Towards Integrating Formal Methods into ML-Based Systems for Networking

ACM HotNets 2023

Fengchen Gong, Divya Raghunathan, Aarti Gupta, Maria Apostolaki

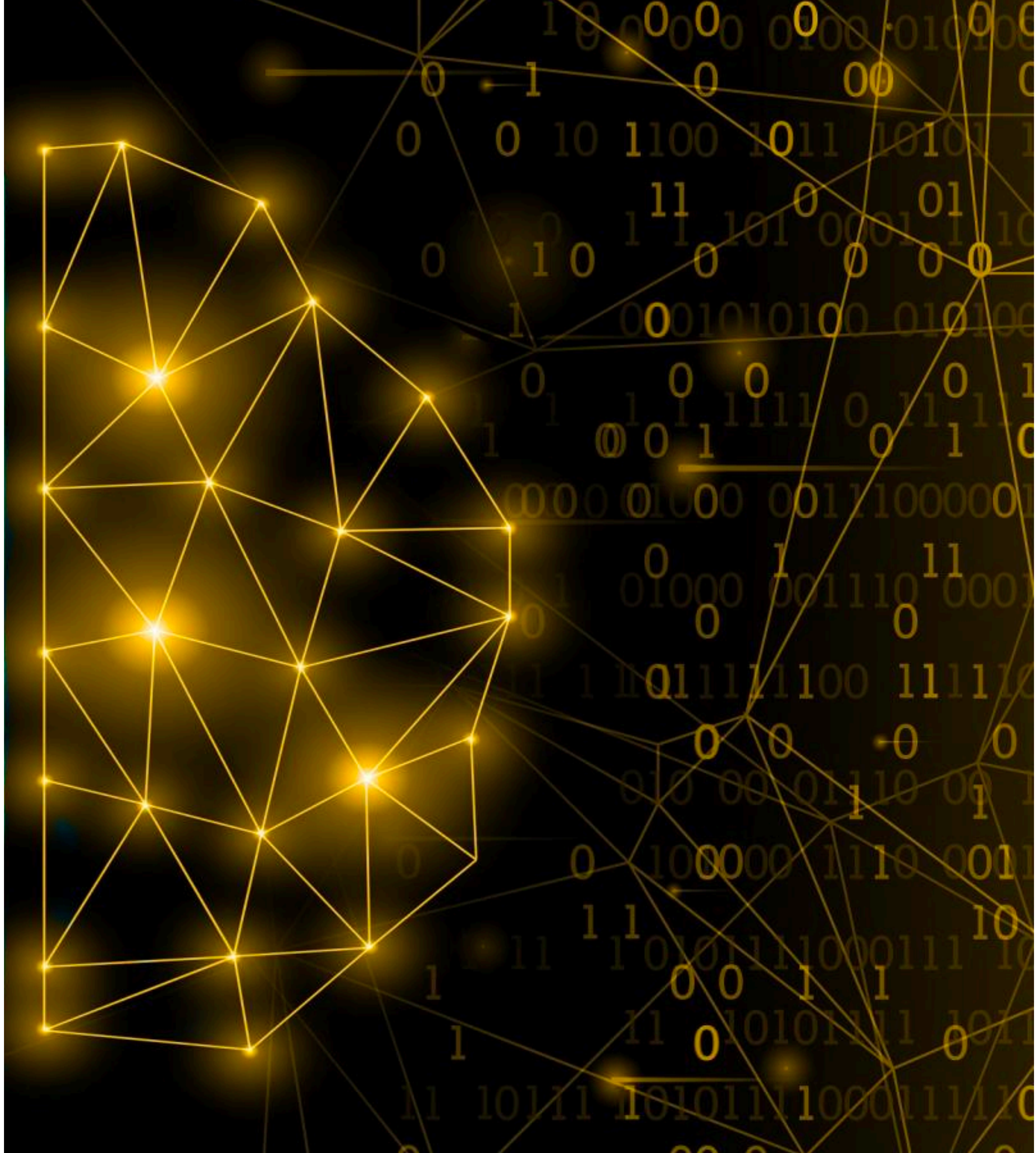
PRINCETON UNIVERSITY

Machine learning is scalable and adaptable

But…

How generalizable is a machine learning model?

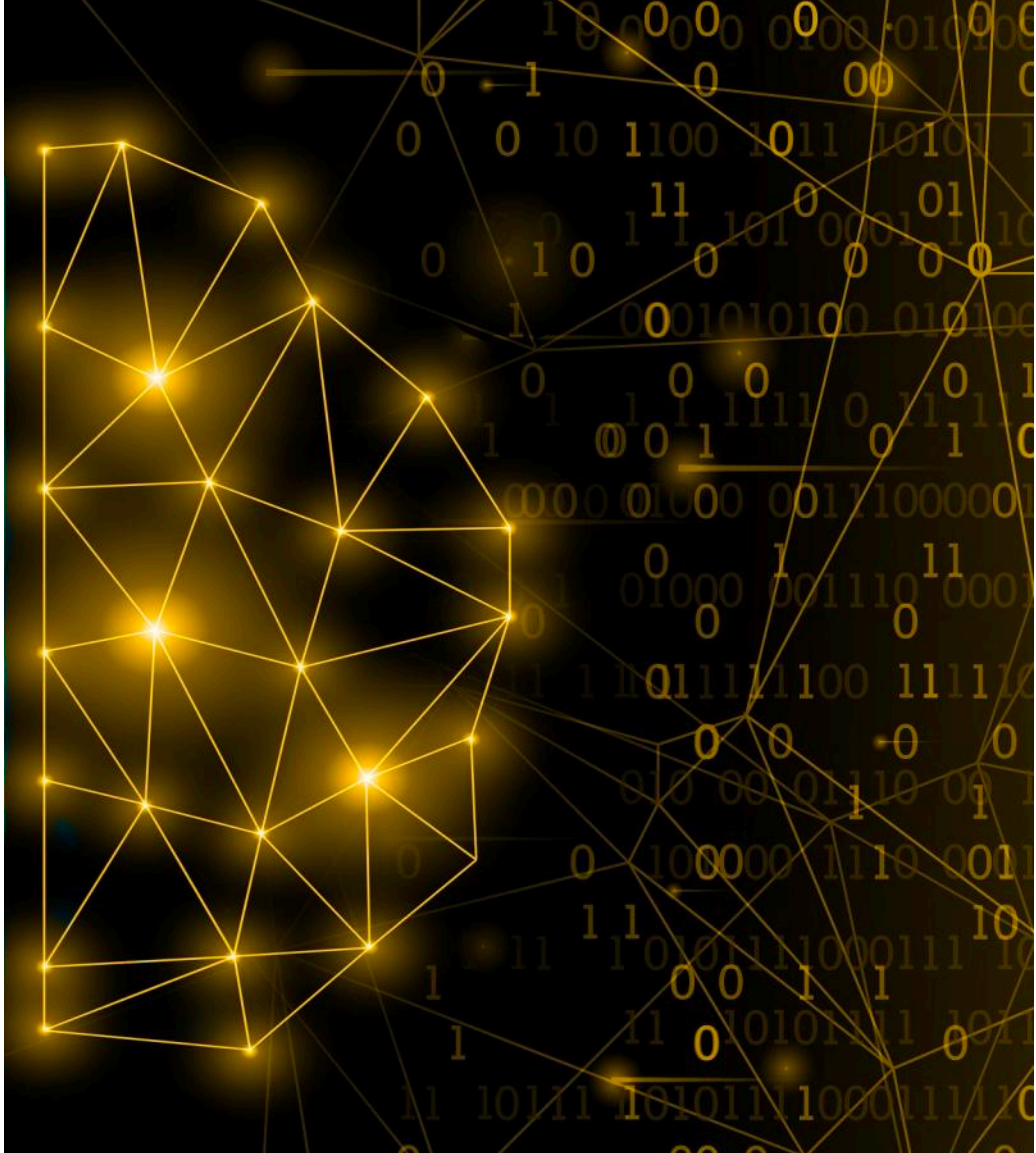
How do we make sure a model is trustworthy when used in real world?

