

File System Data Structures

Pt. 2 - ext

CSC 712 – Data Structures

Goals

- Last lecture we learned about data structures used in the FAT file system.
- But FAT is old and somewhat inefficient
- Let's learn about a modern file system, ext!

File Systems

- *FAT*
 - NTFS
 - HFS
 - *ext (2, 3, 4)*
 - Btrfs
 - ...
- ReiserFS
 - XFS
 - JFS
 - Iso9660
 - NFS
 - ...

Problems with FAT

- In order to “seek” within a file, we must walk the whole FAT chain for that file.
- In order to find free blocks, we must scan the whole FAT

Improvements in ext

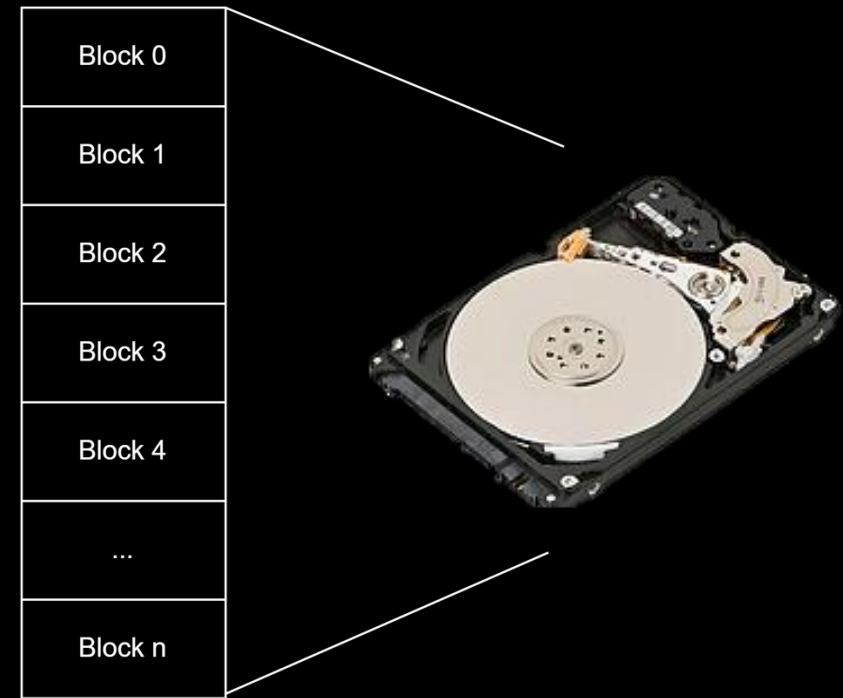
- In order to “seek” within a file, we must walk the whole FAT chain for that file.
 - Ext uses indexing! Index nodes (“inodes”) contain many pointers to file chunks
- In order to find free blocks, we must scan the whole FAT
 - Ext uses bitmaps to indicate block usage, which is much more efficient to search.

Disclaimer

- Again, I'll show you a rough approximation of how ext works
 - Conceptual ideas, not precise technical details
 - Ext has MAAAANY more features than this

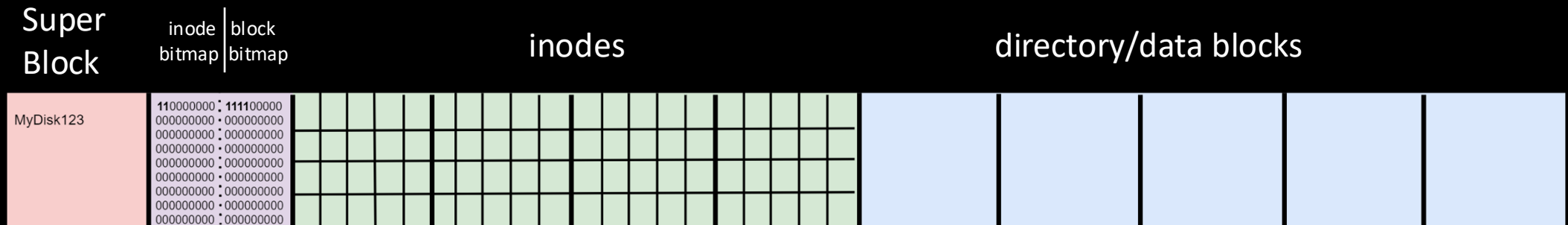
Blocks

- Recall that disks are accessed by block
- Common block sizes
 - 512, 1024 (1K), 4096 (4K)
- Ext usually uses $bs=4096$
- We will still use $bs=512$ for simplicity :)



