

---

# Verification of Neural-Network Control Systems by Integrating Taylor Models and Zonotopes

---

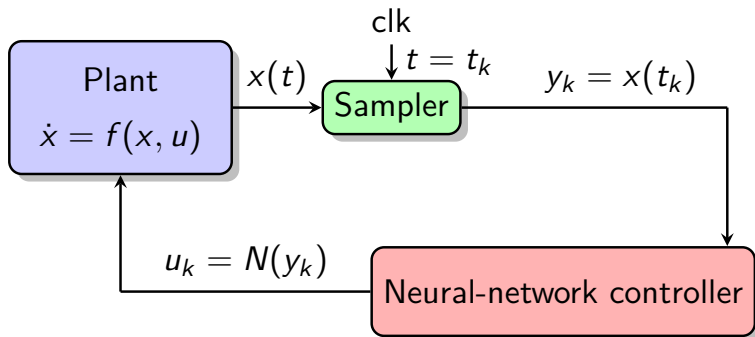
**Christian Schilling**<sup>1</sup> , Marcelo Forets<sup>2</sup> , Sebastián Guadalupe<sup>2</sup>

<sup>1</sup> Aalborg University, Denmark

<sup>2</sup> Universidad de la República, Uruguay

AAAI 2022

## Neural-network control system



# Unicycle model

## Plant:

$$\dot{x} = v \cos(\theta)$$

$$\dot{y} = v \sin(\theta)$$

$$\dot{\theta} = u_2$$

$$\dot{v} = u_1 + w$$

## Specification:

$$x(0) \in \mathcal{X}_0$$

$$x(10) \in \mathcal{X}_T$$

## Controller:

500 hidden units

0.2 s period

# Unicycle model

## Plant:

$$\dot{x} = v \cos(\theta)$$

$$\dot{y} = v \sin(\theta)$$

$$\dot{\theta} = u_2$$

$$\dot{v} = u_1 + w$$

## Specification:

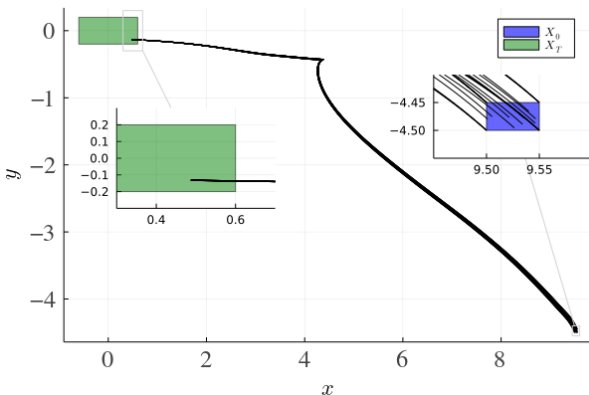
$$x(0) \in \mathcal{X}_0$$

$$x(10) \overset{!}{\in} \mathcal{X}_T$$

## Controller:

500 hidden units

0.2 s period



42 simulations

# Unicycle model

## Plant:

$$\dot{x} = v \cos(\theta)$$

$$\dot{y} = v \sin(\theta)$$

$$\dot{\theta} = u_2$$

$$\dot{v} = u_1 + w$$

## Specification:

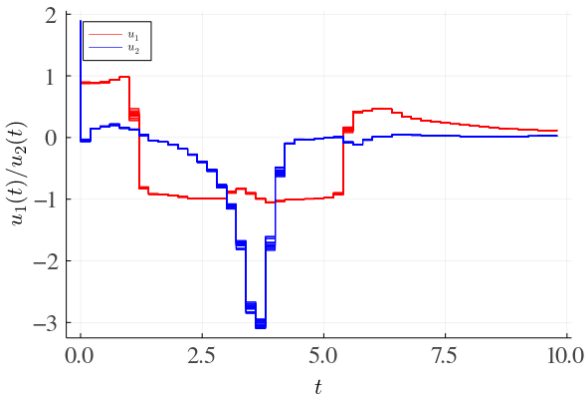
$$x(0) \in \mathcal{X}_0$$

$$x(10) \in \mathcal{X}_T$$

## Controller:

500 hidden units

0.2 s period



control signals (42 simulations)

# Unicycle model

## Plant:

$$\dot{x} = v \cos(\theta)$$

$$\dot{y} = v \sin(\theta)$$

$$\dot{\theta} = u_2$$

$$\dot{v} = u_1 + w$$

## Specification:

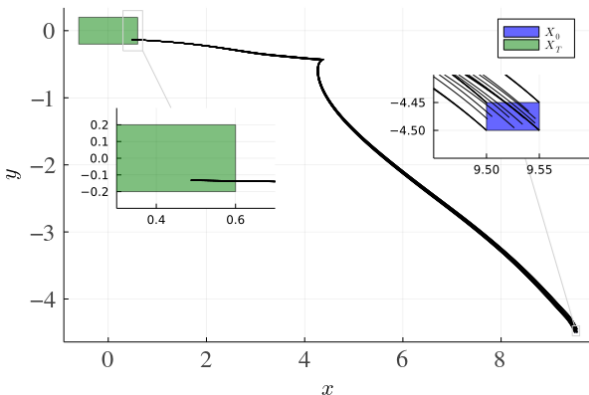
$$x(0) \in \mathcal{X}_0$$

$$x(10) \in \mathcal{X}_T$$

## Controller:

500 hidden units

0.2 s period



# Unicycle model

## Plant:

$$\dot{x} = v \cos(\theta)$$

$$\dot{y} = v \sin(\theta)$$

$$\dot{\theta} = u_2$$

$$\dot{v} = u_1 + w$$

## Specification:

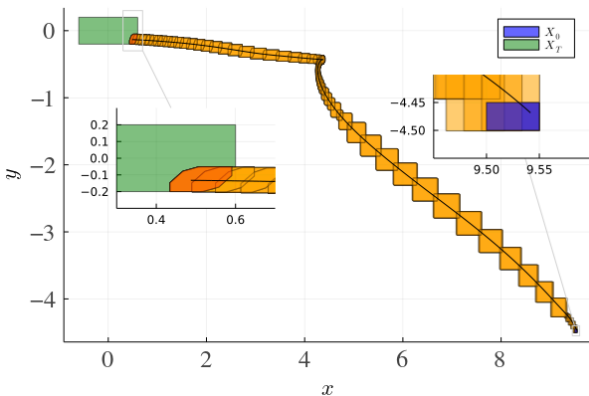
$$x(0) \in \mathcal{X}_0$$

$$x(10) \overset{!}{\in} \mathcal{X}_T$$

## Controller:

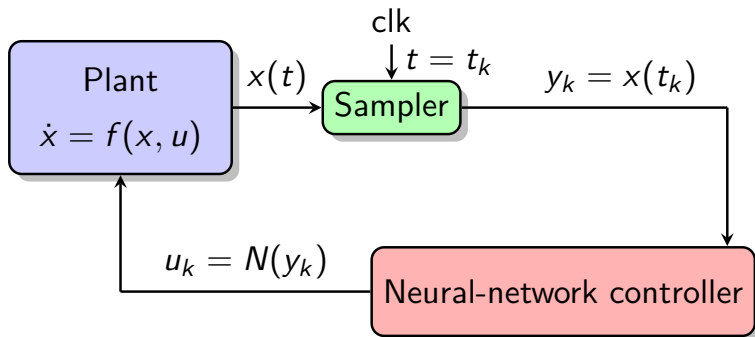
500 hidden units

0.2 s period



set-based simulations/reachability analysis

## Neural-network control system





# Set-based simulation: algorithms and tools used

- **Plant**
  - **Taylor-model** techniques<sup>1</sup>
  - We use an existing implementation<sup>2</sup>
- **Neural network**
  - **Zonotope** techniques<sup>3</sup> (ReLU/sigmoid/tanh activation)
- **Neural-network control system**
  - Combine **Taylor models** and **zonotopes**
  - Implementation in **JuliaReach**<sup>4</sup>

---

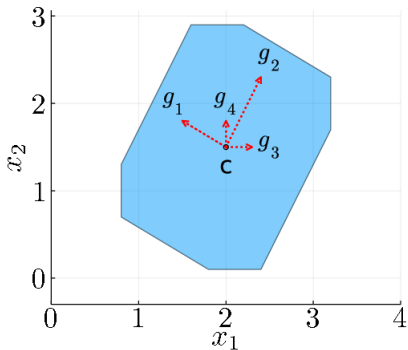
<sup>1</sup>K. Makino and M. Berz. *Int. J. Pure Appl. Math* (2003).

<sup>2</sup>L. Benet and D. P. Sanders. *J. Open Source Softw.* (2019).

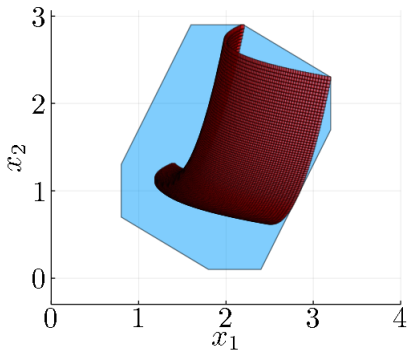
<sup>3</sup>G. Singh et al. *NeurIPS*. 2018.

<sup>4</sup>S. Bogomolov et al. *HSCC*. 2019.

## Taylor models and structured zonotopes



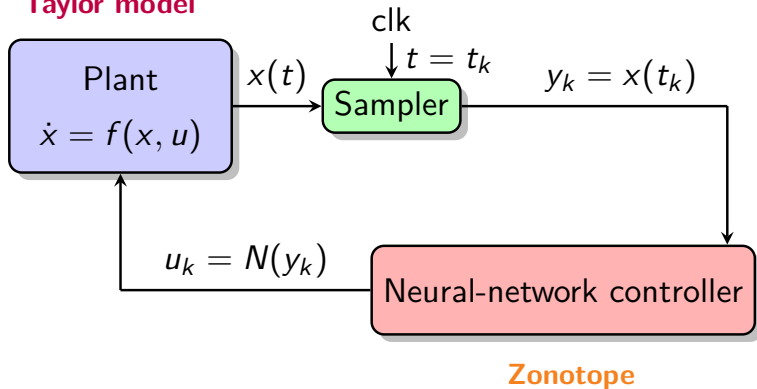
**structured zonotope**



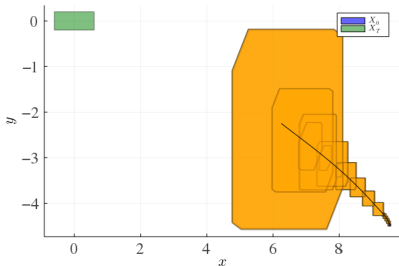
**Taylor model** (red) enclosed by  
**structured zonotope** (blue)

## Trivial problem?

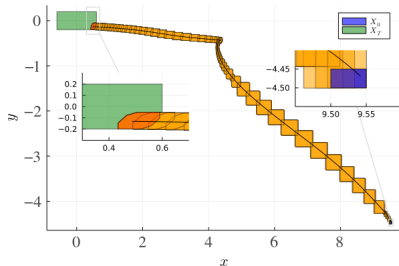
### Taylor model



# Trivial problem?



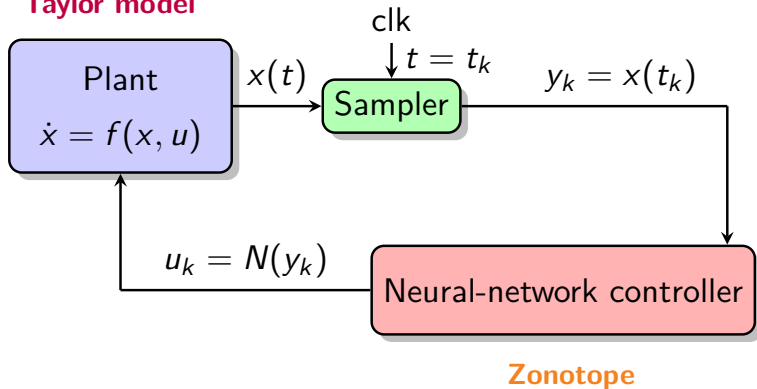
naive combination



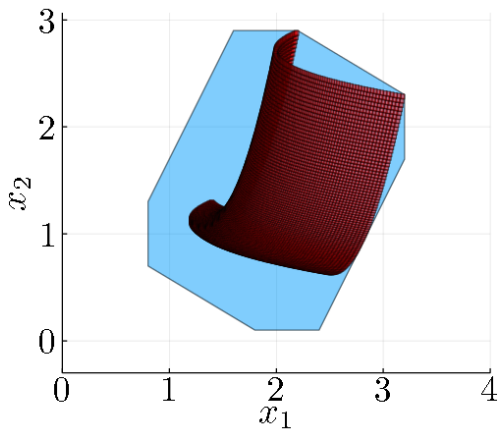
careful combination

## Trivial problem?

### Taylor model

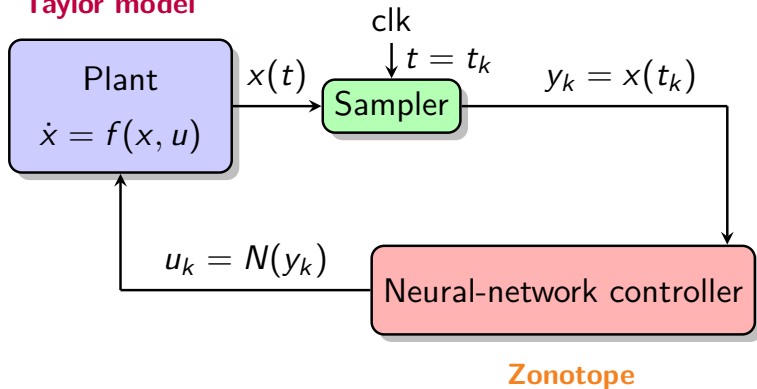


Trivial problem?



