Optimal Delayed Decisions in Encoding Audio Signals

Vinay Melkote and Kenneth Rose

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

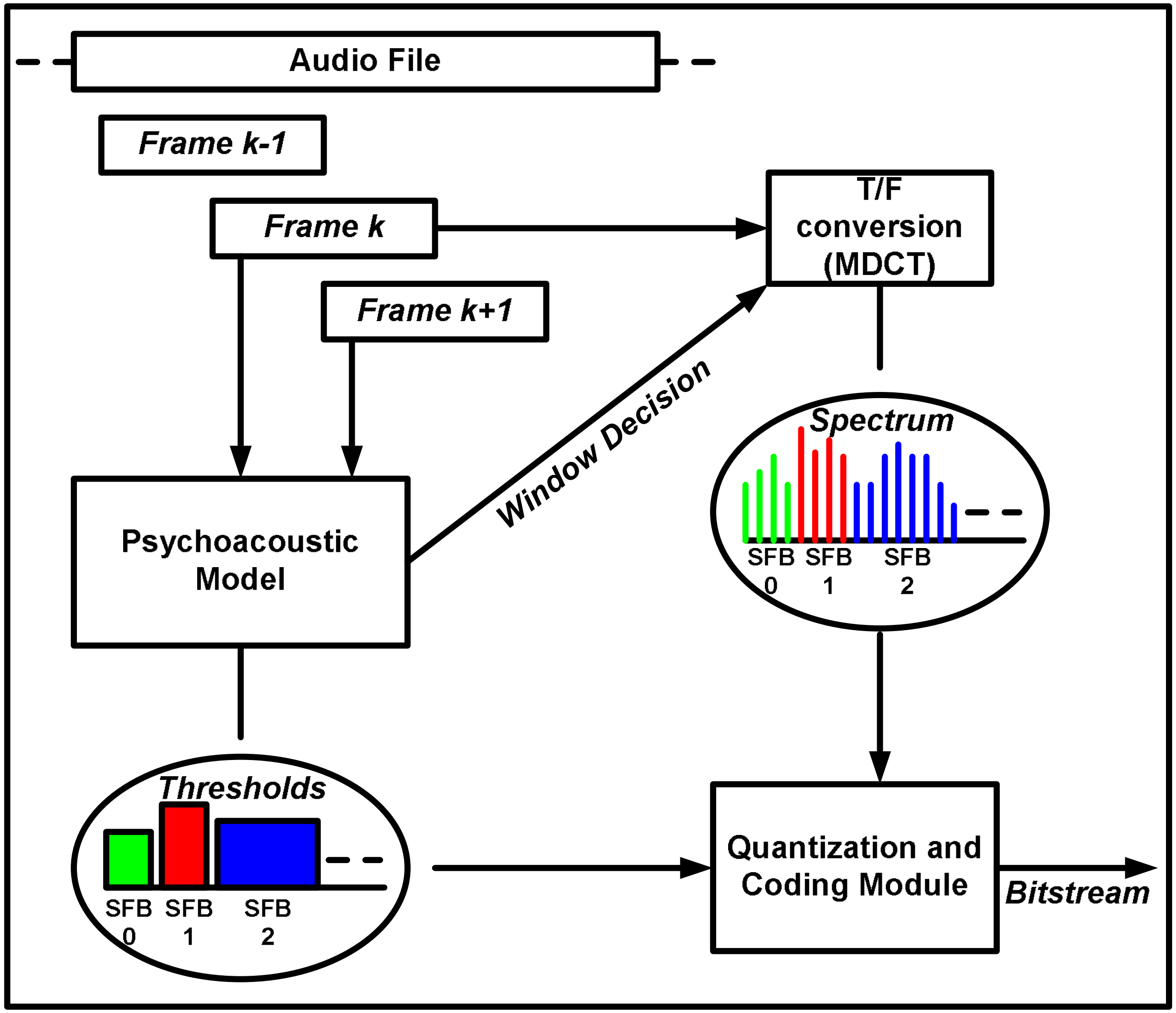
Introduction

- Audio for streaming, storage, gaming, etc., compressed off-line
- Encoding delay not critical to end-user experience
- Encoders typically constrain delay – parameters selected frame after frame
- Can delayed decisions improve the quality of coded audio?

