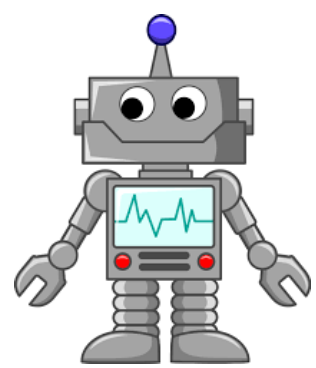
Reinforcement Learning to Control Quantum Systems Away from Equilibrium

A. G.R. Day

D. Sels

A. Polkovnikov



P. Weinberg

P. Mehta

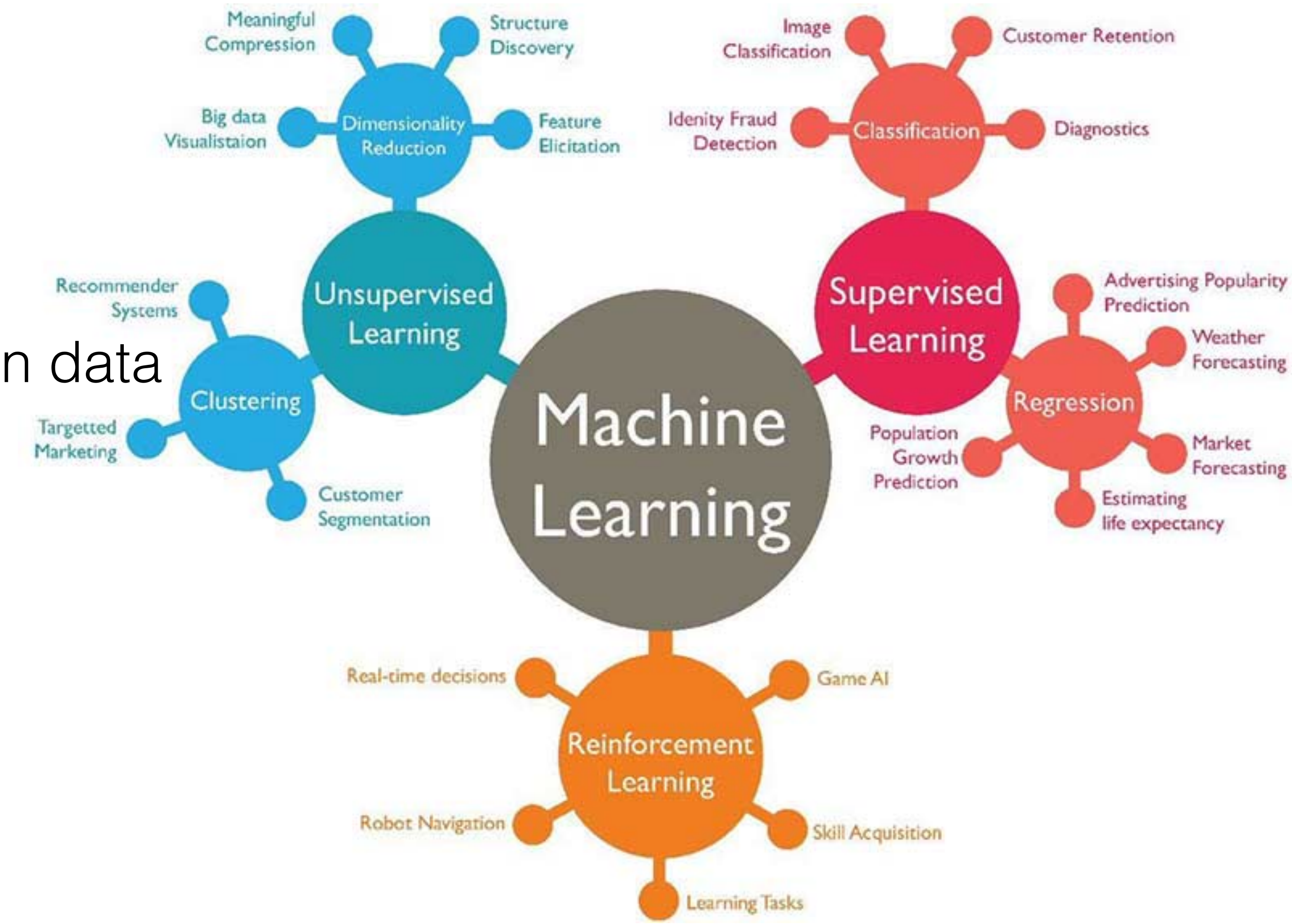
MB et al, PRX 8 031086 (2018)
MB PRB 98, 224305 (2018)

Marin Bukov

Reinforcement Learning (RL) as a branch of ML

→ Supervised Learning

- labelled data
- find approx. model which generalizes beyond known data



Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

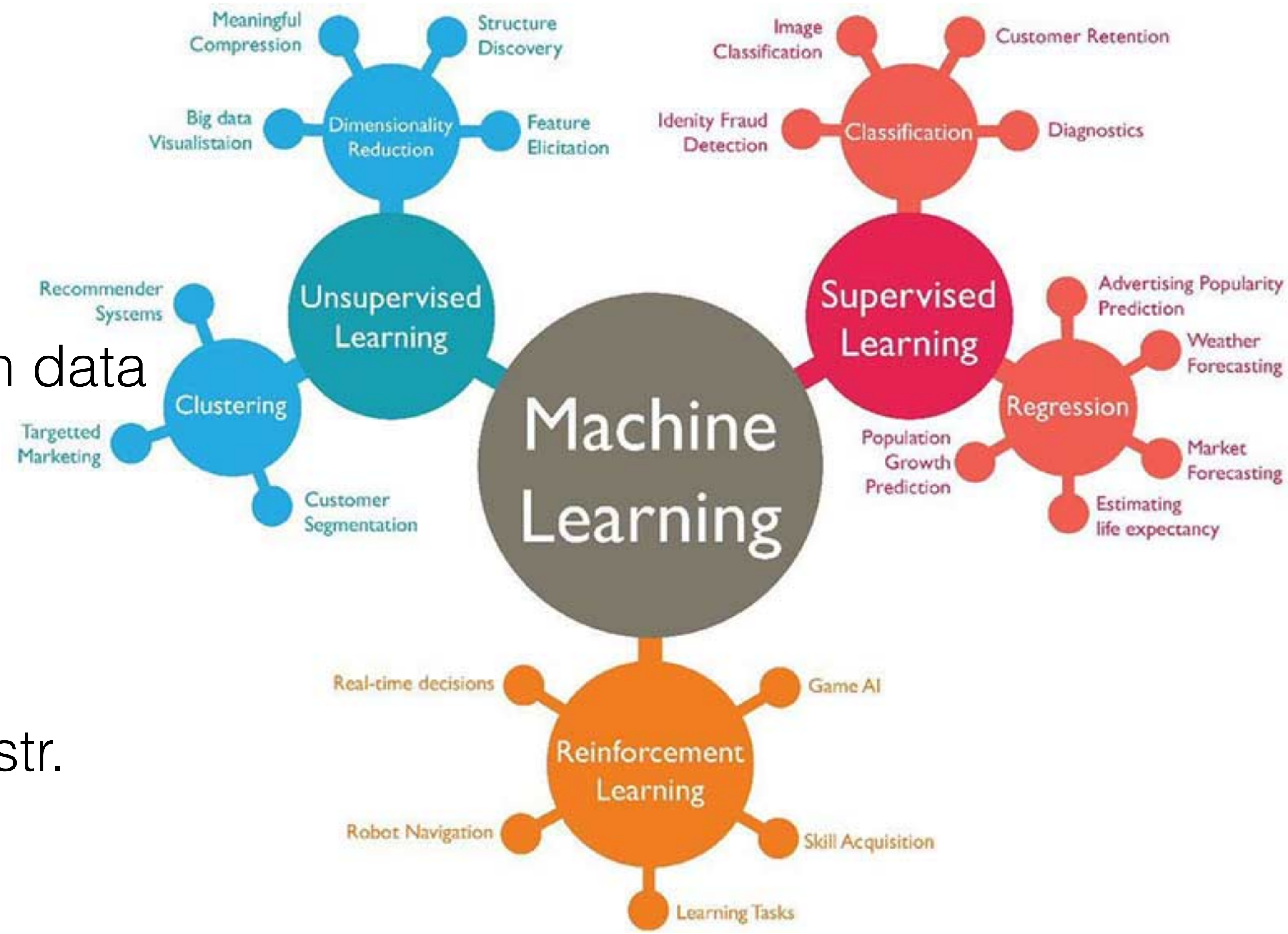
Reinforcement Learning (RL) as a branch of ML

→ Supervised Learning

- labelled data
- find approx. model which generalizes beyond known data

→ Unsupervised Learning

- unlabelled data
- find approx. probability distr. which generated the data



Marin Bukov

Reinforcement Learning (RL) as a branch of ML

→ Supervised Learning

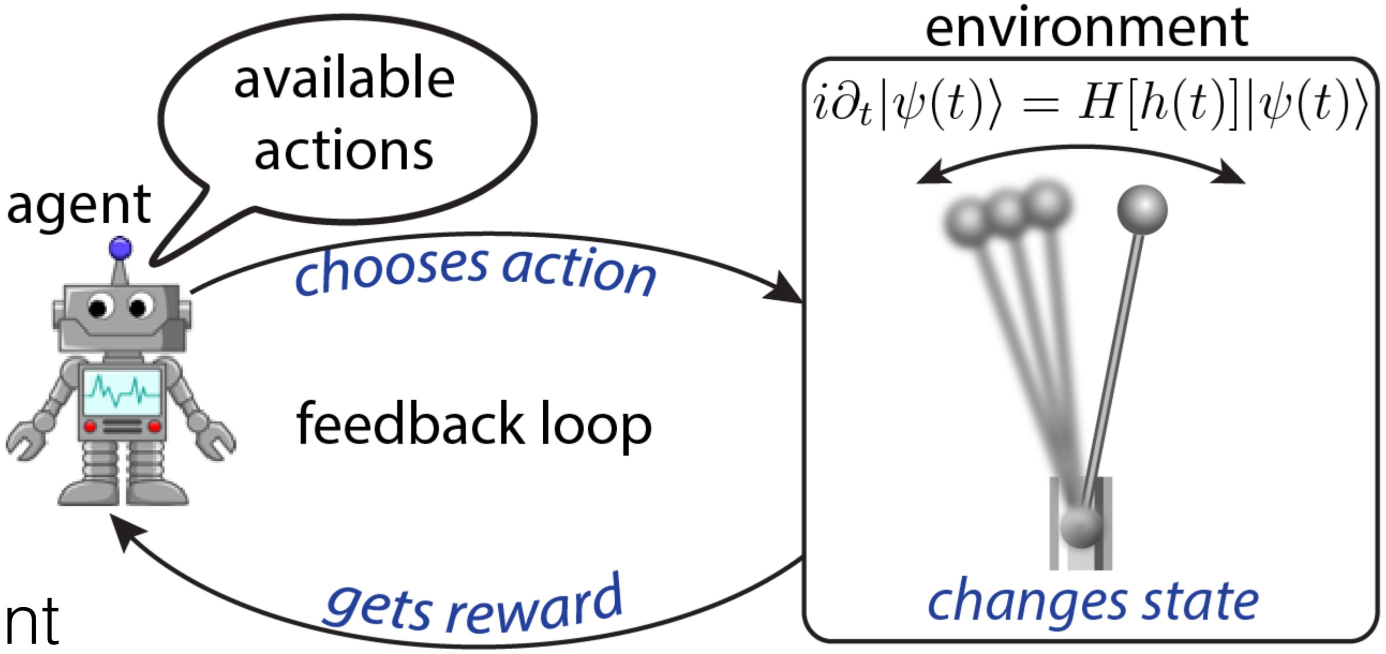
- labelled data
- find approx. model which generalizes beyond known data

→ Unsupervised Learning

- unlabelled data
- find approx. probability distr. which generates the data

→ **Reinforcement Learning**

- agent learns strategy by interactions with its environment
- probability distribution which generates the learning data changes with time due to interaction with the environment



Marin Bukov

Examples of RL Applications

outside physics

video games

Mnih et al, Nature (2015)

board games

Silver et al, Nature (2016)

locomotion

Lillicrap et al, arXiv:1509:02971

Examples of RL Applications

outside physics

video games

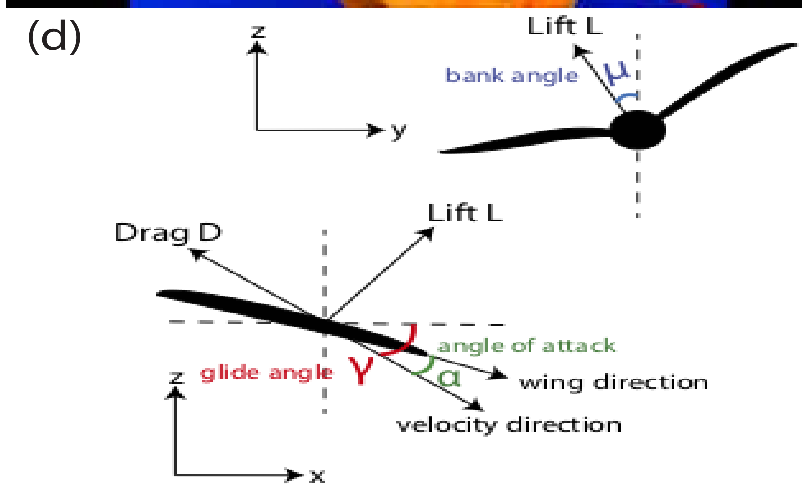
Mnih et al, Nature (2015)

board games

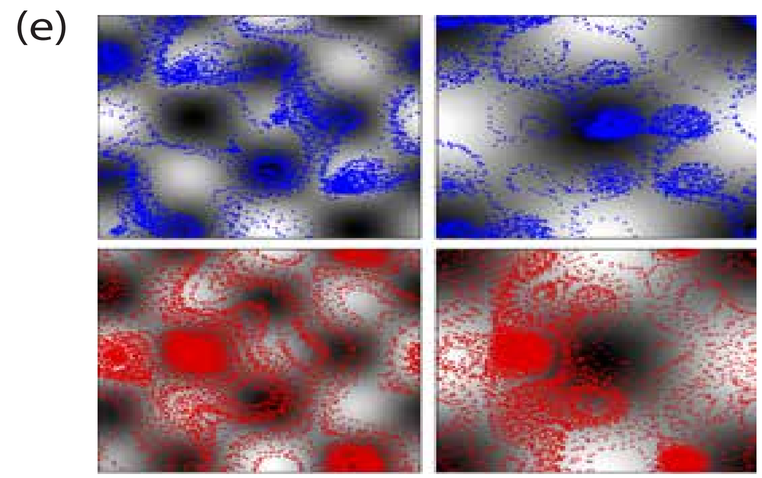
Silver et al, Nature (2016)

locomotion

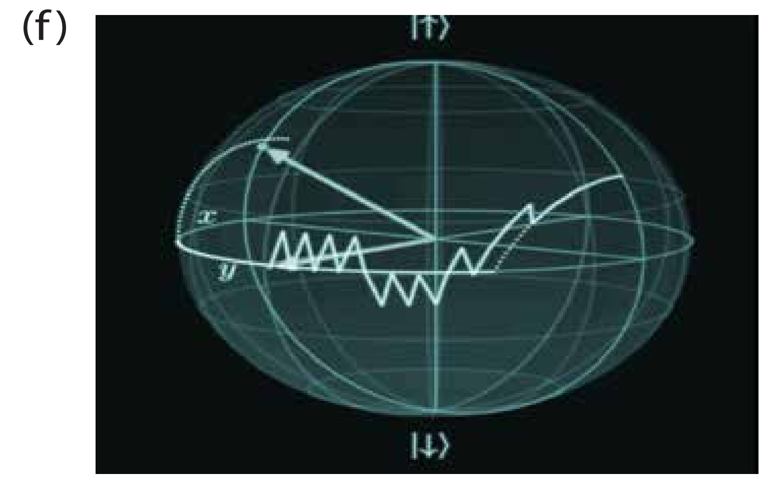
Lillicrap et al, arXiv:1509:02971



Reddy et al, PNAS 113 4877 (2016)



