



# ECFS

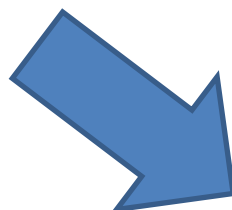
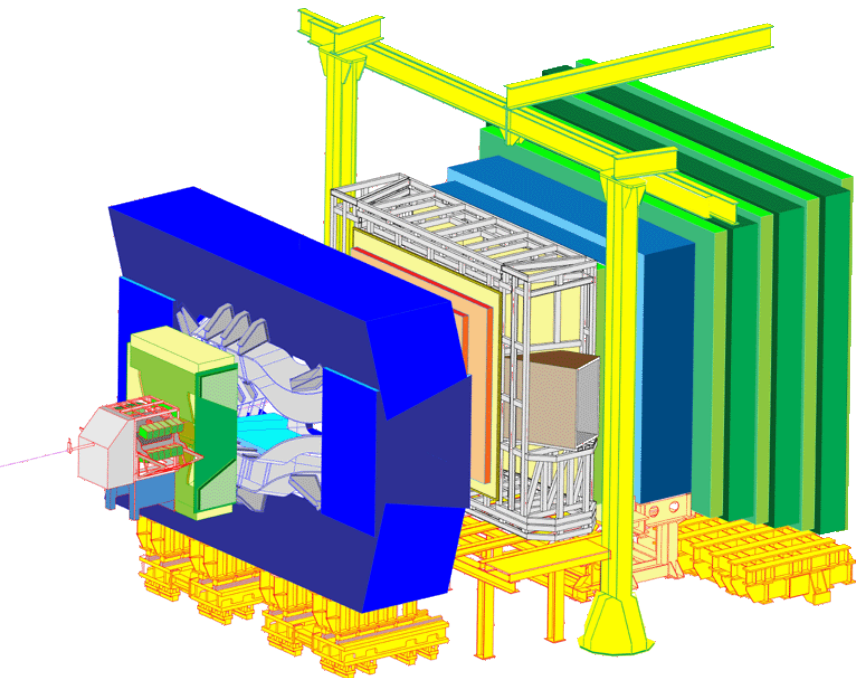
A decentralized, distributed and fault-tolerant FUSE filesystem for the LHCb online farm



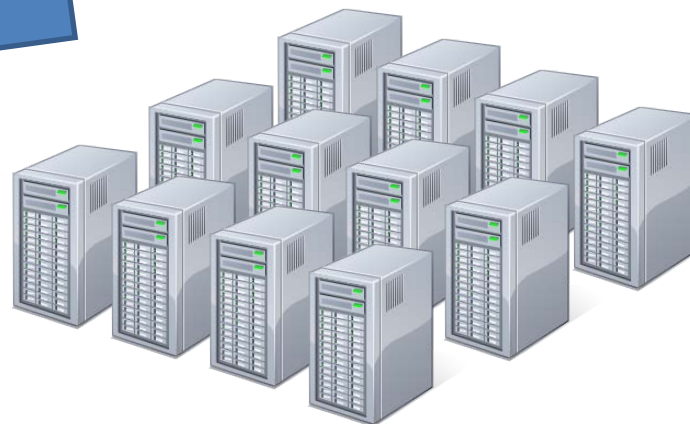
Session: Data Stores, Data Bases,  
and Storage Systems

tomasz.rybczynski@cern.ch\*  
niko.neufeld@cern.ch  
enrico.bonaccorsi@cern.ch

# Introduction



HLT farm  
~2000 nodes



# Requirements

- Data redundancy – fault tolerance
- Single namespace
- POSIX semantics
- Write once read many

