# A Full Bandwidth Audio Codec with Low Complexity and Very Low Delay

Jean-Marc Valin, Octasic Inc.

Timothy B. Terriberry, Xiph.Org Foundation

Gregory Maxwell, Juniper Networks Inc.

EUSIPCO 2009

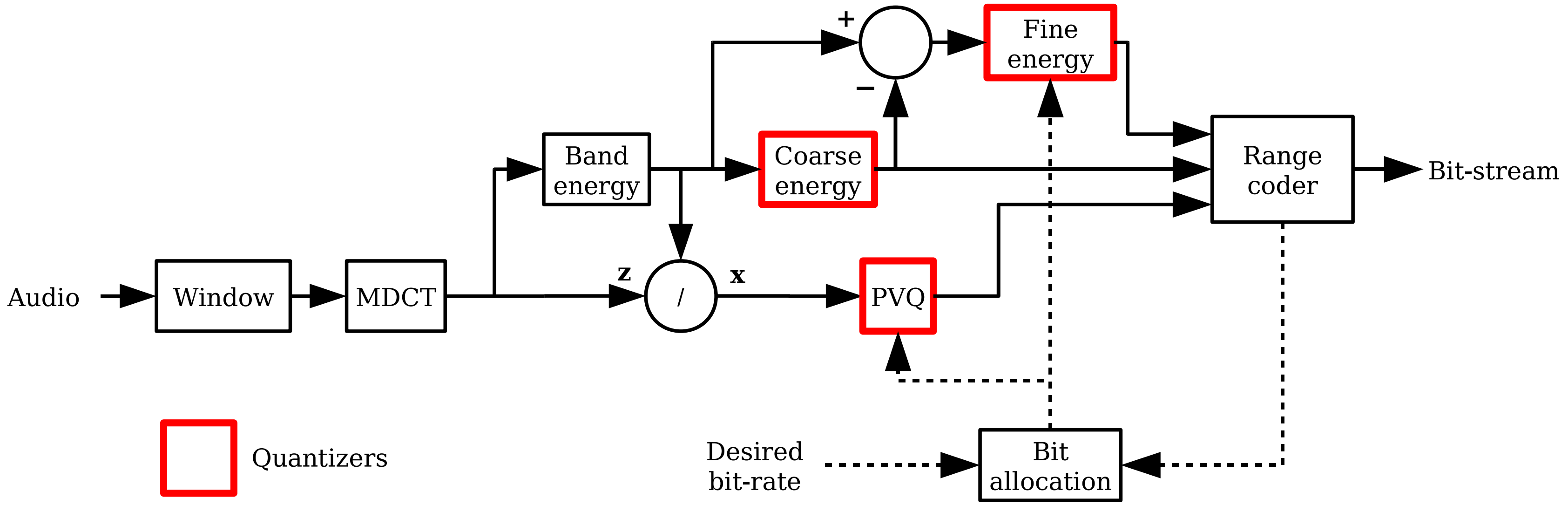
# Introduction

- **Motivations for very low delay**
  - Delay-sensitive applications (e.g. live network music)
  - Reduces perception of acoustic echo
- **Codec characteristics**
  - Speech and music at 48 kHz
  - 5.3 ms frame size (256 samples), 2.7 ms look-ahead
  - 48-128 kb/s per channel (adaptive)
  - Support for frames sizes of 64 – 512 samples

# Overview

- **Constrained-Energy Lapped Transform (CELT)**
- **Basic principles**
    - MDCT spectrum divided into critical bands
    - Band energy explicitly coded, constrained at decoder
    - Spectral “details” coded with spherical codebook
    - Bit allocation based on shared information

# Encoder Block Diagram



Audio

Window

MDCT

Band energy

z

/

x

+

−

Coarse energy

Fine energy

PVQ

Range coder

Bit-stream

Desired bit-rate

Bit allocation

□ Quantizers

# Transform, Bands

- **Modified Discrete Cosine Transform (MDCT)**
  - Low-overlap window
  - Divided into critical bands (except low frequencies)
- **Implications of short frame size**
  - Poor frequency resolution and leakage
  - High cost of “side information”

