Semidefinite Relaxation in Action:

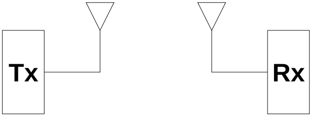
Efficient Soft MIMO Demodulation

Tim Davidson
with Mehran Nekuii, Mikalai Kisialiou and Zhi-Quan (Tom) Luo

Department of Electrical and Computer Engineering,
McMaster University
davidson@mcmaster.ca

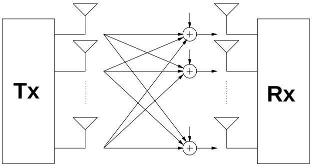
2 November 2010

Classical wireless communication



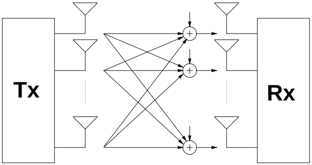
- Narrowband case, received signal: $y = hs + v$
- Coherent reception: h known at Rx, but not at Tx
- Richly-scattered environment, at high SNRs
  - ergodic capacity $\sim \log(\mathrm{SNR})$
  - outage probability $\sim \mathrm{SNR}^{-1}$

MIMO wireless communication



- System: M Tx and N Rx antennas
- In narrowband case, received signal: $\mathbf{y} = \mathbf{Hs} + \mathbf{v}$
- Coherent reception: $\mathbf{H}$ known at Rx, but not at Tx
- Richly scattered environment, at high SNRs,
  - ergodic capacity $\sim \min\{M, N\} \log(\text{SNR})$, or
  - outage probability $\sim \text{SNR}^{-MN}$
  - actually, there is an analytic trade-off between them
- Practical example: Manhattan, 4–16 antennas on back of a laptop, 2.11 GHz (Chizhik *et al*)

MIMO wireless communication

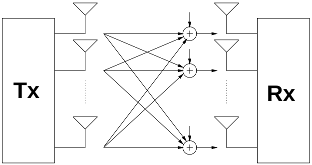


- Capacity and outage benefits obtained by exploiting multiple channels
- **Double edged sword:** interference between symbols
- **Question:** How can we trade benefits against decoding complexity

Reduced complexity MIMO receivers

- An active area of both R and D for $\sim$10 years.
  A number of interesting architectures proposed
- In a simple setting, demodulation problem is a boolean quadratic; suggests semidefinite relaxation (Steingrimsson *et al*)
  - while that has some interesting features
  - not competitive in conventional settings (too expensive)
- Are the other ways to leverage SDR to develop a competitive receiver?

Simple setting



- We will explore ideas in a (very) simple setting
- Coherent channel model

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{v}$$

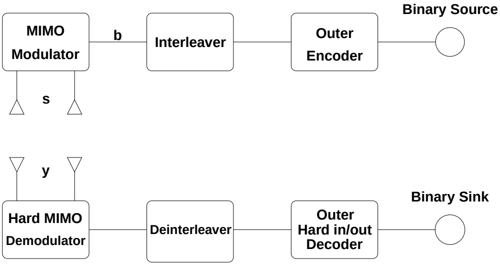
  with $\mathbf{v}$ being AWGN
- QPSK signalling

$$s_i \in \{\pm 1 \pm j\}$$

- Extensions to higher-order QAM discussed later

Outline

Hard MIMO Demod



- $\mathbf{y} = \mathbf{Hs} + \mathbf{v}$, with $\mathbf{s} = \mathcal{M}(\mathbf{b})$
- With QPSK signalling, hard demod solves

$$\hat{\mathbf{b}}_{\mathrm{opt}} = \arg \min_{\mathbf{b} \in \{\pm 1\}^{2M}} \|\mathbf{y} - \mathbf{Hs}\|_2$$

- Comp. cost is exp. in M due to interference
- Goal of hard MIMO demod: *Find a "good" approximation to $\hat{\mathbf{b}}_{\mathrm{opt}}$ at a reasonable computational cost*

Hard MIMO Demod

- Solve or approx: $\arg\min_{\mathbf{b}\in\{\pm 1\}^{2M}} \|\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{b})\|_2$
- **Linear:** (regularized) inversion of **H**;
  scalar decision on each element
- **MMSE-DFE:** Use QR decomposition (re-ordered) **H**:

$$\arg \min_{\mathbf{b}\in\{\pm 1\}^{2M}} \|\check{\mathbf{y}} - \mathbf{R}\mathcal{M}(\mathbf{b})\|_2 \tag{1}$$

