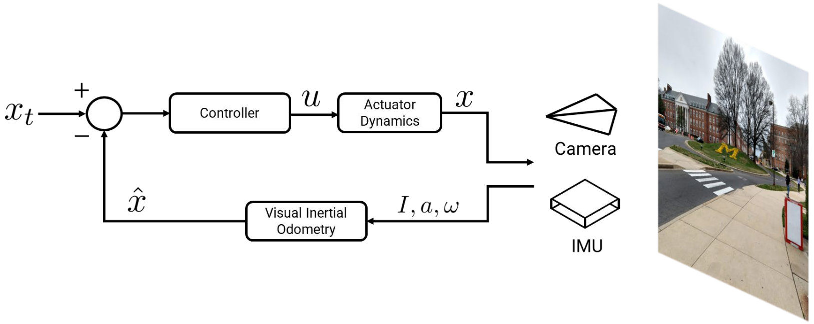
The Advantages of a Control Theoretic Approach to Monocular Computer Vision

Levi Burner, PhD Student

Perception and Robotics Group
Electrical & Computer Engineering Department
University of Maryland, College Park

Agenda

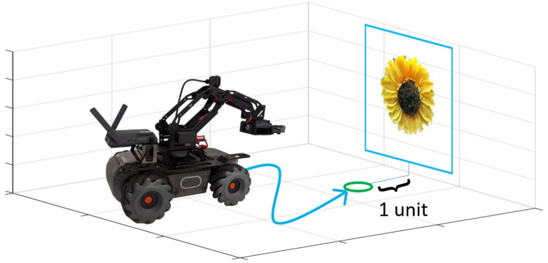
- Traditional Robot Control with Vision



- Control Theoretic Approach
  - Phi and Tau Constraints
  - Stability Invariance

$$\mathbf{F} := \frac{\dot{\mathbf{X}}}{Z} \implies \mathbf{X}(t) = \underbrace{\begin{bmatrix} 1 & 0 & \int_0^t F_X(\lambda)\Phi_{F_Z}(\lambda)d\lambda \\ 0 & 1 & \int_0^t F_Y(\lambda)\Phi_{F_Z}(\lambda)d\lambda \\ 0 & 0 & \Phi_{F_Z}(t) \end{bmatrix}}_{\Phi(t):=} \mathbf{X}_0$$

- Conclusion
  - Summary of how to use in your own projects

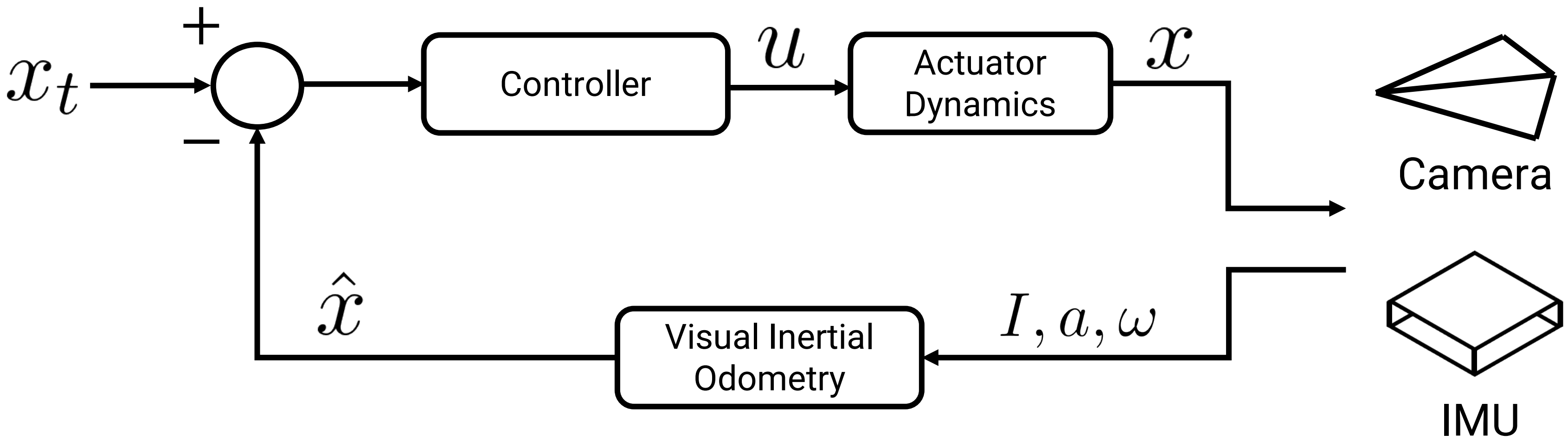


- Full Paper: TTCDist: Fast Distance Estimation From an Active Monocular Camera Using Time-to-Contact, Levi Burner, Nitin J. Sanket, Cornelia Fermüller, Yiannis Aloimonos
  https://arxiv.org/abs/2203.07530

Robots Tracking Trajectories with Cameras

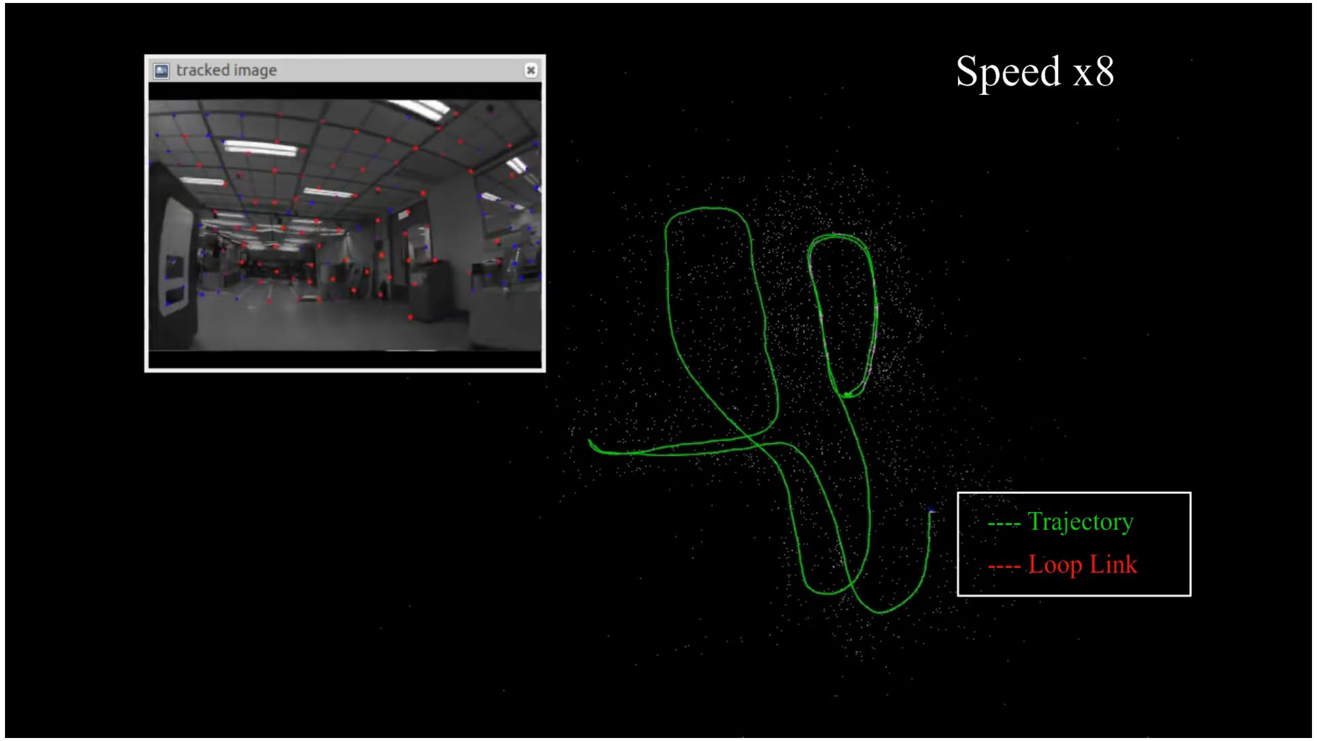
"Introducing Skydio 2" - https://www.youtube.com/watch?v=imt2qZ7uw1s
"With you, Spot can" - https://www.youtube.com/watch?v=VRm7oRCTkjE

Typical Robot Visual Tracking Control (in academia)



Visual Inertial Odometry

- Tracks points/patches scattered across FOV
- Combines IMU and tracked points through optimization
- Some methods use patches instead of points
  - Usually try to minimize a photometric loss with either an optimizer or iterative Extended Kalman Filter

