HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL & ELECTRONIC
ADVANCED PROGRAM

GRADUATION THESIS

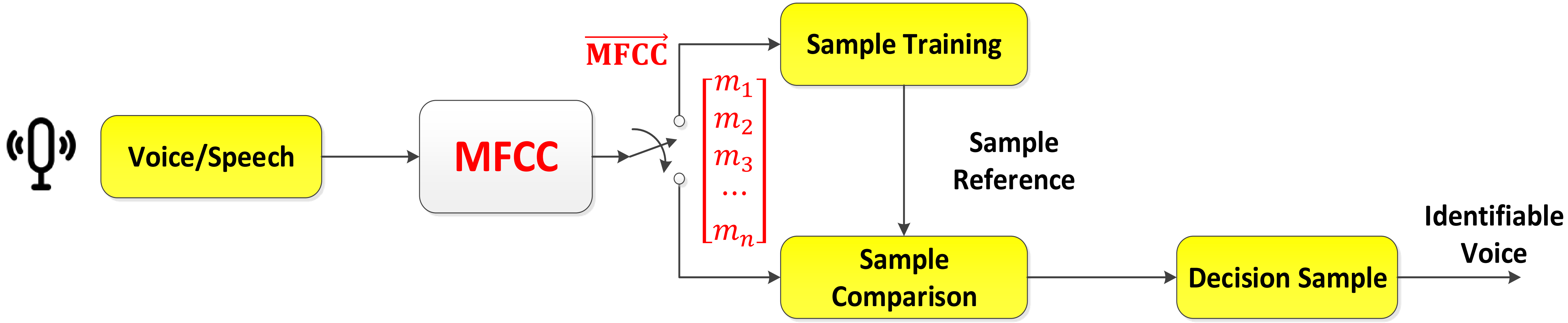
HARDWARE-BASED DESIGN OF DYNAMIC MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

SUPERVISOR : Assoc. Prof . HOANG TRANG
STUDENT : NGO THANH DAT

CONTENTS

CONTENTS

RESEARCH OBJECTIVE

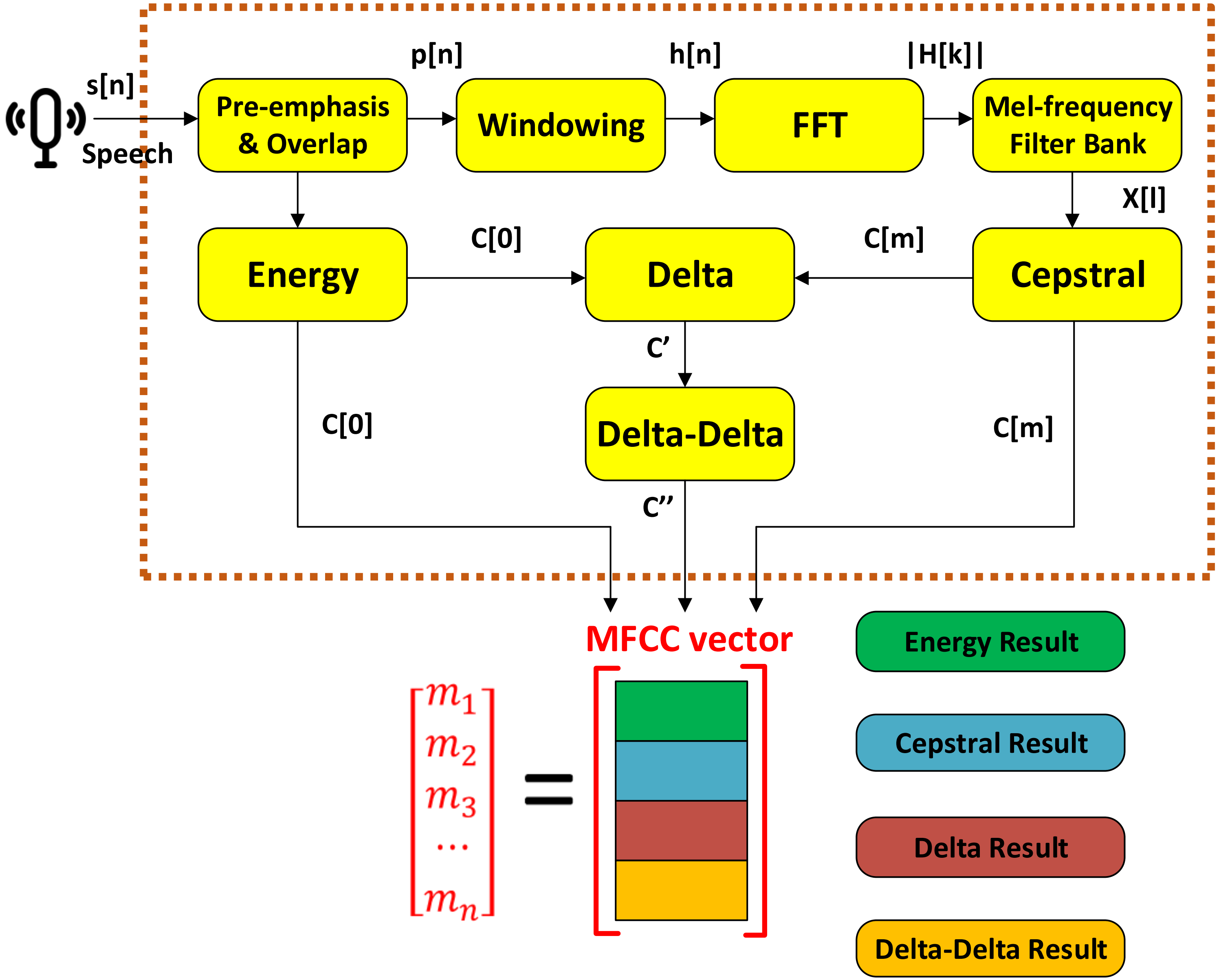


MFCC Feature Extraction:

- ❖ MFCC has been the most essential hardware architecture for ASR (Automatic Speech Recognition) systems [1] - [7]
- ❖ Dynamic MFCC increases 5 % to 6 % of the recognition rate than a fixed one [6]

Reference papers are spread out on the table

RESEARCH OBJECTIVE

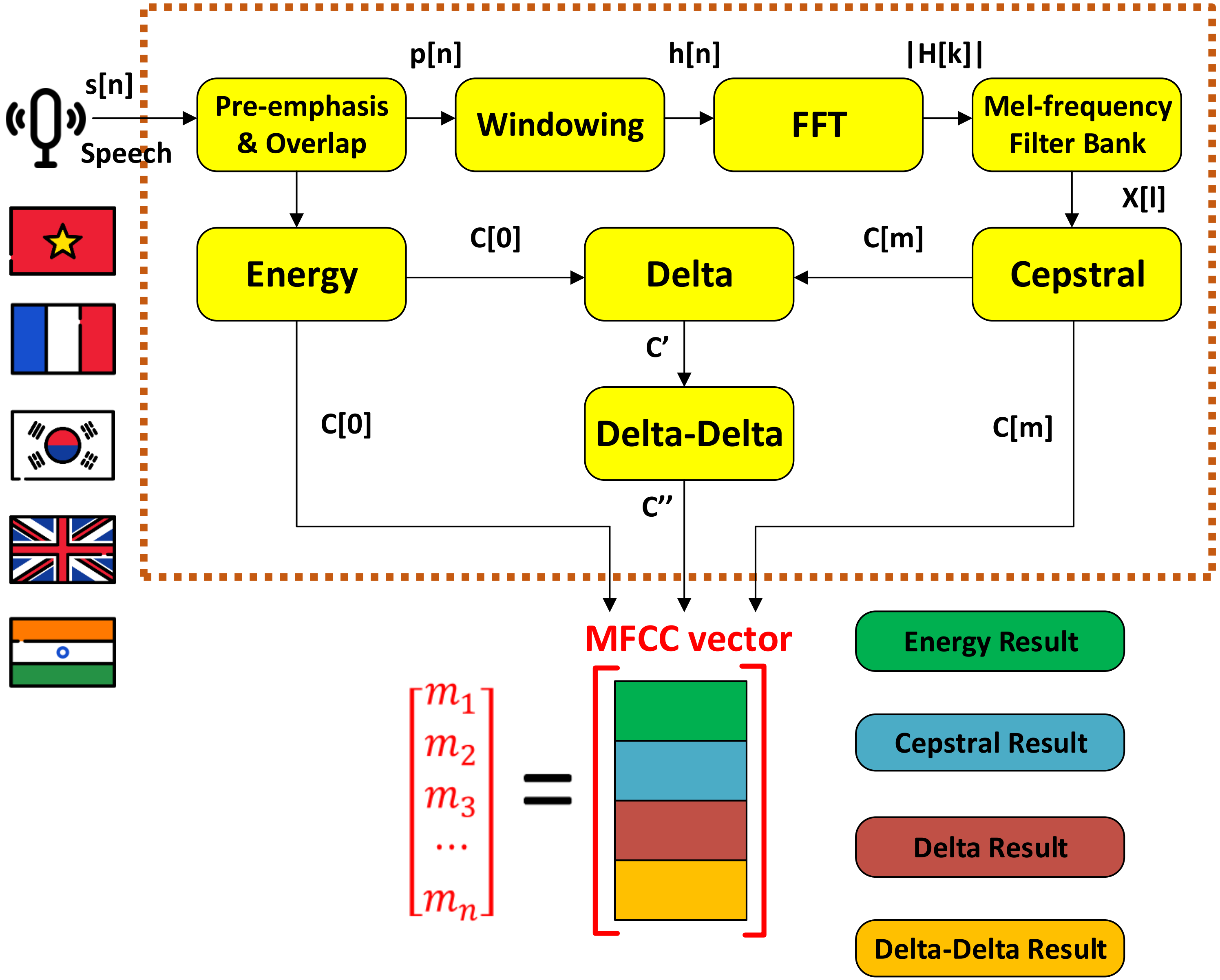


Overall process for MFCC

MFCC vector characteristics :

- MFCC vector's dimension is non-fixed value
- Dimension of MFCC vector is different from languages, word

RESEARCH OBJECTIVE

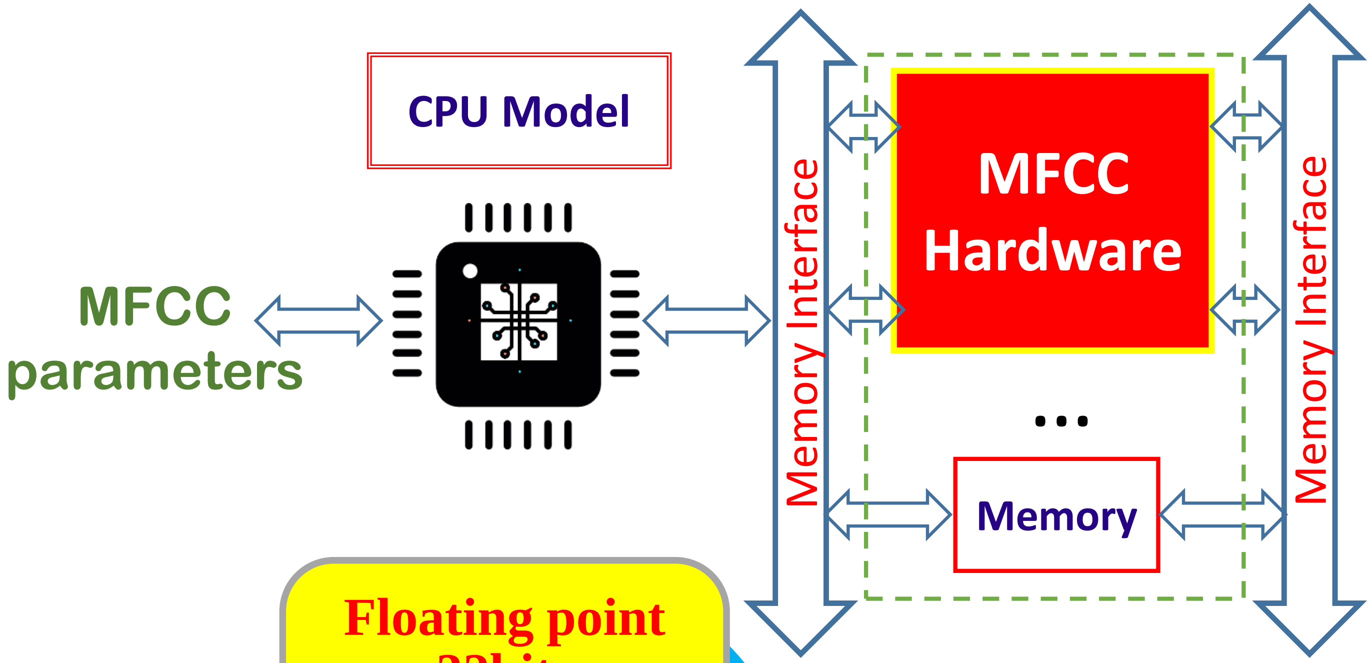


Idea of Dynamic MFCC

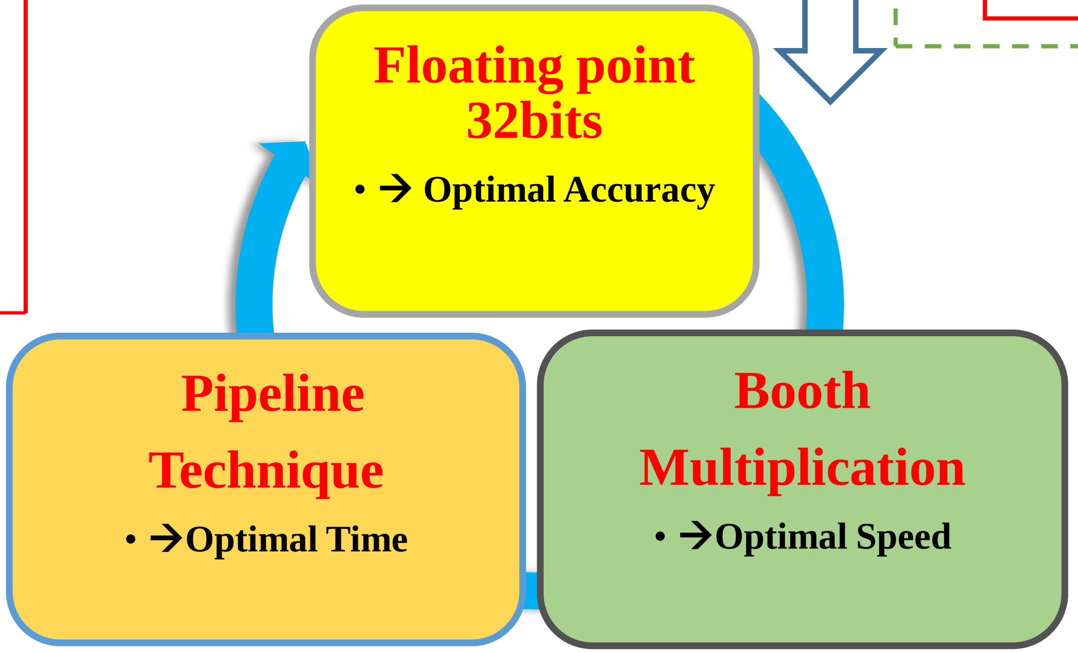
Reference Paper [8] → [15]	
Overlap	50%
Window	128, 160, 256, 512
FFT Point	128, 256, 512
Mel Filter Number	20, 24, 32, 33
Cepstral Number	12, 13, 17, 24
Delta Number	13, 14, 18, 25
Delta-Delta Number	13, 14, 18, 25
Energy number	1

RESEARCH OBJECTIVE

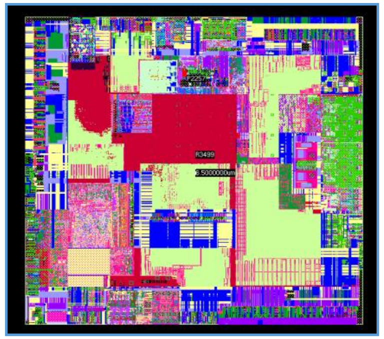
Parameters	Dynamic Range	Reference [8] → [15]
Sample Per Frame	25→512	128, 160, 256, 512
Overlap Ratio (%)	30 → 70	50
FFT Points	8→1024	128, 256, 512
Mel Filters	1→ 63	20, 24, 32, 33
Cepstral Coefficients	1→31	12, 13, 17, 24
Delta Order	1 & 2	1 & 2
MFCC Vector Dimension	1→ 96	26, 28, 36, 50



- ✓ Accuracy problem
- ✓ Real-time issues
- ✓ Ability of reconfiguration



RESEARCH OBJECTIVE

ASIC Design Flow	Output Results
Specification	Matlab Results
RTL Design	Verilog - HDL
RTL Verification	Verify HDL
	Testcases
Synthesis	
Place & Route	

Accuracy Estimation

Physical Performance

Talk Later

...

CONTENTS

MFCC MODEL AND ARCHITECTURE

Pre-emphasis

$$p[n] = s[n] - \boxed{0.97} \cdot s[n-1]$$

Energy

$$C[0] = \boxed{\log}\left(\sum_{n=0}^{N-1} s^2[n]\right)$$

Window

$$h[n] = p[n] \cdot \left\{\boxed{0.54} - \boxed{0.46} \cdot \boxed{cos\left(\frac{2\,\pi\,n}{N-1}\right)}\right\}$$

FFT

$$H[k] = \sum_{n=0}^{N-1} h(n) \cdot \boxed{e^{j\frac{2\pi\,n}{N}}}$$

Amplitude

$$|a + jb| = \max(|a|, |b|) + 1/4\min(|a|, |b|)$$

Mel

$$X[l] = \boxed{\log}\left(\sum_{k=k_{ll}}^{k_{lu}} |H[k]| \cdot \boxed{W_l[k]}\right)$$

Cepstral

$$C[m] = \sum_{l=1}^{L} X[l]\boxed{cos\left(\frac{\pi\,m\,(l-0.5)}{L}\right)}$$

Delta

$$C'_n = 2(C_{n+2} - C_{n-2}) + C_{n+1} - C_{n-1}$$

Delta - Delta

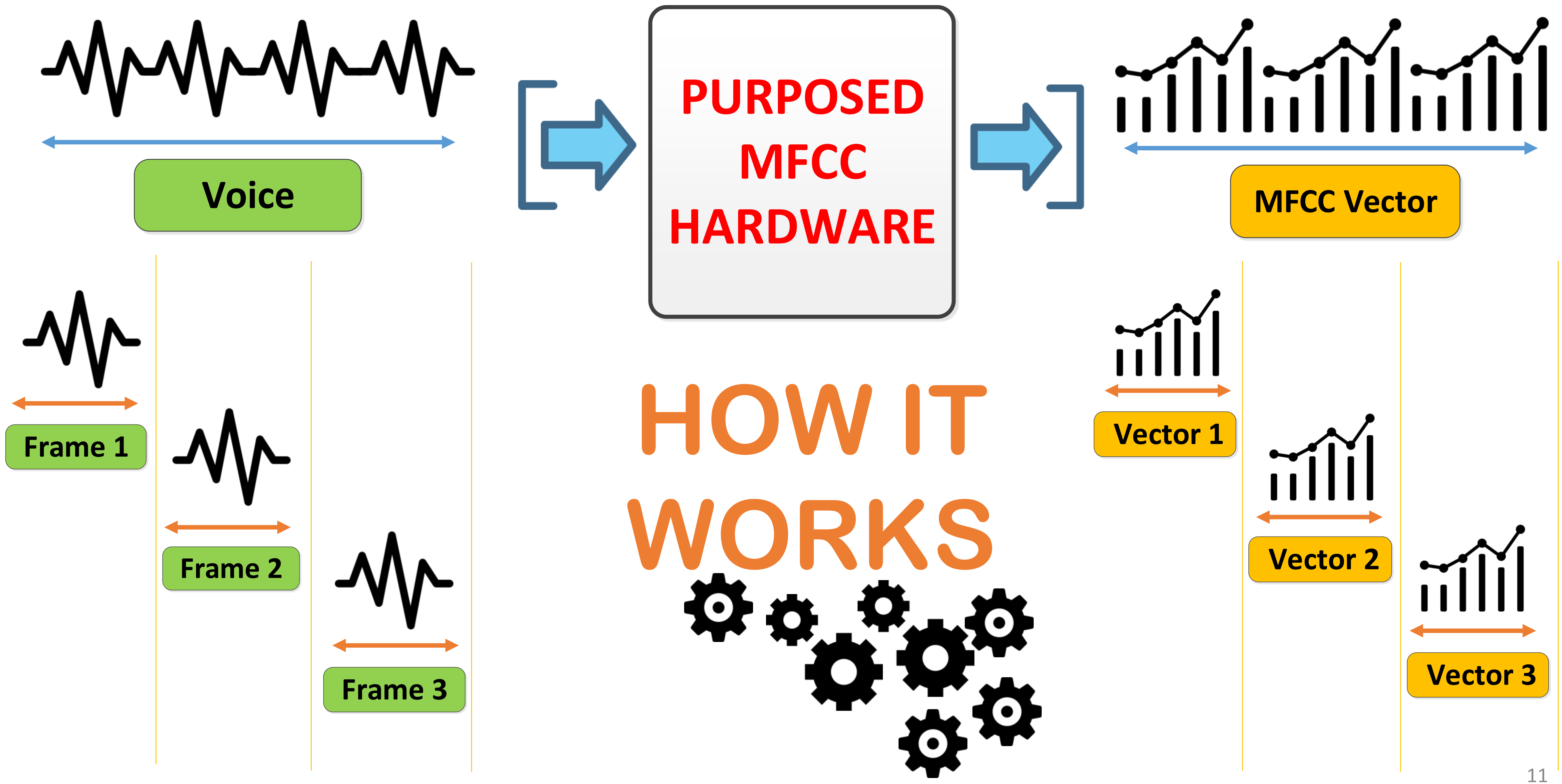
$$C''_n = 2(C'_{n+2} - C'_{n-2}) + C'_{n+1} - C'_{n-1}$$

Real Number Problem

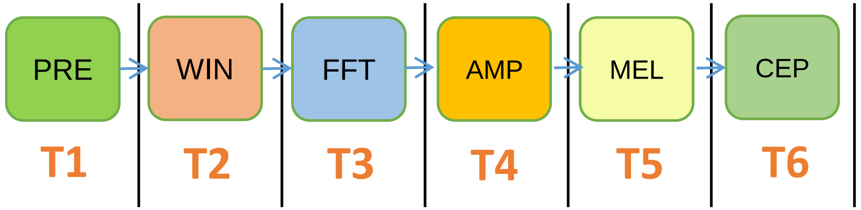
IEEE 754 Floating Point 32bits Standard

31	30:23	22:0
Sign bit (1 bit)	Exponent bits (8 bits)	Mantissa bits (23 bits)

