Scientific Methods in the Era of Big Data and Machine Learning

Ruzena Bajcsy

Zoe Cohen, Isabella Huang, Carolyn Chen, Laura Hallock, Sarah Seko, Robert Matthew

2019.07

HART Lab

Human-Assistive Robotic Technologies

Berkeley
UNIVERSITY OF CALIFORNIA

Talk Roadmap

I. **Definitions**: The Scientific Method, Laws, & Theories

I. **Science in the Era of “Big Data”**

I. **Modeling & Data in the Bajcsy Lab**
   - *Haptic Perception of Liquids Enclosed in Containers*
   - *SOFTCell: A Depth-Camera-Based Soft Fingertip Device*
   - Human modeling overview

I. **Summary & Conclusions**

SECTION I

Definitions: The Scientific Method, Laws, & Theories

Definition: The Scientific Method

The **scientific method** [from Webster/Oxford dictionary] is a method of research in which:

- the problem is **identified**
- relevant data are gathered
- a **hypothesis** is formatted from the data/observation
- the hypothesis is **empirically tested**

The classical model of scientific inquiry comes from Aristotle (344-322 BCE), who distinguished **approximate** and **exact** reasoning. He set out the threefold schema of **abductive**, **deductive**, and **inductive** inference, as well as reasoning by **analogy**.

Bayesian inference is an example of inductive reasoning.

Formal Theory

Again according to Aristotle, **theory** is trying to understand a phenomenon and explain its nature. Theory remains a hypothesis until it is **tested** and **verified** by experimental evidence.

Formal theory is embedded in mathematical logic. It has a **syntax** and a given **semantic** interpretation.

The **truth** is relative to the whole theory. A special case is **axiomatic theory**, which has **axioms** (assumptions) and rules of **inference**. A **theorem** is a statement that can be derived solely from these axioms. (Examples include arithmetics, geometry, and probability.)

Model theory is the study of classes of mathematical structures that obey the rules of mathematical logic.

Theories vs. Theorems vs. Law

A **theorem** is derived deductively from **axioms** (basic assumptions) following the formal syntax of rules within the **theory**.

A **law** is a generalized **statement** of some functions/relations based on many observations.

Example:

- Newton (1687) published the gravity hypothesis. This **law** was put to the test by different scientists in a study of planetary motion and after many tests, the hypothesis became the gravitational **theory**.
- Two hundred years later Einstein developed the **theory of relativity**, which encompasses the **gravitational theory**; hence, **laws** must be inclusive.

Summary: Theory and Law

- A **theory** is an explanation of some observation.
- Theory is the reasoning behind a **law**.
- Physical laws, or “laws of nature”, are statements inferred from **measurable facts** applicable to a class of phenomena.
- Laws are narrower in scope than **scientific theories**.

Some well known laws can be listed:

- Newton’s theories of classical mechanics
- Four Laws of Thermodynamics
- Einstein’s theory of relativity
- Boyle’s Law of gas

SECTION II

Science in the Era of “Big Data”

Observations

Observations depend on the apparatus that performs the observation exercise.

Hence, during the history of science, the sophistication of the **explanatory power** of the theory, laws, and/or models depended on the **capability** of the observing sensors, including range, spatial/temporal resolution, and sensitivity (signal-to-noise ratio).

Initially, we were restricted to our human **perceptual sensors**. Later, magnifying glasses/optics facilitated discoveries of Celestial movement, probing sources such as microwaves outside of the visual spectrum penetrated materials, and high energy accelerators enabled the discovery of atomic structures.

Observations & Measurements in Engineering & the Physical Sciences

It has been accepted in the scientific community that **observation** is more **informal** while **measurements** performed via scientific **instruments** are more **quantitative**.

The output of both observations and measurements is **data**.

In this presentation, I will focus on the engineering sciences.

DATA IS NOT REALITY and its credibility depends on several components:

- the quality and extent of sensors
- the scale (spatial, temporal, material, etc.) of the system

Controlling the Quality of Data

There exist some **standard methods** to control the quality of data:

- Select the sensors with spatial, temporal and signal-to-noise **capabilities** adequate to the problem at hand.
- Perform **calibration** on the sensors to have the the parameters of the interval (upper and lower bounds) of sensitivity and performance.
- Select the proper **sampling** in order to achieve the desired performance.
- Design a **controlled experimental test** for repeatability and as a measure of robustness.

