Worst case additive attack against quantization-based data-hiding methods: joint study of probability of error and mutual information

J.E. Vila-Forcén[a], S. Voloshynovskiy[a], O. Koval[a], F. Pérez-González[b] and T. Pun[a]

[a] **Stochastic Image Processing Group, University of Geneva, Switzerland**

[b] **Signal Processing in Communications Group, University of Vigo, Spain**

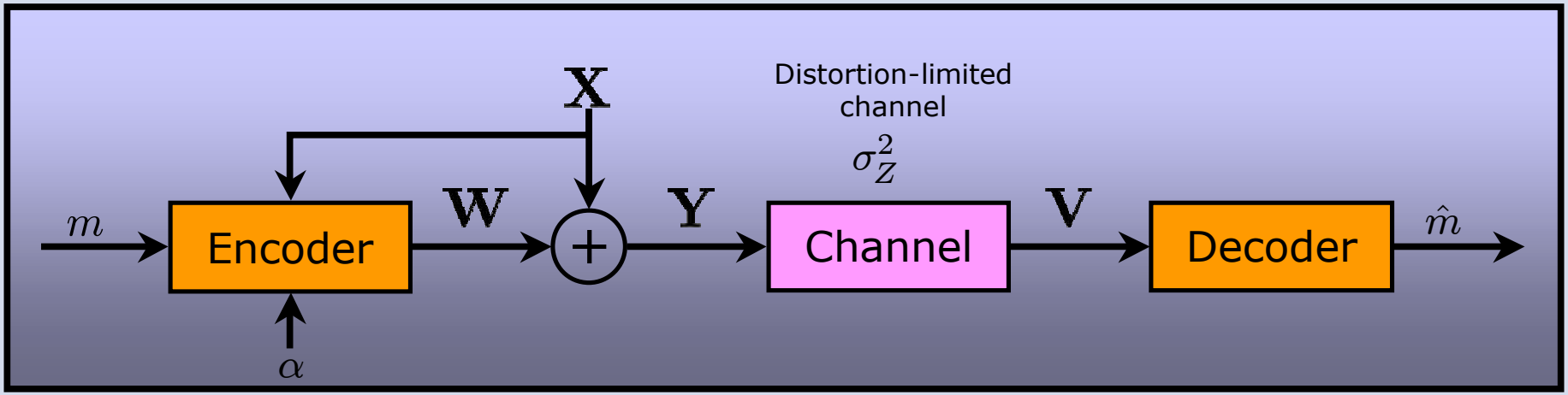
Agenda

- The worst case attack in digital communications and data-hiding;
- Dither modulation (DM) and distortion compensated (DCDM) techniques;
- Probability of error analysis;
- Mutual information analysis;
- Conclusions.

The worst case attack in data-hiding

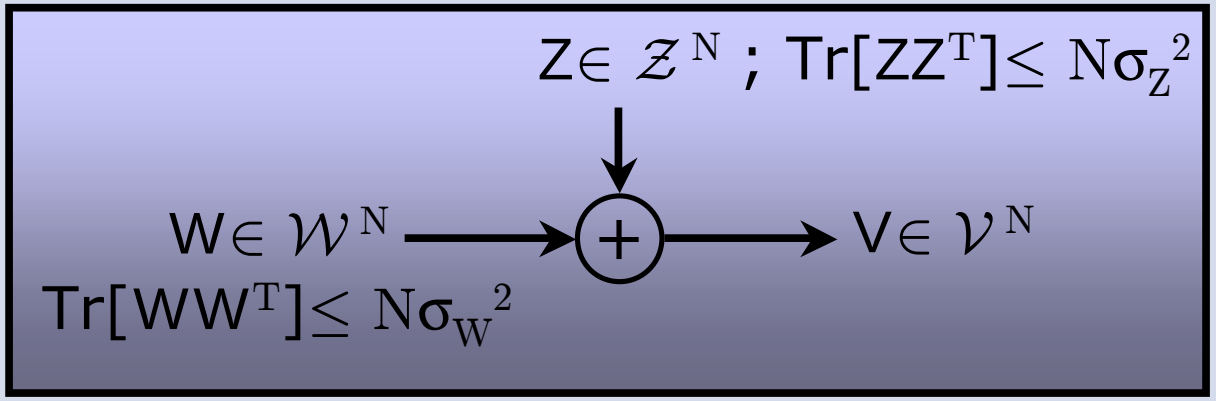
- **Data-hider objective:**
  To reliably communicate the maximum amount of information through the channel.

- **Attacker objective:**
  To impair the reliable communications.



