Oak Ridge National Laboratory

Early Application Experience on the NVIDIA Arm Devkit

Wael Elwasif[1], Sergei Bastrakov[2], Spencer H. Bryngelson[3], Michael Bussmann[2], Sunita Chandrasekaran[4], Florina Ciorba[5], Kate Clark[5], Alexander Debus[2], William Godoy[1], Nick Hagerty[1], Jeff Hammond[5], David Hardy[7], J. Austin Harris[1], Oscar Hernandez[1], Balint Joo[1], Sebastian Keller[8], Paul Kent[1], Henry Le Berre[3], Damien Lebrun-Grandie[1], Elijah MacCarthy[1], Verónnica G. Melesse Vergara[1], Bronson Messer[1], Ross Miller[1], Sarp Oral[1], Jean-Guillaume Piccinali[8], Anand Radhakrishnan[3], Osman Simsek[5], Filippo Spiga[5], Klaus Steiniger[2], Jan Stephan[2], John E. Stone[5], Christian Trott[9], René Widera[2], Jeffrey Young[3]

Oak Ridge National Laboratory[1], Helmholtz-Zentrum Dresden Rossendorf[2], Georgia Institute of Technology[3], University of Delaware[4], University of Basel[5], NVIDIA[6], University of Illinois Urbana-Champaign[7], Swiss National Supercomputing Center (CSCS)[8], Sandia National Laboratory[9]

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725

OAK RIDGE NATIONAL LABORATORY'S FRONTIER SUPERCOMPUTER

- 74 HPE Cray EX cabinets
- 9,408 AMD EPYC CPUs, 37,632 AMD GPUs
- HPE Slingshot 11 interconnect
- 700 petabytes of storage capacity, peak write speeds of 5 terabytes per second using Cray Clusterstor Storage System

TOP500

#1

ORNL’s Frontier supercomputer is #1 on the TOP500.

1.1 exaflops of performance on the May 2022 Top500 list.

GREEN500

#1

ORNL’s Frontier supercomputer is #1 on the GREEN500.

52.23 gigaflops/watt power efficiency.

HPL-AI

#1

ORNL’s Frontier supercomputer is #1 on the HPL-AI list.

6.88 exaflops on the HPL-AI benchmark.

Sources: May 30, 2022 Top500 release

Why Do We Care About ARM Architecture?

- Exploring technologies that might be viable in the next few years
  - Essentially, we're looking for technologies we might use in the next supercomputer
- "AArch64 for HPC" has been a discussion topic for years
  - Astra, Fugaku, Graviton, …
- AArch64 has been on the OLCF's radar since about 2017
  - The Wombat cluster started in 2018
  - Using GPUs on AArch64 started in September of 2019
  - Added A64fx nodes in mid-2020 and NVIDIA ARM DevKit nodes in late-2021
- OLCF doesn't have any AArch64 hardware deployed in production

ARM: Wombat testbed at NCCS

Administrator:
Ross Miller (ORNL)

Hardware:

- 4 HPE Apollo 70 nodes:
  - ThunderX2 CN9980 CPU
  - 2x NVIDIA V100 GPUs
- 16 HPE Apollo 80 nodes
  - Fujitsu A64FX CPUs
- 8 NVIDIA ARM Developer Kit nodes
  - Ampere Altra, 80 Core CPU
  - 2x NVIDIA A100 GPUs
- Infiniband EDR and HDR

Software:

- Centos 8.1
- CUDA 11
- Compilers
  - NVHPC/22.1
  - ARM HPC Compiler 22.0.2
  - LLVM 15.0
  - GCC 10.2 and 11.2
- Open MPI 4.0.5

Website:
https://www.olcf.ornl.gov/olcf-resources/compute-systems/wombat/

Comparison of Summit vs Wombat Nodes

System Specs	Summit	Wombat TX2	Wombat NVDevkit
CPU	IBM POWER9™ (2 Sockets * 21 Cores / socket)	Marvell ThunderX2 (2 Sockets * 28 Cores / socket)	**Ampere Altra Q80-30 (80 Cores, 1 Socket)**
GPU	NVIDIA Volta (6 per node)	NVIDIA Volta (2 per node)	**NVIDIA A100 (2 per node)**
RAM	512 GB DDR + 96 GB HBM (16 GB per GPU)	256 GB DDR + 64 GB HBM (32 GB per GPU)	**512 GB DDR4 + 80 GB HBM (40 GB per GPU)**
On-node interconnect	NVIDIA NVLINK2 (50 GB/s) Coherent memory across the node	PCIe Gen3 (16 GB/s) No coherence across the node	**PCIe Gen4 (32 GB/s) No coherence across the node**
System Interconnect	Mellanox Dual-port EDR IB network 25 GB/s	Mellanox Single-port EDR IB network 12.5 GB/s	**2x NVIDIA BlueField-2 E-Series DPU, 200GbE/HDR single-port QSFP56**
On-node Storage	1600GB NVME	480GB SATA SSD	**2 TB SATA Spinning Disk**

Evaluating ARM + GPU : 20 unique codes

2019 : ThunderX2 + V100

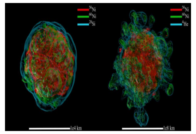
- 12 Codes
- ***LAMMPS, MiniSweep, NAMD/VMD***
- BabelStream
- Comet
- DCA++
- Gamera
- Gromacs
- LSMS
- Patatrack
- SNAP
- Tea Leaf