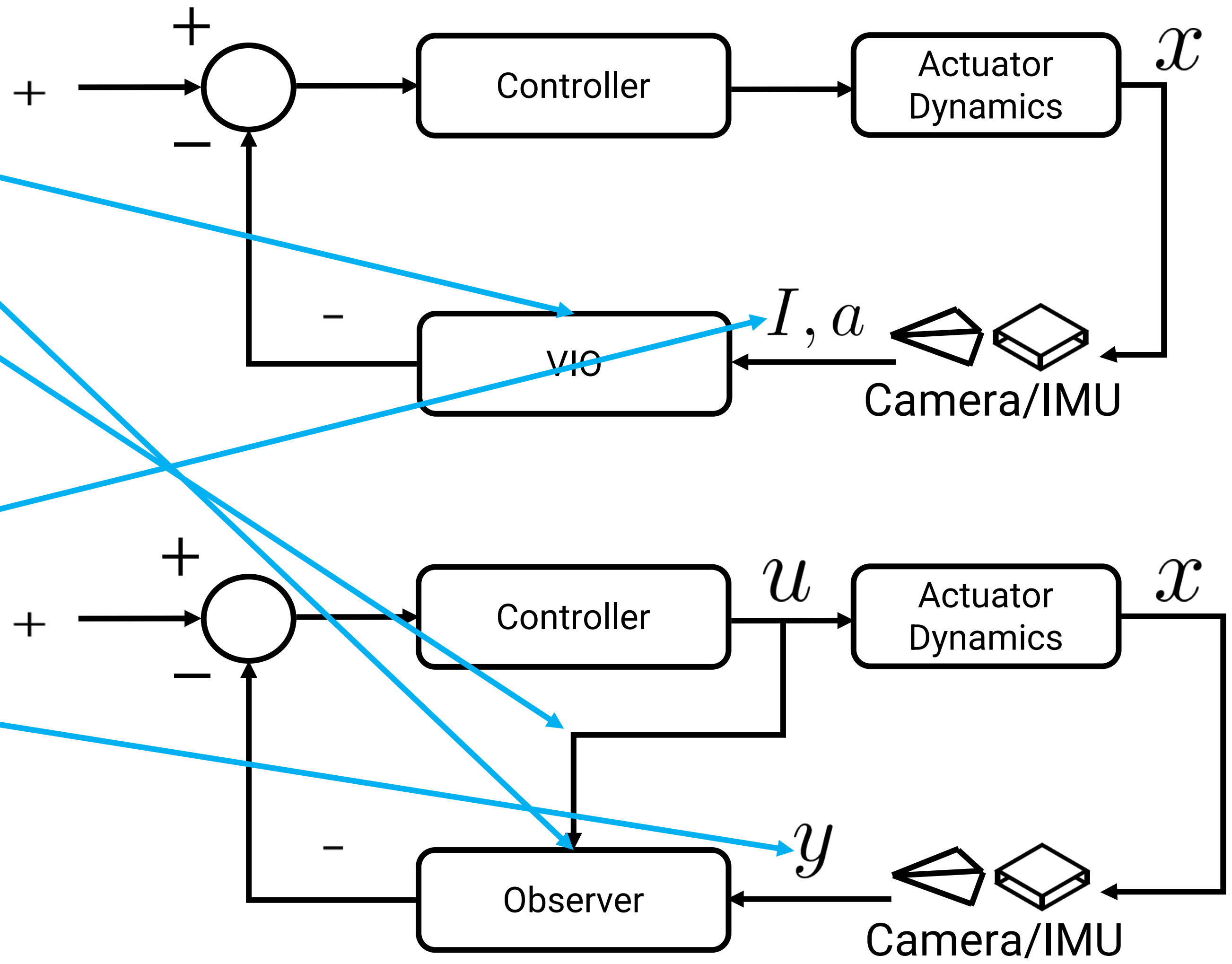


VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator, Tong Qin, Peiliang Li, Zhenfei Yang, Shaojie Shen, *IEEE Transactions on Robotics*

Control Theoretic Approach

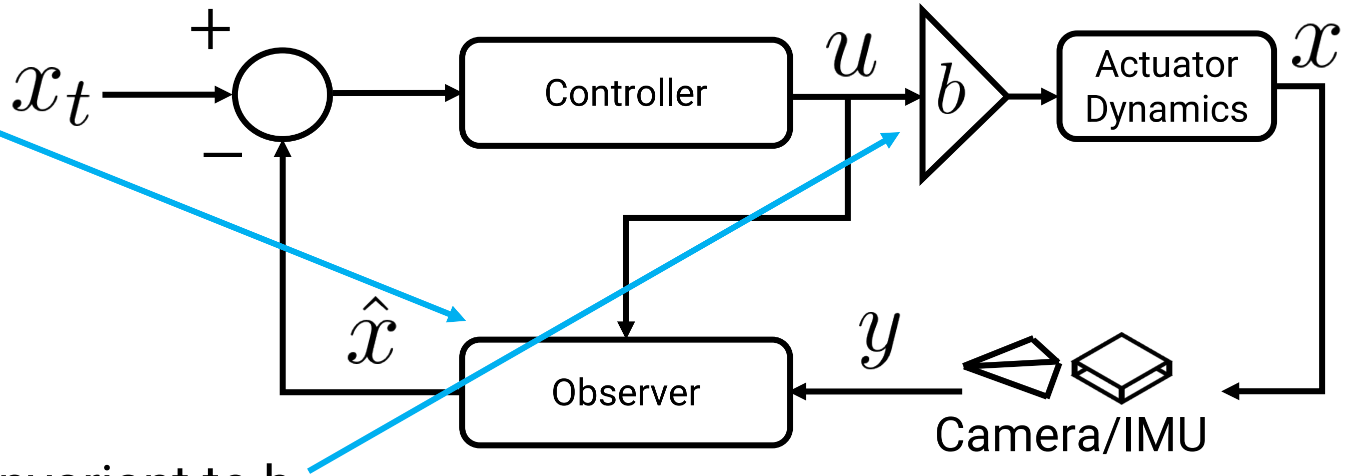
- Replace VIO with observer
- Send control signal to observer
- Treat camera as source of output feedback

$$y = h(x)$$



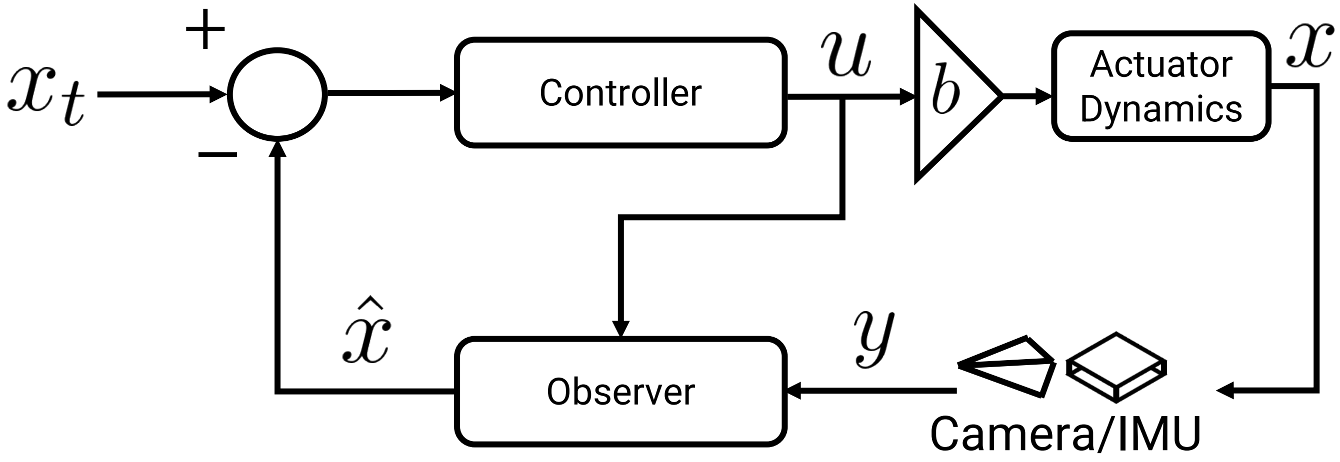
Advantages of Control Theoretic Approach

