary N & M_T can be obtained.

(2) limited feedback SM($M_s > 1$)

- The codebook Ω for limited feedback systems with $M_s > 1$ consists of precoding matrices as $\Omega = [\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_N]$ where $\mathbf{F}_i \in \mathbb{C}^{M_T \times M_s}$

- The chordal distance which represents the subspace distance between two matrices $\mathbf{F}_1$ & $\mathbf{F}_2$ is defined as

$$\begin{aligned} d(\mathbf{F}_1, \mathbf{F}_2) &= \frac{1}{\sqrt{2}} \parallel \mathbf{F}_1\mathbf{F}_1^* - \mathbf{F}_2\mathbf{F}_2^* \parallel_F \\ &= \sqrt{M_s - \sum_{i=1}^{M_s} \lambda_i^2(\mathbf{F}_1^*\mathbf{F}_2)} \end{aligned}$$

where $\lambda_i(A)$: ith singular value of A

- Once a codebook is determined using Grassmanian packing theory, the precoding matrix index is selected based on the design criteria.

Chapter 5

Multiuser MIMO (MU-MIMO)

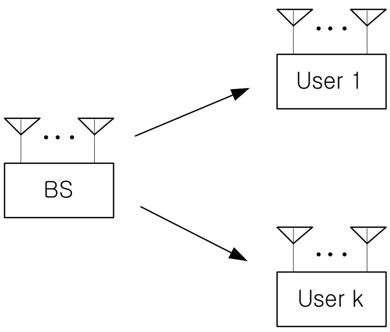

Figure 5.1: Multiuser MIMO

- In MU-MIMO, a basestation supports multiple users simultaneously, where each user receivers other users data as interference.

- It is assumed that users do not cooperate with each other. MU-MIMO systems with user cooperation reduce to SU-MIMO. Thus, the SU-MIMO system serves as a upper-bound of MU-MIMO.

- Two basic MU-MIMO structures(BC)
$\mathbf{y}_k = \mathbf{H}_k \sum_{j=1}^{k} \mathbf{x}_j + \mathbf{n}_k$

- MIMO Multiple Access channel(MAC)

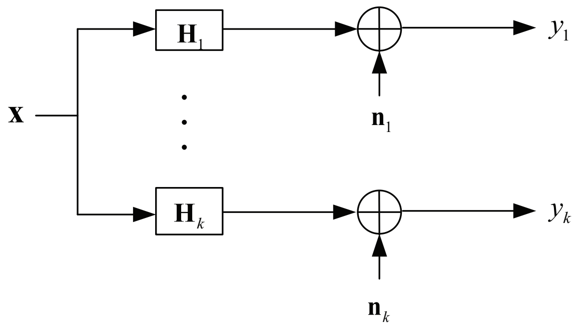

Figure 5.2: MIMO Broadcast Channel

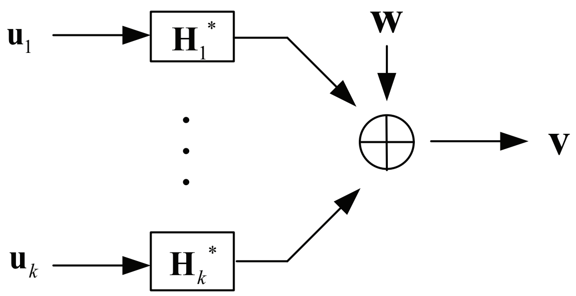

Figure 5.3: MIMO Multiple Access Channel

$\mathbf{V} = \sum_{j=1}^{k} \mathbf{H}_j^* \mathbf{u}_j + \mathbf{w}$

- It was shown that the sum-capacity of MIMO-BC is equal to the sum-capacity of the dual MIMO-MAC.

- Dirty Paper Coding (DPC) was first introduced in Costa's paper in 1983, which showned that if the interference is known at the Tx, the interference does not incur any loss in capacity by employing DPC.

- The capacity region of MIMO BC is achieved by DPC.

