Argonne NATIONAL LABORATORY

DOE USA and MEXT Japan Collaboration on Extreme Computing

The Convergence of HPC and Big Data (Waggle and the Array of Things)

Pete Beckman

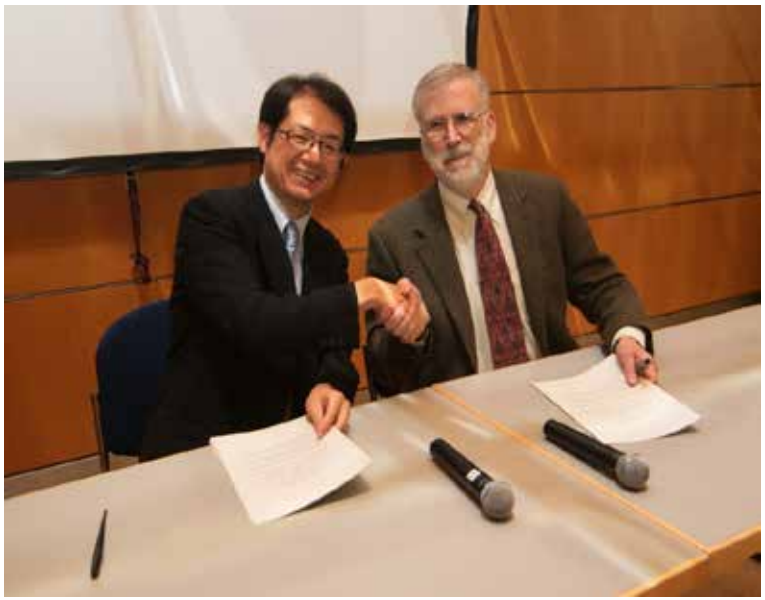
Senior Scientist, Argonne National Laboratory

Co-Director, Northwestern / Argonne Institute for Science and Engineering (NAISE)

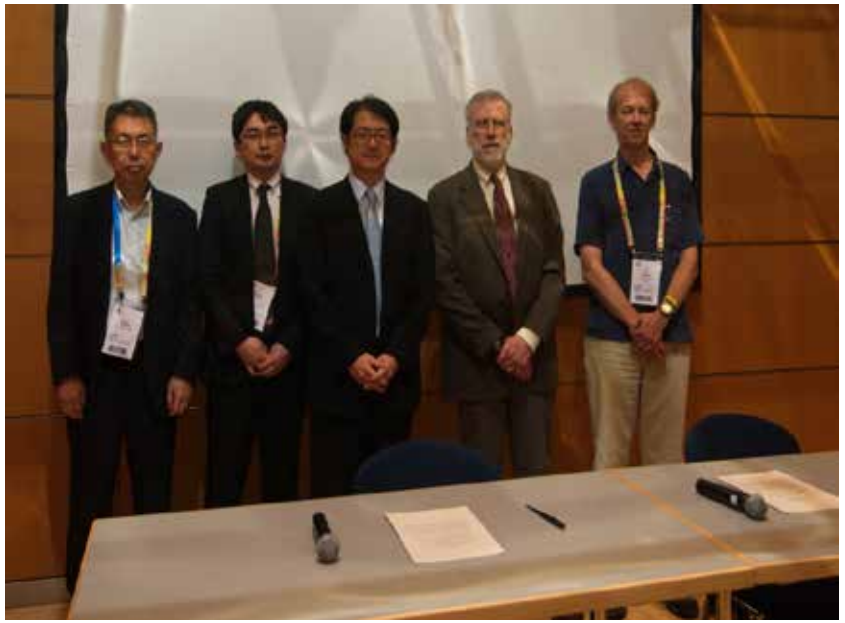
Senior Fellow, University of Chicago Computation Institute

U.S. DEPARTMENT OF ENERGY

Implementing Agreement Signed at ISC 2014



Yoshio Kawaguchi (MEXT, Japan) and William Harrod(DOE, USA)



Yoshio Kawaguchi
Director
Office for the Promotion of Computing Science,
Research Promotion Bureau
Ministry of Education, Culture, Sports, Science and Technology (MEXT)

Shinya TAHATA
Director for Information Science and Technology, Information Division,
Research Promotion Bureau
Ministry of Education, Culture, Sports, Science and Technology (MEXT)

Yutaka Ishikawa, Shinya Tahata, Yoshio Kawaguchi, William Harrod, Peter Beckman

Purpose and Strategy

Purpose: Work together where it is mutually beneficial to expand the HPC ecosystem and improve system capability

- Each country will develop their own path for next generation platforms

- Joint Activities
  - Pre-standardization interface coordination
  - Collection and publication of open data
  - Collaborative development of open source software
  - Evaluation and analysis of benchmarks and architectures
  - Standardization of mature technologies

High Performance Computing
Strong Collaboration between
Key Institutions in the United States and Japan

United States:

Argonne National Laboratory, DOE
Lawrence Livermore National Laboratory, DOE
Oakridge National Laboratory, DOE
Northwestern University
University of Arizona
University of Houston
University of Chicago
Rensselaer Polytechnic Institute

Japan:

RIKEN Advanced Institute for Computational Science
Tokyo Institute of Technology
Tokyo University of Electro-Communications
University of Tsukuba
University of Tokyo
Kyushu University

Committee on HPC Computer Science and System Software for Japan/US open scientific research

- USA DOE: William Harrod, ASCR
- MEXT Japan: Katsu Kudo
- Co-chair person
  - Pete Beckman, ANL
  - Yutaka Ishikawa, RIKEN AICS

Committee Members:

Japan

Satoshi Matsuoka
Tokyo Institute of Technology

Mitsuhisa Sato
RIKEN AICS / University of Tsukuba

Osamu Tatebe
University of Tsukuba

Hiroshi Nakamura
University of Tokyo

USA

Jeffery Vetter
Oakridge National Laboratory

Martin Schulz
Lawrence Livermore National Laboratory

Rob Ross
Argonne National Laboratory

Alok Choudhary
Northwestern University

Forms of Cooperation

- Joint Workshops / Conferences (recent)
  - International Supercomputing 2015, Frankfurt Germany
  - Sept 2015, Chicago IL, USA
  - February 2016, Japan
- Exchange of Personnel
  - scientists, engineers, post-graduates and graduate students on short-term or long-term basis
- Development and Exchange of Technical Information and Experiences, and Collaborative Visits
  - the Participants will jointly or independently develop research results in the areas of cooperation.
  - Data and experience developed by the Participants under this Project Arrangement will be exchanged.
  - The Participants will arrange collaborative visits related to research and development programs for this purpose