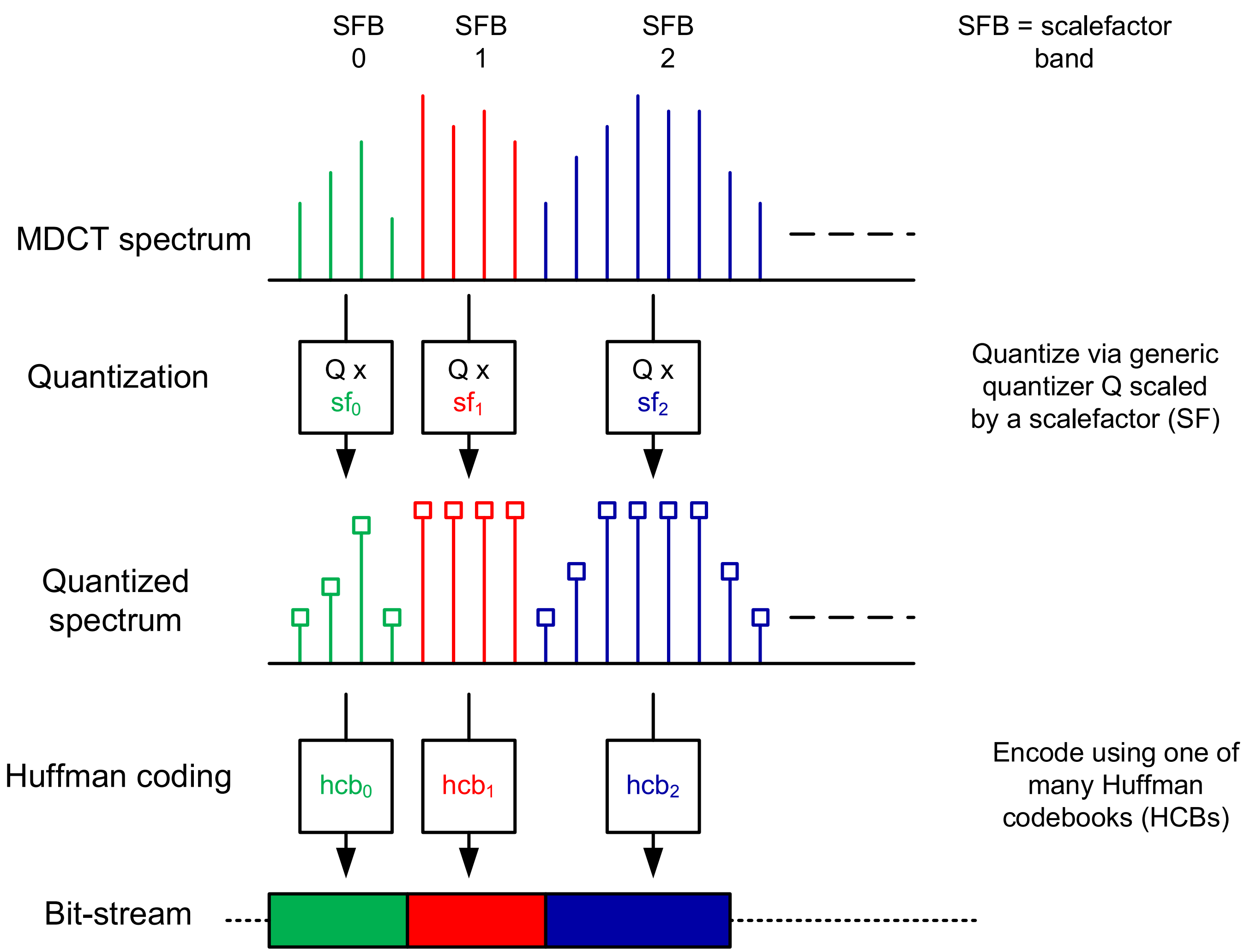
Proposed idea

- Increase encoding delay, and optimize decisions across multiple frames
- Encoder modification: no additional decoding delay
- Compatible with standard decoder
- Encoder parameter selection in a Lagrangian-based RD optimization framework
- Navigate intra- and inter-frame parameter space via two-layered trellis

MPEG Advanced Audio Coding (AAC)

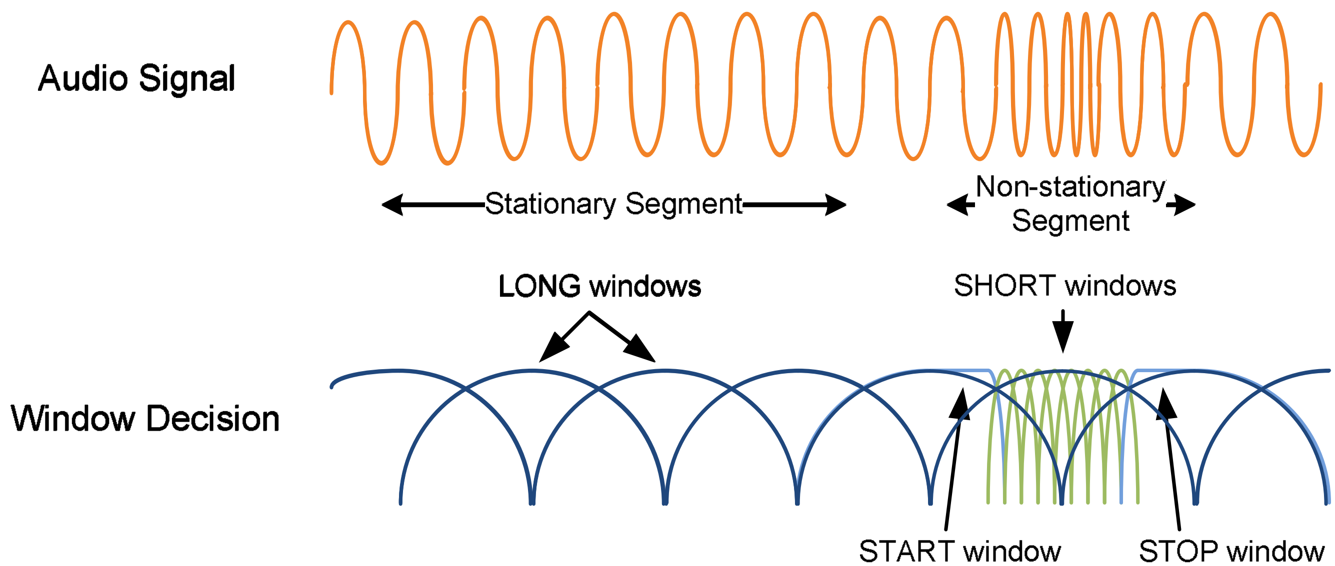


Quantization and coding decisions



- In each frame, SFs and HCBs need to be found for every SFB

Window decisions



- Stationary segments of the audio signal are encoded as LONG frames
- Non-stationary segments: replace LONG frame by 8 SHORT frames
- Insert appropriate transition windows (START and STOP frames)

Sub-optimalities in current encoders

- **Window decisions:**
  - LONG ↔ SHORT switching typically via heuristics of perceptual entropy or transient detection
  - No consideration of effect on neighboring frames that might need to be coded by START/STOP windows
- **Quantization and coding parameters:**
  - Scalefactors (SFs) and Huffman codebooks (HCBs) for each scalefactor band (SFB) found via two-loop search
  - Choice of SFs and HCBs separated into an inner distortion loop and outer rate loop, respectively
  - Fast, but sub-optimal, parameter selection

Sub-optimalities in current encoders

- **Bit-distribution across frames:**
  - Bits can be distributed unevenly to frames, with constraint on the average rate
  - Bit-reservoir technique – save bits when possible
  - Myopic approach results in inefficient bit-distribution

Problem statement

- Let P be the set of encoding decisions – window choice, scalefactors and Huffman codebooks - for *all* frames of the file

- The distortion for the entire file, $D_{overall}(P)$, and bit-rate, $R_{overall}(P)$, are dependent on P

- Objective: find $P^* = \arg \min_P D_{overall}(P)$ s.t. $R_{overall}(P) \le R_t$

- R_t is a target average rate
- Additionally, window switching constraints to be satisfied

Prior work

1. Optimal time segmentations for the MDCT [Niamut & Heudsens, '04]
2. RD optimal block switching [Boehm et al., '06]
3. Optimal bit-reservoir control [Camberlein & Philippe, '05]
4. Trellis-based optimal intra-frame parameter selection [Aggarwal et al., '06]
5. Multiple integer linear programming-based optimal intra-frame parameter selection [Bauer, '04]

Overview of the solution

- Convert the rate-constrained minimization to the minimization of an appropriate alternate cost function $J(P,\lambda)$ governed by a parameter λ
  - Ex: $J(P,\lambda) = D_{overall}(P) + \lambda R_{overall}(P)$, where λ is the Lagrange parameter

- Perform the unconstrained minimization:

$$P^*(\lambda) = \arg \min_P J(P,\lambda)$$

- If $R_{overall}(P^*(\lambda))$ not close to R_t , change λ and repeat minimization

Motivation for a trellis approach

- Size of the encoding parameter space:

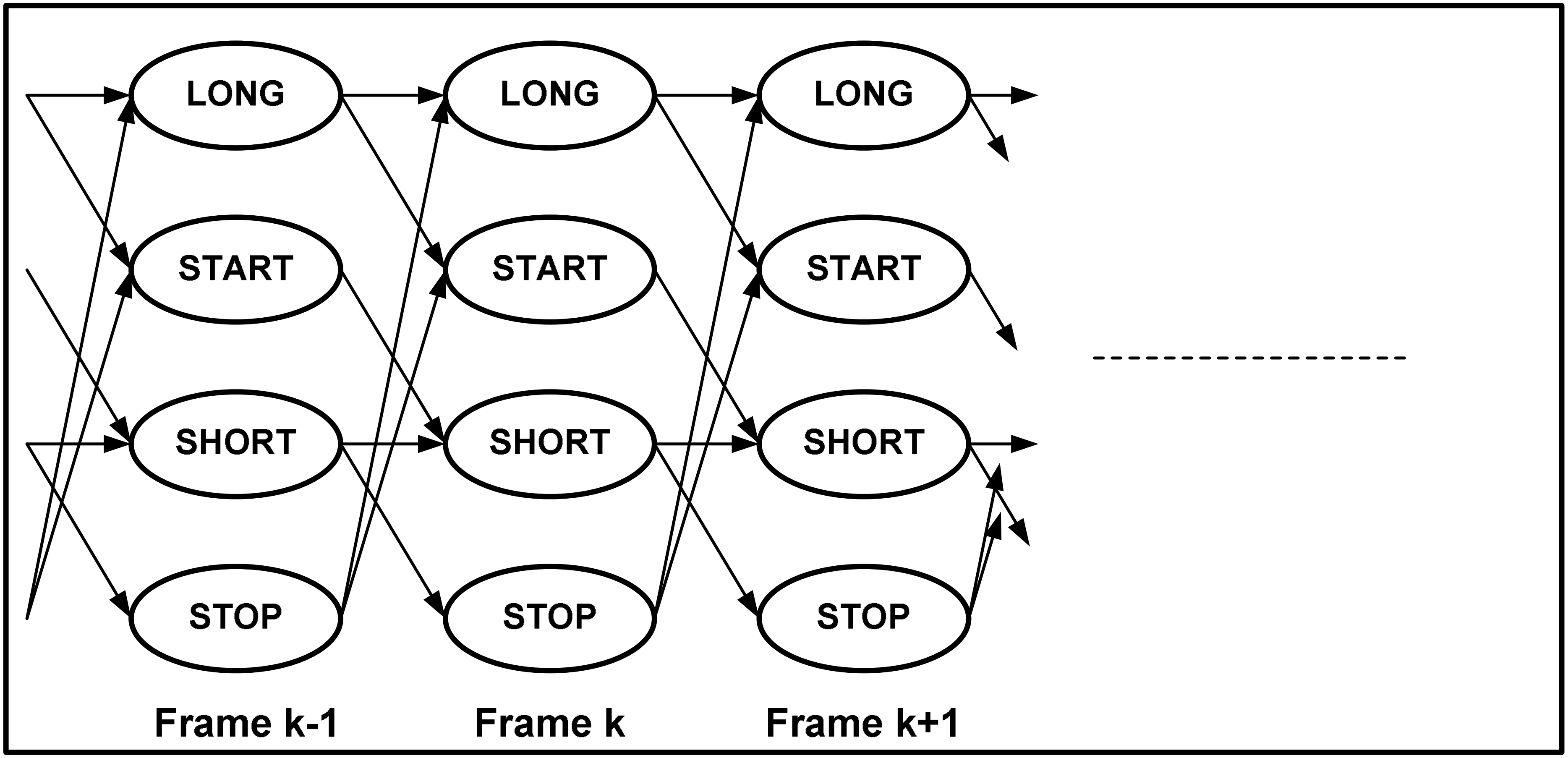
  - SF choices per SFB : 120
  - HCB choices per SFB : 12
  - Window choices per frame : 4

  - Cardinality of the set of values of $P \approx (4 \times (120 \times 12)^L)^K$

    L SFBs/frame
    K frames in the signal

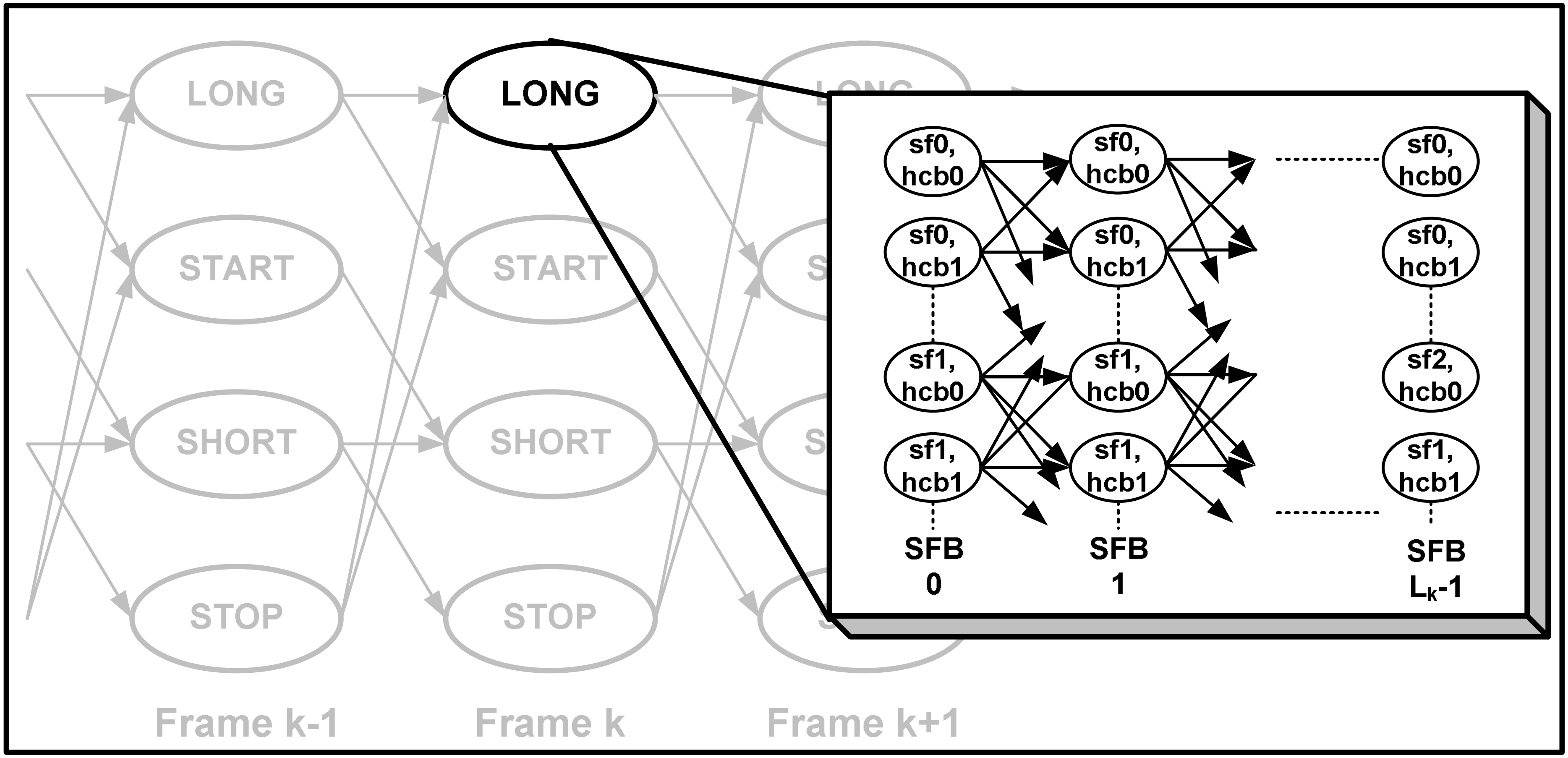
- Naïve search has complexity exponential in the number of frames and SFBs

