Stanford University, Palo Alto, CA, USA
January 2002

Information Embedding and Digital Watermarking

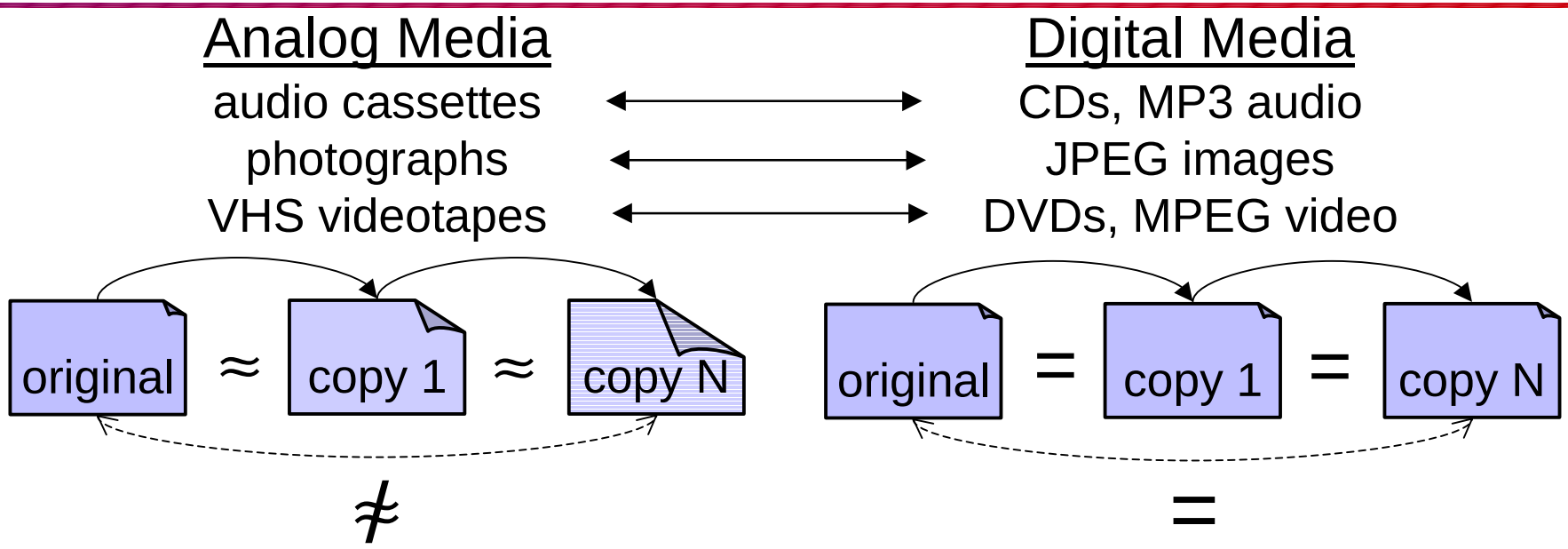
J.J. Eggers

Telecommunications Lab
Univ. of Erlangen-Nuremberg
eggers@LNT.de

Overview

- General concepts and state-of-the-art
- Scalar Costa scheme
- The game between embedder and attacker
- Example application:
  - Payload for SCS watermarks in image data
  - Image integrity verification

Analog and Digital Multimedia

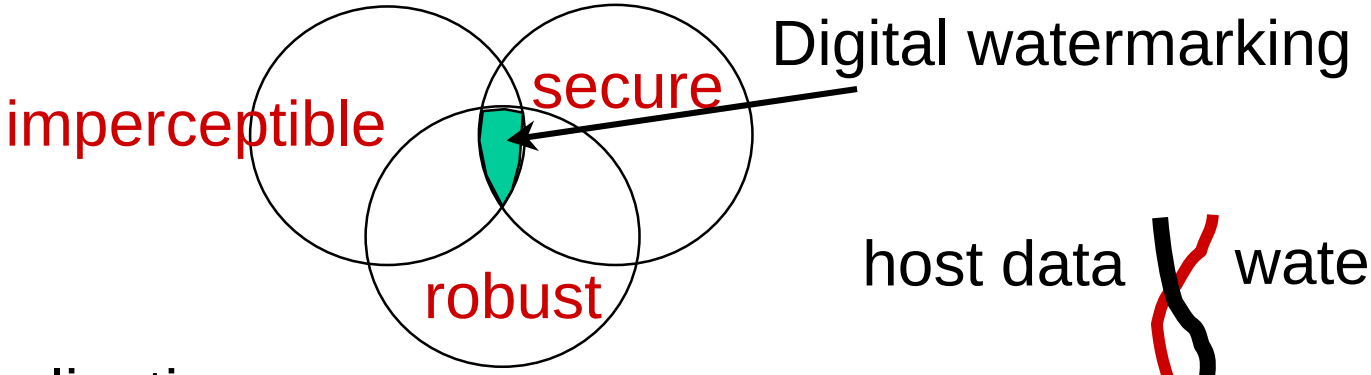


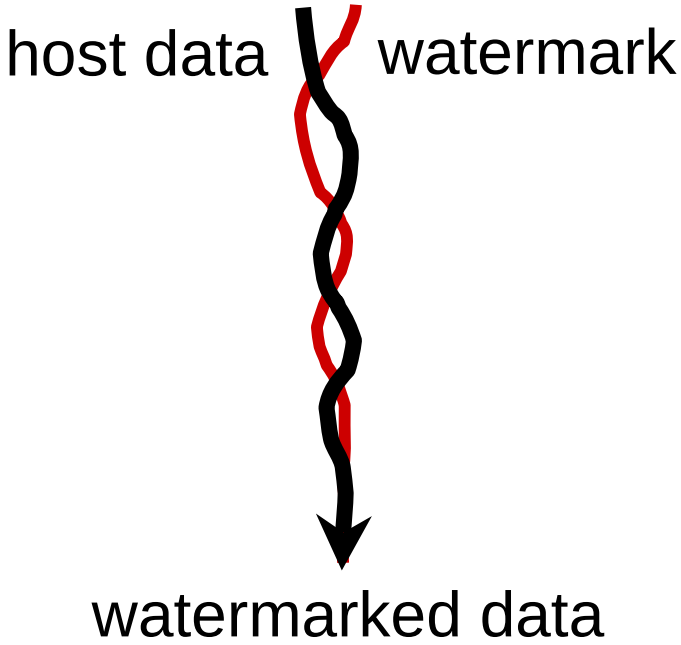
- Distribution net required
- Difficult to edit
- "Built-in" protection against copying, redistribution, editing

- **"Free" distribution net**: Internet
- Simple editing
- **No inherent protection** against copying, redistribution, editing

Digital Watermarking

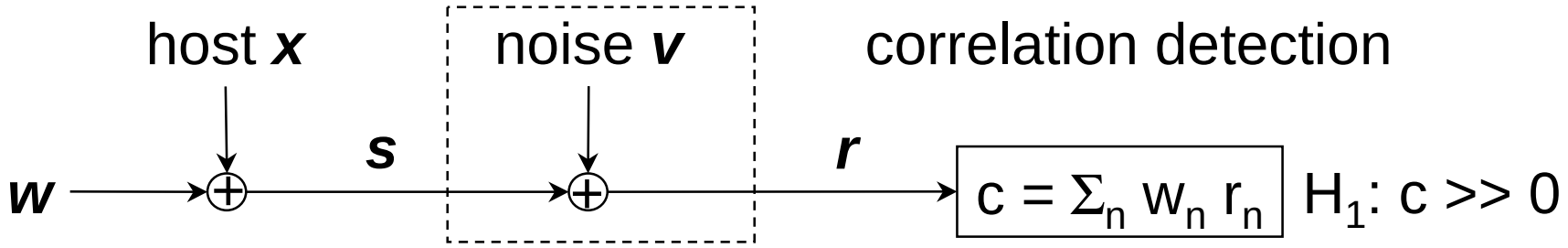
- Information embedding that is





- Applications:
  - copyright protection
  - data integrity verification
  - broadcast control
  - distribution tracing
  - ....

