Tracking and Equalizing Channels in Communication Systems and in Data Compression

Azadeh Vosoughi

University of Rochester

June 28, 2007

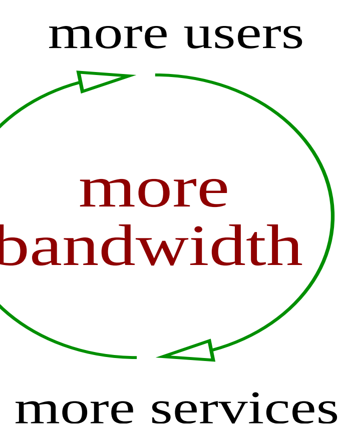
Outline

- Training design for estimating wideband wireless channels with antenna arrays at transmitter and receiver
    - Estimation theoretic perspective
    - Information theoretic perspective
- Distributed data compression for sensor networks

Unifying theme: Channel equalization

Motivation

- The increasing demand for ubiquitous wireless access generates a spiraling demand for bandwidth.
- Developing bandwidth efficient physical layer designs is critical.
- The designs need to be optimized for packet transmission.
- Advances in coding and modulation enable us to approach transmission limits, for a given communication channel.
- The challenging issue in wireless communication is channel estimation.

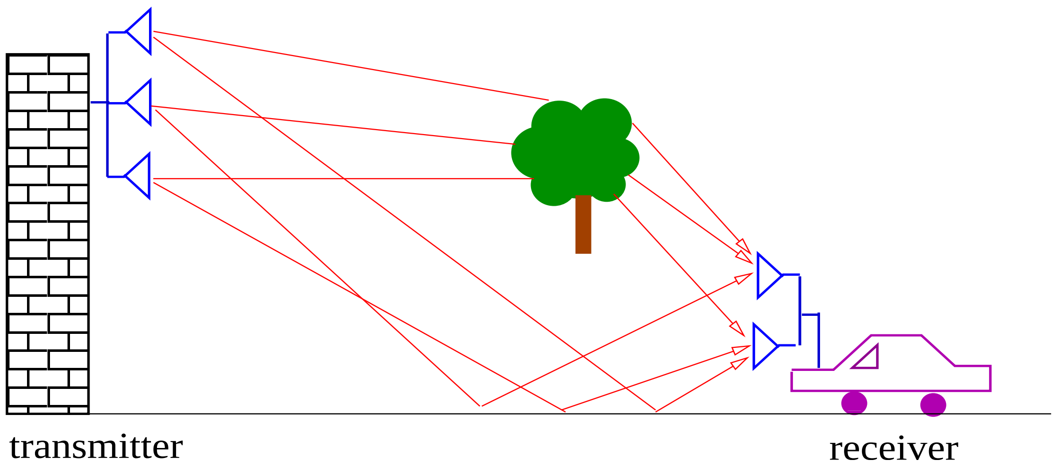


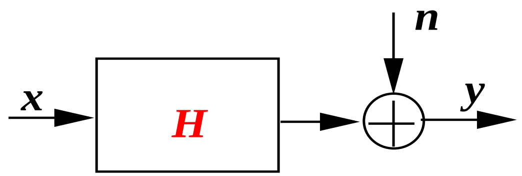
Channel Estimation in Wireless Communication Systems

- Unique characteristics: fading caused by multipath effect.
- The wireless channel is modelled as a linear filter.
- Discrete time representation of the channel input/output relationship:

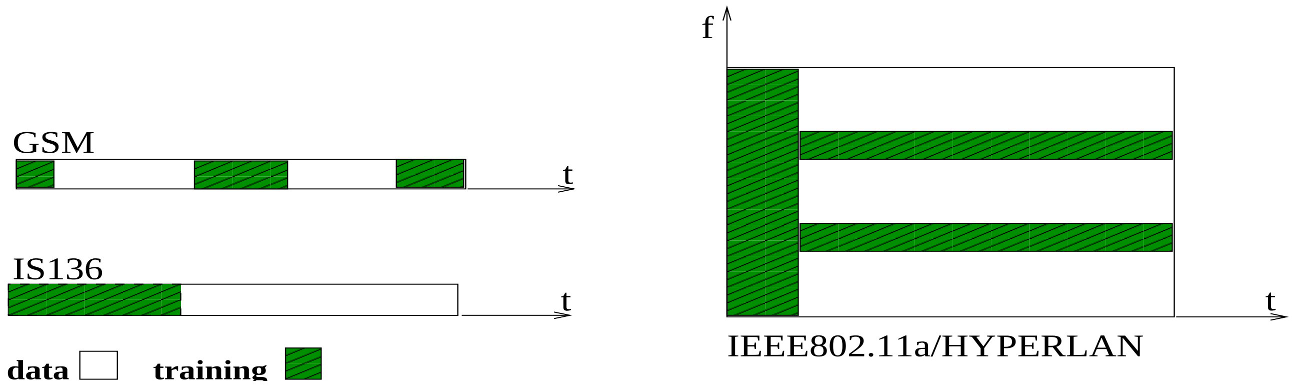
$$\boldsymbol{y}_i = \boldsymbol{H}\boldsymbol{x}_i + \boldsymbol{n}_i$$

- To enable channel estimation transmitter multiplexes data symbols $\boldsymbol{s}$ with known training symbols $\boldsymbol{t}$ $\Rightarrow$ $\boldsymbol{x}$ contains $\boldsymbol{s}$ and $\boldsymbol{t}$.
- Training symbols and the specific multiplexing scheme are known *a priori* at the receiver.





