

# File Systems

---

Martin Děcký

DEPARTMENT OF DISTRIBUTED AND DEPENDABLE SYSTEMS

<http://d3s.mff.cuni.cz/>

CHARLES UNIVERSITY IN PRAGUE  
FACULTY OF MATHEMATICS AND PHYSICS

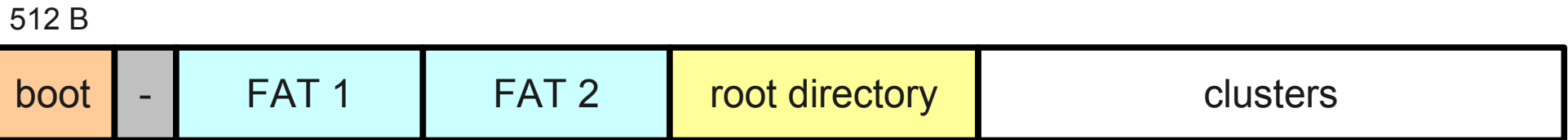


# MS-DOS FAT

- History
  - 1977
    - FAT12 (Gates, McDonald)
  - 1983
    - Subdirectories
  - 1988
    - FAT16
  - 1995
    - Long File Names
  - 1996
    - FAT32

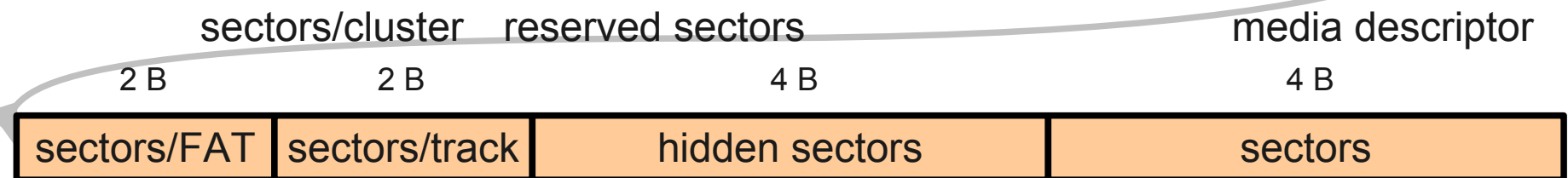
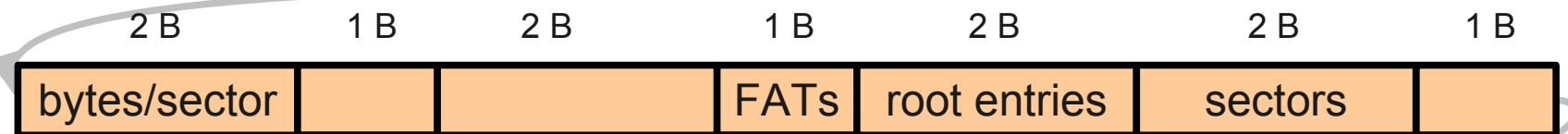


