

# Mục lục

Mục lục.....	1
I. Giới thiệu chung về H264:.....	2
1. Giới thiệu chung về H264:.....	2
2. Luồng dữ liệu vào ra của H264:.....	4
a. Input/output : .....	4
b. Thông số chất lượng trong H264:.....	5
c. So sánh giữa h264 và MPEG-2:.....	5
II. Quá trình nén H264.....	6
1. Sơ đồ khối tổng quát của H264 encoder:.....	6
2. Chức năng chung các khối:.....	7
a. Prediction:.....	7
b. Transform .....	9
c. Lượng tử hóa:.....	9
d. Reordering .....	10
e. Entropy encoding:.....	10
III. Intra prediction:.....	10
1. Giới thiệu chung:.....	10
2. Chọn loại kích cỡ macroblock.....	11
3. Chọn Mode cho mỗi kích cỡ macroblock:.....	12
4. Lưu đồ thuật toán chung:.....	15

# Tổng quan về quá trình nén của H264/AVC(Advance Video Compression):

## I. Giới thiệu chung về H264:

### 1. Giới thiệu chung về H264:

H264 là một chuẩn nén video được phát triển bởi ITU- T Video Coding Experts Group (VCEG) và ISO/IEC Moving Picture Experts Group (MPEG) được chuẩn hóa bởi ITU dưới tên H264 (MPEG4 Part 10 AVC)

H264 gồm 4 profile cho các ứng dụng dành cho người tiêu dùng, các tính chất của profile được liệt kê trong bảng sau:

	Baseline	Extended	Main	High
I and P Slices	Yes	Y	Y	Y
B Slices	No	N	Y	Y
SI and SP Slices	N	Y	N	N
Multiple Reference Frames	Y	Y	Y	Y
In-Loop Deblocking Filter	Y	Y	Y	Y
CAVLC Entropy Coding	Y	Y	Y	Y
CABAC Entropy Coding	N	N	Y	Y
4:2:0 Chroma Format	Y	Y	Y	Y
Monochrome Video Format (4:0:0)	N	N	N	Y
4:2:2 Chroma Format	N	N	N	N
4:4:4 Chroma Format	N	N	N	N

### Ứng dụng:

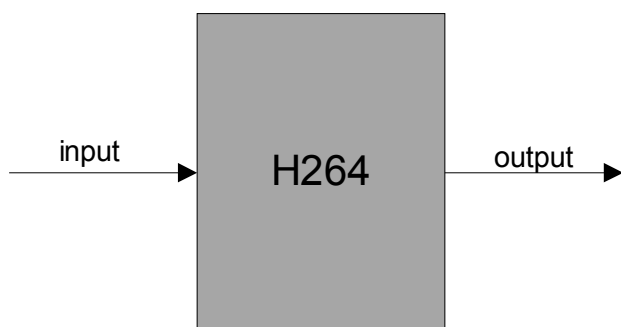
Tên profile	Ứng dụng
Baseline profile	Cho các ứng dụng giá thành thấp, được ứng dụng rộng rãi

	trong video hội nghị và các ứng dụng dùng cho di động
Main Profile	Mục đích ban đầu là hướng tới các ứng dụng broadcast và lưu trữ, tuy nhiên tầm quan trọng của Main Profile bị giảm khi High Profile cũng được phát triển cho các ứng dụng này
Extended Profile	Dùng thông dụng cho truyền dòng video thông qua mạng IP, có tỉ lệ nén cao và có cơ chế sửa lỗi
High Profile	Các ứng dụng broadcast và lưu trữ đĩa, các ứng dụng trong truyền hình HD (HD DVD, blu-ray Disc..)

