List of Publications

8.5 Journal Articles

A.P. Mechev, A. Plaat, J.B.R. Oonk, H.T. Intema, and H.J.A. Röttgering. "Pipeline Collector: Gathering performance data for distributed astronomical pipelines". In: *Astronomy and Computing* 24 (2018), pp. 117–128. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2018.06.005. URL: http://www.sciencedirect.com/science/article/pii/S2213133718300490

A.P. Mechev, T.W. Shimwell, A. Plaat, H.T. Intema, A.L. Varbanescu, and H.J.A. Röttgering. "Scalability model for the LOFAR direction independent pipeline". In: *Astronomy and Computing* 28 (2019), p. 100293. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2019.100293. URL: http://www.sciencedirect.com/science/article/pii/S2213133719300290

Alexandar P Mechev, Aske Plaat, Huib Intema, and Huub Rottgering. "Automated testing and quality control of LOFAR scienctific pipelines with AGLOW". in: *Astronomy and Computing* (2019 in review)

JBR Oonk, **AP Mechev**, N Danezi, F Sweijen, T Shimwell, C Schrijvers, A Drabent, and K Emig. "Radio astronomical reduction on distributed and shared processing infrastructures: a platform for LOFAR". in: *Astronomy and Computing* (2019 in prep)

T. W. Shimwell, Tasse, C., Hardcastle, M. J., **Mechev, A. P.**, Williams, W. L., Best, P. N., Röttgering, H. J. A., Callingham, J. R., Dijkema, T. J., de Gasperin, F., Hoang, D. N., Hugo, B., Mirmont, M., Oonk, J. B. R., Prandoni, I., Rafferty, D., Sabater, J., Smirnov, O., van Weeren, R. J., and White, G. J. et al. "The LOFAR Two-metre Sky Survey - II. First data release". In: *A&A* 622 (2019), A1. DOI: 10.1051/0004-6361/201833559. URL: https://doi.org/10.1051/0004-6361/201833559

K. L. Emig, P. Salas, F. de Gasperin, J. B. R. Oonk, M. C. Toribio, **A. P. Mechev**, H. J. A. Röttgering, and A. G. G. M. Tielens. "Searching for the largest bound atoms in space". In: *A&A* (2019)

8.6 Conference Proceedings

Y. G. Grange, R. Lakhoo, M. Petschow, C. Wu, B. Veenboer, I. Emsley, T. J. Dijkema, **A. P. Mechev**, and G. Mariani. "Characterising radio telescope software with the Workload Characterisation Framework". In: *arXiv e-prints*, arXiv:1612.00456 (Dec. 2016), arXiv:1612.00456. arXiv: 1612.00456 [astro-ph.IM]

A.P. Mechev, J.B.R. Oonk, A. Danezi, T.W. Shimwell, C. Schrijvers, H.T. Intema, A. Plaat, and H.J.A. Röttgering. "An Automated Scalable Framework for Distributing Radio Astronomy Processing Across Clusters and Clouds". In: *Proceedings of the International Symposium on Grids and Clouds (ISGC) 2017, held 5-10 March, 2017 at Academia Sinica, Taipei, Taiwan (ISGC2017). Online at https://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=293, id.2.* Mar. 2017, p. 2. arXiv: 1712.00312 [astro-ph.IM]

A.P. Mechev, J.B.R. Oonk, T.W. Shimwell, A. Plaat, H.T. Intema, and H.J.A. Röttgering. "Fast and Reproducible LOFAR Workflows with AGLOW". in: *2018 IEEE 14th International Conference on e-Science (e-Science)*. Oct. 2018, arXiv:1808.10735. DOI: 10.1109/eScience.2018.00029. arXiv: 1808.10735 [astro-ph.IM]

A.P. Mechev, J.B.R. Oonk, A. Plaat, N. Danezi, and T.W. Shimwell. "Building LOFAR as a Service". In: Astronomical Data Analysis Software and Systems XXVIII, p. 677

Glossary

AGLOW AGLOW is a combination of Apache Airflow with GRID_LRT. This integration allows LOFAR users to build and launch massively parallel workflows. 84

CouchDB CouchDB is a document-based eventually consistent database, that we use to store processing information for distributed jobs. Each CouchDB document corresponds to a single distributed job, and contains a full description of the job required to run on a worker node. As jobs run, they update their status in the CouchDB document, which can be accessed by users through their browsers or a Python client. 24, 45, 91

CVMFS The CERN VM Filesystem is a virtually mounted filesystem that is used to distribute software on multiple clusters, cluster nodes and individual machines. CVMFS allows an institute to host a portable installation of their software, which is distributed and cached by other CVMFS clients. The software is cached locally on the worker nodes as a FileSystem in Userspace (FUSE) module . 26, 54, 67, 109

dCache dCache is a system for storing and retrieving large amounts of data, distributed across heterogenous servers. dCache provides a common virtual filesystem, while also allows data to be located on varied storage devices including SSDs, spinning disks and magnetic tape . 152

Grid Grid computing refers to massively parallel distributed computing introduced in the '90s to tackle the processing challenges processing data from the Large Hadron Collider. A computational grid is a set of compute nodes connected with a high throughput connection, common job scheduler and shared, distributed storage. The computational and storage resources in a Grid are federated, and users are provided a share of those resources by a managing authority . 3, 18, 45, 61, 84, 104, 149

GRID_LRT GRID_LRT is the GRID LOFAR Reduction Tools package. This software consists of a set of tools to easily create and launch processing jobs on a distributed infrastructure. It includes tools to manage LOFAR data stored on the grid filesystem. These tools make it possible to quickly integrate processing scripts with a high throughput environment, accelerating bottleneck steps in LOFAR processing . 18