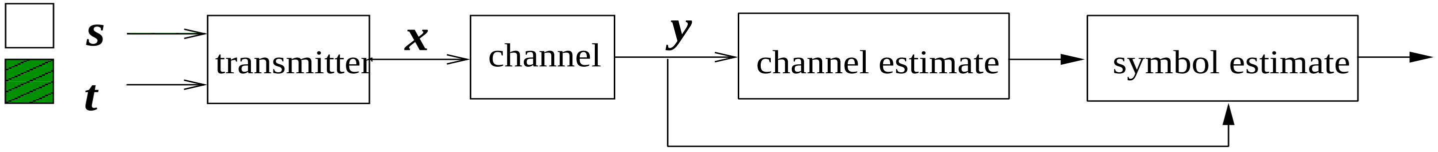
Training Design in Wireless Communication Systems



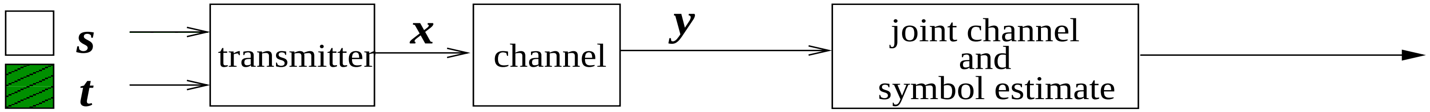
- Current training designs are heuristic.
- Orthogonality between training and data decouples channel and symbol estimation.
- Tradeoff: Training improves the system performance; however, it spends the resources that could have been used for transmitting data.
- Investigating optimal training designs is important.
- Related Work: literature considered optimal training designs, assuming training and data symbols are separated either in time or frequency.

Sequential versus Joint Channel and Symbol Estimation

Sequential Estimate



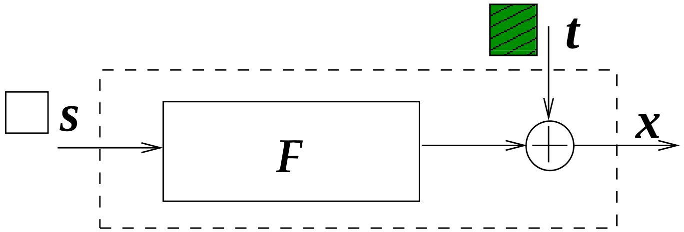
Joint Estimate



Questions

- Is separating training and data symbols for sequential estimation optimal?
- How is optimal training design for sequential estimation different from that of joint estimation?

Affine Precoding: A Unified Transmission Model



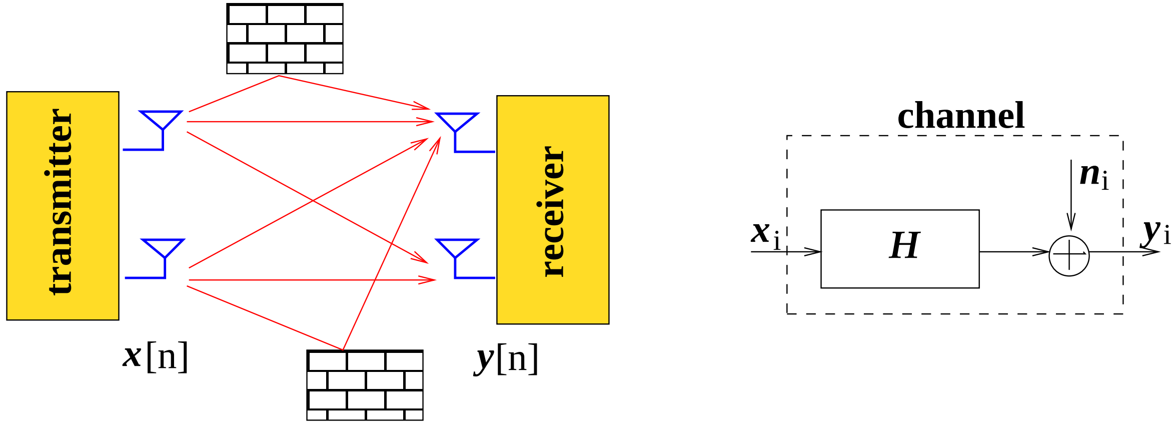
Affine Precoding

- $\boldsymbol{s} \sim \mathcal{CN}(\mathbf{0}, \sigma_{ss}^2 \boldsymbol{I})$ is a block of user's data and $\boldsymbol{t}$ is training vector.
- Affine precoding: $\boldsymbol{x}$ is linear precoded data $\boldsymbol{Fs}$ superimposed on $\boldsymbol{t}$.

$$\boldsymbol{x} = \boldsymbol{Fs} + \boldsymbol{t}$$

- Power constraint on transmit vector: $\mathcal{P} = E\{||\boldsymbol{x}||^2\} = \mathcal{P}_s + \mathcal{P}_t$.
- Affine precoding includes several multiplexing techniques:
  - GSM and IS136: columns of $\boldsymbol{F}$ are unit vectors.
  - IEEE 802.11a/HYPERLAN: $\boldsymbol{F}$ is FFT matrix.

Channel Model: Block Fading frequency Selective MIMO



$$\boldsymbol{y}_i = \boldsymbol{H}\boldsymbol{x}_i + \boldsymbol{n}_i$$

Assumptions

- K transmit and R receive antennas,
- Length of a transmission block is P, length of channel memory is L
- Transmit vector $\boldsymbol{x}_i = vec(\boldsymbol{x}[iP], \boldsymbol{x}[iP+1], \ldots, \boldsymbol{x}[iP+P-1])$,
- Receive vector $\boldsymbol{y}_i = vec(\boldsymbol{y}[iP+L], \boldsymbol{y}[iP+L+1], \ldots, \boldsymbol{y}[iP+P-1])$,
- Channel vector $\boldsymbol{h} = vec\left([\ \mathbf{H}[0]\ \ldots\ \mathbf{H}[L]\ ]^T\right)$,
- Rayleigh fading $\boldsymbol{h} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{R}_h)$, $\boldsymbol{n}_i \sim \mathcal{CN}(\mathbf{0}, \sigma_{nn}^2 \boldsymbol{I})$, $\boldsymbol{n}_i$, and $\boldsymbol{h}$ are independent.

Two Measures of System Performance

- **Minimizing MSE of channel estimation**:
  - We consider an estimation theoretic bound (Cramer-Rao Bound).
  - CRB establishes a lower bound on the MSE of any estimate of $\boldsymbol{h}$:

$$E\{(\boldsymbol{h} - \widehat{\boldsymbol{h}})(\boldsymbol{h} - \widehat{\boldsymbol{h}})^H\} \geq \mathcal{C}_h$$

  - Optimal training is the one that minimizes $trace(\mathcal{C}_h)$, where $\mathcal{C}_h$ is CRB for estimating channel.