### H264 Levels:

Level number	Max macroblocks/s	Max frame size macroblocks	Max video bit rate (VCL) Baseline, Extended and Main Profiles	Max video bit rate (VCL) High Profile
<b>1</b>	1485	99	64 kbit/s	80 kbit/s
<b>1b</b>	1485	99	128 kbit/s	160 kbit/s
<b>1.1</b>	3000	396	192 kbit/s	240 kbit/s
<b>1.2</b>	6000	396	384 kbit/s	480 kbit/s
<b>1.3</b>	11880	396	768 kbit/s	960 kbit/s
<b>2</b>	11880	396	2 Mbit/s	2.5 Mbit/s
<b>2.1</b>	19800	792	4 Mbit/s	5 Mbit/s
<b>2.2</b>	20250	1620	4 Mbit/s	5 Mbit/s
<b>3</b>	40500	1620	10 Mbit/s	12.5 Mbit/s
<b>3.1</b>	108000	3600	14 Mbit/s	17.5 Mbit/s
<b>3.2</b>	216000	5120	20 Mbit/s	25 Mbit/s
<b>4</b>	245760	8192	20 Mbit/s	25 Mbit/s
<b>4.1</b>	245760	8192	50 Mbit/s	50 Mbit/s
<b>4.2</b>	522240	8704	50 Mbit/s	50 Mbit/s
<b>5</b>	589824	22080	135 Mbit/s	168.75 Mbit/s
<b>5.1</b>	983040	36864	240 Mbit/s	300 Mbit/s

## 2. Luồng dữ liệu vào ra của H264:



### a. Input/output :

H264 hỗ trợ sampling format 4:2:0 (và đây chính là sampling format mặc định của H264) ngoài ra, H264 còn hỗ trợ các sampling format 4:4:4 và 4:2:2 để tăng chất lượng ảnh tùy theo từng ứng dụng

Video frame formats với 4:2:0 sampling format:

	Luminance resolution	Kb/s
Format	(Horizon x vertical)	
full HD	1920 x 1080	7000 – 8000
commonly – used HD	1280 – 720	5000 – 6000
standard definition	640 x 480	1000 – 2000
Internet – size content	320 x 240	300 – 500
3G	176 x 144	150 – 200 (for 24-30 frames/s)

File mở rộng cho H264:

.avi, .mp4, .mkv, .h264

**b. Thông số chất lượng trong H264:**

PNSR (peak signal to noise ratio): tỉ số tín hiệu trên tạp âm, được đo dựa trên sự sai khác giữa ảnh gốc và ảnh sau khi nén:

$$PSNR_{dB} = 10 \log_{10} \frac{(2^n - 1)^2}{MSE}$$

MSE: mean square error (giá trị sai khác trung bình giữa ảnh nén và ảnh gốc)

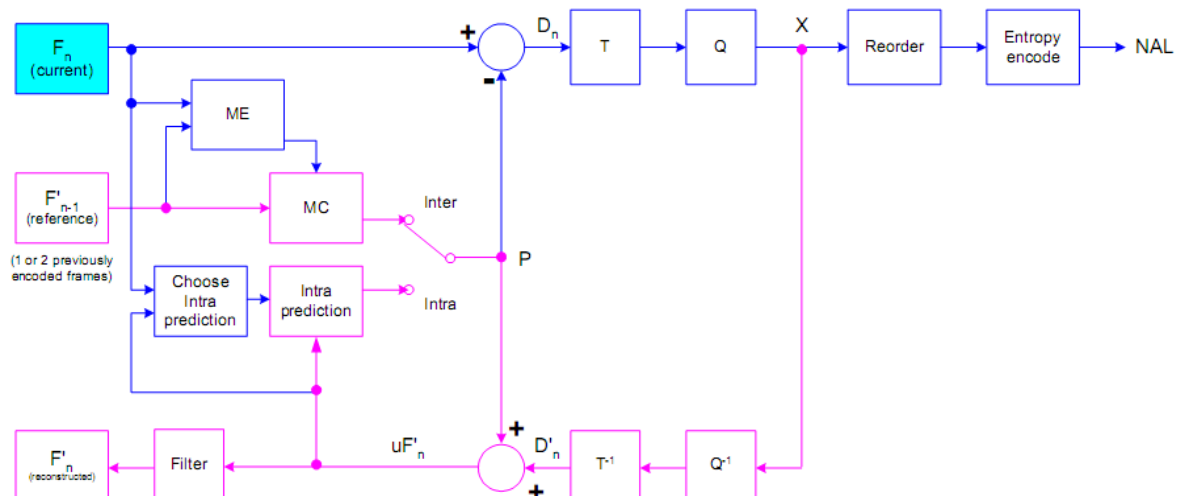
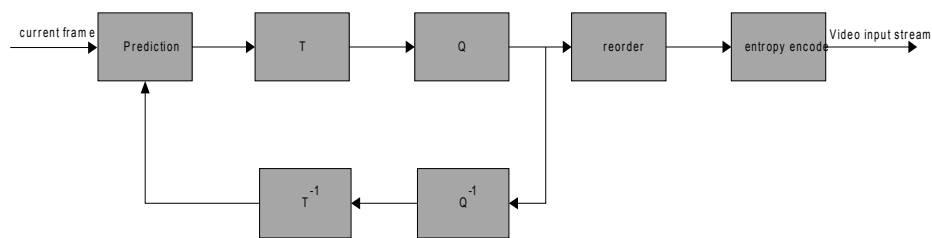
**c. So sánh giữa h264 và MPEG-2:**

