MIMO: Spatial Multiplexing/MU-MIMO/STBC/beamforming

2-3 hr

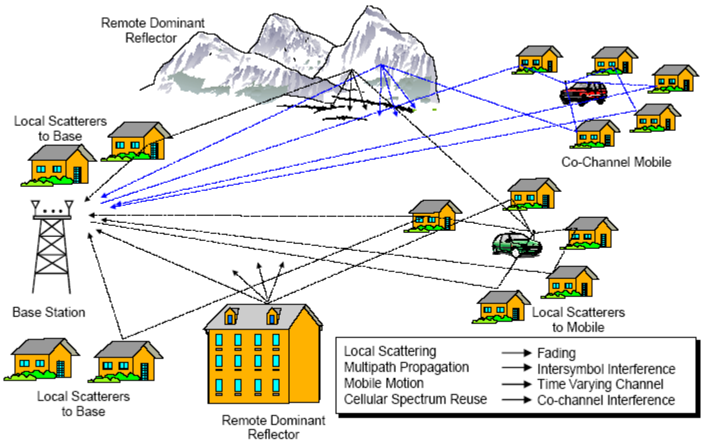
Outline

- Channel model and Channel capacity
- Spatial multiplexing (SM): VBLAST, for high SNR)

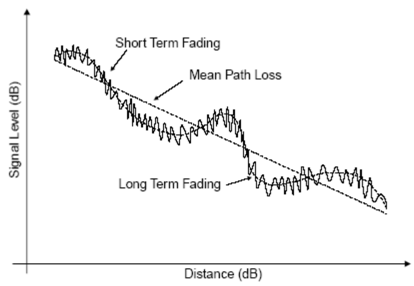
P.S. MU-MIMO: multiple single-antenna MSs treated like one multi-antenna MS, special form of SM

- Spatial transmit diversity: STBC, **for low SNR**
- Spatial filter or Beamforming,

Wireless Channel Model



Signal Level



- Slow fading (shadowing) is caused by large obstructions between transmitter and receiver.
- Fast fading is due to scattering of the signal by object near transmitter.
- Path loss is proportional to $1/r^{\alpha}$, α is between 2.5 and 5.

MIMO Channel Capacity (bits/sec/use)

Let ρ be the SNR

- Channel capacity for single-input single output (SISO) channel:

$C \sim \log_2(1+ \rho|H|^2)$

- Channel capacity for multiple-input multiple-output (MIMO) channel:

$C \sim \log_2 \det(I_N + \rho HH^*/M)$ P.S. total transmit energy=1

Where M is # of TX antennas, N is # of Rx antennas, H is the NXM channel matrix, and superscript * denotes transpose conjugate

- Ignore I_N then

Det(diagonal)=().().() => log det=()+()+()

MIMO Channel Capacity (bits/sec/use)

- the capacity grows linearly with min(M,N) rather than logarithmically
- This result can be intuited as follows: the determinant operator yields a product of min(M,N) nonzero eigenvalues (座標值) of its (channel-dependent) matrix argument, each eigenvalue characterizing the SNR over a so-called channel eigenmode. An eigenmode corresponds to the transmission using a pair of right and left singular vectors of the channel matrix as transmit antenna and receive antenna weights, respectively. Thanks to the properties of the , the overall capacity is the sum of capacities of each of these modes, hence the effect of capacity multiplication.
- eigenvalues (座標值) eigenvectors (座標軸) =>坐標轉換
- P.S. DFT and IDFT are also 坐標轉換

Spatial multiplexing: VBLAST

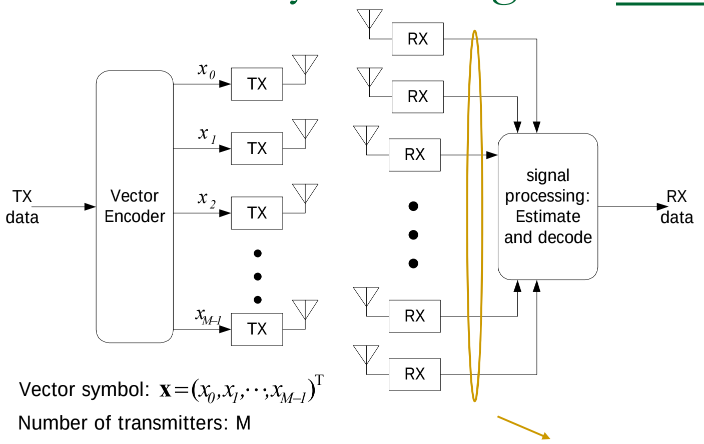
Outline

- Introduction to MIMO system
- V-BLAST (vertical-Bell Laboratories Layered Space-Time):

Spatial Multiplexing for higher data rates is mandatory in 802.11n (2009) 802.11ac (2013) and LTE (R8-)

