论文

SKA低频成像系统在中国SKA区域中心的部署

劳保强, 安涛*

中国科学院上海天文台, 上海 200030

*联系人, E-mail: antao@shao.ac.cn

收稿日期: 2019-09-20; 接受日期: 2019-10-15; 网络出版日期: 2020-02-21

国家重点研发计划(编号: 2018YFA0404603)和国家自然科学基金(编号: U1831204, 11873079, 11703069)资助项目

摘要 平方公里阵列(Square Kilometre Array, SKA)将成为世界上最大的综合孔径射电望远镜. 凭借其高分辨率、超高灵敏度和超快巡天速度, SKA能以前所未有的细节探索宇宙. 连续谱巡天是SKA的几个主要观测模式之一, 为理解和认识射电天空奠定基础. SKA探路者设备已经运行了几年, 积累了丰富的巡天观测数据. 中国天文学家正在为中国SKA区域中心(China SKA Regional Centre, CSRC)的建设进行积极的准备, 已经完成了原型机. 本文详细介绍了在CSRC原型机中部署SKA低频数据处理系统的进展. 默奇森宽场阵列(Murchison Wide-field Array, MWA)是SKA低频阵列的先导项目. 我们使用MWA的观测数据来验证成像管线系统的部署并测试其正确性. CSRC原型机的数据处理结果与已发表的结果一致, 表明在CSRC原型机上部署开发的并行数据处理系统取得了巨大成功. 而且, 该原型机的灵活适配性和可扩展性使其可以轻松扩展, 支持未来SKA第一阶段试运行期间的大规模连续谱成像实验. 这些经验将有助于改善成像管线系统, 并有助于完善CSRC.

关键词 平方公里阵列(SKA), 默奇森宽场阵列(MWA), 连续谱巡天, 中国SKA区域中心, 数据处理

PACS: 47.27.-i, 47.27.Eq, 47.27.Nz, 47.40.Ki, 47.85.Gj

1 引言

超级射电望远镜平方公里阵列(Square Kilometre Array, SKA)将利用其高分辨率和超高的灵敏度实现基础物理学和宇宙学相关的科学实验^[1,2]. 其中, 较为关键的观测实验之一是连续谱巡天观测^[3]. 该观测实验是SKA实现星系演化、宇宙中大规模结构的演化、宇宙磁场等关键科学目标的重要手段. 由于SKA巡天观测的视场较大, 观测时间长, 完成一

次巡天任务将会产生PB量级的原始数据和TB量级的图像数据, 这将给巡天数据存储与处理带来巨大的挑战^[4,5].

目前, 许多射电望远镜已经开展了连续谱巡天观测, 例如巨米波射电望远镜(Giant Metrewave Radio Telescope, GMRT)^[6]和甚大阵(Very Large Array, VLA)^[7]. 还有过去十年里以宇宙再电离的21 cm探测为主要科学目标驱动下建造的新一代射电望远镜, 包括低频阵列(Low Frequency Array, LO-

引用格式: 劳保强, 安涛. SKA低频成像系统在中国SKA区域中心的部署. 中国科学: 物理学 力学 天文学, 2020, 50: 059501

Lao B Q, An T. Deployment of SKA low frequency imaging system in China SKA Regional Centre (in Chinese). Sci Sin-Phys Mech Astron, 2020, 50: 059501, doi: 10.1360/SSPMA-2019-0332

FAR)^[8]、探测再电离时代的精密阵列(Precision Array to Probe the Epoch of Reionization, PAPER)^[9]、长波长阵列(Long Wavelength Array, LWA)^[10]和默奇森宽场阵列(Murchison Wide-field Array, MWA)^[11].