- **Maximizing information rate**:
  - We consider a lower bound on mutual information $I(\boldsymbol{s}; \widehat{\boldsymbol{s}})$.
  - Most receivers make the decision about the symbols based on soft estimate of the modulated symbols and $I(\boldsymbol{s}; \widehat{\boldsymbol{s}}) \leq I(\boldsymbol{s}; \boldsymbol{y})$.
  - $I(\boldsymbol{s}; \widehat{\boldsymbol{s}})$ measures the reduction in the uncertainty about $\boldsymbol{s}$ due to the knowledge of $\widehat{\boldsymbol{s}}$.
  - Optimal training is the one that maximizes this lower bound.

Minimizing Channel Estimation Error

Notation $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{H}(\boldsymbol{F}\boldsymbol{s} + \boldsymbol{t}) + \boldsymbol{n} \Rightarrow \boldsymbol{H}\boldsymbol{t} = \boldsymbol{\Phi}(\boldsymbol{t})\boldsymbol{h}$ and $\boldsymbol{H}\boldsymbol{f}_i = \boldsymbol{\Phi}(\boldsymbol{f}_i)\boldsymbol{h}$.

- **Optimality condition for sequential estimation**

Minimizing $tr(\mathcal{C}_h)$ requires:

$$\boldsymbol{\Phi}(\boldsymbol{t})^H \boldsymbol{\Phi}(\boldsymbol{f}_i) = \boldsymbol{0} \quad i = 1, 2, \ldots, N$$

▶ **Remark 1**: To satisfy orthogonality constraint it is sufficient to separate data and training in frequency domain.

▶ **Remark 2**: Under orthogonality constraint $\mathcal{C}_h = (\sigma_{nn}^{-2}\boldsymbol{\Phi}(\boldsymbol{t})^H\boldsymbol{\Phi}(\boldsymbol{t}) + E_h\{\boldsymbol{\Xi}\} + \boldsymbol{R}_h^{-1})^{-1}$.

- **Optimality condition for joint estimation**

Minimizing $trace(\mathcal{C}_h)$ requires:

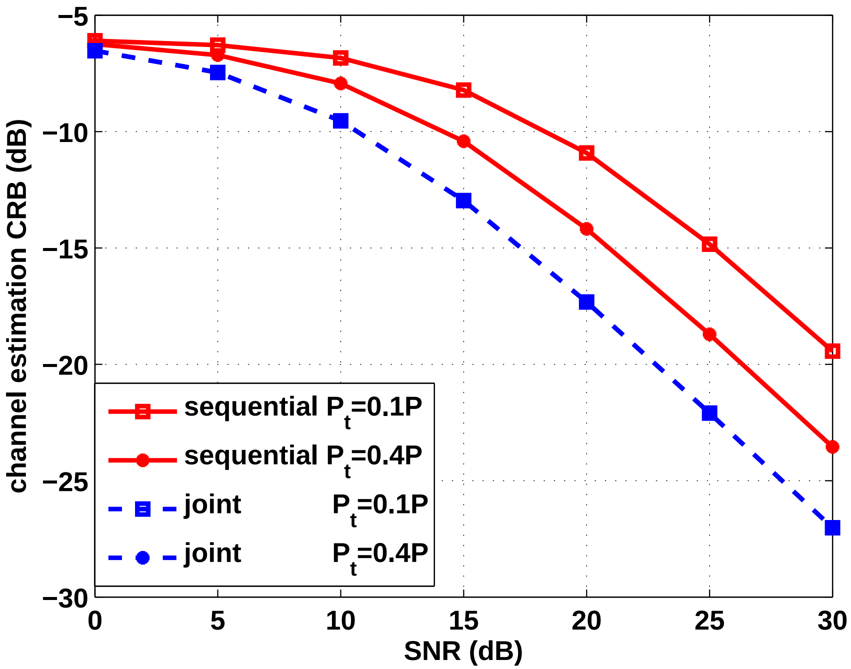
$$\sigma_{ss}^2 \sum_{i=1}^{N} \boldsymbol{\Phi}(\boldsymbol{f}_i)^H \boldsymbol{\Phi}(\boldsymbol{f}_i) + \boldsymbol{\Phi}(\boldsymbol{t})^H \boldsymbol{\Phi}(\boldsymbol{t}) = \boldsymbol{\Lambda}$$

where $\boldsymbol{\Lambda}$ is diagonal.

▶ **Remark 3**: For diagonalization no orthogonality constraint is required to be satisfied, and $\mathcal{C}_h = (\sigma_{nn}^{-2}\boldsymbol{\Lambda} + \boldsymbol{R}_h^{-1})^{-1}$.

Power Allocation Minimizing Channel Estimation Error

- $K = R = 2, L = 3,\ \sigma_{ss}^2 = 1$



- For sequential estimate minimizing $trace(\mathcal{C}_h)$ requires $\mathcal{P} = \mathcal{P}_t$. For joint estimate minimum of $trace(\mathcal{C}_h)$ depends on $\mathcal{P}$, not $\mathcal{P}_t$.

- Training design cannot be only based on minimizing channel estimation error $\Rightarrow$ we consider mutual information $I(\boldsymbol{s}; \hat{\boldsymbol{s}})$.

A Lower Bound on Mutual Information $I(s; \widehat{s})$

- We link the lower bound on $I(\boldsymbol{s}; \widehat{\boldsymbol{s}})$ with symbol estimation performance.

Let $\widehat{\boldsymbol{s}}$ be any symbol estimate, we have $E\{(\boldsymbol{s}-\widehat{\boldsymbol{s}})(\boldsymbol{s}-\widehat{\boldsymbol{s}})^H\} \geq \mathcal{C}_s$, where $\mathcal{C}_s$ is CRB for estimating the symbols. $I(\boldsymbol{s}; \widehat{\boldsymbol{s}})$ can be lower bounded as:

$$I(\boldsymbol{s}; \widehat{\boldsymbol{s}}) \quad \geq \quad log(|\sigma_s^2 \, \mathcal{C}_s^{-1}|)$$

▶ **Remark 1**: Maximizing lower bound is equivalent to minimizing $|\mathcal{C}_s|$.

▶ **Remark 2**: A better symbol estimate results in a higher transmission rate.

