4.2 Error Correction Coding

Motivation

In the set of all multisymbol sequences $\mathcal{S}$.

- $\mathbf{w}_{m_a}, \mathbf{w}_{m_b}, m_a, m_b \in \mathcal{S}, a \neq b$ may be similar.

Sample

- $L = 3, |\mathcal{A}| = 4, \mathcal{W}_{\mathcal{AL}}[i,j] \cdot \mathcal{W}_{\mathcal{AL}}[i,j] = N$
  - $\mathbf{w}_{312} = \mathcal{W}_{\mathcal{AL}}[1,3] + \mathcal{W}_{\mathcal{AL}}[2,1] + \mathcal{W}_{\mathcal{AL}}[3,2].$
  - $\mathbf{w}_{314} = \mathcal{W}_{\mathcal{AL}}[1,3] + \mathcal{W}_{\mathcal{AL}}[2,1] + \mathcal{W}_{\mathcal{AL}}[3,4].$
  - Inner product:

$$\begin{aligned} \mathcal{W}_{\mathcal{AL}}[i,a] \cdot \mathcal{W}_{\mathcal{AL}}[j,b] = 0, \quad i \neq j \\ \Longrightarrow \qquad \mathbf{w}_{312} \cdot \mathbf{w}_{314} &= \mathcal{W}_{\mathcal{AL}}[1,3] \cdot \mathcal{W}_{\mathcal{AL}}[1,3] \\ &+ \mathcal{W}_{\mathcal{AL}}[2,1] \cdot \mathcal{W}_{\mathcal{AL}}[2,1] \\ &+ \mathcal{W}_{\mathcal{AL}}[3,2] \cdot \mathcal{W}_{\mathcal{AL}}[3,4] \\ &\geq N + N - N = N \end{aligned}$$

