ALICE Mass storage system simulation

mr. sc. Linda Vicković

Technical University of Split, FESB

Split, Croatia

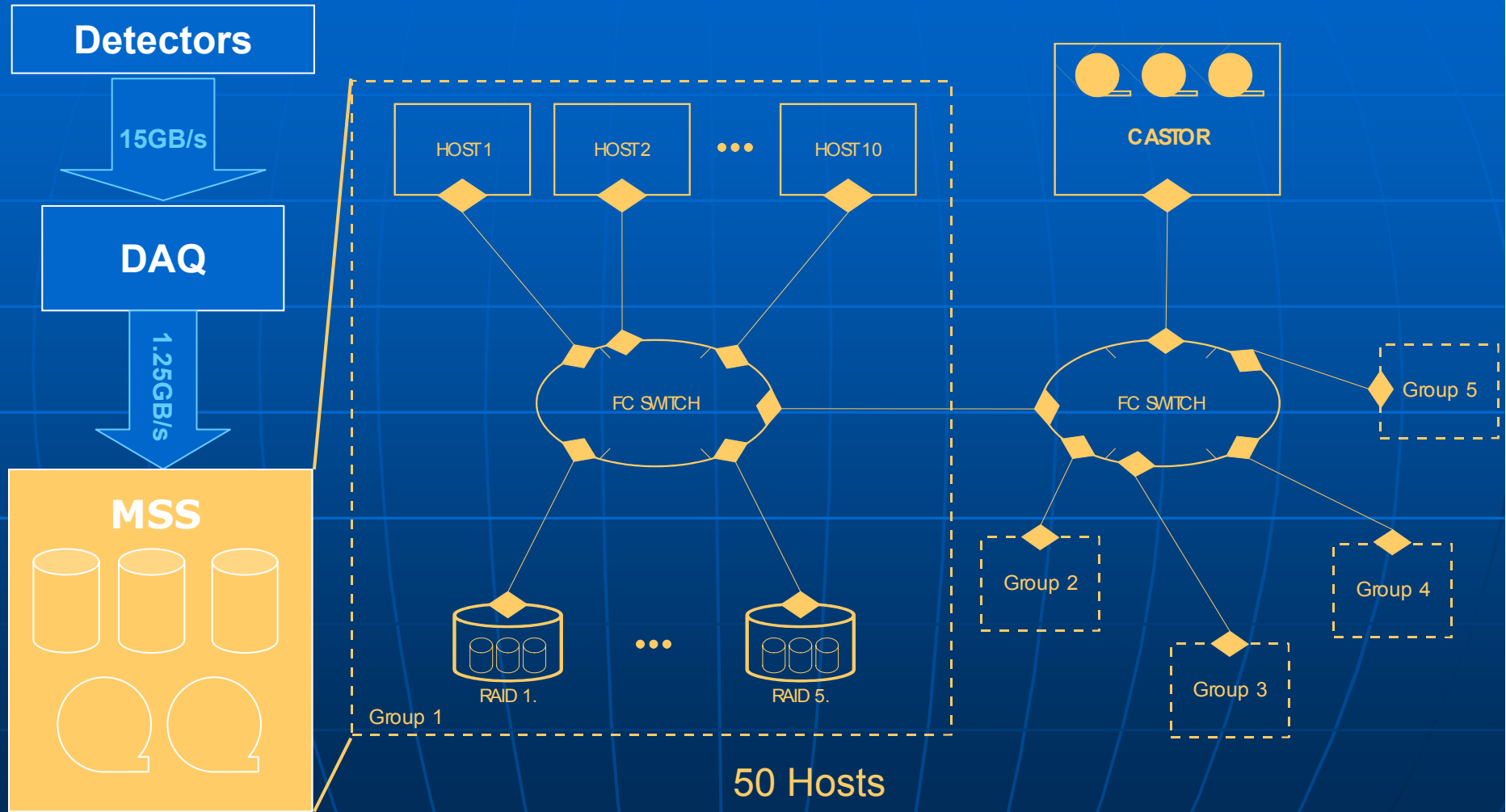October 2004

- Introduction
- Simulated components
- Measurement & simulation results
- Conclusion & future work

Introduction – ALICE data flow

Detectors

15GB/s

DAQ

1.25GB/s

MSS

HOST 1   HOST 2   • • •   HOST 10

CASTOR

FC SWITCH

FC SWITCH

Group 5

Group 2

Group 4

Group 3

RAID 1.   • • •   RAID 5.

