Article can be accessed online at http://www.publishingindia.com

# The Square Kilometer Array - Some Notes Regarding the Largest Telescope Being Planned and Why it is the Ultimate Big-Data Challenge?

**Avadhesh Gupta*, Rajanish Jain**, Gend Lal Prajapati*****

**Abstract**

"The Square Kilometer Array (SKA)" is the largest radio telescope being developed in South Africa/Australia and planned to be operational by 2024. SKA will consist of millions of coherently connected antennas spread over an area of about 3000 km in extent with a total collecting area of approximately one square km. SKA will collect and decipher a deluge of radio signals from deep outer space to help us better understand the evolution of the universe and the nature of the matter. In this paper we provide an introduction to the SKA science project, its basic objectives, the engineering challenges it entails and why SKA is being touted as an ultimate big data challenge.

## Introduction

Advances in science and technology in the past decades have brought the international scientific community to the verge of charting a complete history of the universe. A number of scientific projects have been commissioned to study various aspects of the universe. Some of the well-known projects in this stellar list are Large Hadron Collider (LHC), Cosmic Background Explorer (COBE), Hubble Telescope, Spacecrafts like Cassini, Voyager, Galileo, Search for extraterrestrial intelligence (SETI), Radio-telescopes like LOFAR and MeerKAT etc. "The Square Kilometer Array (SKA)" is another such science project which will collect radio signals from deep outer space and help scientists understand a number of issues regarding the evolution of the universe and the nature of the matter. SKA will be an ultra-sensitive radio telescope and one of the most ambitious science projects ever undertaken by mankind. The project will involve 10 nations (headquarter in UK) as well as numerous universities and industries. With a budget of 1.5 billion euros, the construction of SKA is scheduled to begin in 2016 for initial observations by2019 and for full operations by 2024. When operational, SKA will be 50 times more sensitive than any other radio instrument. It will survey the sky more than ten thousand times faster than ever before and will be powerful enough to collect faint radio waves from objects billions of light-years away from earth. SKA will consist of millions of coherently connected antennas over an area of 3000 km scattered across South Africa and Australia. The collecting areas of all antennas making up SKA will add up to a square kilometer and hence the project is named "Square Kilometer Array". SKA will scan the sky and generate approximately 14 exabytes of data every day. SKA will be sensitive enough to also pick up radio waves emanating from the Big-Bang event which happened 13 billion years back. This data will help scientists understand a number of questions regarding the evolution of the universe, the nature and the formation of matter, the nature of space-time, the presence of life elsewhere in the universe etc.

The 14 exabyte of data will be cleaned, analyzed and whittled down to 1 petabyte of usable data. Yet even at this rate it will add up quickly and become an exabyte of data within three years and three exabytes in a decade. The volume of the SKA data will be unprecedented and will require the development of altogether new algorithms,

* Professor, Department of Information Technology,Institute of Management Studies, Ghaizabad, India. Email: avadhesh.gupta@imsgzb.com

** Assistant Professor, Department of Information Technology,Institute of Management Studies, Ghaizabad, India. Email: rajnesh.jain@imsgzb.com

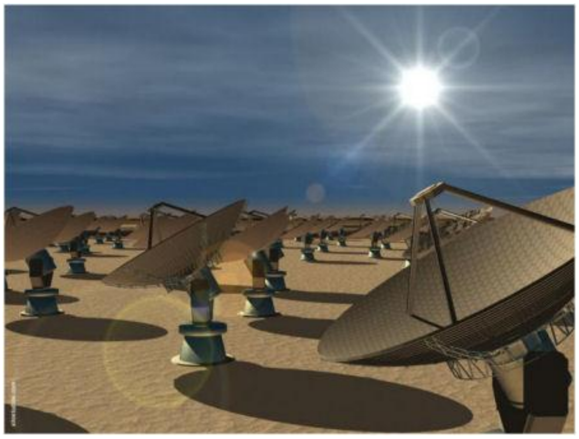
*** Reader, Department of Computer Engineering, Institute of Engineering & Technology, Devi Ahilya University, Indore, India. Email: glprajapati1@gmail.com

technologies in various areas of computer science e.g., data-management, storage, communication, cognitive-computing etc. It is because of these reasons that SKA is being talked about as an ultimate big-data challenge. Below are outlined some facts which indicate the sheer magnitude of the resources SKA will employ and the data it will analyze (www.skatelescope.org).

- The data collected by SKA in a single day will take two million years to play on an IPod.
- SKA central computer will have a processing power of about one hundred million PCs.
- Each day, the dishes of SKA will produce 10 times more data as produced by Internet in a day.
- SKA will use enough optical fiber to wrap around earth twice.