Machine Learning & the Quality of Data

One cannot avoid using **machine learning** in today's climate of "big data".

Thanks to inexpensive hardware, today we can collect large data sets, which enables us to apply statistics and optimization techniques to discover **patterns** and perhaps some **laws** from these measurements.

Needless to say, that credibility and validity of these discoveries will very much depend on the **quality of the data**.

Scientific Discoveries as an Exploration

While I advocate **careful** (perhaps controlled) **experimental** data collection, ***this does not imply negation of exploration***.

History teaches us that **some theories came about because of exploratory experimentation**. For example, thermodynamics came about because of the experimentation of James Watt.

Similarly, mathematicians said the integration method used on the ENIAC machine were wrong and would never work, but Eckert (one of the engineer-designers of ENIAC) was able to use much smaller time intervals than had ever been used before, and the field of numerical integration and simulation was never the same.

Machine learning is in an **empirical phase**.

There is **still a lot of mystery** in why and how neural networks work, which causes some discomfort. However, deep learning **opens many new problems and ideas as opposed to refinements of classical results**. (Signal processing has been for many years extensions of the **Shannon-Weiner** framework.)

We should strive for **reproducibility** and **repeatability** of results that come from deep learning.

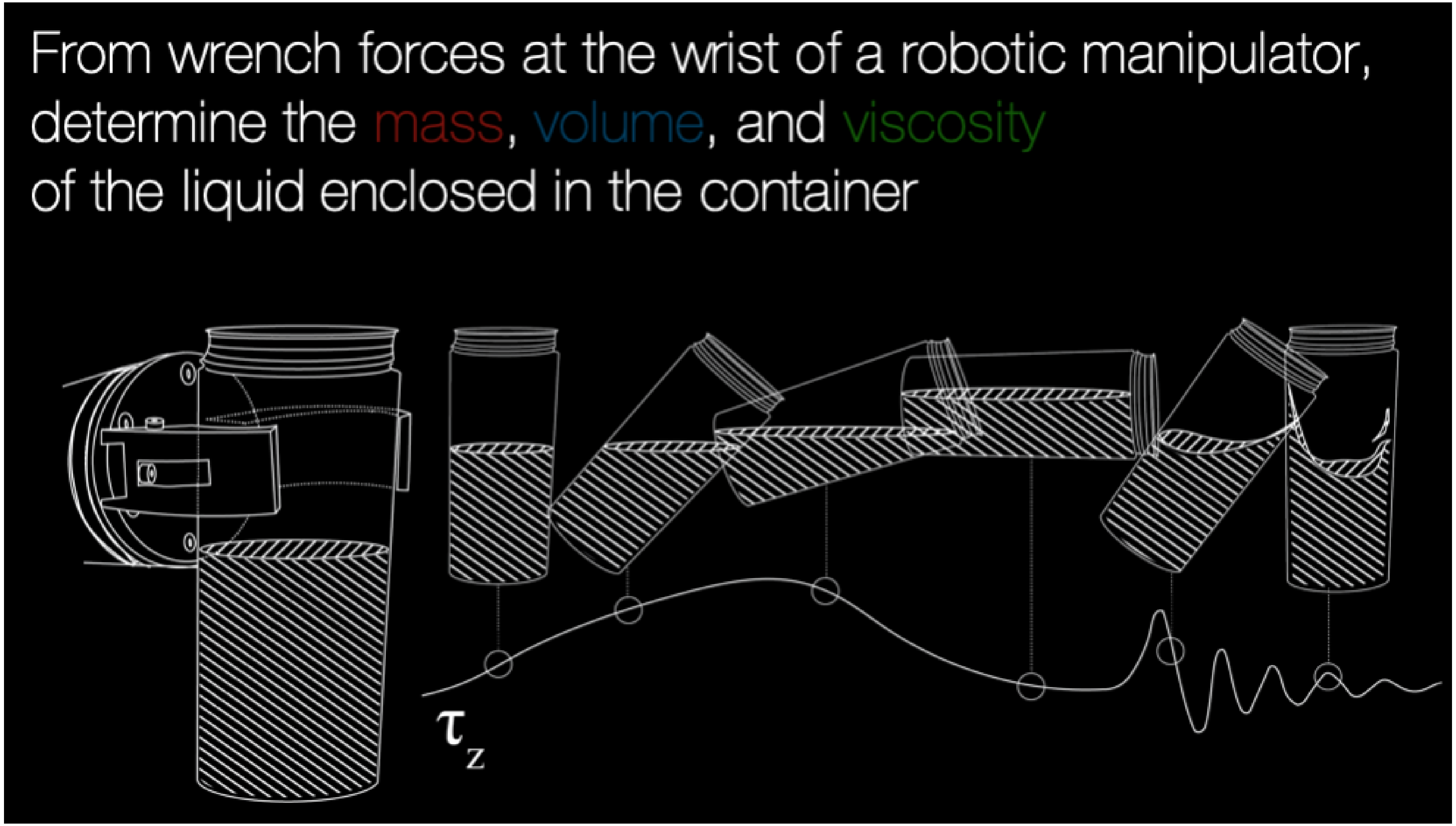
SECTION III

Modeling & Data in the Bajcsy Lab

Carolyn Chen Matl, Robert Matthew, Ruzena Bajcsy

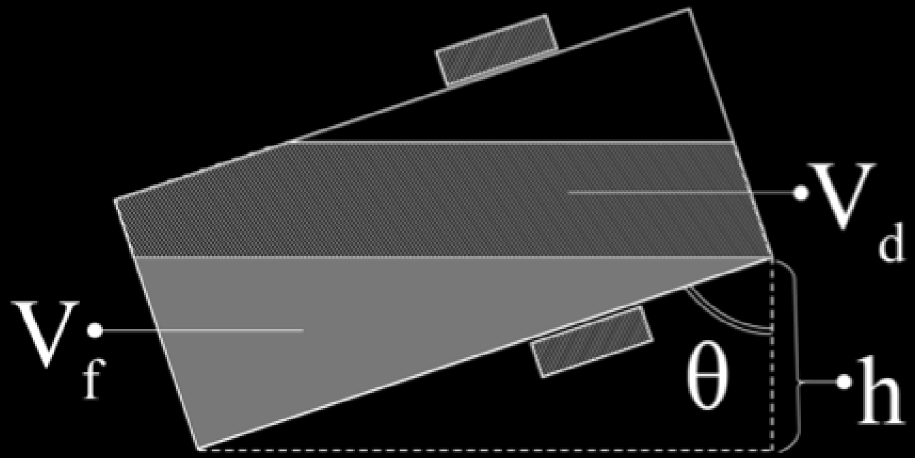
Haptic Perception of Liquids Enclosed in Containers

Objective


From wrench forces at the wrist of a robotic manipulator,
determine the mass, volume, and viscosity
of the liquid enclosed in the container
τ_z

Precision Pouring

Using the estimation of liquid volume, determine the angle at which to pour a precise volume of liquid



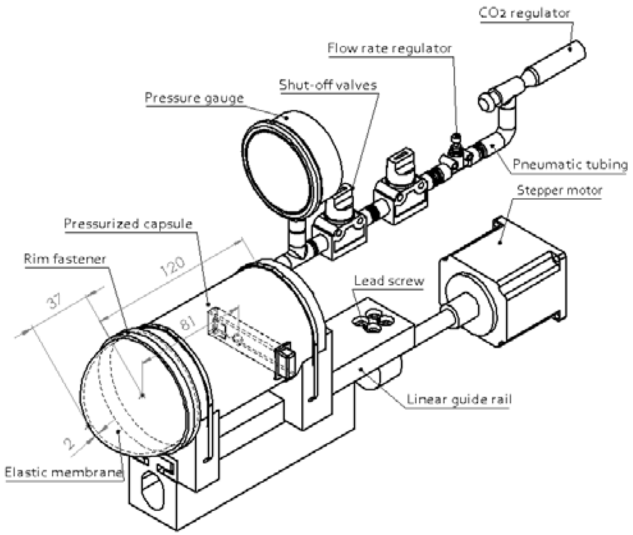
$$\theta = arg\ \min_{\theta\in[0,\frac{\pi}{2}]} ||V_d - (V - V_f(L\cos\theta)||_2^2$$

