Optimal Delayed Decisions in Decoding of Predictively Encoded Sources

Vinay Melkote and Kenneth Rose

Signal Compression Lab
Department of Electrical and Computer Engineering
University of California, Santa Barbara

Introduction

- Decoders simply reconstruct data, no parameter choices to make
- Can decoder delay, and thus accrued future coded data, improve current reconstruction?
- Feasible if adequate correlation exists between coded data units
- Predictive coding systems provide the right setting: assume an underlying correlation model for the source

Introduction

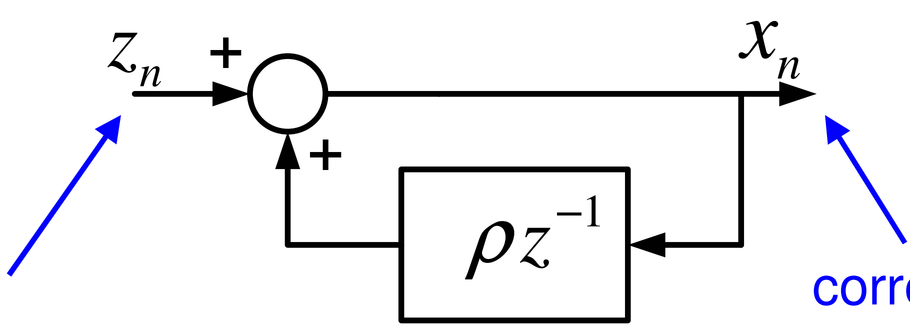
- Predictive coding widely employed in signal compression standards:
  - Motion-compensated video coding (H.264)
  - Speech coding via adaptive differential pulse code modulation (G.726, G.722)
  - Continuously variable slope delta modulation (Bluetooth hands-free profile)

- Attractive for low-delay/low-complexity applications

Introduction

- Assume a scalar first-order autoregressive (AR) source: a sequence of zero-mean random variables $\cdots x_{n-1}, x_n, x_{n+1} \cdots$ that evolve as

$$x_n = \rho x_{n-1} + z_n$$

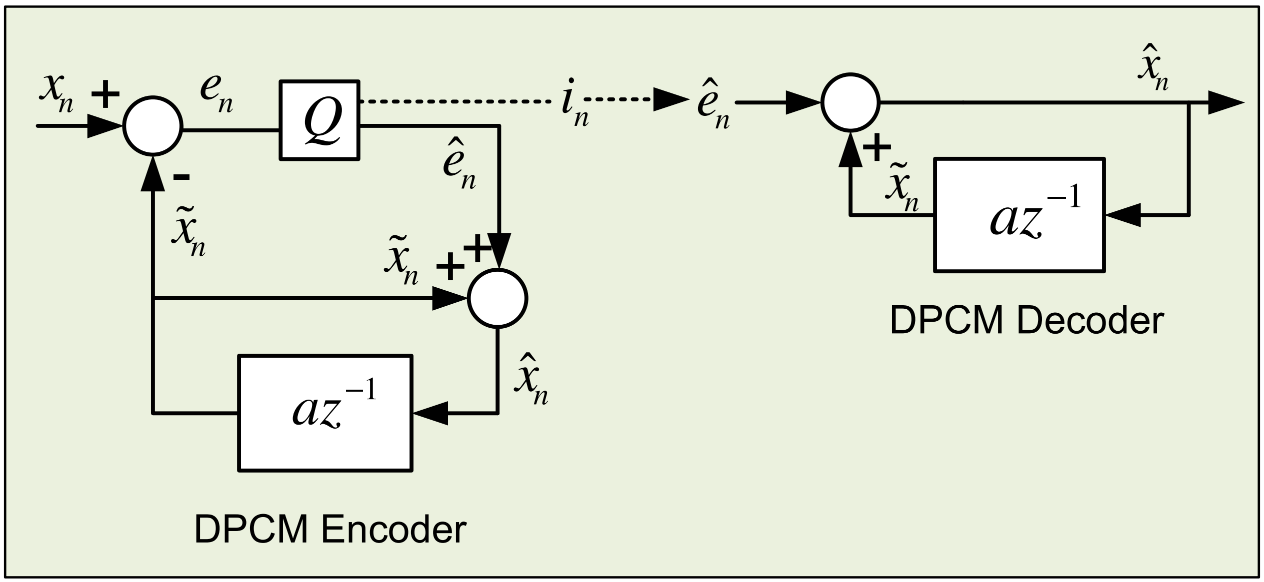


i.i.d innovations with zero-mean pdf $p_Z(z)$

correlated source samples with inter-sample correlation coefficient ρ

Introduction

- Consider coding $\cdots x_{n-1}, x_n, x_{n+1} \cdots$ with a differential pulse code modulation scheme (DPCM)



- The prediction here is $\tilde{x}_n = a\hat{x}_{n-1}$
- Generally, $a = \rho$, i.e., predictor *matched* to source

Introduction

- DPCM encoder and decoder operate at zero delay
- At asymptotically high bit-rates:
  - $\hat{x}_{n-1} \approx x_{n-1}$
  - Matched predictor $\tilde{x}_n = \rho \hat{x}_{n-1} \approx \rho x_{n-1}$ is optimal
  - $\therefore e_n = x_n - \rho \hat{x}_{n-1} \approx x_n - \rho x_{n-1} = z_n$
  - Hence indices $\cdots i_{n-1}, i_n, i_{n+1} \cdots$ are approximately i.i.d

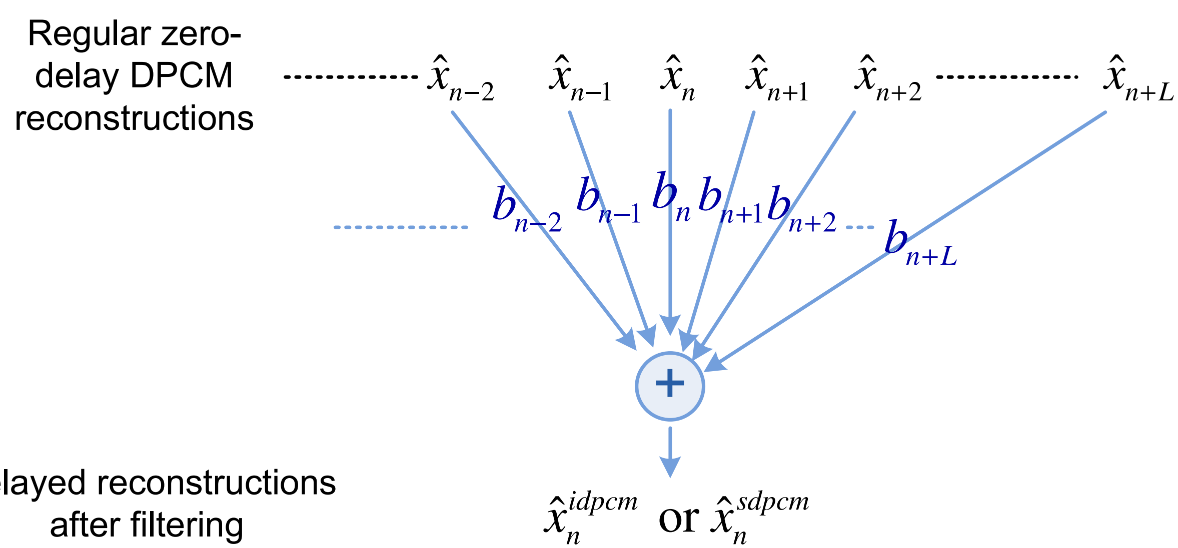
  - Future indices $i_{n+1}, i_{n+2}, \cdots$ provide no information on x_n
  - Zero-delay decoder optimal for the given encoder

Introduction

- At low bit-rates, prediction errors are *correlated*, and the indices $\cdots i_{n-1}, i_n, i_{n+1} \cdots$ as well
- Future indices $i_{n+1}, i_{n+2}, \cdots$ contain information on x_n
- Can this be exploited, by appropriate decoding delay, to improve the reconstruction of x_n ?

Prior work

- Interpolative DPCM (IDPCM) [Sethia & Anderson, '78] and Smoothed DPCM (SDPCM) [Chang & Gibson, '91]
- Apply a *non-causal* post-filter to smooth the zero-delay reconstructions: non-causality implemented by delay



Prior work

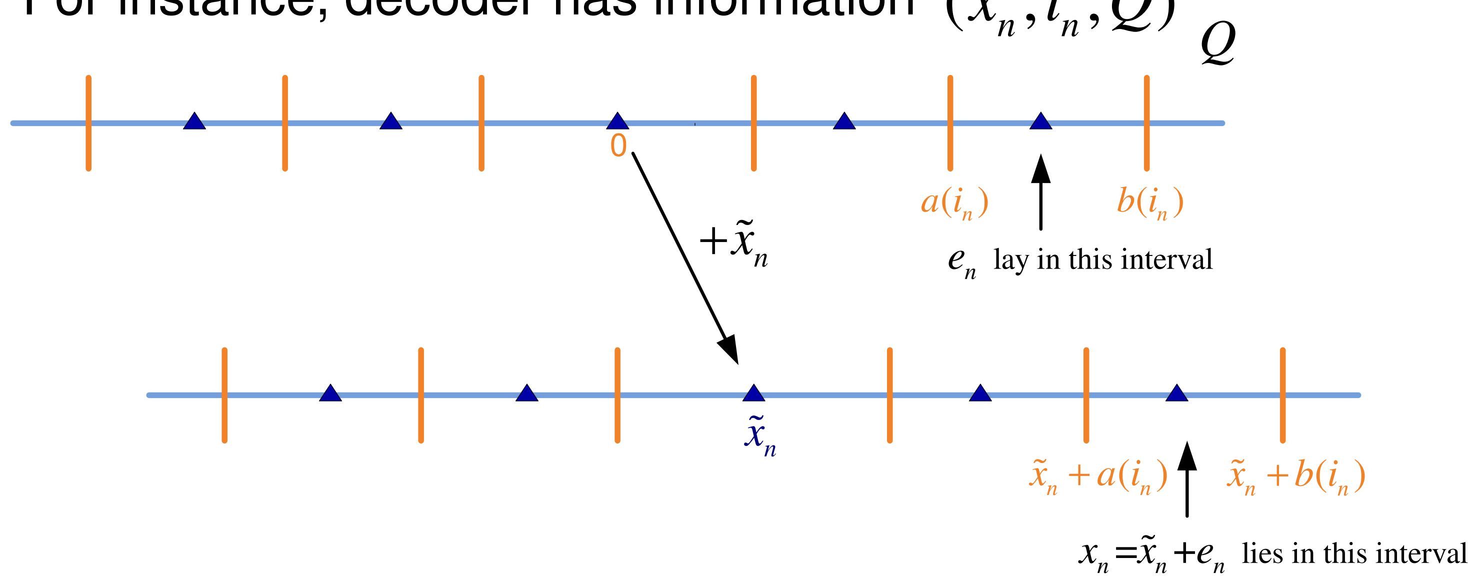
- IDPCM and SDPCM differ in the design of the non-causal filter
- The IDPCM design:
  - Filter taps determined by minimization of an *expected mean squared error* that involves statistics of *unquantized* samples
  - Process autocorrelation determines filter taps
  - Ignores bit-rate and innovation densities
  - No gains by increasing look-ahead beyond process order

Prior work

- The SDPCM design:
  - Employs a Kalman fixed-lag smoother
  - The AR process provides the 'plant' model with source samples viewed as the 'plant state'.
  - Quantizer operation provides the 'observation' model, with quantized source samples ($\hat{x}_n$) perceived as 'observations'
  - The model assumes that the quantization noise is white and uncorrelated with the source
  - Kalman filter optimal for linear Gaussian model: ignores the true innovation pdf

Sub-optimalities

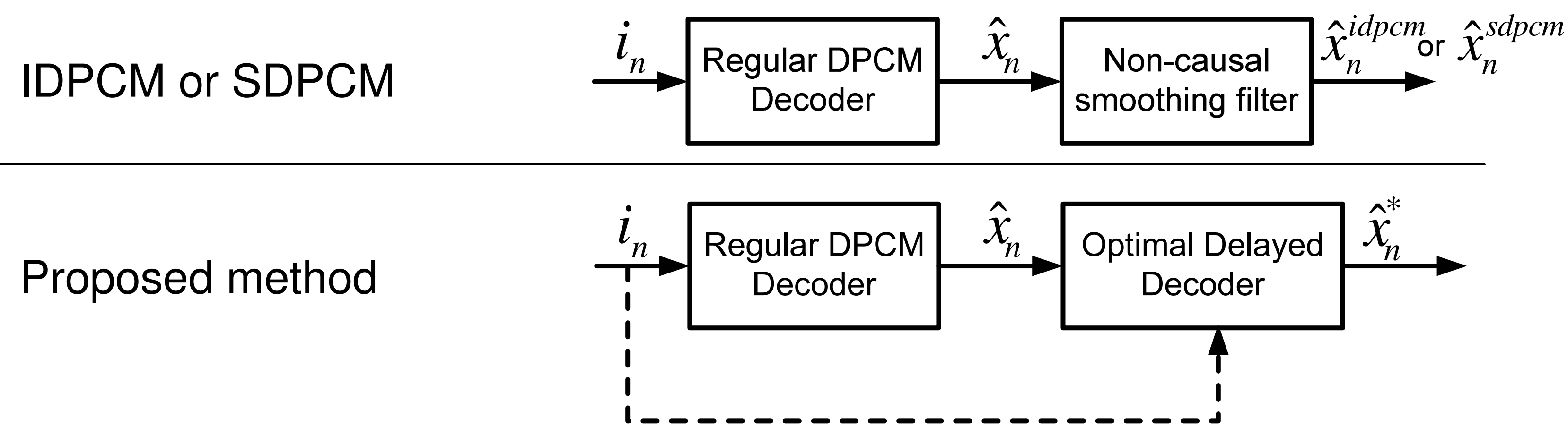
- Decoder has more information: unused by mere averaging of the zero-delay reconstructions
  - For instance, decoder has information $(\tilde{x}_n, i_n, Q)$



  - Smoothed reconstructions need not lie in $I_n = [\tilde{x}_n + a(i_n) \quad \tilde{x}_n + b(i_n))$ which is *known to the decoder*

Proposed method

- Estimation-theoretic approach that optimally combines the information $\cdots, i_{n-1}, i_n, i_{n+1}, \cdots, i_{n+L}$ to obtain the L - sample delayed reconstruction of x_n
- Recursively calculates the pdf of x_n conditioned on all available information



Optimal Delayed Decoder

- Distortion criterion - mean squared error (MSE)
- The optimal estimate of x_n at the decoder, with delay L :

$$\hat{x}_n^* = E[x_n \mid \cdots, i_{n-1}, i_n, \cdots, i_{n+L}]$$ [Gibson & Fischer, '82]

$$= E[x_n \mid \cdots, I_{n-1}, I_n, \cdots, I_{n+L}]$$

- Intervals $I_n = [\tilde{x}_n + a(i_n) \quad \tilde{x}_n + b(i_n))$ are an equivalent representation of information available to the decoder

- Expectation over the conditional pdf $p(x_n \mid \cdots, I_{n-1}, I_n, \cdots, I_{n+L})$

Optimal Delayed Decoder

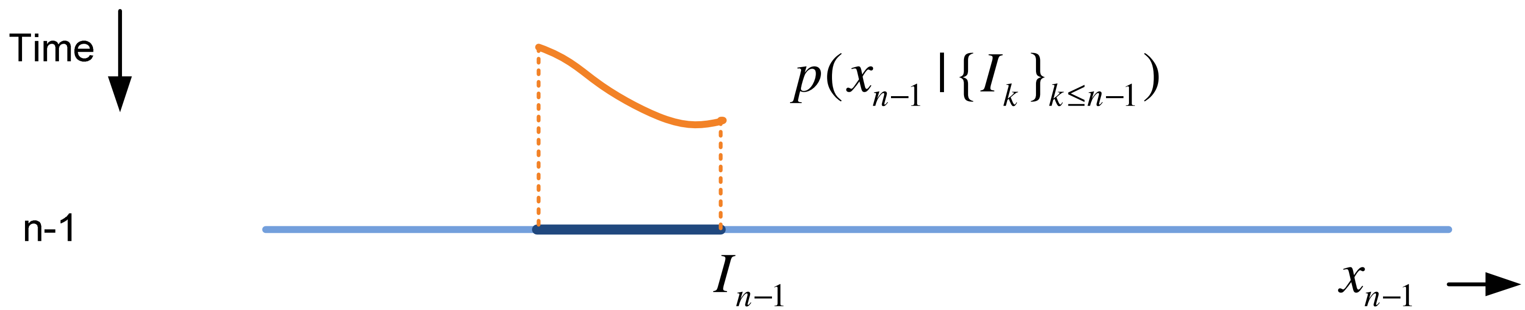
- By application of Bayes' rule and Markov property of the process

$$p(x_n \mid \{I_k\}_{k\le n+L}) = \frac{p(x_n \mid \{I_k\}_{k\le n})\, p(\{I_k\}_{n<k\le n+L} \mid x_n)}{\int p(x_n \mid \{I_k\}_{k\le n})\, p(\{I_k\}_{n<k\le n+L} \mid x_n)\, dx_n}$$

- $p(x_n \mid \{I_k\}_{k\le n})$ is the *zero-delay pdf* – combines all information up to time n
- $p(\{I_k\}_{n<k\le n+L} \mid x_n)$ weighs the zero-delay pdf to incorporate future information

Forward recursion

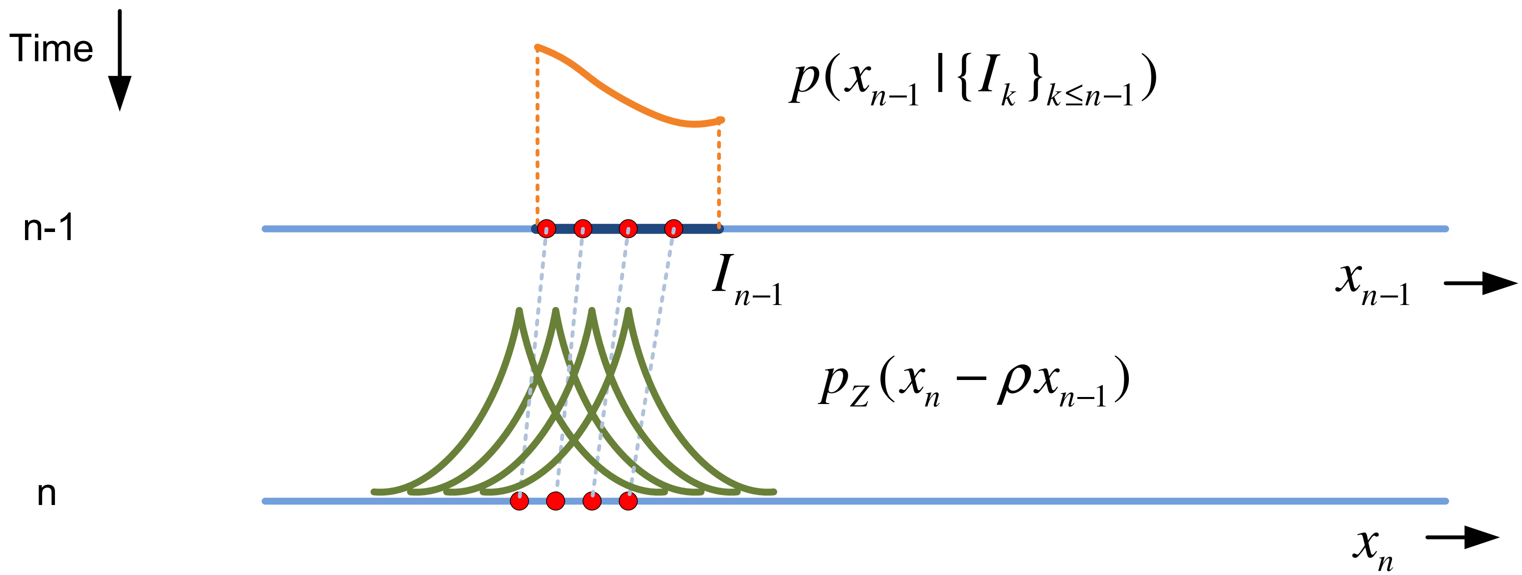
- Recursion for the zero-delay pdf: update from time $n-1$ to n



