



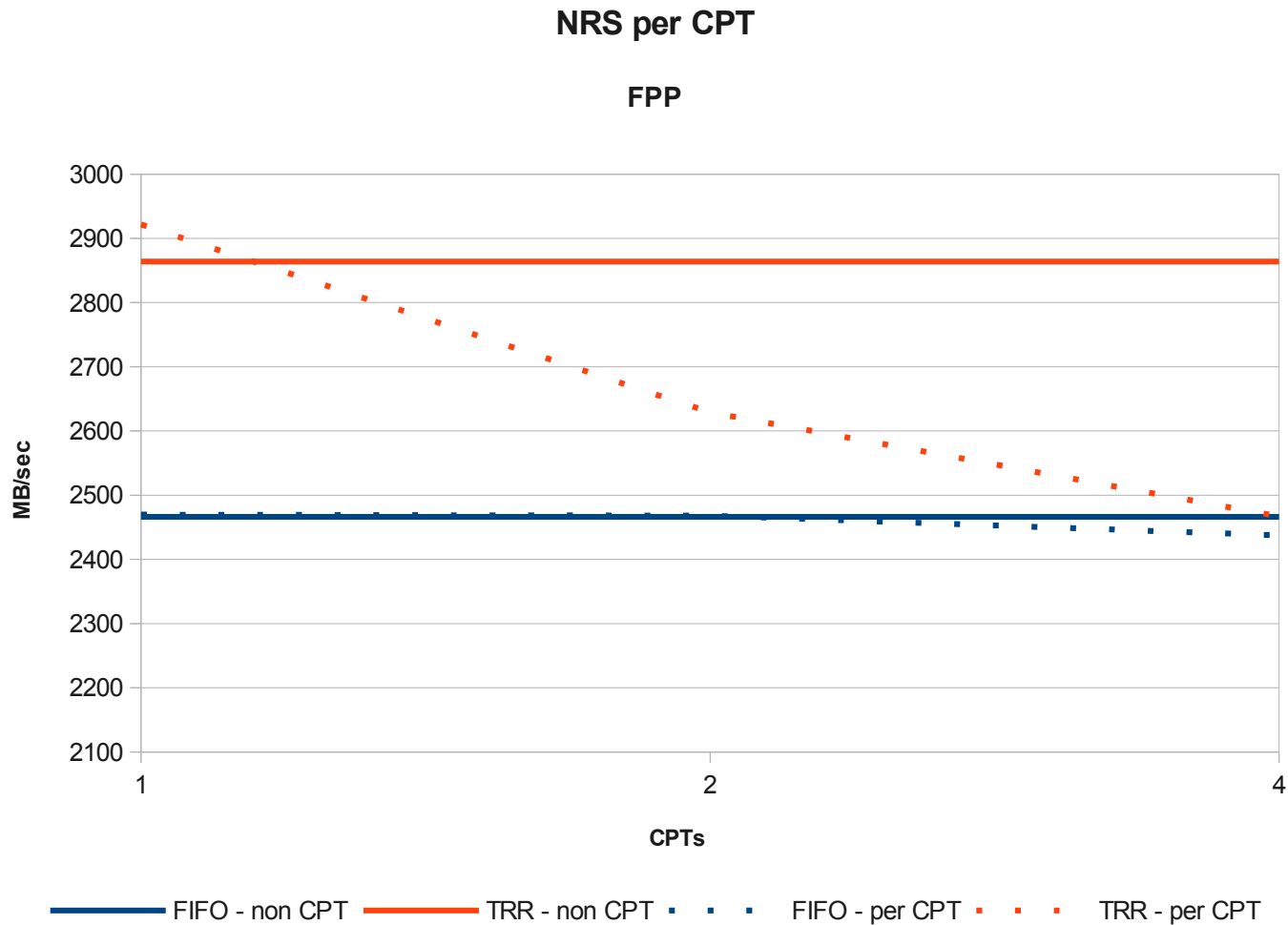
Advancing Digital Storage Innovation



