Trellis-Coded Modulation [TCM]

- Limitations of conventional block and convolutional codes on bandlimited channels
- Basic principles of trellis coding: state, trellis, and set partitioning
- Coding gain with trellis codes
- System mechanization: application of the Viterbi Algorithm for decoding
- Systems issues
- Applications to wired and wireless channels
- Advanced concepts: multi-dimensional trellis codes
- Three aspects of TCM [or any code]
  - **Design the code**
  - **Determine the performance of the code (coding gain)**
  - **Mechanize the encoder and decoder (Viterbi Algorithm**

Classic Coding

- Information theory tells us that for optimal communications we should design long sequences of signals, with maximum separation among them; and at the receiver we should perform decision making over such long signals rather than individual bits or symbols.
- If this process is done properly, then the message error probability will decrease exponentially with sequence length, n

$$P_e < Ke^{-(R_0 - R)n}$$

  provided that the rate R is less than R_0, which in turn is less than the Shannon Capacity.
- This is the idea behind coding. In conventional coding, the **coding is separate from modulation**. Coding occurs at the digital level, before modulation and generally involves adding bits to the input sequence. The resultant redundancy requires added bandwidth.
- At the receiver, **hard** decoding occurs after demodulation. The decoding operation is based on hard decisions, since a digital bit (or symbol) stream fees the decoder and is either in error or not. Decoding can also be done based on the analog received samples, and this is called **soft** decoding. The theoretical loss due to hard [vs soft] decoding leads to a ~2dB performance loss.

Coded Modulation

- Optimum 2-D modulation uses dependency between in-phase and quadrature symbols.
- 4-D modulation introduces dependency between symbols of two successive intervals.
- Trellis coding introduces dependency between every successive symbol.
- Trellis and multi-dimensional codes are designed to maximize the Euclidean distance between possible sequences of transmitted symbols
- Distance between the closest possible [ie, minimum distance] **sequences** of transmitted symbols in signal space determines the performance

$$P_e \sim e^{-d_{\min}^2/2s^2},$$

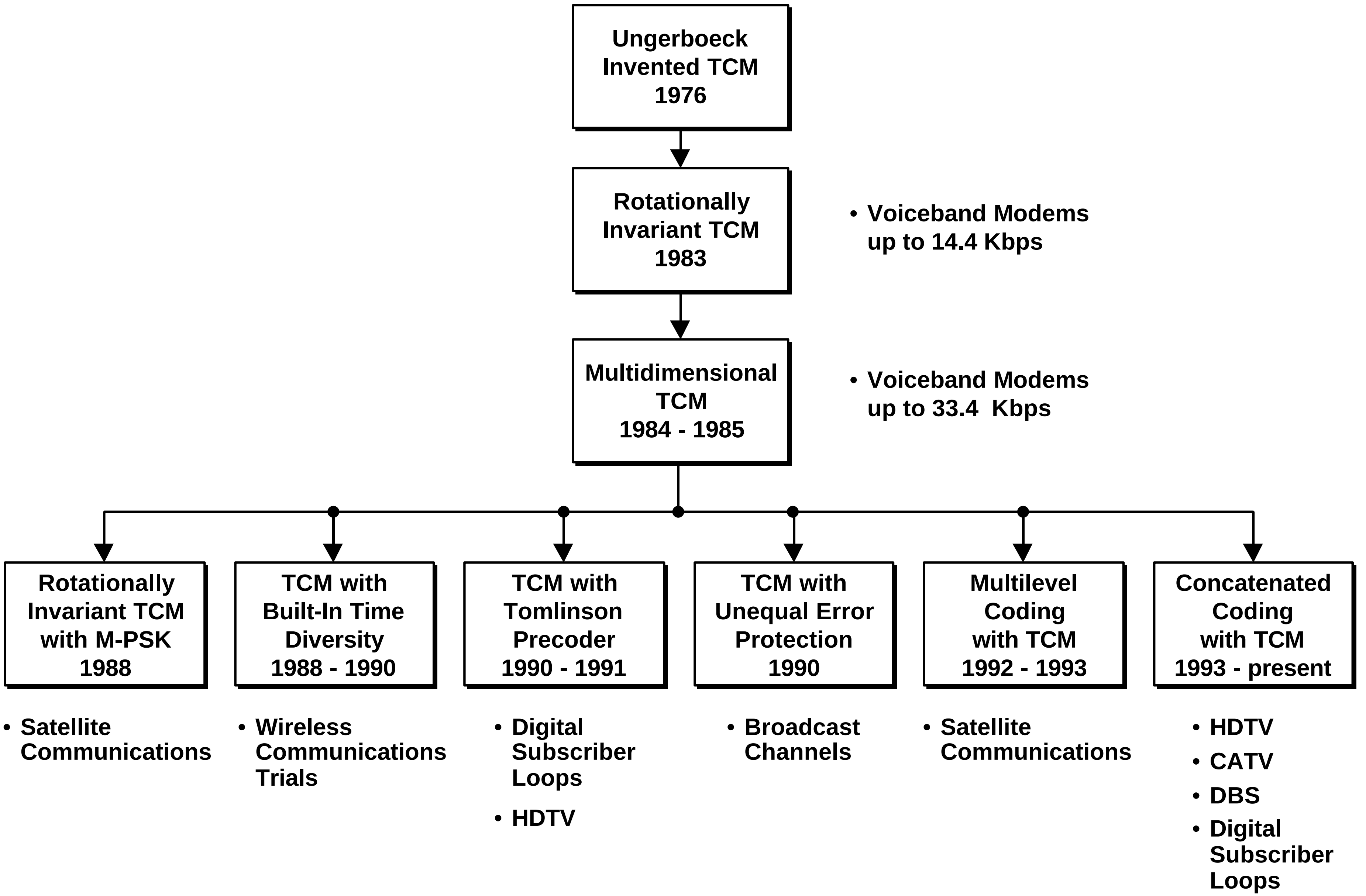
where $d_{\min}^2$ is the minimum distance between signal sequences

and s^2 is the noise power

Trellis Coded Modulation

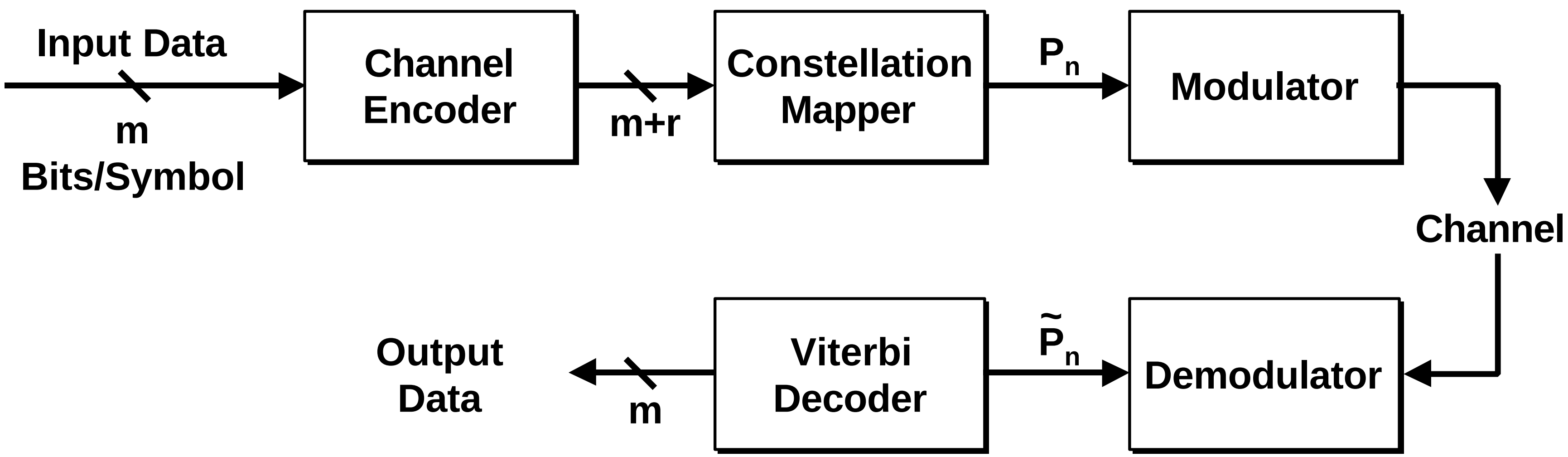
- The key idea is that the operations of [baseband] modulation and coding are combined.
- The **bandwidth is not expanded:** same symbol rate, but redundancy is introduced by using a constellation with more points than would be required without coding.
  - Typically, the number of points is doubled
  - The symbol rate is unchanged
  - The power spectrum is unchanged
- Since there are more possible points per symbol, it may appear that the error probability would increase for a given S/N.
- As in conventional coding, dependencies are introduced among different symbols ---only certain sequences of successive constellation points are allowed.
- By properly making use of these constraints during reception, the error probability actually decreases.
- A measure of performance improvement is the **coding gain**, which is the difference is S/N between a coded and uncoded system of the same information rate that produces the same error probability.

History of Trellis-Coded Modulation



Key Elements of Trellis Coded Modulation

- To improve power efficiency, use channel coding to introduce memory into $\{P_n\}$
- To compensate for the redundant bits introduced by the channel encoder, use a larger constellation with more than 2^m symbols;
- To reduce the decoder complexity, design the channel encoder and constellation mapper jointly.