- Careful choice of y allows observer to be linear
  - Due to two closely related linear equality constraints
    - Tau constraint
    - Phi constraint
  - Can easily prove stability
- Sometimes u can replace IMU
  - IMU acceleration is noisy
  - u is not noisy
  - The closed loop dynamics become invariant to b
  - Robot becomes “very stable”



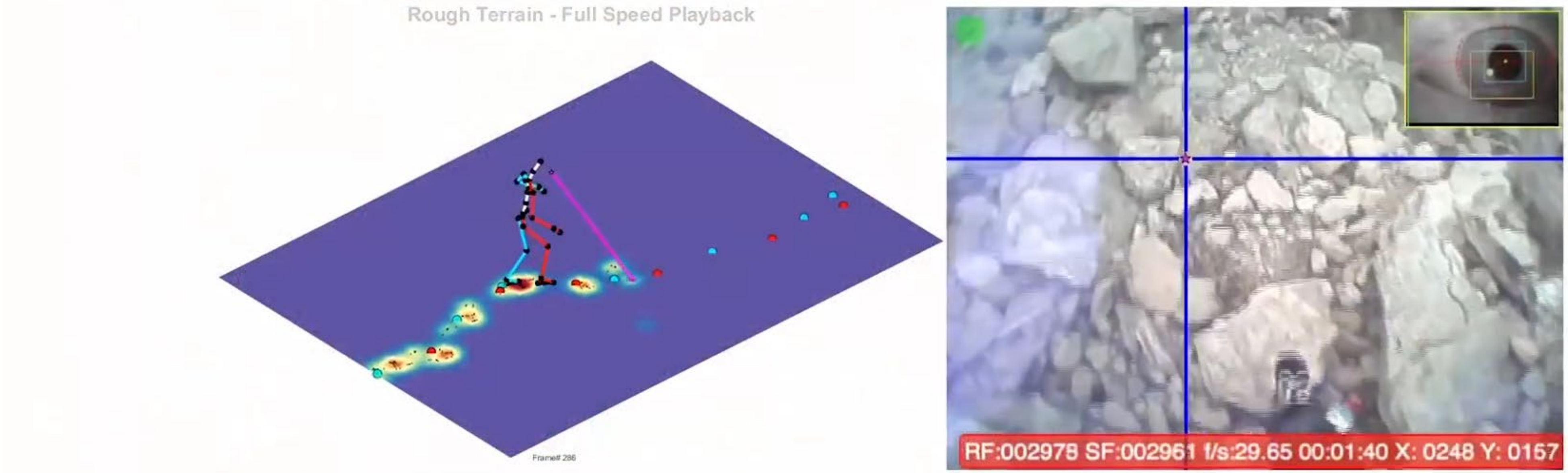
What should y be?

- y = h(x)
- We have the luxury of choosing h
- Bio-inspiration?
  - What might humans measure?



Mammalian Vision

- When human's walk they switch their gaze



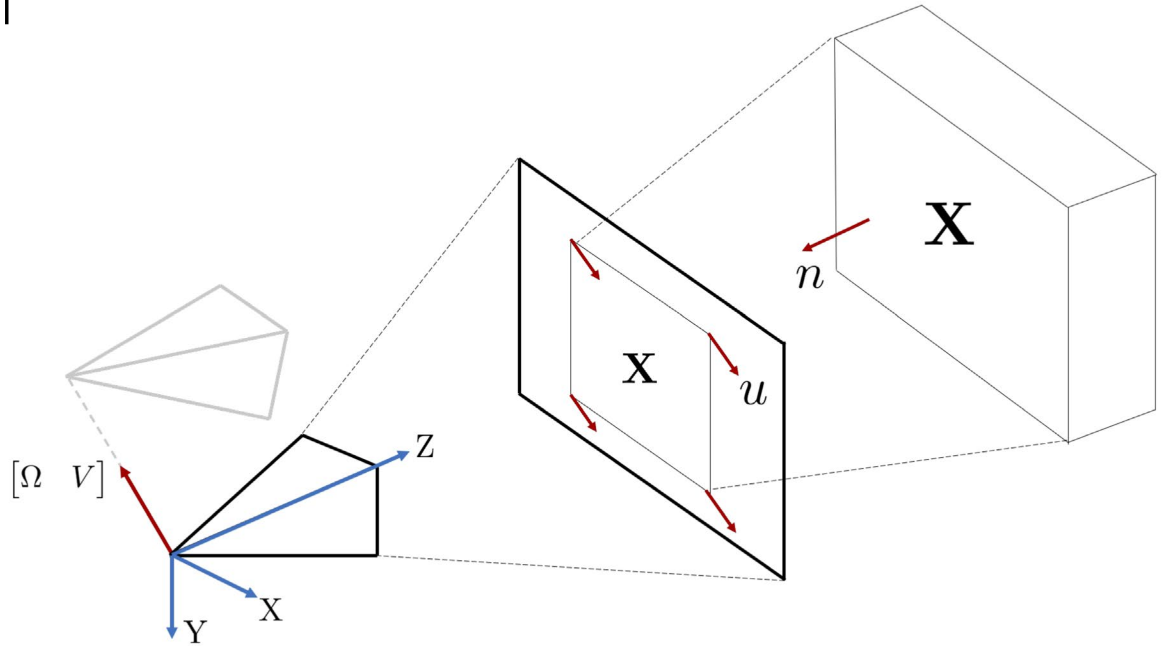
Matthis, J.S., Yates, J.L., and Hayhoe, M.M. (2018), "Gaze and the control of the foot placement when walking in natural terrain". *Current Biology*

Time to Contact

- Intuitively:
  - Rate of change of object size
  - Sense of when an object will l

- Mathematically:

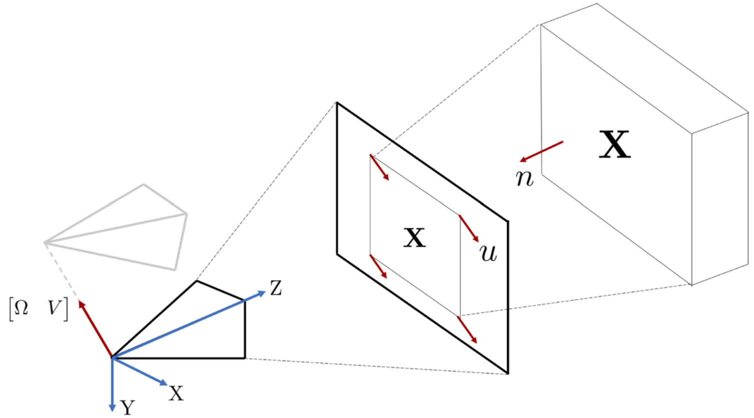
$$\tau(t) = \frac{Z(t)}{\dot{Z}(t)}$$



Measuring Time to Contact

- We can describe where pixels in a patch with an affine "warp"
  - For those taking 733 one way to measure a warp is with the LK algorithm

