Lecture 4

Capacity, Separation Thm, & C-OFDM

April 11, 2024

JOHN M. CIOFFI

Hitachi Professor Emeritus of Engineering

Instructor EE379B – Spring 2024

Announcements & Agenda

- Announcements
  - Problem Set #2 due Wednesday April 17 at 17:00
  - Sections 2.3-2.5
  - Need to leave off hours around 10:40 to catch flight today.

- Agenda
  - Capacity Continued
  - Chain Rule
  - Separation Theorem
  - Coded MultiTone

Capacity Continued

Sections 2.4 – 2.5

See PS2.3 (Prob 2.10)

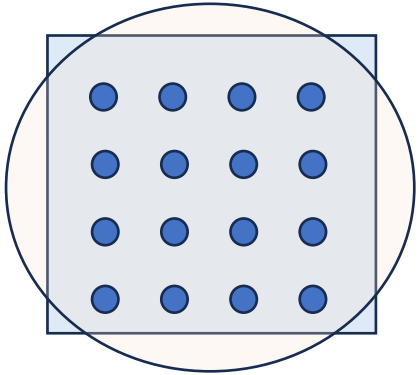
General Capacity Theorem from L3

$$\frac{|A_N^\epsilon(\boldsymbol{x})|}{|A_N^\epsilon(\boldsymbol{x}/\boldsymbol{y})|} \to 2^{\mathcal{I}(\boldsymbol{x};\boldsymbol{y})} \qquad \text{since } \mathcal{I}(\boldsymbol{x};\boldsymbol{y}) = \mathcal{H}_{\boldsymbol{x}} - \mathcal{H}_{\boldsymbol{x}/\boldsymbol{y}}$$

- Good codes will have only 1 codeword per conditional entropy subset.
- MAP detector decision region is then ~ $A_N^\epsilon(\boldsymbol{x}/\boldsymbol{y})$ - on average; but we can find it for one good code.
- If $A_N^\epsilon(\boldsymbol{x})$ were any larger, all codes (good or bad) will have at least one $A_N^\epsilon(\boldsymbol{x}/\boldsymbol{y})$ that contains 2+ codewords, which mean the MAP has to "flip a coin" – not good (high error prob).
- SHANNON's CAPACITY THEOREM
  - Number of codewords is limited by mutual info $b \leq \mathbb{I}(\boldsymbol{x}\,;\boldsymbol{y}\,)$.
  - Which is per-subsymbol equivalent with random code $\tilde{b} \leq \mathbb{I}(\tilde{\boldsymbol{x}}\,;\tilde{\boldsymbol{y}}\,)$.
  - If maximized over input distributions $\tilde{b} < \tilde{C} \leq \max_{p_{\tilde{\boldsymbol{x}}}} \mathbb{I}(\tilde{\boldsymbol{x}}\,;\tilde{\boldsymbol{y}}\,) \frac{\text{bits}}{\text{subsymbol}}$.

The uniform part is most important (from L3)

- The Gaussian distribution corresponds to marginal of uniform distribution over a hypersphere.
  - This uniform distributions marginals are asymptotically Gaussian.
  - This is a special case where uniform and Gaussian are basically the same.
  - Because all the Gaussian infinite-length vectors (codewords) have same energy (zero variance of the energy).
- All the points (really volume) are (is) at the surface.
- The Gaussian marginal dist'n is important only for shaping gain (< 1.53 dB).
- The (AEP) uniform spacing of points (no matter where the majority of them sit, surface or otherwise) remains for the fundamental gain.



The uniform spacing separates codewords in the union of the hypersquare (orthotope) and hypersphere.

Thus, good codes can be based on sequences from uniformly spaced PAM/QAM subsymbols.

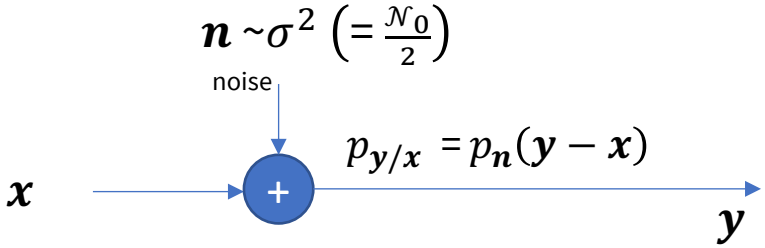
And the rest is MMSE estimation,

With a chain-rule twist in some situations

Vector Coding is always MAP, ML, & MMSE, a special case.

AWGN Capacity Review (379A)

- Simple formula says a lot.



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

- Often "gain" $\|h\|^2$ is absorbed into energy, really $g = \frac{\|h\|^2}{\sigma^2}$ so a "channel gain" $\bar{C} = \frac{1}{2} \cdot \log_2(1 + g \cdot \bar{\mathcal{E}}_x)$.
  - Note g here is per real dimension, but if complex- baseband channel, it would be $\tilde{C}_x = \log_2(1 + g \cdot \bar{\mathcal{E}}_x)$.
  - Know context and be consistent with numerator/denominator dimensionality.
- SNR=4.7 dB (3 and g=1), then $\bar{C} = 1$ 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.