1st Day

9:00 – 09:30 Opening
Katsu Kudo, Peter Beckman, Yutaka Ishikawa

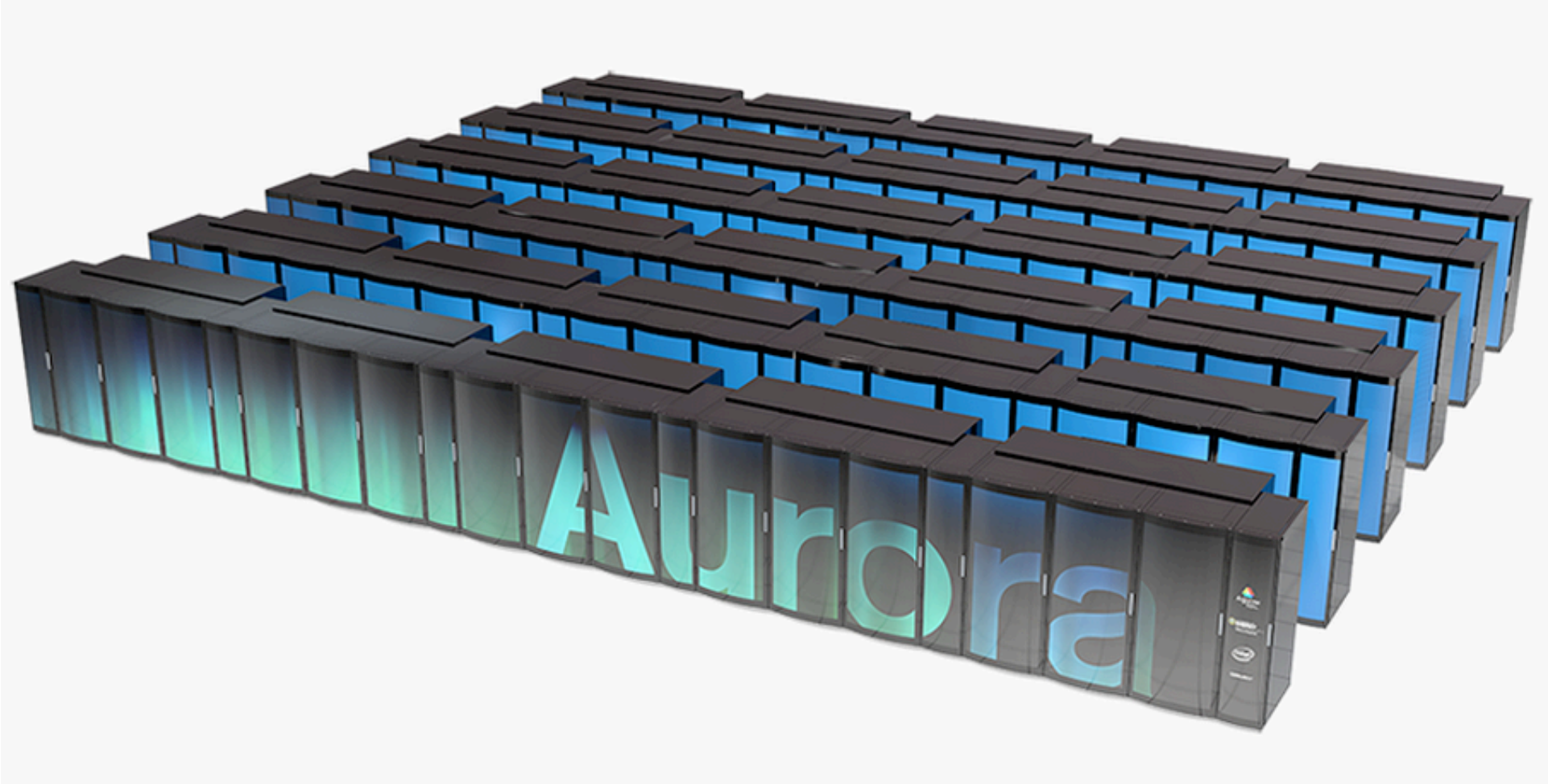
- *10–15min talk, 20–15min discussion*

09:30 – 10:00 **Codesign**, Jefery Vetter (ORNL) and Miwako Tsuji (RIKEN), Taisuke Boku (U. Tsukuba), Mitsuhisa Sato(RIKEN), Martin Schulz (LLNL)
10:00 – 10:30 **Mini-/Proxy-Apps for Exascale Codesign**, Jefery Vertter (ORNL) and Naoya Maruyama (RIKEN)
10:30 – 11:00 **Advanced PGAS runtime and thread-parallelism (Argobots)**, Mitsuhisa Sato(AICS), Pavan Balaji(ANL),Laxmikant Kale (UIUC), Barbara Chapman (U. Huston), Sangmin Seo (ANL), and Jinpil Lee (RIKEN)
11:00 – 11:30 **Advanced programming models for exascale architecture**, Mitsuhisa Sato(AICS), Jefery Vertter (ORNL) ,Taisuke Boku (U. Tsukuba) and Jinpil Lee (RIKEN)
11:30 – 12:00 **Operating System and Runtime**, Kamil Iskra (ANL) and Balazs Gerofi (RIKEN)
12:00 – 12:30 **Extreme-Scale Resilience for Billion-Way Parallelism**, Satoshi Matsuoka (TITECH), Kathryn Mohror (LLNL), and Martin Schulz (LLNL)

12:30 – 13:30 *Lunch meeting*
13:30 – 14:00 **Enhancements for Data Movement in Massively Multithreaded Environments**, Pavan Balaji (ANL) and Satoshi Matsuoka (TITECH)
14:00 – 14:30 **Scalability and performance enhancements to communication library**, Pavan Balaji (ANL) and Masamichi Takagi (RIKEN)
14:30 – 15:00 **Communication Enhancements for Irregular/Dynamic Environments Pavan**, Balaji (ANL) and Atsushi Hori (RIKEN)
15:00 – 15:15 *Break*
15:15 – 15:45 **Storage and I/O**, Rob Ross (ANL) and Osamu Tatebe (U. Tsukuba)
15:45 – 16:15 **I/O Benchmarks and netCDF implementations for Scientific Big Data**, Wei-keng Liao, Alok Choudary (North Western U.), Jianwei Liao and Yutaka Ishikawa (RIKEN)
16:15 – 16:45 **Power Monitoring, Analysis and Management**, Martin Schulz (LLNL), Barry Rountree (LLNL), Masaaki Kondo , Hiroshi Nakamura (U. tokyo), and Satoshi Matsuoka (TITECH)
16:45 – 17:00 Summary


