VP Wissenschaftliche Arbeitstechniken und Präsentation

Vorbis

Robert Harsan
Thomas Loch
Jocelyn Ratac

Eigenschaften

- Verlustbehafteter Audiocodec
- Bei allen Bitraten kompetitiv
- Variable Bitrate
- Samplingraten von 8 bis 192 KHz
- 255 Audiokanäle möglich
- Rechnerisch einfacher zu Dekodieren als MP3
- Viele Platzhalter für Künftige Erweiterungen

Geschichte

- Chris Montgomery begann 1993 mit Arbeit an Audiokompression
- Intensive Entwicklung begann 1998 nachdem Fraunhofer angekündigt hat, Lizenzgebühren für alle MP3 Encoder, kommerzielle Player und auf in MP3 Format verkaufte Musik zu verlangen
- Version 1.0 des Formats in 2000
- LibVorbis 1.0 in 2002

Lizenz

- Format public domain
- Bibliotheken unter BSD-Lizenz
- Die meisten Tools unter GPL-Lizenz

Header

- Identification Header, Comment Header, Codec Setup Header
- Für Dekodierung müssen alle drei in der korrekten Reihenfolge vorhanden sein
- Danach kann beliebiger Paket des Streams wiedergegeben werden

Identification Header

- Identifiziert Vorbis-Version (immer 0) und Anbieter
- Anzahl der Audiokanäle
- Samplingrate
- Blockgrößen
- Bitrate (unverbindlich)

Comment Header

- Kurze Metadaten
- Liste von Vektoren
- UTF-8
- Feldnamen (Charaktere bis =) können nur case-insensitve ASCII Charaktere sein und können öfters vorkommen
- FELDNAME=Feldinformation

