# DOME

## Towards the ASTRON & IBM Center for ExaScale Technology

P. Chris Broekema[†] broekema@astron.nl
Albert-Jan Boonstra[†] boonstra@astron.nl
Victoria Caparrós Cabezas[⋆] vca@zurich.ibm.com
Ton Engbersen[⋆] apj@zurich.ibm.com
Robert Haas[⋆] rha@zurich.ibm.com
Hanno Holties[†] holties@astron.nl
Jens Jelitto[⋆] jje@zurich.ibm.com
Ronald P. Luijten[⋆] lui@zurich.ibm.com
Peter Maat[†] maat@astron.nl
Rob V. van Nieuwpoort[†] nieuwpoort@astron.nl
Ronald Nijboer[†] rnijboer@astron.nl
John W. Romein[†] romein@astron.nl
Bert Jan Offrein[⋆] ofb@zurich.ibm.com

[†]Netherlands Institute for Radio Astronomy (ASTRON)
Postbus 2, 7990 AA, Dwingeloo, The Netherlands

[⋆]IBM Research GmbH — Zürich
Säumerstrasse 4, 8803 Rüschlikon, Switzerland

## ABSTRACT

The computational and storage demands for the future Square Kilometer Array (SKA) radio telescope are significant. Building on the experience gained with the collaboration between ASTRON and IBM with the Blue Gene based LOFAR correlator, ASTRON and IBM have now embarked on a public-private exascale computing research project aimed at solving the SKA computing challenges. This project, called DOME, investigates novel approaches to exascale computing, with a focus on energy efficient, streaming data processing, exascale storage, and nano-photonics. DOME will not only benefit the SKA, but will also make the knowledge gained available to interested third parties via a Users Platform. The intention of the DOME project is to evolve into the global center of excellence for transporting, processing, storing and analyzing large amounts of data for minimal energy cost: the *'ASTRON & IBM Center for Exascale Technology'*.

## Categories and Subject Descriptors

J.2 [**Physical Sciences and Engineering**]: Astronomy

## General Terms

Management, Legal Aspects


*AstroHPC'12,* June 19, 2012, Delft, The Netherlands.


## Keywords

Square Kilometer Array, Dome, ExaScale, Green Supercomputing, Extreme Streaming, Nano-Photonics

## 1. INTRODUCTION

It is clear that the computational requirements for the SKA are beyond the capabilities of even the fastest supercomputers currently available. Even with supercomputer technology expected to keep increasing in performance according to the trends observed over the last two decades for the foreseeable future, the streaming and real-time nature of SKA computing make it unlikely that these systems, or its software, will be ideally suited for this application. A significant research and development effort is therefore required.

While the SKA community consists of dozens of institutes, covering every corner of the world, it doesn't have the experience and manpower needed to tackle the SKA exascale problems on its own. Fortunately, the SKA also provides an interesting, challenging, and real-world problem for industry to base research and development efforts on. In particular, for IBM, with its focus on future Big Data and Big Data analytics, this is a highly interesting field of research. It promises to make analytics low cost and energy efficient.

In this paper we present unique public-private exascale computing research collaboration between ASTRON and IBM, designed specifically to tackle the energy efficient processing of huge amounts of streaming data and everything that requires. This presents a unique opportunity for the SKA community, industry and, via the Users Platform, interested third parties to benefit from each other's experience and facilities.

## 2. COMPUTING CHALLENGES IN SKA

The SKA will be built in two phases. Phase 1 is scheduled to start construction in 2016, and will contain around 20% of the total collecting area. Phase 2 will start construction in 2020. The computational challenges in the SKA are enormous. Although Moore's law is expected to provide the compute power needed by the time the SKA will be built, it is expected that future compute hardware will be markedly different from todays systems. Supercomputers in the SKA timeframe are expected to have very limited I/O bandwidth per floating point operation (flop)[4], making them not ideally suited for radio astronomical applications, which typically have very low computational intensity. The massive parallelism of these future systems, and the streaming nature of SKA processing, also make programmability of future systems a major concern.

