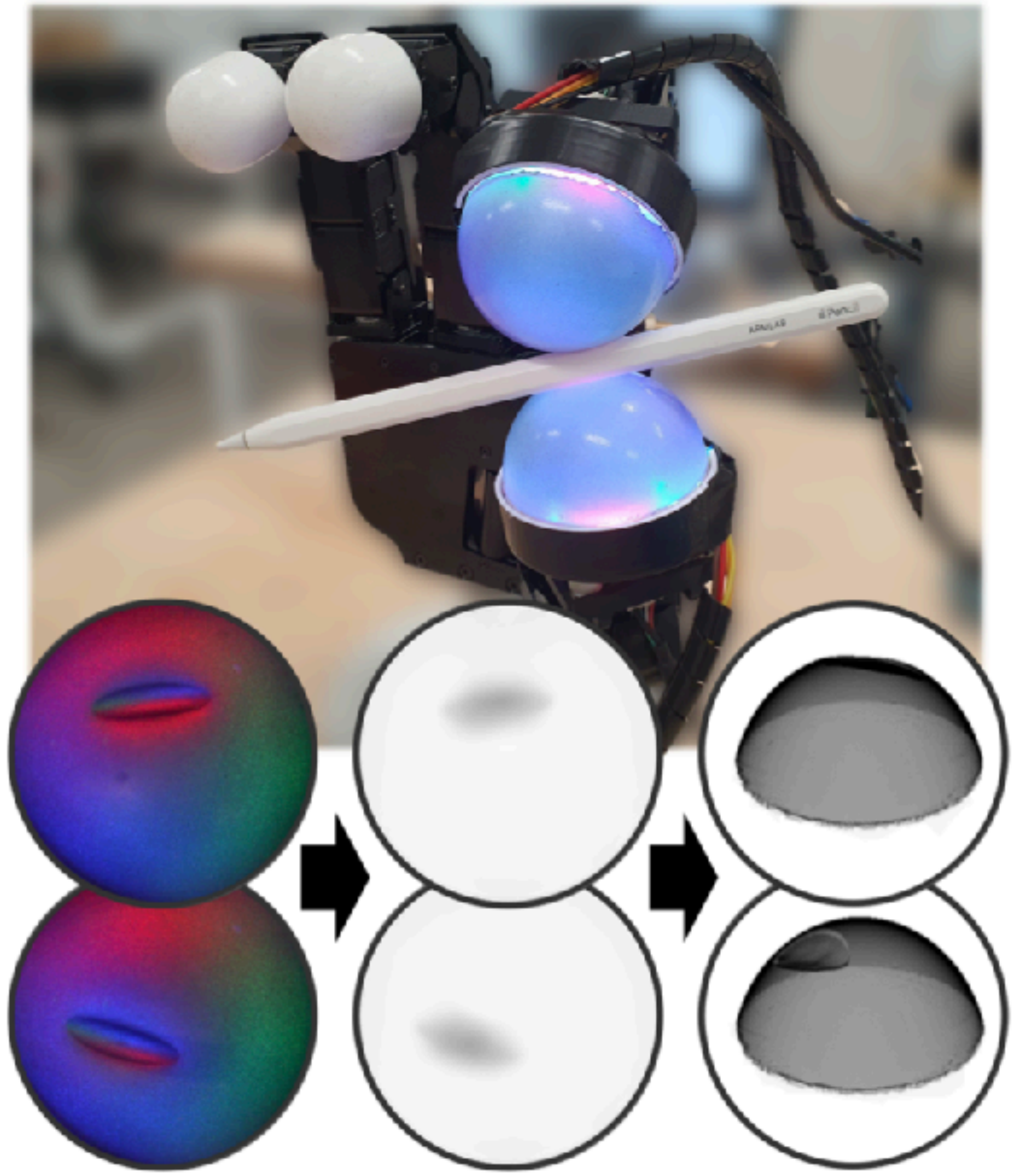
Robotic Manipulation and the Sense of Touch



Background for Monroe Kennedy's DenseTact

Jeannette Bohg - Stanford University
Interactive Perception and Robot Learning lab

Apollo @ MPI for Intelligent Systems

Kids Making Robots Jealous

Outline

Why is robot manipulation hard?

The role of multiple sensor modalities

The sense of touch in nature

The sense of touch in robots

Outline

Why is robot manipulation hard?

The role of multiple sensor modalities

The sense of touch in nature

The sense of touch in robots

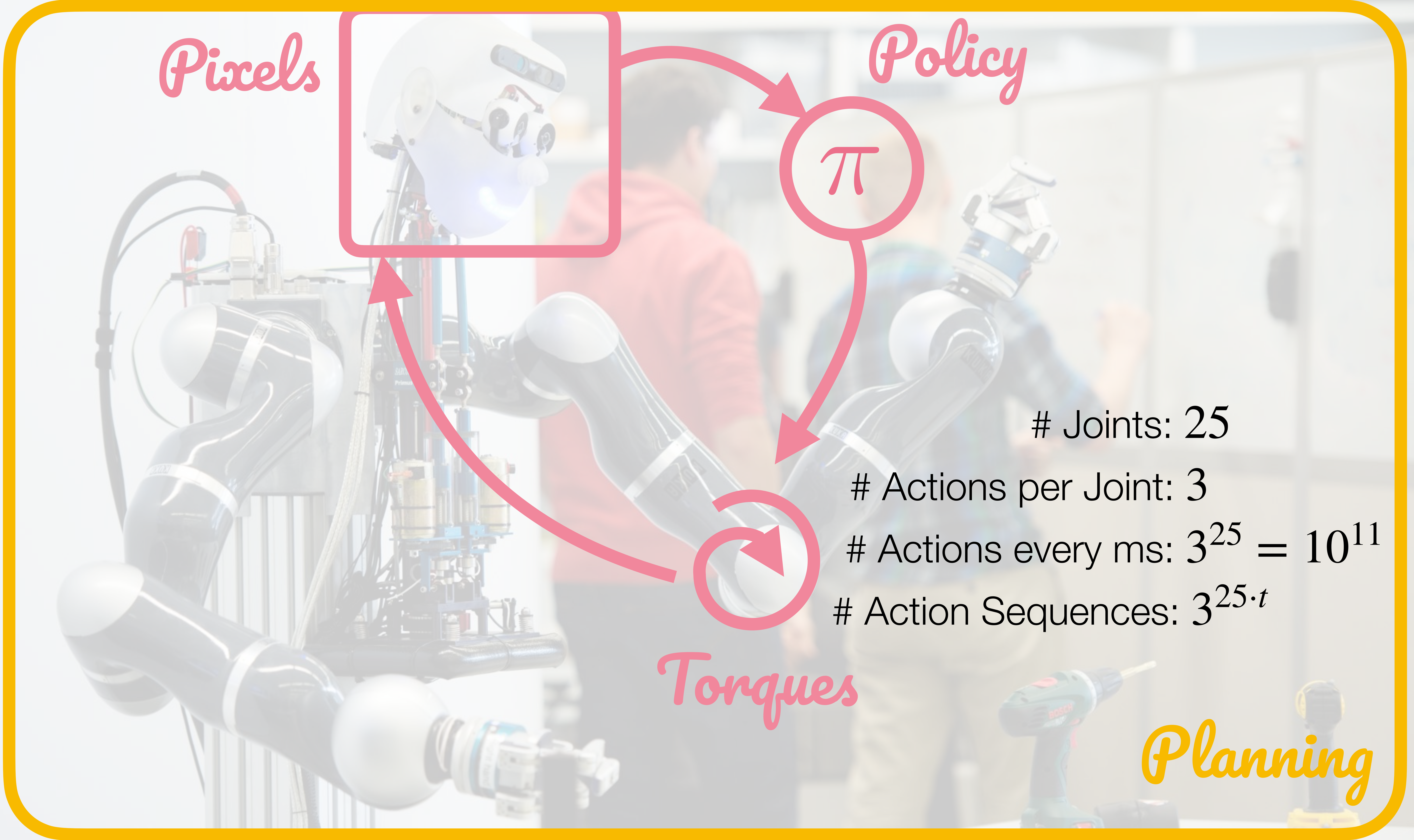
Pixels
Policy
π
Torques
Planning

Joints: 25

Actions per Joint: 3

Actions every ms: $3^{25} = 10^{11}$

Robot Manipulation – Why is it hard?


Pixels
Policy
π
Torques
Joints: 25
Actions per Joint: 3
Actions every ms: $3^{25} = 10^{11}$
Action Sequences: $3^{25 \cdot t}$
Planning


Pixels
Policy
π
Torques
Actions every ms: $3^{25} = 10^{11}$

Bremermann limit

$$1.36 \times 10^{50}$$

bits per kilogram per second

Computer of Size of Earth

10^{75} Ops per second

10^{79} Seconds for 512 crypto key

Searching the space of possible actions

10^{75} Ops per second

10^{64} Seconds to brute-force search actions


Pixels
Policy
π
Torques
Actions every ms: $3^{25} = 10^{11}$


We need some kind of search bias!
Data
Priors

Outline

Why is robot manipulation hard?

The role of multiple sensor modalities

The sense of touch in nature

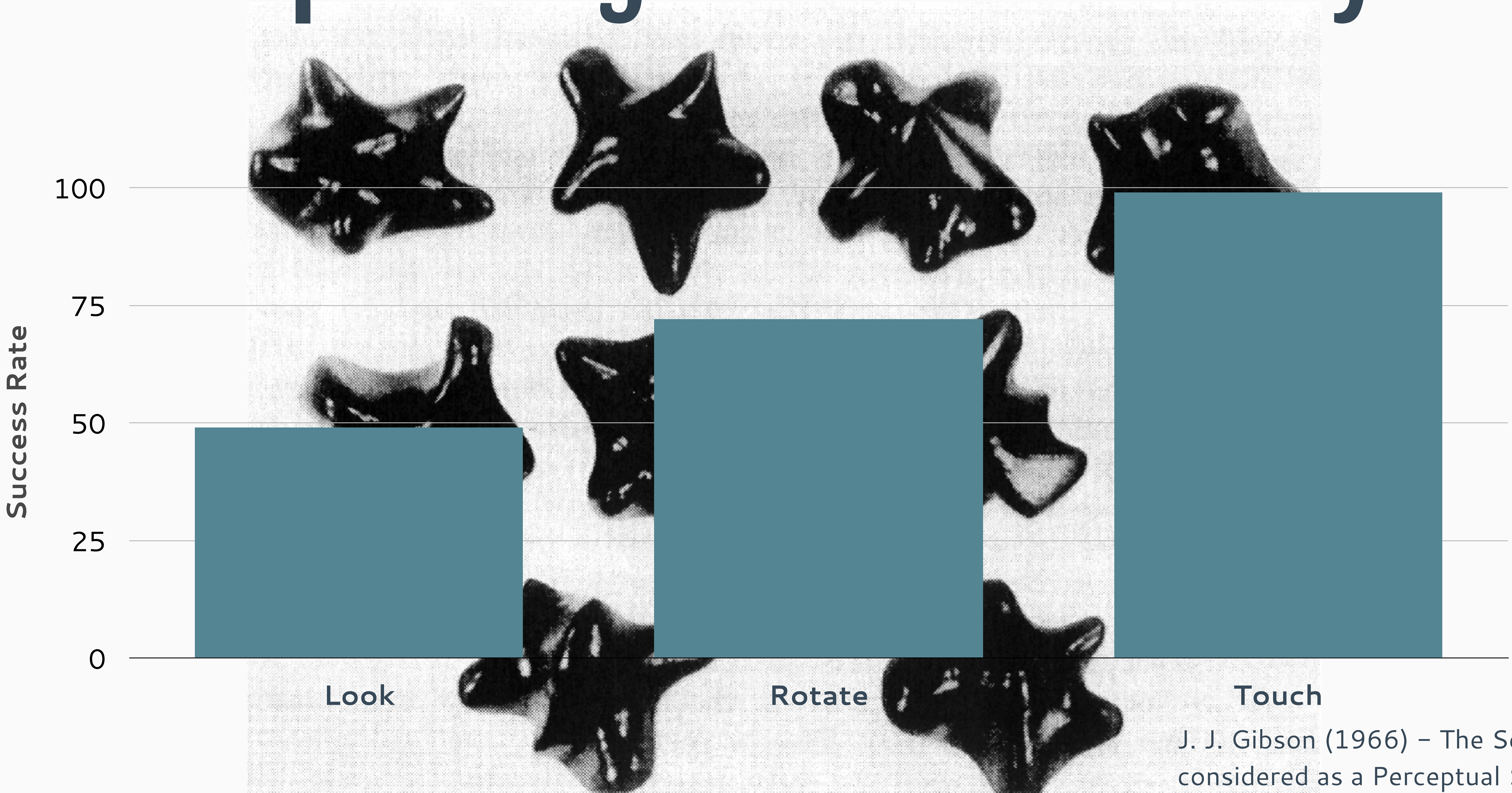
The sense of touch in robots


Pixels
Policy
π
Torques


Tactile
Audio
RGB & Depth
Force/Torque
Active Stereo
Joint Torque

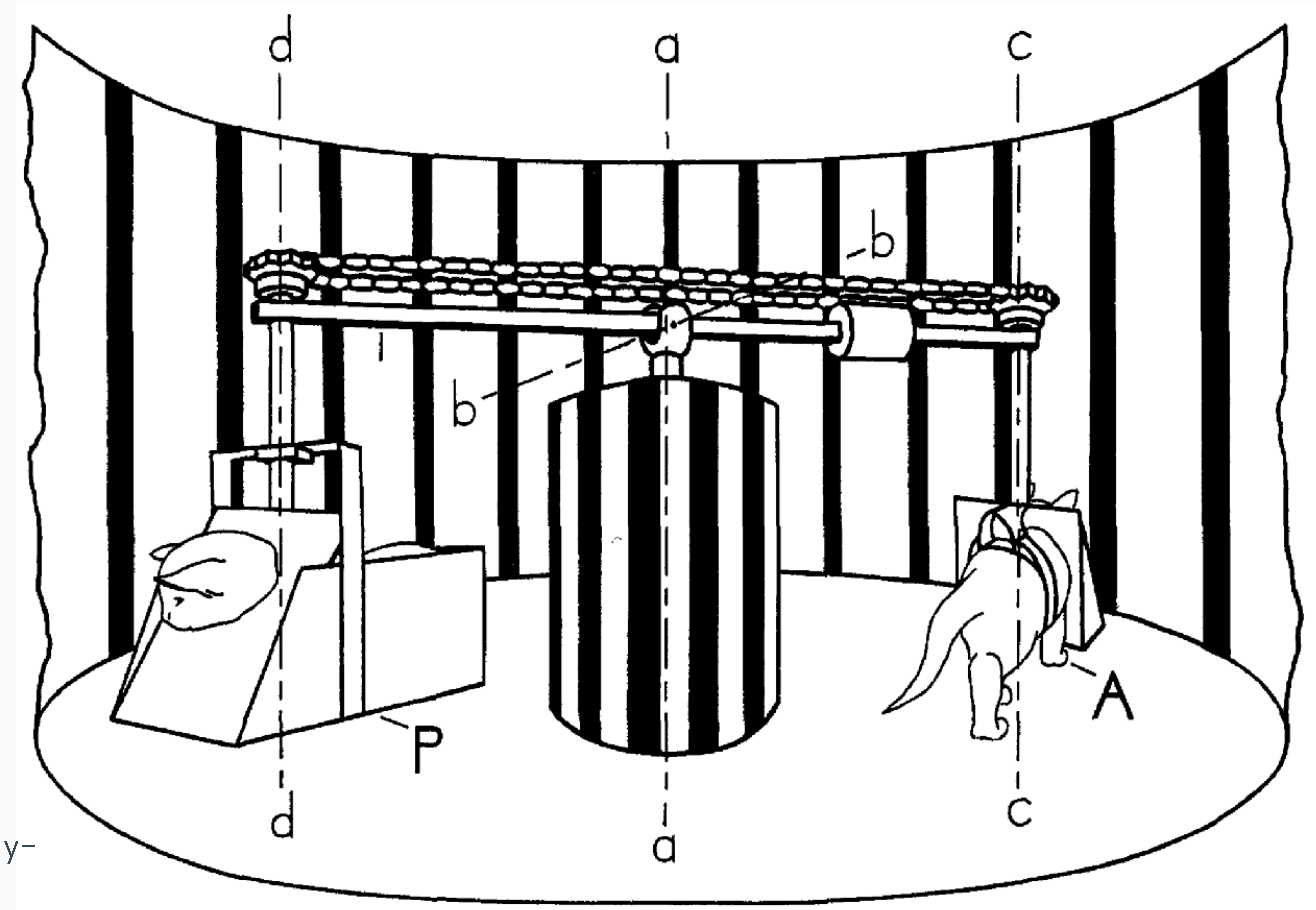

Jenga
Jenga
Jenga
Jenga

Exploiting Multi-Modality



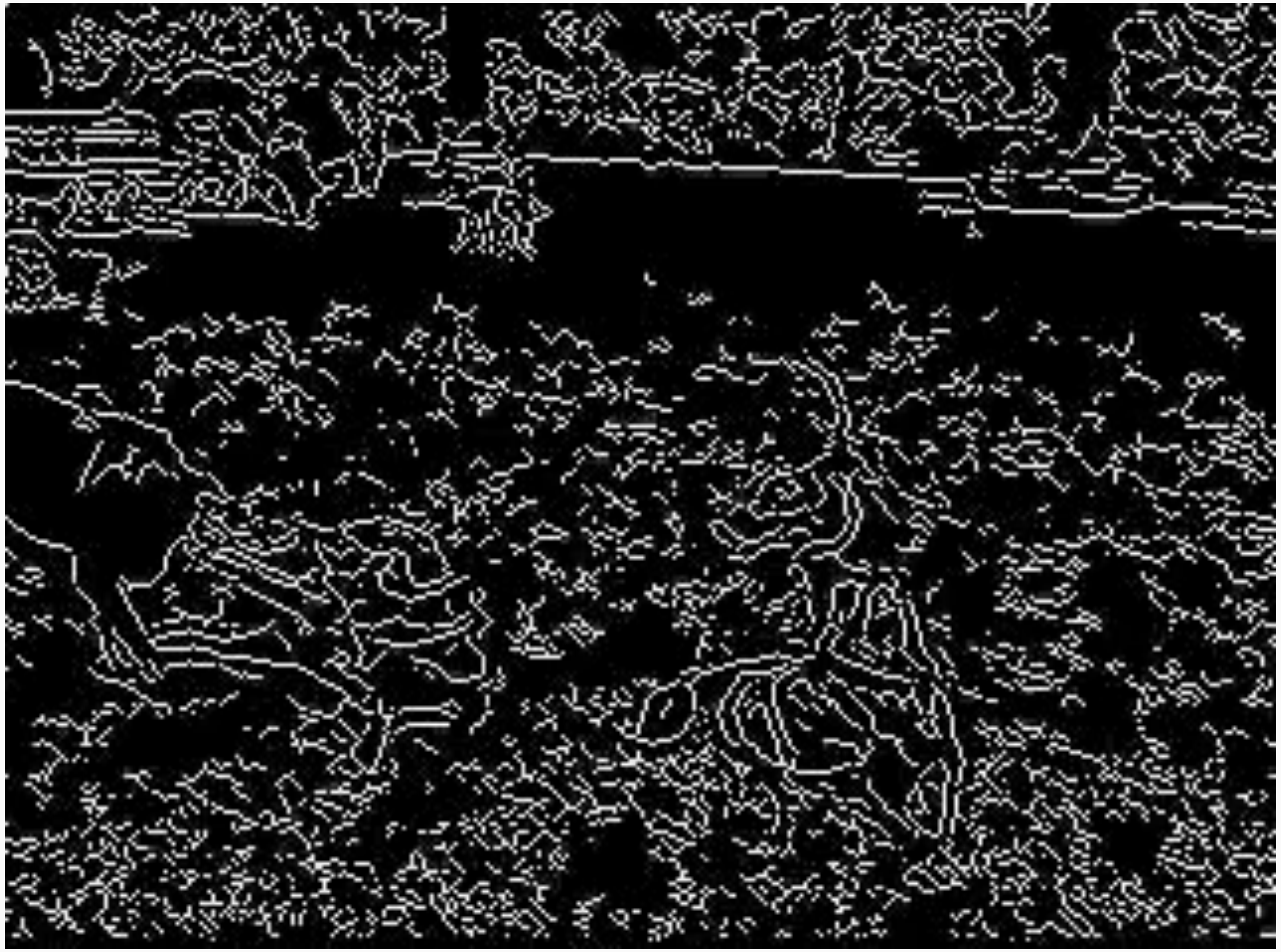
J. J. Gibson (1966) – The Senses considered as a Perceptual System.

Concurrency of Motion and Sensing



Held and Hein (1963). Movement-Produced Stimulation in the Development of Visually-Guided Behaviour

Accumulation over Time



Thanks to Prof. Octavia Camps at Northeastern University, Boston

Outline

Why is robot manipulation hard?

The role of multiple sensor modalities

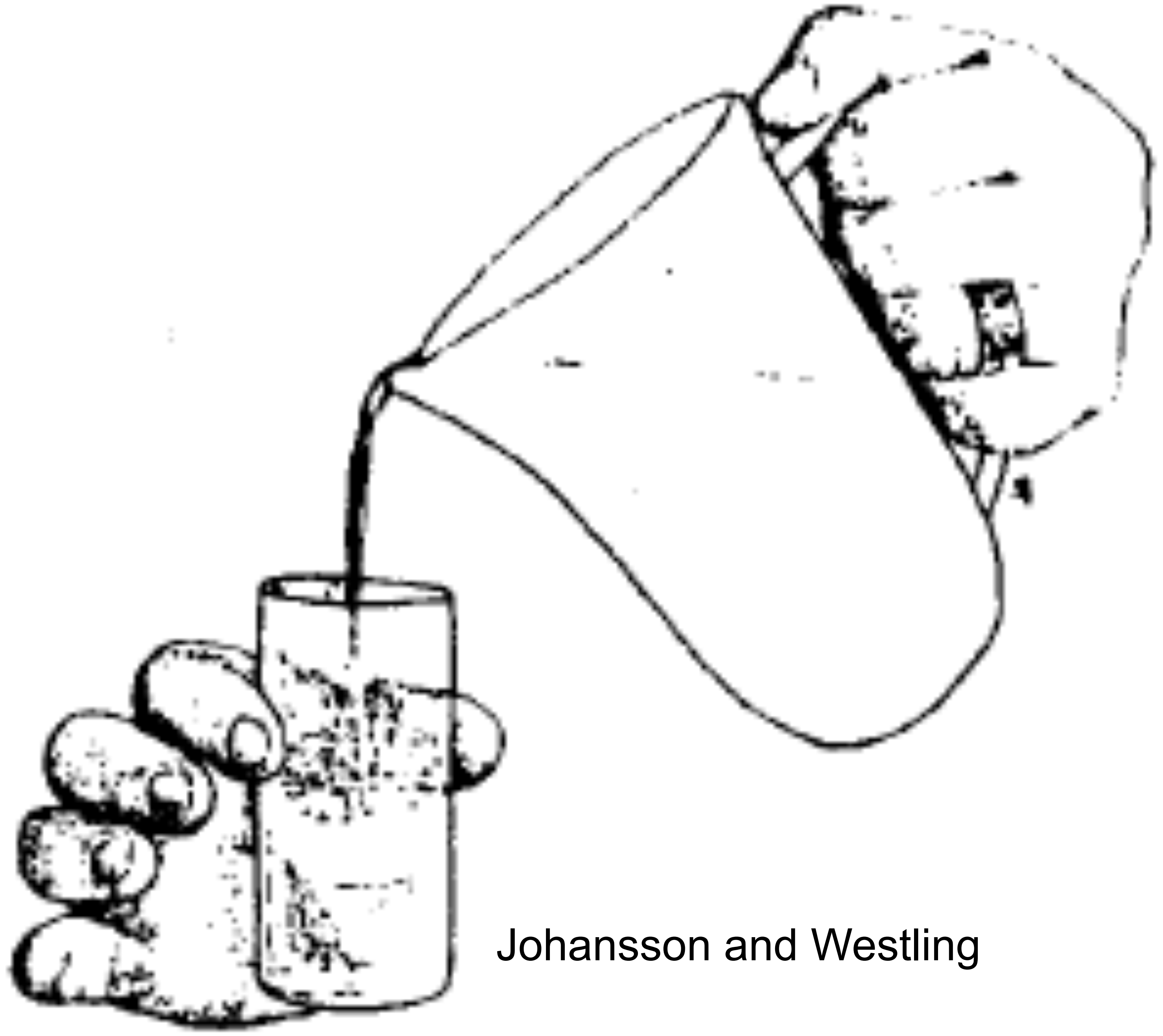
The sense of touch in nature

The sense of touch in robots

Tactile sensing is a component of haptics

Cutaneous (Tactile)

Temperature
Texture
Slip
Vibration
Force



Johansson and Westling

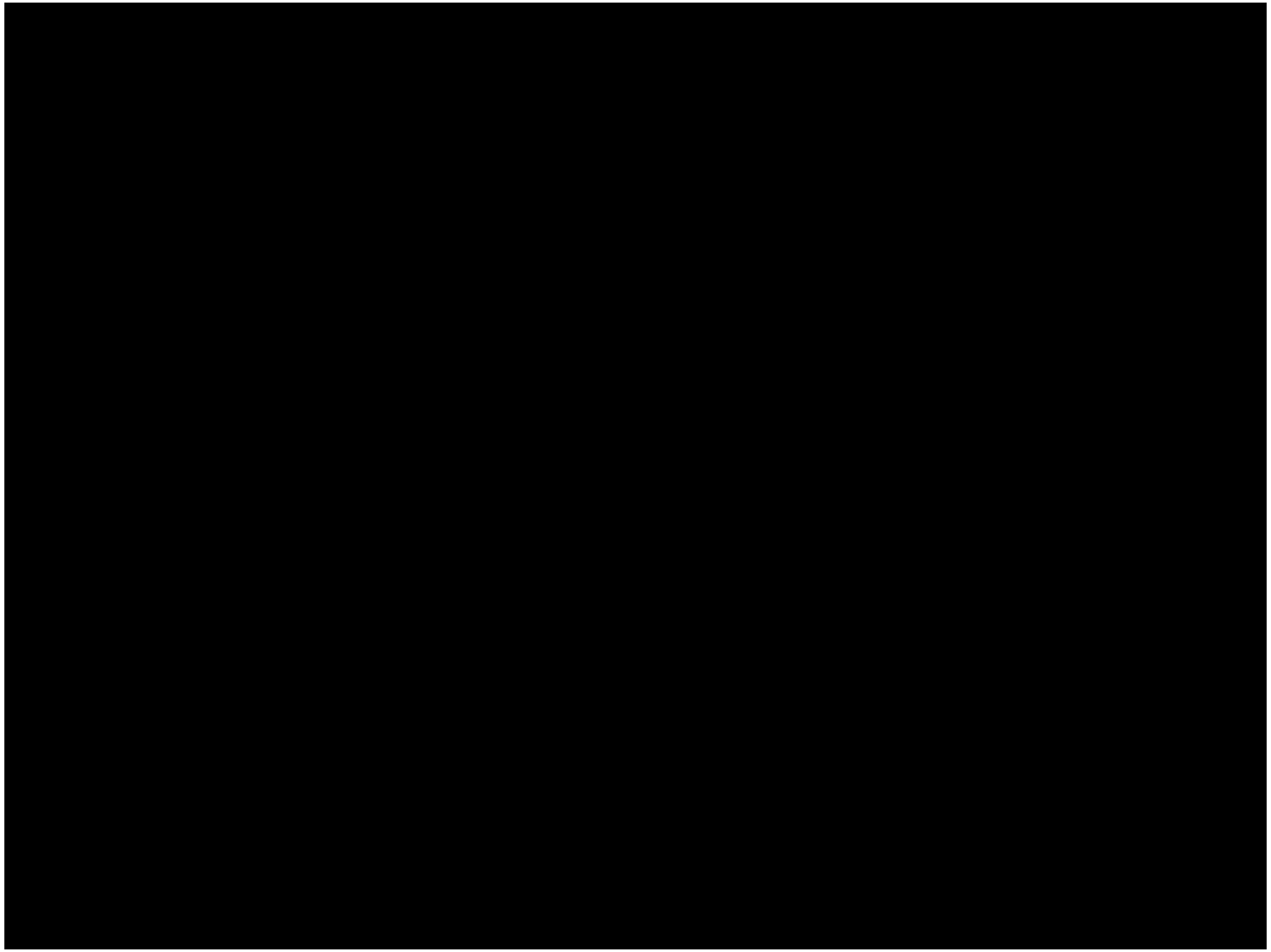
Kinesthesia

Location
Configuration
Motion
Force
Compliance

The haptic senses work together with the motor control system to:
- Coordinate movement
- Enable perception

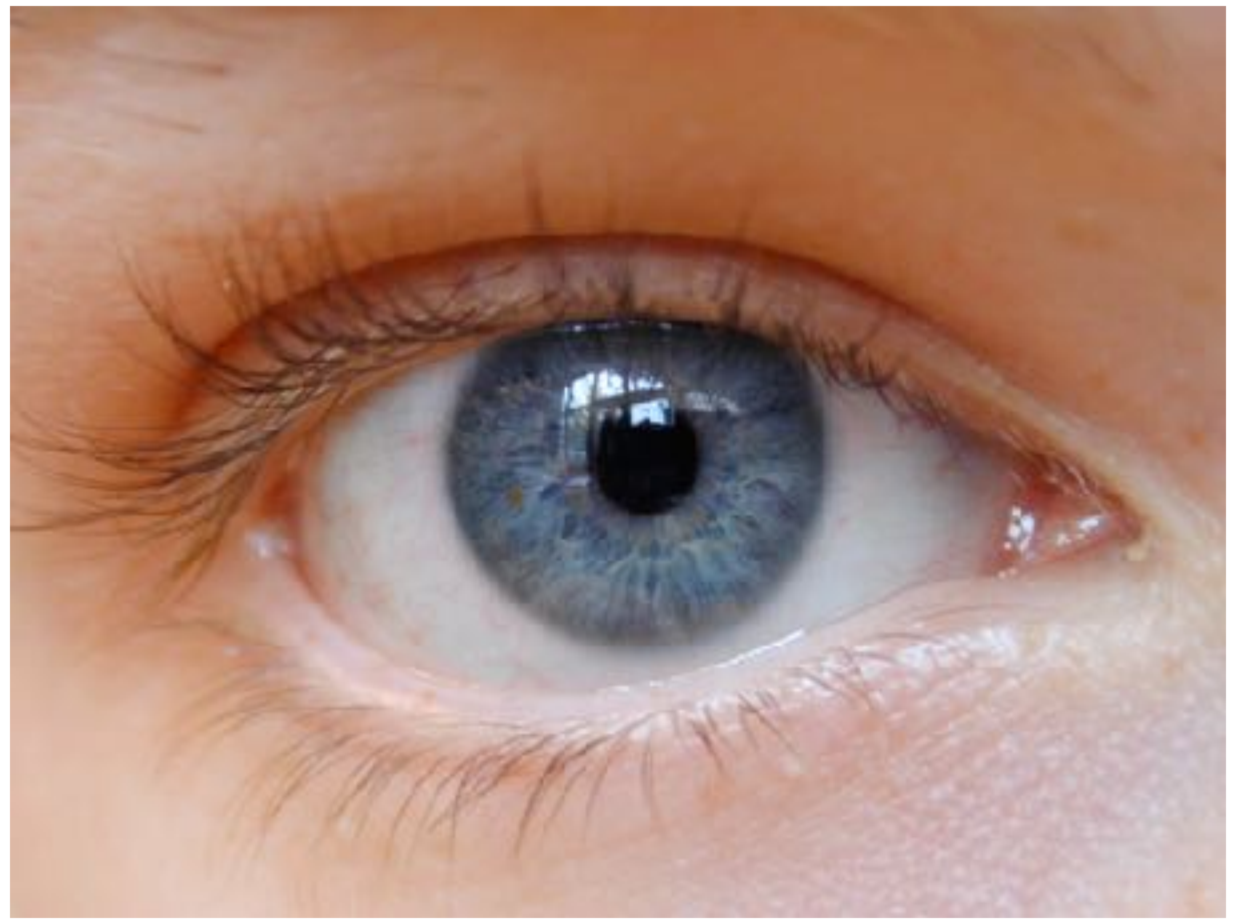
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

What happens without tactile sensing?



Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Vision vs. Touch

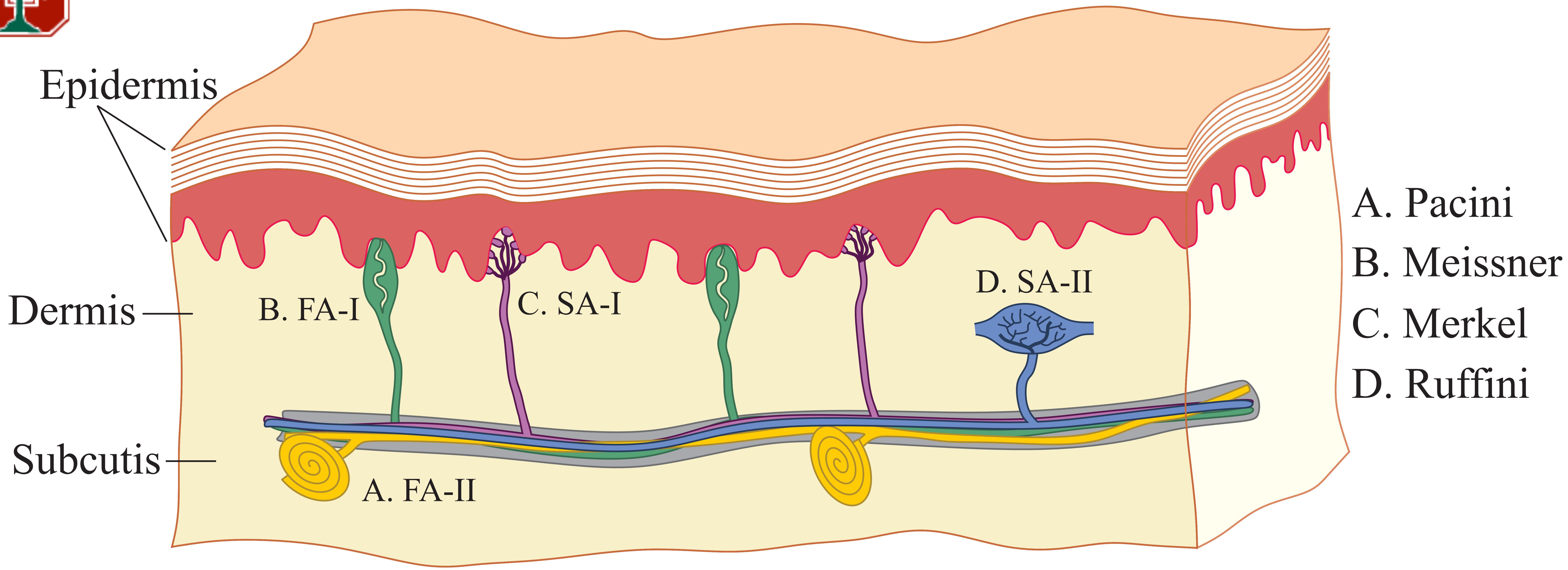


centralized
broad
passive
cognitive

distributed
narrow
active
physical

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Human mechanoreceptors

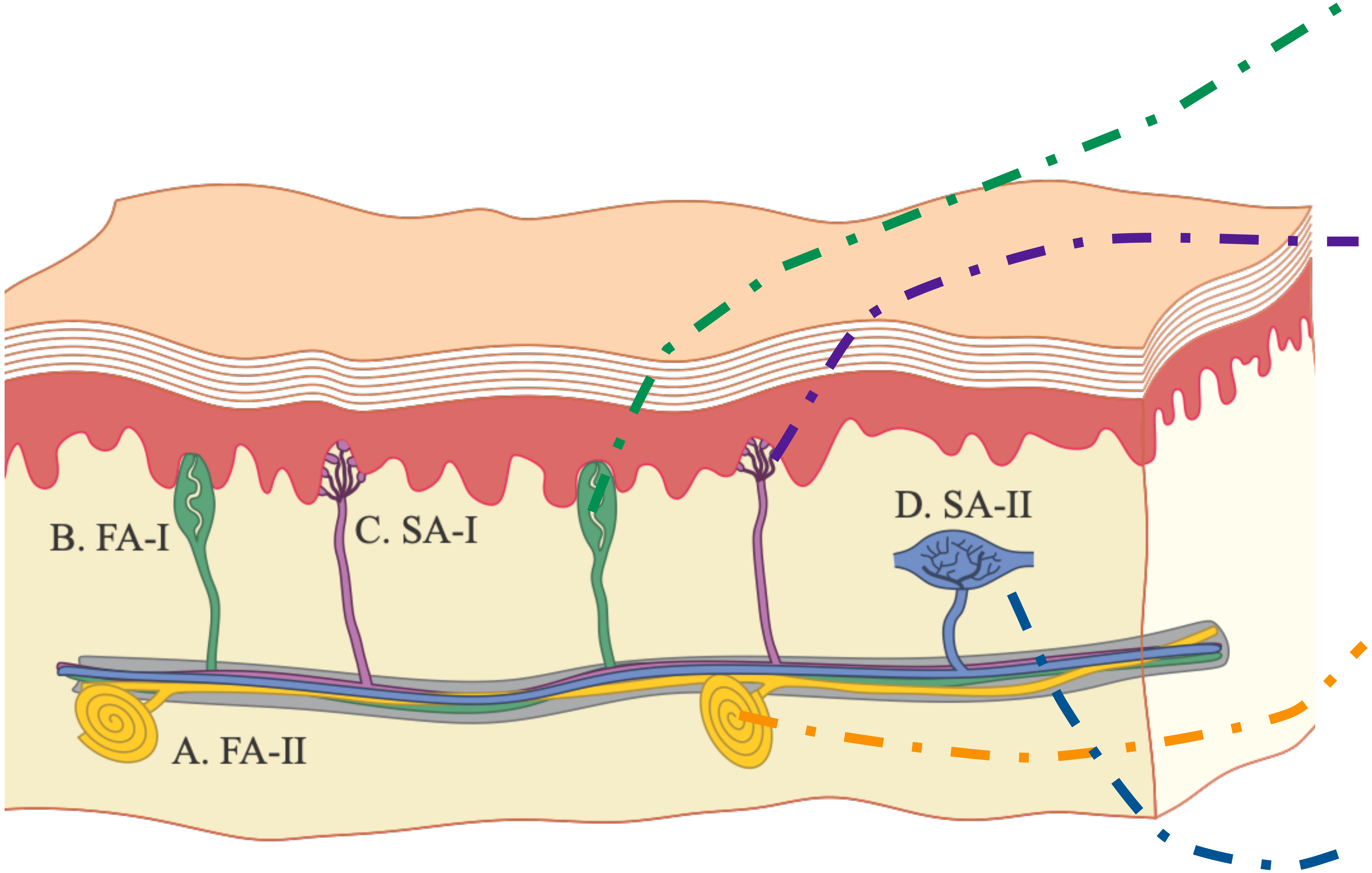


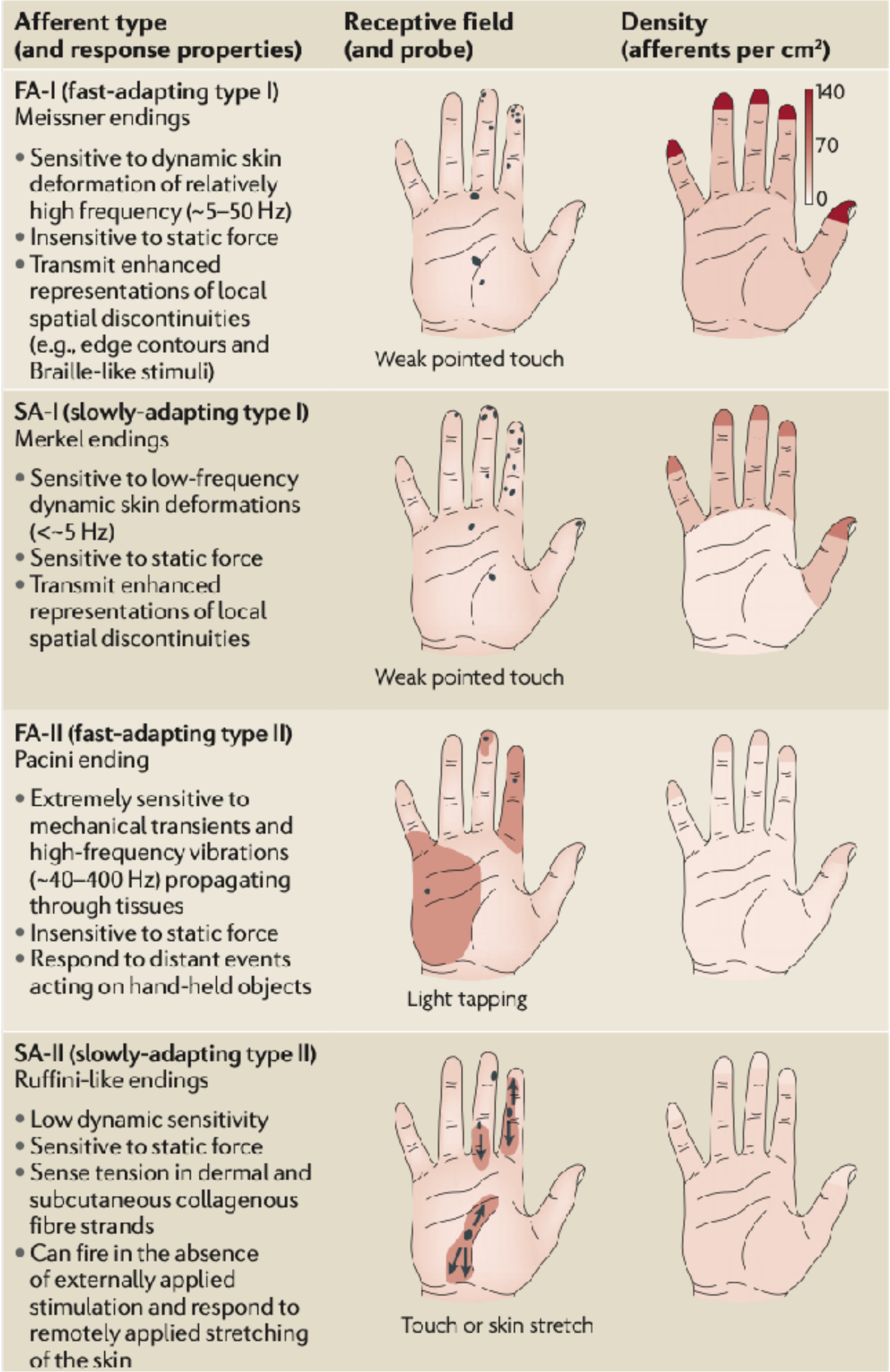
Cutkosky M.R. and Ulmen, J., "Dynamic tactile sensing," in The Human Hand as an Inspiration for Robotic Hands, R. Balasubrmanian and V. Santos, eds., Springer Verlag.

Receptor	receptive field	frequency range	sensed quantity
FA-I Meissner corpuscles 140/cm^2	3-4mm	5-60 Hz	dynamic skin deformation
SA-I Merkel endings 70/cm^2	3-4mm	0-5 Hz	compressive stresses
FA-II Pacini corpuscles	>20mm	50-500+ Hz (peak sensitivity ~250 Hz)	vibration
SA-II Ruffini endings	>10mm	0-10 Hz	directional skin stretch

table data from various sources (see Johansson & Flanagan 2009 for review)

Mapping mechanoreceptors



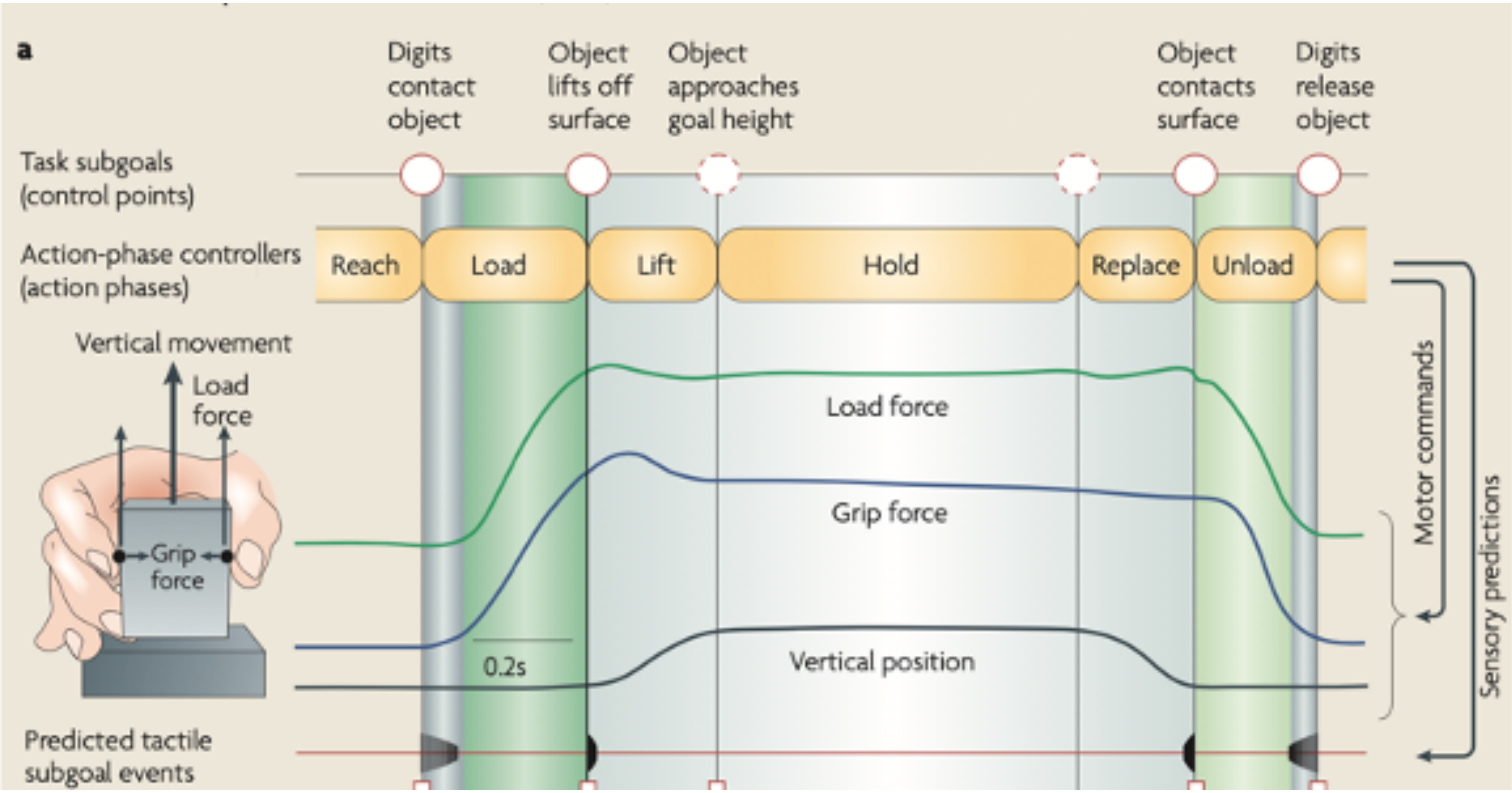


Afferent type (and response properties)	Receptive field (and probe)	Density (afferents per cm²)
FA-I (fast-adapting type I) Meissner endings • Sensitive to dynamic skin deformation of relatively high frequency (~5–50 Hz) • Insensitive to static force • Transmit enhanced representations of local spatial discontinuities (e.g., edge contours and Braille-like stimuli)	Weak pointed touch	140 / 70 / 0
SA-I (slowly-adapting type I) Merkel endings • Sensitive to low-frequency dynamic skin deformations (<~5 Hz) • Sensitive to static force • Transmit enhanced representations of local spatial discontinuities	Weak pointed touch	
FA-II (fast-adapting type II) Pacini ending • Extremely sensitive to mechanical transients and high-frequency vibrations (~40–400 Hz) propagating through tissues • Insensitive to static force • Respond to distant events acting on hand-held objects	Light tapping	
SA-II (slowly-adapting type II) Ruffini-like endings • Low dynamic sensitivity • Sensitive to static force • Sense tension in dermal and subcutaneous collagenous fibre strands • Can fire in the absence of externally applied stimulation and respond to remotely applied stretching of the skin	Touch or skin stretch	

[Cutkosky, Ulmen, '14]

R. S. Johansson, J. R. Flanagan, Coding and use of tactile signals from the fingertips in object manipulation tasks, *Nature Reviews,* 2009

Tactile events in manipulation tasks

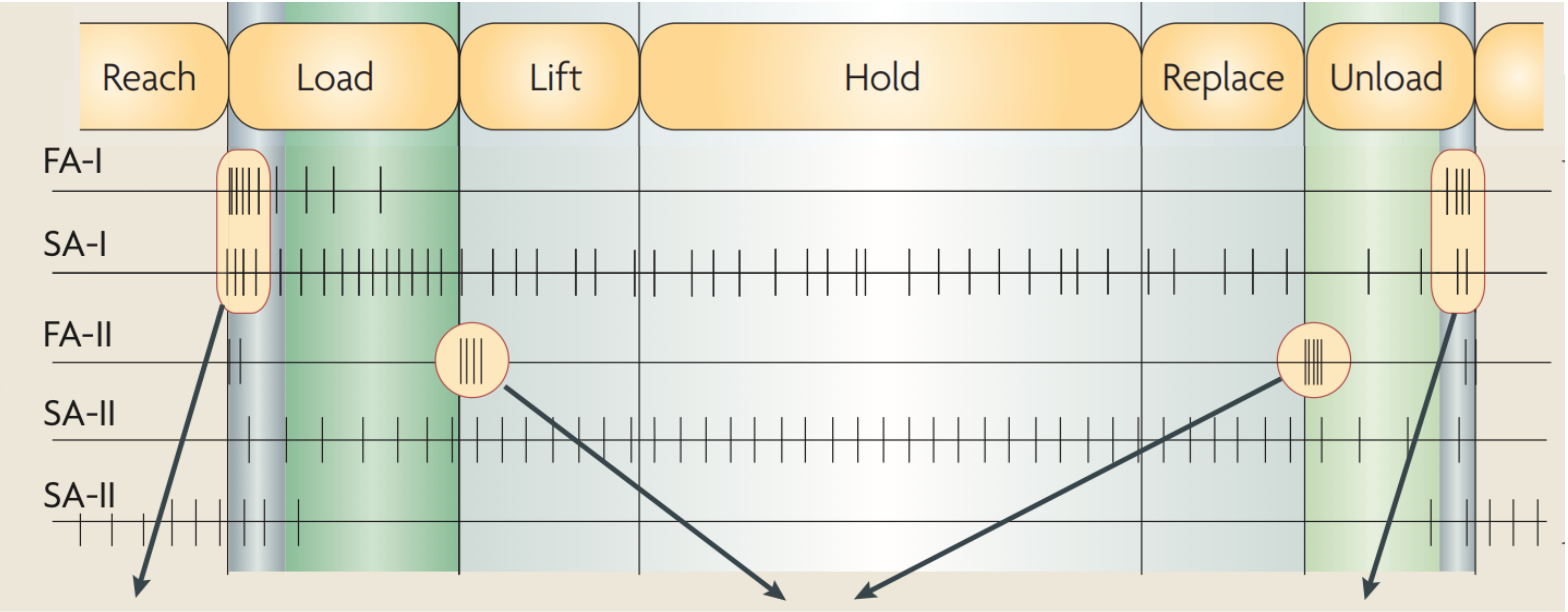


Dexterous manipulation is about balancing *grip* and *load forces* to object surface properties

[Johansson, Flanagan, '09]

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Tactile events in manipulation tasks



fingers contact object

object makes or breaks contact with surface

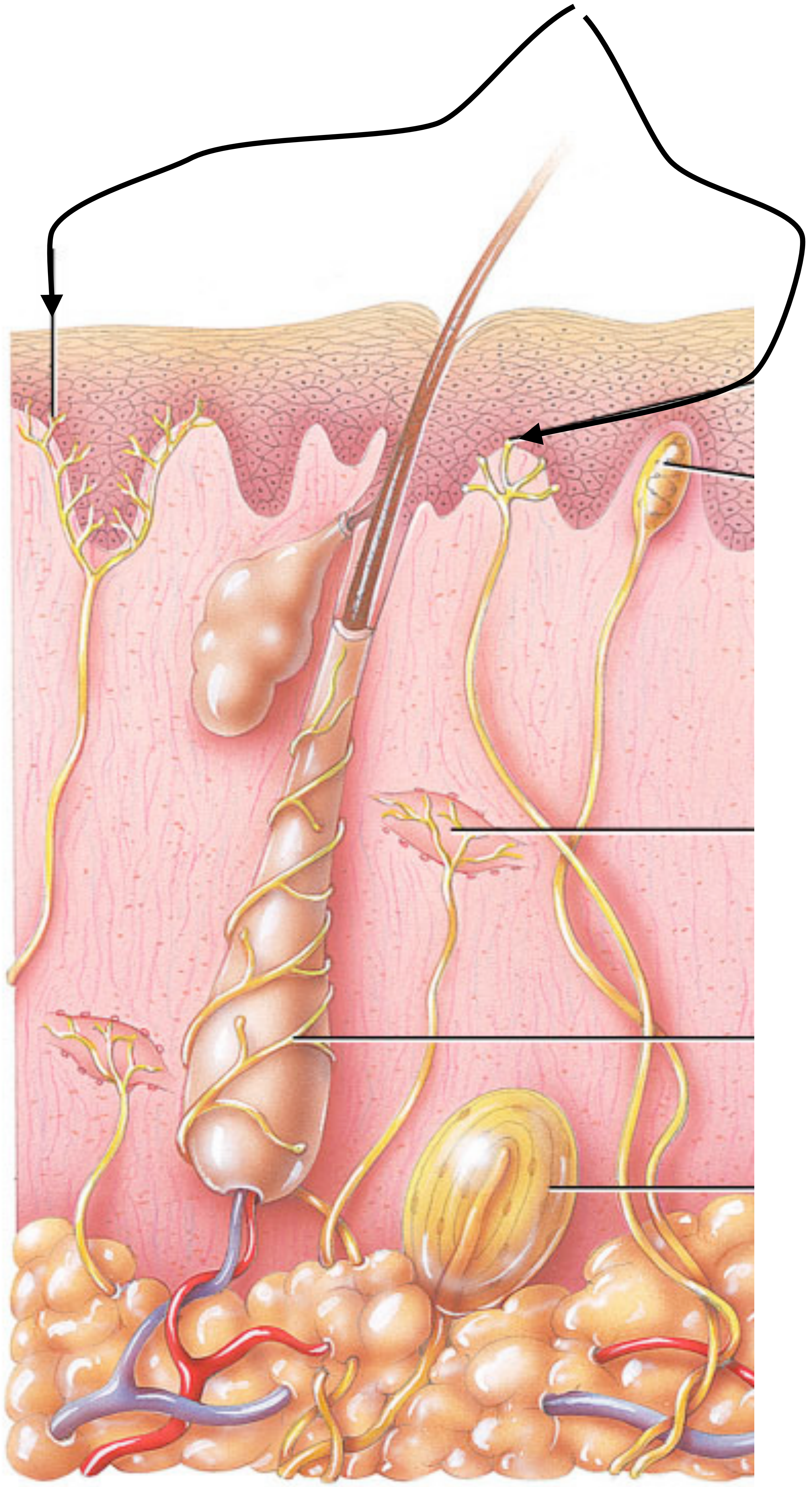
contact released

[Johansson, Flanagan, '09]

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Thermal sensing

- separate warm and cold receptors whose firing rate depends on magnitude of difference w.r.t body temperature
- both slowly adapting (SA) and rapidly adapting (FA) characteristics, so depends on both T and dT/dt
- perception strongly affected by body temperature versus temperature at surface of skin (aluminum feels cooler at room temperature than wood) -- an important component of material identification

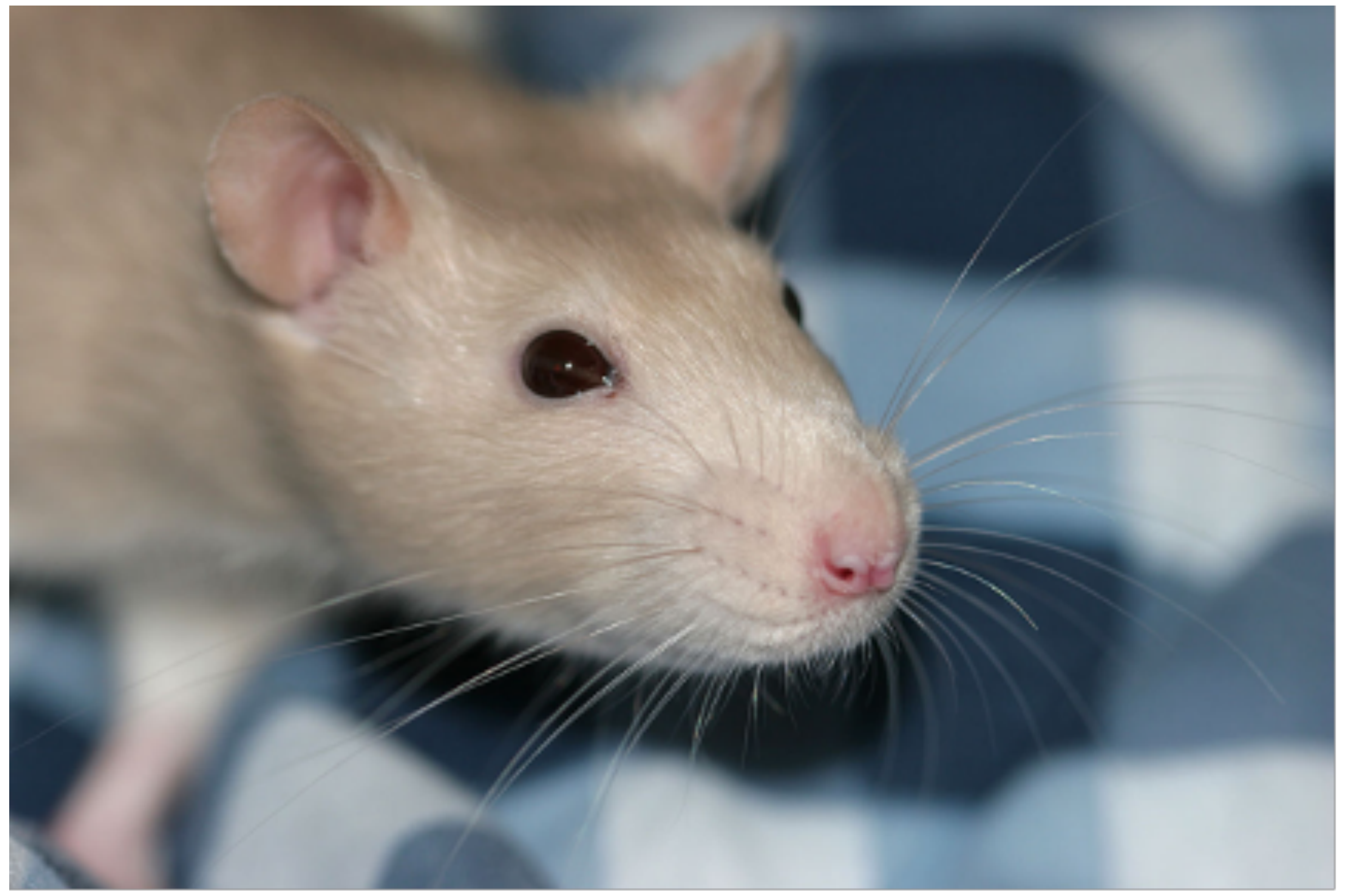


R.K. Adair. A model of the detection of warmth and cold by cutaneous sensors through effects on voltage-gated membrane channels, ***PNAS*** 1999 96 (21).