Chain Rule

Subsection 2.3.2

Chain Rule

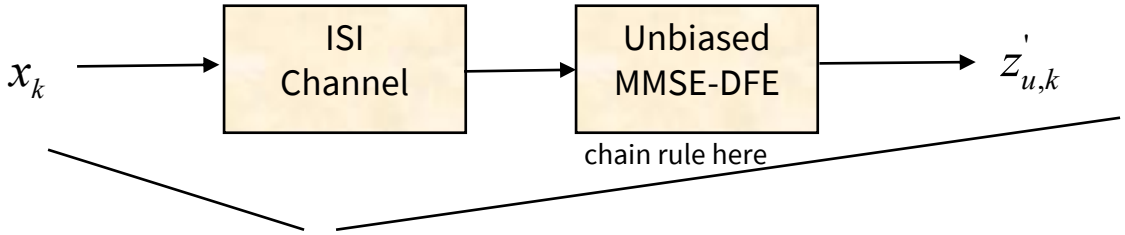
$$\mathcal{I}(\boldsymbol{x};\boldsymbol{y}) = \sum_{n=1}^{N} \mathcal{I}(\tilde{\boldsymbol{x}}_n;\boldsymbol{y}\,/[\tilde{\boldsymbol{x}}_{n-1} \quad \cdots \quad \tilde{\boldsymbol{x}}_1])$$

- If parallel independent channels (we know this by now!), the $[\tilde{\boldsymbol{x}}_{n-1} \quad \cdots \quad \tilde{\boldsymbol{x}}_1]$ provide no help;
  - just sum the individual channels $\mathcal{I}(\tilde{\boldsymbol{x}}_n;\tilde{\boldsymbol{y}}_n)$.
- But suppose not: each term represents a code (MMSE-related if Gaussian) problem with SNR, capacity, etc.

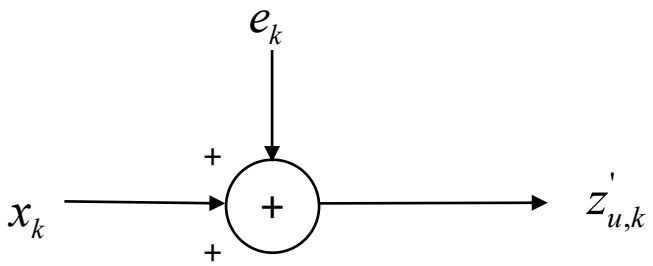
Matrix AWGN: GDFE (sometimes also called "successive decoding")

- 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 (nonlinear).
- It's parallel channels, but with a twist to make them independent step by step ("decision-feedback").

CDEF Example EE379A, L18



equivalent to



$$\mathcal{I}(\tilde{x};\tilde{\boldsymbol{y}}) = \mathcal{H}_{\tilde{x}_k} - \underbrace{\mathcal{H}_{\tilde{x}_k/[\tilde{\boldsymbol{y}} \quad \tilde{x}_{k-1.....-\infty}]}}_{MMSE-DFE}$$

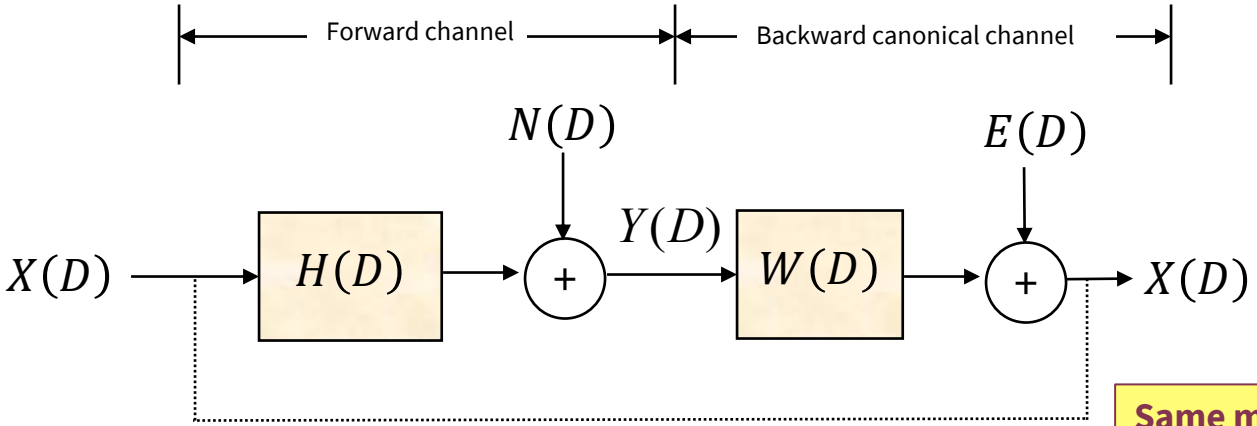
$$SNR = SNR_{mmse-dfe,u} = 2^{2\bar{\mathcal{I}}(\tilde{x};\tilde{y})} - 1$$

- The MMSE-DFE achieves the highest rate (with $\Gamma = 0$ dB) also:
  - $I = C$ if water-filling spectrum is at transmitter.
  - But, this spectra may be impossible with a single DFE, so can be several parallel DFES (see Section 3.12).
  - No error propagation (true if Pe is zero), and **canonical** (reliably achieves capacity).

There are many chain rule orders

- $\mathbb{I}(\boldsymbol{x};\boldsymbol{y}) = \sum_{n=1}^{N} \mathbb{I}\left(\tilde{\boldsymbol{x}}_{\pi(n)}; \boldsymbol{y} / [\tilde{\boldsymbol{x}}_{\pi(n-1)} \quad \cdots \quad \tilde{\boldsymbol{x}}_{\pi(1)}]\right) = \sum_{n=1}^{N} \mathbb{I}\left(\tilde{\boldsymbol{y}}_{\pi(n)}; \boldsymbol{x} / [\tilde{\boldsymbol{y}}_{\pi(n-1)} \quad \cdots \quad \tilde{\boldsymbol{y}}_{\pi(1)}]\right)$
- $N!$ orders exist for each of $\mathbb{I}(\boldsymbol{x};\boldsymbol{y}) = \mathbb{I}(\boldsymbol{y};\boldsymbol{x})$.
- Every order corresponds to different set of parallel channels (some with feedback, a few without).
- But they all produce the same maximum data rate (achieved with good code that has zero gap).
- Thus, not only are there a lot of good codes – there are a lot of good MMSE-based modulation/demodulation designs also!

