Machine Learning and Decision Making for Sustainability

Stefano Ermon
Department of Computer Science
Stanford University

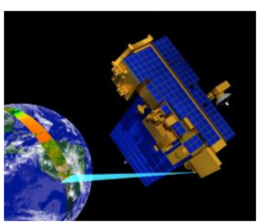
April 12

Overview

Stanford Artificial Intelligence Lab

Big Data

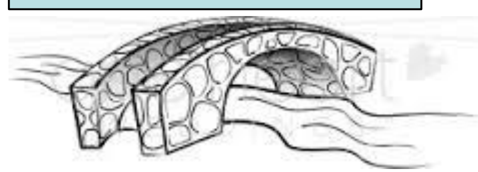
Technology Push



Sensing revolution

Artificial Intelligence

Computational Sustainability



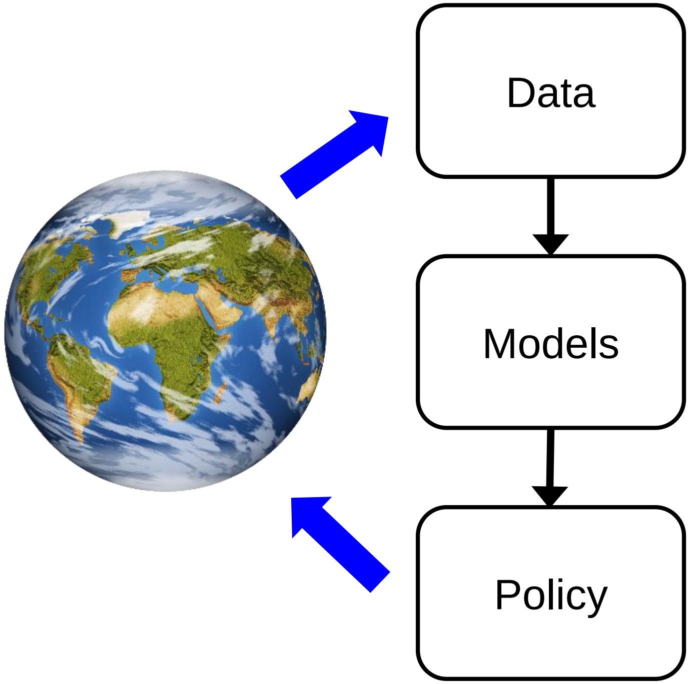
Fellow, Woods Institute for the Environment

Society Pull

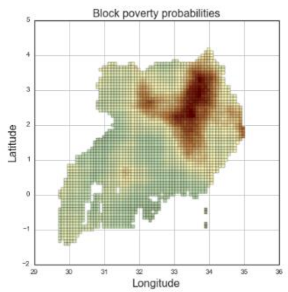
ML and Decision Making for Sustainability

Vision: sustainability challenges as control problems

Algorithmic challenges and opportunities at every step
- Data acquisition and interpretation
- Model fitting
- Decision making and policy optimization



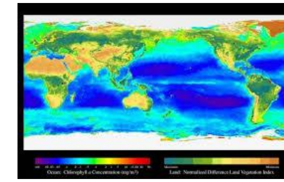
Computational Sustainability



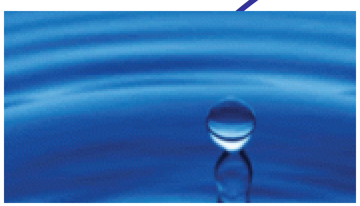
Poverty mapping

Poverty traps

Decision making and optimization



natural resources management

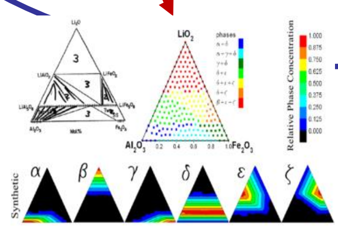


Water and weather systems modeling

Optimization of energy systems

Large unstructured datasets

Machine Learning



Materials discovery for energy applications

Summary

- Introduction
- Machine Learning for Public Policy
- AI for Sustainable Energy
- Conclusion

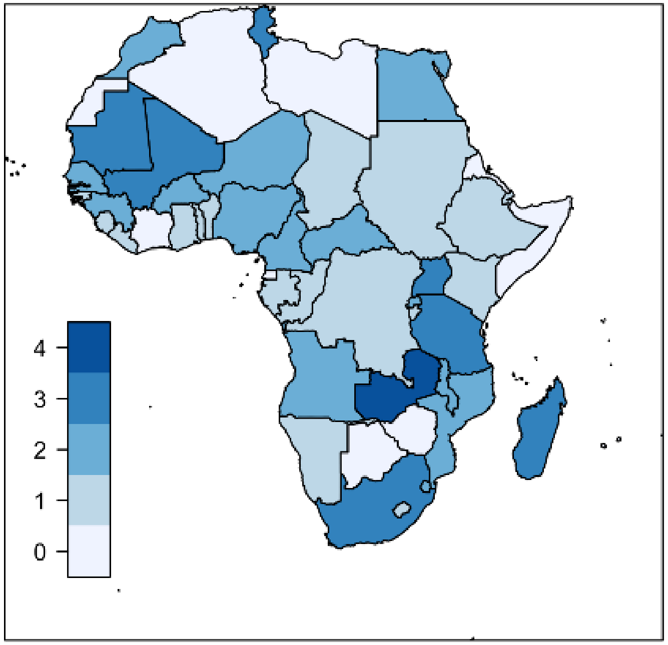
UN’s Global Goals for Sustainable Development

The 2030 Development Agenda (*Transforming our world)*

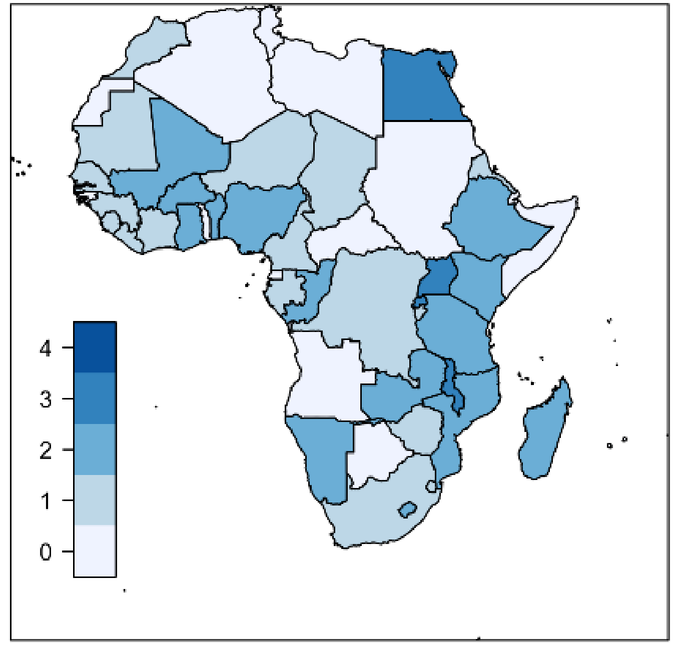
1. End extreme poverty
2. Fight inequality & injustice
3. Fix climate change

Data scarcity

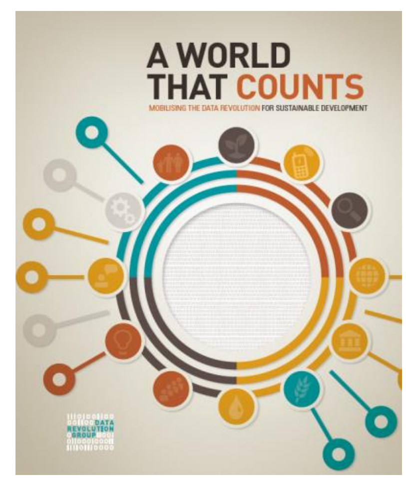
Consumption/Income Survey Availability, 2000-2010



