Native Resolution Detection of Video Sequences

Ioannis Katsavounidis
Sr. Research Scientist
Encoding Technologies
Netflix
ikatsavounidis@netflix.com

A presentation to the VQEG meeting, Glasgow, Sep. 14-18, 2015

Image & Video Resolution

- Video resolution has long been an important aspect of video quality
- Video codecs have been largely driven by the desire to increase and eventually associated with certain video frame resolutions; MPEG1 – VCD (352x240), MPEG – DVD (720x480), H.264 – HDTV (1920x1080), H.265 – UHD (3840x2160)
- Digital still cameras have played the "resolution game" for over 15 years, in an attempt to convince customers to upgrade their earlier cameras
- Content providers can differentiate service based on resolution, but
  - Users can't always tell the difference between a high-quality DVD and a poor-quality Blu-Ray disk
  - Many companies sell up-sampling "technology"

Is 4K truly 4K?

- Netflix licenses and receives pristine video content from various studios around the world, in a variety of formats (lately – mostly IMF/MXF)
- Encoding technologies is the group responsible for
  - Inspecting and assuring quality of video assets delivered to Netflix by content providers
  - Encoding each asset according to a multitude of "recipes"; each recipe consists of a choice of video/audio and timed text codec alongside all parameters that determine an encoder's video/audio quality and associated bitrate
  - We need to establish whether a certain content delivered to us is true 4K or it has been upsampled from a lower resolution

Problem statement

- Given a video sequence in a certain (“apparent”) resolution and color format, how to determine the minimum spatial resolution that this content existed in its past life (“native video resolution”)?
- Many reasons why native resolution is not equal to apparent resolution:
  - Original video camera sensor had lower resolution than apparent resolution
  - Special effects overlaid at different resolution
  - Post-production software can’t handle apparent resolution
  - User error

An example – “Birds in Cage”

“Birds in Cage”– Test 1

“Birds in Cage”– Test 2

“Birds in Cage”– Test 3

“Birds in Cage”– Test 4

“Birds in Cage” – Test 5

Our Approach

- No-reference image/video quality problem
- Frequency-domain analysis
- Obtain spectral signatures of known image downsample/upsample pairs to detect evidence of upsampling in given video spectra
- Isolate active video lines from input video sequence (horizontal/vertical black bar detection)
- Split video sequence in scenes, since upsampling is typically applied on selected scenes (mostly due to special effects/post-production)

System diagram

Video Sequence, uncompressed YUV

Horizontal/Vertical black bar detection

FFT, DC-removal, AC-energy normalization, log(), accumulation

Scene change detection

Average normalized spectra in log domain (1 per scene)