Say, zero-delay pdf at time $n-1$ is known

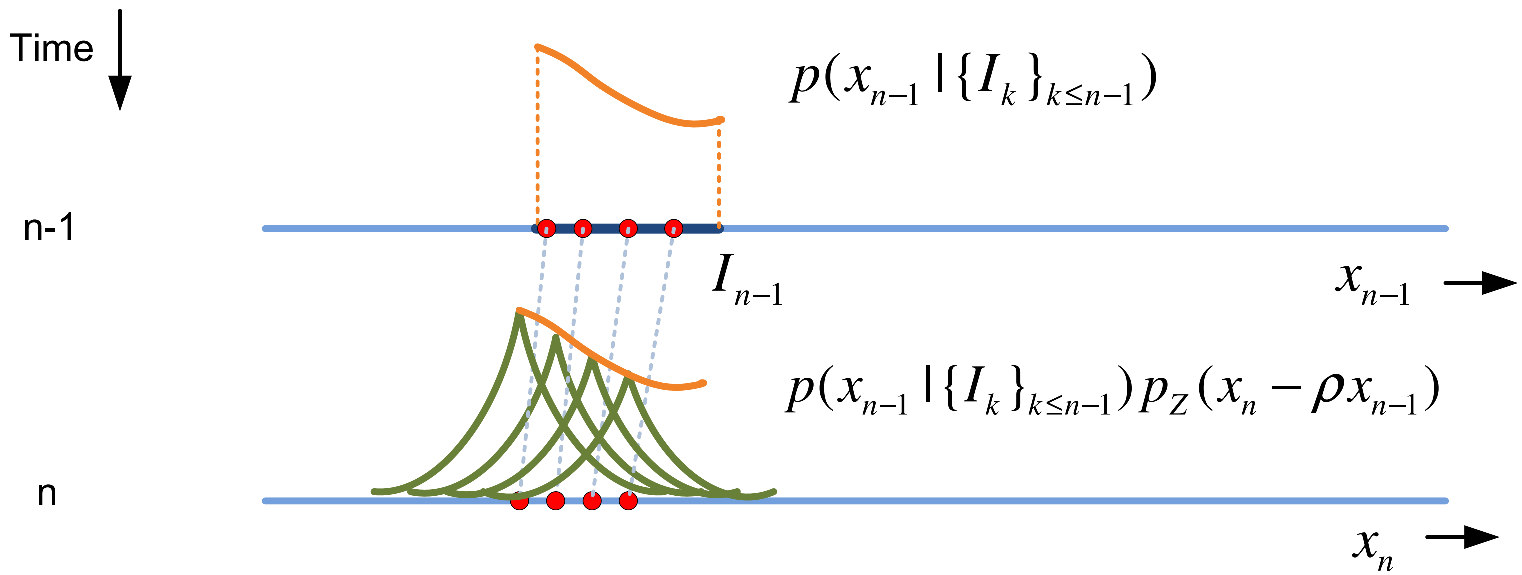
Forward recursion

- Recursion for the zero-delay pdf: update from time $n-1$ to n



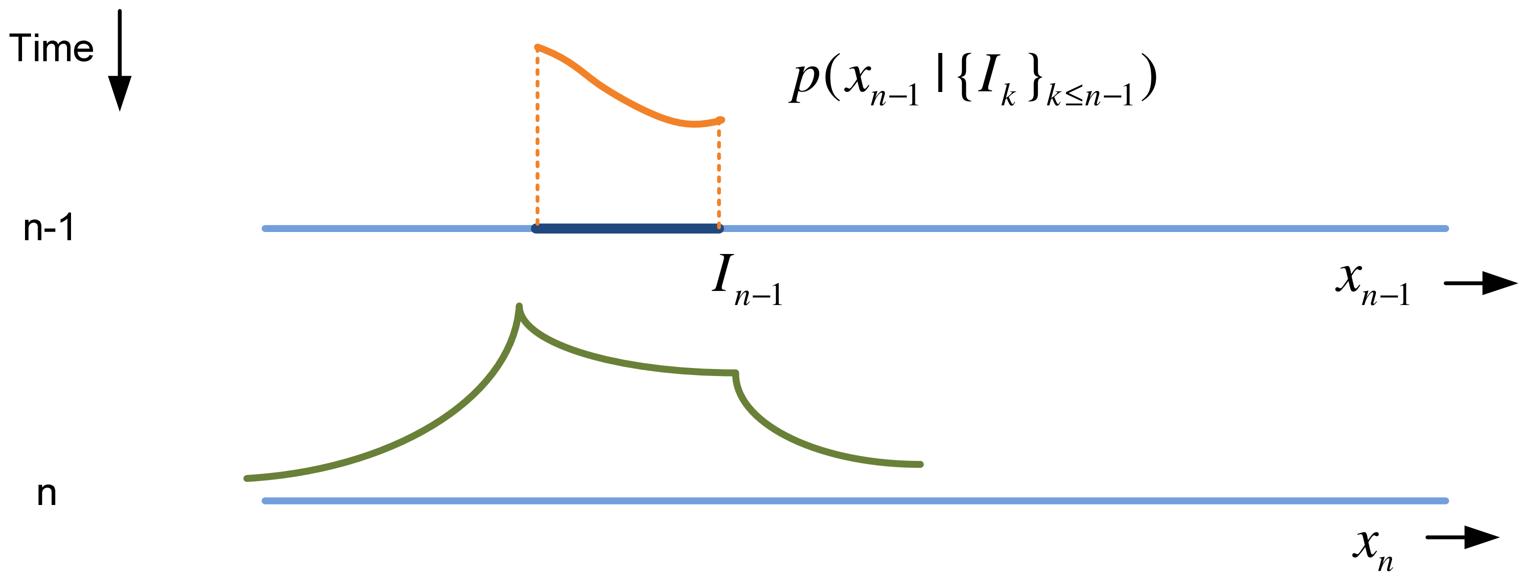
Forward recursion

- Recursion for the zero-delay pdf: update from time $n-1$ to n



Forward recursion

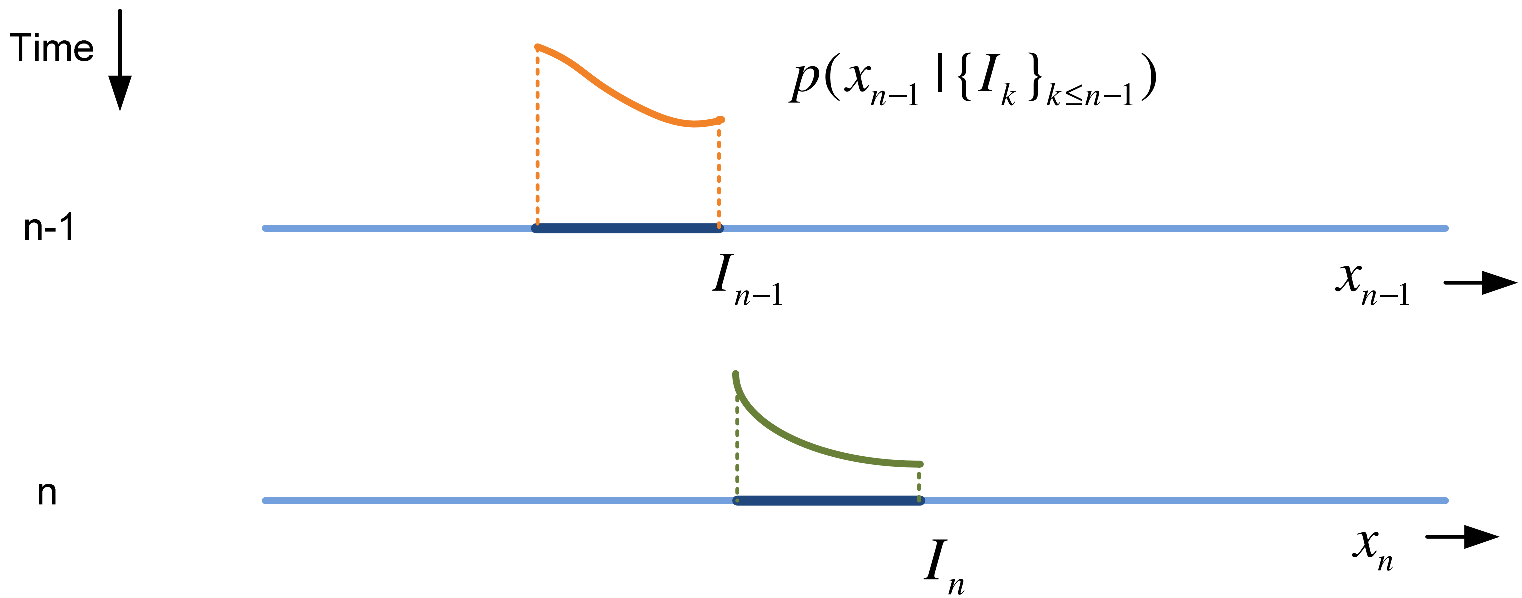
- Recursion for the zero-delay pdf: update from time $n-1$ to n



$$p(x_n | \{I_k\}_{k \le n-1}) = \int p(x_{n-1} | \{I_k\}_{k \le n-1}) \, p_Z(x_n - \rho x_{n-1}) dx_{n-1}$$

Forward recursion

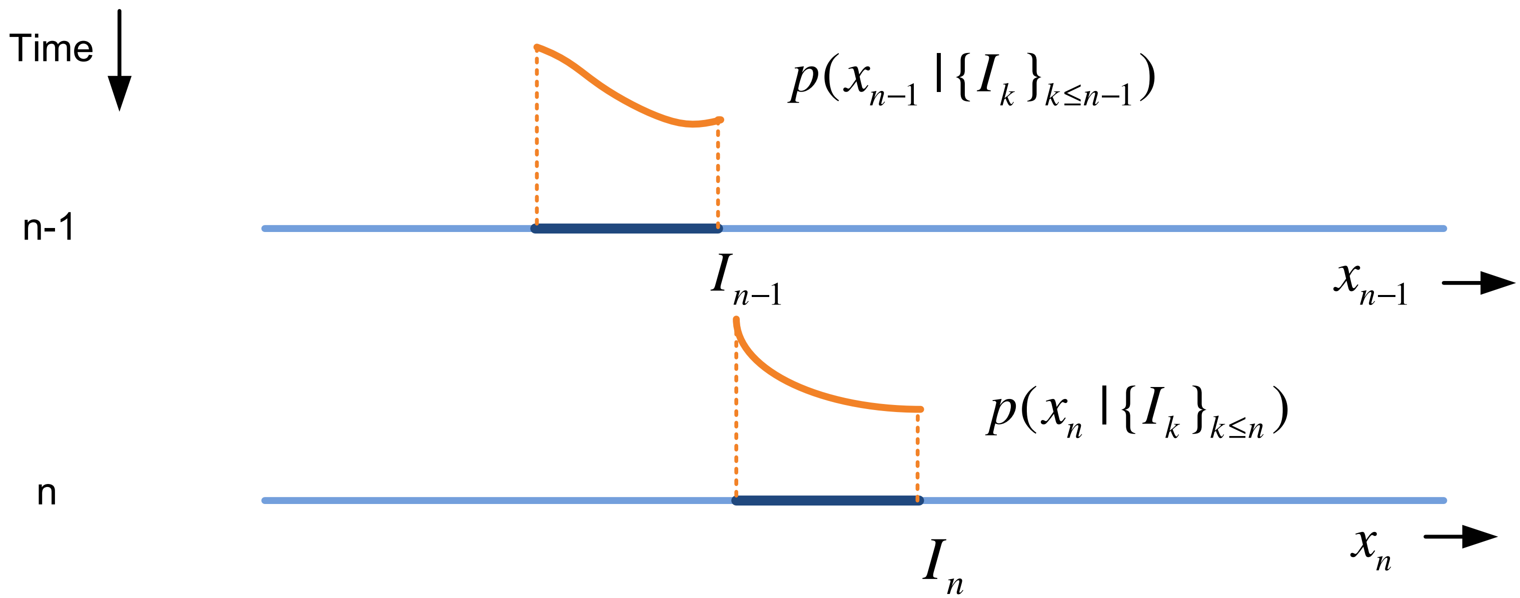
- Recursion for the zero-delay pdf: update from time $n-1$ to n



$$\begin{cases} p(x_n \mid \{I_k\}_{k \le n-1}) & x_n \in I_n \\ 0 & otherwise \end{cases}$$

Forward recursion

- Recursion for the zero-delay pdf: update from time $n-1$ to n



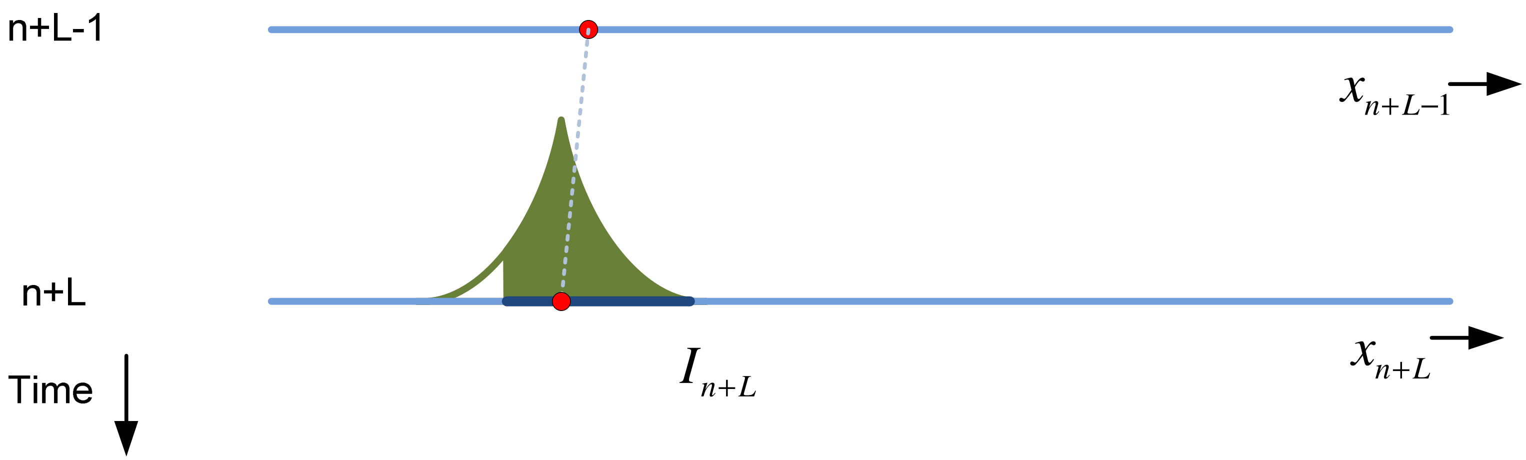
Zero-delay pdf at time n

$$p(x_n | \{I_k\}_{k \le n}) = \begin{cases} \dfrac{p(x_n | \{I_k\}_{k \le n-1})}{\int_{I_n} p(x_n | \{I_k\}_{k \le n-1}) dx_n} & x_n \in I_n \\ 0 & otherwise \end{cases}$$

Backward recursion

- Recursion for the probability of future outcomes: step back from time $n+L$ to n

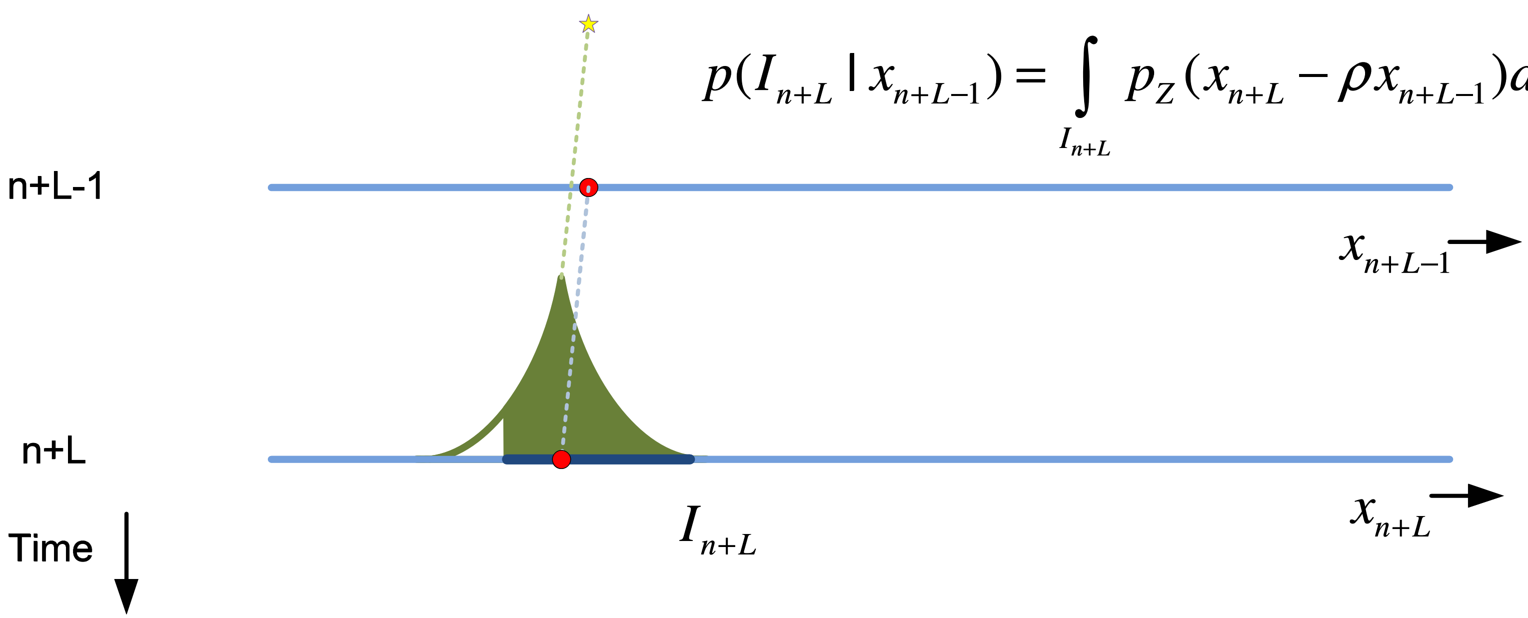
$$p(I_{n+L} \mid x_{n+L-1}) = \int_{I_{n+L}} p_Z(x_{n+L} - \rho x_{n+L-1}) dx_{n+L}$$



Backward recursion

- Recursion for the probability of future outcomes: step back from time $n+L$ to n

