Argonne NATIONAL LABORATORY

ALCF AI Testbed

Argonne Leadership Computing Facility – Enabling Breakthroughs in Science and Engineering

Venkatram Vishwanath
Argonne Leadership Computing Facility
venkat@anl.gov

March 9, 2022

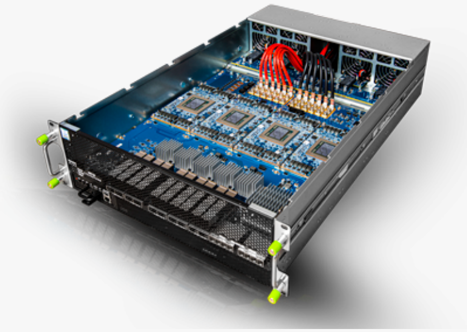
ALCF AI Testbeds

https://www.alcf.anl.gov/alcf-ai-testbed

Cerebras (CS-2)

SambaNova

Graphcore



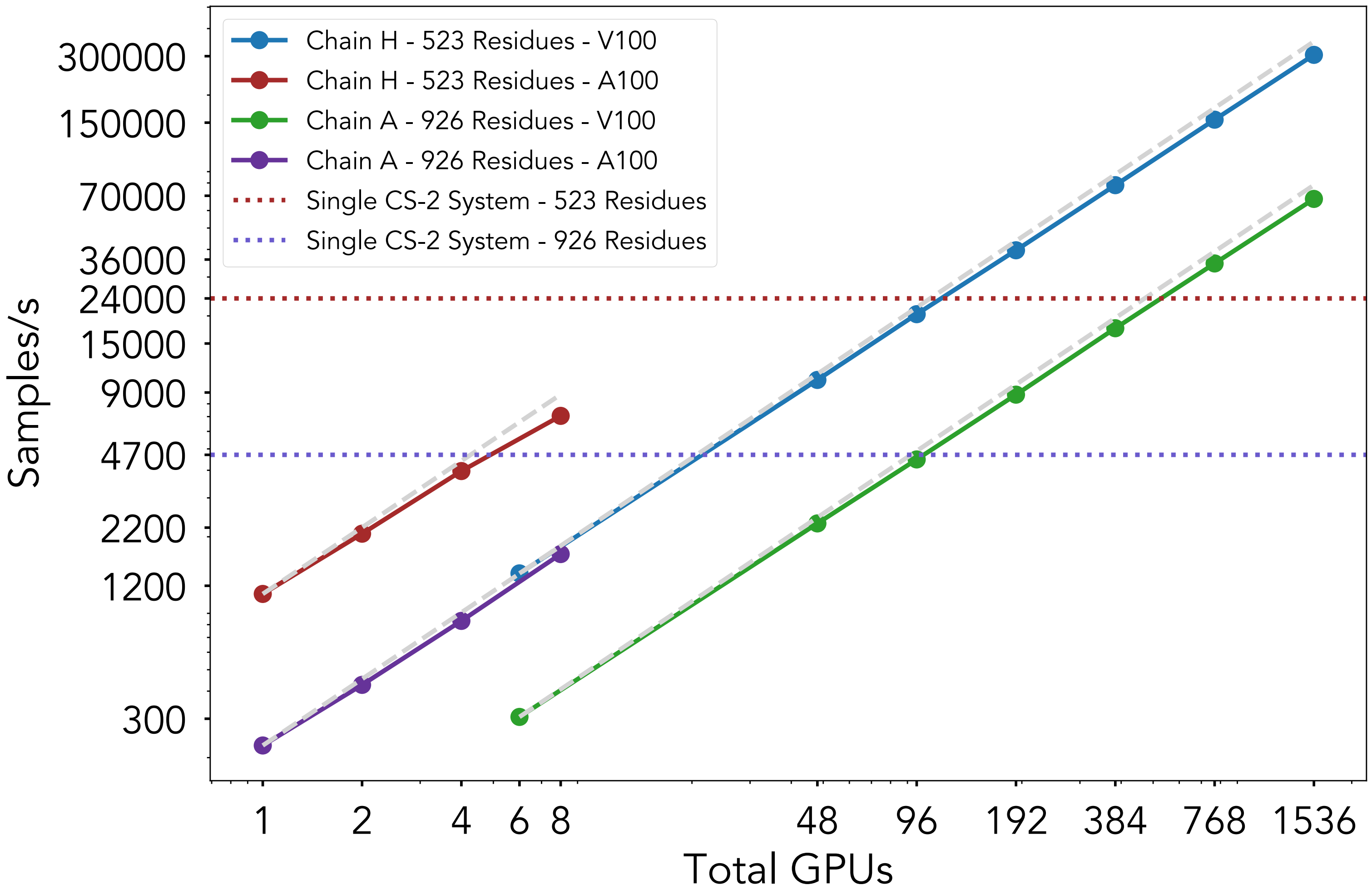
Habana

Groq

- Infrastructure of next-generation machines with hardware accelerators customized for artificial intelligence (AI) applications.
- Provide a platform to evaluate usability and performance of machine learning based HPC applications running on these accelerators.
- The goal is to better understand how to integrate AI accelerators with ALCF's existing and upcoming supercomputers to accelerate science insights

	Cerebras CS2	SambaNova	Groq	GraphCore (MK1)	Habana Gaudi	NVIDIA A100
Compute Units	850,000 Cores	640 PCUs	5120 vector ALUs	1472 IPUs	8 TPC + GEMM engine	6912 Cuda Cores
On-Chip Memory	40 GB	>300MB	230MB	900MB	-	192KB L1 40MB L2
Process	7nm	7nm	14nm	7nm	7nm	7nm
System Size	2 Nodes	2 nodes (8 cards per node)	4 nodes (8 cards per node)	1 node (8 cards per node)	2 nodes (8 cards per node)	1 card
Estimated Performance of a card (TFlops)	>80,000	>300 (BF16)	>205 (FP16)	>125 (FP16)	>150 (FP16)	312 (FP16), 156 (FP32)
Software Stack Support	Tensorflow, Pytorch	SambaFlow, Pytorch	GroqAPI, ONNX	Tensorflow, Pytorch, PopArt	Synapse AI, TensorFlow and PyTorch	Tensorflow, Pytorch, etc