# Energy Quantization

- **Energy computed for each critical band**
- **Coarse-fine strategy**
  - Coarse energy quantization
    - Scalar quantization with 6 dB fixed resolution
    - Prediction in time (previous frame) and frequency
    - Range-coded with Laplacian probability model
  - Fine energy quantization
    - Variable resolution (based on bit allocation)
    - Not entropy-coded
- **Any error in the energy quantization is <u>not</u> compensated in the later quantization stages**

# PVQ Codebook

- **Quantizing *N*-dimentional vectors of unit norm**
  - *N*-1 degrees of freedom (hyper-sphere)
- **Pyramid Vector Quantizer [Fischer, 1986]**
  - Algebraic codebook (no table stored)
  - Combinations of *K* signed “pulses”
  - Set of vectors *y* such that $|| y ||_{L1} = K$
  - Mapped onto the hyper-sphere: $x = y \,/\, || y ||_{L2}$
- **Fast search and indexing algorithms**
- **Index is range-coded (flat probability)**

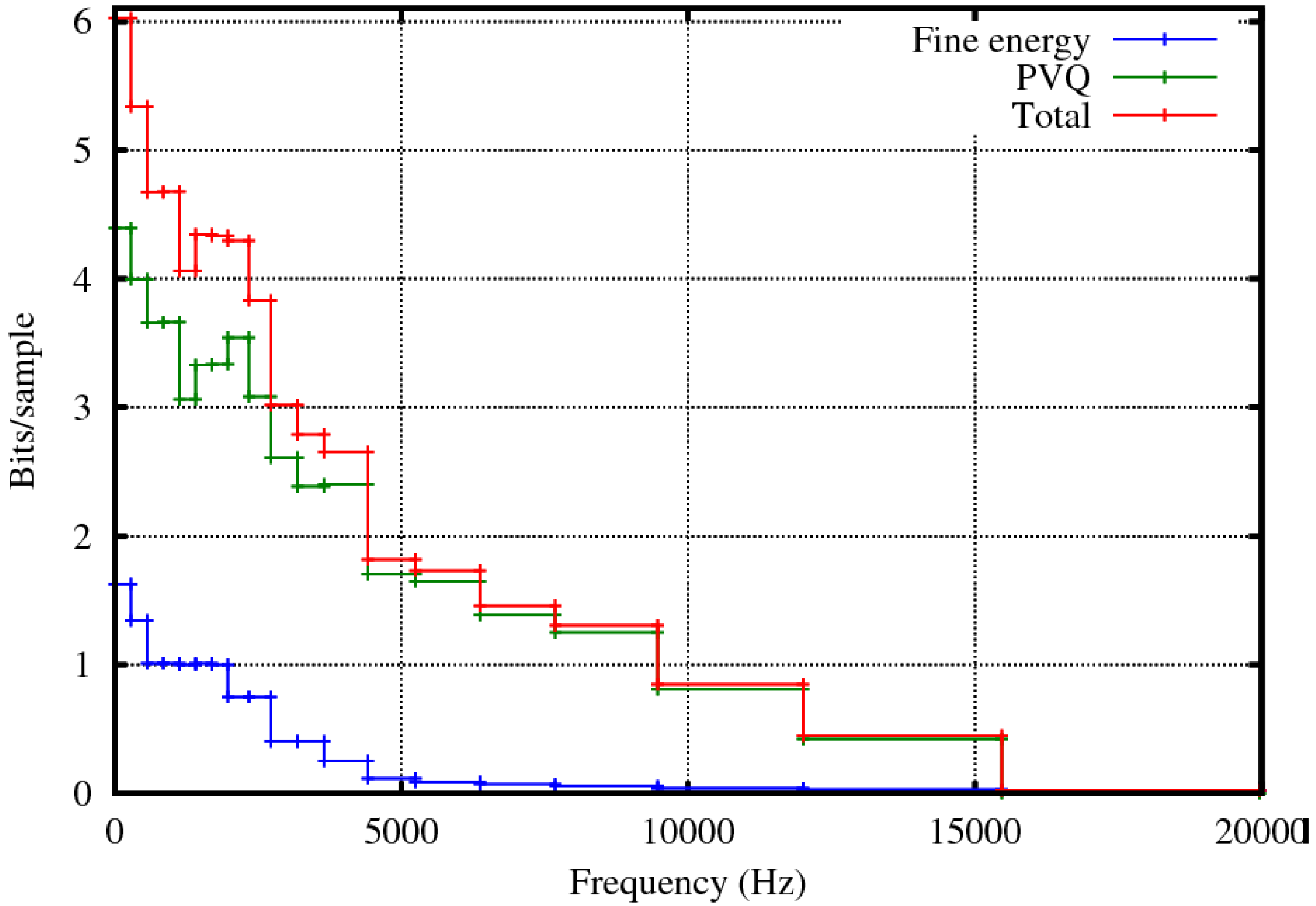
# Perceptual Improvements

- **Pre-echo control**
  - Multiple smaller MDCTs, interleaved spectra
  - Energy computed as if a single MDCT
- **“Birdie” avoidance**
  - Adding an “offset” to PVQ quantization
  - Based on lower part of the spectrum
  - Gain = $N / (N + 6K)$

# Bit Allocation

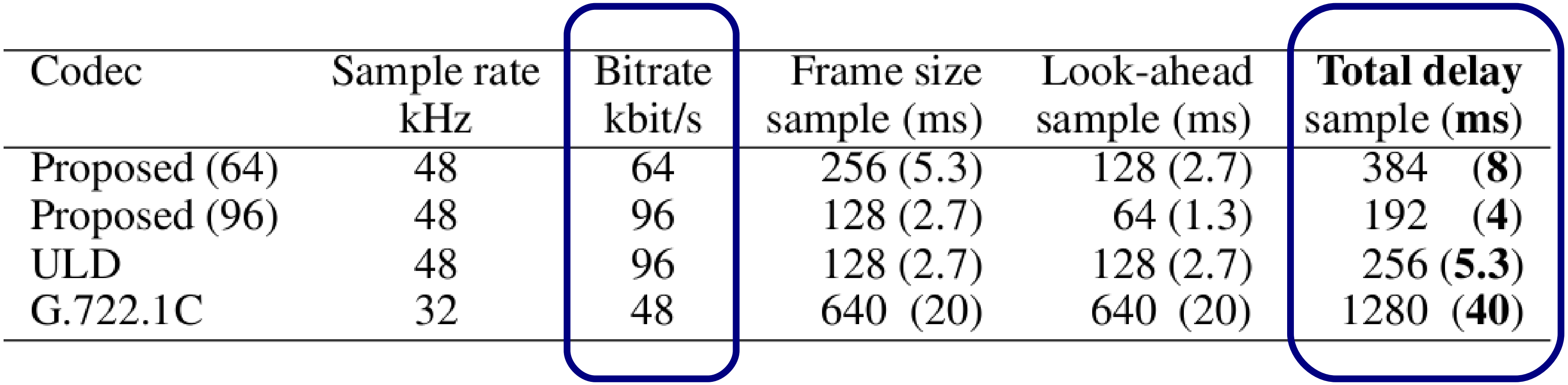
- **Fundamentally a CBR codec (VBR supported)**
- **Synchronized allocator in encoder and decoder**
  - Allocates fine energy bits and PVQ bits
  - Depends only on shared information
    - Number of compressed bytes
    - Number of bits used so far by the range coder
  - Near-constant bits per band in time
    - Models within-band masking with near-constant SMR
    - Does not model inter-band masking, tone vs noise
  - <u>Implicit</u> psycho-acoustic model (not coded)