2022 : Ampere Altra + A100

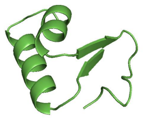
- 11 Codes
- ***LAMMPS, MiniSweep, NAMD/VMD***
- Exastar
- GPU-I-TASSER
- MFC
- MILC
- PIConGPU
- QMCPACK
- SpecHPC
- SPH-EXA2

NVIDIA ARM DevKit Evaluation Phase 1 : HPC mod/sim

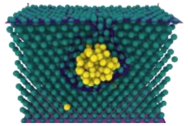
Scientific Applications:



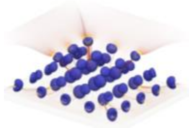
Exastar
Stellar Astrophysics



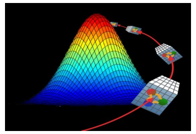
GPU-I-TASSER
Bioinformatics



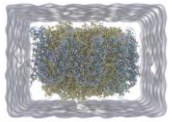
LAMMPS
Molecular Dynamics



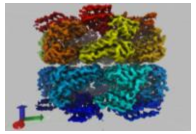
MFC
Fluid Dynamics



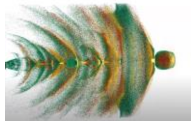
MILC
QCD



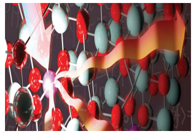
NAMD
Molecular Dynamics



VMD
MD Viz.



PIConGPU
Plasma Physics

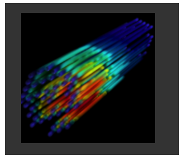


QMCPACK
Chemistry

SPH-EXA2
Hydrodynamic

Benchmarks & Mini-apps:

SpecHPC 2021
Benchmark



MiniSweep
Radiation Transport

Parallel Programming Models and Tools

kokkos

Kokkos
C++ Prog. Model

OPEN MPI

Open MPI
Distributed Prog. Model

NVIDIA

CUDA / NVHPC, NSIGHT
Prg. Env and Tools

arm

Arm Compilers Performance Libraries, Allinea
Prog Env and Tools

LLVM COMPILER INFRASTRUCTURE

GCC

UCX

OpenACC Directives for Accelerators

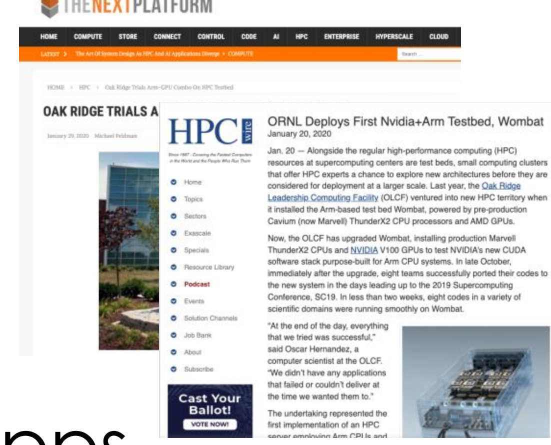
OpenMP

Evaluation done by:

Oak Ridge National Laboratory · I-TASSER Protein Structure & Function Predictions · Georgia Tech · Sandia National Laboratories · Illinois University of Illinois at Urbana-Champaign · SKACH · Universität Basel · CSCS Centro Svizzero di Calcolo Scientifico Swiss National Supercomputing Centre · HZDR Helmholtz Zentrum Dresden Rossendorf · University of Delaware · NVIDIA

Evaluation Overview

- Builds on the work done for the 2019 evaluation
  - https://www.olcf.ornl.gov/wp-content/uploads/2019/11/ARM_NVIDIA_APP_EVALUATION-1.pdf



THE NEXT PLATFORM

HOME COMPUTE STORE CONNECT CONTROL CODE AI HPC ENTERPRISE HYPERSCALE CLOUD

OAK RIDGE TRIALS A

HPCwire

Home · Topics · Sectors · Exascale · Specials · Resource Library · Podcast · Events · Solution Channels · Job Bank · About · Subscribe

Cast Your Ballot! VOTE NOW!

ORNL Deploys First Nvidia+Arm Testbed, Wombat

January 20, 2020

Jan. 20 — Alongside the regular high-performance computing (HPC) resources at supercomputing centers are test beds, small computing clusters that offer HPC experts a chance to explore new architectures before they are considered for deployment at a larger scale. Last year, the Oak Ridge Leadership Computing Facility (OLCF) ventured into new HPC territory when it installed the Arm-based test bed Wombat, powered by pre-production Cavium (now Marvell) ThunderX2 CPU processors and AMD GPUs.

Now, the OLCF has upgraded Wombat, installing production Marvell ThunderX2 CPUs and NVIDIA V100 GPUs to test NVIDIA's new CUDA software stack purpose-built for Arm CPU systems. In late October, immediately after the upgrade, eight teams successfully ported their codes to the new system in the days leading up to the 2019 Supercomputing Conference, SC19. In less than two weeks, eight codes in a variety of scientific domains were running smoothly on Wombat.

"At the end of the day, everything that we tried was successful," said Oscar Hernandez, a computer scientist at the OLCF. "We didn't have any applications that failed or couldn't deliver at the time we wanted them to."

The undertaking represented the first implementation of an HPC

2019

- Apps were typical HPC modeling and simulation apps
  - Phase 1: 11 HPC apps from 9 institutions
- Emphasis mostly on the Ampere nodes, but some teams tested on the other nodes, too.
  - We avoided direct comparisons of results from different node types because the hardware was too disparate
  - Some teams were interested in the SVE support on the A64fx CPU's
- *Phase 2 (planned): AI & ML workloads*
- *Phase 3 (planned): Focus on Bluefield DPUs*

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

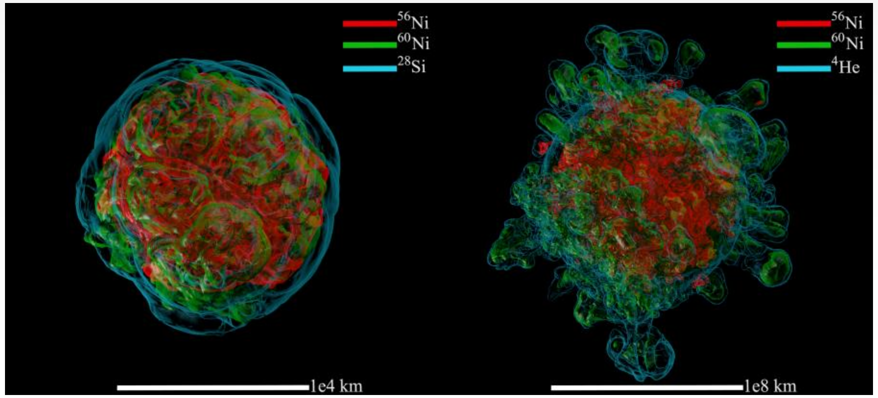
ExaStar (thornado)

Austin Harris, Bronson Messer
Oak Ridge National Laboratory

Application information

- Science Domain & Method
  - Computational stellar astrophysics, spectral neutrino radiation transport
- Programming Models and languages
  - Fortran and OpenACC
- HPC systems where app is used in production
  - Summit, Crusher, Frontier*, Perlmutter*, Aurora*

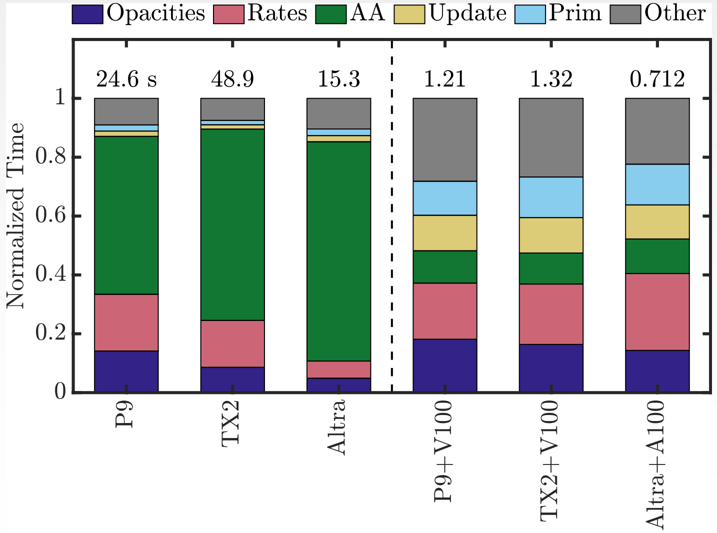
  *Planned



Isotope distributions for early and late stages of core-collapse supernovae simulation

WOMBAT ARM+NVIDIA DevKit Experience

- Easy to transition Summit to Wombat (NVHPC)
- Minimal modifications of build system to use OpenBLAS instead of IBM's ESSL for CPU libraries
- Wombat speedup relative to Summit (1 rank):
  - Altra to POWER9 (multicore): 1.6x – 2.2x
  - Altra+A100 to POWER9+V100: 1.3x – 1.7x



Normalized time of bottleneck kernels. Wall-clock times above each bar.