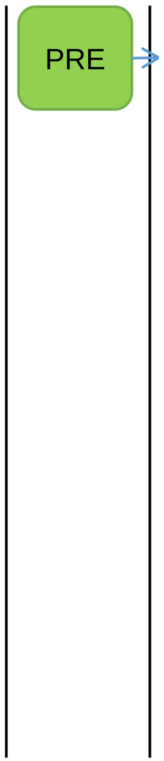
MFCC MODEL AND ARCHITECTURE

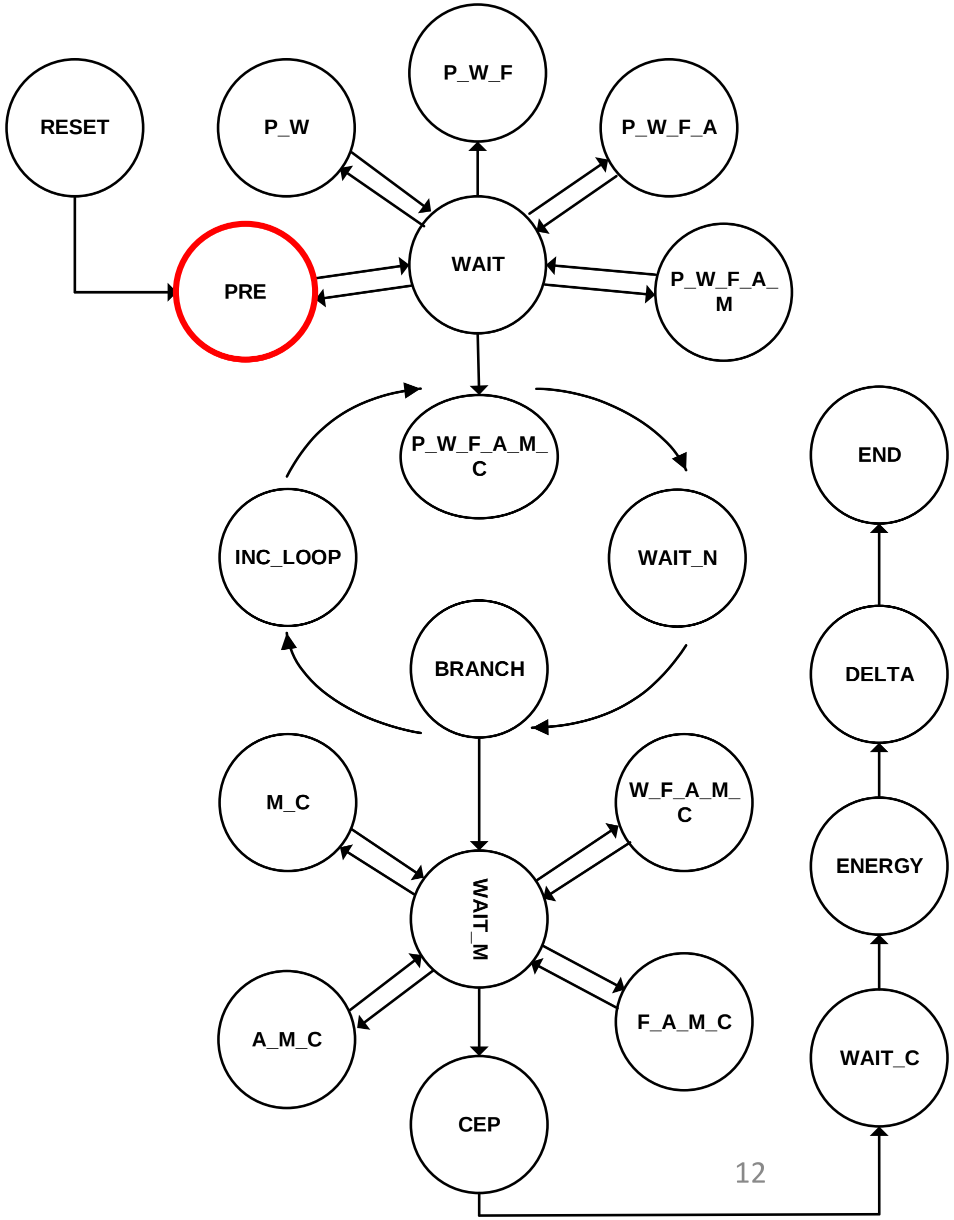


MFCC MODEL AND ARCHITECTURE

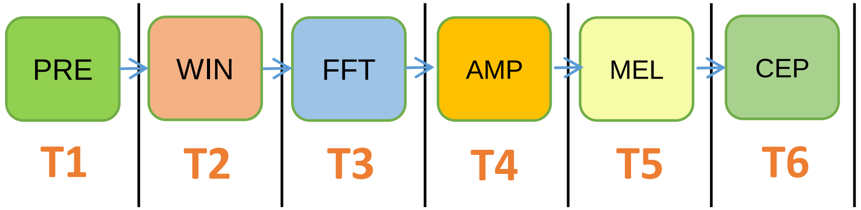


Without pipeline $T_{serial} = 6n \times T_{stage}$

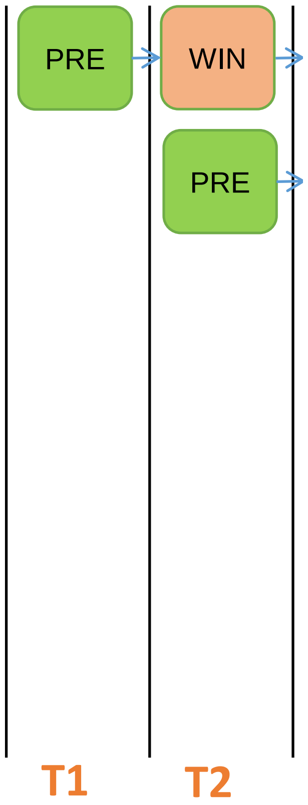


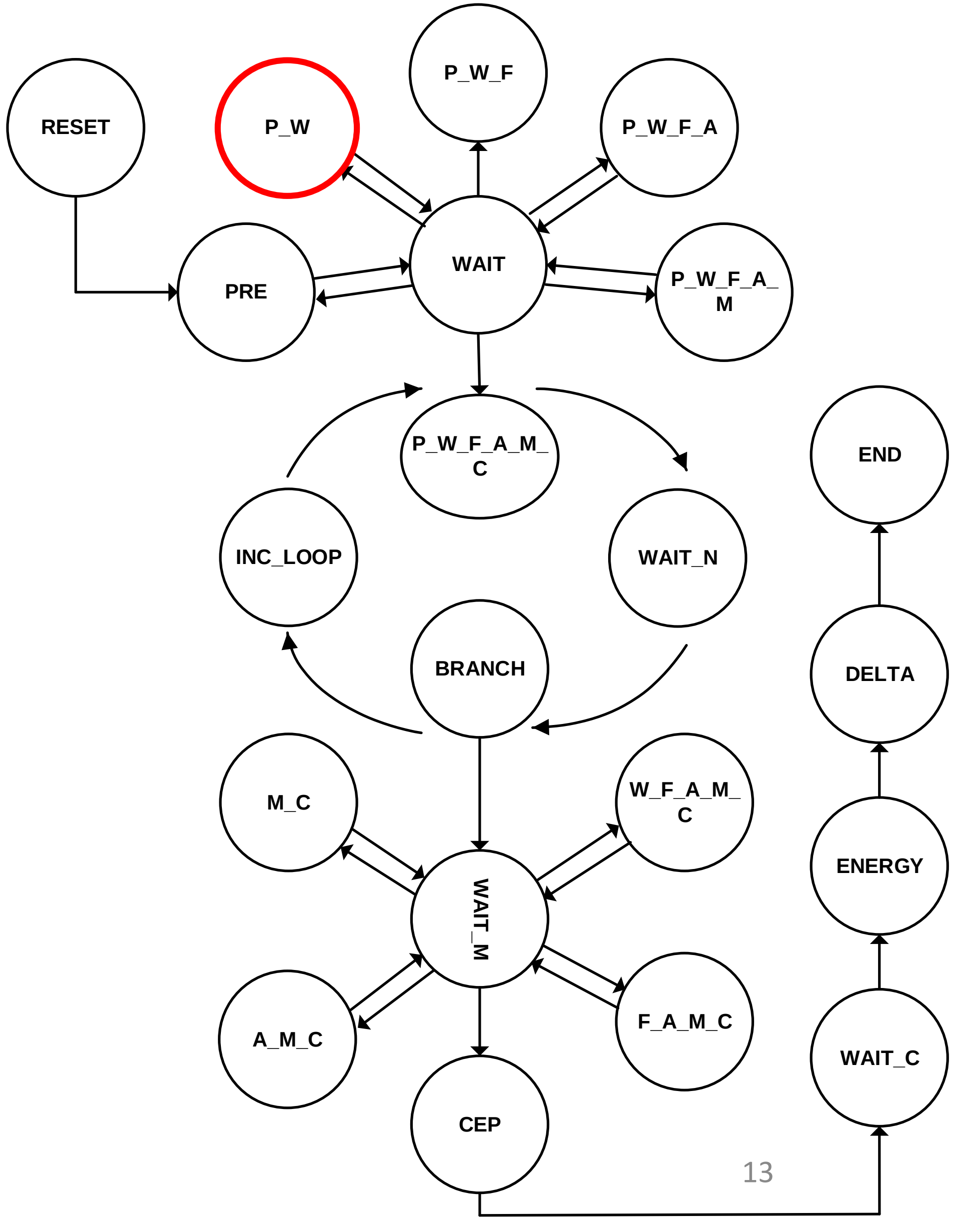


MFCC MODEL AND ARCHITECTURE

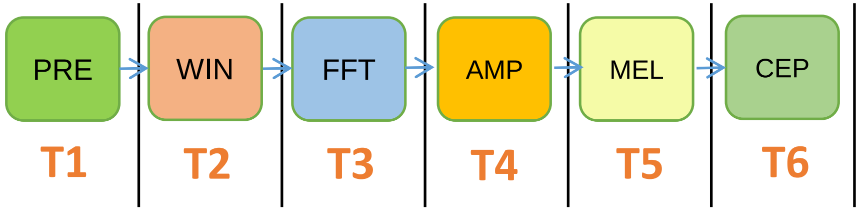


Without pipeline $T_{serial} = 6n \times T_{stage}$

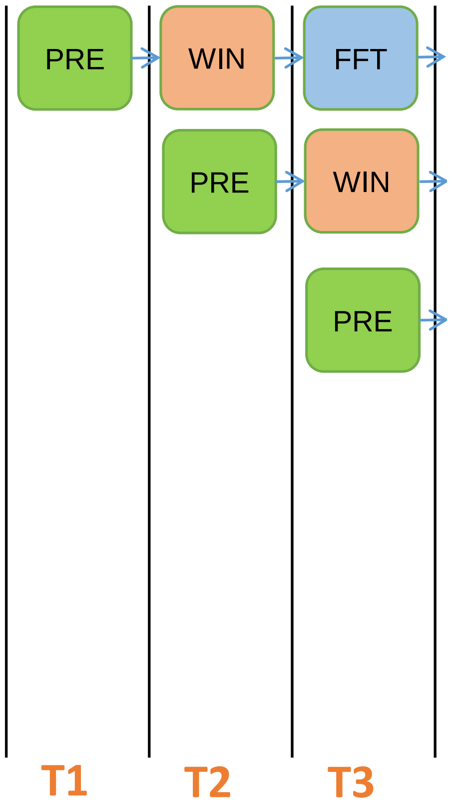


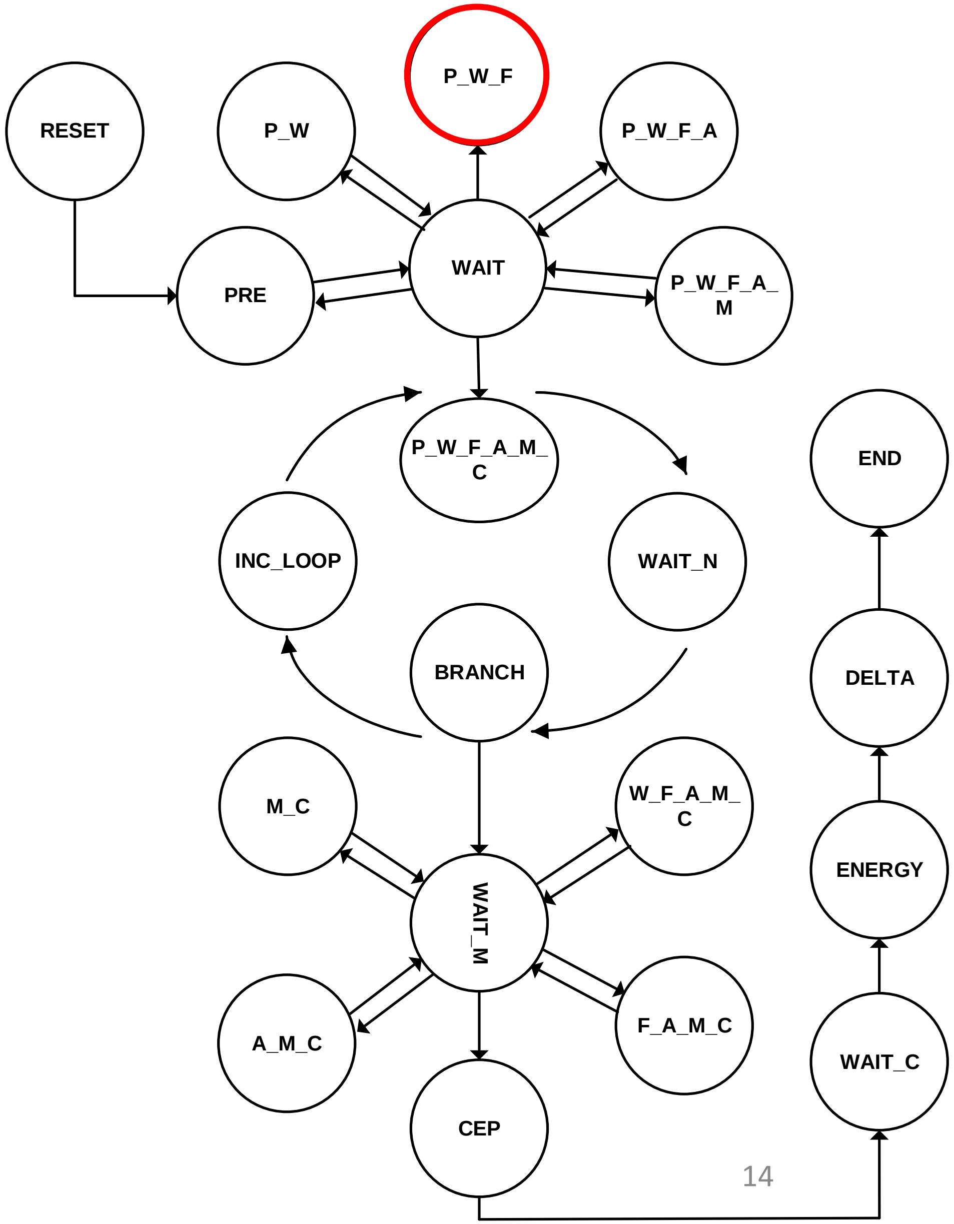


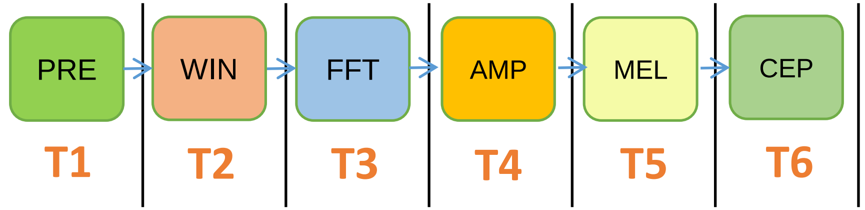
MFCC MODEL AND ARCHITECTURE



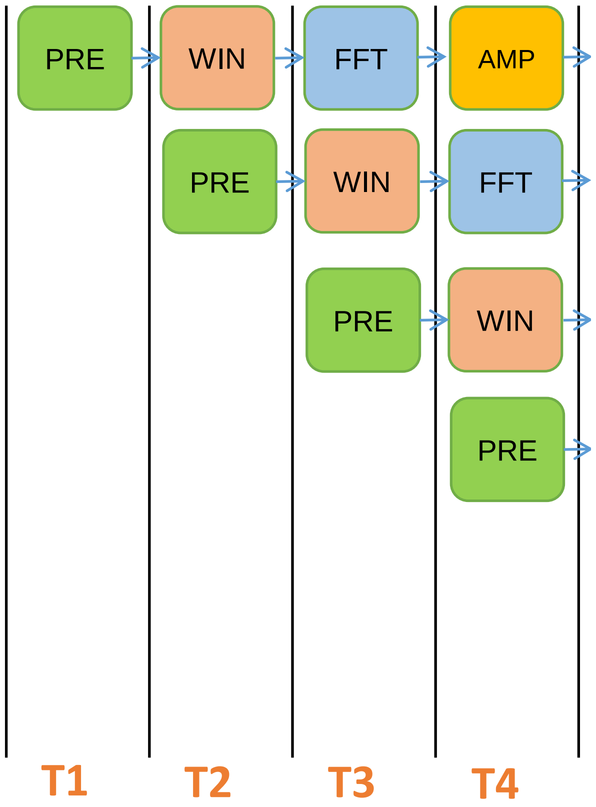
Without pipeline $T_{serial} = 6n \times T_{stage}$

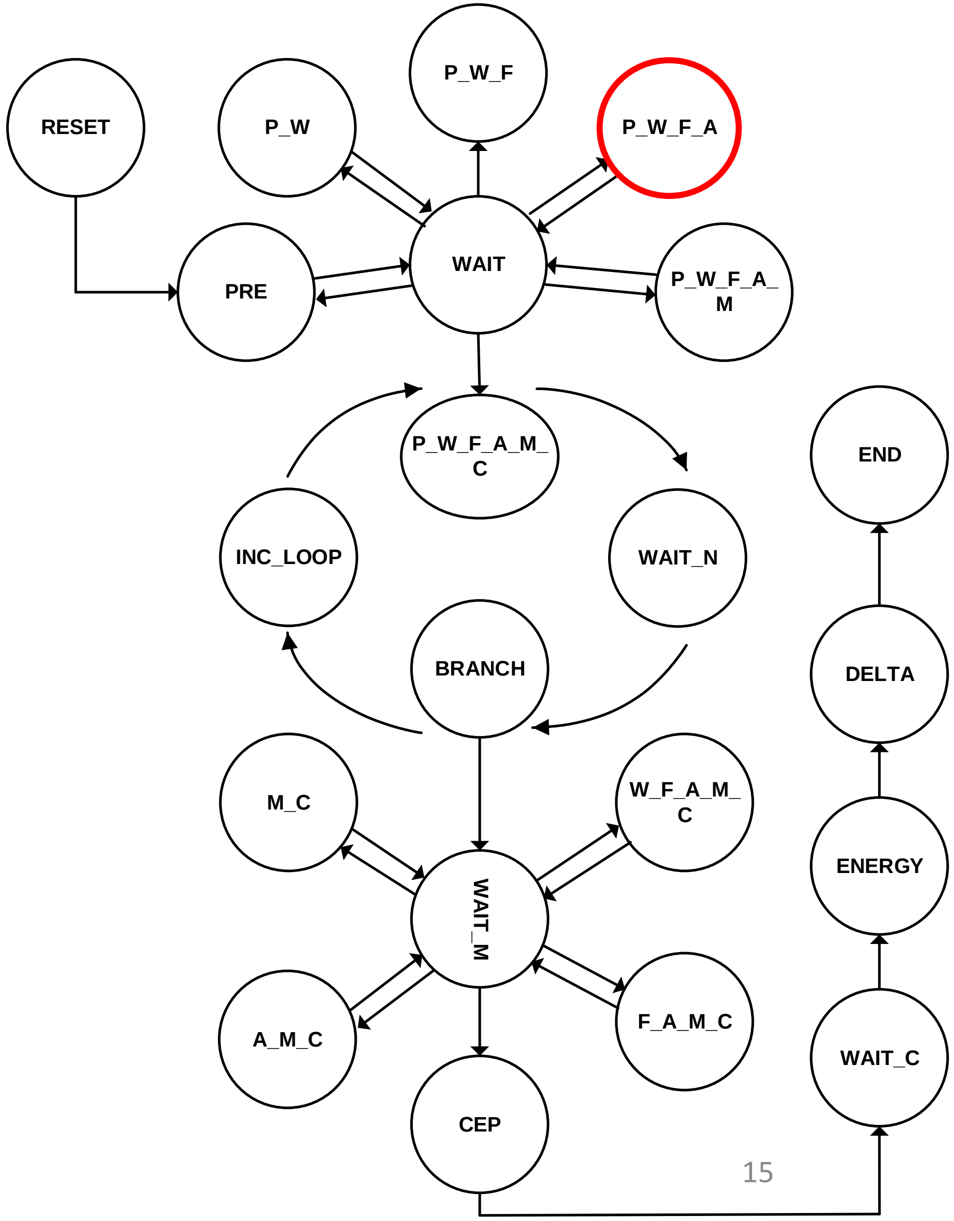




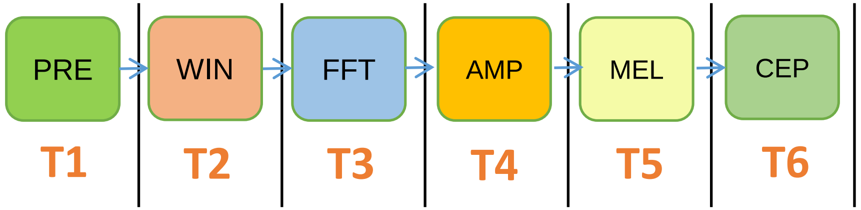


Without pipeline $T_{serial} = 6n \times T_{stage}$

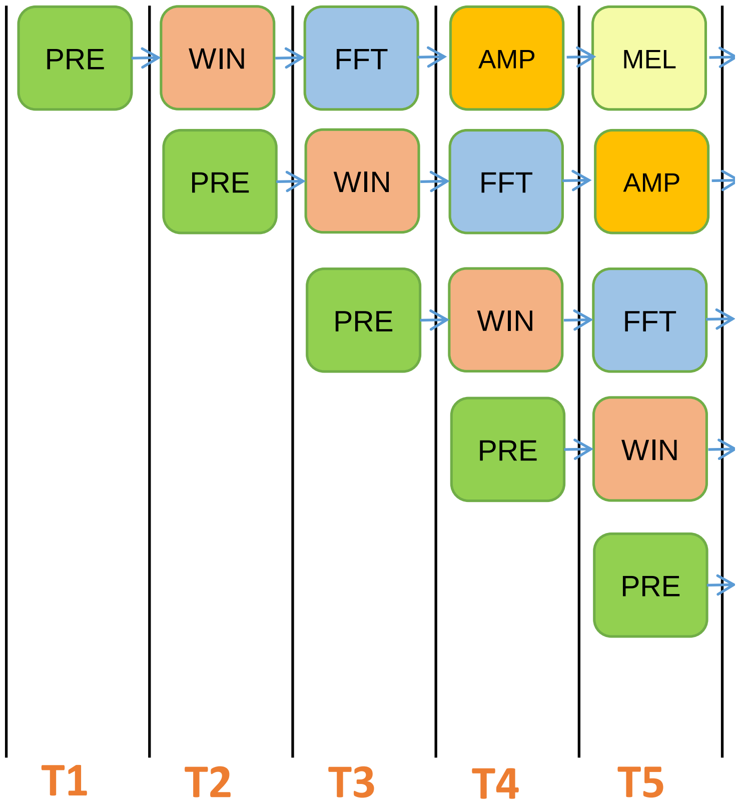


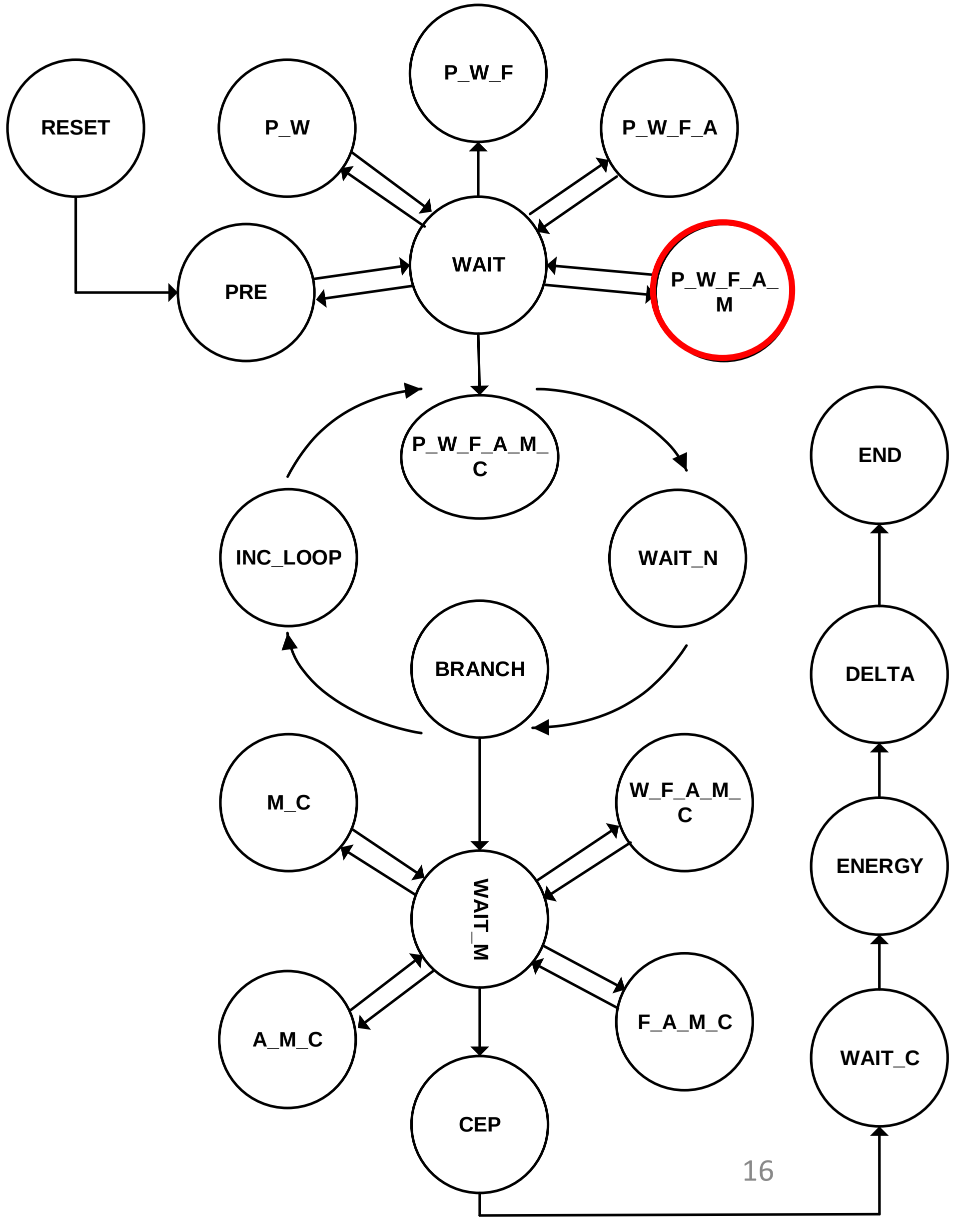


MFCC MODEL AND ARCHITECTURE

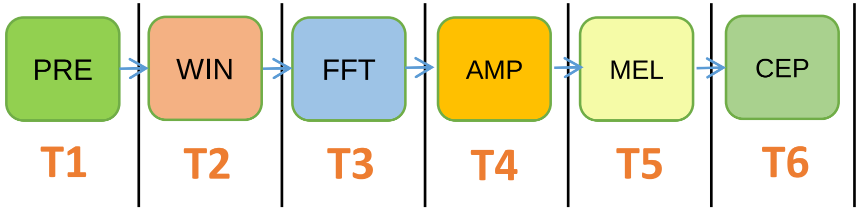


Without pipeline $T_{serial} = 6n \times T_{stage}$



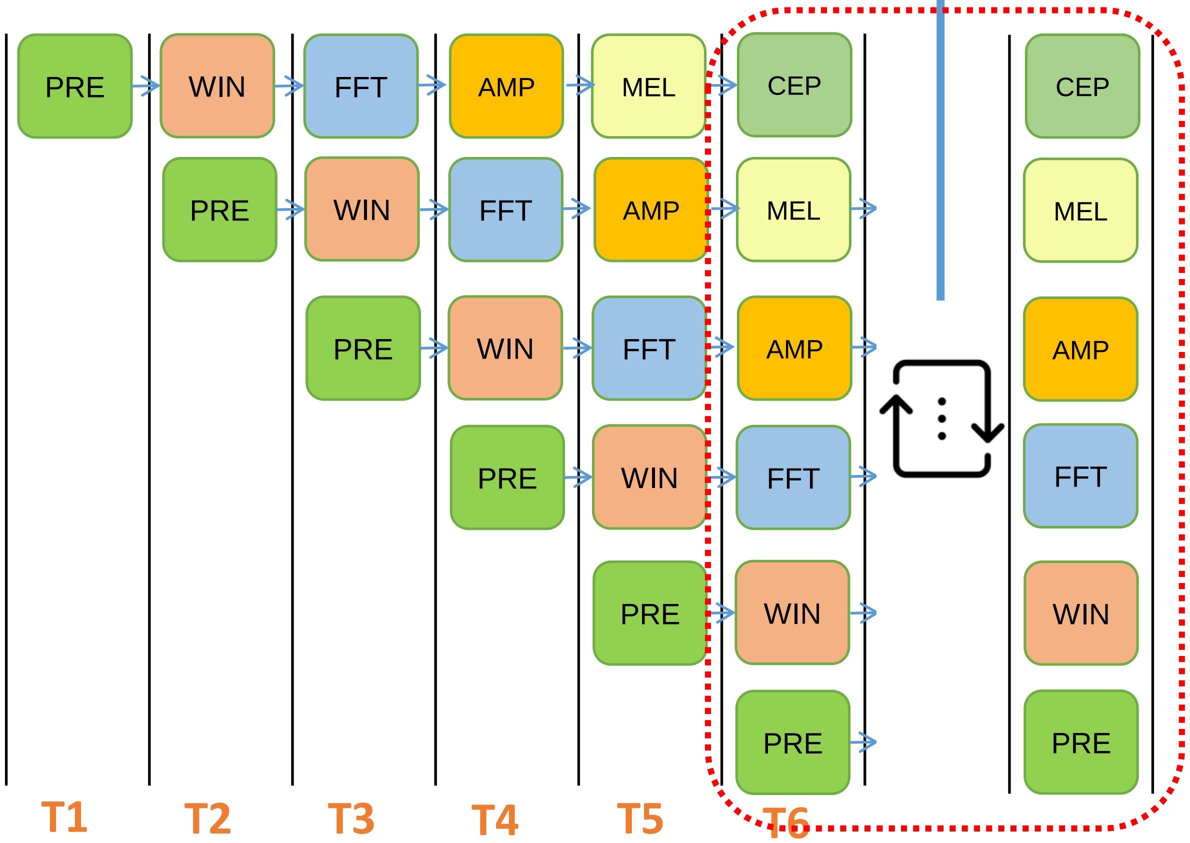


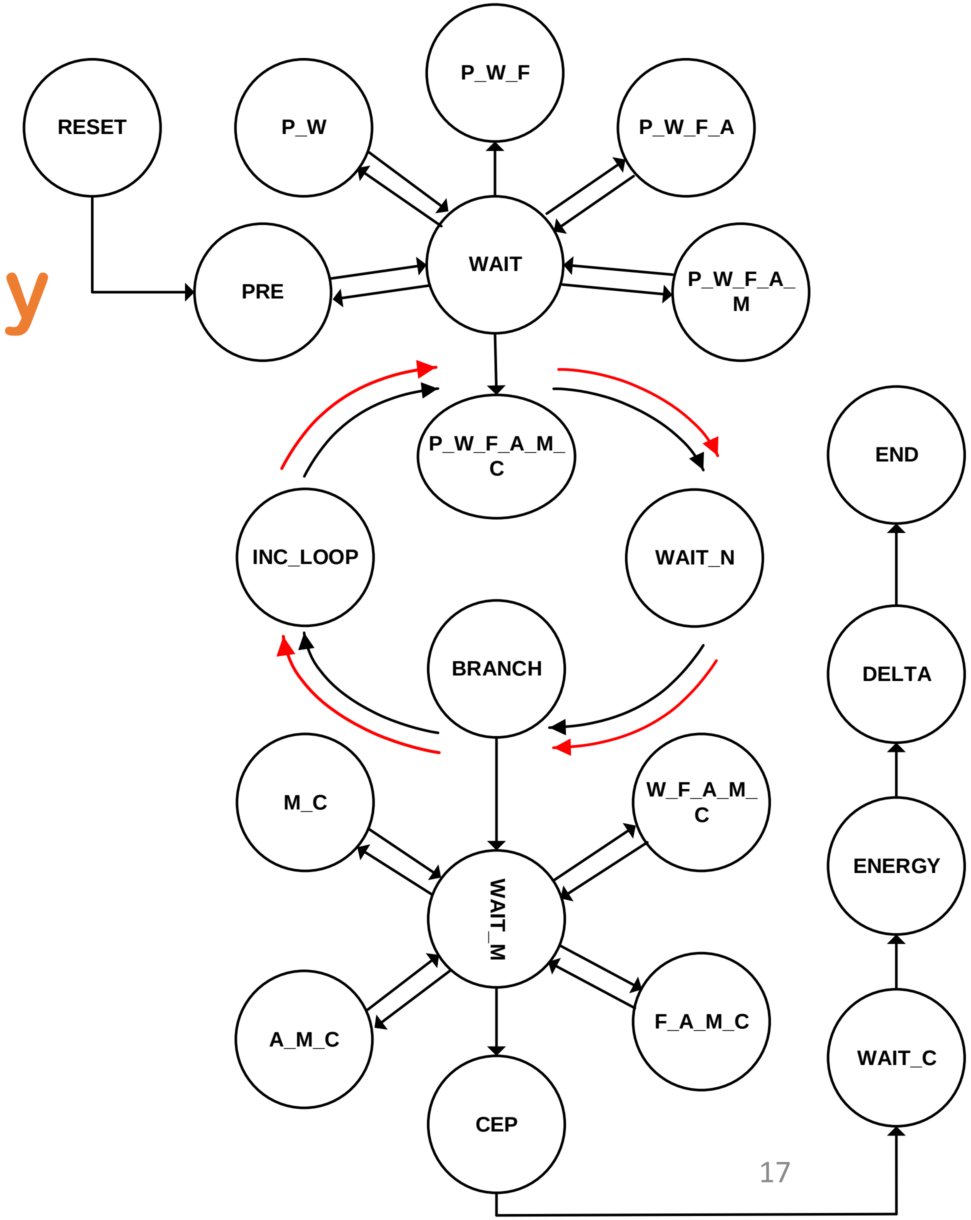
MFCC MODEL AND ARCHITECTURE



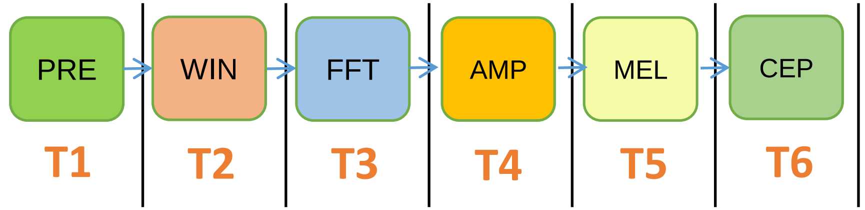
Without pipeline $T_{serial} = 6n \times T_{stage}$

