

## EHT-HOPS pipeline for millimeter VLBI data reduction

LINDY BLACKBURN,<sup>1</sup> CHI-KWAN CHAN,<sup>2</sup> GEOFFREY B. CREW,<sup>3</sup> VINCENT L. FISH,<sup>3</sup> SARA ISSAOUN,<sup>4</sup> MICHAEL D. JOHNSON,<sup>1</sup>  
MACIEK WIELGUS,<sup>1,5</sup> KAZUNORI AKIYAMA,<sup>6,3,7</sup> JOHN BARRETT,<sup>3</sup> KATHERINE L. BOUMAN,<sup>1</sup> ROGER CAPPALLO,<sup>3</sup>  
ANDREW A. CHAEL,<sup>1,5</sup> MICHAEL JANSSEN,<sup>4</sup> COLIN J. LONSDALE,<sup>3</sup> AND SHEPERD S. DOELEMAN<sup>1</sup>

<sup>1</sup>*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*

<sup>2</sup>*University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*

<sup>3</sup>*Massachusetts Institute of Technology, Haystack Observatory, 99 Millstone Rd, Westford, MA 01886, USA*

<sup>4</sup>*Department of Astrophysics/IMAPP, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*

<sup>5</sup>*Black Hole Initiative, Harvard University, 20 Garden Street, Cambridge, MA 02138, USA*

<sup>6</sup>*National Radio Astronomy Observatory, 520 Edgemont Rd, Charlottesville, VA, 22903, USA*

<sup>7</sup>*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*

### ABSTRACT

We present the design and implementation of an automated data calibration and reduction pipeline for very-long-baseline interferometric (VLBI) observations taken at millimeter wavelengths. These short radio-wavelengths provide the best imaging resolution available from ground-based VLBI networks such as the Event Horizon Telescope (EHT) and the Global Millimeter VLBI Array (GMVA), but require specialized processing due to the strong effects from atmospheric opacity and turbulence as well as the heterogeneous nature of existing global arrays. The pipeline builds upon a calibration suite (HOPS) originally designed for precision geodetic VLBI. To support the reduction of data for astronomical observations, we have developed an additional framework for global phase and amplitude calibration which provides output in a standard data format for astronomical imaging and analysis. We test the pipeline on observations taken at 3.5 mm (86 GHz) by the GMVA joined by the phased Atacama Large Millimeter/submillimeter Array in April 2017, and demonstrate the benefits from the specialized processing of high frequency VLBI data with respect to classical analysis techniques.

*Keywords:* techniques: high angular resolution – techniques: interferometric

### 1. INTRODUCTION

In the technique of very-long-baseline interferometry (VLBI), signals from an astronomical source are recorded independently at multiple locations, and later brought together for pairwise correlation. This process samples the coherence function of the incident radiation at separations corresponding to the baseline vectors between sites. The resolution probed by a baseline is determined by the interferometric fringe spacing,  $1/|\mathbf{u}| = \lambda/D_{\text{proj}}$ , in angular units on the sky, where the 2-dimensional spatial frequency  $\mathbf{u} = (u, v)$  corresponds to the projected baseline vector in units of observing wavelength  $\lambda$ . Thus, the highest resolutions are achieved when sites have the widest possible separation  $D$  and observe at the highest possible frequencies  $\nu = c/\lambda$ .

Two global networks exist for millimeter VLBI observations. The Global Millimeter VLBI Array<sup>1</sup> (GMVA) operates at 3.5 mm (86 GHz) and includes the Very Long Baseline Array<sup>2</sup> (VLBA) and a number of large-

aperture dishes with the required surface accuracy and sufficiently good local weather to operate at 3.5 mm. The Event Horizon Telescope<sup>3</sup> (EHT) operates as an array at 1.3 mm (230 GHz), a wavelength at which only a handful of existing sites globally are able to observe. In April 2017, both networks participated in science observations for the first time with the Atacama Large Millimeter/submillimeter Array (ALMA). ALMA acted as a phased array of  $\sim 37$  dishes (Matthews et al. 2018; Goddi et al. 2019), providing a highly sensitive anchor station that greatly expanded the sensitivity, resolution, and baseline coverage of the VLBI networks. In particular, the EHT 2017 array, operating over six geographical locations and including ALMA, is anticipated to have the necessary sensitivity and coverage in order to image horizon-scale features of the supermassive black-hole candidates at the Galactic Center, Sgr A\*, and at the center of the M87 galaxy.

At the heart of the VLBI technique is the correlation of the raw station data using either dedicated hardware or software. The correlation is manifest as an inter-

<sup>1</sup> <https://www3.mpifr-bonn.mpg.de/div/vlbi/globalmm>

<sup>2</sup> <https://science.nrao.edu/facilities/vlba>

<sup>3</sup> <https://eventhorizontelescope.org>

ference fringe that changes in an expected way as the Earth rotates. This is a simple, but computationally expensive process that requires good, but nevertheless approximate, models in order to measure the interferometric fringe. Some post-correlation processing is then required to detect and analyze the fringes to obtain scientifically useful results.

The VLBI correlator estimates the complex correlation for signals  $x_1$  and  $x_2$  between pairs of antennas,

$$r_{12} = \frac{\langle x_1 x_2^* \rangle}{\eta_Q \sqrt{\langle x_1 x_1^* \rangle \langle x_2 x_2^* \rangle}} = \frac{e^{i\phi_1} e^{-i\phi_2} \mathcal{V}_{12}}{\sqrt{\text{SEFD}_1 \times \text{SEFD}_2}}. \quad (1)$$

In this expression,  $\eta_Q$  is a correction factor of  $\sim 0.88$  accounting for the introduction of quantization noise during 2-bit digitization and  $\mathcal{V}$  ( $\sim 1$  Jy for continuum sources) is the correlated flux density. The *system-equivalent flux density* (SEFD  $\sim 10^4$  Jy, see Section 3.1) reflects the original analog system noise  $\eta_Q \langle x x^* \rangle$  in effective flux units of an astronomical source above the atmosphere, and the  $e^{i\phi}$  are station phase terms corresponding to residual geometric, atmospheric, and instrumental phase suffered by the signal before it is recorded. We adopt the convention of Rogers et al. (1974) where positive delay (and unwrapped phase) corresponds to the signal arriving at station 2 after station 1.

The primary residual systematics after correlation are small errors in *delay* and *delay-rate*, which are related to the first-order variation of the baseline phase,  $\psi = \text{Arg}[r]$ , of the complex correlation between two sites in time and frequency ( $t, \nu$ ):

$$\Delta\psi = \frac{\partial\psi}{\partial\nu} \Delta\nu + \frac{\partial\psi}{\partial t} \Delta t. \quad (2)$$

Since phase error  $\psi = 2\pi\nu\tau$ , the delay and delay-rate are given by (Thompson et al. 2017, A12.28, A12.22)

$$\tau = \frac{1}{2\pi} \frac{\partial\psi}{\partial\nu} \quad \dot{\tau} = \frac{1}{2\pi\nu} \frac{\partial\psi}{\partial t}. \quad (3)$$

For linear phase drift, the coherence has a sinc profile,

$$\frac{1}{\Delta\psi} \int_{-\Delta\psi/2}^{\Delta\psi/2} d\psi \cos\psi = \frac{\sin(\Delta\psi/2)}{\Delta\psi/2}, \quad (4)$$

as a function of accumulated phase drift,  $\Delta\psi$ , so that maximum coherence occurs at the fringe solution where data is compensated for fringe phase rotation and the accumulated  $\Delta\psi \rightarrow 0$ . First-order fringe searches vary the two parameters, delay and delay-rate, and search for maximum coherence in excess correlated signal power over the full bandwidth and up to the length of a scan. The original signals are highly noise dominated ( $|r| \lesssim 10^{-4}$ ), and generally at least the first-order fringe correction must be applied in order to coherently average a sufficient number of samples and produce level of correlated flux above the statistical (thermal) noise.

The EHT and GMVA are comprised of heterogeneous collections of individual stations with varying sensitivities and characteristics, and they target high observing frequencies over wide bandwidths. For both VLBI networks, non-linear phase systematics beyond the first-order fringe solution are important. These include phase variations over the observing band due to small variations in path delay versus frequency prior to digitization, and stochastic phase fluctuations in time due to achromatic path variations from atmospheric turbulence. The instrumental phase bandpass is typically constant over long timescales and can be solved using bright calibrator sources. Atmospheric phase is more difficult, as it is continuously varying and must be solved on-source. At millimeter wavelengths, the atmospheric phase can have a decoherence timescale of seconds, and compensating for it requires that the source be detectable on a baseline to within just some fraction of the decoherence time. The need to be able to measure and compensate for the atmosphere on-source at rapid timescales has been a primary driver of the wide recording bandwidths targeted by the EHT.

In Section 2 which follows, we introduce overall structure and algorithms behind the iterative phase calibration applied during the EHT-HOPS pipeline. In Section 3, we describe a suite of postprocessing tools that perform absolute flux calibration and polarization gain ratio calibration, enabling the formation of calibrated Stokes  $I$  visibility coefficients in a standard UVFITS file format. Section 4 describes the overall EHT-HOPS computing software organization and workflow. The EHT-HOPS pipeline is tested on a representative 3.5 mm GMVA+ALMA data set in Section 5, and the output of the pipeline is compared against a classical reduction pathway for low-frequency VLBI in terms of fringe detection, consistency of measured phase and amplitude, and similarity of derived images on blazar NRAO 530.

## 2. EHT-HOPS PIPELINE

The current Haystack Observatory Postprocessing System<sup>4</sup> (HOPS) was born from the efforts of Alan Rogers in the late 70's with a program called FRNGE, which was written in Fortran and designed to be efficient on an HP-21MX (later renamed HP-1000) minicomputer (Rogers 1970; Rogers et al. 1974). With improvements in hardware and software, a rewrite and augmentation of the toolset was launched in the early 90's by Colin Lonsdale, Roger Cappallo and Cris Niell as driven by the needs of the geodetic community and of a move to higher frequencies in astronomical VLBI. The basic algorithms were adopted from FRNGE; but there was a complete rewrite of the code into (K&R) C and substantial revisions of the input/output, control and file structures, and graphical and summary analysis

<sup>4</sup> <https://www.haystack.mit.edu/tech/vlbi/hops.html>

tools resulting in the framework of the current HOPS system. This was followed by a substantial effort in the early-mid 00’s to develop tools for optimizing signal-to-noise (S/N) and deriving correction factors for data with imperfect coherence, based on analysis of amplitude with coherent averaging time (Rogers et al. 1995). Further evolution in the late 00’s was provoked by the re-emergence of software correlation (DiFX; Deller et al. 2011), and in the 10’s by the needs of EHT-scale millimeter VLBI which brings us to HOPS in its current form.

Acknowledging its geodetic heritage, HOPS was optimized for precision on per-baseline delay and delay-rate measurements which are the fundamental quantities of interest for geodetic analysis programs. Consequently, it is somewhat light on support for some routine calibration processes found in some other astronomical software packages, such as the Astronomical Image Processing System (AIPS; Greisen 2003) and the Common Astronomy Software Applications package (CASA; McMullin et al. 2007). Nevertheless, it provides a good framework for the reduction and analysis of millimeter VLBI data, where the complexities of atmospheric effects require ever more specialized processing to obtain reliable astronomical results.

Over the last decade, HOPS has been used extensively for the analysis and reduction of early EHT data (e.g., Doleman et al. 2008, 2012; Fish et al. 2011, 2016; Johnson et al. 2015; Akiyama et al. 2015; Lu et al. 2018). The HOPS suite grew to support the evolving EHT instrument, with steadily increasing bandwidth (Whitney et al. 2013; Vertatschitsch et al. 2015), dual-polarization observations (Johnson et al. 2015), and a move from the Mark4 hardware correlator (Whitney et al. 2004) to the DiFX software correlator (Deller et al. 2011). Calibration strategies were also developed and implemented within HOPS in order to support the segmented averaging of amplitudes and bispectra (Johnson et al. 2015; Fish et al. 2016) as well as on-source phase stabilization (Johnson et al. 2015). The techniques improved the ability to build S/N of visibility amplitude and phase information for high-frequency EHT observations, in the presence of rapid atmospheric phase fluctuations.

For the needs of the EHT campaigns of 2017 and subsequent years, we have extended the basic HOPS framework with Python-based packages, included within the EHT Analysis Toolkit<sup>5</sup> (`eat` library). The Python libraries provide a convenient Python-based interface to the underlying HOPS binary and ASCII file formats via Python `ctypes` and Pandas DataFrames, and they provide a community-standard UVFITS output data format for downstream processing. The `eat` routines are also able to enforce a global (station-based) calibration solution across the VLBI array, locking together the

baseline-based fringe solutions provided by the HOPS `fourfit` fringe fitter. The HOPS and `eat` software suites are packaged together into a EHT-HOPS pipeline, with a set of driver scripts that run an automated end-to-end calibration and reduction of EHT or GMVA correlated data given a minimal basic configuration.

The first five stages of the pipeline run several iterations of `fourfit` (Capallo 2017), while solving for non-linear phase corrections and a global fringe solution. The pipeline workflow is shown in Figure 1 and specific details of the `fourfit` stages are below. Examples of various steps are provided via application of stages of the pipeline on a representative 3.5 mm GMVA+ALMA data set from 2017 (project code MB007), the scientific results of which are published in Issaoun et al. (2019). Details of the observations and data reduction are given in Section 5.1.

