How To Know When You Don't Know: AI for Engineering and 6G

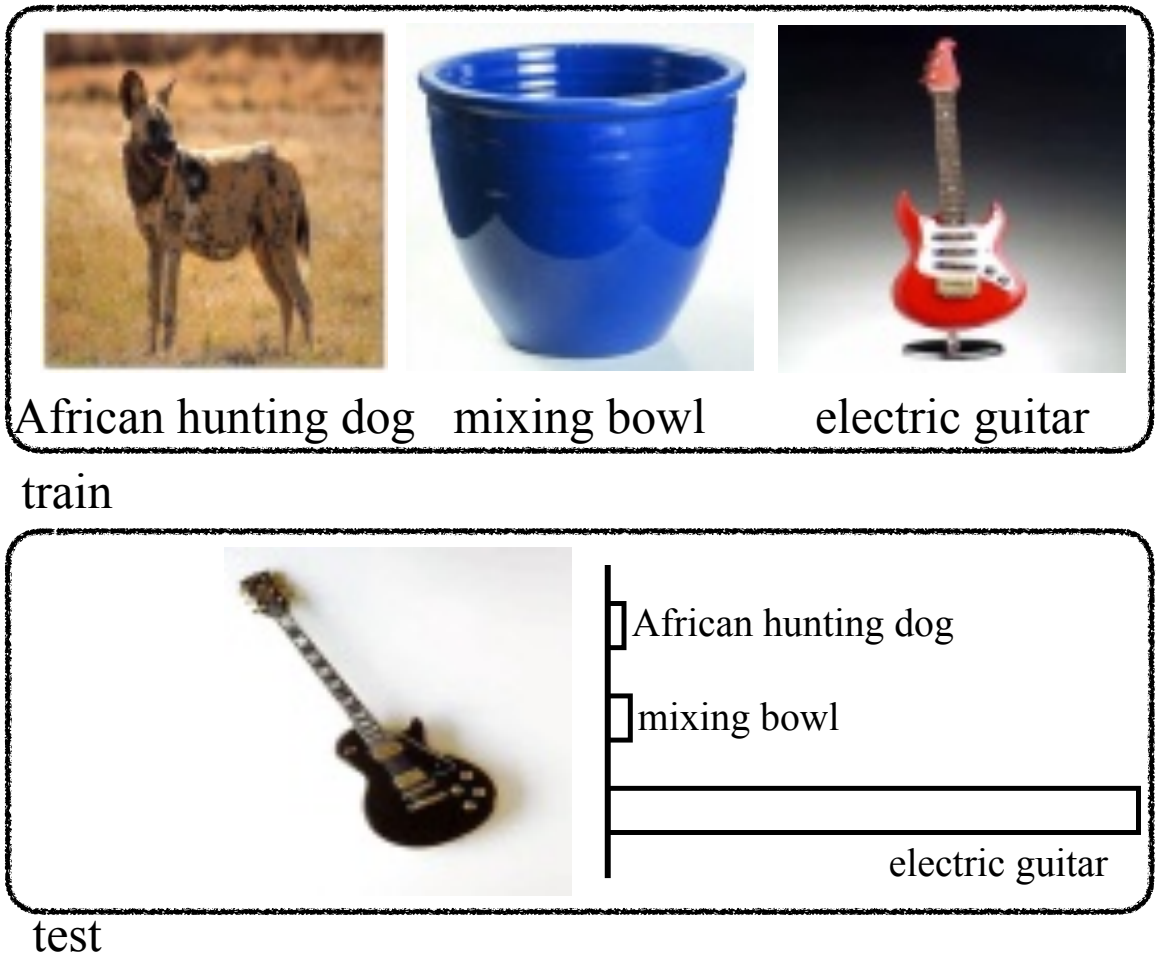
Osvaldo Simeone

King's College London

6G Wireless Foundations Forum 2023, 10/7/2023

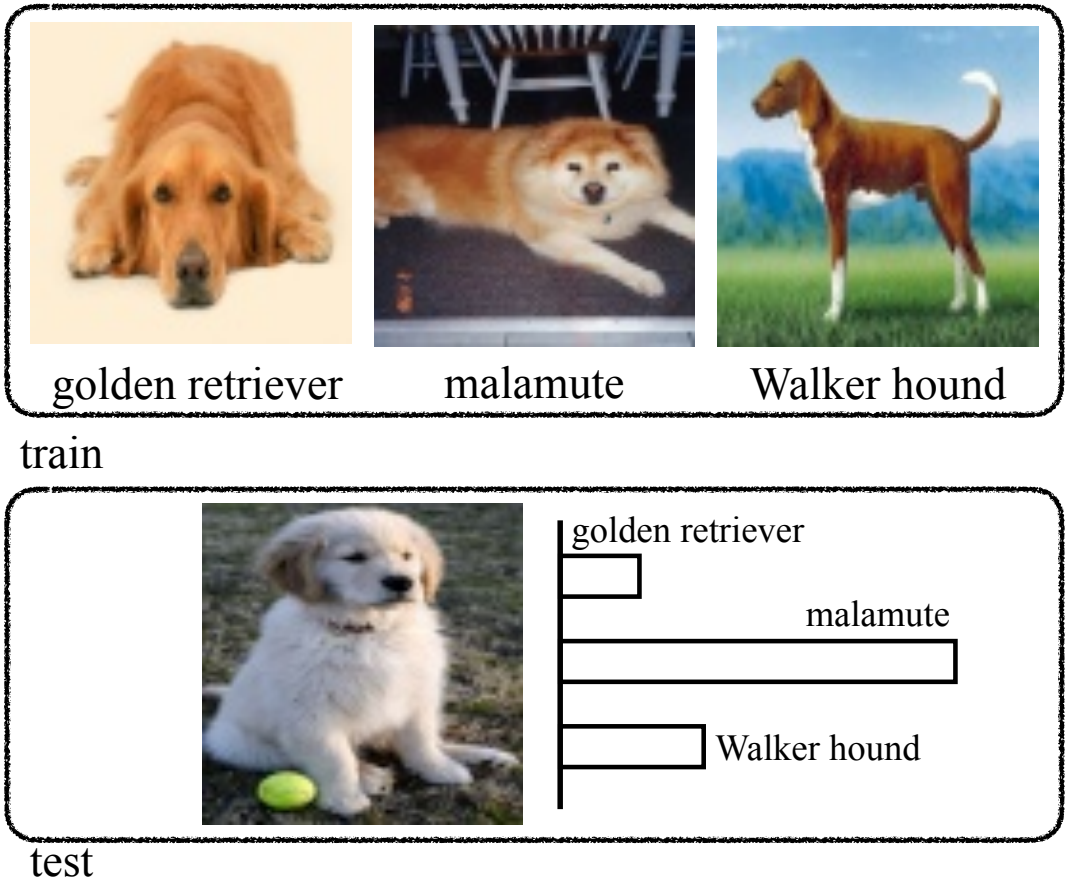
Motivation

Calibration of AI



- AI models typically output a hard decision, along with a **confidence level** (or, conversely, an **uncertainty level**).

Calibration of AI

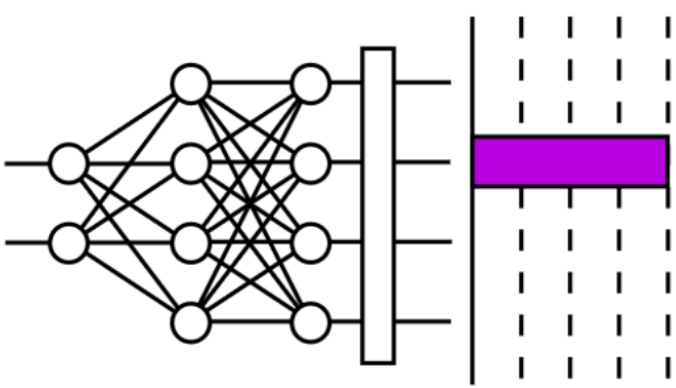


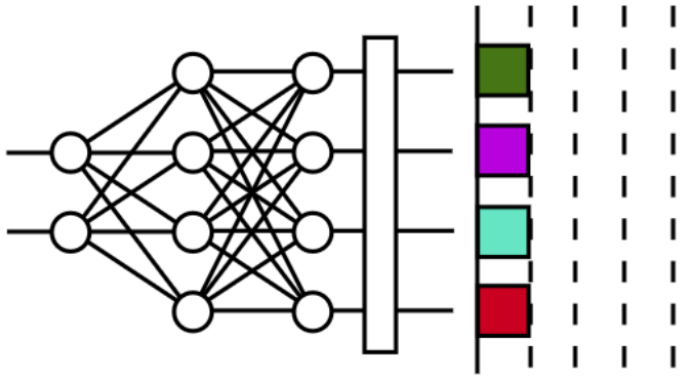
- When failing, conventional **deep learning**-based AI systems tend to make **incorrect** decisions **confidently**.[1]

[1] G. Guo, et al, "On calibration of modern neural networks," in Proc. International conference on machine learning (ICML), 2017.

Calibration of AI

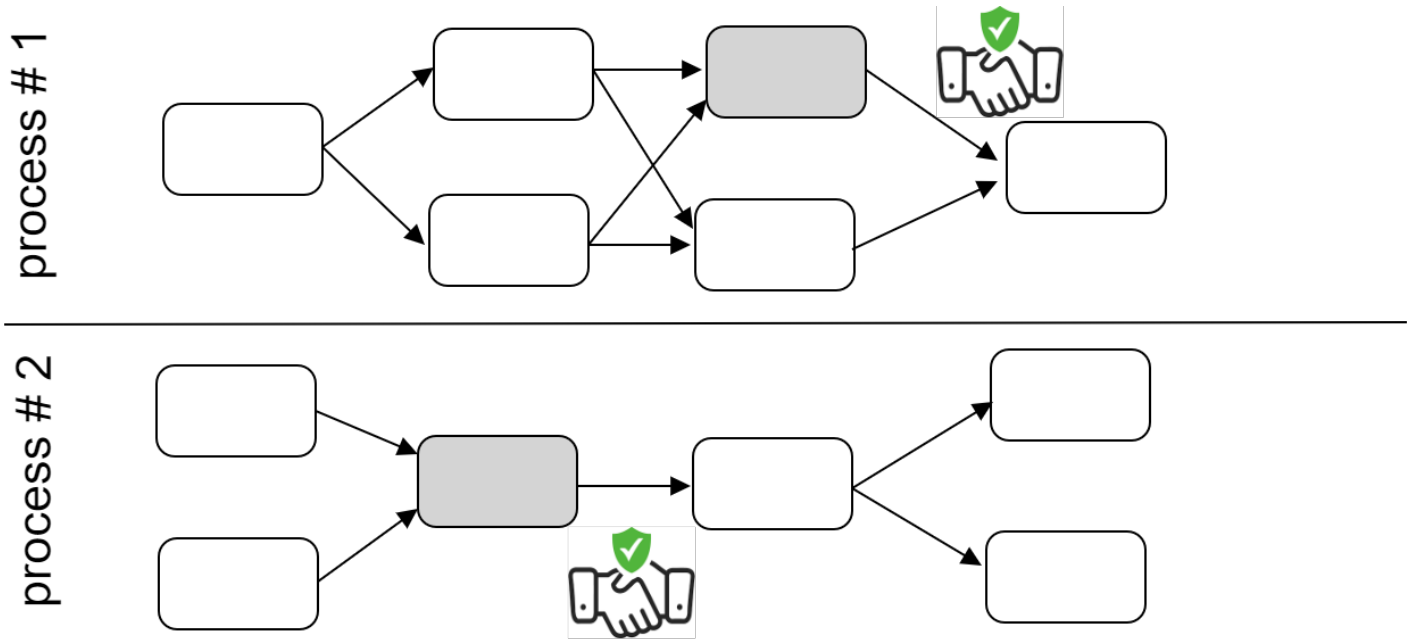
- Conventional deep learning-based AI provides unreliable **estimates of uncertainty**: **poor calibration**.
- A well-calibrated AI is one that "**knows when it knows and knows when it does not know**".





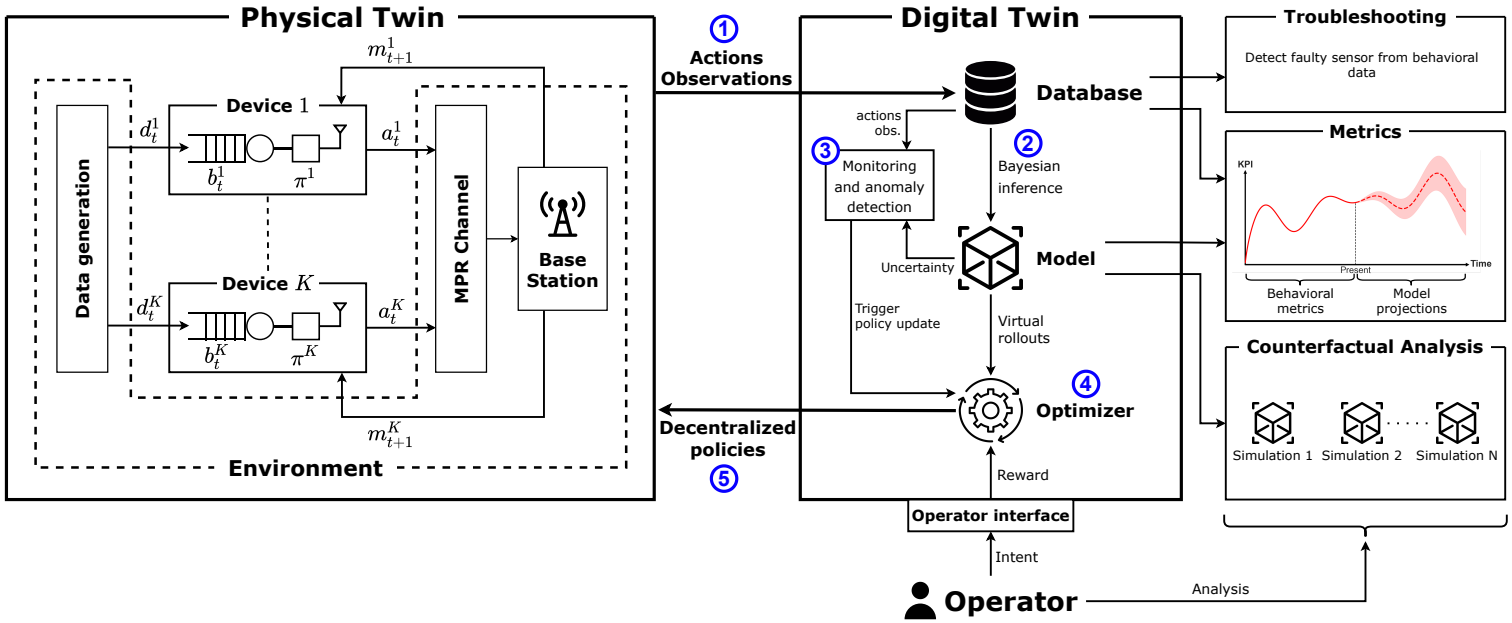
Calibration and Engineering

- Calibration of AI is a **key requirement** for applications to **engineering**:
  - The value of an AI model is often not in its local accuracy, but rather in the role it plays within a number of **decision-making processes**.
  - The output of an AI model should be **trusted** by other models, modules, and subsystems, requiring **calibration**.



Calibration and Engineering

- In **digital twin** platforms for the control and monitoring of 6G networks, calibration supports:[2,3]
  - **safety** by mitigating **model exploitation**
  - functionalities such as **directed exploration**

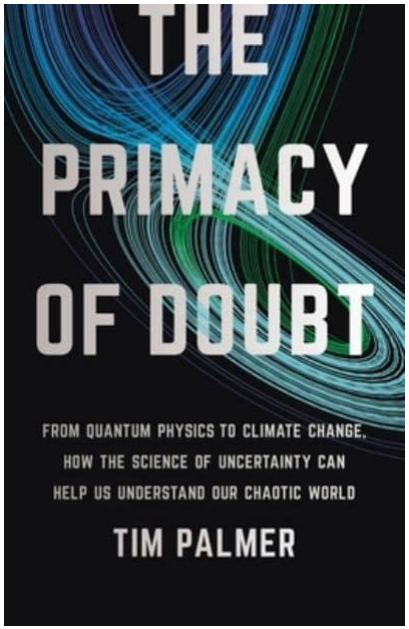


2 A. Thelen, et al, "A Comprehensive Review of Digital Twin – Part 2: Roles of Uncertainty Quantification and Optimization, a Battery Digital Twin, and Perspectives," arXiv:2208.12904, 2022.

3 C. Ruah, O. Simeone, B. Al-Hashimi, "Digital Twin-Based Multiple Access Optimization and Monitoring via Model-Driven Bayesian Learning," arXiv:2210.05582, 2022.

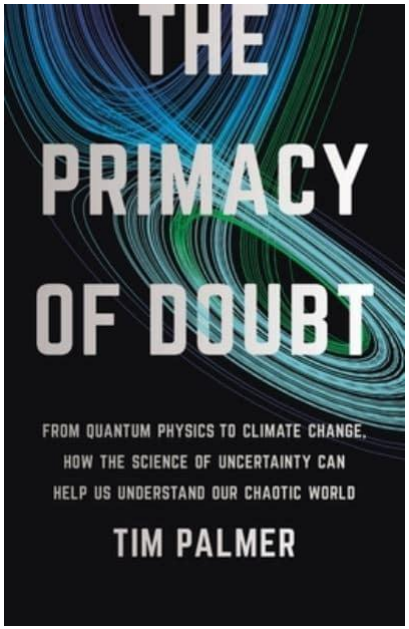
Calibration of AI

- The poor calibration of AI models stems from the fact that they are designed with the goal of **maximizing accuracy**, not of **managing uncertainty**.
- *"If a machine is expected to be infallible, it cannot also be intelligent."* (Alan Turing)



Calibration of AI

- The poor calibration of AI models stems from the fact that they are designed with the goal of **maximizing accuracy**, not of **managing uncertainty**.
- "*If a machine is expected to be infallible, it cannot also be intelligent.*" (Alan Turing)



Calibration of AI: Challenges

- Obtaining effective and efficient AI solutions that can quantify their uncertainty requires advances in
  - **algorithm design**: **beyond optimization** towards statistics and information theory
  - **hardware-algorithm co-design**: from the reproduction of deterministic processes to the efficient control of **stochastic** processes

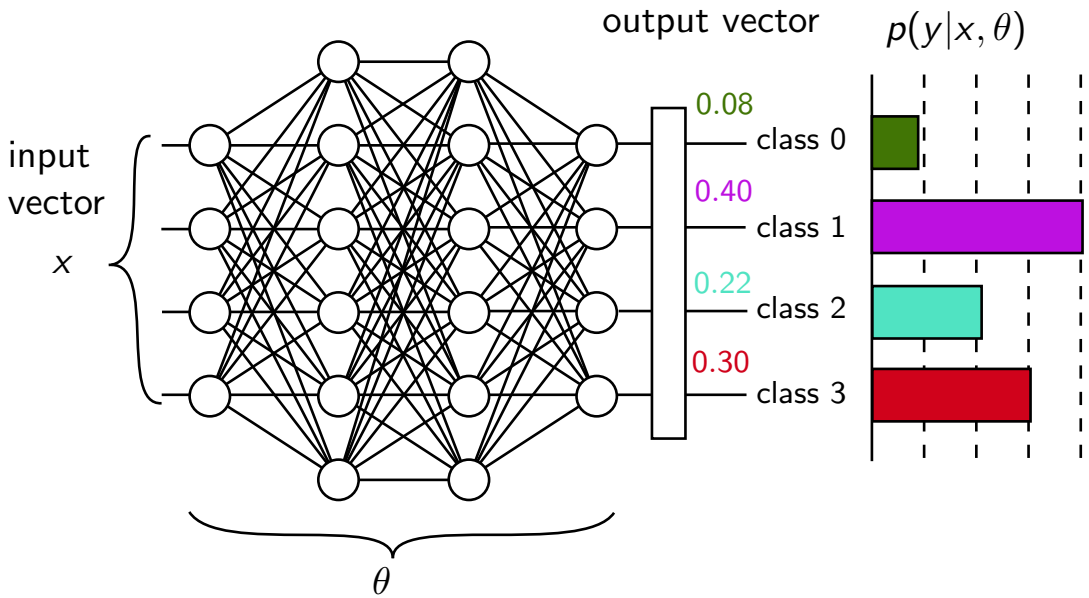
This Talk

- **Part I: Algorithms**
  - Training-based calibration
  - Post-hoc calibration
- **Part II: Hardware-algorithm co-design**
  - Neuromorphic computing
- Conclusions

Part I: Algorithms

Predictive Uncertainty

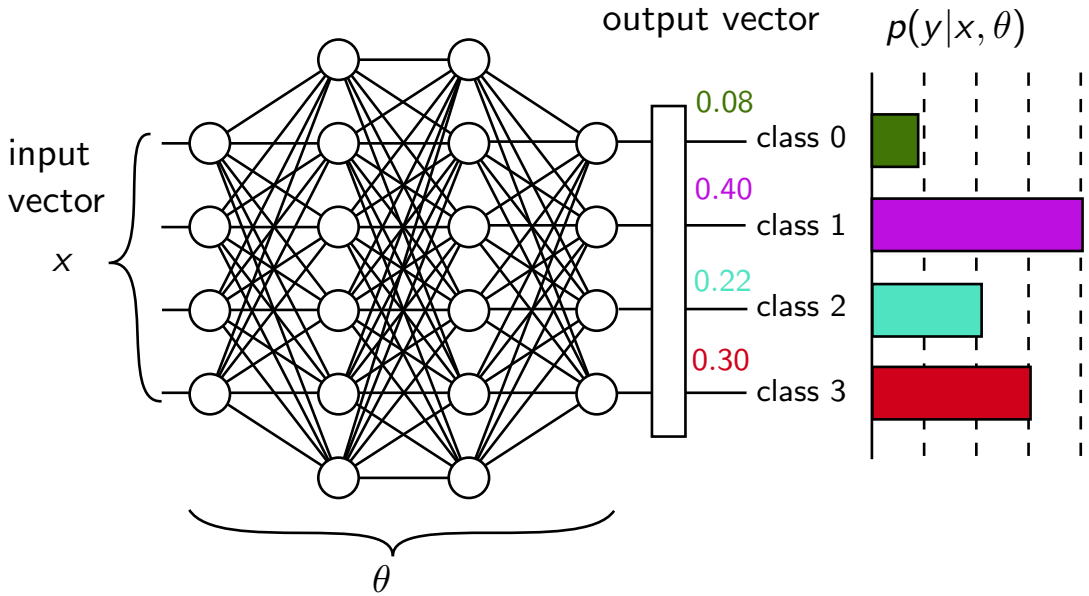
- Typical AI models output **hard decisions** and associated **confidence levels**.



