Research at the Speed of Thought: Combining HPC, networking, storage, i/o, visualization with robotic automated virtual laboratories

Updates from the National Supercomputing Centre (NSCC) Singapore

Presented by:
A/Prof Tan Tin Wee
Director NSCC Singapore

https://www.nscc.sg
“Soon there will come a time when the biologist will not need to travel to the library to research the literature. She will be able to access a wide range of informational resources from her Internet-connected desktop. Knowledge will be at the tip of her fingers.”

Tan Tin Wee, 1993

After winning the MEDINFO’92 Gold Medal, Geneva where Internet access to a biological WAIS and Gopher database was demonstrated live in the conference venue in Palexpo, Geneva, and building the first Singapore Website http://biomed.nus.sg in Sep 1993 (today’s www.bic.nus.edu.sg maintained by Mark De Silva and Lim Kuan Siong in NUS)
2010 InCoB

We discussed:
Building the Robotic Automated Laboratory of the Future

2016 InCoB

Let’s predict:
“The time will come when at the tip of the biologist’s fingers, he will be able to design experiments, have them executed robotically and in automated pipelines, and the results gathered for computerised analysis.”

Tan Tin Wee
After retiring from APBioNet
https://www.youtube.com/watch?v=FfnxB9DDmEg

All images are copyright of Internet Society.
http://www.internethalloffame.org/inductees/year/2012
The future of data can be green | Tan Tin Wee | TEDxSingapore
TEDx Talks
9 months ago • 332 views
In an era of burgeoning data and an increasing need for computing power, there has been an explosive expansion of data centers ...

https://www.youtube.com/watch?v=rc_Ed7VPvBY

http://www.tedxsingapore.sg
RUN EXPERIMENTS IN A CENTRAL LAB
FROM ANYWHERE IN THE WORLD

1 DESIGN
Ship your samples to Emerald, then design your experiments over the web using our software

2 EXECUTE
Emerald conducts your experiments in an automated lab exactly as you specified

3 EXPLORE
Emerald organizes your data into a database on the cloud which you can access from anywhere

4 ANALYZE
Ship samples back to your lab and analyze your results using our extensive data analysis suite

Images are copyright of Emerald Cloud Lab.
LAB EXPERIMENTS

ACCESS A STATE-OF-THE-ART LABORATORY WITHOUT LEAVING YOUR DESK

Labs in South San Francisco

Images are copyright of Emerald Cloud Lab.
Order your experiments

Analytical Balance Readings
Apoptosis Assays
Autoclaving
Buffer Prep
Centrifugation
DNA/RNA Synthesis
Epifluorescence Microscopy
Fast Protein Liquid Chromatography (FPLC)
Flash Chromatography
Flow Cytometry
Fluorescence Kinetics
Fluorescence Polarization
Fluorescence Spectroscopy
Fluorescence Thermodynamics
Genomic DNA Prep
HPLC (Ion Exchange)
HPLC (Reverse Phase)
Light Microscopy
Liquid Handling
Lyophilization
MALDI Mass Spectroscopy
Mammalian Cell Culture
pH Readings
Polyacrylamide Gel Electrophoresis (PAGE)
Polymerase Chain Reaction (PCR)
Protein Extraction
Quantitative Real Time PCR (qPCR)
RNA Extraction/cDNA Prep
Rotary Evaporation
Solid Phase Extraction
Speedvac Concentration
Thermometer Readings
Thin Layer Chromatography (TLC)
Total Protein Quantification
Transfection
UV/Vis Kinetics
UV/Vis Spectroscopy
UV/Vis Thermodynamics
Vacuum Filtration
Viral Prep
Volume Check
Western Blot

Somebody will do it for you!

## What’s coming up Next!

<table>
<thead>
<tr>
<th>Agarose Gel Electrophoresis</th>
<th>Dynamic Light Scattering (DLS)</th>
<th>NMR (2D / Structural)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arabidopsis</em> Studies</td>
<td>Electron Microscopy</td>
<td>NMR (Carbon)</td>
</tr>
<tr>
<td>Atomic Absorption Spectroscopy</td>
<td>Electroperoration</td>
<td>NMR (Proton)</td>
</tr>
<tr>
<td>Atomic Emission Spectroscopy</td>
<td>Electrospray Ionization (ESI) Mass Spectrometry</td>
<td>Organic Synthesis (Milligram to Gram Scale)</td>
</tr>
<tr>
<td>Atomic Force Microscopy</td>
<td>Enzyme-Linked Immunosorbent Assay (ELISA)</td>
<td>Patch Clamp Recordings</td>
</tr>
<tr>
<td>Bacterial Cell Culture</td>
<td>Flow Chemistry</td>
<td>Peptide Synthesis</td>
</tr>
<tr>
<td>Bio-Reactor</td>
<td>Fluorescence Activated Cell Sorting (FACS)</td>
<td>Photostimulated Luminescence (PSL)</td>
</tr>
<tr>
<td>Bomb Calorimetry</td>
<td>Fluorescence In Situ Hybridization (FISH)</td>
<td>Plasmid Construction</td>
</tr>
<tr>
<td><em>C. Elegans</em> Studies</td>
<td>Gas Chromatography</td>
<td>Refractometry</td>
</tr>
<tr>
<td>Capillary Electrophoresis</td>
<td>Gas Chromatography Mass Spectrometry (GC-MS)</td>
<td>Scanning Tunneling Microscopy</td>
</tr>
<tr>
<td>Circular Dichroism (CD)</td>
<td>HPLC (Normal Phase)</td>
<td>Solubility Testing</td>
</tr>
<tr>
<td>Colony Picking</td>
<td>HPLC (Preparative)</td>
<td>Sonication</td>
</tr>
<tr>
<td>Confocal Microscopy</td>
<td>Immunoprecipitation</td>
<td>Supercritical Fluid Chromatography (SFC)</td>
</tr>
<tr>
<td>Crossflow Filtration (TFF)</td>
<td>Inductively Coupled Plasma Mass Spectrometry (ICP-MS)</td>
<td>Surface Plasmon Resonance (SPR)</td>
</tr>
<tr>
<td>Crystallization</td>
<td>Infrared Spectroscopy</td>
<td>Tandem Mass Spectrometry (MS-MS)</td>
</tr>
<tr>
<td>Dialysis (Equilibirum)</td>
<td>Isothermal Titration Calorimetry (ITC)</td>
<td>Tissue Homogenization</td>
</tr>
<tr>
<td>Dialysis (Preparative)</td>
<td>Liquid Chromatography Mass Spectrometry (LC-MS)</td>
<td>Total Internal Reflection Fluorescence (TIRF)</td>
</tr>
<tr>
<td>Differential Scanning Calorimetry (DSC)</td>
<td>Liquid-Liquid Extraction</td>
<td>Microscopy</td>
</tr>
<tr>
<td>Digital Droplet PCR</td>
<td>Melting Point Determination</td>
<td>Ultracentrifugation</td>
</tr>
<tr>
<td>DNA Sequencing (Next Generation)</td>
<td>Microarray Analysis</td>
<td>X-Ray Crystallography</td>
</tr>
<tr>
<td>DNA Sequencing (Sanger)</td>
<td>Microwave Reactions</td>
<td>Yeast Cell Culture</td>
</tr>
<tr>
<td><em>Drosophila</em> Studies</td>
<td>Molecular Cloning</td>
<td></td>
</tr>
</tbody>
</table>

## ARMCHAIR BIOLOGIST?

*Images are copyright of Emerald Cloud Lab, [http://emeraldcloudlab.com](http://emeraldcloudlab.com).*
Transcriptic Inc, Menlo Park
the robotic cloud laboratory

A robotic arm operating in one of the Transcriptic lab’s work cells. Image credit: Transcriptic

Thierry Diagana
Novartis Singapore

Images are copyright of Novartis.
https://www.novartis.com

Images are copyright of Transcriptic.
National Petascale Facility

The National Supercomputing Centre Singapore is a national petascale facility established to support high performance science and engineering computing needs for the academic, research and industry communities in Singapore.
Our facility is linked by high bandwidth multi-gigabit networks (STAR-N) to provide high speed access to users everywhere locally & globally.
Historical Trajectory

2015  | National Supercomputing Centre (NSCC)

2013  | 100TFLOPS Supercomputer at A*CRC

2000s | TFLOPS Supercomputer at IHPC and NTU

1995  | CRAY Vector Supercomputers at NSRC/NUS

1988  | NEC SX1 at NCB’s Advanced Computation Centre
# Top500 List - June 2016

**R_{\text{max}}** and **R_{\text{peak}}** values are in TFlops. For more details about other fields, check the [TOP500 description](#).

**R_{\text{peak}}** values are calculated using the advertised clock rate of the CPU. For the efficiency of the systems you should take into account the Turbo CPU clock rate where it applies.

<table>
<thead>
<tr>
<th>RANK</th>
<th>SITE</th>
<th>SYSTEM</th>
<th>CORES</th>
<th>RMAX (TFLOP/S)</th>
<th>RPEAK (TFLOP/S)</th>
<th>POWER (KW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>93</td>
<td>National Supercomputing Centre</td>
<td>NSCC - PRIMERGY CX2550 M1 LC &amp; CX2570 M1 LC, Xeon E5-2690v3 12C 2.6GHz, Infiniband EDR Fujitsu</td>
<td>30,912</td>
<td>1,010.0</td>
<td>1,285.9</td>
<td></td>
</tr>
</tbody>
</table>
Vision & Objectives

Vision of NSCC

“Democratising Access to Supercomputing”

Objectives of NSCC

1. Support National R&D Initiatives
2. Attract Industrial Research Collaborations
3. Enhance Singapore’s Research Capabilities
NSCC Data Centre @ Fusionopolis

Images are copyright of A*CRC.

Level 17
NSCC Data Centre

HPC Racks

Storage Racks
NSCC Data Centre – Cooling System

Combination of 3 cooling systems to achieve max. efficiency

Air Cooling:
Computer Room Air Handler (CRAH) units

Chilled water Cooling:
Rear door heat exchangers

Liquid Cooling:
Warm water cooling direct-to-chip

L18S Warm water dry coolers & pumps
Direct-to-Chip Cooling Technology

- Direct-to-chip hot water (40 °C / 105 °F) based Cool-Central® Liquid Cooling captures between 60-80% of the servers heat.

- Helps to reduce data centre cooling costs by over 50% and allows for 2.5-5x higher data center density.
Network Operations Centre

- To monitor the technical operations of the data centre complex
- Parameters monitored include: Relative Humidity, Temperature, CPU utilisation, power loads, etc.
- Schneider Struxureware DataCentre Expert 7.3
- Integrated DCIM-BMS
HPC Hardware

**1 PFLOP System**
- **1,288 nodes** (dual socket, 12 cores/CPU E5-2690v3)
- **128 GB DDR4 RAM/node**
- **10 Large memory nodes** (1x6TB, 4x2TB, 6x1TB)

**13PB Storage**
- HSM Tiered, 3 Tiers
- I/O 500 GB/s flash burst buffer
- 10x Infinite Memory Engines (IME)

**EDR Interconnect**
- EDR (100Gbps) Fat Tree within cluster
- InfiniBand connection to remote login nodes at stakeholder campuses

**Accelerator nodes**
- **128 nodes** with NVIDIA GPUs
- NVIDIA Tesla K40 (2,880 cores)
- 368,640 total GPU cores

**Visualization nodes**
- **2 nodes** R940 graphic workstations
- Each with 2 x NVIDIA Quadro K4200
- NVIDIA Quadro Sync support
Long list of installed software