Group 1

50 Hosts
5x16 ports switchs
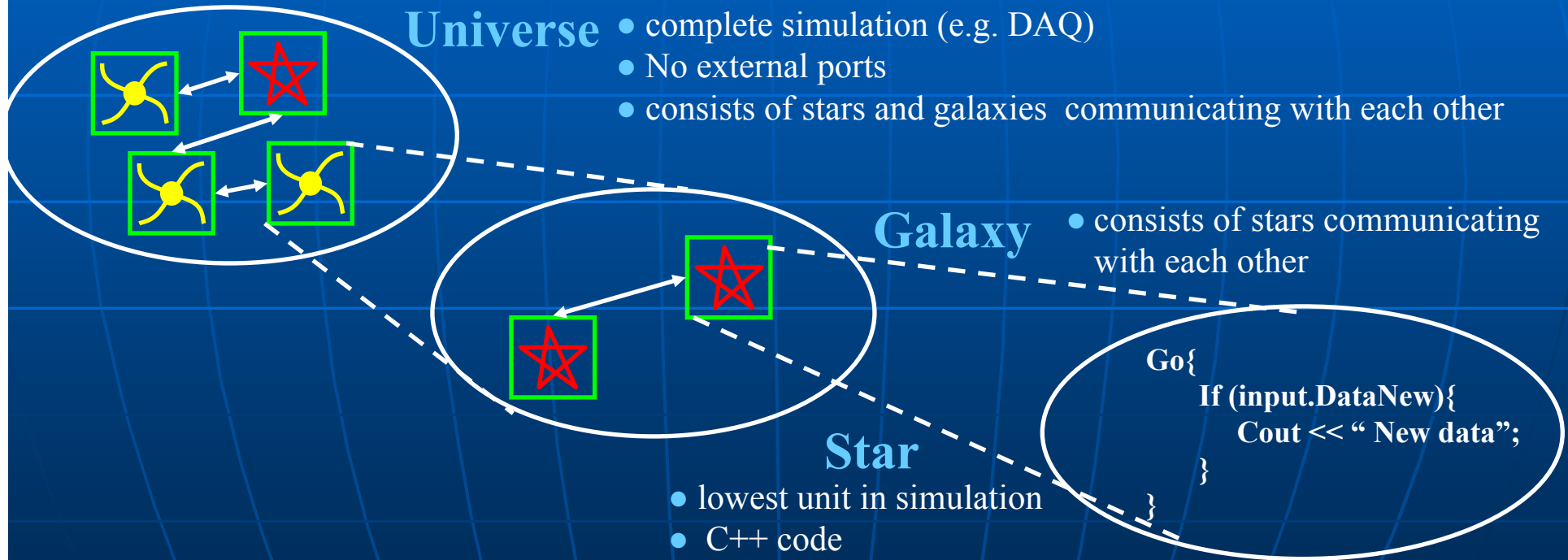25 RAIDs

- Prevent potential I/O bottlenecks
- To optimize system performances by exploring the influence of different parameters on system performances
  - file size, record size, RAID stripe size, …