HBA The High Band Array is an array of LOFAR antennas sensitive to 110-240MHz. Each lofar station in the Netherlands has 48 HBA elements, with core stations having two separate sub-arrays of 24; while international stations have 96 HBA elements. The naming schele of these antenasis as such: LLNNNHBAS, where LL is the location (CS for core stations, RS for remote stations, and the ISO 3166-1 2-letter country code for the international stations); NNN is the station number, starting at 000, HBA/LBA denotes whether it's a HBA or LBA antenna and S refers to the core station sub-array. 6, 19

HPC High Performance Computing, as opposed to High Throughput Computing (HTC), refers to computation on one or multiple machines with many, fast CPUS; large quantities of RAM and fast disks. Often HPC jobs are done on clusters where multiple of these machines are connected with a fast network used to pass messages and to synchronize workloads. 20

HTC High Throughput Computing, as opposed to High Performance Computing (HPC), focuses on minimizing the time taken processing large amounts of data by using techniques in streaming, parallelizing and distributing many small jobs on a cluster . 20, 60, 84, 135

LOFAR The LOw Frequency ARray: A large, low-frequency aperture synthesis radio telescope. 6, 17, 44, 60, 84, 102, 134

LoTSS The LOFAR Two-Meter Sky Survey is a whole-sky study of the low-frequency radio sky at 120-168MHz. LoTSS is composed of a broad, Tier 1, survey of the entire sky, as well as deeper tiers targeted at specific fields of interest . 7, 18, 43, 44, 105, 136, 149

srm Storage Resource Manager, a system to co-ordinate data storage. This system defines the protocols for referencing data on a distributed system, accessing metadata and changing data locality (e.g., moving from tape to disk). . 20, 50, 92

Subband Broadband LOFAR observations are stored in separate 'Subbands', splitting the frequency range into several individual files, before storing at the Long Term Archive. Depending on the observation mode, one observation can have 230-480 Subbands. This splitting makes it easier for users to request, download and process a fraction of observation's entire bandwidth. 8, 31, 47, 85, 104

visibilities Radio Astronomers use the term 'UV plane' or 'visibilities' interchangably to refer to the Fourier Transform of the final image. Each baseline of an aperture synthesis telescope corresponds to one measurement in this space. The letters U and V refer to the two orthogonal components of a baseline with respect to an observation's phase center. The u and v vectors are defined in a plane orthogonal to the direction towards the phase center, and are typically in units of wavelength. To obtain an image, the UV data needs to be 'cleaned' by iteratively removing the point spread function of the telescope. 5, 20, 107

VOMS The Virtual Organization Membership Service is a service that manages the access to data and computational resources provided to each Grid user. It handles authentication and authorisation of job launching and access to data on the grid filesystem. More information can be found at `https://italiangrid.github.io/voms/` . 24

Bibliography

[1] John von Neumann. *First Draft of a Report on the EDVAC*. Tech. rep. 1945.

[2] William W. Wood. *Early history of computer simulations in statistical mechanics*. Tech. rep. Los Alamos National Lab., NM (USA); Carroll Coll., Helena, MT (USA), 1985.

[3] Edward N Lorenz. "Empirical orthogonal functions and statistical weather prediction". In: (1956).

[4] Francis H Harlow and J Eddie Welch. "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface". In: *The physics of fluids* 8.12 (1965), pp. 2182–2189.

[5] Frances E. Allen. "The history of language processor technology in IBM". In: *IBM Journal of Research and Development* 25.5 (1981), pp. 535–548.

[6] Christopher Bingham, M Godfrey, and J Tukey. "Modern techniques of power spectrum estimation". In: *IEEE Transactions on audio and electroacoustics* 15.2 (1967), pp. 56–66.

[7] Jack Dongarra and Francis Sullivan. "Guest Editors' Introduction: The Top 10 Algorithms". In: *Computing in Science & Engineering* 2.1 (2000), pp. 22–23. DOI: 10.1109/MCISE.2000.814652. eprint: https://aip.scitation.org/doi/pdf/10.1109/MCISE.2000.814652. URL: https://aip.scitation.org/doi/abs/10.1109/MCISE.2000.814652.

[8] Gerard Tel. *Introduction to Distributed Algorithms*. 2nd ed. Cambridge University Press, 2000. DOI: 10.1017/CBO9781139168724.

[9] E. Fomalont. "Astronomical Image Processing System / AIPS". In: *National Radio Astronomy Observatory Newsletter* 3 (1981), p. 3.

[10] Doug Tody. "The IRAF data reduction and analysis system". In: *Instrumentation in astronomy VI*. Vol. 627. International Society for Optics and Photonics. 1986, pp. 733–748.

[11] Donald Carson Wells and Eric W Greisen. "FITS-a flexible image transport system". In: *Image Processing in Astronomy*. 1979, p. 445.

[12] Shang-Wen Cheng et al. "Software architecture-based adaptation for grid computing". In: (1989).

[13] BG Clark. "An efficient implementation of the algorithm'CLEAN'". In: *Astronomy and Astrophysics* 89 (1980), p. 377.

[14] Anthony Richard Thompson, James M Moran, George Warner Swenson, et al. *Interferometry and synthesis in radio astronomy*. Wiley New York et al., 1986.

[15] Wilbur Norman Christiansen and Jan A Högbom. *Radiotelescopes*. CUP Archive, 1987.

[16] F. de Gasperin et al. "Systematic effects in LOFAR data: A unified calibration strategy". In: *Astronomy and Astrophysics* 622, A5 (Feb. 2019), A5. DOI: 10.1051/0004-6361/201833867. arXiv: 1811.07954 [astro-ph.IM].

[17] WL Williams et al. "LOFAR 150-MHz observations of the Boötes field: catalogue and source counts". In: *Monthly Notices of the Royal Astronomical Society* 460.3 (2016), pp. 2385–2412.

[18] R. J. van Weeren et al. "LOFAR Facet Calibration". In: *ApJS* 223.1, 2 (Mar. 2016), p. 2. DOI: 10.3847/0067-0049/223/1/2. arXiv: 1601.05422 [astro-ph.IM].

