STANFORD

Lecture 4

Separation Th^{m} & Coded-OFDM Loading

April 8, 2026

JOHN M. CIOFFI

Hitachi Professor Emeritus of Engineering - RECALLED

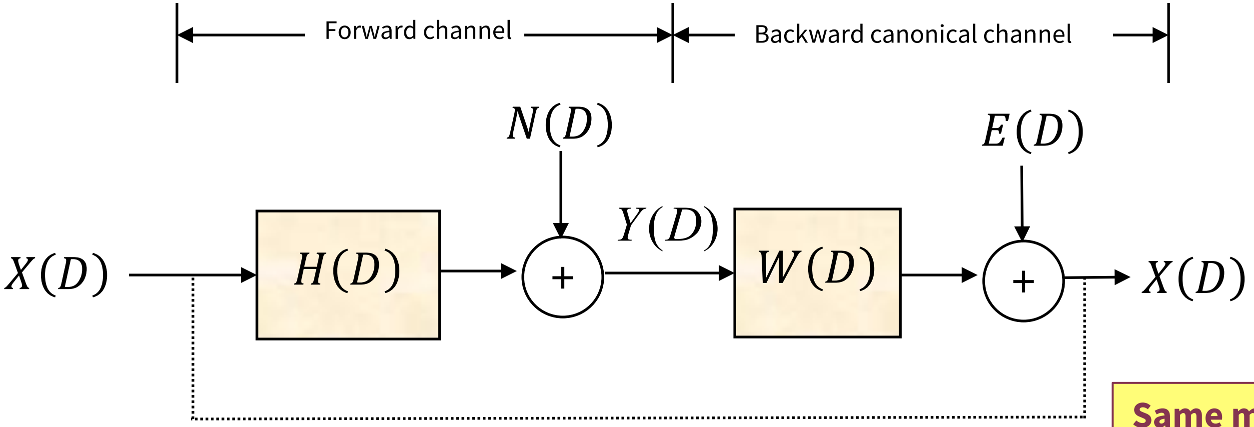
Instructor EE379B – Spring 2026

Announcements & Agenda

- Announcements
  - Problem Set #2 due Wednesday April 15 at 18:00
  - Sections 2.3-2.5, 4.3

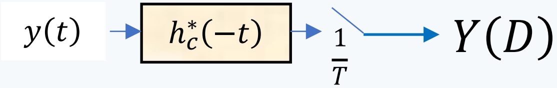
- Agenda
  - Separation Theorem and Modulation
  - Dynamic Coded-OFDM/MT Loading
  - Ergodic Coded-OFDM Loading
  - Quasi-Static Statistical Loading

Forward and its Backward Canonical Models



Same mutual information

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



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

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

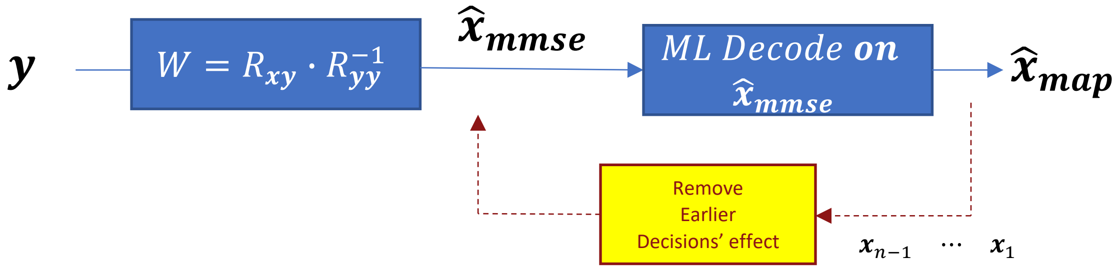
- The MAP and MMSE determine the performance, and also the chain rule suggests a simpler decoder:

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

$= \log_2(SNR_{mmse})$

The $R_{\tilde{x}\tilde{x}}$ must be nonsingular, equivalent to PWC.

Forward
Backward



Still MAP if "previous" decisions are correct sequentially decodes

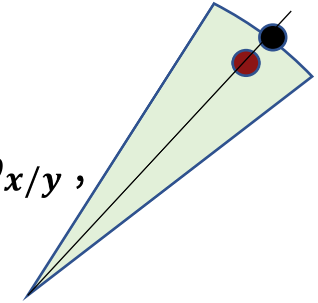
Optimal Detectors for Good Codes

- $\boldsymbol{x}$ 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

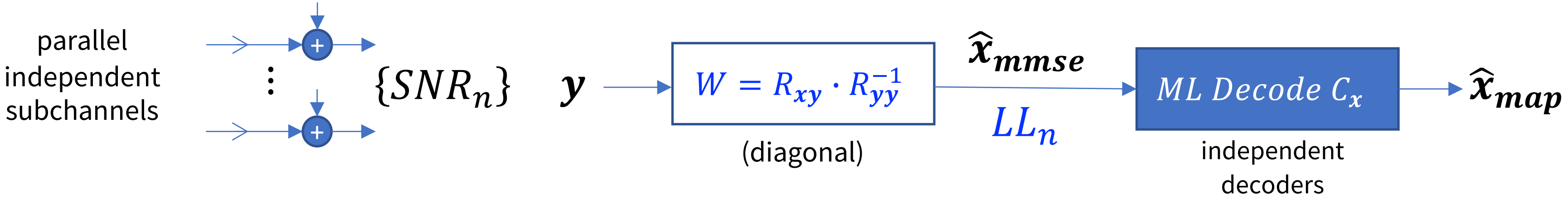
W is MMSE filter

- 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 $\mathbb{I}(x_n; y_n) = \log_2 SNR_{mmse,n}$, since they are independent.

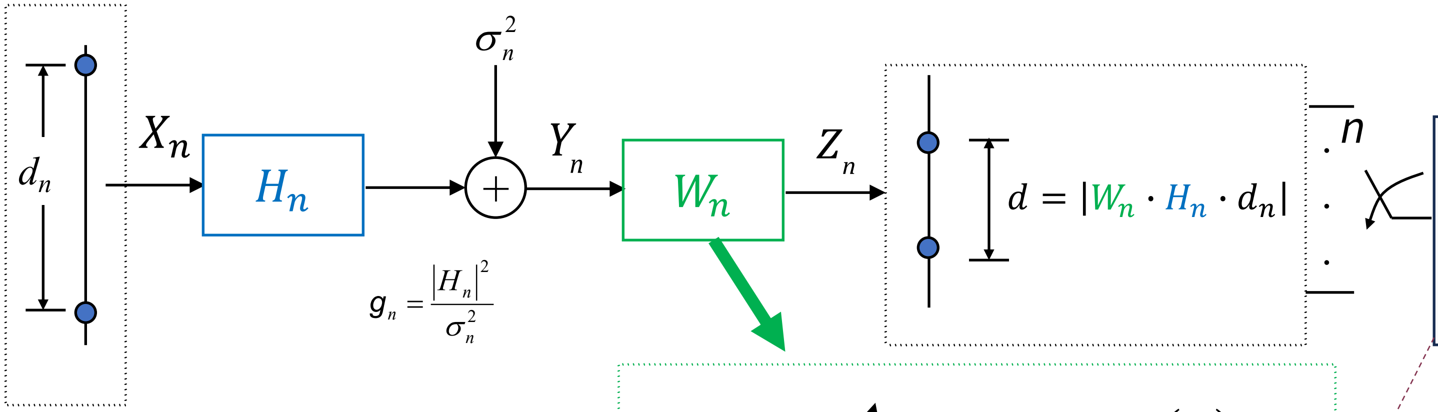
- Suppose each dimension is a dimension within the **same** code?
  - The dimensional subsymbols remain uncorrelated (but not independent because they arise from same encoder).
    - On average over all codewords, these dimensions are independent, but not necessarily for a specific code.