Simultaneously

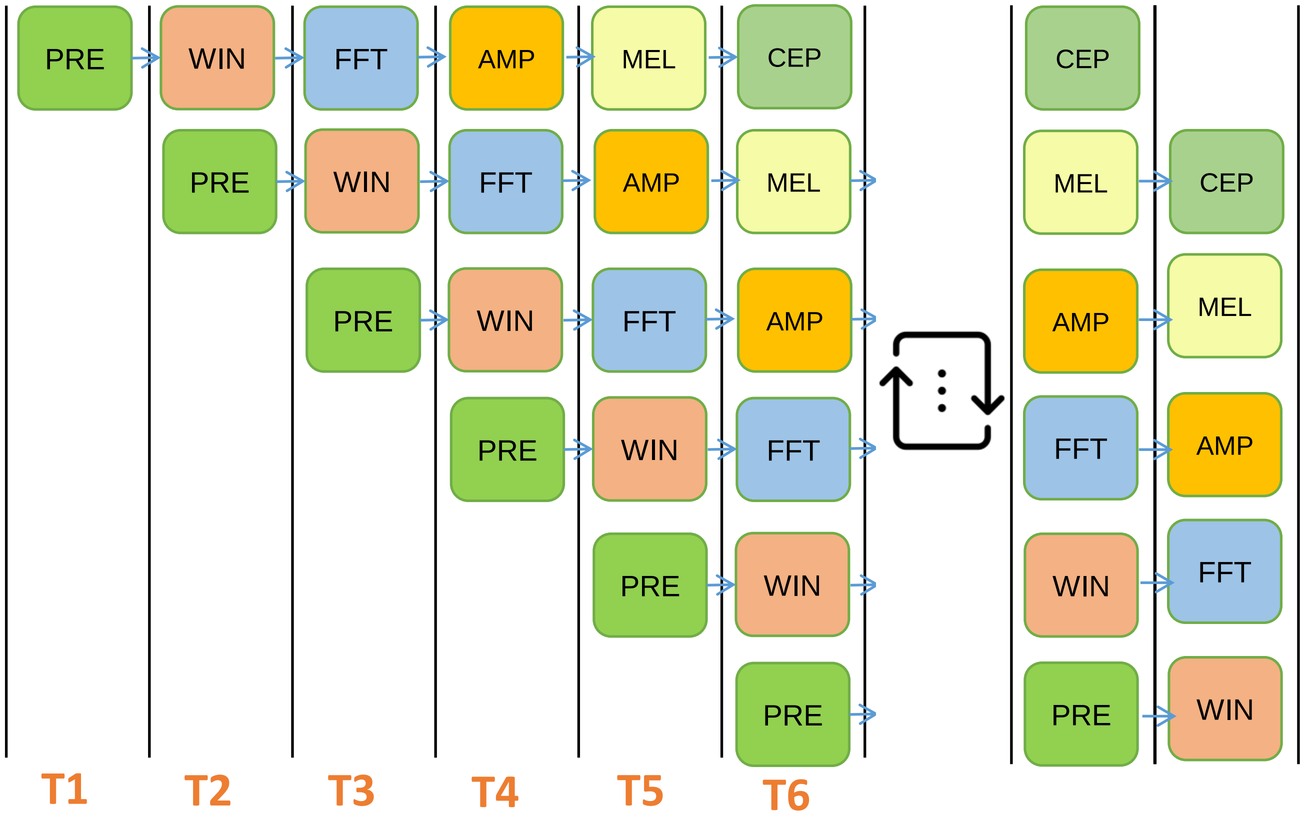


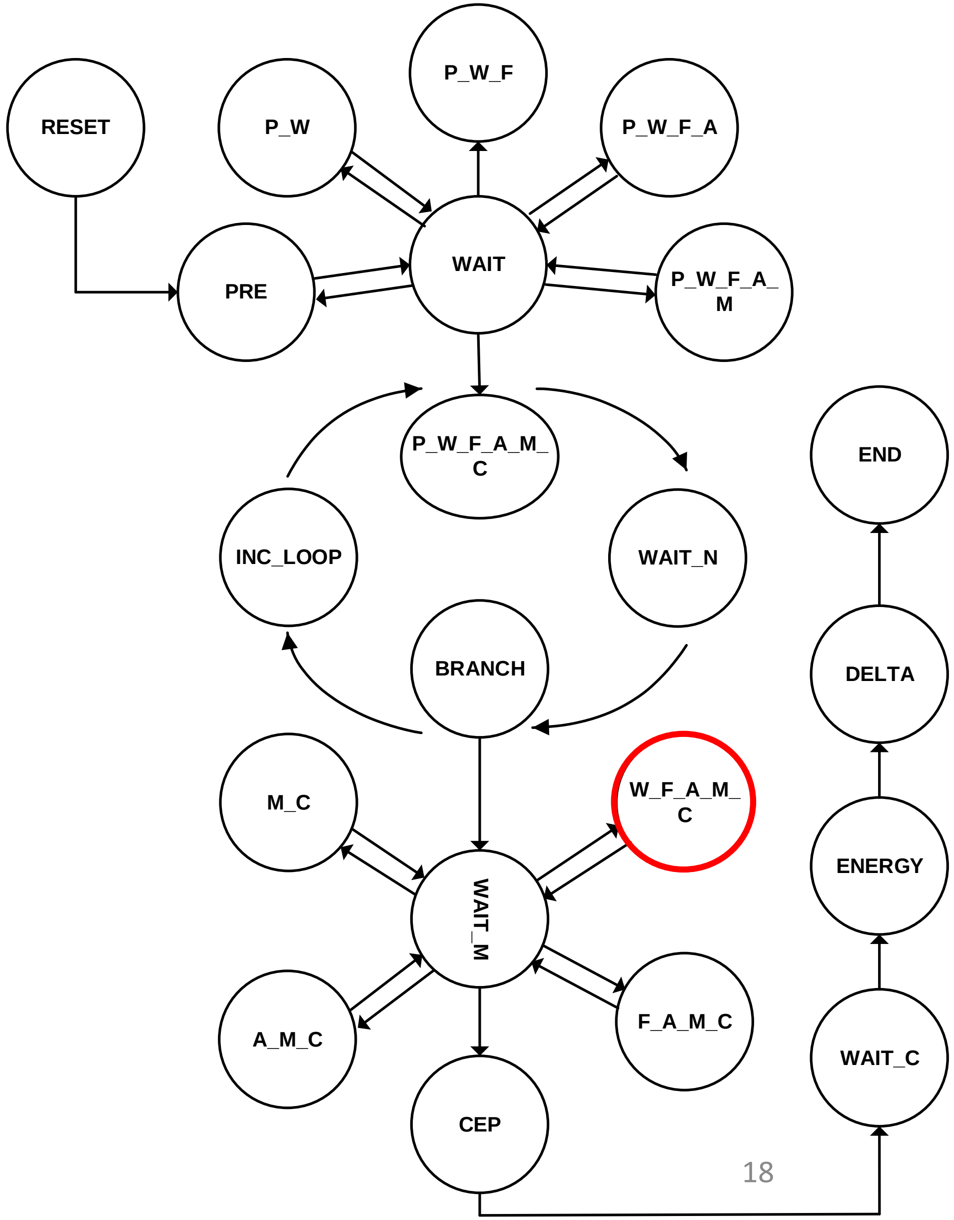


MFCC MODEL AND ARCHITECTURE

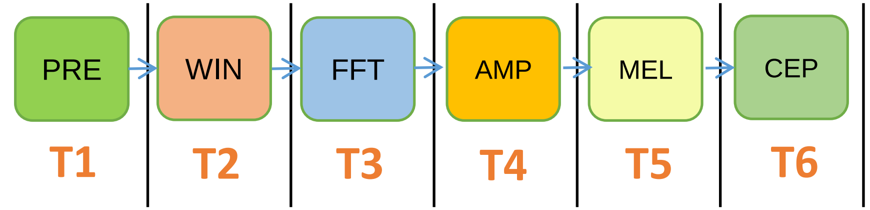


Without pipeline $T_{serial} = 6n \times T_{stage}$

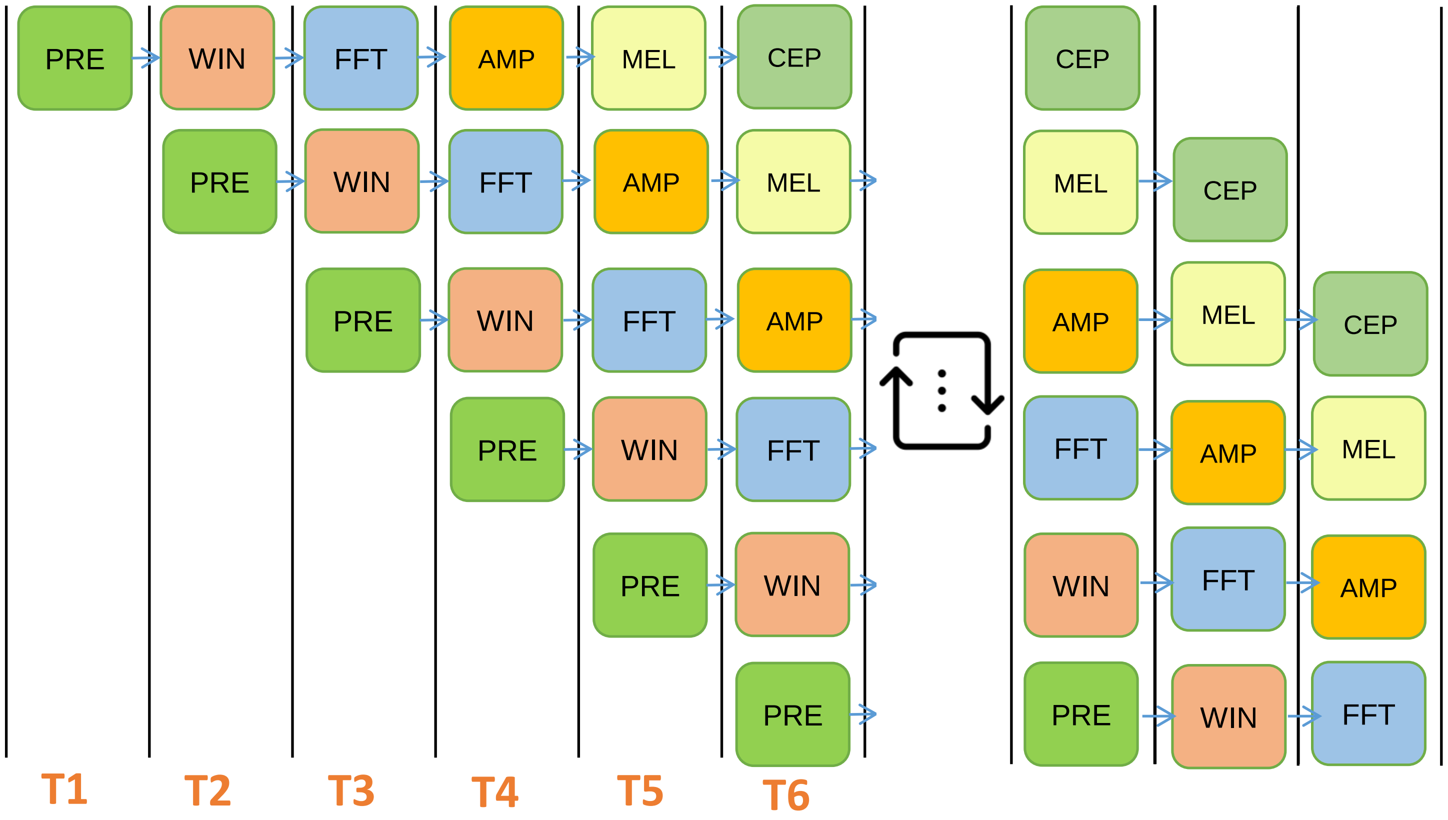


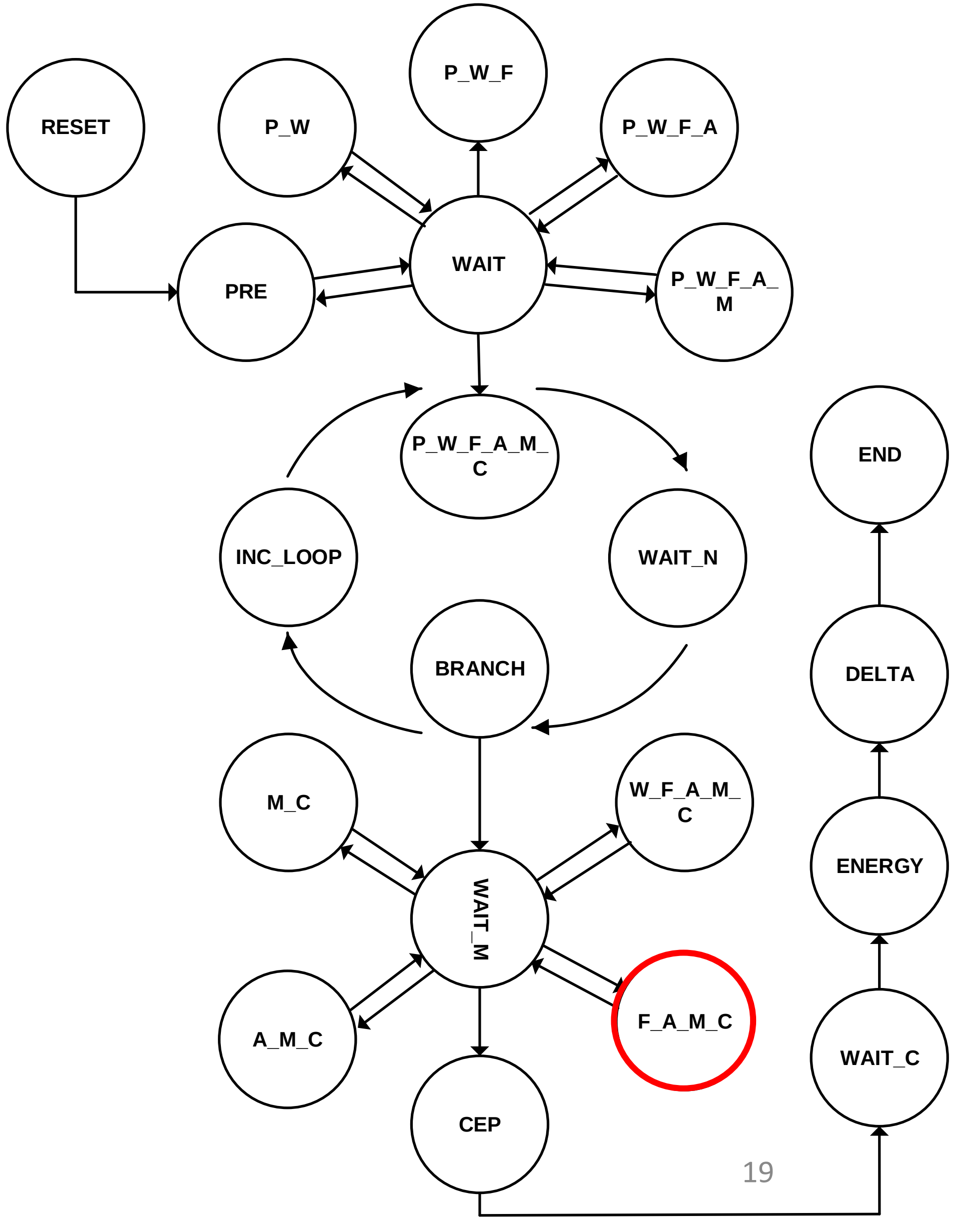


MFCC MODEL AND ARCHITECTURE

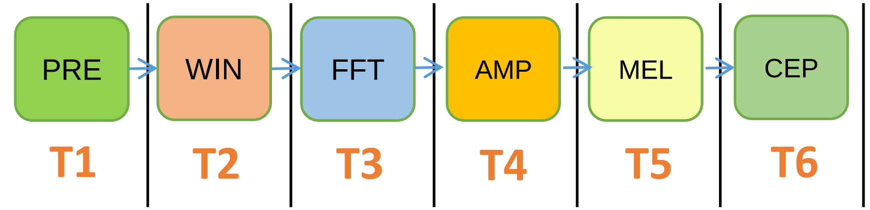


Without pipeline $T_{serial} = 6n \times T_{stage}$

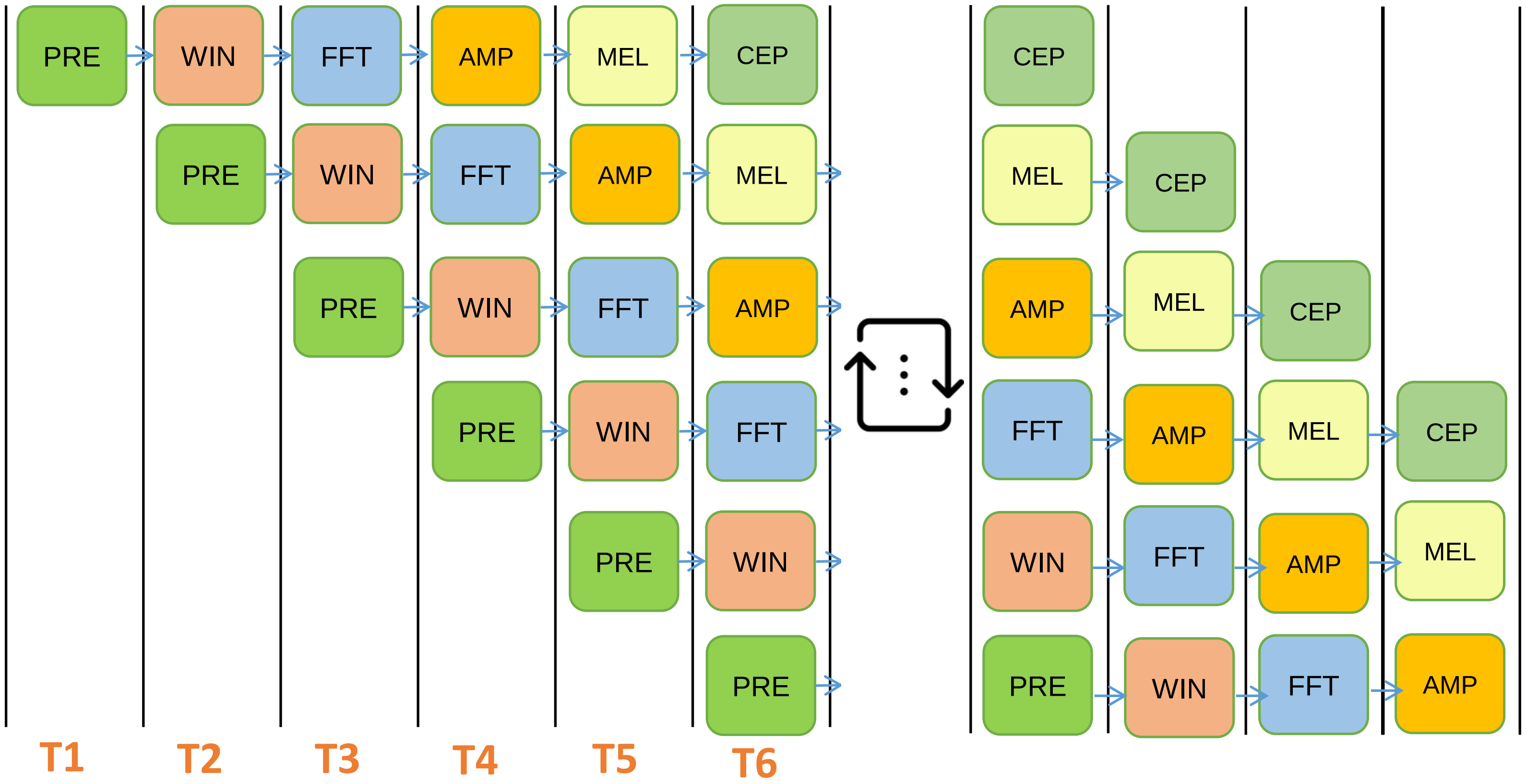


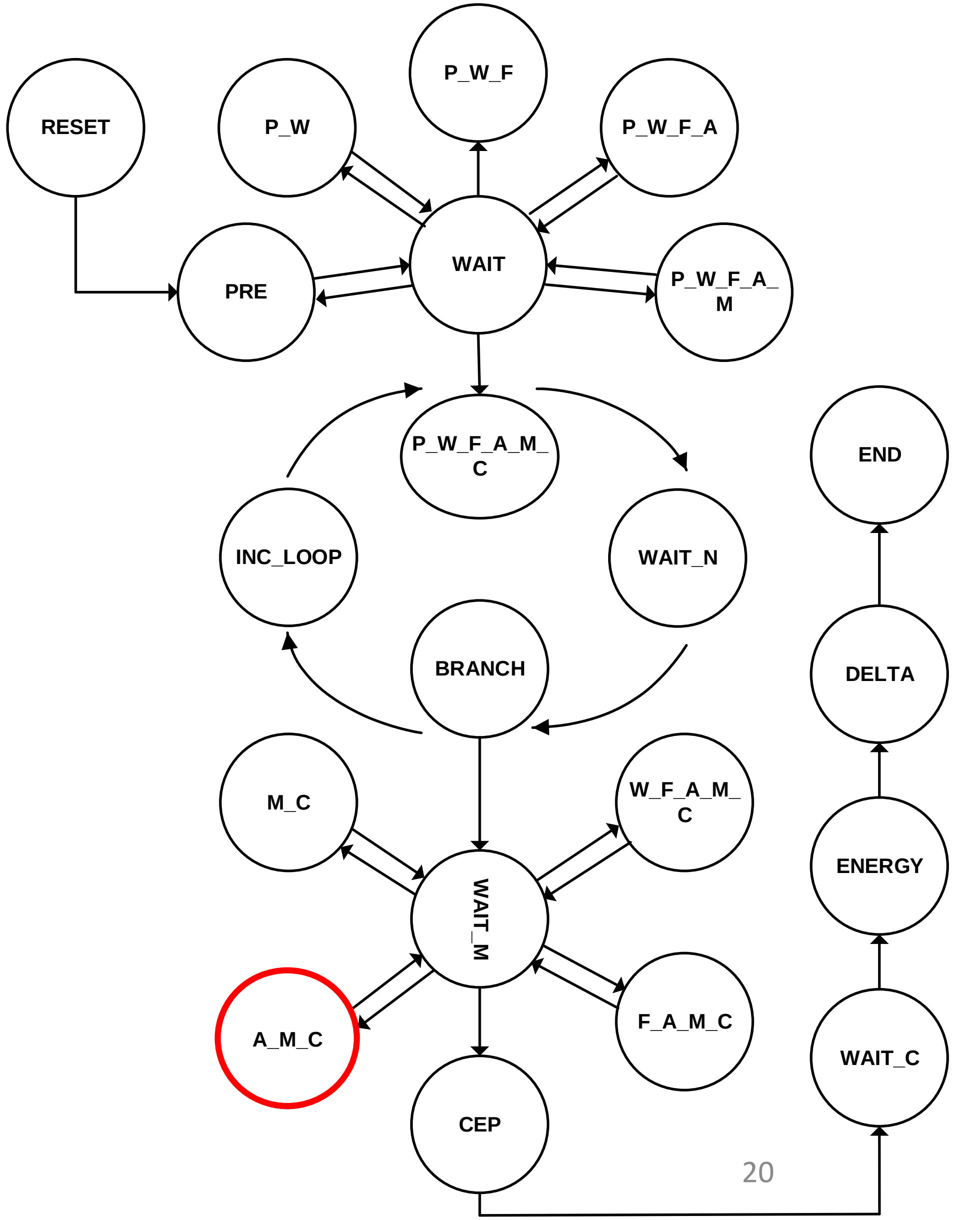


MFCC MODEL AND ARCHITECTURE

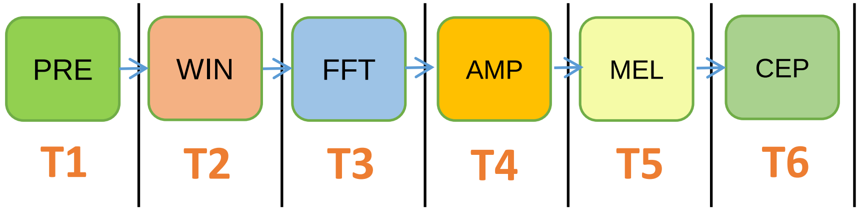


Without pipeline $T_{serial} = 6n \times T_{stage}$

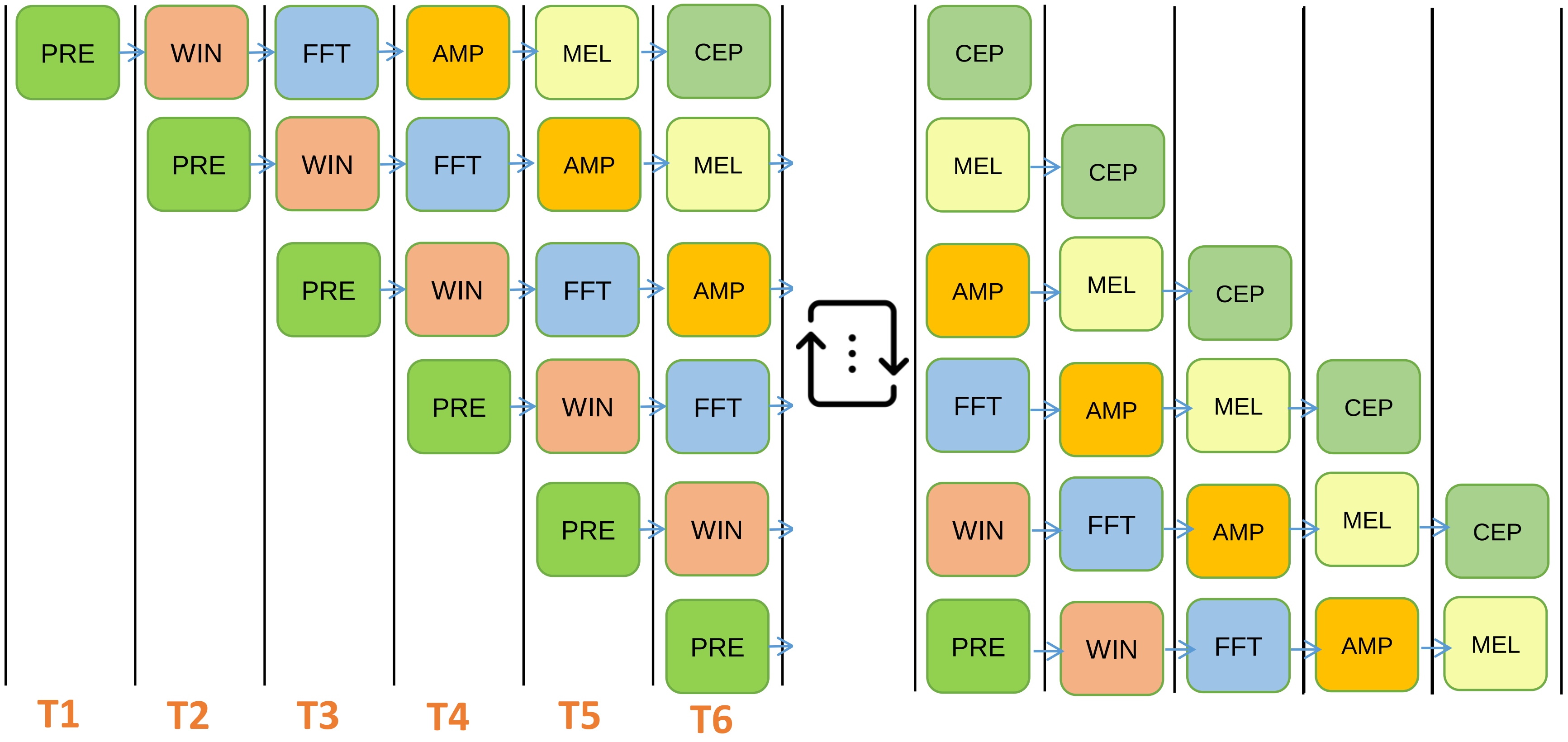


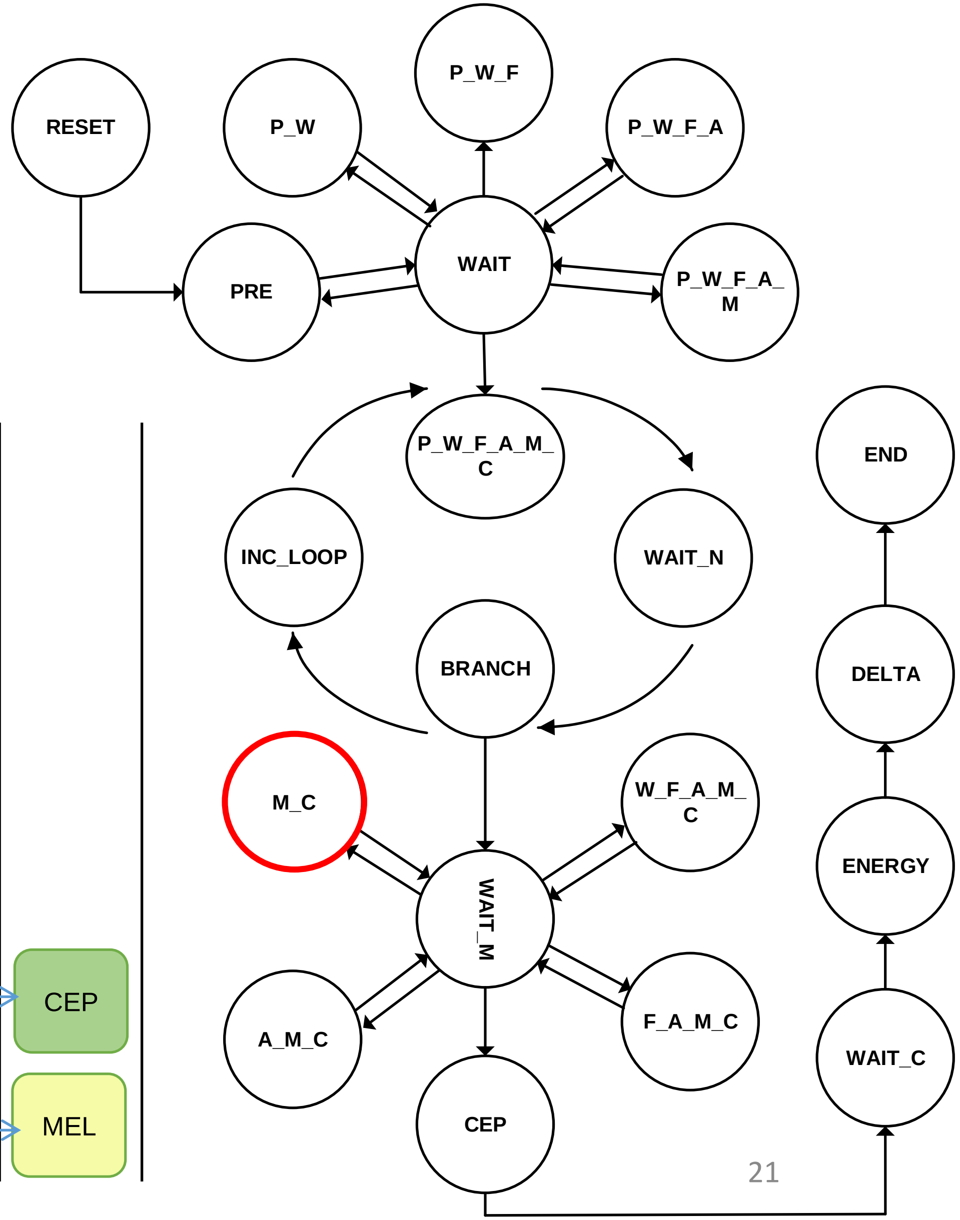


MFCC MODEL AND ARCHITECTURE

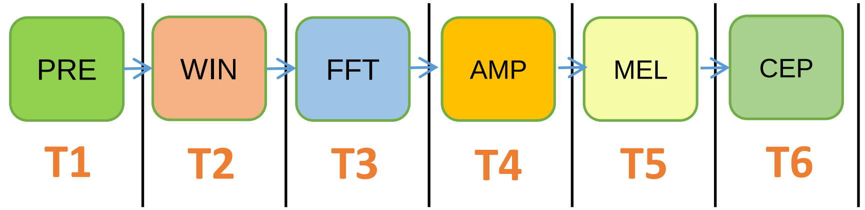


Without pipeline $T_{serial} = 6n \times T_{stage}$

