A perceptually enhanced scalable-to-lossless audio coding scheme and its optimization in a trellis

Emmanuel Ravelli, Vinay Melkote and Prof. Rose
Signal Compression Lab
Dept. of Electrical and Computer Engineering
University of California, Santa Barbara

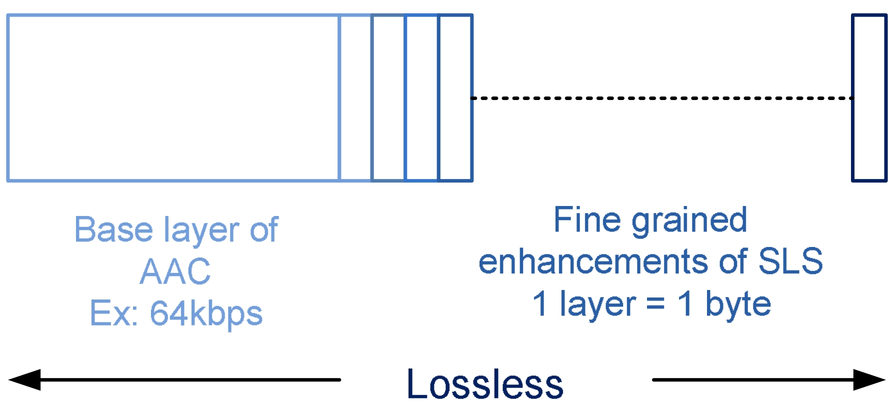
Outline

- Introduction to MPEG Scalable-to-Lossless (SLS) audio
- Drawbacks of the SLS structure
- Proposed perceptually enhanced SLS scheme
- Optimization of the proposed scheme using a trellis
- Results
- Conclusions

Introduction

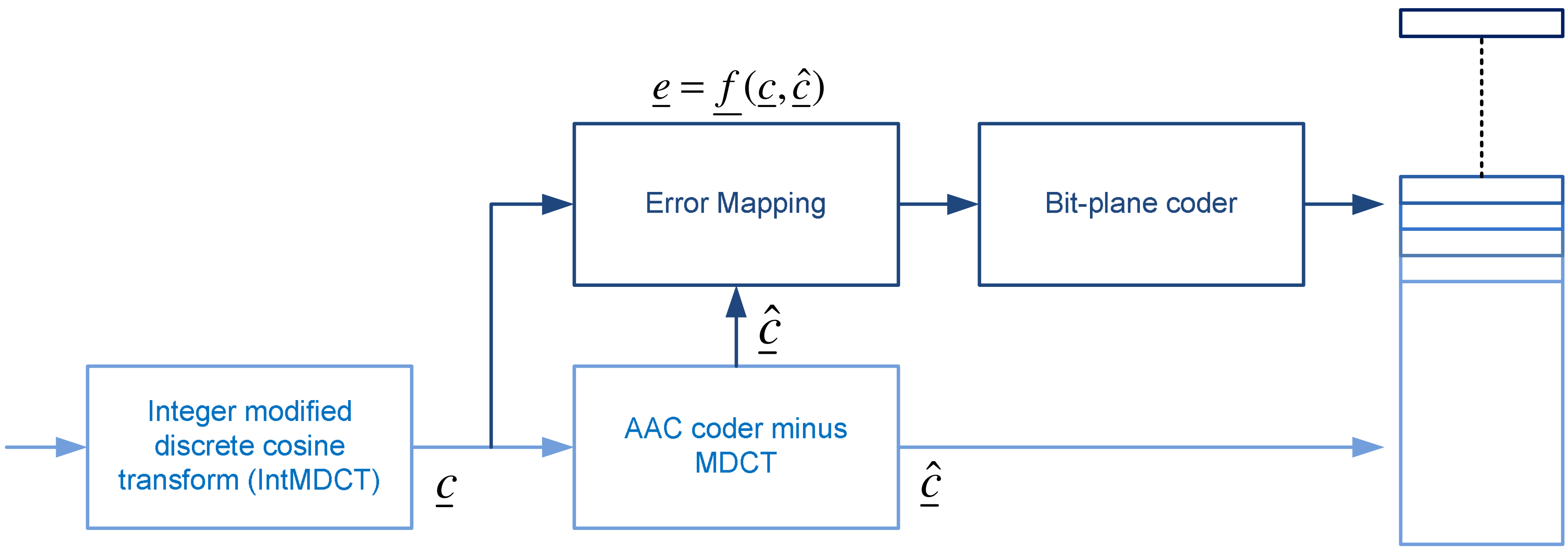
- MPEG SLS standard:
  - Lossless compression of audio
  - Bitstream scalability with Advanced Audio Coding (AAC) base layer
  - Enhancement layers are fine grained