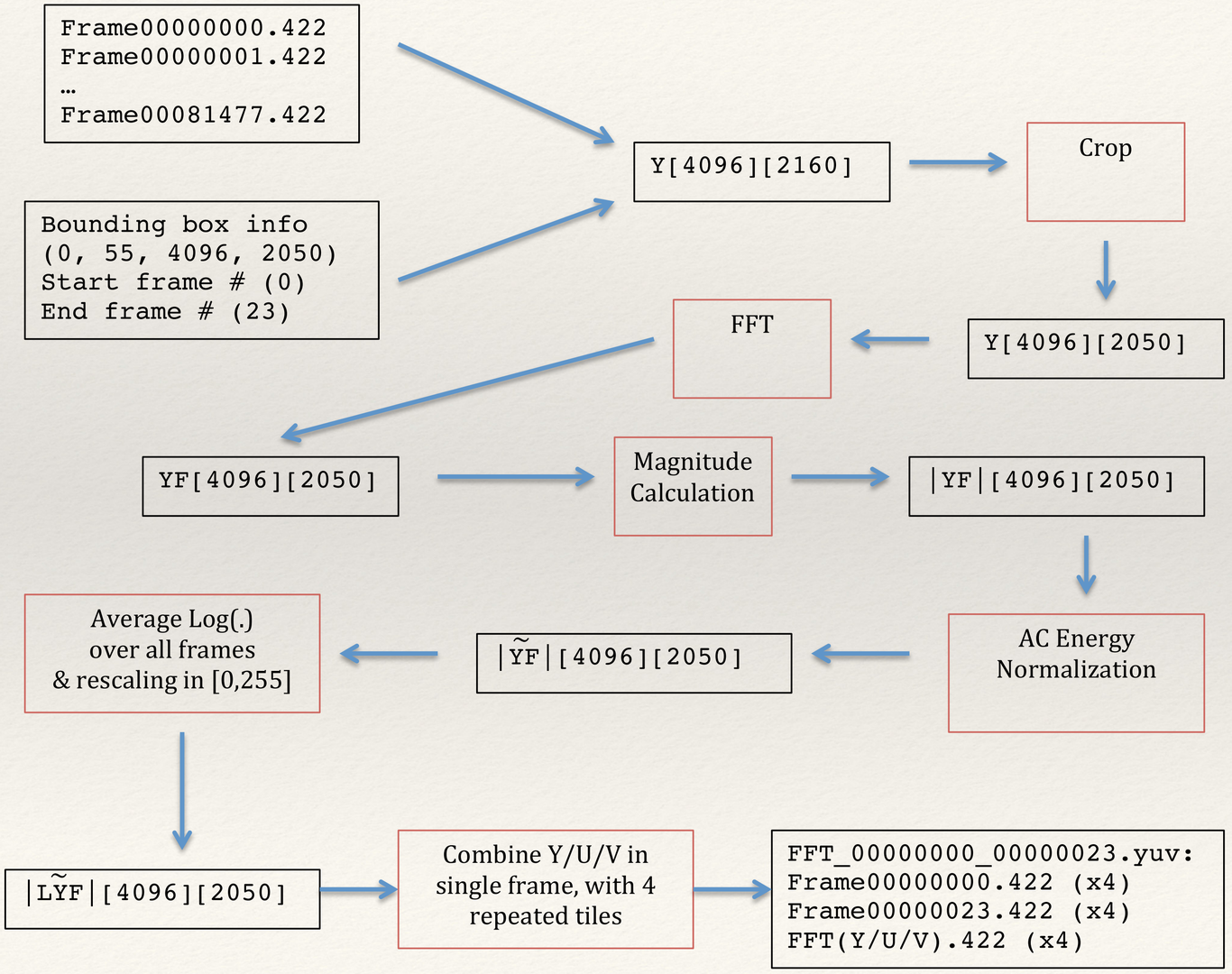
Horizontal/Vertical knee point detection

Low/High frequency Energy density ratio in log domain

Threshold obtained through supervised training

Hor/Ver native resolution estimation with quality/likelihood metrics

Core detection algorithm



Computational complexity dominated by Fourier Transform (FFTW library)

Parallel processing makes it suitable for cloud computing

Back to our example – “Birds in Cage”

