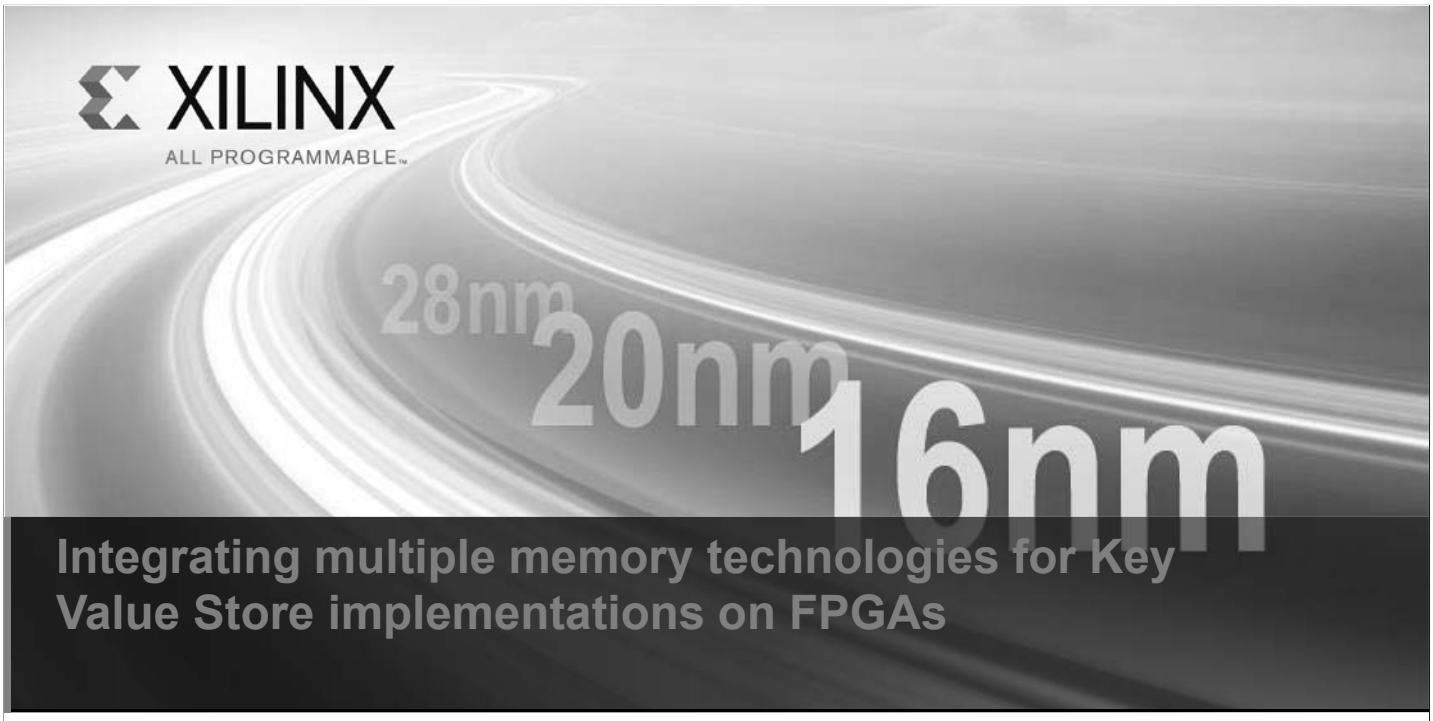


Integrating multiple memory technologies for Key Value Store implementations on FPGAs

Kees Vissers

Xilinx

kees.vissers@xilinx.com

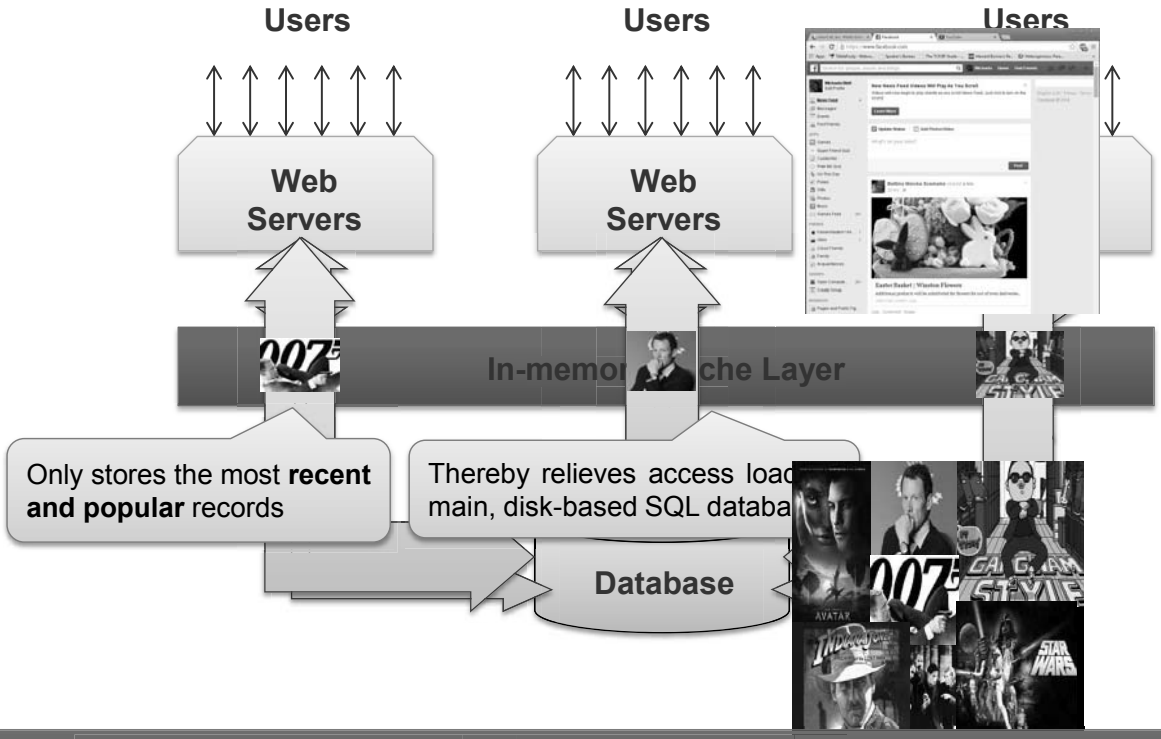
MPSoC'15

Contents

- Memcached basics
- FPGA based Memcached implementation
- FPGA based hybrid memory implementation
- Scaling to disruptive levels
- Xilinx' MPSoC
- Conclusions