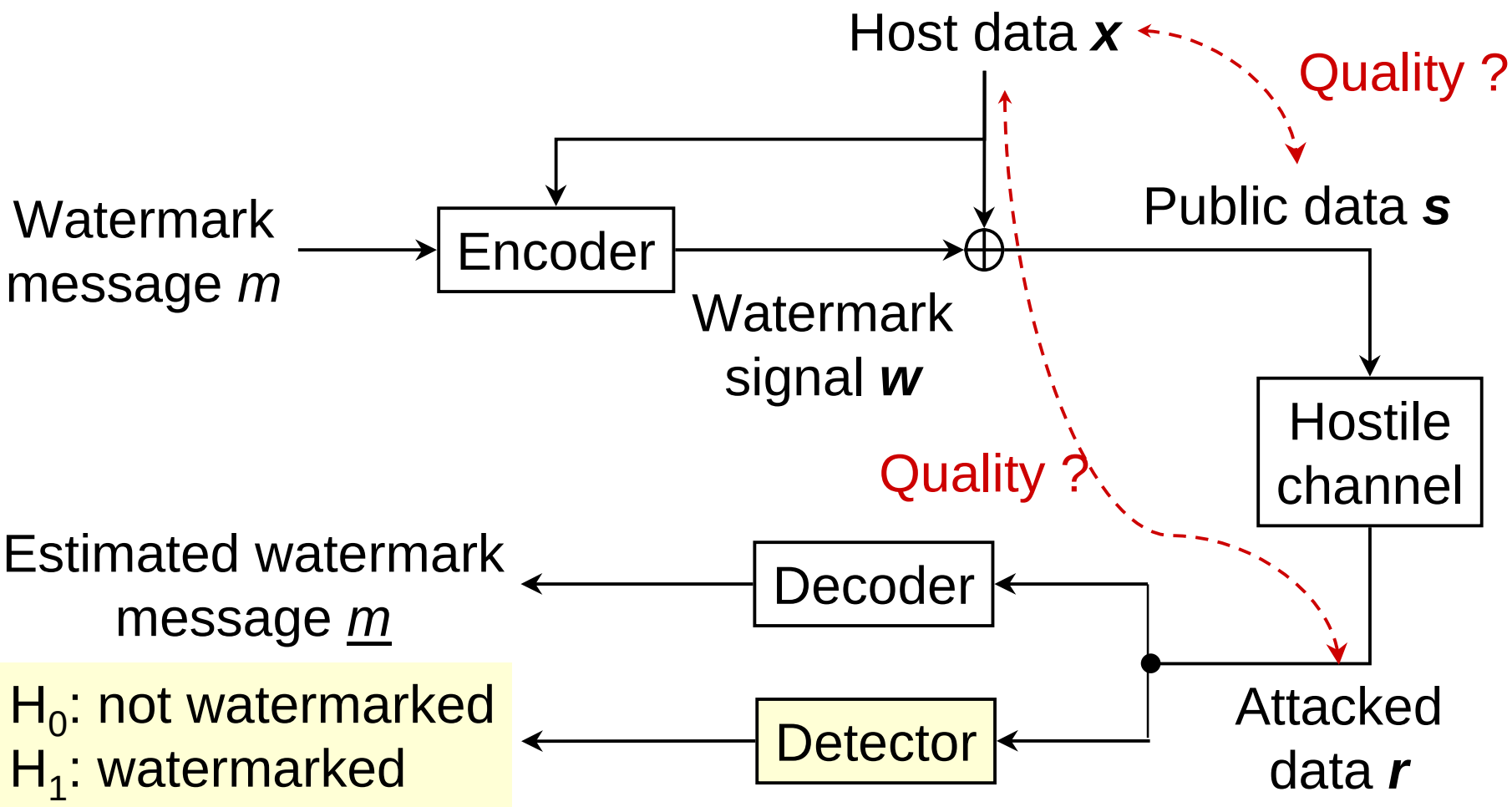
Spread-Spectrum Watermarking



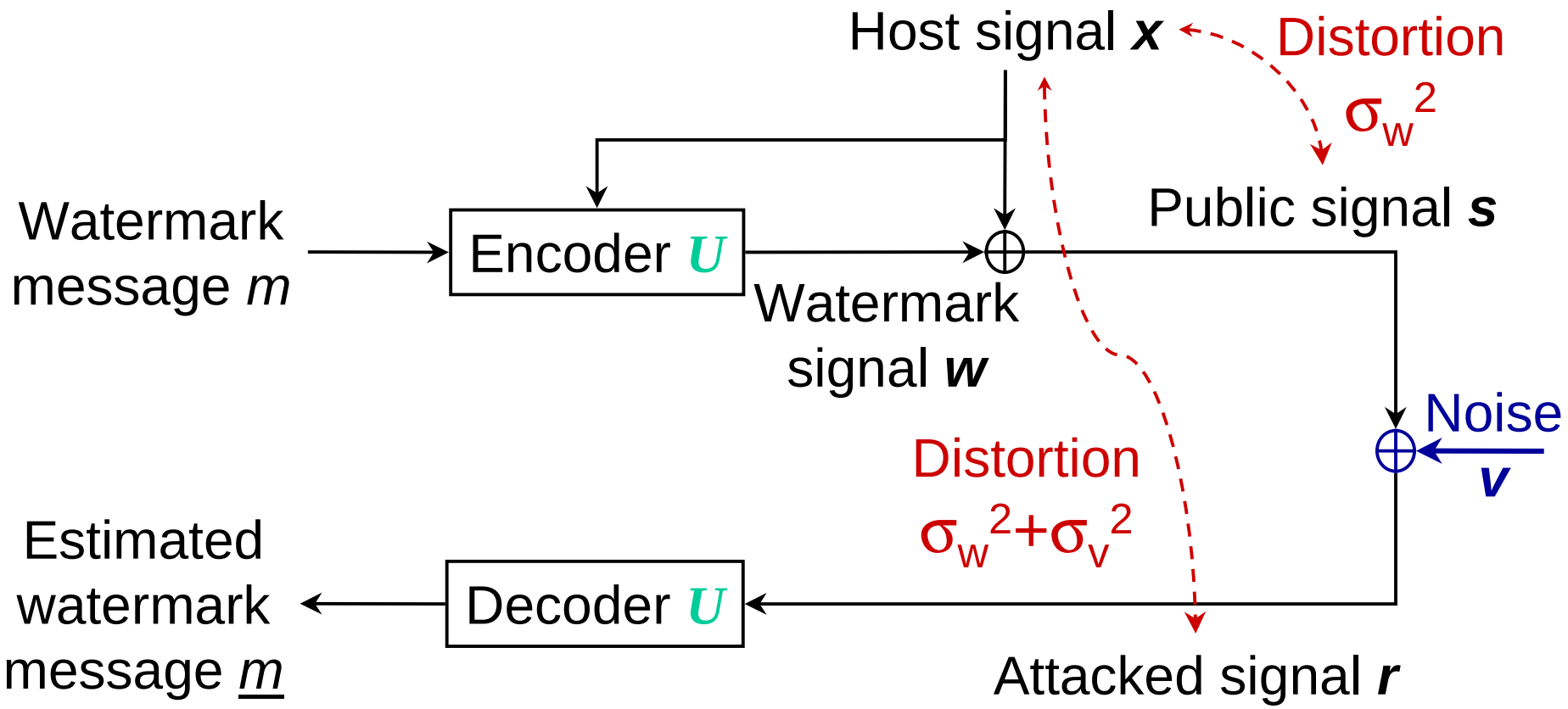
– Side information about the host data is not exploited!