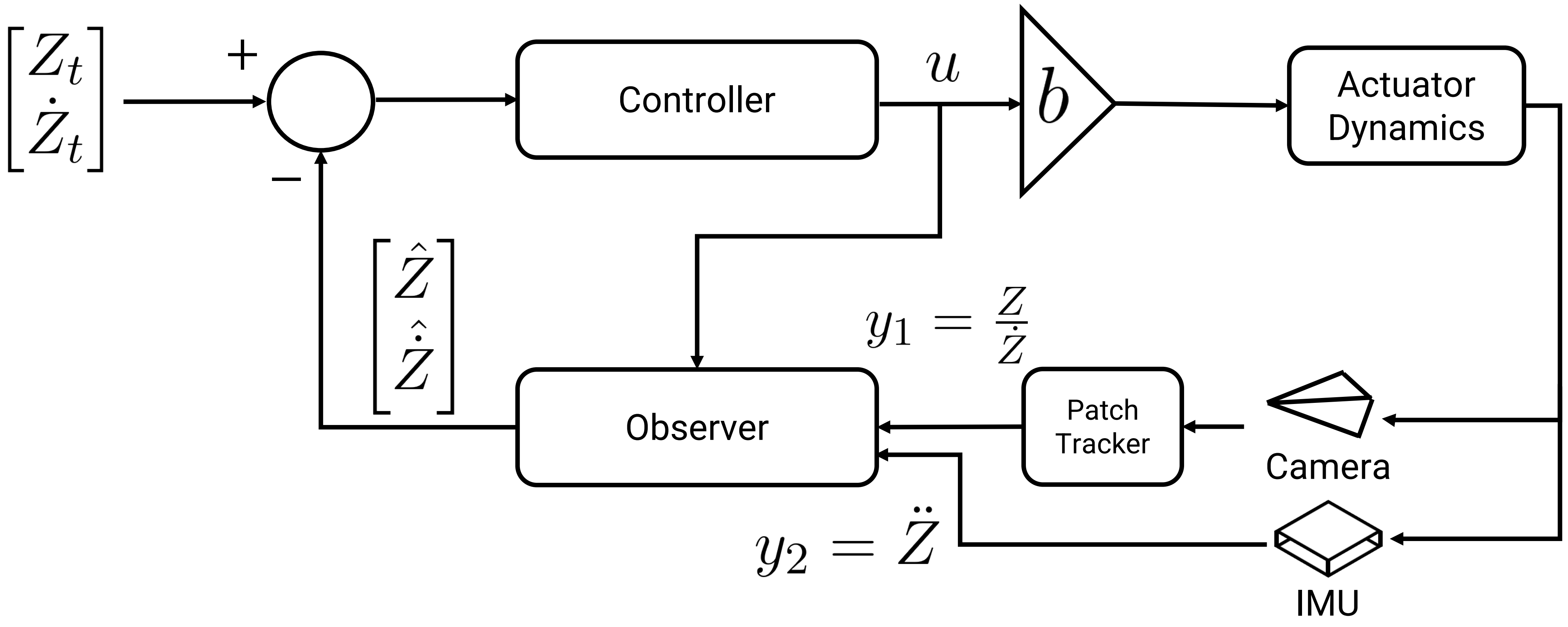
$$\mathbf{x}(t) = \underbrace{\begin{bmatrix} Z_0/Z & 0 & (X - X_0)/Z \\ 0 & Z_0/Z & (Y - Y_0)/Z \\ 0 & 0 & 1 \end{bmatrix}}_{A:=} \mathbf{x}(0)$$



- Differentiating gives optical flow
  - But the affine terms are time-to-contact!

$$\mathbf{u_x} = \frac{d\mathbf{x}(t)}{dt} = \dot{A}A^{-1}\mathbf{x} = -\begin{bmatrix} \dot{Z}/Z & 0 & \dot{X}/Z \\ 0 & \dot{Z}/Z & \dot{Y}/Z \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x}$$

Simplified Case – Z Only



Full Case – Tau/Phi Constraint

- Generalize to frequency of contact

$$\mathbf{F} := \frac{\dot{\mathbf{X}}}{Z} \implies \dot{\mathbf{X}} = \mathbf{F}Z$$

- Recognize F defines an ODE

$$\mathbf{X}(t) = \underbrace{\begin{bmatrix} 1 & 0 & \int_0^t F_X(\lambda)\Phi_{F_Z}(\lambda)d\lambda \\ 0 & 1 & \int_0^t F_Y(\lambda)\Phi_{F_Z}(\lambda)d\lambda \\ 0 & 0 & \Phi_{F_Z}(t) \end{bmatrix}} \mathbf{X}_0$$

- ODE has closed form solution
  - Linear Time Varying System
  - Covered in ENEE660

$$\Phi_{F_Z}(t) = \exp\left(\int_0^t F_Z(\lambda)d\lambda\right)$$

- Write X as function of acceleration

$$\mathbf{X}(t) - \mathbf{X}_0 = t\dot{\mathbf{X}}_0 + \underbrace{\int_0^t \left(\int_0^\lambda \ddot{\mathbf{X}}(\lambda_2)d\lambda_2\right) d\lambda}_{\mathcal{I}\{\ddot{\mathbf{X}}\}(t):=}$$

- Set both sides equal to each other!

Full Case – Phi/Tau Constraint

- Setting both sides equal results in Phi-constraint

$$(\Phi(t) - I)\begin{bmatrix} 0 \\ 0 \\ Z_0 \end{bmatrix} - t\dot{\mathbf{X}}_0 = \mathcal{J}\{\ddot{\mathbf{X}}\}(t) \qquad (\Phi\text{-constraint})$$

- Substitute $\dot{\mathbf{X}} = \mathbf{F}Z_0$ to get Tau-constraint

$$\underbrace{\left(\Phi(t) - I - t\begin{bmatrix} 0 & 0 & \mathbf{F}(0) \end{bmatrix}\right)}_{E(t):=}\begin{bmatrix} 0 \\ 0 \\ Z_0 \end{bmatrix} = \mathcal{J}\{\ddot{\mathbf{X}}\}(t). \qquad (\tau\text{-constraint})$$

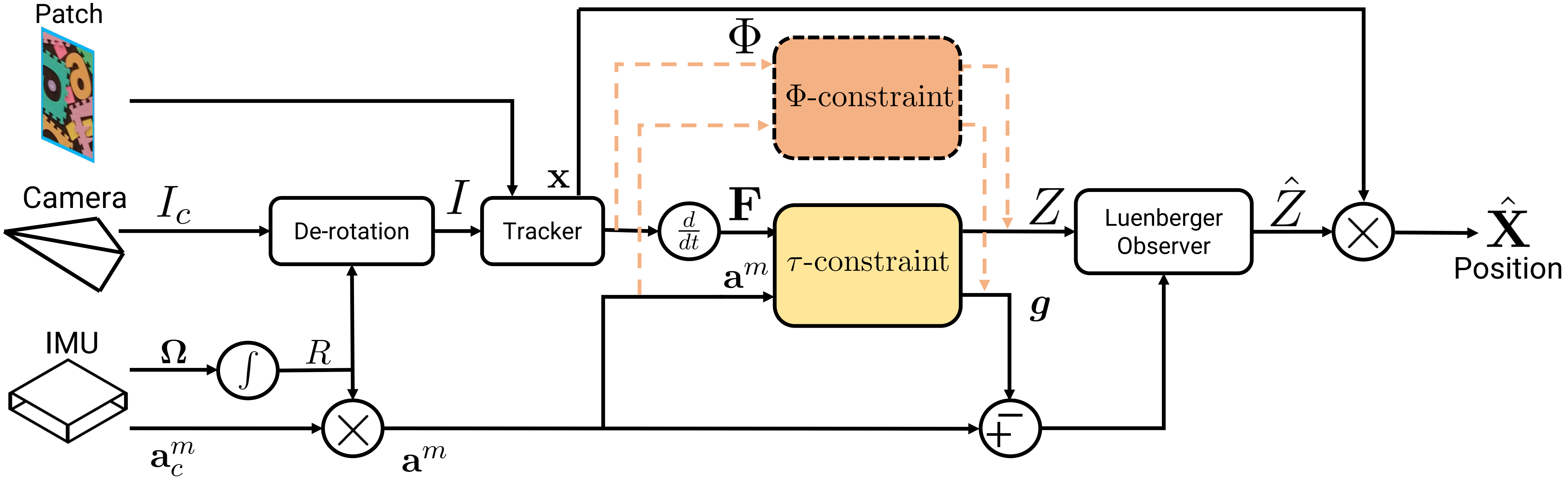
Estimating Distance Becomes a Linear

- Suppose acceleration with an unknown gravitational bias is available
- Can do linear least squares over time

$$\underset{Z_0,\dot{Z}_0,g_Z}{\operatorname{argmin}} \left\| \left(\Phi_{F_Z} - 1\right) Z_0 - r\dot{Z}_0 + \mathcal{J}\{a_Z^m + g_Z\} \right\|_2^2 \qquad (\Phi\text{-constraint})$$

$$\underset{Z_0,g_Z}{\operatorname{argmin}} \left\| E_Z Z_0 + \mathcal{J}\{a_Z^m + g_Z\} \right\|_2^2 \qquad (\tau\text{-constraint})$$