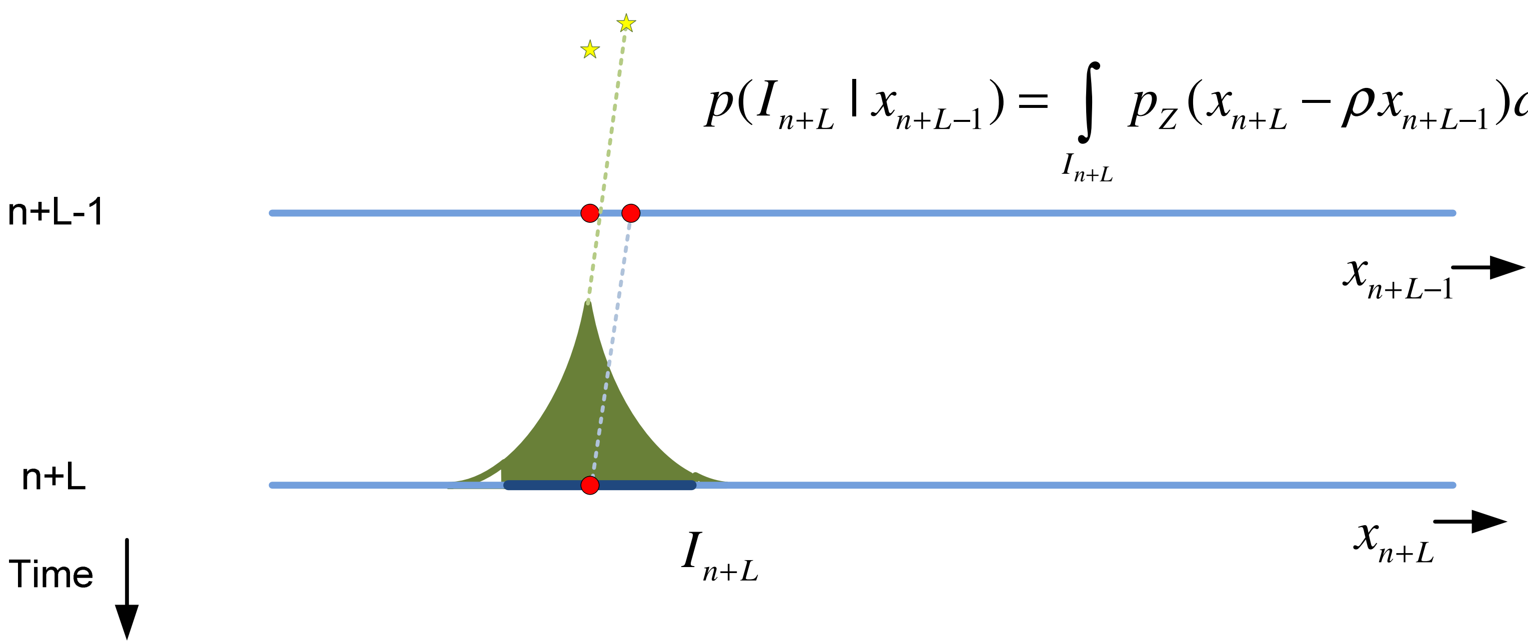
$$p(I_{n+L} \mid x_{n+L-1}) = \int_{I_{n+L}} p_Z(x_{n+L} - \rho x_{n+L-1}) dx_{n+L}$$



Backward recursion

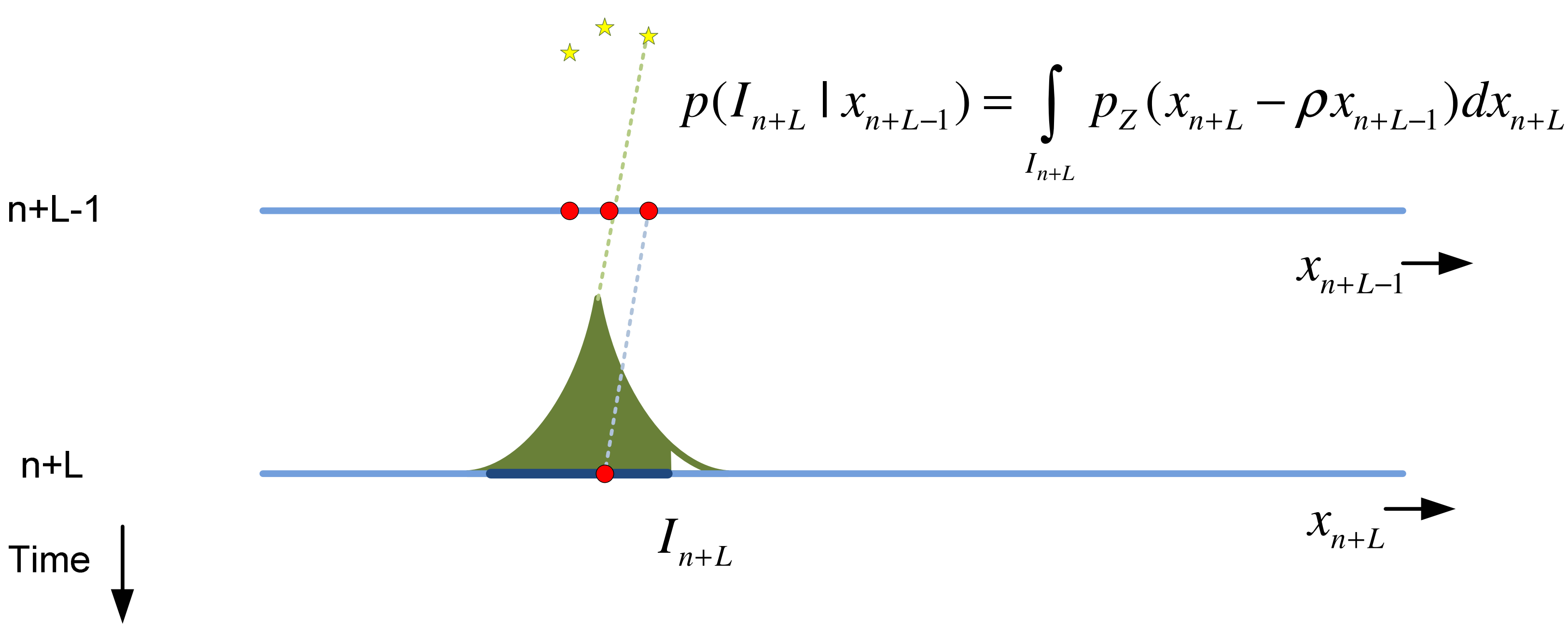
- Recursion for the probability of future outcomes: step back from time $n+L$ to n

$$p(I_{n+L} \mid x_{n+L-1}) = \int_{I_{n+L}} p_Z(x_{n+L} - \rho x_{n+L-1})dx_{n+L}$$



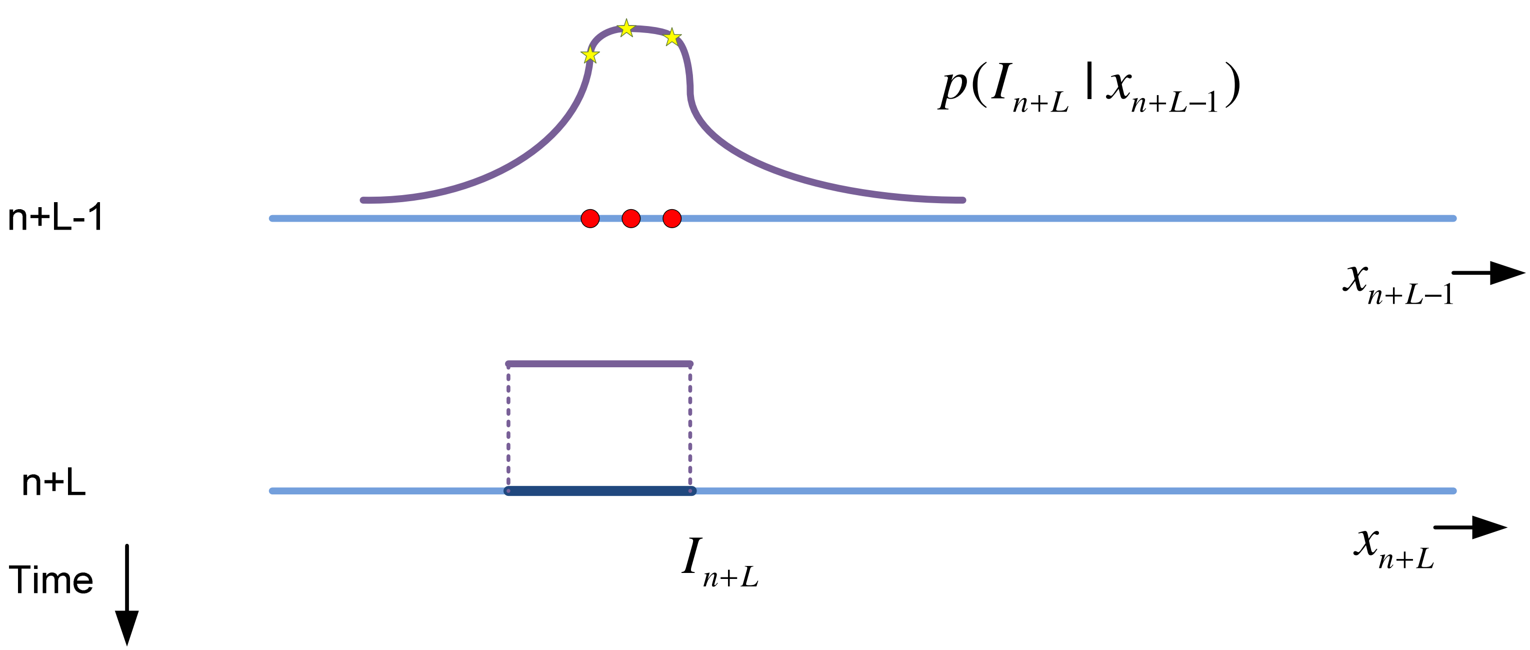
Backward recursion

- Recursion for the probability of future outcomes: step back from time $n+L$ to n



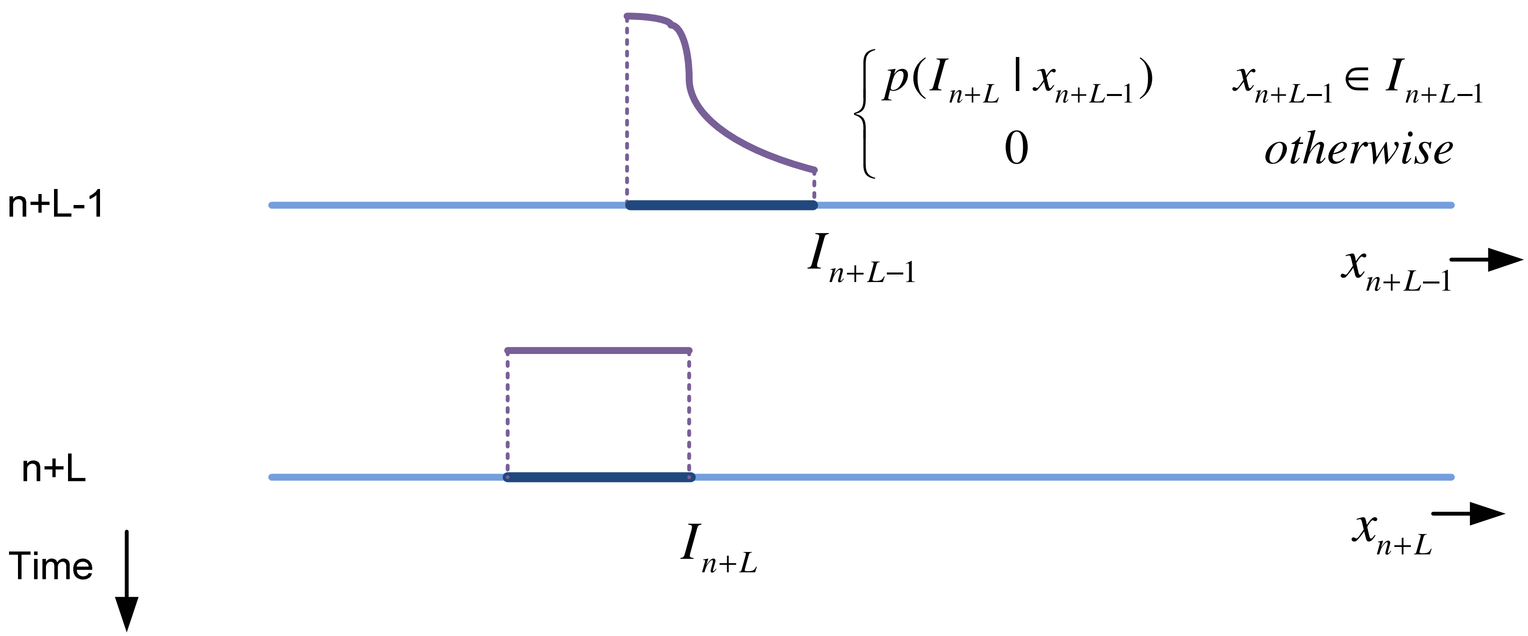
Backward recursion

- Recursion for the probability of future outcomes: step back from time $n+L$ to n