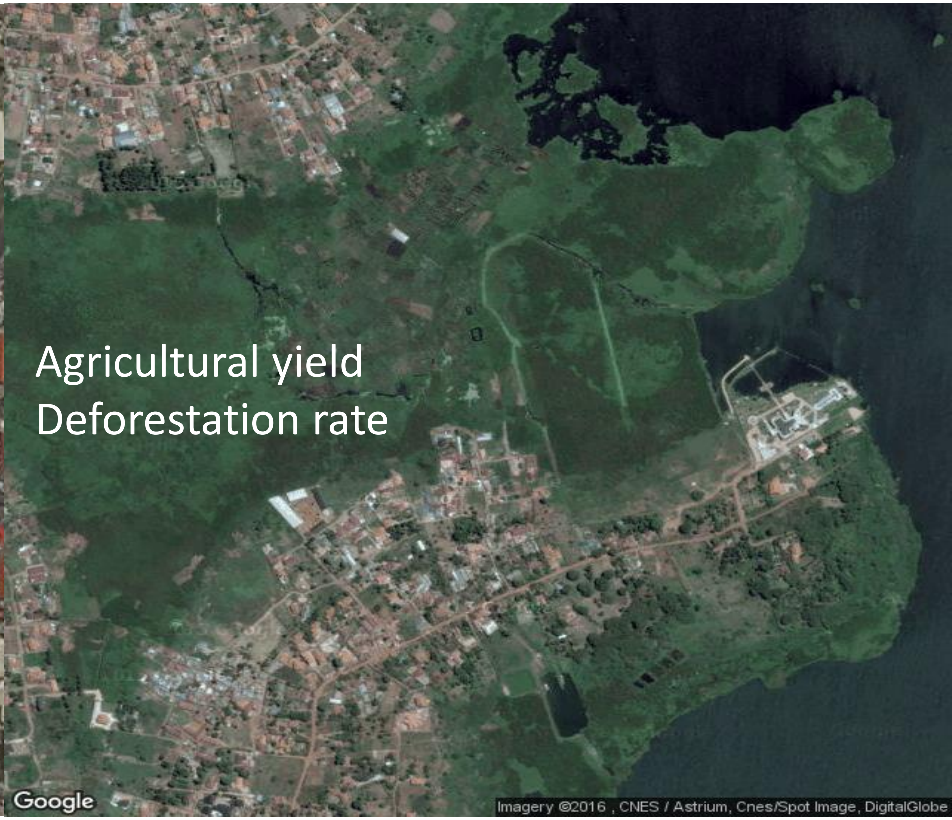
Wealth Survey Availability, 2000-2010



- **Expensive** to conduct surveys
- **Poor** spatial and temporal **resolution**
- **Questionable** data quality



Satellite imagery is low-cost and globally available



Simultaneously becoming **cheaper** and **higher resolution**
(DigitalGlobe, Planet Labs, Skybox, etc.)

What if…

we could **infer** socioeconomic indicators from large-scale, remotely-sensed data?

Standard supervised learning won’t work

Input

Output

Model

Poverty, wealth,
child mortality, etc.

- **Lots** of unlabeled data (images)
- Very little **labeled training** data (few thousand data points)
- Nontrivial for humans (hard to crowdsource labels)

Transfer learning overcomes data scarcity

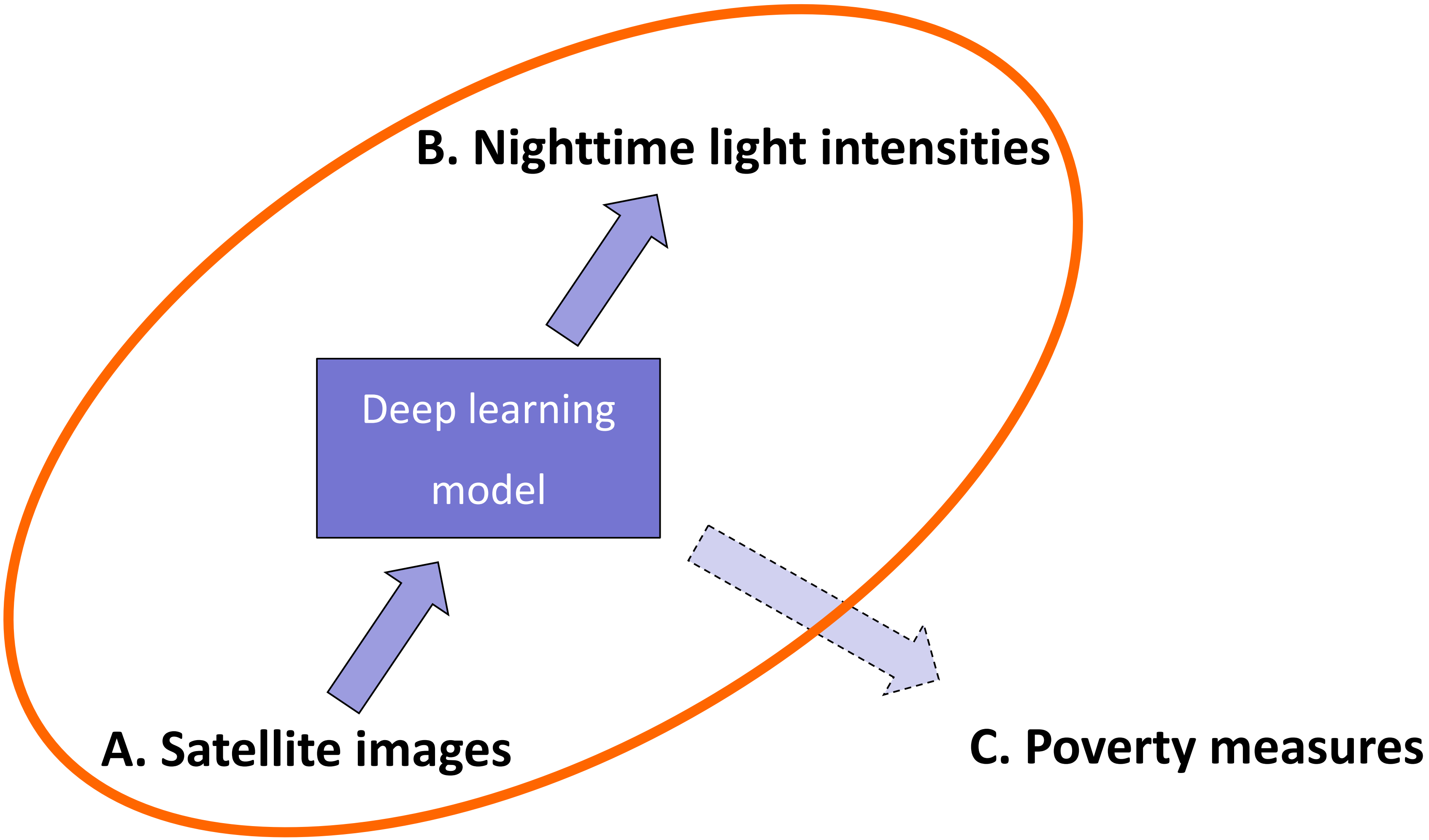
Transfer learning: Use knowledge gained from one task to solve a different (but related) task

Train here

Perform here

Nighttime lights as proxy for economic development

Step 1: Predict nighttime light intensities



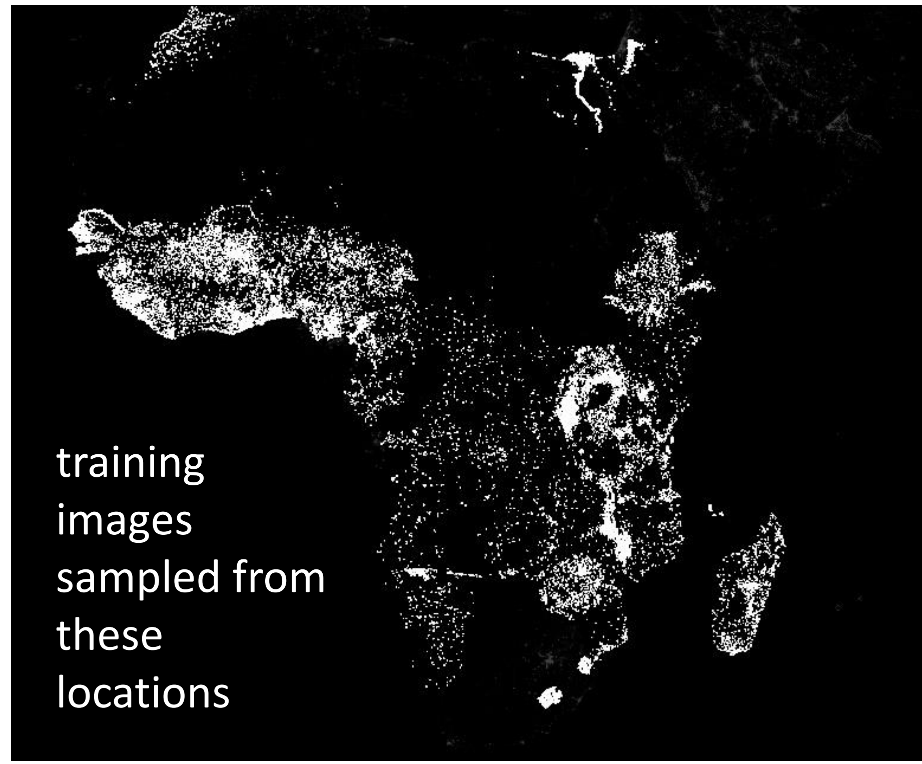
Training data on the proxy task is plentiful

Labeled input/output training pairs

( , Low nightlight intensity)