- **Hard decision** $= 1$ (class with largest score)
- **Confidence level** $= 0.4$ (self reported)
- **Calibration** measures the extent to which the confidence level matches the **true (test) accuracy** of a decision.

Predictive Uncertainty

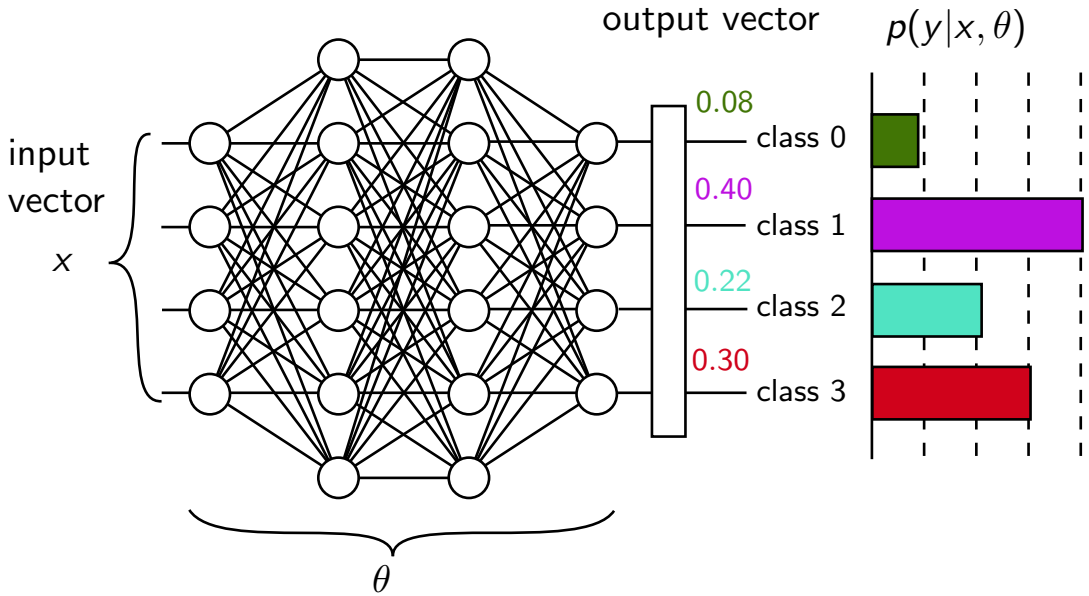
- Typical AI models output **hard decisions** and associated **confidence levels**.



- **Hard decision** $= 1$ (class with largest score)
- **Confidence level** $= 0.4$ (self reported)
- **Calibration** measures the extent to which the confidence level matches the **true (test) accuracy** of a decision.

Predictive Uncertainty

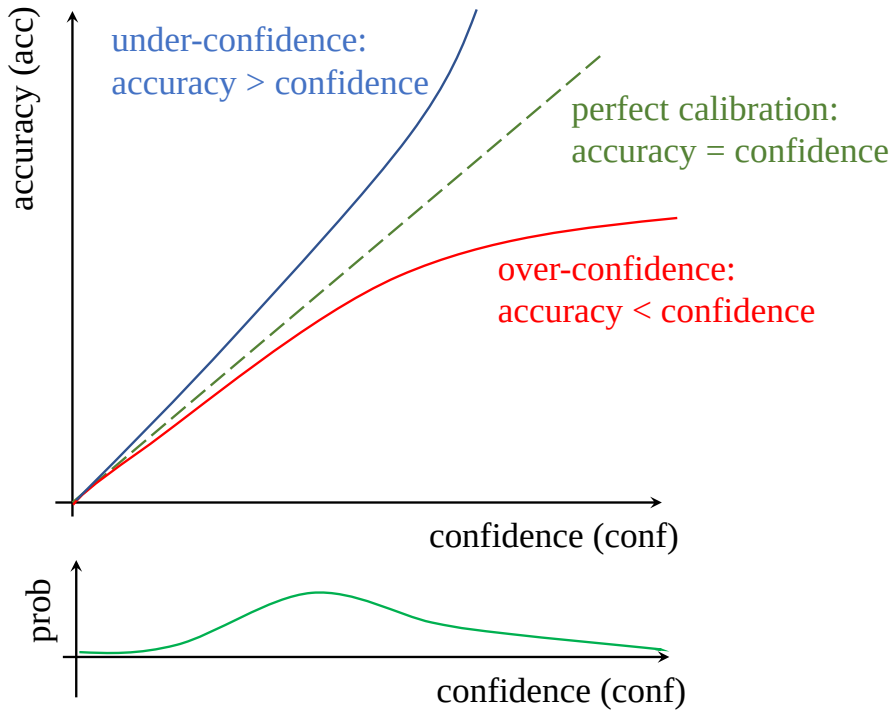
- Typical AI models output **hard decisions** and associated **confidence levels**.



- **Hard decision** $= 1$ (class with largest score)
- **Confidence level** $= 0.4$ (self reported)
- **Calibration** measures the extent to which the confidence level matches the **true (test) accuracy** of a decision.

Measuring Calibration

- Calibration measures are an active subject of research.[4,5]
- Standard approach: **Reliability diagrams** plot accuracy vs. confidence, providing a visual depiction of calibration performance.[6]



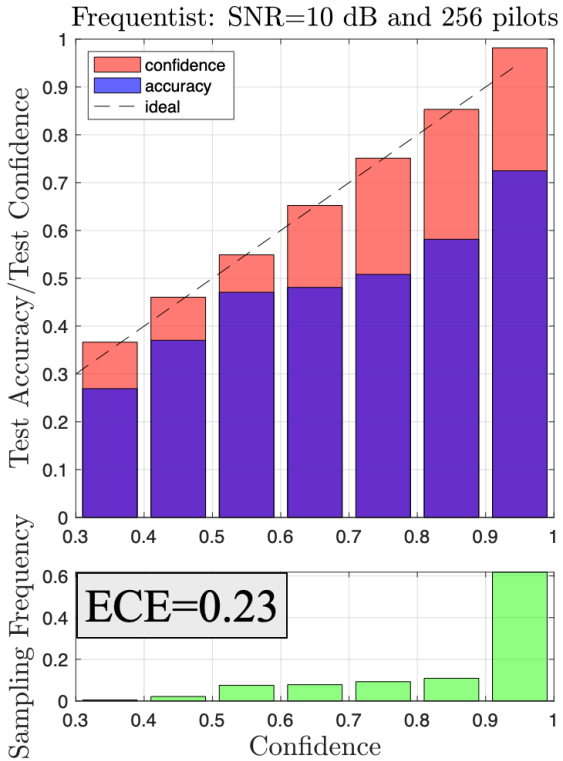
4 A. Perez-Lebel et al, "Beyond calibration: estimating the grouping loss of modern neural networks," arxiv, 2022.

5 B. Holtgen, R. Williamson, "On the Richness of Calibration," arXiv:2302.04118, 2023.

6 C. Guo, et al, "On calibration of modern neural networks," ICML 2017.

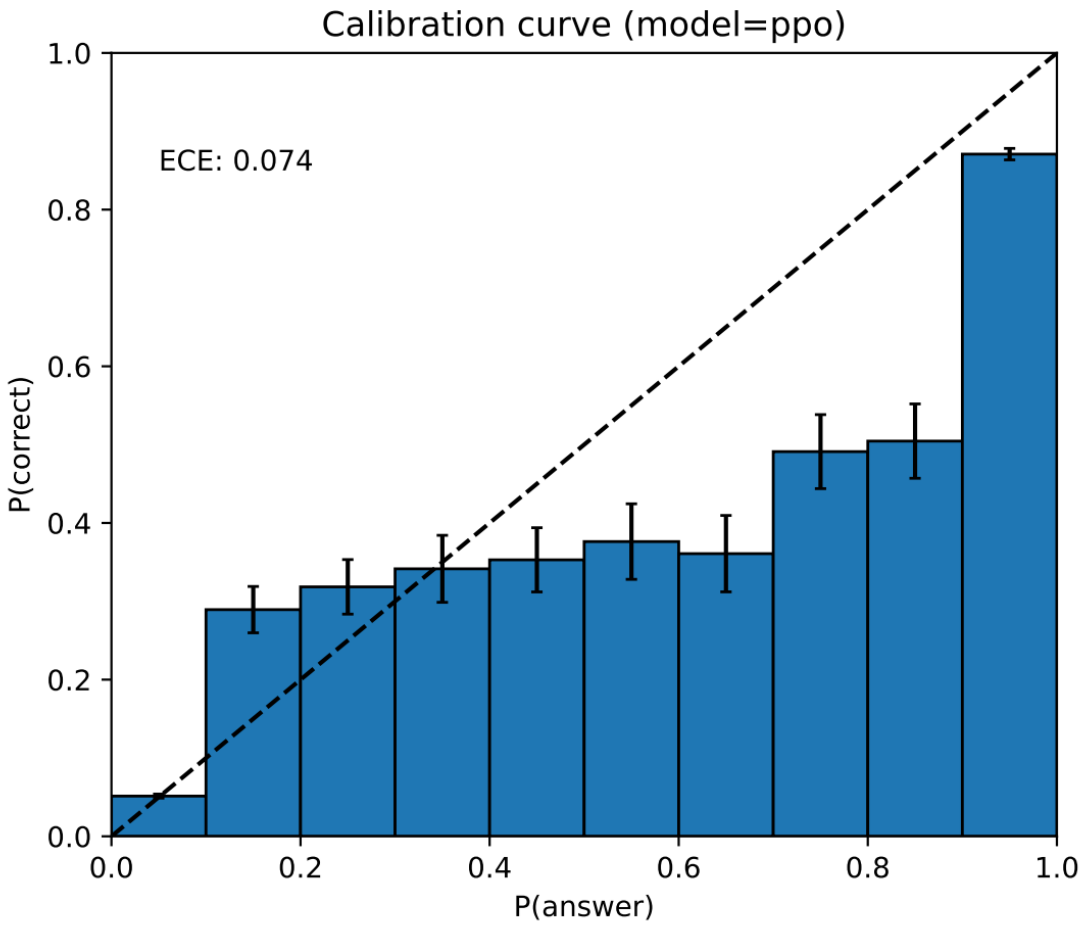
Calibration for Conventional Learning

- **Conventional (frequentist) learning** yields poorly calibrated, typically overconfident, decisions.[7]



[7] S. Park, K. Cohen, O. Simeone, and S. Shamai, "Bayesian Active Meta-Learning for Reliable and Efficient AI-Based Demodulation", IEEE Trans. Signal Processing, to appear.

Calibration for Conventional Learning



[OpenAI, GPT-4 Technical Report, 2023]

Taxonomy of Calibration Techniques

- **Training-based calibration**:
  - Train probabilistic models by accounting for calibration performance
- **Post-hoc calibration**:
  - Use a validation set to "recalibrate" a probabilistic model
  - Can be combined with training-based calibration

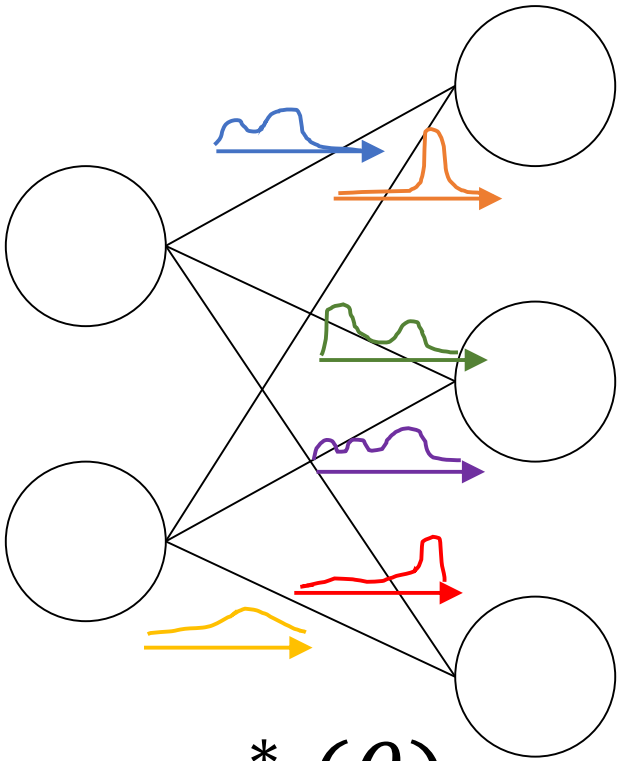
Taxonomy of Calibration Techniques

- **Training-based calibration**:
  - Train probabilistic models by accounting for calibration performance
- **Post-hoc calibration**:
  - Use a validation set to "recalibrate" a probabilistic model
  - Can be combined with training-based calibration

Part I: Algorithms
Training-Based Calibration

Bayesian Learning

- **Bayesian learning** is the gold standard for training-based calibration:
  - Treat parameters as random variables...
  - ... whose distribution is updated as a function of prior knowledge and data to encode **epistemic uncertainty**.

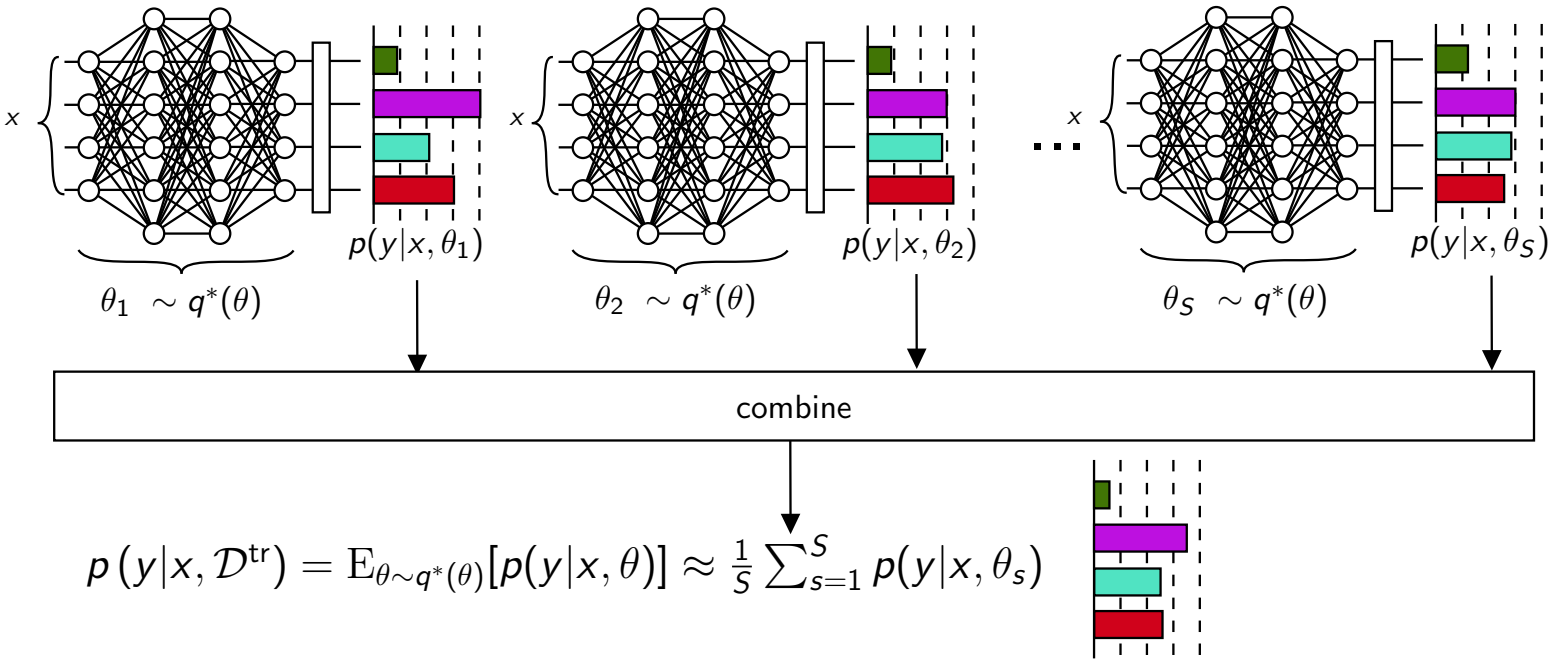


Ensemble Prediction

- Decision obtained via **ensembling**, i.e., via

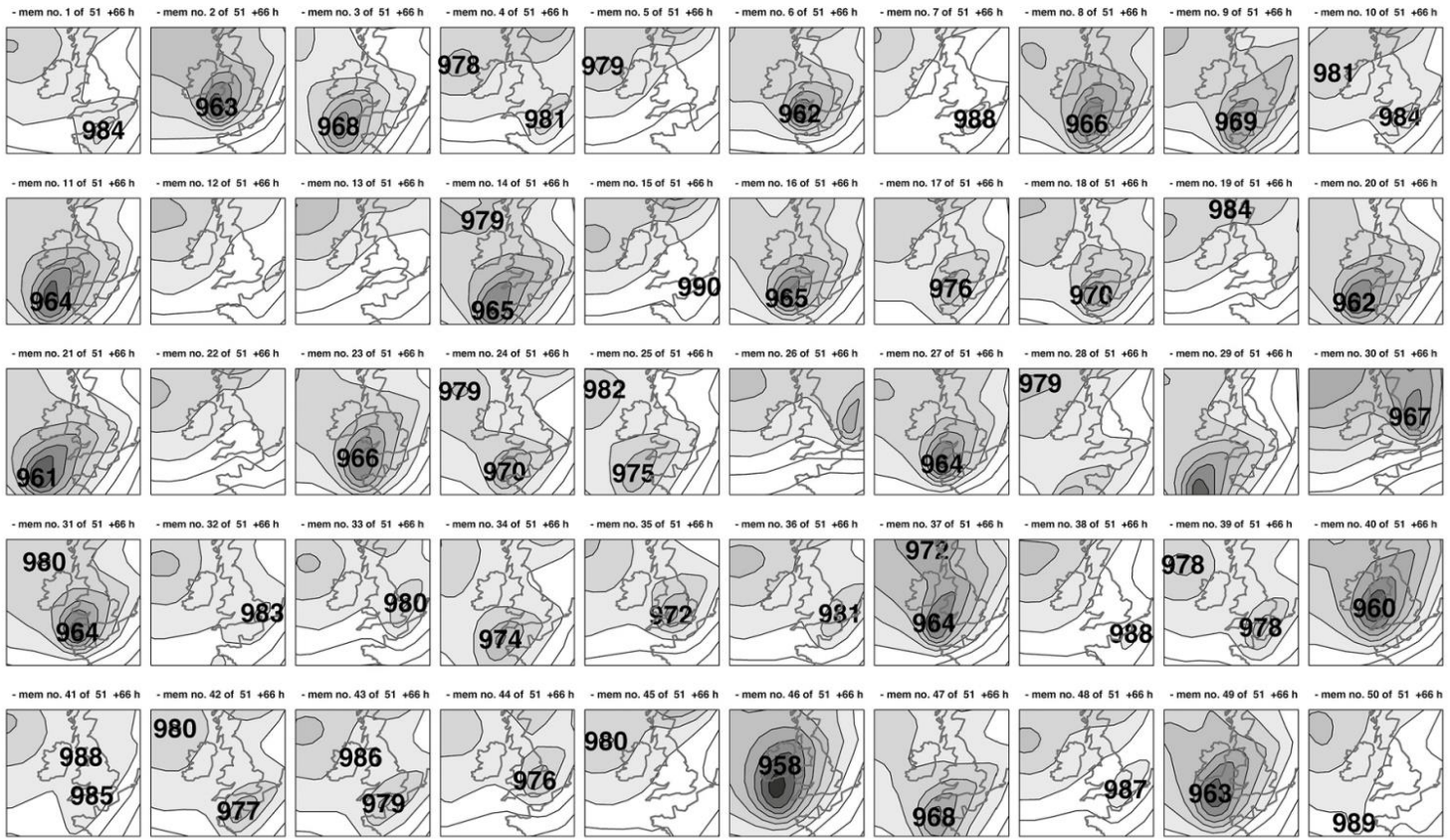
$$\mathrm{E}_{\theta \sim q^*(\theta)} \left[p(y|x, \theta) \right],$$

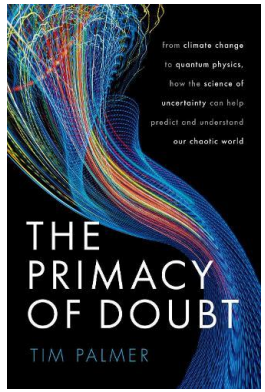
accounting for the "opinions" of multiple models.



