Quality Assessment of the Mizusawa Software and GPU Correlators

Takaaki Jike, Tomoaki Oyama, Aya Yamauchi

Abstract Mizusawa has multiple software correlation systems. These currently operating correlation processing systems have several variations in CPU and GPU. Data evaluation was performed to verify the consistency and reliability of the correlation data obtained from these individual correlation processing systems. From these test results, it was found that the differences in the number of CPU cores and the number of FFT points changes the magnitude of the additional error to the correlation data. The magnitude of this additional error was six to 19 times that of the noise artificially added during the correlation processing.

Keywords additional error, number of cores, FFT points

1 Introduction

We recognize that next-generation astronomy, geodesy, and space-time measurements target parameter estimation accuracy at less than 1 mm, 1 microarcsecond, attosecond levels of clock tempo, and 1 micro Jy. At the ninth IVS General Meeting, we reported the results of the delay measurement and geodetic parameter estimation with theoretical white noise delay error of 2 picoseconds root-mean-square (rms) [1]. However, the post-fit residual rms was more than 15 picoseconds, and therefore, it is acknowledged that noisy or systematic compound errors were added during the observation and analysis processes. Guaranteeing required accuracy in the complex of focus, receiving, transfer, digitizing, recording, and correlation is a necessary task to reduce the magnitude of the post-fit residual rms and to achieve the target accuracy. As an investment in the expected goal, we inspected the delay estimation accuracy guaranteed by the current Mizusawa Software Correlator specification. The purpose of this research is to accumulate effective inspection techniques for investigating the performance of next-generation electromagnetic wave-based space-time measurement techniques.

Mizusawa VLBI Observatory, National Astronomical Observatory of Japan

2 Status Parameters of the Mizusawa Software and GPU Correlators

In Mizusawa, two FX-type correlators are implemented; one is the Software Correlator (MSC), and the other is the GPU Correlator (MGC). The MSC is under regular operation, and multiple PC units are available depending on the application. These operating PCs are built with multiple CPUs, motherboards, and operating systems. The engine of correlation processing, called "gico3", is developed by NICT and is installed with optimization according to the number of CPU cores. The MGC is newly developed as the high-end model of the MSC and is now in the test phase. The correlation engine, called "kfxcom", is developed by NAOJ and ©KIMSOFT, and parallel processing is executed by multiple cores arranged inside the GPU, achieving a higher processing speed.

In this study, we estimated the delay from the same observation data using multiple correlators with different specifications and correlation processing modes, and we examined the difference in the additional error

corresponding to each correlator and the change of the correlation mode. Table 1 shows the types and specifications of correlators used for validation. Even if the CPUs have the same number of cores, the CPU and motherboard type are different in each PC.

Table 1 Specifications of the MSC and MGC used for verification.

Correlator PC Name	IP25	IP90	IP114	IP76	IP61
Core type	CPU				GPU
Number of cores	6			8	4352
FFT math library	FFTW332				
Matrix math library	avx2	avx	avx2	avx512	none
Operation engine	gico3-2.6.8				kfxcom-2.0

3 Correlation Data Set

The observation data used in the correlation test is shown in Table 2. Several correlation processes with different settings were performed using common data. The parameters changed during the correlation test are the number of FFT points and the type of PCs, substantially the number of cores. Hereafter, each correlation processing name will be *IPaaa-bbbk* according to this setting, where *IPaaa* is the correlator name and *bbb* is the number of FFT kilo-points. For example, in the case of 512-point FFT correlation processing using PC "IP90", the processing mode is named *IP90-0.5k*.

4 Estimating the Reference Additional Error Scale

In the correlation processing by gico3, white noise error where the size falls within a certain distribution is artificially added at the time of FFT processing. Figure 1 shows the difference between the first cross-correlation coefficient and the other nine correlation coefficients when the same 40 seconds of data is correlated ten times in *IP90-0.5k* mode. This short correlation process is named *IP90-0.5k40*. Moreover, kfxcom does not have such a noise addition function. This is a comparison of the absolute values of the normalized correlation coefficients represented as a single precision complex number. The individual coefficients are obtained from the integral of the time of one second and bandwidth of 2 MHz. Most of the scatter is within $\pm 2 \times 10^{-5}$. It is considered that the magnitude of scattering when converted to this fringe function decreases to $1/102.4$ in the integrated value for 40 seconds. Based on this result, the delay estimates are compared between *IP90-0.5k* and *IP25-0.5k*, and the results are shown in Figure 2. When the absolutes of the delay differences are arranged along the theoretical white noise error, a proportional line can be drawn at the edge of the delay difference distribution. The delay differences are scattered between the proportional line and the y_0 line. We speculate that the reason for this scattering is the noise added during the FFT processing. When performing a fringe peak search with a three-point quadratic function fitting, even when noise of the same magnitude is added, the smaller the signal-to-noise ratio is, the larger the deviation of the estimated position of the fringe peak becomes. Figure 3 explains this property as an image.