Bitstream for a frame of audio



Introduction

- Structure of SLS encoder

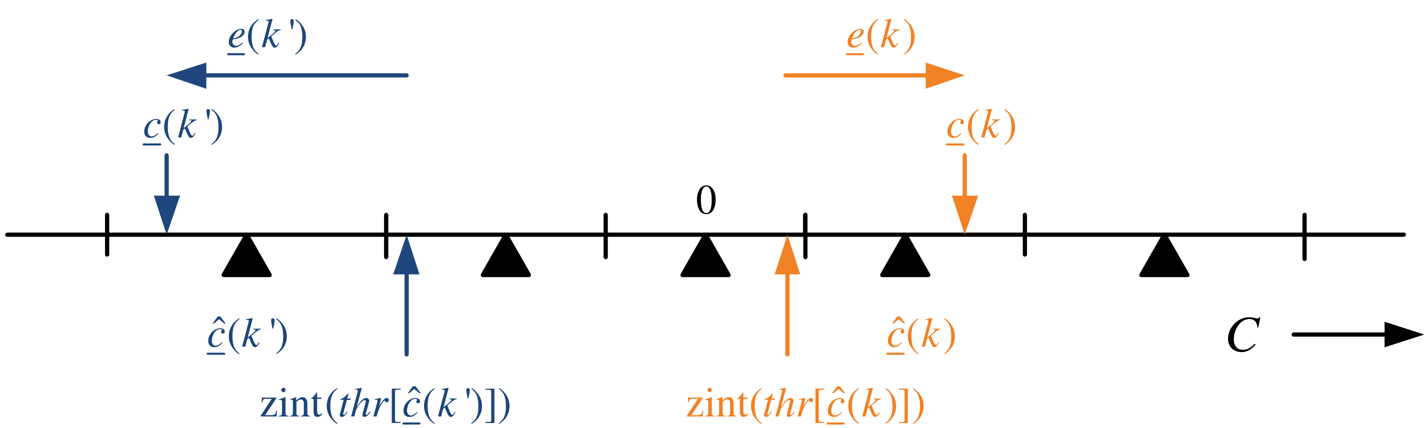


Introduction

- Error mapping:
  - If $\underline{\hat{c}}(k) = 0,$

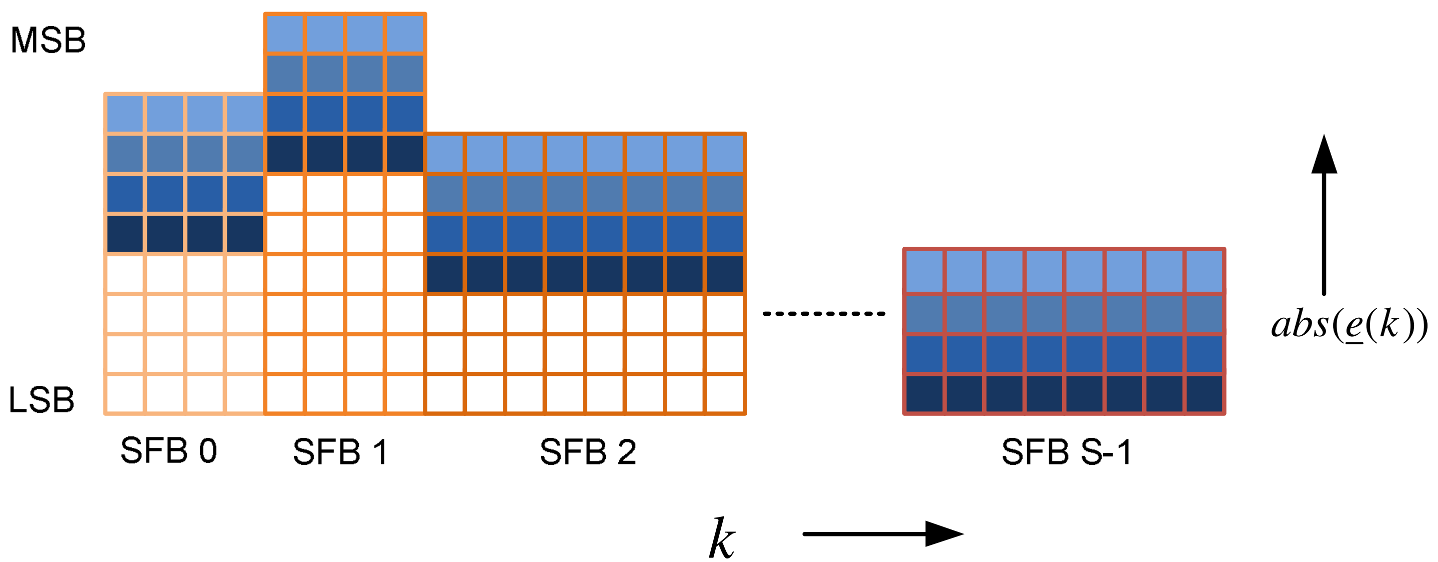
$$\underline{e}(k) = \underline{c}(k) - \underline{\hat{c}}(k) = \underline{c}(k)$$