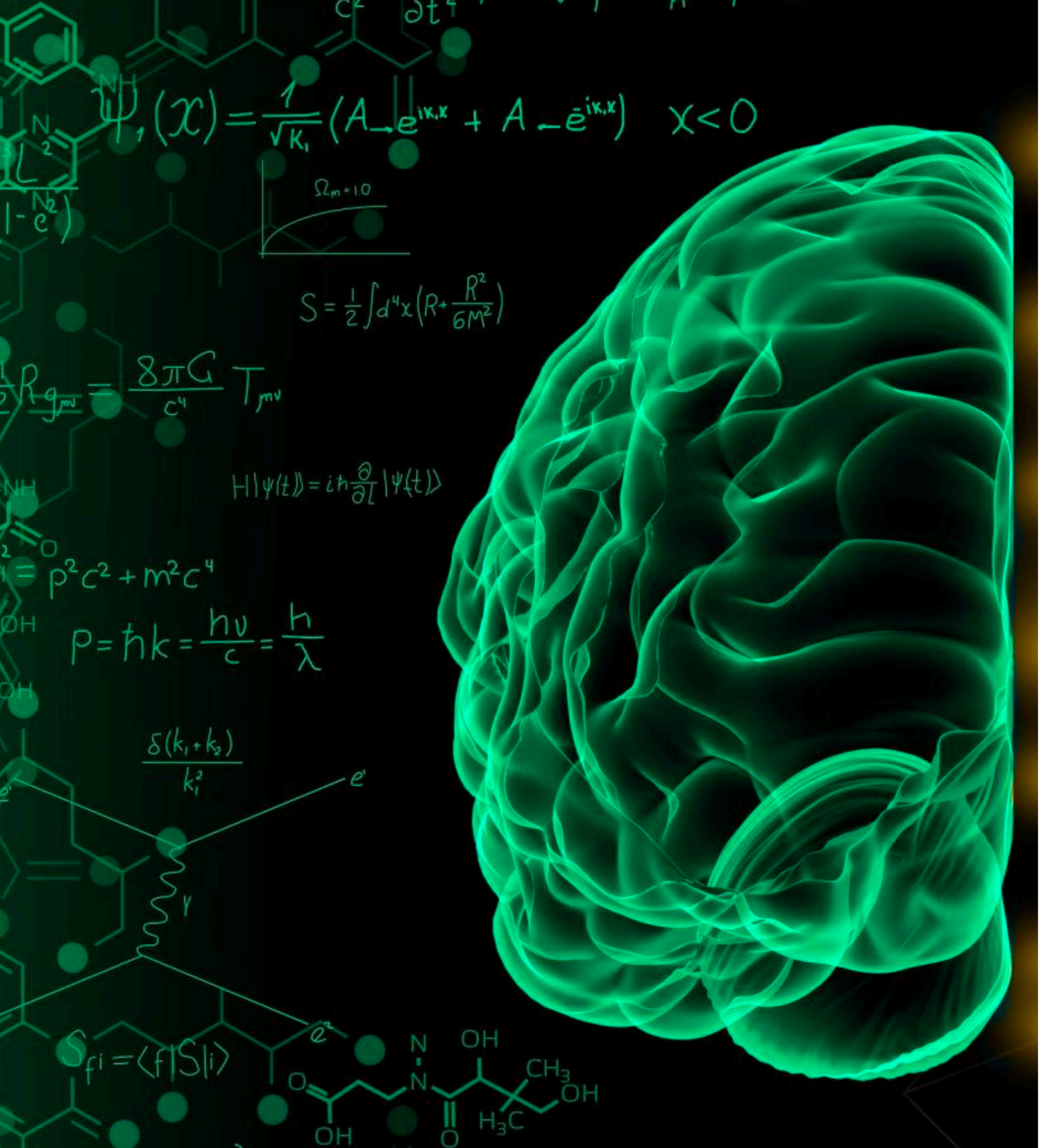


ML-based solutions lack correctness guarantees

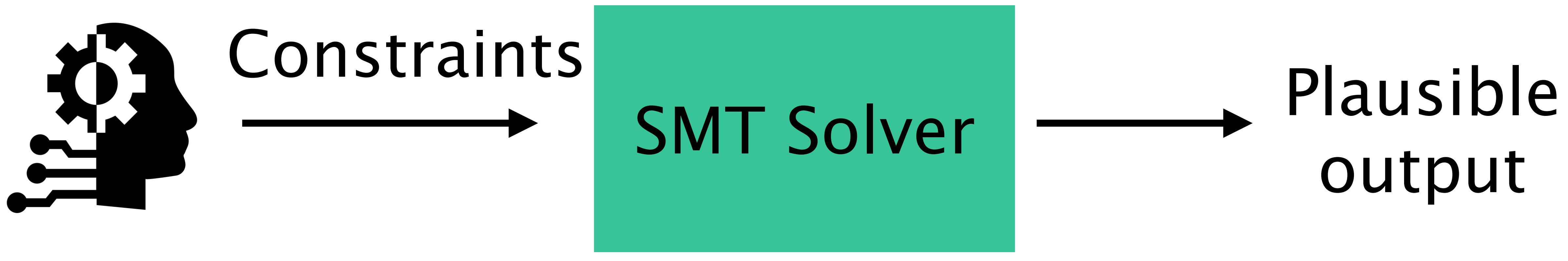
ML-based solutions may…

- result in implausible outputs
- perform worse than simple heuristics
- contradict common sense





Formal Methods leverages knowledge to generate correct results



…but, it doesn't scale

Image by Donald Jorgensen | Pacific Northwest National Laboratory

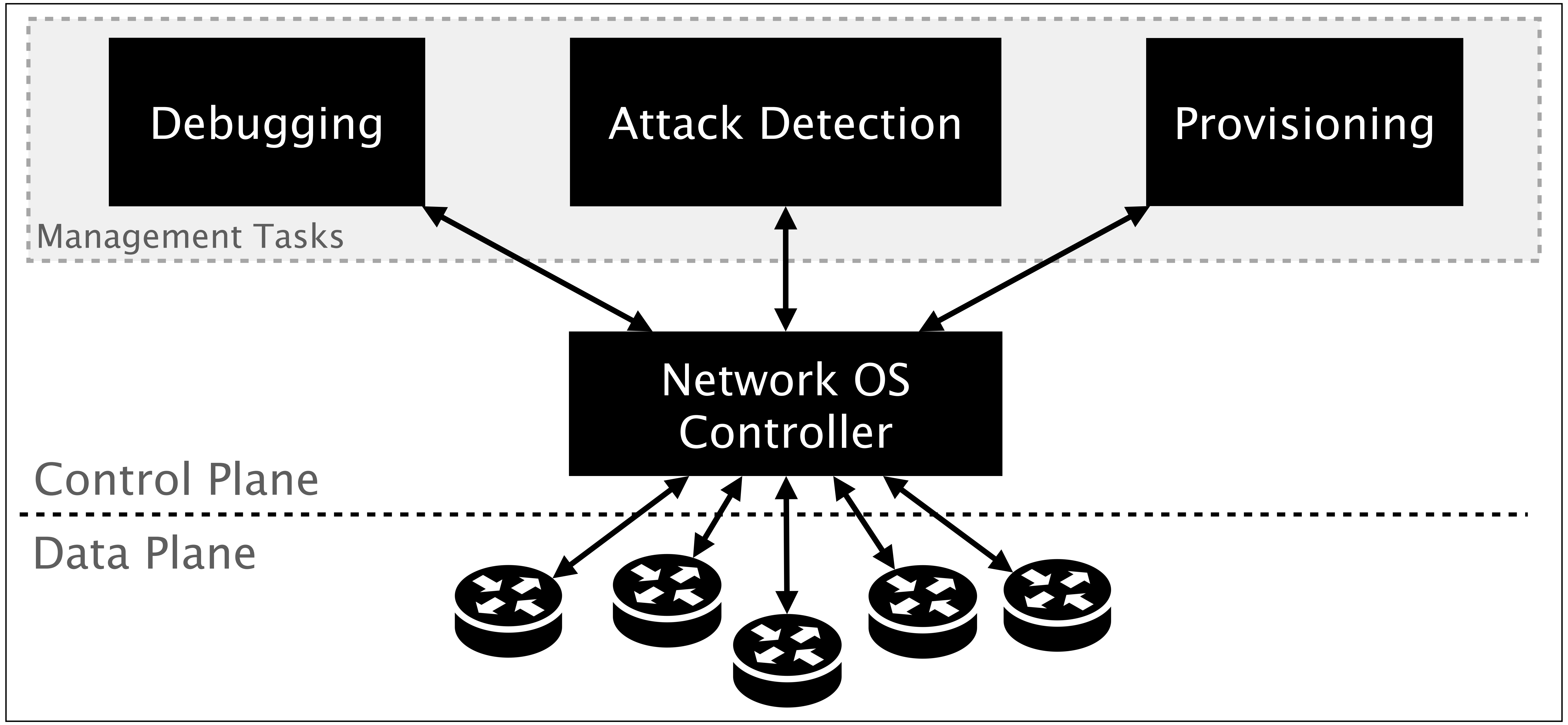
Towards Integrating Formal Methods into ML-Based Systems for Networking