	MPEG-2 (H262)	MPEG – 4 (H264)
Intra Prediction	Không, Macroblock encoded	Spatial 4x4, 16x16
Mã hóa ảnh	- Frame - Field - Picture AFF	- Frame - Field - Picture AFF - MB AFF
Kích thước block cho bù chuyển động	- 16x16 - 16x8, 8x16	-16x16 -16x8, 8x16 -8x8 - 8x4, 4x8 - 4x4
Tiên đoán vector chuyển động	- Full Pel _ Half Pel	- Full Pel - Half Pel - Quarter Pel
Reference Frame (với P frame)	- Single	- Single

		-Multiple
Entropy Coding	- VLC	- CAVLC - CABAC
Transform	- 8x8 DCT	- 4x4 Interger “DCT” - 8x8 Interger “DCT”

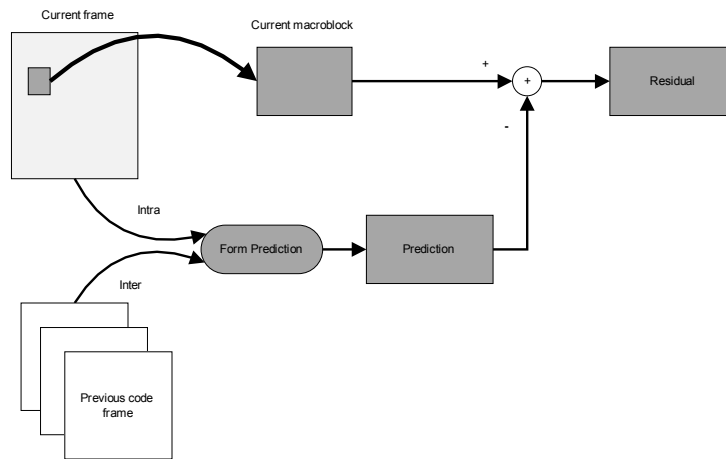
## II. Quá trình nén H264

### 1. Sơ đồ khối tổng quát của H264 encoder:



## 2. Chức năng chung các khối:

### a.Prediction:



Quá trình prediction giúp loại trừ một số lượng lớn độ dư thừa về không gian và thời gian tồn tại trong chuỗi video. Về cơ bản, prediction tính toán ra được một predicted frame, sau đó so sánh và tính toán sự sai khác giữa predicted frame với frame hiện tại và chỉ mã hóa và truyền đi sự sai khác đó. Như vậy, quá trình prediction càng tốt thì sự sai khác càng nhỏ, làm tăng hiệu quả quá trình nén

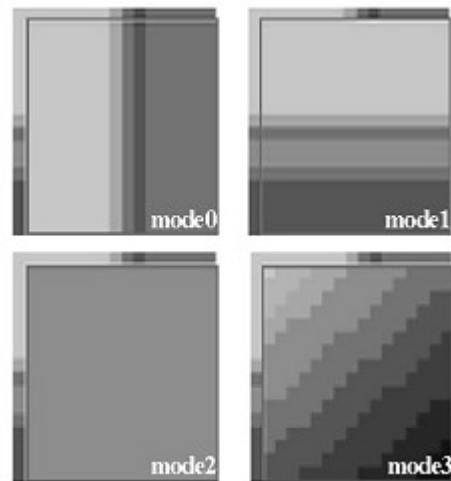
Trong prediction được chia làm 2 mode: intra coding và inter coding tương ứng làm giảm độ dư thừa về không gian và thời gian

