High-Quality, Low-Delay Music Coding in the Opus Codec

Jean-Marc Valin
Gregory Maxwell
Koen Vos
Timothy B. Terriberry

The Xiph.Org Foundation & The Mozilla Corporation

What is Opus?

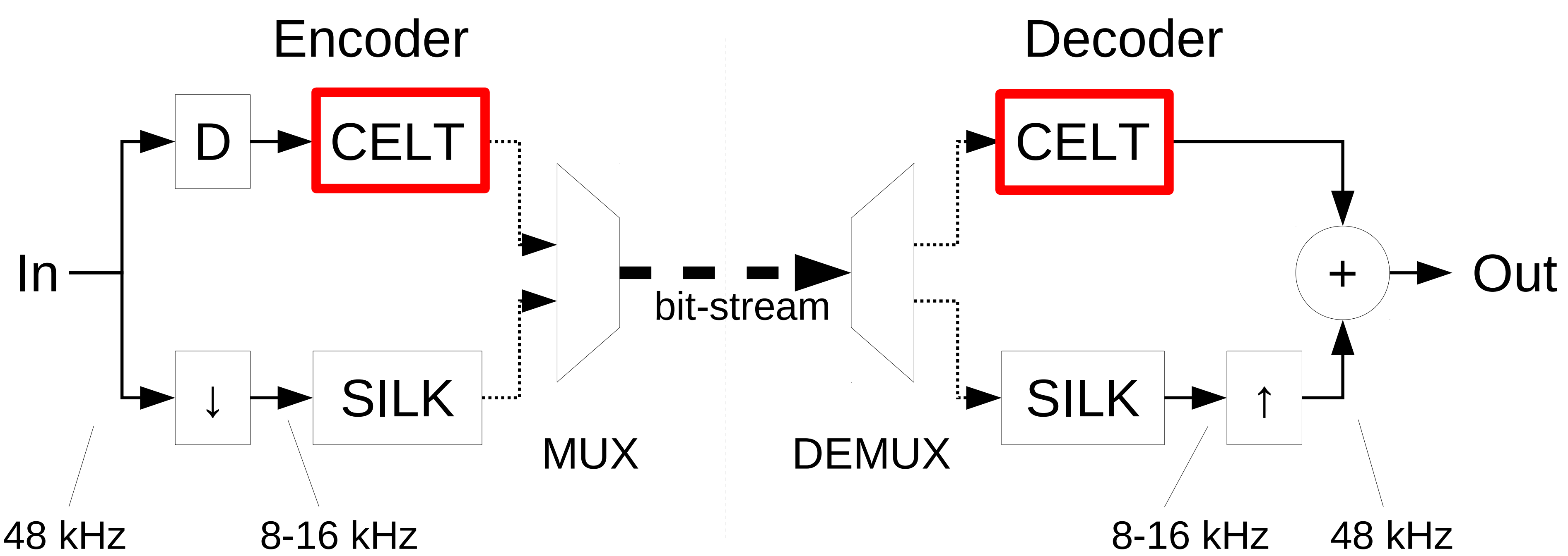
- New highly-flexible speech and audio codec
- Completely free
  - Royalty-free licensing
  - Open-source implementation
- IETF RFC 6716 (Sep. 2012)

Features

- Highly flexible
  - Bit-rates from 6 kb/s to 510 kb/s
  - Narrowband (8 kHz) to fullband (48 kHz)
  - Frame sizes from 2.5 ms to 60 ms
  - Speech and music support
  - Mono and stereo
  - Flexible rate control
  - Flexible complexity
- All changeable dynamically

Opus Operating Modes

- **SILK-only**: Narrowband, Mediumband or Wideband speech
- **Hybrid**: Super-wideband or Fullband speech
- **CELT-only**: Narrowband to Fullband music

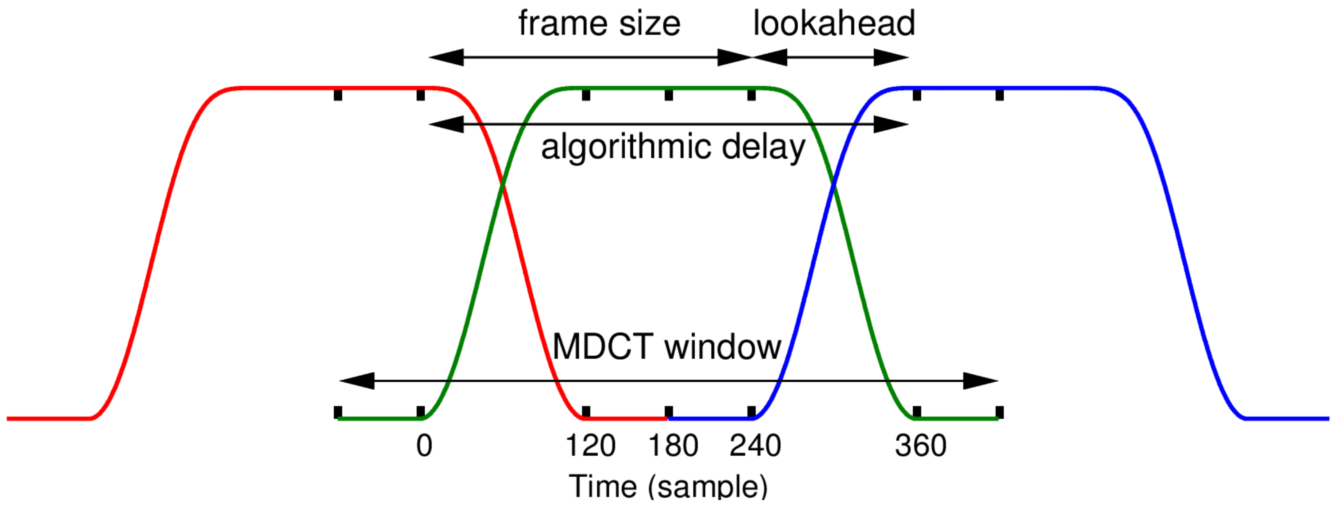


CELT: "Constrained Energy Lapped Transform"

- Transform coding with Modified Discrete Cosine Transform (MDCT)
- *Explicitly* code energy of each band of the signal
  - Spectral envelope preserved no matter what
- Code remaining details using algebraic VQ
  - Gain-shape quantization
- Implicit psychoacoustics and bit allocation
  - Built into the format