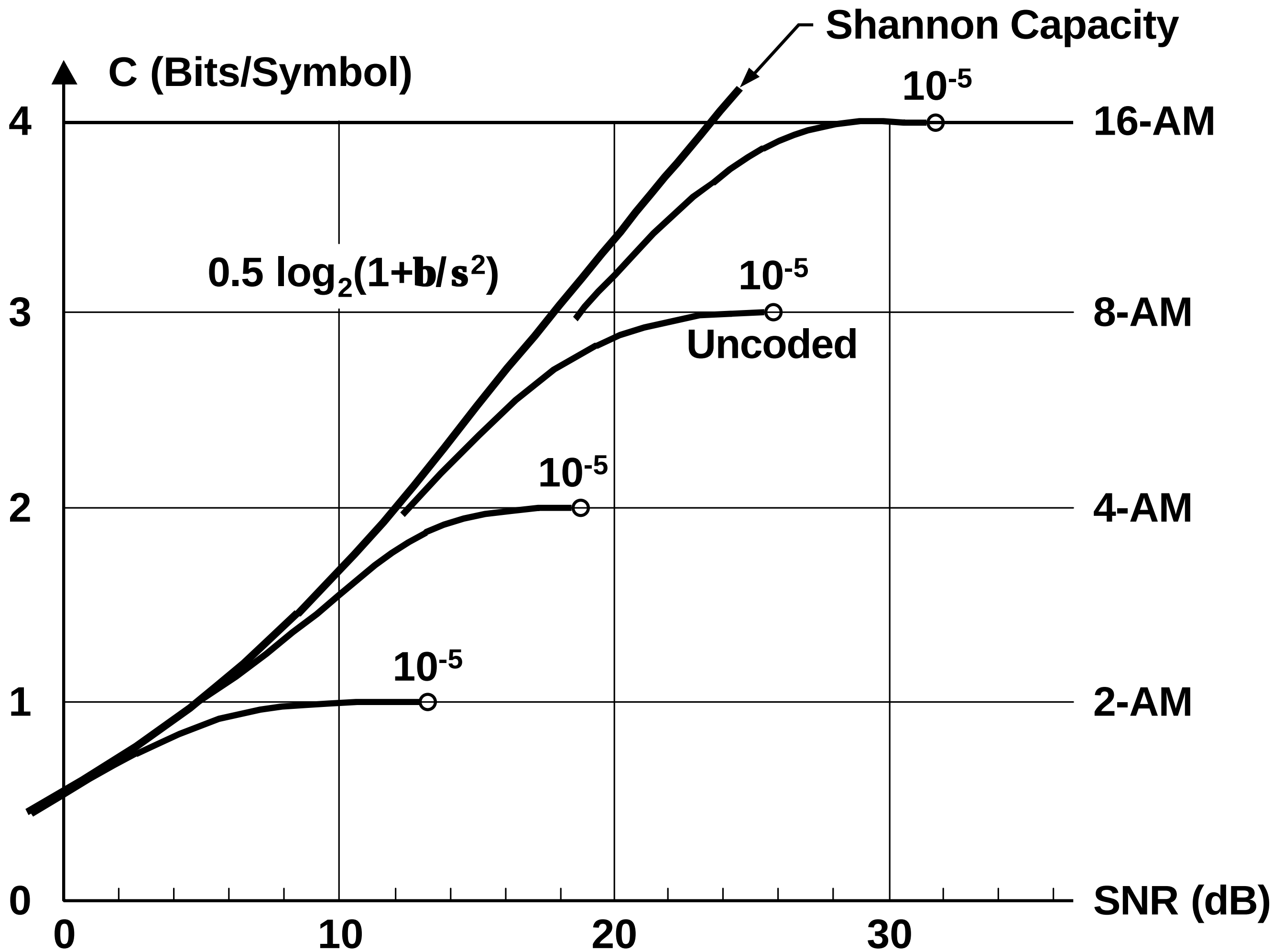


Trellis Coded Modulation

Trellis Coding---the basics

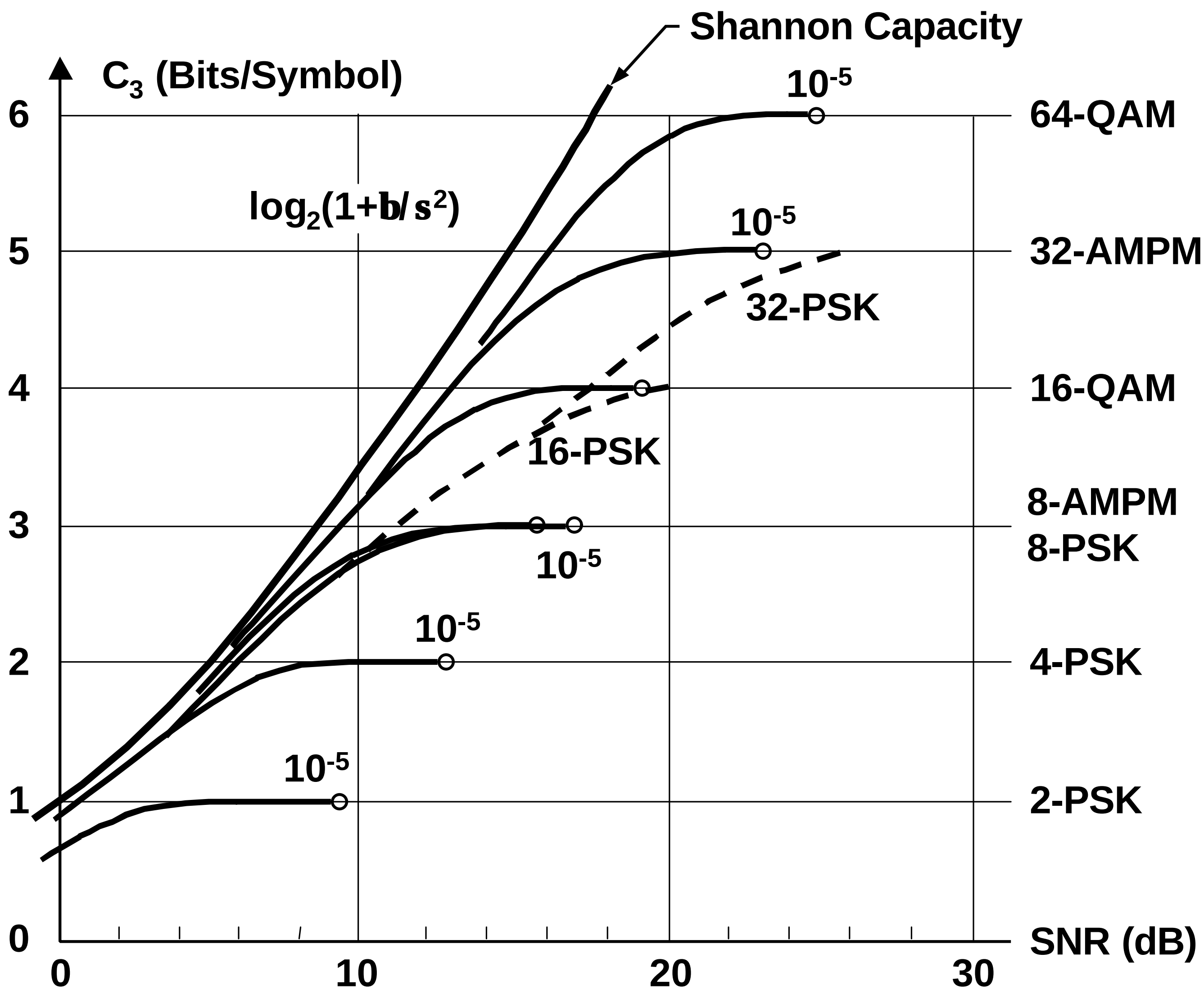
- It can be shown for the Gaussian channel that there is an input discrete alphabet that has capacity very close to the capacity with continuous inputs
- As shown on the next chart, an **eight- level** system can achieve a capacity of **2 bits/symbol**
- This suggests that it is only necessary to **double** the signal constellation to get good coding gains (increasing the signal alphabet will not improve the coding gain)
- Note that at about 19 dB we can achieve 10^{-5} with a four-level constellation
- With coding, using an 8-level constellation we can theoretically transmit 2 bits/symbol error free down to about 13dB
- Hence using coded modulation we could gain as much as 19-13 = 6 dB
- The bandwidth has not been expanded (same symbol rate)

Trellis Coding – The Basics



The Information Conveyed by a Real-Valued Discrete time Channel with Additive Gaussian Noise

Trellis Coding – The Basics (QAM modulation)



Trellis Coding ---QAM modulation

- Given a channel with a bandwidth limitation, first determine the maximum symbol rate that can be transmitted.
- Determine the size of the alphabet, 2^L , that is needed to produce the desired bit rate.
- Double the size of the constellation and introduce a channel coder that produces one extra bit
- The coder need not code all the incoming bits
- There are many ways to map the coded bits into symbols. The choice of mapping will drastically affect the performance of the code.
- Ungerboeck produced a good heuristic technique called ***mapping by set partitioning***
  - The encoding philosophy is to first partition the larger 2^{L+1} constellation into smaller subsets
  - The Euclidean distance between sequences of signal points in different subsets is substantially increased (and may be on the order of the distance between points in the same subset)
  - Performance will be determined by the distance between sequences in different subsets.
- Trellis coding produces a dramatic increase in the Euclidean (free) distance between **sequences** of signal points and the Viterbi Algorithm is used to detect the signal
- Results also hold for 2-dimensional modulation

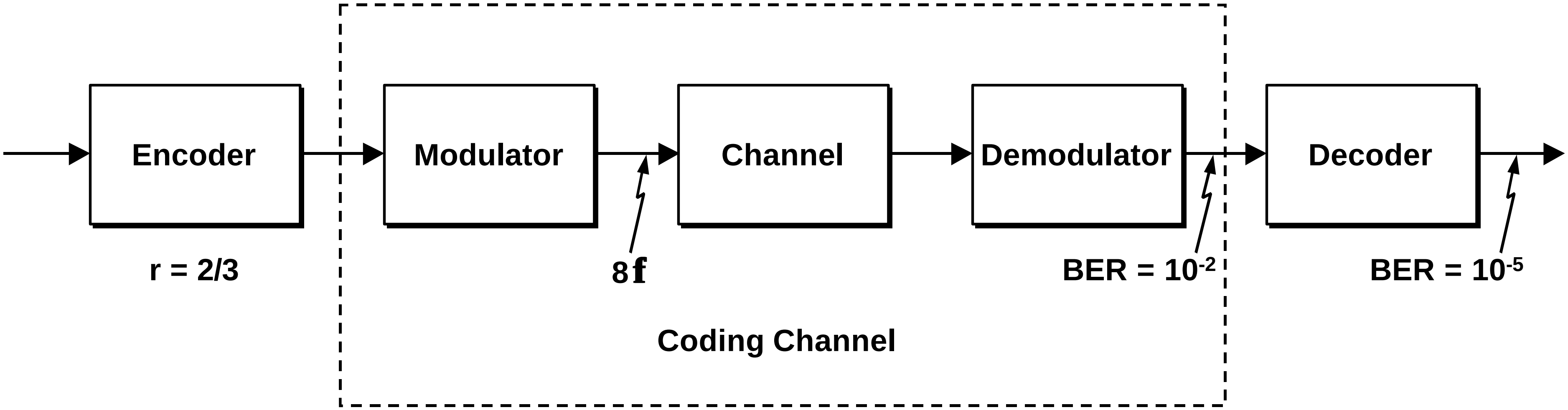
Trellis Coding Summary

- Ungerboeck (1976) showed that for bandlimited channels substantial coding gains could be achieved by **convolutional coding of signal levels** (rather than coding of binary source levels)]
- He joined modulation and coding to increase the Euclidean **distance between signal sequences**
- Called channel “trellis coding” because the sequence of states in the finite state machine which encodes the signal levels follows a trajectory in a trellis of possible trajectories
- The larger Euclidean distance between signal sequences, the lower the error rate, which for moderate to large SNR is

$P_e = N_{free} Q(d_{free}/2s)$, where N_{free} = the average number of sequences at d_{free}, and $Q(x) \approx e^{-x^2/2}$

- Codewords consist of modulated level sequences. Trellis coding uses dense signal sets but restricts the sequences that can be used. This provides a gain in minimun distance and the code imposes a time dependency on the allowed signal sequences that allows the receiver to ride through “noise bursts” as it is estimating the transmitted sequence.
- Since the pulse shape and symbol rate are unchanged ---> no bandwidth expansion

Trellis Coded Modulation: Example System



Uncoded System: 4ϕ PSK, BER = 10^{-5}, Encoder & Mod are Independent

Coded System: r = 2/3, 8ϕ PSK, Demod BER = 10^{-2}, Decoded BER ~ 10^{-5}

Limitations of Conventional Coding Techniques for Bandlimited Channels

- For the example uncoded 4-PSK and rate 2/3 coded 8-PSK system [same **customer** date rate]
  - if the uncoded system has a BER = 10^{-5}, the coded system will have an error rate at the demodulator output of worse than 10^{-2}. What sort of code is needed to make the 8-PSK system have a 10^{-5} ***decoded*** BER?
- A t-error correcting code of block length n, with k information bits must satisfy the Hamming bound [see Weldon and Peterson]--- which provides a lower bound on the code block length

$$2^{n-k} \geq \sum_{i=0}^{t} \binom{n}{i}$$

- Suppose t=2 and k/n = R= 2/3 [as per our example] we find that n > 24.
- With such a code we need a binary block length of 24 bits or eight 8-PSK symbols. Each of the symbols has an error rate of 10^{-2}. With Gray coding [ie, a symbol error --->one bit error], the code will correct two symbol errors.
- An error will be made if 3 [or more] of the 8 symbols are in error, and the decoded BER

$$\binom{8}{3} * (10^{-2})^3 \approx 5 * 10^{-5}$$

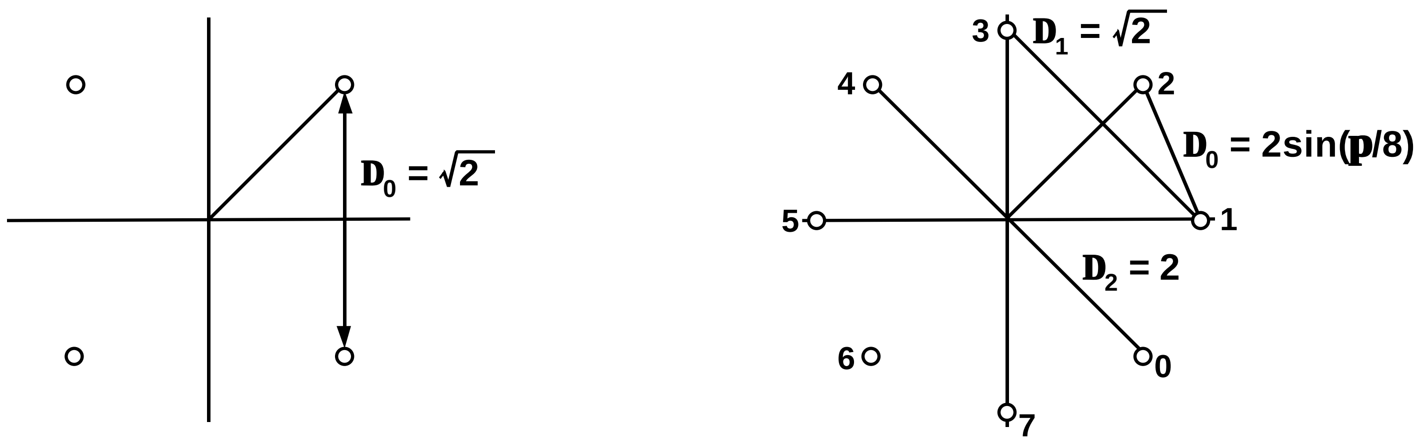
- Thus relatively complex codes [n=24] are required and this code did not provide "gain."
- For a **non-bandlimited** channel the coded system has a DEMOD error rate of 10^{-4} and the BER is

$$\binom{8}{3} * (10^{-4})^3 \approx 10^{-12}$$

So in a bandlimited channel can a simpler code can be used to achieve large coding gain?