To ensure we can effectively use these future systems, it is essential that the SKA project develops and maintains expertise in streaming exascale computing. None of the pathfinder or precursor instruments to the SKA are expected to scale beyond a few hundred teraflops, which leaves a scaling gap of at least three orders of magnitude.

The sheer volume of data coming from the SKA central correlator or beamformer makes it impossible to store intermediate data products. We therefore need to move to a streaming data model, possibly with buffering to allow for multi-pass calibration algorithms. This streaming nature, combined with an extremely low computational intensity of the algorithms involved, make the SKA unique among exascale applications.

This means that the SKA community has to actively engage both industry and the High Performance Computing (HPC) research community to ensure future systems, both hardware and software, can efficiently be used to process SKA data.

## 3. DOME

The DOME project is an initial 32.9 million Euro, five-year collaboration between IBM Research — Zürich (ZRL) and ASTRON, and is funded by the Dutch Ministry of Economic Affairs, Agriculture and Innovation (EL&I) and by the Province of Drenthe. The project investigates low-power exascale computing architectures and systems in support of the next generation radio telescope, the SKA. The project involves 25 research positions, located at ASTRON and at ZRL.

The objective of the DOME project is to investigate novel SKA exascale computing technologies and concepts, with a focus on energy reduction, data processing, data storage, and nano-photonics. The computing challenges for the SKA are several orders of magnitude higher than those of the current generation of radio telescopes. The processing requirements for the SKA therefore can only be met by basing the design on architectures and technologies that are beyond the current state of the art. DOME's ambition is to become the global center of excellence with respect to transporting, processing, storing and analyzing of large amounts of streaming data in the most energy efficient way possible.

The DOME project also aims to collaborate with Small and Medium Enterprises (SME's) and other parties in the (Northern) Netherlands in a Users Platform, in order to stimulate economic activity by supporting the development and testing of new high performance computing applications. The project is concerned with fundamental and industrial research. In fundamental research, capital is converted to knowledge, in industrial research, the newly gained knowledge will be converted to new capital.

## 4. RESEARCH LINES

IBM and ASTRON will collaborate on research in energy efficient data processing, data storage, and nano-photonics, applicable to the SKA. Research will focus on three main subjects:

- "Green supercomputing"
- "Extreme streaming"
- "Nano-photonics"

Within these three main themes, a number of research tracks have been identified. Five of these will start immediately and are detailed below. Two more will start after 24 months (*Compressive sampling* and *Realtime communication patterns for exascale computing*). We will not mention these in great detail.

### 4.1 Algorithms and machines

The SKA computational requirements are well beyond the capabilities of even the most powerful supercomputers today. The design of the systems that will accommodate such workloads with unprecedented levels of computing demands, requires new holistic analyses that consider both the particular requirements of the target applications, and the availability of hardware platforms now, and at the deployment stage. The Algorithms and Machines workpackage is intended to design a whole-system bounds framework that will enable the system design space exploration in early phases of the SKA implementation, and guide the design decisions of the platforms that will hold the future exascale systems.

This research project is based on a methodology, whose development has already started at the IBM Zürich Research Lab, that consists of three main components. First, a set of analytical models and equations that tie application properties, device technology and compute architecture trends, and make predictions on performance[7], power, and hardware costs. The second component is a simulation framework that quantifies inherent application properties in an architecture- and microarchitecture-independent manner[1, 2]. Finally, the analysis is validated with performance, power and thermal measurements across a set of actual hardware platforms that expand from high-performance processors typically used in HPC systems, to low-power processors employed in the embedded domain[6].

The project will start with a retrospective study of the LOFAR and, at a second stage, the analysis will be extended to the main computational kernels and data flow throughout the lifecycle of a SKA processing pipeline. Based on the bounds on projected performance, power and cost, these analyses will identify a number of architectural design points that make up the ideal computing platform for the SKA. Some of the insights will be unique to radio astronomy, while others may be shared by more conventional HPC applications, or even the so called Big Data or Business Analytics applications, relevant to IBM business.