Ensemble Prediction

- Epistemic uncertainty can be quantified via the **disagreement** or agreement of models within the ensemble.

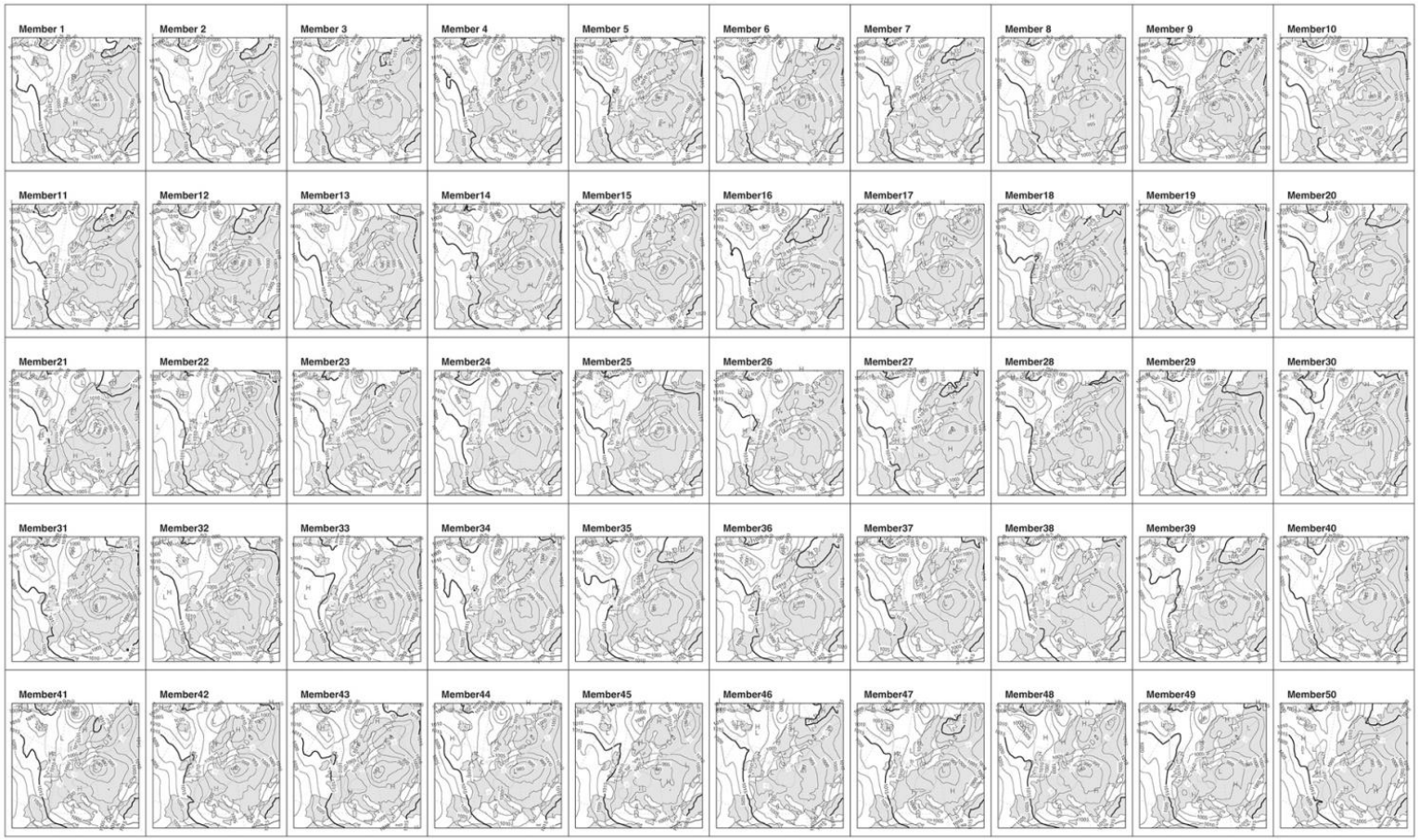


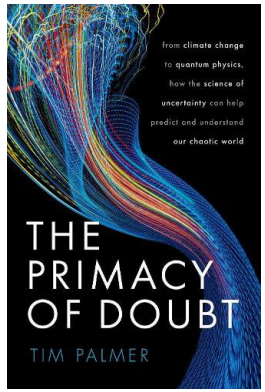


[T. Palmer '22]

Ensemble Prediction

- Epistemic uncertainty can be quantified via the disagreement or **agreement** of models within the ensemble.





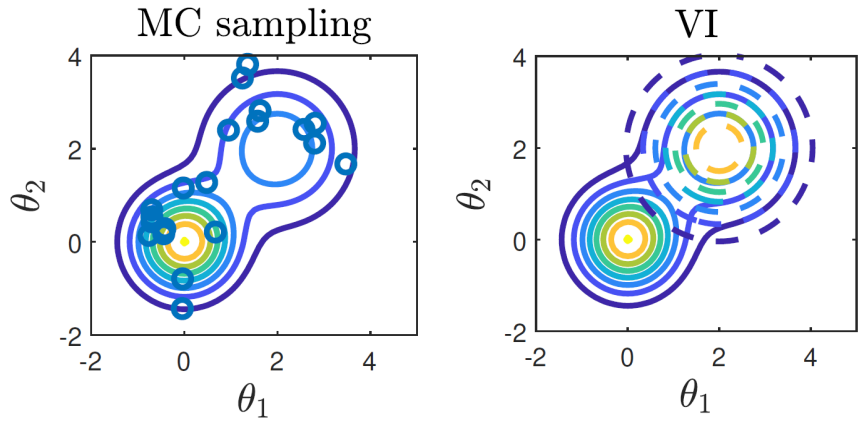
[T. Palmer '22]

Bayesian Learning

- Bayesian learning minimizes the **free energy**[8]

$$\min_{q(\theta)} \mathrm{E}_{\theta \sim q(\theta)}[\text{training loss}(\theta)] + \beta \cdot \text{distance}\left(q(\theta), p(\theta)\right)$$

- Exact minimization of the free energy yields the **posterior distribution**.
- Practical solutions are based on **approximate** optimization (variational inference), (Monte Carlo) **sampling**, and hybrid versions thereof.



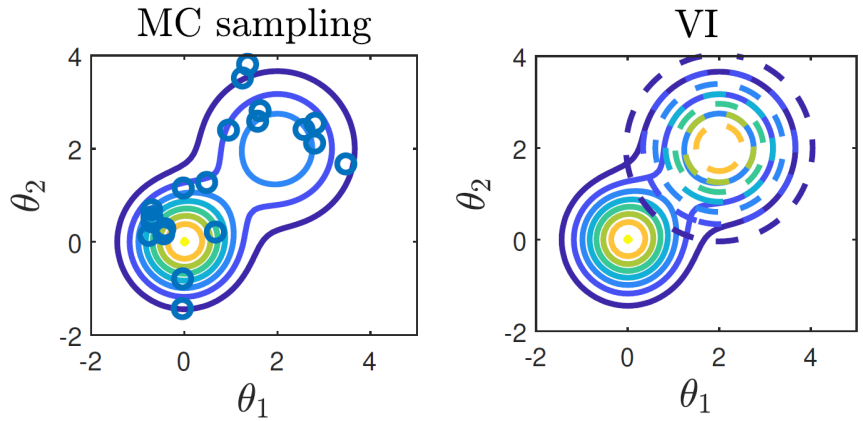
8 J. Knoblauch, et al, "Generalized variational inference: Three arguments for deriving new posteriors," arXiv:1904.02063, 2019.

Bayesian Learning

- Bayesian learning minimizes the **free energy**[8]

$$\min_{q(\theta)} \mathrm{E}_{\theta\sim q(\theta)}[\text{training loss}(\theta)] + \beta \cdot \text{distance}\left(q(\theta), p(\theta)\right)$$

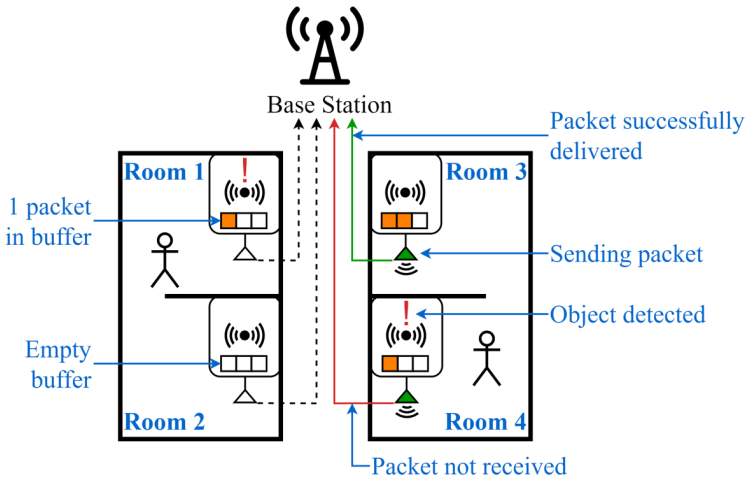
- Exact minimization of the free energy yields the **posterior distribution**.
- Practical solutions are based on **approximate** optimization (variational inference), (Monte Carlo) **sampling**, and hybrid versions thereof.



8 J. Knoblauch, et al, "Generalized variational inference: Three arguments for deriving new posteriors," arXiv:1904.02063, 2019.

Application: Uncertainty-Aware Digital Twins

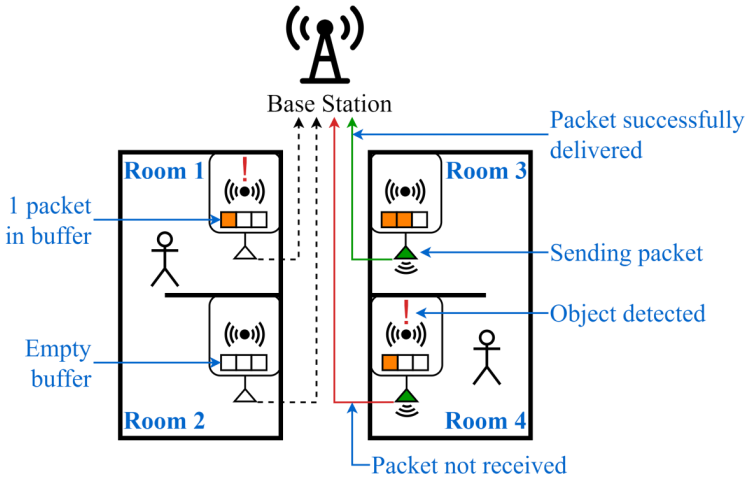
- **Digital twin** platform for the control and monitoring of a wireless **IoT** network.[9]

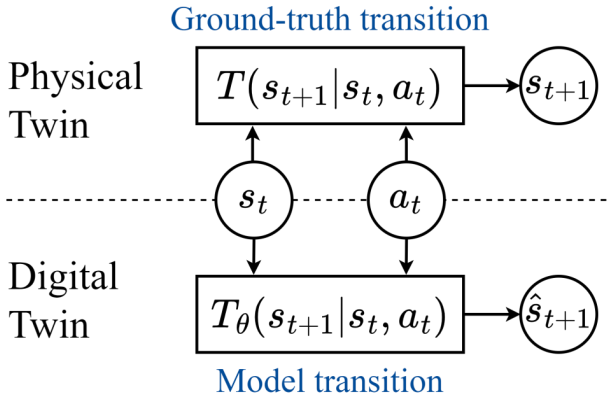


9 C. Ruah, O. Simeone, B. Al-Hashimi, "A Bayesian Framework for Digital Twin-Based Control, Monitoring, and Data Collection in Wireless Systems," arXiv preprint arXiv:2212.01351, 2022.

Uncertainty-Aware Digital Twins

- The digital twin maintains a **model** of the physical twin using **domain knowledge** and **data**.[10]

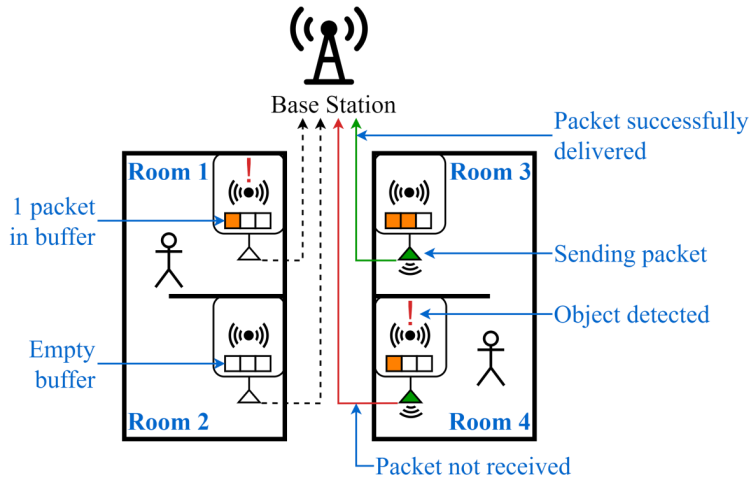


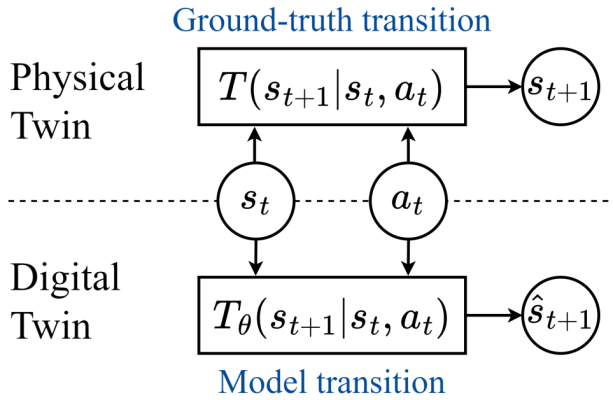


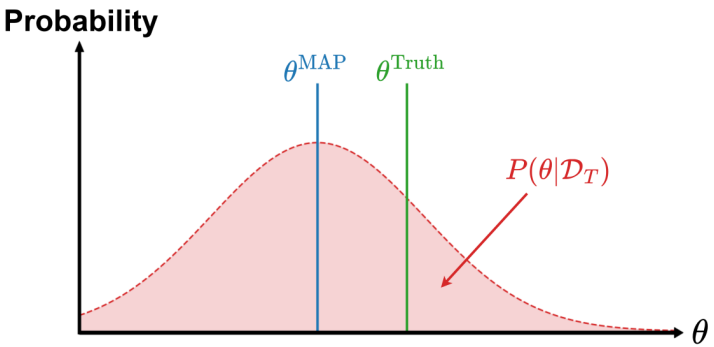
10 C. Ruah, O. Simeone, B. Al-Hashimi, "A Bayesian Framework for Digital Twin-Based Control, Monitoring, and Data Collection in Wireless Systems," arXiv preprint arXiv:2212.01351, 2022.

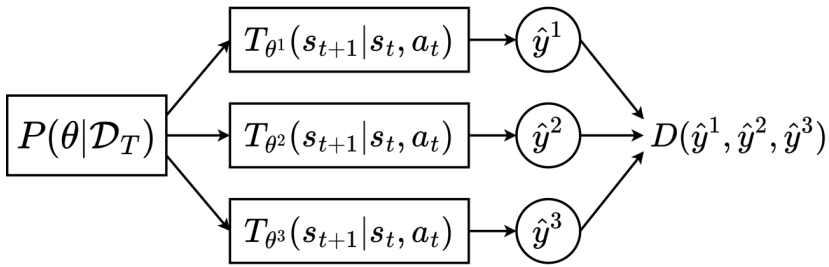
Uncertainty-Aware Digital Twins

- **Bayesian model-based** design (via multi-agent reinforcement learning), prediction, anomaly detection, and directed exploration.[11]



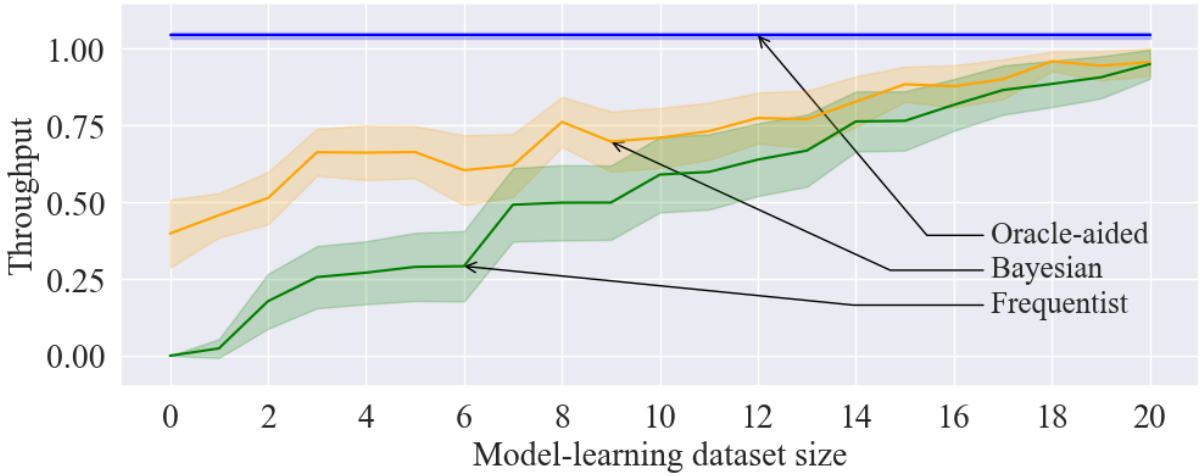




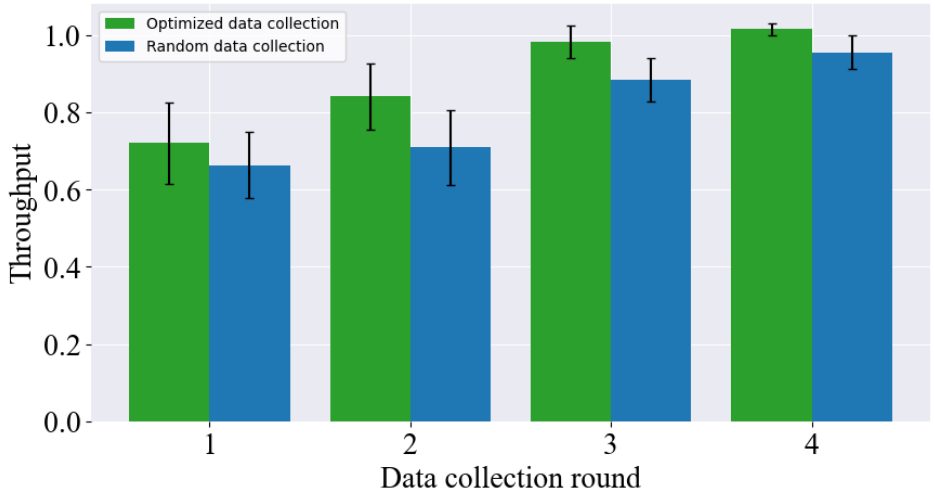


11 C. Ruah, O. Simeone, B. Al-Hashimi, "A Bayesian Framework for Digital Twin-Based Control, Monitoring, and Data Collection in Wireless Systems," arXiv preprint arXiv:2212.01351, 2022.

