Workshop: The Taming of the Wolf - or how to use the Ookami Cray Apollo 80 system and its Fujitsu A64FX processors

EVA SIEGMANN, Stony Brook University, USA
DAVID CARLSON, Stony Brook University, USA
TONY CURTIS, Stony Brook University, USA
ROBERT HARRISON, Stony Brook University, USA
NIKOLAY SIMAKOV, University at Buffalo, USA
JOSEPH WHITE, University at Buffalo, USA
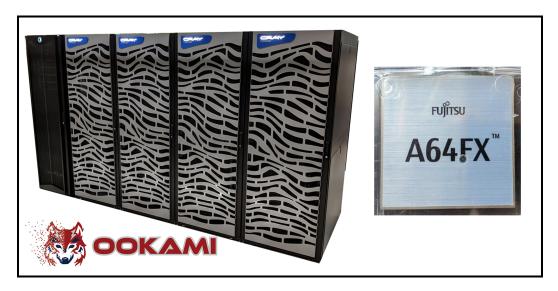MATTHEW D. JONES, University at Buffalo, USA

Fig. 1. The Ookami Cray Apollo80 system and the Fujitsu A64FX processor.

CCS Concepts: • **Applied computing** → *Education*.

Additional Key Words and Phrases: A64FX, high-performance computing, computer systems, exascale

1 INTRODUCTION

The team running the Ookami HPE/Cray Apollo 80 system at Stony Brook University and from the University at Buffalo is proposing to hold a full-day 6 hour workshop at the PEARC'23 conference. The workshop will focus on the characteristics of the Ookami testbed and the specific architecture of the Fujitsu A64FX FX700 processor, developed by RIKEN and Fujitsu for the Japanese path to exascale computing. The processor is also deployed in the until June 2022 fastest computer in the world, Fugaku. "Ookami" is a Japanese word for wolf. This is an homage to the origin of the processors in Ookami, our research relationship with RIKEN[1], to Stony Brook's seawolf[2] mascot, and inspired the title of this workshop.

Ookami first became available to researchers in January 2021, and is open to eligible researchers *worldwide.* Since October 2022, Ookami has also been an official ACCESS[3] resource provider, as well as a Stony Brook resource. The Ookami scientific-computing and user-support team employs on a multi-modal outreach approach. This includes a closely-monitored slack channel, twice-weekly office hours in which anyone can address any related issues, a ticketing system (both via Stony Brook and ACCESS), and by directly funding graduate students and working with them on science codes used by their research groups. The students are encouraged to participate in slack, office hours and to interact at all times with the expert Ookami staff; and indeed, to become the local experts.

Our experience during these 2.5 years since early 2021 has shown that many users, regardless of their seniority or experience, need a substantial amount of support or advice, mainly because of the novel architecture of the A64FX processor, and unfamiliar toolchains. The key features of A64FX are the SVE capability (scalable vector extension), 32GB of high-bandwidth memory, and its low power consumption. Many recipes that you might expect to work on standard x86 systems do not produce optimal results on this architecture, and achieving good performance often requires a more sophisticated and nuanced methodology. Moreover, many of these issues and solutions are relevant to other important architectures including the now-available Intel Sapphire Rapids+HBM, low-power but high-performance ARM platforms emerging from both exascale systems and cloud hyperscaler, and RISC-V architectures.

The focus of this workshop is thus to enable people to get a closer look at this interesting architecture, understand the challenges and opportunities it presents, become familiar with the software toolchains and performance analysis tools, and through hands-on access with experts in the room find ways to tackle and take advantage of them.

2 STRUCTURE AND FORMAT

The workshop will consist of the full-day format (6 hours) and various presentations about the system, its toolchains, and the architecture, practical hands-on examples, and an interactive discussion. Participants will receive guest accounts and can explore Ookami themselves. This assumes that attendees will bring their own laptops. We will make sure to highlight this fact to everybody registering for the workshop beforehand, so they are prepared. We will prepare use cases that highlight the specifics and advantages of the architecture, including its associated challenges and unique potential. Attendees will also have the opportunity to test their own applications and mini-apps on Ookami and get feedback.

---

[1]https://www.riken.jp/en/

[2]https://en.wikipedia.org/wiki/Anarhichadidae

[3]https://access-ci.org/

The team from The Center for Computational Research[4] at the University at Buffalo, will also demonstrate their XDMoD web-service on Ookami[5]. XDMoD not only contains overall metrics (utilization by users, groups, etc.) but also allows users to closely monitor their jobs in terms of power consumption, number of SVE instructions, and overall performance (for example CPU utilization, memory usage, and i/o).

Several instructors from the Ookami team will be on hand to help users troubleshoot any issues. The instructors will be the authors of this proposal plus several experienced graduate students. We are also in discussion with the vendors and are expecting to get contributions from ARM and CPE.

In the past, we did offer short one hour webinars for researchers interested in learning more about Ookami, as well as hands-on workshops which were limited to the Ookami user community. The Ookami team also partnered with ARM and offered participants of their hackathons guest accounts for the duration of the event. A workshop like the proposed one, targeting a broad audience, with a full day of presentations and hands-on exercises has never been offered so far.

3 TARGET AUDIENCE

In this workshop, we will welcome a broad community of researchers, professional cyberinfrastructure staff, and, in general, anybody interested in the specific A64FX architecture and our testbed. We strongly encourage students to attend this workshop. Participants are expected to be familiar with standard Linux commands and the basic use of standard compilers (e.g., GCC).

Interested participants also will have the opportunity to get allocations on the system after the conference. We will show them how to create and submit an allocation request via ACCESS.

4 EXPECTED ATTENDANCE

The workshop will be advertised via the outreach mechanisms of both Ookami and ACCESS. Although it is hard to predict the number of attendees we are prepared to work with around 30 people. The bottleneck will be the support during the hands-on exercises. More people can attend, and we will have guest accounts ready for them, but personal support cannot be guaranteed for more than 30 participants.

5 RELEVANCE TO PEARC'23

One facet of our mission within the Ookami project is to make cutting-edge technology available to a broad community in the field of research computing. For example, as well as workshops and webinars for the Stony Brook and XSEDE/ACCESS research community, we also hold HPC workshops for high school students that leverage the Ookami system. The PEARC conference gives us a perfect opportunity to reach a broad and diverse group of researchers. Since there are no requirements for this workshop (other than knowing basic Linux and standard compiler usage), interested researchers will not be deterred by this new architecture. Although this architecture can introduce unexpected complexities, it can also produce performance gains. By proposing this workshop, we want to encourage a new audience to take advantage of this innovative ecosystem, of our expertise gained over the last 2.5 years; and then use it for their own research.

ACKNOWLEDGMENTS

---

[4]https://www.buffalo.edu/ccr.html

[5]https://ookami.ccr.xdmod.org/

Eva Siegmann, David Carlson, Tony Curtis, Robert Harrison, Nikolay Simakov, Joseph White, and Matthew D. Jones

innovative high-performance Ookami computing system, which was made possible by a $5M National Science Foundation grant[6](#1927880).