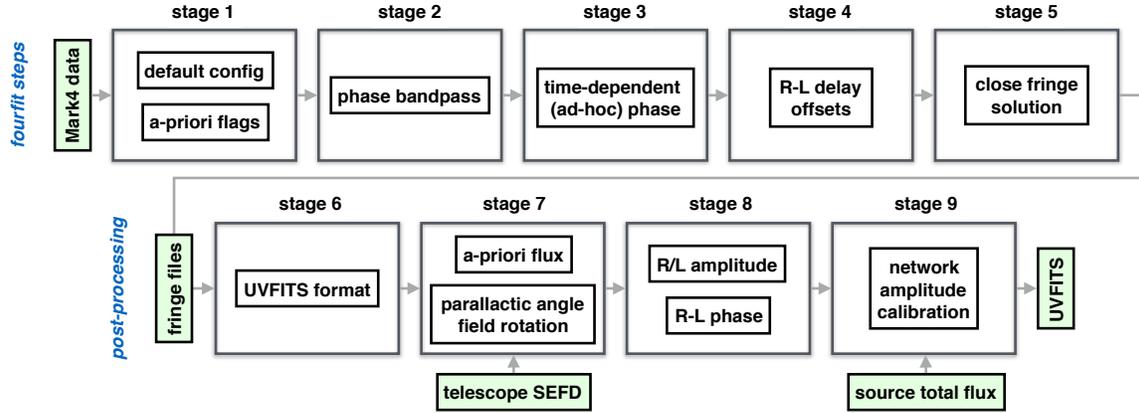
### 2.1. Data Flagging

Data selection and flagging are defined using HOPS ASCII control codes. Data selection involves setting the start and stop time of processing, as well as which frequency channels are processed. Flagging defines small intervals of time within the processed segment, and small frequency ranges within a channel (notches) that have their data weights set to zero and are thus ignored when fringe fitting and visibility averaging. The EHT-HOPS pipeline does not currently implement automated flagging in either time or frequency, and these must be defined by hand from data inspection and telescope logs. However HOPS tool `aedit` and custom timeseries and spectral plotting tools within the `eat` library are available to assist with identifying time and frequency ranges, as well as programmatic manipulation of the relevant HOPS control codes.

### 2.2. Bandpass Calibration

Bandpass response of an antenna can be understood in context of the signal path from Figure 3. In the simplified picture, the recorded signal  $x(f)$  is composed of the sum of received source signal  $HS$  and system noise  $n$ , subject to a common transfer function  $G$  and additive quantization noise  $q$  before being digitally recorded to disk:  $x = G(HS + n) + q$ .  $H$  includes effects such as atmospheric attenuation, dish characteristics, and receiver response. System noise  $n$  includes contributions such as receiver thermal noise, atmospheric emission, and radio-frequency interference. The common transfer function  $G$  accounts for components like cable transmission and back-end electronics. Finally the effects of low-bit quantization can be approximated as additive quantization noise that depends on the signal profile prior to recording. For noise-dominated signals with a flat spectrum, quantization noise is white and uncorrelated with the source signal, and the effect on the data is modeled in a straightforward way by the correlation amplitude efficiency factor  $\eta_Q$  from Equation 1.

<sup>5</sup> <https://github.com/sao-eh/eat>



**Figure 1.** Stages of the EHT-HOPS pipeline and post-processing steps. Stages 1–5 represent iterations of HOPS fringe fitter `fourfit`, where the input data for each stage are the original correlator output files (converted from DiFX native to Mark4 format), and the output data are a series of reduced HOPS native *fringe files* (averaged visibility data plus fringe solutions) and auxiliary calibration parameters (described in the inner boxes) used to refine the fringe search for successive stages. The order of the stages is not fundamental to the calibration process, but is largely determined by which up-front corrections are needed to provide more precise downstream estimation of calibration parameters. After an initial run with a-priori fringe search windows, channel configuration, and data flags, the residual phase bandpass and differential phase-vs-time (*ad-hoc phase*) are calibrated to a reference station in the array during stages 2 and 3. At stage 4, precise delays are measured and aligned between RCP and LCP feeds at each station, so that a single global (station-based) fringe solution in delay and delay-rate can be solved for and applied in stage 5. The output of stage 5 is converted to UVFITS format, and a remaining suite of post-processing tools provide amplitude calibration and time-and-polarization dependent phase calibration, as these cannot currently be performed within `fourfit`. A final stage of network calibration folds in a-priori information about array redundancy and total flux density to self-calibrate co-located sites in a model-independent way.

The *system-equivalent flux density* (SEFD) is defined as the source flux necessary to contribute equal signal power to the system noise. In terms of elements from Figure 3,

$$\text{SEFD} = \frac{\langle |n^2| \rangle}{\langle |Hs|^2 \rangle} S \quad (5)$$

where  $S$  is the flux density of the (unpolarized) source that generates  $s$ . Ignoring quantization and assuming a noise-dominated signal, the auto-correlation spectrum of the received signal  $x$  is

$$\langle xx^* \rangle = \langle |Gn|^2 \rangle \quad (6)$$

and the cross-correlation spectrum is

$$\langle x_1 x_2^* \rangle = \langle s_1 s_2^* \rangle H_1 G_1 H_2^* G_2^* \quad (7)$$

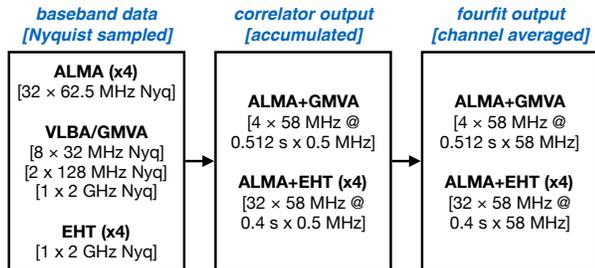
as the system noise between sites is uncorrelated.

On a single baseline, the bandpass from both antennas 1 and 2 will directly affect the correlation coefficient measured (Equation 1). The DiFX software correlator (Deller et al. 2011), used for both EHT and GMVA correlation, computes  $\langle x_1 x_2^* \rangle$  averaged over 1 sub-channel ( $\sim 0.5$  MHz) and 1 AP (accumulation period,  $\sim 0.5$  seconds), as illustrated in Figure 2. The values for each AP are then normalized by their channel-average autocorrelation power during the DiFX  $\rightarrow$  Mark4 data conversion stage (using DiFX conversion tool `difx2mark4`). This

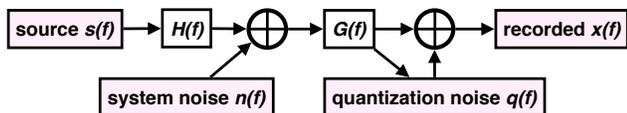
step removes the “auto-correlation” amplitude bandpass  $|G_1 G_2^*|$  (at the resolution of a full channel), but leaves the residual cross-power amplitude bandpass from  $|H_1 H_2^* / \langle n_1 n_2^* \rangle|$  that reflects changes in SEFD over frequency. Also left is the combined phase bandpass,  $\text{Arg}[H_1 G_1 H_2^* G_2^*] = \phi_1 - \phi_2$ , that reflects very small and stable changes in instrumental path length as a function of frequency.

Stage 2 in the EHT-HOPS pipeline estimates and provides corrections for the relative *phase* bandpass over a baseline by averaging over an ensemble of high S/N cross-correlation measurements to a common reference station. High S/N fringes from the reference station (generally ALMA) to other stations in the network are taken from stage 1 output to estimate a single baseline phase and phase slope per 58 MHz channel by direct S/N-weighted average. Baselines that do not contain the reference antenna (station 0) can then be assumed to be subject to phase bandpass  $\psi_{ij} = \psi_{0j} - \psi_{0i}$ .

Because `fourfit` output is already channel-averaged, it is not possible to directly measure *intra-channel* phase bandpass from detected fringes, regardless of S/N. Generally the phase evolution across each 58 MHz channel is small ( $< 10$  degrees), as is any possible coherence loss from residual intra-channel phase variation. To track situations of more rapid intra-channel phase variation, particularly near the 2 GHz band edge of the EHT, the first-order phase slope  $\partial\psi_{0i}/\partial\nu$  is also estimated using



**Figure 2.** Time and frequency resolution of data, covering a single ALMA spectral window, as it is reduced. This represents 1/4 of the total recorded bandwidth for EHT+ALMA at each station as of 2018, where each  $\sim 2$  GHz spectral window is correlated and reduced independently. Correlation parameters when ALMA is present are largely driven by the configuration of ALMA tunable filter bank (TFB) channels which are 62.5 MHz wide, overlap slightly, and have starting frequencies aligned to  $1/(32 \mu\text{s})$ . Correlation for GMVA and EHT must therefore use an FFT window of at least  $32 \mu\text{s}$  to align to the MHz, and currently use  $64 \mu\text{s}$  to also center the channel. The  $64 \mu\text{s}$  FFT window determines available correlation accumulation periods (AP), which must be an integer number of FFT window lengths. GMVA has chosen 0.512 s accumulation periods, while the EHT uses 0.4 s. Frequency accumulation is 0.5 MHz for both networks. The raw output of HOPS fringe fitter *fourfit* maintains the original AP, but averages over each 58 MHz channel. This resolution is maintained throughout the EHT-HOPS post-processing stages, until it is time/band averaged after network calibration (not shown) for a more manageable data volume.

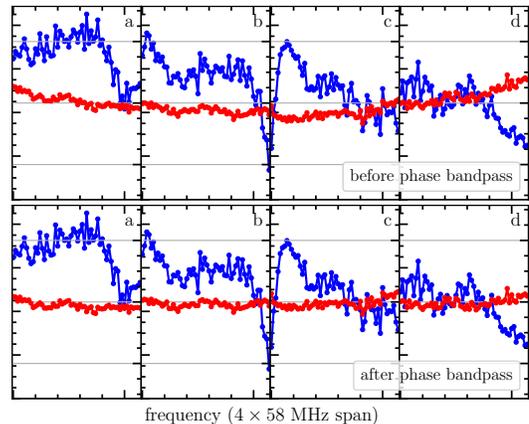


**Figure 3.** Simplified signal path reflecting the bandpass response of one antenna given input source  $s(f)$  and system noise  $n(f)$  signals represented in the frequency domain. Transfer functions  $H(f)$  and  $G(f)$  represent the scaling and shaping of signals as they pass through components of the environment and instrument. The recorded digitized signal  $x(f) = G(Hs + n) + q$ .

the differences between nearby channels, and a linear phase slope correction is implemented as a channel-by-channel “single-band-delay” (SBD) offset referenced to the center of each channel.

The total instrumental phase attributed to each station  $j$  relative to the reference station is

$$\phi_j(f) = \psi_{0j,c} + 2\pi(f - f_{\text{ref},c})\tau_{0j,c}, \quad (8)$$

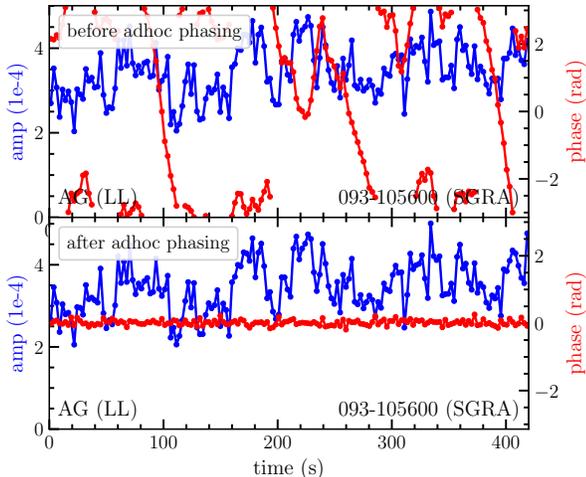


**Figure 4.** Amplitude (blue) and phase (red, covering  $\pm\pi$ ) spectrum of the correlation coefficient between the Fort Davis VLBA station (FD) and the Green Bank Telescope (GBT) during a scan on calibrator source 1749+096, before and after phase bandpass correction. The spectrum is shown across the four 58 MHz channels (labeled a,b,c,d) that are defined at correlation. The phase bandpass correction over the GMVA 256 MHz bandpass is small, but the correction is more pronounced near the edge of the much wider 2 GHz-wide EHT bands. Using HOPS control codes, the pipeline is able to correct for an offset as well as one slope (single-band delay) per channel.

where within the range of each channel  $c$ ,  $\psi_{0j,c}$  is the average measured instrumental phase for channel  $c$  taken at the channel reference frequency  $f_{\text{ref},c}$ , and  $\tau_{0j,c}$  is a small single-band delay used to track phase variation within each 58 MHz channel. Due to the available tuning parameters in *fourfit*, the  $\psi_{0j,c}$  contribution is polarization dependent, while the  $\tau_{0j,c}$  contribution is taken as an average over both polarizations.

### 2.3. Atmospheric Phase

The phase evolution over time captured by the first-order fringe fit is insufficient for millimeter VLBI, where atmospheric turbulence causes non-linear, stochastic phase evolution on timescales of seconds, much shorter than a typical VLBI scan length of minutes. Unlike the non-linear corrections in phase from stable instrumental bandpass mentioned previously, atmospheric phase is continually changing and must be measured and corrected *on-source*. HOPS provides the ability to pre-correct non-linear phase evolution over time using *ad-hoc* phases, where the term *ad-hoc* is used to distinguish these arbitrary atmospheric phase corrections from the modeled linear phase drift due to delay-rate. These non-linear corrections are estimated and applied at stage 3 in the pipeline, resulting in an overall increase in scan-average S/N as well as increased precision and



**Figure 5.** Amplitude (blue) and phase (red) time series of the correlation coefficient between ALMA and GBT during a scan on Sgr A\*. Atmospheric phase compensation is done using the round-robin implementation from Section 2.3 that prevents self-tuning; with an automatically-chosen effective integration timescale  $\mathcal{T}_{\text{dof}} = 2.5$  seconds.

overall self-consistency of the linear fringe solutions across the array.

### 2.3.1. Reference Station Selection

Similar to instrumental phase, atmospheric phase corrections are assigned to each station relative to a reference station. However, since one reference antenna may not be present in all scans, the choice of reference antenna is made scan-by-scan by maximizing a statistic designed to capture the total measurable phase degrees-of-freedom using only baselines to the reference antenna. The scoring depends on the S/N from the proposed reference antenna to all remaining antennas  $\rho_i$ , the S/N required for a good phase measurement  $\rho_{\text{dof}} \sim 10$ , a S/N threshold below which false fringes appear  $\rho_{\text{thr}} \sim 7$ , and an assumed phase coherence timescale  $\mathcal{T}_{\text{coh}} \sim 6$  seconds at 1.3 mm and  $\sim 18$  seconds at 3.5 mm, characteristic of challenging weather. Here,  $\mathcal{T}_{\text{coh}}$  is defined as the expected time span over which phase drifts by 1 radian (as defined later in Equation 14).