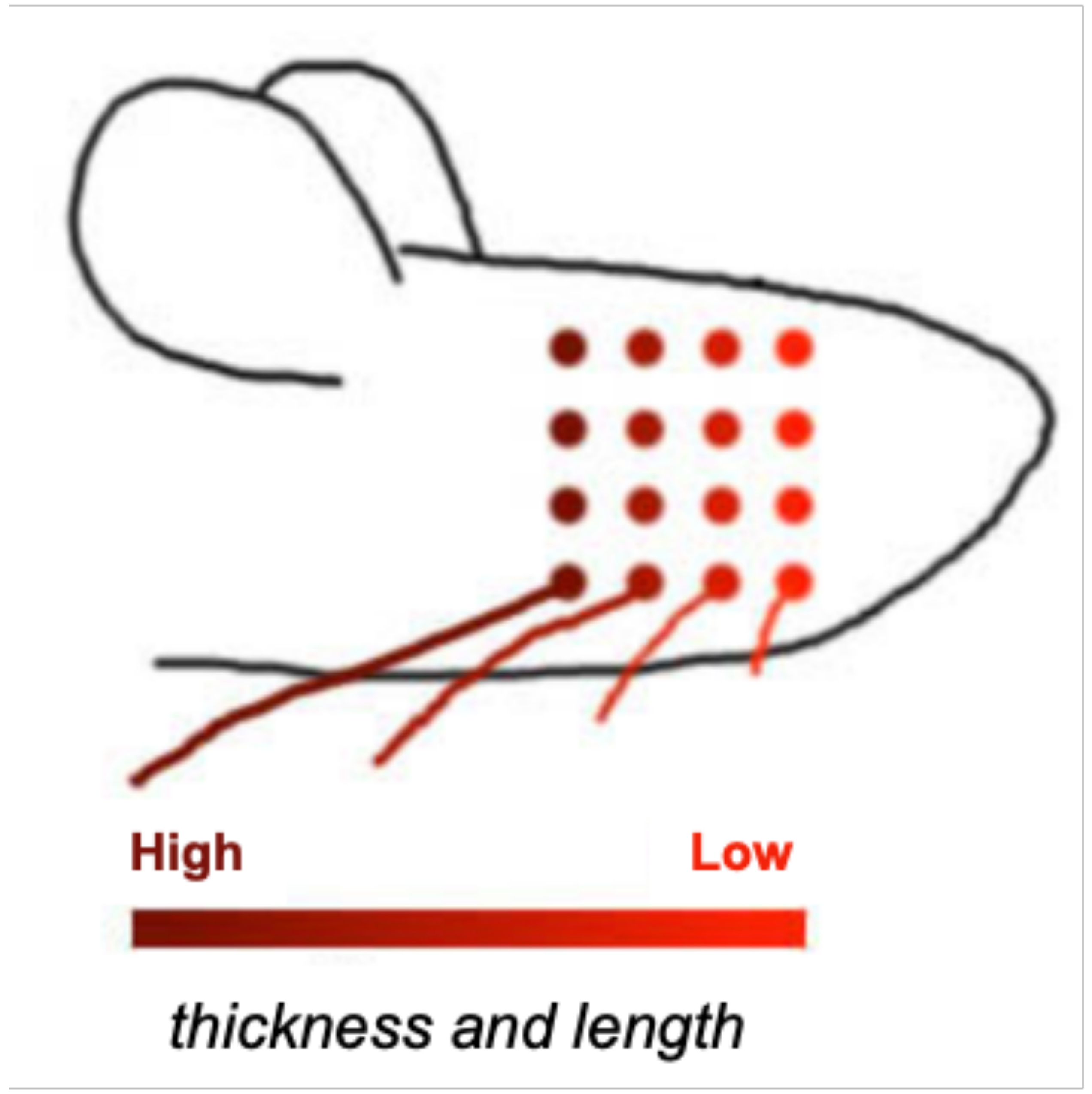
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Tactile Sensing in Nature

Antennae



Rat's whiskers



systematic map of separate tactile channels governed by the mechanical properties of the whiskers

[Gugig, 2020]

Star-nosed mole

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Outline

Why is robot manipulation hard?

The role of multiple sensor modalities

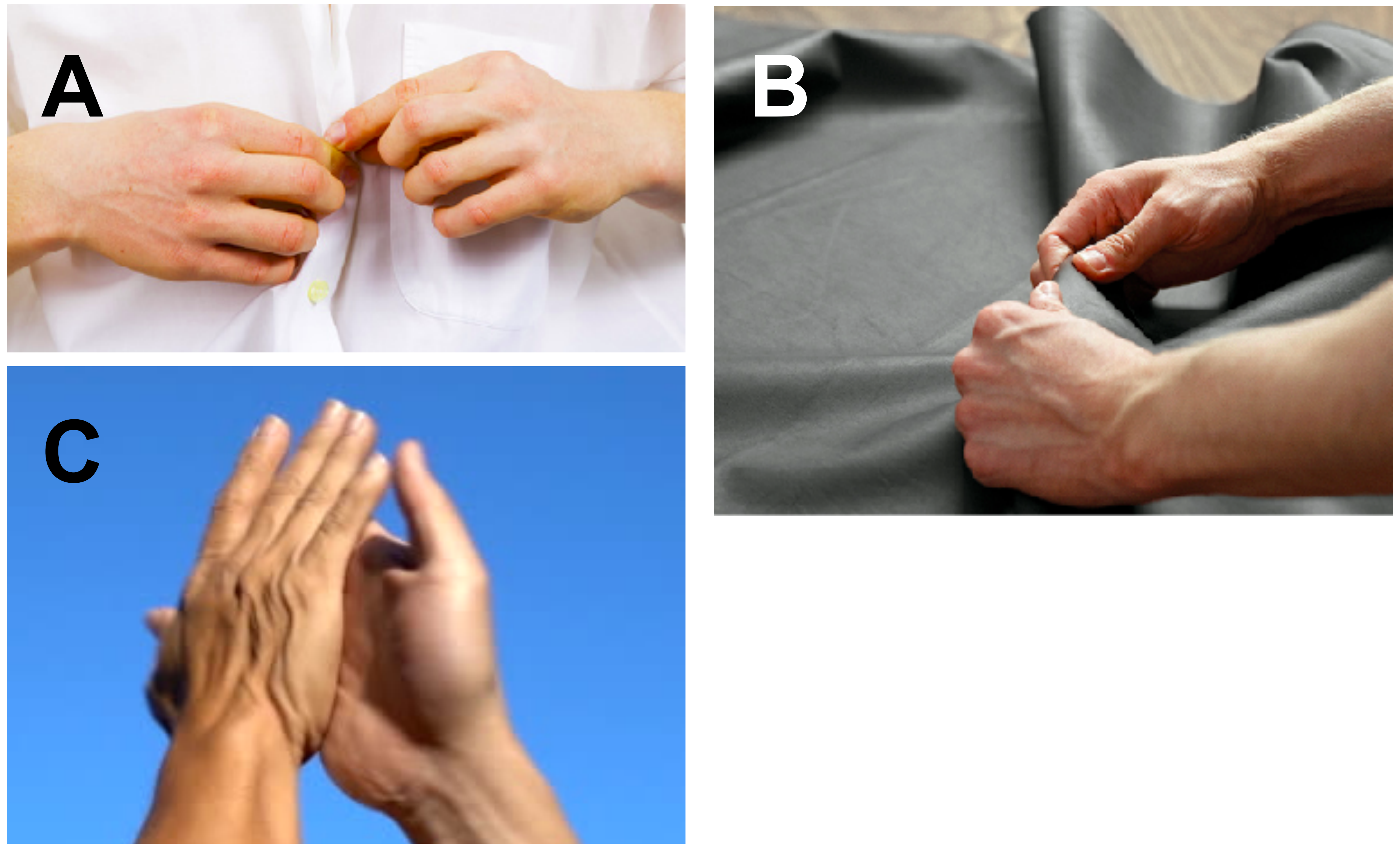
The sense of touch in nature

The sense of touch in robots

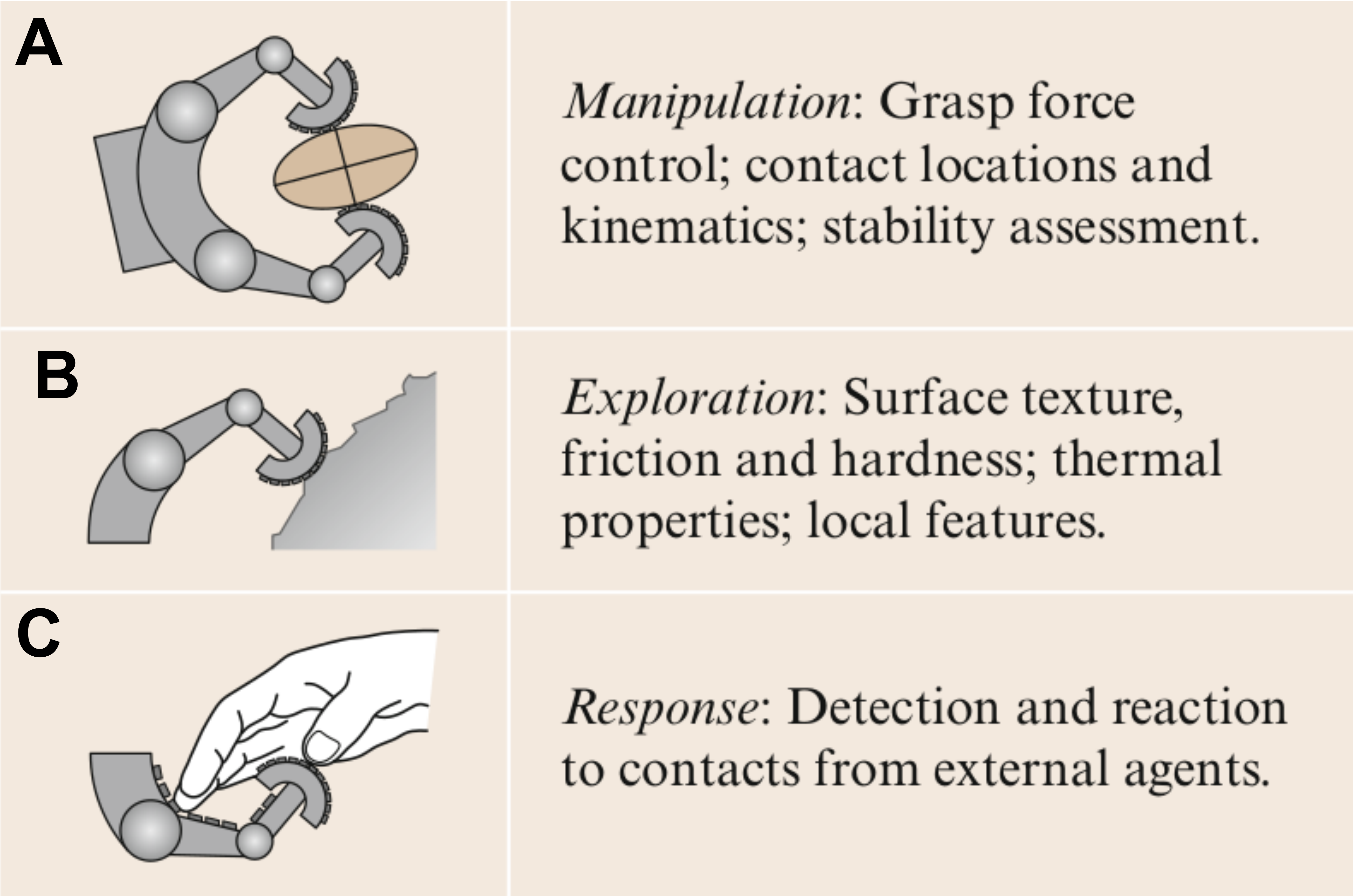
Why Tactile Sensing?

Three main activities:

A. Manipulation
B. Exploration
C. Response



Uses of tactile sensing in humans



Uses of tactile sensing in robotics

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Types of Sensors

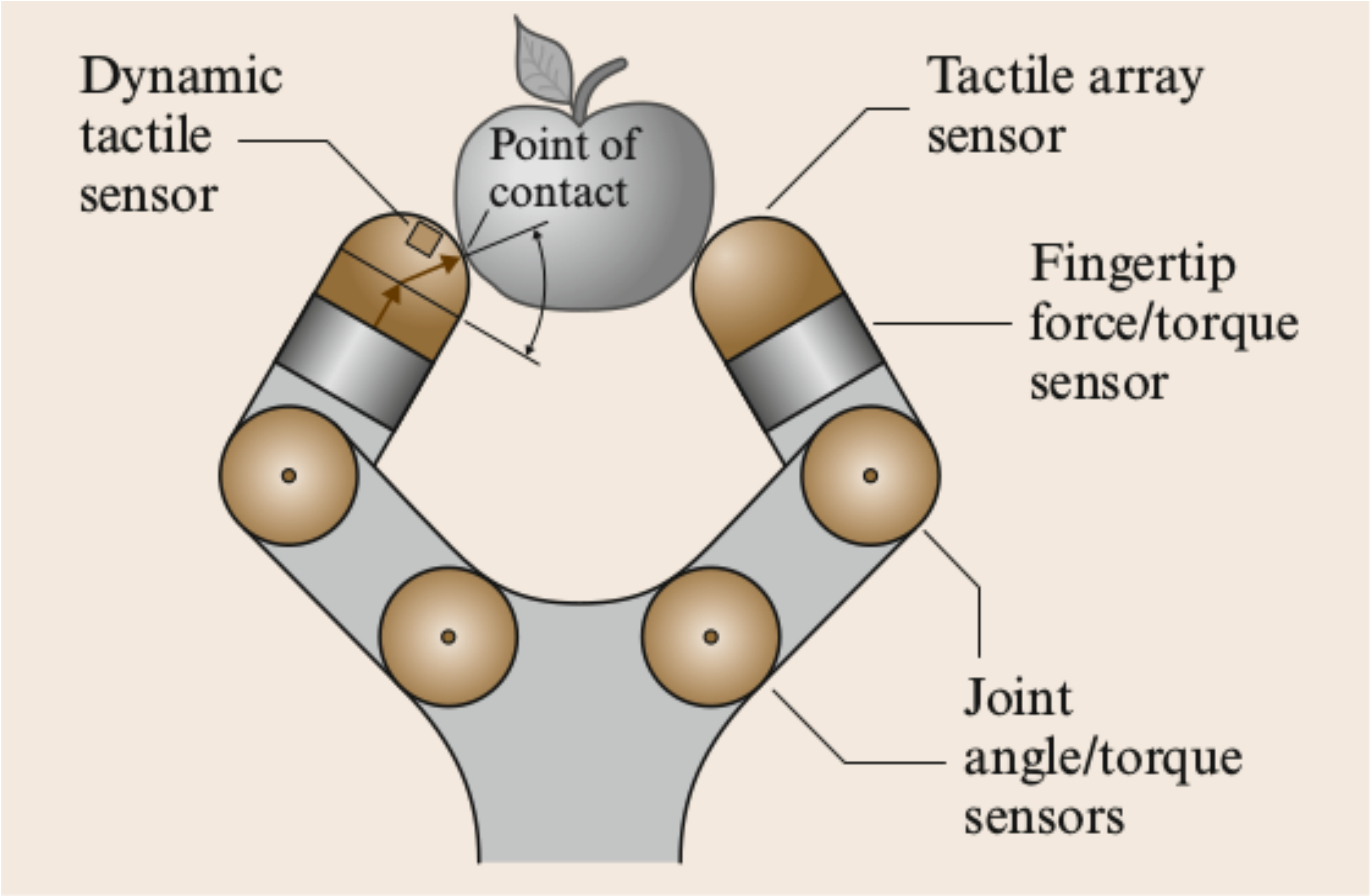
- Most important measured quantities: **force** and **shape**
  - Average or distribution across contact area

5 main types of sensors

- Proprioceptive
- Kinematic
- Force
- Dynamic tactile
- Array tactile

Other sensors:

- Thermal, material composition, etc.



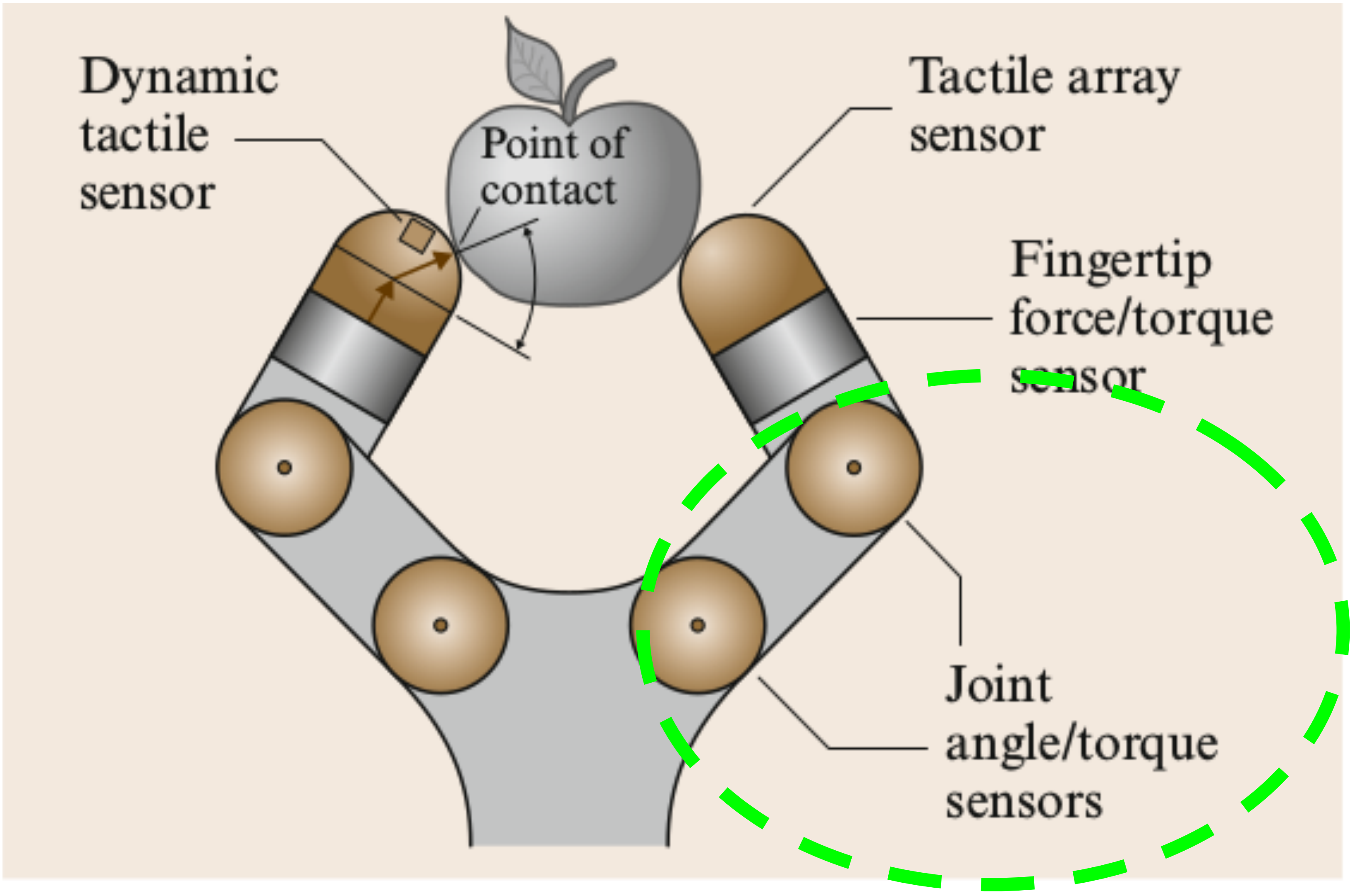
Sensors integrated in a robotic hand.

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Proprioceptive and Proximity Sensing

- **Spatial proprioception**
  To measure the net force or motion of an appendage.
  - Joint angle
  - Force-torque sensors

- **Proximity sensors**
  To explore the environment, and detect collisions.
  - Contact (Whiskers & Antennae)
  - Non contact (Infrared, Ultrasonic, ...)



Robot hand with joint angle sensors.

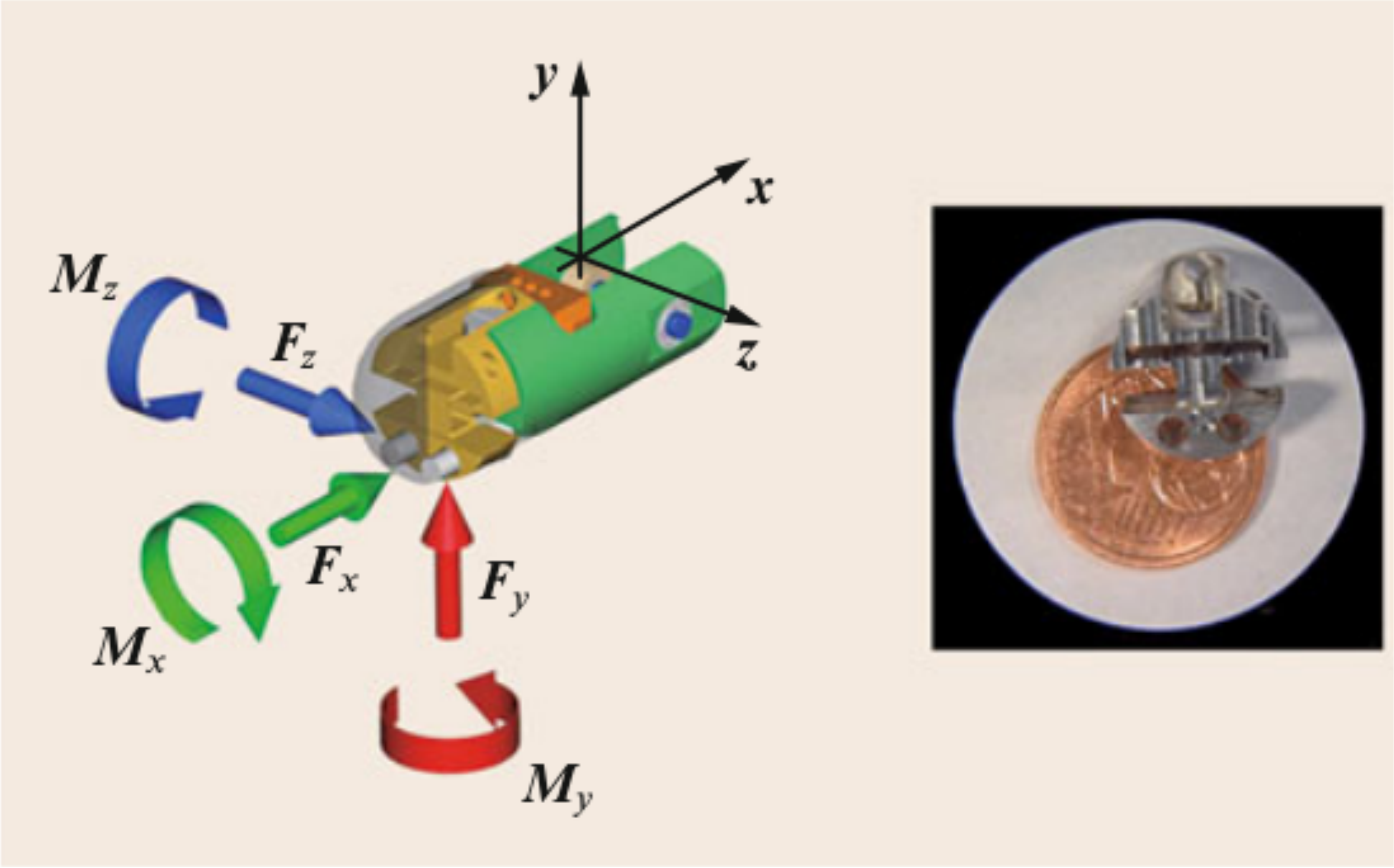
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Force and Load Sensing

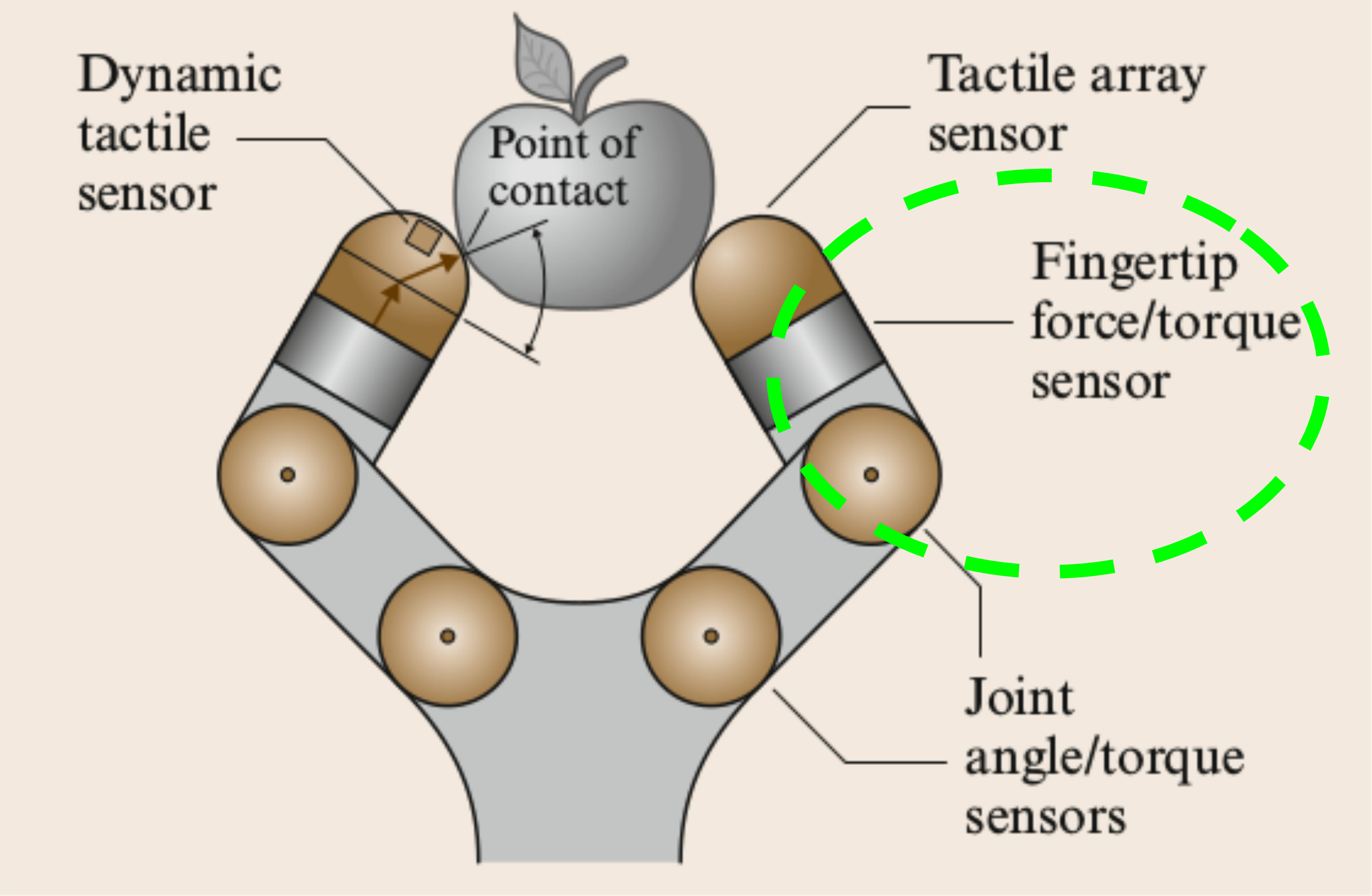
To measure contact forces.

- **Actuator Effort Sensors**
  Servo motors → motor current
  Cables → cable tension

- **Force Sensors**
  - Mounted at the base joint, wrist, or distributed.



Multi-axis fingertip force-torque sensor.