“Birds in Cage”– Test 1 Spectrum

“Birds in Cage”– Test 2 Spectrum

“Birds in Cage”– Test 3 Spectrum

“Birds in Cage”– Test 4 Spectrum

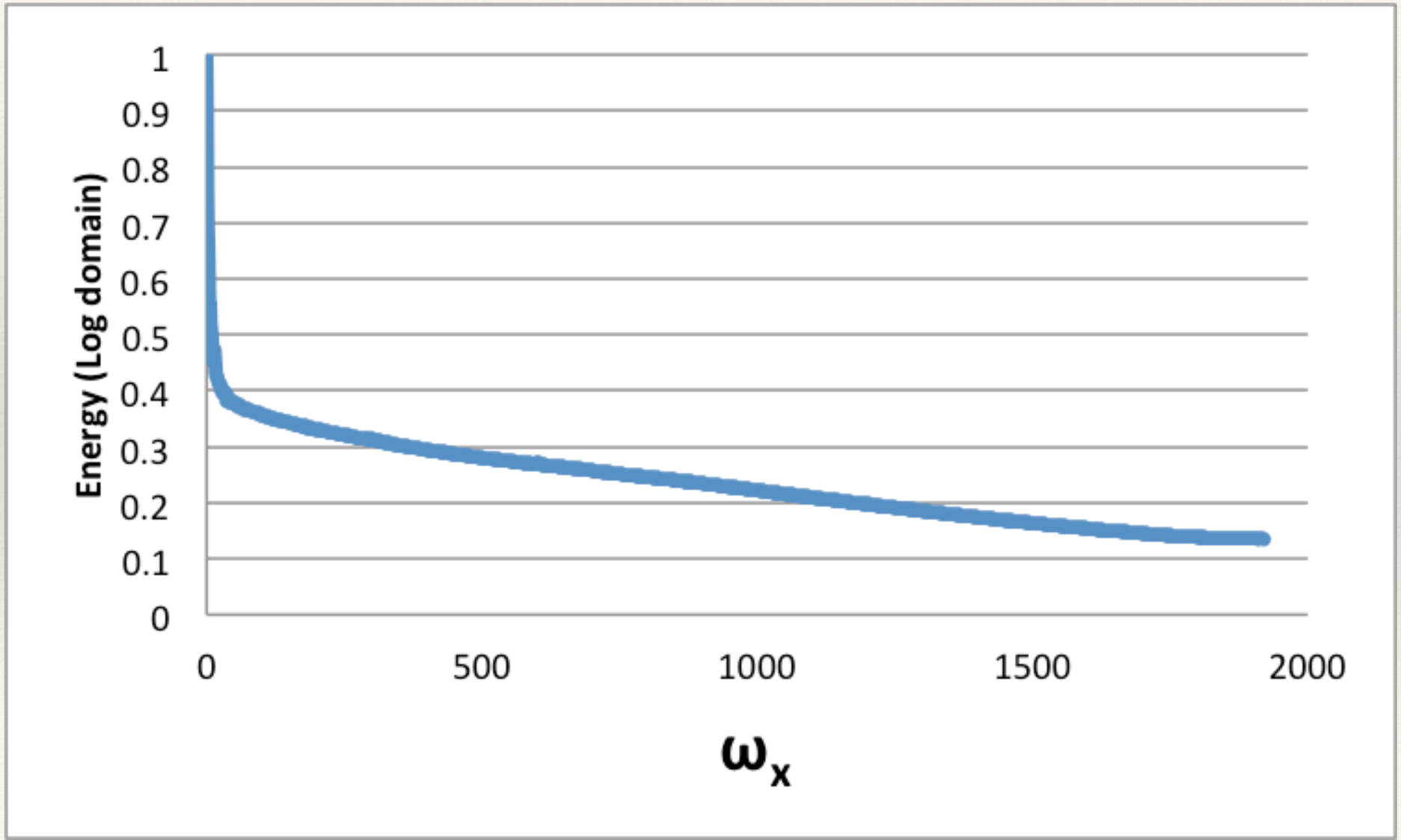
“Birds in Cage” – Test 5 Spectrum

Putting our method to the test

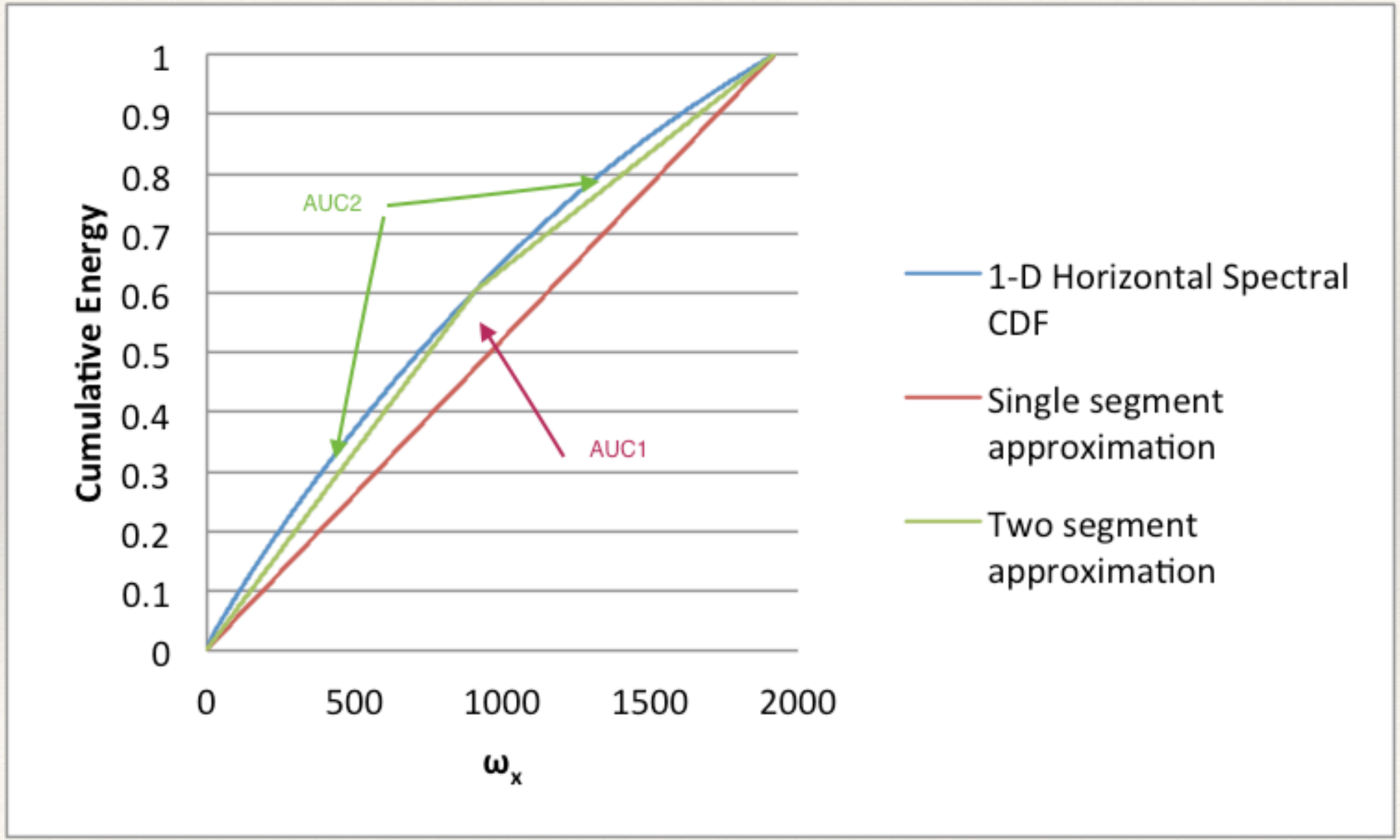
- Tested our native resolution detection method on 2 commercial 4K assets:
  - Asset #1, YUV422, 10-bit/sample, 3840x2160, 29.97fps
  - Asset #2, YUV422, 10-bit/sample, 4096x2160, 29.97fps
- Results:
  - Asset #1 is for the most part true UHD
  - Asset #2 has true 4K resolution in about 75% of its duration; rest 25% is a mix of 1080p and 720p content, upsampled to 4K

