ASTRONOMY WORKSHOP 2016

ASTRONOMICAL DATA AND COMPUTATION

September 5: Day One	
10:00-11:30	**Registration and orientation** 48 Kuybysheva st.
11:30-11:45	**Workshop opening** Room 700, 48 Kuybysheva st.
11:45-14:00	**Session Ⅰ: Modeling and analysis** (Chair: H. P. Singh) Room 700, 48 Kuybysheva st.
Thebe Medupe	Stellar astrophysics research in South Africa
Maria Kirsanova	Morphology of infrared bubbles around HII regions
Ilfan Bikmaev	Optical identifications of X-ray sources with RTT-150
Shazrene Mohamed	Computational astrophysics in South Africa
14:00-15:00	**Lunch at Panorama Business Hotel** 44 Kuybysheva st.
15:15-17:30	**Session Ⅱ: Data exchange and presentation** (Chair: S. Mohamed) Room 700, 48 Kuybysheva st.
Jonathan Sievers	Long-wavelength Data Analysis in South Africa
Ulisses Barres	Implementation of the Brazilian Science Data Center (BSDC)
Oleg Malkov	Binary star database BDB
Ricardo Ogando†	LIneA: towards a Brazilian e-astronomy center
19:00	**Opening banquet at Park Inn Hotel** 98 Mamina-Sibiryaka st.

† Remote presentation.

September 6: Day Two	
10:00-11:45	**Session Ⅲ: Big Data on the national scale** (Chair: O. Malkov) Room 700, 48 Kuybysheva st.
Harinder Singh	Big Data in astronomy – Indian perspective
Bruno Castilho†	Overview on Brazilian projects on computational and Big-Data Astronomy
Sandeep Sirothia	MeerKAT and upcoming data challenges
11:45-12:00	**Tea/coffee break** Room 700, 48 Kuybysheva st.
12:00-14:00	**Session Ⅳ: Databases for astrophysics** (Chair: U. Barres) Room 700, 48 Kuybysheva st.
Angelo Fausti Neto†	Analyzing large astronomical data sets: The Science Portal solution
Nikolay Samus	Big Data and variable stars
Aleksey Shlyapnikov	The project of the Crimean Astronomical Virtual Observatory (CrAVO)
14:00-15:00	**Lunch at Panorama Business Hotel** 44 Kuybysheva st.
15:15	**Excursion to Yeltsin Center (Museum of Modern Russian History)** Meeting point at 48 Kuybysheva st. (main entrance)

† Remote presentation.

September 7: Day Three	
09:15-11:45	**Session Ⅴ: Data logistics** (Chair: X. Chen) Room 700, 48 Kuybysheva st.
Olga Zhelenkova	Problems of long-term preservation of observation data on the example of SAO RAS archive system
Alexei Pozanenko	Cosmic gamma-ray bursts and the V's of Big Data
Brian van Soelen	Multi-wavelength astronomy in South Africa
Russ Taylor†	IDIA and the Big Data Challenge in South African Astronomy
11:45-12:00	**Tea/coffee break** Room 700, 48 Kuybysheva st.
12:00-14:15	**Session Ⅵ: Computation and algorithms** (Chair: S. Sirothia) Room 700, 48 Kuybysheva st.
Liang Gao	Computational astrophysics in China
Xuelei Chen	Prospects of 21 cm cosmology
Maxim Borisyak	Deep Learning for Image Processing in Astronomical Experiments
Reinaldo Rosa†	Improving intensive data analysis in astronomy using many-core heterogeneous computing
14:15-15:15	**Lunch at Panorama Business Hotel** 44 Kuybysheva st.
15:30-16:00	**Closing session: Summary and remarks from SOC**
16:00-17:00	**Roundtable discussion on cooperative projects**
19:00	**Closing banquet at Park Inn Hotel** 98 Mamina-Sibiryaka st.

† Remote presentation.

Ⅰ : Modeling and analysis

Stellar astrophysics research in South Africa

Prof. Thebe Medupe
North-West University, South Africa

The history of stellar astrophysics research in South Africa started in 1820 with the first Astronomical Observatory in Cape Town. The early days were marked by accurate astrometry and measurement of distances to stars. In the last 50 years the main work has been studies of pulsating stars. Pulsating stars pulsate because of seismic waves inside them. In this talk I will review some of the important works that have been done in South Africa on this, and other contributions to Stellar research work both from theoretical and observational angles.