Backward recursion

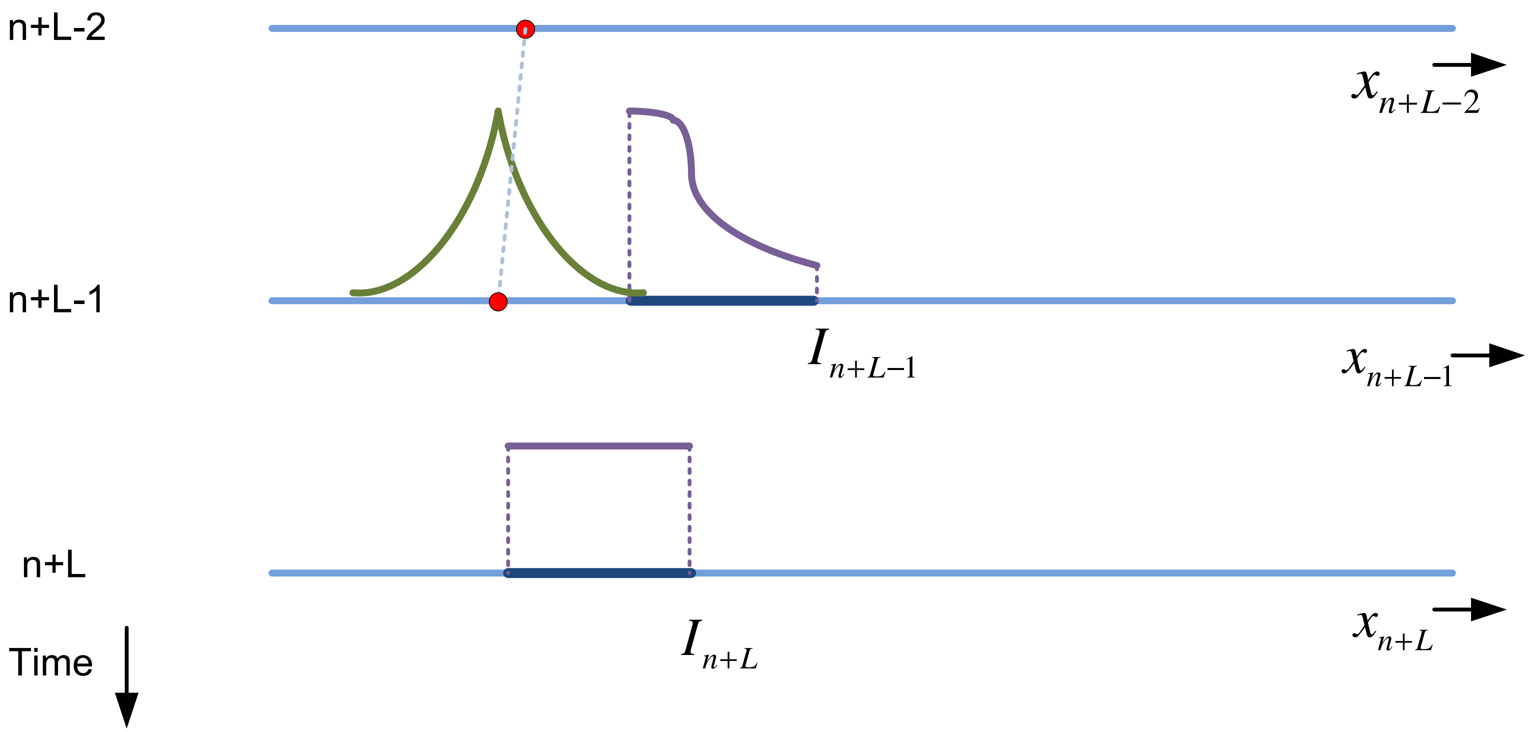
- Recursion for the probability of future outcomes: step back from time $n+L$ to n



Backward recursion

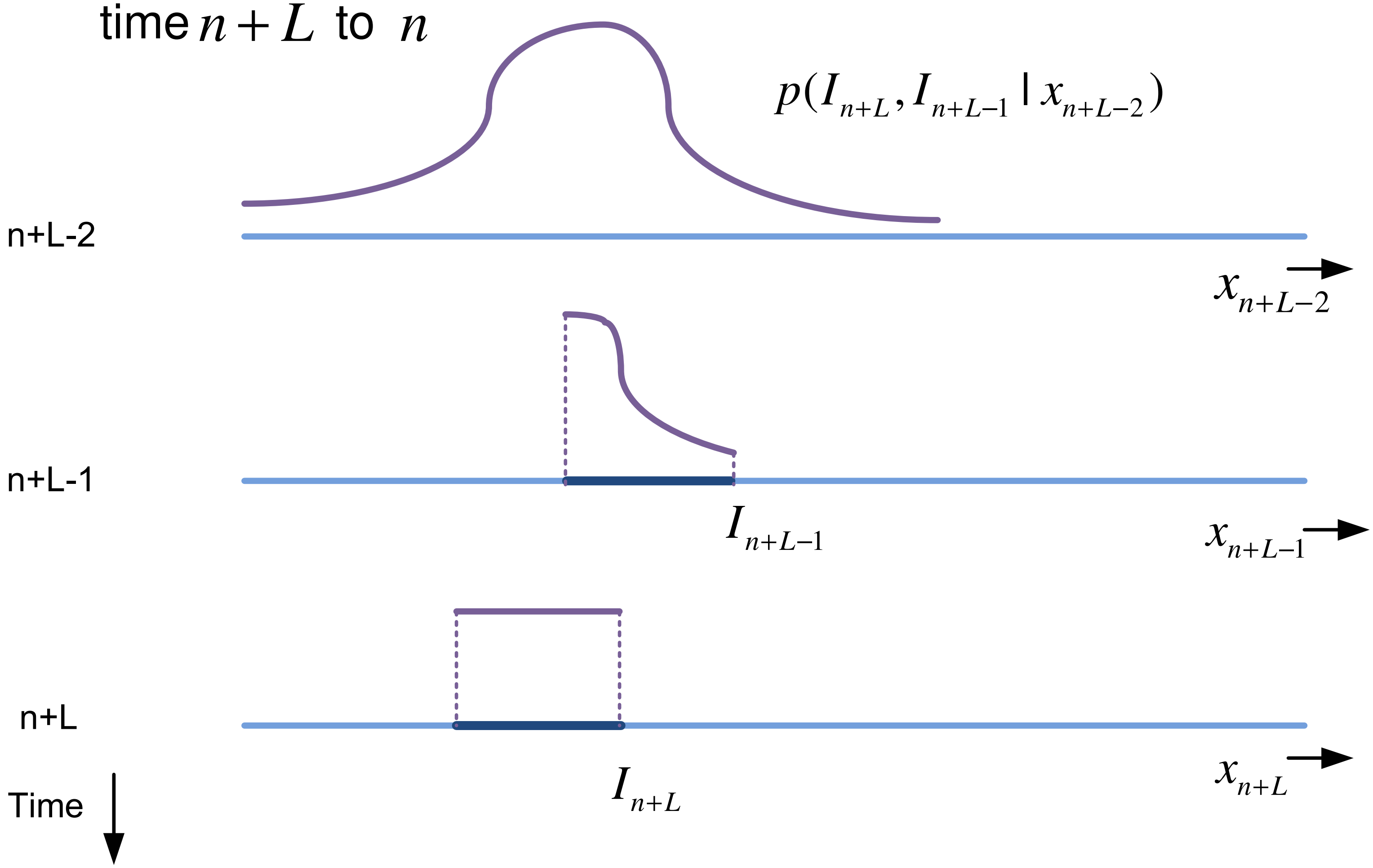
- Recursion for the probability of future outcomes: step back from time $n+L$ to n

$$p(I_{n+L}, I_{n+L-1} \mid x_{n+L-2}) = \int_{I_{n+L-1}} p(I_{n+L} \mid x_{n+L-1})\, p_Z(x_{n+L-1} - \rho x_{n+L-2})\, dx_{n+L-1}$$



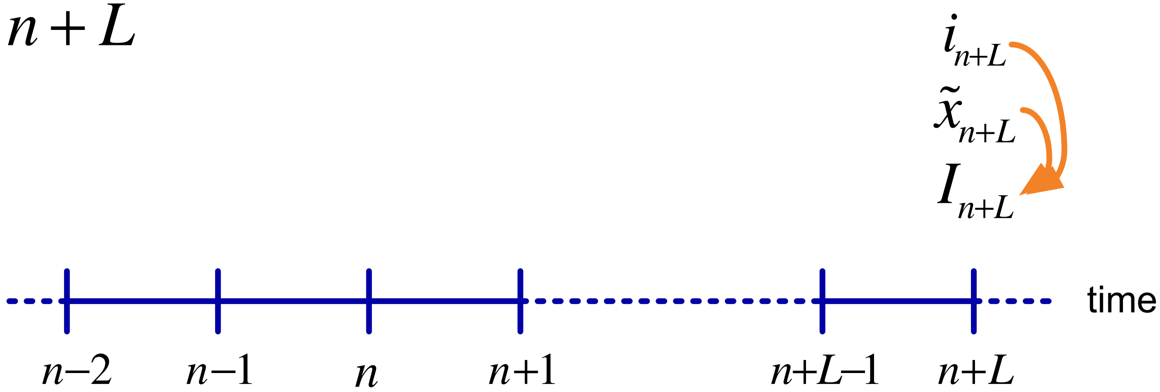
Backward recursion

- Recursion for the probability of future outcomes: step back from time $n+L$ to n



