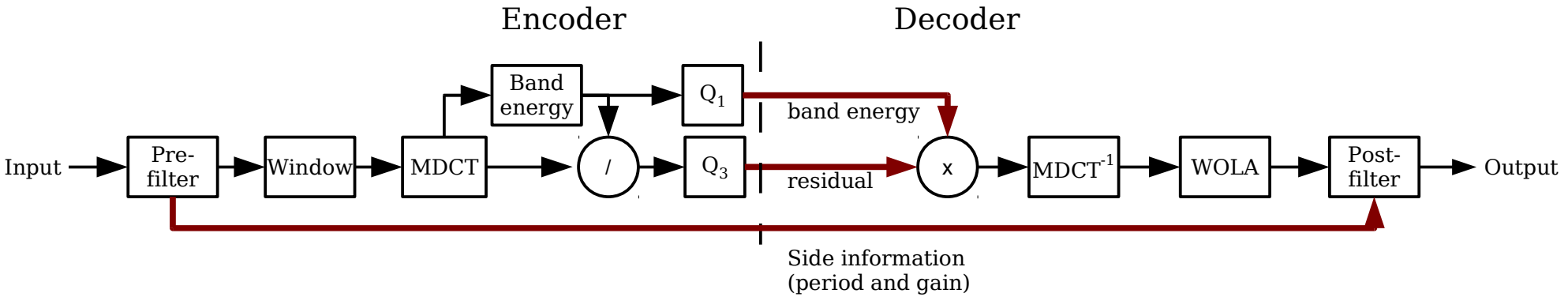
Overview of CELT

Jean-Marc Valin
Timothy B. Terriberry
Gregory Maxwell

CELT: Constrained Energy Lapped Transform

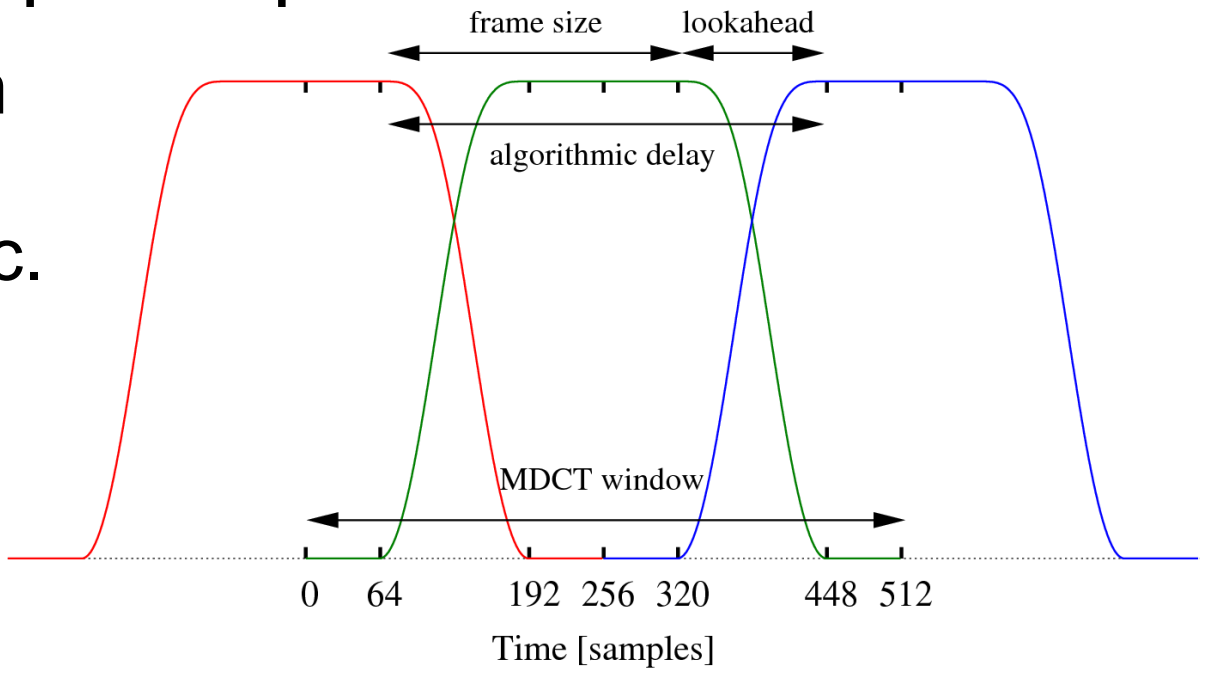
- Transform codec (MDCT, like MP3, Vorbis)
  - Short windows (5-22 ms), poor frequency resolution
- *Explicitly* code energy of each band of the signal
  - Coarse shape of sound preserved no matter what
- Code remaining details using vector quantization
- Variable time-frequency resolution for transients
- Now uses pitch post-filtering