### Intra coding:

Quá trình intra coding được thực hiện chỉ trong một frame. So với MPEG-2, H264 nén ảnh I hiệu quả hơn nhờ quá trình “spatial prediction” trong intra coding. Hơn nữa H264 còn có khả năng thay đổi mềm dẻo kích cỡ của macro block từ 16x16 tới 4x4 tùy thuộc vào tính chất ảnh, trong mỗi loại macro block còn được chia làm nhiều mode, giúp việc tiên đoán ảnh hiệu quả hơn mà không tăng nhiều thông tin overhead và độ phức tạp khi tính toán

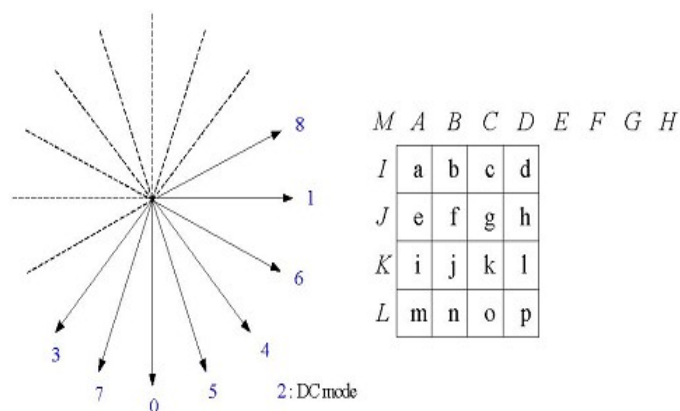
Trong intra coding có 2 loại kích thước macroblock, ứng với mỗi loại là các mode như sau:

16x16 type: 4 mode:



Mode 0: dọc, Mode 1: ngang, Mode 2: DC, Mode 3: plane

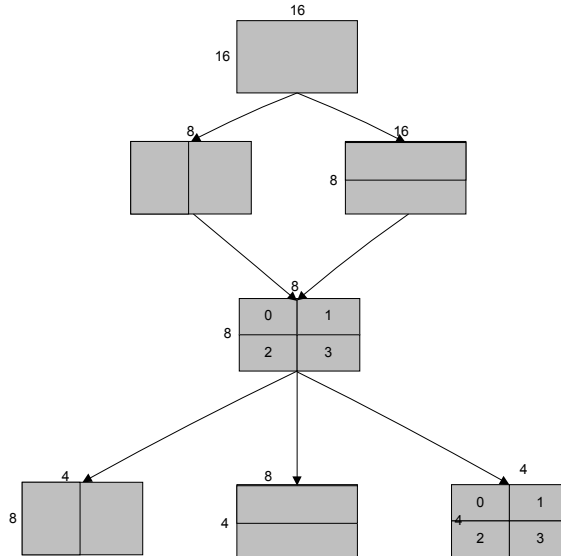
4x4 type: 9 mode



## Inter coding

Quá trình inter coding giúp giảm độ dư thừa về thời gian trong chuỗi video. Được thực hiện giữa các frame với nhau. So với MPEG-2, kích cỡ macro block trong H264 được chia phân lớp, để dễ dàng xử lý với các tính chất ảnh khác nhau, hơn nữa, quá trình tham chiếu inter coding được thực hiện từ không chỉ một mà một vài khung phía trước





Quá trình inter coding được thực hiện chủ yếu dựa trên tính chất: chuỗi frame liên nhau liên tiếp có độ sai khác rất nhỏ. Để giảm độ dư thừa, inter coding thực hiện ước lượng chuyển động : đây chính là bước chiếm nhiều thời gian và có độ phức tạp nhất trong encoder. Ước lượng chuyển động: chọn ra được ứng block matching từ frame tham chiếu, tính toán ra vector chuyển động. Để tăng hiệu quả nén, inter coding thực hiện tiên đoán vector chuyển động từ các macroblock lân cận và chỉ mã hóa độ sai khác giữa hai frame và vector chuyển động.

