Iterative detection and decoding to approach MIMO capacity

Jun Won Choi

In this term project, I studied on MIMO receiver techniques which can achieve near-capacity performance by applying an iterative detection and decoding (IDD) process. This report will provide the brief summary on what I have learned in this project, which will supplement the presentation material. The responses to the list of questions made during the presentation were also included in this report.

- Explanatory Notes

In this section, I will present some supplementary explanations on the topic I studied for the term project.

1. What kind of transmit architecture can approach the MIMO capacity?

    - Assume that channel state information is known only at the receiver. It was proved in [1] and [3] that, for fast fading, the V-BLAST architecture can approach MIMO capacity when combined with minimum mean square error-successive interference cancellation (MMSE-SIC) technique. However, this is not the case for slow fading. Since there always exist non-zero probability that a channel state is in outage, the concept of MIMO capacity is somewhat obscure. Instead, outage capacity is widely used as a maximum achievable rate wherein an outage probability is less than ϵ. The relationship between outage capacity and outage probability can be translated to multiplexing-diversity tradeoff. Various space-time coding techniques such as orthogonal codes [4], D-BLAST [3], and TURBO-BLAST [7] have been proposed to minimize outage probability given transmission data rate by coding across transmit antennas. Among them, I described two transmit architectures designed to take advantage of IDD process to approach near-optimal performance, which are the space time bit-interleaved coded modulation (STBICM) [6] and the TURBO-BLAST [7]. Both techniques employ the concatenation of AWGN channel encoders and random (interleaved) space-time multiplexing for data transmission. Note that the use of interleaver allows the receiver to perform MIMO detection and channel decoding independently.

2. Which receiver architecture is optimal in performance?

    - From now on, I assume that the STBICM transmit architecture is employed at transmitter side. Assuming that all sequences of information bits are equally likely, the optimal receiver would be a join MAP decoder, which maximizes the probability of sequence of information bits given received signal. It may be implemented by first modeling a receiver signal as Markov process whose trellis includes channel encoder, modulator and

MIMO channel and then finding a sequence of information bits which minimizes the MAP probability. It can be easily shown that the joint MAP decoder leads to minimum sequence error probability. However, it requires huge amount of computational complexity since the number of states (in the trellis) increases exponentially with constraint length (of channel code), modulation order, and number of transmit antenna. High complexity of joint MAP decoder motivates the development of iterative receiver which separates the MIMO detection and channel decoding tasks and performs both tasks in an iterative fashion for low complexity. Such IDD receiver yields significant performance gain compared to separate detection and decoding without any iterations.

3. How does IDD process work?

    - The key mechanism of the IDD process is the information exchange between MIMO detector and channel decoder, leading to successive performance improvement. They exchange soft information, which has a form of log likelihood ratio (LLR) of a certain bit. The illustration of IDD process is included in [5] and [8] in detail. First, the MIMO detector processes the received signal and the soft information delivered from the channel decoder to obtain the LLRs of all coded bits (called extrinsic information). Such extrinsic information is delivered to the SISO channel decoder through deinterleaver. The extrinsic information is seen as a priori information at the side of the channel decoder. Based on such a priori information, the channel decoder computes the LLRs of coded bits, which form another extrinsic information, which can be used for better MIMO detection. Such extrinsic LLRs are interleaved and fed back to the MIMO detector as a priori information. The procedure mentioned so far completes one cycle of iteration and the iterations continue until a desired performance is met.

