

The quest for good QML models

Elies Gil-Fuster, Freie Universität Berlin
QCOW, University of Warwick
24.04.2026

Why: broad field

- **Want:** Quantum advantage in learning for *Parametrized Quantum Circuits* (PQCs) with real-world classical data, without hardware.
- **Have:** “ML theory” for PQC-based models 
 - Expressivity
 - Trainability
 - Generalization
- **Compulsory pedantic take:**
 - *Hey Elies, what's the difference between quantum learning theory and quantum machine learning theory?*
 - *Well, surely, it must be the machine!*
 - Oversimplified: (prescriptive, not descriptive, just for this talk)
 - QLT “is about” sample complexity (of the problem)
 - QML “is about” expressivity, trainability, generalization, etc. (of the model)

Why: context

Obstacle 1:

- it's difficult to prove things about messy (Q)ML models (neural networks)
- it's difficult to prove things about messy real-world learning tasks
- Difference QML vs classical ML: we don't have hardware

Obstacle 2:

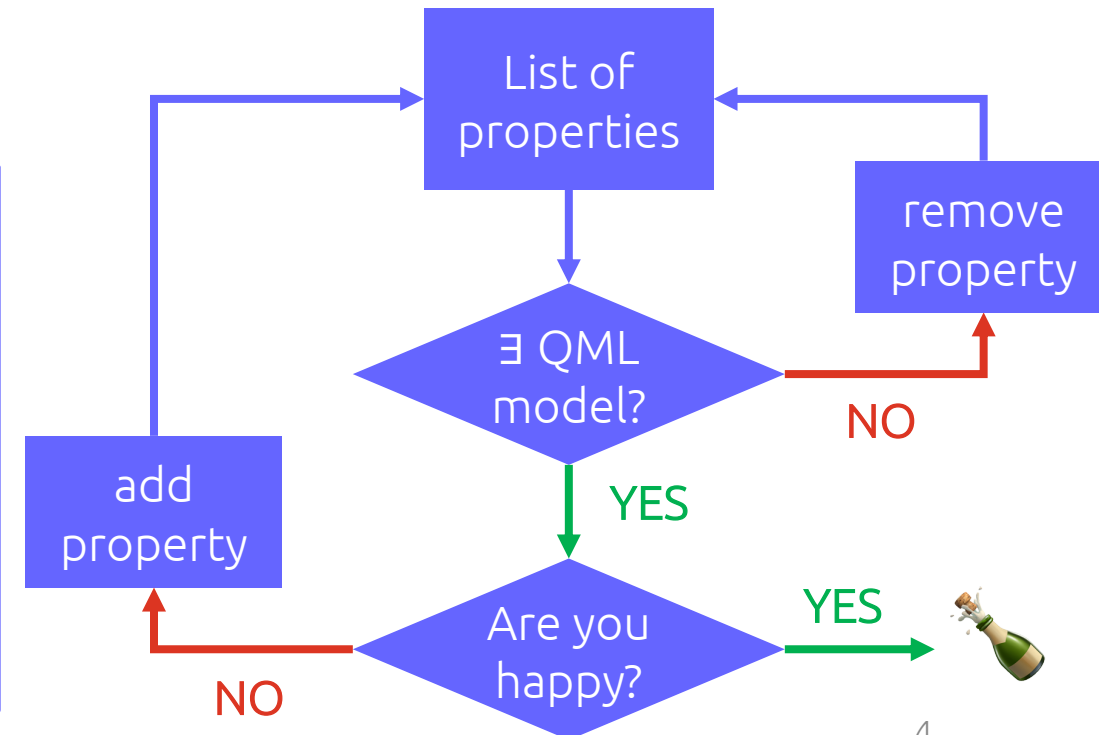
- most proposed heuristics seem to be fundamentally flawed
- folklore / conspiracy: *"no quantum advantage with trainable QML models"*.

Why: motivation

- **Guiding question:** *when should we use quantum alternatives to neural networks, and what advantages might they offer?*

- **Wishlist-based approach:**

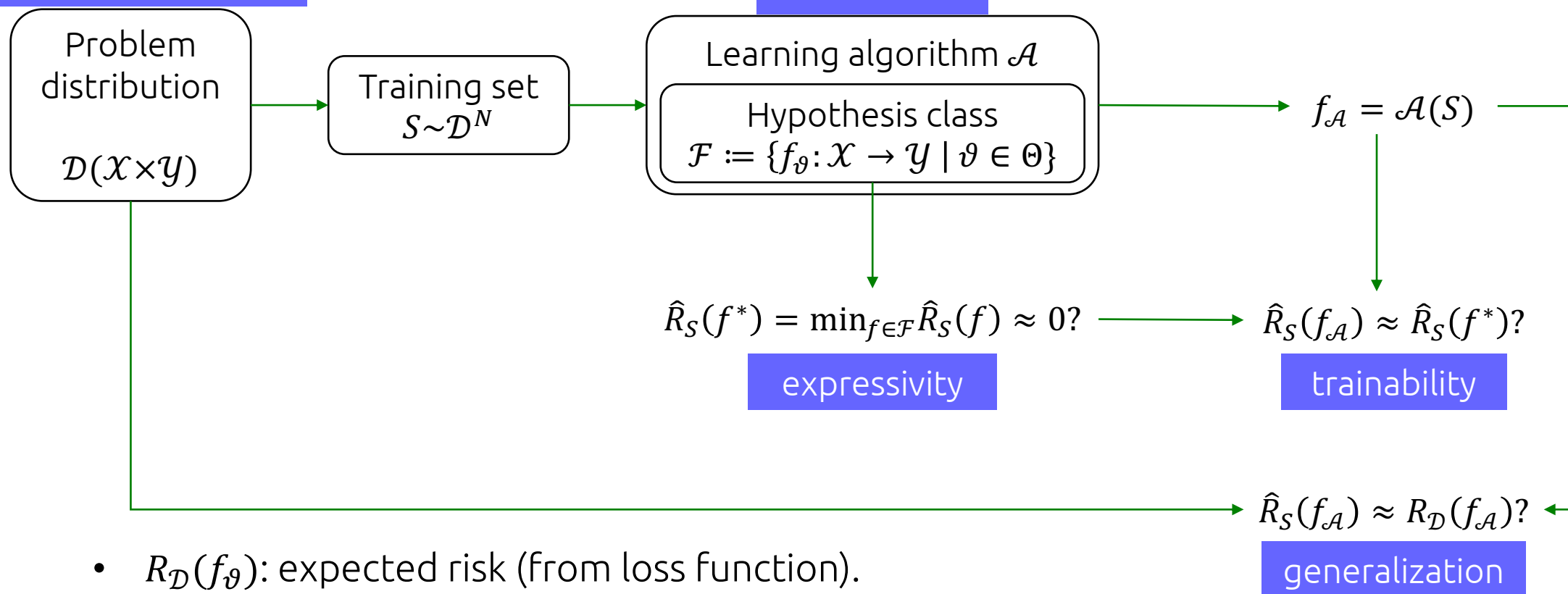
```
while is_happy = False:  
  1. List formal properties  $P = [P_1, \dots, P_L]$   
  2. Question:  $\exists$  QML model fulfilling all the properties?  
    a) YES: Question: are you happy now?  
      i. YES:  
        1.  $\text{is\_happy} = \text{True}$   
        2. Return "pop the champagne!"  
      ii. NO: Add a new property  $P_{L+1}$  to  $P$   
    b) NO: Think of a property  $P_\ell$  and remove it from  $P$ 
```



What: Supervised learning

“The task” (with $R_{\mathcal{D}}$)

“The model”

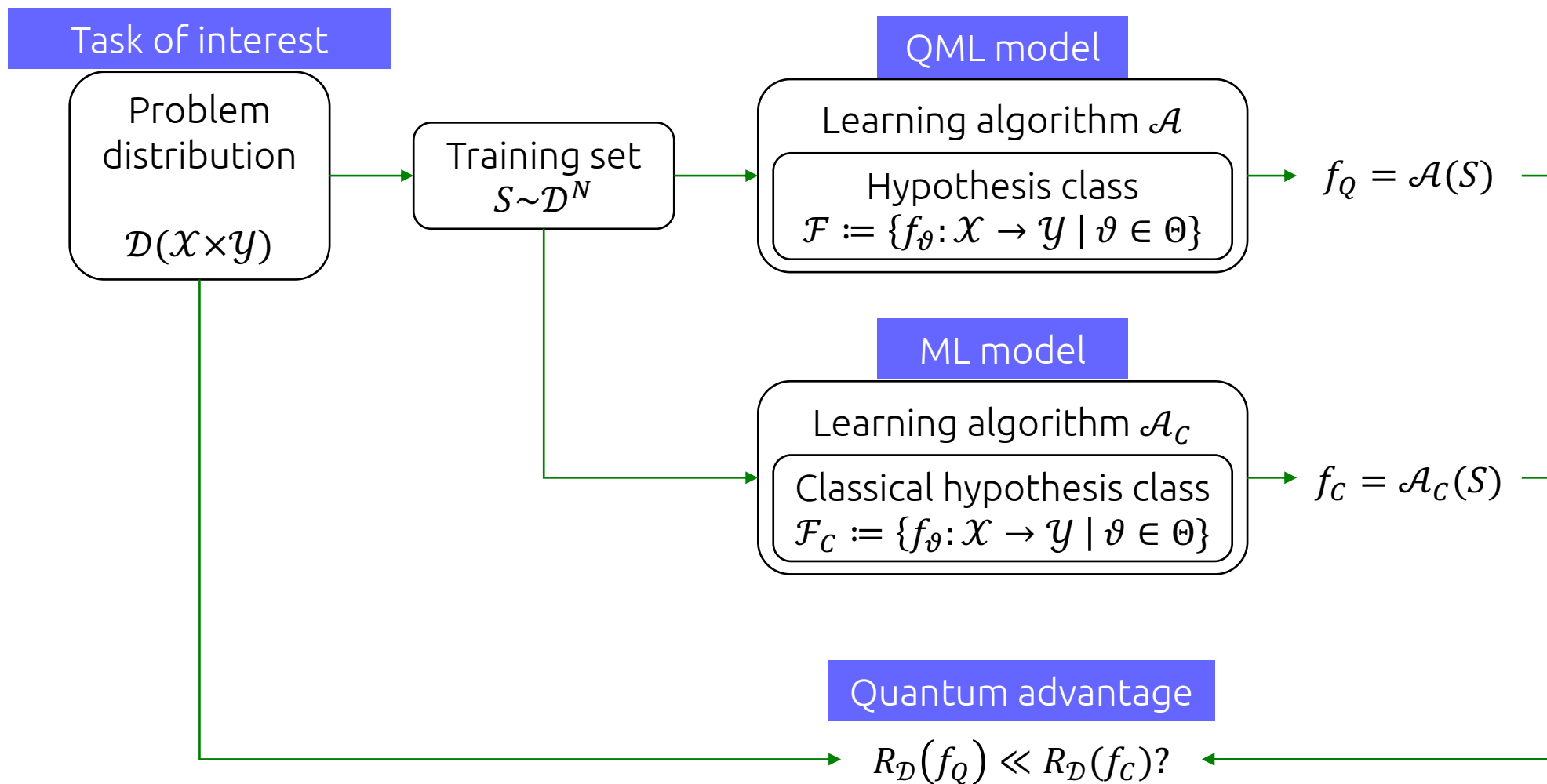


- $R_{\mathcal{D}}(f_{\vartheta})$: expected risk (from loss function).
- $\hat{R}_S(f_{\vartheta})$: empirical risk with respect to S .
- f^* : function in the class with the lowest empirical risk.

What: Examples (jargon alert)

- Good expressivity?
 1. Try to approximate all functions [Pérez-Salinas, Quantum 2020]
[Yu, NeurIPS 2024]
 2. Make sure you can evaluate classically-hard functions [Liu, Nat Phys 2021]
[Gyurik, 2306.16028]
- Good trainability?
 1. Set up the problem so that it's convex optimization (kernels) [Havlíček, Nature 2019]
[Schuld, PRL 2019]
 2. Set up the problem so that training reduces to computing classically-hard functions [Gyurik, 2306.16028]
[Jerbi, Nat Comms 2024]
- Good generalization?
 1. Do good-ol'VC theory (PAC learning, sample complexity) [Caro, QMI 2020]
[Caro, Quantum 2021]
[Caro, Nat Comms 2022]
[Caro, Nat Comms 2023]
 2. Do fun "modern" generalization [GE, Nat Comms 2024]
[Kempkes, PRX Q 2026]
[Rodríguez-Grasa, 2603.22964]
[More to come! 🤓]

What: Q. Adv. in supervised learning



Must specify notions of efficiency

What: Examples (jargon alert)

- Quantum advantage “from expressivity”:
 - Make all valid solutions be classically-hard functions

[Liu, Nat Phys 2021]
[Gyurik, 2306.16028]

- Quantum advantage “from trainability”:
 - Make training reduce to a classically-hard function

[Gyurik, 2306.16028]
[Jerbi, Nat Comms 2024]

- Quantum advantage “from generalization”:
 - A bit more difficult to say, but kind of all of them.

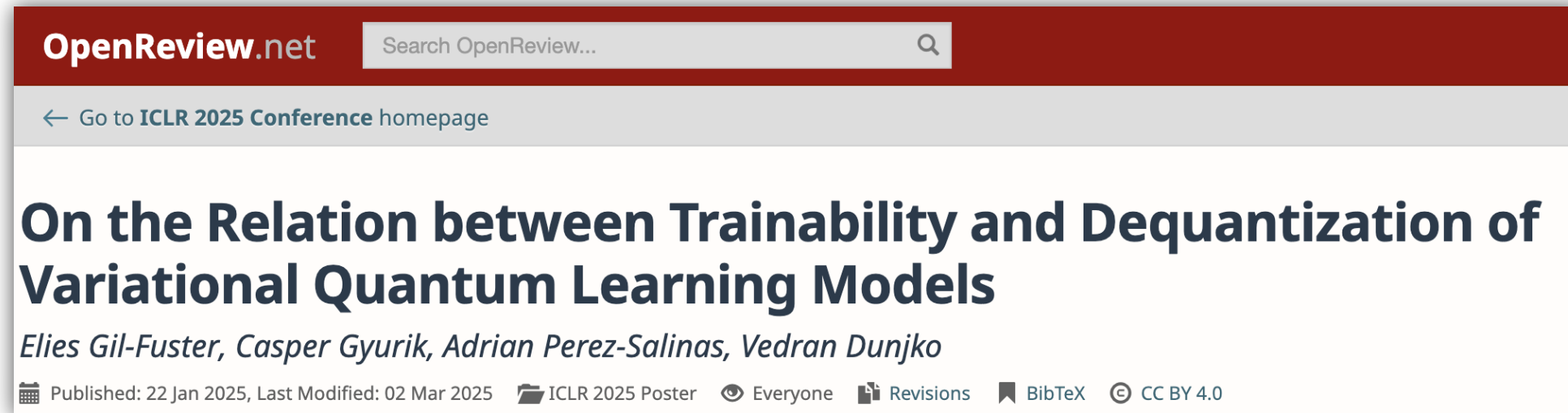
Questions from Matthias

- How do “quantum learning theory” tools like classical shadows or shadow tomography impact QML?
 1. Cheat your way into quantum advantage by having “quantum input” problems.
 2. Make quantum functions more implementation-friendly (classical memory).
 3. Argue that some problems will be quantum-hard.
- What kinds of quantum advantages can we hope for / expect in QML?
 1. Strictly speaking: hope for everything, expect nothing 😅
 2. At least: polynomial advantages for evaluation, super-poly for sampling tasks.
 3. Anything you want if you're happy with contrived examples
- What are the current big challenges in QML and why should you care?
 1. Provide evidence for practical advantage.
 2. Come up with good proxies to overcome lack of hardware and hardness of everything.

How: Wishlist-based approach in action

folklore / conspiracy: “no quantum advantage with trainable QML models”.

Wishlist-based approach: [GF et al. ICLR 2025].



The screenshot shows the OpenReview.net interface. At the top is a dark red header with the OpenReview.net logo and a search bar. Below the header is a light gray navigation bar with a link to the ICLR 2025 Conference homepage. The main content area features the title of the paper, the authors' names, and a row of metadata icons including a calendar, a folder, an eye, a document, a bookmark, and a Creative Commons license.

OpenReview.net Search OpenReview... 

[← Go to ICLR 2025 Conference homepage](#)

On the Relation between Trainability and Dequantization of Variational Quantum Learning Models

Elies Gil-Fuster, Casper Gyurik, Adrian Perez-Salinas, Vedran Dunjko

 Published: 22 Jan 2025, Last Modified: 02 Mar 2025  ICLR 2025 Poster  Everyone  Revisions  BibTeX  CC BY 4.0



Casper Gyurik



Adrián Pérez-Salinas



Vedran Dunjko

How: Wishlist-based approach in action

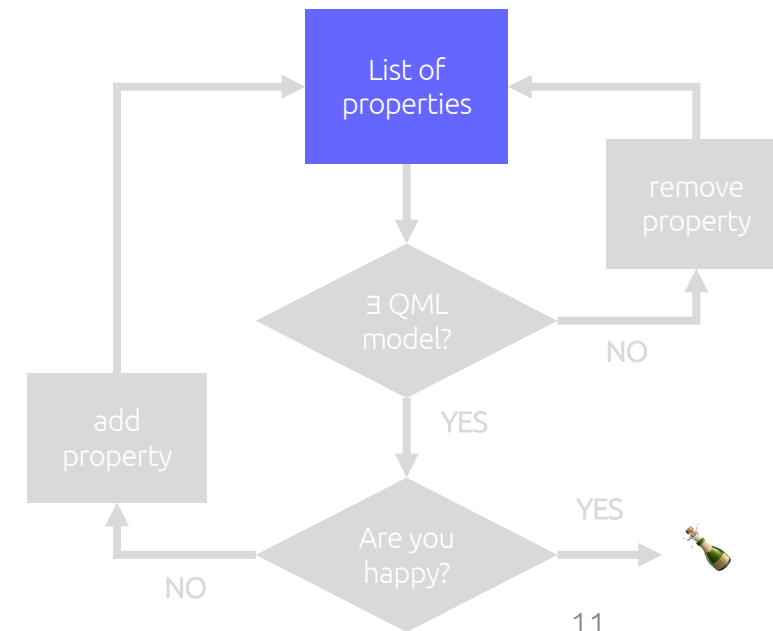
folklore / conspiracy: “*no quantum advantage with trainable QML models*”.

Wishlist-based approach: [GF et al. ICLR 2025].

Let's use our algorithm!

Properties:

1. Trainability
2. Non-dequantization
3. “Variationalness”
4. “Practical relevance”



How: Attempts at formalization

Trainability

A QML model $(\mathcal{F}, \mathcal{A})$ is trainable if

$$\hat{R}_S(f_{\mathcal{A}}) \approx \hat{R}_S(f^*)$$
 with high probability over the training set S .

Barren Plateaus (BPs) of a function family

A parametrized function family $\mathcal{F}$ has a BP if:

$$\text{Var}_{\theta \sim P}[f_{\theta}(x)] \approx 0$$
 for a distribution over parameters $P(\theta)$.

Dequantization of a QML model

A QML model $(\mathcal{F}, \mathcal{A})$ is dequantized if there exists a classical model $(\mathcal{F}', \mathcal{A}')$ such that

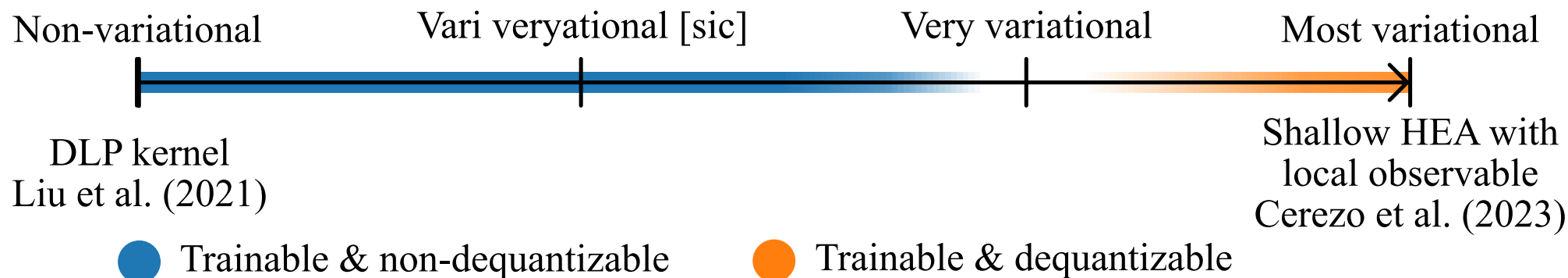
$$R_{\mathcal{D}}(f_{\mathcal{A}'}) < R_{\mathcal{D}}(f_{\mathcal{A}}).$$

Classical simulation of a function family

A parametrized function family $\mathcal{F}$ is classically simulable if the functions are classically efficiently evaluable.

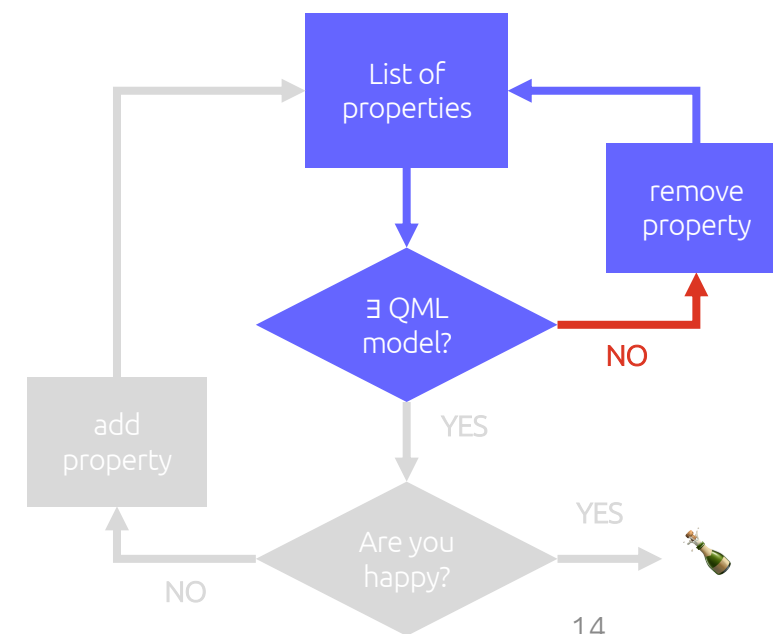
How: The essence of “variationalness”

- Vari veryational QML models:
 - Parameters appear “inside the circuit”, not classical post-processing.
 - Training algorithm must be gradient based.
 - There must be a template for repeated layers, size left free.
- This is a social definition



How: “Practical relevance”

- We tried and failed to formalize practical relevance 😞
- We cannot even tackle the question! 😐
- We prepared for this 😊
- Remove “practical relevance” from the list 🙄
- Properties:
 1. Trainable
 2. Non-dequantizable
 3. “Variational”

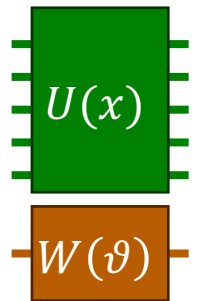


How: Main result

Theorem: there exist **vari veryational** QML models which are (gradient-based) trainable and non-dequantizable (give quantum advantage).

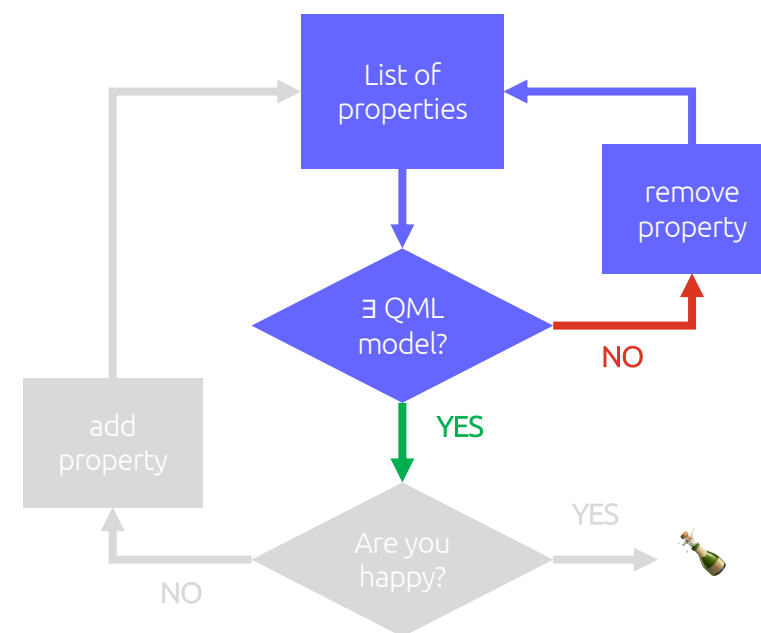
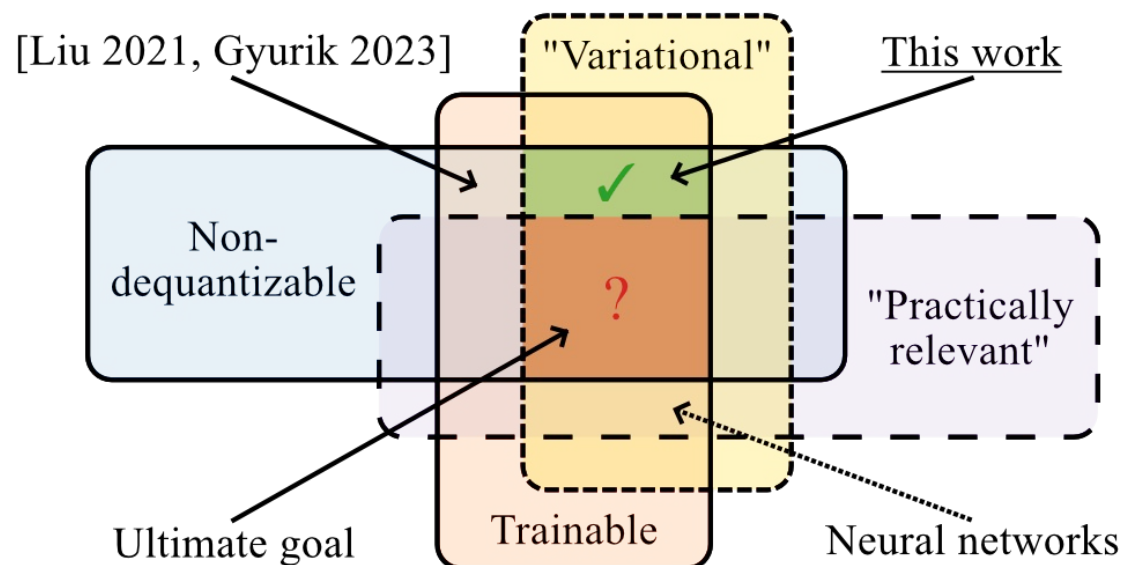
Proof:

1. Combine together circuits corresponding to:
 - a function that is hard to compute and hard to learn.
 - a function that is easy to optimize.
2. Gives gradient-based trainable and non-dequantizable.
3. Massage until it looks vari veryational

