Finding the Capacity of a Quantized Binary-Input DMC

Brian M. Kurkoski
Japan Advanced Institute of Science and Technology

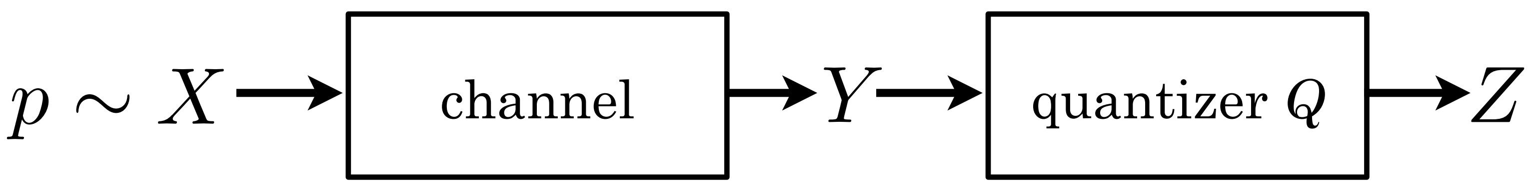
and

Hideki Yagi
University of Electro-Communications
Tokyo, Japan

July 2, 2012
International Symposium on Information Theory
MIT, Cambridge, Massachusetts, USA

Information Theoretic Limits on Channel Quantization

- Detection and decoding algorithms: algorithms with real-valued numbers
- VLSI implementation: numbers are converted to bits — quantization
- Power consumption, cost, etc. of receivers increases with the number of bits
- **Broad goal**: Reduce the number of bits without sacrificing performance.
  - What are the information theoretic limits?



Two simpler questions:

1. Find channel quantizer to maximize the information rate

$$Q^* = \arg\max_{Q} I(X;Z)$$

2. Find the quantized channel capacity

$$p^*, Q^* = \arg\max_{p,Q} I(X;Z)$$

Quantization of Continuous-Output Channels

- 1960s and 70s: The **cut-off rate** as a criterion for channel quantizer design
- **Mutual information** is a better criterion (capacity-achieving LDPC codes)

1. Channel Quantization

- BI-AWGN quantized to 3 levels [Ma et. al, 2002]
- "Locally optimal" quantization algorithm [Liveris and Georghiades 2003]
- Quantization of flash memory channel models [Wang et al 2011]

2. Quantized Channel Capacity

- Singh et al 2007, 2009: Continuous channels quantized to K levels
  - input distribution: K or $K+1$ discrete levels sufficient
  - locally optimal/brute force channel quantization
- Quantization of channels with memory [Zeitler, Singer and Kramer 2010, 2011]

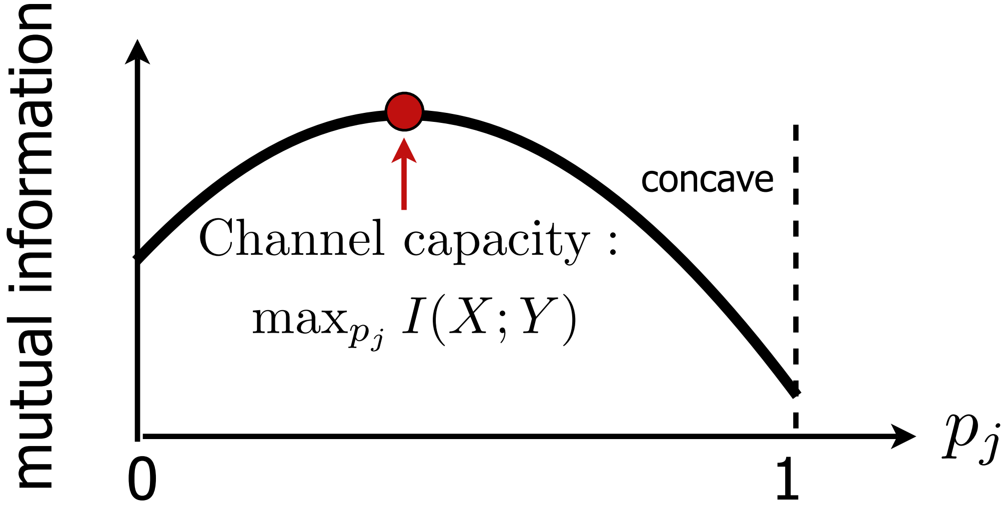
These are hard problems! Few proofs of optimality

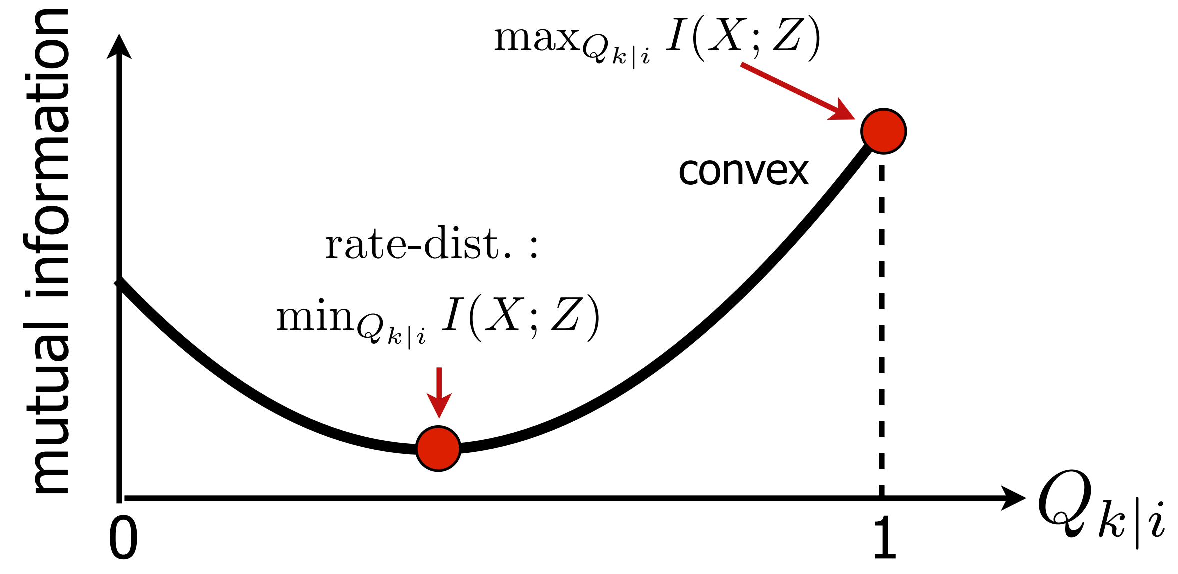
(Highly Simplified)

