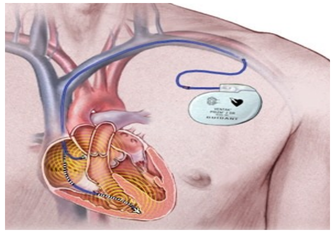
CAREER: Context-Aware Runtime Safety Assurance in Medical Human-Cyber-Physical Systems

Homa Alemzadeh

Dependable Systems and Analytics (UVA-DSA)
Electrical and Computer Engineering
University of Virginia

Medical Cyber-Physical Systems (MCPS)

Pacemakers



Insulin Pumps

Wearable Monitors

Patient Monitors

Infusion Pumps

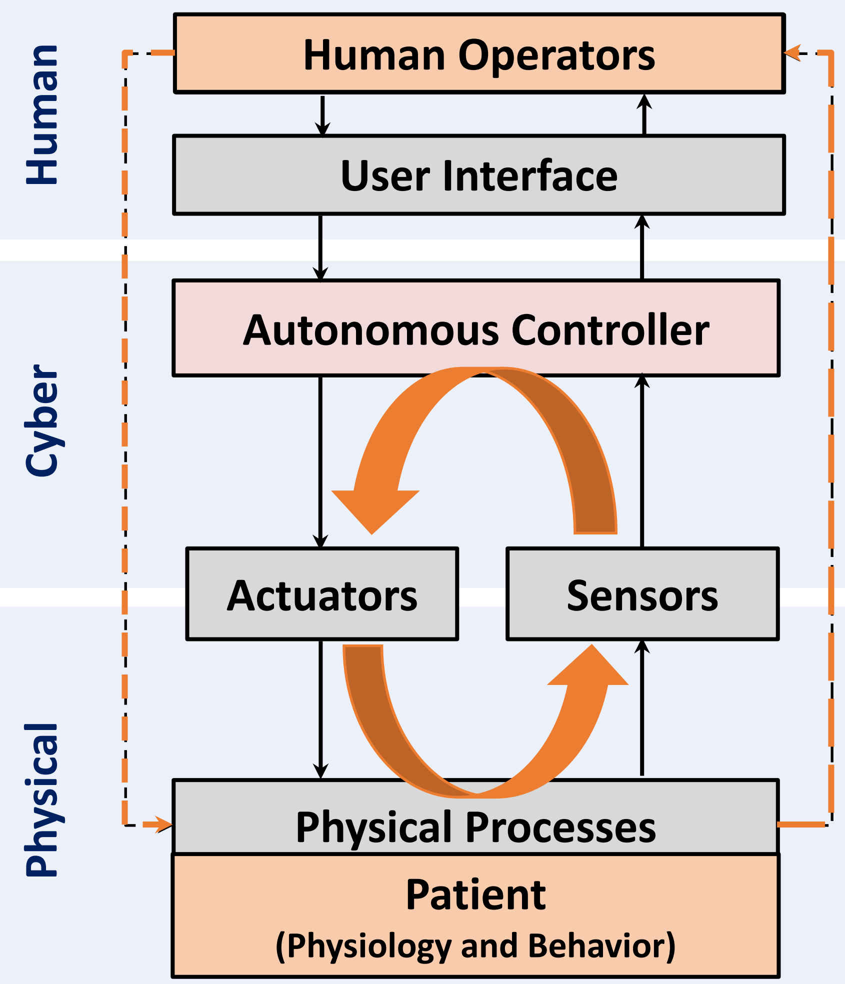
Defibrillators

Surgical Robots

Imaging Systems

Linear Accelerators

Human-in-the-loop MCPS



- Advanced medical expertise
- Human supervision and control
- Tele-operation

- Complex and connected software
- AI-enabled controllers
- Timing or resource constraints

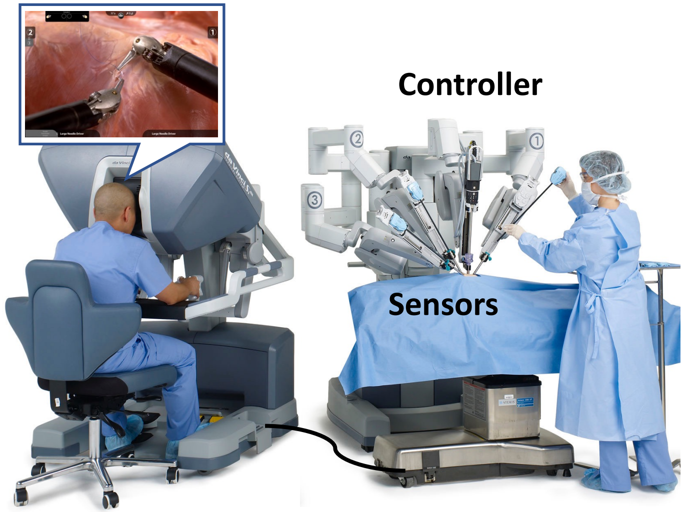
- Variety of patient profiles
- Physiological and behavioral dynamics

Levels of Autonomy in MCPS

Cognitive Assistants for EMS

Level 0 – Decision Support

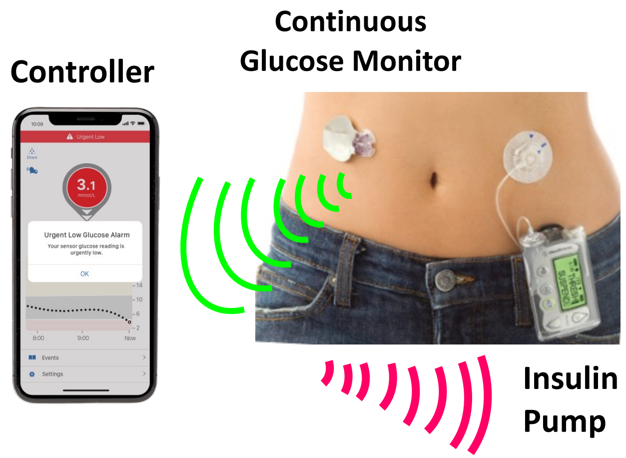
The human actuates the physical system and autonomous controller provides feedback.



Tele-operated Surgical Robots

Level 1 – Tele-Operation

The human actuates the autonomous controller who controls the physical system.