# Structure

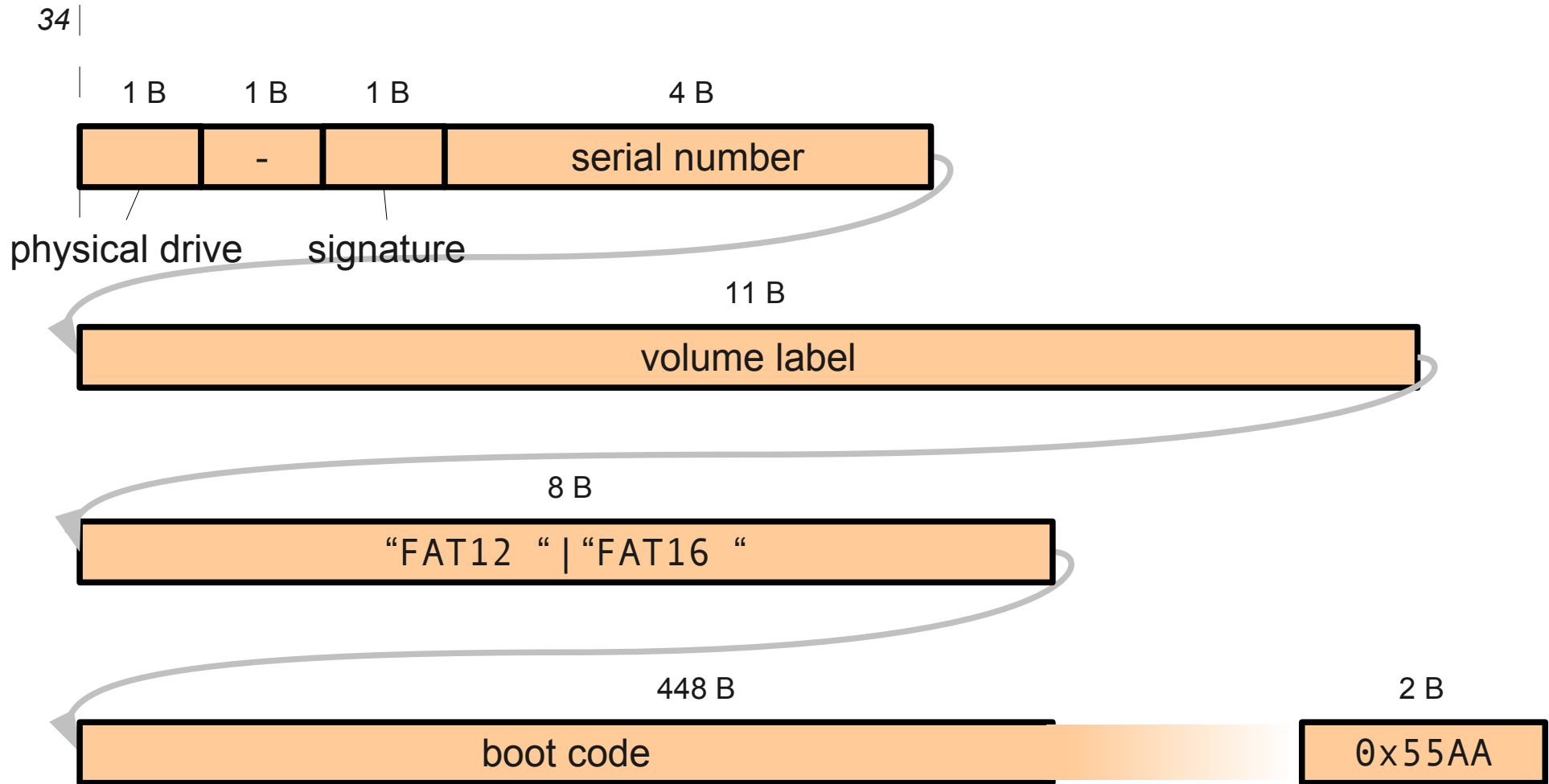


**boot sector**

FAT12, FAT16