# Allocation Example (64 kb/s)



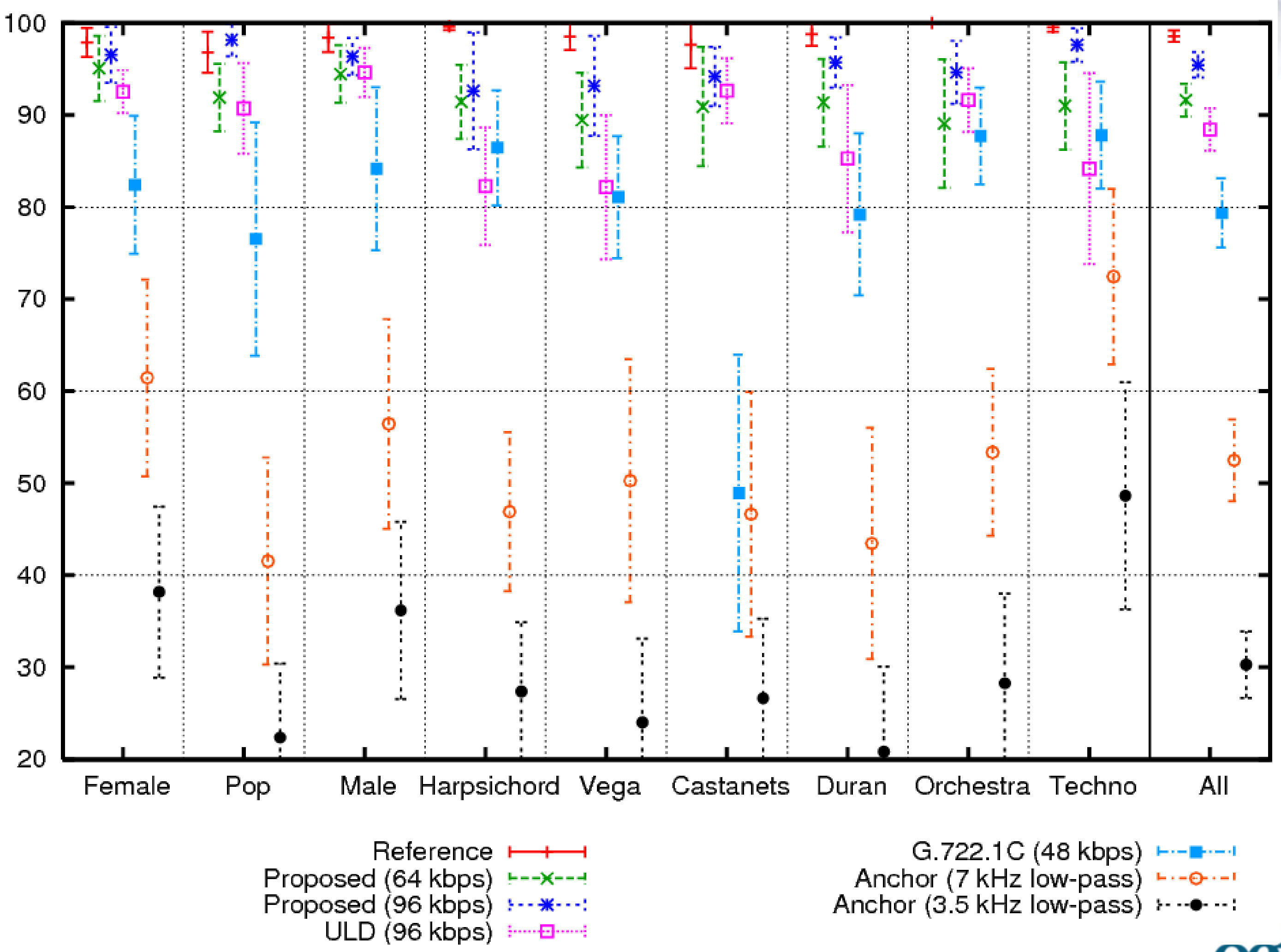
# Evaluation

- **MUSHRA listening tests (10 listeners)**
  - CELT version 0.5.0 (proposed)
  - FhG ULD: warped LPC, pre-filtering
  - G.722.1C: MDCT, scalar quantization, uniform bands



| Codec | Sample rate | Bitrate | Frame size | Look-ahead | **Total delay** |
|---|---|---|---|---|---|
| | kHz | kbit/s | sample (ms) | sample (ms) | sample (**ms**) |
| Proposed (64) | 48 | 64 | 256 (5.3) | 128 (2.7) | 384 (**8**) |
| Proposed (96) | 48 | 96 | 128 (2.7) | 64 (1.3) | 192 (**4**) |
| ULD | 48 | 96 | 128 (2.7) | 128 (2.7) | 256 (**5.3**) |
| G.722.1C | 32 | 48 | 640 (20) | 640 (20) | 1280 (**40**) |

