D R A F T

SonATA Overview

John Dreher
2005 March 07

The SETI on the Allen Telescope Array (SonATA) project is the successor to the current New Search System (NSS). This system is intended for use for large targeted searches using the Allen Telescope Array (ATA) when it is completed. This memo presents a very rough first attempt at an overall system description – it is meant to clarify the thinking of the Sonata team and to provoke discussion. Nothing is set in stone yet, nor even Jell-O®.

Figure 1 shows the three subsystems of SonATA and how they relate to the subsystems of the ATA. The Phased Array Network Data Interface (PANDI) receives the sampled voltages from the phased array beam outputs of the ATA Digital IF Processor. The ATA currently plans to provide four independent RF tunings, denoted “IF channels”, that may be placed anywhere within the 0.5 to 11.2 GHz frequency coverage of the ATA. Each tuning in turn provides four independent dual-polarization beams that can be directed to any point on the sky (although in most applications they would be directed to directions within the primary beam of the ATA elements). The polarizations are orthogonal linear designated “X” and “Y”. PANDI breaks the ~100 MHz bandwidth of each beam into smaller units and places them on a set of 10 Gigabit Ethernet or Infiniband links for use by the detector subsystem, DX. DX is made up of multiple Gigabit Ethernet (GbE) networks and a large number of general-purpose servers that execute the SETI detection algorithms in software and, as economically justified, in additional, specialized hardware. The results of the detectors are passed to the SonATA control system, which also controls the operation of DX and provides the interface to the control software for the ATA. Since PANDI must maintain stringent synchronicity with the ATA Digital IF, this subsystem is controlled by the ATA directly – indeed in a functional sense PANDI is more properly considered a part of the ATA and will undoubtedly be used to support other “backends” than SETI, since its function is very general.

In respect to the NSS, PANDI replaces the current ADC, STX, F2Cu, and DDC. DX evolves from the Programmable Detection Modules and the control system evolves from the Search System Executive (SSE).

The present, very preliminary plan for the large targeted search will use four beams from one IF channel to search four target stars at the same frequencies simultaneously. By comparing the results for each signal detection from one star against those from the other three, we should be able to eliminate almost all RFI. But with four ~100 MHz search units observing, we will certainly have some noise “hits”. A second, smaller set of detectors fed from a second IF channel will reobserve these detections, and, if they were just noise, eliminate them. Confirmation strategies will be discussed in a later memo, but I hope that we will be able to get one of our detectors up and tested on some other

