6. MAXIMUM LIKELIHOOD MULTIUSER DETECTION

- ML takes the use of the finite alphabet property to the limit. We saw in the toy system of Section 2.5 that it can almost eliminate the effect of MAI in AWGN – in fact, no detector can provide a lower BER.

- Because ML-MUD is based on enumeration over all data sequences, its complexity can be high – but there are suboptimal forms with lower.

- We'll see
  - ML-MUD when there's no signal structure (no signatures). It still works, suggesting a way out for CDMA systems with degeneracy.
  - ML-MUD with signatures. Early work, bounds.
  - Complexity management through tree pruning. Is this the way forward?

[Gran98] [Gran01] [Ho01] [Junt00a] [Rapa99] [Vasu86] [Verd86] [Wei97] [Wei98] [Zvon94]

6.1 ML·MUD Without Signatures

- Consider a system without signal structure:
  - all users have same pulse shape
  - flat fading, synchronous transmission

We saw in Section 3 that such a system can be represented by

$$y = CA\underline{b} + \underline{v} \,, \qquad R_v = N_0 I$$

- We know that linear methods, like ZF, cost an order of diversity for each nulled interferer, so that $K \le M$.
- With ML and perfect CSI, the detector performs

$$\hat{\underline{b}} = \arg\max_{\underline{b}} \; p_{y|b}(y|\underline{b}) = \arg\max_{\underline{b}} \frac{1}{(2\pi)^M N_0} \exp\left(-\frac{1}{2N_0} |y - CA\underline{b}|^2\right)$$

$$= \arg\min_{\underline{b}} |y - CA\underline{b}|^2$$

this is the metric for a specific hypothesis

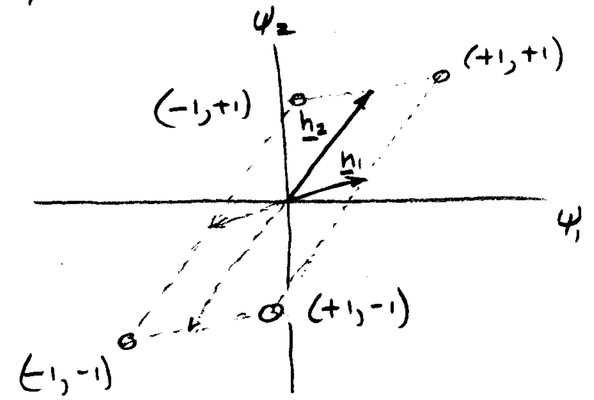
The argmin implies enumeration, so computation is proportional to 2^K. Limited to a small number of users?

- Pictures help to show what happens here.

For simplicity, define $H = CA$, so

$$\underline{y} = H\underline{b} + \underline{n} = \underline{h}_1 b_1 + \underline{h}_2 b_2 + \underline{n} \quad \text{for } K=2$$

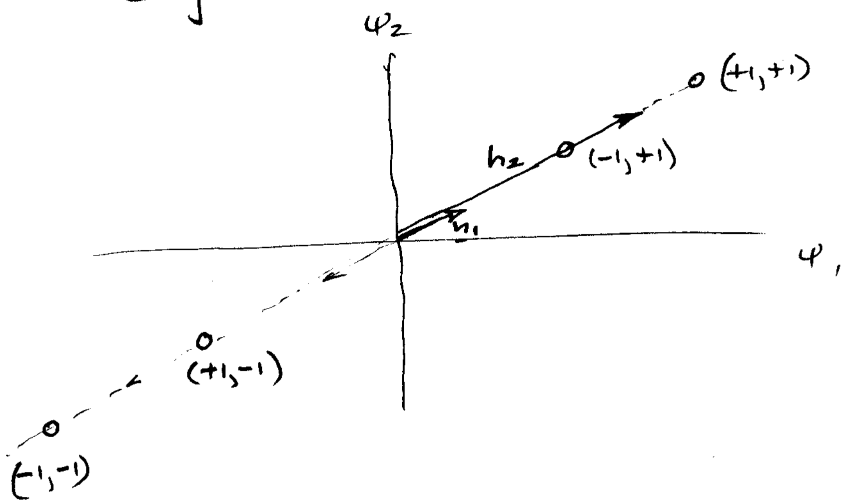
Also, make H real to facilitate sketching.



