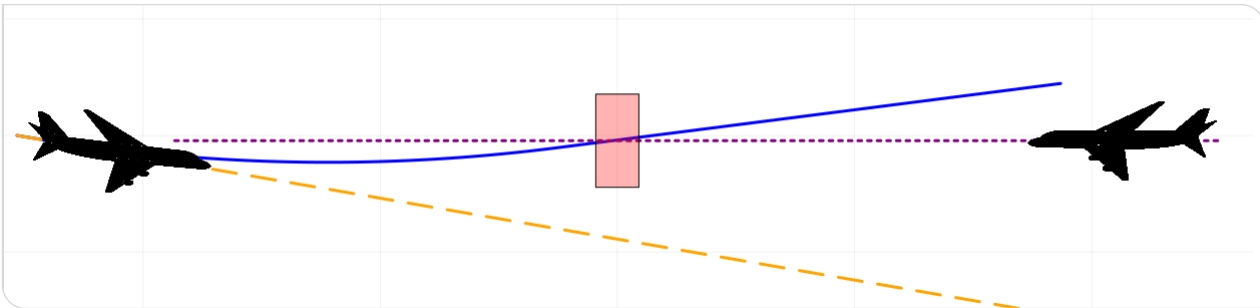


# Provably Safe Neural Network Controllers via Differential Dynamic Logic

NeurIPS 2024

Samuel Teuber, Stefan Mitsch, André Platzer | 2024



## Differential Dynamic Logic

- Program Logic to reason about Cyber-Physical Systems
- Proves **infinite time-horizon** safety for **control envelopes**

# Background

## Differential Dynamic Logic

- Program Logic to reason about Cyber-Physical Systems
- Proves **infinite time-horizon** safety for **control envelopes**

## Open-Loop Neural Network Verification

- Analysis of Neural Network **in isolation**
- Usually supports **linear constraints** with very **simple structure**

# Objective

## Given:

- A **safe** differential dynamic logic model of the system with a **control envelope**  $\alpha_{\text{ctrl}}$
- A **neural network controller**  $g$

# Objective

## Given:

- A **safe** differential dynamic logic model of the system with a **control envelope**  $\alpha_{\text{ctrl}}$
- A **neural network controller**  $g$

## Question:

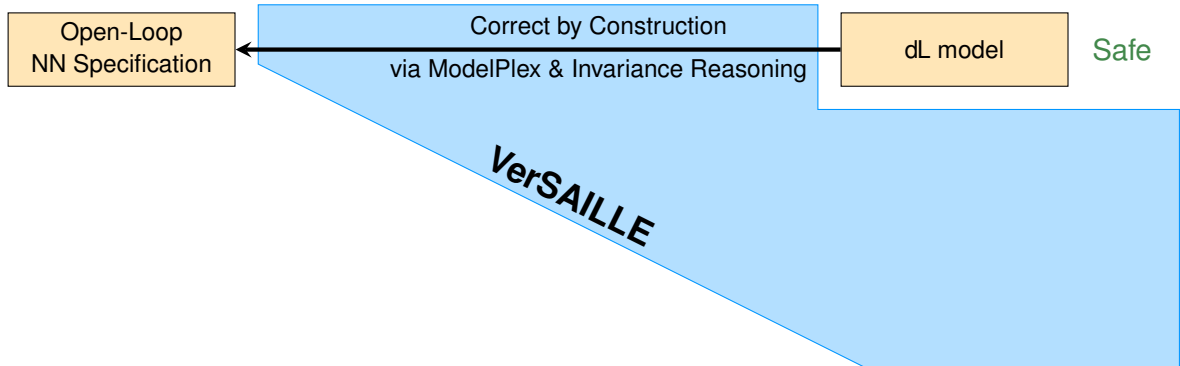
If we **replace** the control envelope  $\alpha_{\text{ctrl}}$  by the NN  $g$ ,  
does the resulting system retain the same safety guarantees?

