STANFORD

Lecture 4

Separation Thm, & C-OFDM

April 8, 2026

JOHN M. CIOFFI

Hitachi Professor Emeritus of Engineering - RECALLED

Instructor EE379B – Spring 2026

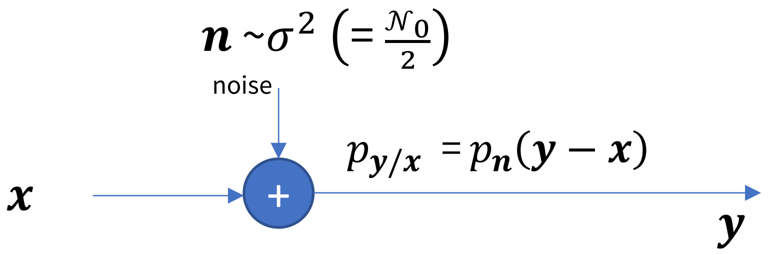
Announcements & Agenda

- Announcements
  - Problem Set #2 due Wednesday April 17 at 17:00
  - Sections 2.3-2.5, 4.3

- Agenda
  - Separation Theorem
  - Coded Multi-Tone
  - Some discrete channels' capacity
  - Statistical Channels' and soft information.

AWGN Good-Code/GAP Review (379A)

- Simple formula says a lot.



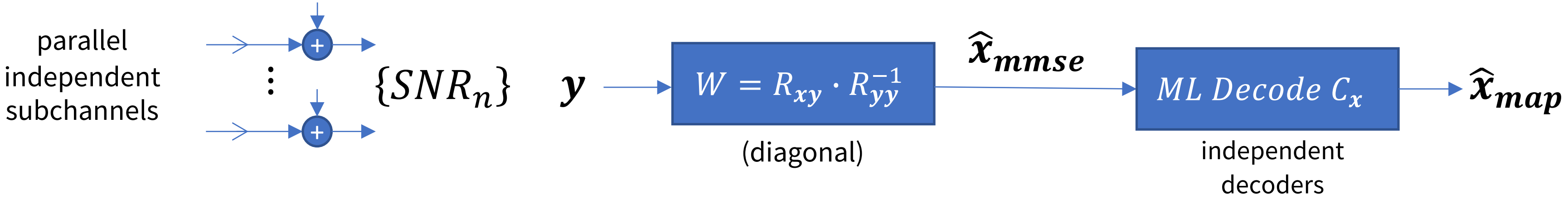
$$\bar{\mathcal{C}} = \frac{1}{2} \cdot \log_2\left(1 + \underbrace{\frac{\bar{\mathcal{E}}_{\boldsymbol{x}}}{\sigma^2}}_{SNR}\right)$$

- Often "gain" $\|h\|^2$ is absorbed into energy, really $g = \frac{\|h\|^2}{\sigma^2}$ so a "channel gain" $\bar{\mathcal{C}} = \frac{1}{2} \cdot \log_2(1 + g \cdot \bar{\mathcal{E}}_{\boldsymbol{x}})$.
  - This gain, g, is per real dimension, but if complex- baseband channel, the capacity is would be $\tilde{\mathcal{C}} = \log_2(1 + g \cdot \bar{\mathcal{E}}_{\boldsymbol{x}})$ bits/complex-symbol.
  - The designer must know context and use the same 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 $\bar{\mathcal{C}} = 3.33$ bits/dimension – and thus $\tilde{\mathcal{C}} = 6.67$ bits/complex subsymbol.
- What SNR gives 7 bits per dimension? $10 \cdot \log_{10}(2^{14} - 1) = 14 \cdot 3 = 42$ dB.
- Good-code families often have a constant gap Γ, so $\bar{b} = \frac{1}{2} \cdot \log_2\left(1 + \underbrace{\frac{\mathcal{E}_{\boldsymbol{x}}}{\Gamma \cdot \sigma^2}}_{SNR/\Gamma}\right)$ - **good** code families also have small gaps.
  - The good code's gap is constant over wide range of $\bar{b}$. (See 379A:L7-12, chapters 7 and 8.)

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 remain uncorrelated (but not independent because they arise from same encoder).
    - On average over all (Gaussian) codewords, these dimensions are independent, but not necessarily for a specific code.

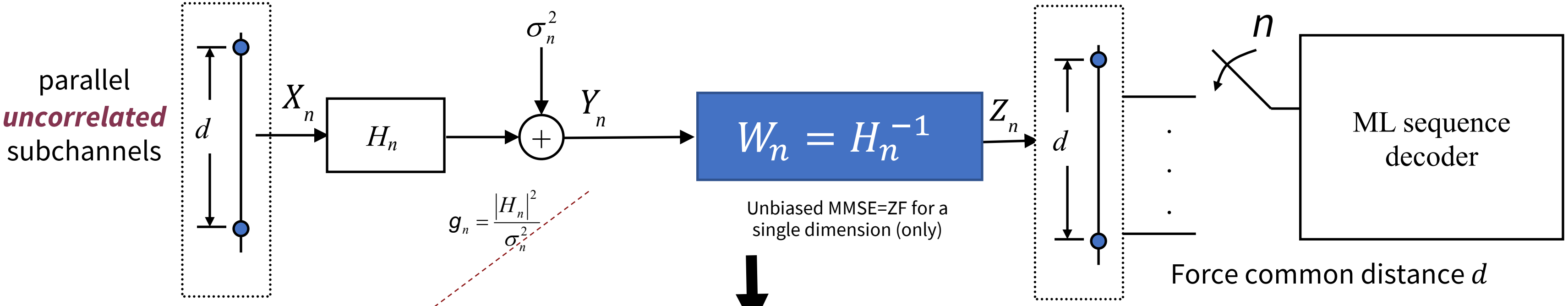
- This W is a scalar multiply for each uncorrelated dimension (that does not change SNR_n).
  - Does use of MMSE matter? (not for VC or DMT with $\tilde{b}_n = \log_2(1 + SNR_n)$)

- **BUT YES, IT DOES MATTER – IF, a constant** bits/subsymbol $b_n \equiv \tilde{b}$ (and/or SNR_{geo}) with **Coded OFDM.**
  - W 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.

