SPALLATION NEUTRON SOURCE:
FUTURE NETWORKING SCENARIOS

J.P. Hodges SNS Division, Oak Ridge National Laboratory

Spallation Neutron Source facility at ORNL.

1.0 INTRODUCTION

Completion of the Spallation Neutron Source (SNS) facility in early 2006 heralds a new era for neutron scattering sciences in the US. The world-class SNS instruments will be used by scientists from diverse fields such as Physics, Chemistry, Materials Science, Earth Science, Engineering, Biology and Pharmacology. Each SNS instrument will offer users at least an order of magnitude performance enhancement over any of today's pulsed spallation neutron source instruments. This great increase in instrument performance is mirrored by an increase in data output from each instrument. In fact, the prevalent use of highly pixelated detector arrays and supermirror neutron guides in SNS instrument designs means that the data output rate for each instrument is likely to be close to two orders greater than a comparable US instrument in use today. Therefore, writing an entire experiment data set to CD and completing data reduction and analysis on a single PC at the facility or user's home institution, as is common today, will not be a viable solution for the SNS. Resolution of this problem will involve at minimum significant data reduction to be completed offsite on a high-performance computer system. Preferably, high-performance network and computer systems will handle all aspects of accurate (few approximations) and real-time (many approximations) data mapping, multi-dimension visualization, and comprehensive data analysis.

Three high-flux neutron scattering facilities within the US (NIST, LANSCE and HFIR each possessing a diverse suite of instruments) have recently or are now undergoing substantial upgrades. Shortly, it is expected (already realized for NIST) that the combination of source, guide/transport and instrument enhancements will greatly improve the usefulness of these facilities. Clearly, any advanced data transform, visualization and analysis capabilities harnessed through the use of high-performance network and computer systems by the SNS will be applicable to these facilities and, indeed, will be requested by the user community. In this expanded role, a versatile user-expandable computing 'toolkit' for the manipulation, visualization and analysis of neutron scattering data would necessitate a high-performance network connecting all the facilities.

2.0 NETWORK USAGE SCENARIOS

The expected demands placed on computer networks by SNS are illustrated below in three usage scenarios. These scenarios cover data mapping, real time visualization and partial analysis during data collection, and comprehensive data analysis. Data rate determining instrument parameters and estimated (peak & time-averaged) data rates are given in Table I for the twelve SNS instruments currently under design and/or construction. The complete SNS instrument suite will include approximately twelve more instruments.

2.1 Data Mapping

All ‘raw’ data collected on the SNS instruments must be processed to yield an intelligible data set. This processed data set called the response function is effectively an *n*-dimensional (n = 1, 2, 3 or 4 depending on the experiment) probability distribution. In a broad measurement the response function is the sample *scattering law* convoluted with the instrument resolution function.

In general, instrument users relate to their experiment through the measured response function and not the raw data. Consequently, with every experiment there is a considerable amount of raw data to be processed. Effectively this raw data processing, typically involving spectra merging, normalization, various corrections, rebinning, etc, is a remapping of the observed data. For the diffractometers and reflectometers, which produce the largest number of raw data sets a day, the data mapping process leads to a considerable amount of data reduction, see Table II. The data mapping step is essential to keep the amount of data produced daily from the diffractometers and reflectometers manageable in terms of storage. The last column of Table II lists the typical amount of experiment data required to be stored daily from each instrument. Across all twelve instruments this amounts to 80 Gb a day or 16 Tb a year (since SNS will operate ~200 days/year).

The accurate determination of the response function is an iterative process and with regard to availability and speed requirements is certainly best performed across a distributed computer network. As envisioned, the twelve SNS instruments listed in Table I would together generate a time averaged data stream of around 10 Mb.s^{-1} for processing across the distributed computer network. Assuming 1 minute duration to distribute a large raw data file, a peak rate across the network of 40 Mb.s^{-1} is estimated. The return data stream of mapped data would be 1 and 15 Mb.s^{-1} time averaged and peak, respectively.

2.2 Real Time Visualization and Analysis

In contrast with the High-Energy Physics and Fusion Energy disciplines where experiments are highly collaborative and secure assess to data is required for tens to hundreds of researchers around the globe, neutron scattering experiments are much smaller affairs and typically access to the data may be required for perhaps five people distributed between the neutron facility and the principal investigators home institution. However, since, neutron scattering instruments operate 24 hrs 7 days a week during facility run periods, real time data visualization, some real time analysis capabilities, and security to modify experiment conditions by a user at his/her hotel via an internet browser is desired.

In this scenario, the combined data transfer between the twelve SNS instruments and a distributed computer network for real time data mapping is estimated to be a constant 140 $Mb.s^{-1}$ (assuming 50% of users using real time visualization). The return data stream to servers managing the visualization and analysis tasks as well as communicating to the users across LAN and/or internet would be around 18 $Mb.s^{-1}$ (dominated by the 4-D and 3-D response maps). The servers (one for each instrument) would generate selected views of the response function as well as send (if requested by the user) the response function back out to the distributed computer network for quick/partial analysis.

Users are given a specific amount of time (0.5 to 2 days) on an instrument. The close to real-time visualization and partial analysis capabilities, therefore, allow a user to refine the experiment during the allotted time. For the majority of SNS user experiments, the material or property being studied is novel and this capability is essential for the experimentalist to focus in on the area of interest and maximize the science accomplished in the limited amount of beam time.

2.3 Comprehensive Data Analysis

Looking forward 5-10 years, it is anticipated that SNS users will be using a common 'tool kit' for the manipulation, visualization and analysis of neutron scattering data. This 'tool kit' will be a logical evolution of the software library that will be initially developed for data mapping (described in §2.1). At first it is expected that the analysis tool kit will provide fairly simple capabilities for fitting neutron scattering data, e.g. least-squares statistics. As the analysis tool kit is developed more computing intensive (Bayesian, maximum likelihood, Monte Carlo, etc) statistical fitting methods will be incorporated. Additionally, analysis of experimental data may be achieved by incorporating a scattering law model within the iterative response function extraction procedure. These advanced analysis methods are expected to require the use of powerful offsite computing systems.

Assuming the fairly simple analysis tool kit scenario, the network data transfer requirements can be estimated from the amount of data stored daily from the instruments (see Table II). Accounting for the fact that an experimental data file will be passed across the network several times (5 times is assumed) before a user is satisfied with their analysis and a maximum of one minute for transferring an experiment data file, then network data transfer rates of 5 and 15 $Mb.s^{-1}$ time averaged and peak, respectively, are required.

3.0 CONCLUSION

The three scenarios described above will be operating simultaneously at the SNS. Considering, therefore, the individual scenario networking demands as cumulative the data transfer requirements for the high-performance network are 200 $Mb.s^{-1}$ peak and 150 $Mb.s^{-1}$ time averaged. The demands are estimated for an SNS facility with only a 50% complement of instruments. A full instrument suite could double the network data transfer requirements.

Table I. SNS instrument data rate determining parameters and estimated data output rates.

Instrument	n[a] (n.cm^{-2}.s^{-1})	Ω[b] (sr)	pixels[c]	count rate[d] (n.s^{-1})	peak data rate[e] (Mb.s^{-1})	data set size[f] (Mb)	average time for data set (s)	data rate[g] (Mb.s^{-1})
Powder Diffractometer	3.4×10^7	3.1	40,000	3.5×10^6	28.0[h]	200[i]	60	3.3
Disordered Materials Diffractometer	5.0×10^8	3.5	50,000	4.2×10^7	13.0	400	3600	0.1
High-Pressure Diffractometer	4.0×10^7	1.8	100,000	3.0×10^5	2.4[h]	800	11000	0.1
Engineering Diffractometer	5.0×10^7	2.7	80,000	2.4×10^6	19.0[h]	240[i]	60	4.0
Single Crystal Diffractometer	4.0×10^7	9.0	5×10^6	3.0×10^5	2.4[h]	1,700[i]	3600	0.5
SANS Diffractometer	4.0×10^8	1.0	40,000	2.0×10^7	64.0	320	500	0.7
Liquids Reflectometer	5.0×10^7	0.03	40,000	7.0×10^6	56.0[h]	320	500	0.7
Magnetism Reflectometer	5.0×10^7	0.04	40,000	9.0×10^6	72.0[h]	320	1200	0.3
Backscattering Spectrometer	2.0×10^7	3.0	4,500	1.3×10^7	8.0	72	900	0.1
ARC Spectrometer	5.0×10^5	3.0	70,000	5.0×10^5	3.8[h]	840	1800	0.5
CNC Spectrometer	1.0×10^7	5.1	15,000	7.0×10^6	6.0	180	3600	0.1
HRC Spectrometer	1.0×10^6	1.6	70,000	4.0×10^5	3.3[h]	840	7200	0.1

a: integrated neutron flux at sample
b: solid angle coverage of detector array
c: no. of pixels in detector array
d: neutron count rate across entire detector array
e: appropriate for real time data harvesting
f: histogram format
g: time averaged data rate for transferring the histogramed data sets only
h: every neutron event streamed
i: sparse matrix

Table II. SNS instruments raw, mapped and grouped data files sizes, and daily data storage requirements.

Instrument	raw data set size (Mb)	Response Function	mapped data set (Mb)	data sets analyzed simultaneously	grouped data size (Mb)	data produced each day[a] (Mb)
Powder Diffractometer	200	1-D $S(Q)$	5 × 0.16	100	80	1,150
Disordered Materials Diffractometer	400	1-D $S(Q)$	0.08	4	0.32	9,600
High-Pressure Diffractometer	800	3-D $S(\boldsymbol{Q})$	216	1	216	1,730
Engineering Diffractometer	240	1-D $S(Q)$	5 × 0.08	100	40	580
Single Crystal Diffractometer	1,700	3-D $S(\boldsymbol{Q})$	216	1	216	5,180
SANS Diffractometer	320	1-D $S(Q)$	4 × 0.08	100	32	280
Liquids Reflectometer	320	2-D $R_{xy}(Q)$	0.72	1	0.72	130
Magnetism Reflectometer	320	2-D $R_{xy}(Q)$	0.72	4	3	50
Backscattering Spectrometer	72	2-D $S(Q,\omega)$	1	50	50	6,900
ARC Spectrometer	840	4-D $S(\boldsymbol{Q},\omega)$	800	1	800	38,400
CNC Spectrometer	180	4-D $S(\boldsymbol{Q},\omega)$	200	1	200	4,800
HRC Spectrometer	840	4-D $S(\boldsymbol{Q},\omega)$	800	1	800	9,600

[a] post data mapping if appropriate for the instrument