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

- **BUT** with **Coded OFDM, IT DOES MATTER** → $b_n \equiv \tilde{b}$ (and/or SNR_{geo}) .
  - W_n impacts the weight of different dimensions before the final ML decoding; this (it turns out) is the same as the earlier xmit "bit-loading," in effect, for same energy distribution/R_{xx}.
  - But C-OFDM reduces channel-state-information (CSI) feedback.

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

Multichannel Normalizer (or “FEQ”)

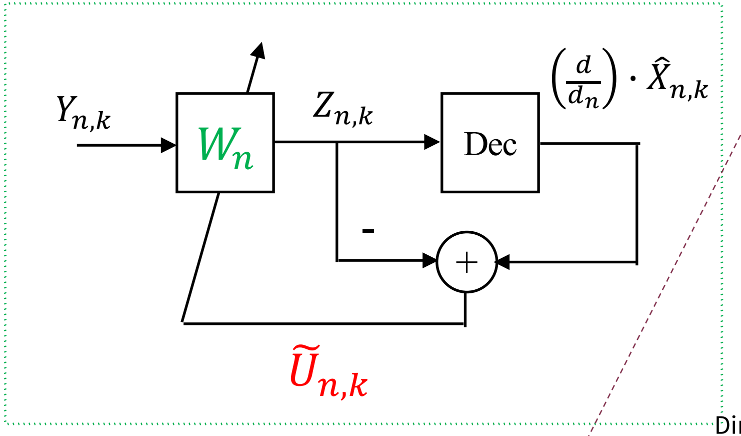


Works Coded DMT
Or Coded OFDM

ML Detector uses

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

Inverse noise scales value.



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

- ~ MMSE/ MAP Decoder is

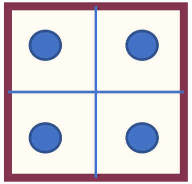
$$\min_{\{X_{n,k}\}} \left\{ \sum_{k=0}^{\infty} \sum_{n=0}^{N-1} \frac{1}{2 \cdot \hat{\sigma}_n^2} \cdot \left| \hat{X}_{n,k} - W_{biased,n} \cdot Y_{n,k} \right|^2 \right\}$$

Dimensions with small MMSE/noise affect the the decoder more significantly.
MMSE for scalar EQ needs scaling for sequence MMSE/MAP

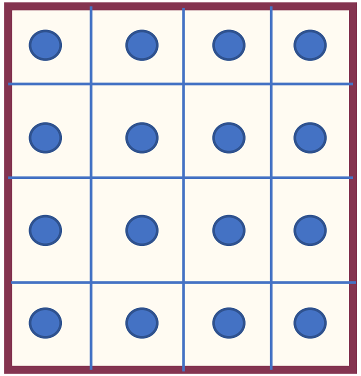
Section 4.3.6.2

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.

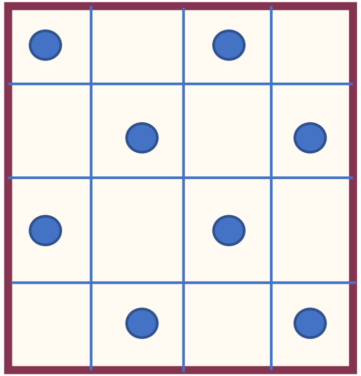


$\tilde{\mathcal{I}}_1 = 1$
$|C_1| = 4$



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

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



- 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

See Backup for another example

Separation Theorem is widely applicable

- This works for partitioning with
  - SVD,
  - Eigenvectors,
  - DFT/FFT (becomes Coded-OFDM here),
  - other bases. &
  - Users individually within a multiuser design.

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

- The ST holds **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).

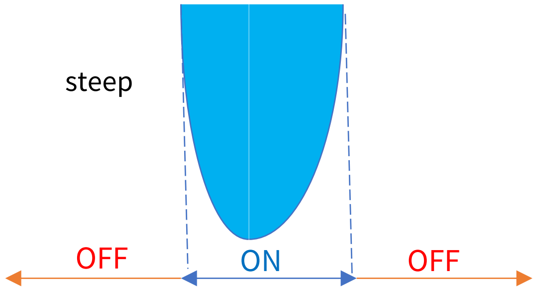
- Earlier examples show that water-fill is close to on/off.
  - So, if the designer guesses well the on/off, the system requires ALMOST no feedback of bit distribution to transmitter.
  - In practice, the $[r\ ,|C|]$ need specification, and thus on some indication of the value of $\mathcal{I}(\tilde{x};\tilde{y})$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

- **Example:** Wireless "MCS" (modulation coding scheme) returns only $[r\ ,|C|]$ only to xmit. & 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.

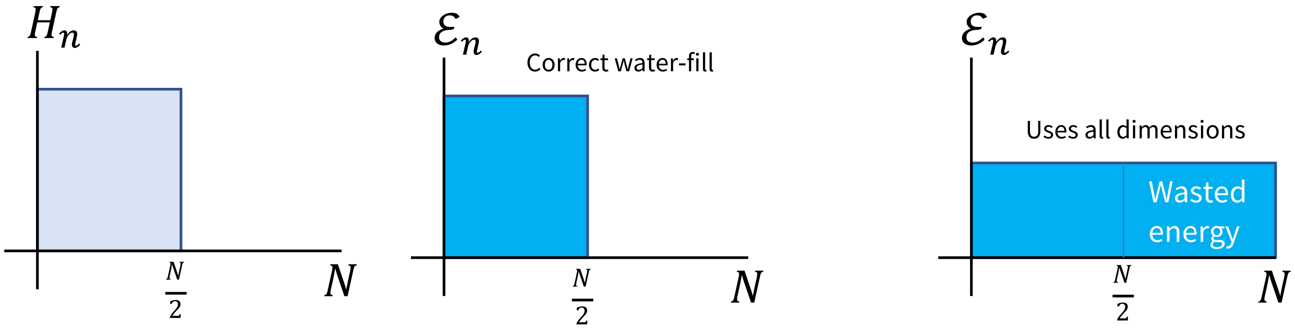


- The energy closely approximates flat:
  - RA: $K-\frac{\Gamma}{g_n}=\frac{\mathcal{E}_x}{L^*}+\frac{\Gamma}{L^*}\cdot\sum_{l=1,l\neq n}^{L^*} {}^{1}/_{g_l}$ is roughly the same (no one dimension dominates).
  - MA: $K-\frac{\Gamma}{g_n}=\Gamma\cdot\left(\frac{2^{\tilde{b}}}{\prod_{l=1}^{L^*} g_l}\right)^{1/L^*}-\frac{\Gamma}{g_n}=\frac{\Gamma}{g_n}\cdot\left\{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

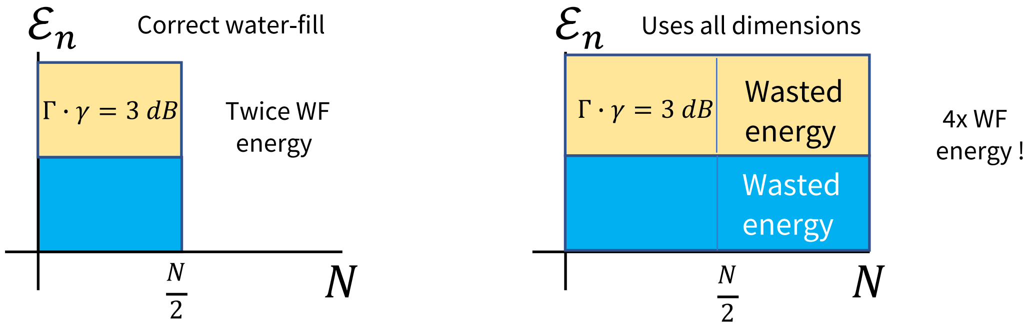
- This is true on waterfill′s ENERGIZED (“*on*”) dimensions, NOT for zeroed dimensions (“off”).

