Opus, the Swiss Army Knife of Audio codecs

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

Mozilla, Xiph.Org Foundation

What Is the Opus Codec?

- IETF standard under development
- Targets interactive audio over the Internet
- Aims to be royalty-free: BSD code with free license to all patents
- Effort involves: Xiph.Org, Mozilla, Skype, Octasic, Broadcom and more
- Combination of the SILK and CELT codecs

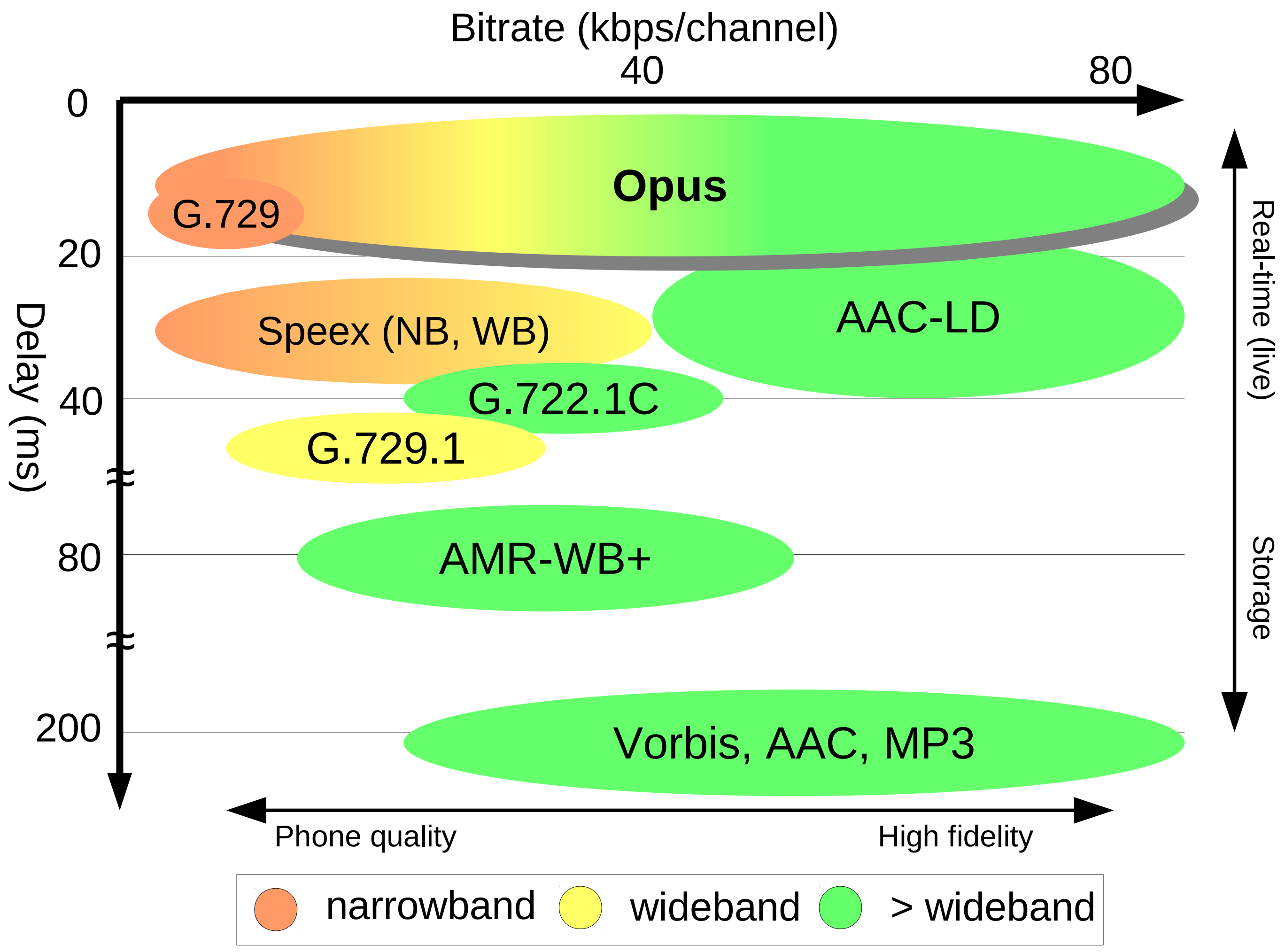
History

- January 2007: SILK codec gets started at Skype
- November 2007: CELT codec gets started
- January 2009: CELT presented at LCA
- March 2009: Skype asks IETF to create a WG to standardize an “Internet wideband audio codec” (SILK)
- February 2010: After heated debate, IETF codec working group created
- July 2010: First prototype of a SILK+CELT hybrid codec
- March 2011: Opus beats HE-AAC and Vorbis in HA test
- Nov 2011: WGLC, last minor bitstream changes

Characteristics

- Sampling rate: 8 – 48 kHz (narrowband-fullband)
- Bitrates: 6 – 510 kb/s
- Frame sizes: 2.5 – 20 ms
- Mono and stereo support
- Speech and music support
- Seamless switching between all of the above
- <u>It just works for everything</u>

Codec Landscape

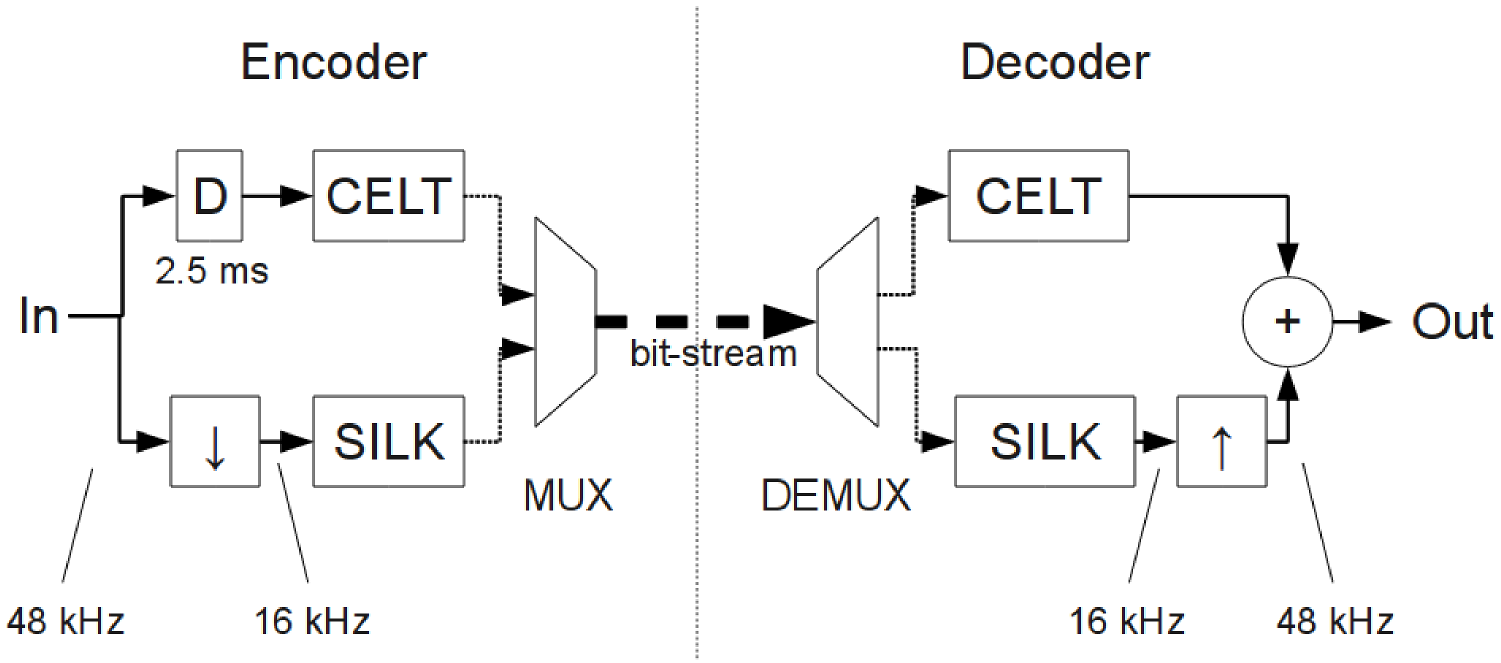


Applications

- VoIP and videoconference
- Music/video streaming and storage
- Remote music jamming
- Wireless speakers/headphones/mic
- Audio books
- Virtualization/sound servers
- Everything except:
  - Lossless (use FLAC)
  - Ultra low bitrate satellite/ham radio (use codec2)

Architecture

- Three operating modes:
  - SILK-only (speech up to wideband)
  - Hybrid (super-wideband/fullband speech
  - CELT-only (music)



Technology (SILK)

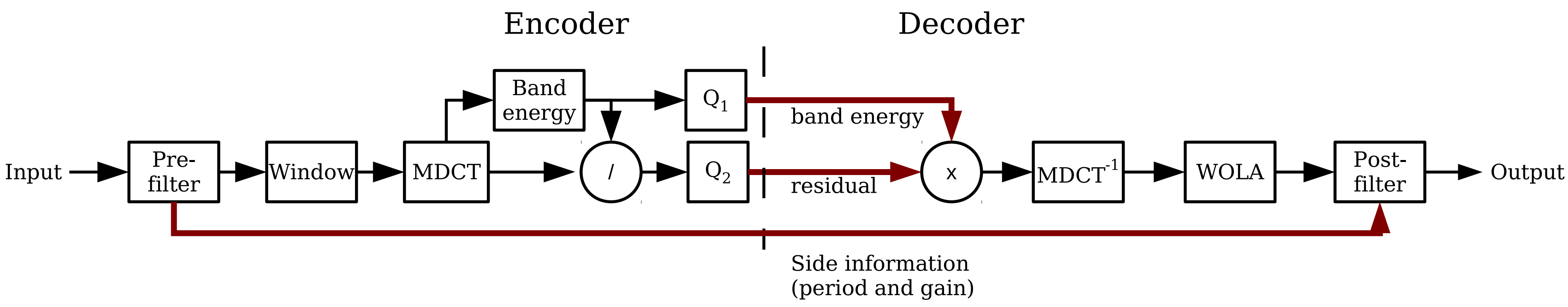
- Speech codec
- Based on linear prediction (LPC)
  - A bit like Speex, but much better
- Very good at coding narrowband and wideband speech
  - Up to ~32 kb/s
- Not very good on music
- Heavily modified to integrate within Opus
  - Not compatible with the original SILK codec

Technology (CELT)

- “Constrained-Energy Lapped Transform”
- Speech+music codec
  - Can work with very low delay
- Uses modified discrete cosine transform (MDCT)
- Most efficient on fullband (48 kHz) audio
  - Useful for 40 kb/s and above
- Not very good on low bit-rate speech