近年, 使用这些已建造的射电望远镜进行的新一代低频连续谱巡天项目有: VLA 74 MHz 低频天区巡天终极版(Low-frequency Sky Survey redux, VLSSr)^[12]、LOFAR 120–180 MHz多快照天区巡天(Multi-Snapshot Sky Survey, MSSS)^[13]、塔塔基础研究所GMRT 150 MHz天区巡天(Tata Institute for Fundamental Research GMRT Sky Survey, TGSS)^[14], 以及MWA 72–231 MHz银河与河外星系全天区(GaLactic and Extragalactic All-sky MWA, GLEAM)巡天^[15]. 仅仅从SKA第一阶段低频阵列(SKA1-LOW)的角度来看, LOFAR, MWA和PAPER作为SKA的先导或探路者项目均能深入揭示SKA1-LOW的设计过程, 给SKA1-LOW的建设与技术攻关提供宝贵的实践经验. 尤其是MWA, 在这方面发挥着至关重要的作用, 因为其台址恰好坐落在SKA1-LOW的台址, 且在该台址环境下取得了大量的低频射电天文科学与技术成果.

SKA能否取得里程碑式的重大科研成果, 取决于SKA科学数据处理能力和科学分析研究水平, 而承担SKA最终的科学处理、面对科学家用户、直接产生科学成果的工作将由SKA区域中心(SKA Regional Centre, SRC)完成. 上海天文台推进了中国SKA区域中心(China SKA Regional Centre, CSRC)建设, 完成了CSRC原型机^[16]. 为了验证CSRC原型机是否具备MWA连续谱巡天整体数据处理的能力, 本文在CSRC原型机上, 借助MWA GLEAM巡天项目原始数据, 开展了GLEAM巡天数据的传输、存储和处理等关键技术的研究. 同时, 为将来参与SKA低频巡天相关研究奠定基础.

2 MWA连续谱巡天项目

MWA作为SKA的先导项目之一, 开展了低频射电连续谱巡天项目即GLEAM巡天. GLEAM巡天主要采用漂移扫描方式, 巡天观测分为两个阶

段: 第1年(2013年8月9日至2014年6月18日)和第2年(2014年8月4日至2015年7月6日). 本文主要关注第1年的巡天观测. GLEAM第1年巡天覆盖了约2.48万平方度天区范围, 具体包括赤纬+30°以南天区和银河系银纬10°以外的天区, 但不包括麦哲伦云等一些区域^[15]. 由Natasha等人发表的GLEAM第1年巡天星表的天区覆盖范围详见图1的有色区域.

第1年巡天观测的频率范围为72–231 MHz, 分为5个连续的频段(编号: 69, 93, 121, 145, 169), 对应的频率范围分别是(72–103), (103–134), (139–170), (170–200)和(200–231) MHz. 为了避免卫星射电频率干扰的影响, 去掉了134–138 MHz范围内的观测数据. 每个频段的带宽为30.72 MHz, 每个频段又分为4个子频段, 带宽为7.68 MHz. 第1年的巡天主要将天区按赤纬分为7条带(+19, +2, -13, -27, -40, -55, -72), 每条带观测1晚, 循环观测直到天区扫描完, 一共观测了28晚. 每晚覆盖8–10 h赤经范围, 具体范围取决于年中的时间. 太阳白天较为明亮且在MWA频率下是时变的, 因此观测仅在夜间进行. 在一个晚上, 每个频段连续观测120 s完成1次快照观测, 10 min内循环则完成5个频段的快照观测. 每晚平均需要进行约45次快照观测才能完成8–10 h范围的天区扫描任务. 每个晚上每隔2 h, 会在所有5个频段上观测校准场, 获得5组120 s的快照观测, 用于目标场的数据校准. 第1年巡天所有快照观测的原始数据总量约为300 TB.

图1 (网络版彩图)已发表的GLEAM第1年巡天星表天区覆盖范围

Figure 1 (Color online) Sky area coverage of published catalog.