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



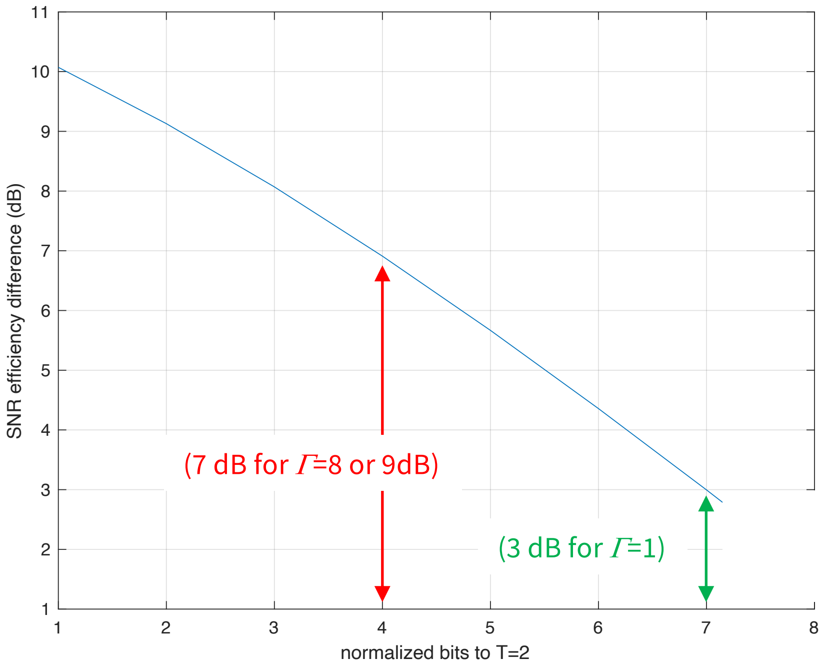
- The geometric SNR for water-fill is 3 dB higher if capacity-achieving codes are used
  - Or could run the water-fill system at same data rate at 3 dB less energy
- This amount is amplified below capacity by non-unity (not 0 dB) gap-margin product



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

See PS2.4 (Prob 2.14)

margin difference for half-band optimum versus full band



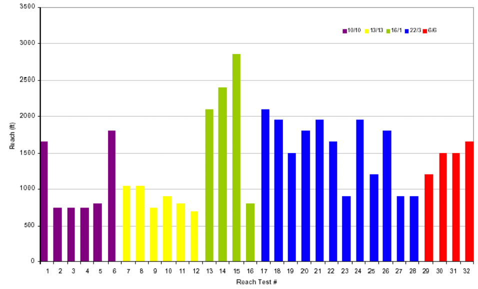
margin difference for half-band optimum versus full band

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

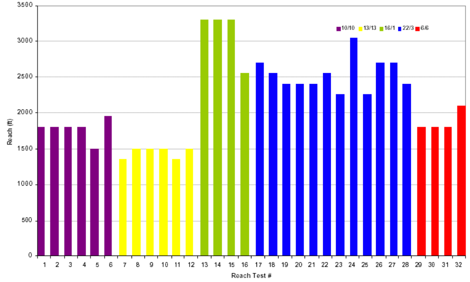
1993 ADSL Olympics – Bellcore
Margin differences at 1.6 Mbps, 4 miles, 11+dB
DMT 4x faster (6 Mbps) at 2 miles

2003 VDSL Olympics - Bellcore

Variable f_c and 1/T single-carrier QAM results



DMT* results – exact same channels as QAM



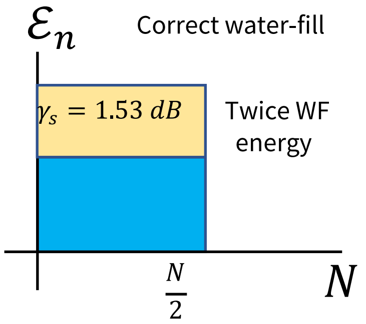
SQ constellations vs “Gaussian”

- There is always a loss for a non-hyper-spherical constellation boundary on the (any matrix/filtered) AWGN within uniform constellation-point probabilities.
  - 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

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

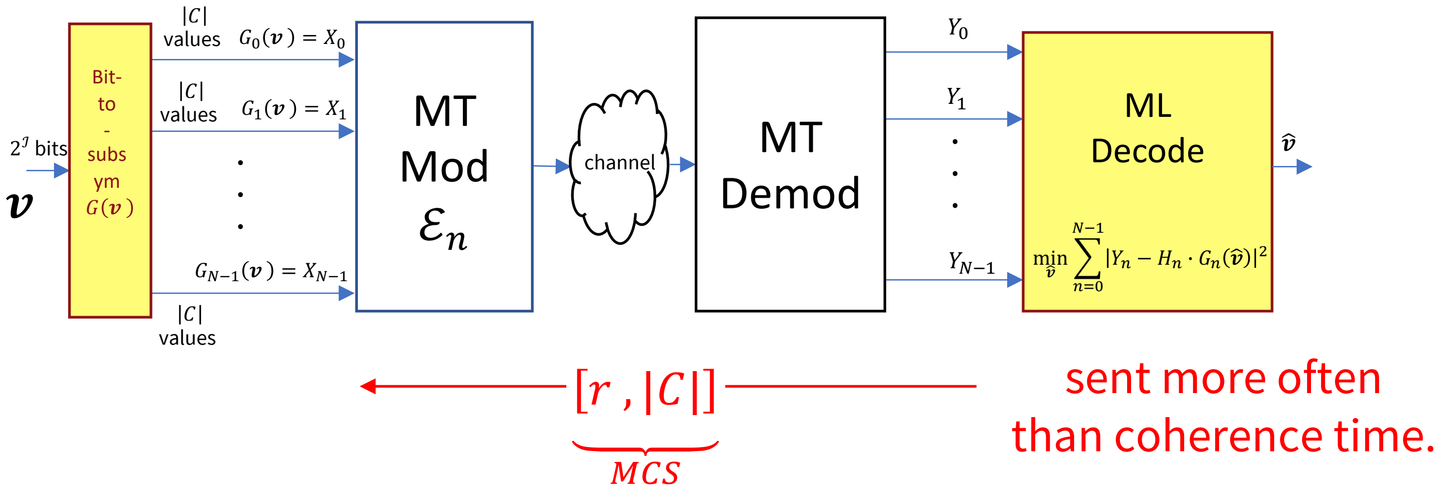


- There is loss of 0.5 bit/complex dimension using SQ constellations.
- **Probabilistic Amplitude Shaping (PAS, see 379A-S12A)** induces the subsymbol Gaussian-like distribution by making “corner” points less likely – requires additional constellation expansion, but recovers the 1.5dB.
  - Separate code of length ~1000 (applied to same subsymbols as the fundamental-gain good code, no extra delay).

Dynamic Coded OFDM/MT Loading

Subsection 4.4.1

Coded-MT/OFDM with Dynamic Loading



- C-OFDM treats a pre-agreed known set of dimensions as repeated constant SNR_{geo} dimensions.
  - There is no transmitter bit loading, and energy is on/off on the pre-agreed set.
  - Only $[r \,, |C|]$ - the MCS (Modulation Coding Scheme) returns to transmitter.