Two-Layered Trellis: Outer Trellis



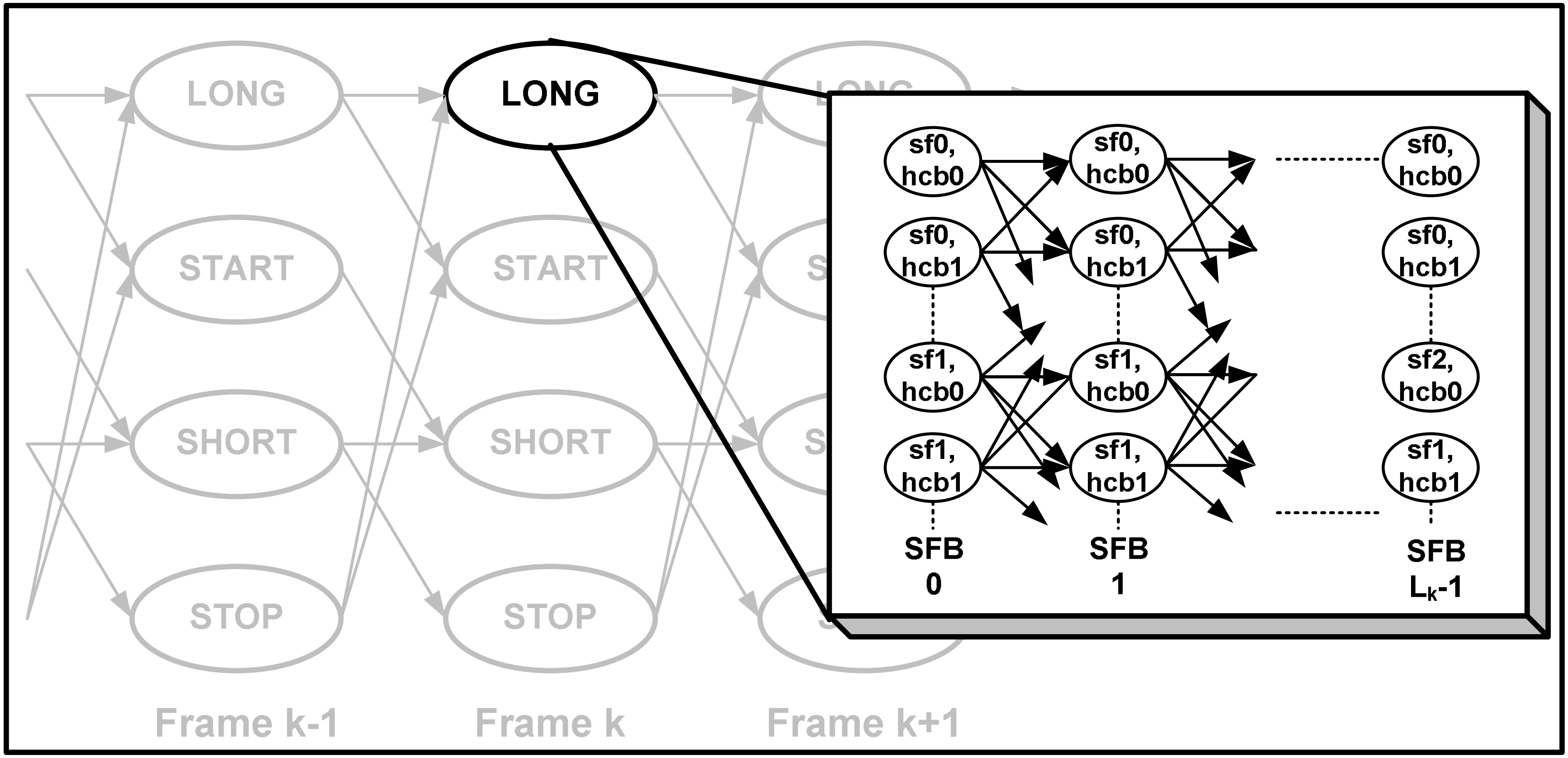
- Window switching trellis: paths correspond to allowed window sequences

Two-Layered Trellis: Inner Trellis



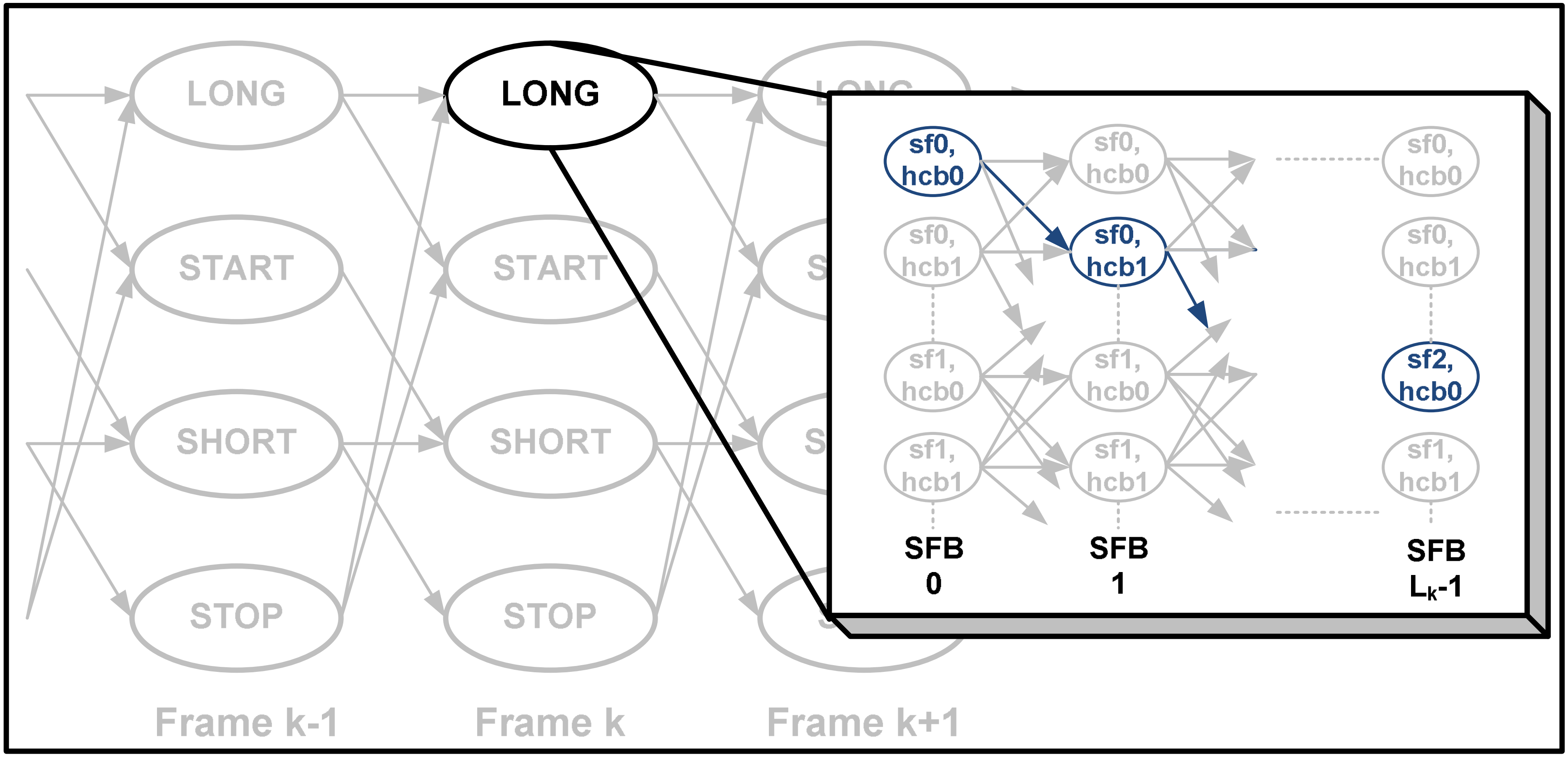
- Quantization and coding trellis: paths correspond to SF and HCB sets for each frame [Aggarwal et al., '06]