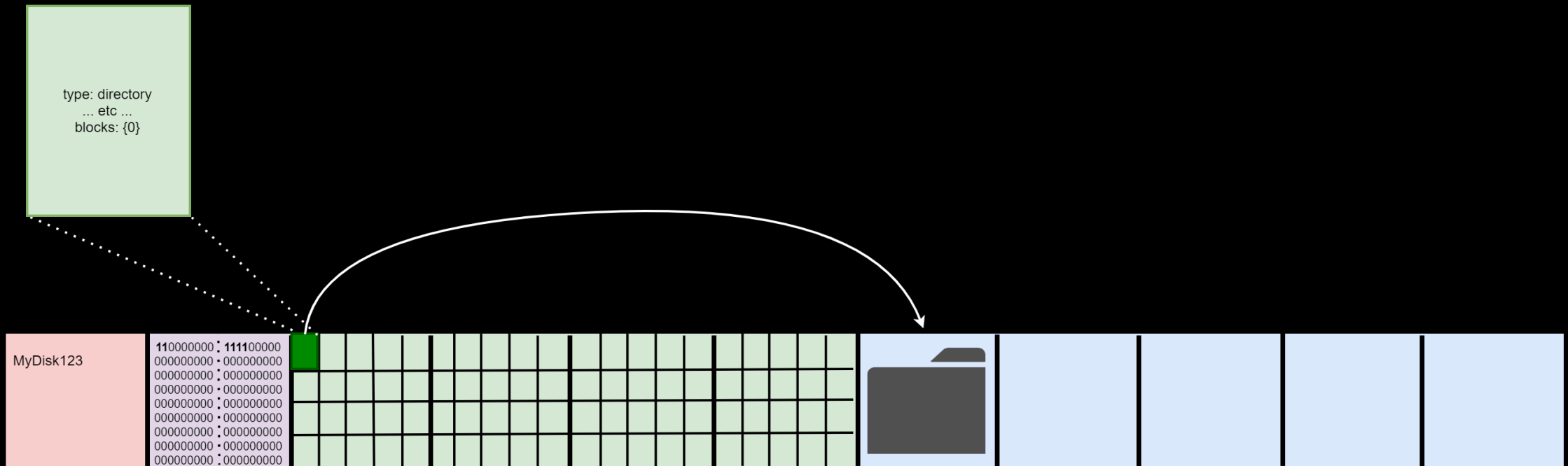
High Level ext Structure

- Superblock
 - Metadata about the disk. Label, telemetry, etc
- Bitmap block(s)
 - inode bitmap / block bitmap, showing free or used
- Inode block(s)
 - Each inode represents one file or directory, and holds its metadata
- Data blocks
 - Actual file data or directory content list



