Information-Theoretic Benchmarks for Coded Modulation

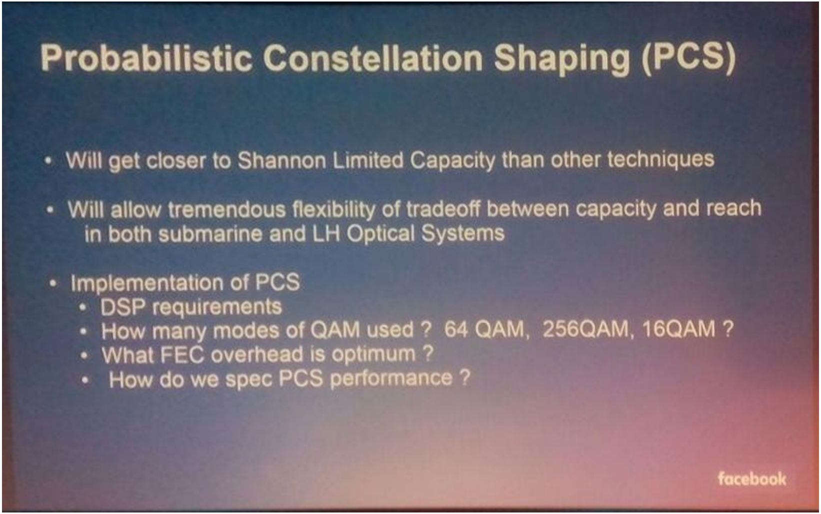
Georg Böcherer
georg.boecherer@tum.de

Technical University of Munich

October 9, 2017

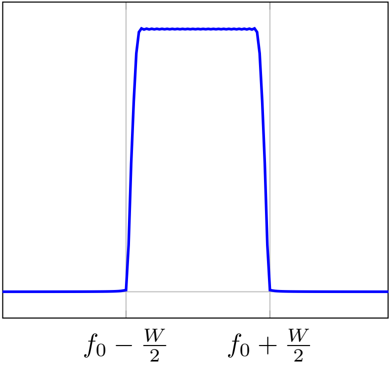
Steve Grubb on Probabilistic Shaping

Global Optical Architecture at **Facebook**, current work focussed on building new worldwide submarine fiber optics networks.



Outline

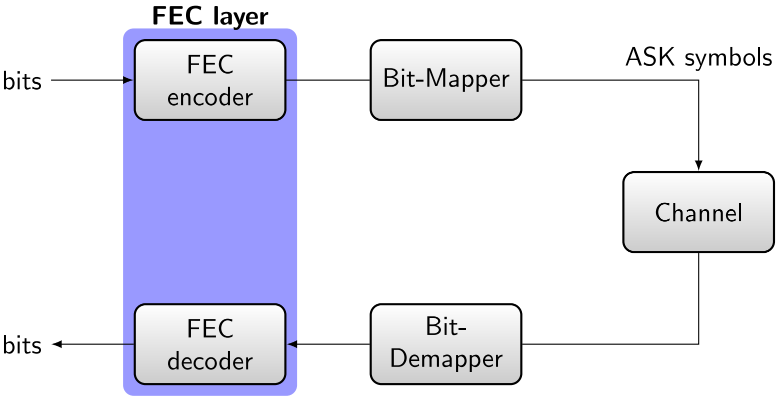
Higher-Order Modulation



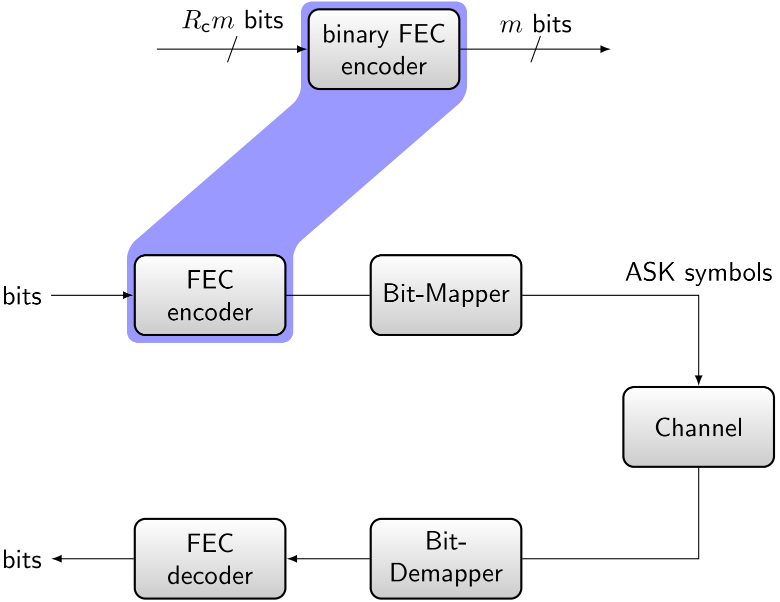
Grid width $W = 50$ GHz, dual polarization.

	QPSK		higher-order modulation			
spectral efficiency $\left[\frac{\text{bits}}{\text{QAM symbol}}\right]$	1	2	3	4	5	6
max net data rate $\left[\frac{\text{Gbit}}{\text{s}}\right]$	100	200	300	400	500	600