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

See PS2.5 (Prob 2.20)

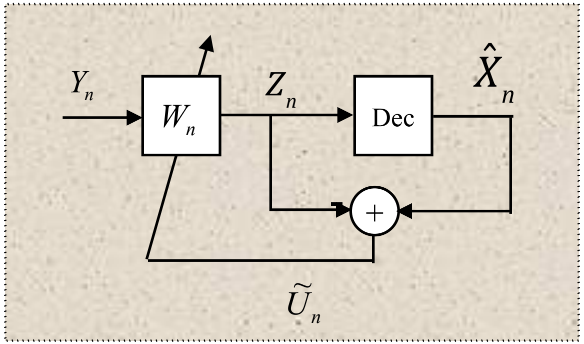
ZF Dimensional Normalizer same as MMSE



Must have each dimension scalar for ZF=MMSE

Zero-Forcing Algorithm

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



- ML Detector uses

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

Bias not present and

$$\sigma_{n,k}^2 = \mu' \cdot \sigma_{n,k-1}^2 + (1-\mu') \cdot \tilde{U}_{n,k}^2$$

- Or MMSE/ MAP Decoder is

$$\min_{\{X_{n,k}\}} \left\{ \sum_{k=0}^{\infty} \sum_{n=0}^{N-1} \frac{1}{2 \cdot \sigma_{mmse,n}^2} \cdot \left| \hat{X}_{n,k} - W_{biased,n} \cdot Y_{n,k} \right|^2 \right\}$$

Dimensions with small MMSE affect the the decoder more significantly.

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*

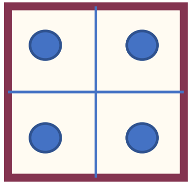
$$\bar{b}_n \leq \bar{\mathcal{I}}_n = \frac{1}{2} \cdot \log_2(1 + \mathcal{E}_n \cdot g_n) \ .$$

If the code does not have $\Gamma \to 0$, then Coded-OFDM will perform worse than DMT (any spectrum)

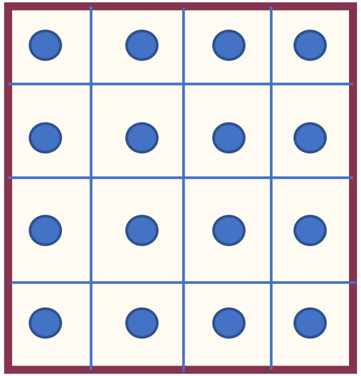
- Critical are:
  - the independent parallel dimensions (not code, the partitioned matrix/filtered AWGN), &
  - the good code $\Gamma \to 0$ dB, for which the input to the parallel dimensions comes from large constant constellation with subsymbol distribution approaching Gaussian ~ random coding,
  - 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.

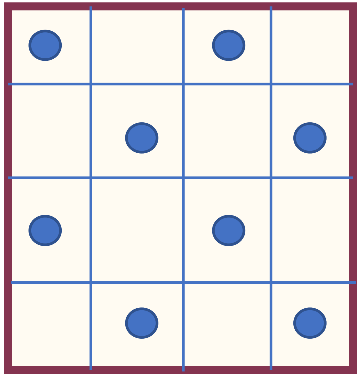


$\tilde{\mathcal{I}}_1 = 1$
$|C_1| = 4$



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

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



- Looks like 2 identical uses of a single AWGN with geometric-average SNR, equivalently $SNR = 2^{\tilde{\mathcal{I}}_{ave}} - 1$.
- There is redundancy, so tacit here is $\Gamma \cong 0$ dB

Separation Theorem 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 $\mathcal{I}(\tilde{x};\tilde{y})$.
  - Water-fill set gives highest data rate (highest mutual information).

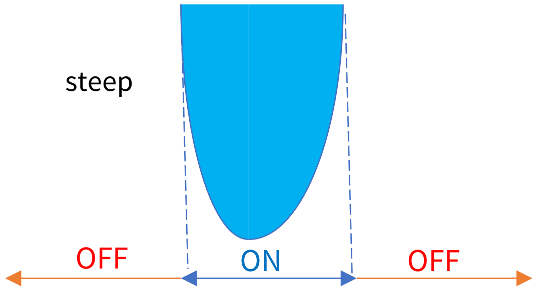
- Earlier examples show that water-fill is close to on/off.
  - So, if the designer guesses well the on/off, the system requires ALMOST no feedback of bit distribution to transmitter.
  - In practice, the constellation size and redundancy need specification, and thus on some indication of the value of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

- **Example:** Wireless "MCS" (modulation coding scheme) specifies code rate and constellation size only in feedback to transmitter. The on/off distribution?
  - 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.

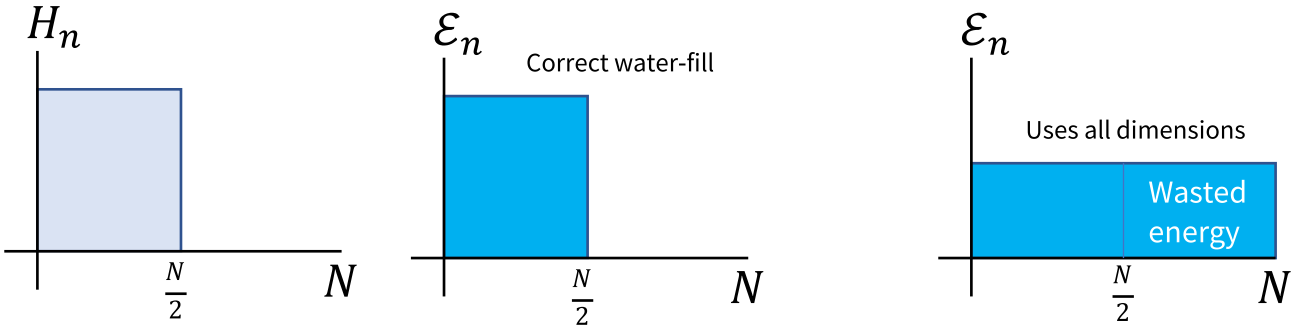


- The energy closely approximates flat:
  - RA: $K - \frac{\Gamma}{g_n} = \frac{\mathcal{E}_x}{L^*} + \frac{\Gamma}{L^*} \cdot \sum_{l=1,l\neq n}^{L^*} {}^1/_{g_l}$ is roughly the same (no one dimension dominates).
  - MA: $K - \frac{\Gamma}{g_n} = \Gamma \cdot \left(\frac{2^{\tilde{b}}}{\prod_{l=1}^{L^*}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