Two-Layered Trellis



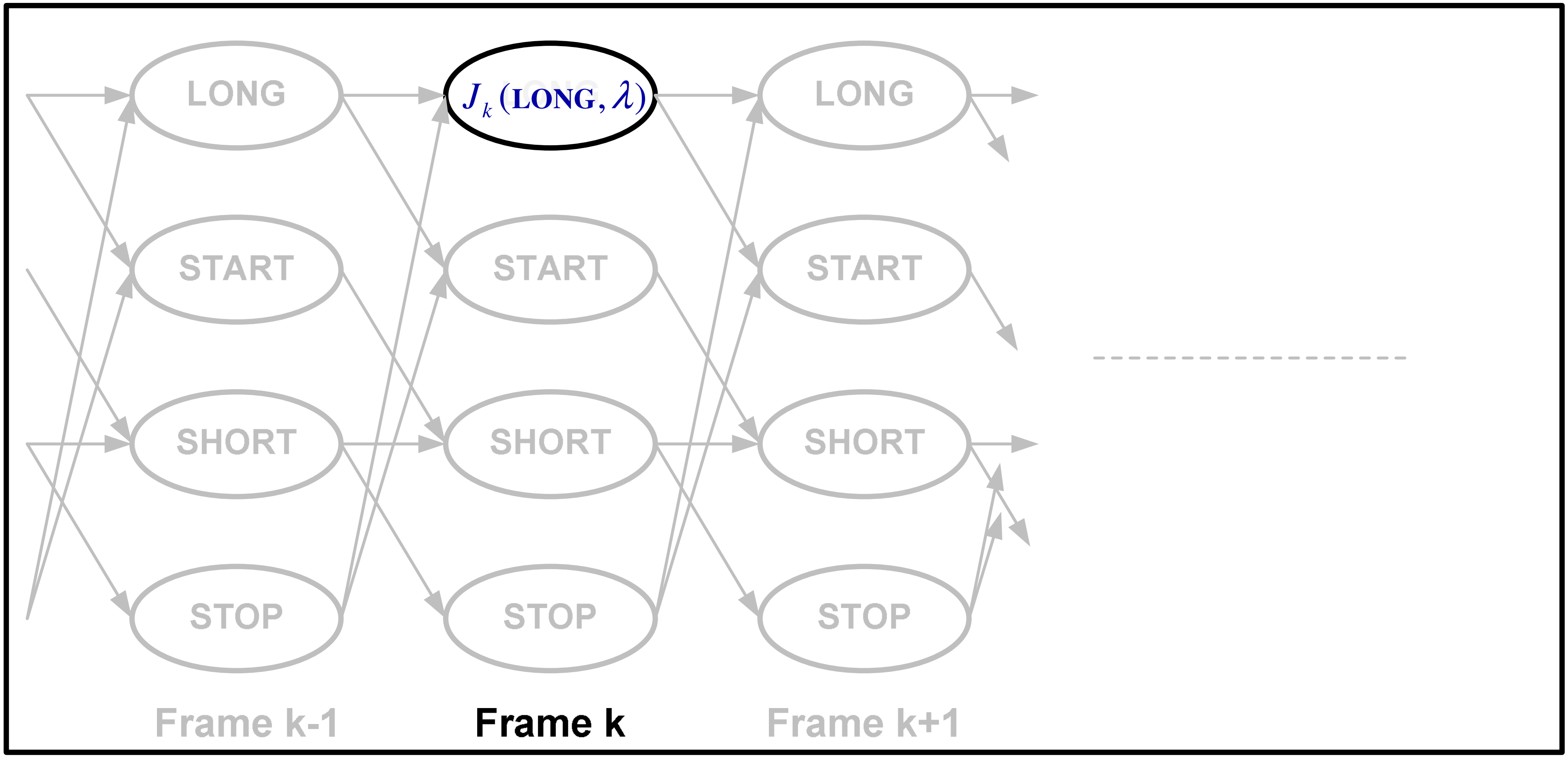
- Split overall cost $J(P,\lambda)$ into per band and per transition costs
- Employs the fact that each inner trellis state/transition is associated with a distortion value and/or number of bits