Colabrese et al, PRL 118 15004 (2017)



M.B. et al, PRX 8 0311086 (2018)

Fossil et al, PRX 8 031084 (2018)

August et al, arXiv:1802.04063

in physics

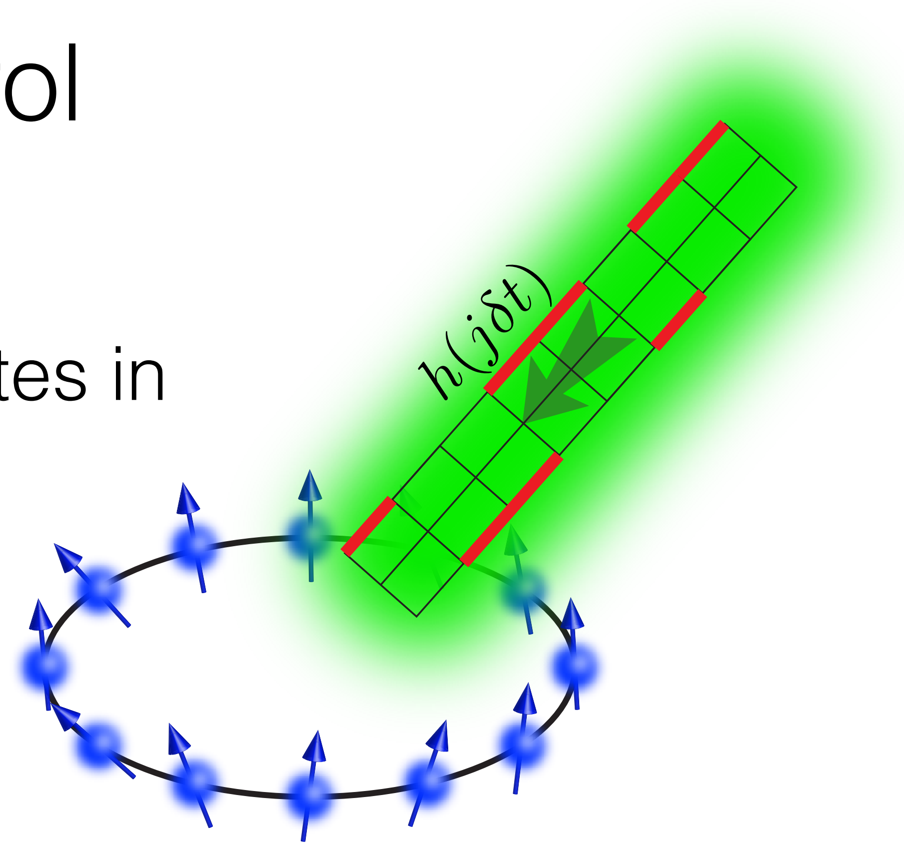
and more: design of molecular properties, quantum optics experiments, error correction, etc.

Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

in this talk:
RL for quantum control

→ Example 1: use RL to prepare many-body states in a nonintegrable spin chain

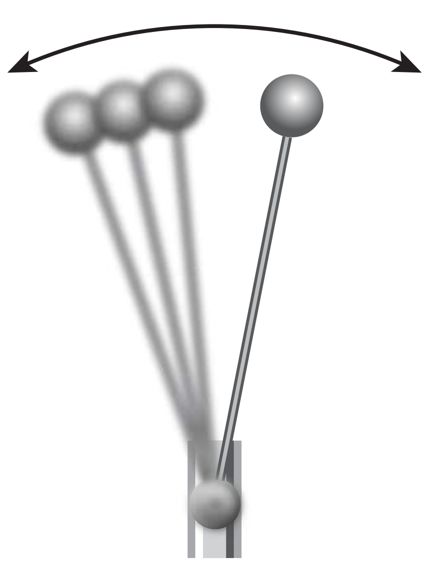


- RL and quantum control
- variational theory for optimal protocols

MB et al, PRX 8 031086 (2018)

→ Example 2: use RL to prepare states on top of strong periodic drives

- simulate quantum experiment

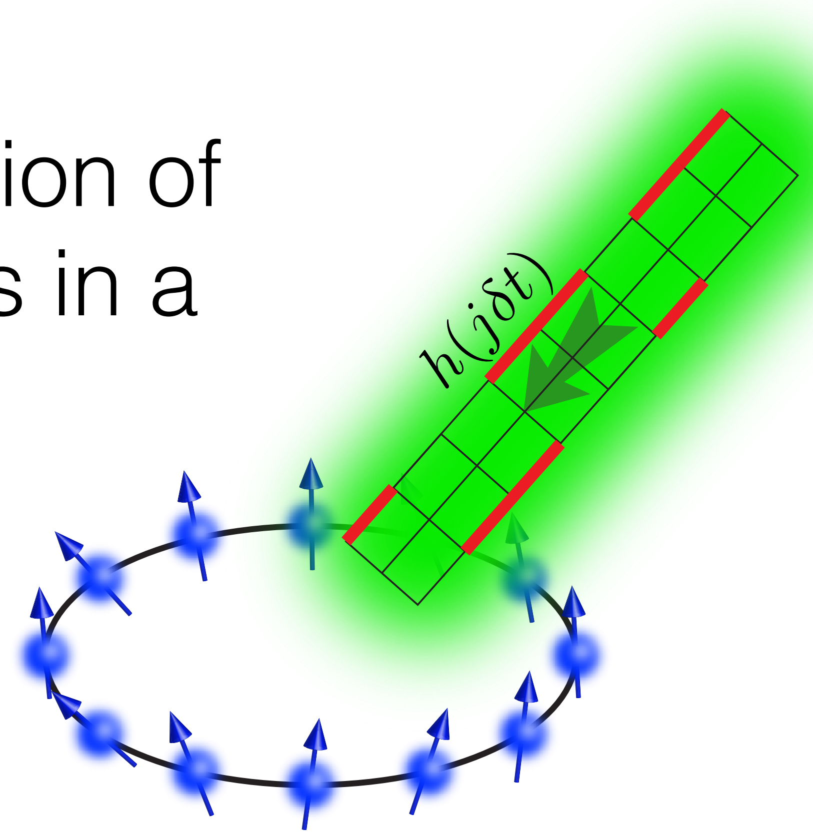


MB PRB 98, 224305 (2018)

Marin Bukov

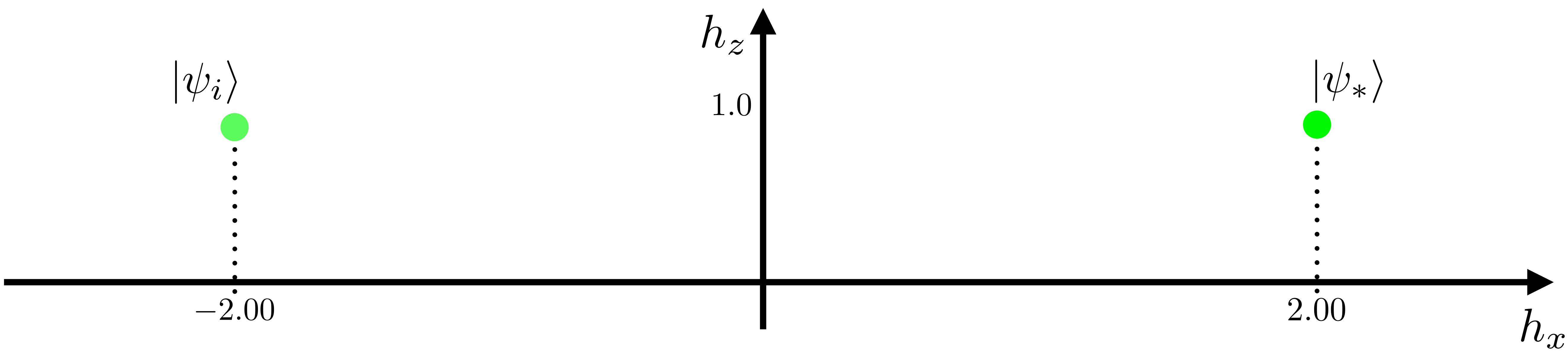
Example 1:

use RL for autonomous preparation of paramagnetic many-body states in a **nonintegrable spin chain**



$$H(t) = -\sum_{j=1}^{L} S^z_{j+1}S^z_j + \underbrace{h_z}_{=1} S^z_j + h_x(t) S^x_j$$

→ initial $|\psi_i\rangle$ and target $|\psi_*\rangle$ states are (paramagnetic) GS at:

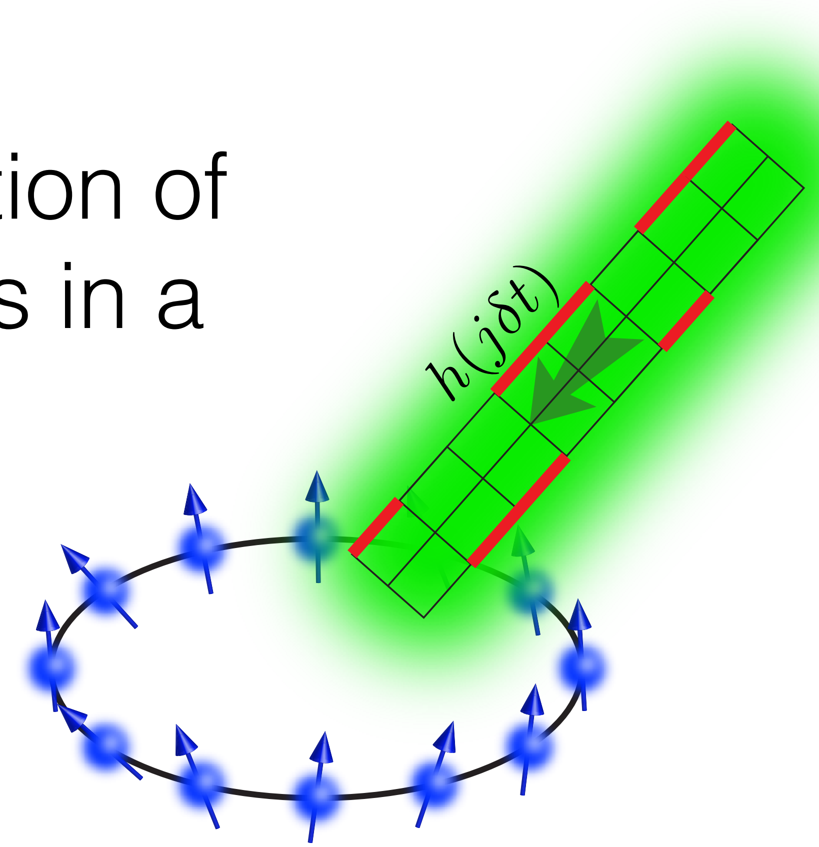


Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

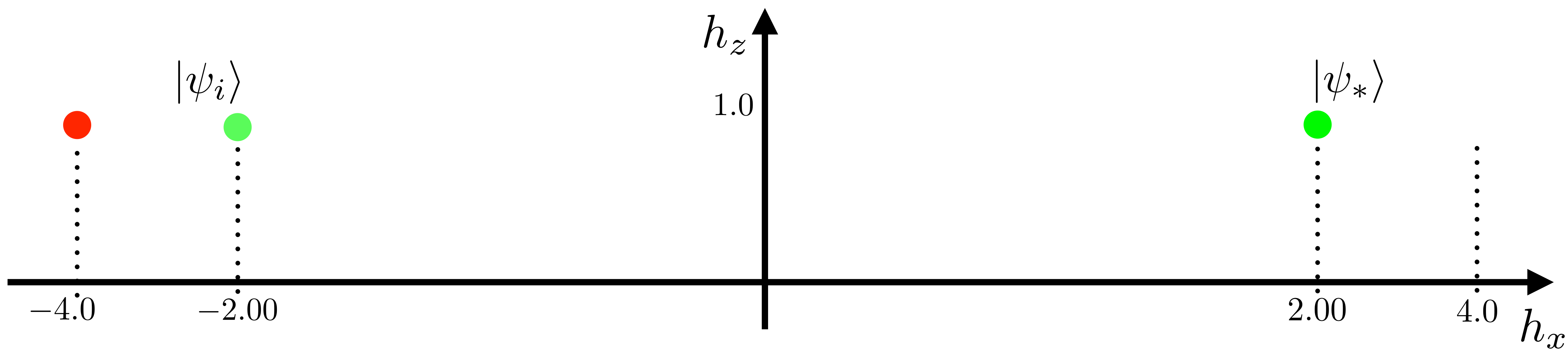
Example 1:

use RL for autonomous preparation of paramagnetic many-body states in a **nonintegrable spin chain**



$$H(t) = -\sum_{j=1}^{L} S^z_{j+1} S^z_j + \underbrace{h_z}_{=1} S^z_j + h_x(t) S^x_j$$

→ initial $|\psi_i\rangle$ and target $|\psi_*\rangle$ states are (paramagnetic) GS at:

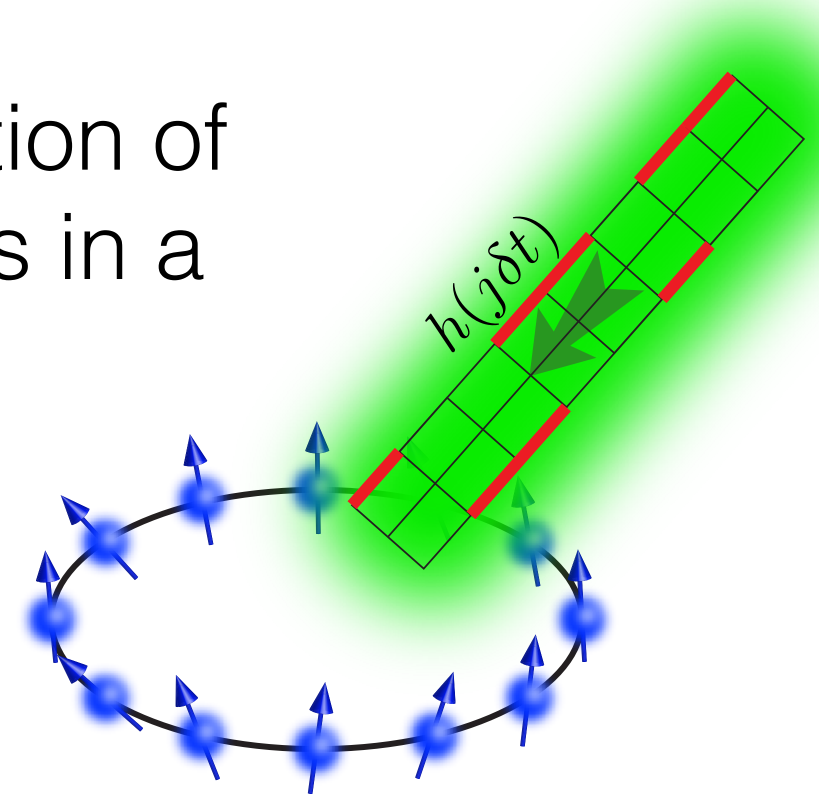


Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

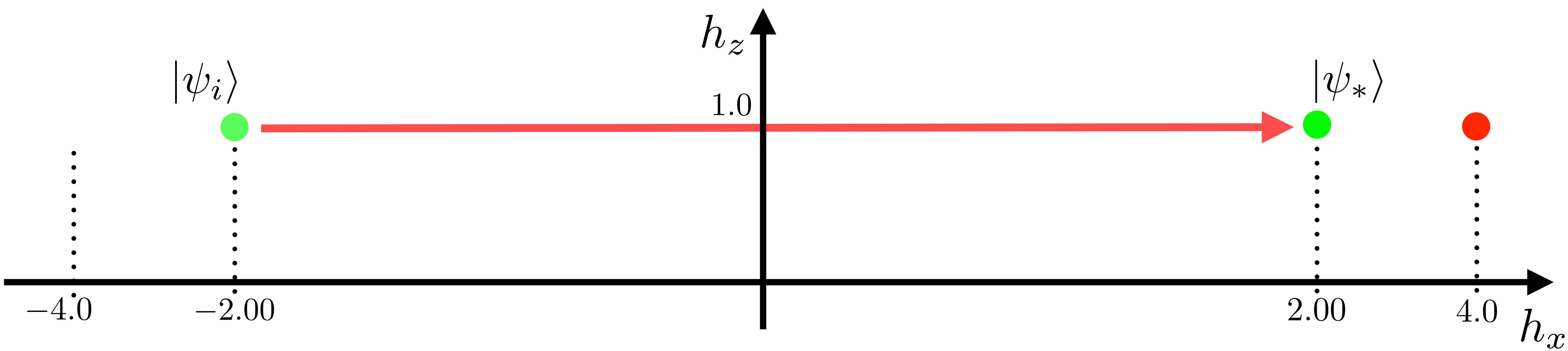
Example 1:

use RL for autonomous preparation of paramagnetic many-body states in a **nonintegrable spin chain**



$$H(t) = -\sum_{j=1}^{L} S^z_{j+1}S^z_j + \underbrace{h_z}_{=1} S^z_j + h_x(t)S^x_j$$

→ initial $|\psi_i\rangle$ and target $|\psi_*\rangle$ states are (paramagnetic) GS at:

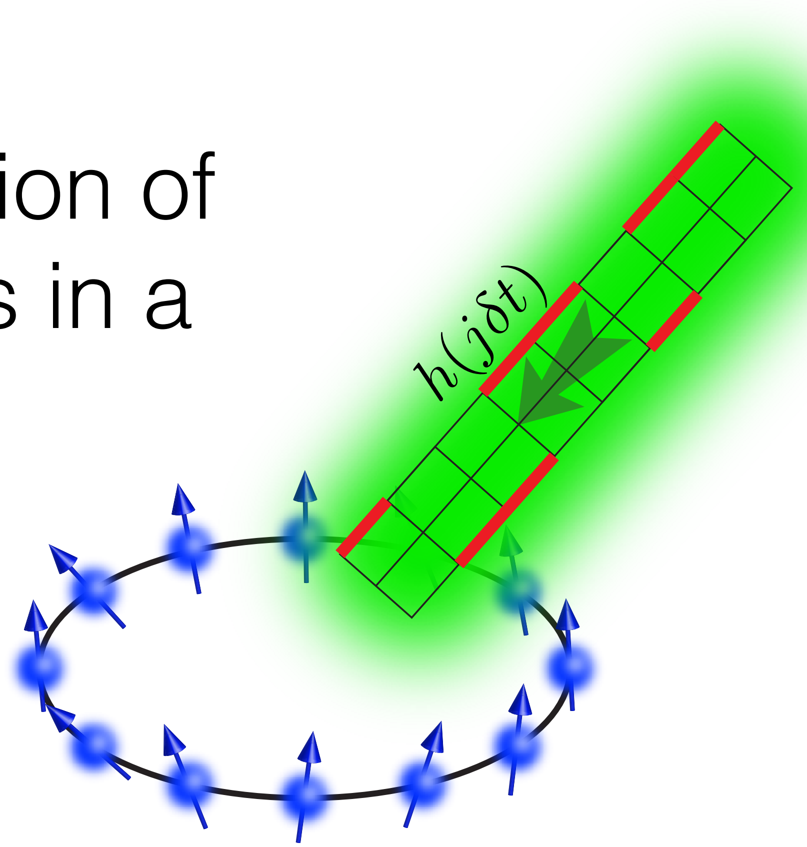


Marin Bukov

Berkeley UNIVERSITY OF CALIFORNIA

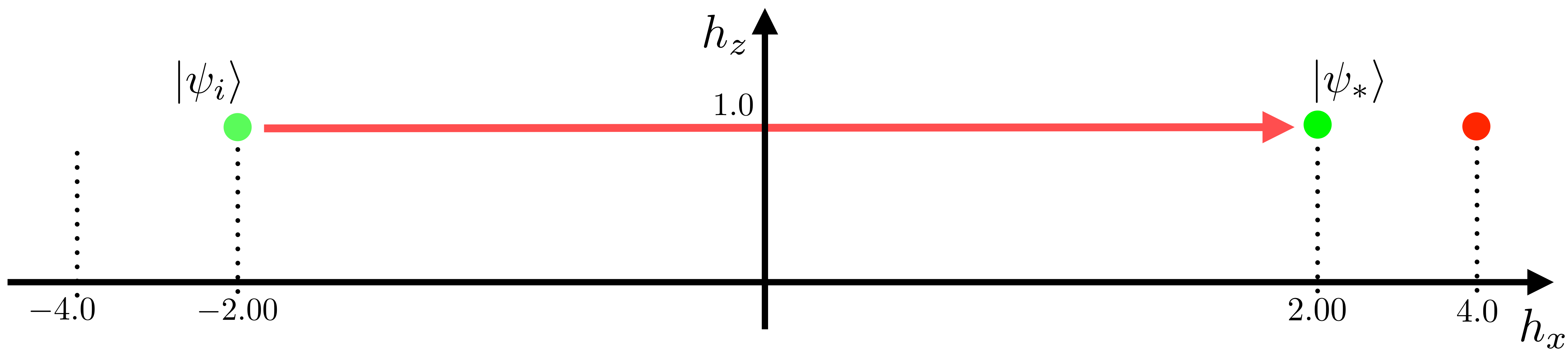
Example 1:

use RL for autonomous preparation of paramagnetic many-body states in a **nonintegrable spin chain**



$$H(t) = -\sum_{j=1}^{L} S^z_{j+1} S^z_j + \underbrace{h_z}_{=1} S^z_j + h_x(t) S^x_j$$

→ initial $|\psi_i\rangle$ and target $|\psi_*\rangle$ states are (paramagnetic) GS at:



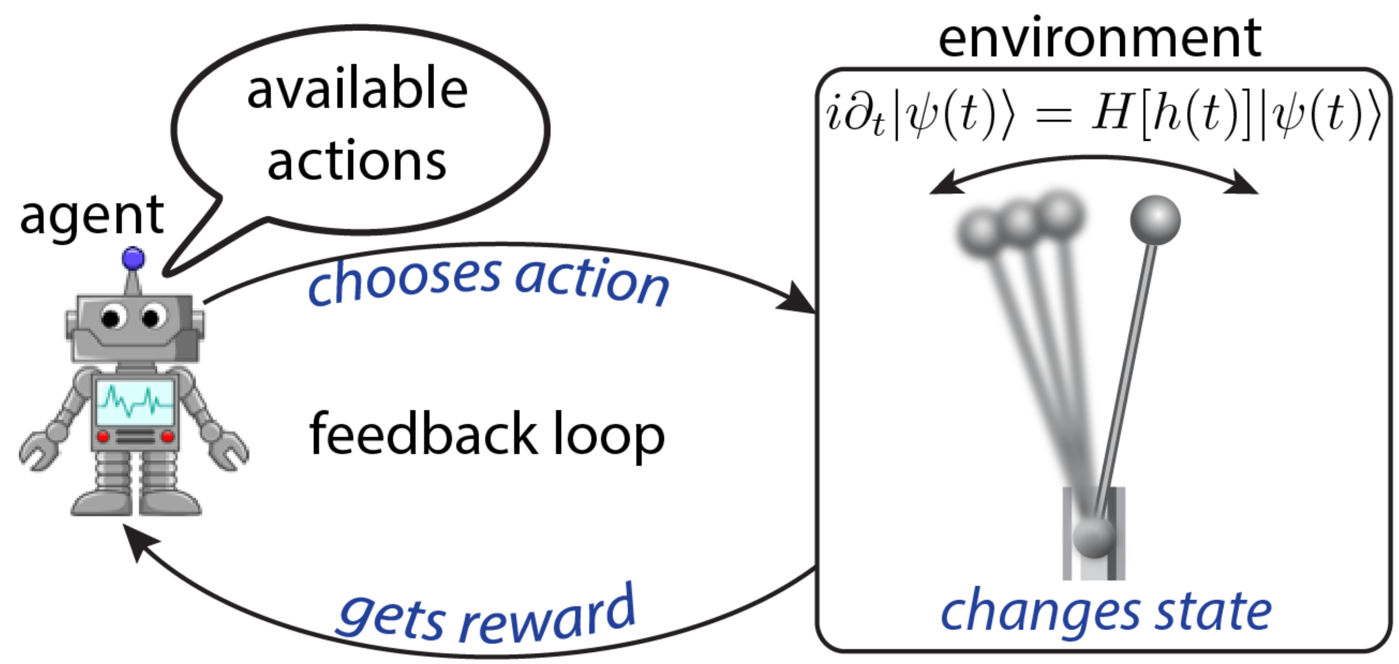
→ for example: $h_x(t) = [+4, +4, -4, +4, -4, -4, \dots]$

$\text{fixed \# of bangs, i.e. fixed total time } t_f$

Marin Bukov

Quantum Control as an RL Problem

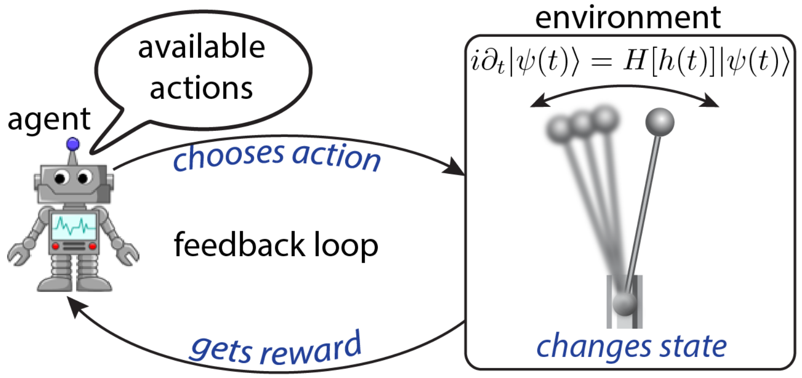
→ RL formalism



Marin Bukov

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

Quantum Control as an RL Problem



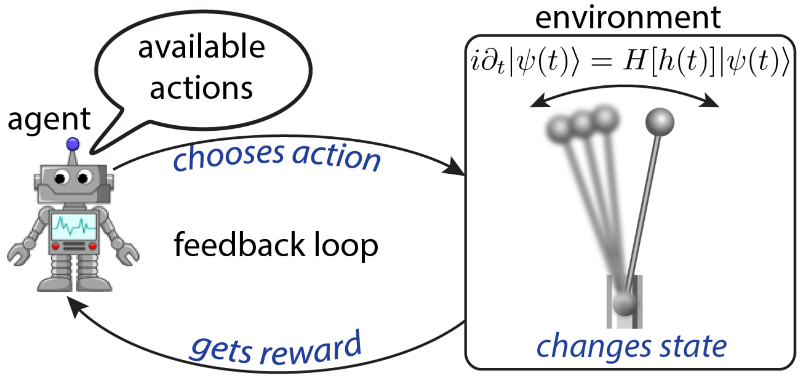
→ RL formalism

- action space $\mathcal{A} = \{+4, -4\}$

Marin Bukov

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

Quantum Control as an RL Problem



→ RL formalism

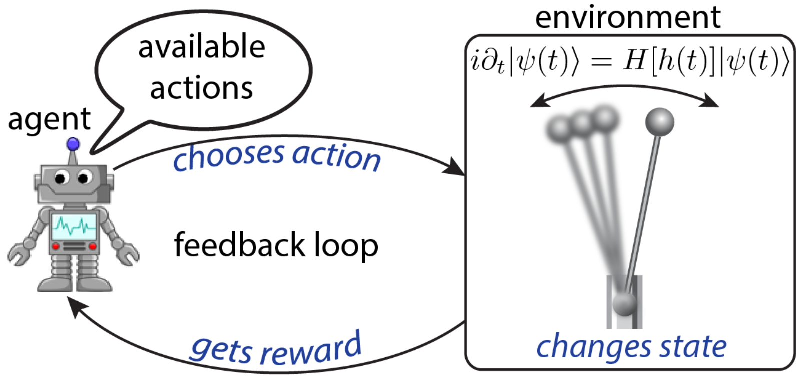
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

$$s = h_x(t) = [+4, +4, -4, +4, -4, -4, \dots]$$

$$\{|\psi(t)\rangle : |\psi(t)\rangle = U_h(t,0)|\psi_i\rangle\} \hat{=} \{h(t) : |\psi_i\rangle\}$$

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

Quantum Control as an RL Problem



→ RL formalism

- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

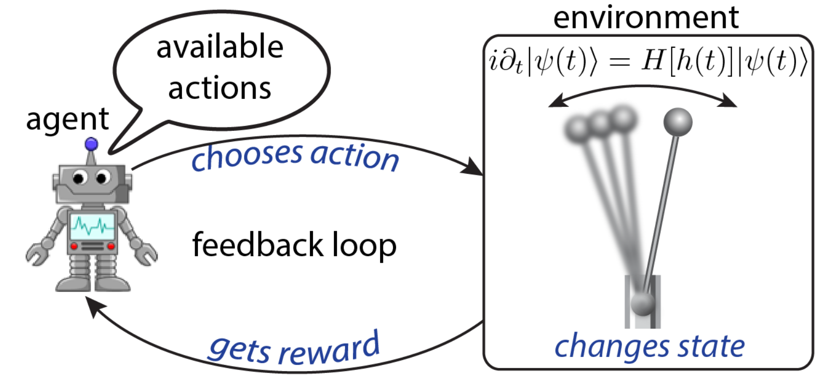
$$s = h_x(t) = [+4, +4, -4, +4, -4, -4, \ldots]$$

- reward space $\mathcal{R} = \{F_h(T) = |\langle\psi_*|U_h(T,0)|\psi_i\rangle|^2\}$