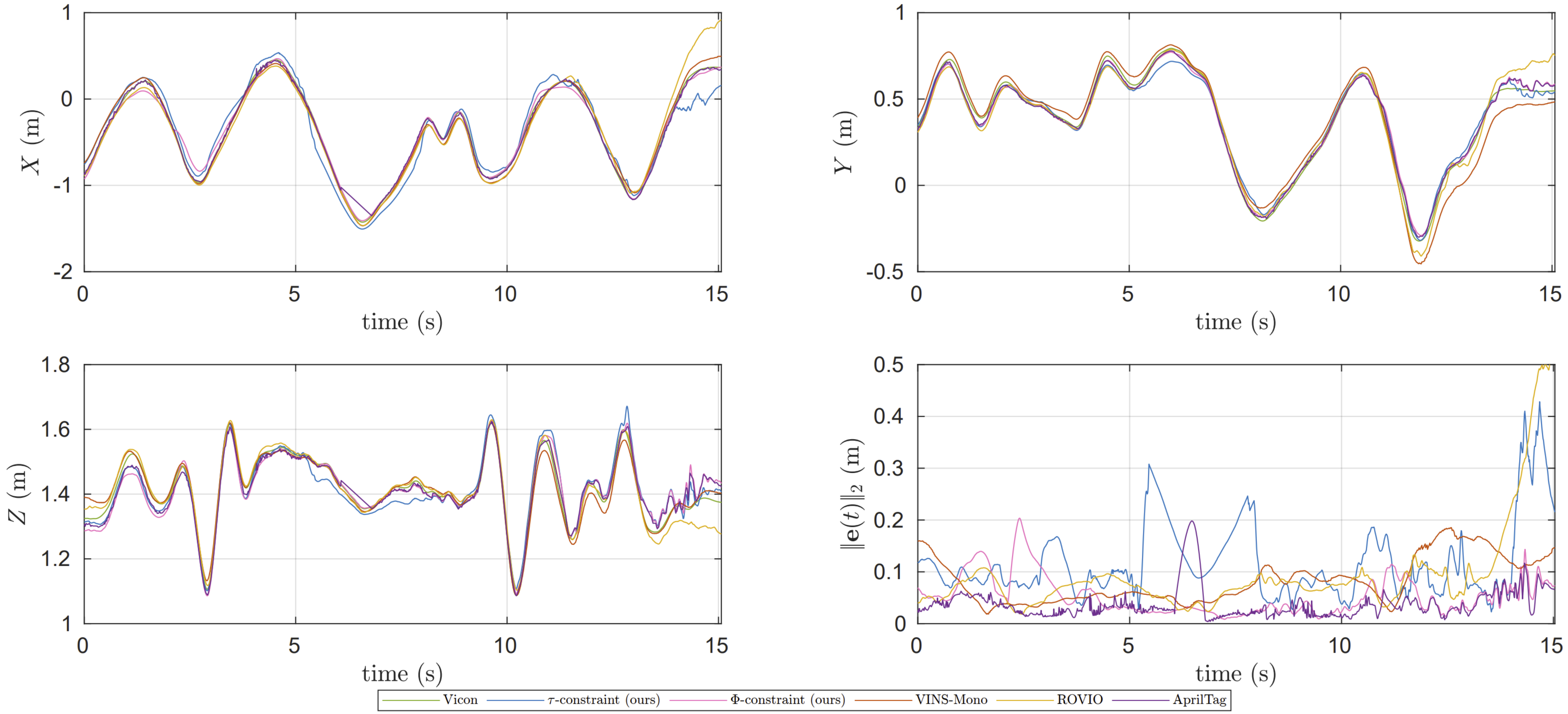
$$\tau := \frac{Z}{\dot{Z}} \Rightarrow \mathbf{F} := \frac{\dot{\mathbf{X}}}{Z}$$