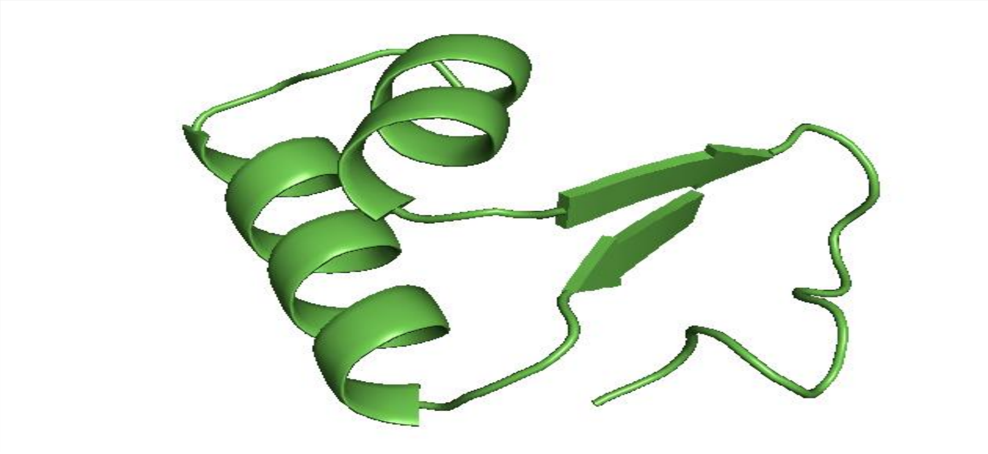
Preliminary Results

GPU-I-TASSER

Elijah MacCarthy, Verónica Melesse Vergara
Oak Ridge National Laboratory

Application information

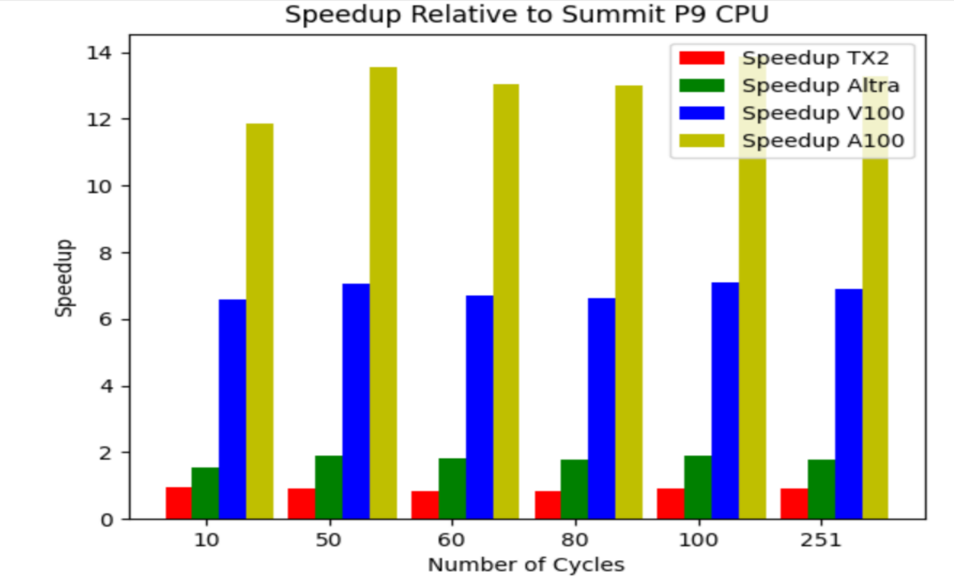
- Science Domain & Method
  - Computational Biology, Bioinformatics
- Programming Models and languages
  - Fortran and OpenACC
- HPC systems where app is used in production
  - PSC Bridges2



Predicted structure of 1YV8 protein from *Crambe hispanica subsp. abyssinica*

WOMBAT ARM+NVIDIA DevKit Experience

- Compiled on Wombat with updated NVHPC 22 compiler.
- Wombat's Ampere Ultra CPU recorded speed-up of 1.8x relative to Summit's P9 CPU.
- A100s on Wombat resulted in 2x speed-up compared to NVIDIA V100s.

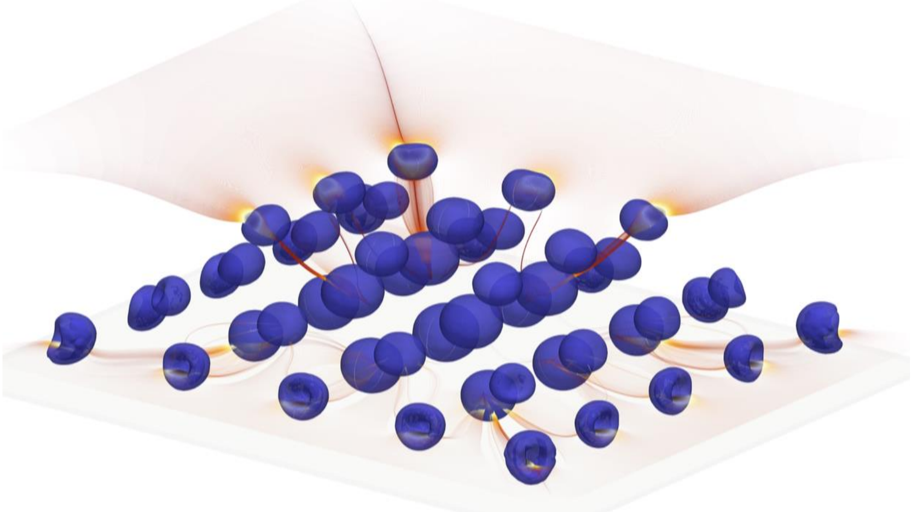


Preliminary Results

A. Radhakrishnan, H. Le Berre, S. H. Bryngelson
Georgia Institute of Technology

Application information

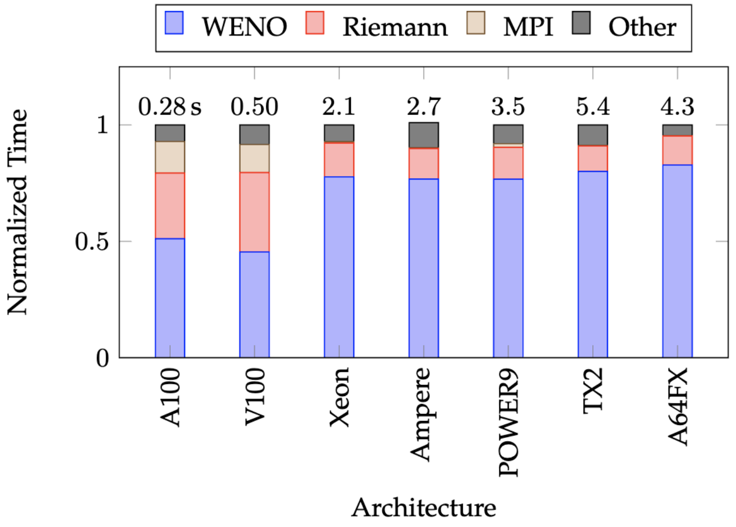
- Science Domain & Method
  - Multiphase Computational Fluid Dynamics
- Programming Models and languages
  - Fortran90, OpenACC 2.6, MPI/RDMA
- HPC systems where app is used in production
  - OLCF Summit, PSC Bridges2, SDSC Expanse



Bubbles cavitating and collapsing near a wall.
Streamlines are shown.