### 4.2 Access patterns

The Access Patterns workpackage focuses on the interaction of users and processes with large scale storage systems. The SKA will be dealing with petabytes of data generated each day. All of this data will have to be kept on storage media and made available for data analysis and distribution purposes. This poses a serious challenge where solutions need to be found that optimize the efficiency and responsiveness of data provisioning mechanisms and minimize the impact in terms of energy consumption, investments and operational costs, while at the same time ensuring a high level of reliability of exabyte scale storage systems.

Within the Access Patterns workpackage some of the most challenging SKA related use cases, with respect to data access patterns, will be defined and used as input to simulate cases in test environments. New storage technologies, both at the physical and the file system level, will be investigated and used in test setups to demonstrate their potential. Since a wide range of storage media with distinctive characteristics exist, now and in the future, particular attention will be given to tiered approaches in conjunction with intelligent data placement and migration mechanisms in order to arrive at cost-effective and efficient storage solutions for the large scale data centers of the future

### 4.3 Nano-photonics

The transport of data is and will remain a major cost factor in future radio-astronomy systems such as the SKA, both for the initial investment and in terms of operational cost. The reduction of these initial investment and operational costs is the main goal in the photonics technology R&D in DOME. In this R&D work, the efforts for reducing the investment costs of the data transport system are focused on the development of low cost, high performance optical interconnects and on an efficient and clever use of copper and fiber-optic cable infrastructures. Important operational cost reductions are obtained by the development and deployment of low power data transport systems.

In the first phase IBM and ASTRON will collaborate on the application and investigation of surface mount optical interconnect technology for high-speed processing board systems and short range Radio over Fibre (RoF) links. In a second step the next-generation of low cost and low power optical interconnect technology will be developed and tested, using the SKA system as application area. In addition, the handling and processing of signals in the optical domain will be investigated in a microwave-photonics technology R&D section.

IBM has a strong research program on optical interconnects and nano-photonics aimed at applications in datacenters. ASTRON has programs on analog optical interconnects and microwave-photonic signal processing. The research described here will maximize the synergy between those lines.

### 4.4 Micro servers

The micro servers project investigates the possibility of using ultra low power hardware, best known from applications in mobile phones and tablet computers, in radio-astronomical processing. The underlying goal is both to limit energy consumption for some or, ideally, all computations in a future telescope, and to investigate the validity of low power hardware for scientific computing.

Recent experience has shown that, given a certain feature size and production process, a flop requires approximately the same amount of energy, regardless of the instruction set or architecture involved. The great difference in energy consumption between mobile and server processor must than lie in vastly greater performance of the latter and, crucially, more efficient peripheral hardware. By carefully selecting appropriate compute hardware and a minimal set of energy efficient peripheral hardware, we may significantly reduce energy required for SKA processing.

### 4.5 Accelerators

A recent trend in High-Performance Computing is to exploit the large-scale parallelism of many-core architectures like GPUs (Graphical Processing Units), to achieve high performance at low energy consumption. Within DOME, we want to study the interaction between algorithms and many-core architectures. The architectures share one common problem: the increase of memory bandwidth does not keep up with the increase in compute power, the gap has already grown to a factor of five. This imbalance forces users to rethink their algorithms. An example of such a newly developed algorithm is an algorithm that creates sky images: the new method needs far less memory bandwidth than existing methods [5].

We also intend to do research on many-core architectures: current many-core architectures differ in core counts, vector lengths, cache structures, I/O capabilities, latency-hiding techniques, programming models, and energy efficiencies. The study will answer the following questions:

- Which properties make an architecture (in)efficient?
- Are the I/O capabilities in balance with the computational power?
- Which architectures are energy-efficient?
- Which architectures are easily programmable?
- Which generic and architecture-specific optimizations and algorithmic changes are required?
- Can we distill a generic approach to program many-cores, or are too many ad-hoc approaches needed?

We did such a study for a single algorithm (computing correlations) [3, 8], and want to extend this to a broader range of algorithms.