Table 2 Specification of correlated data.

Observation	r19336k, VERA K-band Geodesy
Epoch of Data	2019y 336doy 0h - 12h
Network Stations	VERA (Vm, Vr, Vo, and Vs)
Format	1024 Msps - 2 bit - 1 stream
Recording	OCTADISK 2048 Mbps
Accumulation Frequency	1 Hz
Correlation Output Form	CODA FS / FITS-IDI

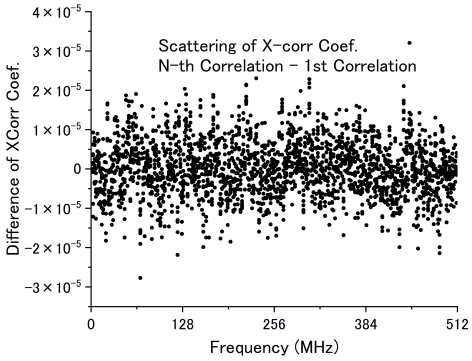

Fig. 1 Reproducibility of the cross-correlation coefficient.

To get evidence to support this speculation, delay estimations were attempted by adding fluctuations of $\pm 2 \times 10^{-7}$ to the quadratic fringe function of *IP90-*

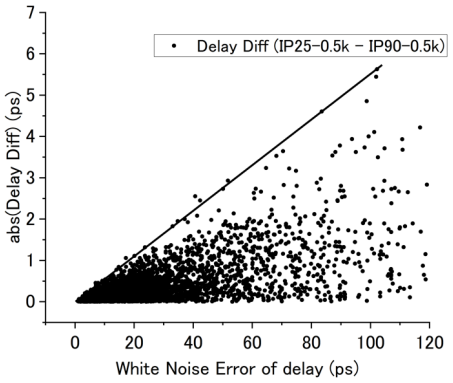

Fig. 2 Distribution of delay differences between *IP25-0.5k* and *IP90-0.5k*.

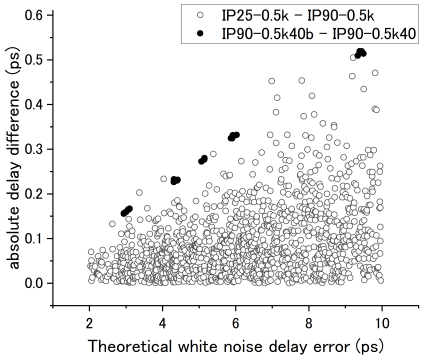

Fig. 4 Estimating the delay deviation with artificial noise added to the fringe.

0.5k40. The name of this process is *IP90-0.5k40b*. Figure 4 is shown by superimposing the delay difference between *IP90-0.5k40* and *IP90-0.5k40b* on the distribution of the delay difference between *IP25-0.5k* and *IP90-0.5k*. The delay differences between *IP90-0.5k40b* and *IP90-0.5k40* is distributed on a proportional straight line along the edge of the distribution of the delay differences between *IP25-0.5k* and *IP90-0.5k*. This is consistent with the case that the maximum fluctuation of the fringe function value is expected to be $\pm 2 \times 10^{-7}$ when the additional error integrated for 40 seconds contributes to the fringe function. Therefore, it was stipulated that the distribution of the delay difference was due to the additional error, and the effect of the additional error was estimated to be 52 femtoseconds/picosecond in terms of the proportionality coefficient. From now on, this additional error will be used as a reference scale. In addition, differences in the CPU model, motherboard, FFT math library, and matrix math library seem to have little effect on the uncertainty of delay estimation.

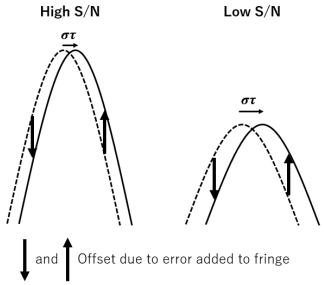

Fig. 3 Relationship between signal-to-noise ratio and deviation of the estimated fringe peak position for the same additional noise.