Each  $\rho_i$  is taken as a quadrature sum over all four polarization products  $j$  on a single baseline, reflecting the fact that changes in atmospheric path delay do not depend on polarization and polarization products can be stacked for better S/N,

$$\rho_i^2 = \sum_j \rho_{i,j}^2 (2/\pi) \arctan [(\rho_{i,j}/\rho_{\text{thr}})^4], \quad (9)$$

with the arctan logistic function applying the threshold at which to ignore potentially false fringes. For a given baseline  $i$  with scan-average S/N of  $\rho_i$ , we estimate the

number of segments that could be formed by splitting the scan in time, while maintaining a S/N above  $\rho_{\text{dof}}$  for each segment. This corresponds to the number of measurable degrees-of-freedom above some nominal statistical precision,

$$N_{\text{meas},i} = (\rho_i/\rho_{\text{dof}})^2. \quad (10)$$

At very high S/N, the number of measurable degrees-of-freedom might be very large, corresponding to a very short segment duration. In this situation, the maximum useful degrees of freedom over total duration  $\mathcal{T}_i$  is limited by the number of phase measurements required to fully characterize any atmospheric variability. Correcting phase more rapidly than  $\sim 5$  times per  $\mathcal{T}_{\text{coh}}$  gives rapidly diminishing returns, so we set

$$N_{\text{max},i} = 5(\mathcal{T}_i/\mathcal{T}_{\text{coh},i}). \quad (11)$$

Finally, we calculate the total useful degrees-of-freedom by summing over all baselines under a scheme which reflects the diminishing returns for measurable degree-of-freedom beyond the maximum useful number,

$$N_{\text{useful}} = \sum_i N_{\text{max},i} \ln(1 + N_{\text{meas},i}/N_{\text{max},i}). \quad (12)$$

The reference station chosen is the one which provides the largest  $N_{\text{useful}}$  for a given scan.

### 2.3.2. Data Alignment

When the time required to accumulate  $\rho_{\text{dof}}$  approaches  $\mathcal{T}_{\text{coh}}$ , performance of on-source phase stabilization increases dramatically by stacking data prior to measuring phase. For example by stacking two equal sensitivity measurements and increasing the S/N by  $\sqrt{2}$ , the timescale over which phases can be reliably estimated is correspondingly reduced by a factor of two.

For the purpose of atmospheric phase estimation, the EHT-HOPS pipeline stacks data from all polarization products by aligning the data empirically before computing a weighted average. First, data are band-averaged and adjusted to a common fringe solution as prior to step 5 fringe globalization (Figure 1), different polarization products may have different delay rate solutions. The empirical phase offset between one of the polarization products  $r_i$  and another  $r_j$  is measured by segmented average,

$$\Delta\psi_{ij} = \text{Arg} \left[ \sum_n r_i[n] r_j^*[n] \right], \quad (13)$$

where the length of each segment should be long enough to accumulate to  $\rho > 1$ .

The measured  $\Delta\psi$  between one of the polarization products and others is then used to align and stack the original visibility data. While it may be challenging to

accumulate sufficient S/N per  $\mathcal{T}_{\text{coh}}$ , the S/N across an entire scan is many times larger so that  $\Delta\psi$  is accurately measured. The same alignment procedure can be used to stack data from multiple independently-processed bands when available.

### 2.3.3. Phase Model

Atmospheric phase is assumed to follow a random stochastic process due to a turbulent cascade. In this section we adopt a phase model appropriate for a single station, although the atmospheric phase corrections will cover the combined effects on a baseline. The model itself is used to set tuning parameters and needs to reflect broadly the ensemble behavior rather than be exact. The phase variation is captured by the phase structure function, which typically follows a power-law profile over a wide range of scales,

$$D_\psi(t) = \langle [\psi(t' + t) - \psi(t')]^2 \rangle \approx (t/\mathcal{T}_{\text{coh}})^\alpha. \quad (14)$$

In this representation, the coherence timescale  $\mathcal{T}_{\text{coh}}$  is the time after which phase is expected to drift on average by 1 radian. The power law index  $\alpha$  will be modified at large scales (where energy is injected) and small scales (where energy is dissipated), but these limits are typically outside the primary timescale range of interest – from the minimum useful integration time up to the duration of a scan. For 2D Kolmogorov turbulence  $\alpha = 2/3$ , and for 3D Kolmogorov turbulence  $\alpha = 5/3$  (Thompson et al. 2017). Measured values of  $\alpha$  generally lie somewhere in-between. The corresponding power spectrum is

$$S_\psi(f) = \Gamma[1 + \alpha] \sin\left(\frac{\pi\alpha}{2}\right) \mathcal{T}_{\text{coh}} (2\pi\mathcal{T}_{\text{coh}}f)^{-1-\alpha}, \quad (15)$$

which is related to the structure function through the autocorrelation

$$C_\psi(t) = \langle [\psi(t' + t)\psi(t')]^2 \rangle = C_\psi(0) - D_\psi(t)/2, \quad (16)$$

and its Fourier transform

$$S_\psi(f) = \int_{-\infty}^{\infty} dt C_\psi(t) e^{-2\pi ift}. \quad (17)$$

Phase estimation is done using an atmospheric phase model drawn from a Savitzky-Golay (`savgol`) filter (Savitzky & Golay 1964) applied to the visibility data, which is a running piecewise polynomial fit that has a convenient implementation in Scipy. The filter acts as a symmetric low-pass linear filter for regularly spaced data (Schafer 2011). Real and imaginary components of the complex visibility time-series are filtered separately, and the filtered visibilities are used to derive a smoothly-varying interpolated phase estimate over time at the location of each data value.

The `savgol` filter fits an  $n$ -degree polynomial over a window length  $\mathcal{T}_{\text{win}}$ , so that the effective integration

time  $\mathcal{T}_{\text{dof}}$  per degree-of-freedom is  $\mathcal{T}_{\text{win}}/(n+1)$ . The statistical phase noise for a measurement taken over  $\mathcal{T}_{\text{dof}}$  is approximately

$$\sigma_{\psi,\text{thermal}}^2 \approx (\rho_1^2 \mathcal{T}_{\text{dof}})^{-1}, \quad (18)$$

where  $\rho_1$  is the S/N in 1-second of accumulation, and we have ignored impact on S/N from coherence loss from atmospheric phase drift over the integration period.

In addition to statistical noise, there is residual phase noise from the inability of the smoothed model to capture true rapid phase variations. The residual noise after filtering by window function  $w(f)$  can be calculated from integrating residual power in the frequency domain:

$$\sigma_{\psi,\text{residual}}^2 = \int_{-\infty}^{\infty} df [1 - w(f)]^2 S_\psi(f). \quad (19)$$

For a boxcar moving-average filter of length  $\mathcal{T}_{\text{dof}}$  (equivalent to `savgol` filter of degree zero), the window function is  $\text{sinc}(\pi f \mathcal{T}_{\text{dof}})$  and the residual power is

$$\sigma_{\psi,\text{residual}}^2 = \frac{2^{-\alpha}(2 + \alpha - 2^\alpha)}{(1 + \alpha)(2 + \alpha)} \left(\frac{\mathcal{T}_{\text{dof}}}{\mathcal{T}_{\text{coh}}}\right)^\alpha. \quad (20)$$

Other window functions such as ideal low-pass and Gaussian give equally simple expressions, and all scale as  $(\mathcal{T}_{\text{dof}}/\mathcal{T}_{\text{coh}})^\alpha$ . The boxcar response is a reasonable approximation to `savgol` filters of low non-zero degree.

The effective averaging time  $\mathcal{T}_{\text{dof}}$  which minimizes total error  $\sigma_{\psi,\text{thermal}}^2 + \sigma_{\psi,\text{residual}}^2$  is

$$\mathcal{T}_{\text{dof}} = \left[ \frac{(1 + \alpha)(2 + \alpha)}{2^{-\alpha}\alpha(2 + \alpha - 2^\alpha)\rho_1^2} \mathcal{T}_{\text{coh}}^\alpha \right]^{1/(\alpha+1)}, \quad (21)$$

which is close to the  $\mathcal{T}_{\text{dof}}$  where  $\sigma_{\psi,\text{thermal}}^2 = \sigma_{\psi,\text{residual}}^2$ . We use this optimal  $\mathcal{T}_{\text{dof}}$  to set the parameters of the `savgol` filter within the constraints of the filter construction (filter length  $N_{\text{savgol}}$  in units of the correlator accumulation period  $\mathcal{T}_{\text{AP}}$  is odd and equal to or greater than polynomial degree  $d$ ),

$$N_{\text{savgol}} = \max \left\{ 1 + d, 1 + 2 \left\lfloor \frac{(1 + d) \mathcal{T}_{\text{dof}}}{2 \mathcal{T}_{\text{AP}}} \right\rfloor \right\}. \quad (22)$$

### 2.3.4. Round-Robin Implementation

Atmospheric phase compensation requires a large number of parameters to be derived from data on-source in a regime that is often S/N limited. By restricting the number of fitted degrees-of-freedom based on available S/N, the previously outlined strategy helps to avoid introducing additional noise from over-fitting to mere statistical variations. However some degree of fitting to thermal noise is inevitable, and this can lead to biases in derived quantities – such as a positive bias in coherently averaged visibility amplitude through the introduction of false coherence.

We employ a round-robin (leave-out-one) scheme for phase corrections that ensures that any phase adjustments are derived from data that are disjoint from the data they are being applied to. Because path variations due to the atmosphere are expected to be achromatic over the observing bandwidth, visibilities for each of the  $N$  frequency channels can be phase-stabilized using a smooth atmospheric phase model derived from the remaining  $N - 1$  channels. As long as the number of channels in the data are large (EHT bands are partitioned into 32 corresponding ALMA channels for correlation), the leave-out-one strategy uses most of the available S/N for estimating a phase model and avoids entirely issues of self-tuning. One drawback to the strategy is that it does not transition naturally to making one stable common phase adjustment to all channels in the limit of low S/N (or no correction at all), which is the desired behavior.

### 2.3.5. Second-Order Corrections

Because it is the atmospheric path length variations and not phase variations which we assume are achromatic, a small frequency-dependent adjustment is made to the original unwrapped phase corrections based on the relative difference of the channel frequency to a reference frequency (typically set to the middle of the entire band, and assumed to be representative of the frequency at which estimates are made),

$$\psi \rightarrow \psi \frac{f_{\text{chan}}}{f_{\text{ref}}}. \quad (23)$$

The adjustment can be interpreted as tracking the small non-linear variations in delay that are inferred from measured phase drift.

Residual frequency offsets in the data can also be corrected at this stage through explicit frequency shifting, so long as the frequency shift  $\delta f$  is small compared to the sampling of the data,

$$\psi \rightarrow \psi + 2\pi\delta f t. \quad (24)$$

If left uncorrected, the effects of the frequency offset will instead be fit through a delay-rate compensation  $\delta\tau$ , through the association  $\delta f \leftrightarrow \nu\delta\tau$ . However since the residual fringe rate  $\nu\delta\tau$  varies with observing frequency while the frequency offset  $\delta f$  is fixed, the corrections are not identical and the compensation through fringe rate (essentially stretching or compressing the data in time) imprints a second-order effect that scales with the fractional bandwidth. Thus it is best to measure any residual frequency offset and correct it at correlation or in data pre-processing prior to fitting delay-rate.

## 2.4. RCP-LCP Delay Calibration

Signals that take different analog paths from the receiver to recording elements will be subject to different

delays from cables, clocks, and electronics. The sensitivity of the measured correlation to relative delay depends on the inverse bandwidth – for 1 GHz of bandwidth, the relative delay should be known to much better than 1 ns for sufficient coherence across the band. It is particularly important to delay align RCP and LCP feeds at each antenna, to be able to stationize the (polarization-independent) atmospheric and geometric delay across all four polarization products. It can also be useful to estimate stable instrumental relative delays between frequency bands so that a station-based set of delays is characterized by only a single free delay parameter per station instead of one per station per band. Because components of the receiving system and electronics are generally locked to the same clock reference, instrumental contributions to delay-rate are generally not polarization-dependent and do not need to be relatively calibrated. For the same reason, the instrumental delay calibration is generally stable over time so long as the setup is not disturbed.

During the initial stages of the EHT-HOPS pipeline, the fringe search is unconstrained within some delay (and delay-rate) search window that is wide enough to accommodate the full range of residual geometric, atmospheric, instrumental, and clock errors. Each baseline and polarization product is fit separately to a relative delay and delay-rate. One strategy to align R-L delay at an antenna  $j$  is to measure the relative delay to RCP and LCP feeds at a site given a common reference signal (e.g. LCP at some other station  $i$ ). This requires some amount of linear polarization in the source to produce a cross-hand fringe in addition to the parallel-hand fringe. Then, for example,

$$\tau_{j,\text{R-L}} = \tau_{ij,\text{LR}} - \tau_{ij,\text{LL}} \quad (25)$$

with  $\tau_{ij,\text{LR}}$  and  $\tau_{ij,\text{LL}}$  as the measured baseline relative delays measured for polarization products LR and LL, and  $\tau_{j,\text{R-L}}$  the inferred relative delay between RCP and LCP at station  $j$ . The measurement can be averaged over all available reference signals for increased accuracy.

One drawback to the reference signal strategy is that detected fringes in the cross-hand polarization products are sensitive to polarization leakage since both the typical magnitude of leaked power and the degree of linear polarization are often of the same magnitude. Therefore, prior to polarization leakage calibration, parallel-hand correlated signal can leak into the cross-hand measurement and introduce significant noise in the delay measurement.

When ALMA is present in the array, it can be used to measure RCP-LCP delay at other stations using only parallel-hand products to ALMA. This is because the ALMA linear feeds are delay and phase aligned through ALMA quality assurance calibration (Goddi et al. 2019). The PolConvert process converts ALMA's mixed-polarization products to circular polarization, maintaining the zero relative delay between ALMA con-

verted RCP and LCP (Martí-Vidal et al. 2016). Then,

$$\tau_{j,R-L} = \tau_{A_j,RR} - \tau_{A_j,LL} \quad (26)$$

with A for ALMA. Since the R–L instrumental delays are generally stable through the night, ALMA only needs to be present in a subset of scans in order to fully R–L delay calibrate the network. The basic strategy at stage 4 of the pipeline is therefore to take an average of ALMA parallel-hand detections to other stations to derive a single RCP–LCP delay offset for each non-ALMA site on each observing night. The average itself is a  $1/\sigma^2$  weighted mean, after accounting for a small amount of systematic delay error and after rejecting  $10\text{-}\sigma$  outliers from the median value. Further validation steps check that the constant offset is a good model to within thermal error plus small systematic tolerances.