The worst case attack in digital communications

Digital communications, additive attack:



- **Continuous input alphabet:** Within the class of additive attacks, the worst case attack is Additive White Gaussian Noise (AWGN);

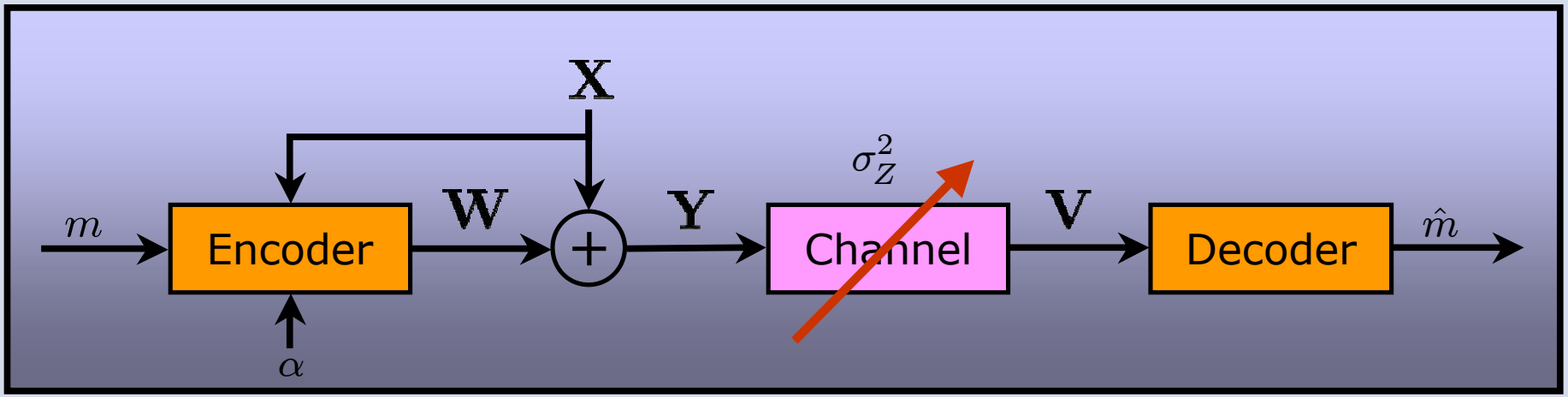
  [Cover and Thomas, Elements of Information Theory, 1991]

- **Discrete input alphabet:** AWGN is not the worst case attack.

  [Mc.Kellips and Verdu, IEEE IT, 1997]

The worst case attack in data-hiding

Problem: To find worst case additive attack (WCAA) against quantization-based techniques (DM and DCDM) using P_e and $I(.;.)$ as cost functions.



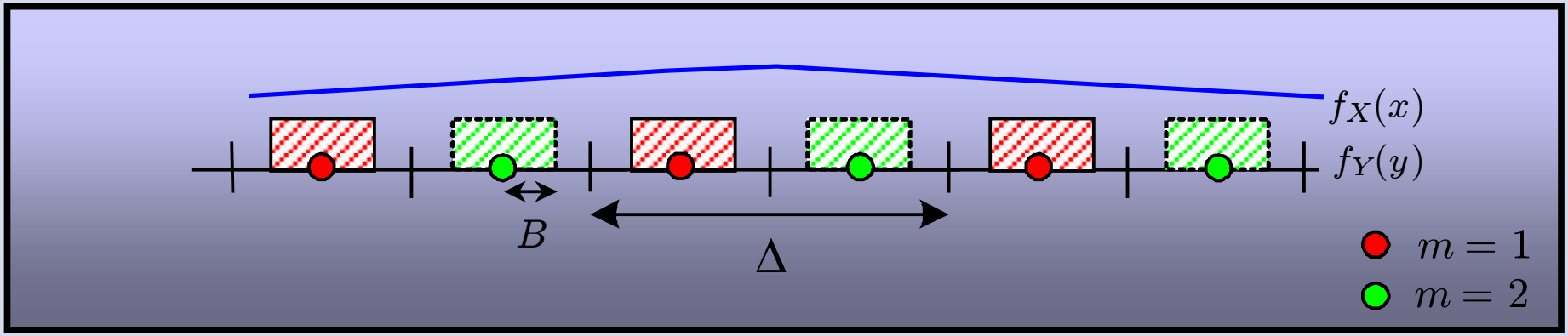
The DM and the DC-DM techniques

DM ($\alpha'=1$) and DCDM:

$$y = x + \alpha'(Q_m(x) - x),$$
$$D_W = \alpha'^2 \frac{\Delta^2}{12}.$$

Binary case

$$m \in \mathcal{M} = \{1, 2\}$$
$$B = \frac{\Delta}{2}(1 - \alpha')$$



Probability of error analysis

- **Assumption:** Encoder and decoder are fixed (practical set-up) for the probability of error as a cost function. $(v = y + z)$

$$P_e = P\{||v - Q_m(v)||^2 > ||v - Q_{m'}(v)||^2; m' \in \mathcal{M}, m' \neq m\}$$

$$P_e = \int_{\bigcup_{m' \neq m} \mathcal{R}_{m'}} f_V(v|M = m)dv$$

where $\mathcal{R}_m$ denotes the decision region associated to the message m.

Problem formulation WCAA:

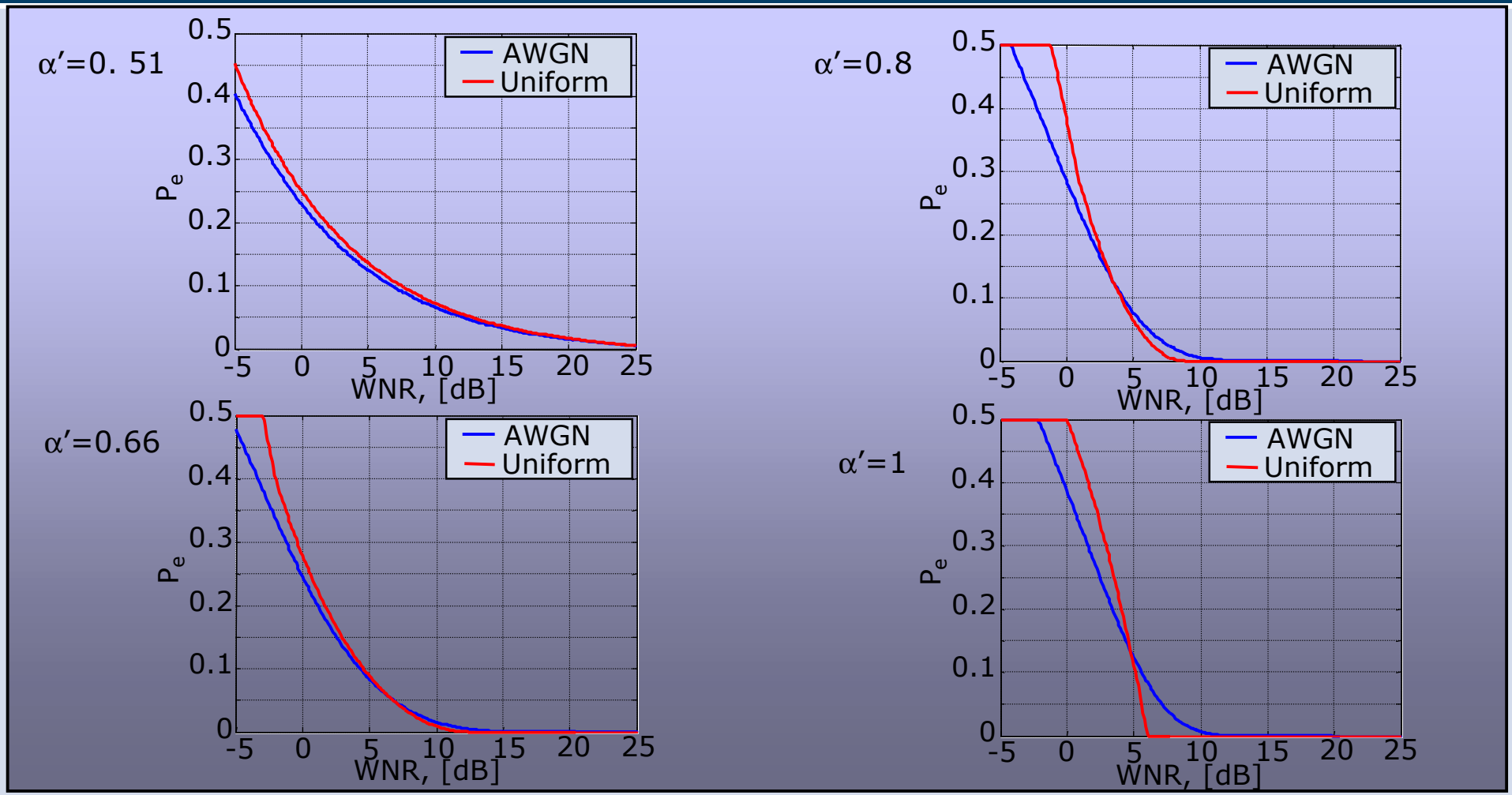
$$\max_{f_Z(\cdot)} P_e = \max_{f_Z(\cdot)} \int_{\bigcup_{m' \neq m} \mathcal{R}_{m'}} f_V(v|M = m)dv$$

Subject to:

$$\int_{-\infty}^{\infty} f_Z(z)dz = 1,$$

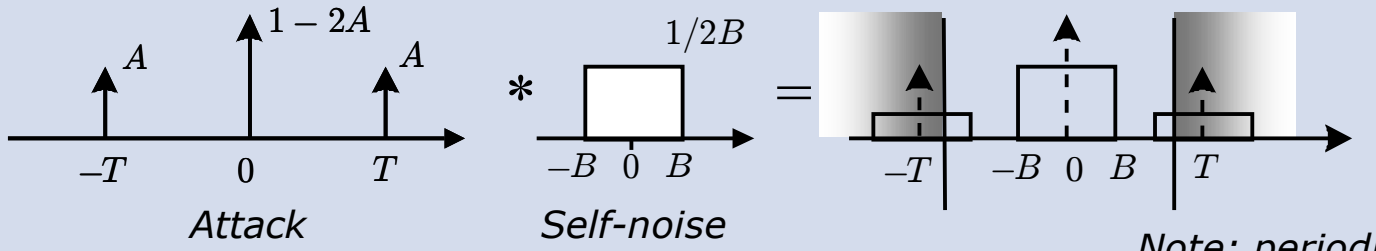
$$\int_{-\infty}^{\infty} z^2 f_Z(z)dz \leq \sigma_Z^2.$$

Probability of error: AWGN and uniform noise attacks



Probability of error: analytical WCAA.