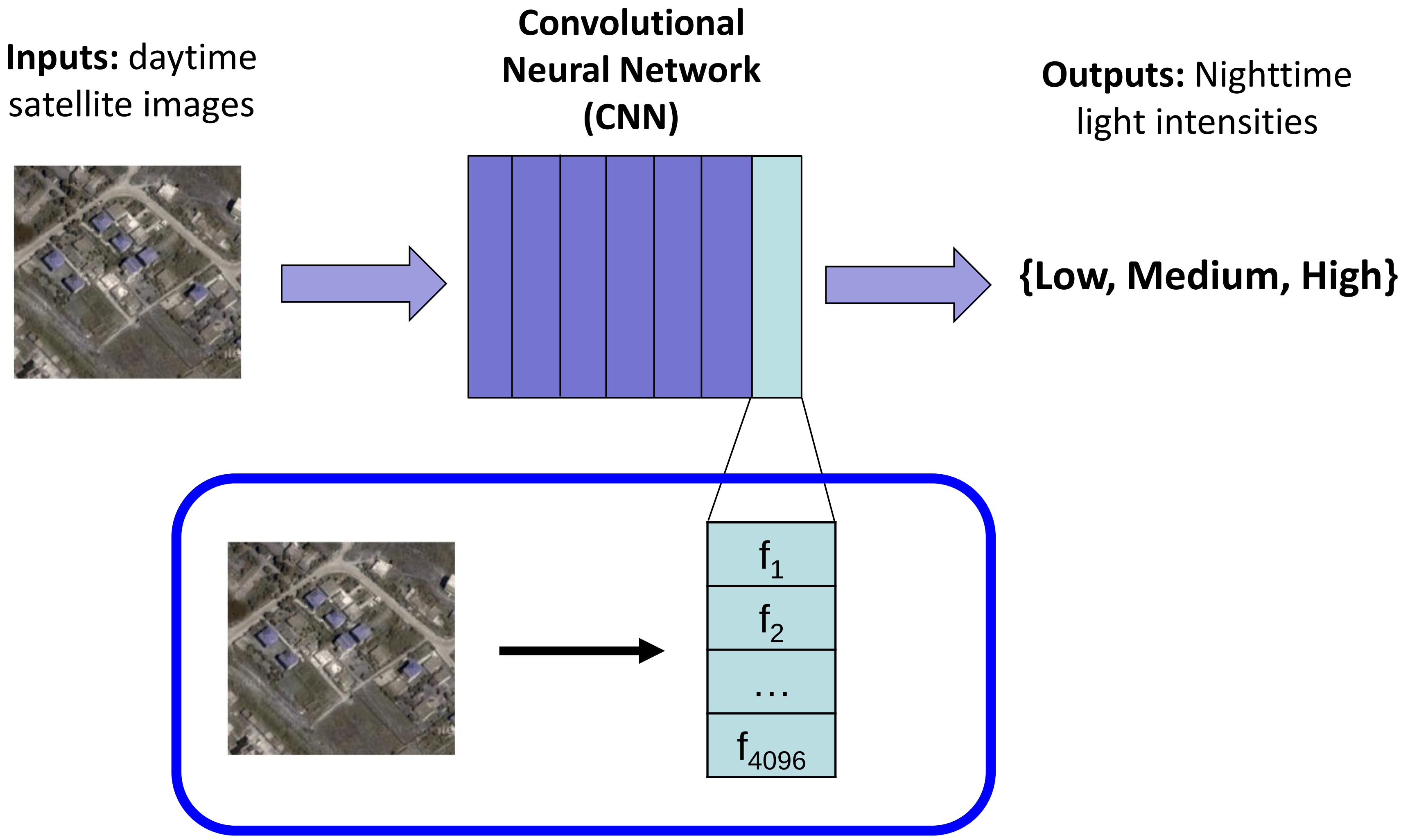
…

( , High nightlight intensity)



Millions of training images

Images summarized as low-dimensional feature vectors



Model learns relevant features automatically

f_1 f_{10}

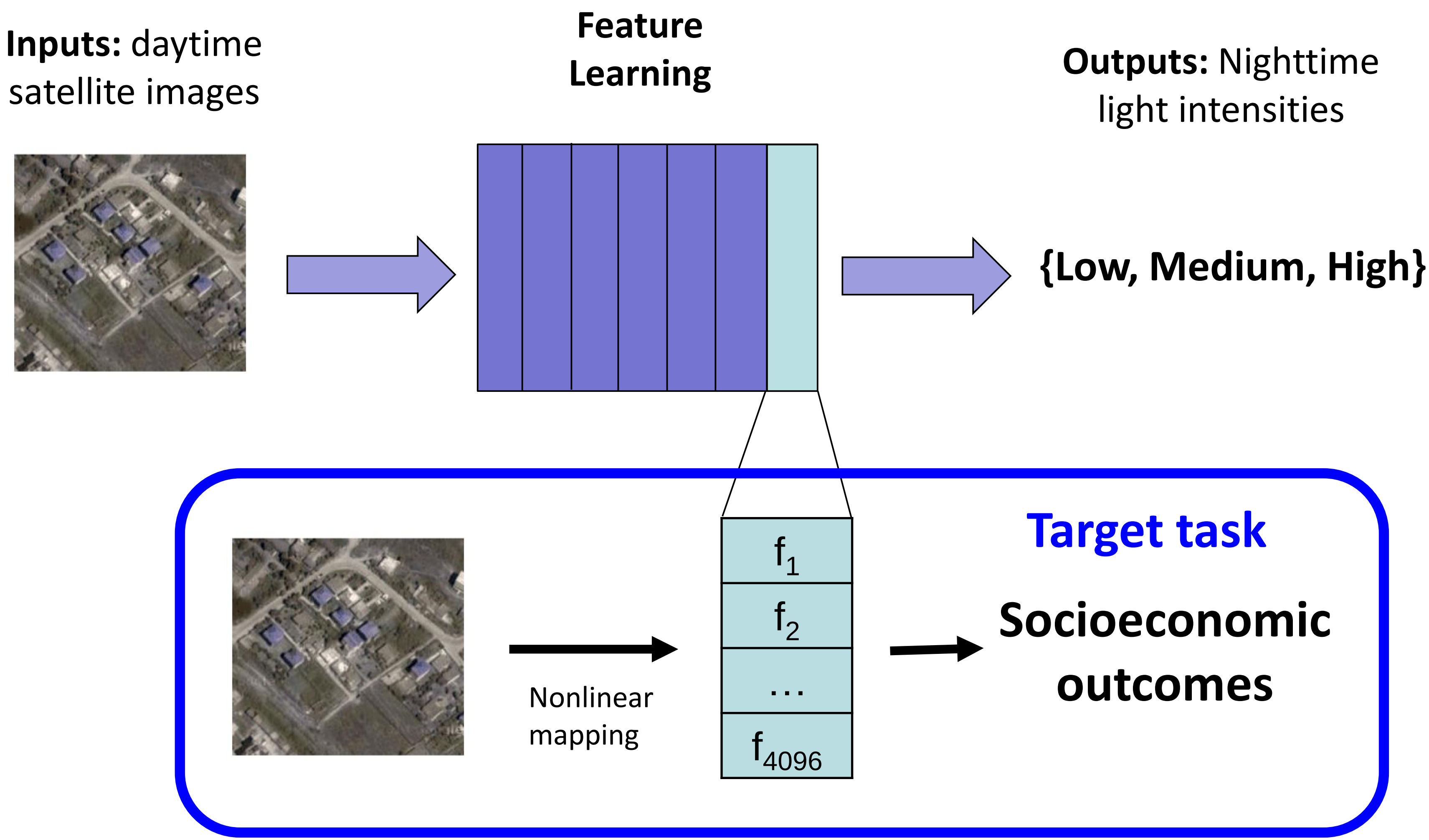
Satellite image

Filter activation map

Overlaid image

No supervision beyond nighttime lights - no labeled example of what a road looks like was provided!