  - Else $\underline{e}(k) = \underline{c}(k) - \text{zint}(thr[\underline{\hat{c}}(k)])$



Introduction

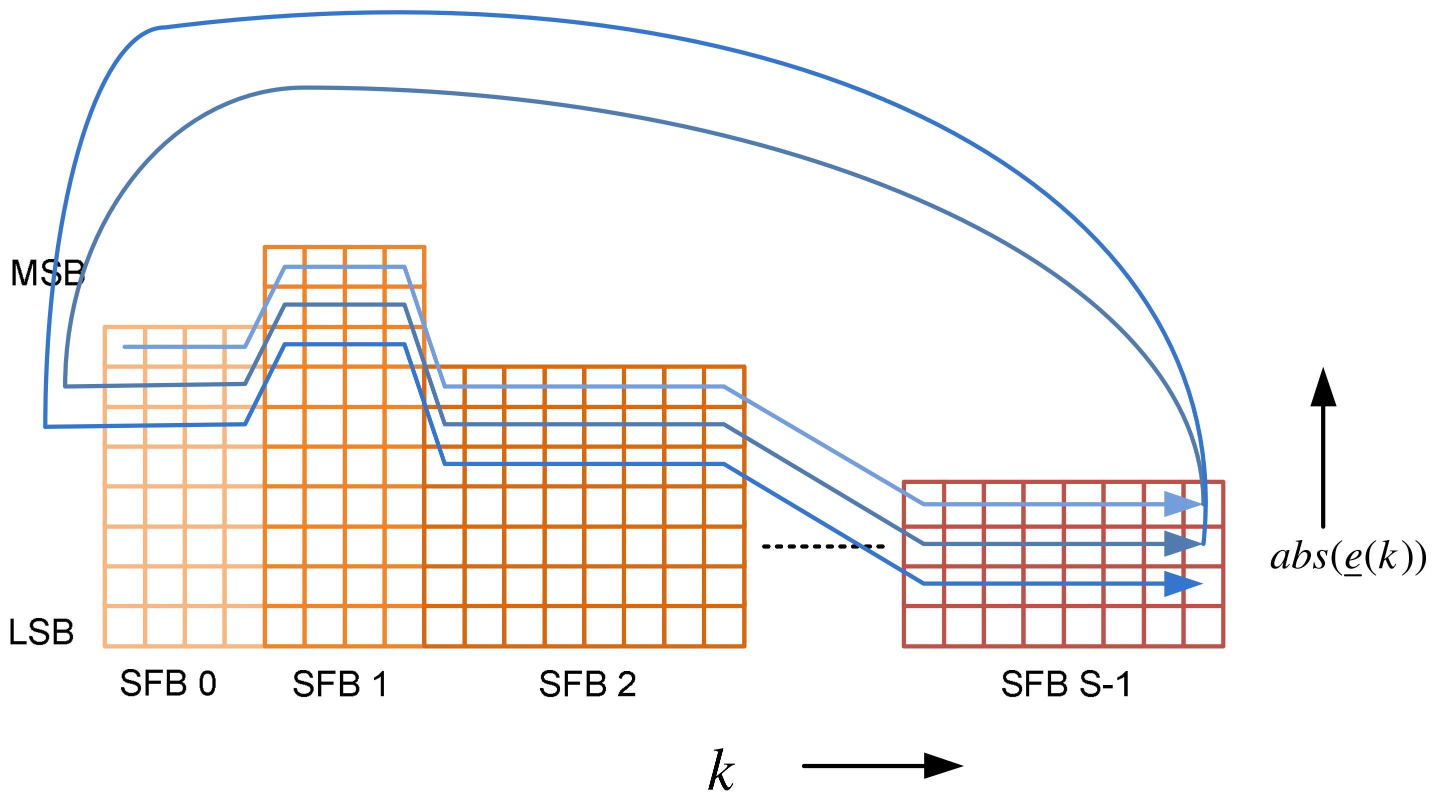
- Bit-plane coding



- SLS header information includes MSB plane number (if necessary)