Result of the problem optimization: 3- d attack:



Note: periodic error region.

DCDM:

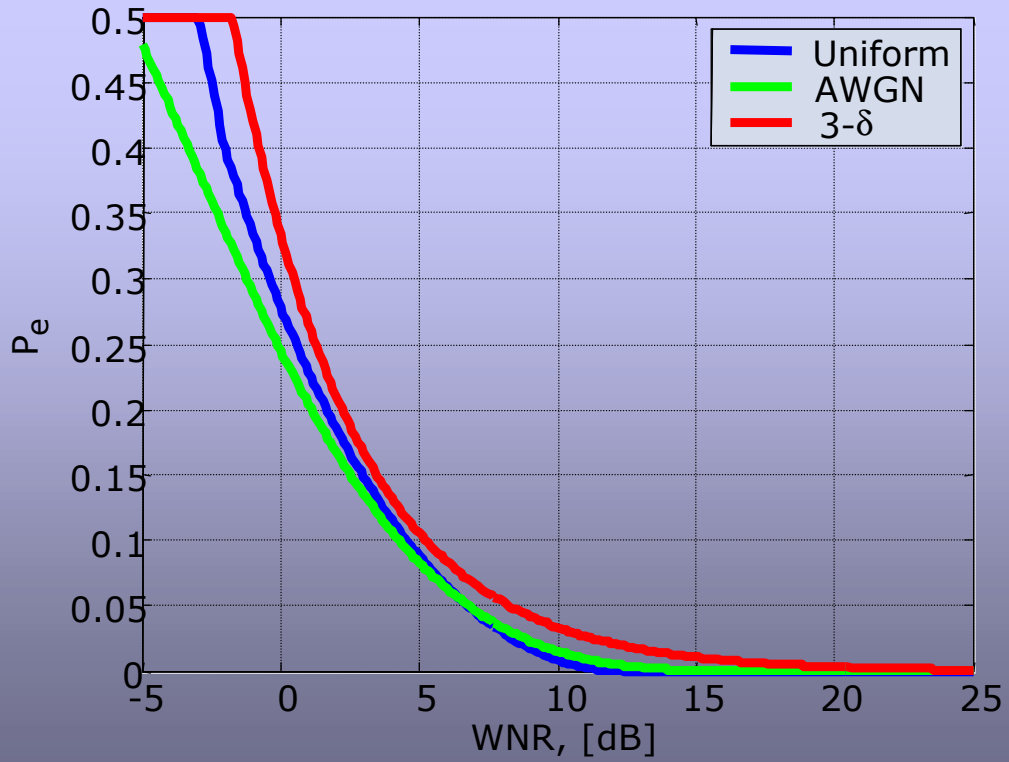
$$T = (2\alpha' - 1)\frac{\Delta}{2}, \alpha' < \frac{5}{6},$$

