STANFORD

Lecture 4

Capacity, Separation Thm, & C-OFDM

April 14, 2023

JOHN M. CIOFFI

Hitachi Professor Emeritus of Engineering

Instructor EE392AA – Spring 2023

Announcements & Agenda

- Announcements
  - Problem Set #2 due Wednesday April 19 at 17:00
  - Sections 2.3-2.5
- Agenda
  - Capacity Examples
  - Chain Rule
    - Relation to MMSE SNR's and Decision Feedback (Successive Decoding)
  - MAP = MMSE on AWGN with good code
  - Separation Theorem
  - Coded MultiTone

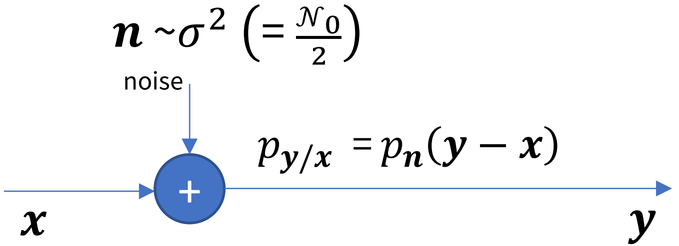
Capacity Examples

Sections 2.4 – 2.5

See PS2.3 (Prob 2.10)

The AWGN Capacity

- Simple formula says a lot



$$\bar{\mathcal{C}} = \frac{1}{2} \cdot \log_2\left(1 + \underbrace{\frac{\bar{\mathcal{E}}_{\boldsymbol{x}}}{\sigma^2}}_{SNR}\right)$$

- Often "gain" $\|h\|^2$ is absorbed into energy, really $g = \frac{\|h\|^2}{\sigma^2}$ so a "channel gain" $\bar{\mathcal{C}} = \frac{1}{2} \cdot \log_2(1 + g \cdot SNR)$
  - Note g here is per real dimension, but if complex noise $\mathcal{N}_0$ were used, it would be $\tilde{\mathcal{C}}_{\boldsymbol{x}} = \log_2(1 + g \cdot SNR)$
  - Know context and be consistent with numerator/denominator dimensionality

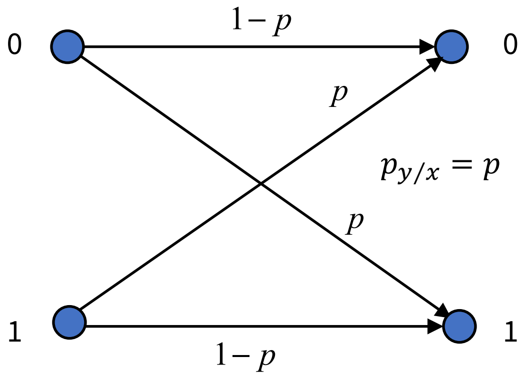
- SNR=4.7 dB (3 and g=1), then $\bar{\mathcal{C}} = 1$ bit/dimension

- SNR=20 dB (100 and g=1), then 3.33 bits/dimension – and thus 6.67 bits/complex subsymbol

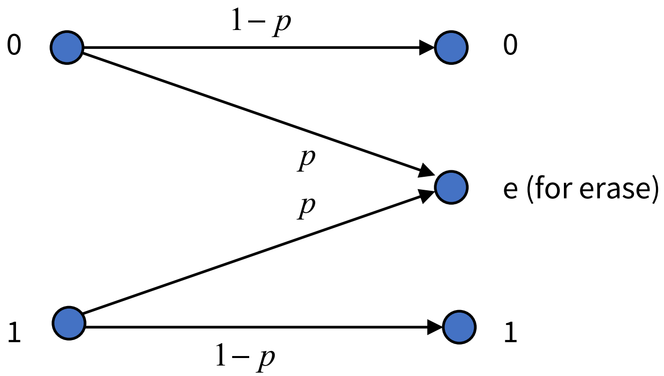
- What SNR gives 7 bits per dimension? $10 \cdot \log_{10}(2^{14} - 1) = 14 \cdot 3 = 42$ dB

BSC and BEC

- **BSC** has $\bar{\mathcal{C}} = 1 - \mathcal{H}(p) = 1 - p \cdot \log_2 p - (1-p) \cdot \log_2(1-p)$
  - $p = 1/2$ → 0 bits possible (makes sense)
  - $p = 0$ → 1 bit/dimension reliably (makes sense)
  - $0 \leq \bar{\mathcal{C}} \leq 1$



- **BEC** has $\bar{\mathcal{C}} = 1 - p$
  - $p = 1/2$ → 1/2 bits/dim reliable (no errors only erasures)
  - $p = 0$ → 1 bit/dimension reliably (makes sense)
  - $0 \leq \bar{\mathcal{C}} \leq 1$
- BEC is better than BSC (higher capacity) – decoders can use erasures with $N > 1$ to improve (reduce) P_e



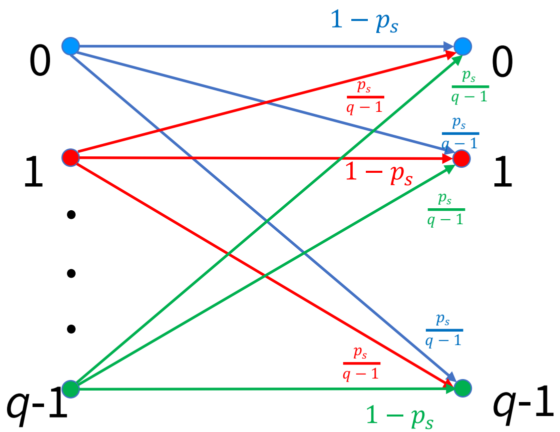
Symmetric DMC

- Generally just a discrete probability transition matrix (Appendix A)

  - q-ary (example 0,…,255 for a byte = subsymbol)

$$\mathcal{C} = b - p_s \cdot \log_2 \frac{2^b - 1}{p_s} + (1 - p_s) \cdot \log_2 (1 - p_s) \leq b \text{ bits.}$$

