Picosecond-Level Error Detection Using PCA in the Hardware Timing Systems for the EISCAT_3D LAAR

Abstract

While developing the timing system for the receiver arrays for the EISCAT_3D system, several approaches for detecting and adjusting for timing errors within the array have been explored. Since the lengths of the radar pulses are shorter than the size of the aperture, time-delay beamforming is necessary, which puts high demands on the timing system. The demand on the timing error among all elements in the array is to have a standard deviation of less than 120 ps. This requires high-quality error-detection systems to guarantee radar operation. This paper investigates the qualities of a secondary error-detection system based on statistical analysis of captured data.

The measurements were assembled with a signal-to-noise ratio (SNR) of $-30$ dB, implying that the elements in a 2112 element array needed to be grouped into subarrays of 48 elements each. The captured data was then evaluated by principal-component analysis (PCA) and averaged over 20,000 measurements, or about half a second. Timing errors between subarrays of down to $\pm 120$ ps and a percentage of faulty subarrays of up to 20% were detectable.

Principal-component analysis can be used as a cheap secondary error-detection system because it only needs a small amount of computer time to implement. In addition, it acts as a detection system for hardware errors in the primary timing system that are undetectable without a secondary system.

1. Introduction

EISCAT_3D is a design study of a new Incoherent-Scatter Radar (ISR) that is based on a Large-Aperture-Array Radar (LAAR) system, where Large Aperture refers to the fact that the incoming radar pulses are shorter than the aperture of the radar. This renders it necessary for the radar to use time-delay beamforming, as opposed to phase-delay beamforming. The addition of high demands on the pointing accuracy, $\pm 0.06^\circ$, and low beamforming loss, $< 0.2$ dB, have put stringent demands on the timing system of the array.

During the development of the receiver hardware for the EISCAT_3D[2] radar test array, one of the main focus areas has been the timing system. Because of the necessity of very low errors in the timing – an error with a standard deviation of less than 120 ps [3] – a continuous timing-calibration system has been built. While this cable-calibration system currently is under evaluation in the test array, another approach for detecting timing errors has been evaluated.

This paper describes the method, which uses principal-component analysis (PCA) to find any subarrays in a digitally sampled array where the timing of the signal differs significantly from the other subarrays. The incoming radar signals are well below the noise floor ($-30$ dB for a single antenna element), and the method is shown to work without prior knowledge of the incoming signal. It can thus be used while the radar is operating.

This error-detection method is not intended to replace the cable-calibration system, but is rather a supplemental detection system to lower the risk of degradation of the functionality of the EISCAT_3D radar. An important difference between the two systems is the capacity in which they are used. The cable-calibration system is only continuous on a large scale, since it is an active calibration system. It injects a signal into different signal paths in the array, effectively drowning out any other signals. It can thus only be used during radar down times, e.g., between pulses or experiments. The method proposed in this paper
is a passive monitoring system that uses statistical methods to find timing errors in signals buried in noise. It can thus be used continuously, during both active radar operation and the radar-calibration mode.

Another important difference is the accuracy of the detection method, itself. The cable-calibration system measures an absolute timing error with a high accuracy. Simulations show that an error with a standard deviation of less than 50 ps should be achievable for the timing, and an amplitude error of less than 8%. The statistical method in this paper does not reveal an absolute error at all, but can only indicate that an element is unsynchronized with the other elements. However, if more than one error is detected, a relative difference in magnitude of the error is discernible. As for amplitude errors, no indication at all is given by the statistical method.

Other methods of sub-noise-signal error detection that were considered included correlation, which, for example, is used in global navigation satellite systems [4]. However, these were not as successful for small errors, and took more than twice the computational time to process.

It is important to note that since the radar pulses are shorter than the aperture of the array, it is essential for the functionality of the principal-component analysis method that enough samples are used in the analysis to ensure that at least one complete radar pulse has been recorded at every element in the array.

The remainder of this paper will go through the methods used, the results, and finally, the conclusions.