CELT Window

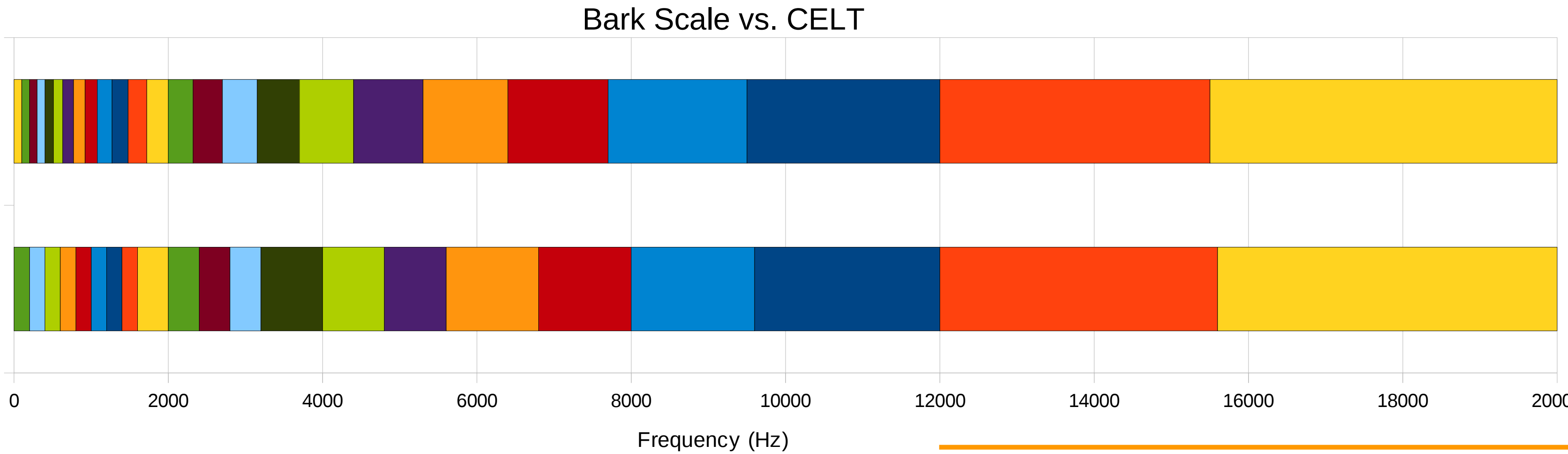
- MDCT with low-overlap window
  - Fixed 2.5 ms overlap for all sizes



- Overlap shape is like the Vorbis window
- Pre-emphasis reduces spectral leakage

Critical Bands

- Group MDCT coefficients into bands approximating the critical bands (Bark scale)
  - Band layout the same for all frame sizes
    - Need at least 1 coefficient for 120 sample frames
    - Corresponds to 8 coefficients for 960 sample frames



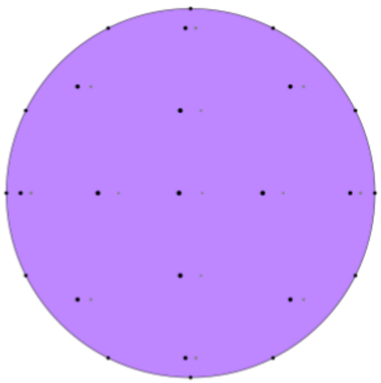
Coding Band Energy

- Energy computed for each band
- Coarse-fine strategy
  - Coarse energy quantization
    - Scalar quantization with 6 dB resolution
    - Predicted from previous frame and from previous band
    - Entropy-coded
  - Fine energy quantization
    - Variable resolution (based on bit allocation)
    - Not entropy coded

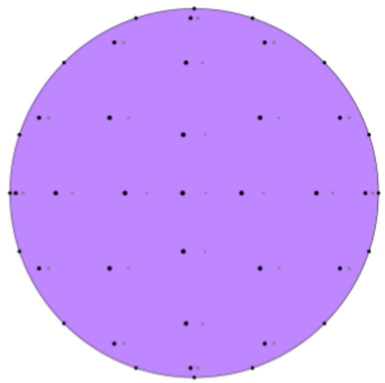
Coding Band Shape

- Quantizing *N*-dimensional vectors of unit norm
  - *N*-1 degrees of freedom (hyper-sphere)
  - Describes "shape" of spectrum within the band
- CELT uses *algebraic* vector quantization
  - Pyramid Vector Quantization (Fischer, 1986)
  - Combinations of *K* signed pulses
  - Set of vectors *y* such that $||y||_{L1} = K$
  - Projected on unit sphere: $x = y / ||y||_{L2}$

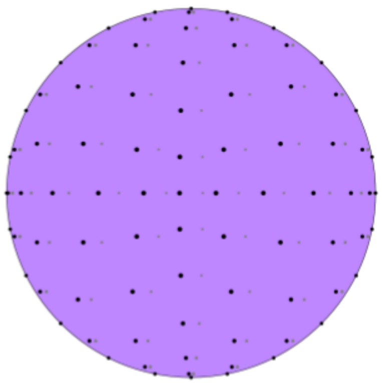
Coding Band Shape
N=3 at Various Rates



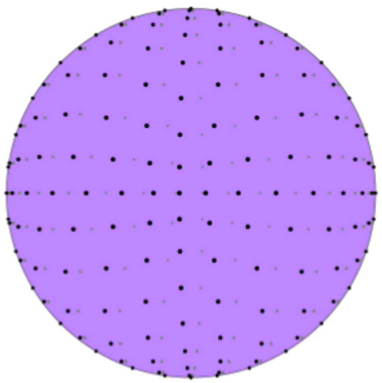
5.25 bits (K=3)



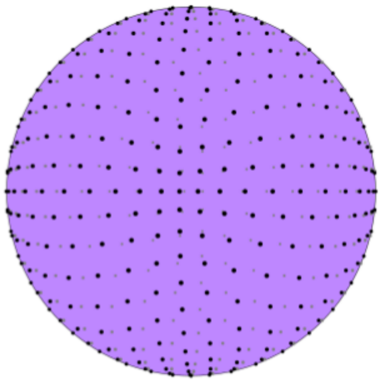
6.04 bits (K=4)



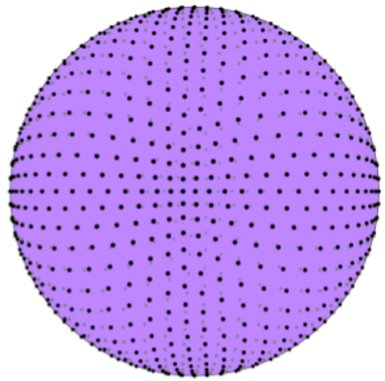
7.19 bits (K=6)