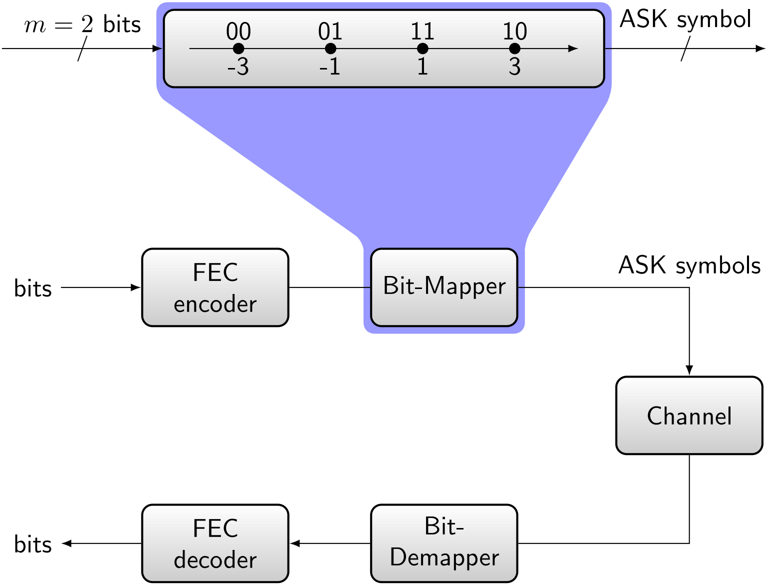
Coded Modulation



Coded Modulation: Binary FEC Code

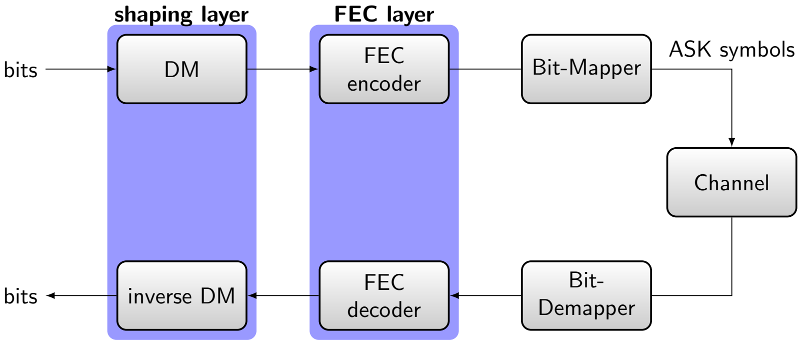


Coded Modulation: Example Bitmapper

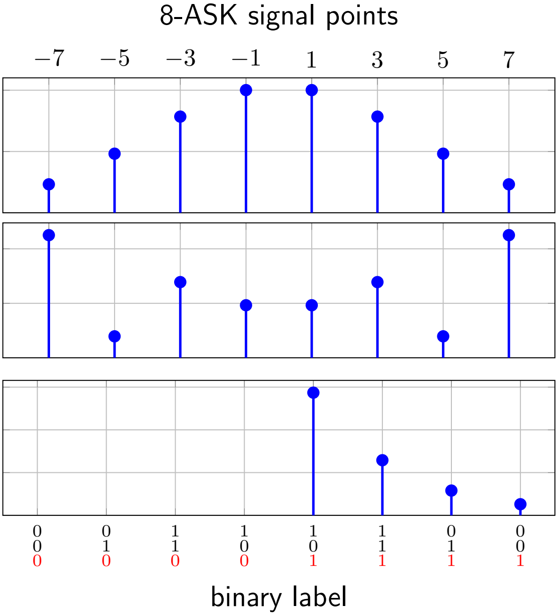


Outline

Layered Probabilistic Shaping



Good Input Distributions



Enabling Component: Systematic Encoder

- Every linear code[1] has a systematic encoder of the form:
  - Copy input unchanged to output.
  - Append parity bits.
- In matrix representation:

$$\boldsymbol{y} = \boldsymbol{u}\boldsymbol{G} = \boldsymbol{u}[\boldsymbol{I}|\boldsymbol{P}] = \boldsymbol{u}\boldsymbol{s}$$

  - Generator matrix $\boldsymbol{G}$, identity matrix $\boldsymbol{I}$, parity forming part $\boldsymbol{P}$.
  - Source bits $\boldsymbol{u}$.
  - Parity bits $\boldsymbol{s} = \boldsymbol{u}\boldsymbol{P}$.

[1] Up to column permutations

Binary FEC Encoder for Probabilistic Amplitude Shaping

Systematically encode the $m-1$ amplitude bits:

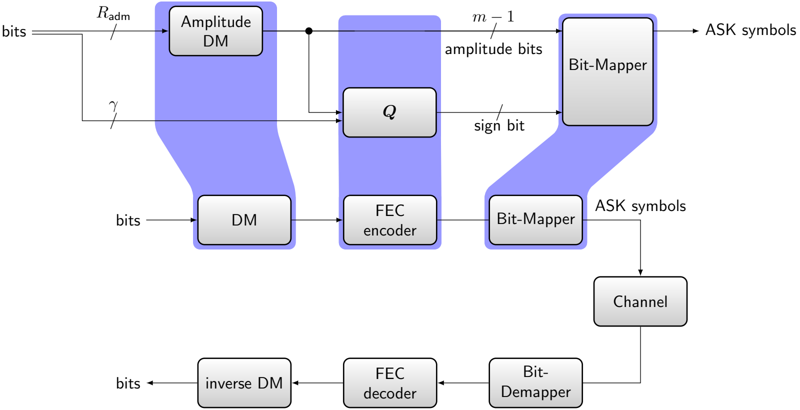
- Binary $(R_{\text{bc}}mn, mn)$ code.
- Dimension $R_{\text{bc}}mn \geq (m-1)n$

$$\Rightarrow R_{\text{bc}} = 1 - \frac{1}{m} + \frac{\gamma}{m}, \quad 0 \leq \gamma \leq 1.$$

- $(m-1)n$ systematic bits $\Rightarrow$ encode with generator matrix

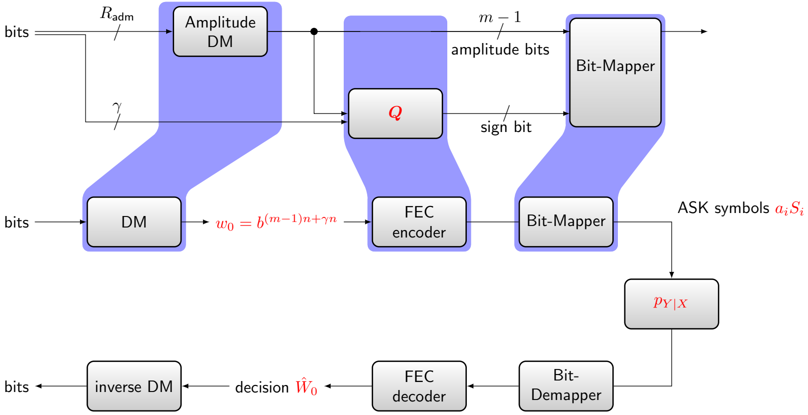
$$\boldsymbol{G} = [\boldsymbol{I}|\boldsymbol{Q}], \quad \boldsymbol{I} \text{ is } R_{\text{bc}}mn \times (m-1)n \text{ upper identity matrix.}$$

Probabilistic Amplitude Shaping



Outline

Random PAS Ensemble



FEC Decoder

- Demapper calculates non-negative metric

$$q(x, y), \quad x \in \mathcal{X}, y \in \mathcal{Y}.$$

- Decoder calculates

$$\hat{W} = \underset{w \in \{0,1\}^{R_{\text{bc}} mn}}{\operatorname{argmax}} \prod_{i=1}^{n} q(X_i(w), y_i)$$

- Decoding error probability

$$P_e = \Pr(\hat{W} \neq w_0).$$

- How large can we make the code rate $R_{\text{c}} = R_{\text{bc}} m$ while ensuring $P_e = 0$ for large n?

Achievable Code Rate

- **Achievable code rate**:

$$R_{\text{ac}} = \underbrace{\log_2 |\mathcal{X}|}_{=m} - \underbrace{\mathbb{E}\left[-\log_2 \frac{q(X,Y)}{\sum_{x\in\mathcal{X}} q(x,Y)}\right]}_{\text{uncertainty}}$$

- $P_X = P_A P_U$
  - P_A = empirical amplitude distribution of a^n.
  - P_U = uniform sign distribution on $\{-1, 1\}$.

Achievable Binary Code (ABC) Rate

- Represent input X by binary label $\boldsymbol{B} = B_1 B_2 \ldots B_m = \text{label}(X)$.
- **ABC rate:**

$$R_{\text{abc}} = 1 - \underbrace{\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}\left[-\log_2 \frac{q_i(B_i, Y)}{\sum_{b \in \{0,1\}} q_i(b, Y)} \right]}_{\text{binary uncertainty (buc)}}$$

Achievable Code Rate Instances

symbol-metric	achievable code rate R_{ac}
optimal	$m - \mathbb{H}(X\|Y)$
bit-metric	**ABC rate** R_{abc}
optimal	$1 - \frac{1}{m}\sum_{j=1}^{m}\mathbb{H}(B_j\|Y)$
bit-interleaved	$1 - \mathbb{H}(B\|Y)$
Hamming metric	$1 - \mathbb{H}(\epsilon)$
⋮	

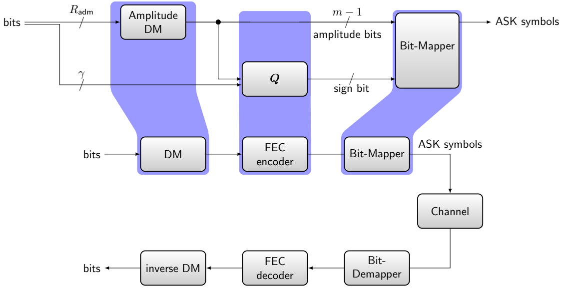
- All follows by $\mathbb{D}(P\|P') \geq 0$ with equality iff $P = P'$.