[19] C. Tasse et al. "Faceting for direction-dependent spectral deconvolution". In: *A&A* 611, A87 (Apr. 2018), A87. DOI: 10.1051/0004-6361/201731474. arXiv: 1712.02078 [astro-ph.IM].

[20] M. P. van Haarlem et al. "LOFAR: The LOw-Frequency ARray". In: *A&A* 556, A2 (Aug. 2013), A2. DOI: 10.1051/0004-6361/201220873. arXiv: 1305.3550 [astro-ph.IM].

[21] *LOFAR Stations: Description and Layout*. URL: https://old.astron.nl/radio-observatory/astronomers/%20users/technical-information/lofar-stations/lofar-stations-description-.

[22] ASTRON. *LOFAR Brochure*. Apr. 2019. URL: https://www.astron.nl/sites/default/files/shared/LOFAR_Brochure_ENG_April2019.pdf.

[23] AH Patil et al. "Upper limits on the 21 cm epoch of reionization power spectrum from one night with LOFAR". In: *The Astrophysical Journal* 838.1 (2017), p. 65.

[24] T. W. Shimwell et al. "The LOFAR Two-metre Sky Survey - II. First data release". In: *arXiv e-prints*, arXiv:1811.07926 (Nov. 2018), arXiv:1811.07926. arXiv: 1811.07926 [astro-ph.GA].

[25] TW Shimwell et al. "The LOFAR Two-metre Sky Survey-I. Survey description and preliminary data release". In: *Astronomy & Astrophysics* 598 (2017), A104.

[26] Heald, G. H. et al. "The LOFAR Multifrequency Snapshot Sky Survey (MSSS) - I. Survey description and first results". In: *A&A* 582 (2015), A123. DOI: 10.1051/0004-6361/201425210. URL: https://doi.org/10.1051/0004-6361/201425210.

[27] J. B. R. Oonk et al. "Carbon and hydrogen radio recombination lines from the cold clouds towards Cassiopeia A". In: *MNRAS* 465.1 (Feb. 2017), pp. 1066–1088. DOI: 10.1093/mnras/stw2818. arXiv: 1609.06857 [astro-ph.GA].

[28] K. L. Emig et al. "The first detection of radio recombination lines at cosmological distances". In: *A&A* 622, A7 (Feb. 2019), A7. DOI: 10.1051/0004-6361/201834052. arXiv: 1811.08104 [astro-ph.GA].

[29] M. Arias et al. "Low-frequency radio absorption in Cassiopeia A". In: *A&A* 612, A110 (Apr. 2018), A110. DOI: 10.1051/0004-6361/201732411. arXiv: 1801.04887 [astro-ph.HE].

[30] P. Salas et al. "Mapping low-frequency carbon radio recombination lines towards Cassiopeia A at 340, 148, 54, and 43 MHz". In: *MNRAS* 475.2 (Apr. 2018), pp. 2496–2511. DOI: 10.1093/mnras/stx3340. arXiv: 1801.05298 [astro-ph.GA].

[31] M. E. Bell et al. "An automated archival Very Large Array transients survey". In: *MNRAS* 415 (July 2011), pp. 2–10. DOI: 10.1111/j.1365-2966.2011.18631.x. arXiv: 1103.0511 [astro-ph.HE].

[32] B. W. Stappers et al. "Observing pulsars and fast transients with LOFAR". In: *A&A* 530, A80 (June 2011), A80. DOI: 10.1051/0004-6361/201116681. arXiv: 1104.1577 [astro-ph.IM].

[33] G.N.J. van Diepen. "Casacore Table Data System and its use in the MeasurementSet". In: *Astronomy and Computing* 12 (2015), pp. 174–180. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2015.06.002. URL: http://www.sciencedirect.com/science/article/pii/S2213133715000621.

[34] M. J. Hardcastle et al. "LOFAR/H-ATLAS: a deep low-frequency survey of the Herschel-ATLAS North Galactic Pole field". In: *MNRAS* 462.2 (Oct. 2016), pp. 1910–1936. DOI: 10.1093/mnras/stw1763. arXiv: 1606.09437 [astro-ph.GA].

[35] **A.P. Mechev** et al. "An Automated Scalable Framework for Distributing Radio Astronomy Processing Across Clusters and Clouds". In: *Proceedings of the International Symposium on Grids and Clouds (ISGC) 2017, held 5-10 March, 2017 at Academia Sinica, Taipei, Taiwan (ISGC2017). Online at https://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=293, id.2*. Mar. 2017, p. 2. arXiv: 1712.00312 [astro-ph.IM].

[36] **A.P. Mechev** et al. "Pipeline Collector: Gathering performance data for distributed astronomical pipelines". In: *Astronomy and Computing* 24 (2018), pp. 117–128. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2018.06.005. URL: http://www.sciencedirect.com/science/article/pii/S2213133718300490.

[37] **A.P. Mechev** et al. "Fast and Reproducible LOFAR Workflows with AGLOW". In: *2018 IEEE 14th International Conference on e-Science (e-Science)*. Oct. 2018, arXiv:1808.10735. DOI: 10.1109/eScience.2018.00029. arXiv: 1808.10735 [astro-ph.IM].

[38] **A.P. Mechev** et al. "Scalability model for the LOFAR direction independent pipeline". In: *Astronomy and Computing* 28 (2019), p. 100293. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2019.100293. URL: http://www.sciencedirect.com/science/article/pii/S2213133719300290.

[39] C.L. Carilli and S. Rawlings. "Motivation, key science projects, standards and assumptions". In: *New Astronomy Reviews* 48.11 (2004). Science with the Square Kilometre Array, pp. 979–984. ISSN: 1387-6473. DOI: https://doi.org/10.1016/j.newar.2004.09.001. URL: http://www.sciencedirect.com/science/article/pii/S1387647304000880.