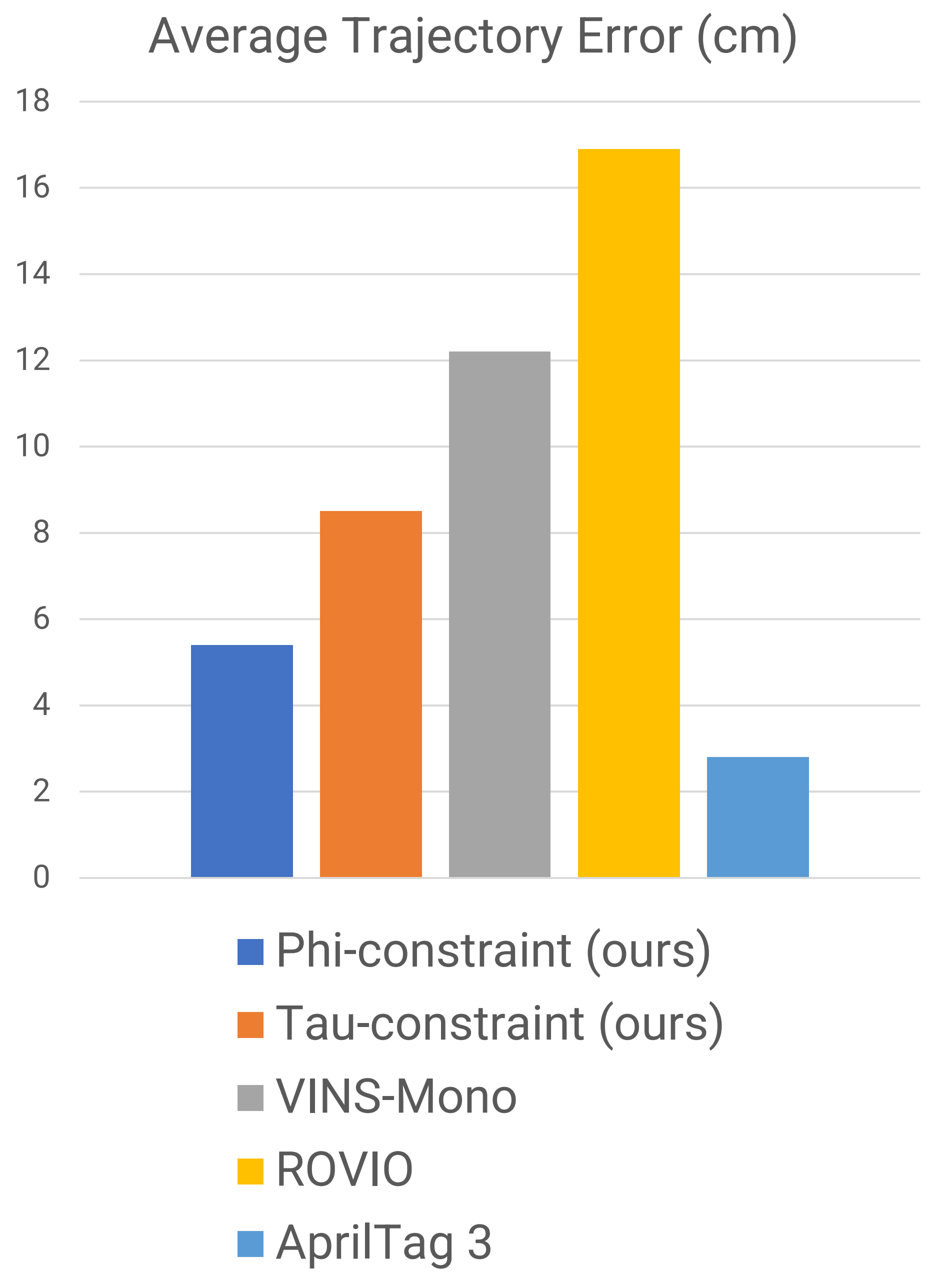


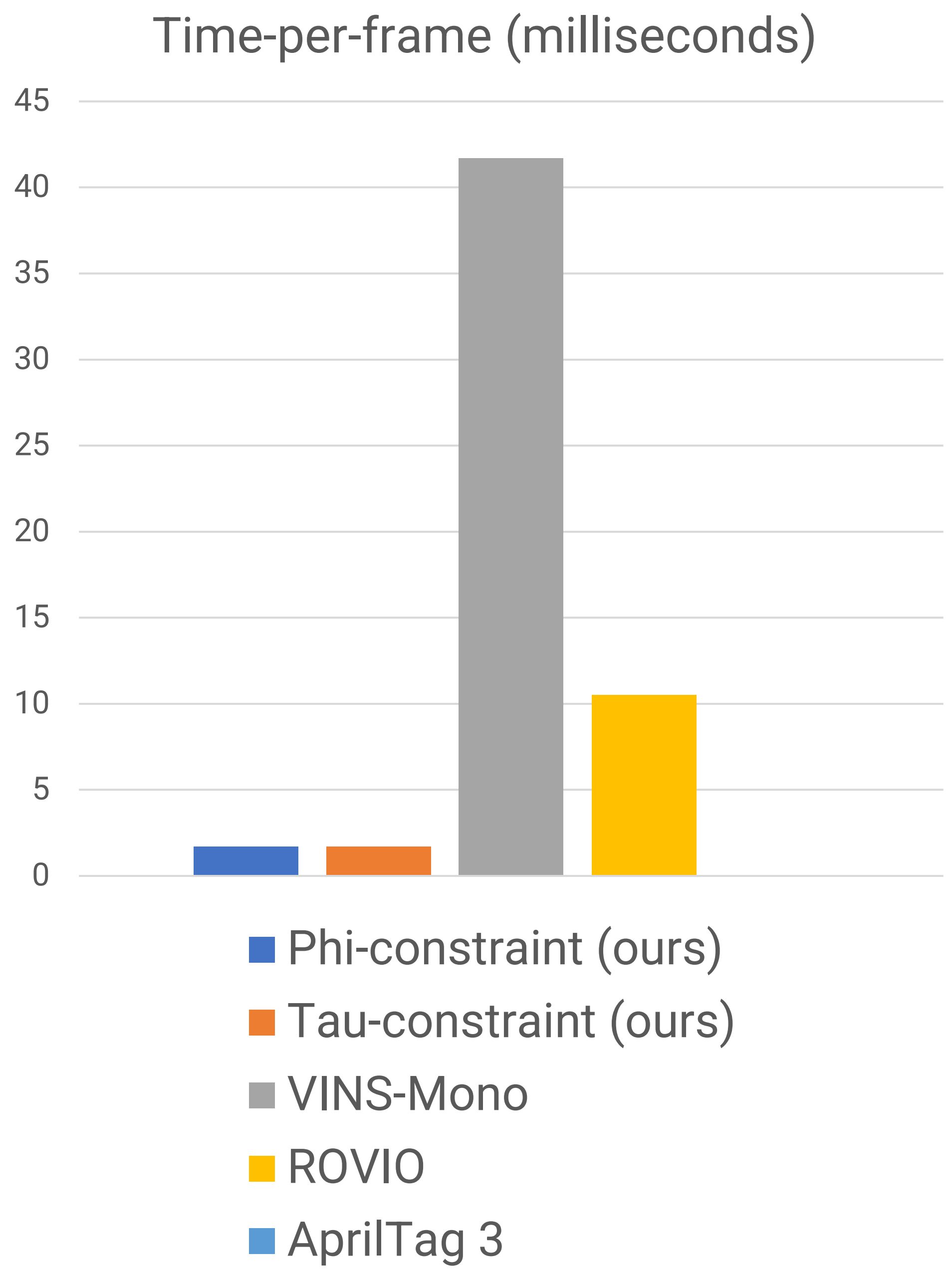
Sequence 2

Sequence 4

Sequence 1 Trajectory/Error

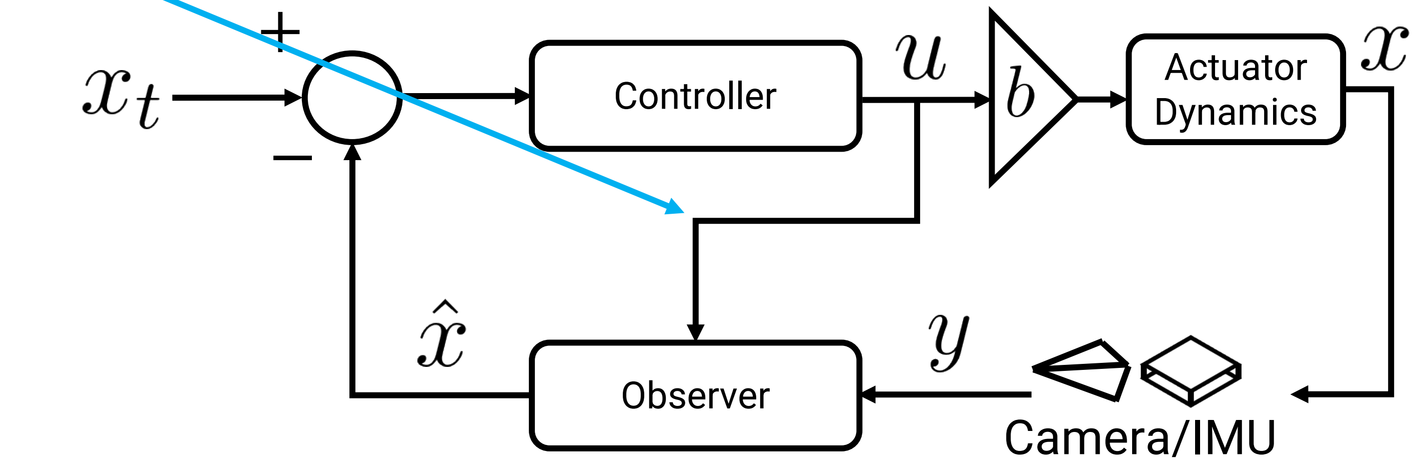



Average Trajectory Error (cm)
18
16
14
12
10
8
6
4
2
0
Phi-constraint (ours)
Tau-constraint (ours)
VINS-Mono
ROVIO
AprilTag 3


Time-per-frame (milliseconds)
45
40
35
30
25
20
15
10
5
0
Phi-constraint (ours)
Tau-constraint (ours)
VINS-Mono
ROVIO
AprilTag 3

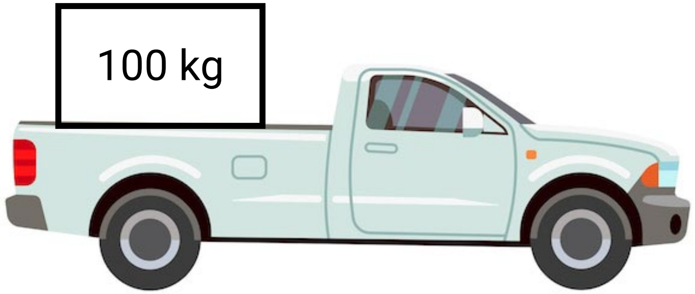
Stability Invariance

- We still have not used control effort in observer

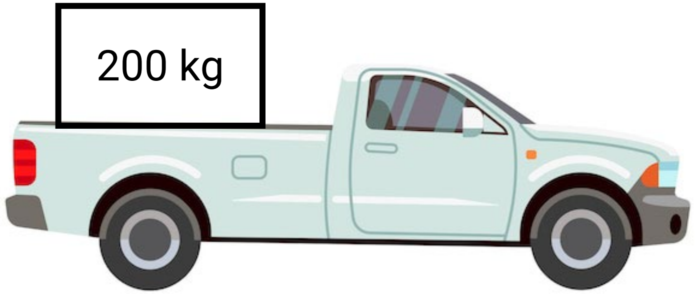


Control effort and acceleration

$$\ddot{x} = \frac{1}{100}u$$



$$\ddot{x} = \frac{1}{200}u$$



Scaled State Estimates Using “u”

- Recall the Phi/Tau constraints result in linear least squares problems

$$\underset{Z_0,\dot{Z}_0,g_Z}{\operatorname{argmin}} \left\| \left(\Phi_{F_Z} - 1\right) Z_0 - r\dot{Z}_0 + \mathcal{J}\{a_Z^m + g_Z\} \right\|_2^2$$

$$\implies \begin{bmatrix} Z \\ \dot{Z} \\ g_Z \end{bmatrix} = (A^T A)^{-1} A^T \ddot{x}$$

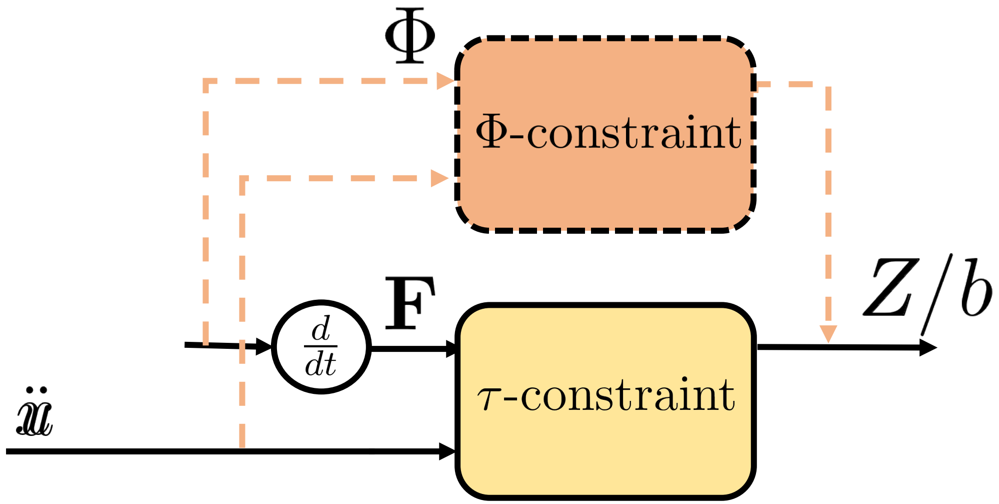
- Using control effort in place of acceleration results in scaled state estimate

