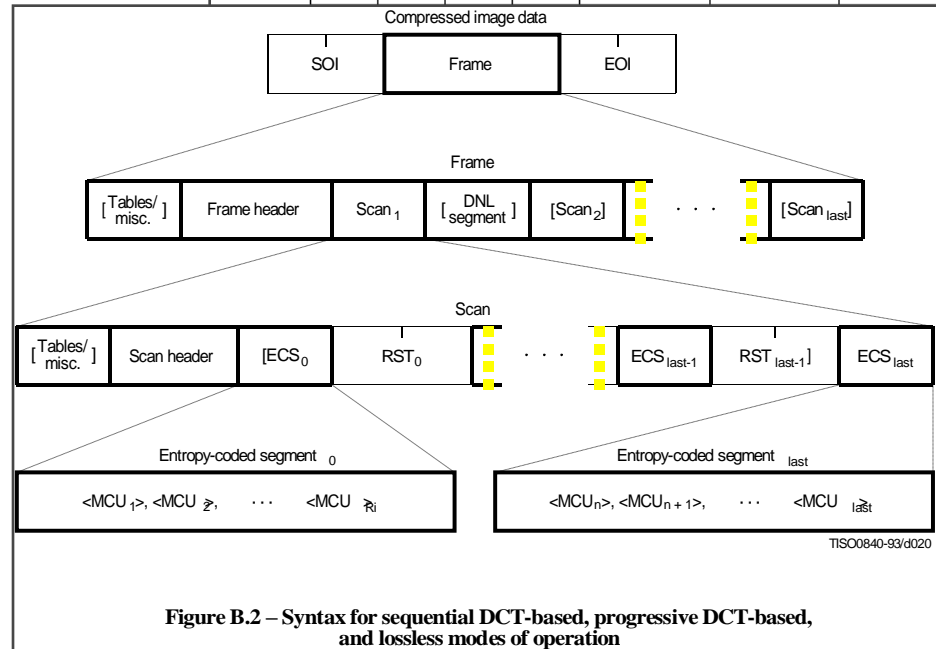


# JPEG Syntax and Data Organization

EC	D	MPS	CX	Qe (hexadecimal)	A (hexadecimal)	C (hexadecimal)	CT	ST	B
1	0	0		5A1D	0000	00000000	11	0	
2	0	0	CE	5A1D	A5E3	00000000	11	0	
3	0	0		2586	B43A	0000978C	10	0	
4	0	0		2586	8EB4	0000978C	10	0	
5	0	0		1114	D25C	00012F18	9	0	