Two-Layered Trellis



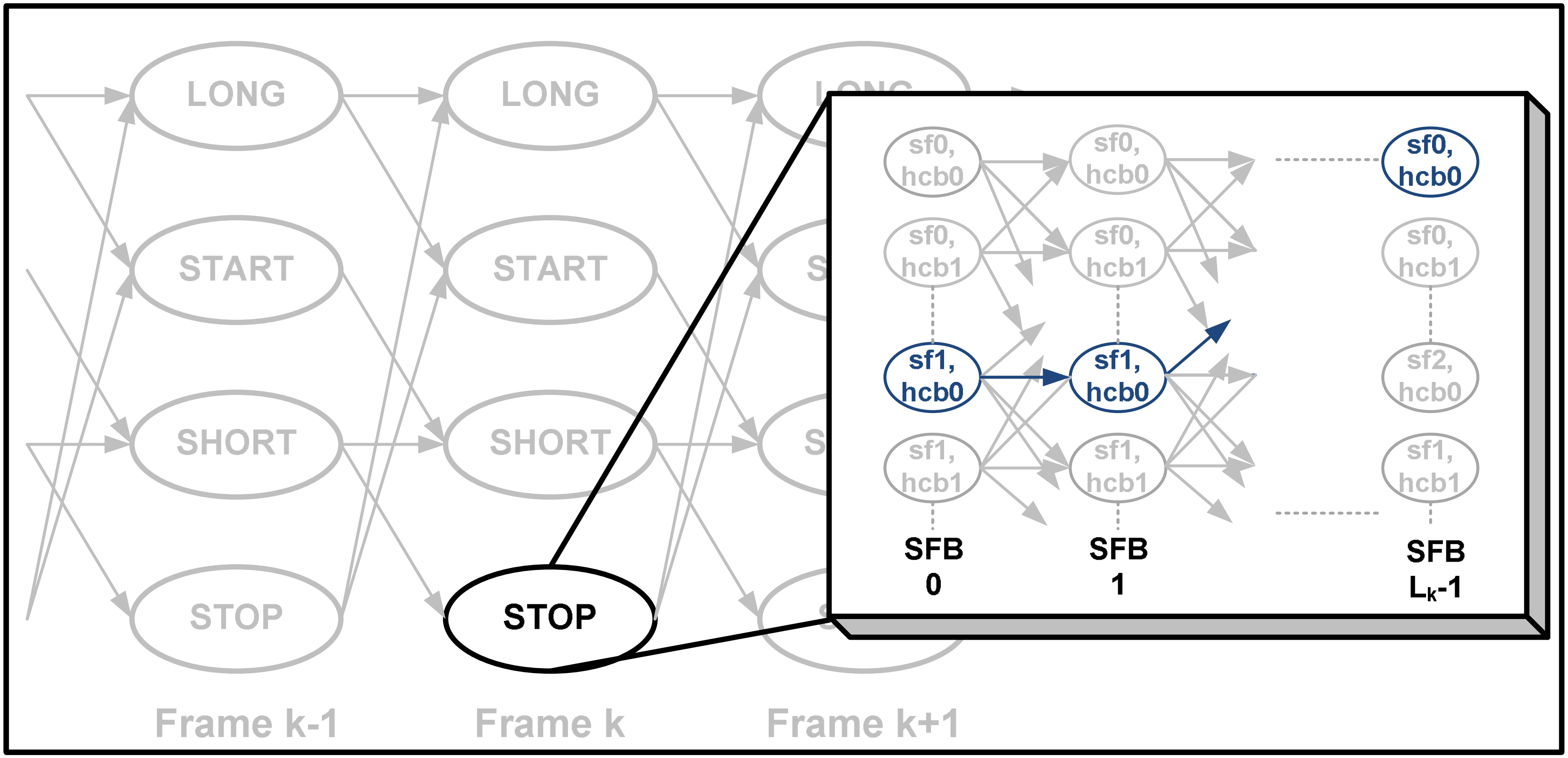
- Inner trellis path with minimum cumulative cost via Viterbi algorithm: optimal SF and HCB sequence for a frame in a particular window configuration
- Search complexity linear in the number of SFBs: $\approx O(L*(120*12)^2)$

Two-Layered Trellis



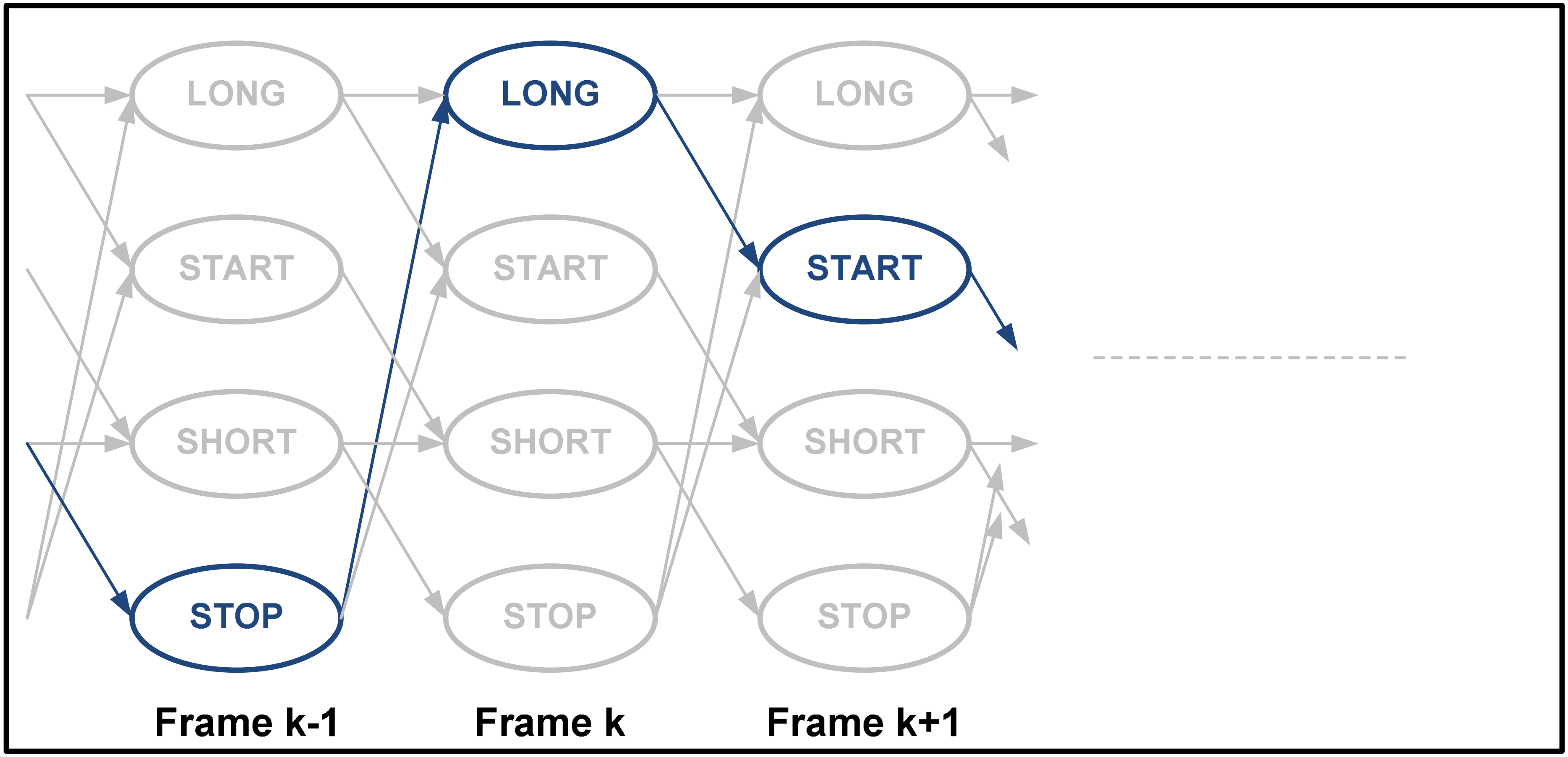
- Populate corresponding outer trellis node with the minimum inner trellis cumulative cost
- Obviously, this cost now depends only on the associated window choice