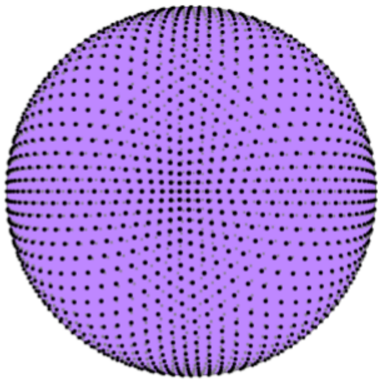
8.01 bits (K=8)



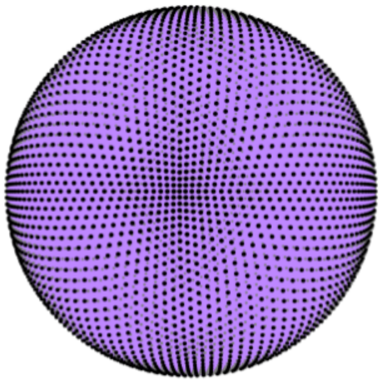
8.92 bits (K=11)



10.00 bits (K=16)



11.05 bits (K=23)



12.00 bits (K=32)

Coding Band Shape
Pyramid Vector Quantization

- PVQ codebook has a fast enumeration algorithm
  - Converts between vector and integer codebook index
- Encoded with flat probability model
  - Range coded but cost is known in advance
- Codebooks larger than 32 bits
  - Split the vector in half and code each half separately

Implicit Psychoacoustics: Bit Allocation

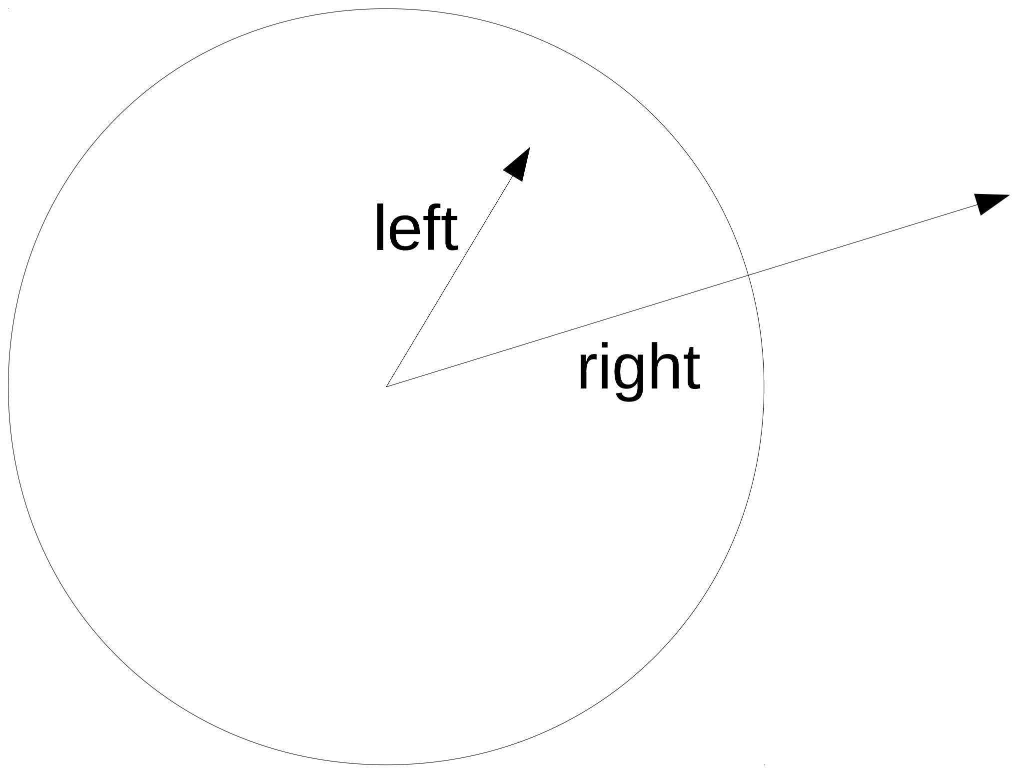
- Sychronized allocator in encoder and decoder
  - Allocates fine energy and PVQ bits for each band
  - Based on shared information (no signaling)
  - Implicit psychoacoustic model
    - Intra-band masking: near-constant per-band SMR
    - Does not model inter-band masking, tone vs noise
- Allocation tuning (signaled)
  - Tilt: balances between LF vs HF bits
  - Boost: Gives more bits to individual bands

CELT Stereo Coupling

- Code separate energy for each channel
  - Prevents cross-talk
- Converts to mid-side after normalization
  - Mid and side coded separately with their relative energy conserved
  - Prevents stereo unmasking
- Intensity stereo
  - Discards side past a certain frequency