Introduction

- Bits are arithmetic coded in a pre-defined sequence



- Naturally the bit-stream can be truncated at any bit

Introduction

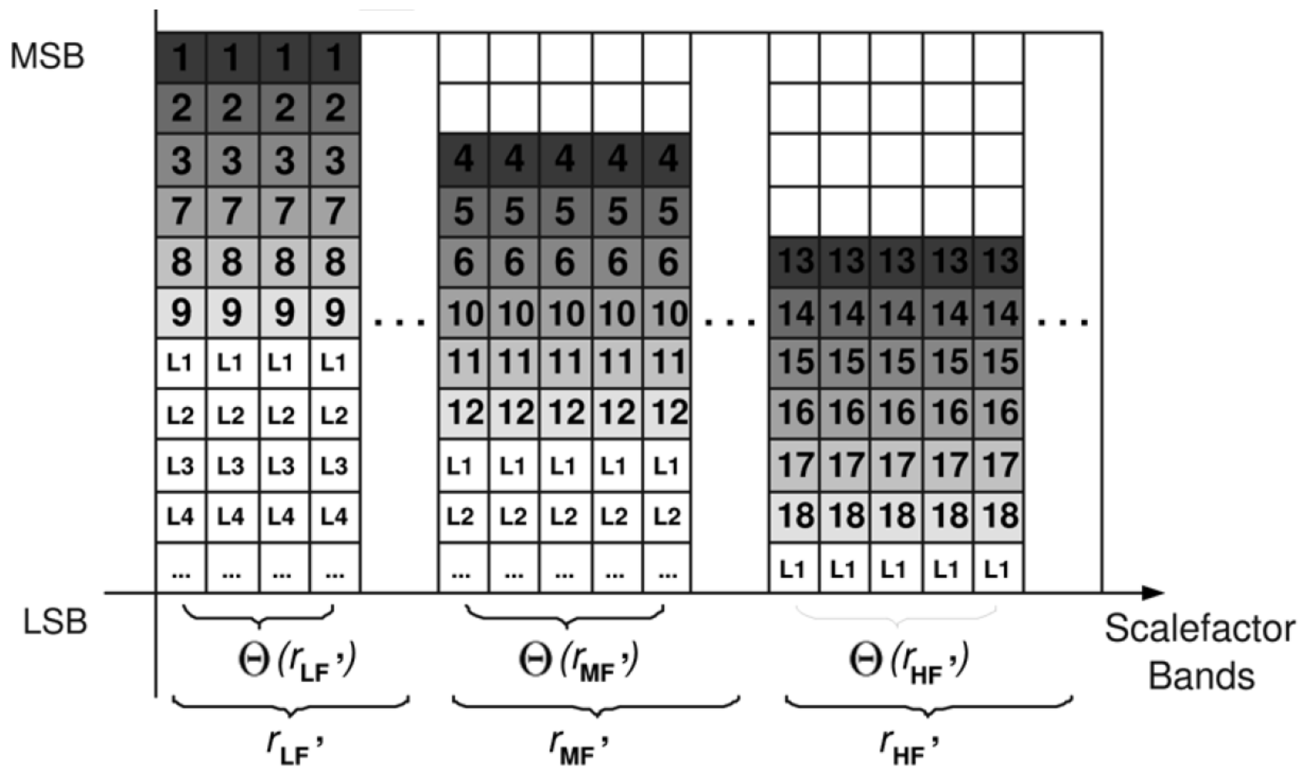
- Two types of probability assignments for arithmetic coding
- Bit-plane Golomb Code (BPGC) and Context-based Arithmetic Coding (CBAC)
- P(bit=1) < P(bit=0) generally, more so at higher bit-planes
- Code parameters sent in SLS header for each SFB

Drawbacks of SLS

- SLS bit coding sequence is very restrictive
- Bit allocation at intermediate layers (i.e., when bitstream is truncated) is perceptually blind
- Poor quality especially when base layer bit-rate is low
- No flexibility for optimization or intelligent bit-allocation at intermediate layers

Prior work

- Frequency region-based prioritized bit-plane coding (Li, Rahardja and Koh) – 2008



Courtesy: Li, Rahardja and Koh (TASLP 2008)

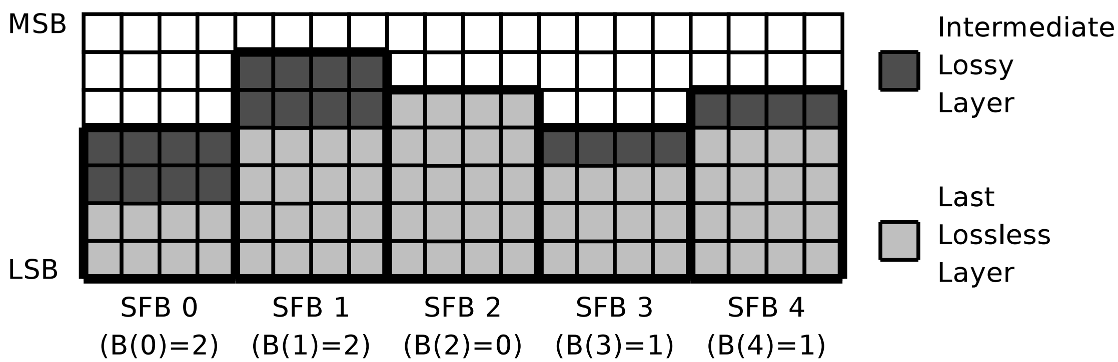
Prior work

- Li et al. found 2 empirical models: each corresponding to different bit-plane sequences
- Model choice is the only extra side-information
- Again, no flexibility for optimization
- Referred to as perceptual SLS (PSLS)

Proposed scheme

- Additional parameter $B(s)$ for each SFB s

- $B(s)$ is the number of bit-planes of the SFB coded consecutively



$$0 \le B(s) \le \min(\ M(s)+1,\ 7\ )$$

Proposed scheme

- Parameters $\mathbf{B} = \{B(0), B(1), \ldots, B(S-1)\}$ can be chosen to guarantee improved quality at a bit-rate other than at the base-layer

- Rate-distortion optimized selection:

$$\mathbf{B}^* = \arg \min_{\mathbf{B}:\mathrm{R}(\mathbf{B}) \leq \mathrm{R}_t} D(\mathbf{B})$$

Proposed scheme

- Distortion per frame - average noise-to-mask ratio (ANMR):

$$D(\mathbf{B}) = \frac{1}{S}\sum_{s=0}^{S-1}\frac{1}{m_s}\left(\sum_{k\in SFB_s}[\underline{c}(k)-\underline{\hat{c}}_{B(s)}(k)]^2\right)$$

Proposed scheme

- Raw encoding of $B(s)$ requires 3 bits per SFB
- Affects lossless compression performance
- Experiments indicate that run-length encoding of $B(s)$ values reduces this to 1.5 bits per SFB
- Coding format: 3 bits for $B(s)$ when $B(s) \neq B(s-1)$ and 5 bits to indicate the run-length

Proposed scheme

- To encode $B(s)$: $F(B(s), B(s-1))$ bits
- To encode MSB plane and coding probability models (if needed) : $G(B(s))$ bits
- Arithmetic coded bits (spectral data):

$$H(s) = \sum_{j=M(s)+1-B(s)}^{M(s)} -\log_2[b[k,j]P(b[k,j]=1)+(1-b[k,j])P(b[k,j]=0)]$$

$$\forall k \in SFB\ s$$

- Total rate at intermediate-layer:

$$R(\mathbf{B}) = \sum_{s=0}^{S-1}[F(B(s), B(s-1) + G(B(s)) + H(B(s))]$$

Proposed scheme

- Lagrangian-based optimization
- The cost function to be minimized is the Lagrangian