So với MPEG-2, H264 tính toán độ chính xác tới  $\frac{1}{4}$  pixel với ước lượng chuyển động do đó tăng hiệu quả nén do tính chất của chuyển động rất ít khi trùng với các đường biên của macroblock

## b. Transform

Một block residual sau khối prediction được biến đổi sử dụng biến đổi nguyên 4x4 hoặc 8x8 (sử dụng một kiểu biến đổi DCT). Đầu ra của biến đổi là một bộ các hệ số, mỗi hệ số là một giá trị trọng số sẽ được sử dụng khi tạo lại block residual.

## c. Lượng tử hóa:

Dựa vào tính chất năng lượng bức ảnh chỉ tập trung vào các thành phần tần số thấp. Đầu ra của khối biến đổi sẽ được lượng tử hóa nhằm loại bỏ các thành phần tần số

cao dựa vào các hệ số lượng tử hóa. Sau khi thực hiện biến đổi lượng tử hóa, thông thường kết quả ra là một block mà hầu hết các hệ số là 0, rất ít các thành phần bằng 0. Do đó, nếu các hệ số lượng tử hóa cao, hiệu quả nén cao, nhưng làm giảm chất lượng của bức ảnh khi giải mã

#### **d. Reordering**

Block sau khi biến đổi lượng tử hóa chứa rất ít các hệ số khác 0. Reordering sắp xếp lại cho các hệ số bằng 0 thành một nhóm và các hệ số khác 0 một nhóm để thực hiện entropy encoding, thực hiện quét zigzag.

#### **e. Entropy encoding:**

Thực hiện mã hóa chuỗi video sau khi đã được reordering. Để tính toán hiệu quả và làm giảm độ phức tạp tại bộ decoder, H264 hỗ trợ 2 mode encoding: mode 1 sử dụng CABAC (context-base adaptive arithmetic coding) CAVLC (context-base adaptive variable length code): được dùng để encode các hệ số sau lượng tử hóa trong khi UVLC (universe variable length code) được phát triển để encode thông tin header: các loại macroblock, loại intra prediction, và các cờ báo hiệu khác

### **III. Intra prediction:**

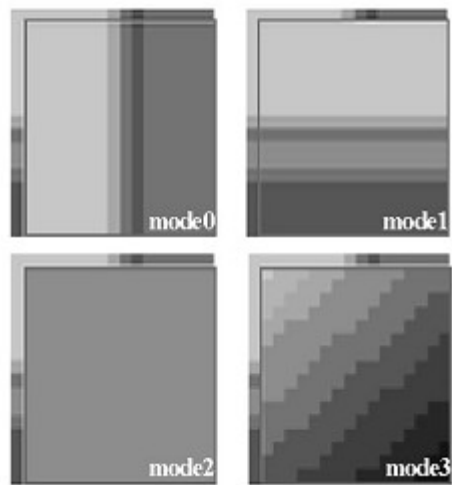
#### **1. Giới thiệu chung:**

Intra prediction là một cải tiến mới trong H264 so với MPEG-2, làm tăng đáng kể hiệu suất nén ảnh I sử dụng phép nội suy từ các block liền kề.

H264 còn rất mềm dẻo trong việc sử dụng kích cỡ block từ 16x16 đến 4x4 ứng với từng tính chất của ảnh. Với những macro block “smooth” độ thay đổi không nhiều, kích thước của macro block lớn (16x16) để vừa không làm giảm hiệu quả nén vừa không tăng độ phức tạp của các phép nội suy. Với những macro block có độ thay đổi nhiều, giảm kích thước macroblock sẽ giúp ích rất nhiều cho việc tăng hiệu quả nén.

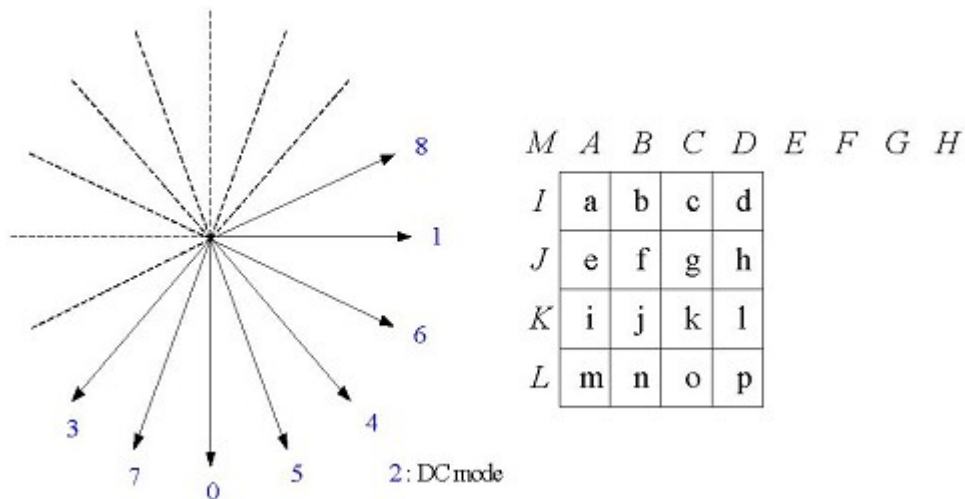
Trong intra prediction có 2 loại kích cỡ macroblock:

16x16: dùng cho các macroblock có độ biến đổi ít. Trong 16x16 có 4 mode tiên đoán:



Mode 0: dọc, Mode 1: ngang, Mode 2: DC, Mode 3: plane

4x4: dùng cho macro block có độ thay đổi nhiều. Do đó trong 4x4 có 9 mode tiên đoán:



Từ đây, có 2 vấn đề chính cần xử lý trong intra prediction: chọn loại và chọn mode cho các macroblock

## 2. Chọn loại kích cỡ macroblock

+ Chọn loại kích cỡ macroblock: phụ thuộc vào tính chất ảnh. Dựa vào tham số MAD (mean absolute deviation)

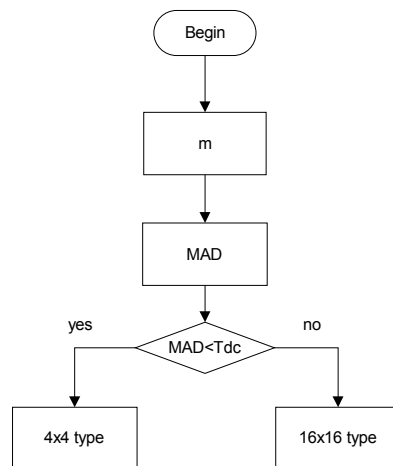
Xét một macroblock 16x16 MAD được tính như sau:

$$MAD = \frac{1}{16 * 16} \sum_{y'=0}^{15} \sum_{x=0}^{15} |p(x,y) - m|$$

$p(x,y)$  giá trị pixel tại tọa độ  $(x,y)$ ,  $m$  là giá trị trung bình của tất cả các pixel trong macroblock

Một giá trị ngưỡng được sử dụng để đánh giá độ “smood” của block. ( $T_{dc} = 3$ )

Ta có sơ đồ thuật toán sau:



### 3. Chọn Mode cho mỗi kích cỡ macroblock:

Chọn mode cho mỗi loại kích cỡ macroblock:

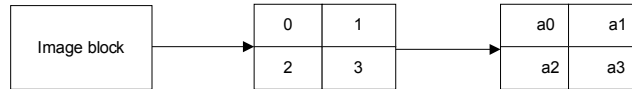
Nhìn chung, để chọn ra best mode cho intra prediction, thì có thể thực hiện tiên đoán cho từng mode, tính độ sai lệch so với macroblock hiện tại và chọn ra độ sai khác nhỏ nhất. Tuy nhiên, khi số frame tăng, độ phức tạp tính toán cũng tăng lên nhiều lần

Đã có nhiều thuật toán đưa ra chọn mode cho intra prediction nhằm làm giảm độ phức tạp tính toán, dựa trên tính chất của ảnh và đặc điểm giới thị giác con người.