2. Method

The signals coming into the EISCAT_3D radar array are very weak. A reasonable assumption is for the signal to have a signal-to-noise ratio (SNR) of about −30 dB for a single antenna element. However, the antennas in the array are grouped into subarrays of 48 antennas each, thereby increasing the SNR to about −10 dB. This is the level that has been used in simulations of the EISCAT_3D system using the Large-Aperture-Array Radar Simulation Environment (LAARSE) [5].

The incoming signal of the EISCAT_3D radar is located in a 30 MHz-wide band centered at around 220 MHz, and contains a number of different frequencies. While real data could be used to test the statistical method, the test array built in Kiruna, Sweden, is too small to test the method, since it only contains one subarray. Instead, the method has been evaluated using the Large-Aperture-Array Radar Simulation Environment to simulate a larger version of the EISCAT_3D test array. The target size of a receiver array in the EISCAT_3D project is between 2,000 to 16,000 antenna elements. The simulations made in this paper were based on the smaller version of these, simulating an 88 × 24 element array, yielding 2,112 antenna elements in the array in $M = 44$ subarrays.

2.1 Statistical Method

Principal-component analysis (PCA) is a well-known statistical method for analyzing multivariate data [6]. Through the analysis, the principal components of the data are calculated, which can be thought of as an alternative way to describe the data by as few variables as possible. From these components, principal-component scores (PCS) can be calculated for each data row of the analyzed data set. This is a single number per component that describes how well the principal components describe that row.

Another way to look at principal-component analysis is through a geometrical interpretation. Each successive component in the principal-component analysis will describe the largest variation in the data not already described by a previous component. In the two-dimensional case, an example set of data is plotted in Figure 1. The data forms a cloud centered around the origin, and when performing

![Figure 1. The geometrical interpretation of the principal-component analysis (PCA) in two dimensions. A data set of 1000 randomly generated points that had an elliptic distribution were analyzed with principal-component analysis. The two first components were plotted with dashed and dash-dot lines, respectively. A solid-line representation of the two-dimensional Gaussian distribution contour that they described is also included.](image-url)
a principal-component analysis on the data set, the two first principal components will point in the directions of the largest variation. The components of the principal-component analysis will in this case describe the contour of a two-dimensional Gaussian distribution plot of the data. It is thus clear that not only the direction of the largest variation is deduced, but also a measure of the relative magnitude of the variation. Increasing the number of components in the principal-component analysis in this simple case would not improve the results any further, since we only had a two-dimensional data set from the start. On the other hand, a data set that would seem to be three-dimensional might very well be described sufficiently well by only two principal components, or by even a single component. When the dimensionality of a problem increases, the geometrical interpretation looses its lucidity, but from an understanding point of view, it gives a clear example in the two- and three-dimensional cases.

### 2.2 Method of Application

When applied to the data from a number of different subarrays in the EISCAT 3D radar, the principal-component analysis will detect any subarray that has a component of data that differs from the rest of the subarrays. In a perfectly synchronized array, the results will not reveal any differences, since the signal part of the data is the same for all subarrays. However, if a subarray is unsynchronized with the others, it will by itself generate a higher principal-component score, since its data have a part that is different from all of the other subarrays.

The data from the subarrays in the Large-Aperture-Array Radar Simulation Environment simulations of the EISCAT 3D radar were collected in a matrix as follows:

\[
X = \begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_M
\end{bmatrix},
\]

where \( M = 44 \) is the total number of subarrays, and \( x_m = [x(0) \cdots x(N)] \) is the sampled data from the \( m \)th subarray, with \( N = 257 \) samples. In \( X \), the mean was subtracted from the columns to center the data. The principal components, \( a_j \) for \( j = 1, 2, \cdots, N \), of \( X \) were calculated. The property of the principal-component analysis, where each successive component describes the largest variation of the data not already described by previous components, was used to describe the data. Since the signals were buried in noise, it was enough to monitor only the first principal component, since this component describes the noise. That is, the variation in the direction of the first principal component, \( a_1 \), was enough to describe the data set. Since the noise was the largest contributor to the analyzed data, it was thus described by the first component.