- Properties
  – pseudo-noise sequence ***w*** = secret key
  – correlation detection is very reliable for long signals
  – host signal is dominating interference source

Model for Blind Watermarking



IID Host Signals & AWGN Attack

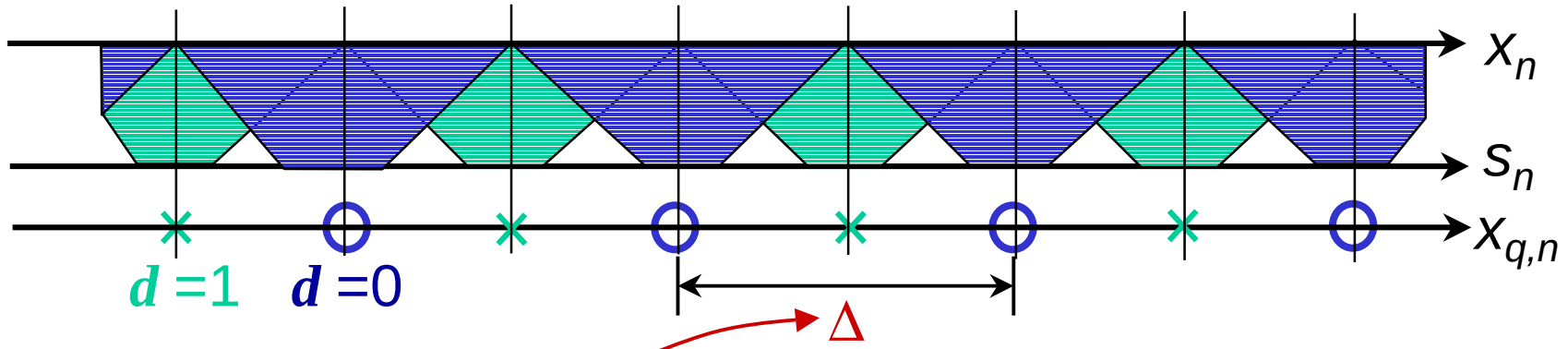


Costa's Scheme

- Costa, 83: "Writing on Dirty Paper"
- Analysis of communication with side information:
  - IID Gaussian noise
  - IID Gaussian host signal
- Information theoretic result:
  - Watermark capacity is independent of host signal!
- Costa's Scheme
  - is not practical
  - gives insights into the problem of communication with side information

Scalar Costa Scheme (SCS): U = uniform scalar quantizer

- Encode message $m = d_1 d_2 \cdots d_N$ & embed in $\boldsymbol{x}=x_1 x_2 \cdots x_N$
- Example: embed $d_n \in \{0,1\}$ (binary SCS)



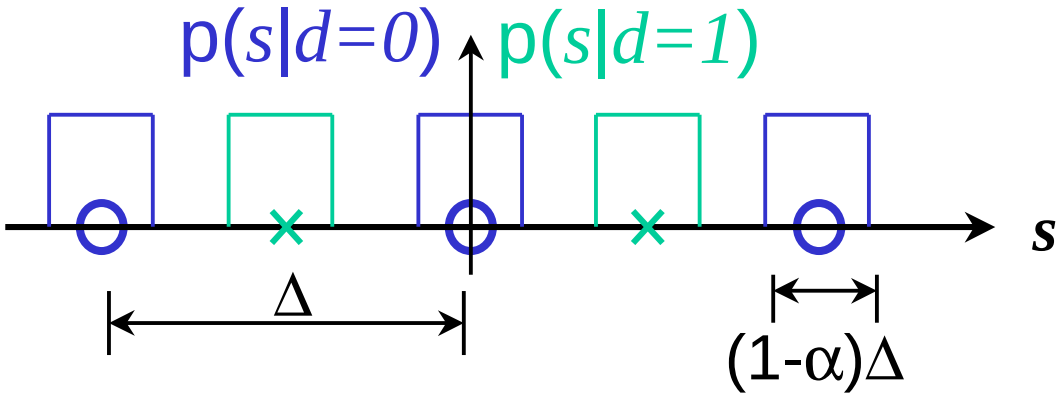
- Embedding rule

$$s_n = x_n + \alpha(x_{q,n} - x_n)$$

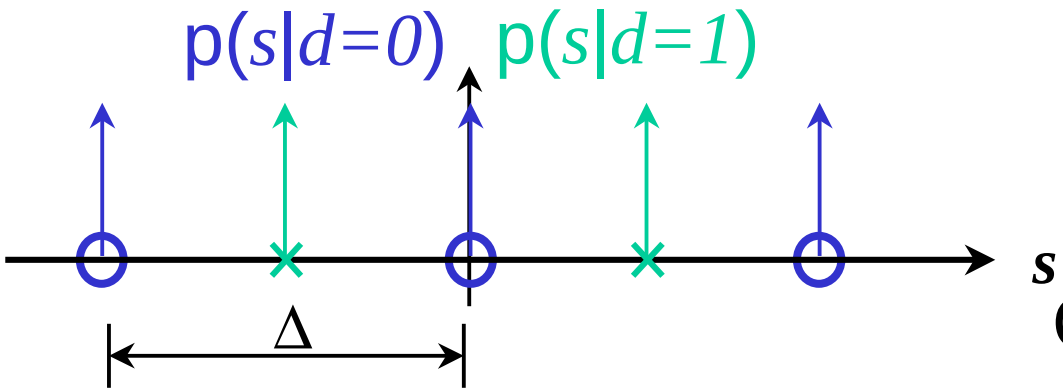
$0 \le \alpha \le 1$

Trade off embedding distortion versus robustness!