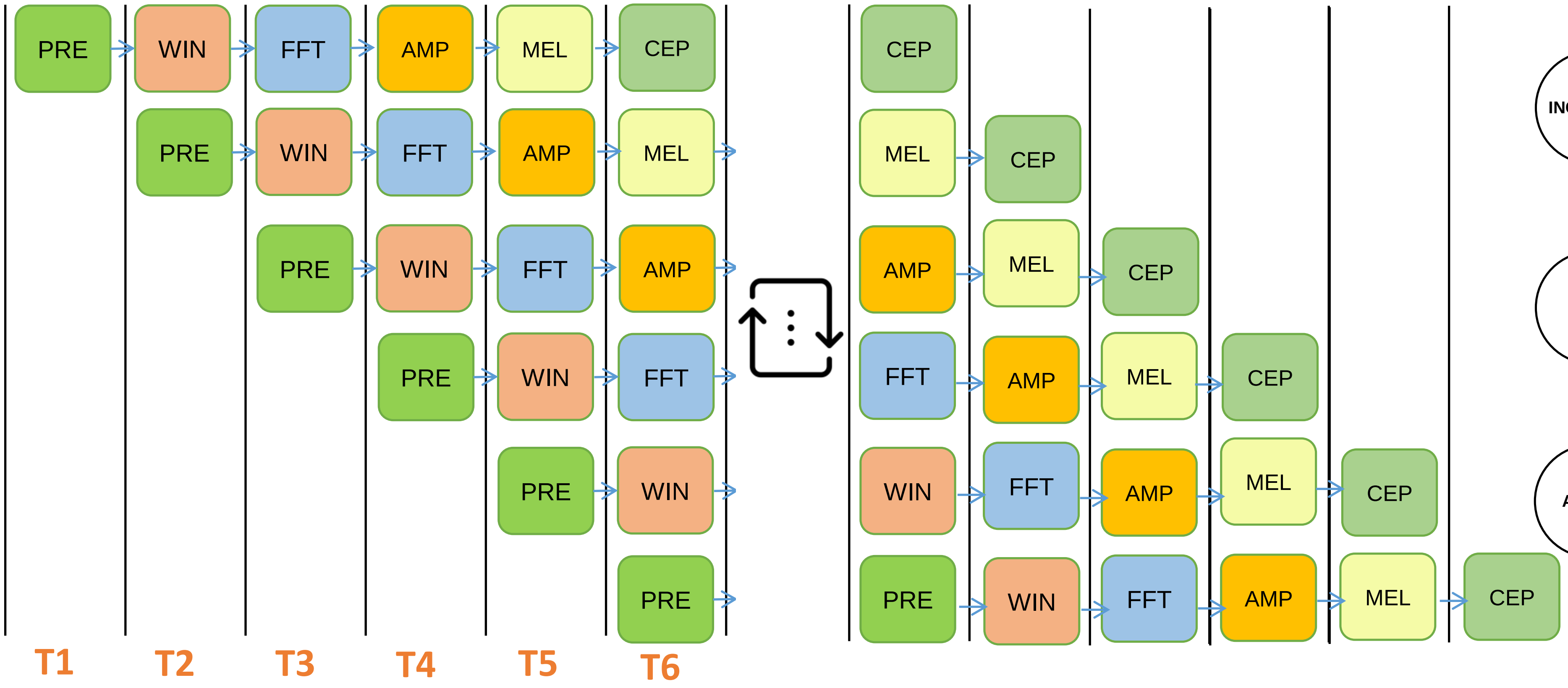


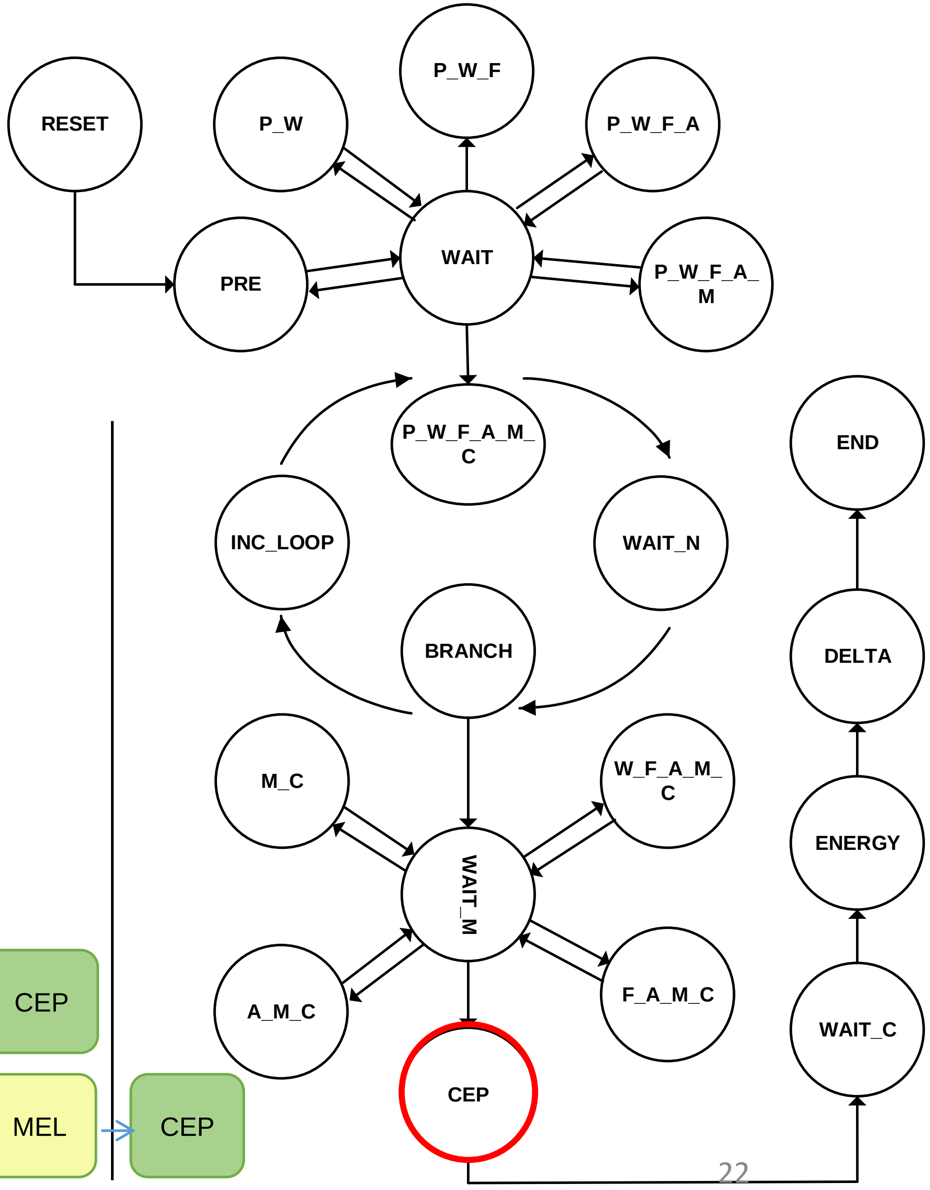


MFCC MODEL AND ARCHITECTURE

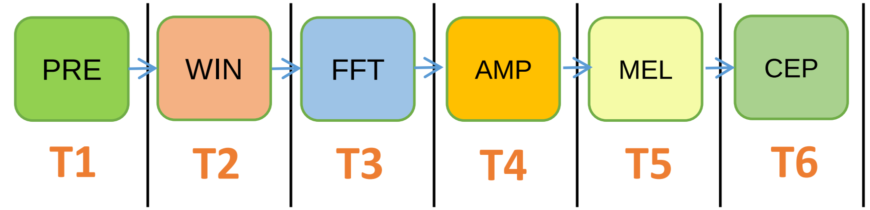


Without pipeline $T_{serial} = 6n \times T_{stage}$





MFCC MODEL AND ARCHITECTURE



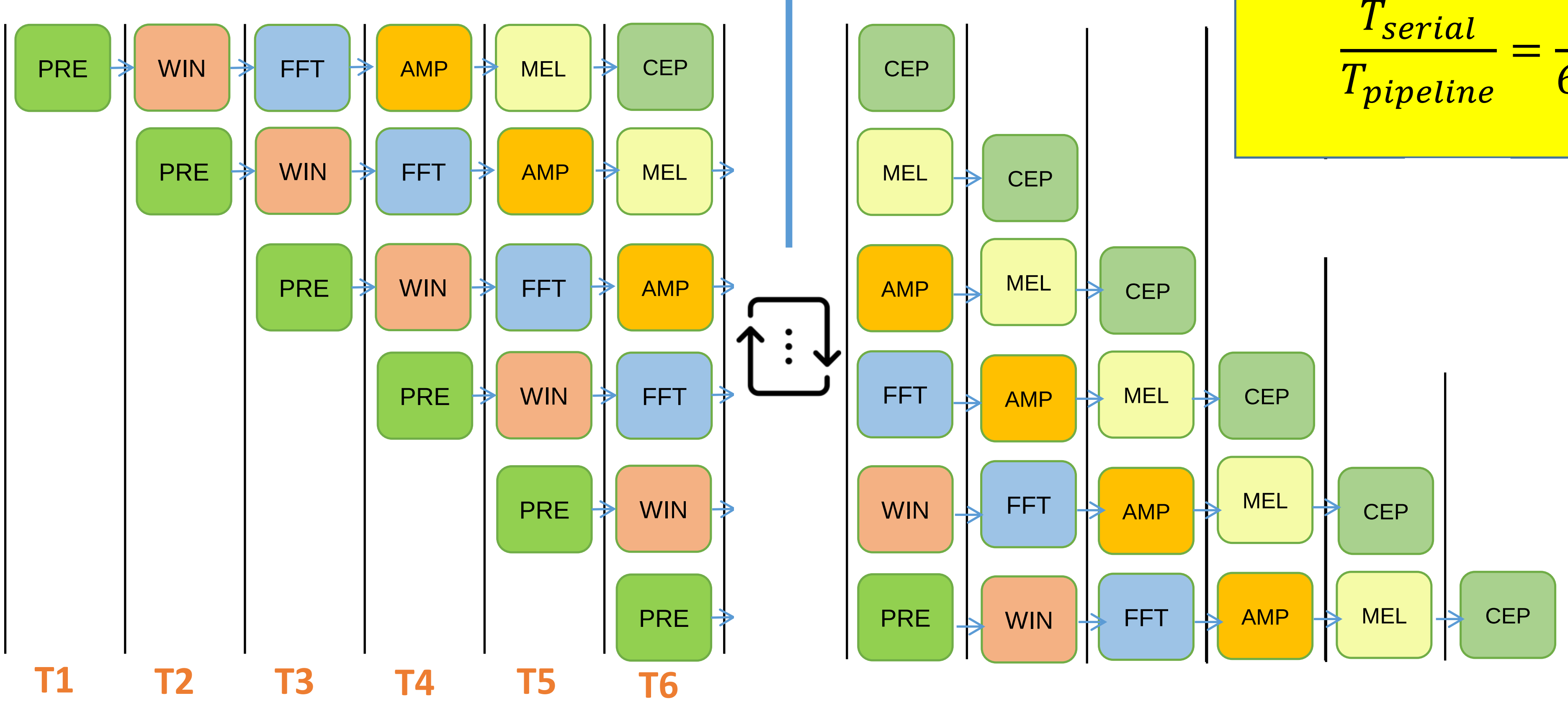
Without pipeline $T_{serial} = 6n \times T_{stage}$

n Loops

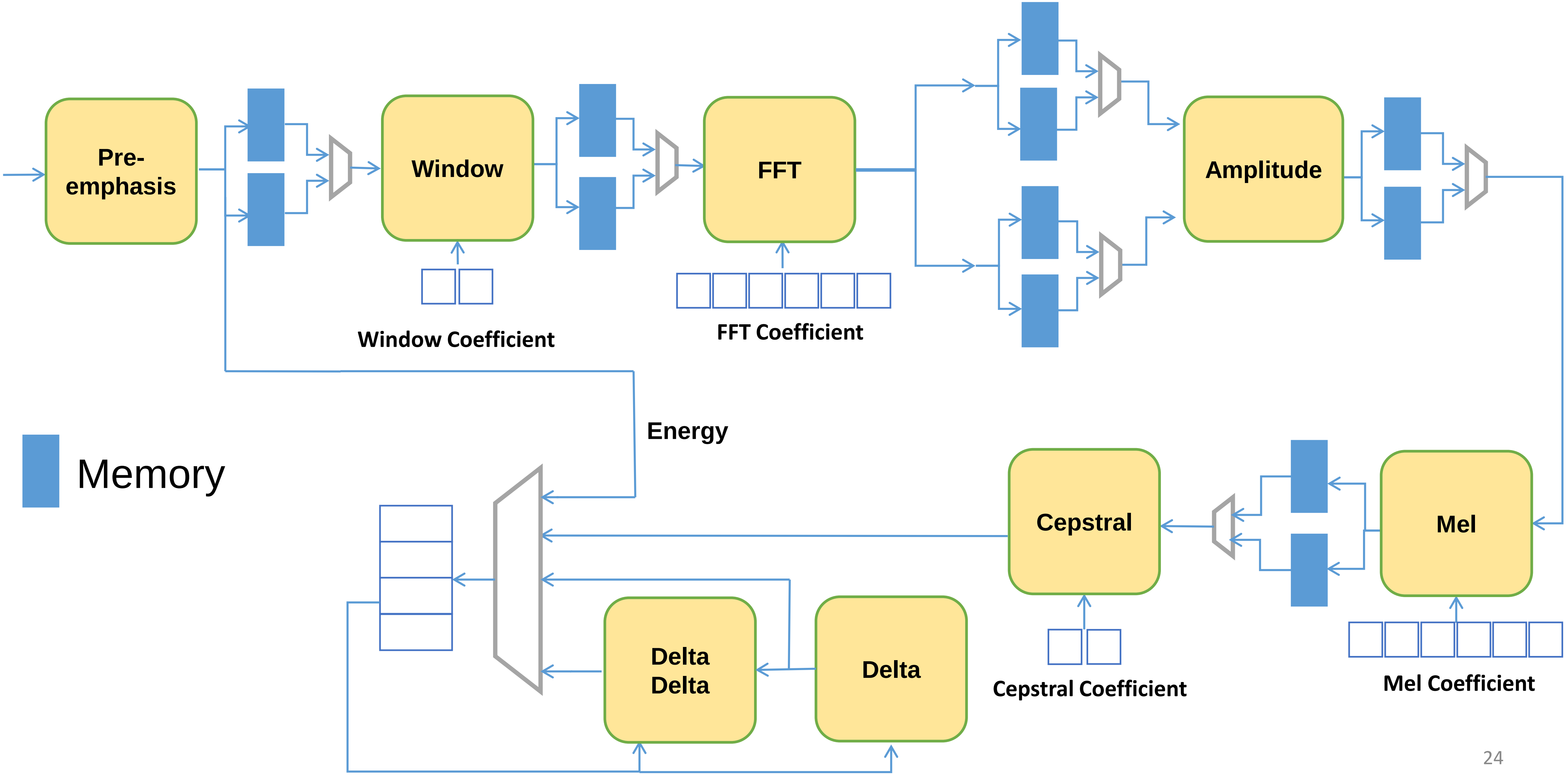
With pipeline technique:

$$T_{pipeline} = [6 + (n-1)] \times T_{stage}$$

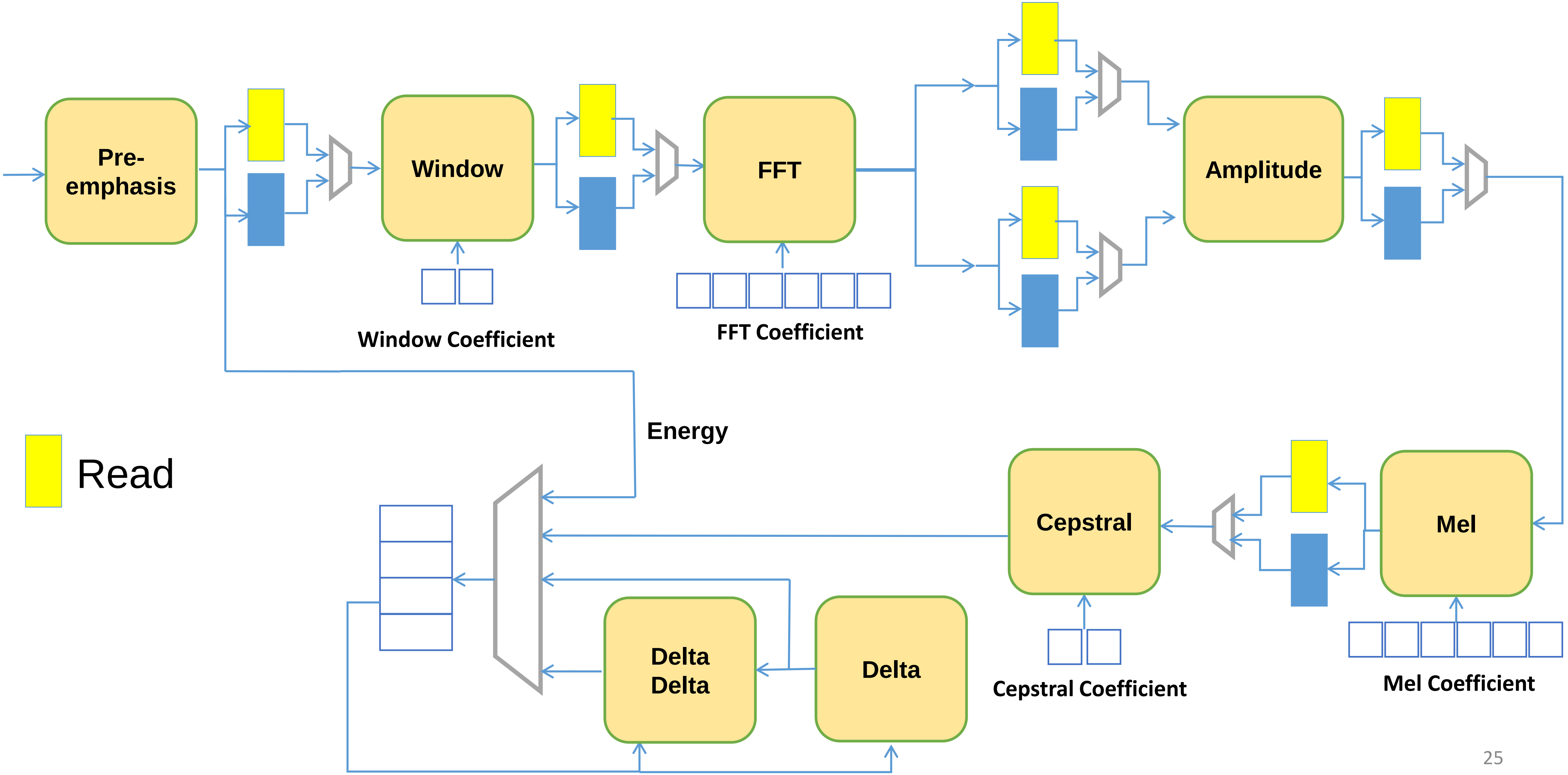
$$\frac{T_{serial}}{T_{pipeline}} = \frac{6n}{6+n-1} \rightarrow 6 \; when \; n \gg 6$$



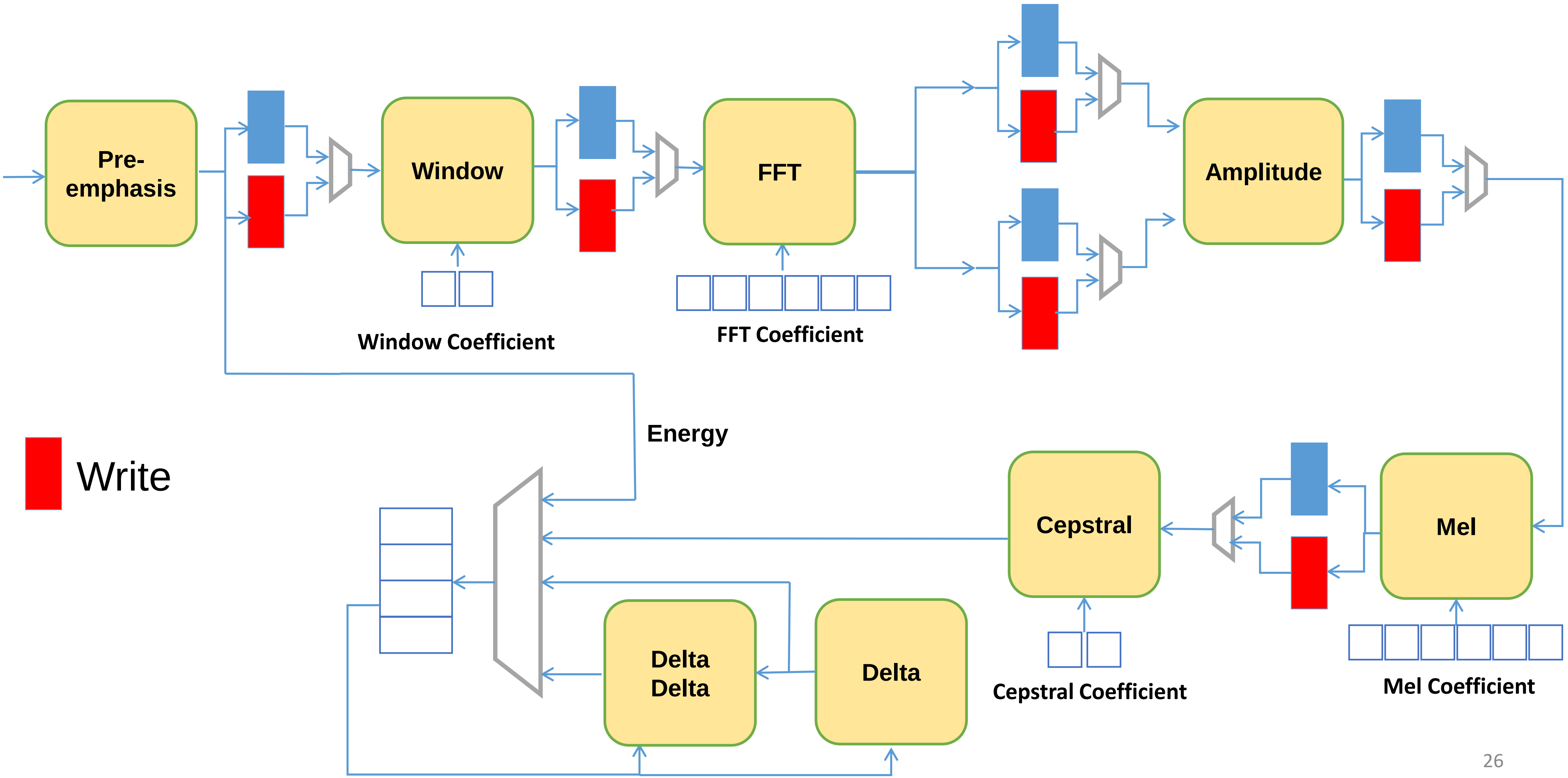
MFCC MODEL AND ARCHITECTURE



MFCC MODEL AND ARCHITECTURE



MFCC MODEL AND ARCHITECTURE



MFCC MODEL AND ARCHITECTURE

PRE

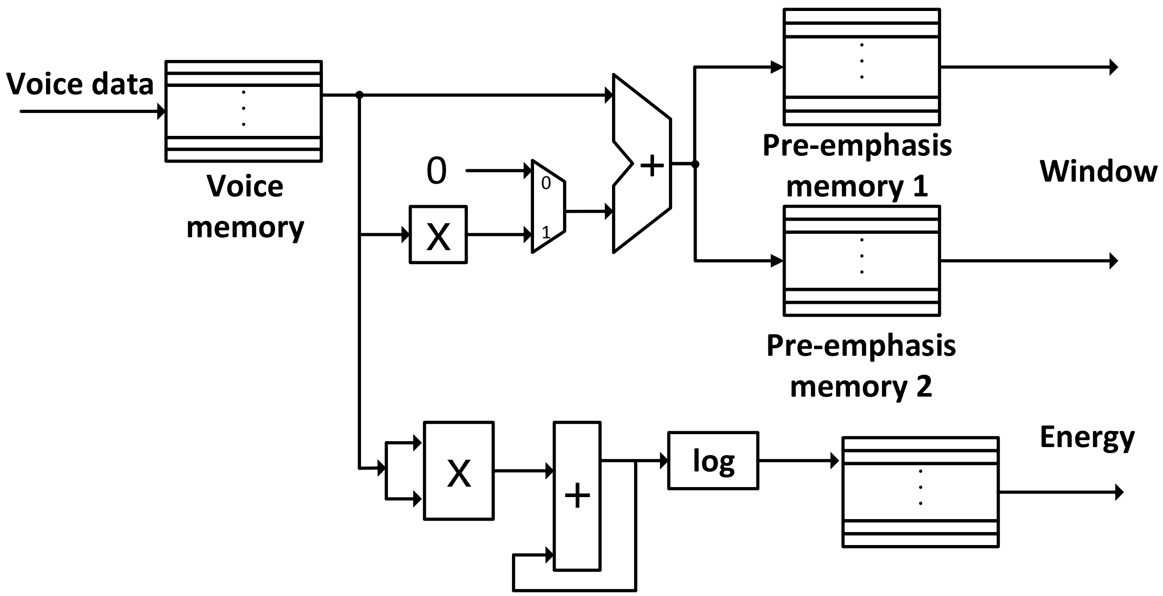
Both Pre-emphasis and Energy use inputs from the Voice data → Be implemented together

