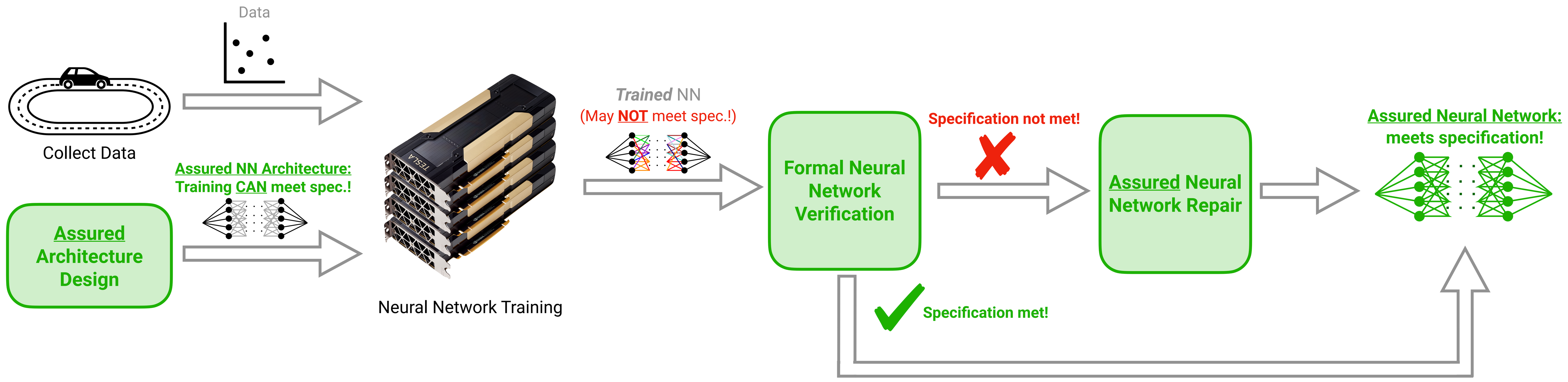
Assured Autonomy for Systems with Neural Network Components

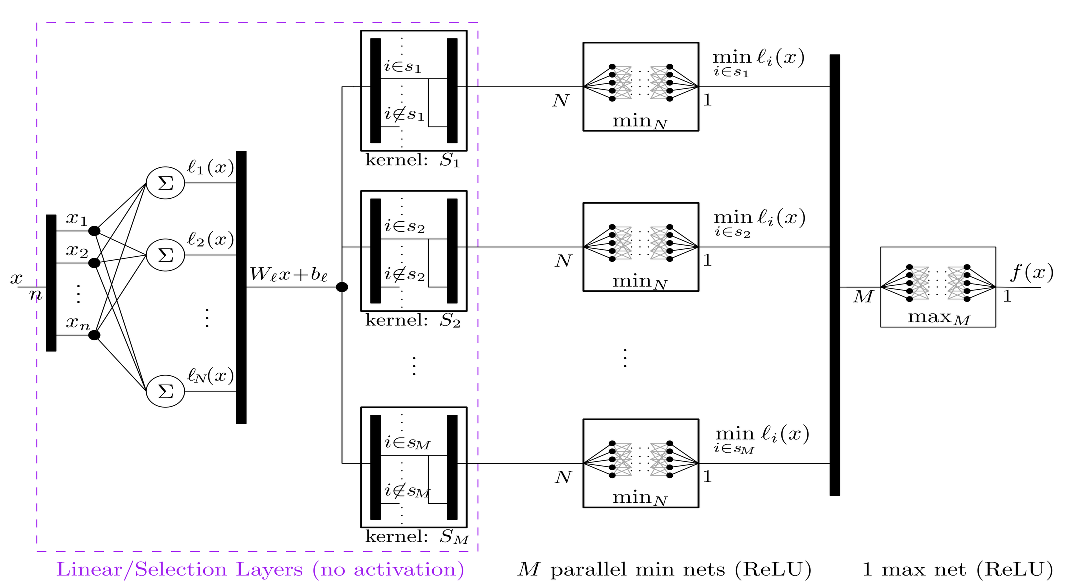
James Ferlez
Postdoctoral Scholar, Resilient Cyber-Physical Systems Lab
University of California, Irvine
jferlez@uci.edu | https://jferlez.github.io