CELT Overview

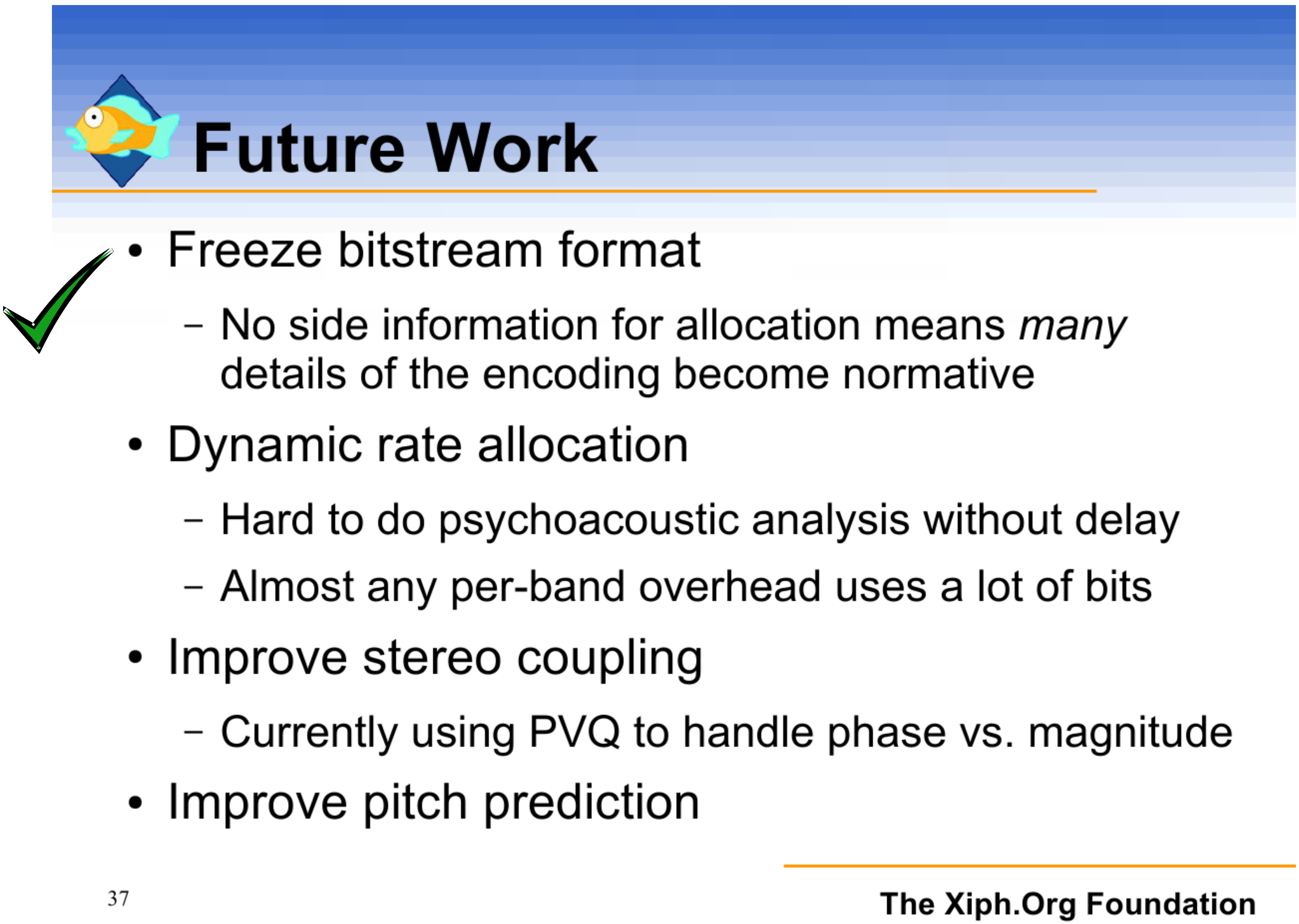
- Transform codec (MDCT)
  - Long blocks up to 20 ms, short blocks of 2.5 ms
- Key is preserving the energy in each Bark band
- Algebraic VQ for band “details”
- Minimal side information



Future Work

- Freeze bitstream format
  - No side information for allocation means *many* details of the encoding become normative
- Dynamic rate allocation
  - Hard to do psychoacoustic analysis without delay
  - Almost any per-band overhead uses a lot of bits
- Improve stereo coupling
  - Currently using PVQ to handle phase vs. magnitude
- Improve pitch prediction

