MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HAYSTACK OBSERVATORY

WESTFORD, MASSACHUSETTS 01886

Telephone: 781-981-5400
Fax: 781-981-0590

27 May 2011

TO: VLBI2010/mm-VLBI development group

FROM: Alan R. Whitney, David E. Lapsley, Mikael Taveniku

SUBJECT: Mark 6 VLBI data system; 16 Gbps recording demo

<u>Introduction</u>

The Mark 6 system is being jointly developed by Haystack Observatory and XCube Communications as a next-generation disk-based VLBI data system, and is designed to support sustained data rates to 16Gbps writing to an array of disks. Begun as a collaborative project in late 2010, the system is based completely on COTS hardware and has demonstrated 16Gbps sustained recording (to 32 disks) from digital-backend systems based on the Casper ROACH board; zero-baseline data of broadband correlated noise have been successfully recorded and correlated on the Haystack DiFX software correlator.

<u>Introduction</u>

The demand for increasing data rates for VLBI observations is particularly acute in two disciplines: 1) mm-VLBI observations, which are typically starved for sensitivity, as well as the need to gather as much data as possible during the atmospheric coherence period of 30-60 seconds, and 2) geodetic-VLBI observations, which must gather as much data as possible over a multi-GHz bandwidth in a period of 5-15 seconds. Both of these disciplines are eager to capture data at 16-64Gbps and record to disk. The Mark 5 series of VLBI data systems is limited to a maximum of 4Gbps (Mark 5C) and cannot support 16Gbps and higher data rates without many systems operating in parallel, which is both expensive and cumbersome. The Mark 6 system, based fully on commercial-off-the-shelf (COTS) hardware, is being jointly developed by MIT Haystack Observatory and XCube Communications, Inc to fulfill these needs. A prototype Mark 6 system has demonstrated VLBI data recording at sustained 16Gbps, with burst-capture capability to ~32Gbps.

<u>Origin of the Mark 6 system</u>

The basis of the Mark 6 system is derived from a disk-based data-logging system developed by XCube Communication of Nashua, NH for testing automated-driving systems for the automotive industry. Cars to be tested are outfitted with XCube 'loggers' to capture real-time technical data from the automobile, as well as from numerous car-mounted HD video cameras pointing in all directions to continuously capture the complete environment around the vehicle; this allows thorough examination of the environment surrounding the vehicle to understand reactions of the automated-driving software, and to diagnose unexpected actions. The sustained data-rate requirement for this application is ~6 to 12 Gbps, with the recording equipment mounted in the trunk of a moving automobile that must operate properly while absorbing all the usual bumps and motions of driving. As a result, the XCube 'logger' was specifically designed to sustain required recording rates in the face of disks whose performance may be degraded by the harsh environment.

Though VLBI-data disks do not typically suffer a bumpy ride during recording, they often suffer a bumpy trip from correlator to station, which sometimes has a negative effect on the performance of some disks; use of such affected disks can threaten required sustained recording rates if used in ordinary RAID disk arrays, which are insensitive to real-time requirements. The techniques used in the XCube 'logger' have been extended to the Mark 6 VLBI data system, but at much higher data rates. Both the XCube 'logger' and the Mark 6 system are based completely on high-end COTS hardware with highly specialized control software.

Figure 1: Front view of Mark 6 controller with disk module

Figure 2: Rear view of Mark 6 controller with disk module

<u>Main Characteristics of the Mark 6 system</u>

Hardware

As indicated above, the Mark 6 controller incorporates only COTS hardware, though components are carefully selected for performance and compatibility with other system elements. A high-end motherboard, CPU and RAM memory are used to maximize performance; a standard Mark 6 system is outfitted with ~12GB of high-speed RAM memory, though up to ~128GB can be accommodated for burst-mode applications. The Mark 6 prototype system is shown in Figures 1 and 2.

Software

The Mark 6 system operates under an Ubuntu Linux OS, though with some modified kernel code and drivers for efficient memory and disk management. Application software is written primarily in C++ and Python. The Mark 6 performance arises from a technique decoupling network and disk datastreams with elastic buffers and kernel-bypass DMA for data transfer.

Input

The Mark 6 can accommodate up to four 10GigE data-input ports, each operating at up to ~7Gbps. Each input may operate at a different data rate. Normally, data are transmitted to the Mark 6 in a UDP packet stream to minimize load on the Mark 6 Ethernet NIC cards and sustain the highest possible data rate.

Disks, disk modules and connection to disks

The Mark 6 system supports only SATA-interface disks; furthermore, it has been discovered that not all disk vendors implement the SATA interface in exactly the same way, which causes some vendor's disks to operate more capably with the chosen disk-interface card. As a result, for maximum performance, it is currently necessary to specify particular disk vendors, though low-cost commodity (i.e. non-enterprise) disks from these vendors are acceptable.