Transfer Learning

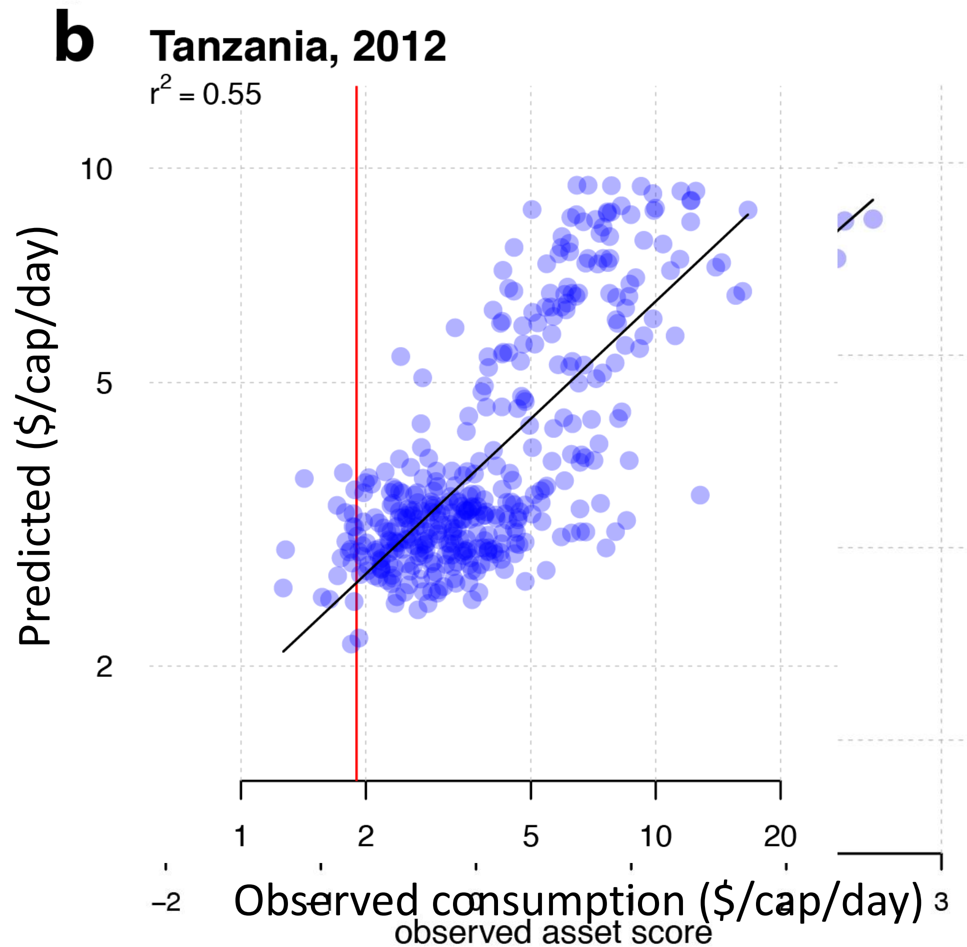


We can differentiate different levels of poverty

2 indicators:

- Consumption expenditures

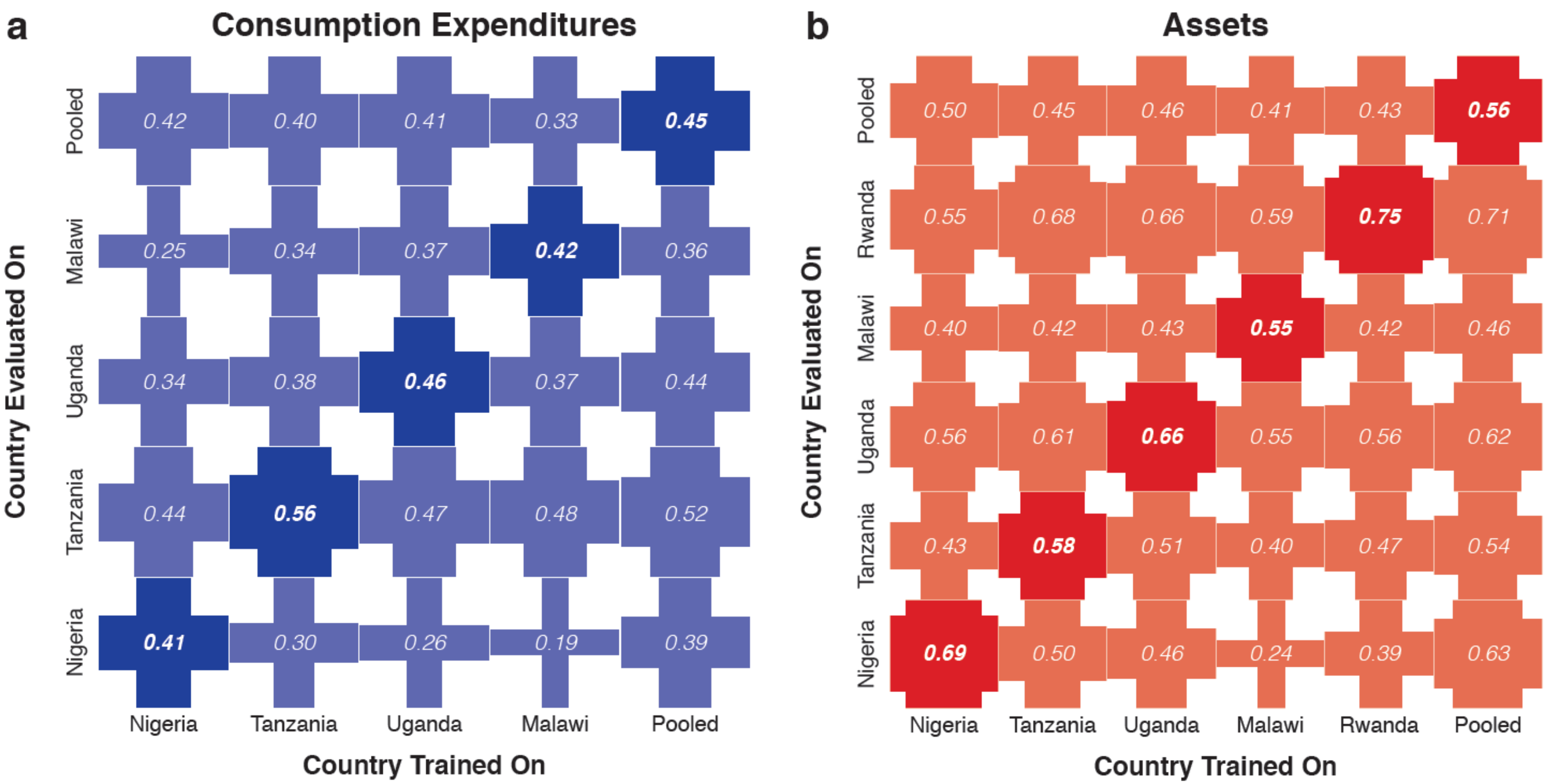
- Household assets



We outperform recent methods based on mobile call record data

Blumenstock et al. (2015) Predicting Poverty and Wealth from Mobile Phone Metadata, *Science*

Models travels well across borders



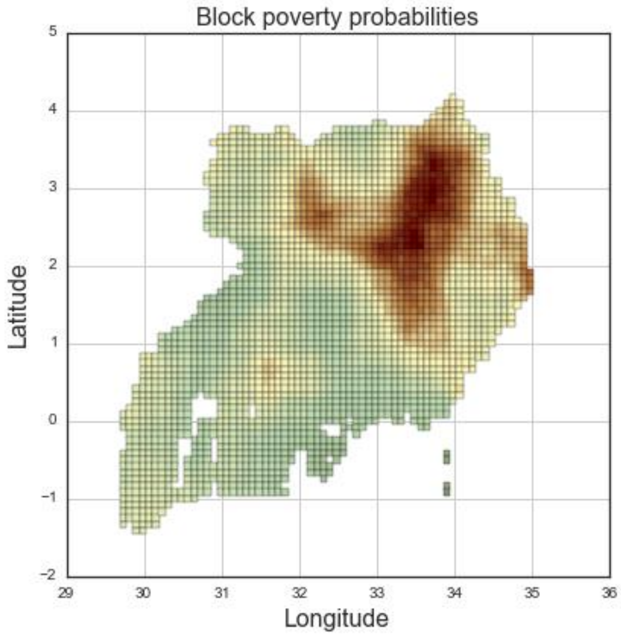
Models trained in one country perform well in other countries

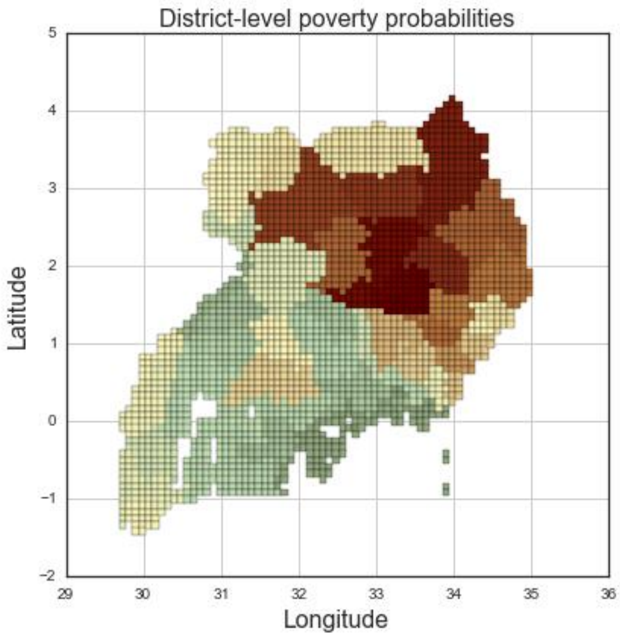
→

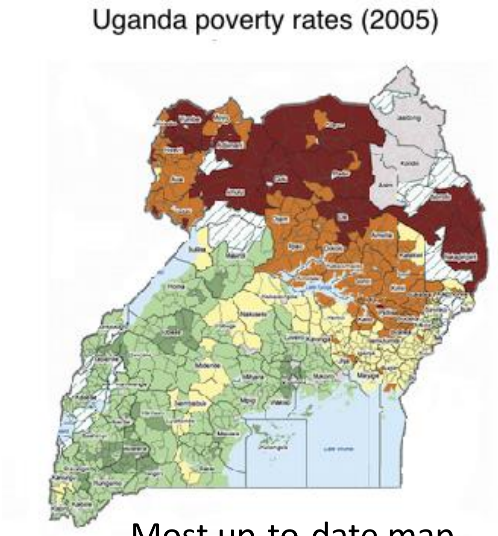
Can make predictions in countries where no training data exists

Scalable High Resolution Poverty Maps

Run the model on about 500,000 images from Uganda:







Most up-to-date map

Scalable and inexpensive approach to generate high resolution maps.