- Parallel multistage detection
- Conclusion

V-BLAST system diagram $N>=M$



Vector symbol: $\mathbf{x}=(x_0,x_1,\cdots,x_{M-1})^T$

Number of transmitters: M

Number of receivers: N

$r=H_{N*M}x+v$, $N\geq M$ (# of eq>=no. of variables)

圖二、作用在1.9GHz，訊號傳輸頻寬為30KHz ，12根傳送
16根接收天線V-BLAST原型機（圖片來源Bell Labs）

MIMO detection is like CDMA multiuser detection

- The goal of signal processing is to separate each x_i from received signal, *(h_{ij}, i=1,…,N, is like N-chip random signature of x_i in DS CDMA, but no bandwidth expansion!)*
- **(column space)** $\mathbf{r} = \mathbf{Hx} + \mathbf{v} = \sum x_i \mathbf{h}_i + \mathbf{v}$, where h_i is the column of $\mathbf{H}$ similar to matrix form of CDMA data detection
- V-BLAST: detects the x_i with largest SNR at each stage, and subtracts its' associated effect from received signal iteratively, until detects all desired signals. (similar to SIC in CDMA multiuser detection)
- How do we choose the strongest signal form received signal?   An optimal detection ordering was proposed in V-BLAST

V-BLAST

- The strongest signal can be selected, by selecting the minimum norm of each row of H^+ (pseudo-inverse of H), because of less noise enhancement.

V-BLAST pseudo code

Initialization: $i=1$,

$\mathbf{G}_1 = \mathbf{H}^{+}$ (pinv in MATLAB)

$= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$,

$k_1 = \arg\min_j \left\|(\mathbf{G}_1)_j\right\|^2$, j-th row of $\mathbf{G}_1$

Recursion:

$y_{k_i} = (\mathbf{G}_i)_{k_i}\mathbf{r}_i$

$\hat{x}_{k_i} = Dec(y_{k_i})$ hard decision

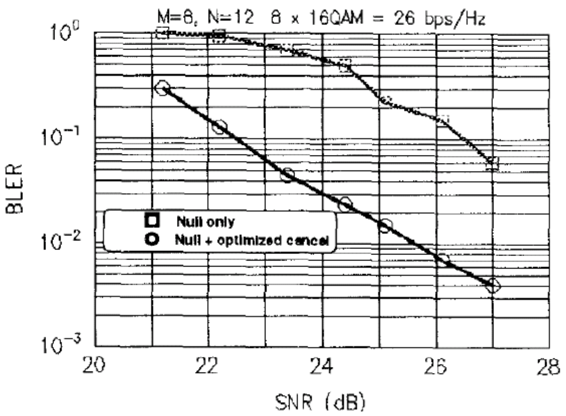
$\mathbf{r}_{i+1} = \mathbf{r}_i - \hat{x}_{k_i}\mathbf{H}_{k_i}$ k_i – th column of $\mathbf{H}$

$\mathbf{G}_{i+1} = (\mathbf{H}_{\bar{k}_i})^{+}$ set $k_1,\ldots,k_i$ – th column of $\mathbf{H}$ to zeros

(so $\mathbf{x}_{k_1},\ldots,\mathbf{x}_{k_i}$ are inactive)

$k_{i+1} = \arg\min_{j \notin \{k_1,\ldots,k_i\}} \left\|(\mathbf{G}_{i+1})_j\right\|^2$

Spectral efficiency: 20~40bps/Hz at SNR 24~34dB



Null only: not optimal ordering

$$E_s = \frac{(8\,\text{xmtrs})\times(4\,\text{b/sym/xmtr})\times(24.3\,\text{ksym/s})}{30\,\text{kHz}}$$

$= 25.9$ bps/Hz

BW

The power of each transmitting Antennas is proportional to 1/M
So the total radiated power is indep of M

the horizontal axis is spatially averaged received SNR=average of SNRi, where SNRi is the ratio of the total received signal power from all M transmitting antennas to noise power at the *i*-th receiving antenna

2 TX antennas example

$\mathbf{r}_1 = \mathbf{Hx} + \mathbf{v} = x_1\mathbf{h}_1 + x_2\mathbf{h}_2 + \mathbf{v}$, where $\mathbf{h}_i$ is the i-th column of $\mathbf{H}$

$\mathbf{G}_1 = \mathbf{H}^+$

Note that $\begin{bmatrix} (\mathbf{G}_1)_1 \\ (\mathbf{G}_1)_2 \end{bmatrix} [\mathbf{h}_1 \quad \mathbf{h}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