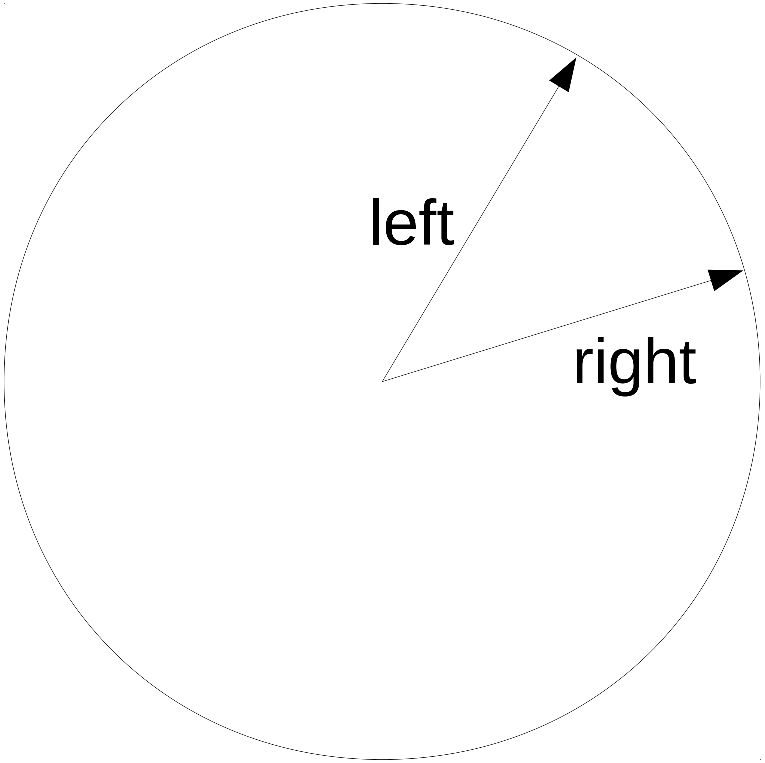
Normalized Mid-Side Stereo

- Input audio



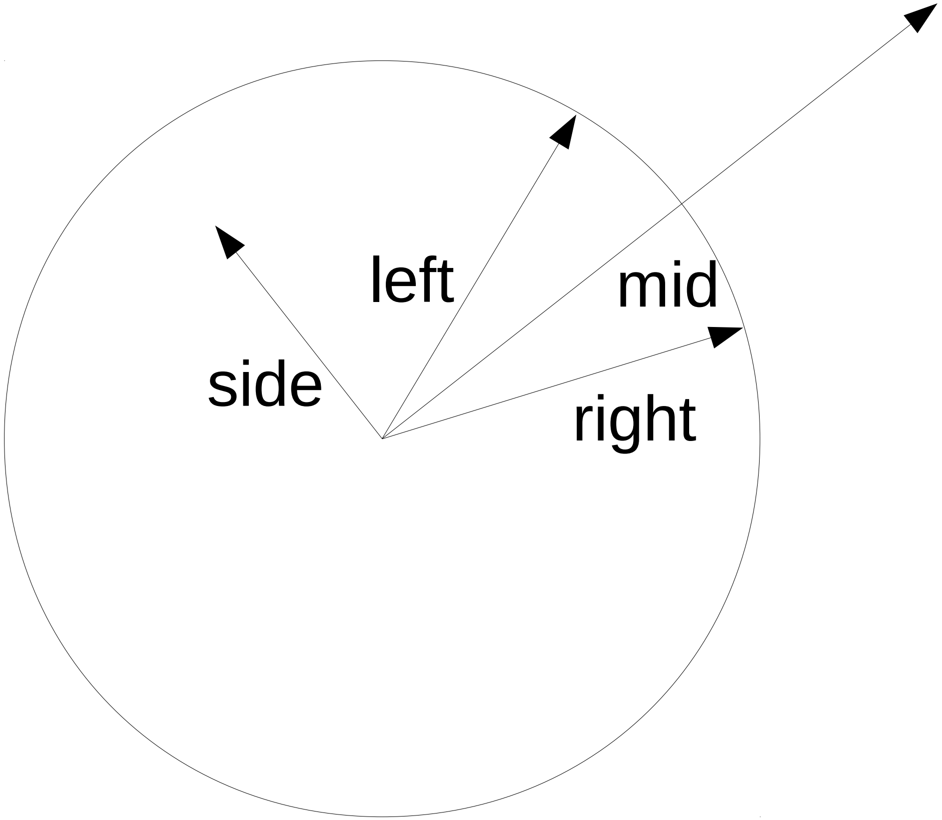
Normalized Mid-Side Stereo

- Channel normalization



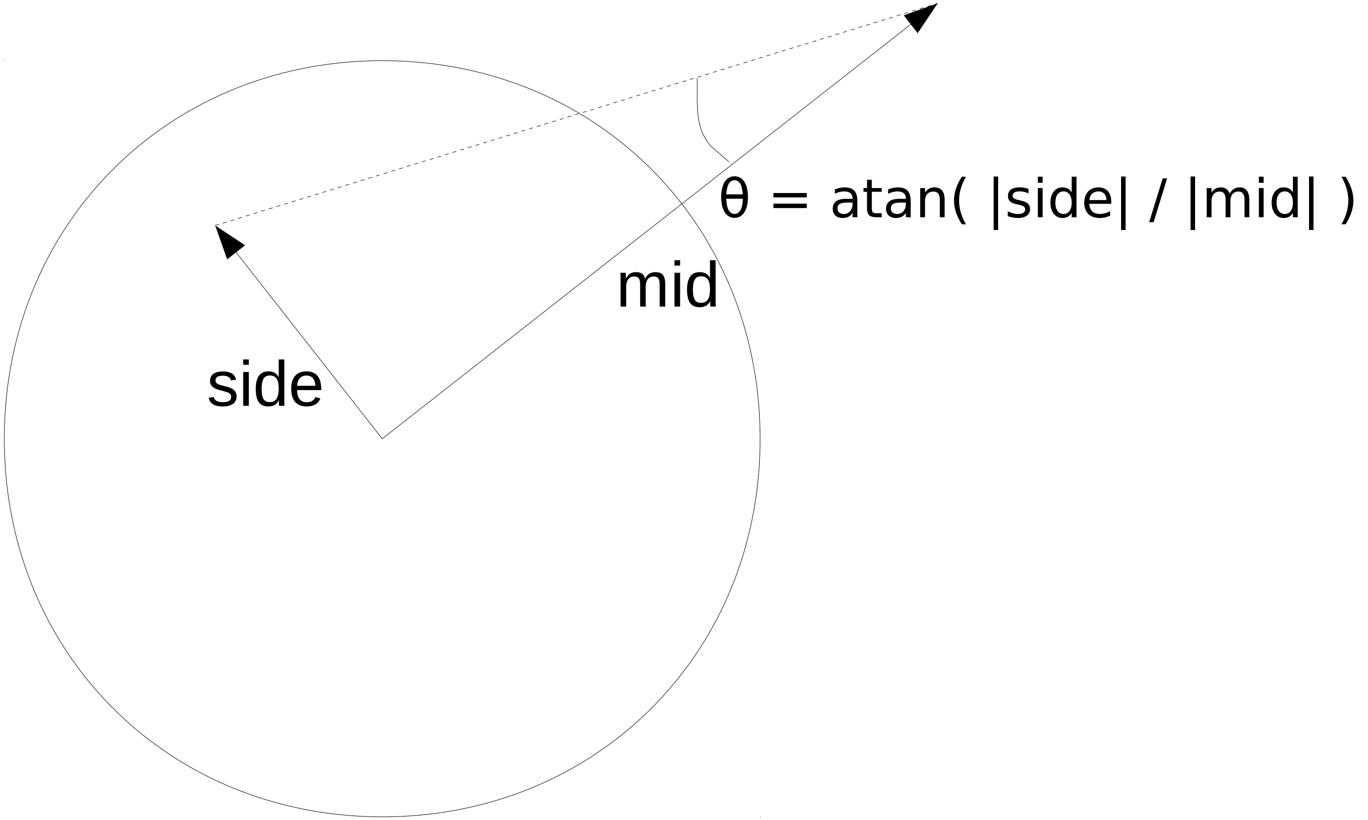
Normalized Mid-Side Stereo

- Mid-side vectors



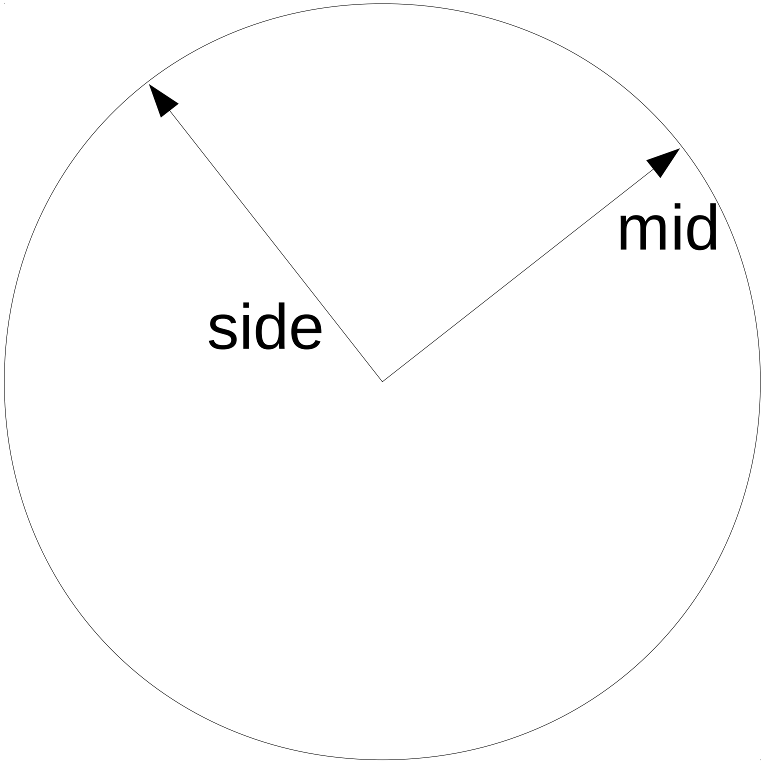
Normalized Mid-Side Stereo

- Mid-side energy ratio