TheUpshot

The New York Times

Satellite Images Can Pinpoint Poverty Where Surveys Can't

Economic View

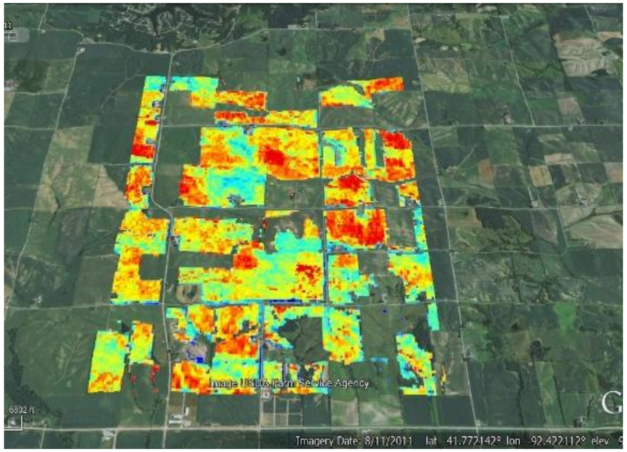
By SENDHIL MULLAINATHAN APRIL 1, 2016

Ongoing work

- Describe, model, and predict **changes over time**

- Incorporate **new data sources** (phone data, crowdsourcing, etc.)

Credit: premise.com

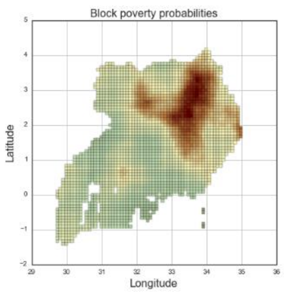


- Mapping and estimating **crop yields**
  - 1st prize at INFORMS yield prediction challenge

Summary

- Introduction
- Machine Learning for Public Policy
- AI for Sustainable Energy
- Conclusion

