Distance-Aware Error Bounds for Reliable Neural Networks and Safe Control : (K-)DAREK

Masoud Ataei
masoud.ataei@maine.edu
Department of Electrical and Computer Engineering, University of Maine

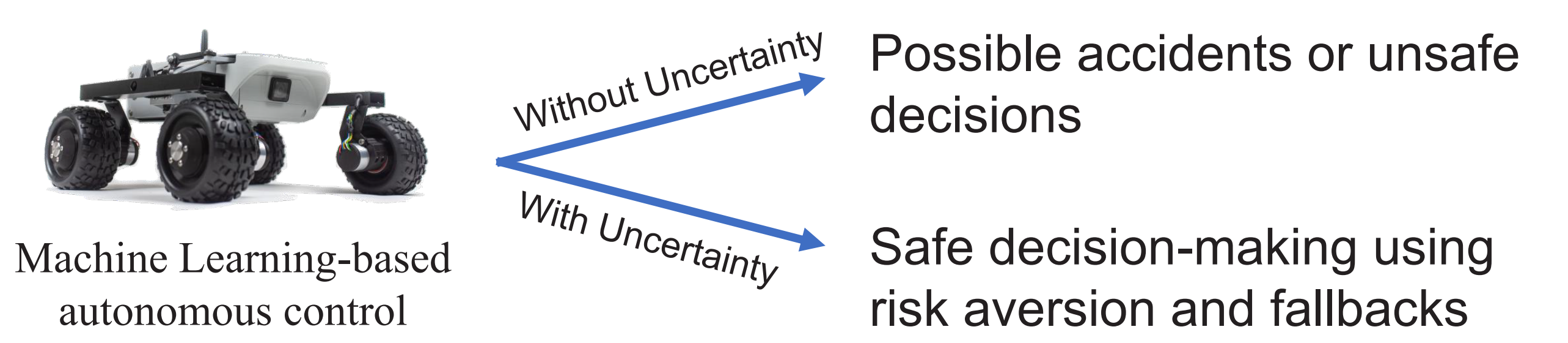
Vikas Dhiman
vikas.dhiman@maine.edu
Department of Electrical and Computer Engineering, University of Maine

Mohammad Javad Khojasteh
mjkeme@rit.edu
Electrical and Microelectronic Engineering, Rochester Institute of Technology

masoud-ataei.github.io/KDAREK/

Introduction

Why uncertainty matter?


Machine Learning-based autonomous control

Without Uncertainty

Possible accidents or unsafe decisions

With Uncertainty

Safe decision-making using risk aversion and fallbacks

When probabilistic methods fall short?

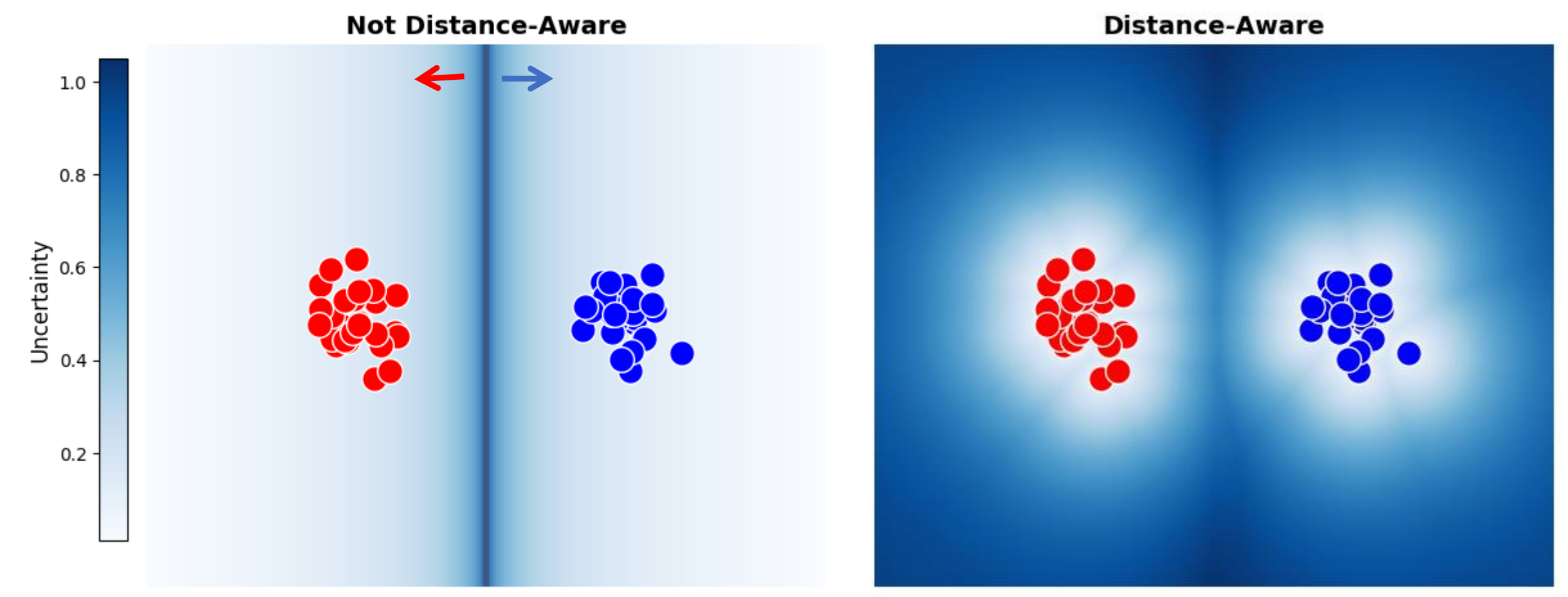
Probabilistic	Worst-case
Performs well on average-case inputs or under typical conditions, however, can fail badly on rare or adversarial inputs.	Guarantees performance even in the hardest possible scenarios, however, may be over conservative.

Semi-parametric

Model	Example	Trade-offs
Parametric	MLPs	Poor uncertainty estimation on rare or adversarial test points
Non-parametric	GPs	Poor scalability for large datasets
Semi-parametric	Spline Neural Networks	Balanced flexibility & efficiency

Distance-awareness

An uncertainty estimator is **distance-aware** when its predicted uncertainty increases as input moves farther from the training data, reflecting higher confidence near familiar samples and more uncertain in unexplored regions.



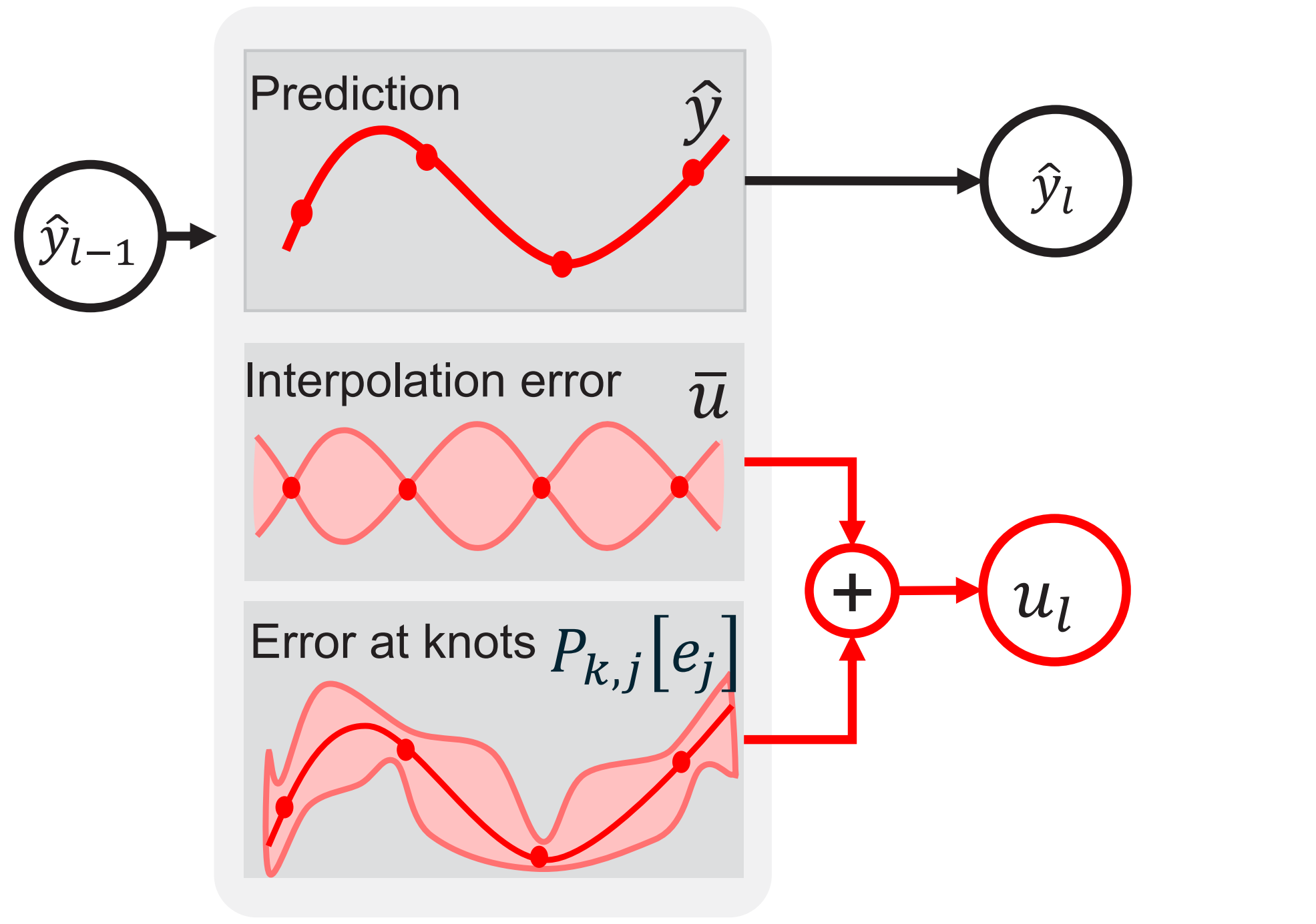
State of the Art & Limitations

	Well-calibrated	Distance Awareness	Efficiency (training, inference)	In/Out-of-Domain
GPs [1]	✅	✅	❌ $\mathcal{O}(n^3), \mathcal{O}(n^2)$	✅
Deep Ensemble [2]	✅	❌	❌ $\mathcal{O}(m\,T), \mathcal{O}(m\,I)$	●
MC-Dropout [3]	●	❌	✅ $\mathcal{O}(T), \mathcal{O}(s\,I)$	❌
DAREK [4] / K-DAREK (Ours)	✅	✅	✅ $\mathcal{O}(T), \mathcal{O}(\log_2(2k)I)$	✅

T: Training one network
I : Inference one network
m: Number of Ensemble models (5-20)
k: Number of knots (5-20)
n: Number of input data

Method

Piecewise Polynomial error (PPE)



$$|f(x) - \hat{f}(x)| \leq u_f(x;\boldsymbol{\tau})$$

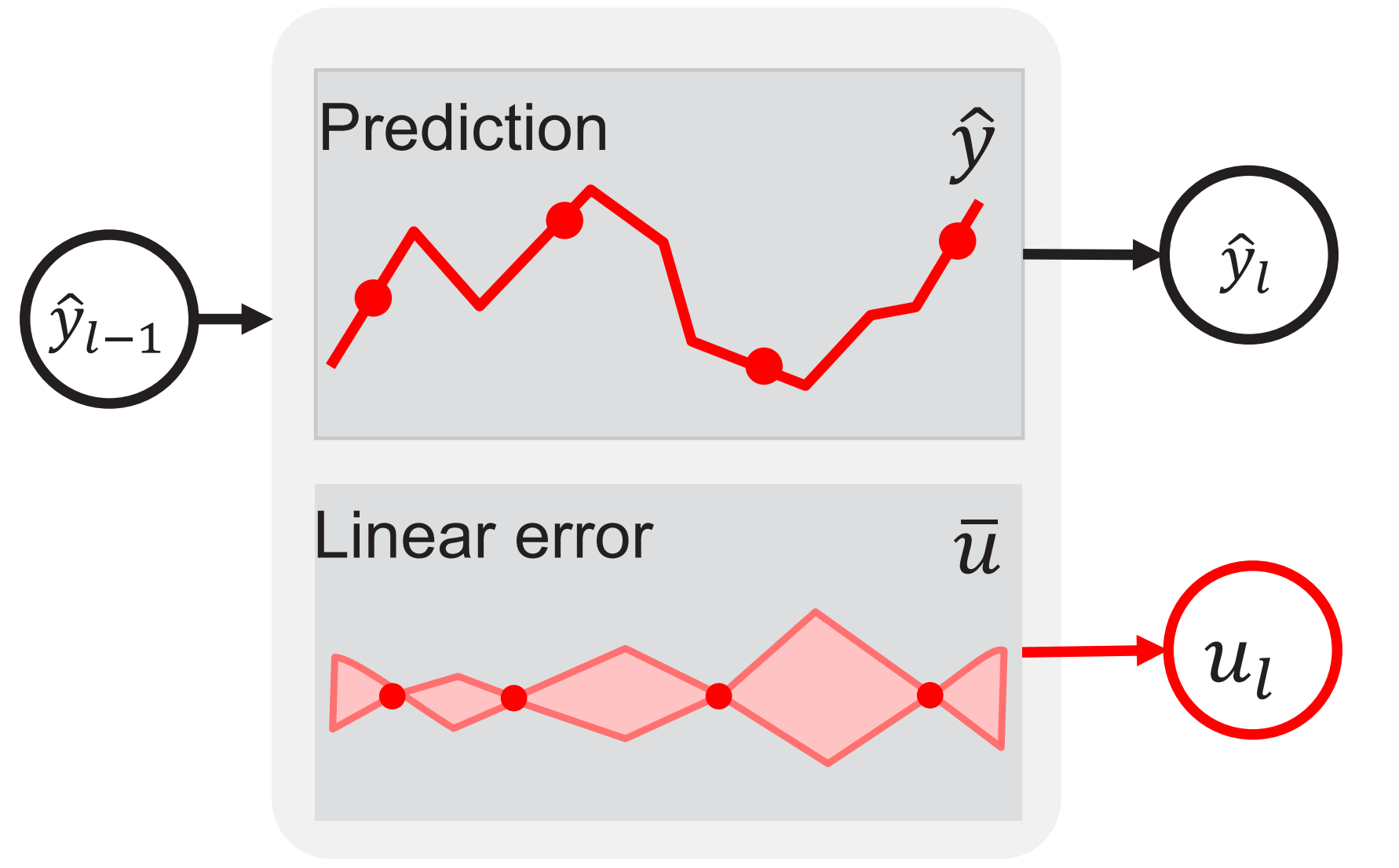
$$\bar{u}_f(x;\boldsymbol{\tau}) := \frac{\mathcal{L}_f^{k+1}}{(k+1)!}\left|\prod_{j=n}^{n+i-1}(x-\tau_j)\right|$$

$$e_n^f(\tau_i) := f(\tau_i) - \hat{f}_{[n]}(\tau_i)$$

$$u_f(x;\boldsymbol{\tau}) = \bar{u}_f(x;\boldsymbol{\tau}) + \left|\mathcal{P}_{k,n}[e_n^f(\boldsymbol{\tau})](x)\right|$$

$$u_l(x) = u_l(\hat{y}_{l-1};\mathcal{T}_l) + \mathcal{L}_l^1 u_{l-1}(\hat{y}_{l-2};\mathcal{T}_{l-1})$$

SNR-MLP Layer



Auxiliary knots act as training representatives that determine the spline knots for later layers.

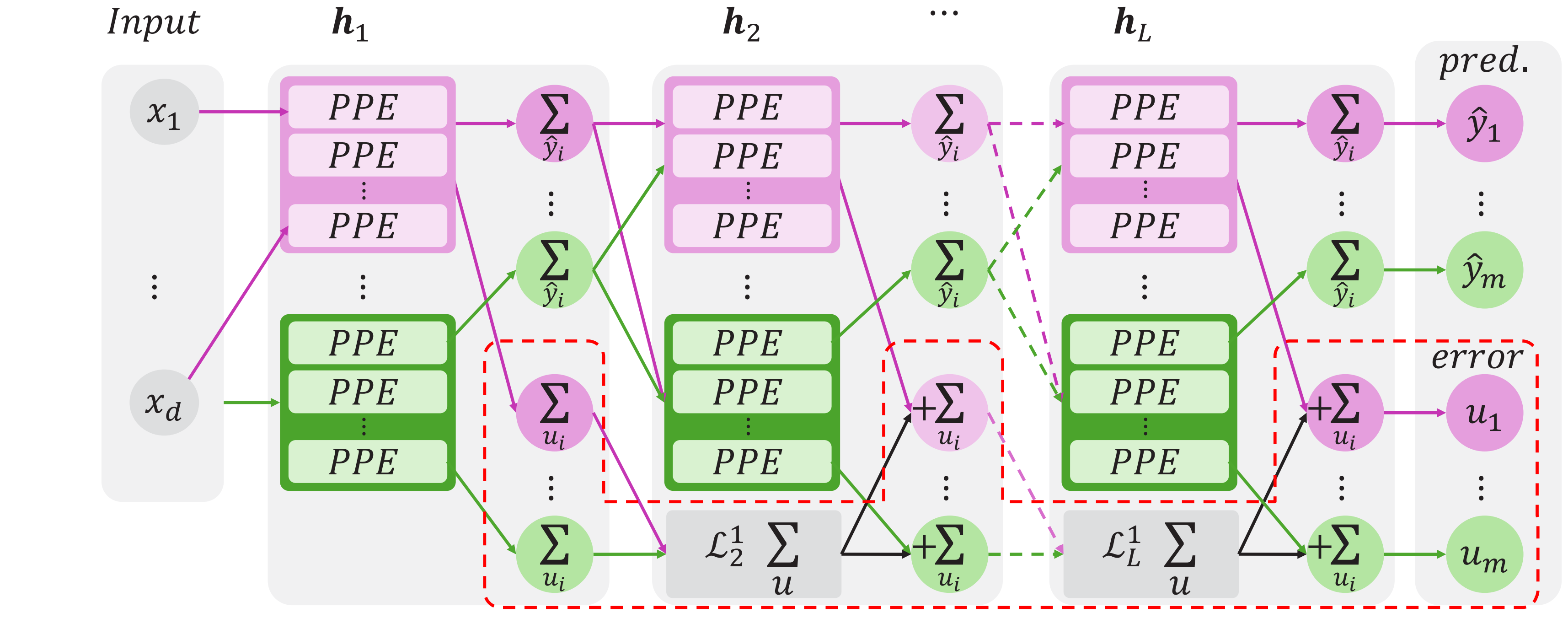
$$u_{\xi_i}(x;\boldsymbol{\tau}_i) = \mathcal{L}_{\xi_i}\Delta\tau_i$$
$$\Delta\tau_i = \min_{\tau'\in\boldsymbol{\tau}_i}|x_i - \tau'|$$

$$u_{\mathrm{MLP}}(x;\mathcal{T}) = \sum_{i=1}^{d} u_{\xi_i}(x;\boldsymbol{\tau}_i)$$

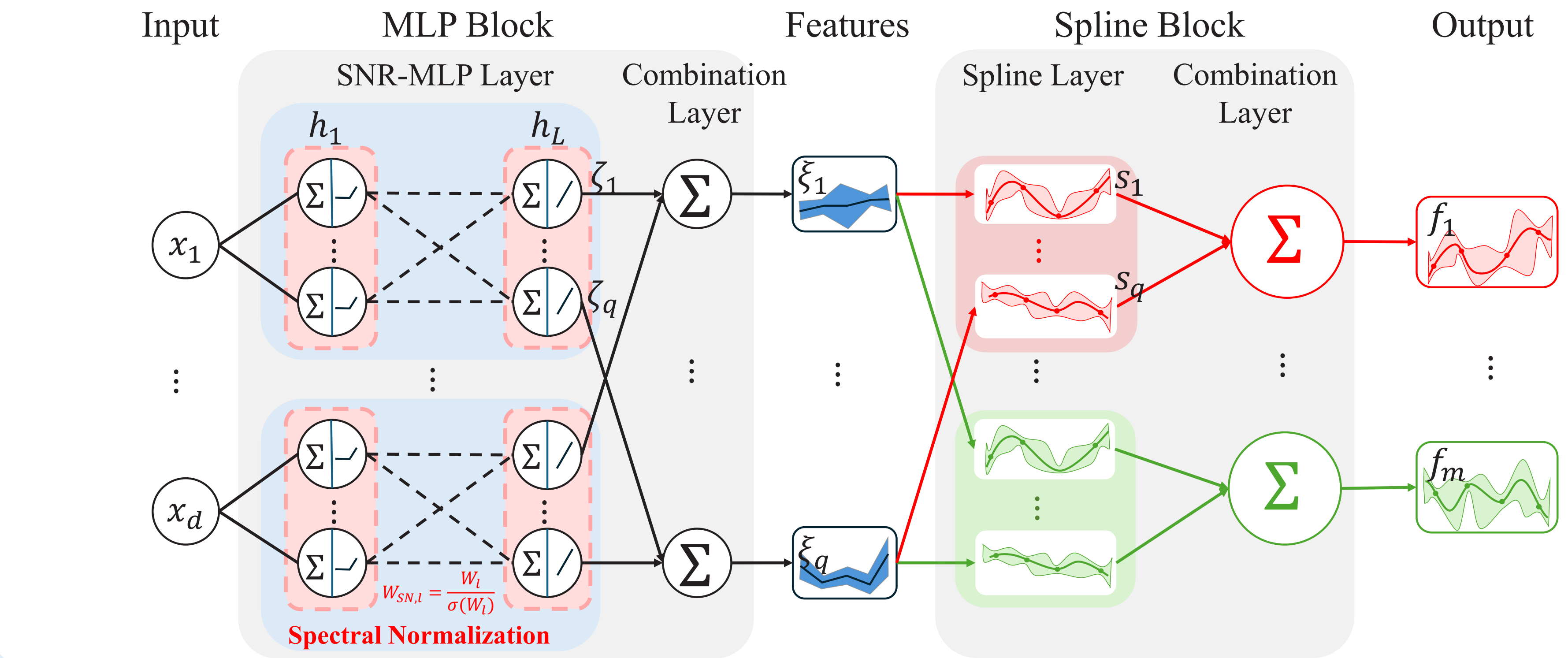
$$u_{\mathrm{sp_r}}(\xi;K) = \sum_{i=1}^{q} u_{\mathrm{sp}_{r,i}}(\xi_i;\boldsymbol{\kappa}_i)$$

$$u_{f_r}(x) = u_{\mathrm{sp_r}}(\xi;K) + \mathcal{L}_{\mathrm{sp}}^1 u_{\mathrm{MLP}}(x;\mathcal{T})$$

DAREK



K-DAREK



Results

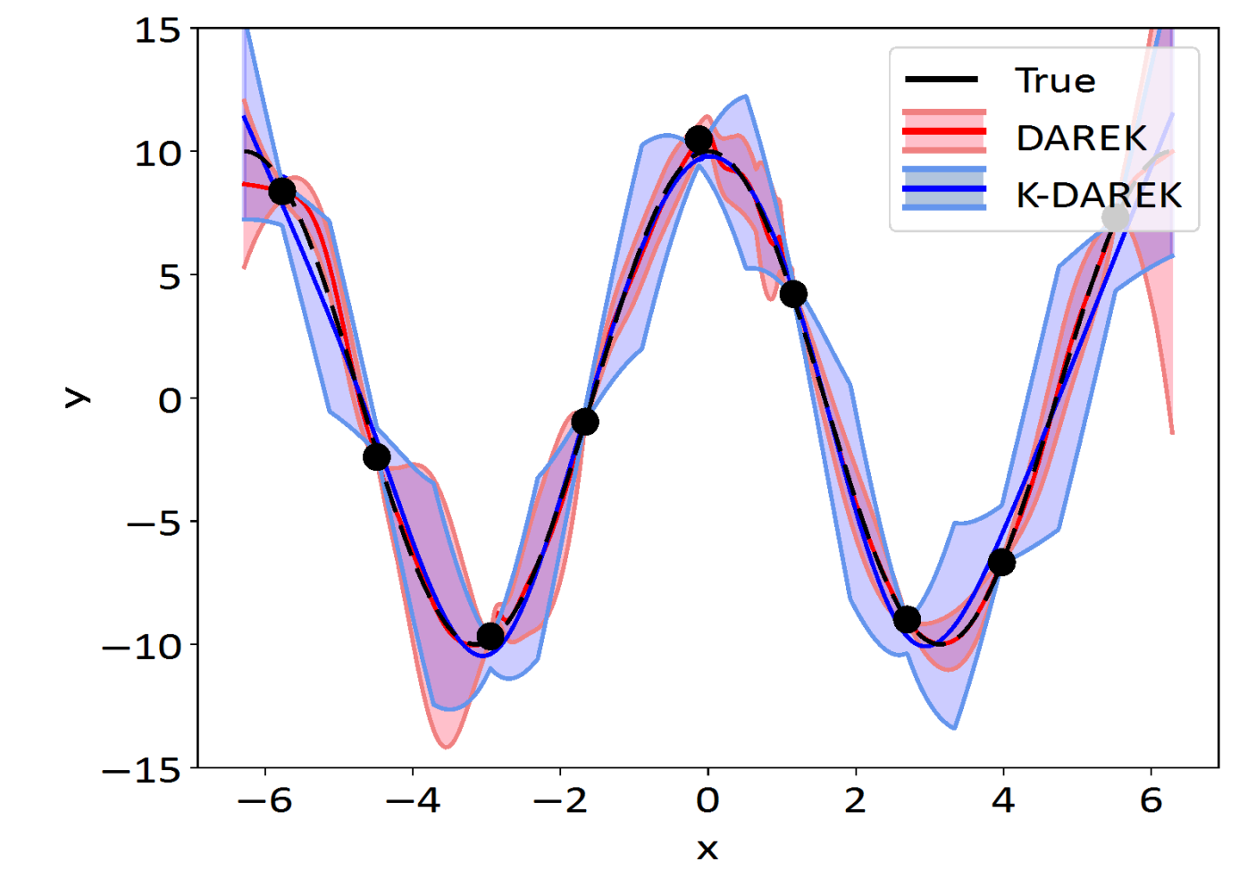


Fig. 1) **Error bound** comparison between DAREK and K-DAREK on *cosine* function dataset. Both models are distance-aware. DAREK is more expressive and interpretable, however may have oscillations, while K-DAREK provides smoother and more conservative results.

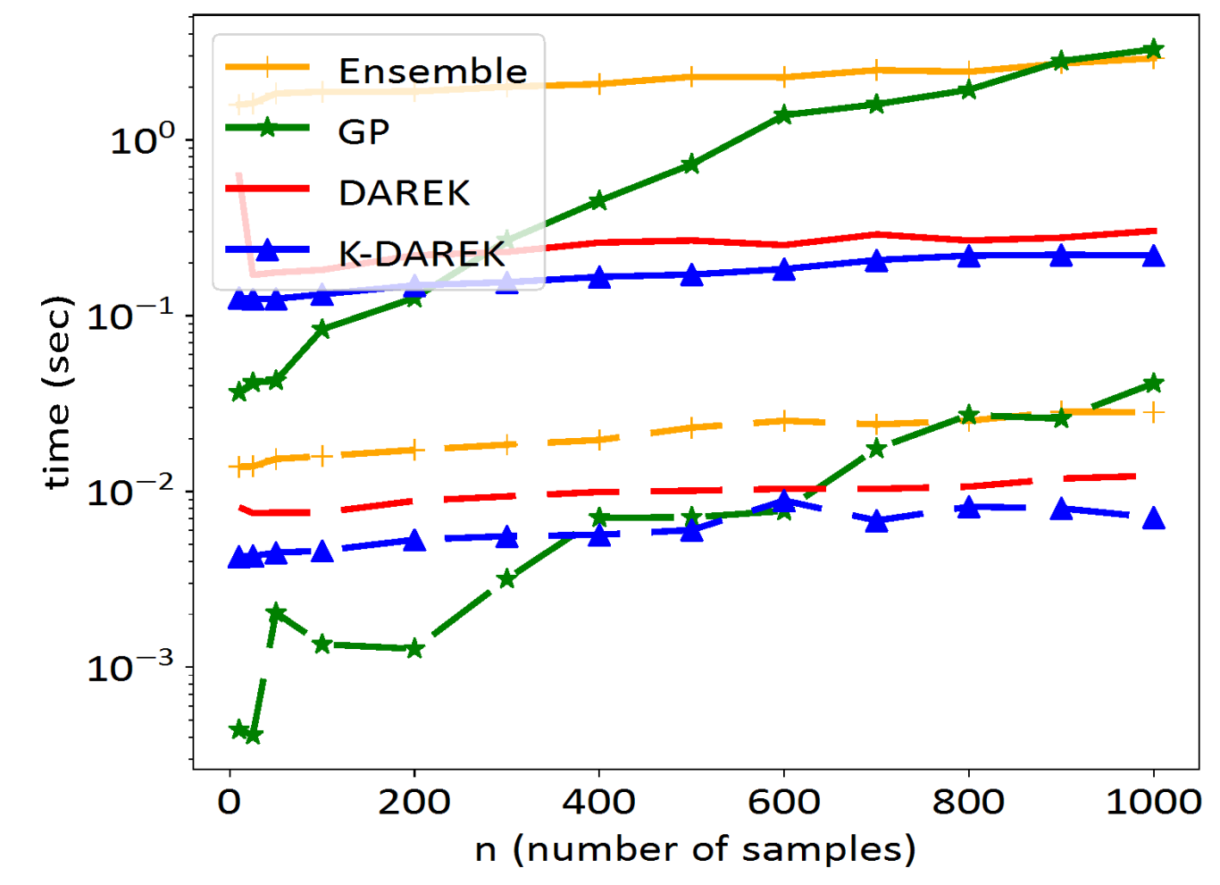
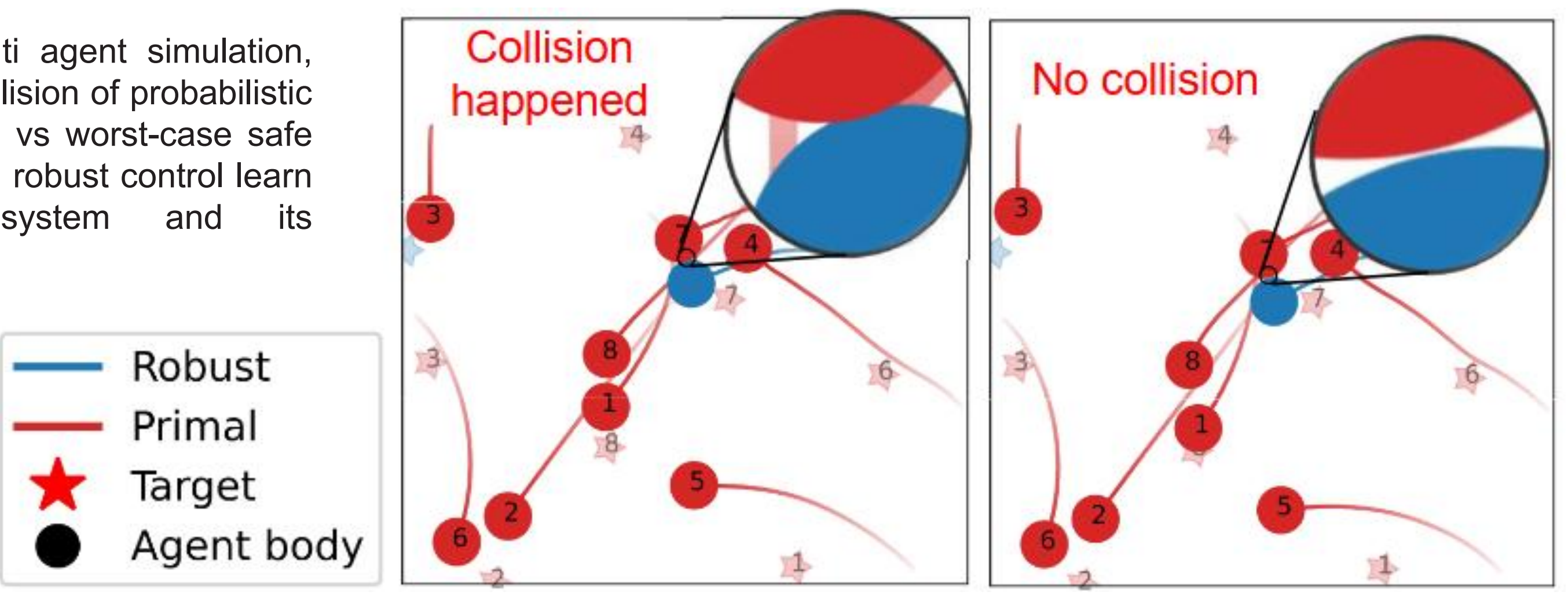


