Active Content Fingerprinting: a marriage of passive content fingerprinting and digital watermarking

S. Voloshynovskiy, F. Farhadzadeh, O. Koval, T. Holotyak

Stochastic Information Processing Group,
University of Geneva,
Switzerland

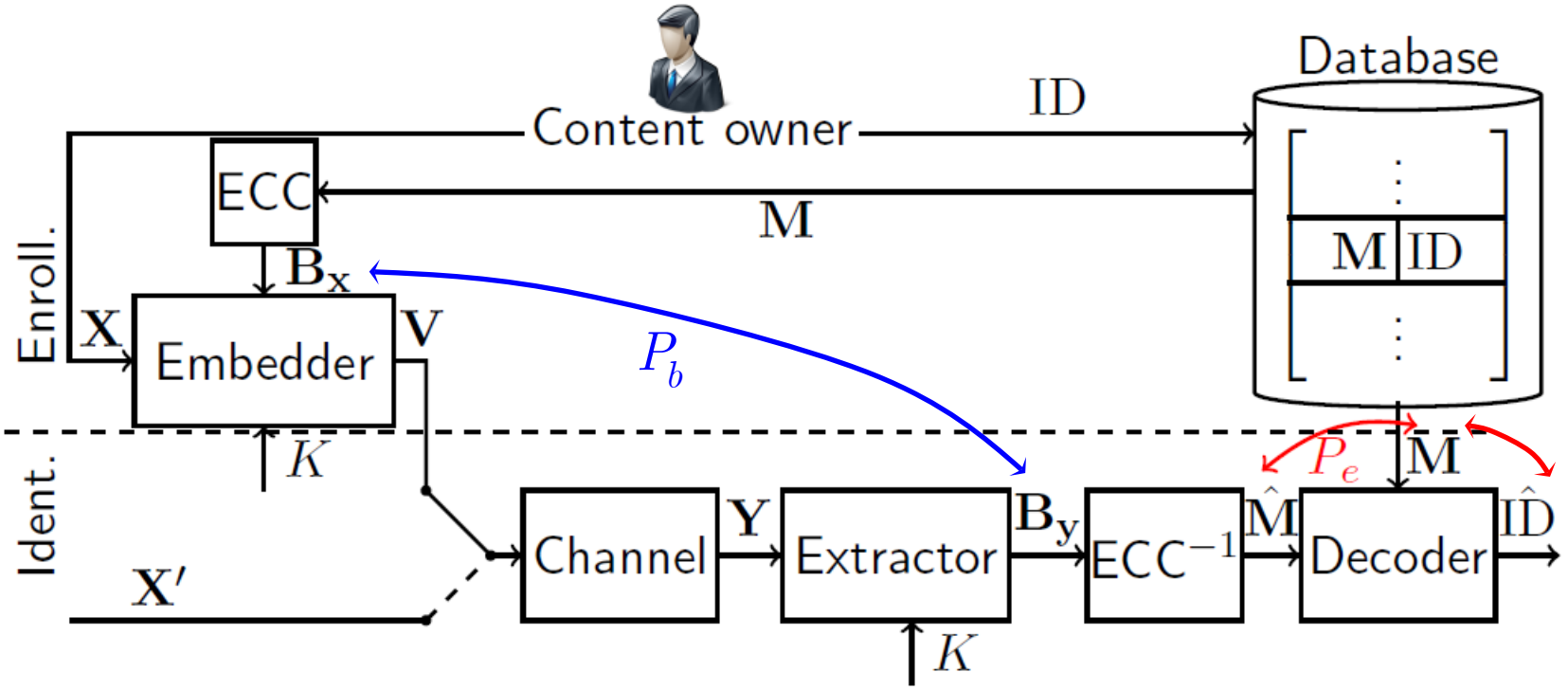
WIFS'12, Tenerife, Spain, December 2-5, 2012

Outline

1. Introduction: Digital watermarking and passive content fingerprinting
2. Active content fingerprinting
3. Simulation results
4. Conclusions