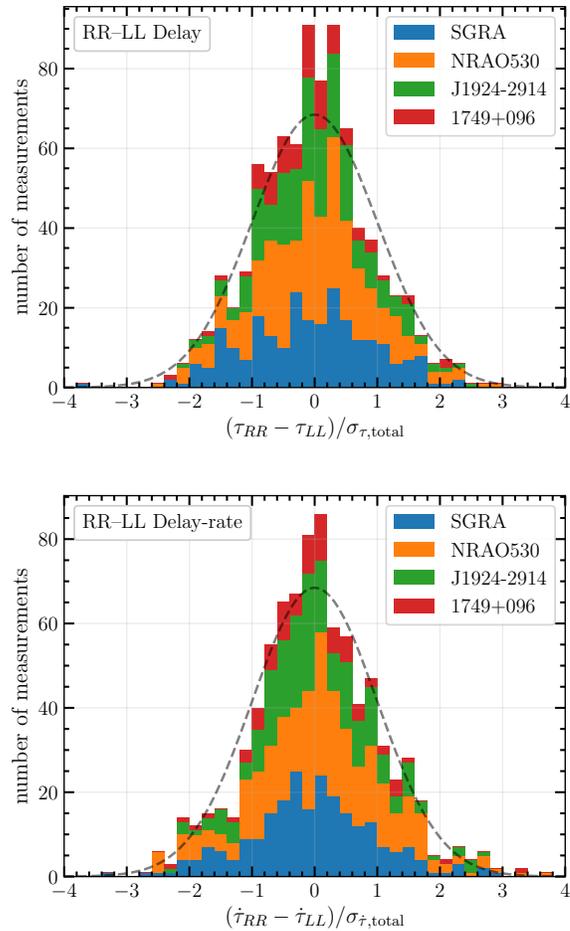
Figure 6 shows the distribution of all measured RR–LL delay differences between ALMA and other stations after calibrating out a constant delay offset between RCP and LCP feeds at non-ALMA sites. The fact that all measured differences are consistent with zero confirms the assumed stability of RCP versus LCP relative delay at each site.

### 2.5. Global Fringe Solution

After stage 4, fitted delays and delay-rates on each baseline are expected to be the same for all polarization products to within measurement thermal noise. This allows us to *stationize* the fitted delay and delay-rate parameters, modeling each as the difference between a pair of station delays and delay-rates. The stationization of the fringe solution provides several benefits: it prevents the first-order fringe correction from introducing non-closing (not station based) phase adjustments to the data, it reduces the total number of free parameters describing the corrections from a number that scales with the  $\mathcal{O}(N^2)$  baselines to a number that goes as  $\mathcal{O}(N)$  stations, and it allows fringe locations to be accurately predicted on baselines which may have no independently detectable correlated signal.

The thermal contribution to errors in the estimation of delay and delay-rate are directly related to the noise in a measurement of total accumulated phase drift across bandwidth  $\Delta\nu$  and time  $\Delta t$ . At moderate S/N and near the true fringe peak, the error is approximately  $\sqrt{12}/\rho$ , where  $1/\sqrt{12}$  is the standard deviation of a uniform distribution corresponding to the flat integration period, and  $1/\rho$  represents thermal noise in the phase measurement (Thompson et al. 2017, A12.25). Therefore,

$$\sigma_{\tau,\text{thermal}} = \frac{1}{2\pi} \frac{\sqrt{12}}{\rho \Delta\nu} \quad \sigma_{\dot{\tau},\text{thermal}} = \frac{1}{2\pi\nu} \frac{\sqrt{12}}{\rho \Delta t}. \quad (27)$$



**Figure 6.** Distribution of differences between calibrated RR and LL measured delays and delay-rates for all scans in the test data set with  $S/N > 7$ , in units of expected total measurement error for the difference (taken in quadrature) from Equation 28 with 10 ps systematic in delay and 0.25 fs/s in delay-rate. The dotted line corresponds to a standard normal distribution, expected if the constant delay model is valid to within total Gaussian measurement uncertainties. 10 ps corresponds to a negligible  $4 \times 10^{-5}$  fractional coherence loss over the 256 MHz GMVA bandwidth, and the same for a residual 0.25 fs/s delay-rate error over 240 seconds of integration at 86 GHz (Equation 35). The successful fitting of parallel hand delay differences using a constant model with zero statistically significant outliers indicates good parallel-hand fringe solutions for both strong and weak sources, as well as the stability of individual station instrumental delays through the night. There are a small number (2%) of delay-rate difference outliers due to the current limitation of ad-hoc phasing to a single reference station, meaning some weaker isolated baselines may not be able to be phase stabilized. A single station delay-rate is still enforced at the global fringe solution, but for the original delay-rate outliers there could be residual coherence loss under a full scan-average.

In addition to the thermal error, we can add some level of systematic error to fitted delay and delay-rate,

$$\sigma_\tau^2 = \sigma_{\tau,\text{thermal}}^2 + \sigma_{\tau,\text{sys}}^2 \quad (28)$$

$$\sigma_{\dot{\tau}}^2 = \sigma_{\dot{\tau},\text{thermal}}^2 + \sigma_{\dot{\tau},\text{sys}}^2, \quad (29)$$

which may be baseline and polarization dependent. The systematic errors arise from search resolution and interpolation accuracy, contamination from leaked signal power (particularly in cross-hand products), or other baseline-dependent processing artifacts, and in general must be estimated from the data.

For a fringe search which fits delay and delay-rate to values which maximize total detected fringe power  $\rho_0^2$ , we must also consider the probability of false-positive, i.e., one minus the probability that all of  $N$  independent noise measurements across the search space in delay and delay-rate are less than the measured value. Thermal noise gives an exponentially distributed random contribution to fringe power, so the probability of a false positive over  $N$  trials (Thompson et al. 2017) is

$$\prod_i^N P(\rho_i < \rho_0) = 1 - \left[ 1 - \exp\left(-\frac{\rho_0^2}{2}\right) \right]^N \quad (30)$$

$$\approx N \exp\left(-\frac{\rho_0^2}{2}\right). \quad (31)$$

This is also the cumulative (survival) distribution of the maximum noise fringe power over  $N$  independent measurements. A requirement that the false positive rate be very low (much less than the number of fits performed) sets a threshold  $\rho_{\text{thr}} \sim 7$  above which detections are considered reliable.

Following Alef & Porcas (1986), we take the estimated baseline and polarization-dependent delay and delay-rate solutions along with their errors from Equation 28, and then perform a least-squares fit to station-based parameters. For each scan, one delay and delay-rate parameter is fit per antenna that minimizes the squared error across all baseline measurements. Measurements with  $\rho_0 < \rho_{\text{thr}}$  are assigned a very large  $\sigma_{\text{sys}}$  so that they are effectively ignored in the presence of any other constraining data. The least-squares minimization is performed in Scipy with an additional `soft_l1` loss function applied at scale  $f = 8\sigma$  to mitigate the effects of outliers.

Specifically the least-squares approach solves for, e.g., model station delays  $\tau_i$  (and delay-rates  $\dot{\tau}_i$ ) by minimizing a chi-square error function,

$$\chi^2 = \sum_{i < j, k} L \left[ \frac{(\tau_{ij,k} - (\tau_j - \tau_i))^2}{\sigma_{\tau,ij,k}^2} \right] \quad (32)$$

where  $i < j$  loops over baselines,  $k$  indexes the four polarization products,  $\tau_{ij,k}$  is the measured delay for each baseline/polarization,  $\tau_i$  and  $\tau_j$  are not dependent

on polarization due to the previous step of delay calibration,  $\sigma_\tau$  is total error as described in Equation 28, and  $L(z^2) = 2f^2(\sqrt{1+z^2/f^2}-1)$  is the `soft_l1` loss function as implemented in Scipy. The best-fit station delays and delay-rates are used to model baseline fringe parameters,

$$\tau_{ij} = \tau_j - \tau_i \quad \text{and} \quad \dot{\tau}_{ij} = \dot{\tau}_j - \dot{\tau}_i, \quad (33)$$

which are then applied to the data for the global fringe solution (as zero-width search windows).

Expanding the fringe amplitude (Equation 4) to second order about zero total phase drift,

$$\frac{|r_{\text{off-fringe}}|}{|r_{\text{ideal}}|} \sim 1 - \frac{(\Delta\psi)^2}{24} \quad (34)$$

so that a total phase drift of  $\sqrt{0.24}$  radians corresponds to a 1% amplitude loss. The expected amplitude loss at a fringe solution based on a measurement of S/N  $\rho$  is  $1/(2\rho^2)$  each for delay and rate errors and not including any noise bias. Propagating fringe solutions with S/N of 7 and above will maintain sub 1% amplitude loss.

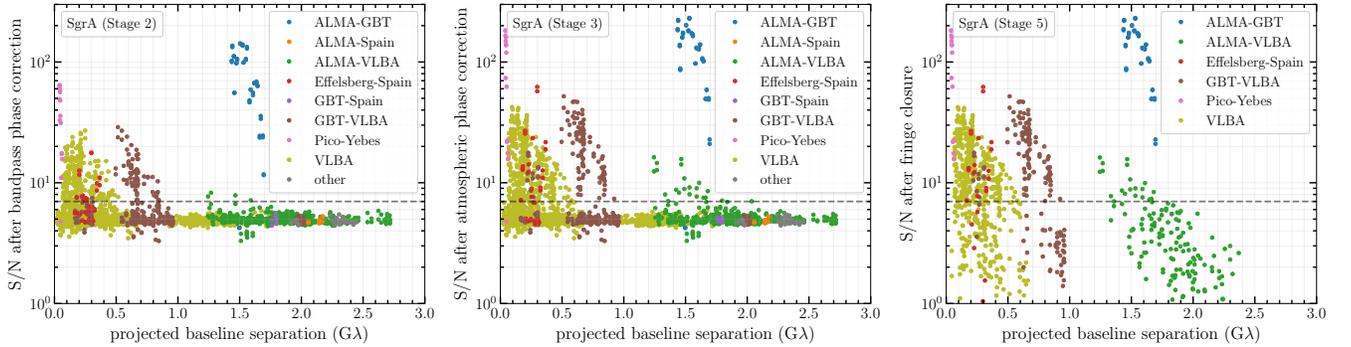
In terms of errors on delay  $\tau$  and rate  $\dot{\tau}$  directly, the amplitude efficiency loss factor is

$$\frac{(2\pi\tau\Delta\nu)^2}{24} \quad \frac{(2\pi\nu\dot{\tau}\Delta t)^2}{24}. \quad (35)$$

To maintain sub 1% amplitude loss for an observing bandwidth of  $\Delta\nu = 2$  GHz at observing frequency  $\nu = 220$  GHz, delay must be within 0.04 ns and rate must be within 0.07 ps/s for  $\Delta t = 10$  second coherent integration. These limits are within typical systematic errors seen in real data (e.g., Figure 6).

Not all baselines are constrained by the global fringe solution. If a station has no reliable ( $\rho_0 > \rho_{\text{thr}}$ ) detections to other stations in the array, its relative delay and delay-rate to other sites remain unconstrained. Stations in the array are partitioned into fringe groups. Each group represents a set of mutually connected stations, where stations are connected through one or more baselines where at least one polarization product gives fringe detection with  $\rho_0 > \rho_{\text{thr}}$ . Baselines between stations which belong to different fringe groups are flagged from the data and removed, so that the only surviving correlation measurements are those which are evaluated at single well-constrained fringe locations.

As noted by Alef & Porcas (1986), the least-squares global fringe fit derived from initial baseline-based fringe solutions is not as powerful (in terms of optimal S/N) as the coherent global fringe search of Schwab & Cotton (1983). However for our purposes it offers a few advantages. For one, the baseline-based fringe search with independent solutions per baseline and per polarization product are extremely useful for using delay consistency to test for instrumental artifacts, data issues, and false fringes, for which Figure 6 provides one example. Second, the baseline-based fringe solution are immune to



**Figure 7.** Scan-average S/N as a function of projected  $(u, v)$  distance for GMVA+ALMA 3.5 mm observations of Sgr A\*. The three panels represent fringe solutions from successive stages of the EHT-HOPS pipeline, before and after atmospheric phase correction and after a station-based global fringe solution has been applied. From the left to middle panel, the removal of non-linear phase variations increases scan-average S/N by preserving phase coherence across the scan. Both stages, however, suffer from a noise floor at  $S/N \sim 5$  due to the large number of trials over the delay and delay-rate search window. The horizontal dashed line at  $S/N = 7$  represents a threshold at which fringe solutions are considered reliable (very unlikely to have arisen from a noise fluctuation). These confident detections are used to solve a global fringe solution across the entire array, reducing the number of effective trials for weak baselines to 1 when the fringe solution can be constrained by other detections. This allows measurement down to an uncertainty of  $S/N \sim 1$  (right panel), where we can see a clear decrease in S/N with increasing baseline length on long ALMA–VLBA baselines (generally connected through ALMA $\leftrightarrow$ GBT $\leftrightarrow$ VLBA).

biases toward an assumed source model (Wielgus et al. 2019), as they do not use a source model. We note that the round-robin strategy as outlined in Section 2.3.4 could also be used to avoid amplitude and phase biases in the Schwab-Cotton method. The difference in sensitivity between the baseline-based search and coherent global fringe search is not large for the EHT and GMVA because both arrays have relatively few stations (the difference between the  $\mathcal{O}(N^2)$  baseline fit parameters and  $\mathcal{O}(N)$  station parameters is not so large, and can be made up by other optimizations such as optimal fringe solution intervals), and because both arrays are highly heterogeneous, with fringe solutions driven primarily by baselines to anchor stations such as ALMA or GBT.