If $\|(\mathbf{G}_1)_2\|$ is minimum

$\hat{x}_2 = Dec((\mathbf{G}_1)_2 \mathbf{r}_1)$

$\mathbf{r}_2 = \mathbf{r}_1 - \hat{x}_2\mathbf{h}_2 + \mathbf{v}$

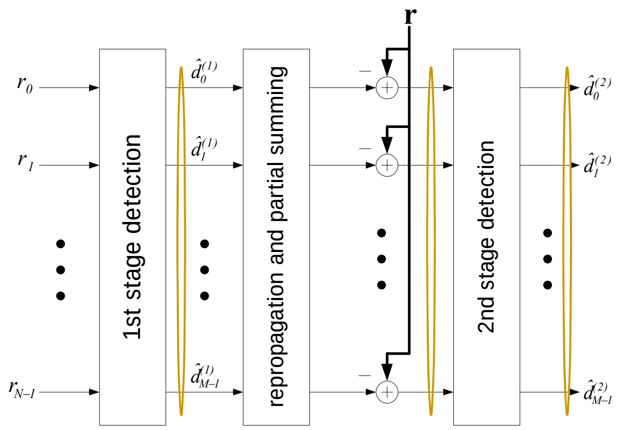
$\mathbf{G}_2 = (\mathbf{H}_{\bar{2}})^+$

$\hat{x}_1 = Dec((\mathbf{G}_2)_1 \mathbf{r}_2)$

Parallel multistage detection

- Disadvantage of V-BLAST:
  power ordering problem
- Relative to V-BLAST (SIC), a parallel multistage detection (PIC) was proposed.
- In MUD point of view, parallel detection can overcome these problems and get better results.

Parallel multistage detection



$\hat{\mathbf{d}}^{(1)} = \mathbf{Dec}(\mathbf{H}^{+}\mathbf{r})$

ZF nulling

$\tilde{\mathbf{r}}_k^{(i)} = \mathbf{r} - \hat{\mathbf{I}}_k^{(i)}$

$\hat{\mathbf{I}}_k^{(i)} = \mathbf{H}\mathbf{J}_k\hat{\mathbf{d}}^{(i-1)}$

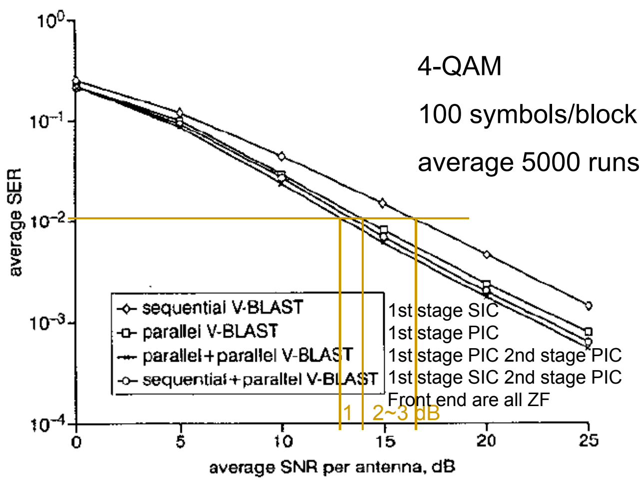
$\hat{\mathbf{d}}^{(i)} = \mathbf{Dec}(\mathbf{H}_k^{+}\tilde{\mathbf{r}}_k^{(i)})$

k=*0,1,…,M*-1

J_k: identity matrix except that the k-th diagonal component is zero

Parallel multistage detection

Assume Rayleigh fading indep between diff times



symbol error rate for 2*2 system

Note

- The total transmitted power of ALL transmitting antennas must be 1. ***(The more transmitting antennas, the less energy one transmitting antenna gets)***

Multiuser MIMO (MU-MIMO)

"Virtual" VBLAST (many single-antenna users' => "one virtual user" with many antennas)

M-fold increase in sum capacity

- Sum capacity of SU-MIMO (like VBLAST) is scaled by min{M,N} at high SNR, where M, N are # of antennas at Tx and RX (M or N can be 1 at MS=> min{M,N}=1).
- M<=N in p.13, but M>N in typical cellular system DL
- Multiplexing n users into MU-MIMO transmission=> sum capacity= min{M,Nn}=M if Nn>M ***M-fold increase!***

Less susceptible to channel

- SU-MIMO susceptible to poor propagation channels such as LOS or strong correlation between antenna elements=> rank decrease=>capacity decreases
- Antenna elements of MU-MIMO are further away, so less susceptible.

Challenges of MU-MIMO

- Channel quality indicator (CQI, to be fed back to BS for co-scheduled MSs and AMC) at MS assume no interference by concurrent transmission from BS to co-schedule MSs=> CQI too optimistic.
- Multiplexing 4 users (max in LTE-A) means ¼ power for each user and inter-user interference.

