Understanding Cities through Measurement and *Embedded Intelligence*:

The Array of Things

Charlie Catlett
Senior Computer Scientist
University of Chicago and
Argonne National Laboratory

Pete Beckman
Senior Computer Scientist
Northwestern University and
Argonne National Laboratory

Collaborators:
Charlie Catlett (ANL/UC)
Pete Beckman (ANL/NU)
Rajesh Sankaran (ANL)
Mike Papka (ANL/NIU)
Kathleen Cagney (UC)
Mark Potosnak (DePaul)
Doug Pancoast (SAIC)
Kate Kusiak Galvin (UC)
Dan Work (Vanderbilt)

catlett @ anl . gov

Outline

- Machine Learning and Urban Science
- The Array of Things Project
- The Waggle Platform
- Survey of Machine Learning Opportunities

This talk is a combination of several of mine and Pete Beckman's, covering work by dozens of collaborators.

Summer 2018 Student Team

2018 Array of Things User Workshop

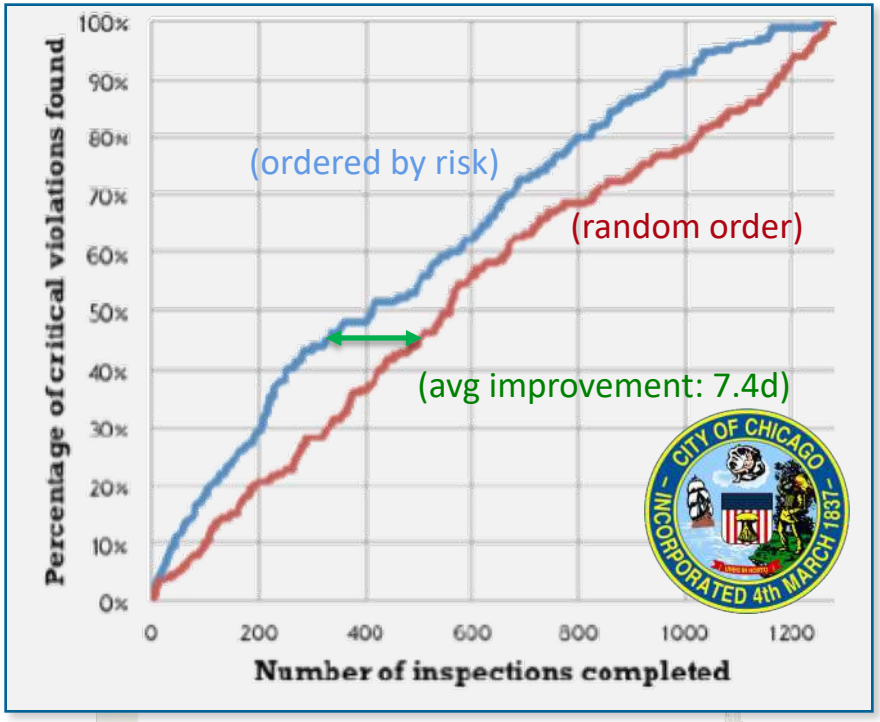
Machine Learning and Urban Science

The Array of Things and Waggle projects were developed in the context of scientific questions about cities and other environments.

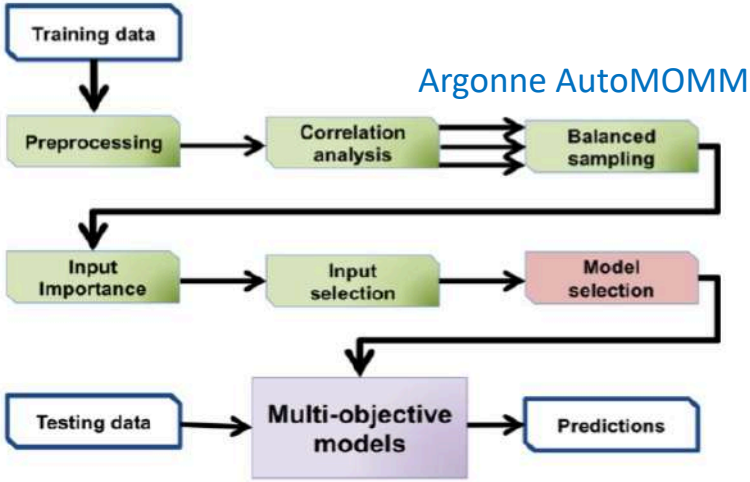
MacArthur Foundation

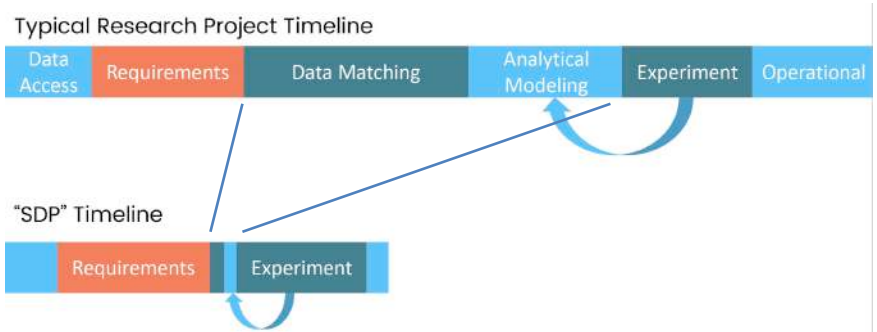
THE UNIVERSITY OF CHICAGO

Improving Food Safety



Argonne AutoMOMML





P. Balaprakash, A. Tiwari, S. M. Wild, L. Carrington, and P. D. Hovland, "AutoMOMML: Automatic Multi-objective Modeling with Machine Learning," in International Conference on High Performance Computing, 2016, pp. 219-239: Springer.

Tom Schenk (City of Chicago); Allstate Insurance; Sven Leyffer, Stefan Wild, Prasanna Balaprakash (Argonne MCS)

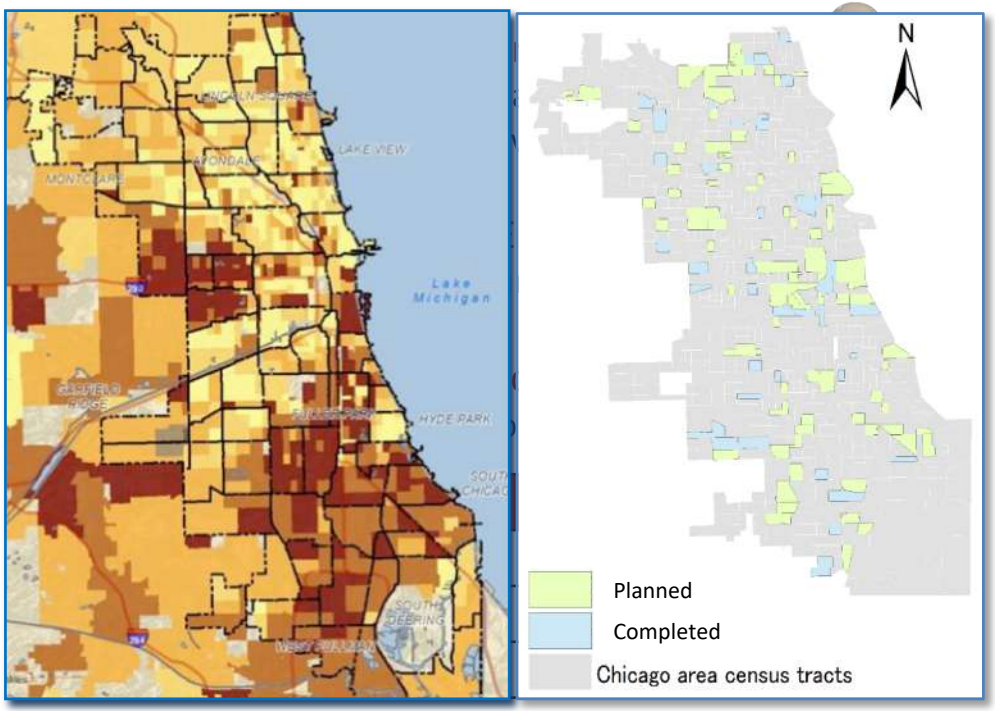
Environmental Health (Risk Prediction)

COMPASS Project

- What factors drive significant differences in life expectancy between neighborhoods?
- Cohort of >5,000 individuals from across Chicago

COMPASS Data

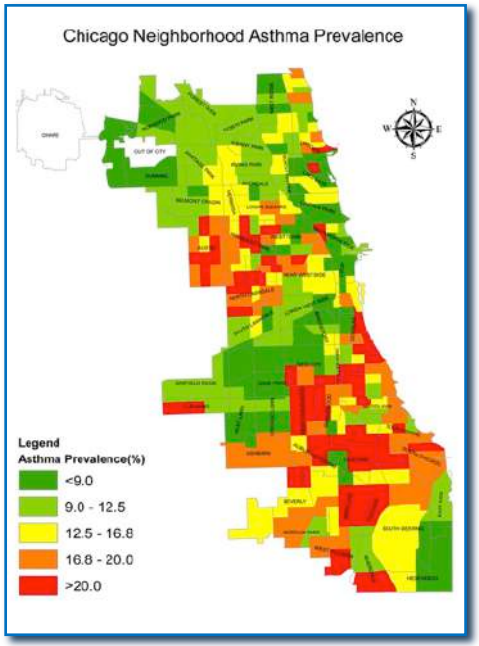
- Standard medical (BP, BMI, immune status, medication, individual and family history)
- Lifestyle (smoking, alcohol, physical activity)
- Memory (recall tests)
- Psychological (depression, hostility, stress)
- Socioeconomic (education, income, occupation)
- Social Environment (sleep, stress, work hours)