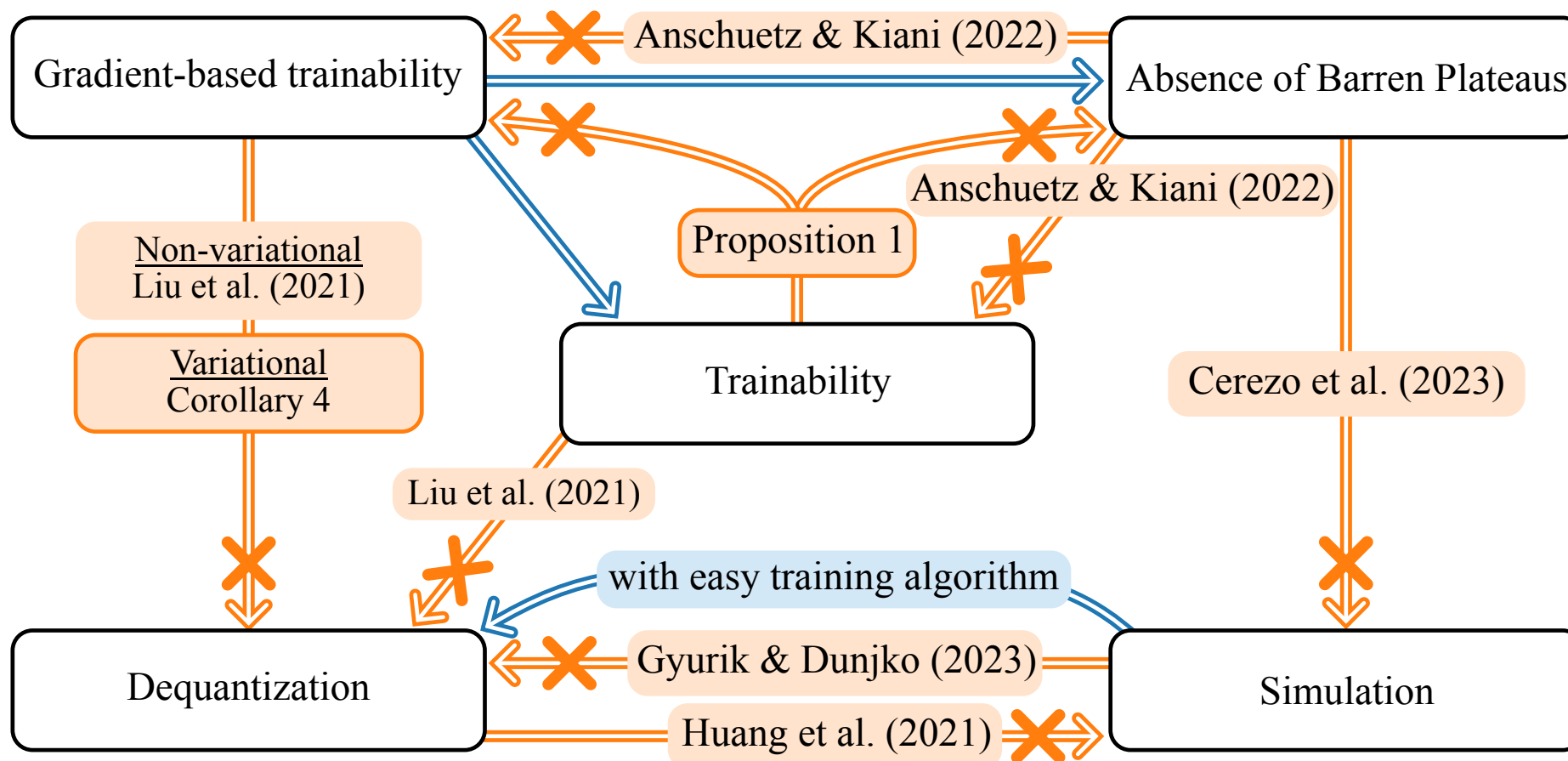


How: Main result

Theorem: there exist **variational** QML models which are **(gradient-based) trainable** and **non-dequantizable** (give quantum advantage).

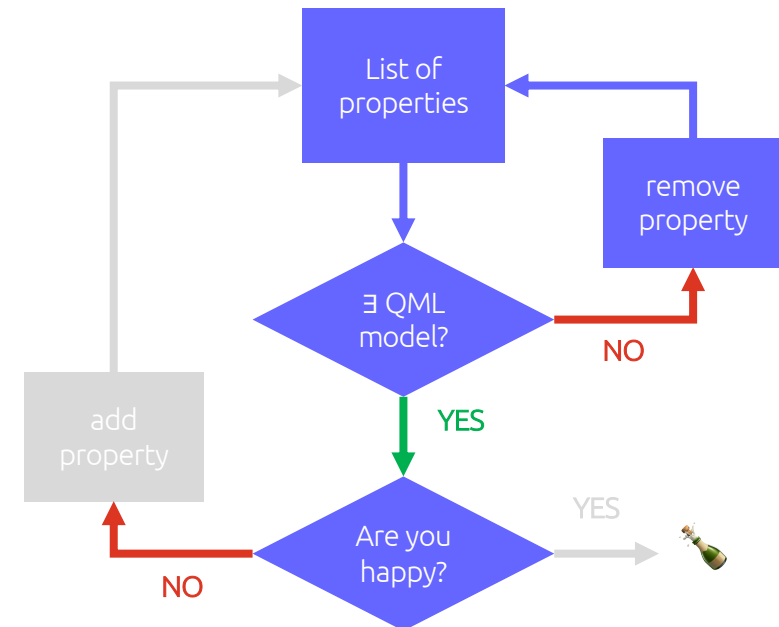


How: Cleaning up the house



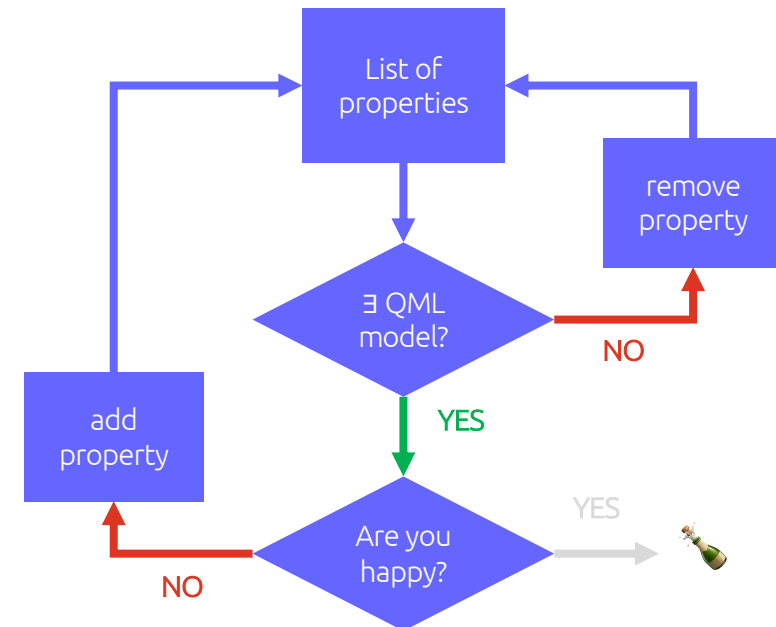
What's next?

- Are we happy? NO
 - Our recipe gives a QML model specific to a single learning task.



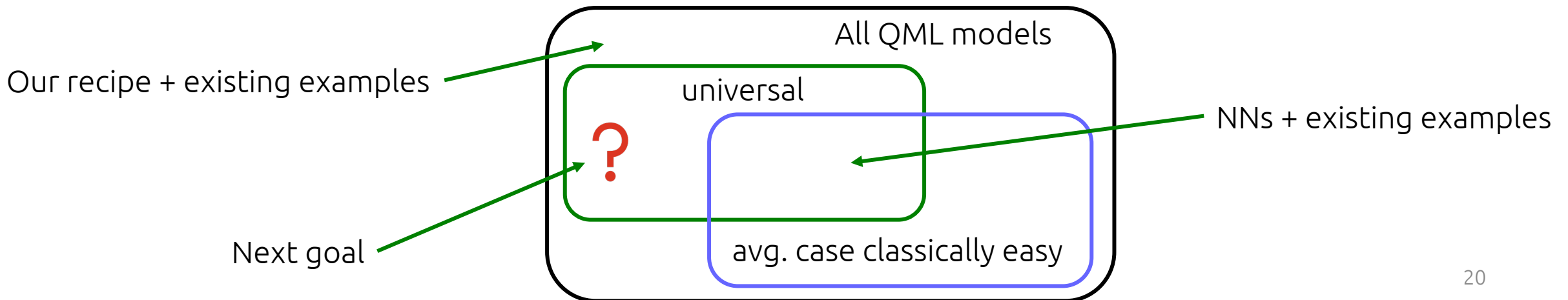
What's next?

- Are we happy? NO
 - Our recipe gives a QML model specific to a single learning task.
- Think of another property:
 - NNs are great because they are “general purpose”.



What's next?