PDF of Public Signal s

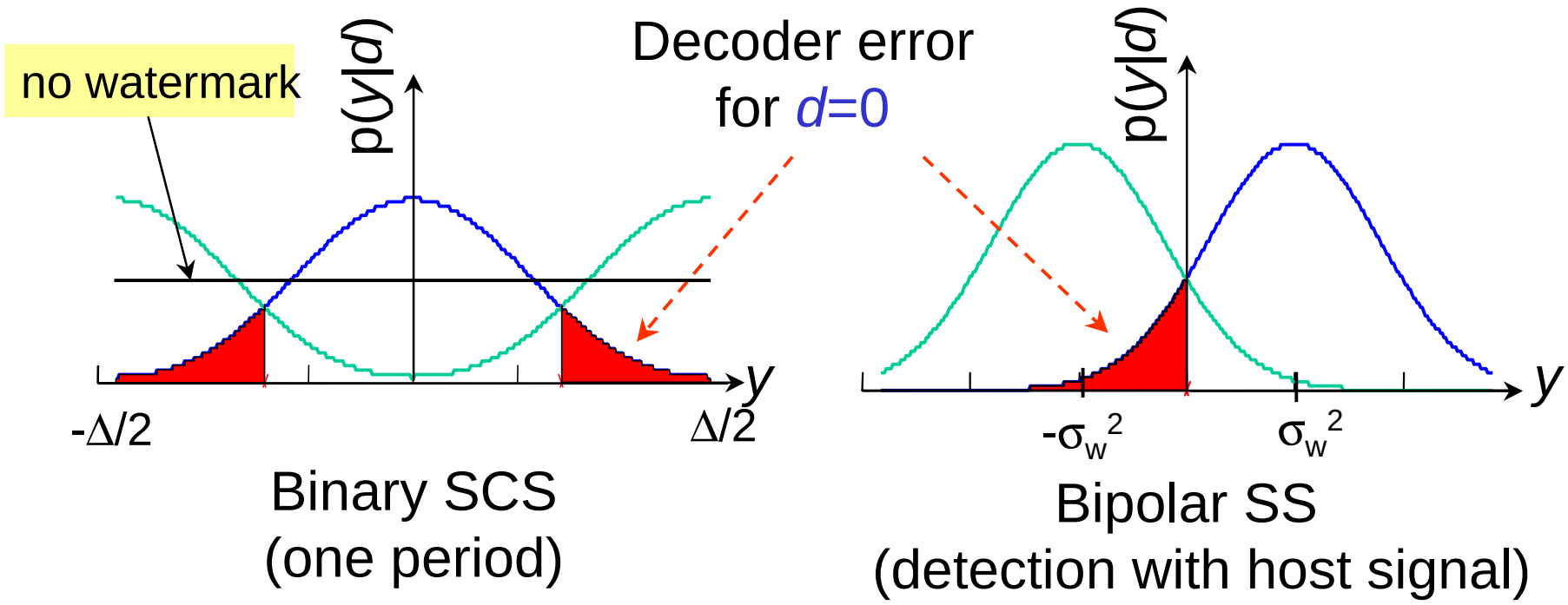


Binary SCS



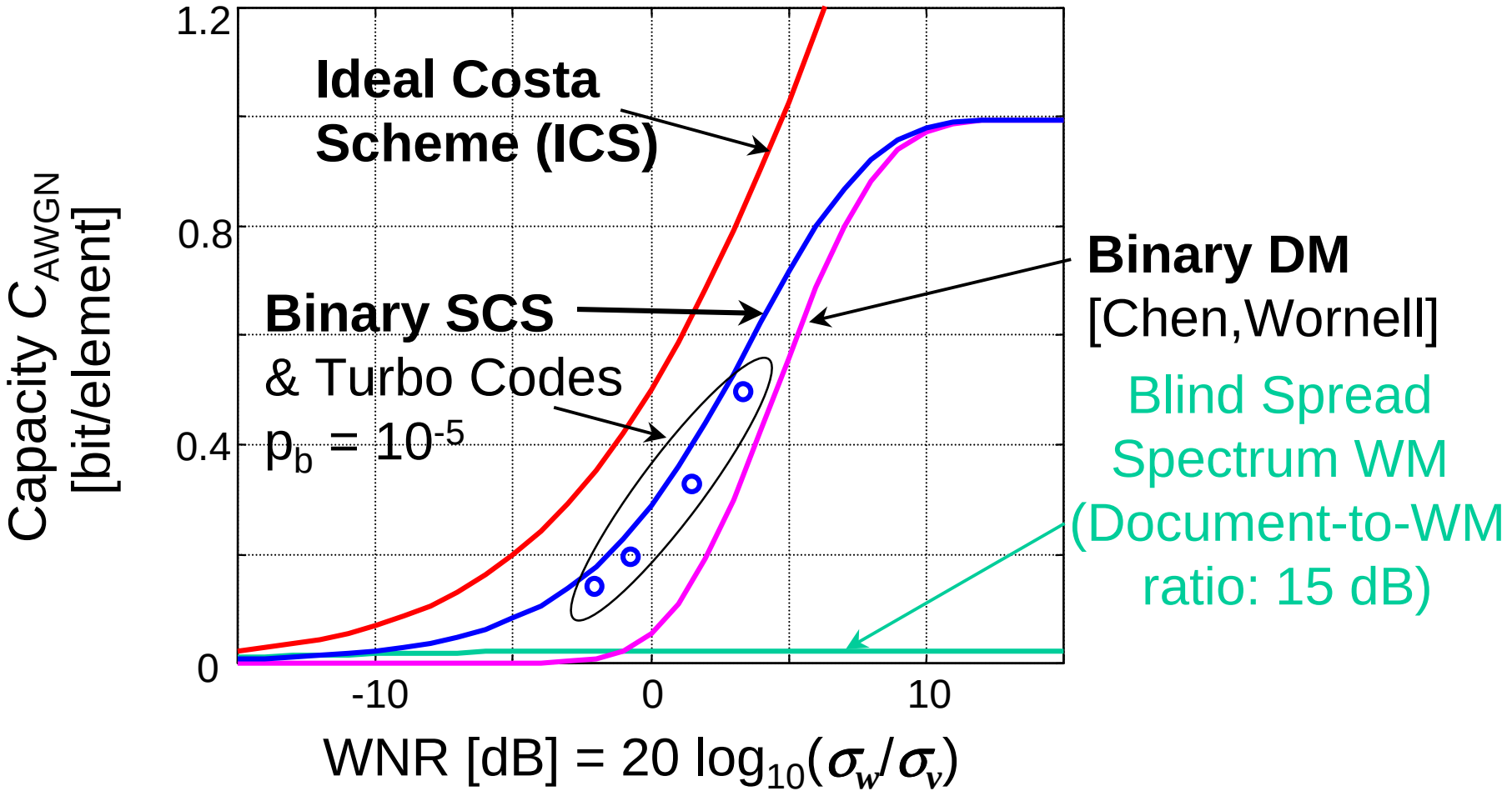
Dither Modulation (DM)
Chen & Wornell (´98)

PDF of Extracted Signal *y*



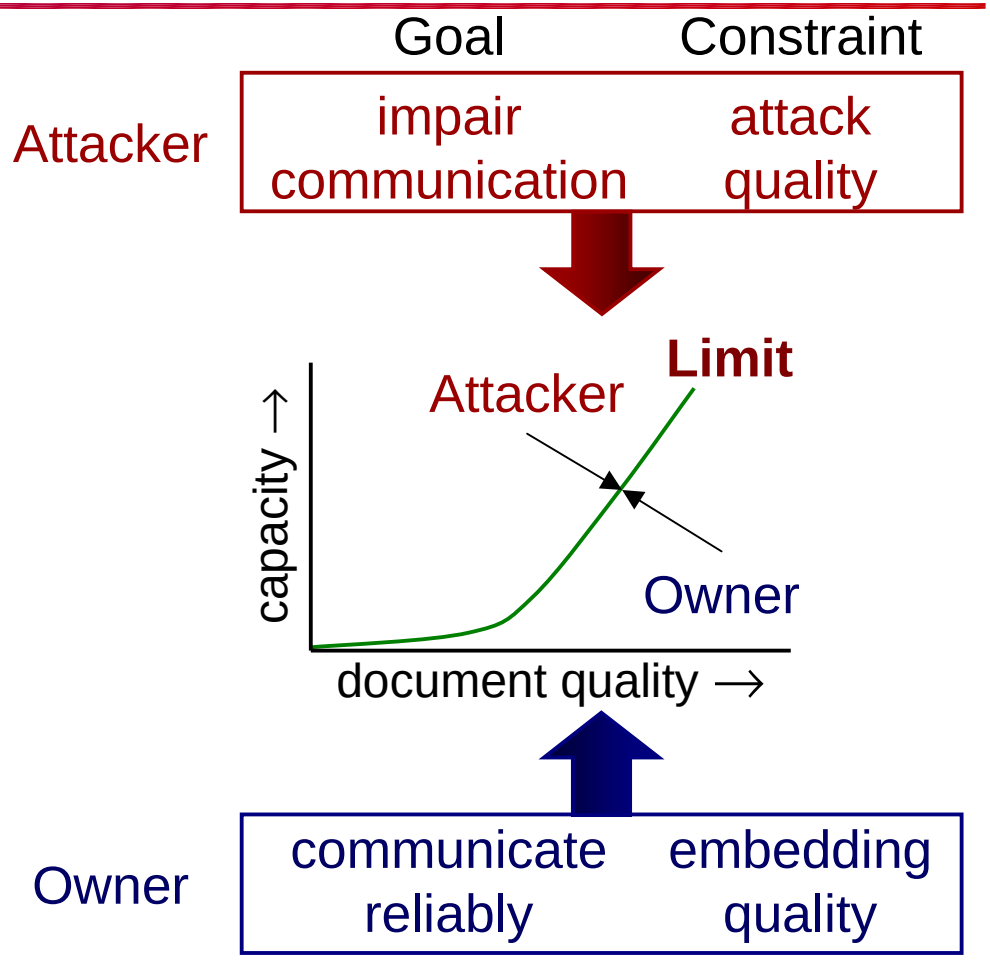
SCS: $p(y|d=0)$ and $p(y|d=1)$ computed numerically