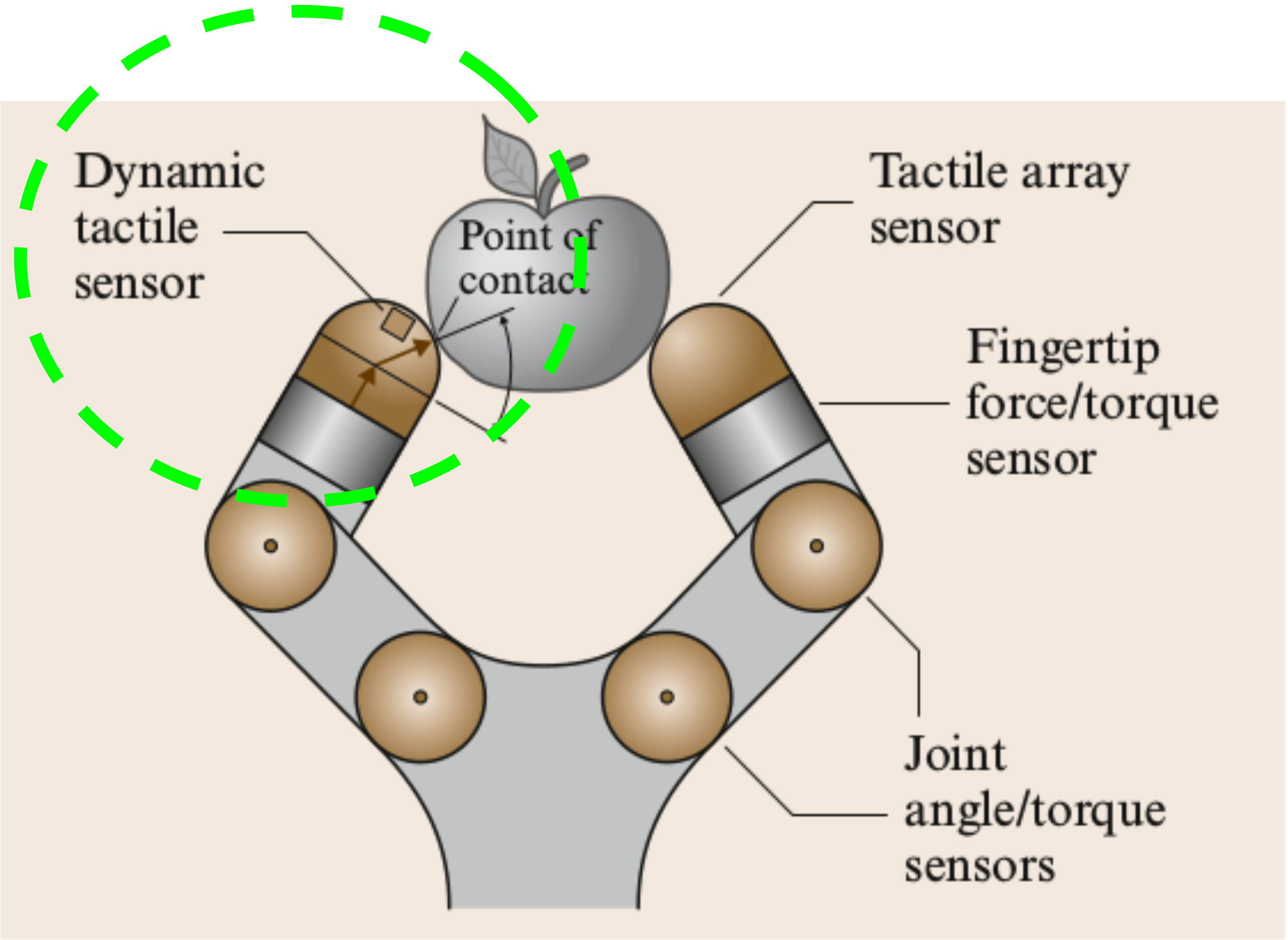
Force sensors + fingertip geometry to estimate contact location.

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

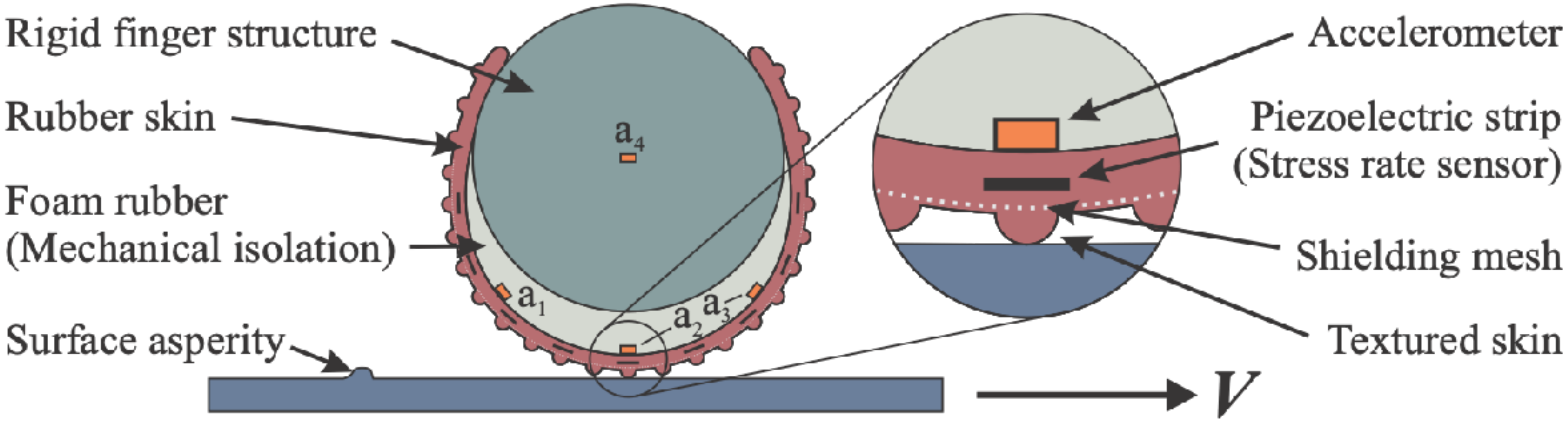
Dynamic Tactile Sensors

To detect slips, and to sense textures and features.

Ex. **Capacitive tactile sensing (video)**



Robot hand with dynamic tactile sensor.



Cutkosky, Mark R., and John Ulmen. "Dynamic tactile sensing." In *The Human Hand as an Inspiration for Robot Hand Development*, pp. 389-403. Springer, Cham, 2014.

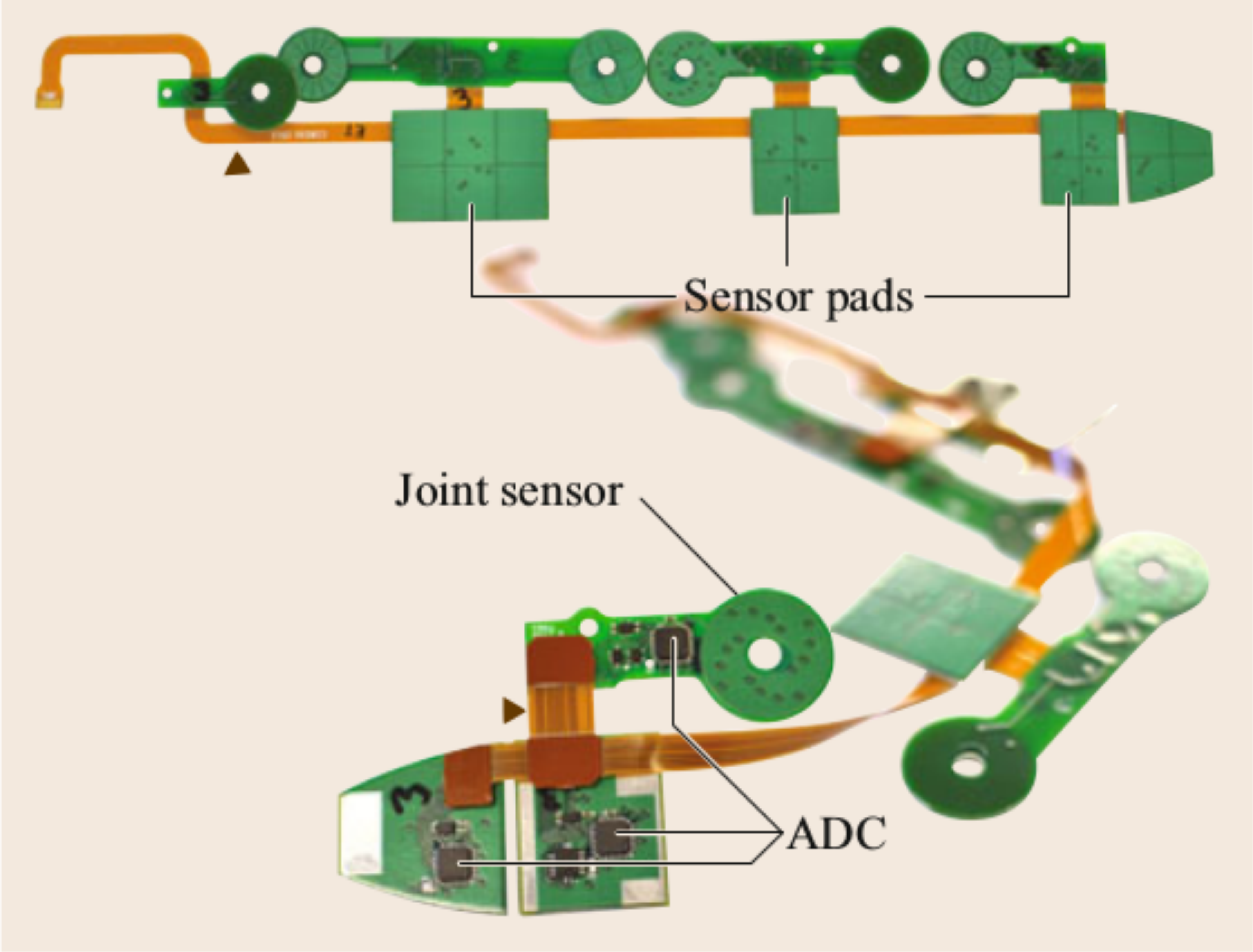
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Array Sensors

To sense shape and pressure.

- **Contact Location Sensors**
  - 2D switch array

- **Pressure Sensing Arrays**



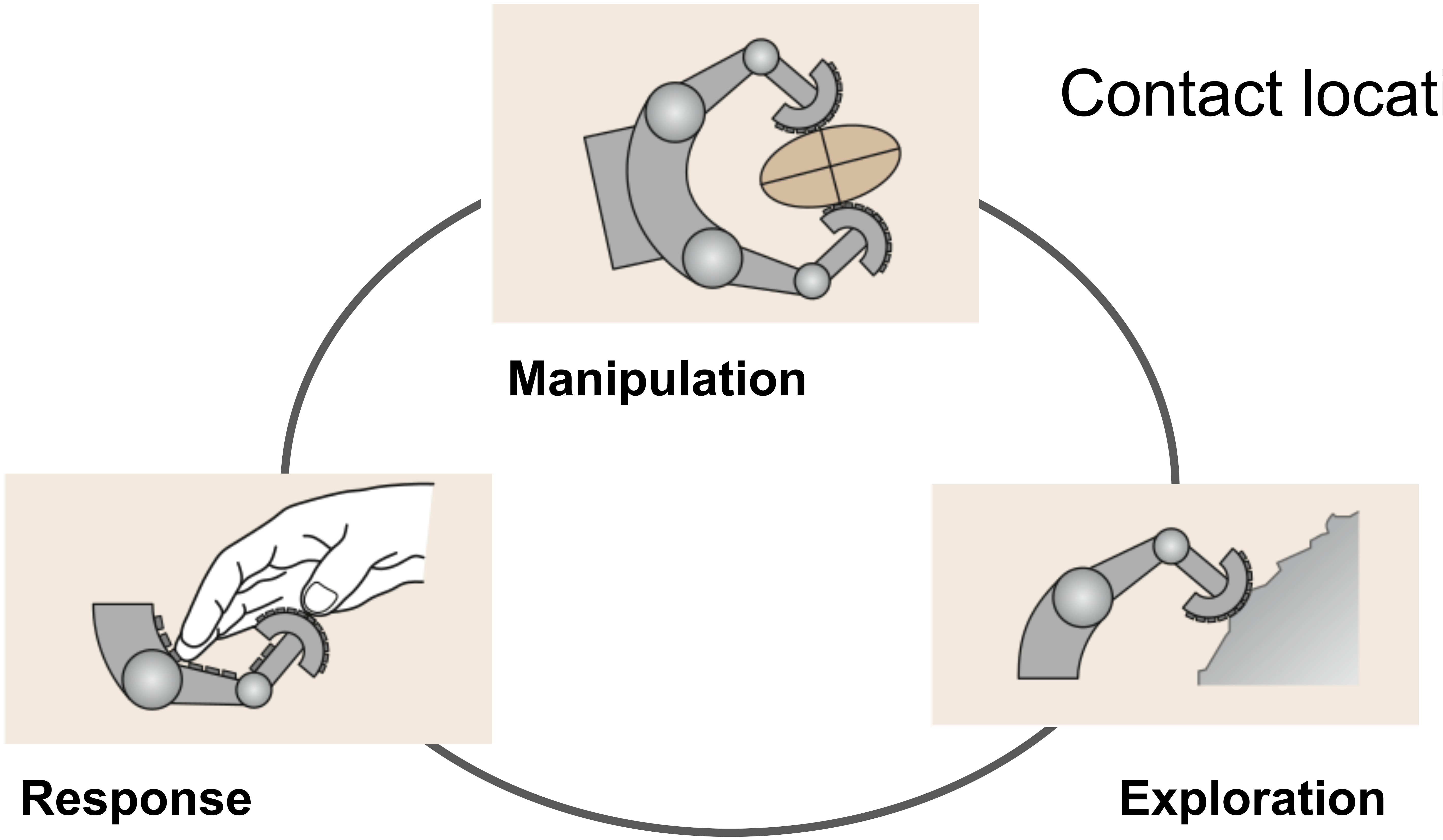
Capacitive touch and joint angle sensors on a flexible circuit for incorporation in a robotic hand.

Ex. **The effect of twice dropping, and then gently placing, a two-gram weight on a small capacitive tactile array.** (video)

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Tactile Information Processing

Contact locations and forces

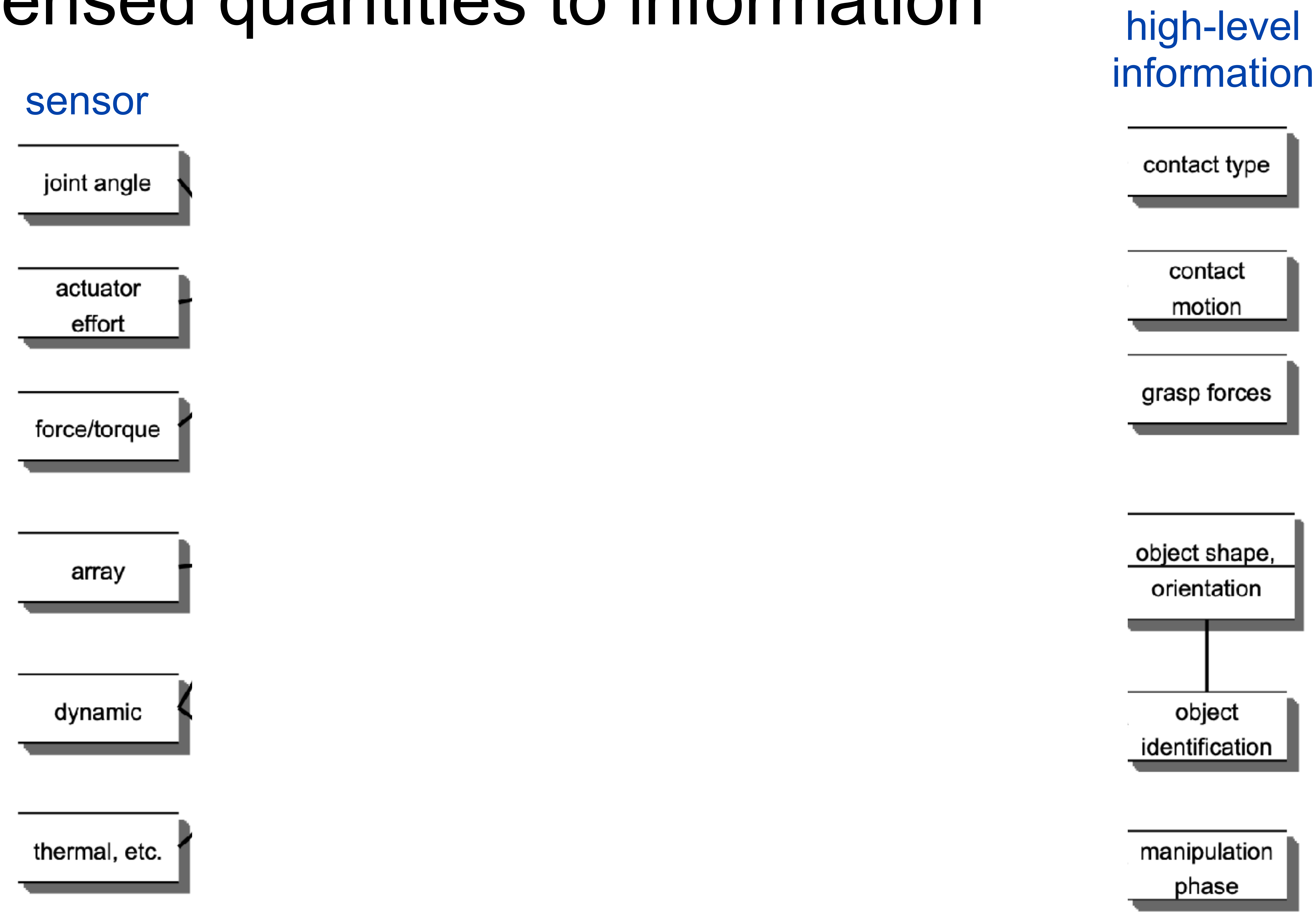


Event detection

Geometric and material properties of the object

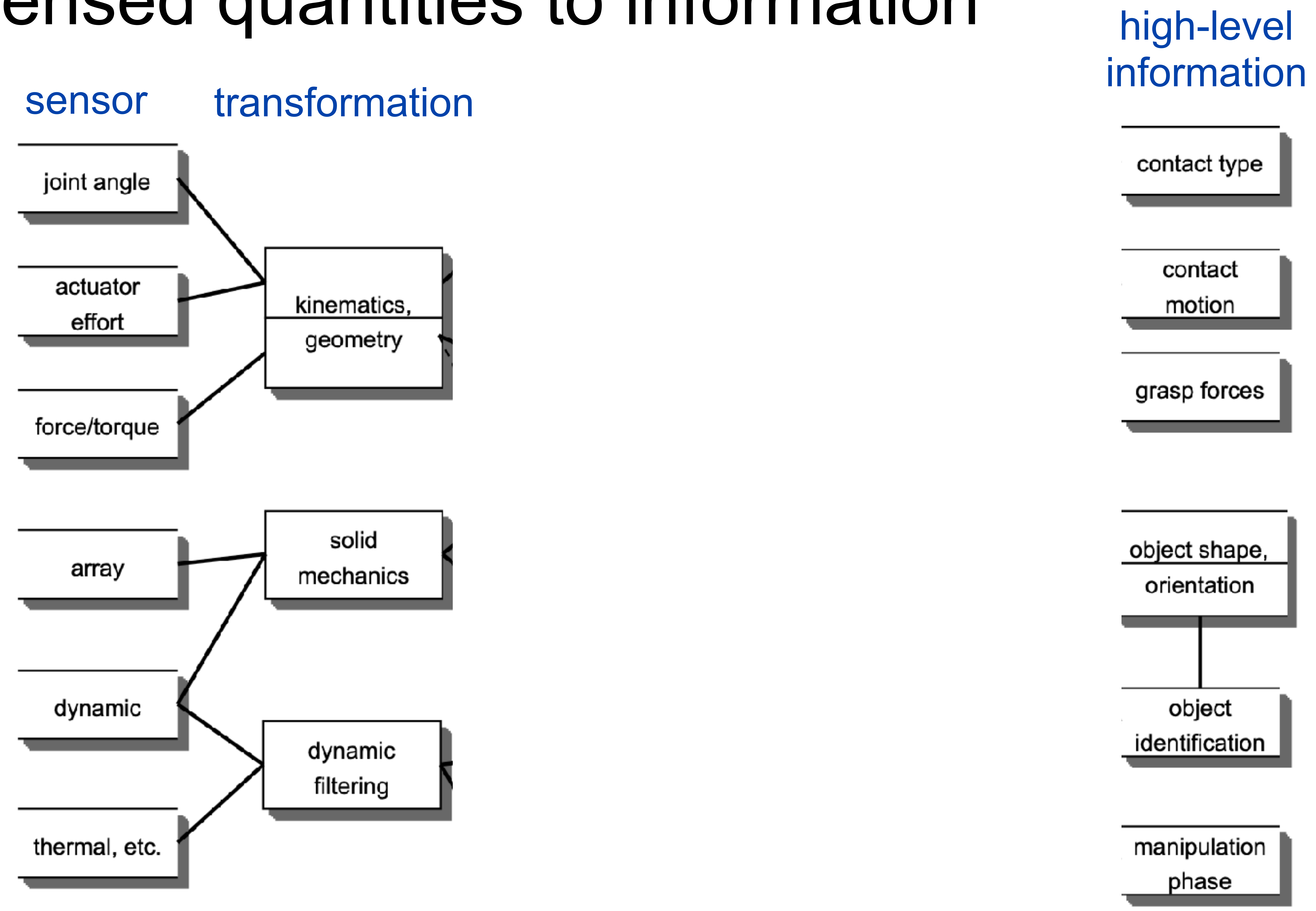
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

From sensed quantities to information



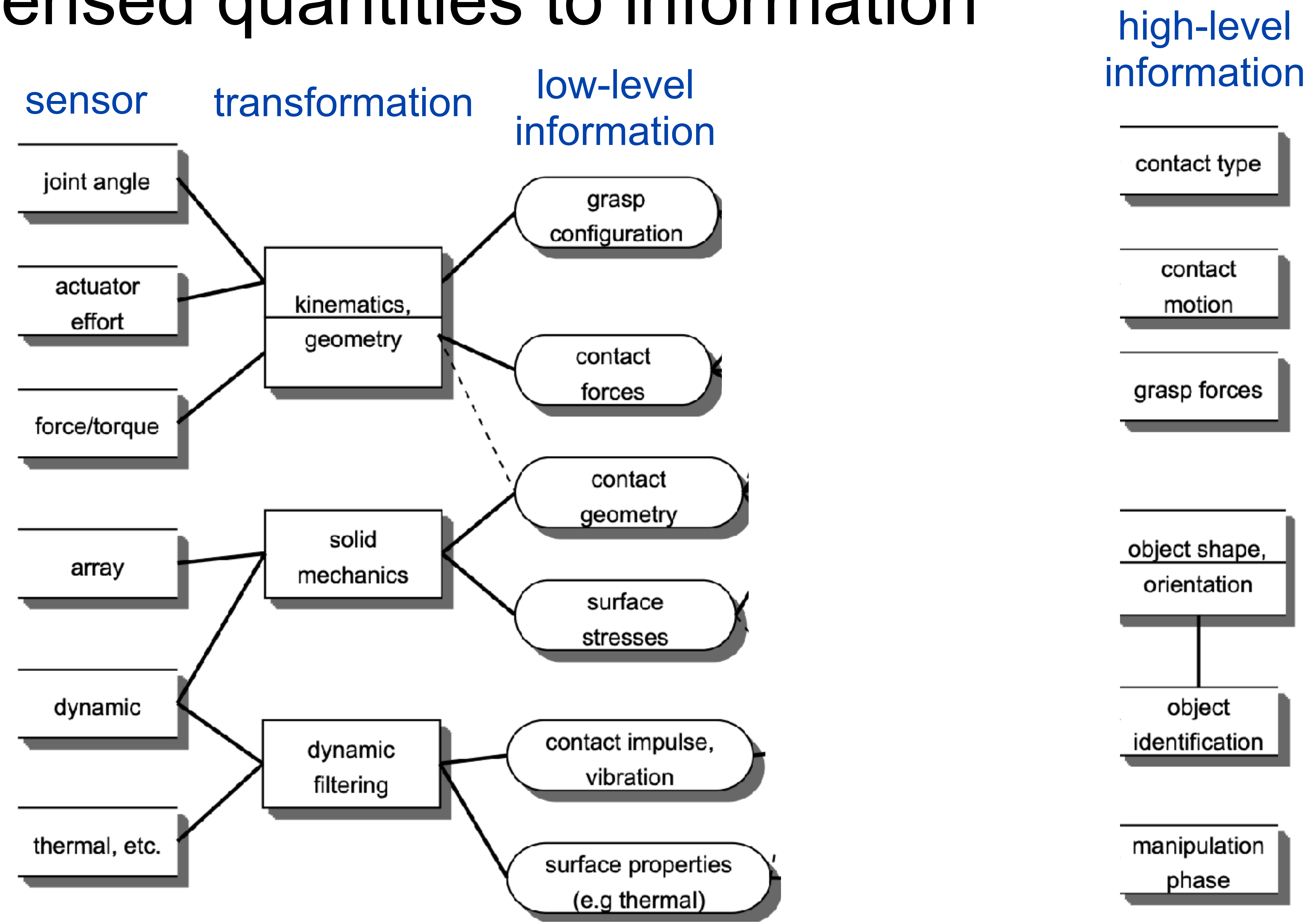
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

From sensed quantities to information



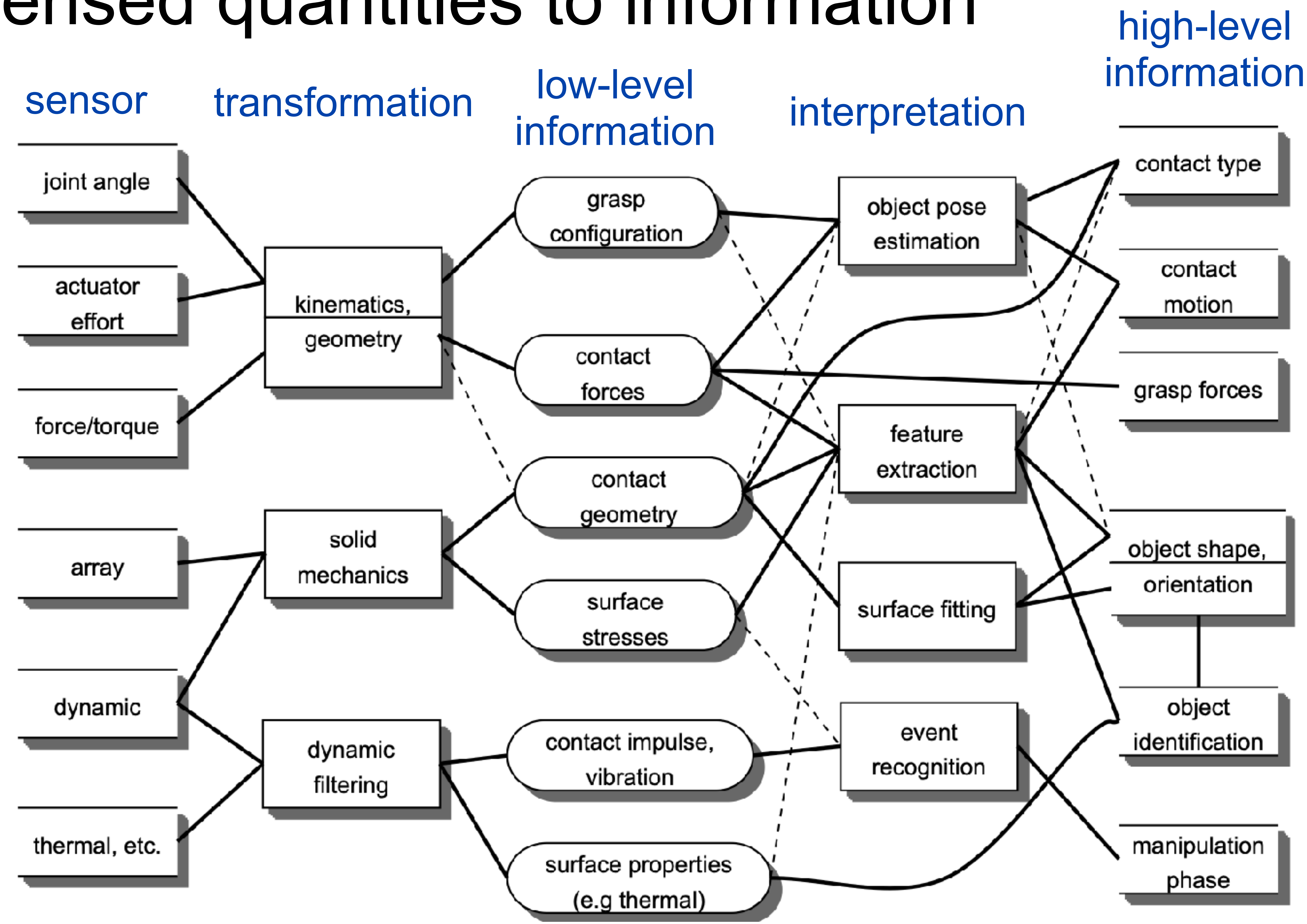
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

From sensed quantities to information



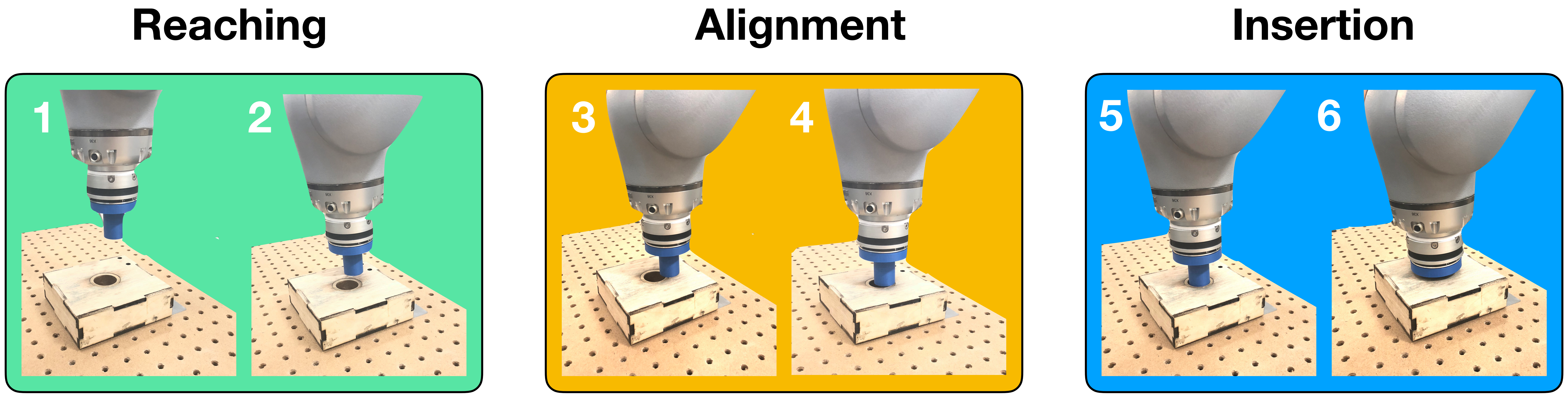
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

