ngVLA Computing Memo #10

Oscilloscope Mode Concept

P. Brandt, M. Schiller

January 17, 2024

Abstract

The purpose of this memo is to clarify SYS3105 and any derived requirements:

SYS3105: Fast-read out modes shall be available for remote engineering diagnostics of all LRUs (i.e., an on-board oscilloscope function).

Anecdotally, the intent behind this requirement has been summarized by R. Selina as “kilohertz to megahertz sampling” of arbitrary monitor points in the system. Sampling at this order of frequency would allow engineers to find issues not typically seen in standard “low rate” (roughly 1Hz - 100Hz) monitor data, for example power supply ripple or transient variations which could indicate impending component failures.

Data Sources

The data acquisition system primarily deals with two types of data: monitor and scientific. In industrial control systems, monitor points are referred to as “tags”, and we will use this terminology from here onwards.

Tag data could be further decomposed into separate categories, such as information required for the Science Data Model (“metadata”) and information which provides insight on system health. However, the distinction between these types of tag data only appears once it has been aggregated at the central system, for example once it has been transmitted to the Central Electronics Building or Array Operations Center, where it is appropriately sent to the tag data storage system, the science archive, or both. That is, though they have different "intents" within the system, there is little difference within the sampling and transport layers of the architecture between these types of tag data; therefore it makes sense to ignore the distinction within this discussion.

Data which flow through the telescope analog and electronic signal path (radio frequency (RF) and intermediate frequency (IF)) are more generally referred to as scientific data. Scientific data are very high rate, can vary greatly in payload size, and can have different structure depending on the science use case. This fits perfectly into the definition of "big data": volume, velocity, and variety. Handling scientific data often involves custom solutions such as a correlator, parallel data paths, and distributed computing environments for processing and visualization.

Sampling of scientific data for diagnostic purposes can be accomplished at a more moderate date rate. In normal operation the DBE will provide low rate signal information such as power measurement statistics that are comparable in data size to other tag sources (such as voltage, current and temperature). But the DBE can also generate higher rate data in the time (oscilloscope) and frequency domains (spectrum) over the data acquisition network and can transmit these directly to the antenna supervisor. These data will typically be used for edge computing to detect system faults that can not be determined directly from the lower rate statistics data, but could also be streamed to a user GUI for engineer inspection. However the high rate of this data (>= 100Mbit/sec is likely) brings up design challenges of treating them like all other tag data as described below.

Tag data are typically low rate and use a small payload. While the structures of tags can vary from one another (giving us the variety component of big data), usually a given tag does not change in structure over its lifetime. Due to ngVLA's limited operational budget, tag data must be handled using cost-effective solutions which implies COTS rather than custom. Even with more forgiving parameters, the size of the tag data over the lifetime of the ngVLA could become non-trivial to store[1]. At ngVLA scale, aggregated over the array and time, even the tag data could meet the definition of big data.

Handling Tag Data

Tag data is anticipated to be sampled at roughly 1Hz under normal operating conditions, though the exact range could be anywhere from intervals of tens of seconds for slowly varying tags, to hundreds of hertz for faster varying values such as currents, voltages,

[1] Based on rough math from ALMA experience, the total number of tags could approach or even exceed one million for the ngVLA. Assuming a (small) one byte payload per tag, sampling each of these at one second intervals over a period of more than twenty years pushes this data into the petabyte regime.

axis encoders, etc. This data will be sent to a central storage system and is planned to be retained for the lifetime of the array. Tools will be developed to provide engineers ways to view and analyze the data for reactive and predictive maintenance.