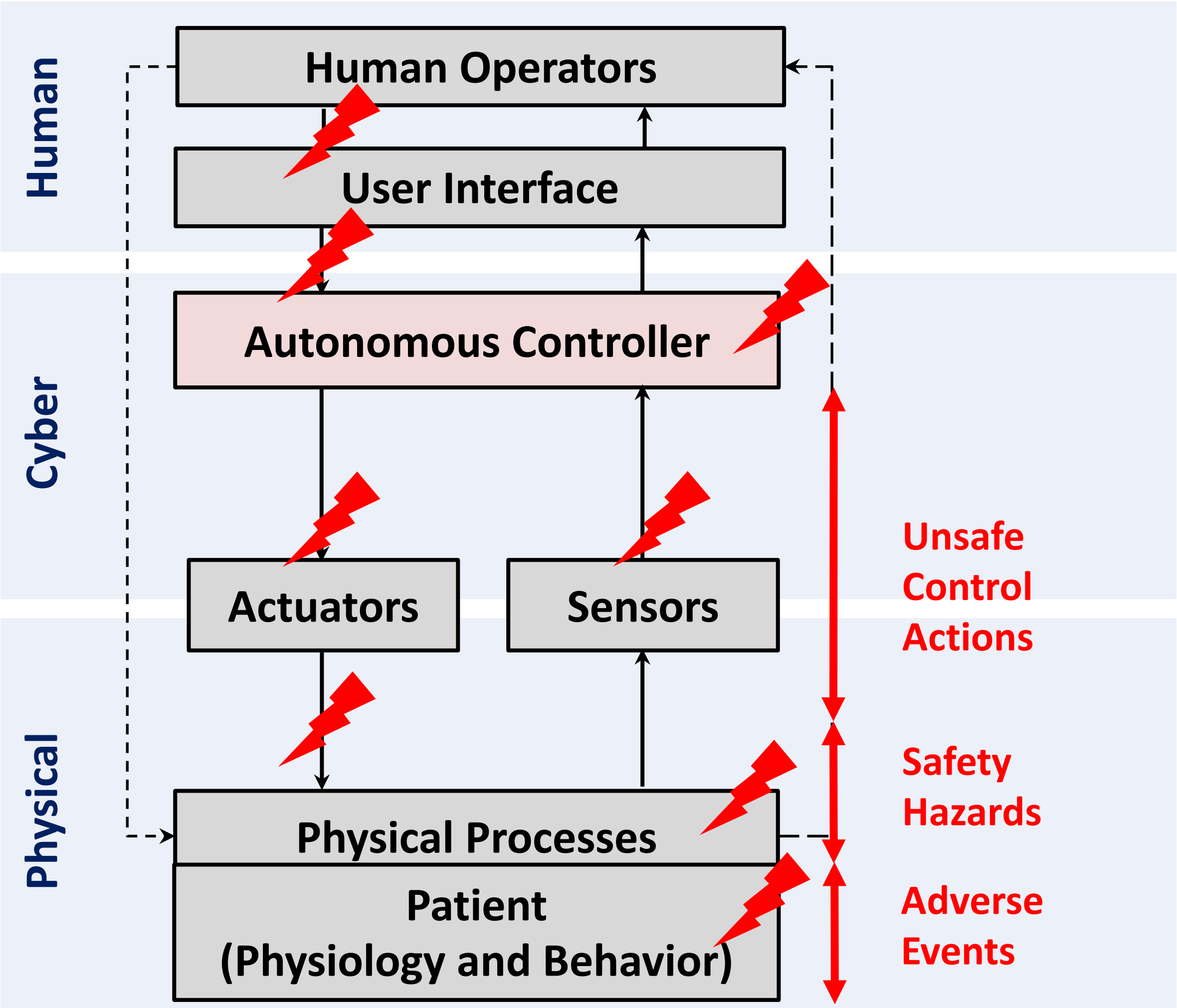
Artificial Pancreas Systems

Level 2 – Autonomous

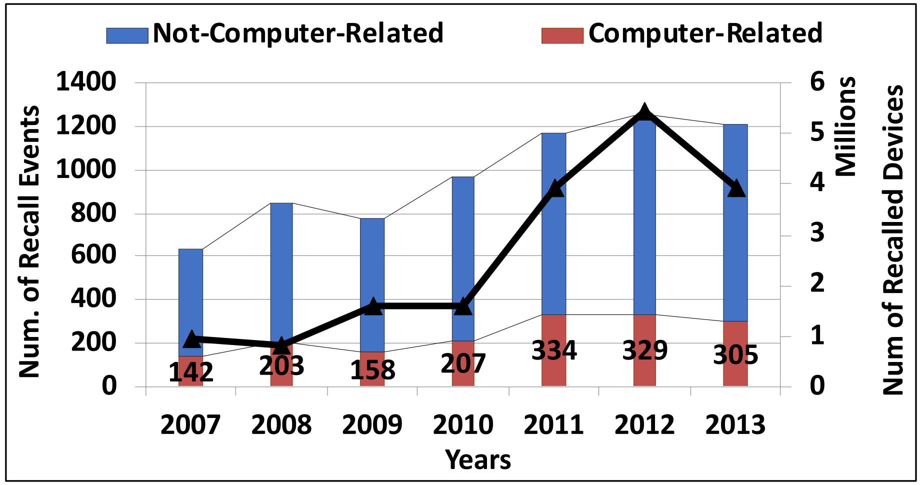
The autonomous controller actuates the physical system and the human monitors it.

Inspired by the SAE Levels of Driving Automation.

Safety and Security Vulnerabilities

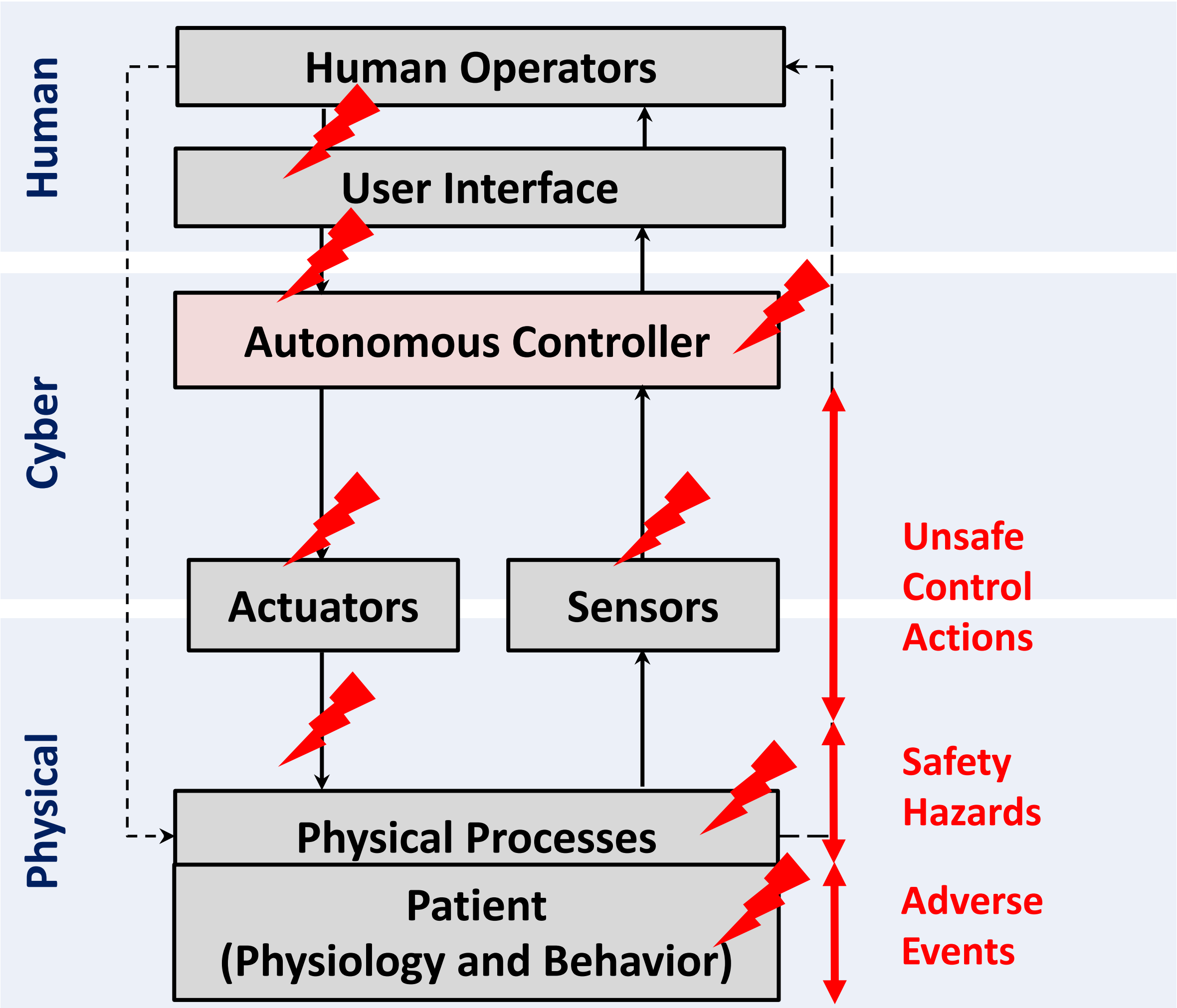


- 2007-2013: Over 6,800 FDA recalls
  - Over 18 million devices
  - ~ 24% computer failures
  - 12% safety-critical
- 2014-2020: Over 7,100 FDA recalls
  - 13% software related