- Modeling techniques can be grouped into two categories:
  - Analytical model – uses mathematical expressions
  - Simulation model – uses specialized computer programs
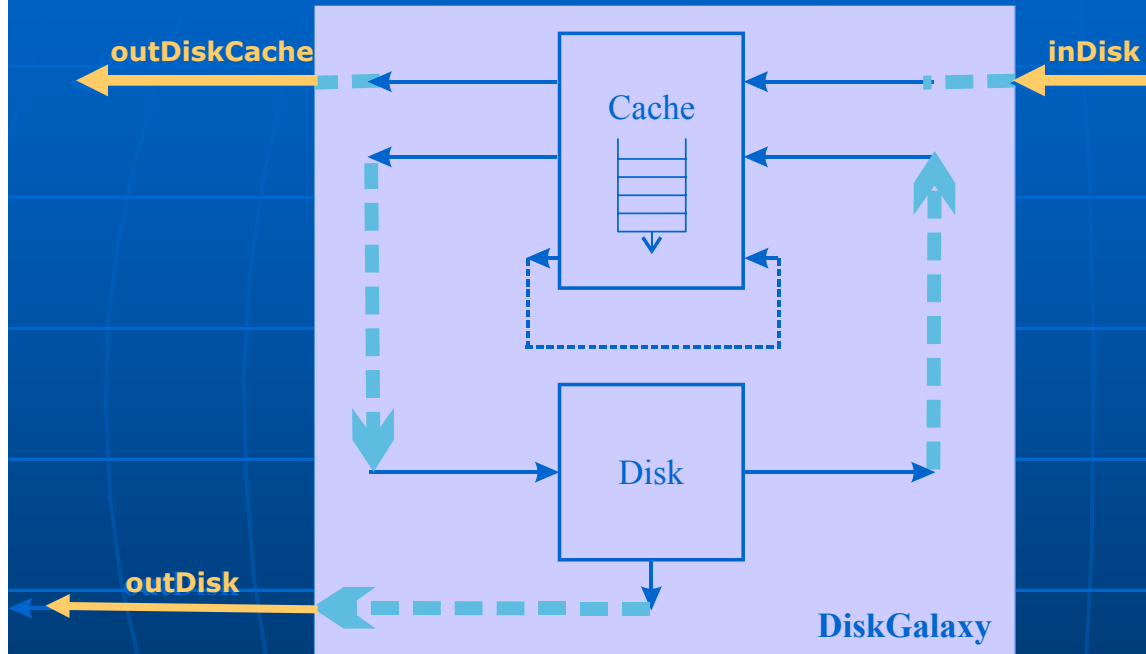 - Combination

Simulation tool – Ptolemy

- Heterogeneous simulation and design environment supporting multiple models of computation.
- Each object can has multiple inputs and outputs, firing in any order

Universe
- complete simulation (e.g. DAQ)
- No external ports
- consists of stars and galaxies communicating with each other

Galaxy
- consists of stars communicating with each other

```
Go{
    If (input.DataNew){
        Cout << " New data";
    }
}
```