A 4-point "constellation"

It depends on the channel gains C, so is constantly changing.

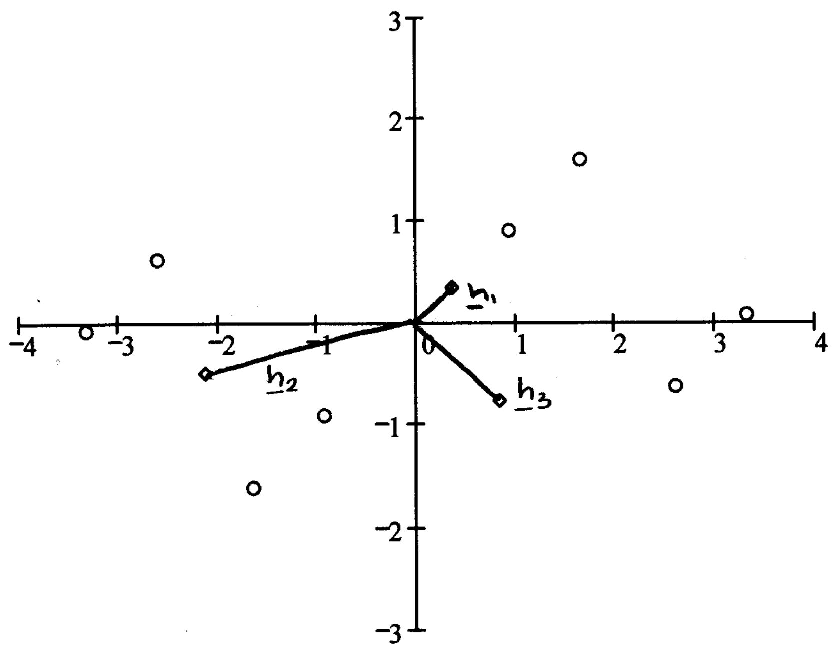
- ZF performance determined by proj'n onto orthonal complement to space spanned by other user(s)
- ML performance is determined by distance to decision boundaries – just a closest match.
- The gain vectors can be collinear



ML works

ZF fails

– We can even have more users than antennas



ZF will not work

MMSE produces an estimate, but it's useless

ML selects the point that is closest to the received $\underline{r}$

- Linear methods try to "undo" the interference caused by the channel gains, and deliver interference-suppressed estimates to a set of single-user detectors.

  In contrast, ML is a joint detection. It accounts for the interference without trying to remove it.

- How can we analyse such a complicated system?
  - Don't look at a specific H, get BER somehow, then average.
  - Instead, recognise that C is random, and focus on the pairwise error probability:

$$P_2(\underline{b}_j, \underline{b}_i) = \Pr[\underline{b}_i \text{ selected if } \underline{b}_j \text{ sent}]$$

$$= \Pr\left[\|\underline{y} - CA\underline{b}_i\|^2 < \|\underline{y} - CA\underline{b}_j\|^2\right] \quad \text{(perfect CSI)}$$

  - Define vectors for antenna m (both length $K+1$)

$$\underline{u}(\underline{b}) = \begin{bmatrix} 1 \\ -A\underline{b}^* \end{bmatrix} \qquad \underline{z}_m = \begin{bmatrix} y_m \\ {}_m\underline{c}^T \end{bmatrix}$$

${}_m\underline{c}$ is row m of C

so component m of $\underline{y} - CA\underline{b}_i$ is $\underline{u}^\dagger(\underline{b})\,\underline{z}_m$...

and metric is

$$\mu(\underline{b}) = \sum_{m=1}^{M} \underline{z}_m^\dagger\, \underline{u}(\underline{b})\, \underline{u}^\dagger(\underline{b})\, \underline{z}_m$$

So that the pairwise error prob is

$$P_2(\underline{b}_j, \underline{b}_i) = \Pr\left[\frac{1}{2}\sum_{m=1}^{M} \underline{z}_m^\dagger\, F(i,j)\, \underline{z}_m < 0\right]$$

where $F(i,j) = \underline{u}(\underline{b}_i)\underline{u}^\dagger(\underline{b}_i) - \underline{u}(\underline{b}_j)\underline{u}(\underline{b}_j)^\dagger$

and we know how to calculate this!