Argonne National Laboratory

Argonne's Next Big Machine: Aurora



Europe-USA-Asia Workshops on Big Data and Extreme Computing

Following the International Exascale Software Initiative
(IESP 2008-2012 ➔**Big Data and Extreme Computing** workshops (BDEC)
http://www.exascale.org/bdec/

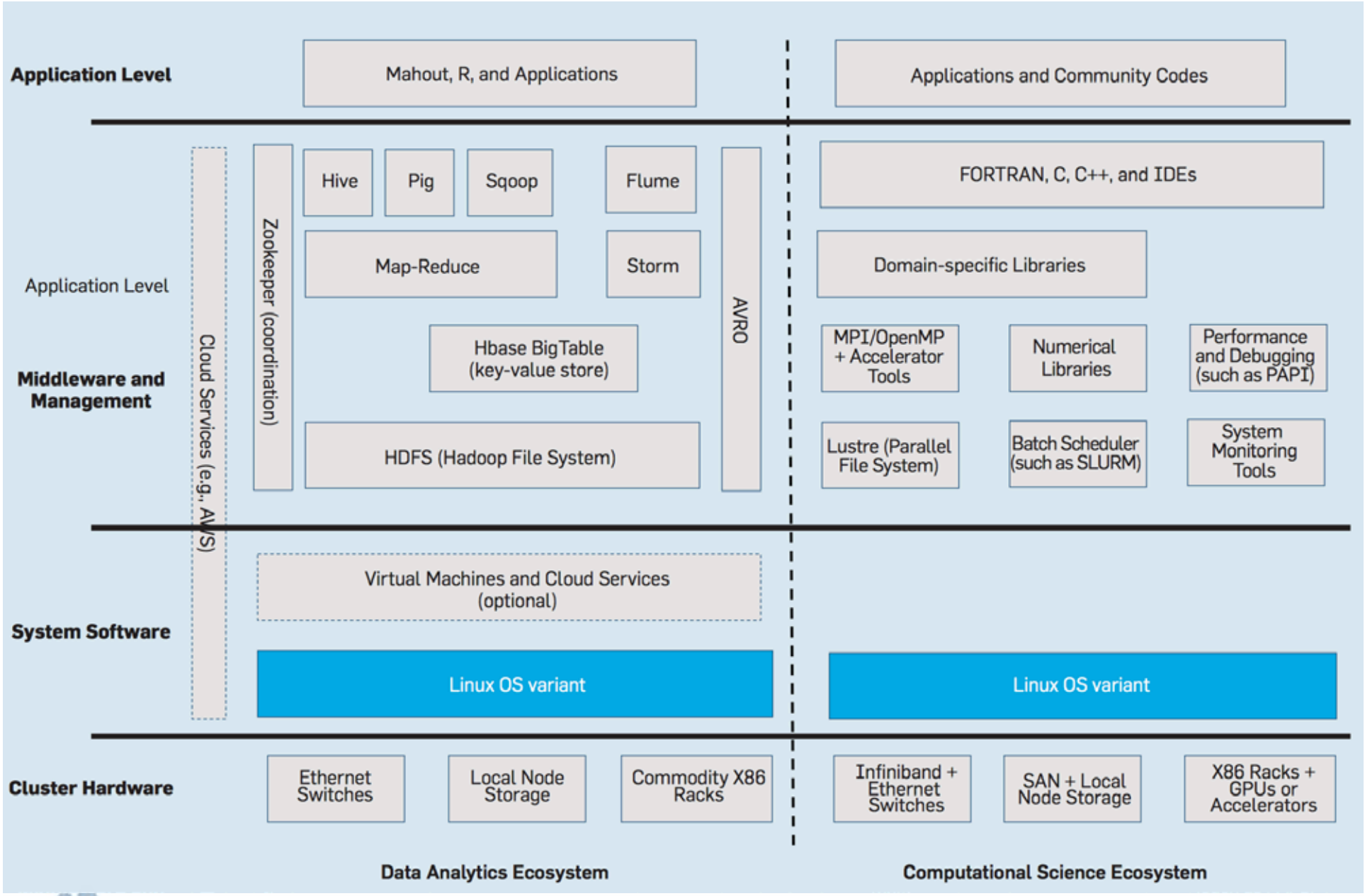
Overarching goal:

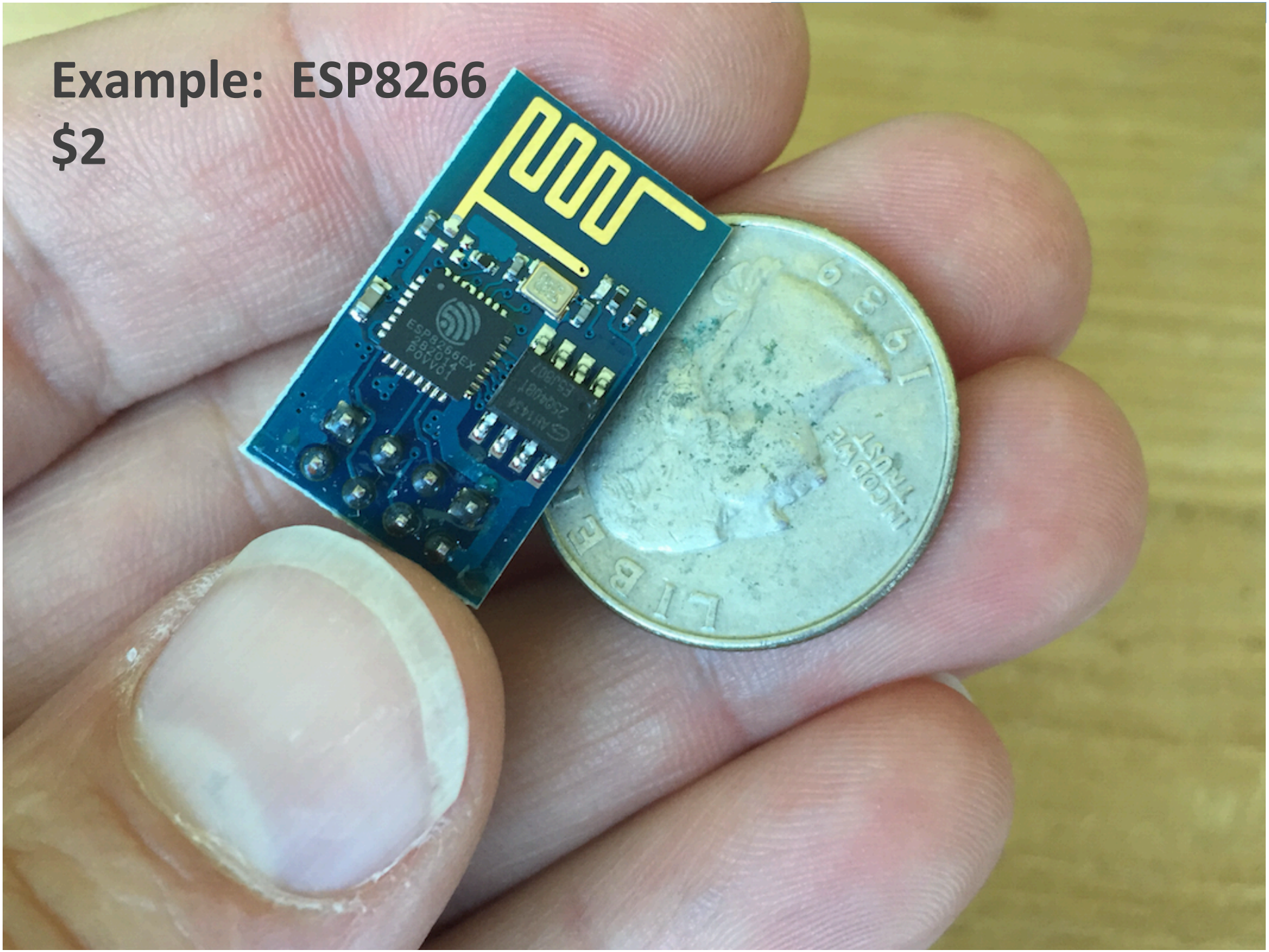
1. Create an international collaborative process focused on the co-design of software infrastructure for extreme scale science, addressing the challenges of both extreme scale computing and big data, and supporting a broad spectrum of major research domains,
2. Describe funding structures and strategies of public bodies with Exascale R&D goals worldwide
3. Establishing and maintaining a global network of expertise and funding bodies in the area of Exascale computing

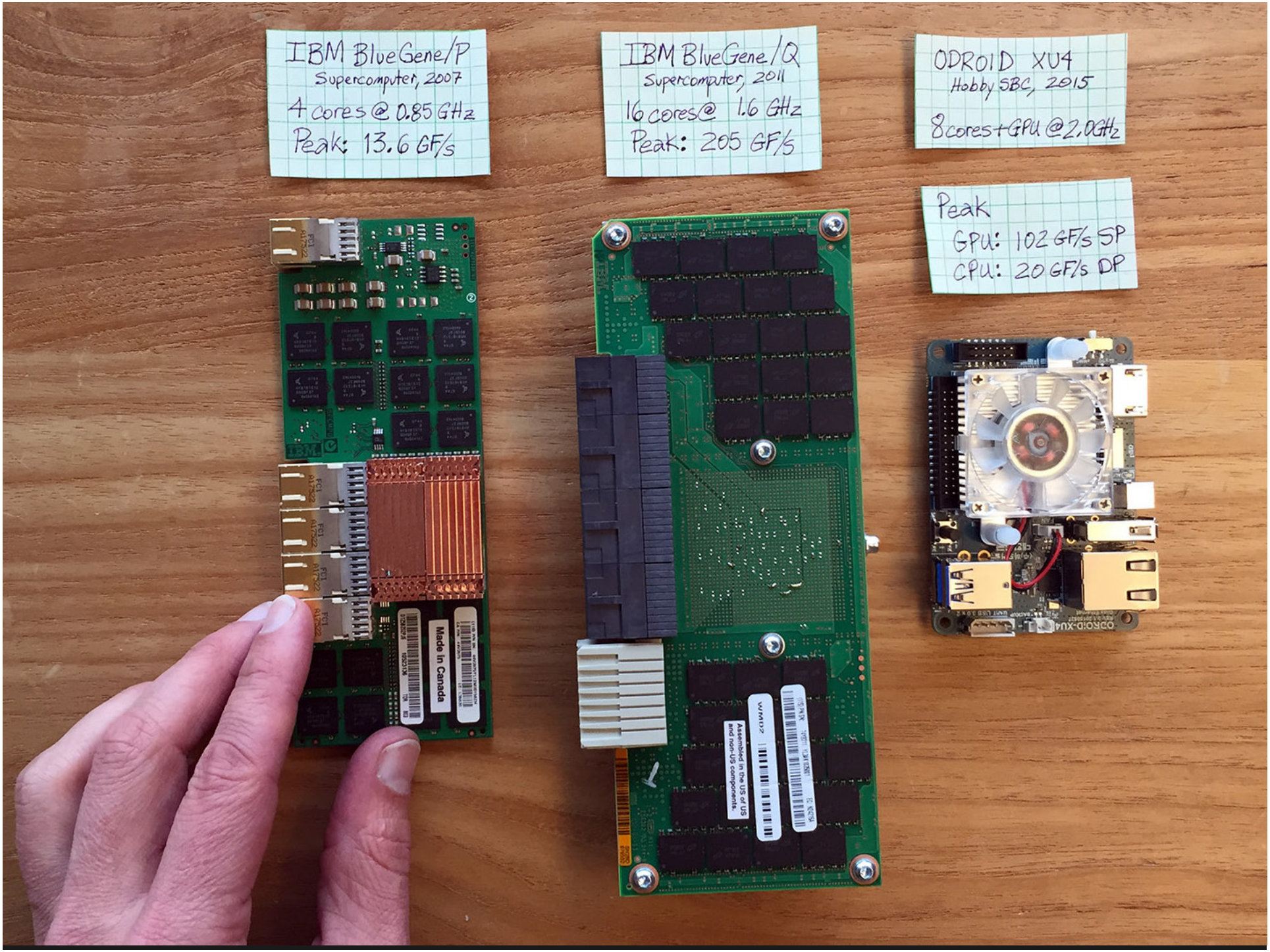
1 – BDEC Workshop, Charleston, SC, USA, April 29-May1, 2013