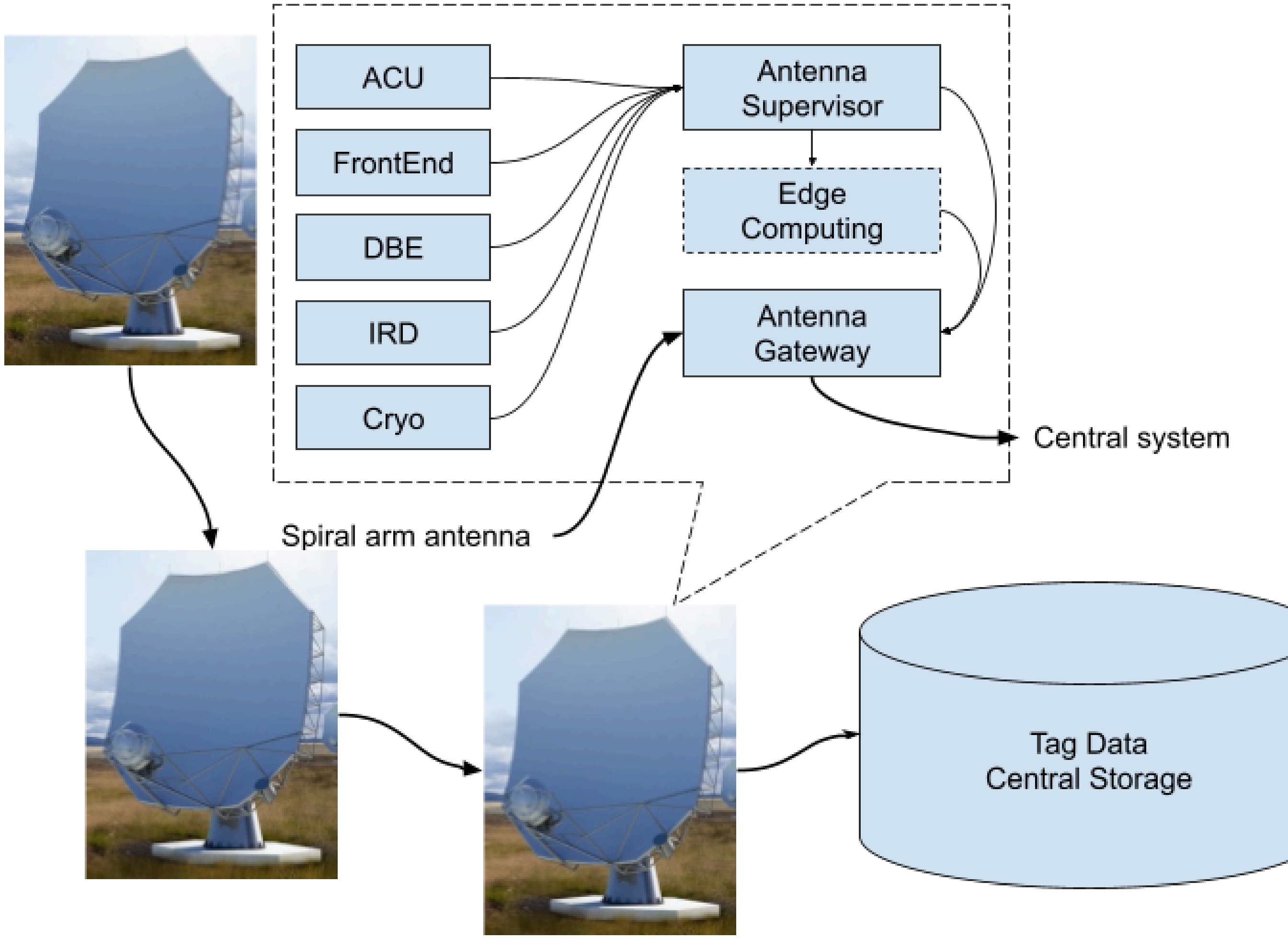

Figure 1: Crude outline of tag data flow in a daisy-chain spiral arm network. Here the antenna gateway is assumed to be a device (e.g. a switch) separating the local network within the antenna from the array network. Spiral arm traffic is passed along through each gateway from antennas further from the center, but does not enter the local antenna network. The aggregate network traffic in this topology increases as we go from the furthest spiral arm antenna inward to the array center.

Oscilloscope mode is an enhancement to normal monitoring, allowing for capture of even higher rate samples (roughly kHz-MHz) that can show transitory effects not visible at the normal sampling rate of a tag. Examples of such effects could include power supply ripple or momentary shorts on a sensor. This data are imagined to be useful primarily for short-term debugging purposes, and as such is not planned to be stored long-term. How oscilloscope mode data may be stored and accessed is discussed in a later section.

Data Retention

In the VLA and ALMA systems, tag data have traditionally been retained forever. For ngVLA, it will almost certainly be necessary to set up data retention policies for tag data, possibly including oscilloscope mode data if they are sent to the tag data storage system. It is certain that some tags will still be stored indefinitely, such as weather information that can be used in reprocessing scientific data or used for historic simulations (e.g. scheduling).

It may be difficult to identify what the useful lifespan of each tag is; at times tag data that are never reviewed can become incredibly useful in identifying a particular problem. But as a general rule, bringing back the tag data subcategories from above, system health data could eventually be decimated, averaged, or discarded after a sufficient time. Even some SDM metadata could potentially be discarded, since they are duplicated in archived data sets.