3 软硬件系统

3.1 硬件系统

CSRC原型机分两个时间段建设, 第一阶段时间为2017年6月至2018年6月, 第二阶段为2018年10月至2019年7月. 本文采用的硬件系统是CSRC原型机第一阶段设备, 如图2所示. 该原型机主要由1个管理与登录节点、8个计算节点和分布式存储系统组成. 其中, 计算节点包括2个Purley计算节点、2个GPU计算节点和4个KNL计算节点. 所有计算节点之间使用英特尔的OPA(Omni-Path Architecture)交换机相连, 用于计算节点之间的数据高速交换. 该OPA交换机(型号: 100SWE48UF2)共48个端口, 每个端口传输速率为100 Gb/s. 同时, 所有计算节点、分布式存储系统和管理与登录节点之间使用千兆交换机相连, 用于计算节点与存储的管理. 分布式存储系统由100个8 TB的希捷(型号: ST8000NM0075)硬盘组成, 总存储容量为800 TB. 其中, 400 TB存储部署了NFS文件系统, 另外400 TB存储部署了Beegfs文件系统, 本文主要使用了Beegfs文件系统的存储. 管理与登录节点为戴尔服务器(型号: Dell R830), 本地硬盘由5个1 TB的希捷硬盘组成. GPU节点配置了4个GPU显卡(型号: Nvidia Tesla V100). 各节点主要参数指标见表1. 原型机中的Purley计算节点和KNL计算节点所有CPU核(316个核心)的计算能力为12.6 TFlops, 平均0.04 TFlops/核; GPU节点的所有GPU的计算能力约为104 TFlops, 平均13 TFlops/GPU.

表1 各节点主要参数指标

Table 1 Main parameter indicators of nodes

节点名称	CPU型号(核数)	内存型号(容量)
管理与登录	2*E5-4610v4(12)	8*32 G DDR4 (256 GB)
Purley	2*Intel Skylake 6152(44)	24*32 GB Registered DDR4 (768 GB)
GPU	2*Intel Xeon 2690 v4(56)	8*32 GB RDIMM DDR4 (256 GB)
KNL	Intel Xeon Phi 7250(68)	6*32 GB RDIMM DDR4 (192 GB)

此外, KNL节点配置了两级高速缓存, 第一级为6个1.2 TB的固态硬盘(SSD), 第二级为4个2 TB的固态硬盘(SSD). Purley节点同样配置了两级高速缓存, 第1级为2个256 GB SSD, 第2级为4个2 TB SSD. GPU节点则自带了本地硬盘(Intel DC P3520 HHHL 4 TB), 总容量4 TB. 这些高速缓存主要用于数据处理过程中数据的缓存, 以提高数据处理的运行效率.

MWA数据处理机器采用澳大利亚Pawsey超算中心的Galaxy超级计算机, 该超级计算机共有472个CPU节点和64个GPU节点. 其中, 每个CPU节点包含2个10核的CPU(型号: Intel Xeon E5-2690V2), 每个CPU节点的内存大小为64 GB; 每个GPU节点包含1个8核CPU(型号: Intel Xeon E5-2670)和1个GPU显卡(型号: NVIDIA K20X), 每个GPU节点的内存大小为32 GB. 该超级计算机所有CPU核9440个核心的计算能力为192.1 TFlops, 平均0.02 TFlops/核, GPU节点的GPU总计算能力为3.95 TFlops, 平均0.062 TFlops/GPU. 综上分析,

图2 (网络版彩图)CSRC原型机第一阶段逻辑结构图

Figure 2 (Color online) The first phase logical structure diagram of CSRC prototype.

本文的原型机单节点的计算性能和内存资源均优于Galaxy超级计算机, 表明本文的原型机满足MWA数据处理的需求.

原型机使用SLURM¹⁾作为资源管理系统, 将原型机所有节点分为三个分区: CPU分区(2个Purely节点)、GPU分区(2个GPU节点)和KNL分区(4个KNL节点), 进行节点资源的统一管理. 同时, SLURM也作为作业或任务调度系统, 方便用户提交任务至计算节点运行, 随时查询各自任务的运行情况.