Blind Watermarking Capacity

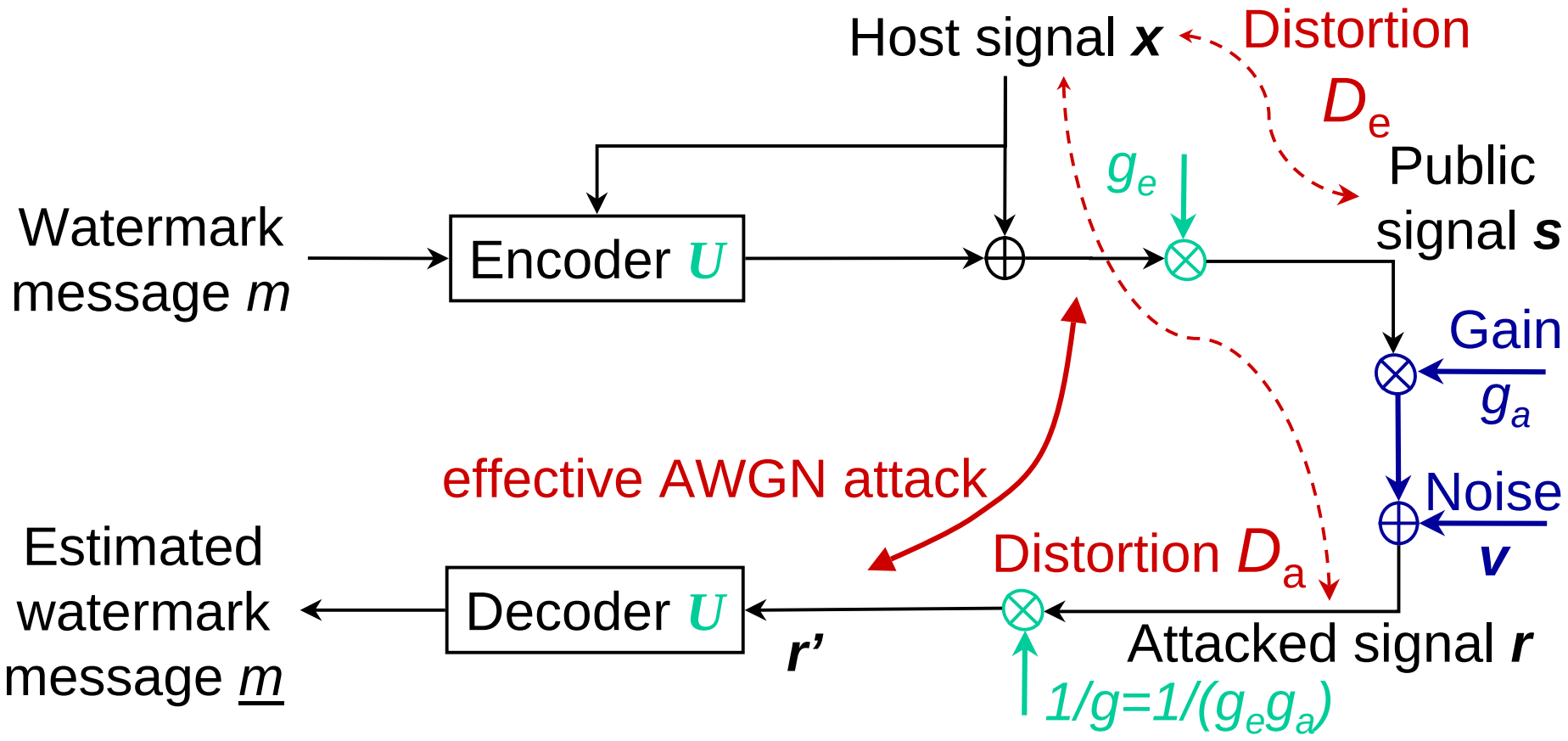


Watermarking as a Game

- Game between embedder and attacker
  - payoff: communication rate
  - penalty: quality loss after watermark embedding and after attack
- Information-theoretic limits
  - robustness well-defined
  - provable watermarking guidelines



Effective AWGN Channel Model for Scaling & AWGN (SAWGN)