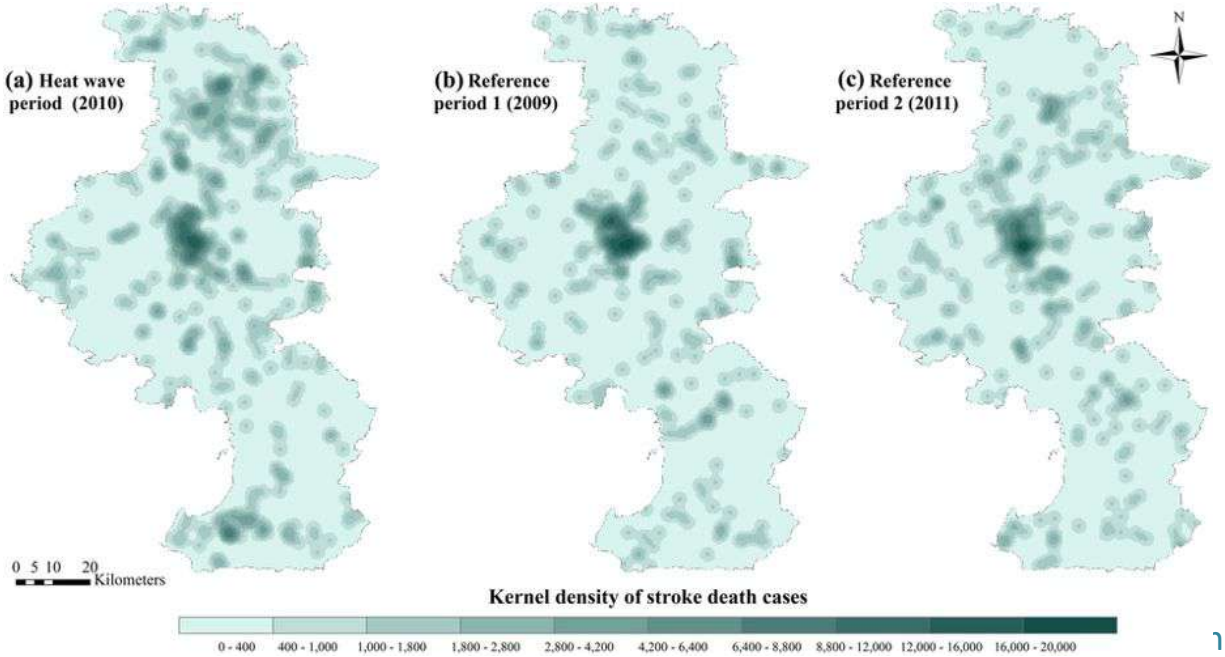
H. Ahsan, S. Olopade, B. Aschebrook-Kilfoy, S. Wild, P. Balaprakash, C. Catlett

Health

Spatial analysis of the effect of the 2010 heat wave on stroke mortality in Nanjing, China



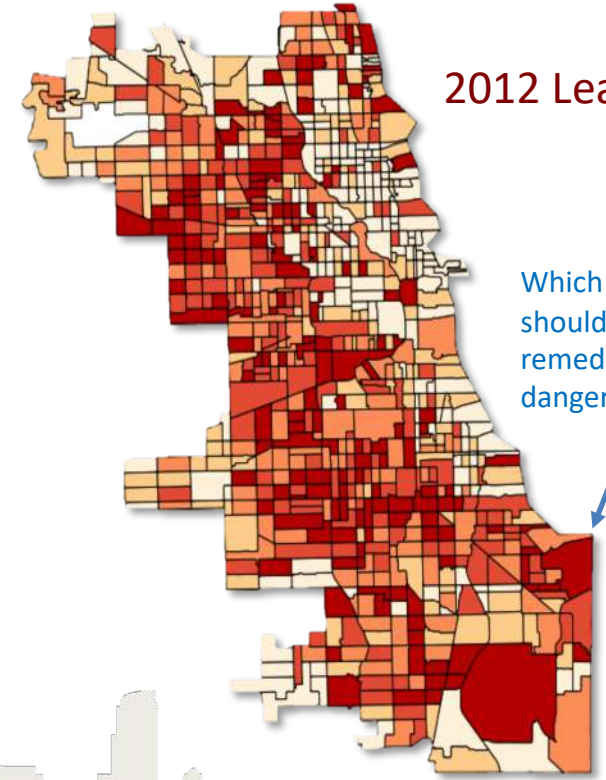
http://www.mentalmunition.com/2011/09/south-side-children-have-greatest.html



significantly increased or decreased adjusted odds ratio (p-value < 0.05).

Spatial analysis of the effect of the 2010 heat wave on stroke mortality in Nanjing, China
Kai Chen, Lei Huang, Lian Zhou, Zongwei Ma, Jun Bi & Tiantian Li, **Nature**,
Scientific Reports volume 5, Article number: 10816 (2015)

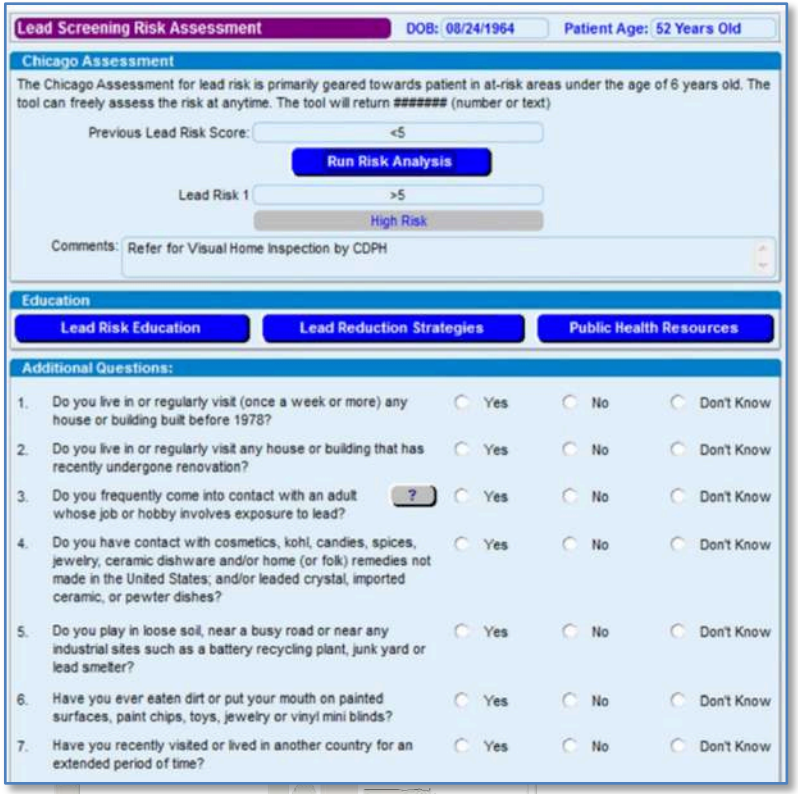
Healthcare



2012 Lead poisoning cases.

Which of the thousands of homes should the city inspect (and remediate) within any given high-danger area?

API to integrate with electronic medical record clinical decision tools



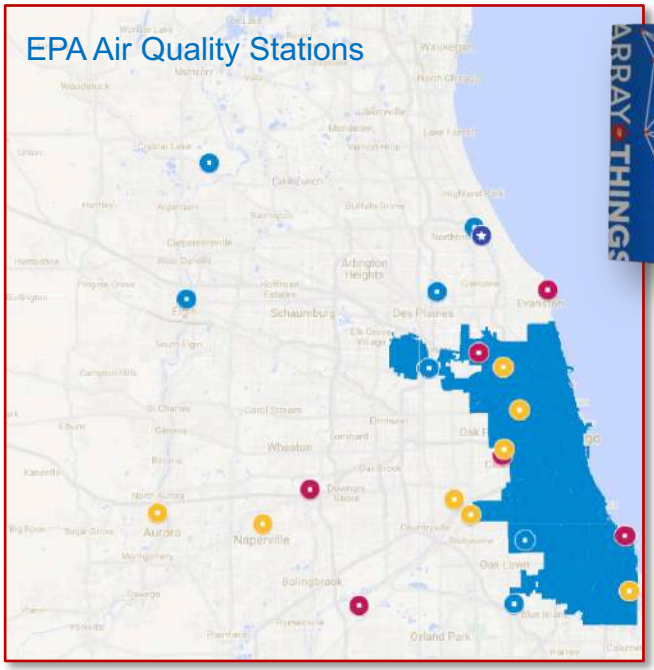
R. Ghani (UChicago Center for Data Science and Public Policy; R. Mansour, Chicago Department of Public Health)

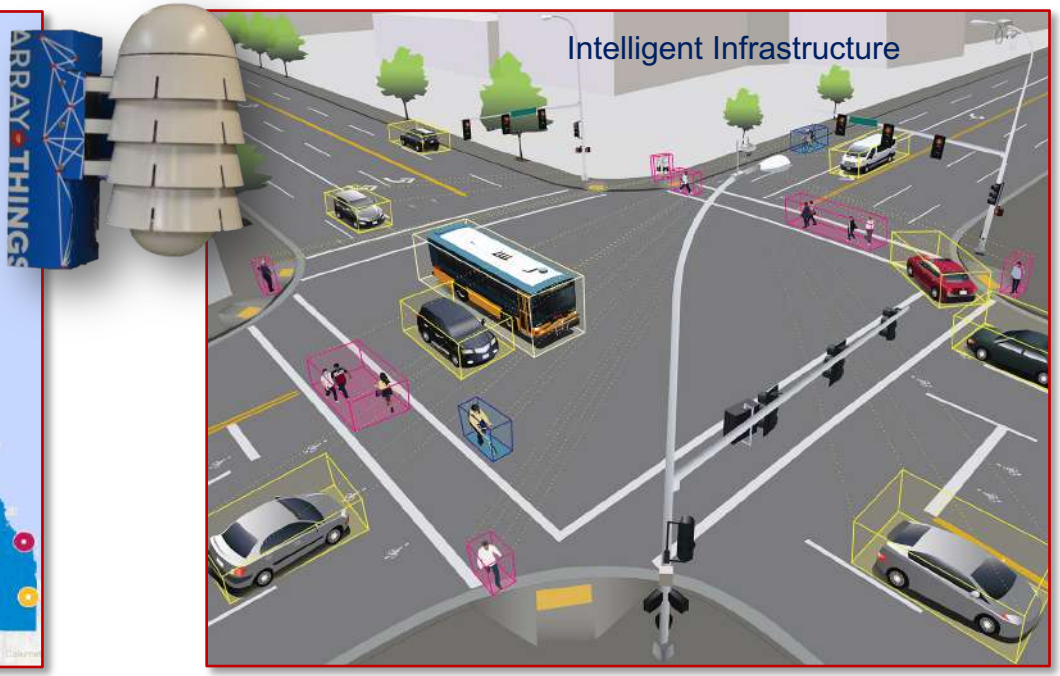
E. Potash, R. Ghani, R. Mansour, et. al., "Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning," ACM KDD'15, August 2015, Sydney, Australia.