Trellis Coded Modulation (Example continued)

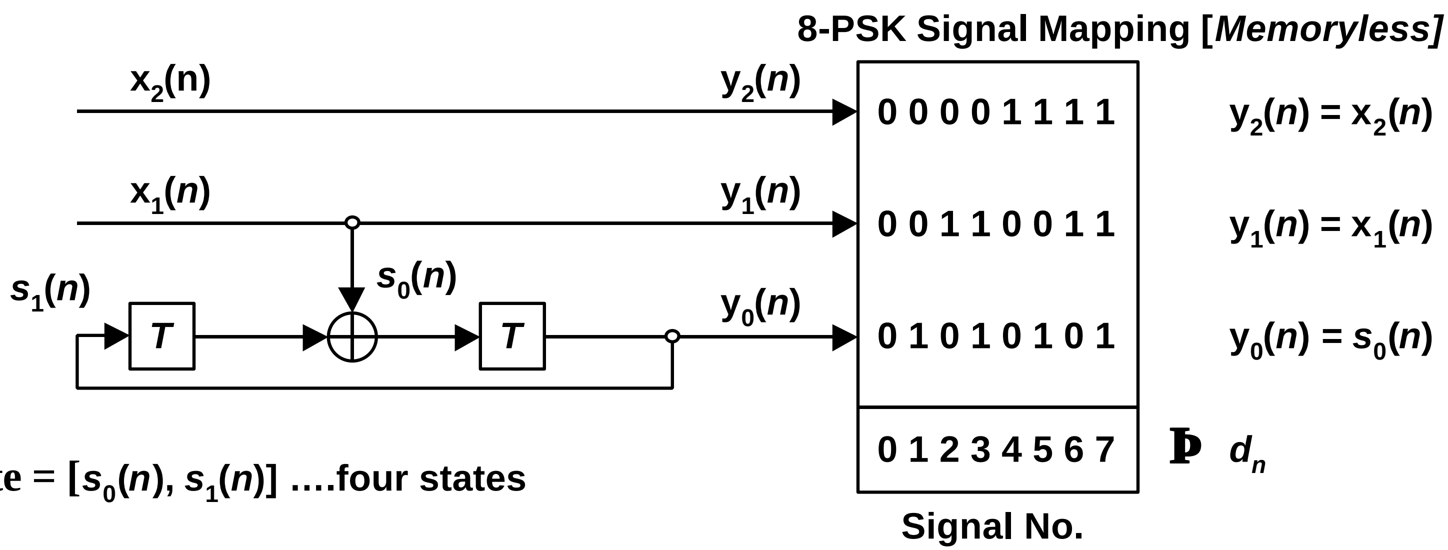
- Uncoded four phase d^2_{free} = 2. Coded d^2 = ?
- Coded system uses 8-point constellation and the signal mapping shown below
- Number of states = $2^{\#SR's}$ = 4
- Number of paths leaving node = $2^{\#\ bits/symbol\ (user)}$
- Tribit determines signal point
- Each new dibit X_1 and X_2 causes a state transition, as well as generating an output (shown are input dibit and output tribit)



Four-point (uncoded) and eight-point (coded) signal constellations.
Each has radius one, and thus an average signal power of unity

Trellis Coding: Implementation

- Even though the minimum distance between signal points in the enlarged constellation is reduced from the original constellation (for the same average power), TCM succeeds in increasing the distance between the transmitted sequences.
- Finding good TCM codes does not follow from knowing how to find good codes with good hamming distances (i.e. conventional convolutional codes)
- The mapping to analog signals is critical
- Recall that the output of the coder is determined by the new input and the current **state** and the state evolves to a (possibly) new state

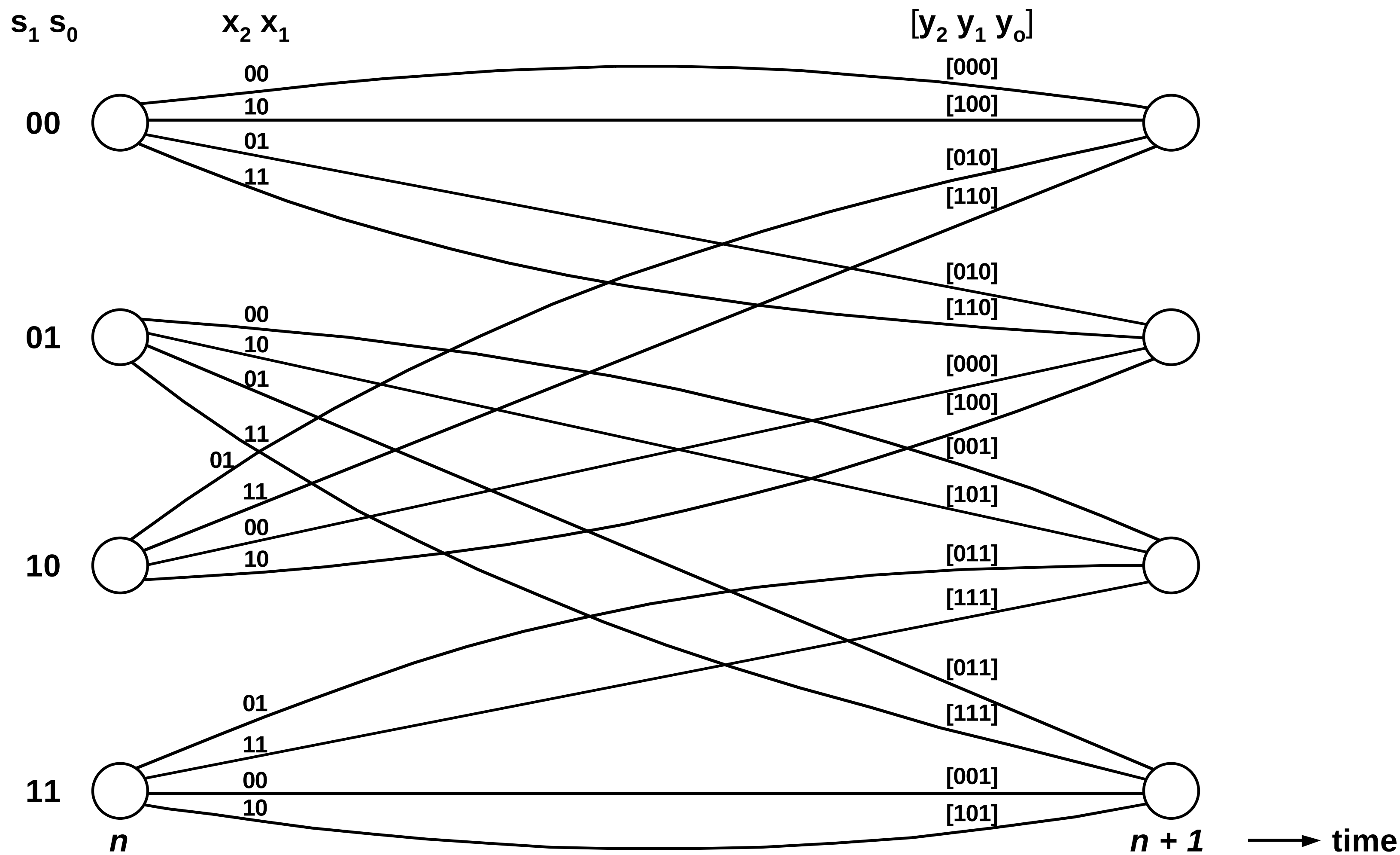


State = [$s_0(n)$, $s_1(n)$] ….four states

Convolutional encoder (rate 2/3 code) mapping a pair of data bits into a trio of bits that defines a signal point, e.g., (y_2, y_1, y_0) - (1, 0, 0) defines point 4 in previous figure.

After the states have been determined, we detect the uncoded **bits by picking the closest (in Euclidean distance)** signal point **to the received sample**

Encoder State Trellis



TCM *does not* protect parallel branches

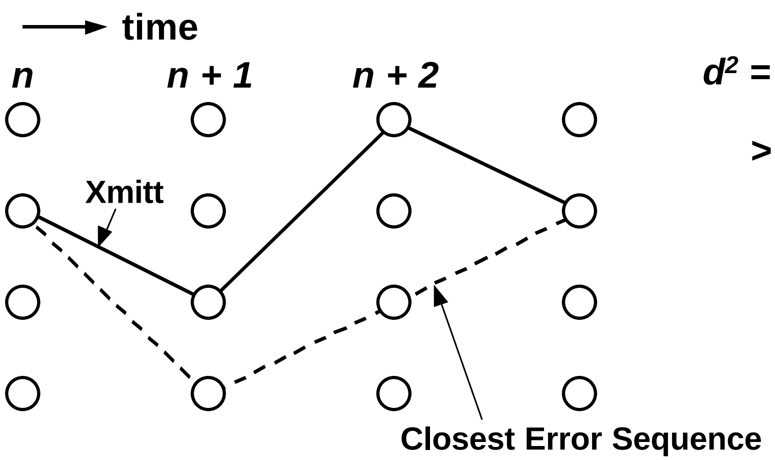
Receiver trellis strives to generate (i.e. "estimate") the transmit path through the trellis

Trellis Encoding

- The new state is a function of the current state and the input
- Transitions are NOT possible between every state
- Parallel branches, corresponding to the two paths associated with each uncoded bit, represent points at at maximum separation in Euclidean distance
- Branches emerging from and converging to a point have the next largest separation
- Each parallel pair represents a pair of points with a Euclidean Distance = 2
- At each node the four arms entering have Euclidean distance $\sqrt{2}$
- The minimum distance separation at the same node is $2\sin\frac{p}{8} = 0.38$

Trellis Coded Modulation Example

- The Viterbi decoder at the receiver “copies” the transmitter trellis and attempts to find the received path through the trellis whose distance is closest to the received signal
- The performance is determined by d^2_{free}
- Suppose two candidate paths (one being the transmitted one) diverge from one node and rejoin late. If they rejoin in one symbol they are parallel arms ($d^2 = 4$)
- If they go to different states in the first transition, they will need at least 3 symbol intervals to rejoin, and



$$d^2 = (d_{leaving\ same\ node})^2 + (d_{entering\ different\ nodes})^2 + (d_{entering\ same\ node})^2$$

$$> 2 + (2\sin\pi/8)^2 + 2 > 4$$

$$Gain = 10\log_{10}\left[\frac{d^2_{min} = (coded\ 8\text{-}\phi)}{d^2_{min} = (uncoded\ 4\text{-}\phi)}\right]$$

$$= 10\log_{10}\frac{4}{2} = 3\ dB$$

Segments of two trajectories leaving a node separately and rejoining three symbol intervals later

Three aspects of coding:

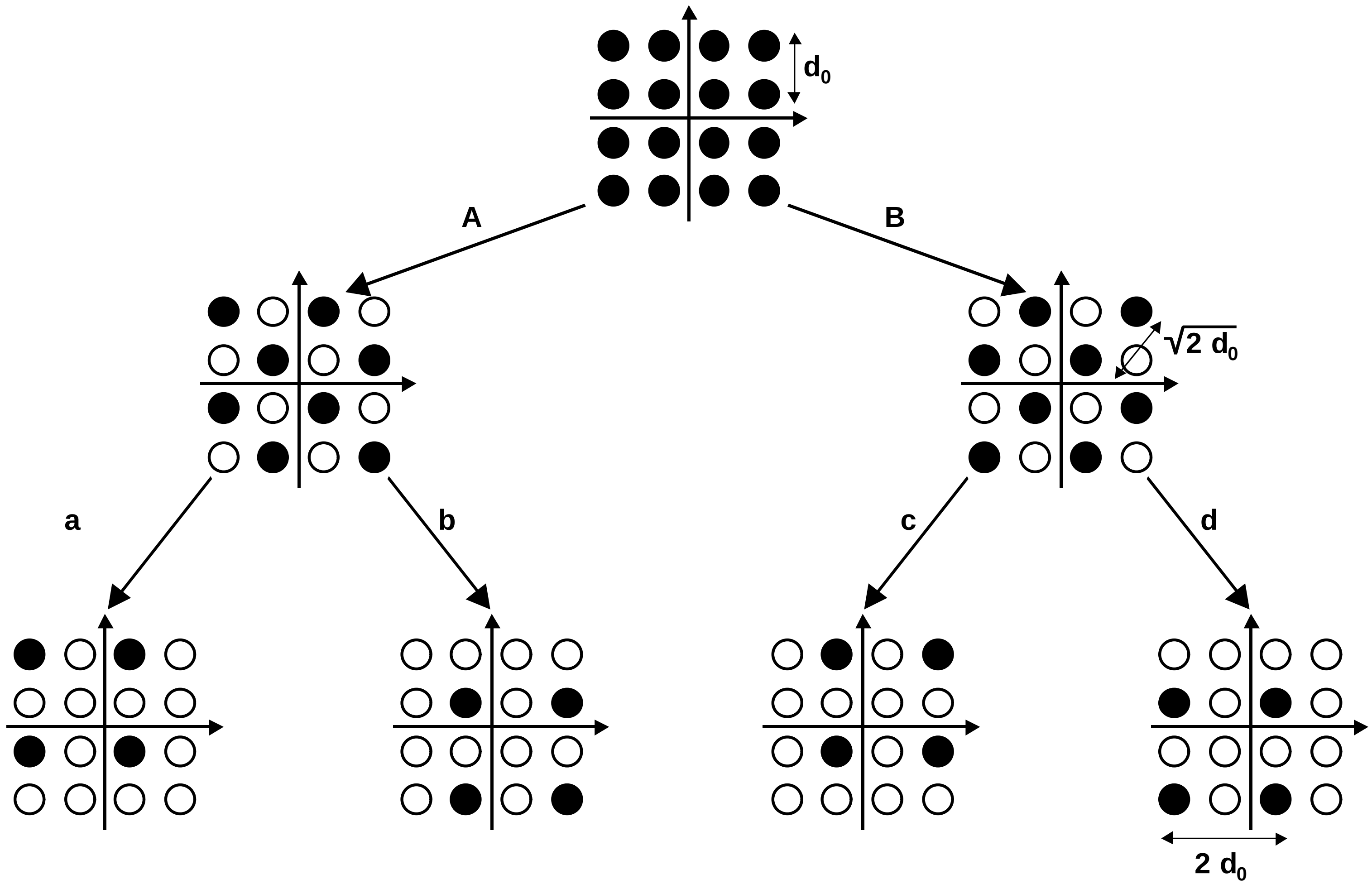
- **Design the code**
- **Determine the performance of the code (coding gain)**
- **Mechanize the encoder and decoder (Viterbi Algorithm)**

Basic Philosophy of TCM

- Form an expanded constellation, with double the number of points
- Partition that constellation into subsets. The points within each subset are far apart in Euclidean distance, and are made to correspond to the uncoded bits.
- The remaining bits determine the choice of the subset. Only certain sequences of subsets are allowed. Typically, the allowed sequences correspond to a simple convolutional code.
- In order to keep the allowed sequences far apart, choose subsets that correspond to branching in and out of each state to have maximum distance separation.
- At the receiver, we find the allowed sequence which is closest in Euclidean distance to the received sequence of signals.
- After the output sequence is decoded, the receiver decides between the uncoded points based on the Euclidean distance to the nearest signal point --- thus the distance between uncoded pairs and closest sequences should be ~ the same.

Partitioning of 16-Point Constellation

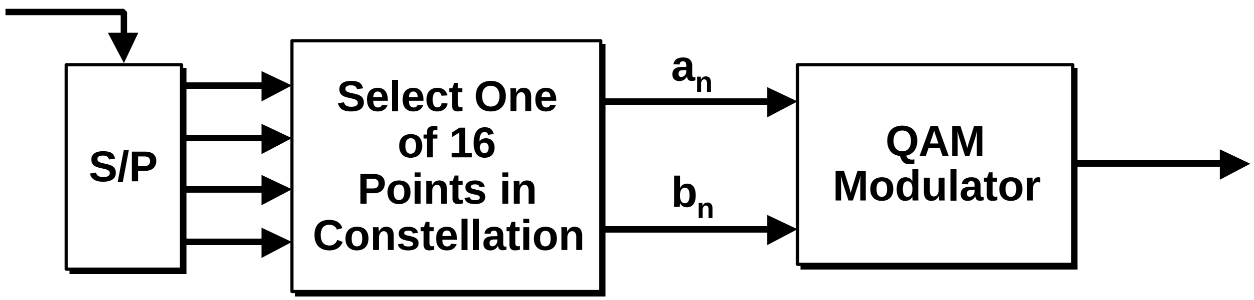
The d_{min}^2 between points in subsets is increased by at least a factor of 2 with each partitioning. Note that this gain has to overcome the larger constellation power.



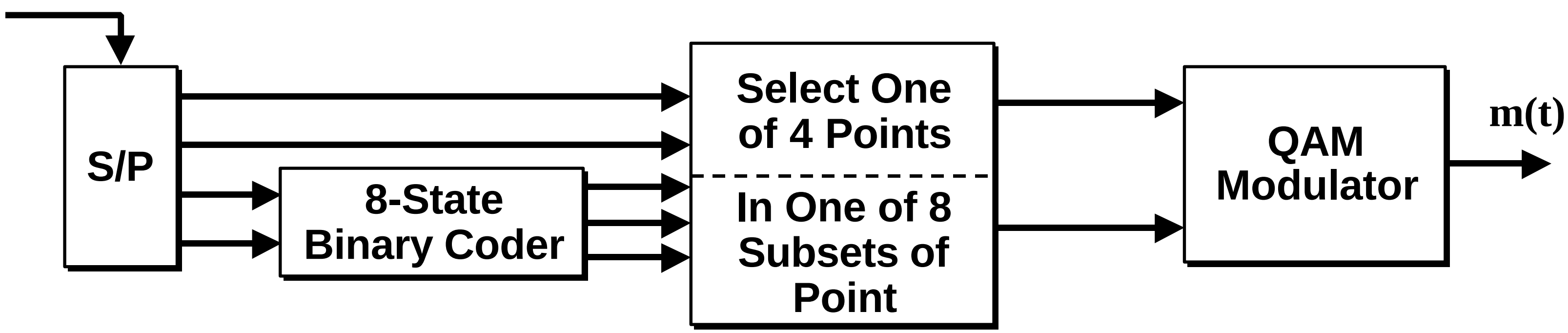
Partitioning of 16-Point Constellation

Trellis Coding for Two-Dimensional Modulation

(1) Example of an Uncoded System (16-Point Constellation)



(2) Example of an 8-State Trellis-Coded System (32-Point Constellation)



(8-State Sequential Logic)

 # States = $2^{\text{\# Shift Registers}}$

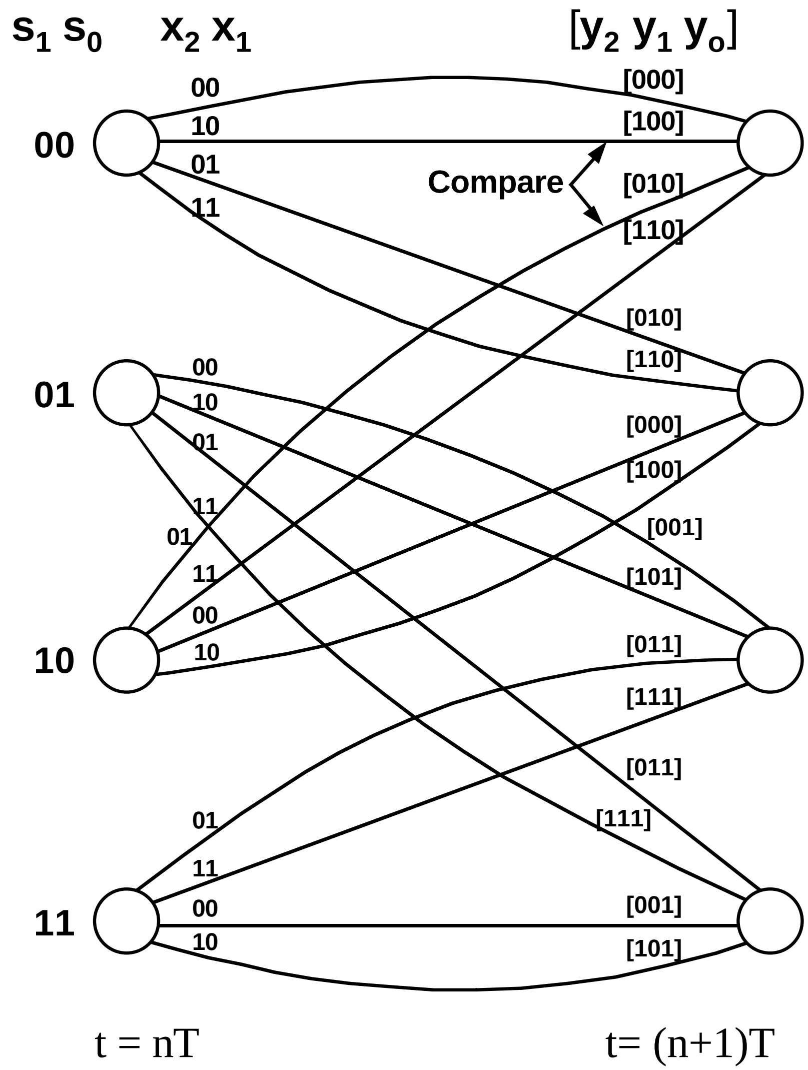
Decoding: Viterbi Algorithm

Note: Number of states in binary coder = number of trellis states

number of subsets = $2^{\text{\# coder output bits}}$

The Viterbi Algorithm

As with decoding of convolutional codes, the **Viterbi Algorithm (VA)** is an application of dynamic programming that finds "shortest paths" [here maximum likelihood sequences]. A critical feature is that **complexity grows linearly with time**, rather than exponentially with time (ie, the number of symbols transmitted). The VA finds the sequence at a minimum Euclidean distance from the received signal ---or equivalently the accumulated squared error is minimized.



1. Keep track of n possible sequences, each terminating in one of the n states.
2. For each new received symbol, calculate the new error value for each allowed continuation of each sequence, and add it to the accumulated error of the sequence up to that time.
3. For each new state, keep only the one sequence with minimum accumulated error. Discard the other sequences. Keep track of the new error and the bits associated with these n survivor sequences.
4. Depend on "merges", and enough delay, to output the past history of the sequence.

The operation of the algorithm is best described by a trellis diagram, which shows the possible sequences of states.