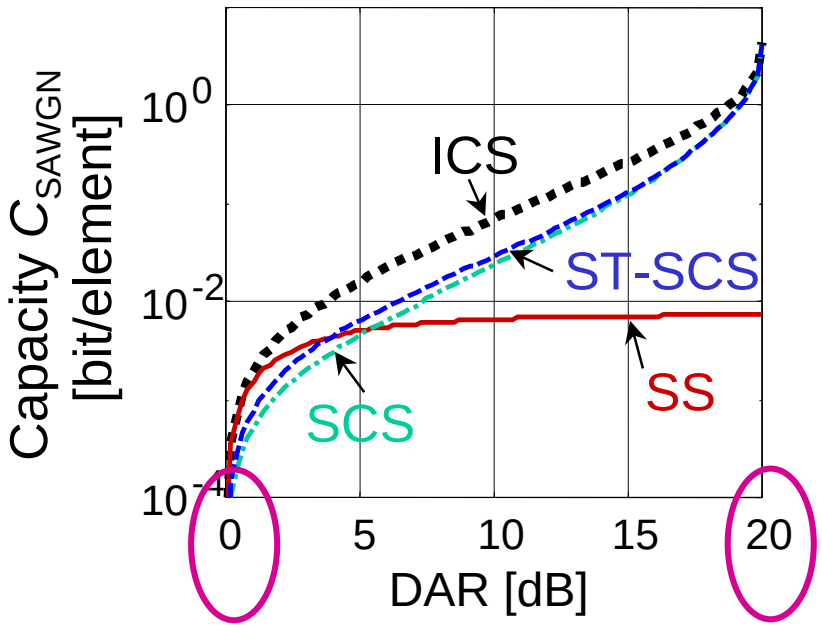


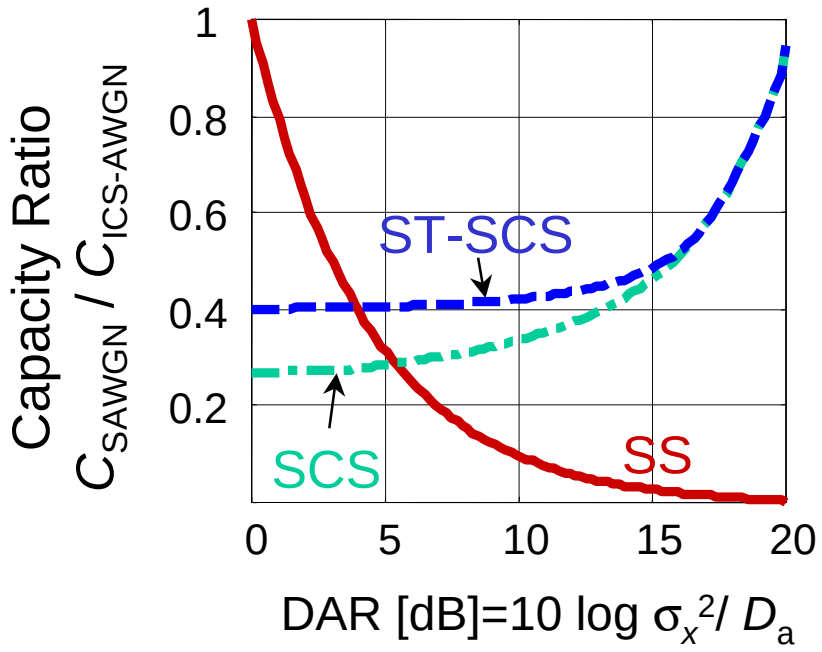
Design U for effective channel noise $\boldsymbol{v' = v/g}$!

Watermark Capacity after SAWGN attack

DWR = Document-to-WM-Power Ratio ~ Quality after embedding

DAR = Document-to-Attack-Power Ratio ~ Quality after attack



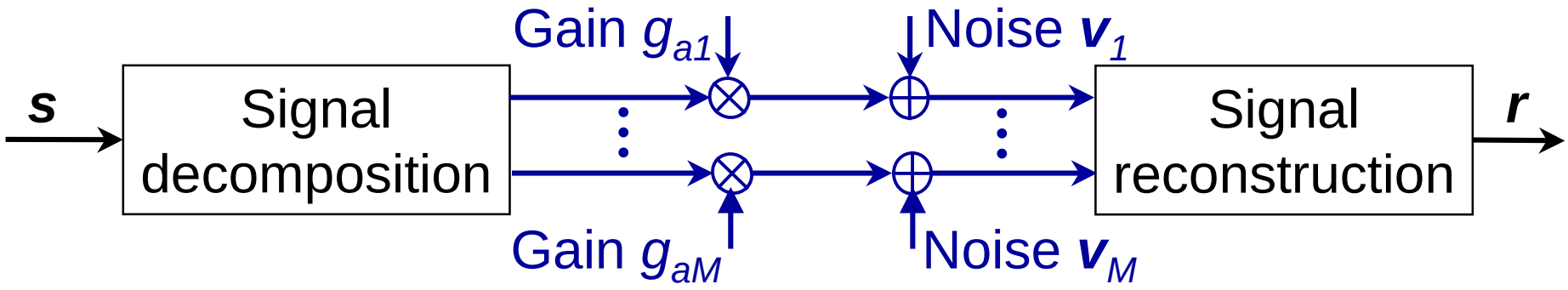


strongest attack

no attack = DWR [dB] = 10 log σ_x^2/D_e

Non-IID Host Signals
Linear Filtering & Additive Noise

- Decompose host signal
  - *M* approximately independent sub-channels
  - white signal statistics within sub-channel
- Linear filtering & additive "colored" noise (FACGN) attack



- Watermark communication over parallel channels

Optimum Allocation of Embedding and Attack Distortion (I)

- Constraints
  - total embedding distortion

$$D_{Embedding} \geq \sum_j rate_j \times weight_j \times emb-distortion_j$$

e_j

  - total attack distortion

$$D_{Attack} \geq \sum_j rate_j \times weight_j \times attack-distortion_j$$

a_j

- Objective function

$$C_{FACGN} = \max_{\{e_j\}} \min_{\{a_j\}} \sum_j rate_j \times C_{SAWGN,j}(host-power_j, e_j, a_j)$$

Optimum Allocation of Embedding and Attack Distortion (II)

- **No unique solution over entire distortion range!**
- Low distortion: **white**
  - attack ~ “add noise”
  - force attack to spread its power over all channels
- High distortion: **PSC**
  - Power-Spectrum-Cond.
  - attack ~ “throw away”
  - attack cannot discard watermark without also destroying original

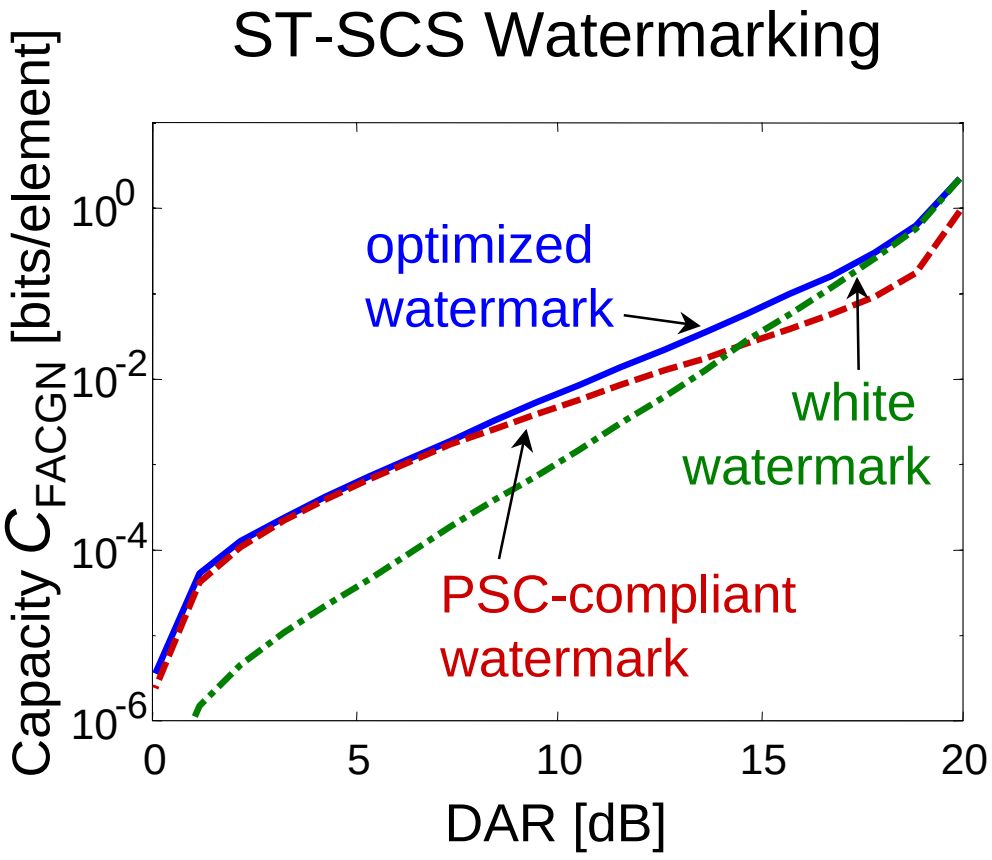
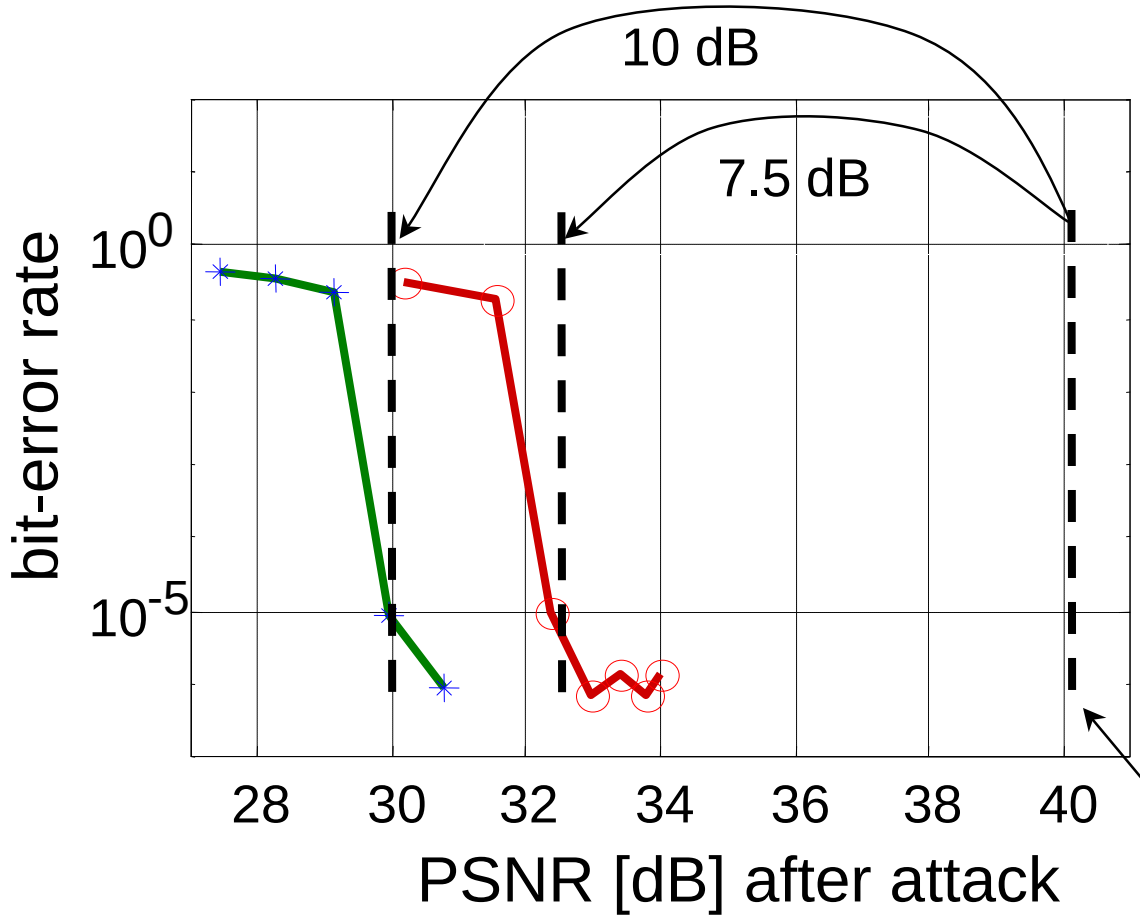


