The Hidden Mysteries of Proxy Applications

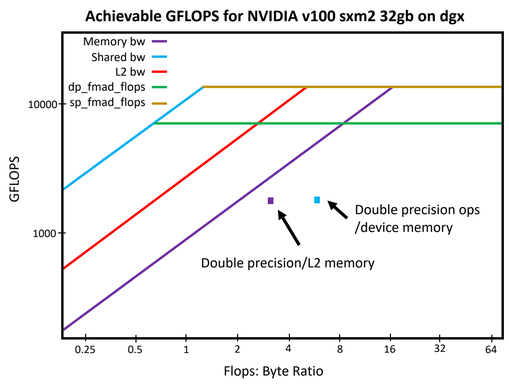
PI: David Richards (LLNL)
Team: Omar Aaziz (SNL), Xavier Andrade (LLNL), Jeanine Cook (SNL), Hal Finkel (ANL), Brian Homerding (ANL), Christoph Junghans (LANL), Nevin Liber (ANL), Peter McCorquodale (LBNL), Tiffany Mintz (ORNL), Shirley Moore (ORNL), Robert Pavel (LANL), Vinay Ramakrishnaiah (LANL), Thomas Uram (ANL), Courtenay Vaughan (SNL), Greg Watson (ORNL)

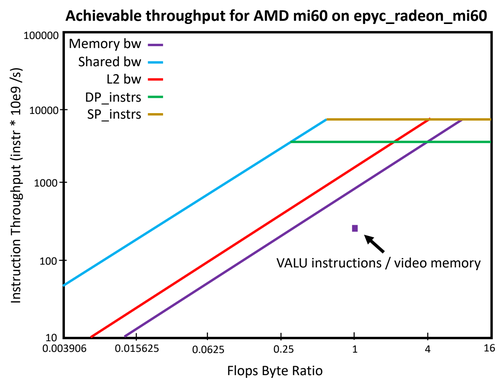
Mission

To curate a suite of proxy applications that are representative of the intended characteristics of their respective parent applications and are easy to obtain and use. Characteristics include hardware bottlenecks (e.g., memory, computation, communication) and programming models.

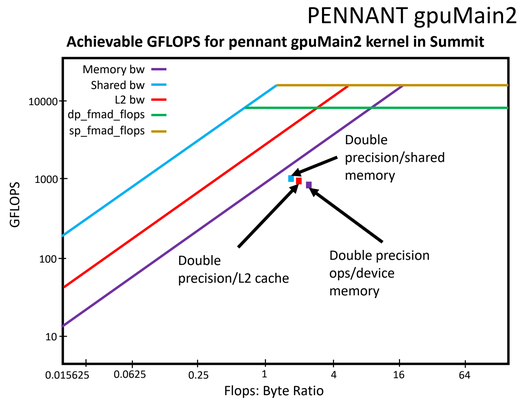
GPU Performance

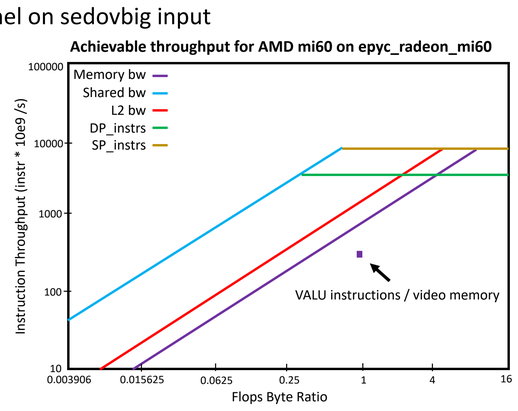
SW4lite curvilinear kernel on gaussianHill input





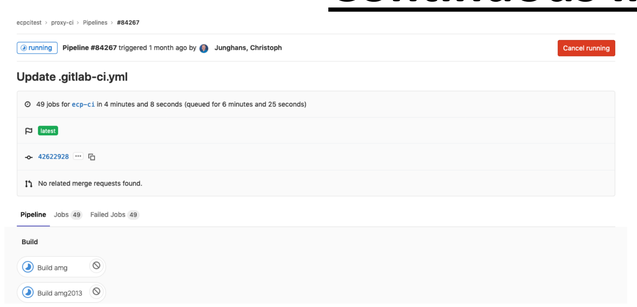
PENNANT gpuMain2 kernel on sedovbig input

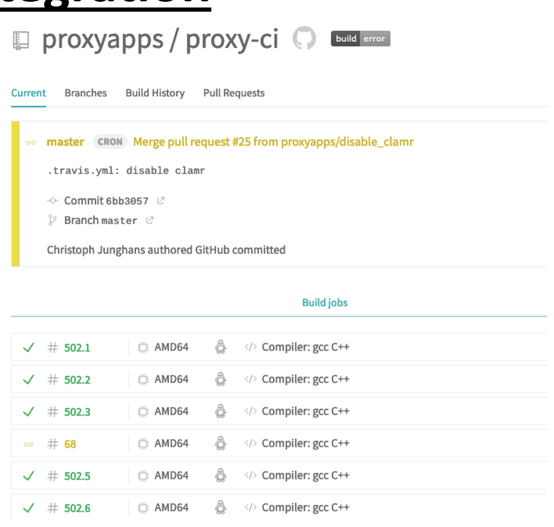




- Memory and cache bandwidths measured using microbenchmarks
  - V100: 829 GB/s achieved HBM BW; MI-60: 806 GB/s achieved HBM BW
- HIP versions of SW4lite and PENNANT provided by AMD
- Arithmetic intensity measured using nvprof on V100 and rocprof on MI-60
- Proxy app kernels 20-30% more slowly on MI-60 consistent with lower achieved HBM BW.