# Finding solution

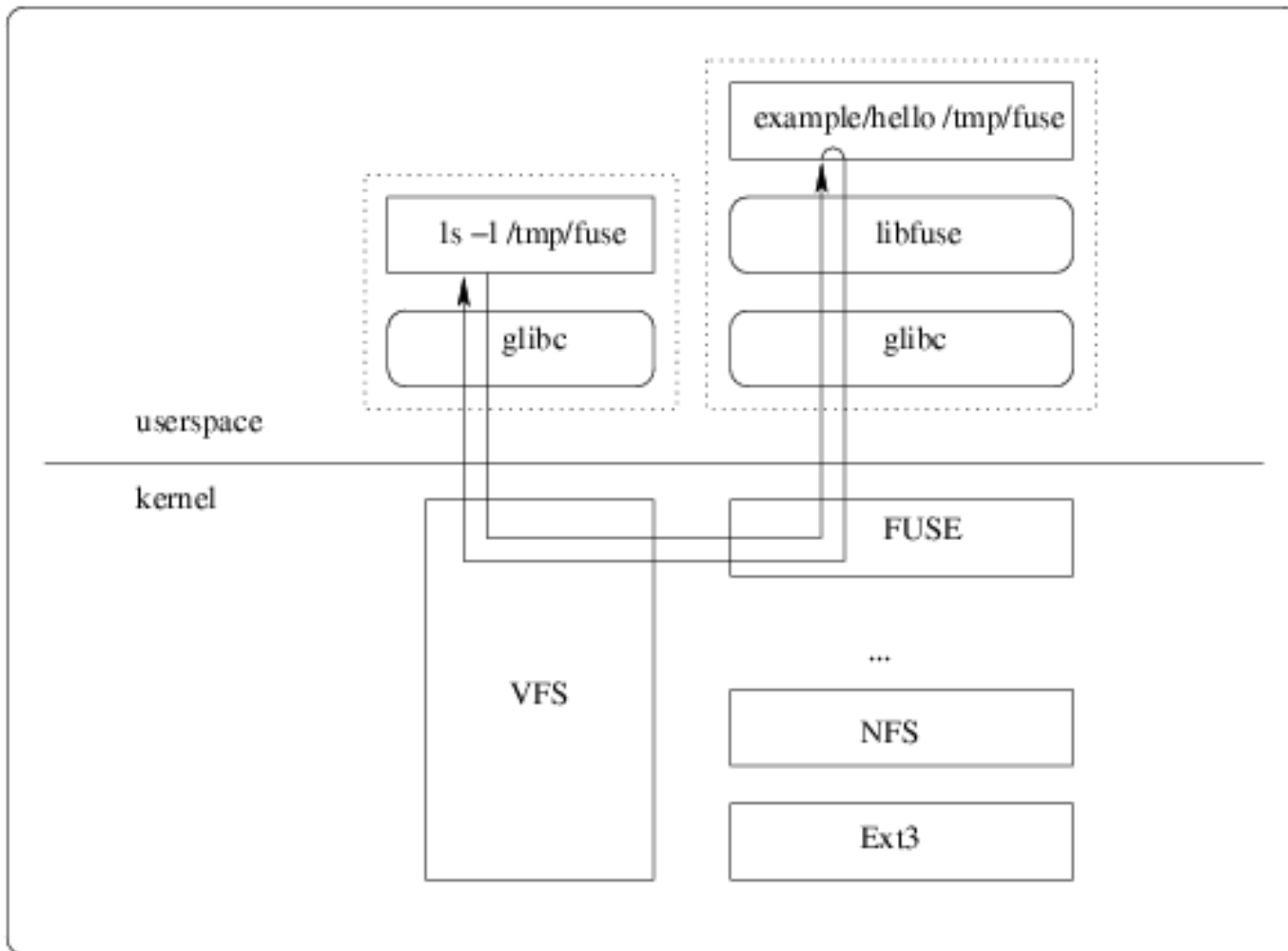
- Linux RAID over HA cluster
- GlusterFS
- Tahoe-LAFS [tahoe-lafs.org](http://tahoe-lafs.org)
- ...