- Are we happy? NO
 - Our recipe gives a QML model specific to a single learning task.
- Think of another property:
 - NNs are great because they are “general purpose”.
- Can we have general-purpose QML models that offer quantum advantage?
 - “General purpose”: universality! (Known examples, no quantum advantage)
 - “Quantum advantage”: function family not average-case easy (Known examples, not universal)



References

QML models with nice properties:

Havlíček, *et al.*, *Nature* **567** (2019)
Schuld, Killoran, *PRL* **122** (2019)
Caro, Datta, *Quantum Machine Intelligence* **2** (2020)
Pérez-Salinas, Cervera-Lierta, GF, Latorre, *Quantum* **4** (2020)
Liu, Arunachalam, Temme, *Nature Physics* **17**(2021)
Caro, *et al.*, *Quantum* **5** (2021)
Caro, *et al.*, *Nature Communications* **13** (2022)
Gyurik, Dunjko, *arXiv:2306.16028* (2023)
Caro, *et al.*, *Nature Communications* **14** (2023)
Yu, *et al.*, *NeurIPS 2024* (2024)
Jerbi, *et al.*, *Nature Communications* **15** (2024)
GF, Eisert, Bravo-Prieto, *Nature Communications* **15** (2024)
Kempkes, *et al.*, *PRX Q* **7** (2026)
Rodríguez-Grasa, *et al.*, *arXiv:2603.22964* (2026)

“Trainability neither necessary nor sufficient for absence of BPs”:

Anschuetz, Kiani, *Nature Communications* **13** (2022)
This work

“Dequantization not sufficient for simulation”:

Schreiber, Eisert, Meyer, *PRL* **131**, 100803 (2023)
Huang *et al.*, *Nature Communications* **12** (2021)

“Folklore”:

Cerezo *et al.*, *arXiv: 2312.09121* (2023)

Our work: “Variational AND gradient-based trainability AND non-dequantization”:

GF, Gyurik, Perez-Salinas, Dunjko, *ICLR 2025* (2025)

Thank you for your attention!

The quest for good QML models

Questions from Matthias you?

How: Main result

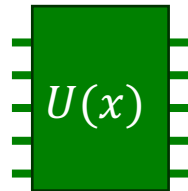
Theorem: there exist **vari veryational** QML models which are **gradient-based trainable** and **non-dequantizable**.

Proof:

Function $Q(x)$ is

1. efficiently quantum evaluable
2. hard classically evaluable (even with data)

$$Q(x) = \langle 0 | U^\dagger(x) Z_0 U(x) | 0 \rangle$$



Important: we know these functions exist.
Examples in [LAT21, GD23]

Circuit $W(\vartheta) = \prod_{j=1}^L R_X(\vartheta_j)$ is

1. easy to optimize gradient-based
2. (efficiently classically evaluable)

$$\vartheta^* \leftarrow \operatorname{argmax}_{\vartheta \in \Theta} \langle 0 | W^\dagger(\vartheta) Z_0 W(\vartheta) | 0 \rangle$$

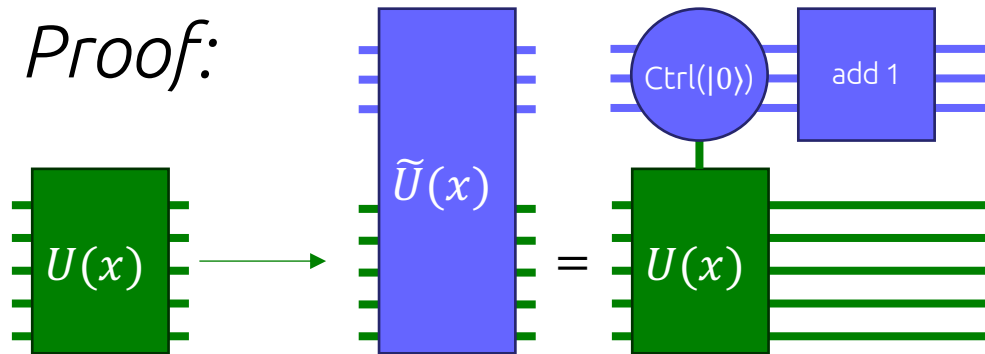


[LAT21]: Liu, Arunachalam, Temme, Nat. Phys. (2021)
[GD23]: Gyurik, Dunjko, 2306.16028 (2023)

How: Main result

Theorem: there exist **variational** QML models which are **gradient-based trainable** and **non-dequantizable**.

Proof:

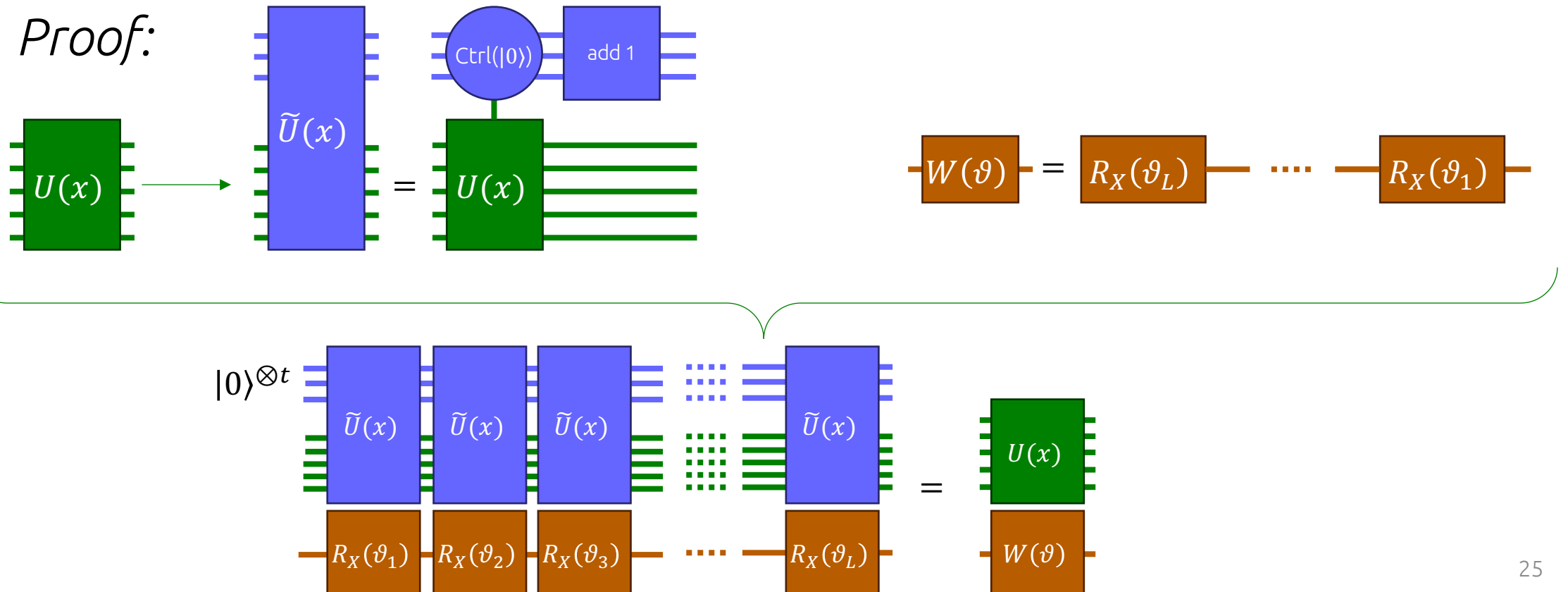


$$W(\vartheta) = R_X(\vartheta_L) \cdots R_X(\vartheta_1)$$

How: Main result

Theorem: there exist **variational** QML models which are **gradient-based trainable** and **non-dequantizable**.

Proof:

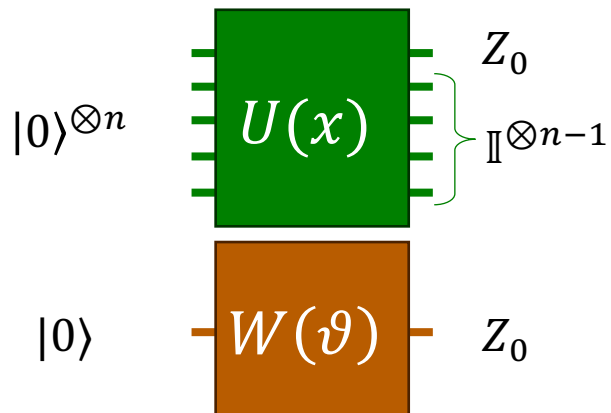


How: Main result

Theorem: there exist **vari veryational** QML models which are **gradient-based trainable** and **non-dequantizable**.

Proof:

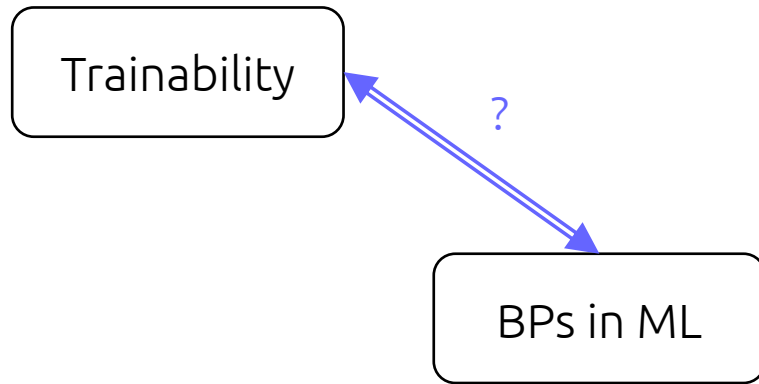
$$f_{\vartheta}(x) = \langle 0 | (U(x) \otimes W(\vartheta))^{\dagger} (Z_0 \otimes Z_0) (U(x) \otimes W(\vartheta)) | 0 \rangle$$



$$f_{\vartheta}(x) = Q(x) \langle Z \rangle_{\vartheta}$$

1. Non-dequantizable: classically intractable to evaluate.
2. Gradient-based trainable: from easy optimization algorithm.
3. Vary veriational: circuit checks all the boxes.

