

High Performance, Scalable and Cost Optimized AWS Cloud Infrastructure for Chip Development

EDA + Cloud + X-Spot = Reduced Design Time + Higher Quality = Competitive advantage - costs

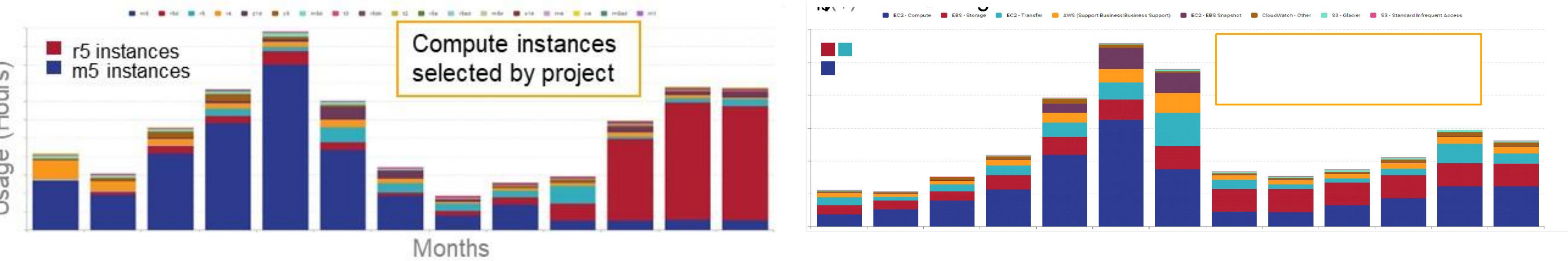
Jitendra Mohan¹, Shantanu Divekar¹, Srimai Inapurapu¹, Tanay Sharma¹, Vikash Tyagi¹, Vilesh Shah¹, Vivek Trivedi¹
Hakim Weatherspoon², Tian Ren², Zhiming Shen², Evan Chen³
¹Astera Labs, ²Exostellar, ³Synopsys



EDA on Cloud – Fast, Scalable ... Not Cheap [1]
Performance + Scale = Faster Time to Tapeout, Higher Quality Product

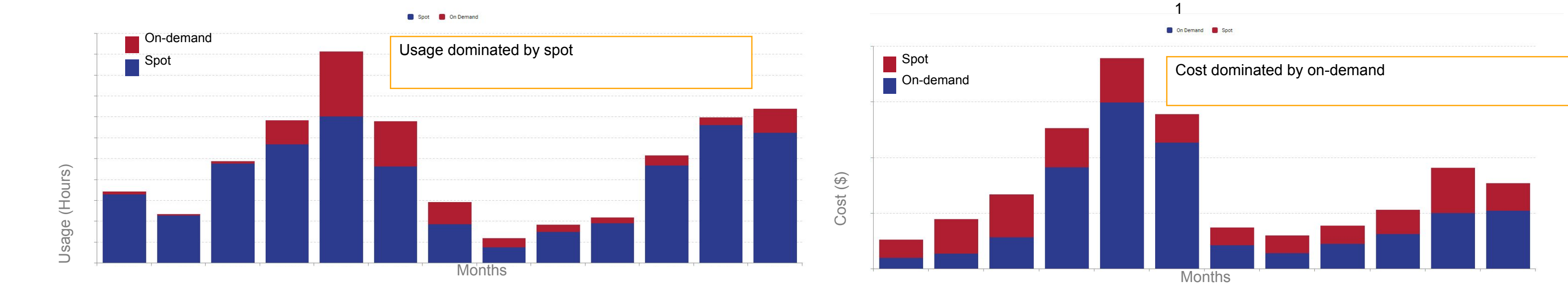
- Astera Labs pioneered 100% EDA on AWS cloud
 - Multiple tapeouts in advanced TSMC process nodes since 2018.
- AWS cloud has been great for performance and scale
 - Compute requirements fluctuate 10x within project timeline
 - Servers get faster every generation making on-prem infrastructure obsolete very quickly
 - c4/m4/r4 (2.4GHz) □ c5/m5/r5 (3.1GHz) □ z1d (4.1GHz) □ x2iezn (4.5GHz) □ r6i (3.5GHz) □ TBD
 - Front end RTL simulations require different compute than back end Physical Design
 - Right sized instances: c(4GB/core) □ m(8GB/core) □ r(16GB/core) □ x2(64GB/core)
 - Storage requirements have grown 100x: Self managed NFS file server □ FSx over OpenZFS