Contribution: In this paper we provide an introduction of the SKA project. We present the goals of the SKA project and also summarize the engineering and technical challenges the project entails. The goal of this paper is to present these issues in a single document without any technical details so that a layman technician (a basic degree in CS and a basic course in Physics) can understand the enormity of the project. To the best of our knowledge, this is the first effort in this direction.

**Figure 1: Artist's Impression of SKA**



Organization: Rest of this paper is organized as follows. The next section mentions the related work and points the readers Figure 1: Artist's Impression of SKA to more advanced resources on SKA. The third section summarizes the four basic goals of the SKA project namely testing out Einstein's theory of relativity in extreme environments, the evolution of galaxies, dark matter/dark energy, probing the Big-Bang event and the search of life elsewhere in the universe. The fourth section summarizes the engineering and technical challenges in the development of SKA. The last section finally concludes this paper.

## Related Work

SKA homepage (www.skatelescope.org ) is the most informative source for getting to know SKA project in detail. It chronicles the project timeline and provides the details regarding member countries, universities and institutes of SKA organization. It describes the basic goals, the technical design and features of the telescope, industrial engagements, technical papers as well as various news and media articles. Most of the notes in this article are picked from this website. Dewdney *et al.*(2009) and Cordes *et al.* (2000) are two good references for knowing the mathematical details behind SKA system and operations.

A number of scientific projects have investigated different aspects of universe in the past few decades using different methodologies. The front-runner among these is the "Large Hadron Collider (LHC)" the highest energy particle collider ever made. It allows physicists to test various theories of particle physics and high-energy physics. A number of issues are currently being explored through LHC like the presence of extra dimensions, inter-relation between quantum mechanics and general relativity, the nature of dark matter etc. Cosmic Background Explorer (COBE) is a satellite the chief aim of which is to measure and investigate cosmic microwave background radiation. These radiations were emitted as a result of Big-Bang event. The measurements taken by the COBE have provided a strong evidence in favour of Big-Bang theory of universe. The COBE project is also seen as the starting-point of cosmology to be considered a precision science.

NASA has launched multiple spacecrafts to study solar systems and its planets, the three of the most important beingGalileo, Cassini and Voyager. Galileowas launched by NASA in 1989 to study Jupiter and its moons.

It reached Jupiter in 1995 and was terminated in 2003. Cassini was launched in 1997 to study Saturn and its moons. It reached Saturn in 2005 and its mission is still undergoing. Voyager spacecrafts, Voyager-1 and Voyager-2, were launched in 1970s to study Jupiter and Saturn. Voyager-1 has currently crossed the solar-system and has entered the inter-stellar space and hence is the furthest man-made object from earth. Voyager-2 is expected to cross the solar-system in 2016. Taken together these space-crafts have studied the geology and mineralogy of the outer planets and helped scientists understand various issues regarding the formation of solar-system.

Hubble's telescope is a space telescope currently in earth's orbit. In past two decades, it has taken a series of spectacular high-resolution images. Hubble has recorded many detailed visible-light images ever, allowing a deep view into space and time. These observations played a crucial role in multiple astrophysics breakthroughs, such as accurately determining the rate of expansion of the universe. SETI project has scanned the electromagnetic radiation from outer space to look for the signs of intelligent life-forms elsewhere.

A number of radio telescopes have also been designed. LOFAR is a radio telescope currently in operation in Netherlands and consists of 25000 small antennas. LOFAR is currently the largest connected as well as the most sensitive radio telescope at low observing frequencies and will be surpassed by SKA in both these aspects when it is completed. The mission of LOFAR is to survey the universe at radio frequencies (10-240 MHz) with greater resolution and sensitivity than previous telescopes like Giant Meter- wave Radio Telescope (GMRT) in India or Very Large Array (VLA) in USA. MeerKATis a radio telescope currently being constructed in South Africa and will consist of 64 antennas , each one 13.5 m in diameter. MeerKAT will be completed in 2018 and will be the most sensitive radio telescope built until SKA's construction is completed in 2024. The missions of MeerKAT include the investigation of cosmic magnetism, galactic evolution, large-scale structure of the cosmos etc.

Similar to these projects, SKA will also investigate various aspects of the universe. It will collect huge amount of radio waves from deep space and analyze these to test various theories and predictions. Secondly the amount of data collected and the subsequent effort in managing and processing this data will be much more significant vis-a-vis past projects and hence requiring a development of altogether new paradigms and fields in Computer Science. The DOME project partnered by IBM and Netherland institute of Radio Astronomy is currently investigating these computing challenges and building a state of the art data management system which will meet SKA requirements.

## Goals of SKA Project

A number of project proposals have been commissioned to be investigated by SKA. In this section we outline four such proposals. For each of these we provide a little background and how SKA plans to carry out its investigation.

### Testing General Theory of Relativity In

Extreme Environments Issac Newton first described the nature of motion of the planets in the solar-system in the seventeenth century. This theory of gravitation described what is the magnitude of attractive force between two celestial bodies. This theory proved to be a phenomenal success in describing the motion of objects on earth and the motion of planets around the sun and the theory went unchallenged for around 300 years. However the major flaw in Newton's theory was that it could not describe how exactly a celestial body like sun transmits the gravitational force to another celestial body like earth (though it could describe the magnitude of the gravitational force). Newton admitted this discrepancy and left it as a question to be pursued by future physicists. Secondly in time the astronomers observed that the orbit of Mercury deviated a little from what is predicted by Newton's gravitational theory.

Both these questions were answered by another giant of science - Albert Einstein in the first decade of the twentieth century. Einstein proposed that the notion of spaceand the notion of time, together called the notion of space-time, are not static but are modified by the presence of mass. A celestial body like sun changes the rate at which the space- time flows in its vicinity. In other words the time (as well as space) flows at different rates at different places in the universe depending on how much matter is present. This curved space-time is what causes another body like earth to go into orbit around the sun. Einstein derived the necessary equations to model the interplay between space-time and the gravitational force and this was called relativity theory. By describing the relationship between space-time and gravity, Einstein's relativity theory replaced Newton's gravitational theory and fundamentally altered the physics world-view of the universe. The proposed theory was successful in explaining the motion of mercury around the sun and hence could do away with the two major flaws of Newton's theory of gravity. Since then Einstein's theory has been checked and verified using multiple methods.

SKA will check the relativity theory in extreme gravity environments in the vicinity of pulsars and black-holes. A star (like sun) in the late stages of its cycle contracts and first converts to a pulsar and then to a black-hole. An object of the size of earth when contracted to the size of a marble will become a black hole. Pulsars and black-holes are hence the densest objects in the universe and they hence significantly modify the nature of space-time in their vicinity. Einstein's relativity theory makes definitive predictions about how the space-time will behave in the vicinity of such extremely dense objects. SKA will be sensitive enough to catch radio waves emanating from distant pulsars and black- holes and hence verify whether the observations tally with the predictions made by the relativity theory. Such tests when completed will be the most stringent tests carried out to verify the relativity theory.

## Evolution of Galaxies, Dark Energy and the Dark Matter

How galaxies form and evolve over time is not well understood. This will be a key question SKA will investigate. SKA will be sensitive enough to pick up signals from galaxies being formed in the deep outer space. SKA will then scan these signals to investigate various aspects regarding the formation of galaxies. Hydrogen is the most abundant element in the universe. From the signals received from newly formed/forming galaxies, SKA will decipher the distribution of Hydrogen. SKA will collect signals from billions of such galaxies and by comparing the distribution of Hydrogen in these billions of galaxies will be able to gain insights regarding how galaxies evolve over time.

Our universe is expanding and the rate of the expansion is also increasing. The detection of this accelerated expansion won the Nobel Prize in Physics in 2011. Exactly why this accelerated expansion is taking place, is not clearly understood. Itis counter-intuitive in some sense as the mutual gravitational attraction among the galaxies should slow down the expansion. Scientists have proposed the existence of two mysterious entities *dark energy* and *dark matter* which form the bulk of the universe. Between two objects, these forces are repulsive and overcome the attractive gravitational force, thereby causing the accelerated expansions. Note that these forces become relevant only over large inter- galactic distances (of the order of millions of light-years) and not on the scale of small inter-planetary distances (which is of the order of few light-minutes). By studying the signals from a vast number of galaxies in various stages of its evolution, SKA may provide evidence for or against the presence of dark-energy and dark-matter.

## Probing the Cosmic Dawn

When the Big-Bang event happened, it sent out signals in the microwave region of the spectrum and these signals are present in all of space and all around us. These signals are named *Cosmic Microwave Background Radiations*. The COBE project took detailed measurements of these signals from various stars and galaxies in space. Two of the scientists working on COBE project received Nobel Prize in Physics for the year 2006. The COBE project helped scientists understand the universe within 300,000 years of the Big-Bang. The first stars and galaxies formed approximately at this time.