Forward and its Backward Canonical Models



Same mutual information

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

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

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

Forward Canonical Model

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

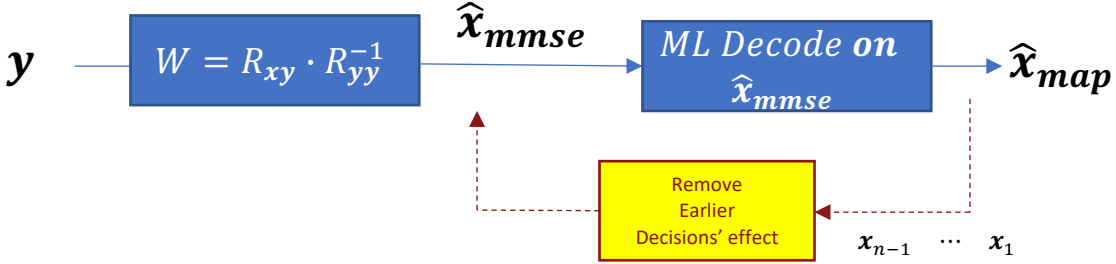
Backward Canonical Model
chain rule helps more here

For the filtered/matrix AWGN

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

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

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



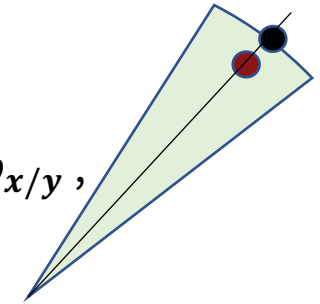
Still MAP if "previous" decisions are correct sequentially decodes

Optimal Detectors for Good Codes

- $\boldsymbol{x}$ codewords/subsymbols selected from Gaussian, $[\boldsymbol{x}\ \ \boldsymbol{y}]$ are jointly Gaussian (as is then $\boldsymbol{n}$ **).**
- **ML = MAP** since all good code's $\boldsymbol{x}$ codewords/symbols 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 $\max_{\boldsymbol{x}} p_{\boldsymbol{x}/\boldsymbol{y}}$,
where $\boldsymbol{x}$ has ∞ length



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

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

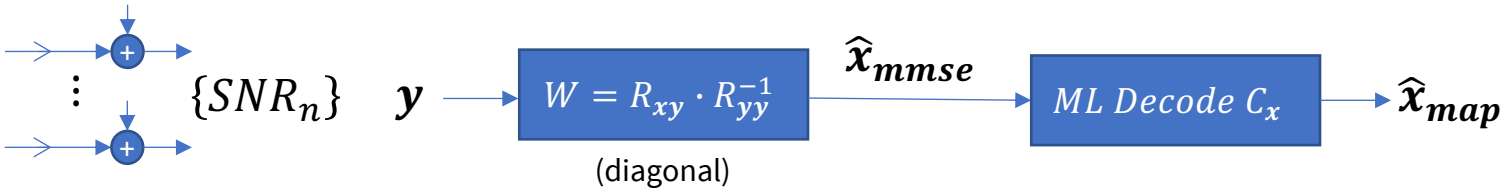
min over entire sum
W is MMSE filter

  - By LLN, this sum is MMSE & MAP and has optimum $\hat{\boldsymbol{x}} = E[\tilde{\boldsymbol{x}}/\tilde{\boldsymbol{y}}]$ …… on average over the random code set.
  - The bias removal is unnecessary because of the hyper-conical decision regions (like QPSK where decision regions don't change) for a zero-gap AWGN code, but we now know MMSE is a "DFE-like" structure (chain rule)

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

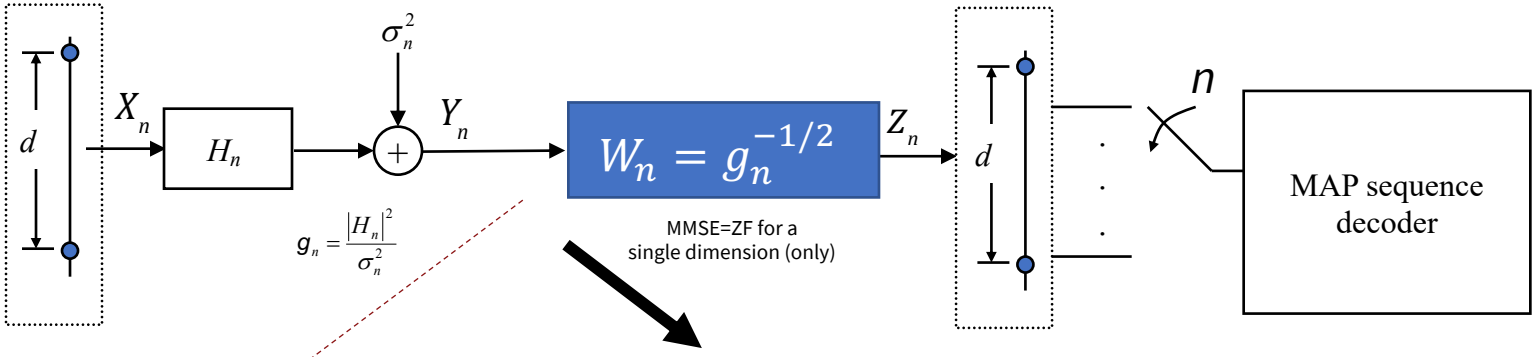
- Suppose each dimension is a dimension within the same code?
  - The dimensional subsymbols will remain uncorrelated (but not independent because of the same code).
    - On average over all (Gaussian) codes these dimensions are independent, not for specific code.