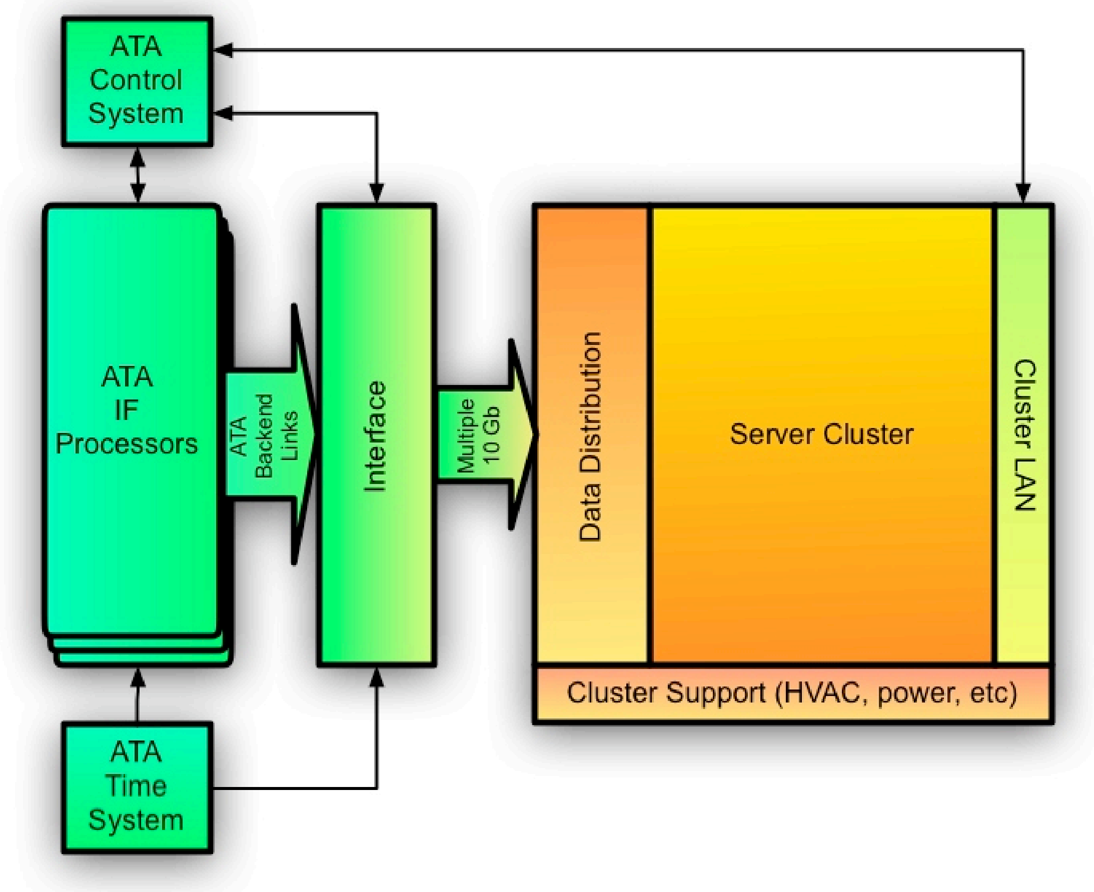

Figure 1. Simplified block Diagram of the SonATA system hardware and its relation to the ATA signal processing.

large telescope, such as the GBT, so that we can use our present, highly-successful very-long-baseline interferometry to demonstrate the signal source is (or is not) a point coincident with the target star to very high accuracy.

PANDI

The ATA Digital IF Processor provides a correctly delayed and phased summation of the complex voltages of the ATA elements as a stream of complex voltage samples on a serial link. Since the bandwidth of the ATA beam is greater than can be processed on a standard server, it is necessary to break this wide bandwidth down into more manageable chunks, of less than 1 MHz, using a Digital Filter Bank (DFB, also known as a polyphase filter bank). For the purposes of this memo I will designate the complex voltage samples representing these narrower chunks as *filaments*.[†] The ATA provides about 50 dB of antialiasing in the digitizers and about 40 dB of image suppression in the quadrature downconverters (after digital tweaking), so it seems reasonable that the DFB be specified to have out of band rejection of 50 dB or better; this will probably require a preconvolver of length 8 or greater. Note that the signal bandwidth of each filament will need to be oversampled by ~20% in order to prevent aliasing. In addition, PANDI also receives timing signals from the ATA that will allow the data packets in each filament to be time stamped. It is also planned to provide a data flagging signal that can be used to blank the signal, for example during a radar pulse. PANDI should also allow for selection of a subset of the filaments, so that we are not forced to send the aliased band edges (only 90% of the Nyquist bandwidth of the ATA output is alias free) and the DC band, and choice of sample requantization. (It may even be possible to choose the number and bandwidth of the filaments.)

The present baseline ATA design has the data organized by IF channel and polarization in order to minimize the connectivity problems associated with summing so many elements. This design has all four beams for one IF channel and polarization sent out on a total of 6 optical fibers and provides 16 bit samples for Re and for Im. Since this organization would be awkward for PANDI, I have suggested changing the size of the complex samples to 12/12 bits, which would allow the four beams to come out on four independent fibers. In that case PANDI might be organized by beam, with each unit accepting two input fibers, one for X and one for Y polarization, and sending the results over the 10 Gb links to routers that will distribute data to the DX units (over GbE) associated with that particular beam and frequency range.

An important unresolved question is where the conversion from dual linear to dual circular polarizations is to take place. In some ways, PANDI is the natural location for this operation, but that choice would imply that each beam be processed as a polarization pair. The alternative would be to leave this to the "backend" itself, in our case the DX

[†] The more 'natural' term might be *thread*, but that term, as well as *stream* are already in widespread use in computing. Similarly, the detectors already use the term *slice*. A more elegant and/or shorter designation would be nice: ideas?

subsystem. This choice would allow the modules of the PANDI subsystem to handle one beam, one polarization each and simplify their design.

Another important decision is how many bits to use in representing the complex voltages for each filament. The DFB will, I suspect, produce 16/16 bit results, but reducing this to an 8/8 or even a 4/4 representation would save in network costs – the tradeoff is in dynamic range, since we only need a few bits to capture the essence of the noise

It seems likely at present that PANDI will be implemented using integer arithmetic in FPGA's. If this proves true, then it is essential that saturating arithmetic with sign preservation and rounding to even be used in all operations. Each DFB will require about 11 Gops/s sustained per beam per polarization; if the implementation of PANDI is delayed for 2-3 years, then it may be possible to implement the DFB using an off-the-shelf server rather than an FPGA; in this case the PANDI might evolve to a server with one or two custom interface cards to handle the ATA synchronous data and timing protocols.

Figure 2 sketches out the data rates for one beam of one IF channel.

