



MFCC: Mel-Frequency Cepstrum Coefficients (MFCC) feature extractor

Stéphane Emery

CSEM White Paper • May 2025

 [ASICs for the Edge](#)

This document is the property of CSEM S.A. Users may not use the work for commercial purposes, they may not redistribute it in modified form, and they must give credit to the author.

This document is the property of CSEM S.A. Users may not use the work for commercial purposes, they may not redistribute it in modified form, and they must give credit to the author.

Table of contents

Empowering Voice Signal Feature Extraction	3
Industry-Standard Voice Signal Processing	3
MFCC Core Block Architecture: Precision and Efficiency	3
Specifications: Designed for Optimal Performance and High Flexibility	4
Reference implementation.....	4
Analysis and Models.....	5
References	5

Empowering Voice Signal Feature Extraction

The Mel-Frequency Cepstrum Coefficients (MFCC) core block is designed to extract features from voice signals with high efficiency. This technology is ideal for applications such as keyword spotting, speaker identification, and voice activity detection. This white paper provides a comprehensive overview of the MFCC core block architecture, specifications, and integration guidelines.

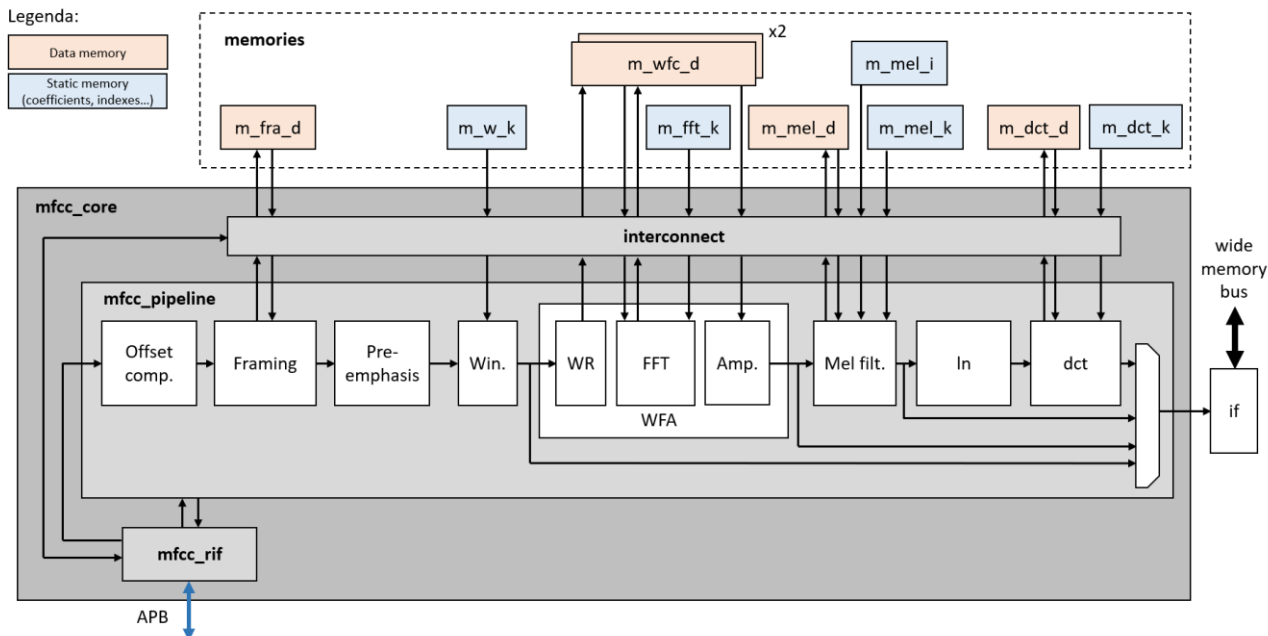
Industry-Standard Voice Signal Processing

The MFCC core block is based on the ETSI standard "ETSI ES 201 108", Version 1.1.3, which outlines the process for extracting MFCC features from audio signals. This standard ensures accuracy and reliability in various voice processing applications.

MFCC Core Block Architecture: Precision and Efficiency

The MFCC core block features a well-designed architecture with several sub-blocks that work together to deliver precise feature extraction. Key components include:

- **Offset Compensation:** Removes DC offset for cleaner audio signals.
- **Framing:** Efficiently divides audio signals into frames.
- **Pre-emphasis:** Enhances high-frequency components.
- **Windowing:** Applies a programmable window to optimize signal processing.
- **FFT:** Performs Fast Fourier Transform for frequency domain analysis.
- **Amplitude Computation:** Utilizes the CORDIC algorithm for accurate magnitude calculation.
- **MEL Band Filter:** Extracts MEL frequency components.
- **Natural Logarithm:** Computes logarithmic transformations.
- **Discrete Cosine Transform (DCT):** Converts MEL frequency components into MFCCs.