Assured Autonomy Neural Network (NN) Design Pipeline

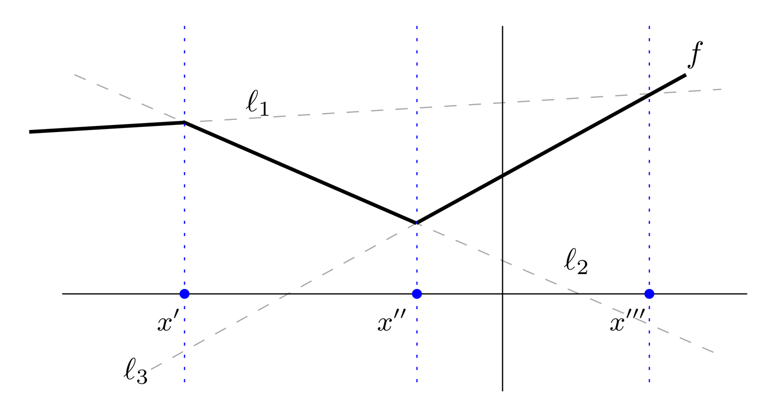
Pipeline of design components to enforce assurances before and after training: **assured architectures**; **formal verification**; and **assured repair**.



Key Idea: *Semantic* NN Architecture — Two-Level Lattice (TLL)

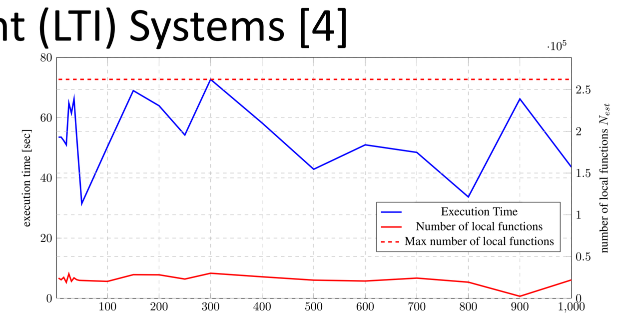


- Based on **Two-Level Lattice (TLL representation** for Continuous, Piecewise Affine (CPWA) functions [10]
- Rectified Linear Unit (ReLU) TLL NNs [4]:
  - Two "levels" of lattice operations: **min** and **max** operations via ReLUs **(all nonlinear neurons in these layers!)**
  - **Local affine functions appear directly as neuron weights** (first layer) [4] (e.g. ℓ_1, ℓ_2 and ℓ_3 in the figure to the right)
  - **Activation region** of local affine functions determined by "selection layer"
- Semantic NN Architecture: **specific neuron weights ⟷ specific properties of NN function**



Assured TLL NN Architectures

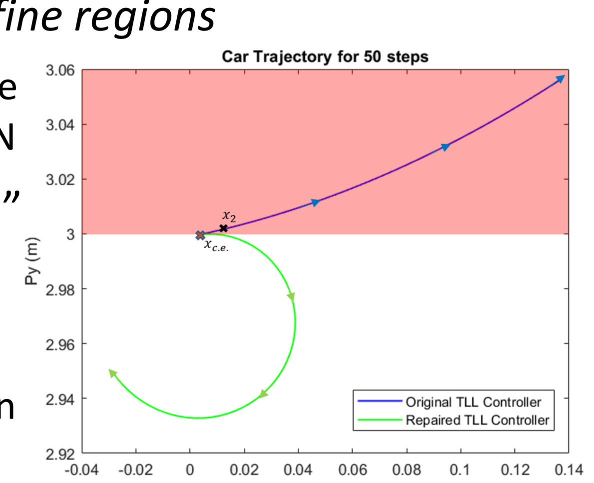
- **Assure that NN training *CAN* be successful**
- Assured NN architectures for Linear-Time Invariant (LTI) Systems [4]
  - Size TLL NN architecture based on Model-Predictive Control
  - Explicit MPC controller **not** required: **fast algorithm** (**assured** architecture in seconds not days like NAS)



- Assured NN architectures for Nonlinear Systems [6,8]
  - Assure more general specifications, too: **bisimulation**
  - **Abstract Disturbance Simulation** → unify/extend robust and disturbance bisimulation
  - **Algorithmic translation of (known) Lipschitz-continuous controller to TLL** [6]

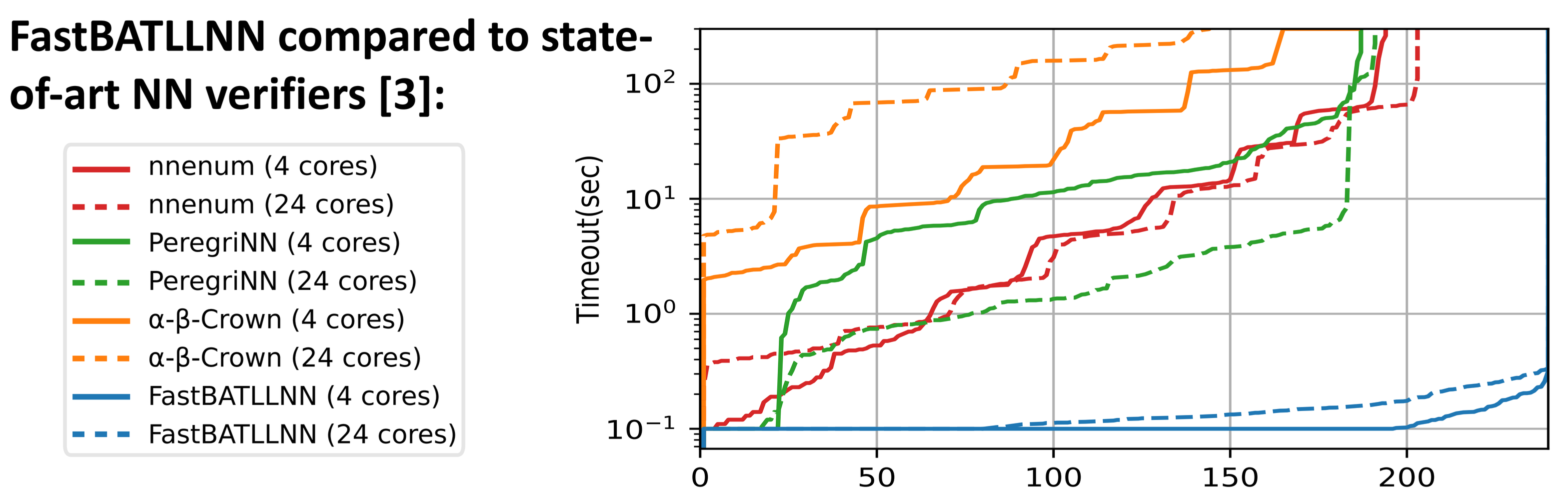
Assured TLL NN Repair

- **Repair counterexample from formal verification: no assurance → assurance**
- Also: repair c.e. while assuring existing safe behavior is retained [1]
- Repair problem is hard: *one neuron affects many affine regions*
  - Change one neuron to repair c.e. → behavior elsewhere in state space is affected → undo original safety of NN
- Solution: TLL ***semantics*** separate "local" and "global" concerns (local linear functions/selector layer) [1]
  - Input-affine dynamics/one-step counterexample [1]:
    - Alternate between Local & Global convex optimization problems