The Array of Things Project

For many questions, much greater resolution is necessary.

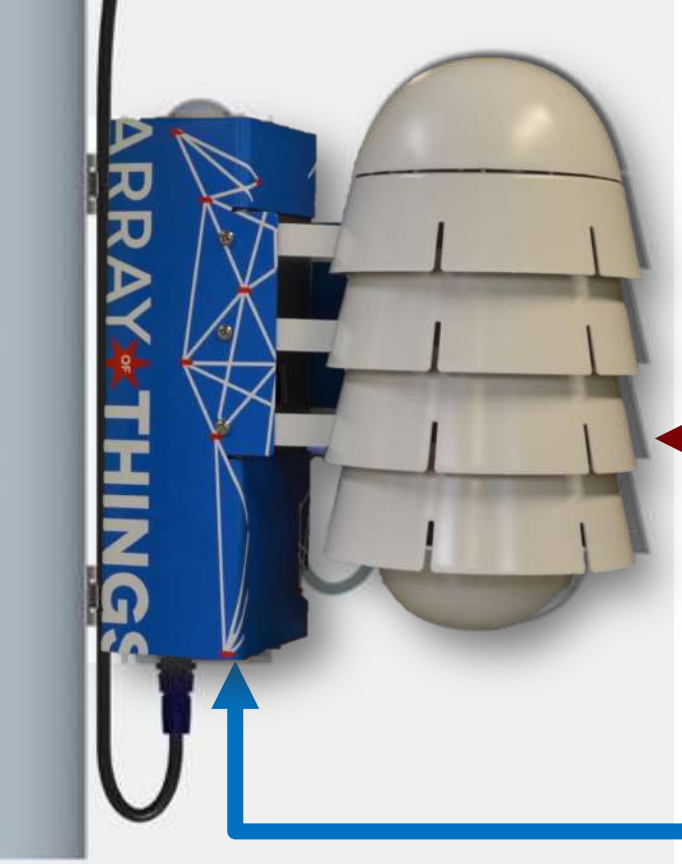
Array of Things (AoT): Beyond Passive Measurement





AoT is an NSF-funded Major Research Instrumentation project to create an urban cyberinfrastructure ”instrument” comprising hundreds of devices in partnership with the City of Chicago.

C. Catlett, P. Beckman (ANL)



AoT Current Configuration

Environment

- Ambient, UV, IR light
- Visibility
- Magnetic Field
- Vibration
- Sound pressure
- Temperature
- Relative humidity
- Barometric pressure

Air Quality

- PM 1, 2.5, 10, 40
- Carbon monoxide
- Ozone
- Sulfur dioxide
- Nitrogen dioxide
- Hydrogen sulfide
- Total reducing gases
- Total oxidizing gases

Edge Computing

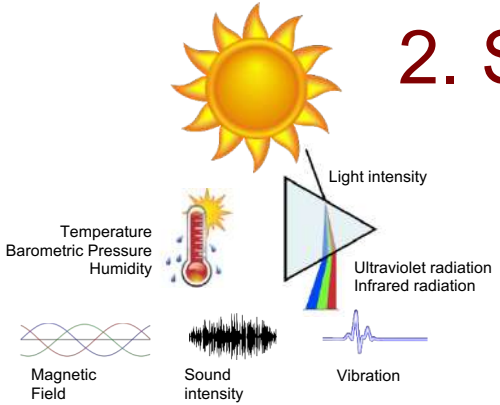
Computer Vision: Flooding, traffic flow, safety (bike helmet use, pedestrian patterns…), use patterns of public spaces, cloud cover

Computer Audio: Noise components, sound events

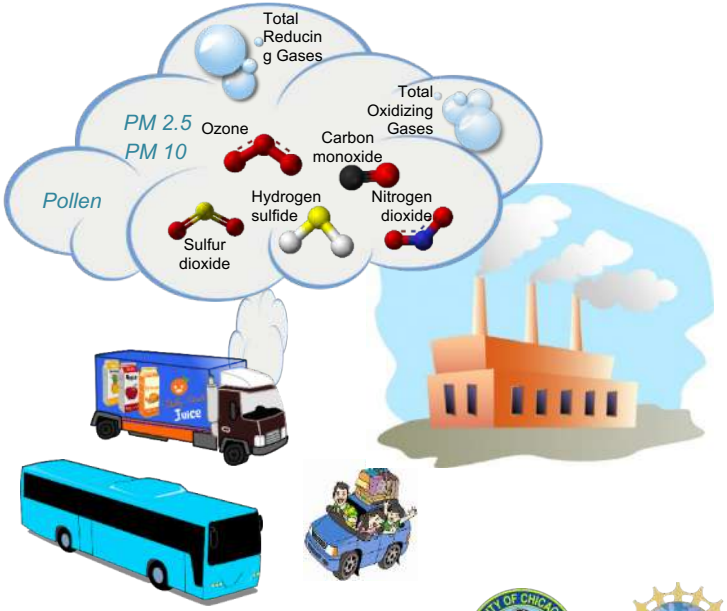
Pete Beckman, Rajesh Sankaran, Charlie Catlett (ANL)
Douglas Pancoast (SAIC), Dave Carhart (Astronics)
Bob Ramos, Michael Buchanan (Chicago Department of Transportation)

AoT: Cyberinfrastructure with Three Functions

1. Sensors



2. Sensor Testbed



3. Artificial Intelligence Testbed

C. Catlett, P. Beckman, R. Sankaran

All data is *open* and *free*.

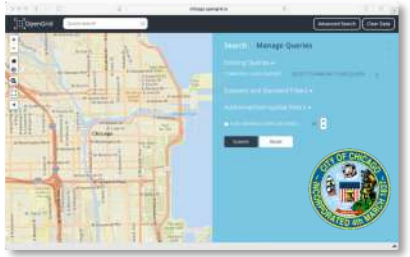
https://aot-file-browser.plenar.io/

Cellular Data Network

Open / Free Data

Download: Scientific Analysis

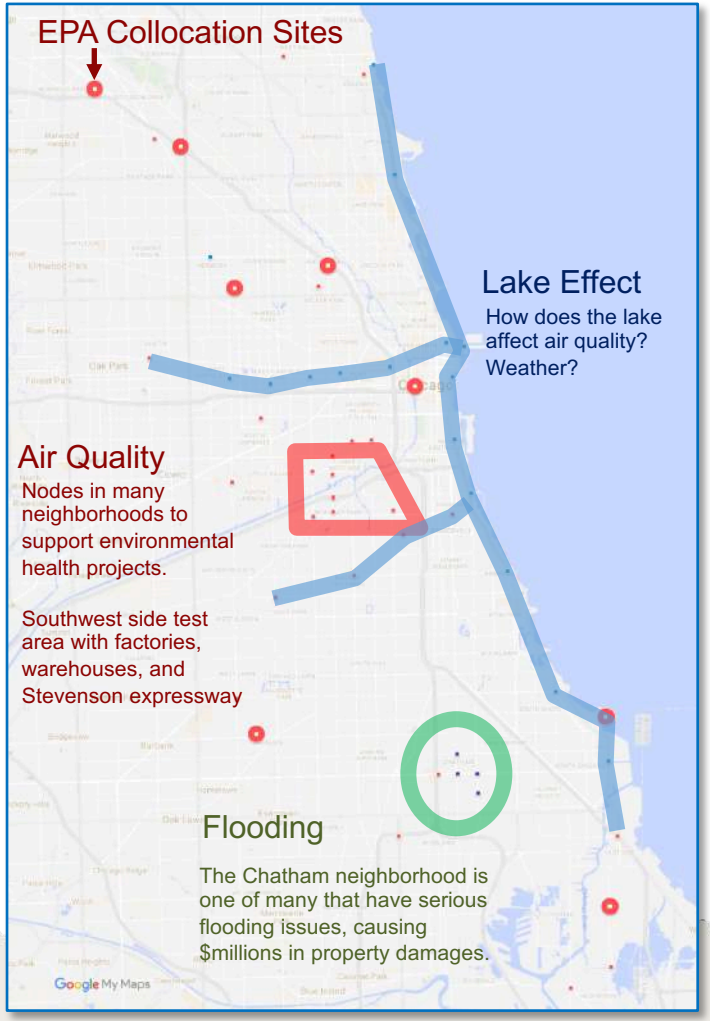
Portals: Data Discovery

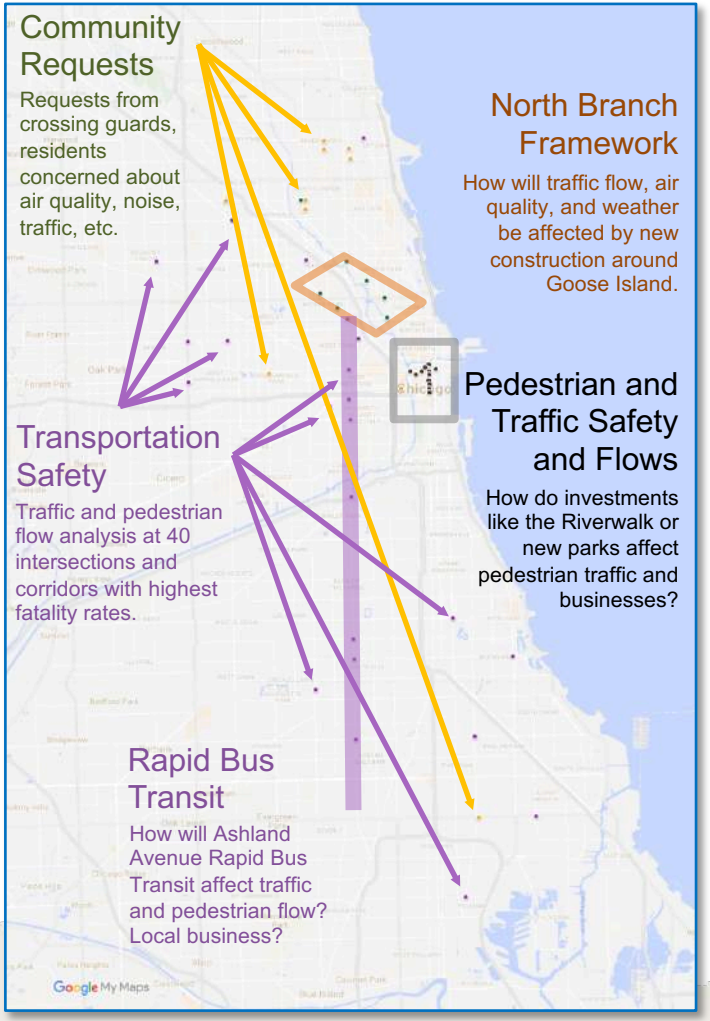


Real-Time Access: New Applications

Pete Beckman, Rajesh Sankaran, Charlie Catlett (ANL)

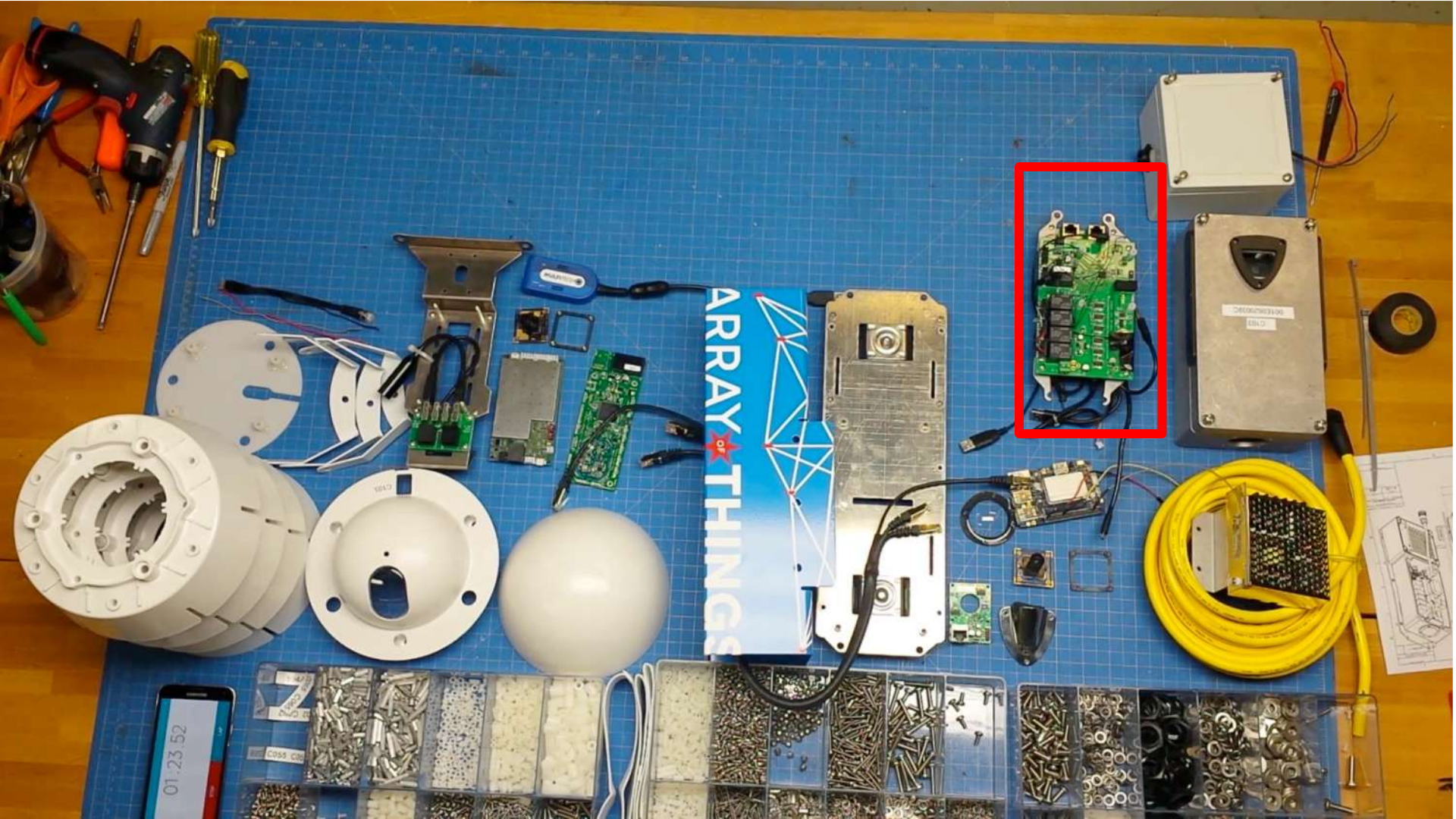
All nodes are identical, but placement is driven by specific science or policy questions.





Argonne NATIONAL LABORATORY

THE UNIVERSITY OF CHICAGO


ARRAY of THINGS
01:23.52

The Waggle Platform

Machine learning ”at the edge” of a continuum.

Northwestern University

THE UNIVERSITY OF CHICAGO

Argonne NATIONAL LABORATORY

Edge+Machine Learning: A Revolution

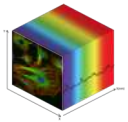
Sensors

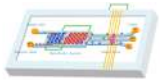
LIDAR

Software Defined Radios

Cameras, microphones

Hyperspectral Imaging





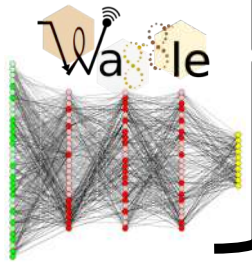
Microfluidic Sensors

Actuators

Servos

Dynamic adaptation

Powerful
Parallel Edge Computing



Artificial Intelligence
Deep Learning Inference

Semantic Output

Reduced, Compressed data

New inference (program code)

Edge computing and deep learning with feedback for continuous improvement

HPC/Cloud

Deep Learning Training

Pete Beckman beckman@anl.gov

The Computing Continuum

IoT/Edge — **Fog** — **HPC/Cloud**

Size	Nano	Micro	Milli	Server	Fog	Campus	Facility
Example	Adafruit Trinket	Particle.io Boron	Array of Things	Linux Box	Co-located Blades	1000-node cluster	Datacenter
Memory	0.5K	256K	8GB	32GB	256G	32TB	16PB
Network	BLE	WiFi/LTE	WiFi/LTE	1 GigE	10GigE	40GigE	N*100GigE
Cost	$5	$30	$600	$3K	$50K	$2M	$1000M

Count = 10^9
Size = 10^1

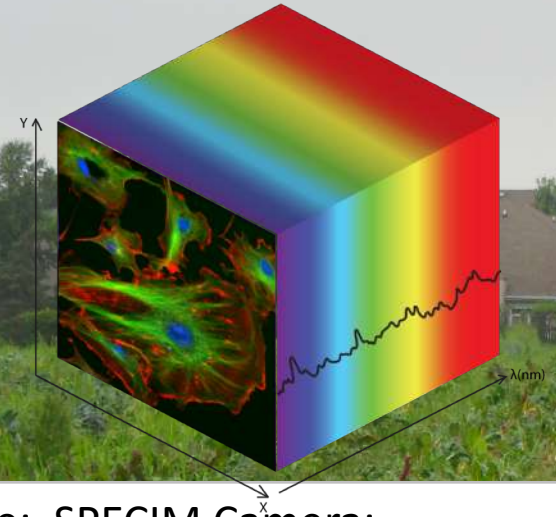
Count = 10^1
Size = 10^9

Pete Beckman beckman@anl.gov

Why Edge Computing?

- **More data than bandwidth**
  - Spallation neutron source, light source, HD Cameras, LIDAR, radar, hyperspectral imaging, grid micro-synchrophasors, etc.
- **Latency is important**
  - Quick local decision & actuation; adaptive sensing & control systems
- **Privacy/Security requires short-lived data: process and discard**
  - Compromised devices have no sensitive data to be revealed
- **Resilience requires distributed processing, analysis, and control**
  - Predictable service degradation, autonomy requires local (resilient) decision
- **Quiet observation and energy efficiency**
  - Vigilant sensors, transmit only essential observations, not big data streams

Pete Beckman beckman@anl.gov



Example: SPECIM Camera:
PFD VNIR with 768 bands
(2734 x 1312) x 768 x 2bytes = **5.1GB image**

1 sample every 5 min
Twilight to twilight on June 21 = **1TB**

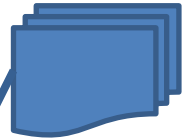
We need a parallel computer with each sensor!

Pete Beckman beckman@anl.gov

Approved Algorithms:

1. Surface water
2. Pedestrian count
3. Vehicle mix

After images are processed they are deleted. They are not transmitted or stored.



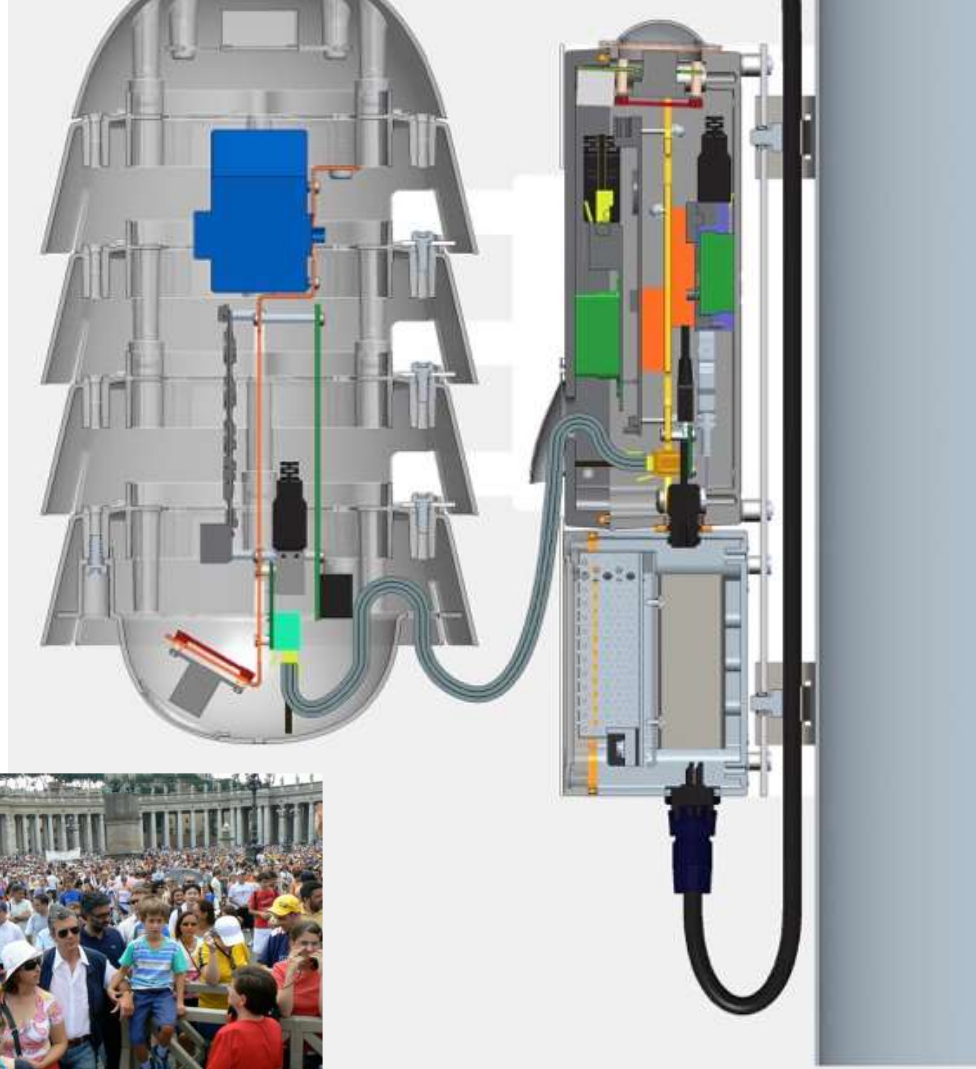
0%

329

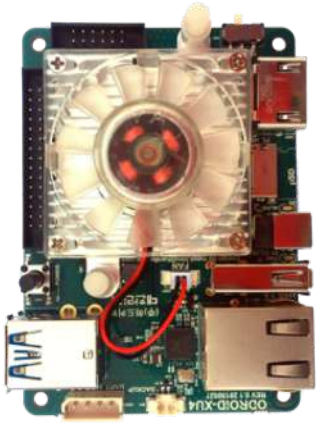
(0,0,0,0)

Encrypted Cellular Data Network

Central Data Repository



When a Computer + Linux is Not Enough…



Challenging Design Contradiction

⊥

- Experimental ML/GPU software fails often
- Edge Devices are remotely deployed

```
[20728464.998518] Kernel panic - not syncing: VFS: Unable to mount root fs
[20728464.998526] CPU: 0 PID: 1 Comm: swapper/0 Not tainted 3.10.0-229.4.2
[20728464.998532]  ffffffff81814288 0000000084f1a4a1 ffff880066eb7d60 ffff
[20728464.998540]  ffff880066eb7de0 ffffffff815fe71e ffffffff00000010 ffff
[20728464.998547]  ffff880066eb7d90 0000000084f1a4a1 0000000084f1a4a1 ffff
[20728464.998554] Call Trace:
[20728464.998565]  [<ffffffff81604eaa>] dump_stack+0x19/0x1b
[20728464.998570]  [<ffffffff815fe71e>] panic+0xd8/0x1e7
[20728464.998579]  [<ffffffff81a4560d>] mount_block_root+0x2a1/0x2b0
[20728464.998585]  [<ffffffff81a4566f>] mount_root+0x53/0x56
[20728464.998590]  [<ffffffff81a457ae>] prepare_namespace+0x13c/0x174
[20728464.998596]  [<ffffffff81a4527b>] kernel_init_freeable+0x203/0x22a
[20728464.998602]  [<ffffffff81a449db>] ? initcall_blacklist+0xb0/0xb0
[20728464.998609]  [<ffffffff815f33f0>] ? rest_init+0x80/0x80
[20728464.998614]  [<ffffffff815f33fe>] kernel_init+0xe/0xf0
[20728464.998620]  [<ffffffff81614d3c>] ret_from_fork+0x7c/0xb0
[20728464.998625]  [<ffffffff815f33f0>] ? rest_init+0x80/0x80
```

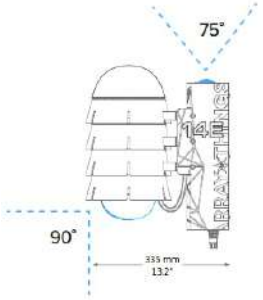
Pete Beckman beckman@anl.gov

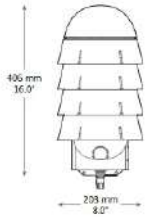
Waggle

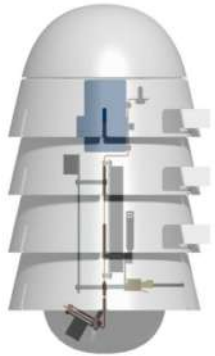
- Support “edge” computing—in situ data analytics for anomaly detection in high-bandwidth sensor steams, machine learning for computer vision, etc.
- Provide modular, secure data pipeline.

- Remotely program edge computing devices (and recover from “bricking”).

- Add sensors easily with minimal re-design (solder ports; shield expansion); rapidly swap devices for upgrades.







C. Catlett, P. Beckman, R. Sankaran

Argonne's Open Waggle Platform

A Pocket-Sized Controller for Edge Computing

The Waggle Manager (WagMan)

- Borrowed BG/Q control system ideas
- Designed mini "rack controller"
  - Devices can be disconnected
  - Devices can be power cycled
- "Deep Space Probe" design
  - Heart beat signals to each device
  - Alternative boot image / safe mode
  - Current and voltage monitoring
  - Environmental monitoring
- Strict cybersecurity design

Pete Beckman beckman@anl.gov

Waggle: Argonne's *Edge Computing* Platform

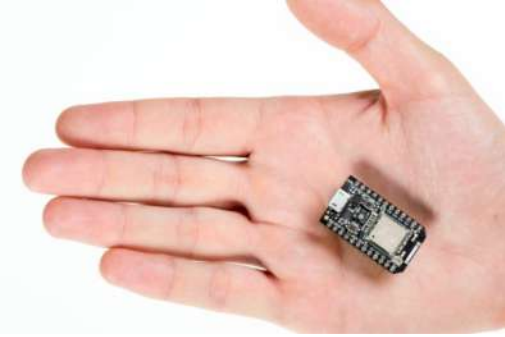
Bring Parallel Computing to the Edge

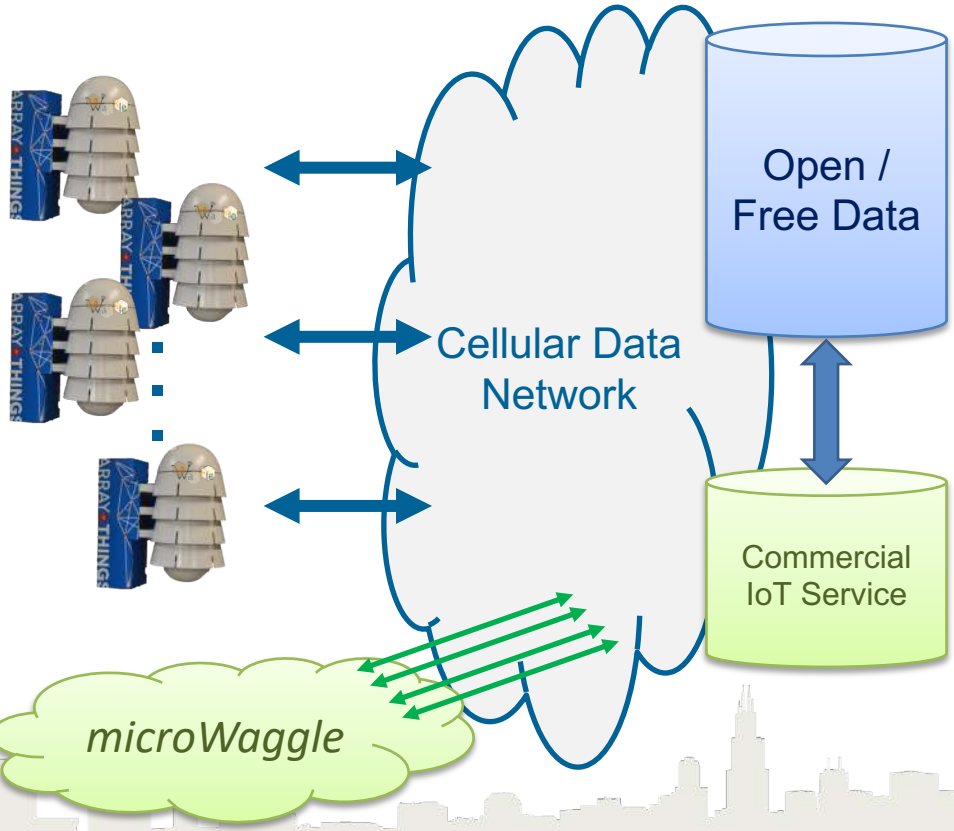
- Supports powerful, parallel computation at the edge
  - Computer vision and deep learning frameworks (Caffé, TensorFlow, OpenCV)
  - Supports edge-optimized & experimental computing
    - ML hardware, GPUs, neuromorphic, FPGAs, etc.
- Open Source, open interfaces
- Integrating advanced sensors: simple plug-in architecture
- Robust remote system management subsystem
- Manufactured at local electronics company

Pete Beckman beckman@anl.gov

~5 years of development by team at Argonne National Laboratory

Extensible Open Architecture



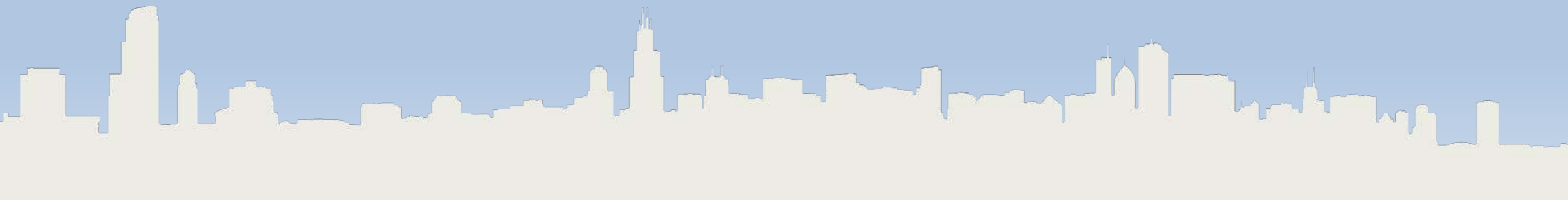


microWaggle uses cellular, mesh, or WiFi to connect lower-cost sensors, reporting data in Waggle format to mix readily with AoT data.

Pete Beckman, Rajesh Sankaran, Charlie Catlett (ANL)

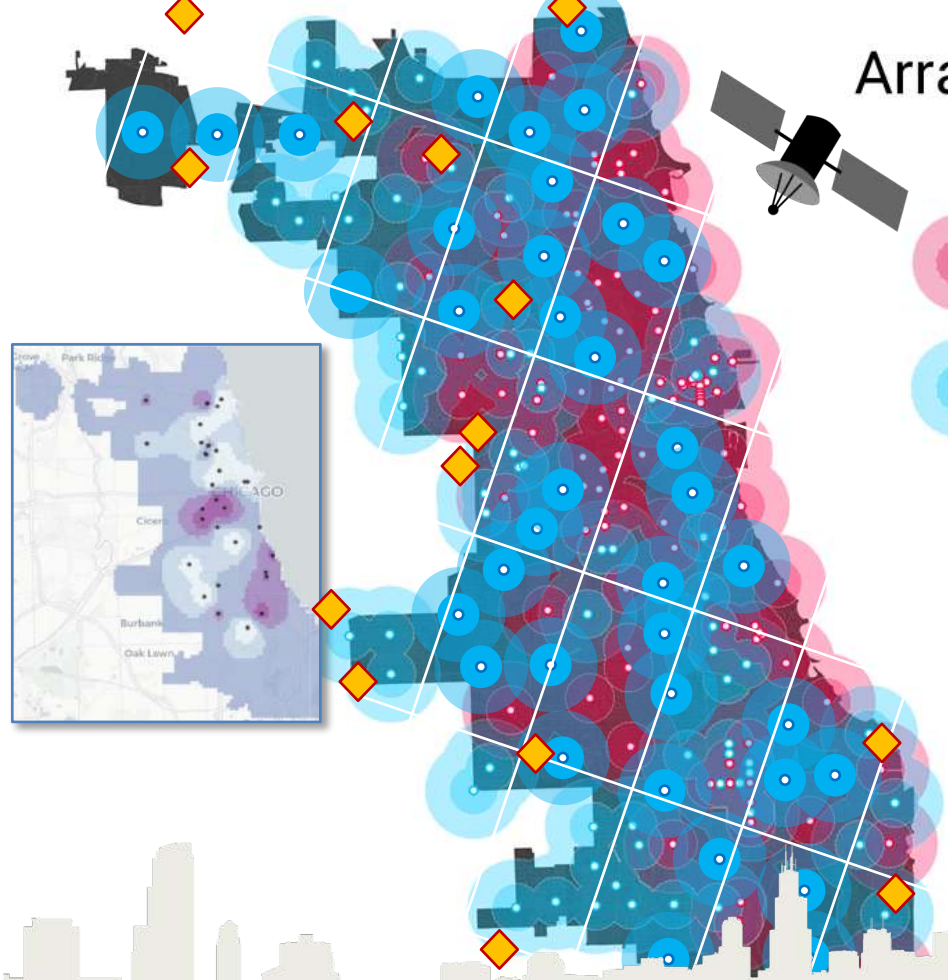
A Survey of Machine Learning Opportunities

Array of Things and Waggle present many interesting opportunities to apply machine learning to diverse science and policy questions.



Array of Things Coverage

July 2018



May 2018 Nodes

Dec 2018 Nodes

https://aot-file-browser.plenar.io/

Today there is an AoT node (red) within 2km of 80% of Chicago's population, and within 1km of 42% of the population

By early 2019 every Chicago resident will have an AoT node within 2km and at least 70% will live within 1km of a node.

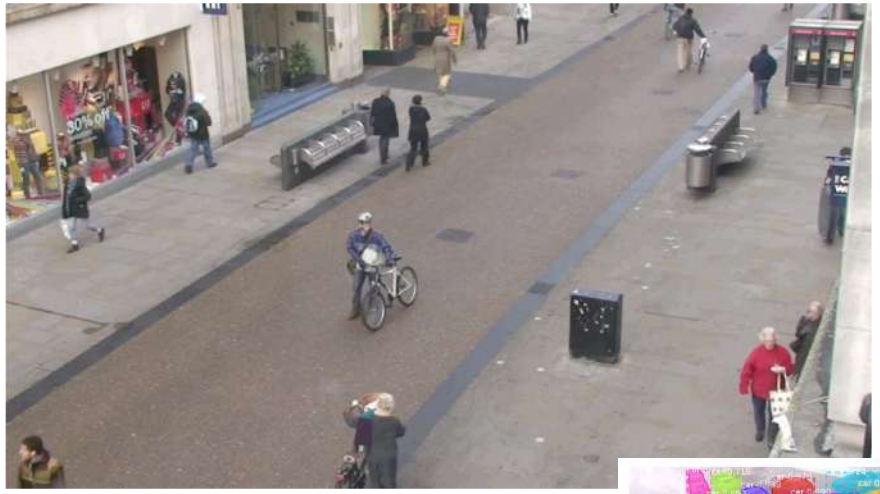
Map: A. Laha (UChicago Spatial Data Science Center)

THE CENTER FOR SPATIAL DATA SCIENCE THE UNIVERSITY OF CHICAGO

Argonne NATIONAL LABORATORY

THE UNIVERSITY OF CHICAGO

Transportation



Advanced computer vision to understand pedestrian movement, eventually to predict dangerous interactions with vehicles

Research Credits:
Zeeshan Nadir (Purdue PhD Student @ ANL, 2017)
Nicola Ferrier (ANL Scientist)

Mask r-CNN [He, 2017]
trained on COCO dataset

Research Credits:
Yongho Kim, Seongha Park
(Purdue PhD Students @ ANL, 2018)
Pete Beckman, Nicola Ferrier (ANL Scientists)

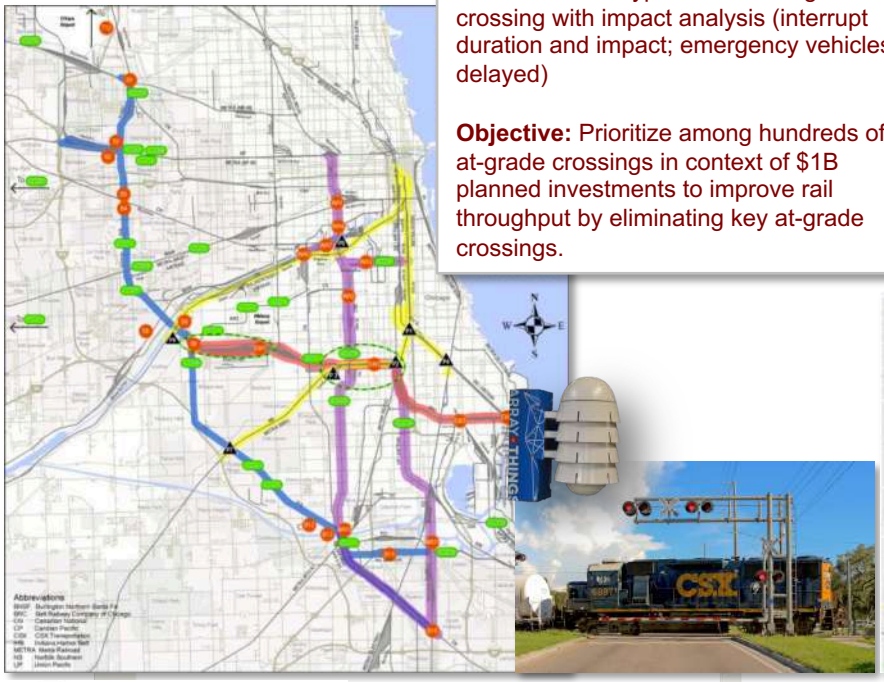
AoT Image from Lake Shore Drive

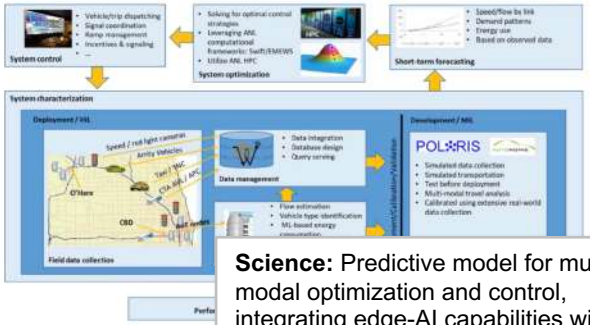
Pete Beckman beckman@anl.gov

Transportation

Science: Prototype model of at-grade crossing with impact analysis (interrupt duration and impact; emergency vehicles delayed)

Objective: Prioritize among hundreds of at-grade crossings in context of $1B planned investments to improve rail throughput by eliminating key at-grade crossings.





Science: Predictive model for multi-modal optimization and control, integrating edge-AI capabilities with traditional transportation data, coupled with HPC models and control systems.

Deployment: Integrate transportation measurements from AoT/Waggle (density, flow, vehicle mix, parking) with live traffic data and traffic model around O'Hare International Airport.

Illinois Department of Transportation

Funding: Illinois DOT

Partners: Argonne, UChicago, Chicago Metropolitan Agency for Planning, Chicago DOT

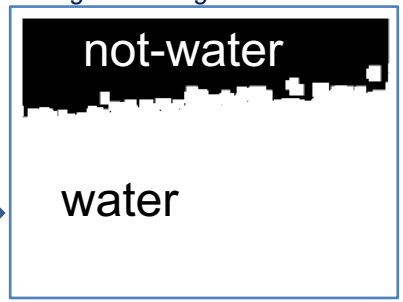
Funding: U.S. Department of Energy

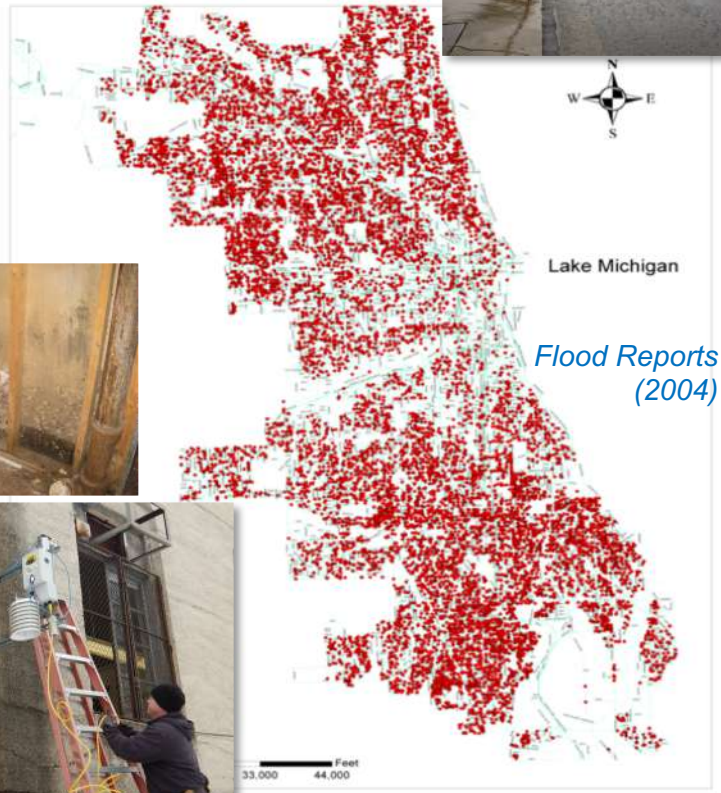
Partners: Argonne, Chicago DOT, Chicago Dept. of Aviation, Chicago Dept. of Innovation and Technology, Arity

C. Catlett, P. Beckman, N. Ferrier

Flooding

50 consecutive frames to flood water and segment image





Flood Reports (2004)

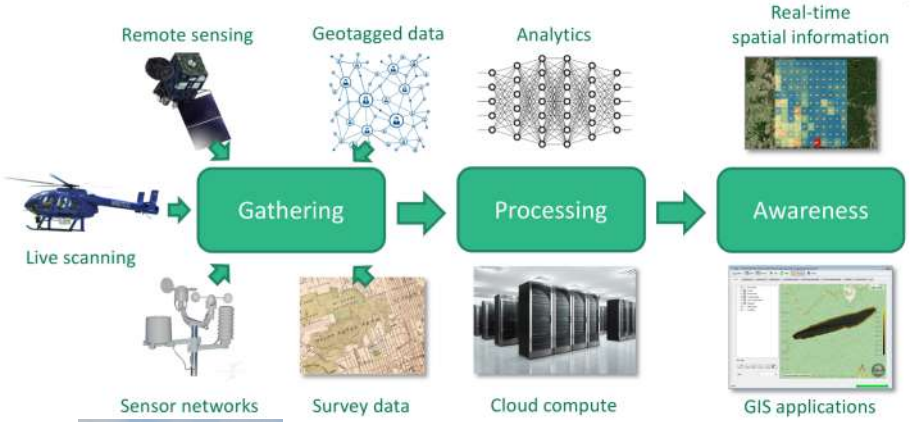
Using advanced computer vision to detect surface flooding

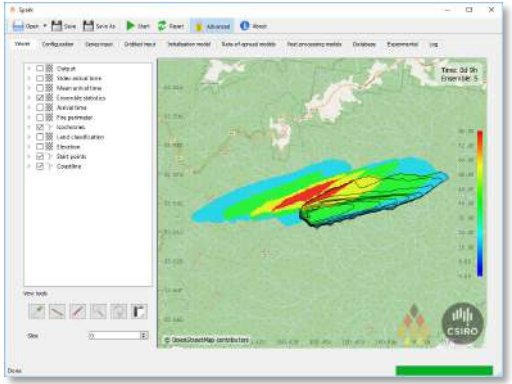
Ethan Trokie (Northwestern Undergrad Student @ ANL, 2017)
Vivien Rivera (Northwestern PhD Student, SCGSR 2018)
Nicola Ferrier (ANL Scientist)
Rajesh Sankaran (ANL Scientist)

Waggle Nodes in Tuley Park, Chicago

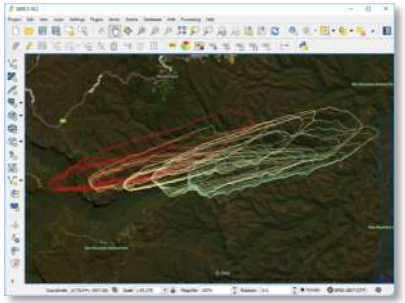
Disaster: Flood, Fire…

Partnership with CSIRO Australia (MOU, visiting postdoc)

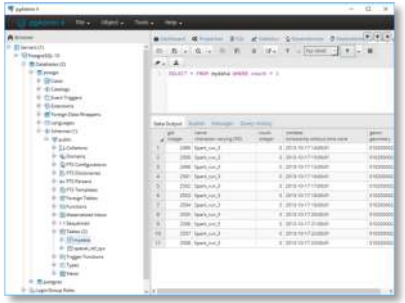




Impact analysis from ensemble wildfire simulations



GIS visualisation



Integration with PostGIS database

Nikhil Garg

DATA 61

CSIRO

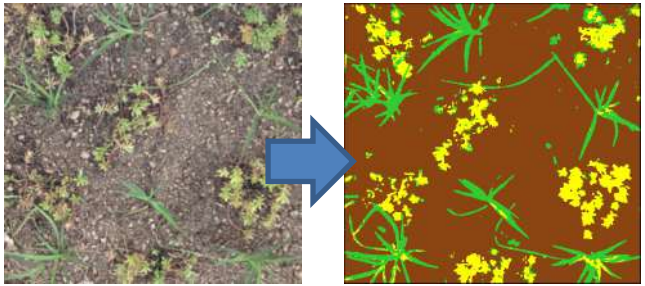
Live HPC Modeling and Prediction…

- Basis of D61 natural hazard applications:
  - Wildfire and wildfire impact (Spark)
  - Flood and coastal inundation (Swift)

Environmental Science

Advanced sensors and computer vision to monitor pristine prairie

Vivien Rivera (Northwestern Univ. PhD Student, 2018)
Aaron Packman, Bill Miller (Northwestern Univ. Professors)
Pete Beckman (ANL Scientist)



Original Auto Segmentation

Using advanced computer vision to monitor plants

Renee Zha (Northwestern Undergrad Student @ ANL, 2017)
Zeeshan Nadir (Purdue PhD Student @ ANL, 2017)
Nicola Ferrier (ANL Scientist)

Chicago Botanic Garden
Conservation Science Center

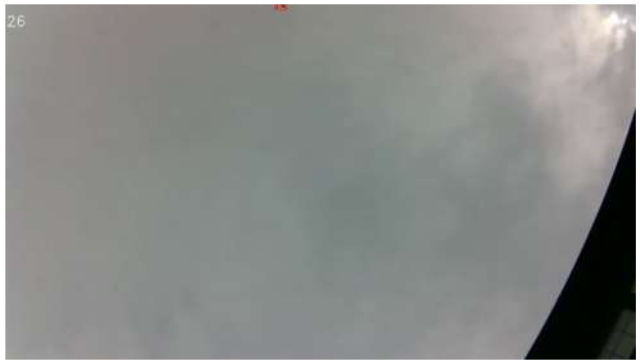
Undergrads Caeley and Jordan developed soil moisture sensor now deployed in Chicago

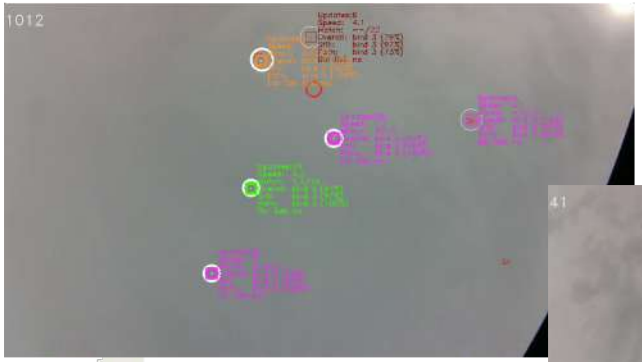
Public Safety

Bird? Plane? Drone?

Advanced computer vision and machine learning to identify drones, birds, or fixed-wing aircraft.

Sean Richardson (USAF visiting ANL)
Adam Szymanski (ANL Scientist)

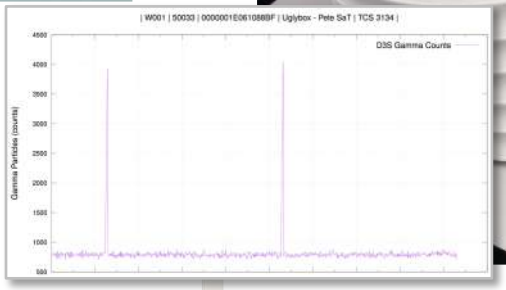


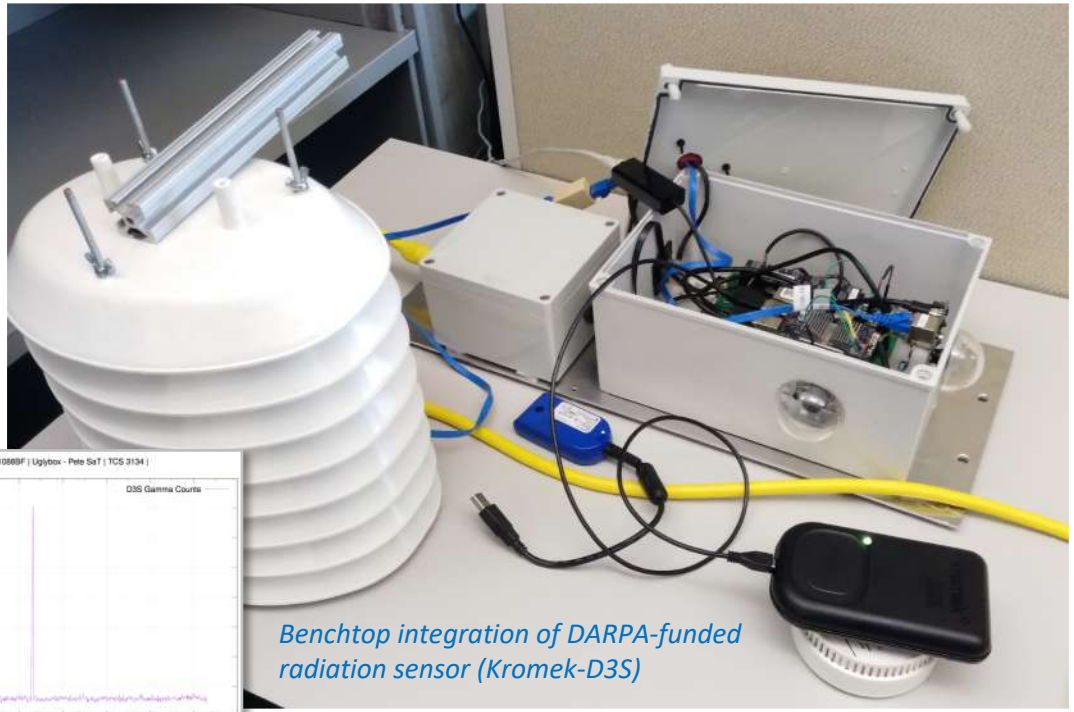


Object Classification

Public Safety

Testing Edge / Waggle
DARPA SIGMA+





Benchtop integration of DARPA-funded radiation sensor (Kromek-D3S)

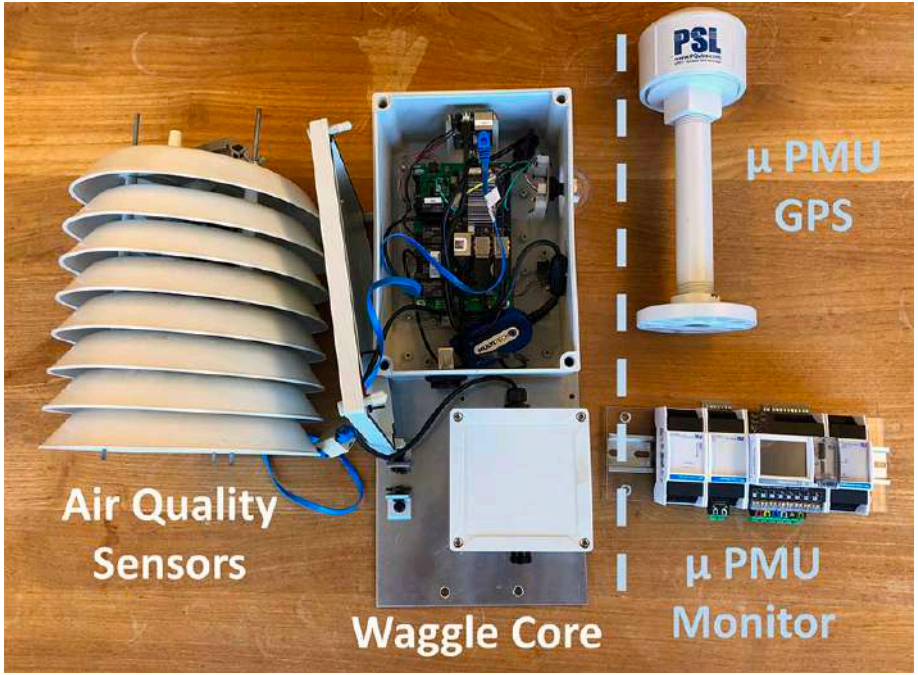
Pete Beckman, Rajesh Sankaran, (Argonne)

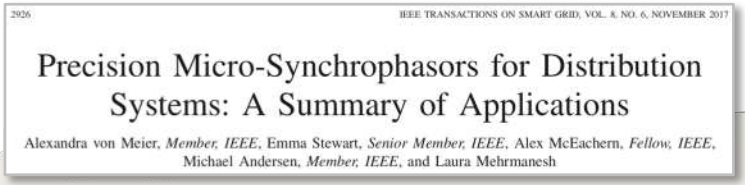
Energy: Power Grid

Argonne NATIONAL LABORATORY

Exelon

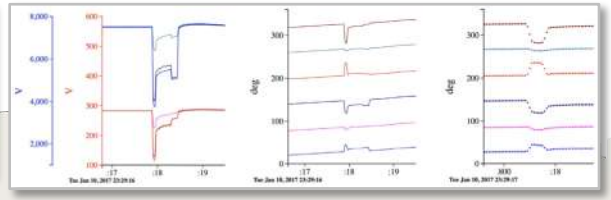
- Load Forecasting
- Grid Stress
- Air Quality




2926 IEEE TRANSACTIONS ON SMART GRID, VOL. 8, NO. 6, NOVEMBER 2017

Precision Micro-Synchrophasors for Distribution Systems: A Summary of Applications

Alexandra von Meier, *Member, IEEE*, Emma Stewart, *Senior Member, IEEE*, Alex McEachern, *Fellow, IEEE*, Michael Andersen, *Member, IEEE*, and Laura Mehrmanesh



DOE ARPA-E

Harnessing The Computing Continuum

Science-driven Problems

e.g.: “Predict urban response to rainfall, trigger intelligent reaction…”

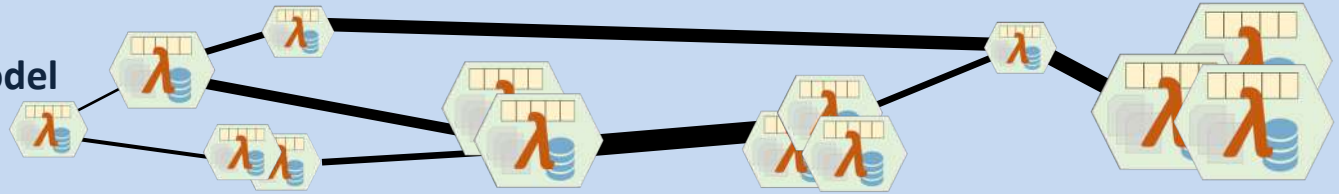
Research Focus

Goal-oriented Annotations

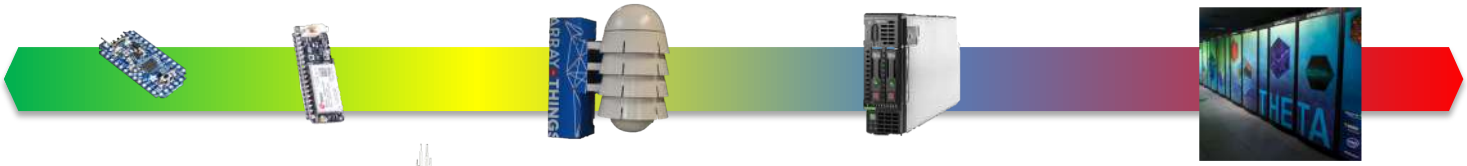
Notional Example:

```
trigger {flood_actuation, resident_warning}
  when {wx_prediction, sewer_model} implies
    (traffic_capacity < 70%) or (home_flooding > 5%)
```

Continuum Abstract Model & Runtime



Existing Resources & Services



Argonne Edge Research: Edge→HPC

- Continually improving Edge-HPC Systems
  - Deep learning + lightweight training + continual improvement
  - Incremental model updates
  - Is Edge really a layer in the model?
- How will the OS/R and system software evolve for Edge-HPC?
  - Scheduling, security, resource management, streaming data
- Programming model & framework for Continuum Computing
- Optimized ML hardware for both Edge & HPC
- Theoretical foundations for failures and correctness of edge/training
- Dynamic resource management and adaptive inference priority
  - AI at the Edge is limited by power and computation – just like HPC
- **Fluid HPC** to support complex and on-demand workflows on future exascale