- System model
where $\mathbf{s}_k$: data vector of length $M_R \leq M_s$ for the kth user

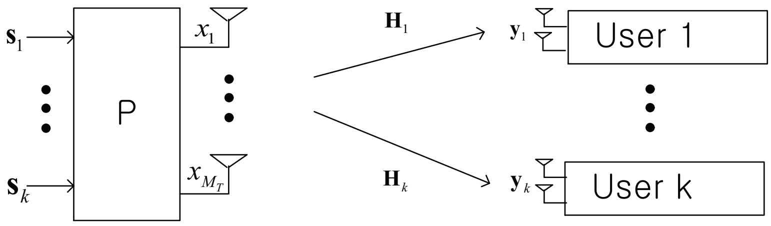

Figure 5.4: Multiuser System model

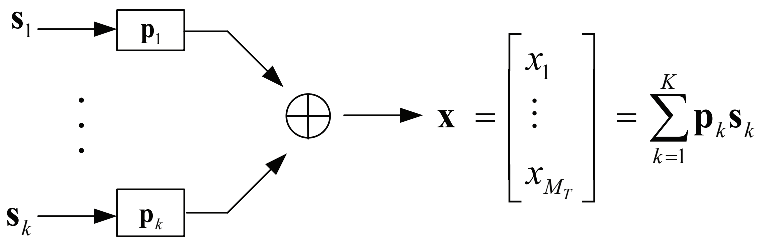

Figure 5.5: Multiuser System in transmitter

where $\mathbf{P}_k$: $M_T \times M_R$ precoding matrix for k^{th} user

- The received signal vector for the kth user is given as

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k \mathbf{x} + \mathbf{n}_k \\ &= \mathbf{H}_k \sum_{i=1}^{k} \mathbf{P}_i \mathbf{s}_i + \mathbf{n}_k \end{aligned}$$

Stacking up all user signals, we have

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_k \end{bmatrix} [\mathbf{P}_1, \cdots, \mathbf{P}_k] \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_k \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_k \end{bmatrix}$$

Let us denote

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{bmatrix}, \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_k \end{bmatrix}, \mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_k], \mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_k \end{bmatrix} \mathbf{n} = \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_k \end{bmatrix}$$

it follows $\mathbf{y} = \mathbf{HPs} + \mathbf{n}$

5.1 Channel inversion (CI)

- CI is applied to MISO systems($M_R = 1$) by employing the right inverse of $\mathbf{H}$ as a precoding matrix

$$\mathbf{P} = \frac{1}{\sqrt{\gamma}}\mathbf{H}^*(\mathbf{HH}^*)^{-1}$$

where γ is a power normalization factor

- The received signal is given by

$$\begin{aligned}\mathbf{y} &= \mathbf{H}\frac{1}{\sqrt{\gamma}}\mathbf{H}^*(\mathbf{HH}^*)^{-1}\mathbf{s} + \mathbf{n} \\ &= \frac{1}{\sqrt{\gamma}}\mathbf{s} + \mathbf{n}\end{aligned}$$

Thus, the inter-user interference is completely eliminated, and the effective channel gain becomes $\frac{1}{\sqrt{\gamma}}$ for all users.

- The power normalization factor γ is determined from the Tx power constraints as

$$\begin{aligned}\mathrm{tr}(\mathbf{PP}^*) = \mathrm{tr}(\mathbf{P}^*\mathbf{P}) = \mathrm{tr}(&\frac{1}{\gamma}(\mathbf{HH}^*)^{-1}\mathbf{HH}^*(\mathbf{HH}^*)^{-1}) \\ &= \frac{1}{\gamma}\mathrm{tr}((\mathbf{HH}^*)^{-1}) \le \mathrm{Px} \\ &\Rightarrow \gamma = \frac{1}{\mathrm{Px}}\mathrm{tr}((\mathbf{HH}^*)^{-1})\end{aligned}$$

- In MISO, $\mathbf{H}_k$ becomes a row vector of length M_T, and $\mathbf{H}$ is a $k \times M_T$ matrix. Thus, the right inverse of $\mathbf{H}$ exists only when $k \le M_T$.

- γ increases when each channel vector $\mathbf{H}_k$ becomes correlated or have small energy.

- Large γ causes power boost which leads to small channel gains.

- CI is the simplest MU-MIMO method which eliminates the inter-user interference. The Rx becomes very simple, and does not need the channel information. Only γ should be reported to all users.

- The CI performance is limited due to the power loss γ, and is applicable only to users with single antenna.