- 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 with $\tilde{b}_n = \log_2(1 + SNR_n)$)

- **BUT 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 for same energy distribution/Rxx.

$$LL_n = -\frac{1}{2\sigma_n^2} \cdot \|y_n - \hat{x}_n\|^2$$

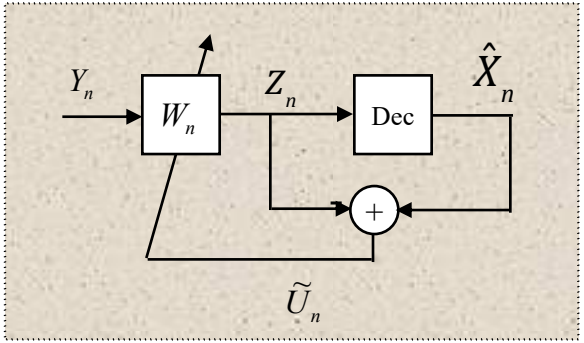
Dimensional Normalizer is MMSE



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 L5:6, L4:30

- MAP Decoder is

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

dimensions with low gains g_n have greater contribution to minimization, higher soft/intrinsic info, and the decoder is consequently affected.

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 $\Gamma \to 0$ *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* $\Gamma \to 0$ *dB each with variable constellation* $|C_n|$ *and bits per tone*

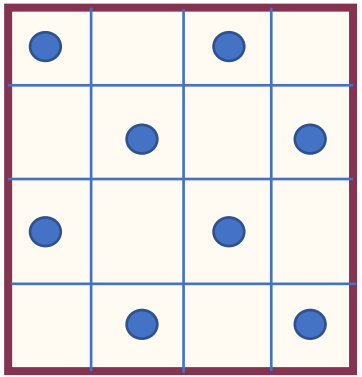
$$\bar{b}_n \leq \bar{\mathcal{I}}_n = \frac{1}{2} \cdot \log_2(1+\mathcal{E}_n \cdot g_n) \ .$$

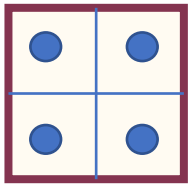
- Critical are:
  - the independent parallel dimensions (not code, the partitioned matrix/filtered AWGN), &
  - the good code for which the input to the parallel dimensions comes from large constellation with subsymbol distribution approaching Gaussian ~ random coding,
  - the code can apply “down the symbol” (over the subsymbols that all have same constellation).

Simple Separation Thm Example

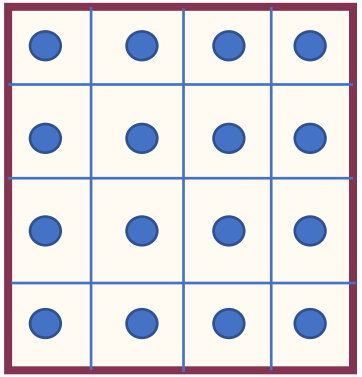
- The two tones' ave is $\tilde{\mathcal{I}}_{ave} = 2$.
- The ST says use of a single constellation with $|C_{ave}| = 8$ is sufficient.
- Decoder must consider channel gains.

$\tilde{\mathcal{I}}_{ave} = 2$
$|C_{ave}| = 8$





$\tilde{\mathcal{I}}_1 = 1$
$|C_1| = 4$



$\tilde{\mathcal{I}}_2 = 3$
$|C_2| = 16$

- Looks like 2 identical uses of a single AWGN with geometric-average SNR, equivalently $SNR = 2^{\tilde{\mathcal{I}}_{ave}} - 1$

Separation T is 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.

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

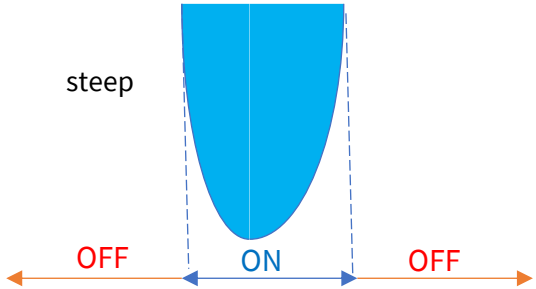
- We've seen in our examples that water-fill is close to on/off.
  - So, if the designer guesses well the on/off, the system requires ALMOST no feedback of bit distribution to transmitter.
  - In practice, the constellation size and redundancy need specification, and thus on some indication of the value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

- **Example:** Wireless "MCS" (modulation coding scheme) specifies code rate and constellation size only in feedback to transmitter. The on/off distribution?
  - 5G/Wi-Fi ignore this for time-frequency and just use flat over the entire band.
  - 5G/Wi-Fi do excite spatial "streams" unequally in that some can carry data and while others are zeroed.

ONLY IF $\Gamma \rightarrow 0$

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.