3.2 软件系统

原型机所有节点安装的操作系统为Centos 7.0. 软件环境采用Modules²⁾进行管理, 便于用户动态地修改自己的软件环境, 且不影响其他用户的软件环境. 目前软件系统部署了一些常规软件和用于MWA数据处理大部分软件. 当进行MWA GLEAM巡天数据处理时, 用户只需要在终端运行: `module use /mnt/beegfs/modulefiles`和`module load MWA_Tools/mwa-sci`, 即可使用MWA数据处理软件环境.

4 数据处理方法与并行实现

如图2所示, 本文主要通过开通高速网络(带宽1 Gbps)完成MWA GLEAM巡天数据的传输, 数据实际的平均传输速率为738 Mbps即92.25 MB/s^[17]. 通过管理与登录节点运行传输程序进行数据传输, 并实时保存数据至分布式存储系统. 目前该分布式存储系统共存储了100 TB GLEAM巡天第1年的原始数据, 包含169频段所有快照观测数据和部分其他频段快照观测数据.

经过反复验证测试, 我们完成了GLEAM巡天第1年数据的处理流程. 首先处理每个快照观测数据; 其次, 按天将每天所有快照图像进行拼接(Mosaic), 拼接成当天扫过天区的图像; 最后, 按照周(7条带)将每个周期内所有天里拼接后的图像

再次拼接, 拼接成当周扫过的天区图像, 获得4个周大天区图像.

具体的快照观测数据处理方法如下.

(1) 读取快照原始数据进行射频干扰标记与削减, 对可见度数据在时间与频率上作平均, 最后生成Measurement Set文件格式数据并压缩数据. 具体方法详见文献[18, 19].

(2) 在已发表的星表中寻找步骤(1)快照数据观测视场范围内对应的射电源信息(流量强度和位置等), 从而制作出相应的天空模型.

(3) 利用步骤(2)的天空模型对步骤(1)生成的Measurement Set数据进行校准, 校准之后的数据保存在原Measurement Set文件中. 这里的校准算法采用Offringa等人^[20]提出的MitchCal算法.

(4) 如果快照视场范围附近有超过50 Jy流量强度的射电源, 则先使用Peeling算法^[21]扣除掉这个亮源, 从而消除亮源对视场范围内的可见度数据的影响. 随后进行深度去卷积与大视场成像, 输出最终的快照图像. 深度成像将会搜寻快照图像射电源个数, 以确保快照图像处理的质量和正确性. 当源的个数小于等于20个时, 我们认为这个快照图像的质量是非常糟糕的, 说明数据校准存在问题. 还需要利用当晚校准源的校准结果进行重新校准, 具体步骤是: 利用步骤(3)中的校准算法对校准源数据进行校准; 将校准源的校准结果应用到该快照数据, 并生成校准后的数据; 重复步骤(4)进行深度成像. 这里的深度去卷积与大视场成像是采用文献[22]中的WSCLEAN软件实现. 这里的亮源为Hydra A, Pictor A, Hercules A, Virgo A, Crab Cygnus A和Cassiopeia A, 称为A-team源.

(5) 使用源搜寻软件在步骤(4)输出的快照图像中寻找射电源, 将找到的射电源的相关信息(如位置(RA, DEC)和峰值强度等)与已经发表的星表结果进行对比, 并进行流量强度的修正, 得到输出修正后的快照图像. 这一步主要为了消除大气、电离层等对快照图像中的源的位置和流量的影响. 源搜寻软件采用文献[23]中的Aegean软件.

1) <https://slurm.schedmd.com/documentation.html>

2) <http://modules.sourceforge.net/>

为了满足SLURM任务提交的模式, 上述所有处理步骤的执行命令和代码均采用Shell脚本进行封装, 形成一套快照观测数据处理流水线(Pipeline).