This period in the early universe (300,000 years after Big-Bang) is not studied in detail. It is hypothesized that at this point, the stars being formed were orders of magnitudes larger than our sun as well as very unstable and hence of much smaller lifetime vis-a-vis sun. The stars forming in this period were the first objects in the universe which produced light. This period has proved difficult to study as the light produced by these stars was very faint and as it travels towards earth, the most of it is absorbed by intervening matter. Such reduced light signals are hence not picked up by less sensitive instruments. It is expected that SKA will be sensitive enough to pick up these signals and hence allow the scientists to study this epoch in the history of universe in detail. This will provide much needed insights into the formation of first stars and galaxies.

## The Search for Life

SKA will look for any signs of life in the universe in a number of ways. Organic molecules like amino acids are the building blocks of life. These molecules have a unique spectral signature at specific frequencies. SKA will be able to identify if the radio signals from some celestial objects consist of these signatures. Presence of the signatures of such organic molecules will greatly enhance the possibility that such celestial objects might harbor life. Secondly, the earth lies in the *Goldilock zone* of the sun i.e., the zone that has all the conditions right for life (but not necessarily contains the life). A number of planets have been discovered which are in the *Goldilock zone* of their star. SKA will be able to analyze signals from billions of such planets and the ones which are currently being formed and are in the *Goldilock zone* of their star. By studying these planets in the various stages of their formation, the scientists may learn a lot about the life as a phenomenon. Thirdly, if there are any intelligent life in the universe, they may be generating various kinds of signals e.g. for communication. SETI project (www.seti.org) has looked for such signals but has found none. SKA will be much more sensitive vis-a-vis SETI and will hence greatly improve over the SETI capabilities.

## Engineering and Technical Challenges

In the previous section we summarized four key goals of SKA project. To meet these goals, SKA researchers are currently working on a number of computing challenges. These activities are being carried out within the ambit of the DOME project (www.dome-exascale.nl) with IBM and Netherlands institute of Radio Astronomy being the two principal partners in this endeavor. The systems developed by DOME project will be first tried out on the data collected by LOFAR and MeerKAT telescopes, the two much smaller telescopes vis-a-vis SKA. In this section we summarize some of these technical goals. Most of the notes in this section are picked up from the DOME project website(www.dome-exascale.nl).

### Algorithms and Machines

SKA project is so extreme that nobody has designed a data management system which handles anything like it or close to exascale level data. The system will require an optimization of multiple parameters like power envelope, performance, hardware cost, workload, CPU, memory size, accelerators, cooling systems etc. The system design will be so complex that it will be beyond what a team of few people can keep "in their head" and reason about. The challenge is to hence develop methodologies based on mathematical principles which address the mathematical equations governing the system and use mathematical optimizations to recommend directions, system design should address.

### Access Patterns

The daily exalevel SKA data will need to be stored at the lowest possible cost, archived as energy efficient as possible and will need to be easily accessible by various researchers world-wide. At this scale, storing and moving data will be very expensive. In computing, data is stored across multiple tiers, depending on how often is it needed and the cost of the storage medium. Magnetic Tape is cheap but very slow to access. Disk drives are expensive but faster to access. Access on memory chips is lightening quick but their cost is very expensive. With the help of various simulations, usage- investigations, modeling etc, Access Pattern technology will need to learn the usage-patterns in which SKA data is accessed. An example of a usage pattern may be selecting a slice of a data based on time, frequency or spatial dimensions. Access Pattern technology will then need to store the SKA data along various tiers (tapes, disks, chips and others) in an optimal fashion by exploiting the SKA data usage patterns. Access pattern technology will also need to proactively learn and adjust data storage and access strategies as user requirements/usage patterns change over time. Smart analysis, modeling, and prediction of these usage patterns hence will be crucial for optimal storage utilization (cost, performance etc.).

### Nano Photonics

The data collected by SKA's antennas will be shipped off to far-away data centers (hundreds or thousands of miles away) for storage and analytics. To transport an exascale data, a fiber-optic communication network will need to be built which is capable of moving data at 100 times the rate of today's Internet traffic. Secondly, the major bottleneck will come when this data reaches the data center where the data will be stored and analyzed. The computers transport data internally via electronic bits moving on copper wires. The current technology is not advanced enough to be able to quickly write so much data and hence with current technology, the data arrival rate will far outstrip the data writing rate on a server. Secondly the computers won't be able to transport so much data internally for processing and analysis at a quick enough rate. The DOME researchers are currently looking to build new systems which will alleviate these problems.

