



## Scaling performance of the SAGECal calibration package: from LOFAR to SKA

J.N. Spreeuw<sup>\*(1)</sup>, S. Yatawatta<sup>(2)</sup>, B.J.C. van Werkhoven<sup>(1)</sup>, and F. Diblen<sup>(1)</sup>

(1) Netherlands eScience Center, Science Park 140, Amsterdam, The Netherlands, <https://www.esciencecenter.nl>

(2) ASTRON, The Netherlands Institute for Radio Astronomy, Dwingeloo, The Netherlands, <https://www.astron.nl/>

### Abstract

This decade, the Square Kilometre Array (SKA) will perform its first observations. Preparations for building dishes, antennas, correlators and infrastructure are well underway. Concurrently, software for the processing of SKA observations is developed at a number of levels. At a more basic level there are the telescope monitoring and control systems and also the correlator software. On top of that, in order to deliver science ready data products, software pipelines are needed for radio frequency interference (RFI) mitigation, averaging, calibration and imaging. Here, we focus on the SAGECal calibration package, in particular on the times needed to obtain calibration solutions. This is an important aspect, since this package is now used for the Epoch of Reionization (EoR) Key Science Project of LOFAR, but will also have to run optimally on SKA1 LOW. In terms of number of stations used for observing this amounts to a factor 10 increase, from 51 to 512 stations. Consequentially, the disk space needed to store an observation will increase by a factor 100, provided the number of frequency channels remains the same. In this paper we investigate the scaling behaviour of SAGECal, whose runtimes should ideally scale linearly with the number of stations. We also explain the algorithms inside SAGECal and use them to explain its scaling behaviour.

### 1 Introduction

A number of radio telescopes with many low-gain elements have recently been deployed, designed and/or funded. However, the history of such telescopes goes back to the Mills Cross Telescopes in Australia <sup>1</sup> and the U.S. <sup>2</sup> in the 1950s and the Clark Lake TPT telescope in the 1960s [1]. Prior to that, in the 1940s, astronomical radio interferometry had begun in Australia with a single radar antenna, a sea-cliff interferometer [2]. Unfortunately for the Clark Lake Telescope, ionospheric effects limited its calibration to short baselines [3]. Self-calibration, a new calibration technique to mitigate these effects, came too late and the Clark Lake TPT could in no way compete with the scientific potential of the VLA, which operated at higher frequencies. These authors [3] explain why two telescopes

with a factor of 10 difference in angular resolution will have a factor of 100 difference in sensitivity. This sophisticated instrument, with a collecting area about 3 times larger than the VLA, closed down in the early 1990s, which marked a period of several decades with little progress in low frequency radio astronomy. This only came to an end with the operation of the Long Wavelength Array (LWA<sup>3</sup>) and the Low Frequency Array (LOFAR<sup>4</sup>). With the advent of these two large telescopes, quite different problems, related to required disk space and compute power arose. These originate from unprecedented available bandwidth and frequency resolution, but also from the large numbers of stations acquiring data for each observation. Since the size, in disk space, of each observation scales with the number of baselines, it scales quadratically with the number of stations. LWA and LOFAR are regarded as pathfinders for the Square Kilometre Array (SKA)<sup>5</sup>. The SKA will operate at two sites, in Australia at low (SKA1 LOW) and in South-Africa at intermediate (SKA1 MID) radio frequencies. For SKA1 LOW 512 stations will be built <sup>6</sup>, which amounts to data sizes a factor 100 larger than LOFAR - 51 Low Band stations as of mid 2019 <sup>7</sup> - if the number of frequency channels would be comparable. For the reduction pipelines that turn the raw observational data into science ready data products such large amounts of data pose severe challenges. Apart from the (temporary) disk space needed, it is essential that the throughput of observational data keeps pace with its collection. The reduction pipelines will perform RFI mitigation, averaging, calibration and imaging. Of these four tasks, calibration is the most compute intensive, and we focus on that aspect in this paper, by investigating the performance of the SAGECal calibration package for numbers of stations comparable to LOFAR up to the number of stations expected for SKA1 LOW. Also, we investigate its dependence on the number of sources in the sky model. In reality, accurate calibration will involve a number of self-calibration loops, which include not only calibration, but also imaging, but we will neglect the latter for now.