Decoding TCM [Baseband/Baseband Equivalent Channel]

- The receiver's (Viterbi Algorithm's) task is to determine the most likely states
- For the most likely state sequence, the uncoded (parallel) branches are resolved by selecting the signal points closest to the received signals
- The TCM decoder estimates the path that the encoded signal traversed through the trellis by associating with each branch of the trellis a branch metric and finding the path whose total metric is minimized
  - Let the sequence of transmitted M-ary symbols be denoted by $y_0, y_1, y_2, \ldots$
  - The received signal is

$$r(t) = m(t) + u(t) = \sum_{k=0}^{K-1} y_k \, p(t - kT) + u(t), \text{ where } u(t) \text{ is the channel noise}$$

  - After processing through the matched filter and sampling at t=NT, the maximum likelihood receiver minimizes

$$D^2(\vec{r}, \vec{s}) = \sum_{k=0}^{K-1} \left[r(kT) - \tilde{m}(\hat{s}_k, \hat{s}_{k-1}; x_k) \right]^2$$

  - where, $\tilde{m}$ is the estimate of the transmitted signal provided by the state estimates $\tilde{s}_k$ and x_k is the current input to the coder

- Note that we cannot detect the symbols, y_k , independently as would be the case for uncoded modulation (since not all state sequences are allowed) ---just as with convolutional codes
- Cannot use "brute force" since this requires searching over M^k possible state sequences (complexity grows exponentially with time, and the decoding delay would increase)

Decoding TCM ---at each sampling instant (t=kT)

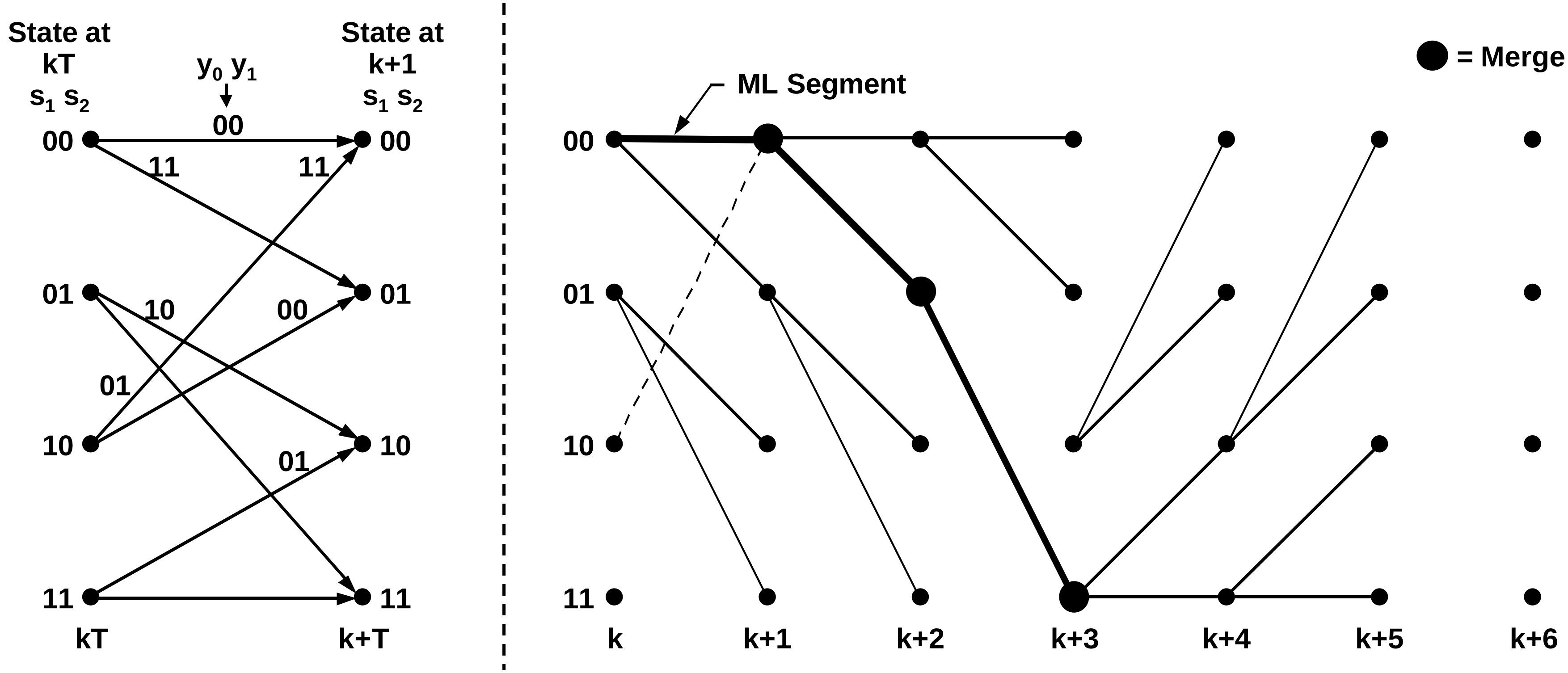
- For each state, the metric $D_{K-1}{}^2(\vec{r},\vec{s}) = \sum_{k=0}^{K-1}\left[r(kT) - \tilde{m}(\hat{s}_k, \hat{s}_{k-1}; x_k)\right]^2$ is stored (start with k=0)

- For each state calculate the [see the text, or EE6712 course notes, for decoding convolutional codes]

$$\min\left(\left[r_k - \tilde{m}_k(s_k, s_{k-1}; x_k)\right]^2 + D_{k-1}(s_k, s_{k-1}; x_k)\right), \quad \text{over } s_{k-1} \text{ for all the states } s_k$$

- Update the metric, store the "winning" path (survivor) segment, and remove the other paths to state
- Keep one survivor path per state
- At each sampling instant, we do not know which state is optimal
- Systems issues.
  - What about delay? Do we have to wait for the last symbol to pick minimum distance (optimum) sequence
  - Two alternatives: wait for a "merge" or force a merge
  - How sensitive is system performance to other impairments (eg, uncompensated phase errors)
  - Significant effort was expended to find TCM codes that are immune to 90 degree phase rotations (data-directed phase locked loops lock at multiples of 90 degrees for constellations with 90 degree symmetry) and do not lose any coding gain

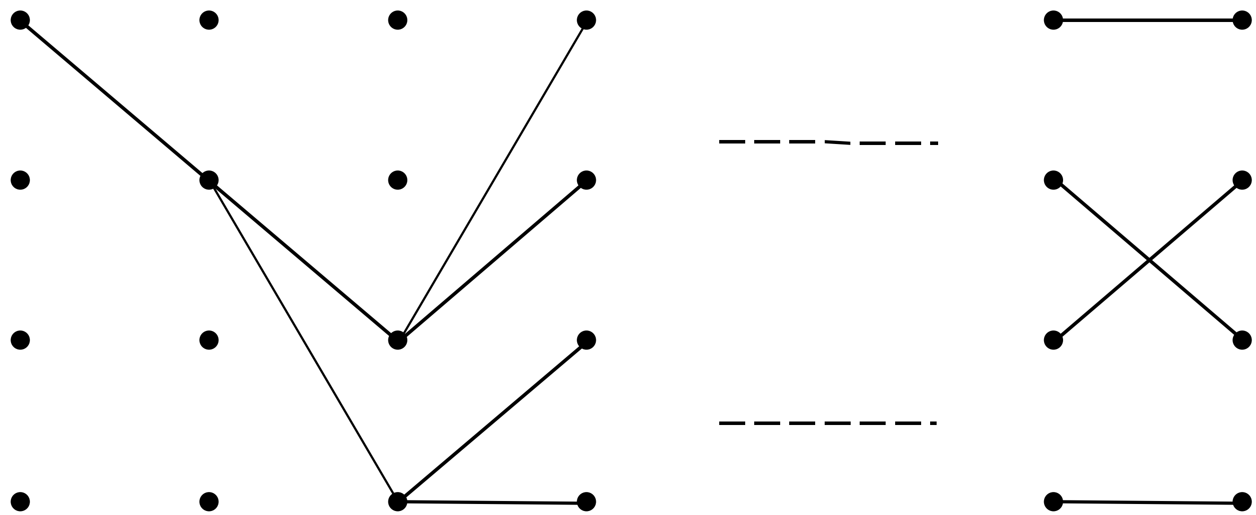
Viterbi Decoding



- Compare $(r_k - A_{solid})^2$ vs $(r_k - A_{dashed})^2$ to determine survivor for each state. We show only the survivors.
- Note that "00" is a merged node at k (evident at k+2); similarly "00" is a merged node at k+1 (evident at k+3). "11" merged at k+3 bold segment is part of max. likelihood sequence.

Merges

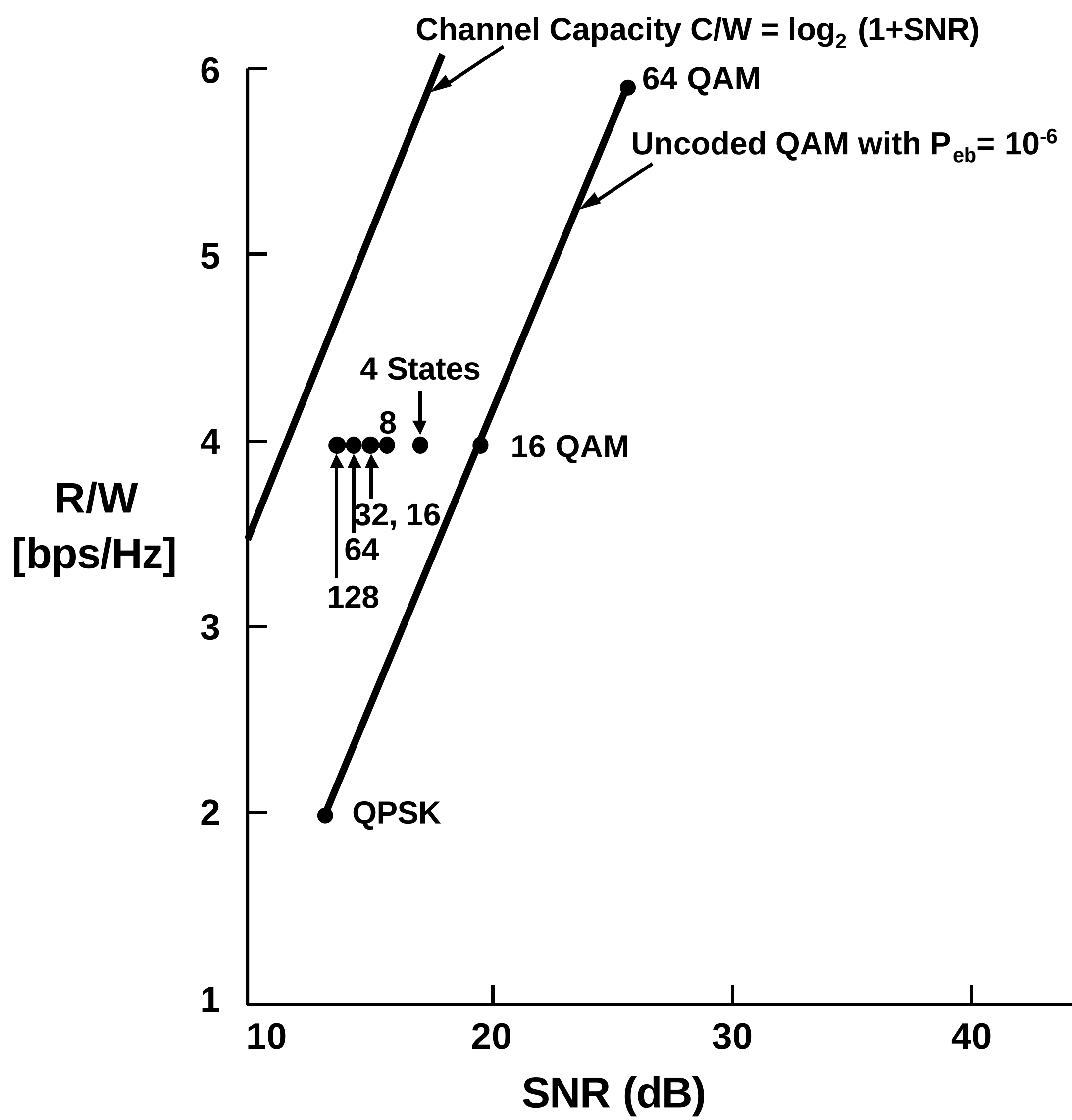
At any time, we retain one survivor path leading to each state. It appears that we never make a final decision. In practice we must depend on merging, that is, all survivors having a common past if we go back far enough.