4. What are the examples of iterative receiver algorithm using IDD?

    - In the presentation, I described three kinds of SISO MIMO detector. First one was the MAP MIMO detector, which obtains the LLRs of coded bits directly (Note that the MAP MIMO detector is different from the joint MAP decoder I mentioned for optimal receiver). The complexity of MAP MIMO detector is exponential with modulation size and number of transmit antenna. Second one was the list sphere decoding technique, which reduces the complexity of MAP MIMO detector [8]. The basic idea of list sphere decoding technique is to reduce the size of summation involved for computation of LLR values by restricting all symbol candidates to ones inside a sphere centered at observed data. Hence, we only count highly likely symbol candidates (in the sense of ML criterion) to reduce the computations. Note that list sphere decoding is different from conventional sphere decoding [2] in that it finds multiple closest points inside a sphere rather than one closest point. Clearly, the size of candidate list is determined of a radius of sphere, offering the performance and complexity trade-off. Third MIMO detection technique provided in the presentation is TURBO-BLAST detector. Its complexity is much simpler than the previous two methods since it employs MMSE estimation with linear structure. The TURBO-BLAST detector tries to minimize the MSE between the desired symbol to be esitmated and the estimator output. Such MMSE symbol estimation can be divided into two steps: 1) interference cancellation step and 2) interference nulling step. Given a priori probabilities of each coded bit, both the interference cancellation and nulling steps can be derived. By using the symbol estimates obtained by the MMSE symbol estimator, the LLRs of coded bits can be computed in a straightforward manner, as shown in the presentation. Note that the complexity of computing an extrinsic LLRs increases only with modulation order. Hence, the complexity of TURBO-BLAST detector is reduced substantially compared to the MAP and list sphere decoding methods.

Responses to the Questions Made in the Presentation

1. (From Sreekanth) You have prepared MAP decoder, list sphere decoder, and linear MMSE detector. How does complexity trade-off with the performance in the above receiver designs, i.e., when do you pick a particular design?

   - There exists performance and complexity trade-off in the choice of one among the algorithms I presented. As I mentioned, the MAP MIMO dectector is most complex but best performed technique. The list sphere decoding reduces the complexity of MAP detector by sacrificing the performance. The complexity of linear MMSE detector (TURBO-BLAST detector) is simplest but its performance is much worse than the MAP detector and list sphere decoding.

2. (From Guanfeng Liang) What is the delay of the IDD?

   - The delay of the IDD might be determined by the block size and how many iterations you executed. Considering that many present wireless communication systems are employing turbo code which requires block by block processing for decoding, I expect that the IDD can also be realized with reasonable delay in the application with loose delay requirement.

3. (From Abdullah) • In turbo-blast decoder, is there a drawback for ignoring prior information? • What is the complexity of IDD? How does it compare to ML?

   - Yes, definitely. If we ignore the priori information, no information exchange occurs between the MIMO detector and channel decoder. This leads to separate detection and decoding, which causes loss of performance gain that could be achieved by joint detection and decoding.
   - The ML decoder searches all possible combinations of symbol vector to find an ML solution. Hence, it suffers from high complexity exponentially increasing with modulation order and transmit antenna size. Based this observation, the ML decoder has same order of complexity compared to the MAP detector. However, due to the feature of iterative receiver, the MAP detector needs the computational overhead for computing LLR values and for repeating the same detection procedure at every iteration.

4. (From Hoa) • What is the trade-off between number of iteration and performance? • What is the relation between frame length and block length?

   - As we increase the iteration number, it can achieve better performance. But it will cause increasing delay and computational complexity. How much performance can be improved with iteration can be interpreted by extrinsic information transfer (EXIT) chart [9] analysis. According to that, the performance improvement get marginal upon certain number of iterations. Hence, the iteration number should be determined based on the analysis on performance limit and complexity requirement.

- Commonly, block length is the same as frame length. Frame length is the certain number of channel uses over which the corresponding received data is processed block-by-block in an iteratively fashion. It is common that a frame length is set to a few hundred to thousand depending on a delay constraint. Note that larger frame size leads to better performance.

5. (From Rami) • How would performance compares between iterative MIMO decoding and those without iterative technique, just one shot detection? • What will happen with orthogonal coding?

   - Clearly, the joint detection and decoding always yields better performance than the separate detection and decoding. The IDD process tries to approach the performance of joint detection and decoding via iterative receiver algorithm. Intuitively, the MIMO detector can do better job by exploiting the information produced by channel decoder. It is shown in [8] that the performance gain of the IDD could be more than 5 dB in the case of 8-by-8 QPSK MIMO system.
   - Orthogonal space-time coding transforms MIMO channels into separate parallel single input single output channels. Hence, it is unnecessary to suppress interferences from other transmit antennas in such case. Since each sub-channel can be considered as single input single output fading channel, only diversity techniques over parallel channels can be employed with AWGN channel decoder.

6. (From Vijay) • What is the complexity of this system? Will the scheme converge? • How do the detection schemes affect the complexity + the convergence properties?

   - The MAP MIMO detector has the complexity increasing exponentially with modulation order and transmit antenna size. The list sphere decoding reduces this complexity by computing a LLR over the candidate list inside a sphere. The complexity of TURBO-BLAST detector is much smaller in that it increases only with modulation order.