Morphology of infrared ring-like nebulae around HII regions

Dr. Maria Kirsanova
Institute of Astronomy of the Russian Academy of Sciences, Moscow, Russia

Regions of ionized hydrogen are often visible at infrared (IR) wavelengths as ring-like nebulae. This is because the emission at IR wavelengths arises from heated dust surrounding the ionized region, and this dust has been partially pushed away (or evaporated) by stellar wind and radiation pressure from the massive central star. We conducted a search for IR ring-like nebulae around H II regions using the New GPS 20cm survey from the VLA and the GLIMPSE (8 μm) and MIPSGAL (24 μm) surveys from Spitzer. Objects were selected which appear ring-like at 8 μm, contain extended emission at 20 cm, and in many cases show an interior extended ring-like structure at 24 μm. We made a catalog of 92 ring-like nebulae which are well approximated by an ellipse at both 8 and 24 μm. In this catalog 28 objects are nearly circular, with an eccentricity $e < 0.3$.

Optical identifications of X-ray sources with RTT-150

Prof. Ilfan F. Bikmaev
Kazan Federal University, Kazan, Russia

In this report we will present results of optical identifications of hard X-ray sources (discovered by space missions INTEGRAL, RXTE, SWIFT) by using 1.5-meter Russian-Turkish optical telescope (RTT-150). Angular resolutions of X-ray telescopes are not enough in many cases to give direct identifications with corresponding optical counterparts. Electronic databases (archives) and astronomical catalogues play an important role for cross-identifications with X-ray maps and help to identify optical counterpart candidates. Results of optical identifications of cluster of galaxies among candidates discovered by PLANCK space mission will be presented too.

Computational astrophysics in South Africa

Dr. Shazrene Mohamed
South African Astronomical Observatory, Cape Town, South Africa

Owing to its geographical advantage and the availability of large telescopes, astronomy in South Africa has largely been dominated by observations. Over the past decade however, computational astrophysics has grown steadily and is set to become a major part of astronomical research in the era of SKA and Big Data. In this talk I will describe the major computational resources currently available in South Africa, as well as the facilities that are due to come online within the near future. I will also highlight the areas of computational research being carried out in the country including simulation and visualization of data for the SKA, galaxy and cosmology simulations, gravitational wave astronomy and stellar astrophysics.

Ⅱ: Data exchange and presentation

Long-wavelength Data Analysis in South Africa

Prof. Jonathan Sievers
University of KwaZulu-Natal, Westville, South Africa

A variety of projects in South Africa are using or producing data in the radio/sub-mm. Besides MeerKAT, South Africa hosts or South Africans are involved in HERA/PAPER, C-BASS, many CMB experiments, and regularly use other radio telescopes such as JVLA/GMRT. We present some of the ongoing projects and how their data are handled/analyzed.

Implementation of the Brazilian Science Data Center (BSDC)

Dr. Ulisses Barres de Almeida
Brazilian Center for Research in Physics, Rio de Janeiro, Brazil

Astrophysics is increasingly characterised by what became known as "big data" or "data-intensive science." To make the most of the available information, one of the bottlenecks is the cataloging activities and access to scientific data of all kinds. In astrophysics, it is increasingly necessary to combine data from instruments operating in different bands of the electromagnetic spectrum, and even different messengers. It is necessary therefore to draw strategies and initiatives that will make the most of the data, and build the necessary bridges between areas of specialization.

The Brazilian Science Data Center (BSDC) is an initiative held in conjunction with the Italian Space Agency (ASI) for the implementation of an astronomical database system as complete and comprehensive as possible, following the principle of "science-ready data". This concept is based on cataloging and availability of scientific data in a format ready for direct scientific application. The availability of data is made through agreements with the collaborations and observatories responsible for them, so that the information is "certified" in quality.

We will present the concept and the current status of the BSDC project. Current development focus on software development for visualization, and statistical data processing, as well as a VHE gamma-ray database and polarimetry database. There is also the intention, once the project is matured, to expand the tool to serve the Brazilian community in other ways.