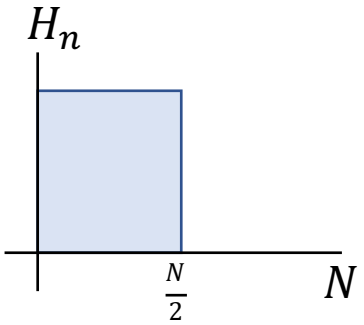


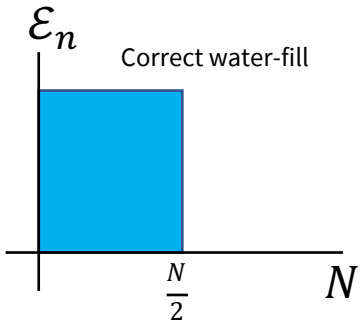
- The energy closely approximates flat:
  - RA: $K - \frac{\Gamma}{g_n} = \frac{\mathcal{E}_x}{L^*} + \frac{\Gamma}{L^*} \cdot \sum_{l=1,l\neq n}^{L^*} {}^1/_{g_l}$ is roughly the same (no one dimension dominates).
  - MA: $K - \frac{\Gamma}{g_n} = \Gamma \cdot \left(\frac{2^{\tilde{b}}}{\prod_{l=1}^{L^*}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

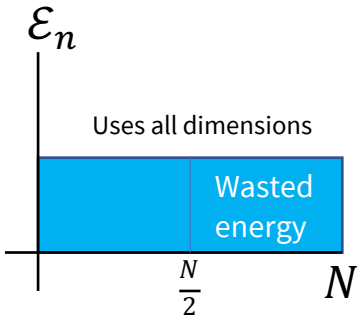
- This is true on waterfill's ENERGIZED ("*on*") dimensions, NOT for zeroed dimensions ("off").

So, it is NOT true that wireless' C-OFDM is the same as DMT, UNLESS the used dimensions are close to the same and $\Gamma \to 0$!

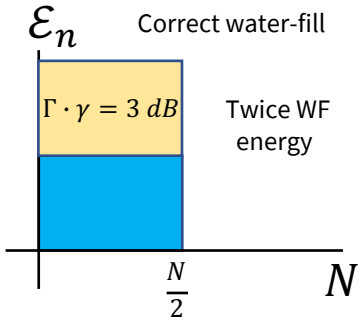
Half-Band Example: Revisit 379A's L18:12-13

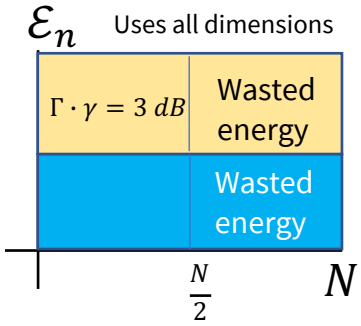






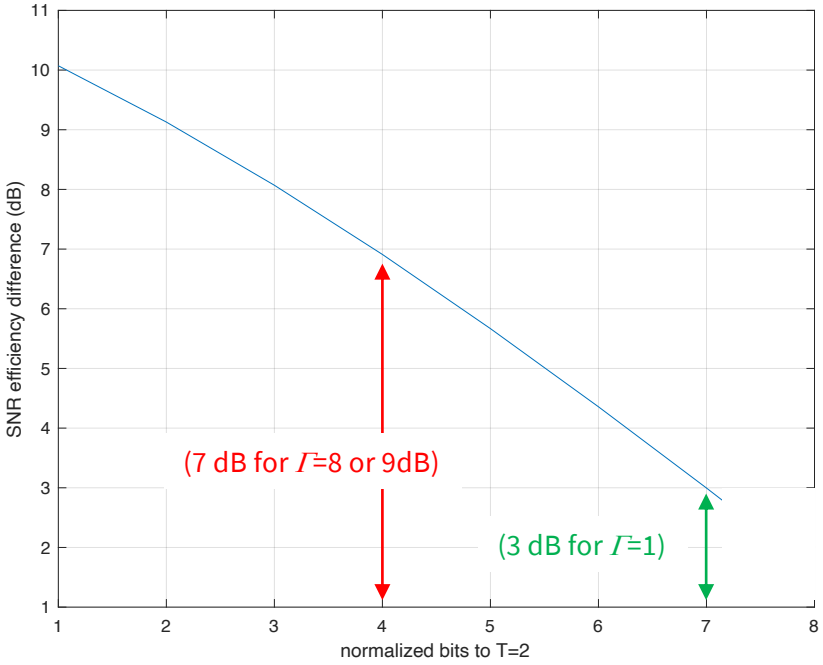
- The geometric SNR for water-fill is 3 dB higher if capacity-achieving codes are used
  - Or could run the water-fill system at same data rate at 3 dB less energy
- This amount is amplified below capacity by non-unity (not 0 dB) gap-margin product





4x WF energy !

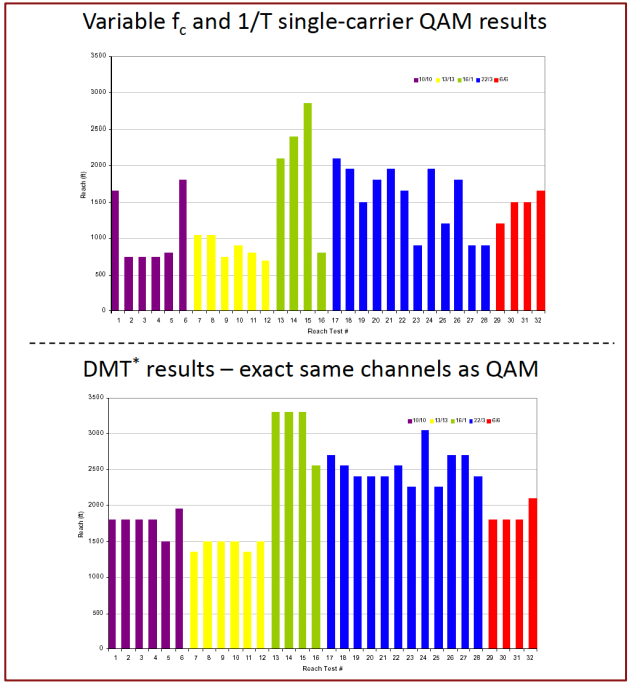
margin difference for half-band optimum versus full band