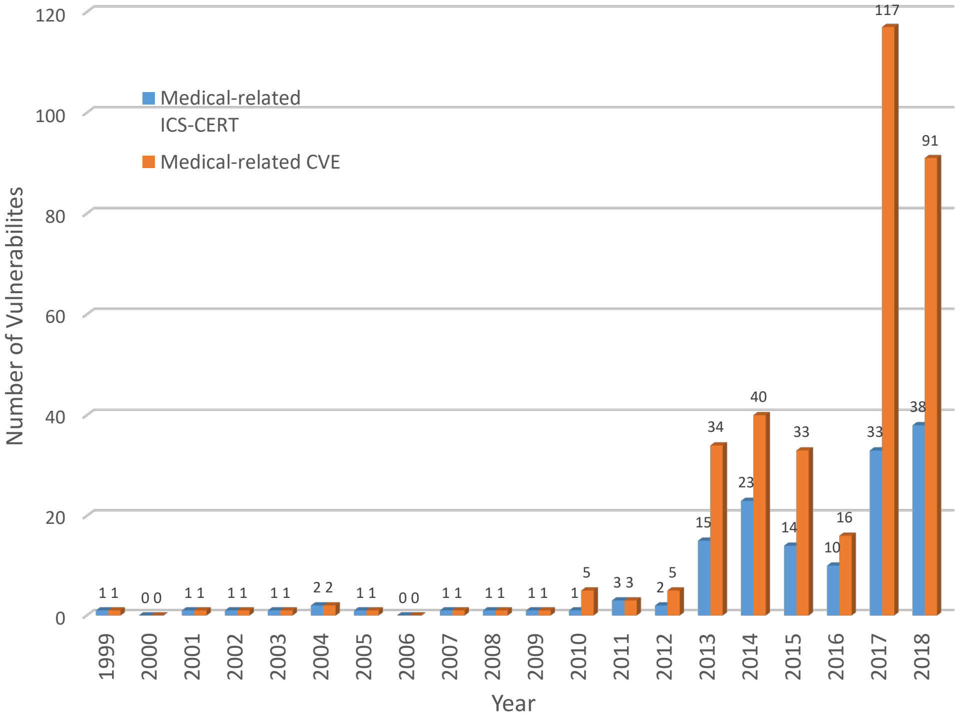


H. Alemzadeh, *et al*, "Analysis of Safety-Critical Computer Failures in Medical Devices," *IEEE Security & Privacy Magazine*, 2013.
Y. Xu, *et al.*, "Analysis of Cyber-Security Vulnerabilities of Interconnected Medical Devices, " *IEEE/ACM CHASE*, 2019.

Safety and Security Vulnerabilities

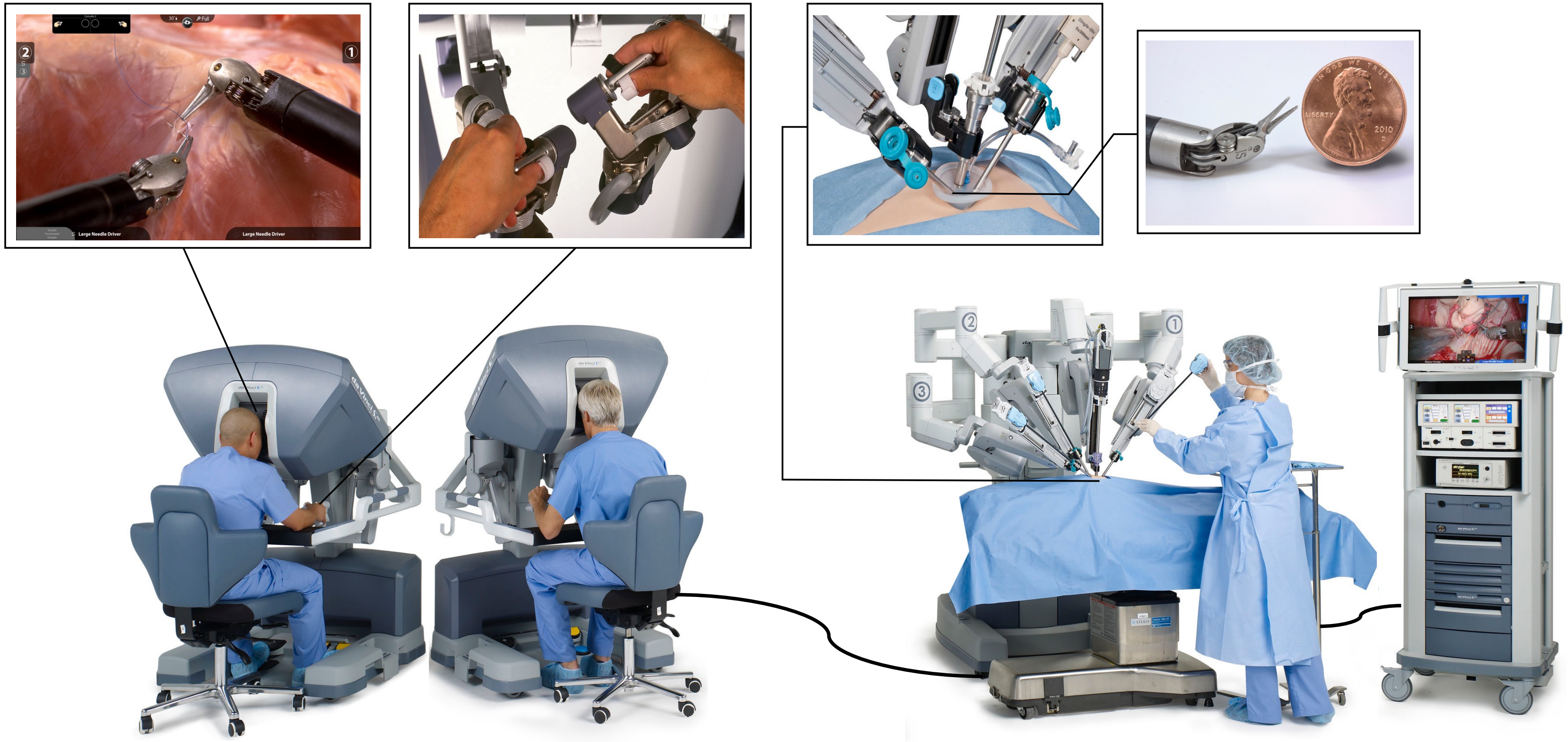


- **1999-2018: 354 CVEs were reported to affect interconnected medical devices.**
- **Steady increase, 2.5 times since 2013**
  - **38 ICS-CERT and 91 CVE records in 2018.**



H. Alemzadeh, *et al*, "Analysis of Safety-Critical Computer Failures in Medical Devices," *IEEE Security & Privacy Magazine*, 2013.
Y. Xu, D. Tran, Y. Tian, H. Alemzadeh, "Analysis of Cyber-Security Vulnerabilities of Interconnected Medical Devices," *IEEE/ACM CHASE,* 2019.