Formal Methods
closed-forms

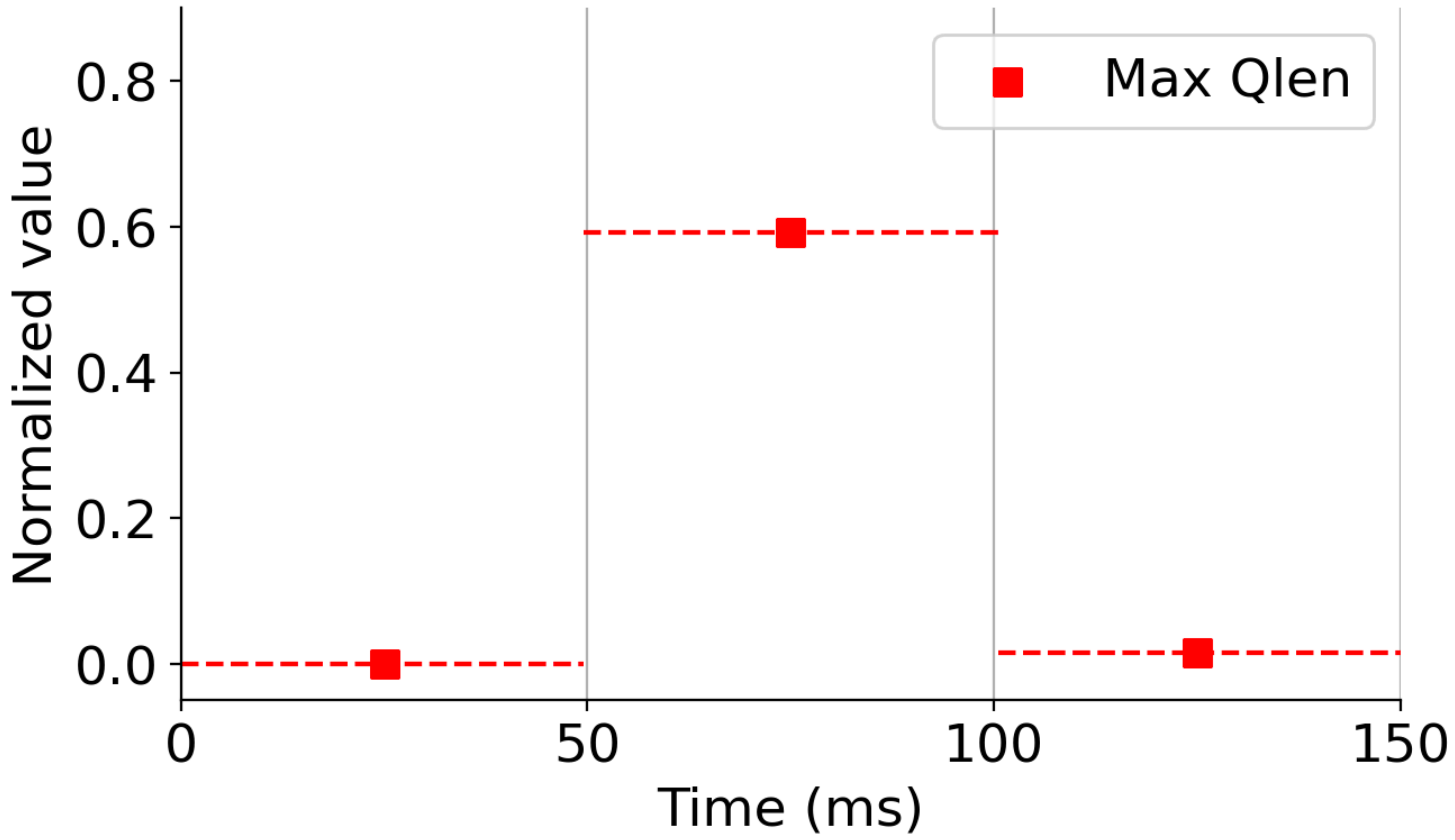
Machine Learning
correlations

Solutions with both correctness and scalability

Fine-grained telemetry is required for network managements, but hard to get

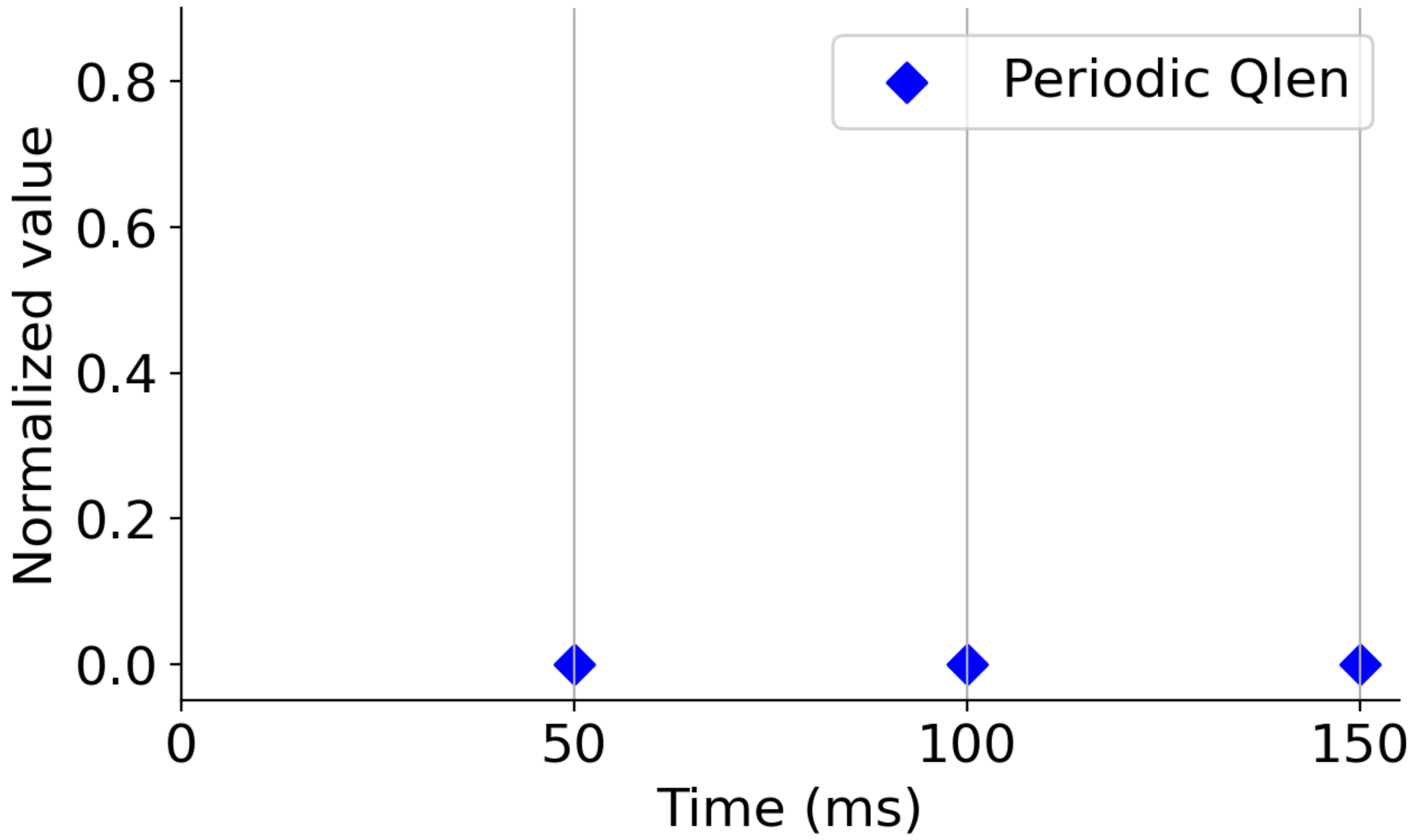