  Make scalar decisions sequentially
- **"Sphere decoding":** Search inherent tree-structure of (1);
  e.g., Breadth-first; Depth-first; Best-first;
  Optimal, but hard, even on average
  Tree-search termination yields suboptimal variants
- **Semidefinite relaxation:** Rewrite as:

$$\arg \max_{\tilde{\mathbf{b}}\in\{\pm 1\}^{2M+1}} \tilde{\mathbf{b}}\tilde{\mathbf{Q}}\tilde{\mathbf{b}}$$

  Apply standard semidefinite relaxation techniques

Properties

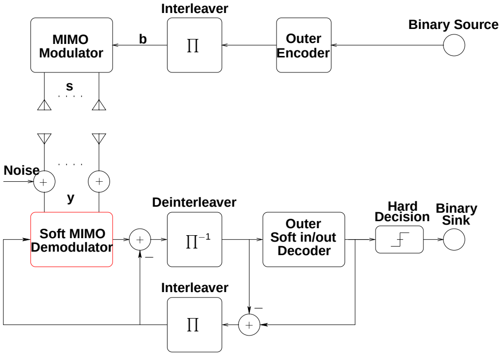
- Linear:
  - Cheap, but poor performance when **H** ill-cond.
- MMSE-DFE:
  - Usually better performance, but
  - Error propagation, ordering issues.
  - Problems remain on ill-conditioned channels
- Sphere decoder:
  - Optimal, but hard, even on ave.
  - Heavy tail on complexity distribution
  - Ave complexity is pretty good for reasonable SNRs
  - Early termination can yield good performance-complexity trade-offs
- Semidefinite relaxation:
  - Good performance; polynomial cost;
  - Bulk of complexity distrib. of sphere decoder typically lies below complexity of SDR

Hard MIMO Demod.

- Despite demodulation effort performance falls well below that suggested by capacity

- Observe that hard demodulation destroys “reliability” information
- Suggests soft demodulation
- If prior information available (via a previous decoding iteration) that can impact “reliability”
- Many approaches to soft demod. exploit aspects of hard demodulation

MIMO BICM with Iterative Soft Receiver



- ‘Turbo’ iterative receiver:
  - suboptimal, but common pragmatic choice.
- Soft MIMO Demodulator
  - Computes posterior log likelihood ratio of each bit.
  - Exploits the memoryless nature of the channel

MIMO Soft Demodulation

- $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{v}$, with $\mathbf{s} = \mathcal{M}(\mathbf{b})$
- Post. log likelihood ratio:

$$\lambda_i \triangleq \log \frac{P\{b_i = +1|\mathbf{y}\}}{P\{b_i = -1|\mathbf{y}\}}$$

- Bayes Theorem:

$$\lambda_i = \log \frac{\sum_{\mathbf{b}\in\mathcal{L}_{i,+1}} p(\mathbf{y}|\mathbf{b})p(\mathbf{b})}{\sum_{\mathbf{b}\in\mathcal{L}_{i,-1}} p(\mathbf{y}|\mathbf{b})p(\mathbf{b})}.$$

- $\mathcal{L}_{i,+1}$ contains all binary vectors with $+1$ in ith position. Comp. cost is exponential in # bits per channel use

MIMO Soft Demodulation, cont.

- We wish to find:

$$\lambda_i = \log \frac{\sum_{\mathbf{b}\in\mathcal{L}_{i,+1}} p(\mathbf{y}|\mathbf{b})p(\mathbf{b})}{\sum_{\mathbf{b}\in\mathcal{L}_{i,-1}} p(\mathbf{y}|\mathbf{b})p(\mathbf{b})}.$$

- In additive white Gaussian noise,

$$\lambda_i = \log \frac{\sum_{\mathcal{L}_{i,+1}} \exp(-D(\mathbf{b})/(2\sigma^2))}{\sum_{\mathcal{L}_{i,-1}} \exp(-D(\mathbf{b})/(2\sigma^2))},$$

where $D(\mathbf{b}) \triangleq \|\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{b})\|_2^2 - 2\sigma^2 \log(p(\mathbf{b}))$

- with good interleaving, $p(\mathbf{b}) \simeq \prod p(b_i)$, which can be easily estimated from decoder output at prev. iter.