5.2 Regularized CI (RCI)

- RCI employs a regularized inverse as

$$\mathbf{P} = \frac{1}{\sqrt{\gamma}}\mathbf{H}^*(\mathbf{H}\mathbf{H}^* + \alpha\mathbf{I})^{-1}$$

where the optimum α is derived as $\alpha = \frac{k}{\mathrm{SNR}}$

- The RCI incurs residual inter-user interference, but the performance gain is obtained over CI due to decreased γ.

- The sum rate of CI approaches that of RCI for high SNR.

$$\begin{aligned}\mathbf{y} &= \mathbf{H}_k \sum_{j=1}^{k} \mathbf{P}_j \mathbf{s}_j + \mathbf{n}_k \\ &= \mathbf{H}_k \mathbf{P}_k \mathbf{s}_k + \underbrace{\mathbf{H}_k \sum_{j \neq k} \mathbf{P}_j \mathbf{s}_j}_{\text{multi}-\text{user interference (MUI)}} + \mathbf{n}_k\end{aligned}$$

5.3 Vector perturbation (VP)

- VP employs the idea of Tomlinson-Harashima precoding (THP) to reduce the power loss.

- Each element of the input vector $\mathbf{s}$is perturbed by an integer vector $\mathbf{l}$ by

$$\tilde{\mathbf{s}} = \mathbf{s} + \tau \mathbf{l}$$

where the scalar τ is choosen large enough so that the Rx apply the modulo operation.

- The perturbation vector $\mathbf{l}$ is selected which minimizes $\gamma = ||\mathbf{x}||^2$ as

$$\mathbf{l} = \arg\min_{\mathbf{l}'} ||\mathbf{H}^*(\mathbf{H}\mathbf{H}^*)^{-1}(\mathbf{s} + \tau \mathbf{l})||^2$$

where sphere encoder can be applied to perform search over all possible candidates.

- The received signal for user k is given by

$$y_k = \frac{1}{\sqrt{\gamma}}(s_k + \tau l_k) + n_k$$

To remouve the effect of integer multiples of τ the modulo operation is applied at the Rx.

- Since the power loss is reduced, the VP performance is largely improved over CI at the high complexity in the Tx which requires a closet-point search in lattice

5.4 Block diagonalization (BD)

- BD was proposed for MU-MIMO systems where each user equipped with multiple antennas gets M_R independent data streams

- The received signal within one user should be properly combined

- There is no need for complete diagonalization of the channel, and BD requires only block diagonalization where MUI is completely eliminated.

- The effective channel is given by

$$\mathbf{HP} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_k \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & \cdots & \mathbf{P}_k \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1\mathbf{P}_1 & \mathbf{H}_1\mathbf{P}_2 & \cdots & \mathbf{H}_1\mathbf{P}_k \\ \vdots & & & \vdots \\ \mathbf{H}_k\mathbf{P}_1 & \mathbf{H}_k\mathbf{P}_2 & \cdots & \mathbf{H}_k\mathbf{P}_k \end{bmatrix}$$

where $\mathbf{H} \in \mathbb{C}^{kM_R \times M_T}$
To have full streams of M_R for k users, we have $M_T \leq kM_R$

- Denoting $\mathbf{M}_j$ as the receive filter for user j, the overall Rx filter $\mathbf{M}$ is defined as $\mathbf{M} = \text{diag}\{\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_k\}$, which implies no cooperation among users.

- Then the receive filter output is given as

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \end{bmatrix} = \mathbf{My} = \begin{bmatrix} \mathbf{M}_1\mathbf{y}_1 \\ \vdots \\ \mathbf{M}_k\mathbf{y}_k \end{bmatrix} = \mathbf{MHPs} + \mathbf{Mn}$$

where $\mathbf{z}_j$ is written by

$$\mathbf{z}_j = \mathbf{M}_j\mathbf{H}_j\mathbf{P}_j\mathbf{s}_j + \mathbf{M}_j\mathbf{H}_j \sum_{i \neq j} \mathbf{P}_i\mathbf{s}_i + \mathbf{M}_j\mathbf{n}_j$$

- All the MUI becomes zero for BD when

$$\mathbf{H}_i\mathbf{P}_j = 0 \qquad \text{for} \;\; i \neq j$$

$\Rightarrow$ The effective channel becomes block-diagonal.

- BD consists of two stage process : MUI suppression and single user processing.

