Enabling Confidence in AI

BRYAN JACOBS

DARPA MTO Program manager

Bryan.Jacobs@DARPA.mil

Bryan.Jacobs@DARPA.IC.gov

Bryan
Jacobs

Dr. Bryan Jacobs joined the DARPA Microsystems Technology Office in May of 2020, where he manages a research portfolio focused on realizing transformational change from alternative models of computation. Individual programs address hardware accelerators to enable computing on encrypted data (DPRIVE), quantum inspired computing to drastically lower the energy cost of optimization (QuICC), and uncertainty estimation in machine learning systems to radically increase the effectiveness of multi-sensor fusion applications at the edge (Enabling Confidence).
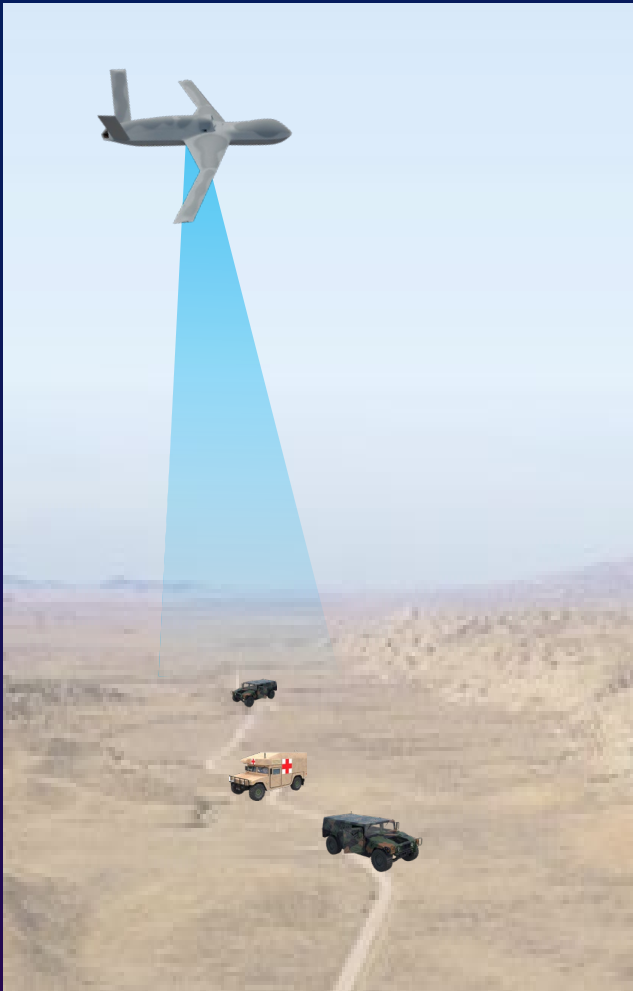
Vision

Displace conventional and quantum computing architectures by developing analog and physics-based processing hardware concepts that are several orders of magnitude more energy efficient for challenging optimization and neural networking applications.
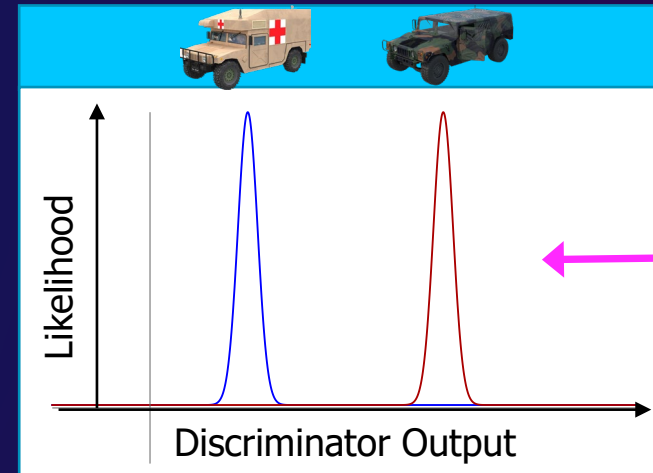
Mission