    - $p_s = .01$
    - $\mathcal{C} = 7.88$ bits/subsymbol



Chain Rule

Subsection 2.3.2

Chain Rule

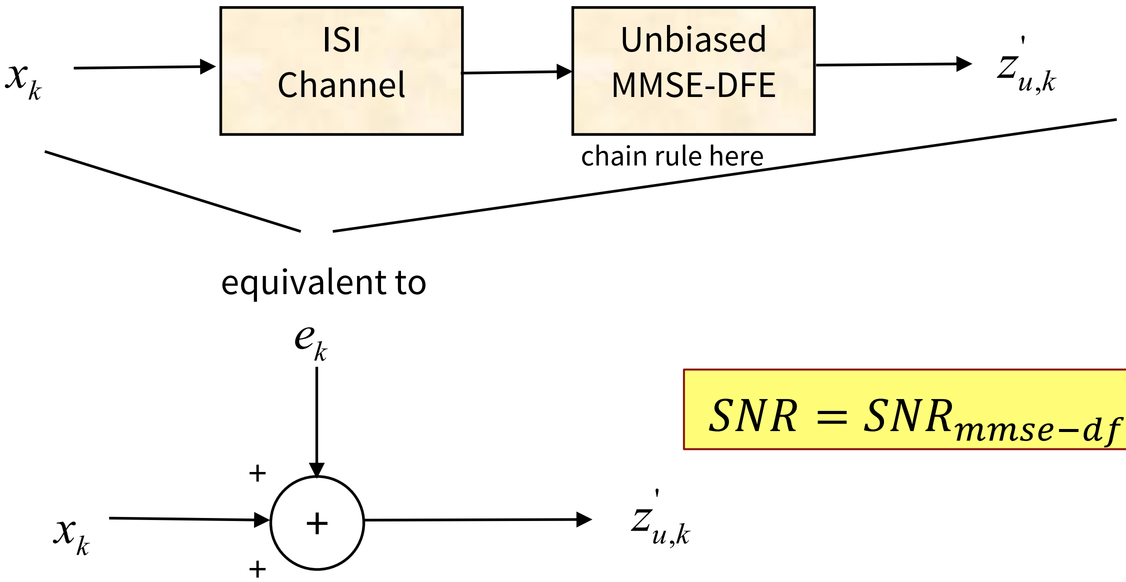
$$\mathcal{I}(\boldsymbol{x};\boldsymbol{y}) = \sum_{n=1}^{N} \mathcal{I}(\tilde{\boldsymbol{x}}_n;\boldsymbol{y}\,/[\tilde{\boldsymbol{x}}_{n-1} \quad \cdots \quad \tilde{\boldsymbol{x}}_1])$$

- If the subsymbols are independent, then parallel channels (we know this by now!)
- But suppose not: each term is itself a coding (MMSE-related if Gaussian) problem with SNR, capacity, etc.

Matrix AWGN: GDFE (sometimes also called "successive decoding")

- Estimate (MMSE) and decode $[\tilde{\boldsymbol{x}}_{n-1} \quad \cdots \quad \tilde{\boldsymbol{x}}_1]$ first, then simpler single component problem
  - So not just linear MMSE, linear MMSE + subtract "earlier" subsymbols' effect
- It's parallel channels, but with a twist to make them independent step by step ("decision-feedback")

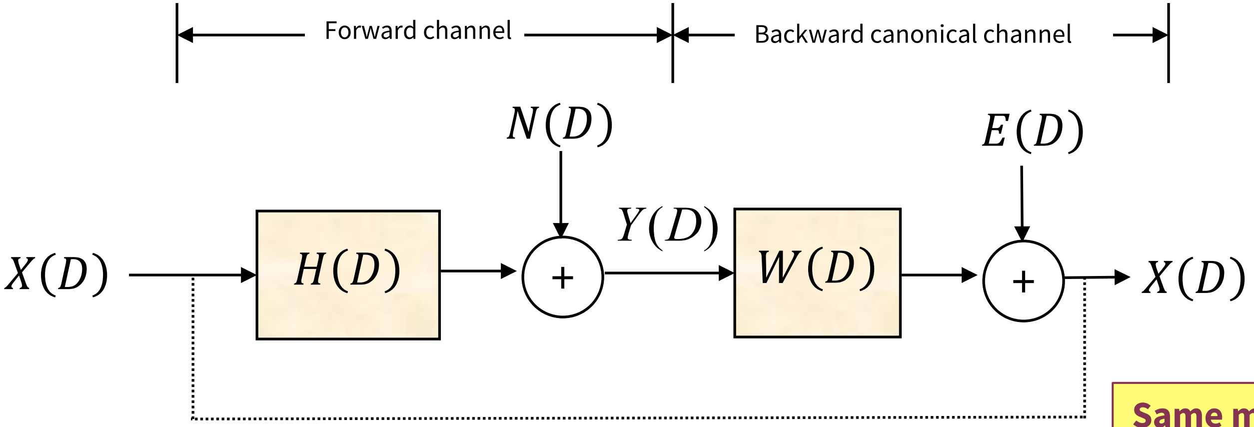
CDEF



$$SNR = SNR_{mmse-dfe,u} = 2^{2\bar{\mathcal{I}}(\tilde{x};\tilde{y})} - 1$$

- This one gets highest rate (with $\Gamma = 0$ dB) also
  - $I = C$ if water-filling spectrum is at transmitter
  - But, this is spectra is hard to do with DFE, so can be several parallel DFES (see Section 3.12)

Forward and its Backward Canonical Models



Same mutual information

$$r(t) = h_c(t) * h_c^*(-t) = \|h\|^2 \cdot q(t)$$

$y(t) \rightarrow h_c^*(-t) \rightarrow \frac{1}{T} \rightarrow Y(D)$

$$Y(D) = R(D) \cdot X(D) + \underbrace{N(D)}_{\frac{\mathcal{N}_0}{2} \cdot R(D)}$$

Forward Canonical Model

$$X(D) = \underbrace{W(D)}_{MMSE-LE} \cdot Y(D) + \underbrace{E(D)}_{\frac{\mathcal{N}_0}{2} \cdot W(D)}$$