- The theory of Gaussian quadratic forms gives us the characteristic function of

$$\Delta\mu(i,j) = \sum_{m=1}^{M} \tfrac{1}{2}\, \underline{z}_m^{+} F(i,j)\, \underline{z}_m$$

as

$$\Phi_{ij}(s) = \frac{1}{\det(I + RF(i,j))^M}$$

where

$$R = \tfrac{1}{2} E\left[\underline{z}_m \underline{z}_m^{+}\right] = \begin{bmatrix} \sum A_m^2 \sigma_g^2 + N_0 & A_1 b_1 \sigma_g^2 & \cdots & A_K b_K \sigma_g^2 \\ A_1 b_1^* \sigma_g^2 & \sigma_g^2 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ A_K b_K^* \sigma_g^2 & 0 & \cdots\ 0 & \sigma_g^2 \end{bmatrix}$$

or

$$\Phi_{ij}(s) = \prod_{\ell=1}^{K+1} \left(\frac{-p_{ij\ell}}{s - p_{ij\ell}}\right)^M \quad \text{where pole } p_{ij\ell} = -1/\lambda_{ij\ell}$$

$$\lambda_{ij\ell} = \text{eigenvalue } \ell \text{ of } RF(i,j)$$

But the structure of F (difference of two rank-one matrices) makes its rank 2; consequently rank$(RF(i,j)) = 2$.

$$\Phi_{ij}(s) = \left[\frac{p_{ij1}\, p_{ij2}}{(s - p_{ij1})(s - p_{ij2})}\right]^M \qquad \begin{array}{ll} p_{ij1} < 0 & \text{left half plane} \\ p_{ij2} > 0 & \text{right half plane} \end{array}$$

and

$$P_2(i,j) = \frac{1}{(1 - r_{ij})^{2M-1}} \sum_{m=0}^{M-1} \binom{2M-1}{m} (-r_{ij})^m \qquad \text{pole ratio } r_{ij} = -\frac{p_{ij2}}{p_{ij1}}$$

- The SER of user k is then

$$P_{sk} = \Pr\left[\bigcup_{i \in E_k} \left[\Delta\mu(i,j) < 0 \right] \right] \qquad E_k = \{ i \mid \underline{b}_i \neq \underline{b}_j \text{ in pos'n } k \}$$

$$\leq \sum_{i \in E_k} P_2(i,j) \qquad \text{allow } b_j \text{ to be "all ones" by symmetry.}$$

- [Gran 98], [Gran 00] shows that the difference of metrics between true ($\underline{b}_j$) and erroneous ($\underline{b}_i$) data can be expressed as quadratic form.

$$D_{ij} = \sum_{m=1}^{M} \underline{z}_m^+ F_{ij} \underline{z}_m$$

$$\underline{z}_m = (y_m, c_{m1}, \ldots, c_{mK})^T$$

$$F_{ij} = (\underline{u}_i \underline{u}_i^+ - \underline{u}_j \underline{u}_j^+)^*$$

$$\underline{u}_i = (1, -A_1 b_{i1}, \ldots, -A_K b_{iK})^T$$

$$\underline{u}_j = (1, -A_1 b_{j1}, \ldots, -A_K b_{jK})^T$$

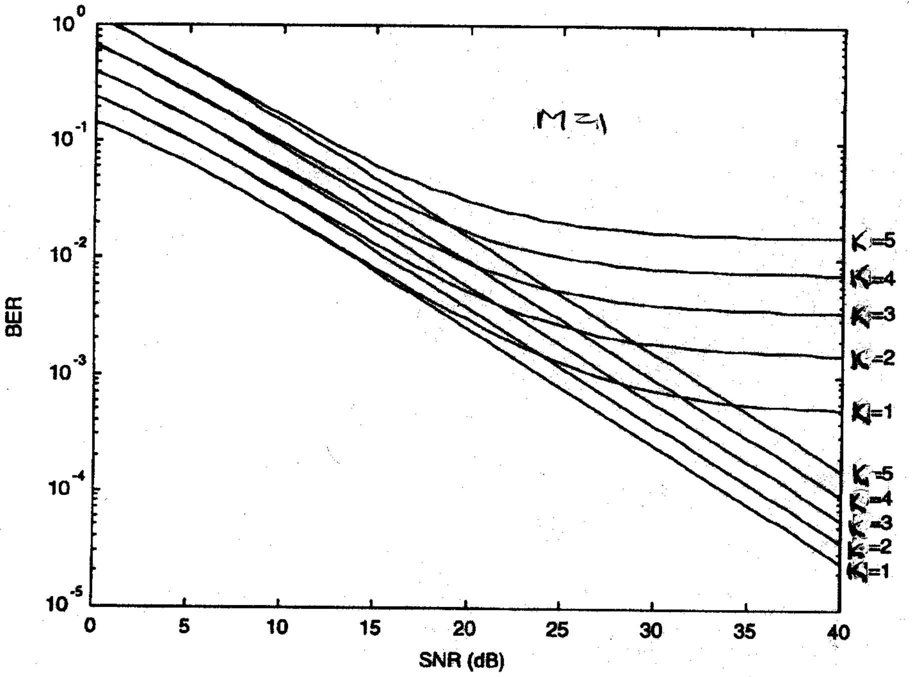
- Pairwise error prob ($\underline{b}_j \to \underline{b}_i$) is given by $Pr[D_{ij} < 0]$, evaluated in usual way for quad forms.

Overall error prob for user k is union bounded by

$$P_{ek} \leq \sum_{i \in \tilde{B}_k} Pr\left[\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \to \underline{b}_i \right]$$

where $\tilde{B}_k$ is set of sequences in which user k has $b_k = -1$ and the all-ones sequence is sent (w.o.l.o.g., since symmetric).

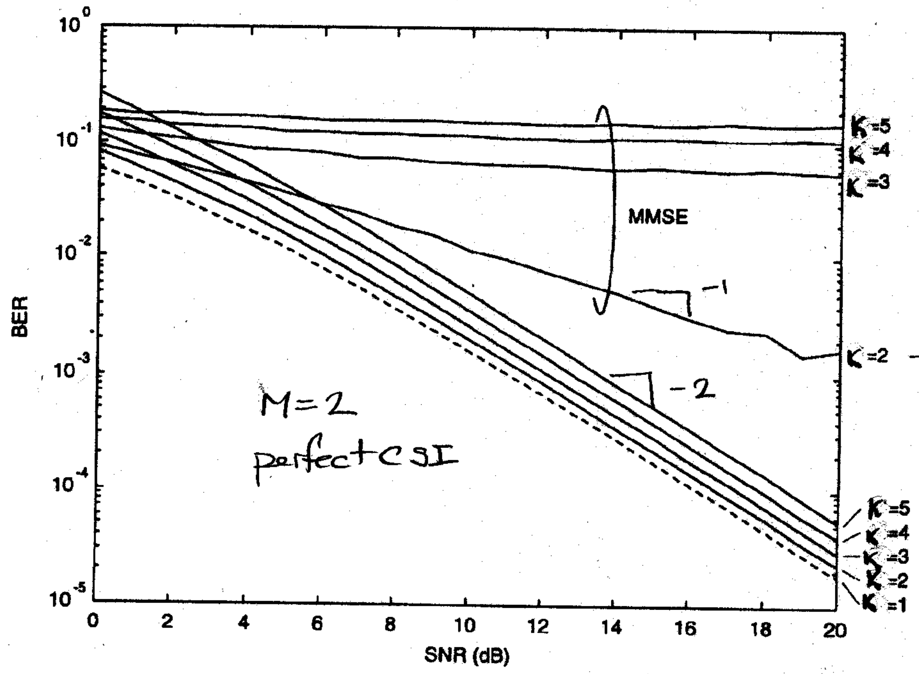
- Numerical results



imperfect CSI
$\rho = 0.999$

perfect CSI
no limit on number of users!

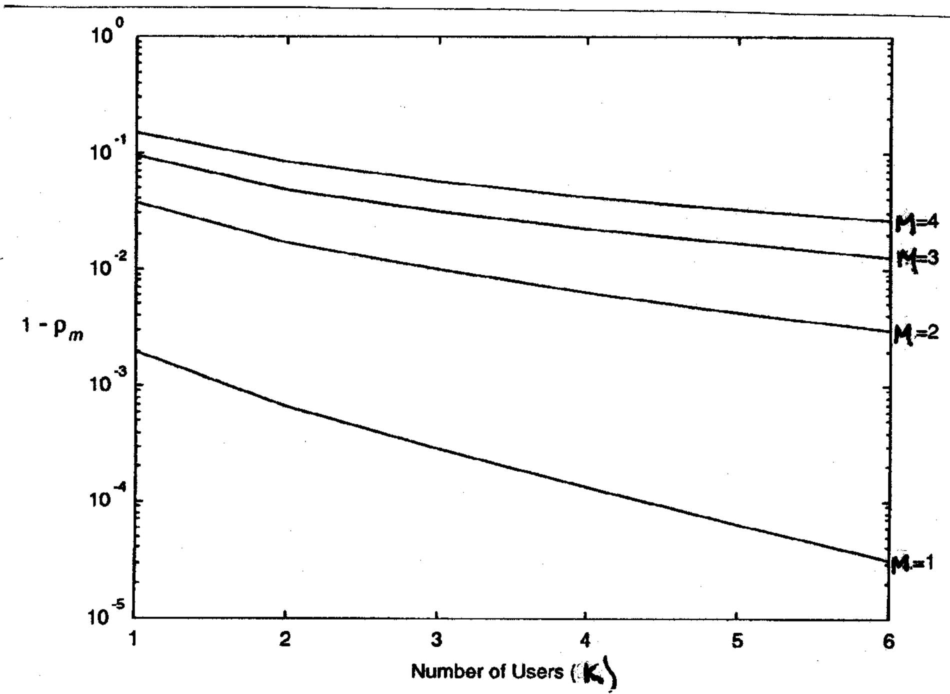
[Gran98]



broken

– lost one order of diversity

- No limit on number of users
- Retains diversity order
- SNR penalty is lower



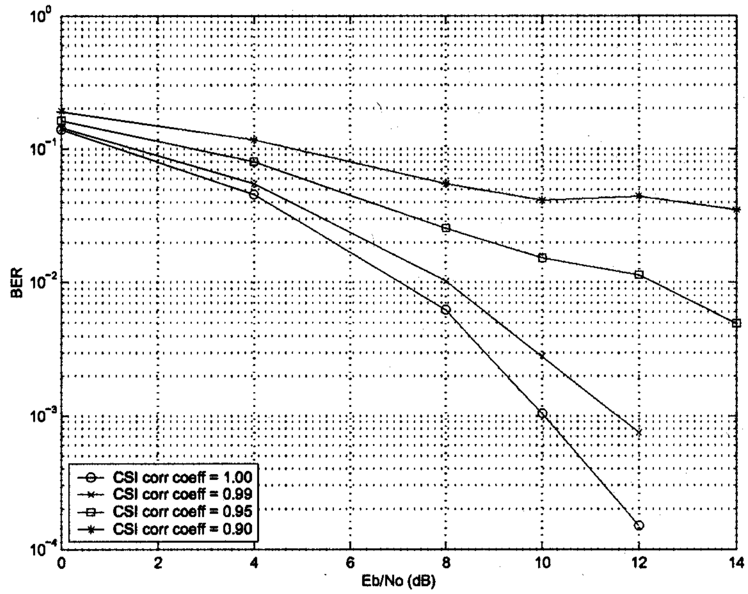
Channel estimation accuracy required for BER = 10^{-3}

Needs at least two antennas, realistically

- For asynch and/or delay spread, complexity shoots up through use of Viterbi algorithm:

  2^{KN_v} N_v symbol times to define Viterbi state.

