Designing Next-Generation Intelligent CyberInfrastructure: An Overview of the NSF-AI ICICLE Institute

NIRTD MAGIC Seminar Series (June ‘22)

by

Dhabaleswar K. (DK) Panda
The Ohio State University
E-mail: panda@cse.ohio-state.edu
http://www.cse.ohio-state.edu/~panda
High-End Computing (HEC) has been evolving over the last three decades with multiple stages.
Stage 1 (1975 - ): Scientific Computing with Supercomputing/High-Performance Computing (HPC)
Stage 2 (2000 - ): HPC + Big Data Analytics

• Big Data changed the way people understand and harness the power of data, both in the business and research domains

• Big Data and High-Performance Computing (HPC) started converging to meet large scale data processing challenges

• Running High Performance Data Analysis (HPDA) workloads in the cloud has been gaining popularity
  • According to the latest OpenStack survey, 27% of cloud deployments are running HPDA workloads

• Has evolved into Data Science
Stage 3 (2010 - ): HPC + AI (Machine Learning/Deep Learning)

- Machine Learning (ML)
  - “the study of computer algorithms to improve automatically through experience and use of data”
- Deep Learning (DL) – a subset of ML
  - Uses Deep Neural Networks (DNNs)
  - Perhaps, the most revolutionary subset!
- Based on learning data representation
- DNN Examples: Convolutional Neural Networks, Recurrent Neural Networks, Hybrid Networks
- AI-Enabled Science, Arts, Health, and Business

Stage 4 (2015 - ): Emergence of the Computing Continuum

HPC Systems & Data Centers

1. Scientific Computing
2. Big Data & Data Science
3. Artificial Intelligence

Computing Continuum

Network Based Computing Laboratory
MAGIC (June ‘22)
AI-Driven Precision Agriculture

ARTIFICIAL INTELLIGENCE IN AGRICULTURE

Digital tools for:
- Visualization
- Crop growth analysis
- Monitoring spatial field variability
- In-season management tips
- Yield prediction

Ground data: Soil and plant analysis

Satellite imagery

Artificial Intelligence (AI)

Unmanned Aerial System (UAS)

https://ccag.tamu.edu/research-project/digital-agriculture/
Increasing Usage of HPC, AI, and Data Science in multiple Disciplines with Distributed Data and Heterogeneous Computing

Convergence of HPC, Deep/Machine Learning, and Data Science!

HPC (MPI, PGAS, etc.)

Deep/Machine Learning (TensorFlow, PyTorch, cuML, etc.)

Big Data (Hadoop, Spark), Data Science (Dask)

Increasing Need to Run these applications on the Cloud!!
Broad Challenge:

How to design next-generation intelligent cyberinfrastructure with plug-and-play capabilities to handle societal problems while taking advantage of heterogeneous (CPU + GPU) High Performance Computing (HPC) and Cloud resources?
Follow us @icicleai

http://icicle.ai

NSF-Funded AI Institute
($20M USD for Five Years)

Intelligent Cyberinfrastructure
With Computational Learning in the Environment (ICICLE)
The Vision

A national infrastructure that enables AI at the flick of a switch, ICICLE will:

• Democratize AI through integrated plug-and-play AI.
• Catalyze foundational AI/CI and transform application domains.
• Transparent and trustworthy infrastructure for AI-enabled future,
• Address societal problems (conservation, food insecurity) and national priorities
• Grow new generations of workforce and Incubate sustainable and inclusive communities
Participation (14 Organizations, 46 Investigators, and many Collaborators)
Objectives: Intelligent CyberInfrastructure for Computing Continuum

Integrating a broad range of:
- Scientists-in-the-field
- Engineers
- Educators
- Collaborative partners
- Institutions under one roof enables democratized, adaptable, plug-and-play AI and long-tail science.

**ICICLE: Intelligent CyberInfrastructure with Computational Learning in the Environment**

- Systems AI Foundational Research for CI
- Intelligent Cyber Infrastructure
  - CI for AI
  - AI for “CI for AI”

**Use Inspired Science Domains**
- Smart Foodsheds
- Animal Ecology
- Digital Agriculture

**Emerging Computing Continuum**
- On Field Sensors
- Edge & Near Edge
- Clouds
- HPC Systems & Data Centers

Network Based Computing Laboratory
MAGIC (June ‘22)
Research Plan: Overall Vision

Heterogeneities

- Interface Heterogeneity
- Model Heterogeneity
- Data Heterogeneity
- Context Heterogeneity
- Compute Heterogeneity

Abstraction Layers

- Conversational AI
- Model Commons
- Knowledge Graph
- Adaptive AI + Federated Learning
- Intelligent CI
The DNA: Foundational Systems AI

Knowledge Graphs
- Multimodal
- Spatio-temporal
- Auto construction
- Knowledge-based reasoning and pre-training

Model Commons
- KG-supported
- Precise profiling
- Flex Composition
- Versioning and provenance

Adaptive AI
- Context-aware
- Interactive
- Continual learning
- Distillation-based compression

Federated Learning
- Heterogeneity
- Applicability to a variety of models
- Context-aware
- Privacy-preserving and robustness

Conversational AI
- KG- and model-commons-aware
- Bootstrapping and adaptivity
- Multimodal contextual response

Crop Yield Model
Midwest Corn, Strawberry

Network Based Computing Laboratory
MAGIC (June ‘22)
The Enabler@Edge: CI4AI

High Perf. Training
- Deep Learning Communication Optimization
- Deep Learning I/O Improvement
- Multi-level data/model/spatial parallelism

High Perf. Data Management
- Unified storage of data, model and hyperparameters
- Model lifecycle management for AI orchestration
- Data location transparency with migration