1) MUI suppression
. Denoting $\tilde{\mathbf{H}}_j = \left[\mathbf{H}_1^T \cdots \mathbf{H}_{j-1}^T \mathbf{H}_{j+1}^T \cdots \mathbf{H}_k^T\right]^T$. Then in order to have zero MUI, the precoding matrix for user j should be in the null space of $\mathbf{H}_j$.

. Defining L_j as the rank of $\tilde{\mathbf{H}}_j$, the SVD of $\tilde{\mathbf{H}}_j$ is written as

$$\tilde{\mathbf{H}}_j = \tilde{\mathbf{U}}_j \tilde{\mathbf{\Lambda}}_j \begin{bmatrix} \tilde{\mathbf{V}}_j^{(1)*} \\ \tilde{\mathbf{V}}_j^{(0)*} \end{bmatrix}$$

where $\tilde{\mathbf{\Lambda}}_j$ has L_j non-zero diagonal elements, $\tilde{\mathbf{V}}_j^{(1)}$ is composed of the first L_j right singular vectors, and $\tilde{\mathbf{V}}_j^{(0)}$ holds the last $M_T - L_j$ right singular vectors.

. Since $\tilde{\mathbf{V}}_j^{(0)}$ forms an orthogonal basis for the null space of $\tilde{\mathbf{H}}_j$, we have

$$\tilde{\mathbf{H}}_j \tilde{\mathbf{V}}_j^{(0)} = \tilde{\mathbf{U}}_j \tilde{\mathbf{\Lambda}}_j \begin{bmatrix} \tilde{\mathbf{V}}_j^{(1)*} \\ \tilde{\mathbf{V}}_j^{(0)*} \end{bmatrix} \tilde{\mathbf{V}}_j^{(0)} = \tilde{\mathbf{U}}_j \tilde{\mathbf{\Lambda}}_j \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} = 0$$

. Applying $\tilde{\mathbf{V}}_j^{(0)}$ as the precoder for user j, the effective channel becomes $\mathbf{HP} = \text{diag}\{\mathbf{H}_1\tilde{\mathbf{V}}_1^{(0)}, \cdots, \mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}\}$. Thus the j^{th} user has a block channel $\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}$.

2) Single user processing

. After the MUI is eleminated with the precoder $\tilde{\mathbf{V}}_j^{(0)}$ the received signal at the user j is given by

$$\mathbf{y}_j = \mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}\mathbf{s}_j + \mathbf{n}_j = \mathbf{U}_j^{(b)}\mathbf{\Lambda}_j^{(b)}\mathbf{V}_j^{(b)*}\mathbf{s}_j + \mathbf{n}_j$$

. In order to decouple this block channel into M_R paralled subchannels, the SVD of $\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}$ is computed by

$$\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)} = \mathbf{U}_j^{(b)}\mathbf{\Lambda}_j^{(b)}\mathbf{V}_j^{(b)*}$$

. Employing $\tilde{\mathbf{V}}_j^{(0)}\mathbf{V}_j^{(b)}$ and $\mathbf{M}_j = \mathbf{U}_j^{(b)*}$, the j^{th} user signal vector becomes

$$\mathbf{z}_j = \mathbf{\Lambda}_j^{(b)}\mathbf{s}_j + \mathbf{U}_j^{(b)*}\mathbf{n}_j$$

. To maximize the sum rate, the optimal power loading matrix $\mathbf{\Phi}_j$ should be applied using the water-filling method.

- Finally, the precoding matrix for BD is given by

$$\mathbf{P} = \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)}\mathbf{V}_1^{(b)} & \cdots & \tilde{\mathbf{V}}_k^{(0)}\mathbf{V}_k^{(b)} \end{bmatrix} \mathbf{\Phi}^{\frac{1}{2}}$$

$$\begin{aligned}
\mathbf{y} &= \mathbf{HPs} + \mathbf{n} \\
&= \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_k \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)}\mathbf{V}_1^{(b)} & \cdots & \tilde{\mathbf{V}}_k^{(0)}\mathbf{V}_k^{(b)} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_k \end{bmatrix} + \mathbf{n} \\
&= \begin{bmatrix} \mathbf{H}_1\tilde{\mathbf{V}}_1^{(0)}\mathbf{V}_1^{(b)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}\mathbf{V}_k^{(b)} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_k \end{bmatrix} + \mathbf{n}
\end{aligned}$$