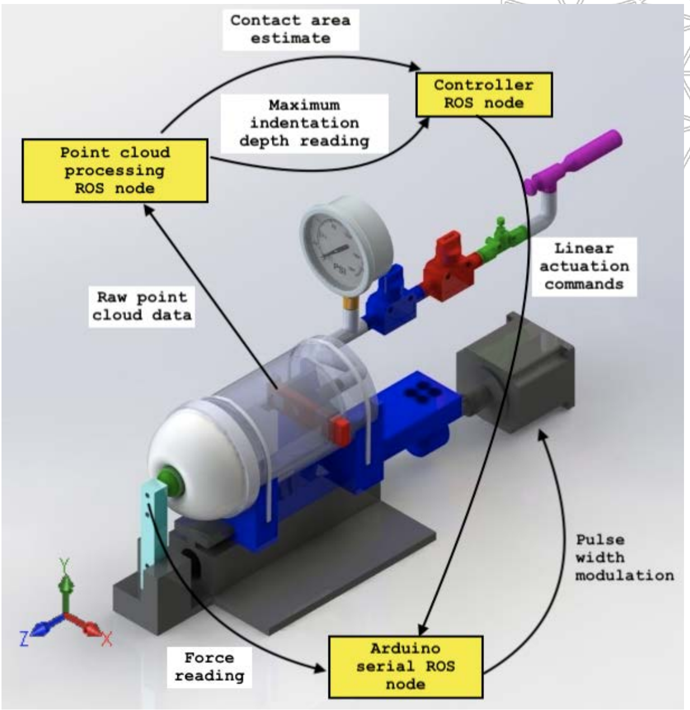
Isabella Huang, Jingjun Liu, Ruzena Bajcsy

SOFTCell: A Depth-Camera-Based Soft Fingertip Device

Design of novel soft tactile sensor

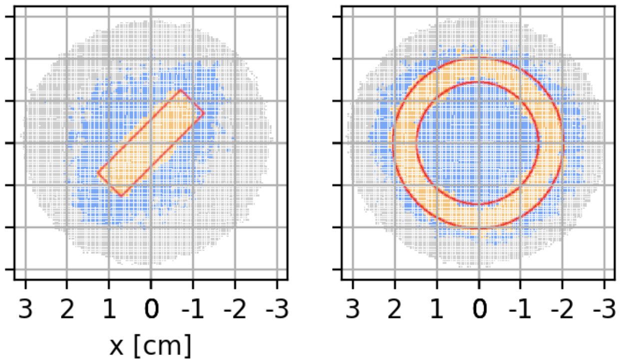
- Novel usage of miniature depth camera
- Characterization and demonstration of **action-perception coupling**
- Modulation of stiffness with pneumatics
- Imaging of elastic membrane deformation gives us information about contact **obstacle geometry** and **applied forces**



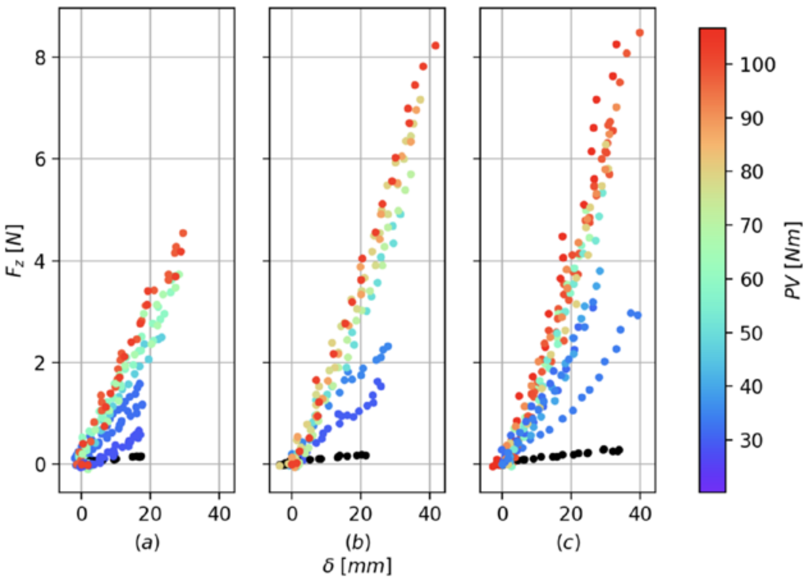


Soft sensor can read geometry and forces

- We are able to distinguish the contact geometry of obstacles
- Force-deformation characteristics were also characterized in the coaxial normal case
- Modulation of internal pressure changes sensitivity to contact interactions



Shape of obstacle (yellow) can be read accurately

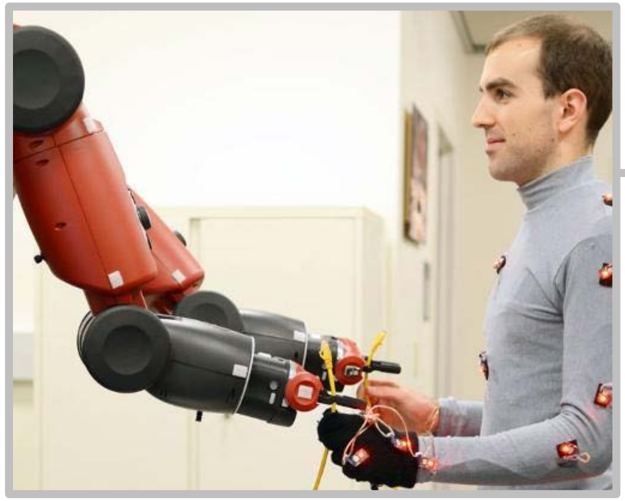


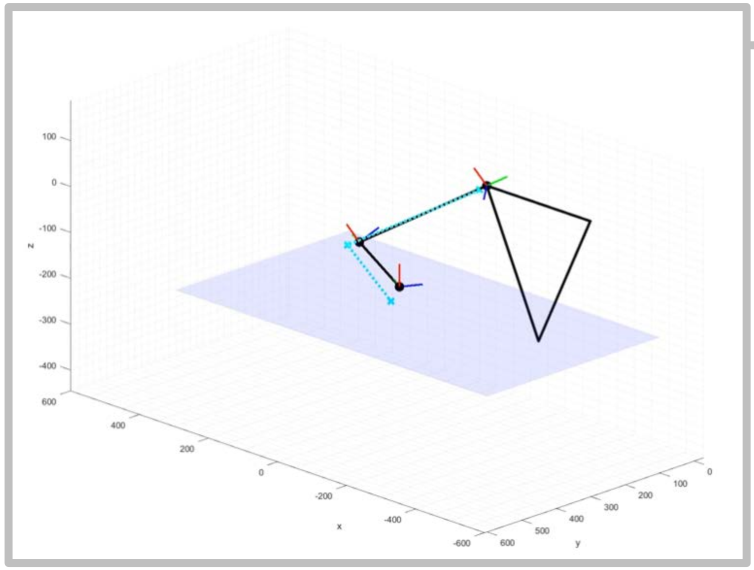
Force-deformation characteristics were measured for different internal pressures

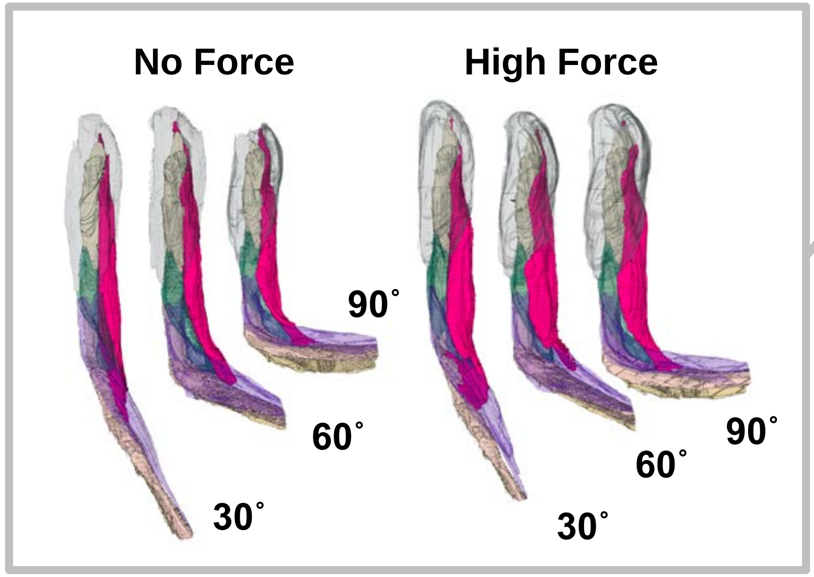
Robert Matthew, Sarah Seko, Laura Hallock, Zoe Cohen, Aaron Bestick, Ruzena Bajcsy