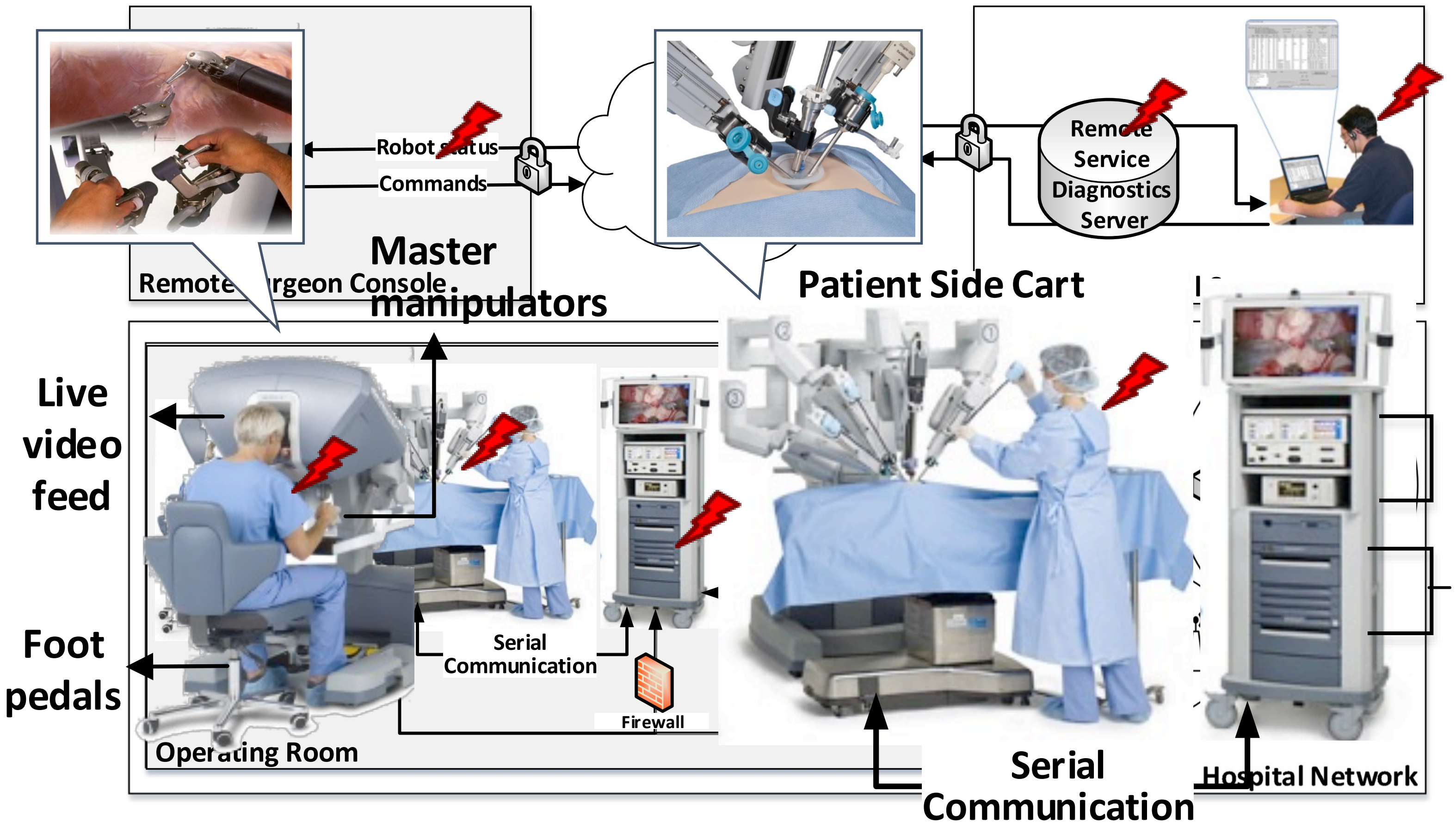
Tele-operated Surgical Robots



Loosely Closed-loop Semi-autonomous: No haptics, limited vision feedback

Credits: da Vinci Surgical System, Intuitive Surgical, Inc.

Tele-operated Surgical Robots



Unintended Human Errors *[IJMRCAS 2022]*

DOS and MITM Attacks *[ICCPS 2015]*

Software, hardware, mechanical faults *[PLOS ONE 2016]*

Faulty firewalls *[WIRED 2014]*

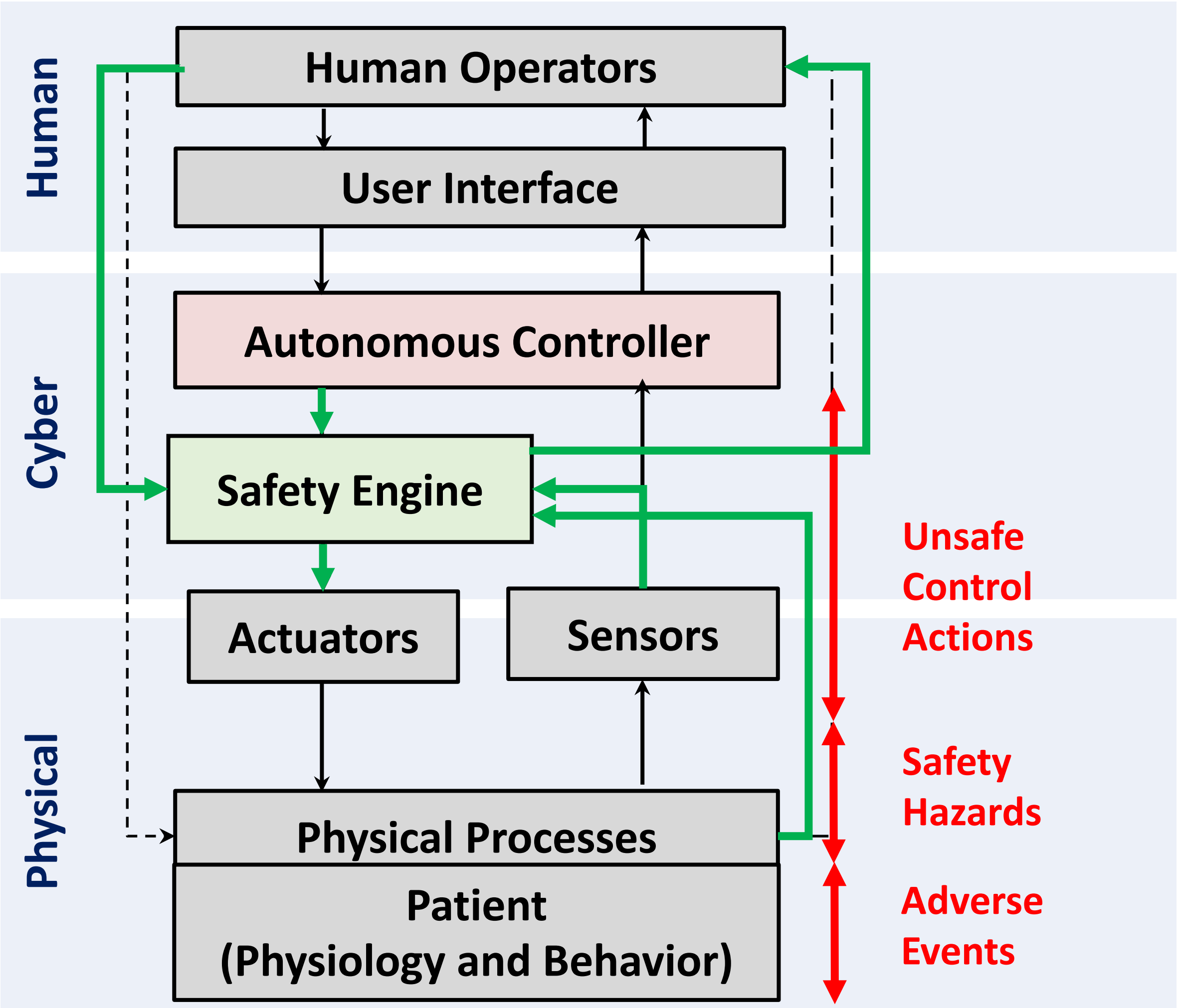
Malware targeting control software *[DSN 2016]*