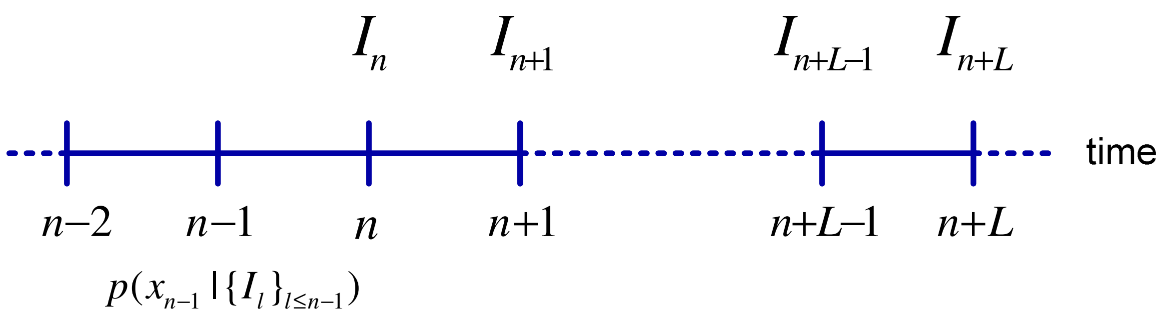
Summary

- At time $n+L$



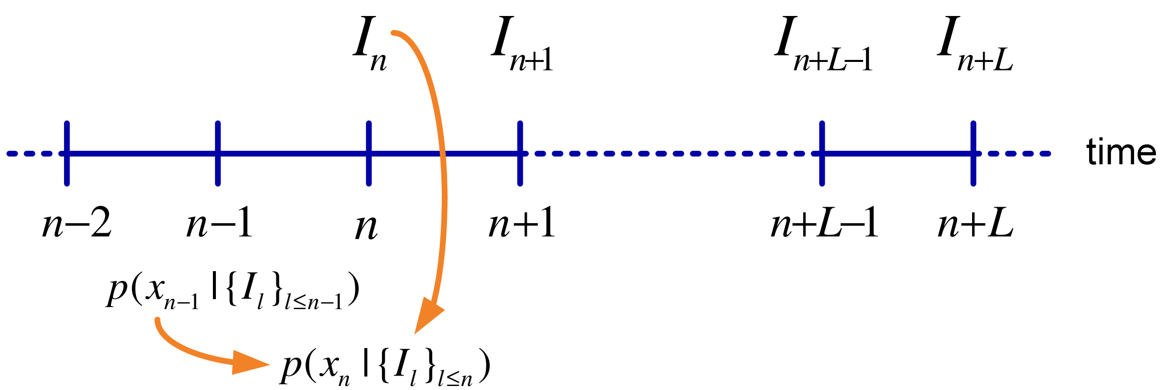
Summary

- At time $n+L$



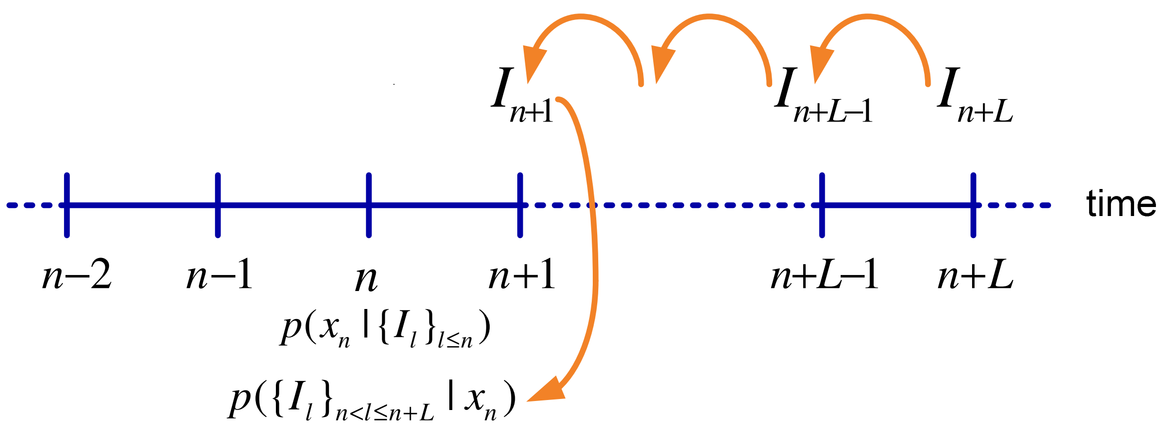
Summary

- At time $n+L$



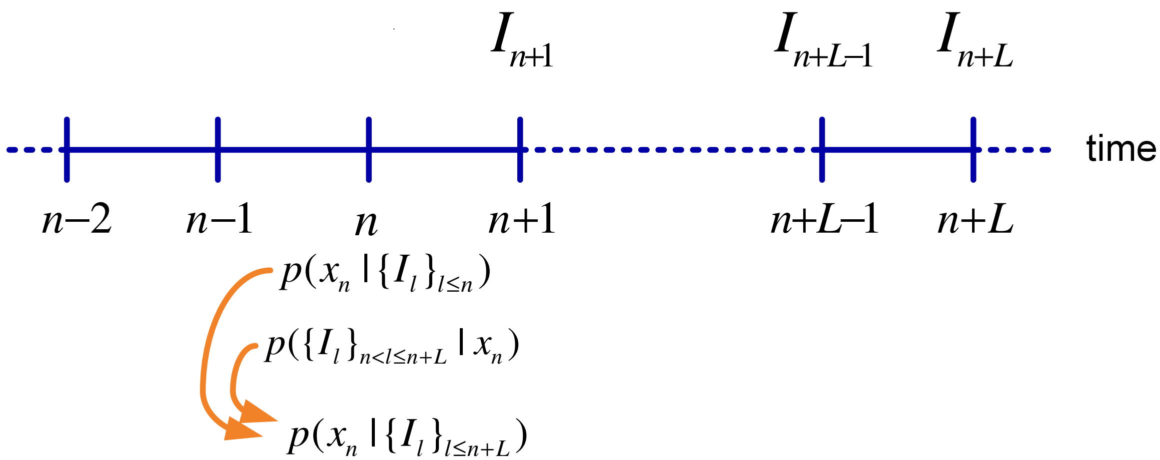
Summary

- At time $n+L$



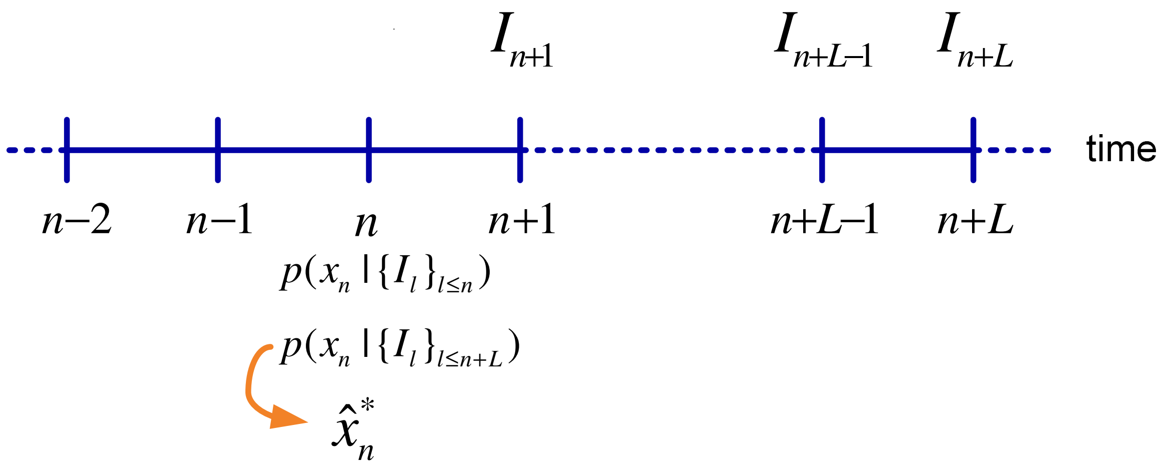
Summary

- At time $n+L$



Summary

- At time $n+L$



Special case: matched predictor $\tilde{x}_n = \rho \hat{x}_{n-1}$

- The L-step recursion for future probabilities can be simplified
- There exists function $\Lambda_{i_1,\cdots,i_L}(x)$ such that,

$$p(\{I_k\}_{n<k\leq n+L} \mid x_n) = \Lambda_{i_{n+1},\cdots,i_{n+L}}(x_n - \hat{x}_n)$$

- A codebook of the functions $\Lambda_{i_1,\cdots,i_L}(x)$ can be constructed
- Recursion can be replaced by codebook access with $i_{n+1},\cdots,i_{n+L}$, and translation of the function by $\hat{x}_n$

Codebook-based Delayed Decoder

- Henceforth, we exclusively consider the matched predictor

$$\tilde{x}_n = \rho \hat{x}_{n-1}$$

- Optimal delayed estimate:

$$\hat{x}_n^* = E[x_n \mid \cdots, I_{n-1}, I_n, \cdots, I_{n+L}]$$

$$p(x_n \mid \{I_k\}_{k \le n})$$

$$p(\{I_k\}_{n<k\le n+L} \mid x_n)$$

$$\Lambda_{i_{n+1},\cdots,i_{n+L}}(x_n - \hat{x}_n)$$

Table look-up via $\cdots, i_{n-2}, i_{n-1}, i_n$?

Table look-up via $i_{n+1}, \cdots, i_{n+L}$

Codebook-based Delayed Decoder

- Henceforth, we exclusively consider the matched predictor

$$\tilde{x}_n = \rho \hat{x}_{n-1}$$

- Optimal delayed estimate:

$$\hat{x}_n^* = E[x_n \mid \cdots, I_{n-1}, I_n, \cdots, I_{n+L}]$$

$$p(x_n \mid \{I_k\}_{k \le n})$$

$$p(\{I_k\}_{n<k\le n+L} \mid x_n)$$

$$\Lambda_{i_{n+1},\cdots,i_{n+L}}(x_n - \hat{x}_n)$$

~~Table look-up via $\cdots, i_{n-2}, i_{n-1}, i_n$?~~

Growing history of indices precludes an optimal look-up table for the zero-delay pdf

Table look-up via $i_{n+1}, \cdots, i_{n+L}$

Codebook-based Delayed Decoder

- Henceforth, we exclusively consider the matched predictor

$$\tilde{x}_n = \rho \hat{x}_{n-1}$$

- Optimal delayed estimate:

$$\hat{x}_n^* = E[x_n \mid \cdots, I_{n-1}, I_n, \cdots, I_{n+L}]$$

$$p(x_n \mid \{I_k\}_{k \le n})$$

$$p(\{I_k\}_{n<k\le n+L} \mid x_n)$$

$$\Lambda_{i_{n+1},\cdots,i_{n+L}}(x_n - \hat{x}_n)$$

A good approximation is still feasible !

Table look-up via $i_{n+1}, \cdots, i_{n+L}$

Codebook-based Delayed Decoder

- Henceforth, we exclusively consider the matched predictor

$$\tilde{x}_n = \rho \hat{x}_{n-1}$$

- Optimal delayed estimate:

$$\hat{x}_n^* = E[x_n \mid \cdots, I_{n-1}, I_n, \cdots, I_{n+L}]$$

$$p(x_n \mid \{I_k\}_{k \le n}) \qquad p(\{I_k\}_{n<k\le n+L} \mid x_n)$$

$$\Lambda_{i_{n+1},\cdots,i_{n+L}}(x_n - \hat{x}_n)$$

A good approximation is still feasible !

Table look-up via $i_{n+1}, \cdots, i_{n+L}$

A codebook-based approximation for the optimal delayed estimate

Codebook-based Delayed Decoder

- Approximation for the zero-delay pdf:
  - Let $p_E(e)$ denote the *stationary marginal prediction error pdf* - a fixed (time invariant) pdf [Farvardin & Modestino, '85]
  - The pdf of x_n conditioned on past indices is approximated as:

$$p(x_n \mid \cdots, i_{n-2}, i_{n-1}) \approx p(x_n \mid \rho \hat{x}_{n-1}) = p_E(x_n - \rho \hat{x}_{n-1})$$

$$\because e_n = x_n - \rho \hat{x}_{n-1}$$

  - Thus the zero-delay pdf is just:

$$p(x_n \mid \{I_k\}_{k \le n}) \approx \begin{cases} \dfrac{p_E(x_n - \rho \hat{x}_{n-1})}{\int_{I_n} p_E(x_n - \rho \hat{x}_{n-1}) dx_n} & x_n \in I_n \\ 0 & otherwise \end{cases}$$

Codebook-based Delayed Decoder

- Approximate delayed estimate:

$$\hat{x}_n^* = E[x_n \mid \cdots, I_{n-1}, I_n, \cdots, I_{n+L}] = \frac{\int x_n p(x_n \mid \{I_k\}_{k\le n}) p(\{I_k\}_{n<k\le n+L} \mid x_n) dx_n}{\int p(x_n \mid \{I_k\}_{k\le n}) p(\{I_k\}_{n<k\le n+L} \mid x_n) dx_n}$$

$$\begin{cases} \dfrac{p_E(x_n - \rho\hat{x}_{n-1})}{\int_{I_n} p_E(x_n - \rho\hat{x}_{n-1}) dx_n} & x_n \in I_n \\ 0 & otherwise \end{cases}$$

$$I_n = [\rho\hat{x}_{n-1} + a(i_n) \quad \rho\hat{x}_{n-1} + b(i_n))$$

$$\Lambda_{i_{n+1},\cdots,i_{n+L}}(x_n - \hat{x}_n)$$

$$\hat{x}_n = \rho\hat{x}_{n-1} + \hat{e}_n(i_n)$$

$$\hat{x}_n^* \approx \rho\hat{x}_{n-1} + c(i_n, \cdots, i_{n+L})$$

Look-up table/codebook

Codebook design

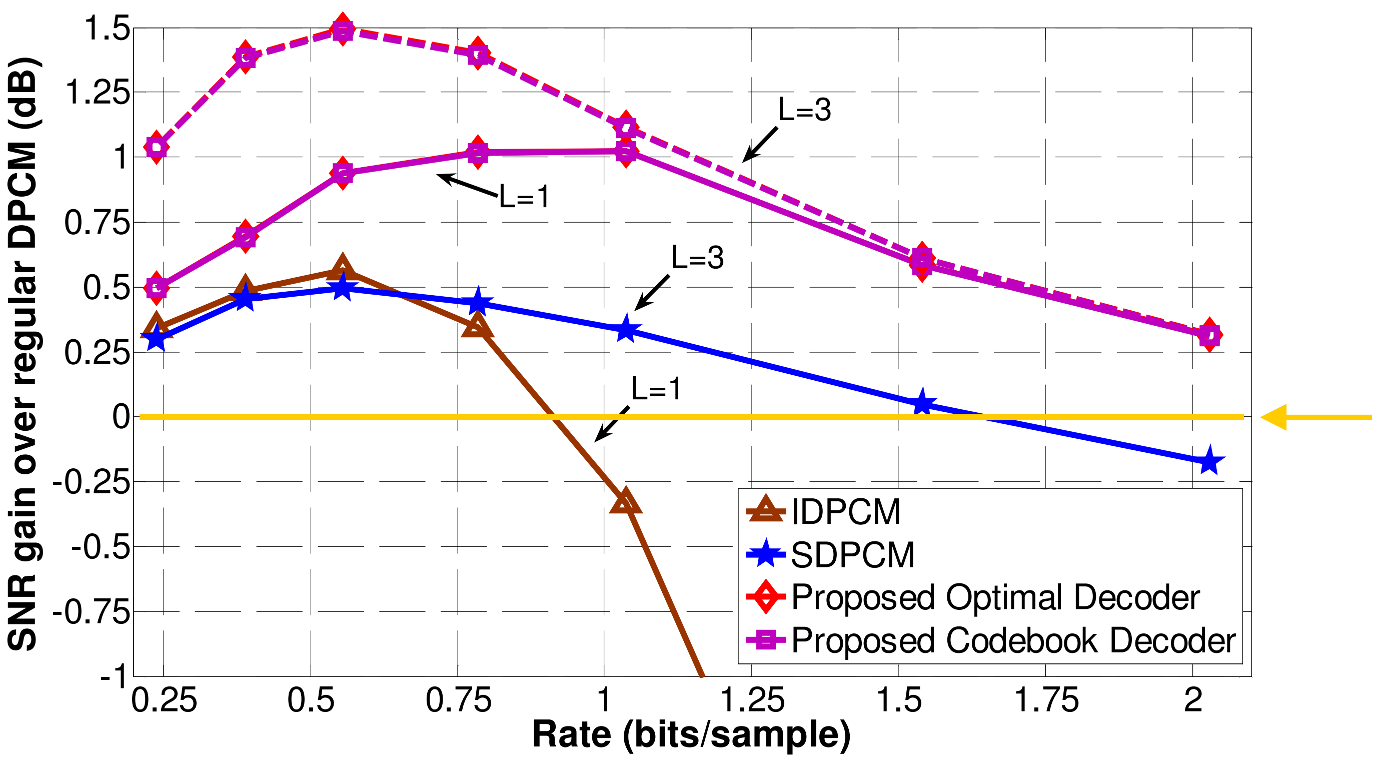
- Numerical evaluation via $p_E(e)$ and $\Lambda_{i_{n+1},\cdots,i_{n+L}}(x_n - \hat{x}_n)$
- Alternative - a training-set based design:
  - $\hat{x}_n^* \approx \rho \hat{x}_{n-1} + c(i_n,\cdots,i_{n+L})$ => $c(i_n,\cdots,i_{n+L})$ is the estimate of the prediction error at time n given the window of indices $i_n,\cdots,i_{n+L}$
  - Encoder is fixed: run it on a long enough training set of the source, and obtain prediction error training set and indices
  - Train delayed decoding codebook

Results

- Source is first order AR
- DPCM Encoder:
  - Rate: first order entropy of output indices
  - Employs uniform threshold quantizer: scaled suitably to achieve different rates
  - Thresholds fixed by scale-factor, reconstructions optimized iteratively similar to [Farvardin & Modestino, '85]
  - Iterative optimization also provides $p_E(e)$ for codebook approach
  - Predictor matched to source