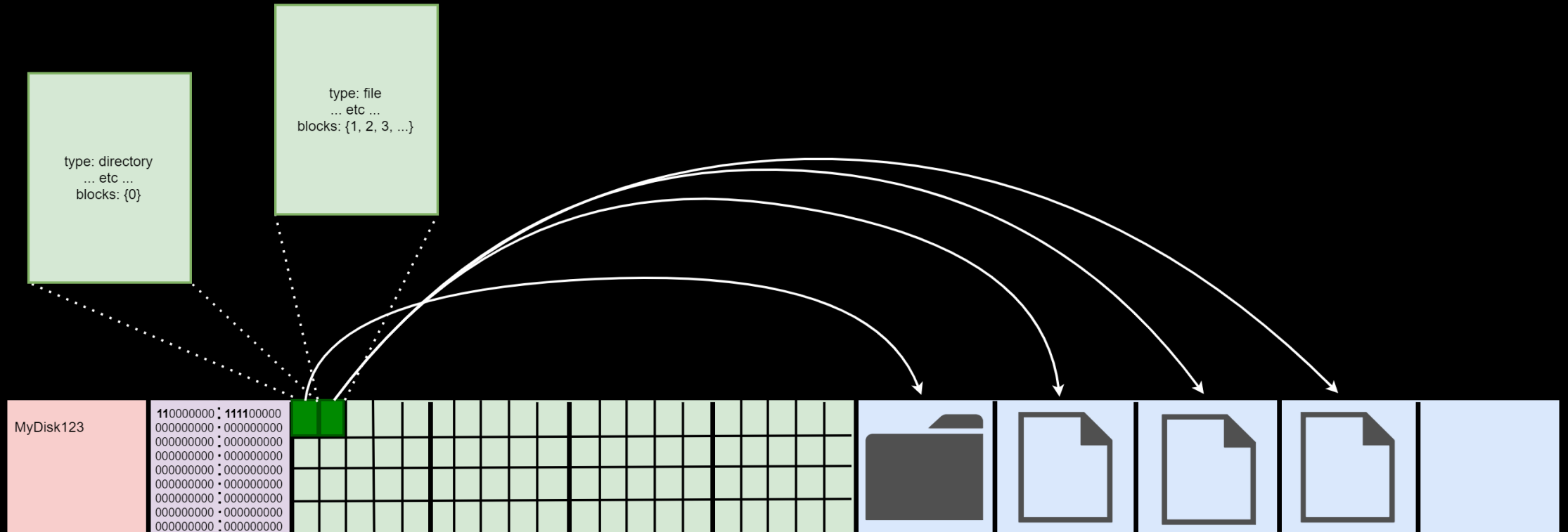
Inodes

- inodes represent files/directories
- Also contain meta-data about that file/directory