2 – BDEC Workshop, Fukuoka, Japan, February 26-28, 2014

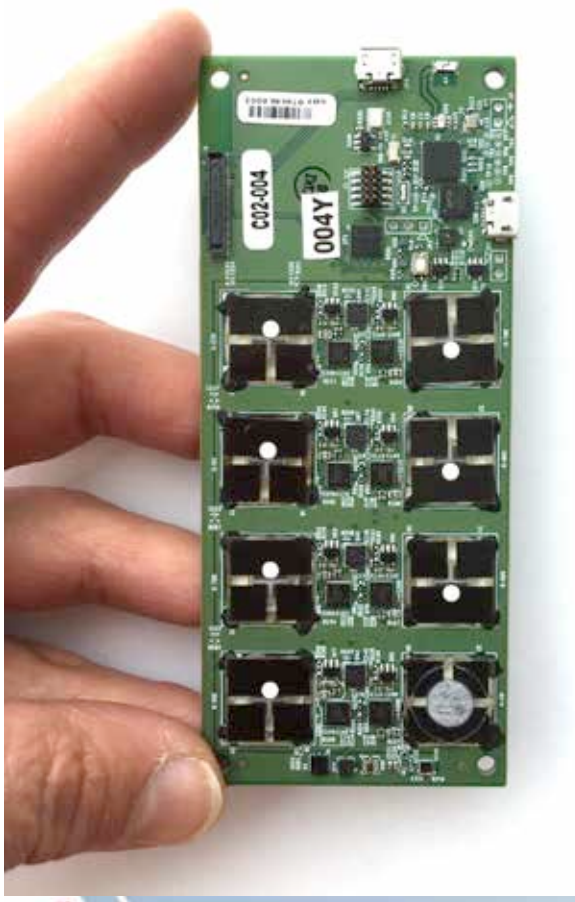
3 – BDEC Workshop, Barcelona, Spain, January 28-30, 2015


Application Level
Mahout, R, and Applications
Applications and Community Codes
Application Level
Middleware and Management
System Software
Cluster Hardware
Cloud Services (e.g., AWS)
Zookeeper (coordination)
Hive
Pig
Sqoop
Flume
Map-Reduce
Storm
AVRO
Hbase BigTable (key-value store)
HDFS (Hadoop File System)
FORTRAN, C, C++, and IDEs
Domain-specific Libraries
MPI/OpenMP + Accelerator Tools
Numerical Libraries
Performance and Debugging (such as PAPI)
Lustre (Parallel File System)
Batch Scheduler (such as SLURM)
System Monitoring Tools
Virtual Machines and Cloud Services (optional)
Linux OS variant
Linux OS variant
Ethernet Switches
Local Node Storage
Commodity X86 Racks
Infiniband + Ethernet Switches
SAN + Local Node Storage
X86 Racks + GPUs or Accelerators
Data Analytics Ecosystem
Computational Science Ecosystem


Example: ESP8266
$2


IBM BlueGene/P
Supercomputer, 2007
4 cores @ 0.85 GHz
Peak: 13.6 GF/s
IBM BlueGene/Q
Supercomputer, 2011
16 cores @ 1.6 GHz
Peak: 205 GF/s
ODROID XU4
Hobby SBC, 2015
8 cores + GPU @ 2.0 GHz
Peak
GPU: 102 GF/s SP
CPU: 20 GF/s DP
Made in Canada
Assembled in the US of US and non-US components.

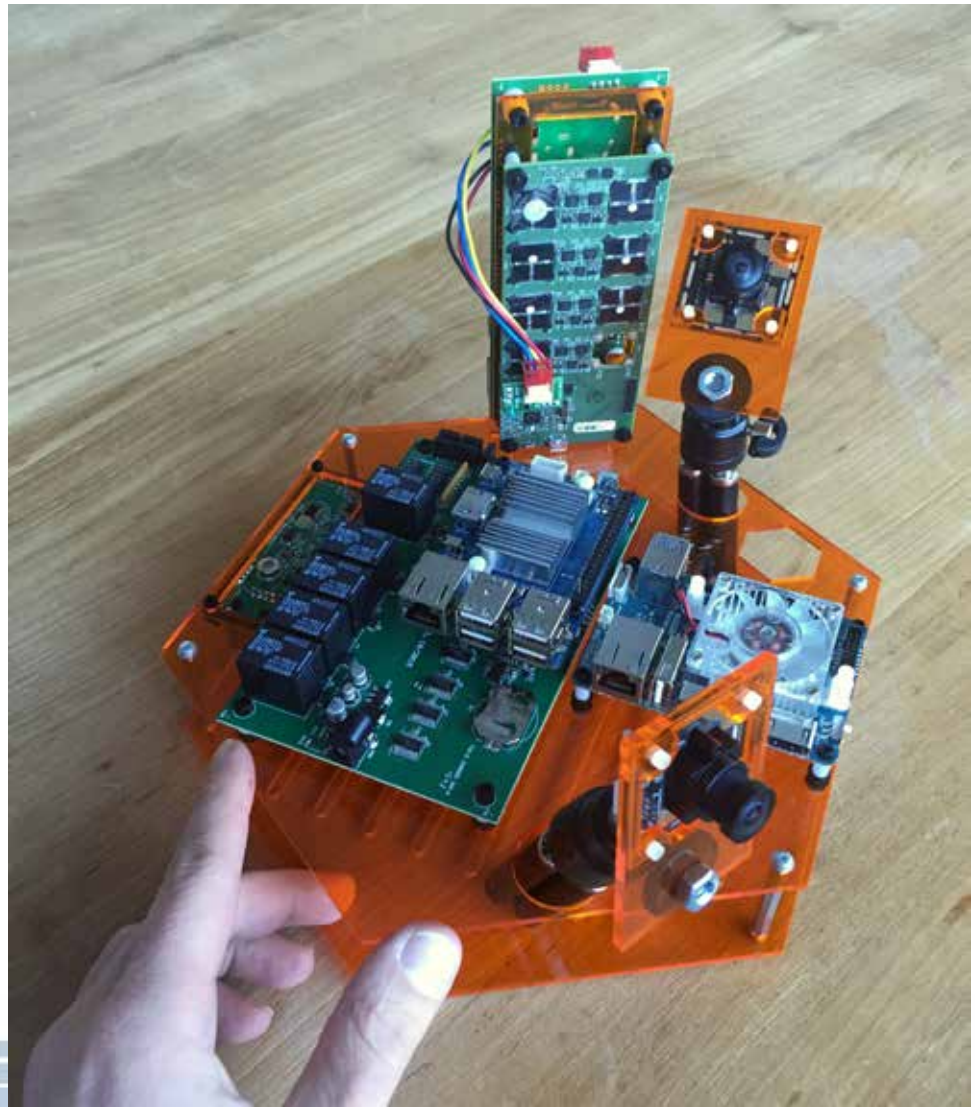
New Advanced Sensors
(via a partnership with Intel & SPEC)