Backward Canonical Model
chain rule helps more here

MAP = MMSE on AWGN

(asymptotically, Subsection 2.3.6)

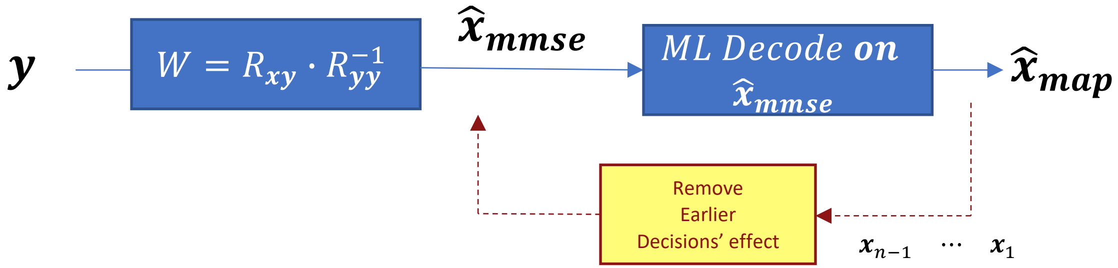
See PS2.5 (Prob 2.20)

For the filtered/matrix AWGN

- The MAP and MMSE determine the performance, and also the chain rule suggests a simpler decoder

$$\begin{aligned}
\mathcal{I}(\tilde{\boldsymbol{x}};\tilde{\boldsymbol{y}}) &= \mathcal{H}_{\tilde{\boldsymbol{y}}} - \mathcal{H}_{\tilde{\boldsymbol{y}}/\tilde{\boldsymbol{x}}} \\
&= \log_2\left(\frac{\left|R_{\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}}\right|}{\left|R_{\tilde{\boldsymbol{n}}\tilde{\boldsymbol{n}}}\right|}\right) \text{ bits/subsymbol} \\
&= \mathcal{H}_{\tilde{\boldsymbol{x}}} - \mathcal{H}_{\tilde{\boldsymbol{x}}/\tilde{\boldsymbol{y}}} \\
&= \log_2\left(\frac{\left|R_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}}\right|}{\left|R_{\boldsymbol{ee}}\right|}\right) \text{ bits/subsymbol} \\
. &= \log_2\left|I - W\cdot H\right| \\
&= \log_2\left|I - H\cdot W\right| \quad .
\end{aligned}$$

$$= \log_2(SNR_{mmse})$$



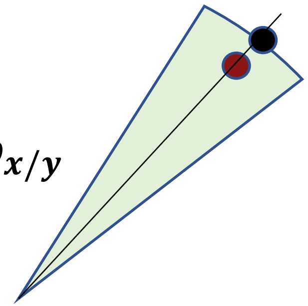
Still MAP if "previous" decisions are correct sequentially decodes

ML Detector for the Good Code

- ML = MAP since all good code's $\boldsymbol{x}$ are equally likely (uniform, AEP)

$$\frac{MAP}{ML} \ni \min_{\{\tilde{\boldsymbol{x}}_k\}} \sum_{k=-\infty}^{\infty} \|\tilde{\boldsymbol{y}}_k - H \cdot \tilde{\boldsymbol{x}}_k\|^2 \neq \sum_{k=-\infty}^{\infty} \|\tilde{\boldsymbol{n}}_k\|^2$$

Same as $\max_{\boldsymbol{x}} p_{\boldsymbol{x}/\boldsymbol{y}}$



- The smallest sum will reduce $\{\tilde{\boldsymbol{x}}_k\}$ magnitude slightly because it also shrinks noise (trade-off in sum)

$$MMSE \ni \min_{\{\tilde{\boldsymbol{x}}_k\}} \left\{ \lim_{K\to\infty} \frac{1}{2K+1} \sum_{K=-K}^{K} \|\tilde{\boldsymbol{x}}_k - W \cdot \tilde{\boldsymbol{y}}_k\|^2 \right\}$$

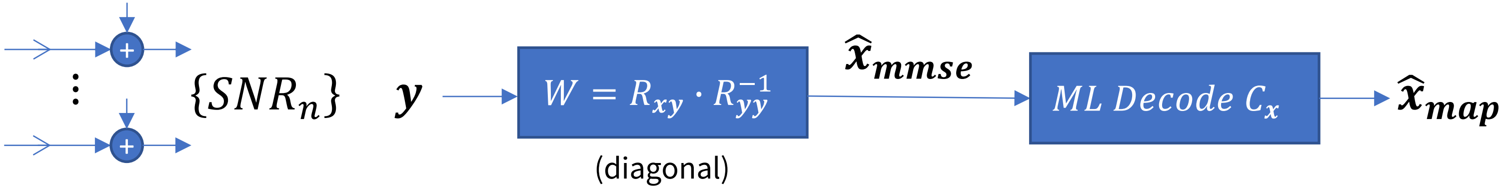
min over entire sum

- By LLN, this sum is MMSE and has solution $\hat{\boldsymbol{x}} = E[\tilde{\boldsymbol{x}}/\tilde{\boldsymbol{y}}]$ …… on average over the random code set
- But this is the conditional (a posteriori) mean that also uses the $\grave{a}$ posteriori (MAP) probability dist'n
  - Any single specific code's optimum receiver begins with an MMSE estimate of the channel input, but then does need to find the closest codeword to the MMSE estimate