Traditional approaches: Digital watermarking

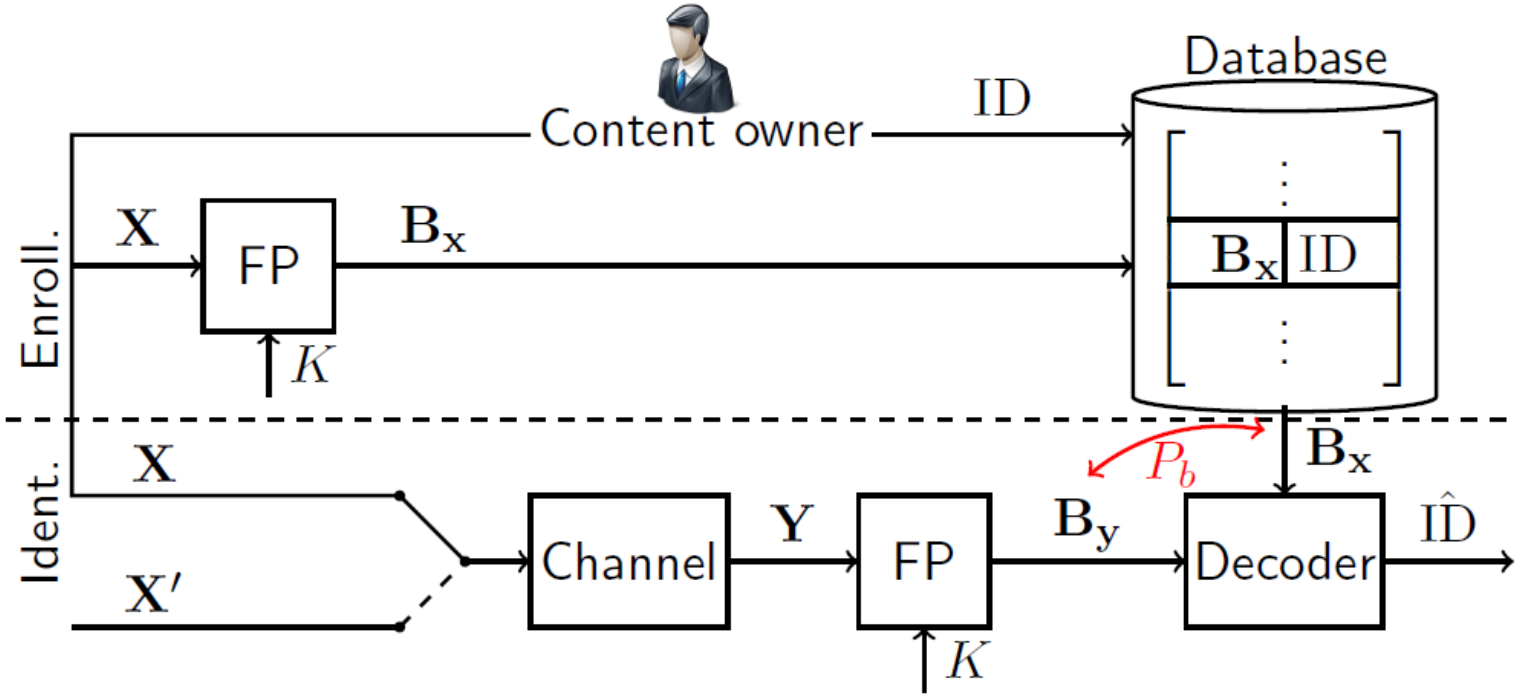
Digital Watermarking (DWM):



- embed host-independent message: $\mathrm{ID} \rightarrow \mathrm{M}$
- host interference
- distortions due to watermark embedding and host interference cancellation
- structured codes ($P_e \approx 0$) $\Rightarrow$ low complexity $\overset{\text{BUT}}{\Longrightarrow}$ security concerns

Traditional approaches: Passive content fingerprinting

Passive Content Fingerprinting (pCFP):



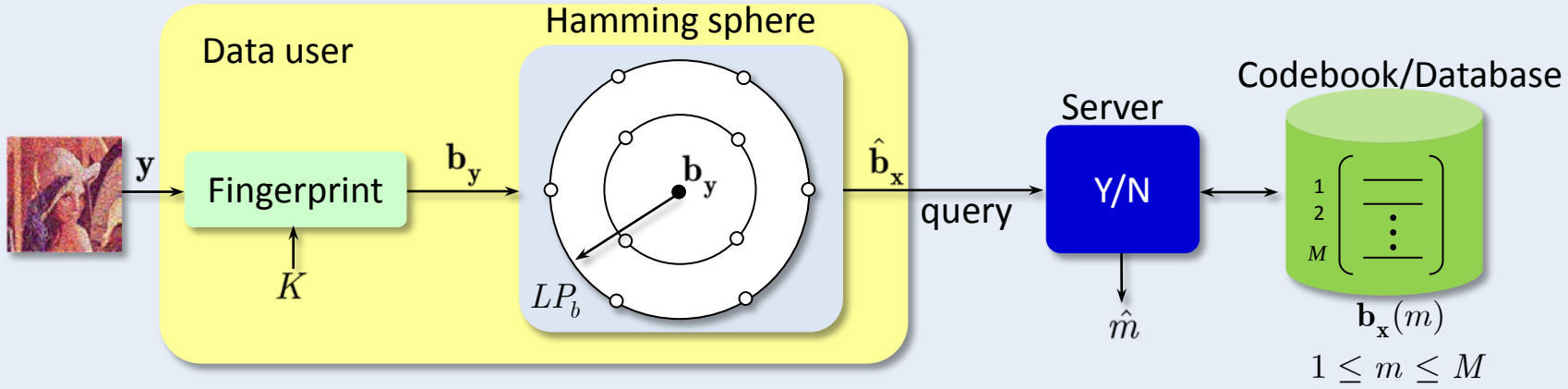
- no modifications
- random codes $\Rightarrow$ exhaustive search $\Rightarrow$ high complexity
- high probability of average bit error P_b $\Rightarrow$ low performance

Complexity of fingerprint based identification (Cont.)