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

Berkeley UNIVERSITY OF CALIFORNIA

Quantum Control as an RL Problem



→ RL formalism

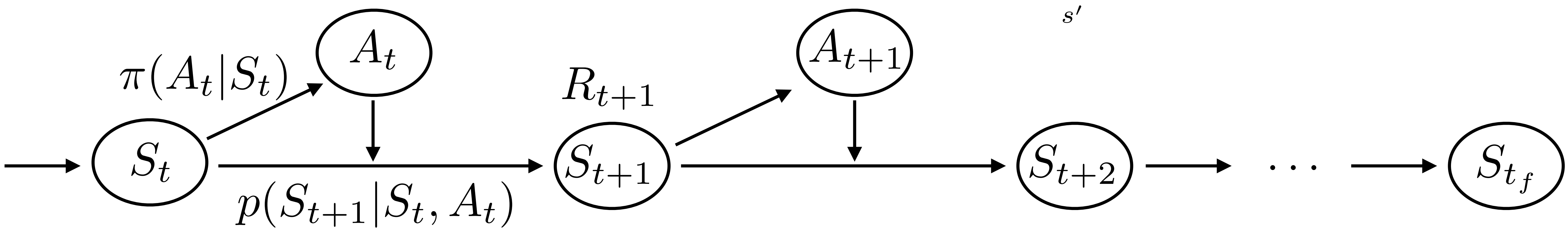
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

$$s = h_x(t) = [+4, +4, -4, +4, -4, -4, \dots]$$

- reward space $\mathcal{R} = \{F_h(T) = |\langle\psi_*|U_h(T,0)|\psi_i\rangle|^2\}$

→ RL as Markov decision process

$$R_{t+1} = \sum_{s'} p(s'|S_t, A_t)r(s', S_t, A_t)$$

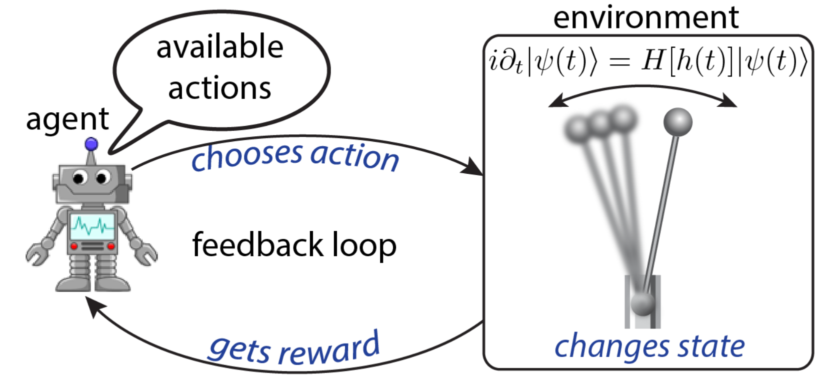


Marin Bukov

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

Berkeley UNIVERSITY OF CALIFORNIA

Quantum Control as an RL Problem



→ RL formalism

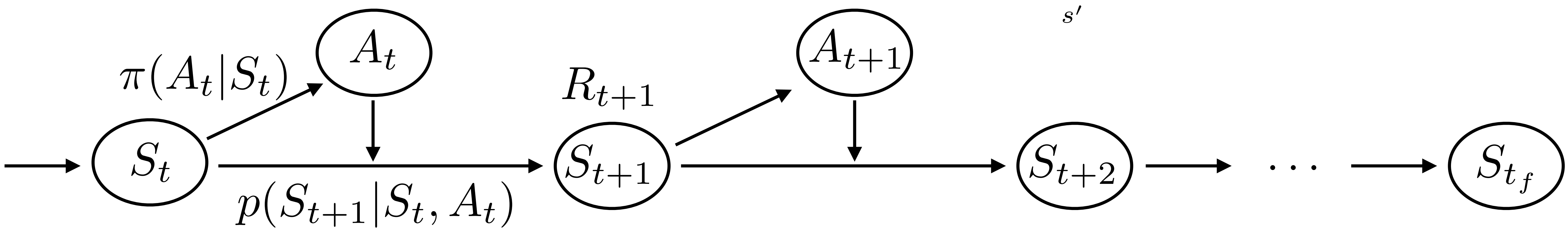
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

$$s = h_x(t) = [+4, +4, -4, +4, -4, -4, \dots]$$

- reward space $\mathcal{R} = \{F_h(T) = |\langle\psi_*|U_h(T,0)|\psi_i\rangle|^2\}$

→ RL as Markov decision process

$$R_{t+1} = \sum_{s'} p(s'|S_t, A_t) r(s', S_t, A_t)$$

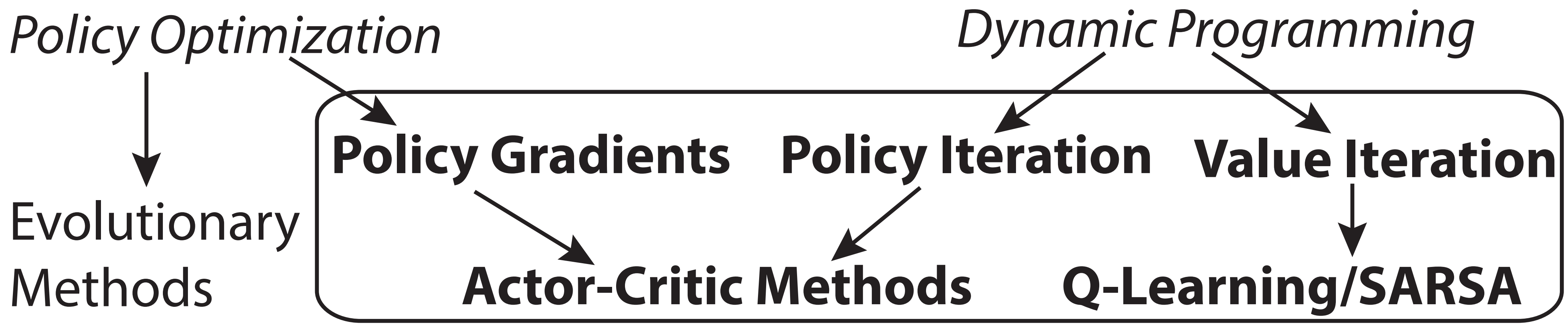


→ RL **objective**: maximize total *expected return* from step t onwards

$$Q(s,a) = \mathbb{E}_{a\sim\pi(a|s)}[R_{t+1} + \cdots + R_{t_f}|S_t = s, A_t = a]$$

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

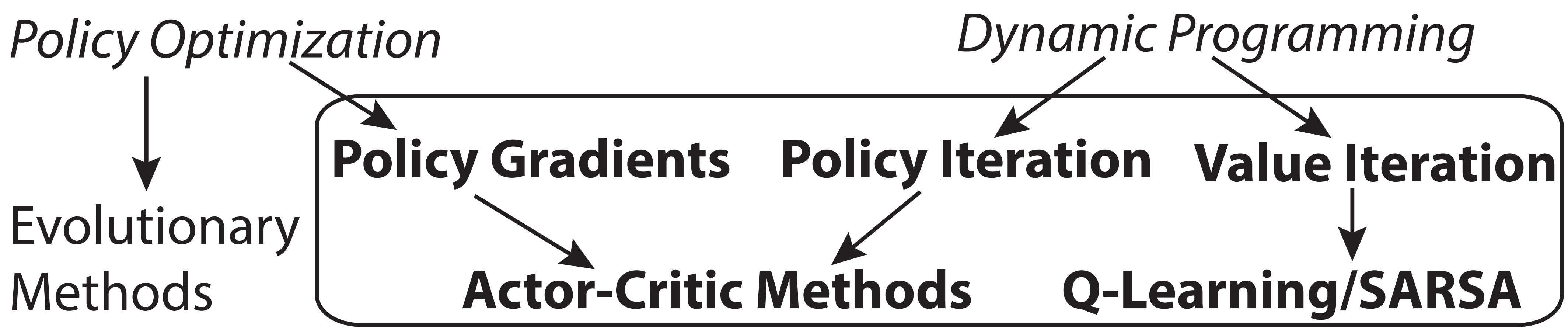
Overview of RL Algorithms



Marin Bukov

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

Overview of RL Algorithms



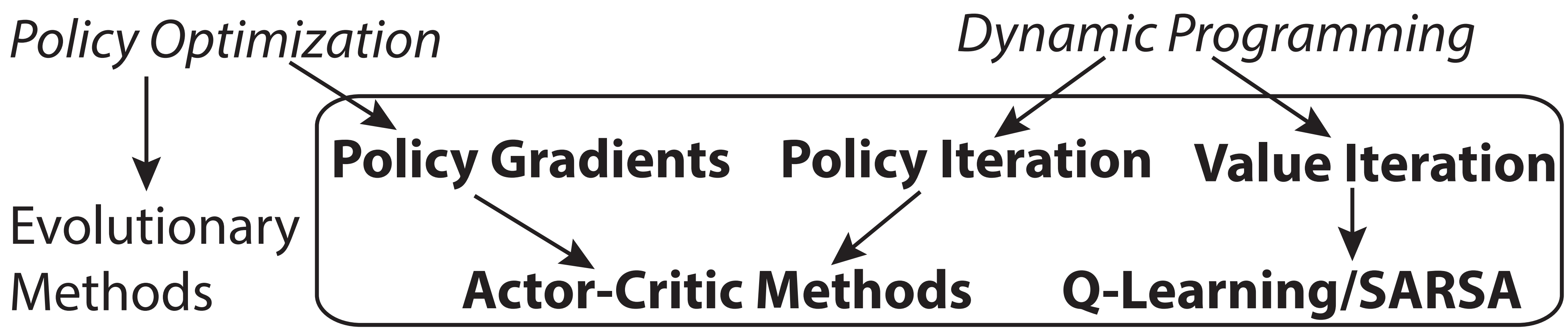
→ Value Iteration methods

- value function: ***expected*** total return under the policy $\pi(a|s)$ from state s

$$v_\pi(s) = \mathbb{E}_{a\sim\pi(a|s)}[G_t|S_t = s]$$

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

Overview of RL Algorithms



→ Value Iteration methods

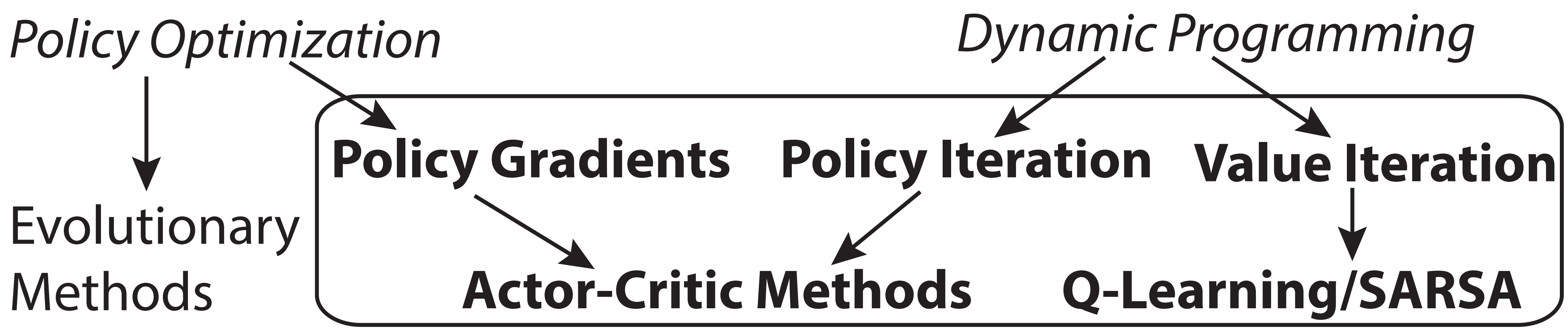
- value function: ***expected*** total return under the policy $\pi(a|s)$ from state s

$$v_\pi(s) = \mathbb{E}_{a\sim\pi(a|s)}[G_t|S_t = s]$$

- action-value (or Q-) function: ***expected*** total return under the policy $\pi(a|s)$ starting from state s and taking action a:

$$Q_\pi(s,a) = \mathbb{E}_{a\sim\pi(a|s)}[G_t|S_t = s, A_t = a]$$

Overview of RL Algorithms



→ Value Iteration methods

- value function: ***expected*** total return under the policy $\pi(a|s)$ from state s

$$v_\pi(s) = \mathbb{E}_{a\sim\pi(a|s)}[G_t|S_t = s]$$

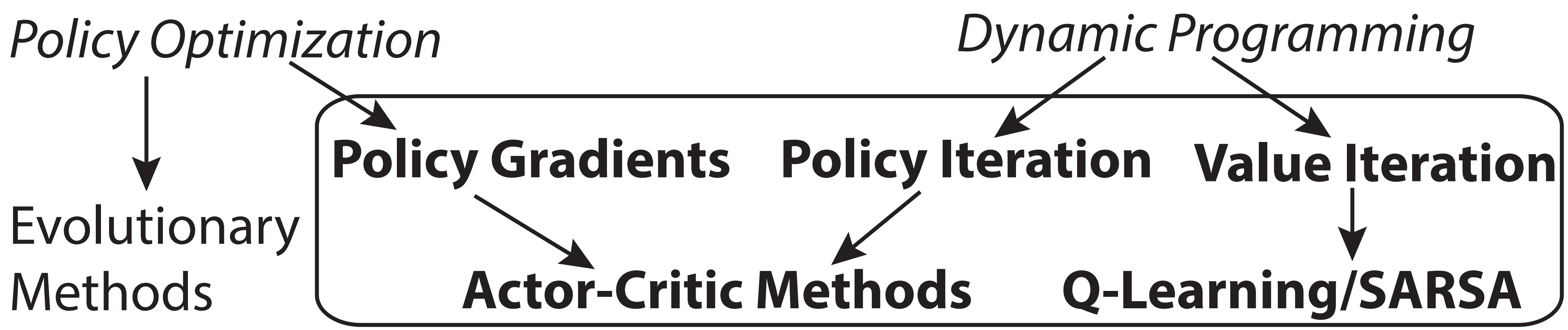
- action-value (or Q-) function: ***expected*** total return under the policy $\pi(a|s)$ starting from state s and taking action a:

$$Q_\pi(s,a) = \mathbb{E}_{a\sim\pi(a|s)}[G_t|S_t = s, A_t = a]$$

→ optimal action-value function: $Q_*(s,a) = \max_\pi Q_\pi(s,a)$

$$\pi_*(a|s) = \mathrm{argmax}_a Q_*(s,a)$$

Overview of RL Algorithms



→ Value Iteration methods

- value function: ***expected*** total return under the policy $\pi(a|s)$ from state s

$$v_\pi(s) = \mathbb{E}_{a\sim\pi(a|s)}[G_t|S_t = s]$$

- action-value (or Q-) function: ***expected*** total return under the policy $\pi(a|s)$ starting from state s and taking action a:

$$G_t = R_{t+1} + G_{t+1}$$

$$Q_\pi(s,a) = \mathbb{E}_{a\sim\pi(a|s)}[G_t|S_t = s, A_t = a]$$

→ optimal action-value function: $Q_*(s,a) = \max_\pi Q_\pi(s,a)$

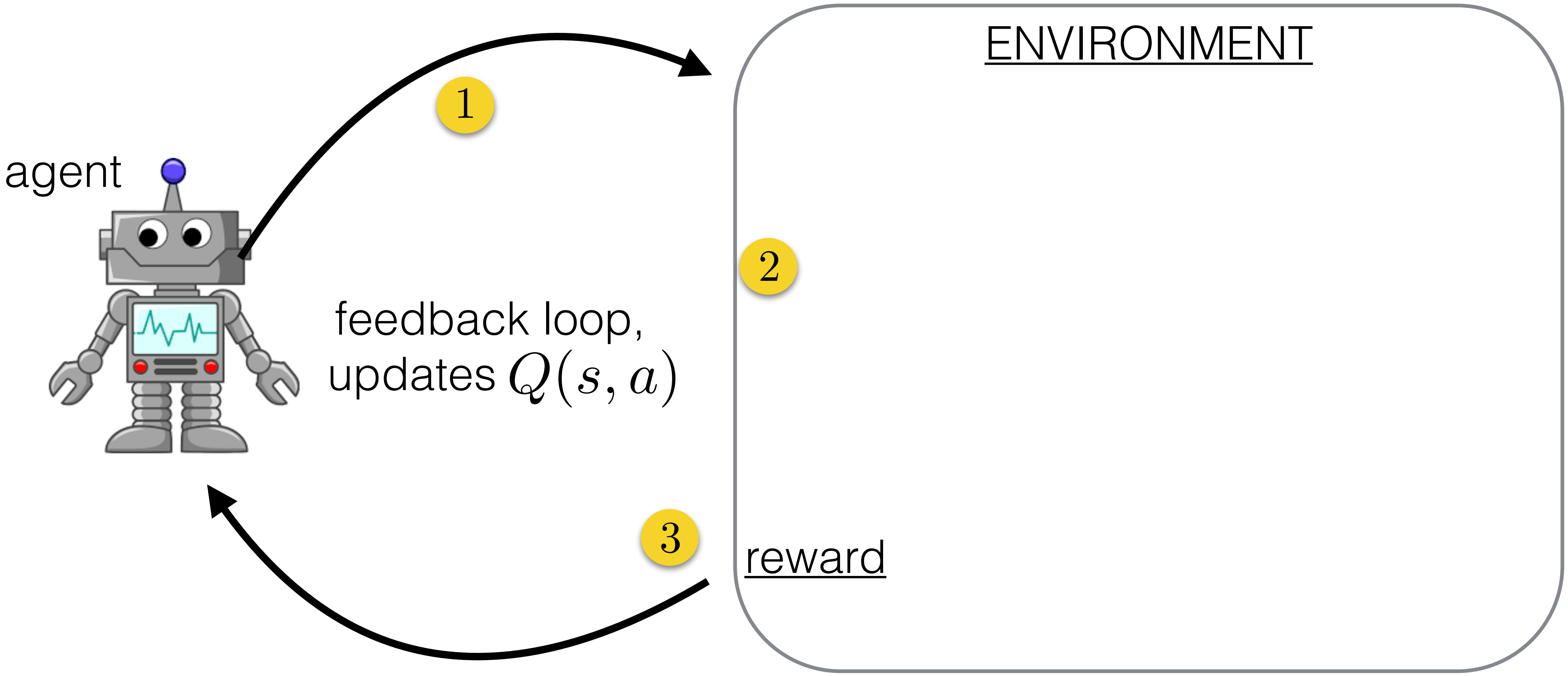
$$\pi_*(a|s) = \mathrm{argmax}_a Q_*(s,a)$$

Bellman's equation: $Q_*(s,a) = \sum_{s'} p(s'|s,a)\left[r(s,s',a) + \max_{a'} Q_*(s',a')\right]$

Marin Bukov

Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

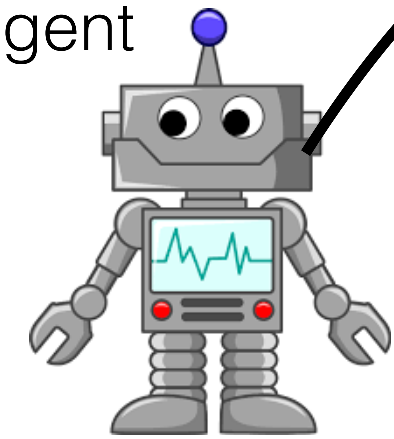
RL Applied to Quantum State Preparation



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

feedback loop, updates $Q(s,a)$

ENVIRONMENT

2

$H(t) = H_0 + H_{\mathrm{ctrl}}(t)$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

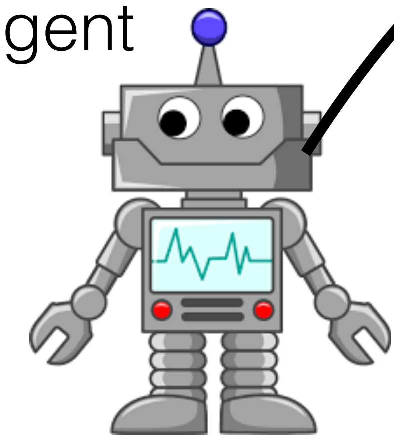
$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, t_f]$

3

reward

RL Applied to Quantum State Preparation

agent



1

feedback loop, updates $Q(s,a)$

ENVIRONMENT

2

$$H(t) = H_0 + H_{\rm ctrl}(t)$$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

$$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t\in[0,t_f]$$

3

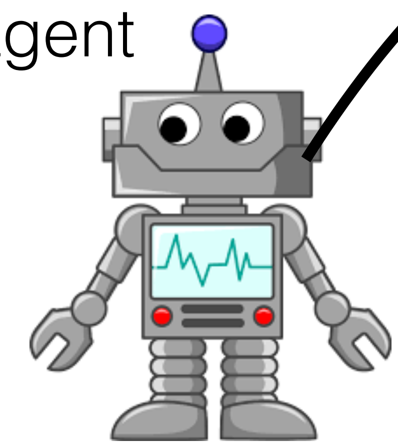
reward $r = \begin{cases} 0 & , 0 \le t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

Marin Bukov

Berkeley UNIVERSITY OF CALIFORNIA

RL Applied to Quantum State Preparation

agent



1

feedback loop, updates $Q(s,a)$

ENVIRONMENT

2

$H(t) = H_0 + H_{\mathrm{ctrl}}(t)$

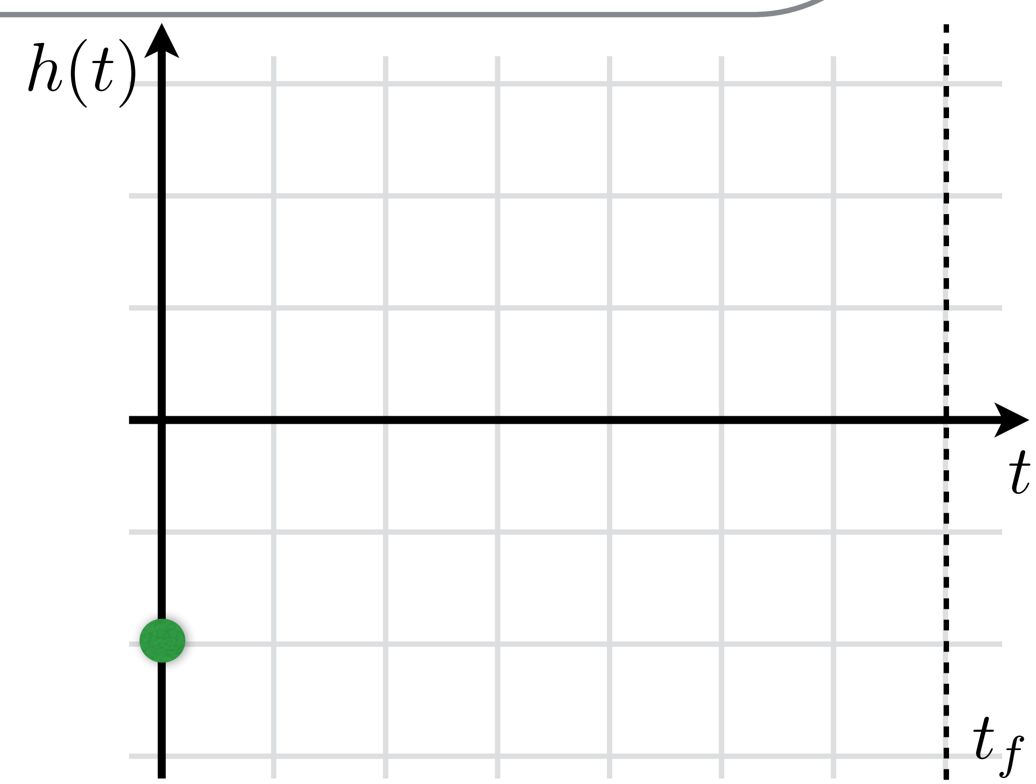
$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, t_f]$

3

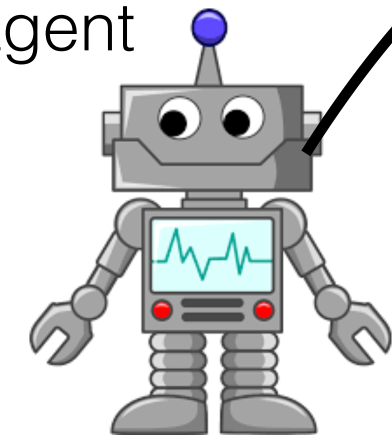
reward $r = \begin{cases} 0 & , 0 \leq t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

feedback loop, updates $Q(s,a)$

3

ENVIRONMENT

2

$$H(t) = H_0 + H_{\rm ctrl}(t)$$

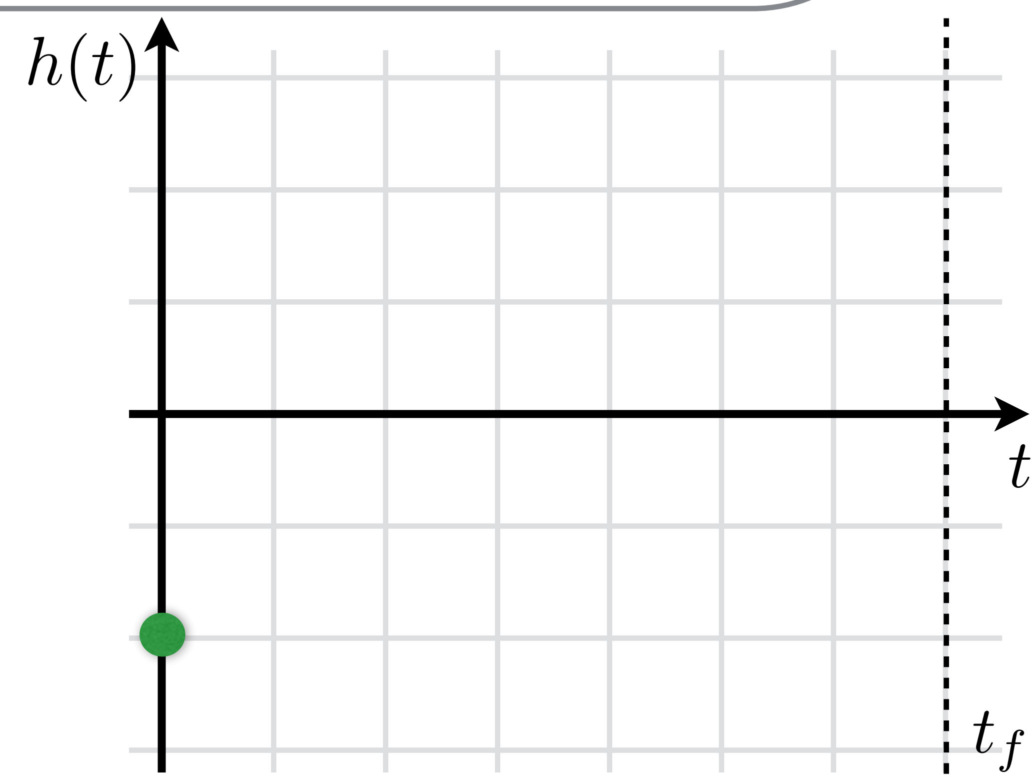
$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

$$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, t_f]$$

reward $r = \begin{cases} 0 & , 0 \le t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

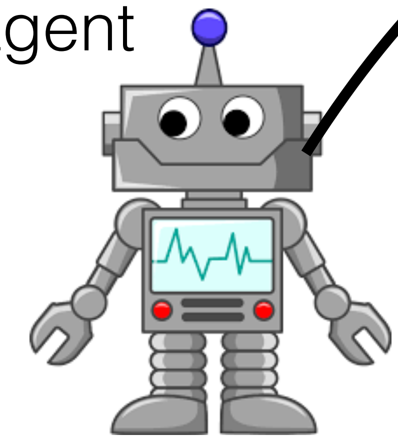
1 start from state $s_0 = [h(0)] = [-4]$



Marin Bukov

RL Applied to Quantum State Preparation

agent



(1)

feedback loop, updates $Q(s,a)$

(2)

ENVIRONMENT

$H(t) = H_0 + H_{\text{ctrl}}(t)$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, t_f]$

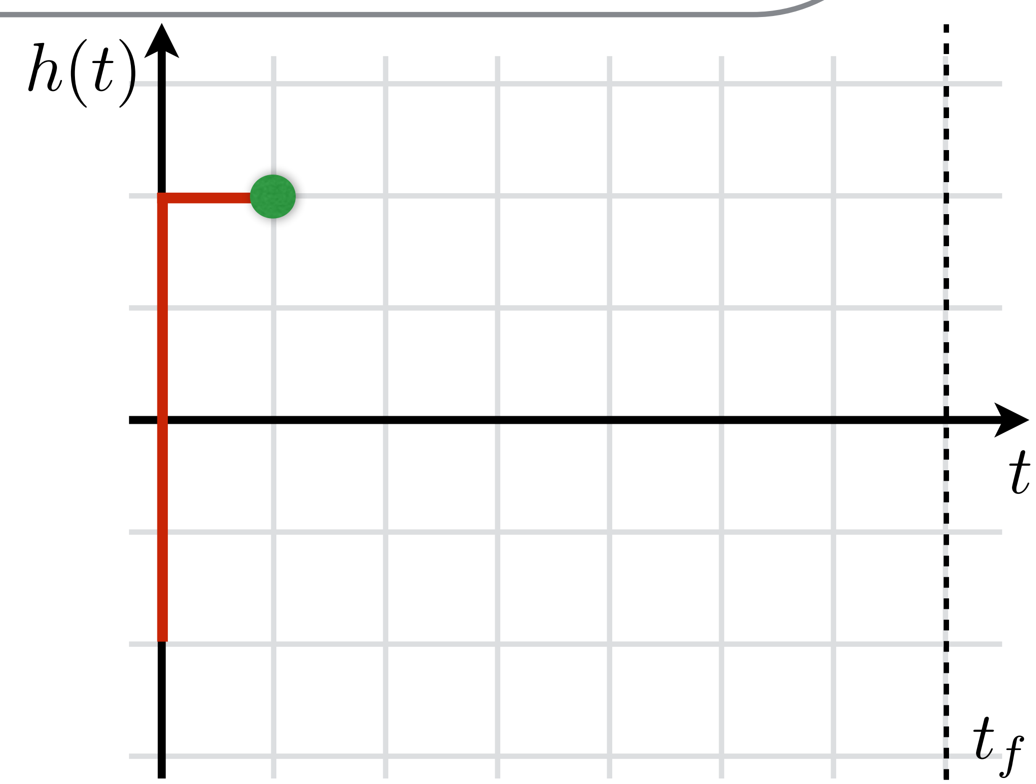
(3)

reward $r = \begin{cases} 0 & , 0 \le t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

(1) start from state $s_0 = [h(0)] = [-4]$

take action $a_0 : \delta h = +4$

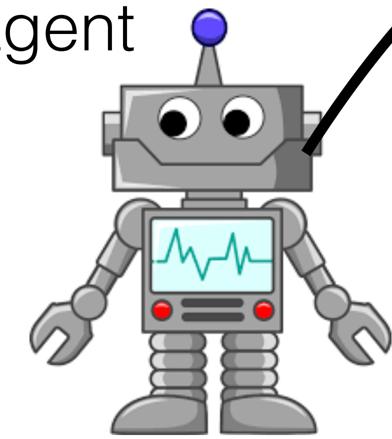
go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$



Berkeley
UNIVERSITY OF CALIFORNIA

RL Applied to Quantum State Preparation

agent



feedback loop, updates $Q(s,a)$

1

ENVIRONMENT

$H(t) = H_0 + H_{\text{ctrl}}(t)$

$|\psi_i\rangle$: GS of H_0

2

$|\psi_*\rangle$: target state

$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, t_f]$

3

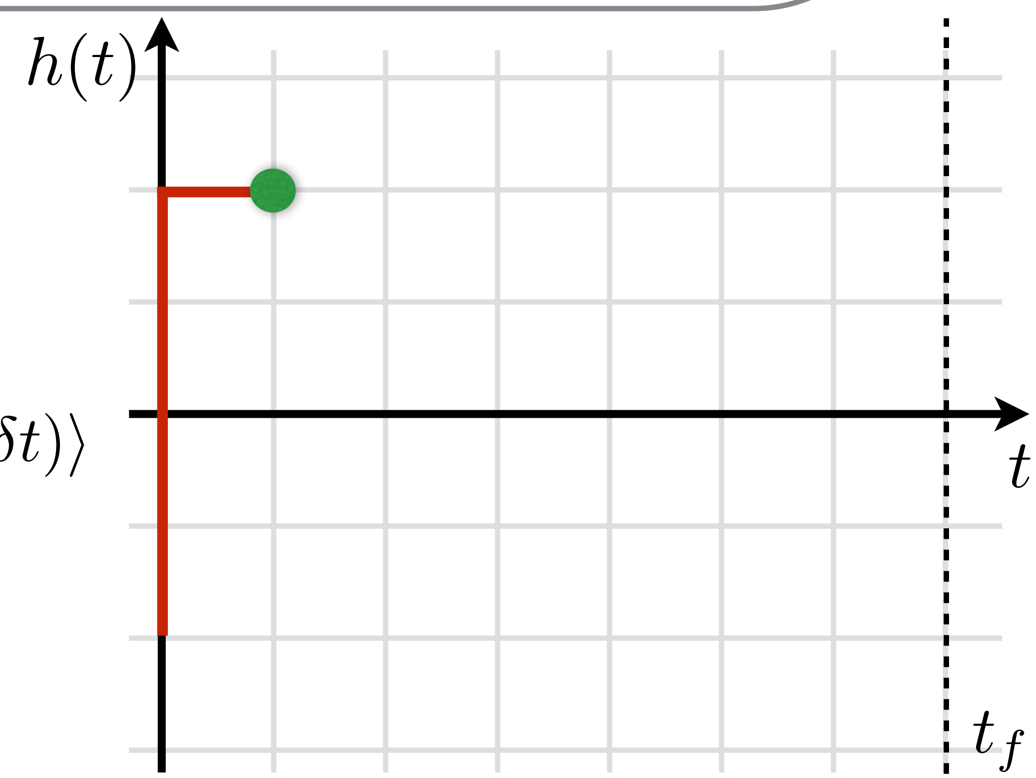
reward $r = \begin{cases} 0 & , 0 \leq t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

1 start from state $s_0 = [h(0)] = [-4]$

take action $a_0 : \delta h = +4$

go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$

2 solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$

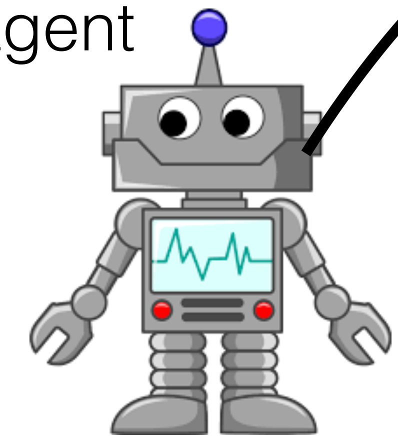


Marin Bukov

RL Applied to Quantum State Preparation

Berkeley UNIVERSITY OF CALIFORNIA

agent



1

feedback loop,
updates $Q(s,a)$

ENVIRONMENT

$H(t) = H_0 + H_{\mathrm{ctrl}}(t)$

$|\psi_i\rangle$: GS of H_0

2

$|\psi_*\rangle$: target state

$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, t_f]$

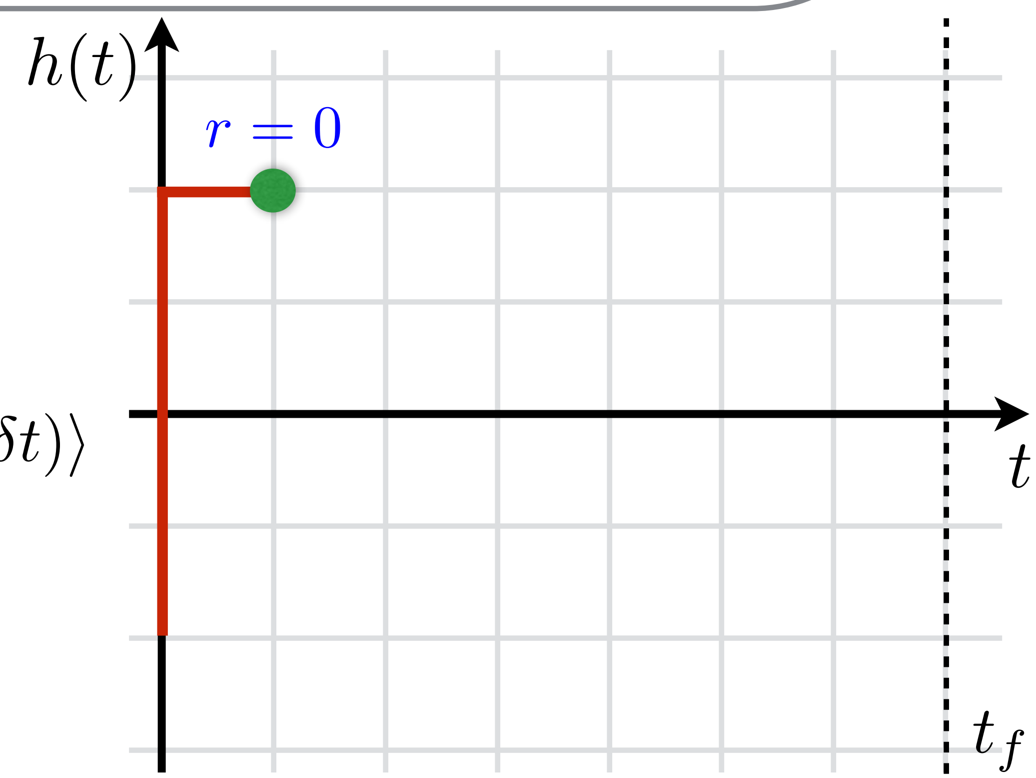
3

reward $r = \begin{cases} 0 & , 0 \le t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

1. start from state $s_0 = [h(0)] = [-4]$
   take action $a_0 : \delta h = +4$
   go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$

2. solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$

3. calculate reward r
   and use it to update $Q(s,a)$
   which in turn is used to choose subsequent actions

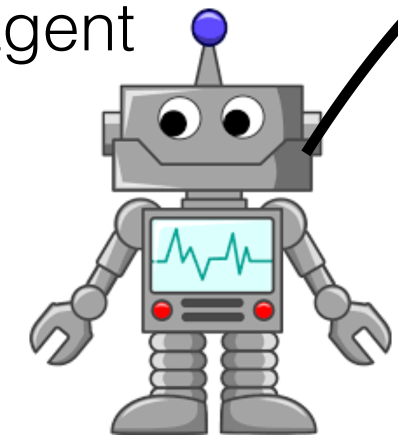


Marin Bukov

Berkeley UNIVERSITY OF CALIFORNIA

RL Applied to Quantum State Preparation

agent



1

feedback loop, updates $Q(s,a)$

ENVIRONMENT

2

$H(t) = H_0 + H_{\mathrm{ctrl}}(t)$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, t_f]$

3

reward $r = \begin{cases} 0 & , 0 \le t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

1 start from state $s_0 = [h(0)] = [-4]$

take action $a_0 : \delta h = +4$

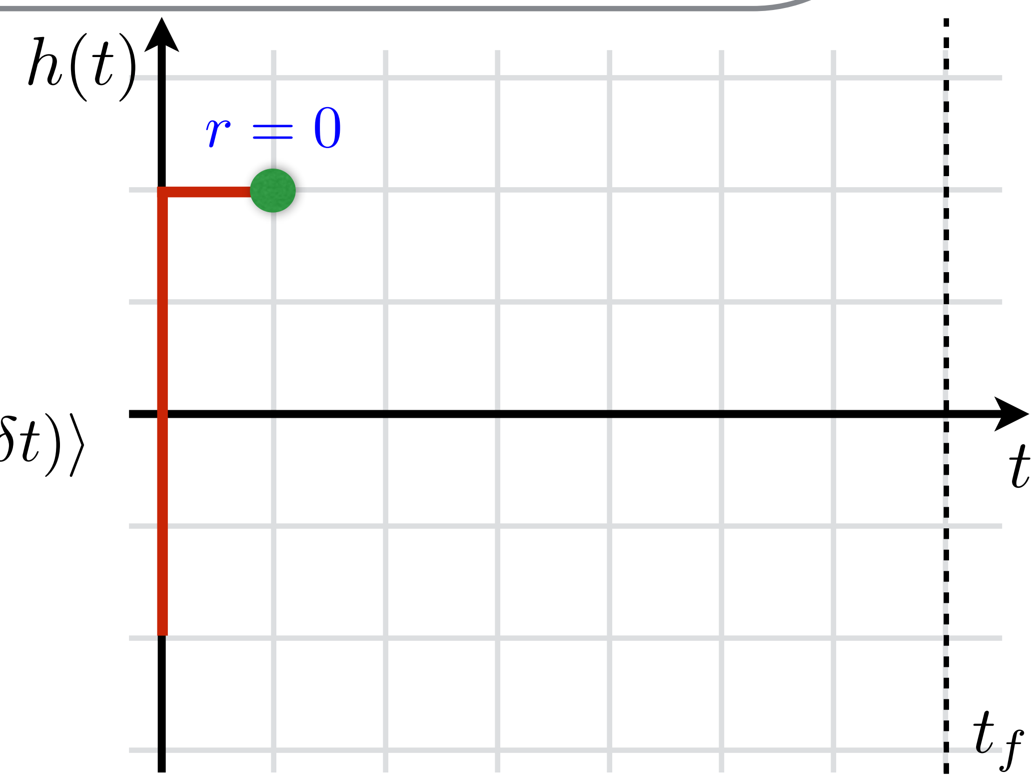
go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$

2 solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$

3 calculate reward r

and use it to update $Q(s,a)$

which in turn is used to choose subsequent actions

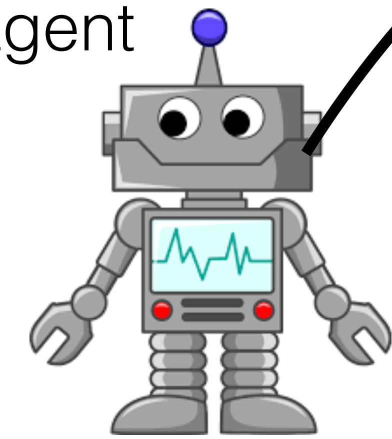


Marin Bukov

Berkeley UNIVERSITY OF CALIFORNIA

RL Applied to Quantum State Preparation

agent



feedback loop, updates $Q(s,a)$

ENVIRONMENT

$$H(t) = H_0 + H_{\mathrm{ctrl}}(t)$$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

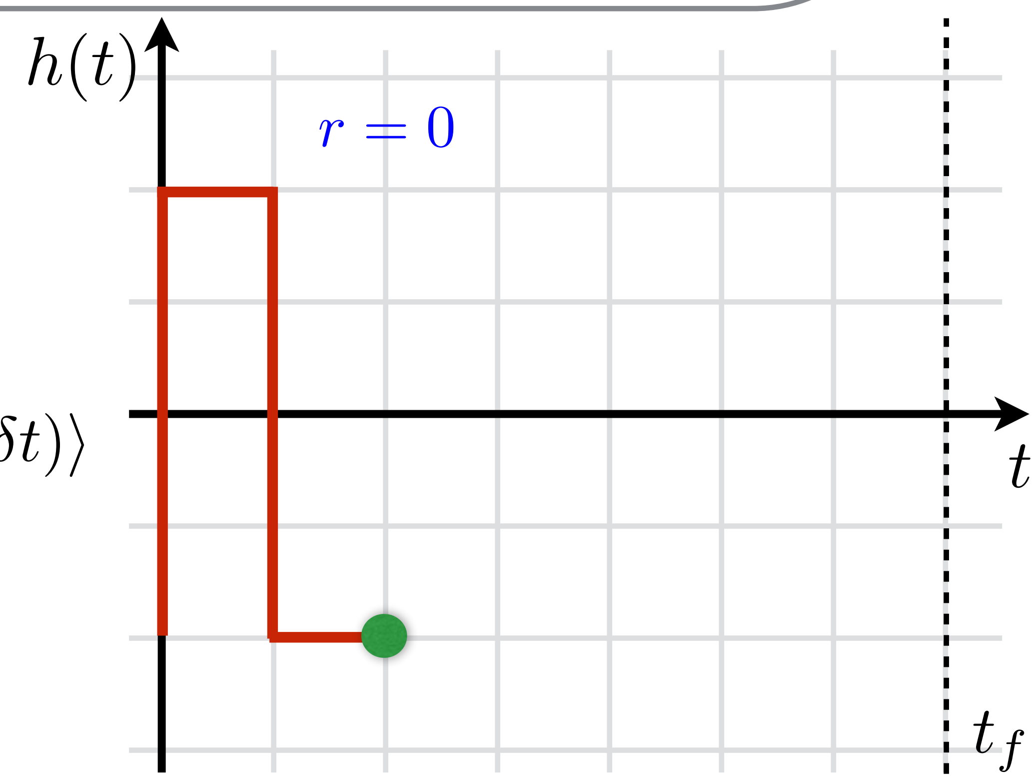
$$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t\in[0,t_f]$$

reward $r = \begin{cases} 0 & , 0\le t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

1. start from state $s_0 = [h(0)] = [-4]$
   take action $a_0 : \delta h = +4$
   go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$

2. solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$

3. calculate reward r
   and use it to update $Q(s,a)$
   which in turn is used to choose subsequent actions

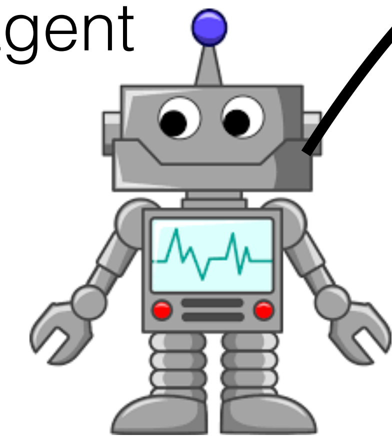


Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

RL Applied to Quantum State Preparation

agent



1

feedback loop, updates $Q(s,a)$

3

ENVIRONMENT

2

$$H(t) = H_0 + H_{\mathrm{ctrl}}(t)$$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