- h different symbols in a length L sequence

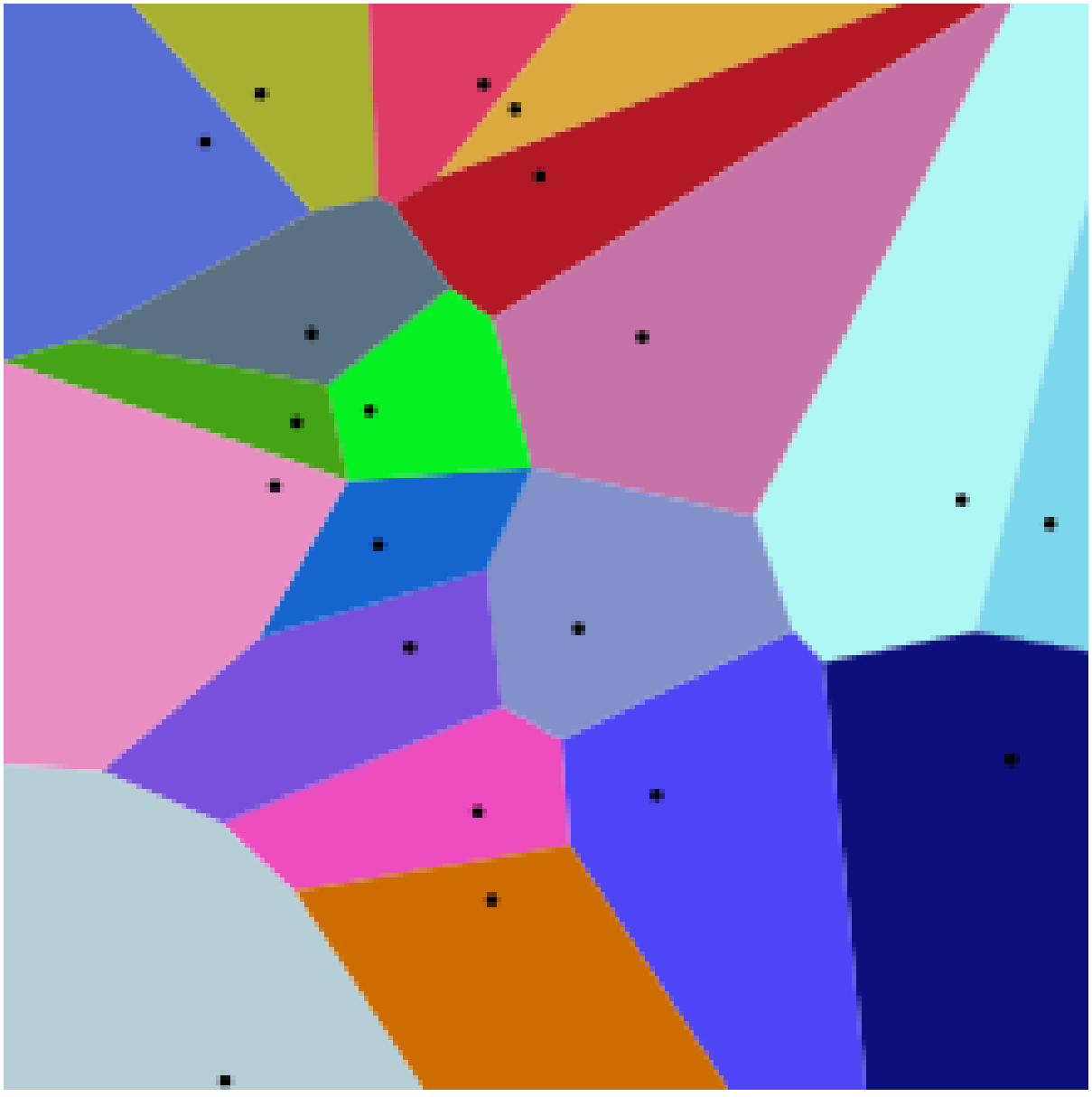
$$(L - 2h)N.$$

The Idea of Error Correction Codes

Decompose all possible sequences $\mathcal{S}$ into $\mathcal{S}_c \cup \bar{\mathcal{S}}_c$.

- $\mathcal{S}_c$: Code words
  - Messages to encode.
  - Well separate to each other.
- $\bar{\mathcal{S}}_c$: Corrupted code words
  - Polluted messages.
  - Associated with the closest code word.

$\mathcal{S}_c \cup \bar{\mathcal{S}}_c$



Error Correction Code (ECC)

To preserve the capacity

- Increase the length of sequence.
- Expand the alphabet.

Increase the Length of Sequence

Sample

- 4-bits message set $\mathcal{M}$
  - Length 4 binary sequence, 16 messages.
- 7-bits word space $\mathcal{S}$
  - Length 7 binary sequence, 128 words.
  - $|\mathcal{S}_c| = |\mathcal{M}| = 16$.
  - $a, b \in \mathcal{S}_c, a \neq b$ have at less 3 different bits.
    - Why 3? Flip one bit for each of the two.
  - Decode $s \in \mathcal{S}$: find $c \in \mathcal{S}_c$ has at most one different bit.

Question: ECC

What is the minimal length of a sequence that can be used as ECC for a length L binary sequence with h bit error tolerance.

Performance

- Without ECC
  - Length 4, 1 bit difference for different message.
  - Min inner product: $N(4 - 2 \times 1) = 2N$.
- With ECC
  - Length 7, at least 3 bit differences for different message.
  - Min inner product: $N(7 - 2 \times 3) = N$.