Star
- lowest unit in simulation
- C++ code

Single disk model

outDiskCache

inDisk

Cache

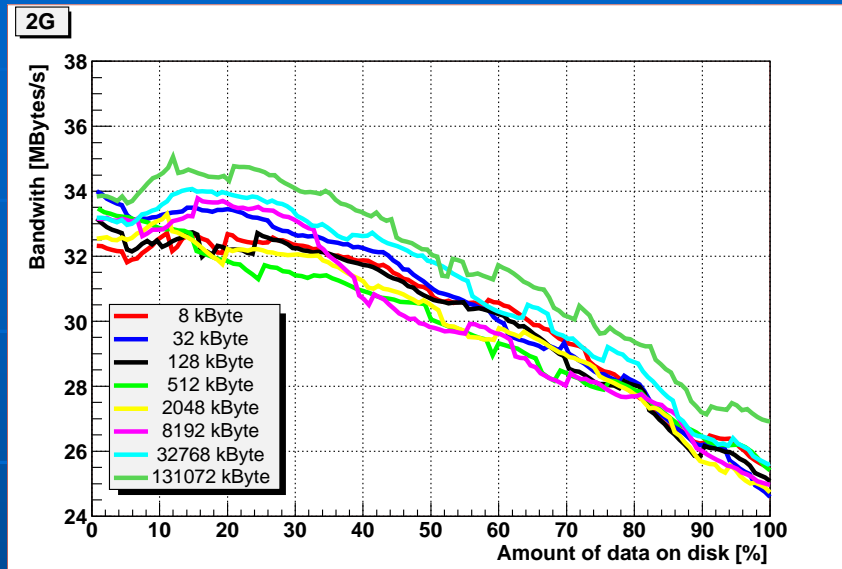Disk

DiskGalaxy

outDisk

- **Required parameters:**
  - Transfer rate from disk interface to disk cache
  - Disk cache size
  - Transfer rate from disk cache to media
  - Rotations per minute (RPM)
  - Full stroke seek
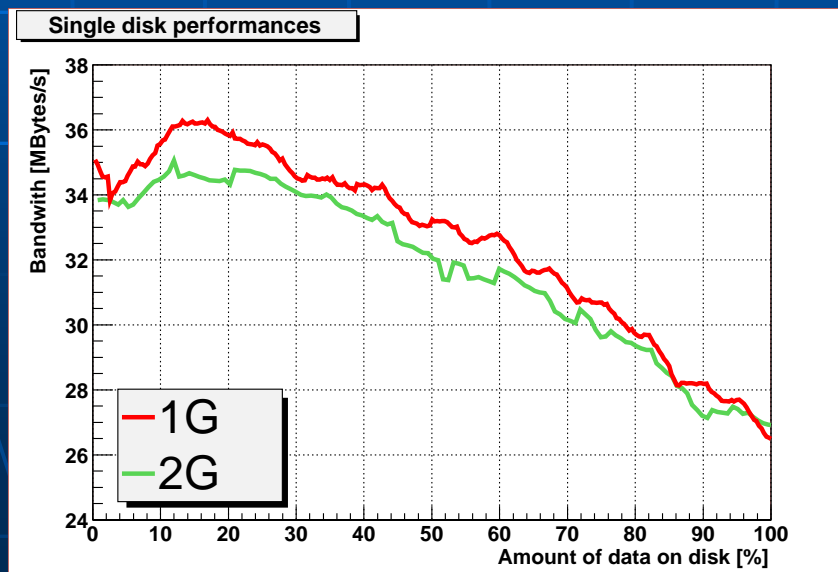 - Controller overhead

$$delay = \frac{record\_size}{transfer\_rate} + \frac{full\_stroke\_seek}{2} + \frac{60}{2*RPM} + cont\_overhead$$
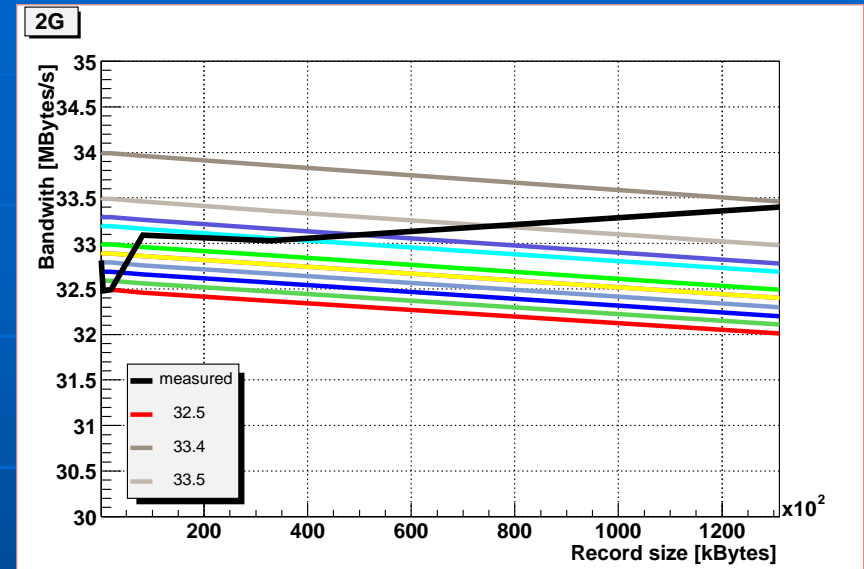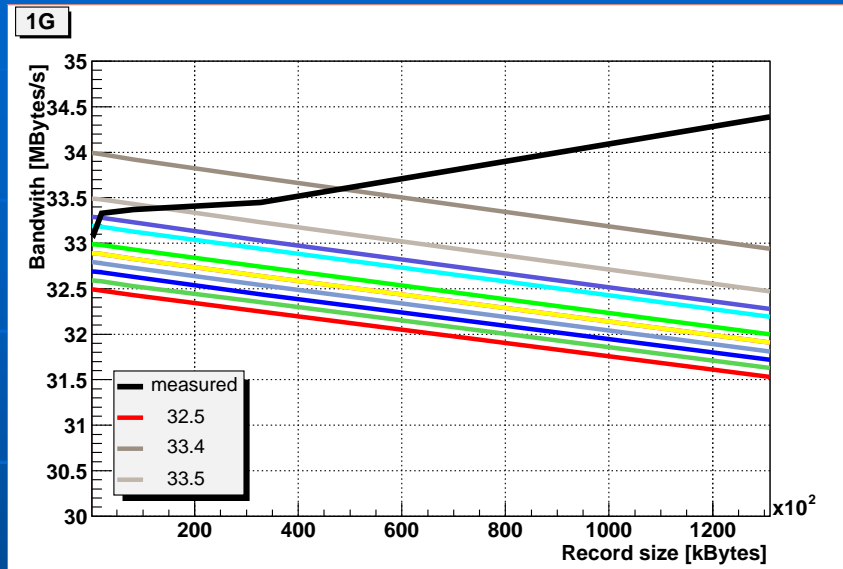
Single disk measurements