Translate the lessons learned from the past two decades of research in quantum information processing into real-word advantages for computing today.
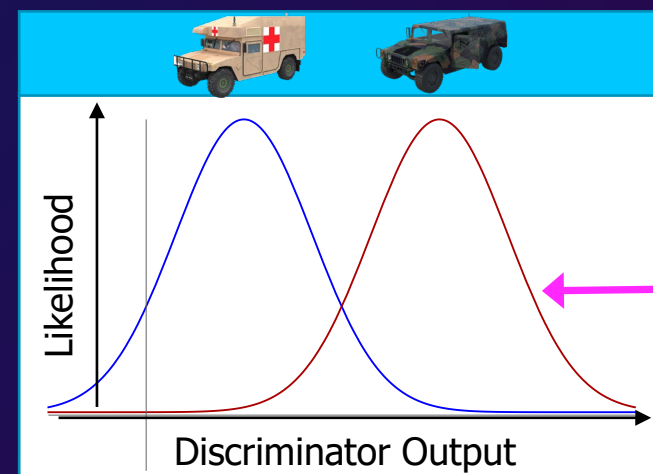
The need for confidence in AI

Measures of uncertainty are critical for decision making – especially at the edge

Sensor Processing
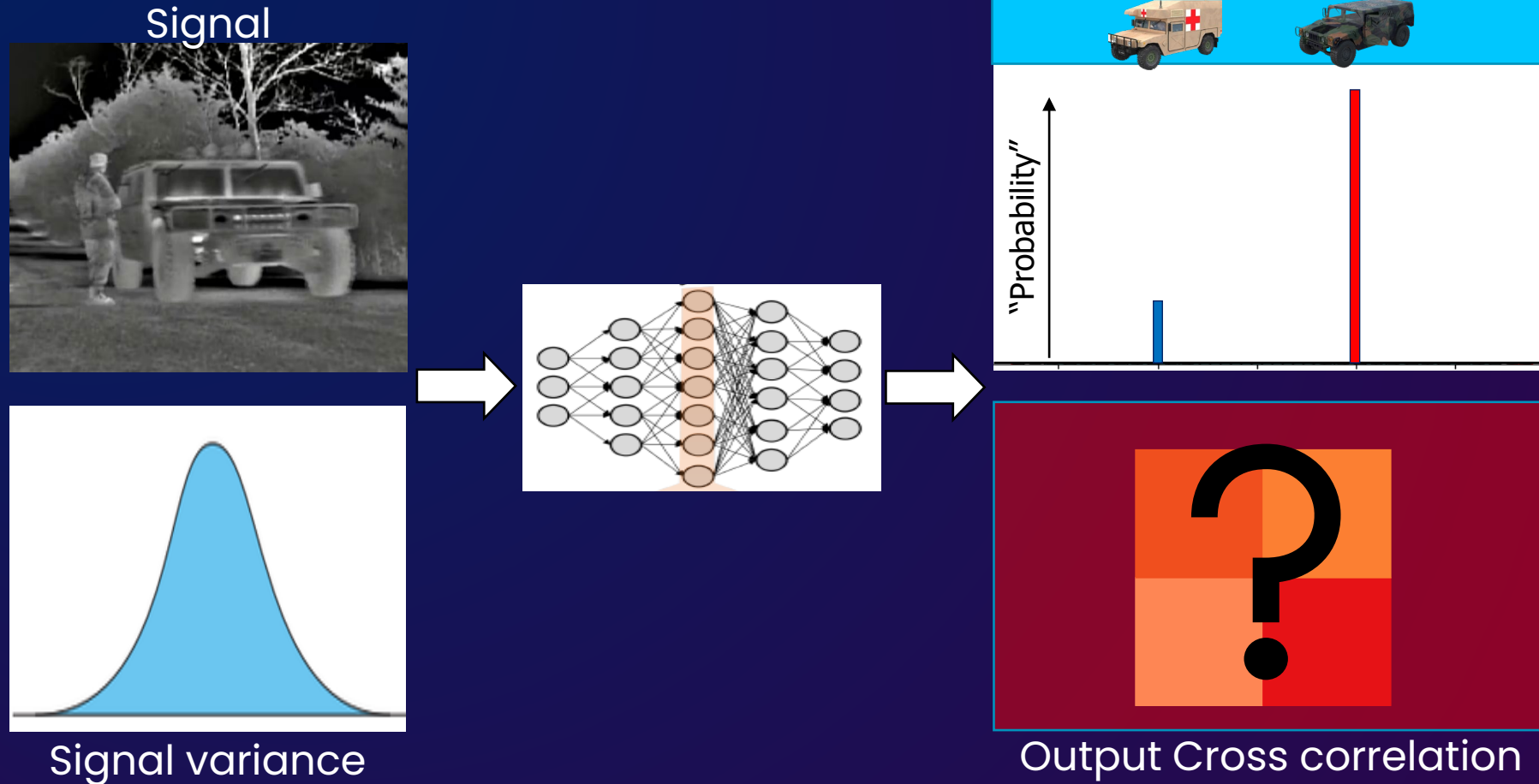
Easy to make mission critical decisions