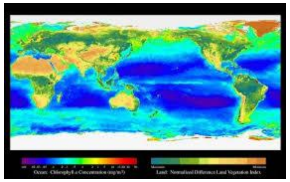
Computational Sustainability



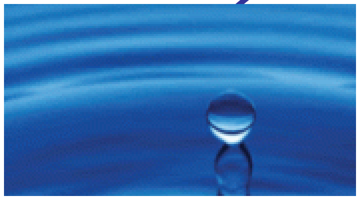
Poverty mapping

Poverty traps

Optimization



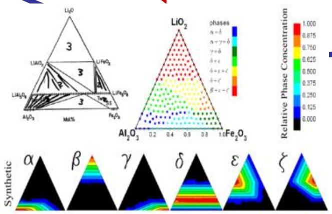
natural resources management



Groundwater and weather systems modeling

Optimization of energy systems

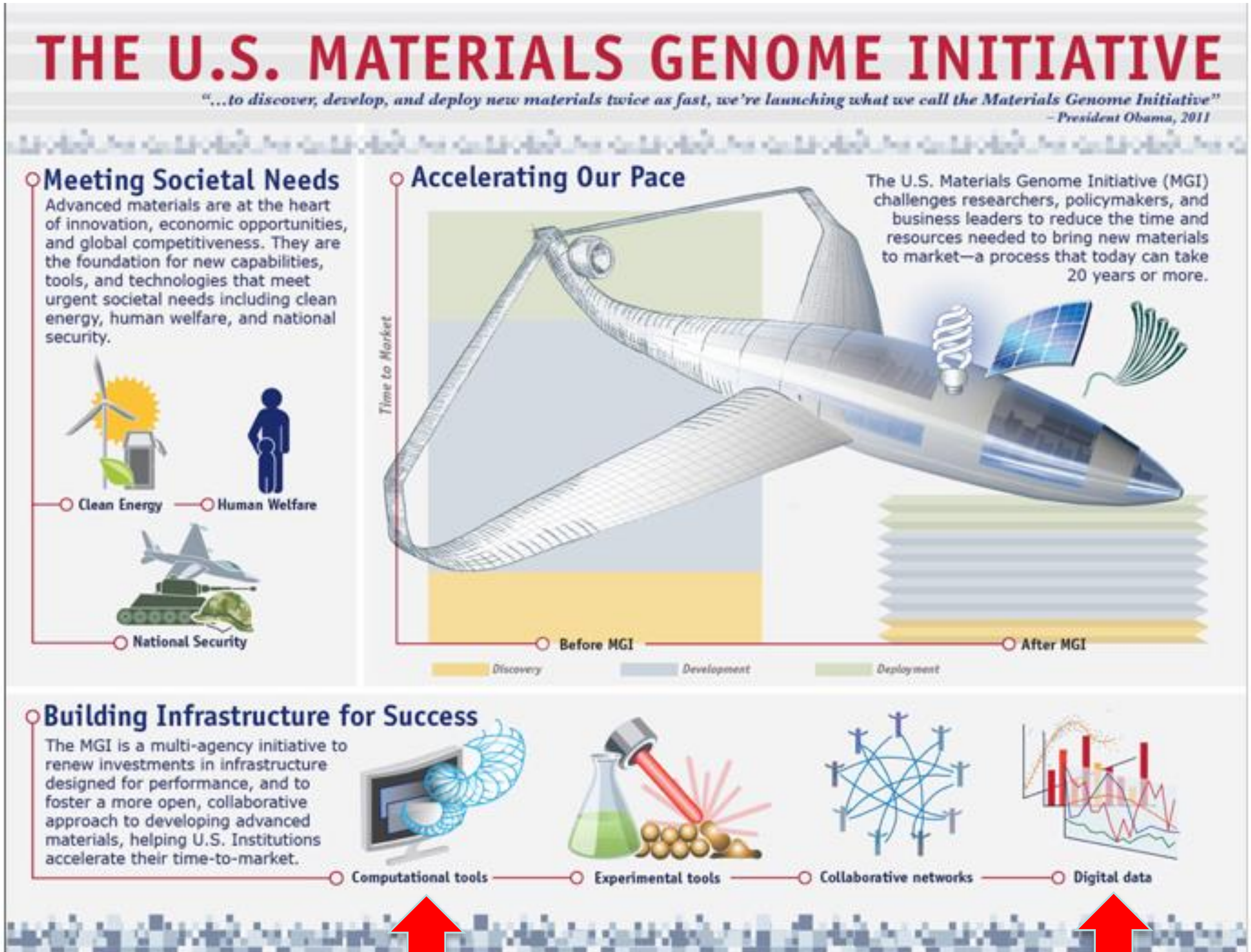
Large Datasets



Energy Materials discovery

Artificial Intelligence and Machine Learning

White House Materials Genome Initiative



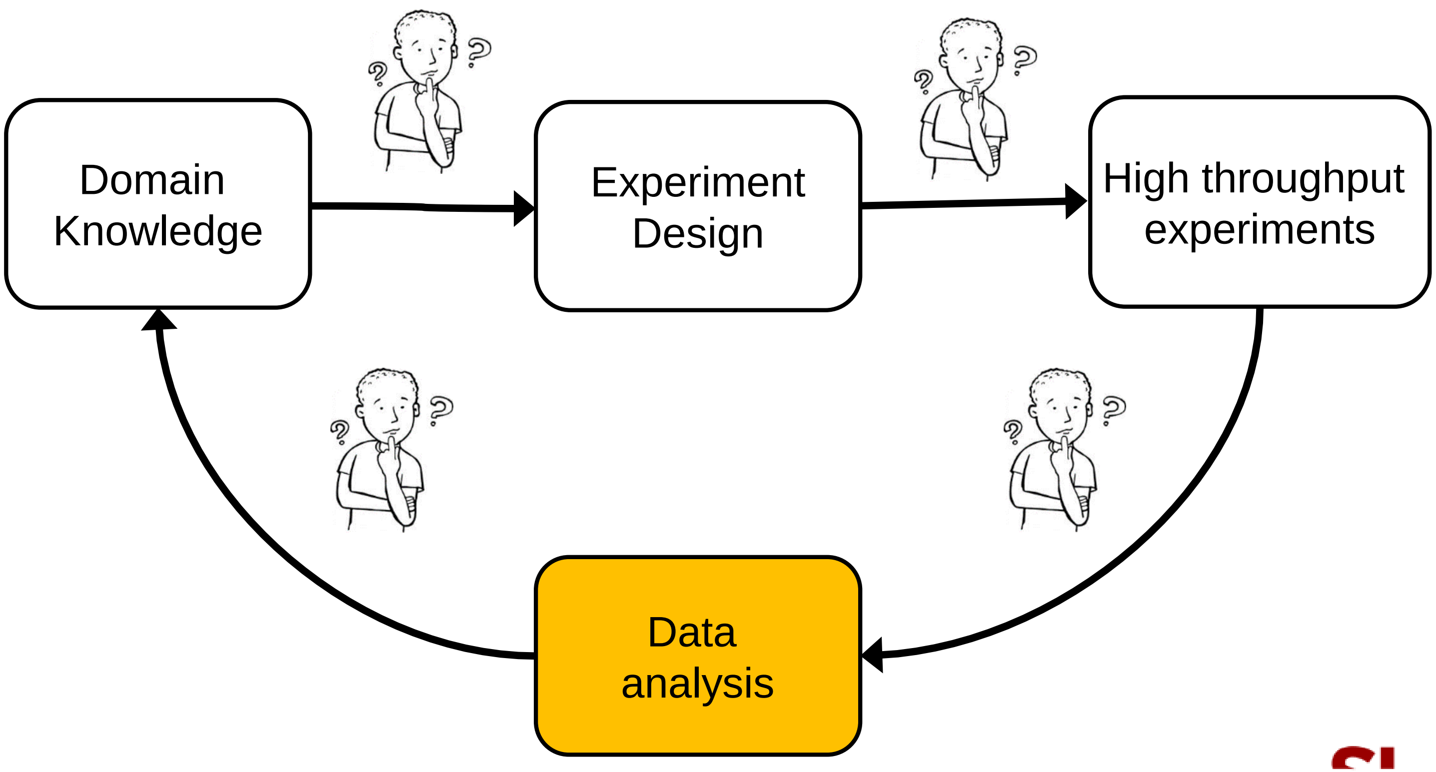
<u>Goal</u>

Accelerate the pace and reduce the cost of discovery, and deployment of advanced material systems

20 years → 5 years

Very exciting new research area for Computer Science and Big Data techniques

Vision: AI for materials research



Automatic Data Analysis

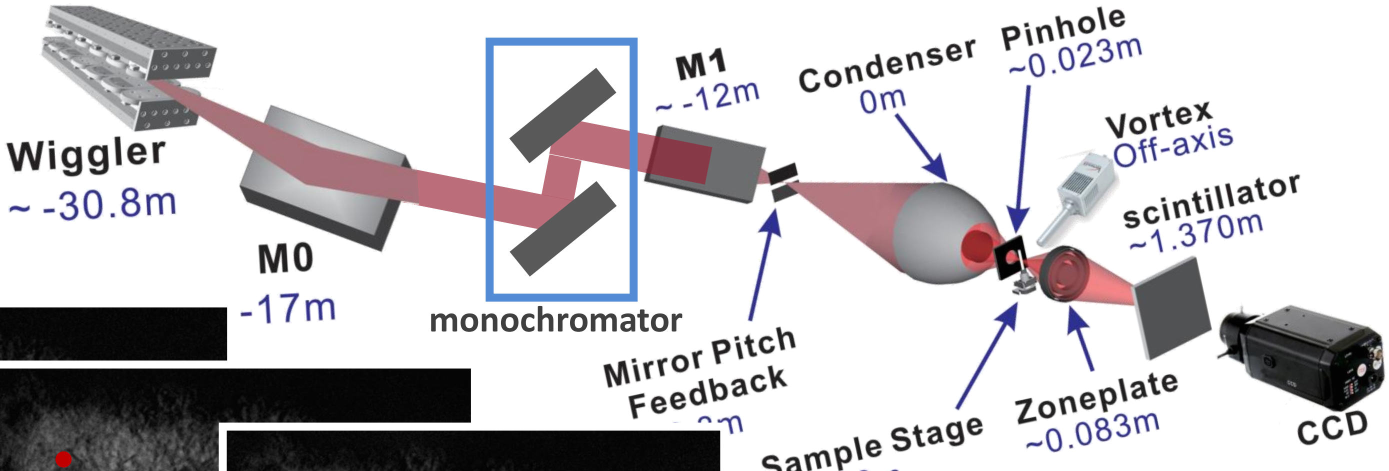
SLAC
Stanford Linear Accelerator

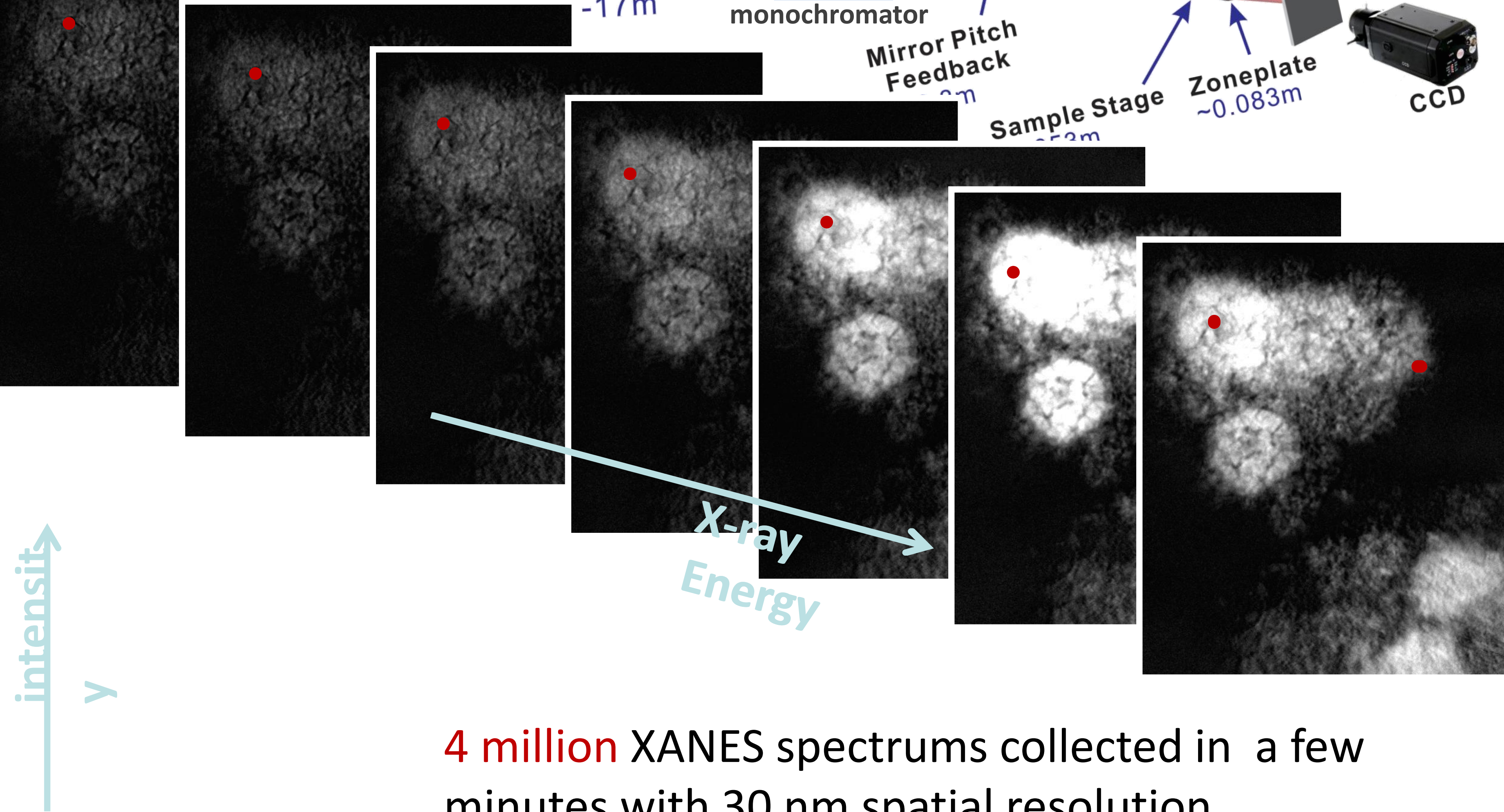
CHESS
Cornell High Energy Synchrotron Source