WOMBAT ARM+NVIDIA DevKit Experience

- Sufficient out-of-the-box support for Arm environments with NVHPC 22, GCC11
- Ampere Arm CPU competitive with latest Xeon CPUs, A100s offer expected speedups
- 2x speed-up over a Summit node with same number of GPUs



Most expensive routines. **MPI** is all communication costs.
Wall time is atop each architecture bar.

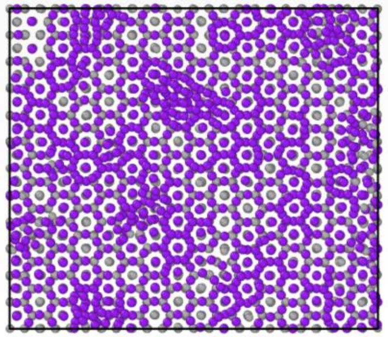
Preliminary Results

LAMMPS and Kokkos

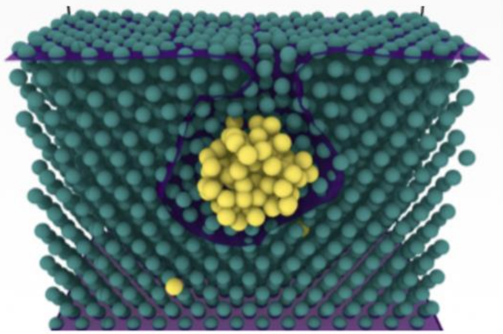
Damien Lebrun-Grandie (ORNL)
Christian Trott (SNL)

Application information

- Science Domain & Method
  - Classical molecular dynamic code with focus on material modeling
  - Domain decomposition + shared memory parallelism
- Programming Models and languages
  - MPI+Kokkos and C++
- HPC systems where app is used in production
  - OLCF Summit, NERSC Perlmutter, etc.



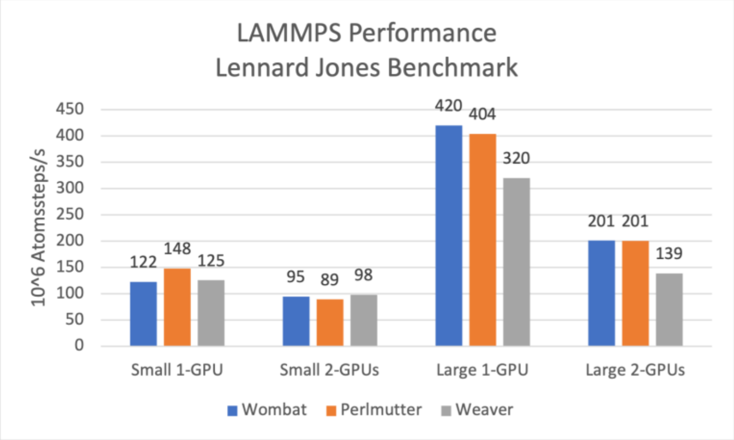
Beryllium depositing on tungsten surface (SNAP Wbe)



Helium bubble forming near tungsten surface

WOMBAT ARM+NVIDIA DevKit Experience

- Kokkos already supported ARM CPUs and all current NVIDIA GPU architectures
  - Wombat: `KOKKOS_ARCH="Ampere80"`
  - Weaver: `KOKKOS_ARCH="Power9;Volta70"`
  - Perlmutter: `KOKKOS_ARCH="ZEN3;Ampere80"`
- Noteworthy observations on the platform and software stack
  - Latency-sensitive scenario performs worse on Wombat
  - Otherwise perform similarly on Wombat and X86