MU-MIMO is NOT always better than SU-MIMO

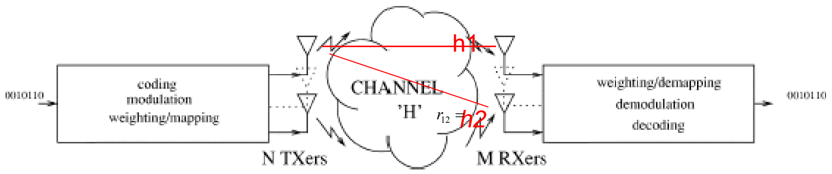
- MU-MIMO is viable if many MSs have high SNR and the rate gain due to spatial multiplexing and multiuser diversity outweighs the per-MS rate loss ($C \sim \log_2(1+\rho|H|^2)$) caused by power splitting and increased interference.

Spatial transmit diversity: STBC

Outline

- Space-time channel model
- MIMO Capacity
- Space-time code model
- Performance analysis
- Space-time block codes (STBC)
- Cooperative communication systems

Diagram of a MIMO wireless transmission system



Virtual user in CDMA

[h1 h2] like "code" or "signature" in CDMA
MIMO can't work in AWGN channel
(because all spreading sequence are the same).
It exploits multipath fading, not mitigate it!

MIMO and smart antenna

- smart antenna emphasizes the weight selection algorithm rather than in the coding
- A key concept in smart antennas is that of beamforming (by weight selection) by which one increases the average signal-to-noise ratio (SNR) through focusing energy into desired directions, in either transmitter or receiver.

MIMO and CDMA

- A strong analogy can be made with code-division multiple-access (CDMA) transmission in which multiple users sharing the same time/frequency channel are mixed upon transmission and recovered through their unique codes
- the advantage of MIMO is that the unique signatures of input streams ("virtual users") are provided by nature in a close-to-orthogonal manner (depending however on the fading correlation) without frequency spreading, hence at no cost of spectrum efficiency.
- A key feature of MIMO systems is the ability to turn multipath fading, traditionally a pitfall of wireless transmission, into a benefit for the user.
- Another advantage of MIMO is the ability to jointly code and decode the multiple streams since those are intended to the same user.
- the isomorphism between MIMO and CDMA can extend quite far into the domain of receiver algorithm design

- any proposed MUD algorithm can be recast in the MIMO context if the input of the MIMO system are seen as virtual users.
- A difference here is that the separation is carried out in the spatial channel domain rather than the code domain, making its success dependent on channel realizations.
- On the other hand, the complexity of CDMA-SIC is much higher than in the MIMO case since the number of CDMA users may go well beyond the number of virtual users/antennas in a single MIMO link.

Space-time code model

- N transmit and M receive antennas
- NXM ***independent*** ***slowly-varying*** Rayligh fading channels
- At any given time, N signals are transmitted simultanenously, one from each transmit antenna.
- The fading are assumed to be fixed during a ST code block and independent from block to blocks

STC model

Denote l for time index

transmitted code vector

$$\mathbf{c}_l = [c_1(l), c_2(l), \ldots, c_N(l)]^T$$

channel matrix

$$\mathbf{H} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & & \alpha_{1N} \\ \alpha_{21} & & & \\ & & & \\ & & & \alpha_{MN} \end{bmatrix}$$

received signal vector

$$\mathbf{r}_l = \mathbf{H}\mathbf{c}_l + \mathbf{n}(l)$$

$$\text{Mx1} = (\text{MxN})(\text{Nx1}) + (\text{Mx1})$$

Performance Analysis

- Optional, can be skipped

transmitted code vector sequence (subscript $l = 1,\ldots,\mathrm{L}, =$ slot index)

$\mathbf{C} = [\mathbf{c}_1,\ldots,\mathbf{c}_\mathrm{L}]$

Probability of error : assume perfect channel knowledge

$$\begin{pmatrix} \text{[Stuber, p.136] pairwise error probability} \le \exp(-\Gamma_s \left\| \mathrm{C} - \tilde{\mathrm{C}} \right\|^2) \\ \text{Chernoff bound approxmiation, so a product form can be obtained.} \end{pmatrix}$$

To the same rx antenna j

$$\Pr\{\mathbf{C} \to \tilde{\mathbf{C}} \mid \mathbf{H}\} \le \exp\{-d^2(\mathbf{C}, \tilde{\mathbf{C}})\}$$

$$d^2(\mathbf{C}, \tilde{\mathbf{C}}) = \Gamma_s \sum_{j=1}^{M} \sum_{l=1}^{L} \left| \boxed{\sum_{i=1}^{N} \alpha_{ji}[c_i(l) - \tilde{c}_i(l)]} \right|^2 = \Gamma_s \sum_{j=1}^{M} \mathbf{h_j A h_j}^*$$

where $\Gamma_s = E_s / 4N_0$, $\mathbf{h_j} = [\alpha_{j1},\ldots,\alpha_{jN}]$

$\mathbf{A} = \mathbf{BB}^*$

$$\mathbf{B} = \begin{bmatrix} c_1(1) - \tilde{c}_1(1) & c_1(2) - \tilde{c}_1(2) & & c_1(L) - \tilde{c}_1(L) \\ & & & \\ & & & c_N(L) - \tilde{c}_N(L) \end{bmatrix}_{NXL} \quad \text{error matrix}$$