Outline

- Achievable code rate holds for every amplitude sequence a^n with empirical distribution P_A.

$\Rightarrow$ Encode into permutations of a^n.

Distribution Matching (DM) Rate



- ADM maps k_{adm} input bits to n_{adm} amplitudes:

$$d_1 d_2 \dots d_{k_{\text{adm}}} \text{ bits} \rightarrow \boxed{\text{ADM}} \rightarrow a_1 a_2 \dots a_{n_{\text{adm}}} \text{ amplitudes}$$

- Empirical output distribution P_A.
- Rate $R_{\text{adm}} = \frac{k_{\text{adm}}}{n_{\text{adm}}}$.

PAS Rate

- Code rate $R_{\mathrm{c}} = mR_{\mathrm{bc}} = m - 1 + \gamma$.
- Achievable code rate

$$R_{\mathrm{ac}} = m - \text{uncertainty}.$$

- Rate $R = R_{\mathrm{adm}} + \gamma$.
- Achievable rate

$$\boxed{R_{\mathrm{pas}} = [R_{\mathrm{adm}} + 1 - \text{uncertainty}]^{+}}.$$

CCDM Rate Loss

- Constant Composition Distribution Matcher (CCDM) indexes all length n sequences of distribution P_A.
- CCDM Rate loss

$$\mathbb{H}(P_A) - R_{\text{ccdm}}(P_A, n) \in \Theta\left(\frac{\log n}{n}\right).$$

$\Rightarrow$ For large n, PAS rate is

$$\boxed{R_{\text{pas}} = [\mathbb{H}(P_X) - \text{uncertainty}]^+}.$$

where $\mathbb{H}(P_X) = \mathbb{H}(P_A) + 1$.

PAS Rate Instances

symbol-metric	PAS Rate R_{pas}
optimal	$[\mathbb{H}(X) - \mathbb{H}(X\|Y)]^+ = \mathbb{I}(X;Y)$
bit-metric	
optimal	$\left[\mathbb{H}(\boldsymbol{B}) - \sum_{j=1}^{m} \mathbb{H}(B_j\|Y)\right]^+$
bit-interleaved	$[\mathbb{H}(\boldsymbol{B}) - m\,\mathbb{H}(B\|Y)]^+$
Hamming metric	$[\mathbb{H}(\boldsymbol{B}) - m\,\mathbb{H}(\epsilon)]^+$
⋮	

- **Information-theoretic remark:**
  By $\mathbb{I}(X;Y)$, PAS can achieve capacity with
  1. non-uniform P_X,
  2. linear code,
  3. no alphabet extension.

  Gallager (1968) can do 1, 2 or 1, 3 or 2, 3, but not 1, 2, 3.

Outline

Parametrization of AWGN Channel

- 8-ASK constellation:

$$\mathcal{X} = \{\pm 1, \pm 3, \pm 5, \pm 7\}.$$

- Input X with distribution P_X on $\mathcal{X}$.
- Binary label

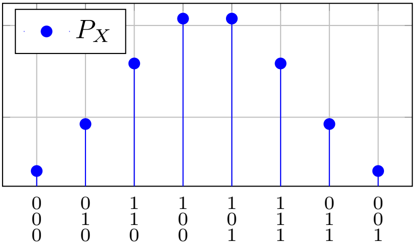
$$\boldsymbol{B} = B_1 B_2 B_3 = \text{label}(X).$$

- Channel output

$$Y = \Delta X + Z.$$

- Noise Z with variance σ^2.
- Signal-to-noise ratio (SNR)

$$\text{SNR} = \frac{\mathbb{E}\left[(\Delta X)^2\right]}{\sigma^2}.$$



Fixed Overhead

- Binary code rate $R_{\text{bc}} = 5/6$ (20% FEC overhead)
- For each SNR, choose P_X, Δ:
  - Maximize $\mathbb{H}(A)$.
  - **Subject to**
    1. SNR constraint:

$$\frac{\mathbb{E}[(\Delta X)^2]}{\sigma^2} = \text{SNR}$$

    2. Uncertainty constraint:

$$\frac{1}{m}\sum_{j=1}^{m} \mathbb{H}(B_j|\Delta X + Z) = 1 - R_{\text{bc}}.$$

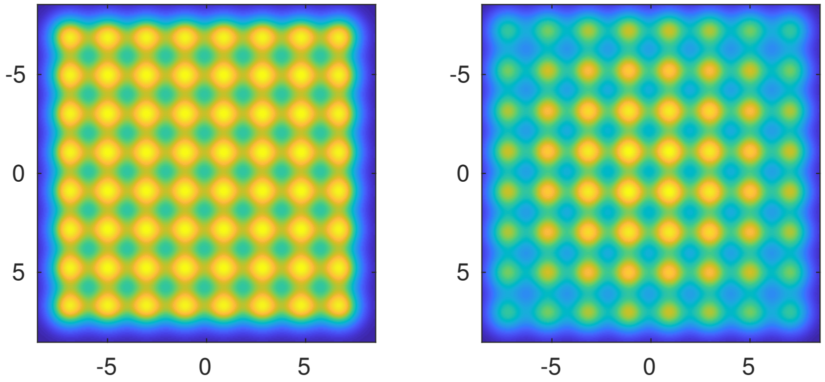
- Uncertainty is the same for all SNRs $\Rightarrow$ **fixed FEC overhead**.

Scatterplot of Received 64-QAM

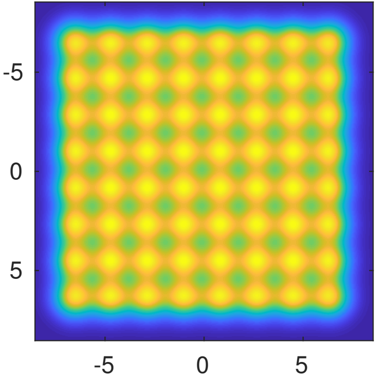
On the next slides:

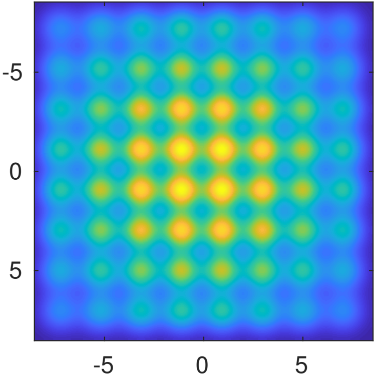
- Scatterplot as heatmap of 64-QAM constellation superposed with AWGN.