     According to EXIT chart analysis, the performance consistently improves with iterations but the improvement get marginal as the number of iteration exceeds a certain level. Hence, we can tell it converges.
   - Clearly, three MIMO detectors offer trade-off in terms of performance, complexity, and convergence speed. It can be shown via EXIT chart analysis that the MAP detector is fastest and TURBO-BLAST detector is slowest. On the other hand, the MAP detector is worst and TURBO-BLAST detector is the best in terms of computational complexity.

7. (From Ram) • Is there some performance comparison for the different schemes you proposed?

   - Yes, definitely. The MAP detector is optimal in terms of performance in the sense that it computes LLRs of coded bits directly. On the contrary, the list sphere decoding is sub-optimal since it reduces all possible combination of symbols into the set of highly likely symbol combinations. Note that the performance loss of the list sphere decoding reduces as the candidate size gets larger. The performance of TURBO-BLAST scheme is worst compared to previous two schemes since linear MMSE estimator is used to simplify the complexity needed for LLR computations.

References

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels,". *European Transactions on Telecommunications*, vol. 10 pp. 585-596, Nov. 1999.

[2] M. O. Damen, H. E. Gamel, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Information Theory*, vol. 49, pp. 2389-2402, Oct. 2003.

[3] J. G. Foschini, "Layered space-time architecture for wireless communication in a facing environment when using multi element antennas," *Bell Labs Tech. J.*, vol. 2, pp. 41-59, 1996.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance analysis and code construction," *IEEE Trans. Inform. Theory,* vol. 44, pp. 744-765, Mar. 1998.

[5] X. Wang and H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1046-1061, July 1999.

[6] A. M. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy," *Proc. IEEE Veh. Technol. Conf.*, pp. 473-478, Sept. 2000.

[7] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: theory and experiments," *IEEE Trans. Signal Processing*, vol. 50, pp. 2538-2546, Oct. 2002.

[8] B. Hochwald and S. T. Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, pp. 389-399, Mar. 2003.

[9] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 40, pp. 1727-1737, Oct. 2001.

[10] D. Tse and P. Viswanath, "Fundamentals of wireless communication," *CAMBRIDGE University Press*, 2005.

Iterative detection and decoding to approach MIMO capacity

Jun Won Choi

Introduction

- MIMO capacity (CSI only at receiver)
  - Fast fading scenario – Ergodic I.I.D. Rayleigh fading Channel

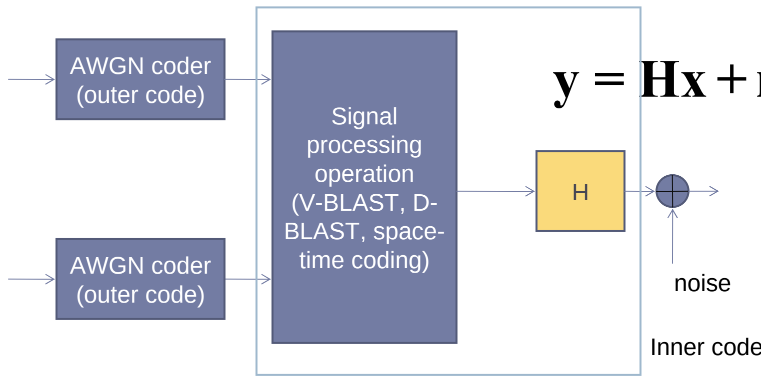
$$E\left[\log\det\left(I_r + \frac{P}{t}\mathbf{H}\mathbf{H}^H\right)\right]$$

Teletar, 1999

    - Under fast fading assumption, transmission of independent data stream with same power is sufficient to achieve capacity (V-BLAST).
    - Capacity achieving Gaussian codes are used at each antenna as outer code.
  - Slow fading scenario
    - Coding across transmit antennas is needed → space-time coding, advanced layering