Acceleration of CVAE on Summit and Cerebras CS-2

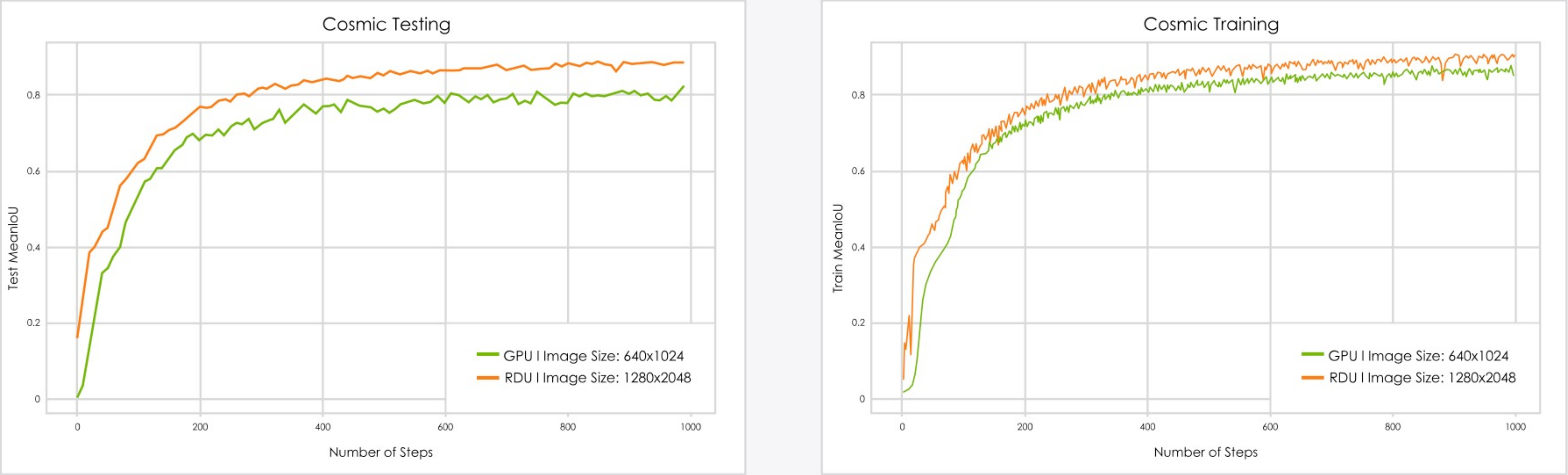


- Single CS-2 delivers performance of over 100 GPUs on CVAE
- Results are for **out-of-the-box performance** based on model config not optimized for CS-2.

Performance	**523 X 523**	**926 X 926**
Throughput (samples/sec)		
1x CS-2 System	24,000	4700
1x V100 GPU	228	51
1x A100 GPU	~1100	~150
Speedup (CS2 vs. GPU ideal/actual)		
1 x V100 GPU	105x/113x	92x/101x
1 x A100 GPU	~22X	~32X

Intelligent Resolution: Integrating Cryo-EM with AI-driven Multi-resolution Simulations to Observe the SARS-CoV-2 Replication-Transcription Machinery in Action, SC21 COVID19 Gordon Bell Finalist, To appear in IJHPCA 2022
https://www.biorxiv.org/content/10.1101/2021.10.09.463779v1.full.pdf

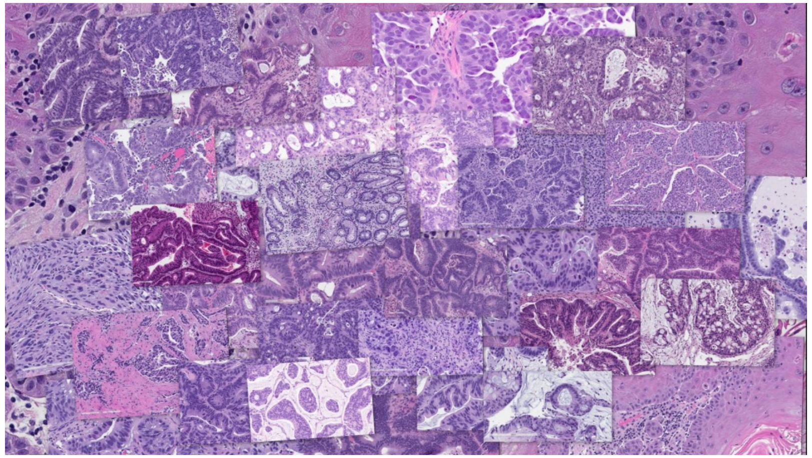
COSMIC TAGGER ON SAMBANOVA DATASCALE



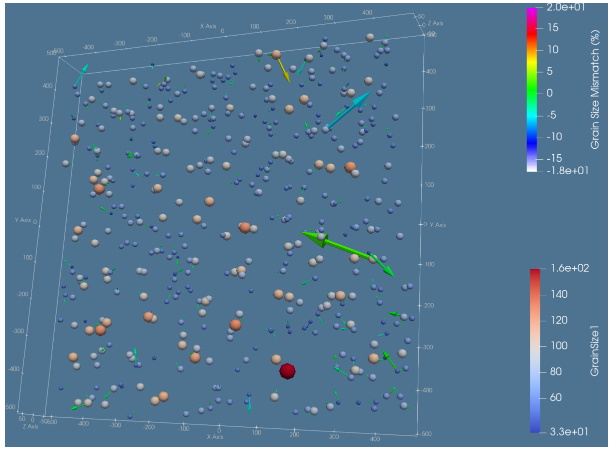
Sambanova RDUs able to accommodate larger image sizes and achieve higher accuracy

M. Emani et al., "Accelerating Scientific Applications With SambaNova Reconfigurable Dataflow Architecture," in Computing in Science & Engineering, vol. 23, no. 2, pp. 114-119, 1 March-April 2021, doi: 10.1109/MCSE.2021.3057203.