Uncertainty-Aware Digital Twins for IoT



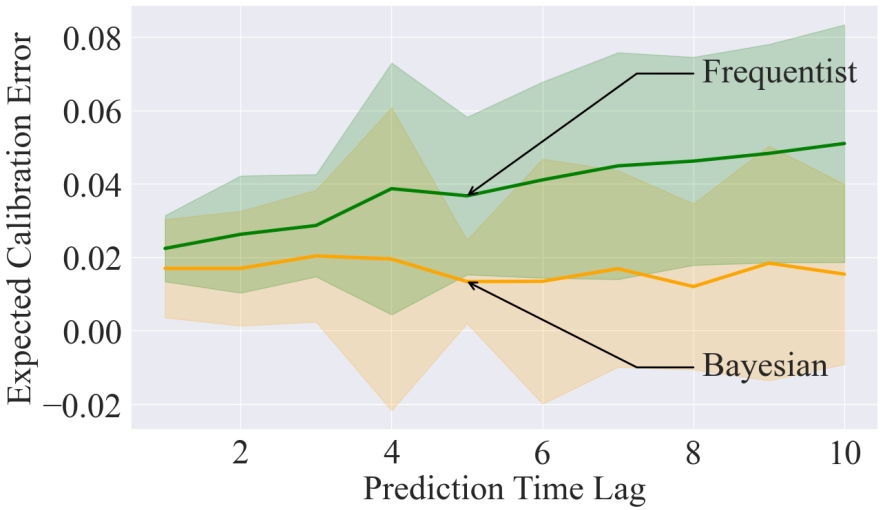
Policy optimization



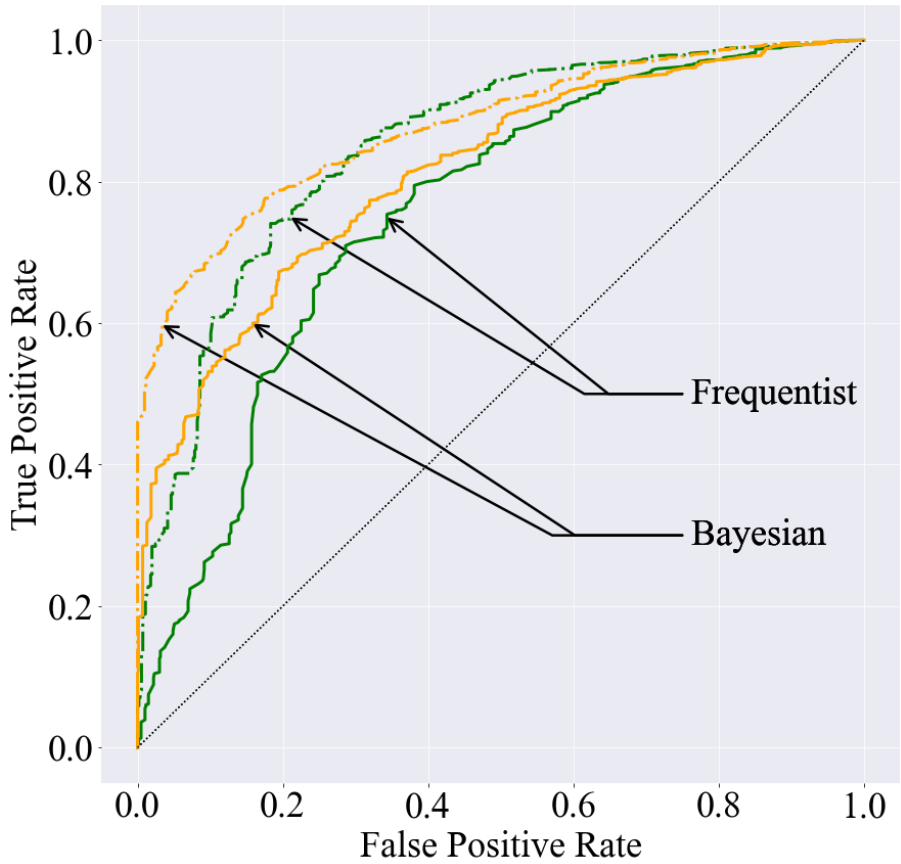
Directed exploration

Uncertainty-Aware Digital Twins for IoT

Prediction

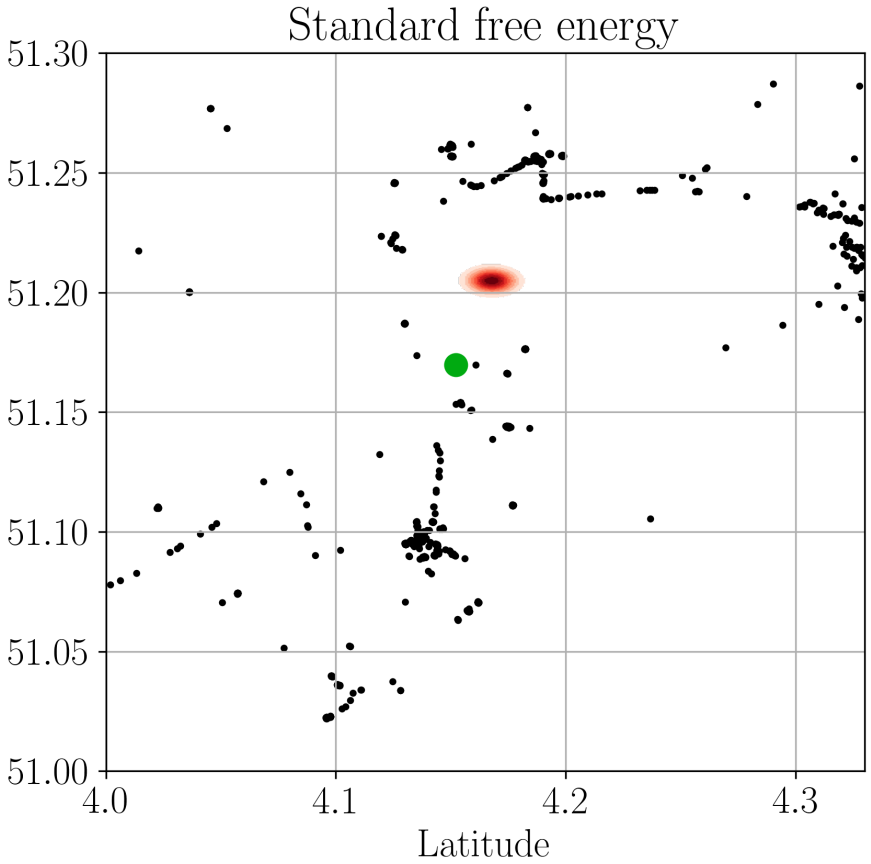


Anomaly detection



The Importance of Model Specification

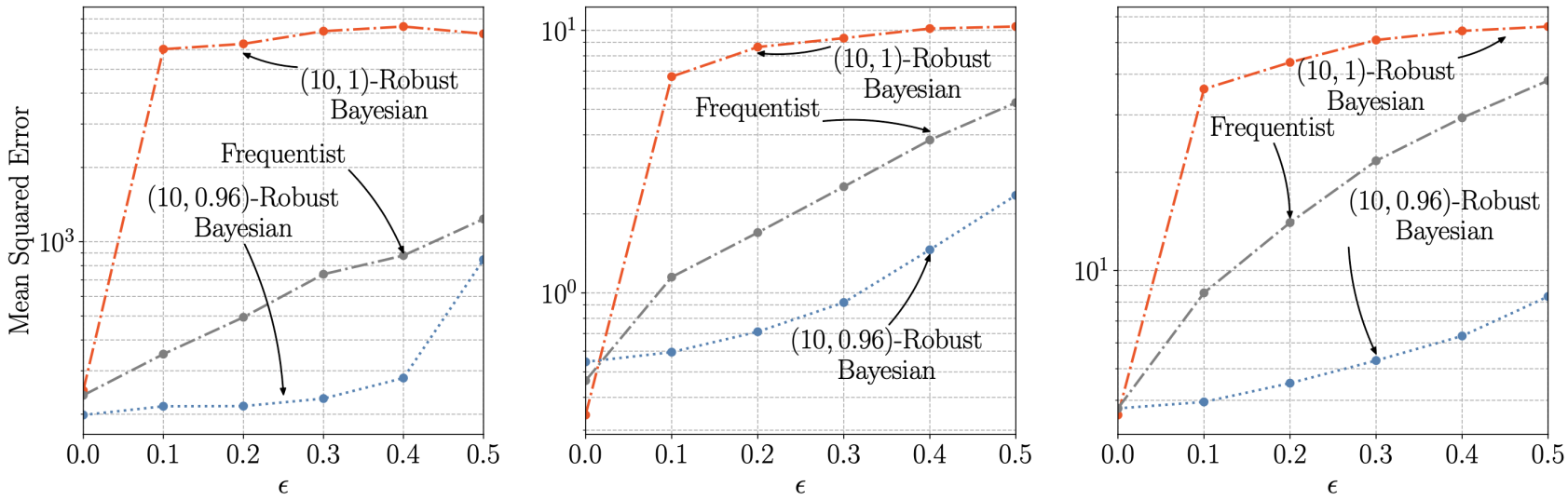
- (Exact) Bayesian learning is only optimal in terms of calibration if the model is **well specified** and if there are no **outliers**.[12]



12 M. Zecchin, S. Park, O. Simeone, M. Kountouris, D. Gesbert, "Robust Bayesian learning for reliable wireless AI: Framework and applications," arXiv preprint arXiv:2207.00300, 2022.

Robust Bayesian Learning

- Robust Bayesian learning:[13]
  - use ensembling not only to quantify uncertainty, but also to increase **expressivity** and mitigate model misspecification
  - adopt an **outlier-robust** loss functions (from KL to α-divergence)
  - derivation of a **generalized free energy** criterion using PAC-Bayes theory



(a) *SigfoxRural* (b) *UTSIndoor* (c) *UJIIndoor*

13 M. Zecchin, S. Park, O. Simeone, M. Kountouris, D. Gesbert, "Robust PAC^m: Training Ensemble Models Under Model Misspecification and Outliers," arXiv preprint arXiv:2203.01859, 2022.

Part I: Algorithms
Post-Hoc Calibration

Post-Hoc Calibration

- Bayesian learning, and robust versions thereof,
  - increase **computational complexity** as compared to conventional learning during both training and inference (due to ensembling)
  - do not provide formal **finite-sample** calibration **guarantees**
- Post-hoc calibration schemes
  - address complexity by operating on a **pre-trained model**
  - can provide formal **finite-sample** calibration guarantees (**conformal prediction**[14,15])

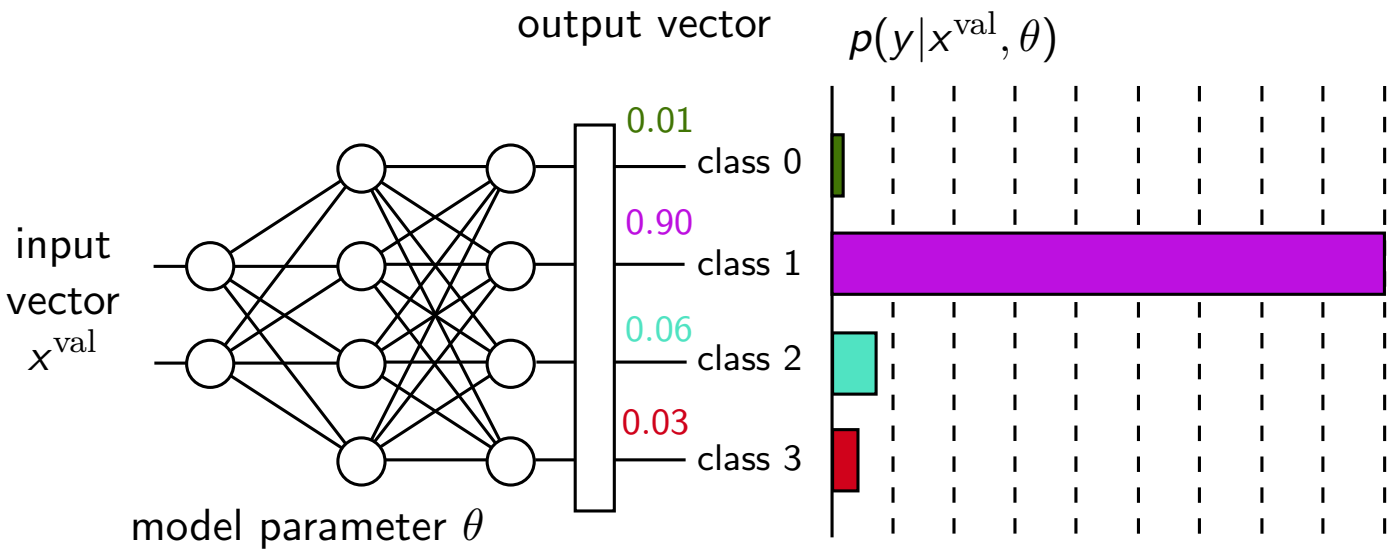
.

[14] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic learning in a random world, Springer, 2023.

[15] J. Cherian and L. Bronner, "How the Washington Post estimates outstanding votes for the 2020 presidential election".

Post-Hoc Calibration

- Some algorithms recalibrate a probabilistic model by matching accuracy estimated on a **validation set**.



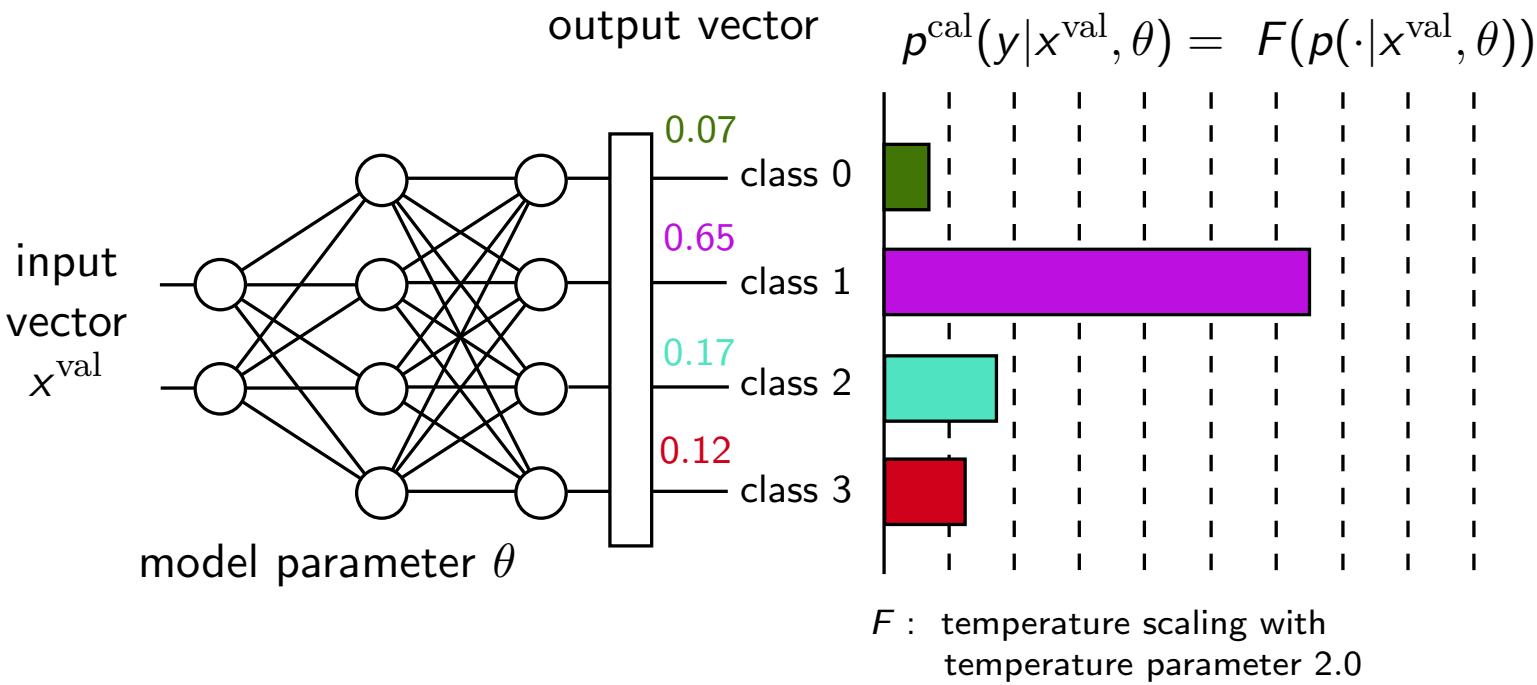
  - Temperature scaling, Platt scaling, isotonic regression
  - No guarantee of calibration: may **overfit the validation set**[16,17]

16 A. Kumar, et al, "Verified Uncertainty Calibration," NeurIPS 2019.

17 X. Ma and M. B. Blaschko, "Meta-Cal: Well-controlled Post-hoc Calibration by Ranking," ICML 2021.

Post-Hoc Calibration

- Some algorithms recalibrate a probabilistic model by matching accuracy estimated on a **validation set**.



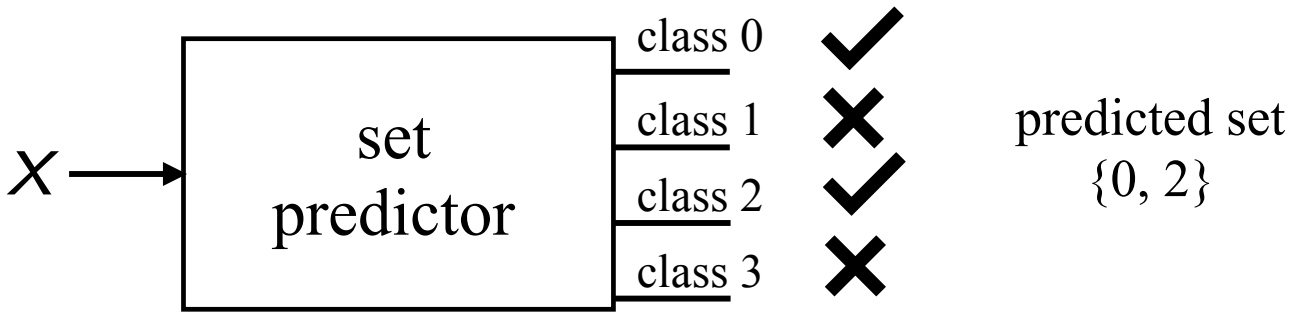
  - Temperature scaling, Platt scaling, isotonic regression
  - No guarantee of calibration: may **overfit the validation set**[16,17]

16 A. Kumar, et al, "Verified Uncertainty Calibration," NeurIPS 2019.

17 X. Ma and M. B. Blaschko, "Meta-Cal: Well-controlled Post-hoc Calibration by Ranking," ICML 2021.

Conformal Prediction

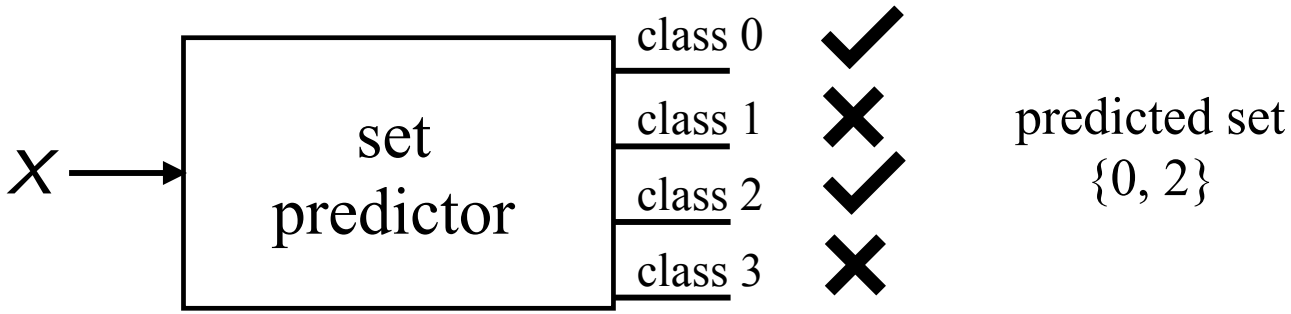
- **Conformal prediction** produces set predictors.
- A **set predictor** is less informative than a probabilistic predictor:
  - Coarser, but easily **interpretable**, measure of **uncertainty** via **set size**
- **Conformal prediction** aims at extracting well-calibrated **set predictors** from probabilistic predictors.[18]



[18] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic learning in a random world, Springer, 2023.

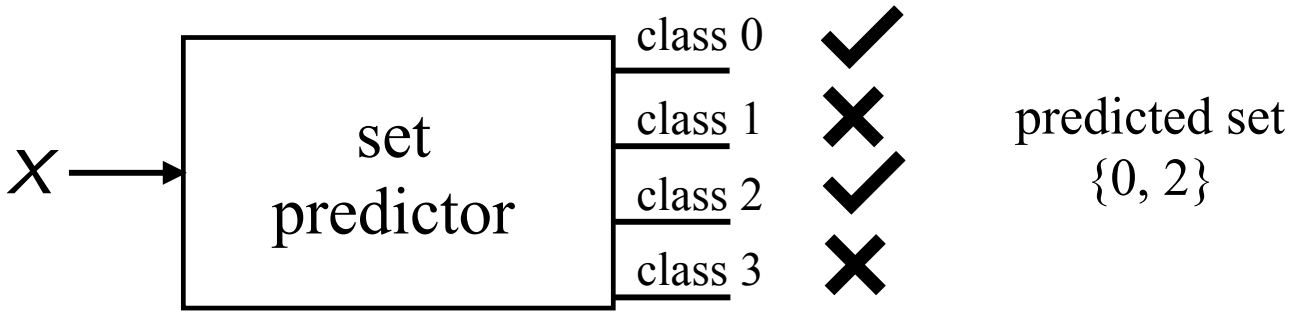
Conformal Prediction

- **Conformal prediction** produces set predictors.
- A **set predictor** is less informative than a probabilistic predictor:
  - Coarser, but easily **interpretable**, measure of **uncertainty** via **set size**
- **Conformal prediction** aims at extracting well-calibrated **set predictors** from probabilistic predictors.[18]



[18] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic learning in a random world, Springer, 2023.