Specifications: Designed for Optimal Performance and High Flexibility

The MFCC core block supports various configurations for audio signal input and feature extraction. Key specifications include:

- **Audio Signal from ADC:**
 - Sampling rate: 8-16 kSps
 - Precision: 16 bits

- **MFCC Feature Extraction:**
 - Input precision: 16 bits
 - Output precision: 16 bits
 - Frame length: 1..1024 samples
 - Frame stride: 1..1024 samples
 - FFT size: 256-512-1024 points
 - Number of MEL bands: 1..64
 - Number of Features (DCT outputs): 1..64

- In the MFCC core, a multiplexer is placed at the end of the pipeline to select the internal block output to be sent on the MFCC output. The options are:
 - the output of the windowing block
 - the norm of the signal spectrum
 - the MEL band filter output (MEL bands)
 - the DCT output (Features)

The MFCC core block has been tested with various configurations based on the ETSI standard and Hello Edge paper [1] that can be used to benchmark its performance:

Specifications	Test cases					Unit	Comment
	ETSI_1	ETSI_2	ETSI_3	HelloEdge_1	HelloEdge_2		
Sampling rate	8	11	16	16	16	kSps	
Window length	1	1	1	1	1	s	As in Google Speech Command data set
Frame length	8,000	11,000	16,000	16,000	16,000	samples	
	25.0	23.3	25.0	40.0	40.0	ms	
Frame stride	200	256	400	640	640	samples	
	10.0	10.0	10.0	20.0	40.0	ms	
FFT size	80	110	160	320	640	samples	
	256	256	512	1024	1024		
Number of MEL bands	23	23	23	23	23		HelloEdge does not specify the value, 23 is often used
Number of features (MFCC size)	13	13	13	10	10		
Number of frames	98	98	98	49	25		
Output Feature matrix	13x98	13x98	13x98	10x49	10x25		
Total Latency	6,113	6,228	12,350	24,950	24,950	cycles	
Minimum clock frequency (Fmin)	515	515	1,073	1,120	560	kHz	Rate given by the number of cycles to execute the FFT, CORDIC and the MEL output.

The latency for each test case is computed for the worst-case scenario, ensuring that framed samples are always within the given window.

Reference implementation

The MFCC IP has been integrated in CSEM's [Fibonacci](#) ML SoC with ULP standard cells targeting 200MHz performance in GF 22FDX technology. MFCC requires the memory instances listed below to be provided externally to the IP (in this example, summing to a total of 24kB)^a :

^a CSEM can provide a standalone wrapper for the MFCC block including all memories for a given technology.

Global variables		Memory name	Block	Access from MFCC	Entries #	Bitwidth bit	Size Bytes	Type	Description
Bitwidth	16 bits	m_fra_d	Framing	Write + Read	2 048	16	4 096	Single-port SRAM	One frame+stride samples per input source (mics) are stored here
Max frame length	1024 samples	m_w_k	Windowing	Read	512	16	1 024	Single-port SRAM	Coefficients of the windowing function
Max frame stride	1024 samples	m_wfc_d[0]	Windowing/FFT/CORDIC	Write + Read	512	32	2 048	Single-port SRAM	Data memory shared by Windowing, FFT, and CORDIC (2 banks)
Max input sources (mics)	1	m_wfc_d[1]	Windowing/FFT/CORDIC	Write + Read	512	32	2 048	Single-port SRAM	Data memory shared by Windowing, FFT, and CORDIC (2 banks)
Max FFT size	1024 points	m_fft_k	FFT	Read	256	32	1 024	Single-port SRAM	FFT twiddle factors
Max mel bands	64 bands	m_mel_d	Mel filter	Write + Read	512	16	1 024	Single-port SRAM	Stores CORDIC output data as mel filter has specific access patterns
Natural Log. addr. bitwidth	11 bits	m_mel_i	Mel filter	Read	64	32	256	Flip-flops	Mel filter indexes on how to access m_mag_d
Max MFCCs	64 values	m_mel_k	Mel filter	Read	2 048	16	4 096	Single-port SRAM	Mel filter coefficients
		m_dct_d	DCT	Write + Read	64	16	128	Flip-flops	Output natural logarithmic data as DCT has specific access patterns
		m_dct_k	DCT	Read	4 096	16	8 192	Single-port SRAM	DCT coefficients
						Total		23 936	

The following table presents the PPA results for different standard-cell flavors (including the Fibonacci reference implementation), for the core of the MFCC IP (without memories), for a scenario in which the SoC runs voice activity detection with the MFCC pipeline active every clock cycle.

Scenario	Fibonacci Reference	ULP Version	ULL Version	ULL Version
Tech Flavour	ULP^b+ULPSL^c	ULP	ULL^d	ULL
Voltage	0.65V	0.65V	0.80V	0.80V
Target Synthesis Frequency	200MHz	100MHz	50MHz	50MHz
Gate Equivalent Area (kGE) ^e	38.0	38.0	39.6	39.6
Benchmark	HelloEdge_1 (KWS with small LSTM)			
Benchmark Frequency (MHz)	50	50	50	1.2 ^f
MFCC Core Dynamic Power μ W	164.4	164.7	315.6	7.76
MFCC Core Leakage Power μ W	78.1	4.76	0.046	0.046
MFCC Core Power (excl. memories) μ W	242.5	169.4	315.7	7.81

Analysis and Models

A reference bit-true python model is available, usable for PTQ / QAT (Post-Training Quantization / Quantization-Aware Training) training and deployment with PyTorch framework.

References

[1] Zhang, Yundong, et al., "Hello Edge: Keyword spotting on microcontrollers." arXiv preprint arXiv:1711.07128 (2017).

^b Ultra-low-power standard cell flavor (7.5T) [Low- V_T]

^c Ultra-low-power standard cell flavor (7.5T) [Super-low V_T]

^d Ultra-low-leakage standard cell flavor (8T) [High- V_T]

^e Gate area divided by the area of the drive-strength-one NAND2 gate.

^f Running at the minimum frequency required for the HelloEdge_1 benchmark