Asset #1– Frames [17751,17825]

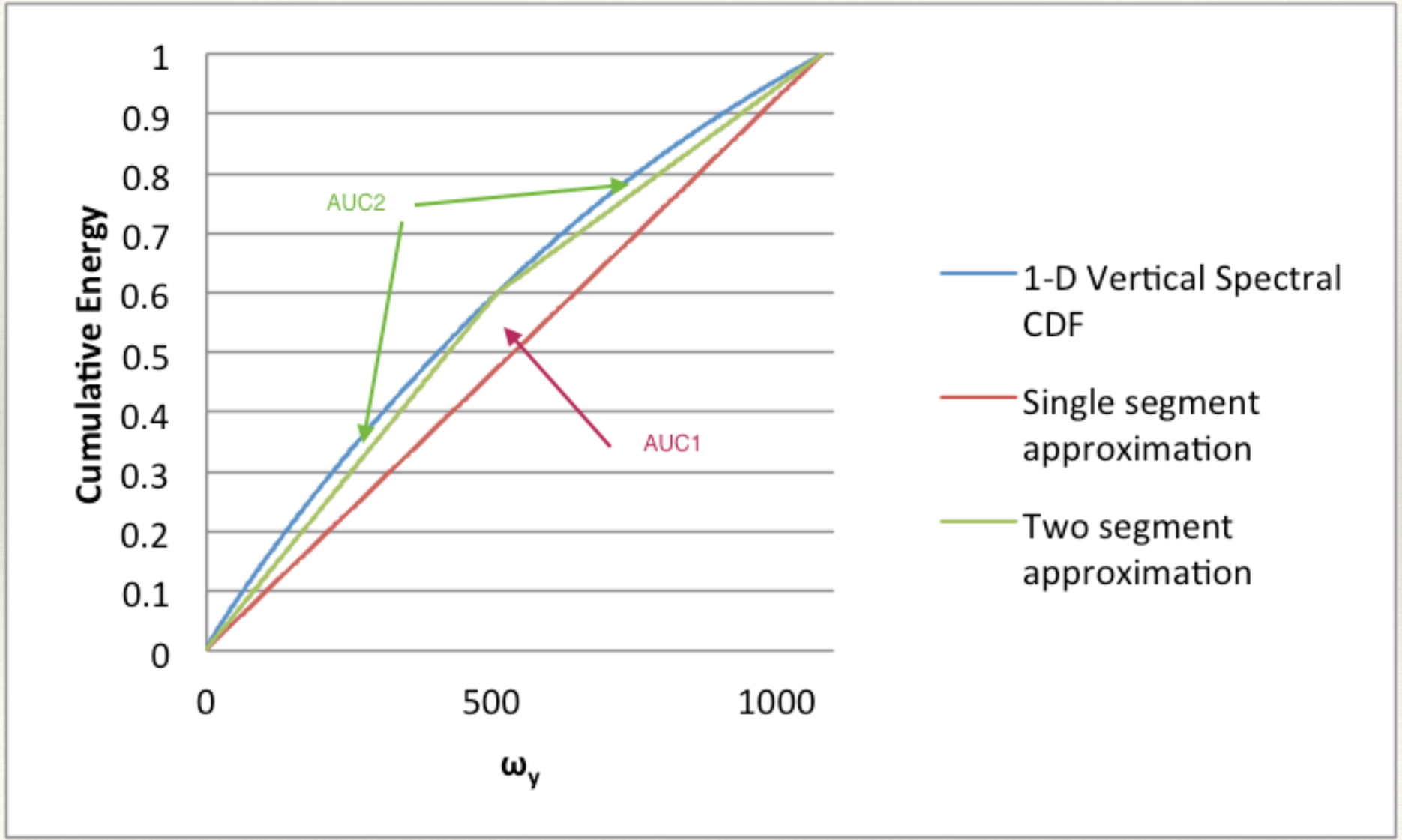
Horizontal spectrum energy



Horizontal spectrum knee-point fit



Vertical spectrum knee-point fit



Candidate low-frequency bounding box

HFD = High Frequency Energy/High Frequency Area

LFD = Low Frequency Energy/
Low Frequency Area

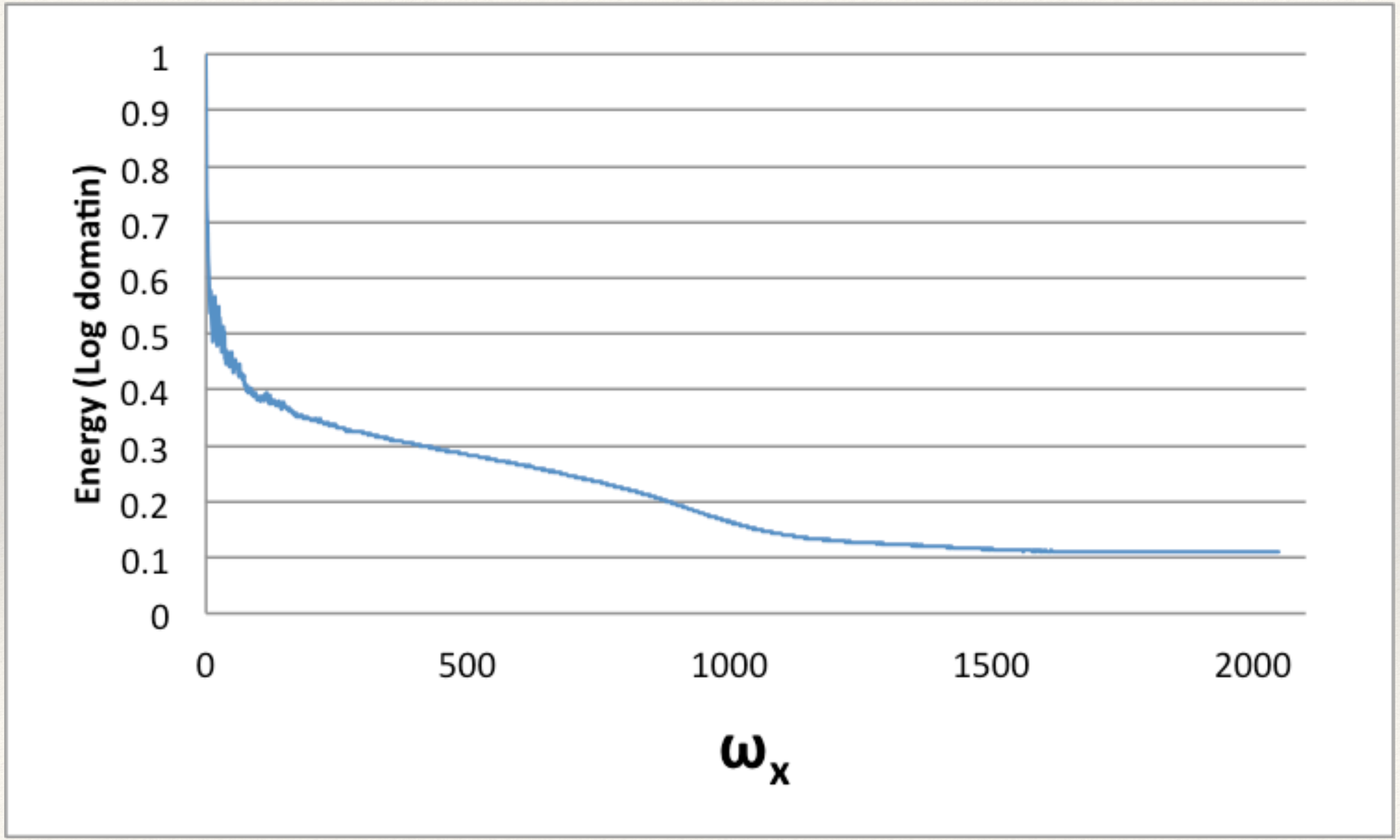
LRD / HRD = 2
Bounding box =
1814 x 1018