- The MT/OFDM 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), and the M and F matrices require address also.
- Best uses a good code – Wireless systems try to use LDPC or Polar (probably transitioning with GRAND to good product codes)

Example MCS loading using Sep Thm

- Base code is the $r = \frac{1}{2}$, $d_{free} = 10$, 64-state code from Wi-Fi (see 379A-L7:17-18), $\bar{b}$ and $\tilde{N}$ for C
  - $k, n, p = n - k$ are for the binary convolutional code. (Wi-Fi-like)

- 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, $r = \frac{3}{4}$, $d_{free} = 5$

BICM Pe (Sec 2.4) with Gray Coding

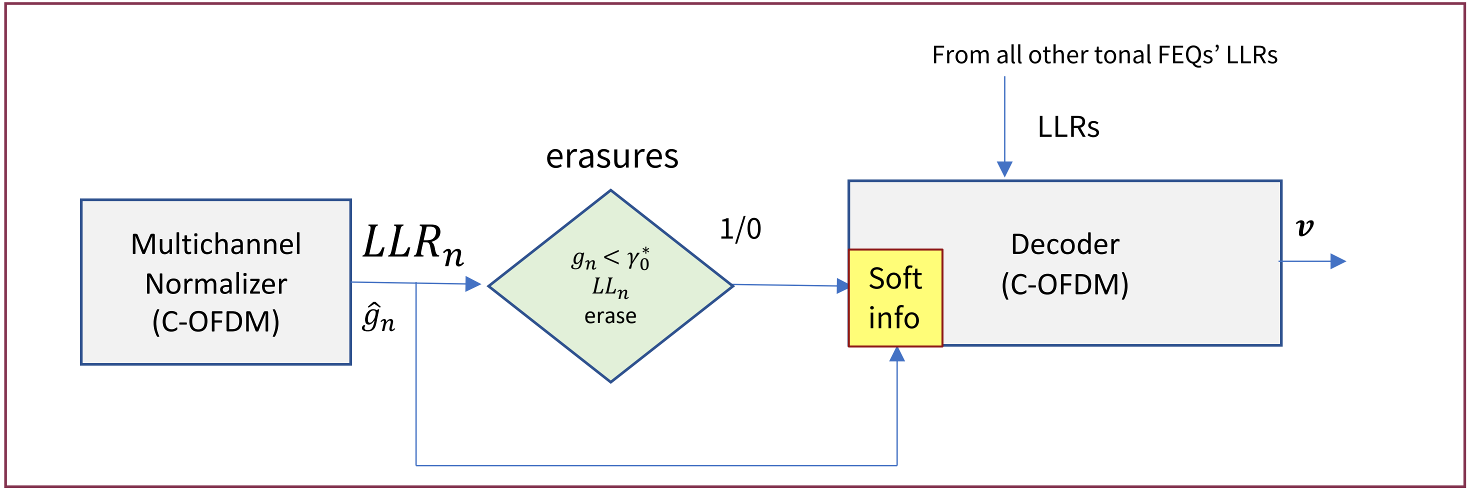
$$\bar{P}_e \approx \bar{N}_e \cdot Q\left[\sqrt{\frac{3 \cdot 2^{2 \cdot \frac{p}{N}} \cdot r \cdot d_{free} \cdot SNR}{2^{2 \cdot \left[\bar{b} + \frac{p}{N}\right]} - 1}}\right] \quad \text{for } \bar{b} \geq 1$$

  - 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: **This is 1.5 bits/dimension.** The rate 3/4 code has $d_{free} = 5$ (7dB), but increase 8SQ to 16QAM constellation costs 7 dB/2 from 8SQ→16SQ with 3dB BICM distance gain and 1.3 dB BICM noise loss so the net 7-.5-1.3 = **gain 5.2 dB**, so 14.5-3.5+5.2 =**16.2 dB** Q-func arg, and easily obtains $P_e \cong 10^{-7}$.
    - Coded 64QAM **at also 1.5 bits/dimension** with r=1/2 has $d_{free} = 10$ (10 dB), but the increase 8SQ to 64QAM constellation costs 9.8dB, and there is 9 dB BICM distance gain and 3 dB noise increase, so 10-9.8-3+9 **gain 6.2 dB** also at **1.5 bits/dimension**. So, 14.5-3.5 +6.2 =**17.2 dB**. Both meet target.
    - Maybe 64SQ transmission at **2 bits/dim with 2/3 code** (assuming 3 dB/bit loss, so 14.2 dB), just barely? (No, because weaker code at 2/3).
    - Not really getting 7 dB from this code in data rate doubling
    - Double symbol rate would achieve it, but 20 MHz is fixed. So here lowest r for largest $|C|$ that meets rate/Pe.

- Each use may need a new MCS solution with time-variation, which means SNR_{geo} varies →$[r, |C|]$ varies.
  - Wi-Fi trains/updates on each channel use and can be tabulated versus SNR_{geo}.

LLs generally with FEQ & Decoder

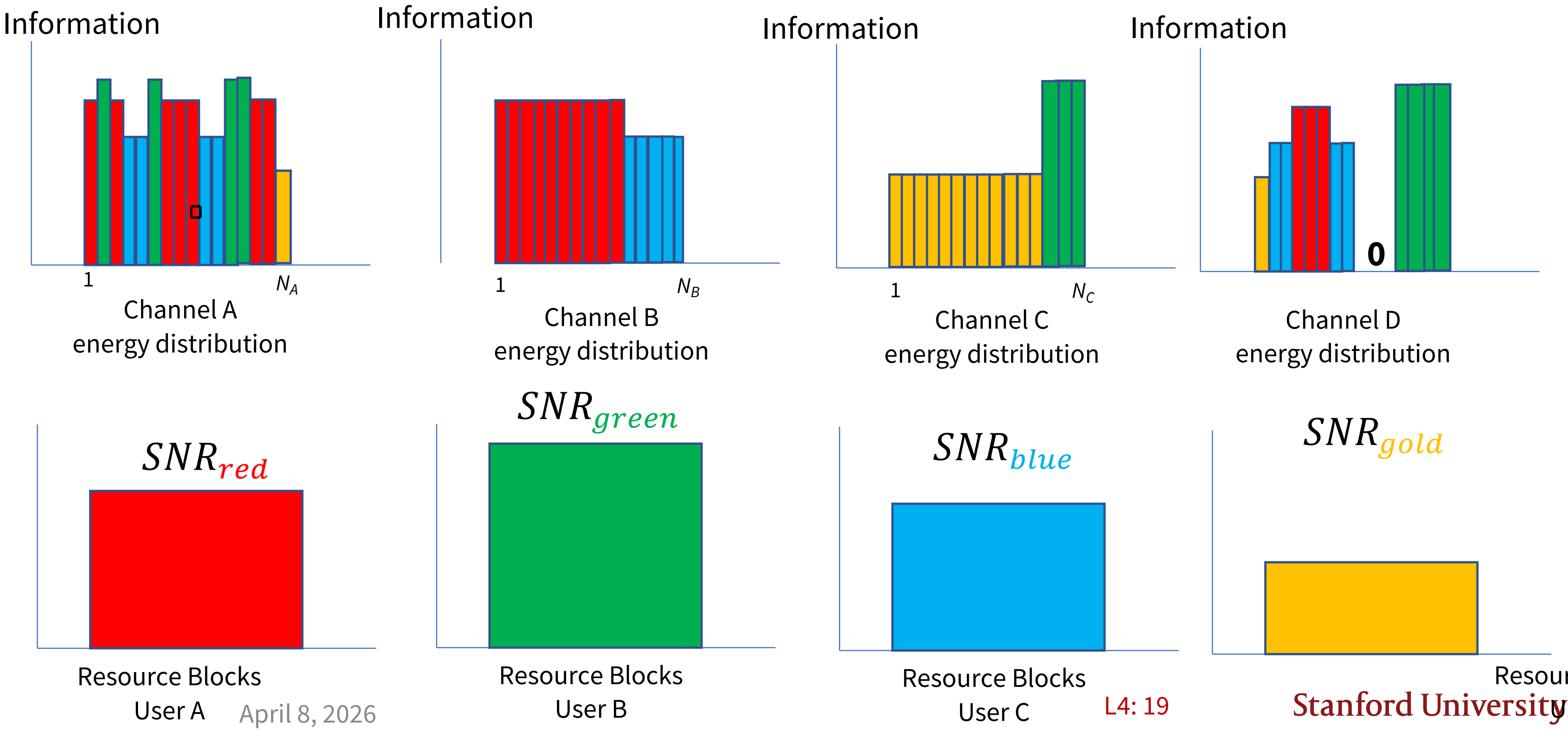
- FEQ provides LL/soft information to the decoder (or directly consequent LLRs if |C|> 1 bit/dim).