Coarse-grained telemetry in switches



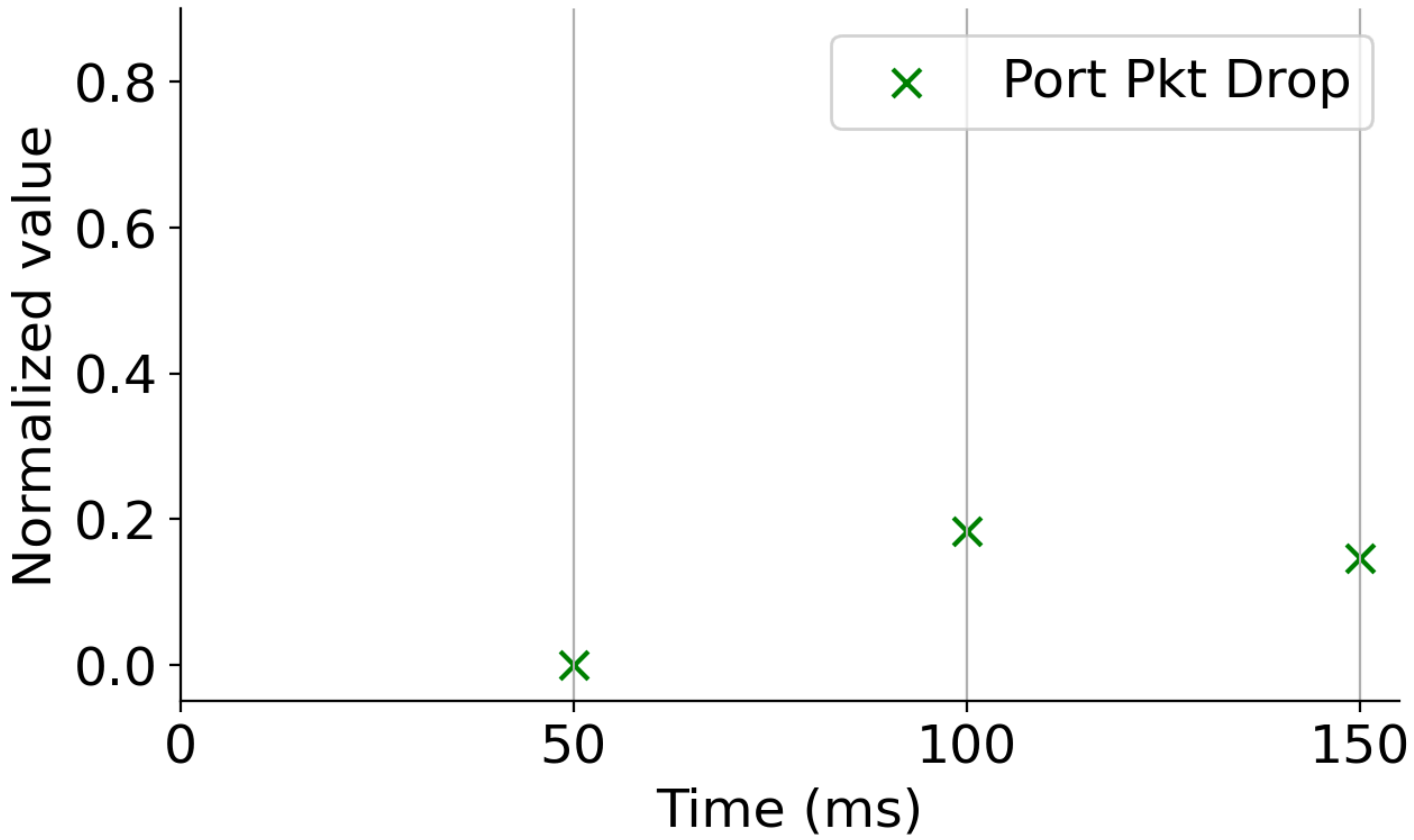
Maximum queue length
with 50ms granularity

Coarse-grained telemetry in switches



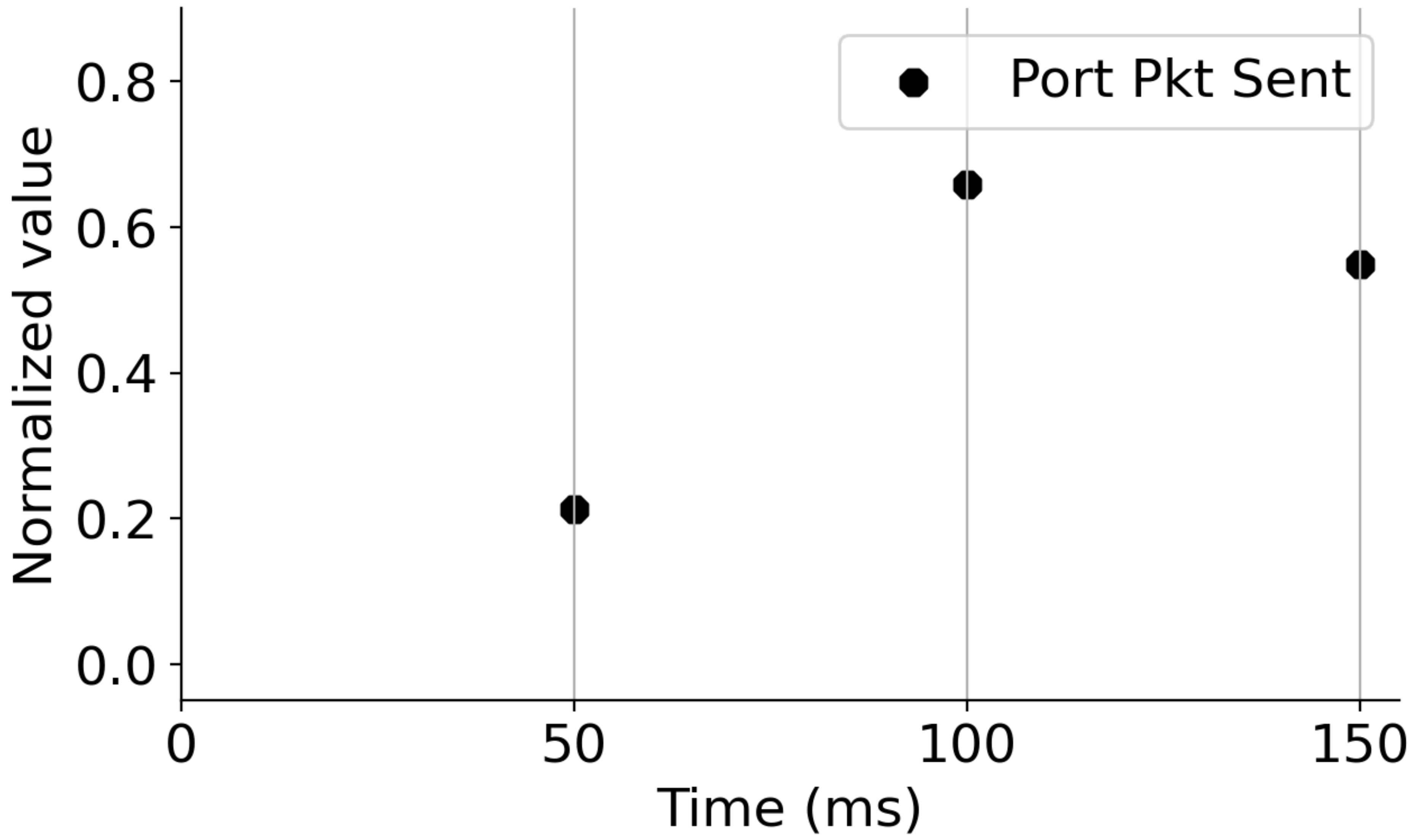
Periodic queue length
with 50ms granularity

Coarse-grained telemetry in switches



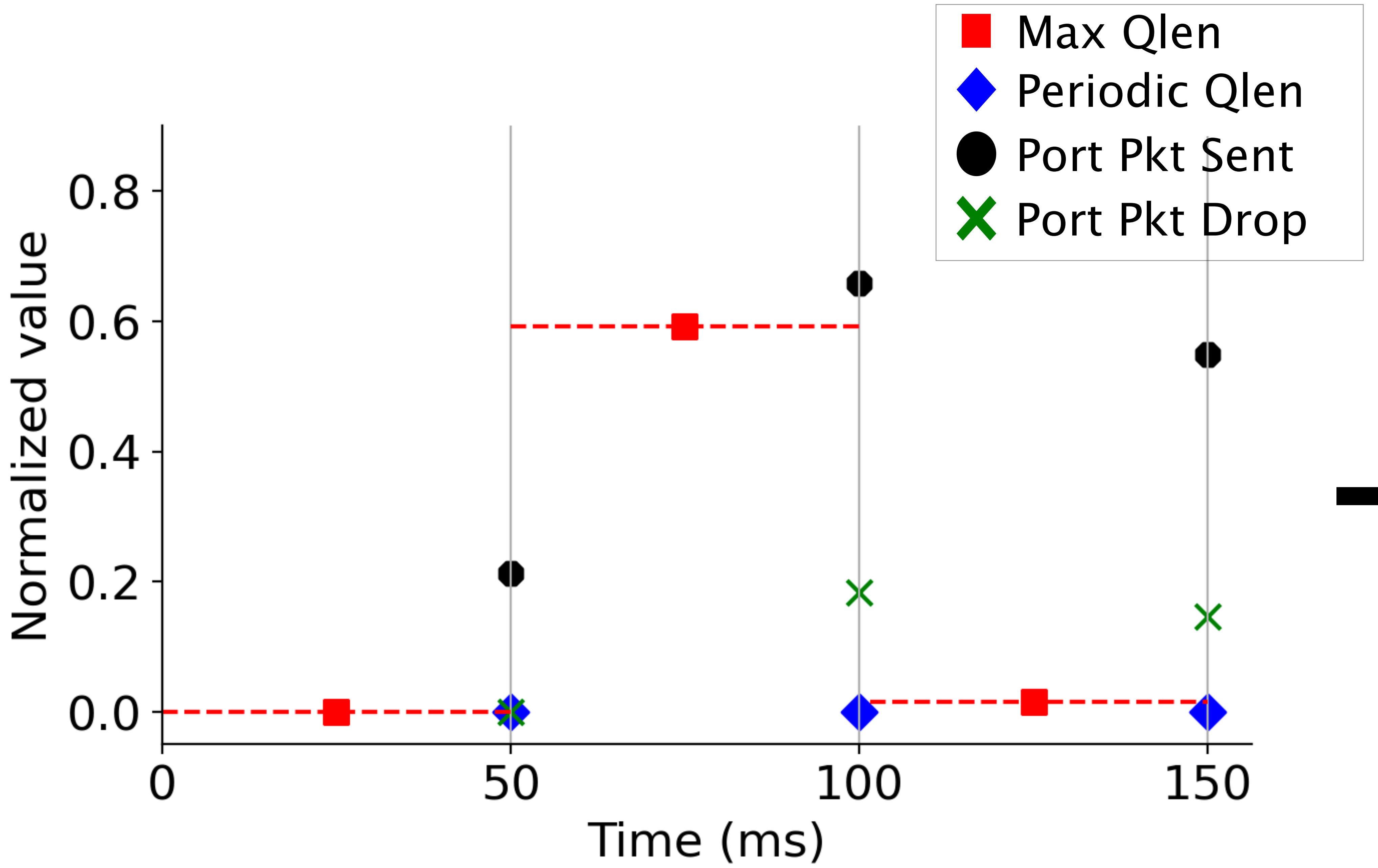
Per-port packet drop count
with 50ms granularity