Continuous Integration





CI on ECP CI @ORNL:
Builds 49 apps using Spack on ORNL's Ascent machine

- Power9 with GPUs
- XL compiler chain

Upsides:

- Same node-level architecture and toolchain as Sierra/Summit

Downsides and Pain Points:

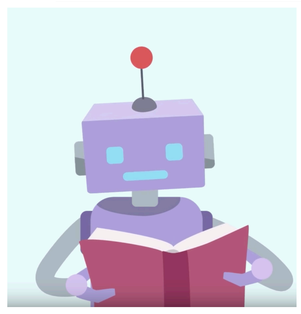
- Debugging over CI is hard – no interactive login
- Need to regularly update CI and Spack configuration to account for system updates
- CI runs as user: User configuration (e.g. user level Spack installation and environment) can "break" CI

CI on Travis: Every night we build 48 proxy apps using the latest Spack in Ubuntu and Fedora containers

- Upsides: free, in the open, early warnings on Spack and spackage breakages
- Downsides: only 2 cores, no MPI, no GPUs, no testing against TOSS etc.

→ Use ECP-CI when more than a smoke test is needed

Coming Soon: Machine Learning Proxy Suite

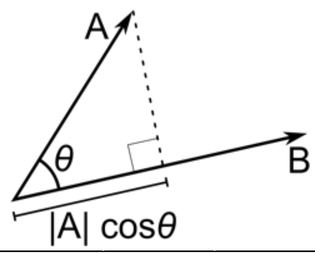


The Proxy App Project is currently collecting candidate proxy apps and use cases for a new proxy app suite focused on applications of machine learning to scientific applications.

Please send your suggestions to ECP-PROXY-APPS@listserv.llnl.gov

Cosine Similarity

- Quantifies the similarity of two applications' signatures
- Identifies gaps and artifacts in a proxy set representation of a workload
- Defines a minimal set of applications that span the observable underlying behaviors (computation, memory, and communication)



- Based on the inner product in vector spaces of 2 or more dimensions
- Uses cosine of θ as an angular distance metric
- Quantifies distance between A and B independent of their magnitude

Application similarity θ in degrees

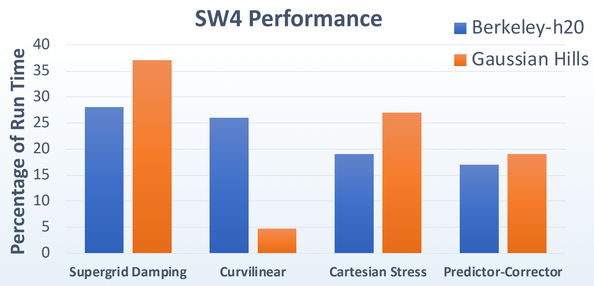
	ExaMiniMD	LAMMPS	MiniQMC	QMCPack	sw4lite	sw4	SWFFT	HACC	pennant	snap
ExaMiniMD	0.00	10.24	84.61	83.55	61.94	64.17	86.71	85.58	75.88	44.50
LAMMPS	10.24	0.00	75.12	73.95	53.63	56.50	79.66	78.51	70.97	34.97
MiniQMC	84.61	75.12	0.00	5.97	42.91	47.75	51.57	51.28	66.16	43.41
QMCPack	83.55	73.95	5.97	0.00	37.71	42.28	45.85	45.52	60.31	40.89
sw4lite	61.94	53.63	42.91	37.71	0.00	6.47	27.99	26.86	30.17	24.55
sw4	64.17	56.50	47.75	42.28	6.47	0.00	23.59	22.42	23.83	29.89
SWFFT	86.71	79.66	51.57	45.85	27.99	23.59	0.00	1.22	18.65	51.79
HACC	85.58	78.51	51.28	45.52	26.86	22.42	1.22	0.00	18.14	50.70
pennant	75.88	70.97	66.16	60.31	30.17	23.83	18.65	18.14	0.00	51.63
snap	44.50	34.97	43.41	40.89	24.55	29.89	51.79	50.70	51.63	0.00

Problem Selection Matters

SW4/SW4lite

Berkeley-h20: Production topography

Gaussian Hills: Synthetic topography



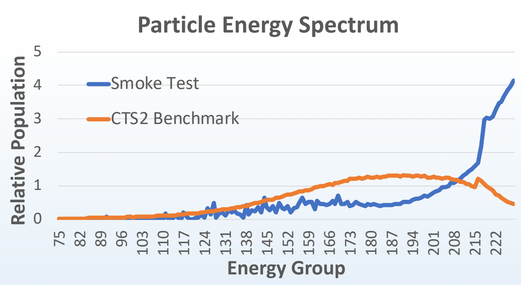
Gaussian Hills spends much less time in Curvilinear vs. Berkeley-h20 (real problem).

SW4lite runs only Gaussian Hills.
Corrections must be applied to SW4lite results to match the performance profile of SW4.

Quicksilver

CTS2 Benchmark: CTS2 problem

Smoke Test: Default problem



Smoke test overpopulates high energies compared to intended benchmark.

Energy spectrum determines memory access pattern for cross section lookups.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. SAND2020-0944 D