Results

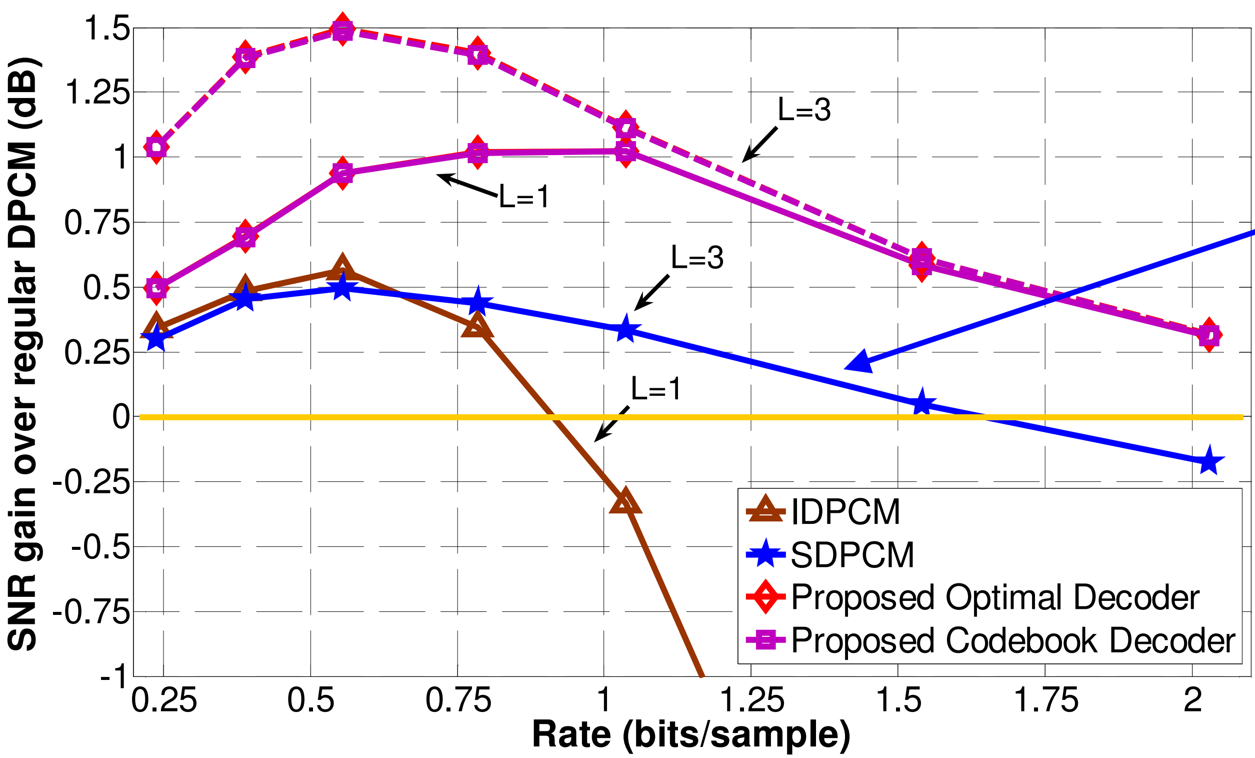
Performance comparison of competing delayed decoders for a Gaussian source with $\rho = 0.95$



Performance of zero-delay DPCM at different bit-rates

Results

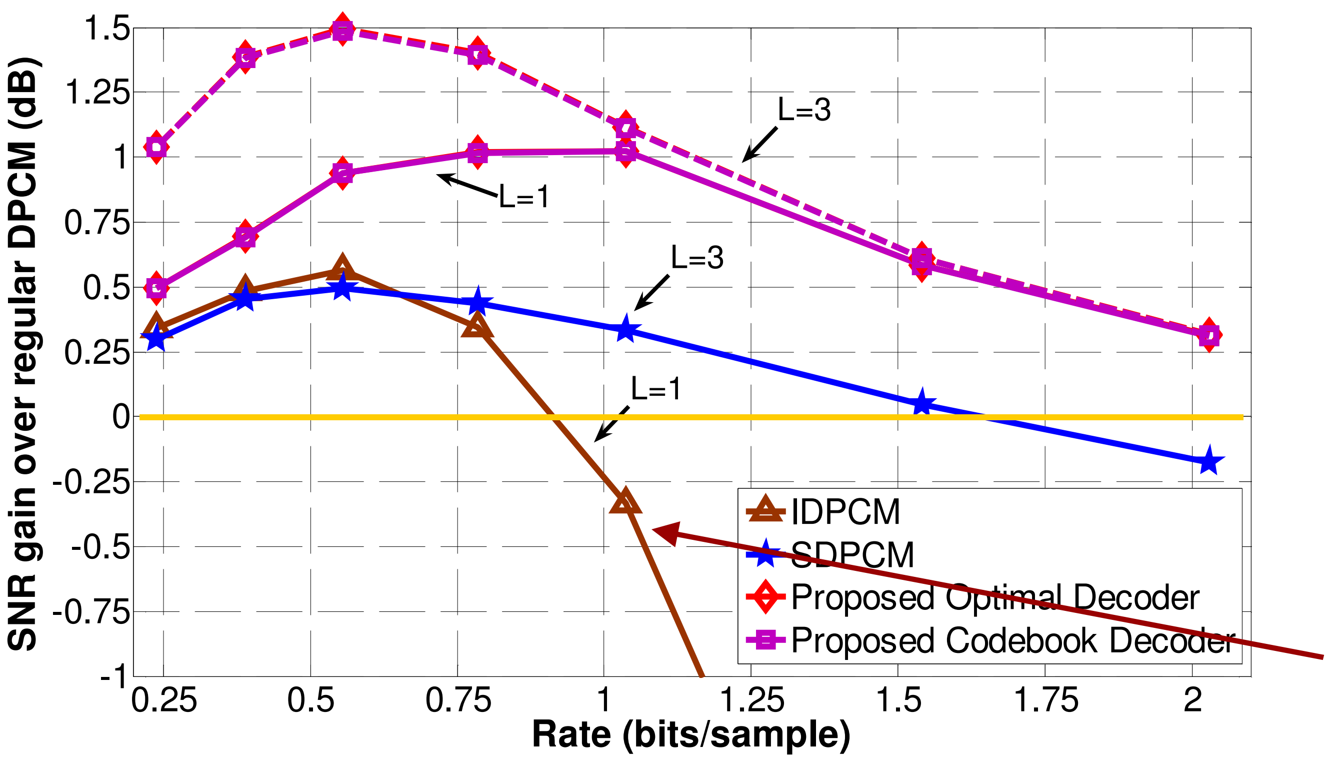
Performance comparison of competing delayed decoders for a Gaussian source with $\rho = 0.95$



SDPCM with lag of 3 samples, worse at lower delays

Results

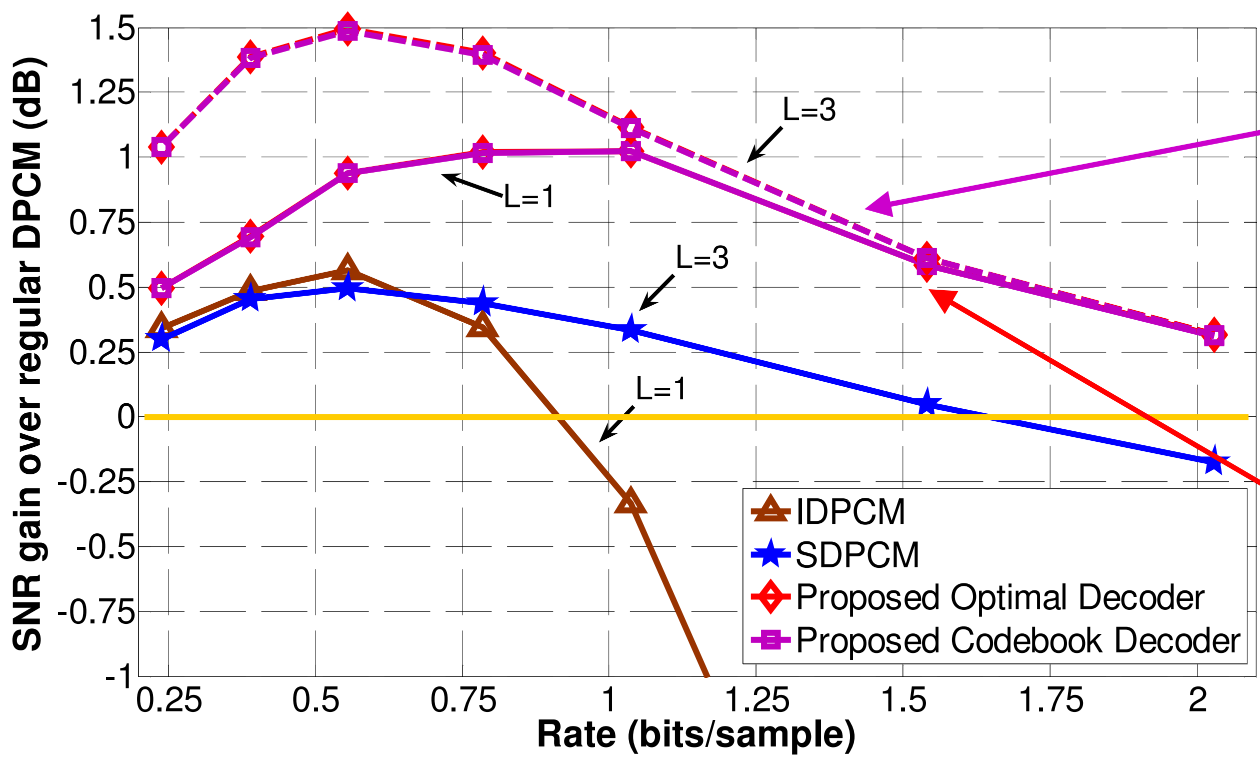
Performance comparison of competing delayed decoders for a Gaussian source with $\rho = 0.95$



IDPCM, delay limited to 1 sample automatically

Results

Performance comparison of competing delayed decoders for a Gaussian source with $\rho = 0.95$

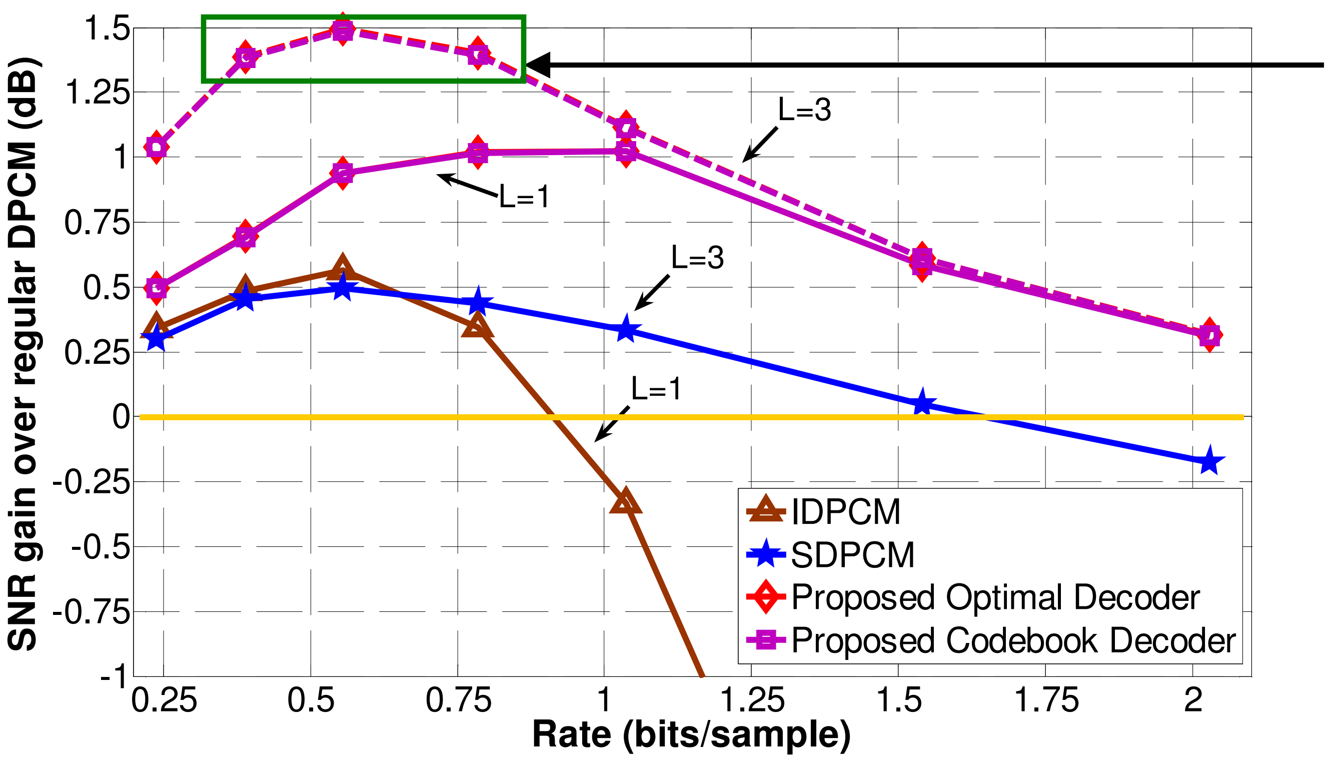


Codebook-based approach using 1 and 3 future indices

Performance curves for the optimal delayed decoder hidden beneath plots for the codebook approach