### 3. POSTPROCESSING

The EHT-HOPS pipeline is naturally divided into the initial stages 1–5, where iterations of HOPS `fourfit` fringe fitter are performed with increasing refinement of the initial phase calibration, and a series of post-processing stages 6–9 which operate on the `fourfit` output. The first step (stage 6) in post-processing is to convert the `fourfit` native binary output data into standard UVFITS (Greisen 2012) format, using interfaces developed as part of the `eat` library for accessing and interpreting the HOPS Mark4 fileset, as well as UVFITS interfaces originally developed for use in the `eht-imaging` library (Chael et al. 2016, 2018). The data conversion routines are packaged as part of the `eat` library, and provide a direct conversion of the HOPS “type-2 fringe” files into corresponding UVFITS format. The fringe files include all calibration corrections from stages 1–5 of the EHT-HOPS pipeline applied, includ-

ing the fringe solution and atmospheric phase corrections, and are provided at the channel-averaged “fourfit output” time and frequency resolution described in Figure 2. This level of averaging is maintained until the final network calibration stage 9, at which data are further averaged (typically full band, 10-second averages) for a more convenient data volume. The post-processing stages 7–9 which follow read and write UVFITS formatted data directly, and apply amplitude calibration and polarization-and-time dependent phase corrections which currently cannot be applied upstream within the HOPS framework. Because they operate on standard UVFITS output, the postprocessing routines have some general utility even outside the EHT-HOPS pipeline.

#### 3.1. *A Priori Flux Density and Field Angle Calibration*

The sensitivity of each telescope is expressed by the *system-equivalent flux density* (SEFD), which represents the flux (Jy) of an unpolarized astronomical source that would be necessary to produce a received power equal to the system noise power (as in Equation 5). It can be estimated from observations of bright primary calibration targets (such as planets), and a calibrated measurement of atmospheric and receiver noise. At millimeter wavelengths, atmospheric noise and attenuation due to opacity are often substantial, so that SEFD can have a strong dependence on elevation.

The EHT-HOPS pipeline relies on SEFD information delivered from each telescope, provided in the form of ANTAB<sup>6</sup> formatted data tables. The ANTAB ta-

<sup>6</sup> <http://www.aips.nrao.edu/cgi-bin/ZXHLP2.PL?ANTAB>

bles provide SEFD information in the form of a constant degrees-per-flux-unit (DPFU) value encoding dish area and efficiency, an elevation-dependent parameterized gain curve efficiency correction  $\eta_{\text{el}}$ , and a time-dependent effective system noise temperature  $T_{\text{sys}}^*$  scaled according to the expected level of atmospheric attenuation through line-of-sight opacity  $\tau$ . While millimeter observatories generally estimate  $T_{\text{sys}}^*$  directly via the “hot-load” calibration technique (Penzias & Burrus 1973; Ulich & Haas 1976), centimeter-wave observatories, such as the majority of stations in the GMVA (even while observing at  $\sim$ several mm), measure the system noise temperature  $T_{\text{sys}}$  directly from calibrating to a noise diode injection, which does not account for atmospheric attenuation. In this case, opacity  $\tau$  is estimated in the line of sight using the measured  $T_{\text{sys}}$  and an estimate of the receiver noise temperature  $T_{\text{rx}}$  and physical temperature of the atmosphere  $T_{\text{atm}}$  (e.g., Altshuler et al. 1968),

$$T_{\text{sys}} \approx T_{\text{rx}} + (1 - e^{-\tau}) T_{\text{atm}}, \quad (36)$$

and then used to scale  $T_{\text{sys}}^* = e^{\tau} T_{\text{sys}}$  appropriately, as described in Issaoun et al. (2017).

For each source, the  $T_{\text{sys}}^*$  values are interpolated into the observation times, and SEFD is calculated as

$$\text{SEFD} = \frac{T_{\text{sys}}^*}{\text{DPFU} \times \eta_{\text{el}}}. \quad (37)$$

The SEFD calibration tables are then used to amplitude calibrate visibilities in the UVFITS formatted data to a physical flux scale within the eat postprocessing framework,

$$V_{ij} = \eta_Q^{-1} \sqrt{\text{SEFD}_i \times \text{SEFD}_j} r_{ij}, \quad (38)$$

where  $r_{ij}$  is the correlation coefficient as in Equation 1.

Apart from flux scaling of visibility amplitudes, this stage of calibration also corrects for the a priori polarimetric field rotation angle, i.e., the relative orientation of the feed with respect to a fixed direction on the sky. The effect manifests as a nonlinear, source and time dependent phase offset between RCP and LCP components at each station, see Figure 8, top panel. The field rotation angle  $\varphi_j$  is generally a combination of the source parallactic angle at the location of the  $j$ -th station and a possible contribution from elevation-dependent rotation due to the receiver mount type. The correction takes the form of a station-based polarization-dependent phase correction to Equation 38 to align RCP and LCP to a fixed orientation on the sky. Figure 8, middle panel, shows the R–L phase offsets after applying the field rotation correction.

### 3.2. RCP/LCP Polarimetric Gain Ratios Calibration

In order to form total intensity Stokes  $I$  visibilities, it is necessary to calibrate the phase and amplitude

mismatch between the measured LCP and RCP components. For small Stokes  $V$  component and small leakage coefficients, the LCP and RCP visibilities are approximately related as (e.g., Roberts et al. 1994)

$$V_{jk,\text{RR}} \approx \left( \frac{g_{j,\text{R}}}{g_{j,\text{L}}} \right) \left( \frac{g_{k,\text{R}}}{g_{k,\text{L}}} \right)^* e^{-2i(\varphi_j - \varphi_k)} V_{jk,\text{LL}}. \quad (39)$$

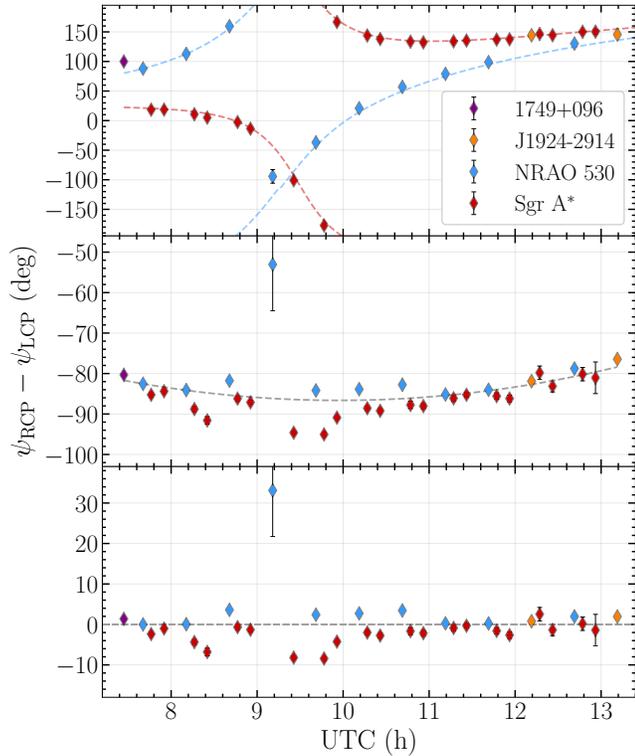
The field rotation component  $\exp[-2i(\varphi_j - \varphi_k)]$  is removed at the a priori calibration stage, Section 3.1, leaving the complex gain ratios  $g_{\text{R/L}} \equiv g_{\text{R}}/g_{\text{L}}$  to be accounted for. Polarimetric gain ratio phases are particularly important to calibrate. The RCP vs LCP phase stability is analogous to the RCP vs LCP delay stability (Section 2.4), but more sensitive by a factor corresponding to the inverse fractional bandwidth (e.g., two orders of magnitude more sensitive). Thus, while relative RCP vs LCP instrumental delays are generally constant throughout the night, relative instrumental phases can exhibit some residual drift.

The station based phase offsets,  $\phi_{\text{R-L}}(t) \equiv \text{Arg}[g_{\text{R/L}}]$ , are modeled as polynomial functions of time and are estimated directly from Equation 39 with respect to a reference station. If the reference station gain ratio  $g_{\text{R/L}}$  is known, or can be derived a priori, as is the case for ALMA, this enables absolute calibration of the electric vector position angle on the sky for the entire array. While the polynomial fit parameters are estimated from the data using robust, S/N-weighted statistics, the algorithm requires a manual selection of a polynomial degree used for a phase offset fit  $\phi_{j,\text{R-L}}(t)$  for particular station. In a heterogeneous VLBI array, the type of fit depends on particular properties of each station, which may vary from a constant offset for multiple subsequent nights to a nonlinear trend varying on timescales of an hour. When available, we jointly analyze observations of multiple sources (e.g., scientific target and calibrators) when estimating source-independent station phase gain offsets over the course of a campaign. This makes the estimate robust against tuning to specific intrinsic source properties, such as contamination from a nonzero Stokes  $V$  circular polarization component.

As an illustration,  $\phi_{\text{R-L}}(t)$  of the GBT, estimated from the GMVA+ALMA data set (see Section 5.1) is shown in Figure 8, middle panel, with a dashed line. Here the offset was modeled as a second order polynomial and estimated using data set consisting of four observed sources. For the phase gain offset calibration, one polarization component is chosen as a reference (LCP by default) and the other one is calibrated to match the first one. As an example we have

$$\begin{aligned} V_{jk,\text{RR}} &\rightarrow V_{jk,\text{RR}} \exp[-i(\phi_{j,\text{R-L}} - \phi_{k,\text{R-L}})], \\ V_{jk,\text{LL}} &\rightarrow V_{jk,\text{LL}}. \end{aligned} \quad (40)$$

A final product of the polarimetric phase offset calibration is shown in Figure 8, bottom panel. In the top



**Figure 8.** Steps of RCP–LCP phase offset calibration illustrated for the ALMA–GBT baseline in the GMVA+ALMA data set (Section 5.1). Top: phase offset after the fringe fitting step; center: phase offset after parallactic angle correction; bottom: phase offset after global fitting of polarimetric gain ratio phases. The GBT phase gain is fitted as a second order polynomial to a data set consisting of all four sources. In each case, the dashed line represents the full RCP–LCP phase model.

panel we also show the full polarimetric phase offset calibration model fit (field rotation plus gains) for Sgr A\* and NRAO 530 as dashed lines.

While the flux calibration, described in Section 3.1, is performed separately for different polarization products, and is expected to account for a priori known differences in sensitivity between RCP and LCP at each site, the `eat` postprocessing framework also offers an option to calibrate residual differences in RCP/LCP amplitude gain. If the option is selected, amplitude gain is estimated as a S/N-weighted median RCP/LCP amplitude ratio  $|g_{j,R/L}|$ . Calibrating polarimetric amplitudes for the  $jk$  baseline yields

$$\begin{aligned} V_{jk,RR} &\rightarrow V_{jk,RR} |g_{j,R/L}|^{-1/2} |g_{k,R/L}^*|^{-1/2}, \\ V_{jk,LL} &\rightarrow V_{jk,LL} |g_{j,R/L}|^{1/2} |g_{k,R/L}^*|^{1/2}. \end{aligned} \quad (41)$$

The presented framework assumes a negligible influence of polarimetric leakage, the calibration of which is not yet a standard part of the EHT-HOPS data calibration pipeline. Proper calibration of leakage necessarily

relies on the joint modeling of leakage terms together with both polarized and unpolarized source structure (Leppanen et al. 1995).

### 3.3. Network Amplitude Calibration

The a priori amplitude calibration of the EHT-HOPS pipeline (Section 3.1) can be improved by determining station-based corrections that produce visibility amplitude relationships that are expected from array redundancy. While array redundancy has regularly been used to improve calibration of connected element arrays, it has not been commonly used for VLBI (see, e.g., Pearson & Readhead 1984; Cornwell & Fomalont 1989). However, the EHT differs from standard VLBI arrays by including a number of co-located sites that introduce significant redundancy (e.g., three different facilities on Mauna Kea have participated in EHT observations: the Submillimeter Array (SMA), the James Clerk Maxwell Telescope (JCMT), and the Caltech Submillimeter Observatory), and this redundancy has been routinely utilized to derive amplitude calibration corrections (e.g., Fish et al. 2011; Johnson et al. 2015; Lu et al. 2018). We refer to the procedure of deriving these corrections as *network calibration*, and the EHT-HOPS implementation of network calibration is an extension of techniques developed in Johnson et al. (2015). We now outline the assumptions, procedure, and limitations of network calibration.

#### 3.3.1. Network Calibration Assumptions

Consider a VLBI array that contains one or more pairs of stations at a single geographic site (e.g., ALMA/APEX and SMA/JCMT for the EHT). Because intra-site baselines do not resolve the compact emission structure sampled on inter-site baselines, they introduce consistency relationships that are weakly dependent on source structure. For example, letting  $\mathcal{V}_x$  denote the ideal source visibility on the baseline  $x$ ,

- Intra-site visibilities should be equal to each other and to measurements of the total flux density  $V_0$  made with connected element interferometers that sample the same angular scales; e.g.,

$$\mathcal{V}_{\text{SMA-JCMT}} = \mathcal{V}_{\text{ALMA-APEX}} = V_0 \in \mathbb{R}^+. \quad (42)$$

- Inter-site visibilities to intra-site stations should be equal; e.g.,

$$\mathcal{V}_{\text{LMT-SMA}} = \mathcal{V}_{\text{LMT-JCMT}}. \quad (43)$$

Both of these properties follow from the assumption that intra-site baselines do not resolve the source; the first relationship integrates an additional measurement ( $V_0$ ), which is routinely recorded in parallel with VLBI observations.

### 3.3.2. Network Calibration Procedure

To motivate the network calibration procedure, we first consider visibility measurements with no thermal noise. Under the assumption that all systematic errors are station-based, we can write a measured visibility  $V_{ij}$  on a baseline between sites  $i$  and  $j$  as

$$V_{ij} = g_i g_j^* \mathcal{V}_{ij}, \quad (44)$$

where  $g_i$  and  $g_j$  are the station-based residual gains.

Suppose that stations  $i$  and  $j$  are co-located, so that  $\mathcal{V}_{ij} = V_0$ . Knowledge of  $V_0$  is not sufficient to determine  $g_i$  and  $g_j$ , but measurements to a third site  $k$  break the degeneracy:

$$\begin{aligned} |g_i| &= \sqrt{\left| \frac{V_{ij}}{V_0} \times \frac{V_{ik}}{V_{jk}} \right|}, \\ |g_j| &= \sqrt{\left| \frac{V_{ij}}{V_0} \times \frac{V_{jk}}{V_{ik}} \right|}. \end{aligned} \quad (45)$$

As these equations suggest, network calibration only determines the amplitudes of the gains of stations with co-located partners; it does not modify the gains of other stations or determine absolute phase corrections. In the limit of all stations having a co-located partner, network calibration yields absolute amplitude calibration for all stations.