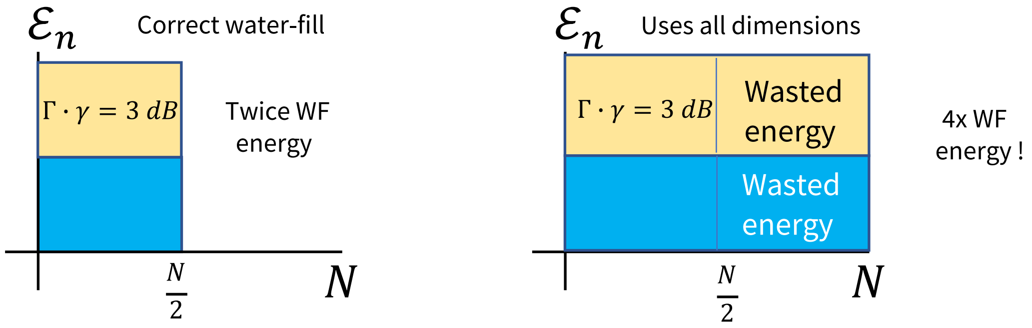
- This is true on waterfill's ENERGIZED ("*on*") dimensions, NOT for zeroed dimensions ("off").

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

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

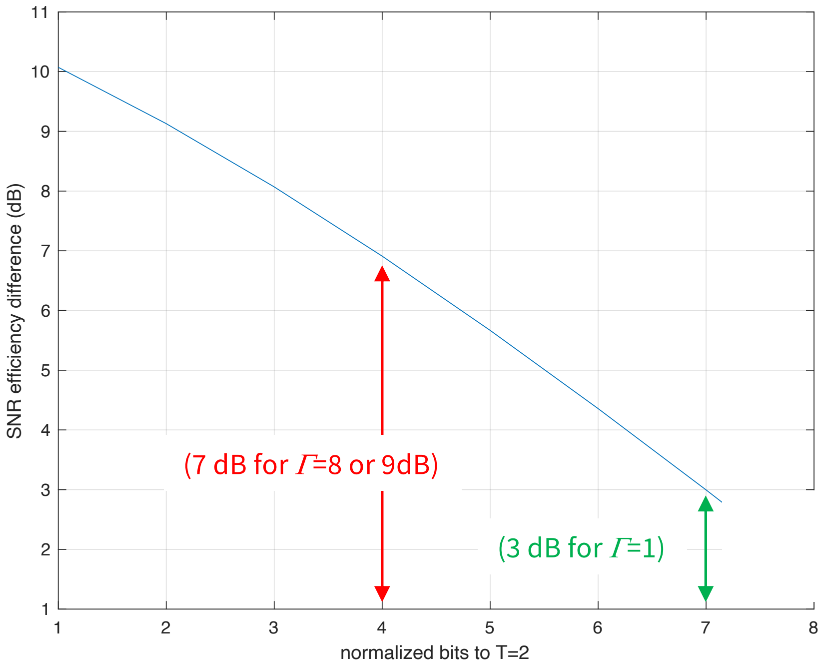


- The geometric SNR for water-fill is 3 dB higher if capacity-achieving codes are used
  - Or could run the water-fill system at same data rate at 3 dB less energy
- This amount is amplified below capacity by non-unity (not 0 dB) gap-margin product



This magnification by gap is for use of wrong bandwidth. For nonzero-gap water-fill, the loss is the non-zero-gap.

margin difference for half-band optimum versus full band

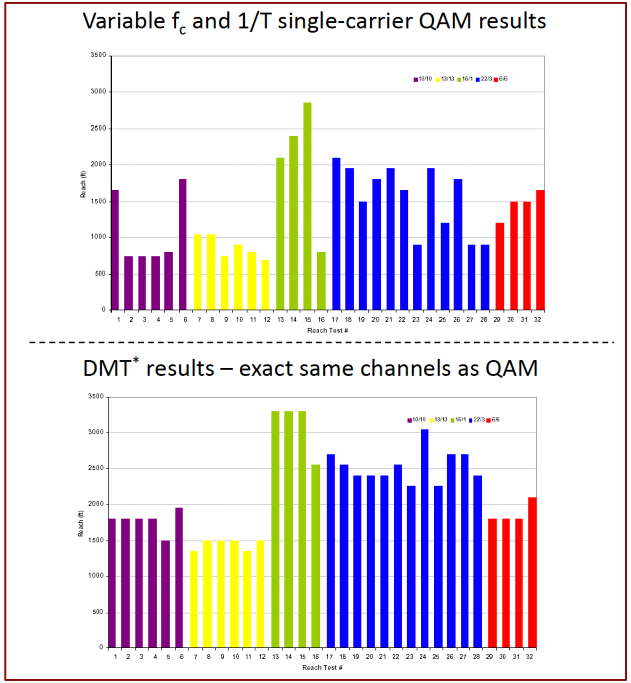


margin difference for half-band optimum versus full band

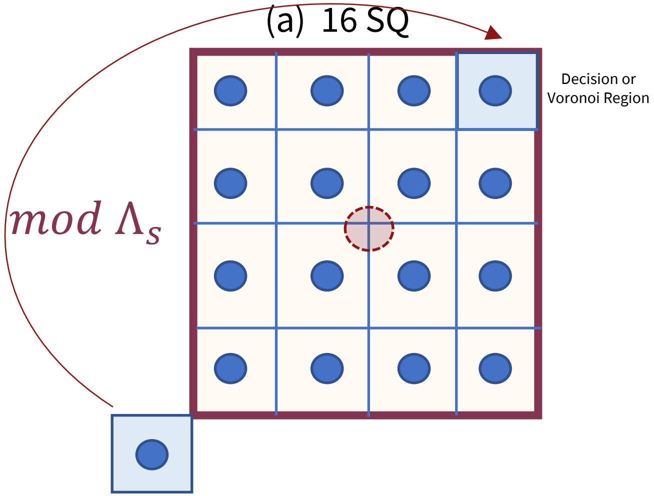
- Capacity of AWGN with WF is 8 bits/subsymbol (4 bits/dimension)