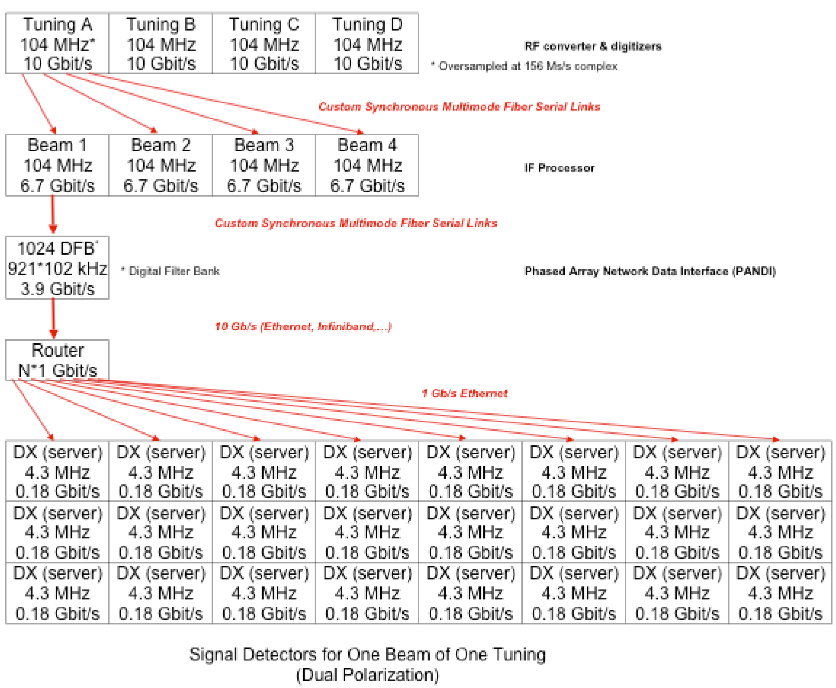

Figure 2. Approximate data rates for one of the 4+ beams analyzed by SonATA.

DX

The DX subsystem does the actual detection work. It includes the network hardware needed to route the packetized filament sample streams from PANDIs (on 10 Gbit/s links) to multiple GbE links that feed the general-purpose servers that will implement the detection algorithms. The tasks of each module of the detectors is similar to those of the NSS PDM with the following differences: the voltage samples arrive via (probably one) GbE and are placed into a circular buffer in memory. The "brickwall filter" task of the current DDC and all the tasks of the "Blue Wave" DSP board migrate to software implementations running on the host processors. We also need to place a high priority on developing ~100-second coherent non-drifting CW and pulse detectors to comply with the recommendations contained in SETI 2020. An important open question is whether "non-drifting" must include some search range in drift and/or curvature. A slightly lower priority should be given to the current NSS detectors, including multi-resolution pulse detection. Development of CW detectors for higher drift rates is worthy, but of lower priority. Minimal computing requirements will be of order 1 to 3 Gops/s per MHz of bandwidth processed. In order to maximize throughput a large amount (5-10 Gbyte) of RAM would be preferred, but it is not clear whether we can afford that; the alternative is the use of disk as in our current implementation, but this might have adverse effects on the bandwidth per server.

The overwhelming driver is cost reduction. The hardware of DX should be viewed as an expendable with a lifetime of 3-5 years. In order to keep NRE under control, we need to consider the relative costs of optimizing code versus just using more hardware. In order to maximize the possibilities for donations of the server "farm" that will form the core of DX, we should also strive to use generic hardware from large vendors rather than the cheapest hardware. For the same reason we should, as we have done in the NSS, stick with common operating systems and vendor supplied processor-dependent libraries wherever possible.

Last but not least we must pay close attention to racking, power, and cooling issues to provide a low-maintenance system that we can afford to run.

SonATA Control System

The interfaces to the units of the DX subsystem will be via another Ethernet link from each host. Management of the DX units should be very similar to the tasks of the current SSE software system, extended for the new detector classes. The interface between the Sonata system and the ATA control system needs to be defined in conjunction with the Science team and the ATA software team so that we can develop suitable observing protocols both for the situation in which Sonata is selecting targets as the primary observer and for the case in which Sonata is running in a complementary mode with an astronomy program (probably a survey) as the primary observer. At first the latter case will be facilitated by mutual design of the two observing programs by the relevant science teams; in future we may evolve the ATA control software to act as a broker in a "market oriented" resource allocation system.

The user interface to the SonATA control system will need to be readily used over the network, since we do not plan to have a permanent observing staff at Hat Creek. Indeed, the goal should be to have the observing proceed with no continuous human supervision in order to reduce the operating cost of the SETI effort. The ATA control system will provide a set of suitable Java-based applications and applets that can be used for secure control of the array over the Internet – these should probably be the basis of the user interfaces to SonATA.