... not so good on costs. AWS spend has increased 10x since 2018



EDA on Cloud – Fast, Scalable ... AND Cheap (relatively!) [2]
Increased use of Spot Instances for Front End Simulations and Physical Design (PD)

- AWS costs dominated by on-demand instances
- AWS spot instances usually 70% cheaper
 - Can be interrupted anytime resulting in complete loss of compute job
 - Workflow must be adapted to tolerate machine failures
- Frontend simulation workflow: Astera uses VCS by Synopsys
 - Almost exclusively transitioned to use AWS Spot instances
 - Operationalize VCS built-in Save-Restore functionality
- Backend PD workflow: Various Synopsys tools
 - Almost always use on-demand instances
 - Require large memory instances and generate huge files
 - Conventional checkpointing slow and costly □ tools do not support checkpointing natively
- Use X-Spot technology by Exostellar for PD workflows
 - AI based algorithm to predict spot instance interruption
 - Automatically migrate jobs between Spot and On-Demand instances

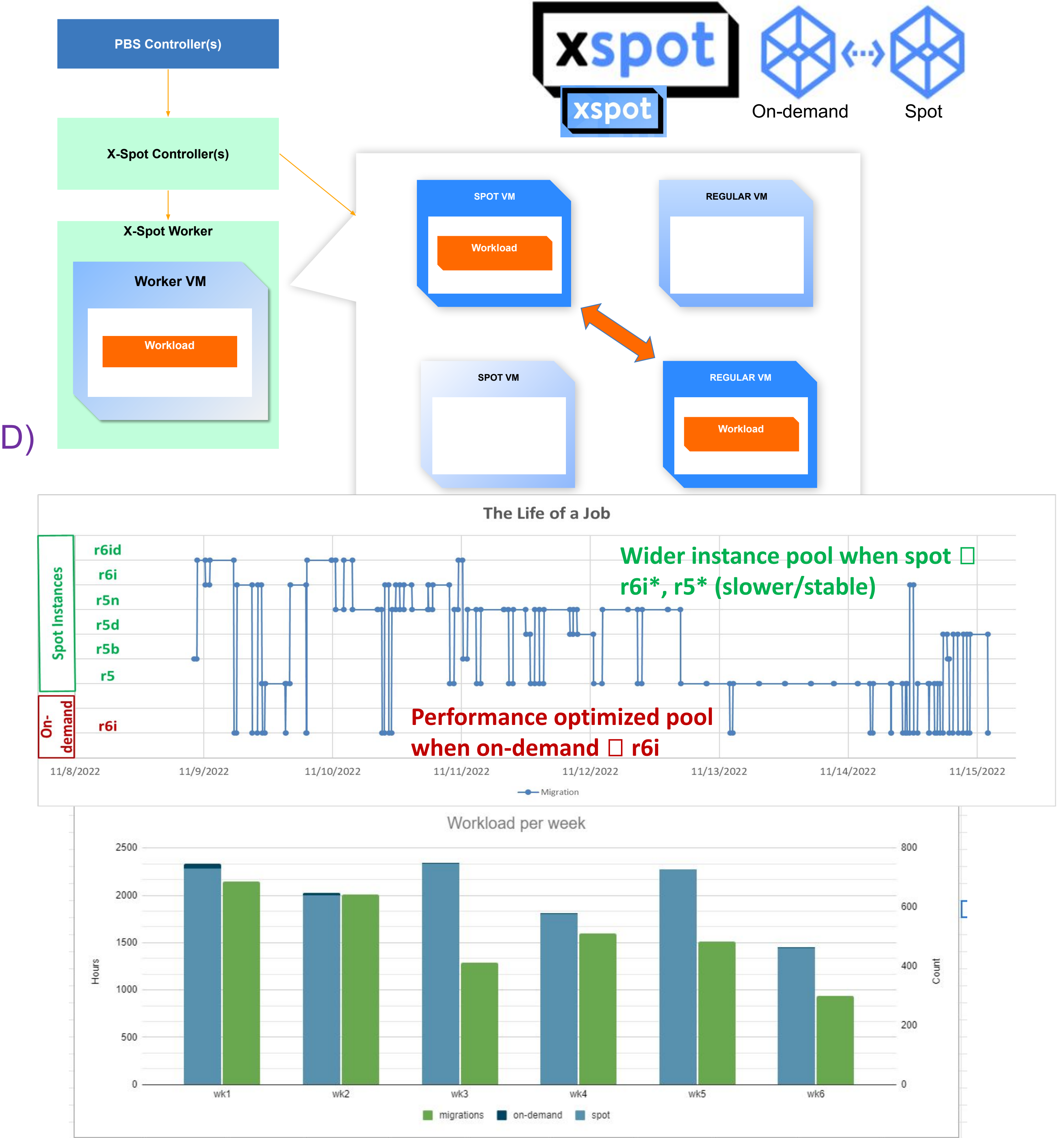


More X-Spot Performance and Reliability Results [5]

Tool	Version	VM	X-Spot	X-Spot performance
		Time (s)	Time (s)	
Design Compiler	R-2020.09-SP3a	2,796	2,866	98%
IC Compiler-II	Q-2019.12-SP6	11,364	11,506	99%
Fusion Compiler	T-2022.03	352,822	363,041	97%
Star-RC	P-2019.03-SP5	5,740	5,727	100%
Prime Time	Q-2019.12-SP5-2	14,974	15,224	98%
IC Validator	S-2021.06-SP2	28,600	28,406	101%
Back-to-back 49 migrations between r6i and r5 instances – minimal runtime impact		r6i VM	r6i-r5 X-Spot	X-Spot performance
Tool	Version	Time (s)	Time (s)	