As a rule of thumb, if we go back at least 5 times the memory of code, we will find that all survivor paths have merged with very high probability.

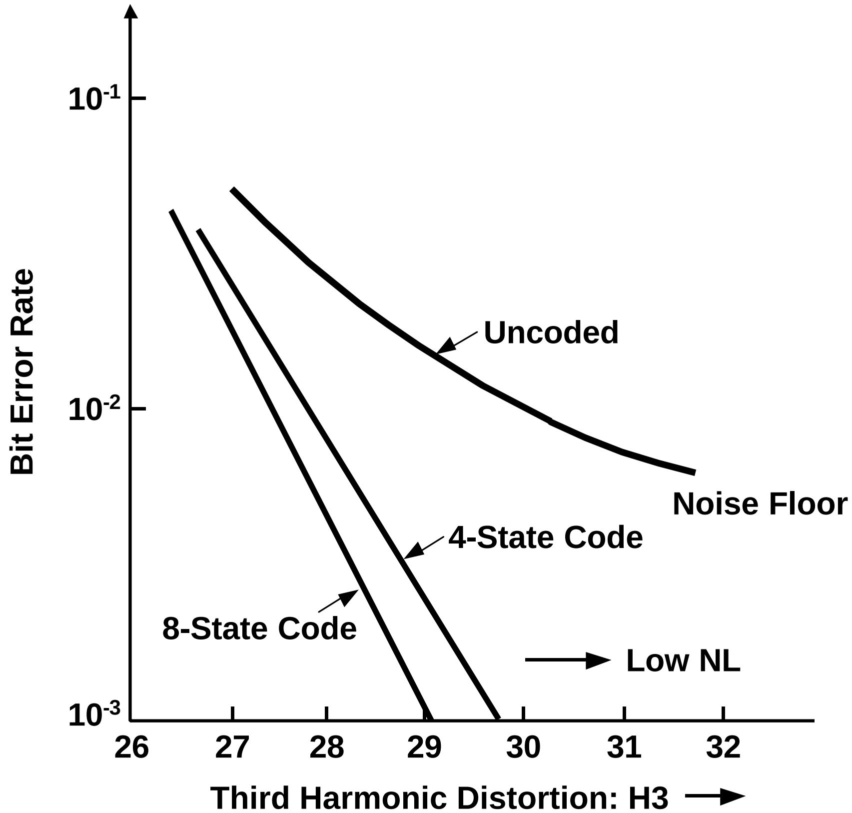
The final output is delayed by this fixed amount. If a merge is not achieved, we must output something, the best choice being the delayed symbol on the survivor path with the lowest error metric. If there is sufficient delay, this will not be the principal error mechanism.

Gain Due To Trellis Coding

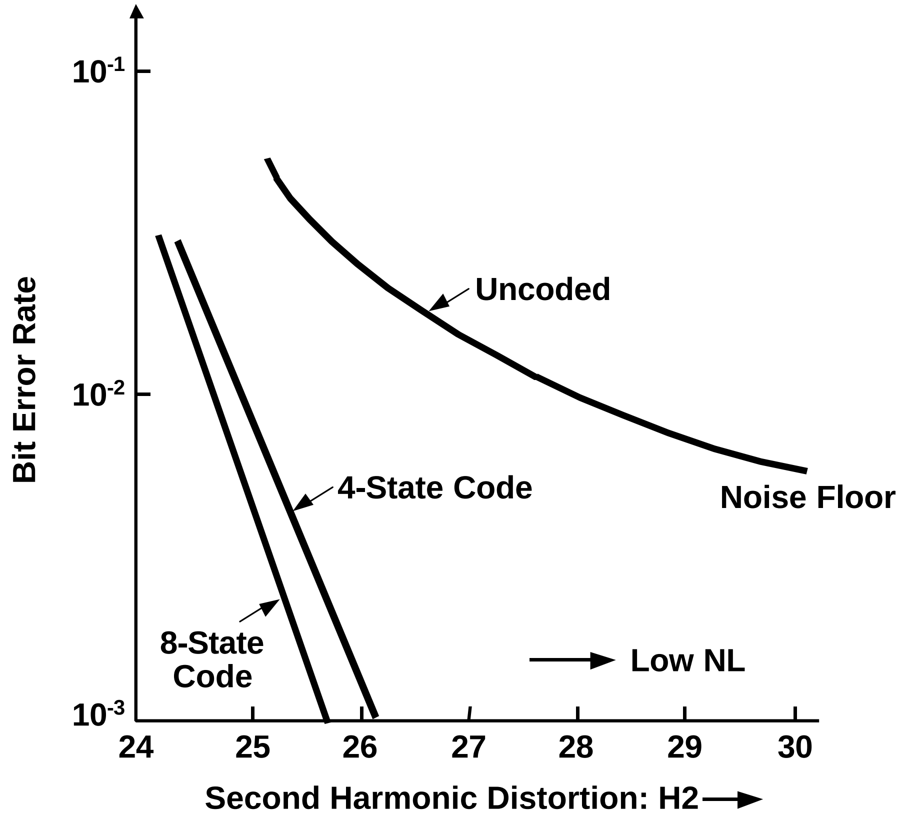


Number of States	Gain (Reduction in Required SNR in dB)
4	3.0
8	4.0
16	4.8
32	5.2
64	5.4
128	6.0

Trellis Coding Experiments



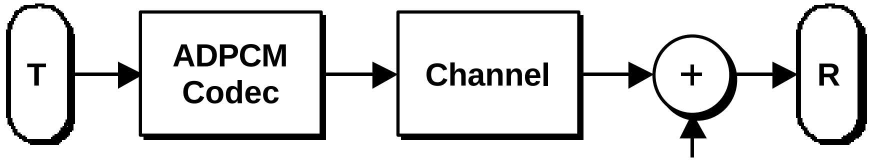
Effect of third harmonic distortion (H3) (background impairments: linear and nonlinear distortion, and phase jitter).

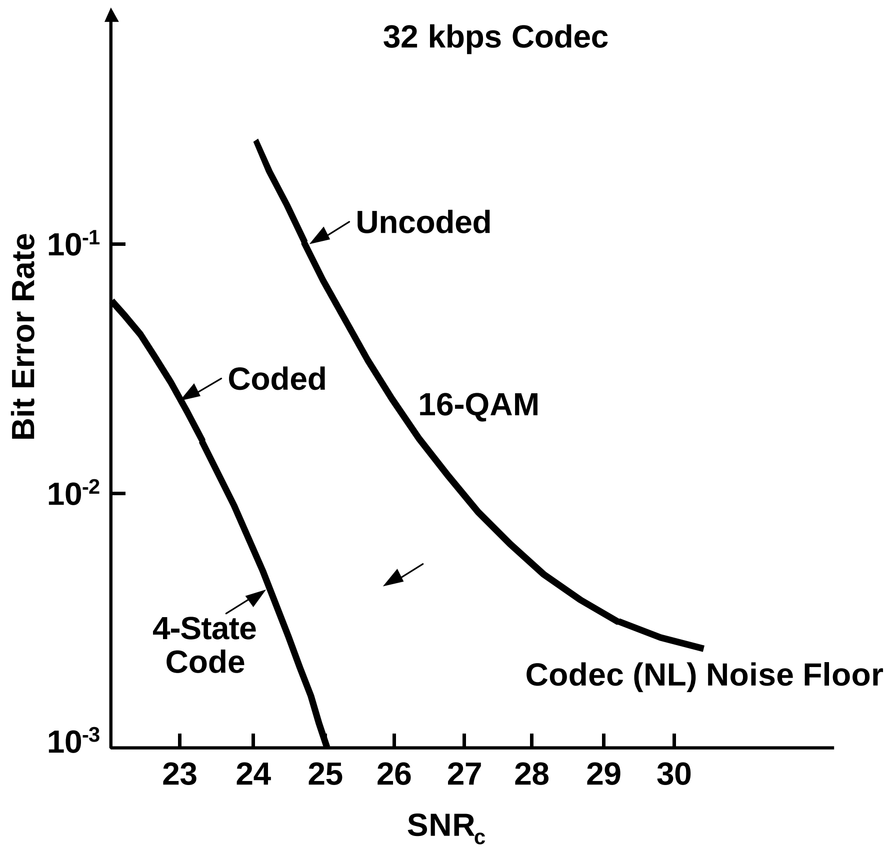


Effect of second harmonic distortion (H2) (background impairments: linear and nonlinear distortion, and phase jitter).

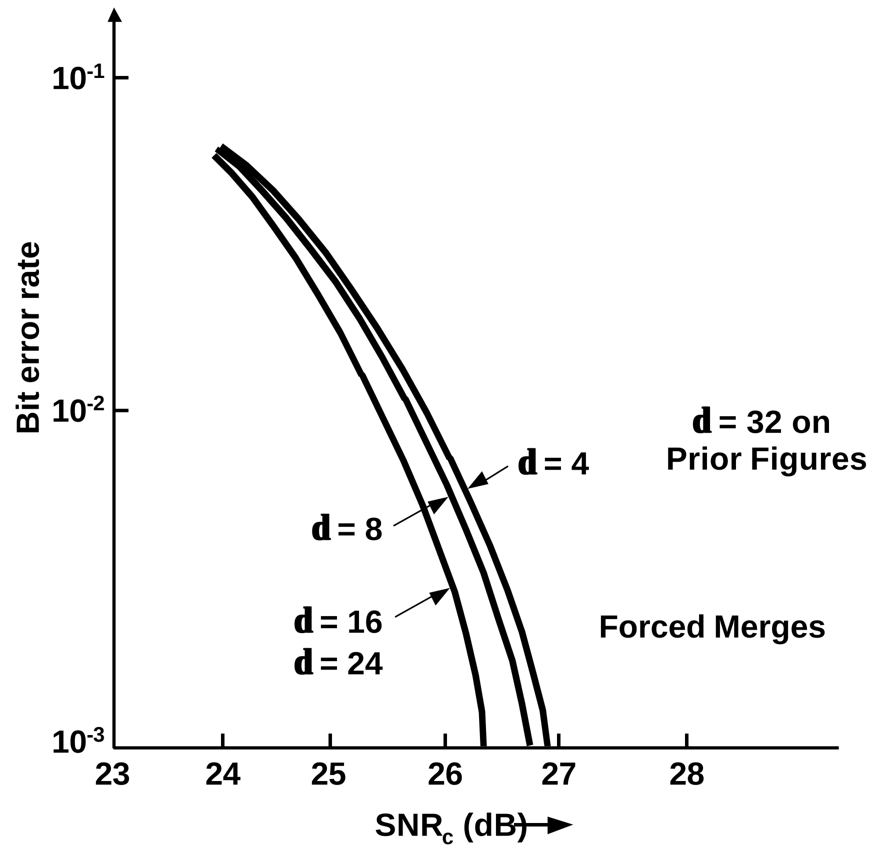
Trellis Coding Experiments

Performance of 9600 bps Transmission





Measured performance over ADPCM codec, linear and nonlinear distortion, noise, and jitter.



Effect of decoding depth (d) on the four-state coded system.