Approximation by Hard Demod.

$$\lambda_i \simeq \log \frac{\sum_{\mathcal{L}_{i,+1}} \exp(-D(\mathbf{b})/(2\sigma^2))}{\sum_{\mathcal{L}_{i,-1}} \exp(-D(\mathbf{b})/(2\sigma^2))},$$

- “max-log” approximation

$$\lambda_i \approx \frac{1}{2\sigma^2}\Big(\min_{\mathbf{b}\in\mathcal{L}_{i,-1}} D(\mathbf{b}) - \min_{\mathbf{b}\in\mathcal{L}_{i,+1}} D(\mathbf{b})\Big),$$

  - 2 length($\mathbf{b}$) hard demodulation problems, each of size length($\mathbf{b}$) $-$ 1
  - Tree search (sphere decoding): Jalden, Giannakis
  - Semidefinite relaxation (for QPSK): Steingrimsson *et al.*
  - These approaches have rather different complexity characteristics
  - In raw form, still quite expensive
  - Some attempts to modify tree search to approx. both problems

List Approximation

$$\lambda_i \simeq \log \frac{\sum_{\mathcal{L}_{i,+1}} \exp(-D(\mathbf{b})/(2\sigma^2))}{\sum_{\mathcal{L}_{i,-1}} \exp(-D(\mathbf{b})/(2\sigma^2))},$$

- List approximation:
  - Determine a list $\hat{\mathcal{L}}$ containing dominant components of LLRs,
  - insert in either above or max-log approx thereof
  - modify likelihood to account for approx; e.g., “clip”
- Existing approaches
  - Tree search: Hochwald & ten Brink; Vikalo *et al*; Hagenauer *et al*
  - Have proved to be quite effective
- Can we develop a competitor based on SDR?

Semidefinite relaxation

- Real-valued model for QPSK: $\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\mathbf{b} + \tilde{\mathbf{v}}$.
- $D(\mathbf{b}) = \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{b}\|_2^2 - \sigma^2 \boldsymbol{\lambda}_A^T \mathbf{b} = \tilde{\mathbf{b}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{b}}$,

  where $\tilde{\mathbf{b}}^T = [c\mathbf{b}^T, c]$, $c \in \{\pm 1\}$, $\lambda_{A,i} = \log\left(\frac{p(b_i=+1)}{p(b_i=-1)}\right)$
- This is a boolean quadratic. Hence max/min is NP hard
- Semidefinite relaxation (Lovász & Schrijver, 1990): efficient generation of approx. solns via convex SDP
- Goemans and Williamson (1996), Nesterov (1997): Using randomization, w.p.$\rightarrow$1 SDR generates a solution within a certain constant factor of optimal
- Also: Luo *et al*, *IEEE Signal Process. Mag.*, May 2010

The SDR trick

- $\min_{\tilde{\mathbf{b}} \in \{\pm 1\}^{2M+1}} \tilde{\mathbf{b}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{b}}$
- Setting $\mathbf{X} = \tilde{\mathbf{b}}\tilde{\mathbf{b}}^T$ this can be rewritten as

$$\begin{aligned} \min_{\mathbf{X}} \quad & \text{Trace}(\mathbf{X}\tilde{\mathbf{Q}}) \\ \text{s.t.} \quad & \mathbf{X} \succeq \mathbf{0} \\ & \mathbf{X}_{ii} = 1 \\ & \text{rank}(\mathbf{X}) = 1 \end{aligned}$$

- Dropping the rank 1 constraint yields an SDP
- That SDP can be efficiently solved: $O(M^{3.5} \log \epsilon^{-1})$
- This SDP is a relaxation.
- Must project back to the feasible set.
- Projection from real semidefinite matrix to binary vector