H. Alemzadeh, *et al.* "Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data". *PLOS ONE*, 2016.
K. Hutchinson, Z. Li, N. Schenkman, L. Cantrell, Homa Alemzadeh,"Analysis of Executional and Procedural Errors in Dry-lab Robotic Surgery Experiments," *IJMRCAS*, 2022.
H. Alemzadeh, *et al.*, "Targeted Attacks on Teleoperated Surgical Robots: Dynamic Model-Based Detection and Mitigation", DSN, 2016.

Challenges

- **Offline safety assurance techniques**
  - Hazard analysis and risk assessment
  - Model-based design, verification, and controller synthesis
  - **Inadequate in detecting residual faults, attacks on controller, and preventing adverse events**
- **Runtime anomaly detection and recovery methods**
  - Joint cyber-physical modeling and monitoring
  - Cyber-physical checkpointing, roll-forward recovery, simplex architectures, and ML controllers
  - **Focused on fully autonomous, not considering humans in the loop (operators and patients)**
- **Solely model-based or data-driven techniques**
  - Fixed rules based on domain knowledge and medical guidelines
  - Model-predictive Control (MPC) based on simple linear or complex non-linear models
  - Black-box machine learning using limited data and non-transparent logic
  - **Issues with generalizability, robustness, and transparency**

Context-Aware Runtime Safety Assurance

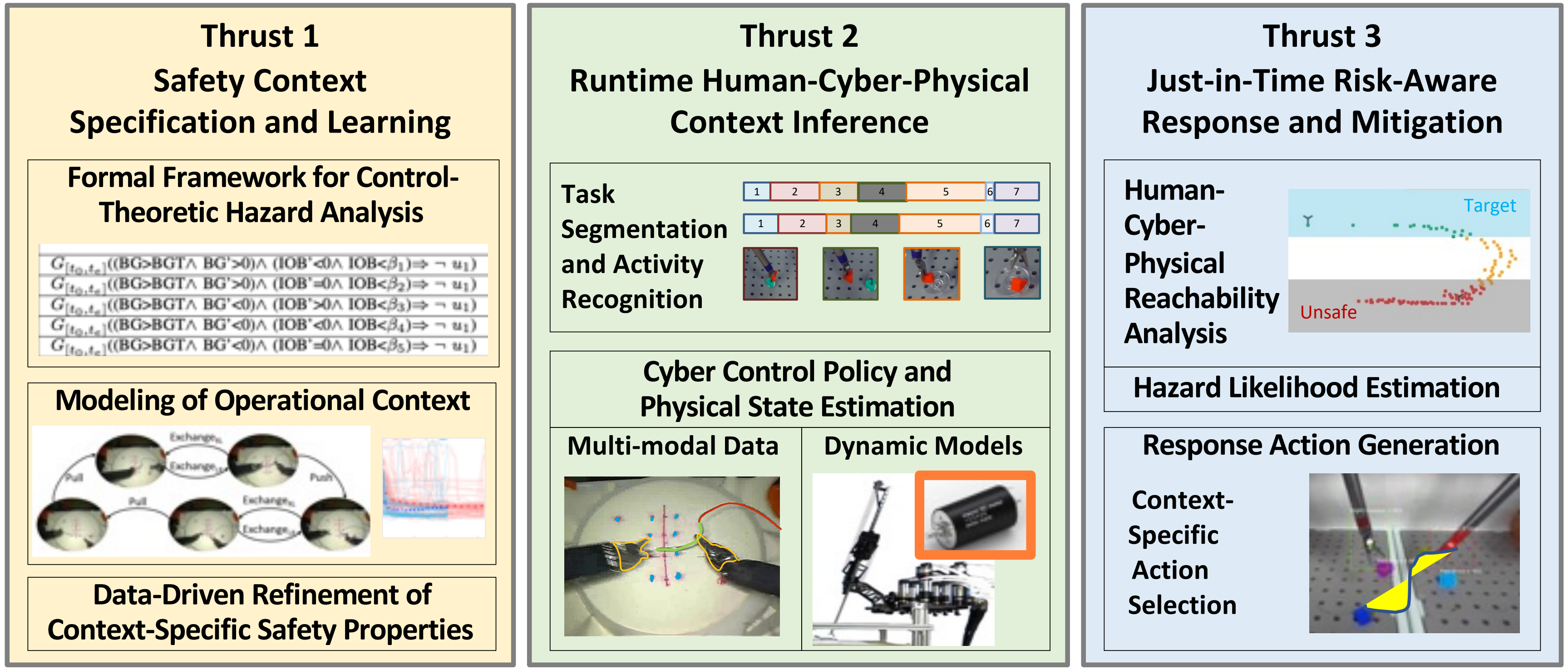


- **Preemptive Detection of Unsafe Control Actions**
  - Based on sensor data and control commands
  - Just before execution in physical layer
- **Prediction of Safety Hazards**
  - Time and likelihood estimation
- **Prevention of Adverse Events**
  - Hazard mitigation and recovery
  - Feedback to human operators

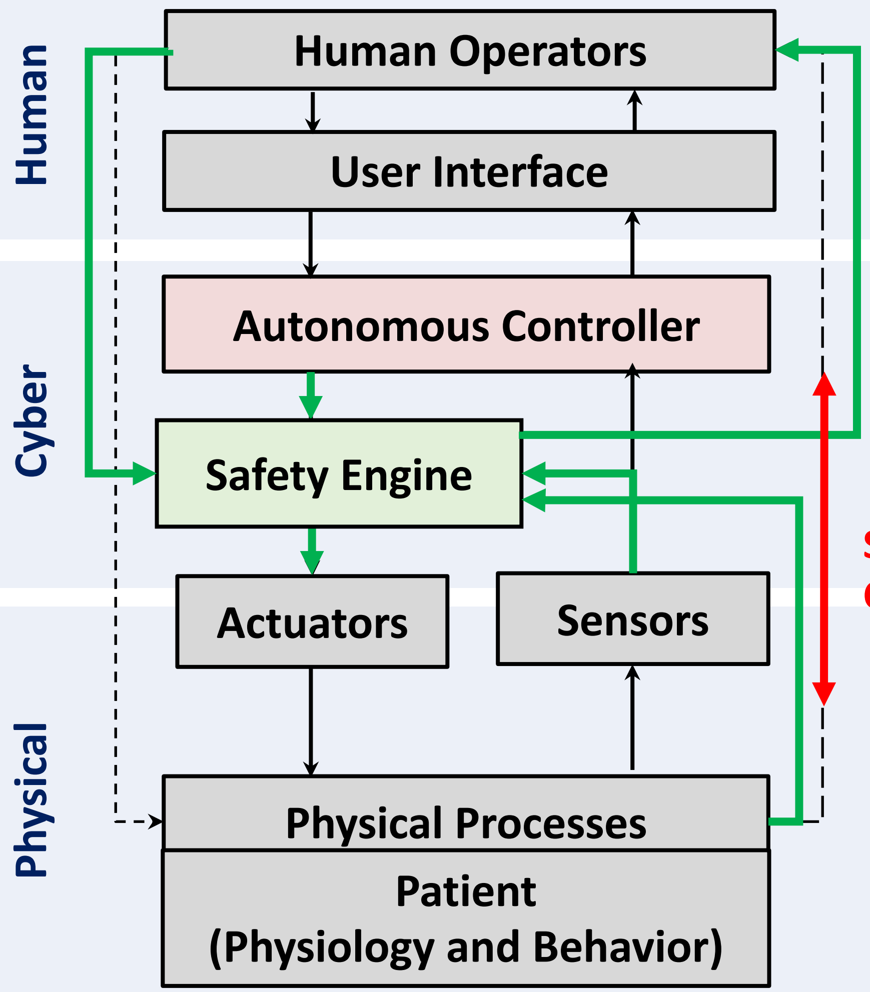
Context-Aware Runtime Safety Assurance

An integrated model and data-driven approach:

- Bridge the gap between offline formal modeling and runtime monitoring
- Consider domain knowledge, human-cyber-physical context, and operator/patient profiles
- Design principles for safety engines applicable to medical, robotics, and autonomous systems

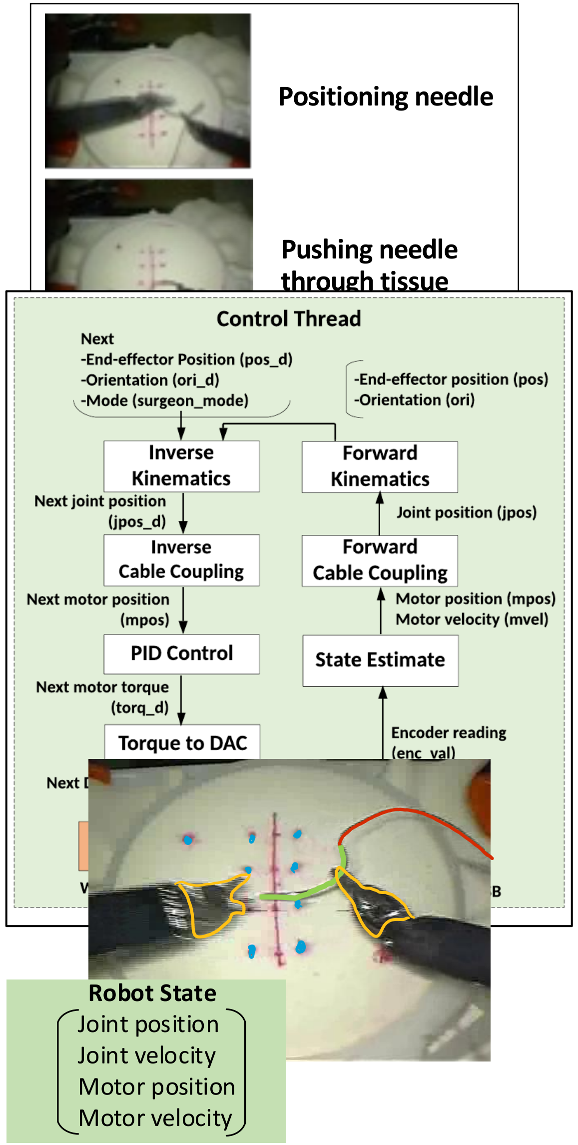


Human-Cyber-Physical Context



Safe or Unsafe Control Action?

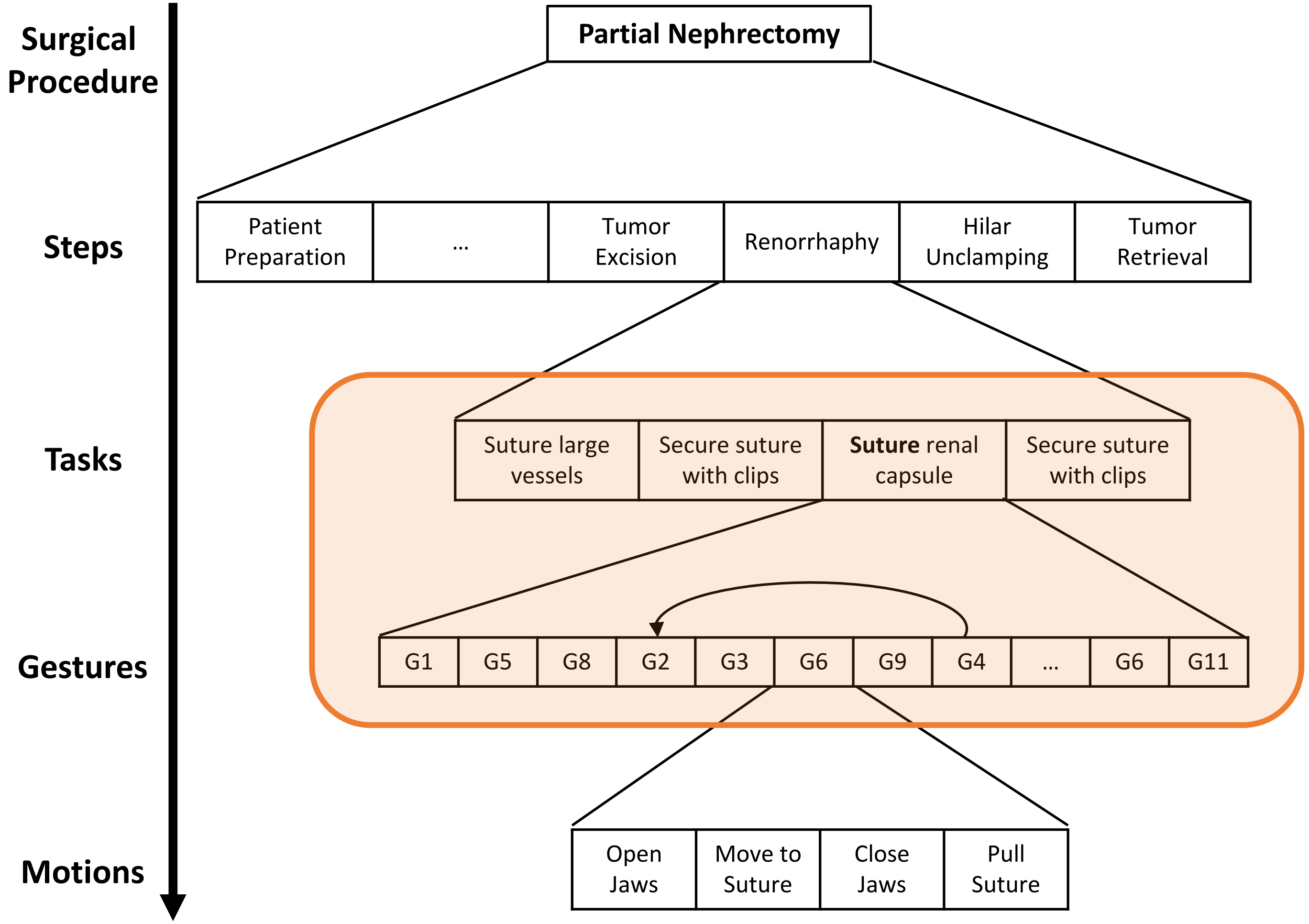
- **Operational Context**
  - Operator workflows, tasks, actions
  - User intent, style, and skill level
- **Cyber Context**
  - Control system and software states
- **Physical Context**
  - Physical and environmental states
  - Patient profiles and dynamics