Mark 6 disk modules are similar to Mark 5 disk modules (8 disks per module) except that Mark 6 disk modules connect to the controller via COTS external-SATA cables. Each external-SATA cable supports 4 disks, so that a module requires the connection of two such cables plus one power cable. Any cable may be connected to any disk module; the controller sorts out which disks it is connected to. XCube also offers a module that houses 16 2.5-inch disks that requires the connection of four external-SATA cables. All XCube modules include an internal fan.

We are investigating the feasibility of developing a conversion kit that will permit existing Mark 5 disk modules to be modified to operate with the Mark 6 system, allowing significant cost recovery of the original module expense (though existing PATA modules will need to be outfitted with SATA disks). A conversion kit would include a new module backplane, plus a new front panel with appropriate connectors. Because the Mark 5 modules have built-in cooling, an external cooling fan must be provided, which could be provided by an (otherwise idle) Mark 5 chassis or a simple chassis that accepts the modules and provides fan cooling.

The use of standard external-SATA cables to connect the Mark 6 controller to the modules is different from Mark 5 but is quite workable. The heaviest wear is on the cable ends connecting to the modules. Cables can be reversed to double their lifetime (projected ~500 connect/disconnect cycles for each cable end). When a cable connector wears out, the cable is easily and inexpensively replaced.

Monitor & Control and disk management

A VSI-S command set has been specified for the Mark 6 and is being built as a layer on top of the native XML interface provided by XCube.

Depending on recording data rate, different numbers of simultaneously-operating disk modules are required. A single 8-disk module with modern disks will support 4Gbps; two modules (16 disks) are required for 8Gbps, 4 modules (32 disks) are required for 16Gbps. In order to accommodate these different requirements, the concept of a 'volume' has been created to identify the collection of one or more disk modules needed to support a particular observing requirement. A 'volume' is created by 'bonding' a specified set of modules together for the duration of existence of a particular data set. When that data set is erased, the constituent modules become individual volumes again.

In the case that multiple volumes are simultaneously mounted to the Mark 6 (for example, four volumes of single 8-disk modules, or two volumes of two 8-disk modules), they are organized into a logical 'volume stack' that specifies the order of usage; when the currently active (recording) volume nears capacity, the Mark 6 can automatically re-direct subsequent scans to the next volume in the 'volume stack'. The full volume may be then disconnected and a new empty volume attached and added to the volume stack, all while recording continues on the active recording volume. Enhancements have had to be made to the XCube software in order to support this mode of operation required for VLBI operations.

Care is being taken to continuously collect disk-performance statistics to identify the module and serial-number of poor-performing disks so that such disks can be easily physically identified and replaced.

Data format on disks; playback

The Mark 6 writes a single standard Linux file to each data disk for each input data stream. The data stream from each Ethernet interface is gathered into a buffer of order 64MB before attempting to write that data to the next disk in a round-robin sequence; if a disk is not ready at the time the data buffer is ready to be written, the write request is immediately diverted to the next disk in the sequence; this allows data distribution among disks according to their individual performance. Each buffer is identified by a header with a sequence number that, on replay, allows properly ordered data to be reconstructed from the multiple files to which is has been scattered.

Software is being written that, on replay, presents the data from each recorded data stream as a single Linux file; this greatly simplifies interfacing the Mark 6 to a software correlator.

In order to maintain the maximum recording-rate capability, the Mark 6 records the entire Ethernet packet from all data streams and is oblivious to the format of the actual data frame within the Ethernet packet. On replay, Ethernet header and trailer data are normally stripped so that the user sees only the actual payload packet (VDIF or Mark 5B, etc).

e-VLBI capabilities

Full e-VLBI capabilities are planned for the Mark 6.

<u>16 Gbps Mark 6 demonstration at Haystack TOW meeting</u>

On the occasion of the VLBI Technical Operations Workshop (TOW) held at Haystack Observatory in May 2011, a demonstration of the Mark 6 system was held. Figure 3 shows the schematic of the system setup for recording. Two broadband correlated noise sources are fed into a set of four up-down converters (UDC). Each UDC selects a different 512MHz slice of spectrum from each of the two correlated noise sources and translates it to 512-1024MHz (second Nyquist zone) for input into two RDBE digital backend units (based on the ROACH general-purpose signal-processing board developed at Berkeley, NRAO and South Africa, with firmware developed at Haystack Observatory). The output of each RDBE is two 10GigE data streams, each at 4Gbps, for an aggregate data rate into the Mark 6 of 16Gbps. Data are recorded to a set of 32 disks for several minutes.

Following recording (see Figure 3), ~1 second of data were reconstructed from the Mark 6 disks and transferred to a standard Linux file for correlation on the Haystack DiFX software correlator (see Figure 4). The correlation (from the 'fourfit' fringe-fitting program) results from one 512MHz slice of the recorded bandwidth are shown in Figure 5, and are as expected. Ripples on the cross-correlation bandpass amplitude and phase are due to both non-flatness of the noise sources and differences in analog filters along the parallel signal-processing paths across the 512MHz channel bandwidth.

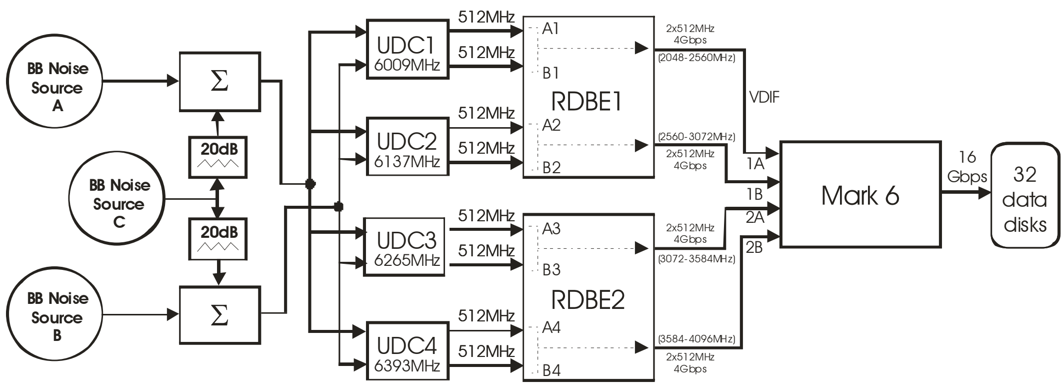

Figure 3: System used for 16Gbps recording

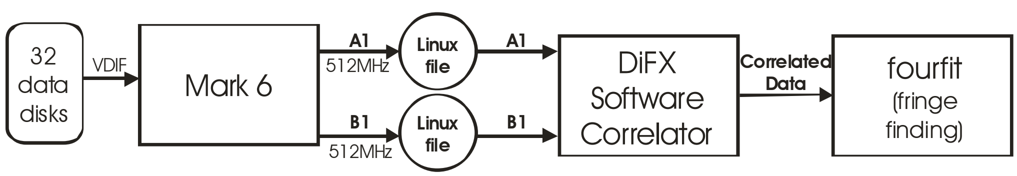

Figure 4: System used for replay and correlation

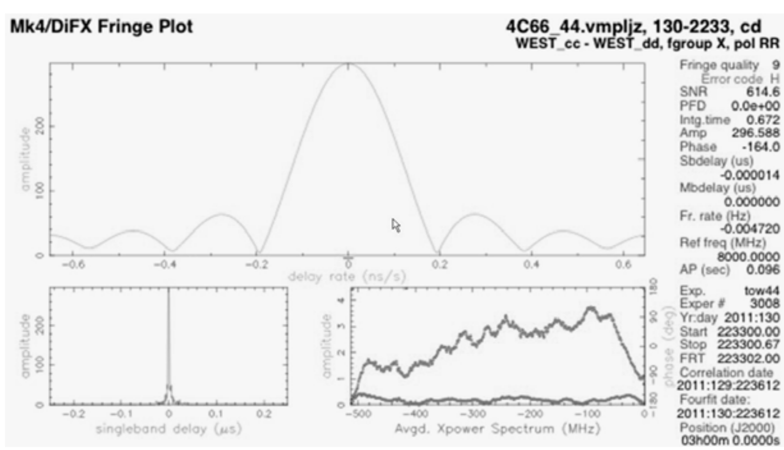

Figure 5: Cross-correlation results through 'fourfit' fringe-fitting program

Summary

The Mark 6 VLBI data system is a major step forward in data-rate capability over previous VLBI data systems, and is the first high-performance system to use fully COTS hardware. The high performance and low cost of the Mark 6 system, as well as future improvements due to the normal progress of COTS technology, will help to maintain its long-term viability. In the short term, mm-VLBI and geodetic VLBI will be the major beneficiaries of this new capability, but in the longer term it will also enable much higher sensitivity over a board range of VLBI observations. The Mark 6 system is expected to be tested in several VLBI experiments in mid-2011 and be available to the general VLBI community in late 2011. More information about the Mark 6 system is available at http://www.haystack.edu/tech/vlbi/mark6/index.html. The Mark 6 system is expected to be available to the general VLBI community in late 2011.