16,0 dB

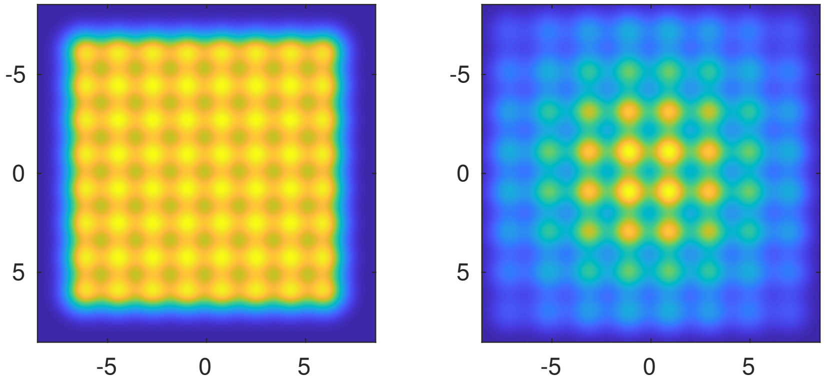


15,5 dB

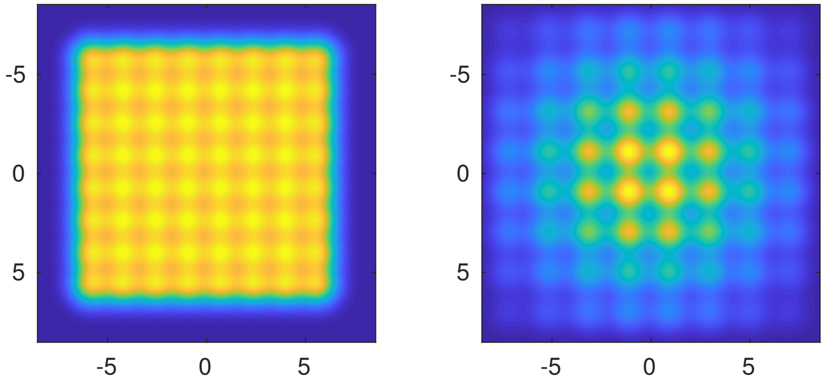




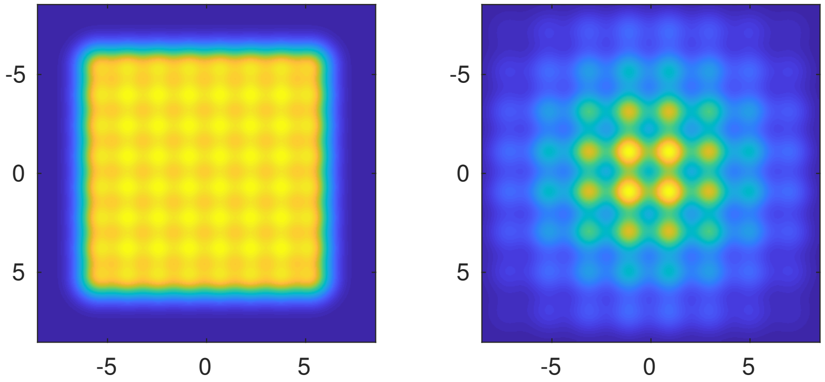
15,0 dB



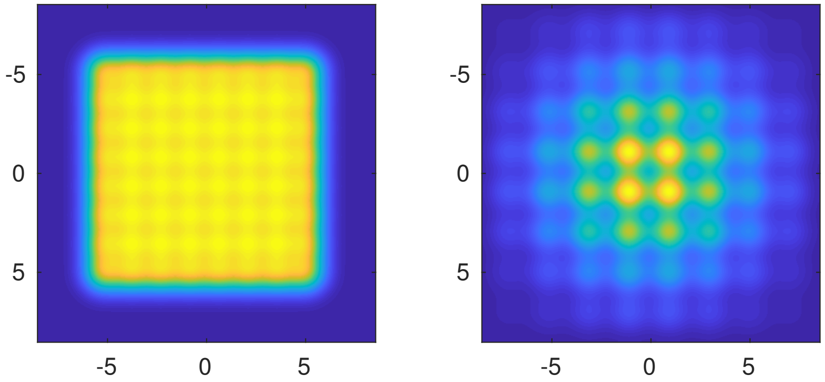
14,5 dB



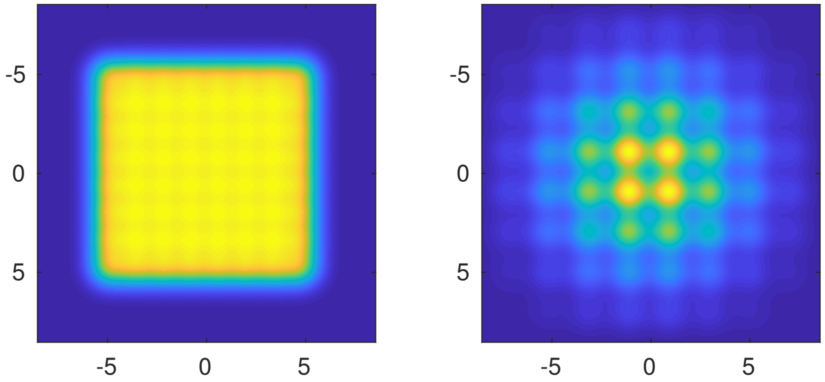
14,0 dB



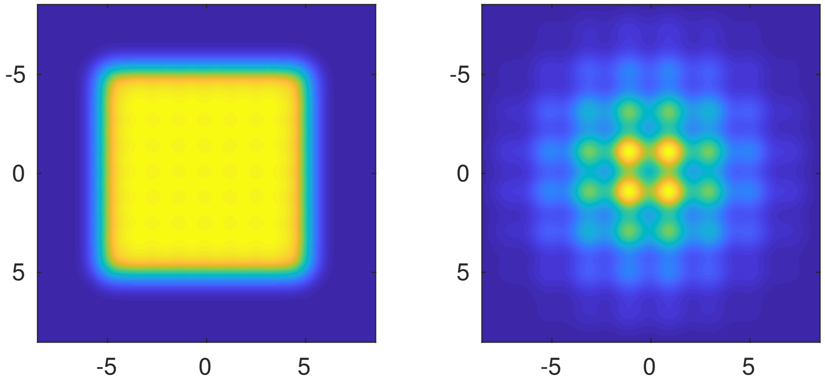
13,5 dB



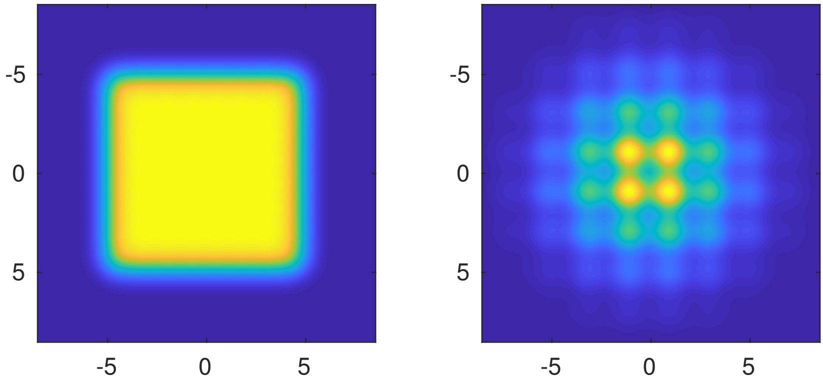
13,0 dB



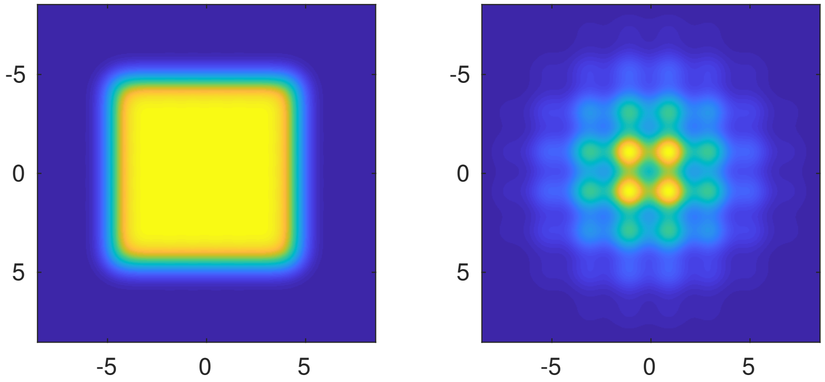
12,5 dB



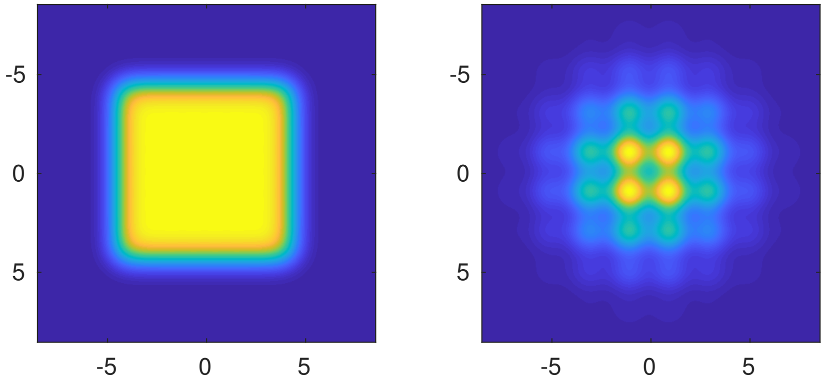
12,0 dB



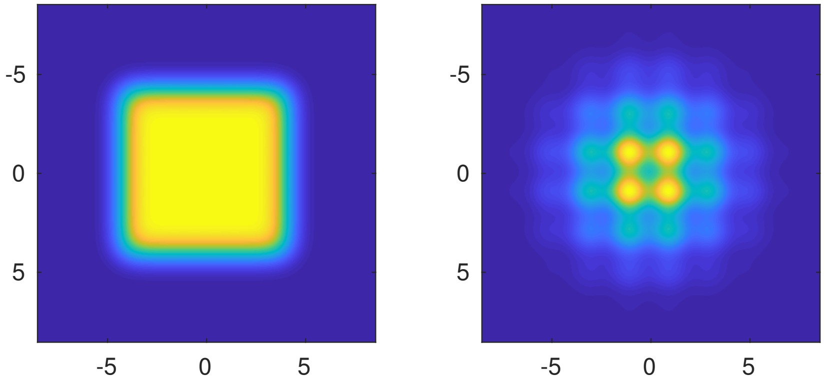