From sensed quantities to information



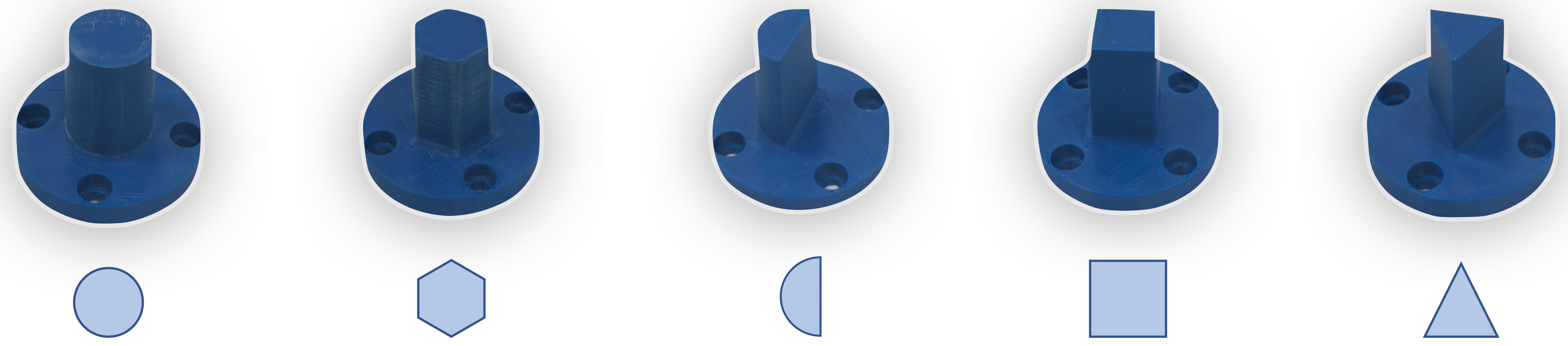
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Multimodal Representation for Manipulation

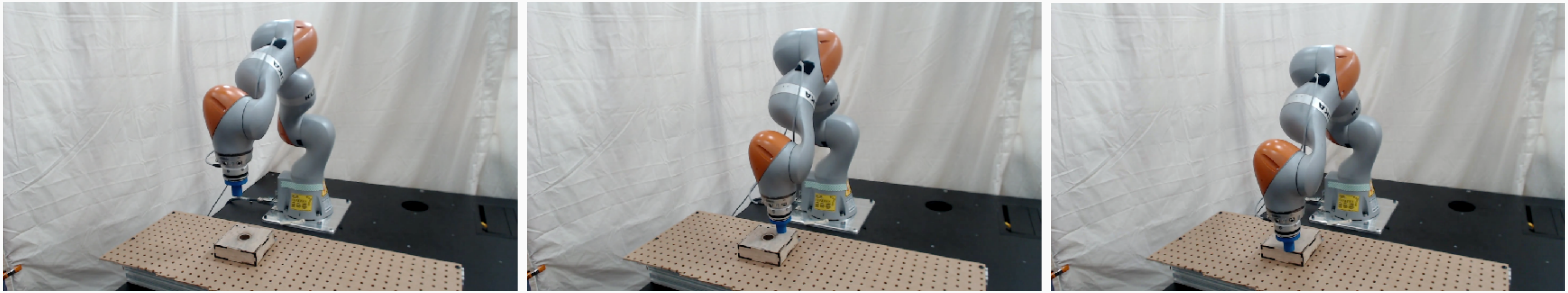


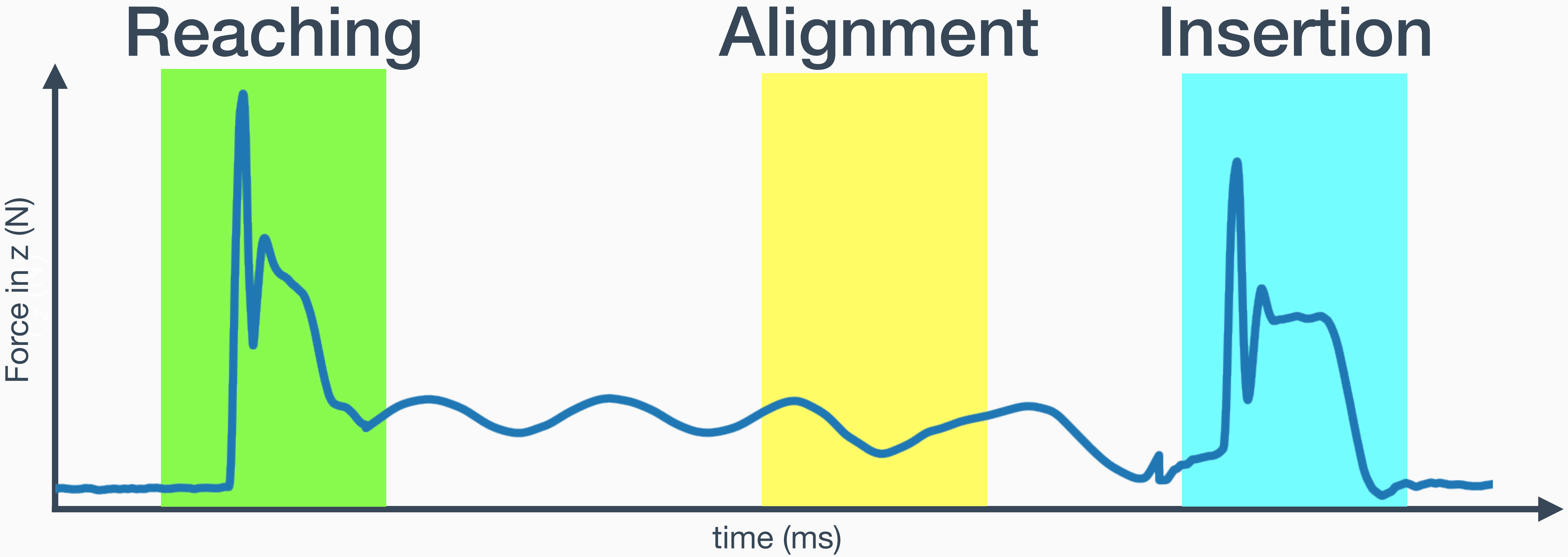
Peg Insertion

RGB Images + End Effector F/T

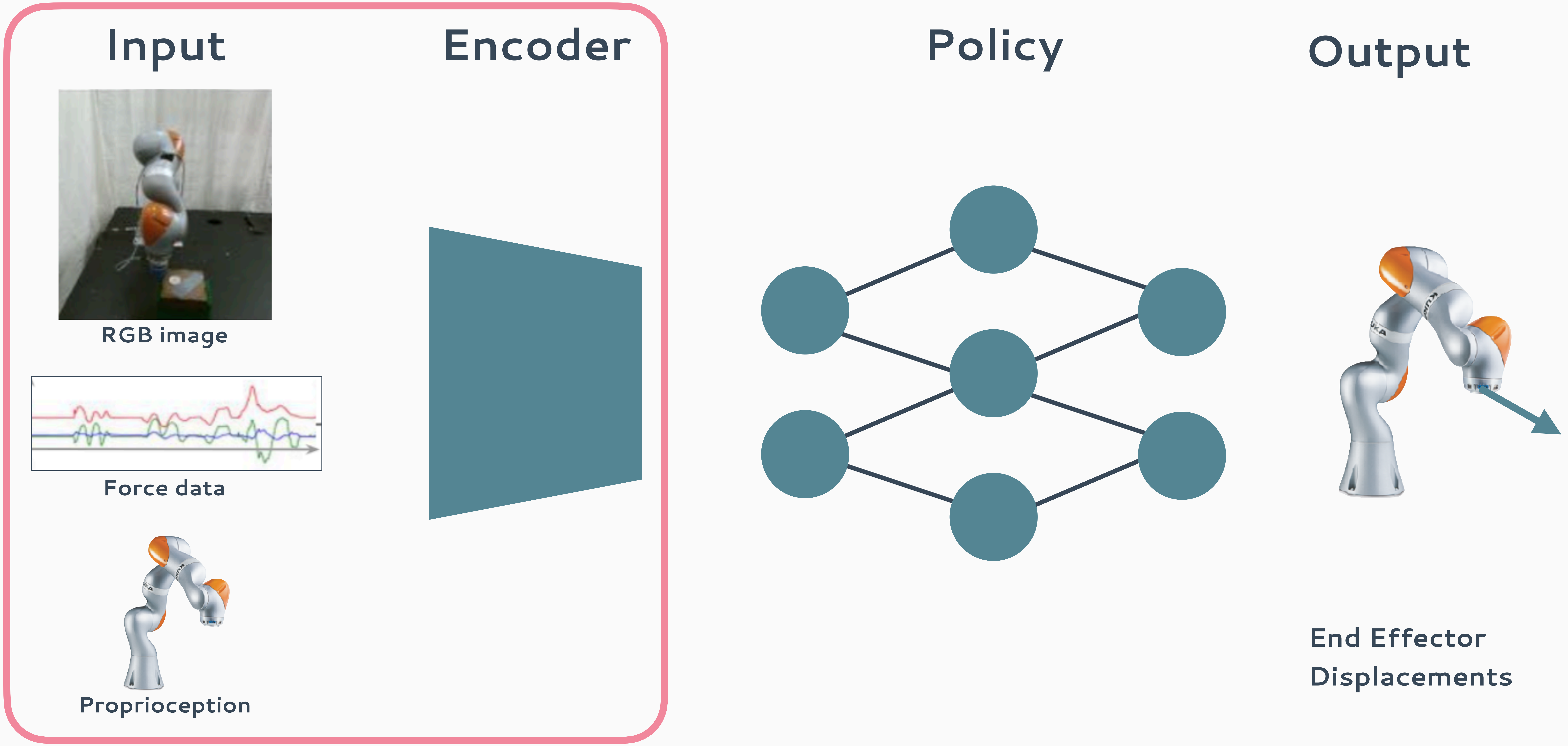


Vision and Touch are complementary

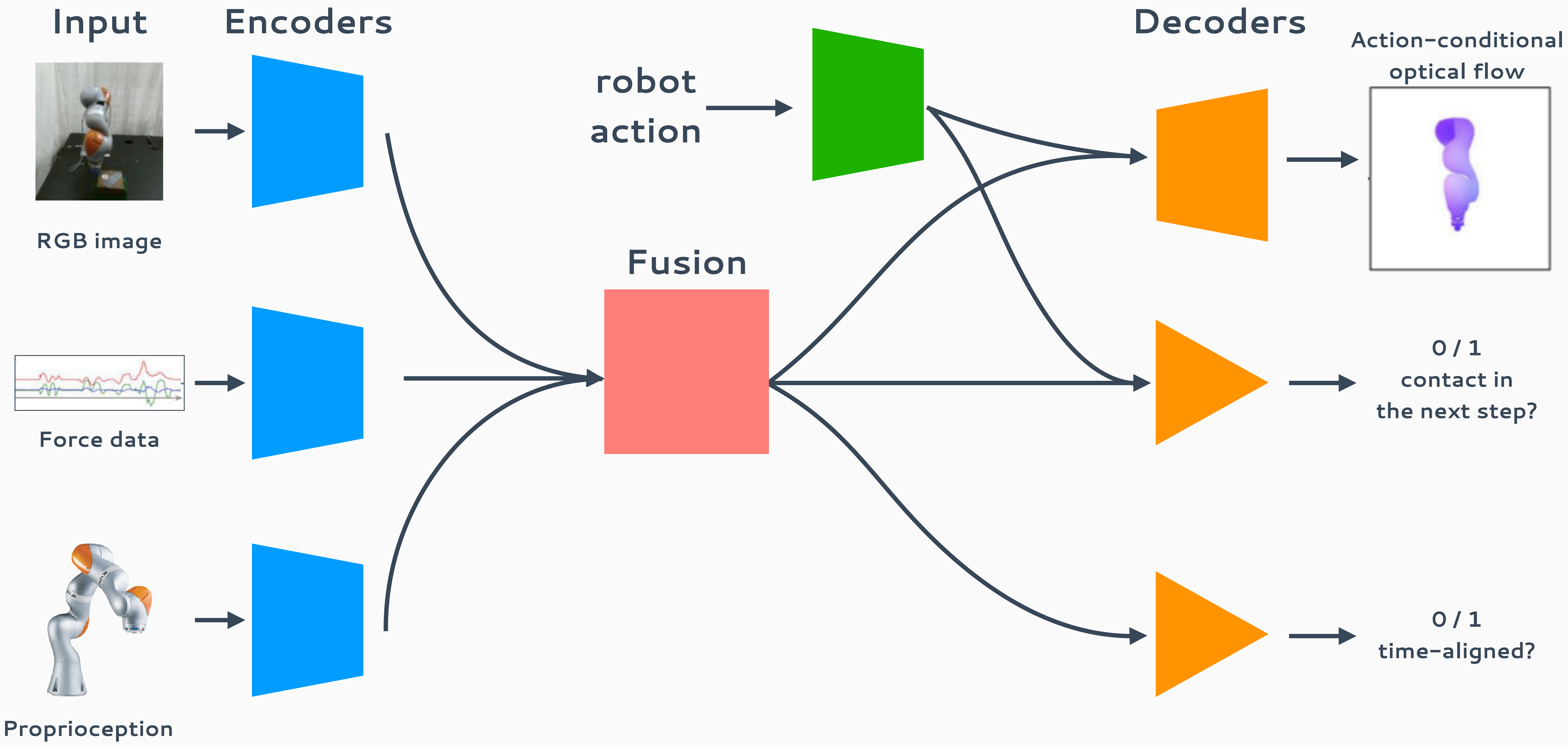




Learning a policy that leverages Vision & Touch



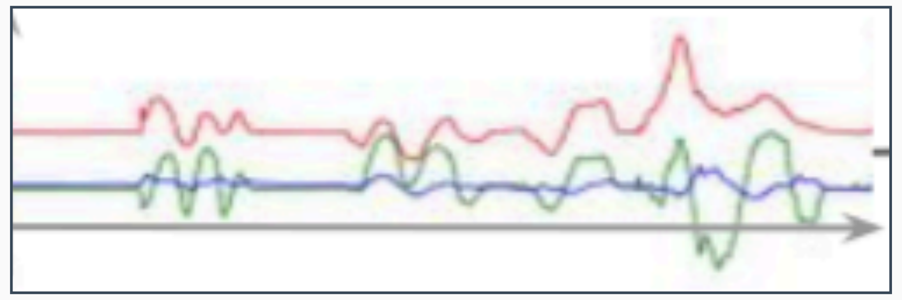
Representation Learning



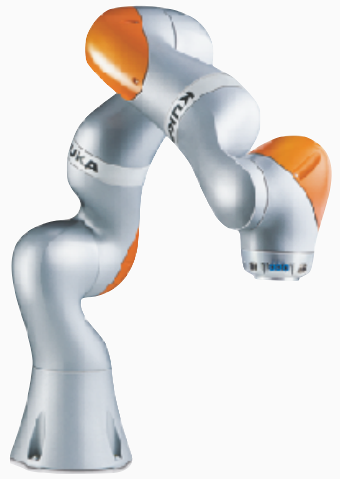
Learning a Policy based on this Representation