Introduction

- Optimal transmitter structure



$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

Coding across transmit antennas is needed in slow fading

Introduction

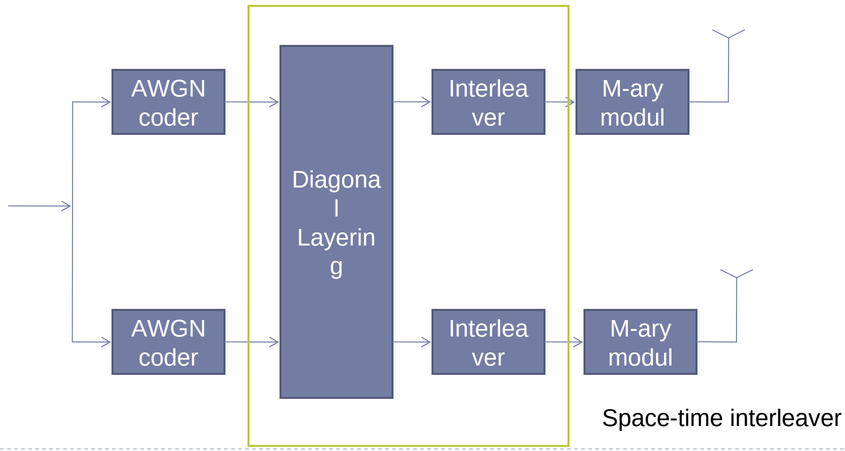
- Optimal receiver structure
  - Maximum a posteriori (MAP) decoder
    - Model a received signal as Markov process whose trellis is formed to include AWGN code, space-time coding, and MIMO channel.
    - Map decoding rule is optimal.

$$b_i = \arg\max P\left(b_i \mid \mathbf{y}_1, \cdots, \mathbf{y}_T\right)$$

    - Computationally infeasible !
  - Iterative detection and decoding (IDD)
    - Divide decoding job into MIMO detection (inner code) and AWGN channel decoding (outer code).
    - Approximately approach to optimal performance via information exchange between two constitutional blocks.

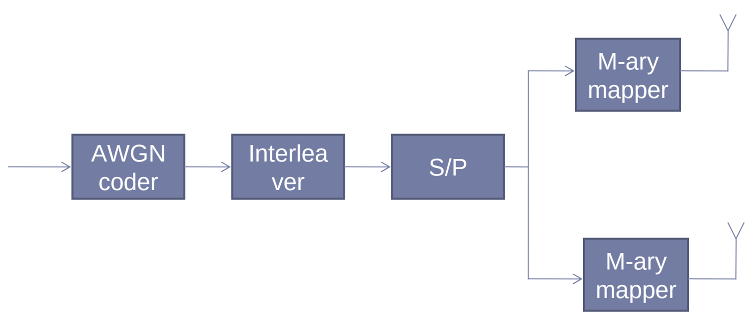
Transmitter design example 1 (IDD)

- Turbo-Blast (Haykin 2002)
  - Random layered space time coding



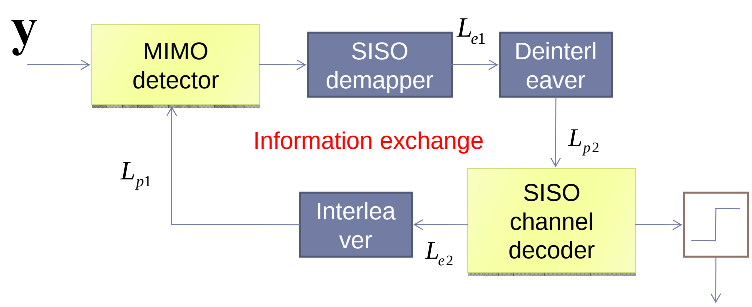
Transmitter design example 2

- Space-time bit interleaved coded modulation (Tonello, 2000)



Principle of IDD

- Iterative (MIMO) detection and (channel) decoding



Soft information is expressed as L-value

$$L_p = \frac{p(x=+1)}{p(x=-1)}$$ Priori LLR

L_e Extrinsic LLR

IDD

- SISO Channel decoder
  - BCJR algorithm – based on trellis-based search
  - Low-complexity APP decoder - LOG-MAX algorithm, Soft output viterbi algorithm (SOVA)
- MIMO detector
  - Complexity and performance trade-off
    - MAP versus Sub-optimal detector with linear structure

Definition (Space-time bit interleaved coded modulation)

b_i : information bit

c_i : coded bit

$\tilde{c}_{n,m}$: interleaved coded bit

$x_n : M_c - ary$ symbol

$$c_{n,1}, \cdots, c_{n,M_t} \rightarrow x_n$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{N_t} \end{bmatrix}$$