Calibration of Set Predictors



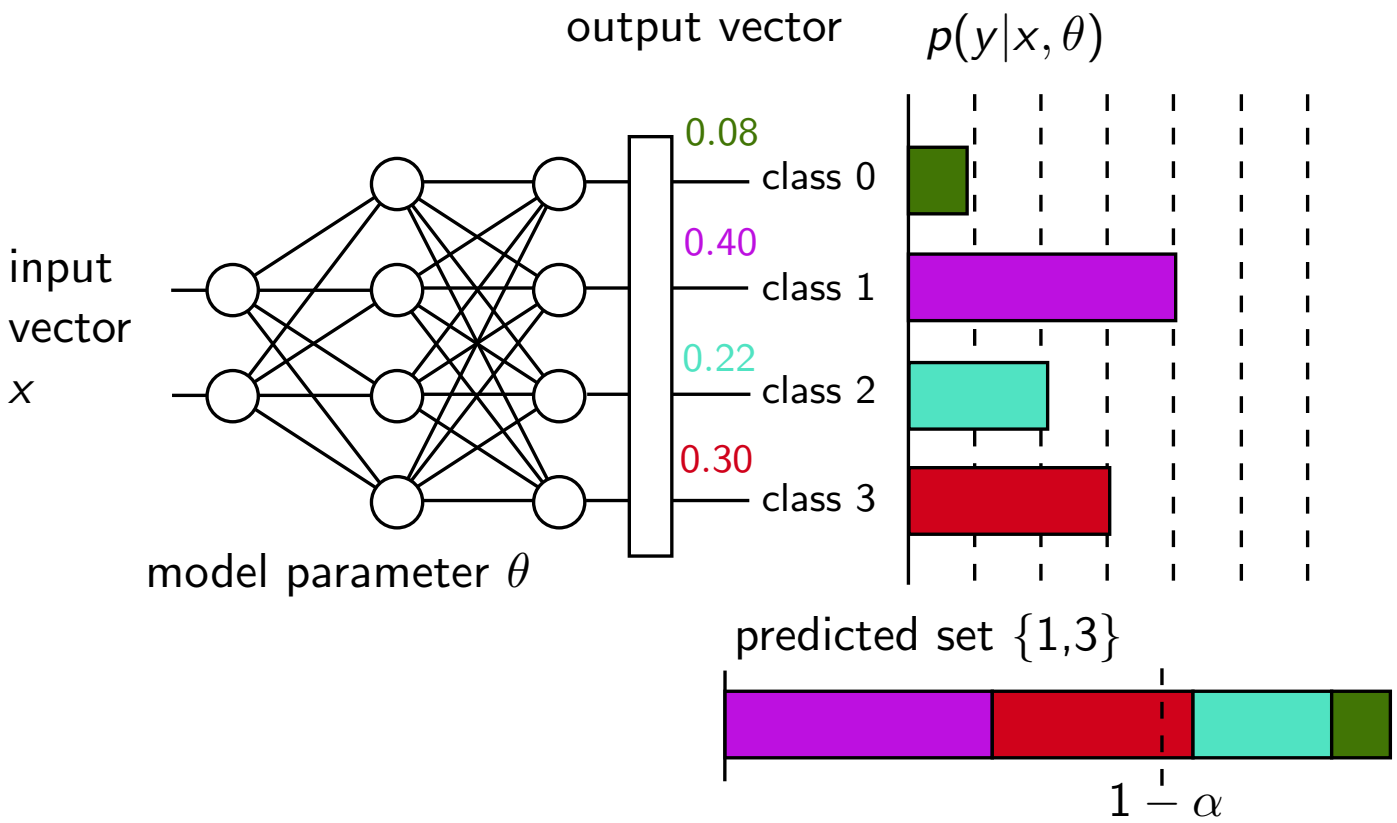
- A set predictor is **well calibrated** if

$$\Pr[\text{true label} \in \text{predicted set}] \geq 1 - \alpha$$

for some desired **coverage** probability $1 - \alpha$.

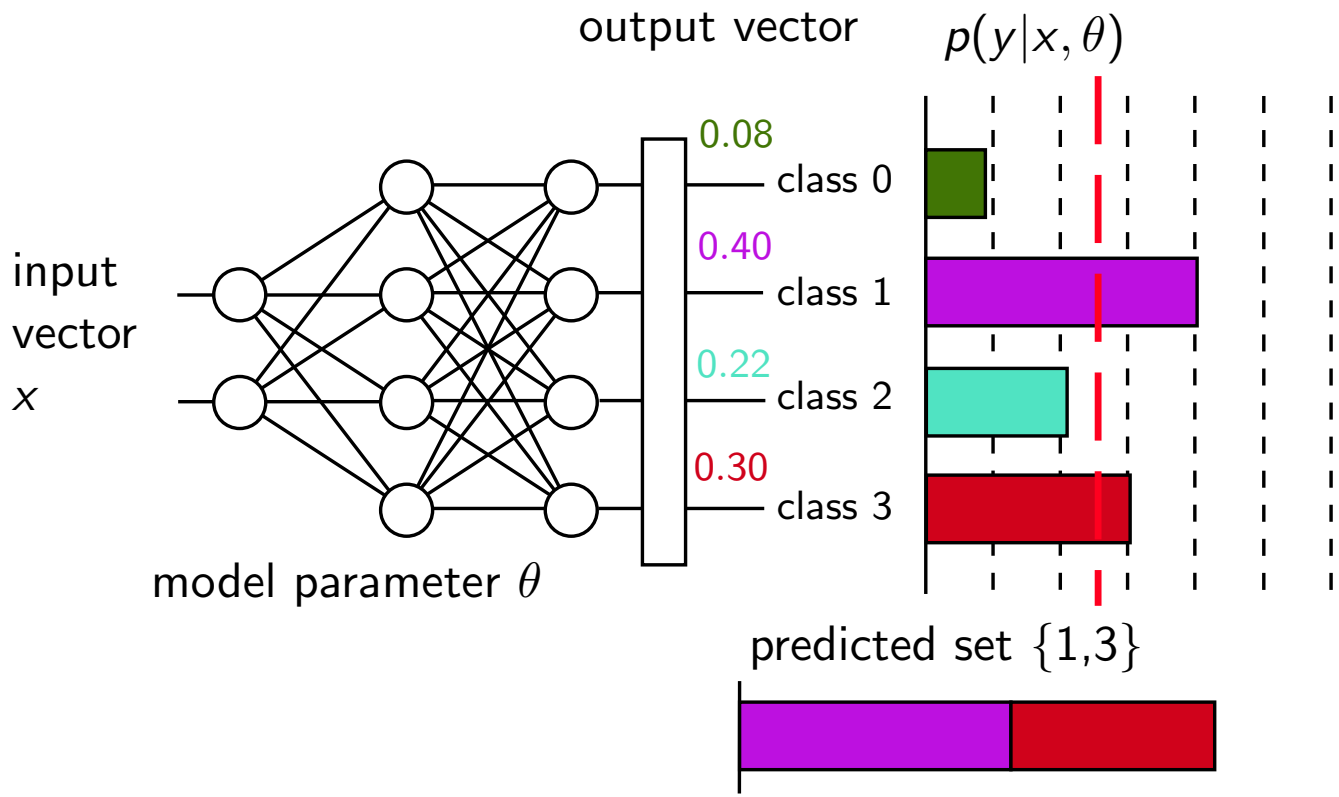
Set Predictors from Probabilistic Predictors

- Well-calibrated probabilistic predictor $\Longrightarrow$ well-calibrated set predictor



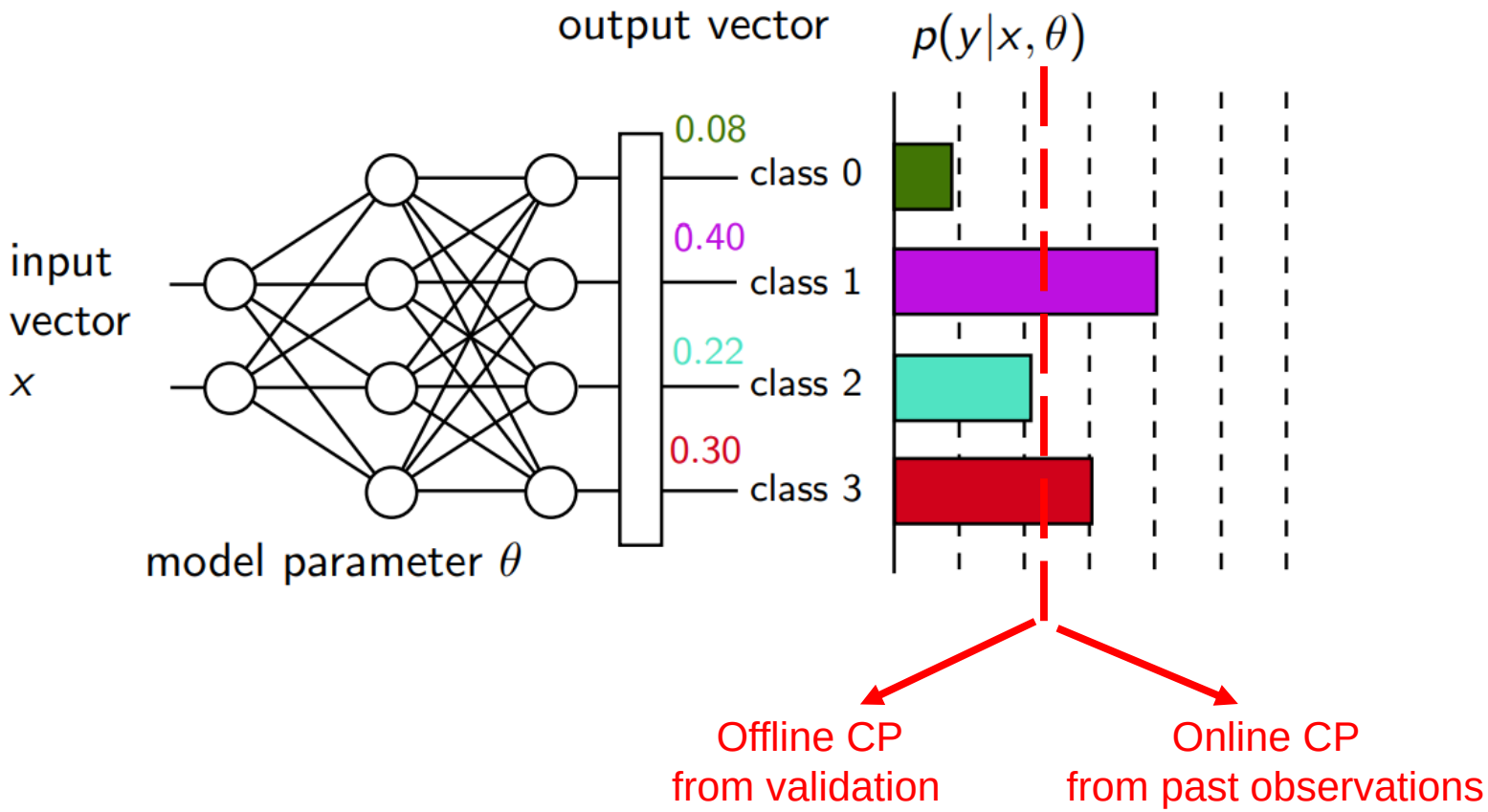
Set Predictors from Probabilistic Predictors

- Well-calibrated probabilistic predictor $\Longrightarrow$ well-calibrated set predictor



Conformal Prediction

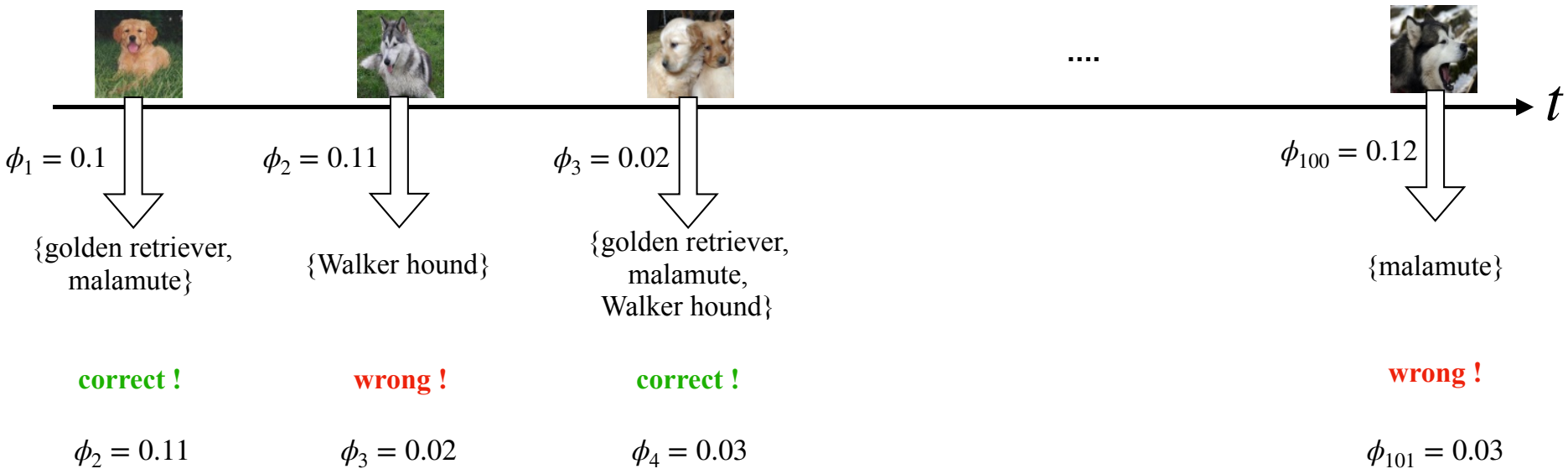
- Conformal prediction applies an **adaptive threshold** to the confidence levels based on additional information.



Online Conformal Prediction

- **Online conformal prediction** adjusts the threshold adaptively based on past errors to minimize the **regret**.[19,20]
  - It can be interpreted as a form of **online gradient descent**.

$$\phi_{t+1} = \phi_t + \gamma(\alpha - \text{err}_t)$$



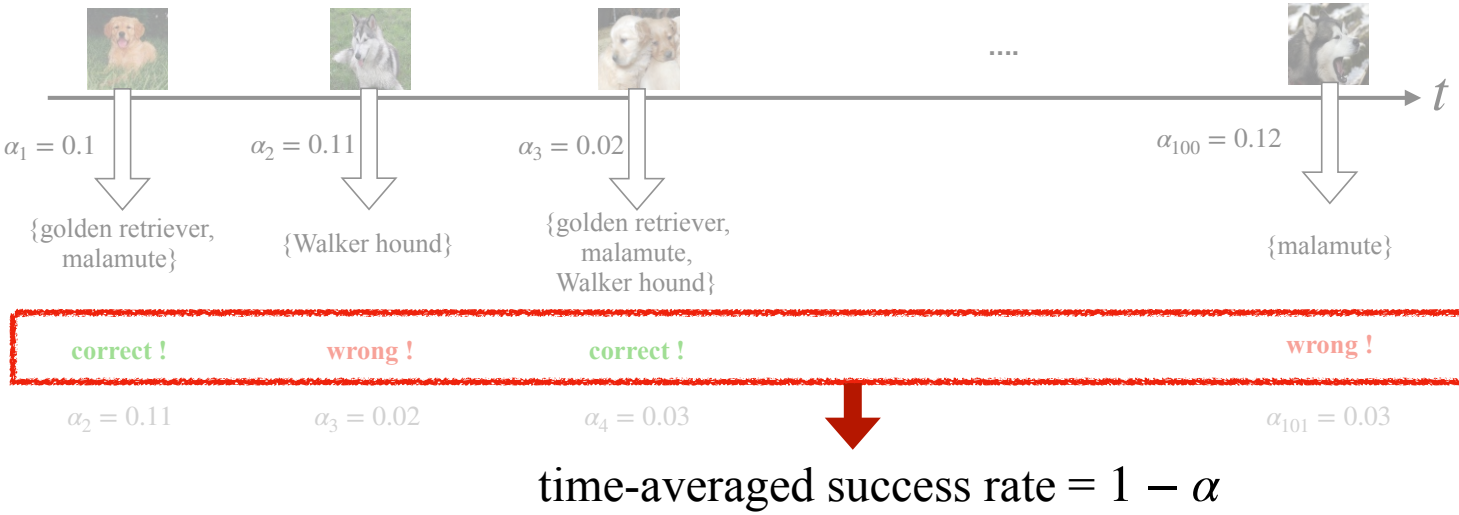
19 I. Gibbs and E. Candes, "Adaptive Conformal Inference Under Distribution Shift," arxiv.org/abs/2106, 2021.

20 S. Feldman, S. Bates, and Y. Romano, "Conformalized Online Learning: Online Calibration Without a Holdout Set," arxiv.org/abs/2205.09095, 2022.

Calibration Guarantees

- Offline conformal yields set predictors that are well-calibrated on average over the generation of test and validation data (for **exchangeable (e.g., i.i.d.)** data).
- Online conformal prediction guarantees calibration on average **over time**.[21]

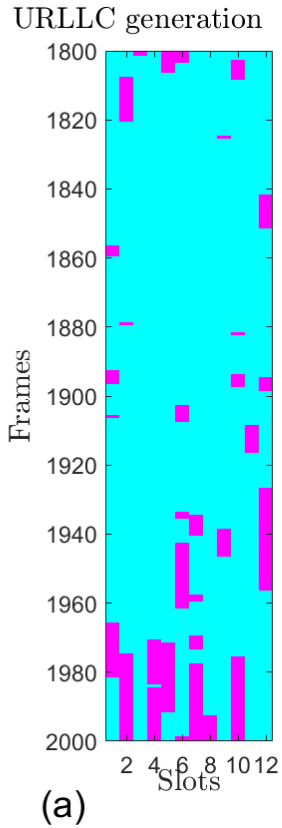
$$\alpha_{t+1} = \alpha_t + \gamma(\alpha - \text{err}_t)$$
$$\iff \alpha_{t+1} \propto t(\alpha - \text{avg. miscoverage rate})$$



21 A. Bhatnagar, H. Wang, C. Xiong, Y. Bai, "Improved Online Conformal Prediction via Strongly Adaptive Online Learning," arXiv:2302.07869, 2023.

Prediction-Based Proactive Resource Allocation

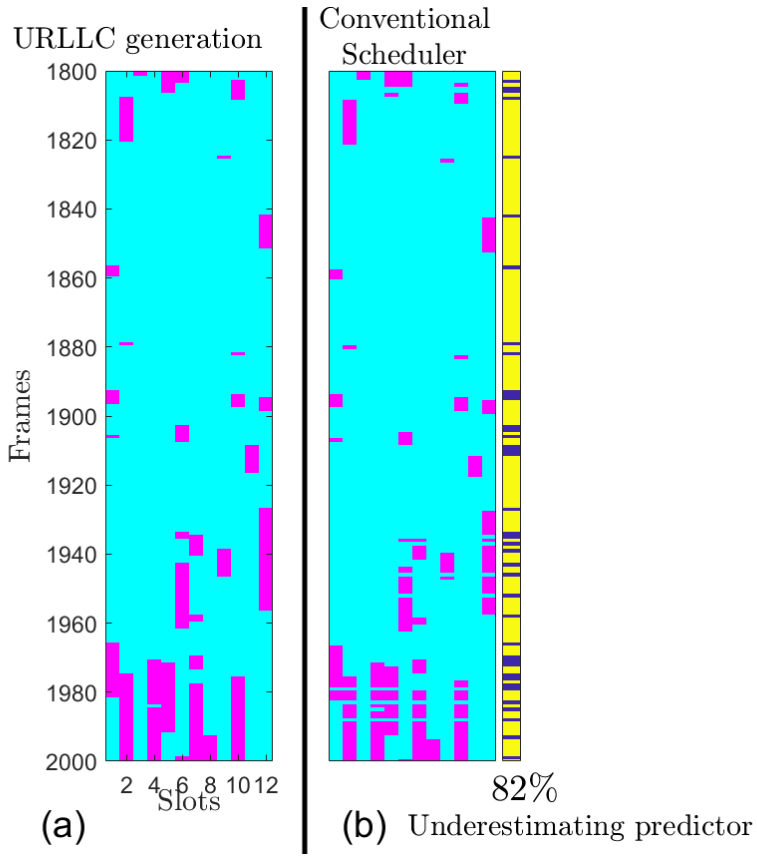
- **Resource allocation** based on the **prediction** of **stochastic** traffic.[22]



22 K. Cohen, S. Park, O. Simeone, P. Popovski, S. Shamai, "Guaranteed Dynamic Scheduling of Ultra-Reliable Low-Latency Traffic via Conformal Prediction," arXiv:2302.07675, 2023.

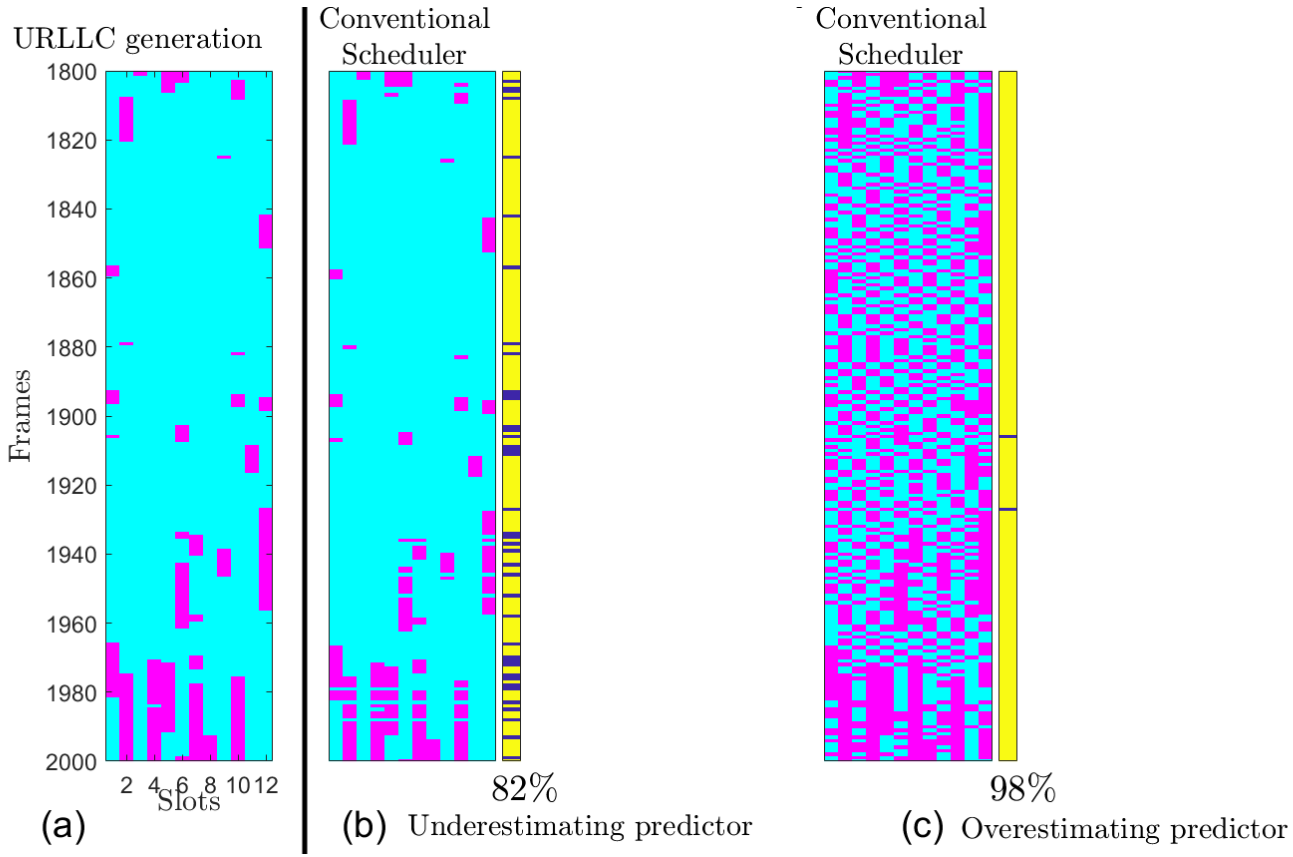
Prediction-Based Proactive Resource Allocation

- A scheduler that **trusts** a non-calibrated predictor may be **unreliable** or inefficient.