Coarse-grained telemetry in switches

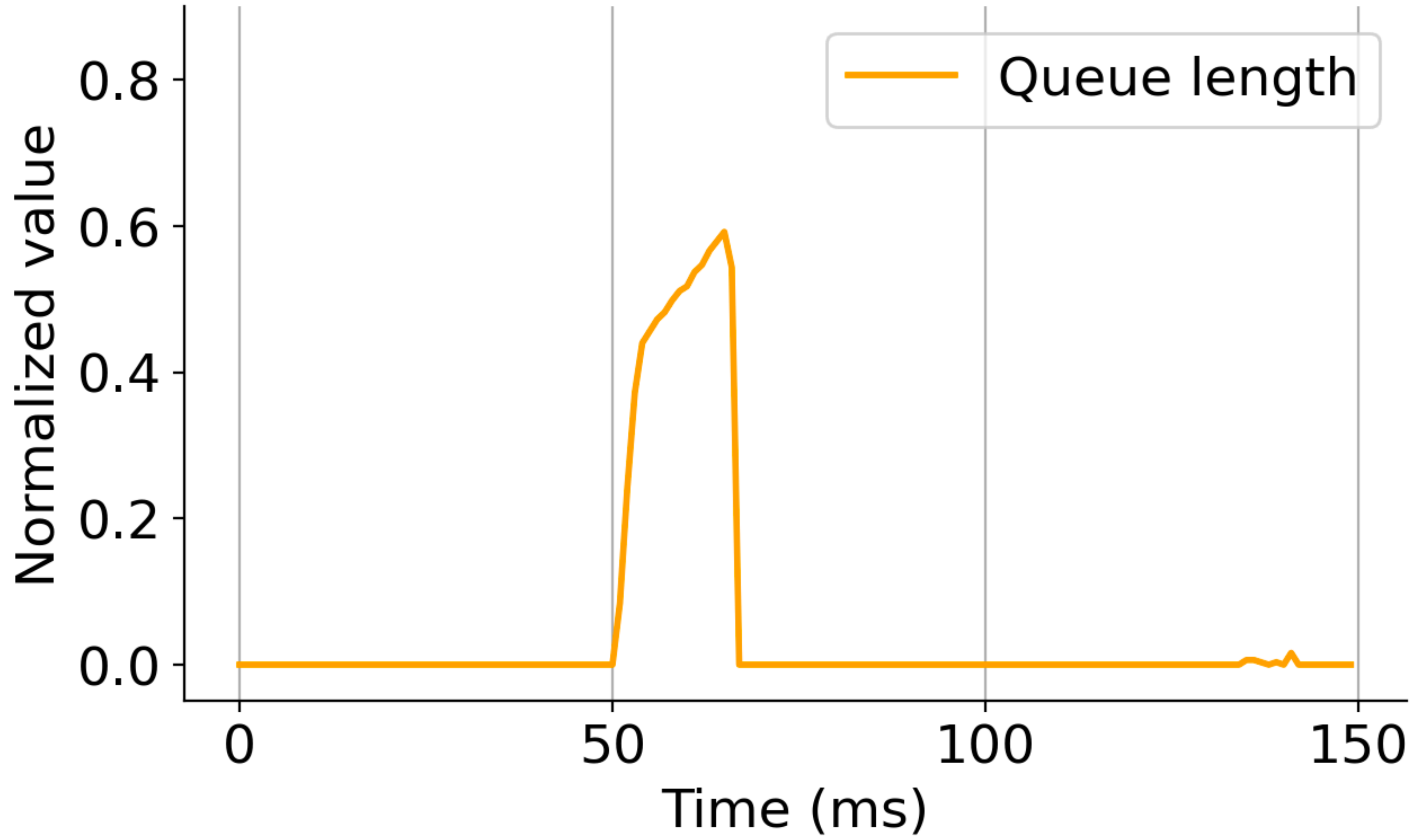


Per-port packet sent count
with 50ms granularity

Network telemetry imputation

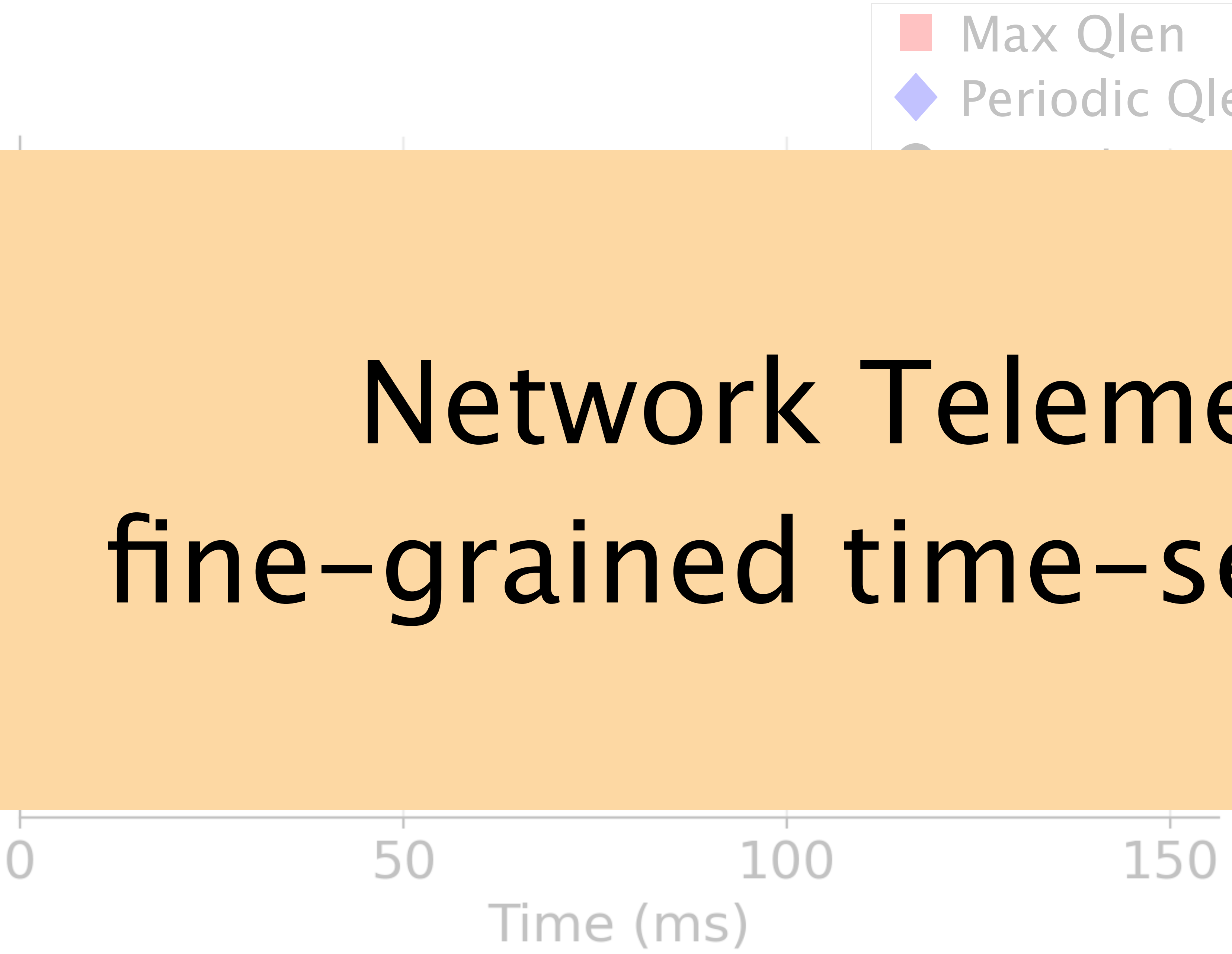


Coarse-grained measurements
with 50ms granularity



Fine-grained queue length
with 1ms granularity

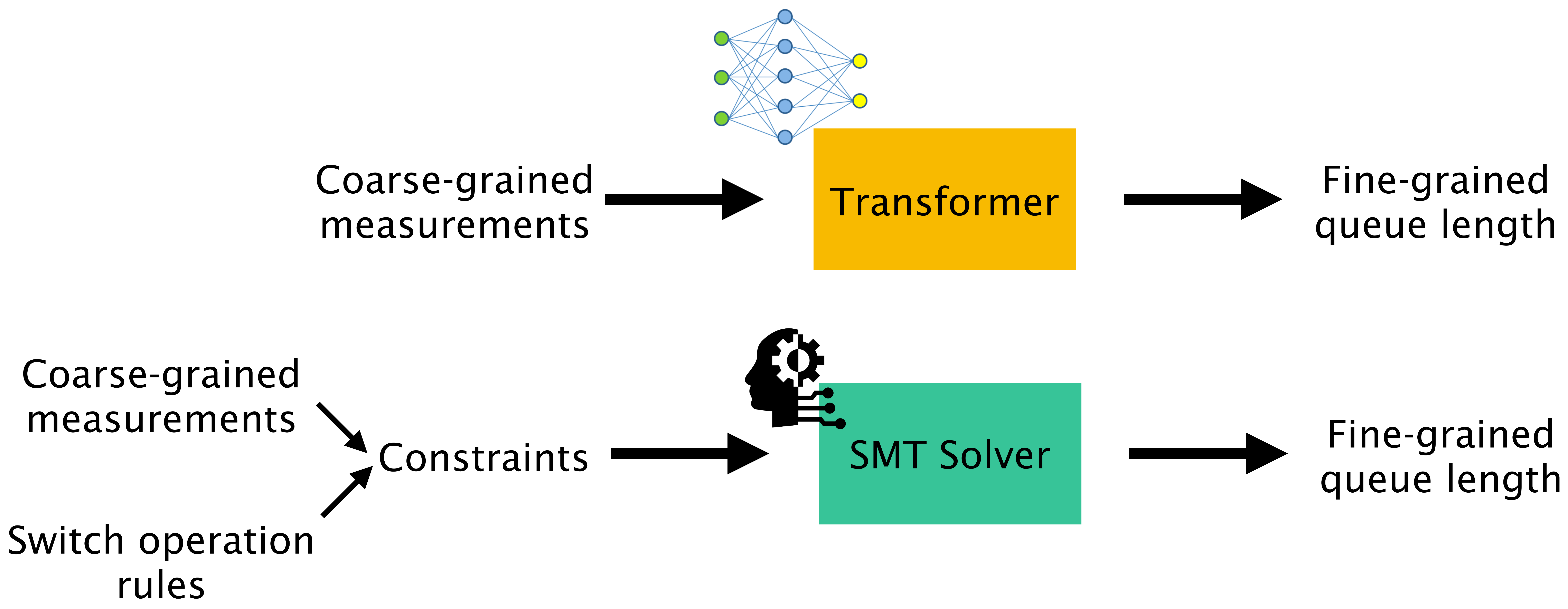
Network Telemetry Imputation: recovering fine-grained time-series from coarse-grained ones



