Constrained-Energy Lapped Transform (CELT) codec

Jean-Marc Valin
Octasic Inc.

CELT Characteristics

- Speech and music at 32 kHz and above
- 32 kb/s to 128 kb/s (scales to very high quality)
  - Sweet spot: 48 kb/s for speech, 64 kb/s for music
- Tunable delay down to 2 ms (8 ms typical)
- Complexity: 11 + 6 WMOPS (enc + dec)
- State RAM: 0.5 + 8.5 kB
- Scratch RAM: 7 kB

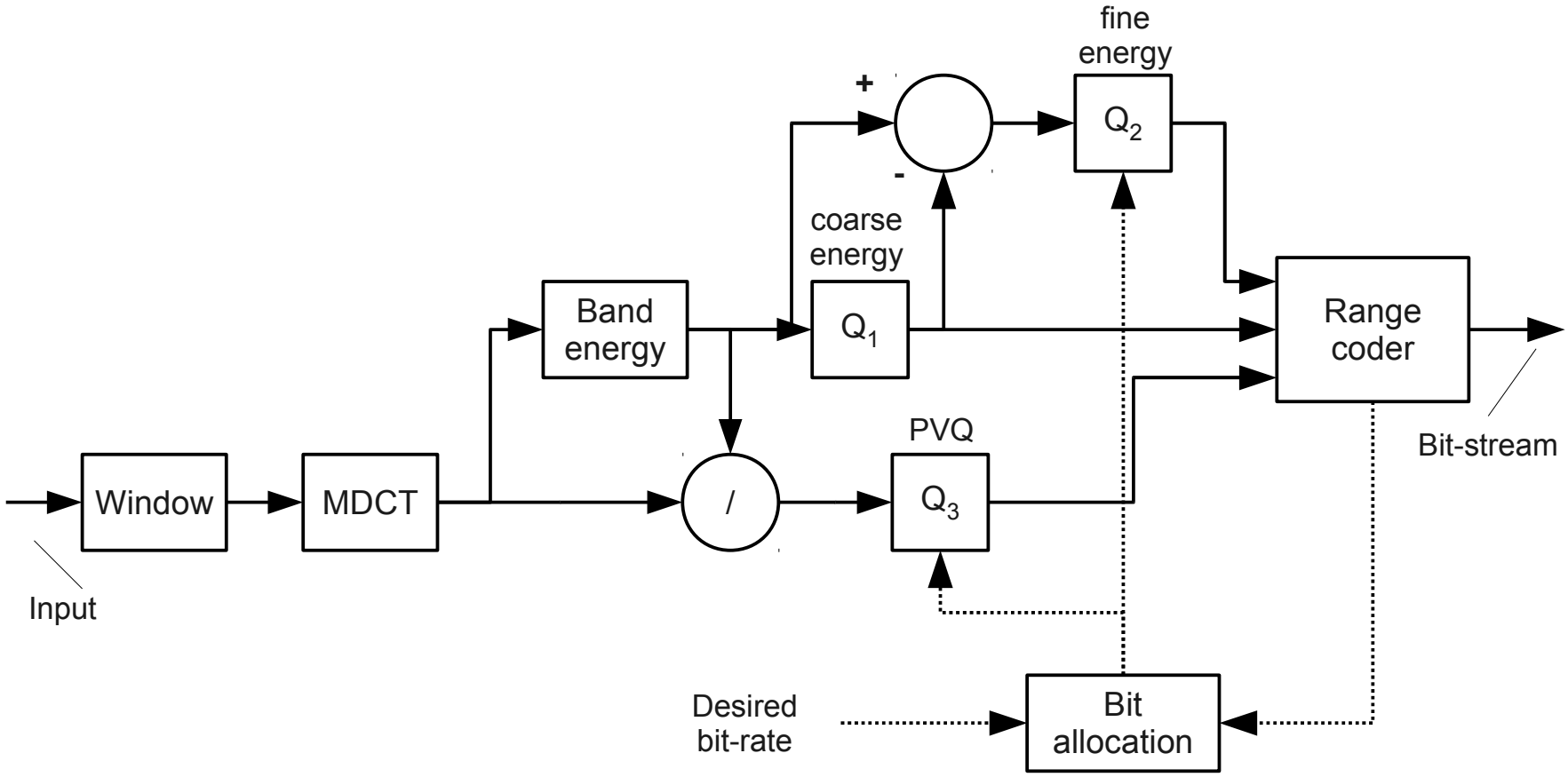
Very Low-Delay Coding

- Benefits
  - Reduces acoustic echo problems (even w/o AEC)
  - Enables new applications
    - Collaborative network music performances
    - Transparent network sound services
  - Better loss robustness (smaller losses)
- Challenges
  - Limited frequency resolution
  - Must minimize overhead in bit-stream