Because the gains of an intra-site pair are fully determined by a third site, additional sites can be combined to reduce thermal uncertainties in the estimated gains. The EHT-HOPS pipeline uses all baselines simultaneously to solve for the set of unknown model visibilities  $\mathcal{V}_{ij}$  and station gains  $g_j$  by minimizing an associated  $\chi^2$ . Specifically, for each solution interval, we find the set of gains  $\{g_i\}$  and source visibilities  $\{\mathcal{V}_{ij}\}$  connecting each pair of sites by minimizing

$$\chi^2 = \sum_{i < j} \frac{|g_i g_j^* \mathcal{V}_{ij} - V_{ij}|^2}{\sigma_{ij}^2}, \quad (46)$$

where  $\sigma_{ij}$  is the thermal uncertainty on  $V_{ij}$ . In practice, the only gains that must be included are those of sites with intra-site partners; also, visibilities connecting two sites that each lack an intra-site partner can be excluded as they provide no additional constraints for the network calibration. Thus, for  $N$  sites of which  $N_{\text{intra}}$  have intra-site partners, network calibration requires solving for at most  $N_{\text{intra}}$  gains and  $(N - N_{\text{intra}}/2)(N - N_{\text{intra}}/2 - 1)/2$  model visibilities. In 2017, the EHT had  $N = 8$  and  $N_{\text{intra}} = 4$ , requiring solutions for at most 4 gains and 15 model visibilities in each solution interval.

We implemented this procedure for network calibration within the `eht-imaging` library (Chael et al. 2016, 2018). This calibration can be used for any VLBI array and only requires specifying the total source flux density  $V_0$  and a maximum baseline separation for which a

pair of sites is considered co-located (i.e., a threshold to define baseline lengths that do not resolve the observed source).

### 3.3.3. Network Calibration Error Budget

The outcome of network calibration has both thermal errors from thermal noise on the input visibilities and systematic errors from broken assumptions in the network calibration procedure. We now briefly assess expected elements of this error budget.

Thermal errors are straightforward to compute from analysis of the  $\chi^2$  hypersurface explored in the minimization procedure discussed in Section 3.3.2. From Equation 45, it is clear that thermal errors have contributions from intra-site baselines and from inter-site baselines; the latter typically have lower S/N because these baselines can heavily resolve the source, so they typically dominate the thermal error budget. For each pair of co-located sites  $\{i, j\}$ , the fractional uncertainty from thermal noise will be dominated by their strongest inter-site baselines to another site  $k$ , with

$$\frac{\Delta g_i}{g_i}, \frac{\Delta g_j}{g_j} \sim \frac{1}{2} \sqrt{\left( \frac{\sigma_{ij}}{|V_{ij}|} \right)^2 + \left( \frac{\sigma_{ik}}{|V_{ik}|} \right)^2 + \left( \frac{\sigma_{jk}}{|V_{jk}|} \right)^2}. \quad (47)$$

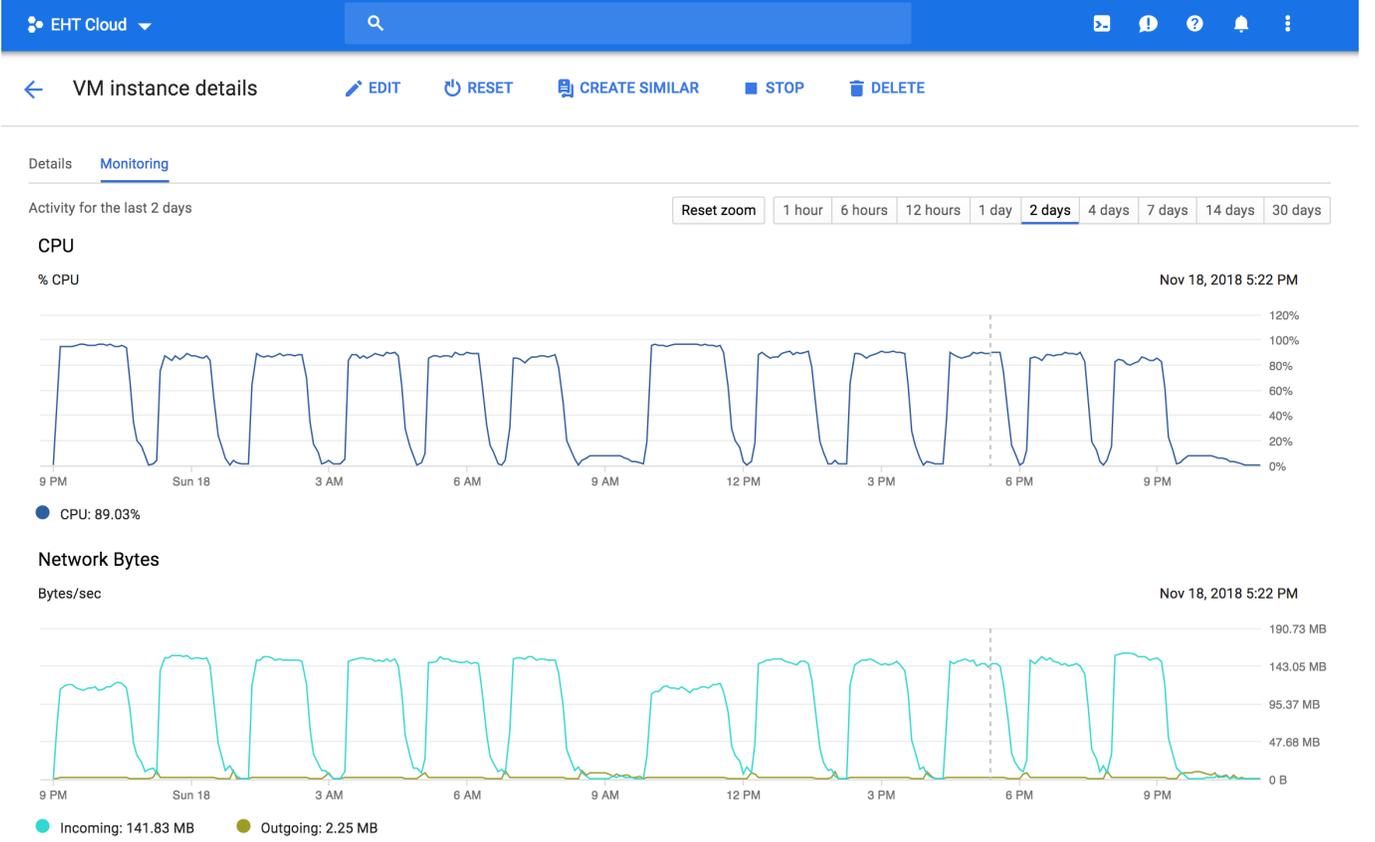
For instance, in the case of ALMA and APEX, APEX will always have much lower S/N than ALMA and dominates the thermal uncertainties for the derived gains of both stations. If the maximum S/N from APEX to another site is 10 in the network calibration solution interval, then the network calibration will have a fractional uncertainty for  $g_{\text{ALMA}}$  and  $g_{\text{APEX}}$  of approximately 5% from thermal noise.

Systematic errors in the network calibration solution arise from incorrect or broken assumptions (see Section 3.3.1). We now estimate the magnitude of three primary expected errors; additional sources of error (e.g., from baseline-dependent systematic errors) can be assessed using Equation 45.

**Incorrect assumed total flux density:** Errors in the assumed total flux density  $V_0$  lead to a constant multiplicative factor for the derived gains. Suppose that  $V_0 = \mathcal{V}_0 + \Delta V_0$ , where  $\mathcal{V}_0$  is the true total flux density. Then  $\Delta g/g \approx \sqrt{V_0/\mathcal{V}_0} \approx 1 + \frac{1}{2} \Delta V_0/V_0$ .

**Intra-site baselines partially resolve the source:** In practice, intra-site baselines may partially resolve the emission structure. In some cases, it may be possible to model this effect and correct for it (e.g., using an ALMA image to predict the flux density seen on the SMA-JCMT baseline). However, even in the limit of unmodelled losses, the measured flux density on a short baseline will differ from the true value by some amount  $\Delta V_0$ , and the error propagation is identical to the case of an incorrect assumed total flux density.

**Inter-site baselines to intra-site partners are not equal:** Suppose that the intra-site stations are sepa-



**Figure 9.** A screenshot of CPU and network usages from running the EHT-HOPS pipeline on a single 64-core virtual machine on Google Cloud Platform. The data volume processed is 1.7 Terabytes, covering two 2 GHz bands from the EHT (up to 8 stations, dual-polarization). Each pass processes one of the bands in about 12 hours, which includes one bootstrap stage with generic parameters and wide search windows, followed by the five fringe fitting stages of the EHT-HOPS pipeline (Figure 1). Each stage performs a fringe search over the complete data set. Because the `Mark4` formatted correlator input data is separated by scan, the processing at each stage can be naturally parallelized over multiple scans.

rated by a vector displacement  $\mathbf{u}_{\text{intra}}$ , and let  $\mathbf{u}_{\text{inter}}$  denote an inter-site baseline to a site that has an intra-site pair. In this case, network calibration relies on the approximation that the two inter-site visibilities for the pair are approximately equal:  $\mathcal{V}(\mathbf{u}_{\text{inter}}) \approx \mathcal{V}(\mathbf{u}_{\text{inter}} + \mathbf{u}_{\text{intra}})$ . By the van Cittert-Zernike theorem, this condition can be expressed in terms of the sky brightness distribution (Thompson et al. 2017):

$$\begin{aligned} & \int d^2\mathbf{x} I(\mathbf{x}) e^{-2\pi i\mathbf{x} \cdot \mathbf{u}_{\text{inter}}} \\ & \approx \int d^2\mathbf{x} I(\mathbf{x}) e^{-2\pi i\mathbf{x} \cdot (\mathbf{u}_{\text{inter}} + \mathbf{u}_{\text{intra}})} \\ & \approx \int d^2\mathbf{x} I(\mathbf{x}) e^{-2\pi i\mathbf{x} \cdot \mathbf{u}_{\text{inter}}} (1 - 2\pi i\mathbf{x} \cdot \mathbf{u}_{\text{intra}}), \end{aligned} \quad (48)$$

where  $\mathbf{x}$  is an angular coordinate on the sky and  $I(\mathbf{x})$  is the sky brightness distribution. The second approximation requires  $u_{\text{intra}} \ll 1/(2\pi x_{\text{max}})$ , where  $x_{\text{max}}$  is the maximum extent of non-zero source brightness. The fractional error in the first approximation is then of

order  $2\pi\mathbf{x} \cdot \mathbf{u}_{\text{intra}}$ . For the EHT, the longest intra-site baselines are shorter than inter-site baselines by a factor of  $u_{\text{inter}}/u_{\text{intra}} \gtrsim 10^3$ . For sources that are weakly resolved on the shortest inter-site baselines (i.e.,  $2\pi u_{\text{inter}} x_{\text{max}} \lesssim 1$ ) the fractional error on a derived gain from breaking this assumption will then be  $\Delta g/g \lesssim 0.01$ .

#### 4. COMPUTING WORKFLOW

The EHT-HOPS pipeline is designed to be automated and provide reproducible output. The pipeline is conceptually structured in three layers: 1) The software libraries/modules layer consist with the core software packages HOPS, `eat`, and `eht-imaging`. 2) The driver scripts layer consists of `BASH` scripts for preparing input files, running programs from the software layer, creating logs and summary Jupyter notebooks, and cleaning up data products. 3) The pipeline repository layer is made up of multiple directory structures that contain both configuration files for different processing stages (see Figure 1) and a master run script that enables running the full pipeline and packing output data products

in a single step. The software and driver script layers are generic, and are suitable for being applied to different VLBI data sets. The pipeline repository, including summary notebook templates, is associated with the data set being processed.

To ensure reproducibility, software libraries and module layers that are developed within the EHT, such as `eat`, are version controlled by `git` and publicly available on GitHub. Furthermore, we use Docker, an operating-system-level virtualization software, to freeze the entire software environment, which includes many libraries and software packages distributed in binary format. The recipes to build the Docker images, i.e., the Dockerfiles, are also version controlled and available on GitHub.

Although the entire pipeline can be run and debugged interactively on the native host operating system, production runs make use of Docker environments. The associated hash-based Docker image identification numbers allow us to keep track of the exact versions of software, down to system libraries, and the specific image used for each production run is tagged along with its output. This allows us to go back and repeat any previously tagged analysis.

The correlated data are separated by scan in relatively small “type-1 corel” individual files in the Mark4 fileset. When we run the EHT-HOPS pipeline, within each stage, all CPU cores on the (virtual) machine are made available to a single Docker container. Inside the Docker container, we use GNU `parallel` to start multiple `fourfit` tasks, with one scan mapped per task, to maximize CPU utilization. When fringe fitting is done, we use the HOPS `alist` program to reduce the fringe fitting results into a single summary text file. Further tools from `eat` process this output to generate HOPS calibration control codes for the next stage. Figure 9 is a screenshot of CPU and network usage during a production run on a 64-core virtual machine on the Google Cloud Platform. The periods of high CPU and network utilization correspond to the parallel `fourfit` tasks, while the periods of low utilization correspond to the `alist` and `eat` reduction tasks. From the utilization cycles, it is easy to read off from Figure 9 that there are two passes of the data, one for each of the two 2 GHz bands from the EHT. Each pass includes one bootstrap stage with generic parameters and wide search windows, followed by the five fringe fitting stages with refined HOPS control files.