Block Diagram



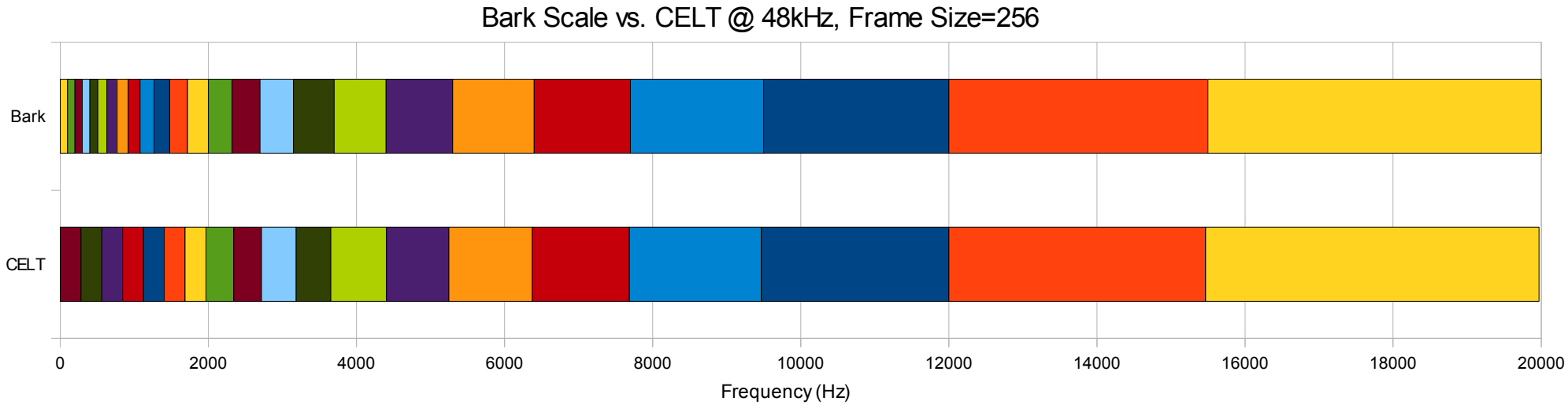
"Lapped Transform" Modified DCT

- The normal DCT causes coding artifacts (sharp discontinuities) between blocks, easily audible
- The "Modified" DCT (MDCT) uses a decaying window to overlap multiple blocks
  - Same transform used in MP3, Vorbis, AAC, etc.
  - But with much smaller blocks, less overlap



"Constrained Energy" Critical Bands

- Group MDCT coefficients into bands approximating the critical bands (Bark scale)



"Constrained Energy" Coding Band Energy

- Most important psychoacoustic lesson learned from Vorbis:

  Preserve the energy in each band

- Vorbis does this implicitly with its "floor curve"
- CELT codes the energy explicitly
  - Coarse energy (6 dB resolution), predicted from previous frame and from previous band
  - Fine energy, improves resolution where we have available bits, not predicted

Coding Band Shape

- After normalizing, each band is represented by an *N*-dimensional unit vector
  - Point on an *N*-dimensional sphere
  - Describes "shape" of energy within the band
- Code this shape using with vector quantization
  - Pyramid Vector Quantizer (Fischer, 1986)
- Vectors with integer coordinates whose magnitudes sum to *K*

$$S(N,K) = \{\mathbf{y} \in \mathbb{Z}^N : \sum_{i=1}^{N} |y_i| = K\}$$

Psychoacoustic Tricks

- Avoiding "birdie" artifacts
  - *K* may be small, giving a sparse spectrum > 8 kHz
  - Use N-D rotation to “spread” pulses
    - Makes the signal sound more “noisy” (less tonal)
    - Inverse rotation applied in the encoder
- Avoiding "pre-echo" artifacts
  - When a transient is detected, split the frame and do a smaller MDCT on each piece
  - Interleave the results and continue as normal