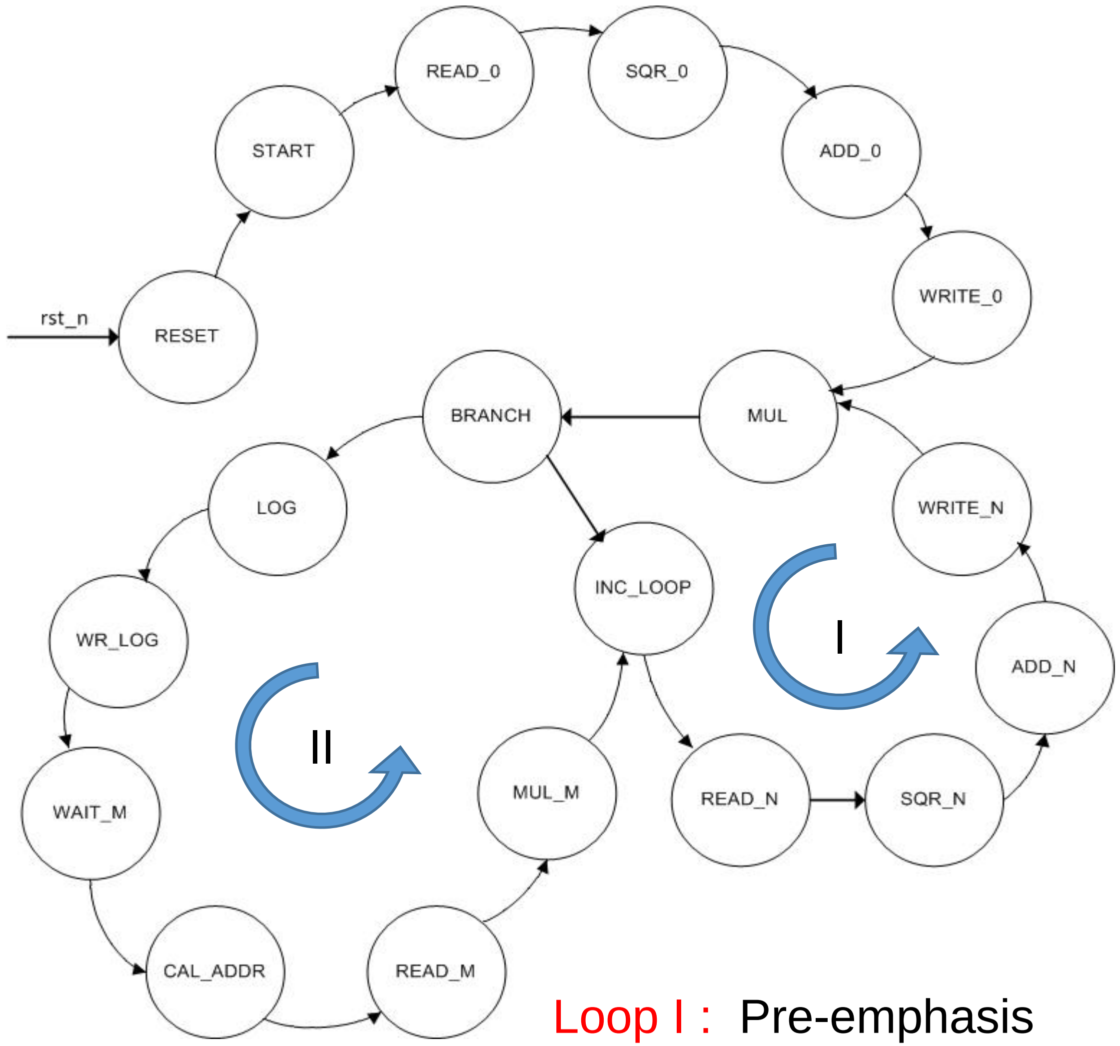
Pre-emphasis

$$p[n] = s[n] - 0.97 \cdot s[n-1]$$

Energy

$$C[0] = \log\left(\sum_{n=0}^{N-1} s^2[n]\right)$$





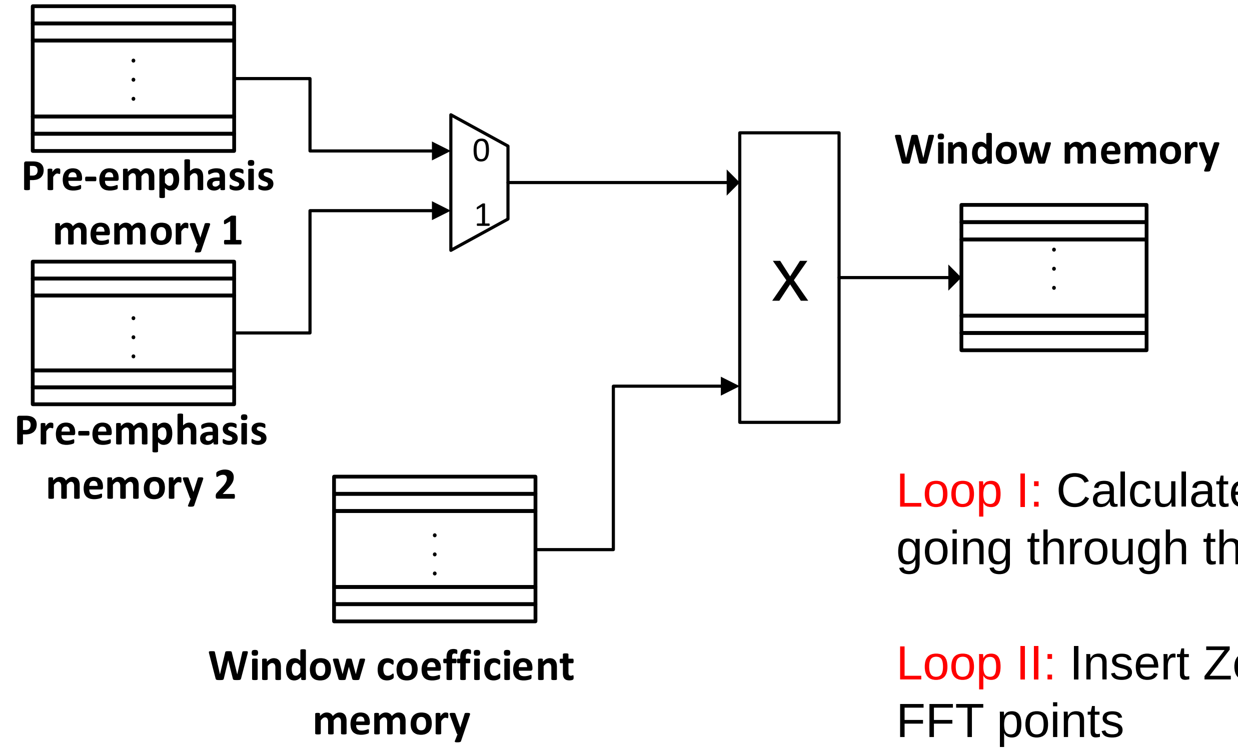
Loop I : Pre-emphasis
Loop II: Energy

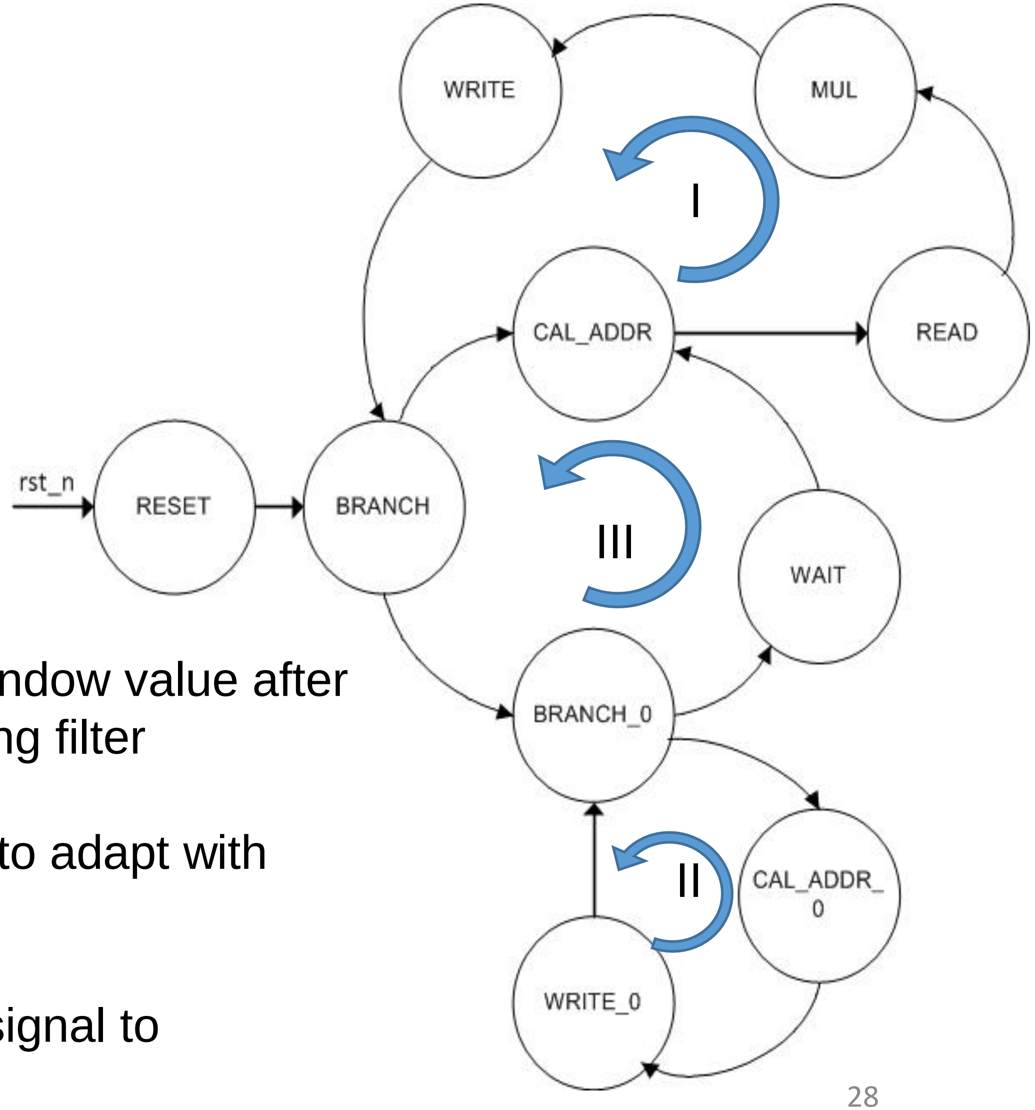
MFCC MODEL AND ARCHITECTURE

WIN

Window

$$h[n] = p[n] \cdot \left\{0.54 - 0.46 \cdot cos\left(\frac{2\,\pi\,n}{N-1}\right)\right\}$$





Loop I: Calculate each Window value after going through the Hamming filter

Loop II: Insert Zero value to adapt with FFT points

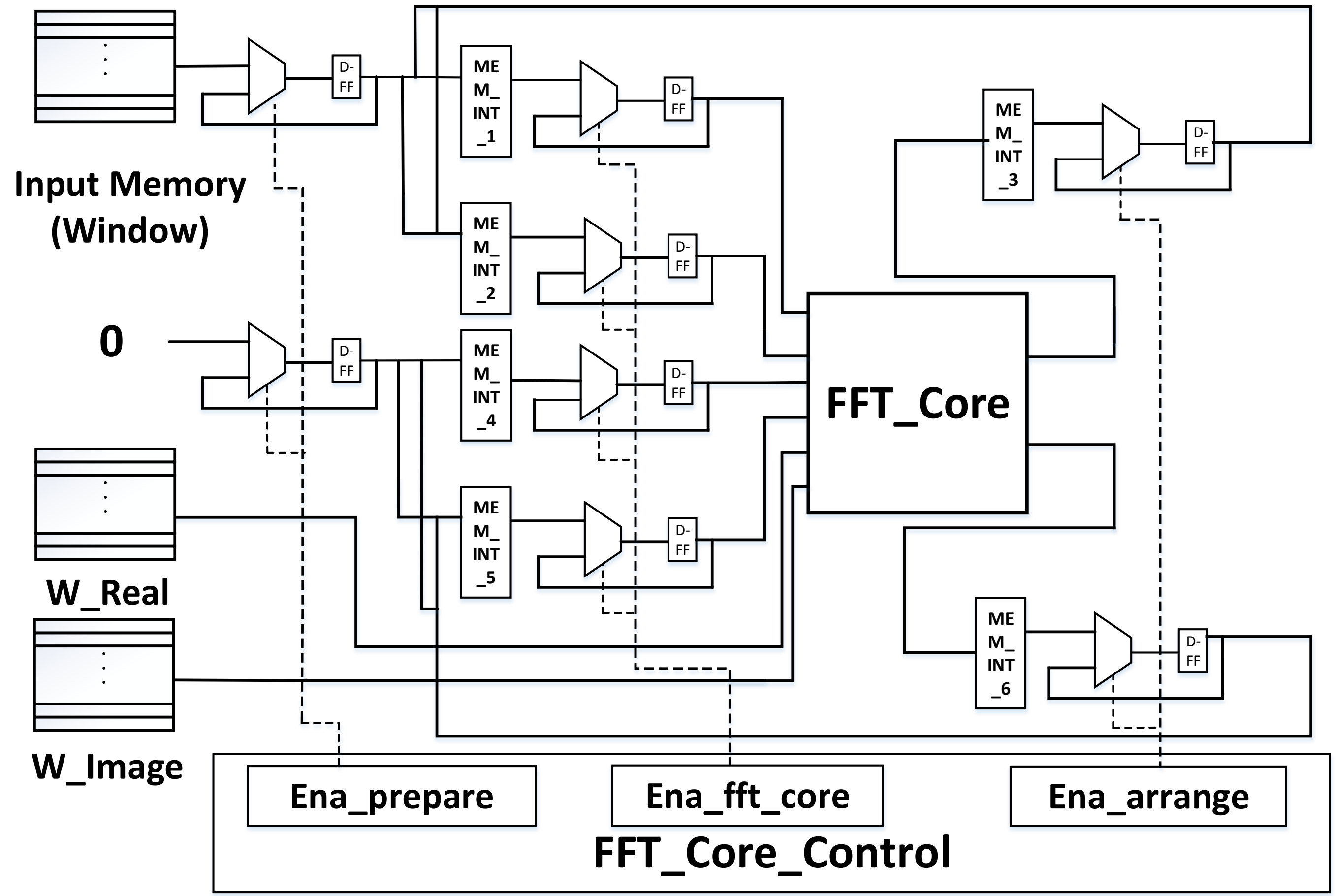
Loop III: Wait the enable signal to calculate a next frame

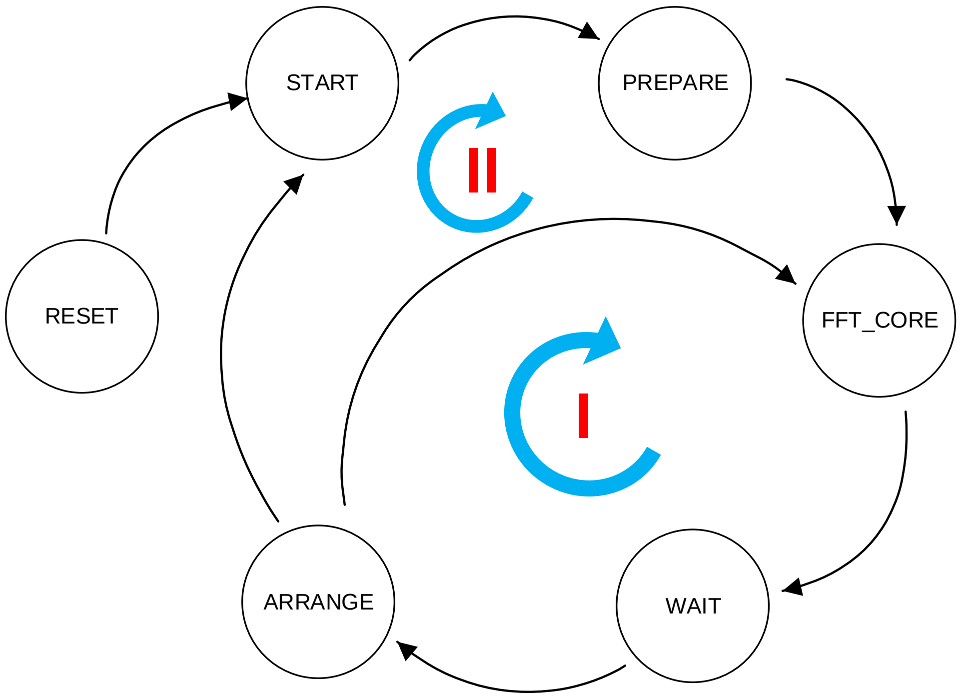
MFCC MODEL AND ARCHITECTURE

Dynamic FFT

FFT

$$H[k] = \sum_{n=0}^{N-1} h(n) \cdot e^{j\frac{2\pi n}{N}}$$

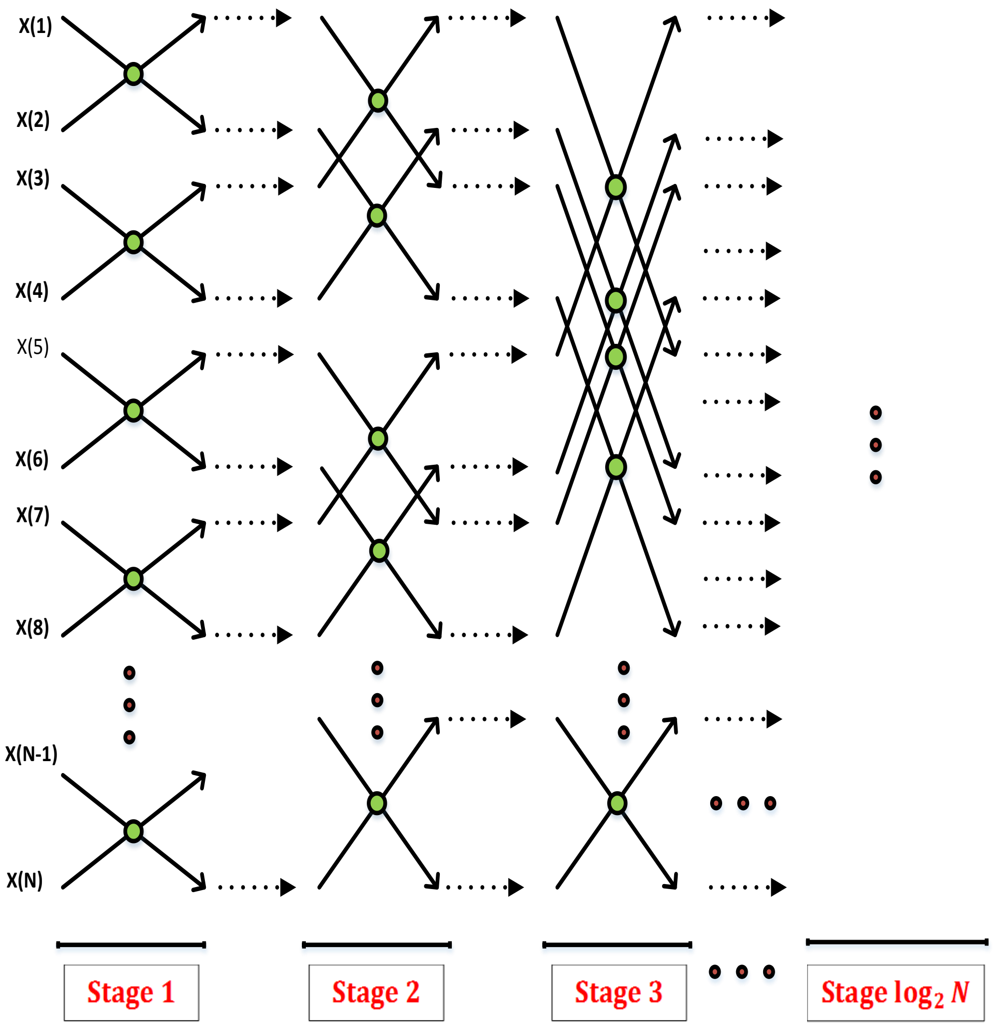


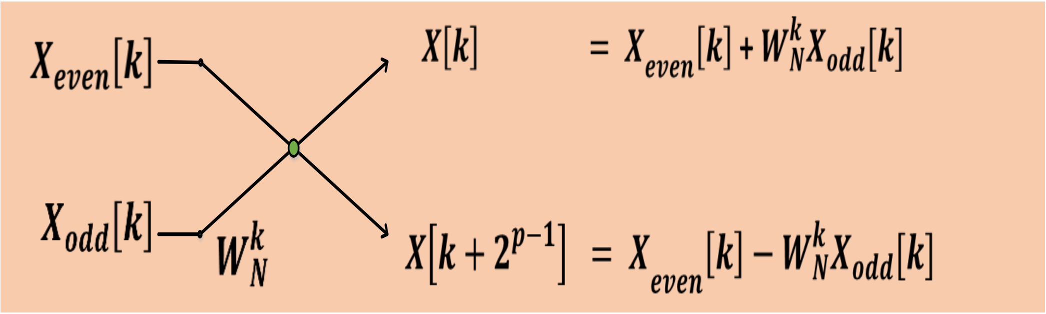


- Loop I: Repeat the Butterfly computation
- Loop II: Wait the enable signal to calculate a next frame

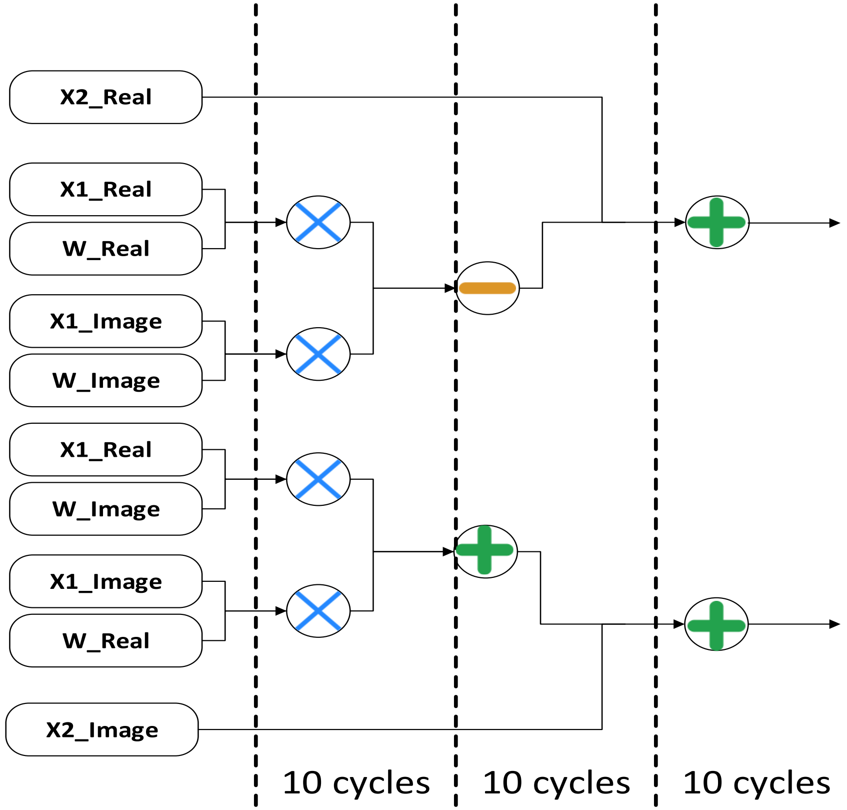
MFCC MODEL AND ARCHITECTURE

Dynamic FFT

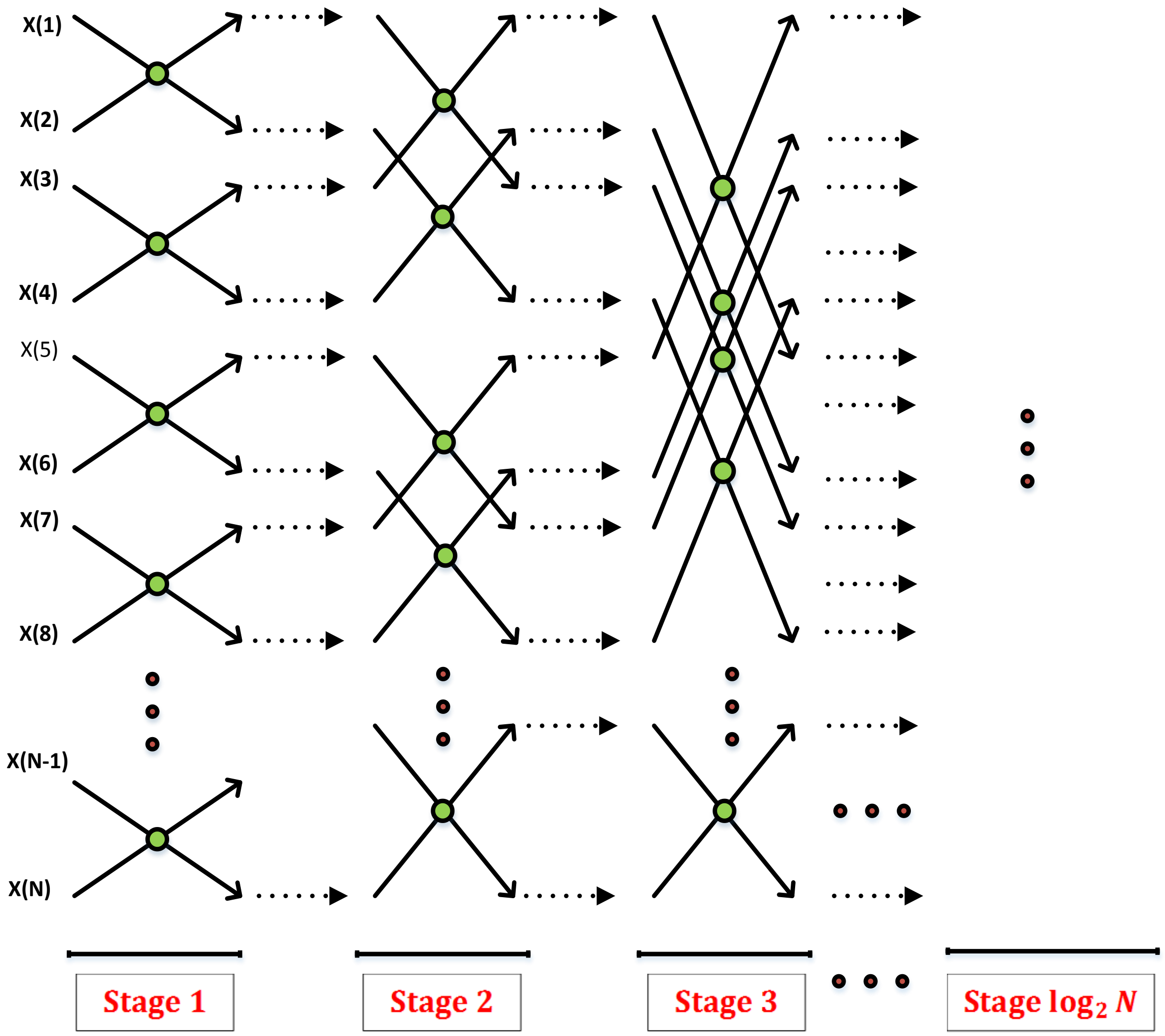




Recursive Butterfly Unit



Dynamic FFT



The results of this work published at ECIT – REV conference 2015

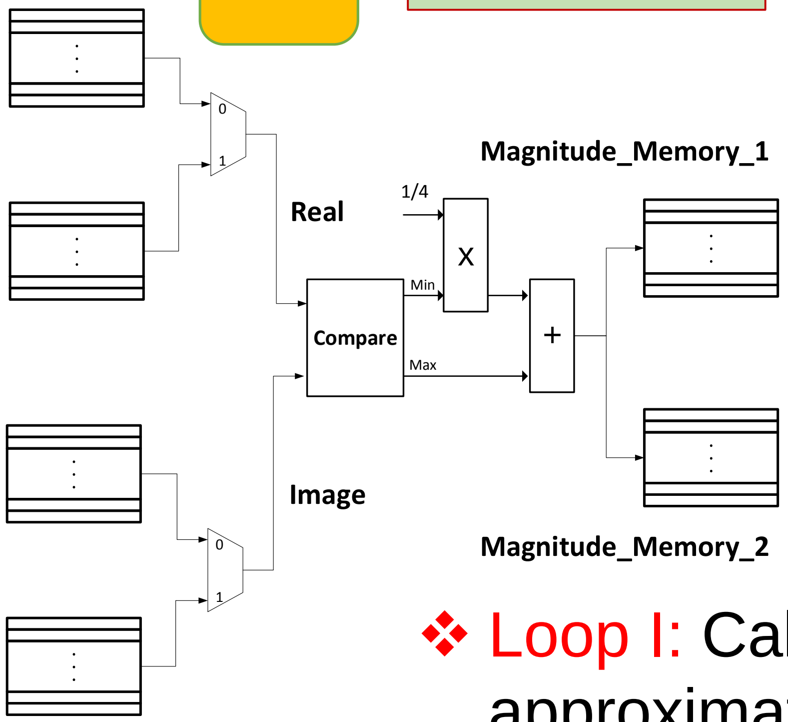
FFT Points	Clocks	Latency at 500MHz (ms)
8	780	1.56E-3
16	1710	3.42E-3
32	3760	7.52E-3
64	8370	0.0167
128	18740	0.0374
256	41910	0.0838
512	93240	0.1864
1024	206010	0.4120