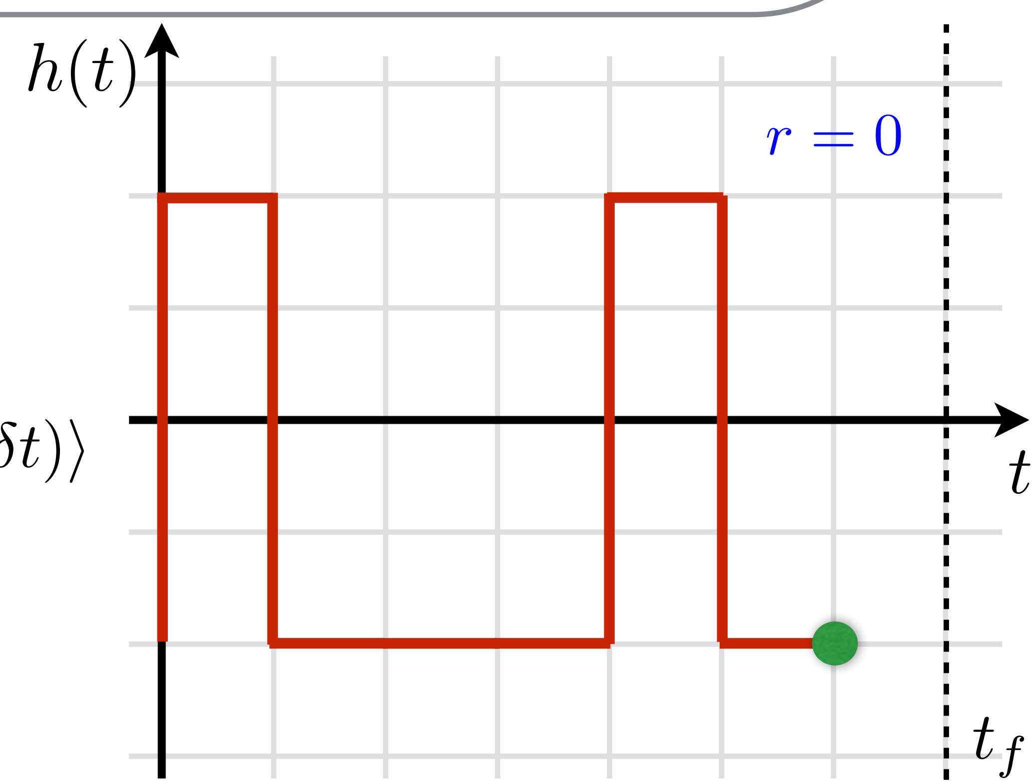
$$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \quad t \in [0, t_f]$$

reward $r = \begin{cases} 0 & , 0 \le t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

1. start from state $s_0 = [h(0)] = [-4]$
   take action $a_0 : \delta h = +4$
   go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$

2. solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$

3. calculate reward r
   and use it to update $Q(s,a)$
   which in turn is used to choose subsequent actions

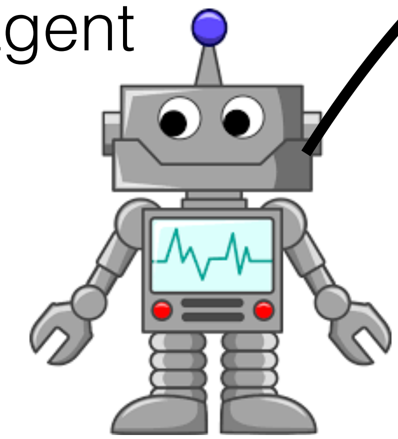


Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

RL Applied to Quantum State Preparation

agent



1

feedback loop, updates $Q(s,a)$

3

ENVIRONMENT

2

$$H(t) = H_0 + H_{\text{ctrl}}(t)$$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

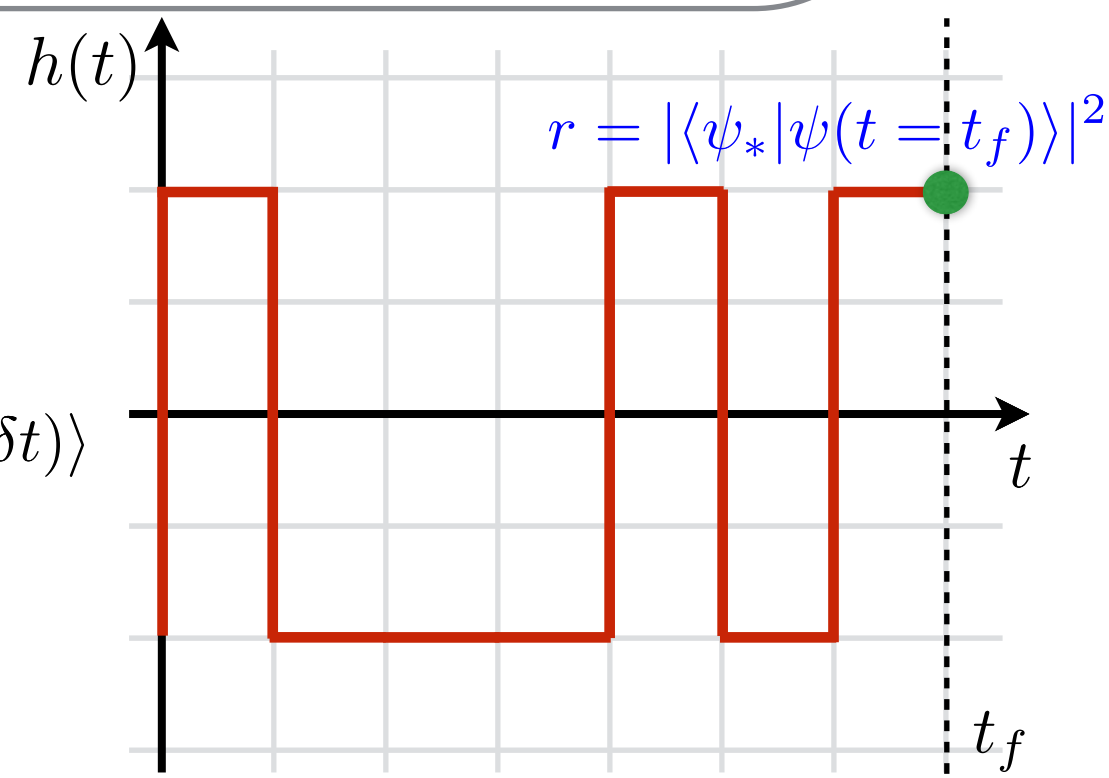
$$i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle \; t\in[0,t_f]$$

reward $r = \begin{cases} 0 & ,0\leq t < t_f \\ |\langle\psi_*|\psi(t=t_f)\rangle|^2 & , t = t_f \end{cases}$

1. start from state $s_0 = [h(0)] = [-4]$
take action $a_0 : \delta h = +4$
go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$

2. solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$

3. calculate reward r
and use it to update $Q(s,a)$
which in turn is used to choose subsequent actions



episode completed

Marin Bukov

RL with Function Approximation

→ problem: state space has exponentially many configurations $|\mathcal{A}|^{N_T}$

- can we estimate values of not yet encountered states?

RL with Function Approximation

- problem: state space has exponentially many configurations $|\mathcal{A}|^{N_T}$
  - can we estimate values of not yet encountered states?
- YES, via extrapolation: parametrize the Q-function/policy

$$Q(s,a) \to Q_\theta(s,a) \qquad\qquad \pi(a|s) \to \pi_\theta(a|s)$$

Berkeley UNIVERSITY OF CALIFORNIA

RL with Function Approximation

→ problem: state space has exponentially many configurations $|\mathcal{A}|^{N_T}$

- can we estimate values of not yet encountered states?

→ YES, via extrapolation: parametrize the Q-function/policy

$$Q(s,a) \to Q_\theta(s,a) \qquad \pi(a|s) \to \pi_\theta(a|s)$$

- typical approach: use deep neural network **(Deep RL)**
- caveat: value-iteration RL algorithms have convergence guarantees only for linear function approximators
- lots of empirical tricks to combine Deep Learning and RL

RL with Function Approximation

→ problem: state space has exponentially many configurations $|\mathcal{A}|^{N_T}$

- can we estimate values of not yet encountered states?

→ YES, via extrapolation: parametrize the Q-function/policy

$$Q(s,a) \to Q_\theta(s,a) \qquad\qquad \pi(a|s) \to \pi_\theta(a|s)$$

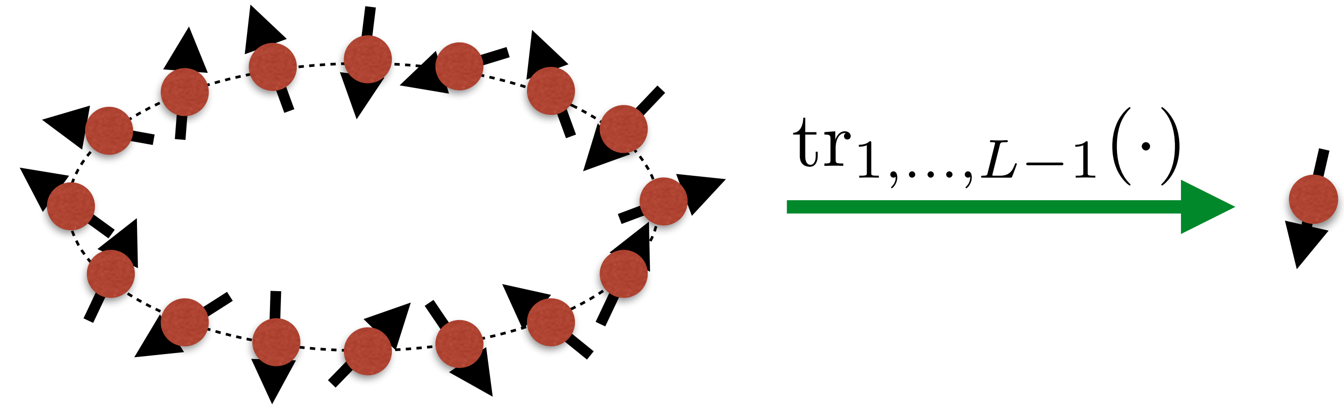
- typical approach: use deep neural network **(Deep RL)**
- caveat: value-iteration RL algorithms have convergence guarantees only for linear function approximators
- lots of empirical tricks to combine Deep Learning and RL

→ examples of Deep RL:

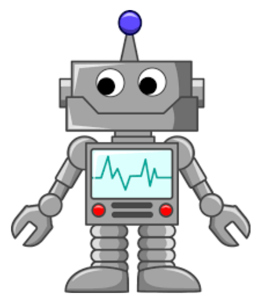
- Tesauro's Backgammon RL player (1992)
- DeepMind: Atari games, AlphaGo, etc.
- self-driving cars, autonomous drone/helicopter hovering, etc.