## 5. COMPARISON TO AIPS

### 5.1. Data Set

The data set used for validation of the EHT-HOPS pipeline is the result of observations of Sgr A\* on 2017 April 3 (project code MB007) with the GMVA, composed of the eight VLBA antennas operating at 86 GHz, the Robert C. Byrd Green Bank Telescope (GBT), the Yebes 40m telescope (YS), the IRAM 30m telescope (PV), the Effelsberg 100m telescope (EB), and the

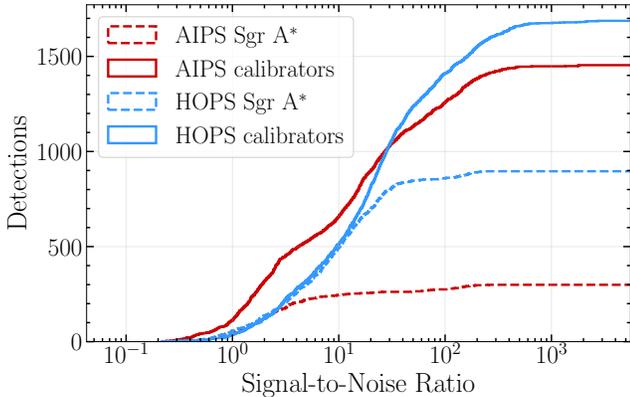
ALMA phased array of 37 antennas. A total bandwidth of 512 MHz (256 MHz per polarization) was recorded at each station except YS, which recorded a single left circular polarization component. At correlation, the bandwidth per polarization was divided in four channels of 116 sub-channels each. The observations included three calibrator sources: 1749+096, a bright quasar for bandpass and instrumental phase and delay calibration; NRAO 530 and J1924–2914, two quasars only  $\sim 10^\circ$  away from Sgr A\* on the sky, for differential phase, delay, and rate calibration. Several of the VLBA stations (NL, OV, PT) observed in difficult weather conditions, such as frost, strong winds or rain, leading to limited detections to those stations. Observations at PV suffered from phase instability and coherence losses in the signal chain, which led to poor quality data and lower visibilities on its baselines. See Issaoun et al. (2019) for further details on the observations.

The data were reduced via the EHT-HOPS pipeline (Figure 1), with additional validation from the NRAO Astronomical Image Processing System (AIPS; Greisen 2003). Reduction in HOPS utilized ALMA, the most sensitive station in the array, as the reference antenna to derive stable instrumental phase bandpass and RR–LL delay relative to other stations (Sections 2.2 and 2.4). Depending on S/N, either ALMA or GBT baselines were used to correct for intra-scan stochastic differential atmospheric phase, which varies on a timescale of a few seconds for this data set. The integration time for rapid phase corrections was determined automatically for each scan, taking into account the amount of random thermal variation and thus depending on S/N (Section 2.3). In the final `fourfit` stage, fringe solutions per scan were constrained to a single set of station-based delays and rates, or global fringe solutions, obtained from a least-squares solution to robust baseline detections (Section 2.5). A priori calibration was performed in post-processing, where all stations apart from YS, PV and ALMA required an additional opacity correction to calibrate the visibility amplitudes (Section 3.1).

The AIPS reduction followed a classical procedure for low-frequency VLBI, with additional steps for fringe fitting refinement. After loading the data set into AIPS, during which digital sampler corrections are applied, we inspected the data interactively via the tasks `BPEDT` and `EDITA` and removed spurs in frequency domain accumulated bandpass tables and time domain amplitude plots. We then normalized the amplitudes via `ACCOR` and applied field rotation angle corrections via `VLBAPANG` (correcting for source parallactic angle and receiver mount type of each antenna) prior to fringe fitting. The standard instrumental phase calibration, with the station-based fringe fitter `KRING`, corrects for experiment-wide correlator model phase and delay offsets using the full bandwidth and scan coherence. These solutions were derived using a scan on 1749+096, the brightest calibrator of the experiment, where twelve out of the thirteen

**Table 1.** Stokes  $I$  detections in the GMVA+ALMA data.

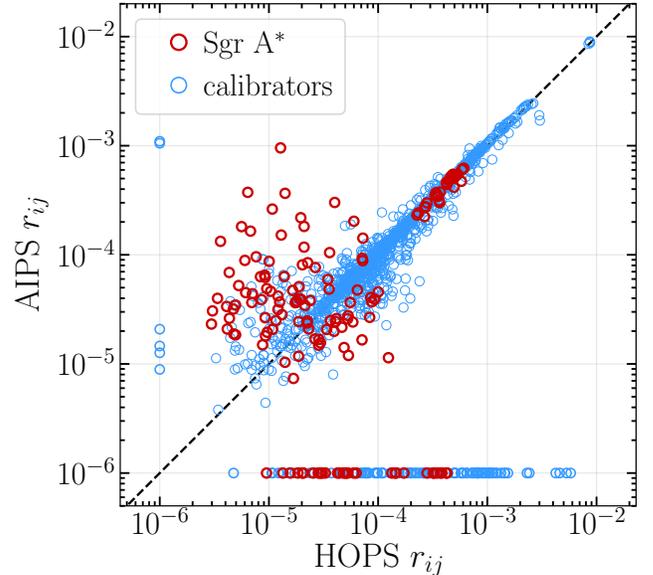
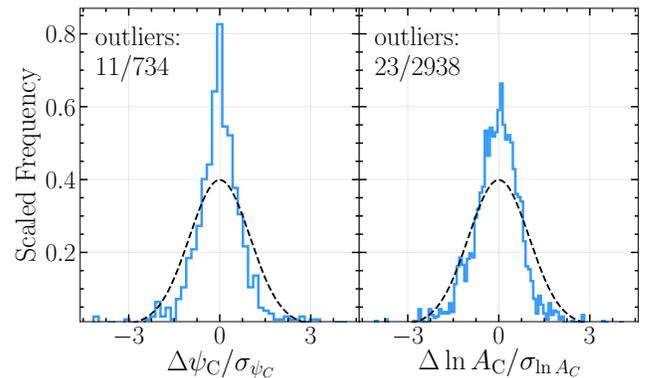
Source	AIPS	HOPS
1749+096	120	123
J1924-2914	309	304
NRAO 530	415	443
Sgr A*	196	461
Total	1040	1331

**Figure 10.** Comparison of cumulative S/N histograms for HOPS and AIPS RR and LL detections. HOPS recovers a significant number of detections that are not present in the AIPS data product. Particularly for Sgr A\*, there are just 79 parallel hands detections with S/N larger than 5 in the AIPS data set, and 596 in the HOPS data set. Possible reasons for the differences in fringe recovery are discussed in Section 5.2.

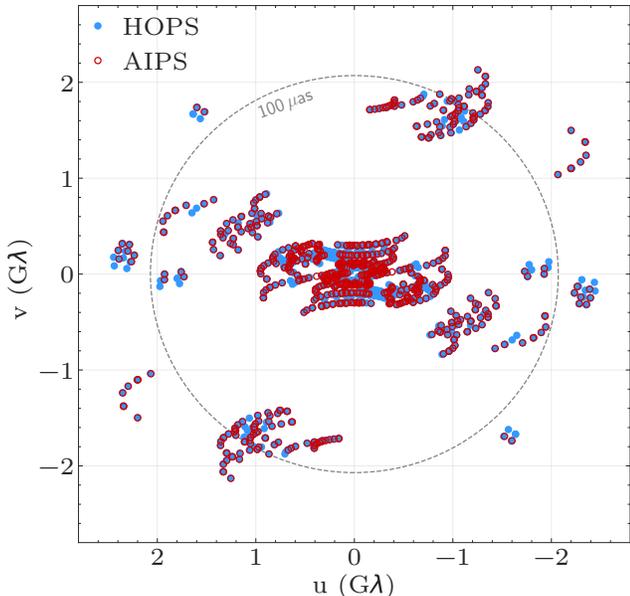
stations are present. A later scan on J1924-2914, the second brightest calibrator, was used to derive solutions for the MK VLBA station, not present in the 1749+096 scans. ALMA was also used as the reference antenna for this processing. The instrumental phase calibration was applied to all scans before proceeding to finer fringe fitting, where either ALMA or GBT was used as a reference antenna, depending on the source. We solved for fringe rates and residual phase and delay offsets per channel, using full scan coherence, for each individual scan. We ran a third fringe fitting step to solve for stochastic atmospheric phase variations in time across the full bandwidth, with a fixed solution interval of 10 seconds. A final fringe fitting step was used to solve for further scan-based residual delays and phases per channel to realign the channels. A priori calibration was performed with APCAL, ignoring the opacity correction ( $\text{DOFIT} = -1$ ) for YS, ALMA and PV.

### 5.2. Pipeline Comparison

We have performed a comparative analysis of the GMVA+ALMA data set processed by the EHT-HOPS pipeline and a classic AIPS reduction. The EHT-HOPS

**Figure 11.** Scatter plot of correlation coefficient  $r_{ij}$  magnitude in HOPS and AIPS data sets for RR and LL detections. The horizontal line at AIPS  $r_{ij} = 10^{-6}$  corresponds to detections present exclusively in the HOPS data set.**Figure 12.** Histogram of differences HOPS - AIPS, normalized by the combined uncertainties of the pipelines for closure phases (left) and log closure amplitudes (right). The dashed line corresponds to a standard normal distribution.

pipeline has recovered a significantly larger number of detections, as summarized in Table 1, as well as in Figure 10. This likely reflects a more efficient use of free parameters for phase calibration in the EHT-HOPS pipeline. The EHT-HOPS pipeline calibration is driven by purpose-designed tasks targeting the characteristics of high-frequency VLBI data, while the AIPS processing relies on standard tasks available in the AIPS environment. A significant difference is in the handling of atmospheric phase, where the EHT-HOPS pipeline parameterizes phase variations as a smooth function using a flexible variability timescale that can accommo-



**Figure 13.** The  $(u, v)$  GMVA+ALMA coverage of NRAO 530 following data reduction. Each symbol denotes a scan-averaged measurement: filled blue circles are detections via the EHT-HOPS pipeline; and hollow red circles are detections via the AIPS processing.

date the available S/N in the data (Section 2.3). In our AIPS reduction, rapid phase variation is captured using a fixed 10-second fringe solution interval, which may be too long (in the case of rapidly-varying atmosphere for poor weather conditions or low elevation), or too short (in the case of low S/N). Benefits in sensitivity from the coherent Schwab-Cotton global fringe search in AIPS may not make up for the other inefficiencies due to the arguments presented at the end of Section 2.5.

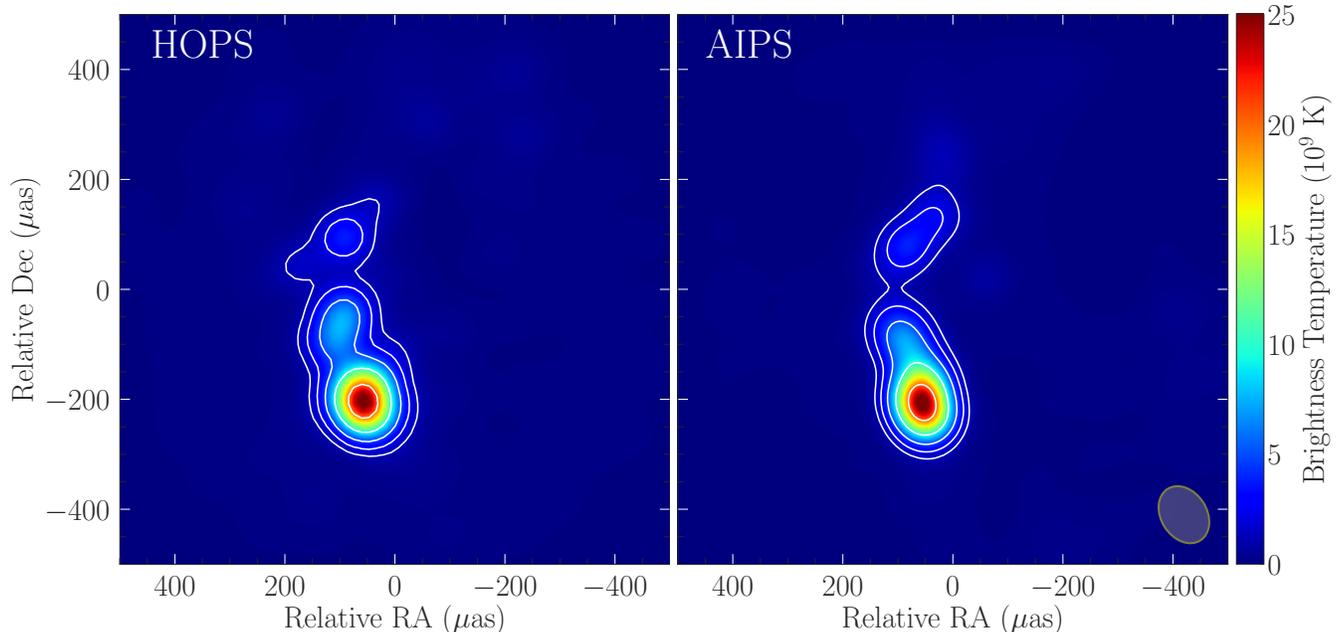
A broad consistency between the pipelines can be seen in Figure 11 for the common set of detections, showing the scatter plot of the correlation coefficient amplitude after the fringe fitting. While a certain amount of variation is seen in the lower S/N part of the data set, particularly for Sgr A\*, the high S/N data show a high level of consistency between the two reductions. Additionally, we directly compare closure quantities in Figure 12. Here we consider maximum sets of closure phases (left panel) and log closure amplitudes (right panel). We construct differences of scan-averaged RR and LL closure products matched between the pipelines and normalize them by the combined error for both pipelines. In the case of two independent measurements of the true underlying quantity, with normally distributed uncertainties, we would expect the result to be a standard normal distribution, plotted with dashed line for reference. The fact that the measured spread of the normalized differential quantity is smaller than that of a standard normal distribution indicates that differences between the pipelines are of subthermal magnitude, even after full

scan-averaging. This is not surprising as the pipelines are analyzing the same (not independent) thermal noise realizations. In Figure 12, we also note a small number of  $3\sigma$  outliers.

