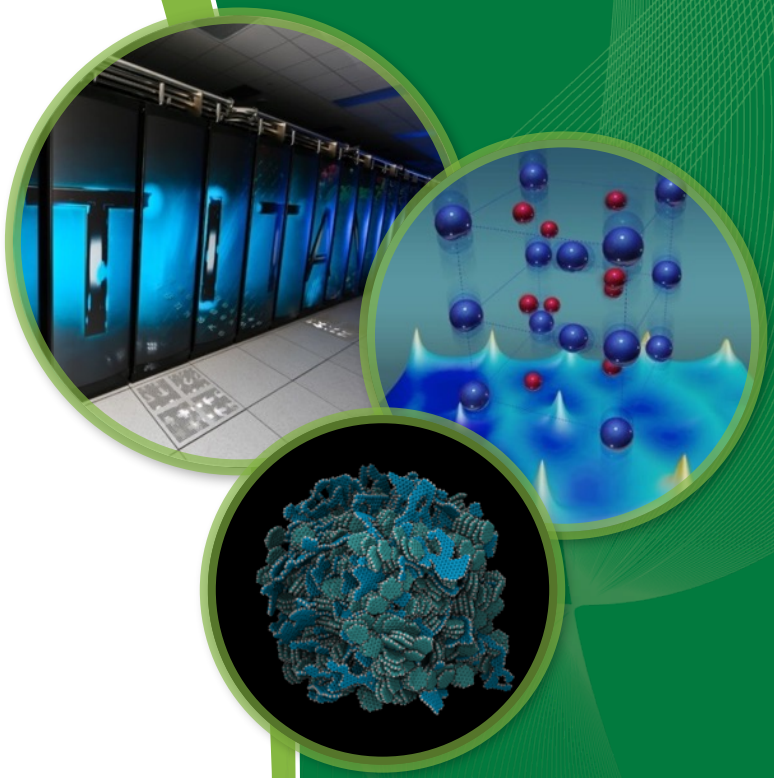
Vendor Connections

Robert Varner



Why Vendor Connections

- Affordability
  - NRE costs at national labs too high for one-off modules
  - NP is way too small a market to amortize costs
- Capability
  - Businesses can leverage
    - market size
    - supplier relationships
  - Respond more quickly to new technology
- Quality
  - Custom modules - one pays again and again until it works
  - Happy customers require quality modules
- Endurability
  - Custom modules break - who fixes them?
  - Components go obsolete

SBIR - topic submitted by FRIBDAQ WG

23.NUCLEAR PHYSICS SOFTWARE AND DATA MANAGEMENT

Maximum Phase I Award Amount: **$150,000**

Maximum Phase II Award Amount: **$1,000,000**

b. Software-Driven Network Architectures for Data Acquisition

Future FRIB experiments anticipate event rates of a few per second to tens of thousands per second. Each event will contain hundreds to millions of bytes of data from the digitizers. The large latencies possible in highly buffered flash ADC architectures can be used to advantage in the design of a new data acquisition architecture. This will have digitizers or digitizer systems connected by commercial network fabrics, moving data to general purpose processors for software “trigger” determination and event building. The hardware architecture is simplified, composed of digitizers, networks and fast general purpose processors. What used to be a largely hardwired logic system is now a software-driven system. The fundamental requirement for success of this is that the data from each detector element be labeled with a precisely synchronized time and location before transmission on the network.

SBIR 28.b (continued)

Applications are invited in the following areas;

1. The development of streaming data acquisition system and control systems:
    1. protocols and data formats to maximize throughput, decrease latencies, facilitate event building, improve efficiency of data retrieval from permanent storage, and facilitate real time monitoring of the detector performance;
    2. trigger decision systems that may be fully software based or have hardware assists (e.g. FPGA accelerators),
    3. data flow systems that are capable of responding to trigger accepts by reading data from the digitizers and making it available to interested clients,
    4. data flow systems that are capable of responding to trigger accepts by reading data from the digitizers and making it available to interested clients,
    5. scalable event builders that accept data from the data flow system and inject back to that system built events for online analysis and, if rate permits, logging,
    6. protocols and middle-ware that can tie this system together and provide relatively high level interfaces to user software;
2. Soft core FPGA module(s) to implement the network protocol for Ethernet and/or Infiniband, able to drive existing and emerging commercial network chips, with sufficient buffering to support data aggregation using a commercial network switch, and with sufficient speed to drive up to 40 Gb/sec network links;
3. Time distribution protocols and hardware to support fine-grained time tagging (with 1 nanosecond or better precision) of each network packet for later correlation and event or frame assembly, possibly integrated with FPGAs and possibly exploiting the commercial network for some aspects of tagging,
4. Data streaming (from up to 1 gigasample per second digitizers) and distribution hardware and software capable of routing time tagged data from several thousands of data sources to temporary storage where the data can be accessed in time order.

SBIR (M. Cromaz)

c. Data Science / Distributed Computing Applications

As discussed above, analysis of experimental data from accelerator-based detector systems is a central task in the NP community. In the case of medium scale experiments such as those planned for FRIB, data sets will be collected with each event having a large number of independent parameters/attributes. The manipulation of these complex datasets into summaries suitable for the extraction of physics parameters and model comparison is a difficult and time-consuming task. Currently, both the accelerator facilities and university-based groups carrying out analysis maintain local computing clusters running domain specific software, often written in an ad-hoc way by the experimentalists themselves, to carry out these tasks. Recently, the data science community has developed tools and techniques for handling such tasks at scale in an efficient and more generic manner. These tools are generally open-source and can be deployed on inexpensive, distributed computing resources provided by commercial web services firms which provide computing resources that are both inexpensive and scalable on demand. Furthermore, these tools hide many of the implementation details required to run efficiently on distributed systems allowing the experimenter to focus on the physics analysis task at hand while fully utilizing a modern computing infrastructure.

Adaption of these new technologies to the analysis of nuclear physics data requires the development of domain specific tools. Specifically, we seek applications for (1) the development high-throughput, low- latency methods to parse and securely transfer binary experimental data to commercial cloud services (e.g., AWS, Google Cloud), (2) distributed data analysis for experimental physics applications implemented using data processing systems such as Apache Spark, (3) the application of machine-learning techniques with standard frameworks (Google TensorFlow, Spark GraphX) to automate analysis tasks and provide intelligent diagnostics, (4) the creation of lightweight packages, leveraging libraries in modern, widely- adopted analysis environments (e.g., python/pandas, r/dplyr), to facilitate common physics analysis methodologies. Applicants are expected to address a specific application domain in experimental nuclear physics data analysis. Proposals should address performance and plan to demonstrate feasibility with working prototypes.

Communication with Vendors

- Readout Software and hardware depends on commercial components
- They can influence us - what technologies do they anticipate
  - FPGA's with 20GbE links
  - 16 bit ADC's
  - Directions we do not even know
- Can we influence them?
  - What changes would make system coordination easier
    - event buffer standards?
    - clock synchronization

Communication with vendors

- These are significant vendors in nuclear data acquisition
  - CAEN Technology - Marco Locatelli, Greg Kibilko, (Ron Fox)
  - Wiener, Plein and Baus; Mesytec - Andreas Ruben
  - Struck Innovative Systems - Matthias Kirsch
  - XIA - Hui Tan
- All were invited to this meeting; We need to learn how to communicate.