Expand the Alphabet

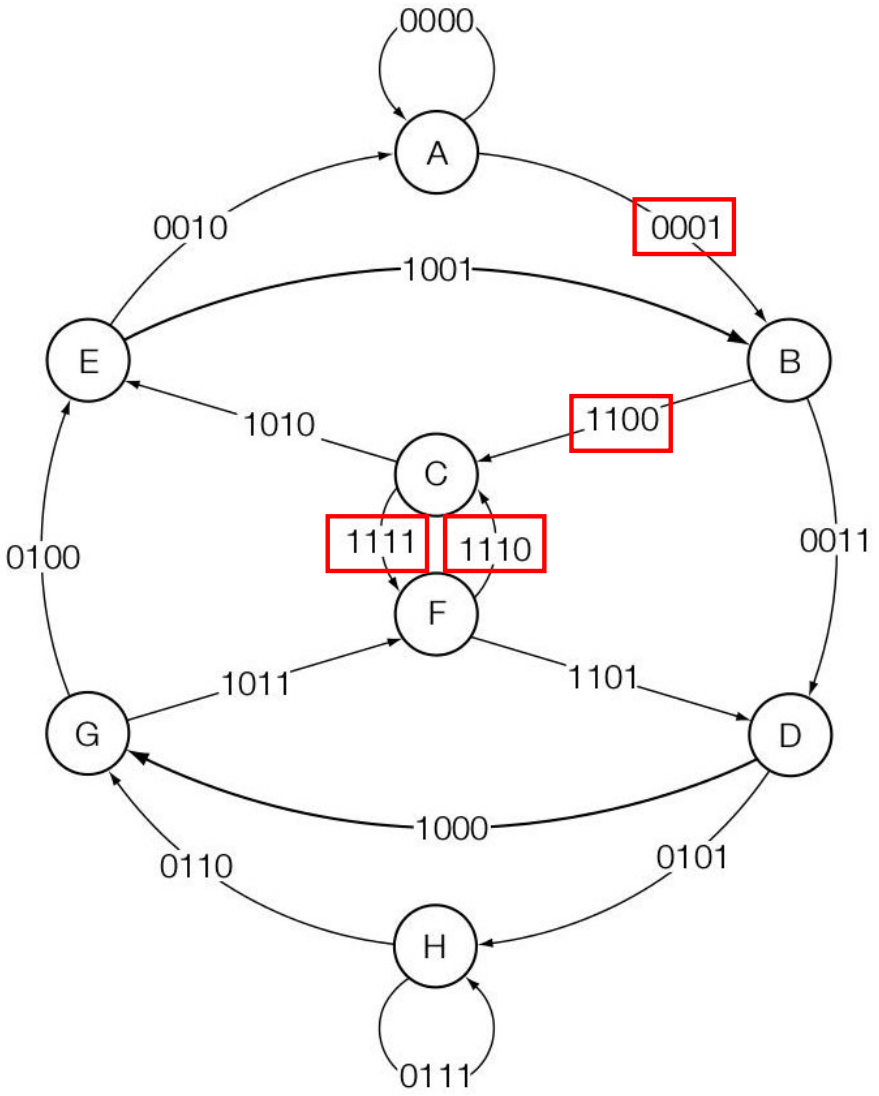
From $|\mathcal{A}| = 2$ to $|\mathcal{A}'| = 4$.

- Less typical.
- Equivalent to increase length in capacity.
- But different in modulation.