Channel Quantization is Concave Optimization

Mutual information $I(X;Z)$: function of

- input distribution p
- and quantizer Q

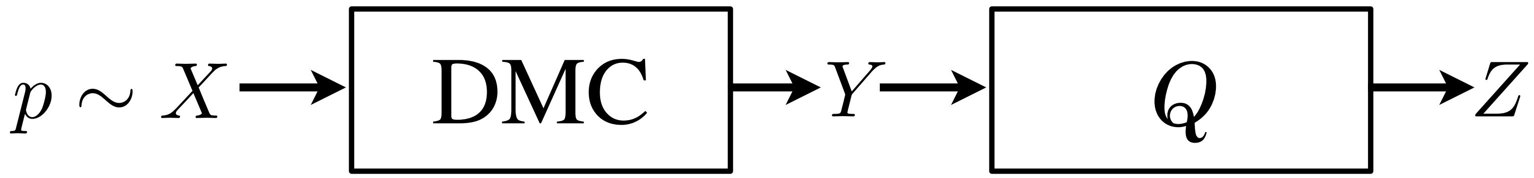




Optimization conditions	Known as	Class of problem
Maximize over p, fixed channel	Classical channel capacity	convex optimization Arimoto-Blahut algorithm [1972]
Fixed p, minimize over Q	Classical rate-distortion	
Fixed p, maximize over Q	1. Channel quantization	concave optimization, NP hard
Maximize over p and over Q	2. Quantized channel capacity	convex-concave optimization

Contributions of This Work

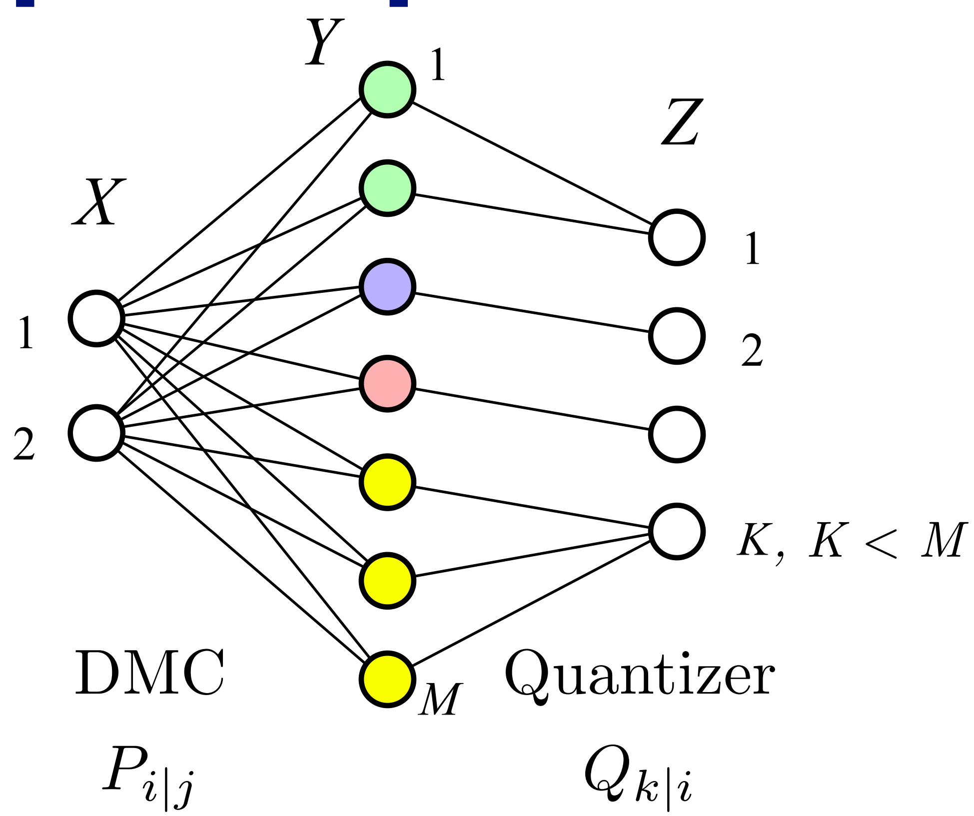
Consider discrete memoryless channels, rather than continuous channels



1. Channel Quantization
   - Maximize mutual information $I(X;Z)$, provably optimal
   - Polynomial-complexity algorithm, dynamic programming approach
2. Quantized Channel Capacity
   - Find the jointly optimal input distribution/quantizer or declare failure
   - also polynomial complexity

Applies to arbitrary DMCs with **binary inputs**

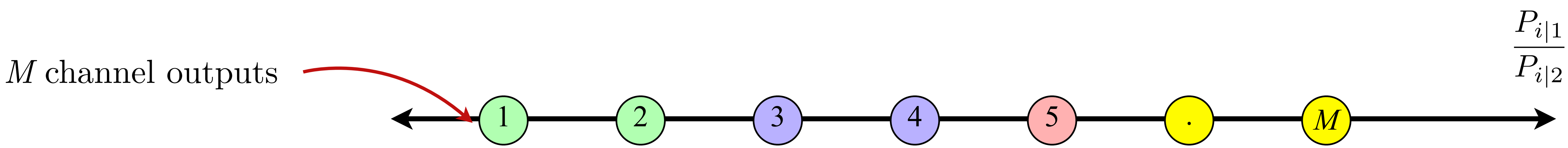
Optimal quantizer is convex



Consider a DMC with $P_{i|j}$. Goal: $\max I(X;Z)$
The channel outputs are points in a 1-D space