emc2
Energy Materials Center at Cornell

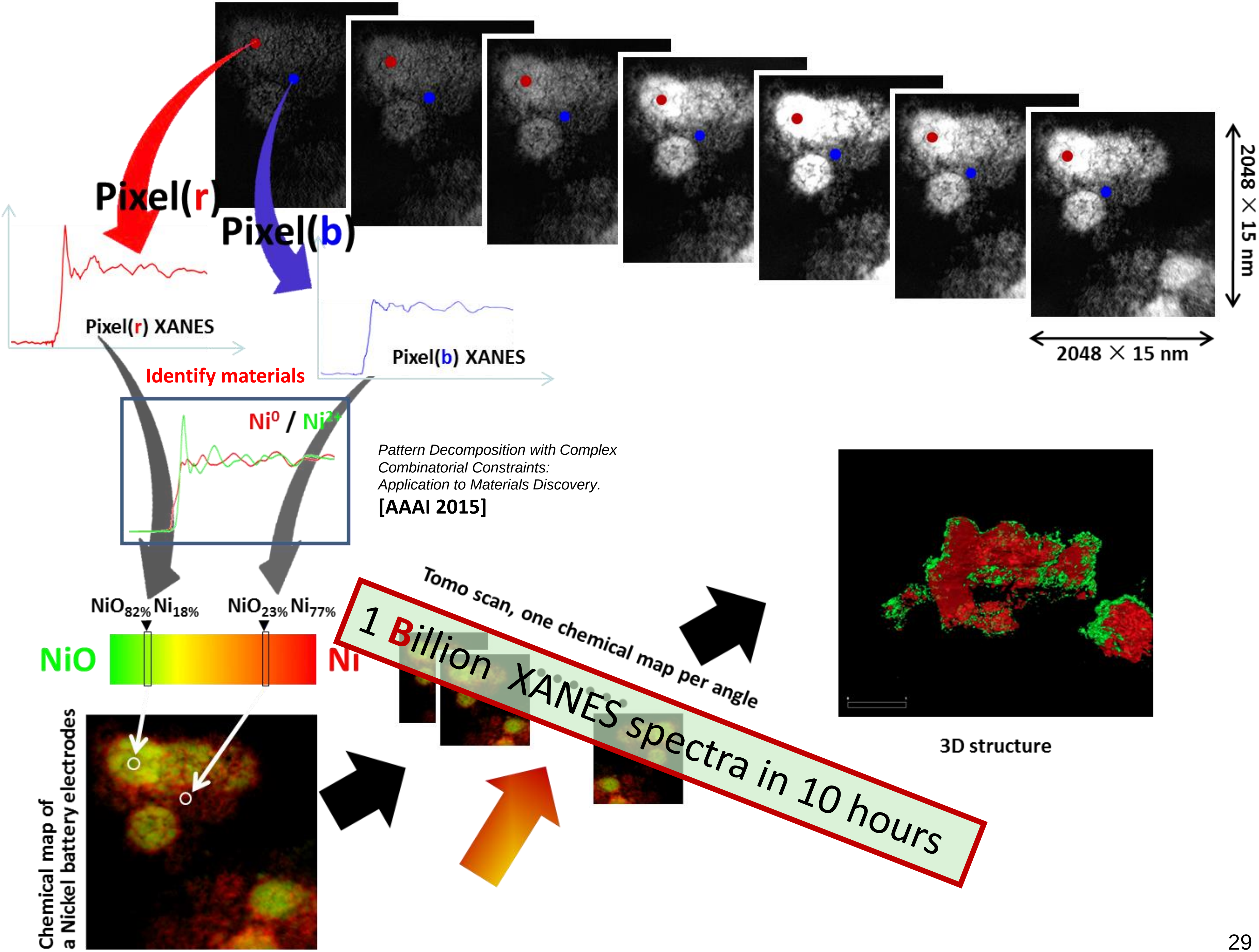
JCAP
JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS
Caltech

Slide courtesy of
Apurva Mehta and Yijin Liu,
SLAC

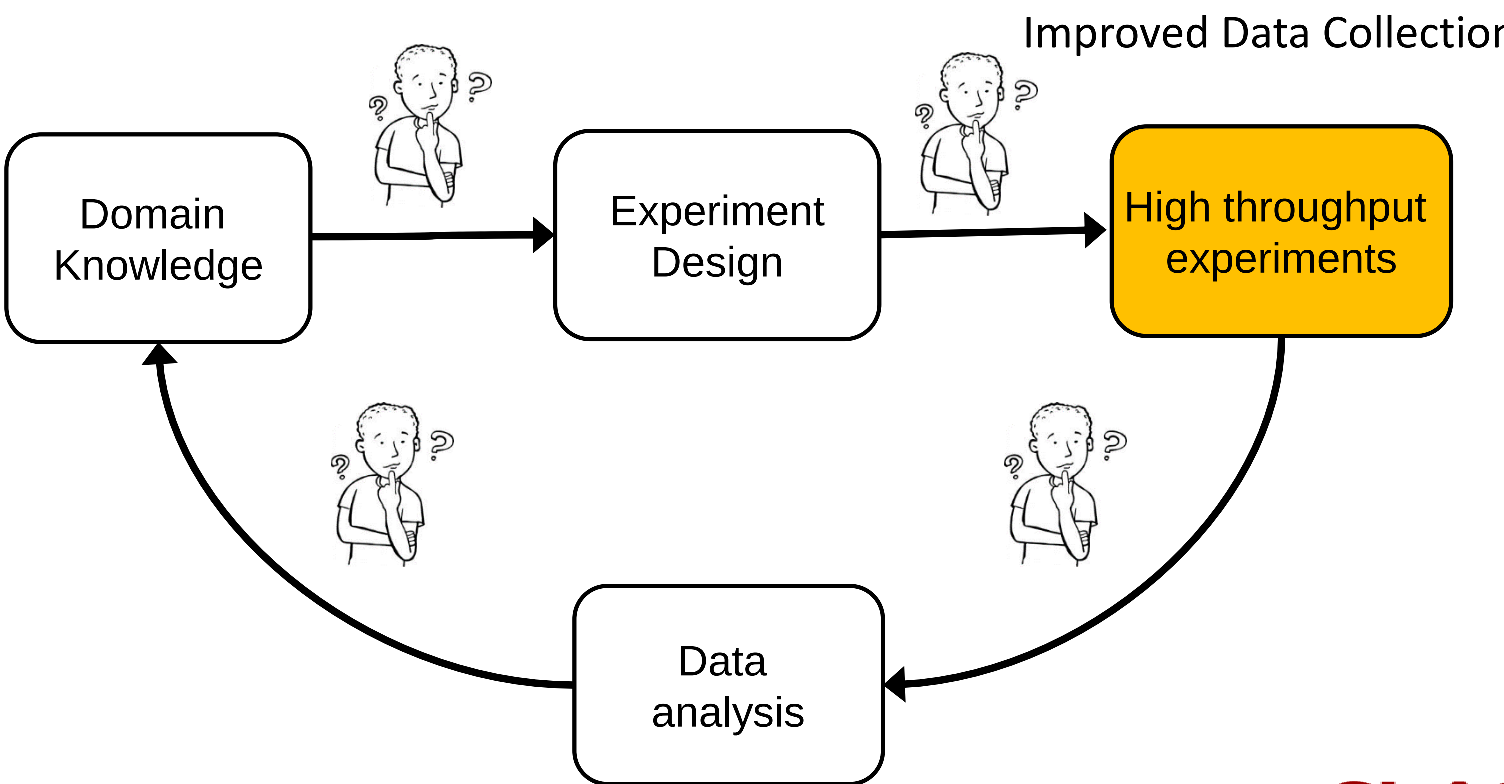




4 million XANES spectrums collected in a few minutes with 30 nm spatial resolution.