Key Value Store – Memcached

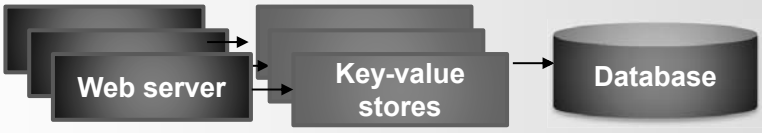
- Many popular websites share a similar basic architecture:



FPGAs will increase performance and reduce power & latency

Memcached

- Common middleware application to alleviate access bottlenecks on databases
  - Most popular and most recent database contents are cached in main memory of a tier of server platforms



- Used by many well-known websites
  - up to 30% of servers in data centers run memcached or similar

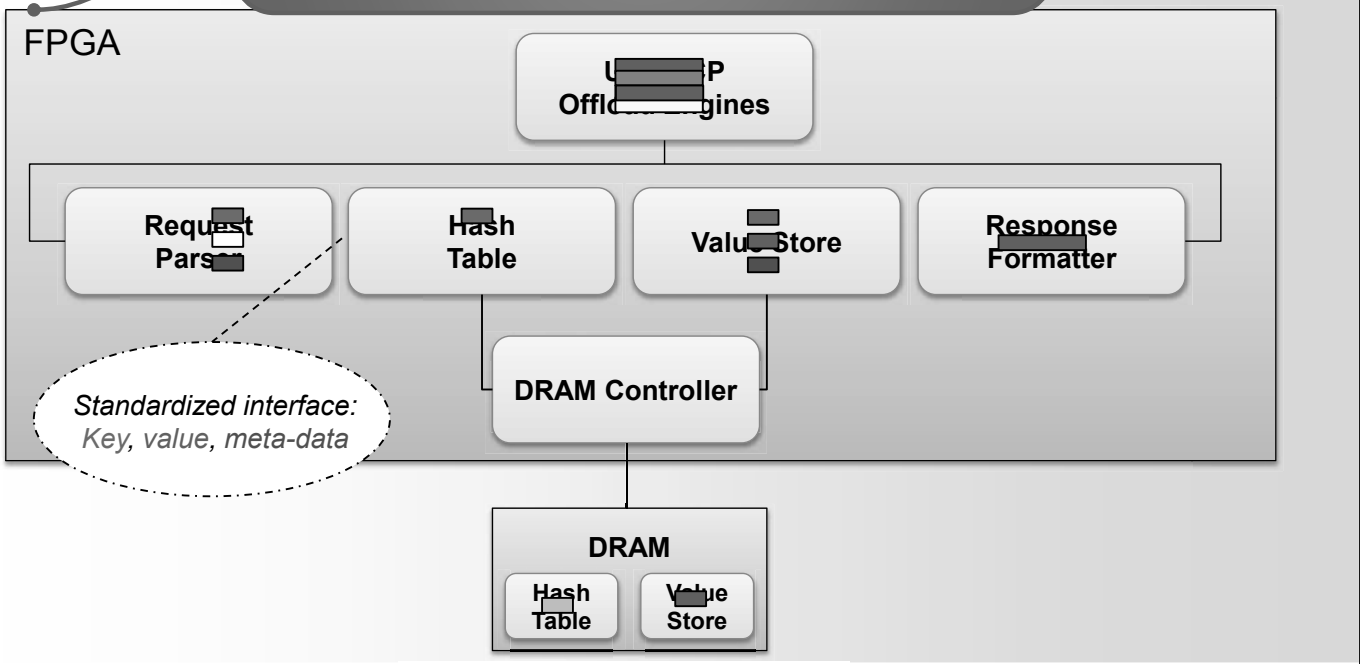
- Current server-based implementations are limited, cannot keep up with 10Gbs network speed and won't scale with more cores

- Investigated using dataflow architectures on FPGAs to dramatically increase performance and lower power and latency