- Maxtor MaXLine Plus II
  - 300GB
  - 5400RPM
  - 2M cache
  - 133MB/s interface tr. Rate
  - 24–42 MB/s int. tr. rate
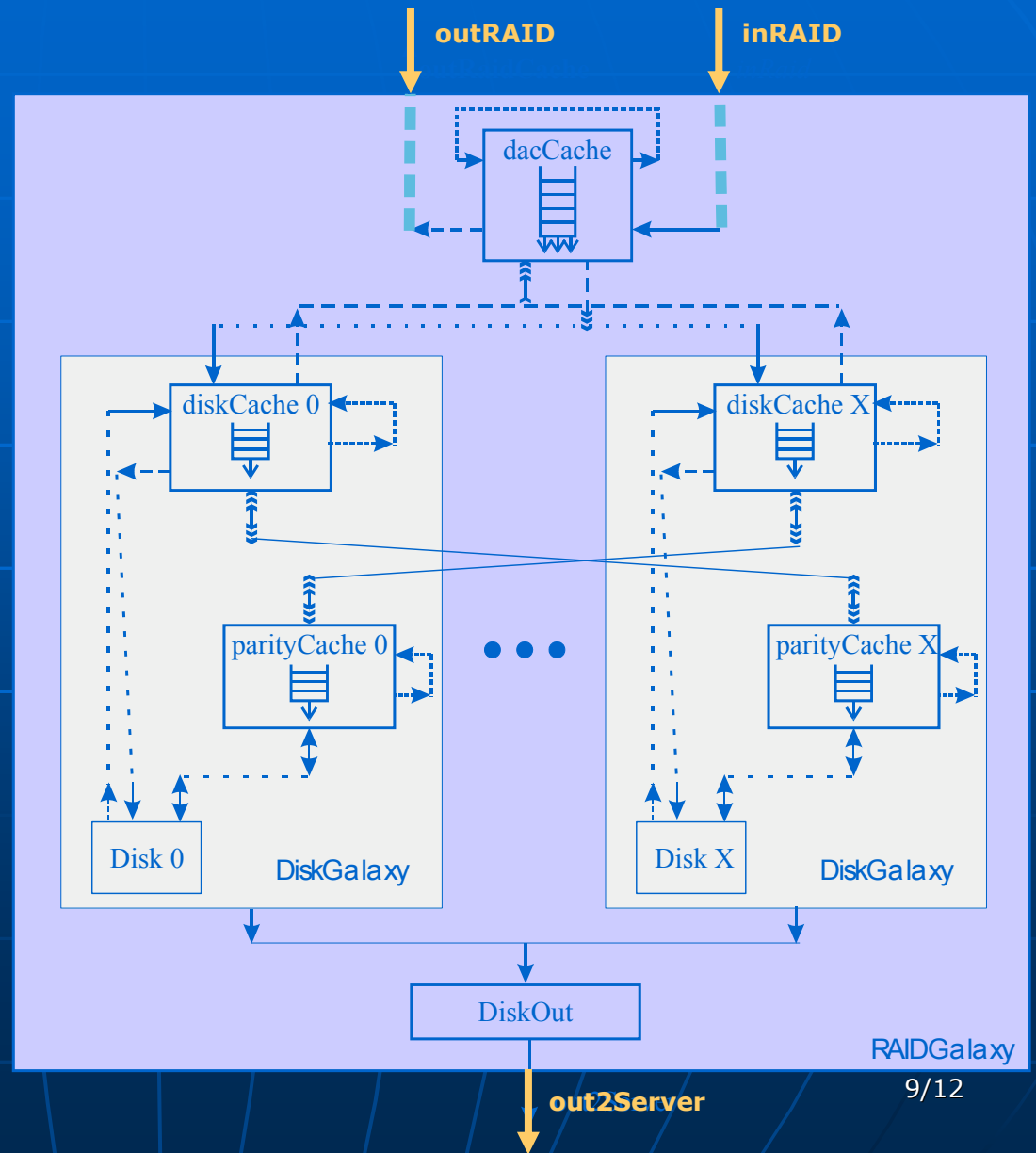
Results for single disk model