Separation of Coding and Modulation

Subsection 4.4

The best (MAP) receiver



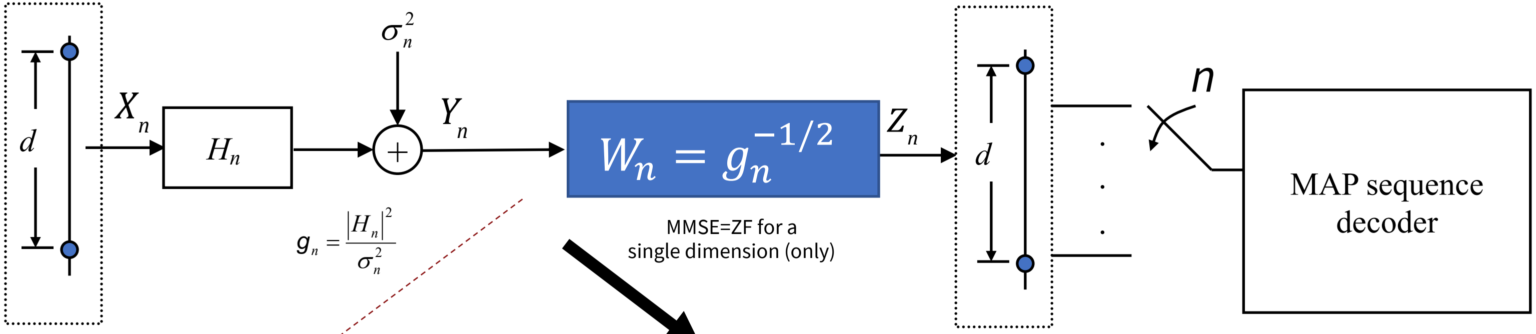
- Each parallel channel has $\mathcal{I}(x_n; y_n) = \log_2 SNR_{mmse,n}$, since they are independent

- Suppose each dimension is a dimension within the same code?
  - The dimensional signals will remain uncorrelated (but not independent because of the same code)
    - On average over all (Gaussian) codes these dimensions are independent, not for specific code.

- W is a scalar multiply for each such uncorrelated dimension (so does not change signal to noise)
  - Does use of MMSE matter? (not for VC or DMT)

- **YES, IT DOES MATTER** – **IF, a constant** bits/subsymbol $b_n \equiv \tilde{b}$ (and/or SNR_{geo}) – **Coded OFDM**
  - because it impacts the weight of different dimensions before the final ML decoding; this (it turns out) is the same as the earlier bit-loading, in effect

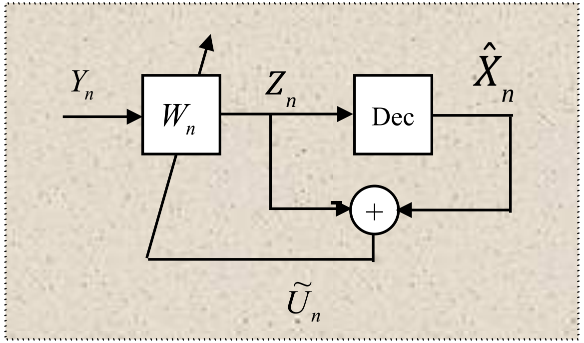
Dimensional Normalizer is MMSE



Must have each dimension scalar for ZF=MMSE

Zero-Forcing Algorithm

$$W_{n,k+1} = W_{n,k} + \mu \cdot \tilde{U}_n \cdot \hat{X}_n$$



see also L5 S18, S19

- MAP Decoder is

$$\min_{\{X_{n,k}\}} \left\{ \sum_{k=0}^{\infty} \sum_{n=0}^{N-1} \left| X_{n,k} - W_n \cdot Y_k \right|^2 \right\}$$

dimensions with low gains have greater contribution to minimization and decoder must apply more code redundancy there

Separation Theorem

Theorem 4.4.1 **[Separation of Coding and Modulation]** *Given a set of independent partitioned AWGN-channel dimensions with energies/dimension $\bar{\mathcal{E}}_n$ and gains g_n with equivalent*

$$SNR_{geo} = \left[\prod_{n=1}^{N}(1+\bar{\mathcal{E}}_n \cdot g_n)\right]^{1/N} - 1 \ ,$$

N repeated uses of a single good code with $\Gamma \rightarrow 0$ dB and

$$\bar{b} \leq \bar{\mathcal{I}} = \frac{1}{2} \cdot \log_2(1 + SNR_{geo})$$

and corresponding constant constellation $|C| = 2^{\bar{b}+\bar{\rho}}$ achieves the same performance as using N instances of that same good code with $\Gamma \rightarrow 0$ dB each with variable constellation $|C_n|$ and bits per tone

$$\bar{b}_n \leq \bar{\mathcal{I}}_n = \frac{1}{2} \cdot \log_2(1 + \mathcal{E}_n \cdot g_n) \ .$$

- Critical are:
  - the independent parallel dimensions (not code, the partitioned matrix/filtered AWGN)
  - the good code for which the input to the parallel dimensions comes from large constellation with subsymbol distribution approaching Gaussian

Widely applicable

- This works for partitioning with
  - SVD
  - Eigenvectors
  - DFT/FFT (becomes Coded-OFDM here)
  - Other bases

- The transmitter does not need to know individual $\tilde{b}_n$, just the sum for any symbol/subsymbol

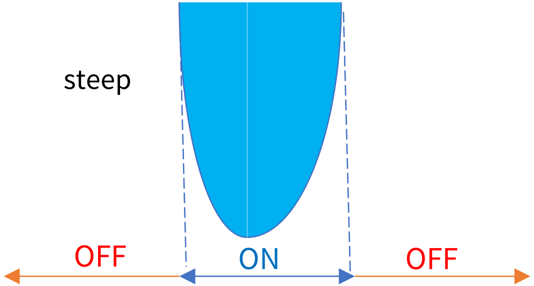
- It works for any $\mathcal{E}_n$ and leads to highest rate for those energies $\mathcal{I}(\tilde{x};\tilde{y})$.
  - Water-fill set gives highest data rate (highest mutual information)

- We've seen in our examples that water-fill is pretty close to on/off
  - So if the designer guesses well the on/off, ALMOST no feedback of bit distribution to transmitter
  - In practice, the constellation size and redundancy need specification, and thus on some indication of the value of $\mathcal{I}(\tilde{x};\tilde{y})$ for the channel.

- **Example:** Wireless "MCS" (modulation coding scheme) specifies code rate and constellation size only in feedback to transmitter. The on/off distribution?
  - They ignore this for time-frequency and just use flat over the entire band
  - They do excite spatial "streams" that can carry data and zero others.