Installed software

- Bioinformatics
  - Bcl2fastq 2.17.1.14
  - Bedtools 2.25
  - BWA 0.7.13
  - Bowtie 1.1.2
  - Bowtie2 2.2.9
  - Cufflinks 2.2.1
  - GATK
  - Hmmer 3.0
  - MPIBlast 1.6.0
  - mutect
  - NCBI Blast 2.3.0+
  - Samtools 1.3
  - Tophat 2.1.1

- Chemistry and Materials Science
  - AmberTools 15update6
  - CPMD 3.15.3 (requires individual user registration, helpdesk for access)
  - GAMESS
  - Gromacs 5.1.2
  - LAMMPS
  - NAMD
  - NWChem 6.6
  - Quantum Espresso 5.3.0

- Deep Learning
  - Caffe
  - Digits
  - Tensorflow
  - Theano
  - Torch

- Engineering
NSCC leverages on long-haul InfiniBand switches which deliver high-bandwidth and low-latency between different geographic locations.

- Between stakeholder remote login nodes/fat nodes and NSCC HPC centre.
- Also for Data Transfer Nodes overseas.
Connectivity: Remote Login Nodes / Fat Nodes

- **Remote Login nodes @ NTU & NUS** – allow students & staff to utilize NSCC HPC resources as if they were locally at NSCC due to the high-speed InfiniBand connectivity
- **Large memory node @ GIS Sequencing Farm** – MediaFlux application server, allows sequencing data to be stored into NSCC’s GPFS for immediate computation

Leveraging on Singapore Advanced Research and Education Network (SingAREN) and SingAREN-Lightwave Internet Exchange (SLIX)
Federated Identity Management (FIM) for Stakeholders

• Leveraging on SingAREN’s Singapore Access Federation (SGAF) service, NSCC enables seamless access to its resources for our stakeholders.

• SGAF uses Shibboleth-based authentication and authorisation systems to enable scalable, trusted collaborations among Singapore's R&E community.

• The authentication is performed by the user's home institution (maintains control of its users' information). NSCC grants authorised service access based on specific user attributes.

https://user.nscc.sg
Federated Identity Management (FIM)

1. Welcome to NSCC!
2. Please select the identity you want to authenticate:
3. Sign In
4. Information that will be sent to
5. Affiliation at home organizat
Building a National High-Speed InfiniBand Fabric

- **WOOLANDS**
  - Republic Polytechnic
  - Rolls Royce

- **SELETAR**
  - Rolls Royce

- **CHANGI**
  - SUTD

- **NOVENA**
  - Lee Kong Chian School of Medicine
  - Tan Tock Seng Hospital

- **JURONG**
  - National Cancer Centre Singapore
  - SingHealth Academy

- **BIOPOLIS**
  - Genome Institute of Singapore

- **OUTRAM**
  - Duke-NUS
  - Integrated Health Information System
  - Ministry of Health
  - National Cancer Centre Singapore

- **WOODLANDS**
  - Up to 500Gbps InfiniBand
  - Up to 100Gbps InfiniBand
  - 10/40/100Gbps InfiniBand/IP

- **55m**

- **OUTRAM ONE-NORTH**
  - A*STAR Fusionopolis

- **NTU**

- **NUS**

- **SUTD**

- **KEPEL**

- **ONE-NORTH**
  - A*STAR Biopolis

- **30m**

- **30km**

- **42km**
NSCC Co-Funded International Links for Inter-Connecting Supercomputers

- London, UK; Europe
- Singapore
- Los Angeles, USA

- 100Gbps co-funded with Internet 2
- 10Gbps co-funded with Internet 2
- Japan NICT/NII

GÉANT
INTERNET
TEIN
SingAREN
Regional Network Connectivity (via SingAREN)

Updates

1. Connectivity with Japan: 20Gbps (NII and NICT circuits);
2. New KOREN 10Gbps circuits;
3. Upgrade of the internet2 Hong Kong-Singapore connectivity from 1 Gbps cct to 10Gbps;
4. India-Singapore NKN cct at 10Gbps planned for September;
5. Sri Lanka - Singapore LEARN (Lanka Ed Advanced Research Network) 2.5Gbps cct;
6. UAE - Singapore Ankabut 2.5Gbps