Input

RGB image



Force data

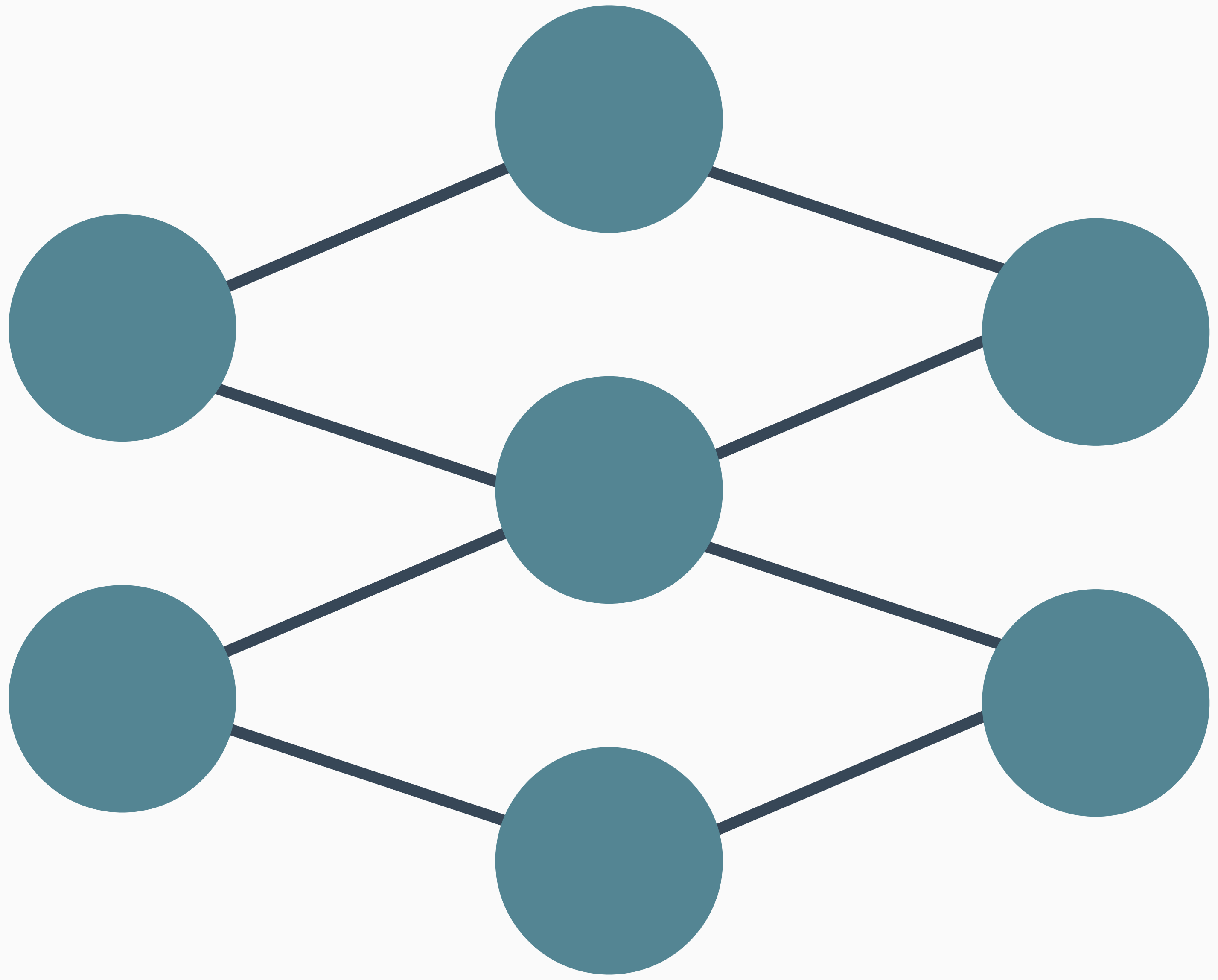


Proprioception

Encoder

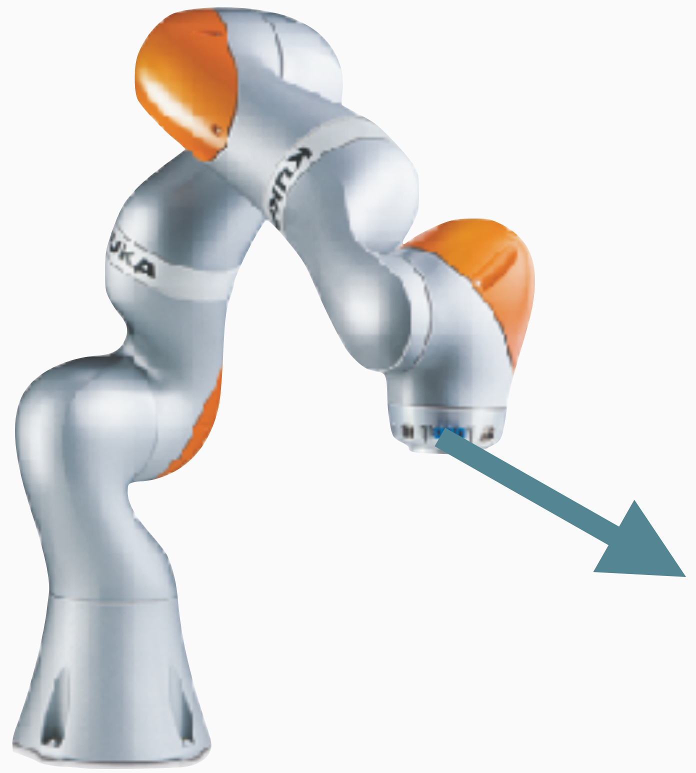
Learned Representation

Policy



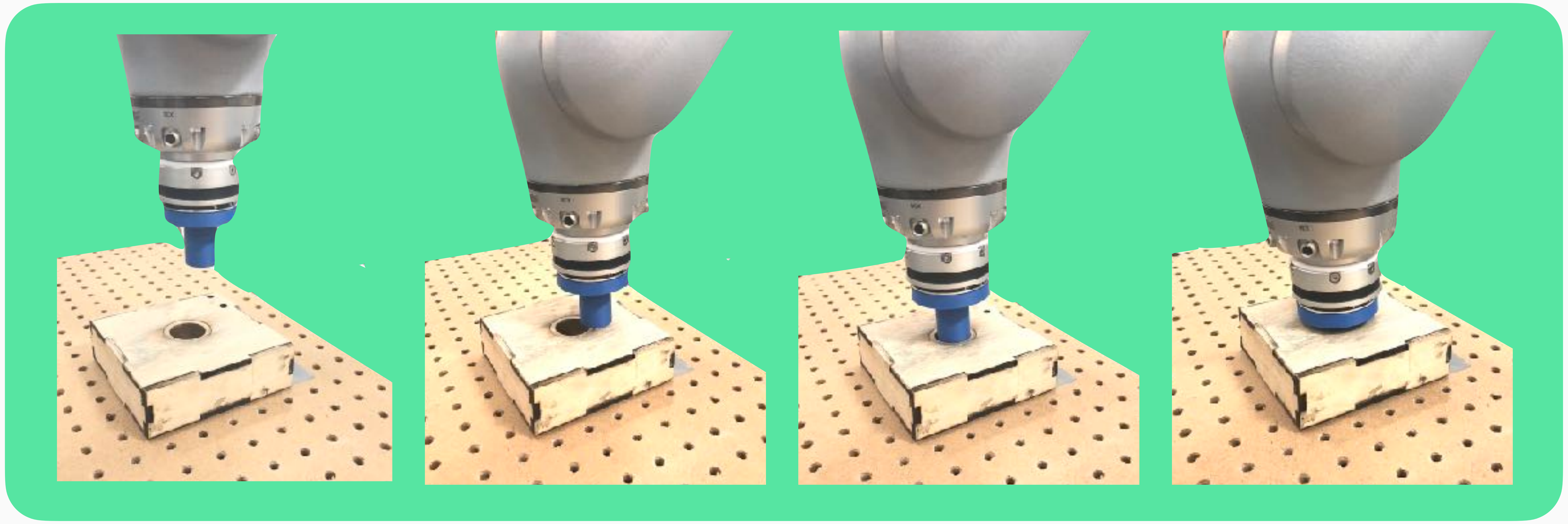
TRPO

Output

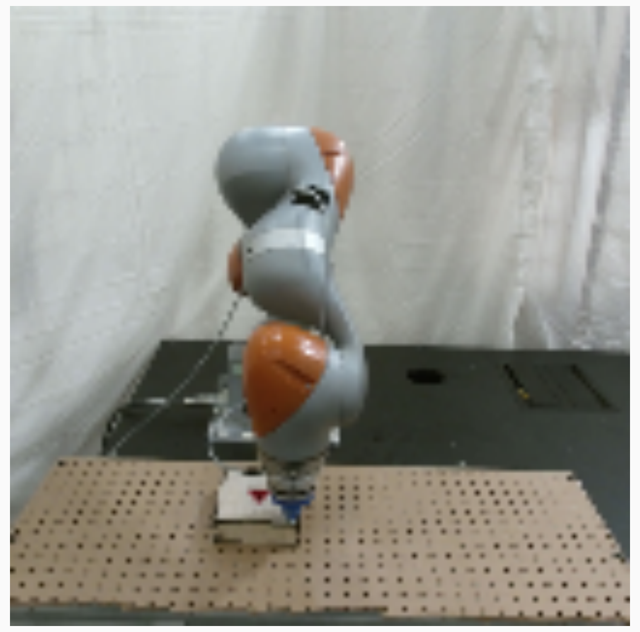


End Effector Displacements

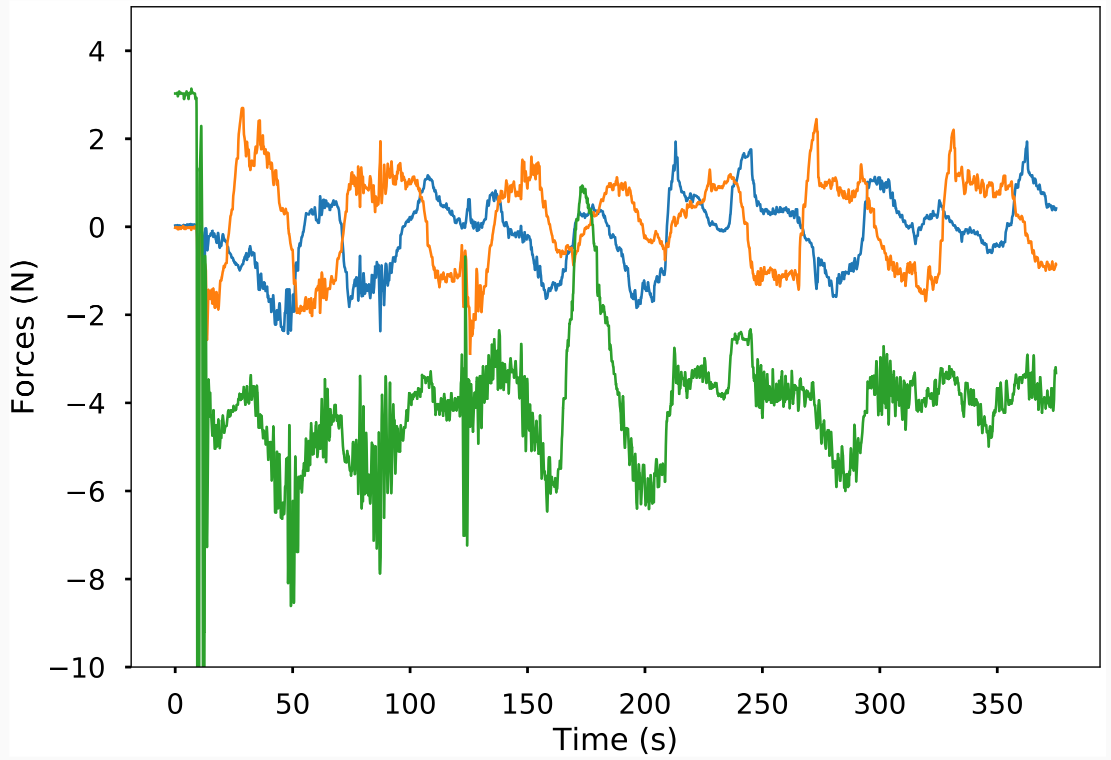
Experimental setup



Multimodal sensory inputs



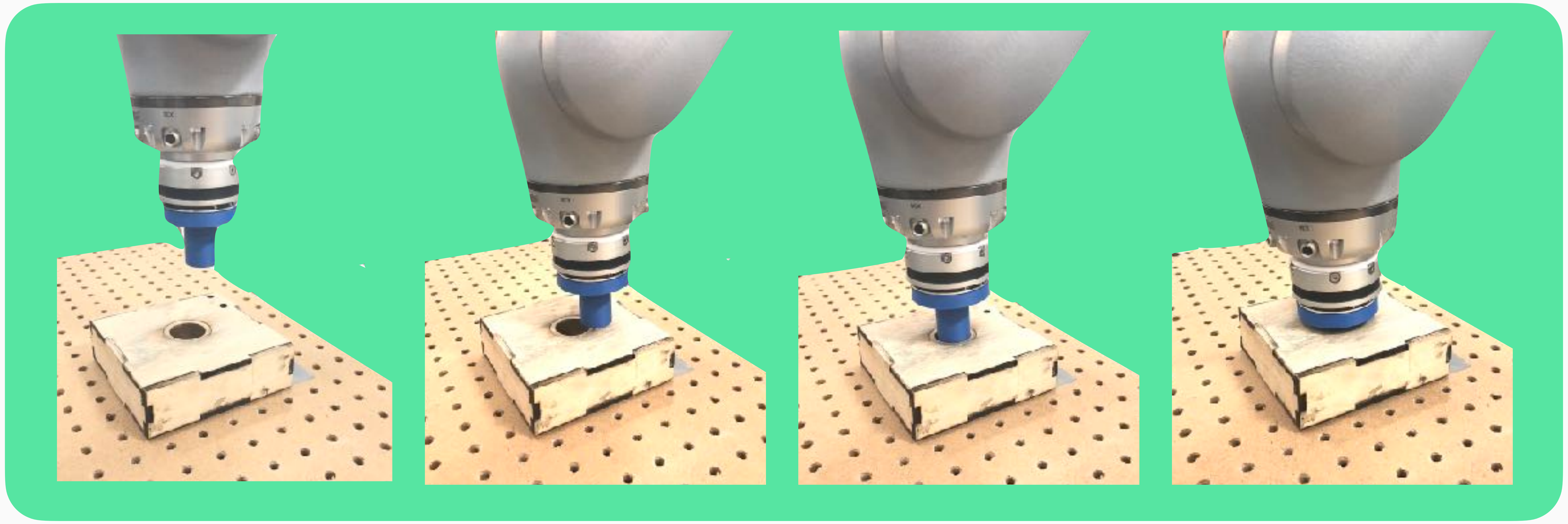
RGB



force/torque

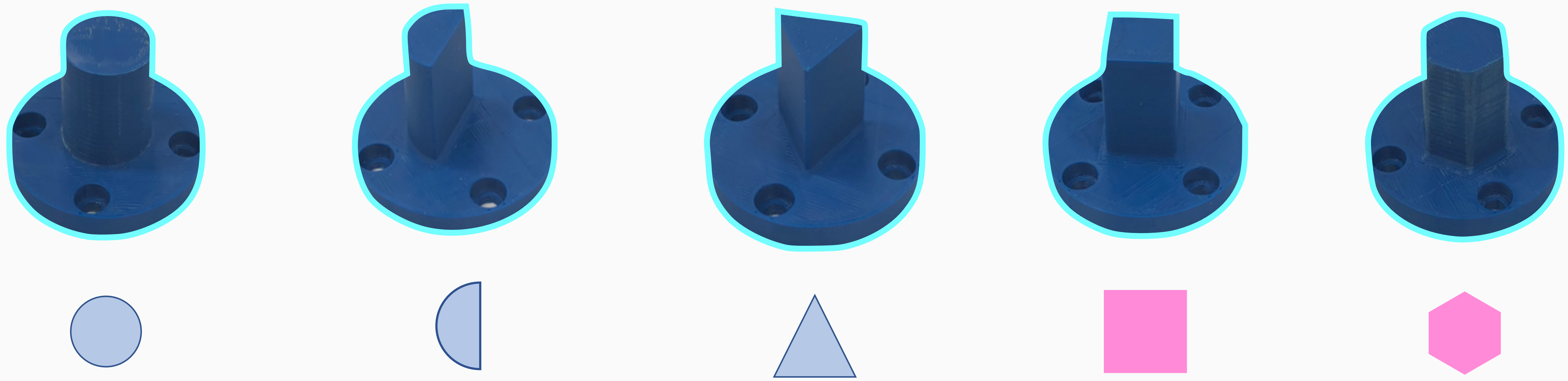
robot states

Experimental setup



Training Testing

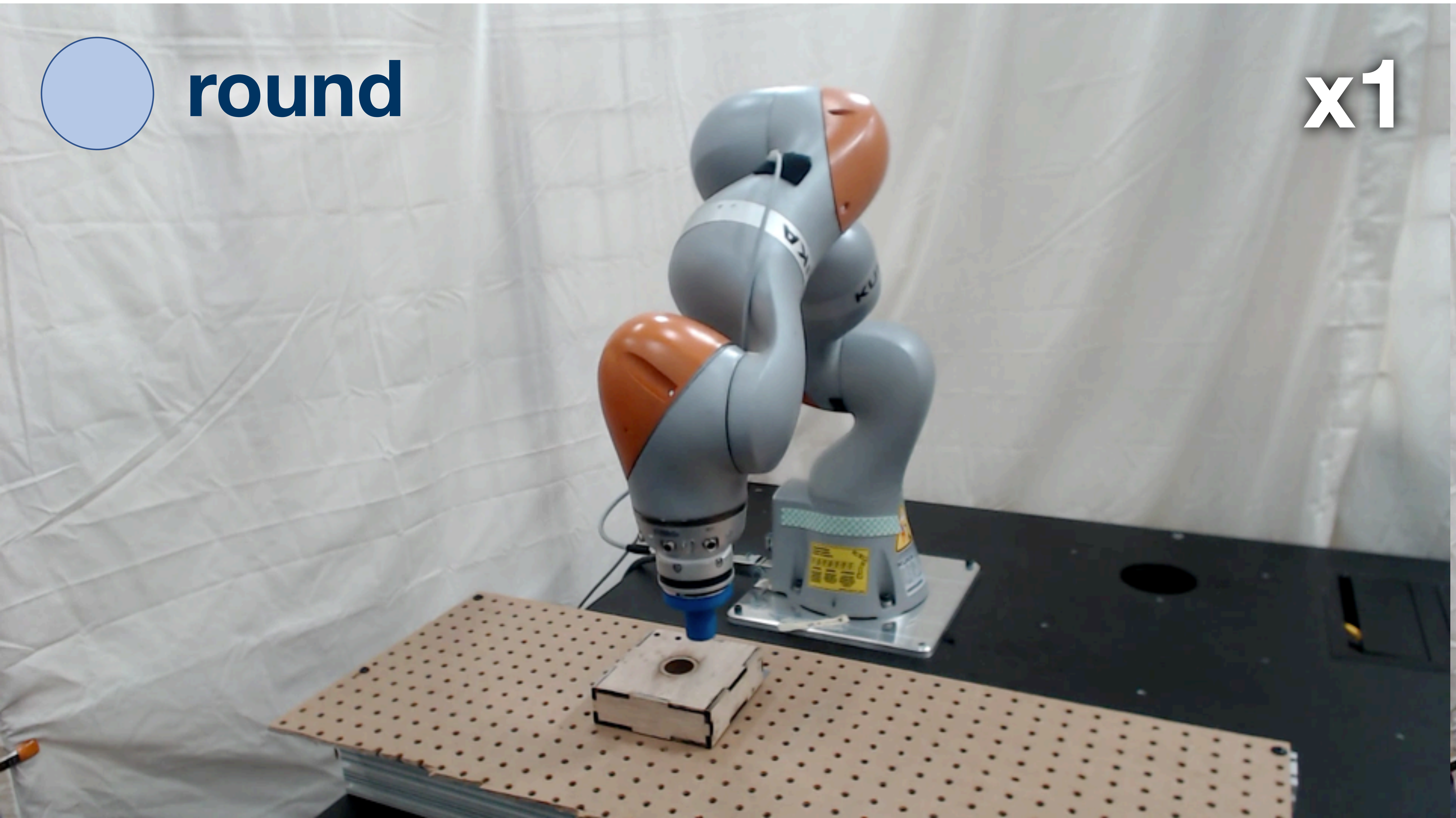
Peg geometry



We evaluate our representation with policy learning

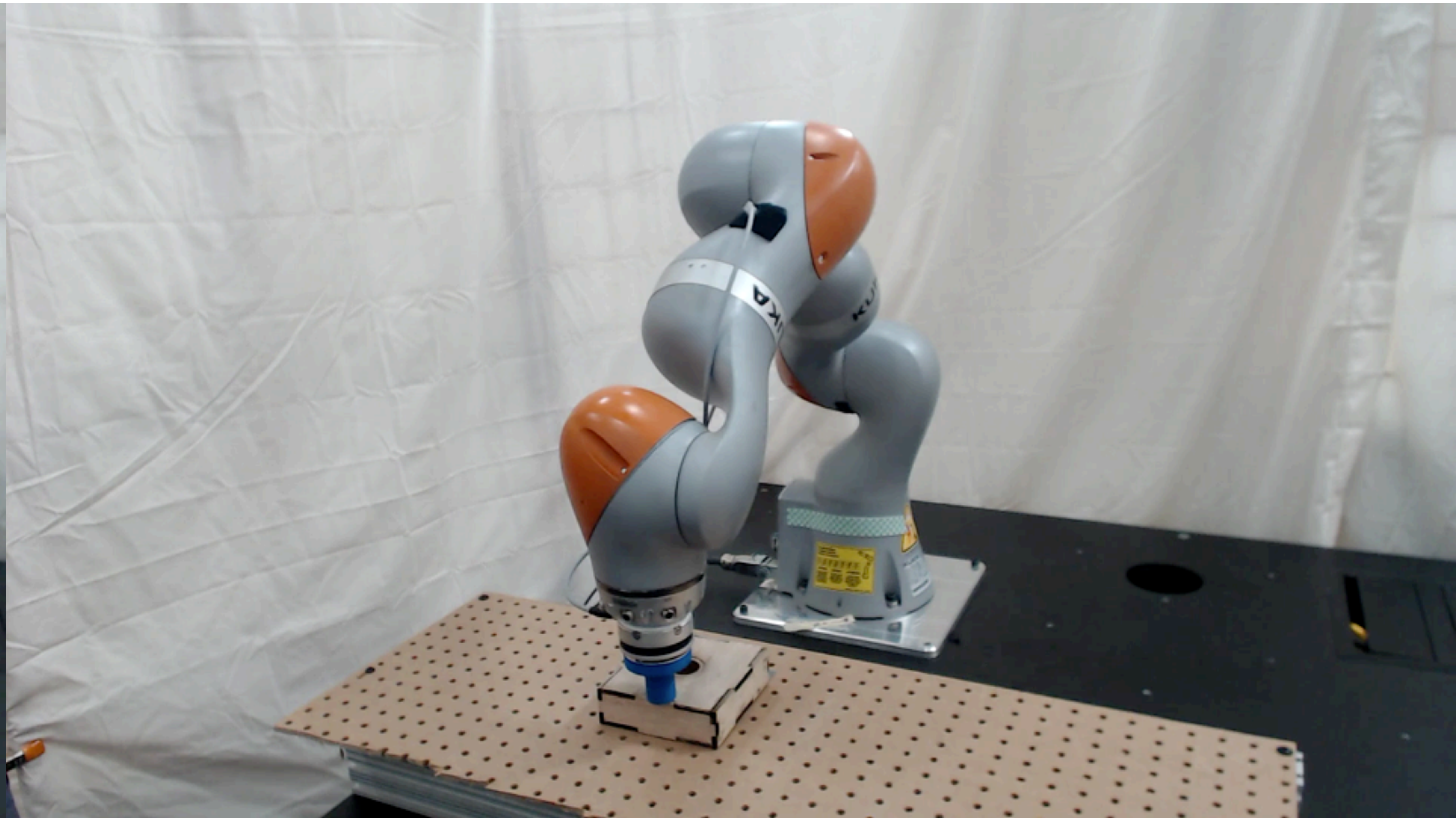
Episode 0

0% success rate

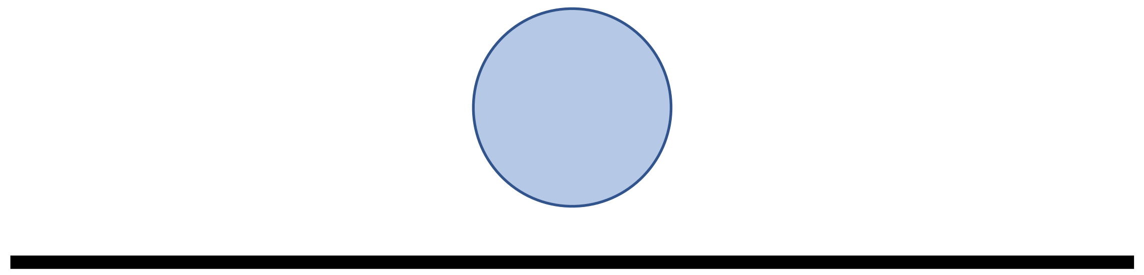


Episode 100

21% success rate

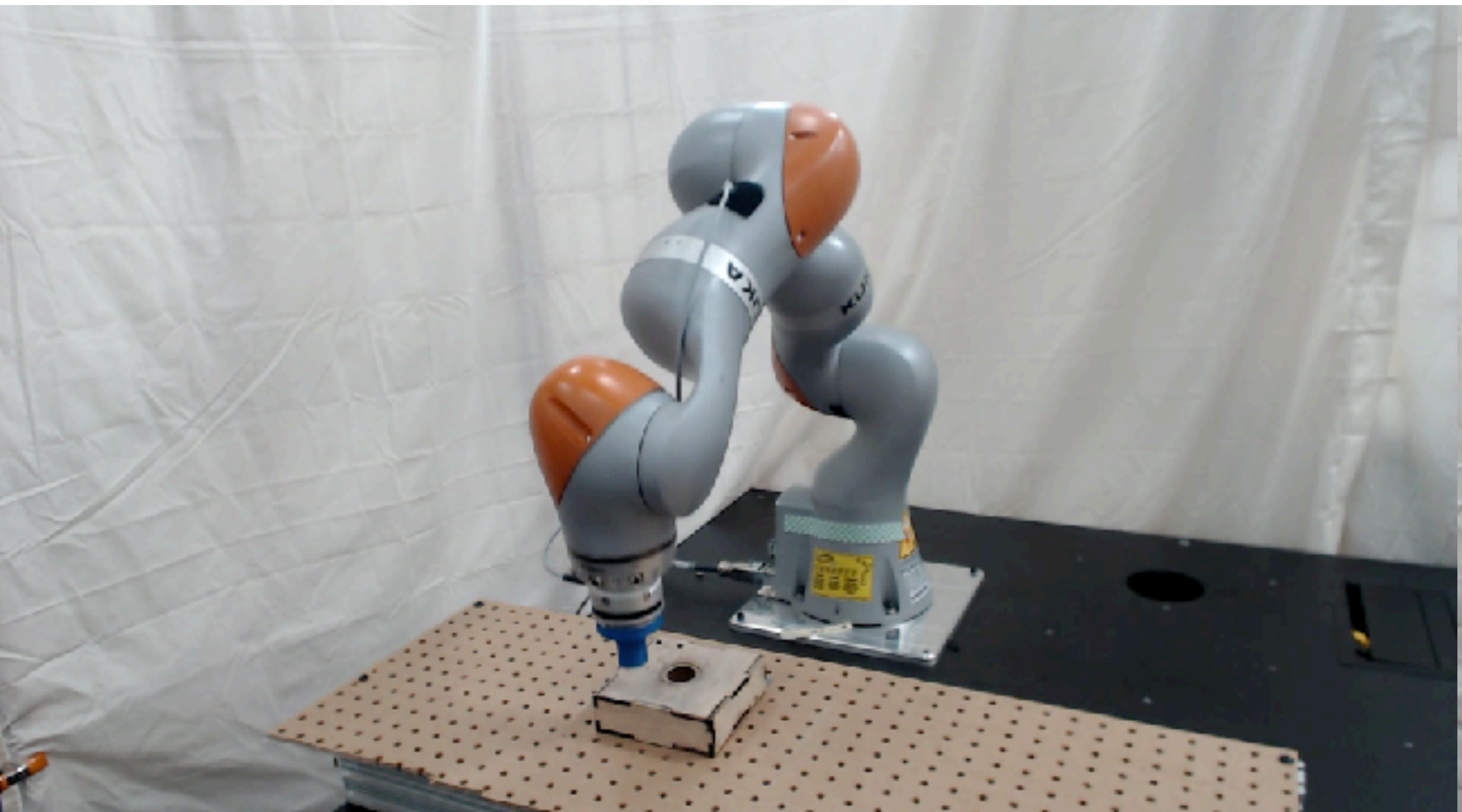


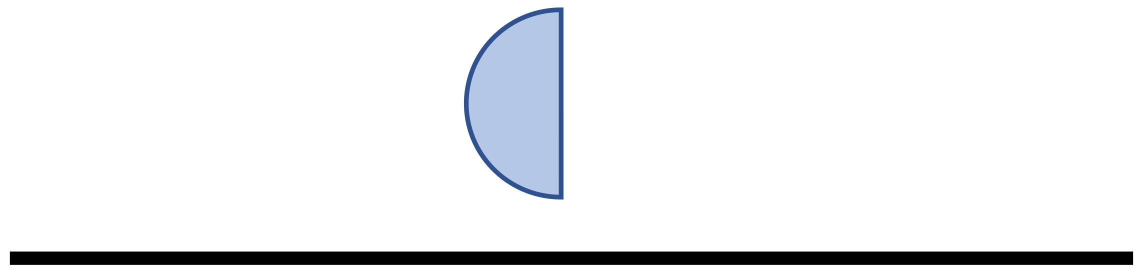
We efficiently learn policies in 5 hours



Episode 300

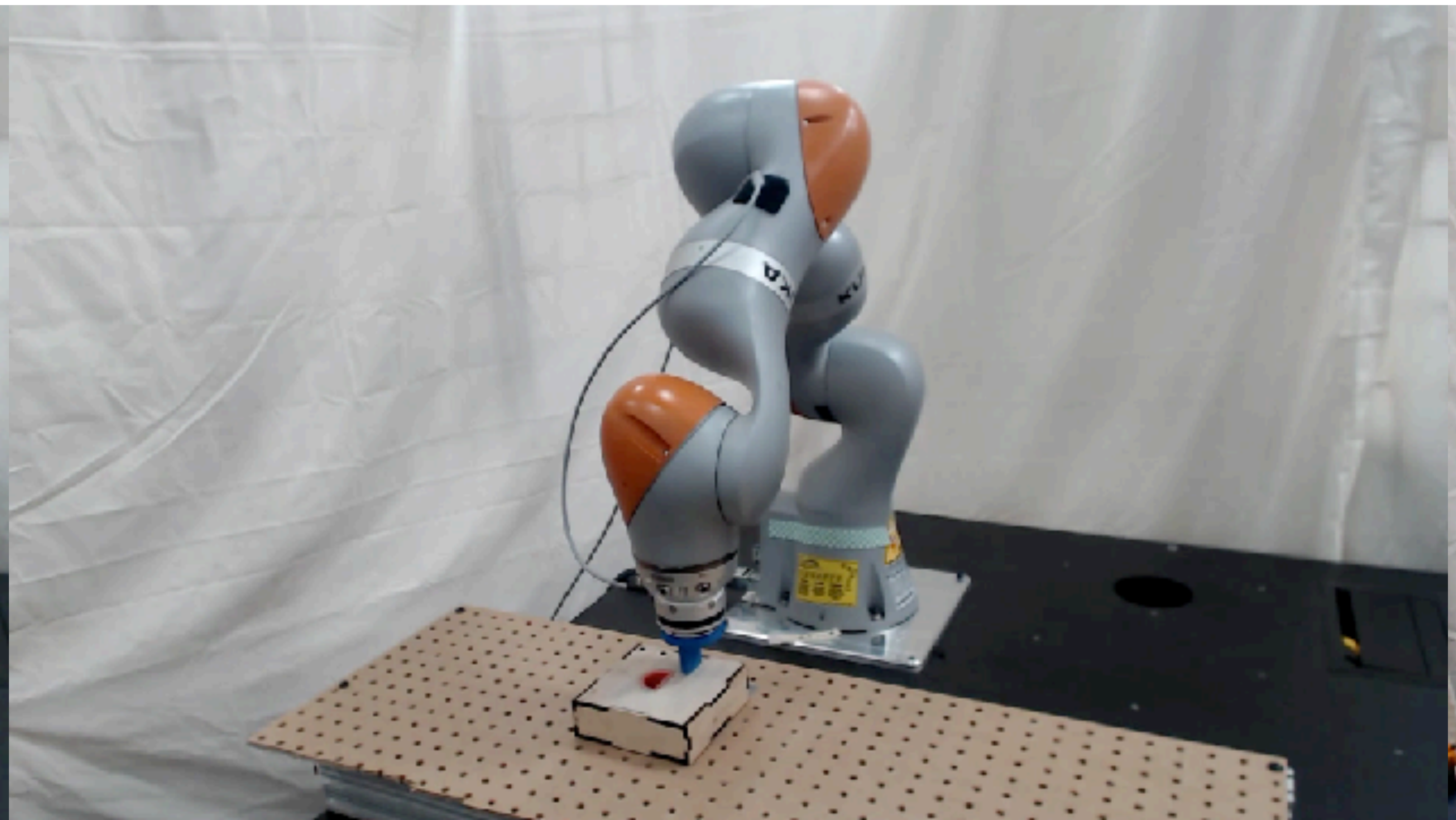
73% success rate

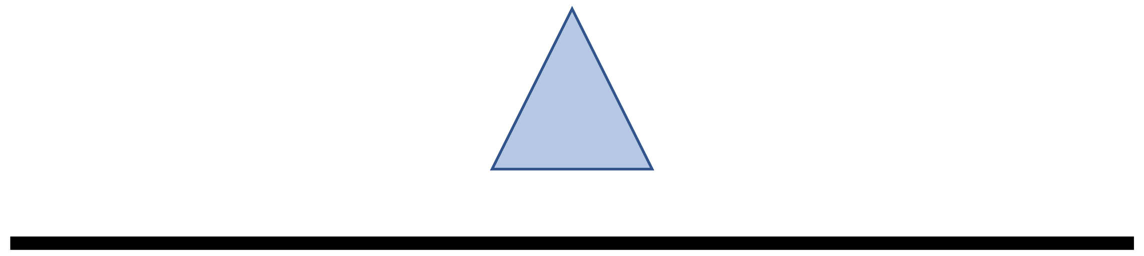




Episode 300

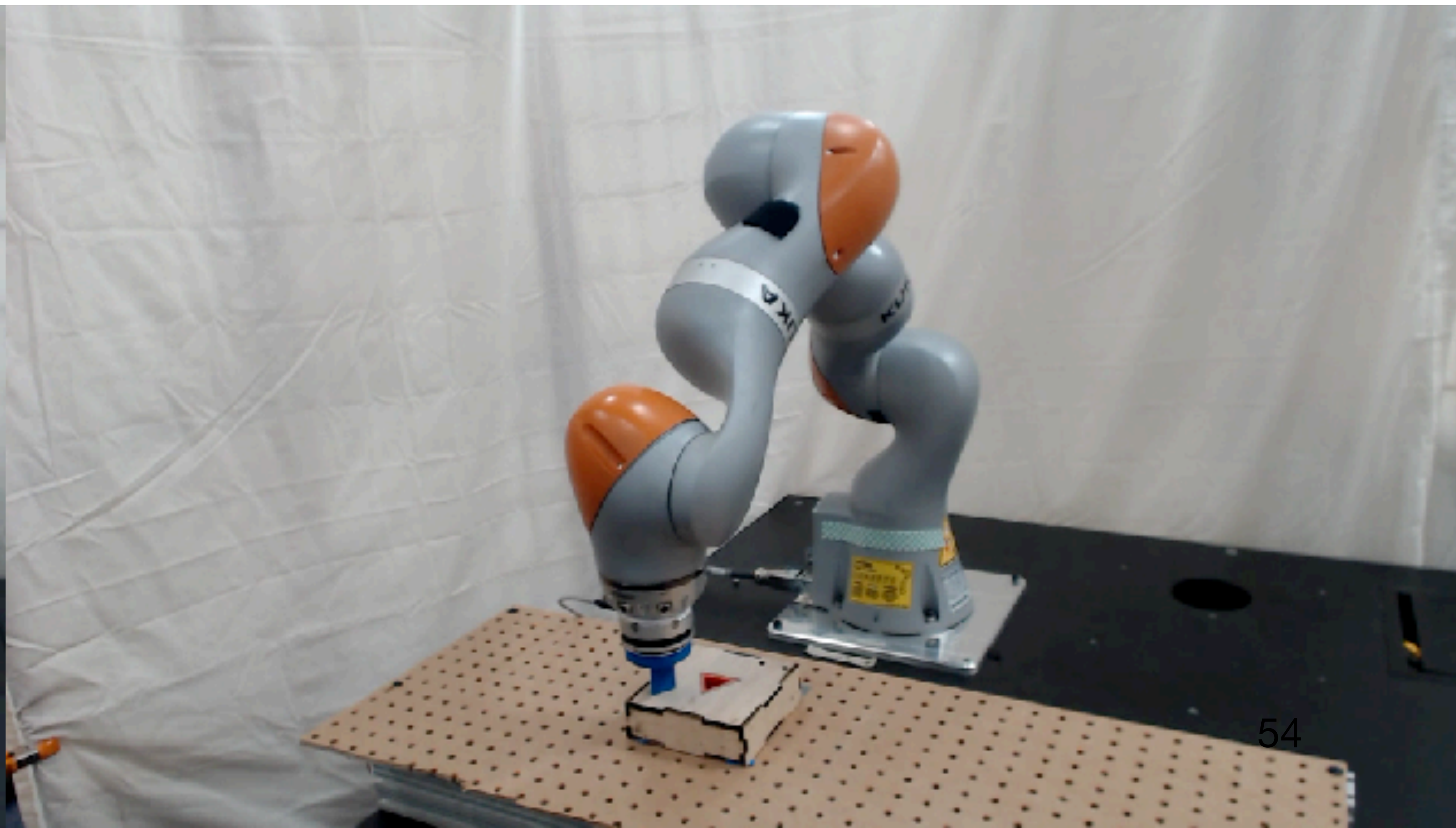
71% success rate



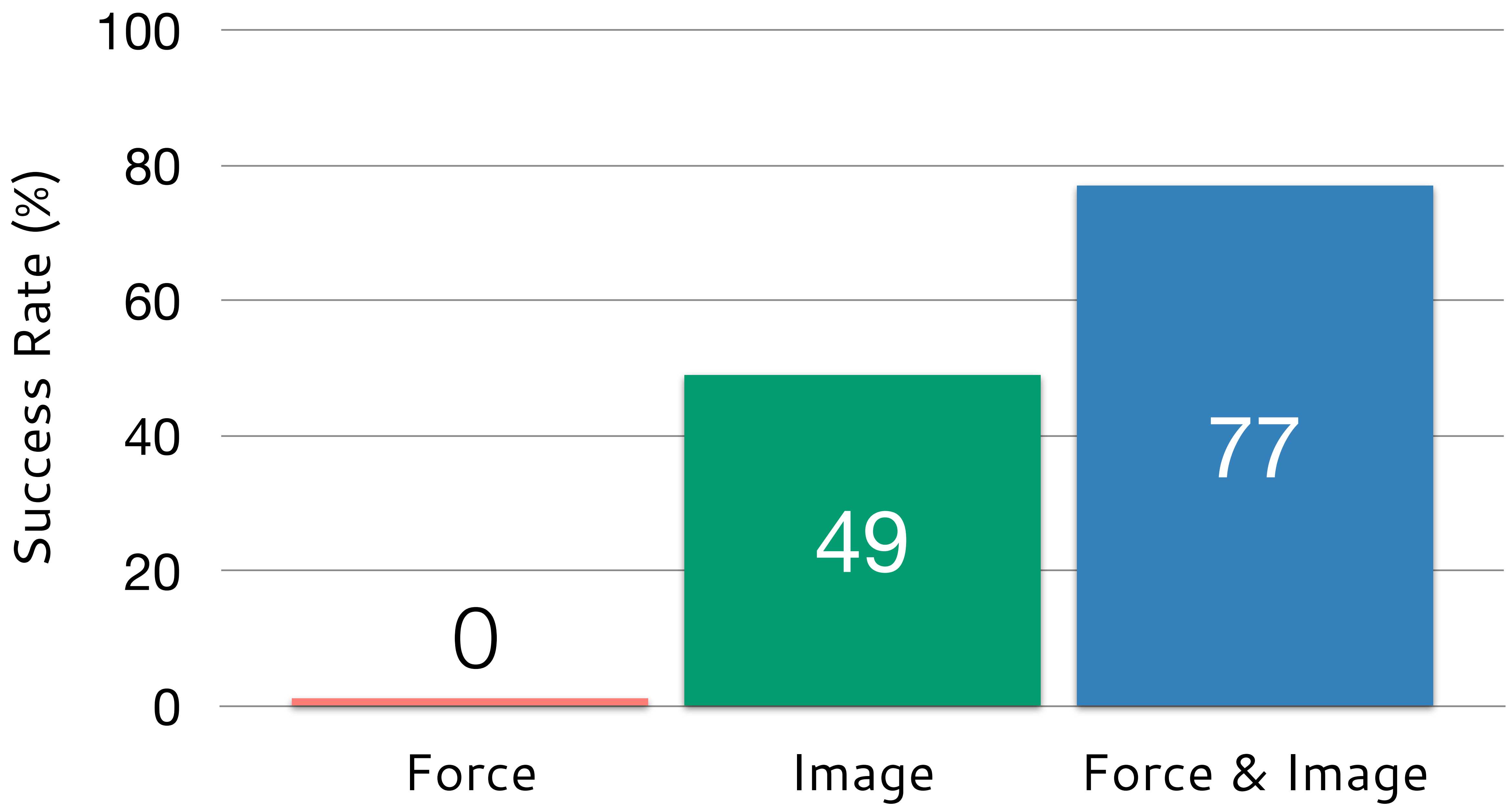


Episode 300

92% success rate



How is each modality used?



Simulation Results
(Randomized box location)

Force Only: Can't find box

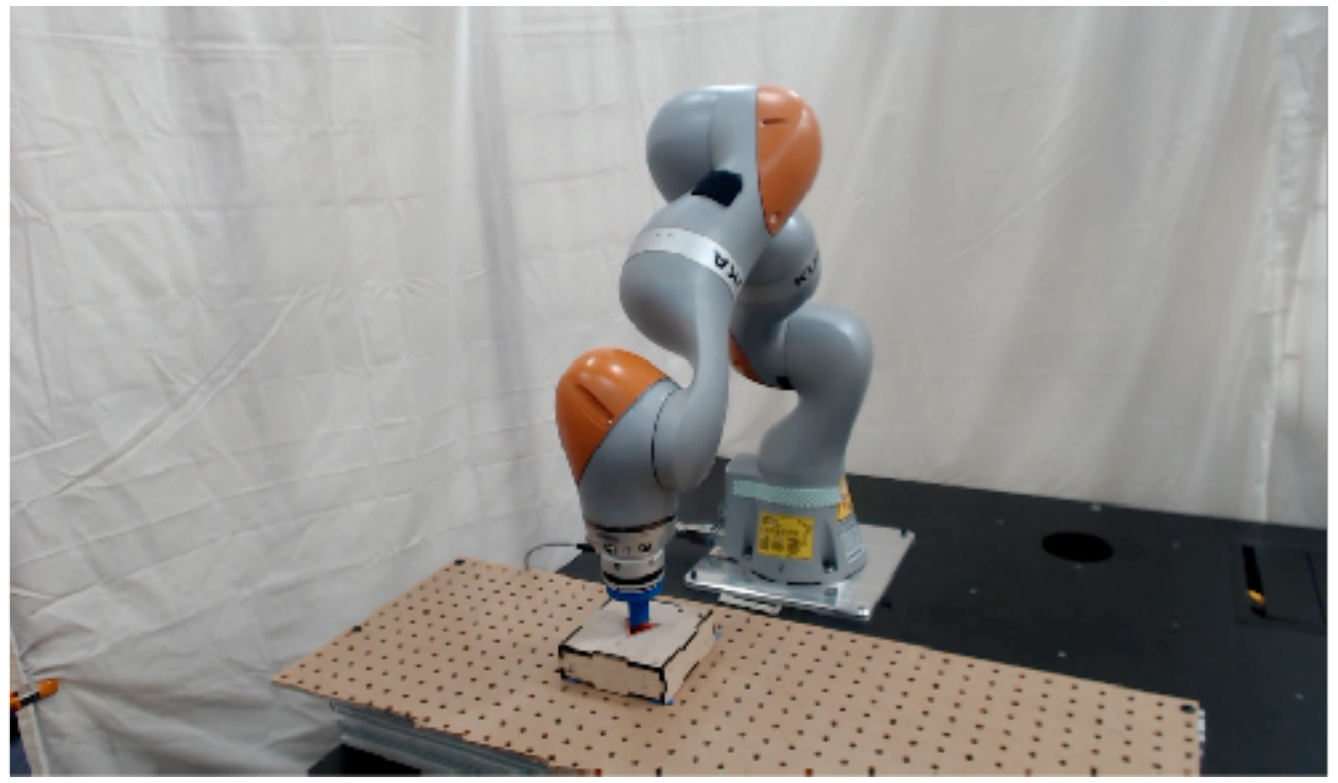
Image Only: Struggles with peg alignment

Force & Image: Can learn full task completion

Does our representation generalize to new geometries?