Two-Layered Trellis



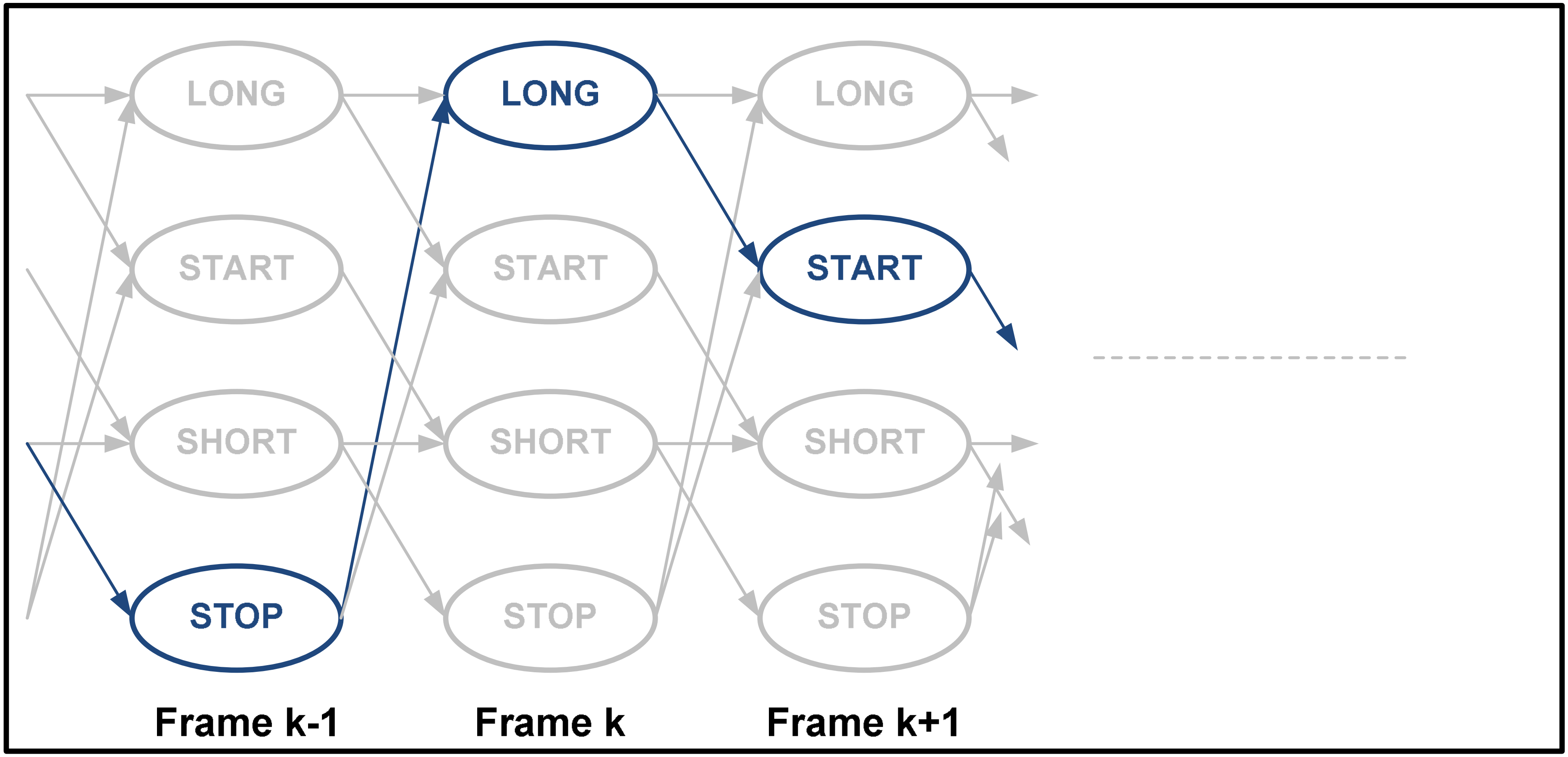
- Repeat the inner trellis algorithm in each window configuration for a frame

Two-Layered Trellis



- Viterbi algorithm in outer trellis for path/window decisions with minimum overall cost $J(P,\lambda)$: provides $P^*(\lambda)$ and $R_{overall}(P^*(\lambda))$
- Outer trellis complexity linear in the number of frames: $\approx O(K*7)$

Two-Layered Trellis



- If rate constraint not satisfied by $R_{overall}(P^*(\lambda))$, repeat search through the two-layered trellis with different λ

Results

- Overall distortion is defined as MTNMR: maximum over frames, of the total noise-to-mask ratio in each frame

$$D_{overall}(P) = \max_k \sum_l d_{k,l} \qquad 0 \le k \le K-1 \qquad 0 \le l \le L-1$$