Randomization

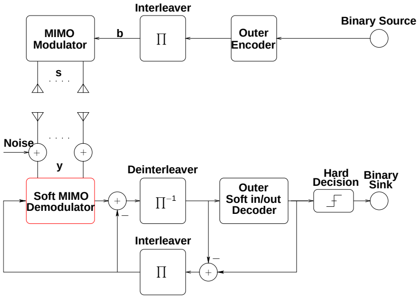
- Let $\mathbf{X}_{\text{opt}} = \mathbf{V}^T\mathbf{V}$ be a Cholesky decomposition of the solution of SDP.
- Generate random vectors $\mathbf{u}$ uniformly on the unit circle.
- Compute $\tilde{\mathbf{x}} = \text{sign}(\mathbf{V}^T\mathbf{u})$,
- Let $\mathbf{x}$ be the all except the last element of $\tilde{x}_{2M+1} \times \tilde{\mathbf{x}}$
- Repeat, retaining the bit vector $\mathbf{x}$ with the smallest $D(\mathbf{x})$ as the current estimate of $\mathbf{b}_{\text{opt}}$

Analysis of randomization [GW96,N97]

- $\tilde{\mathbf{x}} = \text{sign}(\mathbf{V}^T\mathbf{u})$, and $\mathbf{x} = \tilde{x}_{2M+1} \times \tilde{\mathbf{x}}_{1:2M}$
- $E\{x_i\} = \frac{2}{\pi}\arcsin(\mathbf{v}_i^T\mathbf{v}_{2M+1})$.
- Furthermore, can calculate $E\{D(\mathbf{x})\}$.
- In the worst case this expectation lies at least 4/7 of the way from the worst solution to the optimal solution.
- Probability that a randomization procedure fails to do better than the expected value decays rapidly with number of randomizations

- **Insight:** Some of the vectors generated by the randomization procedure are “good”
- Can this insight be used to generate an efficient list demodulator?

SDR Soft Demodulators



- For QPSK signalling,
  $D(\mathbf{b}) = \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{b}\|_2^2 - \sigma^2 \boldsymbol{\lambda}_A^T \mathbf{b} = \tilde{\mathbf{b}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{b}}$,
- List approach:
  - identify bit vectors $\mathbf{b}$ for which $D(\mathbf{b})$ is small.
- Analysis of randomization suggests that some of the generated vectors will have small values.

SDR Soft Demodulator I

At each demodulation iteration,

- obtain $\boldsymbol{\lambda}_A$ from decoder, and compute $\tilde{\mathbf{Q}}$
- Solve the SDP
- Run the randomization procedure, store all generated bit-vectors as the preliminary list
- Enrich that list by adding all other vectors within Hamming distance 1
  - mitigates the incomplete list problem implicit in list approximations.
- Use this list to perform a list approx of the LLRs.

SDR Soft Demodulator I

Advantages

- Only one SDP per demodulation iteration;
  SDR-based hard demod. approach requires 2M+1
- Computational cost is polynomial in the worst case;
  tree-search approaches are not
- SDP requires $O(M^{3.5})$;
  randomization only requires $O(M^2)$ per vector

Unfortunately, still not competitive in typical instances

What can we do?

- List-SDR solves one SDP per demodulation iteration.
  Can we reduce to one SDP per channel use?
- How accurately do we need to “solve” the SDP?

SDR Soft Demodulator II

Let's examine the use of randomization in more detail

- Recall, randomization is being used to generate elements of the list.
- The LLRs are then approximated using the list.
- Randomization: $x_i = \text{sign}(\mathbf{v}_{2M+1}^T \mathbf{u})\ \text{sign}(\mathbf{v}_i^T \mathbf{u})$
  $\mathbf{u}$ uniformly distrib. on unit sphere
- Elements are correlated
- $E\{x_i\} = \frac{2}{\pi} \arcsin(\mathbf{v}_i^T \mathbf{v}_{2M+1})$.
- Approx. randomization using independent Bernoulli's with this mean
- No performance guarantee, but much simpler.
- However, still requires one SDP per demodulation iteration

SDR Soft Demodulator II, cont.

- Extrinsic information from channel ($\boldsymbol{\lambda}$) and decoder ($\boldsymbol{\lambda}_A$) are independent
- Hence, use one SDP to generate Bernoulli probs from the channel.
- At subsequent demodulation iterations, compute

$$\boldsymbol{\lambda}_B = \boldsymbol{\lambda} + \boldsymbol{\lambda}_A$$

- Regenerate list via independent Bernoulli trials ($\boldsymbol{\lambda}_B \to \mathbf{m}_B$)
- Now only one SDP per channel use.