Map detector

- Map detector
  - A posteriori L-value of the bit

$$L_{a1}\left(\tilde{c}_{n,m} | \mathbf{y}\right) = \ln \frac{P\left[\tilde{c}_{n,m} = 1 | \mathbf{y}\right]}{P\left[\tilde{c}_{n,m} = -1 | \mathbf{y}\right]}$$

Extrinsic information (output)

$$p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{\left(\pi\sigma_w^2\right)^{N_r}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{Hx}\|^2}{\sigma_w^2}\right)$$

$$L_{e1}\left(\tilde{c}_{n,m} | \mathbf{y}\right) = L_{a1}\left(\tilde{c}_{n,m} | \mathbf{y}\right) - L_{p1}\left(\tilde{c}_{n,m}\right)$$

$$= \ln \frac{\sum_{\mathbf{x} \in X_{n,m,+1}} p(\mathbf{y} \mid \mathbf{x}) \exp\left(\frac{1}{2} \sum_{k=1(\neq n)}^{N_t} \sum_{l=1(\neq m)}^{M_c} \tilde{c}_{k,l} L_{p1}\left(\tilde{c}_{k,l}\right)\right)}{\sum_{\mathbf{x} \in X_{n,m,-1}} p(\mathbf{y} \mid \mathbf{x}) \exp\left(\frac{1}{2} \sum_{k=1(\neq n)}^{N_t} \sum_{l=1(\neq m)}^{M_c} \tilde{c}_{k,l} L_{p1}\left(\tilde{c}_{k,l}\right)\right)}$$

Map decoder

- Approximation
  - Log-Max approximation rule

$$\ln\left(e^{a1}+e^{a2}\right)\approx \max(a1,a2)$$

$$L_{e1}\left(\tilde{c}_{n,m}\,|\mathbf{y}\right)=\max_{\mathbf{x}\in X_{n,m,+1}}\left(-\frac{1}{\sigma_w{}^2}\|\mathbf{y}-\mathbf{Hx}\|+\frac{1}{2}\sum_{k=1(\neq n)}^{N_t}\sum_{l=1(\neq m)}^{M_c}\tilde{c}_{k,l}L_{p1}\left(\tilde{c}_{k,l}\right)\right)$$
$$-\max_{\mathbf{x}\in X_{n,m,-1}}\left(-\frac{1}{\sigma_w{}^2}\|\mathbf{y}-\mathbf{Hx}\|+\frac{1}{2}\sum_{k=1(\neq n)}^{N_t}\sum_{l=1(\neq m)}^{M_c}\tilde{c}_{k,l}L_{p1}\left(\tilde{c}_{k,l}\right)\right)$$

$$X_{n,m,+1}=\left\{\mathbf{x}:\tilde{c}_{n,m}=1\right\}$$

- Complexity
  - Complexity of MAP decoder is exponential in modulation size, antenna size.

There are $2^{M_t N_t}-1$ combinations for each hypothesis.

List sphere decoding

- Idea (Hochwald, 2003)
  - Find the combinations of symbol vector that are highly likely to be transmitted. It is called candidate list.

$$p\left(\mathbf{y}\,|\,\mathbf{x}\right)\propto -\|\mathbf{y}-\mathbf{Hx}\|^2$$

  - Define the candidate list, L as $L=\left\{\mathbf{x}:\|\mathbf{y}-\mathbf{Hx}\|<B\right\}$
  - Then, extrinsic L-value can be find over such candidate list, i.e.,

$$L_{e1}\left(\tilde{c}_{n,m}\,|\mathbf{y}\right)=\max_{\mathbf{x}\in X_{n,m,+1}\cap L}\left(-\frac{1}{\sigma_w{}^2}\|\mathbf{y}-\mathbf{Hx}\|+\frac{1}{2}\sum_{k=1(\neq n)}^{N_t}\sum_{l=1(\neq m)}^{M_c}\tilde{c}_{k,l}L_{p1}\left(\tilde{c}_{k,l}\right)\right)$$
$$-\max_{\mathbf{x}\in X_{n,m,-1}\cap L}\left(-\frac{1}{\sigma_w{}^2}\|\mathbf{y}-\mathbf{Hx}\|+\frac{1}{2}\sum_{k=1(\neq n)}^{N_t}\sum_{l=1(\neq m)}^{M_c}\tilde{c}_{k,l}L_{p1}\left(\tilde{c}_{k,l}\right)\right)$$