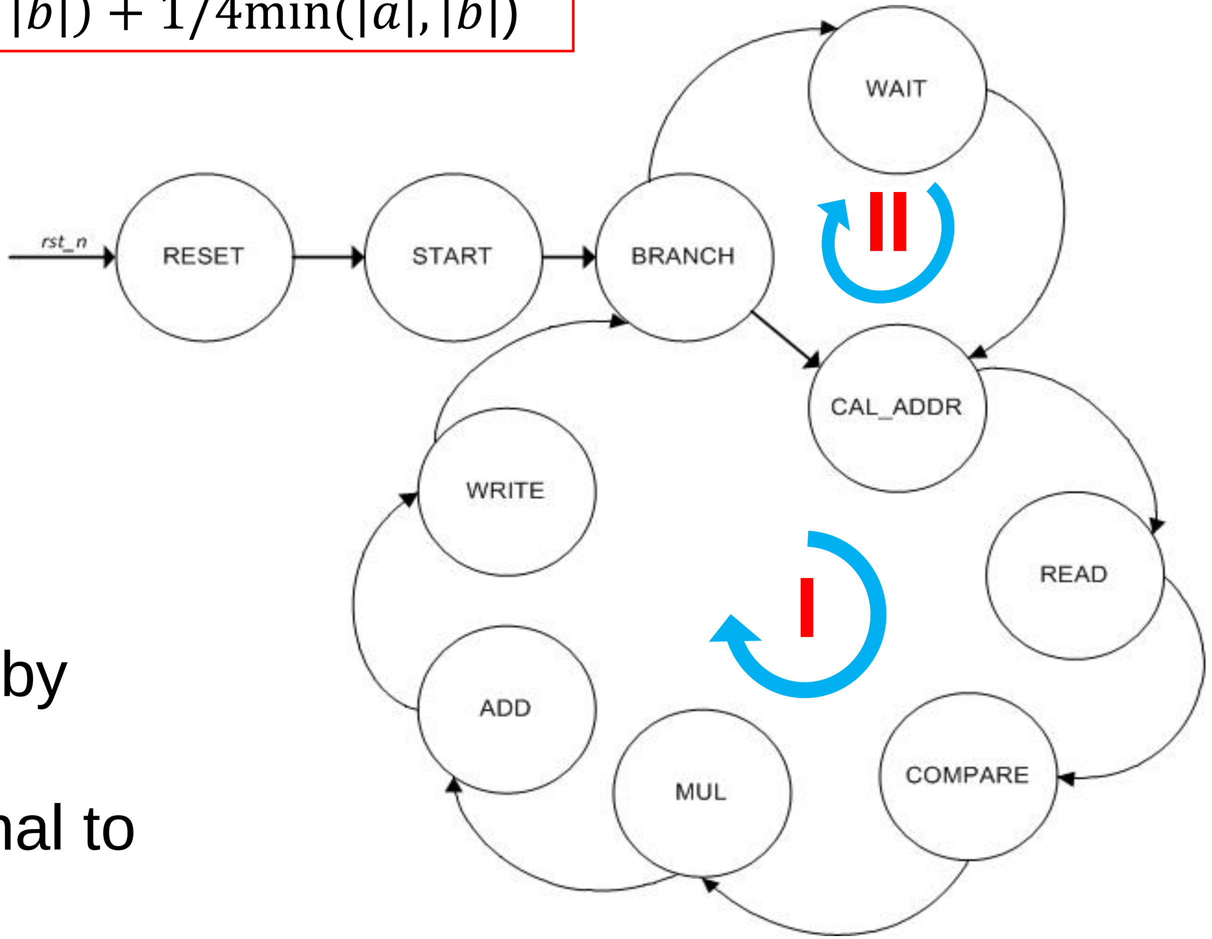
MFCC MODEL AND ARCHITECTURE

AMP

Amplitude

$$|a + jb| = \max(|a|, |b|) + 1/4\min(|a|, |b|)$$





- **Loop I:** Calculate amplitude by approximate formula
- **Loop II:** Wait the enable signal to calculate a next frame

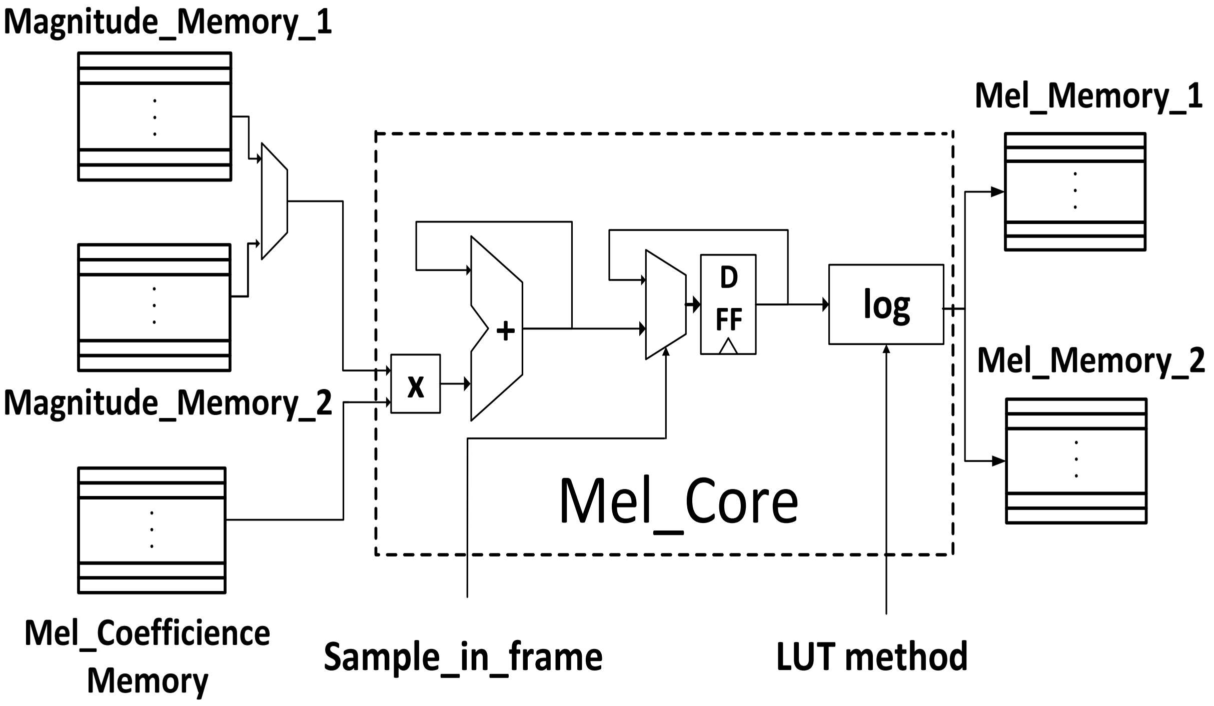
Reference: Hoang Trang, Nguyen Ly Thien Truong "VLSI Architecture Of Magnitude Estimation Algorithm For Speech Recognition System," *Chuyên san Công nghệ thông tin và Truyền thông,* vol. 5, pp. 92-101, 10-2014

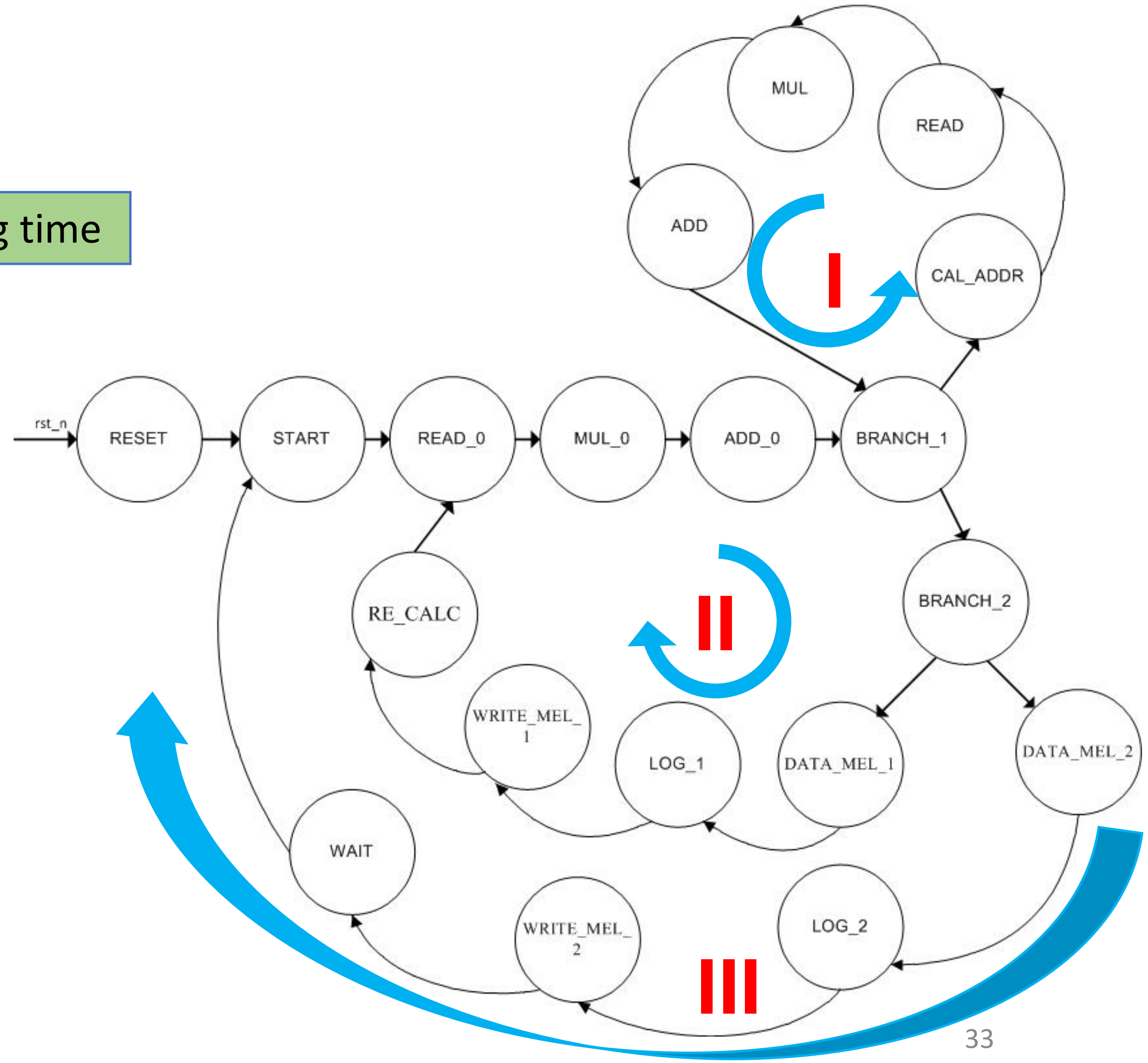
Mel

MEL

$$X[l] = \log\left(\sum_{k=k_{ll}}^{k_{lu}} |H[k]| \cdot W_l[k]\right)$$

Quite a complicated task and take a long time





MFCC MODEL AND ARCHITECTURE

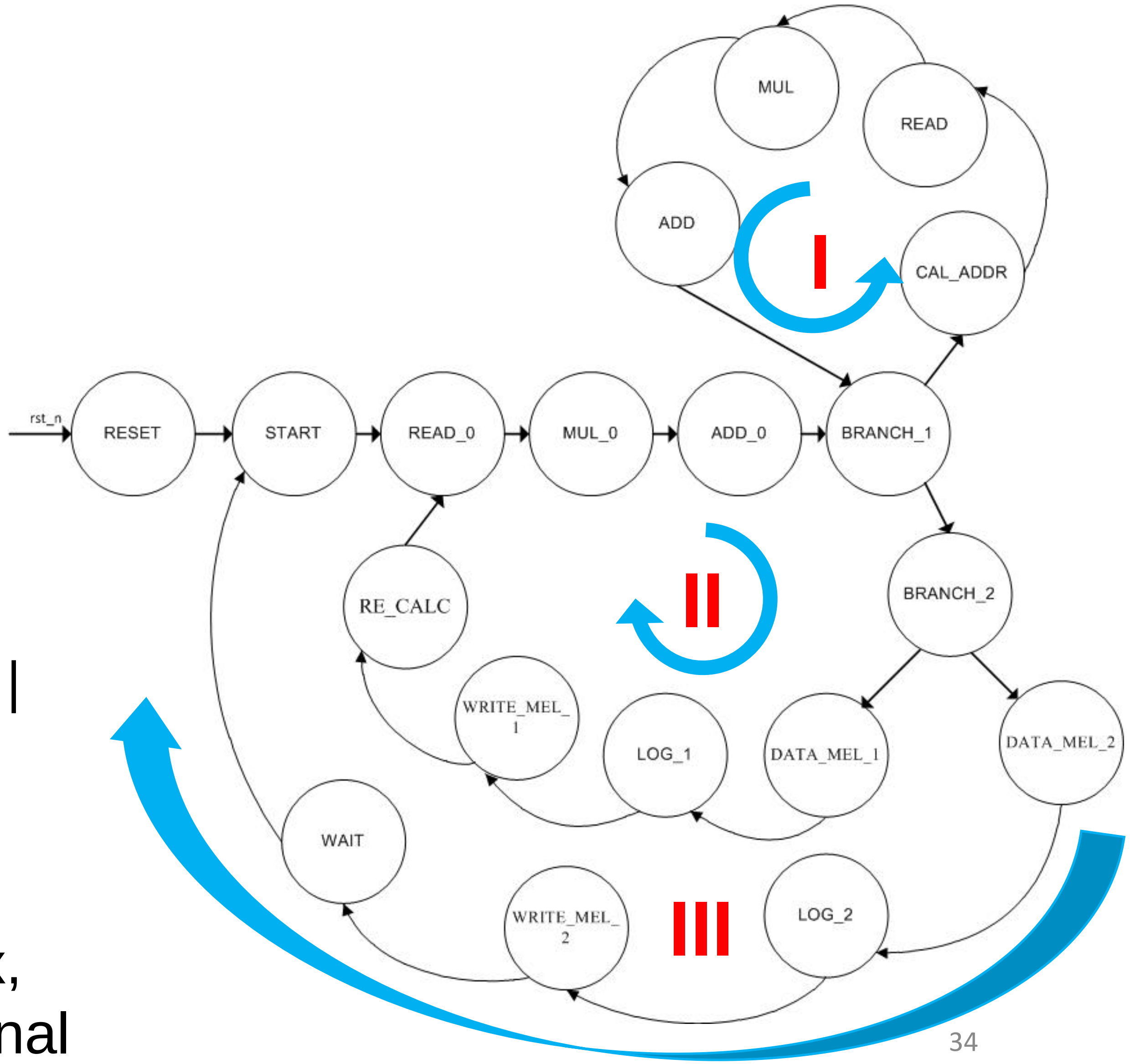
Mel

$$X[l] = \log\left(\sum_{k=k_{ll}}^{k_{lu}} |H[k]| \cdot W_l[k]\right)$$

$$\begin{bmatrix} m_{1,1} & \cdots & m_{1,80} \\ \vdots & \ddots & \vdots \\ m_{23,1} & \cdots & m_{23,80} \end{bmatrix} . \begin{bmatrix} x_1 \\ \vdots \\ x_{80} \end{bmatrix}$$

$W_l[k]$ — Mel coefficients

$|H[k]|$ — Amplitude



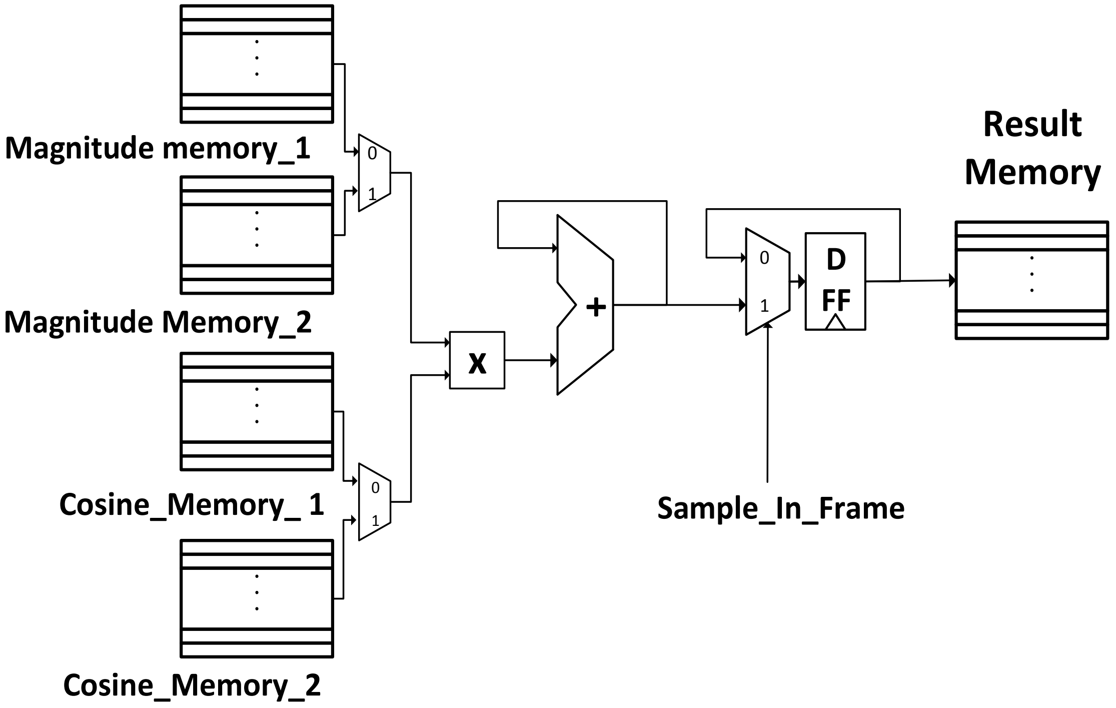
- **Loop I:** Multiply each arrow value of $|H[k]|$ with each column value of $W_l[k]$ and the take the sum of their results
- **Loop II:** Call a new arrow of $|H[k]|$
- **Loop III:** Finish the multiplication of matrix, take logarithm and wait a next enable signal

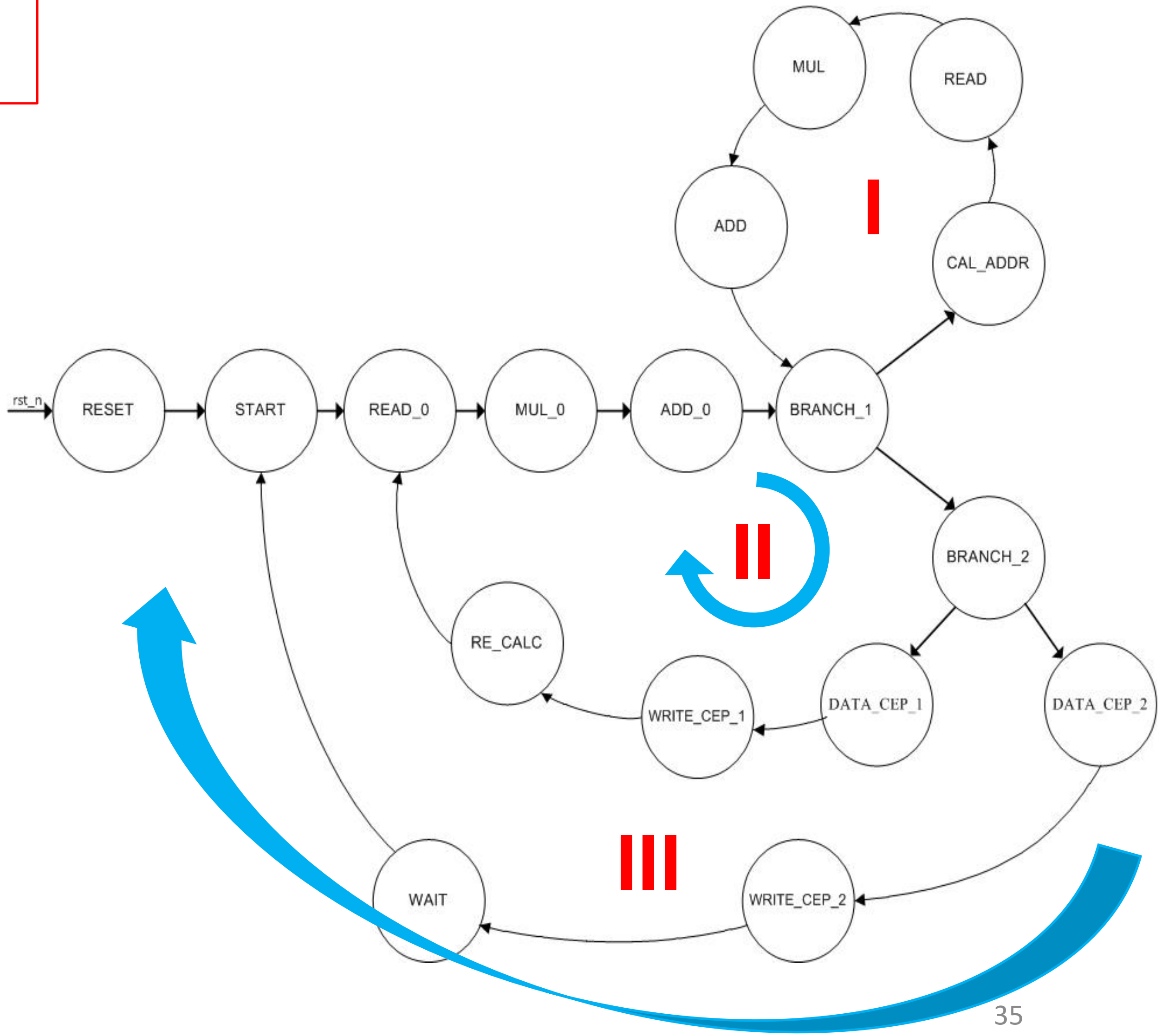
MFCC MODEL AND ARCHITECTURE

Cepstral

$$C[m] = \sum_{l=1}^{L} X[l] cos\left(\frac{\pi\, m\, (l - 0.5)}{L}\right)$$

CEP





MFCC MODEL AND ARCHITECTURE

Cepstral

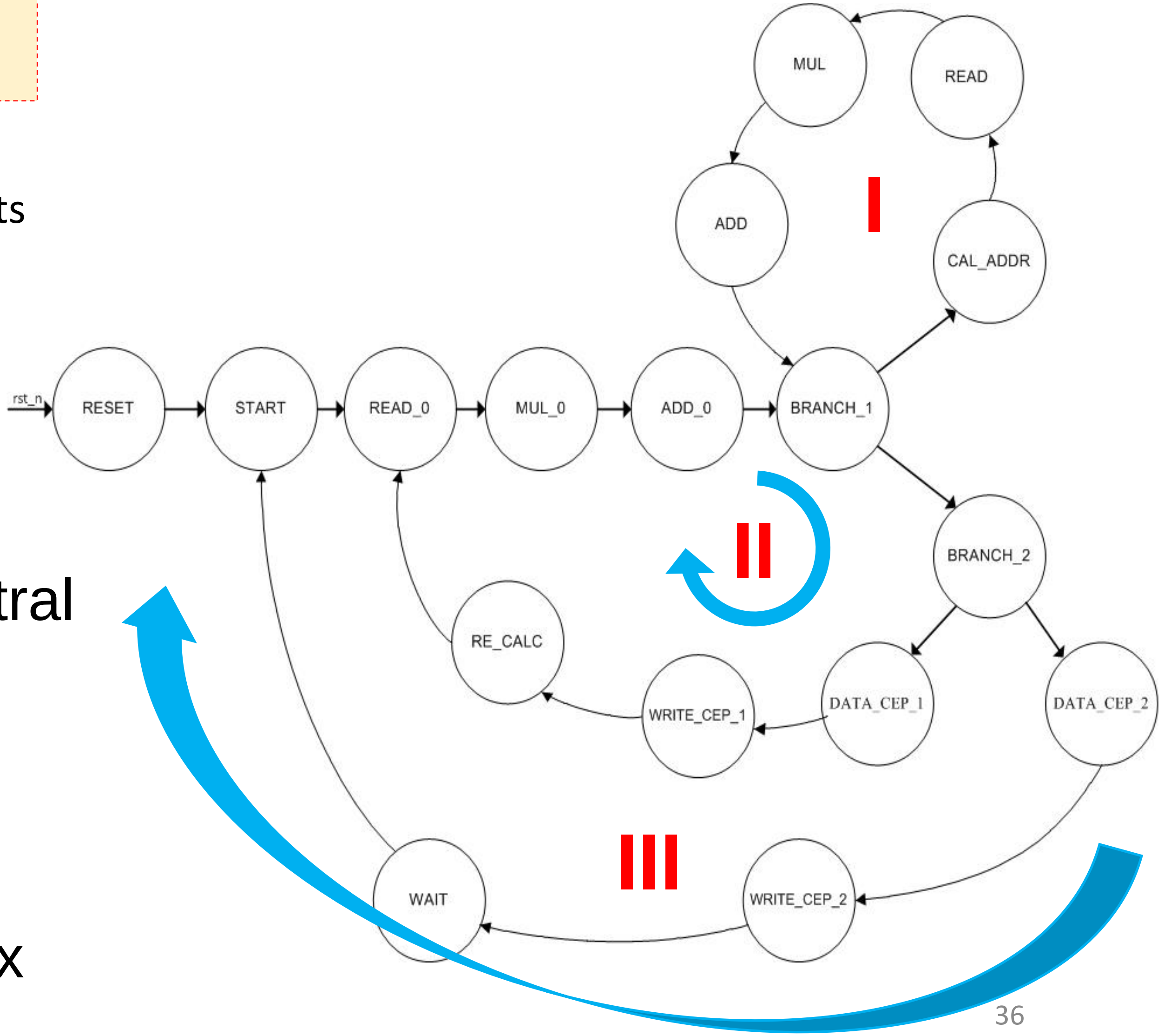
$$C[m] = \sum_{l=1}^{L} X[l] cos\left(\frac{\pi\, m\, (l - 0.5)}{L}\right)$$