- This works if code is good $\Gamma \rightarrow 0$ 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 $\hat{g}_n < \gamma_0^*$.
- **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.

Crude "multi-user" 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)
  - Or if continuous resource blocks, then use geometric average for each user.



Ergodic C-OFDM Loading

Subsection 4.4.2

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

Fast Fading – too fast to send r and $|C|$

- The receiver can compute SNR_{geo} but the variation occurs too fast to tell transmitter

- This requires a statistical approach to a presumed stationary distribution on the time-variation's parameter(s).
  - Elegant theory, but may be hard to use in practice.

- Requires "channel prediction"
  - Constant velocity on a known path can allow prediction of the channel gain's changes
  - E.g. satellite or other aerial vehicle
  - Or if velocity and location updated frequently enough that the prediction can work.

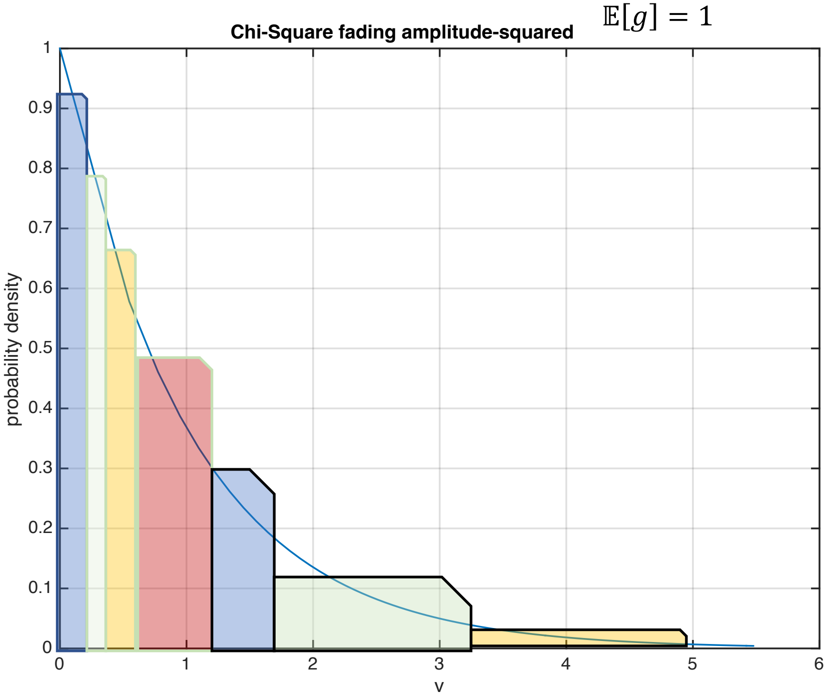
- Perhaps an idea well before its time …

Average and Outage Capacities

- **Average Capacity** is $\langle \tilde{\mathcal{C}} \rangle = \sum_g p_g \cdot \log_2\left(1 + \mathcal{E}_{x,g} \cdot g\right)$ bits/complex-subsymbol.
  - $\langle \tilde{\mathcal{C}} \rangle$ depends on energy distribution, which would be ergodic water-fill for maximum value.

- **Outage Capacity** is $\tilde{\mathcal{C}}_{out} \triangleq (1 - P_{out}) \cdot \langle \tilde{\mathcal{C}} \rangle$ bits/complex-subsymbol.
  - Basically. capacity reduces by data that would have been transmitted during outage
  - This may need retransmission, so would then be lower yet by average number of retransmissions + 1.

Approximate the distribution as discrete



$$p_g(v) \cdot dv \rightarrow p_{g,n}$$

May care about lower ranges more and so more finely divide distribution there into samples

- This simplifies calculations for loading, and the Rayleigh model was only approximate anyway.

Discrete Distributions

- All distributions (Rayleigh, Ricean, Log-Normal, etc) are gross approximations in wireless.
- Approximate by discrete distributions (which can be learned): $p_g(v) \cdot dv \to p_{g,n}$
- May need to renormalize so probabilities sum to 1 $p_{g,n} = \frac{p_g(n \cdot \Delta) \cdot \Delta}{\sum_n p_g(n \cdot \Delta) \cdot \Delta}$

- The probability $p_{g,n}$ represents the fraction of time that $g_n - \frac{\Delta_n}{2} < g \leq g_n + \frac{\Delta_n}{2}$.
  - Intervals Δ_n do not need to be the same size (may want to align with constellation-size choices).

- Simplify so that g is discrete set of center values $p_g\ (n\Delta)$ and discrete index $g \in \mathcal{G}$, size $|\mathcal{G}|$.
- g takes place of dimension index, almost (weight not 1 as with integer index).

Ergodic rate-adaptive MCS

Definition 4.4.5 **[Ergodic Rate-Adaptive Coded MT Loading with constant energy]** *Ergodic Rate-adaptive coded MT loading* **with constant energy** *solves*

$$\textbf{objective:} \quad \max_{r,|C|,g_{out}} \quad r \cdot \log_2 |C| \qquad (4.155)$$

$$(4.156)$$

$$\textbf{subject to:} \quad \langle \bar{P}_e \rangle = \sum_{g > g_{out}} p_g \cdot \bar{N}_e \cdot Q\left[\sqrt{\frac{3 \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g \cdot d_{free}(r)}{|C|-1}}\right] \leq \bar{P}_e \qquad (4.157)$$

$$r \leq 1 - \sum_{g \leq g_{out}} p_g , \qquad (4.158)$$

where the algorithm selects the code rate $0 < r \leq 1$ from among the allowed code rates, and $|C|$ is the selected (usually square) QAM constellation size The outage threshold g_{out} characterizes the two sums that are computed for each candidate ordered pair of $[r, |C|]$. The fraction $\frac{3}{|C|-1}$ can be adjusted to κ with non-square constellations, but the concept is the same.

Q-func is correct with quasi-static assumption.

Also BICM causes the $d_{free}(r)$ example later.

Definition 4.4.3 **[Geometric-Average Channel-Gain]** *The* **geometric-average channel gain** *is*

$$\gamma^*_{geo} \triangleq \prod_{g \in \mathcal{G}^*} (g)^{\left[\frac{p_g}{\sum_{g \in \mathcal{G}^*} p_g}\right]} . \qquad (4.143)$$

- Finds the cut-off channel gain g_0 that maximizes data rate ($\tilde{b} = r \cdot |\mathcal{C}|$).
  - The $[r, |C|]$ choice is over allowed constellations and codes with ($d_{free}(r)$), for the average error probability.
  - Assumes $r \leq 1 - P_{out}$. So applied code with rate r can correct the subsymbols lost to outage (Reed Solomon can at least do this, soft-decoded codes may be yet better).