<sup>1</sup>[https://www.atnf.csiro.au/news/newsletter/jun02/Flowering\\_of\\_Fleurs.htm](https://www.atnf.csiro.au/news/newsletter/jun02/Flowering_of_Fleurs.htm)

<sup>2</sup>[https://radiojove.gsfc.nasa.gov/library/sci\\_briefs/discovery.html](https://radiojove.gsfc.nasa.gov/library/sci_briefs/discovery.html)

<sup>3</sup><http://www.phys.unm.edu/~lwa/index.html>

<sup>4</sup><https://www.astron.nl/telescopes/lofar>

<sup>5</sup><https://www.skatelescope.org/precursors-pathfinders-design-studies/>

<sup>6</sup><https://www.skatelescope.org/technical/info-sheets/>

<sup>7</sup><http://old.astron.nl/radio-observatory/astronomers/technical-information/lofar-technical-information>

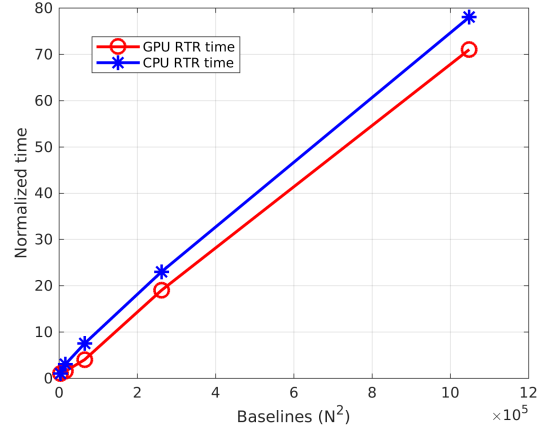
## 2 Description of the SAGECal calibration package

SAGECal is an amalgamation of several key algorithms for optimal performance and accuracy in calibration. We briefly discuss these algorithms in the following and highlight their unique characteristics.

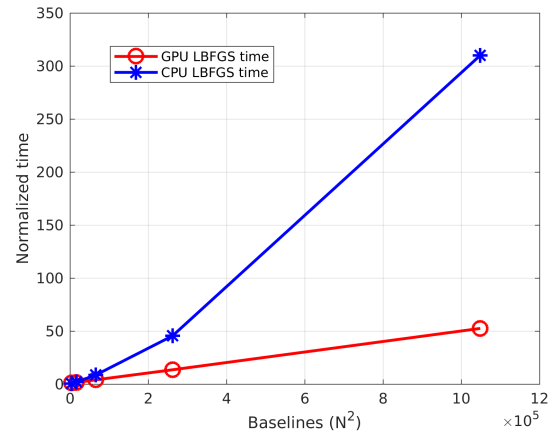
- Expectation maximization (EM) and space alternating generalized expectation maximization (SAGE) algorithms [4]: These algorithms enable us to decompose calibration along multiple directions in the sky into a set of uni-directional calibration problems [5]. Therefore, we get linear scaling with the number of directions being calibrated.
- Expectation-conditional maximization either (ECME) algorithm [6]: We use this algorithm to minimize a Student's T cost function [7] in robust calibration.
- Limited memory, Broyden Fletcher Goldfarb Shanno (LBFGS) algorithm [8]: By approximating the (inverse) Hessian using a set of vectors, we can calibrate with a large number of unknowns, without running out of memory. Moreover, by using a stochastic LBFGS algorithm [9], we can calibrate data at the highest resolution.
- Riemannian trust region (RTR) algorithm [10]: Combined with the truncated conjugate gradient method [11, 12], we get a solver that scales linearly in memory usage with the number of stations [13] unlike quadratic scaling of other Newton methods. Moreover, because we calibrate on a Riemannian quotient manifold, we get faster convergence compared to Euclidean space calibration.
- Alternating direction method of multipliers (ADMM) [14]: We can calibrate data taken over a wide bandwidth and also data that are stored across a compute cluster by formulating it as a distributed consensus optimization problem [15, 16].

All the above algorithms have been accelerated using GPUs [17, 18]. We show the scaling of the workhorses of SAGECal, i.e., the optimization algorithms RTR and LBFGS, using simulated data. We simulate mock radio telescopes having the number of stations  $N$  from 64 (LOFAR) to 1024 (super SKA). The calibration is performed along one direction only because the linear scaling with more than one direction is well illustrated in our previous work [17, 18].

For the comparison, we use a 12 core Intel Xeon E5-2670 CPU and an NVIDIA Tesla K40c GPU. We plot the time taken by RTR and LBFGS as well as the total time taken to calibrate 10 time samples. We plot the time taken (normalized by the GPU time for  $N = 64$ ) against  $N^2$  (instead of  $N$ ) in Figs. 1, 2 and 3. This is to illustrate the scaling with the number of baselines which is proportional to  $N^2$ .



**Figure 1.** The scaling of RTR algorithm with  $N^2$ . Compared to LOFAR, SKA will require about  $\times 20$  more time.