$$\frac{P_{i|1}}{P_{i|2}},$$

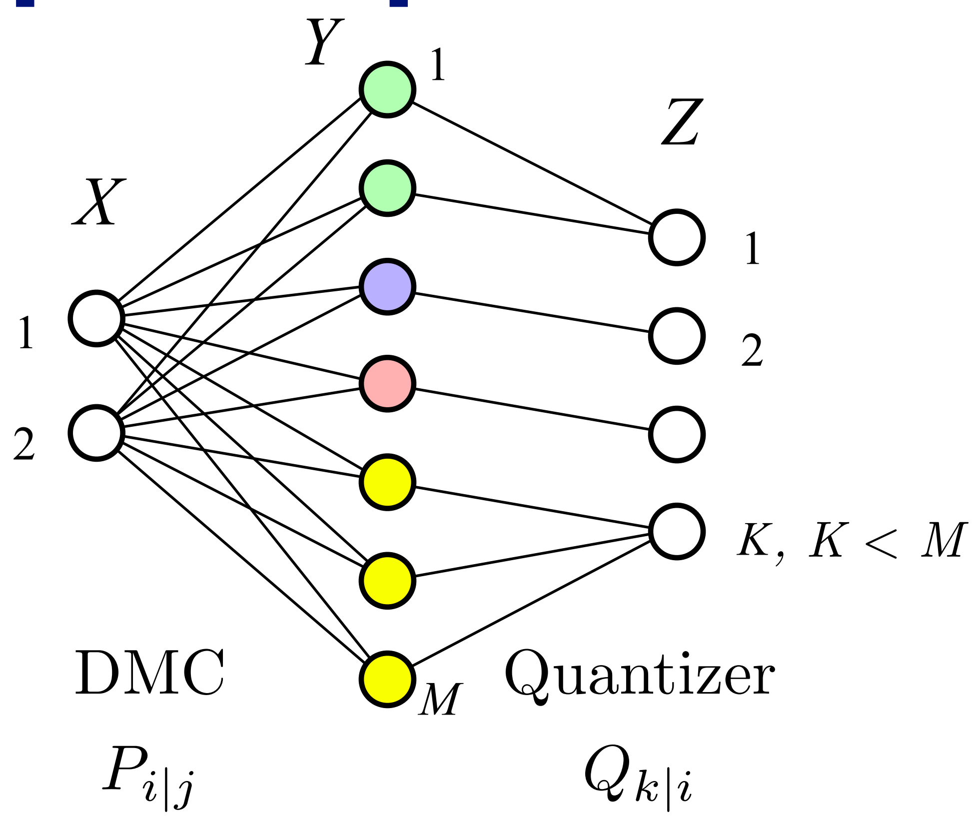
for $i = 1, 2, \ldots, M$



Lemma: If the outputs are labeled according to: $\frac{P_{1|1}}{P_{1|2}} < \frac{P_{2|1}}{P_{2|2}} < \cdots < \frac{P_{M|1}}{P_{M|2}}$,

then the mutual-information maximizing quantizer is convex on {1,2,..., M}.

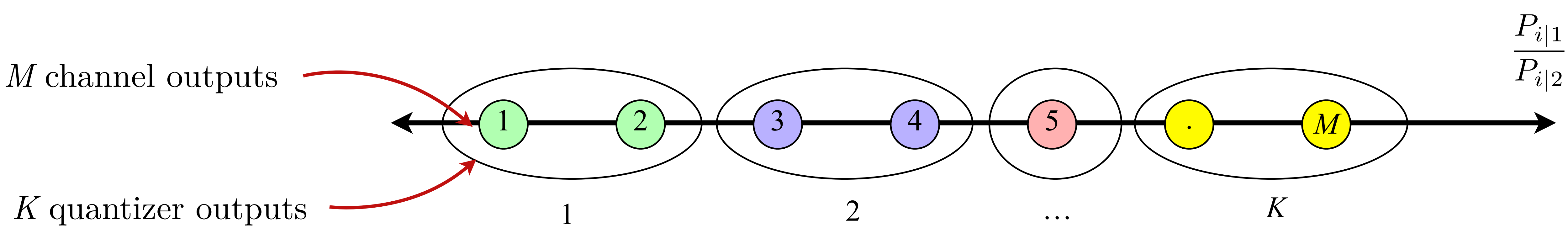
Optimal quantizer is convex



Consider a DMC with $P_{i|j}$. Goal: $\max I(X;Z)$
The channel outputs are points in a 1-D space

$$\frac{P_{i|1}}{P_{i|2}},$$

for $i = 1, 2, \ldots, M$



Lemma: If the outputs are labeled according to: $\frac{P_{1|1}}{P_{1|2}} < \frac{P_{2|1}}{P_{2|2}} < \cdots < \frac{P_{M|1}}{P_{M|2}}$,

then the mutual-information maximizing quantizer is convex on {1,2,..., M}.

Optimal Quantizer is Convex: Proof Sketch

Proof sketch: $$I(X;Z) = H(X) - \underbrace{H(X|Z)}$$

$$H(X|Z) = -\mathrm{E}_Z\Big[\mathrm{E}_{X|Z}\big[\log p(X|Z)\big]\Big]$$

Statistical/Machine Learning Theory

- Classification: from observation, make a classification (e.g. optical character recognition)
- Minimize some loss function (supervised learning)
- Impurity (or risk) is the expectation of an expectation of a loss function:

partition of the sample space. Hence the risk (1) can be rewritten

$$R(q) = \sum_{t \in \tilde{T}} P(t) \cdot E[\ell(\boldsymbol{Y}, \hat{\boldsymbol{y}}(t)) \mid t], \qquad (2)$$

[Chou, 1991]

Broad class of loss functions, preimage of optimum classification mapping forms a convex set:

2. Results.

THEOREM 1. *For any* $C\colon \mathscr{X} \to \mathscr{C}$ *there exists a* $\tilde{C}\colon \mathscr{U} \to \mathscr{C}$ *such that* $\Psi(\tilde{C}(Y)) \le \Psi(C)$ *and such that* $\tilde{C}^{-1}(c)$ *is convex for all* $c \in \mathscr{C}$.

[Burshtein et al., 1992]

Identify $H(X|Z)$ as an impurity. Quantizer (mapping) image forms a convex set.

