Trellis Coded Modulation

- Earlier channel coding schemes (block coding and convolutional coding) provided improvements in error performance at the cost of bandwidth expansion.

- Transforming each input $k$-tuple into a larger output codeword $n$-tuple requires additional transmission bandwidth.

- Thus, in the past, coding was not popular for bandlimited channels (e.g. telephone channels), where signal bandwidth expansion is not practical.

- Since 1984, there was active research interested in combined modulation and coding schemes, called trellis coded modulation (TCM).

- TCM schemes achieve error–performance improvements without the expansion of signal bandwidth.
Trellis Coded Modulation

- TCM schemes use redundant nonbinary modulation in combination with a finite–state–machine (the encoder)

- A finite–state–machine (FSM) is the general name given to a device that has a memory of past signals

- The adjective *finite* refers to the fact that there are only a finite number of unique states that the machine can encounter

- A state consists of the smallest amount of info, that together with a current input to the machine, can predict the output of the machine

- For each symbol interval, a TCM FSM encoder selects one of a set of waveforms to be transmitted

- The noisy received signals are detected and decoded by a soft–decision maximum likelihood detector/decoder
Hard Versus Soft Decisions

- A demodulator output can be configured into a variety of ways, including a hard decision or a soft decision.

- In case of making hard decisions, the output of the demodulator is quantized into two levels, zero and one, and fed into the decoder.

- Since the decoder operates on hard decisions made by the demodulator, the decoding is called hard–decision decoding.

- The demodulator can also be configured to feed the decoder with a quantized value of $z(T)$ greater than two levels.

- When the quantization level of the demodulator output is greater than two, the decoding is called soft–decision decoding.

- 8 levels (3–bits) of quantization are shown in Fig. 1. (slide 6)
Hard Versus Soft Decisions

- When the demodulator sends a hard binary decision to the decoder, it sends it a single binary symbol.

- When the demodulator sends a soft binary decision, quantized to 8 levels, it sends the decoder a 3-bit word describing an interval along $z(T)$.

- Sending a 3-bit word instead of a single binary symbol is equivalent to sending the decoder a *measure of confidence* along with the code-symbol decision.

- Using soft decoding, if the demodulator sends 111 to the decoder, this means that it declares the code symbol to be a 1 with very high confidence.

- But sending 100 would mean that it declares the code symbol to be a 1 with very low confidence.

- For a Gaussian channel, 8-level quantization results in performance improvement of about 2 dB in SNR compared to 2-level quantization.
Hard Versus Soft Decisions

- Analog (or infinite-level) quantization results in about only 2.2 dB performance improvement!
- Keep in mind that soft-decoding provides improved performance at the expense of more processing and higher memory requirement at the decoder.

Fig. 1. Hard and soft decoding decisions.
The Idea Behind TCM

- The development of TCM started with the notion that “not all signal subsets (in a constellation) have equal distance properties”

- For example, in MPSK, antipodal signals have the best distance properties for easily discriminating one signal from the other, while nearest neighbor signals have poor distance properties

**Fig. 2.** Constellation diagrams of BPSK and 8-PSK.
The Idea Behind TCM

- The wizard analogy makes understanding TCM very simple: Imagine that there is a wizard at the transmitter!

- As the message bits enter the system, the wizard recognizes that some of the bits are most vulnerable to the degradation effects of channel impairments; hence they are assigned modulation waveforms associated with the best distance properties.

- Similarly, other bits are judged to be very robust, and hence, they are assigned waveforms with poorer distance properties.

- Modulation and coding take place together!