The principal-component score, \( y_m \) for \( m = 1, 2, \cdots, M \) for the data were calculated as

\[
y = Xa_1,
\]

resulting in an \((M \times 1)\) vector with the scores \( y_m \) for the \( M \) subarrays by using only the first principal component. Looking at the principal-component score of the first component, a score that differs from the average score could be attributed to a difference in the underlying signal with respect to the data set. However, a single evaluation would not reveal this difference with any certainty, thus requiring a larger set of data on which to perform the analysis.

To provide the larger data set, all simulations were made \( K = 20,000 \) times, and thereafter the analysis was conducted. With a data length of \( N = 257 \) samples at 80 MHz sampling frequency, each data set corresponded to \( \sim 3.2 \mu s \). Thus, 20,000 sets of data at 12.5% duty cycle of the radar were collected in 514 ms.

### 2.3 Experimental Setup

In the Large-Aperture-Array Radar Simulation Environment, a signal is generated at the incoming frequencies and was, in the data used in this paper, constructed from four randomly placed signals within the signal band. The Large-Aperture-Array Radar Simulation Environment simulates every step of the receiving array, from input filters to analog-to-digital converters (ADC) and beamforming. The output from the simulations used for further investigation was the beam-formed signal from each antenna subarray, since this was what could be used for analysis in a future system. White Gaussian measurement noise and a distributed timing error over the array with a standard deviation of 160 ps were added to each simulation run to match reality as closely as possible.

Known but random timing errors were introduced into a number of elements to provide an error that could be analyzed. A number of different timing errors were tested to give a wider range of indication of the accuracy of the method.

To increase experimental robustness, the pointing of the array was randomized \( \pm 10^\circ \). The signal structure described above was also randomly generated for each simulation, both in frequency and amplitude, so that every signal was unique. This was to simulate a real experiment where the changing ionosphere was measured, and to make sure that the results from the statistical analysis were not influenced by a signal correlation between simulations that would not exist in reality. Every simulation run was thus unique in every way, expect for the introduced error, which was kept constant over the simulations. This enabled constant errors in the timing hardware to be tracked and detected.
3. Results

To evaluate the performance of principal-component analysis for detecting timing errors, two main tests were performed. The first of these tested how small the errors were that could be detected with the method. The second tested how many faulty subarrays could be present in the array without the detector breaking down.

The first step was to analyze the case without any introduced errors to get a reasonable value for the threshold for detecting faulty subarrays. Figure 2 shows the mean value, $\bar{y}$, for each subarray of the first principal component score over $K = 20,000$ runs:

$$\bar{y} = \frac{1}{K} \sum_{k=1}^{K} |y_k|,$$

where the absolute value of the scores, $|y_k|$, was used. Using the absolute value was necessary since any subarray that stood out from the rest was to be detected, regardless of the sign. The result from Figure 2 and Equation 1 was used in conjunction with five different sets of measurements where the number of faulty elements was varied. The target of the threshold was to provide a simple detector capable of detecting small errors with a very low probability of false alarm ($P_{FA}$). To achieve this, the threshold was swept from $\gamma = \mu_y + 6.8\sigma_y$ to $\gamma = \mu_y$ over the different measurements, and the minimum threshold possible with $P_{FA} = 0\%$ was calculated. The resulting thresholds are collected in Table 1. Since the largest of these approached $\gamma = \mu_y + 3\sigma_y$ and a margin was desirable, that threshold level was chosen for all consecutive simulations done in this paper.

As can be seen in Table 1, increasing the number of faulty subarrays in the analysis decreased the lowest possible threshold with $P_{FA} = 0\%$. This was because when the number of faulty subarrays increased, the number of correct subarrays decreased, causing the difference between errors and the correct part of the array to grow. A more-

<table>
<thead>
<tr>
<th># of Faulty Subarrays</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td>$\mu_y + 2.80\sigma_y$</td>
</tr>
<tr>
<td>10%</td>
<td>$\mu_y + 2.93\sigma_y$</td>
</tr>
<tr>
<td>23%</td>
<td>$\mu_y + 0.10\sigma_y$</td>
</tr>
<tr>
<td>41%</td>
<td>$\mu_y + 0.00\sigma_y$</td>
</tr>
<tr>
<td>43%</td>
<td>$\mu_y + 0.00\sigma_y$</td>
</tr>
</tbody>
</table>

Table 1. The lowest possible threshold level with $P_{FA} = 0\%$ for five different levels of errors.
sophisticated threshold setting, deduced from each case, could therefore be preferred. However, since the goal of the method is to detect errors in the array at an early stage, it is likely that only a single or a few faulty subarrays are to be detected at one time. Thus, it was reasonable to use a fixed threshold based on the zero-error case.