由于所有快照观测数据的处理步骤几乎相同, 为了提高数据处理效率, 我们设计了批量数据处理方法, 即同时并行处理多个快照观测数据. 通过充分利用作业调度系统中的任务数组功能, 如图 3 所示, 为GLEAM巡天一晚观测的所有数据(假设共有 N 个快照数据, 且 $N \geq 43$)进行并行处理. 用户与管理与登录节点提交(Sbatch)8个任务脚本, 8个任务分别在不同的计算节点上运行, 实现多节点的并行处理. 同时, 在每个任务中设置任务数组(Array), 根据数组的长度(L)与数值大小(ID), 将每个任务又分成 L 个子任务并行运行. 数组中的数值作为编号选取不同的快照数据进行处理, 从而实现单节点的并行数据处理. 如图 3 中, 计算节点1任务中的任务数组设置为Array=1-6, 即是计算节点1的任务将分成6个子任务并行运行, 数值为1-6的自然数, 将会选取第1到第6个快照数据进行处理. 每个计算节点处理完快照数据后, 利用1个计算节点完成该晚所有快照图像的拼接, 图 3 是提交拼接任务到计算节点1上运行.

一晚快照图像的拼接处理方法是: 首先, 获取当晚各个快照图像的波束(Beam)参数, 包括波束的半长轴(BMAJ)、半短轴(BMIN)和位置角(BPA), 并计算出这些参数的最大值. 以BMAJ, BMIN和BPA最大值的1.2倍获得一个新的beam. 其次, 当晚所有快照图像与这个新的beam进行卷积, 使得所有快照图像的分辨率相同, 并输出卷积后的快照图像. 卷积通过Miriad³⁾软件实现. 最后, 将所有卷积后的图像拼接起来, 主要采用Swarp软件实现⁴⁾.

5 数据处理结果与分析

数据处理与分析主要分为三类: 第一类是整晚观测天区范围内没有包含A-team源和银道面的延展源即只包含点源; 第二类是整晚观测天区范围内有A-team源, 但没有银道面的延展源; 第三类是整晚观测天区范围内有银道面的延展源. 已经公开的星表数据不包含延展源和A-team源的信息, 因此第二类和第三类的数据校准时, 不能采用该星表制作的天空模型进行校准, 必须使用当晚亮源的校准结果来进行校准, 相对第一类的数据处理较为复杂.

第一类选取了2013年8月10日这晚的观测数据,

图 3 (网络版彩图)一晚观测数据并行处理

Figure 3 (Color online) Data parallel processing for observations of one night.

3) <https://www.atnf.csiro.au/computing/software/miriad/>

4) <https://www.astromatic.net/software/swarp>

该晚的观测天区范围内没有包含A-team源和银道面的延展源. 一共有49个快照观测数据, 其中有3个是校准源的观测, 46个是目标场的观测. 目标场的快照数据校准, 均是使用已发表的星表制作天空模型来校准. 在深度成像的时候, 均去掉短基线数据(30 m以下), 提高点源成像分辨率. 最终拼接的总强度(Stokers I)图像如图4所示, 图像的RMS噪声约为 37 mJy beam^{-1} , 该值与文献[15]给出的结果相符. 该图像使用Aegean软件搜寻出的射电源共有25297个, 其中与已发表的GLEAM星表匹配的源个数为23367, 匹配的源个数占源总数92.4%. 该结果表明, 本文采用的数据处理方法基本正确, 第一类的处理结果与已发表的星表结果基本一致. 这里及以下使用到的星表匹配工具为TOPCAT⁵⁾和STILTS⁶⁾软件.

多出了7.6%源的主要原因在于: (1) 数据处理方法存在差异, 例如数据校准时, 已发表的星表使用CASA⁷⁾的BANDPASS任务对校准源的观测数据进行校准, 然后将校准源的校准结果应用于快照数据的校准. 在流量修正时, 已发表的星表是将快照图像与VLSSr巡天的星表结果进行对比完成的. (2) 大部分数据处理软件已经升级, 数据处理结果得到进一步提高. 例如, 用于射频干扰标记与削减的软件, 相比已发表的星表当时使用的软件, 能够标记与削减掉更多的射频干扰数据. 用于大视场成像的wscclean软件, 本文采用的版本是2.6, 相对之前的版本, 成像质量也有一定的提高.

