Reinforcement Learning to Stabilize Nonequilibrium Phases of Matter

Dr. Giovanni Cemin (MPIPKS)

Quant25 Conference Notes

Théo HUET

Max-Planck-Institut für
Chemische Physik fester Stoffe

27 octobre 2025

Table des matières

1 Introduction

During my participation in the **Quant25** summer program, I had the opportunity to attend a series of advanced lectures and conferences on modern topics in quantum physics and quantum information. Among them, the talk given by **Dr. Giovanni Cemin** at the Max Planck Institute for the Physics of Complex Systems (MPIPKS) particularly caught my attention.

The presentation, entitled *"Reinforcement Learning to Stabilize Nonequilibrium Phases of Matter"*, combined two areas that I find especially fascinating : the physics of quantum many-body systems and the use of artificial intelligence to understand and control their dynamics. It provided a concrete example of how modern computational techniques can go beyond passive data analysis to become tools for **active control and feedback** in complex quantum systems.

This conference was also an excellent opportunity for me to deepen my understanding of **quantum information theory**, **entanglement dynamics**, and the emerging field of **machine learning applied to quantum control**. Through this talk, I learned how concepts from reinforcement learning (RL) can be transferred to the quantum domain, where agents interact with partially observable environments and learn how to influence quantum evolution in real time.

The talk by Dr. Cemin explored how **reinforcement learning (RL)** can be applied to control and stabilize quantum many-body systems that are out of equilibrium. In particular, RL was used to learn active feedback strategies that prevent uncontrolled **entanglement growth** in large stabilizer circuits.

This line of research belongs to a broader effort to understand how machine learning can be used not only to analyze data from quantum systems, but also to *actively control* their dynamics using limited information in real time.

2 Reinforcement Learning and Quantum Control

In classical machine learning, reinforcement learning refers to a process in which an agent interacts with an environment and learns to maximize a **reward** through trial and error. At each step, the agent :

— observes a partial state of the system,
— chooses an action according to its policy,
— receives a reward that reflects how good the action was.

In the quantum setting, the agent can act on a quantum circuit or Hamiltonian, applying control operations (measurements, gates, feedback pulses, etc.) to steer the system toward a desired state.

The specific algorithm used here is called **Proximal Policy Optimization (PPO)**, a popular and stable RL method used in continuous or discrete action spaces.

3 The Unitary Circuit Game

Dr. Cemin introduced the concept of the Unitary Circuit Game, an intuitive framework for visualizing how reinforcement learning can control entanglement dynamics in quantum circuits. In this model, the system is represented as a one-dimensional chain of qubits evolving under a combination of entangling and disentangling operations applied between neighboring sites.

Two types of operations :

— **Entangling gates :** Random unitary operations that increase correlations and spread entanglement across the chain. These mimic chaotic dynamics or local two-qubit interactions that naturally generate entanglement.
— **Disentangling operations :** Local measurements or feedback actions that reduce entanglement on selected bonds. These operations are typically *non-unitary*, as they rely on measurement outcomes and conditional control to remove correlations between subsystems.

The interplay between these two competing processes, entanglement generation and active disentanglement, produces rich nonequilibrium dynamics. Depending on how often and where disentangling actions are applied, the circuit can enter different dynamical regimes characterized by distinct scaling laws for the entanglement entropy.

Tetris analogy. To illustrate this, Dr. Cemin used the analogy of a **Tetris game** : each entangling operation adds a "block" to a growing pile, representing the accumulation of entanglement. Disentangling actions correspond to removing blocks in selected locations. If the agent fails to remove blocks in time, the entanglement "pile" overflows, corresponding to a highly entangled, chaotic state obeying a **volume law**. The reinforcement learning agent learns how to remove blocks strategically so that the pile remains bounded, effectively maintaining an **area-law** scaling of entanglement over long times.

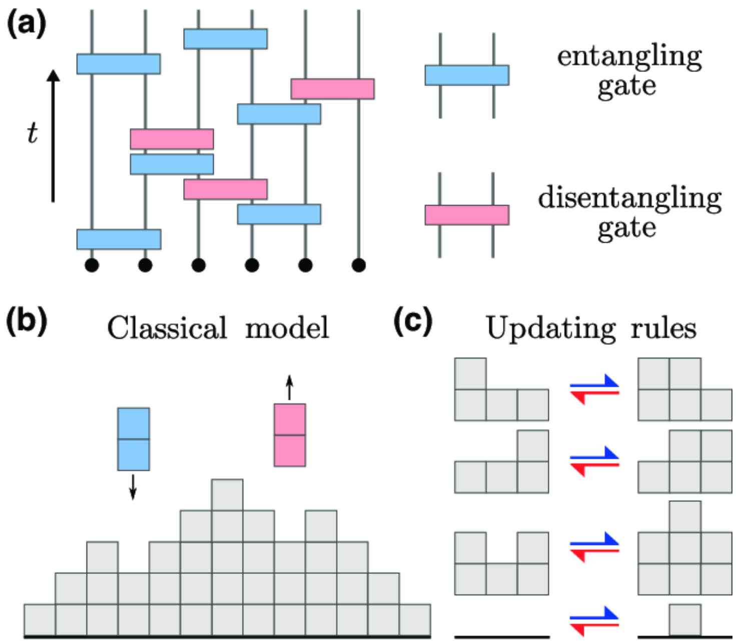

Figure 1 – Illustration adapted from G. Cemin *et al.*, *Entanglement Transitions in Unitary Circuit Games* (2024), available on ResearchGate.

— (a) Illustration of the Unitary Circuit Game : blue boxes represent random entangling unitaries, while red markers indicate local disentangling operations (measurements or feedback actions).

— (b) Classical mapping of the circuit dynamics onto a surface growth model, where adding or removing blocks corresponds to entangling or disentangling events.

— (c) Update rules in the classical picture : the entangler (blue arrow) adds blocks and the disentangler (red arrow) removes them, with the constraint that adjacent heights differ by at most one and remain non-negative.

This game-like representation makes it visually clear how local feedback, even with limited information, can prevent unbounded entanglement growth and stabilize structured, low-entropy regimes in a nonequilibrium quantum circuit.

4 Entropy and the Control Objective

The target quantity in this learning problem is the total entanglement entropy of the circuit, The maximum entanglement entropy of a subsystem of size n is :

$$S_{\max} = n \ln(2),$$

for a system of n maximally entangled qubits. the reward function is designed to penalize excessive entanglement, e.g. proportional to the total or local entanglement entropy. In practice, the RL agent is trained to minimize S_{total} over time, leading to the following reward function :

$$R = -S_{\text{total}}.$$

By minimizing the entropy, the agent effectively learns how to maintain a controlled, **area-law entangled state** instead of a volume-law one.

- **Volume-law scaling :** $S(L) \propto L$, entanglement grows linearly with subsystem size.
- **Area-law scaling :** $S(L) \propto$ const., entanglement saturates, indicating stability.

This shift from volume-law to area-law reflects a transition to a more structured, low-entropy phase, achieved not by static design, but through active feedback.

5 Partial Observability and Active Feedback

One key challenge emphasized in the talk is that the RL agent has access only to **partial information** about the system, not the full quantum state. This makes the control problem much more realistic, since real quantum experiments often cannot reconstruct the complete wavefunction due to measurement constraints.

Despite this limitation, the agent can learn approximate, near-optimal strategies by recognizing patterns in the observable quantities it can measure. This is an example of an **out-of-equilibrium learning process** : the control policy is dynamically adapted in real time, not precomputed from static data.

The success of such strategies shows that active feedback based on limited information can still stabilize complex quantum dynamics, something that human-designed static control rules cannot easily achieve.

6 Emergent Structures and Bottlenecks

An intriguing result discussed by Dr. Cemin is that the learned RL strategies spontaneously created spatial structures that were not explicitly programmed. In particular, the control actions tend to form **bottlenecks** in the circuit, regions that restrict entanglement flow between different parts of the system.

This leads to characteristic **pyramidal patterns** in the spatial entanglement profile. The base of the pyramid corresponds to regions of high entanglement, while the top corresponds to disentangled zones created by the agent's feedback.

These emergent structures act like "walls" that split the system into smaller subsystems, effectively limiting how far correlations can spread.

Such behavior demonstrates that machine learning can discover novel control mechanisms that go beyond what we could intuitively design by hand.

7 Stabilizing Nonequilibrium Phases

The broader goal of this research is to use learning-based feedback to stabilize exotic nonequilibrium entanglement regimes (effective phases), quantum states that would otherwise be transient or unstable.

By keeping entanglement under control, the system can remain in a metastable area-law regime even though its natural dynamics would drive it toward thermalization and chaos. This is conceptually related to the idea of a "dynamical phase transition" driven by information feedback rather than by a change in the Hamiltonian.

The learned strategies therefore act as an additional **information channel** that continuously extracts entropy from the system, a remarkable bridge between information theory, control, and quantum thermodynamics.

8 Summary

To summarize, the main ideas from Dr. Cemin's talk are :

— Reinforcement learning can discover control strategies that prevent excessive entanglement growth in quantum circuits.

— Partial observability does not prevent efficient control when the agent uses active feedback.

— The agent's learned actions generate spatial structures that act as effective entanglement bottlenecks.

— These mechanisms stabilize out-of-equilibrium phases that would otherwise thermalize rapidly.

This work illustrates how artificial intelligence can become a powerful tool to **design and maintain nontrivial quantum states**, even in regimes where human intuition or analytical control fail.

9 Conclusion

Attending Dr. Cemin's talk was a highly enriching experience that deepened my understanding of how modern machine learning tools can interface with quantum physics. Beyond the specific results on entanglement control, this work illustrates how ideas from information theory, feedback, and artificial intelligence are becoming essential for exploring and stabilizing complex quantum systems.

This conference also gave me valuable insight into the broader landscape of quantum information science, where theoretical physics, computation, and data-driven methods now interact in powerful and unexpected ways. It strengthened my interest in the intersection between **quantum control, information, and learning**, and highlighted how such approaches could eventually contribute to the development of robust quantum technologies.

Remerciements

Je voulais simplement remercier Mr Honecker et Mme Rollet qui m'ont permis d'aller a cette ecole d'été en Allemagne.

Références

[1] G. Cemin, *Reinforcement Learning to Stabilize Nonequilibrium Phases of Matter with Active Feedback Using Partial Information*, Conference talk, Quant25, Max Planck Institute for the Physics of Complex Systems (MPIPKS), Dresden, Germany, 2025. https://www.pks.mpg.de/de/quant25/scientific-programme

[2] G. Cemin, M. Kiefer-Emmanouilidis, J. Wilming, J. Eisert, and A. Chan, *Entanglement Transitions in Unitary Circuit Games*, (2024). Available at : https://www.researchgate.net/publication/377671404_Entanglement_Transitions_in_Unitary_Circuit_Games