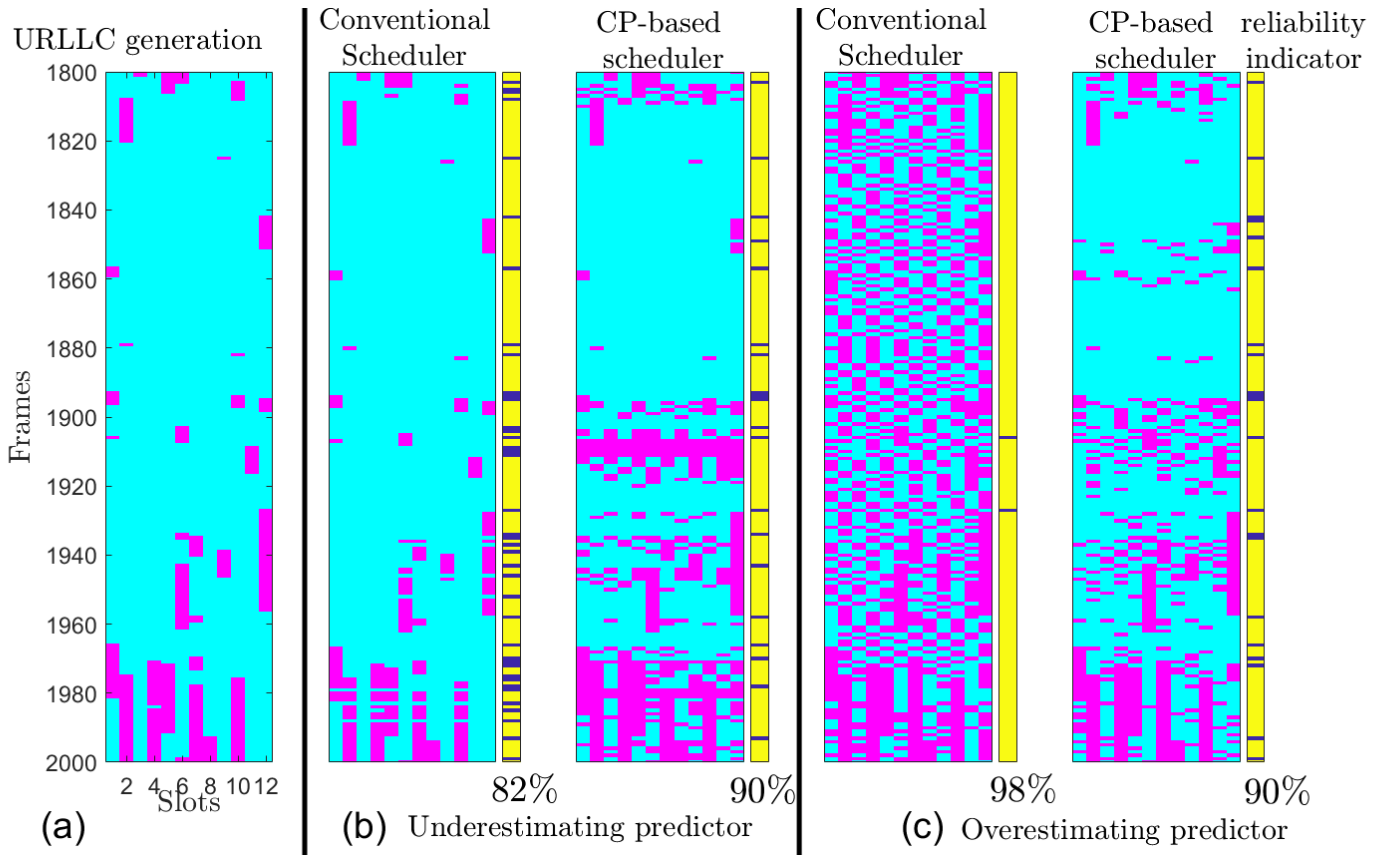
Prediction-Based Proactive Resource Allocation

- A scheduler that **trusts** a non-calibrated predictor may be unreliable or **inefficient**.



Prediction-Based Proactive Resource Allocation

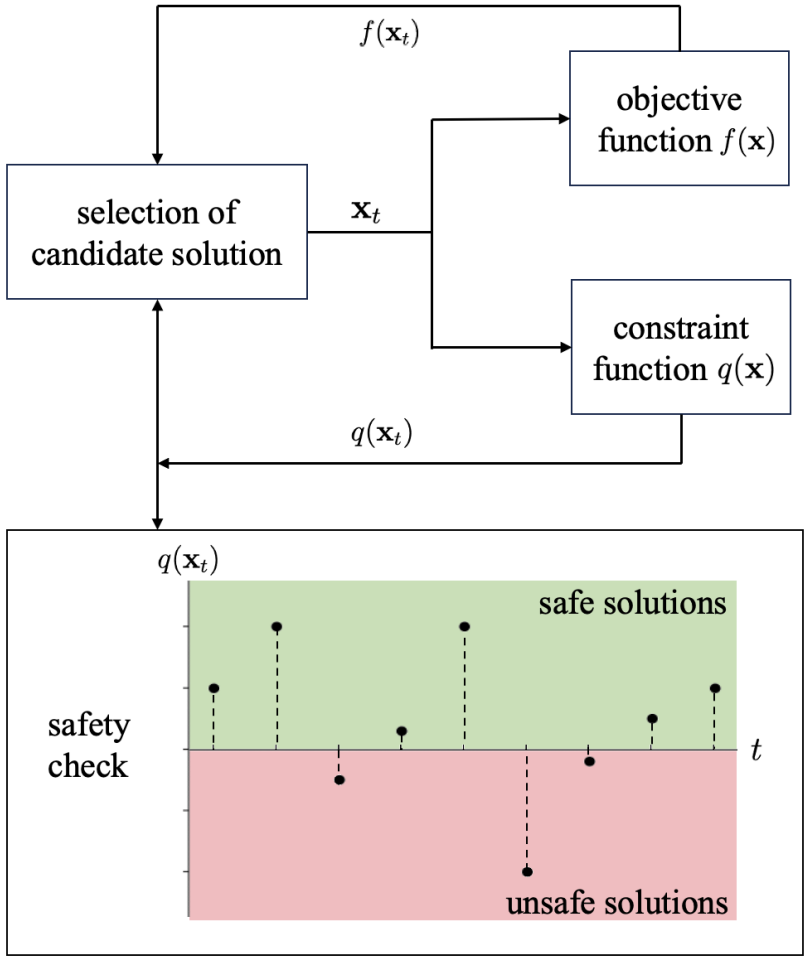
- **Online conformal prediction** allows the implementation of **reliable and efficient** resource allocation schemes irrespective of the calibration of the predictor.[23]



23 K. Cohen, S. Park, O. Simeone, P. Popovski, S. Shamai, "Guaranteed Dynamic Scheduling of Ultra-Reliable Low-Latency Traffic via Conformal Prediction," arXiv:2302.07675, 2023.

Safe Black-Box Optimization

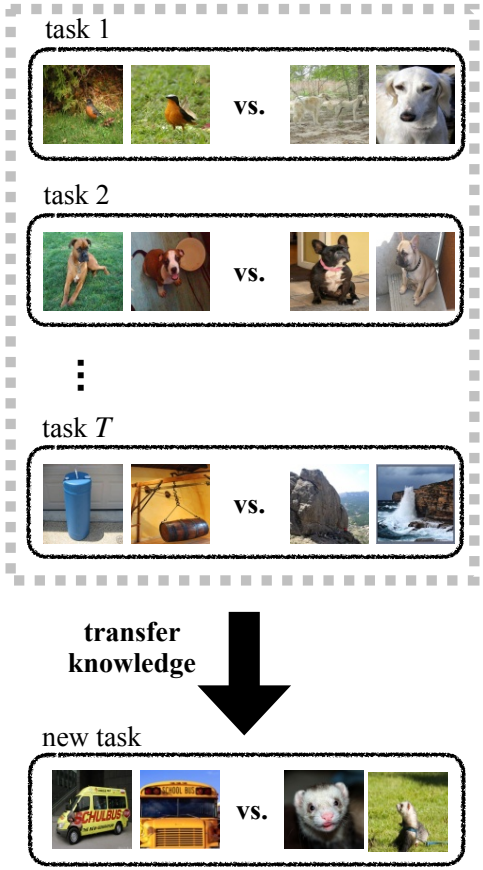
- Black-box optimization with **safety guarantees**.[24]



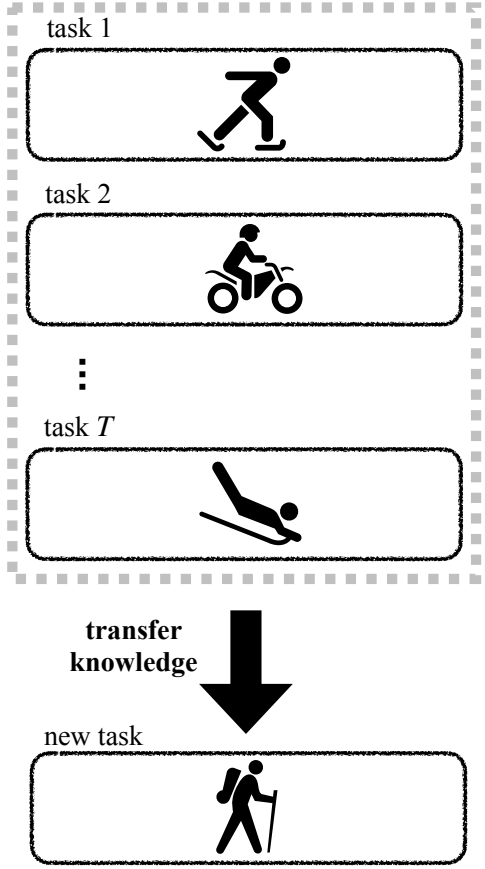
[24] Y. Zhang, S. Park, O. Simeone, "Bayesian Optimization with Formal Safety Guarantees via Online Conformal Prediction,ââ arXiv:2306.17815, 2023.

Meta-Learning for Conformal Prediction

- **Meta-learning** transfers knowledge from multiple tasks to optimize the inductive bias (e.g., model class) for new, related, tasks.[25]



goal: classify better with few samples

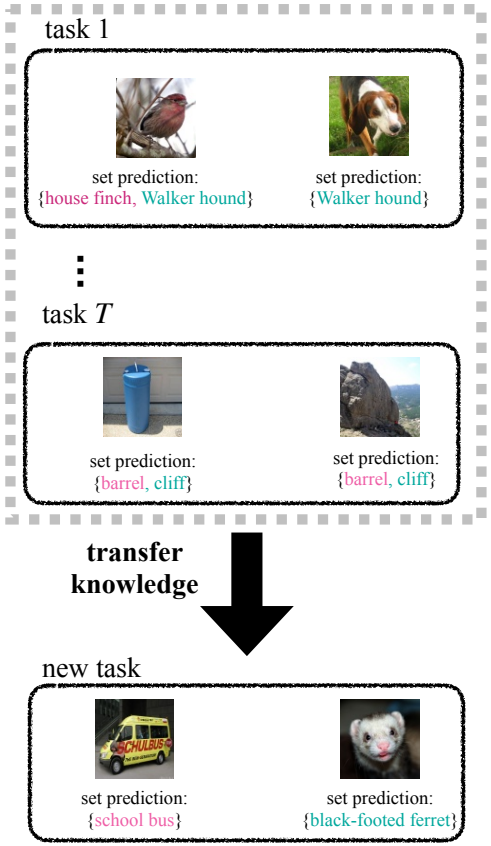


goal: play sports better with less practice

25 L. Chen, et al, "Learning with Limited Samples–Meta-Learning and Applications to Communication Systems," FnT in Signal Processing, 2023.

Meta-Learning for Conformal Prediction

- Meta-learning can enhance **efficiency** for set prediction, while maintaining **per-task formal coverage guarantees**.[26]

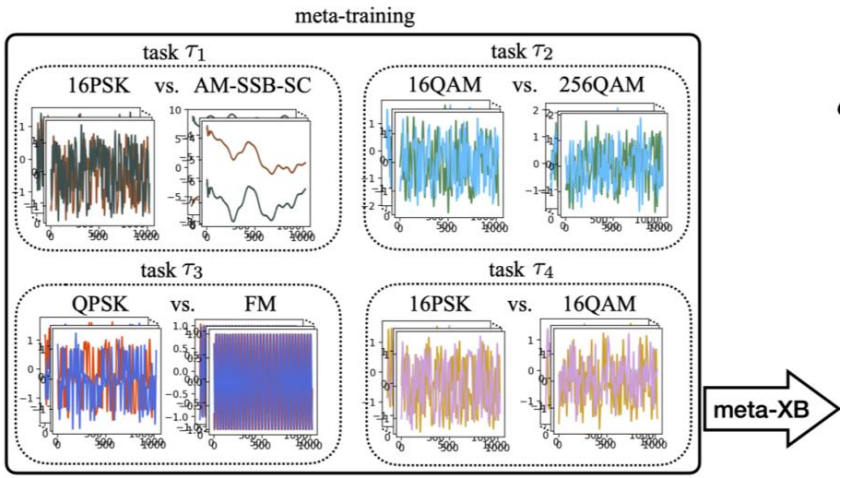


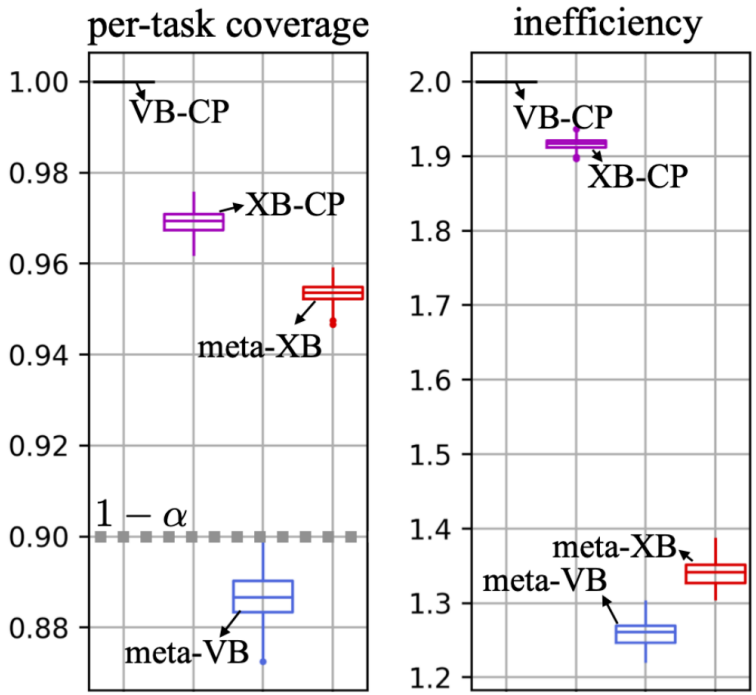
goal: reliable set prediction with small average set prediction size with few samples

26 S. Park, K. Cohen, and O. Simeone, "Few-Shot Calibration of Set Predictors via Meta-Learned Cross-Validation-Based Conformal Prediction," in NeurIPS 2022 Workshop on Meta-Learning.

Meta-Learning for Conformal Prediction

- Application to **modulation classification** (RadioML 2018.01A data set)
- Other applications of CP to communications[27]





27 K. Cohen, S. Park, O. Simeone, S. Shamai, "Calibrating AI Models for Wireless Communications via Conformal Prediction," arXiv:2212.07775, 2022.

Part II: Hardware-Algorithm Co-Design

Hardware-Algorithm Co-Design

- **Ensembling**, which requires the **random** generation of models, enables quantification of uncertainty.
- Implementing ensembling on conventional digital hardware requires the addition of computational resources in order to generate randomness.
- Randomness, however, is inherently present in the physical computing substrate, e.g., thermal noise.

Hardware-Algorithm Co-Design

- Key idea: **Hardware-generated randomness as a computational resource**:
  - In-memory/ neuromorphic computing – noise from analog devices
  - Quantum computing – measurement "noise"
- Same principle applies beyond reliable AI for tasks such as **generative (e.g., diffusion) models**[28]
- A similar principle is to leverage **communication noise** in decentralized AI platforms.[29]

[28] P. Coles, "Thermodynamic AI and the fluctuation frontier", arXiv:2302.09664.

[29] D. Liu and O. Simeone, "Wireless federated Langevin Monte Carlo: Repurposing channel noise for Bayesian sampling and privacy, IEEE Transactions on Wireless Communications, 2022.

Hardware-Algorithm Co-Design

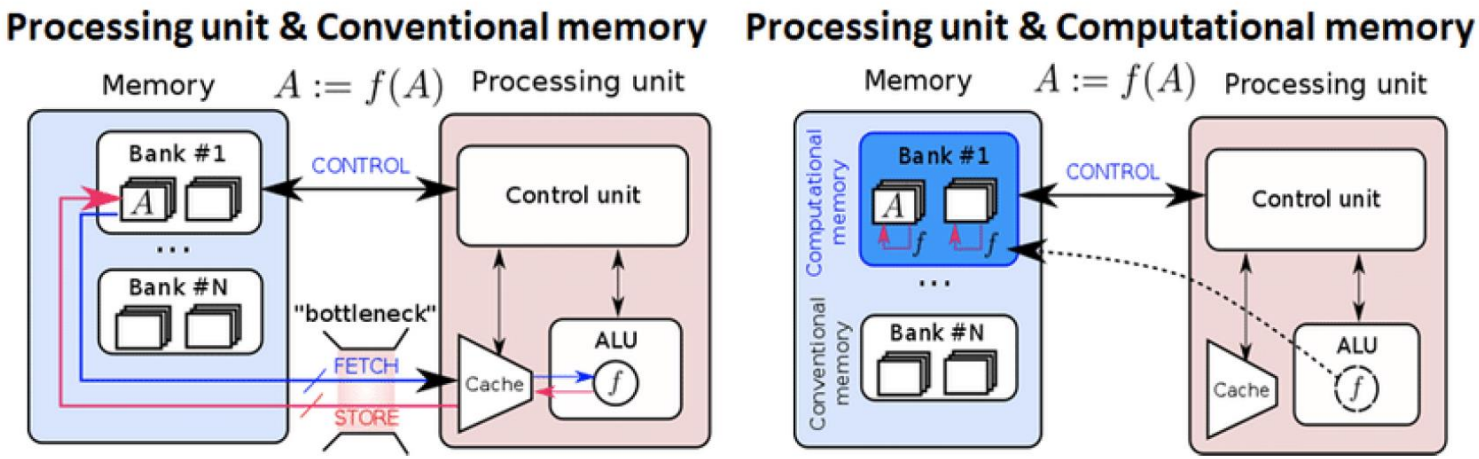
- Key idea: **Hardware-generated randomness as a computational resource**:
  - In-memory/ neuromorphic computing – noise from analog devices
  - Quantum computing – measurement "noise"
- Same principle applies beyond reliable AI for tasks such as **generative (e.g., diffusion) models**[28]
- A similar principle is to leverage **communication noise** in decentralized AI platforms.[29]

28 P. Coles, "Thermodynamic AI and the fluctuation frontier", arXiv:2302.09664.

29 D. Liu and O. Simeone, "Wireless federated Langevin Monte Carlo: Repurposing channel noise for Bayesian sampling and privacy, IEEE Transactions on Wireless Communications, 2022.

