OOKAMI PROJECT APPLICATION

Date: **22/04/2021**

Project Title: Testing and Porting ALMA and CGYRO to A64FX architecture

Usage:

☒ Testbed

☐ Production

Principal Investigator: Federico D. Halpern

University/Company/Institute: General Atomics

Mailing address including country: 3550 General Atomics Ct, San Diego, CA 92121

Phone number: 858-455-2847

Email: halpernf@fusion.gat.com

Names & Email of initial project users:

Dr Tess Bernard, bernardt@fusion.gat.com
Dr Alessandro Marinoni, marinoni@mit.edu

Usage Description:
We would like to request a 15k node-hour resource at the Ookomi cluster for testing, porting, and scaling ALMA and CGYRO into the new A64FX architecture.

ALMA is a finite difference code based on a new anti-symmetric formulation of the classic models of fluid and plasma dynamics [1,2]. The formulation of the models is such that the force operator is completely anti-symmetric, which gives exact energy conservation, a numerical stability theorem, and retains the self-adjoint property of the ideal MHD force operator. ALMA is capable of carrying out simulations in different application domains, including aerodynamics (Navier-Stokes model), magnetized target implosions in inertial confinement fusion (Hall-MHD), and tokamak turbulence (drift-Braginskii model).

CGYRO [3] is a state-of-the-art 5D kinetic turbulence code based on the delta-f formulation of gyrokinetics (GK). Compared to typical GK codes, CGYRO is geared towards studying turbulence in plasmas with strong rotation, resolving multi-scale turbulence (encompassing both ion and electron spatial scales), and unravelling the effect of different isotopes on transport [4]. CGYRO is also being used to study negative triangularity tokamaks as an accelerated path to fusion energy [5]. CGYRO uses spectral representation in 2 spatial coordinates, uses finite differences in the spatial coordinate parallel to the magnetic field, and is semi-spectral in the two energy coordinates.

Being both applications memory-bound (in view of the so-called “roofline model”), they are likely to extract a large benefit from using a high memory bandwidth architecture.

CGYRO and ALMA both already scale up to thousands of nodes in modern clusters such as NSF's Frontera (Intel Xeon) and DOE's Theta (Intel KNL) and Summit (Nvidia V100). They have been subject of intense optimization to perform in these platforms. We welcome the opportunity to use Ookomi as a testbed. The amount of resources requested corresponds to a factor of 1/5 of the resources that were needed to optimize and scale the two codes at Frontera up to 1000 nodes.

Computational Resources:

Total node hours per year: 15k

Size (nodes) and duration (hours) for a typical batch job: Short jobs (<30mins) ranging from 1 node to all nodes available.

Disk space (home, project, scratch): There will be no long-term storage of simulation data (porting/optimizing/scaling project). The default home disk space space allocation, plus access to high-speed scratch space would suffice.

Personnel Resources:

The three participants will partake in carrying out benchmarks and optimization, however with the PI performing a larger fraction of the work.

Required software:

Modern C/C++/Fortran compilers, git, hdf5, netcdf, fftw, blas, lapack

If your research is supported by US federal agencies:

Agency: US DOE, Office of Science, Office of Fusion Energy Science

Grant number(s): DE-FG02-95ER54309

References:

[1] F.D. Halpern and R.E. Waltz, Physics of Plasmas 25 (6), 060703

[2] F.D. Halpern, Physics of Plasmas 27 (4), 042303

[3] J Candy, EA Belli, RV Bravenec, Journal of Computational Physics 324, 73-93

[4] EA Belli, J Candy, RE Waltz, Physical Review Letters 125 (1), 015001

[5] https://frontera-portal.tacc.utexas.edu/media/filer_public/09/58/09580be6-2cba-465c-b710-3794111cedd8/2021_lrac_fact_sheet.pdf

Production projects:

Production projects should provide an additional 1-2 pages of documentation about how (a) the code has been tuned to perform well on A64FX (ideally including benchmark data

comparing performance with other architectures such as x86 or GPUs)
(b) it can make effective use of the key A64FX architectural features (notably SVE, the high-bandwidth memory, and NUMA characteristics)
(c) it can accomplish the scientific objectives within the available 32 Gbyte memory per node