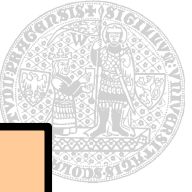
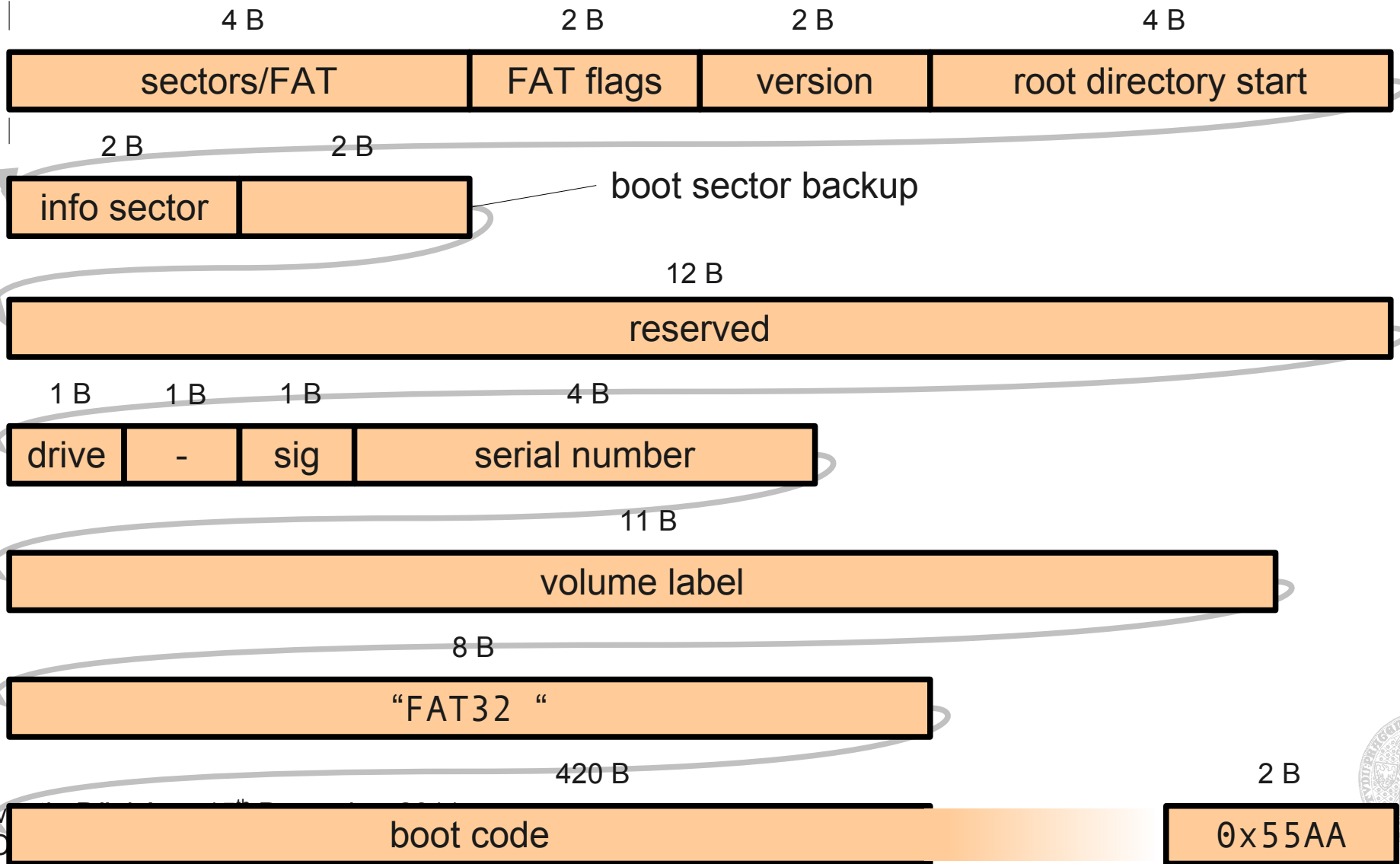