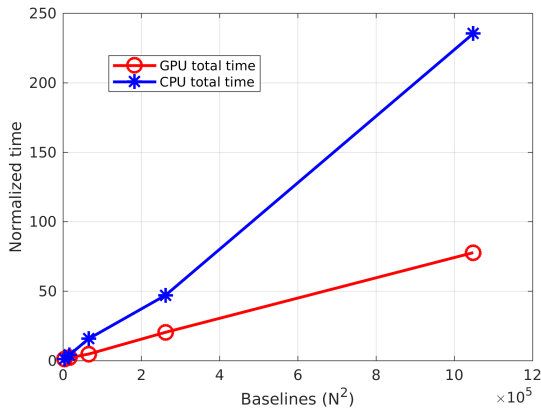


**Figure 2.** The scaling of LBFGS algorithm with  $N^2$ . Compared to LOFAR, SKA will require about  $\times 20$  ( $\times 50$  for CPU) more time.

From Fig. 1, we see that the performance of RTR algorithm is similar in both GPU and CPU versions. The CPU is not overwhelmed while running the RTR algorithm and can keep up with the GPU. In contrast, we see a significant overhead in the CPU version of LBFGS in Fig. 2. The total time in Fig. 3 is dominated by the LBFGS algorithm. Overall, we clearly see the scaling that is dependent on the data, which is in turn dependent on the number of baselines. Using mini-batches of data as in [9] will improve this and we pursue this as future work. Note also that we have only shown the scaling of the optimization algorithms and for full calibration, additional time is required for model computation and data reading and writing [18].

## 3 Conclusions

We have shown the scaling of the core optimization algorithms used in SAGECal. With GPU acceleration, we require about  $\times 20$  more computations when we scale from LOFAR to SKA. Considering that there is  $\times 8$  increase in



**Figure 3.** The scaling of both RTR and LBFSG algorithms with  $N^2$ . Compared to LOFAR, SKA will require about  $\times 20$  ( $\times 50$  for CPU) more time.

the number of stations and  $\times 64$  increase in the number of baselines, we see that the scaling is superlinear.

#### 4 Acknowledgements

This work is supported by Netherlands eScience Center (project DIRAC, grant 27016G05).

#### References

- [1] Erickson, W. C., Mahoney, M. J., & Erb, K. 1982, *Astrophysical Journal*, Supplement, 50, 403
- [2] Sullivan, W. T. 1991, *IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications*, 132
- [3] Kassim, N. E., & Polisensky, E. J. 2005, *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science*, 114
- [4] J.A. Fessler and A.O. Hero, "Space alternating generalized expectation maximization algorithm," *IEEE Trans. on Sig. Proc.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
- [5] S. Kazemi, S. Yatawatta, S. Zaroubi, P. Labropoulos, A.G. de Bruyn, L. Koopmans, and J. Noordam, "Radio interferometric calibration using the SAGE algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 2, pp. 1656–1666, June 2011.
- [6] Chuanhai Liu and D.B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, , no. 5, pp. 19–39, 1995.
- [7] S. Kazemi and S. Yatawatta, "Robust radio interferometric calibration using the t-distribution," *Monthly Notices of the Royal Astronomical Society*, vol. 435, pp. 597–605, Oct. 2013.
- [8] R. Fletcher, *Practical Methods of Optimization; (2Nd Ed.)*, Wiley-Interscience, New York, NY, USA, 1987.
- [9] S. Yatawatta, L. De Clercq, H. Spreeuw, and F. Diblen, "A stochastic LBFSG algorithm for radio interferometric calibration," in *2019 IEEE Data Science Workshop (DSW)*, June 2019, pp. 208–212.
- [10] P.-A. Absil, C. G. Baker, and K. A. Gallivan, "Trust-region methods on Riemannian manifolds," *Found. Comput. Math.*, vol. 7, no. 3, pp. 303–330, July 2007.
- [11] Trond Steihaug, "The conjugate gradient method and trust regions in large scale optimization," *SIAM Journal on Numerical Analysis*, vol. 20, no. 3, pp. 626–637, 1983.
- [12] Philippe Toint, "Towards an Efficient Sparsity Exploiting Newton Method for Minimization," pp. 57–88, Academic press, 1981, Publication editors : I.S. Duff.
- [13] S. Yatawatta, "Radio interferometric calibration using a Riemannian manifold," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3866–3870.
- [14] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [15] S. Yatawatta, "Distributed radio interferometric calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 449, no. 4, pp. 4506–4514, 2015.
- [16] S. Yatawatta, F. Diblen, H. Spreeuw, and L. V. E. Koopmans, "Data multiplexing in radio interferometric calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 475, pp. 708–715, Mar. 2018.
- [17] S. Yatawatta, S. Kazemi, and S. Zaroubi, "GPU accelerated nonlinear optimization in radio interferometric calibration," in *Proceedings of Innovative Parallel Computing (InPar'12)*, 2012.
- [18] Hanno Spreeuw, Ben van Werkhoven, Sarod Yatawatta, and Faruk Diblen, "GPU acceleration of the SAGECal calibration package for the SKA," *arXiv e-prints*, p. arXiv:1910.13908, Oct 2019.