Fig. 2) **Computation time** comparison among Ensemble, GP, DAREK, and K-DAREK models. The computational cost of GP increases with the number of data samples, while K-DAREK demonstrates the highest efficiency for large datasets. The solid line shows training time and dashed line shows inference time.

Model	MSE	Violations(%)	Width	Size
Ensemble	0.378	0.0	10×[1,5,1]	700
GPs	**0.195**	0.0	N. A.	N. A.
DAREK	0.406	7.2	[1,5,1]	70
K-DAREK	0.623	**0.0**	[1,5]+[1,5]	**45**

Table 1) **Model characteristics**: The number of hidden units is kept consistent across all models. The Ensemble requires training multiple models, making it less efficient and not distance aware. DAREK may have violations due to converging to non-smooth approximations, K-DAREK achieves the lowest violation with fewest parameters.

Fig. 3) Multi agent simulation, illustrate collision of probabilistic safe control vs worst-case safe control. The robust control learn dynamic system and its uncertainty.



Conclusion

- **DAREK** and **K-DAREK** are novel frameworks developed for quantifying **error estimation** in spline-based networks and hybrid models, respectively.
- The frameworks provides **deterministic, structured, interpretable, and computationally efficient distance-aware worst-case error bounds.**
- The core mechanism employs Lipschitz continuity assumption over function approximation and Newton's polynomial error bound to ensure tight error bounds in neuron level.
- **DAREK is 3 times and K-DAREK 4 times more computationally efficient and about 8.6 times more scalable than traditional GPs.**

References

1) Williams and Rasmussen. "Gaussian processes for regression." Advances in neural information processing systems 8 (1995).
2) Lakshminarayanan, Pritzel, and Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles." *Advances in neural information processing systems* (2017).
3) Gal and Ghahramani. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning." *international conference on machine learning*. PMLR, 2016.
4) Ataei, Dhiman, and Khojasteh, "DAREK-Distance Aware Error for Kolmogorov Networks." ICASSP, 2025.

Acknowledgements

This work is supported by the National Science Foundation under Grant No. 2218063 and the Gleason Endowment at RIT.