The results are of high importance to the computer-science, radio telescope, and signal-processing communities: they teach us how to build and use many-core hardware, and how to process data for the Square Kilometre Array (SKA).

## 5. ASTRON & IBM CENTER FOR EXASCALE TECHNOLOGY

The intention of the DOME project is to evolve into the global center for processing, transporting, storing, analyzing, integrating, implementing, and applying large amounts of data for minimal energy cost. This center, the *'ASTRON & IBM Center for Exascale Technology'* will initially be located at the ASTRON buildings in Dwingeloo, and finally at the Sensor System Technology Campus in Assen. The center will form the basis for a growing cluster of companies and knowledge institutes, and will stimulate economic activity in fields related to sensor system technologies.

## 6. USERS PLATFORM & FACILITY SHARING

In the course of the project new knowledge will be generated in the area of the three main research themes. This knowledge, initially residing at the partners, will be made available to a wide range of companies via the Users Platform.

In general, the platform is to be used to identify and test new technologies for market applications. The Users Platform will support valorisation by several means, including facility sharing, expert meetings, access to ASTRON laboratory facilities for start-ups, and regulated access to DOME technology demonstrators. The Platform is open access under the conditions that there is mutual interest, that additional investments are being made by the joining partner, and that an appropriate Joined Development Agreement (JDA) is signed in which Intellectual Property (IP) rights are covered.

Using the Users Platform, the goal is to strengthen the SME's and to prepare companies for a role in SKA development. Companies, involved in early collaborative research will have an advantage when the tendering process will start in 2017 and beyond. The Users Platform also aims to prepare SME's for a role in hosting, servicing and application development for next generation datacenters. One of the ambitions of ASTRON and the SKA industrial consortium is to host the European Data Center for the SKA.

# 7. CONCLUSIONS

In this paper we've introduced a unique public-private collaboration initiative between IBM and ASTRON, aimed at technology development for the SKA. The extreme scaling challenges presented by the SKA are significant, but at the same time provide an interesting challenging and realistic target application for the development of novel techniques and systems.

# 8. REFERENCES


[1] Victoria Caparrós Cabezas and Phillip Stanley-Marbell. Parallelism and data movement characterization of contemporary application classes. In *ACM Symposium on Parallelism in Algorithms and Architectures (SPAA'11)*, pages 95–104, New York, NY, June 2011.

[2] Victoria Caparrós Cabezas and Phillip Stanley-marbell. Quantitative analysis of parallelism and data movement properties across the berkeley computational motifs. In *ACM International Conference on Computing Frontiers (CF'11)*, New York, NY, May 2011.

[3] R. V. van Nieuwpoort and J. W. Romein. Correlating Radio Astronomy Signals with Many-Core Hardware. *International Journal of Parallel Processing*, 1(39):88–114, February 2011.

[4] Peter M. Kogge et al. ExaScale Computing Study: Technology Challenges in Achieving ExaScale Systems, September 2008.

[5] John W. Romein. An Efficient Work-Distribution Strategy for Gridding Radio-Telescope Data on GPUs. In *ACM International Conference on Supercomputing (ICS'12)*, Venice, Italy, June 2012.

[6] Phillip Stanley-Marbell and Victoria Caparrós Cabezas. Performance, power, and thermal analysis of low-power processors for scale-out systems. In *IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW'11)*, pages 863–870, Shanghai, China, May 2011.

[7] Phillip Stanley-Marbell, Victoria Caparrós Cabezas, and Ronald P. Luijten. Pinned to the walls - impact of packaging and application properties on the memory and power walls. In *IEEE International Symposium on Low Power Electronics and Design (ISLPED'11)*, pages 51–56, Fukuoka, Japan, August 2011.

[8] Rob V. van Nieuwpoort and John W. Romein. Building Correlators with Many-Core Hardware. *IEEE Signal Processing Magazine (special issue on "Signal Processing on Platforms with Multiple Cores: Part 2 – Design and Applications")*, 27(2):108–117, March 2010.