1. Channel Quantization
Possible algorithmic approaches

Algorithms (M channel outputs, K quantizer outputs):

1. Brute force (ignore convexity): complexity K^M
2. Search all convex sets. Complexity:

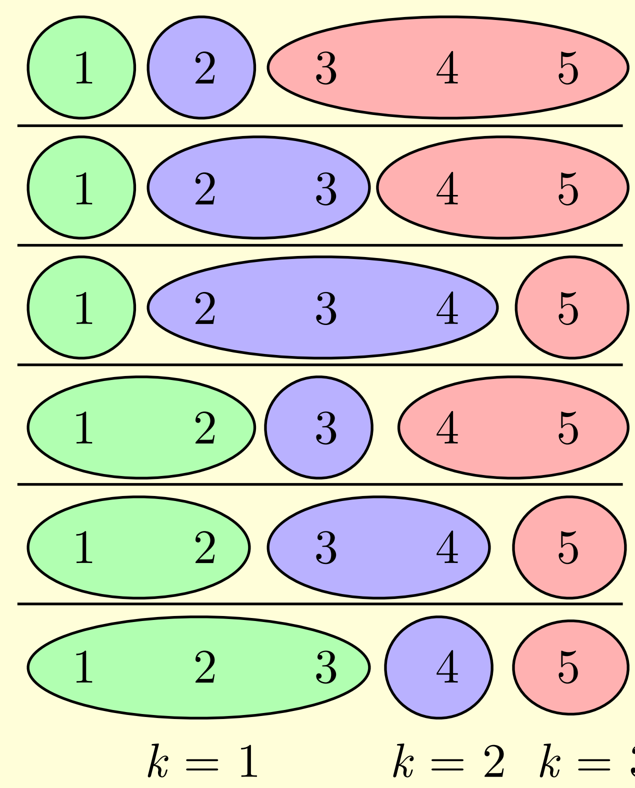
$$\binom{M-1}{K-1}$$

   - Worst-case complexity: exponential in K

3. *Proposed* Dynamic programming approach:
   - Avoid recomputing part of the total sum
   - Parts are called "partial mutual information"
   - Complexity M^3

example: $M = 5$ channel outputs,
$K = 3$ quantizer outputs

$$\binom{4}{2} = 6$$



$$I(X;Z) = \sum_{k=1}^{K} \left(\sum_{j=1}^{2} p_j T_{k|j} \log \frac{T_{k|j}}{\sum p'_j T_{k|j}} \right)$$

$$I(X;Z) = \sum_{k=1}^{K} (\iota_k)$$

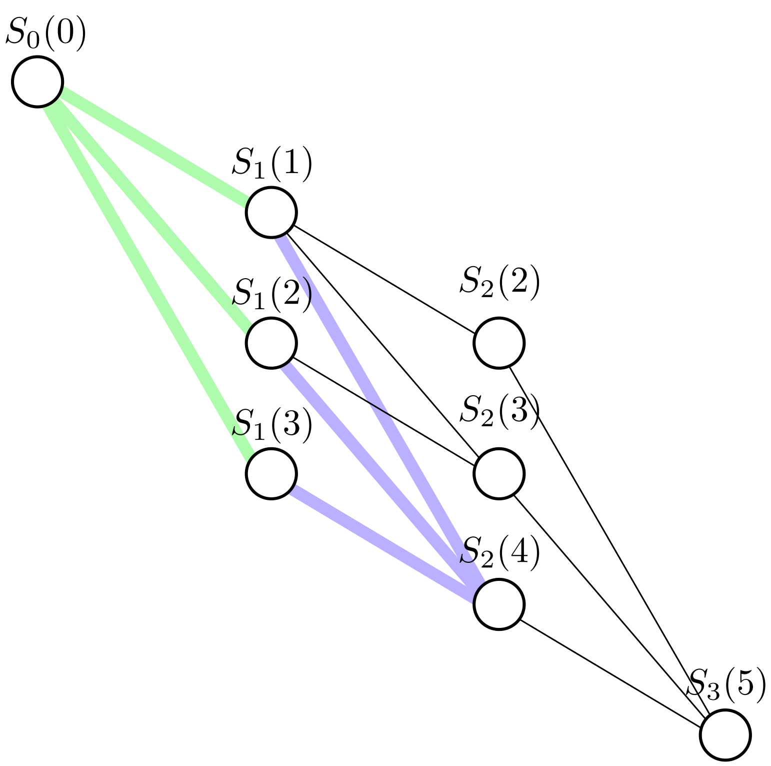
1. Channel Quantization
Algorithm for Optimal quantization

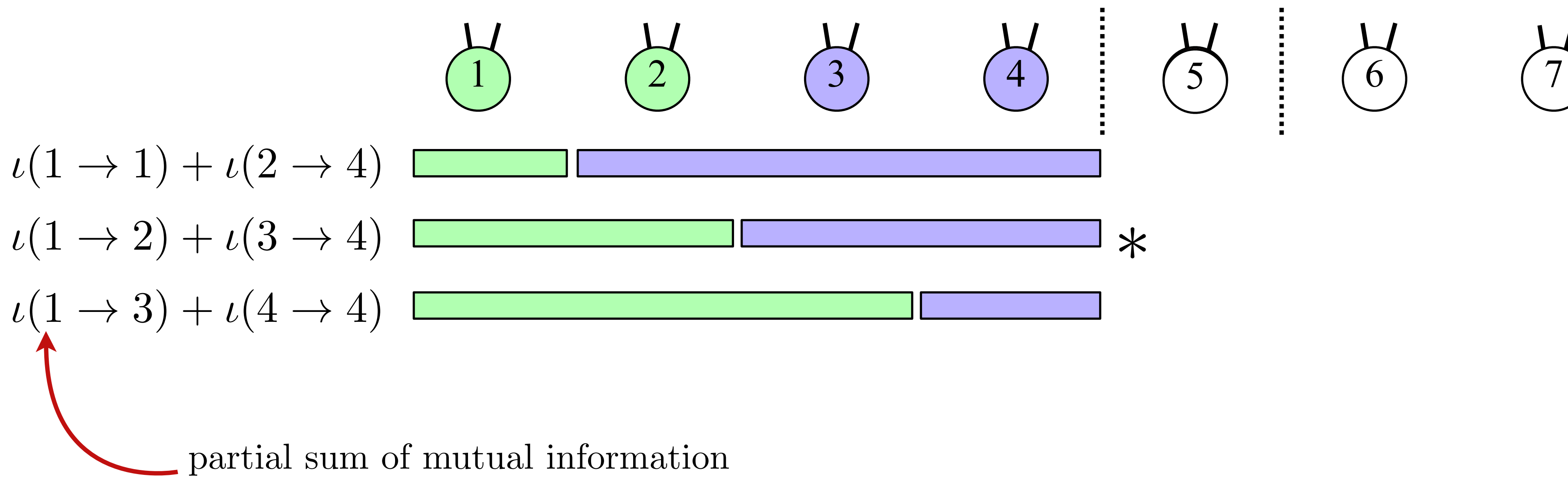
Dynamic programming: Optimal solution contains optimal solutions to subproblems

Subproblem find optimal quantization of ch. outputs 1...m to quantizer outputs 1...k

- Key: use the optimal quantization 1...k-1 to find optimal quantization of 1...k

Example: how to quantize outputs 1-4 to two values





$$\sum_{i=a'+1}^{a} \sum_{j} p_j P_{i|j} \log \frac{\sum_{i'=a'+1}^{a} P_{i'|j}}{\sum_{j'} p_{j'} \sum_{i'=a'+1}^{a} P_{i'|j'}}.$$

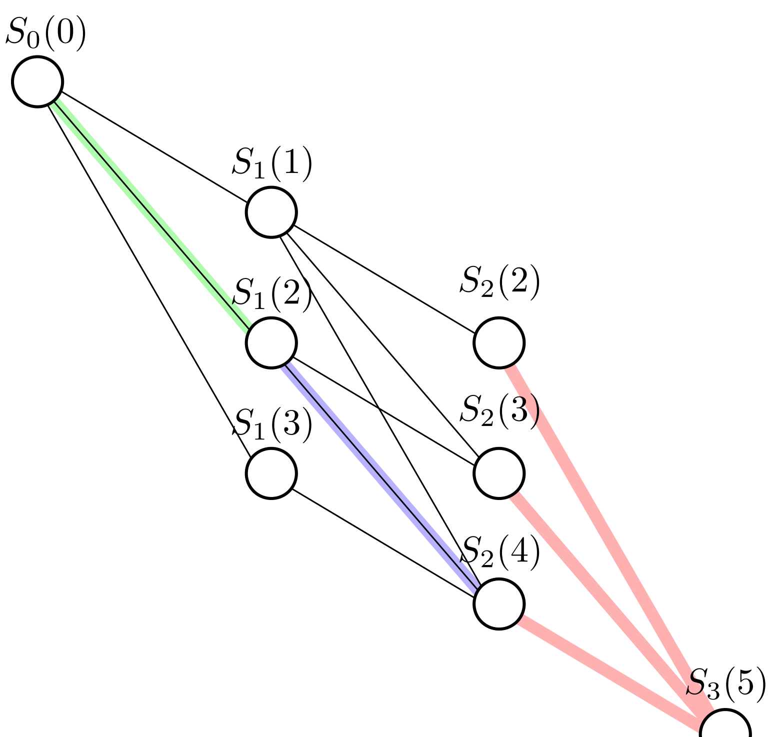
1. Channel Quantization
Algorithm for Optimal quantization

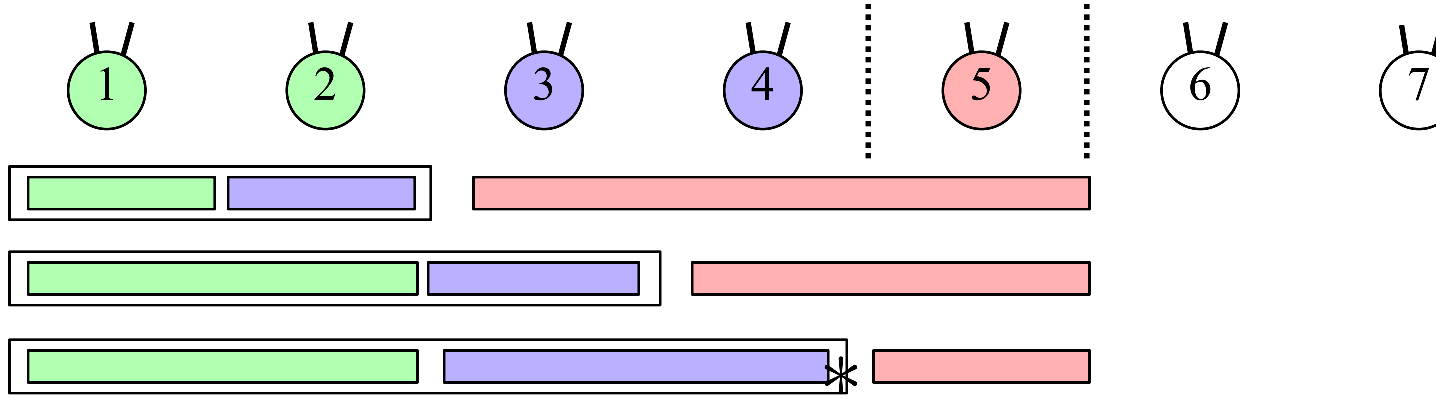
Dynamic programming: Optimal solution contains optimal solutions to subproblems

Subproblem find optimal quantization of ch. outputs $1 \ldots m$ to quantizer outputs $1 \ldots k$

- Key: use the optimal quantization $1 \ldots k$-1 to find optimal quantization of $1 \ldots k$