Difficult to make mission critical decisions

Modern warfare depends on accurate discrimination of targets

The problem with conventional machine learning (ML)

ML output "probabilities" are notional – not tied to underlying distribution

Signal

Signal variance
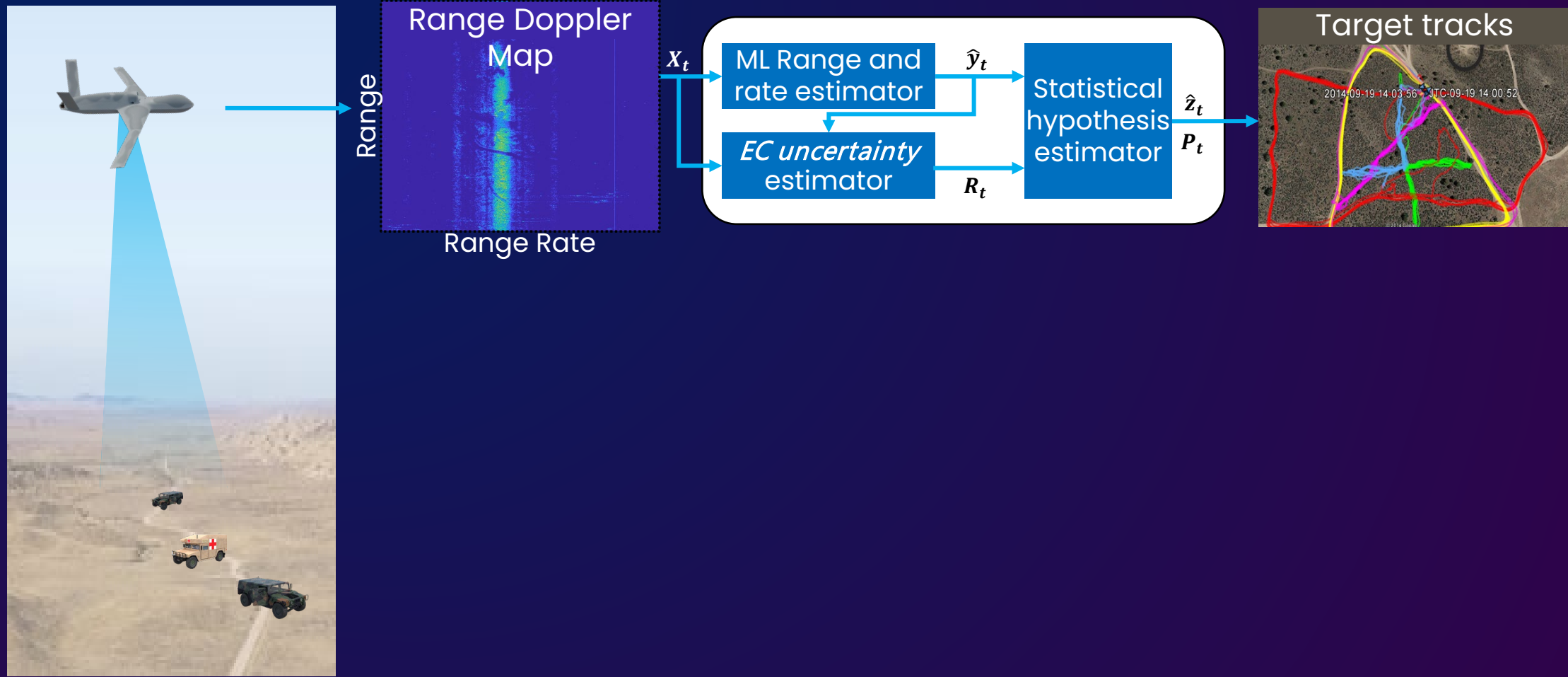
"Probability"