- The wizard is assigning waveforms to bits (modulation), but the assignment is performed according to the criterion of better or worse distance properties (channel coding).
TCM: An Illustrative Example

- Consider in Fig. 3 this trellis encoder example with an 8-state FSM driving a 3-bit to 8-PSK signal mapper

- $a_n$ denotes a complex-valued discrete channel signal transmitted at modulation time $nT$

Fig. 3. Realization of an 8-PSK code by means of a systematic convolutional encoder with feedback.
TCM: An Illustrative Example

- Uncoded 4–PSK is regarded as a reference system

The encoder consists of two parts:

- The first is a FSM with 8 states, such that each of the states is defined by the contents of the memory cells, $M$

- The second part is the signal mapper. Its function is memoryless mapping of the three bits $(y_n^2, y_n^1, y_n^0)$ into one of the eight symbols of an 8–PSK signal set

- The FSM accepts two bits $(x_n^2, x_n^1)$ at each symbol time $n$ and moves from a state $s_n$ to one of four next states $s_{n+1}$

- Assuming that the encoder in Fig. 3 operates continuously, there are four choices at each time instant $n$, which allows us to transmit two information bits $(x_n^2, x_n^1)$ per symbol, the same as the reference uncoded 4–PSK system
TCM: An Illustrative Example

- The trellis shown in Fig. 4 is the two dimensional representation of the operation of the encoder.

- It shows all possible state transitions starting from an originating state and terminating in a final state.

![Figure 4. Eight-state trellis diagram for coded 8-PSK. (Sklar, Digital Communications, 2nd edition)](image-url)
TCM: An Illustrative Example

- The three successive horizontal lines, labelled 000 and the three successive lines labelled 676 show possible paths with their associated signal sequences through the trellis.

- It is apparent here that the term trellis–coded modulation originates from the fact that these encoded sequences consist of modulated symbols \( \{a_n\} \), rather than bits.

So, what is the advantage of using TCM in this example, instead of uncoded 4-PSK?
TCM: An Illustrative Example

- In case of 4–PSK, a decoder is most likely to make an error due to noise, by confusing two neighboring signals, say $a^1$ and $a^2$, with probability

$$P_{a^1 \rightarrow a^2} = Q\left(\sqrt{\frac{d_E^2 E_s}{2N_0}}\right)$$

where $d_E^2 = 2$ for a unit energy 4–PSK signal constellation

- In case of TCM, the FSM puts restrictions on the symbols that can be transmitted in a sequence, and these restrictions can be exploited by the Viterbi decoder

- Consider Fig. 4 and assume that one sequence is correct $\{a^c\} \rightarrow 000$, and the other is erroneous, $\{a^e\} \rightarrow 676$. The Viterbi decoder will make an error between these two sequences with a probability as in (1); With $d_E^2$ as follows

$$d_E^2 = (\sqrt{2})^2 + (0.765)^2 + (\sqrt{2})^2 = 4.586,$$
TCM: An Illustrative Example

\[
d_{\text{min}} = \frac{\sin(\pi/2)}{\sin(\pi/4)} = \sqrt{2}
\]

\[
d_{\text{min}} = 2\sin(\pi/8) \approx 0.765
\]

Fig. 5. \(d_{\text{min}}\) calculation for 4–PSK and 8–PSK, assuming unit energy for each signal.
TCM: An Illustrative Example

where $(\sqrt{2})^2$, $(0.765)^2$ and $(\sqrt{2})^2$ are the Euclidean distances between the 8–PSK constellation points, labelled 0 and 6, 0 and 7, and 0 and 6, respectively *(examine Fig. 5)*

- Notice that the squared Euclidean distance equal to 4.586 for the TCM is much larger than that of uncoded 4–PSK which is equal to only 2

Fig. 6. A plot of the Q function.
TCM: Increasing Signal Redundancy

- TCM maybe implemented with a convolutional encoder (as seen in the illustrative example)

- $k$ current bits and $K - 1$ prior bits are used to produce $n = k + p$ code bits, where $K$ is the encoder constraint length and $p$ is the number of parity bits

- Such an encoding increases the signal set size from $2^k$ to $2^{k+p}$

- Ungerboeck investigated the increase in channel capacity achievable by signal set expansion, and concluded that most of the achievable coding gain over conventional uncoded multilevel modulation could be achieved by a factor of two ($p = 1$)

- This is accomplished by encoding with a rate $k / (k + 1)$ code and subsequently mapping groups of $(k + 1)$ bits into the set of $2^{k+1}$ waveforms
Fig. 7 (next slide) shows a number of constellation sets before and after increasing their sizes to accommodate for TCM

In each of these cases, the system is configured to use the same average signal power before and after coding.