Image Watermark Payload



- Test image “Lenna“ grayscale 512x512
- 912 bits & linear filtering and colored noise attack
- 645 bits & JPEG compression attack

Image Integrity Verification (I)

What is the original? Where are the differences?

- Watermark entire image
- Local detection
  - H_0: no watermark = modified content
  - H_1: watermark = no content modification

detect from all elements r_i in small image region

$$\frac{\Pi_i \; p(r_i \mid H_1)}{\Pi_i \; p(r_i \mid H_1) + \Pi_i \; p(r_i \mid H_0)}$$

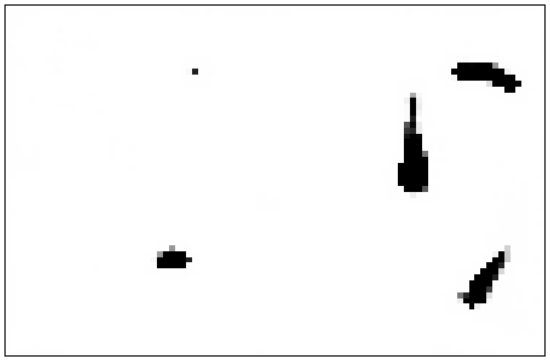
$< 0.5 : H_0$

$> 0.5 : H_1$

Image Integrity Verification (II)

SCS watermarked

manipulated and
JPEG compressed (Q=70)



detected non-authentic regions

Detection with sliding window of size 32x32.

Correct detection of manipulated image regions.

Detection error in flat image region due to compression.

Image Integrity Verification (III)