Mutual Information Lower Bound For Gaussian Symbols

- **Lower bound for sequential estimation**

Receiver models $\boldsymbol{h} = \hat{\boldsymbol{h}} + \tilde{\boldsymbol{h}}$. *Assuming* $\hat{\boldsymbol{h}}$ *is the LMMSE channel estimate:*

$$I(\boldsymbol{s};\hat{\boldsymbol{s}}) \geq E_{\hat{\boldsymbol{h}}}\{I(\boldsymbol{s};\hat{\boldsymbol{s}}|\hat{\boldsymbol{h}})\}$$

$$= E_{\hat{\boldsymbol{h}}}\{\log(|\sigma_{ss}^2 E_{\boldsymbol{s}}\{\boldsymbol{F}^H \hat{\boldsymbol{H}}^H \boldsymbol{R}_{v|s}^{-1} \hat{\boldsymbol{H}}\boldsymbol{F} + \boldsymbol{\mathcal{E}}^H(\boldsymbol{R}_{v|s}^{-1} \otimes \boldsymbol{D}^H \boldsymbol{R}_{v|s}^{-1} \boldsymbol{D})\boldsymbol{\mathcal{E}}\} + \boldsymbol{I}|)\}$$

▶ **Remark 1**: bound depends of $\hat{\boldsymbol{H}}$ and estimation error covariance.

- **Lower bound for joint estimation**

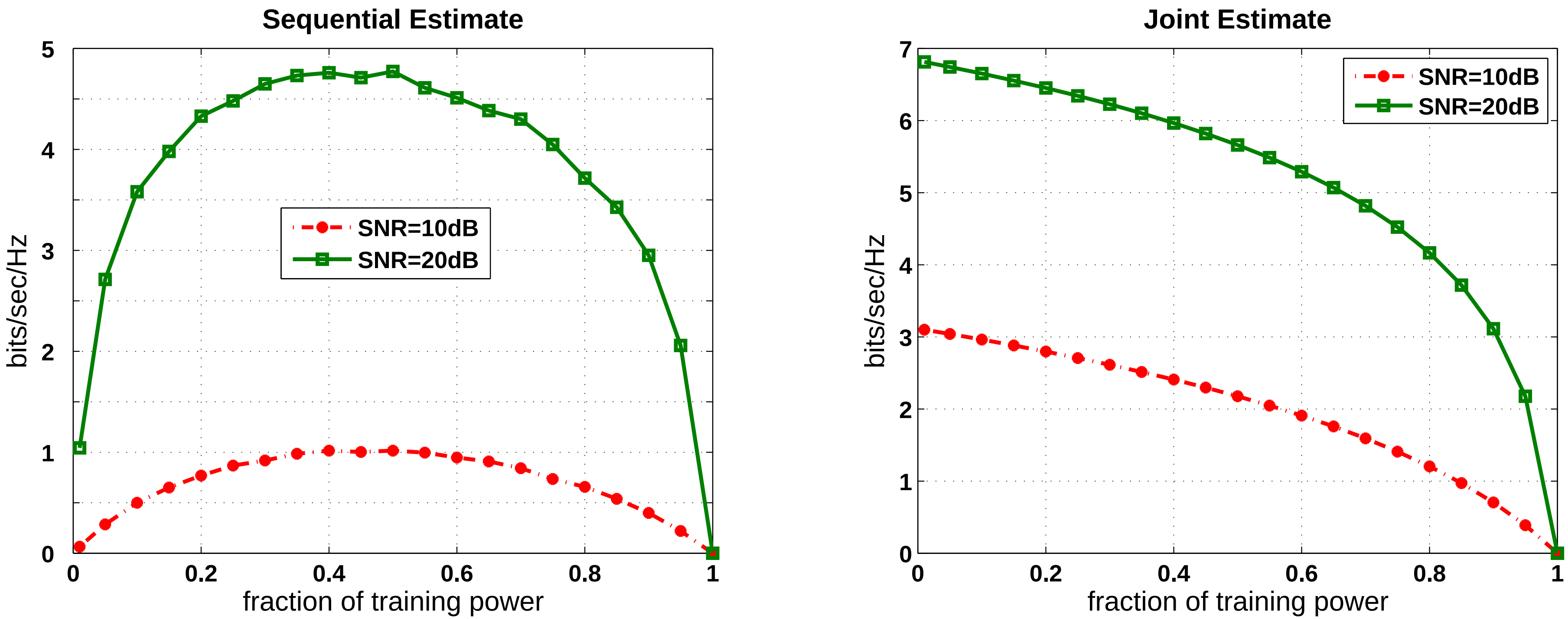
$$I(\boldsymbol{s};\hat{\boldsymbol{s}}) \geq \log(|\sigma_{ss}^2 \boldsymbol{\mathcal{C}}_s^{-1}|)$$

$$\boldsymbol{\mathcal{C}}_s^{-1} = \sigma_{nn}^{-2} \begin{bmatrix} tr(\boldsymbol{R}_h \mathbf{F}_1^H \mathbf{F}_1) & \cdots & tr(\boldsymbol{R}_h \mathbf{F}_1^H \mathbf{F}_N) \\ \vdots & \ddots & \vdots \\ tr(\boldsymbol{R}_h \mathbf{F}_N^H \mathbf{F}_1) & \cdots & tr(\boldsymbol{R}_h \mathbf{F}_N^H \mathbf{F}_N) \end{bmatrix} + \sigma_{ss}^{-2}\boldsymbol{I}$$

▶ **Remark 2**: bound is independent of $\hat{\boldsymbol{H}}$ and estimation error covariance.

▶ **Remark 3**: maximizing bound leads to the same optimality conditions derived for minimizing channel estimation error.

Power Allocation Maximizing Mutual Information Lower Bound



- Minimizing the bound requires $\mathcal{P}_t \simeq 0.4\mathcal{P}$ for sequential estimate and $\mathcal{P}_t = 0$ for joint estimate.

Summary

- We investigated optimal training for transmission over wideband wireless MIMO channels.
- We adopted affine precoding and considered channel estimation error and information rate as figures of merit.
- We showed that optimal training design depends on the receiver structure.
- For joint estimate training is not needed.
- Implementing joint estimate is computationally expensive.
- For sequential estimation training is needed for minimizing channel estimation error and maximizing mutual information lower.
- For sequential estimation separating data and training in frequency is optimal.