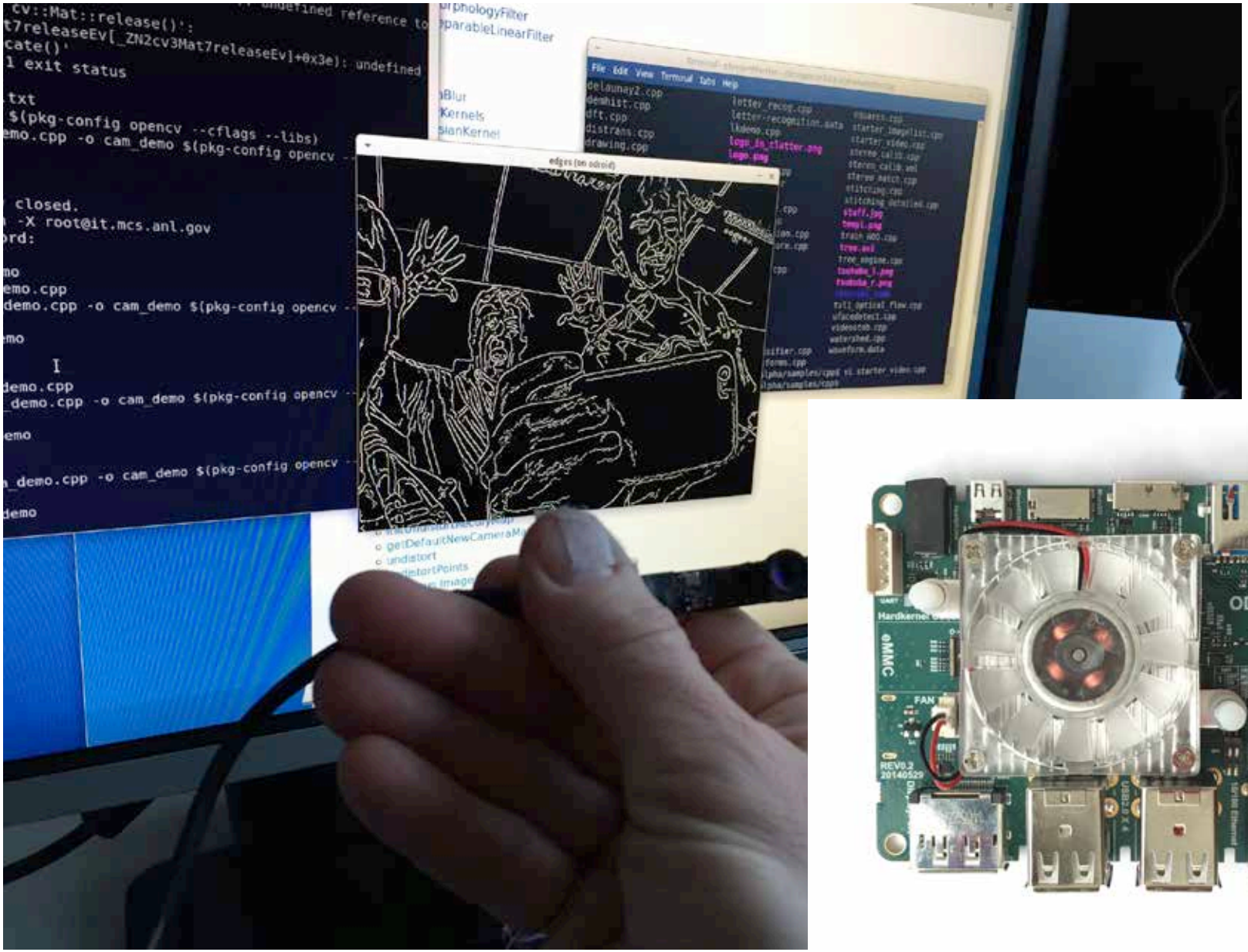
- NO2 (Nitrogen Dioxide): <2 ppb
- O3 (Ozone) < 5 ppb
- CO (Carbon Monoxide) < 1 ppm
- SO2 (Sulfer Dioxide) < 15 ppb
- H2S (Hydrogen Sulfide) < 2 ppb
- TOX (total oxidizing index) < 1 ppm CO equiv
- TOR (total reducing index) < 2 ppb NO2 equiv
- Future:
  - HCHO (Formaldehyde)
  - VOC (Volatile Organic Compound)
  - CH4 (Methane)

An open software and hardware, hackable, wireless sensor platform with in-situ computing.

http://www.wa8.gl



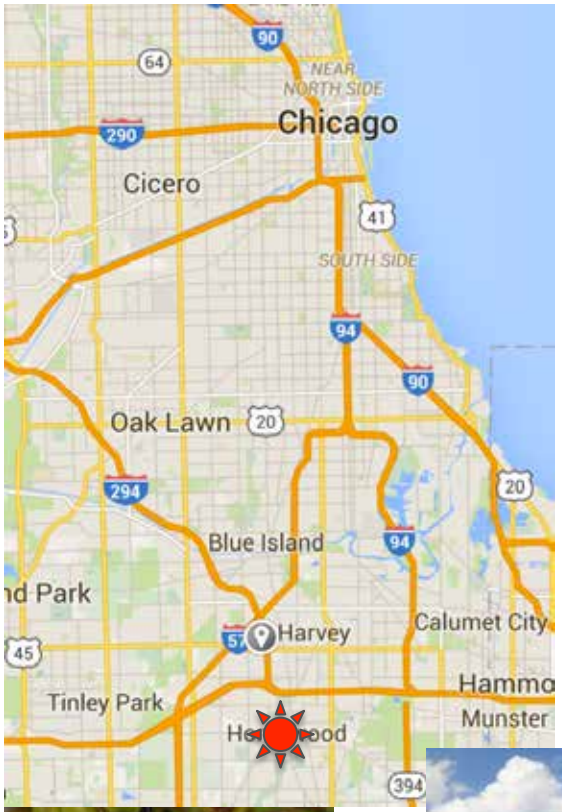
In-Situ Analysis and Feature Recognition