In-Memory Computing

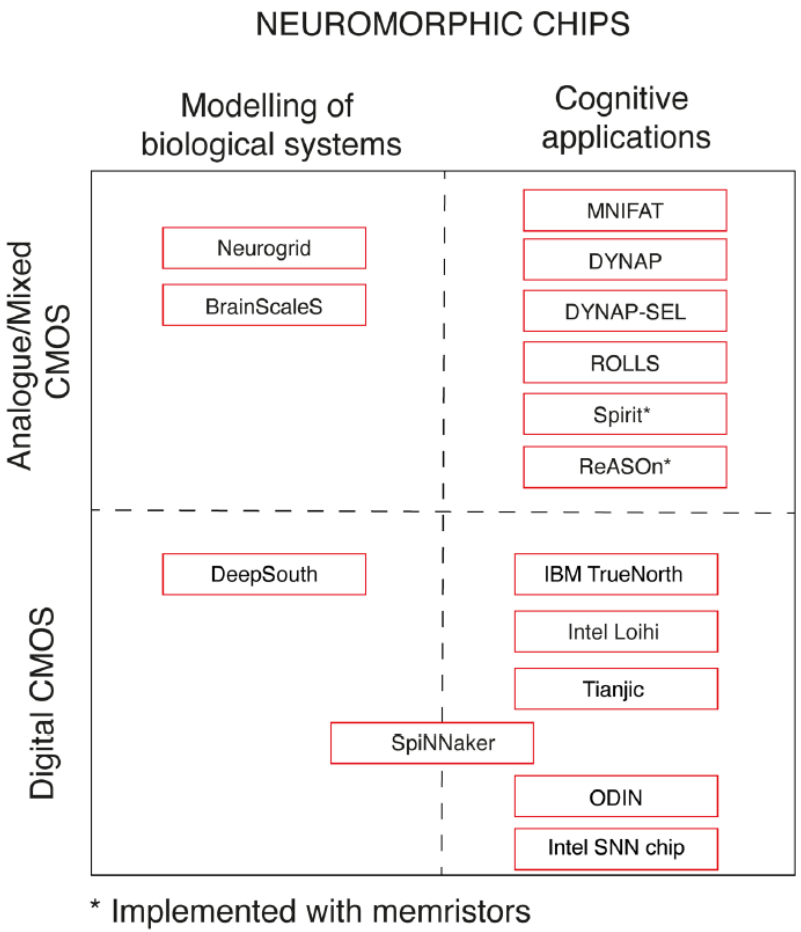
- For **data-intensive** workloads, the conventional separate memory-processor architecture is limited by the **"von Neumann bottleck"**...
- ... and **in-memory computing** can enhance energy, area, and time efficiency.[30]
- Also referred to as **neuromorphic** computing.



30 A. Mehonic, A. Sebastian, B. Rajendran, O. Simeone, E. Vasilaki, A. Kenyon, "Memristors – From In-Memory Computing, Deep Learning Acceleration, and Spiking Neural Networks to the Future of Neuromorphic and Bio-Inspired Computing," Advanced Intelligent Systems, 2020.

Neuromorphic Computing

- In-memory computing systems adopt **digital** or **mixed analog-digital** technology to implement **nonvolatile memory devices**.



PROPERTIES:

- In-memory computing
- Fine-grained parallelism
- Learning in hardware
- Event based and asynchronous communication
- Reduced precision
- Spike-based processing
- Adaptability
- Leveraging noise and stochasticity

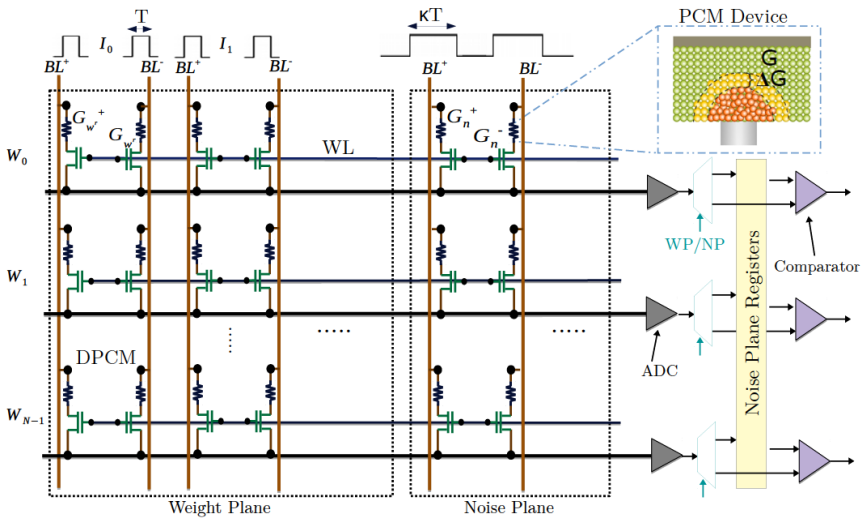
.
.
.

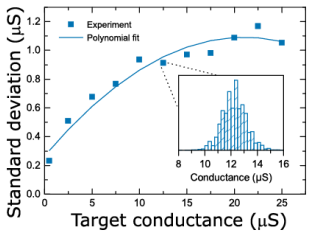
- Brain-inspired

[Mehonic and Kenyon '21]

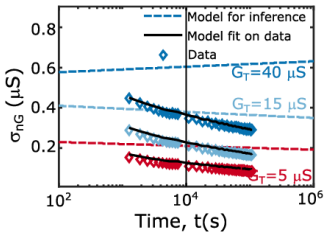
Hardware Noise as a Computational Resource

- **Mixed analog-digital systems** can be more energy and area efficient, but they are subject to hardware noise arising from analog devices.
- Noise is typically **mitigated** via, e.g., averaging.
- For **crossbar** arrays of PCM devices, leverage hardware noise as a resource for **synaptic sampling** implementing Bayesian ensembling.[31]





(a) Programming noise

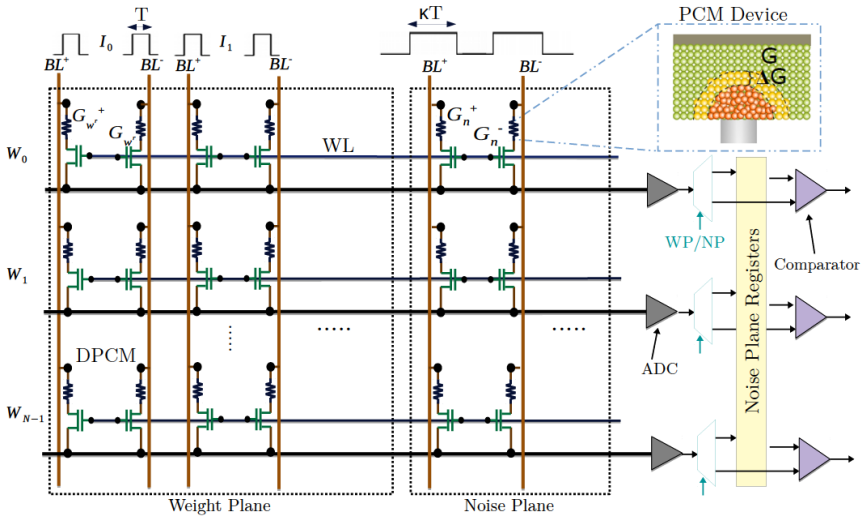


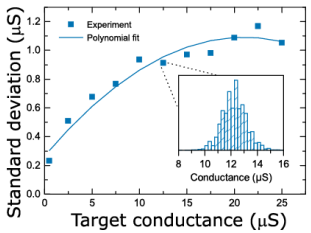
(b) Read noise

[31] P. Katti, N. Skatchkovsky, O. Simeone, B. Rajendran, B. M. Al-Hashimi, "Bayesian Inference on Binary Spiking Networks Leveraging Nanoscale Device Stochasticity," in Proc. ISCAS 2023.

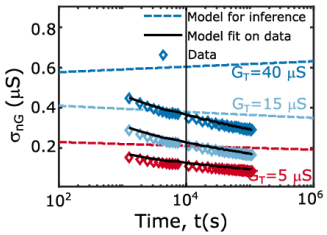
Hardware Noise as a Computational Resource

- **Mixed analog-digital systems** can be more energy and area efficient, but they are subject to hardware noise arising from analog devices.
- Noise is typically **mitigated** via, e.g., averaging.
- For **crossbar** arrays of PCM devices, leverage hardware noise as a resource for **synaptic sampling** implementing Bayesian ensembling.[31]





(a) Programming noise

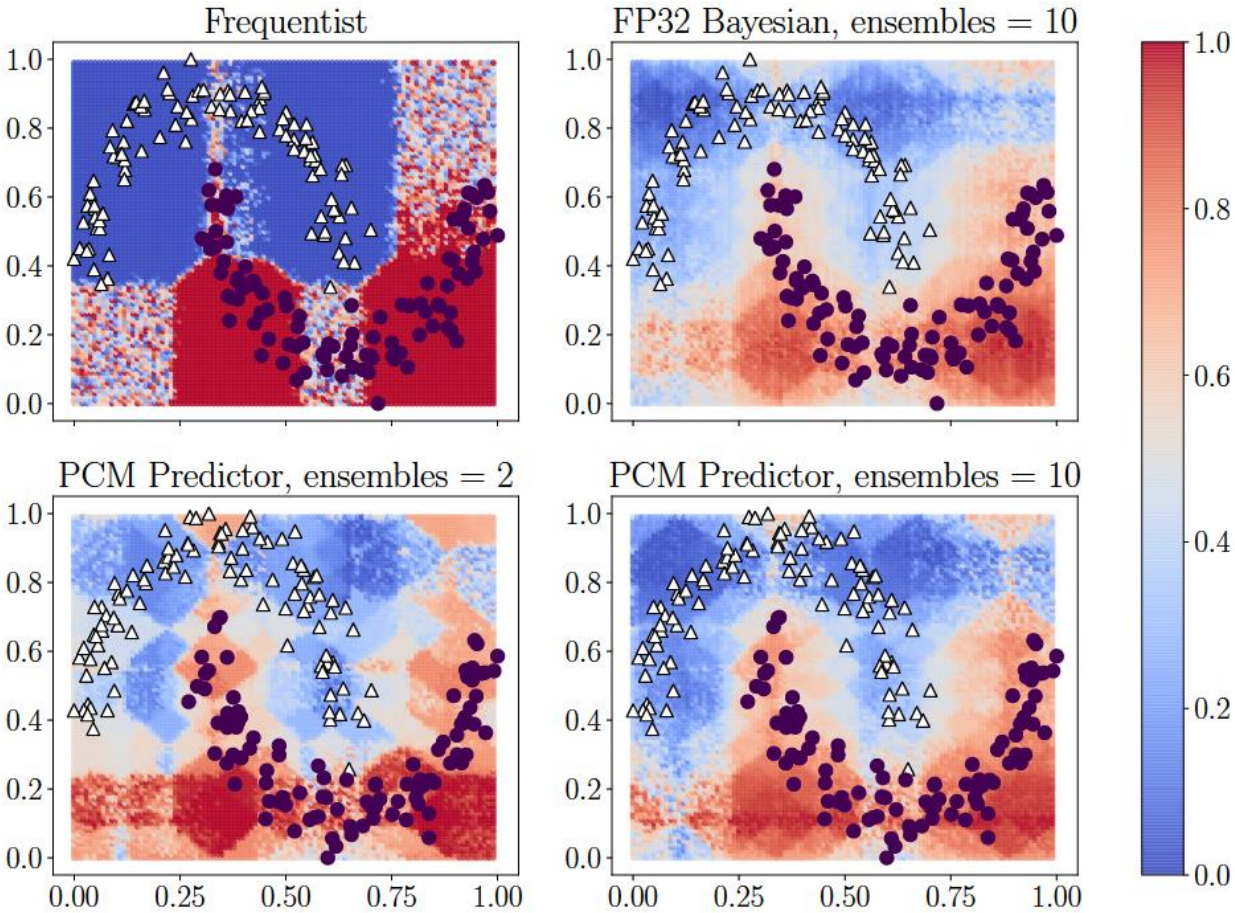


(b) Read noise

31 P. Katti, N. Skatchkovsky, O. Simeone, B. Rajendran, B. M. Al-Hashimi, "Bayesian Inference on Binary Spiking Networks Leveraging Nanoscale Device Stochasticity," in Proc. ISCAS 2023.

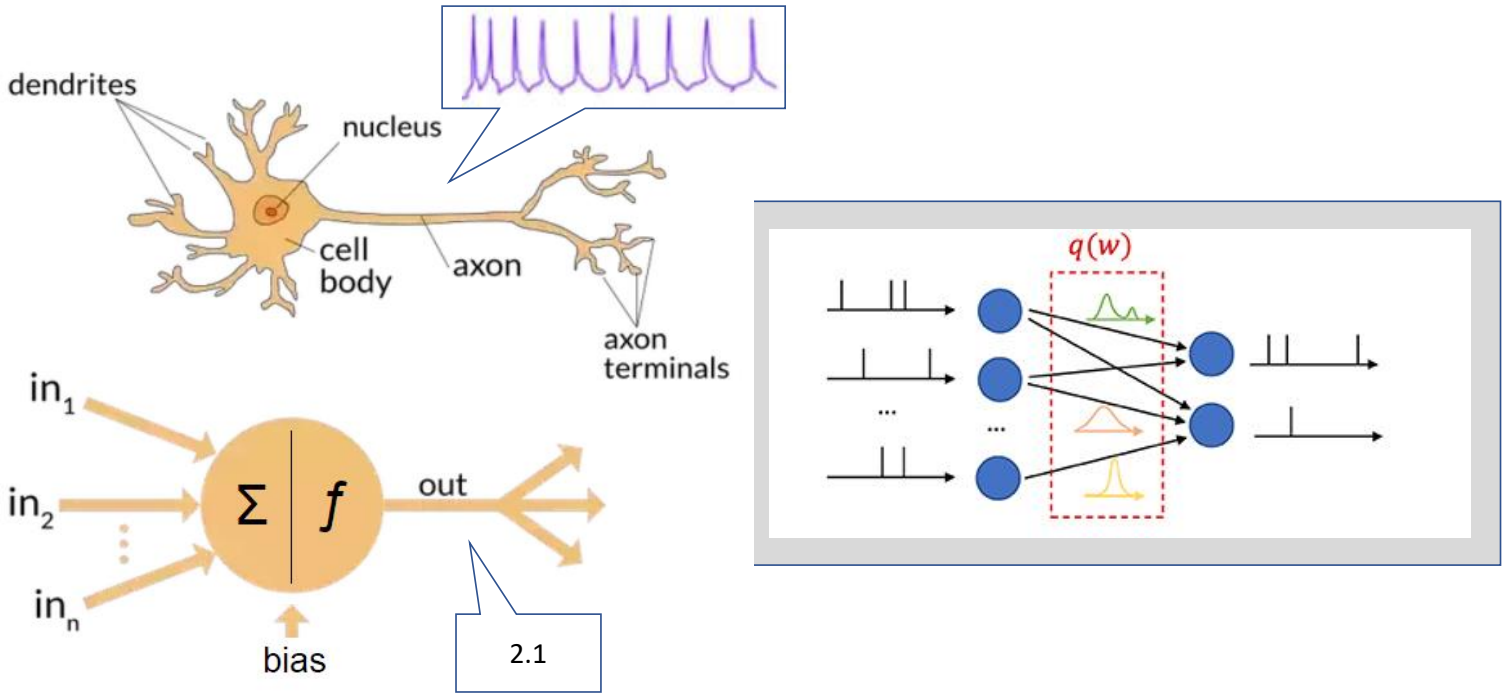
Hardware Noise as a Computational Resource

- Through **hardware-algorithm co-design**, accuracy and calibration performance matches digital (SRAM) baseline implementation, with significant projected savings in core area transistor count



Spiking Neural Networks

- Many neuromorphic chips implement **spiking neural networks (SNNs)**.[32,33]



32 H. Jang, O. Simeone, B. Gardner, and A. Gruning, "An Introduction to Spiking Neural Networks," IEEE Signal Processing Magazine, 2019.

33 N. Skatchkovsky, H. Jang, and O. Simeone, "Bayesian Continual Learning via Spiking Neural Networks," Frontiers in Neuroscience, Nov. 2022.

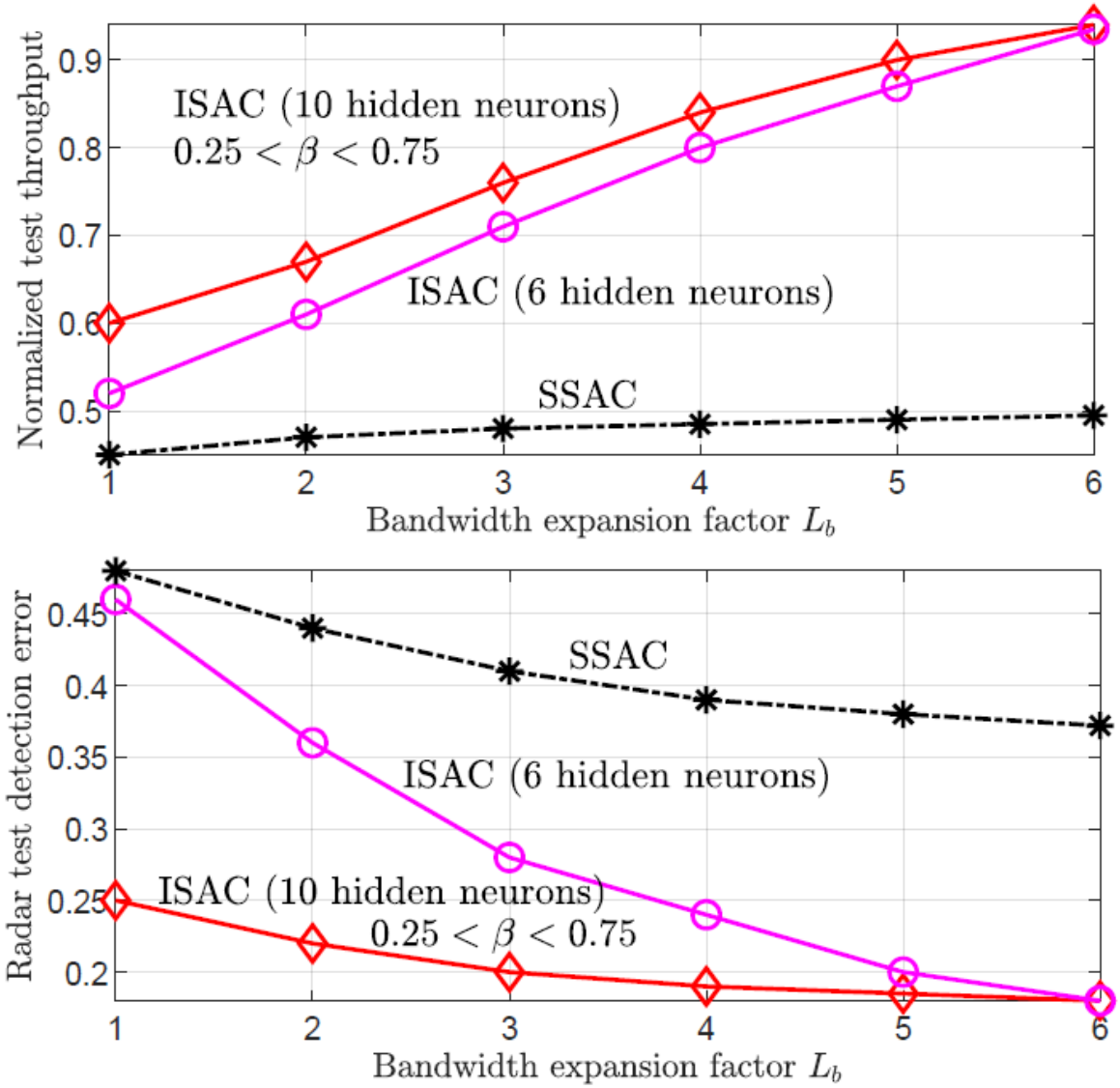
Spiking Neural Networks for 6G

- Synergy with neuromorphic sensors (e.g., Sony's event-based vision sensors) and with impulse radio[34]
- Application to integrated sensing and communications[35]

34 J. Chen, N. Skatchkovsky, O. Simeone, "Neuromorphic wireless cognition...," IEEE Transactions on Cognitive Communications, 2023.

35 J. Chen, N. Skatchkovsky, O. Simeone, "Neuromorphic Integrated Sensing and Communications," IEEE Wireless Communications Letters, 2022.

Spiking Neural Networks for 6G



Conclusions

Conclusions

- AI systems in engineering should "know when they don't know".
- Ensuring calibration calls for the development of novel algorithmic frameworks and hardware-algorithm co-design paradigms.
- Directions for future work:
  - Interplay between calibration and **privacy**[36], **fairness**[37], and **explainability**[38]
  - Calibration and **large language models**, with applications[39]
  - Uses cases for the adoption of **neuromorphic** and quantum computing in engineering[40]

36 D. Liu and O. Simeone, "Privacy for Free: Wireless Federated Learning ...," IEEE JSAC, 2020.

37 G. Pleiss et al, "On fairness and calibration," in Proc. NeurIPS 2017.

38 L. Schut, et al, "Generating interpretable counterfactual explanations...," in AISTATS 2021.

39 V. Quach, "Conformal Language Modeling," arXiv:2306.10193

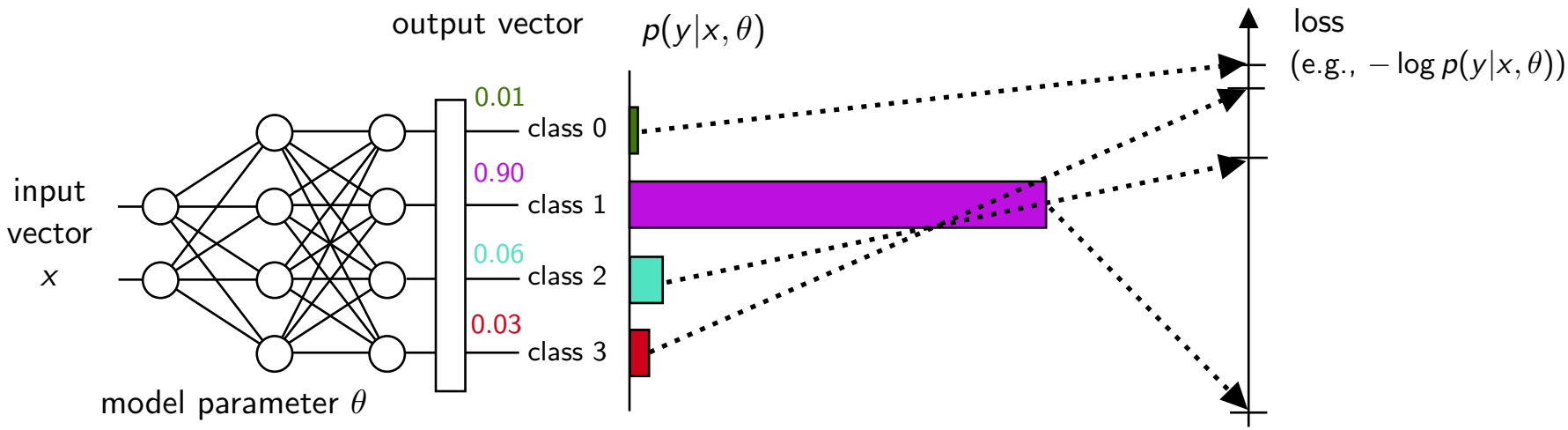
40 J. Chen, "Neuromorphic Wireless Cognition: Event-Driven Semantic Communications for Remote Inference ," IEEE Trans. Cognitive Comm. and Networking, 2022

Acknowledgements

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 725731), by the European Union's Horizon Europe project CENTRIC (101096379), and by an Open Fellowship of the EPSRC (EP/W024101/1).

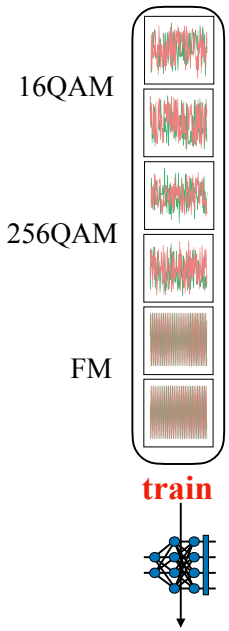
Offline Conformal Prediction

- Loss (scoring function) = decreasing function of the confidence
- **Offline conformal prediction** includes in predictive set labels with loss smaller than a fraction $1 - \alpha$ of examples in the validation set.



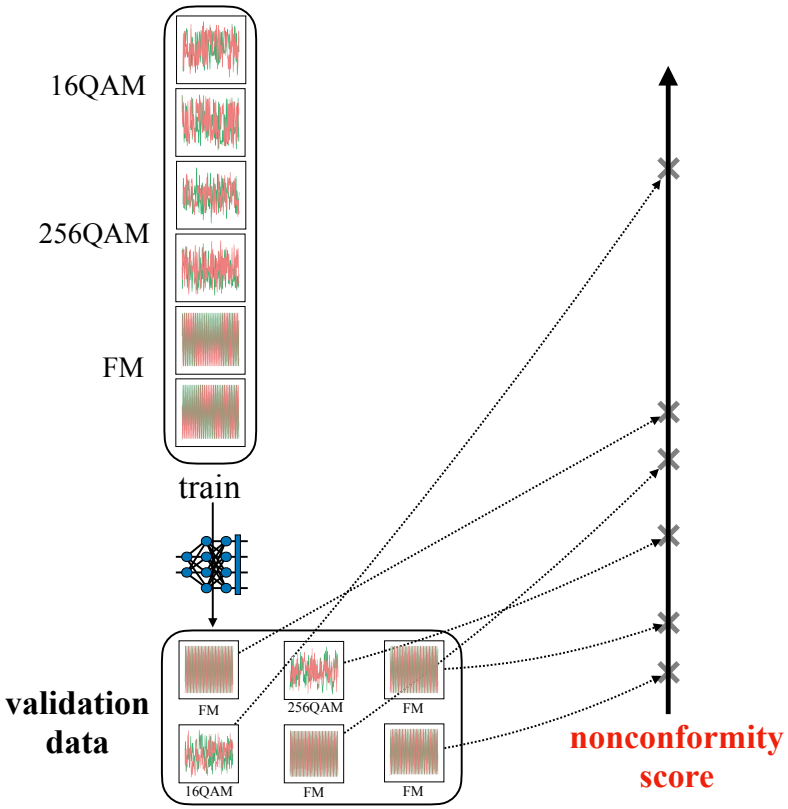
Offline Conformal Prediction

- **1.** Train a probabilistic model with **any** algorithm



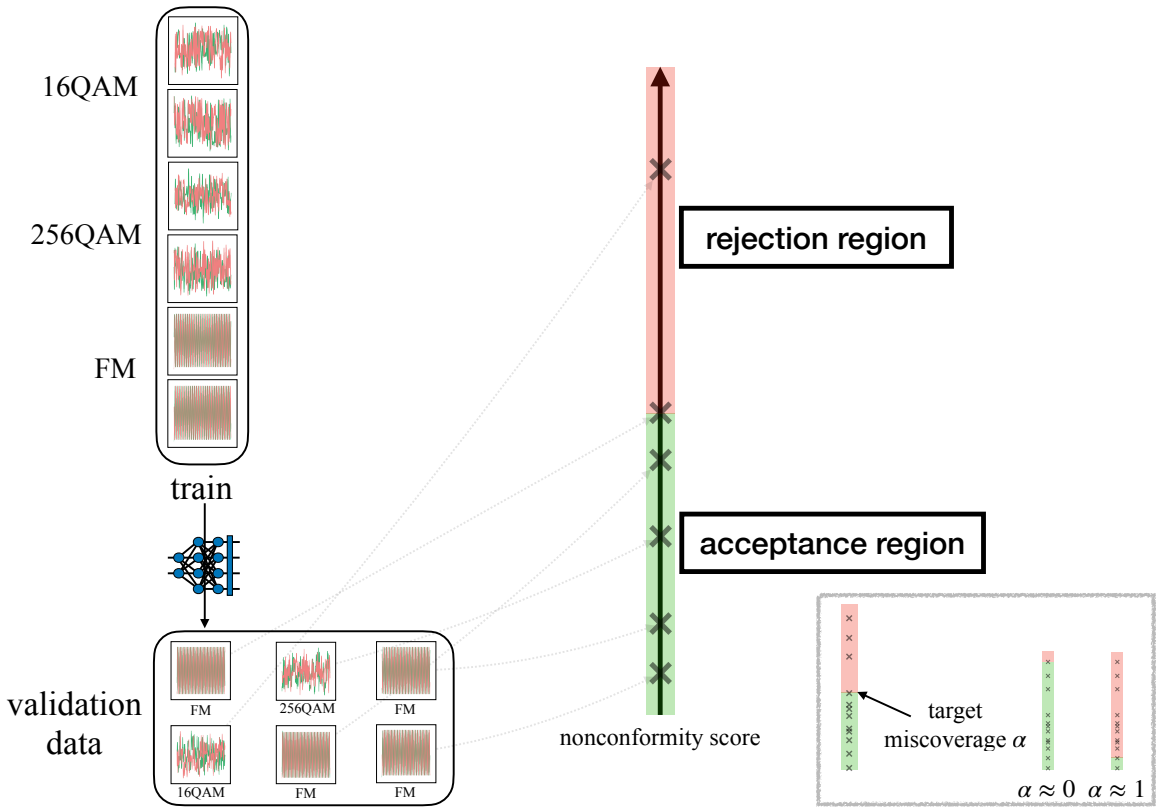
Offline Conformal Prediction

- **2.** Using the trained model, evaluate a **nonconformity score** via a scoring rule (e.g., log-loss) for each validation example



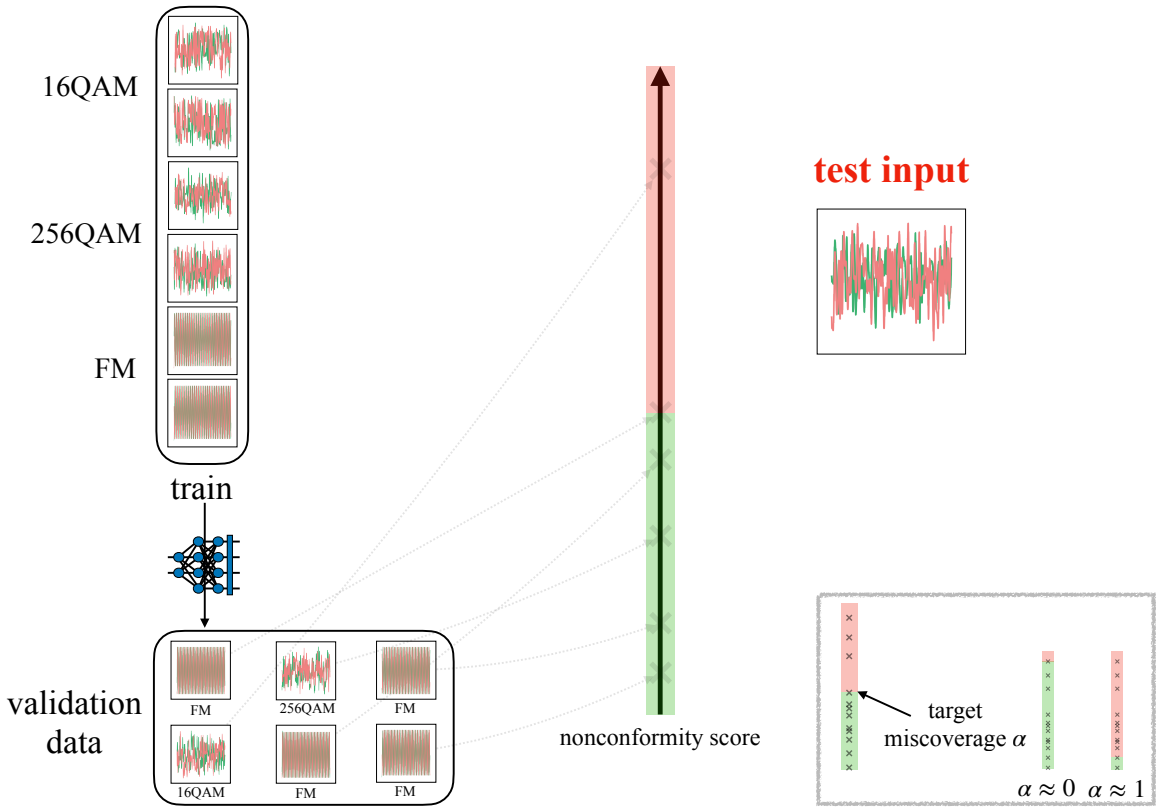
Offline Conformal Prediction

- **3.** Based on the target coverage level $1 - \alpha$, determine **acceptance/rejection** regions using empirical quantile of the nonconformity scores.



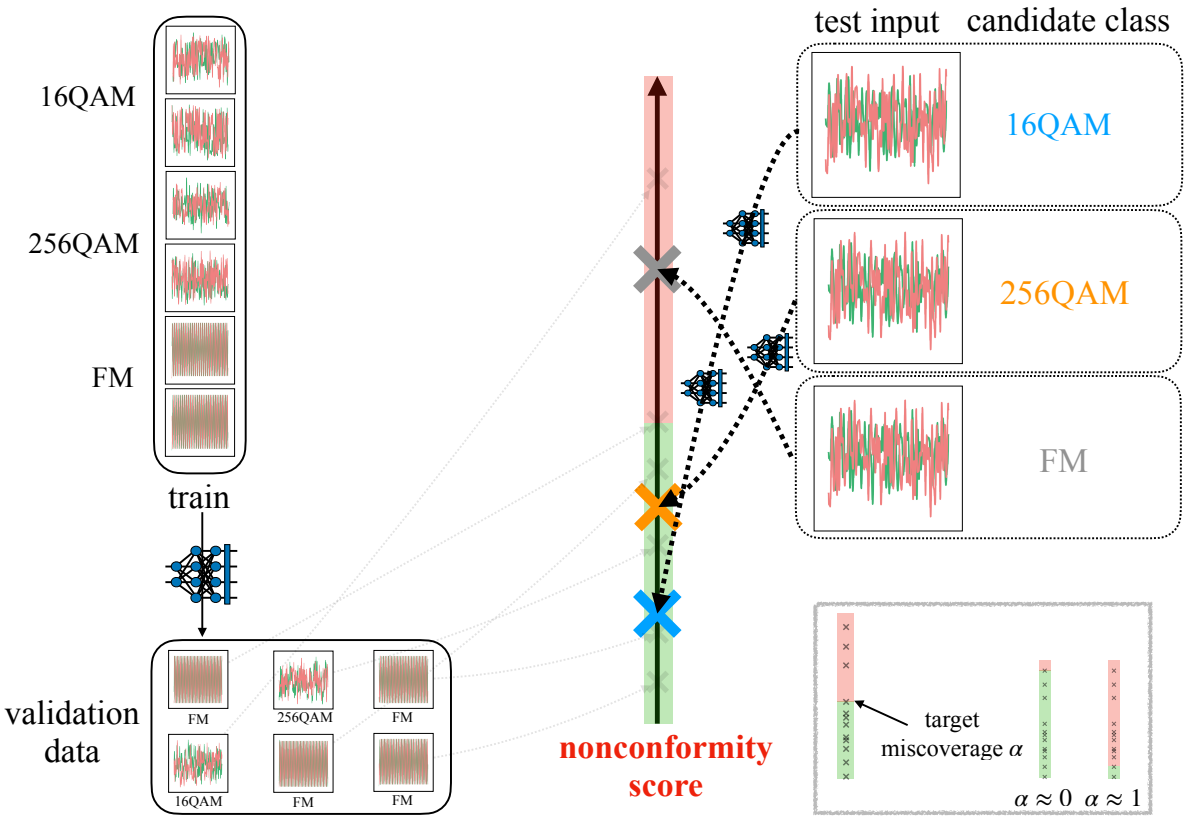
Offline Conformal Prediction

- **4.** At test time, given a test input...



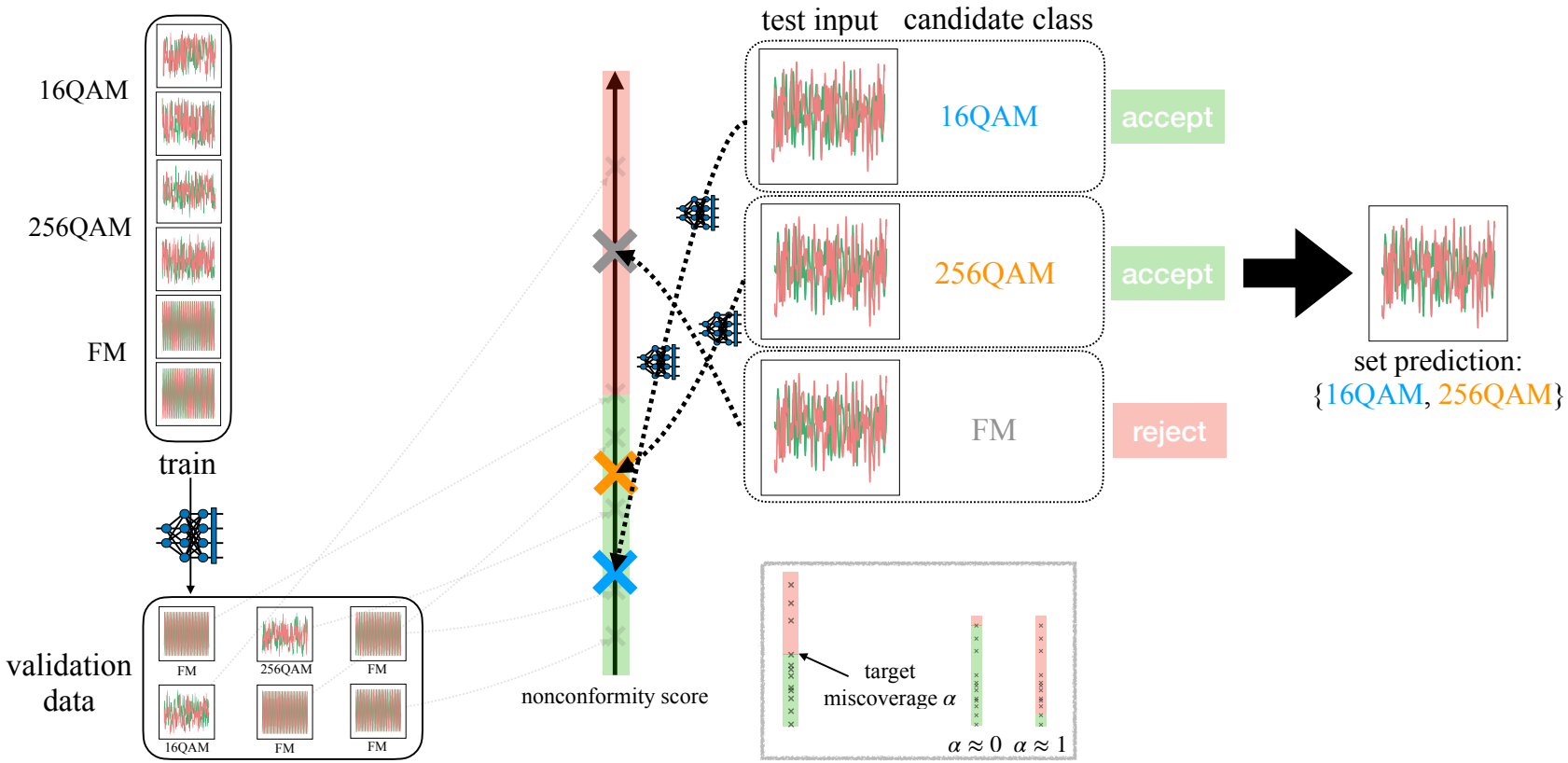
Offline Conformal Prediction

- ... compute the nonconformity score **for each candidate label** with the same trained probabilistic model...



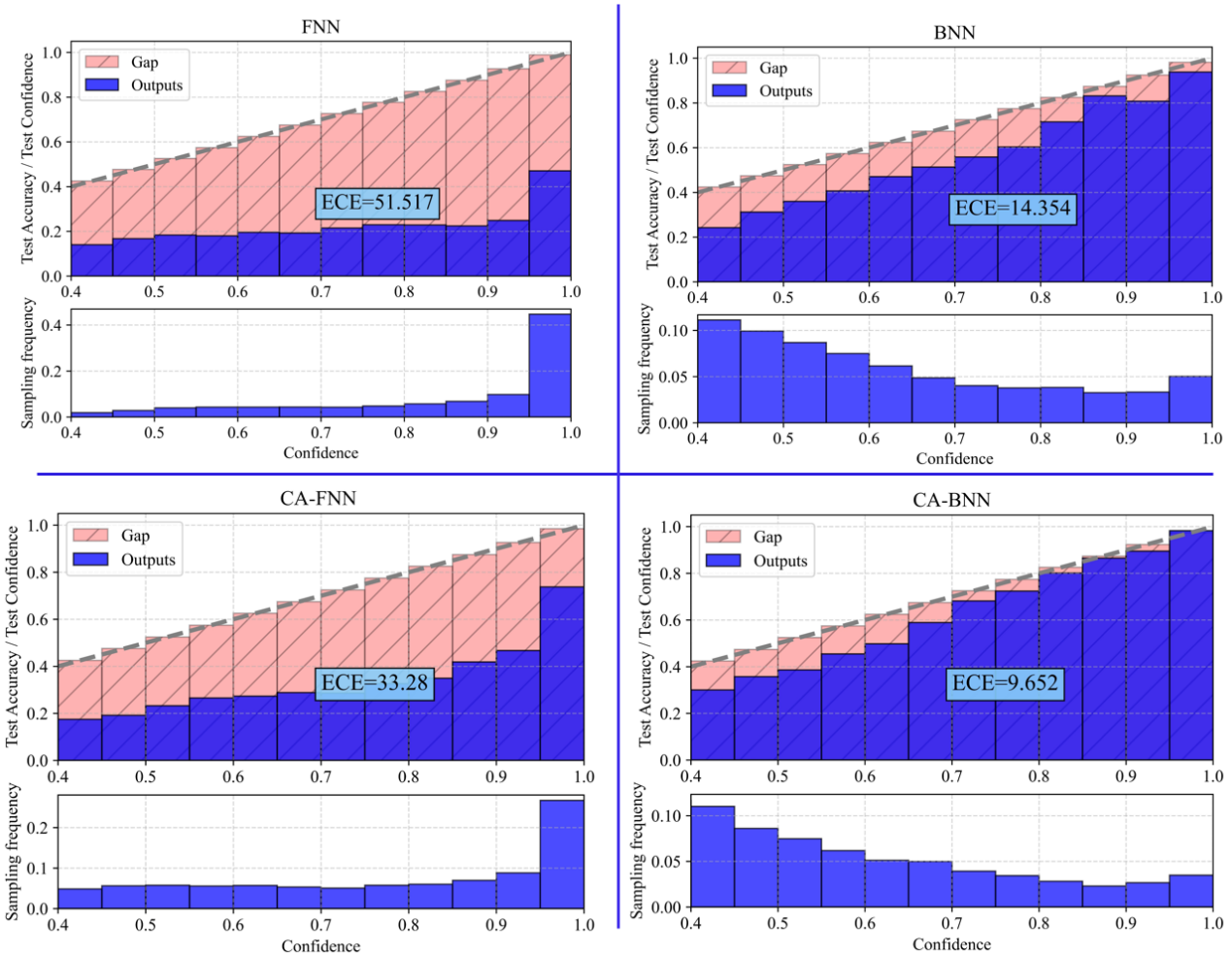
Offline Conformal Prediction

- ... and include in the prediction set all the candidate labels that lie in the acceptance region.



Calibration for OOD Data

- Calibration-aided training[41]



41 J. Huang, S. Park, and O. Simeone, "Calibration-Aware Bayesian Learning," in Proc. IEEE MLSP 2023