Edge Computing

As part of the ngVLA maintenance strategy, we plan to use edge computing to assist in early detection of hardware and software issues in the system. Edge computing is a concept in the Internet of Things (IoT) and Industrial IoT (IIoT) architectures, where computation and storage are moved closer to data sources. Under this model, a machine within each antenna (e.g. the M501), or machines near central components such as the correlator and central LO, would receive tag data in tandem with the central monitoring system to perform data science to detect problems, generate alarms, or flag and transmit tag data to the central system for human review. One advantage of Edge computing is that tag data, either from the Hardware Interface Layer (HIL) or from a "smart" device directly (such as the DBE), could flow at a higher rate to the Edge system than is possible to the tag storage system. This could be leveraged to improve detection algorithms implemented on the edge computer.

This functionality is envisioned to be useful for detecting conditions which may be harder to deal with elsewhere in the data acquisition system, for example:

- Long term slopes in values, such as decreasing power, which could indicate impending LRU failures[2].

---

[2] ALMA engineers currently use tag data analysis tools to detect this exact condition within IF modules for predictive maintenance.

- Complex configurations which indicate a problem state; i.e. a particular set of values or a trend of values across multiple devices representing a known failure mode.
- “Triggered” capture, where a predefined condition could enable either oscilloscope mode or simply increase “normal”-rate tag data capture, in order to localize a prior failure or problematic state in the system.

It should be noted that a catalog of these failure modes will need to be accumulated over time, and many will not be known when the ngVLA first enters operations. However, once proper rules are established, edge computing should greatly reduce the need for direct intervention in failures, and will hopefully assist in predictive maintenance to decrease array downtime and time to repair.

Oscilloscope Mode

The oscilloscope functionality as defined in SYS3105 does not fit into the scientific category as it is explicitly for engineering diagnostics. There may be use cases where tag data could be used for other purposes that skew closer to scientific; this will be discussed in a later section. However, it is unclear whether oscilloscope mode data fits into the monitoring category either.

Below are some considerations for whether oscilloscope mode data should be handled the same as tag data, or utilize a different mechanism. Edge processing could be combined with any of these approaches, though the relative ease or difficulty would vary depending on how much additional software is required.

O-Mode as Monitoring

If oscilloscope mode were integrated into the system via a monitoring channel, it would come with a number of caveats:

1. Data retention policies would need to be put into place with respect to the tag data storage system. Oscilloscope mode data would be averaged or deleted after a short period (e.g. a few weeks) in order to avoid overwhelming the tag data storage system over the lifetime of the ngVLA.
2. Sanity checks would need to be installed such that when oscilloscope mode data is enabled, it is not left turned on for extended periods of time.
3. The system would also need to ensure that reasonable bandwidth limits are enforced based on the physical network topology. For example, if a spiral arm of antennas had a daisy-chained data acquisition system network, the system would not allow more than a certain number of tags to have oscilloscope mode enabled on that spiral arm.

The main benefit of treating oscilloscope mode the same as other tag data is that engineers could use the same tools to query, inspect, and visualize the data. It could also make use of edge processing with little to no modifications to the edge processing pipeline for normal tag data.

Independent O-Mode

If oscilloscope mode is separated from the tag data acquisition functionality of the online system, different technological solutions would need to be developed to support the transfer, storage, query, inspection, and visualization of this data. Bandwidth caps would also need to be enforced on whatever network the data utilizes, as it will most likely still travel on the data acquisition system network.

Edge processing may be more complicated in this scenario due to the need for different ways to ingest or store the data in a central system for review.

O-Mode as Edge Processing

Another possibility is to use oscilloscope mode data solely as input to the edge computing nodes; that is, tag data would not be sent to the tag data storage system by default. Engineers could define rules within the edge computing system to dictate how to react to specific conditions. Such rules could include anything from automatically resetting a device based on detected features in a spectrum[3], to recording data of interest for human review.

This edge computing approach would alleviate concerns of overwhelming both the data acquisition system network and the tag data storage system. There would still need to be bandwidth limits between the edge computing nodes and the data sources (e.g. within each antenna), but that is an easier problem than accounting for a larger collection of devices and the network topology within the array.

Non-Maintenance Applications

While the primary requirement for fast read-out data is for maintenance type purposes, there are possible scenarios where this data could be useful for other applications.

---

[3] One example raised in discussions was the VLA digitizers losing word alignment, leading to a characteristic shape in the bandpass plotter.

One planned capability of the DBE is to provide lower rate data that represents a spectrum analyzer functionality via the data acquisition system. This spectrum analyzer functionality could surely detect maintenance problems, but could also detect signals that might be considered RFI to astronomy science, and might in and of themselves be interesting.