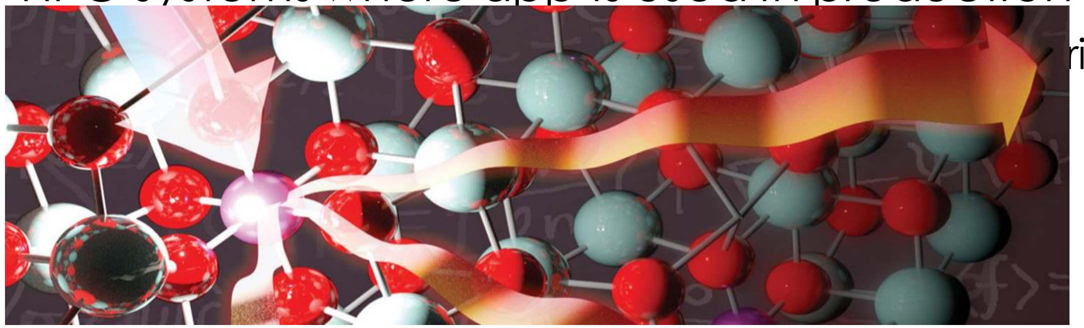


Preliminary Results

William F Godoy, Paul Kent
Oak Ridge National Laboratory

Application information

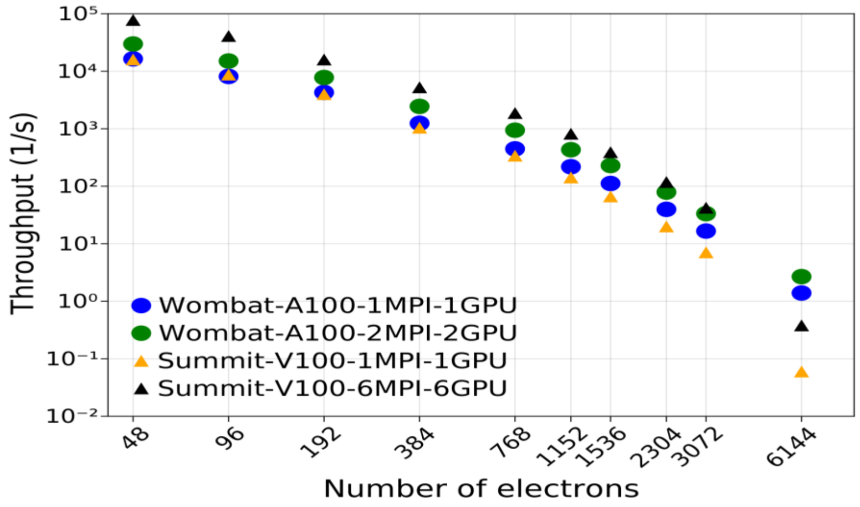
- Science Domain & Method
  - Highly-accurate Quantum Monte Carlo for predicting the quantum-mechanical properties of materials, molecules, and atoms. Open source.
- Programming Models and languages
  - C++17, OpenMP target offload, CUDA/HIP/SYCL wrappers, Python workflow automation (Nexus). Older CUDA-only implementation tested.
- HPC systems where app is used in production



Optical Properties of Correlated Defects via QMC

WOMBAT ARM+NVIDIA DevKit Experience

- Sufficient out-of-the-box support for Arm environments with NVHPC22 for GPU
- Ampere Arm CPU competitive with latest Xeon CPUs on single threads, need to evaluate Arm performance libraries with OpenMP
- Competitive with Summit node performance (6 V100s) with fewer number of GPUs (2 A100s)



Greater scalability (higher is better) on Wombat vs Summit nodes using GPUs for NiO benchmark problem

Preliminary Results

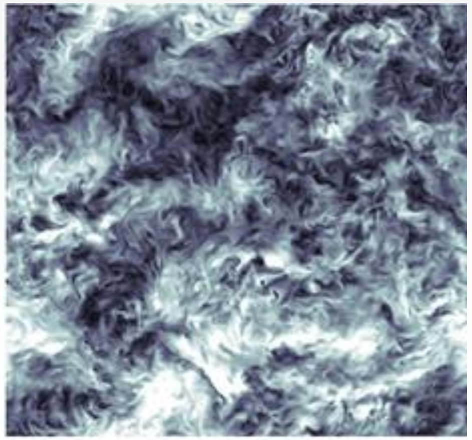
SPH-EXA2

https://github.com/unibas-dmi-hpc/SPH-EXA

F. M. Ciorba (UniBas), S. Keller (CSCS), J.G. Piccinali (CSCS), O. S. Simsek (UniBas)

Application information

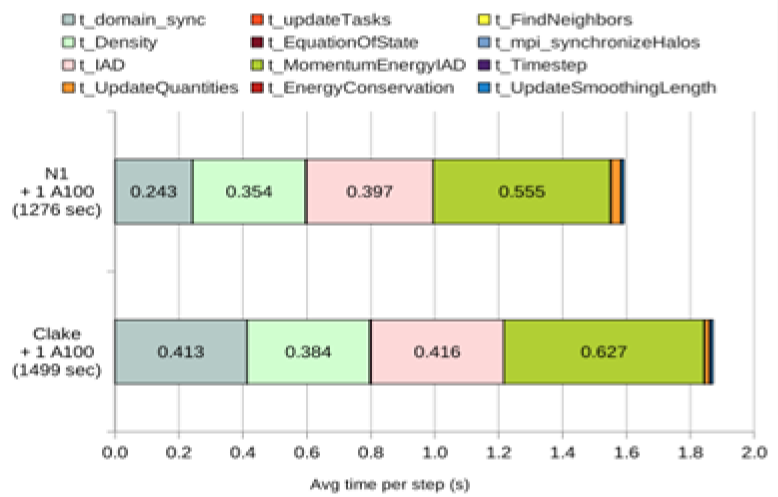
- Science Domain & Method
  - Smoothed Particle Hydrodynamics and gravity for Astrophysics and Cosmology
- Programming Models and Languages
  - C++ code, parallelized with MPI and OpenMP, and accelerated with CUDA and HIP
- HPC systems where app is used in production
  - Piz Daint, LUMI-G



High resolution subsonic turbulence

WOMBAT ARM+NVIDIA DevKit Experience

- Straightforward porting w/ GCC, ARM, NVCC
- Performance experience and comparisons
  - CPU-only execution performance is limited on ARM CPUs compared to Intel Xeon CPUs
  - ARM Ampere CPUs deliver competitive performance on the GPU-accelerated version
- HPC tools for ARM are quite mature and useful for future ARM systems



Execution times of SPH-EXA2 kernels on ARM N1 and Intel Clake CPUs with NVidia GPUs

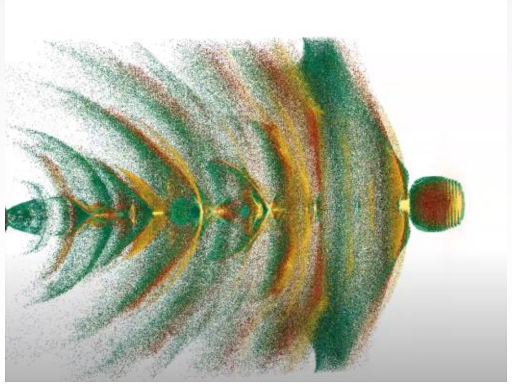
Preliminary Results

PIConGPU

S. Bastrakov[1], M. Bussmann[1], S. Chandrasekaran[2], A. Debus[1], K.Steiniger[1], J. Stephan[1], R. Widera[1]
[1]Helmholtz-Zentrum Dresden-Rossendorf, [2]University of Delaware

