Title	Turbo Equalization: Fundamentals, Information Theoretic Considerations, and Extensions
Author(s)	Matsumoto, Tad; Anwar, Khoirul; Ahmad, Norulhusna
Citation	A Tutorial on the 75th IEEE Vehicular Technology Conference (VTC-Spring 2012)
Issue Date	2012-05-06
Type	Presentation
Text version	author
URL	http://hdl.handle.net/10119/10523
Rights	Copyright © 2012 Authors. Tad Matsumoto, Khoirul Anwar and Norulhusna Ahmad, Turbo Equalization: Fundamentals, Information Theoretic Considerations, and Extensions, A Tutorial on the 75th IEEE Vehicular Technology Conference (VTC-Spring 2012), Tutorial Handout ; Place : Yokohama, Japan ; Date : 6 - 9 May, 2012.
Description	IEEE Vehicular Technology Conference （電気電子学会移動体技術国際学会）VTC 2012-SpringでのTutorial Handout

Turbo Equalization: Fundamentals, Information Theoretic Considerations, and Extensions

A Tutorial on IEEE VTC-Spring 2012

Tad Matsumoto, Khoirul Anwar and Norulhusna Ahmad

Information Theory and Signal Processing Lab., School of Information Science,
Japan Advanced Institute of Science and Technology (JAIST),
1-1 Asahidai, Nomi, Ishikawa, 923-1211 JAPAN,
E-mail: {*matumoto,anwar-k*}*@jaist.ac.jp*

Yokohama, 6 May 2012

Our History: Block Wise Time Domain SC-MMSE

Problem of Complexity: $O[N^3(L+K)^3] \leftarrow$ Matrix inversion of $\mathbf{\Sigma}^{-1}$, N: antennas, L: path length, K: Frame length.

My 2002 Equation

$$\mathbf{z}_m = (\mathbf{I}_K + \mathbf{\Gamma}_m \mathbf{S}_m)^{-1} \left[\mathbf{\Gamma}_m \hat{\mathbf{s}}_m + \mathbf{H}_m^H \mathbf{\Sigma}^{-1} \tilde{\mathbf{r}}\right], \quad (1)$$

where

$$\mathbf{\Gamma}_m = \overline{\text{diag}} \left[\mathbf{H}_m^H \left(\mathbf{H}\mathbf{\Lambda}\mathbf{H}\right)^{-1} \mathbf{H}_m\right], \quad (2)$$

$$\mathbf{S}_m = \text{diag}[|\hat{\mathbf{s}}_m|^2], \quad (3)$$

$$\mathbf{\Sigma} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}^H + \sigma^2 \mathbf{I}, \quad (4)$$

$$\mathbf{\Lambda} = \mathbb{E}[(s-\hat{s})(s-\hat{s})^H] = \text{diag}[\mathbb{E}[|\mathbf{s}|^2] - |\hat{s}|^2]. \quad (5)$$

Our History: Block Wise Freq. Domain SC-MMSE

Finally, the equalizer output is given by:

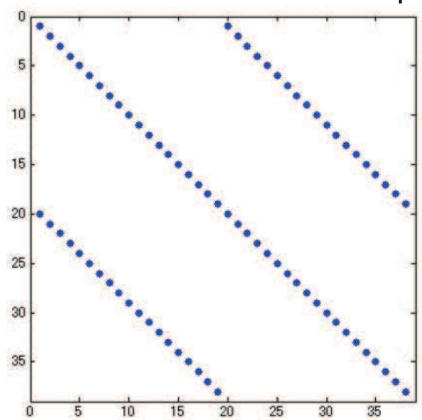
My 2005 Equation

$$\mathbf{z}_m = (\mathbf{I}_K + \mathbf{\Gamma}_m \mathbf{S}_m)^{-1} \left[\mathbf{\Gamma}_m \hat{\mathbf{s}}_m + \mathbf{F}^H \mathbf{\Xi}_m^{fH} (\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK})^{-1} \tilde{\mathbf{r}}^f \right], \tag{6}$$

where

$$\mathbf{\Gamma}_m \approx \overline{\mathrm{diag}} \left[\mathbf{F}^H \mathbf{\Xi}^{fH} (\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I})^{-1} \mathbf{\Xi}_m^f \mathbf{F} \right] \tag{7}$$

Diagonal-Block in the case $N = 2$:



Goals and Objectives of This Course

- The primary goal is to provide the audience with the in-depth knowledge of turbo equalization, its design techniques, and practical applications and its extensions.
- To provide lay knowledge for deriving the frequency domain processing, with which the computational complexity of the turbo equalization is independent of the number of the multipath components in the channel.
- To provide the course takers with the information theoretic background of the turbo equalization techniques.
- To introduce concatenation turbo equalizers neighboring in time in the absence of Cyclic Prefix (CP) or Guard Interval (GI)
- To introduce the n-OFDM technique, and then to provide the audience with the knowledge that how nOFDM systems with equalization can achieve high spectrum efficiency and improve the BER performance, compared to the conventional OFDM system.

Course Contents

1. Introduction of Instructors
2. Goals and Objectives, and Methodologies
3. Introduction to Frequency Domain Turbo Equalization (*by Prof. Tad Matsumoto*)
4. Information Theoretic Consideration of Turbo Equalization
5. Time-Concatenated Turbo Equalization systems without cyclic prefix (*by Prof. Khoirul Anwar*)
6. non-Orthogonal Frequency Division Multiplexing (nOFDM) (*by Ms. Norulhusna Ahmad*)
7. Conclusions

Part I
Turbo Equalization Basics
- Sliding Window -

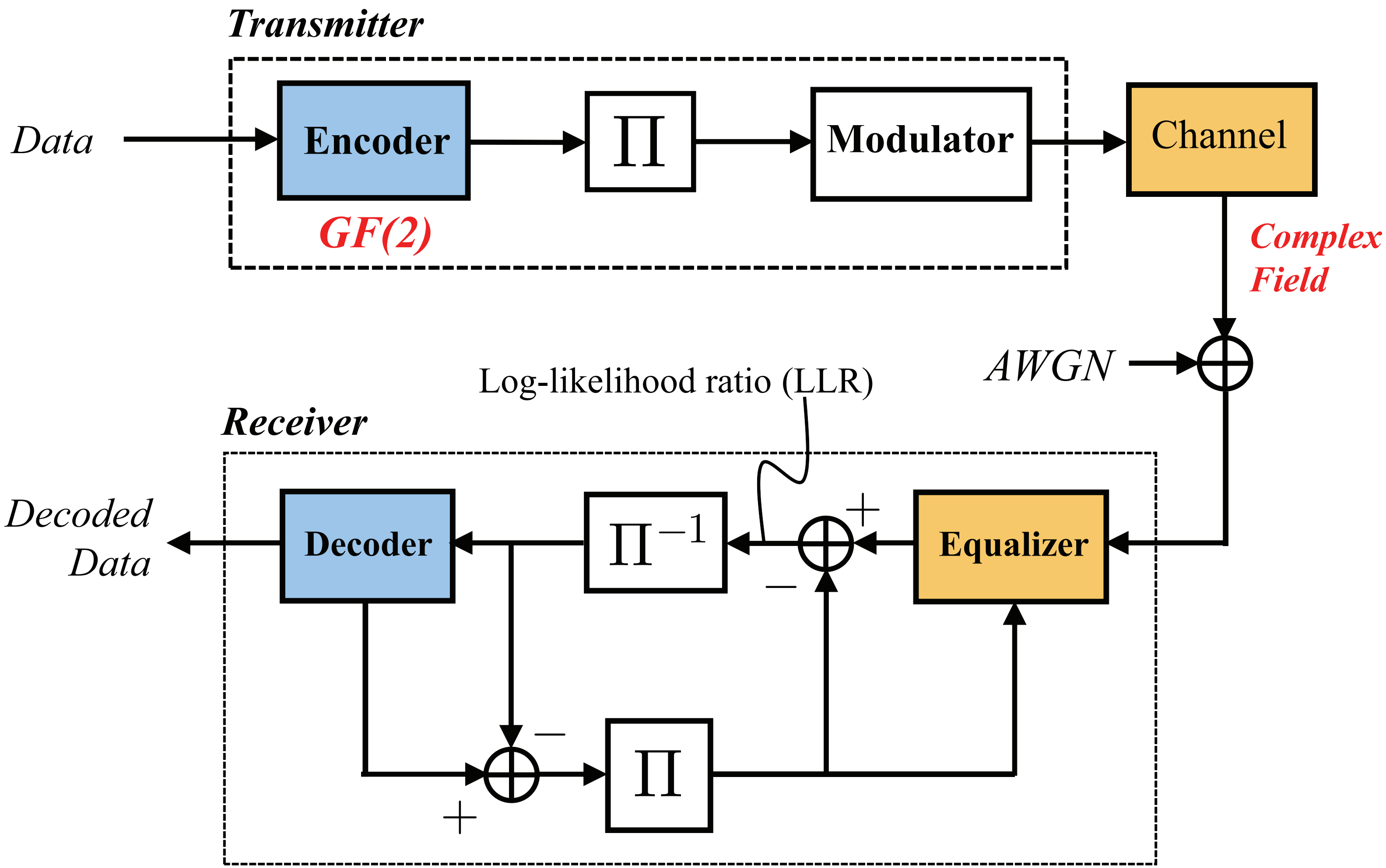
Tad Matsumoto

CWC, University of Oulu
Japan Advanced Institute of Science and Technology (JAIST)
e-mail: tadashi.matsumoto@ee.oulu.fi, matumoto@jaist.ac.jp

P.O. BOX 4500, University of Oulu, FIN-90014, FINLAND,
http://www.cwc.oulu.fi

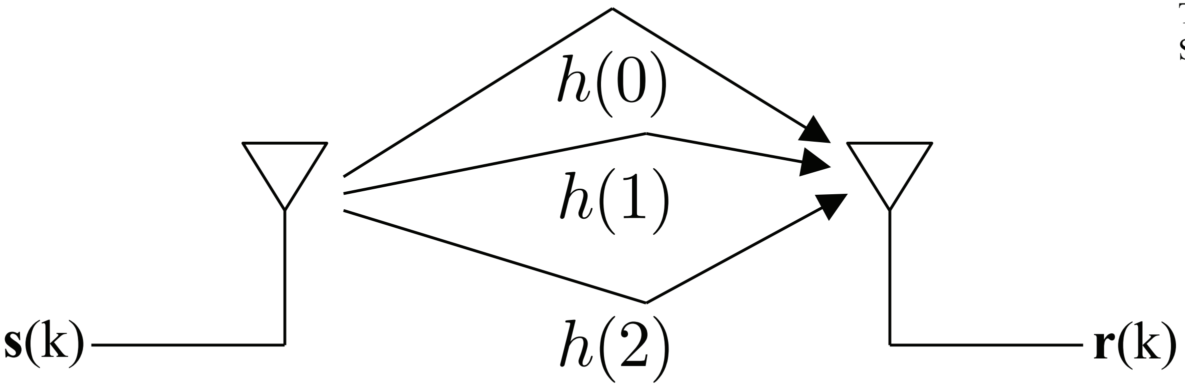
1-1 Asahidai, Nomi-shi, Ishikawa, 923-1292, JAPAN
http://www.jaist.ac.jp/is/labs/matsumoto-lab

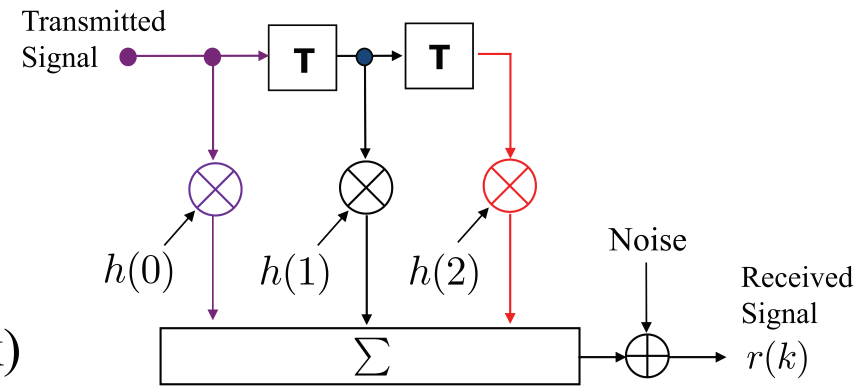
Single User's Case: Turbo Equalization (1)



Note: The definition field of the code should not necessarily be GF(2), but if the code is not binary, the exchanged information is probability vector. This is very complex.

The Channel

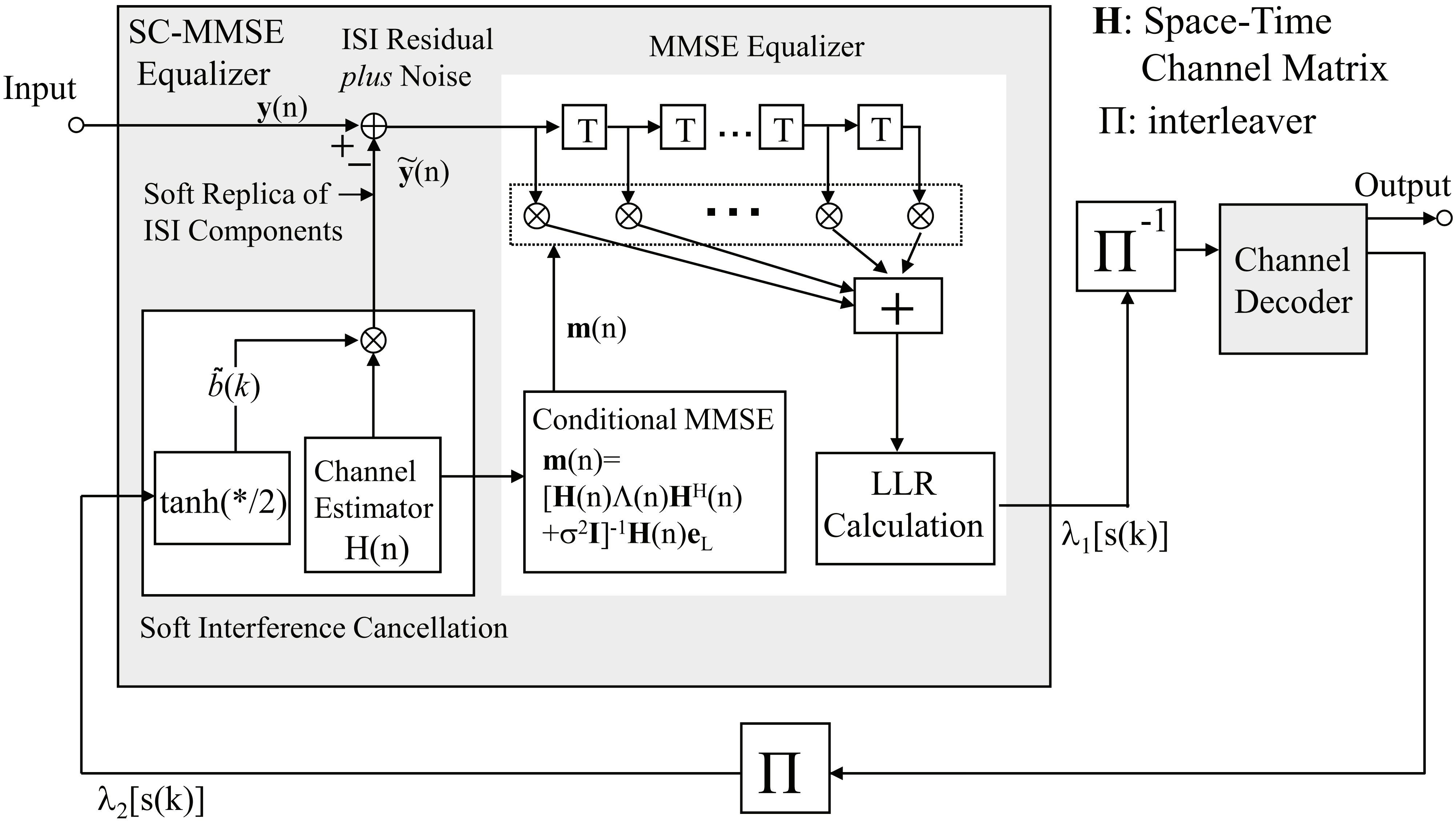




$$\mathbf{r}(k) = \begin{bmatrix} h(2) & h(1) & h(0) & 0 & 0 \\ 0 & h(2) & h(1) & h(0) & 0 \\ 0 & 0 & h(2) & h(1) & h(0) \end{bmatrix} \begin{bmatrix} s(k-2) \\ s(k-1) \\ s(k) \\ s(k+1) \\ s(k+2) \end{bmatrix} + \begin{bmatrix} v(1) \\ v(2) \\ v(3) \end{bmatrix}$$

$$= \mathbf{H}\mathbf{s}(k) + \mathbf{v} \quad (8)$$

SC-MMSE: Soft Cancellation with MMSE



Symbol's Expected Value

Why $\tanh(\lambda/2)$?

- Can anyone answer this question?

Why $\tanh(\lambda/2)$?

- Extrinsic information of bits (decoder feedback) $\lambda_2 = \ln \frac{\Pr(+1)}{\Pr(-1)}$.
- Using $\Pr(+1) + \Pr(-1) = 1$, we have

$$\Pr(-1) = \frac{1}{1 + e^{\lambda_2}}. \tag{9}$$

- The soft estimate of symbols $\hat{s} = \mathbb{E}[s]$ can be calculated from the probabilities $\Pr(+1)$ and $\Pr(-1)$, as

$$\begin{aligned} \hat{s} &= (-1) \cdot \Pr(-1) + 1 \cdot \Pr(+1), \\ &= -\Pr(-1) + \Pr(+1) e^{\lambda_2}, \\ &= \Pr(-1)(e^{\lambda_2} - 1) \end{aligned} \tag{10}$$

- Using (9) and (10), we have $\hat{s} = \frac{e^{\lambda_2} - 1}{e^{\lambda_2} + 1} = \tanh(\lambda_2/2)$.