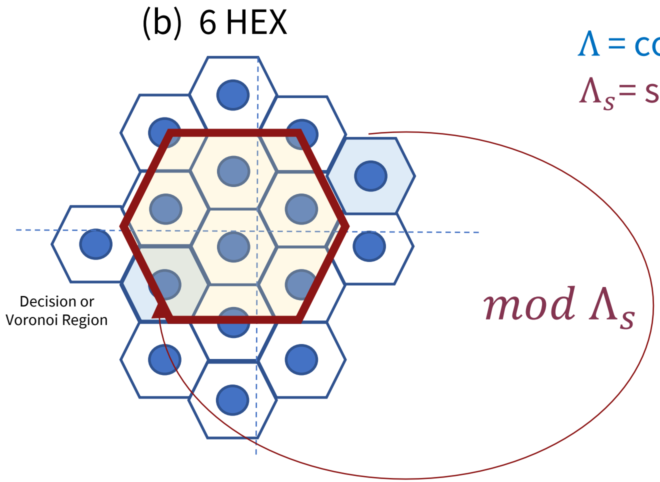
1993 ADSL Olympics – Bellcore
Margin differences at 1.6 Mbps, 4 miles, 11+dB
DMT 4x faster (6 Mbps) at 2 miles

2003 VDSL Olympics - Bellcore



Coding Gain Refresher





Λ = coding lattice for d_{min}

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

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

fundamental gain (γ_f)

shaping gain (γ_s)

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

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

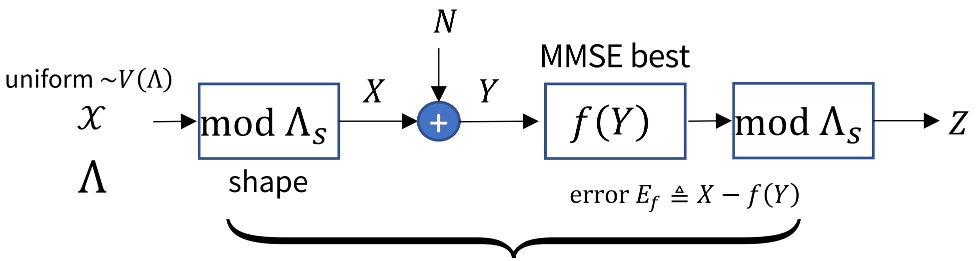
- There is loss of 0.5 bit/complex dimension using SQ constellations.

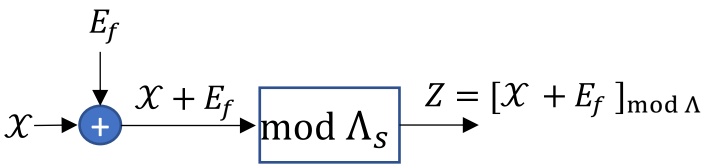
Forney's Crypto MMSE equivalence

1.53 dB (max) loss

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

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





So, the design only loses (up to) 1.53 dB and $\Gamma > 0$ does not magnify losses if the used code has good fundamental gain (uniform over the Voronoi shaping region), AND the waterfill is run for this gap.

Coded OFDM/MT

Subsection 4.4.1

SQ constellations vs "Gaussian"

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

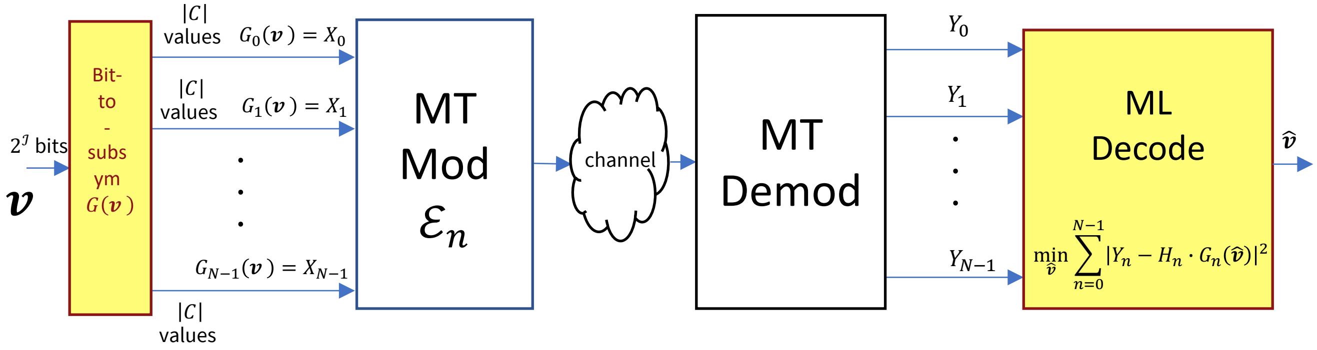
$\gamma_s \leq 1.53$ dB **shaping gain**

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

- 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 independent subchannels.
  - Nonlinear decision feedback needed when NOT parallel independent channels (chain rule).