11,5 dB



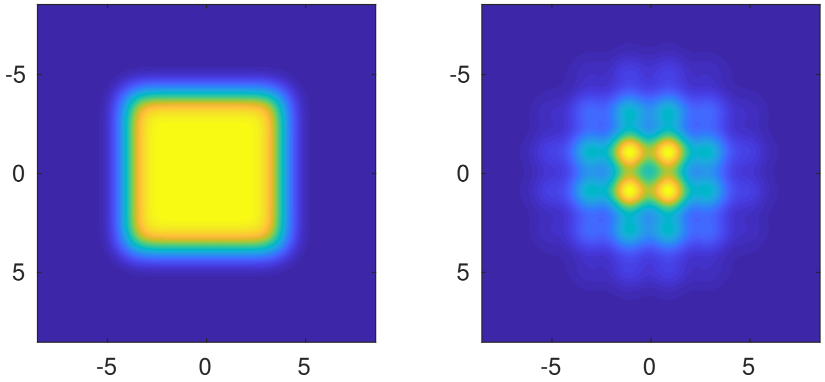
11,0 dB



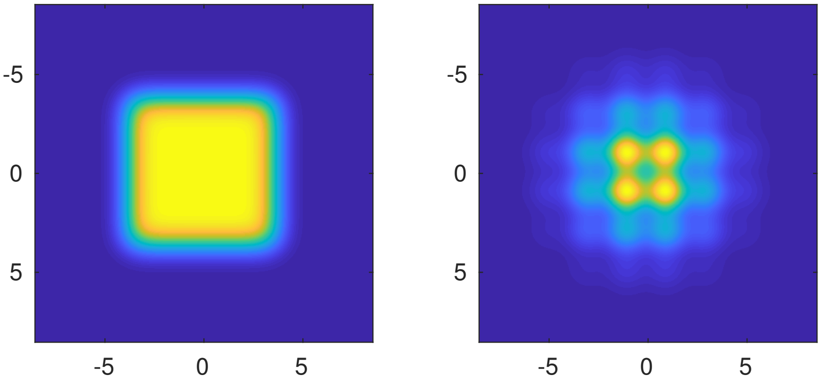
10,5 dB



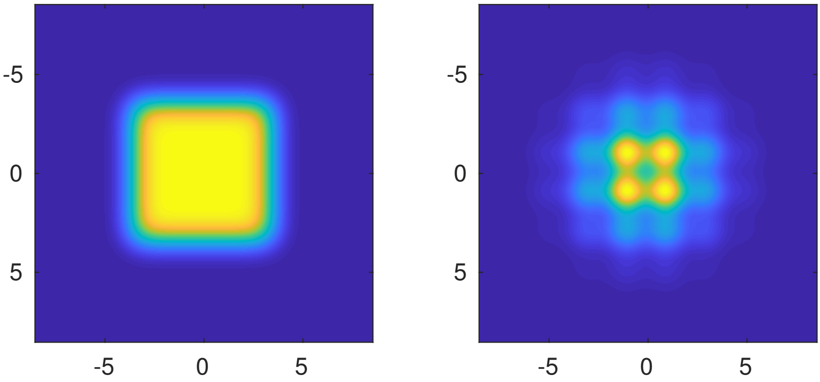
10,0 dB



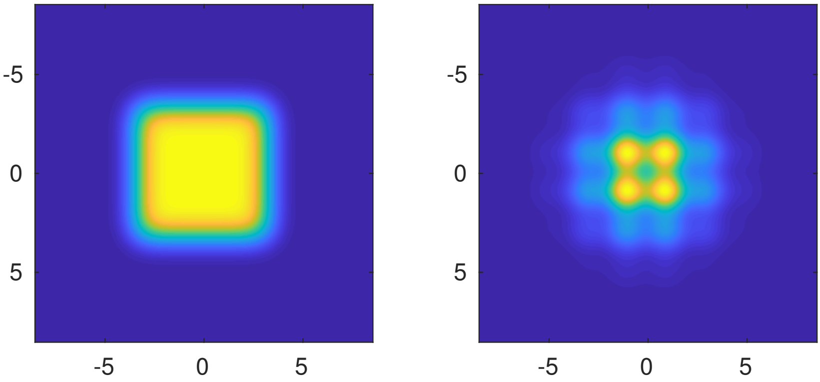
9,5 dB



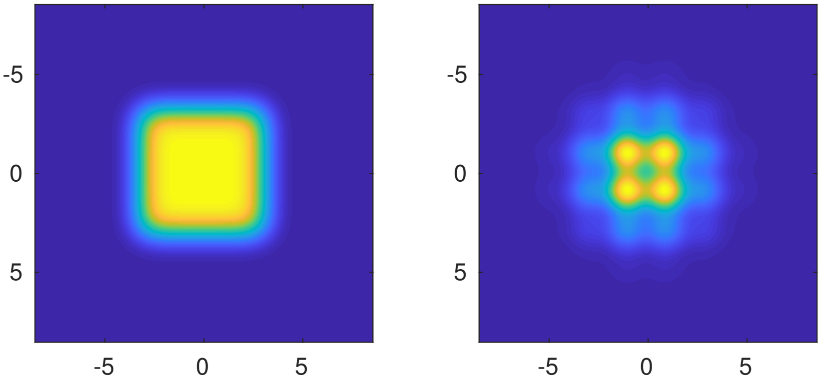
9,0 dB



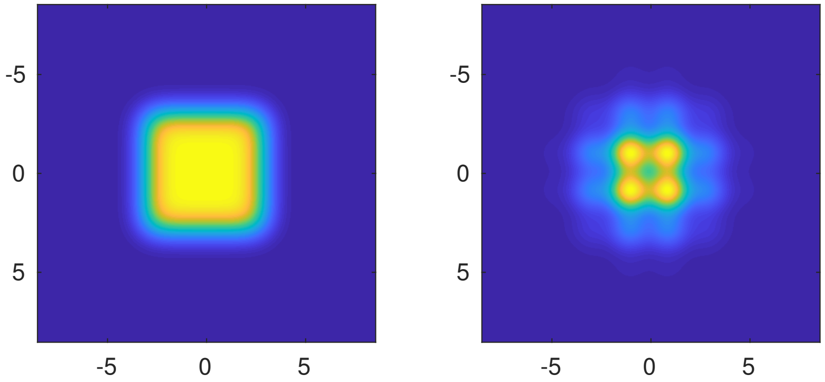
8,5 dB



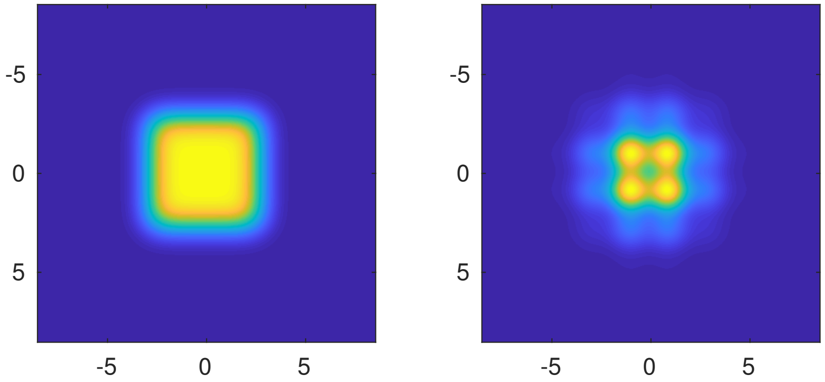
8,0 dB



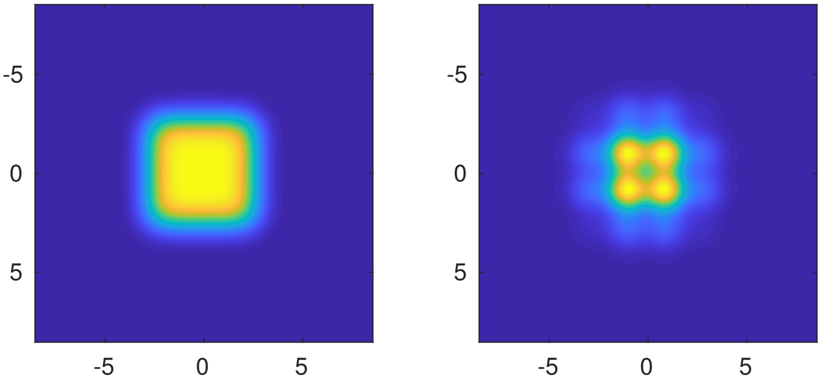
7,5 dB



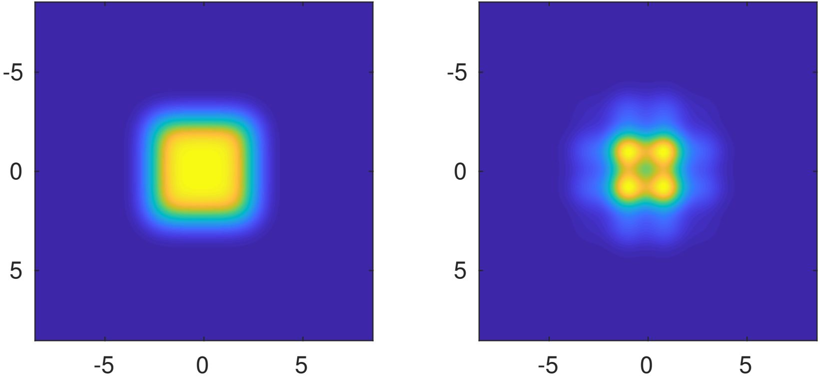
7,0 dB