[40] Andreas Horneffer et al. *apmechev/prefactor: LOTSS Data Release 1*. Nov. 2018. DOI: 10.5281/zenodo.1487962. URL: https://doi.org/10.5281/zenodo.1487962.

[41] T. W. Shimwell et al. "The LOFAR Two-metre Sky Survey. I. Survey description and preliminary data release". In: *A&A* 598, A104 (Feb. 2017), A104. DOI: 10.1051/0004-6361/201629313. arXiv: 1611.02700 [astro-ph.IM].

[42] A. Wilber et al. "LOFAR discovery of an ultra-steep radio halo and giant head-tail radio galaxy in Abell 1132". In: *MNRAS* 473.3 (Jan. 2018), pp. 3536–3546. DOI: 10.1093/mnras/stx2568. arXiv: 1708.08928 [astro-ph.GA].

[43] H. J. A. Rottgering et al. "LOFAR - Opening up a new window on the Universe". In: *arXiv e-prints*, astro-ph/0610596 (Oct. 2006), astro–ph/0610596. arXiv: astro-ph/0610596 [astro-ph].

[44] Offringa, A. R. et al. "The LOFAR radio environment". In: *A&A* 549 (2013), A11. DOI: 10.1051/0004-6361/201220293. URL: https://doi.org/10.1051/0004-6361/201220293.

[45] Tammo Jan Dijkema. "RFI Flagging, Demixing and Visibilities Compression". In: *Astrophysics and Space Science Library*. Vol. 426. Astrophysics and Space Science Library. Jan. 2018, p. 55. DOI: 10.1007/978-3-319-23434-2_4.

[46] Jamie Shiers. "The Worldwide LHC Computing Grid (worldwide LCG)". In: *Computer Physics Communications* 177.1 (2007). Proceedings of the Conference on Computational Physics 2006, pp. 219–223. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2007.02.021. URL: http://www.sciencedirect.com/science/article/pii/S001046550700077X.

[47] S. . van der Tol, B. D. Jeffs, and A. -J. . van der Veen. "Self-Calibration for the LOFAR Radio Astronomical Array". In: *IEEE Transactions on Signal Processing* 55.9 (Sept. 2007), pp. 4497–4510. DOI: 10.1109/TSP.2007.896243.

[48] Hanno Holties, Adriaan Renting, and Yan Grange. "The LOFAR long-term archive: e-infrastructure on petabyte scale". In: *SPIE Astronomical Telescopes+ Instrumentation*. International Society for Optics and Photonics. 2012, pp. 845117–845117.

[49] Domingos Barbosa et al. "Power Monitoring and Control for Large Scale projects: SKA, a case study". In: *SPIE Astronomical Telescopes+ Instrumentation*. International Society for Optics and Photonics. 2016, pp. 99100L–99100L.

[50] JBR Oonk et al. "Radio astronomical reduction on distributed and shared processing infrastructures: a platform for LOFAR". In: *Astronomy and Computing* (2019 in prep).

[51] Shayan Shams et al. "A Scalable Pipeline For Transcriptome Profiling Tasks With On-demand Computing Clouds". In: *Parallel and Distributed Processing Symposium Workshops, 2016 IEEE International*. IEEE. 2016, pp. 443–452.

[52] Anjani Ragothaman et al. "Developing ethread pipeline using saga-pilot abstraction for large-scale structural bioinformatics". In: *BioMed research international* 2014 (2014).

[53] JD Van Horn et al. *Grid-Based Computing and the Future of Neuroscience Computation, Methods in Mind*. 2005.

[54] William K Michener and Matthew B Jones. "Ecoinformatics: supporting ecology as a data-intensive science". In: *Trends in ecology & evolution* 27.2 (2012), pp. 85–93.

[55] Ewa Deelman et al. "Pegasus: A framework for mapping complex scientific workflows onto distributed systems". In: *Scientific Programming* 13.3 (2005), pp. 219–237.

[56] Yong Zhao et al. “Swift: Fast, reliable, loosely coupled parallel computation”. In: *Services, 2007 IEEE Congress on*. IEEE. 2007, pp. 199–206.

[57] *PiCaS Overview - Grid Documentation v1.0*. `http://doc.grid.surfsara.nl/en/latest/Pages/Practices/picas/picas_overview.html`. 2017.

[58] C Aguado Sanchez et al. “CVMFS-a file system for the CernVM virtual appliance”. In: *Proceedings of XII Advanced Computing and Analysis Techniques in Physics Research*. Vol. 1. 2008, p. 52.

[59] Jan Bot. *PiCaS: Python client using CouchDB as a token pool server*. `https://github.com/jjbot/picasclient`. 2017.

[60] Chirs ANDERSON. “Apache couchdb: The definitive guide”. In: *http://couchdb. Apache. org/index. htm Acessado em* 5.06 (2009), p. 2009.

[61] Joris Borgdorff, Harsha Krishna, and Michael H Lees. “SIM-CITY: An e-Science framework for Urban Assisted Decision Support”. In: *Procedia Computer Science* 51 (2015), pp. 2327–2336.

[62] Y Li. “Reliability of long heterogeneous slopes in 3d: Model performance and conditional simulation”. In: (2017).

[63] Marco Clemencic and B Couturier. “A New Nightly Build System for LHCb”. In: *Journal of Physics: Conference Series*. Vol. 513. 5. IOP Publishing. 2014, p. 052007.

[64] Tom SANTE. “Development of (graphical) web applications for the processing and interpretation of arrayCGH data”. In: (2010).

[65] Grigory Rybkin. “ATLAS software packaging”. In: *Journal of Physics: Conference Series*. Vol. 396. 5. IOP Publishing. 2012, p. 052063.

[66] GS Davies et al. “Software Management for the NOνAExperiment”. In: *Journal of Physics: Conference Series*. Vol. 664. 6. IOP Publishing. 2015, p. 062011.