Caution on Water-filling and on/off

- Most water-fill will satisfy $\left(1+\frac{SNR_n^*}{\Gamma}\right) \cong \frac{SNR_n^*}{\Gamma}$
  - IF dimension carries nonzero energy

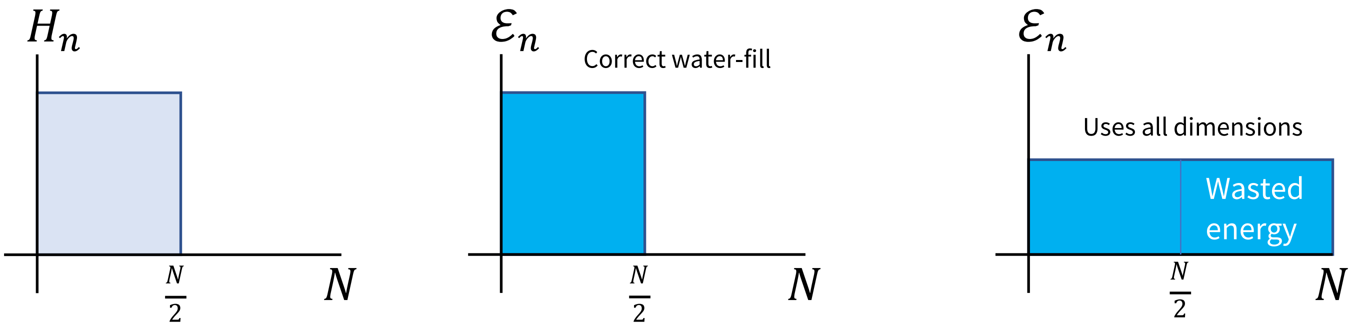


- The energy closely approximates flat
  - RA: $K - \frac{\Gamma}{g_n} = \frac{\mathcal{E}_x}{L^*} + \frac{\Gamma}{L^*} \cdot \sum_{l=1,l\neq n}^{L^*} {}^1/_{g_l}$ is roughly the same (no one dimension dominates)
  - MA: $K - \frac{\Gamma}{g_n} = \Gamma \cdot \left(\frac{2^{\tilde{b}}}{\prod_{l=1}^{L^*} g_l}\right)^{1/L^*} - \frac{\Gamma}{g_n} = \frac{\Gamma}{g_n} \cdot \left\{\left(\frac{SNR_{geo}}{\prod_{l=1,l\neq n}^{L^*} g_l}\right)^{1/L^*} - 1\right\}$ is roughly the same (again no one dim dominates)

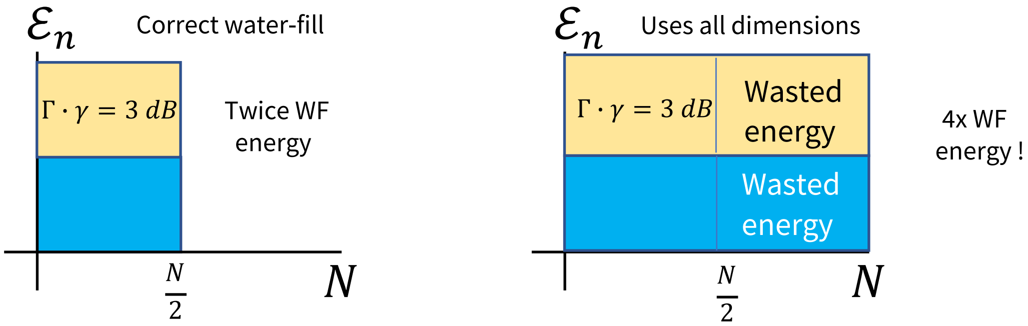
- This is true on water − fill′s ENERGIZED (“on”) dimensions, NOT for zeroed dimensions (“off”)

So it is NOT true that wireless’ C-OFDM is the same as DMT, UNLESS the used dimensions are close to the same!

Half-Band Example

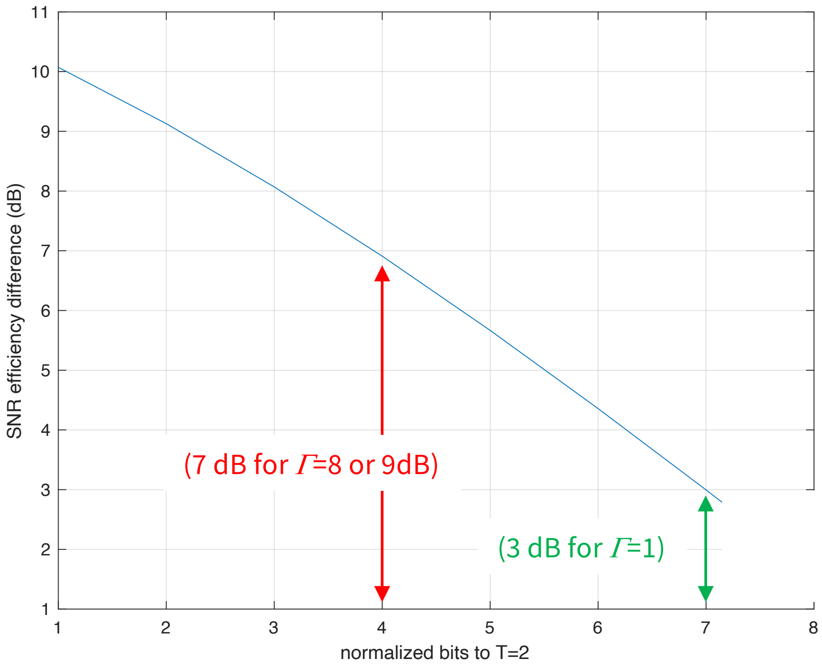


- The geometric SNR for water-fill is 3 dB higher if capacity-achieving codes are used
  - Or could run the water-fill system at same data rate at 3 dB less energy
- This amount is amplified below capacity by non-unity (not 0 dB) gap-margin product