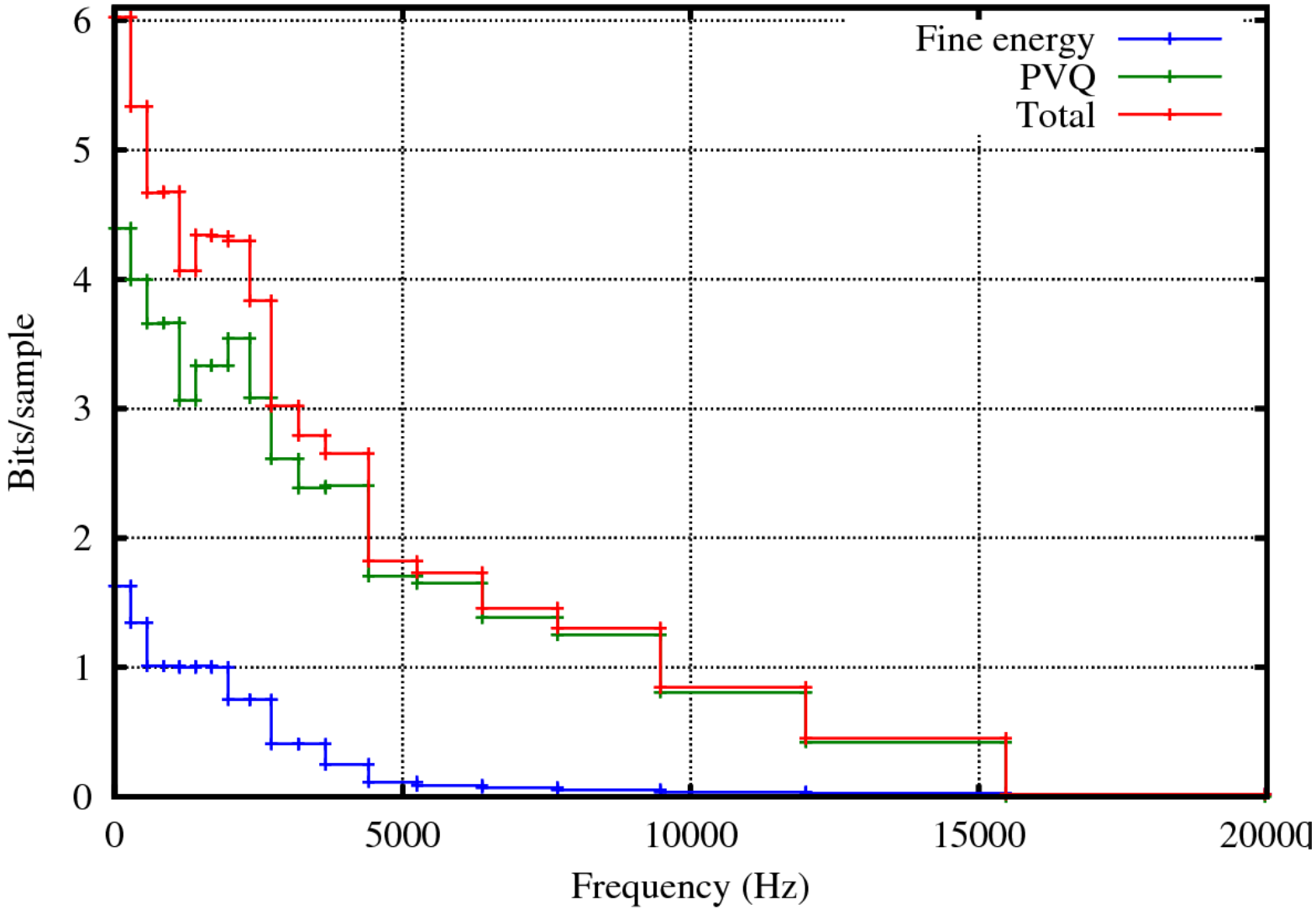
Technology

- Using the Modified Discrete Cosine Transform (MDCT)
- Dividing (roughly) into critical bands
- Explicitly coding the energy in each band with an entropy coder
  - Spectral envelope is preserved
- Using a spherical quantizer for encoding each band

Encoder Block Diagram



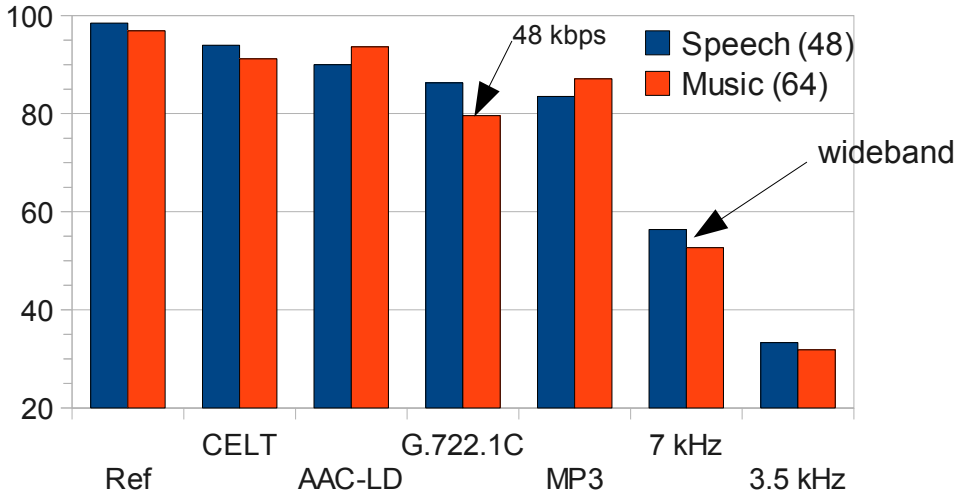
Bit allocation (64 kb/s)



Quality

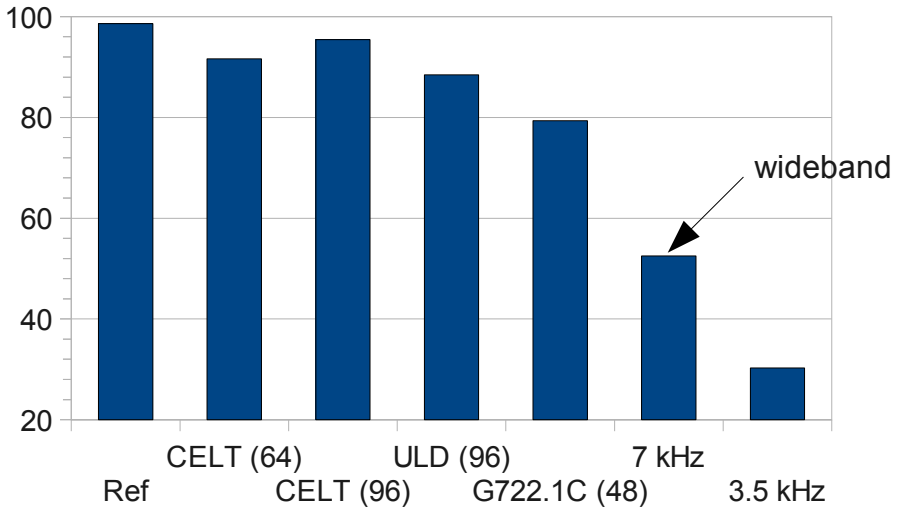
- Internal MUSHRA (ITU-R BS.1534) test

V0.3.2



Delay (ms) **8.7** 34.8 40 >100

V0.5.1



Delay (ms) **8** **4** 5.3 40

Resources

- Website: http://www.celt-codec.org/
  - Source code
  - Papers/presentations
- Mailing list: celt-dev@xiph.org
- IRC: irc.freenode.net #celt