CELT Presentation, LCA 2009

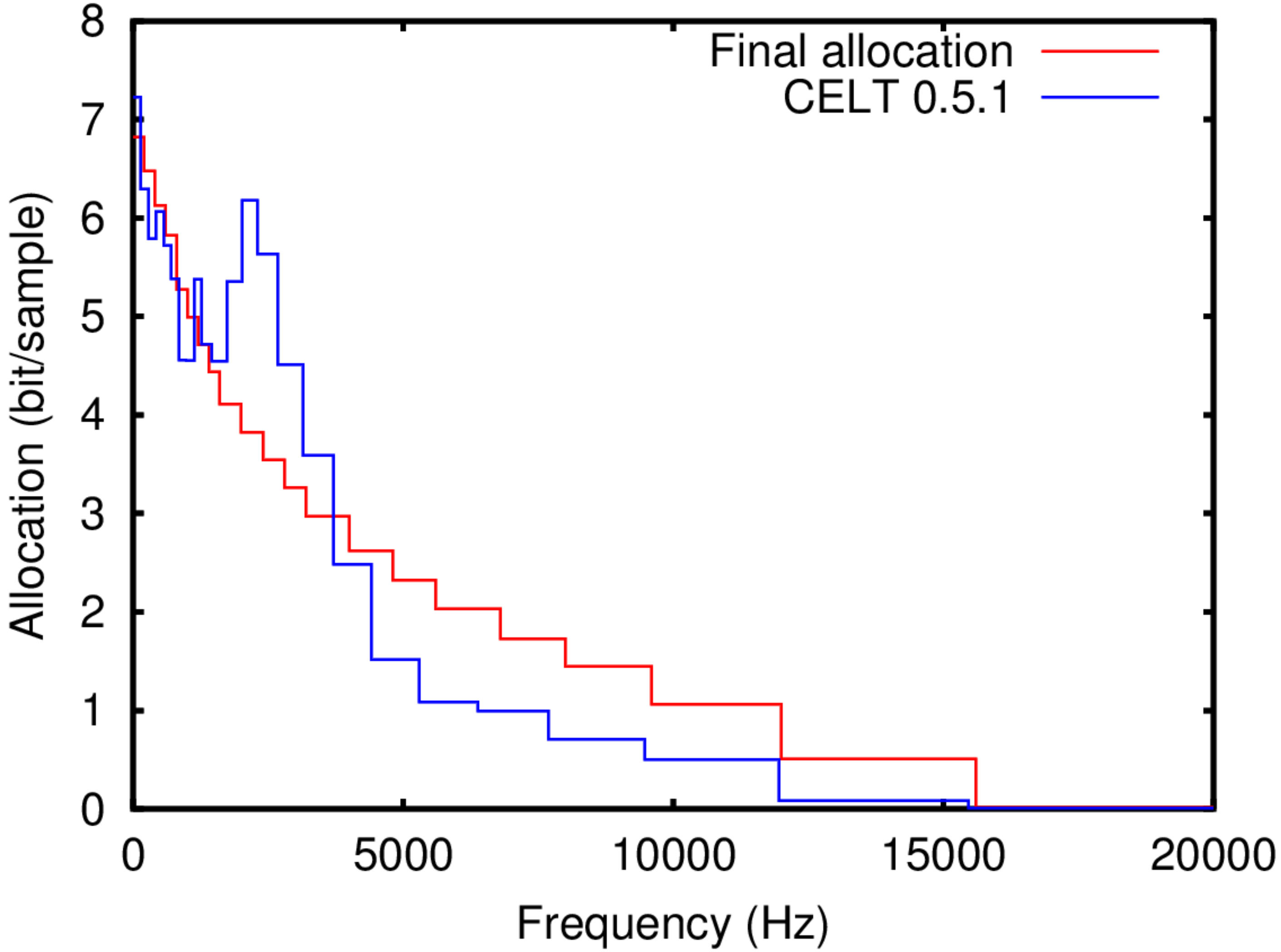


Bitstream Changes

- Many changes required by Opus
  - Changes to band layout
  - 20 ms frames
- Static bit allocation tuning
  - Stop starving the high frequencies

Static Bit Allocation Tuning

- Comparison for 64 kb/s stereo



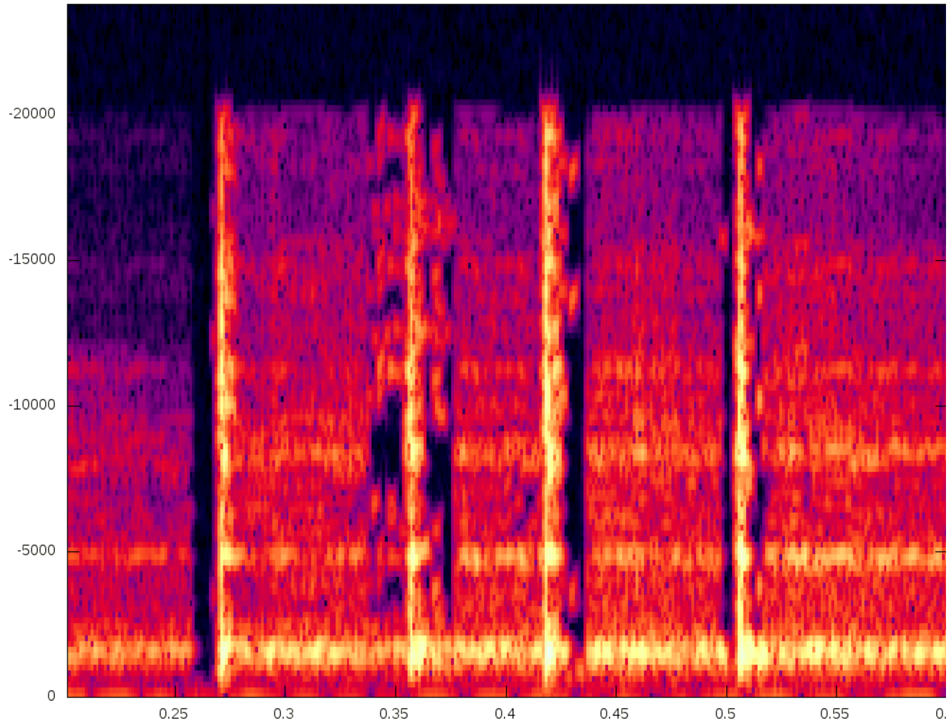
Bitstream Changes

- Many changes required by Opus
  - Changes to band layout
  - 20 ms frames
- Static bit allocation tuning
  - Stop starving the high frequencies
- Anti-collapse