# BIOS Parameter Block



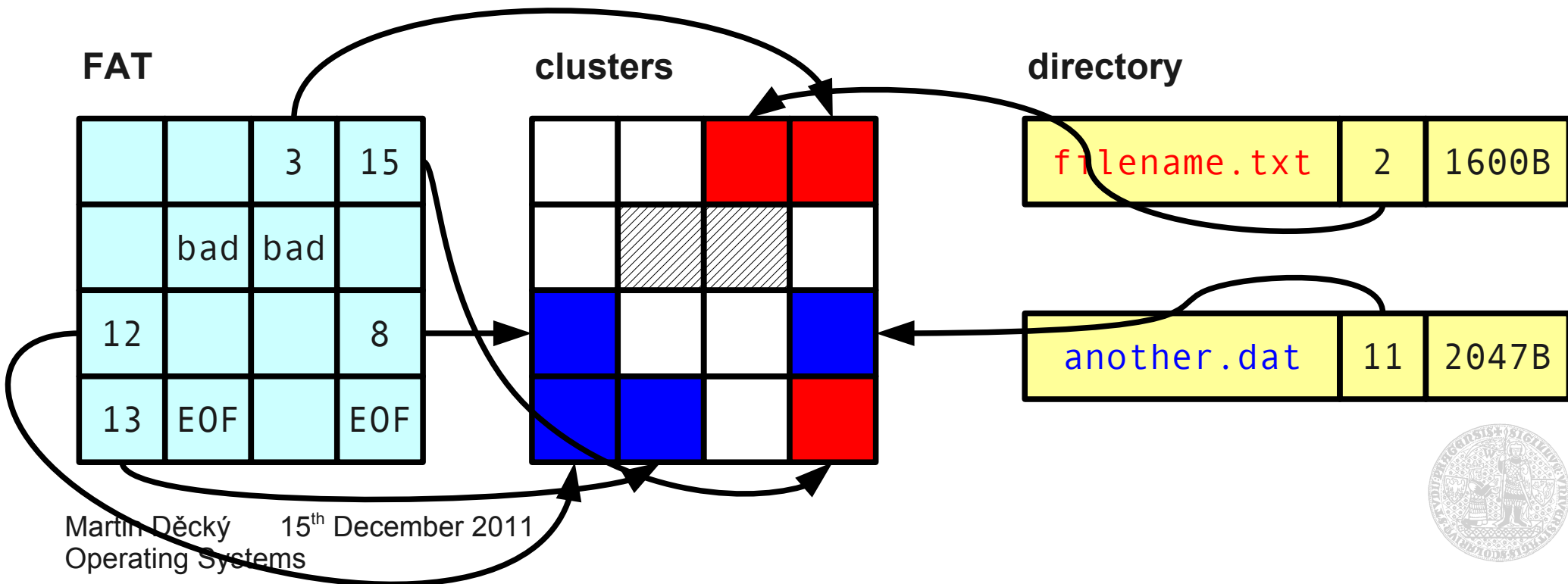
# Extended FAT32 Structure

34 |



# FAT

- Cluster: 512 B – 32 KB (64 KB)
  - File is a singly linked list of clusters
    - First cluster stored in directory record
    - Linked list stored in FAT