Complexity mitigation through per-survivor processing in [Ho 01].



$M = 3$

$K = 2$

$N_v = 1$

[Ho 01] also reports relaxation of CSI accuracy as M increases.

- These results are <u>big</u>, for narrowband, since the normal value for K (# users) is 1.
- But how does it compare against MMSE combining, the other way of increasing simultaneous users?

From [Gran99] — outage prob, accounts for other cells also increasing user base (e.g. Section 1.2), increased cluster size

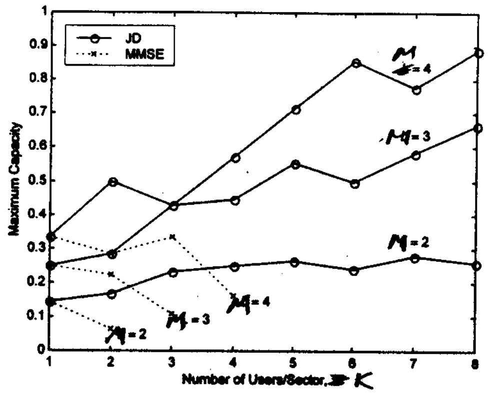


Figure 5: Maximum capcity η_{max} of joint detection (JD) and MMSE combining systems with perfect CSI.

← For 1% outage.
Capacity with ML increases
Capacity with MMSE <u>decreases</u>.

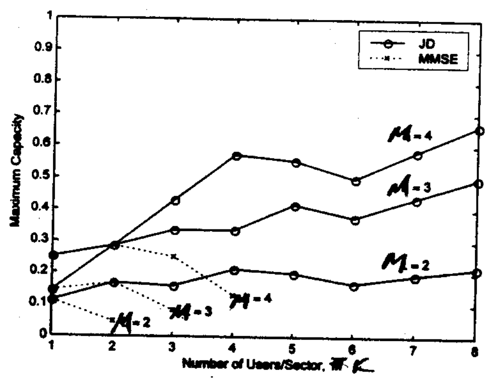


Figure 6: Maximum capcity η_{max} of joint detection (JD) and MMSE combining systems for a representative example of imperfect CSI.

← For imperfect CSI, too

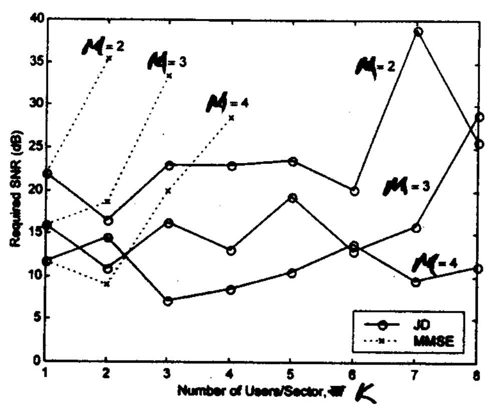


Figure 7: SNR $\Gamma_{N,r}$ required to achieve 1% outage probability for joint detection (JD) and MMSE combining systems with perfect CSI.

← Required SNR shoots up in MMSE as # users increases.

- Importance of these studies:
  - With ML, there is no limit to the number of users (at least for perfect CSI) and the order of diversity is retained.
  - It suggests that, with signature sequences and CDMA, the degeneracy of R may result only in an SNR penalty, not a hard limit on number of users.
  - Channel estimation accuracy needs to be high with several users (although it's reduced when number of antennas increases). Does use of signatures ease channel estimation requirement?
  - These are <u>speculations</u> w.r.t. ML-MUD in CDMA.