Block Coding - Four Dimensional QAM (4D-QAM)

$$X(t) = \sum_{k} [[A_k \cos\omega_c t - B_k \sin\omega_c t]\ \ p(t-kT)$$

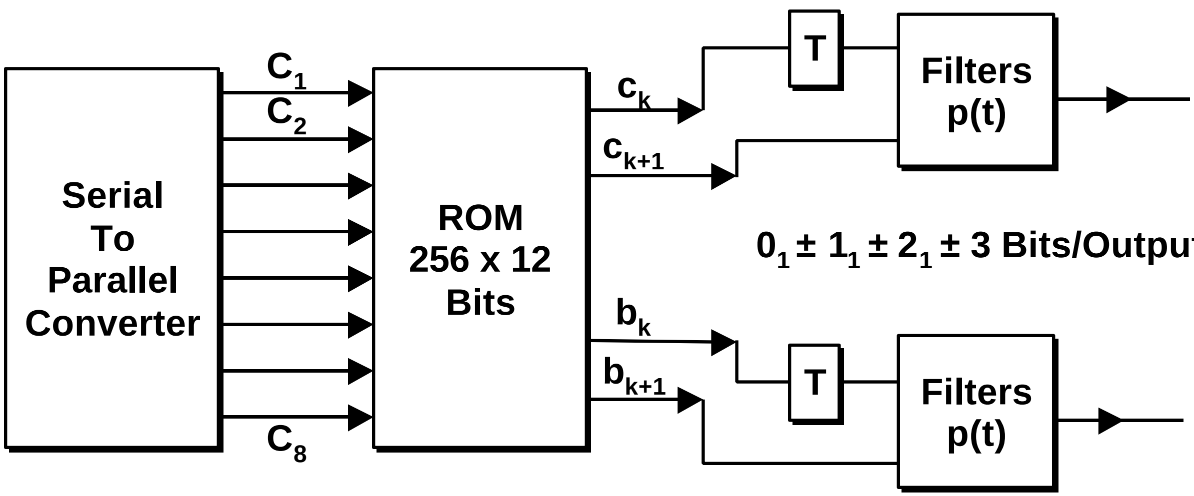
4-D Constellation of Points

$$(P_{2N}, P_{2N+1}) = (A_{2N}, B_{2N}, A_{2N+1}, B_{2N+1})$$

Symbol Interval = 2T = 2 Time Slots

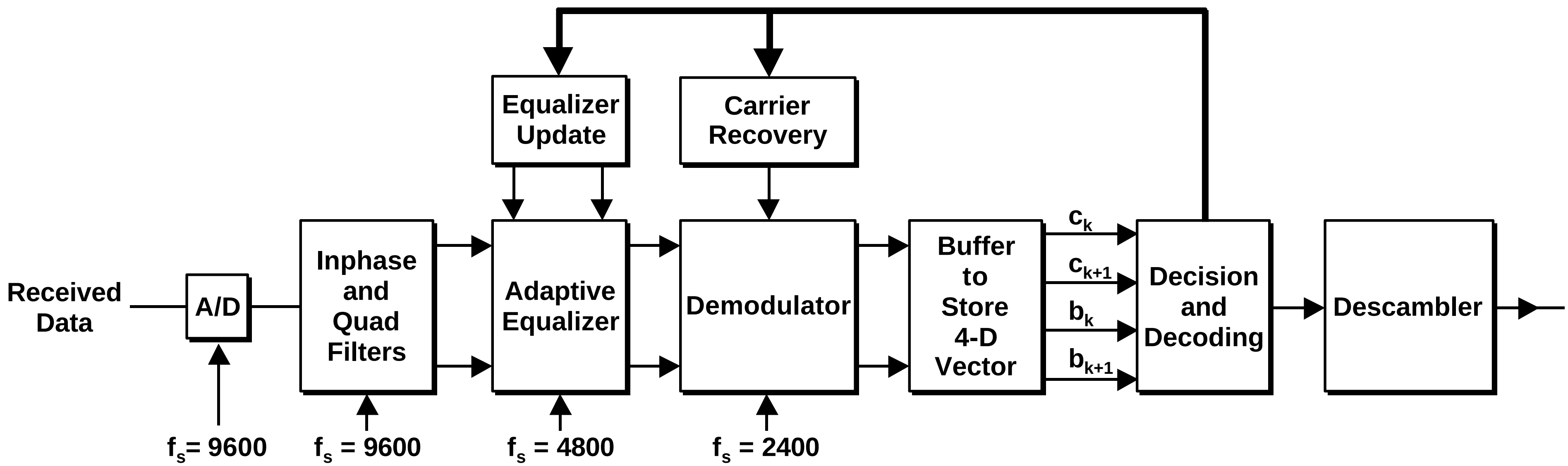
Receiver Sampling Interval = T

256 Points → 8 Bits/Symbol Interval or 4 Bits/Sample Interval



Encoder Block Diagram

Block Coding - Four Dimensional QAM (4D-QAM) (continued)



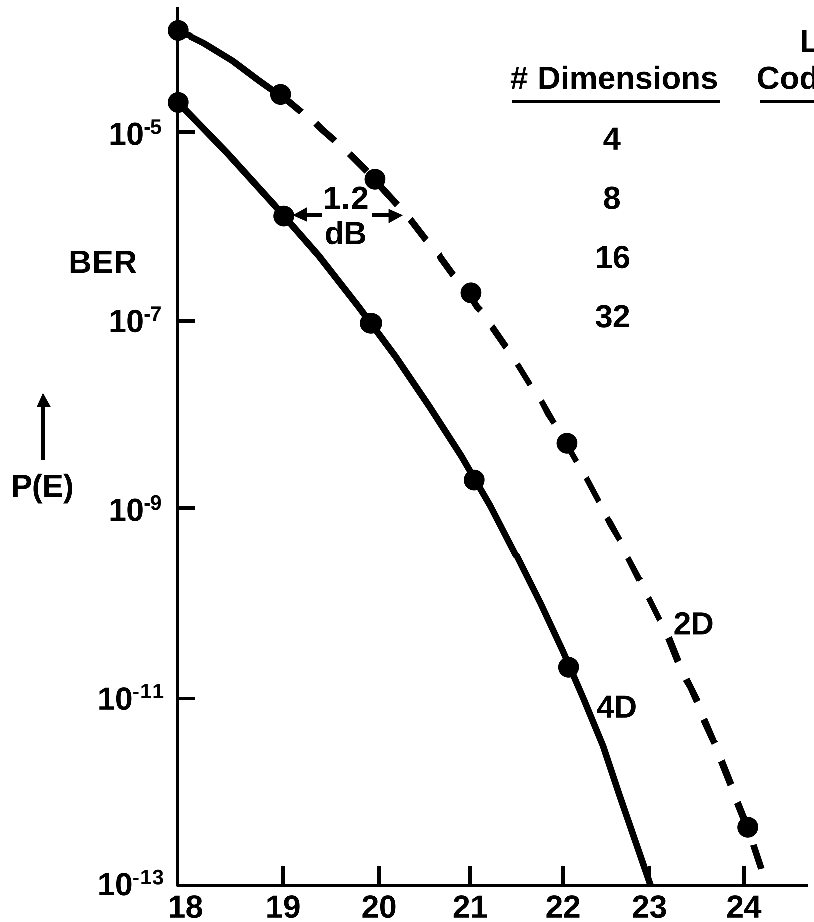
Receiver Block Diagram

Block Coding Using Lattices

- Lattices: (a_1, b_1, a_2, b_2) = either all even or all odd integers
- Use lattices to construct multi-dimensional signal constellation
- Simple decodings exist
- Modest coding gains achievable

Examples of Bit Assignment

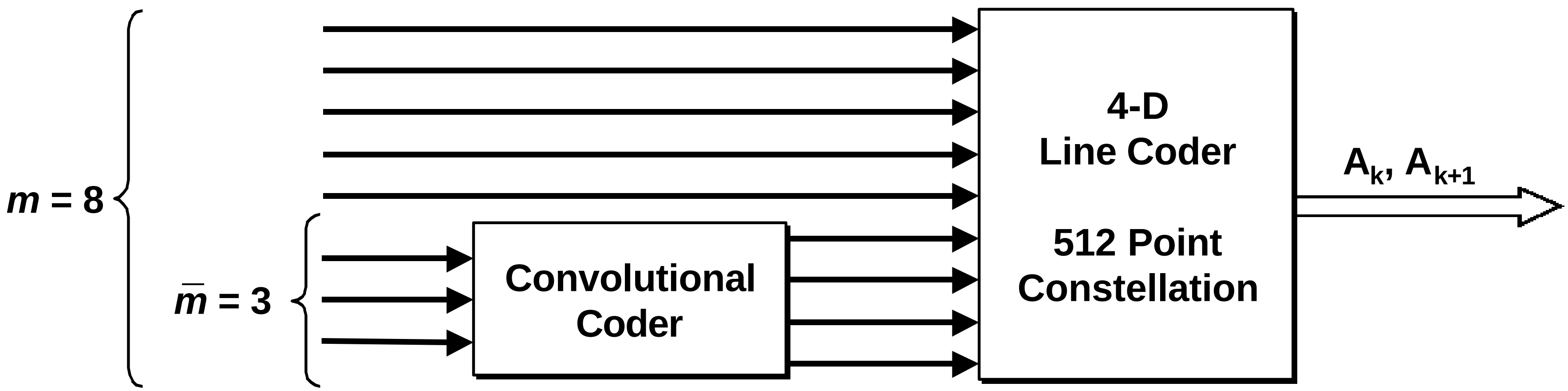
00000110	1	-1	-1	1
00001110	-1	-1	-1	1
00011101	-2	-2	2	-2
00010101	2	-2	2	-2
00111111	0	-2	-2	-2
00101101	-2	2	0	-2
01110111	1	-1	1	3
01110000	1	1	1	3
11001110	-3	-3	-1	1
11000001	3	3	1	-1
10000000	2	0	0	0
10000111	0	0	0	-2
10010001	2	-2	0	0
10011111	0	-2	-2	0



# Dimensions	Lattice Coding Gain
4	1.2
8	2.0
16	2.83
32	4.00

Error Probability P(E) versus SNR for Both 4-D and 2-D Signal Constellations

Multi-Dimensional Trellis Codes

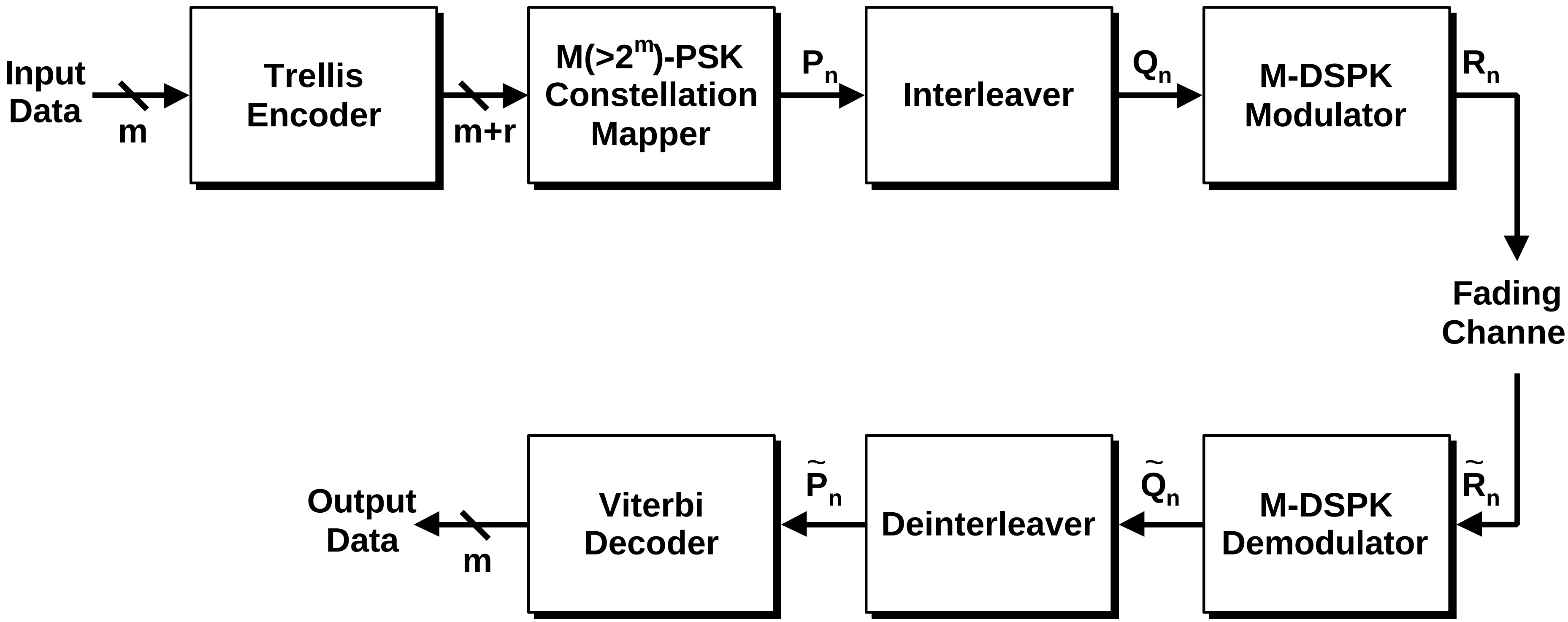


•A four-dimensional trellis coder. For every set of 8 bits that come in at the left, one four-dimensional symbol is produced by the line coder. These are actually transmitted, however, as two successive two-dimensional symbols

•Multi-dimensional coding overcomes one of the main limitations of trellis codes --- doubling the constellation size [doubling reduces the noise margin by ~ 3dB]. In more than two dimensions, doubling the size of the alphabet entails less than a 3 dB loss.