Application information

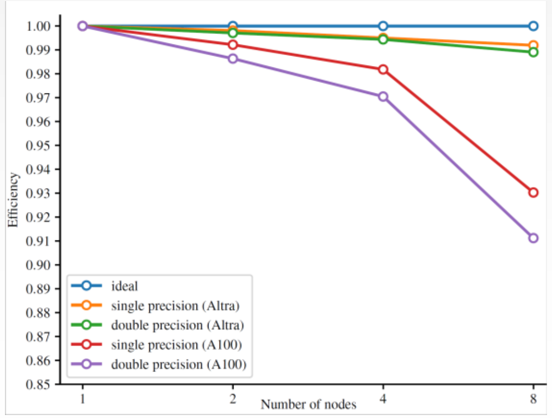
- Science Domain & Method
  - Radiation physics & particle-in-cell simulations
- Programming Models and languages
  - C++ & *alpaka* abstraction layer over compute APIs (CUDA, HIP, SYCL, OpenMP, …)
- HPC systems where PIConGPU is used for production
  - ORNL Frontier, Summit, Crusher, NERSC's Perlmutter, JSC JUWELS, TUD Taurus

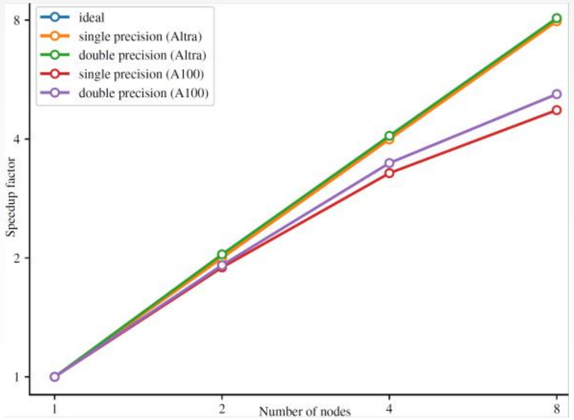


Still image from an uncut LWFA simulation video using Summit and 48 V100s using ISAAC 1.5

WOMBAT ARM+NVIDIA DevKit Experience

- ARM CPUs and NVIDIA GPUs seamlessly supported by *alpaka* (using the OpenMP back-end) → no porting effort required
- Modification of build system required to use *armclang* compiler
- Weak and strong scaling working almost ideal for Ampere Altra CPUs, stronger MPI impact with GPUs due to GPU computation speed





Preliminary Results

SPEChpc

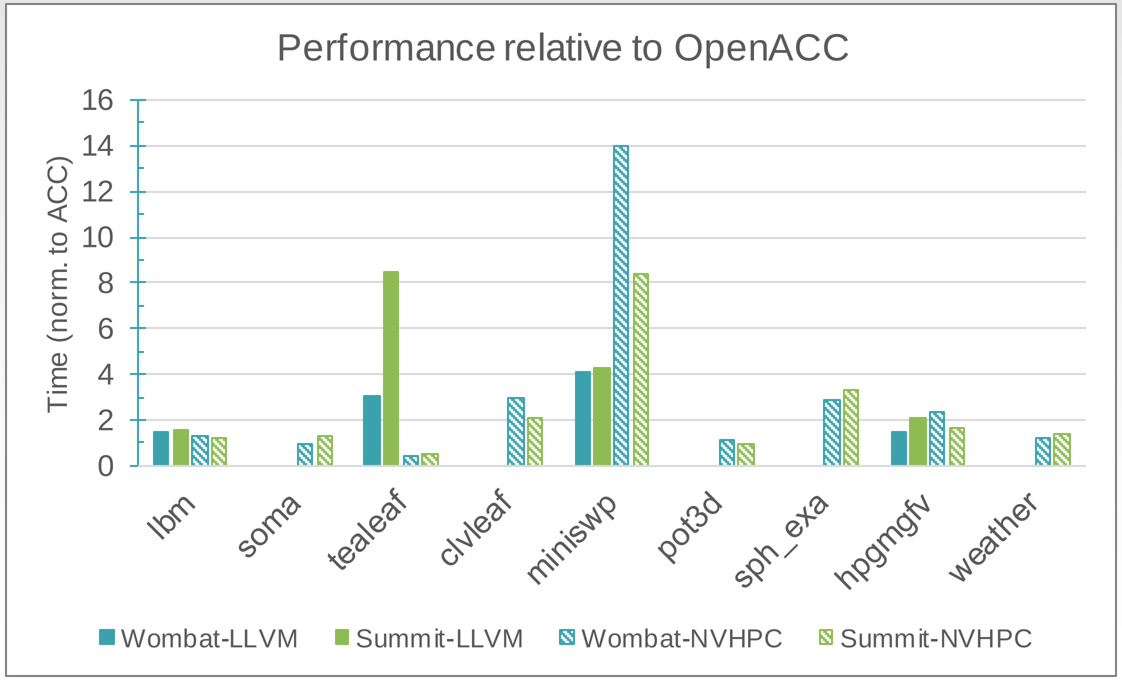
Nick Hagerty, Verónica Melesse Vergara
ORNL

Application information

- Science Domain & Method
  - Portable benchmarking of real-world applications
- Programming Models and languages
  - C, C++, and Fortran (NVHPC only)
  - OpenACC offload via NVHPC 22.1
  - OpenMP TGT via LLVM 15.0.0, NVHPC 22.1
- HPC systems where app is used in production
  - Summit (ORNL), Frontera (Texas Advanced Computing Center), Hemera (Helmholtz-Zentrum Dresden – Rossendorf)

