Fraunhofer USA

University of Maryland 1856

SHOULD YOU RELY ON THAT AI?

Session 2: Moving to a Full-lifetime Testing Approach

Dr. Adam Porter

Panel Members

Dr. Douglas C. Schmidt is the Cornelius Vanderbilt Professor of Computer Science, Associate Provost for Research Development and Technologies, Co-Chair of the Data Science Institute, and a Senior Researcher at the Institute for Software Integrated Systems, all at Vanderbilt University.

Dr. Mikael Lindvall Dr. Mikael Lindvall, is the Technology Director of Fraunhofer USA, Center Mid-Atlantic (CMA) and an adjunct professor at UMCP for more than 10 years. He recently created and delivered one of the first courses on Software Engineering for AI-based systems.

Dr. Jeffrey W. Herrmann is a professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the director of the Reliability Engineering Graduate Program and the director of the Systems Engineering Graduate Program at the University of Maryland.

Dr. Laura Freeman is a Research Associate Professor of Statistics and the Director of the Intelligent Systems Lab at the Virginia Tech Hume Center. Previously, she was the Assistant Director of the Operational Evaluation Division at the Institute for Defense Analyses.

No One Should Rely on Artificial Intelligence

Software safety is hard to achieve, even for systems that don't use AI

AI is used to manage dynamism, complexity, and uncertainty

There's no reason to believe that AI makes the safety problem easier

Rely on the System, not on System Components

AI is a system component

Don't rely on AI; try to build a system you can rely on

Requires that systems be architected to include:

- Effective safeguards and harm mitigation,
- Monitoring and testing throughout the full life cycle, and
- Human accountability

Eternal Vigiliance is the Price of Liberty

AI can learn over time; data sets grow and change rapidly

No point in time where we can safely turn off monitoring and surveillance

Full lifecycle monitoring and testing is a fundamental requirement

This is a big data problem, not a pass/fail testing problem

The System Includes People, Organizations, and Missions

AI systems make decisions that affect people

Poor AI systems risks:

- Malpractice
- Unfairness
- Diminishing agency
- Rejection by humans

Panel Members

Dr. Douglas C. Schmidt is the Cornelius Vanderbilt Professor of Computer Science, Associate Provost for Research Development and Technologies, Co-Chair of the Data Science Institute, and a Senior Researcher at the Institute for Software Integrated Systems, all at Vanderbilt University.

Dr. Mikael Lindvall Dr. Mikael Lindvall, is the Technology Director of Fraunhofer USA, Center Mid-Atlantic (CMA) and an adjunct professor at UMCP for more than 10 years. He recently created and delivered one of the first courses on Software Engineering for AI-based systems.

Dr. Jeffrey W. Herrmann is a professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the director of the Reliability Engineering Graduate Program and the director of the Systems Engineering Graduate Program at the University of Maryland.

Dr. Laura Freeman is a Research Associate Professor of Statistics and the Director of the Intelligent Systems Lab at the Virginia Tech Hume Center. Previously, she was the Assistant Director of the Operational Evaluation Division at the Institute for Defense Analyses.

Challenges of Certifying Adaptive Dynamic Computing Environments

Douglas C. Schmidt
d.schmidt@vanderbilt.edu
www.dre.vanderbilt.edu/~schmidt

Institute for Software
Integrated Systems
Vanderbilt University
Nashville, Tennessee, USA

Motivating Example: DDG 1000

Software Engineering Institute
Carnegie Mellon

Motivating Example: DDG 1000

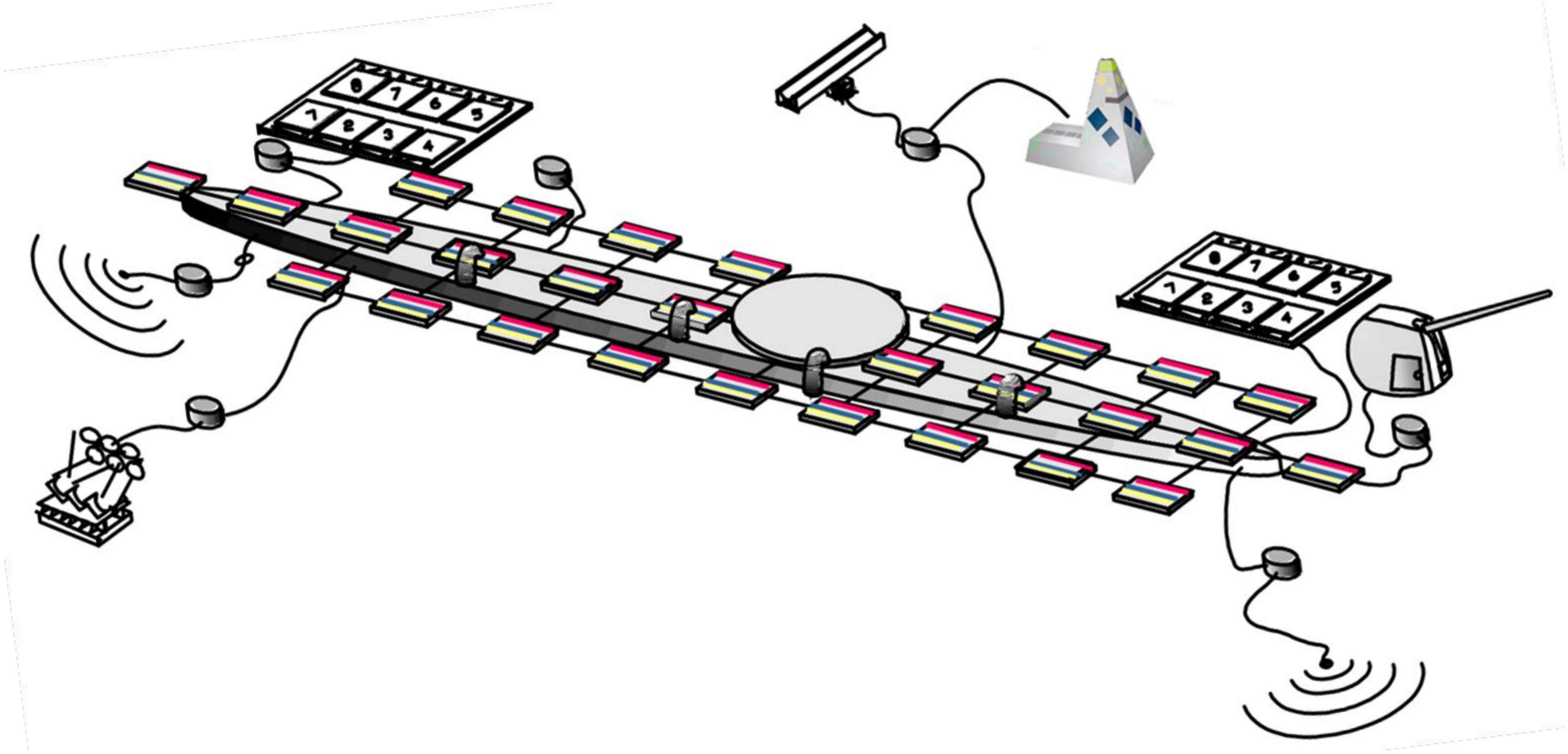
"The right answer delivered too late becomes the wrong answer"

Motivating Example: DDG 1000

"Captain James Kirk Takes Command of the Navy's New $4 Billion Destroyer"

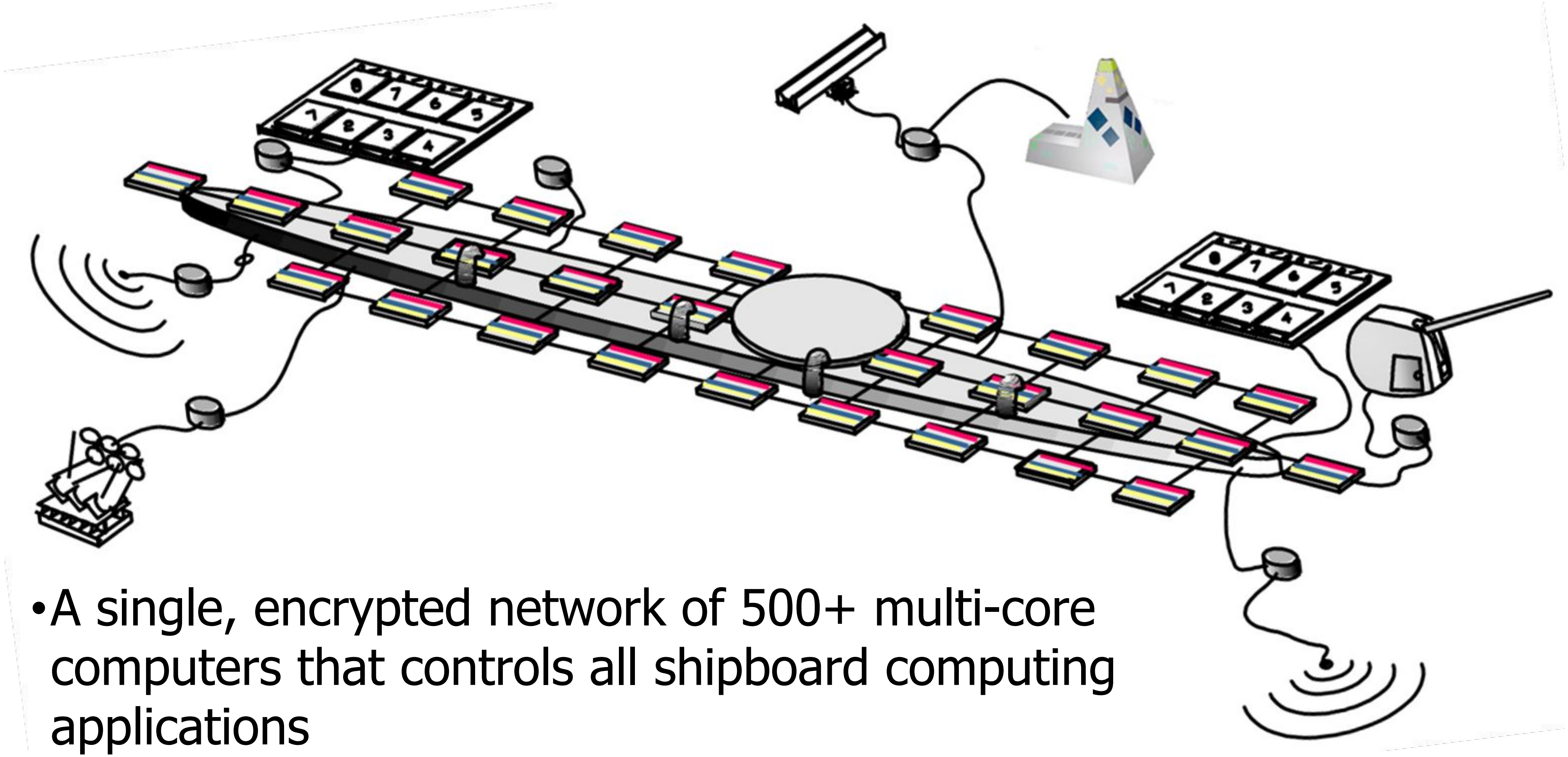
www.thefiscaltimes.com/2016/05/20/Real-Captain-Kirk-Takes-Command-Navy-s-New-4

DDG 1000 has a “Total Ship Computing Environment”



www.raytheon.com/capabilities/products/zumwal

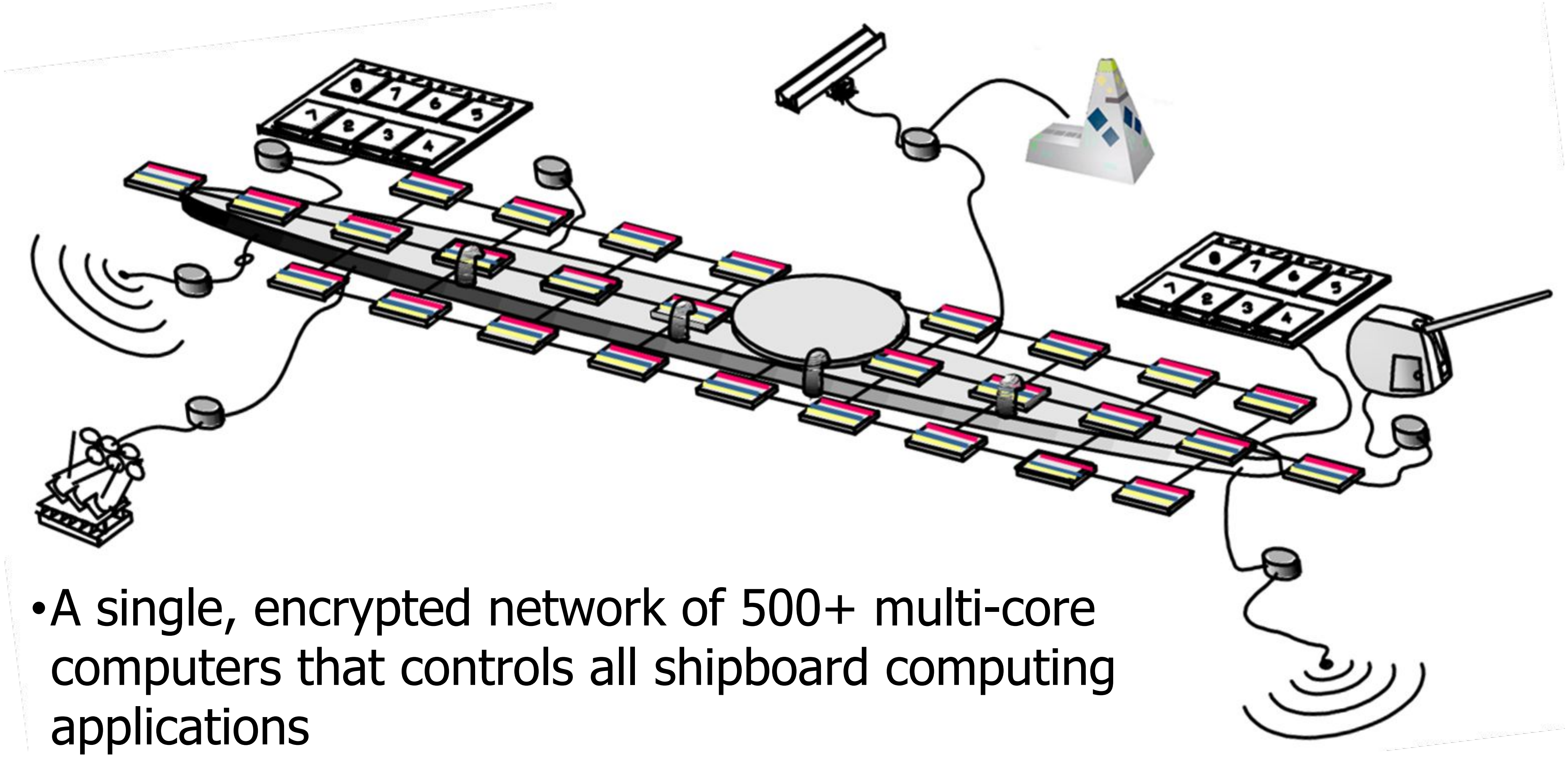
DDG 1000 has a "Total Ship Computing Environment"



- A single, encrypted network of 500+ multi-core computers that controls all shipboard computing applications

www.raytheon.com/capabilities/products/zumwal

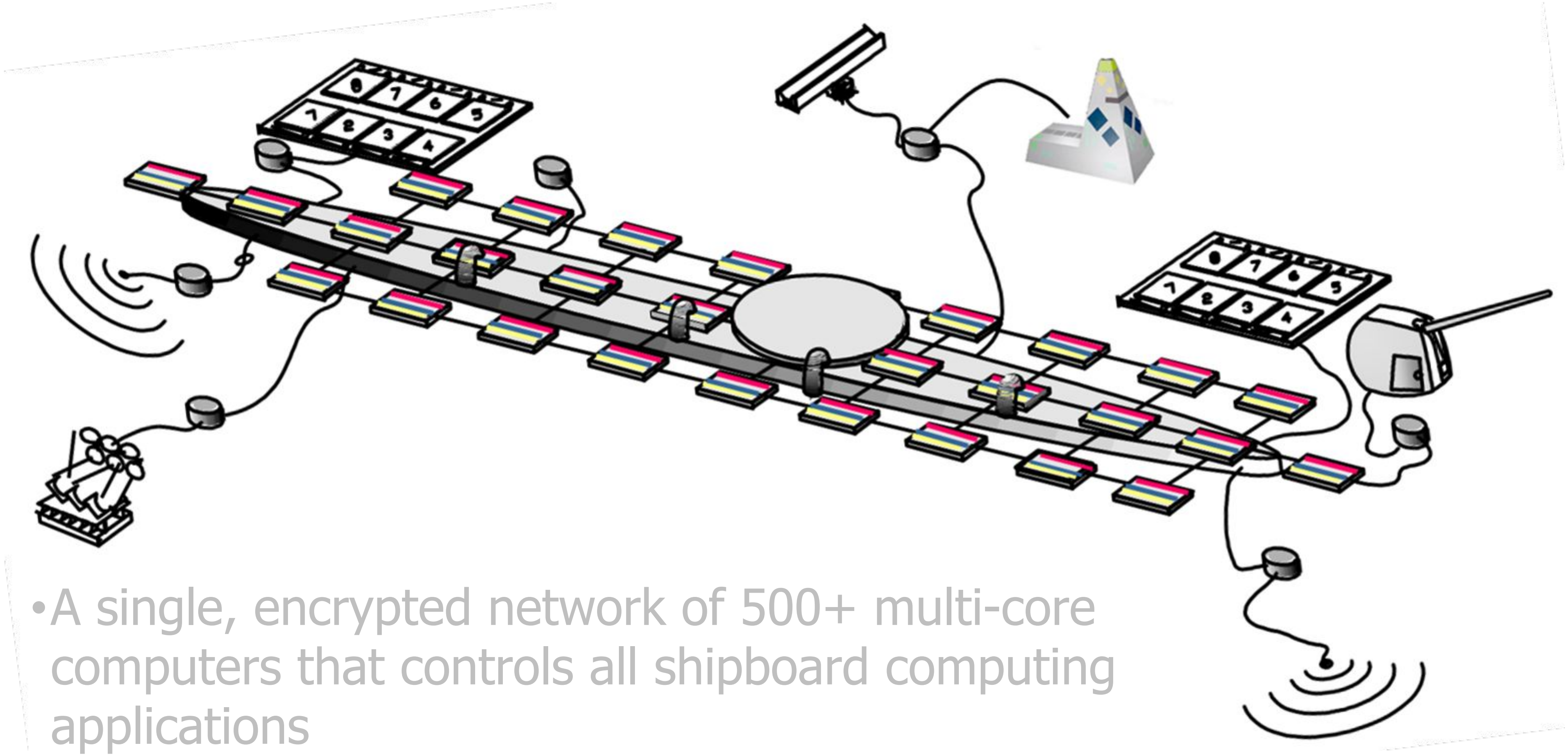
DDG 1000 has a "Total Ship Computing Environment"



- A single, encrypted network of 500+ multi-core computers that controls all shipboard computing applications
  - e.g., ranging from the ship's lights & machinery control to its radars, weapon systems, & crew entertainment

www.raytheon.com/capabilities/products/zumwal

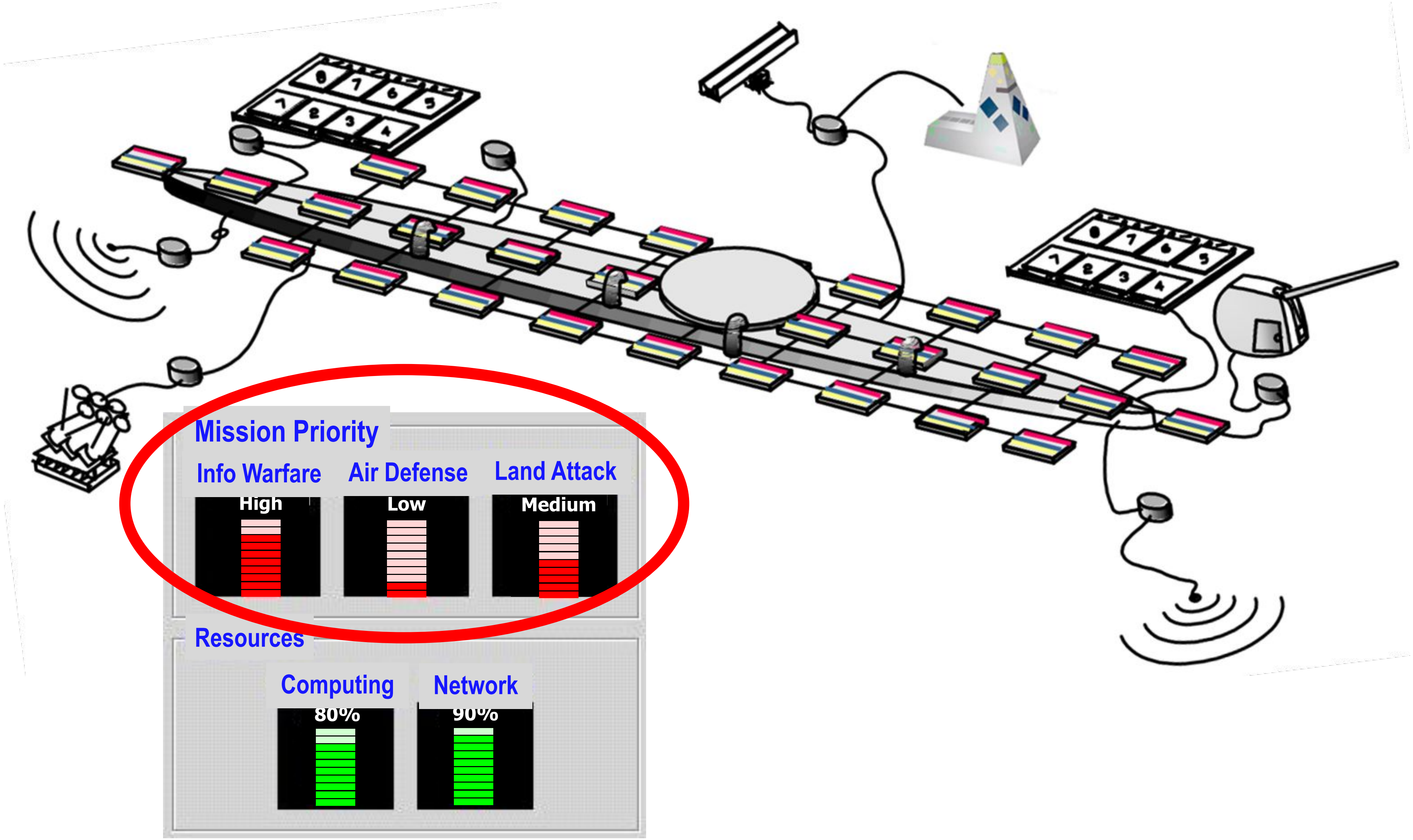
DDG 1000 has a "Total Ship Computing Environment"



- A single, encrypted network of 500+ multi-core computers that controls all shipboard computing applications
- The TSCE's high degree of automation enables the ship to run more effectively & efficiently & dramatically reduces manning requirements

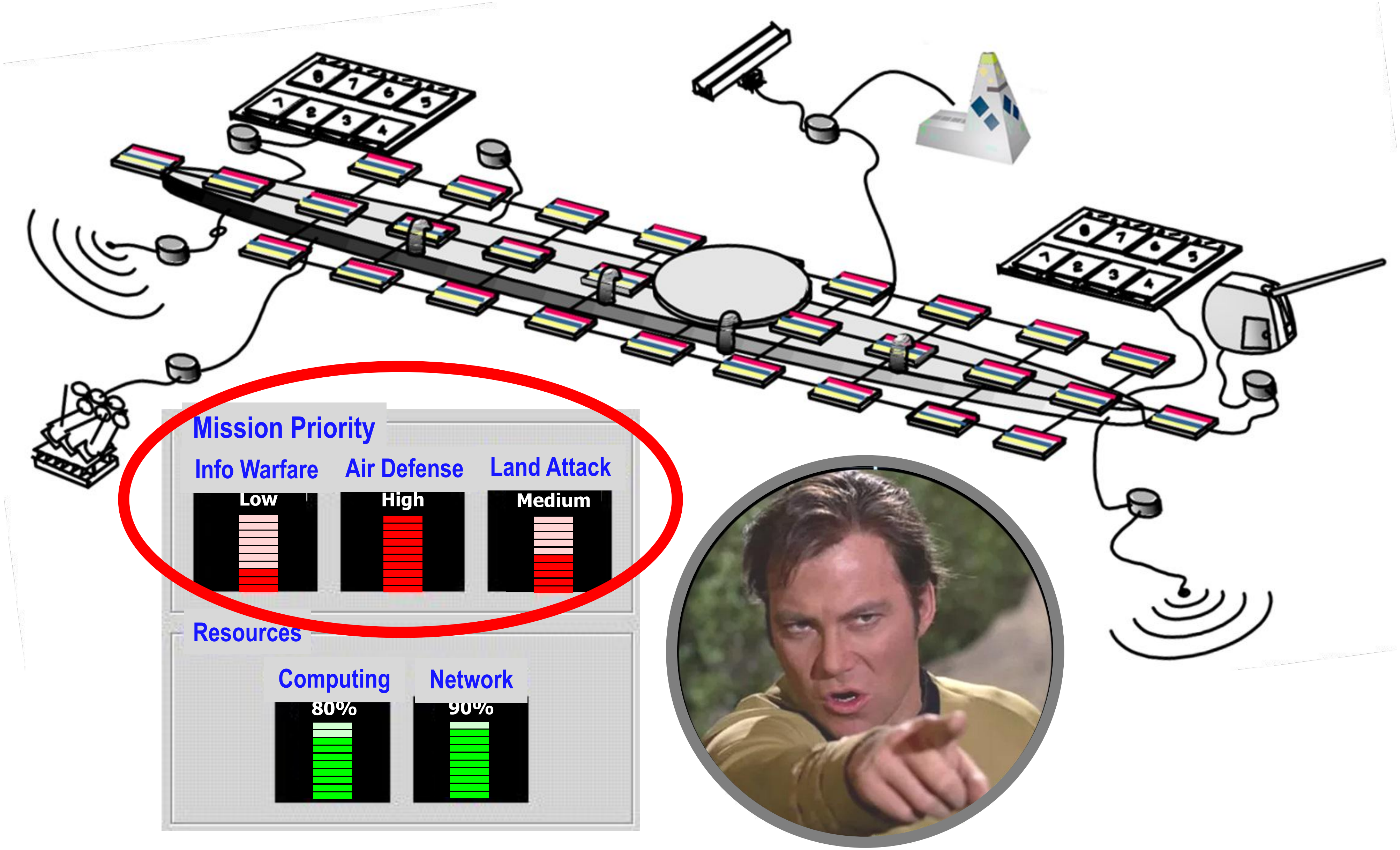
www.raytheon.com/capabilities/products/zumwal

The DDG 1000 TSCE Enables Dynamic Resource Management



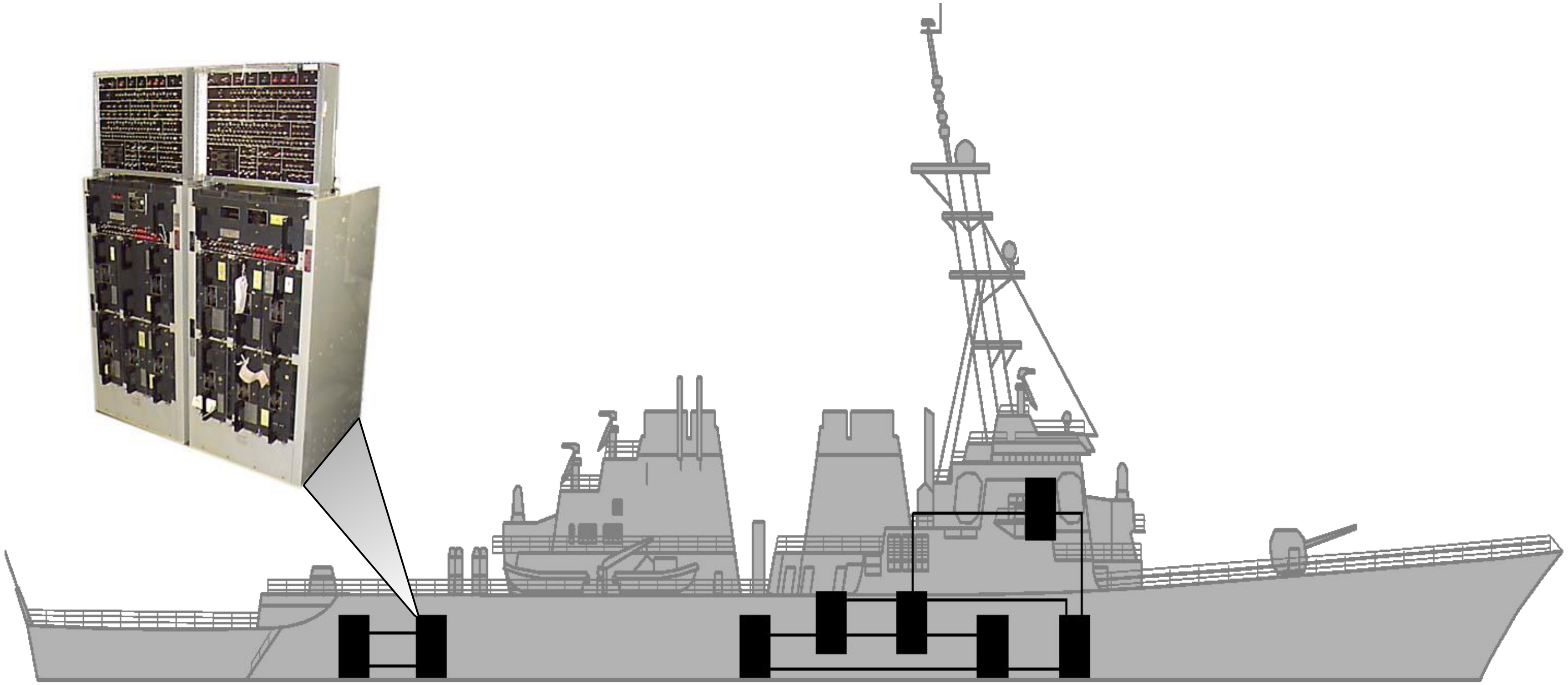
Resource allocation can be optimized dynamically across the TSCE

The DDG 1000 TSCE Enables Dynamic Resource Management



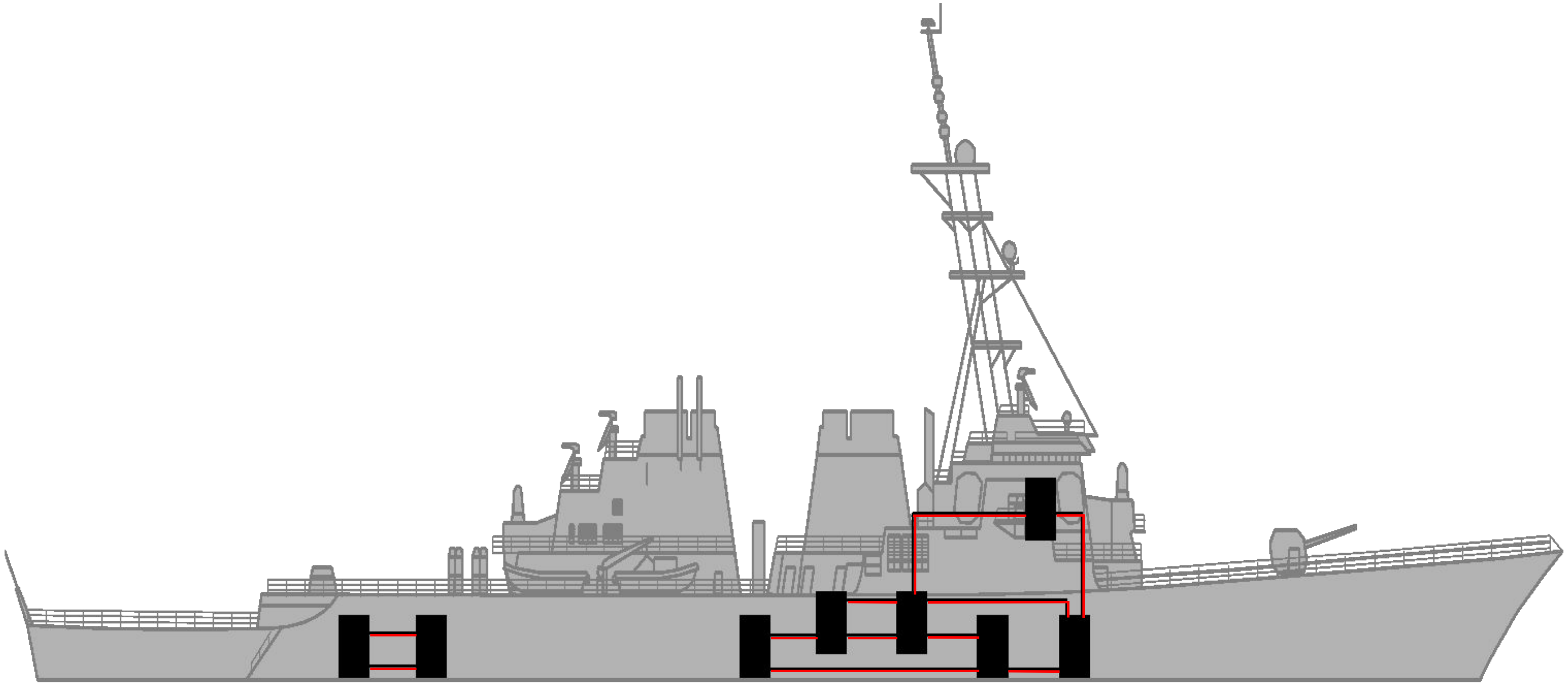
Resource allocation can be optimized dynamically across the TSCE

TSCE is a Major Improvement on Legacy Combat Systems



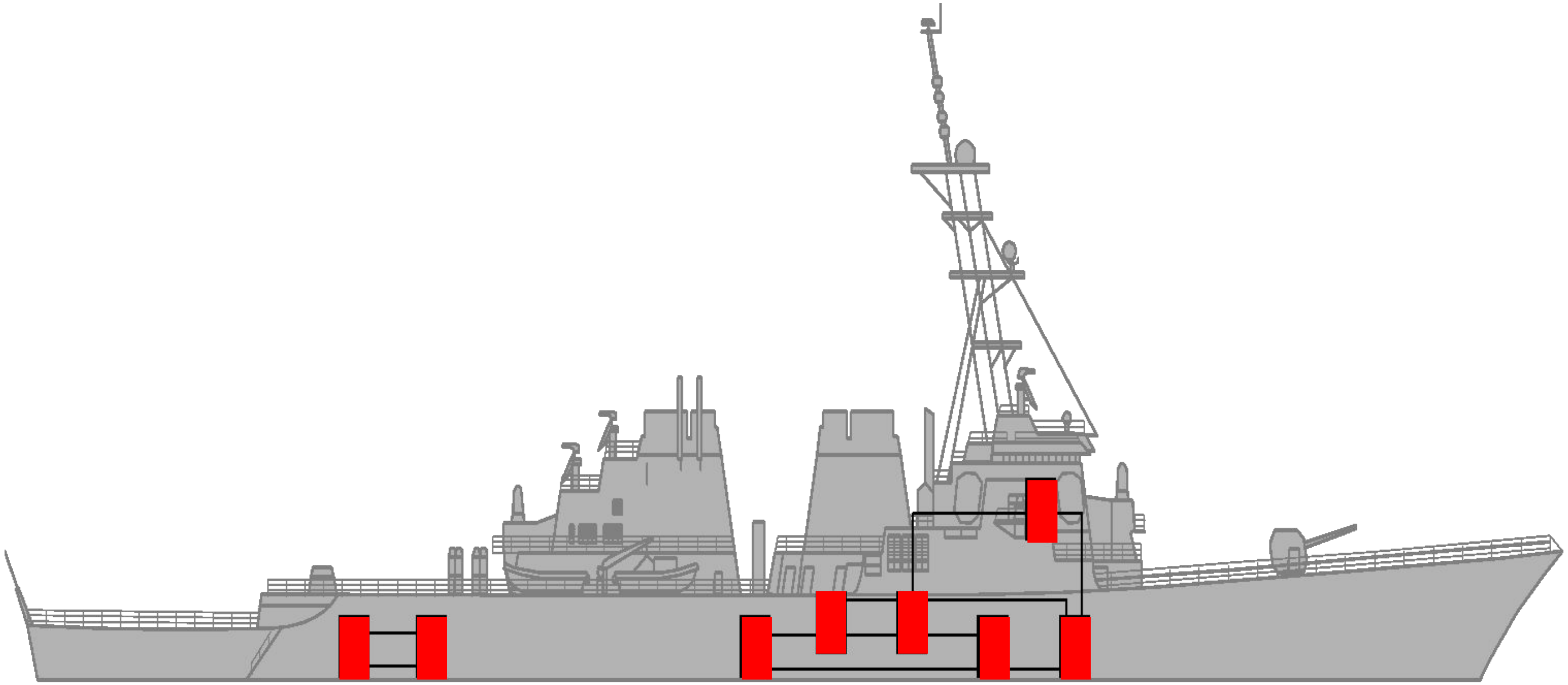
- Proprietary UYK-43s computers

TSCE is a Major Improvement on Legacy Combat Systems



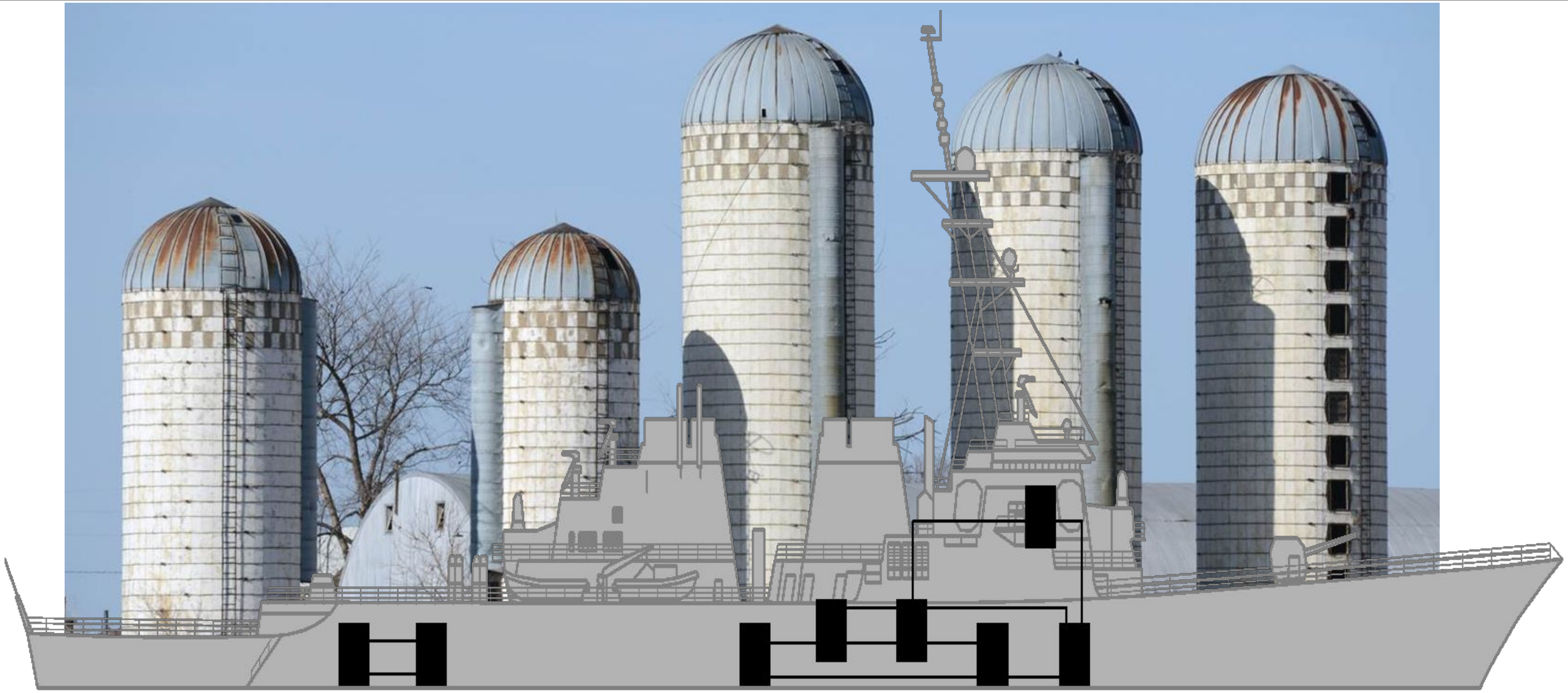
- Proprietary UYK-43s computers
- Point-to-point interconnects

TSCE is a Major Improvement on Legacy Combat Systems



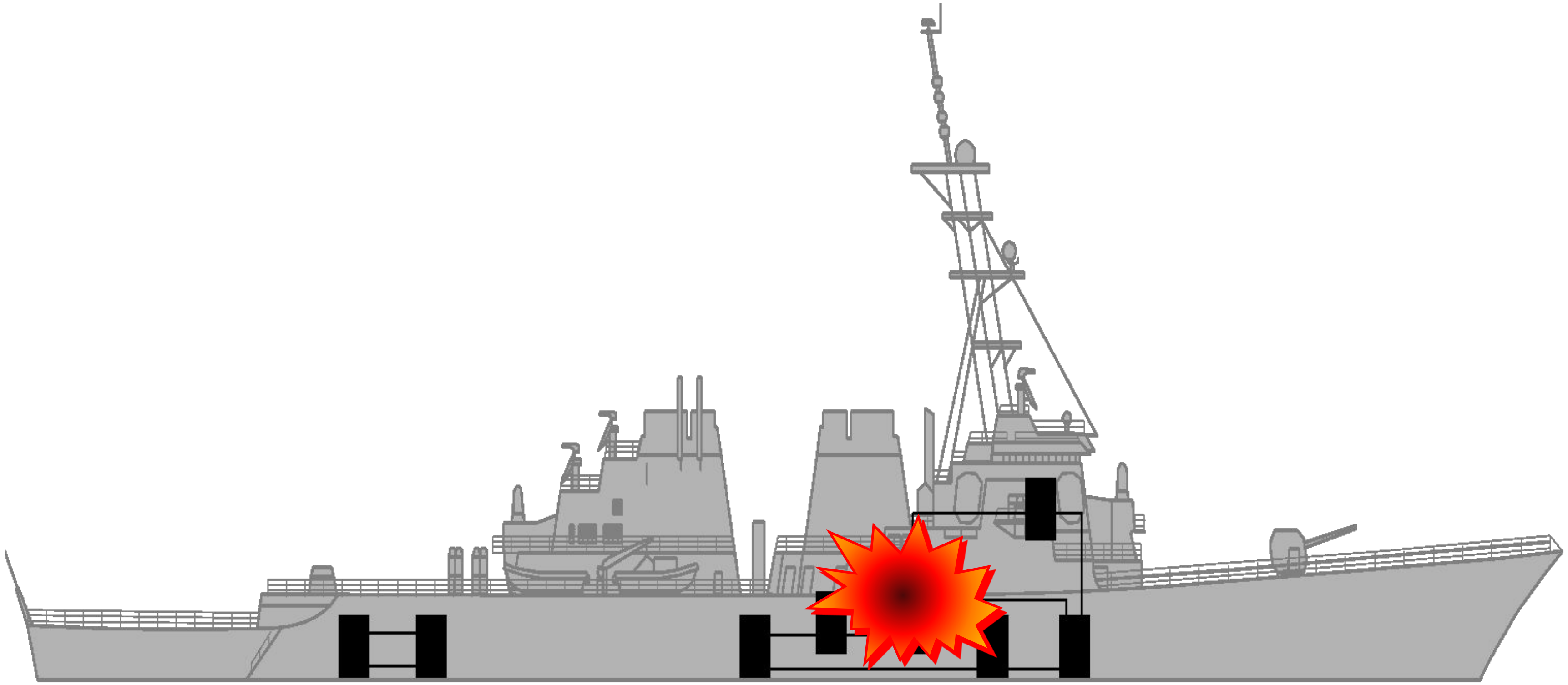
- Proprietary UYK-43s computers
- Point-to-point interconnects
- Limited growth capability

TSCE is a Major Improvement on Legacy Combat Systems



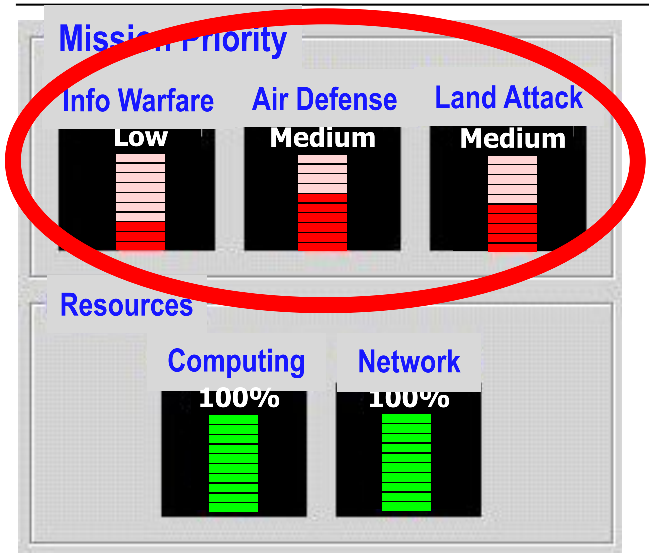
- Proprietary UYK-43s computers
- Point-to-point interconnects
- Limited growth capability
- Tightly constrained by stove-pipe subsystems

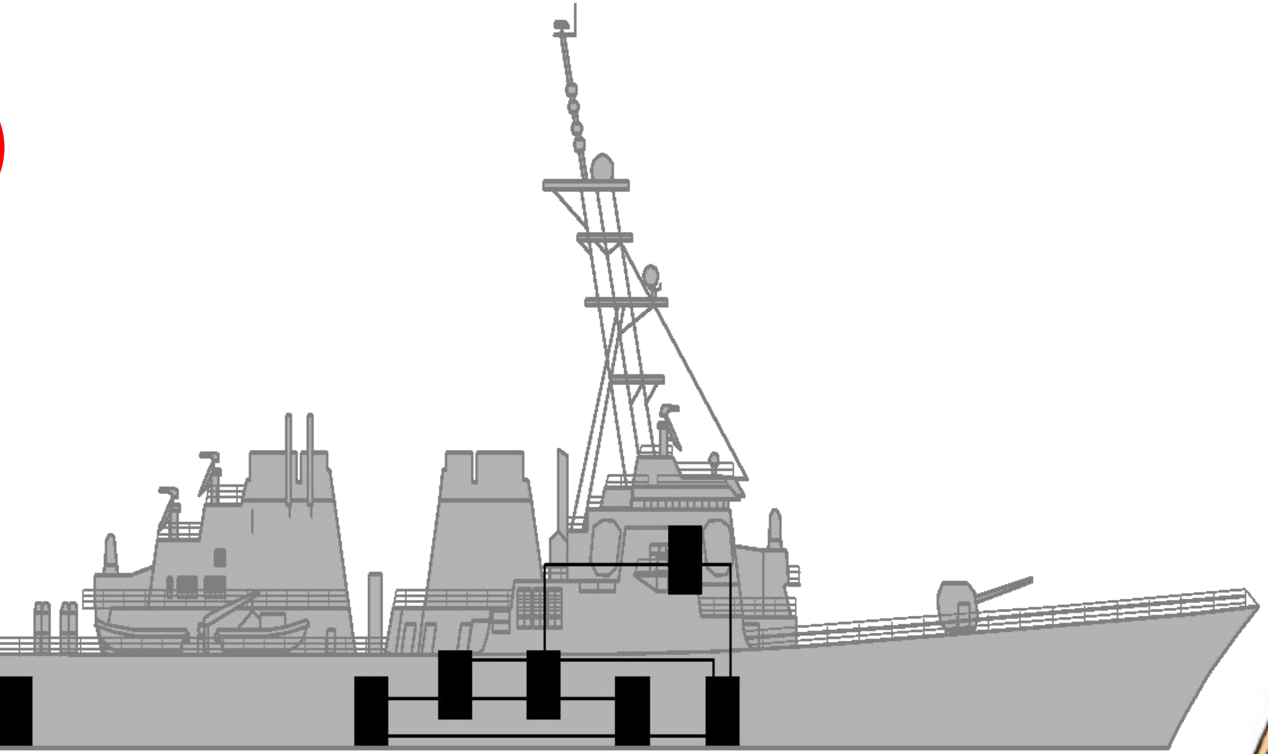
TSCE is a Major Improvement on Legacy Combat Systems



- Proprietary UYK-43s computers
- Point-to-point interconnects
- Limited growth capability
- Tightly constrained by stove-pipe subsystems
- Highly vulnerable to damage

TSCE is a Major Improvement on Legacy Combat Systems

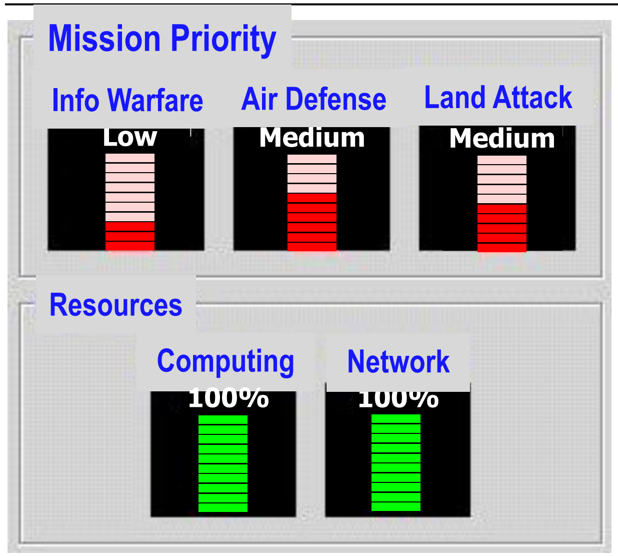


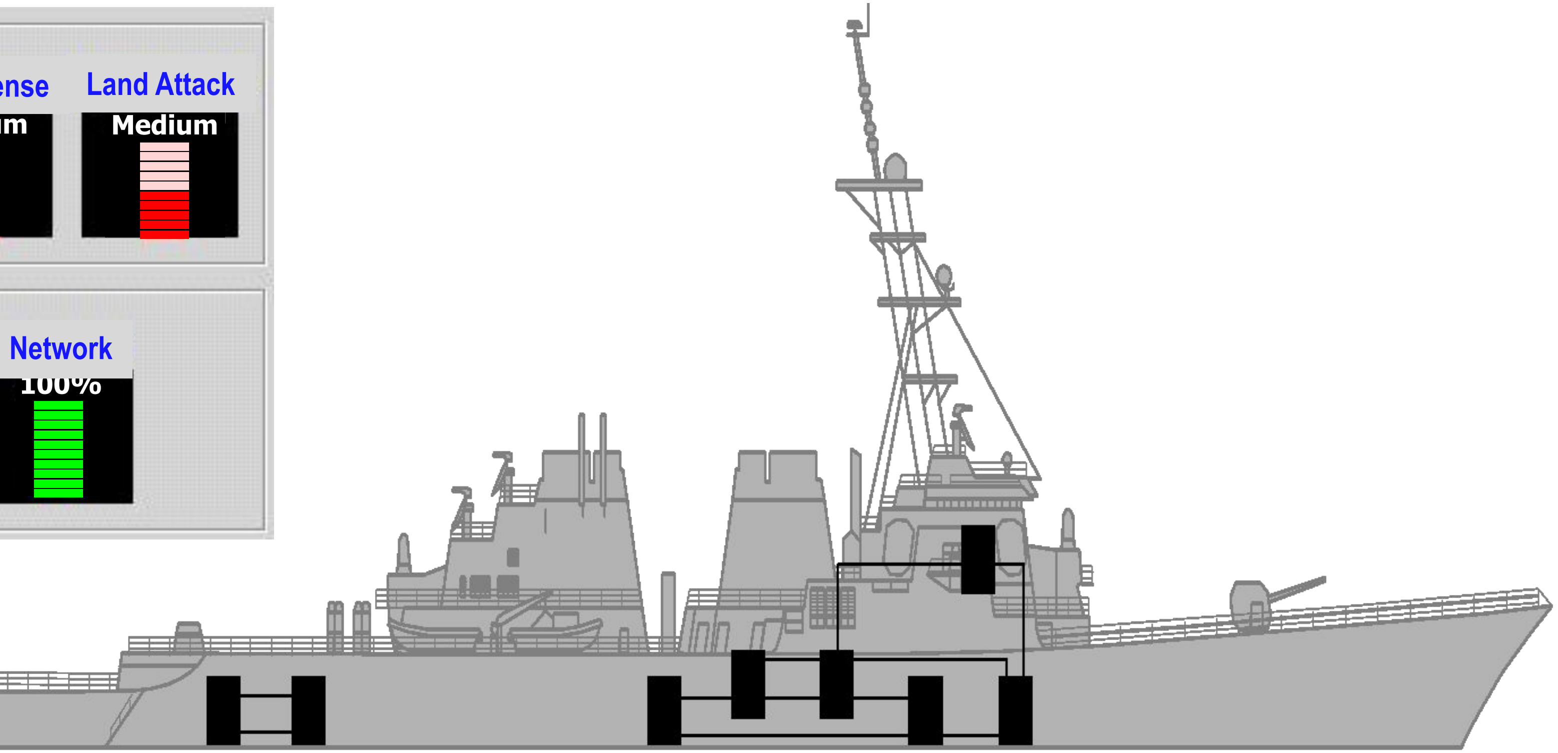


- Proprietary UYK-43s computers
- Point-to-point interconnects
- Limited growth capability
- Tightly constrained by stove-pipe subsystems
- Highly vulnerable to damage

Resources allocated statically at initial system configuration!

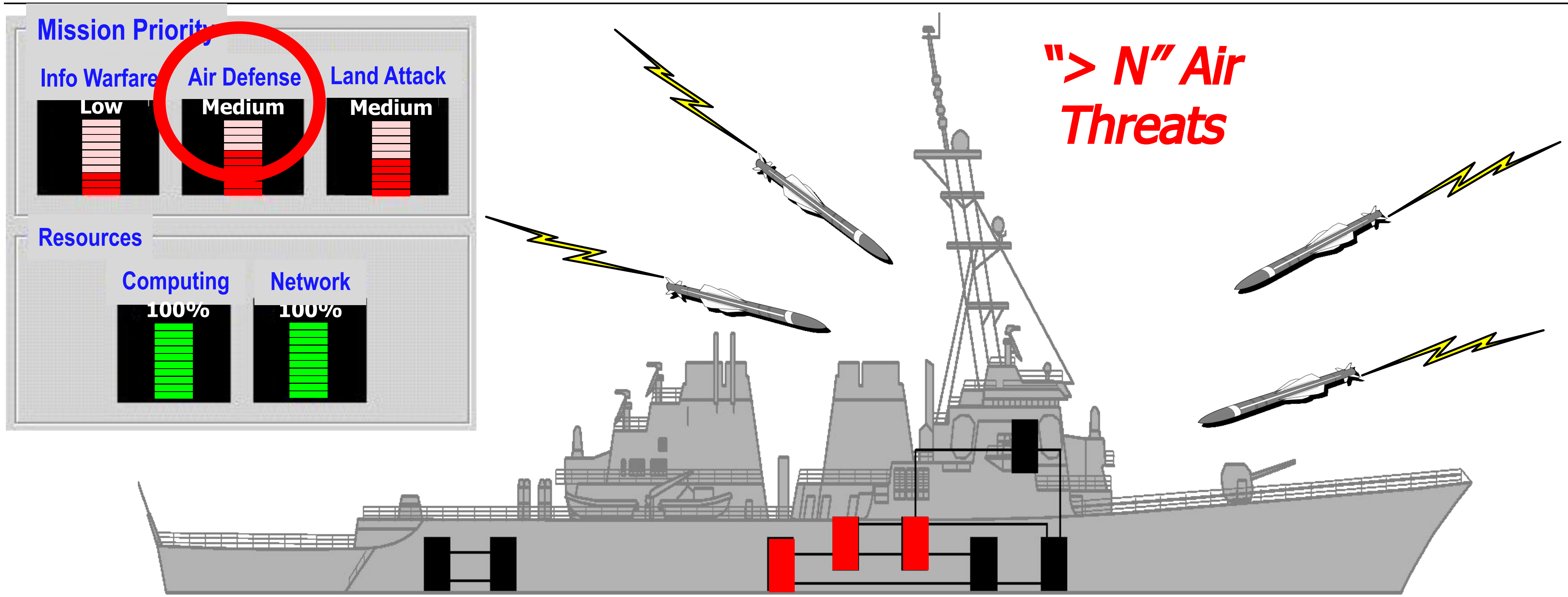
TSCE is a Major Improvement on Legacy Combat Systems





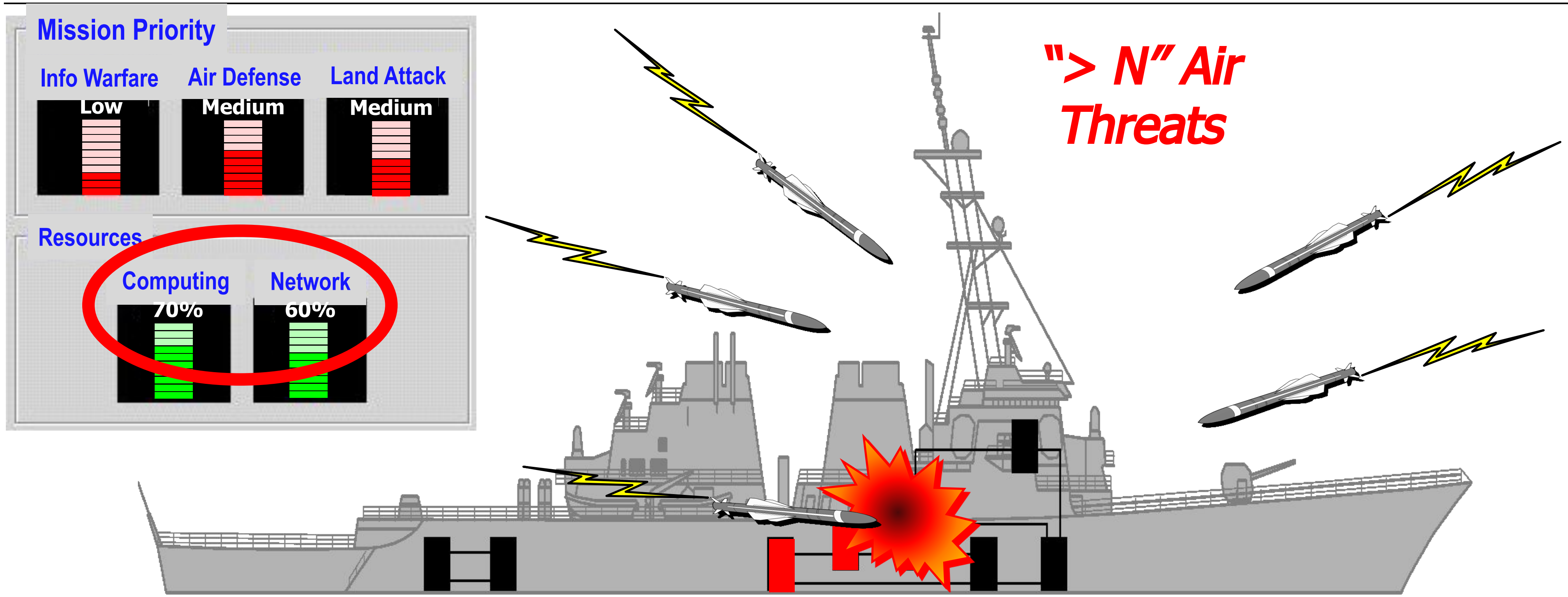
- Static allocation is problematic in several scenarios

TSCE is a Major Improvement on Legacy Combat Systems



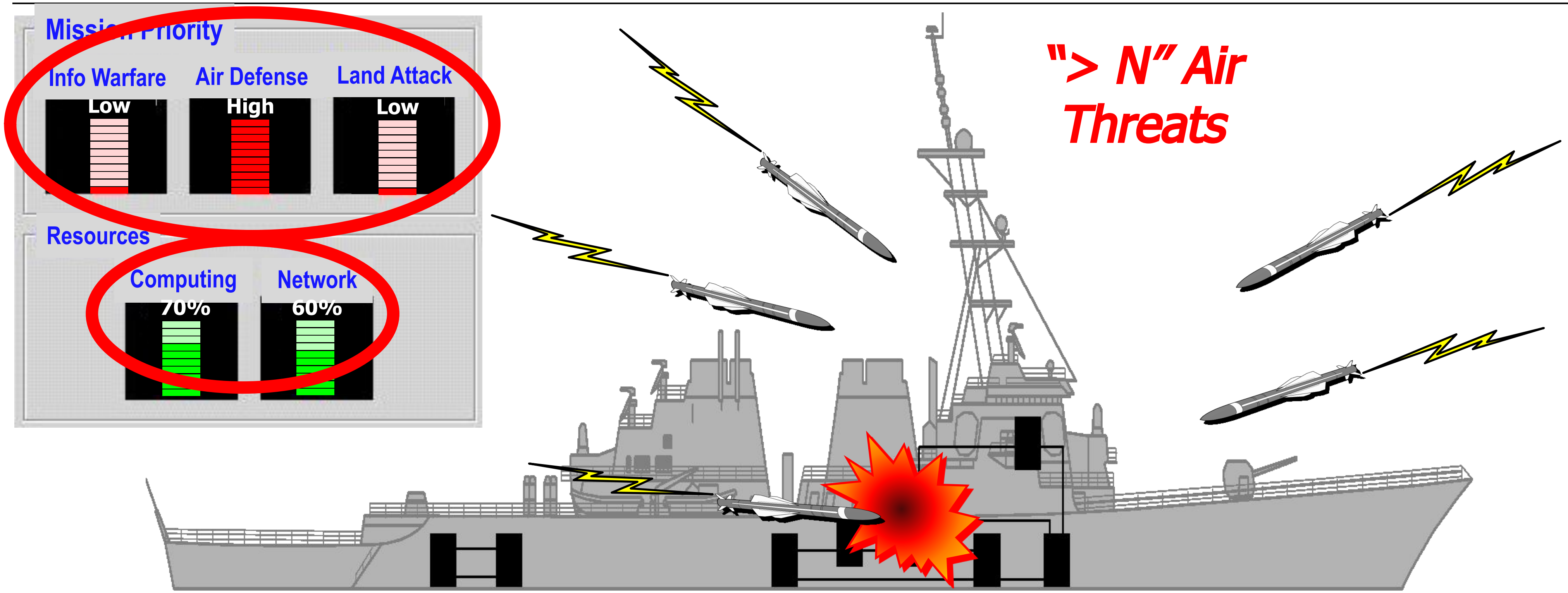
- Static allocation is problematic in several scenarios
  - When # of threats exceed design parameters

TSCE is a Major Improvement on Legacy Combat Systems



- Static allocation is problematic in several scenarios
  - When # of threats exceed design parameters
  - When resources are damaged/degraded

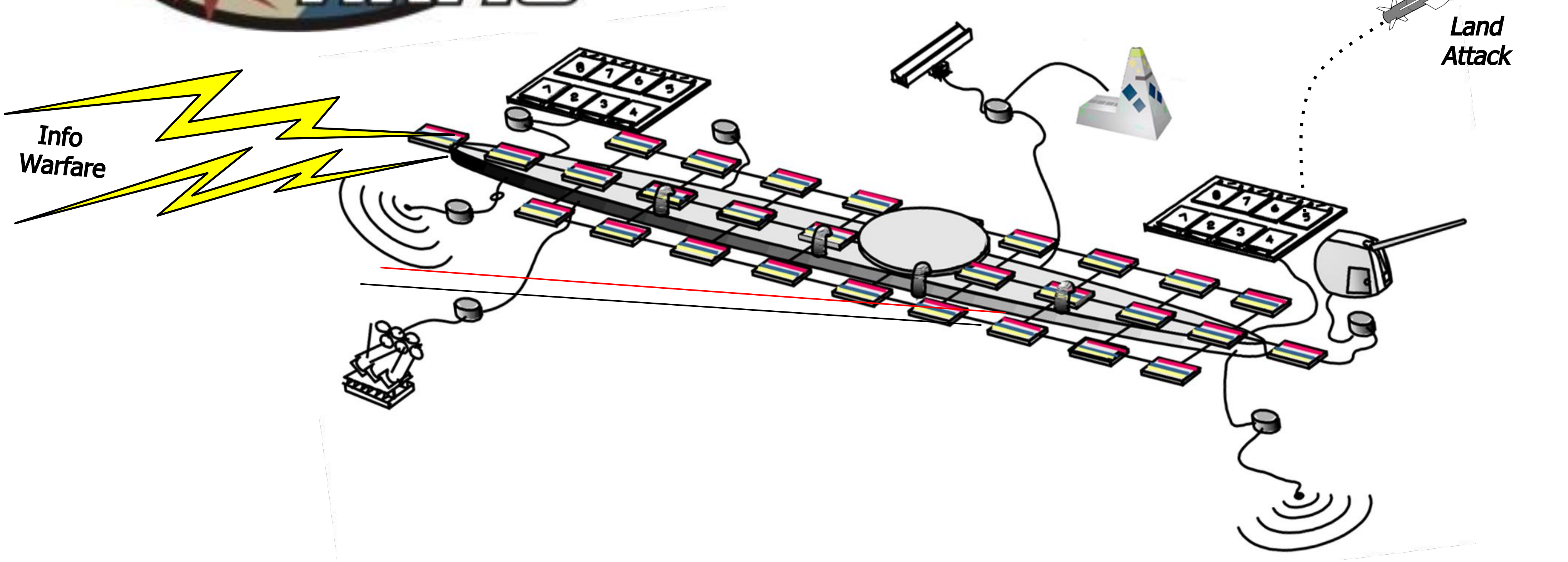
TSCE is a Major Improvement on Legacy Combat Systems



- Static allocation is problematic in several scenarios
  - When # of threats exceed design parameters
  - When resources are damaged/degraded

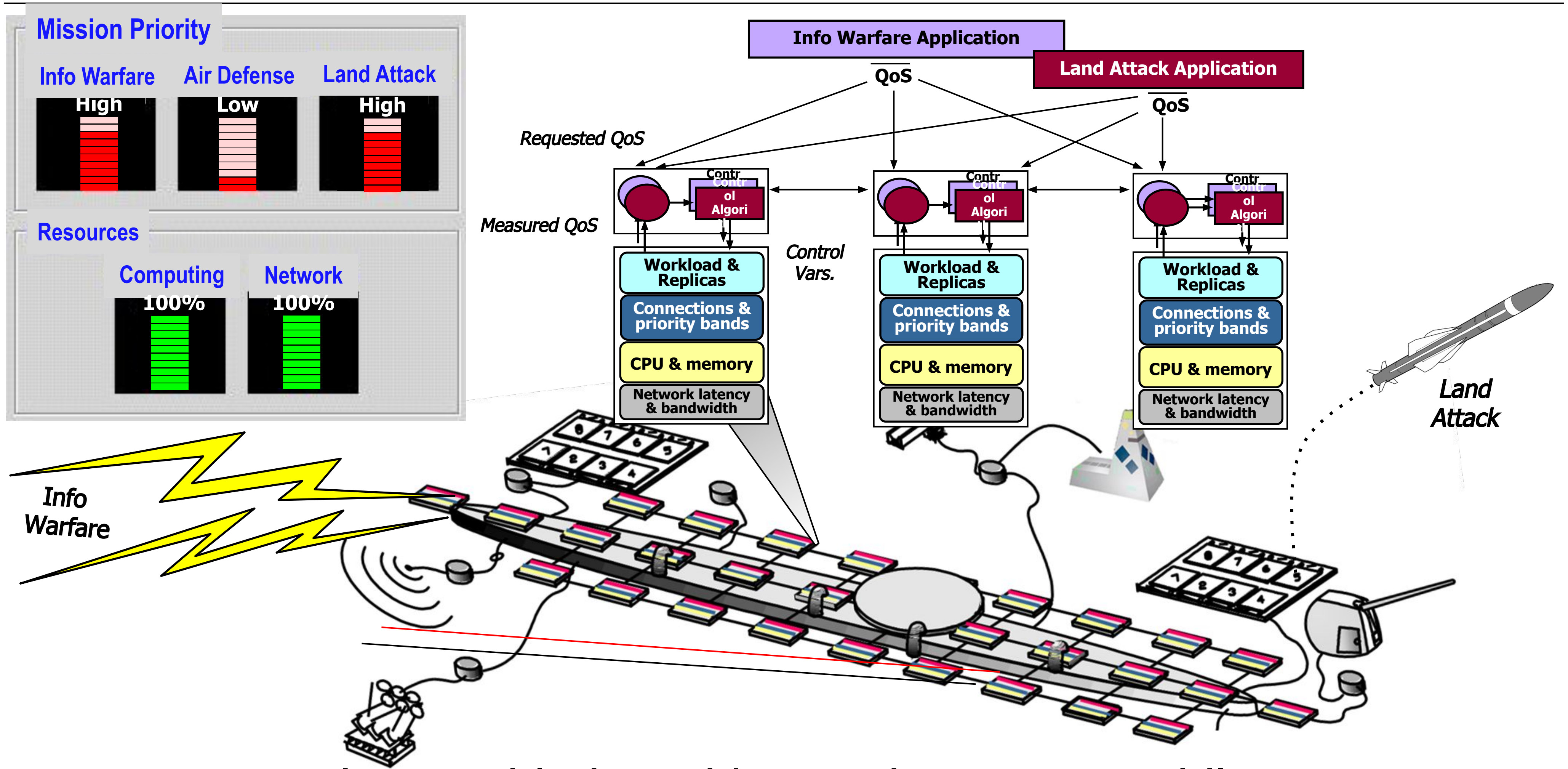
Dynamic resource management (DRM) can help both these scenarios

DARPA ARMS Program Created DRM for DDG 1000 TSCE



www.atl.external.lmco.com/programs/arms.ph

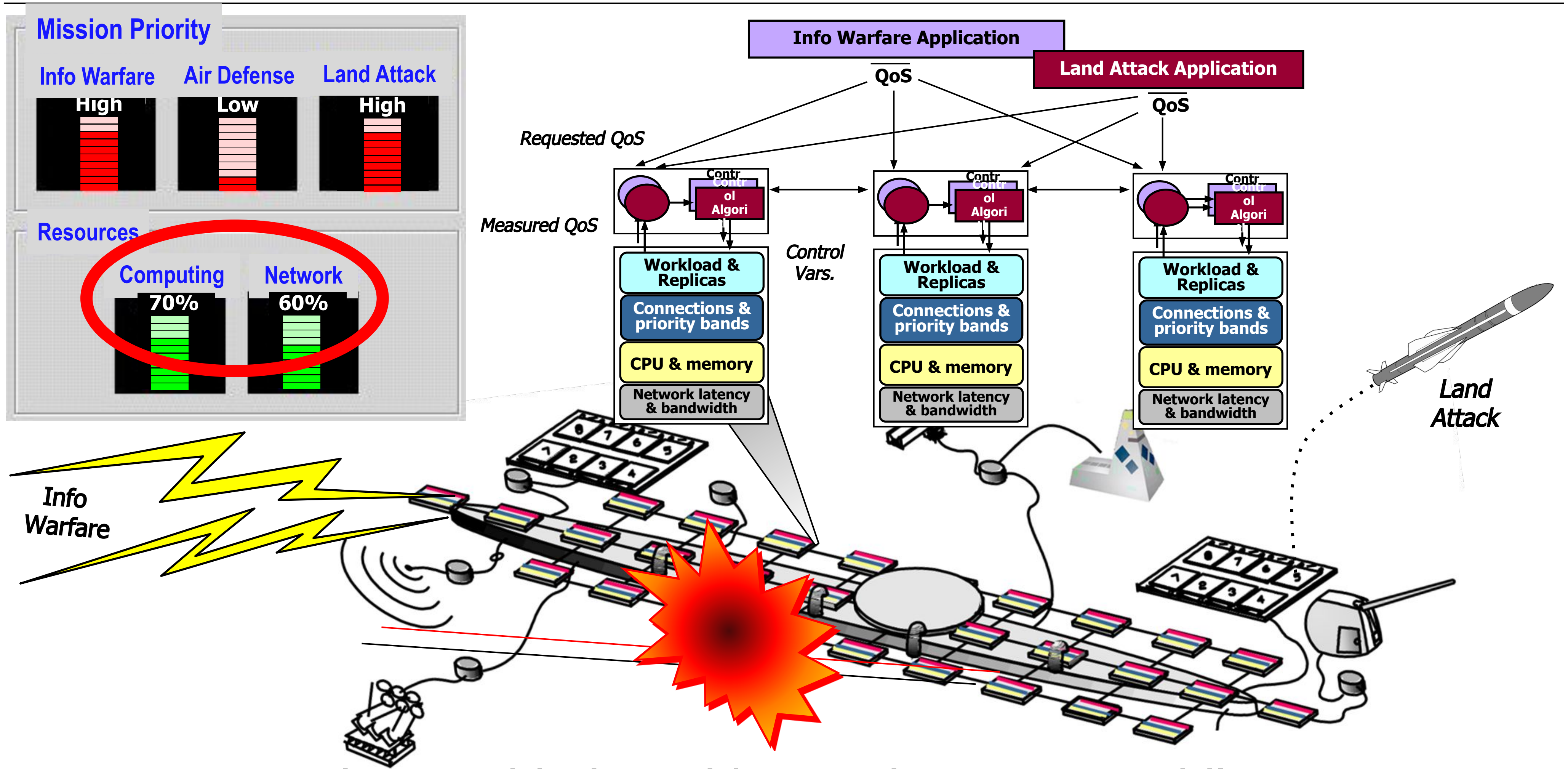
DARPA ARMS Program Created DRM for DDG 1000 TSCE



- ARMS created ensemble-based bin-packing DRM middleware

See www.dre.vanderbilt.edu/~schmidt/PDF/JSS-2006.pdf

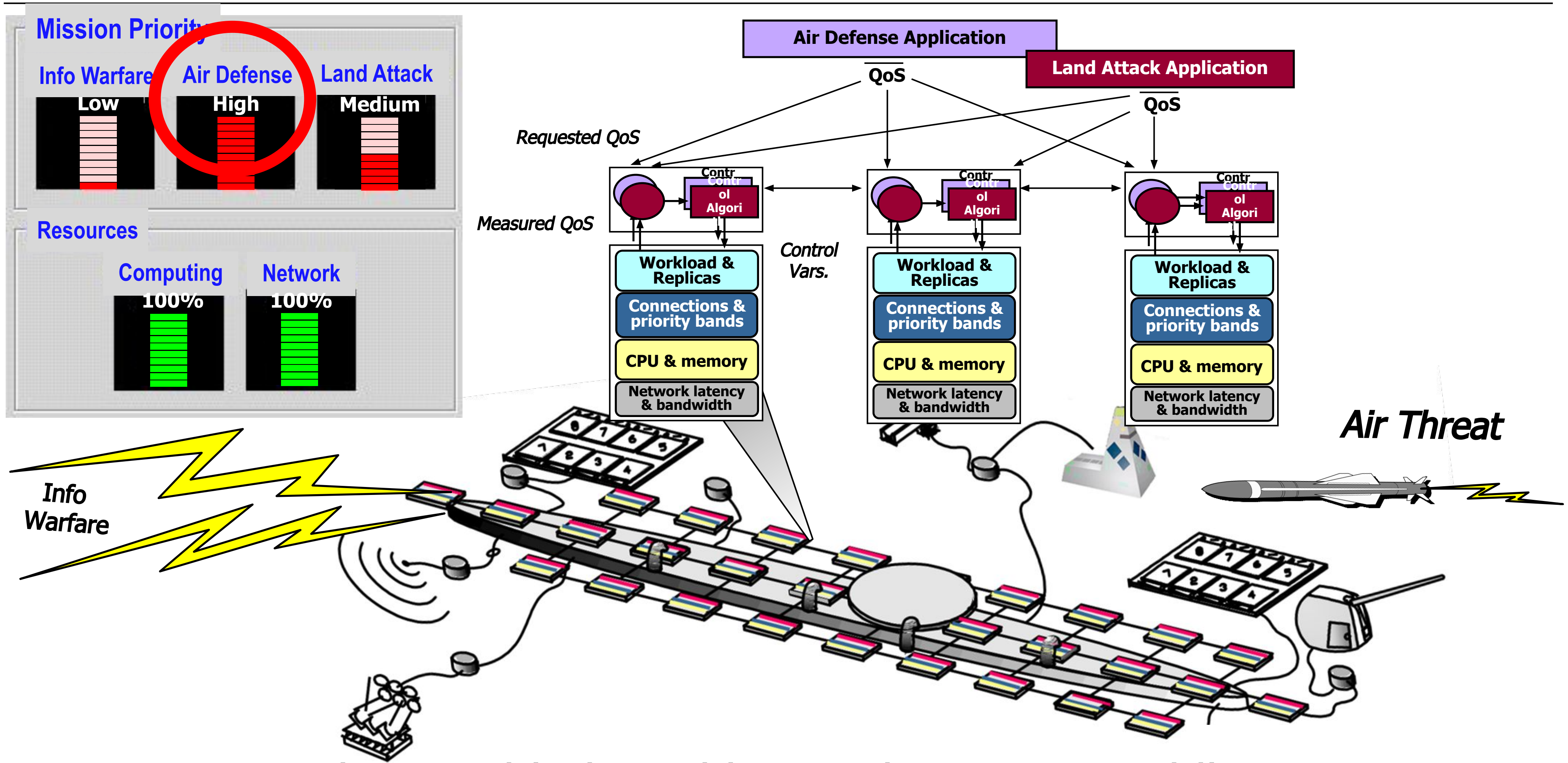
DARPA ARMS Program Created DRM for DDG 1000 TSCE



- ARMS created ensemble-based bin-packing DRM middleware
- Automatically adapts to changes in mission conditions

See www.dre.vanderbilt.edu/~schmidt/PDF/autonomic-journal.pdf

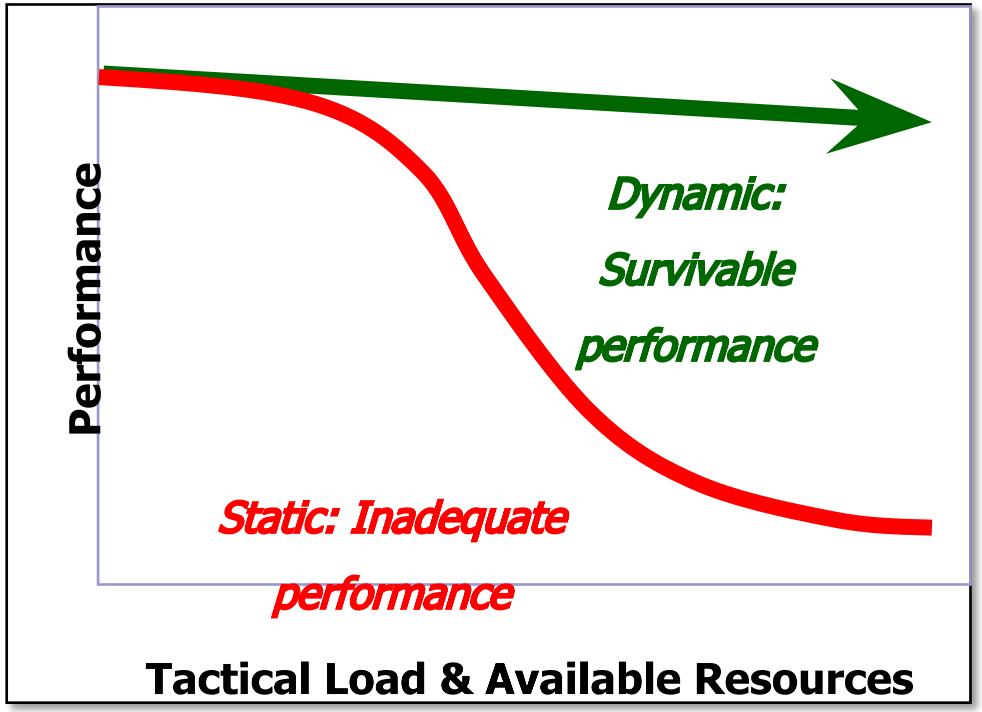
DARPA ARMS Program Created DRM for DDG 1000 TSCE



- ARMS created ensemble-based bin-packing DRM middleware
- Automatically adapts to changes in mission conditions
- Ensures the allocation of computing & network resources accurately matches changing priorities of mission requirements

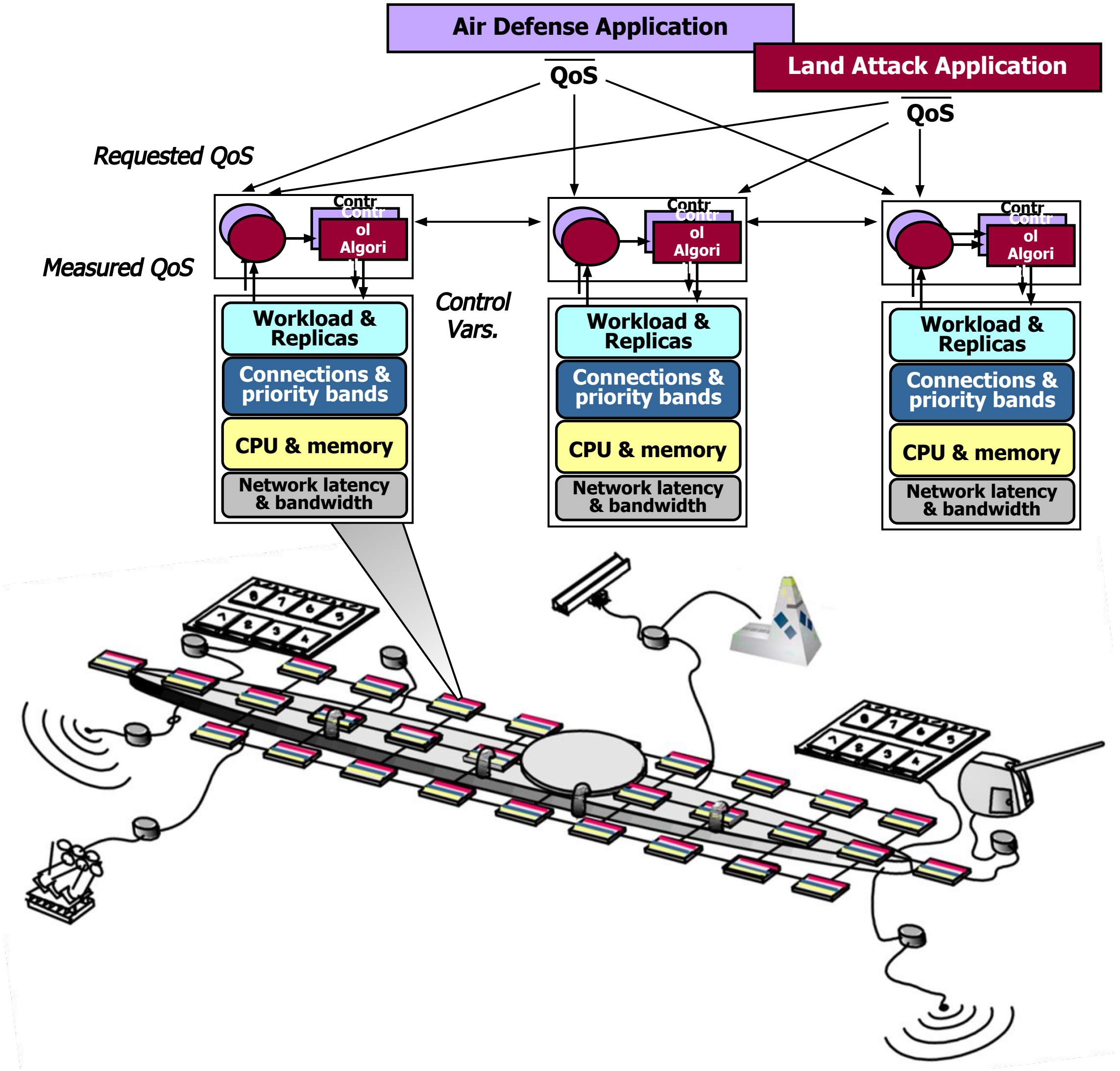
Contributions of the DARPA ARMS Program

- Proved that dynamic resource allocation delivers significantly greater survivability of combat system functionality



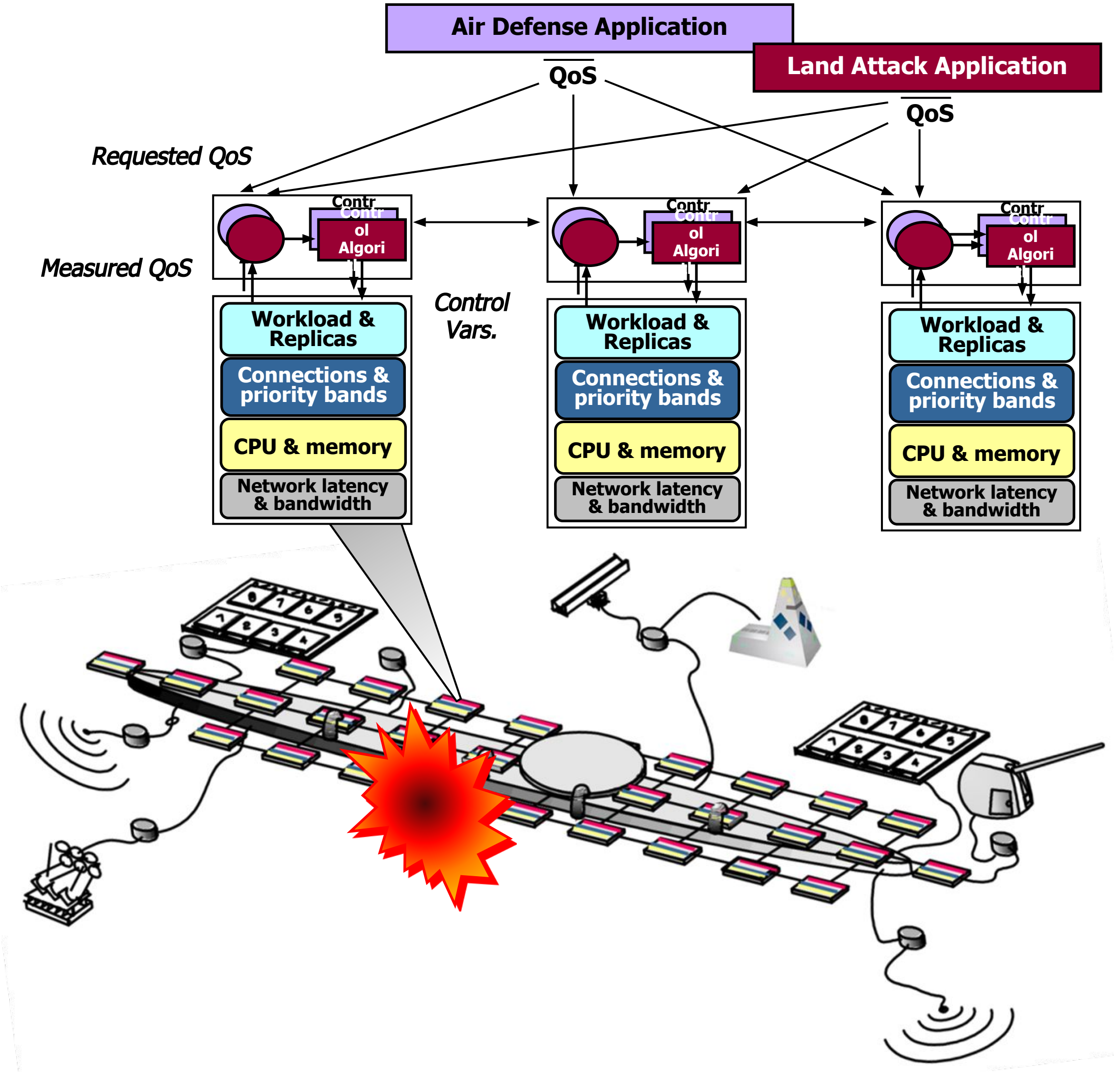
Contributions of the DARPA ARMS Program

- Proved that dynamic resource allocation delivers significantly greater survivability of combat system functionality
- Demonstrated effective & reliable dynamic allocation of processes/tasks to processors