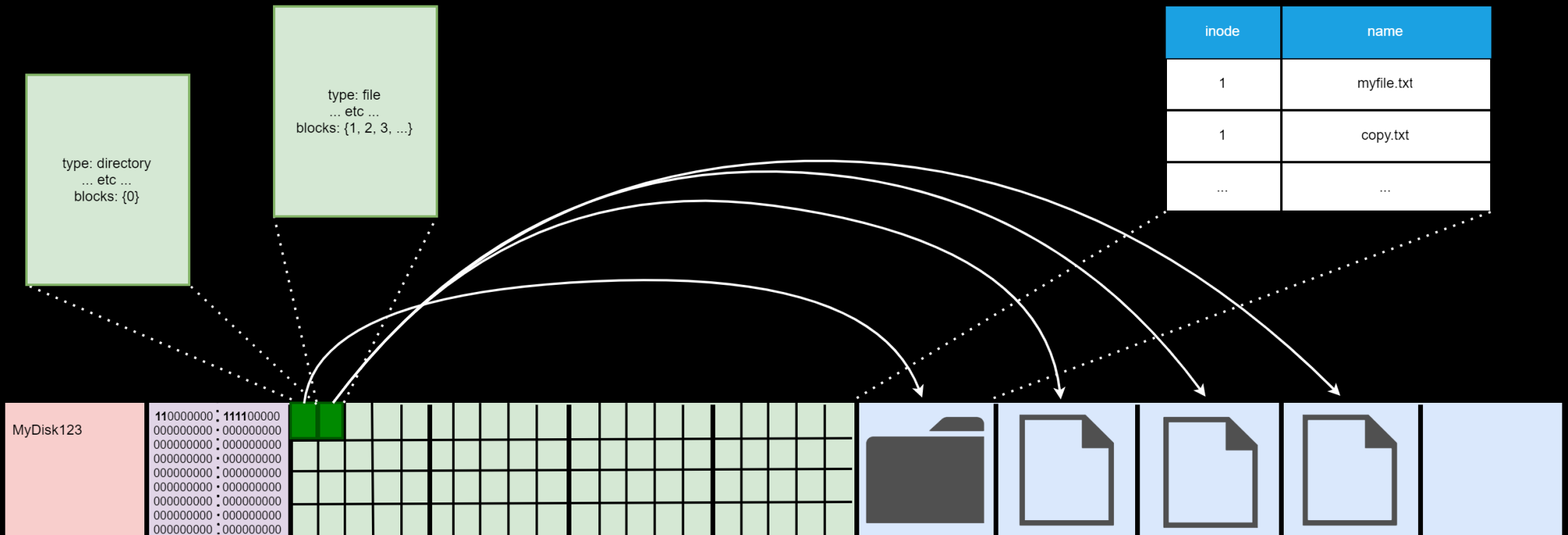
Files

- May span one or more blocks



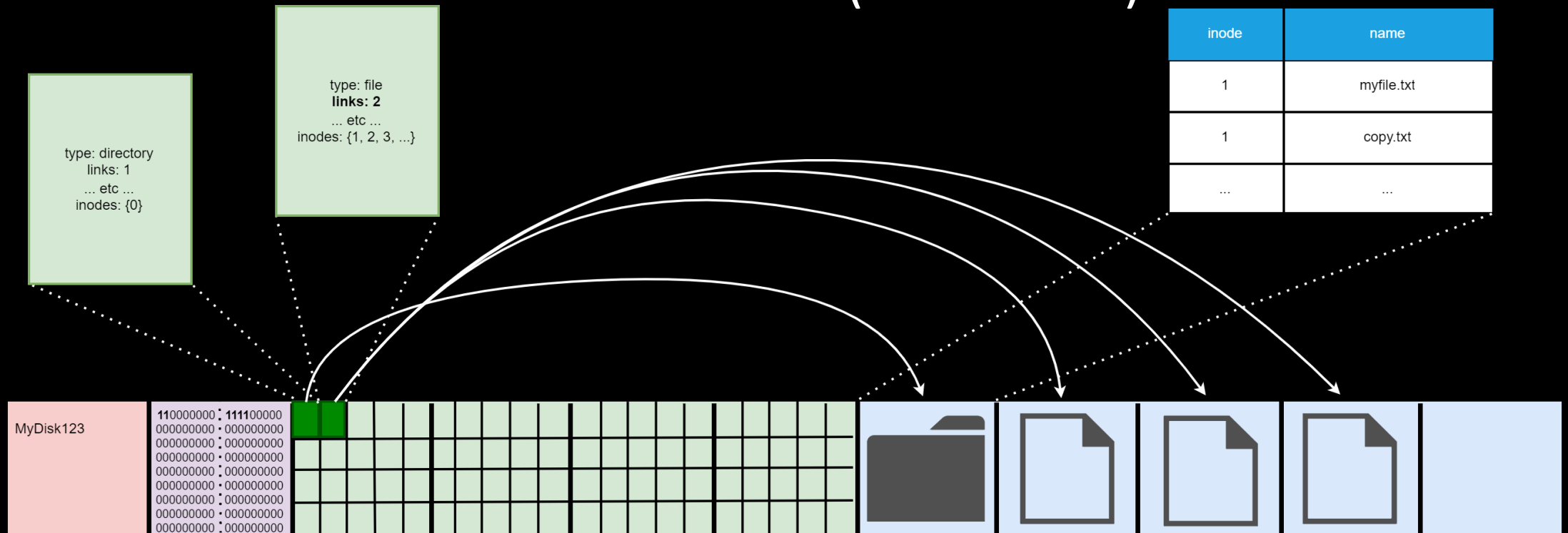
Directory Entries

- Directories hold inode:name pairs



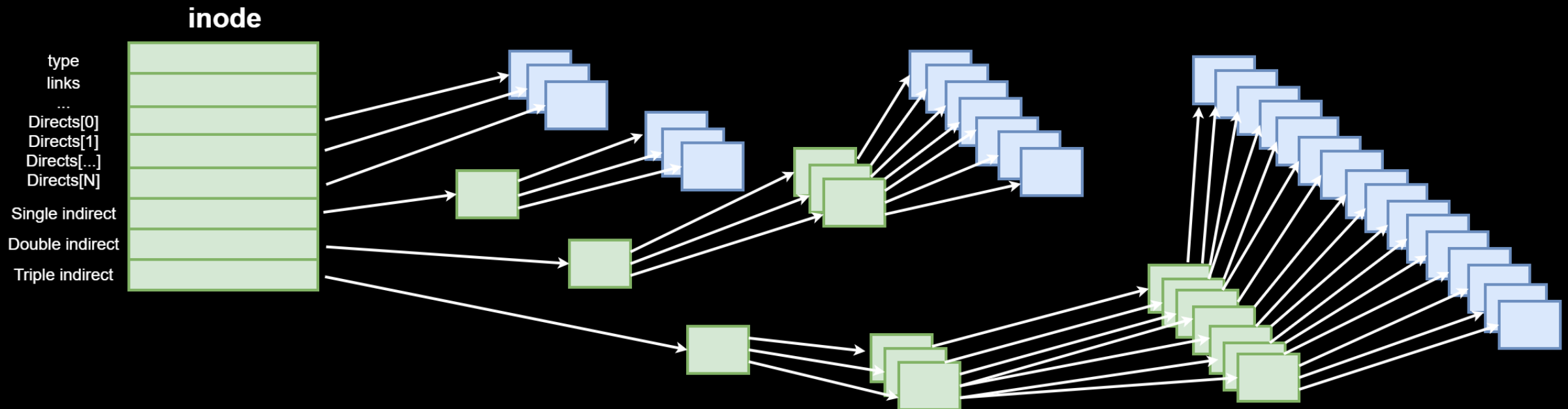
Hard Links

- Many directory listings may point to the same inode! :)
- Need to track number of links (references)



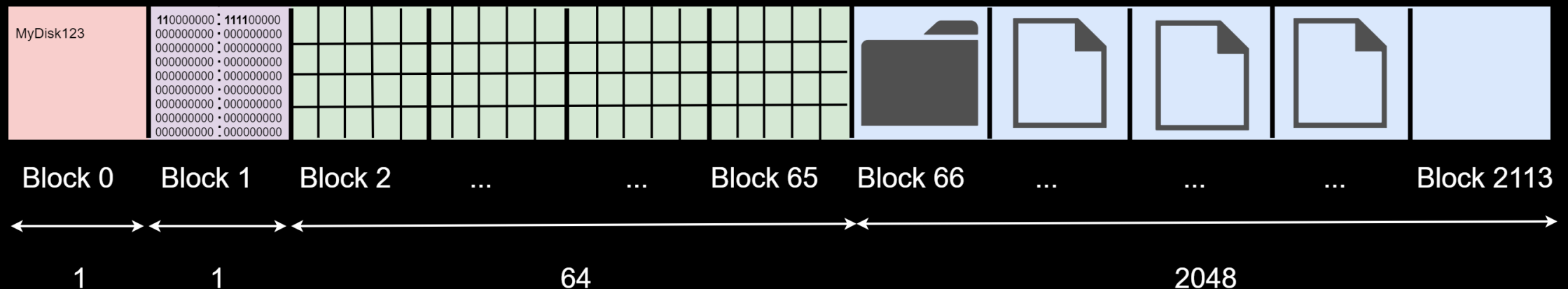
Block Indirection

- Files might be huge! How do we point to 1TB of blocks using only one inode?
 - directs = KB, single = MB, double = GB, triple = TB



Our Ext implementation (bs=512)

- Block 0: superblock
- Block 1: bitmaps $512 * 8 = 4096$ (2048 inodes & 2048 data)
- Blocks 2-65: inodes size = 16 ($16 * 2048 = 32768$, $32768 / 512 = 64$)
- Blocks 66-2113: data



