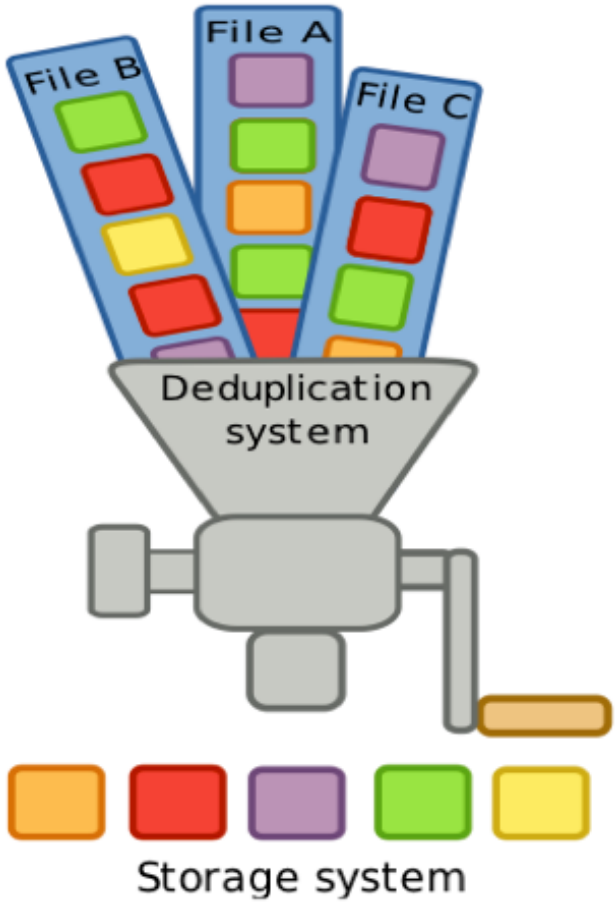
PC² PADERBORN CENTER FOR PARALLEL COMPUTING

dedupv1
Improving Deduplication Throughput using Solid State Drives

Dirk Meister, André Brinkmann
Paderborn Center for Parallel Computing
University of Paderborn

Motivation

- Data deduplication eliminates duplicate blocks
- Used for D2D backup (95-%-98% redundancy)
- Fingerprints are often used to identify duplicate blocks
- Fingerprint lookups limit throughput (disk seek time)



Question:

Are Solid State Drives a way out of the

disk bottleneck?

Solid State Drives

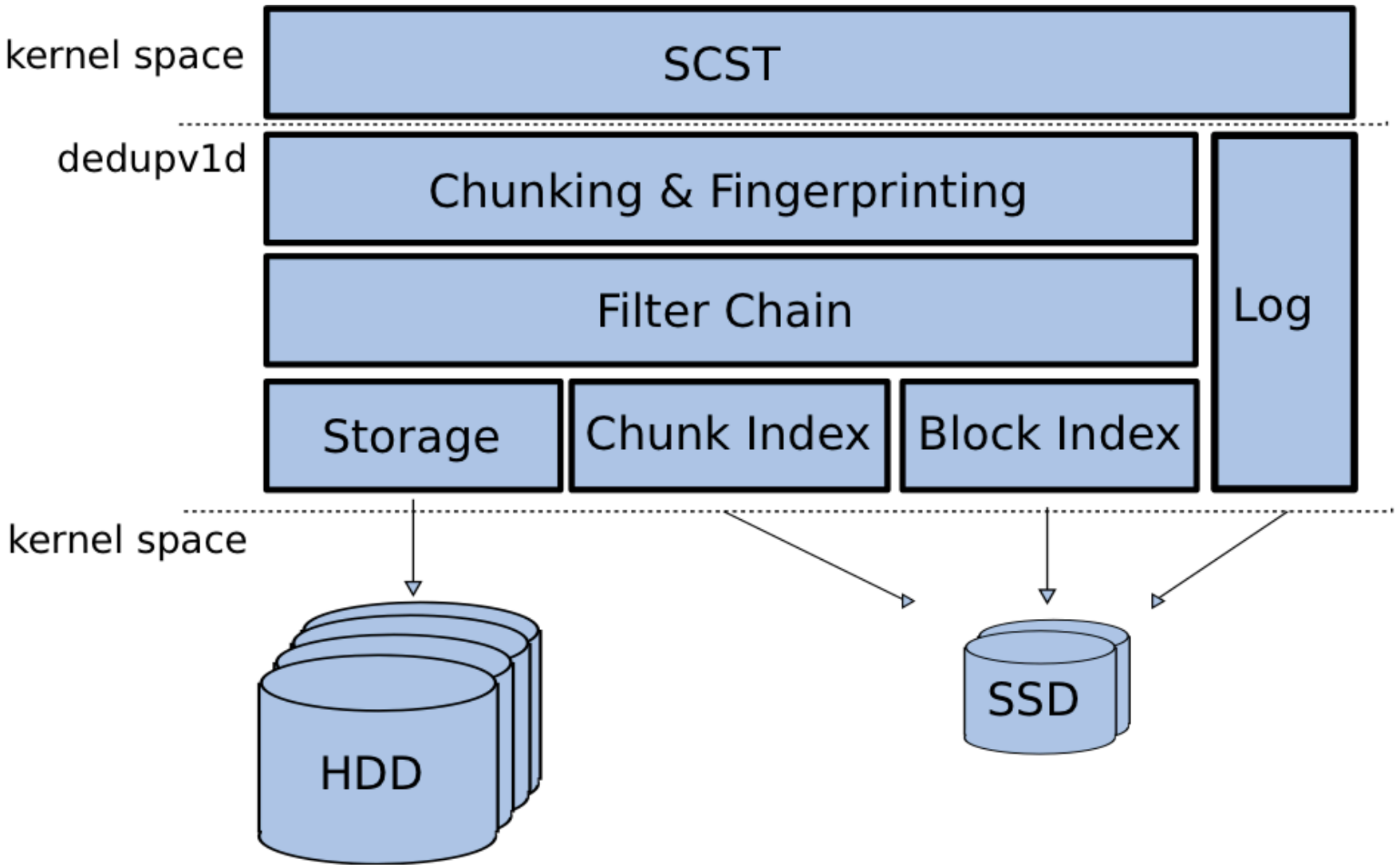
- Here: based on Flash memory
- High random read performance
- Usually: slow random writes

- Naive approach of simply storing index on SSD would fail

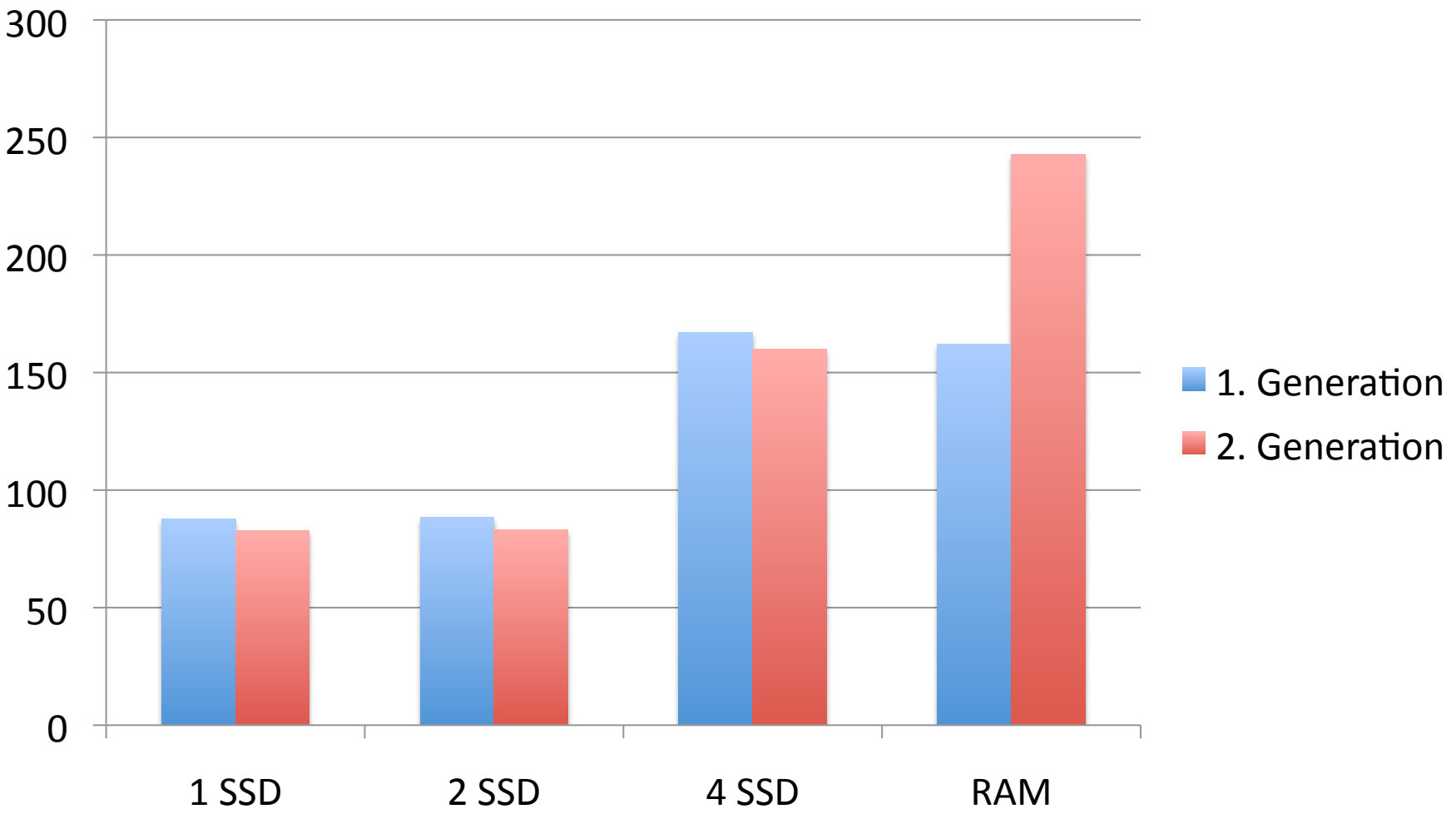
  „An alternative approach—storing the index in Flash memory— would eliminate seek cost for reads but greatly increase it for writes.“ [1]

[1] Rhea, Cox, Pesterev: Fast Inexpensive Content-Addresse Storage in Foundation, USENIX, 2008

Architecture



Performance (MB/s)



Conclusion

- SSD can improve the throughput of deduplication systems
- Naive approach fails
  - large write latency of SSD
- Move write operations into the background

Thank you