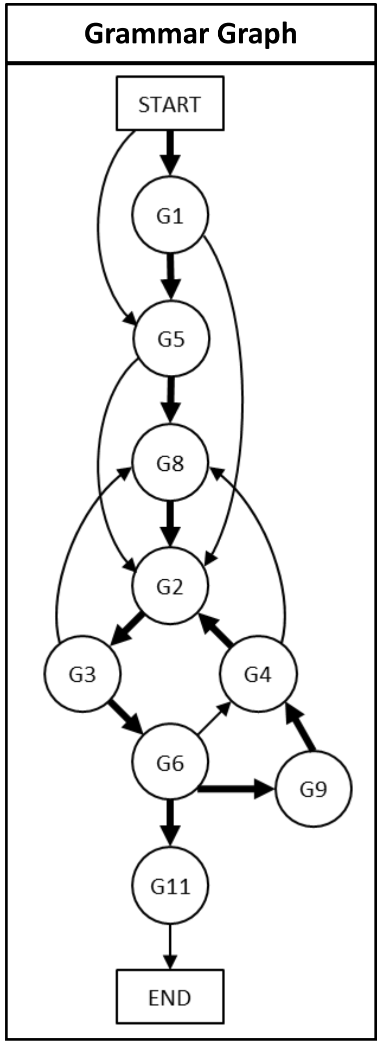


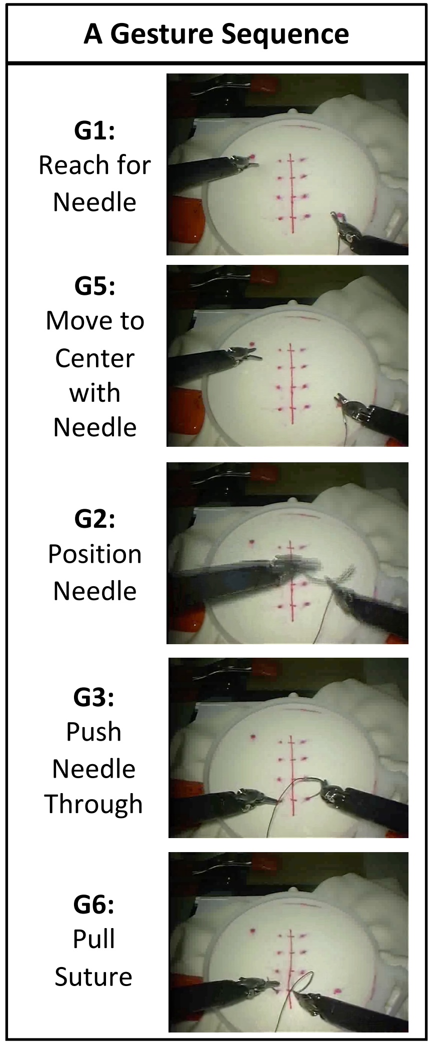
Thrust 1:
Safety Context Specification and Learning

- Framework for formal specification of human-cyber-physical safety context
  - Control-theoretic hazard analysis for specification of unsafe control actions based on multi-dimensional system context
  - Template temporal logic formulas for context-dependent hazard prediction and mitigation
- Modeling of operational context and relationship to cyber-physical context
  - Hierarchical and generalized modeling of surgical tasks and context
- Data-driven refinement of safety context specifications
  - Optimization of logic formulas using fault-free and faulty patient-specific data
  - Guided adversarial model training based on safety specifications

Modeling of Operational Context



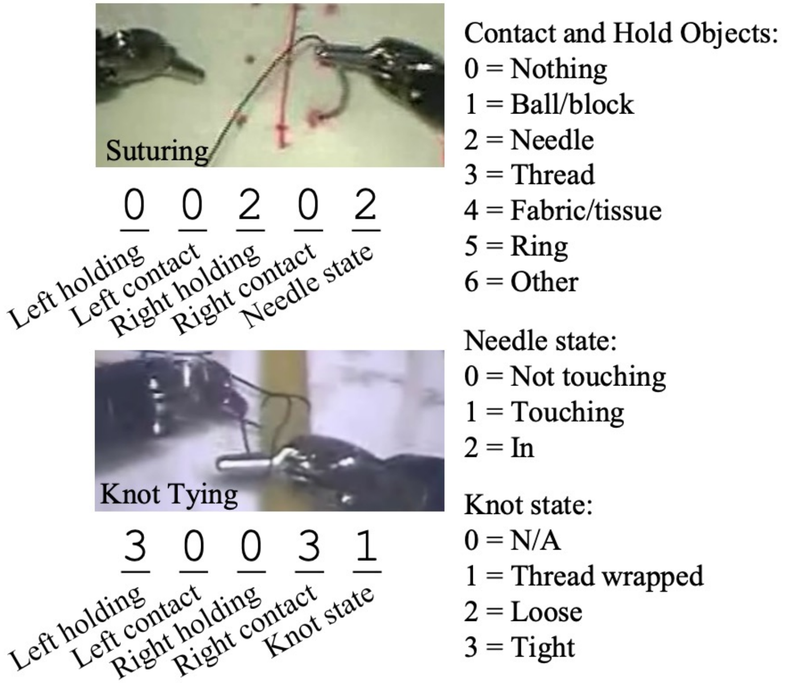


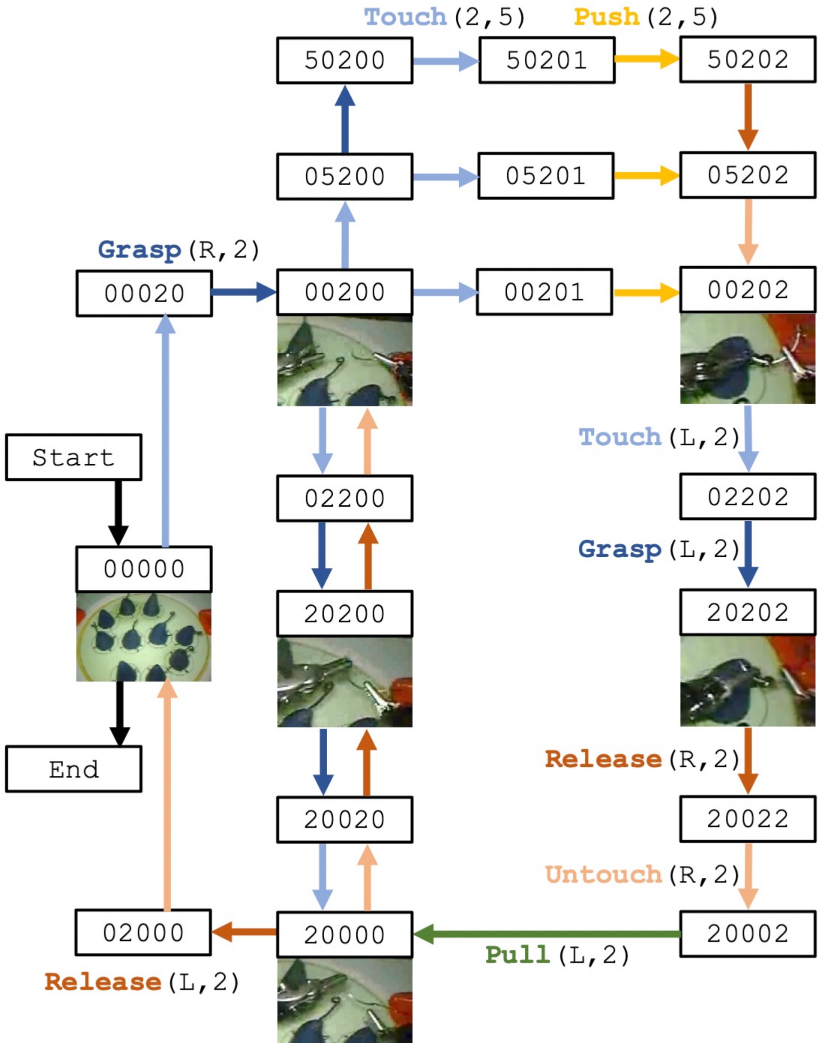


K. Hutchinson, I. Reyes, Z. Li, H. Alemzadeh, "COMPASS: A Formal Framework and Aggregate Dataset for Generalized Surgical Procedure Modeling," *Under Review.*

Modeling of Operational Context

- **Operational to physical context mapping**
  - Surgical tasks as finite-state MDPs
  - Change in physical context by the execution of motion primitives





K. Hutchinson, I. Reyes, Z. Li, H. Alemzadeh, "COMPASS: A Formal Framework and Aggregate Dataset for Generalized Surgical Procedure Modeling," *Under Review.*

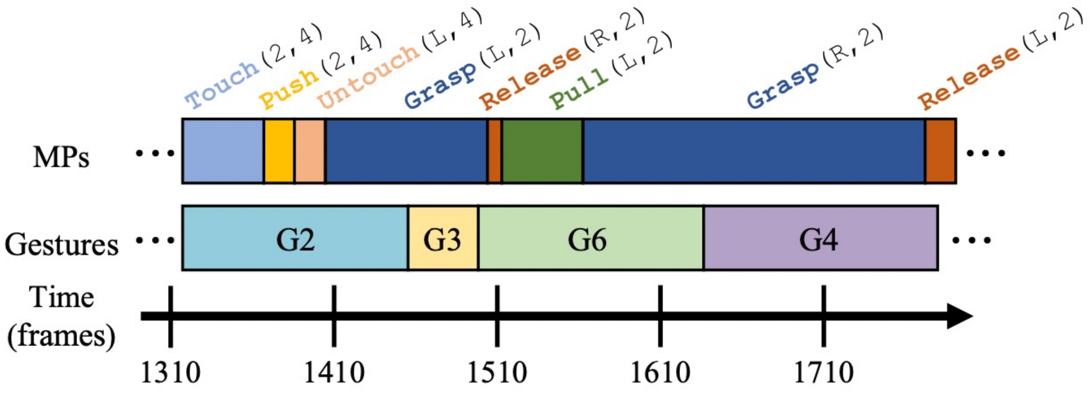
Thrust 2:
Runtime Human-Cyber-Physical Context Inference

- Surgical Context Detection and Segmentation
  - Real-time surgical action recognition
  - Multi-modal surgical scene segmentation

- Cyber Controller State Estimation
  - Mapping low-level control commands to kinematic state variables
  - Dynamic modeling of robotic joints and motor controllers

Operational Context Inference

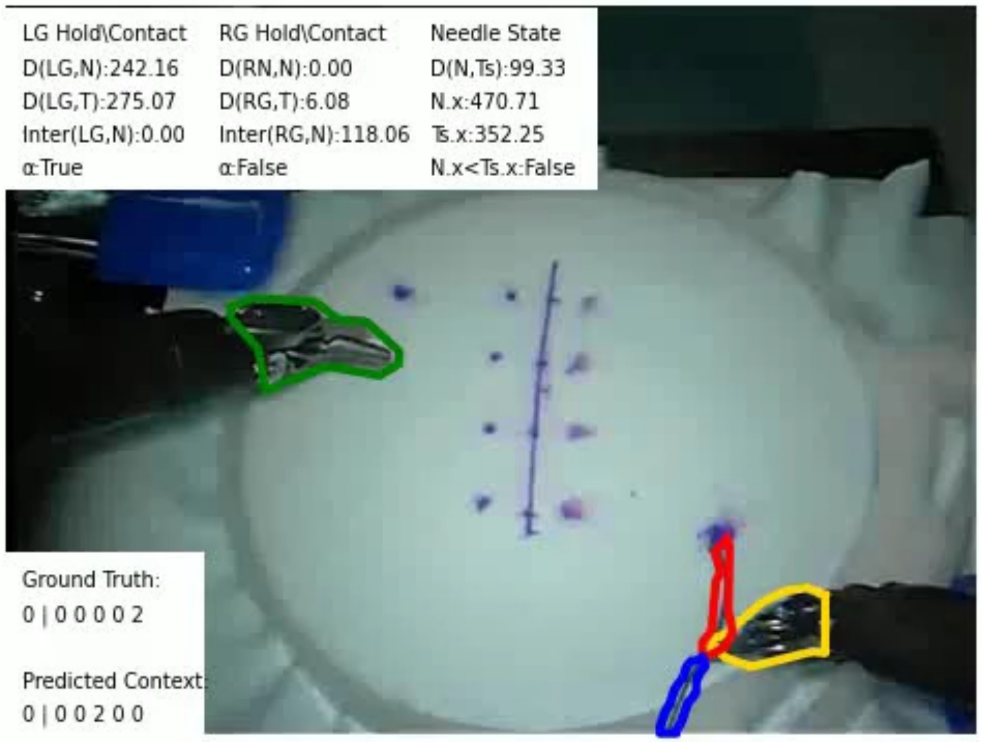
- **Task Segmentation and Action Recognition**
  - Gesture and motion primitive prediction
  - Aggregate surgical dataset for robust model training
  - Leave-One-Task-Out (LOTO) cross validation for generalizability



K. Hutchinson, I. Reyes, Z. Li, H. Alemzadeh, “COMPASS: A Formal Framework and Aggregate Dataset for Generalized Surgical Procedure Modeling,” *Under Review.*

Physical Context Inference

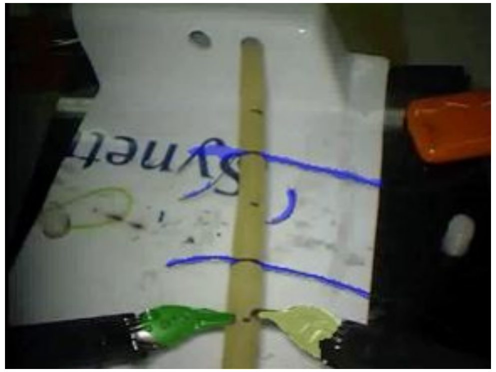
- Semantic segmentation: Surgical tool and object localization
- Context detection: Contour extraction and overlap detection



Suturing

Needle Passing

Knot Tying



K. Hutchinson, Z. Li, I. Reyes, H. Alemzadeh, “Towards Surgical Context Inference and Translation to Gestures,” *Under Review*.

Thrust 3:
Just-in-Time Risk-Aware Response and Mitigation

- Human-Cyber-Physical Reachability Analysis for Hazard Prediction
  - Prediction of human, cyber, and physical states
  - Estimate likelihood and timing of hazards

- Context-Dependent Response Action Prioritization and Decision Making
  - Prioritize and select sequences of corrective actions
  - Timely and safe execution of motion primitive trajectories
  - Real-time feedback to human operators

Education and Outreach

- Promote participation of undergraduate researchers and K-12 students from diverse backgrounds in the areas of engineering and robotics in medicine.

- **Autonomous manipulation of a robotic arm**
  - Basic training task of “Pick and place” in robotic surgery
- **K-12 and public outreach events**
  - Biomed-Tech-Girls: Robotic programming challenge
  - UVA Engineering Open House
- **Hands-on projects for a core course in NRT CPS curriculum**
  - Real-time embedded computing systems
  - C programming on RTOS and TI micro-controller/launchpad