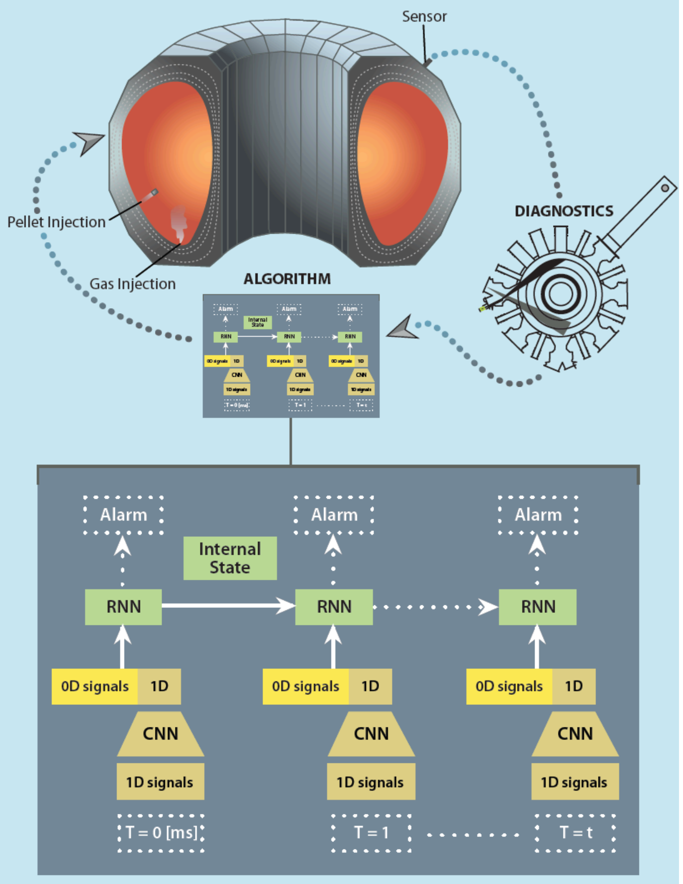
AI FOR SCIENCE APPLICATIONS ON AI TESTBED



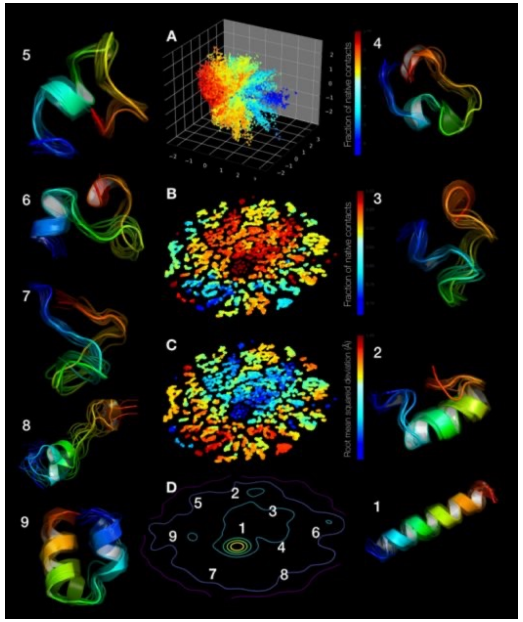
Cancer drug response prediction



Imaging Sciences-Braggs Peak



Tokomak Fusion Reactor operations



Protein-folding(Image: NCI)

and more..

Getting Started on ALCF AI Testbed:

Apply for a Director’s Discretionary (DD) Award

Director’s Discretionary (DD) awards support various project objectives from scaling code to preparing for future computing competition to production scientific computing in support of strategic partnerships.

https://www.alcf.anl.gov/science/directors-discretionary-allocation-program

Acknowledgements

- This research was funded in part and used resources of the Argonne Leadership Computing Facility (ALCF), which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

- William Arnold, Bruce Wilson, Varuni Sastry, Sid Raskar, Murali Emani, Corey Adams, Rajeev Thakur, Arvind Ramanathan, Alex Brace, Hyunseung (Harry) Yoo, Ryan Aydelott, Craig Stacey, Mike Papka and others contributed to the material

venkat@anl.gov