Binary star database BDB

Prof. Oleg Malkov
Institute of Astronomy of the Russian Academy of Sciences, Moscow, Russia

The fully operational version of the world's principal database of binary and multiple systems of all observational types is presented. The Binary star DataBase, BDB, is available at http://bdb.inasan.ru and is created as a source joining the comprehensive information on binary stars of all observational and evolutionary types. It provides the user with synthesis of data of large variety of catalogues and databases of binaries of different types: visual, orbital, astrometric, eclipsing, spectroscopic, photometric, etc. The BDB contains all data from catalogues on about 50,000 stellar systems of multiplicity 2 to 22: positional, photometric, spectroscopic, orbital and astrophysical parameters are provided when available. Organization of the information is based on the careful cross-identification of the objects. This allows the user, in particular, to search data on binaries having certain sets of parameters within the complete catalogued data set.

LIneA: towards a Brazilian e-astronomy center

Dr. Ricardo Ogando
Observatório Nacional/MCT, Rio de Janeiro, Brazil
Laboratório Interinstitucional de e-Astronomia (LIneA), Rio de Janeiro, Brazil

We describe how LIneA was created, what it has produced in terms of software and hardware infrastructure in the last 10 years of work on Dark Energy Survey and in preparation to many other projects, such as LSST, when an e-science center in Brazil will be needed to handle huge data volume, velocity, and variability. Parallelism, provenance, and visualization are some of the challenges faced at LIneA in order to generate products and achieve scientific results.

Ⅲ: Big Data on the national scale

Big Data in astronomy – Indian perspective

Prof. Harinder P. Singh
University of Delhi, New Delhi, India

India has recently joined mega projects in Astronomy including TMT (Thirty Meter Telescope) and the LIGO gravitational wave detector. A multi wavelength Astronomy satellite (ASTROSAT) was launched less than a year ago. A ~4m class ARIES optical telescope has become operational. I shall discuss the big data challenges from an Indian perspective.

Overview on Brazilian projects on Computational and Big-data Astronomy

Dr. Bruno Castilho
Laboratório Nacional de Astrofísica, Itajubá, Brazil

In the last years Brazil has invested both at the local infrastructures and in international collaborations in the field of big data and computational astronomy. This investments are crucial to foster the general development of this extreme important field. The work of several Brazilian groups in the last decade either in computer infrastructure and software or in surveys are now consolidating a more national effort. Recently Brazil has inaugurated the fastest computer in Latin America and one of its uses will be astronomy, this computer is/will be linked to a broader net of computations infrastructure the will allow the effective participation in SLOAN III and IV, DES, JPAS and LSST supported by data and software portals and national databanks. We will present the latest developments in the field and give some perspectives.

MeerKAT and upcoming data challenges

Dr. Sandeep Sirothia
Square Kilometre Array South Africa, South Africa

The South African MeerKAT radio telescope will be an array of 64 interlinked receptors is a precursor to the Square Kilometre Array (SKA) telescope and will be integrated into the mid-frequency component of SKA Phase 1. We will present highlights and recent milestones.

In addition, we also present our steps in preparation of an extraordinary era characterised by huge amounts of data from new and upcoming observational facilities including the Southern African Large Telescope (SALT), the Karoo Array Telescope (MeerKAT) and the SKA.

Ⅳ: Databases for astrophysics

Analyzing large astronomical data sets: The Science Portal solution

Dr. Angelo Fausti Neto
Laboratório Interinstitucional de e-Astronomia (LIneA), Rio de Janeiro, Brazil
Large Synoptic Survey Telescope

Large multi-wavelength photometric surveys have had a profound impact in astronomy forcing new methods in computer science and statistics due to the large data sets and the complexity involved in the science analysis. In this talk we describe the Science Portal: a web-based infrastructure to support science analysis for the Dark Energy Survey (DES) collaboration being developed in Brazil since 2007 through the DES-Brazil consortium and LIneA. The Science Portal is being used for several applications in DES including real-time assessment of the survey image quality, visualization and data exploration, and more recently to prepare science-ready catalogs and run science analysis pipelines. We discuss the challenges and solutions to analyze DES Year 1 data and perspectives for the full DES data release planned for 2018/19. While the Science Portal has being designed for DES it can be extended to other surveys. We discuss, in particular, the importance of having such an infrastructure for the Large Synoptic Survey Telescope (LSST) in the next decade.