Future Work

- Optimal affine precoding design for practical communication systems:
  - specific channel and symbol estimation algorithms,
  - practical measures for system performance (e.g., BER),
  - specific channel coding and modulation schemes.
- Tradeoff between training and error correction channel coding.

Outline

- Training design for estimating wideband wireless channels with antenna arrays at transmitter and receiver
  - Estimation theoretic perspective
  - Information theoretic perspective
- Distributed data compression for sensor networks

Motivation

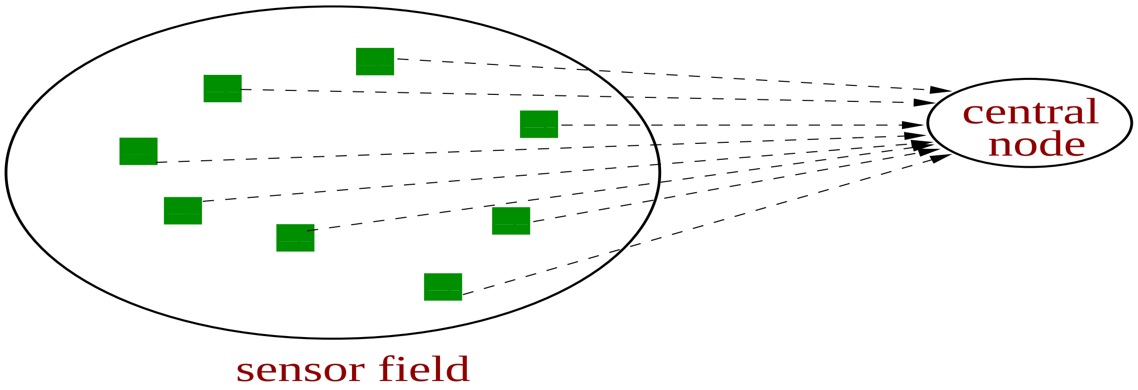
- Sensor networks consist of many low power and cheap sensors with limited computational capabilities.

- Neighboring sensors measure highly correlated data.

- Sensors send their measured data to a central node for joint decoding.

- The sensors could remove redundant data, if joint encoding is allowed.

- **Question**: is there a way of removing the redundancy in a completely distributed manner (without requiring the sensors to inter-communicate)?



Coding with Side Information: Example

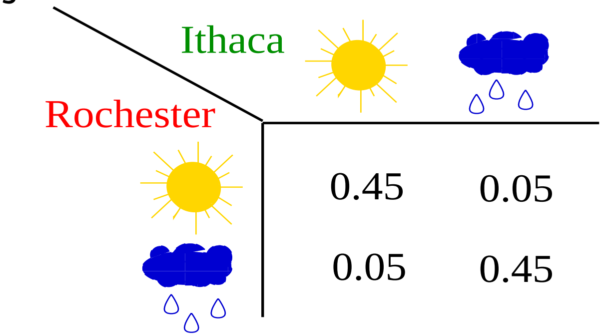
- Coding with side information enables distributed data compression.

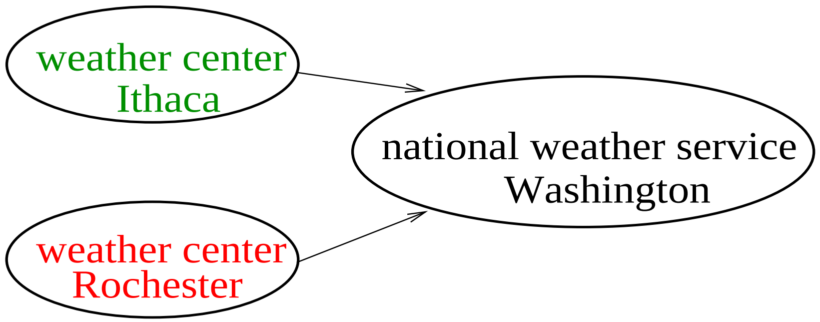
Example

▶ Rochester and Ithaca report the weather of 100 days in Summer to Washington.

▶ Washington knows the weather at Rochester, and the joint statistics.

▶ Ithaca is not aware of the weather in Rochester; however, it knows the joint statistics.

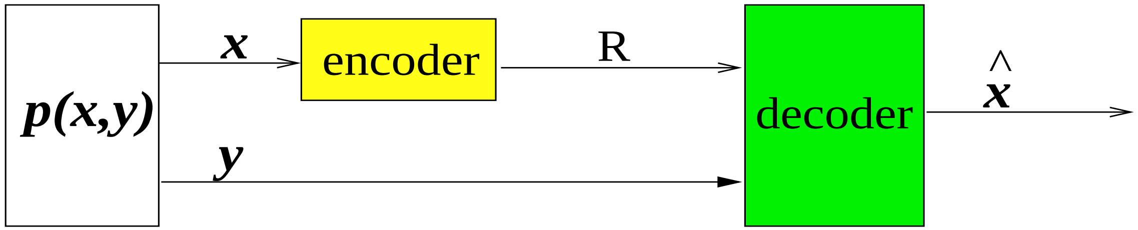




Question

▶ How many bits are needed to convey the Ithaca weather to Washington?

Coding with Side Information: Model



Set up: x, y are related through $p(x, y)$, known at encoder and decoder.

- (**Slepian and Wolf coding**) for discrete x, y one can compress x into $R < R_x$ bits to reconstruct x perfectly.

▶ Answer to the question: for independent coding $R_x = 100$ bits, and for coding with side information $R = 50$ bits are needed.