To provide the needed redundancy for coding, the signaling set is increased from $M = 2^k$ to $M' = 2^{k+1}$. Thus $M' = 2M$.

However, this increase in the constellation size does not result in an increase in the required bandwidth.

An expanded signal set results in a reduced distance between adjacent symbol points.

In uncoded systems this results in a degradation of error performance.

However in TCM, because of the introduced redundancy, this reduced distance does not matter.
Fig. 7. Increase of signal set size for trellis coded modulation.
(Sklar, *Digital Communications*, 2nd edition)
• Instead, the free distance, which is the minimum distance between members of the set of allowed code sequences, determines the error performance.

• The objective of TCM is to assign waveforms to trellis transitions so as to increase the free distance between the waveforms that are most likely to be confused (recall the wizard analogy).
Ungerboeck Partitioning for 8–PSK

- Assume that the receiver uses soft decision, so that the appropriate distance metric is Euclidean

- In order to maximize the Euclidean distance (ED), we employ the code–to–signal mapping pioneered by Ungerboeck

- Ungerboeck proposes successive partitioning of a modulation–signal constellation into subsets having increasing minimum distance \(d_0 < d_1 < d_2 \ldots\) between the elements of the subsets

- If the average signal power is chosen to be 1, then \(d_0\) between any two adjacent signals is seen to be \(2 \sin(\pi/8) \approx 0.765\)
Ungerboeck Partitioning for 8–PSK

- The first level of partitioning results in subsets $B_0$ and $B_1$, where the distance between adjacent signals is $d_1 = \sqrt{2}$.

- The second level of partitioning results in subsets $C_0$, $C_1$, $C_2$, and $C_4$, where the distance between adjacent signals is $d_2 = 2$. 
Ungerboeck Partitioning for 8–PSK

Fig. 7. Ungerboeck partitioning of an 8–PSK signal set. (Sklar, Digital Communications, 2nd edition)
TCM: Rules for Mapping of Waveforms to Trellis Transitions

1. If $k$ bits are to be encoded per modulation interval, the trellis must allow for $2^k$ possible transitions from each state to the next.

2. More than one transition may occur between pairs of states.

3. All waveforms should occur with equal frequency and with a fair amount of regularity and symmetry.

4. Transitions originating from the same state are assigned waveforms either from subset $B_0$ or $B_1$ – never a mixture between them.

5. Transitions joining into the same state are assigned waveforms either from subset $B_0$ or $B_1$ – never a mixture between them.
6. Parallel transitions are assigned waveforms either from subset $C_0$ or $C_1$ or $C_2$ or $C_3$ – never a mixture between them

- These rules guarantee that codes constructed in this way will have a regular structure and a free ED that will always exceed the minimum distance between signal points of the uncoded reference modulation

![Four-state trellis with parallel paths.](Fig. 8)
Fig. 8 shows a possible code–to–signal mapping, using a four–state trellis with parallel paths.

The code–to–signal assignments are made by examining the partitioned signal space in Fig. 7 in relation to the trellis diagram in Fig. 8 and the six mapping rules.

Written on the trellis transitions are the waveform numbers that have been assigned to those transitions by following the partitioning rules.

Note that for 8–PSK, the waveform assignments comply with rule 1: There are $k + 1 = 3$ code bits and thus $k = 2$ information bits, and there are $2^2 = 4$ transitions into and out of each state.

The waveform transitions comply with rule 6 because each parallel pair of transitions have been assigned waveforms from subset $B_0$ or $B_1$.

In Fig. 8, the states of the trellis have been designated according to the waveform types that may appear on the transition leaving that state.
• Note that the states can be referred to in terms of the signal subsets as the $C_0 C_1$ state or the $C_2 C_3$ state, and so forth.