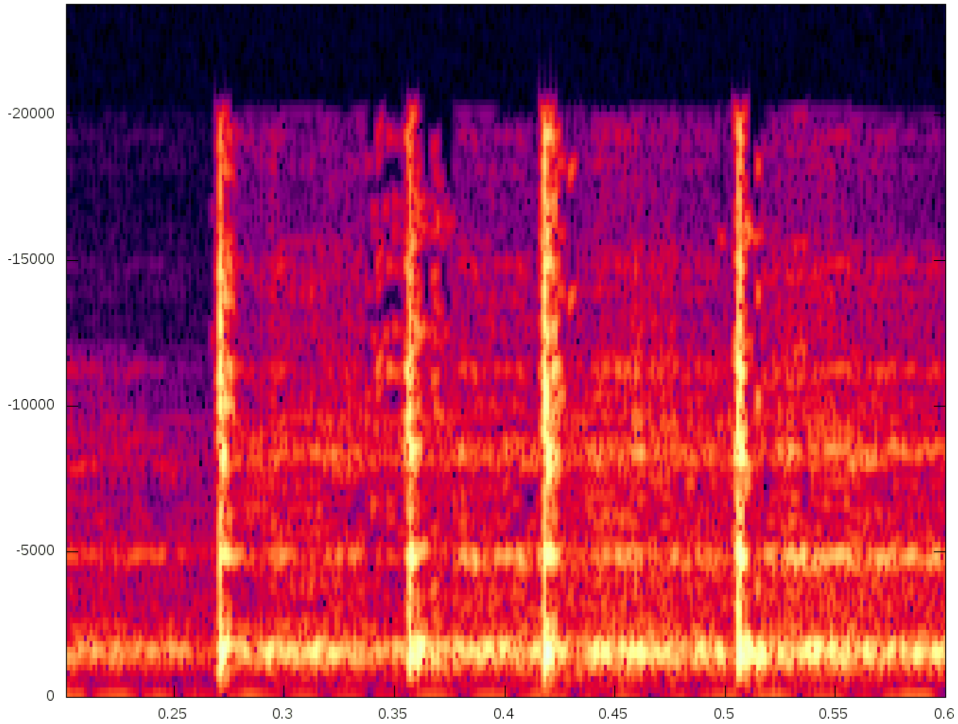
Anti-Collapse

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

No anti-collapse



With anti-collapse



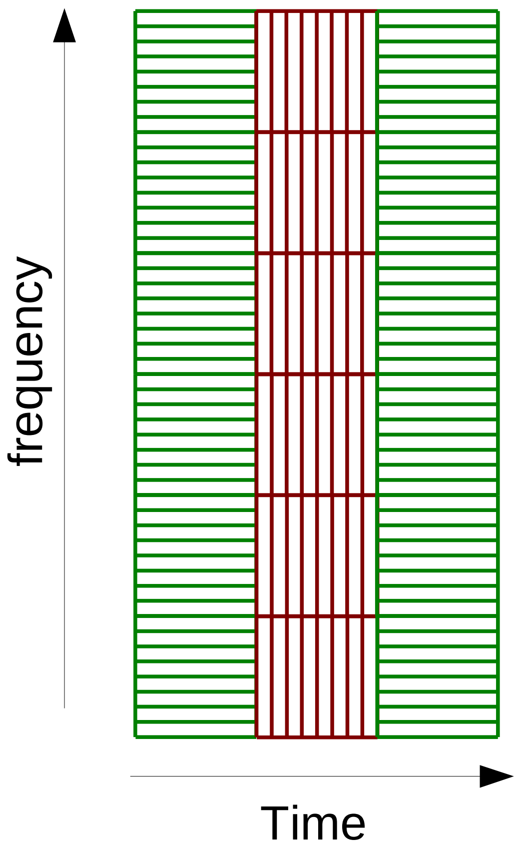
Bitstream Changes

- Many changes required by Opus
  - Changes to band layout
  - 20 ms frames
- Static bit allocation tuning
  - Stop starving the high frequencies
- Anti-collapse
- Per-band time-frequency modifications
  - Long vs short blocks on a per-band basis

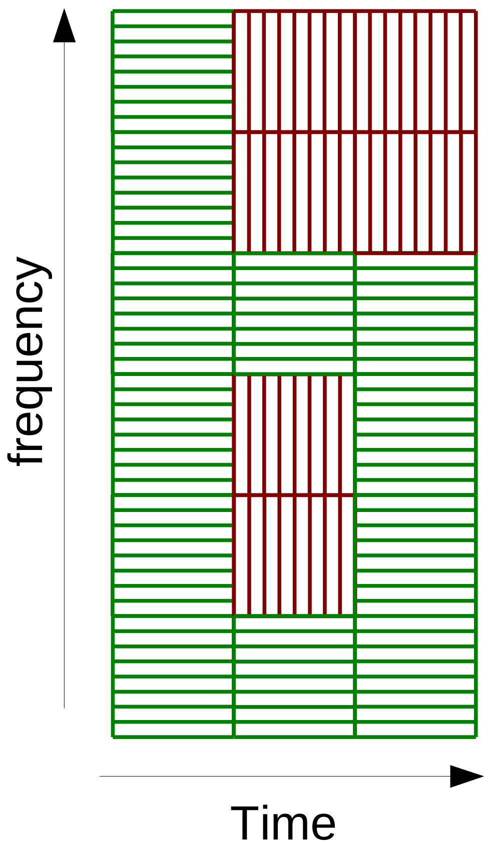
Time-Frequency Resolution

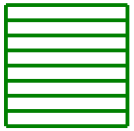
- Tones and transients **can** happen simultaneously