Pixel(r)
Pixel(b)
Pixel(r) XANES
Pixel(b) XANES
2048 × 15 nm
2048 × 15 nm
Identify materials
Ni0 / Ni2+
Pattern Decomposition with Complex
Combinatorial Constraints:
Application to Materials Discovery.
[AAAI 2015]
NiO82% Ni18%
NiO23% Ni77%
NiO
Ni
Chemical map of
a Nickel battery electrodes
Tomo scan, one chemical map per angle
1 Billion XANES spectra in 10 hours
3D structure

Vision: AI for materials research



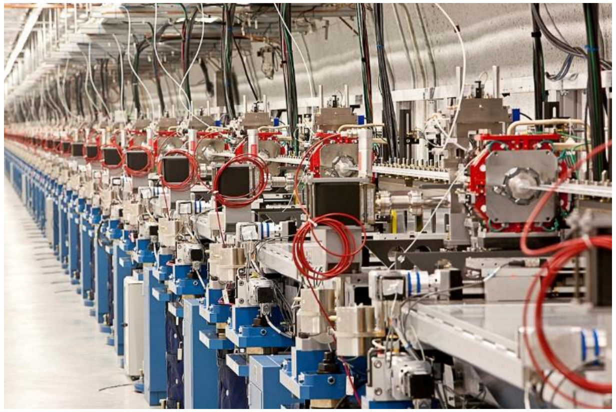
CHESS
Cornell High Energy Synchrotron Source

emc2
Energy Materials Center at Cornell

JCAP
JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS
Caltech

SLAC
Stanford Linear Accelerator

LCLS tuning at SLAC



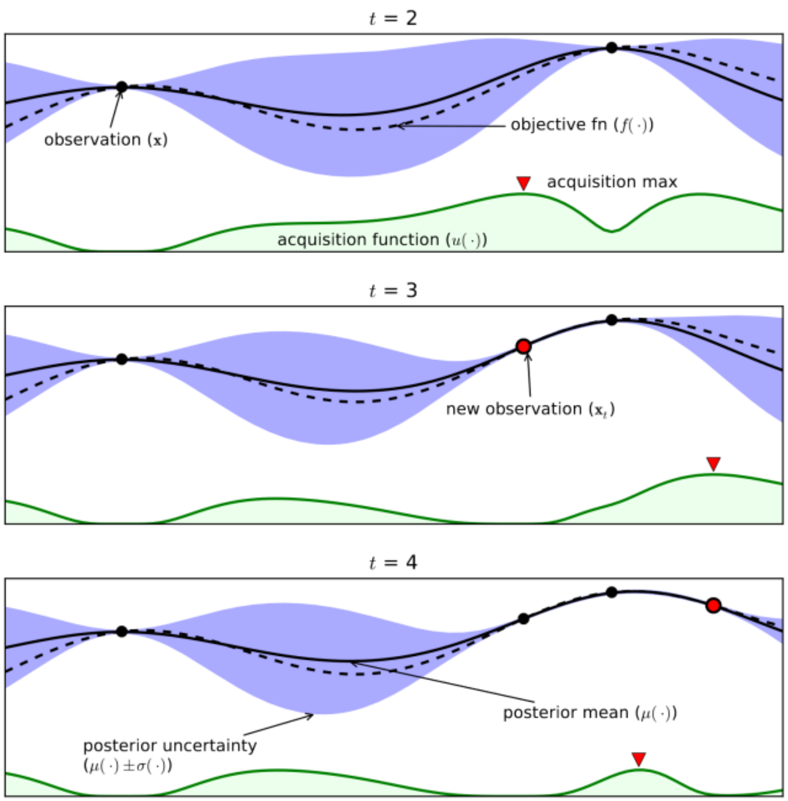
Linac Coherent Light Source (LCLS) is the world's first X-ray laser.
10 billion times brighter than any other X-ray source before it

Very complex machine, difficult to operate, requires **manual tuning (hundreds of hours per year)**

Operating cost close to **$1,000 per minute** – want to make parameter tuning as robust and as quick as possible

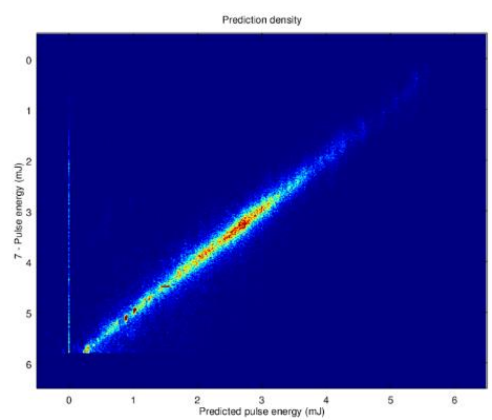
Bayesian Optimization for LCLS

Archiving system: records almost 200,000 independent variables once a second, and goes back several years



Bayesian optimization:

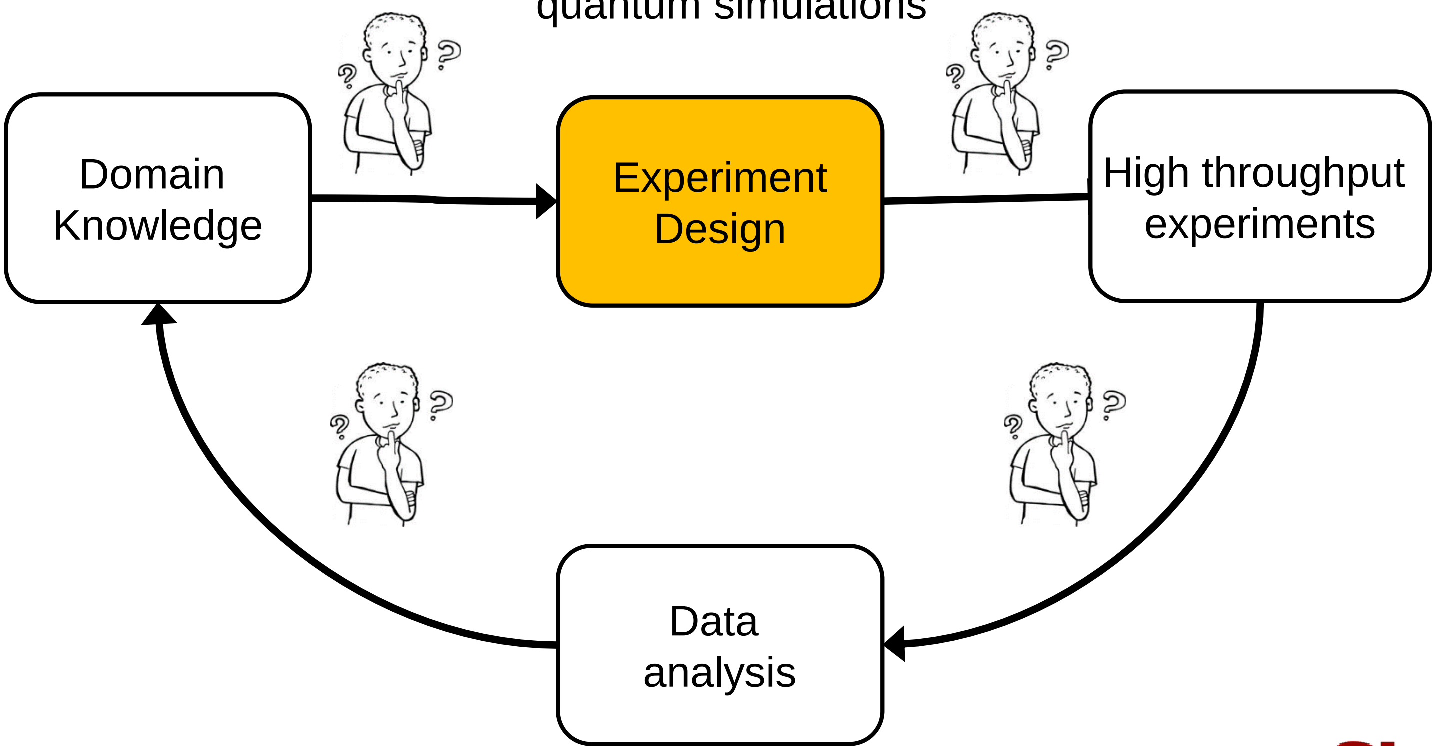
- Works by seeking promising points that aren't already explored
- Sound way to deal with the classic **exploration** vs **exploitation** tradeoff



Sparse Gaussian Processes for Bayesian Optimization
[under review at UAI-16]

Vision: AI for materials research

Preliminary work on dieletric screening via quantum simulations



Stanford Linear Accelerator

Cornell High Energy Synchrotron Source

Energy Materials Center at Cornell

Caltech

Summary

- Introduction
- Machine Learning for Public Policy
- AI for Sustainable Energy
- Conclusion

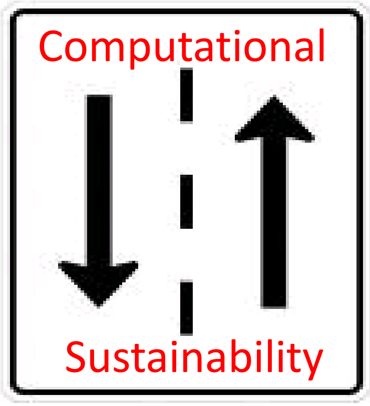
Conclusions

- Growing concerns about the threats of Artificial Intelligence to the future of humanity

- Recent advances in AI also create **enormous opportunities for having deeply beneficial influences on society** (energy, sustainability, …)

- Exciting opportunities for Computer Science research

Sustainability Sciences



Computational Sciences