# Results


100
90
80
70
60
50
40
30
20
Female
Pop
Male
Harpsichord
Vega
Castanets
Duran
Orchestra
Techno
All
Reference
Proposed (64 kbps)
Proposed (96 kbps)
ULD (96 kbps)
G.722.1C (48 kbps)
Anchor (7 kHz low-pass)
Anchor (3.5 kHz low-pass)

# Complexity and RAM

- **Complexity (encoder+decoder average)**
  - 17 WMOPS in fixed-point
  - 27 MHz on Intel Core2 (unoptimised floating-point C)
- **State data (per channel)**
  - Encoder: 0.5 kB
  - Decoder: 0.5 kB (+ 4 kB for PLC)
- **Scratch space**
  - Encoder+decoder: ~7 kB

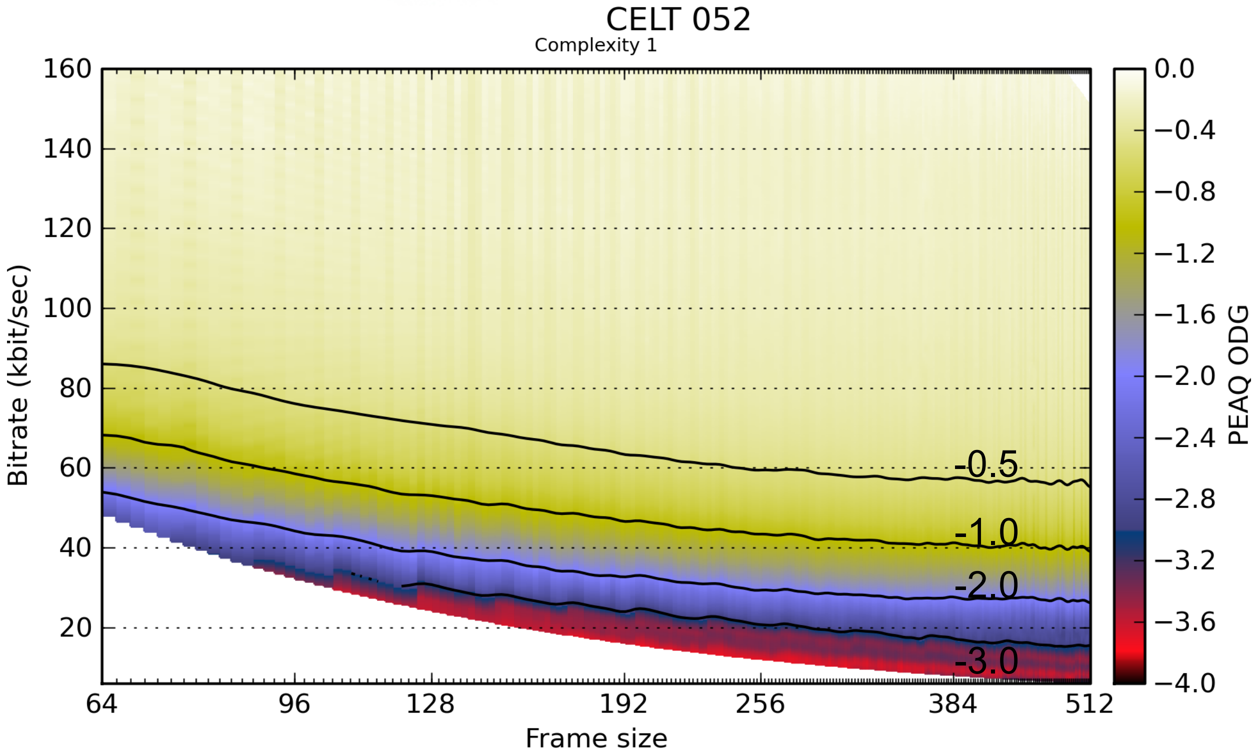
# Conclusion

- **Low-delay coded, explicit energy constraint**
- **Work in progress**
  - Pitch prediction
  - Stereo coupling
- **Submitted to IETF as Internet codec proposal**
- **Resources**
  - Source code: http://www.celt-codec.org
  - Mailing list: celt-dev@xiph.org

# Questions?

Ask me for audio samples after the session

# Other Frame Sizes



Overhead is about 42 bits/frame