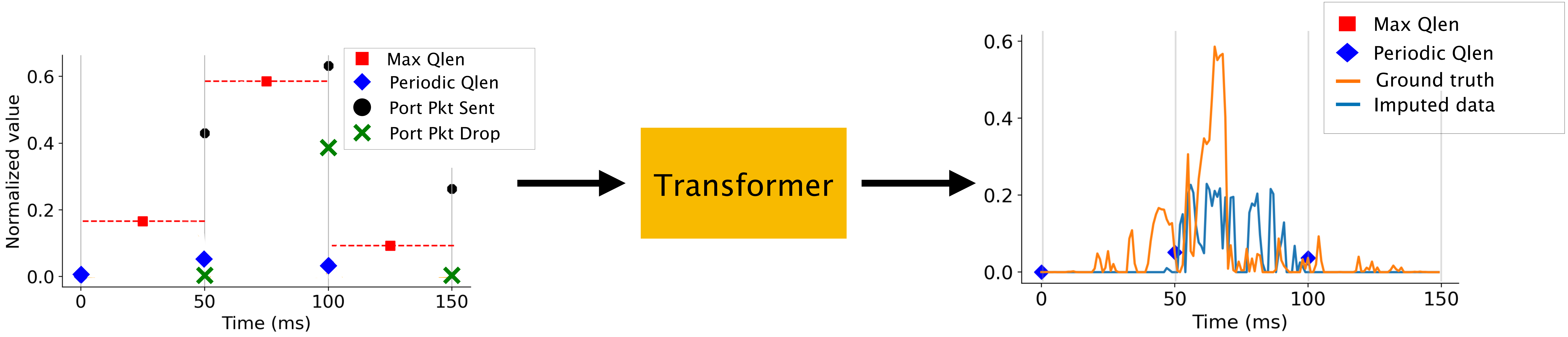
Coarse-grained measurements
with 50ms granularity

Fine-grained queue length
with 1ms granularity

Potential solutions: ML or Formal Methods



A transformer can learn correlations, but the output lacks correctness

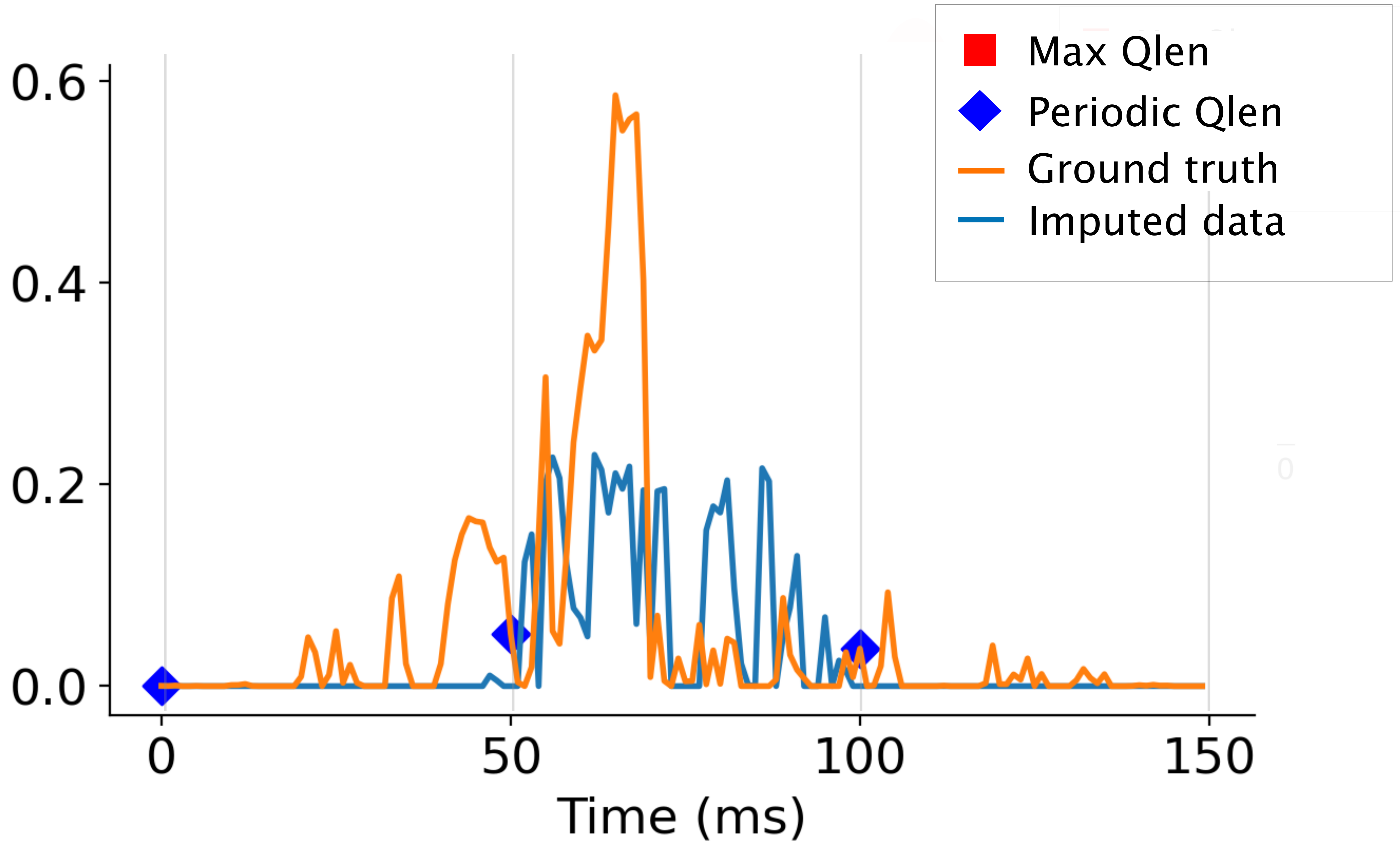


Coarse-grained measurements

Imputed fine-grained queue length

A transformer can learn correlations, but the output lacks correctness

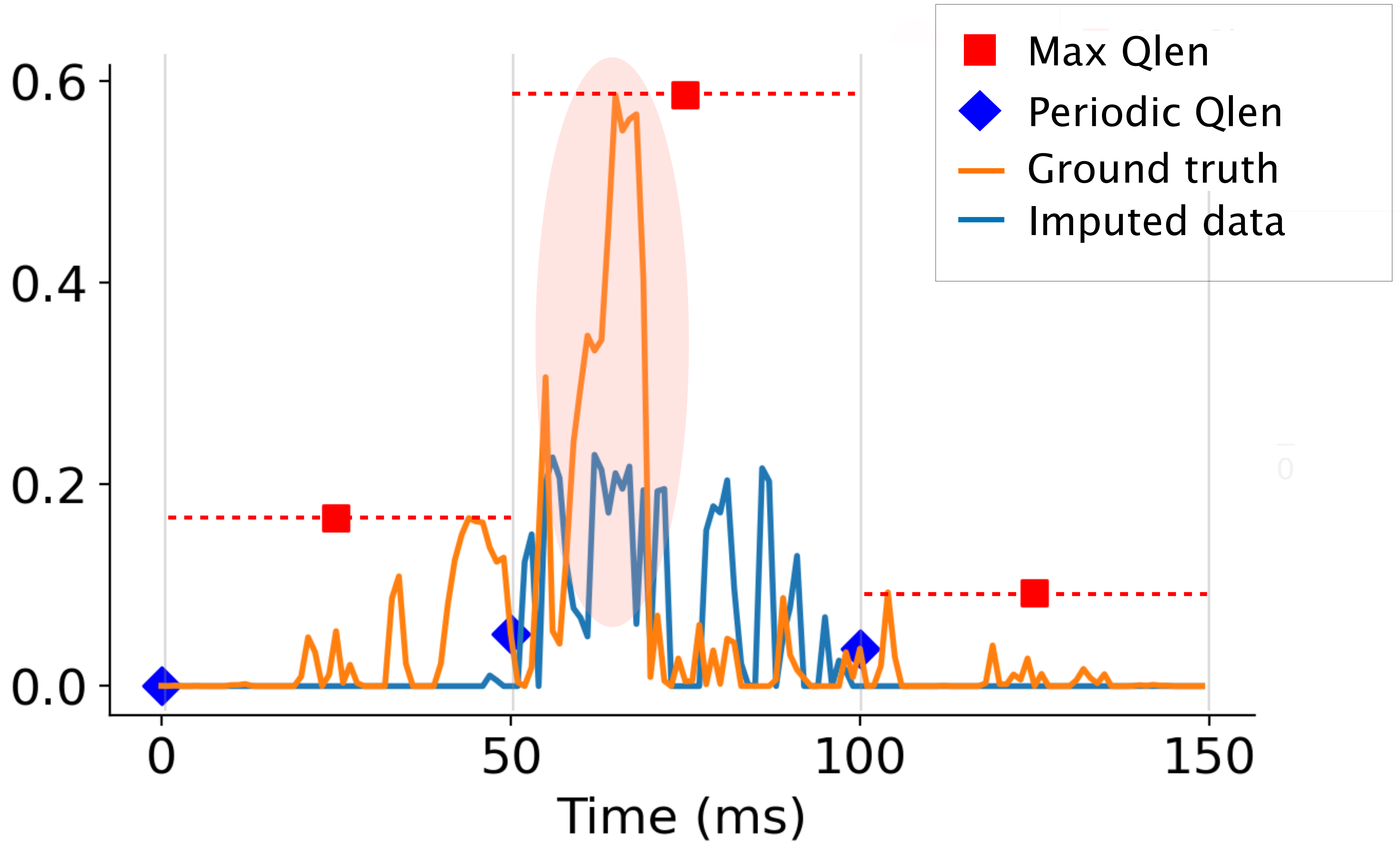
Coarse-grained measurements



Imputed fine-grained queue length

A transformer can learn correlations, but the output lacks correctness

Coarse-grained measurements



Imputed fine-grained queue length

Formal Methods can find a plausible solution, but is hard to scale

Operation constraints:

- Scheduling algorithm
- Buffer management algorithm

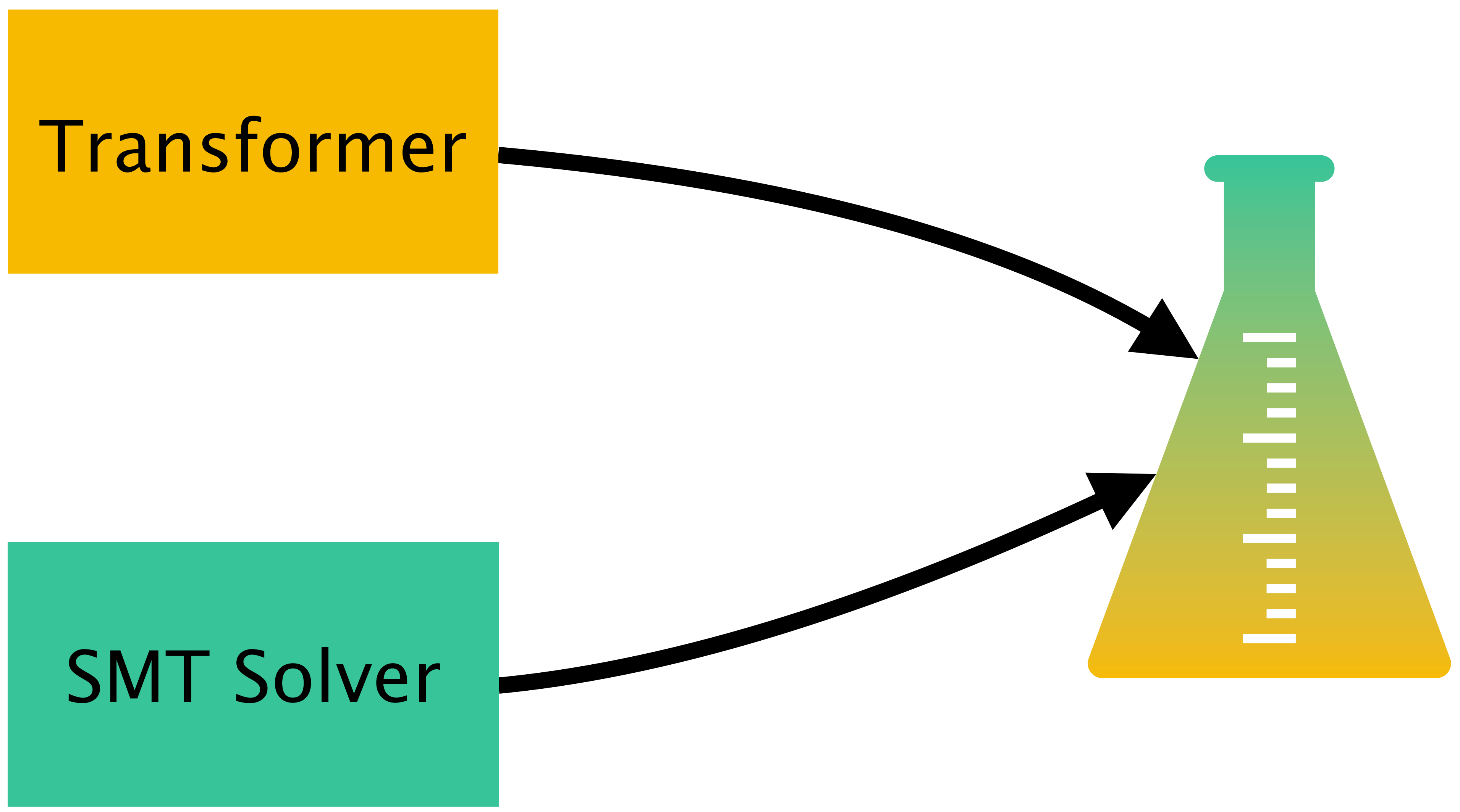
Measurement constraints:

- The maximum queue length
- The periodic queue length

→ SMT Solver

Couldn’t finish in 24hrs
for 1Gbps bandwidth

How to integrate them?



Challenges:

No standard way

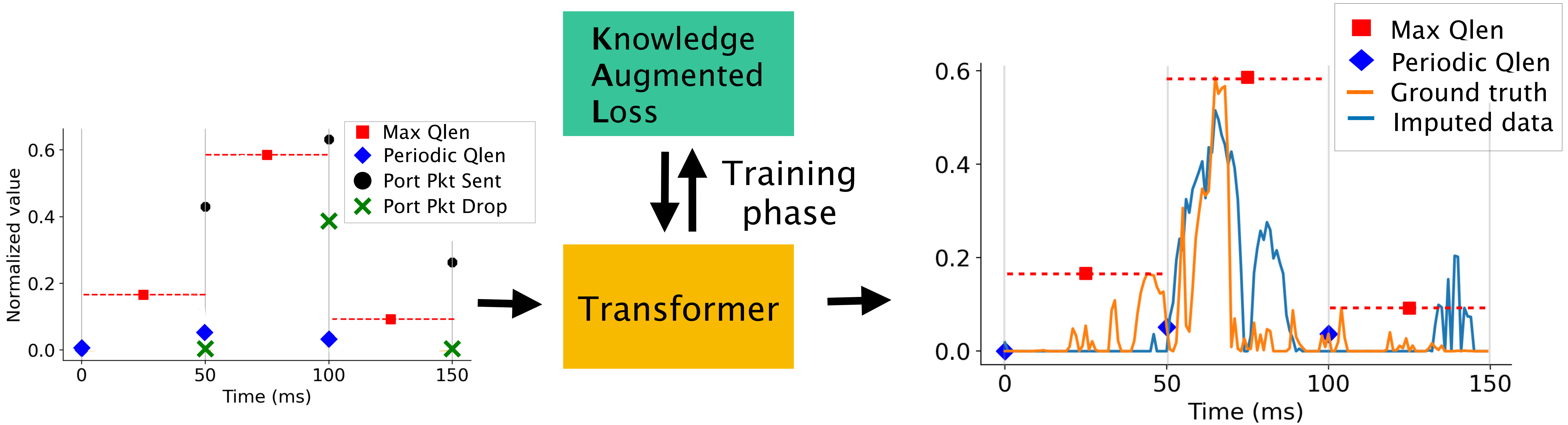
ML models cannot easily ingest traditional rules or relationships

Incorporating knowledge can increase the complexity of the learning process

Start from transformer

Transformer

Knowledge augmented loss



Coarse-grained Time Series

Imputed fine-grained queue length

Constraint enforcement module

Coarse-grained
Time Series

Imputed fine-grained
queue length

What next?

- Integrating Formal Methods and ML
  - Is there a better way of integrating them?
  - What other network problems can benefit?

- Generalize Network Telemetry Imputation
  - What other telemetry metrics can be imputed?
  - How do we impute real-time?

netsyn.princeton.edu