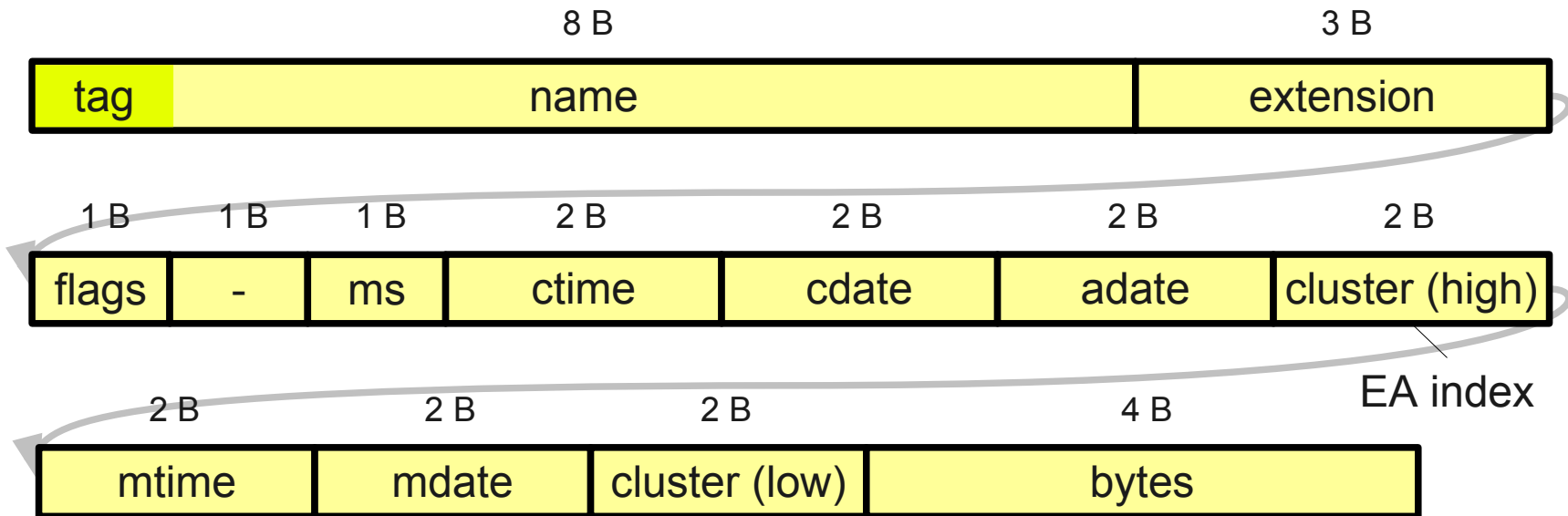
# Cluster Numbers

	FAT12	FAT16	FAT32
free cluster	0x000	0x0000	0x_0000000
reserved	0x001	0x0001	0x_0000001
next cluster	0x002 - 0xFE	0x0002 - 0xFFE	0x_0000002 - 0x_FFFFFFFE
reserved	0xFF0 - 0xFF6	0xFFFF0 - 0xFFFF6	0x_FFFFFFF0 - 0x_FFFFFFF6
bad cluster	0xFF7	0xFFFF7	0x_FFFFFFF7
EOF	0xFF8 - 0xFFF	0xFFFF8 - 0xFFFFF	0x_FFFFFFF8 - 0x_FFFFFFFF

- Cluster 0
  - Copy of media descriptor
- Cluster 1
  - EOF, dirty bit



# Directory Entry



- **flags**

- bit 0: read-only
- bit 1: hidden
- bit 2: system
- bit 3: volume label
- bit 4: directory
- bit 5: archive
- bit 6: *device*
- bit 7: unused

- **tag**

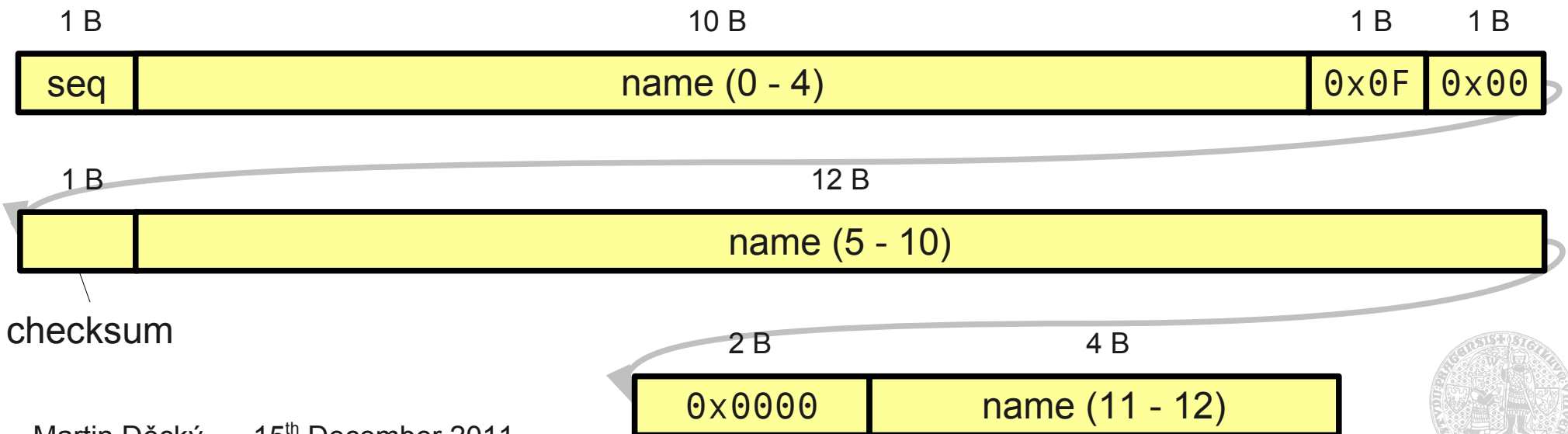
- 0x00: last & empty entry
- 0x05: alias for 0xE5
- 0x2E: special entry (., ..)
- 0xE5: erased entry





# Long File Names

- Backward compatible
  - Stored as Volume label directory entry
    - Each entry has space for 13 UTF-16 characters
    - Multiple entries chained
  - Paired with 8+3 name using a checksum