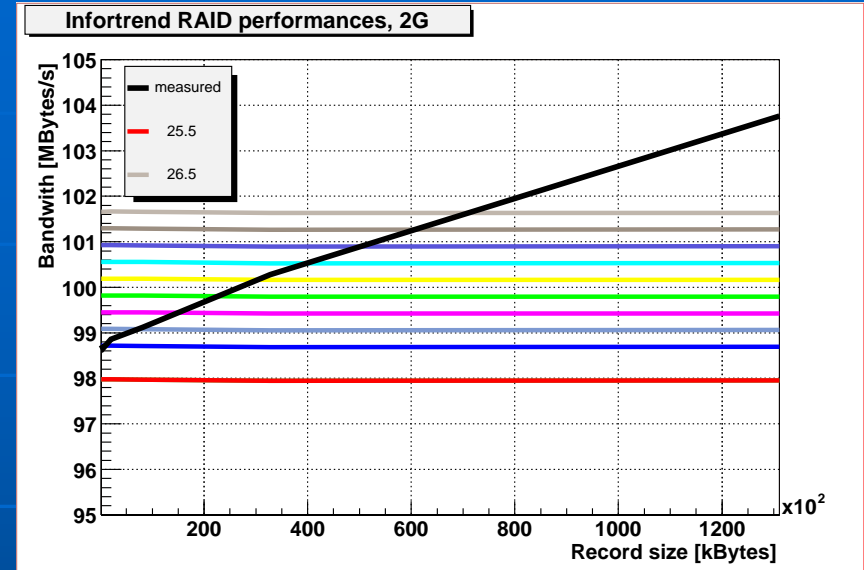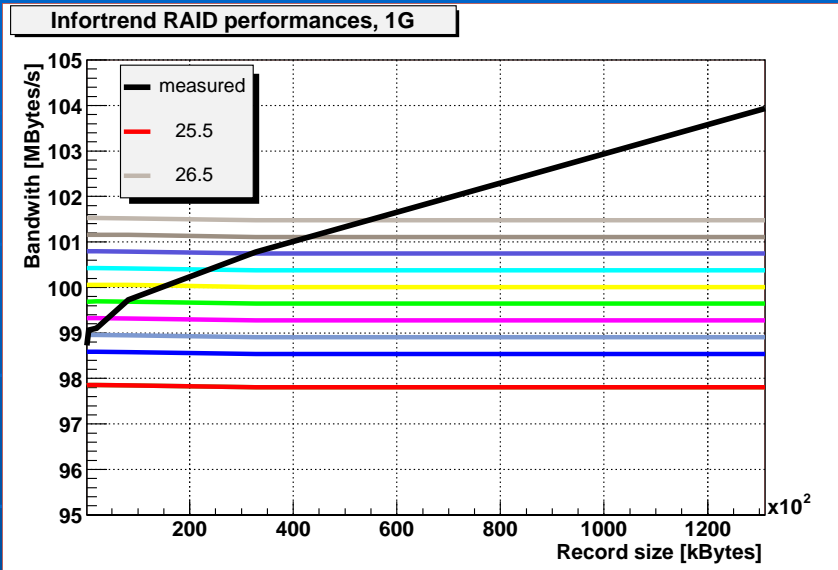| Record size	Difference between measured and simulated values for different internal transfer rate, for 2G files [%]										
[kB]	32.5	32.6	32.7	32.8	32.9	33	33.1	33.2	33.3	33.4	33.5
128	-0.96	-0.65	-0.34	-0.04	0.27	0.57	0.27	1.17	1.46	3.49	2.05
512	0.05	0.36	0.66	0.97	1.27	1.57	1.27	2.16	2.45	4.46	3.04
2048	0.00	0.31	0.61	0.92	1.22	1.52	1.22	2.11	2.40	4.41	2.99
8192	-1.92	-1.60	-1.29	-0.99	-0.68	-0.37	-0.68	0.23	0.53	2.57	1.12
32768	-2.04	-1.73	-1.42	-1.11	-0.81	-0.50	-0.81	0.10	0.40	2.44	0.99
131072	-4.35	-4.04	-3.72	-3.41	-3.10	-2.80	-3.10	-2.19	-1.88	0.18	-1.29

RAID model

- **Required parameters:**
  - All single disk parameters

  - Number of disks in RAID
  - RAID interface transfer rate
  - Size of the DAC cache
  - RAID stripe size