So Coded-DMT/OFDM is still optimum for SQ Constellations, minus small γ_s

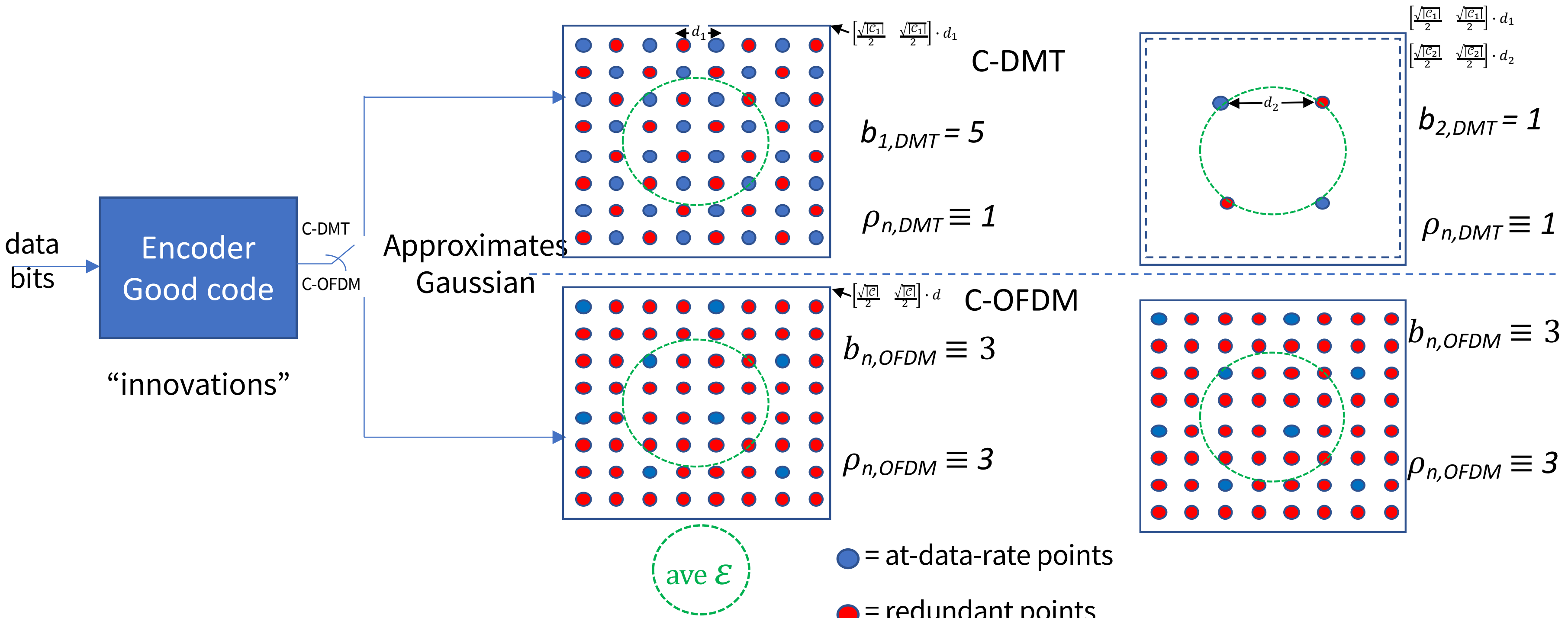
Coded-MT/OFDM



- 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 allows “water-fill” over spatial dimensions (but not temporal).
- Better use a good code – Wireless systems try to use LDPC or Polar (probably transitioning with GRAND to good product codes)

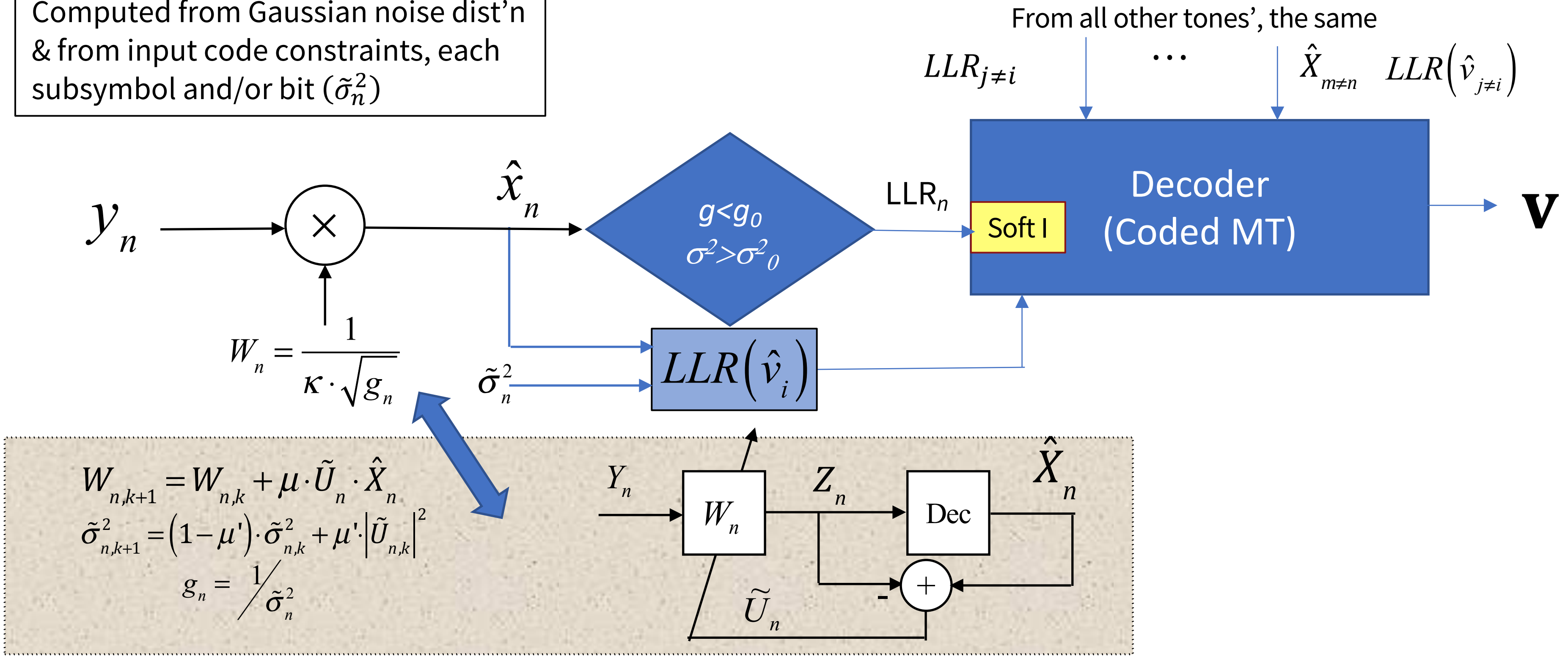
Comparison of variable and fixed constellation



- These types of system are heavily used in practice
- DMT uses a variable bit distribution with slightly simpler FEQ ; Coded OFDM needs to scale soft information in receiver prior to decoder
- DMT works is optimum, given any Γ with Γ loss ; Coded-OFDM only works if $\Gamma \cong 0$ dB – it will magnify a nonzero gap.