[67] WN Brouw. “Aperture synthesis”. In: *Image Processing Techniques in Astronomy*. Springer, 1975, pp. 301–307.

[68] Cyril Tasse et al. “LOFAR calibration and wide-field imaging”. In: *Comptes Rendus Physique* 13.1 (2012), pp. 28–32.

[69] RF Pizzo et al. “The Lofar Imaging Cookbook v2.0”. In: *internal ASTRON report* (2010).

[70] *Gina specifications - GRID Documentation v1.0*. Available at `http://docs.surfsaralabs.nl/projects/grid/en/latest/Pages/Service/system_specifications/gina_specs.html`. 2017.

[71] *Lofar Software Stack*. http://www.lofar.org/wiki/doku.php?id=public:software_stack_installation. 2017.

[72] *Softdrive on the GRID*. Available at http://docs.surfsaralabs.nl/projects/grid/en/latest/Pages/Advanced/grid_software.html#softdrive.

[73] Jason Maassen et al. *Xenon: Xenon 1.1.0*. Dec. 2015. DOI: 10.5281/zenodo.35415. URL: https://doi.org/10.5281/zenodo.35415.

[74] Software Foundation Apache. *TCollector: OpenTSDB documentation*. Available at http://opentsdb.net/docs/build/html/user_guide/utilities/tcollector.html. 2017.

[75] P. Chris Broekema et al. "Cobalt: A GPU-based correlator and beamformer for LOFAR". In: *Astronomy and Computing* 23 (2018), pp. 180–192. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2018.04.006. URL: http://www.sciencedirect.com/science/article/pii/S2213133717301439.

[76] Y Gupta et al. "The upgraded GMRT: opening new windows on the radio Universe". In: *Current Science* 113.4 (2017), p. 707.

[77] MP Van Haarlem et al. "LOFAR: The low-frequency array". In: *Astronomy & astrophysics* 556 (2013), A2.

[78] Colin J Lonsdale et al. "The murchison widefield array: Design overview". In: *Proceedings of the IEEE* 97.8 (2009), pp. 1497–1506.

[79] S J Tingay et al. "The Murchison widefield array: The square kilometre array precursor at low radio frequencies". In: *Publications of the Astronomical Society of Australia* 30 (2013).

[80] Justin L Jonas. "MeerKAT-The South African array with composite dishes and wide-band single pixel feeds". In: *Proceedings of the IEEE* 97.8 (2009), pp. 1522–1530.

[81] Chen Wu et al. "Optimising NGAS for the MWA Archive". In: *Experimental Astronomy* 36.3 (2013), pp. 679–694.

[82] David B Davidson. "Potential technological spin-offs from MeerKAT and the South African Square Kilometre Array bid". In: *South African Journal of Science* 108.1-2 (2012), pp. 01–03.

[83] OM Smirnov and Cyril Tasse. "Radio interferometric gain calibration as a complex optimization problem". In: *Monthly Notices of the Royal Astronomical Society* 449.3 (2015), pp. 2668–2684.

[84] J. B. R. Oonk et al. “Discovery of carbon radio recombination lines in absorption towards Cygnus A”. In: *MNRAS* 437 (Feb. 2014), pp. 3506–3515. DOI: 10.1093/mnras/stt2158. arXiv: 1401.2876.

[85] SURF. *Grid at SURFsara*. https://www.surf.nl/en/services-and-products/grid/index.html. 2018.

[86] R Centeno et al. “The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: optimization of the spectral line inversion code”. In: *Solar Physics* 289.9 (2014), pp. 3531–3547.

[87] Stephen Strother et al. “Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis”. In: *Neuroimage* 23 (2004), S196–S207.

[88] S. Ott. “The Herschel Data Processing System — HIPE and Pipelines — Up and Running Since the Start of the Mission”. In: *Astronomical Data Analysis Software and Systems XIX*. Vol. 434. Dec. 2010, p. 139.

[89] Jens-S Vöckler et al. “Kickstarting remote applications”. In: *2nd International Workshop on Grid Computing Environments*. 2006, pp. 1–8.

[90] Matthew L Massie, Brent N Chun, and David E Culler. “The ganglia distributed monitoring system: design, implementation, and experience”. In: *Parallel Computing* 30.7 (2004), pp. 817–840.

[91] Terrehon Bowden. *THE /proc FILESYSTEM v1.3*. https://www.kernel.org/doc/Documentation/filesystems/proc.txt. 2009.

[92] Philip J Mucci et al. “PAPI: A portable interface to hardware performance counters”. In: *Proceedings of the department of defense HPCMP users group conference*. Vol. 710. 1999.

[93] Brendan Gregg and Jim Mauro. *DTrace: Dynamic Tracing in Oracle Solaris, Mac OS X and FreeBSD*. Prentice Hall Professional, 2011.

[94] Nicholas Nethercote and Julian Seward. “Valgrind: a framework for heavyweight dynamic binary instrumentation”. In: *ACM Sigplan notices*. Vol. 42. 6. ACM. 2007, pp. 89–100.

[95] *LOFAR Imaging Cookbook*. Available at http://www.astron.nl/sites/astron.nl/files/cms/lofar_imaging_cookbook_v19.pdf.

[96] B Sigoure. *OpenTSDB scalable time series database (TSDB)*. 2012.

[97] Dong H Ahn. *Measuring FLOPS using hardware performance counter technologies on LC systems*. Tech. rep. Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 2008.

[98] GM Loose. "LOFAR self-calibration using a blackboard software architecture". In: *Astronomical Data Analysis Software and Systems XVII*. Vol. 394. 2008, p. 91.

[99] Ger van Diepen and Tammo Jan Dijkema. *DPPP: Default Pre-Processing Pipeline*. Astrophysics Source Code Library. Apr. 2018. ascl: 1804.003.