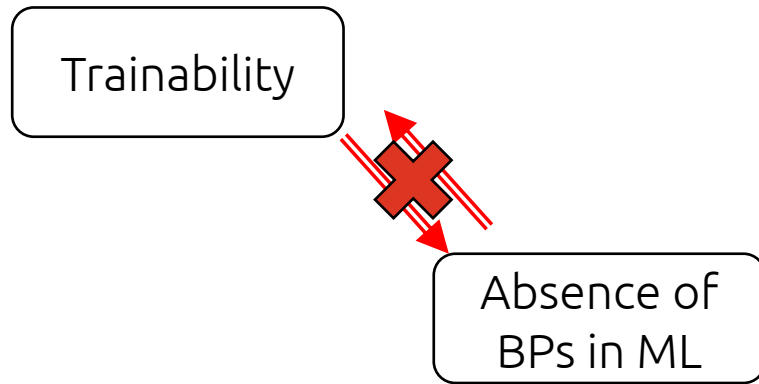
How: Definitions



Intuition?

Concentration versus trainability?

How: Definitions



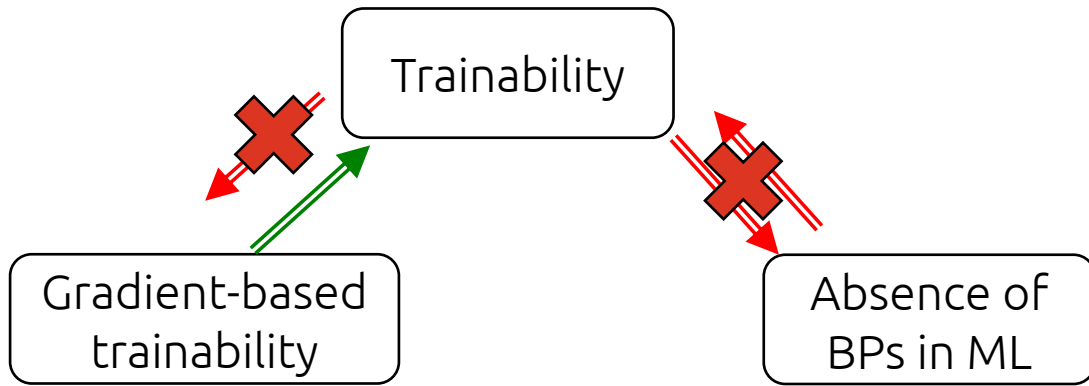
Trainability is neither necessary nor sufficient for absence of BPs.

1. Without BP we can get stuck in bad local minima [AK22].
2. With BP we can use a contrived training algorithm [This work].

[AK22]: Anschuetz, Kiani, *Nat. Comms.* (2022)

[This work]: GF *et al.*, 2406.07072 (2024)

How: Definitions

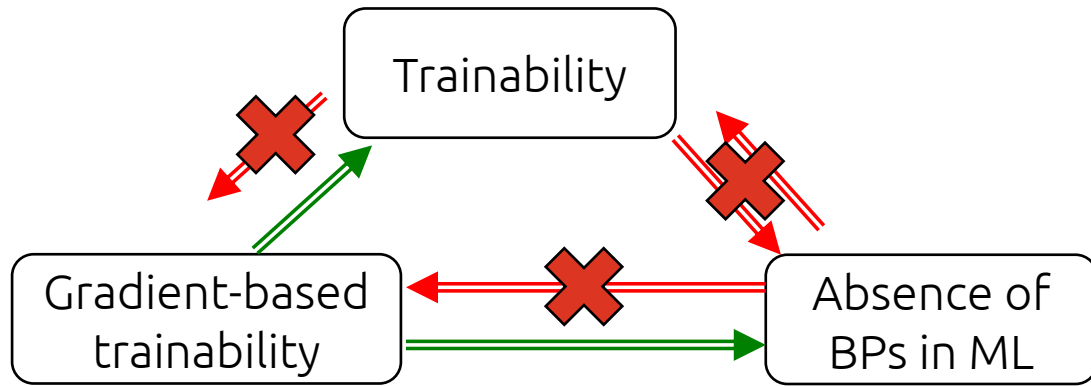


Restrict to less powerful training algorithms.

Capture more realistic scenarios.

(More difficult to prove things).

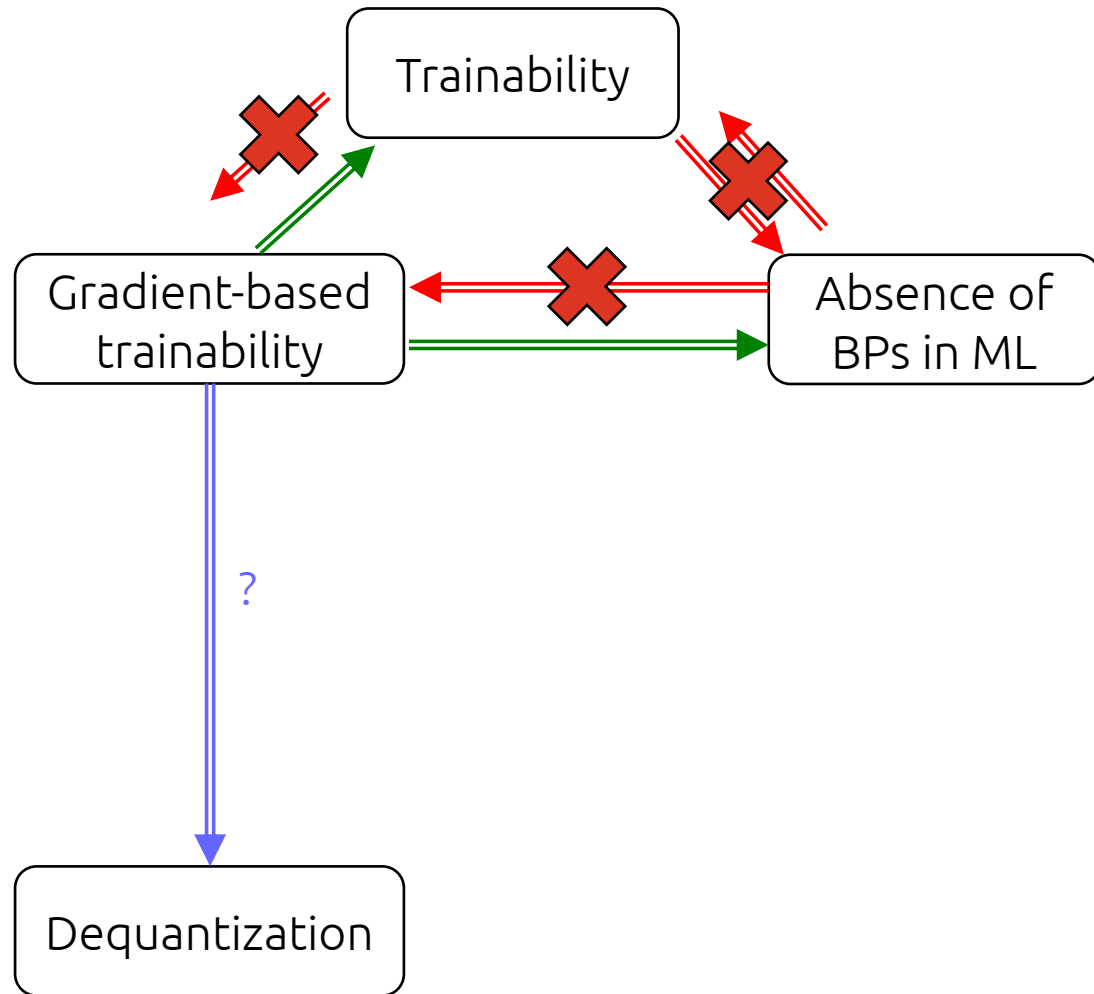
How: Definitions



Gradient-based trainability is sufficient for absence of BPs, but not necessary.