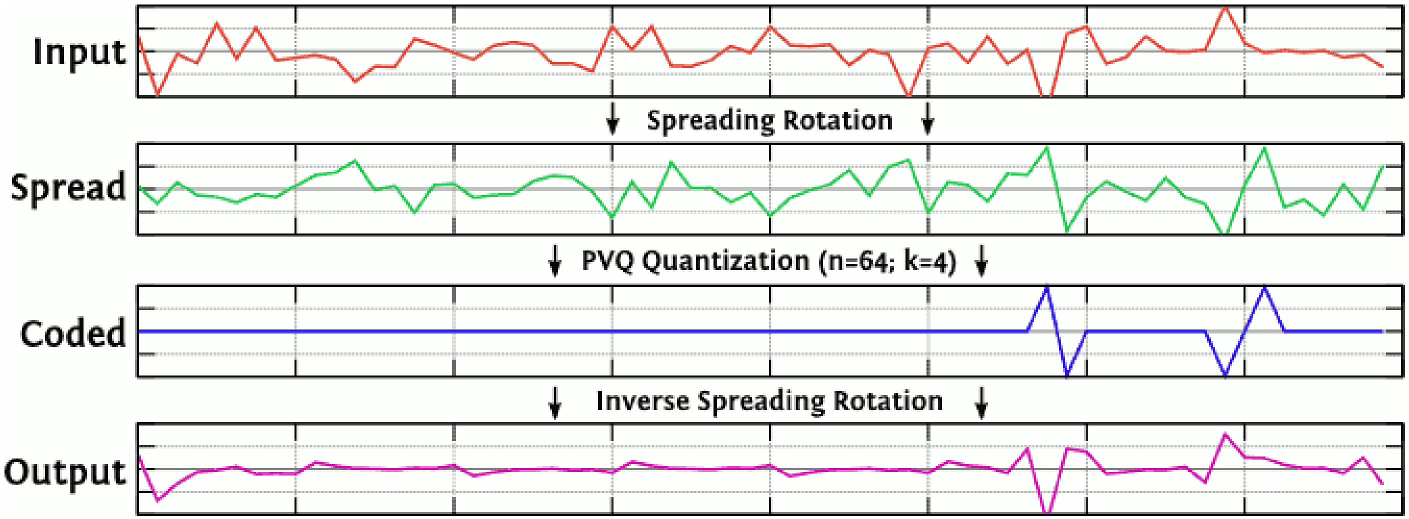
Normalized Mid-Side Stereo

- Normalized mid and side, coded separately



Avoiding Birdie Artifacts

- Small K → sparse spectrum after quantization
  - Produces tonal “tweets” in the HF
- CELT: Use pre-rotation and post-rotation to spread the spectrum
  - Completely automatic (no per-band signaling)



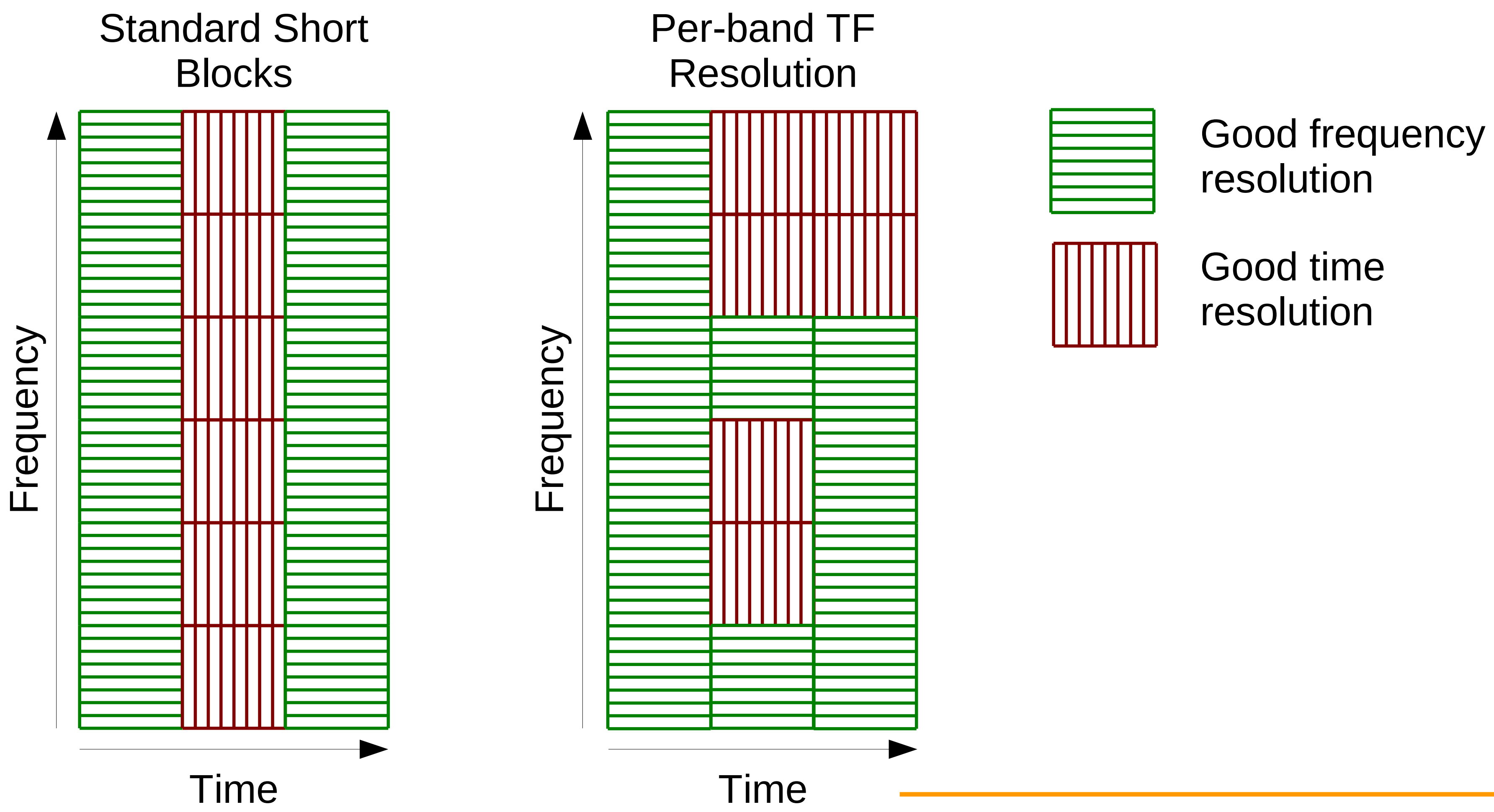
Spectral Folding

- When rate in a band is *too* low, code nothing
  - *Spectral folding*: copy previous coefficients
  - Preserves band energy
  - Gives correct temporal envelope
  - Better than coding an extremely sparse spectrum
- Partial signaling
  - Hard threshold at 3/16 bit per coefficient
  - Encoder can choose to skip additional bands

Transients (avoiding pre-echo)

- Quantization error spreads over whole window
  - Can hear noise before an attack: pre-echo
- Split a frame into smaller MDCT windows
  - Up to 8 “short blocks”
  - Interleave results and code as normal
    - Still code one energy value per band for all MDCTs
- Simultaneous tones and transients
  - Use adaptive time-frequency resolution
  - Per-band Walsh-Hadamard transform