Output Cross correlation

Without accurate output uncertainty distributions, ML results are suspect

Example approach from Enabling Confidence program

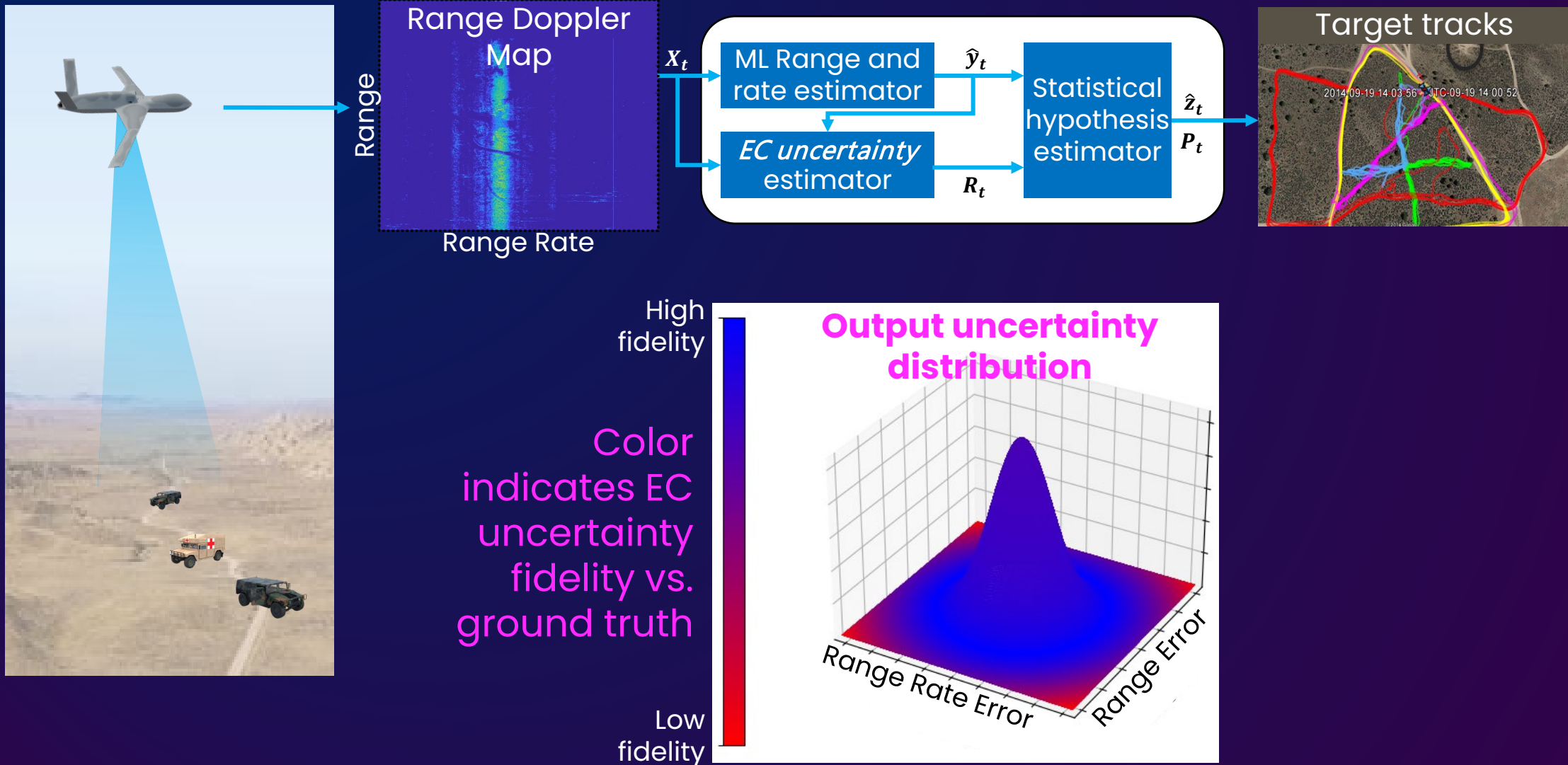## Science & Technology Research (STR) multi-object detection and tracking scenario

Range Doppler Map

Range

Range Rate

X_t

ML Range and rate estimator → $\hat{y}_t$

EC uncertainty estimator → R_t

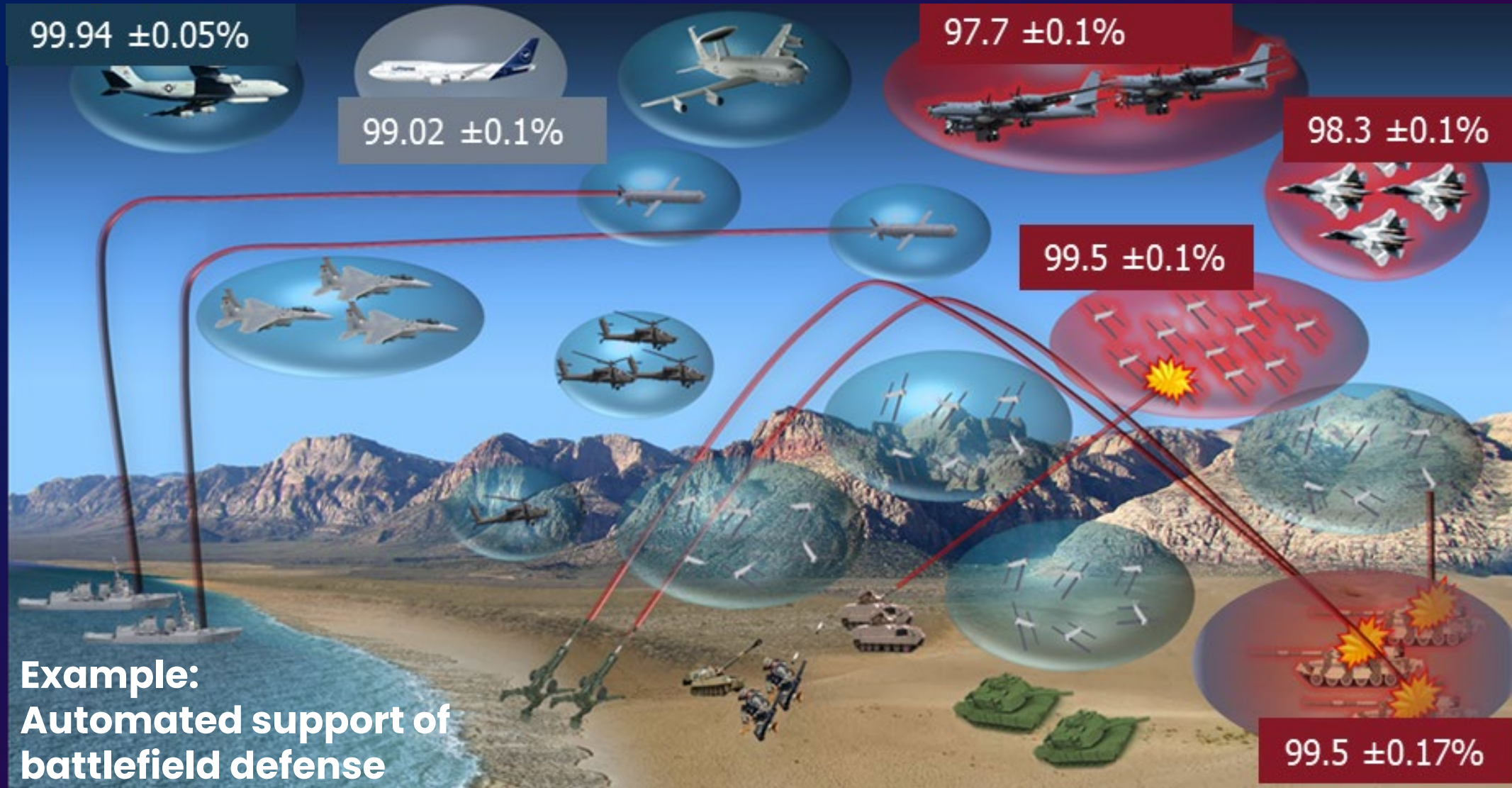Statistical hypothesis estimator → $\hat{z}_t$ P_t

Target tracks

Example approach from Enabling Confidence program

Science & Technology Research (STR) multi-object detection and tracking scenario

Range Doppler Map

Range

Range Rate

X_t

ML Range and rate estimator

$\hat{y}_t$

EC uncertainty estimator

R_t

Statistical hypothesis estimator

$\hat{z}_t$
P_t

Target tracks

2014-09-19 14 03 56   JTC-09-19 14 00 52

Color indicates EC uncertainty fidelity vs. ground truth

High fidelity

Low fidelity

Output uncertainty distribution

Range Rate Error

Range Error

Statistical confidence in AI will enable disruptive capabilities

99.94 ±0.05%

99.02 ±0.1%

97.7 ±0.1%

98.3 ±0.1%

99.5 ±0.1%

99.5 ±0.17%

Example: Automated support of battlefield defense

Could move human oversight of tactical defense systems to a higher level

THANK YOU