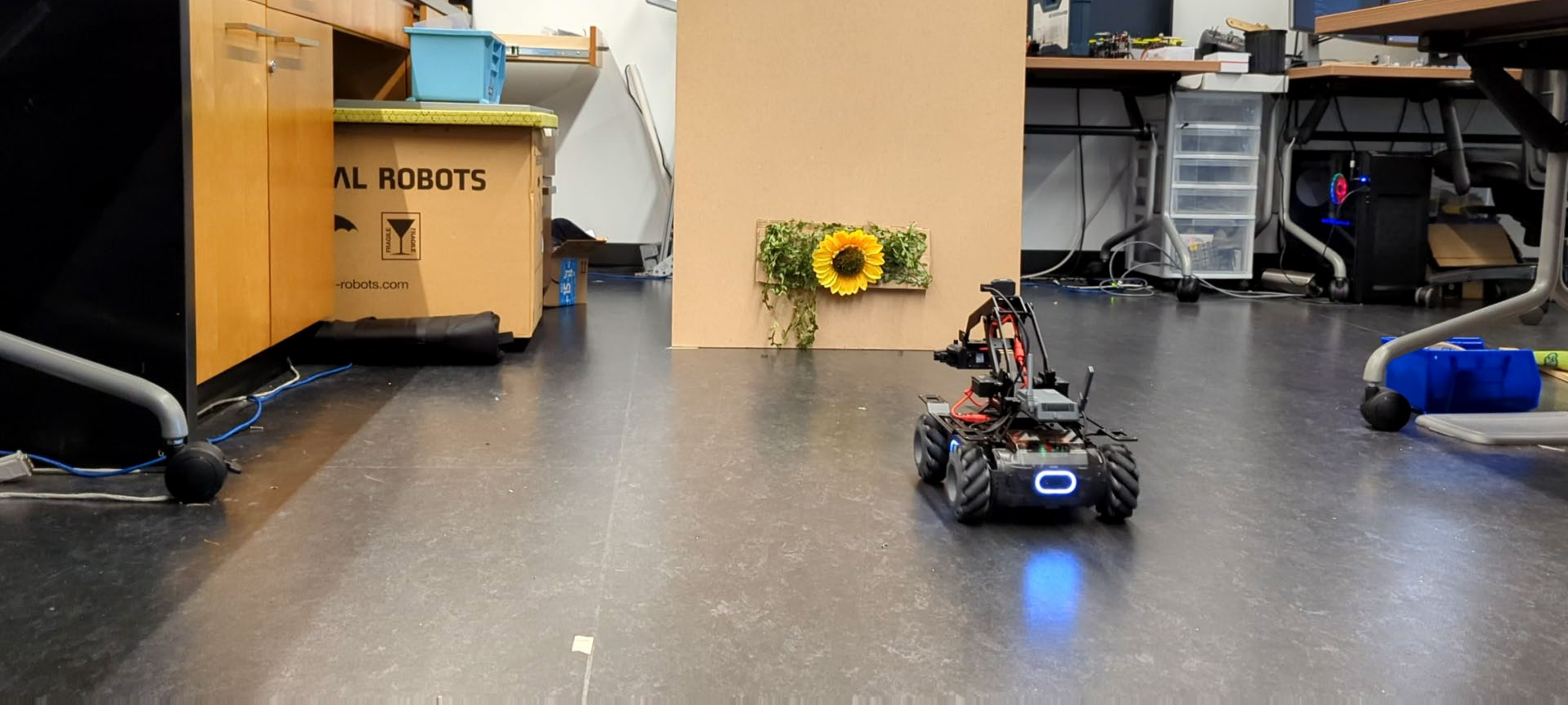
$$\ddot{x} = bu \implies \ddot{x}/b = u$$

$$\implies \begin{bmatrix} Z/b \\ \dot{Z}/b \\ g_Z/b \end{bmatrix} = (A^T A)^{-1} A^T u$$

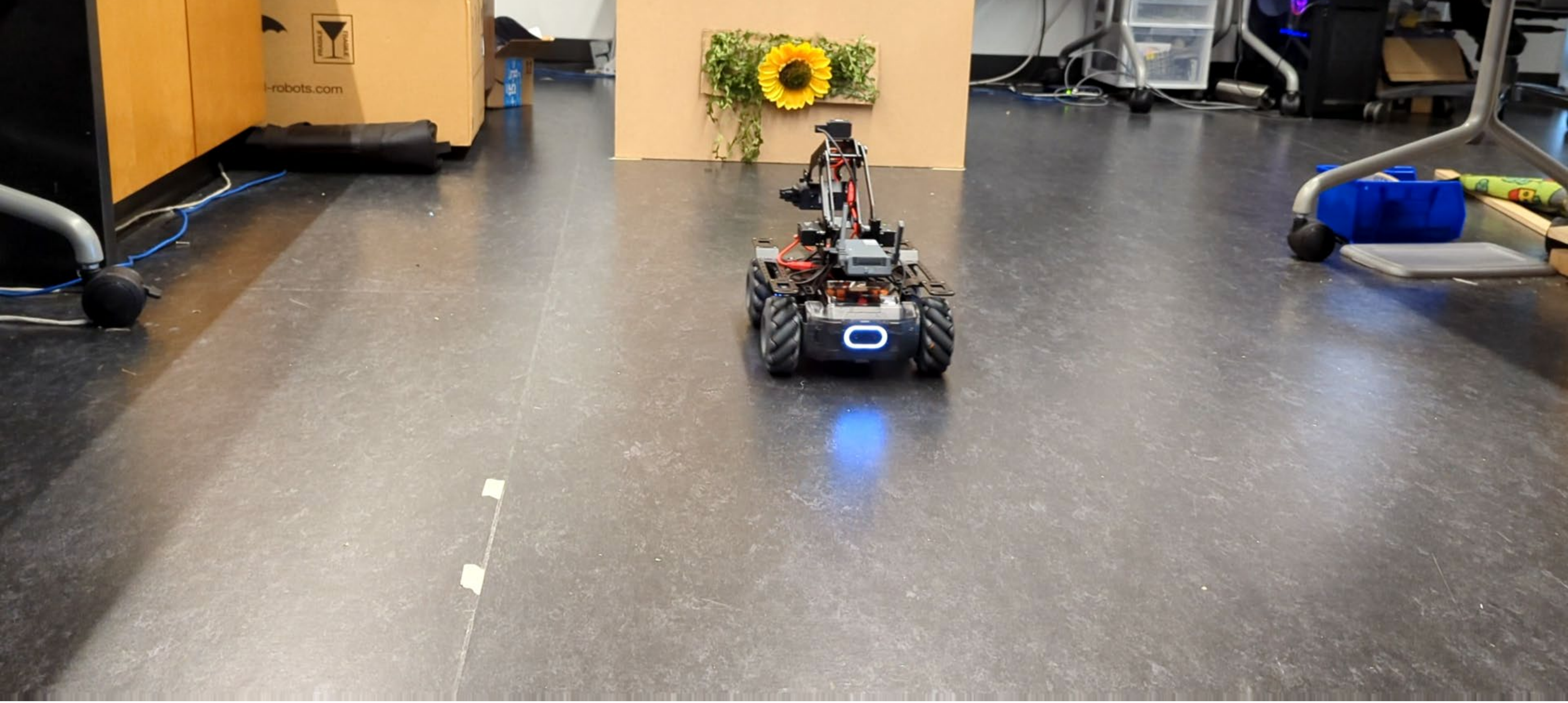
Scaled State Estimates Using “u”