•Typically, a four-dimensional encoding can add ~ 1dB to the gain of a trellis code.

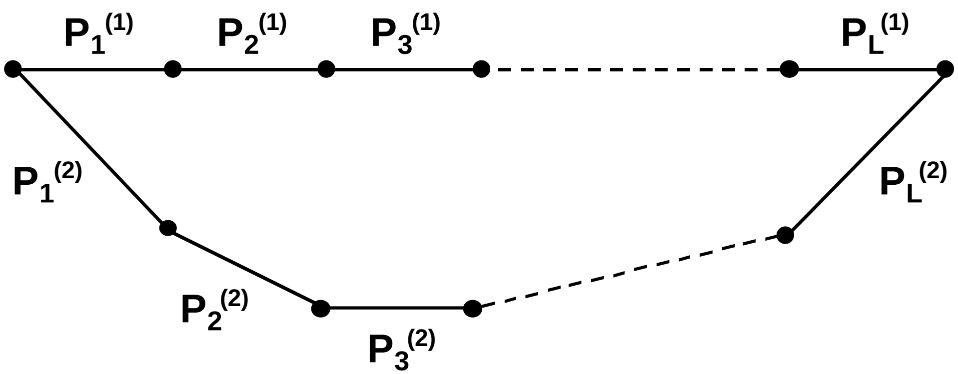
Transceiver Structure of Trellis-Coded M-DPSK for Fading Channels



- Systems optimized for the Gaussian channel are *not* optimized for fading channels.
- This is primarily due to the need for built-in time diversity in the coded modulation.
- Traditional L-fold time diversity can be achieved by repeating a symbol in L different time slots
    - Each slot should experience independent fading
    - Standard method is interleaving

Design Criteria of Trellis-Coded M-PSK for Fading Channels

- Maximize the minimum amount of built-in time diversity between valid sequences of signal point of the code;
- For any two valid sequences of signal points with the minimum amount of time diversity, maximize further the minimum squared product distance between them.

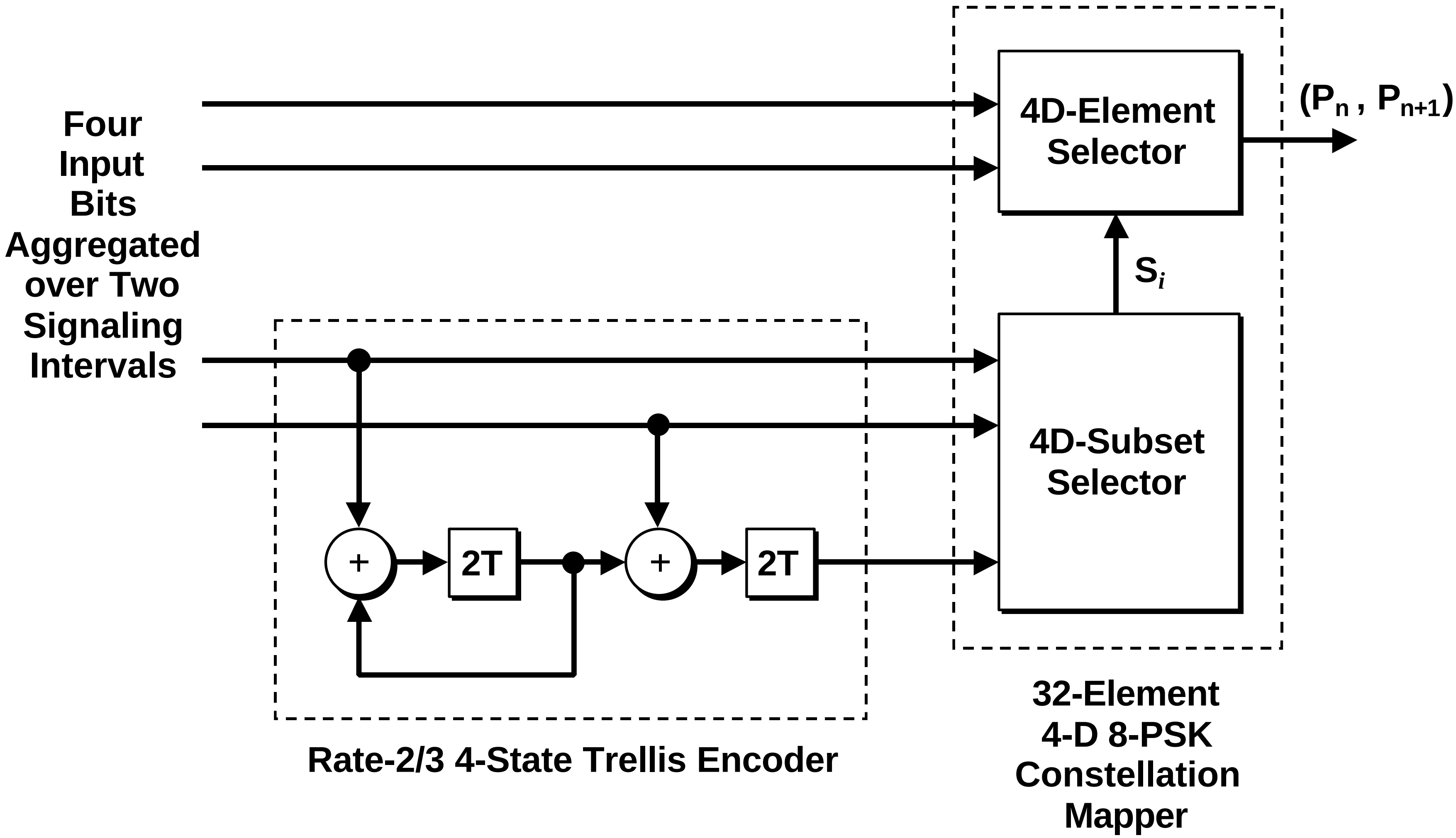


$$A \triangleq \left\{ i \,\middle|\, i \in \{1,2,3,\ldots,L\},\ P_i^{(1)} \neq P_i^{(2)} \right\}$$

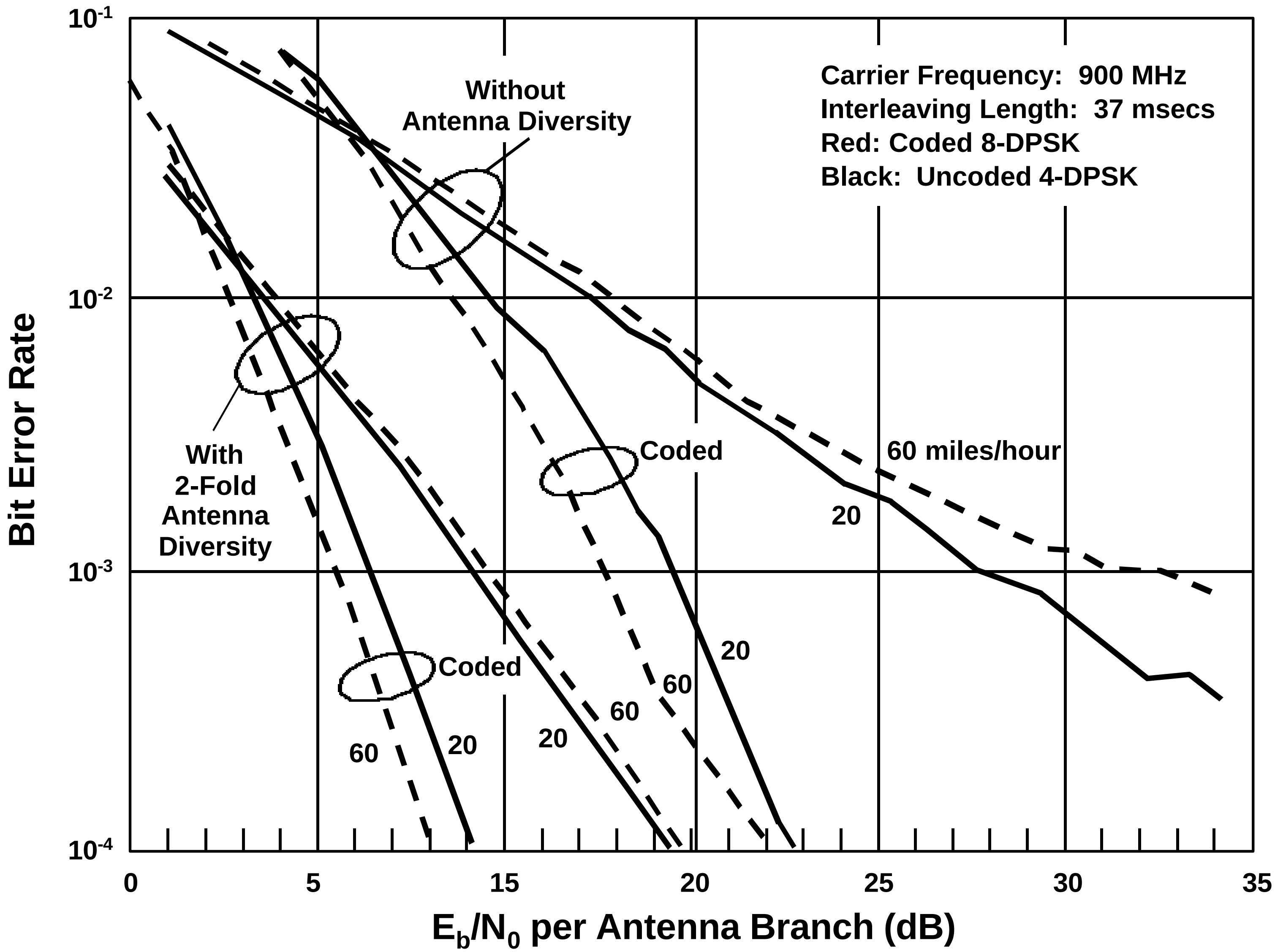
Amount of Time Diversity $\triangleq$ **Number of Elements in Set A**

$$\text{Squared Product Distance (SPD)} \triangleq \prod_{i \in A} \left| P_i^{(1)} - P_i^{(2)} \right|^2$$

4-Dimensional 4-State Trellis-Coded 8-PSK with 2 Bits/Symbol



Fading-Channel Performance of 4-Dimensional 4-State Trellis-Coded 8-DPSK with 2 Bits/Symbol



Code Comparisons

Scheme	Bandwidth Efficiency (bits/symbol	Type of Interleaver	Minimum Amount of Time Diversity	Minimum Squared Product Distance	Minimum Decoding Depth (symbols)	Coding Gain without Antenna Diversity (dB)*		Coding Gain with 2-Fold Antenna Diversity (dB)*	
						20 miles/hour	60 miles/hour	20 miles/hour	60 miles/hour
4D 8-State Trellis-Coded 4-DPSK	1.5	Bit	4	16	11	9.1	16.3	17.9	23.1
4D 4-State Trellis-Coded 8-DPSK	2	Signal Point	2	4	5	8.2	14.1	16.2	21.2
2D 16-State Trellis-Coded 12-DPSK	2.5	Signal Point	2	3	5	6.5	13.1	14.8	20.3

* Relative to uncoded 4-DPSK without antenna diversity at BER = 10^{-3}.
The carrier frequency is 900 MHz. The interleaving length is 37 msecs.

Conclusions

- Just as for Gaussian channels, trellis-coded modulation is shown to be power-efficient and bandwidth-efficient modulation technique for fading channels. A large amount of coding gain on the order of 10 dB can be easily obtained in this case.
- Using a two-fold antenna diversity can improve the performance of trellis-coded modulation by another 8 dB.