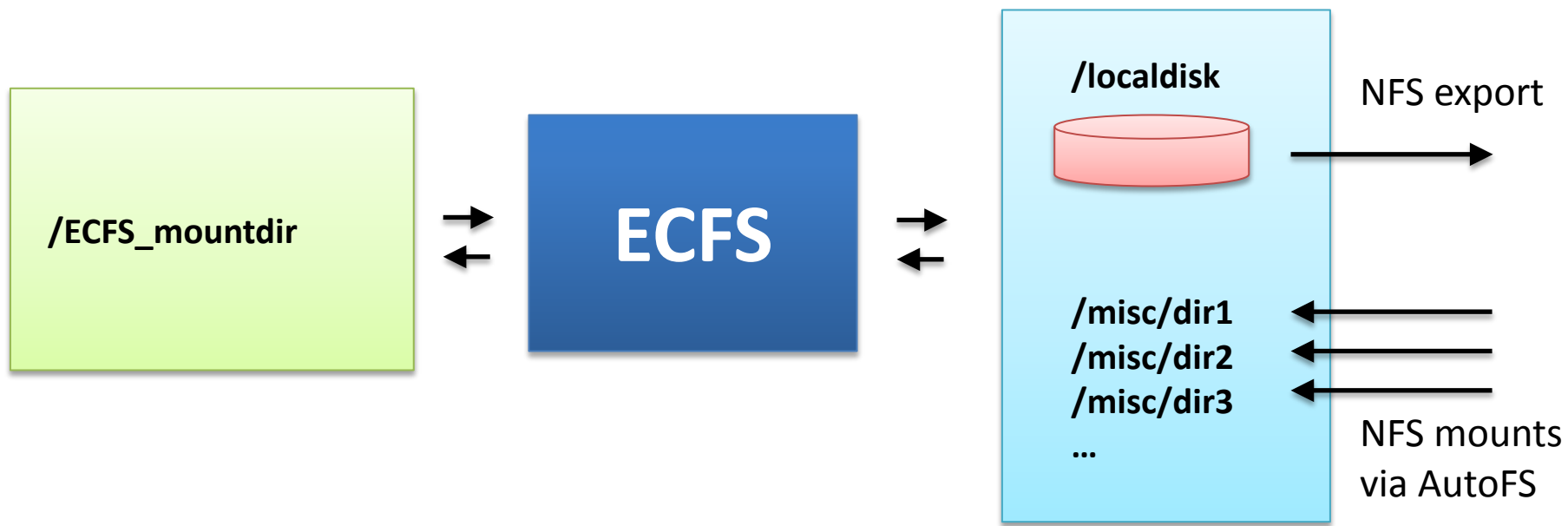
**ECFS** - distributed, fault-tolerant file system

Project based on:

- FUSE
- NFS
- Erasure Codes

# FUSE





**Coding parameters:**

$$k = 6$$

$$n = 2$$

Original message of a size of  $k$  symbols