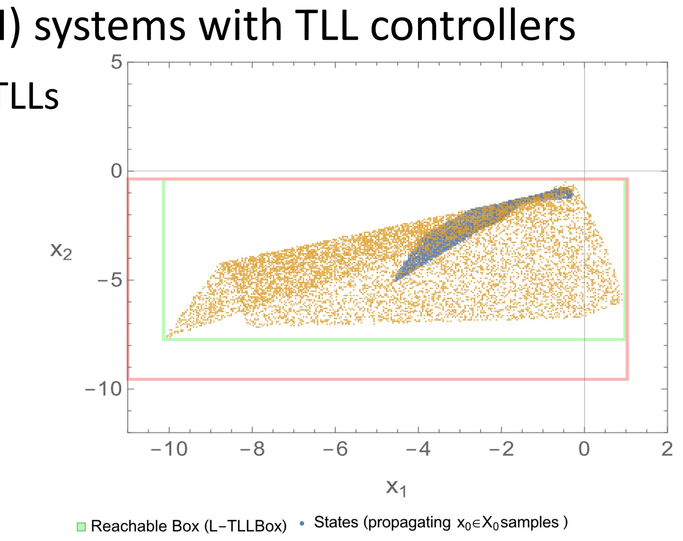


Formal TLL NN Verification

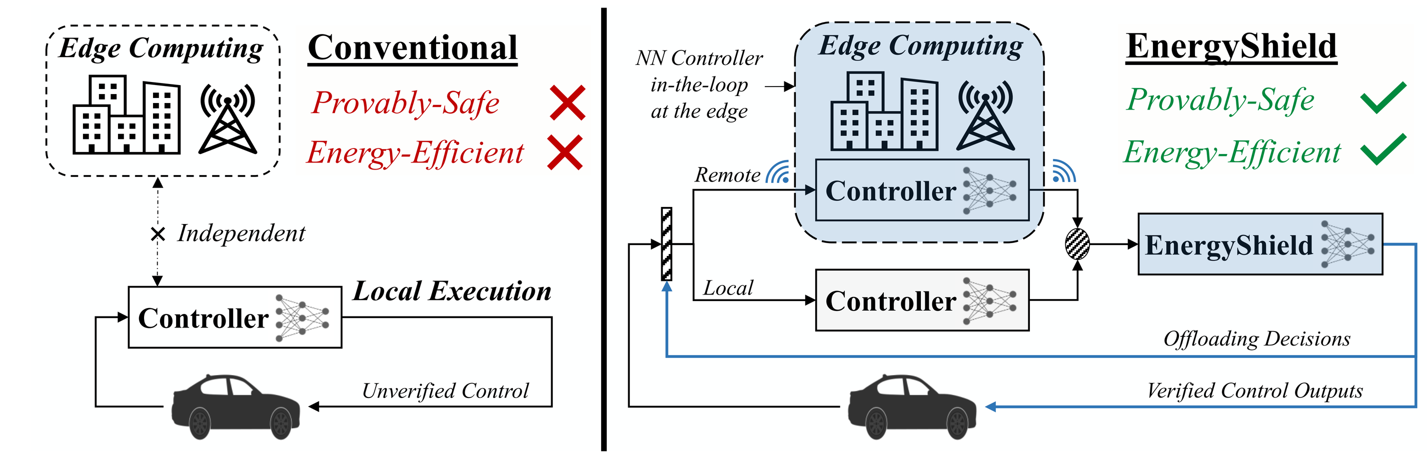
- Fast **polynomial** complexity algorithms for TLL Verification (in # neurons) [5]
  - Restricted architectures → faster verification [5]
- Fast Box Analysis of Two Level Lattice NNs: **FastBATLLNN** [3]
  - Restricted, "Box-like" (hypercube) output constraints → even faster [3]
  - Polynomial complexity (in # neurons)
  - Exploit constraints and min/max semantics
- **FastBATLLNN compared to state-of-art NN verifiers [3]:**



