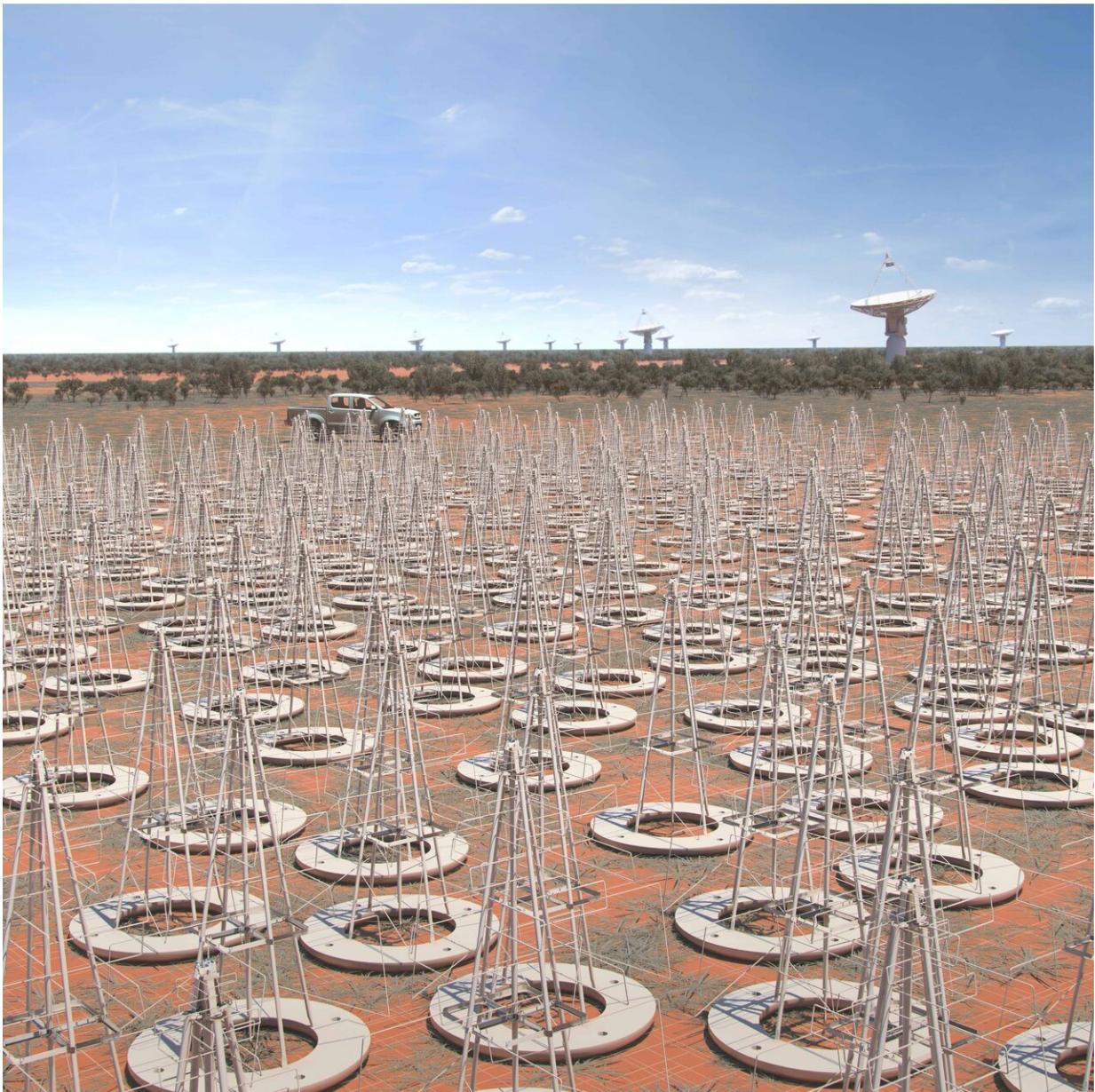


With ADIOS, Summit processes celestial data at scale of massive future telescope

December 19 2019, by Scott S Jones



An artist rendering of the SKA's low-frequency, cone-shaped antennas in Western Australia. Credit: SKA Project Office

For nearly three decades, scientists and engineers across the globe have worked on the Square Kilometre Array (SKA), a project focused on designing and building the world's largest radio telescope. Although the SKA will collect enormous amounts of precise astronomical data in record time, scientific breakthroughs will only be possible with systems able to efficiently process that data.

Because construction of the SKA is not scheduled to begin until 2021, researchers cannot collect enough [observational data](#) to practice analyzing the huge quantities experts anticipate the telescope will produce. Instead, a team from the International Centre for Radio Astronomy Research (ICRAR) in Australia, the Department of Energy's (DOE's) Oak Ridge National Laboratory (ORNL) in the United States, and the Shanghai Astronomical Observatory (SHAO) in China recently used Summit, the world's most powerful supercomputer, to simulate the SKA's expected output. Summit is located at the Oak Ridge Leadership Computing Facility, a DOE Office of Science User Facility at ORNL.

"The Summit supercomputer provided a unique opportunity to test a simple SKA dataflow at the scale we are expecting from the telescope array," said Andreas Wicenec, director of Data Intensive Astronomy at ICRAR.

To process the simulated data, the team relied on the ORNL-developed Adaptable IO System (ADIOS), an open-source input/output (I/O) framework led by ORNL's Scott Klasky, who also leads the laboratory's scientific data group. ADIOS is designed to speed up simulations by increasing the efficiency of I/O operations and to facilitate data transfers

between high-performance computing systems and other facilities, which would otherwise be a complex and time-consuming task.

The SKA simulation on Summit marks the first time radio astronomy data have been processed at such a large scale and proves that scientists have the expertise, software tools, and computing resources that will be necessary to process and understand real data from the SKA.

"The scientific data group is dedicated to researching next-generation technology that can be developed and deployed for the most scientifically demanding applications on the world's fastest computers," Klasky said. "I am proud of all the hard work the ADIOS team and the SKA scientists have done with ICRAR, ORNL, and SHAO."

Using two types of radio receivers, the telescope will detect radio light waves emanating from galaxies, the surroundings of black holes, and other objects of interest in outer space to help astronomers answer fundamental questions about the universe. Studying these weak, elusive waves requires an army of antennas.

The first phase of the SKA will have more than 130,000 low-frequency, cone-shaped antennas located in Western Australia and about 200 higher frequency, dish-shaped antennas located in South Africa. The international project team will eventually manage close to a million antennas to conduct unprecedented studies of astronomical phenomena.

To emulate the Western Australian portion of the SKA, the researchers ran two models on Summit—one of the antenna array and one of the early universe—through a software simulator designed by scientists from the University of Oxford that mimics the SKA's data collection. The simulations generated 2.6 petabytes of data at 247 gigabytes per second.

"Generating such a vast amount of data with the antenna array simulator

requires a lot of power and thousands of graphics processing units to work properly," said ORNL software engineer Ruonan Wang. "Summit is probably the only computer in the world that can do this."

Although the simulator typically runs on a single computer, the team used a specialized workflow management tool Wang helped ICRAR develop called the Data Activated Flow Graph Engine (DALiuGE) to efficiently scale the modeling capability up to 4,560 compute nodes on Summit. DALiuGE has built-in fault tolerance, ensuring that minor errors do not impede the workflow.

"The problem with traditional resources is that one problem can make the entire job fall apart," Wang said. Wang earned his doctorate degree at the University of Western Australia, which manages ICRAR along with Curtin University.

The intense influx of data from the array simulations resulted in a performance bottleneck, which the team solved by reducing, processing, and storing the data using ADIOS. Researchers usually plug ADIOS straight into the I/O subsystem of a given application, but the simulator's unusually complicated software meant the team had to customize a plug-in module to make the two resources compatible.

"This was far more complex than a normal application," Wang said.

Wang began working on ADIOS1, the first iteration of the tool, 6 years ago during his time at ICRAR. Now, he serves as one of the main developers of the latest version, ADIOS2. His team aims to position ADIOS as a superior storage resource for the next generation of astronomy data and the default I/O solution for future telescopes beyond even the SKA's gargantuan scope.

"The faster we can process data, the better we can understand the

universe," he said.

Provided by Oak Ridge National Laboratory

Citation: With ADIOS, Summit processes celestial data at scale of massive future telescope (2019, December 19) retrieved 2 May 2024 from <https://techxplore.com/news/2019-12-adios-summit-celestial-scale-massive.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.