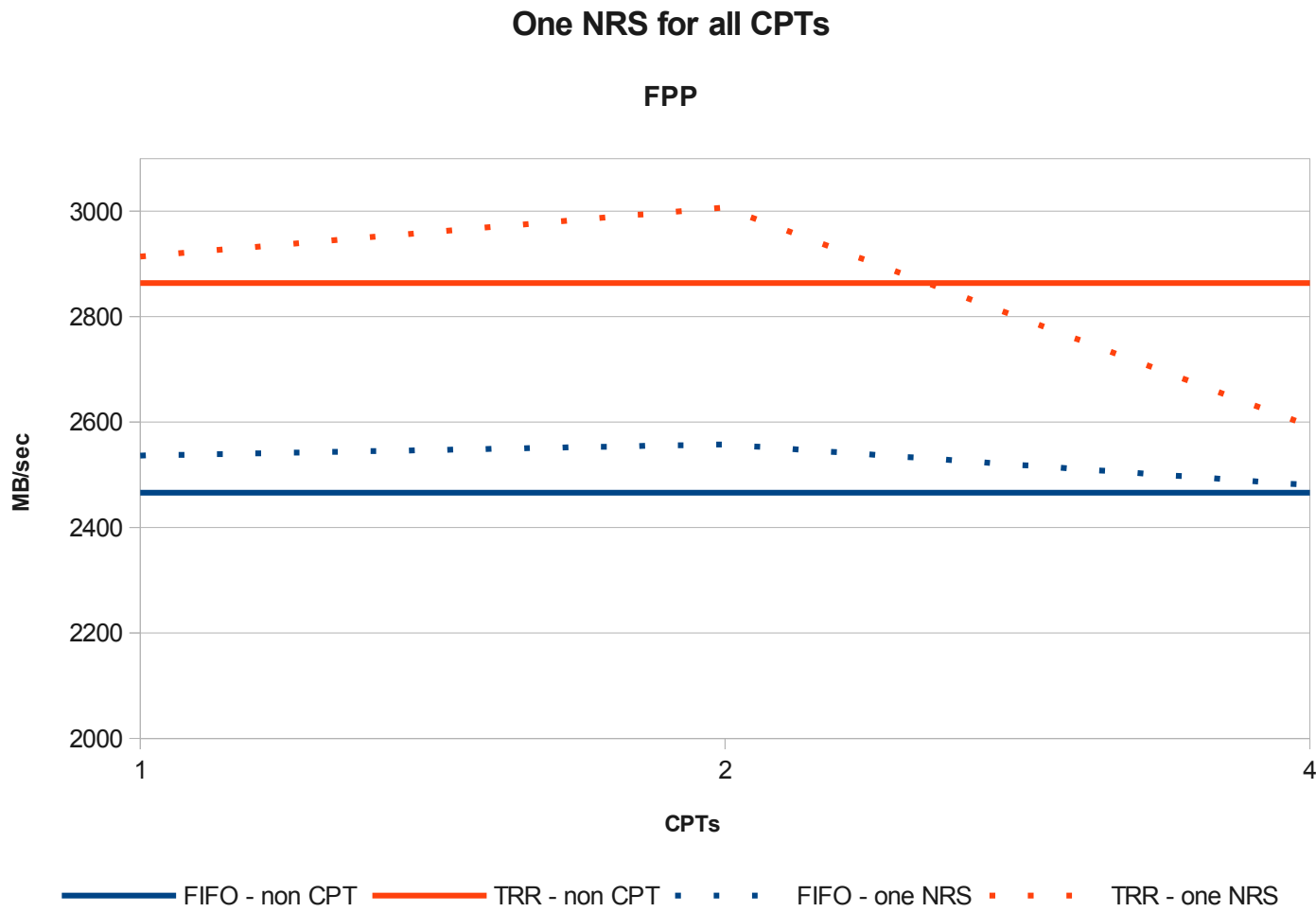
Network Request Scheduler (NRS) Bandwidth Policies

Nikitas Angelinas
nikitas_angelinas@xyratex.com

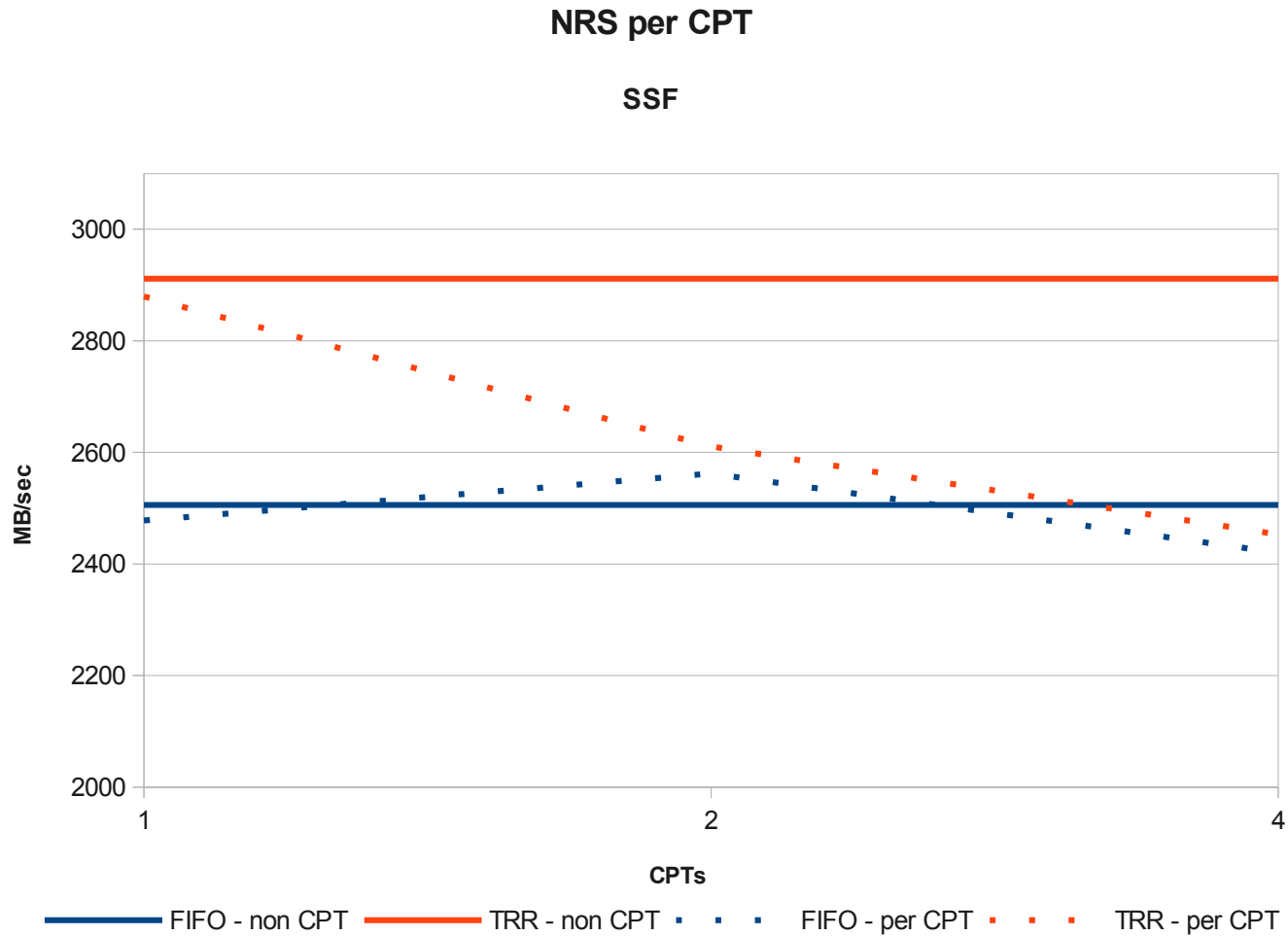
- ORR serves bulk I/O RPCs in a Round Robin manner over available backend-fs objects
 - RPCs are placed in per-object groups of 'RR quantum' size; lprocfs tunable
 - Sorted within each group by logical or physical disk offset
 - Physical offsets are calculated using extent information obtained via fiemap calls
 - Support for OST_READ and/or OST_WRITE RPCs; lprocfs tunable
- TRR is equivalent, but schedules RPCs in a Round Robin manner over available OSTs
- The main aim is to minimize drive seek operations, thus increasing read performance
- May help with load balancing across OSTs
- May take advantage of locality of reference



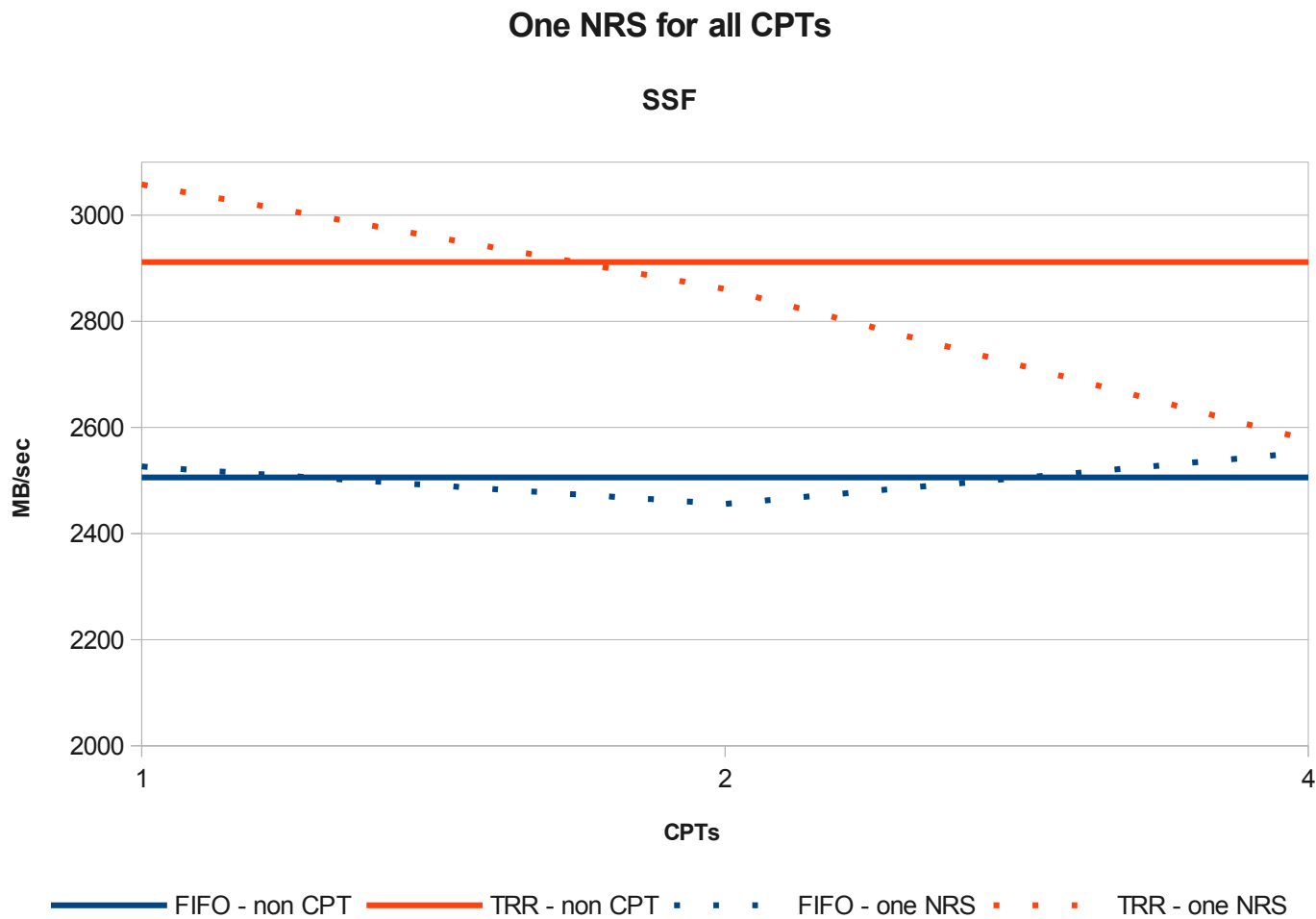
FFP Read - One NRS for all CPTs



SSF Read - NRS per CPT



SSF Read - One NRS for all CPTs



Notes on ORR and TRR policies

- ORR and/or TRR may help improve:
 - Some generic read use cases
 - Small and/or random reads
 - Widely striped file reads
 - Backward reads
 - Cases in which OSTs are underutilized; this has not been tested yet
 - Reads by aligning writes
- ORR will need an LRU-based or similar method for object destruction; TRR much less so
- TRR and ORR should be less (if at all) beneficial on SSD-based OSTs

Increase read performance by aligning writes

- Possibly increase read performance by aligning writes
- Write performance takes a hit
- But this may be useful in read-important cases
- Quick, small scale test
 - 14 clients, ost_io.threads_max = 128, stripped directories

| Test | policy writing | policy reading | write (MB/s) | read (MB/s) |
|------|----------------|----------------|--------------|-------------|
| FPP | FIFO | FIFO | 2013.72 | 2735.63 |
| | ORR | FIFO | 1074.25 | 3937.05* |
| | ORR | ORR | 1074.25 | 3966.07* |
| SSF | FIFO | FIFO | 2094.56 | 2832.48 |
| | ORR | FIFO | 1115.28 | 3226.53 |
| | ORR | ORR | 1115.28 | 3186.26 |

* value is >> quoted system maximum



Advancing Digital Storage Innovation



Cheers

Nikitas Angelinas
nikitas_angelinas@xyratex.com