- $d_{k,l}$ is the noise-to-mask ratio in SFB l of frame k

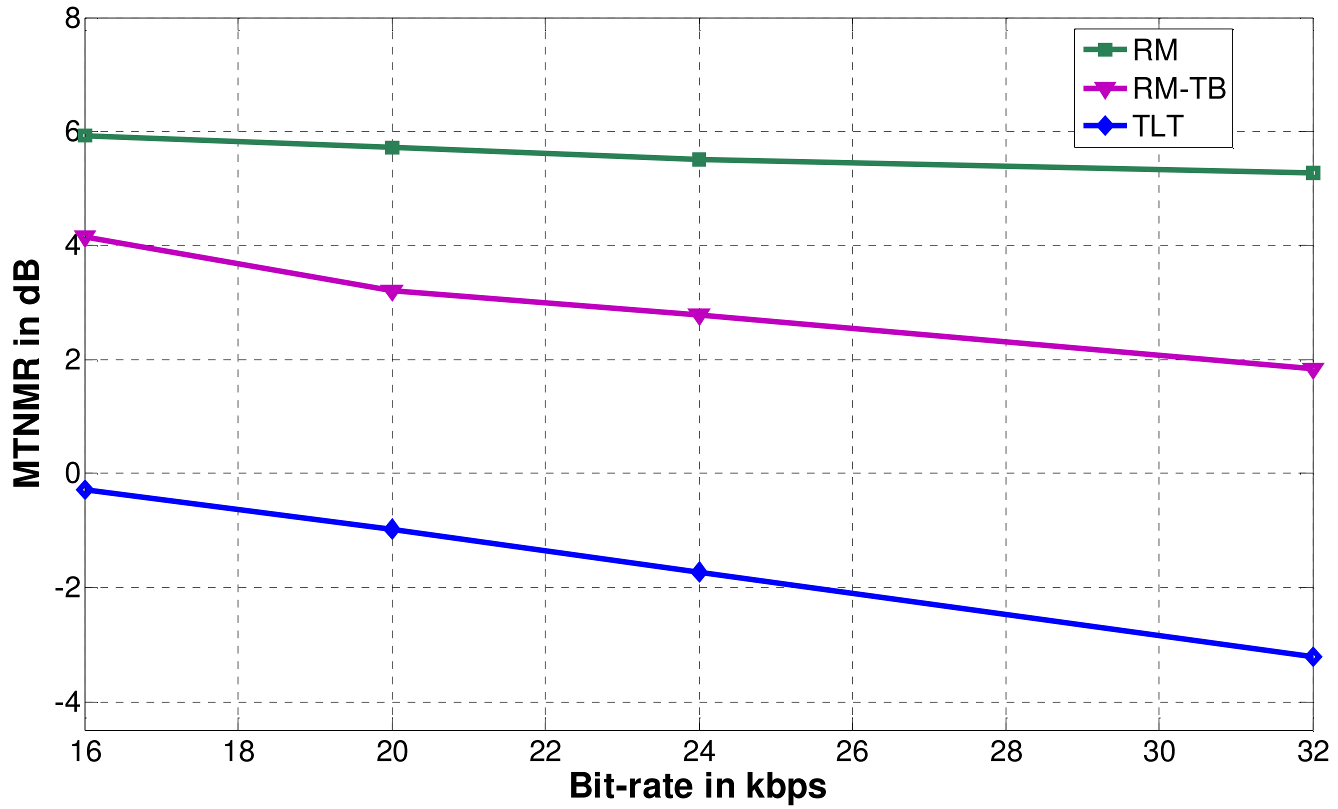
- An appropriate cost function can be defined, although ***not*** the Lagrangian $J(P,\lambda) = D_{overall}(P) + \lambda R_{overall}(P)$

Results

- Codecs under comparison:
  - MPEG reference model (RM): two loop search for SFs and HCBs, transient-detection based windows, bit-reservoir – no delayed decisions
  - Inner trellis-only model (RM-TB): inner trellis-based optimization of SFs and HCBs, transient-detection based windows, bit-reservoir – no delayed decisions
  - Two-layered trellis (TLT): overall optimization – delayed decisions

Results

Objective evaluation: in terms of distortion metric, MTNMR

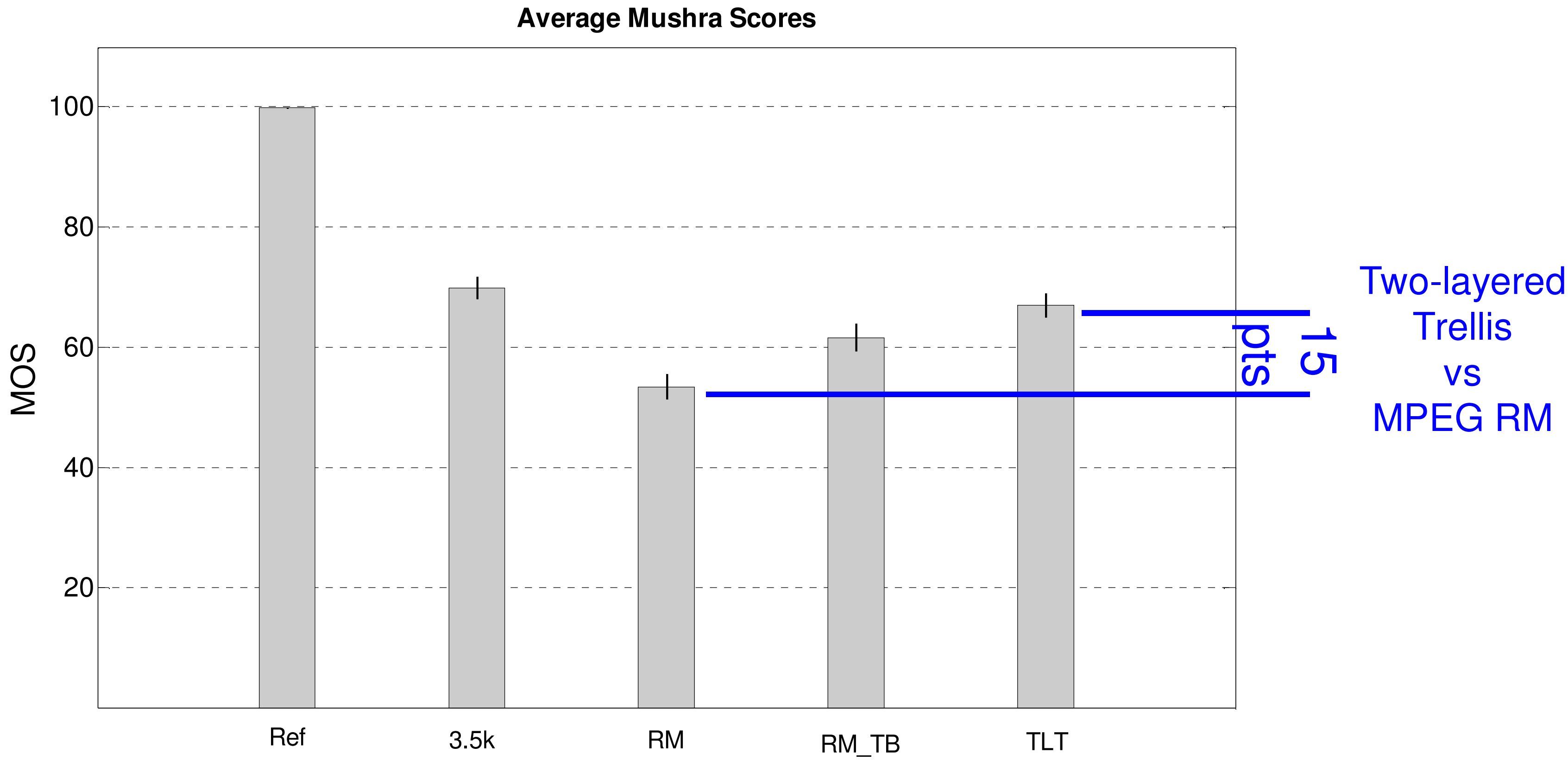


All this without changing the decoder !

- Distortion averaged over 10 different audio samples

Results

Subjective evaluation: multiple stimulus with hidden reference and anchor (MUSHRA) tests



- Scores averaged over 6 different audio samples: bit-rate = 16kbps, sampling rate = 44.1kHz, number of channels = mono

Some samples

Codec → / Sample ↓	Original	RM	RM - TB	TLT
Orchestra				
Accordion				
Glocken-spiel				

Summary

- Proposed a two-layered trellis approach for optimal delayed encoding of audio
- Bit-stream compatibility with the standard
- No additional decoding delay
- Substantial gains in objective and subjective quality metrics can be obtained by delayed decisions
- Particularly useful for applications that employ off-line compression