Dataflow Architecture

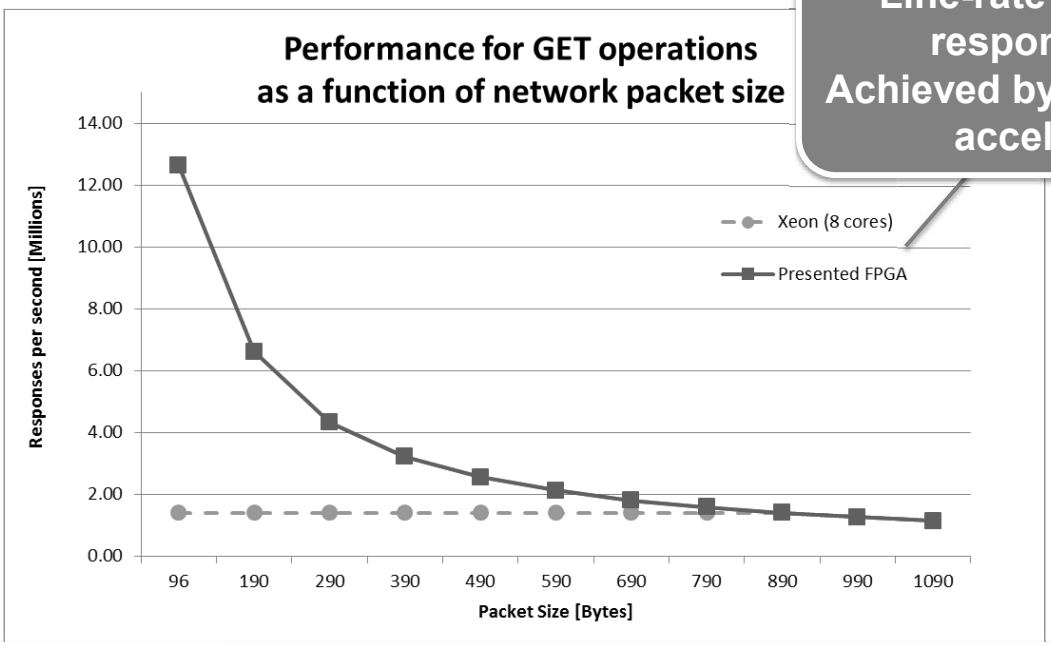
Streaming architecture:

Flow-controlled series of processing stages which manipulate and pass through packets and their associated state



Key Value Store Acceleration with FPGAs

Line-rate maximum response rate Achieved by Xilinx FPGA accelerator



Demonstrator:

Up to 36x in performance/power demonstrated
10-100x reduction in latency
Scalability to higher rates possible

Leverage Hybrid Memory Systems

Calculated probability of value sizes									
Value size [Bytes]	128	256	512	768	1014	2048	4096	22000	32000
Facebook: USR*	1	0	0	0	0	0	0	0	0
Twitter	0	0	0	0.1	0.85	0.05	0	0	0
Wiki	0	0	0	0	0.58	0.02	0.1	0.25	0.05
Flicker	0	0	0	0	0	0	0	0.1	0.9
Youtube	0	0	0	0	0	0.75	0.11	0.11	