Average bit rate and ergodic capacity ($\Gamma = 0$ dB)

- Average bit rate is: $$\langle b \rangle = \sum_{g \in \mathcal{G}} p_g \cdot \log_2\left(1 + \frac{\mathcal{E}_{x,g} \cdot g}{\Gamma}\right)$$

Ave Mutual Info has Γ=0 dB.

- Average energy is: $$\mathcal{E}_x = \sum_{g \in \mathcal{G}} \mathcal{E}_{x,g} \cdot p_g$$

g takes n's place in DMT RA.

See PS3.2 (Prob 4.14)

- Rate Adaptive and Margin-Adaptive Water-fill have $\mathcal{E}_g = K - \Gamma / g$.

$$\mathcal{E}_{x,g} = K_{ra} - \frac{\Gamma}{g} \;,$$

$$\begin{aligned}
\mathcal{E}_{\boldsymbol{x}} &= \sum_{g \in \mathcal{G}^*} p_g \cdot \left(K_{ra} - \frac{\Gamma}{g}\right) \\
&= K_{ra} \cdot \sum_{g \in \mathcal{G}^*} p_g - \sum_{g \in \mathcal{G}^*} p_g \cdot \frac{\Gamma}{g} \\
K_{ra} &= \frac{\mathcal{E}_{\boldsymbol{x}} + \Gamma \cdot \sum_{g \in \mathcal{G}^*} \frac{p_g}{g}}{\sum_{g \in \mathcal{G}^*} p_g}
\end{aligned}$$

RA

$$g_{geo}(*) \triangleq \prod_{g \in \mathcal{G}^*} (g)^{\left[\frac{p_g}{\sum_{g \in \mathcal{G}^*} p_g}\right]}$$

$$\begin{aligned}
< b > &= \log_2 \prod_{g \in \mathcal{G}^*} \left(1 + \frac{\mathcal{E}_{x,g} \cdot g}{\Gamma}\right)^{p_g} \\
&= \log_2 \prod_{g \in \mathcal{G}^*} \left(\frac{K_{ma} \cdot g}{\Gamma}\right)^{p_g} \\
2^{<b>} &= \left(\frac{K_{ma}}{\Gamma}\right)^{\sum_{g \in \mathcal{G}^*} p_g} \cdot \prod_{g \in \mathcal{G}^*} g^{p_g} \\
K_{ma} &= \Gamma \cdot \left(\frac{2^{<b>}}{\prod_{g \in \mathcal{G}^*} g^{p_g}}\right)^{\frac{1}{\sum_{g \in \mathcal{G}^*} p_g}} \\
&= \Gamma \cdot \frac{\left(2^{<b>}\right)^{\left[\frac{1}{\sum_{g \in \mathcal{G}^*} p_g}\right]}}{g^*_{geo}}
\end{aligned}$$

MA

Ergodic Water-filling - Goldsmith

- The amount of energy is either zero or a value given by the water-fill equation for $g \in \mathcal{G}^*$.

$$\mathcal{E}_{x,g} = \begin{cases} \Gamma \cdot \frac{2^{<b>}}{g_{geo}} - \frac{\Gamma}{g} & g > \frac{\Gamma}{K_{ma}} \\ 0 & g \leq \frac{\Gamma}{K_{ma}} \end{cases}$$

$$\text{transmit if } g > g_0 = \frac{\Gamma}{K_{ma}} .$$

- Direct water-fill calc from time-domain g values is non-causal; can't really be made causal with delay.
- Instead, the stationary statistics have been exploited above.
  - Those statistics (p_g) may need to be estimated by counting g values though over time.
  - If the system is ergodic (predictable path/channel), this will get better and better.
- Instantaneous transmit energy often has limit (less than the $\mathcal{E}_{x,g}$ value above) for small g.

If so, reduce $\langle b \rangle$ -- this is a kind of margin also

This is not practical – use erasures instead, but provides bound.

Quasi-Static Statistical Loading

Subsection 4.4.2.3

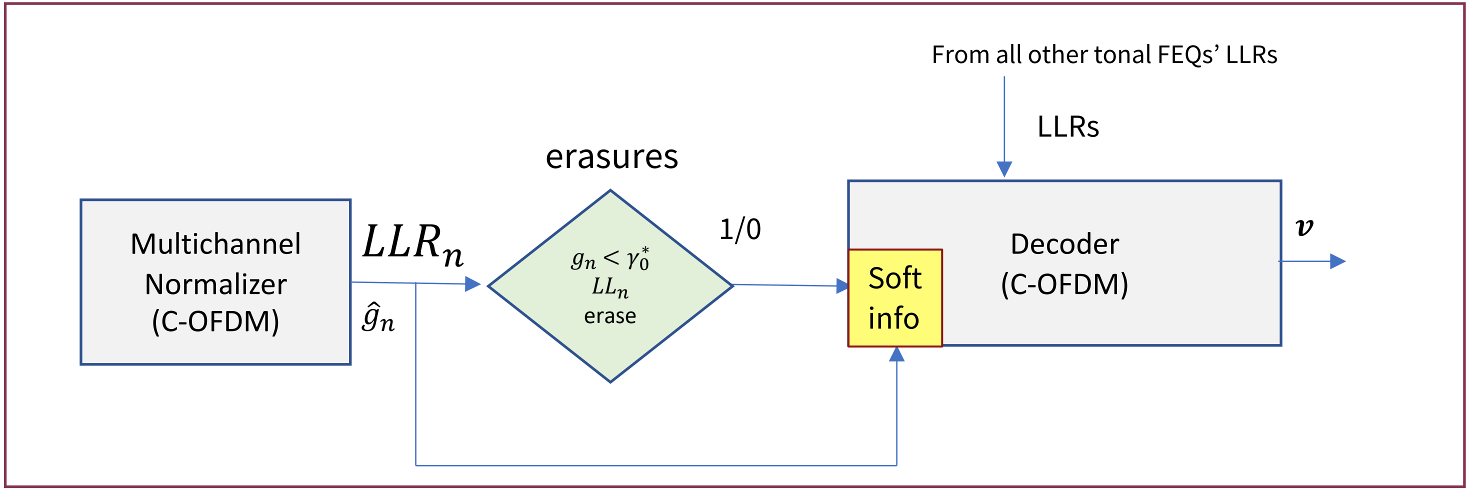
PS3.1 (Prob 4.13), PS3.3 (Prob 4.22) and PS3.4 (Prob 4.16)

r and $|C|$ sent once or infrequently

- **Statistical loading** also uses the channel distribution,
  - but is more practical for fast-fading situation.
  - Exploits Erasures and Interleaving

- Only the receiver needs to know when the channel gain is “too low.”

- The transmitter of course needs to know $[r\ , C]$, but one chance (or effectively so)



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)*

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

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

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 $\langle P_b \rangle^{\left\lceil \frac{d_{free}+1}{2} \right\rceil}$ for a BSC.

Loading on Actual Wireless Channel with Flat Energy

- Compute data rate for both average and max (over all tones) with rough gap estimate.

$$b_{flat-geo} = \frac{1}{N} \cdot \sum_{n=1}^{N} \log_2\left(1 + \frac{\bar{\mathcal{E}}_x \cdot g_n}{\Gamma}\right) \qquad b_{max} = \log_2 (\text{max constellation size}) \leq \max_n \left\{\log_2\left(1 + \frac{\bar{\mathcal{E}}_x \cdot g_n}{\Gamma}\right)\right\}$$

- Test allowed code rates r, which then → d_{free} e.g.

$$r = \frac{b_{flat-geo}}{b_{max}} \leq 1$$

- Estimate "correctable outage probability" for each *r* as $\bar{P}_{out} = 1 - r$.

- Estimate p_g by **binning** or counting of *g* values in 6 dB – *10 log10 ($d_{free\text{-}new}/d_{free\text{-}old}$).*
  - In this range, the *b* value does not change the constellation-size choice.
  - Binning counts FEQ noise-estimate values in each range (current symbol or over many symbols).

- Solve (L4:30) for g_0 -- tones with $g<g_0$ will be "erased."

- Demod indicates "erasure" (delete from sum is one way) to ML detector.
  - This could be an erasure in Reed-Solomon.
  - It can also be zero LLR in iterative decoder for bits corresponding to that tone.

$$\bar{P}_{out} = \sum_{g<g_0} p_g$$

Wireless Example 64-state CC with BICM

- **Extension** of well-known code as example (64-state rate-1/2 code with puncturing) has:
  - >> r = 0.9 0.8 0.75 0.67 0.5 0.25 0.2
  - >> dfree = 2 4 5 6 10 20 25
- Given (measured) channel-gain distribution
  - >> g =[0.2000 2.0000 20.0000 40.0000 80.0000 160.0000 320.0000 640.0000];
  - >> pg=[.11 .1 .03 .05 .35 .2 .1 .06];

```
>> prob2 = kron(ones(7,1),pg);
```

```
Pout=cumsum(prob2(1,1:8)) % =
    0.1100    0.2100    0.2400    0.2900    0.6400    0.8400    0.9400    1.0000

>> ones(1,8)-Pout =
    0.8900    0.7900    0.7600    0.7100    0.3600    0.1600    0.0600         0

>> r =
    0.9000    0.8000    0.7500    0.6700    0.5000    0.2500    0.2000
```

Not correctable | correctable | Not correctable

Wireless Example continued for <Pe>

- SNR component (with $\tilde{\mathcal{E}}_x = 1$) is

```
>> SNR=kron(dfree',g);
>> 10*log10(SNR) =  (in dB)
 -3.9794   6.0206  16.0206  19.0309  22.0412  25.0515  28.0618  31.0721
 -0.9691   9.0309  19.0309  22.0412  25.0515  28.0618  31.0721  34.0824
       0  10.0000  20.0000  23.0103  26.0206  29.0309  32.0412  35.0515
  0.7918  10.7918  20.7918  23.8021  26.8124  29.8227  32.8330  35.8433
  3.0103  13.0103  23.0103  26.0206  29.0309  32.0412  35.0515  38.0618
  6.0206  16.0206  26.0206  29.0309  32.0412  35.0515  38.0618  41.0721
  6.9897  16.9897  26.9897  30.0000  33.0103  36.0206  39.0309  42.0412
```

$$\tilde{\mathcal{E}}_x \cdot d_{free} \cdot g$$

Only rows 3-4 are eligible.

- Compute $\langle P_e \rangle$ for several SNRs and SQ QAM Constellations
  - 4QAM -- >> prob1=qfunc(sqrt(SNR(1:4,1:4)))
  - 16 QAM -- >>prob1=2*qfunc(sqrt((3/15)*SNR(1:4,1:4)))

Compute <Pe> tables

4SQ QAM

```
C=4;
SNR4=zeros(7,8);
bicmscale=(3/(C-1))*[ 2 2 2 2 2 4 8];
SNR4(3,:)=bicmscale(3)*r(3)*SNR(3,:);
SNR4(4,:)=bicmscale(4)*r(4)*SNR(4,:);
>> prob1=qfunc(sqrt(SNR4(3:4,:)))
    0.2113   0.0056   0.0000   0.0000   0.0000   0.0000   0.0000        0
    0.2032   0.0043   0.0000   0.0000   0.0000   0.0000   0.0000        0
>> avePe100=100*cumsum((prob2(3:4,1:8).*prob1)','reverse')'
    2.3809   0.0561   0.0000   0.0000   0.0000   0.0000   0.0000        0
    2.2788   0.0433   0.0000   0.0000   0.0000   0.0000   0.0000        0
```

$r = 3/4$

$r = 2/3$

r

.75

.67

$\bar{b} = \frac{3}{4} \cdot 1 = .75$

16SQ QAM

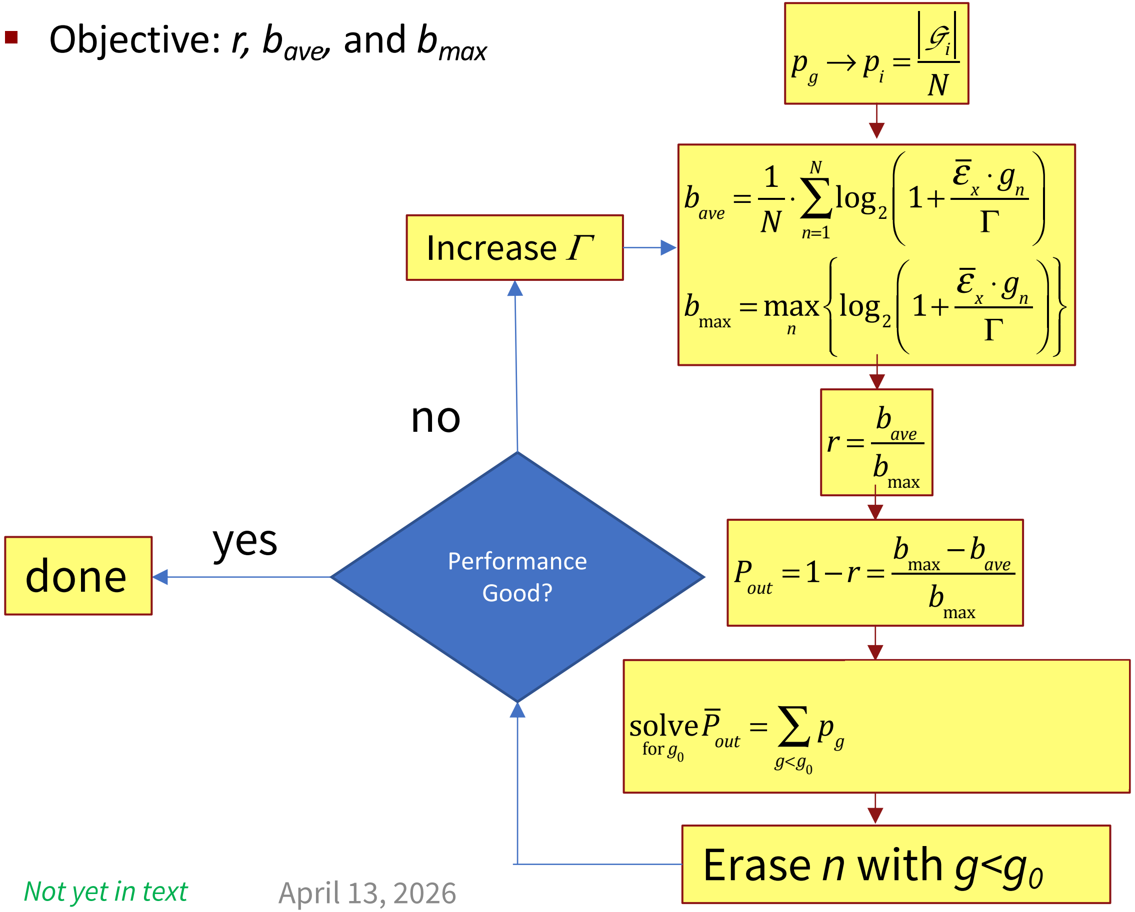
```
C=16;
SNR16=zeros(7,8);
bicmscale=(3/(2*C-1))*[ 2 2 2 2 2 4 8];
SNR16(3,:)=bicmscale(3)*r(3)*SNR(3,:);
SNR16(4,:)=bicmscale(4)*r(4)*SNR(4,:);
>> prob1=qfunc(sqrt(SNR16(3:4,:))) % =
   0.3516   0.1141   0.0001   0.0000   0.0000   0.0000   0.0000   0.0000
   0.3466   0.1061   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000

>> avePe100 = 100*cumsum((prob2(3:4,1:8).*prob1)','reverse')'% =
   5.0092   1.1416   0.0002   0.0000   0.0000   0.0000   0.0000   0.0000
   4.8740   1.0613   0.0001   0.0000   0.0000   0.0000   0.0000   0.0000
```

$\bar{b} = \frac{3}{4} \cdot 2 =$ 1.5 bits/dim

- 64 QAM doesn't work, so 16SQ with r=3/4 is best design, see Section 4.4.2.3.

Flow Chart for Quasi-Static Statiscal Loading

- Objective: r, b_{ave}, and b_{max}



Calculate Distribution

Compute averate and max bits/tone

For the used code find r and d_{free}

Code should be able to correct for P_{out} % erased

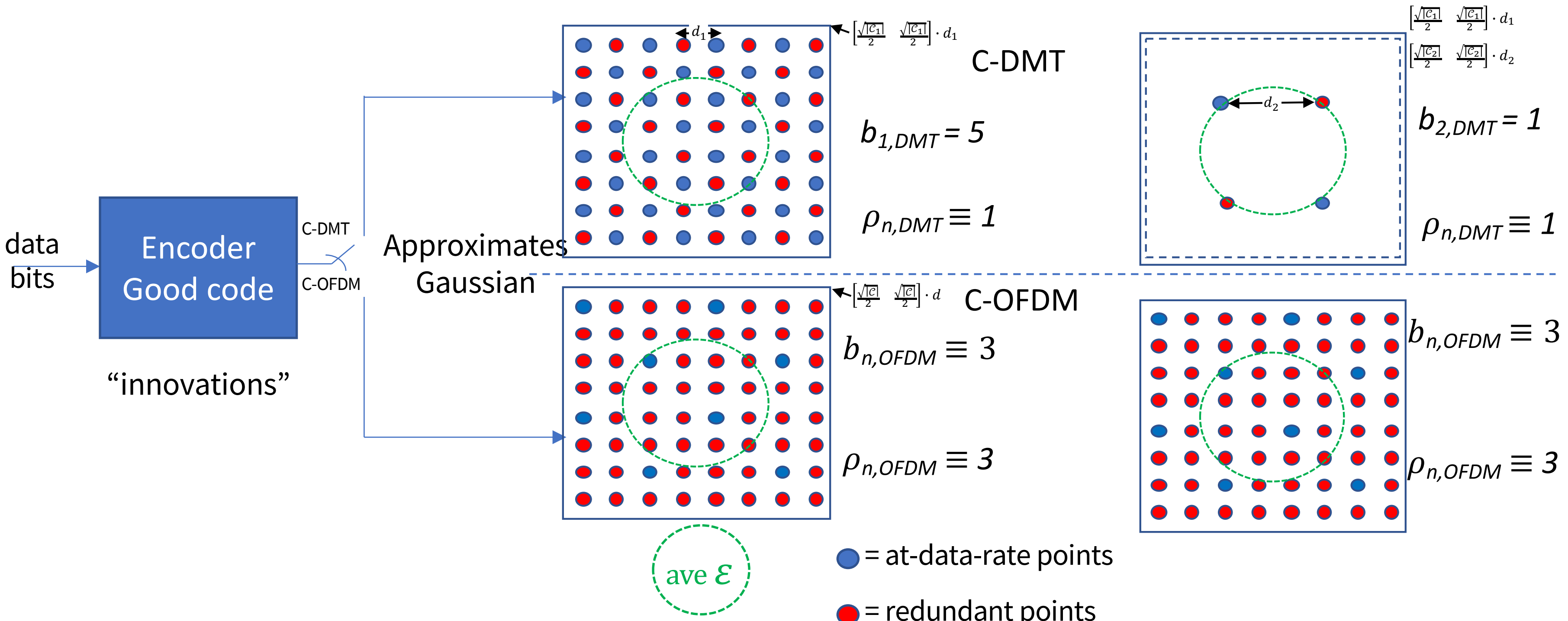
Need to approximate Q function or store it for each i value

untested by instructor

STANFORD

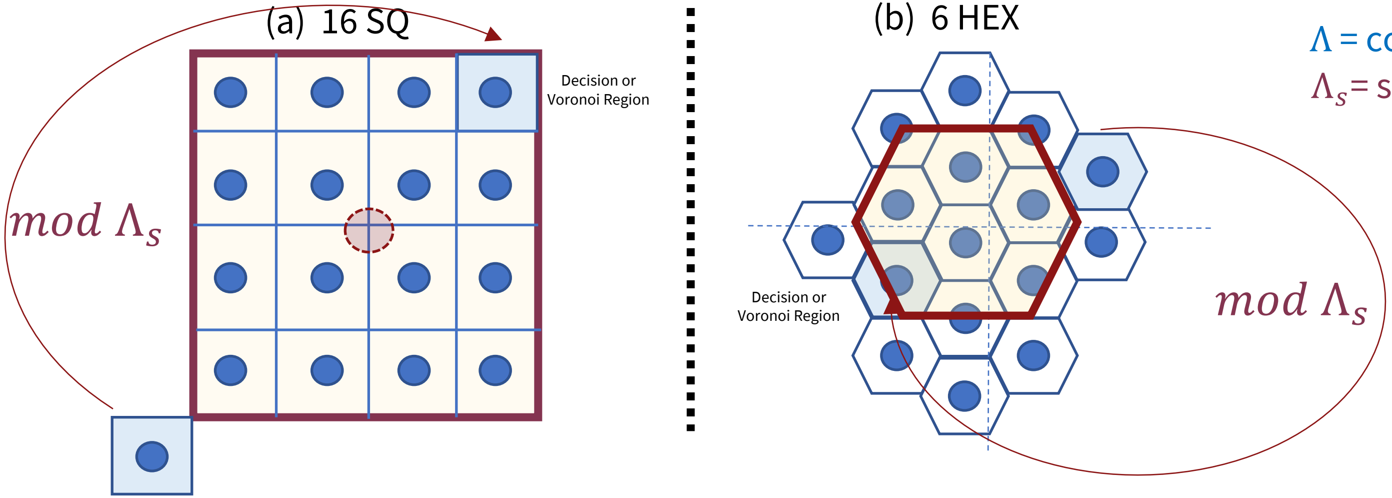
End Lecture 4

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.

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 | **shaping gain**

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

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

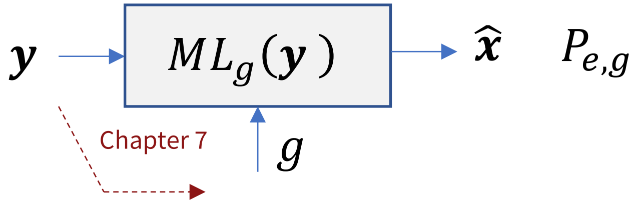
The random-gain AWGN

- 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