See PS2.4 (Prob 2.14)

margin difference for half-band optimum versus full band

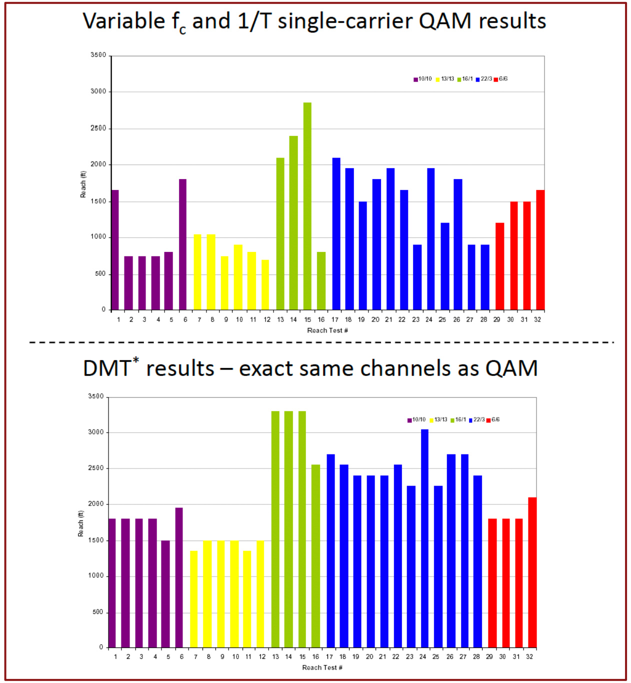


margin difference for half-band optimum versus full band

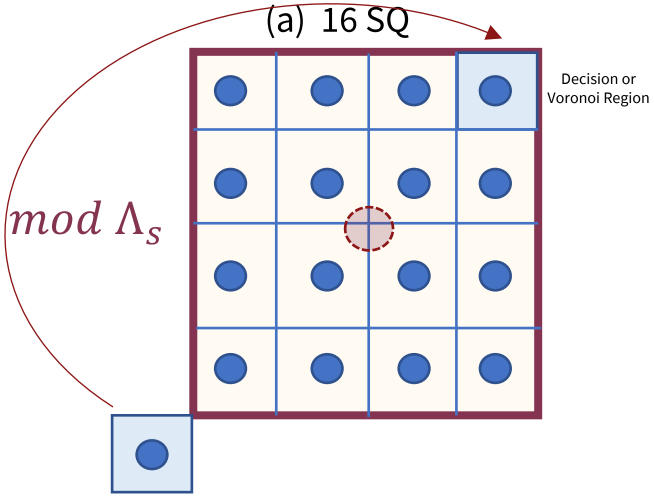
- Capacity of AWGN with WF is 8 bits/subsymbol (4 bits/dimension)

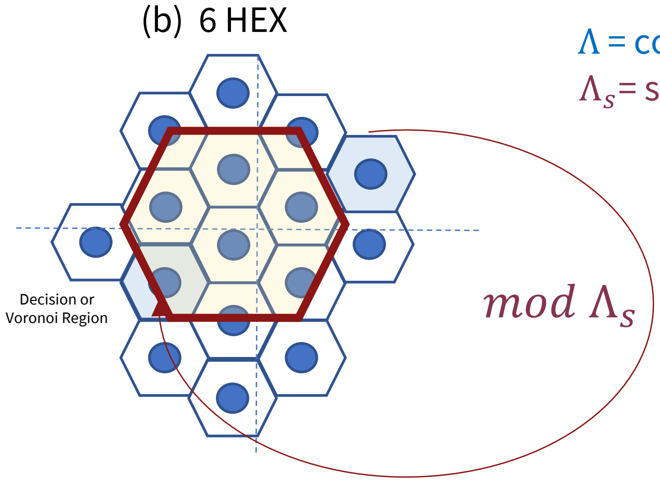
1993 ADSL Olympics – Bellcore
Margin differences at 1.6 Mbps, 4 miles, 11+dB
DMT 4x faster (6 Mbps) at 2 miles

2003 VDSL Olympics - Bellcore



Coding Gain and Constellation/Code





Λ = coding lattice for d_{min}

Λ_s = shaping lattice for $\mathcal{E}_x$

$$\gamma \triangleq \frac{\left(d_{\min}^2(\boldsymbol{x})/\bar{\mathcal{E}}_{\boldsymbol{x}}\right)}{\left(d_{\min}^2(\breve{\boldsymbol{x}})/\bar{\mathcal{E}}_{\breve{\boldsymbol{x}}}\right)} = \underbrace{\frac{\left(\frac{d_{\min}^2(\boldsymbol{x})}{V^{2/N}(\Lambda)}\right)}{\left(\frac{d_{\min}^2(\breve{\boldsymbol{x}})}{V^{2/N}(\breve{\Lambda})}\right)}}_{\gamma_f} \cdot \underbrace{\frac{\left(\frac{V^{2/N}(\Lambda)}{\bar{\mathcal{E}}_{\boldsymbol{x}}}\right)}{\left(\frac{V^{2/N}(\breve{\Lambda})}{\bar{\mathcal{E}}_{\breve{\boldsymbol{x}}}}\right)}}_{\gamma_s}$$

fundamental gain | **shaping gain**

Basic principle extends $\bar{N} \rightarrow \infty$
Hexagon→ hypersphere (Gaussian marginals)

good codes can follow from $^{\Lambda_s}/_{\Lambda} = |C|$

SQ constellations vs "Gaussian"

- There is always a loss for a non-hyper-spherical constellation boundary on the (any matrix/filtered) AWGN
  - The max shaping gain, $\gamma_{s,max}$=1.53 dB (when $\tilde{b} \geq 1$), relative to hypercube
  - Hypercube is often the assumed reference system (so Λ for fundamental and scaled Λ_s for shaping)

- All of random coding/AEP can repeat with the input distribution being uniform in any dimension (instead of Gaussian) – hypercube-energy constraint