3.1 Minimum Detectable Error

By inducing time errors with increasing size in different antennas in one simulation setup, an indication of the minimum detectable error with the method investigated could be found. Figure 3 shows the results from a setup where errors between 25 ps and 200 ps were induced at a step size of 25 ps. In this setup, 18% of the subarrays contained errors. The figure shows that errors down to 150 ps were clearly detectable with the method. With a more-sophisticated method of setting the threshold, even the 125 ps error would be detectable.

In a situation with only a single faulty subarray, the detector would probably be able to detect errors down to 100 ps (see Figure 4).

3.2 Multiple Number of Faulty Subarrays

To investigate how large a part of the array can be erroneous and have the detection capabilities maintained, a number of different setups were used. Each setup increased the number of induced errors in the array, and was then evaluated for missed and false detection. The position, size, and sign of the error was randomized for each setup. The results for a single faulty subarray are shown in Figure 4. The induced error was about 200 ps, and was clearly detectable. Increasing the number of faulty subarrays to about 10%, Figure 5 still showed a 100% detection rate with no false alarms, even though one of the erroneous subarrays had a timing error of ~105 ps.

Even at 23% faulty subarrays the principal-component-score method did not fail, as can be seen in Figure 6. All errors were clearly detected, without false alarms. With this amount of faulty subarrays, the question arose as to whether there was a point in further pursuing this investigation. After all, a continuous monitoring system such as this is meant to be should alert the operator at the
first faulty subarray, and thus trigger a repair of the faulty subarray, rather than continued operation. The probability that as much as 23% of the subarrays would have erroneous timing-hardware errors simultaneously is negligible. Regardless, for the completeness of the investigation, even more faulty subarrays were added.

At 41% faulty subarrays, the method started to miss-detect. As can be seen in Figure 7, two out of 18 faulty subarrays were not detected. However, the missed errors were all below 130 ps in amplitude, so the question was whether they were not detected because of their magnitude or because of the high number of faulty subarrays. The authors are convinced of the former, since when increasing the number of faulty subarrays to 43% (see Figure 8), the six out of 19 faulty subarrays that were not detected were again all below 130 ps. It thus seems likely that what sets the detectability of the faulty subarray is a tradeoff between the number of faulty subarrays and the magnitude of the errors. However, the simulations showed the algorithm to be more sensitive to the magnitude of the error.

4. Conclusions

The use of principal-component analysis to detect timing errors in a subarray of the EISCAT_3D radar is viable under certain circumstances. The magnitude of the errors has to be larger than ~120 ps, which is concurrent with the standard deviation of the acceptable timing error throughout the array, and the ratio of faulty subarrays has to be below 20% of the subarrays in the radar. In these situations, the principal-component analysis method is capable of detecting 100% of the faults with a 0% false-alarm rate.

As a secondary timing-error detection system, the method described in this paper is easy and cheap to implement, as it only needs a small amount of computation time to work. It can be used to detect hardware errors foremost in the primary timing calibration system itself, as this might cause otherwise undetectable errors in the array.

5. Acknowledgments

The work presented in this paper was funded by the European Community under the “Structuring the European Research Area” Specific Programme Research Infrastructure action.

The EISCAT Scientific Association is supported by the Suomen Akatemia of Finland, the Chinese Institute of Radiowave propagation, the Deutsche Forschungsgemeinschaft of Germany, the National Institute for Polar Research of Japan, Norges Forskningsråd of Norway, Vetenskapsrådet of Sweden, and the Particle Physics and Astronomy Research Council of the United Kingdom.

6. References