- Fast reachability for Linear Time-Invariant (LTI) systems with TLL controllers
  - One-step exact LTI reachability **polynomial** for TLLs (exponential for DNNs) [7]
  - **L-TLLBox** [7]: faster speed using bounding box propagation (leverage **FastBATLLNN**)
- Example bounding box after T=3 steps [7]
  - **L-TLLBox (green): 25 seconds**
  - **NNV (red): 139,000 seconds**



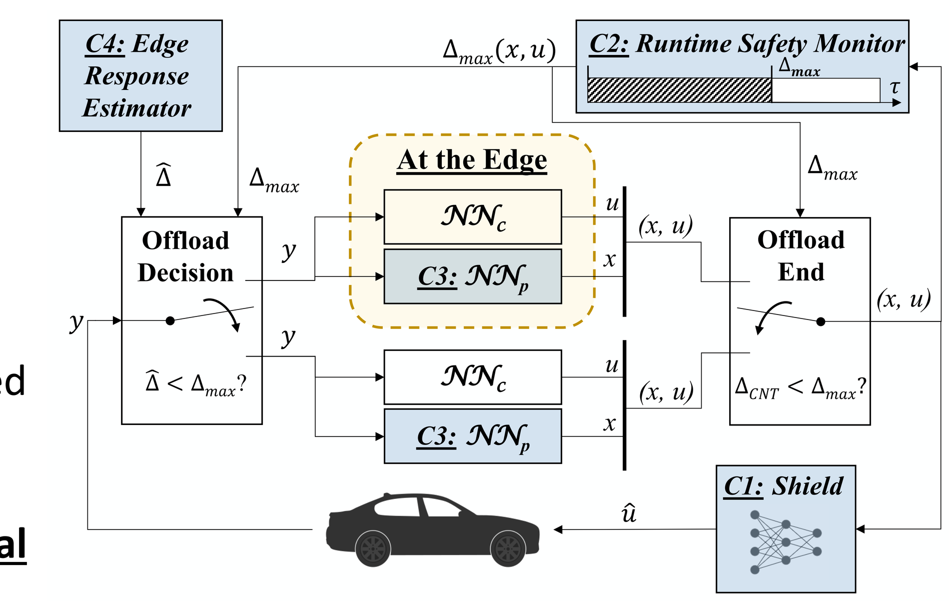
Safe Vehicle-to-Edge NN Offloading



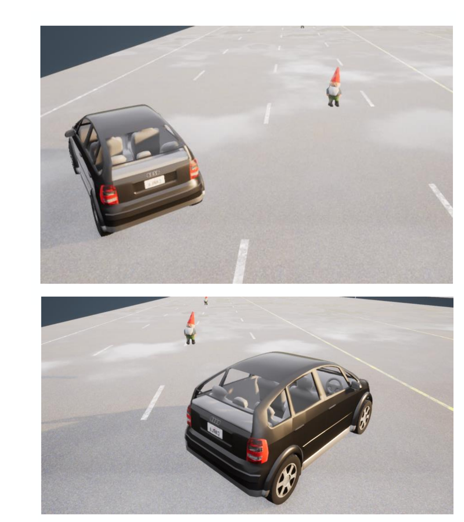
- Energy used by **NN hardware reduces Electric Vehicle range by up to 15% !**