- The MMSE Estimator can still be used with decoder, and it's basically

$$\tilde{C} = \log_2\left(1 + SNR_{mmse,u}/\gamma_{s,max}\right)$$

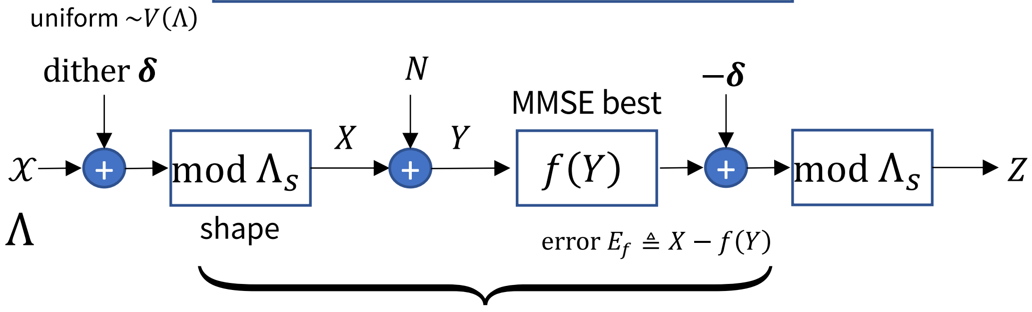
- Loss of 0.5 bit/complex dimension

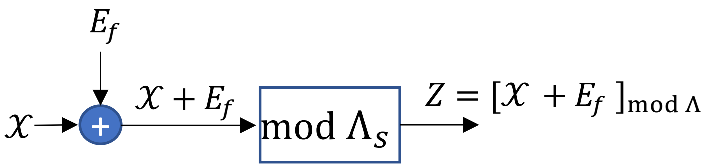
Forney's Crypto MMSE equivalence

1.53 dB (max) loss

$$\tilde{J} \geq \tilde{C} - \log_2\left(\pi \cdot e \cdot \frac{\varepsilon}{V(\Lambda_s)}\right) - \underbrace{\log_2\left(\frac{\sigma^2_{E_f}}{\sigma^2_{mmse}}\right)}_{0}$$

$$\Lambda_s = \sqrt{\frac{|C|}{2}} \cdot Z^2$$





- See also Section 2.8 – there is a shaping loss with any Λ_s that is not a hypersphere (SQ is worst in practice) so various shaping methods can apply; however the separation theorem still applies to them all, with random coding used on uniform over Λ_s's Voronoi region

Coded OFDM/MT

Subsection 4.4.1

SQ constellations vs "Gaussian" - REPEAT S23

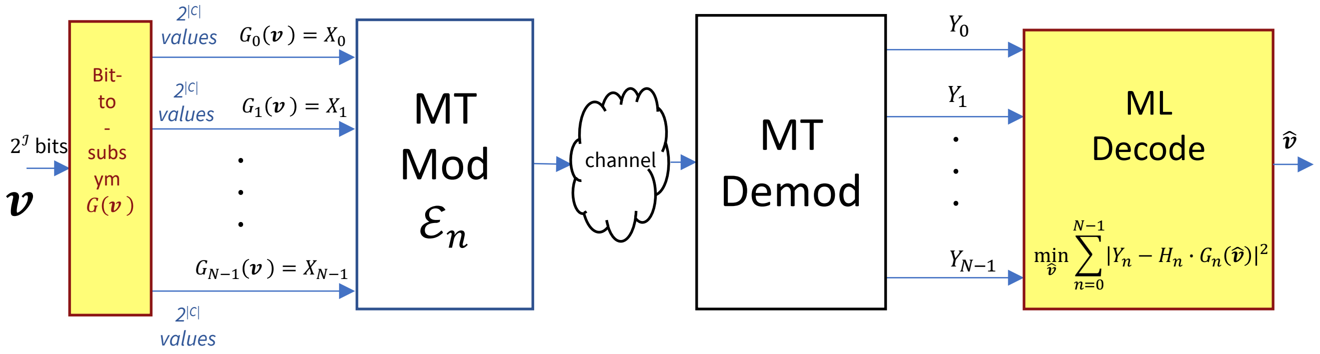
- Matrix/filtered-AWGN loss for "square" constellations

$\gamma_s \leq \mathbf{1.53\ dB}$ **shaping gain**

$$\tilde{C} = \log_2\left(1 + SNR_{geo}/\gamma_{s,max}\right)$$

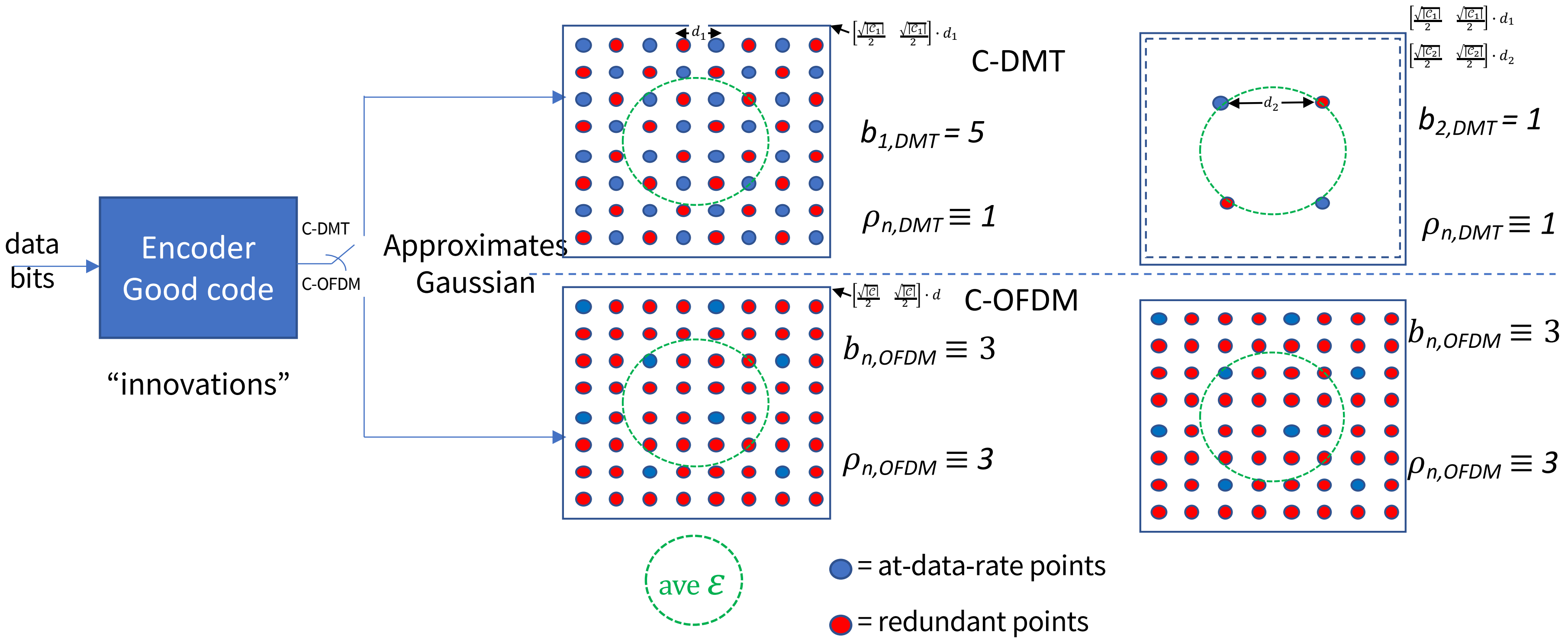
- When $\tilde{C} \leq 1$, $\boldsymbol{\gamma_s}$=0 dB
  - there is no **low-SNR** shaping loss for binary codes
- AEP applies to hypercube (with shaping loss) boundary and random codes
- MMSE estimator precedes MAP decoder for **original** code
  - ISI/crosstalk optimally handled linearly with parallel ind subchannels
  - slight nonlinear decision feedback when NOT parallel independent channels

Coded-MT/OFDM



- Treats a pre-agreed known set of dimensions as repeated constant SNR_{geo} dimensions
  - No transmitter bit loading, and energy is on/off on the pre-agreed set
- The MT could be replaced by space-time MIMO, "Coded-Vector-Coding" – same basic principle

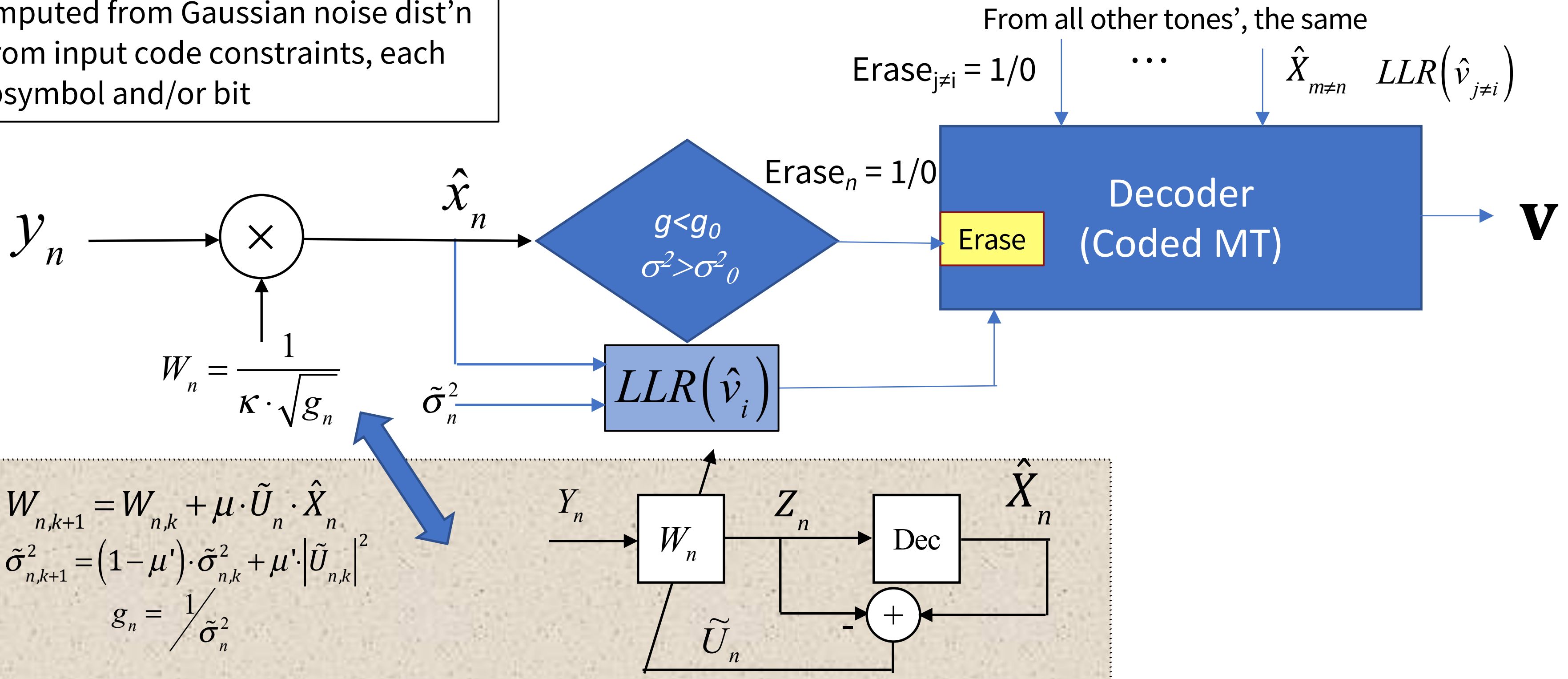
Comparison of variable and fixed constellation



- These types of system are heavily used in practice

Full MAP Decoder

LLR = log likelihood ratio
Computed from Gaussian noise dist'n & from input code constraints, each subsymbol and/or bit



$$W_{n,k+1} = W_{n,k} + \mu\cdot\tilde{U}_n\cdot\hat{X}_n$$
$$\tilde{\sigma}^2_{n,k+1} = \left(1-\mu'\right)\cdot\tilde{\sigma}^2_{n,k} + \mu'\cdot\left|\tilde{U}_{n,k}\right|^2$$
$$g_n = 1/\tilde{\sigma}^2_n$$

STANFORD

End Lecture 4