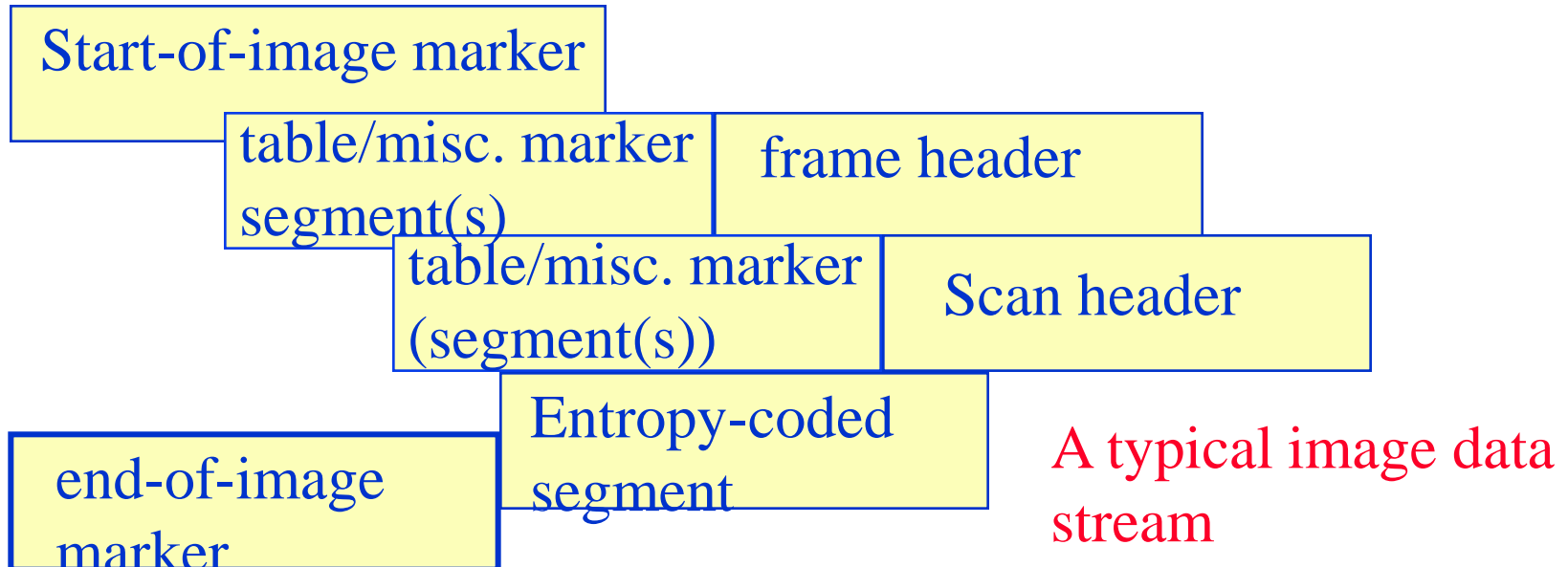


Nimrod Peleg  
Update: Dec. 2006

Figure B.2 – Syntax for sequential DCT-based, progressive DCT-based, and lossless modes of operation

# Control and Data Structure

- Two classes of segments:
  - Entropy coded segments: data
  - Marker segments: headers, tables and other general information
- Markers start with a unique two-byte code.



# Interchange/Abbreviated Data Formats

3 formats for JPEG compressed data:

- Interchange format for compressed data:  
includes **all** required tables (for the decoder)
- abbreviated format for compressed data:  
**may omit some or all tables** (decoder must have them some other way)
- abbreviated format for table specification:  
**only tables are sent** (no frames, entropy coded data etc.)

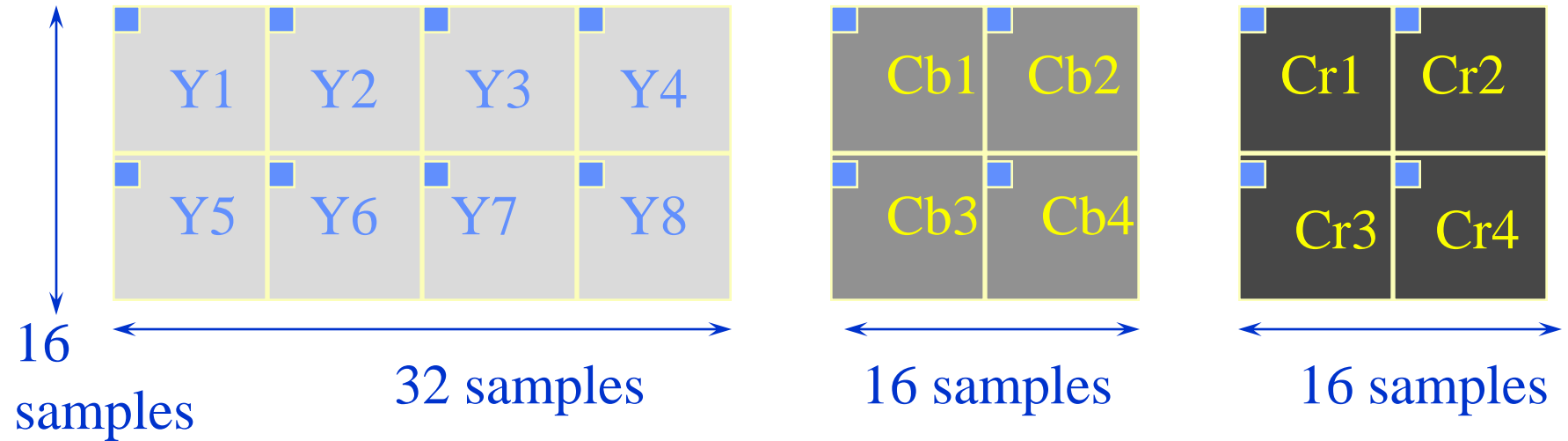
# Image Data Ordering

- Internal representation:
  - Upto 255 unique components for each image
  - Each component is represented as a rectangular array of samples
  - All processing of those rectangulars are from left to right, top to bottom
- Those are merely convenience, except the rectangular array which must be.