Results for RAID model

Infortrend RAID performances, 1G

Infortrend RAID performances, 2G

- Infortrend RAID
  - 200MB/s interface tr. Rate
  - 128 MB cache size
  - 133 MB/s internal tr.rate to disks
  - 5 Maxtor MaXLine Plus II disks

Record size [kB]	Difference between measured and simulated values for different internal transfer rate, for 2G [%]										
	25,5	25,6	25,7	25,8	25,9	26	26,1	26,2	26,3	26,4	26,5
128	-0,77	-0,77	-0,02	0,35	0,72	1,09	1,45	1,81	2,17	2,53	2,88
512	-0,70	-0,70	0,05	0,43	0,80	1,16	1,53	1,89	2,24	2,60	2,95
2048	-0,90	-0,90	-0,14	0,23	0,60	0,97	1,33	1,69	2,05	2,41	2,76
8192	-1,17	-1,17	-0,42	-0,04	0,33	0,70	1,06	1,42	1,78	2,14	2,49
32768	-2,38	-2,38	-1,61	-1,24	-0,86	-0,49	-0,12	0,25	0,61	0,97	1,33
131072	-5,92	-5,92	-5,13	-4,74	-4,35	-3,97	-3,59	-3,21	-2,83	-2,46	-2,09

Results for RAID model

- DotHill RAID
    - 200MB/s interface and internal tr. rate
    - 2G MB cache size
    - 5 Cheetah X15 36LP FC disks
        - 36.7GB
        - 15kRPM
        - 8M cache
        - 200MB/s interface tr. Rate
        - 51.8–61.8 MB/s internal tr. rate
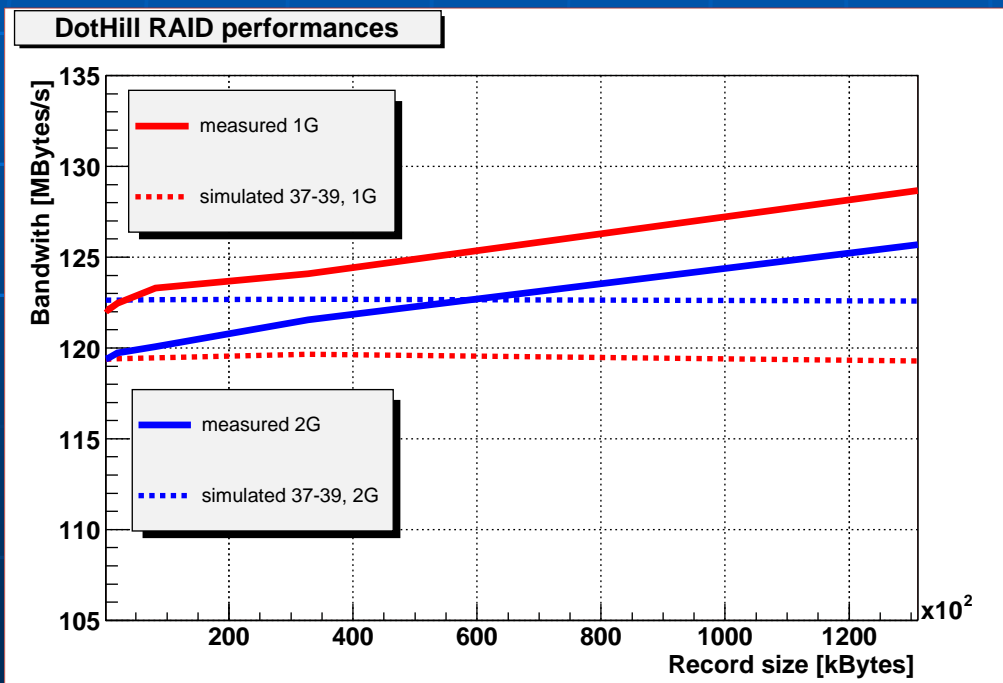
DotHill RAID performances

Record size [kB]	Difference between measured and simulated values for different internal transfer rate [%]	
	37 − 39 1G	37 − 39 2G
128	-2,36	2,62
512	-2,28	2,61
2048	-2,57	2,37
8192	-2,21	2,09
32768	-3,71	0,93
131072	-7,88	-2,53

Conclusion & future works

- It is possible to simulate the behavior of such a complex system.

- Change the simulation to use variable transfer rate to disk drive.

- Find the distribution of the transfer rate that will simulate the real system behavior as realistic as possible.

- Make the simulation of the whole system.