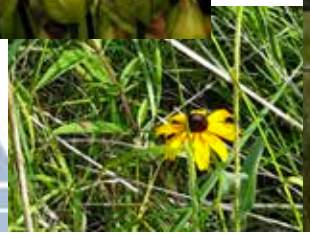


al Laboratory / Northwestern University

Gensburg-Markham Prairie

370 acres, owned/managed by Nature Conservancy and Northeastern IL Univ.
Registered as National Natural Landmark





A Science-Driven Instrument: The Array of Things

Climate, Environmental and Life Sciences
(Robert Jacob, ANL)

Potosnak (DePaul); Niyogi (Purdue); Gilbert, Graham, Kotamarthi, (UC/ANL); Fernando (Notre Dame)

Urban Infrastructure Systems
(Danie Work, UIUC)

Markoupolou (IaaC); Negri, Snyder (UC/ANL); Buttlar, Peschel, Garcia (UIUC), Gonzales (MIT), Pancoast (SAIC), Guzowski, Rosner (UC/ANL), Claudel (UT); Liu (UMich), Chen (UW)

Education, Health, Social and Behavioral Sciences
(Kathleen Cagney, UChicago)

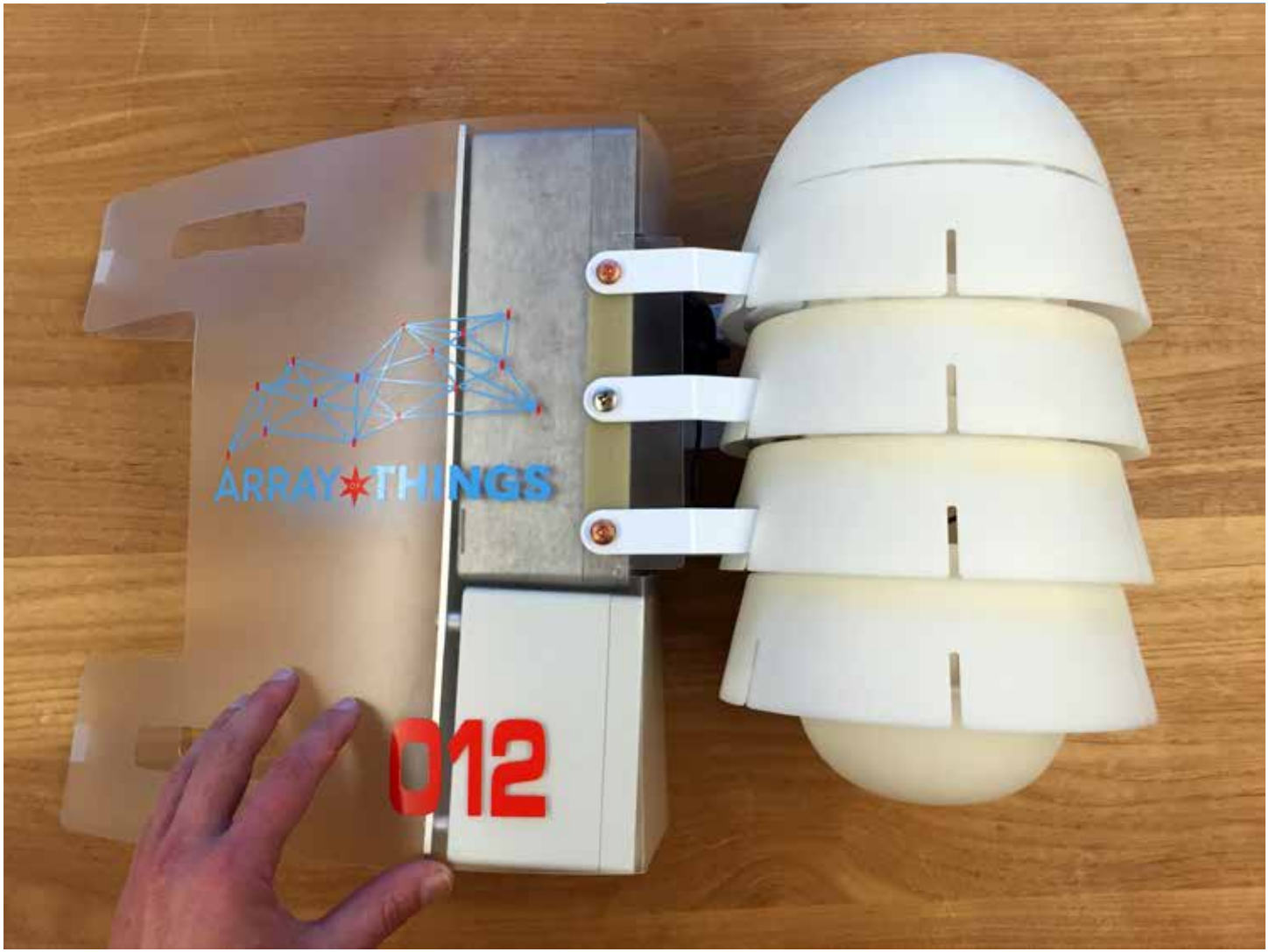
Diez (UCL/IaaC); Contractor (Northwestern); Epley, Gilliam, Lindau, Meltzer, Hampton-Marcel, Zarraonaindia (UC); Bellingham (Strathclyde)

Computer Science and Cyber-Physical Systems
(Michael Papka, UC/NIU/ANL)

Derrible, Lin, Eriksson (UIC); Alok Choudhary (NU); Beckman, Sankaran, Chien (UC/ANL)

ARRAY THINGS


GERALD



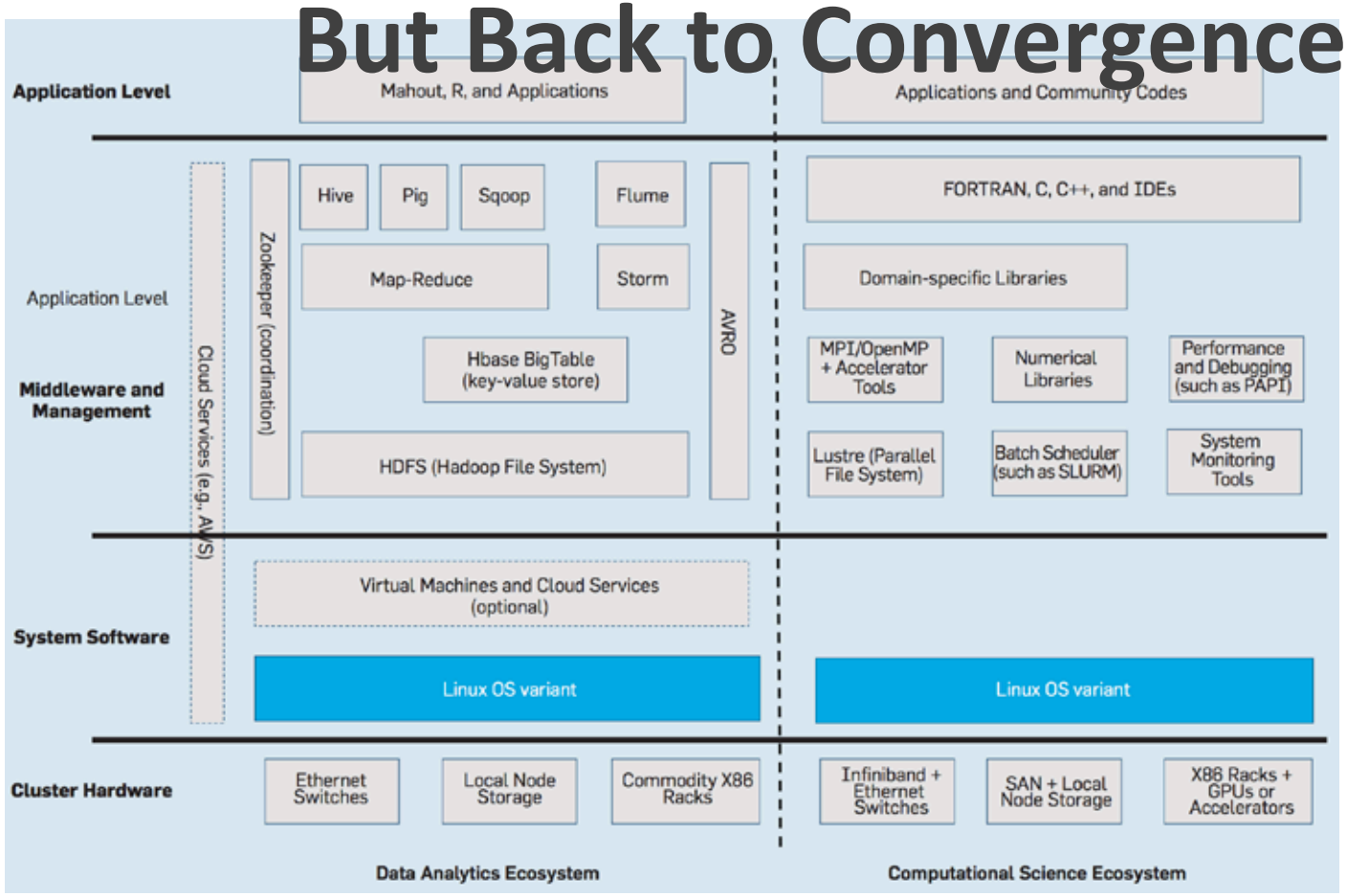
Array of Things

Beta1 Prototype

Why HPC Geeks Should Care

- New sensors are ***programmable*** **parallel computers**
  - Multicore + GPUs & OpenCL or OpenMP
  - New algorithms for in-situ data analysis, feature detection, compression
  - Need new progmod for "stackable" in-situ analysis (for sensors and HPC)
  - Need advanced OS/R resilience, cybersecurity, networking, over-the-air programming
- 1000s of nodes make a **distributed computing "instrument"**
  - New streaming programming model needed
  - New techniques for machine learning for scientific data required
    - Both for within a "node" and collectively across time series
- How will **HPC streaming analytics and simulation** be connected to live data?
  - Can we trigger HPC simulations after first approximations? (weather, energy, transportation)
  - Unstructured database with provenance and metadata for QA/collaboration
- Use novel HPC hardware to solve power issue?
  - Can we use neuromorphic or FPGAs to reduce power for in-situ analysis & compression?
- We are trading precision & cost for greater spacial resoluton: What is possible?

But Back to Convergence



Smart Sensors / BDEC Needs:

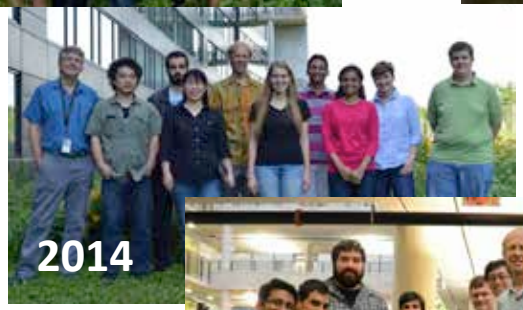
- First-level in-situ analysis & cache
- Streaming Data to Cloud
- Event-driven HPC Simulations
- Multiple Data Source Integration
- Large complex data queries

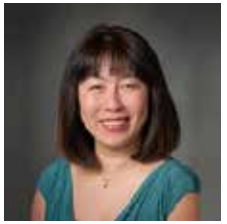
Some Thoughts on How…

- Allow Sharing
  - Increase interactivity, allow extremely long-running jobs
- Data and Compute Must Be Co-Located (Scheduled)
  - We are getting NVRAM everywhere. We must build new resource managers that can handle persistent data left behind and favored job positioning
- Support Containers and VMs
  - Not that hard, but we have to add some hardware support for security….
- Get a Real Workflow System

Finally: Load up ***Cool Big Data software stacks….***

Waggle Team & Collaborators





Questions?