margin difference for half-band optimum versus full band

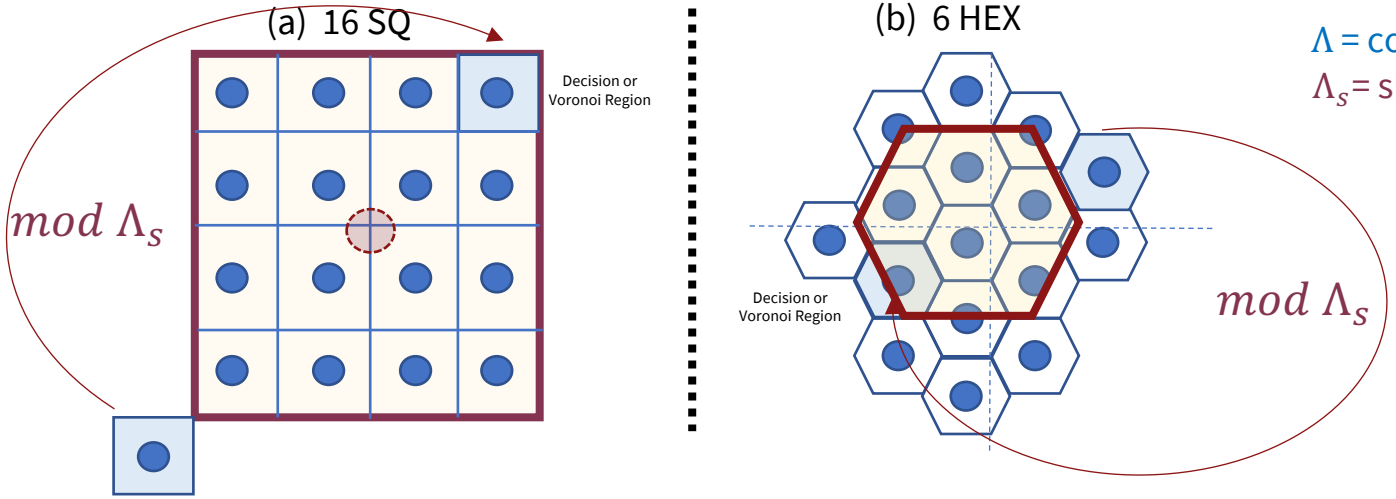
1993 ADSL Olympics – Bellcore
Margin differences at 1.6 Mbps, 4 miles, 11+dB
DMT 4x faster (6 Mbps) at 2 miles

2003 VDSL Olympics - Bellcore



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

Coding Gain Refresher



Λ = coding lattice for d_{min}
Λ_s = shaping lattice for $\mathcal{E}_x$

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

γ_f: fundamental **gain**

γ_s: **shaping gain**

Basic principle extends $\bar{N} \to \infty$
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

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

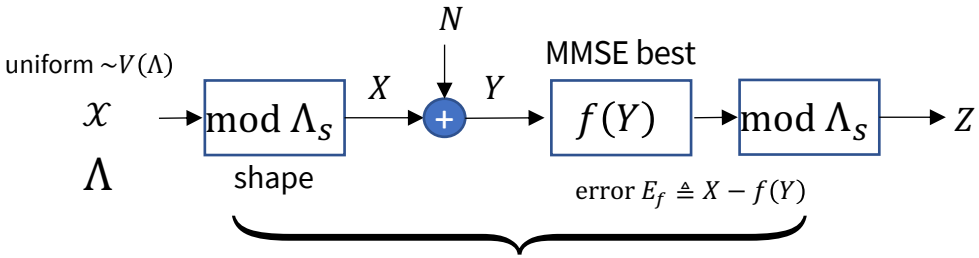
- There is loss of 0.5 bit/complex dimension using SQ constellations.

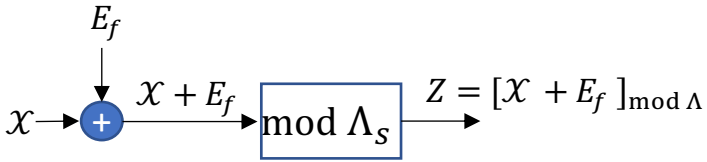
Forney's Crypto MMSE equivalence

1.53 dB (max) loss

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

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





- 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"