• Another possible designation is in terms of the waveform numbers, as the 0426 state or the 1537 state, and so forth.
TCM: Coding Gain

- The general equation describing the gain provided by TCM is

\[
G(\text{dB}) = 10 \log_{10} \left( \frac{d_{f,c}^2 / E_{s,c}}{d_{\text{ref}}^2 / E_{s,\text{ref}}} \right)
\]  
(3)

- For the specific case of MPSK, where all the signals have the same energy, (3) simplifies to

\[
G(\text{dB}) = 10 \log_{10} \left( \frac{d_{f,c}^2}{d_{\text{ref}}^2} \right)
\]  
(4)

- Thus, for the illustrative example, at the beginning of this lecture, the gain provided by TCM through the use of an 8–state trellis (as in Fig. 4) is

\[
G(\text{dB}) = 10 \log_{10} \left( \frac{4.585}{2} \right) = 3.6 \text{ dB}
\]  
(5)
While the gain provided by TCM through the use of a 4–state trellis (with parallel paths, as in Fig. 8) would be

\[
G(\text{dB}) = 10 \log_{10} \left( \frac{d^2_{f,c}}{d^2_{\text{ref}}} \right) = 10 \log_{10} \left( \frac{4}{2} \right) = 3 \text{ dB}
\]

It is clear that for the case of using TCM to code 4–PSK signals into 8–PSK signals, it is best to utilize an 8–state trellis.
Fig. 9. The gain achieved by TCM encoding of 4–PSK into 8–PSK, using a 4–state trellis (with parallel paths, as in Fig. 8).
Exercise 1

- Perform Ungerboeck partitioning on the following constellation and calculate the Euclidean distance between any two signals at the most partitioned level

\[ A_0 \]

\[ \begin{align*}
&\begin{array}{cccccccc}
7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \\
-7 & -5 & -3 & -1 & 1 & 3 & 5 & 7 \\
\end{array}
\end{align*} \]

\[ \text{Waveform number} \]
\[ \text{Euclidean distance} \]
\[ d_0 = 2 \]

**Fig. 10.** An 8–PAM signal constellation set. (Sklar, *Digital Communications*, 2nd edition)
Exercise 1 Solution

Fig. 11. First level of Ungerboeck partitioning of 8–PAM signals. (Sklar, *Digital Communications*, 2nd edition)
Exercise 1 Solution

Fig. 12. Ungerboeck partitioning of 8–PAM signals.
(Sklar, Digital Communications, 2nd edition)
Exercise 2

- Perform Ungerboeck partitioning on the following constellation and calculate the Euclidean distance between any two signals at the most partitioned level.

![Fig. 13. A signal constellation set.](image-url)
Exercise 3

- Perform Ungerboeck partitioning on the following constellation and calculate the Euclidean distance between any two signals at the most partitioned level.

![A 16-QAM signal constellation set. (Sklar, Digital Communications, 2nd edition)](image)
Exercise 3 Solution

Fig. 15. First level of Ungerboeck partitioning of 16-QAM signals.
(Sklar, Digital Communications, 2nd edition)
Exercise 3 Solution

Fig. 16. Second level of Ungerboeck partitioning of 16-QAM signals. (Sklar, Digital Communications, 2nd edition)
Exercise 3 Solution

Fig. 17. Ungerboeck partitioning of 16-QAM signals.
(Sklar, *Digital Communications*, 2nd edition)
TCM Notes

- The mapping of bits into waveforms is done in such a way such that the squared Euclidean distance between symbol sequences that constitute an error event is maximized.

- The *free distance*, which is the minimum distance between members of the set of *allowed* code sequences, determines the error performance.

- Where is the proper place for evaluating allowed code sequences and distance properties? It is the trellis diagram. Using this diagram, the objective of TCM is to assign waveforms to trellis transitions so as to increase the free distance between the waveforms that are most likely to be confused.

- In convolutional coding, the trellis states are labelled with code bits. For TCM, the trellis transitions are labelled with modulation waveforms.