Images are copyright of TEIN*CC and TEIN4 PROJECT.
NSCC-GIS Integration – Piloting Genomic Big Data with HPC for Precision Medicine

**NSCC**

**STEP 1:** Sequencers stream data directly to NSCC Storage *(NO footprint in GIS)*

**STEP 2:** Automated pipeline analysis once sequencing completes. Processed data resides in NSCC

**STEP 3:** Data manager indexes and annotates processed data. Replicate metadata to GIS. Allowing data to be search and retrieved from GIS

**GIS**

**NSCC Sequencers at B2 (Illumina + PacBio)**

Images are copyright of GIS A*STAR.
Global Connectivity
as Anchor Participant of InfiniCortex Project
to create a Galaxy of Computers

Objective
To create a Galaxy of Computers by enabling a number of globally-dispersed HPC facilities to collaborate and function as ONE distributed concurrent supercomputer, bringing the capability to address and solve grand challenges to the next level of efficiency and scale.

Images are copyright of A*CRC.
InfiniCortex Project

Largest spanning InfiniBand network: Ring-around the world with SG-EU@10G & SG-US@100G

InfiniBand routing across 7 subnets over 4 continents

Scalable InfiniBand connected HPC cloud instances - InfiniCloud across four countries (Australia, Singapore, USA, France)

Images are copyright of A*CRC.
Long Range InfiniBand

1 TB in 24 minutes
DATA “WORMHOLE”

Facilitating Data Intensive Collaboration Between NSCC and the National Center for Microscopy and Imaging Research (NCMIR) at UC San Diego
Mark Ellisman’s 3D Reconstructions from NCMIR X-ray Microscopic Computed Tomography Facilitates Development of Bioinspired “Tough” Materials

UCR researchers are modeling the teeth (radula) of marine snail, *Cryptochiton Stelleri*, to engineer new biomimetic abrasion resistant composites.

MicroCT reconstructions of Chiton radula. Chiton radula have evolved to incorporate an iron oxide mineral, *magnetite*, making them extremely hard and magnetic. Images courtesy of Steven Herrera, Ph.D., Kisailus Biomemetics and Nanostructured Materials Laboratory, UC Riverside.
Cell Image Library Designed For “Big Data”
Leverages High Bandwidth Connected High Performance Storage and Computing Resources

Images are courtesy of Prof Mark Ellisman.

Collaboration with Mark Ellisman UCSD/NTU

Mirror Cell Image Library Infrastructure and Data Management Workflows at NSCC
The future of Biology is Data-intensive and Data-driven

Search, Retrieve, Interrogate
Simulation and Modeling
Machine Learning, Deep Learning, AI

Building the Infrastructure for:
1. High Performance Computing
2. High Throughput Computing
3. High Performance Data Analytics
4. High Performance Data Transfer
5. High Performance Visualisation
6. Integration with Robotic Automated Labs

Images are copyright of Dassault System
http://www.slideshare.net/DassaultSystemes/enter-into-the-4th-industrial-revolution-28626426

Image copyright of Asmirov Institute and taken from Cool Infographics, Randy Krum
## From Industry 1.0 to Industry 4.0

<table>
<thead>
<tr>
<th></th>
<th>18th Century</th>
<th>Driven by water and steam power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>19th Century</td>
<td>Use of electrical energy.</td>
</tr>
<tr>
<td>2.0</td>
<td>20th Century</td>
<td>Use of electronics and IT</td>
</tr>
<tr>
<td>3.0</td>
<td>21st Century</td>
<td>Use of Cyber-physical systems</td>
</tr>
</tbody>
</table>

Join me and enter the future of Biology

**RESEARCH AT THE SPEED OF THOUGHT!**
13-17 March 2017
Singapore
Matrix Building @ Biopolis One-North
https://supercomputingfrontiers.com/2017