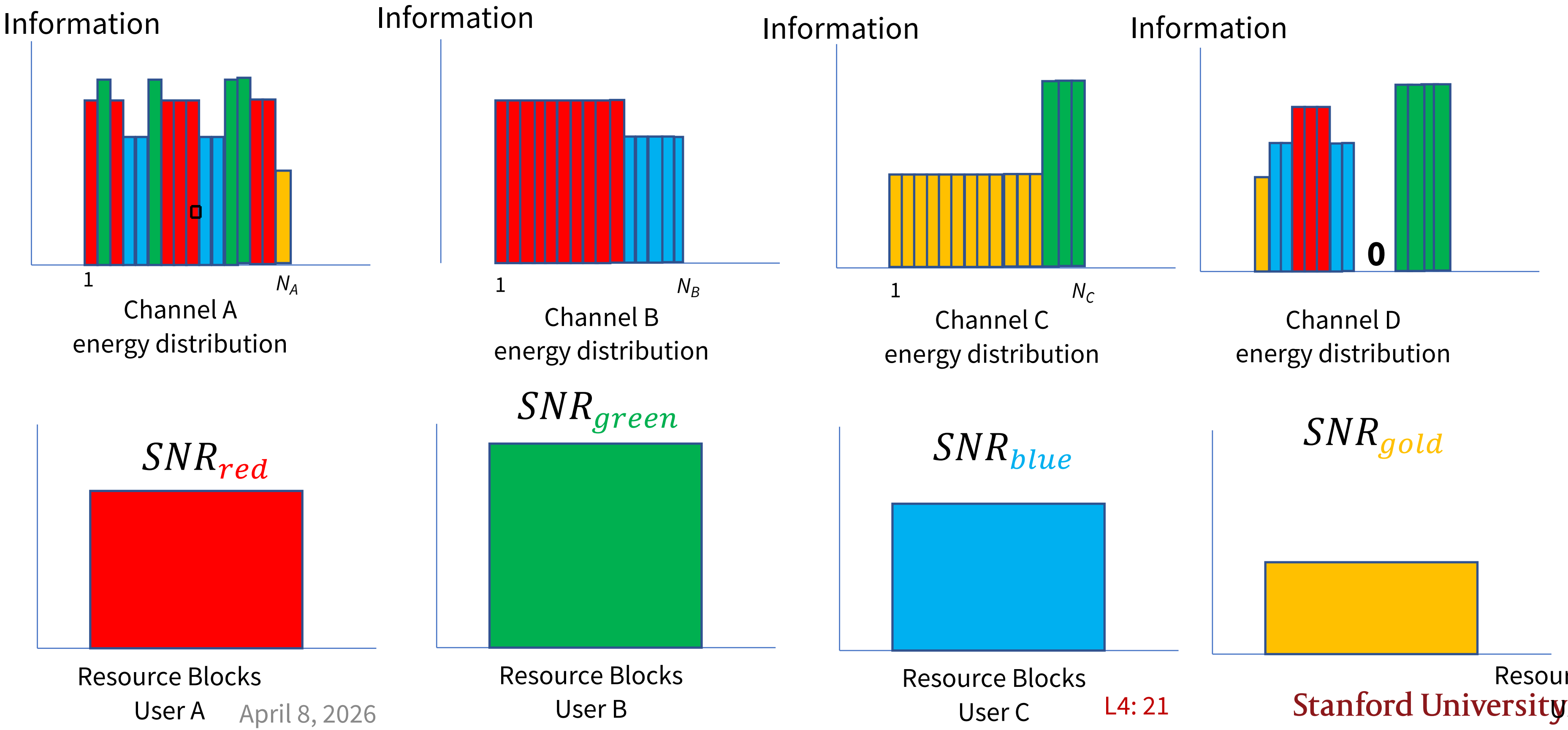
Full MAP Decoder – Coded MT

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



Crude Multiuser Separation Theorem

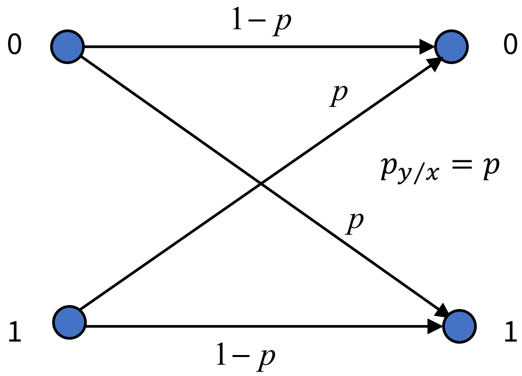
- Basically sort channels by SNR into resource blocks
  - Assign such blocks to different users (who would use different codes if transmitter and/or receiver in different places)



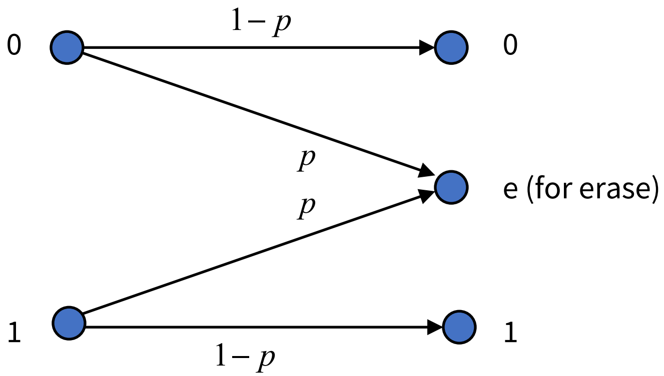
Some discrete channels’ capacity

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).
  - $0 \le \bar{\mathcal{C}} \le 1$.



- **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).
  - $0 \le \bar{\mathcal{C}} \le 1$.