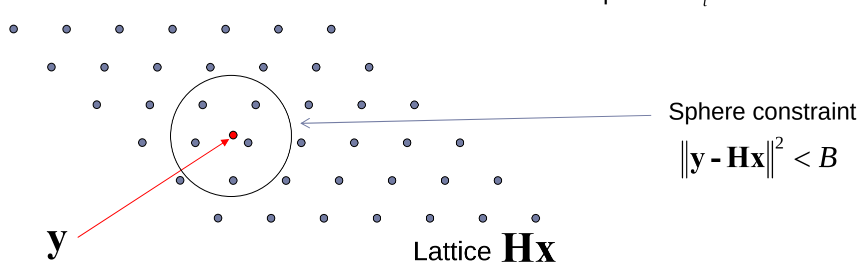
List sphere decoding

- List sphere decoding
  - Efficient tree pruning problem

Form skewed lattice

$$\mathbf{y} = \mathbf{Hx} + \mathbf{w}$$

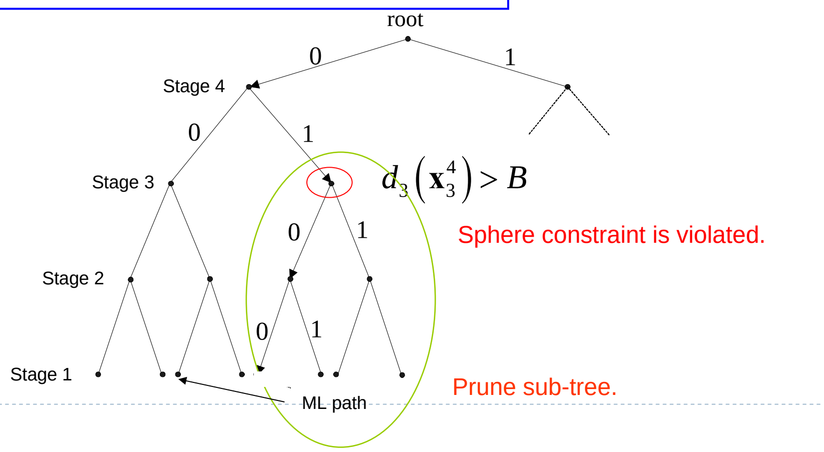
Number of points $M_t^{N_t}$



$$\|\mathbf{y} - \mathbf{Hx}\|^2 = \|\mathbf{y}' - \mathbf{Rx}\|^2 = d_{n_t}\left(\mathbf{x}_1^{n_t}\right) = \sum_{i=1}^{n_t}\left(y_i' - \sum_{k=i}^{n_t} r_{i,k} x_k\right)^2$$

Define the cost metric $d_q\left(\mathbf{x}_q^{n_t}\right) = \sum_{i=q}^{n_t}\left(y_i' - \sum_{k=i}^{n_t} r_{i,k} x_k\right)^2$

$$d_{q-1}\left(\mathbf{x}_{q-1}^{n_t}\right) = d_q\left(\mathbf{x}_q^{n_t}\right) + \left(y_{q-1}' - \sum_{k=q-1}^{n_t} r_{i,k} x_k\right)^2$$



List sphere decoding

- Procedure
  - 1. Find the points inside sphere by tree search.
  - 2. Select N_{cand} closest points. (when number of points found is larger than predefined list size)
  - 3. Increase radius and restart the search. (when number of points found is less than list size)
  - 4. If candidate list has no common entry with $X_{n,m,+1}$ or $X_{n,m,-1}$, the extrinsic info is set to –inf or inf depending on its sign of entries.
- How to choose B?, For true x

$$\|\mathbf{y} - \mathbf{Hx}\|^2 = \|\mathbf{w}\|^2 \sim \sigma_W^2 \chi_{2N_t}^2 \qquad p\left(\|\mathbf{w}\|^2 < B\right) = \mathrm{CDF}\left(B, N_t\right) = P_B$$

$$P_B = 0.9$$

Turbo-Blast detector

- Turbo-Blast detector
  - Sub-optimal detector with linear structure
  - Derive based on linear MMSE criterion

Assume that $L_{p1}(c_{[1:N_t],[1:M_t]})$ are available.