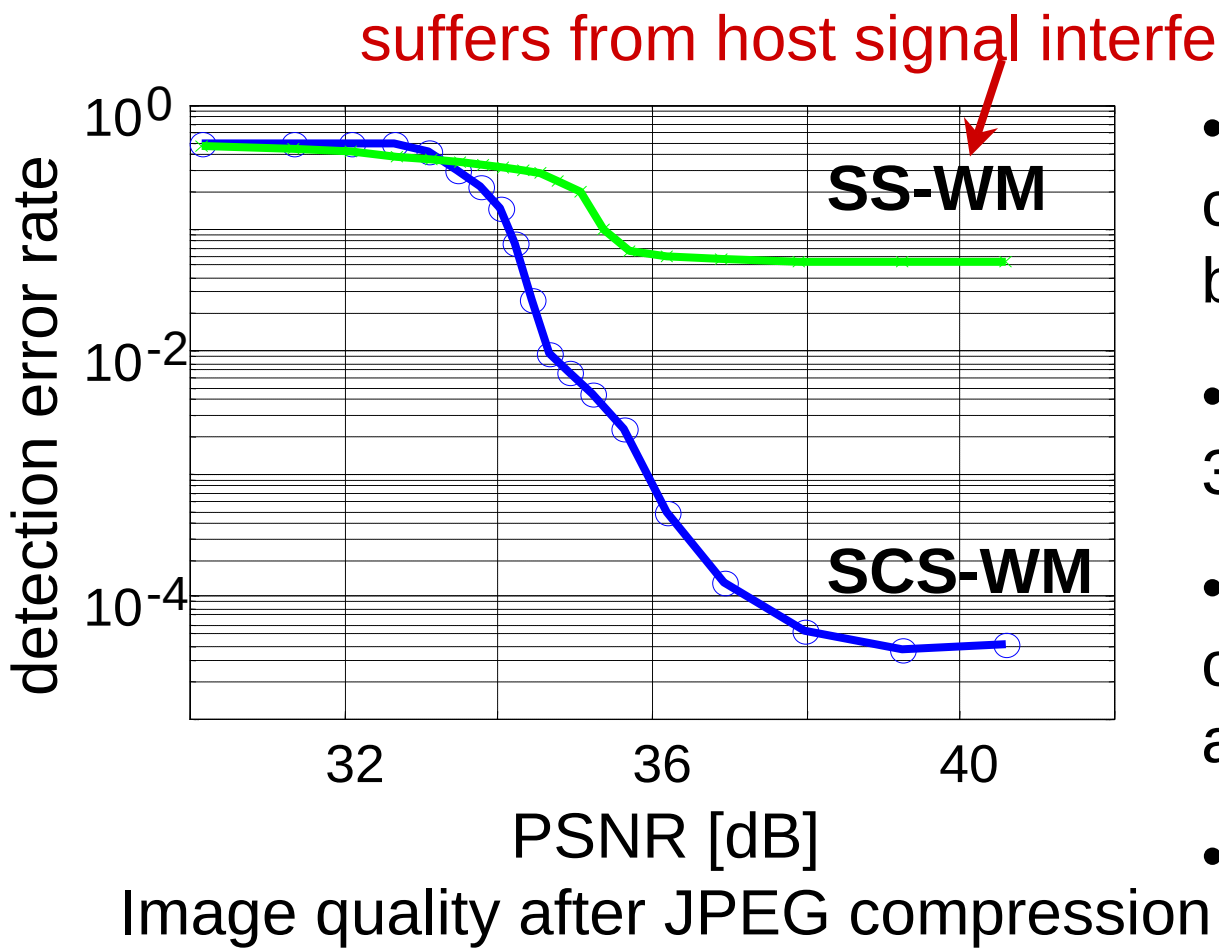


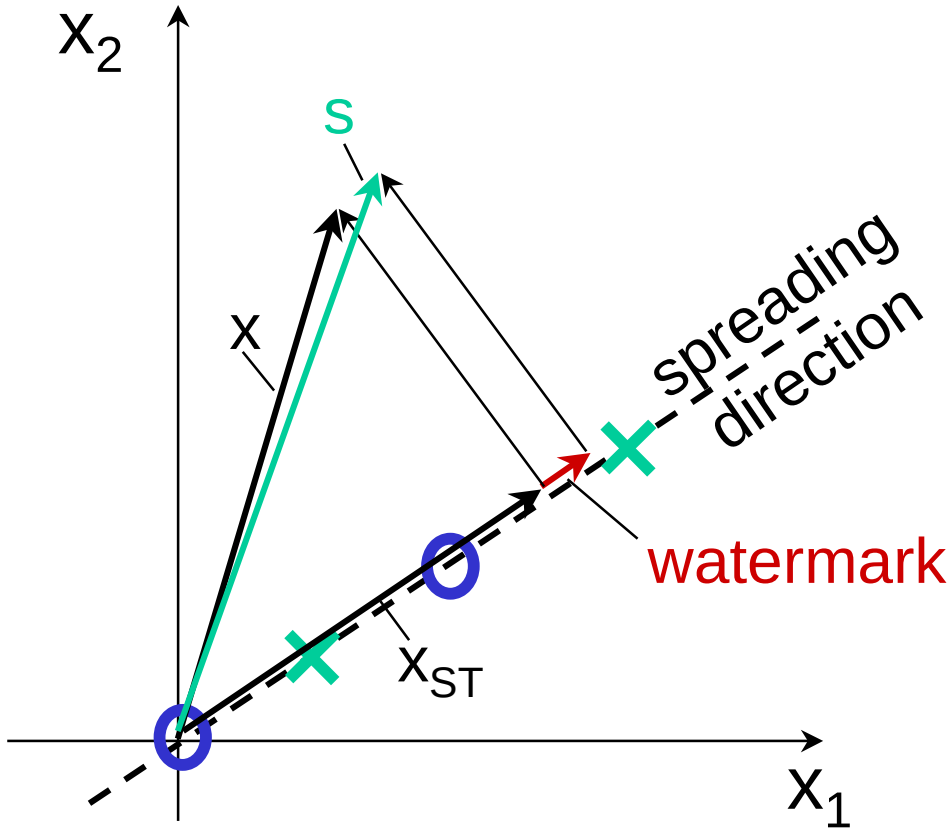
Image quality after JPEG compression

- Embedding into coefficients of 8x8 block DCT
- Detect from 32x32 pixel blocks
- Measure average of false positive and false negative
- Test image “Girl“

Summary

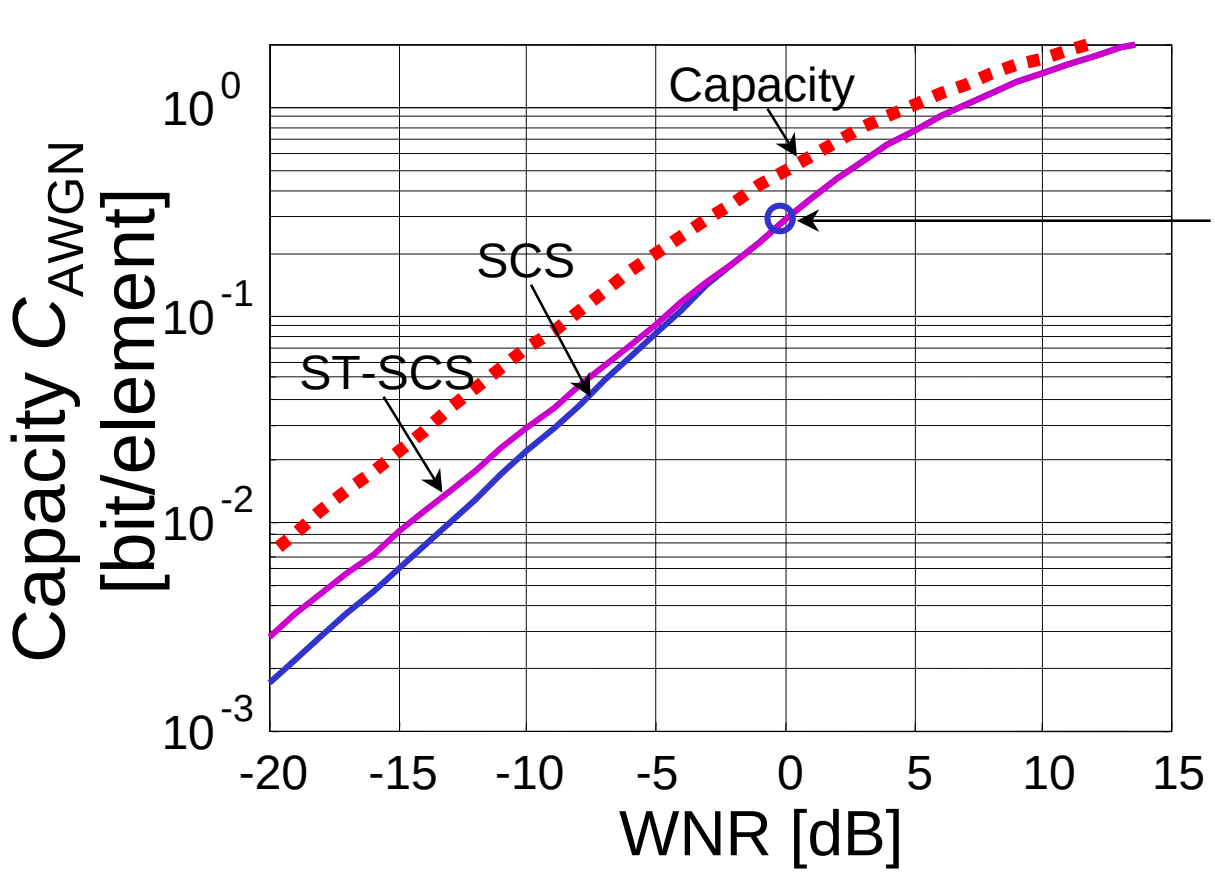
- Blind watermarking
  - original data is useful side information
  - Scalar Costa Scheme (SCS): practical & performs close to capacity limits
- Analysis of watermarking via game theory
- Some open problems
  - efficient synchronization algorithms
  - robustness dependent on host PDF

Spread-Transform Watermarking



- Example: spreading factor $\nu=2$
- x - host signal
- s - public signal
- Noise orthogonal to the spreading direction does not affect the detection performance
- 3 dB gain for $\nu=2$

Capacity of ST-Watermarking



- ST-Watermarking is useful for $WNR<WNR_{crit}$
- $WNR_{crit,SCS}$= 0.01dB
- SCS requires lower spread-transform length and achieves higher rates than comparable schemes at the same WNR!