# Data Units

- For lossless modes: arrays are processed sample at a time, left-to-right, top-to-bottom
- For DCT modes: a block of 8x8 samples is a basic unit
- Minimum Coded Units (MCU), are groups containing interleaved/non-interleaved data from different components (if more than one).

# MCU Example



- Each block is 8x8 samples
- The preceding DC value is always the predictor for the current one (first is 0)

# Data units ordering example:

## Non-Interleaved data units

### Component block MCU

Scan 1:      Y1    1  
              Y2    2  
              .  
              Y8    8

Scan 2:      Cb1   1  
              .  
              Cb4   4

Scan 3:      Cr1   1  
              .  
              Cr4   4

## Horizontally Interleaved

### Component block MCU

Scan 1:      Y1    1  
              Y2    1  
              Cb1   1  
              Cr1   1

.  
. .  
Y7    4

Y8    4  
Cb4   4  
Cr4   4

# Marker Definitions

- Each **marker segment** begins with X'FF and a non-zero one byte 'marker code' to identify its function.
- An occasionally created X'FF (in the entropy coded data) is followed by a **stuffed zero byte**.
- All marker segments and entropy coded segments contain an **integer number of bytes**, so in Huffman coding one-bits used to pad data to achieve byte alignment for the next marker.



# Start-of-Frame (SOF) Markers

For example: **Huffman coded frames:**

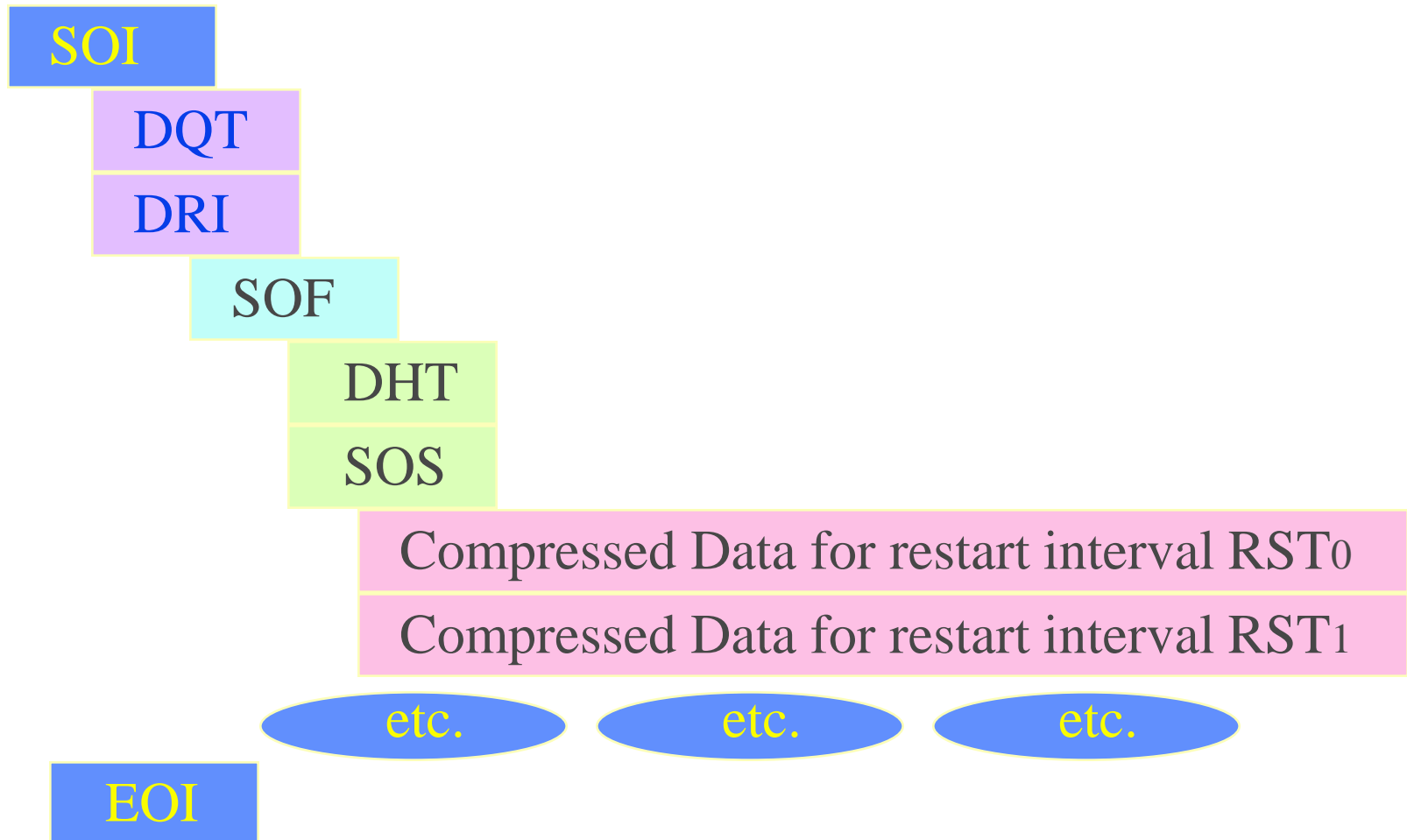
SOF <sub>0</sub>	X'FFC0	Baseline DCT
SOF <sub>1</sub>	X'FFC1	Extended Sequential DCT
SOF <sub>2</sub>	X'FFC2	Progressive DCT
SOF <sub>3</sub>	X'FFC3	Lossless (Sequential)
....		
SOF <sub>15</sub>	X'FFCF	Differential Lossless (Arithmetic Coding)