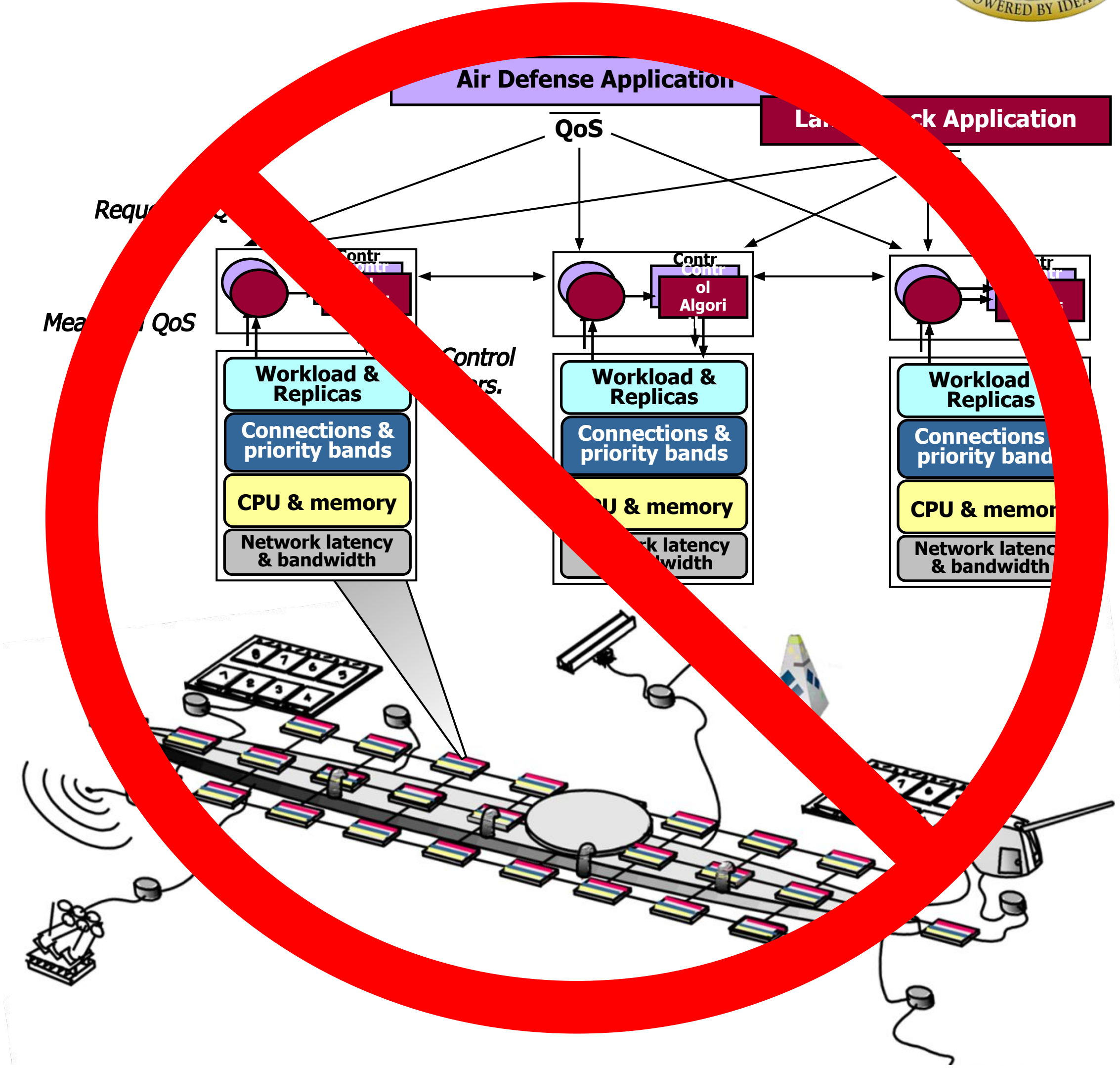
Contributions of the DARPA ARMS Program

- Proved that dynamic resource allocation delivers significantly greater survivability of combat system functionality
- Demonstrated effective & reliable dynamic allocation of processes/tasks to processors
- Demonstrated survivability of dynamic resource allocation system itself



Contributions of the DARPA ARMS Program

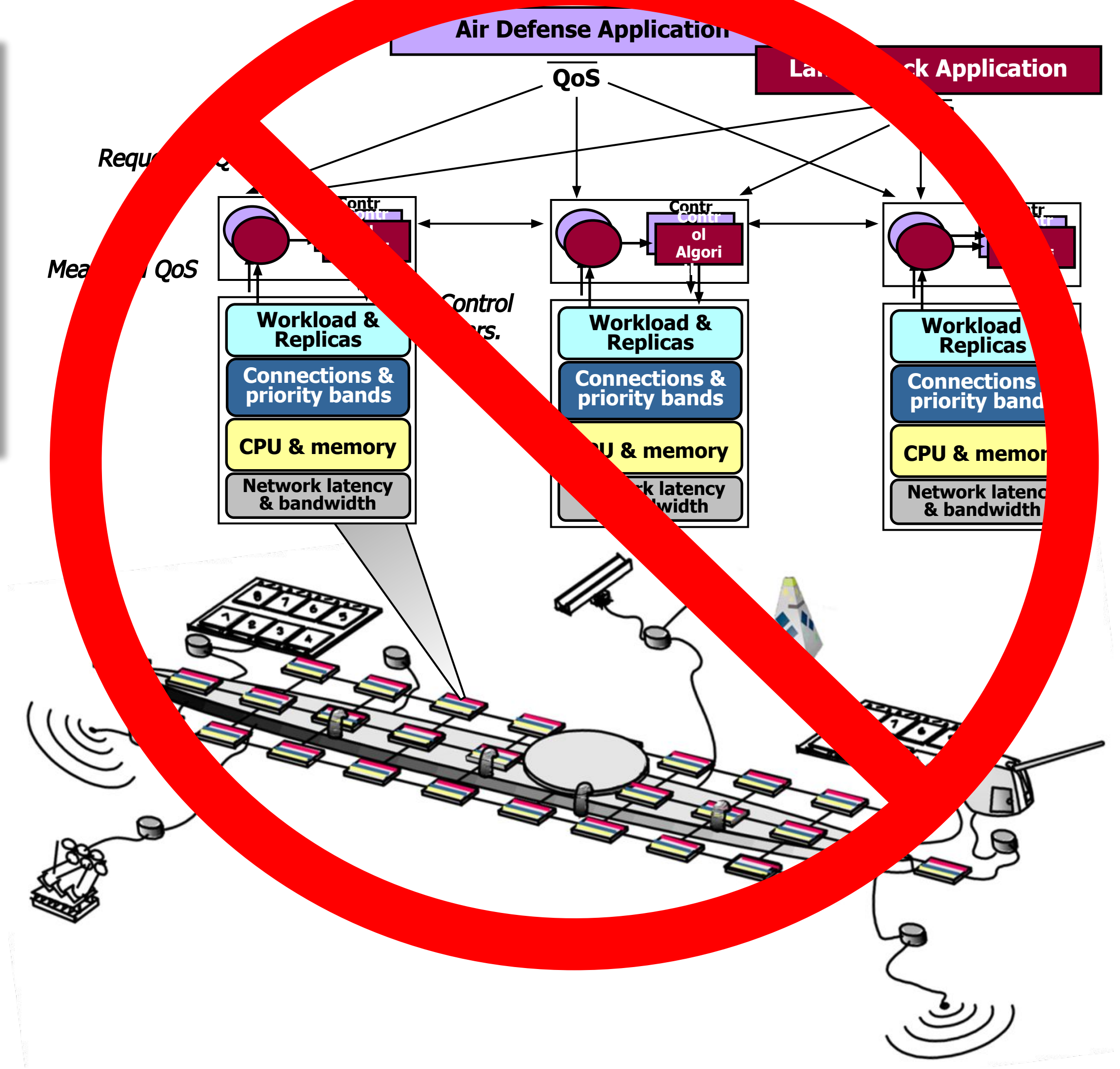
- Proved that dynamic resource allocation delivers significantly greater survivability of combat system functionality
- Demonstrated effective & reliable dynamic allocation of processes/tasks to processors
- Demonstrated survivability of dynamic resource allocation system itself



However, the Navy could not deploy ARMS DRM-based systems due to challenges in certifying adaptive systems

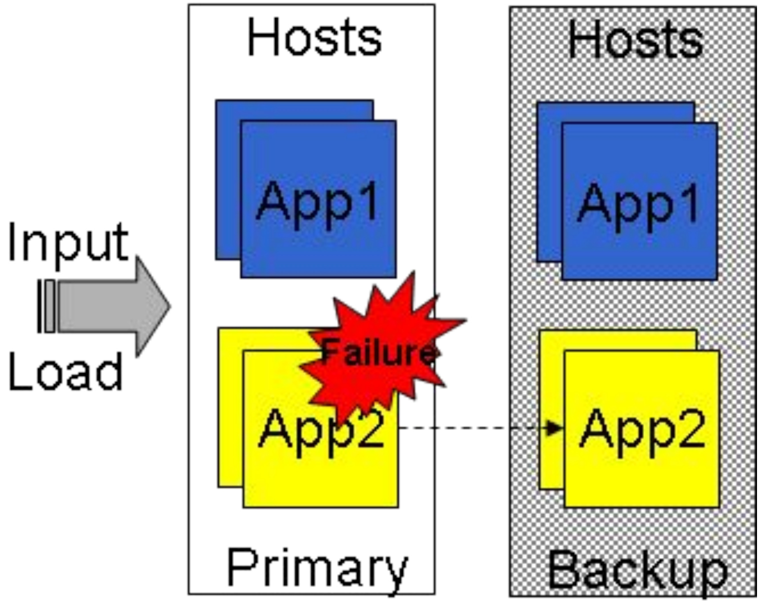
Contributions of the DARPA ARMS Program

"We can't certify that DDG-1000 doesn't move into unstable, incorrect, or unsafe operating configurations during system operations"



Challenges of Adaptive Dynamic Computing Environments

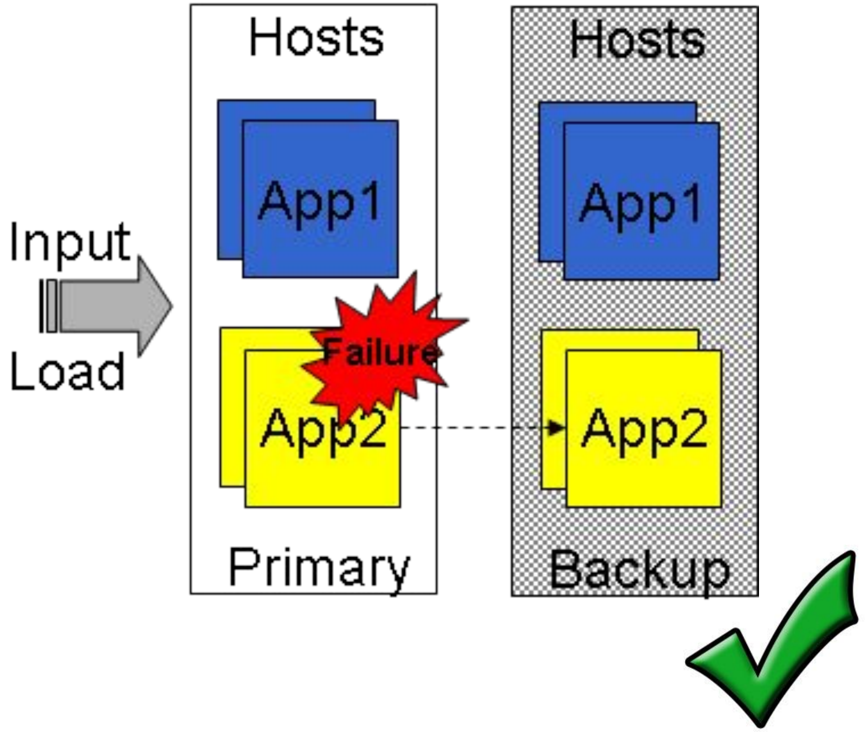
- To ensure real-time predictable quality of service mission-/safety-critical DoD systems today are statically configured



www.acq.osd.mil/se/docs/DoD-Safety-Data-Package-Preparation-Guide.pdf

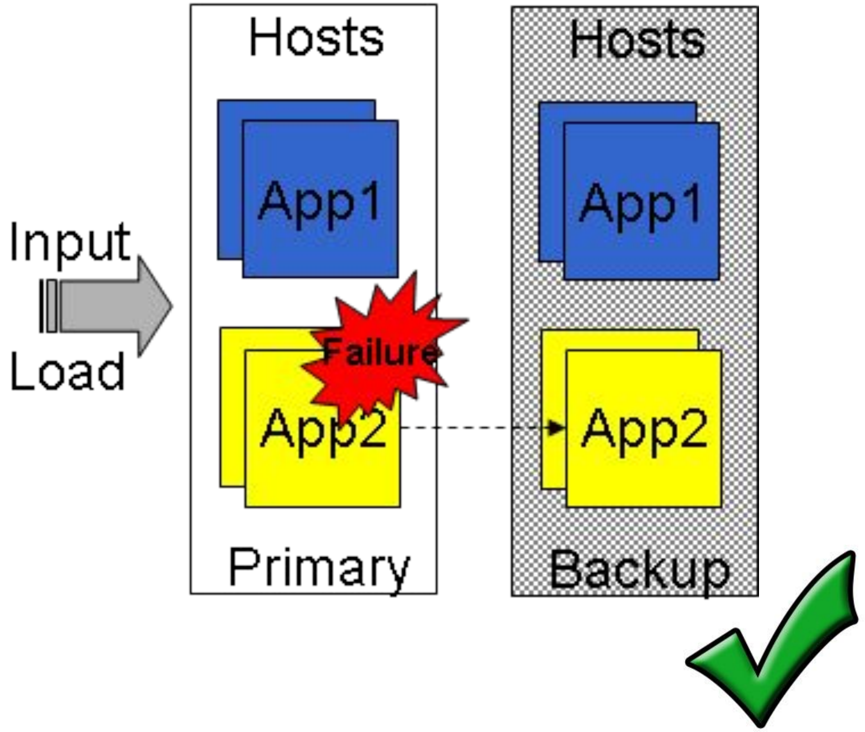
Challenges of Adaptive Dynamic Computing Environments

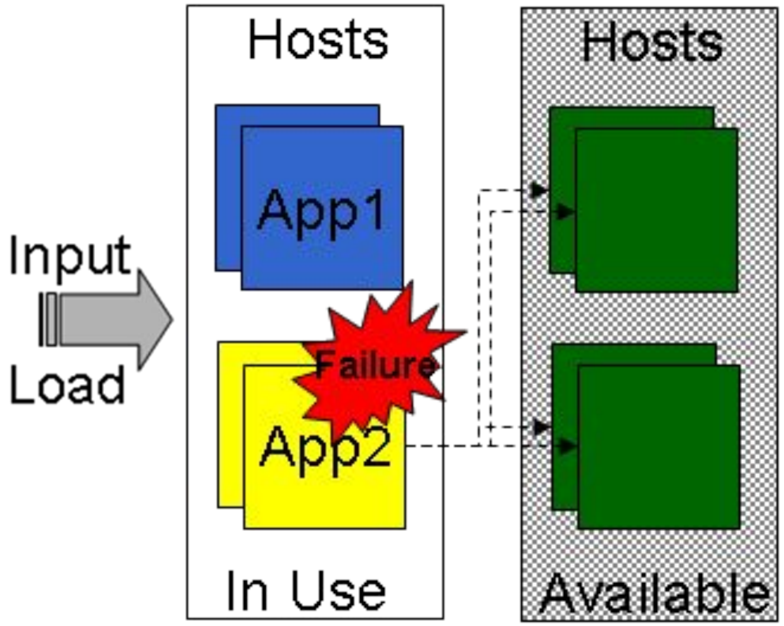
- To ensure real-time predictable quality of service mission-/safety-critical DoD systems today are statically configured
- There are a range of time-proven techniques for certifying statically configured systems



Challenges of Adaptive Dynamic Computing Environments

- To ensure real-time predictable quality of service mission-/safety-critical DoD systems today are statically configured
- There are a range of time-proven techniques for certifying statically configured systems

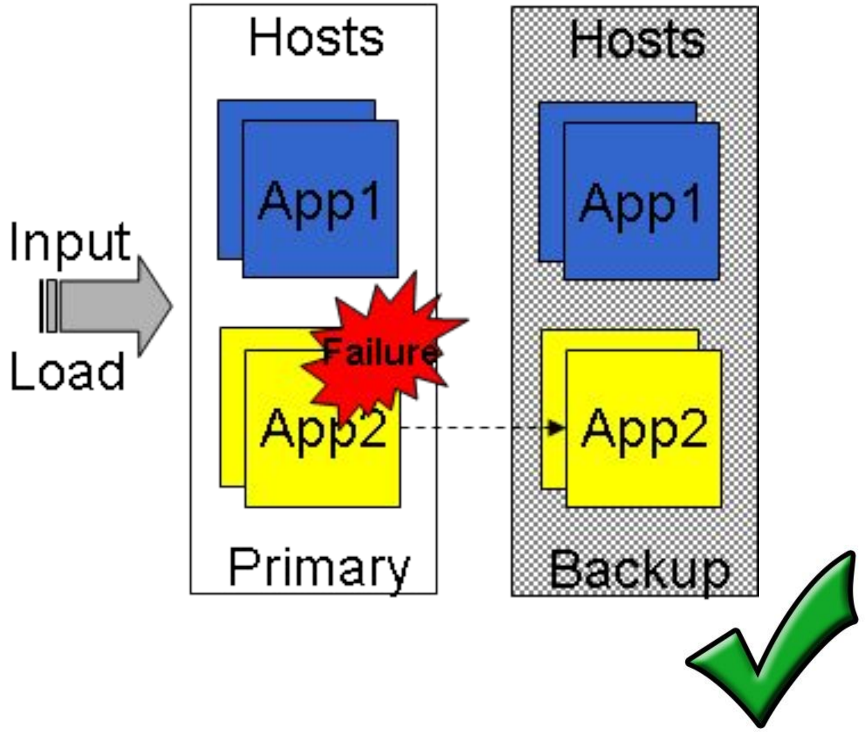


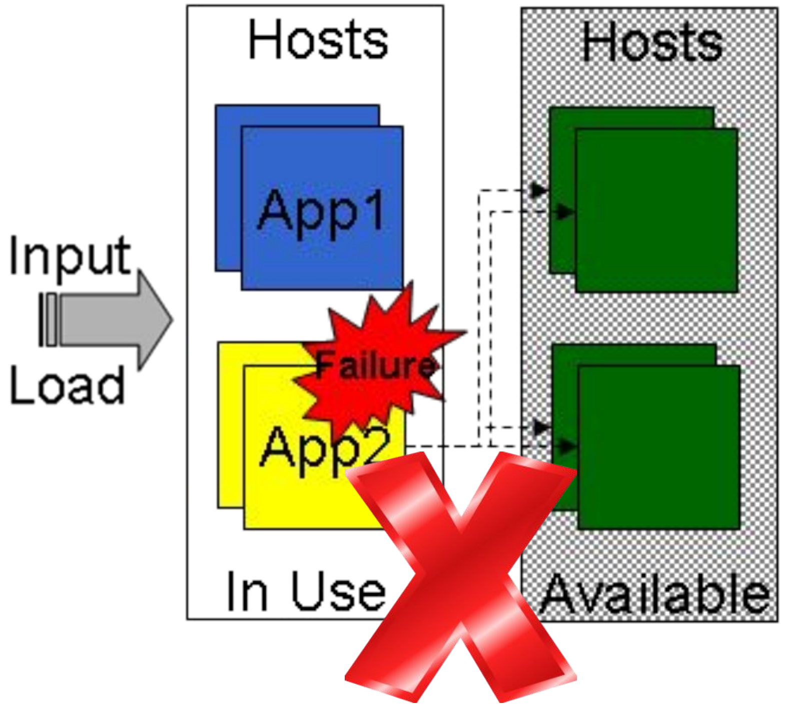


- Dynamically managed systems deliver far greater efficiency & survivability

Challenges of Adaptive Dynamic Computing Environments

- To ensure real-time predictable quality of service mission-/safety-critical DoD systems today are statically configured
- There are a range of time-proven techniques for certifying statically configured systems

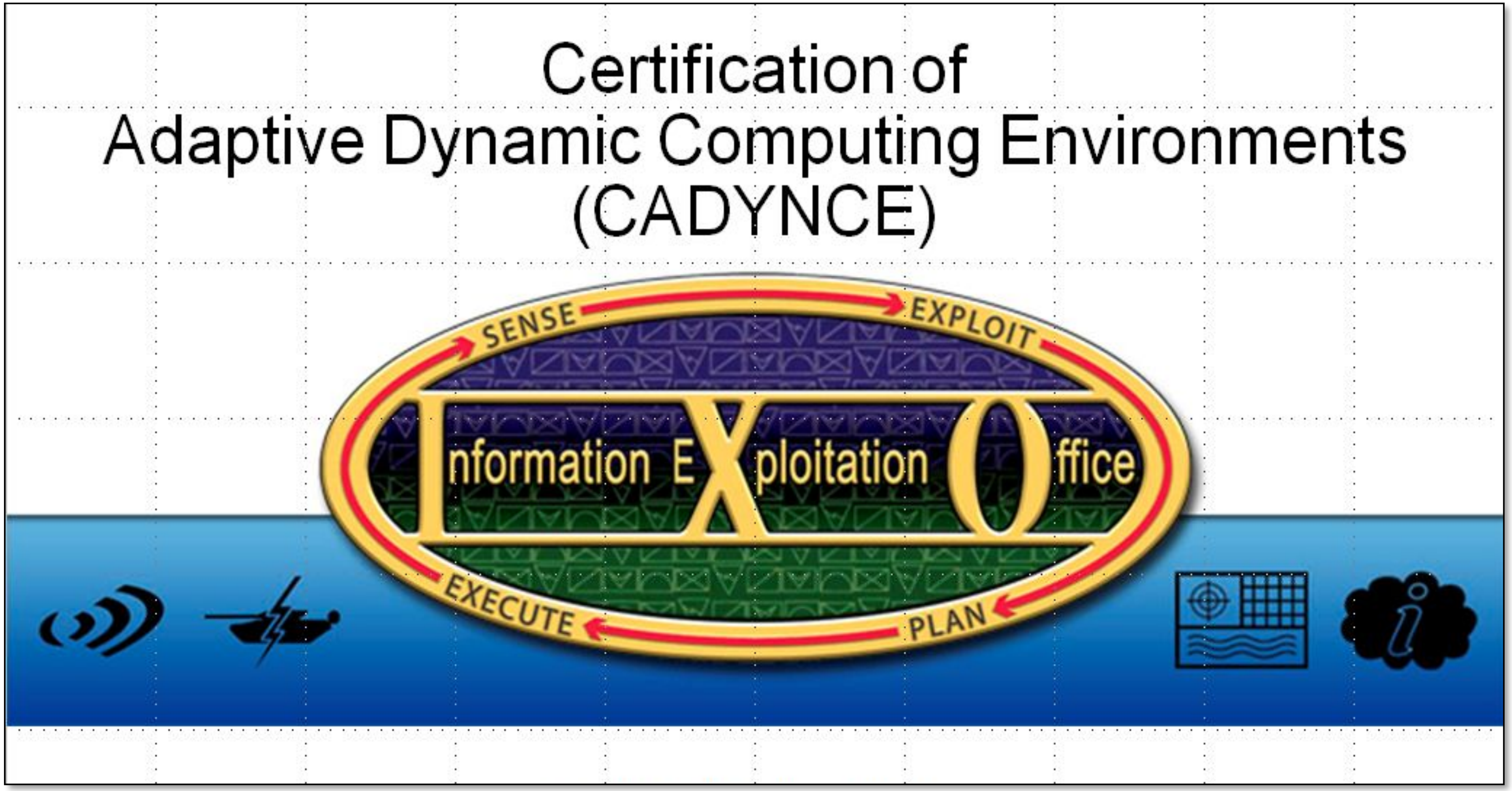




- Dynamically managed systems deliver far greater efficiency & survivability
- However, DRM techniques can't currently be deployed in mission-/safety-critical DoD combat systems because they are not certifiable via conventional methods

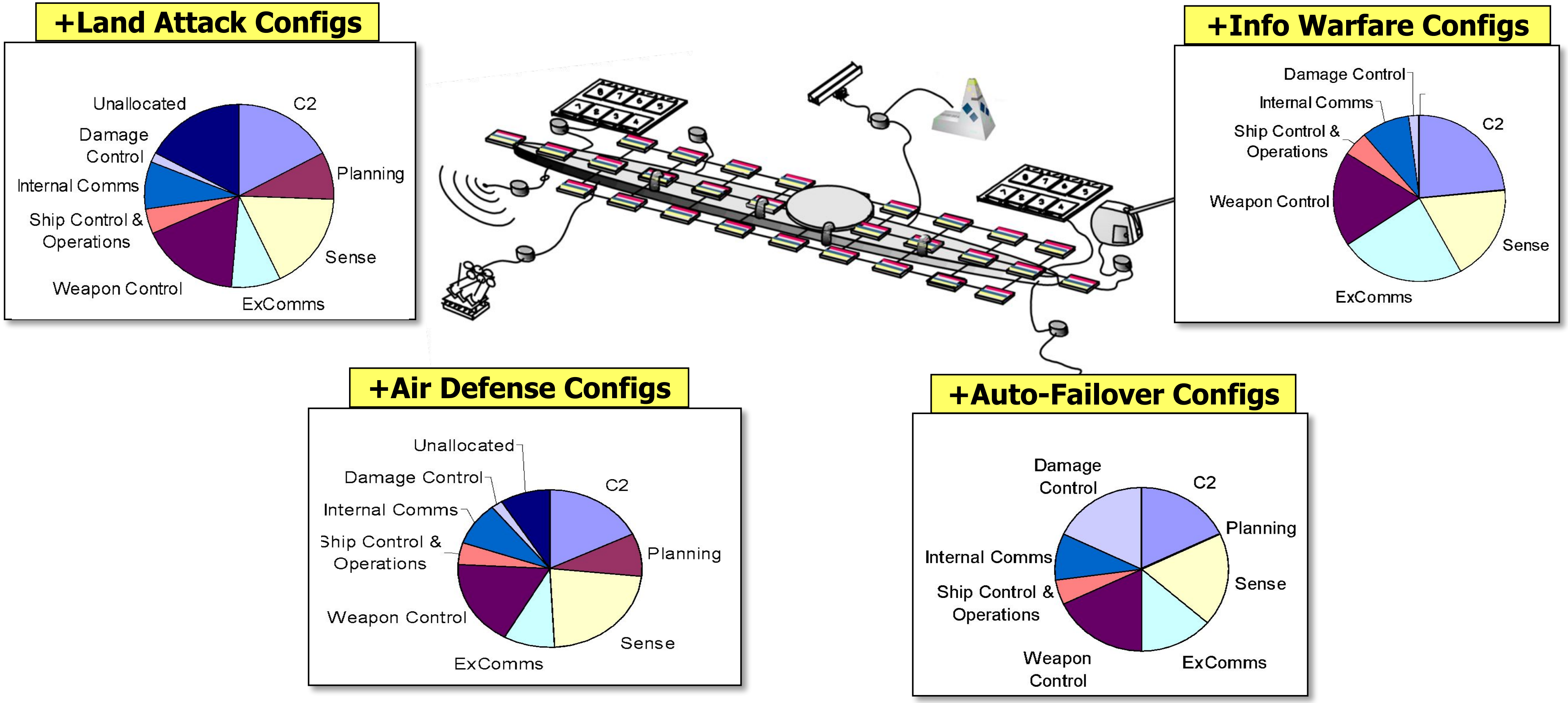
Towards Certification of Adaptive Dynamic Computing Environments

- The *Certification of Adaptive Dynamic Computing Environments* (CADYNCE) was a follow-on to ARMS that focused on two topics:



Towards Certification of Adaptive Dynamic Computing Environments

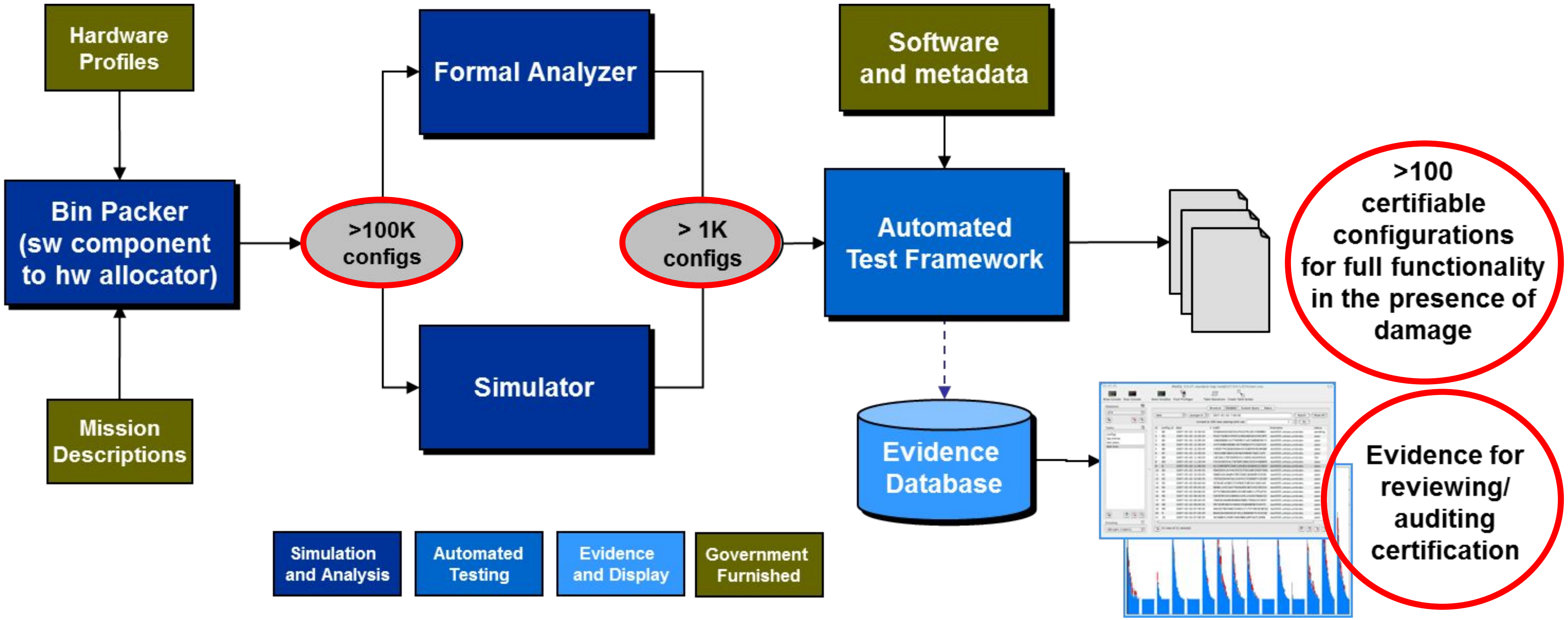
- The *Certification of Adaptive Dynamic Computing Environments* (CADYNCE) was a follow-on to ARMS that focused on two topics:
  - How to constrain ARMS DRM to make it more "certification friendly"



Pre-compute 100's of certified configs that are then used dynamically

Towards Certification of Adaptive Dynamic Computing Environments

- The *Certification of Adaptive Dynamic Computing Environments* (CADYNCE) was a follow-on to ARMS that focused on two topics:
  - How to constrain ARMS DRM to make it more "certification friendly"
  - How to create & integrate an assisted certification tool chain



Towards Certification of Adaptive Dynamic Computing Environments

Increasingly Static Configuration Management ⟵⟶ **Increasingly Dynamic Configuration Management**

- **At runtime select from a few manually generated pre-certified configurations**
- **At runtime select from many automatically generated, analyzed, tested, & pre-certified configurations**
- **At runtime generate new configurations based on conditions that affect current configuration**
- **Continuous adaptation through fine-grained monitoring of resources & applications**

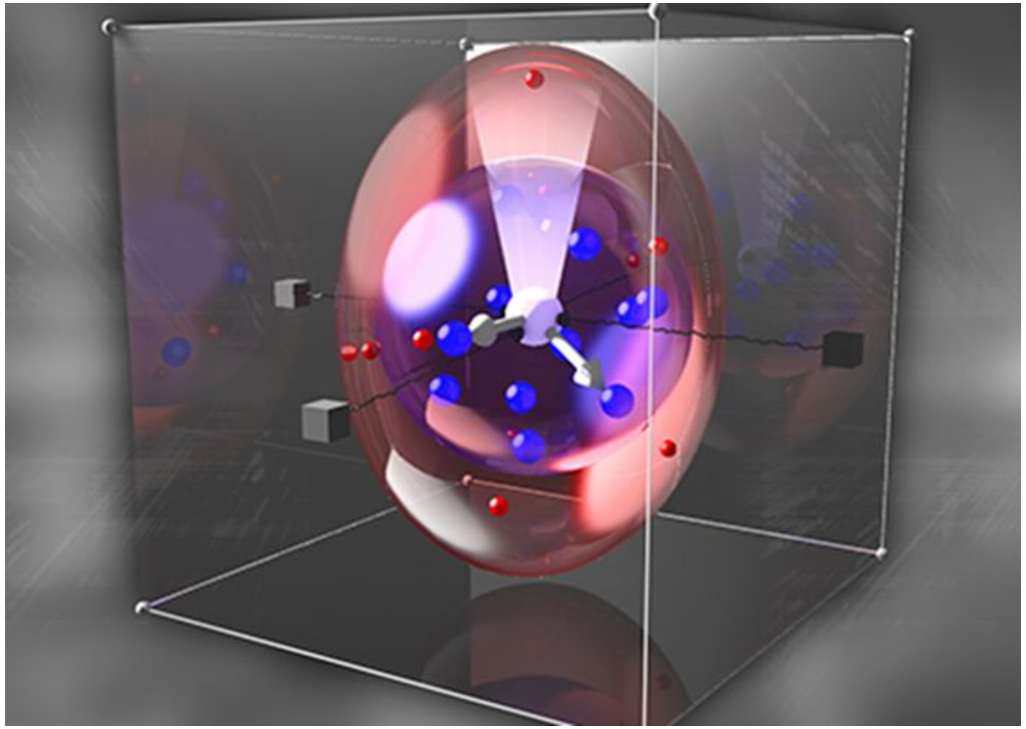
CADYNCE focused on enabling certification of 100's of configurations that can be selected at runtime

No technical approach enabled certification of more dynamic configuration management approaches

Availability of 100's of certified configurations demo'd benefits of DRM behavior, while maintaining the assurance of certification, but more R&D is needed

Concluding Remarks

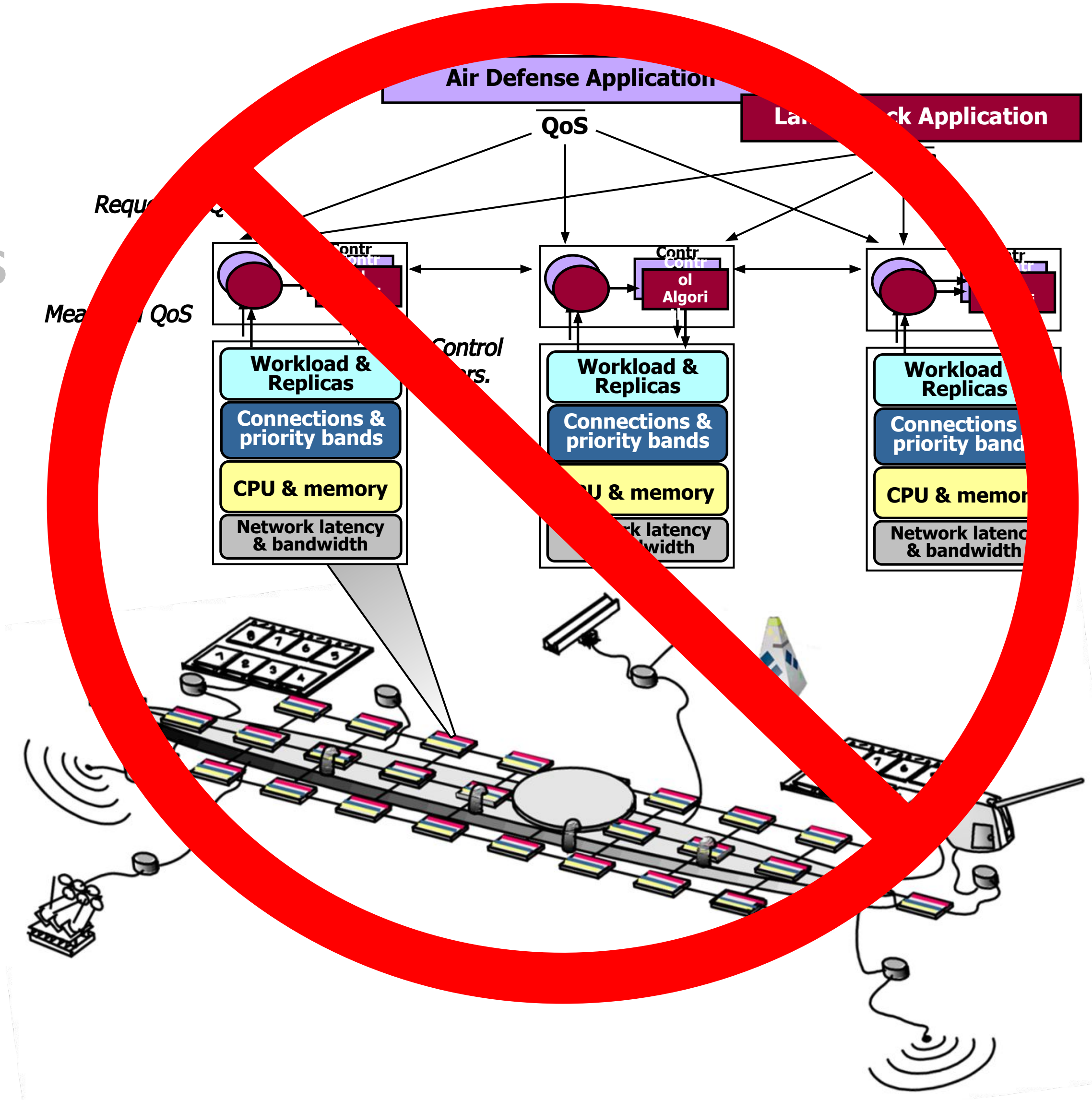
- Adaptive dynamic computing environments remain a key topic for research & practice in mission-/safety-critical systems



www.darpa.mil/program/building-resource-adaptive-software-systems

Concluding Remarks

- Adaptive dynamic computing environments remain a key topic for research & practice in mission-/safety-critical systems
- Adaptive dynamic computing environments aren’t deployed in DoD combat systems since they aren’t yet certifiable via conventional methods



Concluding Remarks

- Adaptive dynamic computing environments remain a key topic for research & practice in mission-/safety-critical systems
- Adaptive dynamic computing environments aren't deployed in DoD combat systems since they aren't yet certifiable via conventional methods
- It's easier to pitch programs on adaptive computing than to pitch programs on *certification* of adaptive computing..

Panel Members

Dr. Douglas C. Schmidt is the Cornelius Vanderbilt Professor of Computer Science, Associate Provost for Research Development and Technologies, Co-Chair of the Data Science Institute, and a Senior Researcher at the Institute for Software Integrated Systems, all at Vanderbilt University.

Dr. Mikael Lindvall Dr. Mikael Lindvall, is the Technology Director of Fraunhofer USA, Center Mid-Atlantic (CMA) and an adjunct professor at UMCP for more than 10 years. He recently created and delivered one of the first courses on Software Engineering for AI-based systems.

Dr. Jeffrey W. Herrmann is a professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the director of the Reliability Engineering Graduate Program and the director of the Systems Engineering Graduate Program at the University of Maryland.

Dr. Laura Freeman is a Research Associate Professor of Statistics and the Director of the Intelligent Systems Lab at the Virginia Tech Hume Center. Previously, she was the Assistant Director of the Operational Evaluation Division at the Institute for Defense Analyses.

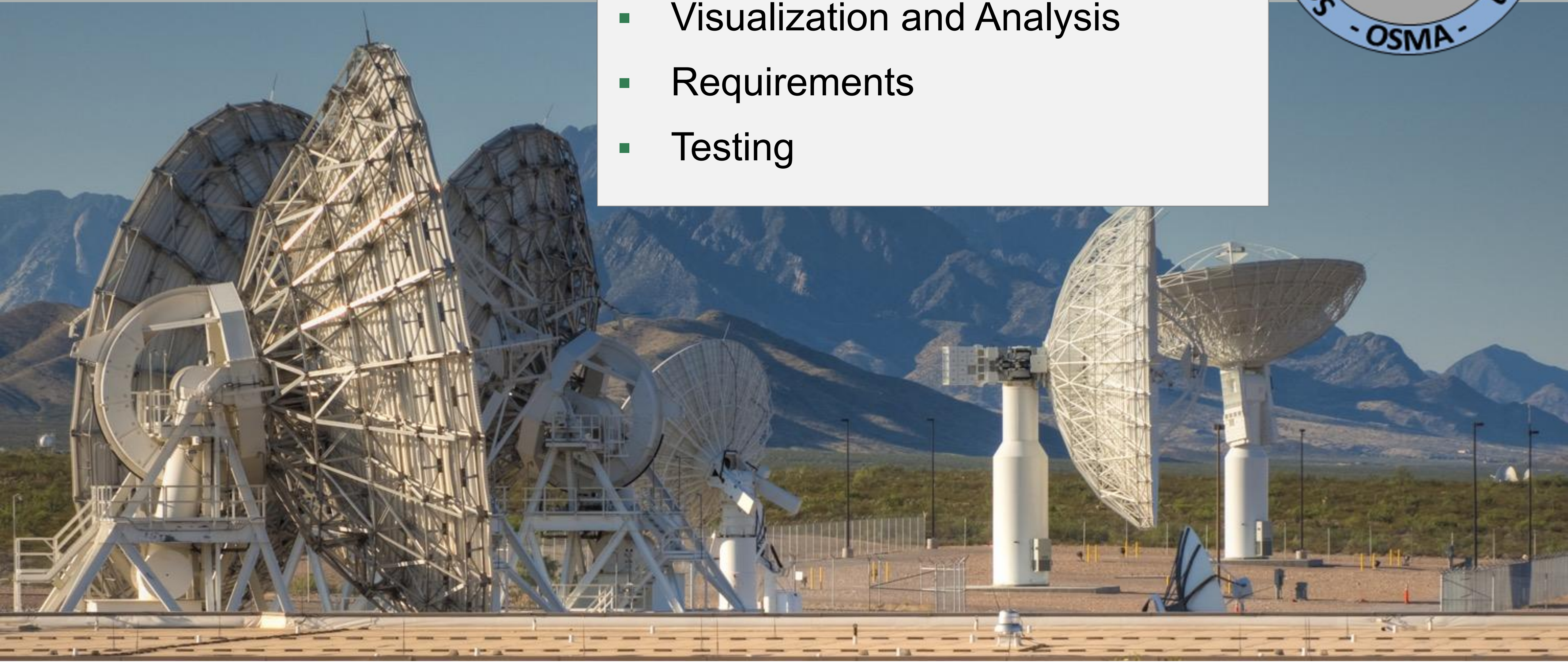
AI-BASED SYSTEMS NEED BETTER ENGINEERING TOOLS

Dr. Mikael Lindvall

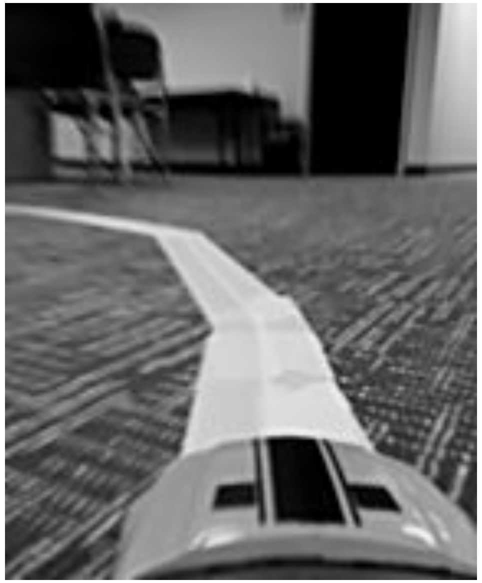
Fraunhofer USA CMA

Traditional Tools and Methods Don't Always Work With AI

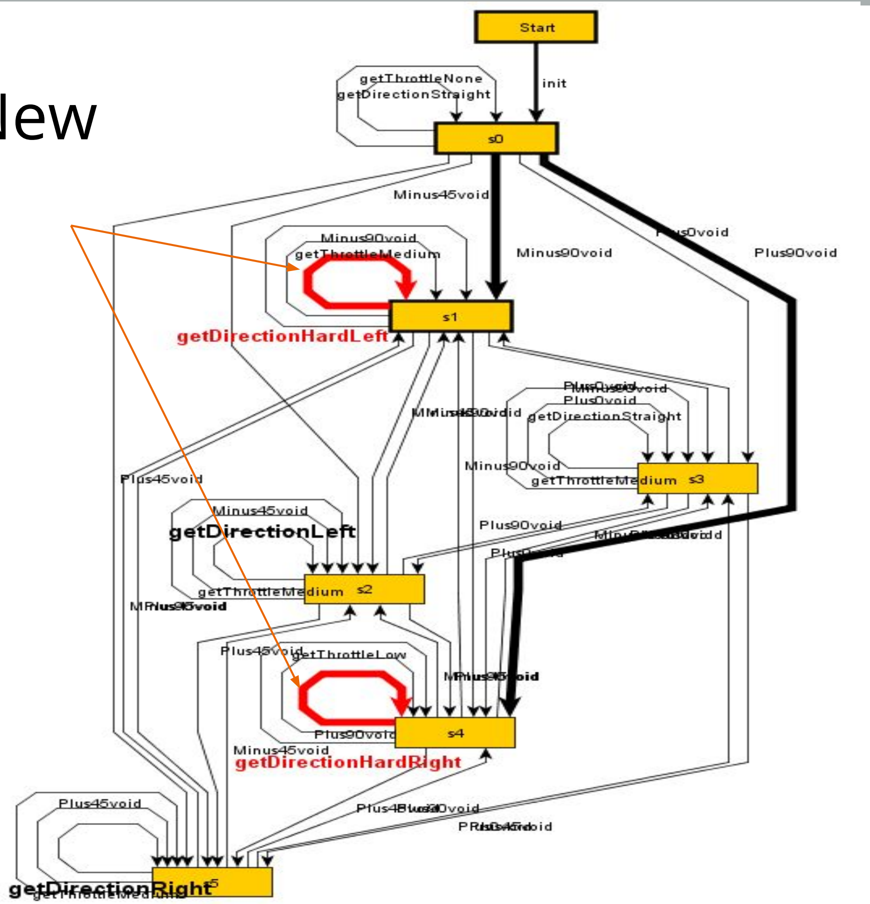
- Visualization and Analysis
- Requirements
- Testing



Visualization and Analysis of Learned Behavior



Incorrect/New behavior

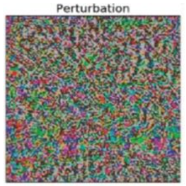


Requirements and Testing of AI and Autonomous Behavior

Testing and Protecting AI-Based Image Recognition Systems

Clean

Adversarial

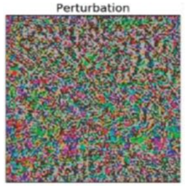


Rotation – 0.5 degrees

Ocean Cruiser

Submarine

Rotation – 0.5 degrees



Ocean Cruiser

Carrier

AI-Based Systems Need Better Engineering Tools and Methods

- That provide visibility into neural networks
- That allow us to model and simulate requirements
- That allow us to generate test cases and identify oddities in system behavior
  - During testing and runtime

Fraunhofer USA

WANT TO HEAR MORE?

For more information

mlindvall@fraunhofer.org

Panel Members

Dr. Douglas C. Schmidt is the Cornelius Vanderbilt Professor of Computer Science, Associate Provost for Research Development and Technologies, Co-Chair of the Data Science Institute, and a Senior Researcher at the Institute for Software Integrated Systems, all at Vanderbilt University.

Dr. Mikael Lindvall Dr. Mikael Lindvall, is the Technology Director of Fraunhofer USA, Center Mid-Atlantic (CMA) and an adjunct professor at UMCP for more than 10 years. He recently created and delivered one of the first courses on Software Engineering for AI-based systems.

Dr. Jeffrey W. Herrmann is a professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the director of the Reliability Engineering Graduate Program and the director of the Systems Engineering Graduate Program at the University of Maryland.

Dr. Laura Freeman is a Research Associate Professor of Statistics and the Director of the Intelligent Systems Lab at the Virginia Tech Hume Center. Previously, she was the Assistant Director of the Operational Evaluation Division at the Institute for Defense Analyses.

AI Systems as Organizations

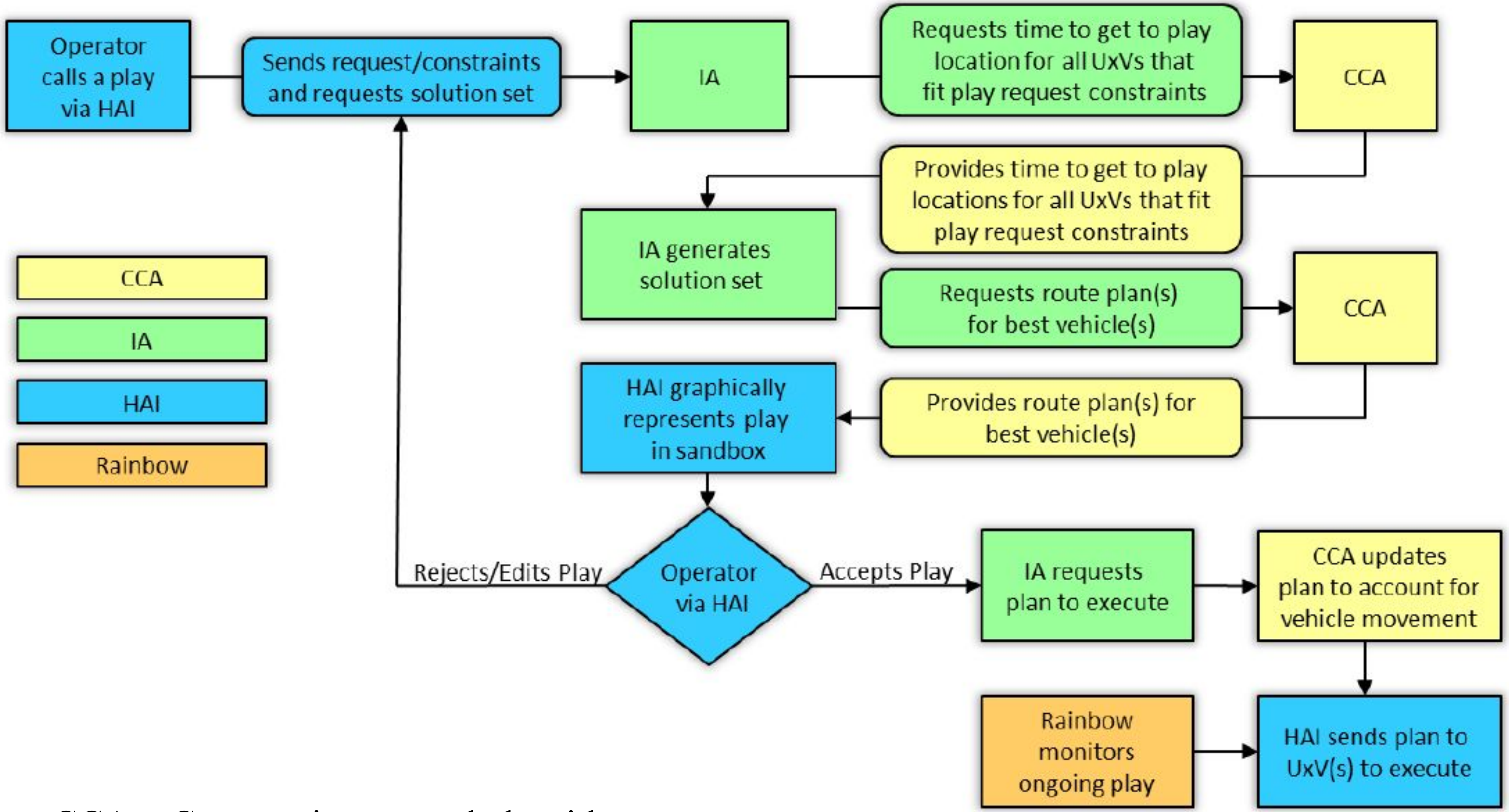
Jeffrey W. Herrmann

1. Overview
2. Case study: IMPACT (Behymer *et al*., 2017)
3. Implications

Behymer, Kyle, Clayton Rothwell, Heath Ruff, Michael Patzek, Gloria Calhoun, Mark Draper, Scott Douglass, Derek Kingston, and Doug Lange. *Initial Evaluation of the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT)*. Infoscitex Corp. Beavercreek, Ohio, 2017.
AFRL report number: AFRL-RH-WP-TR-2017-0011

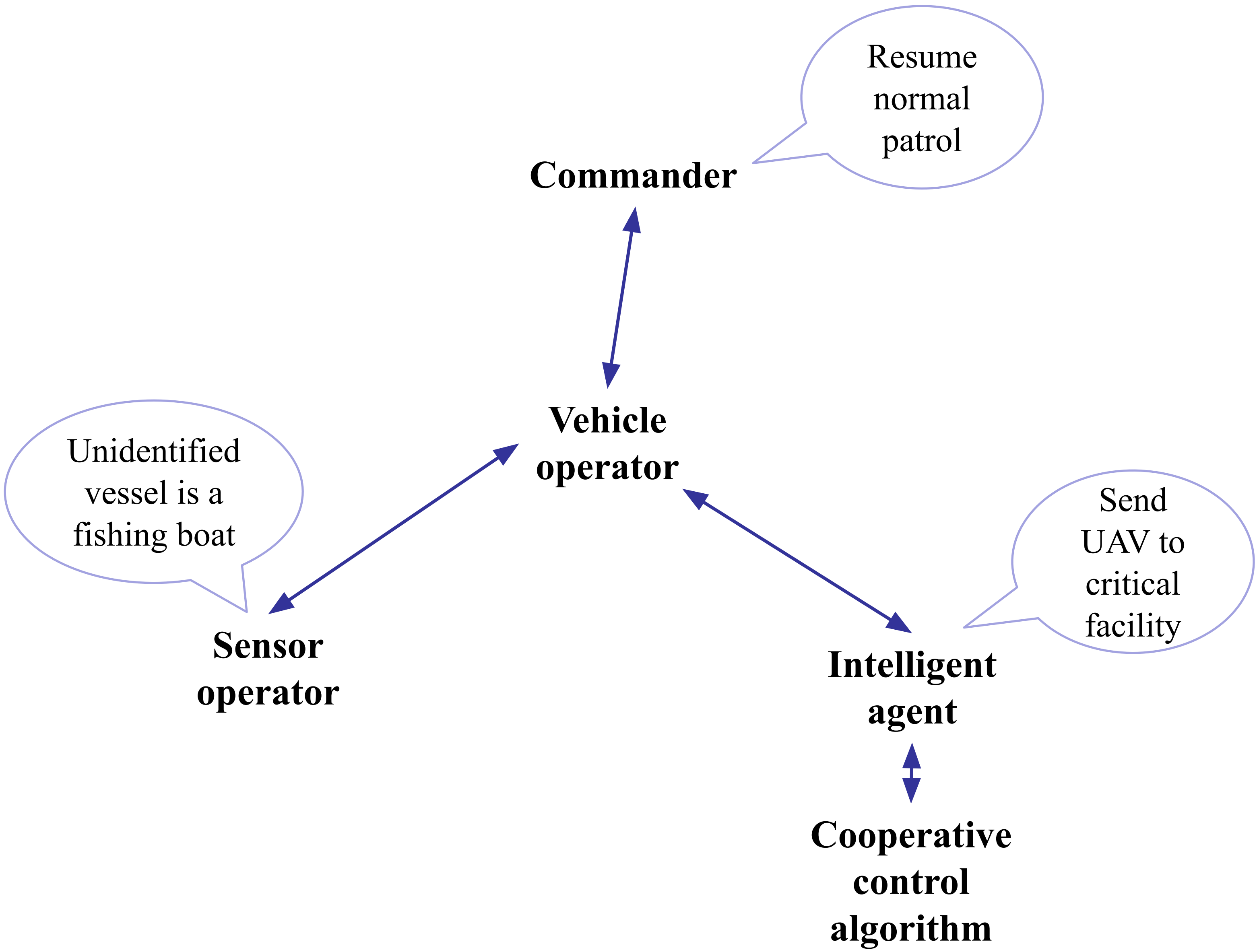
The IMPACT user interface (Behymer *et al.*, 2017)

Play calling flowchart (Behymer *et al*., 2017)

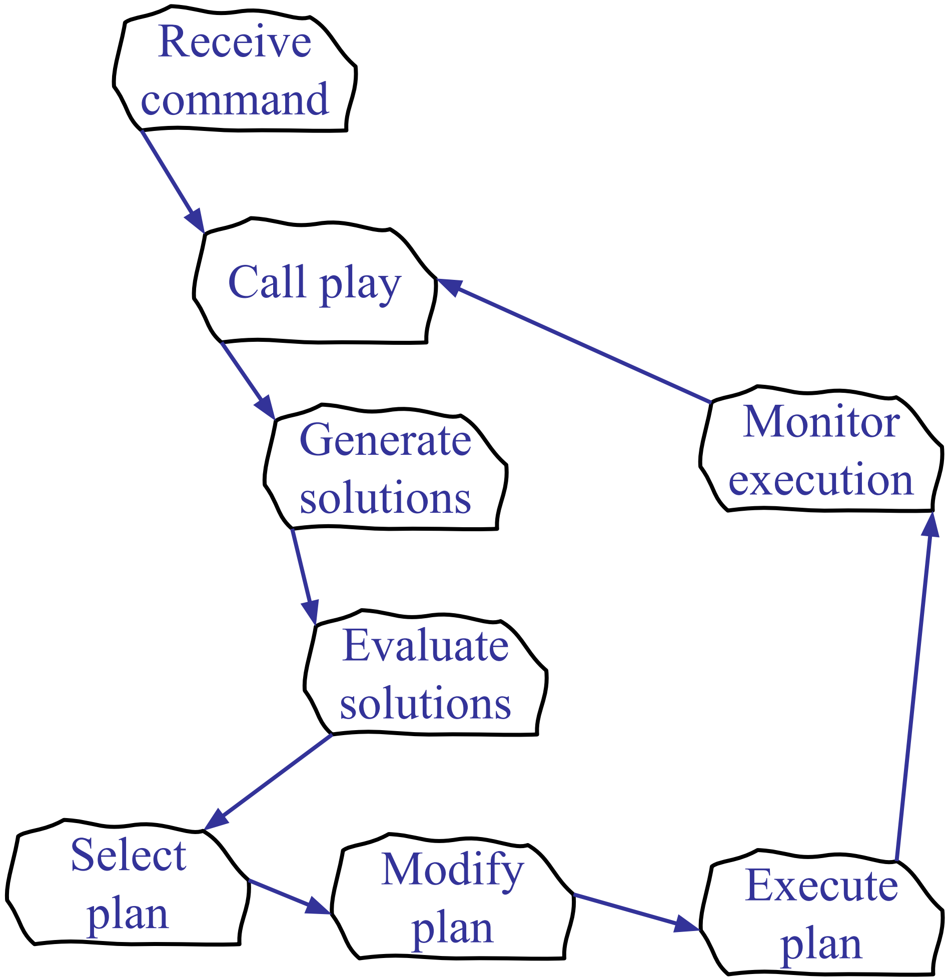


CCA = Cooperative control algorithm
IA = Intelligent agent
HAI = Human-autonomy interface

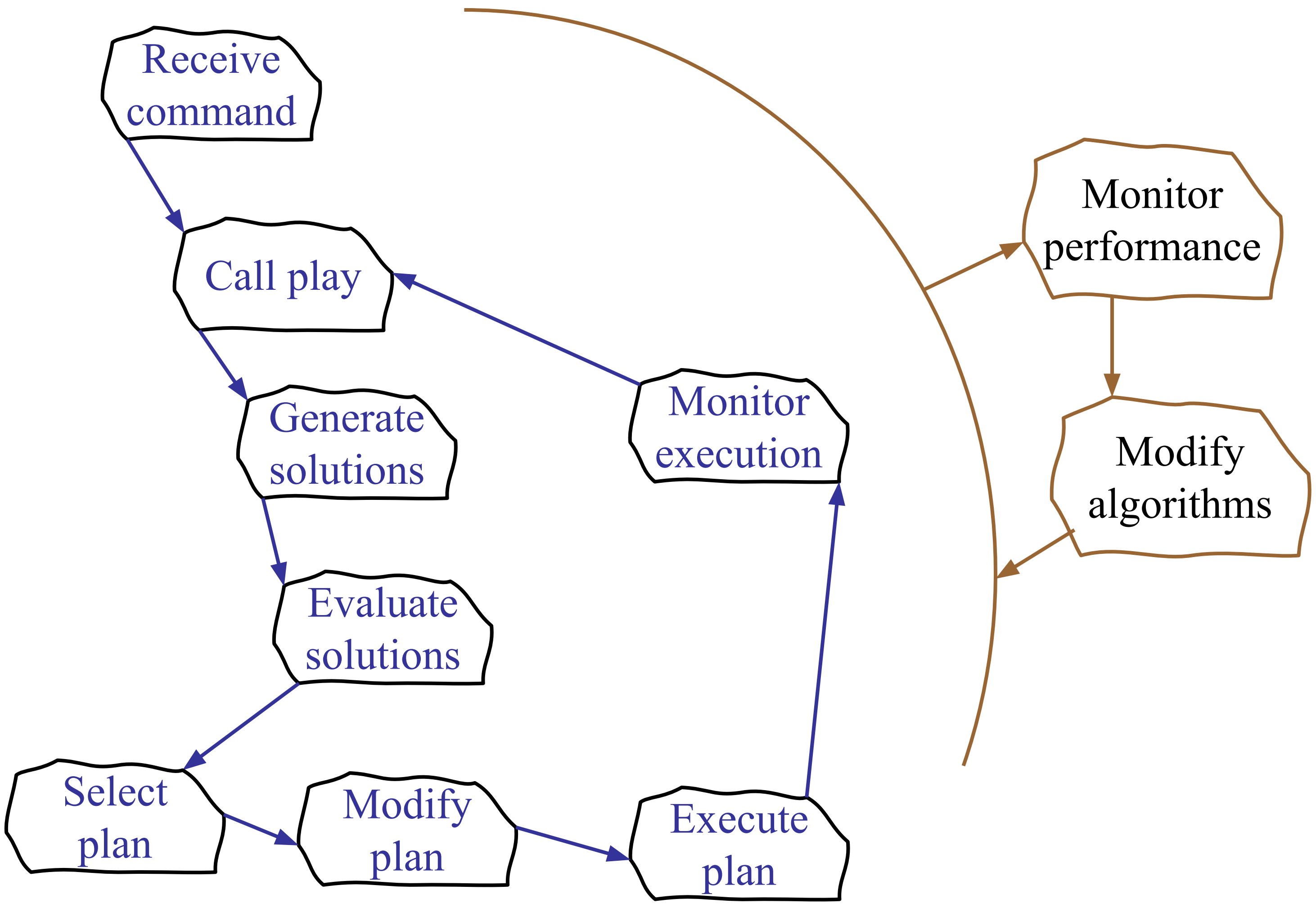
The organization included three personnel, an intelligent agent, and an algorithm.



This conceptual model describes the key activities in the decision-making system.



What additional activities are needed to control system performance?



Panel Members

Dr. Douglas C. Schmidt is the Cornelius Vanderbilt Professor of Computer Science, Associate Provost for Research Development and Technologies, Co-Chair of the Data Science Institute, and a Senior Researcher at the Institute for Software Integrated Systems, all at Vanderbilt University.

Dr. Mikael Lindvall Dr. Mikael Lindvall, is the Technology Director of Fraunhofer USA, Center Mid-Atlantic (CMA) and an adjunct professor at UMCP for more than 10 years. He recently created and delivered one of the first courses on Software Engineering for AI-based systems.

Dr. Jeffrey W. Herrmann is a professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the director of the Reliability Engineering Graduate Program and the director of the Systems Engineering Graduate Program at the University of Maryland.

Dr. Laura Freeman is a Research Associate Professor of Statistics and the Director of the Intelligent Systems Lab at the Virginia Tech Hume Center. Previously, she was the Assistant Director of the Operational Evaluation Division at the Institute for Defense Analyses.

VIRGINIA TECH
Hume Center for National Security and Technology

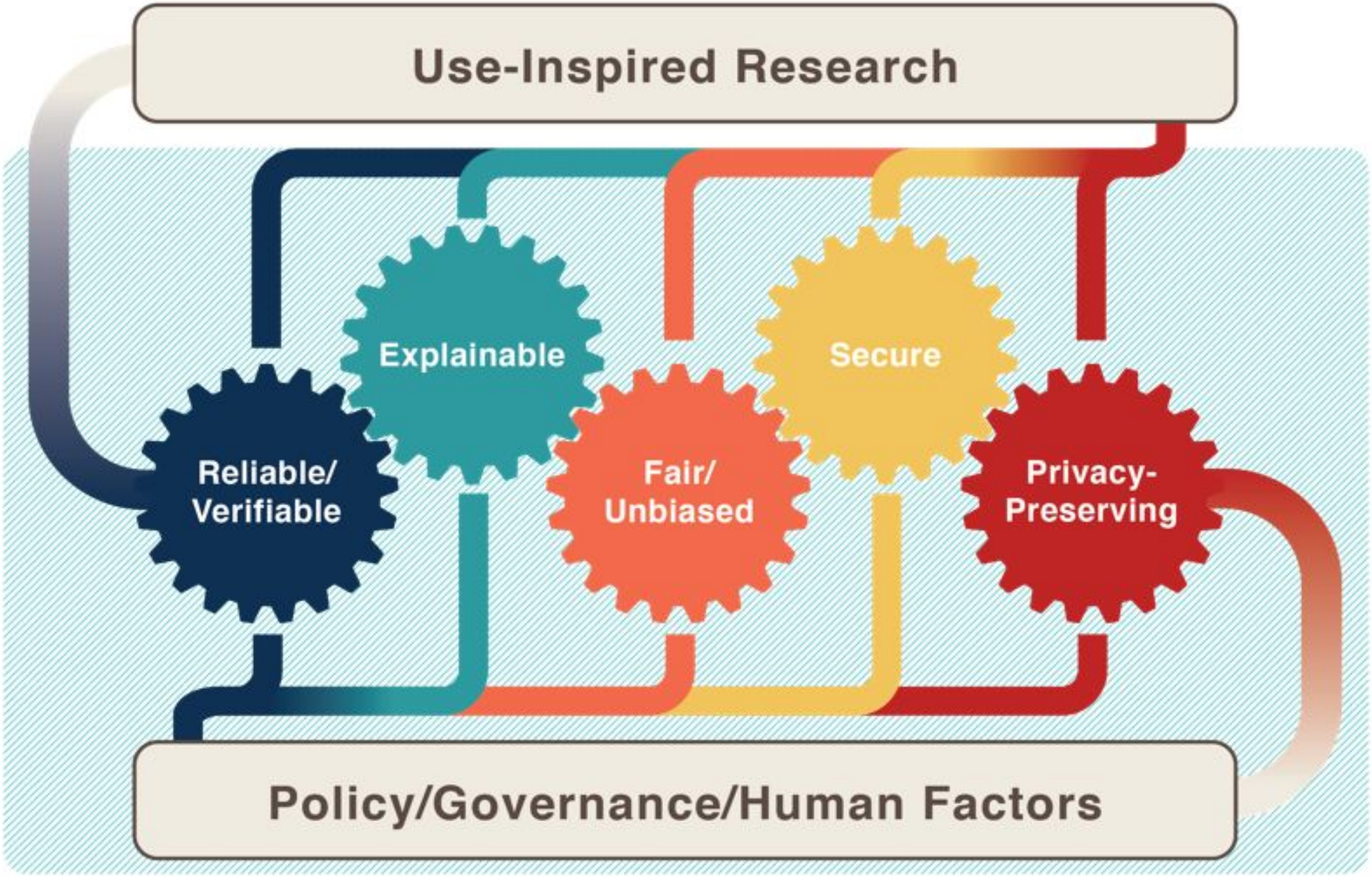
Artificial Intelligence Assurance & Enabling AI Adoption via T&E

Dr. Laura Freeman
Assistant Dean for Research, College of Science
Director, Intelligent Systems Lab, Hume Center
Lead, Artificial Intelligence Program, Commonwealth Cyber Initiative
Research Associate Professor, Statistics

hume@vt.edu
www.hume.vt.edu

What is AI Assurance? An evolving definition.

- AI Assurance provides the necessary information to enable AI adoption into mission critical systems.
- The level of confidence that a system leveraging AI algorithms functions only as intended and is free of vulnerabilities throughout the life cycle of the ***system***. This level of confidence stems from all the planned and ***systematic activities*** implemented within the algorithm and system development that provide confidence that a product or service will fulfill requirements for performance.



We need a framework for AI Assurance

- Lots of Data Types
  - Image, Video, Geospatial, Network Traffic, Time Series, etc.
  - Labeled, unlabeled
- Lots of Analysis / Prediction Goals
  - Detect, Classify, Track, Forecast, Optimize, Visualize
- Lots of Algorithm Types
  - Supervised - labeled records with known output variable and algorithm learns how to predict the output variable
  - Unsupervised - no labels are provided and algorithm makes inferences from input data
  - Bagging – model development in parallel followed by some aggregation (typically averaging)
  - Boosting – sequential model development
  - Stacking – parallel model development and combines with a trained meta-model
- What makes a good algorithm depends on:
  - Problem that we are trying to address
  - Type of data
  - Available data (amount, balance, quality, etc.)
  - System complexity
  - Integration with the system (to include the humans)

Test and Evaluation: Framework and Feedback Loop

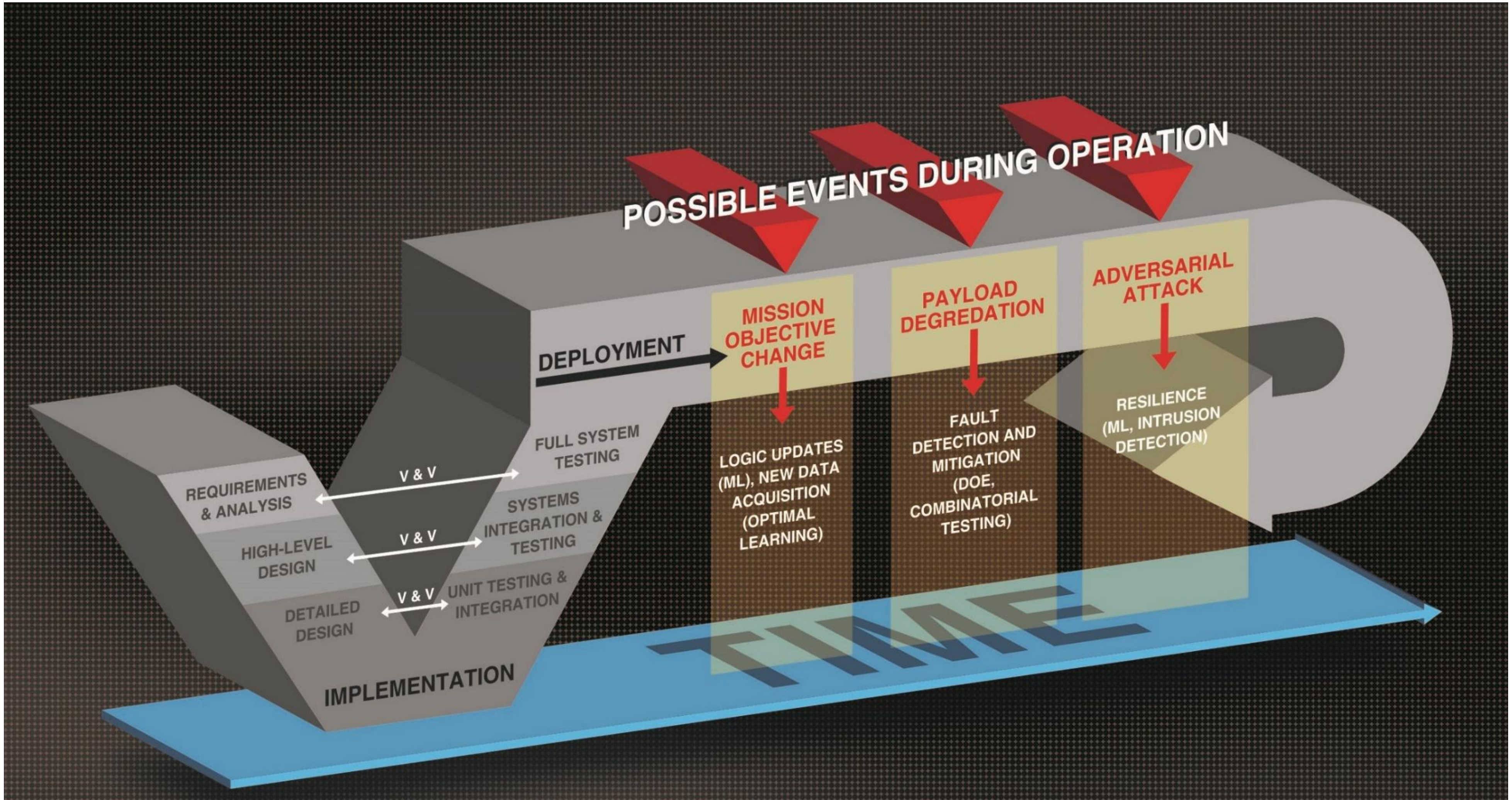
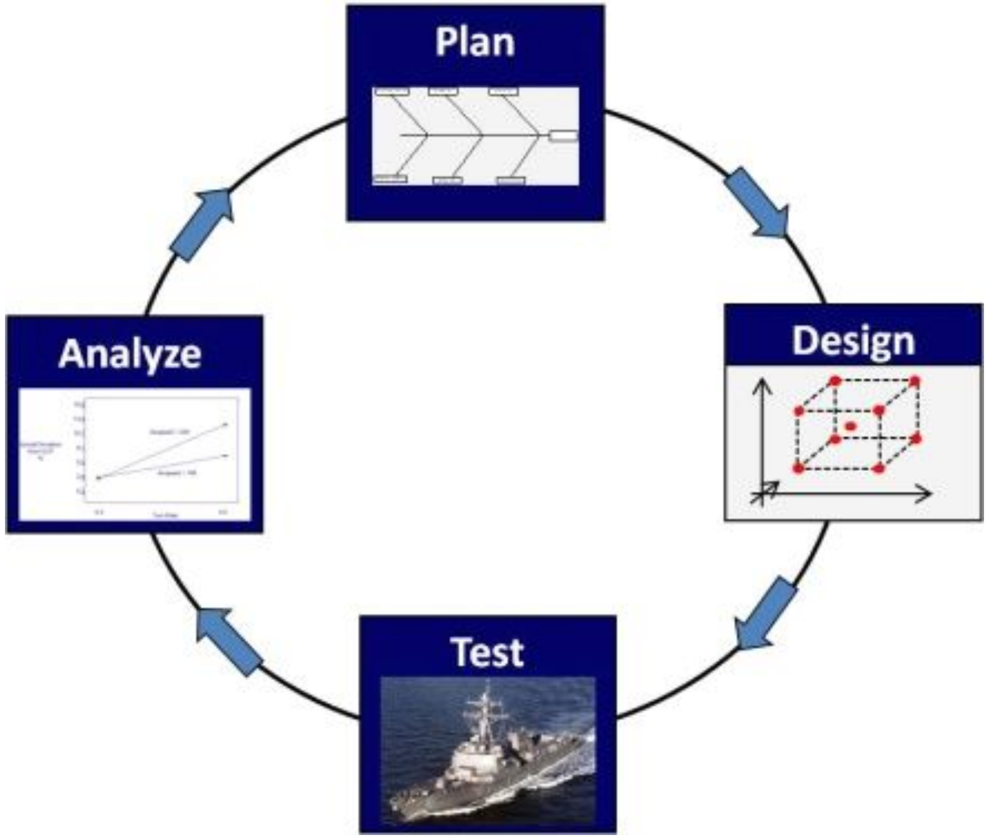


Image Credit: John Matthew Guy, Chief of Multimedia, AFOTEC

Test and Evaluation Process

1. Describe mission, systems, functions
2. Determine what questions must be answered
3. Derive evaluation, performance measures, metrics
4. Determine factors and levels for testing
5. Define testing scenarios and experiments
6. Conduct experiments
7. Analyze data
8. Feedback into future test planning



Research Challenges in AI Assurance

- Problem 1: Defining Operating Envelopes for Artificial Intelligence
- Problem 2: Understanding Models Robustness for Available Data
- Problem 3: “Optimizing” Complex Configurations for Machine Learning Models
- Problem 4: Developing Comprehensive Metrics for AI Assurance to Include AI integration with Human Teams
- Problem 5: Characterizing AI Capabilities in Complex Systems

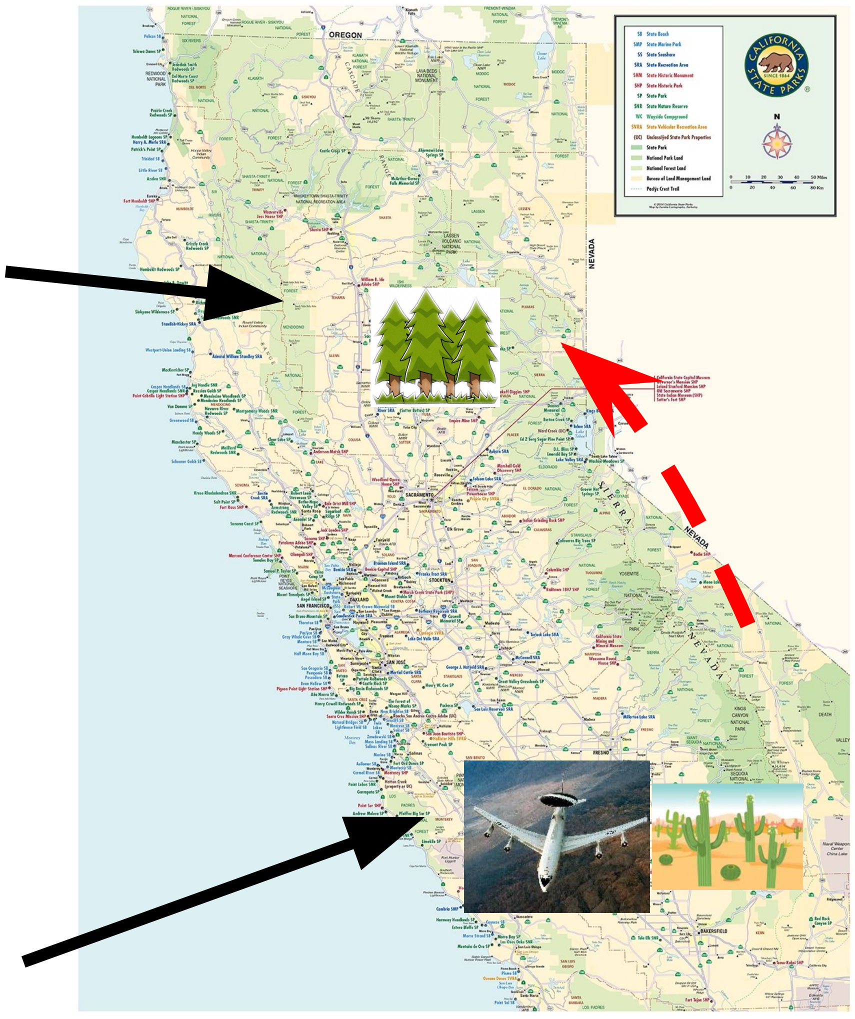
Defining Operating Envelopes: Motivating Problem

Questions

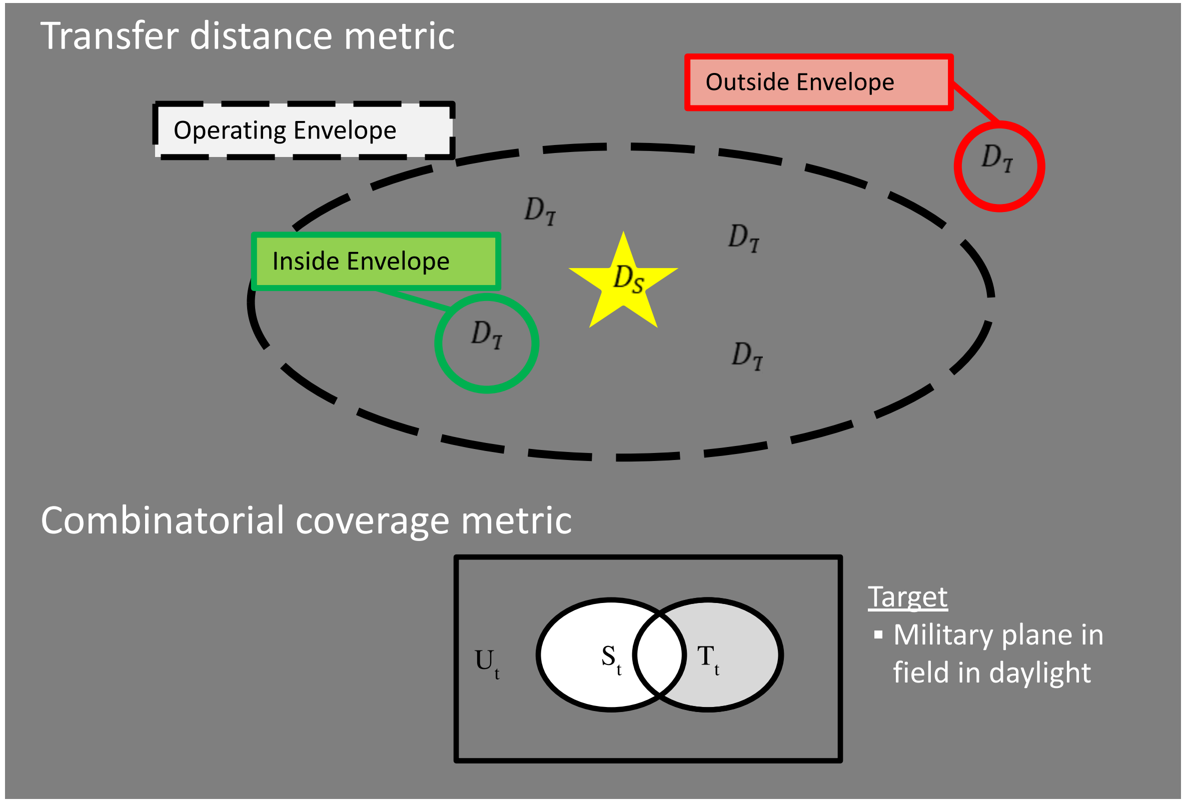
- How can we ensure that a model continues to perform as expected after it is deployed?
- Can we detect a drop in performance and react to it?
- Can we anticipate a drop in performance and prevent it?
- How can we certify a model with quantifiable, reliable guarantees of expected performance?
- How can we accomplish all this efficiently?

Mission to detect in Northern California

Trained to detect planes in Southern California



New Approaches to Defining Operating Envelopes



Define operating envelopes of models allows us to

- Measure distance between source and target data distributions
- Use metadata to measure proportion of target data covered by source data
- Smartly select training, validation, and test splits

Plan for incorporating operating envelopes into an efficient, automated process to ensure certification

1. Search model zoo for model with sufficient predicted performance based on metrics
2. Create ensemble of models from model zoo with sufficient predicted performance
3. Fine-tune model with more target data
4. Identify unlabeled data to collect based on metrics
5. Identify labeled data to collect based on metrics

Increasing Cost

VIRGINIA TECH

UNIVERSITY of VIRGINIA

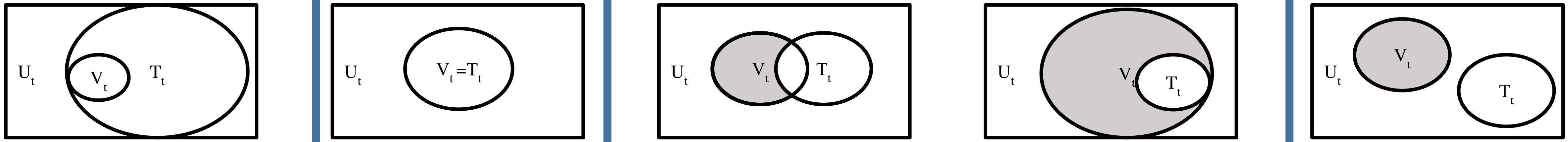
OLD DOMINION UNIVERSITY

Set Difference Combinatorial Coverage Metrics

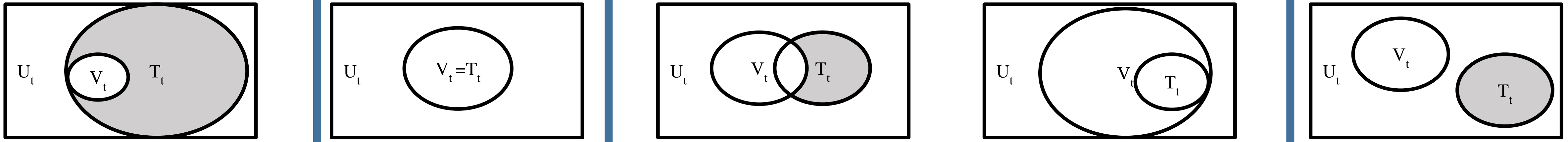
1. Test how well the model performs in contexts it should have learned
   - Prefer SDCCM(Validation \ Test) score close to 0
   - Best when SDCCM(Test \ Validation) also low
     - Suggests representative of same population
     - Avoid confusion in testing contexts might not generalize well
2. Test how well the model generalizes to contexts it hasn't seen
   - Prefer SDCCM(Test \ Validation) close to 1

$$SDCCM_t(V_t \setminus T_t) = \frac{|V_t \setminus T_t|}{|V_t|}$$

$$SDCCM_t(T_t \setminus V_t) = \frac{|T_t \setminus V_t|}{|T_t|}$$



1) $SDCCM_t(V_t \setminus T_t) = 0$ 2) $SDCCM_t(V_t \setminus T_t) = 0$ 3) $0 < SDCCM_t(V_t \setminus T_t) < 1$ 4) $0 < SDCCM_t(V_t \setminus T_t) < 1$ 5) $SDCCM_t(V_t \setminus T_t) = 1$



1) $0 < SDCCM_t(T_t \setminus V_t) < 1$ 2) $SDCCM_t(T_t \setminus V_t) = 0$ 3) $0 < SDCCM_t(T_t \setminus V_t) < 1$ 4) $SDCCM_t(T_t \setminus V_t) = 0$ 5) $SDCCM_t(T_t \setminus V_t) = 1$

Q & A

Adam Porter
aporter@fraunhofer.org