Transients
Time-Frequency Resolution

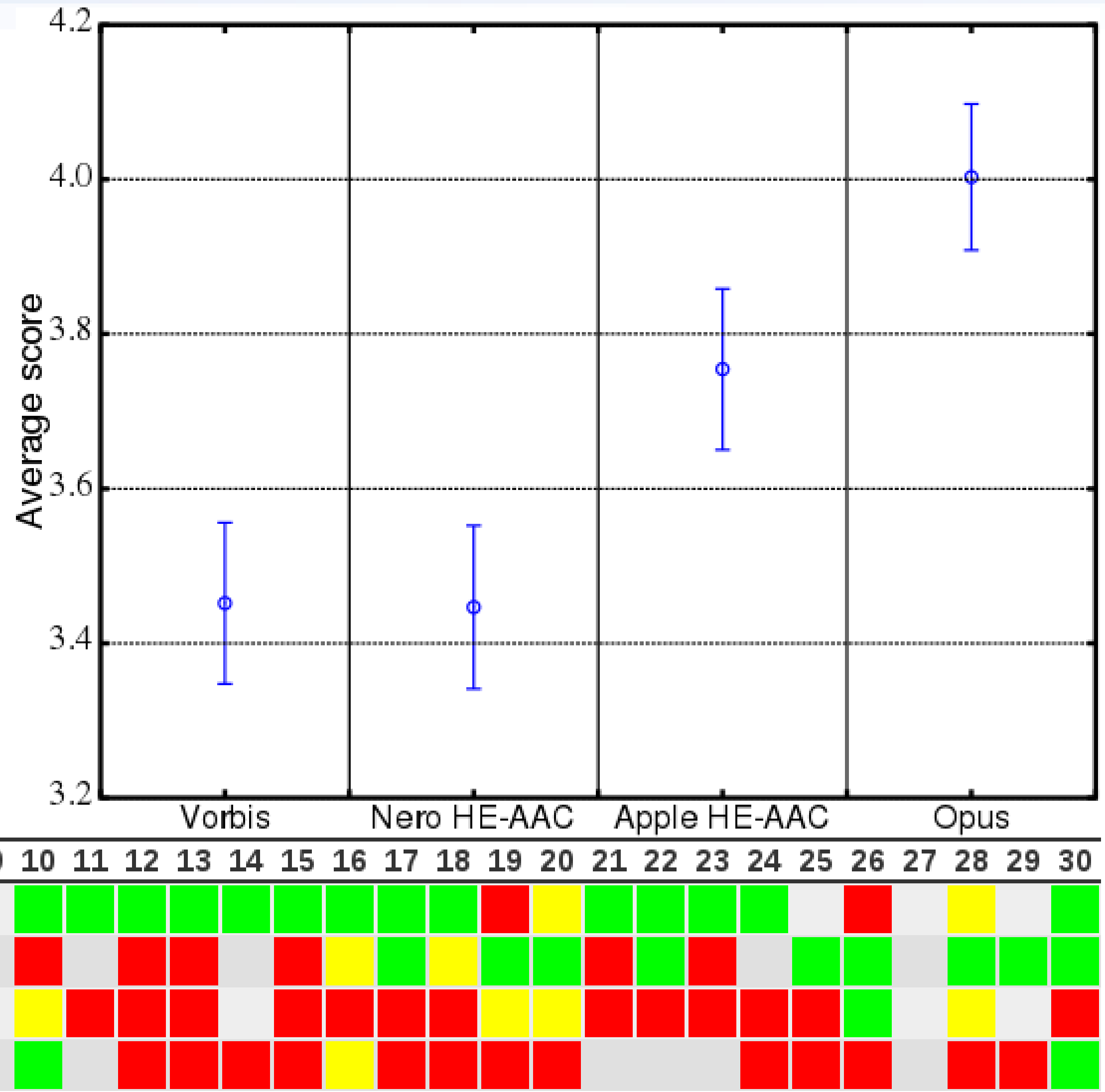


Configuration Switching

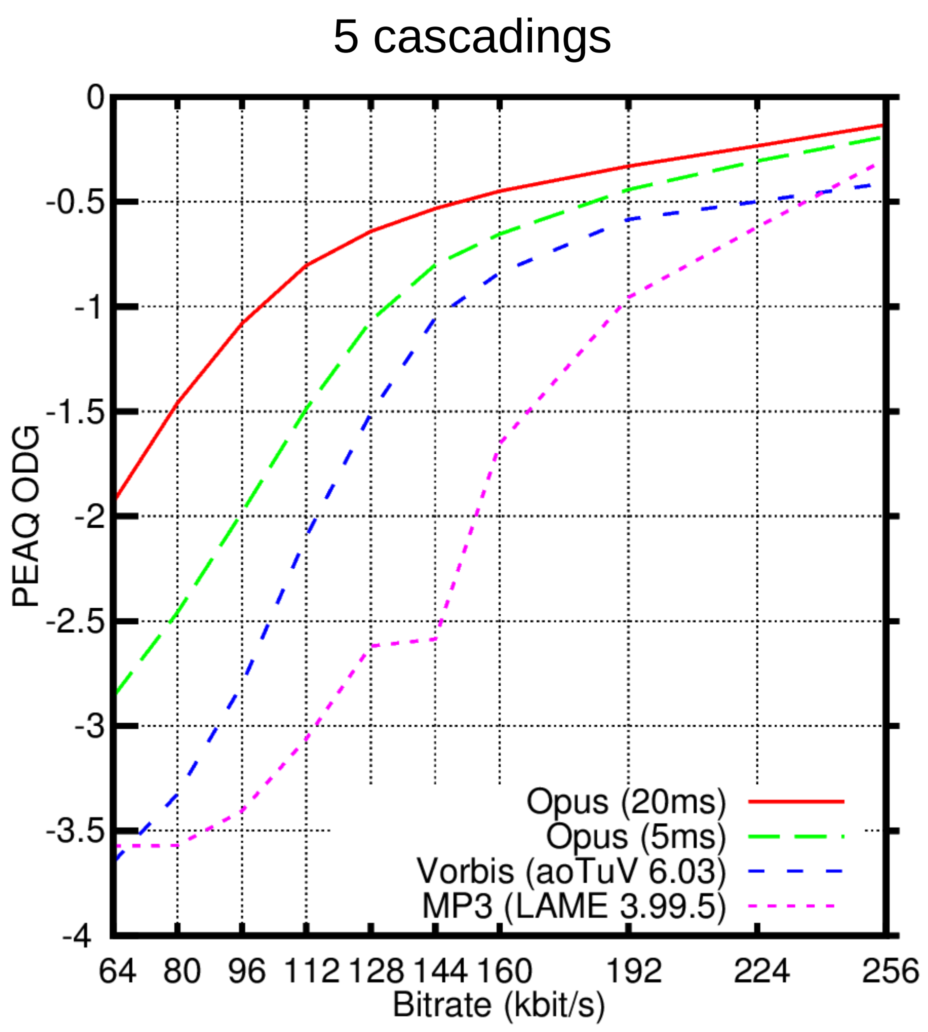
- Mode/bandwidth/framesize/channels changes
- Avoiding *glitches* when we switch
  - All modes can change frame sizes without issue
  - CELT can change audio bandwidth or mono/stereo
  - SILK can change mono/stereo with encoder help
- How about everything else?
  - 5 ms “redundant” CELT frames smooth transition
- Bitrate sweep example: 8 to 64 kb/s