Our Inodes

- Type: 2 bytes
- Links: 2 bytes
- Size: 4 bytes
- Directs[3]: $3 * 2 \text{ bytes} = 6 \text{ bytes}$
- Indirects: 2 bytes
- Total size = $2 + 2 + 4 + 6 + 2 = 16 \text{ bytes}$

Our Directory Entries

- inode: 2 bytes
- Name: 30 bytes
- Total size = 32
- $512 / 32 = 16$ directory listings per block

Block 0: Superblock

- Just contains the disk label
- Null terminated

| | | | | | | | | | | | | | | | | | |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|------------------|
| 00000000 | 45 | 58 | 54 | 5F | 44 | 69 | 73 | 6B | 5F | 66 | 6F | 72 | 5F | 41 | 73 | 73 | EXT_Disk_for_Ass |
| 00000010 | 69 | 67 | 6E | 6D | 65 | 6E | 74 | 5F | 35 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | ignment_5..... |
| 00000020 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000030 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000040 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000050 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000060 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000070 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000080 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000090 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000000A0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000000B0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000000C0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000000D0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000000E0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000000F0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000100 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000110 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000120 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000130 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000140 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000150 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000160 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000170 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000180 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 00000190 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000001A0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000001B0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000001C0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | |
| 000001D0 | 00 | 00 | 00 | 00 | 00 | | | | | | | | | | | | |

Block 1: Bitmaps

- 256 bytes (2048 bits) for inodes
- 256 bytes (2048 bits) for data
- $0x80 = 1\ 0\ 0\ 0\ 0\ 0\ \dots$
 - Root directory entry

[illegible]

Blocks 2-65 (64 total): inodes

- 0x1111 = TYPE_DIR
- 0x0001 = links
- 0x00000000 = size
- Directs[3]:
 - 0x0000: data 0 (block 66)
 - Unused
 - Unused
- Indirects: unused

[illegible]

Blocks 2-65 (64 total): inodes

- 0x2222 = TYPE_FILE
- 0x0001 = links
- 0x00000BB9 = size (3001)
- Directs[3]:
 - 0x0001: data 1 (block 67)
 - 0x0002: data 2 (block 68)
 - 0x0003: data 3 (block 69)
- Indirects: 0x0004
 - Data 4 (block 70) contains an array of 2-byte datablock numbers

[illegible]

Blocks 66-2113 (2048 total): data

- Entry 1
 - 0x0000: inode 0
 - 0x2e 0x00: “.”
- Entry 2
 - 0x0000: inode 0
 - 0x2e 0x2e 0x00: “..”
- Entry 3
 - 0x0001: inode 1
 - 0x67 0x6f ... : “lots_of_A.txt”
- Entry 4 - 15
 - 0xFFFF: unused

```

00008400 00 00 2E 00 00 00 00 00 00 00 00 00 00 00 00
00008410 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008420 00 00 2E 2E 00 00 00 00 00 00 00 00 00 00
00008430 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008440 01 00 6C 6F 74 73 5F 6F 66 5F 41 2E 74 78 74 00
00008450 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008460 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
00008470 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008480 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
00008490 00 00 00 00 00 00 00 00 00 00 00 00 00 00
000084A0 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
000084B0 00 00 00 00 00 00 00 00 00 00 00 00 00 00
000084C0 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
000084D0 00 00 00 00 00 00 00 00 00 00 00 00 00 00
000084E0 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
000084F0 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008500 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
00008510 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008520 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
00008530 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008540 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
00008550 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008560 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
00008570 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00008580 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
00008590 00 00 00 00 00 00 00 00 00 00 00 00 00 00
000085A0 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
000085B0 00 00 00 00 00 00 00 00 00 00 00 00 00 00
000085C0 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
000085D0 00 00 00 00 00 00 00 00 00 00 00 00 00 00
000085E0 FF FF 00 00 00 00 00 00 00 00 00 00 00 00
000085F0 00 00 00 00 00 00 00 00 00 00 00 00 00 00

```

Further Reading

<http://web.mit.edu/tytso/www/linux/ext2intro.html>

<https://www.kernel.org/doc/html/latest/filesystems/ext4/index.html>

Live Demo