# Non-SOF Markers

- APP            X'FFE0 - X'FFEF    Reserved for application use
- DHT            X'FFC4            Define Huffman table
- DQT            X'FFD4            Define Quant. table
- SOS            X'FFDA            Start of Scan
  
- (About 20 non-SOF markers)

# Structure of Compressed Data

- Example of non-hierarchical data:



# Frame Header

- *Frame:* Basic attributes of the image

Frame header length                      16 bits

Sample precision                              8

Number of lines                              16

Number of samples/line                      16

Number of components                      8

## Frame component specification

Identifier                                      8

Horiz. sampling factor                      4

Vertical sampling factor                      4

Quantization table                      8 (4 optional tables)

# Scan Header

- **Many scans** can occur in a frame
- If scan has only one component than data is **non-interleaved** (MCU contains one data unit)
- If more than one component - **interleaved** data
- Components are always **coded independently**
- Scan parameters are: Number of components, DC and AC entropy coding table, Spectral selection and successive approximation parameters etc.

# Number of Data Units in MCU

- Data Unit = Block or samples (in lossless)
- If more than one component in a scan, the total number of data units in the MCU up to 10 units.
- Any combination of components and sampling factors that gives more than 10 is forbidden.

# Other Markers

- Define Huffman Table (DHT) segment
- Arithmetic Conditioning table (DAC) segment
- Quantization table (DQT) segment
- Restart interval (DRI) segment

# Web sites & Reviews

- Official site of JPEG group:

<http://www.jpeg.org/>

- JPEG FAQ:

<http://www.faqs.org/faqs/jpeg-faq/>

Wallace, K. Gregory,

The JPEG Still Picture Compression Standard,

Commun. of the ACM. 34:4, 1991, pp.30–44.

Furht B.

A Survey of Multimedia Compression Techniques and Standards.

Part I: JPEG Standard, Journal of Real-Time Imaging,

vol. 1, no. 1, April 1995, pp. 49-67.