Mistake Quite Often Found in the Literatures (1)

- $\hat{s} = \tanh(\lambda_2/2)$ is correct, only if the decoder feedback is uncorrelated. Otherwise, $\hat{s} = (-1) \cdot \Pr(-1) + 1 \cdot \Pr(+1)$ does not hold.
- Feeding the "soft" equalizer (or detector) output directly back to equalizer does not satisfy "turbo concept" because equalizer (detector) outputs are correlated.
- Shannon said "If the code is random enough and long enough, ..."
  - Interleaving has to be long enough
  - Interleaving has to be random enough

MMSE Weight Calculation – 1

- The whole channel structure can be expressed by the compact expression $\mathbf{y} = \mathbf{H}\mathbf{b} + \mathbf{v}$
- Soft cancellation

$$\tilde{\mathbf{y}} = \mathbf{y}(k) - \mathbf{H}\tilde{\mathbf{s}}(k) = \mathbf{H}(\mathbf{s}(k) - \tilde{\mathbf{s}}(k)) + \mathbf{v}(k), \tag{11}$$

where

$$\tilde{\mathbf{s}}(k) = [\tilde{s}(k-(L-1)), \cdots, \tilde{s}(k-1), 0, \tilde{s}(k+1), \cdots, \tilde{s}(k+(L-1))]^T. \tag{12}$$

MMSE Weight Calculation – 1

- Soft canceller outputs is weighted by the vector $\mathbf{m}(k)$ to output $z(k) = \mathbf{m}(k)^H \tilde{\mathbf{y}}(k)$ for which the criterion is:

$$\begin{aligned} \mathbf{m}(k) &= \arg\min_m \mathbb{E}\{|s(k) - \mathbf{m}^H \tilde{\mathbf{y}}(k)|^2\} \\ &= \arg\min_m \mathbf{m}^H \mathbb{E}\{\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}(k)^H\}\mathbf{m} - 2\mathbf{m}^H \mathbb{E}\{s(k)\tilde{\mathbf{y}}(k)\} + 1 \end{aligned} \tag{13}$$

where $\mathbb{E}\{\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}(k)^H\} = \mathbf{H}\mathbf{\Lambda}(k)\mathbf{H}^H + \sigma^2\mathbf{I}$ and $\mathbb{E}\{b(k)\tilde{\mathbf{y}}(k)\} = \mathbf{H}\mathbf{e}_L$ with

$$\begin{aligned} \mathbf{\Lambda}(k) &= \mathrm{cov}\{\mathbf{s}(k) - \tilde{\mathbf{s}}(k)\} \\ &= \mathrm{diag}\{1 - \tilde{s}(k-L+1)^2, \cdots, 1 - \tilde{s}(k-1)^2, 1, 1 - \tilde{s}(k+1)^2, \cdots, 1 + \tilde{s}(k+L-1)^2\} \end{aligned} \tag{14}$$

MMSE Weight Calculation – 2

- Now,

$$\text{MSE} = \mathbf{m}(k)^H \left[\mathbf{H}\boldsymbol{\Lambda}(k)\mathbf{H}^H + \sigma^2\mathbf{I}\right]\mathbf{m}(k) - 2\mathbf{m}(k)^H\mathbf{H}\mathbf{e}_L + 1 \quad (15)$$

- Setting the gradient of MSE with respect to $\mathbf{m}$ to be zero, we have

$$\mathbf{m}(k) = \left[\mathbf{H}\boldsymbol{\Lambda}(k)\mathbf{H}^H + \sigma^2\mathbf{I}\right]^{-1}\mathbf{H}\mathbf{e}_L. \quad (16)$$

Mistake Quite Often Found in Literatures (2)

- With the soft-feedback, the MMSE filter is time-invariant.
- Using the channel-fixed MMSE

$$\mathbf{m}(k) = \left[\mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I}\right]^{-1}\mathbf{H}\mathbf{e}_L \quad (17)$$

for every iteration does not make sense at all.

Asymptotic Performance Analysis

- If we can assume that decoder feedback is strong enough code, and that iterations are performed,

BPSK

$$\Lambda(k) \rightarrow \tilde{\Lambda} = \mathrm{diag}[0, \cdots, 0, 1, 0, \cdots, 0]. \tag{18}$$

- If this happens, since $\mathbf{H}\tilde{\mathbf{\Lambda}}\mathbf{H}^H$ becomes a rank-one matrix, the MMSE filter taps become

$$\mathbf{m}(k) \rightarrow \mathbf{m} = [\mathbf{h}\mathbf{h}^H + \sigma^2\mathbf{I}]^{-1}\mathbf{h} \tag{19}$$

$$\overset{(a)}{=} \alpha\mathbf{h}, \tag{20}$$

where $\alpha = \frac{1}{\mathbf{h}^H\mathbf{h}+\sigma^2}$.

Note:
(a) Matrix Inversion Lemma.

Matrix Inversion Lemma

Matrix Inversion Lemma:

$$[\mathbf{A}+\mathbf{BCD}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}]^{-1}\mathbf{DA}^{-1} \quad (21)$$

$$\mathbf{m} = [\mathbf{hh}^H + \sigma^2\mathbf{I}]^{-1}\mathbf{h} \quad (22)$$

$$\mathbf{A} = \mathbf{I}, \mathbf{B} = \mathbf{C}^H, \mathbf{D} = \frac{1}{\sigma^2} \quad (23)$$

- Now, by using the Matrix Inversion Lemma, we have

$$\begin{aligned} \mathbf{m} &= \left[\frac{1}{\sigma^2} - \frac{1}{\sigma^2}\mathbf{h}\left(\frac{1}{1+\mathbf{h}^H\mathbf{h}/\sigma^2}\right)\mathbf{h}^H\frac{1}{\sigma^2}\right], \\ &= \left[\frac{1}{\sigma^2} - \frac{1}{\sigma^2}\cdot\frac{\mathbf{h}^H\mathbf{h}}{\sigma^2+\mathbf{h}^H\mathbf{h}}\right]\mathbf{h} = \alpha\mathbf{h}, \end{aligned} \quad (24)$$

where $\alpha = \frac{1}{\mathbf{h}^H\mathbf{h}+\sigma^2}$.

Extrinsic LLR Calculation

- Assuming that the MMSE filter output $z(k) = \mathbf{m}(k)^H \tilde{\mathbf{y}}(k)$ can be approximated as a Gaussian channel output having input $b(k)$,

$$z(k) = \mu(k)s(k) + \eta(k), \tag{25}$$

where $\mu(k) = \mathbb{E}[z(k)s(k)] = \mathbf{e}_L H^H \mathbf{m}(k)$, and the variance of $\eta(k)$ is

$$\begin{aligned} v(k)^2 &= \mathbb{E}[|z(k)|^2] - \mu^2 \\ &= \mathbf{e}_L H^H \mathbf{m}(k) - \mu^2 = \mu - \mu^2 \end{aligned} \tag{26}$$

- The extrinsic LLR $\lambda_1[b(k)]$ at the MMSE filter output is then given by:

$$\begin{aligned} \lambda_1[b(k)] &= \log \frac{\Pr(z(k)|s(k) = +1)}{\Pr(z(k)|s(k) = -1)} \\ &= -\frac{|z(k) - \mu(k)|^2}{v^2} + \frac{|z(k) + \mu(k)|^2}{v^2} = \frac{4\Re(z(k))}{1 - \mu(k)}. \end{aligned} \tag{27}$$

Time Domain Block-Wise Processing

Tad Matsumoto

Revisit MMSE Weight Calculation – 1

- Soft cancellation $\tilde{\mathbf{y}} = \mathbf{y}(k) - \mathbf{H}\tilde{\mathbf{s}}(k) = \mathbf{H}(\mathbf{s}(k) - \tilde{\mathbf{s}}(k)) + \mathbf{v}(k)$, where $\tilde{\mathbf{s}}(k) = [\tilde{s}(k+(L-1)), \cdots, \tilde{s}(k+1), 0, \tilde{s}(k-1), \cdots, \tilde{s}(k-(L-1))]^T$.
- Soft canceller outputs is weighted by the vector $\mathbf{m}(k)$ to output $z(k) = \mathbf{m}(k)^H \tilde{\mathbf{y}}(k)$ for which the criterion is:

$$\begin{aligned} \mathbf{m}(k) &= \arg\min_{m} \mathbb{E}\{|s(k) - \mathbf{m}^H \tilde{\mathbf{y}}(k)|^2\} \\ &= \arg\min_{m} \mathbf{m}^H \mathbb{E}\{\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}(k)^H\}\mathbf{m} - 2\mathbf{m}^H \mathbb{E}\{s(k)\tilde{\mathbf{y}}(k)\} + 1 \end{aligned} \tag{28}$$

where $\mathbb{E}\{\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}(k)^H\} = \mathbf{H}\mathbf{\Lambda}(k)\mathbf{H}^H + \sigma^2\mathbf{I}$ and $\mathbb{E}\{b(k)\tilde{\mathbf{y}}(k)\} = \mathbf{H}\mathbf{e}_L$ with

$$\begin{aligned} \mathbf{\Lambda}(k) &= \mathrm{cov}\{\mathbf{s}(k) - \tilde{\mathbf{s}}(k)\} \\ &= \mathrm{diag}\{1 - \tilde{s}(k-L+1)^2, \cdots, 1 - \tilde{s}(k-1)^2, 1, 1 - \tilde{s}(k+1)^2, \cdots, 1 - \tilde{s}(k+L-1)^2\} \end{aligned} \tag{29}$$

MMSE Weight Calculation – 3: Modification

- Now define
  $\tilde{\tilde{\mathbf{s}}}(k) = [\tilde{s}(k+(L-1)), \cdots, \tilde{s}(k+1), \tilde{s}(k), \tilde{s}(k-1), \cdots, \tilde{s}(k-(L-1))]^T$,
  then

$$\begin{aligned}\mathbf{\Lambda}_0(k) &= \mathrm{cov}\{\mathbf{s}(k) - \tilde{\tilde{\mathbf{s}}}(k)\} \\ &= \mathrm{diag}\{1-\tilde{s}(k+L-1)^2, \cdots, 1-\tilde{s}(k+1)^2, 1-\tilde{s}(k)^2, 1-\tilde{s}(k-1)^2, \cdots, 1-\tilde{s}(k-L+1)^2\}\end{aligned} \quad (30)$$

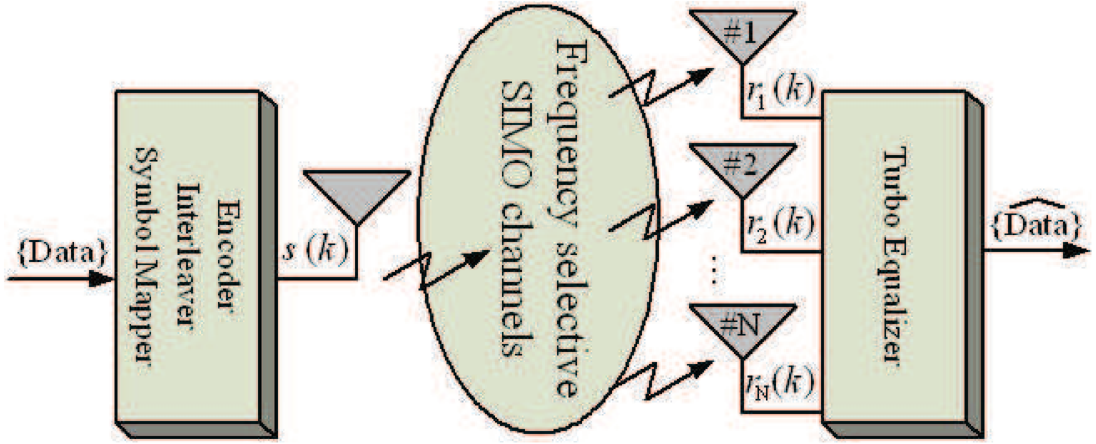
.

- With Λ_0 defined above, the MMSE weight vector is given by:

$$\begin{aligned}\mathbf{m}(k) &= \Bigg[\mathbf{H}\mathbf{\Lambda}_0(k)\mathbf{H}^H + \overbrace{\tilde{\mathbf{s}}(k)^2}^{\text{Restoral term}} \mathbf{h}\mathbf{h}^H + \sigma^2\mathbf{I}\Bigg]^{-1} \mathbf{H}\mathbf{e}_L \\ &= \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\left(\frac{1}{\tilde{s}(k)^2} + \mathbf{h}^H\mathbf{\Sigma}^{-1}\mathbf{h}\right)^{-1} \mathbf{h}^H\mathbf{\Sigma}^{-1}\end{aligned} \quad (31)$$

and $\mathbf{\Sigma}^{-1} = \left(\mathbf{H}\mathbf{\Lambda}_0\mathbf{H}^H + \sigma^2\mathbf{I}\right)^{-1}$ is independent of the symbol index k.

SIMO Channel Model (1)



Received Signal:

$$r_n(k) = \sum_{l=0}^{L-1} h_n(l)s(k-l) + v_n(k)$$

Received Signal (spatially sampled):

$$\mathbf{r}(k) = \sum_{l=0}^{L-1} \mathbf{H}(l)s(k-l) + \mathbf{v}(k)$$

- $r_n(k)$: received signal
- L : channel memory length
- $h_n(l)$: channel impulse response
- $s(k)$: transmitted signal
- $v_n(k)$: complex white Gaussian noise $\mathcal{CN}(0, \sigma^2)$

$$\mathbf{r}(k) = [r_1(k), r_2(k), \cdots, r_N(k)]^T,$$
$$\mathbf{H}(l) = [h_1(l), h_2(l), \cdots, h_N(l)]^T,$$
$$\mathbf{v}(k) = [v_1(k), v_2(k), \cdots, v_N(k)]^T.$$

Channel Model (2)

- Block-wise Space-Time channel matrix

$$\begin{aligned}
\mathbf{H} &= \left[\mathbf{H}_1^T, \mathbf{H}_2^T, \cdots, \mathbf{H}_n^T\right]^T \in \mathbb{C}^{K \times N(K+L)}, &(32)\\
\mathbf{H}_n &= \left[\mathbf{h}_{n,1}, \mathbf{h}_{n,2}, \cdots, \mathbf{h}_{n,K}\right], &(33)\\
\mathbf{h}_{n,k} &= \left[0_{k-1}, h_n(0), h_n(1), \cdots, h_n(L-1), 0_{K-k+1}\right]^T. &(34)
\end{aligned}$$

- $\mathbf{H}_n$ has the Toeplitz structure (linear convolution)
- K: block length, 0_x: all zeros vector of length x
- Example of $\mathbf{H}_n \in \mathbb{C}^{(K+L)\times K}$ with $L = 3$,

$$\mathbf{H}_n = \begin{bmatrix}
h_n(0) & 0 & 0 & 0 & \cdots & 0 & 0 \\
h_n(1) & h_n(0) & 0 & 0 & \cdots & 0 & 0 \\
h_n(2) & h_n(1) & h_n(0) & 0 & \cdots & 0 & 0 \\
0 & h_n(2) & h_n(1) & h_n(0) & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & \cdots & h_n(2) & h_n(1) \\
0 & 0 & 0 & 0 & \cdots & 0 & h_n(2)
\end{bmatrix} \quad (35)$$

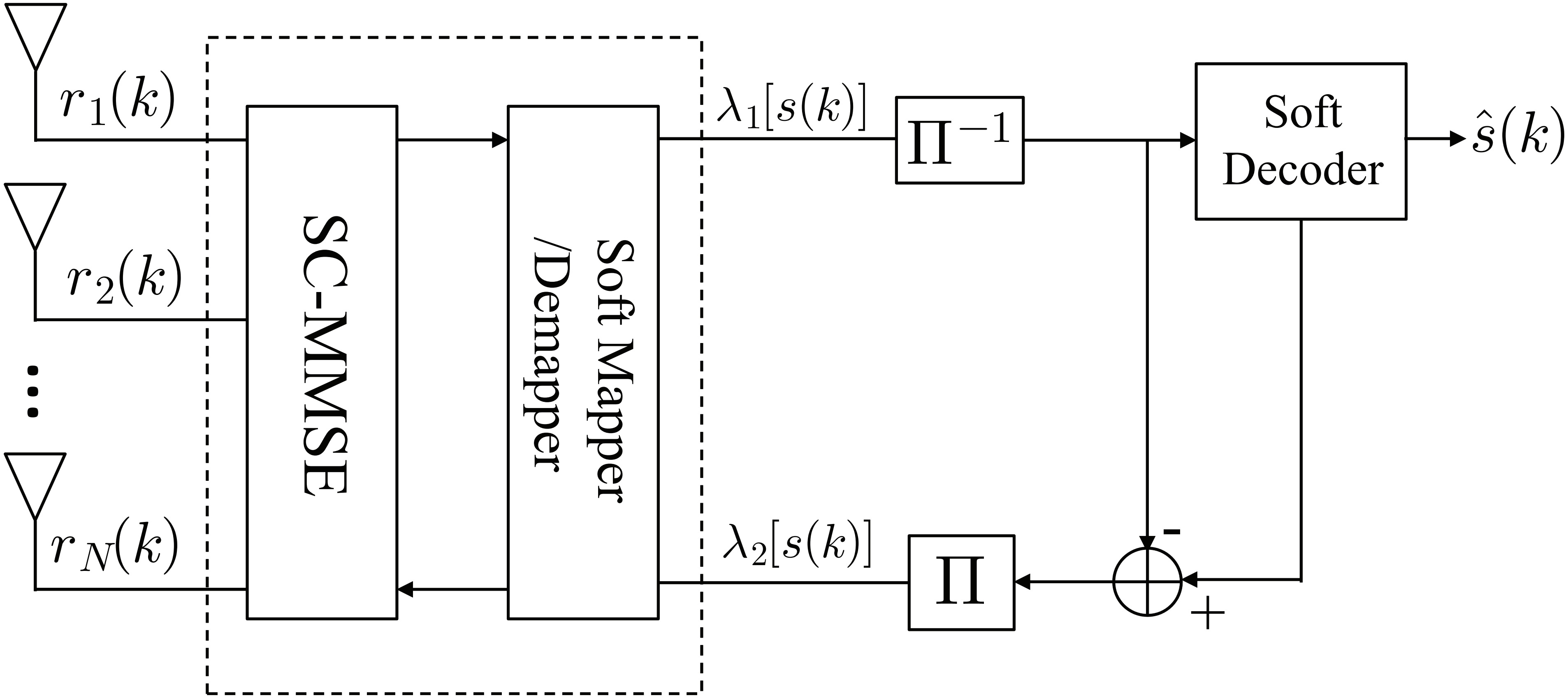
Channel Model (3)