Hamming sphere decoding

Observation: the most likely codewords $\mathbf{b}_{\mathbf{x}}(\hat{m})$ are within Hamming sphere of radius γL around $\mathbf{b}_{\mathbf{y}}$.

Identification = codeword verification [S. Voloshynovskiy et al, ITW2010, Dublin]

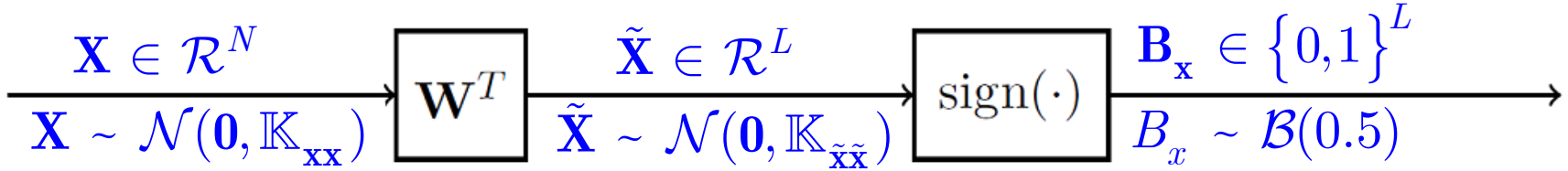


Remarks

- Complexity depends on "data quality": $\mathcal{O}\left(L2^{LH_2(P_b)}\right)$
- In limit $P_b \to 0$, complexity $\Rightarrow \mathcal{O}\left(L\right)$ (noiseless hashing)

pCFP: factors determining performance

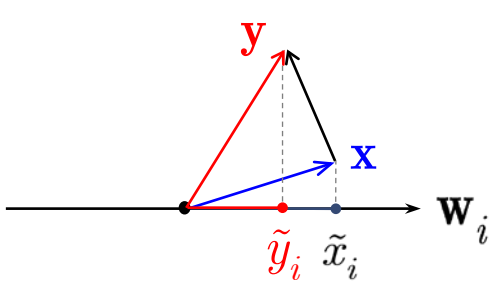
Fingerprinting (FP) block based on random projections:



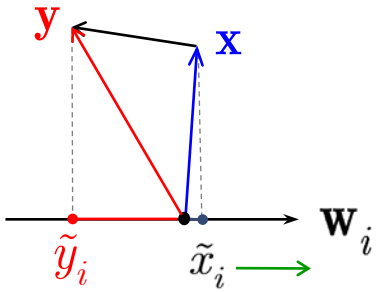
$$\mathbf{W} \in \{\pm 1\}^{N \times L}, W_{ij} \sim \mathcal{B}(0.5)$$

Additive White Gaussian Noise (AWGN): $\mathbf{y} = \mathbf{x} + \mathbf{z}$

$$P_{b|\tilde{x}} = Q\left(\frac{|\tilde{x}|}{\sigma_Z}\right), P_{b-\mathrm{pCFP}} = E\left[P_{b|\tilde{x}}\right] = \frac{1}{\pi}\arccos\left(\rho_{\tilde{X}\tilde{Y}}\right)$$



Reliable bit

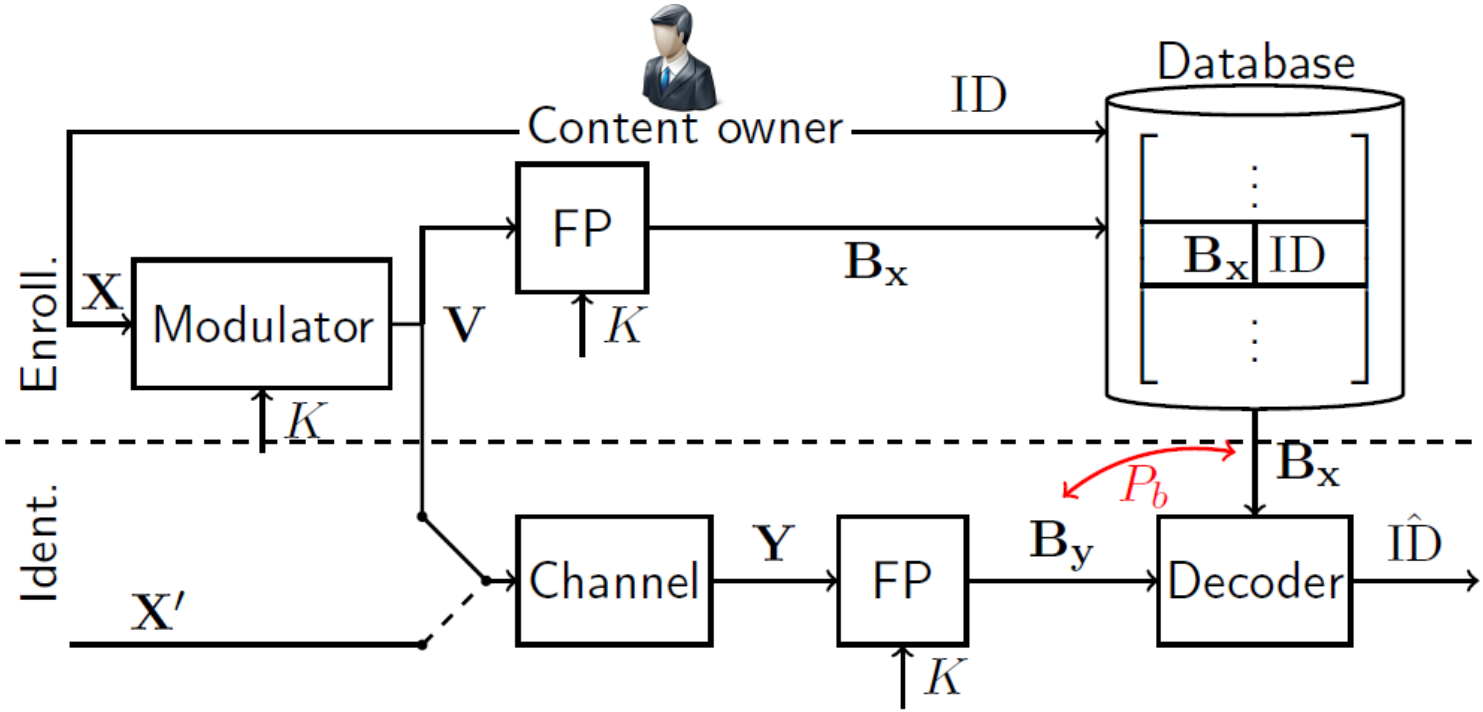


Unreliable bit

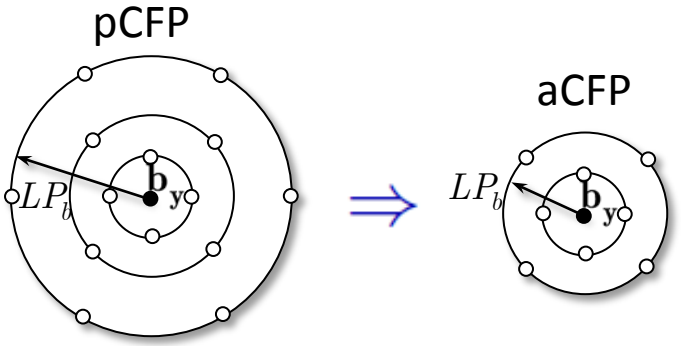
Can one increase $|\tilde{x}|$ by purpose?

Active Content Fingerprint (aCFP)

Proposed Approach:



- **No message will be embedded** (= pCFP)
- Content is modified
- **Main goal:** $P_b \rightarrow 0 \Rightarrow$ **low search complexity**



How it works?

modify images to make features (more) robust

- orthogonal expansion

$$\begin{cases} \tilde{x}_i &= \mathbf{w}_i^T \mathbf{x},\ 1 \le i \le N, \\ \mathbf{x} &= \sum_{i=1}^{N} \tilde{x}_i \mathbf{w}_i = \sum_{i \in \mathcal{K}} \tilde{x}_i \mathbf{w}_i + \sum_{i \notin \mathcal{K}} \tilde{x}_i \mathbf{w}_i \end{cases}$$

$$\mathcal{K} = \{i_1, i_2, \cdots, i_L\}$$

- modified content

$$\mathbf{v} = \sum_{i \in \mathcal{K}} \varphi_i(\tilde{x}_i)\mathbf{w}_i + \sum_{i \notin \mathcal{K}} \tilde{x}_i \mathbf{w}_i,$$

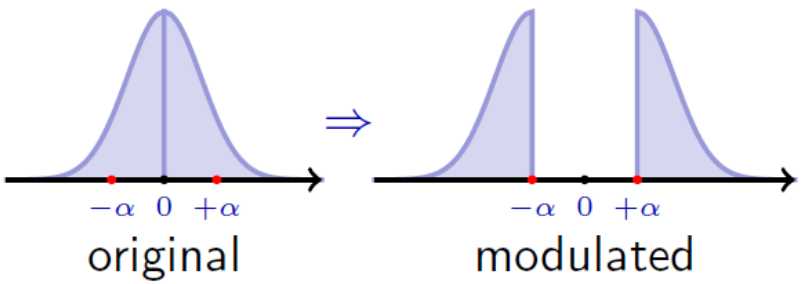
where $\varphi_i(\bullet), i \in \mathcal{K}$ is a modulation function, which can be the same for all i.

Additive aCFP (AddaCFP)

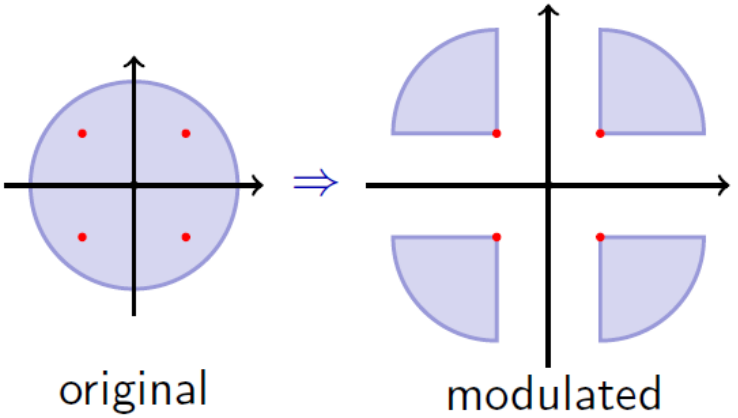
AddaCFP modulation function

$$\varphi_A(\tilde{x}_i) = \tilde{x}_i + \alpha \text{sign}(\tilde{x}_i), \forall i \in \mathcal{K}$$

One dimensional:



Two dimensional:



- Bit error probability (AWGN):

$$P_{b-\text{AddaCFP}} \leq \exp\left(-\frac{\alpha^2}{2\sigma_Z^2}\right) P_{b-\text{pCFP}}$$

With distortion

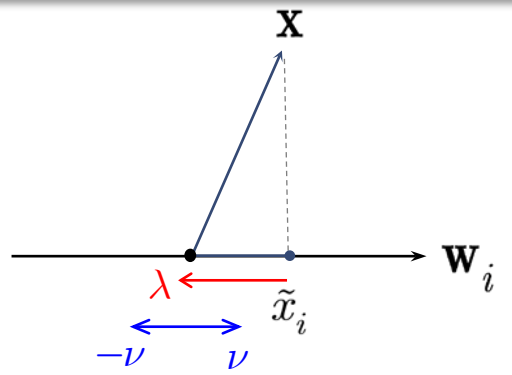
$$D_A = \frac{L}{N}\alpha^2, L = \mid \mathcal{K} \mid$$

Additive aCFP (AddaCFP): link to additive DWM

Link to improved spread spectrum (ISS) watermarking

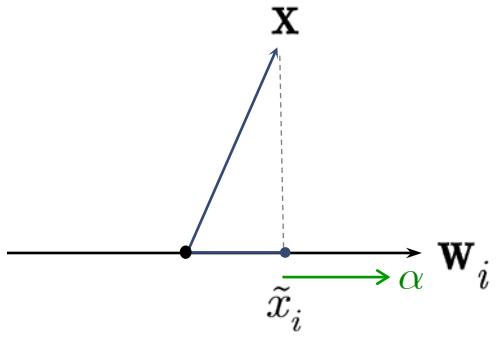
$$\mathbf{v} = \mathbf{x} - \underbrace{\lambda \sum_{i \in \mathcal{K}} \tilde{x}_i \mathbf{w}_i}_{\text{Interference cancellation}} + \underbrace{\nu \sum_{i \in \mathcal{K}} (-1)^{m_i} \mathbf{w}_i}_{\text{DWM embedding}}, \forall i \in \mathcal{K}, m_i \in \{0,1\}$$

- Two types of distortions due to:
  - host interference cancellation
  - WM embedding



$$\mathbf{v} = \mathbf{x} + \underbrace{\alpha \sum_{i \in \mathcal{K}} \text{sign}(\tilde{x}_i) \mathbf{w}_i}_{\text{AddaCFP content modulation}}$$

$\equiv$ Coefficient magnitude increase

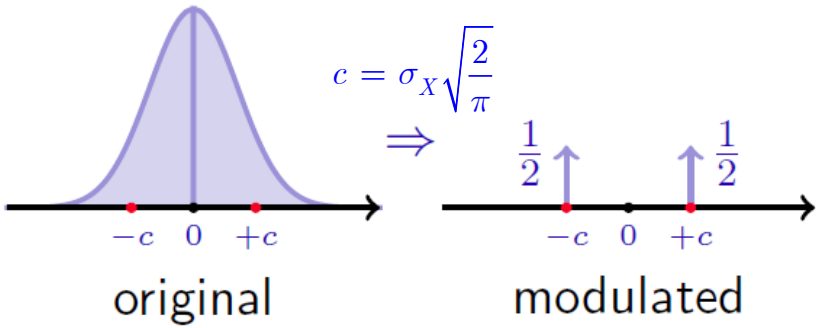


Quantization based aCFP (QbaCFP)

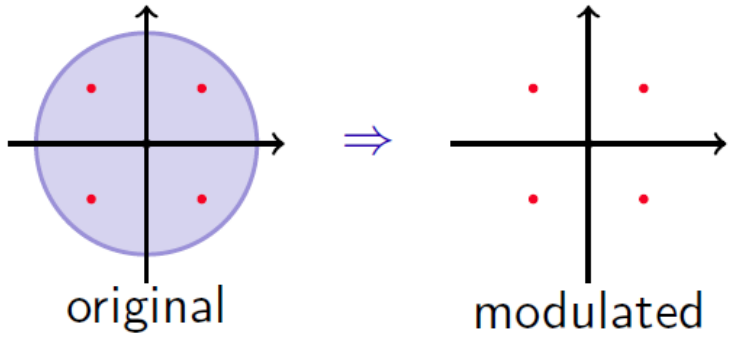
QbaCFP modulation function

$$\varphi_Q(\tilde{x}_i) = c\,\mathrm{sign}(\tilde{x}_i), \forall i \in \mathcal{K}$$

One dimensional:



Two dimensional:



- Bit error probability (AWGN):

$$P_{b-\mathrm{QbaCFP}} = Q\left(\frac{c}{\sigma_Z}\right) \overset{\min D_Q}{\Rightarrow} P_{b-\mathrm{QbaCFP}} = Q\left(\frac{\sigma_X}{\sigma_Z}\sqrt{\frac{2}{\pi}}\right)$$

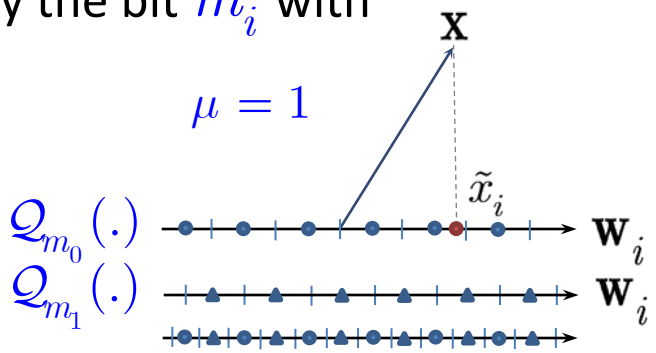
Quantization based aCFP (QbaCFP): link to QIM

Link to spread-transform dither modulation (ST-DM) watermarking

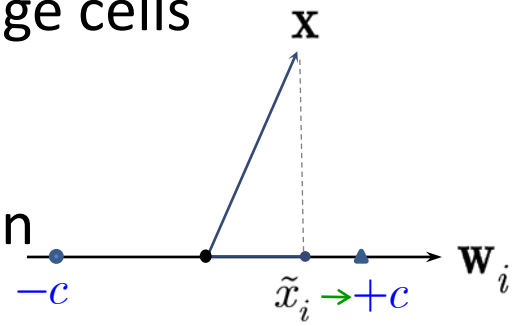
$$\mathbf{v} = \mathbf{x} + \mu \sum_{i \in \mathcal{K}} \left(\mathcal{Q}_{m_i}(\tilde{x}_i) - \tilde{x}_i \right) \mathbf{w}_i, \forall i \in \mathcal{K}, m_i \in \{0,1\}$$

where $\mathcal{Q}_{m_i}(.)$ is a scalar quantizer, which is defined by the bit m_i with the centroids denied by

$$c_i = \Delta \mathcal{Z} + (-1)^{m_i} \frac{\Delta}{4}, \text{for } m_i = 0,1$$



- The distortion is due to quantization
- The quantizers are message-dependent
- Errorless bit extraction, iff noise is within message cells

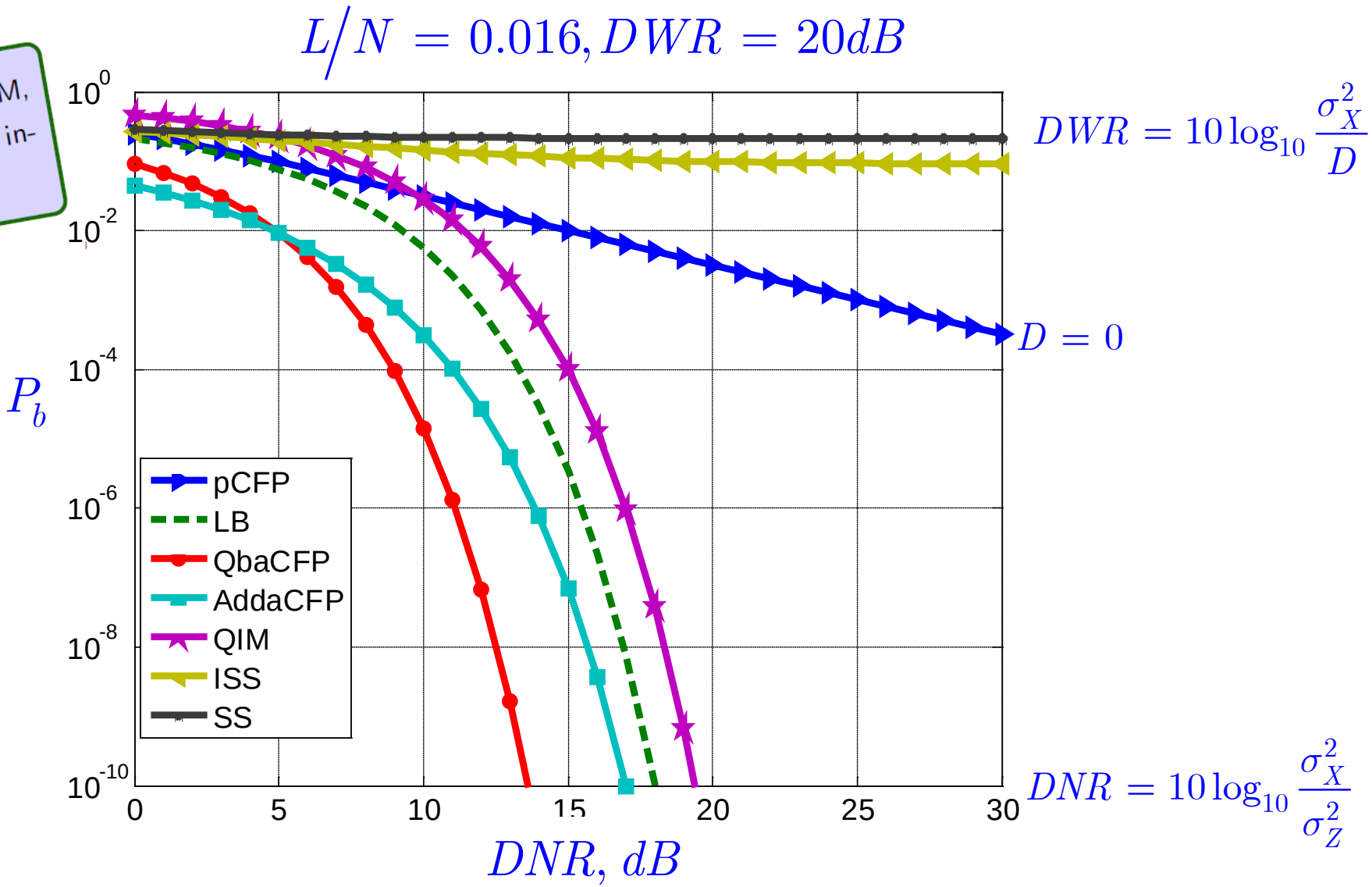


- The quanizer is message independent
- Errorless bit extraction, if noise does not flip sign

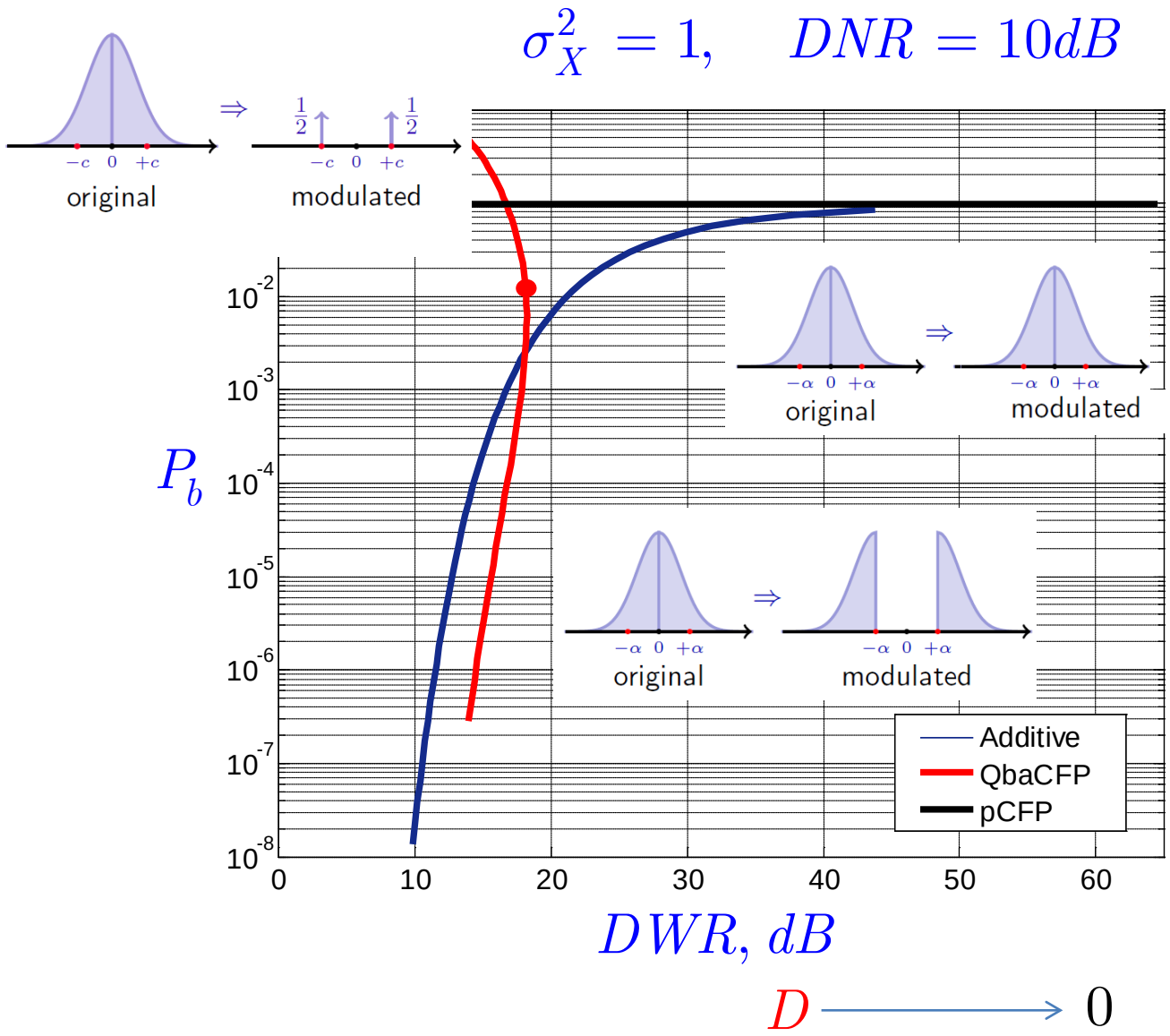
  $\Rightarrow$ higher tolerance to noise distortions

Analytical comparison

LB
Idealized DWM, without host interfrence.



Analytical comparison



Quantization based aCFP (QbaCFP)

Table: Comparison of BERs, $N = 768$, $L = 32$, $\alpha = 45$ and average PSNR=53dB (original to embedded image).

PSDR

Attack	Parameters	pCFP	AddaCFP
		$P_{b-\text{pCFP}}$	$P_{b-\text{AddaCFP}}$
AWGN	PSNR=5 dB	0.21	0.13
	PSNR = 10 dB	0.13	0.06
	PSNR = 15 dB	0.08	0.02
	PSNR = 20 dB	0.05	0.002
JPEG	QF = 1	0.11	0.04
	QF = 10	0.04	0.001
	QF = 25	0.01	0
Histeq		0.1	0.014

Feature vectors extracted from Uncompressed Colour Image Database (UCID), which consists of 1338 images of size 384 by 512.

F. Farhadzadeh, S. Voloshynovskiy, and O. J. Koval, "Performance Analysis of Content-Based Identification Using Constrained List-Based Decoding," *IEEE Transactions on Information Forensics and Security*, vol. 7, No. 5, pp. 1652-1667, 2012.

Conclusions

- aCFP: combination of pCFP and DWM
- Improved performance in terms of bit error probability
- Potential huge gain for lower complexity based on BDD (“crypto hashing”)

Future extensions

- Optimal modulation to trade-off distortion-performance
- New insights for security and privacy
- Multidimensional and lattice extensions
- Application to physical object security
  SNF project: Nano- and micro-structure identification based on controllable randomness