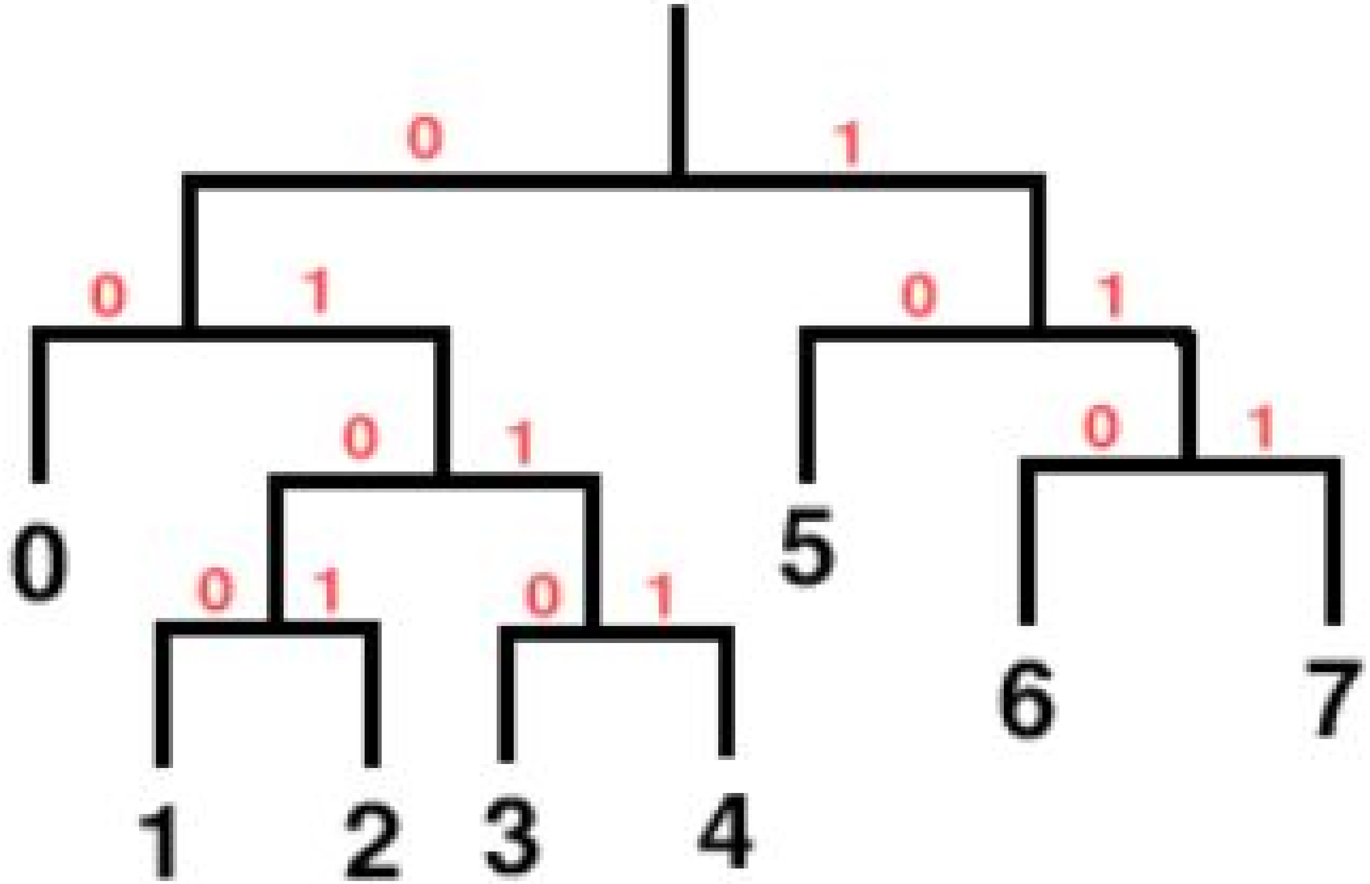
Codec Setup Header

- Codebooks
- Time domain Transforms (Platzhalter)
- Dekodierkonfiguration für Floors und Residues
- Mappings
- Modes (Platzhalter)

Codebooks

- Für Dekodierung der Entropiekodierung benötigt
- Huffman-Kodierung
  - Verlustfreie Kompression
  - Jedem Zeichen wird eine Bitfolge Zugeordnet
  - Größte Häufigkeit → Kürzeste Bitfolge
  - Keine Zuordnung darf Präfix einer anderen Sein

Huffman Binärbaum



Codebooks

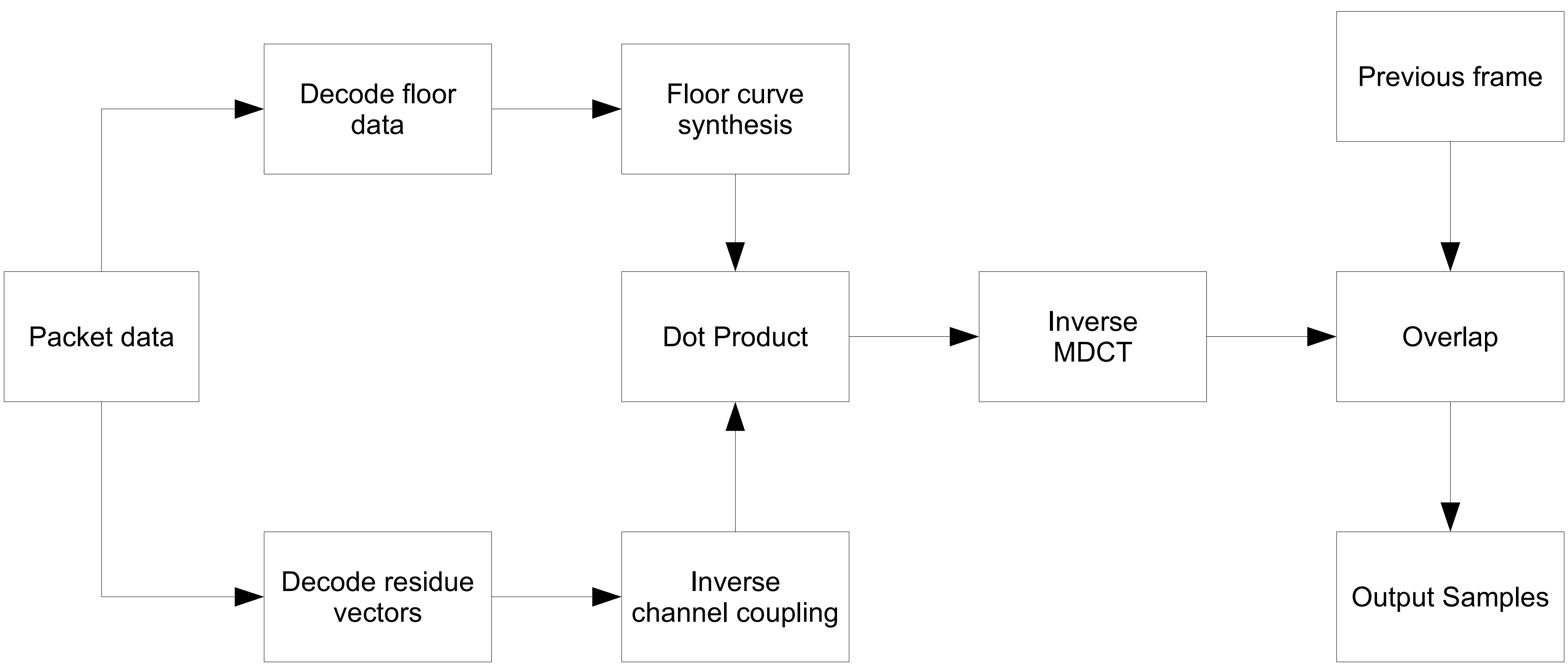
- Vorbis hat im Gegensatz zu fast allen anderen Audiocodecs kein statisches Wahrscheinlichkeitsmodell
- Huffman-Kodierung meist meistens in Verbindung mit einer Vector Lookup Tabelle (2 Methoden)
- Codebücher Verbrauchen nur wenige Kilobytes

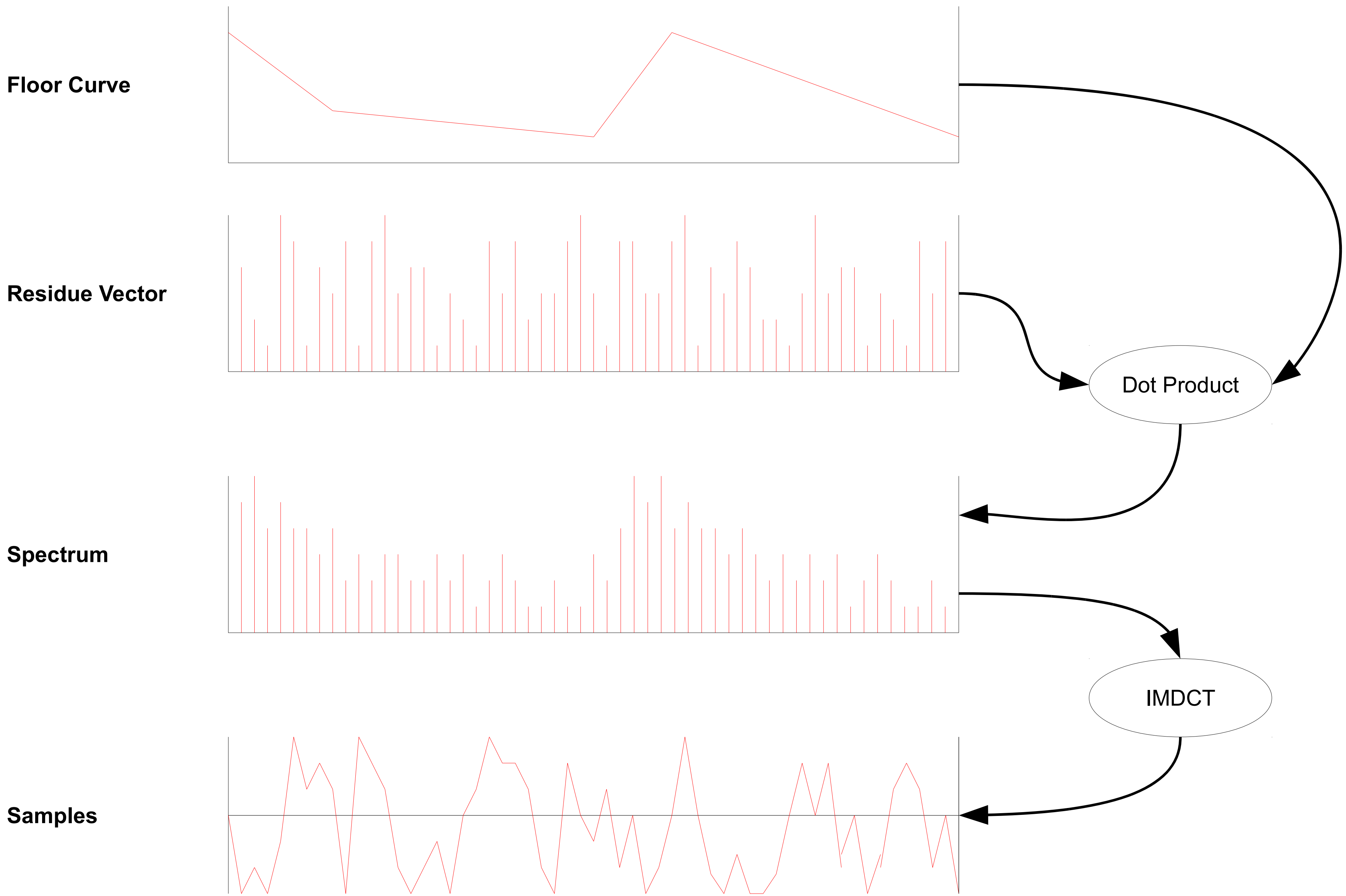
Mappings

- Channel Coupling Description
- Submaps
  - Konfiguration für ein Subset von Floor-und Residue-Vektoren
  - Bsp: bei 5.1 Audio ist der 6. Kanal nur Bass und hat ein anderes Submap als die anderen Kanäle wodurch er einen kleineren Frequenzbereich hat

Audio Packet Structure and Decoder Pipeline

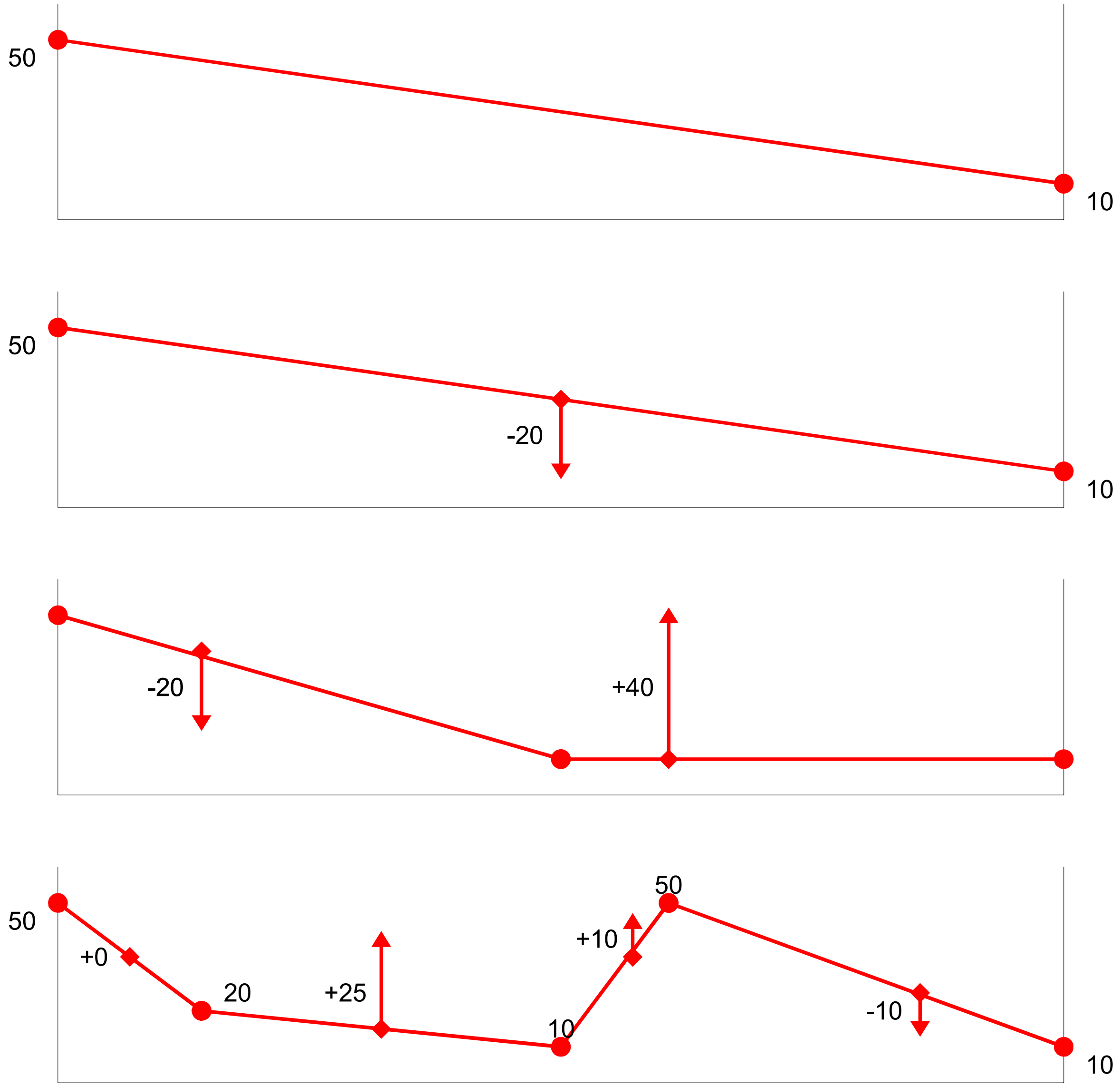
Type Mode Window	Floor curves	Residue vectors





Floor Curve
Residue Vector
Dot Product
Spectrum
IMDCT
Samples

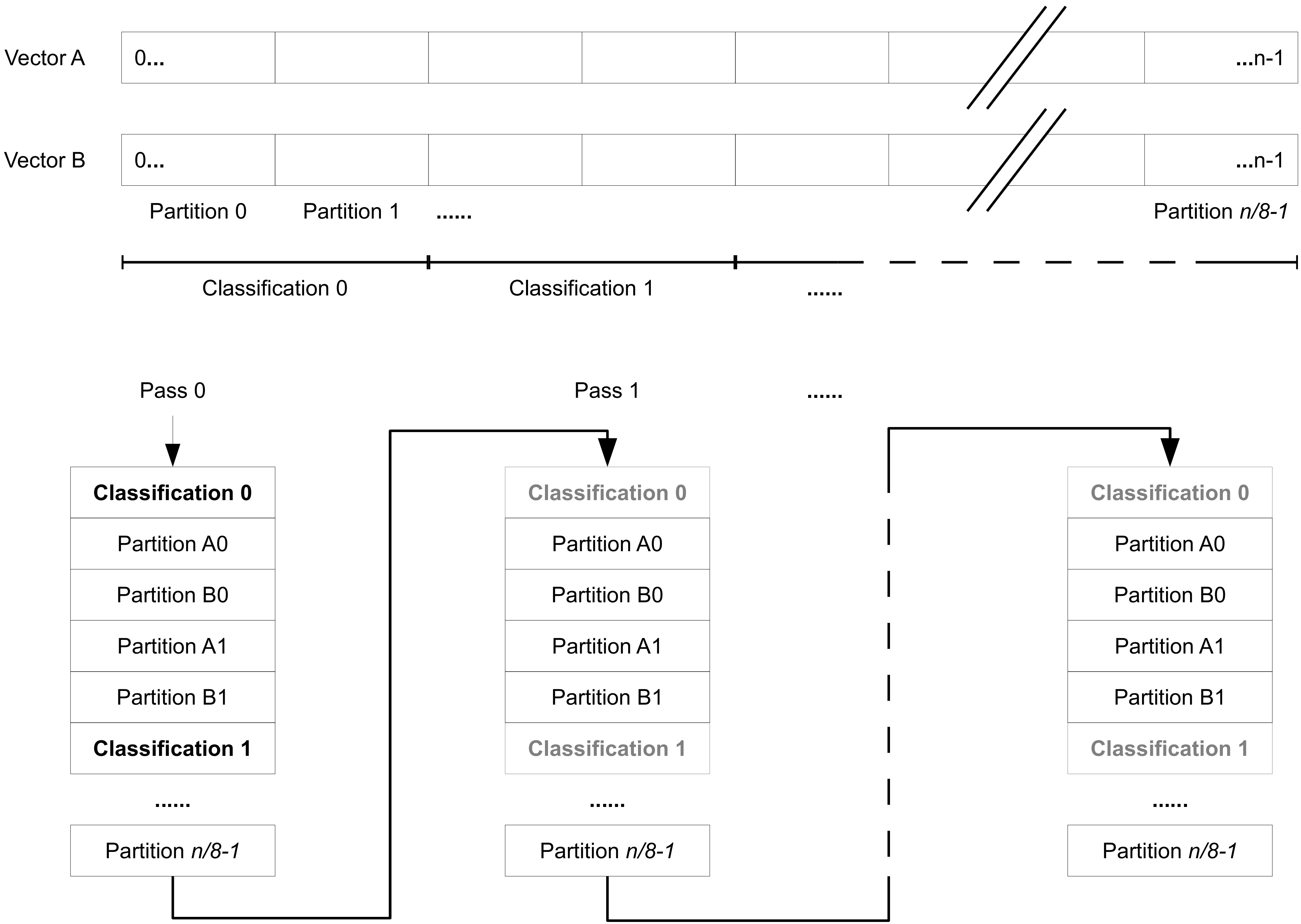
Floor Type 1 Curve Synthesis: Point-wise Linear Function



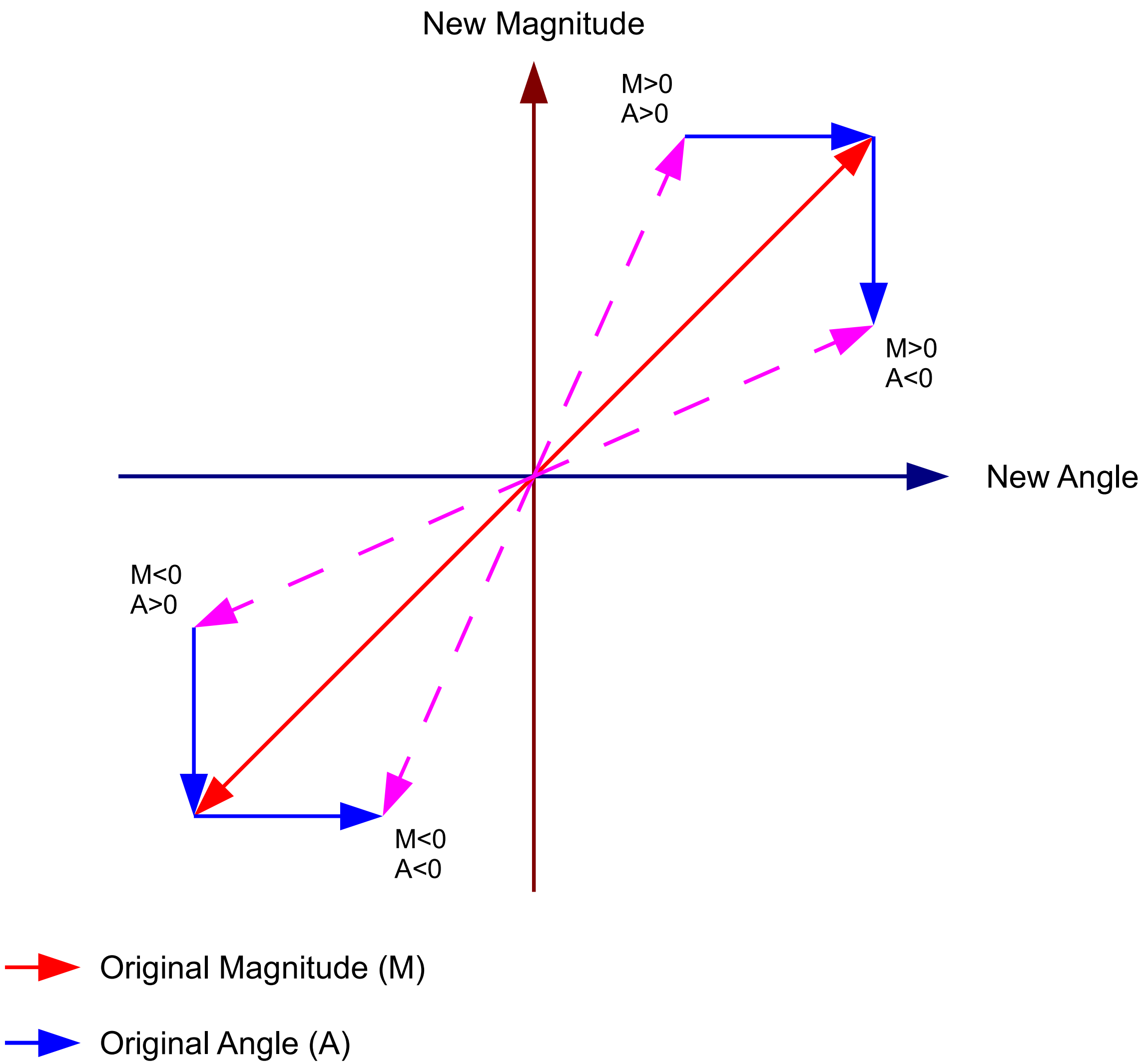
Base case: Function only has two points

Induction step: Function has n points → add (at most) *floor(n/2)* new points

Residue Vectors



Inverse Coupling



If M > 0:

$$\text{New M} = \begin{cases} M & \text{if } A > 0 \\ M + A & \text{otherwise} \end{cases}$$

$$\text{New A} = \begin{cases} M - A & \text{if } A > 0 \\ M & \text{otherwise} \end{cases}$$

Otherwise:

$$\text{New M} = \begin{cases} M & \text{if } A > 0 \\ M - A & \text{otherwise} \end{cases}$$

$$\text{New A} = \begin{cases} M + A & \text{if } A > 0 \\ M & \text{otherwise} \end{cases}$$

dot product

- A 280dB range is approx. 48 bits with sign

  - residue vector 48 bit range

  - dot product must be able to handle 48x24bit multiplication.

- This range may be achieved using large(64bit or larger)integers or implementing a movable binary point representation.

dot product

- Floor vector values can span ~140dB(~24 bit unsigned)
- Audio spectrum vector a minimum of 120dB(~21 bits signed)
- Residue vector – to reach full scale:

  - if the floor is -140dB ---> 0 to +140dB

  - if the floor is 0dB ----> -140dB to +140dB

MDCT(Modified Discrete Cosine Transform)

- A Fourier-related transform based on (DCT-IV) with the addition of being lapped.
- Designed to performed on consecutive blocks of larger dataset, where subsequent blocks are overlapped, so that the last half of one block coincides with the first half of the next block.
- Overlapping, makes MDCT attractive for signal compression applications, since it helps to avoid artifacts stemming from the block boundaries.

MDCT

- Employed in most modern lossy audio formats like MP3, AC-3, Windows Media Audio, Cook and AAC
- Proposed by Princen, Johnson and Bradley in 1987
- Earlier by Princen and Bradley to develop MDCT's underlying principle of time domain aliasing cancellation(TDAC)

MDCT

- Forward Transform

$$X_k = \sum_{n=0}^{2N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

- Inverse Transform

$$y_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

Windowed MDCT

Windows applied to MDCT are different from windows used for other types of signal analysis because of Princen-Bradley condition:

$$w_n^2 + w_{n+N}^2 = 1$$

Window function:

$$w_n = \sin\left(\frac{\pi}{2}\sin^2\left[\frac{\pi}{2N}\left(n+\frac{1}{2}\right)\right]\right)$$