## Trivial problem?

### Taylor model



# Evaluation

- Six benchmark models from **ARCH-COMP**<sup>1</sup>
  - **JuliaReach** solved all problems for first time
- Comparison with **Sherlock**<sup>2</sup>
  - Taylor-model techniques for plant
  - Synthesis of a quadratic approximation for network
  - No set conversion (Taylor model all the time)

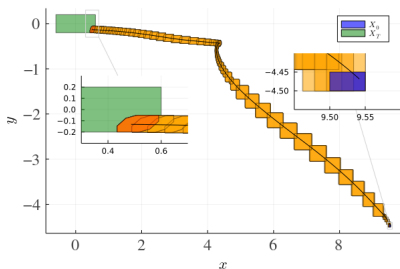
---

<sup>1</sup>T. T. Johnson et al. *ARCH*. 2021.

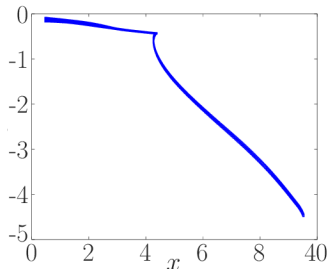
<sup>2</sup>S. Dutta, X. Chen, and S. Sankaranarayanan. *HSCC*. 2019.



# Unicycle model



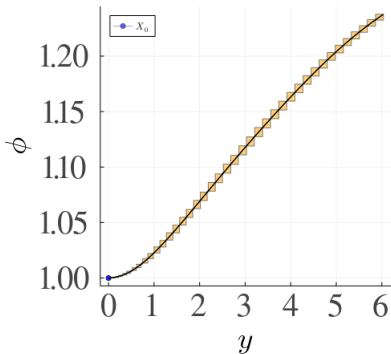
**JuliaReach**, 93 seconds



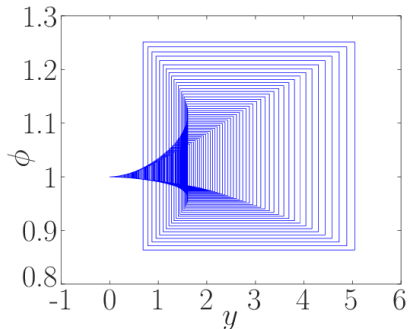
**Sherlock**, 526 seconds

- 4 state dimensions, 2 control dimensions
- Comparable precision

## Airplane model



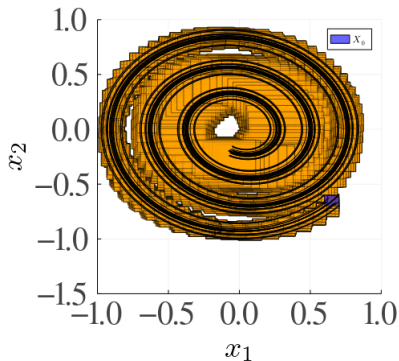
**JuliaReach**, 29 seconds



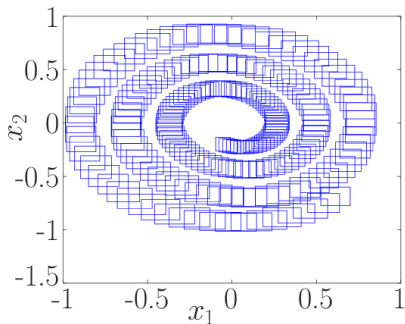
**Sherlock**, 169 seconds  
(terminated)

- 12 state dimensions, 6 control dimensions
- **Sherlock** diverges

## TORA model



**JuliaReach**, 2040 seconds



**Sherlock**, 30 seconds

- 4 state dimensions, 1 control dimension
- **JuliaReach** needs to split the initial states

## Conclusion

- Reachability approach for **neural-network control systems**
- Challenge: repeated **set conversion** incurs precision loss
- Mitigated by **Taylor models** and **structured zonotopes**
- Works well on many models
- Publicly available implementation in **JuliaReach**
- Experimental setup for repeatability evaluation:  
[https://github.com/JuliaReach/AAAI22\\_RE](https://github.com/JuliaReach/AAAI22_RE)