Thuật toán sau dựa trên cơ sở: mắt người rất nhạy với các đặc tính biên. Trong MPEG-7, biên được phân ra thành 5 hướng: dọc, ngang, chéo 45 độ, chéo 135 độ

và biên vô hướng, dựa vào hướng biên sau để xác định được mode nào phù hợp nhất

Chia macroblock như hình vẽ



a0, a1, a2, a3 tương ứng là giá trị trung bình trong các sub pixel được chia

Tính toán các hệ số lọc ứng với các loại biên:

$$S^v = \left| \sum_{n=0}^3 a_n \times f_n^v \right|,$$

$$S^h = \left| \sum_{n=0}^3 a_n \times f_n^h \right|,$$

$$S^{45^\circ} = \left| \sum_{n=0}^3 a_n \times f_n^{45^\circ} \right|,$$

$$S^{135^\circ} = \left| \sum_{n=0}^3 a_n \times f_n^{135^\circ} \right|,$$

$$S^{nd} = \left| \sum_{n=0}^3 a_n \times f_n^{nd} \right|,$$

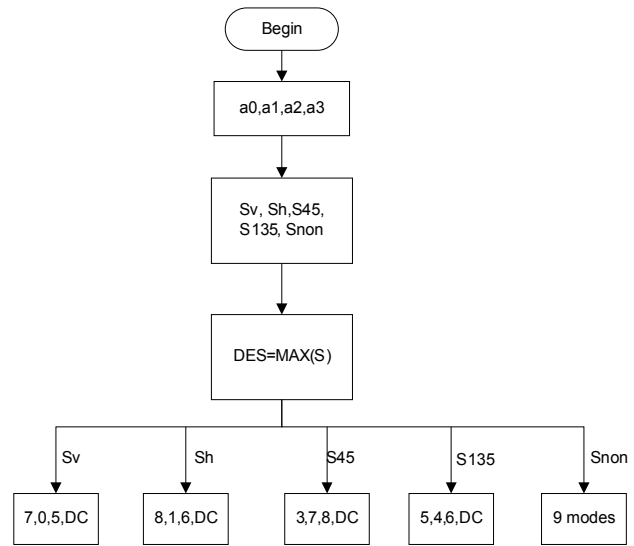
Giá trị các hệ số f như sau:

$f_0^v = 1$	$f_1^v = -1$	$f_0^h = 1$	$f_1^h = 1$	$f_0^{45^\circ} = \sqrt{2}$	$f_1^{45^\circ} = 0$	$f_0^{135^\circ} = 0$	$f_1^{135^\circ} = \sqrt{2}$	$f_0^{nd} = 2$	$f_1^{nd} = -2$
$f_2^v = 1$	$f_3^v = -1$	$f_2^h = -1$	$f_3^h = -1$	$f_2^{45^\circ} = 0$	$f_3^{45^\circ} = -\sqrt{2}$	$f_2^{135^\circ} = -\sqrt{2}$	$f_3^{135^\circ} = 0$	$f_2^{nd} = -2$	$f_3^{nd} = 2$

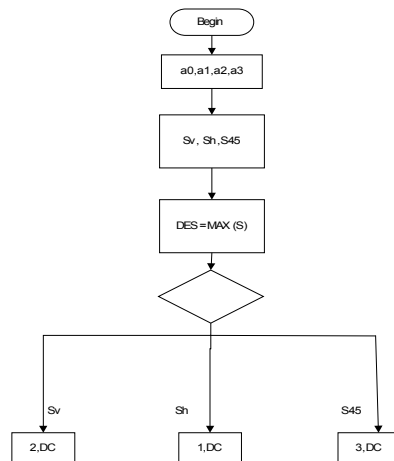
DES = max(S)

Dựa vào giá trị DES sẽ có thể chọn được các mode ứng cử viên so sánh và chọn ra mode tốt nhất (Lưu đồ thuật toán)

## Type 4x4:



## Type 16x16



#### 4. Lưu đồ thuật toán chung:

