Assured Autonomy

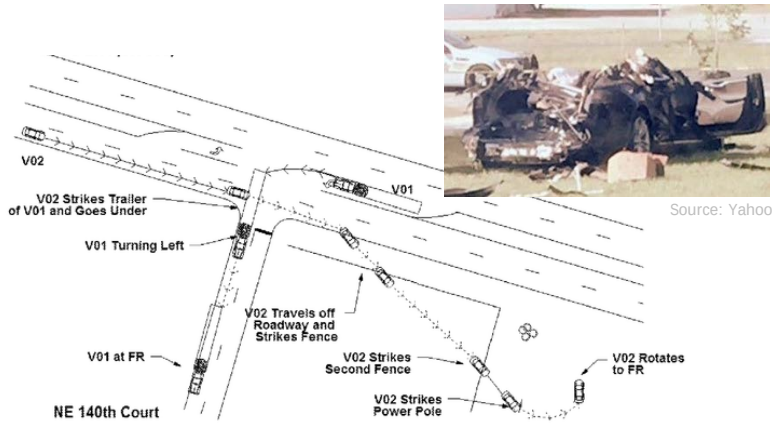
Sandeep Neema, I2O

August 15, 2017

Assurance for Autonomous Systems is Hard

"The decision for DoD to deploy autonomous systems must be based both on trust that they will perform effectively in their intended use and that such use will not result in high-regret, unintended consequences. *Without such trust, autonomous systems will not be adopted*...it is therefore important for DoD to focus on critical trust issues and the *assurance* of appropriate levels of trust."
- DSB Report on Autonomy, June 2016

"The notion that autonomous systems can be fully tested is becoming increasingly infeasible as higher levels of self governing systems become a reality...*the standard practice of testing all possible states and all ranges of inputs to the system becomes an unachievable goal*. Existing TEVV methods are, by themselves, insufficient for TEVV of autonomous systems; therefore *a fundamental change is needed in how we validate and verify these systems*."
- OSD TEV&V Strategy Report, May 2015



Source: Yahoo

Source: https://www.nytimes.com/2016/07/01/business/self-driving-tesla-fatal-crash-investigation.html

Traditional testing will require exorbitant time and money: 11B miles, 500 years, $6B
- Driving to Safety, RAND Corp. Report, 2016

Table 1. Examples of Miles and Years Needed to Demonstrate Autonomous Vehicle Reliability

		Benchmark Failure Rate		
Statistical Question	How many miles (years[a]) would autonomous vehicles have to be driven...	(A) 1.09 fatalities per 100 million miles?	(B) 77 reported injuries per 100 million miles?	(C) 190 reported crashes per 100 million miles?
	(1) without failure to demonstrate with 95% confidence that their failure rate is at most...	275 million miles (12.5 years)	3.9 million miles (2 months)	1.6 million miles (1 month)
	(2) to demonstrate with 95% confidence their failure rate to within 20% of the true rate of...	8.8 billion miles (400 years)	125 million miles (5.7 years)	51 million miles (2.3 years)
	(3) to demonstrate with 95% confidence and 80% power that their failure rate is 20% better than the human driver failure rate of...	11 billion miles (500 years)	161 million miles (7.3 years)	65 million miles (3 years)

[a] We assess the time it would take to compete the requisite miles with a fleet of 100 autonomous vehicles (larger than any known existing fleet) driving 24 hours a day, 365 days a year, at an average speed of 25 miles per hour.

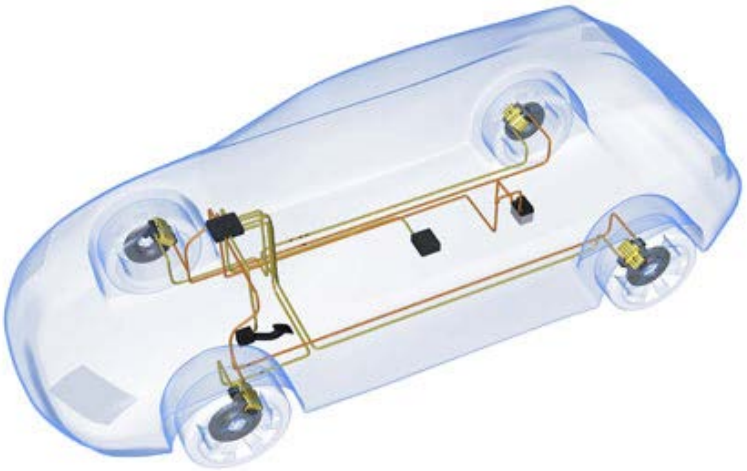
Program Goal

Develop rigorous design and analysis technologies for *continual assurance*[†] of *learning-enabled autonomous systems*, in order to guarantee *safety properties in adversarial environments*

[†]*assurance: a positive declaration intended to give confidence*

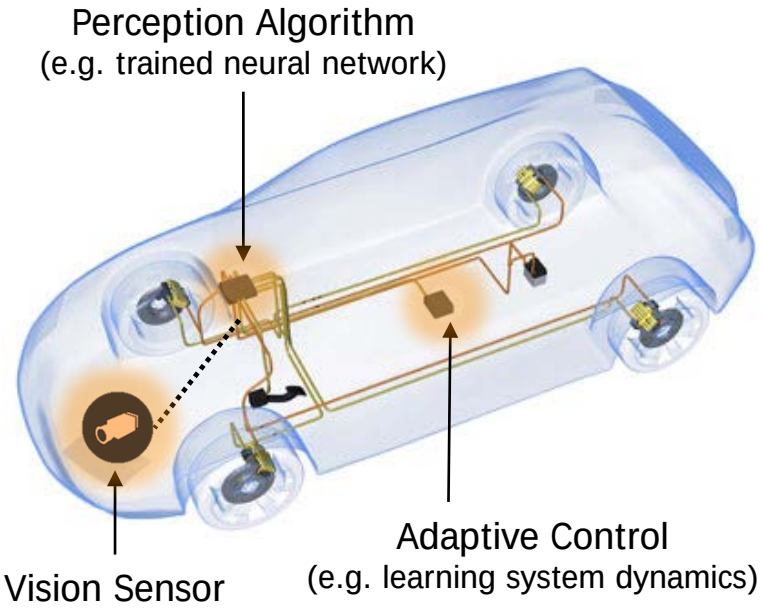
Illustrating the challenge

Non-Learning System
(e.g. manual brake-by-wire)



Safety assurance
can be provided

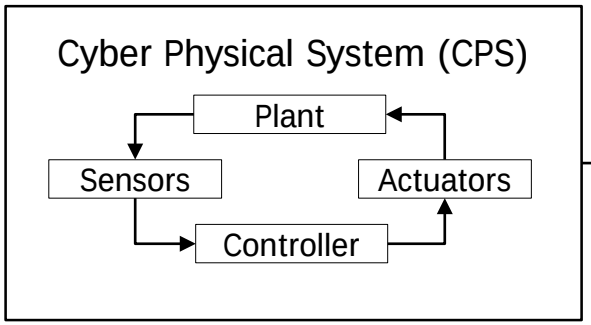
Learning-Enabled Autonomous System
(e.g. automated brake-by-wire for collision avoidance)



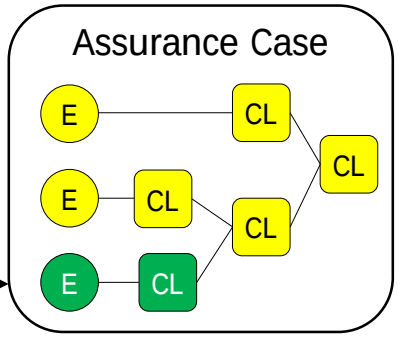
Safety assurance
can NOT be provided

Safety Assurance for Systems - State of Practice

Requirements | Design | Development | Test & Validation | Certs & Assurance



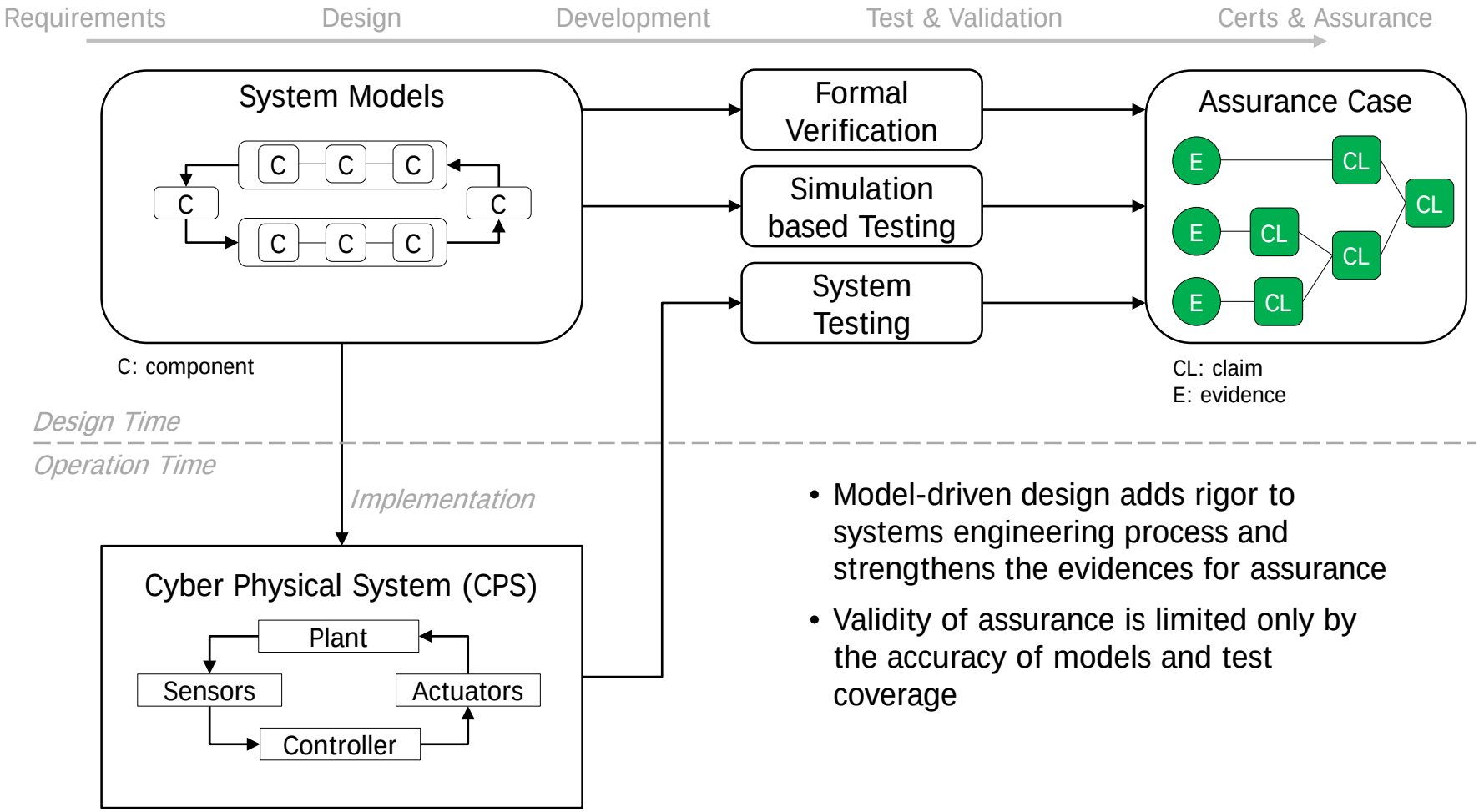
System Testing



CL: claim
E: evidence

- Assurance case: chain of reasoning to establish a claim with given evidence
- Claims specify system safety and security requirements to be assured
- Evidence is extracted from system test and other engineering data in a traditional systems engineering process

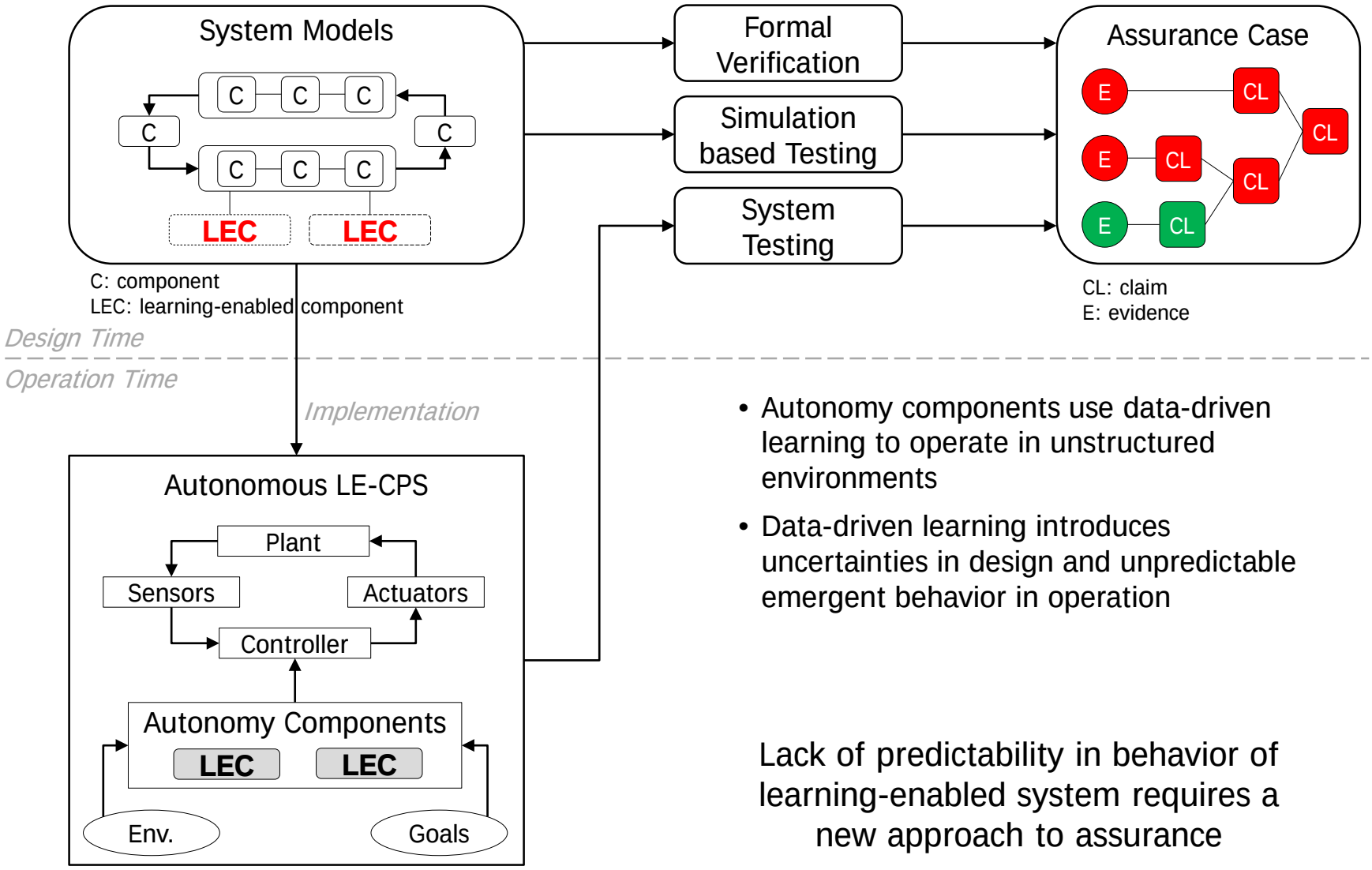
Model-driven Design for Safety Assurance - State of Art



- Model-driven design adds rigor to systems engineering process and strengthens the evidences for assurance
- Validity of assurance is limited only by the accuracy of models and test coverage

Applicable only to non-learning systems operating in well-characterized environments

Learning-Enabled Autonomous Systems Lack Safety Assurance



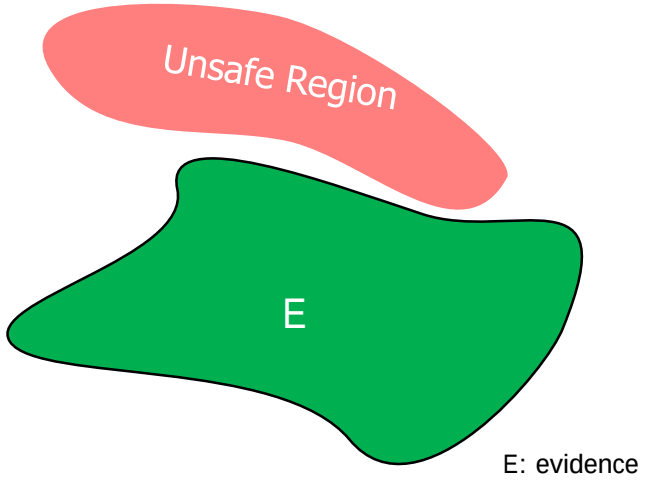
- Autonomy components use data-driven learning to operate in unstructured environments
- Data-driven learning introduces uncertainties in design and unpredictable emergent behavior in operation

Lack of predictability in behavior of learning-enabled system requires a new approach to assurance

Key Insight

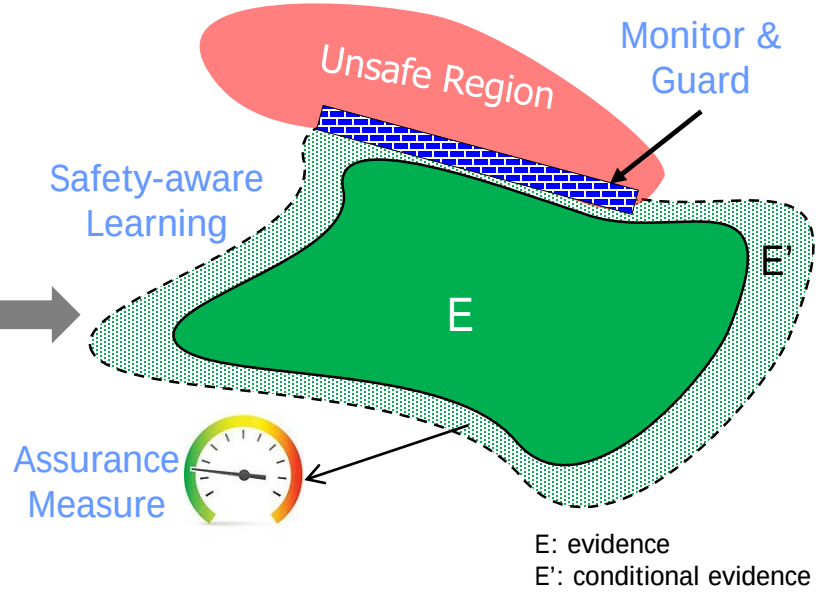
Non-Learning System

System Models → Formal Verification, Simulation based Testing, System Testing →

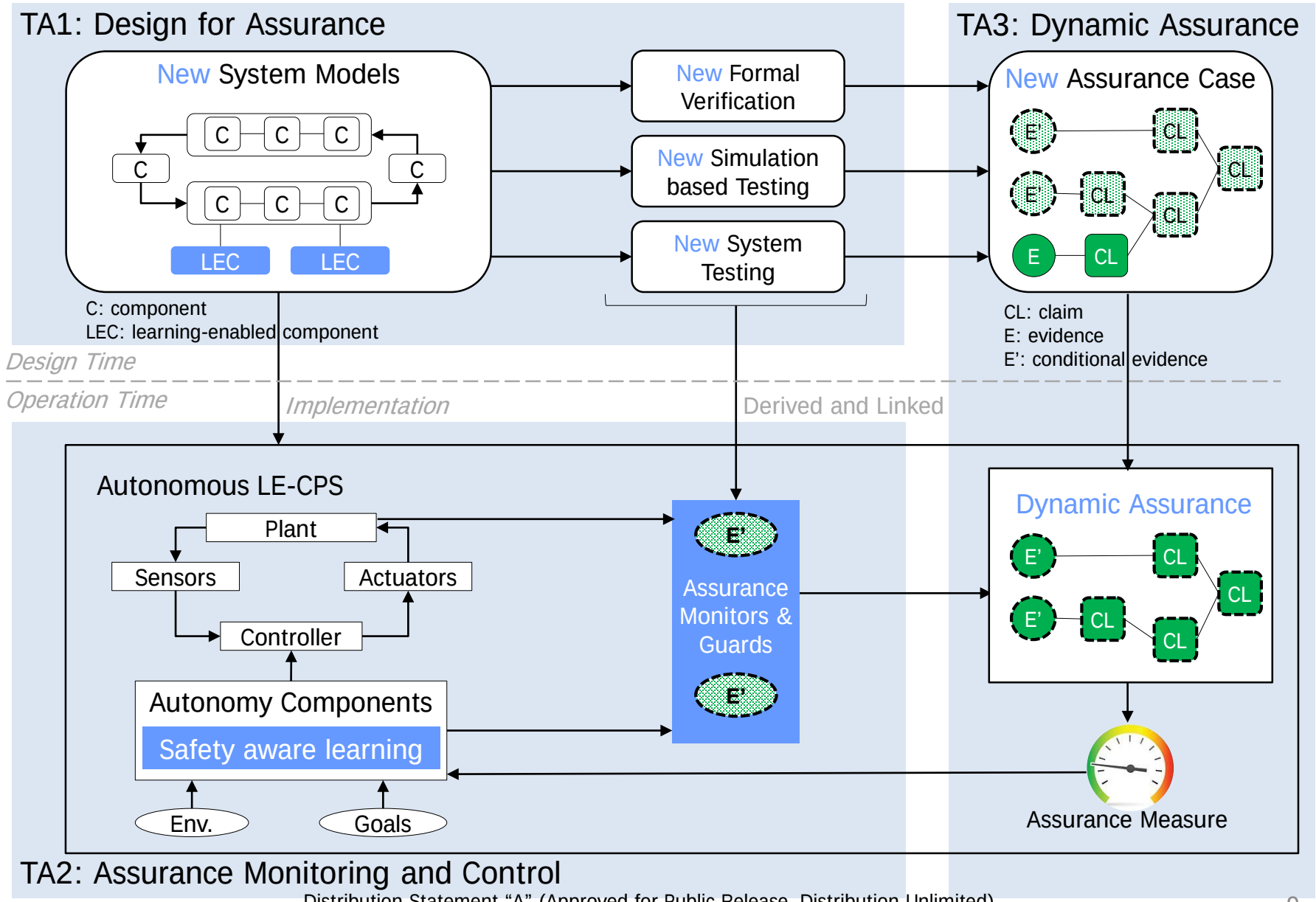


Learning-Enabled System

New System Models → New Formal Verification, New Simulation based Testing, New System Testing →

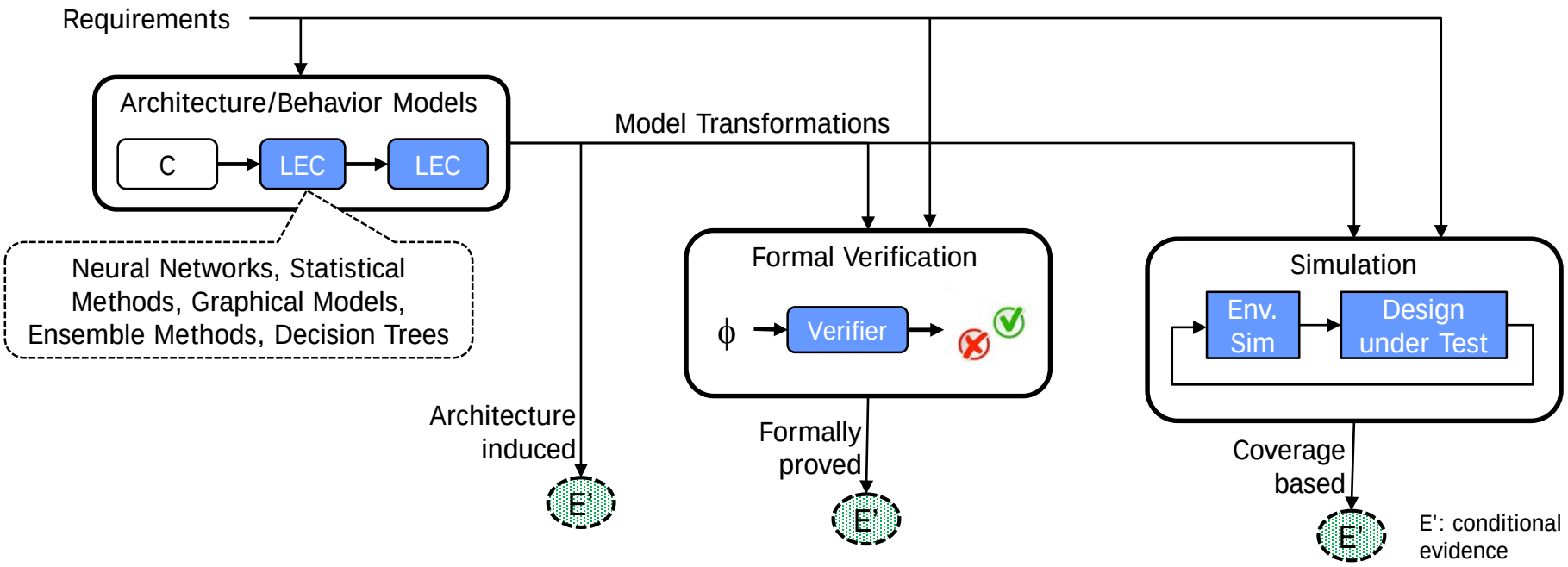


Program Structure



TA1: Design for Assurance

Develop and integrate tools for design and verification of learning-enabled systems, and generate evidence of safety and correctness

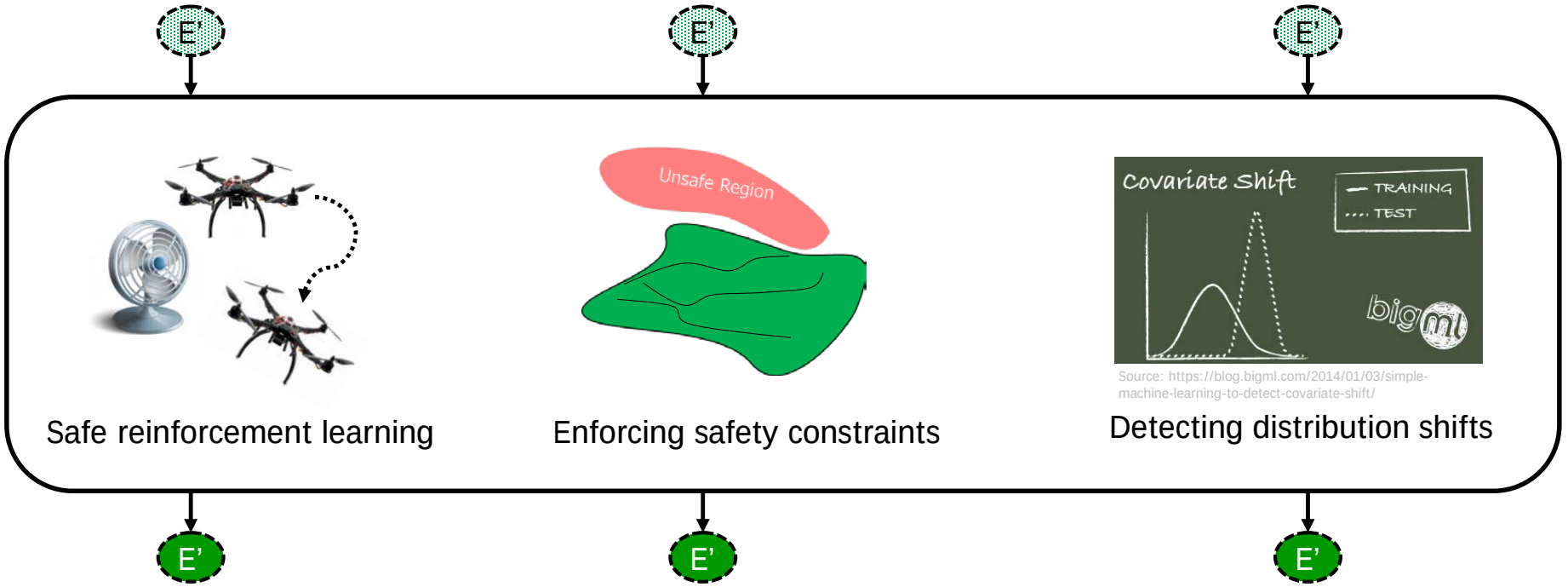


Research challenges include, but are not limited to development in:

- Compositional architectures for learning-enabled systems that guarantee and preserve specified properties
- Formalisms, abstractions, and domain specific modeling languages for representation of learning-enabled components, systems and their dynamics
- Scalable methods addressing formal verification of safety and liveness properties of LE-CPS
- Simulation approaches that drive the learning-enabled system to elicit unanticipated behaviors
- Approaches for maximizing test coverage of LE-CPS
- Transformations for automated synthesis of assurance monitors.

TA2: Assurance Monitoring and Control

Develop algorithms for safety-aware learning and techniques for monitoring and enforcement of safety constraints



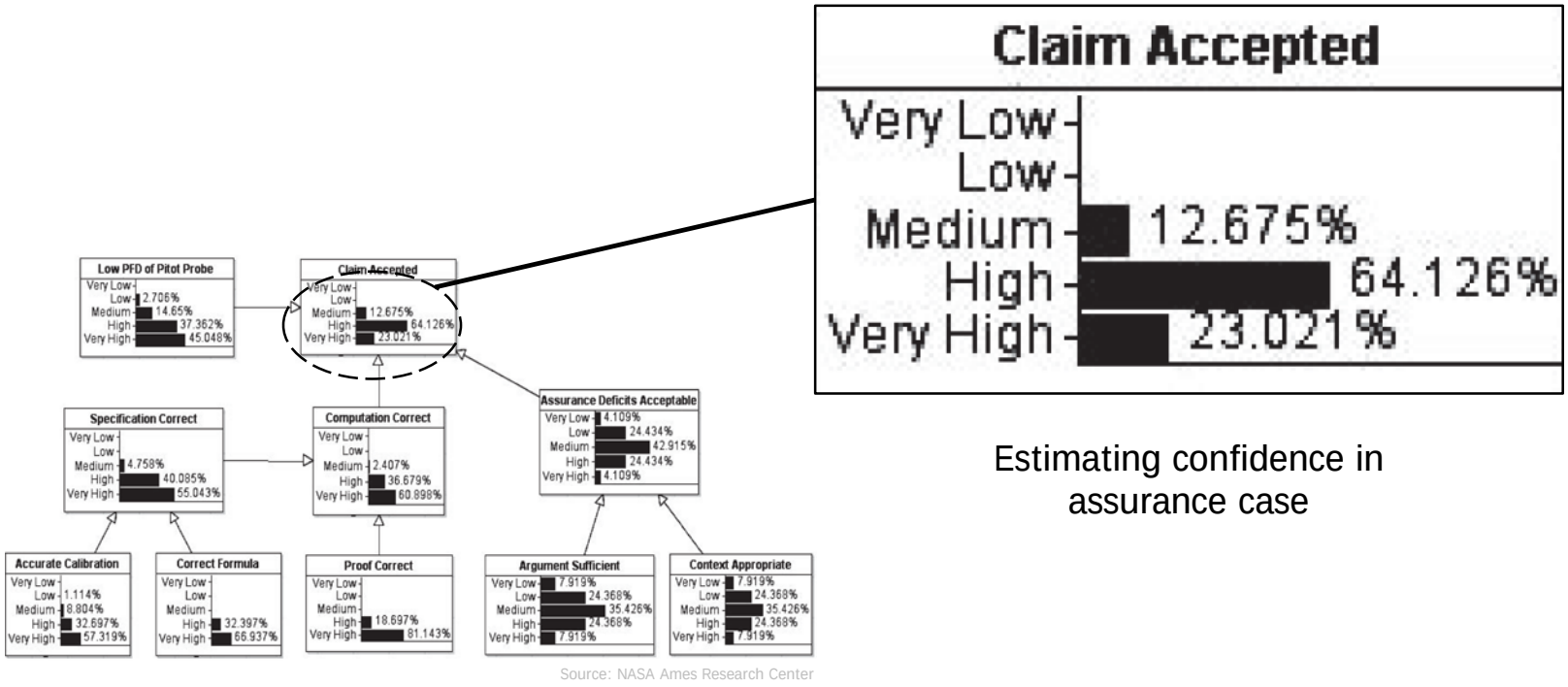
E': conditional evidence

Research challenges include, but are not limited to development in:

- Techniques and algorithms for safety-aware learning
- Monitors for enforcement of hard safety constraints
- Monitors for enforcement of architectural constraints
- Monitors to detect data-distribution shifts, and qualifying the performance of learning algorithms as the operating environment diverges from training environment.

TA3: Dynamic Assurance

Develop tools and algorithms for formal representation and online evaluation of assurance cases



Source: NASA Ames Research Center

Estimating confidence in assurance case

Research challenges include, but are not limited to development in:

- Formal semantics of assurance cases that enable assessment in terms of validity and confidence
- Scalable algorithms for dynamic evaluation of assurance cases consistent with the formal semantics
- Capabilities for modularizing and automatically generating assurance cases from system design descriptions
- Application of developed capabilities to produce dynamic assurance cases for LE-CPSs.

TA4: Integration and Experimentation Platforms

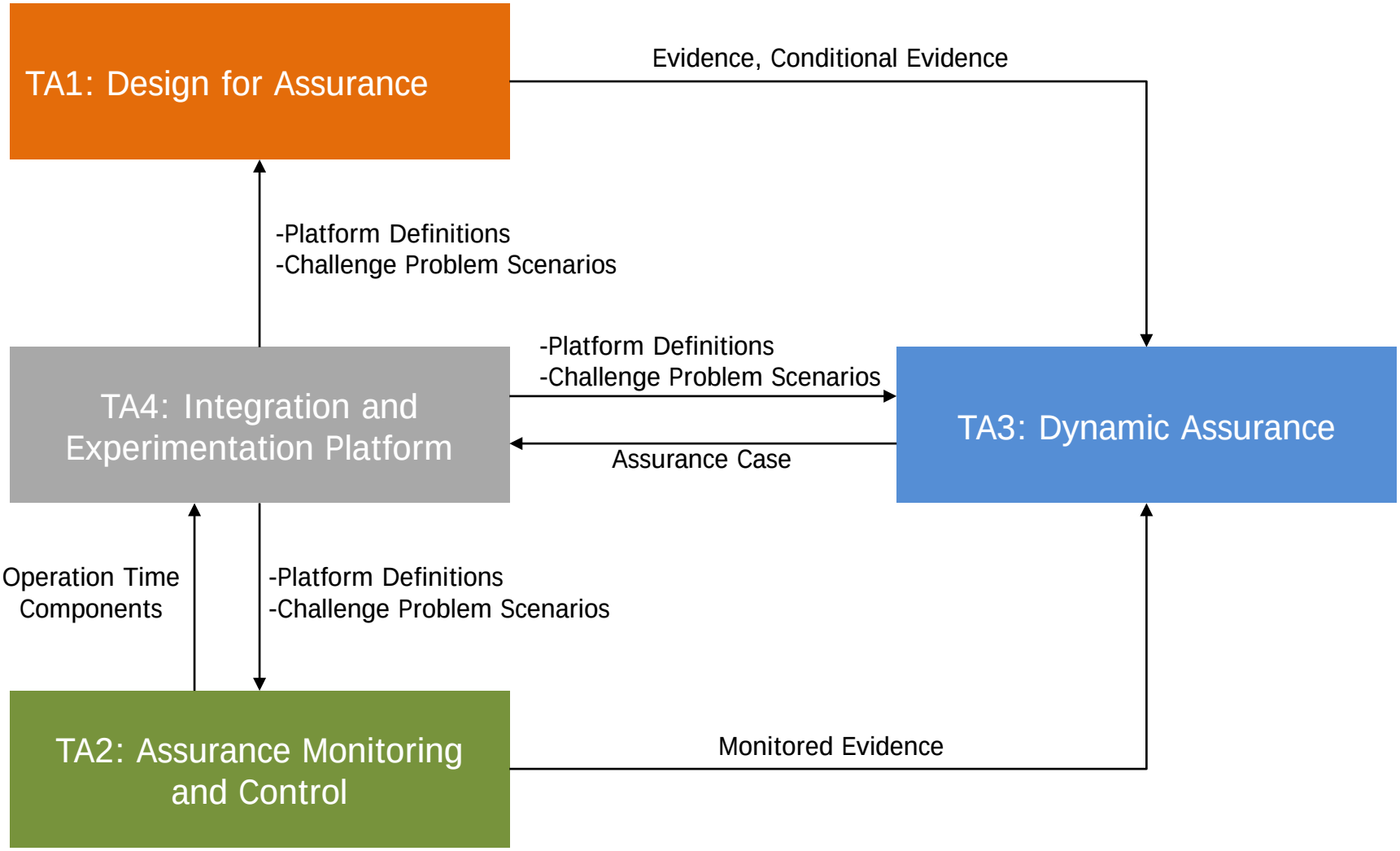
Candidate autonomy platforms to serve as testbeds for experimentation, evaluation, and demonstration

- Platform characteristics
  - DoD relevant platform to demonstrate assured autonomy technologies
  - Autonomy capable or extensible to include autonomy components
  - Accessible to performers, potentially through high-fidelity simulators or surrogate platforms

Notional Platform - Autonomous Surface Ship

	Phase 1	Phase 2	Phase 3
Experimentation Vignette (TA4)	Collision-free navigation, Autonomous trailing		
Learning algorithms (to be integrated) (TA1, TA2, TA4)	Classifier learning • Visually recognize objects • Infer capabilities and COLREGS behavior	Reinforcement/imitation learning • Understand evader tactics • Determine pursuer course of action	System dynamics learning • Learn/predict dynamics of evader • Learn dynamics of own ship
Assurance (TA1, TA2, TA3)	Develop assurance case for COLREGS	Develop assurance case for COLREGS under RL guided behavior	Develop assurance case for trailing under RL guided behavior
Demonstration (TA4)	Simulation, SW-in-the-loop	Simulation, SW-in-the-loop	Physical (surrogate)

Technical Area Interactions



Program objectives

- Increase scalability of design-time assurance
  - What is the baseline capability of the proposed methods, in terms of the hybrid state-space and number and complexity of learning-enabled components
  - How do you plan to scale up by an order of magnitude?
  - How will you characterize the tradeoffs between fidelity of your modeling abstractions and scalability of the verification approach.

- Reduce overhead of operation-time assurance
  - What is the baseline overhead of the operation-time assurance monitoring techniques?
  - How do you plan to minimize it to be below 10% of the nominal system resource utilization?

- Scale up dynamic assurance
  - What is the size and scale of dynamic assurance case that can be developed and dynamically evaluated with your tools?

- Reduce trials to assurance
  - How will your approach quantifiably reduce the need for statistical testing?

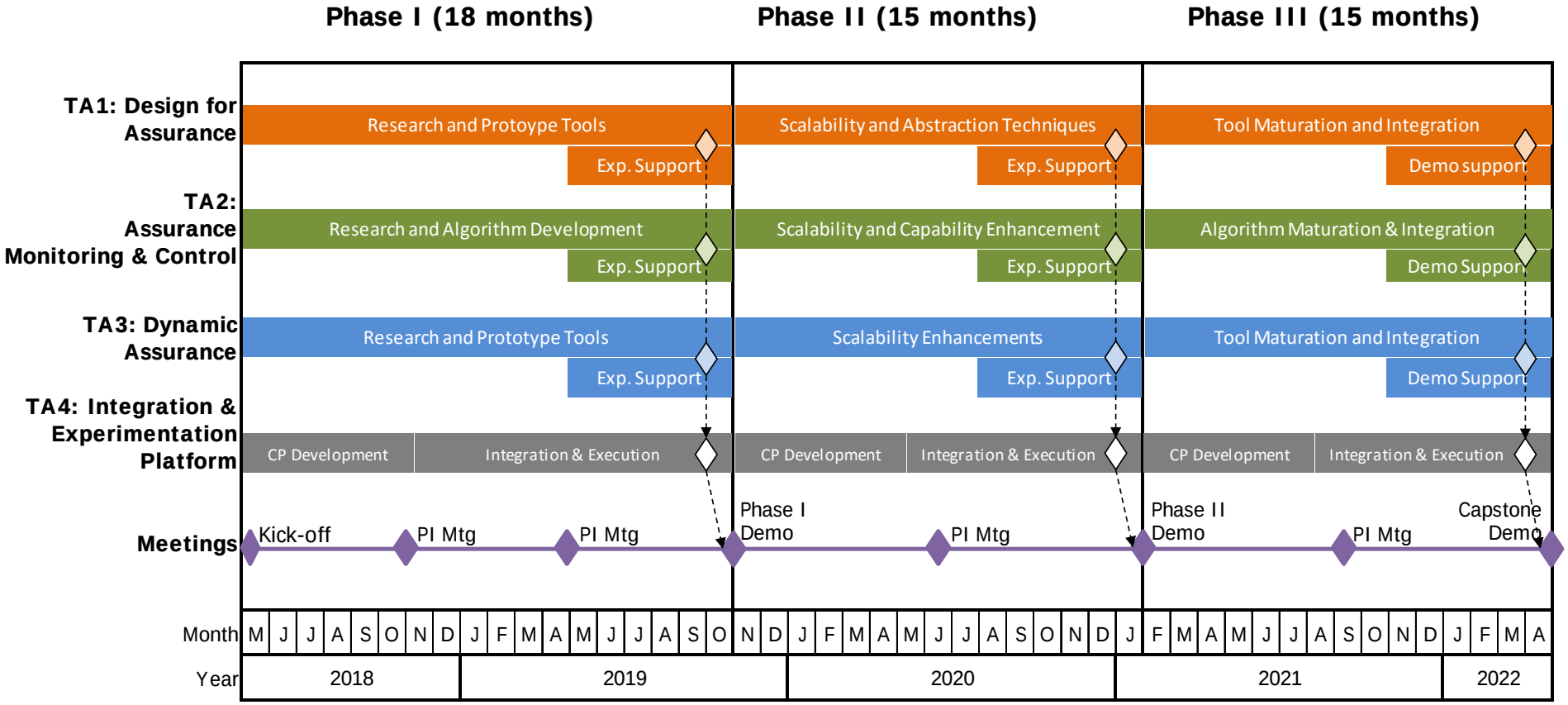
Program objectives (illustration purposes)

DARPA seeks an order of magnitude increase in the scale and complexity of the challenge problems across the Phases, and correspondingly the assurance technologies developed in the program must scale up to address these challenges

TA	Metric	Current	Phase 1	Phase 2	Phase 3
1	Scalability	10+ dimensions N/A for LEC	20 dimensions 1-2 LEC (low)	40 dimensions 2-4 LEC (mid)	100 dimensions 4-6 LEC (high)
2	Monitoring overhead	N/A	50%	30%	10%
3	Dynamic Assurance	N/A	10 conditional evidence	100 conditional evidence	1000 conditional evidence
4	Reduced trials to Assurance	10K-100K physical trials	0.1x	0.01x	0.001x

Proposers should develop and describe quantitative metrics specific to their approach aligning with the objectives listed above.

Schedule and Milestones



Overall Programmatics

- Three Phases – Phase I (18 months); Phases II and III (15 months each)
- Four Technical Areas (TAs): Design for Assurance (TA1); Assurance Monitoring and Control (TA2); Dynamic Assurance (TA3); Integration and Experimentation Platform (TA4)
- Anticipate Multiple Awards in TAs 1-4
  - Anticipate two TA4 platforms to demonstrate domain agnosticism
  - Proposers selected for TA4 will not be selected for award as a performer (prime or subcontractors) on TA1-3
- TA4 Proposal
  - Plan to support 2 TA1-3 teams in their base effort
  - Include options for supporting additional TA1-3 teams
- TA1-3 Proposal
  - Plan to apply technology to both TA4 platforms
  - Include support for the 2nd platform as an option
- DARPA encourages, but doesn't require, integrated TA1-3 proposals
- Proposers not submitting an integrated TA1-3 proposal should anticipate joining an integrated TA1-3 team post program kickoff, and describe their interface and plans for working with other TAs
- Strong interaction among all performers is critical to program success
  - Associate Contractor Agreement (ACA)

TA1 Programmatics

- Proposals must describe at least one example of a challenge problem (as a proxy for challenge problems that will be proposed by TA4), for which the proposed techniques can be applied effectively, clearly describing the metrics and capability milestones reached at the end of each Phase.
- The envisioned outcome of TA1 effort are tools that will need to be integrated into LE-CPS design flows
- Proposals need to explain interfaces provided for integration.

- Work with **TA4** to learn about the target platforms, apply techniques to the provided CPs, and consult on the application of the techniques to the target platform.
- **TA1** performers must demonstrate their tools on appropriate portions of the challenge problems, and provide evidence that **TA3** can use to construct a dynamic assurance case.

TA2 Programmatics

- Proposals must describe at least one example of a challenge problem (a proxy for challenge problems that will be proposed by TA4) for which the proposed algorithms can be applied effectively, clearly describing the metrics and capability milestones reached at the end of each Phase.
- The envisioned outcome of the TA2 effort are operation-time components that need to be integrated into the TA4 platforms.
- Proposals need to explain interfaces provided for integration.

- Work with **TA4** to learn about the target platforms, apply techniques to the provided CPs, and consult on the application of the techniques to the target platform.
- **TA2** will also require close coordination with **TA1**, as the conditions to be monitored and the assumptions to be validated may be provided by the **TA1** performers.

TA3 Programmatics

- Proposals must describe at least one example of a challenge problem (as a proxy for challenge problems that will be proposed by TA4), for which the proposed techniques can be applied effectively, clearly describing the metrics and capability milestones reached at the end of each Phase.
- The envisioned outcome of the TA3 effort is a dynamic assurance case and the tools to build and evaluate the assurance case that needs to be integrated into the TA4 platforms.
- Proposals need to explain interfaces provided for integration.

- **TA3** performers are expected to work closely with the **TA4** performers by learning about the target platforms, formulating a dynamic assurance case, and guiding the **TA1** and **TA2** performers.
- The **TA3** performers will lead the effort to integrate the evidence developed by researchers in TA1 into a coherent assurance case.
- The **TA3** performers will work with the **TA4** performers to deploy the dynamic assurance case in the platform.

TA4 Programmatics

- Demonstrate the tools applied to the platform-specific challenge problems at the demonstration meeting.
- Evolve the experimentation platform, extend with LECs, develop challenge problems and integrated evaluation scenarios, provide a framework and access to platforms for allowing technology developers to integrate LECs, provide models and access to development toolchains, and experimentally validate assurance developed by TA1-3 performers.
- Identified platforms (including sensors, computing hardware, and software) must be unclassified.

- Educate **other performers** about use cases, and related technical challenges in producing assured versions of the platforms
- Develop unrestricted and unclassified challenge problems that abstract the key difficulties for use by **other performers**;
- Provide design models, simulators and/or distributable prototypes and related APIs for **other performers**
- Apply the research results from the **other TAs** to the development of the assured platform.