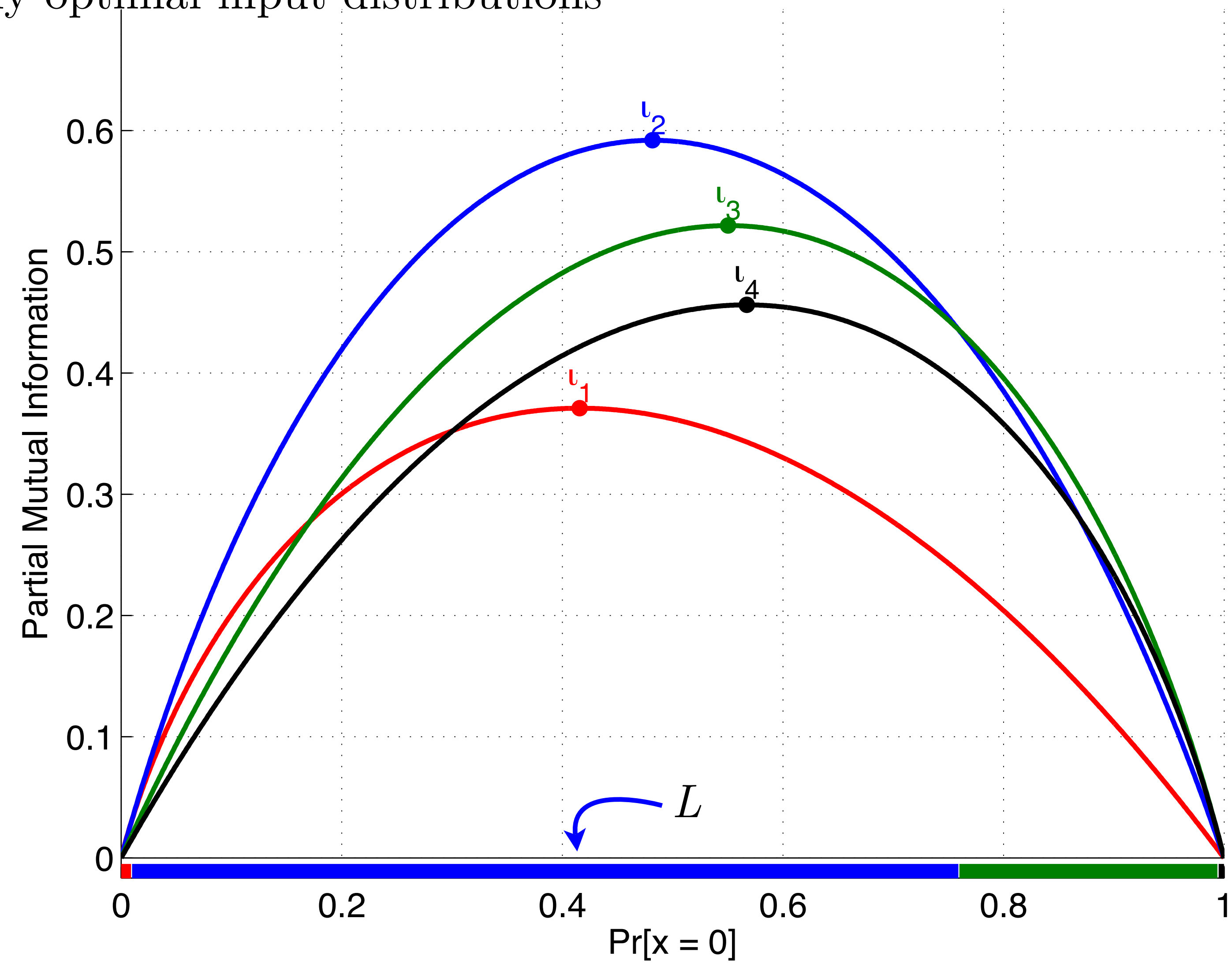
Example: how to quantize outputs 1-5 to three values





2. Quantized Channel Capacity
Jointly Optimal Quantizer and Input Distribution

- Similar procedure to previous. For m outputs quantized to k levels:
  - for each quantizer: maximum partial mutual information
  - select the quantizer with the greatest partial mutual information
  - for this one, find a range L of locally optimal input distributions



- This range is important

  (1) if final p* is range L,
  then it is known optimal.

  (2) otherwise, another quantizer
  may be optimal
  -> Declare a failure

- Continue recursively:
  - locally optimal quantizer **and**
  - the range of input distributions

2. Quantized Channel Capacity
Jointly Optimal Quantizer and Input Distribution

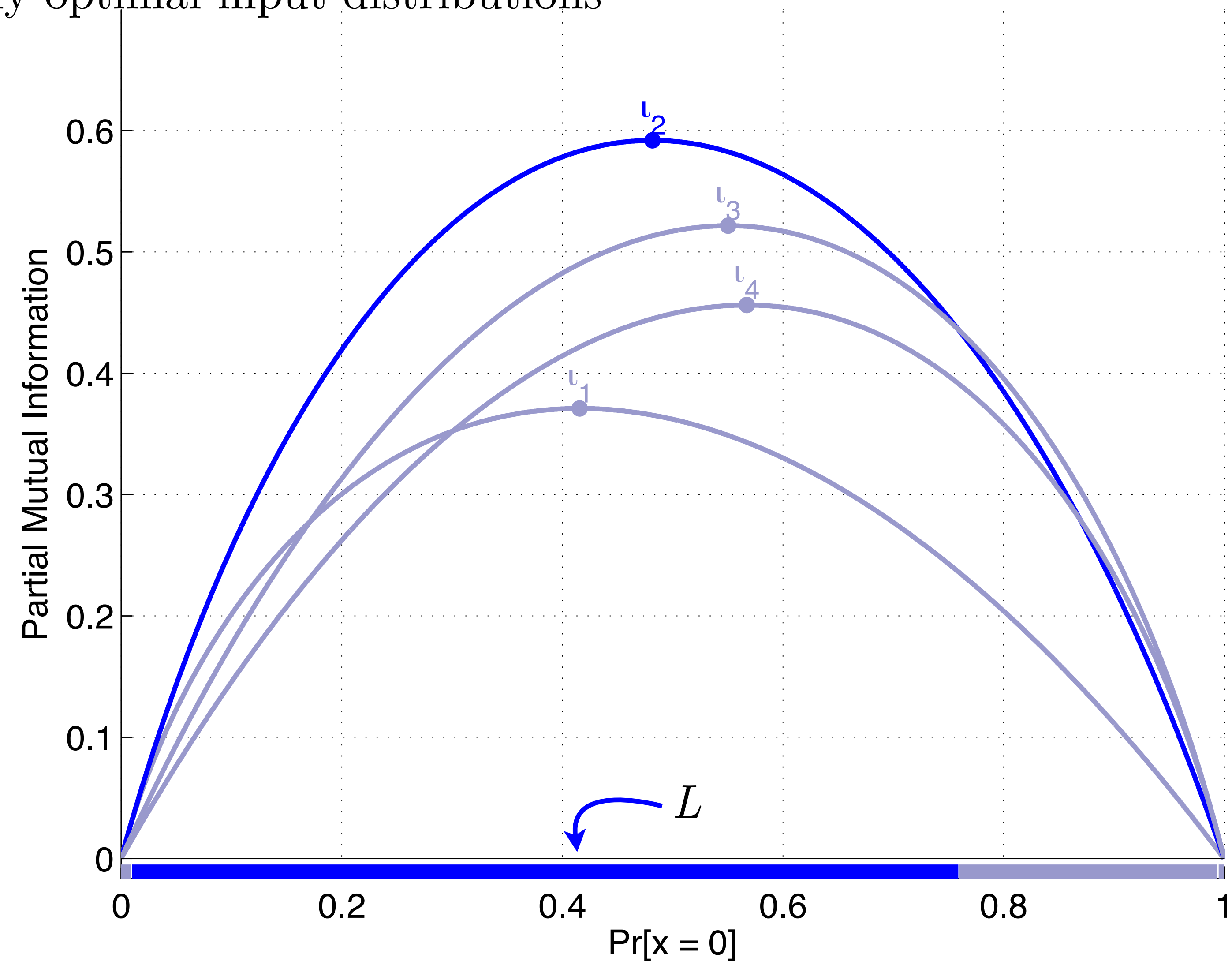
- Similar procedure to previous. For m outputs quantized to k levels:
  - for each quantizer: maximum partial mutual information
  - select the quantizer with the greatest partial mutual information
  - for this one, find a range L of locally optimal input distributions

- This range is important

  (1) if final p* is range L,
  then it is known optimal.

  (2) otherwise, another quantizer
  may be optimal
  -> Declare a failure

- Continue recursively:
  - locally optimal quantizer **and**
  - the range of input distributions



2. Quantized Channel Capacity

Example: Optimal Quantizers

Create a DMC:

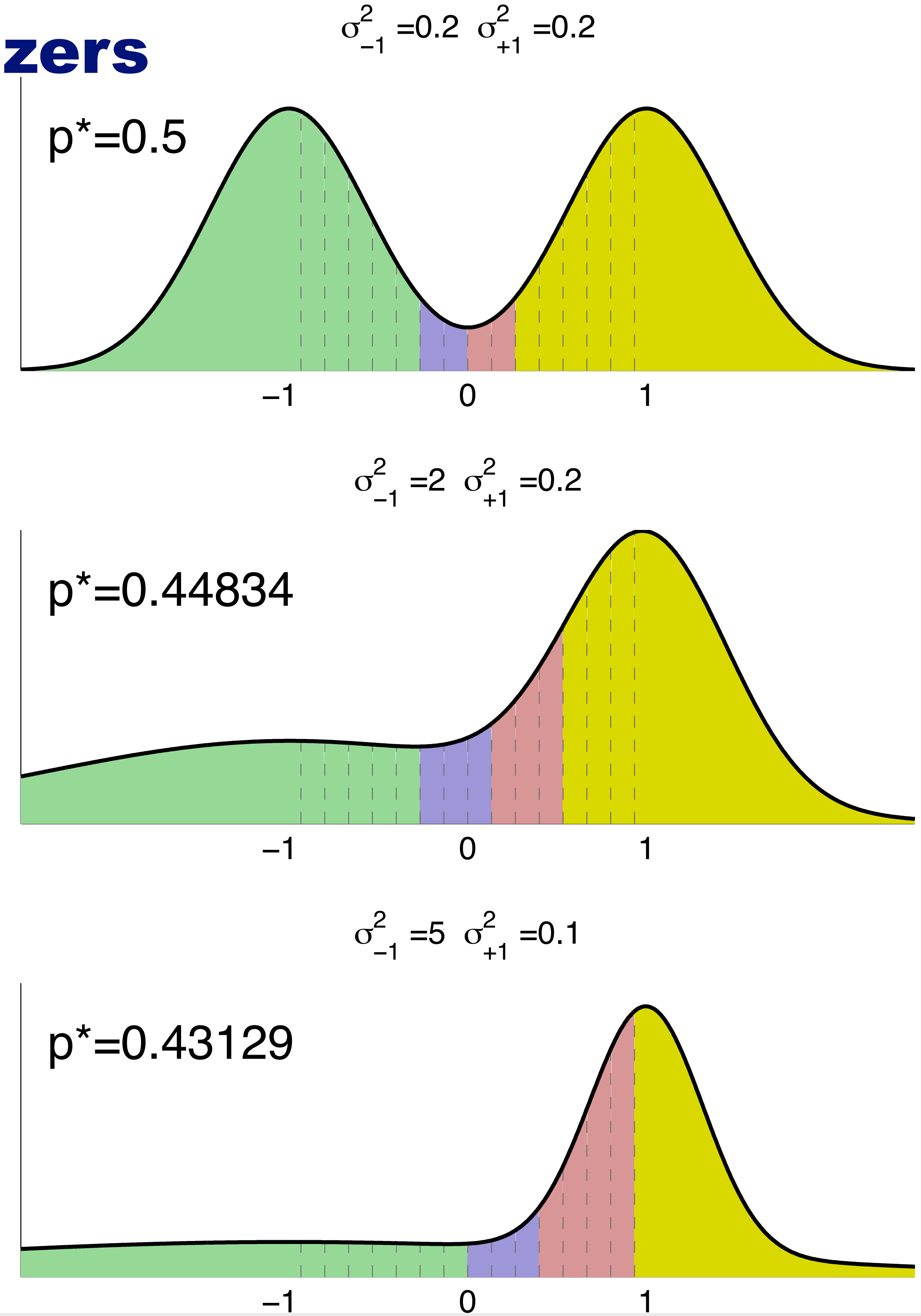
- BPSK +1/–1
- data-dependent AWGN noise
- quantize uniformly M levels

Example:

- $M = 16$ output channel
- quantize to $K = 4$ levels

Observations:

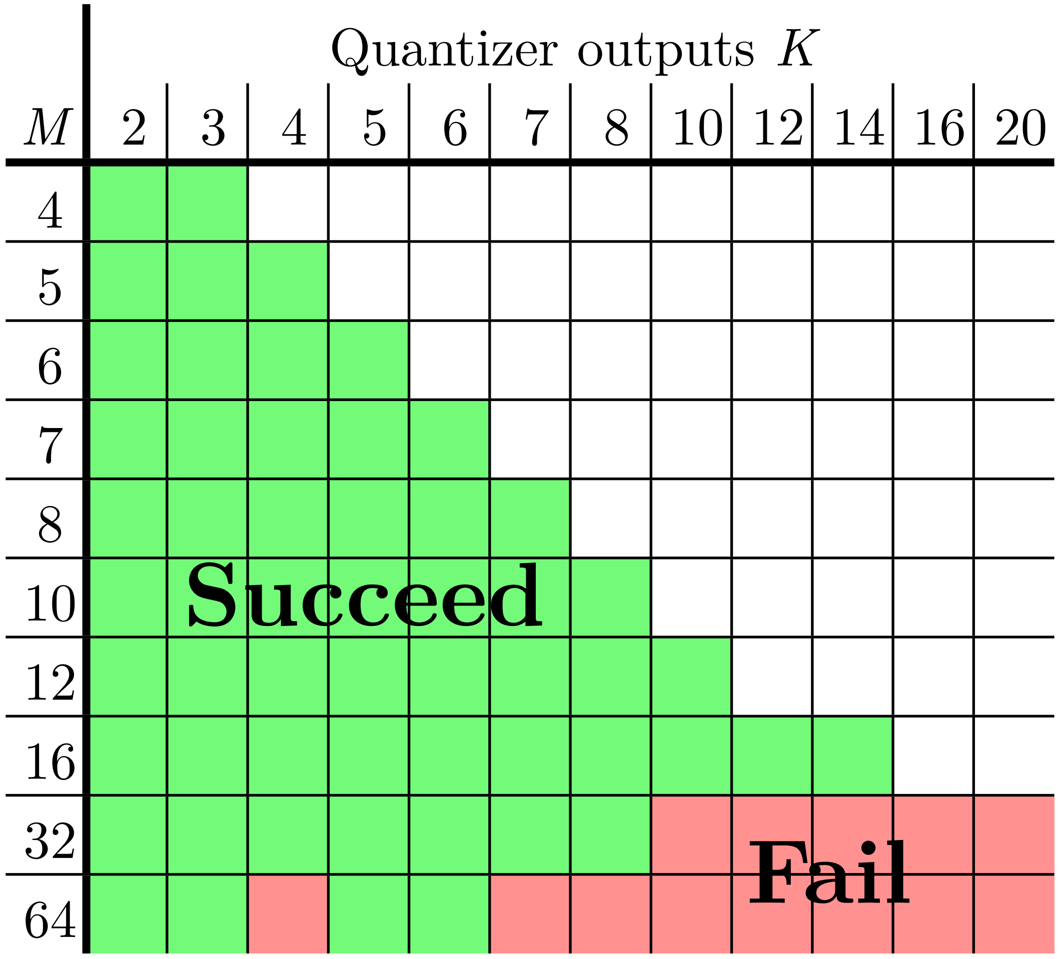
- data-independent noise:
  decisions cluster at cross-over
- data-dependent noise:
  decisions move towards reliable data

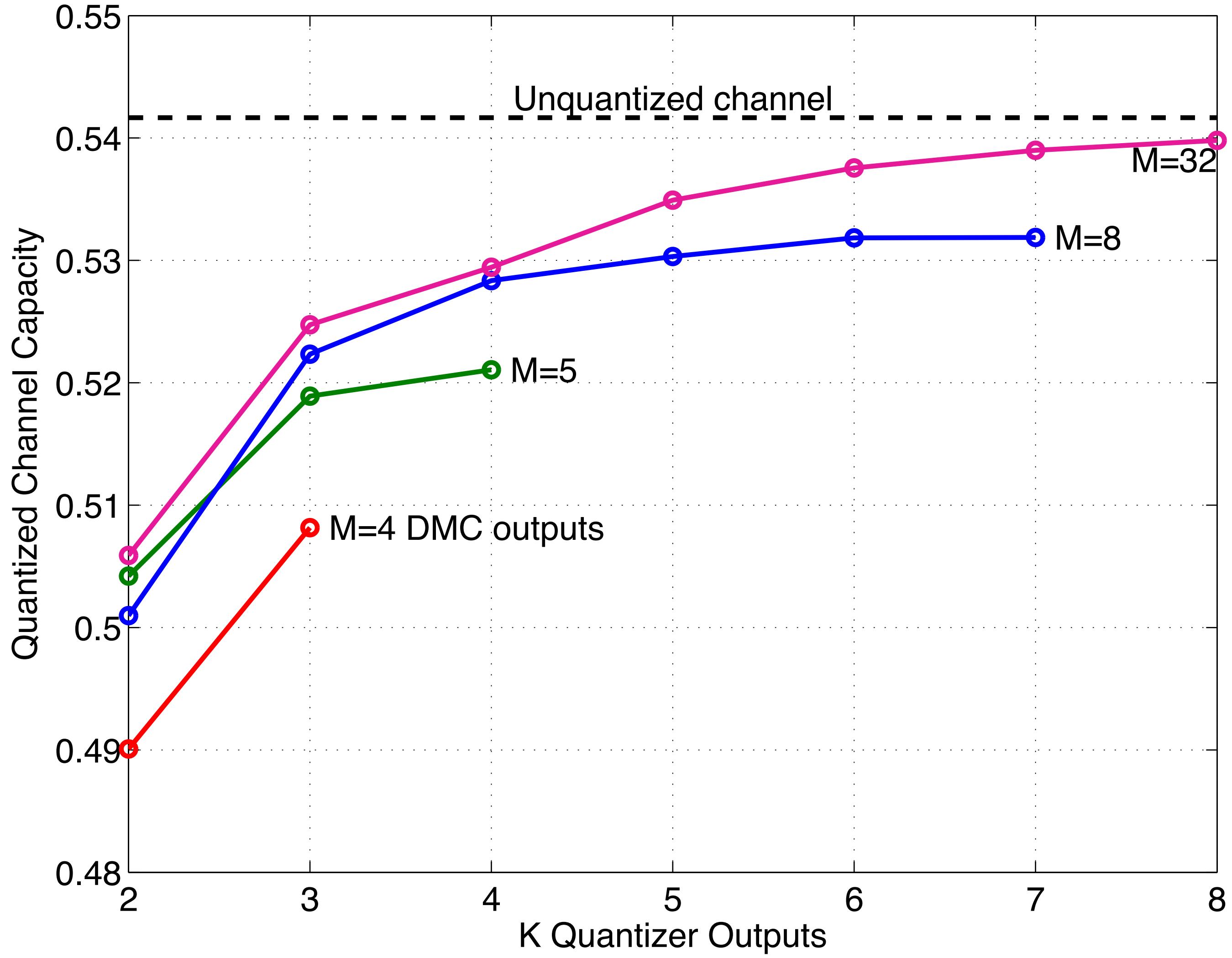


2. Quantized Channel Capacity
Example: Capacity

Asymmetric channel (var 0.1 & 4)

- Capacity increases in K, M
- Asymmetric optimal quantizer





Failure to converge

- Large M: Many good candidate quantizers, so optimal range is narrow

Conclusion

Channel quantization is important for reducing complexity of receivers

- Maximization of mutual information is a highly suitable metric
- These concave optimization problems are NP-Hard

Easier to work with discrete problems than continuous problems

For arbitrary binary-input DMCs:

- 1. Channel quantization (fixed input distribution):
  - Maximize mutual information, provably optimal
  - Polynomial (cubic) complexity
- 2. Quantized channel capacity
  - Find the jointly optimal input distribution/quantizer or declare failure
  - Also polynomial-complexity