Comp. Costs of SDR Demods

- M: number of transmit antennas
- P: number of demodulation iterations
- ϵ: measure of accuracy of the SDP solution
- K: number of randomizations

Cost of SDPs, randomizations, and computation of $D(\mathbf{b})$

- **Hard decision SDR approach:** $2M + 1$ SDPs per iter.

$$O(PM^{4.5} \log \epsilon^{-1}) + O(PKM^3) + O(PKM^3)$$

- **SDR Soft Demod I:** One SDP per iteration; regular randomization

$$O(PM^{3.5} \log \epsilon^{-1}) + O(PKM^2) + O(PKM^3)$$

- **SDR Soft Demod II:** One SDP per channel use; Bernoulli randomization

$$O(M^{3.5} \log \epsilon^{-1}) + O(PKM) + O(PKM^3)$$

Computational comparisons

- Single SDR: (Dominant term)

$$O(M^{3.5} \log \epsilon^{-1})$$

- Tree-search based:
    - not polynomially bounded
    - average for small problems at moderate SNRs:

$$\sim O(PM^4)$$

- MMSE Soft Interference Canceller.
    - Good performance-complexity trade-off in related application (CDMA MUD)

$$O(PM^4)$$

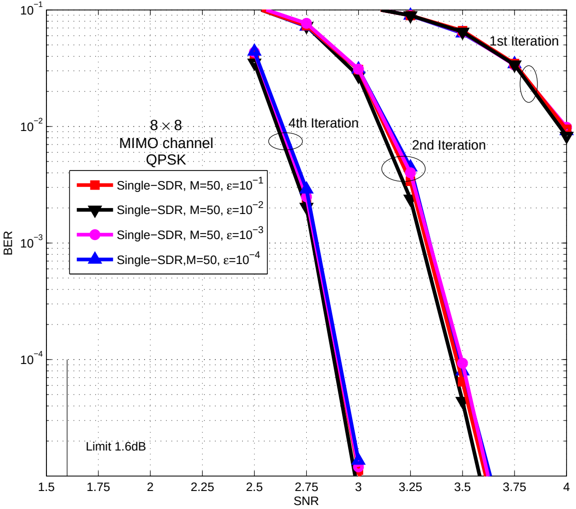
- **Note:** In currently envisioned systems, dimensions are moderate. Constants cannot be ignored

Simulation set up

- Same as that of Hochwald and ten-Brink
- 8×8 iid Rayleigh block fading channel
- V-BLAST transmission with QPSK symbols
- Standard rate 1/2 punctured Turbo code, with (5,7) convolutional codes as constituent
- BCJR based iterative outer decoding; 8 turbo iterations per demodulation

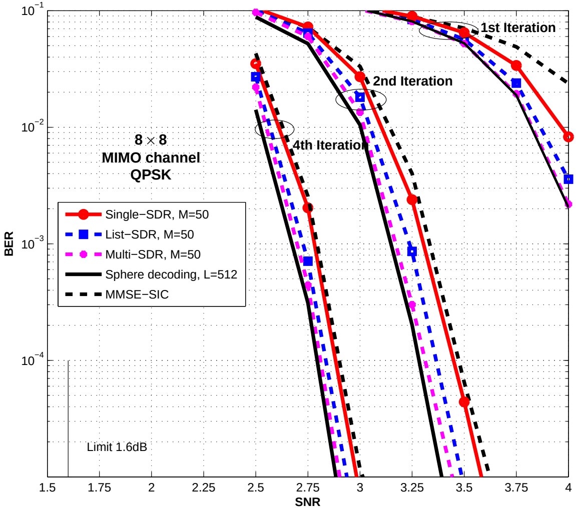
- M will denote number of randomizations

How accurately to solve SDP?