1. Without BP we can get stuck in bad local minima [AK22].
2. With BP we cannot train via gradients.

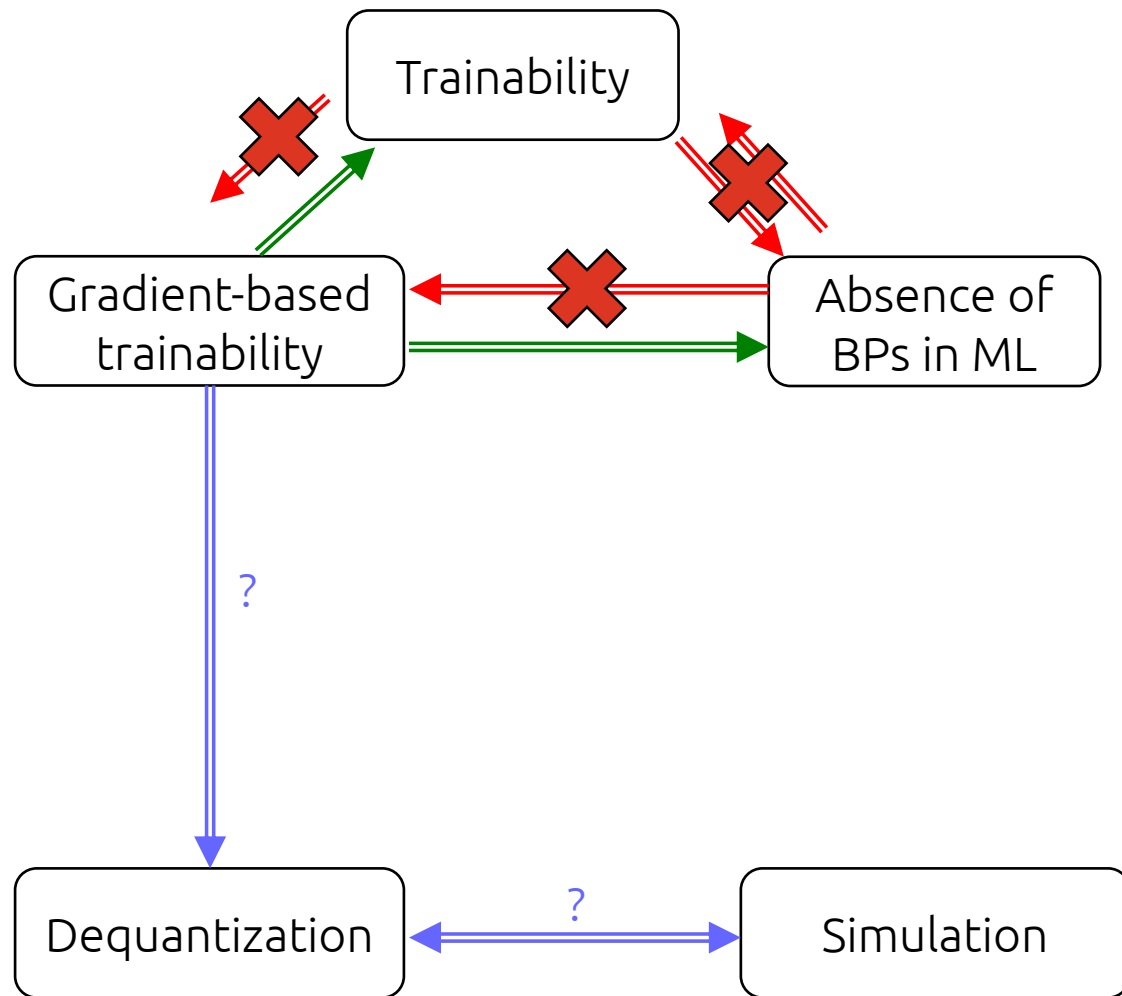
How: Definitions



This is the question we care about.

Dequantization is what kills QML.

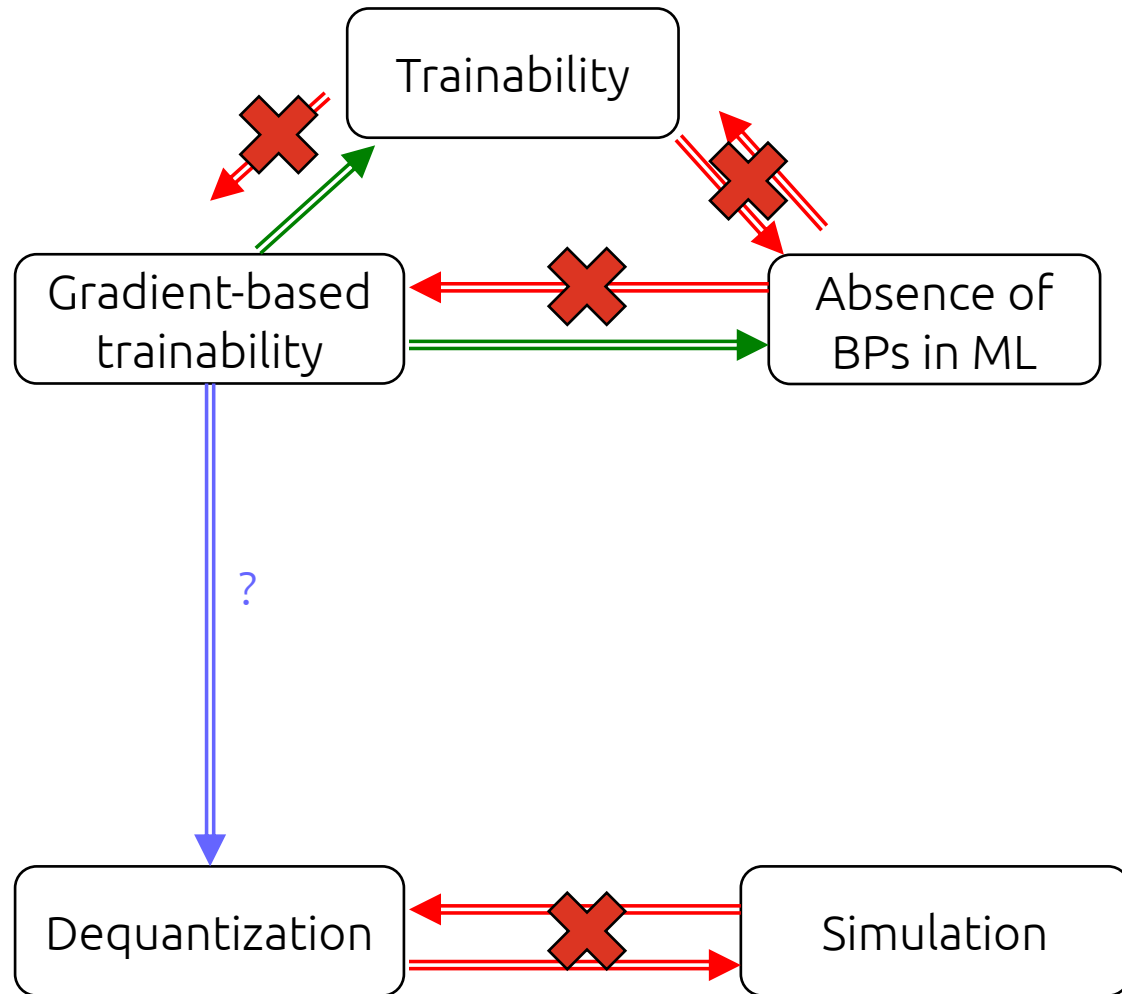
How: Definitions



Intuition?

Dequantization versus simulation?

How: Definitions



Simulation is neither necessary nor sufficient for dequantization *in general*.

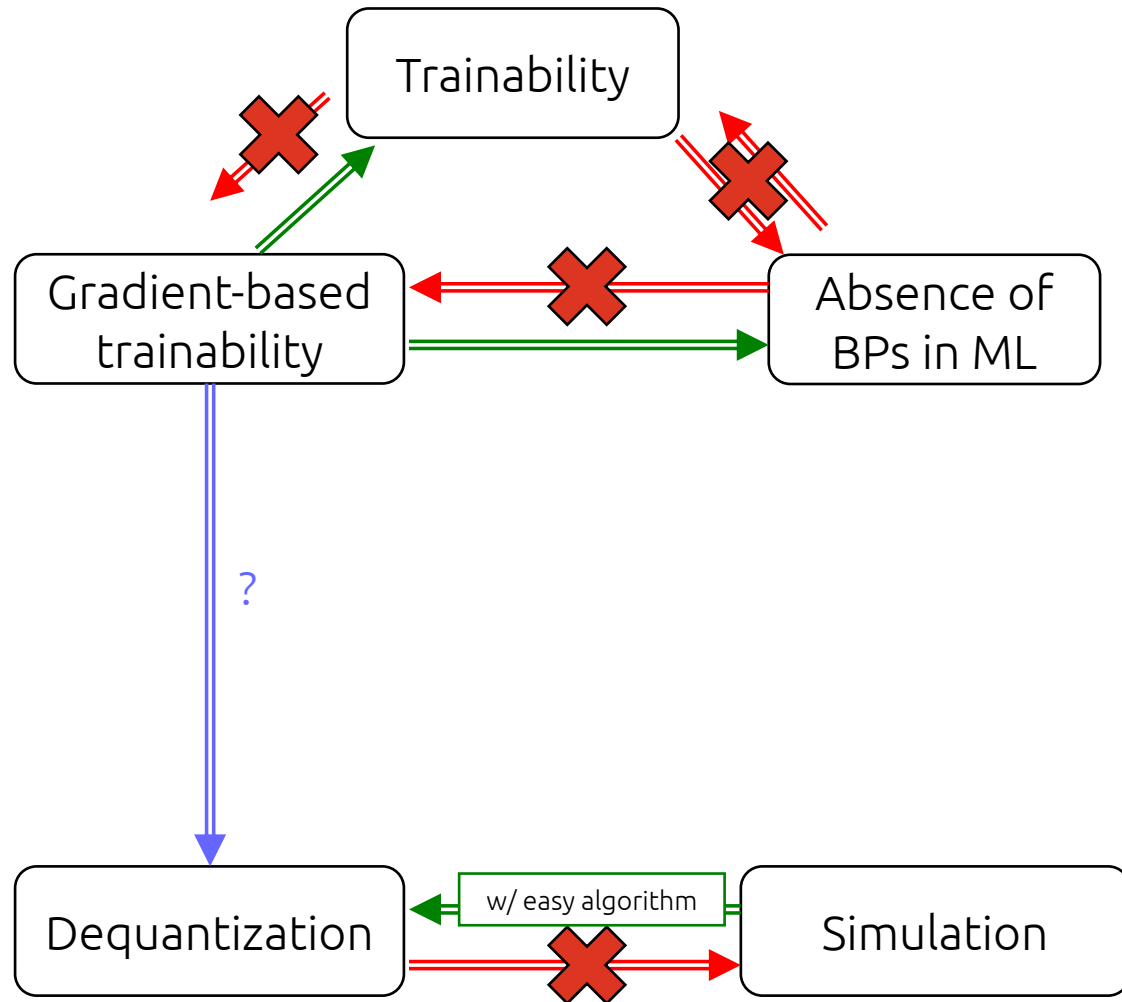
1. Dequantization without simulation as in the power of data & classical surrogates [Hua+21, SEM23].
2. Simulation without dequantization if training algorithm is classically hard [GD23].