# Provably Safe Neural Network Controllers

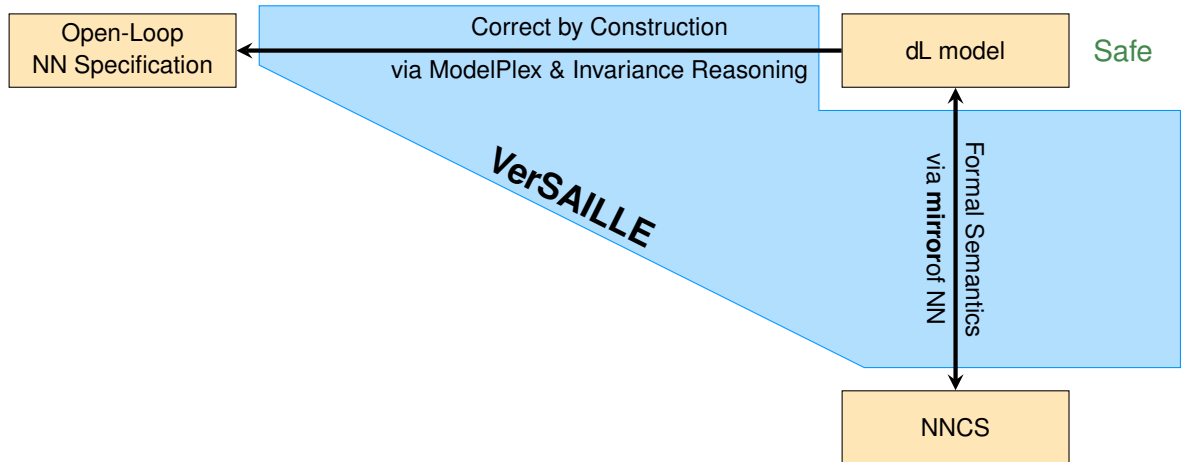
dL model

Safe

# Provably Safe Neural Network Controllers

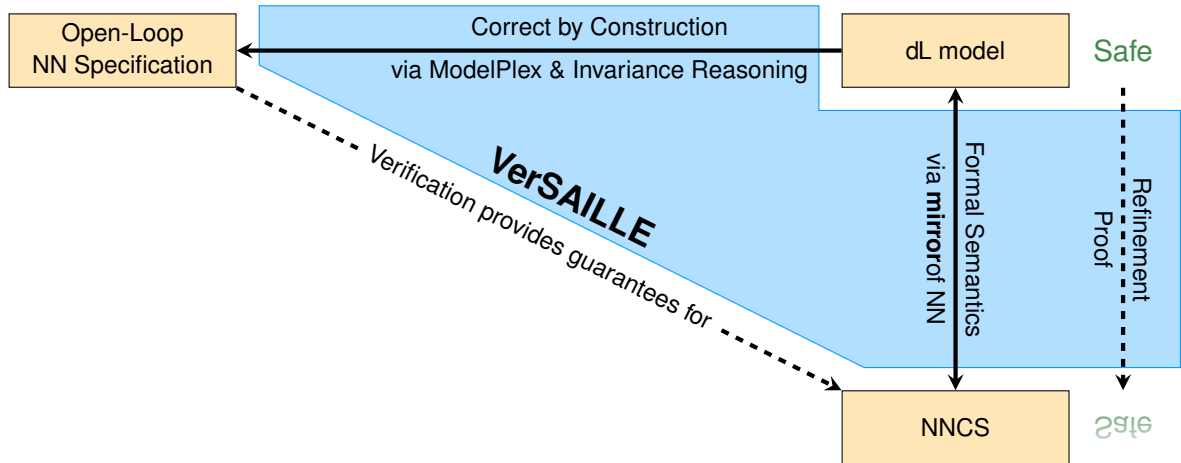


# Provably Safe Neural Network Controllers

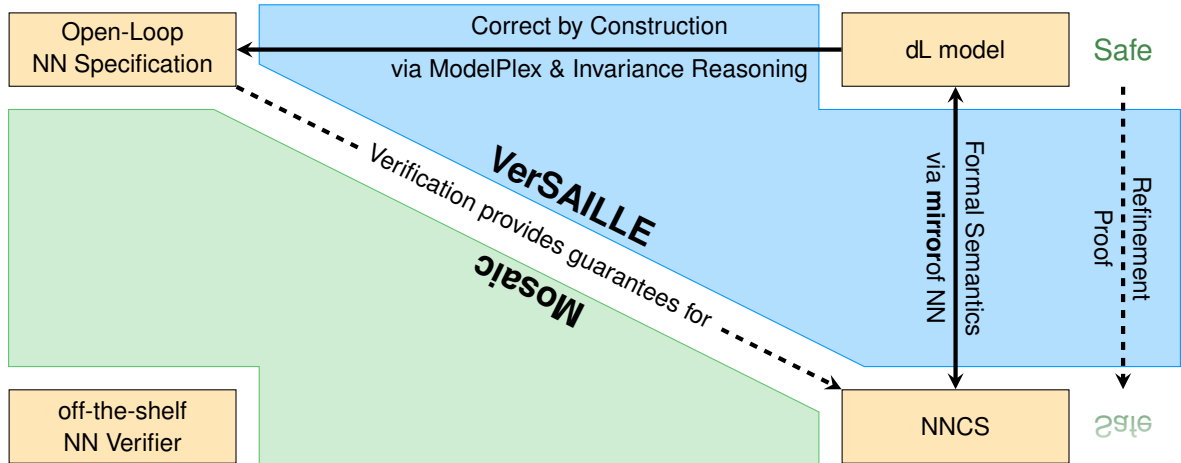




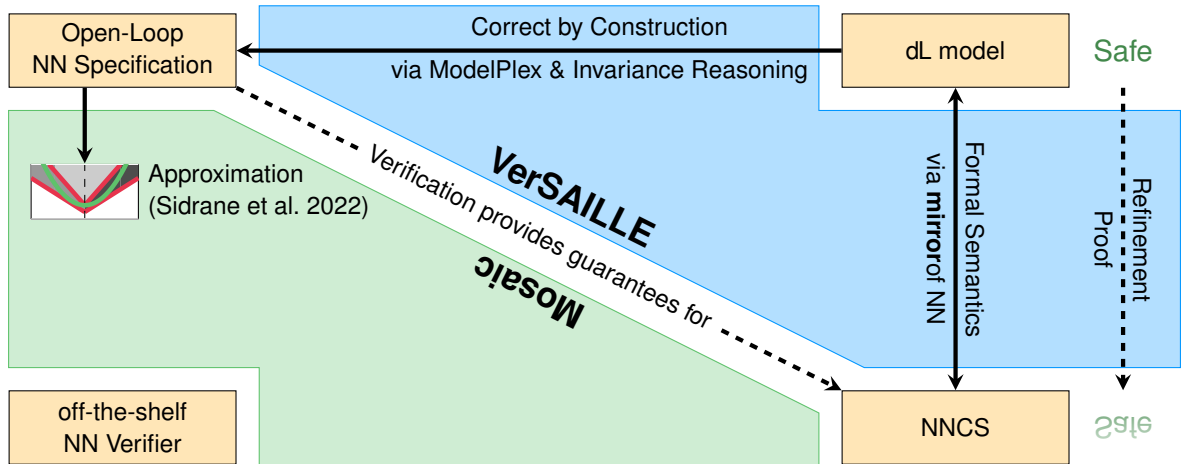
# Provably Safe Neural Network Controllers



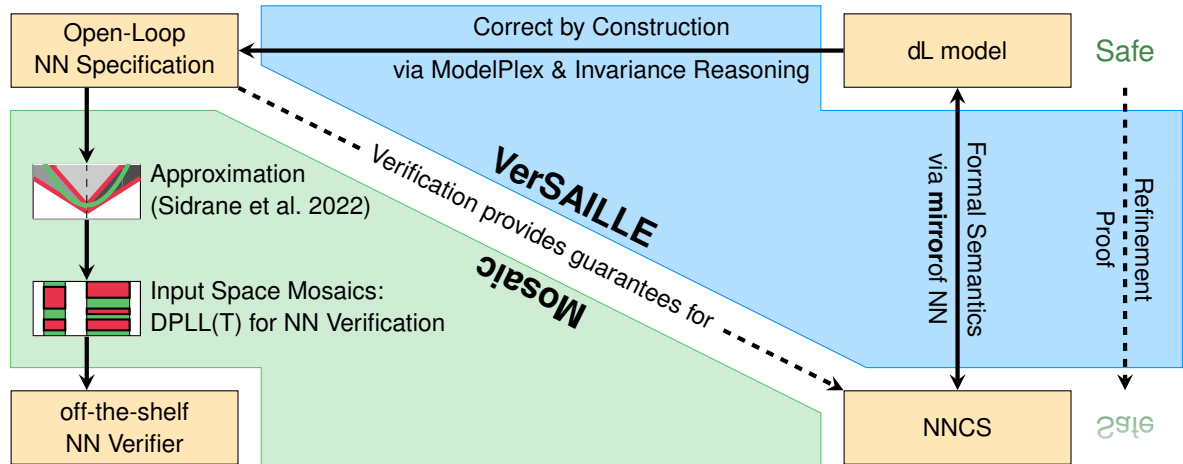
# Provably Safe Neural Network Controllers



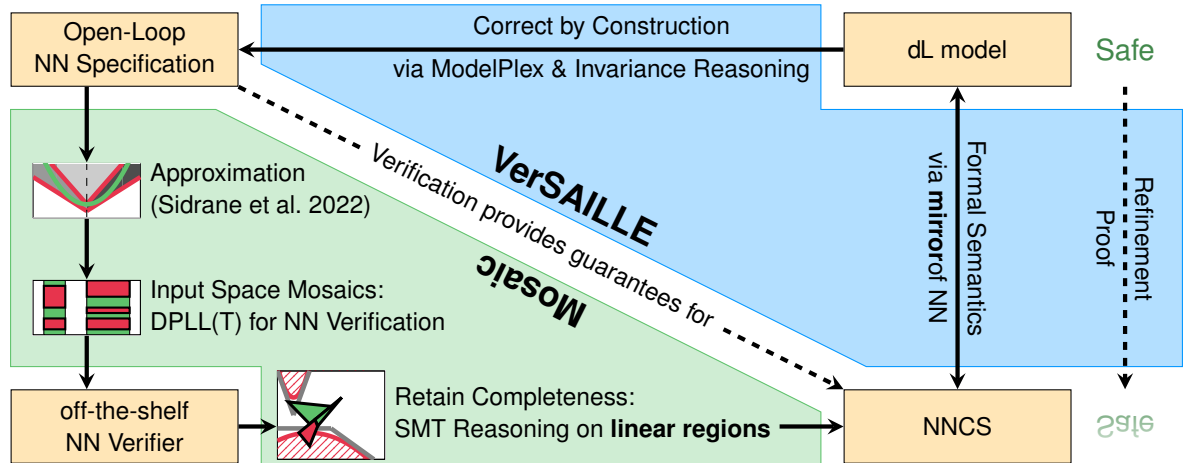
# Provably Safe Neural Network Controllers



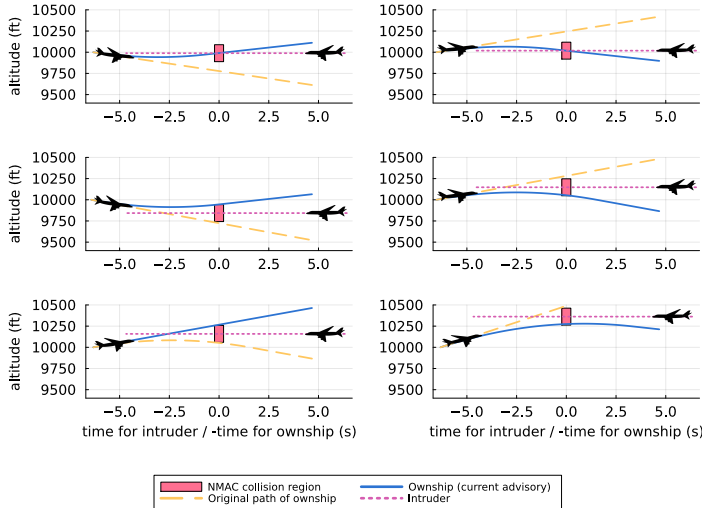
# Provably Safe Neural Network Controllers



# Provably Safe Neural Network Controllers



# Analysis of Vertical Airborne Collision Avoidance



Networks **and** dL formalization **from prior literature**  
 (Julian and Kochenderfer 2019; Jeanin et al. 2017)

## Result

- 6 out of 8 NNs **unsafe**
- Other NNs safe for **intruder in level flight**  
 (but found crashes for non-level flight intruder)

# References I

- [1] Jean Baptiste Jeannin et al. “A formally verified hybrid system for safe advisories in the next-generation airborne collision avoidance system”. In: *International Journal on Software Tools for Technology Transfer* 19.6 (2017), pp. 717–741. ISSN: 14332787. DOI: 10.1007/s10009-016-0434-1. URL: <https://doi.org/10.1007/s10009-016-0434-1>.
- [2] Kyle D. Julian and Mykel J. Kochenderfer. “Guaranteeing Safety for Neural Network-Based Aircraft Collision Avoidance Systems”. In: *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*. 2019, pp. 1–10. DOI: 10.1109/DASC43569.2019.9081748.
- [3] Chelsea Sidrane et al. “OVERT: An algorithm for safety verification of neural network control policies for nonlinear systems”. In: *Journal of Machine Learning Research* 23.117 (2022), pp. 1–45.