Cepstral Coefficients

$$\begin{bmatrix} c_{1,1} & \cdots & c_{1,80} \\ \vdots & \ddots & \vdots \\ c_{23,1} & \cdots & c_{23,80} \end{bmatrix} . \begin{bmatrix} x_1 \\ \vdots \\ x_{80} \end{bmatrix}$$

$$cos\left(\frac{\pi\, m\, (l - 0.5)}{L}\right) \qquad X[l]$$

- **Loop I:** Multiply each arrow value of Ceptral coefficient with each column value of $X[l]$ and the take the sum of their results
- **Loop II:** Call a new arrow of Cepstral Coefficient
- **Loop III:** Finish the multiplication of matrix and wait a next enable signal



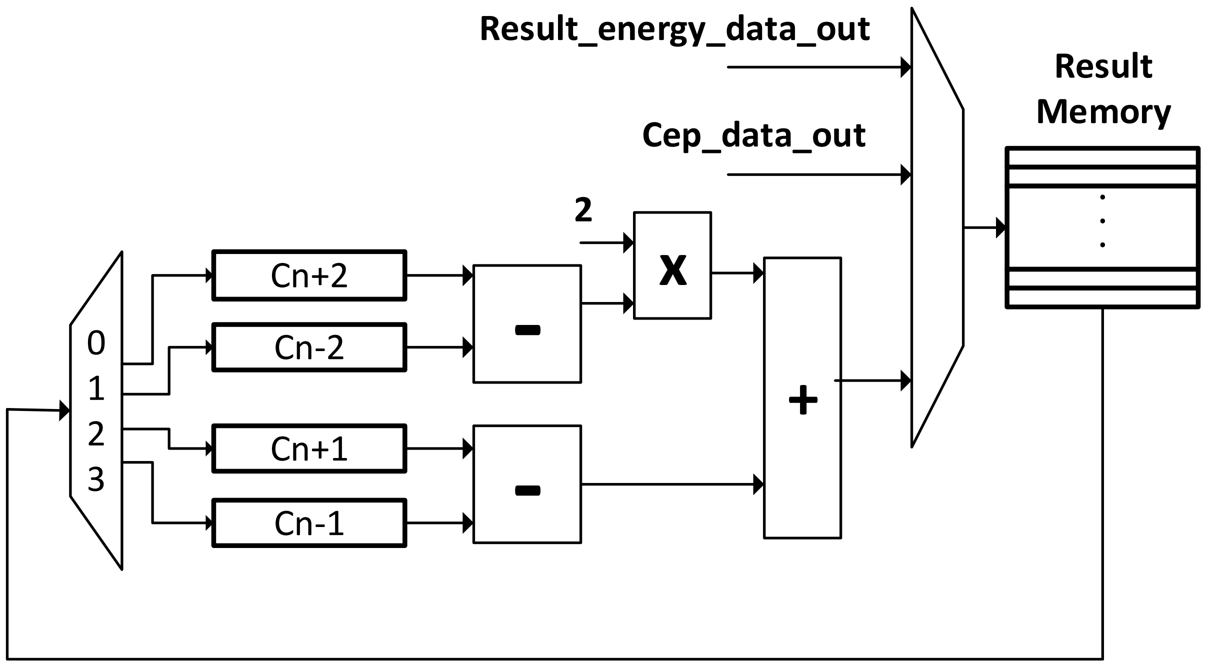
MFCC MODEL AND ARCHITECTURE

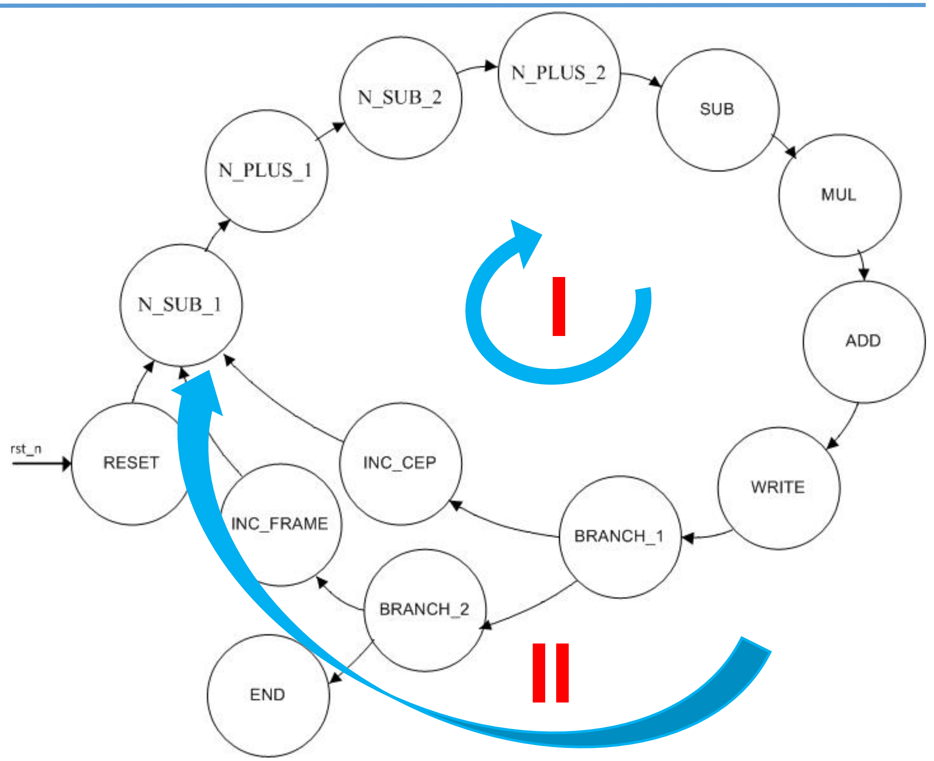
Delta

$$C'_n = 2(C_{n+2} - C_{n-2}) + C_{n+1} - C_{n-1}$$

Delta - Delta

$$C''_n = 2(C'_{n+2} - C'_{n-2}) + C'_{n+1} - C'_{n-1}$$

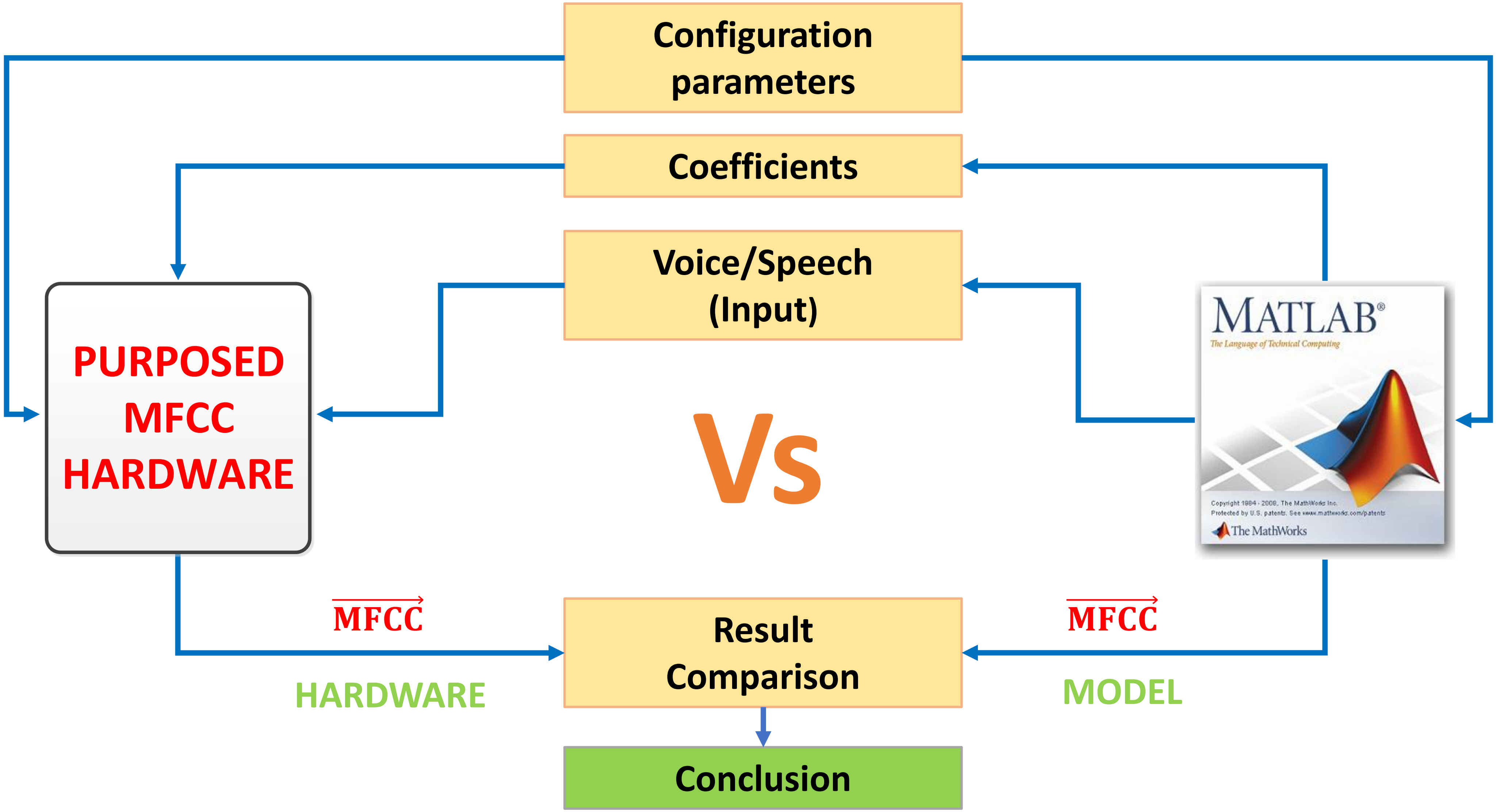




- **Loop I:** Calculate each delta and delta-delta value for 1 MFCC vector
- **Loop II:** Finish and calculate next new frames

CONTENTS

ACCURACY ESTIMATION



ACCURACY ESTIMATION

Testcase	Sample per frame	Overlap Ratio	FFT points	Mel	Cepstral
Maximum	320	50%	512	63	31
Medium	320	50%	512	50	21
Minimum	160	50%	256	30	12

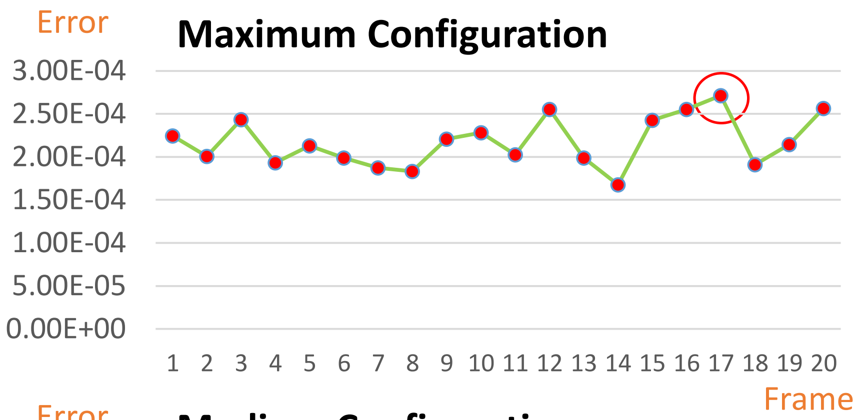
$$E_i = \frac{\sum_{j=1}^{n} \left| x_{matlab_j} - x_{hardware_j} \right|}{n}$$

E_i is an average error of MFCC vector for i^th frame

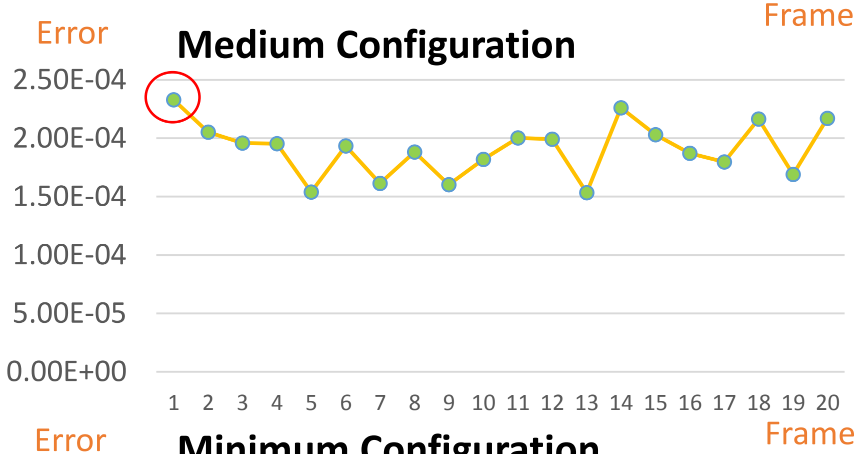
x_{matlab} is the MFCC calculated by Matlab

$x_{hardware}$ is the MFCC calculated by Verilog

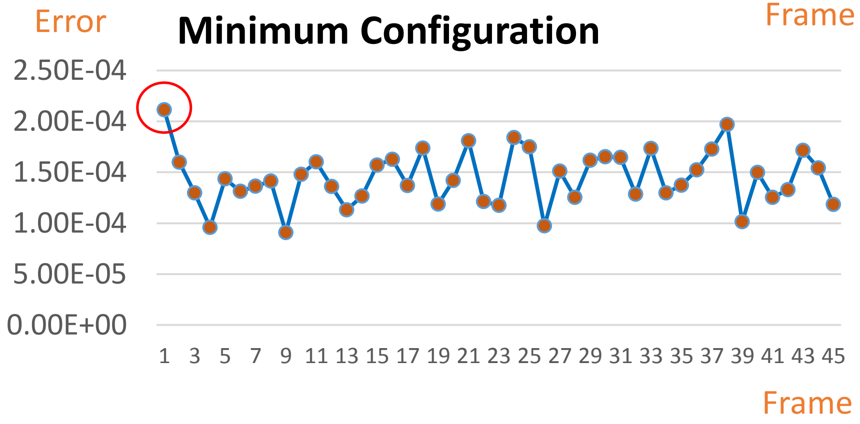
n is the number of energy, cepstral, delta and delta-delta in each frame.



2.71E-4



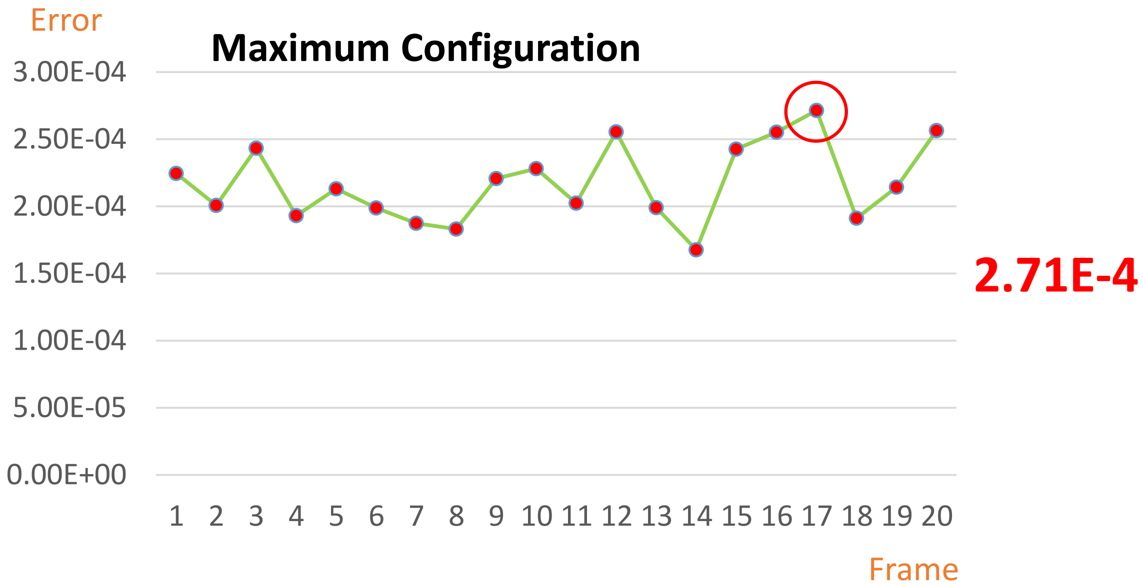
2.33E-4



2.11E-4

ACCURACY ESTIMATION

Testcase	Sample per frame	Overlap Ratio	FFT points	Mel	Cepstral
Maximum	320	50%	512	63	31
Medium	320	50%	512	50	21
Minimum	160	50%	256	30	12



Maximum MFCC configuration :

- Operation Time is 0.0216 s
- Maximum Absolute Error is 2.71E-4
- Maximum Relative Error is 0.0163 %

CONTENTS

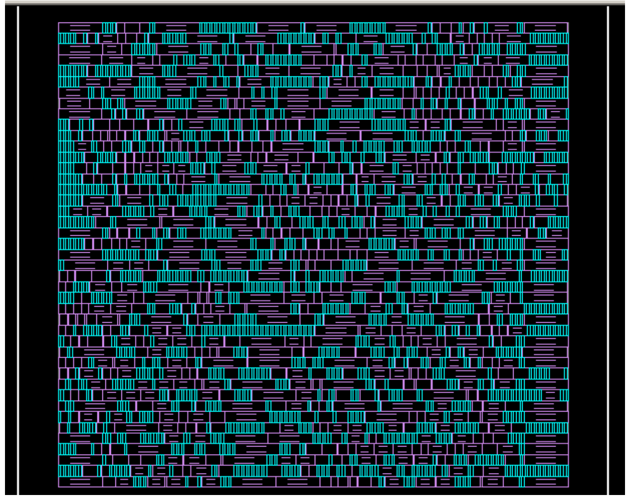
PHYSICAL PERFORMANCE

Frequency (MHz)	Total equivalent gate count (# cells)
100	29 684 971
200	29 803 786
250	29 701 747
500	30 186 098

$$\frac{500\ (\mathrm{MHz})}{100\ (\mathrm{MHz})} = 5$$

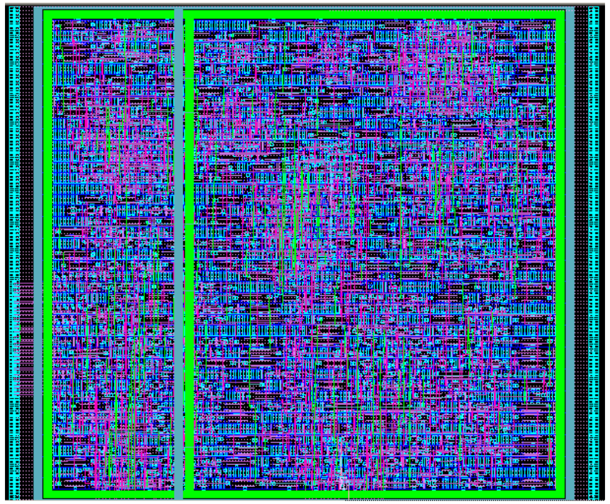
But

$$\frac{30\,186\,098}{29\,684\,971} \approx 1$$



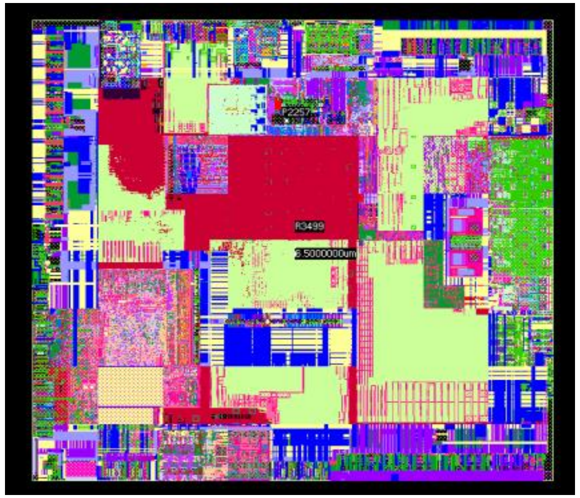
Floorplan

→



Place and Route

→

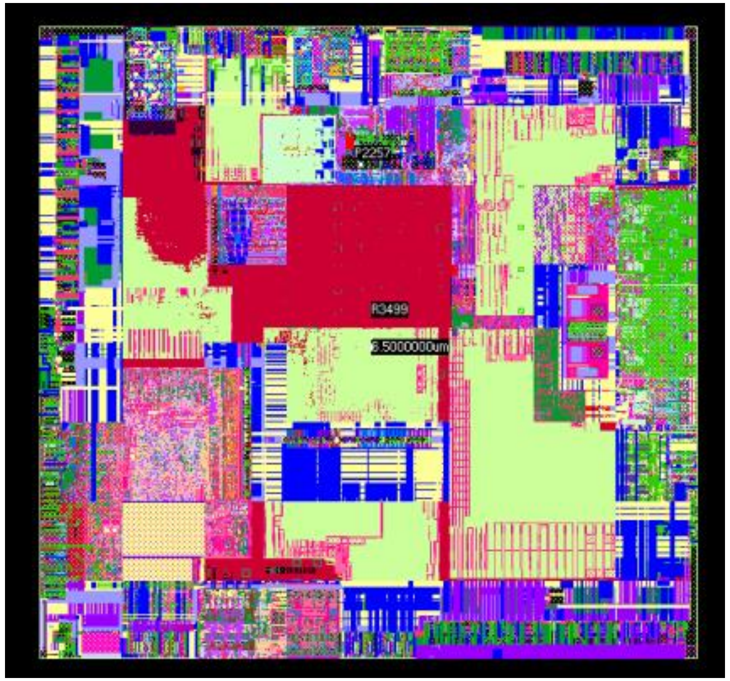


Final GDS File

CONTENTS

CONCLUSIONS

Architecture	Proposed Architecture	[9]	[11]
Technology	ASIC (130nm)	ASIC (0.6µm)	ASIC (0.18µm)
FFT points	8 → 512	256	256
Mel	1 →63	20	32
Cepstral	1 → 31	12	13
Feature Number	12 → 96	12	48
Core Area (mm^2)	1.29x1.29	3.2x3.3	6.5x3.5
Frequency (MHz)	500	50	30



Final GDS file of Purposed MFCC Hardware Architecture

[9] Jia-Ching Wang, Jhing- Fa Wang, Yu-Sheng Weng, "Chip Design Of Mel Frequency Cepstral Coefficients," *Acoustics, Speech, and Signal Processing, IEEE,* vol. 6, pp. 3658 - 3661, 2000.

[11] E. Cornu, "An Ultra Low Power, Ultra Miniature Voice Command System Based On Hidden Markov Models," in *Acoustics, Speech, and Signal Processing, IEEE*, Orlando, FL, USA, 2002.

ACHIEVEMENTS

Lam Pham, Trong Du Nguyen, Dat Thanh Ngo, Hoang Trang, "An Efficient Hardware Architecture for Dynamic FFT Based on Radix 2," in *The 2015 National Conference on Electronics, Communications and Information Technology*, ECIT-REV, Ho Chi Minh, 2015.

Tam Chi Nguyen, Dat Thanh Ngo, Lam Pham, Hieu Minh Nguyen, Bao Gia Bui, Hoang Trang "A High Performance Dynamic ASIC-Based Audio Signal Feature Extraction (MFCC)," in *International Conference on Advanced Computing and Application, ACOM 2016, IEEE*, Cantho, 2016. *Proceedings will be published on November 23-25, 2016*

THANKS !