$$T = (3 - 2\alpha')\frac{\Delta}{4}, \alpha' > \frac{5}{6},$$

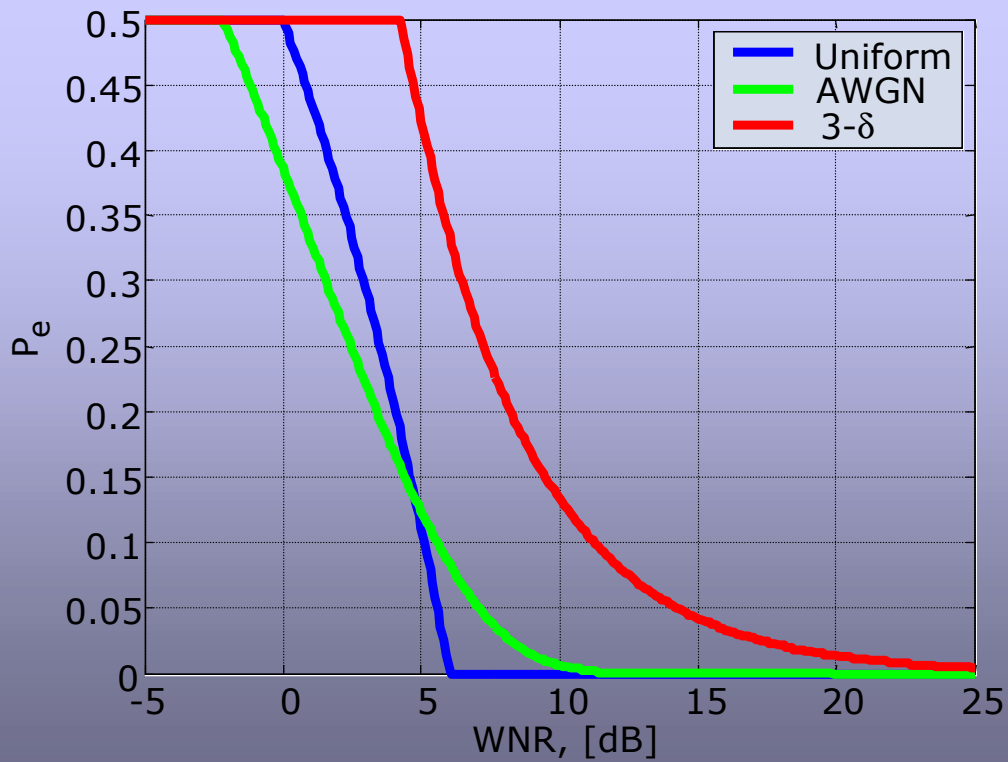
$$A = \frac{\sigma_Z^2}{2T^2}.$$

$$P_e = \begin{cases} \frac{\sigma_Z^2 \alpha'^2}{12\sigma_W^2(\alpha'(3-2\alpha')-1)}, & \alpha' < \frac{5}{6}, \\ \frac{4\sigma_Z^2 \alpha'^2}{3\sigma_W^2(3-2\alpha')^2}, & \alpha' > \frac{5}{6}. \end{cases}$$