5 Differences in Correlation Processing Modes and Their Impact on Additional Errors

In order to confirm that the magnitude of the additional error differs depending on the mode of the correlation processing, the correlation processing was performed in several different modes, and the rate of additional error was estimated from the delay difference between multiple modes. These results are listed in Table 3. The rate of additional error is represented by a proportional coefficient along the edge of the delay difference distribution as in Figure 2.

Because the results in the table are for confirming the tendency of additional errors to occur, more systematic testing is required to estimate the characteristics and causes of additional errors. However, the number of cores of the correlator and the FFT score of the correlation processing have the following characteristics regarding the additional error:

1) When the number of FFT points is the same and the number of cores is the same, the additional error is dominated by the artificial noise added during the FFT of the correlation processing.

Table 3 Relation of correlation mode and rate of additional error.

Mode1 - Mode2	Rate of Additional Error (ps/ps)
EQ. Number of Cores. and EQ. FFT Points	
IP25-0.5k - IP90-0.5k	0.052
IP114-0.5k - IP90-0.5k	0.052
IP114-2k - IP90-2k	0.053
EQ. Number of Cores and NE. FFT Points	
IP61-4k - IP61-2k	0.316
IP61-2k - IP61-1k	0.412
IP61-4k - IP61-1k	0.441
IP76-4k - IP76-2k	0.346
IP76-2k - IP76-1k	0.540
IP76-4k - IP76-1k	0.556
IP76-1k - IP76-0.5k	0.778
IP76-2k - IP76-0.5k	0.801
IP90-2k - IP90-0.5k	0.803
NE. Number of Cores and EQ. FFT Points	
IP76-2k - IP61-2k	0.360
IP90-2k - IP61-2k	0.459
IP76-2k - IP90-2k	0.486
IP76-1k - IP61-1k	0.495
IP76-0.5k - IP90-0.5k	0.977
NE. Number of Cores and NE. FFT Points	
IP76-1k - IP90-2k	0.634
IP76-2k - IP90-0.5k	0.795
IP76-1k - IP90-0.5k	0.834
IP76-0.5k - IP90-2k	0.857

2) When the number of FFT points and the number of cores are changed, the magnitude of the additional error increases from six to 19 times the reference additional error scale.

3) As the ratio of FFT points between processes increases, so does the ratio of magnitude of additional error.

4) The difference in the number of FFT points affects the increase in the additional error, and the rate of increase in the additional error is larger in the MSC than in the MGC.

It is considered that the additional errors that change due to the difference in these correlation modes form a complex generated in the process of various operations. One of the components in this complex appears to be noise-like errors in the frequency and time domains. However, the characteristics and causes of this error are insufficiently identified; rather, it may be a complex of errors with systematic characteristics such that the magnitude of noise changes depending on the number of FFTs and cores. Another main part of the complex is presumed to be the accumulation error from fringe rate mis-tracking. The larger the FFT segment, the longer the time length of data used for one FFT process. At this time, uncorrected frequency fluctuations are accumulated. This error is systematic and can be seen in the geodetic solution as shown in Figure 5, which shows the relationship between the UT1-TAI solution and the number of FFT points. The solutions of UT1-TAI step according to the number of FFT points. The MGC also is represented as steps in the solutions, but the size of the steps is smaller than with the MSC.

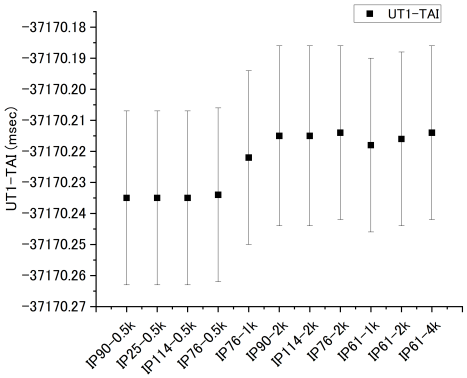

Fig. 5 Relationship of the number of FFT points and step-like differences of UT1-TAI solutions.

6 Characteristics of Additional Error in the Time Domain

The characteristics of the additional error in the time domain were investigated using the MGC. The specifications of the observations are the same as those of Table 2 except that the radio source “3C454.3” was tracked continuously for three hours. Delays were estimated every 32 seconds from each of the correlation processes with FFT points changed to 1k, 2k, and 4k. The statistical parameters of the delay difference compared between 1k and 2k and between 1k and 4k are shown in Table 4. Figure 6 shows the distribution of delay differences used to estimate the statistics in Table 4.