WOMBAT ARM+NVIDIA DevKit Experience

- Availability of both LLVM and NVHPC on both Wombat and Summit eased transition
- Wombat and Summit exhibit consistent OpenMP performance trends

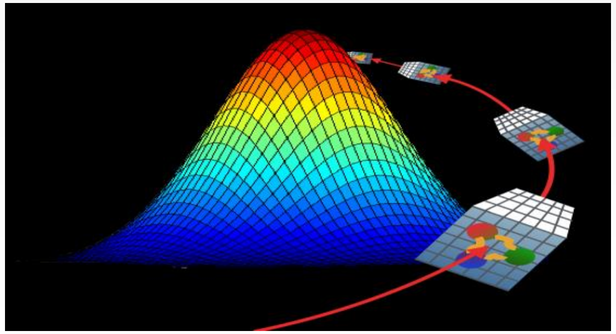


Preliminary Results

Kate Clark[1], Bálint Joó[2]
[1]NVIDIA, [2]Oak Ridge National Laboratory

Application information

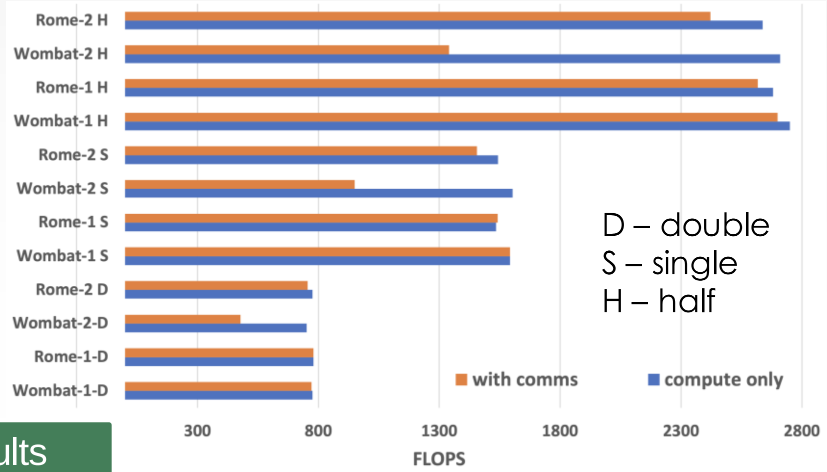
- Science Domain & Method
  - Lattice Quantum Chromodynamics (LQCD)
  - Dominated by iterative linear solve (CG) of discretized PDE (stencil) on 4-d grid
- Programming Models and languages
  - Utilizes QUDA library for GPU acceleration
  - C (MILC) and CUDA C++ (QUDA)
- HPC systems where app is used in production
  - OLCF Summit, NERSC Perlmutter, etc.



Lattice QCD gauge generation codes like the MILC benchmark, sample the strong force field distribution in the vacuum, by generating one gauge field after the other using a process of Hybrid Monte Carlo updates.

WOMBAT ARM+NVIDIA DevKit Experience

- Both MILC and QUDA compiled out of the box on Wombat with no changes required
- Benchmarked with NERSC MILC generation tests
- Compared to AMD Rome + NVIDIA A100 system
  - Wombat's Ampere Ultra CPU faster than AMD ROME for running MILC CPU code
  - Single GPU performance identical
  - Scaling limited by lack of p2p on Wombat
  - Stencil with/without communication:



D – double
S – single
H – half

Preliminary Results

Observations: What Worked, What Didn't, What Are The Gaps

- Good news: Every team was able to build and run. Most teams required no special effort.
  - Most apps have **not** been optimized for every last FLOP of performance, though. (NAMD & VMD are exceptions – they are heavily optimized!)
- The software ecosystem has matured significantly since 2019
  - CUDA on AArch64 is officially supported. (The 2019 work used an early beta.)
  - OpenACC offloading works
  - More mature Spack package management on Arm
- Profiling tools were not exercised much but were used (NSIGHT, etc)
  - More effort needed to fully evaluate platform readiness.
  - Optimization not major focus of this effort

Observations: What Worked, What Didn't, What Are The Gaps

- Containers:
  - Not needed for this phase, more important for AI/ML in Phase 2
- GPU Direct, CPU-GPU Latency, CPU-GPU coherency
  - Latency-sensitive applications underperform (CPU-GPU data transfer, GPUDirect, ..)
  - ***Waiting for Grace-Hopper***
- Offloading programming models
  - Cuda, OpenACC (via NVHPC) main offloading model
  - OpenMP offloading (LLVM, NVHPC) not exercised in this phase
  - Offloading using C++ std library: need production app
- DPU's (in HPC) yet to be explored
  - Use case co-design - focus of Phase 3

Evaluation Team

Wael Elwasif, William Godoy, Nick Hagerty, J. Austin Harris, Balint Joo, Paul Kent, Damien Lebrun-Grandie, Elijah MacCarthy, Verónica G. Melesse Vergara, Bronson Messer, Ross Miller, Sarp Oral, Oscar Hernandez,

Sergei Bastrakov, Michael Bussmann, Alexander Debus, Klaus Steiniger, Jan Stephan, Rene Widera

Spencer H. Bryngelson, Henry Le Berre, Anand Radhakrishnan, Jeffrey Young

Florina Ciorba, Osman Simsek

Sebastian Keller, Jean-Guillaume Piccinali

Christian Trott

David Hardy

Sunita Chandrasekaran

M. A. Clark, Filippo Spiga, Jeff Hammond, John E. Stone

For more details: https://arxiv.org/abs/2209.09731

Questions?

Acknowledgement: This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.