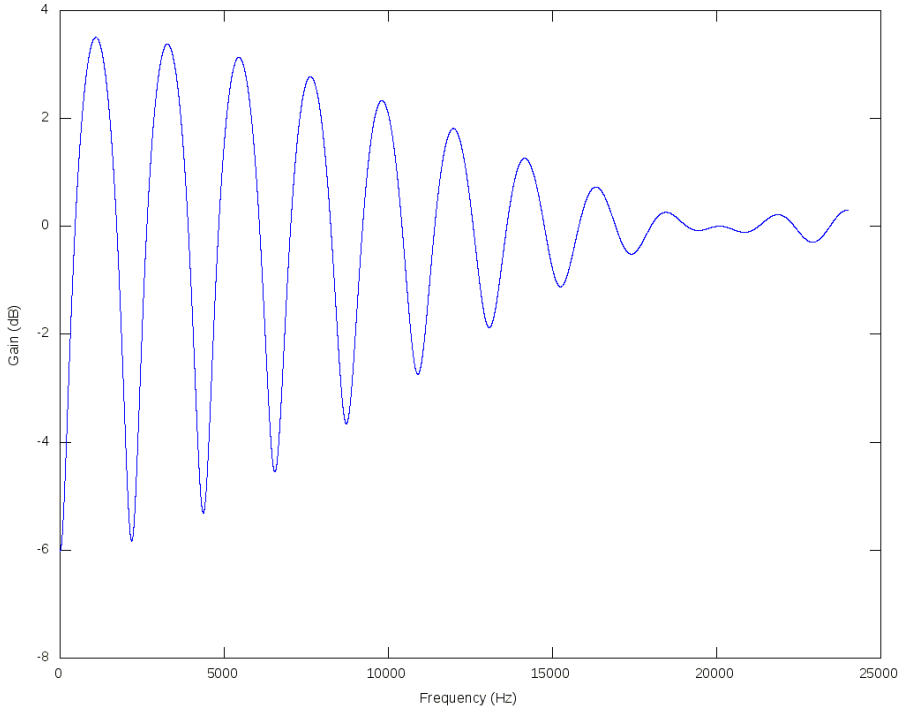
Time-Frequency Adjustment

- Finer transient control: TF resolution switching
  - Apply Hadamard transform in one band over multiple blocks
  - Forward transform: increase frequency resolution
  - Inverse transform: increase time resolution
- Two TF resolutions per frame
  - Decision coded per band
  - Handles mixed tonal/transient content

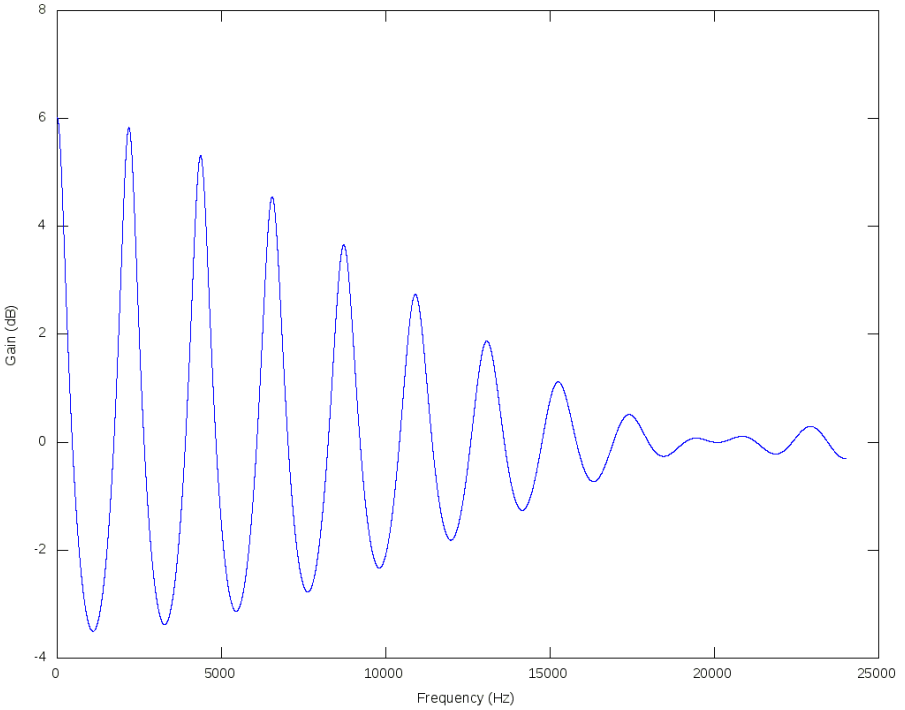
Post-Filter

- Proposed by Raymond Chen (Broadcom)
- Noise shaping for highly harmonic contents

Prefilter



Postfilter



Questions?