Standard short blocks



per-band TF resolution





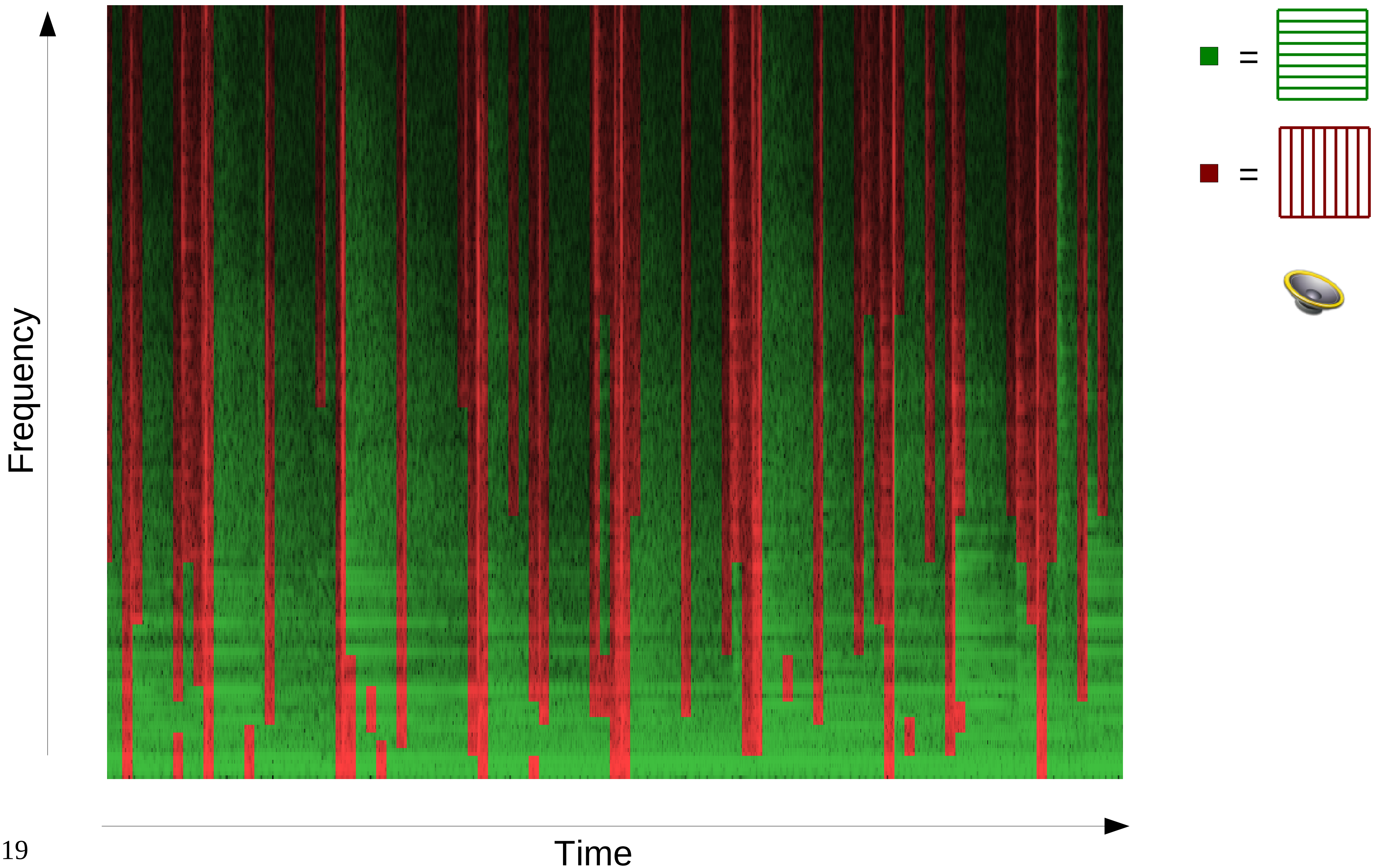
Good frequency resolution

Good time resolution

$\Delta T * \Delta f \geq$ constant

(also known as Heisenberg's uncertainty principle)

Time-Frequency Resolution Example



CELT Presentation, LCA 2009

Future Work

- Freeze bitstream format
  - No side information for allocation means *many* details of the encoding become normative
- Dynamic rate allocation
  - Hard to do psychoacoustic analysis without delay
  - Almost any per-band overhead uses a lot of bits
- Improve stereo coupling
  - Currently using PVQ to handle phase vs. magnitude
- Improve pitch prediction

37

The Xiph.Org Foundation

CELT Presentation, LCA 2009

Future Work

- Freeze bitstream format
  - No side information for allocation means *many* details of the encoding become normative
- Dynamic rate allocation
  - Hard to do psychoacoustic analysis without delay
  - Almost any per-band overhead uses a lot of bits
- Improve stereo coupling
  - Currently using PVQ to handle phase vs. magnitude
- Improve pitch prediction

Dynamic Allocation

- CELT still has mostly static allocation
  - Part of the bit-stream, tuned since 2009
- Now two ways to deviate from static allocation
  - Allocation tilt
    - Controls HF vs LF allocation trade-off
  - Band boost
    - Gives more bits to a band in particular
    - WIP: Use for leakage compensation

CELT Presentation, LCA 2009

Future Work

- Freeze bitstream format
  - No side information for allocation means *many* details of the encoding become normative
- Dynamic rate allocation
  - Hard to do psychoacoustic analysis without delay
  - Almost any per-band overhead uses a lot of bits
- Improve stereo coupling
  - Currently using PVQ to handle phase vs. magnitude
- Improve pitch prediction

CELT Presentation, LCA 2009

Future Work

- Freeze bitstream format
  - No side information for allocation means *many* details of the encoding become normative
- Dynamic rate allocation
  - Hard to do psychoacoustic analysis without delay
  - Almost any per-band overhead uses a lot of bits
- Improve stereo coupling
  - Currently using PVQ to handle phase vs. magnitude
- Improve pitch prediction

Stereo Coupling

- Three modes: Dual, mid-side, intensity
- Mid-side in the normalized domain
  - Safe, cannot cause cross-talk or bad artefacts
  - Based on preservation of the mid/side magnitude ratio
  - $\theta = \mathrm{atan}\frac{\|S\|_{L2}}{\|M\|_{L2}}$
  - Bit allocation depends on theta
- Same mechanism now used to split bands with more bits than largest codebook

CELT Presentation, LCA 2009

Future Work

- Freeze bitstream format
  - No side information for allocation means *many* details of the encoding become normative
- Dynamic rate allocation
  - Hard to do psychoacoustic analysis without delay
  - Almost any per-band overhead uses a lot of bits
- Improve stereo coupling
  - Currently using PVQ to handle phase vs. magnitude
- Improve pitch prediction

CELT Presentation, LCA 2009

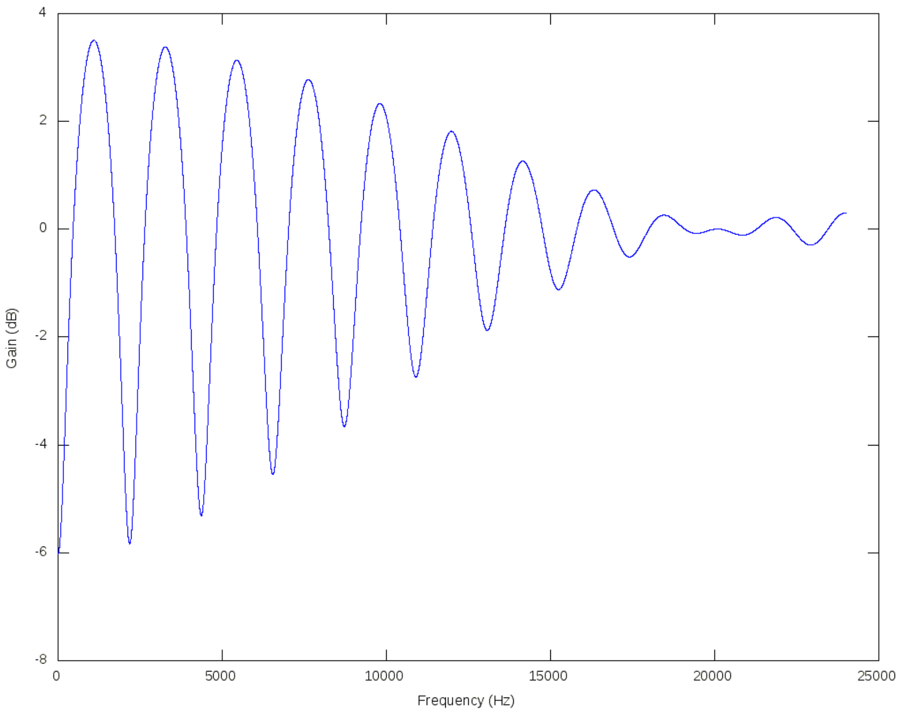
Future Work

- Freeze bitstream format
  - No side information for allocation means *many* details of the encoding become normative
- Dynamic rate allocation
  - Hard to do psychoacoustic analysis without delay
  - Almost any per-band overhead uses a lot of bits
- Improve stereo coupling
  - Currently using PVQ to handle phase vs. magnitude
- Improve pitch prediction