[100] Jakob Blomer et al. "Distributing LHC application software and conditions databases using the CernVM file system". In: *Journal of Physics: Conference Series*. Vol. 331. 4. IOP Publishing. 2011, p. 042003.

[101] Randy H. Katz and David A. Patterson. *Memory Hierarchy, CMPUT429/CMPE382 Winter 2001. University of Calgary*. Available at https://webdocs.cs.ualberta.ca/~amaral/courses/429/webslides/Topic4-MemoryHierarchy/sld003.htm. 2001.

[102] Peter J Denning. "The working set model for program behavior". In: *Communications of the ACM* 11.5 (1968), pp. 323–333.

[103] Kim Hazelwood and James E Smith. "Exploring code cache eviction granularities in dynamic optimization systems". In: *Code Generation and Optimization, 2004. CGO 2004. International Symposium on*. IEEE. 2004, pp. 89–99.

[104] Kazushige Goto and Robert A Geijn. "Anatomy of high-performance matrix multiplication". In: *ACM Transactions on Mathematical Software (TOMS)* 34.3 (2008), p. 12.

[105] Stefano Salvini and Stefan J Wijnholds. "StEFCal-An Alternating Direction Implicit method for fast full polarization array calibration". In: *General Assembly and Scientific Symposium (URSI GASS), 2014 XXXIth URSI*. IEEE. 2014, pp. 1–4.

[106] Kevin Skadron et al. "Branch prediction, instruction-window size, and cache size: Performance tradeoffs and simulation techniques". In: *IEEE Transactions on Computers* 48.11 (1999), pp. 1260–1281.

[107] Tarush Jain and Tanmay Agrawal. "The haswell microarchitecture-4th generation processor". In: *International Journal of Computer Science and Information Technologies* 4.3 (2013), pp. 477–480.

[108] Subhradyuti Sarkar and Dean Tullsen. "Compiler techniques for reducing data cache miss rate on a multithreaded architecture". In: *High Performance Embedded Architectures and Compilers* (2008), pp. 353–368.

[109] Alvin R Lebeck et al. "A large, fast instruction window for tolerating cache misses". In: *Computer Architecture, 2002. Proceedings. 29th Annual International Symposium on*. IEEE. 2002, pp. 59–70.

[110] Shiliang Hu et al. “An approach for implementing efficient superscalar CISC processors”. In: *High-Performance Computer Architecture, 2006. The Twelfth International Symposium on*. IEEE. 2006, pp. 41–52.

[111] P Chris Broekema, Rob V van Nieuwpoort, and Henri E Bal. “The Square Kilometre Array science data processor. Preliminary compute platform design”. In: *Journal of Instrumentation* 10.07 (2015), p. C07004.

[112] A Patterson David and L Hennessy John. “Computer organization and design: the hardware/software interface”. In: *San mateo, CA: M organ Kaufmann Publishers* 1 (2005), p. 998.

[113] Team Apache HBase. “Apache hbase reference guide”. In: *Apache, version* 2.0 (2015).

[114] J Sabater et al. “Calibration of radio-astronomical data on the cloud. LOFAR, the pathway to SKA.” In: *Highlights of Spanish Astrophysics VIII*. 2015, pp. 840–843.

[115] Jeff Templon and Jan Bot. “The Dutch National e-Infrastructure”. In: *International Symposium on Grids and Clouds (ISGC)*. Vol. 13. 18. 2016.

[116] Erwin Laure et al. *Middleware for the next generation Grid infrastructure*. Tech. rep. CERN, 2004.

[117] Jamie Shiers. “The worldwide LHC computing grid (worldwide LCG)”. In: *Computer physics communications* 177.1-2 (2007), pp. 219–223.

[118] Ilkay Altintas et al. “Kepler: an extensible system for design and execution of scientific workflows”. In: *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on*. IEEE. 2004, pp. 423–424.

[119] David Churches et al. “Programming scientific and distributed workflow with Triana services”. In: *Concurrency and Computation: Practice and Experience* 18.10 (2006), pp. 1021–1037.

[120] Ji Liu et al. “A survey of data-intensive scientific workflow management”. In: *Journal of Grid Computing* 13.4 (2015), pp. 457–493.

[121] Lin Wang. “Directed acyclic graph”. In: *Encyclopedia of Systems Biology*. Springer, 2013, pp. 574–574.

[122] David J Pearce and Paul HJ Kelly. “A dynamic topological sort algorithm for directed acyclic graphs”. In: *Journal of Experimental Algorithmics (JEA)* 11 (2007), pp. 1–7.

[123] A. B. Kahn. “Topological Sorting of Large Networks”. In: *Commun. ACM* 5.11 (Nov. 1962), pp. 558–562. ISSN: 0001-0782. DOI: 10.1145/368996.369025. URL: http://doi.acm.org/10.1145/368996.369025.

[124] Jianjun Zhou and Martin Müller. "Depth-first discovery algorithm for incremental topological sorting of directed acyclic graphs". In: *Information Processing Letters* 88.4 (2003), pp. 195–200.

[125] John Vivian et al. "Toil enables reproducible, open source, big biomedical data analyses". In: *Nature biotechnology* 35.4 (2017), p. 314.

[126] Paolo Di Tommaso et al. "Nextflow enables reproducible computational workflows". In: *Nature biotechnology* 35.4 (2017), pp. 316–319.

[127] W. Freudling et al. "Automated data reduction workflows for astronomy. The ESO Reflex environment". In: *Astronomy & Astrophysics* 559, A96 (Nov. 2013), A96. DOI: 10.1051/0004-6361/201322494. arXiv: 1311.5411 [astro-ph.IM].

[128] Peter Amstutz et al. "Common Workflow Language, v1. 0". In: (2016).

[129] Michael Kotliar, Andrey Kartashov, and Artem Barski. "CWL-Airflow: a lightweight pipeline manager supporting Common Workflow Language". In: *bioRxiv* (2018), p. 249243.