Does our representation generalize to new geometries?

92% Success Rate



Tested on

Representation

Policy

Does our representation generalize to new policies?

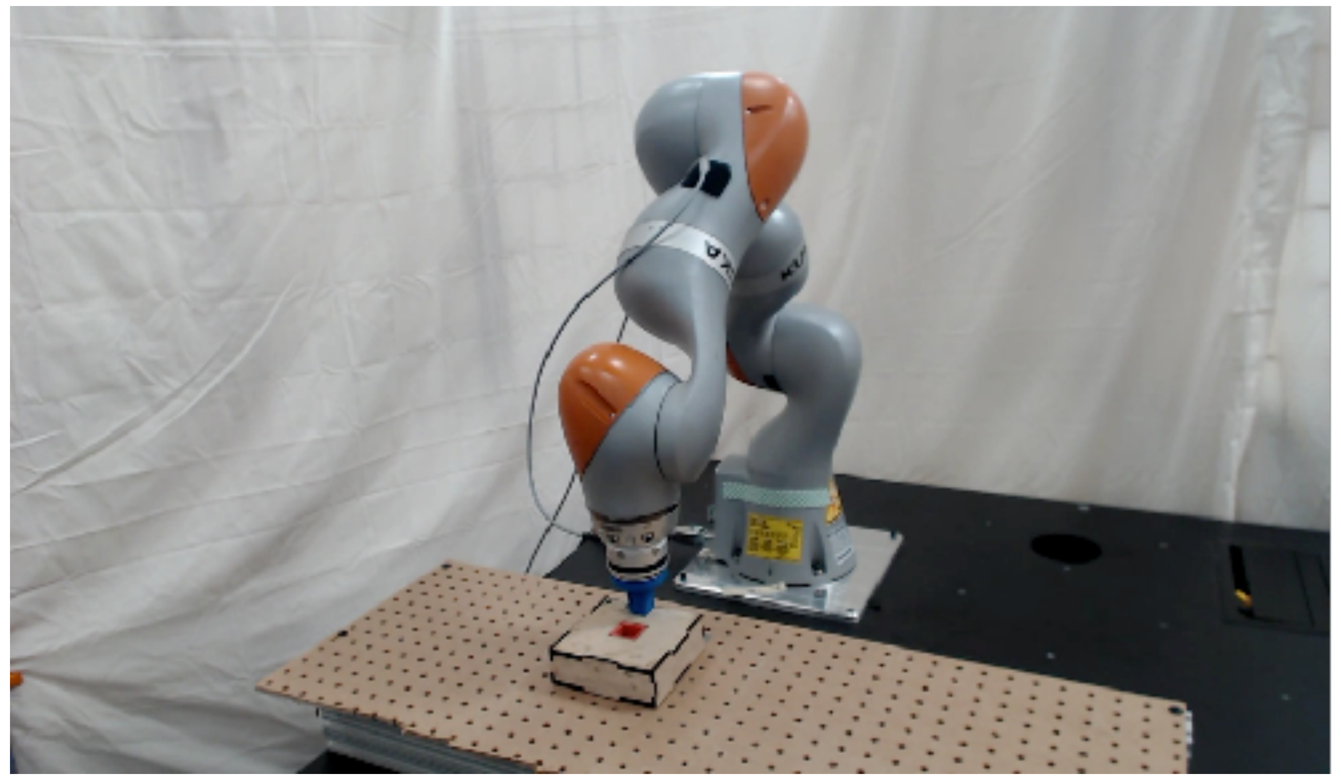
92% Success Rate

Tested on

Representation

Policy

62% Success Rate



Policy does not transfer

Tested on

Representation

Policy

Does our representation generalize?

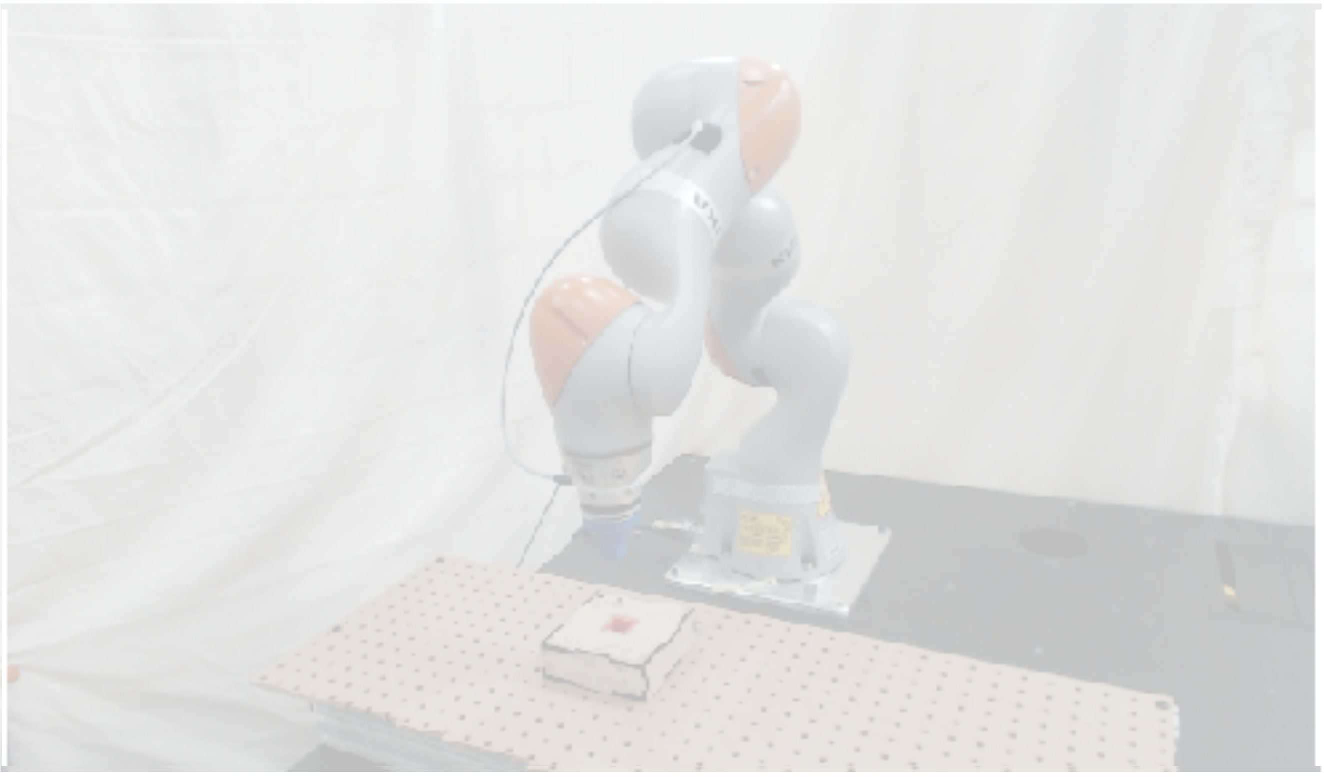
92% Success Rate

Tested on △

Representation △

Policy △

62% Success Rate



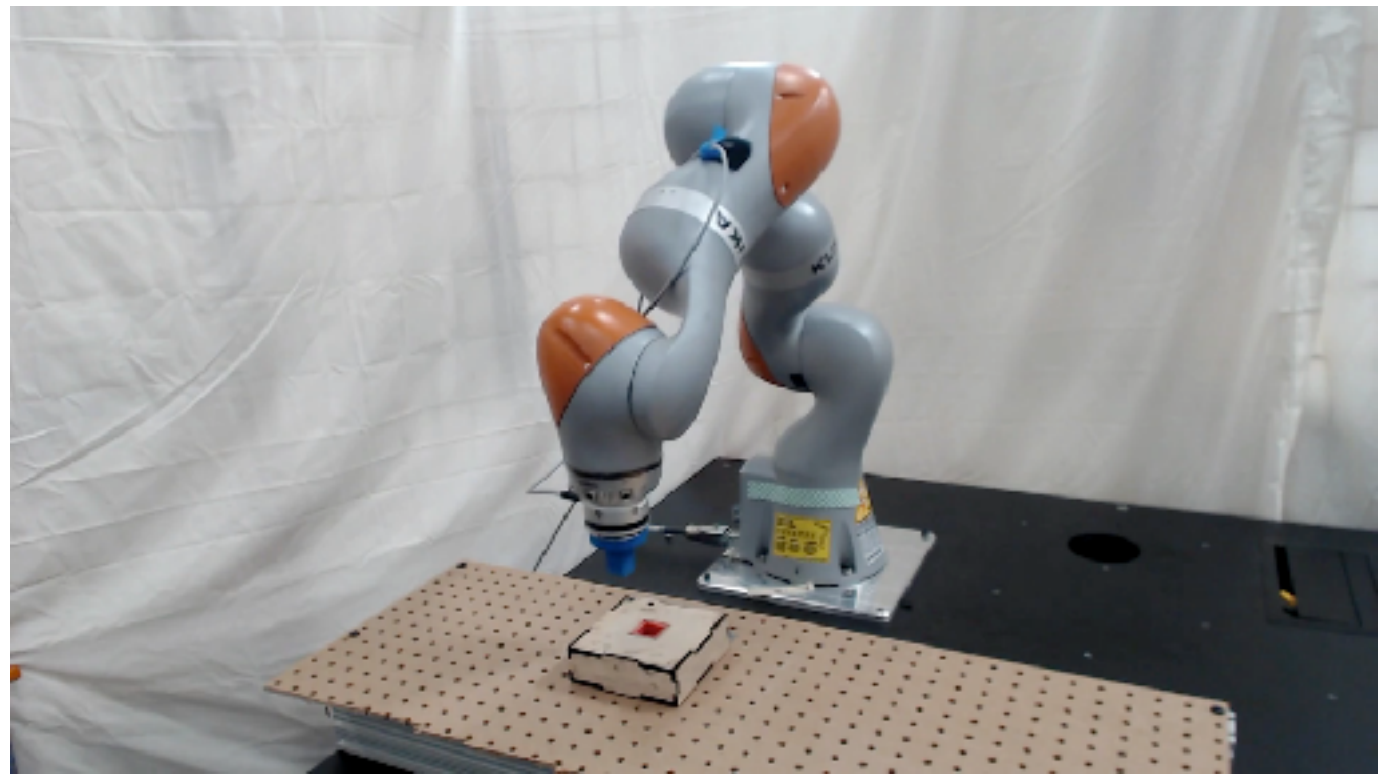
Policy does not transfer

Tested on ■

Representation △

Policy △

92% Success Rate



Representation transfers

Tested on ■

Representation △

Policy ■

Our multimodal policy is robust against sensor noise

1

Force Perturbation

2

Camera Occlusion

External Force

The policy is able to recover from external pushes on the arm.

Target Movement

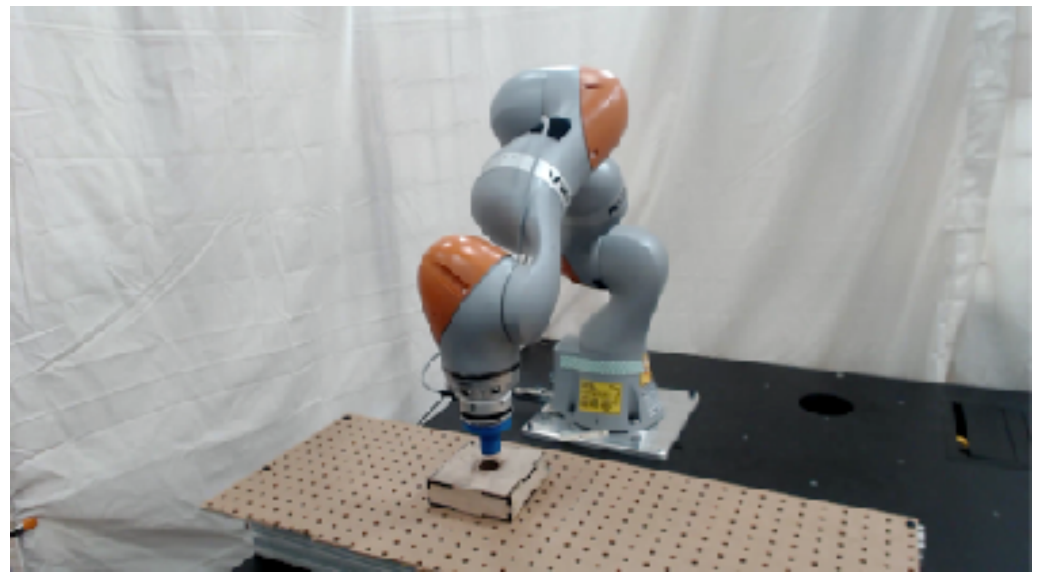
The policy is able to handle small offsets of the position of the box (in new unseen locations)

x1 * no object in environment is tracked

Overview of our method

Self-supervised data collection

$o_{RGB}, o_{force}, o_{robot}$

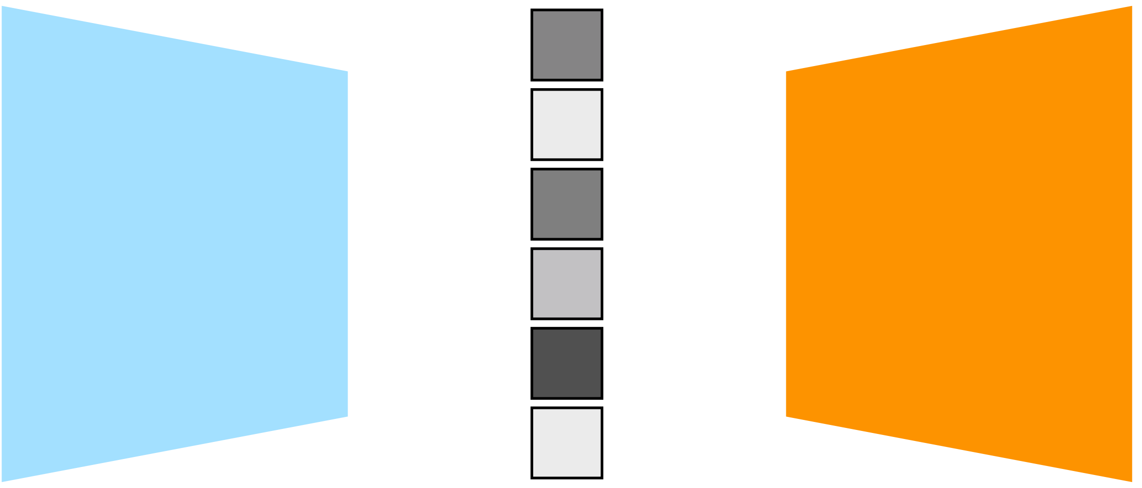


100k data points

90 minutes

Representation learning

$f(o_{RGB}, o_{force}, o_{robot})$

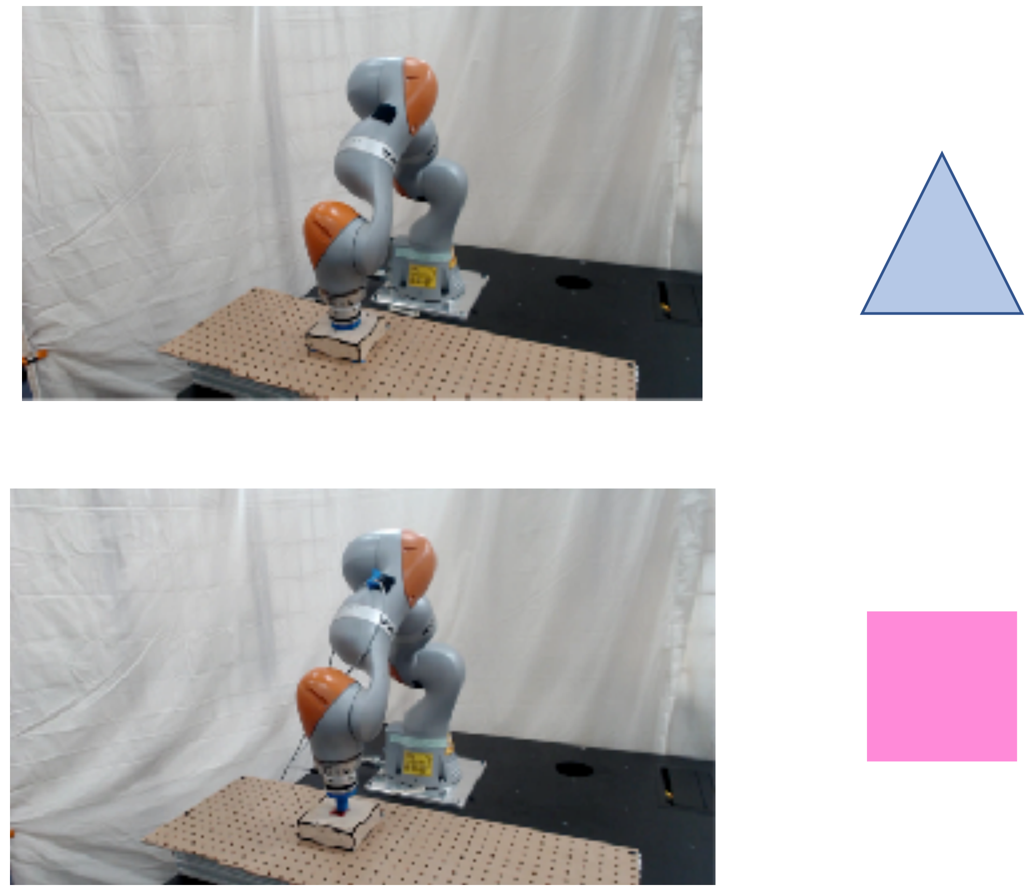


20 epochs on GPU

24 hours

Policy learning

$\pi(f(\cdot)) = a$



Deep RL

5 hours

Lee et al. Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks. ICRA'19. *Best Paper Award.*

Lessons Learned

1. **Self-supervision** gives us **rich** learning objectives

2. Representation that captures **concurrency** and **dynamics** can **generalize** across task instances

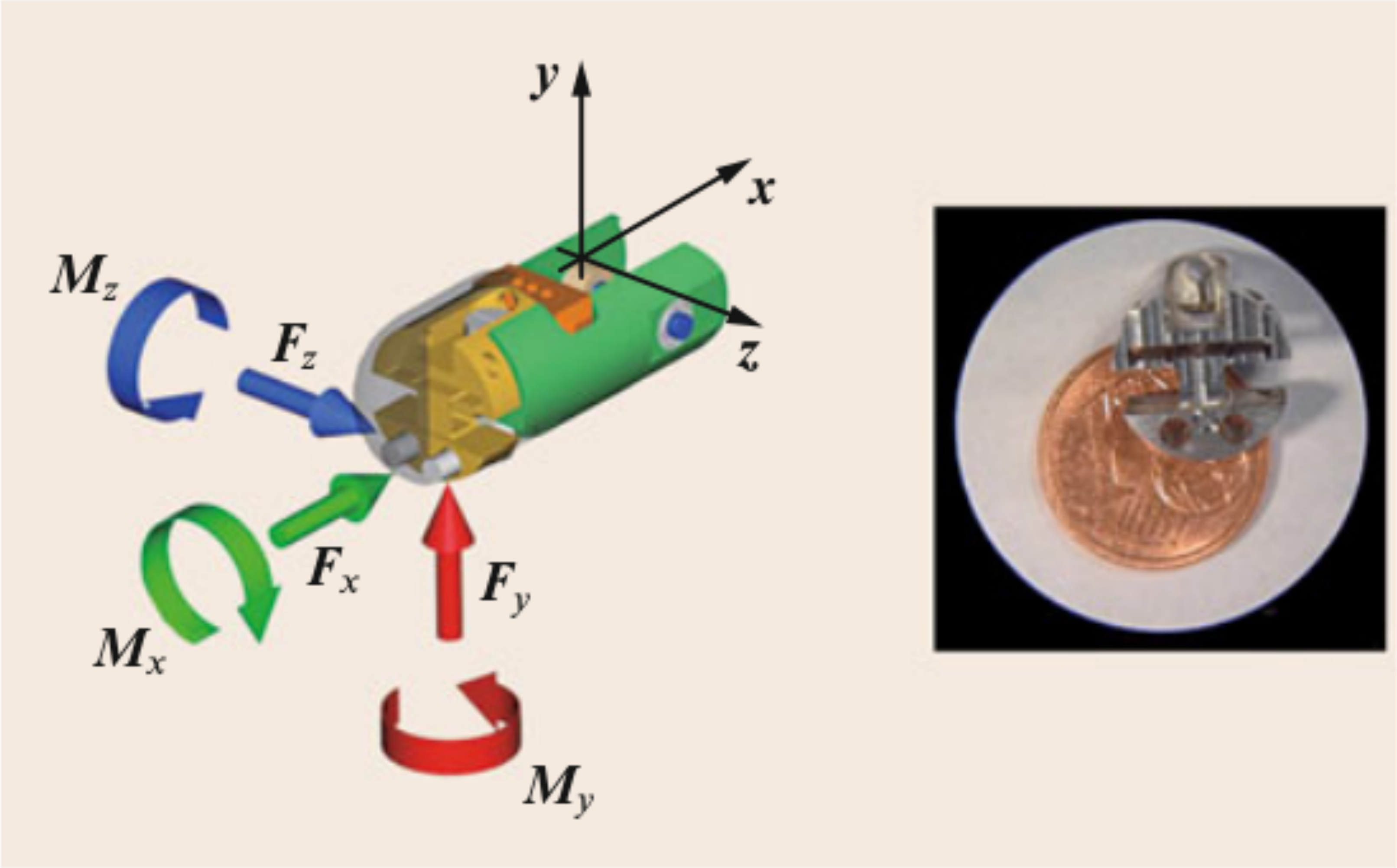
3. Our experiments show that multimodal representation leads to **learning efficiency** and **policy robustness**

Force and Load Sensing

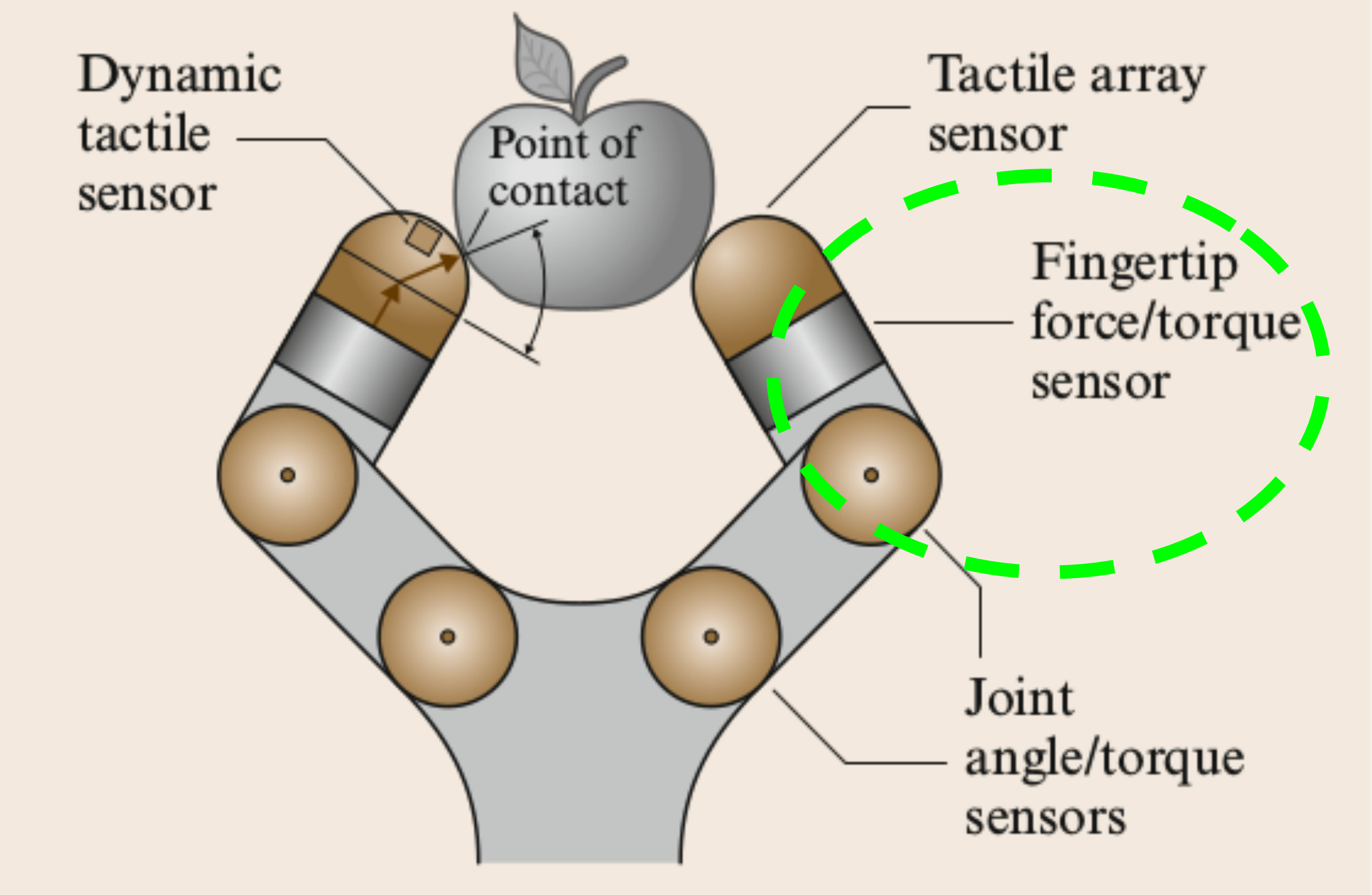
To measure contact forces.

- **Actuator Effort Sensors**
  Servo motors → motor current
  Cables → cable tension

- **Force Sensors**
  - Mounted at the base joint, wrist, or distributed.



Multi-axis fingertip force-torque sensor.



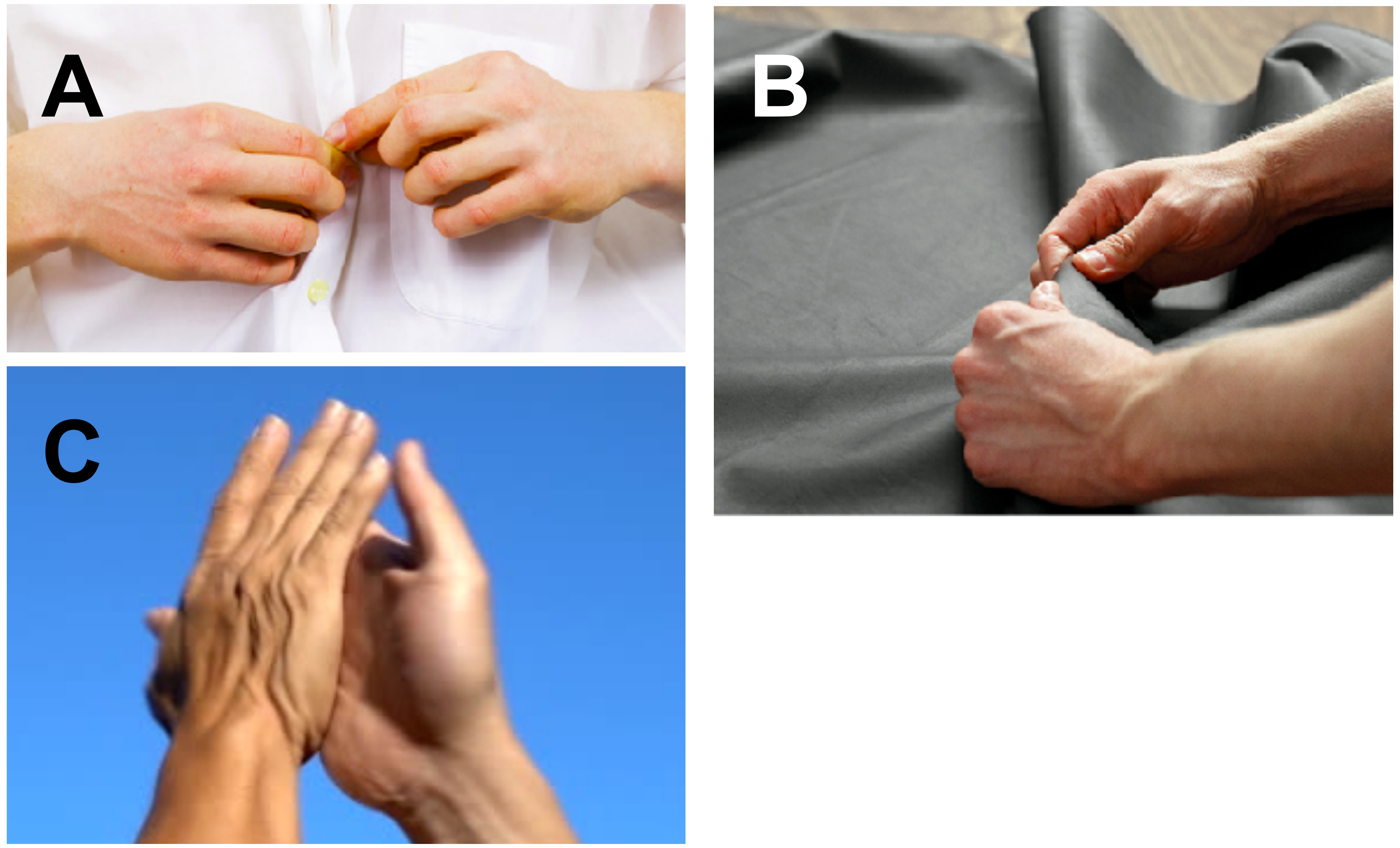
Force sensors + fingertip geometry to estimate contact location.