Big Data and variable stars

Prof. Nikolay Samus
Institute of Astronomy of the Russian Academy of Sciences, Moscow, Russia

Having started just with a handful of stars, the variable star science has grown to about half a million known variable stellar objects, with quite a number of parameters known for each of them. Handling this amount of data with techniques traditional for variable-star catalogs becomes impossible. I will review the history of variable-star list, discuss the current effort in the field of variable-star catalogs, and consider future directions of work in the field.

The project of Crimean Astronomical Virtual Observatory (CrAVO)

Dr. Aleksey Shlyapnikov
Crimean Astrophysical Observatory, Nauchnyj, Russia

An overview of the most important components of the CrAVO is presented, including the main principles of formation of databases with information about astronomical objects and their physical characteristics, derived from observations obtained at the Crimean Astrophysical Observatory (CrAO) and published in the “Izvestiya of the CrAO” and elsewhere. Emphasis is placed on the DBs missing from the most complete global library of catalogs and data tables, VizieR. We specially consider the problem of forming a digital archive of observational data obtained at the CrAO as an interactive DB related to database objects and publications. We present examples of all our DBs as elements integrated into the CrAVO. We illustrate the work with the CrAO DBs using tools of the IVO: Aladin, VOPlot, VOSpec, in conjunction with the VizieR and Simbad DBs.

Ⅴ: Data logistics

Problems of long-term preservation of observation data on the example of SAO RAS archive system

Dr. Olga Zhelenkova
Special Astrophysical Observatory of the Russian Academy of Sciences, Nizhnii Arkhyz, Russia

The data of experiments require archiving and reliable storage with accessibility and semantic reenterability. In the SAO RAS in the early 80-ies for these purposes a digital archive of radio observations was created. Then with the introduction of CCD cameras into observations in the late 80s we started to develop a data bank, which united the different digital collections. Latter we realized an archiving system with a search information system. To date, the archive capacity is 1.5 TB, also there are 0.5 million files and 1 million records in the database. The system supports free web access to 16

local archives with digital collections of different devices used, or been used in the telescopes of the observatory. Two servers with PostgreSQL database each support the system. Each server data storage area has a similar structure and content. One server supports and contains the working version of the system, the second one supports the test version, with which we carry out and test all the new developments. The system was developed so that it was possible to add a new collection. There is a possibility of hosting other Russian telescope archives. In order to provide additional reliability of information keeping, archival data is still stored on the optical CD/DVD disks. There is no doubt in the need for long-term storage of astronomical data. As example, the sky survey conducted by the RATAN-600 in 1980, with the results of which it was obtained the first restriction on the value of CMB fluctuations, then these observations we used to study the variability of the radio sources and transients search. The life cycle of modern digital carriers is usually 5-10 years, that also applies to the read-write hardware and software. So while changing magnetic, data and streamer tapes to hard drives some of our collections, fortunately minor could not be read from the carriers for these reasons. A timely migration of digital files on the modern carriers is required to ensure long-term storage of data. Recently, a new storage media is appeared. It is the M-disc designed for long-term storage of unchanging data (http://millenniata.com). We plan to make migration of the archived data to the M-discs. The manufacturer promises shelf life of the disks up to 100 years but it is a predicted term. Nevertheless, it is not possible to completely abandon from external carriers for long-term data storage and fully transfer the information to the database only. Other way of supporting long-term preservation is based on a multi-tiers hardware architecture and special software systems like iRODS (integrated Rule-Oriented Data System, http://irods.org). It is an open source data management software used by research organizations. iRODS virtualizes data storage resources, so users can take control of their data, regardless of where and on what device the data is stored. We are going to start works for migration our data into the new environment based on iRODS.

Cosmic gamma-ray bursts and "V"s of big data

Dr. Alexei Pozanenko
Space Research Institute, Moscow, Russia

We consider Gamma-Ray Bursts (GRB) as a source of big data. We discuss the pipeline of GRB data generating, receiving, reduction and observations of GRB in the context of a search of a counterpart of gravitational-wave events detected by the LIGO experiment. We also discuss possible cooperation in GRB observations with BRICS countries.