overlap/add data

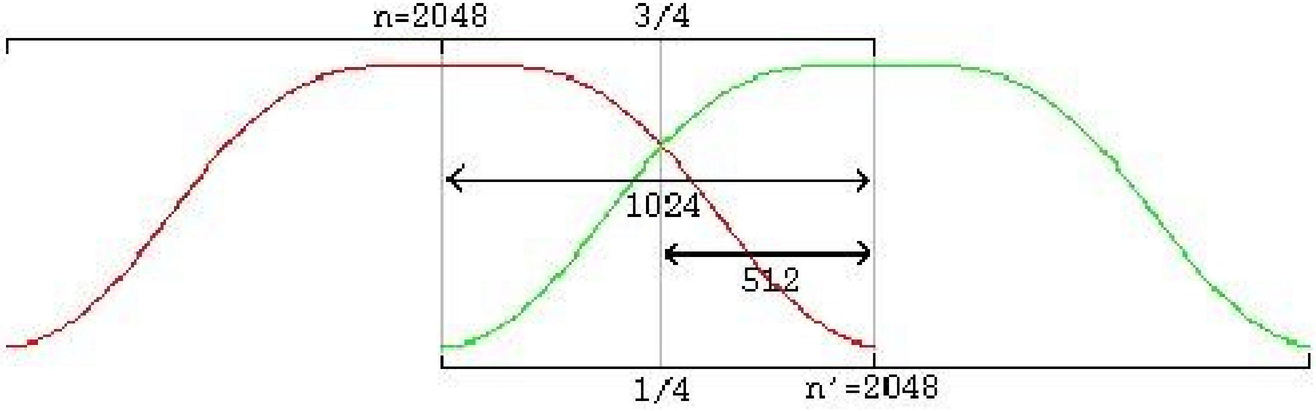
- Windowed MDCT output is overlapped and added with the right hand data of the previous window
- When overlapping, a short and long window, much of the returned range does not actually overlap.
- Amount of the data to be returned is:

```
window_blocksize(previous_window)/4+window_blocksize(current_window)/4
```

Window shape decode

For equal sized window:

- Overlapped 50% with the output of the previous frame and added.

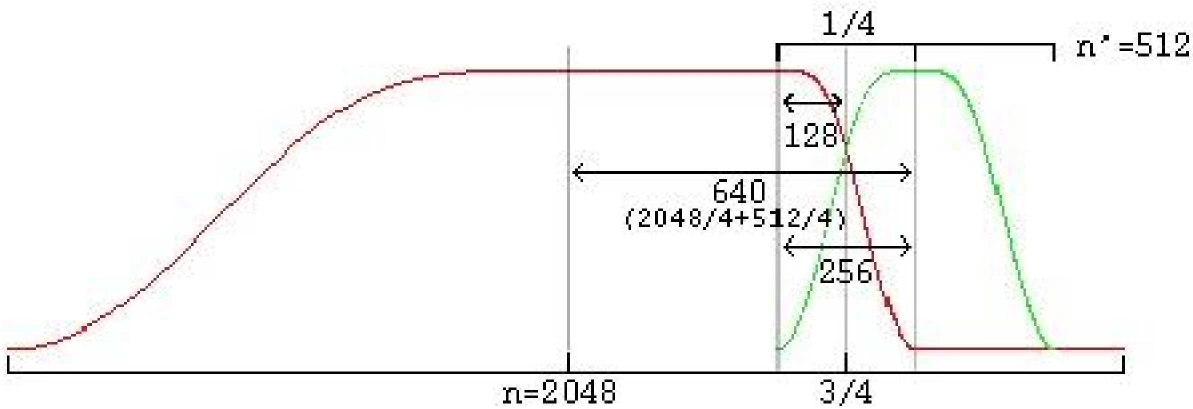


Window shape decode(cont.)

- For unequal sized window case:

  - ¾ point of the previous window is aligned with the ¼ point of the current window

  - the overlapped portion is the finished data to be returned to the decoder.



output channel order

- Vorbis specifies only channel mapping type 0

  Mapping 0 is defined as follows:

  - one channel– the stream is monophonic

  - two channels– stereo, channel order: left, right

  - three channels – 1d-sorround encoding

    left, center, right

  - four channels – quadrophonic sorround

    front left, front right, rear left, rear right

output channel order(cont)

- five channels – five-channel sorround

  front left, front center, front right, rear left, rear right
- six channels – the stream is 5.1 sorround

  front left, front center, front right, rear left, rear right, LFE
- greater than six – channel use and order is defined by application.