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

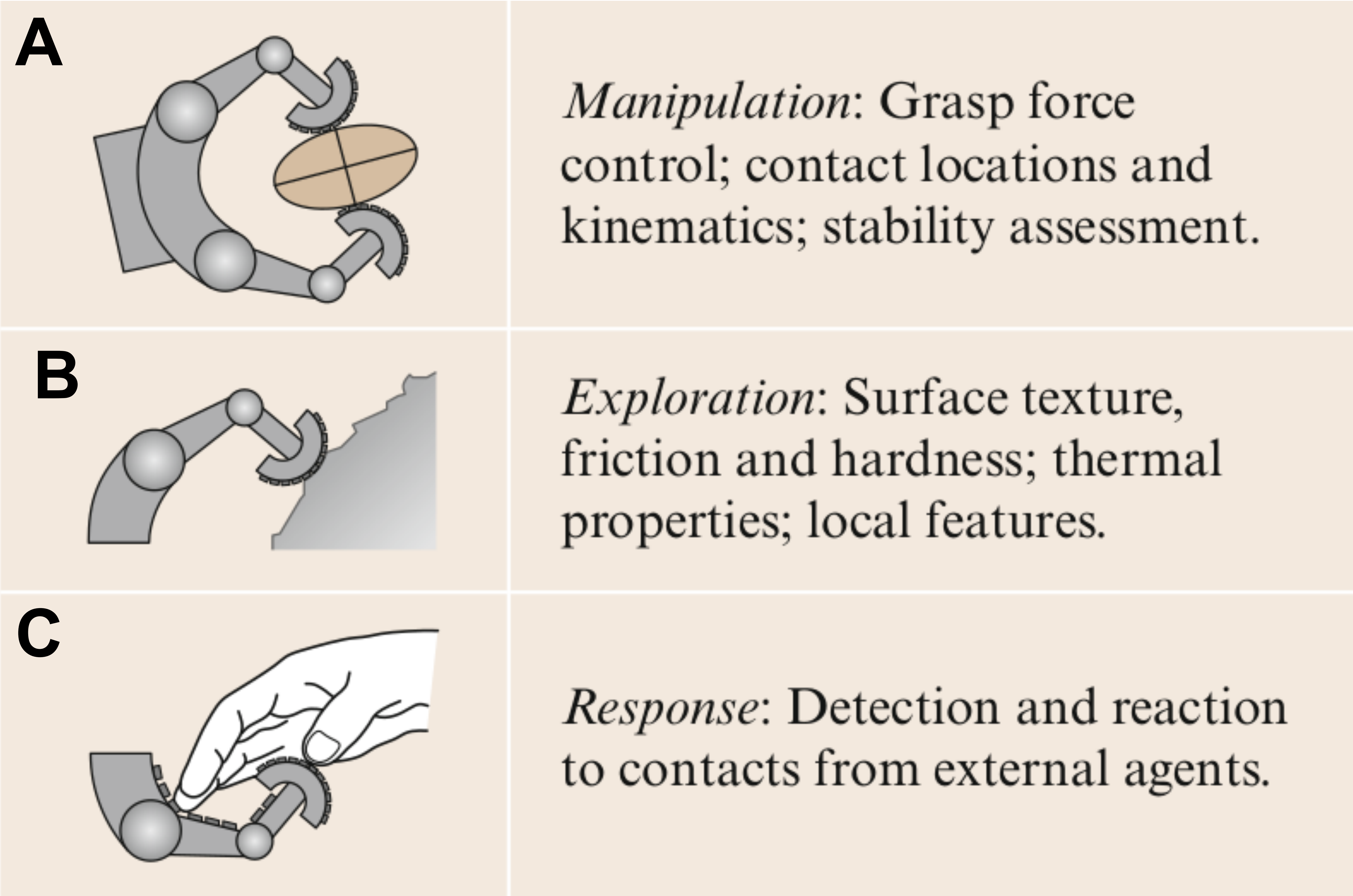
Why Tactile Sensing?

Three main activities:

A. Manipulation
B. Exploration
C. Response



Uses of tactile sensing in humans



Uses of tactile sensing in robotics

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

GelSight

- **Optical sensor** with deformable elastomer
- **Geometry sensing**
- **High spatial resolution**
- **Independent** from optical properties of the object

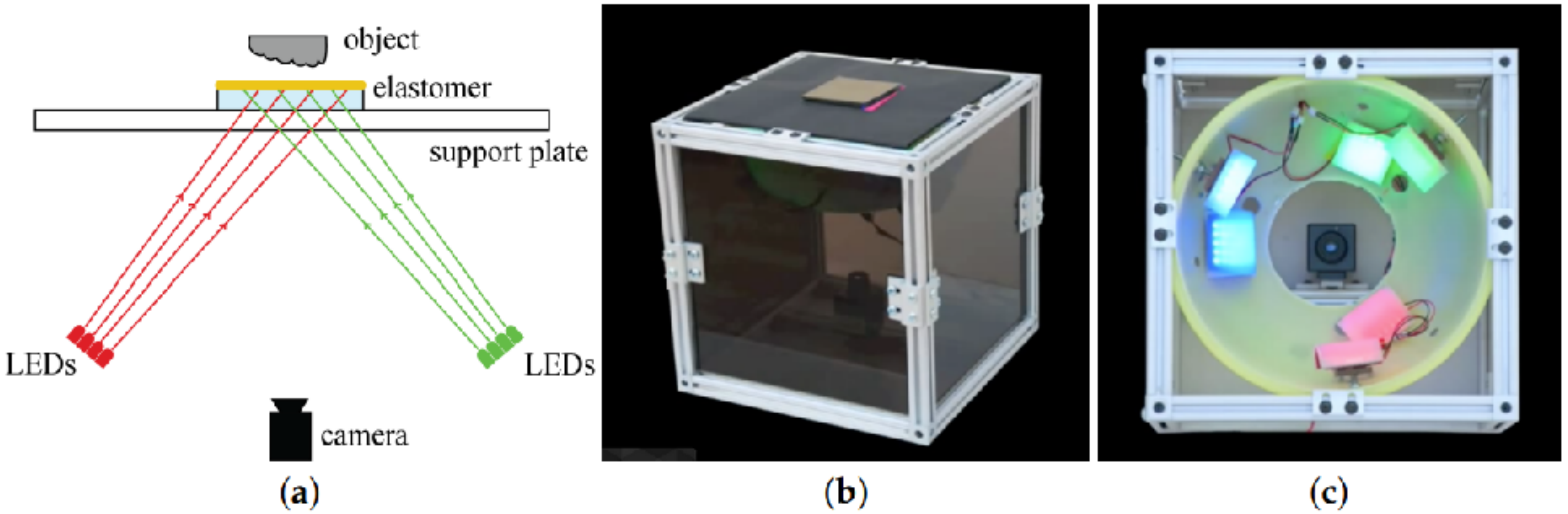
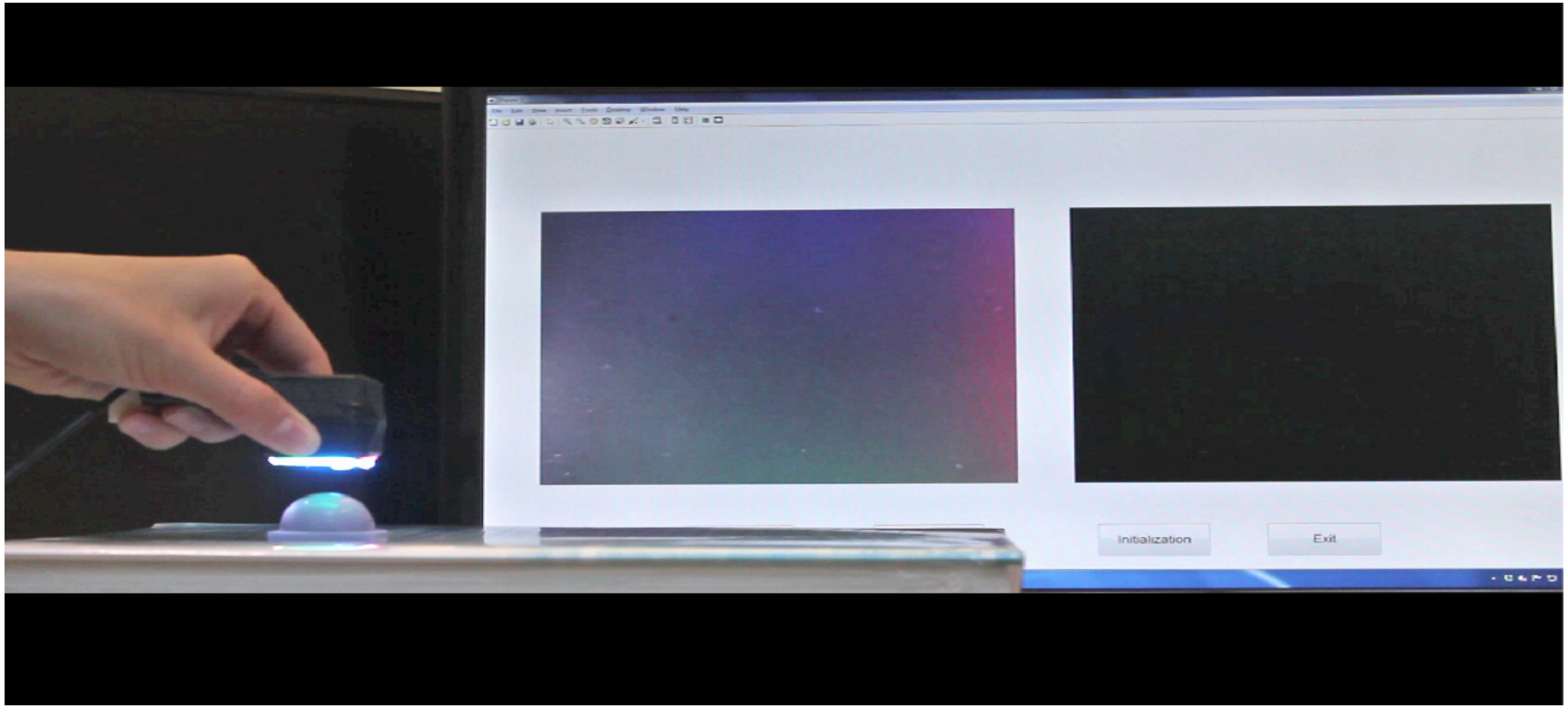


Figure 2. (**a**) basic principle of the Gelsight and the desktop design introduced in [7]. There are four main components for the GelSight sensor: an sensing elastomer piece with the opaque reflective membrane on top, supporting plate, LEDs which provide illumination, and the camera in the bottom to capture the shaded images under the illumination from different directions; (**b**) shows the picture of the sensor, and (**c**) shows the arrangement of the LEDs and camera when viewing from the top.

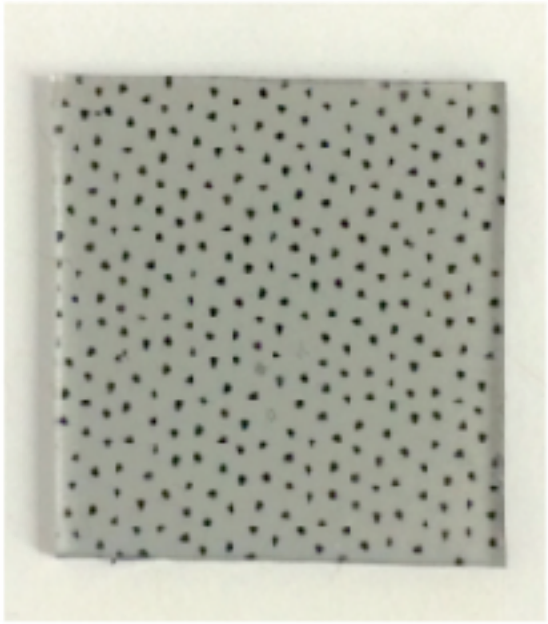
W. Yuan, S. Dong, E. H. Adelson, GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force, 2017.

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

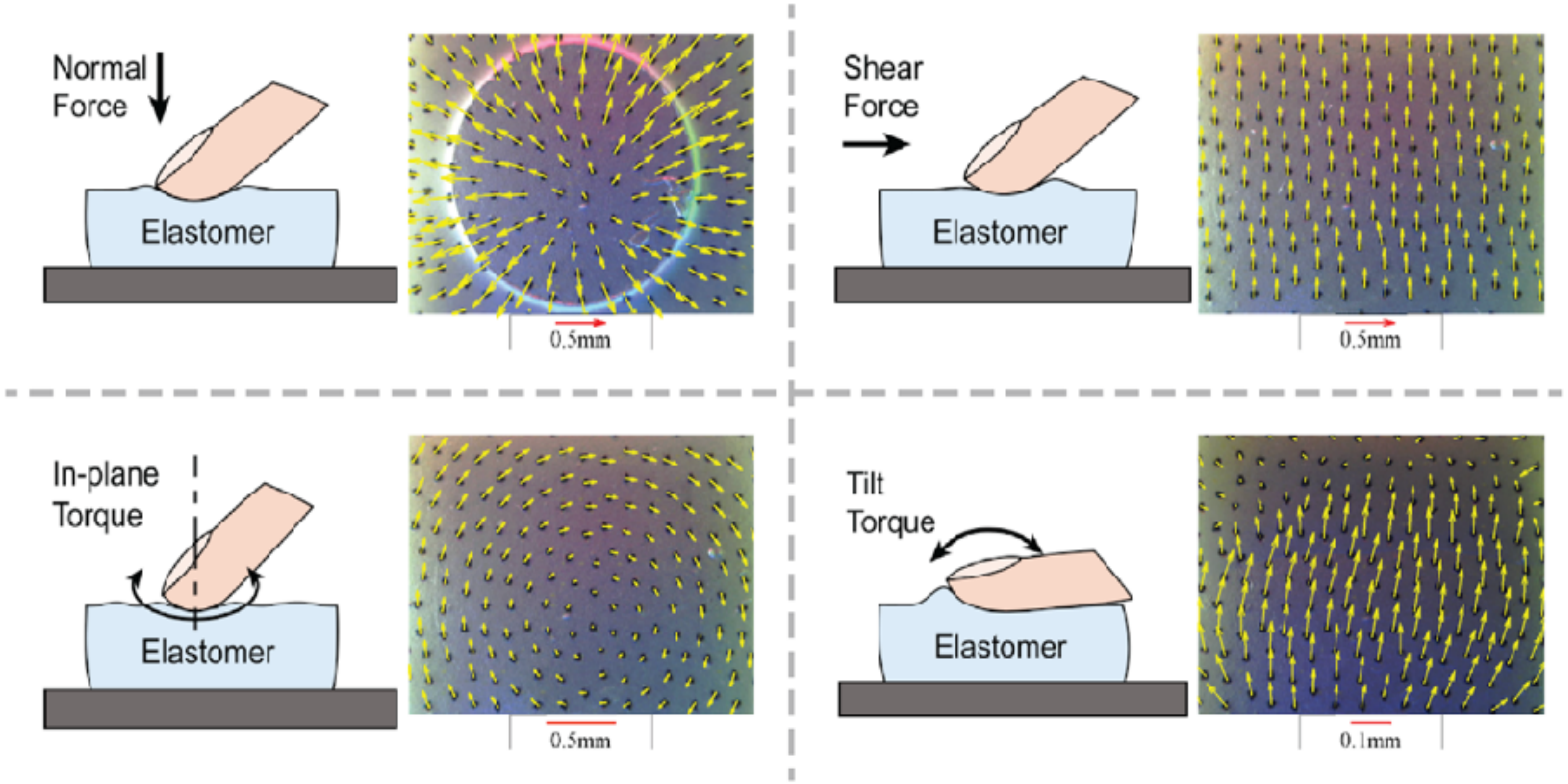

Initialization
Exit

Challenges addressed

1. Measurement of shear force
2. Detecting contact area
3. Hardware optimization
4. Fabrication



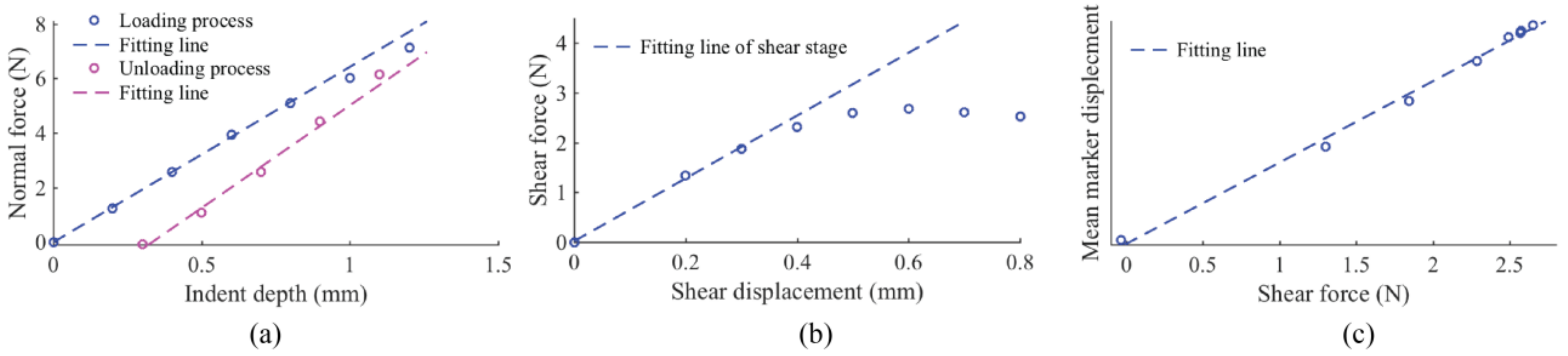
(a)



(b)

Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Marker Motion for Force Measurement

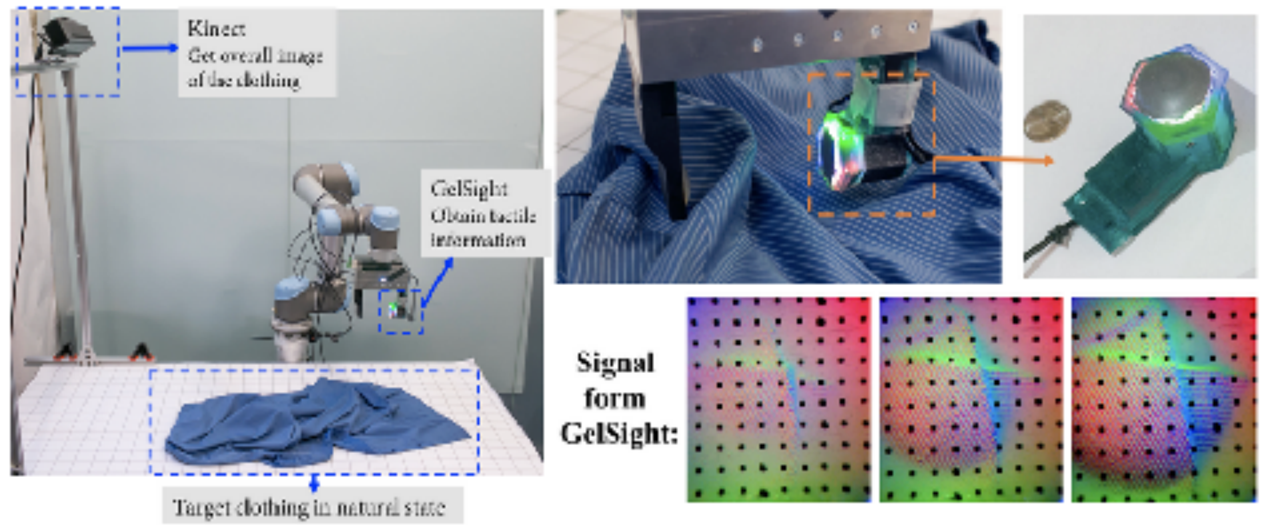


- Magnitude of the motion is roughly proportional to force
- In (b) when the shear load increases until slip occurs

Contact Surface Type	Rigid		Soft (Shore 00-10)	
Contact area	30 mm^2	Flat (>2 cm^2)	30 mm^2	Flat (>2 cm^2)
Using shape measurement	<0.05 N	<0.05 N	<0.05 N	0.08 N
Using marker measurement	<0.05 N	<0.05 N	<0.05 N	<0.05 N

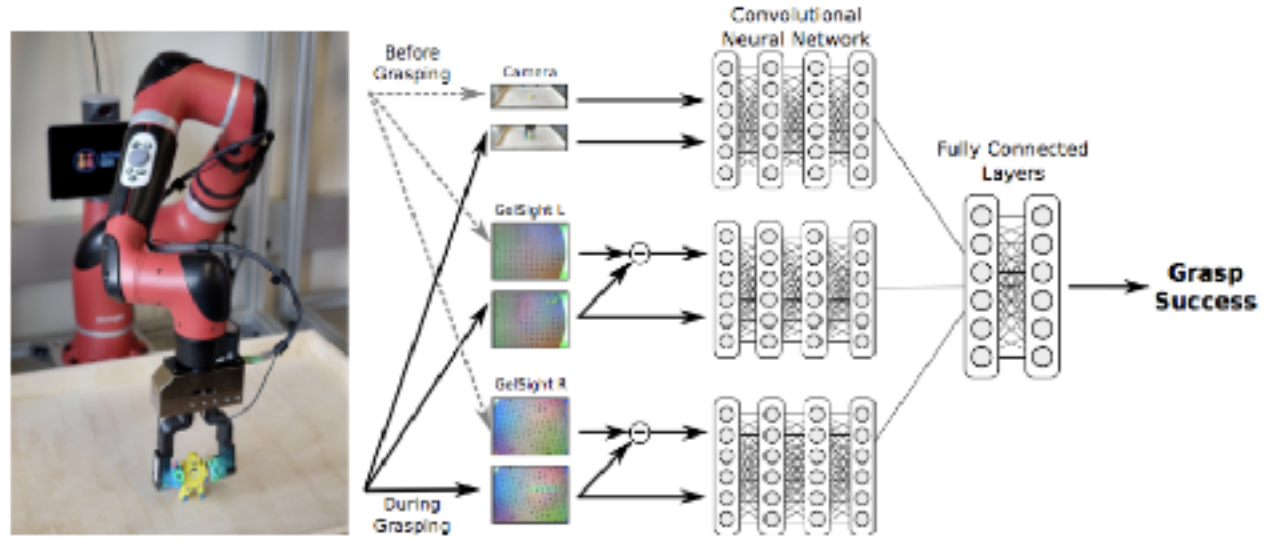
Slide Credit: Allison Okamura and Mark Cutkosky (Stanford ME)

Touch Sensing for Robotics - What's next?



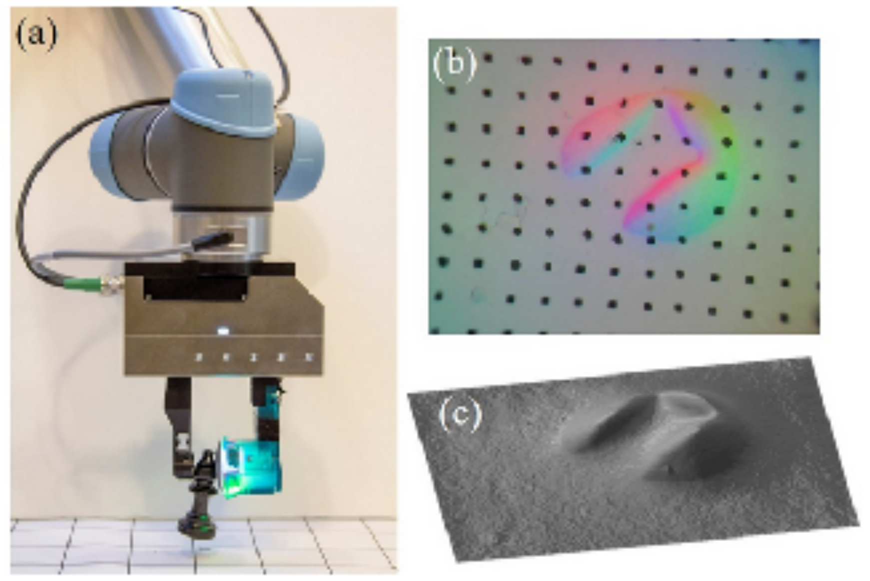
Active clothing perception

The goal of this project is to build a robot system that can autonomously explore the properties of natural clothes. An external Kinect sensor guides the robot to move to the proper positions on the clothing for tactile exploration, and then the robot squeezes the clothing with a GelSight finger. We applied CNN to learn multiple clothing properties from the tactile data. The tactile output was used to improve the robotic exploration as well.



Deep grasping with vision and touch

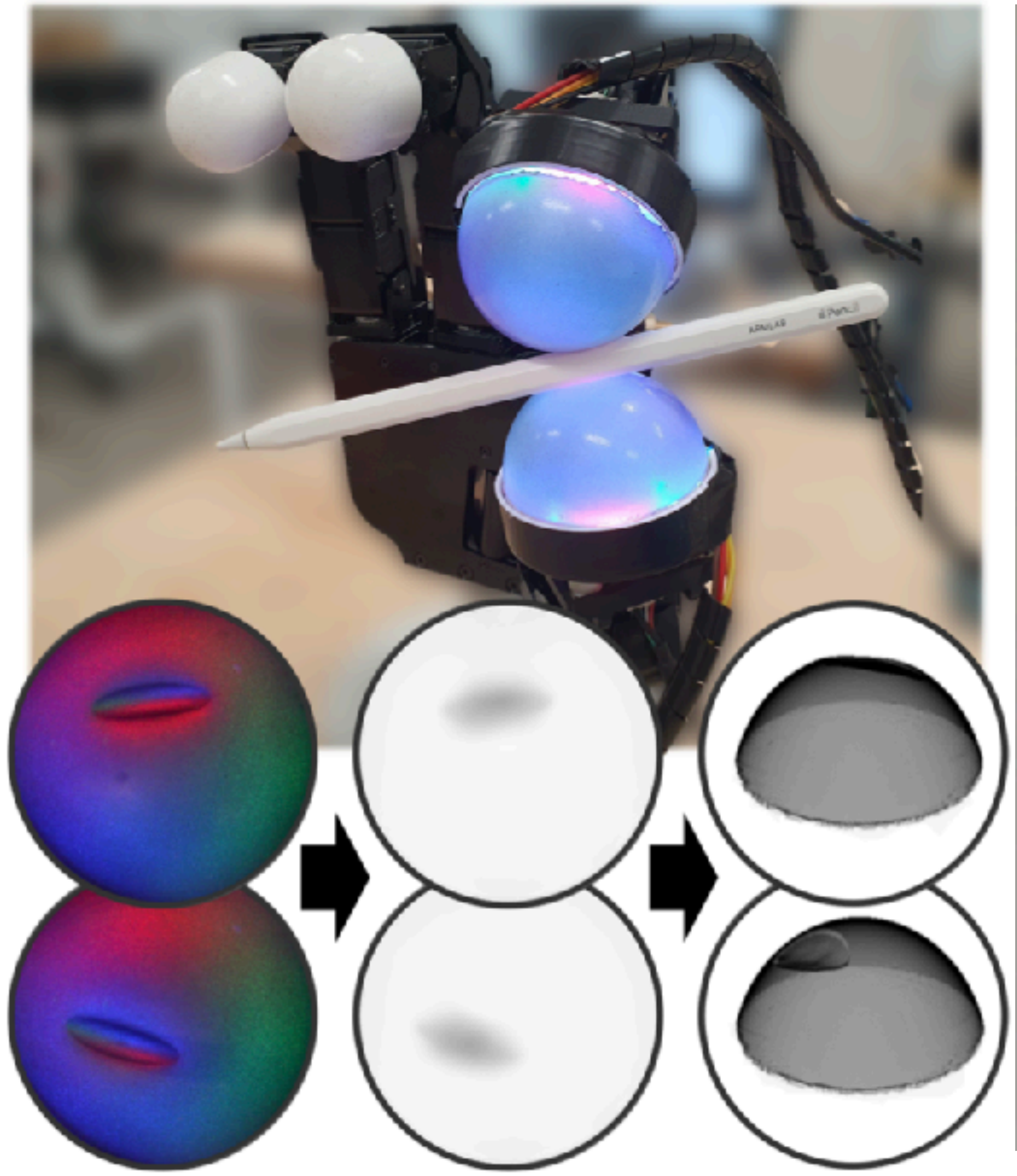
We try to predict the grasping success through both vision and tactile sensing using a deep neural network architecture. We build a dataset of over 9,000 grasping trials on 106 different objects. The experiment results show that incorporating tactile sensors substantially improve grasping performance.



Improved design of fingertip GelSight sensor

We introduce a new design of the fingertip GelSight sensor for robot grippers. The new design measures the geometry of the contact surface with higher 3D precision, and the fabrication is much easier with 3D printing. We publish the drawing and 3D printing files of the sensor, with the hope of the sensor could be accessible to more people.

Thank you for your Attention!



Monday: Prof. Monroe Kennedy on Dense Tact

Appendix