Asset #2– Frames [53041,53100]

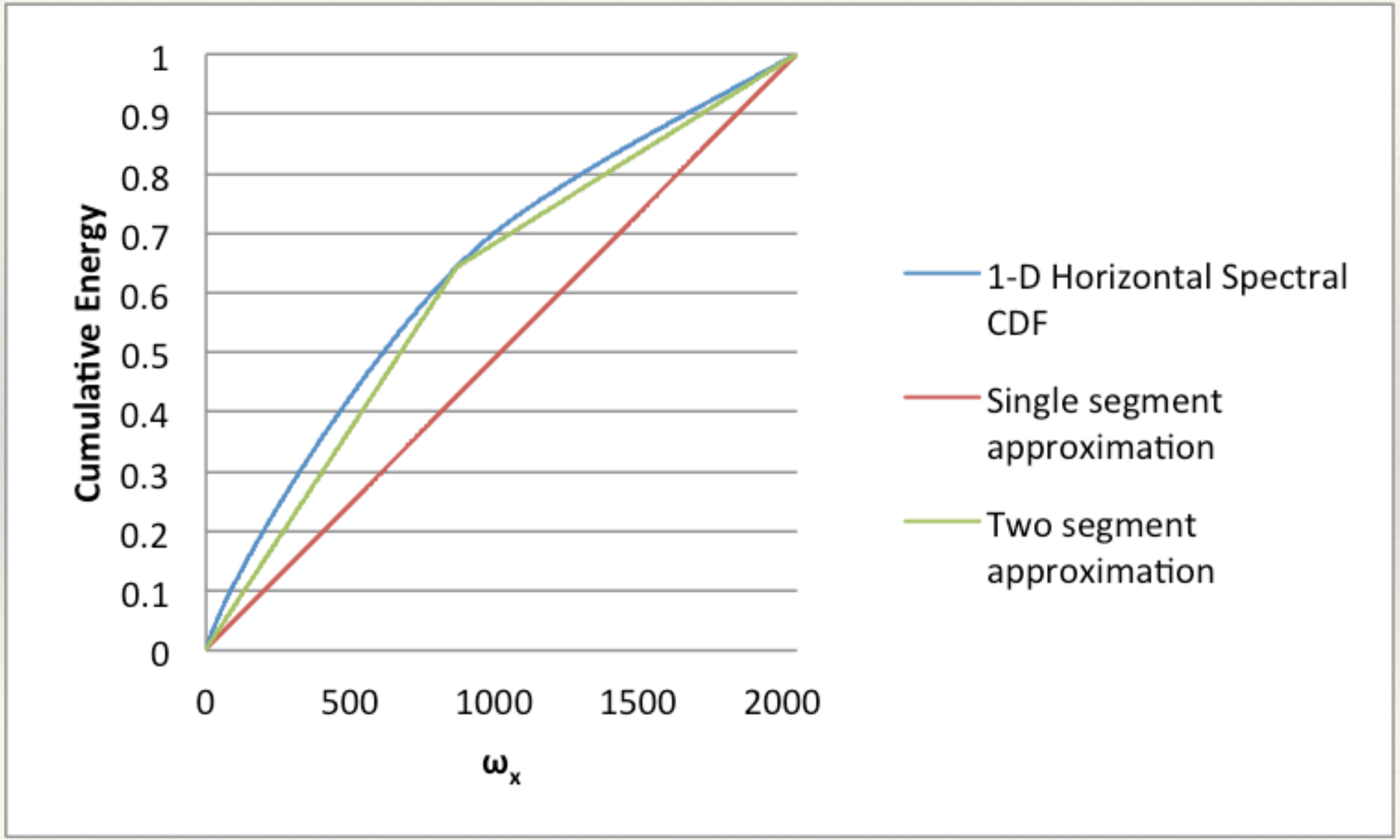
Asset #2– Frames [53154,53188]