DBE Spectrum data could be used for:

1. Research/Development on RFI rejection
    a. Edge computing could be used to process o'scope data provided by the DBE and detect RFI sources. This data could be used to train AI models to detect and mitigate the RFI.
        i. For example some RFI sources are pulsed; a simple approach to mitigate Pulsed RFI sources is to detect the data, and then reject science data during periods of RFI
        ii. AI Models could fit the observed signals to known signals, and once characterized an ideal mitigation could be enabled for Science Data
    b. While RFI rejection is a goal of ngVLA, algorithm development still needs to occur, if ngVLA is built before this occurs o'scope data could be downloaded to help develop algorithms.
2. Extraterrestrials
    a. O'Scope data could potentially be sent directly to a simplified SETI detector. While a SETI detector could operate on the full-rate Science data, it could be possible to operate on lower rate o'scope data if the signal of interest was presumed to be detectable on a single dish. This might allow a low cost (labor and hardware) seti detector to be bolted onto ngVLA without the high cost of hardware capable of processing the main science feeds.
3. ngRadar on ngVLA
    a. If a low bandwidth waveform is used, DBE hosted sample rate conversion could provide I/Q data to a Radar processor (however, this application is probably best implemented over the science network as higher bandwidth would be possible)
4. Defense related applications
    a. Some defense applications rely on detecting "novel" signals. ngVLA could help detect new space based signals of interest and relay information to the relevant authority about the unknown signal. This could have overlap with the RFI Use case, as both involve characterization of known (and unknown) RFI sources.
5. Commissioning
    a. During Commissioning the Spectrum Monitor and other monitoring features can be helpful to determine functionality of a new Antenna.

Conclusion

It is the opinion of the authors of this memo that edge processing is an attractive concept for dealing with oscilloscope mode data. It gives the ngVLA the flexibility to record targeted tag data snapshots without overwhelming the tag storage system, as well as the option to directly analyze and react to predefined events within the system without the need for human review or intervention.

Edge processing also provides a potential path towards accomplishing RFI identification within the time series scientific data (identified as block “A” below).

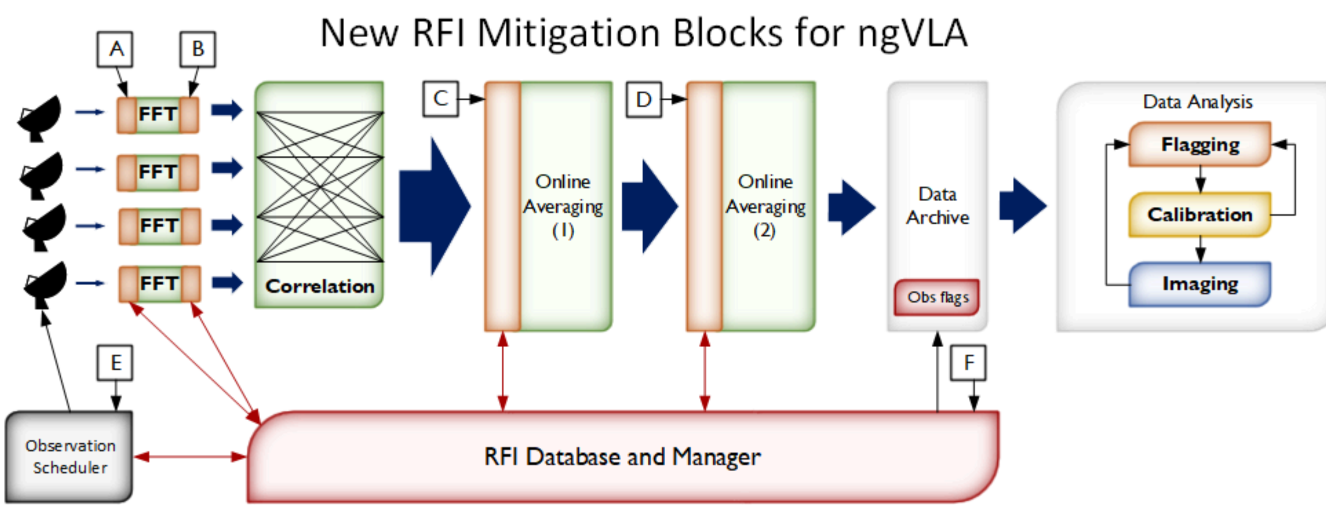

Figure 2: Data acquisition flow diagram illustrating where RFI detection could be implemented. Taken from ngVLA Memo #71, figure 3.

Use cases involving full rate scientific data, such as those identified in (2), (3), and (4) in the previous section, could be accomplished using the existing scientific data path as new backends connected to the correlator switch fabric. We see no reason to reinvent the CBE in edge processing for these applications.