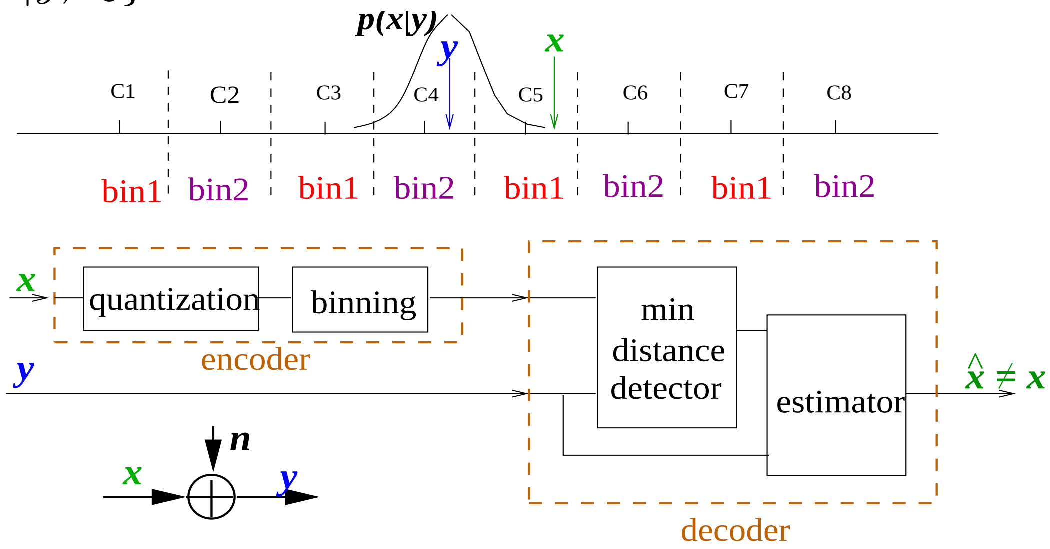
- (**Wyner and Ziv coding**) for continuous x, y one can compress x into $R < R_x$ bits to reconstruct x with MSE distortion D (e.g., $\mathbb{E}\{(x - \hat{x})^2\} = D$)

Conclusion: exploiting correlation at encoder/decoder allows a reduction of transmission rate.

Challenge: proofs are not constructive and are based on the idea of binning.

Most Relevant Work: DISCUS [Pradhan and Ramchandran 2003]

- DISCUS is a constructive approach to the problem of coding with side information.
- Assumption: the sources x, y are related by $y = x + n$.
  - ▶ Encoder quantizes x to c_5, and finds the bin index $p = 1$.
  - ▶ Given p and y decoder decides c_5 based on minimum distance rule, and forms $\hat{x} = \mathbb{E}\{x|y, c_5\}$.

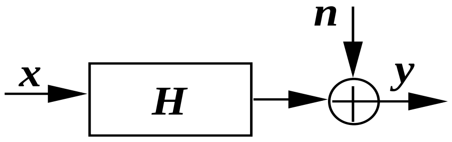


- Question: What if the relationship between y and x cannot be modelled as $y = x + n$?

A General Linear Correlation Model

- We consider two vector sources $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ that are related by

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



Assumptions

- $\boldsymbol{H}$ is constant, $\mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^T\} = \boldsymbol{R}_{xx}$, $\mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^T\} = \boldsymbol{R}_{nn}$, $\boldsymbol{x}$ and $\boldsymbol{n}$ are independent,
- Decoder knows $\boldsymbol{y}$ and the parameters $\boldsymbol{H}$, $\boldsymbol{R}_{xx}$, $\boldsymbol{R}_{nn}$,
- Encoder knows neither $\boldsymbol{y}$ nor $\boldsymbol{H}$, $\boldsymbol{R}_{nn}$.

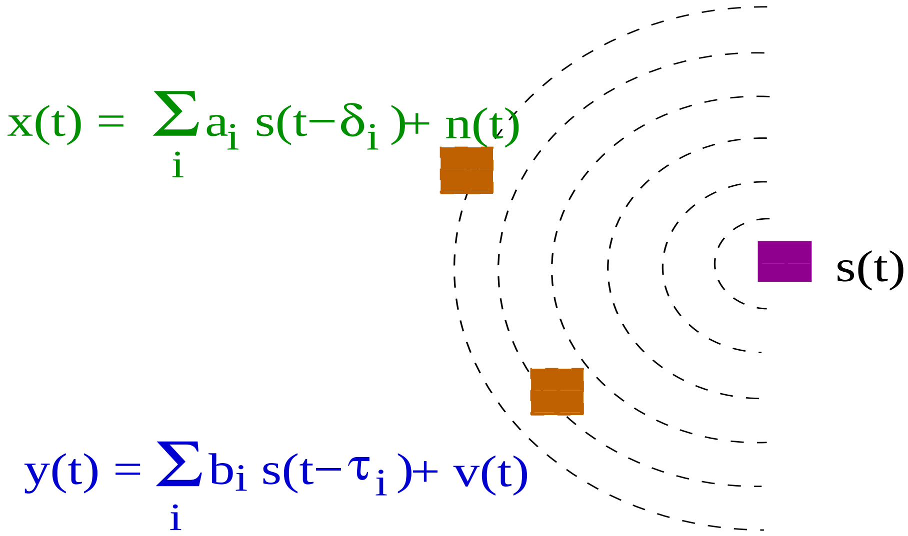
Goal: develop low complexity compression techniques to encode $\boldsymbol{x}$.

Why a Linear Correlation Model?

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

- It is more general than the model $y = x + n$.

Potential Application: audio field sensors



- For jointly Gaussian sources the model is exact.

Relevant Questions

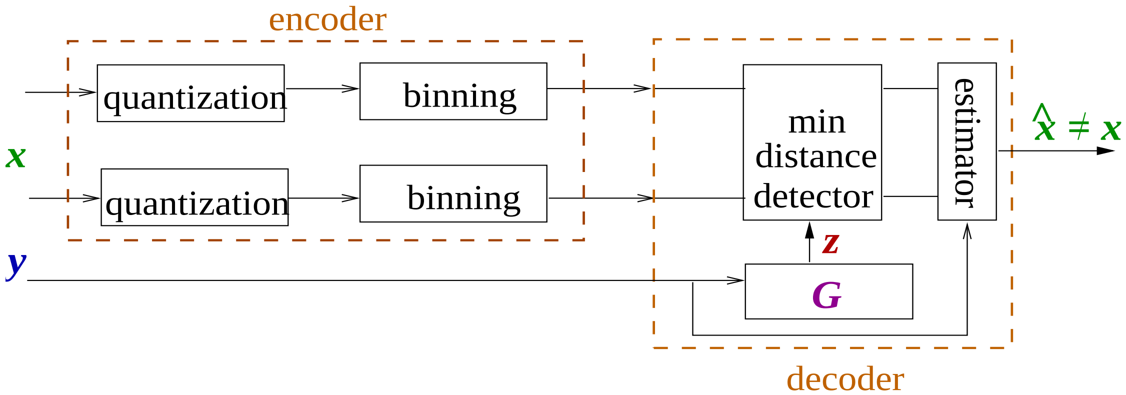
- Can we use the code construction designed for scalar source coding (e.g., DISCUS) to solve vector source coding problem?
- What is the structure of encoder/decoder?
- What is the best rate allocation to encode components of $\boldsymbol{x}$, subject to a sum rate constraint?