$$\ddot{x} = bu = \cancel{b}kZ/\cancel{b}$$

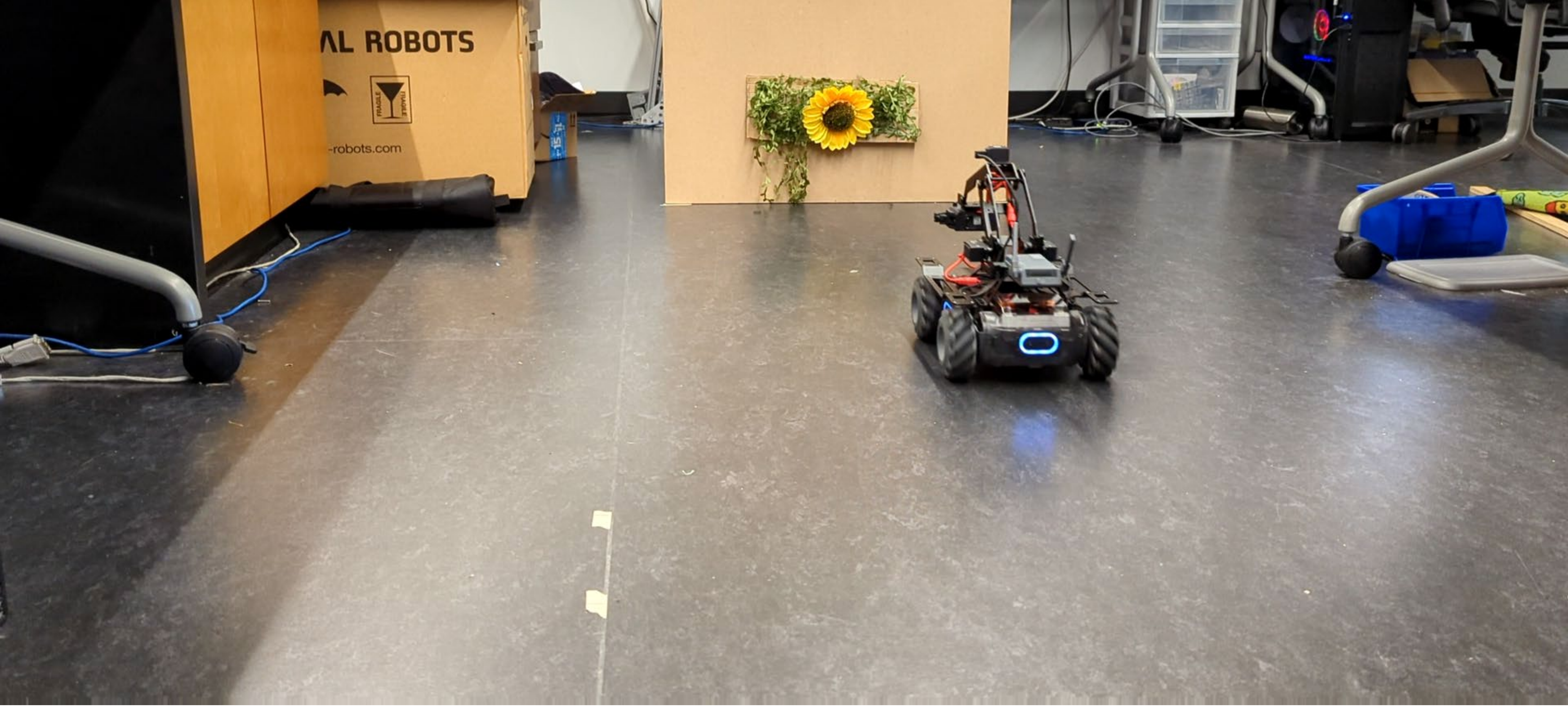




b = 1, measured acceleration



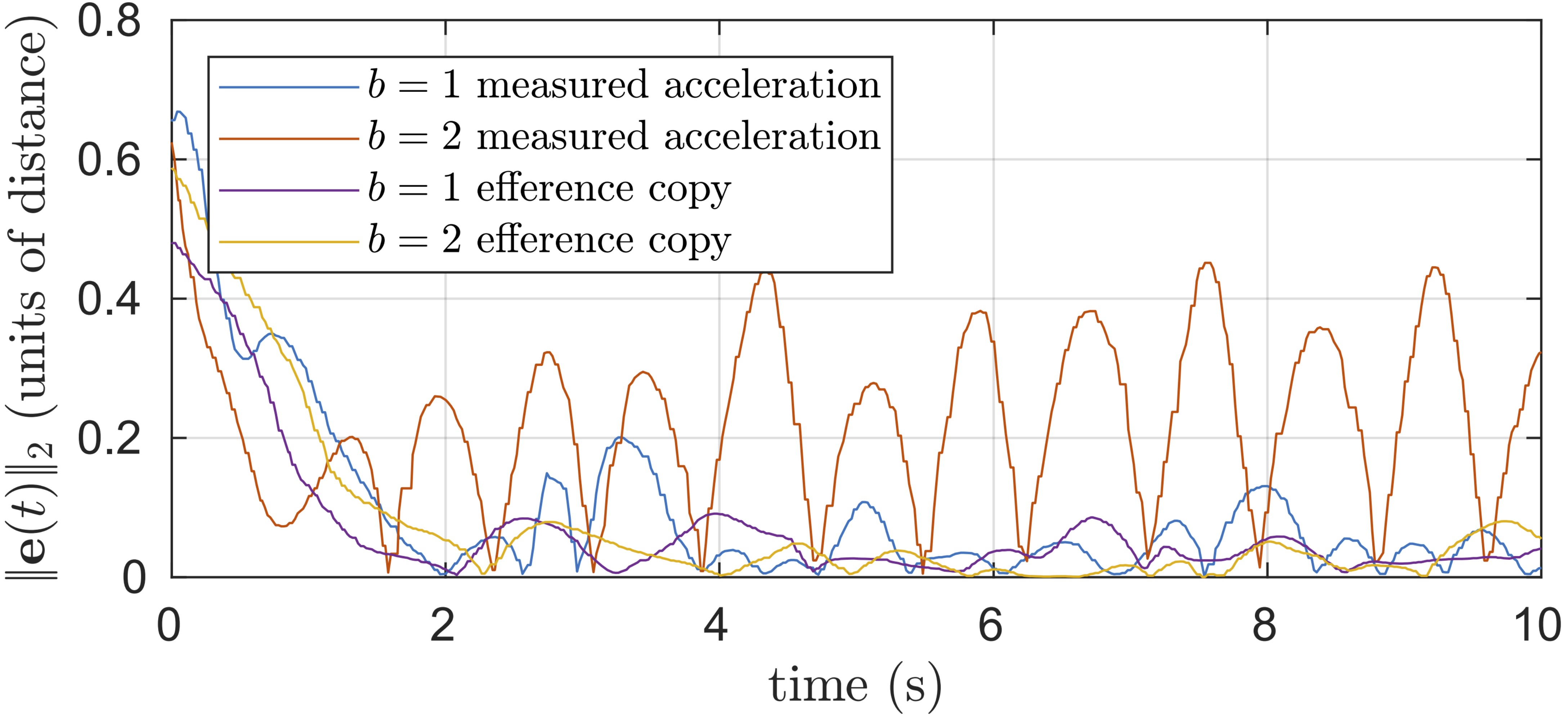
b = 1, efference copy



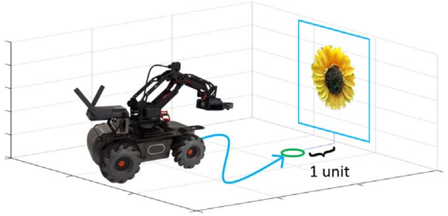
b = 2, measured acceleration

b = 2, efference copy

Oscillations



How to Use This Approach?



$$\implies \mathbf{X}_c = \Phi_c(t)\mathbf{X}_0$$

- Three Easy Steps
  - Measure a bounding box with a camera
    - Bounding box will give you Phi or F
  - Relate bounding box params to motion model
  - Apply optimization to a window of time
    - Can use a recursive observer, pure optimization, etc
    - Feedback linearized models will result in linear problems

$$\dot{\mathbf{X}}_m = f(t, \mathbf{X}_m, u)$$

$$\implies \mathbf{X}_m = \Phi_f(t, \mathbf{X}_0, u)$$

$$\min_{\mathbf{X}_0} \|\mathbf{X}_m - \mathbf{X}_c\|$$

- Full Paper: TTCDist: Fast Distance Estimation From an Active Monocular Camera Using Time-to-Contact, Levi Burner, Nitin J. Sanket, Cornelia Fermüller, Yiannis Aloimonos https://arxiv.org/abs/2203.07530

Conclusion

- Took a control theoretic approach to monocular distance estimation
- Found a linear equality constraint that allows fast and accurate estimation of distance
  - Achieved competitive trajectory estimation performance
  - 6.2x and 25x faster than ROVIO and VINS-Mono
  - 30-70% less centimeters of average trajectory error
- Found that in certain cases, stability margins become invariant
  - Idea should be more fully developed into a form of adaptive control

Thanks to my Collaborators and Sponsors!

Prof. Nitin Sanket

Dr. Cornelia Fermüller

Prof. Yiannis Aloimonos