IC Validator	S-2021.06-SP2	22,537	23,478	96%

Physical Design Workflow Using Exostellar X-Spot [3]
Reliability of On-Demand Instances at the Cost of Spot Instances!

- Jobs run on workers which provide an unchanging virtual environment.
- Jobs are scheduled using existing PBS system and everything else is handled in the background.
- Workers are automatically migrated between on-demand and spot instances
 - On-demand: r6i, Spot: r6i/r5*
- Meanwhile, the job continues to run without interruption!



Performance and Reliability Results [4]
Significant cost savings with minimal performance and reliability impact

- Scalable and cost optimized EDA infrastructure in full production
- Novel approaches to increase usage of AWS spot instances
- Use of Exostellar X-Spot to run long workloads on spot machines
- Reliably run Synopsys EDA tools over X-Spot platform:
 - Average workload performance: 99% of on-demand VMs
 - Longest workload run time: 30+ days
 - Reliability: 99.5%
 - Number of surprise interruptions: 78 jobs (out of 17,019) over 9+ months
- Significant cost savings
 - Workload execution on spot instances: 90+%

Workload Reliability Analysis Over 9+ Months of Deployment		
Job Lifetime Analysis	Total number of jobs	17,019
	Average job lifetime (days)	0.7
	Max job lifetime (days)	30.0
	Percent of time job is on spot instance	91%
	Surprise job terminations	78
Spot Instance Lifetime Analysis	Average lifetime (hours)	3.4
	Longest lifetime (hours)	11.5
	Number of spot instances created	140,490
Migration Analysis	% jobs migrated	55.%
	% jobs migrated over 10 times	16.9%
	Max # of migrations for a single job	229