## Microservers

Server computers are expensive, use a lot of energy and are about the size of a pizza box. SKA will perform the first round of data filtering or analysis close to the antennas, or even on them. This will require very small, inexpensive and highly energy efficient servers to perform those processing jobs. Such servers are named "microservers" and will be about the size of a soap. The DOME researchers are currently working on building the microserver technology. Secondly many such microservers will be required to process the SKA data and these microservers will be densely packed on racks. To prevent such dense packages of electronics from overheating, a novel water cooling technique will need to be developed. The DOME researchers are working on such a technology whereby un-chilled water is directed through microscopic channels across the surfaces of chips in a server. Then the hot water is used for some other purpose. In the SKA scenario, it could be used to supply energy for sea water desalination in southern African and Australian deserts. Thirdly, computing systems are not typically designed to operate in the middle a desert (SKA antennas will be located in the Kalahari desert), where the temperature can reach as high as 50 degree celsius. The DOME researchers are also working on making the microservers "desert proof" to handle the extreme conditions.

## Accelerators

Supercomputers link together thousands of microprocessors so they behave like a single large machine. These supercomputers use brute force computation to get the job done. But the volume of SKA data will be so much that a brute force computation model will fail, even for the immensely powerful supercomputers. The DOME researchers are hence exploring the possibility of creating hybrid systems containing both traditional supercomputer elements and another kind of processor, the accelerator. An accelerator will handle a certain kind of computational task very well. An accelerator might be positioned close to where the SKA data is stored so they can filter the data and send only the useful bits to the main microprocessors for analysis. Another accelerator might be used to remove the distortion to radio waves caused by passing through the earth's atmosphere. These accelerators will need to be programmable i.e., engineers can use software to change how they operate and hence there is no need to install a special accelerator for each processing task.

## Compressive Sampling

One way to reduce the sheer volume of data in the SKA data management system will be to compress the data as it streams in. This approach could result in significant savings in energy use, storage and processing. The DOME researchers are developing specialized signal processing and machine learning algorithms for the capture, processing and analysis of radio astronomy data. The research will result in advancement of the techniques used for radio-astronomical data analysis, and pose new challenges to advance the development of machine learning, compressive sampling and other signal processing algorithms.

Each of these six research directions aims to solve a critical problem faced by the developers of the SKA system. On all these six fronts, the current technologies cannot be used and must be improved so that they can be applied to exascale level data. It is because of these challenges that SKA project is being talked about as the ultimate Big-data challenge. SKA plans to tackle the most ambitious of questions that we've never been able to answer about the universe, with technology that was never before available. As the DOME research continues, it is expected that it will also handle multiple challenges facing the science community and technology industry as we advance into the new era of computing.

# Conclusions

In this paper, we provided an introduction to the "The Square Kilometer Array" radio telescope project. We discussed the astronomical goals of the project and presented these goals in a language, a naive technician can intuitively understand. We finally discussed various engineering and technical challenges in realizing SKA system and why the SKA project is being talked of as the ultimate Big-data challenge?

# References

Cassini Solstice Mission. Retrieved from http://saturn.jpl.nasa.gov/.

Cosmic Background Explorer. Retrieved from http://lambda.gsfc.nasa.gov/product/cobe/.

Galileo Spacecraft. Retrieved from http://en.wikipedia.org/wiki/galileo spacecraft.

Giant Meterwave Radio Telescope.Retrieved from http://gmrt.ncra.tifr.res.in/.

Hubble Space Telescope. Retrieved from http://www.nasa.gov/mission pages/hubble/main/.

Large Hadron Collider. Retrieved from www.lhc.ac.uk.

Low Frequency Array. Retrieved from www.lofar.org.

MeerKAT Array. Retrieved From www.ska.ac.za/meerkat/.

Search for Extraterrestrial Intelligence. Retrieved from http://www.seti.org.

Tests of General Relativity Retrieved from http://en.wikipedia.org/wiki/tests of general relativity.

The DOME Project www.dome-exascale.nl.

The Square Kilometer Array. Retrieved fromhttp://www.skatelescope.org/.

Very Large Array. Retrieved from http://www.vla.nrao.edu/.

Voyager - The InterStellar Mission. Retrieved from http://voyager.jpl.nasa.gov/.

Cordes, J. (2000). The square kilometer array as a radio synoptic survey telescope: Widefield surveys for transient, pulsars and eti. Technical report, Cornell University.

Dewdney, P.E., Hall, P.J., Schilizzi & Lazzio, T.J.L.W. (2009).

The square kilometer array. *Proceedings of the IEEE*, *97*(8).