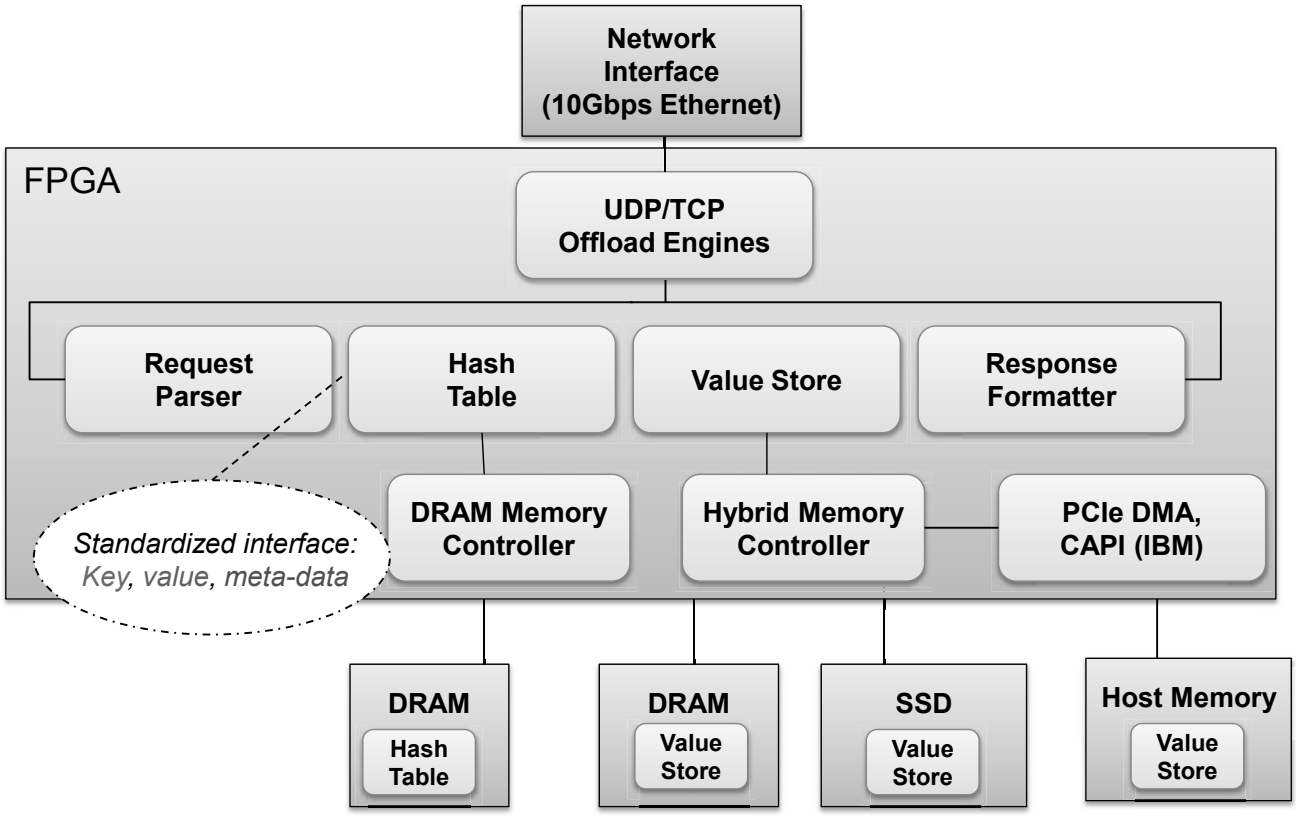
Stored in DRAM

Stored in Host Memory

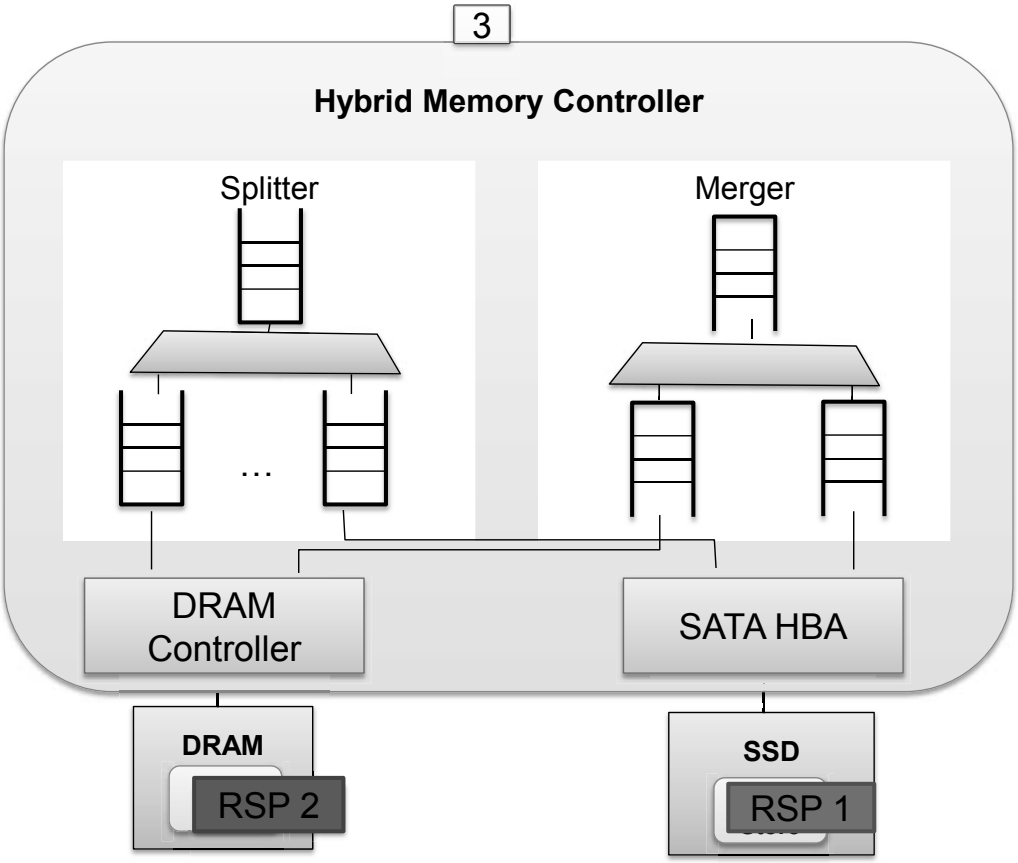
Stored in Flash