Probability of error: results



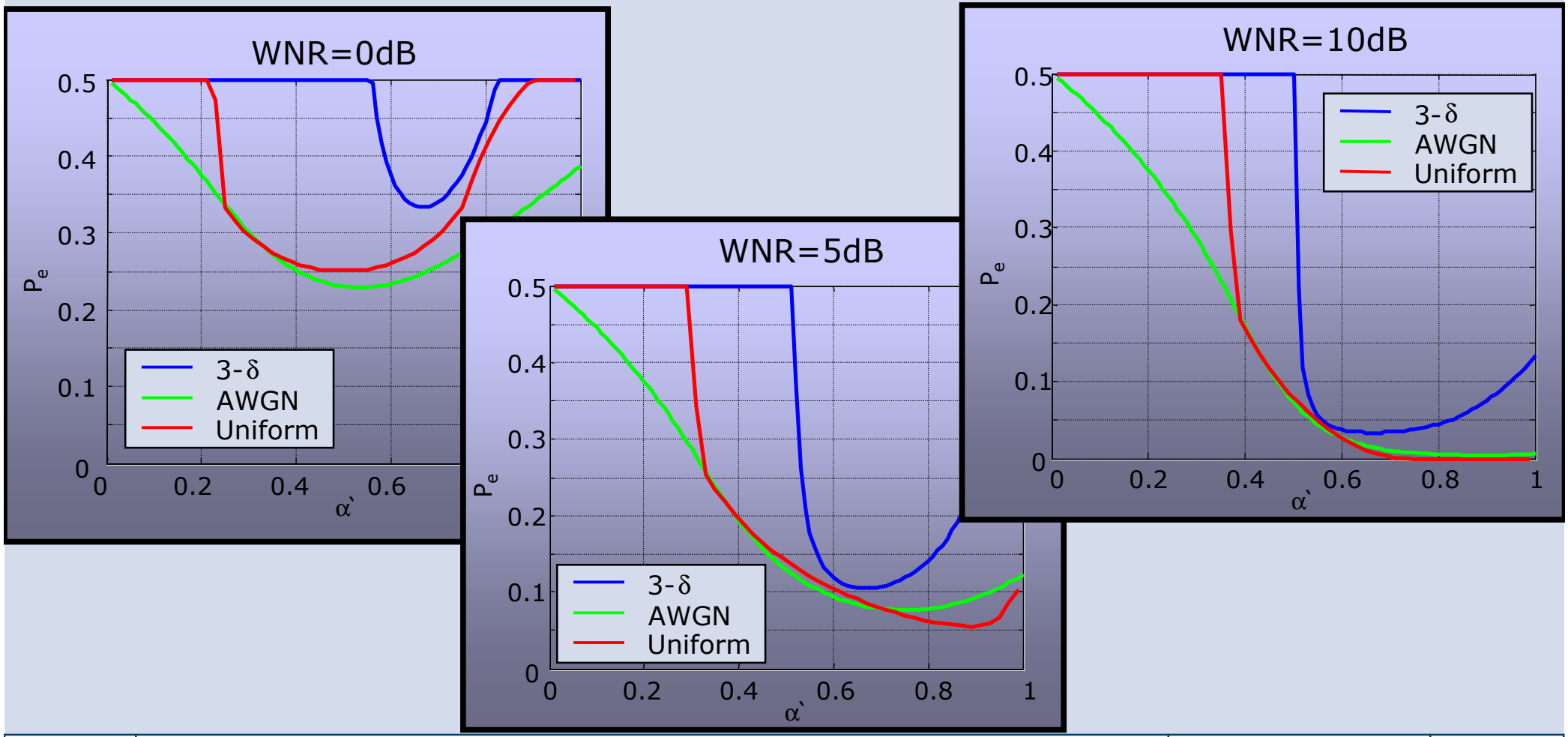
$\alpha' = 0.66$



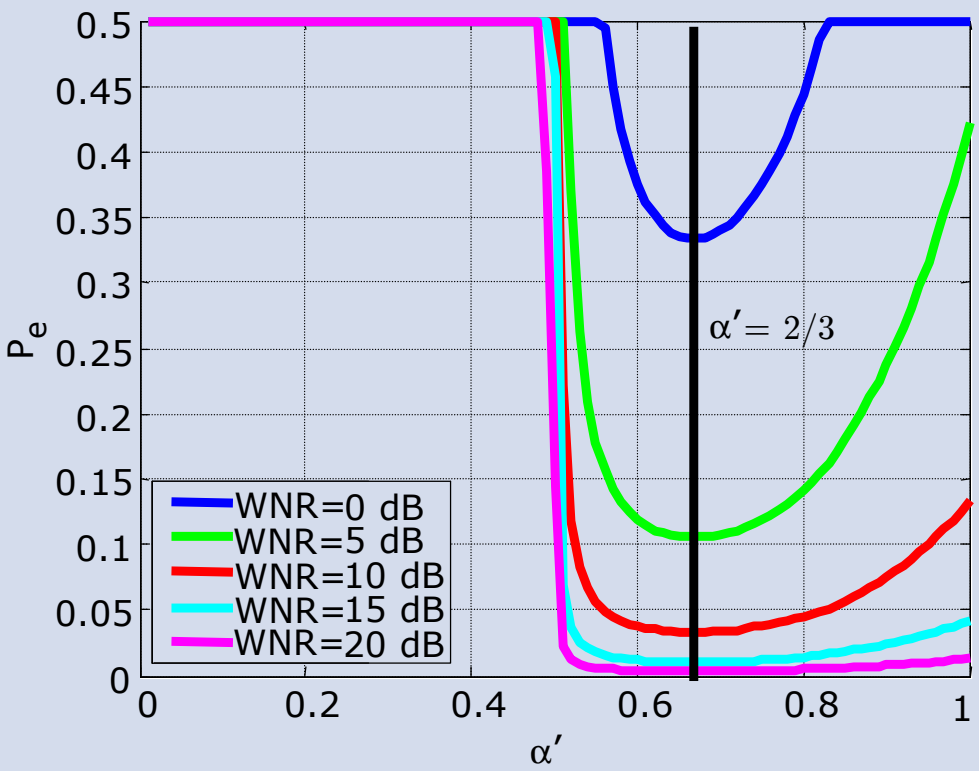
$\alpha' = 1$ (DM)

Conclusion: The 3-δ attack is the WCAA in terms of probability of error.

Probability of error

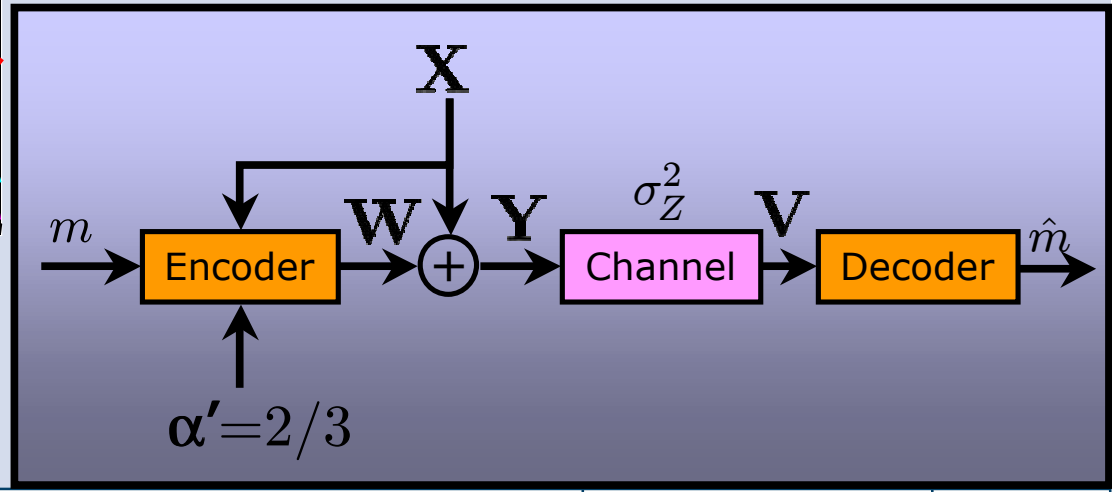


Probability of error: optimal compensation parameter



3-δ attack

- Optimal compensation parameter can be chosen equal to $\alpha'=2/3$ in order to guarantee that the P_e is bounded;
- The use of proper error correction codes leads to reliable communication.