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

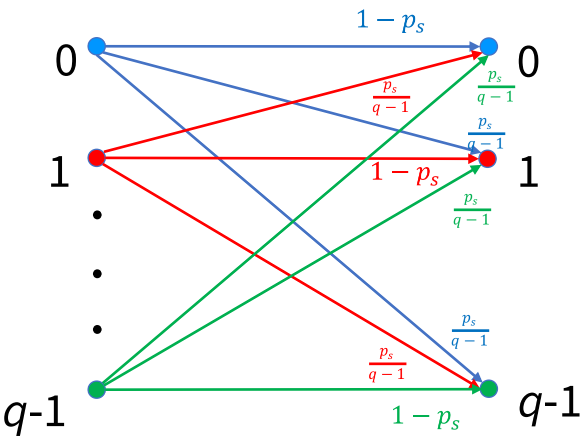
Symmetric DMC

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

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

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

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



Statistical Channels & Soft Information

Subsection 4.4.2

PS3.1 (Prob 4.13), PS3.3 (Prob 4.22) and PS3.4 (Prob 4.16)

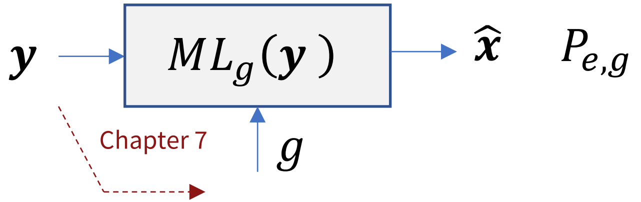
The random-gain AWGN

- See 379A-L5:6-18 for statistical channels.
- The AWGN (tones/space) has a gain parameter g , so $p_{[\boldsymbol{y}\ g]/\boldsymbol{x}}$ where g is random.
  - The gain is $g = \frac{|h|^2}{\sigma^2}$, or with variable gains/dimension, then $g_n \rightarrow {}^{SNR_{geo}}/_{\bar{\mathcal{E}}_x}$, following the Sep Th^m.
  - Effectively, each dimension has random $SNR_n = \mathcal{E}_n \cdot g_n$ where the designer can choose $\mathcal{E}_n$.

$$p_{[\boldsymbol{y}\ g]/\boldsymbol{x}} = p_{\boldsymbol{y}/[\boldsymbol{x}\ g]} \cdot \underbrace{p_{g/\boldsymbol{x}}}_{p_g}$$

$\boldsymbol{x}$ and g are independent.

- The ML/MAP receiver is a function of g.
  - It has error-probability distribution $P_{e,g}$, $\langle P_e \rangle \triangleq \mathbb{E}_g\left[P_{e,g}\right]$.



- For Rayleigh $\langle P_e \rangle = \frac{1}{2} \cdot \left(1 - \sqrt{\frac{\kappa \cdot SNR}{\kappa \cdot SNR + 1}}\right) \cong \frac{1}{4\kappa \cdot SNR}$ For large SNR

$\kappa = \frac{3}{M-1}$ for square QAM

Example MCS loading using Sep Th^m

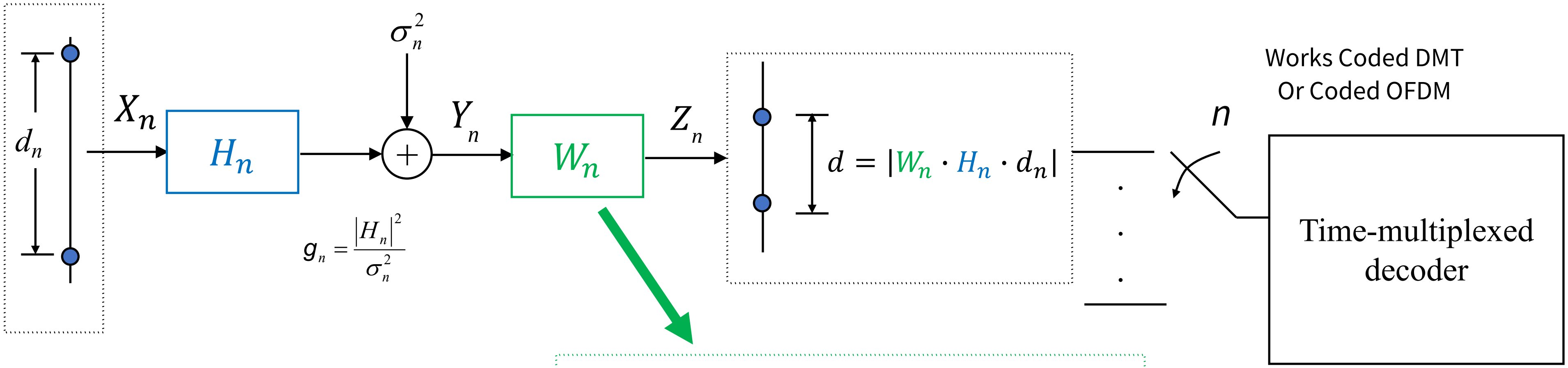
- Base code is the $r = \frac{1}{2},\ d_{free} = 10,$ 64-state code from Wi-Fi (see 379A-L7:17-18):

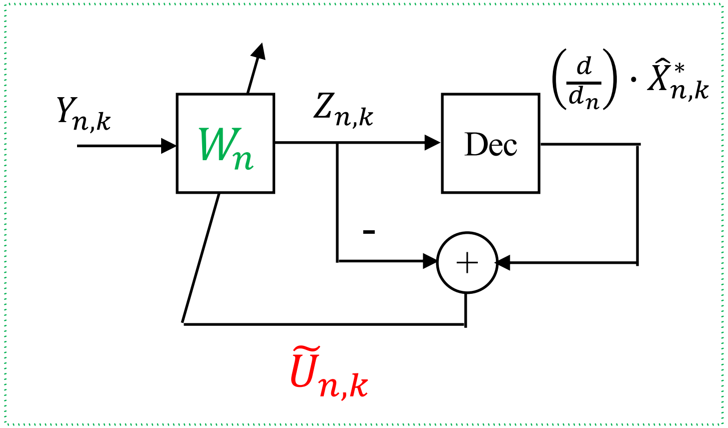
- 48 used tones (20 MHz-wide channel); $SNR_{geo} = 14.5$ dB (computed with 0 dB gap) and $\mathcal{E}_n \equiv \mathcal{E}$.
  - Puncturing options
    - No puncturing $r = \frac{1}{2},\ d_{free} = 10$
    - Puncture 3 bits from 12, $r = \frac{2}{3},\ d_{free} = 6$
    - Pucture 4 bits from 12, $r = \frac{3}{4},\ d_{free} = 5$

  - MCS Options
    - Uncoded 4QAM – nominally $P_e \cong 10^{-7}$ for a single AWGN; however, Sep Thm only applies with good codes – not viable.
    - Coded 16QAM at rate 2/3 has coding gain of 6 dB (although increase to 16QAM constellation costs 7 dB), but still able to get $P_e \cong 10^{-6}$ (Gray coding, BICM).
    - Coded 16QAM at rate ¾ has slightly higher gain, so also would work.
    - The 16QAM system thus has $b = \frac{3}{4} * 2$=1.5 bits/dimension so indeed better rate than uncoded 4 QAM also.

- This presumes the $SNR_{geo} = 14.5$ dB remains constant.
  - Each use may need a new MCS solution when there is time-varying fading, which means SNR_{geo} varies.
  - Wi-Fi trains on each channel use (so when channel is available and there is data to send).
  - Even video streams today are divided into packets and not sent as continuous stream.

Multichannel Normalizer (or “FEQ”), See L4:20





Zero-Forcing Algorithm

$$W_{n,k+1} = W_{n,k} + \mu_n \cdot \tilde{U}_{n,k} \cdot \left(\frac{d}{d_n}\right) \cdot \hat{X}^*_{n,k}$$

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

Channel estimate

$$\hat{H}_n = \frac{d}{d_n \cdot W_n}$$

Noise Estimate

$$\hat{\sigma}_n^2 = \frac{\tilde{\sigma}_n^2}{|W_n|^2}$$

Rcvr can also estimate actual gain distribution p_g from this.

Calculation of average and outage error probs

Definition 4.4.4 [Outage Probability] *The* **outage probability** *differs from the* **random-error probability** *according to (they differ in the sum's value range)*

$$\langle \bar{P}_e \rangle = \sum_{g > g_{out}} p_g \cdot \bar{N}_e \cdot Q\left[\sqrt{\frac{3 \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g \cdot d_{free}(r)}{|C| - 1}}\right] \quad (4.151)$$

$$\bar{P}_{out} = \sum_{g \leq g_{out}} p_g \text{ , respectively,} \quad (4.152)$$

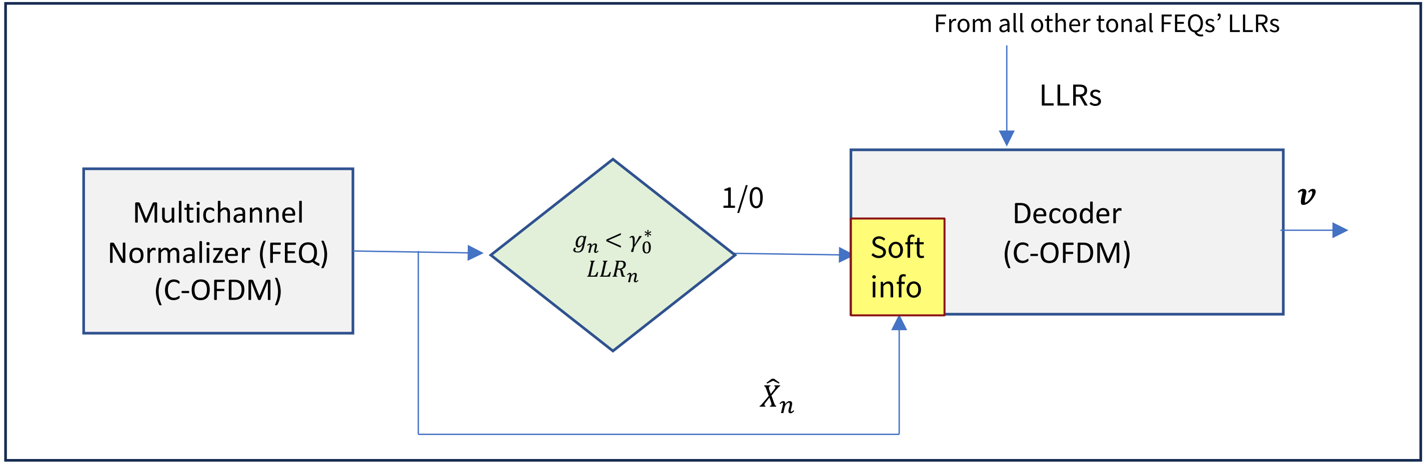
where g_{out} is a threshold channel SNR to be determined so that (4.151) holds, while the "outage" corresponding to lower gains (meaning very poor performance with high error probability) must be accommodated by the receiver's erasure marking in decoding. The fraction $\frac{3}{|C|-1}$ can be adjusted to κ if the design uses non-square constellations, but the concept is the same.

In general, soft decode.

- Average error prob and outage depend on a channel-gain distribution p_g that is given or measured (by receiver).
- The average error probability is weighted by p_g for just those transmissions not in outage.
- **Quasi-stationary assumption** is that the gain g_n is constant for each decoder use (over the codeword) – it can vary over n , but not time.
  - Think tone for multicarrier – what this really means is SNR_{geo} is constant over each decoder use of tones/spatial dimensions in the context of separation theorem.
  - That is, coherence **time** is best longer than a codeword.
  - If not, hard-decoding applies at lower data rate with that $\langle P_b \rangle^{\left\lfloor \frac{d_{free}+1}{2} \right\rfloor}$ for a BSC.

LLRs generally with FEQ & Decoder

- FEQ provides LL/soft information to the decoder.



- This works if code is good $\Gamma \to 0$ dB;
- Even works well if erasures used, then hard (e.g., Reed Solomon) code recovers if < P/2 erred ss.
  - Upcoming Ergodic Water-Fill (both MA and RA) provide theoretical guidance, but not practical.
  - LLR_n=0 if $g_n < \gamma_0^*$.
- **LL/Erasure avoids:**
  - "feedback delay that $g_n < \gamma_0$, so don't transmit" would take too long (issue largely only in wireless)
  - On/off or variable gain that frustrates power amplifiers (on/off transients) and receiver automatic gain-control circuits
- Adaptive power is often feasible for space (MIMO) to use/not use certain spatial channels, but not for time-frequency.

End Lecture 4