$$J(\lambda, \mathbf{B}) = D(\mathbf{B}) + \lambda R(\mathbf{B})$$

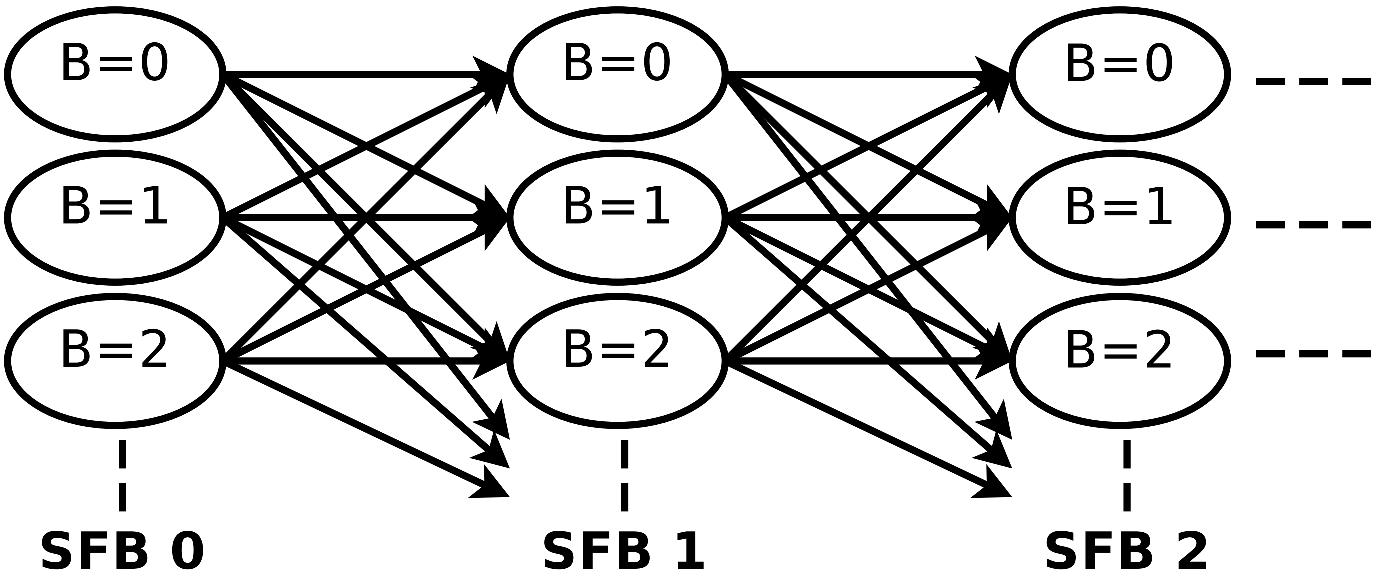
- Find parameters $\mathbf{B}^*$ that minimize $J(\lambda, \mathbf{B})$
- Iterate over λ till rate constraint achieved

Proposed scheme

- Brute force minimization has complexity as large as 8^S
- Cannot separately optimize over each SFB as $B(s)$ is run-length coded
- Trellis-based optimization is thus a natural choice

Trellis-based optimization

- Trellis used for the optimization



Trellis-based optimization

- Associate each state with
  - Distortion costs: $d_s(B(s)) = \frac{1}{S}\left(\frac{1}{m_s} \sum_{k \in SFB\ s} [\underline{c}(k) - \hat{\underline{c}}_{B(s)}(k)]^2 \right)$
  - Rate costs: $\lambda[G(B(s)) + H(B(s))]$

Trellis-based optimization

- Associate transition $B(s-1) \rightarrow B(s)$ with cost:

$$\lambda F(B(s-1), B(s))$$

- Use Viterbi algorithm to find path of minimum cost and hence corresponding $\mathbf{B}^*$

Results

- Codecs under comparison: SLS, PSLS, and proposed method

- AAC base-layer optimized with a trellis approach (Aggarwal et al.) for all 3 codecs

- Comparison of:
  - Quality at intermediate bit-rate
  - Lossless compression performance

- 15 standard test files (48kHz, 16bit PCM, mono)

Results

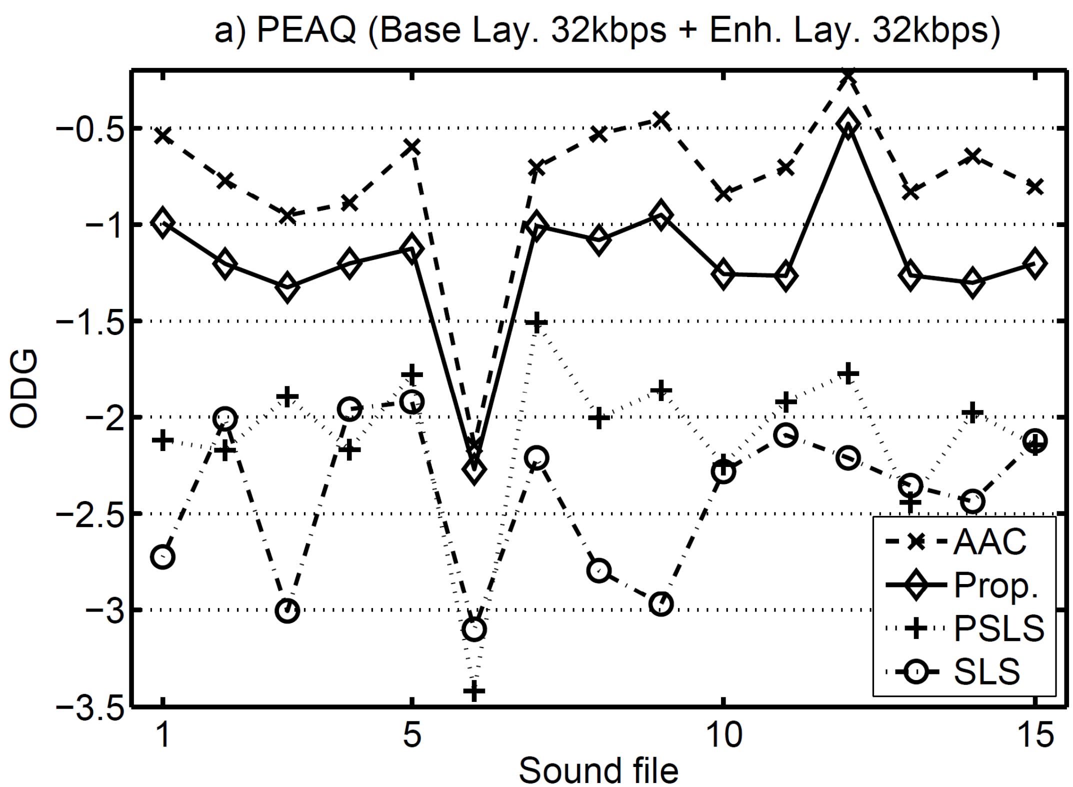
- Performance in terms of distortion metric ANMR (in dB)

	32+32	32+64	64+32	64+64
SLS	-2.18	-5.12	-6.68	-9.69
PSLS	-1.38	-5.74	-5.36	-9.76
Proposal	-3.90	-7.78	-8.09	-11.45
Non-scalable AAC	-5.05	-9.48	-9.48	-13.81

X+Y indicates base-layer bit-rate of X kbps and intermediate-layer bit-rate of Y kbps

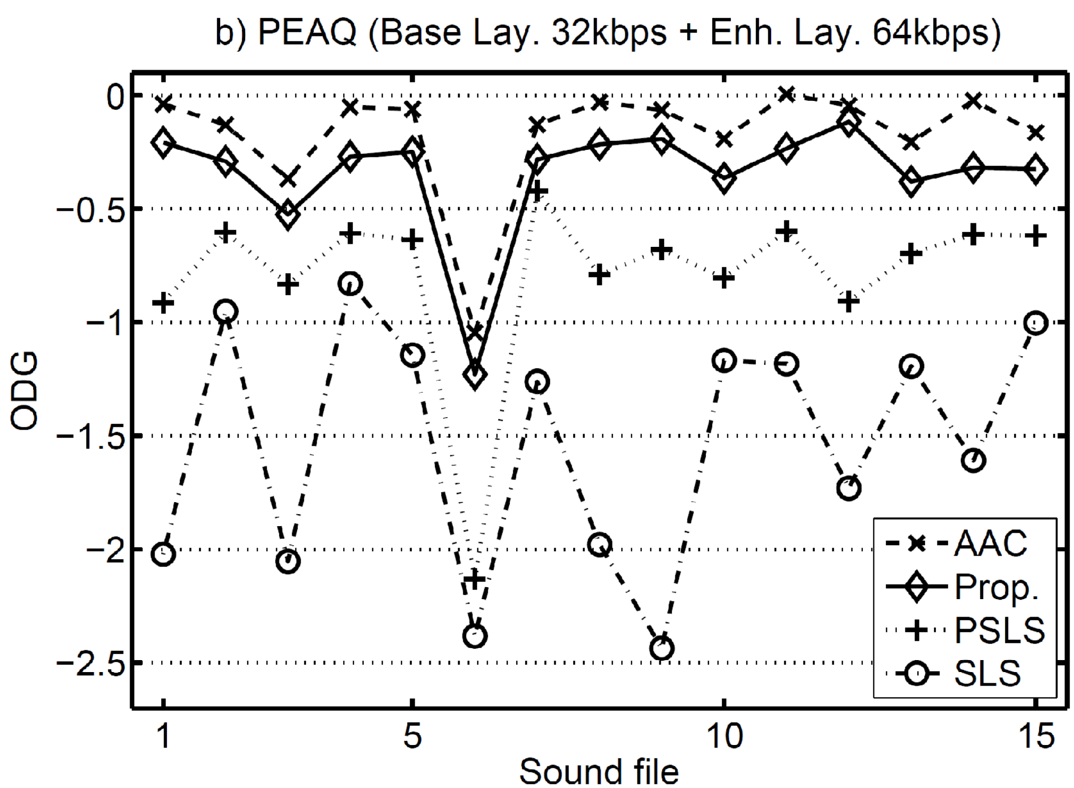
Results

- Performance in terms of objective difference grades (ODGs) by the perceptual evaluation of audio quality (PEAQ) method



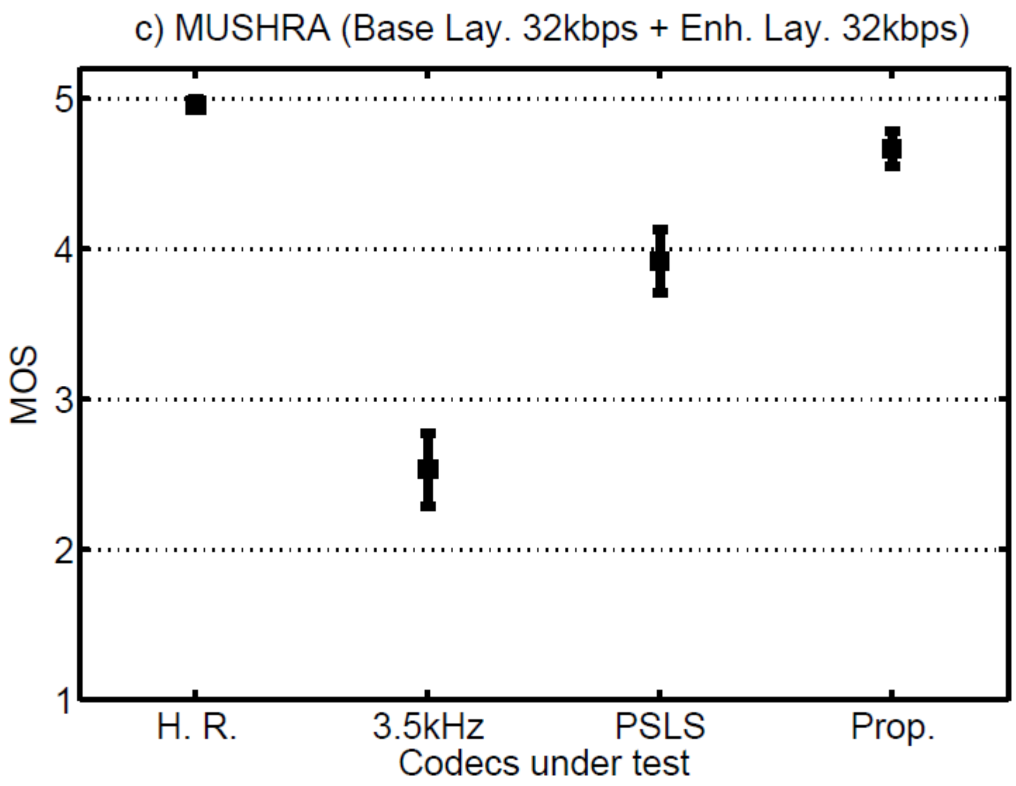
Results

- Performance in terms of objective difference grades (ODGs) by the perceptual evaluation of audio quality (PEAQ) method



Results

- Performance in terms of MUSHRA scores (subjective tests)



Results

- Lossless compression performance in terms of compression ratio

	SLS	PSLS	Proposal R_t=32kbps	Proposal R_t=64kbps
AAC base at 32kbps	2.08	2.08 (-0.01%)	2.06 (-0.93%)	2.06 (-0.99%)
AAC base at 64kbps	2.05	2.05 (-0.01%)	2.03 (-1.06%)	2.03 (-1.08%)

Conclusion

- Proposed an SLS based scheme with perceptually enhancement at intermediate layers
- Flexible coding format that is amenable to optimization
- Trellis-based optimization of the proposed coding scheme resulted in subjective and objective quality gains

Thanks!