**Encoding**



Code word of a size of  $k + n$  symbols

**Successful decoding possible if the number of missing symbols is less or equal  $n$**

Code word with some symbols missing



**Decoding**

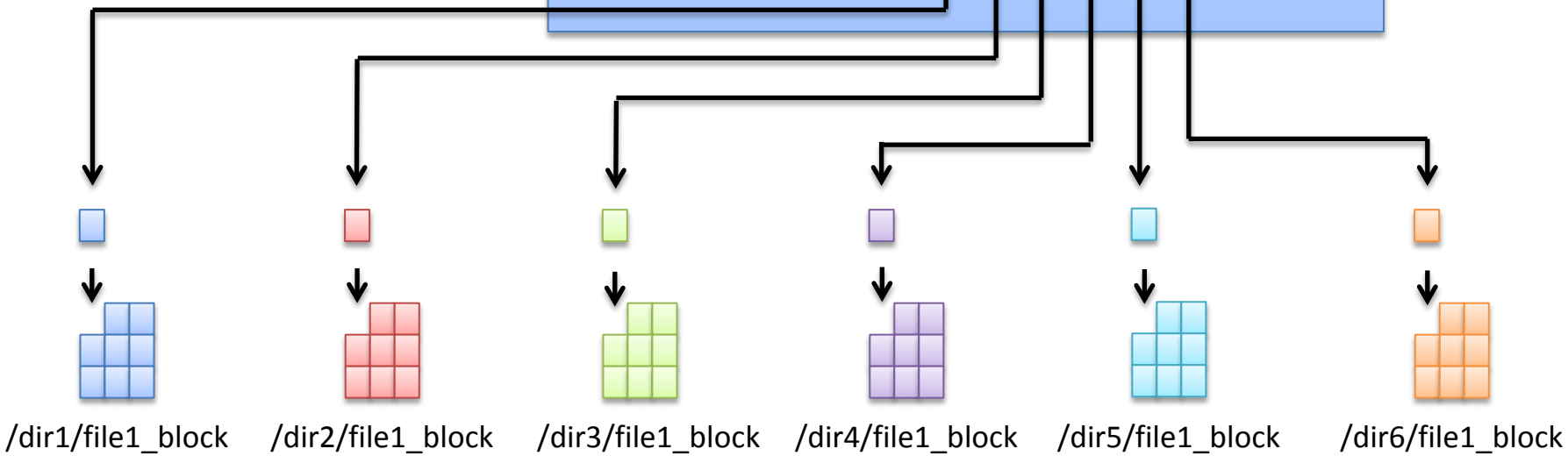
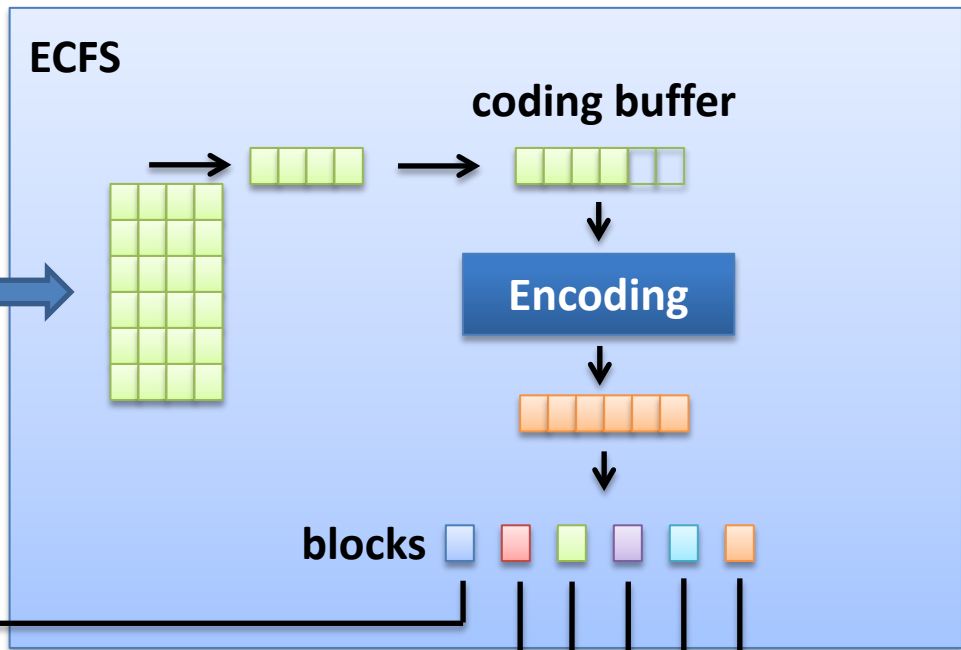