[130] Patrick Fuhrmann and Volker Gülzow. "dCache, storage system for the future". In: *European Conference on Parallel Processing*. Springer. 2006, pp. 1106–1113.

[131] Linus Torvalds and Junio Hamano. "Git: Fast version control system". In: *URL http://git-scm. com* (2010).

[132] A. R. Offringa et al. "WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy". In: *Monthly Notices of the Royal Astronomical Society* 444 (Oct. 2014), pp. 606–619. DOI: 10.1093/mnras/stu1368. arXiv: 1407.1943 [astro-ph.IM].

[133] A. R. Offringa, J. J. van de Gronde, and J. B. T. M. Roerdink. "A morphological algorithm for improving radio-frequency interference detection". In: *Astronomy & astrophysics* 539, A95 (Mar. 2012), A95. DOI: 10.1051/0004-6361/201118497. arXiv: 1201.3364 [astro-ph.IM].

[134] JP McMullin et al. "CASA architecture and applications". In: *Astronomical data analysis software and systems XVI*. Vol. 376. 2007, p. 127.

[135] Niruj Mohan and David Rafferty. "PyBDSF: Python Blob Detection and Source Finder". In: *Astrophysics Source Code Library* (2015).

[136] Cyril Tasse. "Nonlinear Kalman filters for calibration in radio interferometry". In: *Astronomy & Astrophysics* 566 (2014), A127.

[137] P. Salas et al. "LOFAR observations of decameter carbon radio recombination lines towards Cassiopeia A". In: *Monthly Notices of the Royal Astronomical Society* 467 (May 2017), pp. 2274–2287. DOI: 10.1093/mnras/stx239. arXiv: 1701.08802.

[138] J.B.R. Oonk et al. “Radio astronomy on a distributed shared computing platform: The LOFAR case”. In: (2018).

[139] Peter E Dewdney et al. “The square kilometre array”. In: *Proceedings of the IEEE* 97.8 (2009), pp. 1482–1496.

[140] Jeff Templon and Jan Bot. “The Dutch National e-Infrastructure”. To appear in Proceedings of Science edition of the International Symposium on Grids and Clouds (ISGC) 2016 13-18 March 2016, Academia Sinica, Taipei, Taiwan. Oct. 2016. URL: https://doi.org/10.5281/zenodo.163537.

[141] H.A. Sanjay and Sathish Vadhiyar. “Performance modeling of parallel applications for grid scheduling”. In: *Journal of Parallel and Distributed Computing* 68.8 (2008), pp. 1135–1145. ISSN: 0743-7315. DOI: https://doi.org/10.1016/j.jpdc.2008.02.006. URL: http://www.sciencedirect.com/science/article/pii/S0743731508000464.

[142] Zhichen Xu, Xiaodong Zhang, and Lin Sun. “Semi-empirical Multiprocessor Performance Predictions”. In: *Journal of Parallel and Distributed Computing* 39.1 (1996), pp. 14–28. ISSN: 0743-7315. DOI: https://doi.org/10.1006/jpdc.1996.0151. URL: http://www.sciencedirect.com/science/article/pii/S0743731596901513.

[143] Bradley J Barnes et al. “A regression-based approach to scalability prediction”. In: *Proceedings of the 22nd annual international conference on Supercomputing*. ACM. 2008, pp. 368–377.

[144] Michael Kuperberg, Klaus Krogmann, and Ralf Reussner. “Performance prediction for black-box components using reengineered parametric behaviour models”. In: *International Symposium on Component-Based Software Engineering*. Springer. 2008, pp. 48–63.

[145] Carl Witt et al. “Predictive Performance Modeling for Distributed Computing using Black-Box Monitoring and Machine Learning”. In: *CoRR* abs/1805.11877 (2018).

[146] Leo T Yang, Xiaosong Ma, and Frank Mueller. “Cross-platform performance prediction of parallel applications using partial execution”. In: *Proceedings of the 2005 ACM/IEEE conference on Supercomputing*. IEEE Computer Society. 2005, p. 40.

[147] S. Kavulya et al. “An Analysis of Traces from a Production MapReduce Cluster”. In: *2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. May 2010, pp. 94–103. DOI: 10.1109/CCGRID.2010.112.

[148] Laura Carrington, Allan Snavely, and Nicole Wolter. "A performance prediction framework for scientific applications". In: *Future Generation Computer Systems* 22.3 (2006), pp. 336–346.

[149] Alexandru Calotoiu et al. "Using automated performance modeling to find scalability bugs in complex codes". In: *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. ACM. 2013, p. 45.

[150] Aniello Castiglione et al. "Exploiting mean field analysis to model performances of big data architectures". In: *Future Generation Computer Systems* 37 (2014). Special Section: Innovative Methods and Algorithms for Advanced Data-Intensive Computing Special Section: Semantics, Intelligent processing and services for big data Special Section: Advances in Data-Intensive Modelling and Simulation Special Section: Hybrid Intelligence for Growing Internet and its Applications, pp. 203–211. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2013.07.016. URL: http://www.sciencedirect.com/science/article/pii/S0167739X13001611.

[151] HT Intema et al. "The GMRT 150 MHz all-sky radio survey-First alternative data release TGSS ADR1". In: *Astronomy & Astrophysics* 598 (2017), A78.

[152] S Kazemi et al. "Radio interferometric calibration using the SAGE algorithm". In: *Monthly Notices of the Royal Astronomical Society* 414.2 (2011), pp. 1656–1666.

[153] Ronny Levanda and Amir Leshem. "Synthetic aperture radio telescopes". In: *Signal Processing Magazine, IEEE* 27 (Feb. 2010), pp. 14–29. DOI: 10.1109/MSP.2009.934719.

[154] Cecchi Marco et al. "The glite workload management system". In: *Journal of Physics: Conference Series*. Vol. 219. IOP Publishing. 2010, p. 062039.