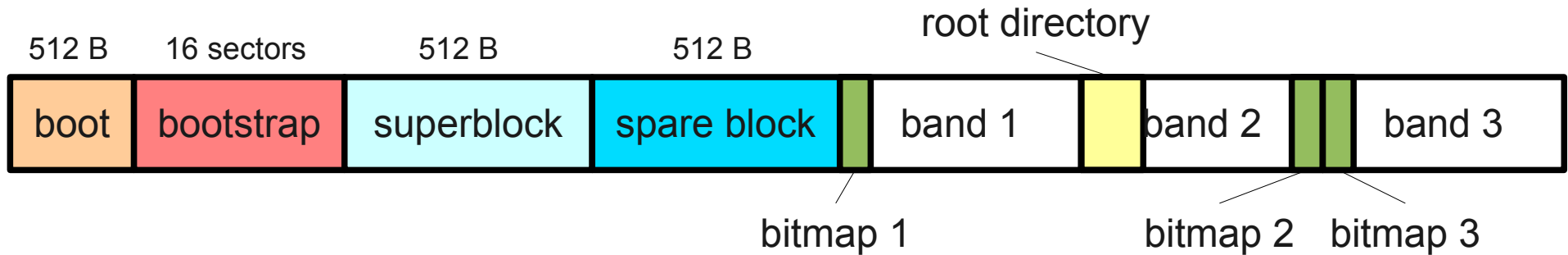


# HPFS

- 1989 (OS/2 1.2)
  - 254 characters in file name
  - Data in sectors
  - Extents (runs) of sectors
  - Bands
  - Bitmap of free sectors
  - Directories as B+ trees
  - Root directory in the middle of the disk
  - Extended attributes (64 KB per file)