$\mathbf{A} = \mathbf{BB}^*$ is Hermitain so $\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^*$

where $\mathbf{\Lambda}$ = diagonal matrix of eigenvalues $\lambda_1, \lambda_2, ..., \lambda_N$,

$\mathbf{U}$ is orthonormal and its columns are the eigenvectors of $\mathbf{A}$

(Gram Schmit procedure)

Let $\boldsymbol{\beta}_j = \mathbf{U}^* * \mathbf{h}_j^*$, then we have

$$d^2(\mathbf{C},\tilde{\mathbf{C}}) = \Gamma_s \sum_{j=1}^{M} \boldsymbol{\beta}_j^* \mathbf{\Lambda} \boldsymbol{\beta}_j = \sum_{j=1}^{M} \left(\sum_{i=1}^{N} \Gamma_s \lambda_i \,|\beta_{ij}|^2 \right)$$

Assume h_{ij} is Rayleigh random variable with unit power,

then x $=|\beta_{ij}|^2$ is exponentional distributed with unit power (Rayleigh^2)

The moment generating function of an **exponential** random variable

$\frac{1}{\mu} e^{-x/\mu}$ is $E[\exp(s\mathbf{x})] = \frac{1}{1 - s\mu}$ [Papoulis,3e,1991]

Therefore $E\left[\exp(-\Gamma_s \lambda_i \,|\boldsymbol{\beta}_{\mathbf{ij}}|^2)\right] = \frac{1}{1+\Gamma_s \lambda_i}$

$P.S.\ E\left[|\beta_{ij}|^2\right] = 1 \Rightarrow \mu = 1$

=> product of moment generating function (Lapace transform)

$$\Pr\{\mathbf{C} \to \tilde{\mathbf{C}} \mid \mathbf{H}\} \le \left(\prod_{i=1}^{N} \frac{1}{1+\Gamma_s \lambda_i} \right)^M$$

Note that $1+\Gamma_s \lambda_i \approx \Gamma_s \lambda_i$

Finally, the average error probability (some eigenvalues = 0)

$$\Pr\{\mathbf{C} \to \tilde{\mathbf{C}} \mid \mathbf{H}\} \le \left(\prod_{i=1}^{r} \lambda_i \right)^{-M} (\Gamma_s)^{-rM}$$

where $r \le N$ is the the rank, $\Gamma_s = E_s / 4N_0$

diversity order = rM

coding gain $= \left(\prod_{i=1}^{r} \lambda_i \right)^{1/r}$

1) Rank criterion : to achieve the maximum diversity NM,

the matrix B should be full rank (r = N) for any pair of different code vectors

P.S. For Alamouti code and 16QAM => 256 code matrices

2) Determinant criterion : to achieve the maximum coding gain,

the min determinant (over all code pairs) should be maximized.

P.S. determinant of diagonal matrix $= \prod_{i=1}^{r} \lambda_i$

Diversity order is more important

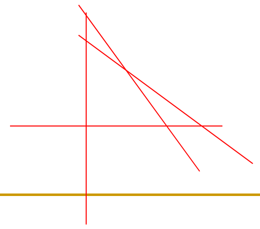
$$\log \Pr\{\mathbf{C} \rightarrow \tilde{\mathbf{C}} \mid \mathbf{H}) \leq -M \log\left(\prod_{i=1}^{r} \lambda_i\right) - rM \log(\Gamma_s)$$

In the $\log\Gamma_s - \log BER$ plot

diversity order = rM is the slope

coding gain $= \left(\prod_{i=1}^{r} \lambda_i\right)^{1/r}$ is the offset along y axis

So diversity order is more important



Alamouti code(1998)

- **Mandatory in LTE downlink transmission mode 2 (TM2)**
- **Optional in DVB-T2(2009)**
- **Optional in IEEE 802.11n(2009), 802.11ac (2013)**
- **802.16e: As of 2005 in WiMax Standard**
- **3G Cellular: 3GPP Release 6 specifies transmit diversity mode**
- first enables full diversity with *linear* processing at the receiver, not *exponentially* with the number of transmit antennas
- Also first open-loop transmit diversity technique which had this capability.

Space-time block codes (STBC): transmitter diversity- Alamouti code(1998)

original idea for 2X1 MISO system:

$$[c_1\ c_2] \rightarrow \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}$$

where $c_1\ c_2$ are two adjacent symbols

(complex-valued such as QAM)

Assume the channels are quasi-static: h_1 and h_2 are fixed during two consecutive symbol intervals

option in WCDMA/cdma2000/wirelessMAN/LTE because the diversity antennas can be put in BS instead of MS in the downlink.

$$\begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix} \begin{matrix} h_1 \\ h_2 \end{matrix} \quad [r_1 \quad r_2]$$

received signal:

$$r_1 = h_1 c_1 + h_2 c_2 + n_1$$

$$r_2 = -h_1 c_2^* + h_2 c_1^* + n_2$$

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix} = \mathbf{H} \cdot \mathbf{c} + \mathbf{n}$$

Note that $\mathbf{H}^H \mathbf{H} = \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} = (|h_1|^2 + |h_2|^2) \cdot \mathbf{I} = \|\mathbf{h}\|^2 \cdot \mathbf{I}$

$$\tilde{\mathbf{r}} = \mathbf{H}^H \mathbf{r} = \|\mathbf{h}\|^2 \cdot \mathbf{c} + \tilde{\mathbf{n}}$$

where $\tilde{\mathbf{n}} = \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}$

$$\text{so SNR gain} = \frac{(|h_1|^2 + |h_2|^2)^2}{|h_1|^2 + |h_2|^2} = |h_1|^2 + |h_2|^2$$

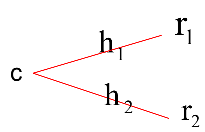
diversity order = 2

$$E[\mathbf{n}\mathbf{n}^H] = N_0 \cdot \mathbf{I} \text{ (complex white noise)}$$

$$E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H] = \mathbf{H}^{\mathbf{H}} E[\mathbf{n}\mathbf{n}^H]\mathbf{H} = (|h_1|^2 + |h_2|^2) \cdot N_0\mathbf{I} \text{ still white}$$

so c_1, c_2 are detected independently => low complexity

Comparison: receiver diversity order 2



By maximum ratio combining (MRC)

$$= h_1^* r_1 + h_2^* r_2$$

$$= c(|h_1|^2 + |h_2|^2) + h_1^* n_1 + h_2^* n_2$$

r_1

With M receive antennas

$$\mathbf{r}_m = \mathbf{H}_m \cdot \mathbf{c} + \mathbf{n}_m$$

where $\mathbf{H}_m = \begin{bmatrix} h_{1m} & h_{2m} \\ h^*_{2m} & -h^*_{1m} \end{bmatrix}$

$$\tilde{\mathbf{r}} = \sum_{m=1}^{M} \mathbf{H}_m^H \mathbf{r}_m = \sum_{m=1}^{M} \begin{bmatrix} h^*_{1m} & h_{2m} \\ h^*_{2m} & -h_{1m} \end{bmatrix} (\mathbf{H}_m \cdot \mathbf{c} + \mathbf{n}_m)$$

$$= \sum_{m=1}^{M} \left(\left(|h_{1m}|^2 + |h_{2m}|^2 \right) \mathbf{c} + \mathbf{H}_m^H \mathbf{n}_m \right)$$

a diversity order of 2M is achieved. (**no timing overhead!**)

only simple linear processing at the receiver (like handsets) is required

complete channel state information is required at the receiver.

In practice channel estimation is used to obtain CSI.

- STBC rate=no of data/no of time slot
- Ultimately, it is possible to code the signals so that the effective data rate is back to that of a single antenna system (rate 1 STBC, for example). Effectively, each TX antenna then sees a differently encoded, fully redundant version of the same signal.
- In this case, the multiple antennas are only used as a source of spatial diversity and not to increase data rate, or at least not in a *direct* manner.

N>2, no rate 1 OSTBC

- for general complex constellationslike M-QAM or M-PSK, it *is not known* whether a orthogonal STBC with transmission rate 1 and simple linear processing that will give the maximum diversity gain with TX antennas *does exist or not*.
- So rate 1 quasi-orthogonal STBC

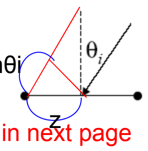
Spatial filter: Beamforming

Beamforming Basics

- The multipath channel:

$$h(t,\tau,\mathbf{z}) = \sum_i \alpha_i(t)\cdot e^{-j\varphi_i(t,\tau_i,\mathbf{z})}\cdot\delta(t-\tau_i)$$

$$\varphi_i(t,\tau_i,\mathbf{z}) = 2\pi\cdot\left[f_c\tau_i(t) - f_{d,i}t + \boxed{(z/\lambda)\cdot\sin\theta_i}\right]$$



= θ in next page

t : *time*

τ : *delay*

z : *distance*

f_d : *Doppler*

The signal s_i at each antenna is:

$$s_0(t) = s(t)$$

$$s_1(t) = s(t-\tau) \approx s(t)e^{-j\theta}$$

$$s_{M-1}(t) = s(t-(M-1)\tau) \approx s(t)e^{-j(M-1)\theta}$$

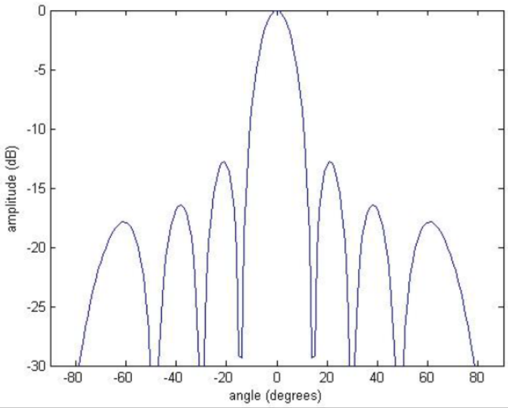
$$\mathbf{s}(t) = \begin{bmatrix} 1 \\ e^{-j\theta} \\ e^{-j2\theta} \\ e^{-j3\theta} \\ \vdots \\ e^{-j(M-1)\theta} \end{bmatrix}\cdot s(t) = \mathbf{a}(\theta)\cdot s(t),$$

Written as a vector:

$$\mathbf{s}(t) = \begin{bmatrix} 1 \\ e^{-j\theta} \\ e^{-j2\theta} \\ e^{-j3\theta} \\ \vdots \\ e^{-j(M-1)\theta} \end{bmatrix}\cdot s(t) = \mathbf{a}(\theta)\cdot s(t),$$

where **a** is the array steering vector.

- Figure below shows an example of the amplitude response of an antenna array with eight elements (uniform linear array, ULA) versus the angle θ. In this example, the maximum is obtained when a signal coming from the boresight direction (θ=0) impinges on the array.



Beamforming is made possible by weighting the magnitude and/or phase of the signal at the individual antennas:

$y(t) = \mathbf{w}^H \cdot \mathbf{a}(\theta) \cdot s(t)$,

where **w** is the weight vector. The signals are weighted so that they can be added constructively in the direction of an intended transmitter/receiver, and destructively in the direction of interferers.

- **One way of determining w is using DoA**
- If the position of the UE is known, the beamforming weightings can be adapted accordingly to optimize transmission for this UE. Therefore, specialized algorithms, such as MUSIC or ESPRIT (數位訊號處理), could be used in the base station to determine the DoA for the UE signal, and thus to determine its location.

(optional) Upper bound: Cramer Rao bound

- MIMO example

Performance Criteria for Estimators: introduce some def. and lemmas before Cramer-Rao bound.

Def: score function $s(\theta, x) = \frac{d}{d\theta} \log L(\theta, x)$

p.s. If θ is a vector $s(\theta, x) = \begin{bmatrix} \frac{\partial \log L(\theta,x)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L(\theta,x)}{\partial \theta_n} \end{bmatrix}$

Claim:If no prior (MAP=>ML), then $E[S(\theta, x)] = 0$

$$E\left[\frac{\partial \log p(x|\theta)}{\partial \theta_i}\right] = \int \left[\frac{1}{p(x|\theta)} \frac{\partial p(x|\theta)}{\partial \theta_i}\right] p(x|\theta) dx$$
$$= \frac{\partial}{\partial \theta_i} \int p(x|\theta) dx = \frac{\partial}{\partial \theta_i} 1$$
$$= 0$$

Def : Fisher information matrix

$$J(\theta) \stackrel{\Delta}{=} E\left[s(\theta,x)s(\theta,x)^T\right]$$

$$J(\theta) = -E\left[\tfrac{\partial}{\partial\theta} s(\theta,x)^T\right] \text{ if } s(\theta,x) = \tfrac{\partial}{\partial\theta}\log p(x|\theta)$$

$$\text{where } J_{ij} = \int \frac{\partial \log p(x|\theta)}{\partial\theta_i}\frac{\partial \log p(x|\theta)}{\partial\theta_j} p(x|\theta)dx$$

proof :

$$s(\theta,x) = \frac{\partial}{\partial\theta}\log p(x|\theta) = \frac{1}{p(x|\theta)}\frac{\partial}{\partial\theta}p(x|\theta)$$

$$\frac{\partial}{\partial\theta}s(\theta,x)^T = \frac{p(x|\theta)\frac{\partial}{\partial\theta}[\frac{\partial}{\partial\theta}p(x|\theta)]^T - [\frac{\partial}{\partial\theta}p(x|\theta)][\frac{\partial}{\partial\theta}p(x|\theta)]^T}{p^2(x|\theta)}$$

$$= \frac{\frac{\partial}{\partial\theta}[\frac{\partial}{\partial\theta}p(x|\theta)]^T}{p(x|\theta)} - [\frac{\partial}{\partial\theta}\log p(x|\theta)][\frac{\partial}{\partial\theta}\log p(x|\theta)]^T$$

$$E\left[\tfrac{\partial}{\partial\theta}s(\theta,x)^T\right] = \int \left(\frac{\frac{\partial}{\partial\theta}[\frac{\partial}{\partial\theta}p(x|\theta)]^T}{p(x|\theta)}\right)p(x|\theta)dx - E\left[s(\theta,x)s(\theta,x)^T\right]$$

Q.E.D.

Thm. Cramer-Rao bound

If $\hat{\theta}(x)$ is an unbiased estimator of θ

$J(\theta) = E\left[s(x,\theta)s(x,\theta)^T\right]$ where $s(x,\theta) = \frac{d}{d\theta}\log p(x|\theta)$, $C = E\left[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T\right]$

Then $C - J^{-1}$ is nonnegative definite

Proof: $E\left[\begin{pmatrix}\hat{\theta}-\theta \\ s(x,\theta)\end{pmatrix}\begin{pmatrix}\hat{\theta}-\theta \\ s(x,\theta)\end{pmatrix}^T\right] = \begin{bmatrix} C & I \\ I & J\end{bmatrix} \geq 0$

$$(x^T, -x^TJ^{-1})\begin{pmatrix} C & I \\ I & J\end{pmatrix}\begin{pmatrix} x \\ -J^{-1}x\end{pmatrix} \geq 0; \quad (x^TC - x^TJ^{-1}, x^T - x^T)\begin{pmatrix} x \\ -J^{-1}x\end{pmatrix} \geq 0$$

$x^T(C - J^{-1})x \geq 0;$ $C - J^{-1}$ nonnegative definite

p.s. Set

$$X = (0,\cdots 0,1,0,\cdots 0)^T$$

$\uparrow$ i-th position

$$\Rightarrow C_{ii} \geq (J^{-1})_{ii}$$

Nuisance parameters :

Suppose we have $(\theta_1, \theta_2, \theta_3, \cdots \theta_n)$

want to estimate θ_1

- If we know $\theta_2, \theta_3, \cdots \theta_n$,

$$Var\left[\hat{\theta}_1(x)\right] = E\left[\left(\hat{\theta}_1(x) - \theta_1\right)^2\right] \geq (J_{11})^{-1} = 1/J_{11}$$

- If we do not know $\theta_2, \theta_3, \cdots \theta_n$,

$$Var\left[\hat{\theta}_1(x)\right] \geq \left(J^{-1}\right)_{11}$$

Claim : $(J^{-1})_{ii} \geq 1/J_{ii}$

proof :

J > 0, so Schwartz inequality becomes

$(u^T Jv)^2 \leq (u^T Ju)(v^T Jv)$

Let $u = J^{-1}v$, $v = [0,...,0,1\ ,0,...,0]^T$

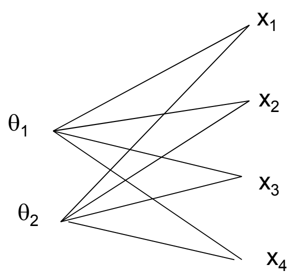
(1 in the i - th position)

$1 \leq \left(J^{-1}\right)_{ii}(J_{ii})$

$\left(J^{-1}\right)_{ii} \geq 1/J_{ii}$

$Q.E.D.$

A 2x4 MIMO system



Ex : (ref. Scharf p.238) signal detection in MIMO systems

$X = H\theta + noise \quad where \quad noise \sim N(0, I) \quad X \sim N(H\theta, I)$

$f_\theta(x) = \frac{1}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2}(X - H\theta)^T (X - H\theta)\right]$

$s(\theta, x) = H^T (X - H\theta)$

Note : you can prove it or

intuitively think it as square and then

make the dimention right for matrix multiplication (H^T not H)

$s(\theta, x) = H^T X - H^T H\theta = 0$

$\hat{\theta}_{ML} = \left(H^T H\right)^{-1} H^T X$

(used in MIMO system)

$$J(\theta) = -E\left[\tfrac{\partial}{\partial\theta} S(\theta, x)^T\right] = -E\left[\tfrac{\partial}{\partial\theta}\left(X^T H - \theta^T H^T H\right)\right] = H^T H$$

Then, let $H = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \\ 4 & 3 \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$

$$J(\theta) = H^T H = \begin{bmatrix} 18 & 14 \\ 14 & 14 \end{bmatrix}, \quad J^{-1}(\theta) = \frac{1}{56}\begin{bmatrix} 14 & -14 \\ -14 & 18 \end{bmatrix}$$

$Var\left[\hat{\theta}_1(x)\right] \geq \dfrac{1}{J_{11}} = 1/18$ if θ_2 is known

$Var\left[\hat{\theta}_1(x)\right] \geq \left(J^{-1}\right)_{11} = 1/4$ if θ_2 is not known