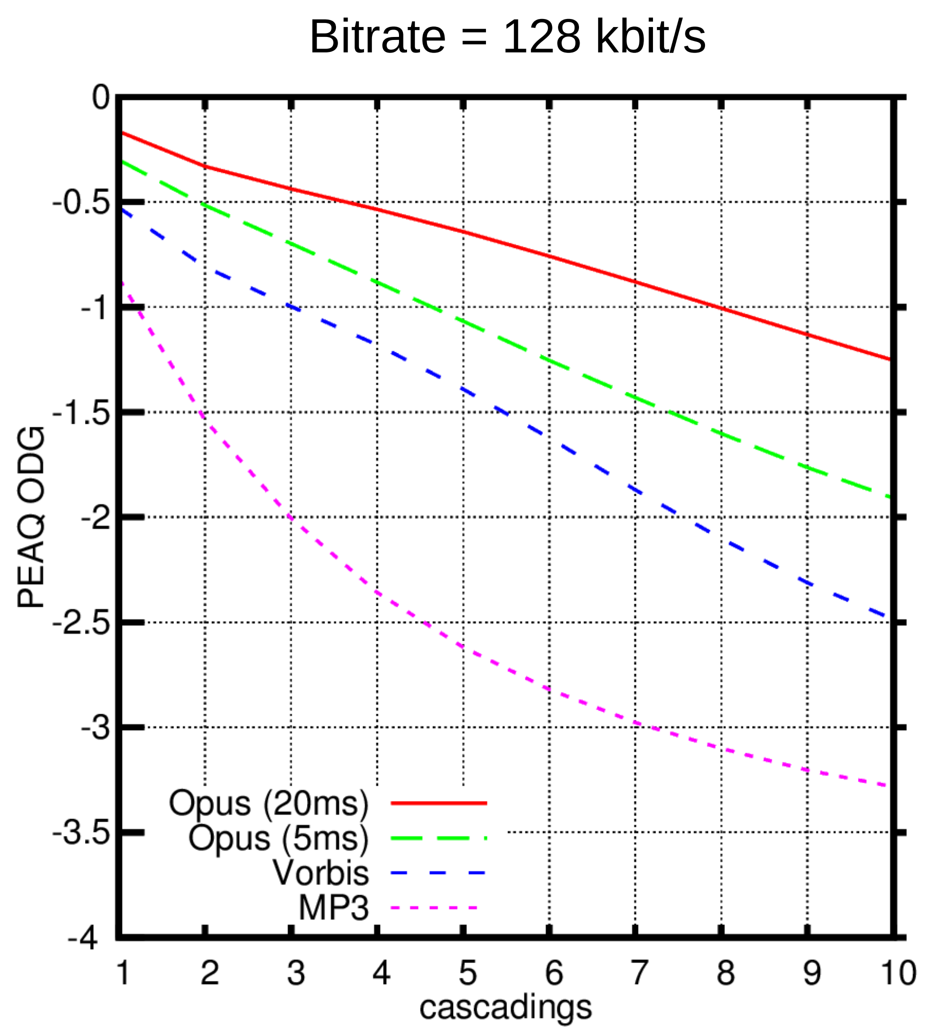
Opus Music Quality

- 64 kb/s stereo music ABC/HR listening test by Hydrogen Audio



Cascading Tests





Future Work

- Upcoming libopus 1.1 release
  - Automatic speech/music detection
  - Better VBR
  - Better surround quality
  - Optimizations
  - https://people.xiph.org/~xiphmont/demo/opus/demo3.shtml
- Specs
  - RTP payload format
  - File format (Ogg, Matroska)

Resources

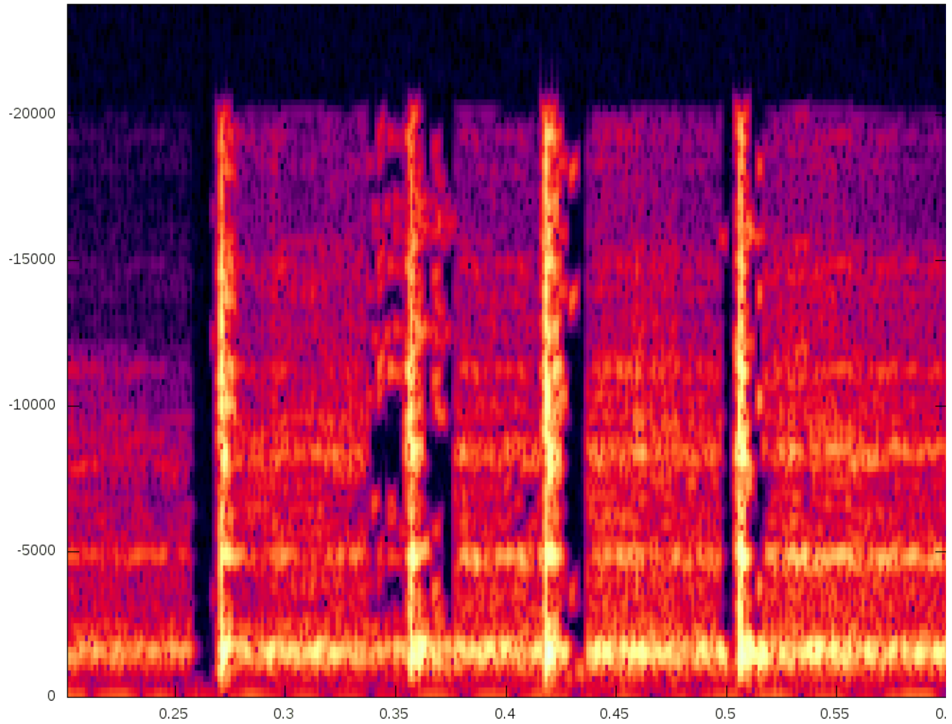
- Website: http://opus-codec.org
- Mailing list: opus@xiph.org
- IRC: #opus on irc.freenode.net
- Git repository: git://git.opus-codec.org/opus.git

Questions?

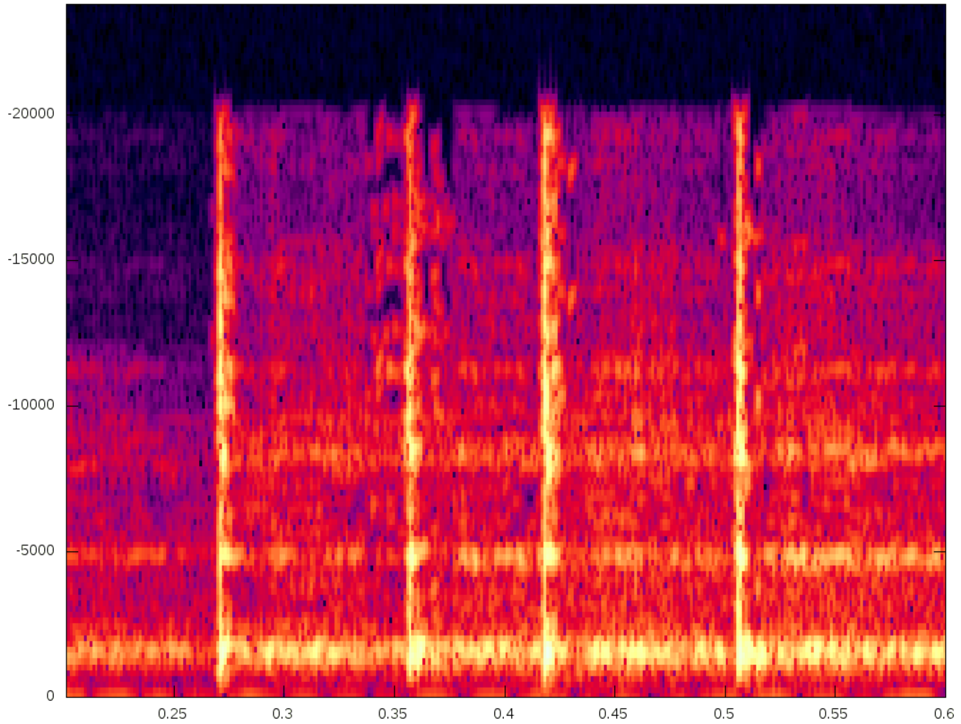
Anti-Collapse

- Pre-echo avoidance can cause collapse
  - Solution: fill holes with noise

No anti-collapse



With anti-collapse



Psychoacoustics
Pitch Prefilter/Postfilter

- Shapes quant. noise (like SILK's LPC filter), but for harmonic signals (like SILK's LTP filter)