[Hua+21]: Huang *et al.*, *Nat. Comms.* 12 (2021)

[SEM23]: Schreiber, Eisert, Meyer, *PRL* 131 (2023)

[GD23]: Gyurik, Dunjko, 2306.16028 (2023)

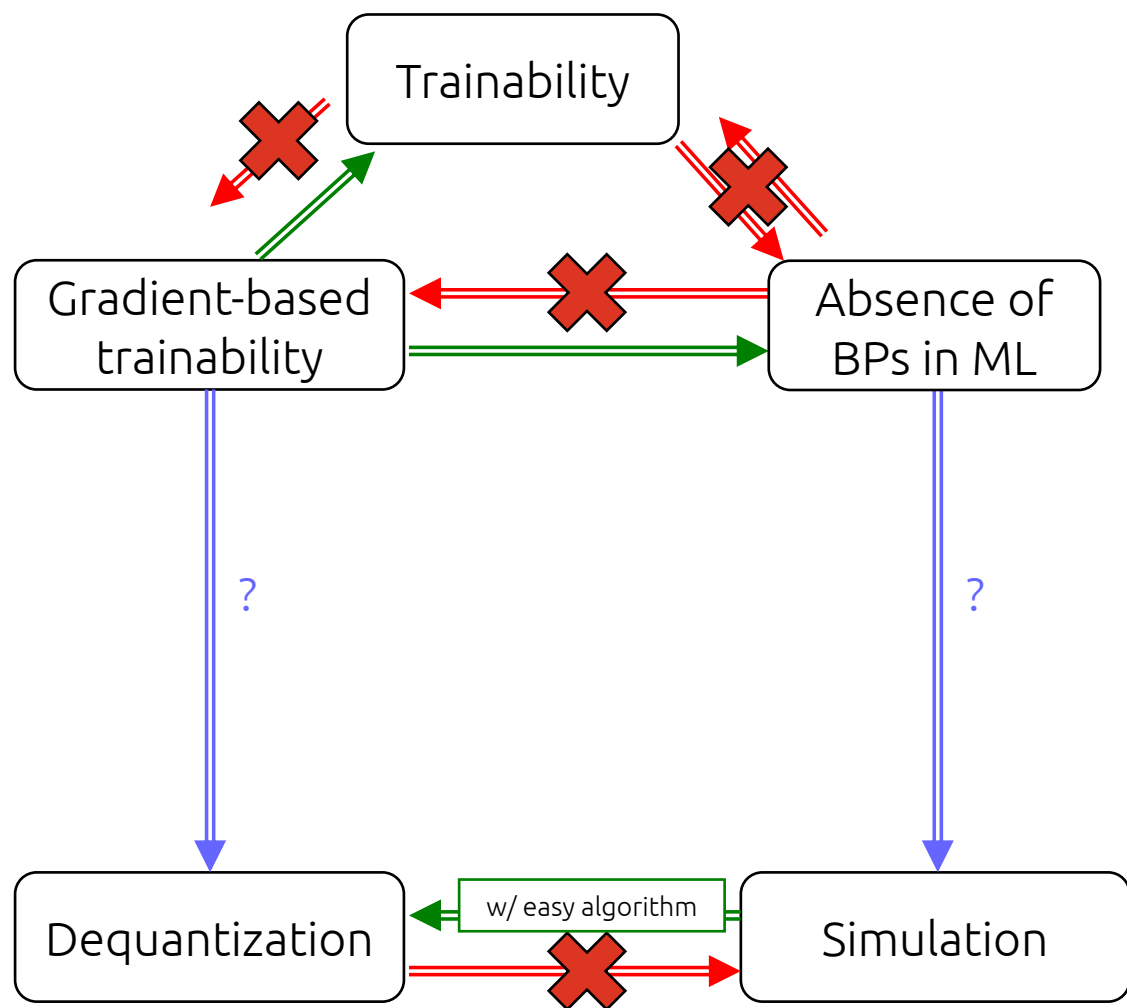
How: Definitions



Simulation is sufficient for dequantization if training algorithm is classically easy.

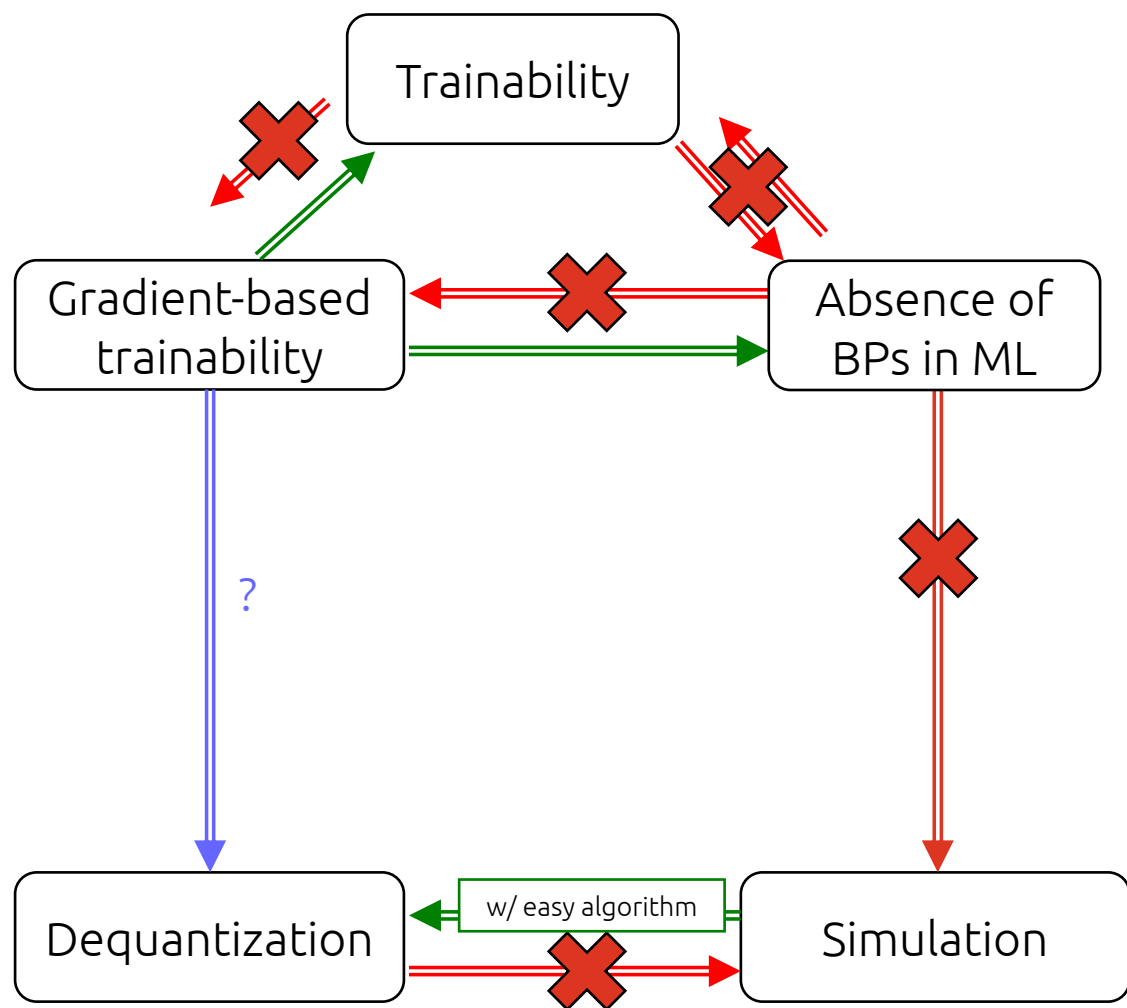
Direct simulation of the circuit gives classical model.

How: Definitions



Folklore? Panic? Intuition?

How: Definitions



Absence of BPs is not sufficient for classical simulation.

[Cer+23]: Proof by counterexample.

Then what's the folklore?

“for all known proofs of absence of BPs¹ we have a proof of classical simulation²” [Cer+23].

¹: for generic, non-contrived PQC.

²: with quantum pre-computation.