第二类选取了2013年11月6日这晚的观测数据, 该晚的观测天区内有一个A-team源: Pictor A (RA=05h19m50s, DEC=-45°46'44"). 该晚一共有49个快照观测, 其中有14个是校准源的观测, 35个是目标场的观测. 其中, 校准源的观测有9个是Pictor A, 其中有4个将用于最后的图像拼接. 在数据校准时, 快照视场范围内包含该Pictor A的数据(8个), 均需要使用校准源的观测数据进行校准. 原因在于已发表的星表中没有包含A-team源的信息, 利用第4节步骤(2)制作的天空模型进行校准, 将获得错误的解.

因此, 先使用Pictor A源独立的天空模型对校准源的观测数据进行校准, 然后使用该校准源的校准结果对包含Pictor A源的快照数据进行校准. 其余不包含Pictor A源的快照数据, 则直接使用已发表星表制作的天空模型进行校准.

图5是从快照图像裁剪出Pictor A源部分的图像结果, 其RMS噪声值约为 $19.7 \text{ mJy beam}^{-1}$. 最终拼接的总强度图像如图6所示, 图中小圆圈的位置即为

图4 (网络版彩图)2013年8月10日该晚169频段拼接的总强度图像

Figure 4 (Color online) Mosaic image of total intensity at 169 band based on all the data on 2013-08-10.

图5 从快照图像中裁剪出Pictor A部分的图像结果

Figure 5 Cut-out image of Pictor A from the snapshot image.

5) <http://www.star.bris.ac.uk/mbt/topcat/>

6) <http://www.star.bristol.ac.uk/mbt/stilts/>

7) <http://casa.nrao.edu/>

Pictor A源的位置. 该图像使用Aegean软件搜寻出的射电源共有16026个, 其中与已发表的GLEAM星表匹配的源个数为13852, 匹配的源个数占源总数的86.4%. 图7是图6使用Aegean搜寻到源的分布和其与已发表的星表匹配的源的分布对比, 图中赤经RA在 120° – 140° 之间的源无法与已发表的星表匹配. 除了第一类所述的原因外, 大量多出源的根本原因是已发表的星表没有包含这部分源的信息.

第三类选取2014年6月18日这晚169频段的观测

图6 (网络版彩图)2013年11月6日该晚169频段拼接的总强度图像

Figure 6 (Color online) Mosaic image of total intensity at 169 band based on all the data on 2013-11-06.

图7 (网络版彩图)图6源搜寻结果与已发表的星表匹配结果的对比

Figure 7 (Color online) Comparison of source finding results in Figure 6 with the published catalogue.

数据. 该晚共有41个快照观测, 其中4个是校准源的观测, 其余均为目标场的观测. 该晚的观测天区包含一部分银道面的弥散辐射结构类型的延展源, 这些延展源是GLEAM数据处理的难点之一. 因为延展源的数据处理必须使用短基线部分(低分辨率)的观测数据才能成像, 然而短基线数据将会大大降低该快照范围内点源的成像分辨率. 本文采用的处理方法是: 包含延展源的快照数据, 设置长基线上限, 去掉上限以上较长基线数据, 保留所有短基线数据, 以确保延展源能够成图且点源的成像分辨率不至于太低; 没有延展源的快照数据, 则设置短基线的上限, 去掉较小的基线部分, 以提高点源的成图分辨率. 除此之外, 包含延展源的快照数据需要使用校准源的观测数据完成校准, 原因与第二类相同.

如图8所示, 是其中1个包含银道面延展源的快照数据获得的总强度图像结果, 图中延展源和附近的点源均能得到很好的成图, 图像的RMS噪声约为 48 mJy beam^{-1} . 如图9所示, 是该晚所有快照图像拼接获得的总强度图像. 图像包含的源总数为6344个, 与已发表的星表匹配的个数为4510个, 匹配个数占源总数的71.1%. 如图10所示, 不匹配的部分均分布在延展源及附近区域, 除了第一类描述的原因外, 最主要的原因在于已发表的星表并没有对

图8 2014年6月18日中1个包含银道面延展源的快照总强度图像

Figure 8 A total intensity snapshot image containing the Galactic plane on 2014-06-18.

图 9 (网络版彩图)2014年6月18日该晚169频段拼接的总强度图像

Figure 9 (Color online) Mosaic image of total intensity at 169 band based on all the data on 2014-06-18.

延展源附近的天区进行深入分析.

通过重新处理MWA GLEAM第1年巡天数据, 探测到已发表星表中的绝大多数天体, 表明对致密源、高亮度源、延展源的数据处理结果是可靠的, 在CSRC原型机上开发的并行数据处理流程是有效的, 展示了CSRC原型机已经具备MWA连续谱巡天数据处理的能力.

6 结论

本文主要介绍了上海天文台CSRC原型机软硬件系统, 详细介绍了硬件的具体情况以及硬件资源

图 10 (网络版彩图)图 9源搜寻结果与和发表的星表匹配结果的对比

Figure 10 (Color online) Comparison of source finding results in Figure 9 with the published catalogue.

的管理情况, 重点介绍了MWA GLEAM第1年连续谱巡天数据处理软件的部署情况. 完成了一套MWA GLEAM第1年的数据处理流程, 并开发了相应的并行数据处理流水线, 大大提高了数据处理速度. 该数据流水线获得的数据处理结果与已发表且公开的结果基本一致, 说明该数据处理流水线基本正确, 也从侧面表明原型机满足MWA GLEAM第1年的数据处理要求. 由于采用了改进的数据处理方法和算法软件, 与已经发表的星表相比, 我们的星表包含了更多的源(7.6%–28.9%). 本文的工作能够为将来参与SKA低频巡天数据处理和CSRC软硬件平台的搭建提供有力的参考.

致谢 感谢科廷大学Nick Seymour教授和云南大学孙晓辉研究员在数据处理方面提供的帮助.

参考文献

- Carilli C, Rawlings S. Science with the Square Kilometre Array. Amsterdam: Elsevier B.V., 2004
- Bourke T, Braun R, Fender R, et al. Advancing Astrophysics with the Square Kilometre Array. Dolman Scott Ltd, SKAO, Published online as PoS [AASKA14], 2015
- Dewdney P, Turner W, Braun R, et al. SKA1 System Baseline Design. Technical Report, SKA-TEL-SKO-0000002, Rev 3. 2016
- Wilman R J, Jarvis M J, Mauch T, et al. An infrared-radio simulation of the extragalactic sky: From the Square Kilometre Array to Herschel. *Mon Not R Astron Soc*, 2010, 405: 447–461
- An T. Science opportunities and challenges associated with SKA big data. *Sci China-Phys Mech Astron*, 2019, 62: 989531
- Swarup G, Ananthakrishnan S, Kapahi V K, et al. The Giant Metre-wave Radio Telescope. *Curr Sci*, 1991, 60: 95–105

- 7 Perley R A, Chandler C J, Butler B J, et al. The expanded very large array: A new telescope for new science. *Astrophys J*, 2011, 739: L1
- 8 van Haarlem M P, Wise M W, Gunst A W, et al. LOFAR: The LOw-Frequency ARray. *Astron Astrophys*, 2013, 556: A2
- 9 Parsons A R, Backer D C, Foster G S, et al. The precision array for probing the epoch of re-ionization: Eight station results. *Astron J*, 2010, 139: 1468–1480
- 10 Ellingson S W, Taylor G B, Craig J, et al. The LWA1 Radio Telescope. *IEEE Trans Antennas Propagat*, 2013, 61: 2540–2549
- 11 Tingay S J, Goeke R, Bowman J D, et al. The Murchison widefield array: The square kilometre array precursor at low radio frequencies. *Publ Astron Soc Aust*, 2013, 30: e007
- 12 Lane W M, Cotton W D, van Velzen S, et al. The very large array low-frequency sky survey redux (VLSSr). *Mon Not R Astron Soc*, 2014, 440: 327–338
- 13 Heald G H, Pizzo R F, Orrú E, et al. The LOFAR Multifrequency Snapshot Sky Survey (MSSS). *Astron Astrophys*, 2015, 582: A123
- 14 Intema H T, Jagannathan P, Mooley K P, et al. The GMRT 150 MHz all-sky radio survey. *Astron Astrophys*, 2017, 598: A78
- 15 Hurley-Walker N, Callingham J R, Hancock P J, et al. GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey—I. A low-frequency extragalactic catalogue. *Mon Not R Astron Soc*, 2017, 464: 1146–1167
- 16 An T, Wu X, Hong X. SKA data take centre stage in China. *Nature Astronomy*, 2019, 3: 1030
- 17 Guo S G, Zheng X Y, Mao Y F, et al. Scheme and prospect of the SKA big data transferring (in Chinese). *E-sci Tech Appl*, 2018, 9: 3–13 [郭绍光, 郑小盈, 毛羽丰, 等. SKA海量数据传输的方案及展望. *科研信息化技术与应用*, 2018, 9: 3–13]
- 18 Offringa A R, Wayth R B, Hurley-Walker N, et al. The low-frequency environment of the Murchison widefield array: Radio-frequency interference analysis and mitigation. *Publ Astron Soc Aust*, 2015, 32: e008
- 19 Offringa A R. Compression of interferometric radio-astronomical data. *Astron Astrophys*, 2016, 595: A99
- 20 Offringa A R, Trott C M, Hurley-Walker N, et al. Parametrizing Epoch of Reionization foregrounds: A deep survey of low-frequency point-source spectra with the Murchison Widefield Array. *Mon Not R Astron Soc*, 2016, 458: 1057–1070
- 21 Intema H T, van der Tol S, Cotton W D, et al. Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. *Astron Astrophys*, 2009, 501: 1185–1205
- 22 Offringa A R, McKinley B, Hurley-Walker N, et al. WSCLEAN: An implementation of a fast, generic wide-field imager for radio astronomy. *Mon Not R Astron Soc*, 2014, 444: 606–619
- 23 Hancock P J, Trott C M, Hurley-Walker N. Source Finding in the Era of the SKA (Precursors): AEGEAN 2.0. *Publ Astron Soc Aust*, 2018, 35: e011

Deployment of SKA low frequency imaging system in China SKA Regional Centre

LAO BaoQiang & AN Tao*

Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China

The Square Kilometre Array (SKA) will be the largest synthesized radio telescope in the world. Thanks to its high resolution, ultra-high sensitivity and super fast survey speed, the SKA can explore the Universe with unprecedented details. The continuum survey, being one of the key observational modes of the SKA, establishes the basis of fundamental understandings of the radio sky. SKA pathfinders have been operational for several years and accumulated rich survey data. Chinese astronomers are enthusiastically preparing for the construction of China SKA Regional Centre (CSRC), and have already completed the prototype. The present paper introduces the detailed implementation of the SKA low-frequency data processing system in the CSRC prototype. The continuum survey data observed by the Murchison Widefield Array (MWA), the SKA-low precursor, were used to verify the pipeline deployment and test the correctness. The data processing results are consistent with the published one, suggesting that the parallel data processing developed in the CSRC prototype is a great success. The flexibility and scalability of the prototype allows it to easily expand for future large-scale continuum imaging survey during the commissioning phase of the SKA. The valuable experience gained in these experiments is helpful for future improving the imaging pipeline, and developing the CSRC.

Square Kilometre Array (SKA), Murchison Wide-field Array (MWA), continuum survey, China SKA Regional Centre (CSRC), data processing

PACS: 47.27.-i, 47.27.Eq, 47.27.Nz, 47.40.Ki, 47.85.Gj

doi: [10.1360/SSPMA-2019-0332](https://doi.org/10.1360/SSPMA-2019-0332)