Results

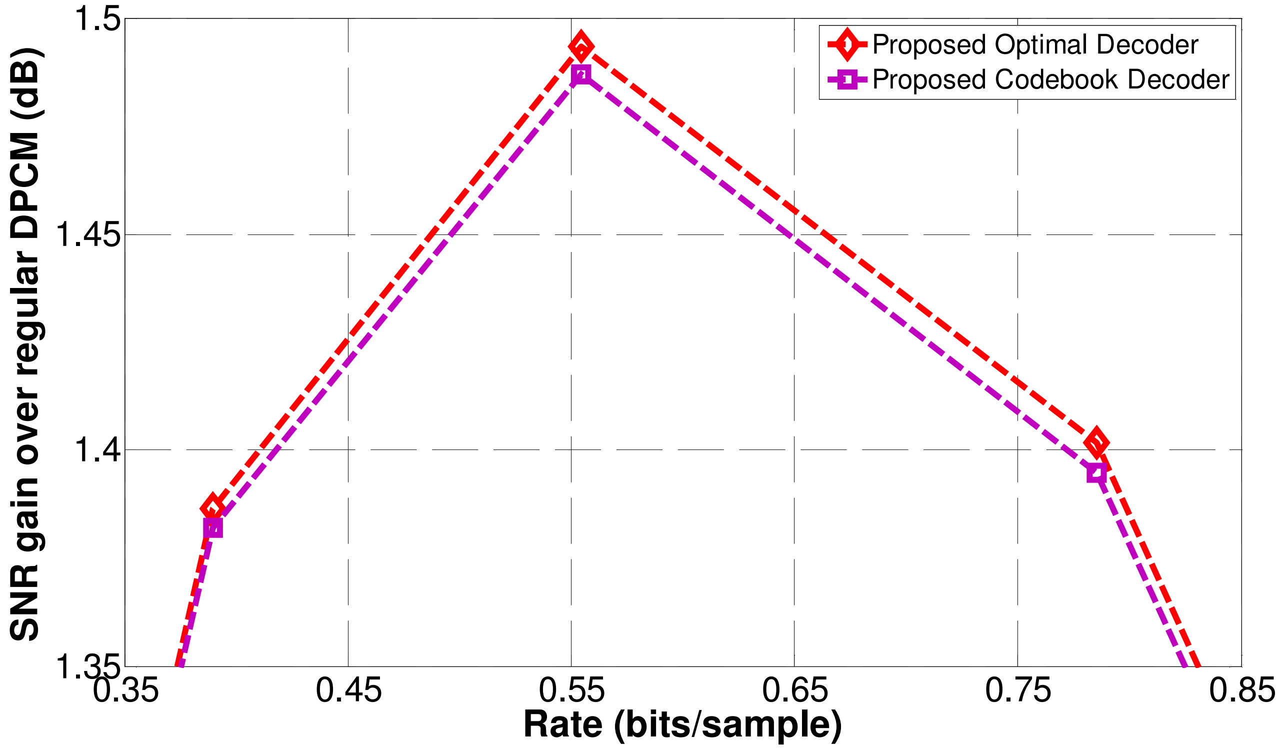
Performance comparison of competing delayed decoders for a Gaussian source with $\rho = 0.95$



Zoom in to see the performance gap between optimal and codebook approaches

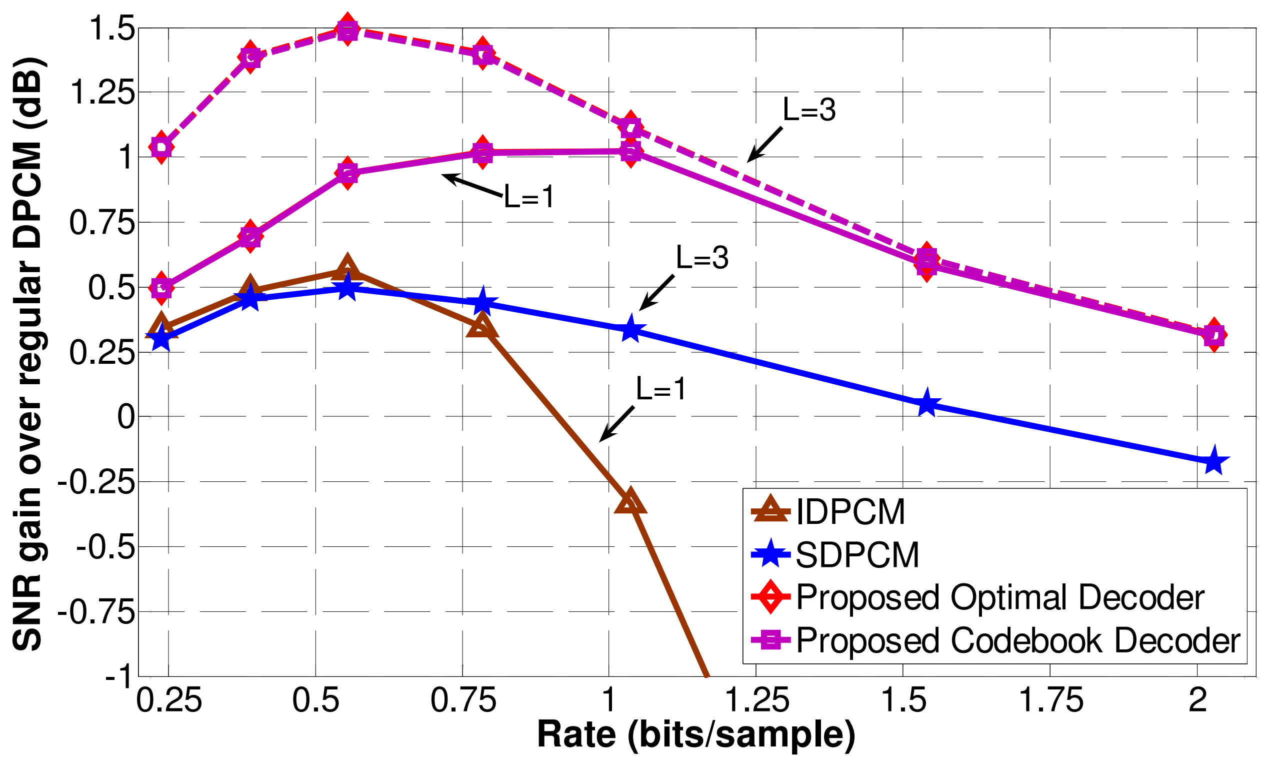
Results

Performance comparison of competing delayed decoders for a Gaussian source with $\rho = 0.95$



Results

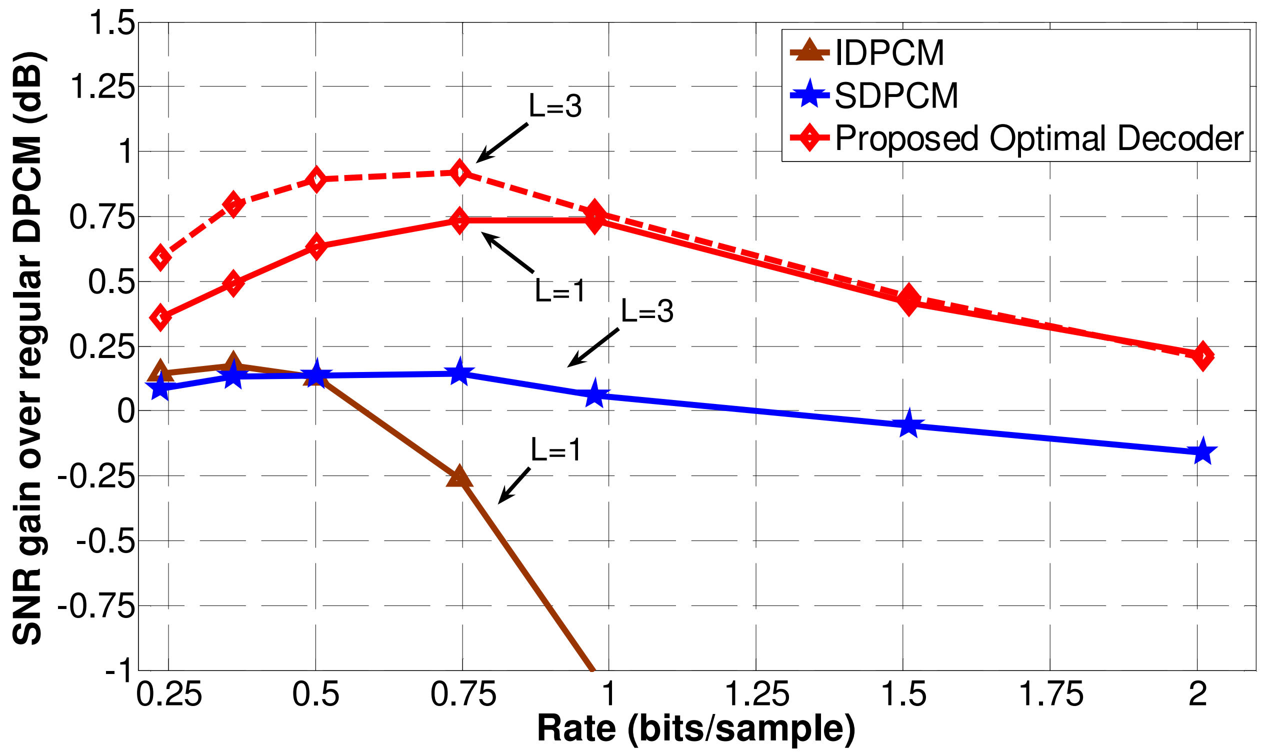
Performance comparison of competing delayed decoders for a Gaussian source with $\rho = 0.95$



- Performance of SDPCM and IDPCM not guaranteed to be better than zero-delay DPCM
- Proposed approaches at 1 sample delay outperform SDPCM at higher delay (3) : indices contain a lot of information
- At low bit-rates increasing delay provides more gains

Results

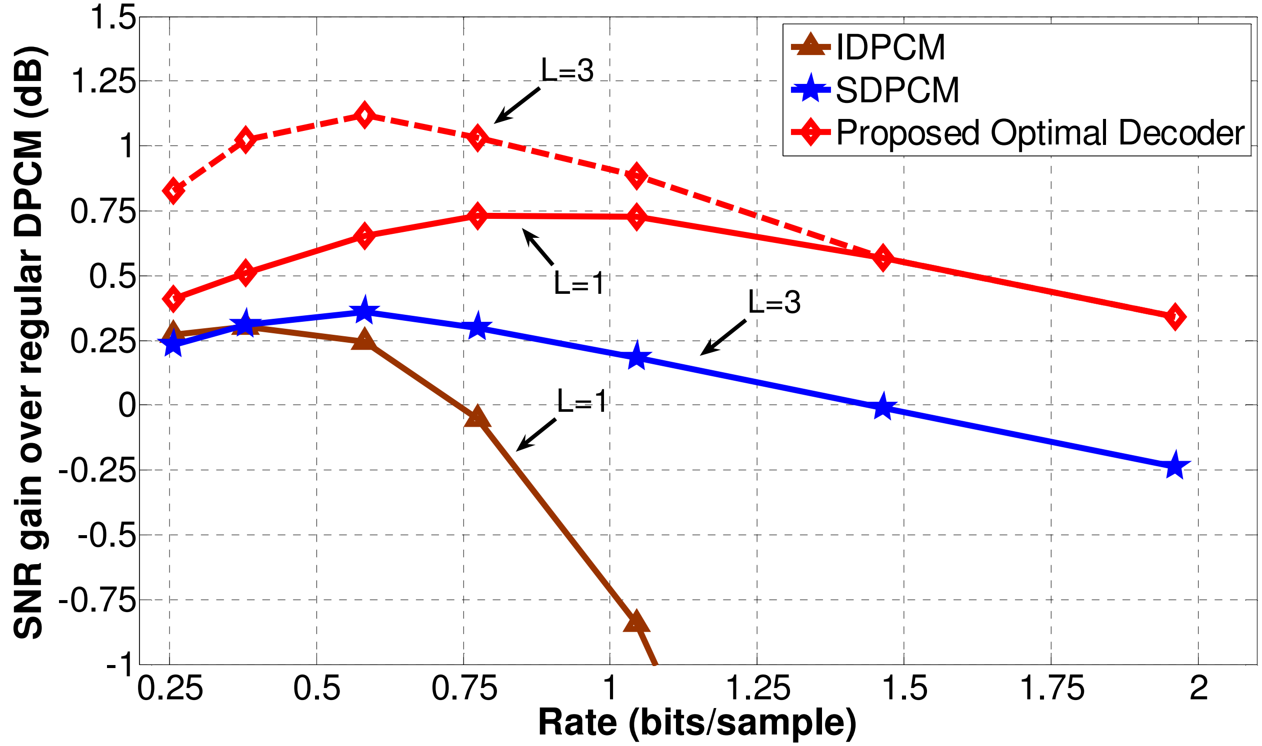
Performance comparison of competing delayed decoders
for a Gaussian source with $\rho = 0.8$



- Lower correlation naturally implies lesser to be gained from looking into the future

Results

Performance comparison of competing delayed decoders for a source with Laplacian innovations with $\rho = 0.95$



Other contributions

- Codebook approach trades computational complexity for memory
- Proposed an approach for codebook-size reduction via an index mapping technique with very minimal performance loss
- Optimal and codebook approaches readily extended to higher order sources (equivalence via an appropriate first-order vector AR process)
- Index window employed in the codebook can be extended to include a few past indices: useful in the case of higher order sources
- Training-set based design, and codebook-based operation, particularly attractive for higher order sources (due to to the higher dimensionality involved)

Summary

- Proposed an estimation-theoretic approach for optimal delayed decoding in predictive coding systems
- Combines all known information at the decoder in a recursively calculated conditional pdf
- Motivates a codebook-based delayed decoder that is nearly optimal even for modest dimensions
- Substantial performance gains compared to prior smoothing/filtering techniques

Future directions

- Encoder optimization based on the proposed delayed decoder
  - Employ delayed reconstructions for prediction via local decoder

- Delayed decoding in adaptive predictive coding scenarios
  - Application for speech/audio coding in Bluetooth systems
  - Delayed decoding codebook adaptation techniques