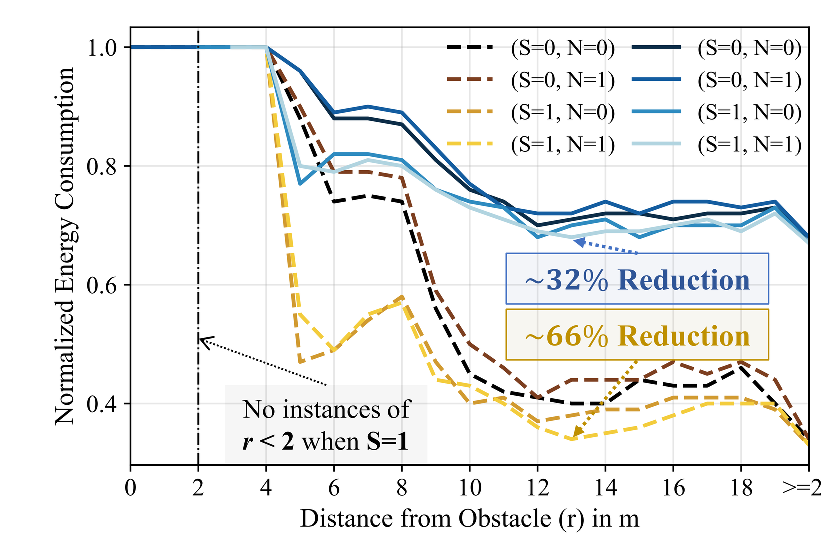
Controller "Shield": Runtime safety monitor

- Use controller shield to **bound safe edge response times** [9]
- Controller shield **assures** safety after receiving response [2]
- On-vehicle computation used as fallback
- **Energy use (and hence simplicity) of shield is critical**



EnergyShield [9]: Safety and Energy Savings





- State dependent energy savings: **more energy saved when "safer"**!

References

[1] Ulices Santa Cruz, James Ferlez, and Yasser Shoukry. Safe-by-Repair: A Convex Optimization Approach for Repairing Unsafe Two-Level Lattice Neural Network Controllers. In *2022 61st IEEE Conference on Decision and Control (CDC)*, 2022. URL: http://arxiv.org/abs/2104.02788, arXiv:2104.02788, doi:https://doi.org/10.48550/arXiv.2104.02788.

[2] James Ferlez, Mahmoud Elnaggar, Yasser Shoukry, and Cody Fleming. ShieldNN: A provably safe NN filter for unsafe NN controllers. 2020. URL: https://arxiv.org/abs/2006.09564.

[3] James Ferlez, Haitham Khedr, and Yasser Shoukry. Fast BATLLNN: Fast Box Analysis of Two-Level Lattice Neural Networks. In *Hybrid Systems: Computation and Control 2022 (HSCC'22)*. ACM, 2022. URL: http://arxiv.org/abs/2111.09293, arXiv:2111.09293.

[4] James Ferlez and Yasser Shoukry. AReN: Assured ReLU NN Architecture for Model Predictive Control of LTI Systems. In *Hybrid Systems: Computation and Control 2020 (HSCC'20)*. ACM, 2020. arXiv:1911.01608.

[5] James Ferlez and Yasser Shoukry. Bounding the Complexity of Formally Verifying Neural Networks: A Geometric Approach. In *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021. doi:https://doi.org/10.1109/CDC45484.2021.9683375.

[6] James Ferlez and Yasser Shoukry. Assured neural network architectures for control and identification of nonlinear systems. 2022. URL: https://arxiv.org/abs/2109.10298.

[7] James Ferlez and Yasser Shoukry. Polynomial-time reachability for LTI systems with Two-Level Lattice Neural Network controllers. *IEEE Control Systems Letters*, 2023 (to appear).

[8] James Ferlez, Xiaowu Sun, and Yasser Shoukry. Two-Level Lattice Neural Network Architectures for Control of Nonlinear Systems. In *59th Conference on Decision and Control (CDC)*, 2020. URL: http://arxiv.org/abs/2004.09628, arXiv:2004.09628.

[9] Mohanad Odema, James Ferlez, Goli Vaisi, Yasser Shoukry, and Mohammad Abdullah Al Faruque. EnergyShield: Provably-safe offloading of Neural Network controllers for energy efficiency. In *International Conference on Cyber-Physical Systems (ICCPS)* [Under review], 2023.

[10] J. M. Tarela and M. V. Martínez. Region configurations for realizability of lattice Piecewise-Linear models. *Mathematical and Computer Modeling*, 30(11):17–27, 1999. URL: http://www.sciencedirect.com/science/article/pii/S0895717799001958, doi:10.1016/S0895-7177(99)00195-8.