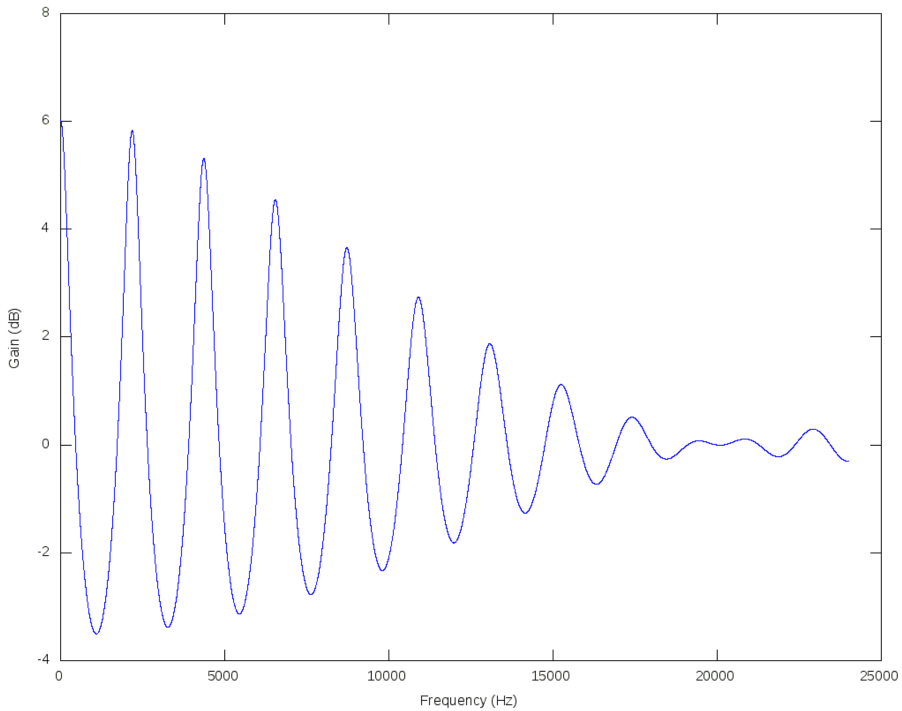
Pitch prefilter/postfilter

- Contributed by Broadcom
- Shapes noise for highly harmonic content

Prefilter



Postfilter



Subjective Testing

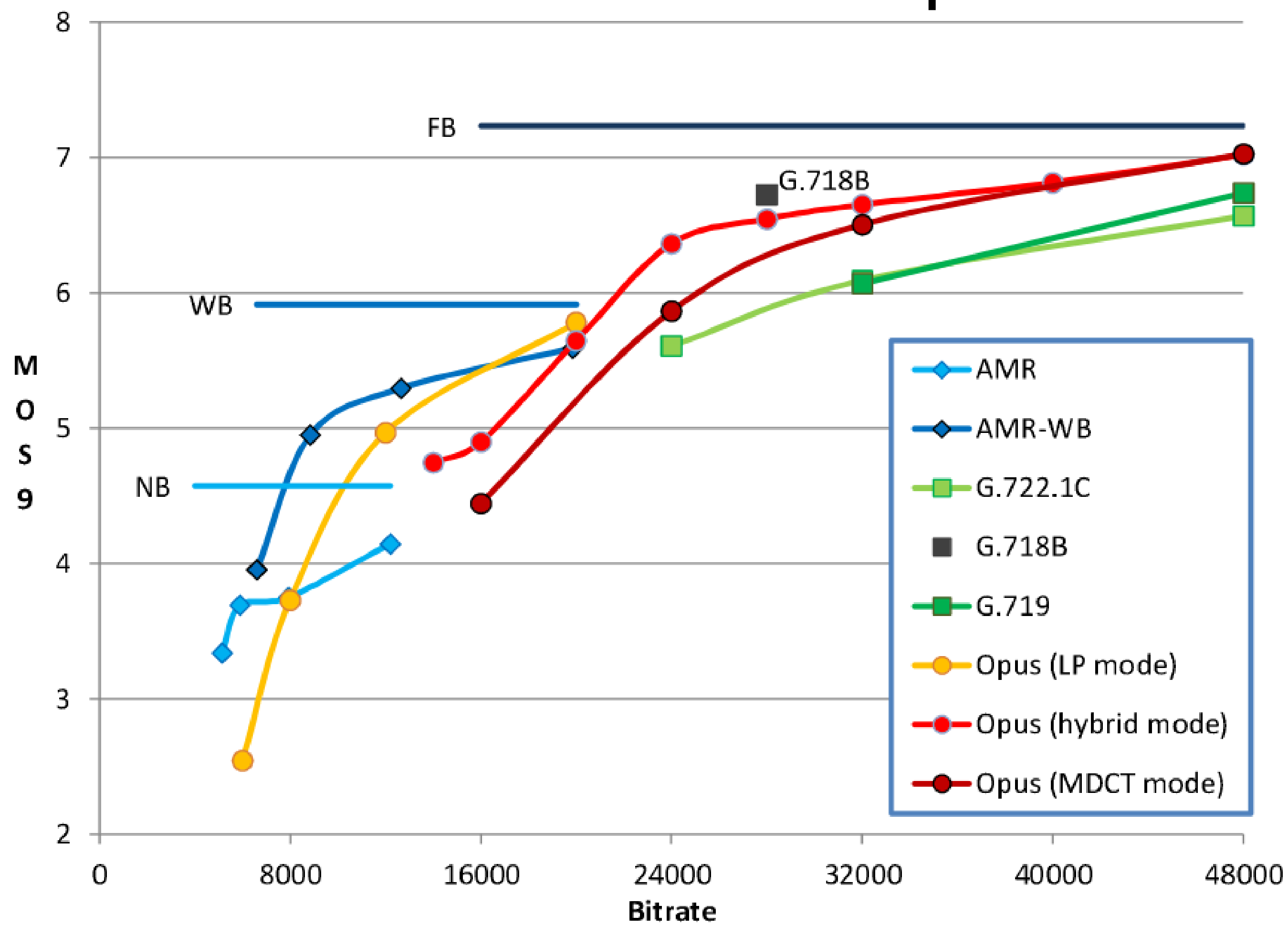
- Comparison with other codecs
  - AMR-NB, AMR-WB, Speex, Vorbis, AAC, ...
- Many tests performed during development
- Tests on the final version:
  - Google (7 MUSHRA tests)
  - Nokia (2 MOS tests)
  - HydrogenAudio (ABC/HR test)