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}$$

Interference nulling step — Interference cancellation step

$$\hat{x}_n = E\left[x_n\right] + \mathrm{Cov}\left(x_n, \mathbf{y}\right)\mathrm{Cov}\left(\mathbf{y}, \mathbf{y}\right)^{-1}\left(y - \mathbf{H}E\left[\mathbf{x}_n\right]\right)$$

Let $L_{p1}(c_{n,[1:M_t]}) = 0 \rightarrow E\left[x_n\right] = 0$

$$\mathrm{var}\left(x_n\right) = 1$$

Turbo-Blast detector

- Interference cancellation step

$$\mathbf{z}_n = \mathbf{y} - \mathbf{H}E[\mathbf{x}] = \mathbf{y} - \mathbf{H}\begin{bmatrix} E[\mathbf{x}_{1:n-1}] \\ 0 \\ E[\mathbf{x}_{n+1:N_t}] \end{bmatrix}$$

$$E[x_k] = \sum_{x\in\Lambda} x \prod_{m=1}^{M_t} \frac{1}{2}\left(1 + \tilde{c}_{n,m} \tanh\left(L_{p1}(\tilde{c}_{n,m} \mid \mathbf{y})\right)\right)$$

- Interference nulling step

$$\hat{\mathbf{x}}_n = \mathbf{a}_n^H \mathbf{z}_n$$

$$\mathbf{a}_n = \left(\mathbf{H}\boldsymbol{\Lambda}_n\mathbf{H} + \sigma_w^2\mathbf{I}\right)^{-1} \mathbf{h}_n$$

$$\boldsymbol{\Lambda}_n = \mathrm{diag}\left(\left[\mathrm{var}(x_1),\cdots,\mathrm{var}(x_{n-1}),1,\mathrm{var}(x_{n+1}),\cdots,\mathrm{var}(x_{N_t})\right]\right)$$

$$\mathrm{var}(x_n) = \sum_{x\in\Lambda}\|x\|^2 \prod_{m=1}^{M_t}\frac{1}{2}\left(1+\tilde{c}_{n,m}\tanh\left(L_{p1}(\tilde{c}_{n,m}\mid\mathbf{y})\right)\right) - \|E[x_n]\|^2$$

Turbo-Blast detector

- Gaussian approximation

$$\hat{x}_n = \mathbf{a}_n^H\mathbf{z}_n = \mathbf{a}_n^H\left(H\begin{bmatrix}\mathbf{x}_{1:n-1} - E[\mathbf{x}_{1:n-1}] \\ x_n \\ \mathbf{x}_{n+1:N_t} - E[\mathbf{x}_{n+1:N_t}]\end{bmatrix} + \mathbf{w}\right) = \mu_n x_n + \lambda_n$$

Interference + noise term

$$\sigma_\lambda^2 = \mu_n - \mu_n^2$$

$$L_{e1}(\tilde{c}_{n,m} \mid \hat{x}_n) = \ln\frac{p(\hat{x}_n \mid \tilde{c}_{n,m} = 1)}{p(\hat{x}_n \mid \tilde{c}_{n,m} = -1)} = \ln\frac{\sum_{x\in X_{m,+1}} \exp\left(-\frac{\|\hat{x}_n - \mu_n x\|}{\sigma_\lambda^2}\right)}{\sum_{x\in X_{m,-1}} \exp\left(-\frac{\|\hat{x}_n - \mu_n x\|}{\sigma_\lambda^2}\right)}$$

$$= \max_{x\in X_{m,+1}}\left(-\frac{\|\hat{x}_n - \mu_n x\|}{\sigma_\lambda^2}\right) - \max_{x\in X_{m,-1}}\left(-\frac{\|\hat{x}_n - \mu_n x\|}{\sigma_\lambda^2}\right)$$

$$X_{m,+1} = \{x_n : \hat{c}_{n,m} = 1\}$$

There are only $2^{M_t} - 1$ combinations for each hypothesis.

Conclusions

- Capacity achieving MIMO architecture
  - Transmitter design
    - V-BLAST + AWGN code for fast fading
    - Coding across transmit antenna for slow fading → space time coding, D-BLAST, Treaded space time coding
  - Receiver design
    - Global MAP decoding
    - Iterative detection and decoding
      - Map decoding
      - List sphere decoder
      - Linear MMSE detector