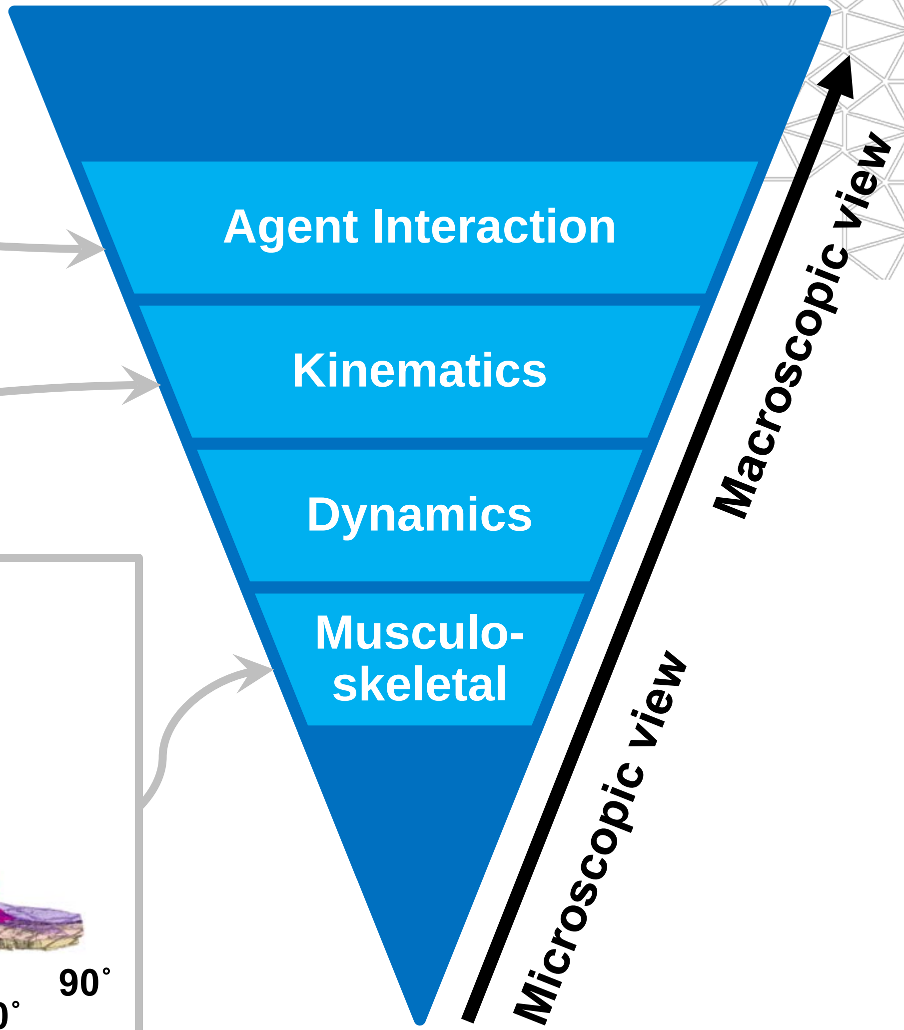
Human Modeling Overview

HART Lab Human Modeling Overview









{rpmatthew, seko, lhallock, cohen.zoe, abestick}@berkeley.edu

SECTION IV

Summary & Conclusions

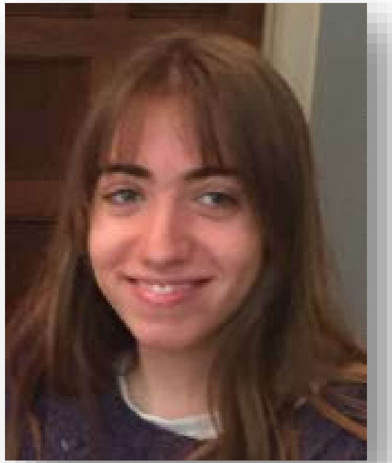
Experiments & Computational Capabilities

It is widely acknowledged that in science we stand on a three-legged stool: **theory**, **experiments**, and **computation**.

Unfortunately, it turns out that ***many theoretical models/equations are not amenable to fast or large-scale computation***.

Hence, it behooves us to **reexamine** these models, **reformulate** them to admit computation, and bring the technology to **new applications**.

Questions?


Zoe Cohen


Isabella Huang


Carolyn Chen


Laura Hallock


Sarah Seko


Robert Matthew

NSF GRFP NSF GRADUATE RESEARCH FELLOWSHIP PROGRAM
BAIR BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
SIEMENS
UCSF