Google Tests

- Narrowband tests (English+Mandarin)
  - Opus clearly better than Speex and iLBC
  - Opus better than AMR-NB at 12 kb/s
- Wideband/fullband tests (English+Mandarin)
  - Opus clearly better than Speex, G.722.1, G.719
  - Opus better than AMR-WB at 20 kb/s
- Opus clearly better than MP3 on music, inconclusive with AAC
- No transcoding issues with AMR-NB/AMR-WB

Nokia (clean+noisy speech)

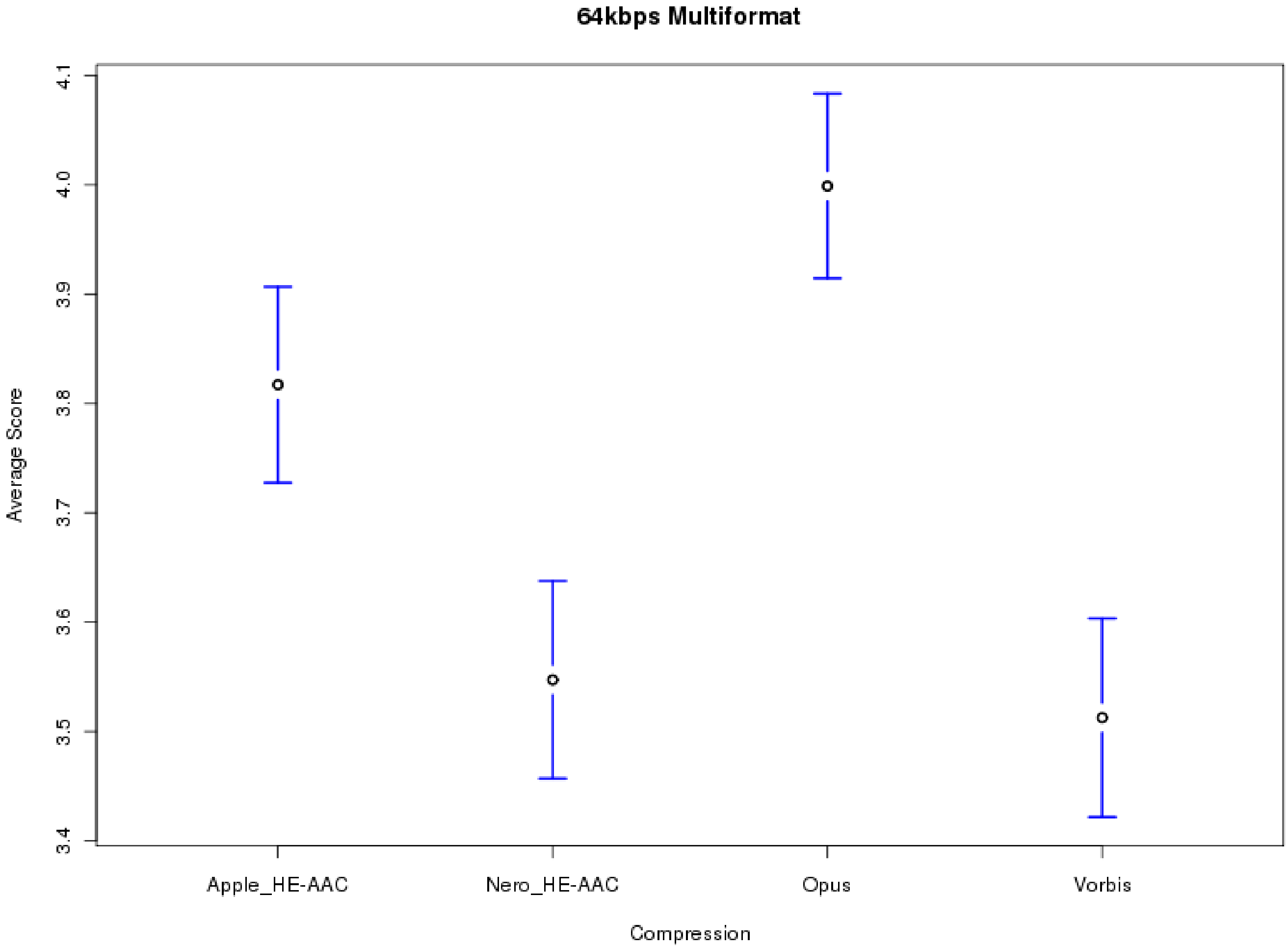
- Narrowband – fullband MOS speech test



Anssi Rämö, Henri Toukomaa, "Voice Quality Characterization of IETF Opus Codec", *Proc. Interspeech*, 2011.

HydrogenAudio

- 64 kb/s stereo music ABC/HR test



Demo

- Music at 64 kb/s
  - u-law (G.711)
  - Opus
  - Reference
  - MP3
- Bitrate sweep
  - 8 kb/s to 64 kb/s

Current Development

- Tools
  - Ogg encoder/decoder
  - Matroska encoder/decoder
  - Firefox support
- Quality improvements
  - Better tuning of encoder decisions
  - Improved unconstrained VBR
  - Automatic speech/music detection

Coming Up

- IETF process
  - IETF Last call
  - RFC
- Industry adoption
  - RTCWeb
  - Browser support (streaming/HTML5)
  - Skype
  - World domination

Resources

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

Questions?