[155] I Virtanen et al. *Station Data Cookbook v1.2*. 2018. URL: http://lofar.ie/wp-content/uploads/2018/03/station_data_cookbook_v1.2.pdf.

[156] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open source scientific tools for Python*. [Online; accessed November 21, 2019]. 2001–. URL: http://www.scipy.org/.

[157] Huub Rottgering et al. "LOFAR and APERTIF Surveys of the Radio Sky: Probing Shocks and Magnetic Fields in Galaxy Clusters". English. In: *Journal of Astrophysics and Astronomy* 32.4 (Dec. 2011), pp. 557–566. ISSN: 0250-6335.

[158] Gregory M. Kurtzer, Vanessa Sochat, and Michael W. Bauer. "Singularity: Scientific containers for mobility of compute". In: *PLOS ONE* 12.5 (May 2017), pp. 1–20. DOI: 10.1371/journal.pone.0177459. URL: https://doi.org/10.1371/journal.pone.0177459.

[159] Martin Fowler and Matthew Foemmel. "Continuous integration". In: *ThoughtWorks) http://www. thoughtworks. com/Continuous Integration. pdf* 122 (2006), p. 14.

[160] Jez Humble and David Farley. *Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation (Adobe Reader)*. Pearson Education, 2010.

[161] Andrew E. Slaughter et al. "Continuous integration for concurrent MOOSE framework and application development on GitHub". In: *Journal of Open Research Software* 3.1 (Nov. 2015). ISSN: 2049-9647. DOI: 10.5334/jors.bx.

[162] Pasquale Salza, Filomena Ferrucci, and Federica Sarro. "Develop, deploy and execute parallel genetic algorithms in the cloud". In: *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*. ACM. 2016, pp. 121–122.

[163] Matthew C Chambers et al. "A cross-platform toolkit for mass spectrometry and proteomics". In: *Nature biotechnology* 30.10 (2012), p. 918.

[164] Eric Jeschke and Takeshi Inagaki. "Lessons learned deploying a second generation Observation Control System for Subaru Telescope". In: *Software and Cyberinfrastructure for Astronomy*. Vol. 7740. International Society for Optics and Photonics. 2010, 77400F.

[165] *Travis CI - Test and Deploy Your Code with Confidence*. https://travis-ci.org/. Accessed: 2019-04-18.

[166] *Continuous Integration and Delivery - CircleCI*. https://circleci.com/. Accessed: 2019-04-18.

[167] *GitLab Continuous Integration & Delivery | GitLab*. https://about.gitlab.com/product/continuous-integration/. Accessed: 2019-04-18.

[168] *Jenkins*. https://jenkins.io/. Accessed: 2019-04-18.

[169] Davide Salomoni et al. "INDIGO-Datacloud: foundations and architectural description of a Platform as a Service oriented to scientific computing". In: *CoRR* abs/1603.09536 (2016). arXiv: 1603.09536. URL: http://arxiv.org/abs/1603.09536.

[170] Carl Boettiger. "An Introduction to Docker for Reproducible Research". In: *SIGOPS Oper. Syst. Rev.* 49.1 (Jan. 2015), pp. 71–79. ISSN: 0163-5980. DOI: 10.1145/2723872.2723882. URL: http://doi.acm.org/10.1145/2723872.2723882.

[171] Oscar Esteban et al. “fMRIPrep: a robust preprocessing pipeline for functional MRI”. In: *Nature methods* 16.1 (2019), p. 111.

[172] Björn Grüning et al. “Practical computational reproducibility in the life sciences”. In: *Cell systems* 6.6 (2018), pp. 631–635.

[173] Andrew J Younge et al. “A tale of two systems: Using containers to deploy hpc applications on supercomputers and clouds”. In: *2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*. IEEE. 2017, pp. 74–81.

[174] Yaron Goland et al. *HTTP Extensions for Distributed Authoring–WEBDAV*. Tech. rep. 1999.

[175] Stefan Van Der Walt, S. Chris Colbert, and Gaël Varoquaux. “The NumPy array: a structure for efficient numerical computation”. In: *Computing in Science & Engineering* 13, arXiv:1102.1523 (2 Mar. 2011), pp. 22–30. DOI: 10.1109/MCSE.2011.37. arXiv: 1102.1523 [cs.MS].

[176] The Astropy Collaboration et al. “The Astropy Project: Building an inclusive, open-science project and status of the v2.0 core package”. In: *ArXiv e-prints* (Jan. 2018). arXiv: 1801.02634 [astro-ph.IM].

[177] Gary J Hill et al. “The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results”. In: *arXiv preprint arXiv:0806.0183* (2008).

[178] Mark D Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific data* 3 (2016).

[179] Melanie Johnston-Hollitt. “Taming the Data Deluge to Unravel the Mysteries of the Universe”. In: *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2017, pp. 1–1.

[180] **Alexandar P Mechev** et al. “Automated testing and quality control of LOFAR scienctific pipelines with AGLOW”. In: *Astronomy and Computing* (2019 in review).

[181] T. W. Shimwell et al. “The LOFAR Two-metre Sky Survey - II. First data release”. In: *A&A* 622 (2019), A1. DOI: 10.1051/0004-6361/201833559. URL: https://doi.org/10.1051/0004-6361/201833559.

[182] K. L. Emig et al. “Searching for the largest bound atoms in space”. In: *A&A* (2019).

[183] Y. G. Grange et al. “Characterising radio telescope software with the Workload Characterisation Framework”. In: *arXiv e-prints*, arXiv:1612.00456 (Dec. 2016), arXiv:1612.00456. arXiv: 1612.00456 [astro-ph.IM].

[184] **A.P. Mechev** et al. “Building LOFAR as a Service”. In: Astronomical Data Analysis Software and Systems XXVIII, p. 677.