Trellis Codes

$$1010 \Rightarrow 0001\ 1100\ 1111\ 1110$$

State machine

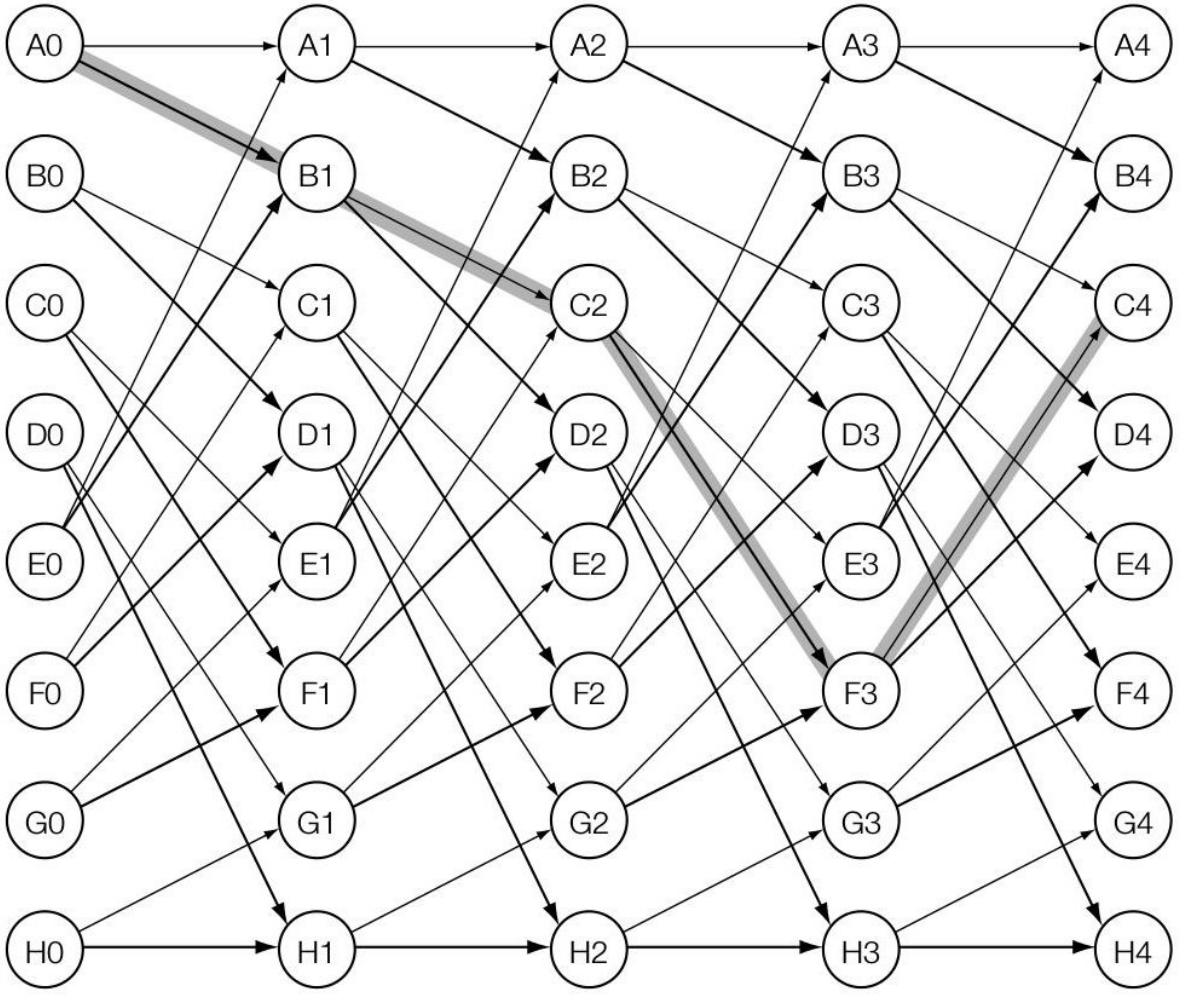


Modulation

- Increase length to $4L$.
- Expand the alphabet to contain $2^4 = 16$ symbols.

Viterbi Decoding

- A greedy method to find most closest code.
- Based on Trellis diagram.



Performance of E_TRELLIS_8/D_TRELLIS_8

The same to E_SIMPLE_8/D_SIMPLE_8:

- 8-bit message instead of 4-bit.
  - Pad two more zero at the end: 10-bit indeed.
  - More redundancy: a priory for accuracy.
- 6 integers in each of 2000 images.

Much better accuracy

- 1 out of 12000 is wrong.

4.3 Detecting Multisymbol Watermarks

False Positive

If there is no watermark

- Direct message encoding
  - The most likely one is still poor in correction.
- Multisymbol system:
  - The corrections for all the symbols are not good enough.
  - How to define “good”.

Valid Messages

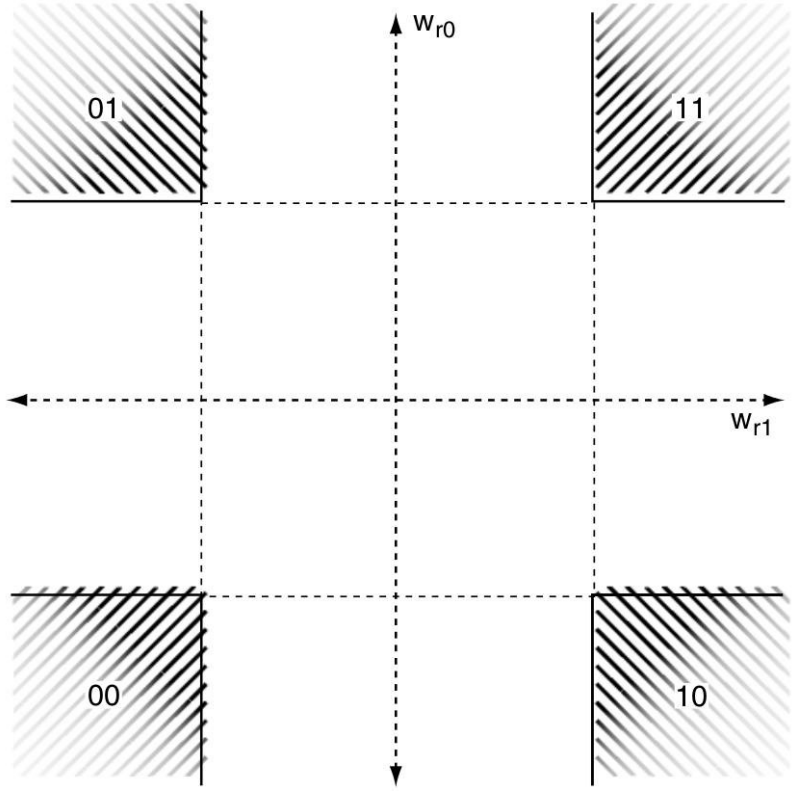
An intelligible message or a garbage.

- Checksum for verification
  - 16-bits message: m.
  - 9-bits checksum: $c = m[1:8] + m[9:16]$.
  - 25-bits watermarking: (m, c).
- Detector
  - Extractor 25-bits watermarking (m, c).
  - Compare c and $m[1:8] + m[9:16]$.
- False positive probability: $P_{fp} = \frac{1}{2^9}$.

Individual Symbols 1

All symbols are reliable (high correlated).

- Watermark presence.



2-bit system in linear correlation.

Individual Symbols 2

False positive probability

- Single reference mark: P_{fp0}.
- In each index/position/order
  - If one mark in $\mathcal{A}$

$$P_{fp1} \approx |\mathcal{A}| P_{fp0}.$$

- For the whole length L sequence.
  - All of them is high

$$P_{fp} = (P_{fp1})^L \approx (|\mathcal{A}| P_{fp0})^L.$$

Normalized Correlation 1

- Multiple-symbol embedding
  - $\mathbf{w}_{ri}$ orthogonal to each other and unit.

$$\mathbf{v}_L = \mathbf{v}_o + \sum_{i=1}^{L} \mathbf{w}_{ri}, \quad \|\mathbf{v}_L\| \approx \sqrt{L}.$$

  - Linear correlation: independent of L

$$z_{lc}(\mathbf{v}_L, \mathbf{w}_{r1}) = \mathbf{v}_o \cdot \mathbf{w}_{r1} + \mathbf{w}_{r1} \cdot \mathbf{w}_{r1} = \varepsilon + 1.$$

  - Normalized correlation: difficult for larger L

$$z_{nc}(\mathbf{v}_L, \mathbf{w}_{r1}) = \frac{\mathbf{v}_L}{\|\mathbf{v}_L\|} \cdot \mathbf{w}_{r1} = \frac{\varepsilon + 1}{\sqrt{L}}.$$

Normalized Correlation 2

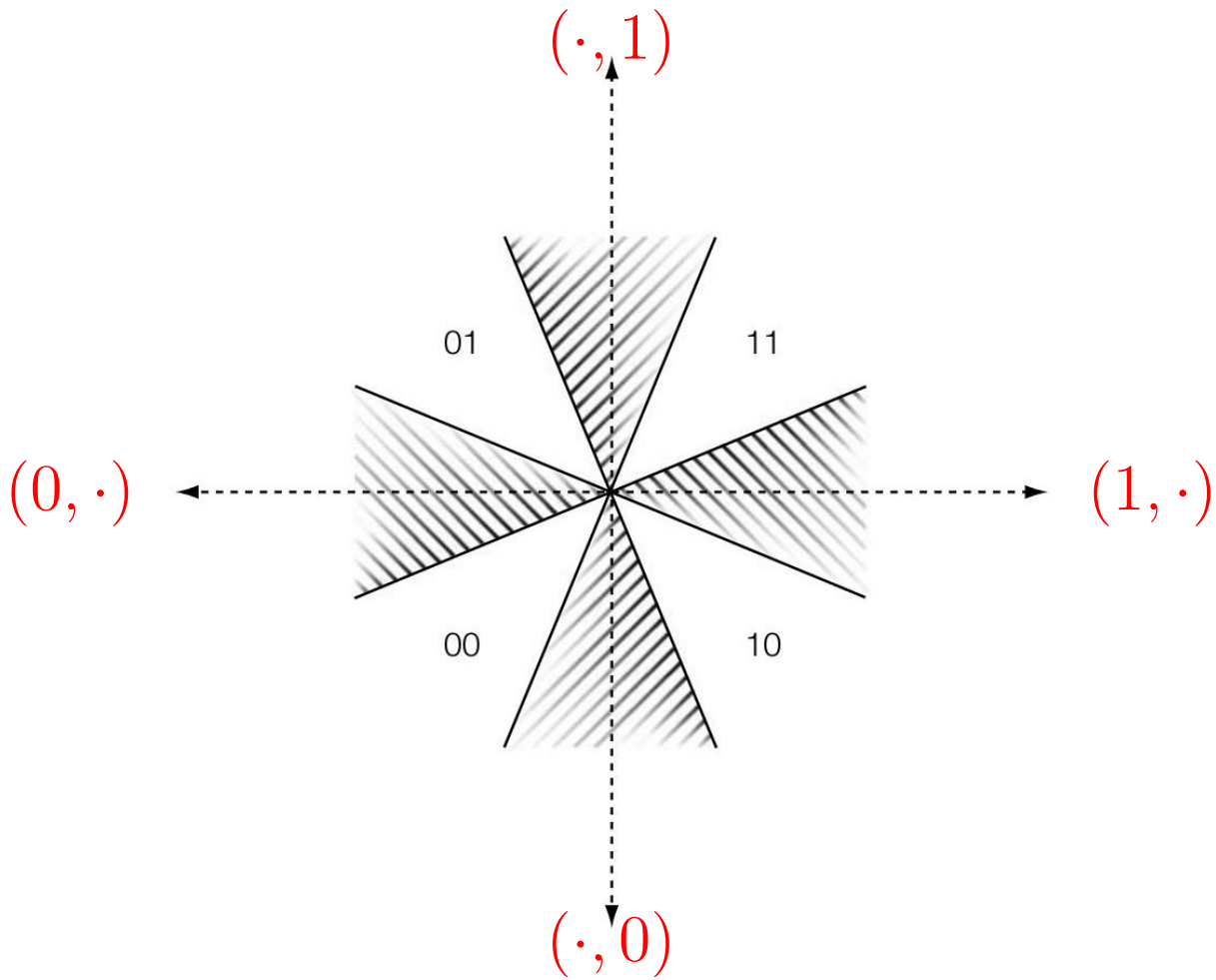
Less distinguishable.

- Large threshold: none is correlated enough, no symbol found.
- Small threshold: High false positive probability.

Geometric Interpretation

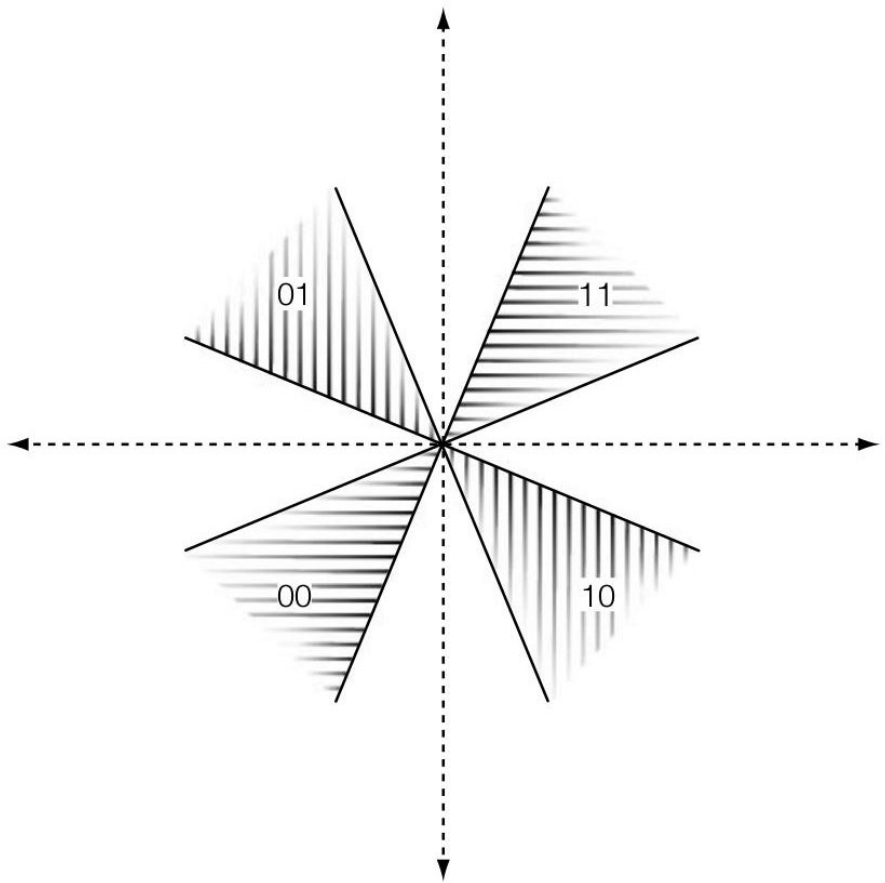
Large threshold: no overlap for the cones.

- No detectable 2-bit message.



Reencode

1. Extract message m.
2. Reencode m into mark $\mathbf{v}_m$.
3. Test the presence of $\mathbf{v}_m$



False Positive Probability

When the detection regions for the different messages do not overlap,

$$P_{fp} = |\mathcal{M}| P_{fp0}.$$

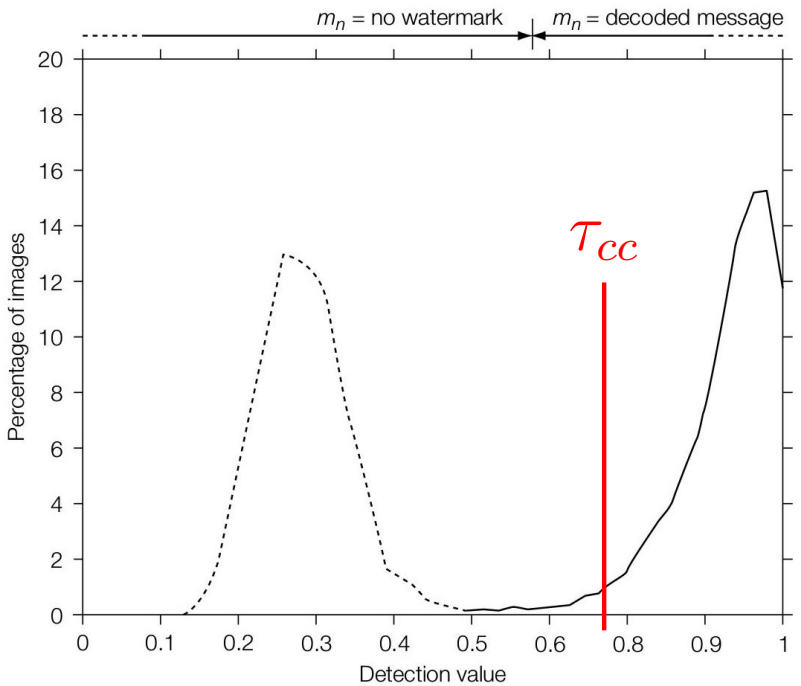
E_BLK_8/D_BLK_8

8-bit message:

- Trellis code with two padding 0 at the end.
  - A sequence of 10 symbols drawn from a 16-symbol alphabet.
- Reference marks:
  - 8×8 (block): low dimensional mark space.
    - So choose seed to reduce max correlation (0.73).
- Embedding strength $\alpha = 2$.
- $\tau_{cc} = 0.65$: false positive probability 10^{-6}.

Performance

- 2000 unwatermarked images (dashed line).
  - No false positive found.
- 12000 unwatermarked images (solid line).
  - 6 messages $\times 2000$ images.
  - 109 fail: effectiveness 99%.



Project: System 6

E_BLK_8/D_BLK_8

- marking space: 8×8 block
- 8-bit message.
- ECC: hamming or optional.
- Reencode check.

Presentation: 7.6 Analysis of Normalized Correlation

Approximate Gaussian Method

- False Positive Analysis
- False Negative Analysis