The displacement of the delay difference distribution over time is considered to be the cumulative error

Table 4 Time domain statistics of delay difference distributions obtained from the same observation, the same MGC (IP61), and different FFT points.

Diff. of FFT points	*2k-1k*	*4k-1k*
Baseline: Vm-Vr		
average displacement rate of distribution (ps/hr)	0.199	0.328
rms of scatter (ps)	0.271	0.277
average delay error (ps)	1.875	1.875
Baseline: Vm-Vo		
average displacement rate of distribution (ps/hr)	0.524	0.815
rms of scatter (ps)	0.331	0.349
average delay error (ps)	2.169	2.169
Baseline: Vr-Vo		
average displacement rate of distribution (ps/hr)	0.542	0.758
rms of scatter (ps)	0.435	0.464
average delay error (ps)	2.870	2.869
Baseline: Vo-Vs		
average displacement rate of distribution (ps/hr)	-0.184	-0.312
rms of scatter (ps)	0.622	0.695
average delay error (ps)	4.168	4.168

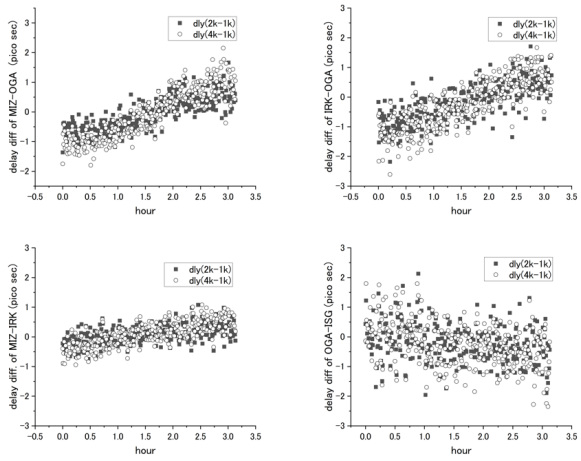

Fig. 6 Distribution of the delay difference of each baseline in the time domain.

caused by fringe rate mis-tracking due to the change in the number of FFT points. There is a difference of 120 to 290 femtoseconds/hour in the displacement rate between *2k-1k* and *4k-1k*. On the other hand, the magnitude of the delay difference scattering is six to 73 femtoseconds between *4k-1k* and *2k-1k*, and the dependence on the FFT point ratio is small, but it can be confirmed. Therefore, it is considered that the cumulative error of the fringe rate mis-tracking is the dominant additional error that depends on the change in the number of FFT points, and the cause of this scattering error is expected to be the bit-real conversion and its correction and operational errors that occur during the FFT and integration process.

7 Conclusions and Outlook

The following is a summary of this report.

The core engine "gico3" of the Mizusawa Software Correlator is equipped with a noise addition function during FFT. Expressing the magnitude of this noise as an additional error in delay estimation, the increment rate is 52 femtoseconds/picosecond as a proportional coefficient to the theoretical white noise error of delay. This additional error is treated here as a reference additional error scale.

The noise addition function implemented in the MSC gives $\pm 2 \times 10^{-5}$ fluctuation to the complex visibility obtained from 512-point FFT correlation processing.

When the correlation processing was performed by changing the number of arithmetic cores of the CPU and GPU and the number of FFT points, the magnitude of the additional error increased from six to 19 times the reference scale.

One of the components constituting this additional error is the cumulative error due to the fringe rate mistracking linked to the number of FFT points. This error indicates a step-like indeterminacy in which the geodetic estimation parameters are linked to the number of FFT points. In order to reduce this error, it is desirable to develop a new method of fringe stopping, such as improving the accuracy of high-order term tracking of the delay before the FFT. The other component is noise-like errors that may occur during the various processes of the correlator. This may be solved by improving the computational precision of the correlator and improving the signal reproducibility by increasing the number of layers of bits.

References

1. T. Jike, "VERA Geodetic VLBI with a Newly Developed High-speed Sampler and Recorder", In D. Behrend, K. D. Baver, and K. L. Armstrong, editors, *IVS 2016 General Meeting Proceedings*, NASA/CP-2016-219016, pp. 159–162, 2016.