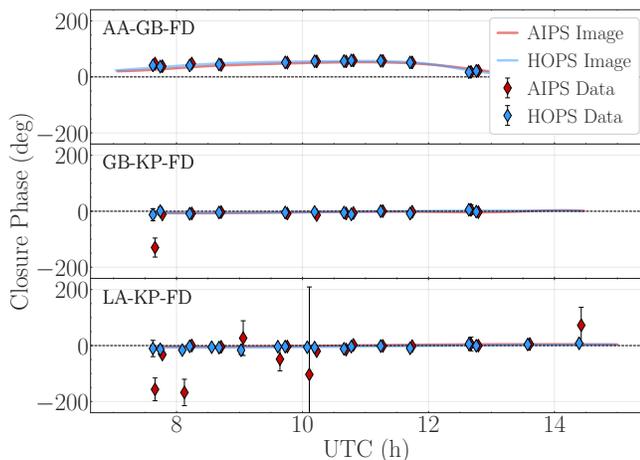
In Figure 13, we illustrate the detections from both reductions on the  $(u, v)$  coverage plot for the blazar NRAO 530, in which the EHT-HOPS pipeline has recovered  $\sim 7\%$  more detections. We proceed to an additional validation of the data sets via image reconstructions. We reconstruct images of NRAO 530 with both the AIPS and EHT-HOPS data sets, constraining the total flux of the source from simultaneous ALMA interferometric measurements (Goddi et al. 2019). For the imaging process, we make use of only closure quantities (amplitudes and phases), as the thirteen stations of the GMVA+ALMA array and their coverage provide a large relative amount of closure information, independent of station gain errors (e.g., Thompson et al. 2017). We image NRAO 530 using the `eht-imaging` library (Chael et al. 2016), following the closure imaging method of Chael et al. (2018). The same script was used for both data sets, and the resulting images are shown in Figure 14. The images show a high degree of consistency with each other, in addition to consistency in both morphology and jet direction with previous observations of NRAO 530 in the literature (Bower et al. 1997; Bower & Backer 1998; Feng et al. 2006; Chen et al. 2010; Lu et al. 2011).

In Figure 15, we inspect individual closure phases on three triangles of various sizes and orientations. The HOPS and AIPS closure phases are generally consistent, but the HOPS data set indicates smoother trends. The HOPS pipeline recovers zero closure phase more consistently on triangles that do not resolve the source, whereas AIPS has some difficulty due to the lower S/N of the intra-VLBA detections (bottom panel of Figure 15). The closure phase trends derived from the two reconstructed images are also shown in Figure 15, and both images result in smooth trends in modeled closure phase that are similar to each other, and either follow both data sets when the detections are well constrained or follow predominantly the HOPS detections when the AIPS detections result in different values.

While the calibrators are bright blazar sources typically reduced through classical AIPS procedures, the case of Sgr A\* presents added difficulty to the calibration process. In particular, the source is subject to interstellar scattering in our line of sight, causing scatter-broadening predominantly in the east-west direction, where a large majority of GMVA baselines lie (Davies et al. 1976; van Langevelde et al. 1992; Frail et al. 1994; Bower et al. 2004, 2006; Shen et al. 2005; Psaltis et al. 2018; Johnson et al. 2018). Additionally, Sgr A\* was  $\sim 2$  Jy in 2017 at 3.5 mm, at the lower end of its typical flux density range at this wavelength, and most stations of the GMVA, in the northern hemisphere, observe Sgr A\* at very low elevations and thus through



**Figure 14.** Closure-only images of NRAO 530 as reduced by HOPS (left) and AIPS (right) and imaged with the `eht-imaging` library (Chael et al. 2018). Total compact flux is determined by the analysis of ALMA interferometer data (Goddi et al. 2019). The equal brightness temperature contour levels start from  $1.25 \times 10^9$  K and increase in factors of two. The observations have a uniform-weighted beam =  $(111 \times 83) \mu\text{as}$ , PA =  $32^\circ$ .



**Figure 15.** Scan-averaged closure phases for NRAO 530 on three triangles (ALMA–GBT–FD, GBT–KP–FD, LA–KP–FD). Each diamond symbol denotes a scan-averaged measurement: blue diamonds are detections via the EHT-HOPS pipeline, red diamonds are detections via the AIPS processing. The AIPS points are offset in time by +2 min for clarity. Closure phase trends from the reconstructed images of the HOPS and AIPS data are shown as blue and red lines respectively.

a large airmass that lowers the chance for strong detections. These conditions add difficulty to a classical AIPS processing, which does not fare as well for Sgr A\* fringe fitting as the EHT-HOPS pipeline. Due to the clear

difference in performance, as shown in Table 1 and Figure 10, the EHT-HOPS pipeline processing was chosen to derive subsequent scientific results on Sgr A\*, presented in Issaoun et al. (2019).

## 6. SUMMARY

We have developed an automated calibration and reduction pipeline for high frequency VLBI data, suitable for processing data from the GMVA (at 86 GHz) and the EHT (at 230 GHz). The pipeline is structured around the Haystack Observatory Postprocessing System (HOPS), which was originally designed for precision geodetic analysis, but has also been widely used for the processing of early data from the EHT. The new EHT-HOPS pipeline was specifically targeted to meet the needs of the developing EHT and GMVA arrays. Specifically, it leverages high-sensitivity anchor stations, such as ALMA acting as a phased array, in order to phase-stabilize the network to atmospheric turbulence. The pipeline also provides reduced data that are phase calibrated to a global fringe solution in a standard UVFITS data format. This allows the HOPS output to be analyzed using a wide variety of downstream tools for VLBI data characterization, imaging, and modeling.

The EHT-HOPS pipeline was successfully used for the analysis of VLBI data taken at 86 GHz on Sgr A\* and associated calibration sources, using the GMVA joined by the ALMA phased array. The scientific analysis of the data was presented in Issaoun et al. (2019). In this work we have used data from the observations to illustrate the

calibration process, and have compared the output from the EHT-HOPS pipeline with a classical data reduction through AIPS.

The current implementation of the pipeline addresses the need for rapid phase calibration at high observing frequencies, and focuses on the robust detection of correlated fringes for the newly expanded VLBI networks. Future developments to the UVFITS postprocessing tool set will support amplitude bandpass corrections and polarization leakage corrections, to reduce non-closing baseline systematic errors and to provide the calibration necessary for polarization analysis.

We thank Thomas Krichbaum for helpful comments and discussion, as well as the EHT Collaboration, specifically members of the Correlation and the Calibration & Error Analysis working groups. We thank the National Science Foundation (AST-1126433, AST-1614868, AST-1716536) and the Gordon and Betty Moore Foundation (GBMF-5278) for financial support leading to this work. This work was supported in part by the Black Hole Initiative at Harvard University, which is supported by a grant from the John Templeton Foundation. The work was also supported by the ERC Synergy Grant “BlackHoleCam: Imaging the Event Horizon of Black Holes”, Grant 610058. This paper makes use of the following ALMA data: ADS/JAO.ALMA2016.1.00413.V.

ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. This research has made use of data obtained with the Global Millimeter VLBI Array (GMVA), which consists of telescopes operated by the Max-Planck-Institut für Radioastronomie (MPIfR), IRAM, Onsala, Metsahovi, Yebes, the Korean VLBI Network, the Green Bank Observatory and the Very Long Baseline Array (VLBA). The VLBA is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The data were correlated at the correlator of the MPIfR in Bonn, Germany. This work is partly based on observations with the 100-m telescope of the MPIfR at Effelsberg. This work made use of the Swinburne University of Technology software correlator (Deller et al. 2011), developed as part of the Australian Major National Research Facilities Programme and operated under licence.

*Facilities:* GMVA, GBT, ALMA

*Software:* HOPS (Whitney et al. 2004), AIPS (Greisen 2003), GNUParallel (Tange et al. 2011), eht-imaging (Chael et al. 2016, 2018), Numpy (van der Walt et al. 2011), Scipy (Jones et al. 2001), Pandas (McKinney 2010), Astropy (The Astropy Collaboration et al. 2013, 2018), Jupyter (Kluyver et al. 2016), Matplotlib (Hunter 2007).

## REFERENCES

- Akiyama, K., Lu, R.-S., Fish, V. L., et al. 2015, ApJ, 807, 150
- Alef, W., & Porcas, R. W. 1986, A&A, 168, 365
- Altshuler, E. E., Falcone, V. J., J., & Wulfsberg, K. N. 1968, Spectrum, IEEE, 5, 83
- Bower, G. C., & Backer, D. C. 1998, ApJL, 507, L117
- Bower, G. C., Backer, D. C., Wright, M., et al. 1997, ApJ, 484, 118
- Bower, G. C., Falcke, H., Herrnstein, R. M., et al. 2004, Science, 304, 704
- Bower, G. C., Goss, W. M., Falcke, H., Backer, D. C., & Lithwick, Y. 2006, ApJL, 648, L127
- Capallo, R. 2017, Fourfit users manual, [https://www.haystack.mit.edu/tech/vlbi/hops/fourfit\\_users\\_manual.pdf](https://www.haystack.mit.edu/tech/vlbi/hops/fourfit_users_manual.pdf), ,
- Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018, ApJ, 857, 23
- Chael, A. A., Johnson, M. D., Narayan, R., et al. 2016, ApJ, 829, 11
- Chen, Y. J., Shen, Z.-Q., & Feng, S.-W. 2010, MNRAS, 408, 841
- Cornwell, T., & Fomalont, E. B. 1989, in Astronomical Society of the Pacific Conference Series, Vol. 6, Synthesis Imaging in Radio Astronomy, ed. R. A. Perley, F. R. Schwab, & A. H. Bridle, 185
- Davies, R. D., Walsh, D., & Booth, R. S. 1976, MNRAS, 177, 319
- Deller, A. T., Brisken, W. F., Phillips, C. J., et al. 2011, PASP, 123, 275
- Doeleman, S. S., Weintroub, J., Rogers, A. E. E., et al. 2008, Nature, 455, 78
- Doeleman, S. S., Fish, V. L., Schenck, D. E., et al. 2012, Science, 338, 355
- Feng, S.-W., Shen, Z.-Q., Cai, H.-B., et al. 2006, A&A, 456, 97
- Fish, V. L., Doeleman, S. S., Beaudoin, C., et al. 2011, ApJL, 727, L36
- Fish, V. L., Johnson, M. D., Doeleman, S. S., et al. 2016, ApJ, 820, 90
- Frail, D. A., Diamond, P. J., Cordes, J. M., & van Langevelde, H. J. 1994, ApJ, 427, L43
- Goddi, C., Marti-Vidal, I., Messias, H., et al. 2019, arXiv e-prints, arXiv:1901.09987

- Greisen, E. W. 2003, in *Astrophysics and Space Science Library*, Vol. 285, *Information Handling in Astronomy - Historical Vistas*, ed. A. Heck (Kluwer Academic Publishers), 109
- Greisen, E. W. 2012, *AIPS Fits File Format*, AIPS Memo 117, NRAO
- Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
- Issaoun, S., Folkers, T. W., Blackburn, L., et al. 2017, *EHT Memo Series*, 2017-CE-02
- Issaoun, S., Johnson, M. D., Blackburn, L., et al. 2019, *ApJ*, 871, 30
- Johnson, M. D., Fish, V. L., Doeleman, S. S., et al. 2015, *Science*, 350, 1242
- Johnson, M. D., Narayan, R., Psaltis, D., et al. 2018, *ApJ*, 865, 104
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open source scientific tools for Python*, ,
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, ed. F. Loizides & B. Schmidt (IOS Press), 87 – 90
- Leppanen, K. J., Zensus, J. A., & Diamond, P. J. 1995, *AJ*, 110, 2479
- Lu, R.-S., Krichbaum, T. P., & Zensus, J. A. 2011, *MNRAS*, 418, 2260
- Lu, R.-S., Krichbaum, T. P., Roy, A. L., et al. 2018, *ApJ*, 859, 60
- Martí-Vidal, I., Vlemmings, W. H. T., & Muller, S. 2016, *A&A*, 593, A61
- Matthews, L. D., Crew, G. B., Doeleman, S. S., et al. 2018, *PASP*, 130, 015002
- McKinney, W. 2010, in *Proceedings of the 9th Python in Science Conference*, ed. S. van der Walt & J. Millman (Austin, TX: Proceedings of the 9th Python in Science Conference), 51 – 56
- McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 376, *Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell, 127
- Pearson, T. J., & Readhead, A. C. S. 1984, *ARA&A*, 22, 97
- Penzias, A. A., & Burrus, C. A. 1973, *ARA&A*, 11, 51
- Psaltis, D., Johnson, M., Narayan, R., et al. 2018, arXiv e-prints, arXiv:1805.01242
- Roberts, D. H., Wardle, J. F. C., & Brown, L. F. 1994, *ApJ*, 427, 718
- Rogers, A. E. E. 1970, *Radio Science*, 5, 1239
- Rogers, A. E. E., Doeleman, S. S., & Moran, J. M. 1995, *AJ*, 109, 1391
- Rogers, A. E. E., Hinteregger, H. F., Whitney, A. R., et al. 1974, *ApJ*, 193, 293
- Savitzky, A., & Golay, M. J. 1964, *Analytical chemistry*, 36, 1627
- Schafer, R. W. 2011, *IEEE Signal processing magazine*, 28, 111
- Schwab, F. R., & Cotton, W. D. 1983, *AJ*, 88, 688
- Shen, Z.-Q., Lo, K. Y., Liang, M.-C., Ho, P. T. P., & Zhao, J.-H. 2005, *Nature*, 438, 62
- Tange, O., et al. 2011, *The USENIX Magazine*, 36, 42
- The Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- The Astropy Collaboration, Price-Whelan, A. M., Sipócz, B. M., et al. 2018, *AJ*, 156, 123
- Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2017, *Interferometry and Synthesis in Radio Astronomy*, 3rd Edition (Springer International Publishing), doi:10.1007/978-3-319-44431-4
- Ulich, B. L., & Haas, R. W. 1976, *ApJS*, 30, 247
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science and Engineering*, 13, 22
- van Langevelde, H. J., Frail, D. A., Cordes, J. M., & Diamond, P. J. 1992, *ApJ*, 396, 686
- Vertatschitsch, L., Primiani, R., Young, A., et al. 2015, *PASP*, 127, 1226
- Whitney, A. R., Cappallo, R., Aldrich, W., et al. 2004, *Radio Science*, 39, RS1007
- Whitney, A. R., Beaudoin, C. J., Cappallo, R. J., et al. 2013, *PASP*, 125, 196
- Wielgus, M., Blackburn, L., Issaoun, S., et al. 2019, *EHT Memo Series*, 2019-CE-02