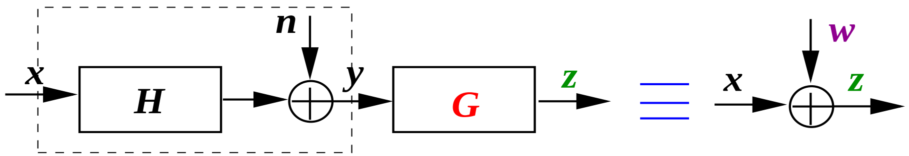
Encoder and Decoder Architecture: Equalization at the Decoder

- We use tools from communications (linear equalizer $\boldsymbol{G}$) to convert vector source coding problem into several scalar source coding problems.

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \quad \Rightarrow \quad \boldsymbol{z} = \boldsymbol{G}\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}$$

Zero Forcing $\boldsymbol{G} : \boldsymbol{G}\boldsymbol{H} = \boldsymbol{I}$ LMMSE equalizer $\boldsymbol{G} = \arg\min \mathbb{E}\{||\boldsymbol{G}\boldsymbol{y} - \boldsymbol{x}||^2\}$





- Given compression rate r_i bits/sample encoder quantizes x_i and provides decoder with bin indices p_i. Given p_i and z_i decoder decides on quantization intervals, and forms $\hat{\boldsymbol{x}}$.

Rate Allocation Policy in Distributed Data Compression

- Performance of the proposed system depends on rate allocation policy.
- Encoder compresses x_i with r_i bits/sample.
- Optimal rate allocation: given $R = \sum_{i=1}^{N} r_i$ what is the best $\{r_i\}_{i=1}^{N}$?

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \Rightarrow \boldsymbol{G}\boldsymbol{y} = \boldsymbol{z} = \boldsymbol{x} + \boldsymbol{w}$$

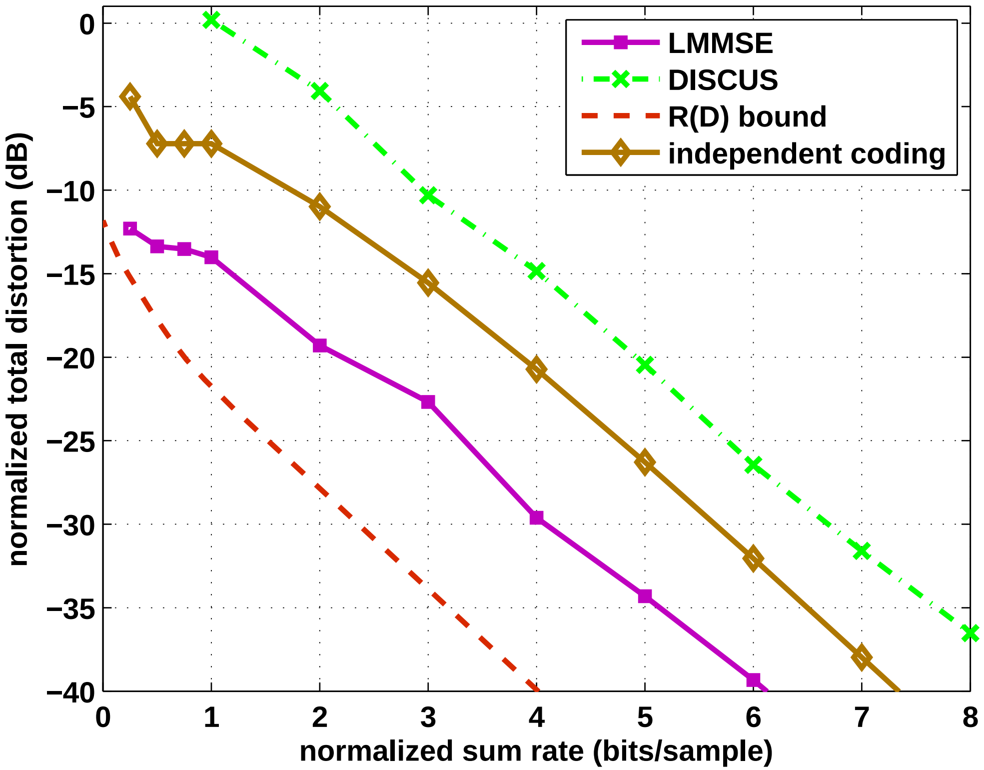
- Statistics of $\boldsymbol{w}$ depends on $\boldsymbol{G}$.
- Each z_i corresponds to a different $var(w_i) \Rightarrow r_i$ should be adapted to $var(w_i)$.
- Proposition: assign bits such that higher $var(w_i) \Rightarrow$ more r_i.

▶ **Remark 1**: the rate allocation policy contradicts the one in communication systems where higher $var(w_i) \Rightarrow$ less r_i.