Edge Intelligence
- Adaptive Training/Inference and FL on Edge
- Novel Edge Offloading/Caching Orchestration
- Intelligent Anomaly Detection to improve QoS

AI-Adaptive Edge Wireless
- AI-adaptive, Predictable Comm. Capacity Allocation
- Predictable Wireless Comm. via Rateless-Coding & Multi-Modal/Path

Control and Coordination
- App/CI Interface Design
- Tapis Integration
- Production-ready Service Hardening and Optimization

Network Based Computing Laboratory
MAGIC (June ‘22)
The Enabler@Scale-and-@Edge: AI4CI

Intelligent Primitives (sparse/dense) Product: High Performance Library (Powered by AI to maximize the utilization of CI)

Innovation: Learn from hardware (e.g., SIMD width), network (e.g., bandwidth/latency), and data sparsity to extract best attainable performance

Utility: (1) Portable high performance on diverse HPC systems (2) Usable at the backend of any other AI system

Heterogeneous and rapidly evolving platform (CPUs, GPUs, Supercomputers, edge devices)

APPLICATIONS
- PERFORMANCE MODELING / PREDICTION

MIDDLEWARE
- HYPERPARAMETER OPTIMIZATION

INTELLIGENT COMPUTE PRIMITIVES

SYSTEMS
- INTELLIGENT SCHEDULING
- INTELLIGENT WIRELESS COMMUNICATION

INTELLIGENT MODELING AND OPTIMIZATIONS
Research Plan: Use-Inspired Science
Crop Care: Nutrient and Pest Management

Demonstrate swarms of small unmanned aerial systems to study crop stressors, such as insect infestations, nutrient stress, disease, etc. and produce crop and soil maps that improve agricultural productivity

- Edge computing setup to speed up AI model training (NSF Funded Tapis@Texas Adv. Comp. Center)
- SoftwarePilot, open-source software developed for ICICLE at OSU, will be used to pilot sUAS
- Produce crop and soil maps that improve agricultural productive
- Demonstrate use-cases on multiple crops (e.g., corn, soybean, wheat) via novel neural network models
- Feedback control for automation of agricultural field machinery (e.g., targeted pesticide application etc.).
- Extraction of actionable information from machine and agronomic data to support cropping decisions.
Crop Care: Water Management and Quality

Capture nutrient stress and moisture deficit in corn and soybean field and provide feedback control for automation of agricultural field machinery, e.g., controlled drainage, robotic irrigation, etc.

- Extraction of actionable information from machine and agronomic data to support water management
- Adaptive AI deployed on edge, IoT and in-field devices will boost efficacy
- Target irrigation scheduling and water table level
- Discharge water quality via controlled drainage
A Deliverable: The ICICLE Software Architecture

END USER APPLICATIONS
- CONVERSATIONAL AI
- DATA VISUALIZATION
- INTERACTIVE NOTEBOOKS, CLI, SDK, WEB APP

WORKER AGENTS
- DATA TRANSFER
- RESOURCE PROVISIONING
- JOB SCHEDULING
- PERSISTENCE
  - SQL
  - NOSQL
  - MESSAGE BROKER

SYSTEMS & FILES
- SYSTEMS & FILES
- STREAMS
- META
- APPS & FUNCTIONS
- JOBS

META
- KNOWLEDGE
- CONVERSATIONAL AI
- CONVERSATIONAL AI
- MODEL COMMONS
- HISTORY & PROVENANCE
- AUTHN & AUTHZ

HTTP FRONT END APIs
- FILE SYSTEMS
- AI DATABASES
- GIT REPOSITORIES
- CONTAINER REGISTRIES
- EXECUTION HOSTS

RESEARCHERS & USERS IN THE FIELD

HPC & CLOUD DATA CENTERS

EDGE & NEAR EDGE

ICICLE GATEWAY

FIELD SENSORS

NETWORK BASED COMPUTING LABORATORY

MAGIC (June '22)
ICICLE As A Whole

Use-Inspired Science
(Smart Foodsheds,
Animal Ecology, Digital Agriculture)

Education and
Outreach

CI for
Plug-and-
Play AI

Intelligent
CI

Field’s
Edge to
HPC/Cloud

BPC/WFD
for CI
driven AI

Collaboration and
Knowledge Transfer

Network Based Computing Laboratory
MAGIC (June ‘22)
ICICLE Enables Global AI leadership

- Integrate into the National CI Ecosystem
- Integrative and Interoperable
- Leverages existing recognized capabilities
  - Centers of Excellence, AI Institutes, Large Facilities
- Collaborative
- Sustainable
  - Workforce Development, Broadening Participation, Collaboration and Knowledge Transfer
  - Benefits other institutes, large facilities, and all sciences beyond lifetime of award
Engaging With ICICLE

---

Co-develop & Adopt ICICLE developed CI!

Contact: panda@cse.ohio-state.edu
Conclusions

• AI solutions can help to solve many societal problems
• Increasing use of HPC, AI and Data Science with heterogenous resources
• Need for plug-and-play-based AI solutions which can democratize AI
• The new ICICLE NSF-AI Institute aims to establish next-generation cyberinfrastructure to provide comprehensive AI-driven solutions to many societal problems
Thank You!

panda@cse.ohio-state.edu

Network-Based Computing Laboratory
http://nowlab.cse.ohio-state.edu/

The High-Performance MPI/PGAS Project
http://mvapich.cse.ohio-state.edu/

The High-Performance Big Data Project
http://hibd.cse.ohio-state.edu/

The High-Performance Deep Learning Project
http://hdl.cse.ohio-state.edu/
"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov