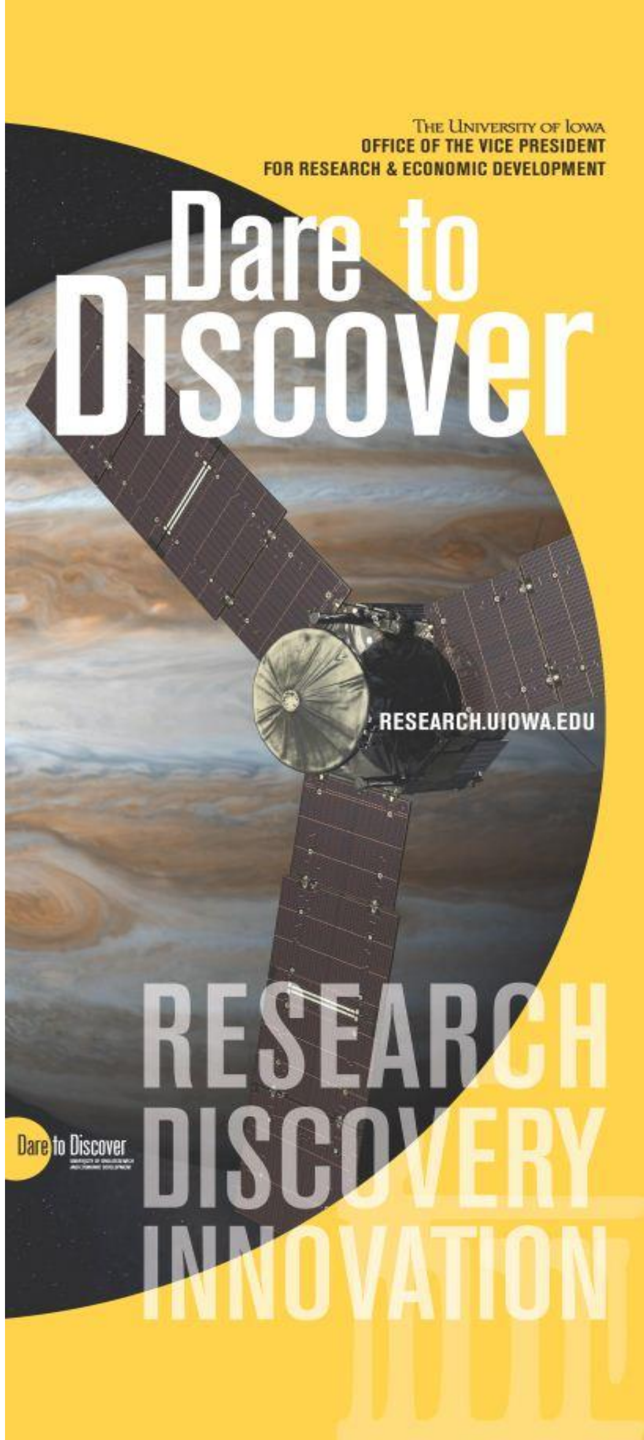
Beyond the Hype: Convergence Technologies

AKA a placeholder title ☺

Dan Reed
dan-reed@uiowa.edu
University of Iowa

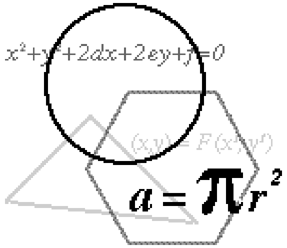


Discussion roadmap

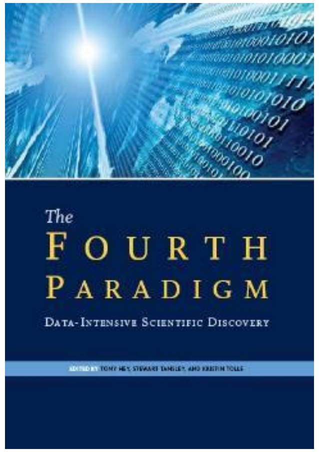
- Big data and science
- Steaming versus batch ecosystems
- Epistemology of discovery
- Inexpensive sensor networks

THE UNIVERSITY OF IOWA

The changing nature of scientific research

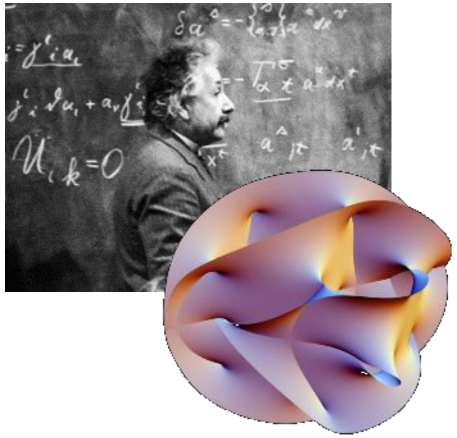


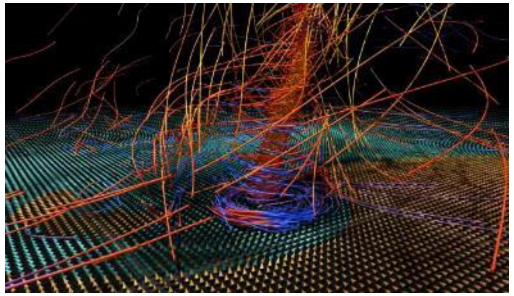
$$H(t)|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle$$



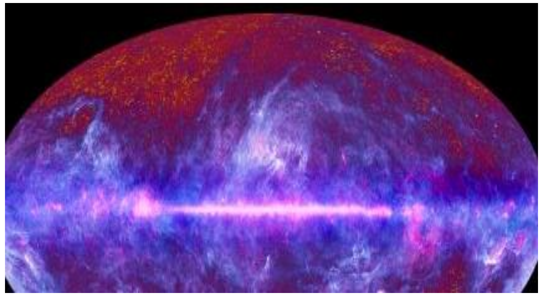
Experimental → Theoretical → Computational → Data Exploratory

Large Hadron Collider





Severe Storm Model (NCSA)



ESA Planck Sky Survey

The University of Iowa

Large Synoptic Survey Telescope (LSST)

Medium, big data

Structure

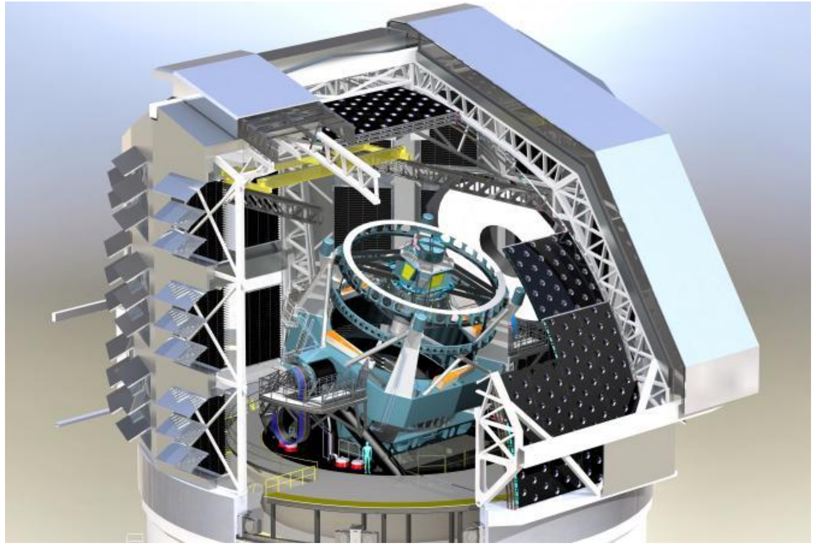
- 8.4 meter telescope with 3.5 degree FOV
- 3.2 gigapixel camera
- Construction underway in Chile

Science

- Dark matter/energy and gravitational lensing
- Outer solar system and NEOs
- Milky Way structure and evolution

Data and analysis

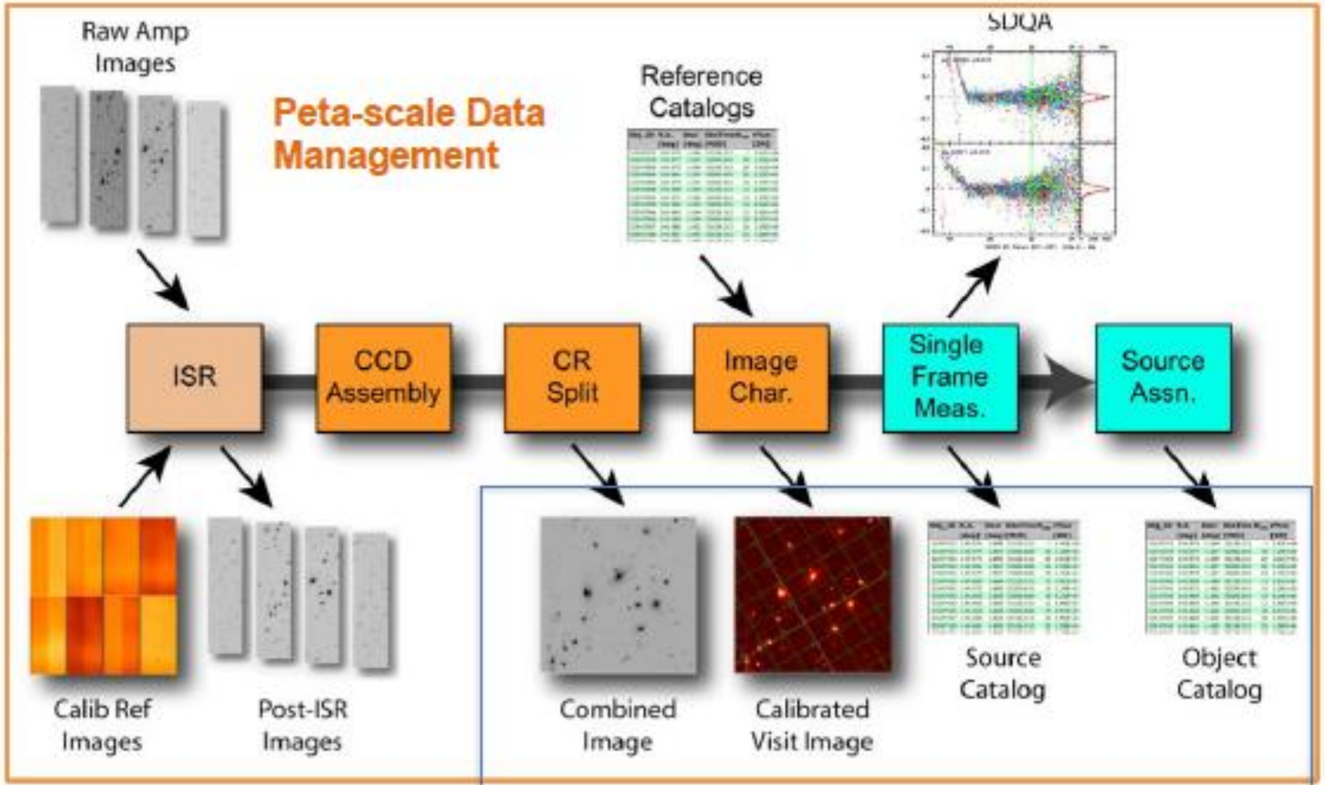
- ***~15 TB/night and 200K images/year (1.3 PB)***
- 60 PB raw and 15 PB catalog (over 10 ears)
- 50-150 TF computing need



Large Synoptic Survey Telescope

THE UNIVERSITY OF IOWA

LSST Basics



Science and EPO user Interfaces

Likely cloud-based

LSST data processing

Data captured in Chile
Buffered in Chilean data center
Transmitted via WAN to NCSA
NCSA does “real-time” processing

- 60 second trigger for interesting events

Two years of data releases on disk at NCSA

- Science workflows run against this data
- External requests satisfied for data subsets

Cooperative processing/storage at IN2P3

Blue Waters is not used

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

cnrs IN2P3 Les deux infinis

THE UNIVERSITY OF IOWA

Stateless versus stateful computation

Stateful (aka streaming/continuous)

- Long term (days, months, years)
- Environmentally responsive (trigger sensitive)
- Discipline-tailored environment

Data and computing co-resident

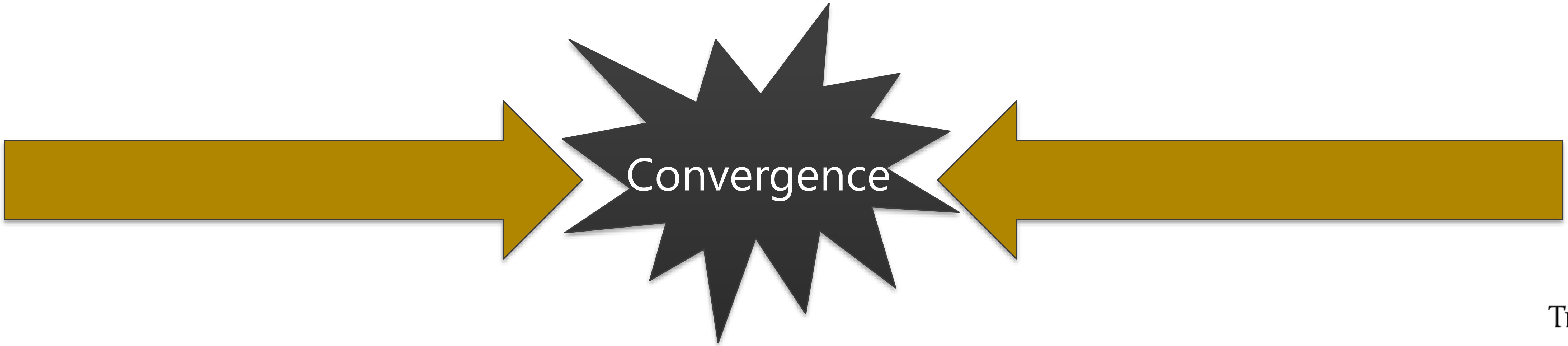
Increasingly complex workflows

Stateless (aka batch)

- Short term (hours, days)
- Oblivious (generally not trigger sensitive)
- Multidisciplinary environment

Data staged for computation

Increasingly complex workflows

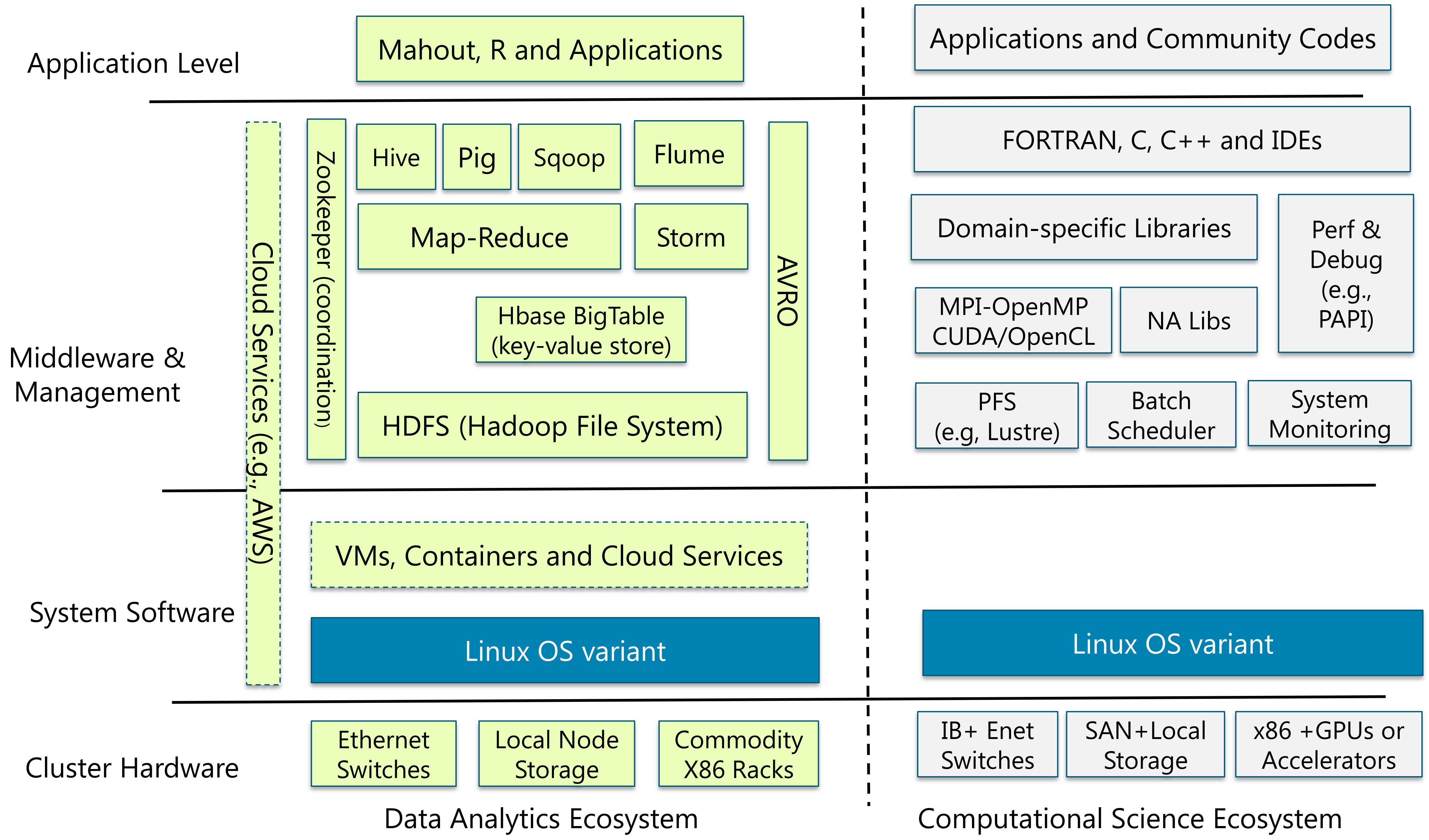


Convergence challenges: among many

Differing culture and tools
Shifting workforce demands
Dependence on retargeted infrastructure
Stream and batch model optimization
Content distribution networks (CDNs)
Clouds and edge computations
Virtualization and containerization
Security and provenance
Performance requirements

While, ironically …
… big data hardware and HPC hardware are converging

Divergent ecosystems (Reed/Dongarra, CACM, July 2016)



The University of Iowa

Have you ever …

Requested 200 nodes and 2 PB for *four years*?

Logged onto a node and killed processes just to see what would happen?

Wished you could load containers rather than just applications?

Found your code performance limited by the I/O bandwidth of a Raspberry Pi?

Thought SAN was just a typo in a message meant for Sam?

Asked your system for recommendations?

Wondered why R came after S and C doesn't matter?

The University of Iowa

What they didn't tell you in school about science

Scientific inference

- Abduction (guessing at an explanation)
- Deduction (determining necessary consequences of set of propositions)
- Induction (making a sampling-based generalization)

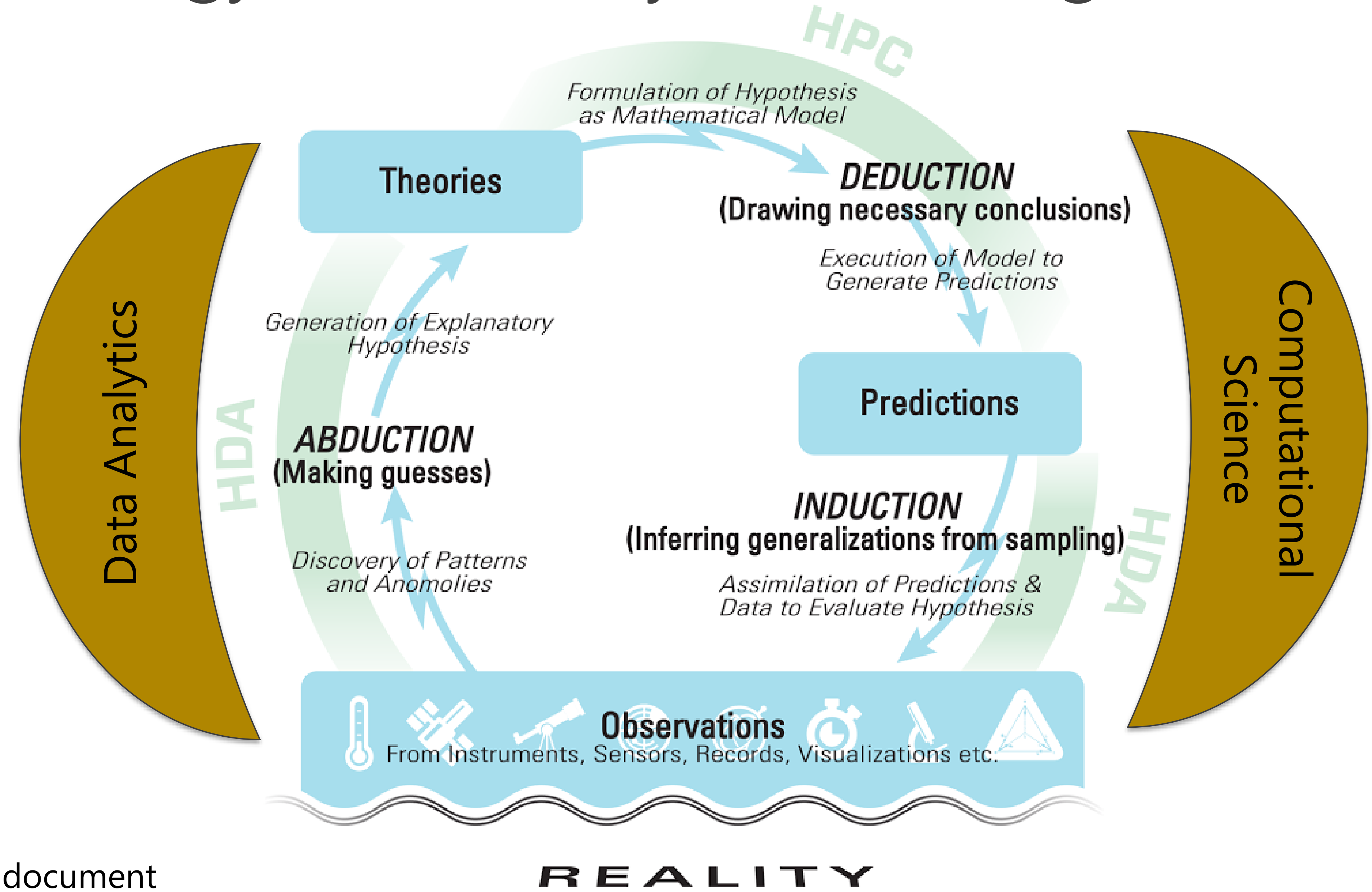
Or as Richard Feynman put it

- We guess at a law that would explain what is currently inexplicable
- We derive the consequences of the law that we guessed
- We make further observations to see if the consequences predicted match the reality we find

It's more than hypothesis, experiment, theory

And most of us in computational science focus on just a subset …

Epistemology – the theory of knowledge



Source: BDEC document

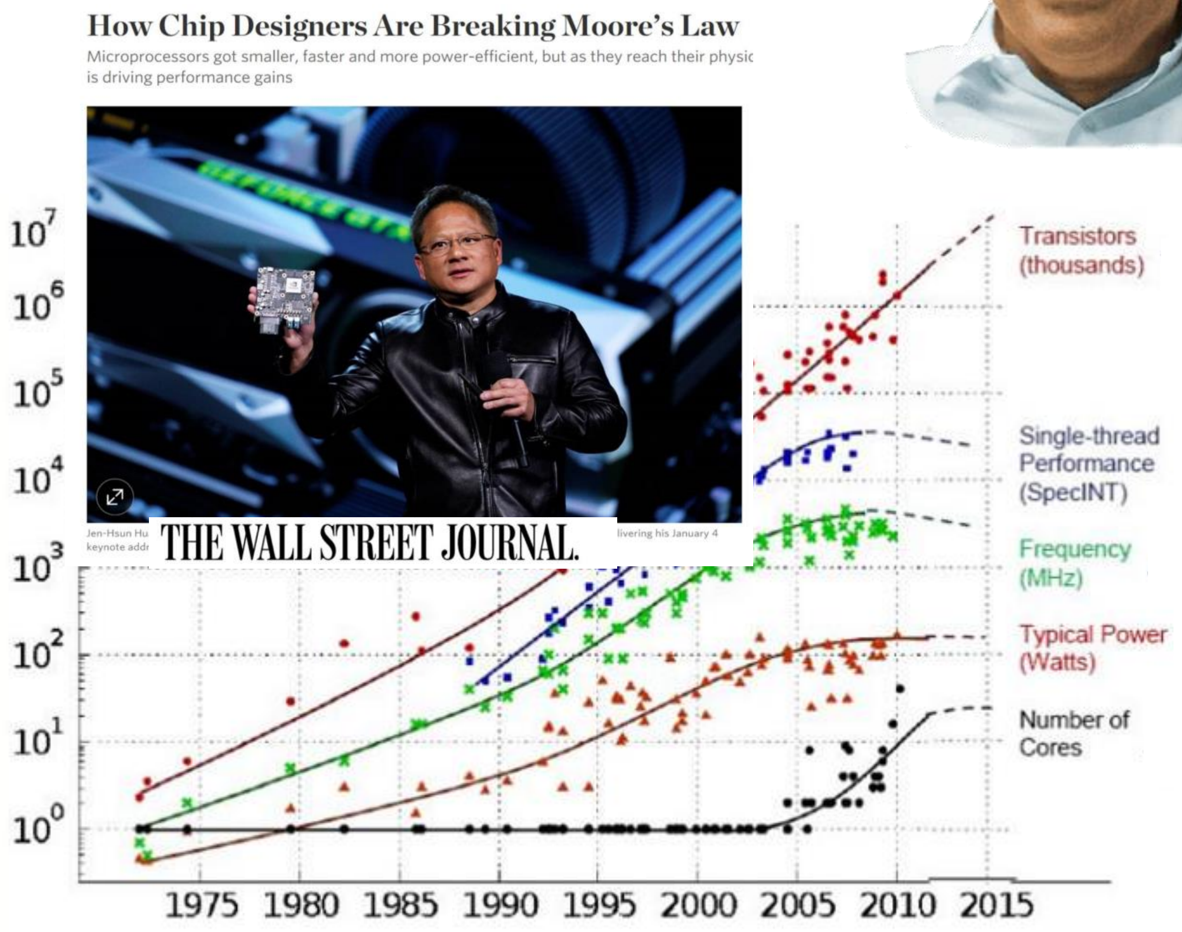
THE UNIVERSITY OF IOWA

Room at the bottom: less and less

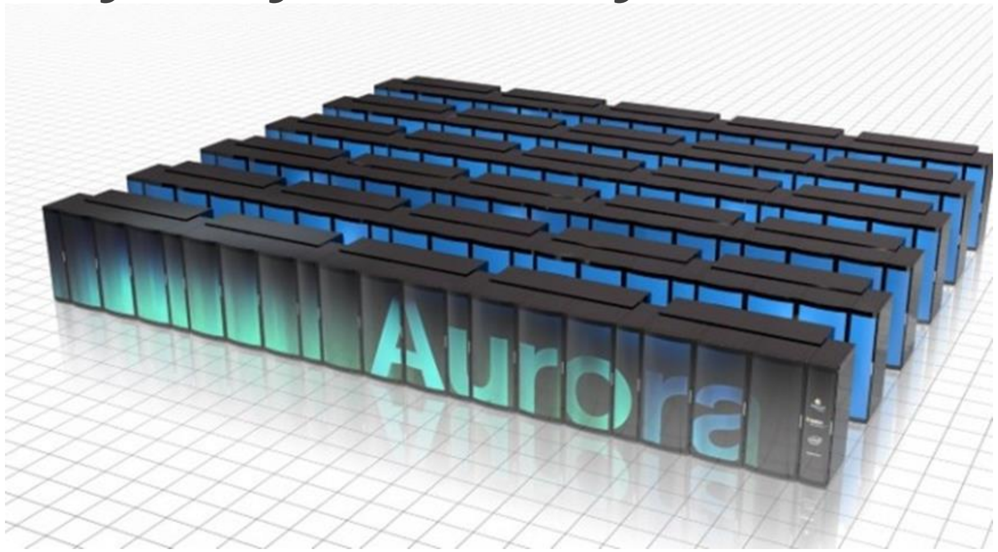
Yeah, we did that – Dennard scaling has come and gone

Why can't we make them [computers] very small, make them of little wires, little elements – and by little, I mean *little*. For instance, the wires should be 10 or 100 atoms in diameter, and the circuits should be a few thousand angstroms across. Everybody who has analyzed the logical theory of computers has come to the conclusion that the possibilities of computers are very interesting – if they could be made to be more complicated by several orders of magnitude. Richard Feynman

December 29, 1959



Room at the top: bounded {mostly} by money



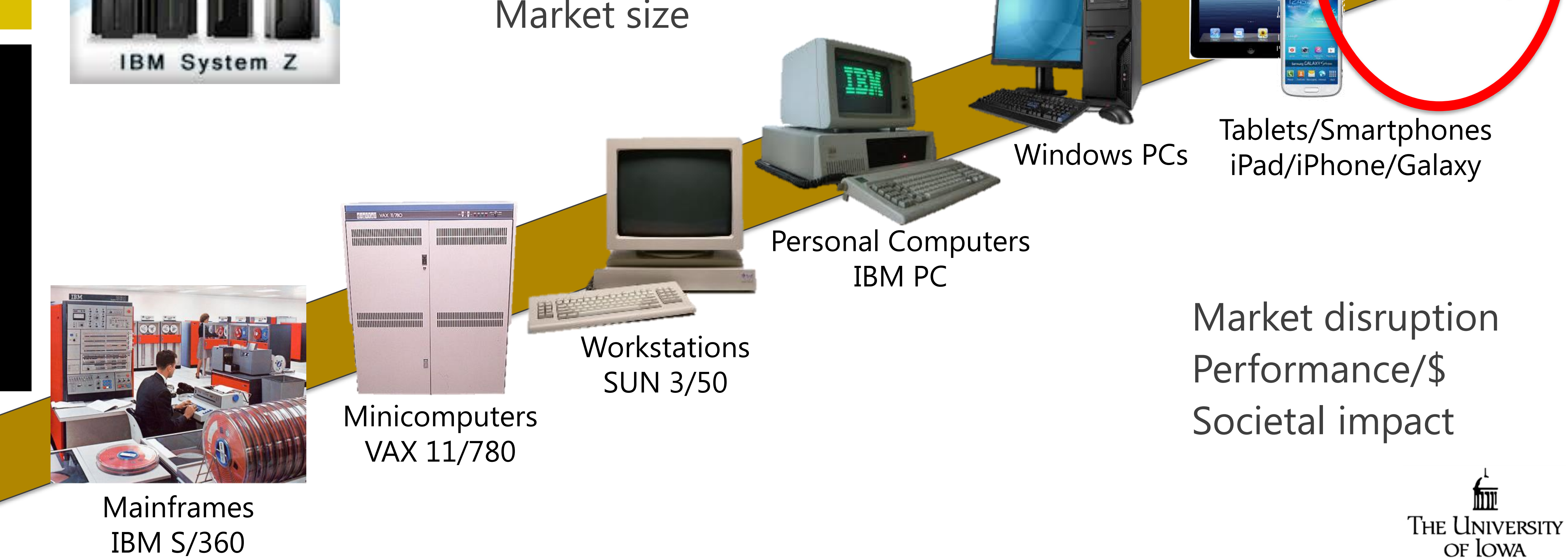
Inspiration: Christian Belady

THE UNIVERSITY OF IOWA

Generation after generation

Disrupted from below by 10X

Volume
Unit price
Market size



Market disruption
Performance/$
Societal impact

Raspberry Pi ecosystem

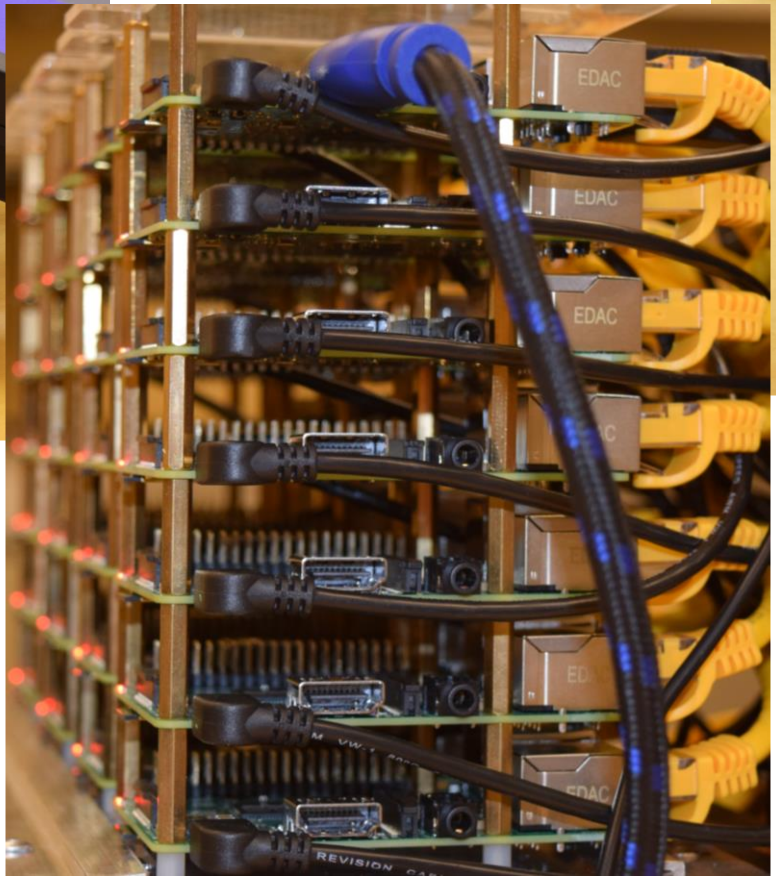
Raspberry Pi Zero

- 1 GHz single-core CPU
- 512 MB RAM
- Mini-HDMI port
- Micro-USB OTG port
- Micro-USB power
- HAT-compatible 40-pin header
- Composite video and reset headers
- CSI camera interface (v1.3 only)
- ***~$10 at better toy stores near you***

Raspberry Pi 3

- 1.2 GHz 64-bit quad-core ARMv8 CPU
- 1 GB RAM
- VideoCore IV 3D graphics core
- 802.11n Wireless LAN plus Bluetooth 4.1 with BLE
- 40 GPIO pins
- 4 USB ports plus HDMI and Ethernet ports
- Combined 3.5mm audio/composite video jack
- Camera (CSI) and display (DSI) interfaces
- ***~$35 at better toy stores near you***

Raspberry Pi2 educational cluster

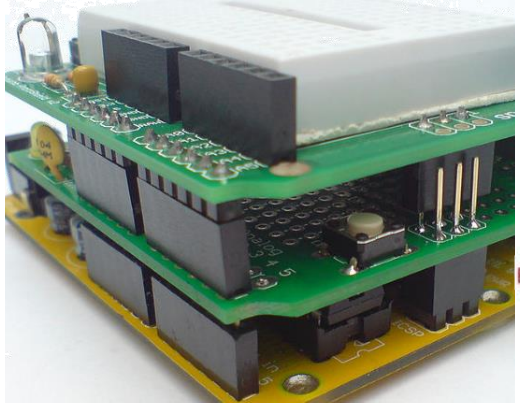


THE UNIVERSITY OF IOWA

Arduino ecosystem: lots more room at the bottom

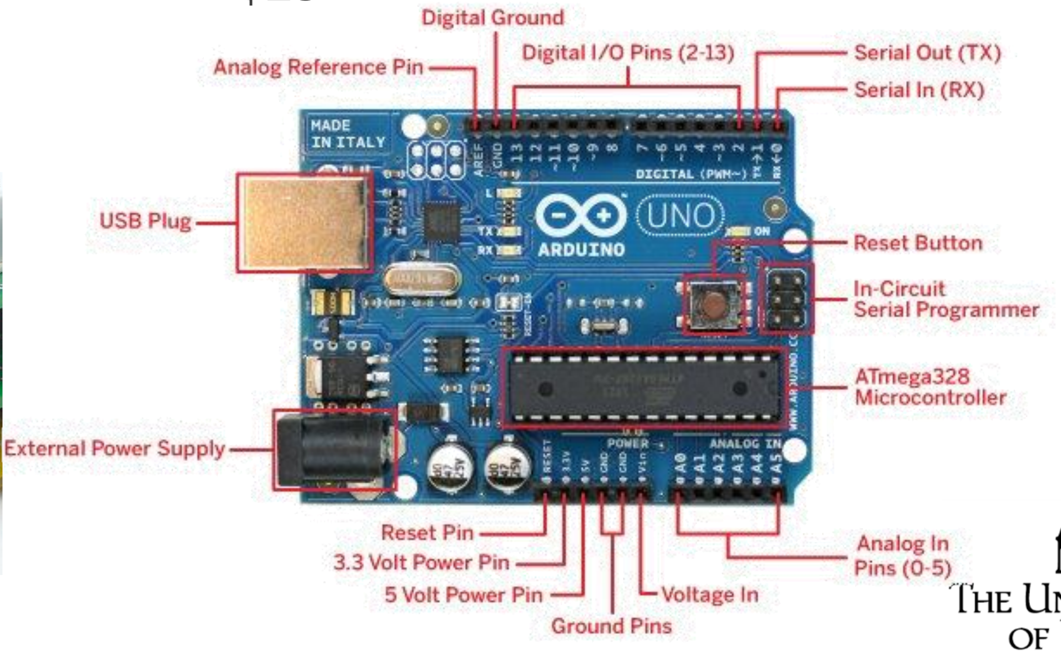
Open source hardware

- (Typically) Atmel 8, 16, or 32-bit AVR microcontroller
  - Also ARM Cortex-M0/M3 (32-bit)
  - Also Intel Quark x86 (32-bit)
- Bootloader to onboard FLASH
- Digital and analog I/O pins
- Add-on modules (shields) via I²C serial bus
  - Just about anything one could imagine
  - GPS, LoraWAN, sensors, actuators, …
- Standard IDE
- ***Prices range from ~$2 up***



Arduino UNO R3

- 16 MHz ATmeta328 chip
- 32 KB FLASH
- 2 KB SRAM and 1 KB EEPROM
- 14 digital I/O and 6 analog I/O pints
- USB port
- ~$25



THE UNIVERSITY OF IOWA

LoRaWAN

Unlicensed Industrial, Scientific and Medical (ISM) band

- 868 MHz (Europe) and 915 MHz (United States)

Low power operation with 0.3 Kb/s to 50 Kb/s data rates

Asynchronous, ALOHA-based protocol

Multiple kilometer range

- ~10-20 open space and ~1-2 built environment

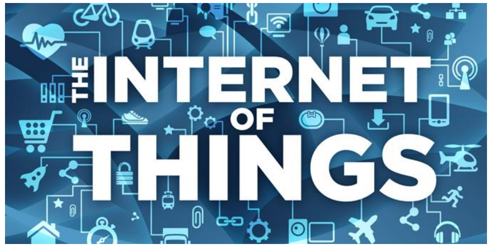
Three device classes

- Class A (bidirectional, unicast)
- Class B (bidirectional, with scheduled receive slots)
- Class C (bidirectional, unicast/multicast)

Originally developed by Semtech

Sources: https://www.lora-alliance.org

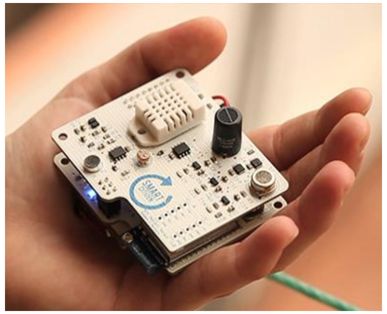
A framing question …



What would you do with 10,000 ~$10 wireless sensors?

- Natural and built environments
- Low bandwidth data streaming
- Complex social and technical questions

… and the data from them?



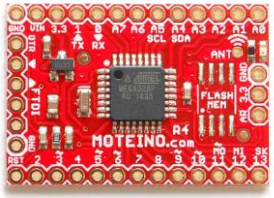
Iowa Quantified: environmental sensor network

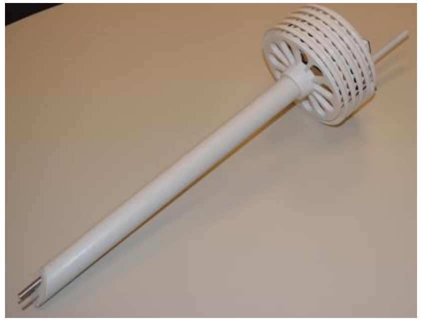
Research problem

- Flooding and environmental damage
- Citizen science engagement

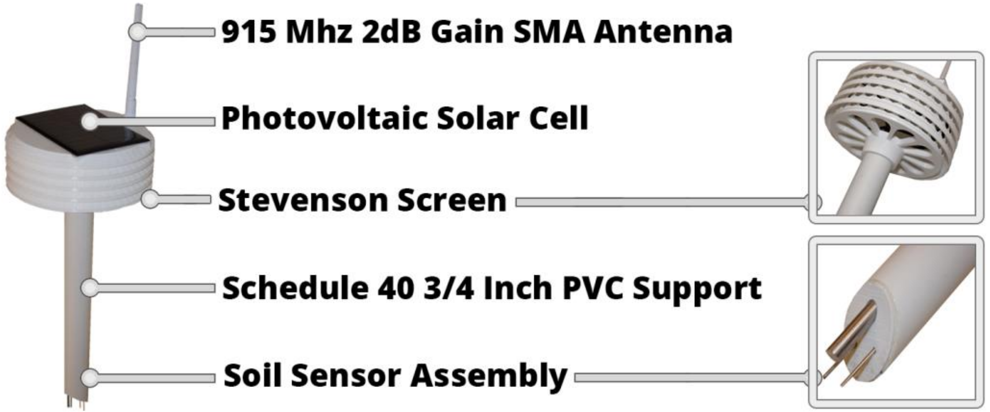
Requirements

- Simple to deploy
- Hands free operation
- Wireless access
- 2+ year lifetime
- CHEAP
- Did I mention CHEAP?



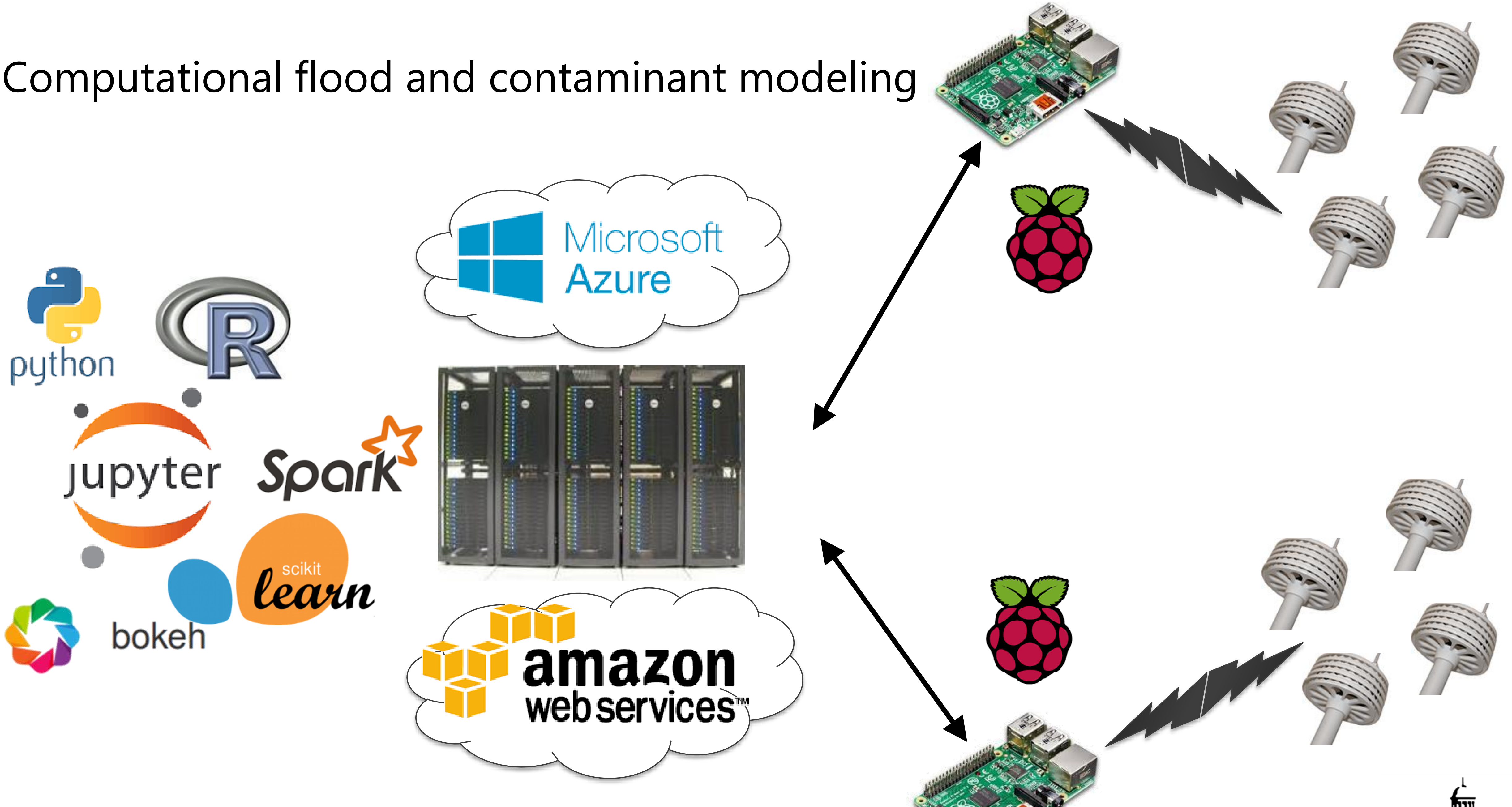


Part	Retail Price
Moteino with Atmel ATmega328P + LoraWAN	$22.95
5V 1W solar cell	$1.00
3.6V 2.1 Ah NiMh battery	$3.00
915 MHz 2dB gain SMA antenna	$0.25
Bosch BME 280 sensor (temperature, humidity, pressure)	$3.20
Soil moisture/temperature sensors	$2.35
3-D printed Stevenson screen + PVC	$1.60
TOTAL	**$34.35**



The University of Iowa

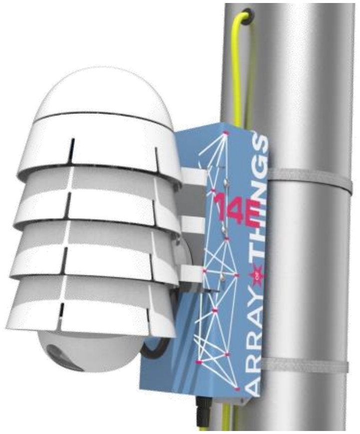
Computational flood and contaminant modeling

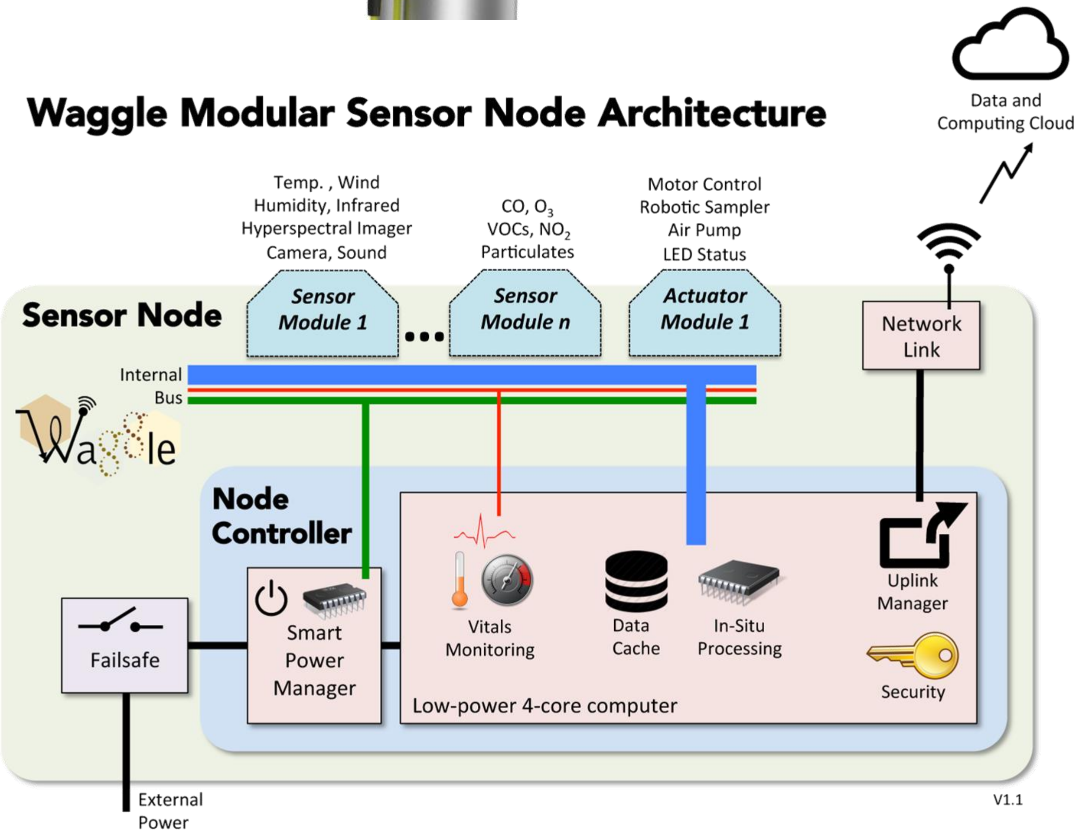


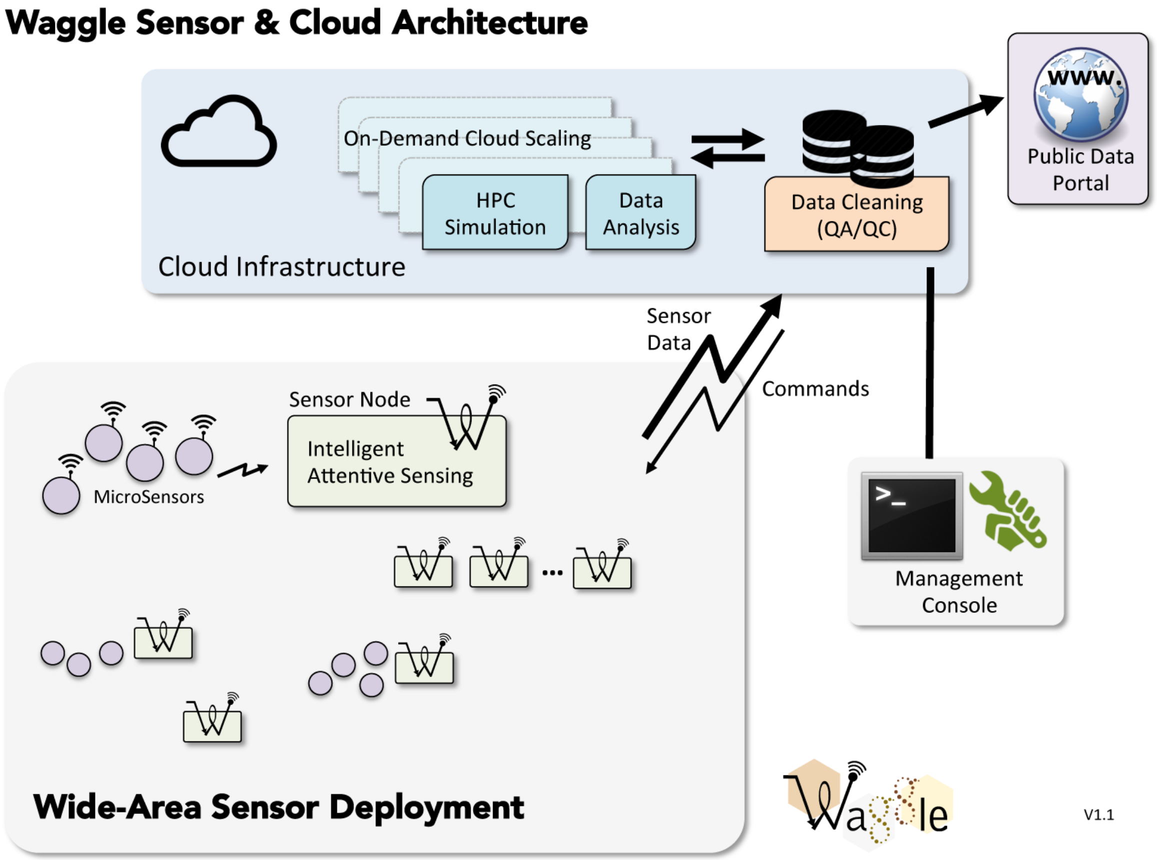
Machine learning and prediction

The University of Iowa

Argonne's Waggle toolkit for the Array of Things

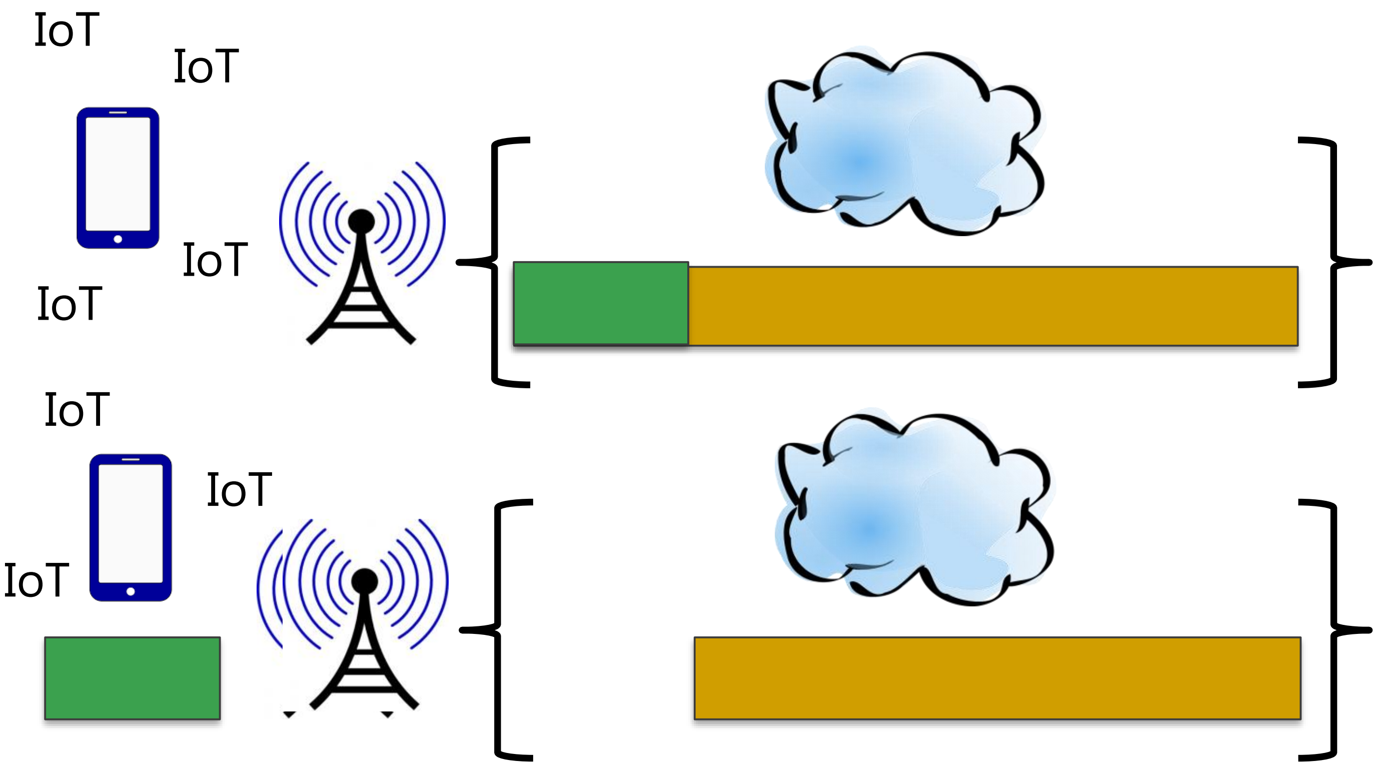


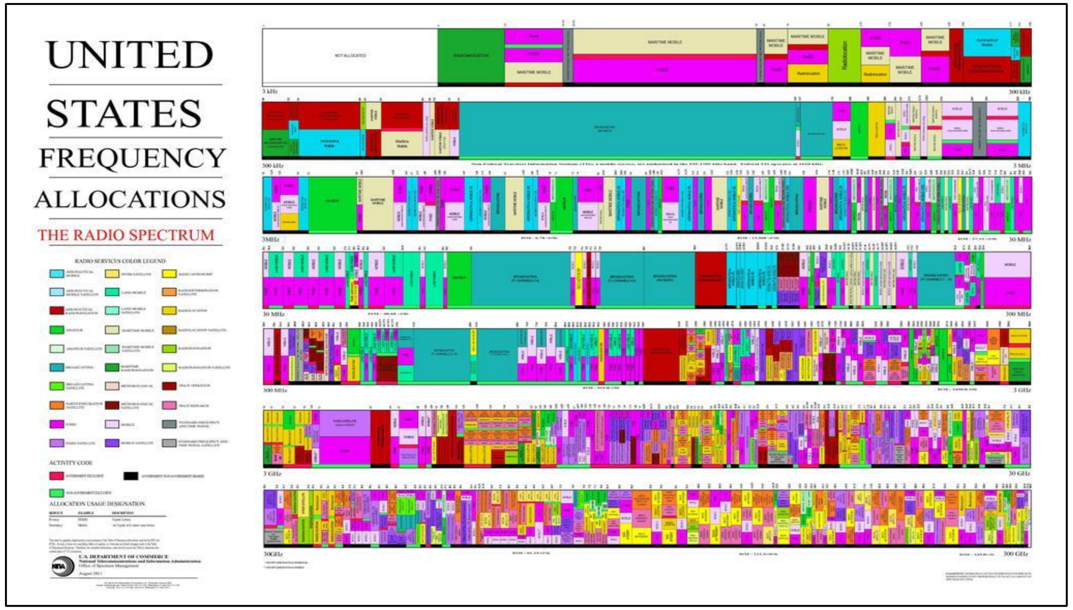




The University of Iowa

End-to-end multivariate optimization





Latency
Bandwidth
Energy

Storage
Knowledge
Context


The University of Iowa
Office of the Vice President for
Research &
Economic Development
RESEARCH.UIOWA.EDU
RESEARCH
DISCOVERY
INNOVATION
Dare to Discover

Discussion


The University of Iowa
Office of the Vice President
for Research & Economic Development
Dare to
Discover
RESEARCH.UIOWA.EDU
RESEARCH
DISCOVERY
INNOVATION
Dare to Discover