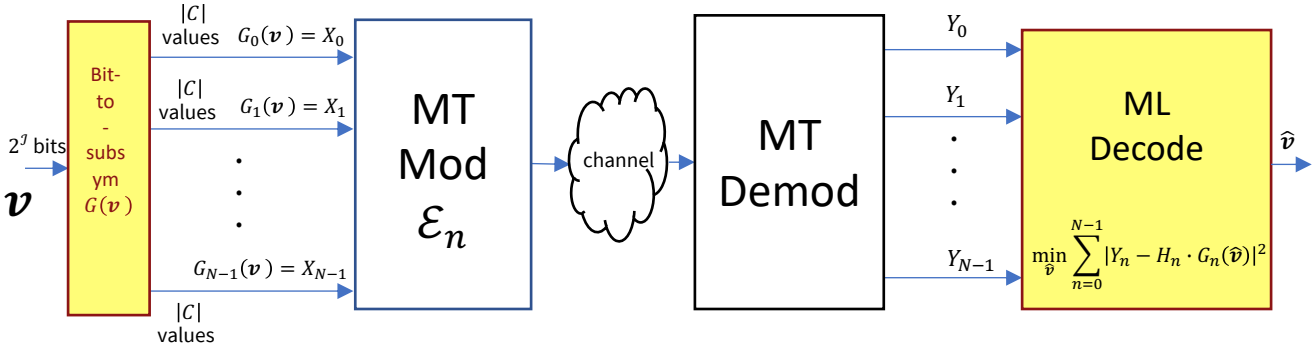
- Matrix/filtered-AWGN loss for "square" constellations

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

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

- When $\tilde{C} \leq 1$, $\gamma_{s,max} \cong 0$ dB.
  - There is tiny **low-SNR-AWGN** 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 is optimally handled linearly with parallel ind subchannels.
  - Nonlinear decision feedback needed when NOT parallel independent channels (chain rule).

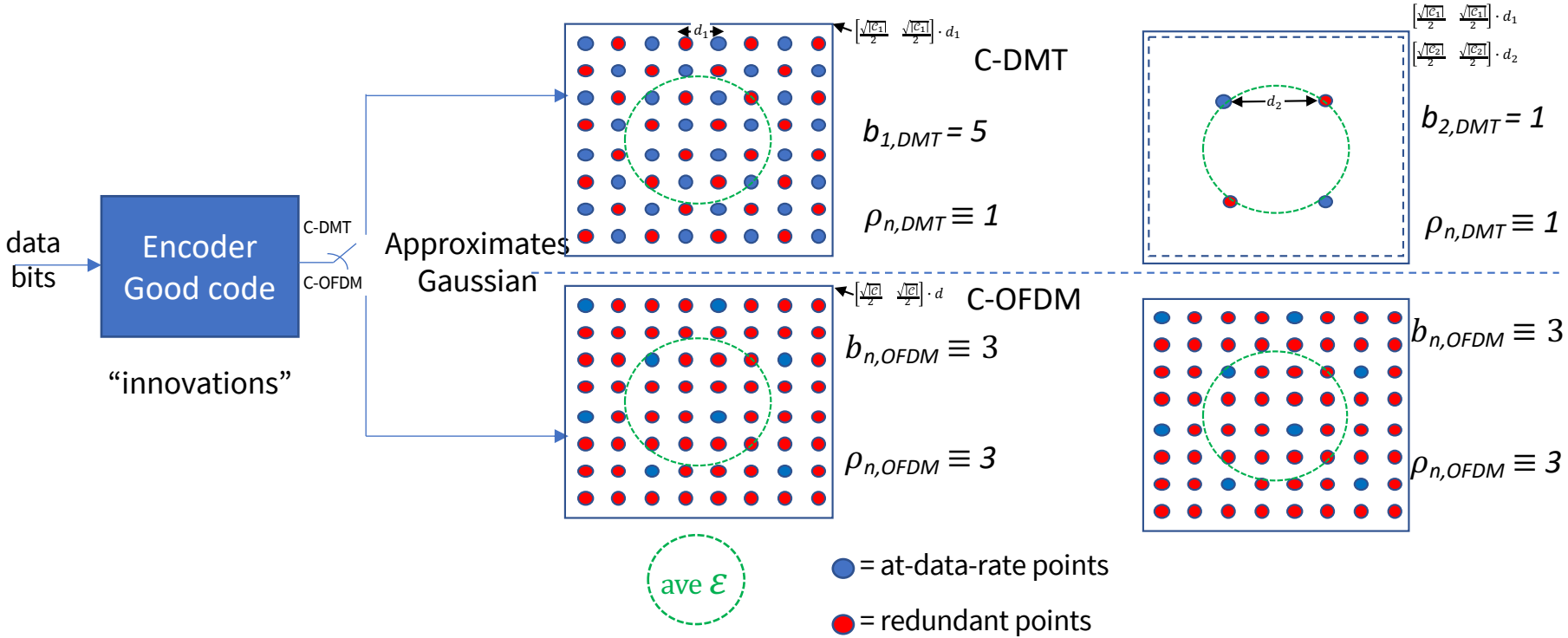
Coded-MT/OFDM



- 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.
- Usually wireless MIMO does allow "water-fill" over spatial dimensions (but not temporal).

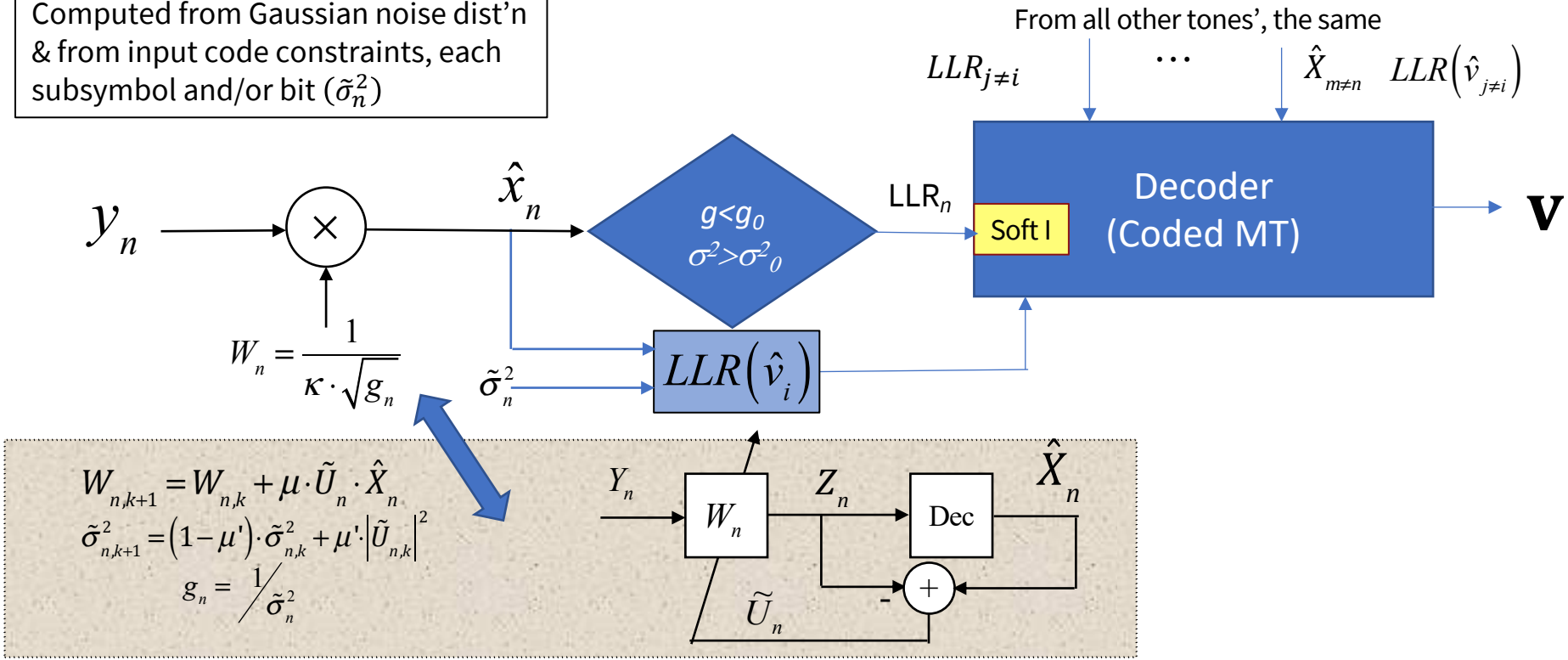
Comparison of variable and fixed constellation



- These types of system are heavily used in practice

Full MAP Decoder – Coded MT

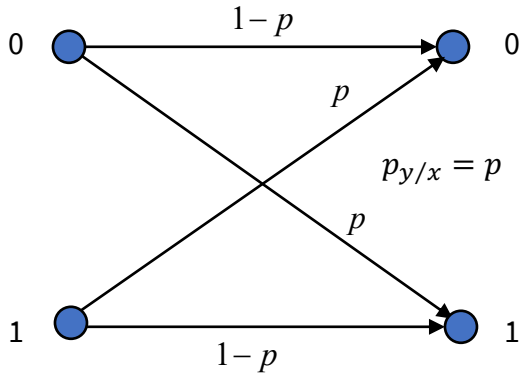
LLR = log likelihood ratio
Computed from Gaussian noise dist'n & from input code constraints, each subsymbol and/or bit ($\tilde{\sigma}_n^2$)



End Lecture 4

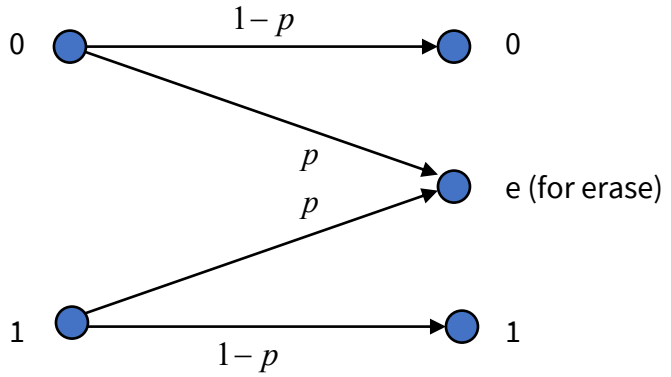
BSC and BEC

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



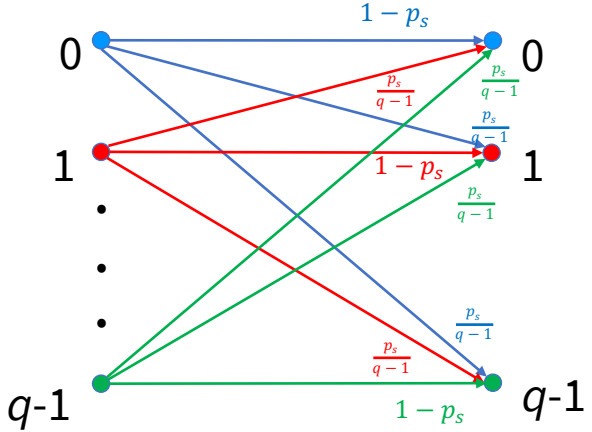
- **BEC** has $\bar{C} = 1 - p$.
  - $p = 1/2$ → 1/2 bits/dim reliable (no errors only erasures).
  - $p = 0$ → 1 bit/dimension reliably (makes sense).
  - $0 \le \bar{C} \le 1$.

- BEC is better than BSC (higher capacity) – decoders can use erasures with $N > 1$ to improve (reduce) P_e (soft info, 379A).



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.