# Structure

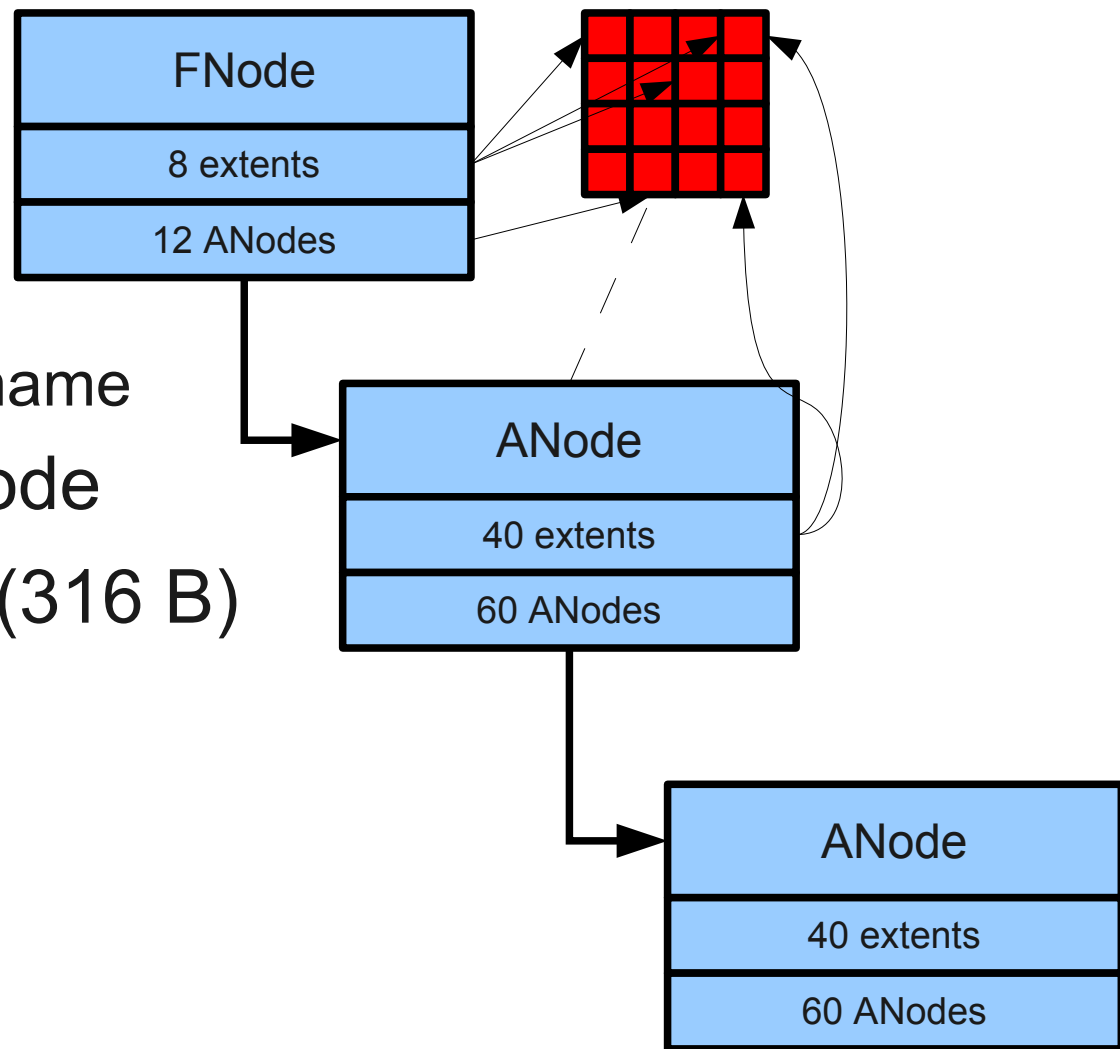


- boot block
  - BIOS Parameter Block
  - volume name
  - serial number
- superblock
  - root directory FNode sector
  - size of the filesystem
  - pointer to list of bitmap sectors
  - pointer to list of bad sectors
  - directory band
- list of bitmap sectors
  - 4 sectors containing the list of groups
  - 4 sectors in each group (16384 bits)
    - Band = 8 MB
- list of bad sectors
  - linked list
- spare block
  - hotfix list
  - code page information sector



# FNode

- Represents a file
  - Recovery measures
    - Length of file name
    - 15 characters from name
  - Parent directory FNode
  - Extended attributes (316 B)
  - Sector information
    - 8 extents (runs)
    - 12 ANodes



# Directory Structure

- Directory Block
  - 4 sectors
  - Primarily in directory band
  - Variable entries
    - size of entry
    - flags, attributes
    - FNode
    - modification date/time
    - access date/time
    - creation date/time
- length
- size of extended attributes
- file name code page
- length of file name
- file name
- Overwritten on deletion
- Stored in B+ tree



# NTFS

- 1993 (Windows NT 3.1)
  - 255 characters in file name (UTF-16)
  - Data in clusters (512 B – 64 KB)
  - Extents (runs) of clusters, sparse files
  - Bitmap of free clusters
  - Directories as B+ trees
  - Transparent compression, encryption
  - Journaling, ACLs, extended attributes
  - *Everything is a file, everything is an attribute*



# Structure

1 cluster



## **\$MFT**

records for all files (inodes)

## **\$MFTMirr**

backup of first 4 records of MFT

## **\$LogFile**

transaction journal

## **\$Volume**

information about the file system

## **\$Bitmap**

bitmap of free clusters

## **\$Boot**

boot cluster

jump, "NTFS ", bytes per sector, sectors per cluster, first cluster of \$MFT, first cluster of \$MFTMirr, ...

## **\$BadClus**



# File Attributes

## **\$STANDARD\_INFORMATION**

ctime, atime, mtime, rtime

file permissions

r/o, hidden, system, archive, device, normal, temporary, sparse, reparse point,  
compressed, offline, not indexed, encrypted

version

owner

...

## **\$ATTRIBUTE\_LIST**

type

length

name of attribute

starting cluster (0 for resident attribute)

...

## **\$FILE\_NAME**

parent directory, size, ...

## **\$DATA**

## **\$BITMAP**

## **\$SYMBOLIC\_LINK**





# Data Runs

- Each non-resident attribute stored in runs of clusters
  - Set of runs stored in Run Elements

size of the offset field ( $F$ )	1 B
size of the length field ( $L$ )	1 B
length of the run	$2L$ B
Offset from previous element	$2F$ B



# Inode

- File record
  - Magic identifier
  - Update sequence numbers
  - Hard link count
  - File size
  - Next attribute ID