- Received signal vector

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

$$\begin{aligned}
\mathbf{r} &= [\mathbf{r}_1^T, \mathbf{r}_2^T, \cdots, \mathbf{r}_N^T]^T \\
\mathbf{r}_n &= [r_n(L-2), \cdots, r_n(0), r_n(1), \cdots, r_n(K-L+1), \cdots, r_n(K)]^T \\
\mathbf{s} &= [s(1), s(2), \cdots, s(K)]^T \\
\mathbf{v} &= [\mathbf{v}_1^T, \mathbf{v}_2^T, \cdots, \mathbf{v}_N^T]^T \\
\mathbf{v}_n &= [v_n(L-2), \cdots, v_n(0), v_n(1), \cdots, v_n(K-L+1), \cdots, v_n(K)]^T
\end{aligned}$$

Principle of Turbo Equalization



Mapper & Soft Interference Cancellation

- Mapper
  A priori log-likelihood ratio (LLR) provided by the decoder (BPSK)

$$\lambda_1[s(k)] = \frac{\Pr(s(k)=+1)}{\Pr(s(k)=-1)} \tag{36}$$

- First moment of the soft symbol

$$\hat{s}(k) = \tanh\left(\frac{\lambda_1[s(k)]}{2}\right) \tag{37}$$

- Soft interference cancellation
  The soft symbol is used in the cancellation of interference components from the received signal to provide a residual

$$\tilde{\mathbf{r}} = \mathbf{r} - \mathbf{H}\hat{\mathbf{s}} \tag{38}$$

$$\hat{\mathbf{s}} = [\hat{s}(1), \hat{s}(2), \cdots, \hat{s}(K)^T \tag{39}$$

- If the a priori information is perfect, only noise components remain.

MMSE Equalizer (Time Domain)

- Adaptive Linear Filter MMSE filter is used to further reduce the residual interference components: The vector $\mathbf{w}^H$ for the filter taps is obtained to minimize the following mean square error (MSE).

$$\mathbf{w}(k) = \arg \min_{\mathbf{w}(k)^H} |\mathbf{w}(k)^H(\mathbf{r} + \overbrace{\mathbf{h}(k)\hat{s}(k)}^{\text{restoral term}}) - s(k)|^2, \quad (40)$$

where $\mathbf{h}(k) = \left[\mathbf{h}_{1,k}^T, \mathbf{h}_{2,k}^T, \ldots, \mathbf{h}_{N,k}^T\right]$.

- Using Wiener-Hopf solutions

$$\mathbb{E}\left[\frac{\partial|\mathbf{w}(k)^H(\mathbf{r} + \mathbf{h}(k)\hat{s}(k)) - s(k)|^2}{\partial\mathbf{w}(k)^H}\right] = 0. \quad (41)$$

MMSE Equalizer (Time Domain)

- After taking partial derivative

$$\underbrace{\mathbb{E}\left[(\tilde{\mathbf{r}} + \mathbf{h}\hat{s}(k))(\tilde{\mathbf{r}} + \mathbf{h}\hat{s}(k))^H\right]}_{\text{Covariance Matrix}}\mathbf{w}(k) - \underbrace{\mathbb{E}\left[(\tilde{\mathbf{r}} + \mathbf{h}(k)\hat{s}(k))s(k)^*\right]}_{\text{Steering Vector}} = 0 \quad (42)$$

About $\mathbf{w}(k)$:

$$\mathbf{w}(k) = [\text{Covariance Matrix}]^{-1}\,[\text{Steering Vector}]$$

- Covariance Matrix

$$\begin{aligned}\mathbb{E}\left[(\tilde{\mathbf{r}} + \mathbf{h}\hat{s}(k))(\tilde{\mathbf{r}} + \mathbf{h}\hat{s}(k))^H\right] &= \mathbb{E}\left[\tilde{\mathbf{r}}\tilde{\mathbf{r}}^H\right] + \tilde{\mathbf{r}}\hat{s}(k)^*\mathbf{h}(k)^H + \mathbf{h}(k)\hat{s}(k)\tilde{\mathbf{r}}^H \\ &\quad + \mathbf{h}(k)\hat{s}(k)\hat{s}(k)^*\mathbf{h}(k)^H \quad (43) \\ &\overset{(a)}{=} \mathbb{E}\left[\tilde{\mathbf{r}}\tilde{\mathbf{r}}^H\right] + \mathbf{h}(k)|\hat{s}(k)|^2\mathbf{h}(k)^H \quad (44)\end{aligned}$$

Note: (a): with $\mathbb{E}[s(k)] = \hat{s}(k)$, then

$$\mathbf{h}(k)\left(\mathbb{E}[s(k)]\hat{s}(k)^* + \hat{s}(k)\mathbb{E}[s(k)^*] - 2|\hat{s}(k)|^2\right)\mathbf{h}(k)^H = 0$$

MMSE Equalizer (Time-Domain): Covariance Matrix

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\mathbf{r}}\tilde{\mathbf{r}}^H\right] &= \mathbb{E}\left[(\mathbf{r}-\mathbf{H}\hat{\mathbf{s}})(\mathbf{r}-\mathbf{H}\hat{\mathbf{s}})^H\right], \\
&= \mathbb{E}\left[(\mathbf{H}(\mathbf{s}-\hat{\mathbf{s}})+\mathbf{v})(\mathbf{H}(\mathbf{s}-\hat{\mathbf{s}})+\mathbf{v})^H\right], \\
&= \mathbf{H}\mathbb{E}\left[(\mathbf{s}-\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}})^H\right]\mathbf{H}^H + \sigma^2\mathbf{I}, \\
&\overset{(a)}{=} \mathbf{H}\mathbf{\Lambda}\mathbf{H}^H + \sigma^2\mathbf{I}_K
\end{aligned}
\tag{45}
$$

$$
\begin{aligned}
\mathbf{\Lambda} &= \mathbb{E}[(\mathbf{s}-\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}})^H], \\
&\overset{(b)}{=} \overline{\mathrm{diag}}\{\mathbb{E}[|\mathbf{s}|^2] - \mathbb{E}[\mathbf{s}]\hat{\mathbf{s}}^H - \mathbb{E}\left[\mathbf{s}^H\right]\hat{\mathbf{s}} + |\hat{\mathbf{s}}|^2\}, \\
&= \overline{\mathrm{diag}}\{\mathbb{E}[|\mathbf{s}|^2] - 2|\hat{\mathbf{s}}|^2 + |\hat{\mathbf{s}}|^2\}, \\
&\overset{(c)}{=} \overline{\mathrm{diag}}\{\mathbb{E}[|\mathbf{s}|^2] - |\hat{\mathbf{s}}|^2\} + 0
\end{aligned}
\tag{46}
$$

Note: (a): $\mathbf{I}_K$ denotes a $K \times K$ identity matrix, (b) : $\overline{\mathrm{diag}}$ is an operator which replaces non-diagonal elements in the matrix with zeros, (c) : $\mathbb{E}[\mathbf{s}(k)] = \hat{s}(k)$, then $\mathbb{E}[\mathbf{s}(k)]\mathbb{E}[\mathbf{s}(k)^*] = |\hat{\mathbf{s}}|^2 \neq \mathbb{E}[|\mathbf{s}(k)|^2]$.

MMSE Equalizer (Time-Domain): Mean Vector

- Covariance Matrix

$$\mathbb{E}\left[(\tilde{\mathbf{r}}+\mathbf{h}\hat{s}(k))(\tilde{\mathbf{r}}+\mathbf{h}\hat{s}(k))^H\right] = \mathbf{\Sigma}+\mathbf{h}|\hat{s}(k)|^2\mathbf{h}(k)^H, \\ \stackrel{(a)}{=} \mathbf{H}\mathbf{\Lambda}\mathbf{H}^H+\sigma^2\mathbf{I}+\mathbf{h}|\hat{s}(k)|^2\mathbf{h}(k)^H \quad (47)$$

- Steering Vector

$$\begin{aligned}\mathbb{E}[(\tilde{\mathbf{r}}+\mathbf{h}(k)\hat{s}(k))s(k)^*] &= \mathbb{E}[\mathbf{H}\mathbf{s}s(k)^* - \mathbf{H}\hat{\mathbf{s}}s(k)^* + \mathbf{v}s(k)^* \\ &\quad +\mathbf{h}(k)\hat{s}(k)s(k)^*], \\ &= \mathbf{h}(k) \end{aligned} \quad (48)$$

- Finally, we can find the optimum filter tap coefficients as

$$\mathbf{w}(k) = (\mathbf{\Sigma}+\mathbf{h}(k)|\hat{s}(k)|^2\mathbf{h}(k)^H)^{-1}\mathbf{h}(k). \quad (49)$$

Note: (a) : $\mathbf{\Sigma}$ is independent of the k-th common covariance.

MMSE Equalizer: $z(k)$

- Equalizer Output

$$
\begin{aligned}
z(k) &= \mathbf{w}(k)^H(\tilde{\mathbf{r}}(k) + \mathbf{h}(k)\hat{s}(k)), \\
&= \mathbf{h}(k)^H \left[\left(\mathbf{\Sigma} + \mathbf{h}(k)|\hat{s}(k)|^2\mathbf{h}(k)^H\right)^{-1}\right]^H \left(\tilde{\mathbf{r}}(k) + \mathbf{h}(k)\hat{s}(k)\right), \\
&\overset{(a)}{=} \mathbf{h}(k)^H \left(\mathbf{\Sigma} + \mathbf{h}(k)|\hat{s}(k)|^2\mathbf{h}(k)^H\right)^{-1} \left(\tilde{\mathbf{r}}(k) + \mathbf{h}(k)\hat{s}(k)\right) \quad (50)
\end{aligned}
$$

Note:
(a) : Because of the "Hermitianity" of the argument $\left(\mathbf{X}\mathbf{X}^H\right)^H = \mathbf{X}\mathbf{X}^H$.

Weighting Factor $\mathbf{w}(k)$

$$
\begin{aligned}
\mathbf{w}(k)^H &= \mathbf{h}(k)^H \left[\boldsymbol{\Sigma} + \mathbf{h}(k)|\hat{s}(k)|^2\mathbf{h}(k)^H\right]^{-1}, \\
&\overset{(a)}{=} \mathbf{h}(k)^H\Big(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{h}(k)\left[|\hat{s}(k)|^{-2} + \mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}\mathbf{h}(k)\right]^{-1}\mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}\Big), \\
&\overset{(b)}{=} \mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1} - \mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}\mathbf{h}(k)\left[|\hat{s}(k)|^{-2} + \gamma(k)\right]^{-1}\mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}, \\
&= \mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1} - \gamma(k)\left[|\hat{s}(k)|^{-2} + \gamma(k)\right]^{-1}\mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}, \\
&= \mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1} - \frac{\gamma(k)|\hat{s}(k)|^2}{1+\gamma(k)|\hat{s}(k)|^2}\mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}, \\
&= \left(1 - \frac{\gamma(k)|\hat{s}(k)|^2}{1+\gamma(k)|\hat{s}(k)|^2}\right)\mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}, \\
&= \left(1+\gamma(k)|\hat{s}(k)|^2\right)^{-1}\mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1} \qquad (51)
\end{aligned}
$$

Note: (a) : matrix inversion Lemma,
(b) : $\gamma(k) = \mathbf{h}(k)^H\boldsymbol{\Sigma}^{-1}\mathbf{h}(k)$

Equalizer Output: Final $z(k)$

$$z(k) = \left(1+\gamma(k)|\hat{s}(k)|^2\right)^{-1} \mathbf{h}(k)^H \boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{r}}(k) + \mathbf{h}(k)\hat{s}(k)),$$
$$= \left(1+\gamma(k)|\hat{s}(k)|^2\right)^{-1} \left(\mathbf{h}(k)^H \boldsymbol{\Sigma}^{-1}\tilde{\mathbf{r}}(k) + \mathbf{h}(k)^H \boldsymbol{\Sigma}^{-1}\mathbf{h}(k)\hat{s}(k)\right) \tag{52}$$

Finally, the equalizer output can be obtained as:

$$z(k) = \left(1+\gamma(k)|\hat{s}(k)|^2\right)^{-1} \left(\gamma(k)\hat{s}(k) + \mathbf{h}(k)^H \underset{(a)}{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{r}}(k)\right) \tag{53}$$

Sorting all element of $z(k)$ into a vector, we have:

$$\mathbf{z} = (\mathbf{I}_K + \boldsymbol{\Gamma}\mathbf{S})^{-1}\left[\boldsymbol{\Gamma}\hat{\mathbf{s}} + \mathbf{H}^H\boldsymbol{\Sigma}^{-1}\hat{\mathbf{s}}\right] \tag{54}$$

where $\boldsymbol{\Gamma} = \overline{\mathrm{diag}}\left[\mathbf{H}^H(\mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^H + \sigma^2\mathbf{I})^{-1}\mathbf{H}\right], \mathbf{S} = \mathrm{diag}[|\hat{\mathbf{s}}|^2]$. *Note:*

(a) : symbol-wise matrix inversion is NO LONGER required!, however, too complex

BPSK De-Mapper

- Filter Output

$$z(k) = \mathbf{w}(k)^H(\tilde{\mathbf{r}}(k) + \mathbf{h}(k)\hat{s}(k)) = \underbrace{\mu_z}_{\text{Mean}}(k)s(k) + \underbrace{v_z(k)}_{\text{Equivalent noise}} \quad (55)$$

- Mean: $\mu_z(k) = \mathbb{E}\{z(k)s(k)^*\} = \frac{\gamma(k)}{1+\gamma(k)|s(\hat{k})|^2}\mathbb{E}\{|s(k)|^2\}$
- Equivalent Noise:
  $\sigma_2(k)^2 = \mathbb{E}\{z(k)z(k)^*\} - \mu_z(k)^2 = \mu_z(k)(1 - \mu_z(k))$
- Log-likelihood ratio (LLR): Extrinsic information

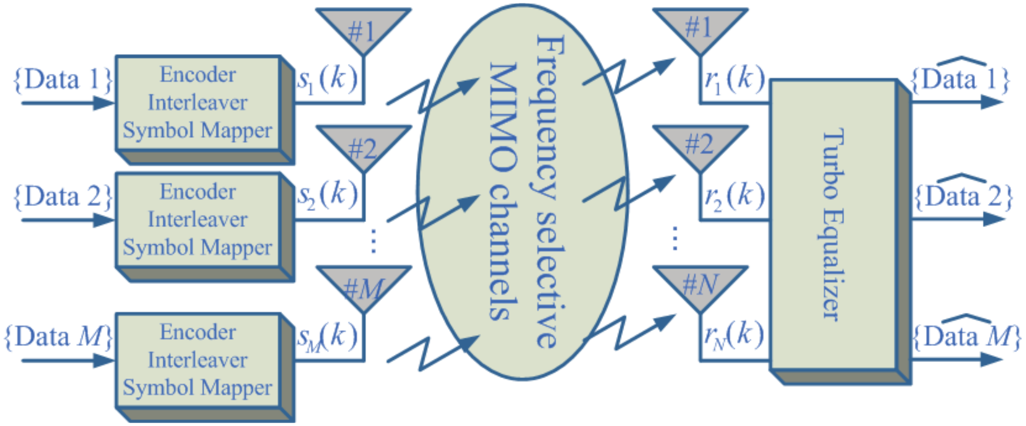
$$\begin{aligned} \lambda_1[s(k)] &= \ln\left(\frac{\Pr[z(k)|s(k) = +1]}{\Pr[z(k)|s(k) = -1]}\right) = \ln\left(\frac{\exp\left[-\frac{|\Re[z(k)]-\mu_z(k)|^2}{2\sigma_z(k)^2/2}\right]}{\exp\left[-\frac{|\Re[z(k)]+\mu_z(k)|^2}{2\sigma_z(k)^2/2}\right]}\right), \\ &= \frac{4\mu_z(k)\Re[z(k)]}{\mu_z(k)(1 - \mu_z(k))} = \frac{4\Re[z(k)]}{1 - \mu_z(k)} \end{aligned} \quad (56)$$

Frequency Domain Block-Wise Processing

Tad Matsumoto

E-mail: *tadashi.matsumoto@ee.oulu.fi*
matumoto@jaist.ac.jp

Channel Model (1): MIMO CP-Transmission



Recv. Signal (spatially sampled):

$$\mathbf{r}(k) = \sum_{l=0}^{L-1} \mathbf{H}(l)s(k-l) + \mathbf{v}(k)$$

Received Signal:

$$r_n(k) = \sum_{m=1}^{M}\sum_{l=0}^{L-1} h_{nm}(l)s_m(k-l) + v_n(k)$$

- $r_n(k)$: received signal
- L : channel memory length
- $h_{nm}(l)$: channel impulse response
- $s_m(k)$: transmitted signal
- $v_n(k)$: $\sim \mathcal{CN}(0, \sigma^2)$

$$\mathbf{r}(k) = [r_1(k), r_2(k), \cdots, r_N(k)]^T,$$

$$\mathbf{H}(l) = \begin{bmatrix} h_{11}(l) & \cdots & h_{1M}(l) \\ h_{21}(l) & \cdots & h_{2M}(l) \\ \vdots & \vdots & \vdots \\ h_{N1}(l) & \cdots & h_{NM}(l) \end{bmatrix}^T$$

$$\mathbf{s}(k) = [s_1(k), \ldots, s_M(k)]^T,$$

$$\mathbf{v}(k) = [v_1(k), \ldots, v_N(k)]^T.$$

Channel Model (2): MIMO CP-Transmission

- Block-wise Space-Time channel matrix

$$\mathbf{H} = \left[\mathbf{H}_1^T, \mathbf{H}_2^T, \cdots, \mathbf{H}_M^T\right] \in \mathbb{C}^{KN \times KM}, \quad (57)$$

$$\mathbf{H}_m = \left[\mathbf{H}_{1m}^T, \mathbf{H}_{2M}^T, \cdots, \mathbf{H}_{Nm}^T\right]^T, \quad (58)$$

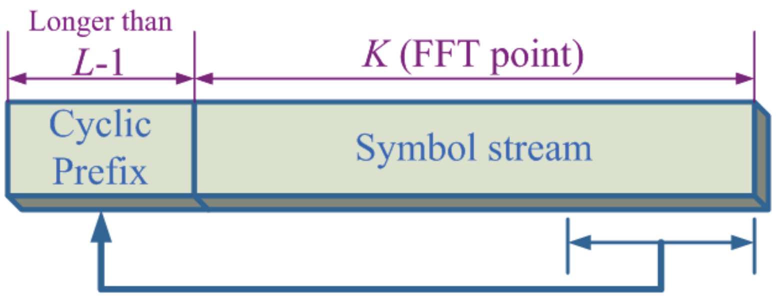
$$\mathbf{H}_{nm} = \left[\mathbf{h}_{n,m,1}, \mathbf{h}_{n,m,2}, \cdots, \mathbf{h}_{n,m,K}\right], \quad (59)$$

$$\mathbf{h}_{n,m,k} = \left[0_{k-1}, h_{nm}(0), h_{nm}(1), \cdots, h_{nm}(L-1), 0_{K-k+1}\right]^T. \quad (60)$$

- $\mathbf{H}_{nm}$ is Topelitz matrix where the last $L-1$ elements are removed.
- K: block length, 0_x: all zeros vector of length x
- Example of $\mathbf{H}_{nm} \in \mathbb{C}^{K \times K}$ with $L = 3$,

$$\mathbf{H}_{nm} = \begin{bmatrix} h_{nm}(0) & 0 & 0 & \cdots & 0 & 0 & 0 \\ h_{nm}(1) & h_{nm}(0) & 0 & \cdots & 0 & 0 & 0 \\ h_{nm}(2) & h_{nm}(1) & h_{nm}(0) & \cdots & 0 & 0 & 0 \\ 0 & h_{nm}(2) & h_{nm}(1) & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & h_{nm}(2) & h_{nm}(1) & h_{nm}(0) \end{bmatrix} \quad (61)$$

Channel Model (3): CP-Transmission



- Append cyclic prefix: Channel matrix becomes circulant, $\mathbf{H}^c_{nm}$.
- Example of Circulant Matrix for $L = 3$

$$\mathbf{H}^c_{nm} = \begin{bmatrix} h_{nm}(0) & 0 & 0 & 0 & h_{nm}(2) & h_{nm}(1) \\ h_{nm}(1) & h_{nm}(0) & 0 & 0 & 0 & h_{nm}(2) \\ h_{nm}(2) & h_{nm}(1) & h_{nm}(0) & 0 & 0 & 0 \\ 0 & h_{nm}(2) & h_{nm}(1) & h_{nm}(0) & 0 & 0 \\ 0 & 0 & h_{nm}(2) & h_{nm}(1) & h_{nm}(0) & 0 \\ 0 & 0 & 0 & h_{nm}(2) & h_{nm}(1) & h_{nm}(0) \end{bmatrix} \quad (62)$$

- $\mathbf{H}^c_m = \left[\mathbf{H}^{cT}_{1m}, \mathbf{H}^{cT}_{2m}, \cdots, \mathbf{H}^{cT}_{Nm}\right]^T$
- $\mathbf{H}^c = [\mathbf{H}^c_1, \mathbf{H}^c_2, \cdots, \mathbf{H}^c_M]$
- Because of the circularity, the channel matrix in frequency domain is expressed as diagonal matrix.

DFT Operator

- DFT Matrix: N_{DFT} is the number of DFT point (corresponds to K).

$$\mathbf{F} = \frac{1}{\sqrt{K}} \begin{bmatrix} e^{-j\frac{2\pi(0)(0)}{K}} & e^{-j\frac{2\pi(0)(1)}{K}} & \cdots & e^{-j\frac{2\pi(0)(K-1)}{K}} \\ e^{-j\frac{2\pi(1)(0)}{K}} & e^{-j\frac{2\pi(1)(1)}{K}} & \cdots & e^{-j\frac{2\pi(1)(K-1)}{K}} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-j\frac{2\pi(K-1)(0)}{K}} & e^{-j\frac{2\pi(K-1)(1)}{K}} & \cdots & e^{-j\frac{2\pi(K-1)(K-1)}{K}} \end{bmatrix} \quad (63)$$

- Frequency Domain Channel Matrix
  - Between m-th Tx and n-th Rx: $\mathbf{H}^c_{nm} = \mathbf{F}^H \mathbf{\Xi}^f_{nm} \mathbf{F}$
    $\mathbf{\Xi}^f_{nm}$ is a diagonal matrix.
  - Between all Tx and all Rx: $\mathbf{H}^c = \mathbf{F}^H_N \mathbf{\Xi}^f_{nm} \mathbf{F}_M$
    $\mathbf{\Xi}^f$ is a diagonal matrix, $\mathbf{F}_N = \mathbf{I}_N \otimes \mathbf{F}$, $\mathbf{F}_M = \mathbf{I}_M \otimes \mathbf{F}$, $\otimes$ is Kronecker product, $\mathbf{I}_x$ identity matrix of dimension x.
  - Between m-th Tx and all Rx: $\mathbf{H}^c_m = \mathbf{F}^H_N \mathbf{\Xi}^f_m \mathbf{F}$
    $\mathbf{\Xi}^f_m$ is a block diagonal matrix.

FD-Turbo Equalization Algorithm (1)

Filter output of block-wise time domain (MIMO):

$$\mathbf{z}_m = (\mathbf{I}_K + \mathbf{\Gamma}_m \mathbf{S}_m)^{-1} \left[\mathbf{\Gamma}_m \hat{\mathbf{s}}_m + \mathbf{H}_m^H \mathbf{\Sigma}^{-1} \tilde{\mathbf{r}}\right], \quad (64)$$

$$\mathbf{\Sigma} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}^H + \sigma^2 \mathbf{I}_{NK},\ \mathbf{\Gamma}_m = \overline{\mathrm{diag}}[\mathbf{H}_m^H \mathbf{\Sigma}^{-1} \mathbf{H}_m],\ \mathbf{S}_m = \mathrm{diag}\left[|\hat{s}_m|^2\right]$$

Frequency Domain $\Leftarrow$ Just by replacing $\mathbf{H}$ by $\mathbf{F}^H \mathbf{\Xi} \mathbf{F}$

$$\mathbf{\Sigma} = \mathbf{F}_N^H \mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} \mathbf{F}_N + \sigma^2 \mathbf{I}_{NK}, \quad (65)$$

$$\begin{aligned}
\mathbf{\Sigma}^{-1} &= \left[\mathbf{F}_N^H \mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} \mathbf{F}_N + \sigma^2 \mathbf{I}_{NK}\right]^{-1}, \\
&\overset{(a)}{=} \left[\mathbf{F}_N^H \left(\mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK}\right) \mathbf{F}_N\right], \\
&\overset{(b)}{=} (\mathbf{F}_N)^{-1} \left(\mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK}\right)^{-1} (\mathbf{F}_N^H)^{-1} \quad (66)
\end{aligned}$$

Note: $(a):\ \sigma^2 \mathbf{I}_{NK} = \mathbf{F}_N^H \sigma^2 \mathbf{I}_{NK} \mathbf{F}_N,$, $(b): (\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$,
$\mathbf{F}$ is a unitary matrix (no noise enhancement).

FD-Turbo Equalization Algorithm (2)

$$
\begin{aligned}
\mathbf{\Gamma}_m &= \overline{\mathrm{diag}}\left[\mathbf{H}_m^H \mathbf{\Sigma}^{-1} \mathbf{H}_m\right], \\
&= \overline{\mathrm{diag}}\left[\mathbf{F}^H \mathbf{\Xi}_m^{fH} \mathbf{F}_N \mathbf{\Sigma}^{-1} \mathbf{F}_N^H \mathbf{\Xi}_m^f \mathbf{F}\right], \\
&= \overline{\mathrm{diag}}\left[\mathbf{F}^H \mathbf{\Xi}_m^{fH} {\color{red}\mathbf{F}_N (\mathbf{F}_N)^{-1}} \left(\mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK}\right)^{-1} {\color{red}(\mathbf{F}_N^H)^{-1} \mathbf{F}_N^H} \mathbf{\Xi}_m^f \mathbf{F}\right], \\
&= \overline{\mathrm{diag}}\left[\mathbf{F}^H \mathbf{\Xi}_m^{fH} \left(\mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK}\right)^{-1} \mathbf{\Xi}_m^f \mathbf{F}\right] \qquad (67)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbf{z}_m &= (\mathbf{I}_K + \mathbf{\Gamma}_m \mathbf{S}_m)^{-1} [\mathbf{\Gamma}_m \hat{\mathbf{s}}_m + \mathbf{H}_m^H \mathbf{\Sigma}^{-1} \tilde{\mathbf{r}}], \\
&= (\mathbf{I}_K + \mathbf{\Gamma}_m \mathbf{S}_m)^{-1} [\mathbf{\Gamma}_m \hat{\mathbf{s}}_m + \mathbf{F}^H \mathbf{\Xi}_m^{fH} {\color{red}\mathbf{F}_N (\mathbf{F}_N)^{-1}} \\
&\qquad (\mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK})^{-1} (\mathbf{F}_N^H)^{-1} \tilde{\mathbf{r}}], \\
&\overset{(a)}{=} (\mathbf{I}_K + \mathbf{\Gamma}_m \mathbf{S}_m)^{-1} [\mathbf{\Gamma}_m \hat{\mathbf{s}}_m + \mathbf{F}^H \mathbf{\Xi}_m^{fH} (\underline{\mathbf{\Xi}^f \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK}})^{-1} \tilde{\mathbf{r}}^f] \\
&\qquad (68)
\end{aligned}
$$

Note: (a) : $\tilde{\mathbf{r}}^f = \mathbf{F}_N \tilde{\mathbf{r}}$, underlined part → computation is still too heavy.

FD-Turbo Equalization Algorithm: Approximation 1

General Formula: If $\mathbf{\Lambda}$ a diagonal matrix, $\mathbf{\Delta}$ block circulant matrix.[a]

[a] In the case that the matrix size (or DFT size) is large.

$$\mathbf{\Delta} = \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H \tag{69}$$

Approximation 1:

$\mathbf{\Delta} \approx \mathrm{diag}[\overline{\lambda}_1 \mathbf{I}_K, \overline{\lambda}_2 \mathbf{I}_K, \ldots, \overline{\lambda}_M \mathbf{I}_K]$ with $\overline{\lambda}_m = \frac{1}{K}\mathrm{tr}\mathbf{\Delta}_m$ and $\mathbf{\Delta}_m = \mathbf{F}\mathbf{\Lambda}_m\mathbf{F}^H$, where $\mathbf{\Lambda}_m$ is the m-th block of $\mathbf{\Lambda}$, therefore

$$\mathrm{tr}\{\mathbf{\Delta}_m\} = \mathrm{tr}\{\mathbf{\Lambda}_m\}. \tag{70}$$

Finally,

$$\begin{aligned} \mathbf{z}_m &= (\mathbf{I}_K + \mathbf{\Gamma}_m \mathbf{S}_m)^{-1} \\ &\cdot \left[\mathbf{\Gamma}_m \hat{\mathbf{s}}_m + \mathbf{F}^H \mathbf{\Xi}_m^{fH} \left(\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK} \right)^{-1} \tilde{\mathbf{r}}^f \right] \end{aligned} \tag{71}$$

FD-Turbo Equalization Algorithm: Approximation 2

About $\mathbf{\Gamma}_m$:

$$\mathbf{\Gamma}_m \approx \overline{\mathrm{diag}} \left[\mathbf{F}^H \mathbf{\Xi}_m^{fH} \left(\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{\Xi}_m^f \mathbf{F} \right] \tag{72}$$

Approximation 2

$$\begin{aligned} \mathbf{\Gamma}_m &\approx \mathbf{F}^H \overline{\mathrm{diag}} \left[\mathbf{\Xi}_m^{fH} \left(\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{\Xi}_m^f \right] \mathbf{F}, \\ &= \frac{1}{K} \mathrm{tr} \left[\mathbf{\Xi}_m^{fH} \left(\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK} \right)^{-1} \mathbf{\Xi}_m^f \right] \mathbf{I}_K \approx \overline{\gamma}_m \mathbf{I}_K \end{aligned} \tag{73}$$

and

$$\mathbf{S}_m \approx \frac{1}{K} \sum_{k=1}^{K} |\hat{s}_m(k)|^2 \mathbf{I}_K = (1 - \overline{\lambda}_m) \mathbf{I}_K = \overline{\delta}_m \mathbf{I}_K \tag{74}$$

FD-Turbo Equalization: Summary

- Frequency domain Turbo equalization output

$$\mathbf{z}_m = (1 + \overline{\gamma}_m \overline{\delta}_m)^{-1} \left[\overline{\gamma}_m \hat{\mathbf{s}}_m + \mathbf{F}^H \mathbf{\Psi}_m \tilde{\mathbf{r}}^f \right], \quad (75)$$

where

$$\overline{\gamma}_m = \frac{1}{K} \mathrm{tr} \left[\mathbf{\Xi}_m^{fH} \left(\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK} \right)^{-1} \mathbf{\Xi}_m^f \right], \quad (76)$$

$$\overline{\delta}_m = \frac{1}{K} \sum_{k=1}^{K} |\hat{s}_m(k)|^2 = (1 - \overline{\lambda}_m), \quad (77)$$

$$\mathbf{\Psi}_m = \mathbf{\Xi}_m^{fH} \left(\mathbf{\Xi}^f \mathbf{\Delta} \mathbf{\Xi}^{fH} + \sigma^2 \mathbf{I}_{NK} \right)^{-1}. \quad (78)$$

- First and second moments of the equalizer output

$$\mu_{z,m} = \overline{\gamma}_m (1 + \overline{\gamma}_m \overline{\delta}_m)^{-1}, \quad (79)$$

$$\sigma_{z,m}^2 = \mu_{z,m}(1 - \mu_{z,m}). \quad (80)$$

Complexity is Independent of the Path Number L

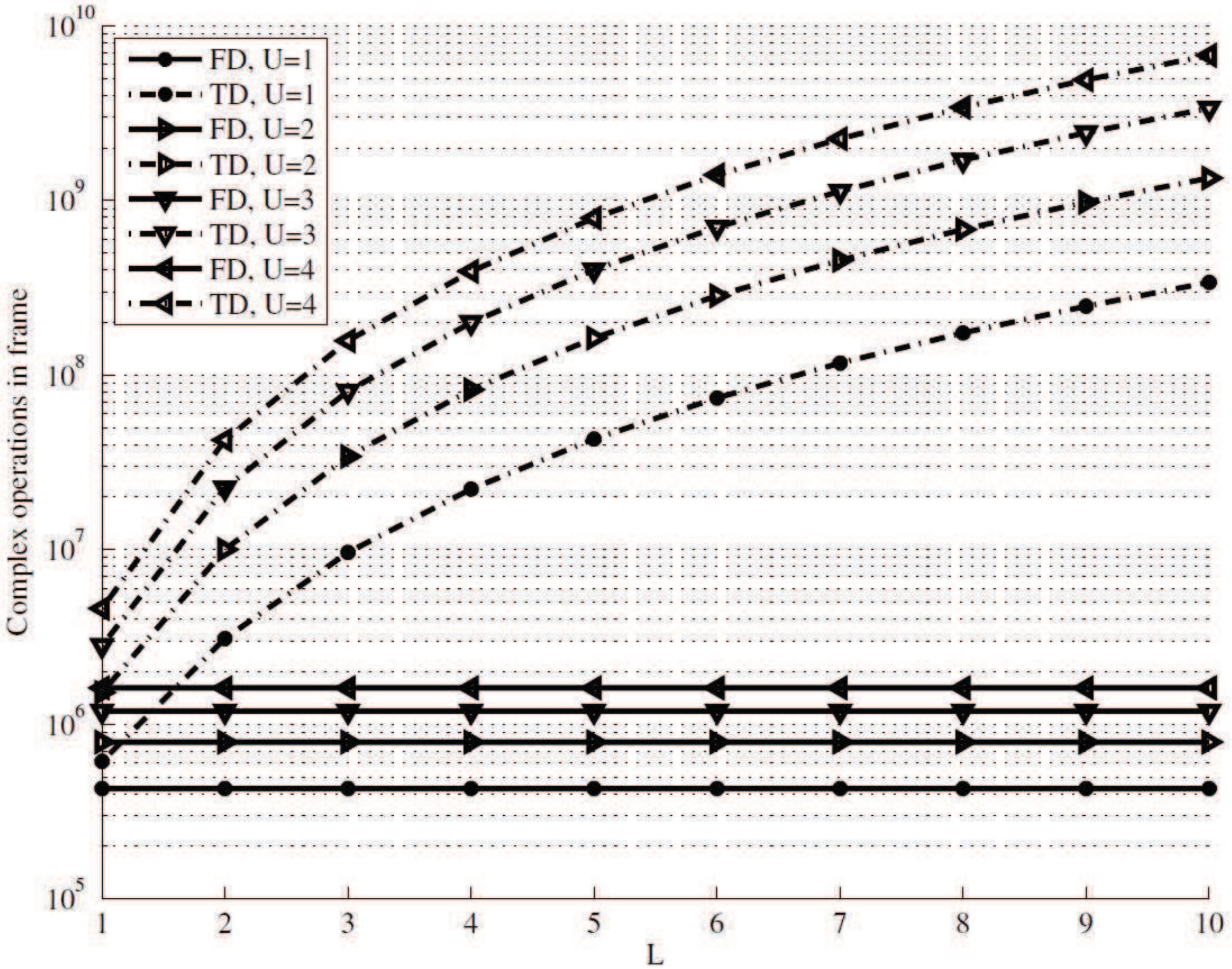


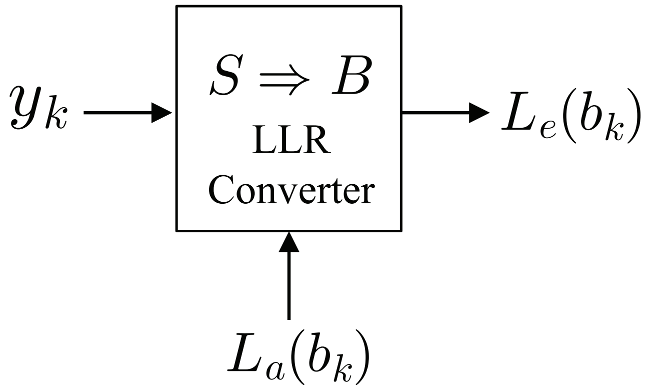
Figure 1: Each user uses 2 Tx antennas and Rx uses 4 antennas with $K = 512$. U is the number of users, TD is time-domain, and FD is frequency domain. Source: J. Karjalainen, N. Veselinović, K. Kansanen, and T. Matsumoto, "*Iterative Frequency Domain Joint-over-Antenna Detection in Multiuser MIMO*", IEEE Trans on Wireless Comm., vol. 6, no. 10, Oct. 2007, pp. 3620–3630.

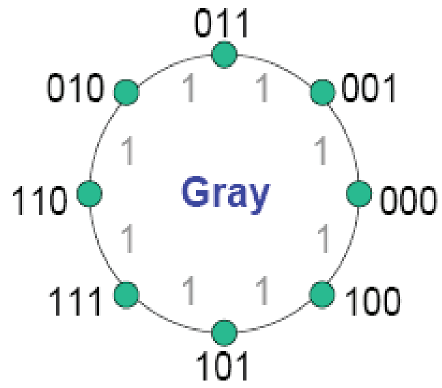
Turbo Equalization for Higher Order Modulation

Tad Matsumoto

E-mail: *tadashi.matsumoto@ee.oulu.fi*
matumoto@jaist.ac.jp

Demapping: Symbol-to-Bit LLR Conversion (1)





$y_k = hs_k + n_k$, n_k is k-th noise component, $L_a(b_k)$ is a priori knowledge of the related symbols, and $L_e(b_k)$ is exrinsic bit-wise LLR.

Note: The following equations hold for arbitrarily mapping rules. Each symbols s_k is constructed by the mapping rule

$$s_k = \mathrm{MAP}\{\mathbf{b} = [b_1, b_2, \ldots, b_\mu, \ldots, b_m]\} \tag{81}$$

$$L[b_\mu(k)] = \ln \frac{\Pr(b_\mu = 0|y_k)}{\Pr(b_\mu = 1|y_k)} = \ln \frac{\Pr(b_\mu = 0, y_k)}{\Pr(b_\mu = 1, y_k)} = \ln \frac{\sum_{\mathbf{b} \in B_0} \Pr(y|\mathbf{b}) \Pr(\mathbf{b})}{\sum_{\mathbf{b} \in B_1} \Pr(y|\mathbf{b}) \Pr(\mathbf{b})} \tag{82}$$

where the sets B_0 and B_1 of the bit vector $\mathbf{b}$ that yield the μ-th bit b_μ being "0" and "1", respectively.

Demapping: Symbol-to-bit LLR Conversion (2)

Property of (*a priori*) LLR L

$$L = \ln \frac{\Pr(b=0)}{\Pr(b=1)} \tag{83}$$

and[1]

$$\Pr(b=1) + \Pr(b=0) = 1 \tag{84}$$

therefore,

$$\Pr(b=0) = \frac{1}{1+e^{-L}}, \tag{85}$$

$$\Pr(b=1) = \frac{e^{-L}}{1+e^{-L}}. \tag{86}$$

In general

$$\Pr(b) = \frac{e^{-bL}}{1+e^{-L}} \tag{87}$$

[1] Note that we can also write it as: $-L = \frac{\Pr(b=1)}{\Pr(b=0)}$.

Demapping: Symbol-to-bit LLR Conversion (3)

$$
\begin{aligned}
L[b_\mu(k)] &= \ln \frac{\sum_{\mathbf{b}\in B_0} \Pr(y|\mathbf{b}) \Pr(\mathbf{b})}{\sum_{\mathbf{b}\in B_1} \Pr(y|\mathbf{b}) \Pr(\mathbf{b}}, \\
&= \ln \frac{\sum_{s\in S_0} \exp\{-\frac{|y-hs|^2}{\sigma_N^2}\} \Pr(s)}{\sum_{s\in S_1} \exp\{-\frac{|y-hs|^2}{\sigma_N^2}\} \Pr(s)}, \\
&= \ln \frac{\sum_{s\in S_0} \exp\{-\frac{|y-hs|}{\sigma_N^2}\} \prod_{\nu=1}^{m} \Pr(b_\nu(s))}{\sum_{s\in S_1} \exp\{-\frac{|y-hs|}{\sigma_N^2}\} \prod_{\nu=1}^{m} \Pr(b_\nu(s))} \qquad (88)
\end{aligned}
$$

where the sets S_0 and S_1 of symbols s that yield the μ-th bit b_μ being "0" and "1", respectively, and $b_\nu(s)$ is the ν-th bit of the vector $\mathbf{b}$ corresponding to the symbol s. h is the channel gain.

Demapping: Symbol-to-bit LLR Conversion (4)

Using the LLR property:

$$
\begin{aligned}
L[b_\mu(k)] &= \ln \frac{\sum_{s\in S_0} \Pr(y|s) \prod_{\nu=1}^{m} \Pr(b_\nu(s))}{\sum_{s\in S_1} \Pr(y|s) \prod_{\nu=1}^{m} \Pr(b_\nu(s))}, \\
&= \ln \frac{\sum_{s\in S_0} \Pr(y|s) \prod_{\nu=1}^{m} \frac{e^{-c_\nu(s)L_a[c_\nu(s)]}}{1+e^{-L_a[c_\nu(s)]}}}{\sum_{s\in S_1} \Pr(y|s) \prod_{\nu=1}^{m} \frac{e^{-c_\nu(s)L_a[c_\nu(s)]}}{1+e^{-L_a[c_\nu(s)]}}}, \\
&= \ln \frac{\sum\limits_{s\in S_0} \exp\{-\frac{|y-hs|^2}{\sigma_N^2}\} \prod\limits_{\nu=1}^{m} e^{-c_\nu(s)L_a[c_\nu(s)]}}{\sum\limits_{s\in S_1} \exp - \frac{|y-hs|^2}{\sigma_N^2}\} \prod\limits_{\nu=1}^{m} e^{-c_\nu(s)L_a[c_\nu(s)]}}
\end{aligned} \quad (89)
$$

where $c_\nu(s) = 1 - 2b_\nu(s)$.

Demapping: Symbol-to-bit LLR Conversion (5)

Therefore,

$$
\begin{aligned}
L_e[b_\mu(k)] &= L[b_\mu(k)] - L_a[b_\mu(k)], \\
&= \ln \frac{\displaystyle\sum_{s\in S_0} \exp\{-\frac{|y-hs|^2}{\sigma_N^2}\} \prod_{\nu=1,\nu\neq\mu}^{m} e^{-c_\nu(s)L_a[c_\nu(s)]}}{\displaystyle\sum_{s\in S_1} \exp -\frac{|y-hs|^2}{\sigma_N^2}\} \prod_{\nu=1,\nu\neq\mu}^{m} e^{-c_\nu(s)L_a[c_\nu(s)]}} \qquad (90)
\end{aligned}
$$

Note:

The mapper receives "symbol" containing a bit b_μ. Therefore, we need $\nu \neq \mu$ to calculate extrinsic LLR.

Soft Mapping: MMSE Equalization for QAM

For generic modulation format,

$$\Delta(k) = \mathrm{diag}\{\mathbb{E}\left[|s_k|^2\right] - |\hat{s}(k)|^2\}, \tag{91}$$

$$\hat{s}_k = \mathbb{E}[s_k] = \sum_{s_i \in S} s_i \Pr(s_k = s_i), \tag{92}$$

$$\mathbb{E}[|s_k|^2] = \sum_{s_i \in S} |s_i|^2 \Pr(s_k = s_i). \tag{93}$$

with

$$\Pr(s_k = s_i) = \frac{1}{2^M} \prod_{m=1}^{M} \left(1 - \hat{c}_m(s_i) \tanh\left\{\frac{L_a[b_m(s_i)]}{2}\right\}\right) \tag{94}$$

and

$$\hat{c}_m(s_i) = 1 - 2b_m(s_i), \tag{95}$$

where $b_m(s_i)$ is the m-th bit of the symbol s_i. M is the modulation multiplicity.

Convergence Property

Tad Matsumoto

E-mail: *tadashi.matsumoto@ee.oulu.fi*
matumoto@jaist.ac.jp

Our Founding Father Said:

1948: Establishment of Information Theory

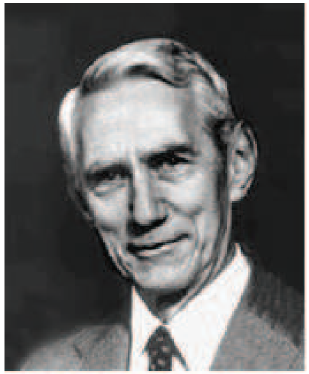
Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

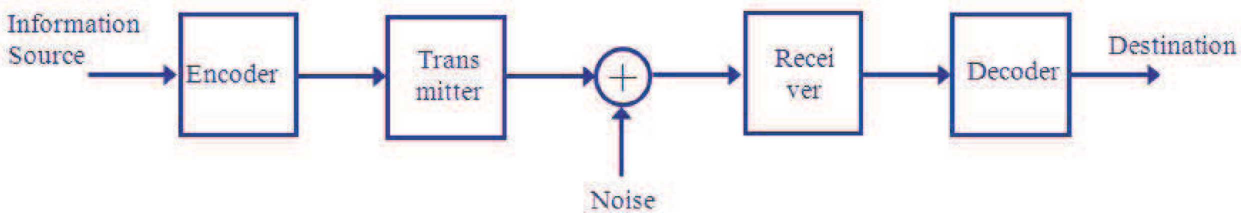
By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.



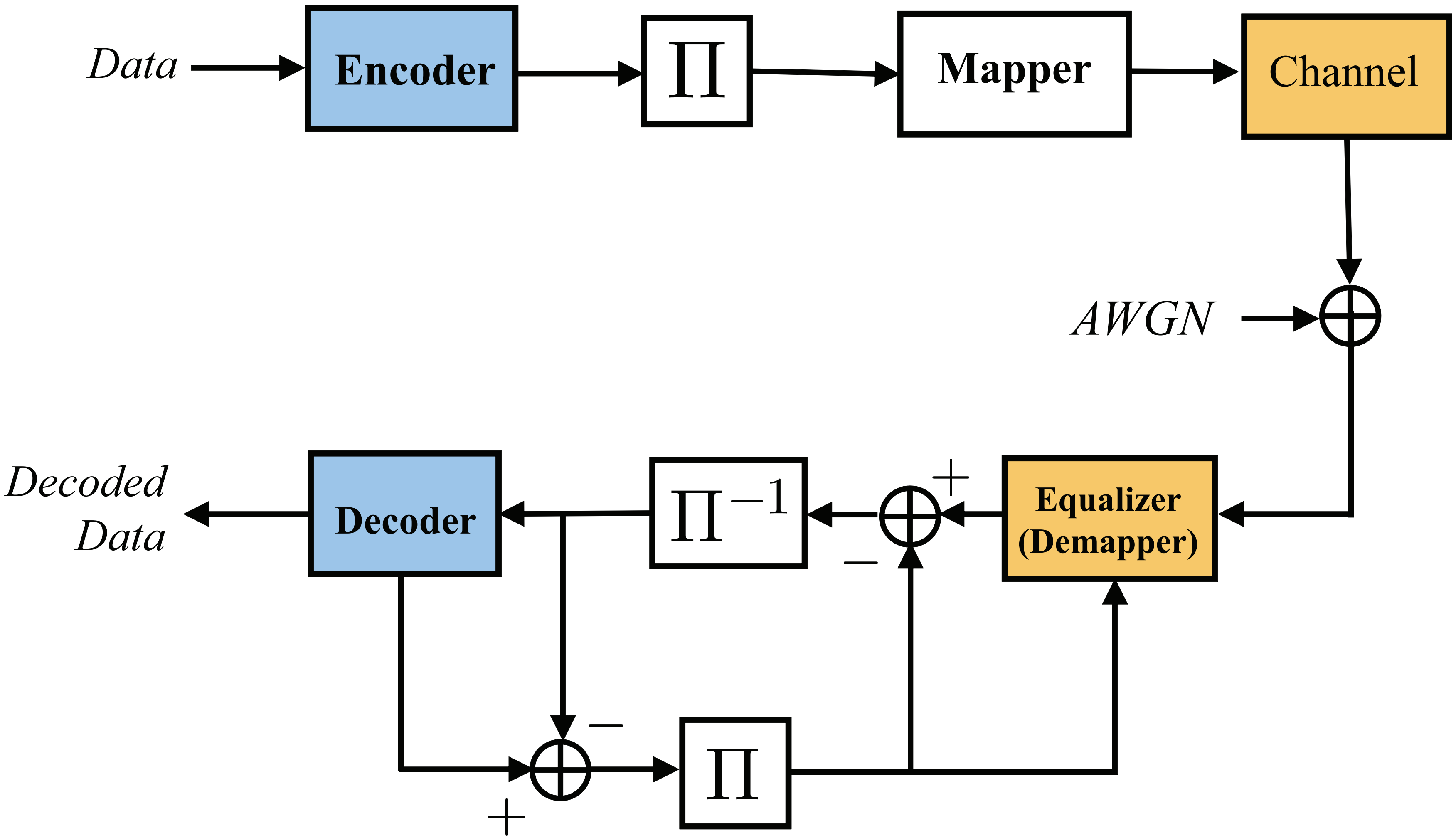
Claude Shannon



Channel Coding Theorem:

1. There exist a length n rate R code $(2^{nR}, n)$ such that the error probability can be made arbitrarily small, if the code rate is lower than the capacity, $R < C$.
2. Conversely, any $(2^{nR}, n)$ rate R code that can achieve arbitrarily small error must satisfy $R < C$.

Turbo Concept



Use very symple codes allows us to use a lot of mathematical tools.

Is BER $\to 0$ Always Guaranteed?

- Fact: "*Conditioning reduces Entropy*" $H(X) \geq H(X|Y)$
- Turbo feedback works as conditioning with LLR $L_i, i \in \{1, 2, \ldots\}$.

$$H(X) \geq H(X|Y) \geq H(X|Y, L_1) \geq H(X|Y, L_1, L_2) \geq \cdots \quad (96)$$

- Mutual information between transmitted coded bit and the LLR

$$I(X;Y) = H(X) - H(X|Y) \quad (97)$$

and

$$I(X;Y, L_1) = H(X) - H(X|Y, L_1) \geq H(X) - H(X|Y), \quad (98)$$

$$I(X;Y, L_1, L_2) = H(X) - H(X|Y, L_1, L_2) \geq H(X) - H(X|Y), \quad (99)$$

$$\vdots$$

$$I(X;Y, L_1, .., L_N) = H(X) - H(X|Y, L_1, .., L_N) \geq H(X) - H(X|Y). \quad (100)$$

This means that:

$$I(X;Y) \leq I(X;Y, L_1, L_2) \leq \cdots \leq I(X;Y, L_1, L_2, ..., L_N) \to 1 \quad (101)$$

$$\text{if} \quad H(X|Y, L_1, ..., L_N) \to 0 \quad (102)$$

EXIT Charts: Equalizer

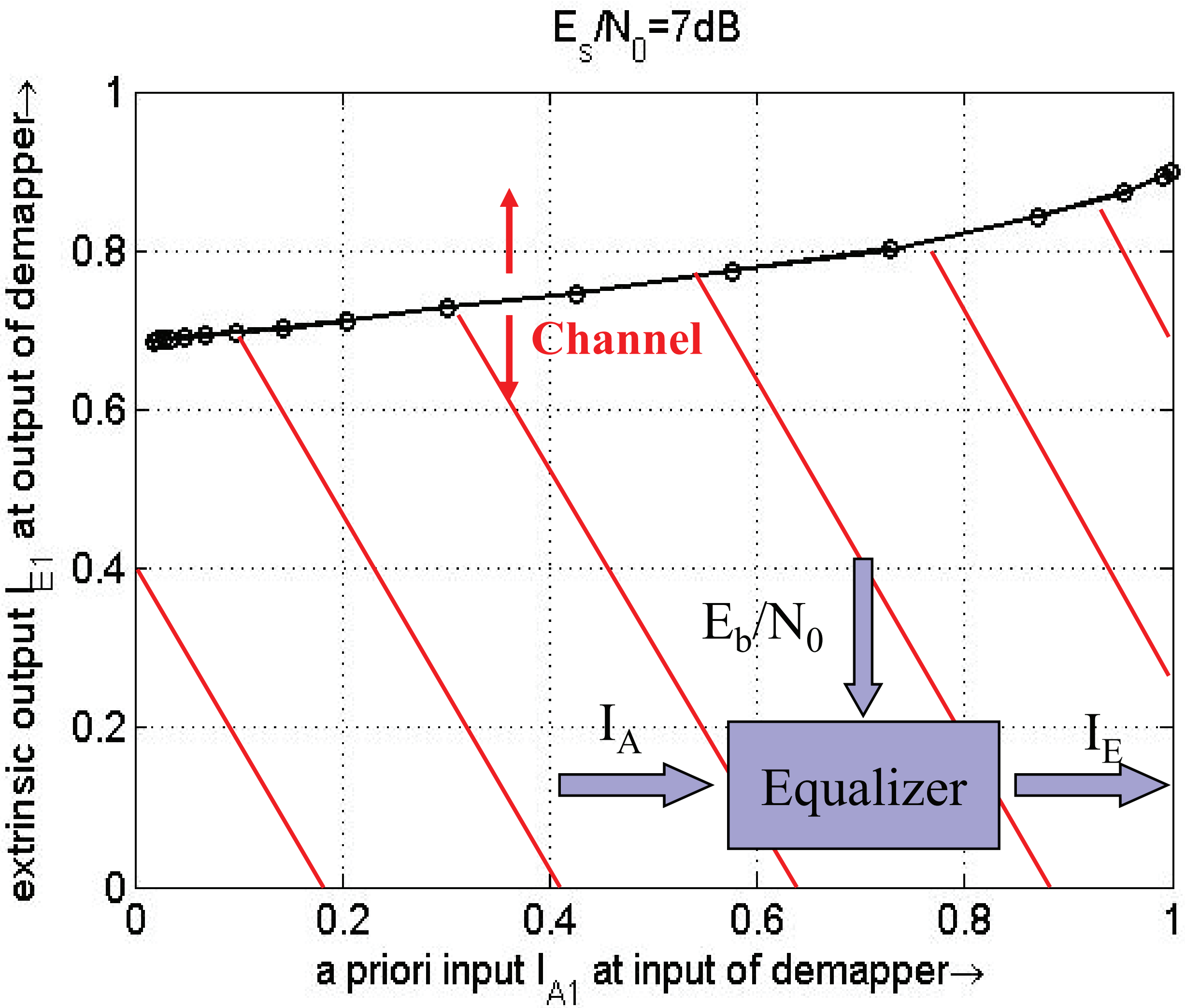


Figure 2: Area under the EXIT curve depends on the constellation constrained capacity

EXIT Charts: Decoder

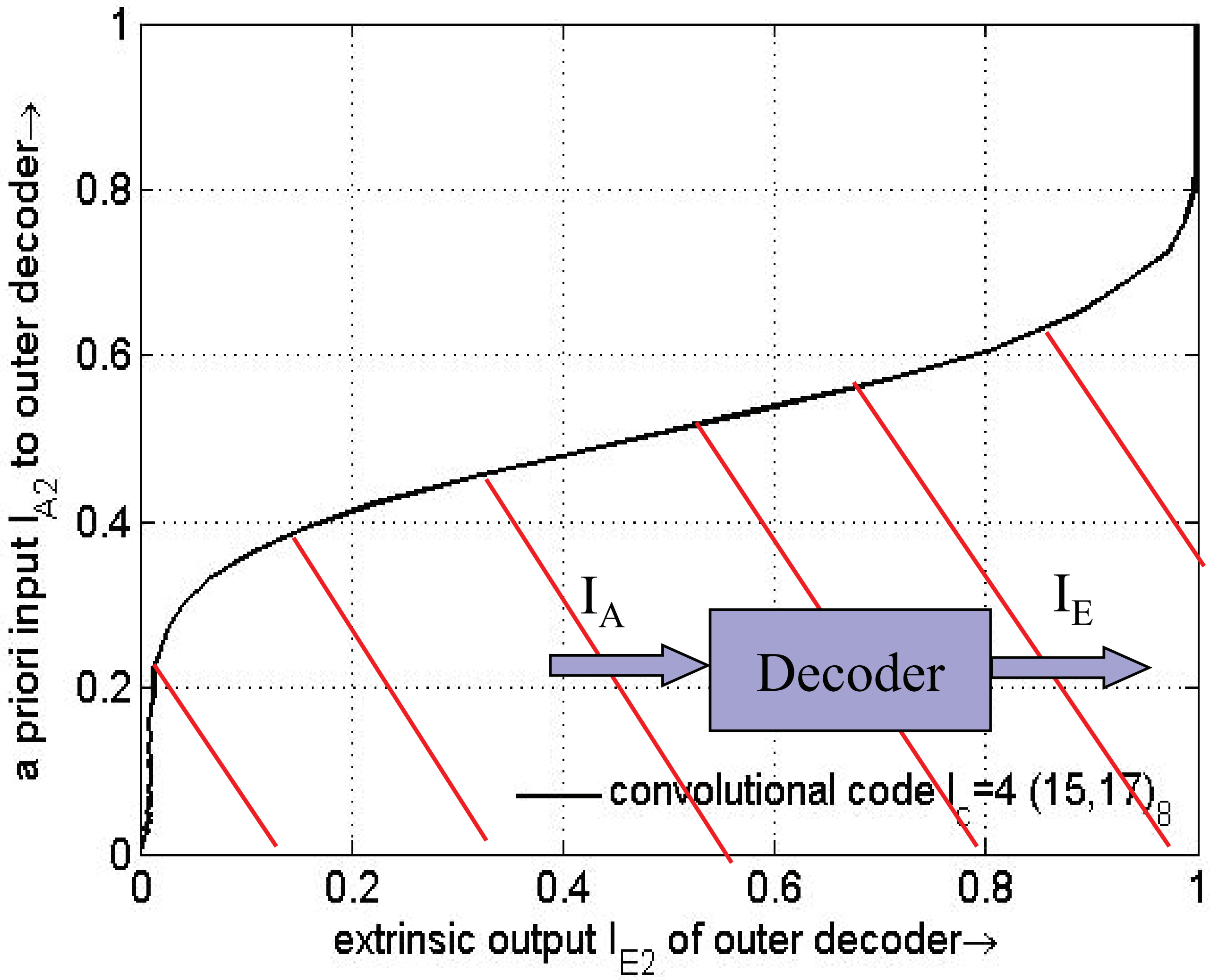


Figure 3: Area under the EXIT equals to the channel coding rate of the used code

EXIT Charts: Exchange Between Equalizer and Decoder

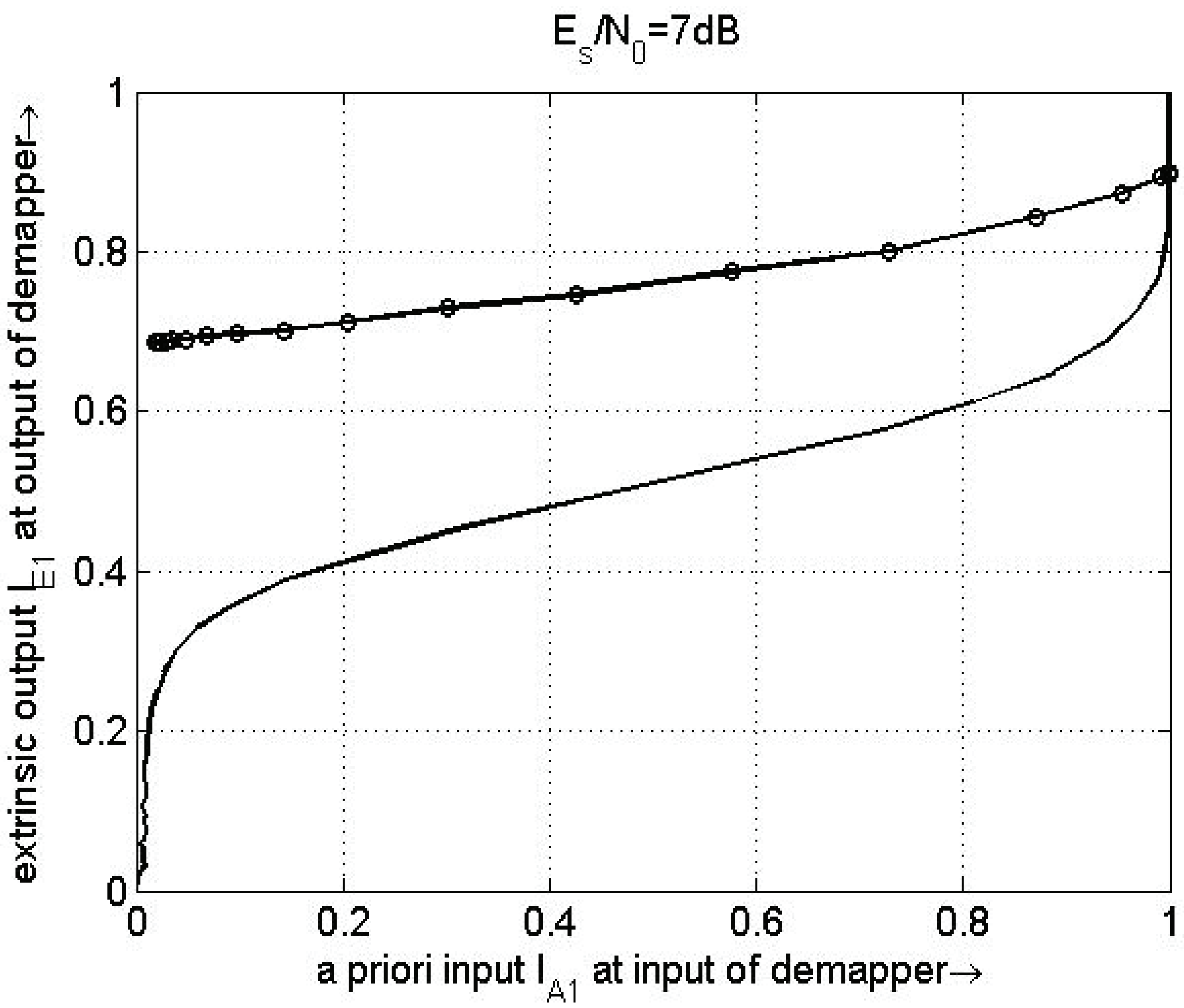


Figure 4: If the both curve do not intersect, the received sequence can be iteratively decoded without any bit-error

EXIT Chart Tells Us the Convergence Story

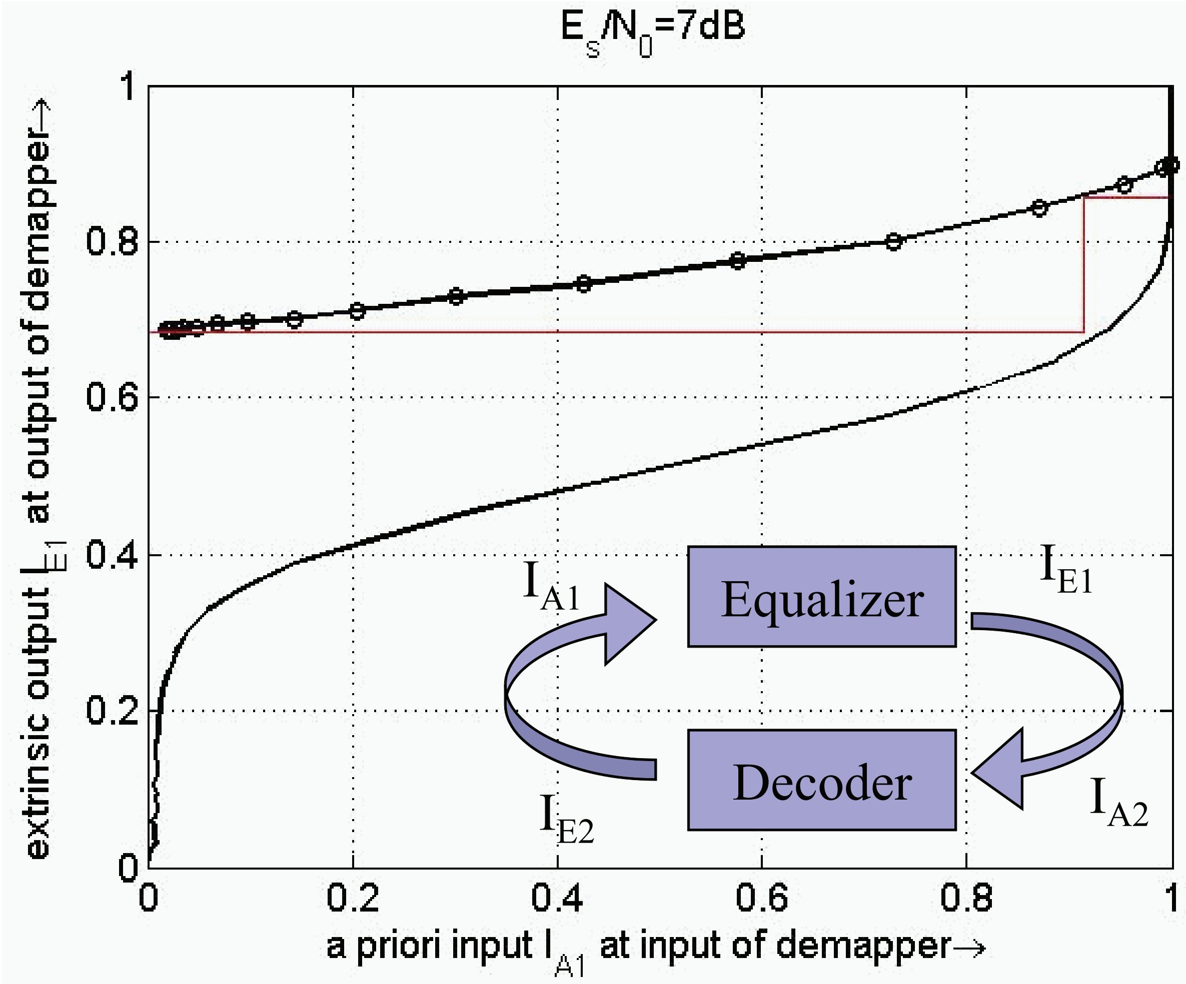


Figure 5: The area under the decoder EXIT curve should be smaller than the area under the demapper EXIT curve to avoid intersection. Curve Fitting: Minimize area between both EXIT curves.

In IEEE ICC 2011, Kyoto,...

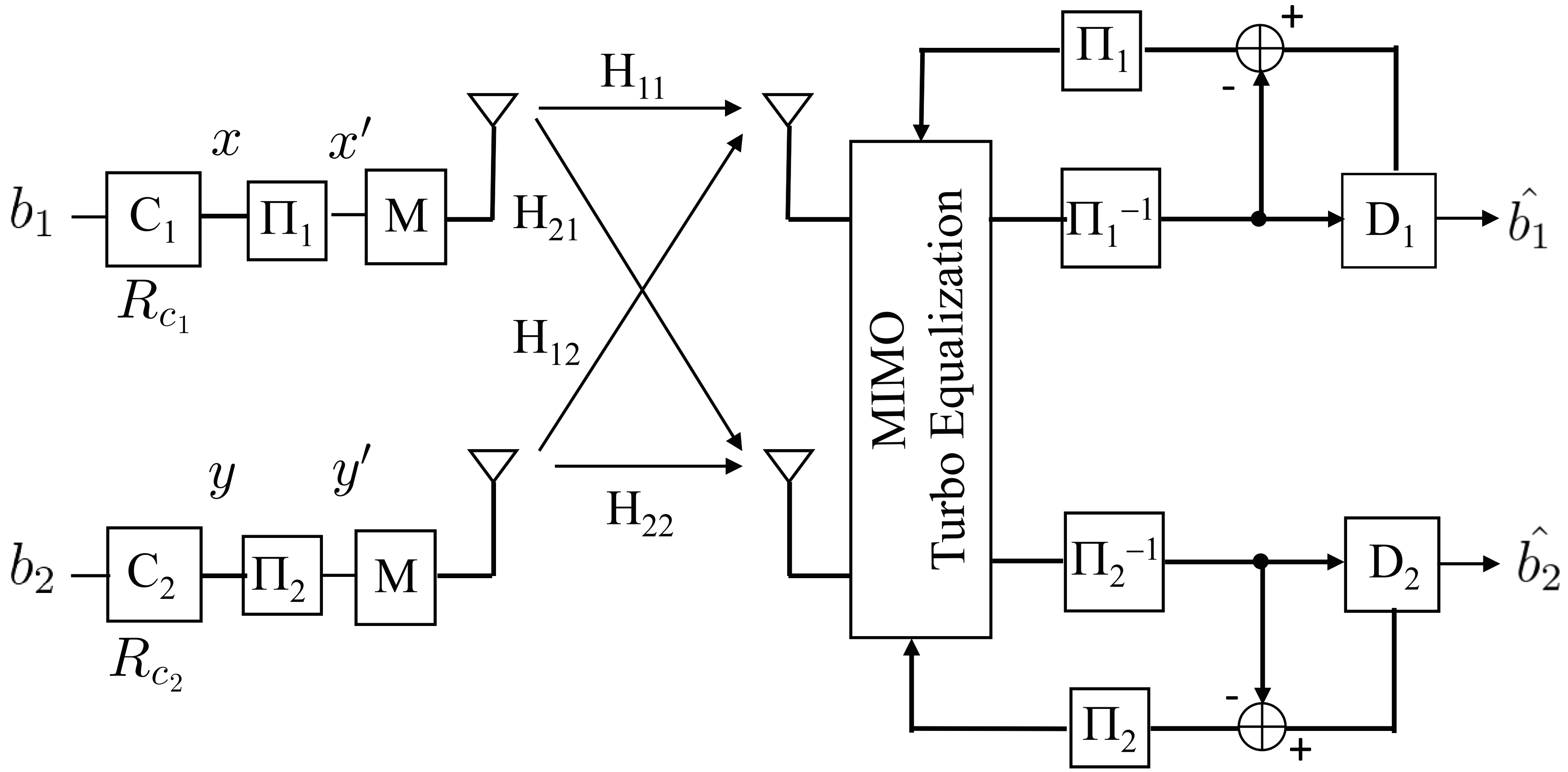
A student of a famous professor came to my place and asked me a question...

"Why the gain is saturated within 1-2 iterations, even though I am using very strong Turbo code?..."

...

"The issue is NOT the code strength, but the matching of the convergence property."

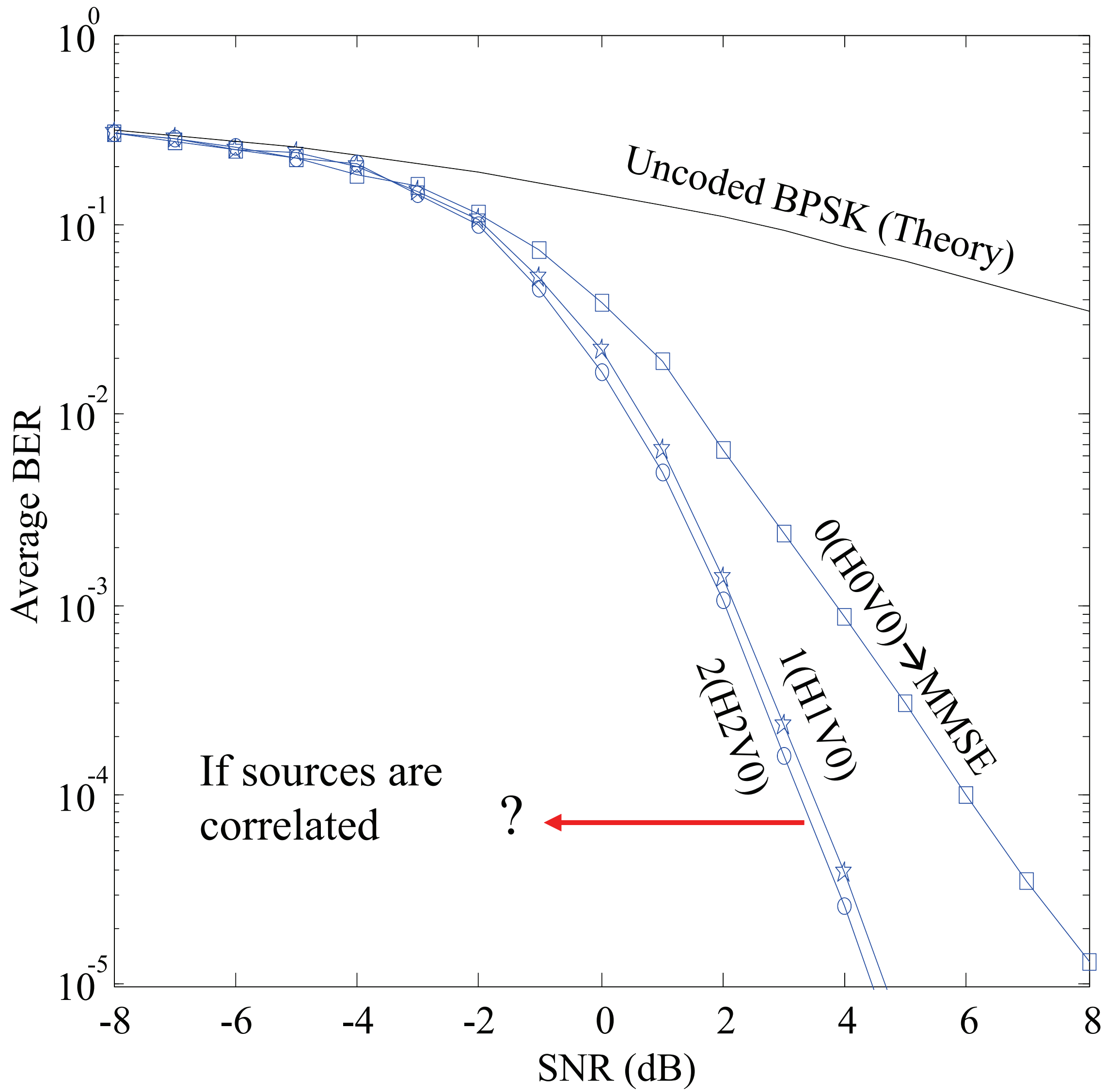
MU-SIMO: Revisit



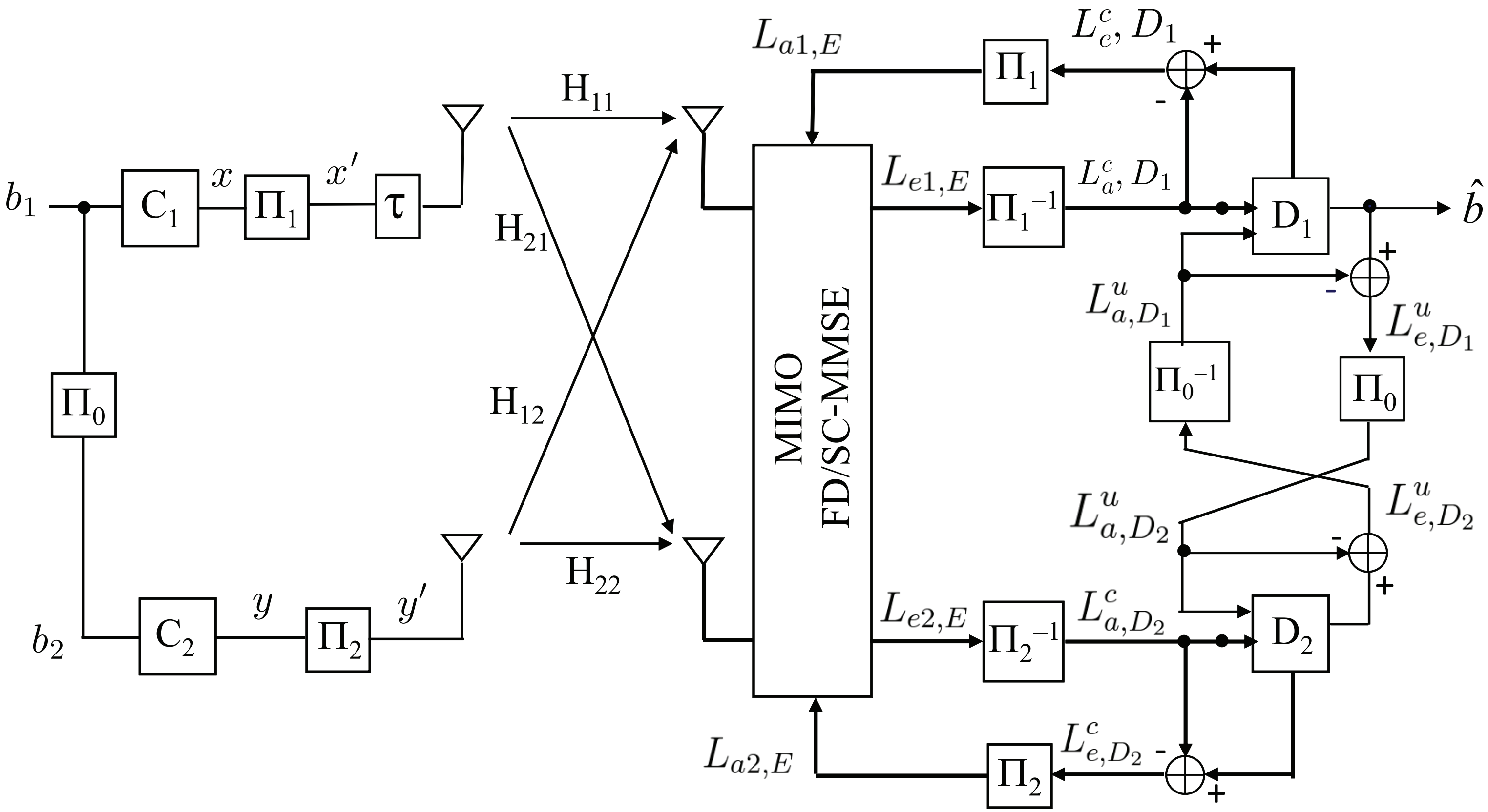
- Block at m-th transmit antenna: $\mathbf{s}_m = [s_m(1), s_m(2), \ldots, s_m(K)]^T$
- Block from all transmit antennas: $\mathbf{s} = [s_1, s_2, \ldots, s_M]^T$

Performance of Multiuser MIMO FD/SC-MMSE

- MIMO FD/SC-MMSE jointly decode information from each user.
- The decoding is performed separately for each user.
- Encoder: NSNRCC $4(17, 15)$, FFT$=$512, MIMO 2×2, Rayleigh 64-path, Decoder: BCJR Log-MAP.
- Ref.: K. Anwar and T. Matsumoto, "MIMO Spatial Turbo Coding with Iterative Equalization", ITG Workshop on Smart Antennas (WSA), Bremen, Germany, Feb. 2010.



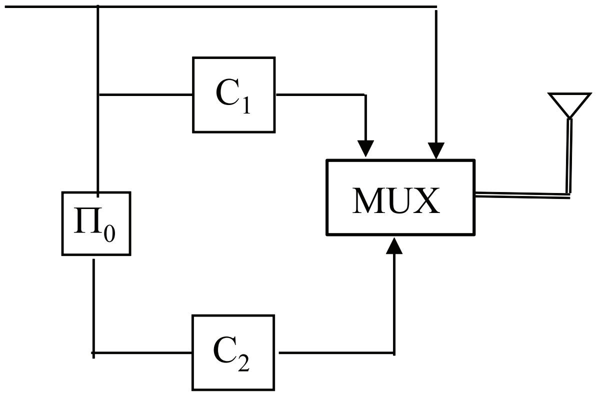
Idea: Vertical Iterations (Spatial Turbo Codes)

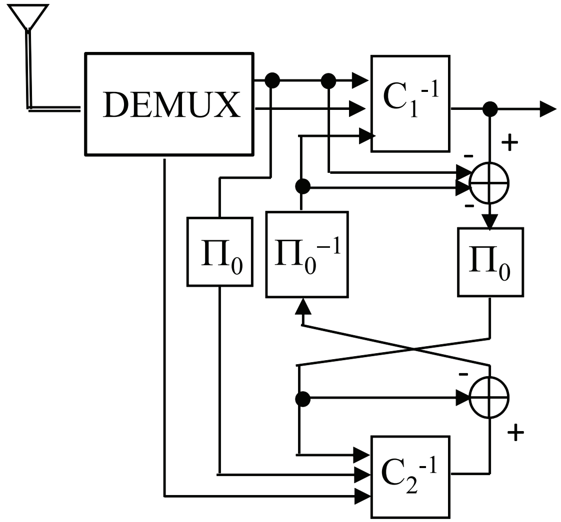


- Vertical iteration is expected to improve the performance because of (1) space diversity utilization and (2) coding gain.
- The design above is called spatial turbo codes (STC) since the output of the second decoder is not multiplexed (in time) but transmitted in parallel over the space.

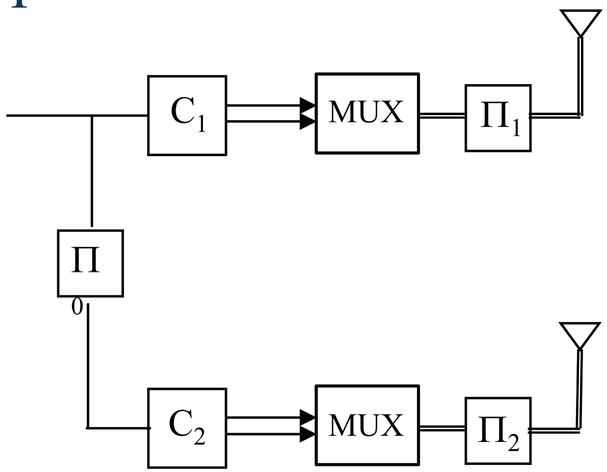
Comparing STC with Turbo Codes

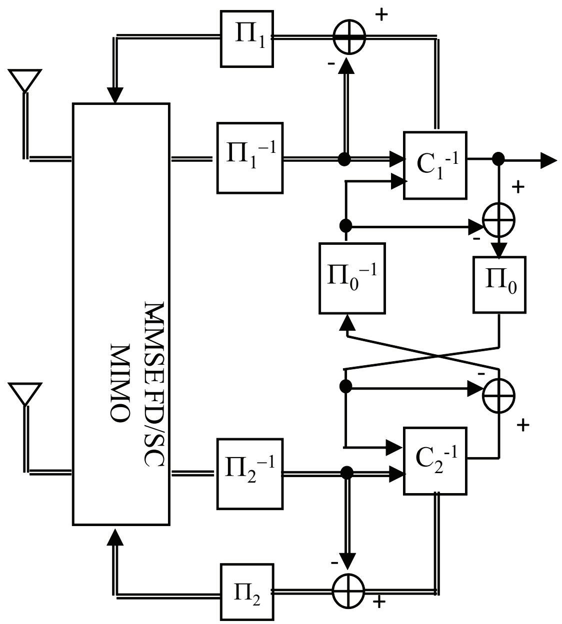
Turbo Code



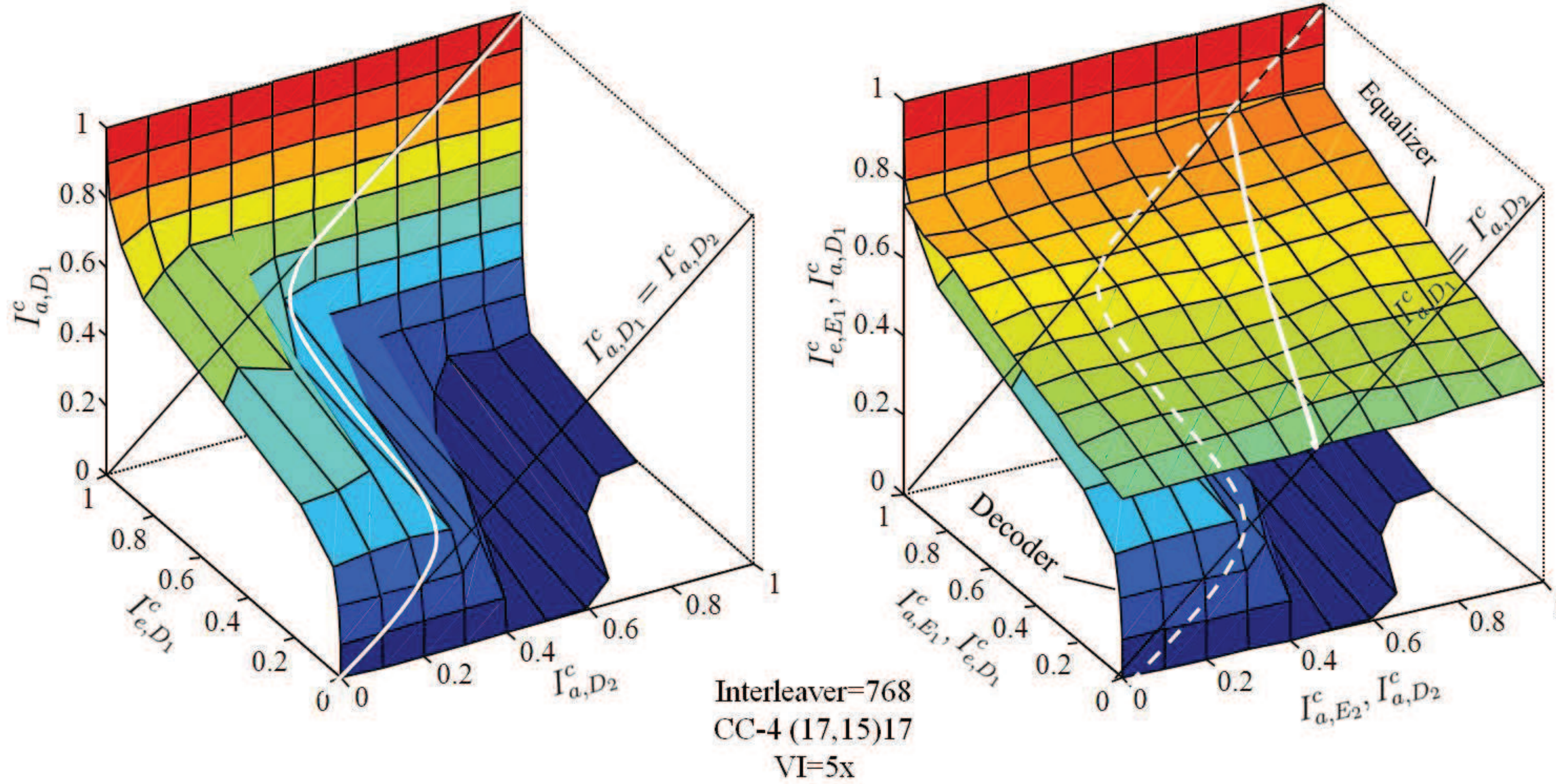


Spatial Turbo Code





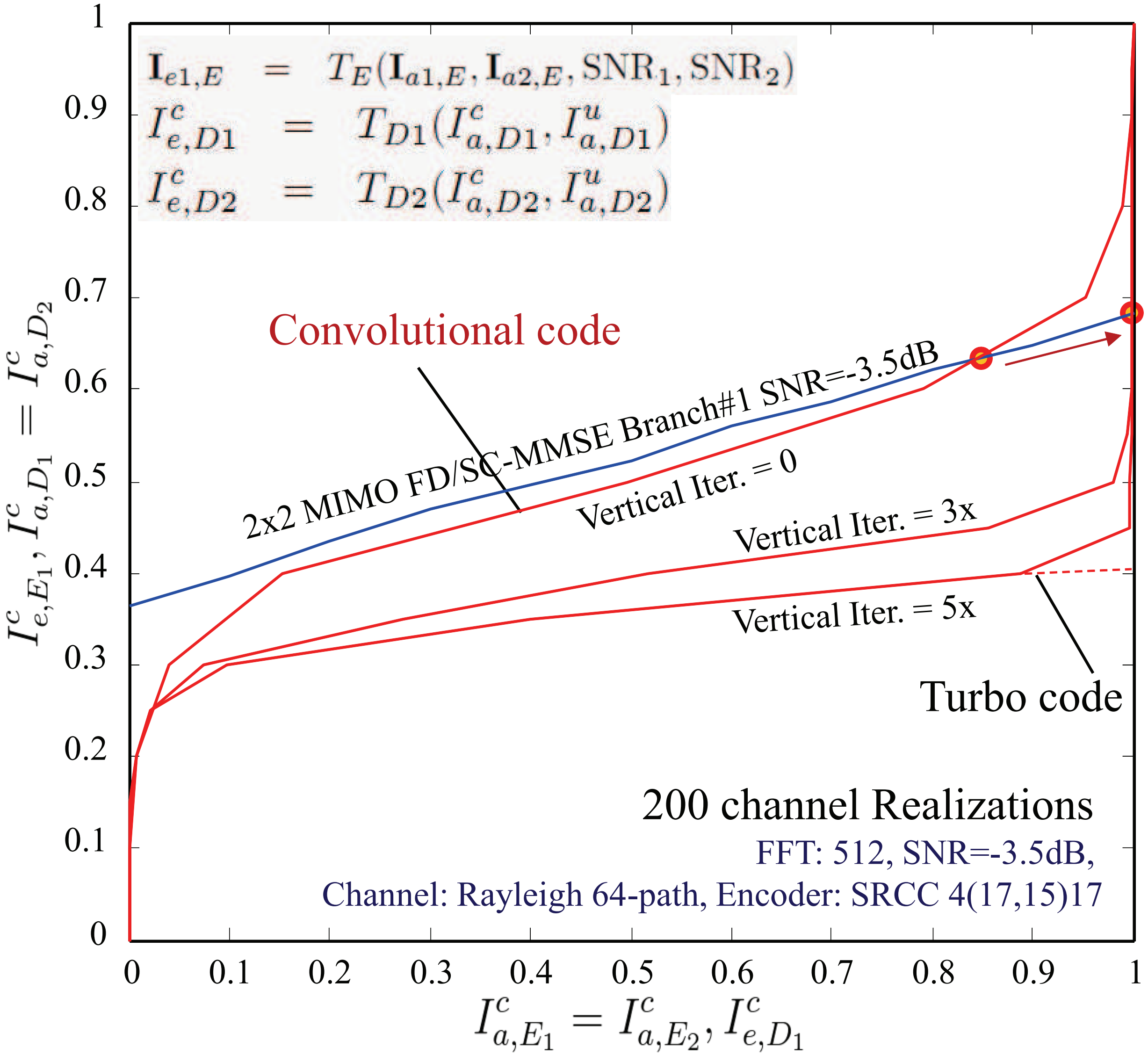
3D EXIT Chart at Antenna 1: SNR=-3 dB



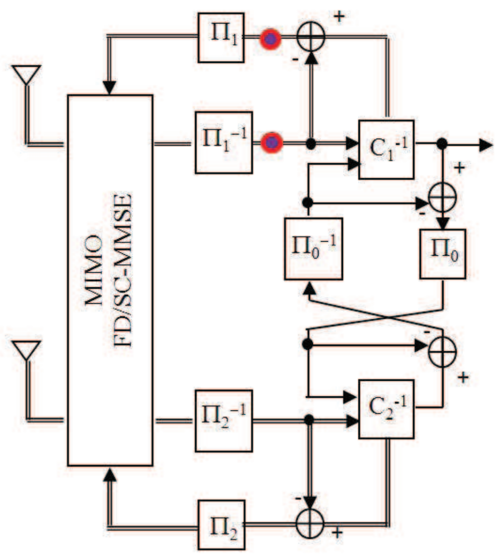
$$I^c_{e,D1} = T_{D1}(I^c_{a,D1}, I^u_{a,D1}), \quad I_{e1,E} = T_E(I_{a1,E}, I_{a2,E}, \mathrm{SNR}_1, \mathrm{SNR}_1), \quad (103)$$

$$I^c_{e,D2} = T_{D2}(I^c_{a,D2}, I^u_{a,D2}), \quad I_{e2,E} = T_E(I_{a2,E}, I_{a2,E}, \mathrm{SNR}_1, \mathrm{SNR}_1). \quad (104)$$

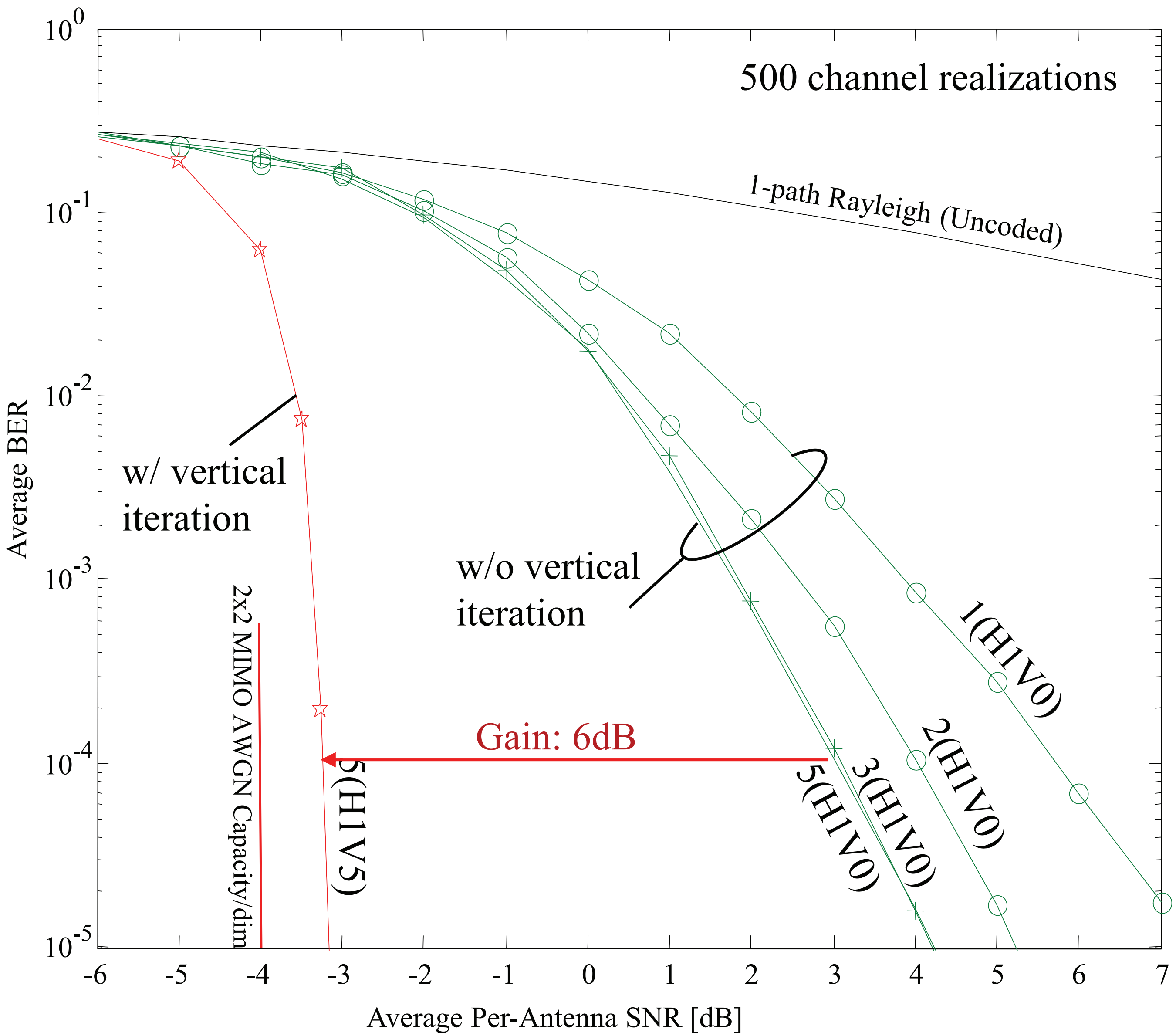
Contribution of Vertical Iterations



- Vertical Iteration converts CC $\rightarrow$ Turbo
- Stuck point is shifted to the right side.

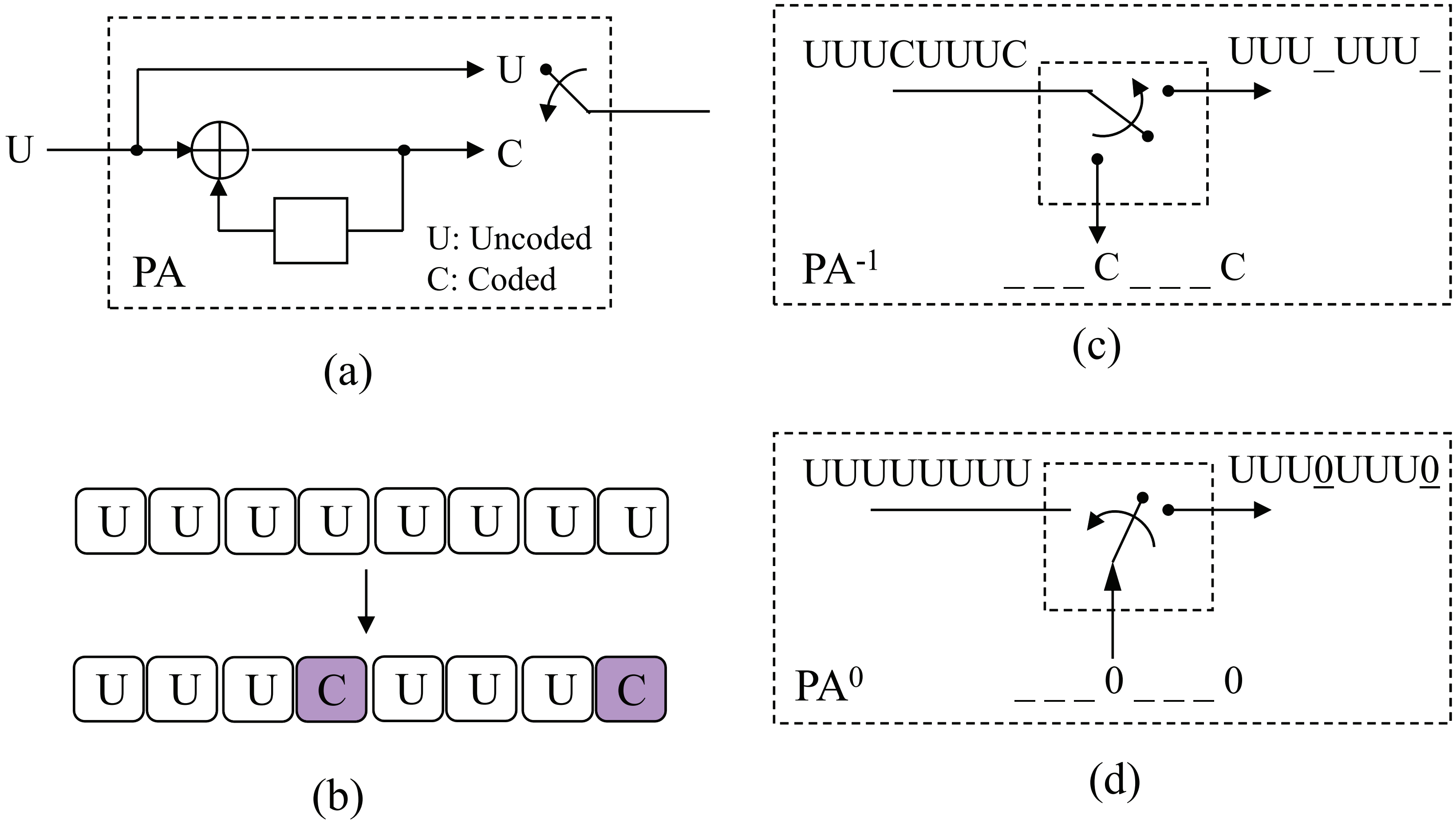


BER Performance of STC



- *Transmitter:* Encoder: SRCC $4(17, 15), 17$, Interleaver $=$1024 (random).
- *Channel:* MIMO 2×2 Equal average power 64-path.
- *Receiver:* Decoder: BCJR Log-MAP, FFT$=$512.

Doped-Accumulator



Doped-Accumulator: EXIT Chart and BER Perf.

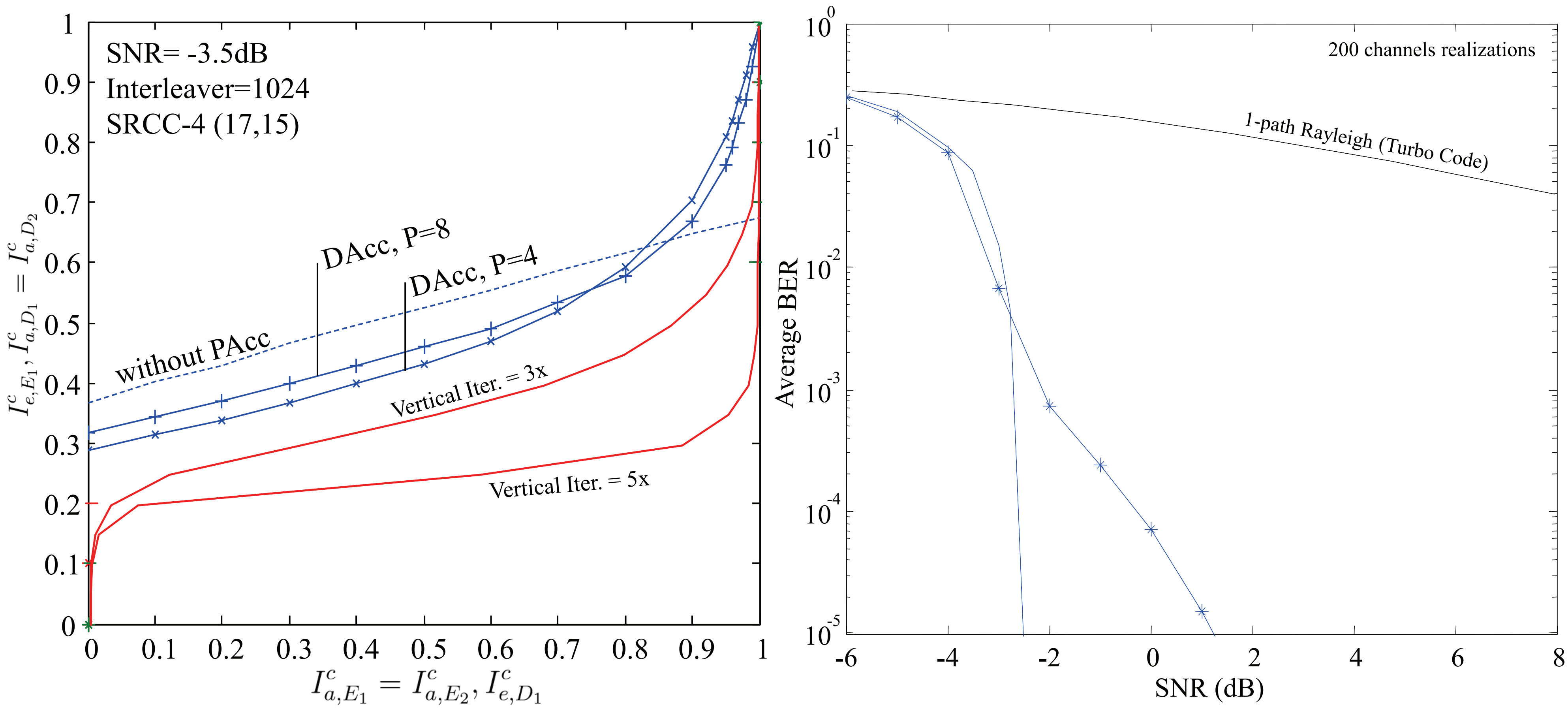


Figure 6: Doped Accumulator "bend" the EXIT curve of turbo equalization such that the error floor is reduced.

Conclusions

- Mathematics required to perform turbo equalization algorithm is no longer too complex. It is at a high school level.
- Single carrier signaling does not need high linearity-amplifier, and hence very energy-efficient.
- The excellent performance achievability of turbo equalization is NOT the matter of code strength. It is the matter of equalizer-decoder EXIT matching.
- Turbo MIMO changes the convolutional code into Turbo code by the vertical iterations.
- Doped accumulator can eliminate the error floor.