MGI37-22   Room:203   Time:May 28 12:15-12:30

Development of High-Performance Data Analysis Library Based on Vector Operation and Its Application to ALMA

NAKAZATO, Takeshi[1*] ; SUGIMOTO, Kanako[1] ; KAWASAKI, Wataru[1] ; KAWAKAMI, Shinnosuke[1] ;
KUNIYOSHI, Masaya[1] ; NAKAMURA, Kohji[1] ; KOSUGI, George[1] ; MAEKAWA, Jun[1]

[1]National Astronomial Observatory of Japan

Atacama Large Millimeter/submillimeter Array (ALMA) is a radio-astronomical interferometer operated by an international partnership of East Asia including Japan, Europe, and North America. With its state-of-the-art capability, ALMA produces a lot of important scientific results for various fields of astronomy that cover from planetary science to early universe. These results are sustained by large amount of data produced by many telescope elements of ALMA. The data size will become 200TB per year in a full operation phase. To analyze such a large data efficiently, it is extremely important to improve data analysis software to take advantage of full capability of given hardware resources as well as to build up powerful hardware resource. In particular, it is crucial that such data analysis software implements each function based on parallel processing since it is one of the key means to improve performance of the data analysis. Among others, vector operation is a promising for high-performance computing because 1) it is universal that is widely available from personal laptop computer to computer cluster environment, and 2) it is relatively handy to implement using auto-vectorization of compilers if the source code is carefully written so that it suits to vector operation. In this context, we have been working on improving the performance of data analysis software for ALMA called Common Astronomy Software Applications (CASA). Our plan is to develop general purpose high-performance data analysis library and utilize it from CASA to improve its performance. We have been developing this library for three years. The library is named Sakura. Noticeable feature of Sakura is to improve the performance by using vector operation as much as possible. In optimizing the code based on vector operation, it is important to take care of an antinomy between optimization and versatility. In general, it is better to use newer instruction set for more efficient processing. However, the library will not work on older CPU if too new instruction set is used. In Sakura, we provide an easy interface for compiling it with various types of instruction set to overcome this antinomy. For specific purpose computer and software, one can optimize Sakura by compiling it against native instruction set for the computer. For general purpose, we can generate several versions of Sakura library that are optimized to various types of instruction set including conservative but inefficient optimization, and select which library is loaded depending on runtime environment. In CASA, we will take the latter option. Based on Sakura, we have developed a prototype CASA command for reduction of single-dish radio telescope data. Thanks to vector operation in addition to improve workflow and reduce unnecessary data input/output from the disk, the prototype command is 20 times faster than existing counterpart in CASA. With this result, Sakura based development has been approved by CASA team. Sakura based commands will be available in CASA 4.4 that is planned to release in April of 2015. Here, we will summarize an implementation of Sakura, and a performance of Sakura based command compared with existing CASA commands. We will also describe a status of Sakura based development for CASA 4.4. Furthermore, we will mention about our future plan for improving CASA performance by extending Sakura to be able to handle interferometry data, which usually deals with complex visibility data.

Keywords: ALMA, Data Analysis, Vector Operation