Original message



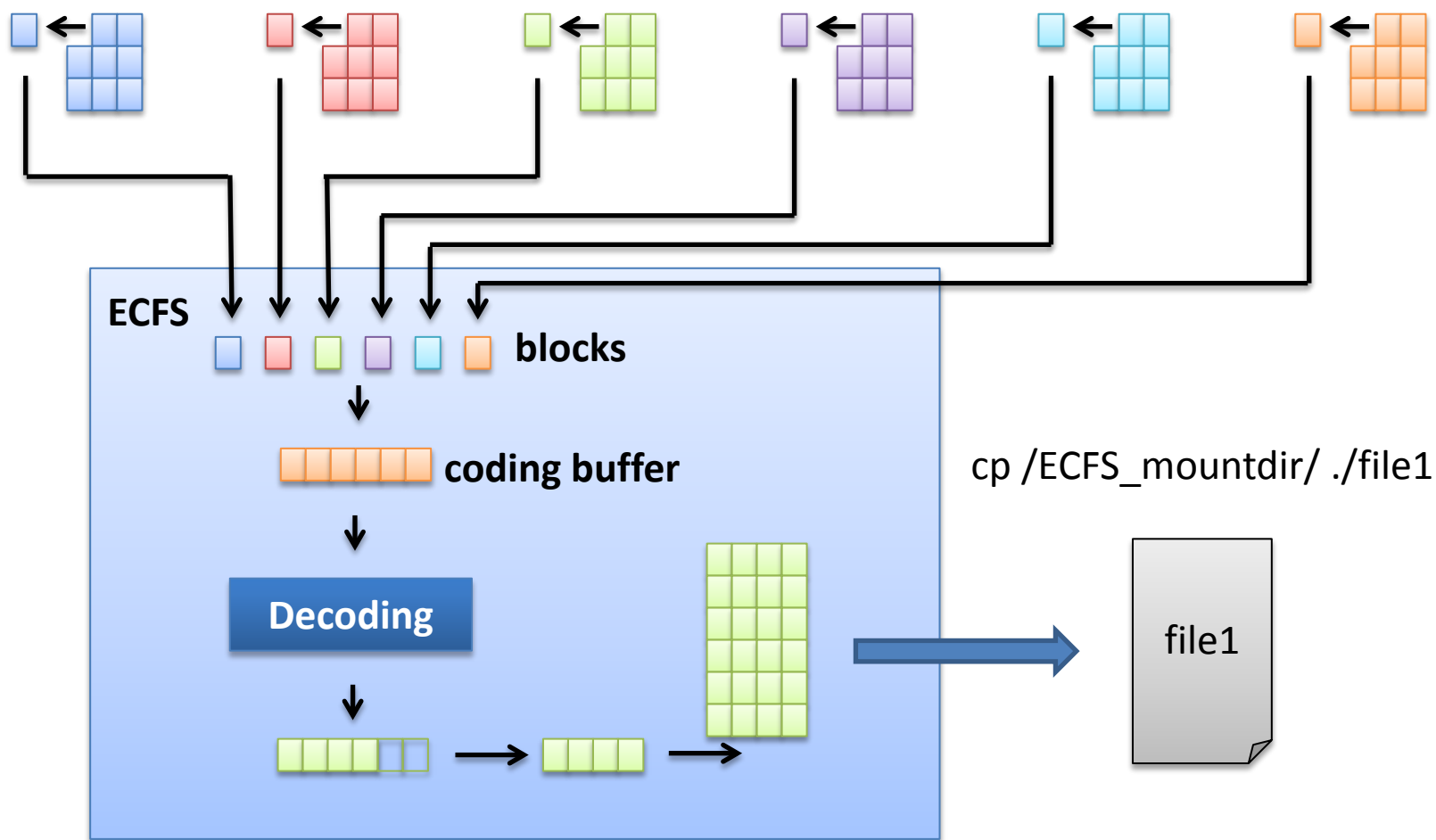
# File encoding

cp file\_1 /ECFS\_mountdir/

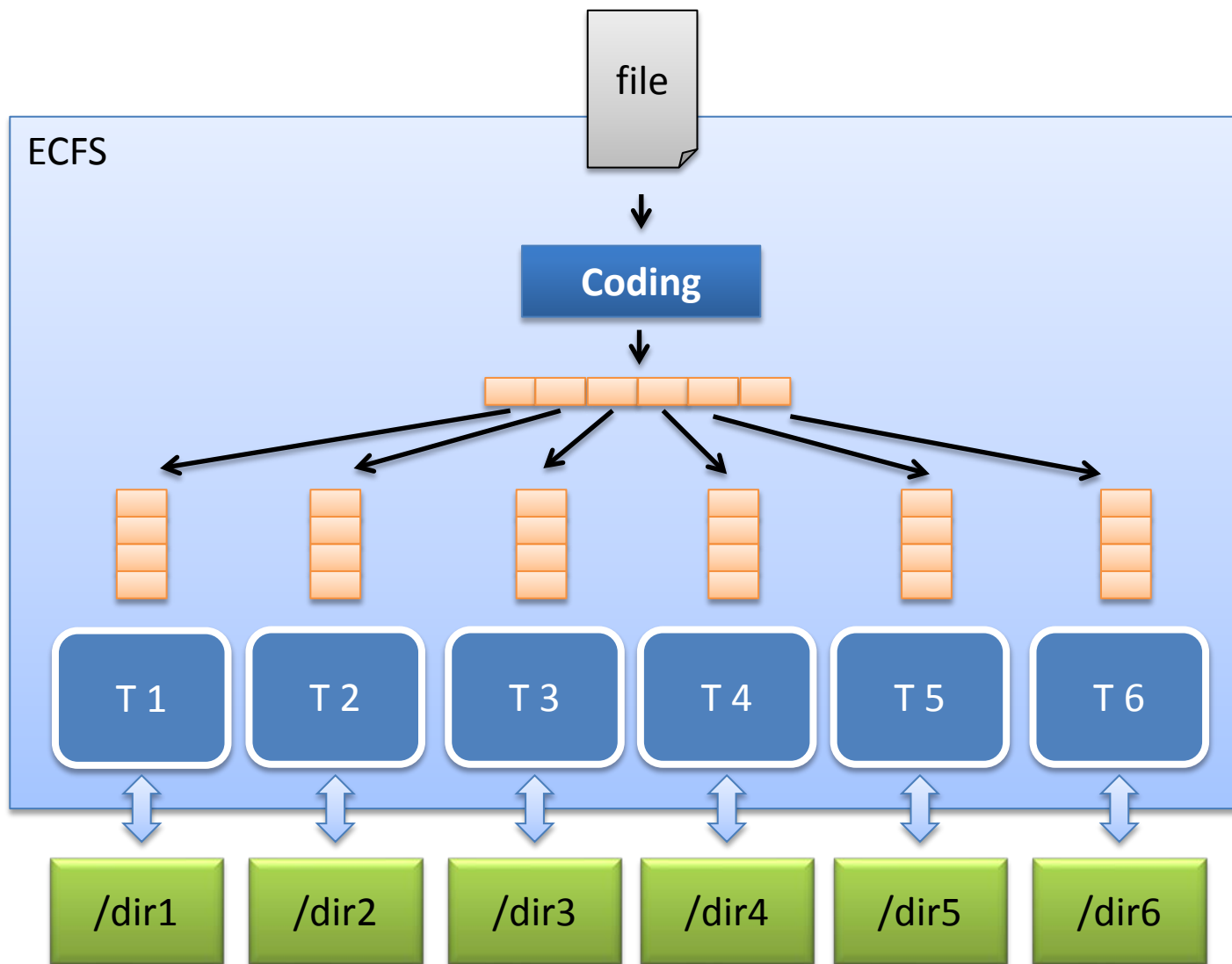


# File decoding

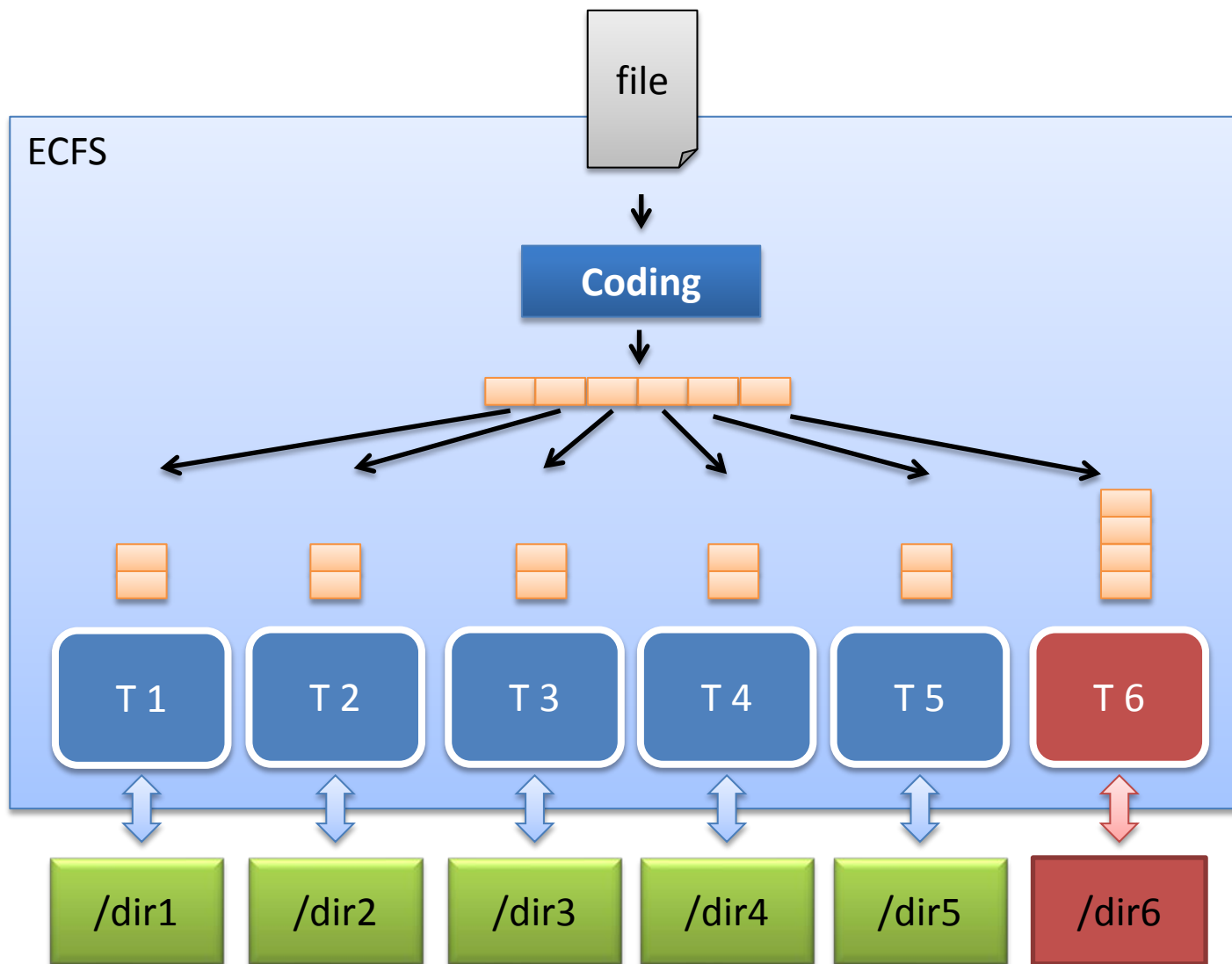
/dir1/file1\_block /dir2/file1\_block /dir3/file1\_block /dir4/file1\_block /dir5/file1\_block /dir6/file1\_block



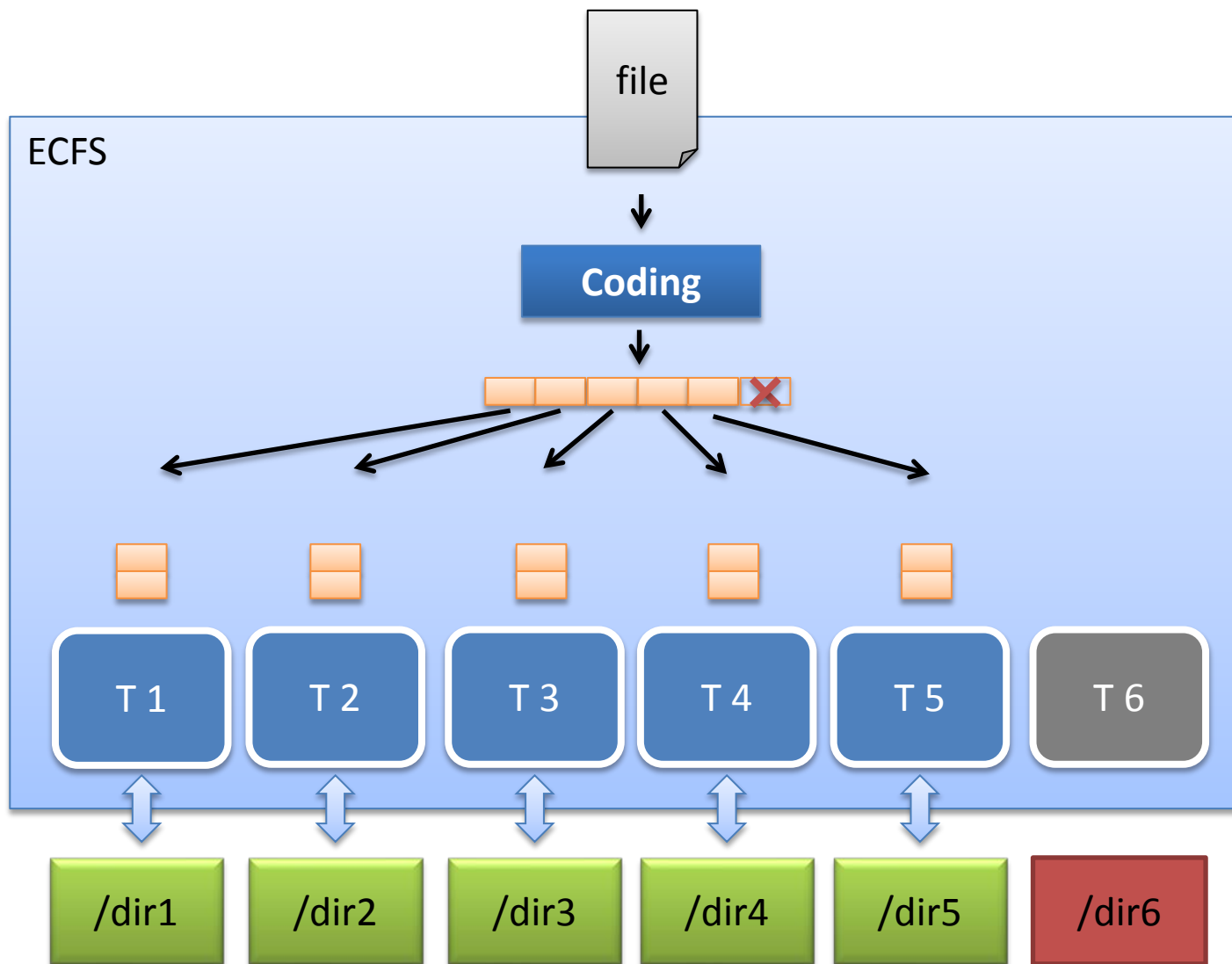
# Fault tolerance



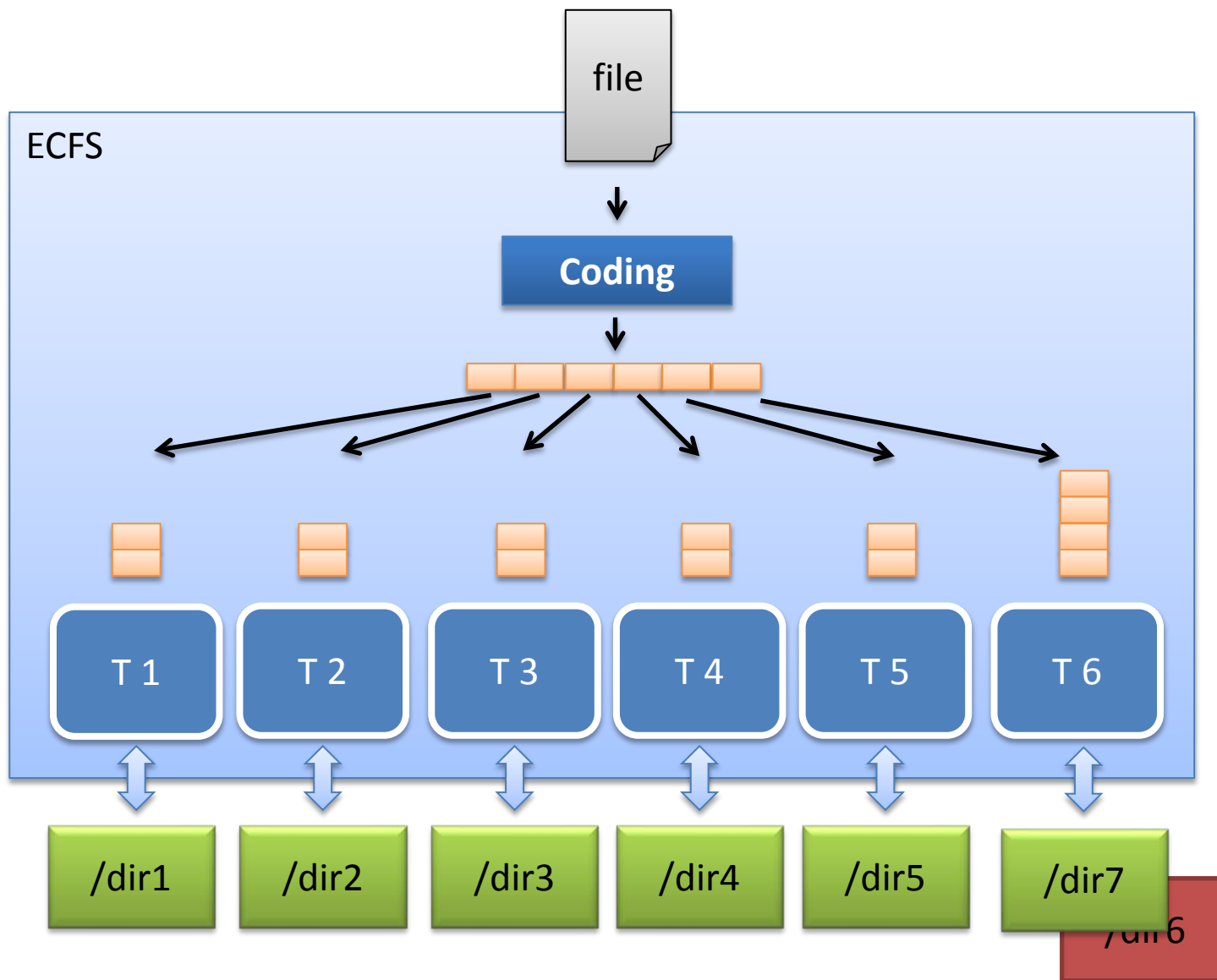
# Fault detection



# Degraded write



# Write to a spare





# Performance





# Conclusion