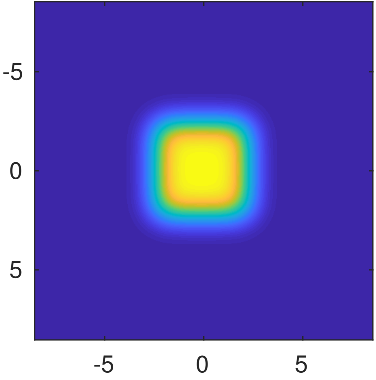
6,5 dB

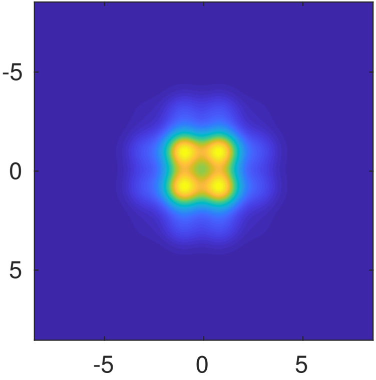


6,0 dB

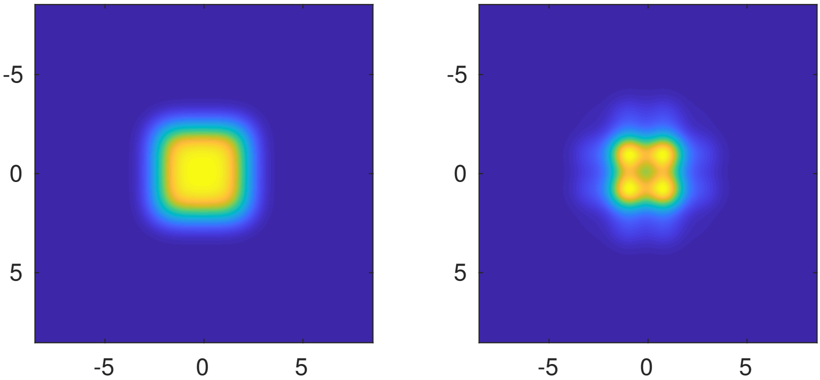


5,5 dB

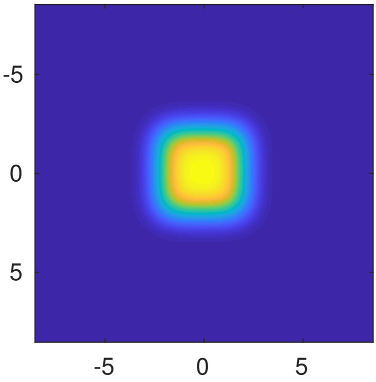


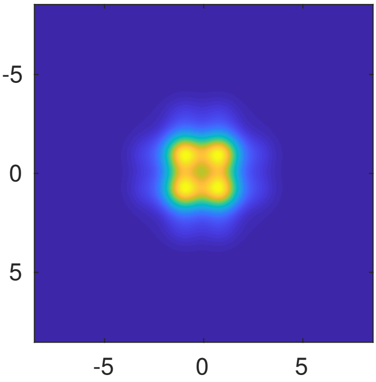


5,0 dB

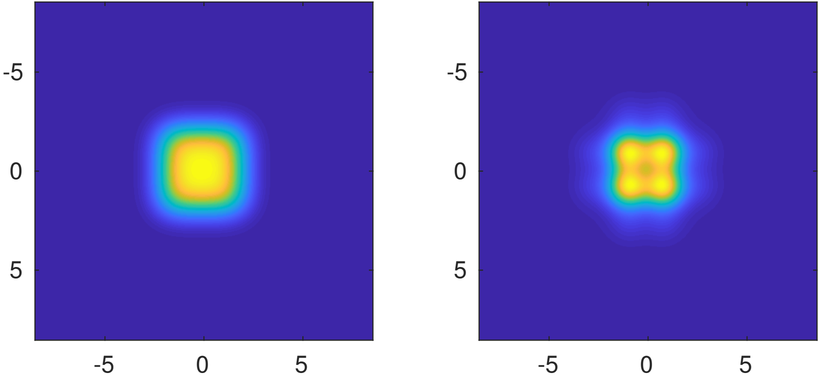


4,5 dB

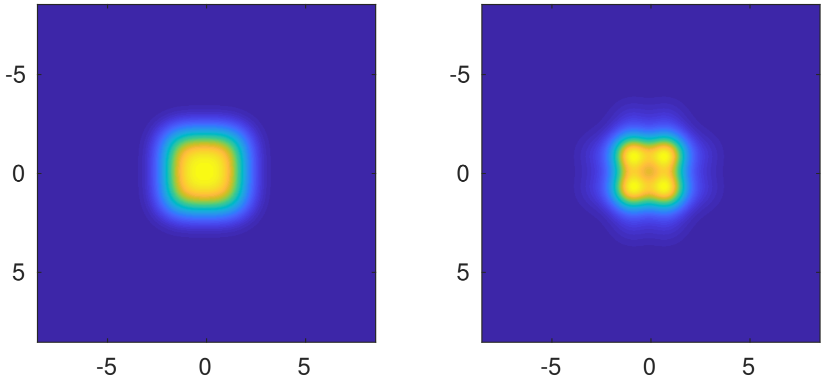




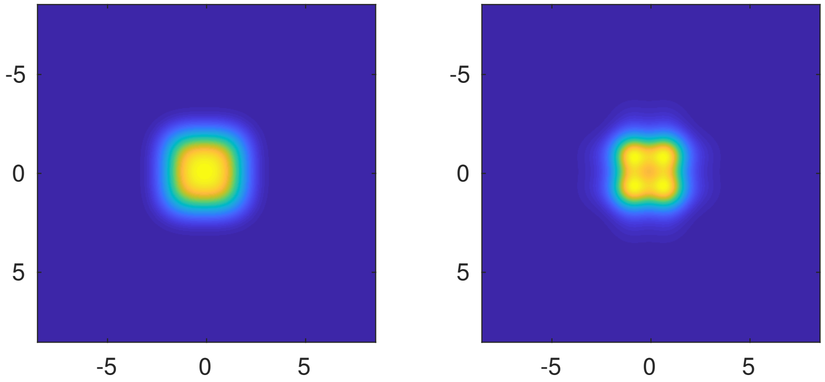
4,0 dB



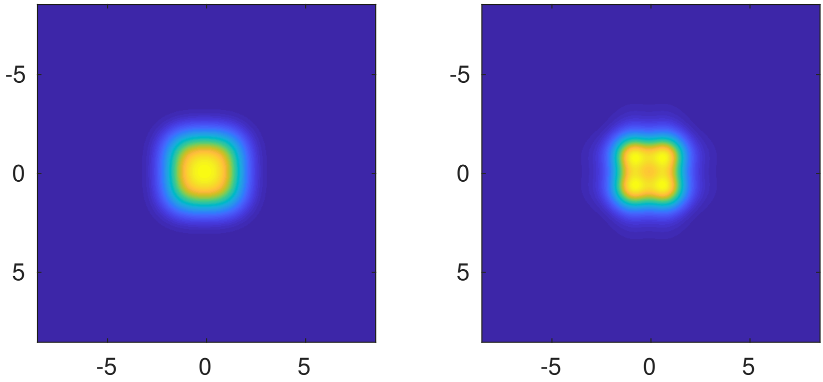
3,5 dB



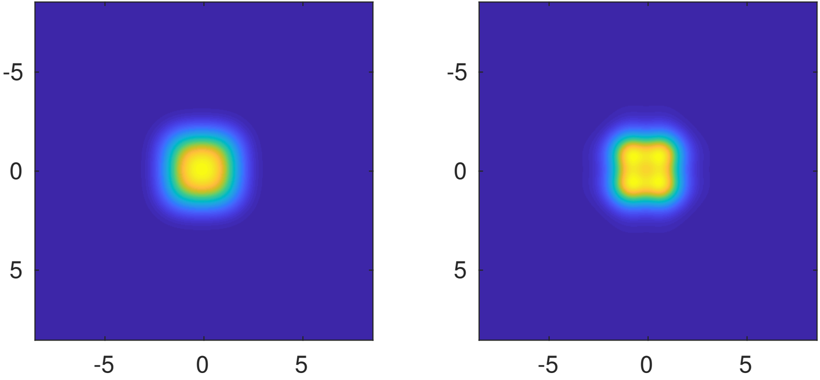
3,0 dB



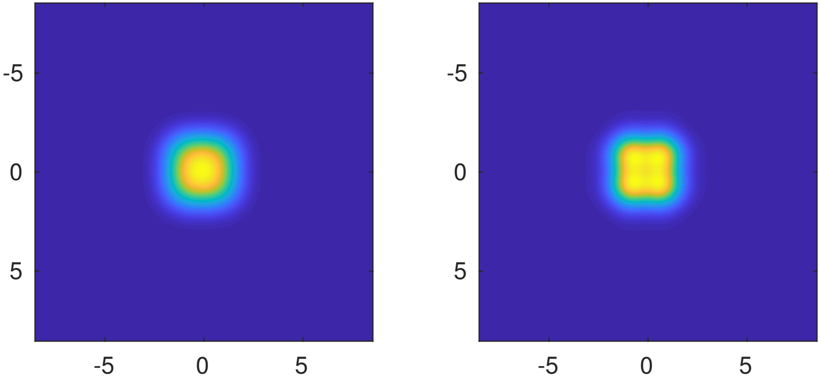
2,5 dB



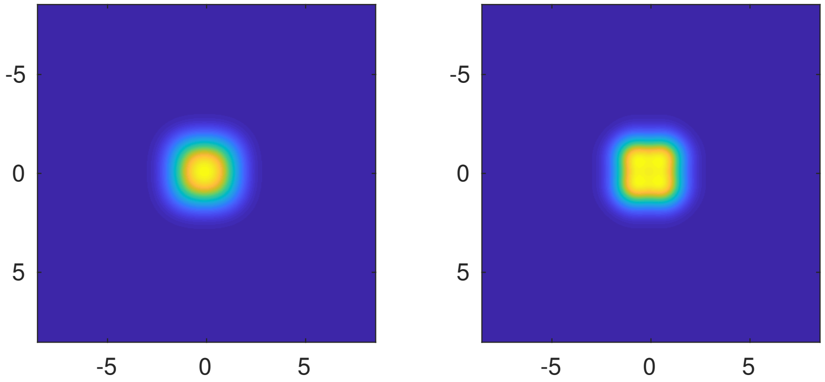
2,0 dB



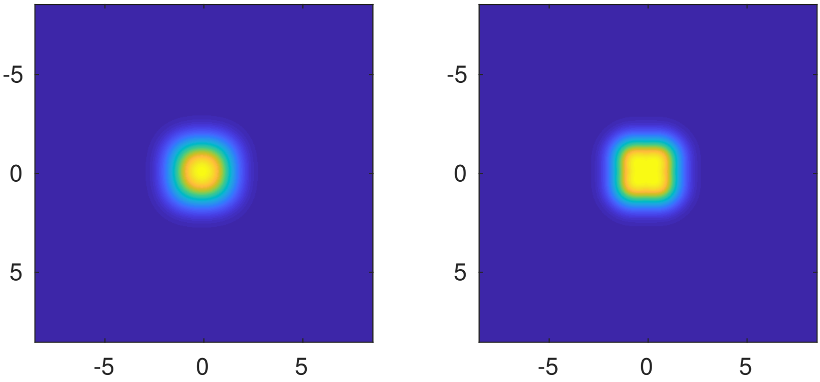
1,5 dB



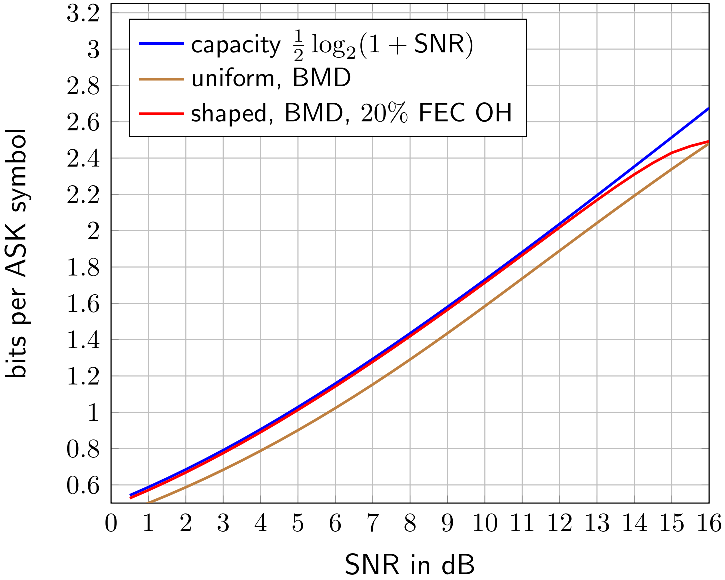
1,0 dB



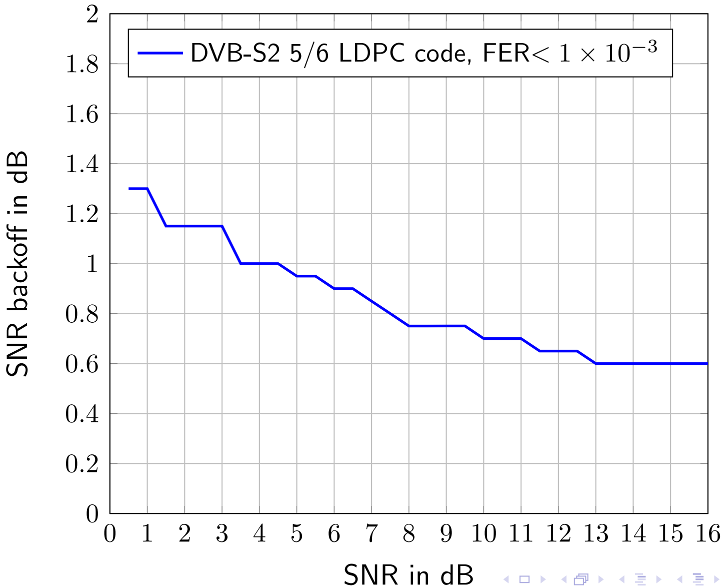
0,5 dB



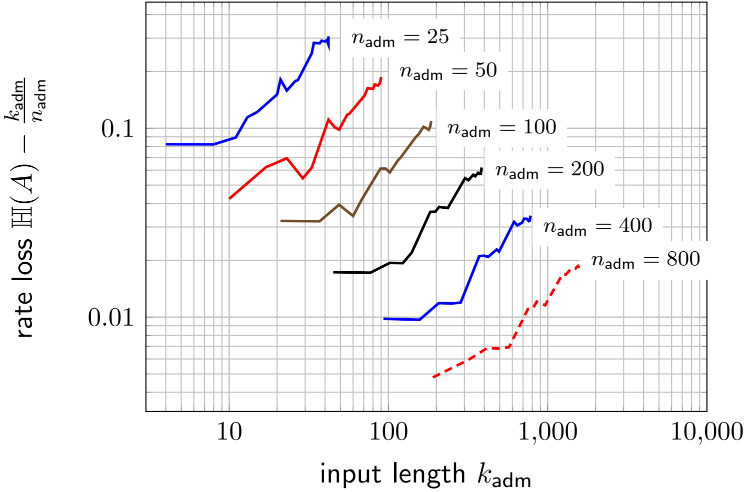
Achievable Rates for PAS with 20% FEC Overhead



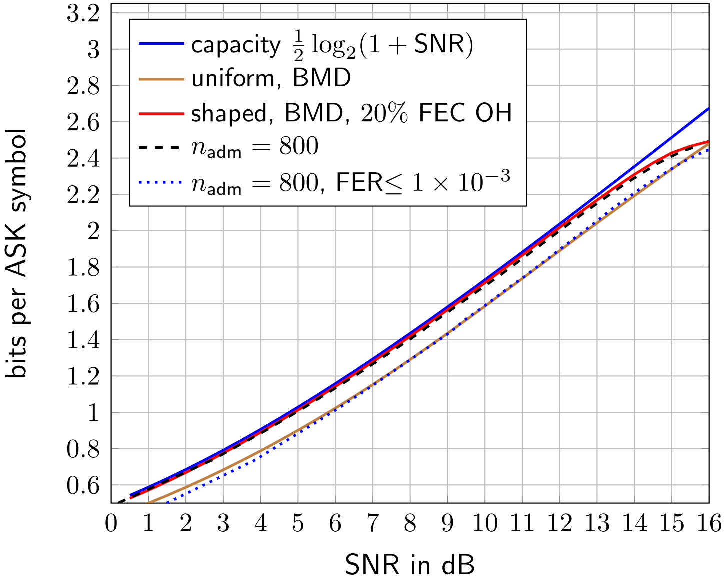
SNR Backoff for 5/6 DVB-S2 LDPC Code at FER$< 1 \times 10^{-3}$



DM Rate Loss


$n_{\text{adm}} = 25$
$n_{\text{adm}} = 50$
$n_{\text{adm}} = 100$
$n_{\text{adm}} = 200$
$n_{\text{adm}} = 400$
$n_{\text{adm}} = 800$
0.1
0.01
10
100
1,000
10,000
rate loss $\mathbb{H}(A) - \frac{k_{\text{adm}}}{n_{\text{adm}}}$
input length k_{adm}

5/6 DVB-S2 LDPC Code, CCDM with $n_{\text{adm}} = 800$



Outline

Conclusions

Presented **Framework** for individually benchmarking

- Decoding metrics: soft/hard/quantized/...
- FEC codes.
- DM algorithms.

Open problem:

- PAS-like architecture for non-coherent transmission.