Mutual information analysis

There is no assumption about the decoder structure:

$$I(V;M) = D(f_{V,M}(v,m)||f_M(m)f_V(v))$$

$$I(V;M) = D\left(\tilde{f}_{V|M}(v|M=1)||\tilde{f}_V(v)\right)$$

$D(\cdot||\cdot)$: Kullback-Leibler distance

$$\tilde{f}_{V|M}(v|M=m) \triangleq \begin{cases} \sum_i f_{V|M}(v-i\Delta|M=m), & v \in \mathcal{R}_m; \\ 0, & \text{otherwise.} \end{cases}$$

$$\tilde{f}_V(v) = \tfrac{1}{2}\sum_{m=1}^{2} \tilde{f}_{V|M}(v|M=m)\ .$$

Problem formulation:

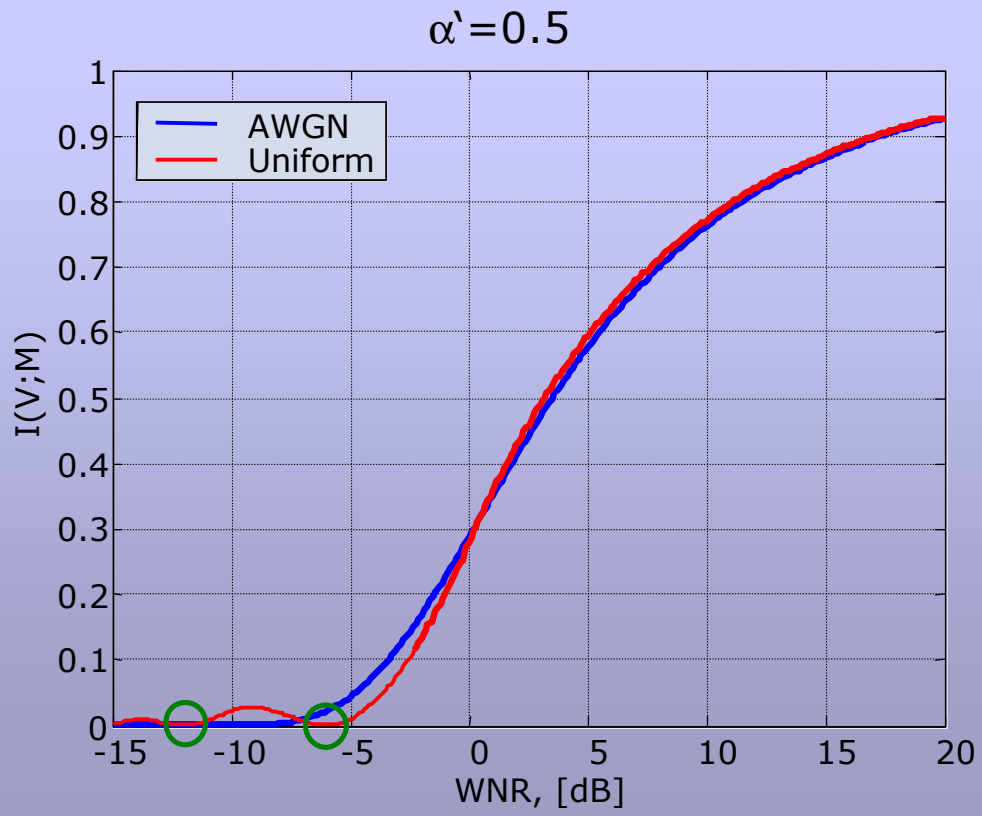
$$\min_{\tilde{f}_Z(\cdot)} I(V;M) = \min_{\tilde{f}_Z(\cdot)} D\left(\tilde{f}_{V|M}(v|M=1||\tilde{f}_V(v)\right)$$

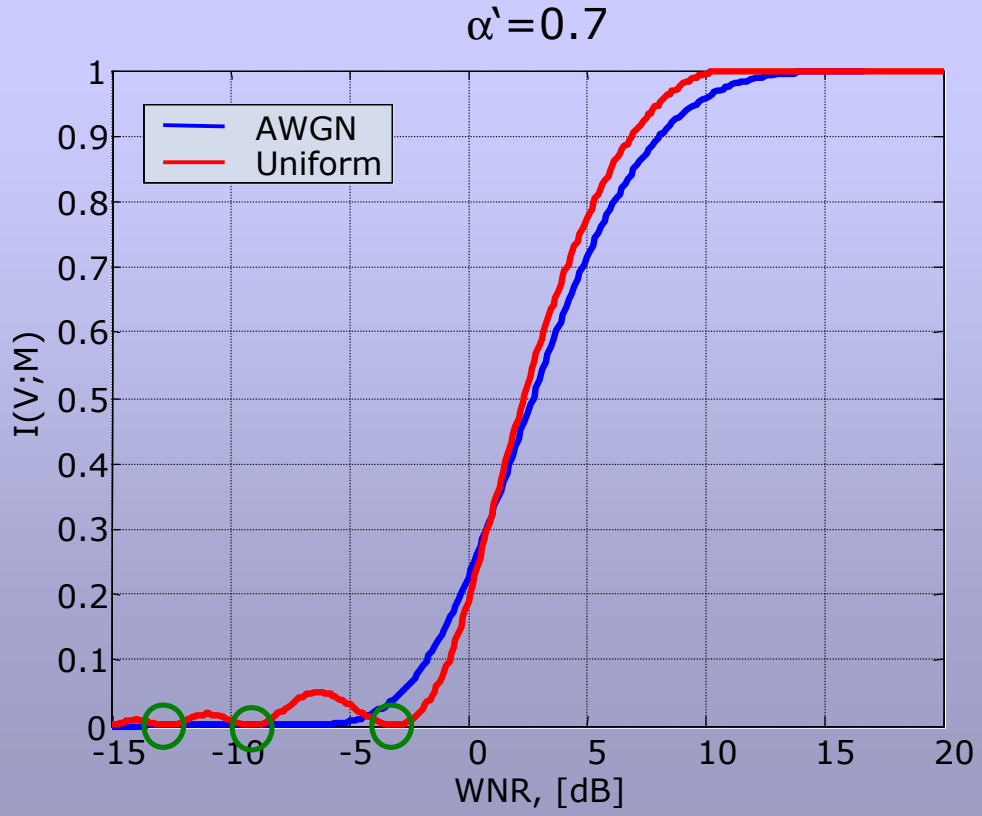
Subject to:

$$\int_{-\infty}^{\infty} f_Z(z)dz = 1,$$

$$\int_{-\infty}^{\infty} z^2 f_Z(z)dz \leq \sigma_Z^2.$$

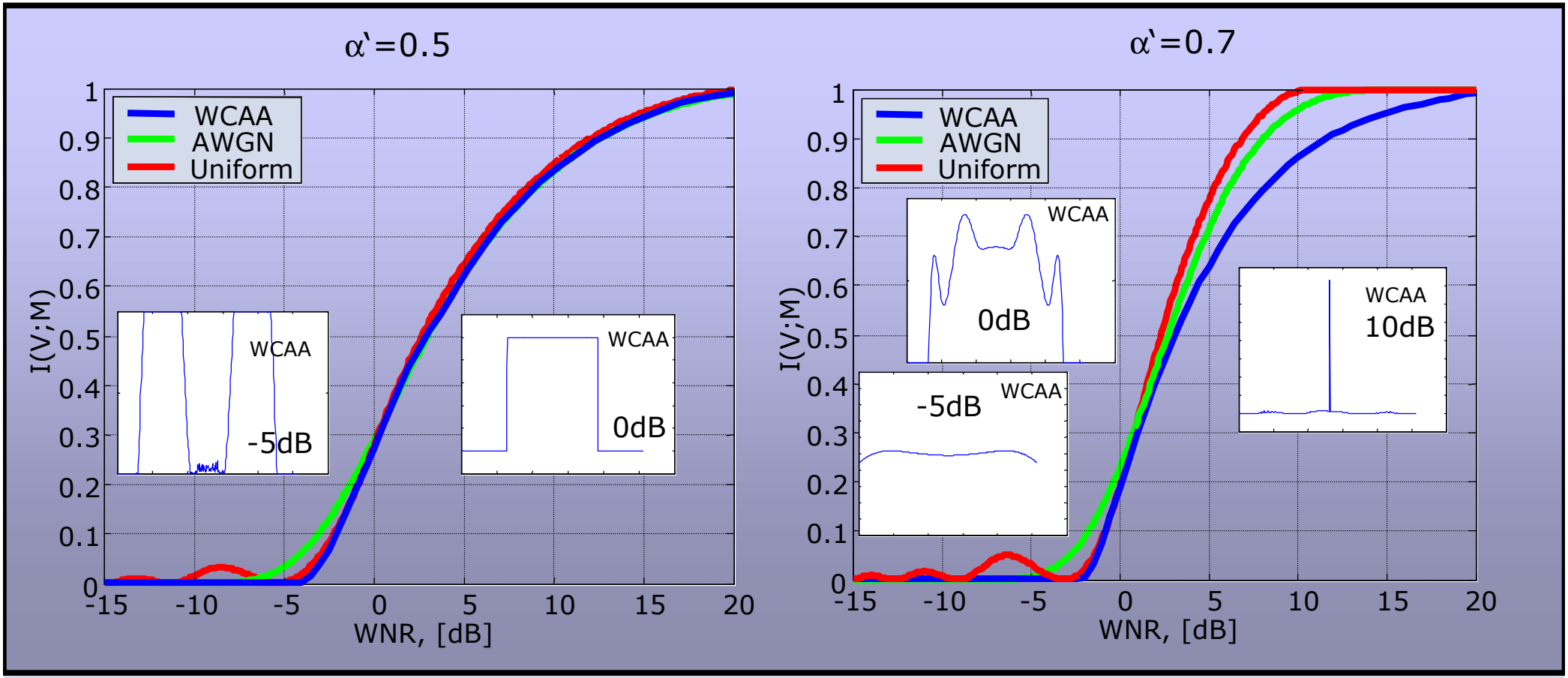
Mutual information: results





$I(V;M) = 0$ when $\mathrm{WNR} = 10 \log_{10}(\alpha'^2/k^2)$, $k \in \mathbb{N}$.

Mutual information: results



Conclusions

- **WCAA problem against quantization-based methods using probability of error and mutual information as cost functions is considered;**
- **The analytical pdf of the WCAA for the probability of error as a cost function has been obtained;**
- **An optimal compensation parameter $\alpha'=2/3$ has been found for the minimum distance decoder;**
- **The particular WCAA pdfs for different WNRs are studied;**
- **Fair benchmarking of quantization-based methods should be performed accordingly.**