Asset #2– Frames [53189,53221]

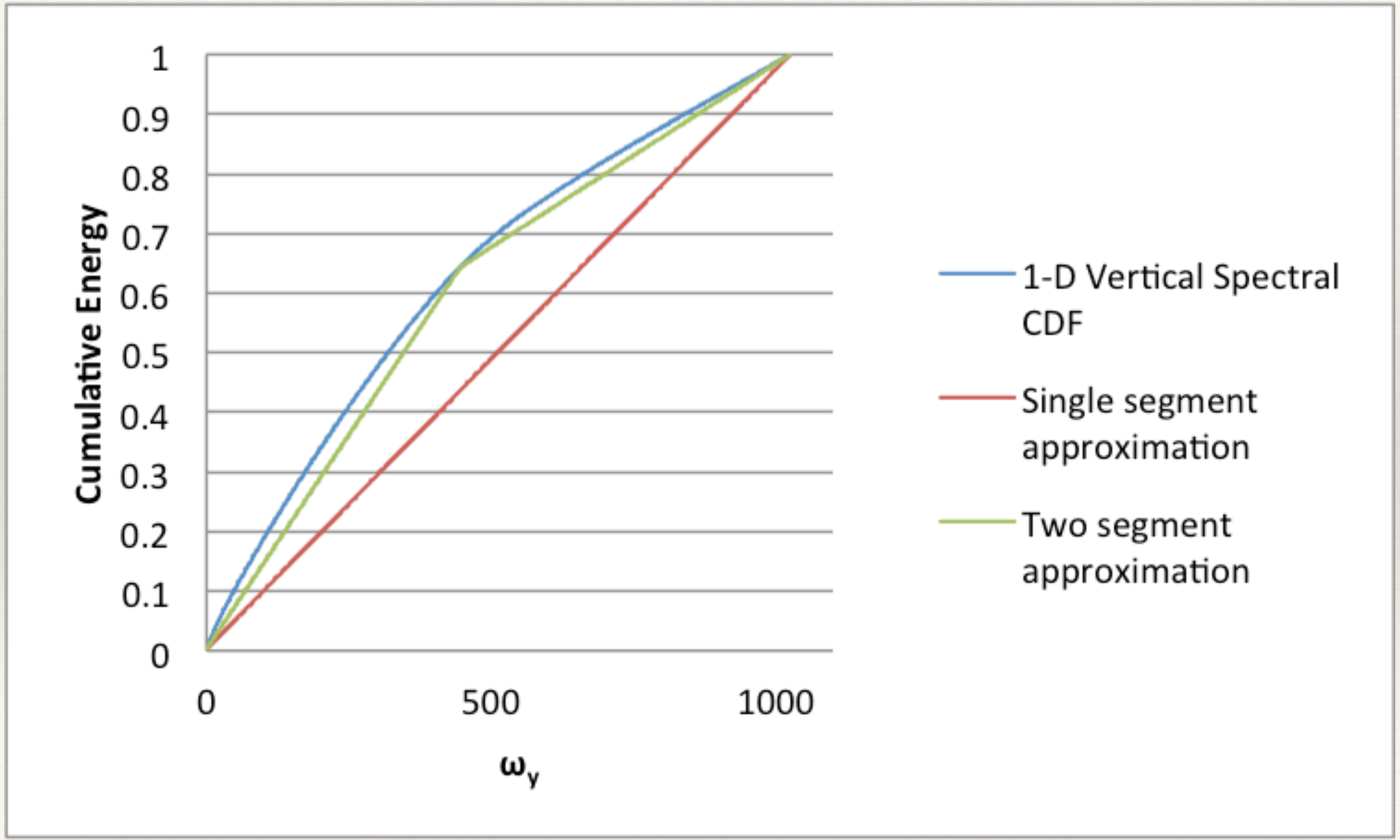
Horizontal spectrum energy



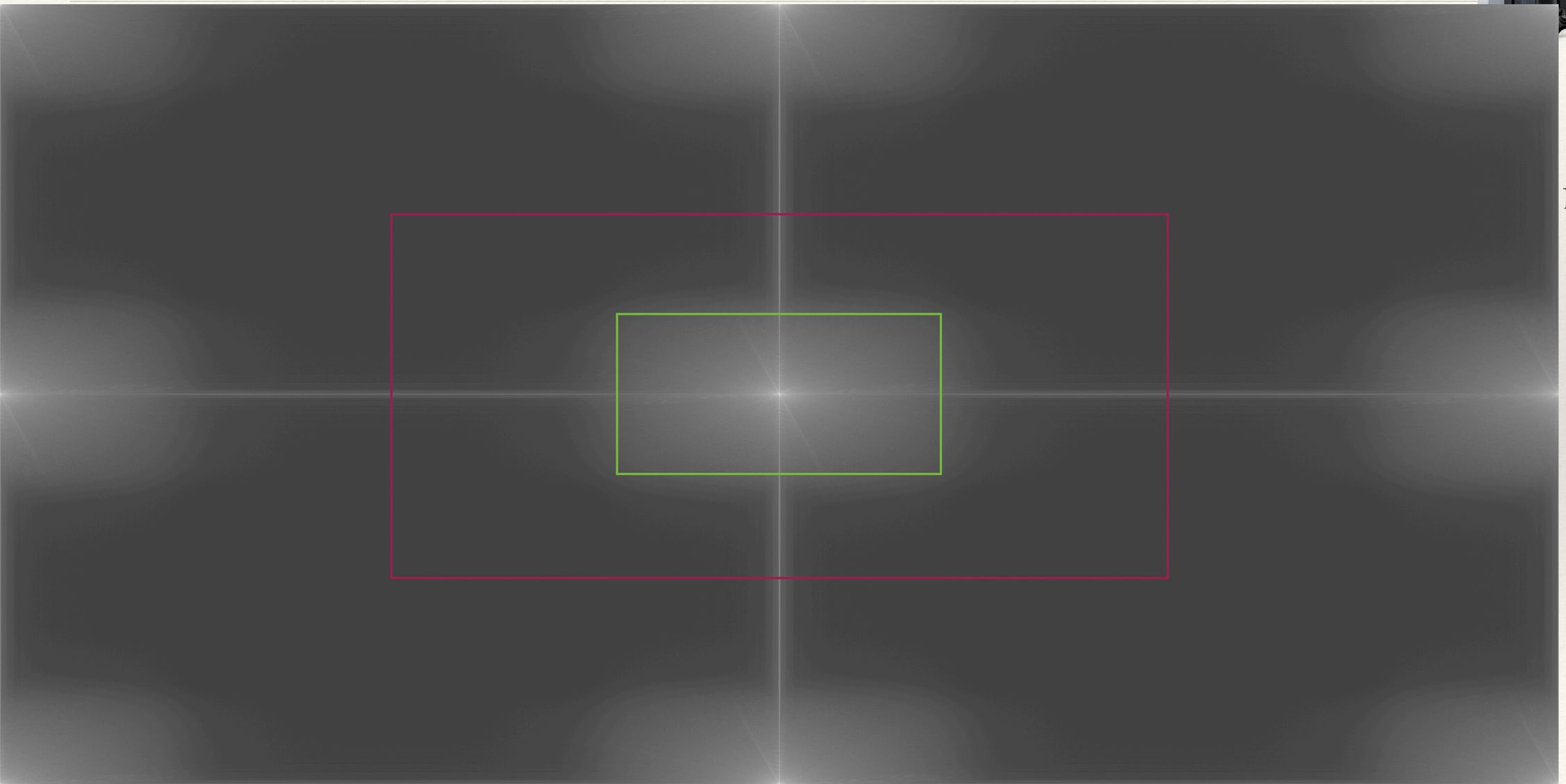
Horizontal spectrum knee-point fit



Vertical spectrum knee-point fit



Candidate low-frequency bounding box



LRD / HRD = 3.51
Bounding box =
1738 x 892

Results

- Method works!
- Many scenes with special effects have been produced in 1080/720p and upsampled
- Image sensor contributes to spectral image (Bayer pattern)
- Scan from film has more high-energy component
- Artistic intent can be mistaken as down/up-sampled content (capture of older TV or camcorder footage)