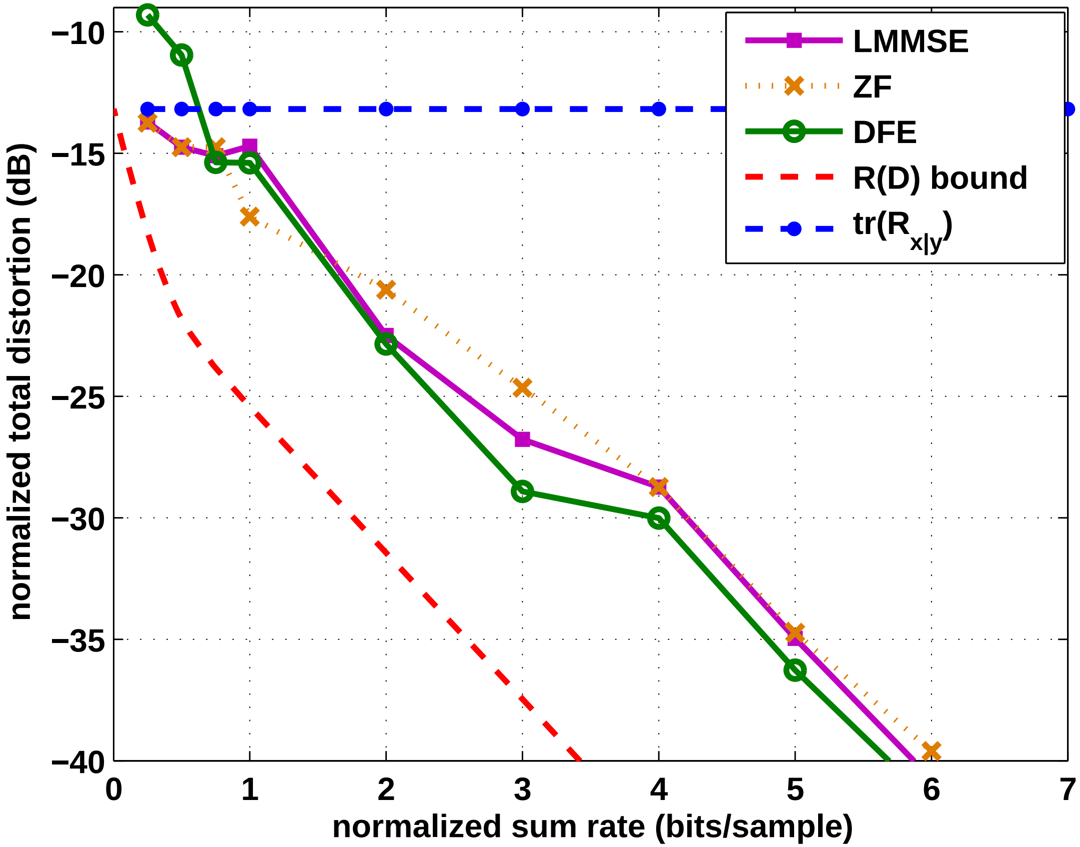
▶ **Remark 2**: higher $var(w_i) \Rightarrow z_i$ is less correlated is to $x_i \Rightarrow$ encoder needs to send more bits to the decoder.

Empirical Rate Distortion Bound: DISCUS vs. Decoder With LMMSE

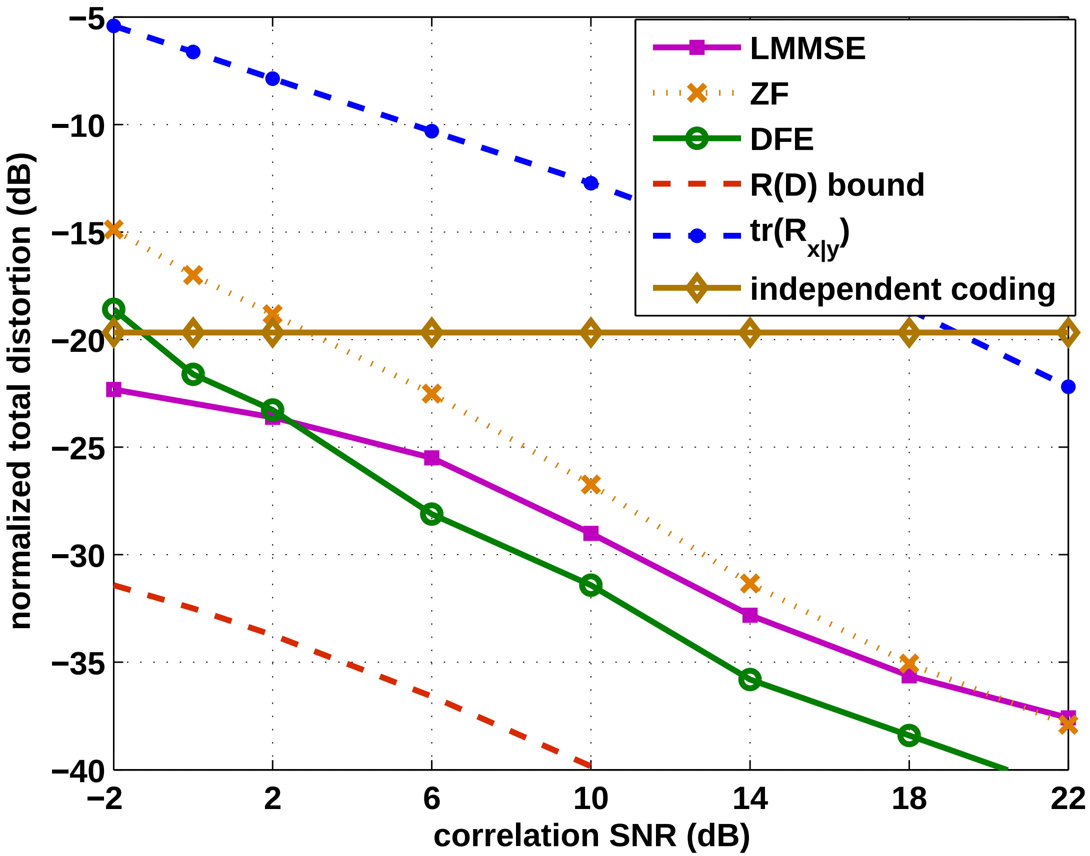
- $N = 4$, $\boldsymbol{H}$ and $\boldsymbol{R}_{xx}$ are Toeplitz , $\boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{R}_{xx})$ and $\boldsymbol{n} \sim \mathcal{N}(0, \sigma_n^2 \boldsymbol{I})$ with 10 $log_{10}\sigma_n^2 =$ -12 dB.



Empirical Rate Distortion Bound: Decoder With LMMSE,ZF,DFE



Distortion vs. Correlation SNR :Decoder With LMMSE,ZF,DFE



Summary

- We considered coding of vector source $\boldsymbol{x}$ with side information $\boldsymbol{y}$ when the correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$ is linear $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$.

- Using the equalization techniques we provided low complexity compression algorithms.

- The compression scheme reduces transmission rate, without explicitly assuming the correlation model at the encoder.

- We proposed a rate allocation policy which minimizes error probability of minimum distance detection.

Future Work

- How can we obtain parameters of the correlation model at the decoder?
- How can we make the bin index transmission robust to possible noise?
- How can we extend this code construction to more than two vector sources?