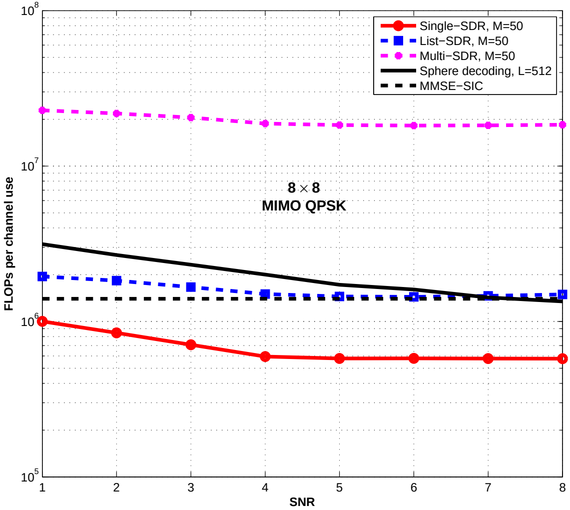


$\epsilon = 10^{-2}$ enough; $\epsilon = 10^{-4}$ too much

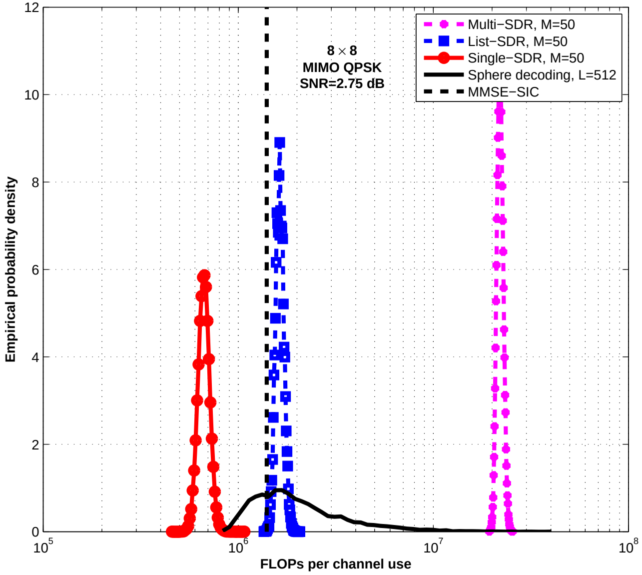
BER Performance



Average Computational Cost



PDF of Computational Cost



Interim Conclusion

- In the case of QPSK signalling, one can modify the SDR randomization procedure to obtain a competitive soft MIMO demodulator.
    - (Slightly) superior trade-off between ave performance and ave computational cost than that of sphere decoder
    - Lighter tail of complexity distribution; implies that it is easier to provision processor to reduce “computational outage”
- However, there is still much to examine

However, . . .

- Proposed signalling schemes use higher-order QAM at higher SNRs; e.g., 16-QAM, $s_i \in \{\pm 1, \pm 3\} + j\{\pm 1, \pm 3\}$
- Tree-search algo's extend directly to trees with higher-order nodes;
- MMSE-SIC receiver also easy to extend
- Extending SDR? several options
  - can use additional polynomial constraints to capture constellation structure
    - good accuracy, but loose cheap SDP algo
    - performs well, but expensive
  - can relax on extreme values
    - less accurate; recall we only use SDP to generate list
    - retains cheap SDP algo
    - remains competitive, with similar desirable properties

However, . . .

- What is the appropriate notion of computational cost?
- In this talk, we discussed flops
- Reasonable for a general purpose processor
- Current systems tend to use ASICs
- Tree search maps well onto an ASIC
- However, in future “software radios” that can implement multiple standards, metrics for general purpose processors may become relevant

However, . . .

- We have assumed coherent demodulation with perfect knowledge of the channel matrix $\mathbf{H}$
- In practice, $\mathbf{H}$ is identified through training
- System designers must trade training resources against real communication resources
- How sensitive are the various methods to imperfect or out-of-date channel estimates?
- Should we "interpolate" channel between training phases, or is it enough to assume block-wise constant channel?

However, . . .

- One of our achievements is to reduce cost to one SDP per channel use
- However, when block-wise constant channel model is used, other techniques, incl. MMSE-SIC and tree-search, can amortize the cost of some pre-processing over several channel uses
- Can we exploit warm-start techniques to amortize some of the cost over several channel uses?

However, . . .

- We have considered a narrowband (memoryless) channel
- In most current proposals, channel is broadband
- Most proposals involve multi-carrier transmission multiple parallel narrowband subcarriers, e.g., 64 or 256
- Channels on neighbouring subcarriers are correlated
- How can we amortize computational effort over subcarriers?
- On-going issue for all methods

Conclusion

- Modification of SDR yields a competitive soft demodulator
- There is still much to be done to determine whether this technique can be promoted as the technique of choice
- Despite the fact that this is not the first technique “on the market”, the current assessment suggests that it has several desirable properties
- Since receiver structure is typically a choice of the (chip) manufacturer, and not directly specified in a standard, many opportunities on the table