Advantages: High capacity (TB vs GB), lower in cost (1$ vs 8$/GB) lower in power

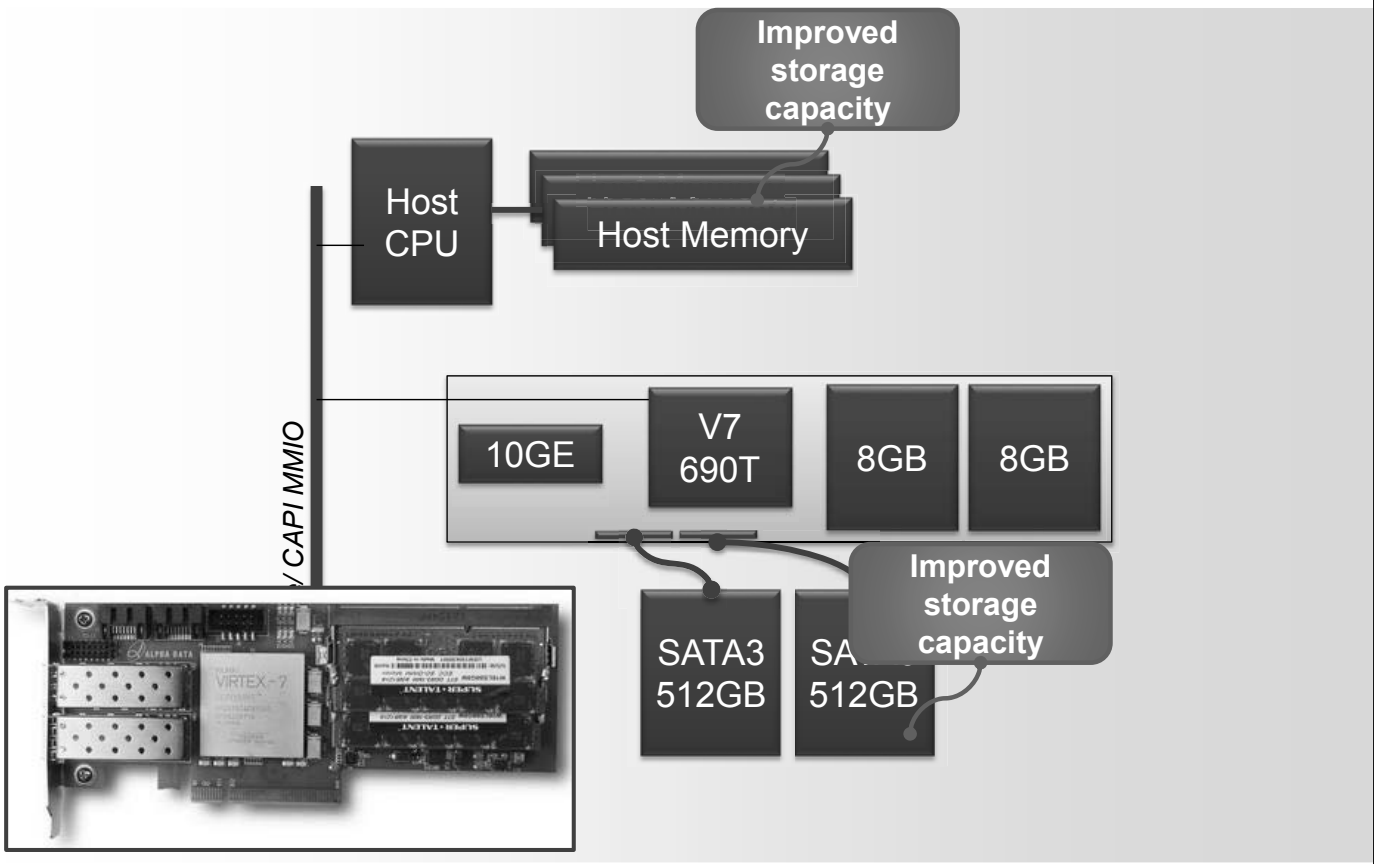
Expanding Memory Subsystem to use Host Memory & SSDs



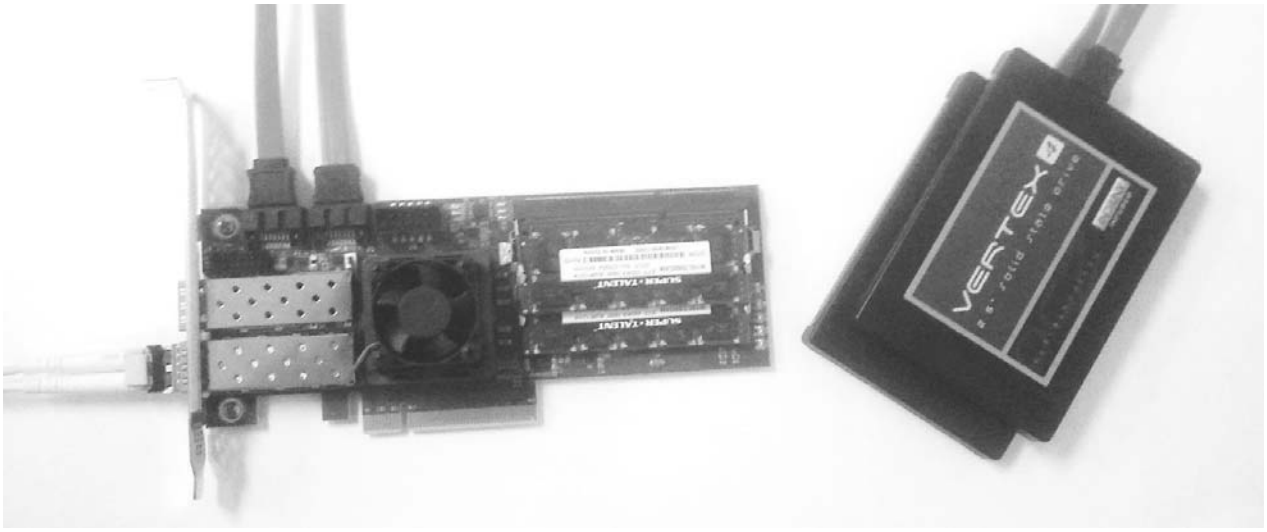
Hybrid Memory Controller



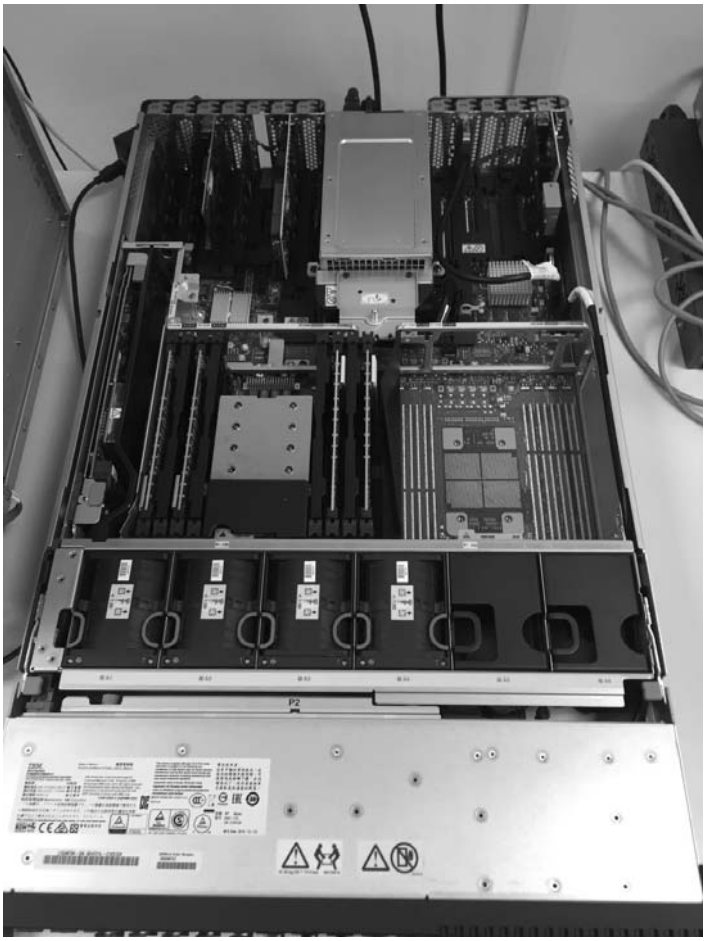
Current System



FPGA card

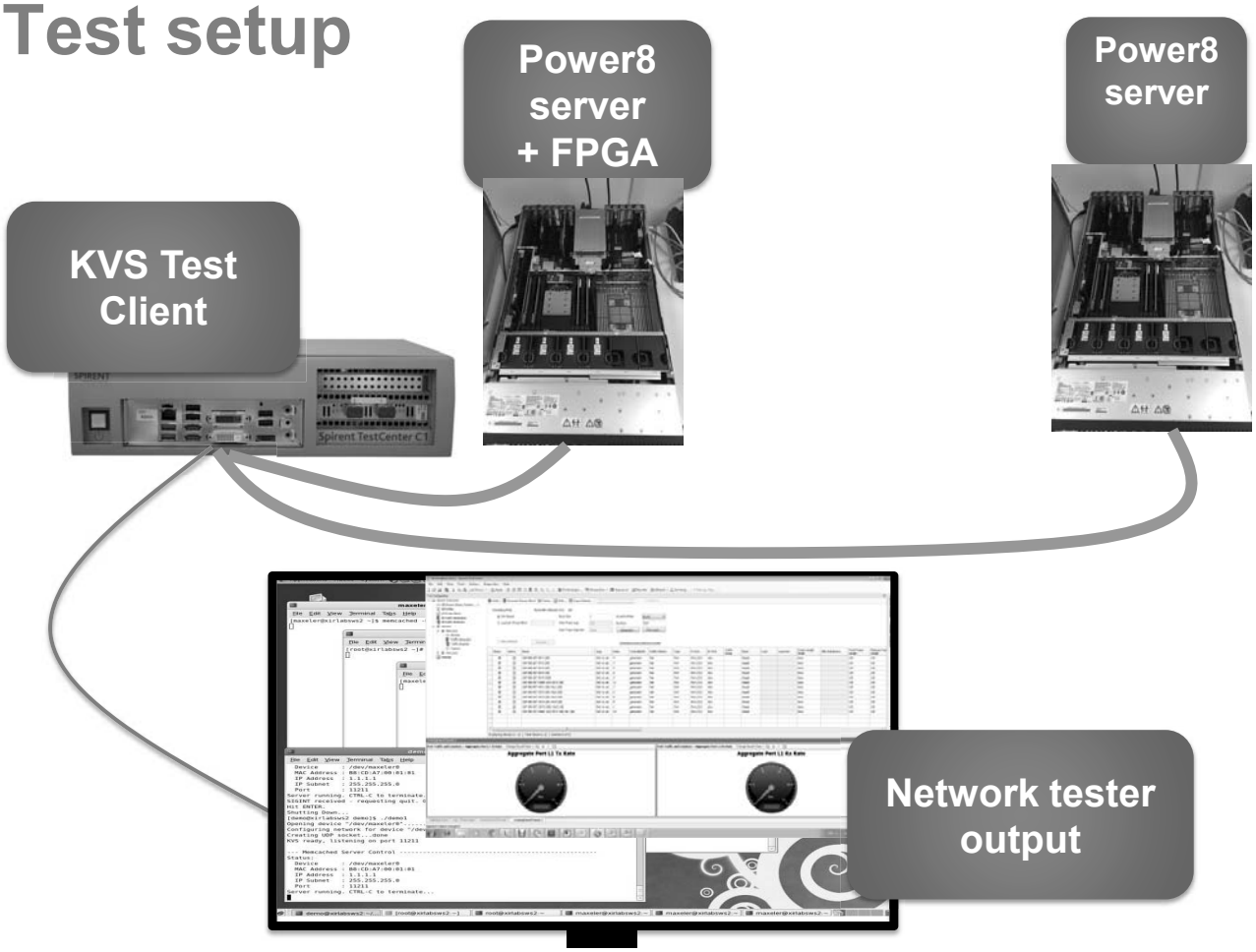


systems

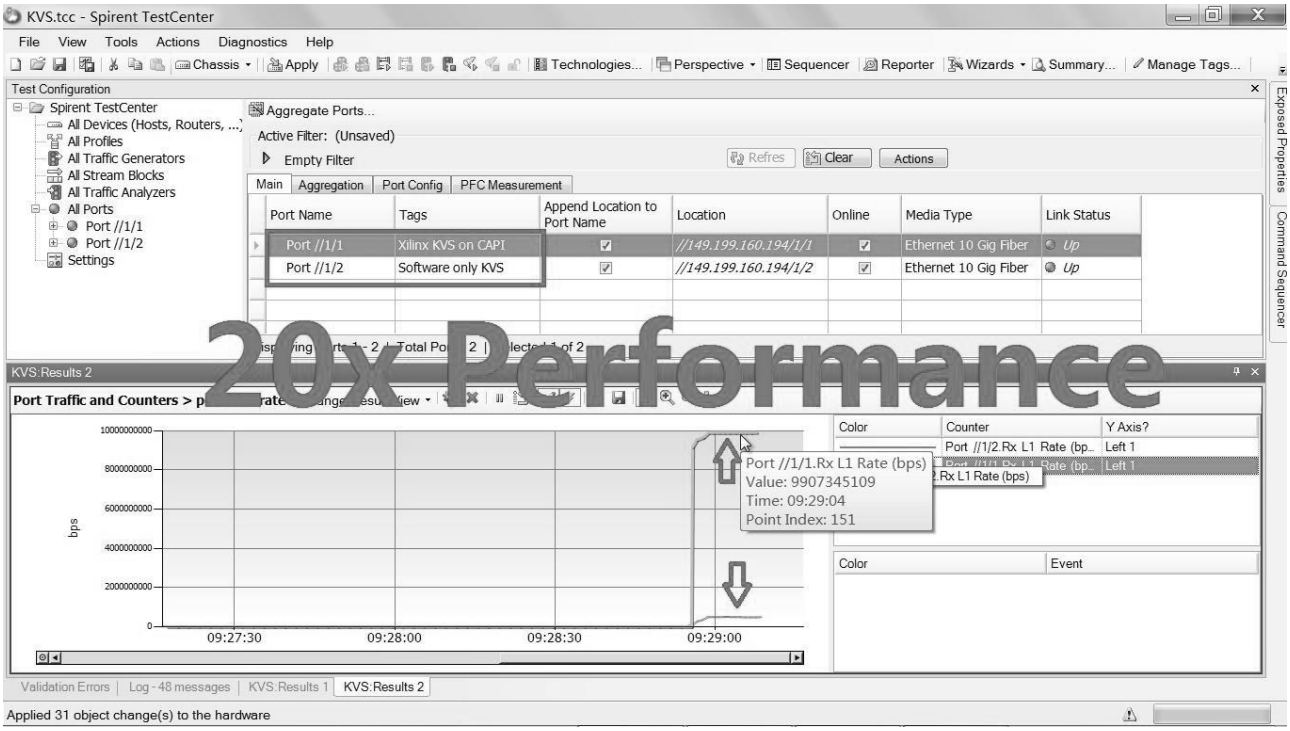


IBM cloud server solution

Test setup



Measured Benefit



Scaling out to 80 Gb/s network and 40TB storage

Access probability distribution over size in Bytes.

Value size (Bytes)	128	256	512	768	1024	4096	8192	32K	1M	$prob_{SET}$
Facebook	0.55	0.075	0.275	0	0	0	0	0	0.1	3%
Twitter	0	0	0	0.1	0.85	0.05	0	0	0	20%
Wiki	0	0	0.2	0.1	0.4	0.29	0.008	0.001	0.001	1%
Flickr	0	0	0	0	0	0.9	0.05	0.03	0.02	0%

Hybrid Memory system evaluation results.

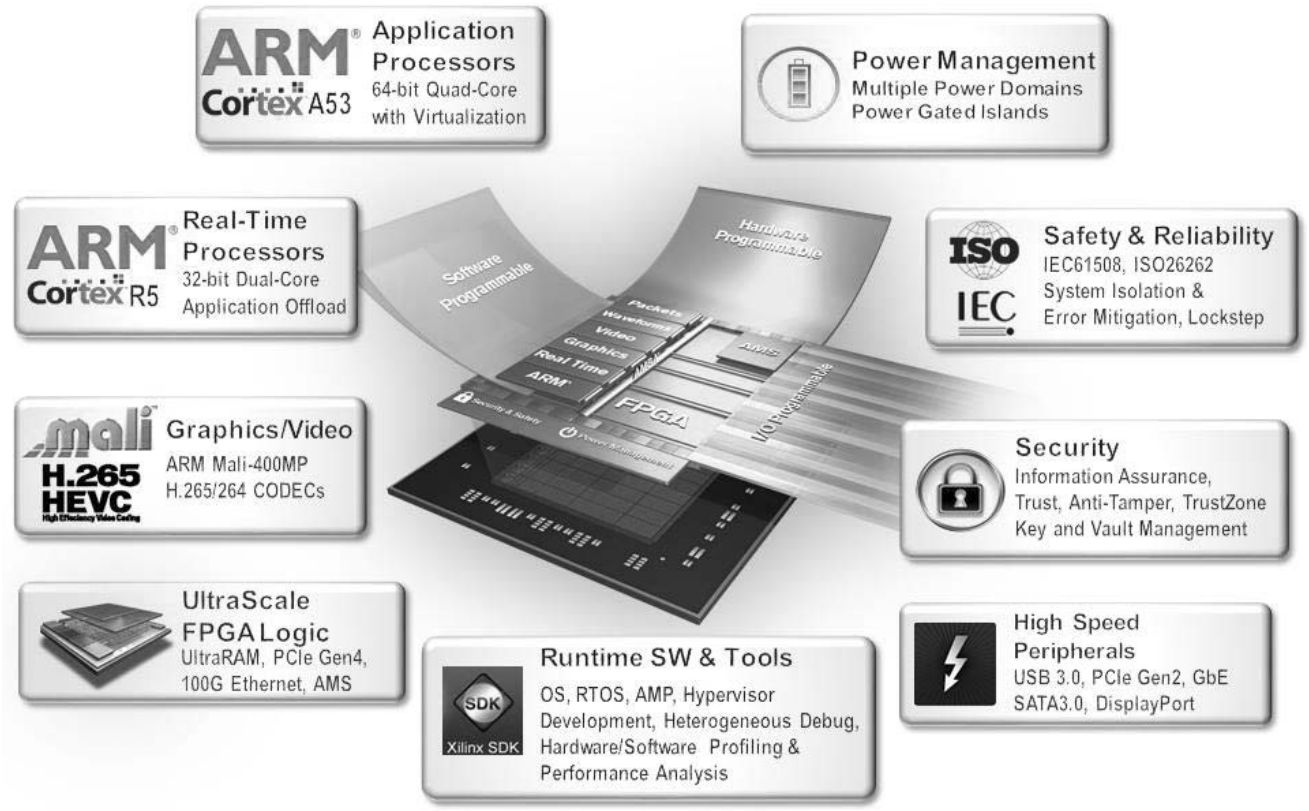
Use case	d(DRAM) [GB]	d(SSD) [GB]	DRAM bw utilization	SSD bw utilization	Maximum entries[M]	SSDs
Facebook	254	20,000	17%	12%	200	20
Twitter	238	25	16%	61%	120	2
Wiki	244	343	12%	92%	150	8
Flickr	250	6,000	1%	98%	240	25

Comparison

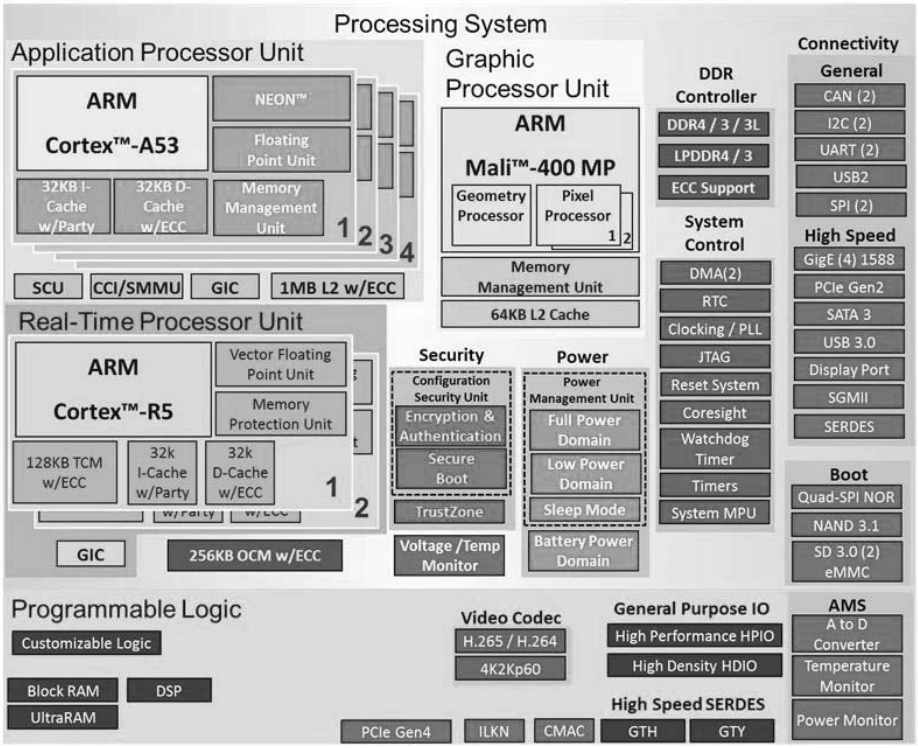
- **Comparison with best published results**

Platforms	GB	KRPS	Watt	KRPS/Watt	GB/Watt
FPGA (80Gbps design)	40,000	104,000	434.8	239.2	92.0
FPGA (80Gbps facebook)	20,254	32,657	343	95.2	59.0
FPGA (10Gps prototype)	272	1,340	27.2	49.2	10.0
Dual x86 (MICA)[12]	64	76,900	477.6	161	0.1
Dual x86 (FlashStore)[6]	80	57.2	83.5	0.7	1.0
FAWN (SSD) [2]	32	35	15	2.3	2.1

Zynq Ultrascale+ MPSoC



Zynq Ultrascale+ MPSoC for Smarter Network



- 64-bit quad cores
- Up to 100 Gb/s network
- 76 High-speed serial transceivers
- PCIe, Gen4x8, Gen3 x16
- > 1 Million Logic Elements
- > 70Mb on chip storage

Acknowledgements

- Michaela Blott, Lisa Liu, Kimon Karras in Xilinx Labs Ireland
- Paul Hartke, Mark Paluszkiewicz in Xilinx Labs USA
- Brian Allison, Lance Thompson , Bruce Wile, IBM USA

Conclusion

- **Memcached and Key-Value stores show significant speedup and power savings on FPGAs, especially with small values**
- **Hybrid memory solutions are a disruptive opportunity in Datacenters for Database storage appliances**
- **Programming FPGAs with higher levels of abstraction is essential: C/C++/OpenCL**
- **For realistic workloads ONE Zynq Ultrascale+ MPSOC can implement a 80Gbs network with 40TB storage, and can run all 64 bit software**
- **MPSoC is a good idea.**

Selected References

- ANDERSEN, D. G., FRANKLIN, J., KAMINSKY, M., PHANISHAYEE, A.,TAN, L., AND VASUDEVAN, V. Fawn: A fast array of wimpy nodes. In Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles (2009), ACM, pp. 1–14.
- DEBNATH, B. K., SENGUPTA, S., AND LI, J. Flashstore: High throughput persistent key-value store. PVLDB 3, 2 (2010), 1414–1425.
- ATIKOGLU, B., XU, Y., FRACHTENBERG, E., JIANG, S., AND PALECZNY, M. Workload analysis of a large-scale key-value store. SIGMETRICS Perform. Eval. Rev. 40, 1 (jun 2012), 53–64.
- BLOTT, M., KARRAS, K., LIU, L., VISSERS, K., ISTVAN, Z., AND BAR, J. Achieving 10Gbps line-rate key-value stores with FPGAs. In Proceedings of HotCloud '13 (5th USENIX Workshop on Hot Topics in Cloud Computing),June 25-26, 2013, (San Jose, CA, USA, 2013).
- ISTV´AN, Z., ALONSO, G., BLOTT, M., AND VISSERS, K. A. A flexible hash table design for 10gbps key-value stores on FPGAS. In FPL (2013),pp. 1–8.
- LIM, H., HAN, D., ANDERSEN, D. G., AND KAMINSKY, M. Mica: A holistic approach to fast in-memory key-value storage. In 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14) (Seattle, WA, Apr. 2014), USENIX Association, pp. 429–444.
- OUYANG, X., ISLAM, N. S., RAJACHANDRASEKAR, R., JOSE, J., LUO, M., WANG, H., AND PANDA, D. K. SSD-assisted hybrid memory to accelerate memcached over high performance networks. In ICPP (2012), pp. 470–479.
- M. BLOTT, L. LIU, K. KARRAS, K. VISSERS, Scaling out to a Single-Node 80Gbps Memcached Server with 40Terabytes of Memory, in Proceedings of HotStorage '15 (7th Usenix Workshop on Hot Topics in Storage and File Systems), July 6-7, Santa Clara, USA