Multi-wavelength astronomy in South Africa

Dr. Brian van Soelen
University of the Free State, Bloemfontein, South Africa

South Africa has a long history in astronomy and while, historically, the focus has been on optical telescopes, astronomers are focused on all aspects from radio up to the highest energy gamma rays. In recent years, the major focus has been in radio astronomy, with on-going development towards hosting the Square Kilometre Array which will be the world's largest radio interferometer. However, South African astronomers are also involved in a number of other multi-wavelength initiatives which have important Big Data requirements. Such projects include the development of Virtual Observatory tools, the management of the data from the Southern African Large Telescope (SALT), observational proposals for the Large Synoptic Survey Telescope (LSST), the development of MeerLICHT, participation in the H.E.S.S. Collaboration, and preparation for collaborations in Cherenkov Telescope Array (CTA). This talk will present an overview of South Africa's current or planned participation in these projects.

IDIA and the Big Data Challenge in South African Astronomy

Prof. Russ Taylor
Inter-University Institute for Data Intensive Astronomy, South Africa

In 2012 the ten member countries of the international Square Kilometer Array Organization awarded the site for the mid-frequency dish array of the first phase of the Square Kilometre Array (SKA) as well as

the innovative technologies of the second phase to a partnership of nine African countries led by South Africa. The SKA is one of the world's largest mega-science projects and drives one of the most significant big data challenges of the coming decade. The construction of MeerKAT at the South African SKA site and the associated ambitious multi-wavelength science programs marks the beginning of the new astronomy big data revolution in Africa. The South African university community is rising to this big data challenge by collaborating to establish the Inter-University Institute for Data Intensive Astronomy (IDIA). I will discuss the data intensive challenges confronting South African astronomers and the strategies and programs that have been initiated within IDIA to meet them.

Ⅵ: Computation and algorithms

Computational astrophysics in China

Prof. Liang Gao
Key Laboratory of Computational Astrophysics, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China

Prospects of 21cm cosmology

Prof. Xuelei Chen
National Astronomical Observatory of China, Beijing, China

The 21cm line of neutral hydrogen can be a powerful cosmological probe throughout the history of the Universe, but at present the detection is still limited to relatively nearby galaxies. I describe a few ongoing and upcoming experiments which attempt to detect the 21cm signal, focusing on the Tianlai experiment in China, and the SKA.

Deep Learning for Image Processing in Astronomical Experiments

Maxim Borisyak
National Research University Higher School of Economics, Moscow, Russia
Yandex School of Data Analysis, Russia

In recent years, Deep Learning has become a powerful tool for Data Analysis including image processing. Notably, it became the first Machine Learning algorithm that surpassed human performance in visual pattern recognition. Today usage of Deep Learning methods in natural sciences, such as High Energy Physics and Astrophysics, is rapidly growing. In this talk, we cover methods for image processing in the astrophysical experiment. Two particular methods, namely, track recognition and learning read-out model from real data, are explained in detail. For illustration purposes, we consider an astrophysical experiment: Cosmic Rays Found In Smartphones, which proposes usage of private mobile phones as a ground detector for Ultra High Energy Cosmic Rays. Unusual structure of the detector and unknown properties of individual sensors lead to a number of challenges which can be bypassed with the help of Deep Learning methods.

Improving intensive data analysis in astronomy using many-core heterogeneous computing

Prof. Rosa Reinaldo
National Institute for Space Research, São José dos Campos, Brazil

Intensive analysis of large data sets from advanced research in astrophysics and cosmology deals with amounts of data flow greater than 1TB/h (Big data workload). In this talk we will discuss, within the astronomy scenario, how to improve performance for digital image analysis in the context of Data Science. Performing an intensive morphometric analysis of digital images obtained from the SDSS projects we propose a quantitative balance between hardware and database algorithms that is able to optimize the analytical performance using heterogeneous computing (based on the general purpose manycores technology) with solutions from NoSQL approach. Furthermore, this heterogeneous computing solution allows resorting to the Machine Learning paradigms for reliable automation in the realization of the most important analytical tasks as classification and pattern recognition of structural

information. In this framework, the minimum automatic heterogeneous architecture (which we call MAHA) solution should provide the lowest energy consumption as a determinant of the HPC system. As a highlight of this study we show that the performance of a generic MAHA depends only on four main variables: amount of cores, number of threads per core, the percentage of parallelized workload, and the energy flux efficiency, even considering the data assimilation and validation of models as 2nd order tasks.