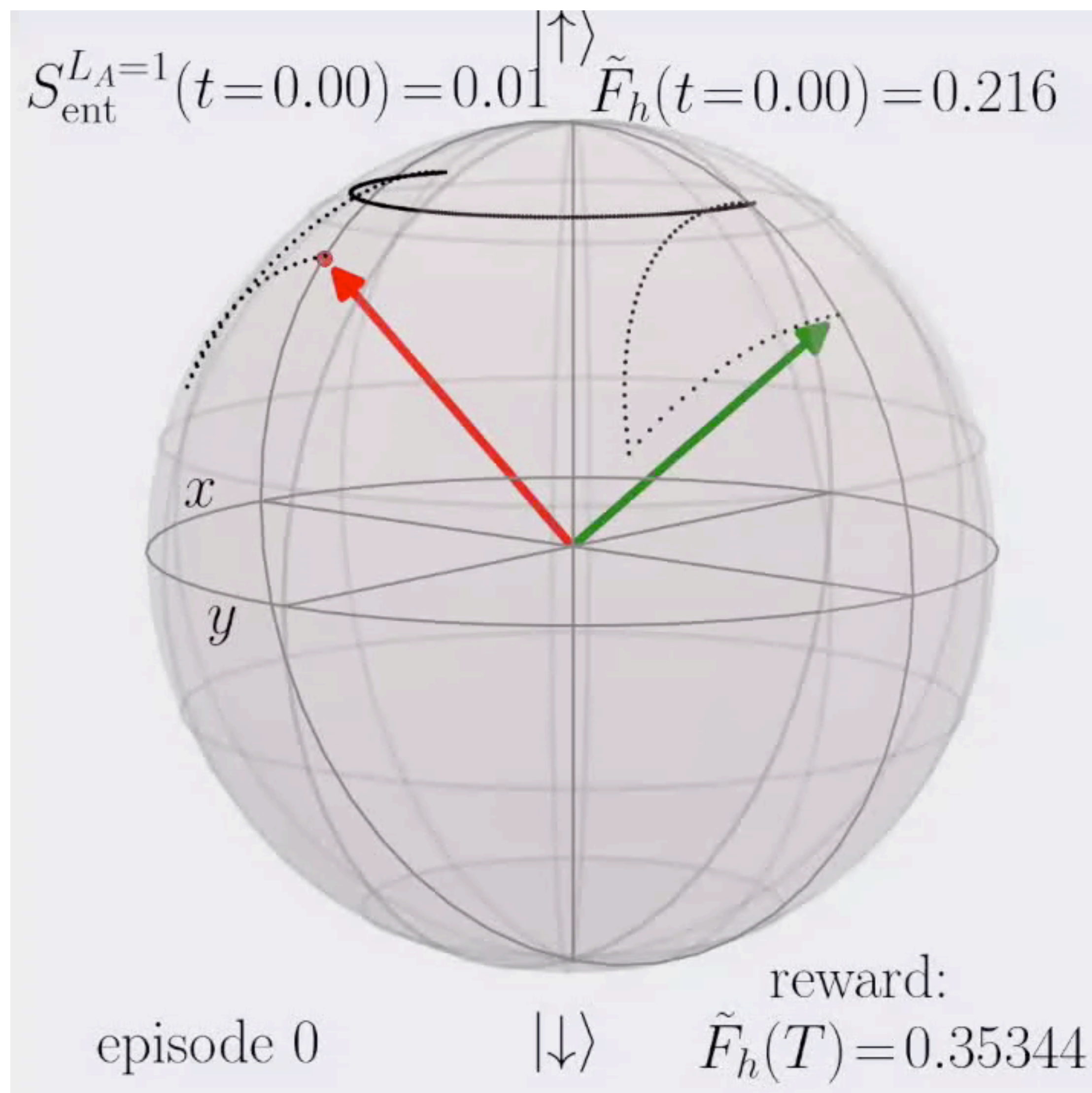
Learning Many-Body Quantum Control

$$H(t) = -\sum_{j=1}^{L} S^z_{j+1} S^z_j + S^z_j + h_x(t) S^x_j$$



$h_x \in \{\pm 4\}$ bang-bang protocols





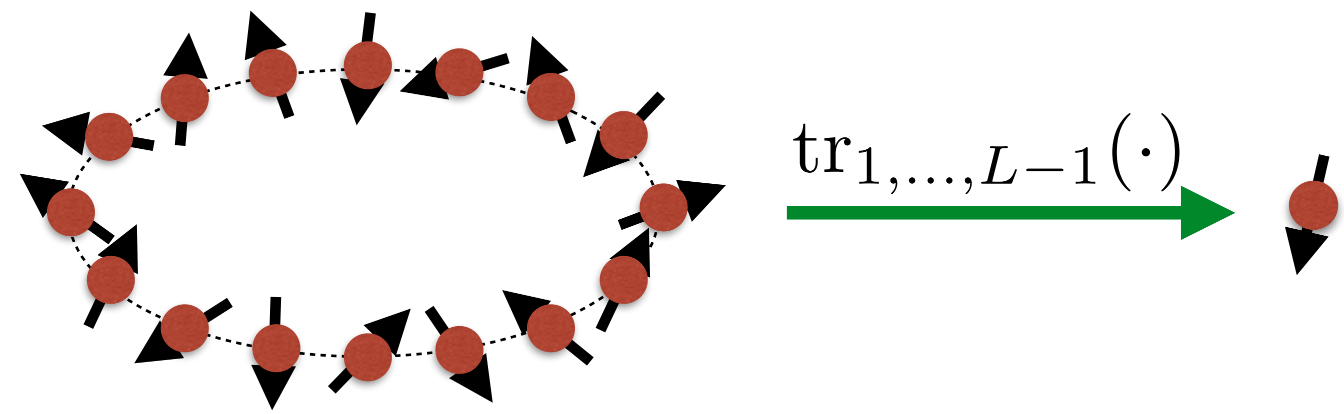
Bloch sphere

$$\tilde{F}_h = \frac{1}{L} \log F_h$$

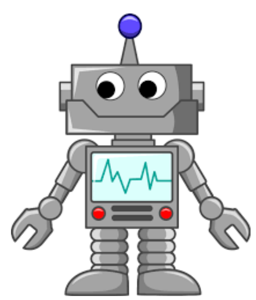
Berkeley UNIVERSITY OF CALIFORNIA

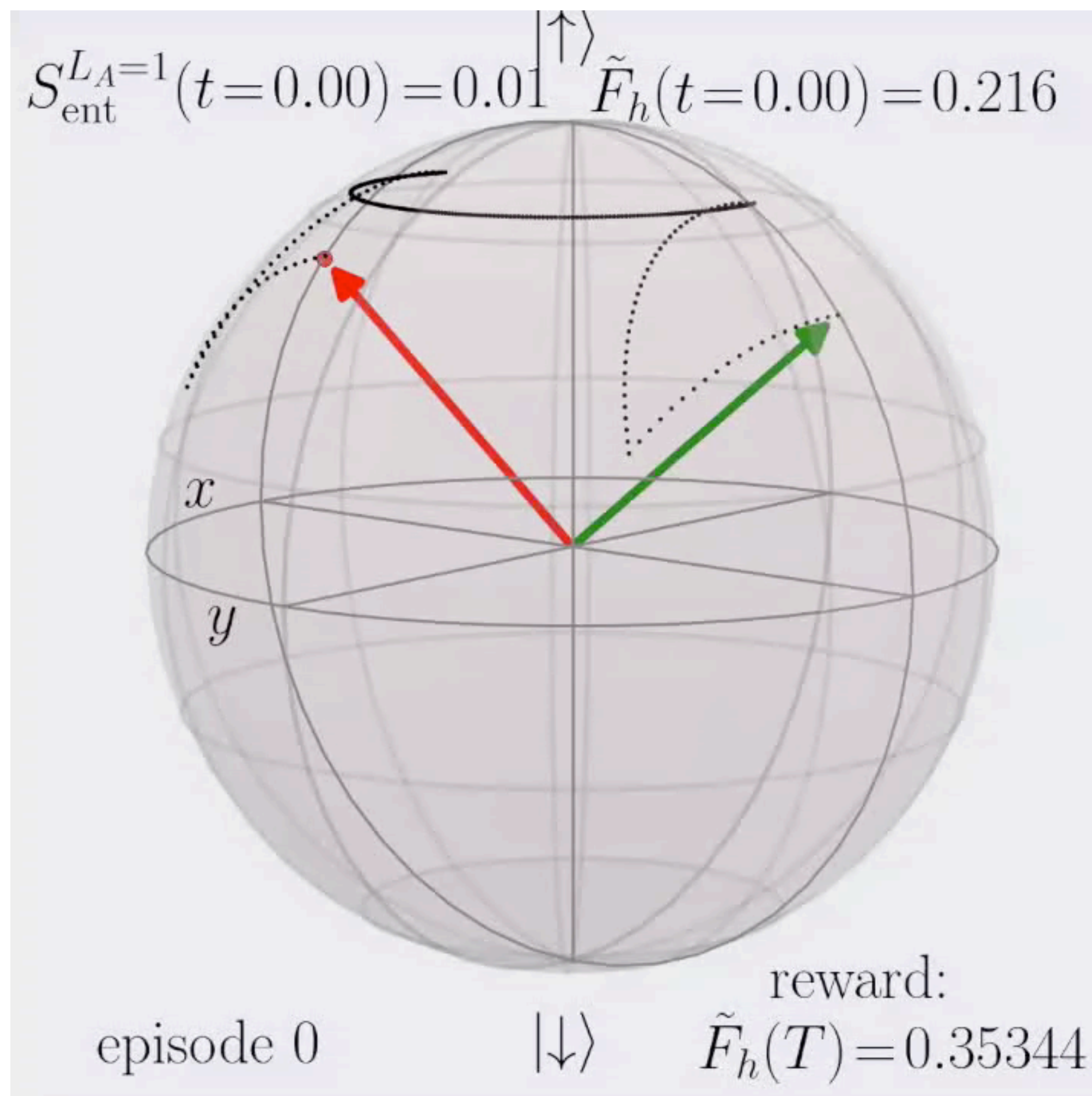
Learning Many-Body Quantum Control

$$H(t) = -\sum_{j=1}^{L} S^z_{j+1}S^z_j + S^z_j + h_x(t)S^x_j$$



$h_x \in \{\pm 4\}$ bang-bang protocols





Bloch sphere

$$\tilde{F}_h = \frac{1}{L}\log F_h$$

MB et al, PRX 8 031086 (2018)

Berkeley
UNIVERSITY OF CALIFORNIA

Why RL in Nonequilibrium Dynamics?

→ **model-free:** find effective control degrees of freedom (dof)

→ **adaptive:** train on one environment, use in a different environment

→ **autonomous:** does not require supervision

Why RL in Nonequilibrium Dynamics?

→ **model-free:** find effective control degrees of freedom (dof)

- microscopic descriptions have extensively many dof
- cannot solve equations of motion
- use (deep) RL to find guiding principles away from equilibrium?
- can RL handle uncertain environments and learn policies in the presence of various (correlated) sources of noise?

→ **adaptive:** train on one environment, use in a different environment

→ **autonomous:** does not require supervision

Marin Bukov

Berkeley UNIVERSITY OF CALIFORNIA

Why RL in Nonequilibrium Dynamics?

→ **model-free:** find effective control degrees of freedom (dof)

- microscopic descriptions have extensively many dof
- cannot solve equations of motion
- use (deep) RL to find guiding principles away from equilibrium?
- can RL handle uncertain environments and learn policies in the presence of various (correlated) sources of noise?

→ **adaptive:** train on one environment, use in a different environment

- Q-function/policy contain knowledge about the environment which can be used after training
- can RL reveal similarities between at first sight unrelated problems?

→ **autonomous:** does not require supervision

Berkeley UNIVERSITY OF CALIFORNIA

Why RL in Nonequilibrium Dynamics?

→ **model-free:** find effective control degrees of freedom (dof)

- microscopic descriptions have extensively many dof
- cannot solve equations of motion
- use (deep) RL to find guiding principles away from equilibrium?
- can RL handle uncertain environments and learn policies in the presence of various (correlated) sources of noise?

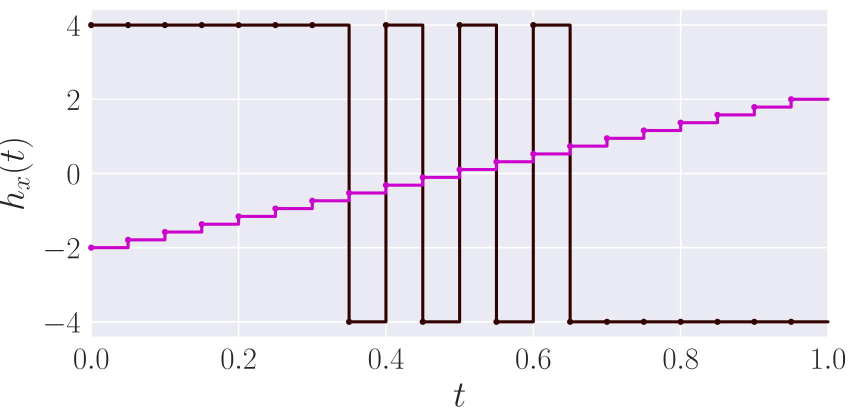
→ **adaptive:** train on one environment, use in a different environment

- Q-function/policy contain knowledge about the environment which can be used after training
- can RL reveal similarities between at first sight unrelated problems?

→ **autonomous:** does not require supervision

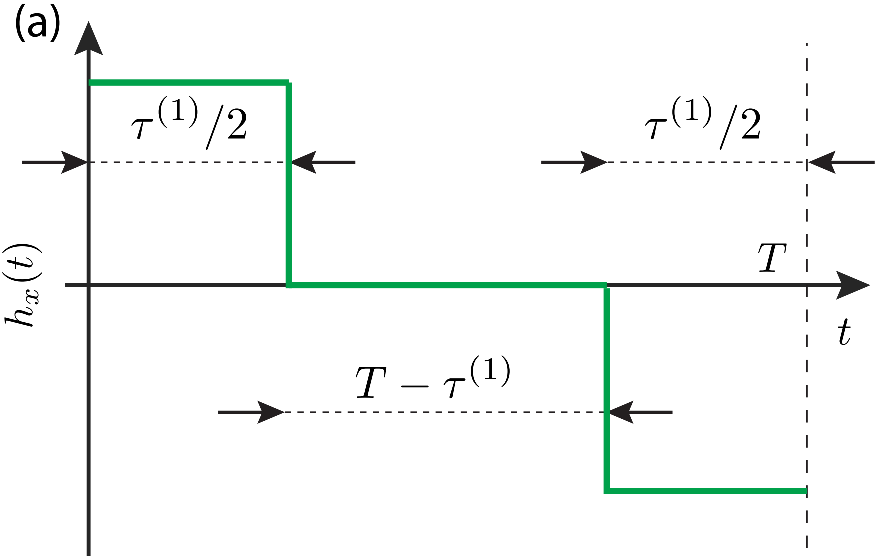
- can RL automate experimental setups?

What do we Learn from the RL Agent?

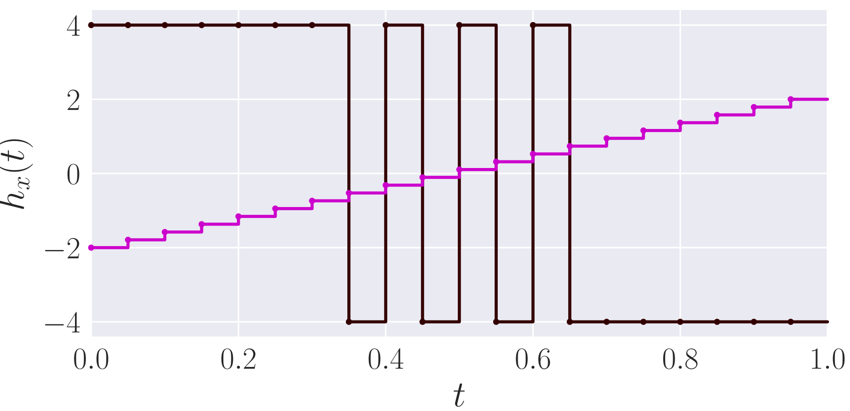


$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

$$-\mathcal{F}_h(T) = \min_{\tau^{(1)}\in[0,T]} \left(-\mathcal{F}_h(T;\tau^{(1)})\right)$$



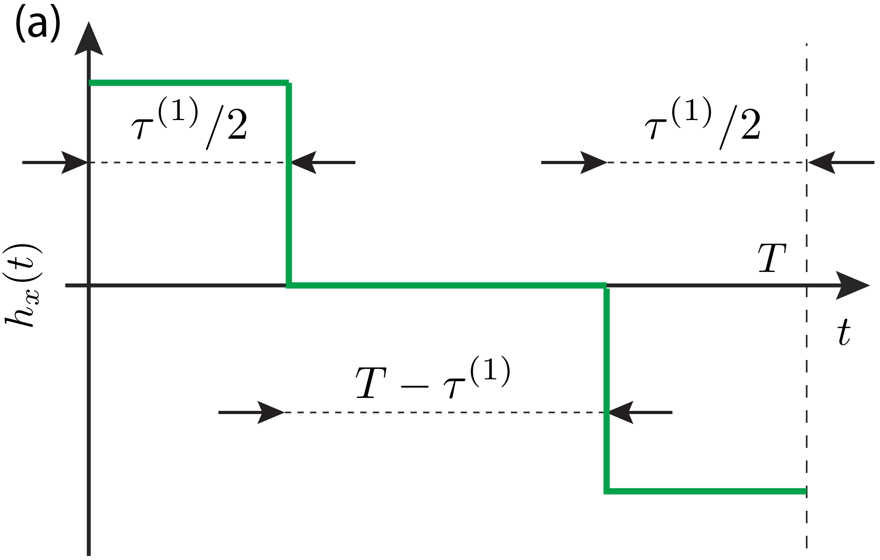
What do we Learn from the RL Agent?

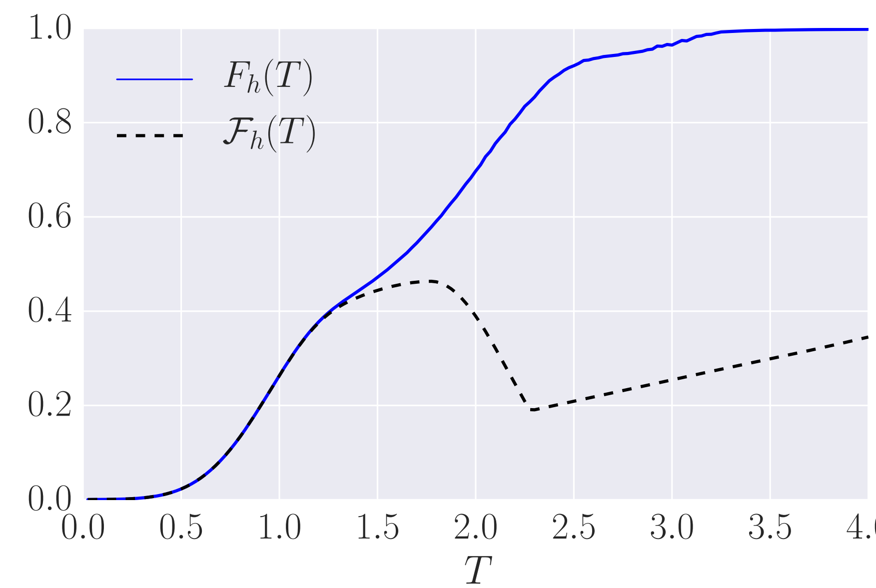


$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t)S^x_j$$

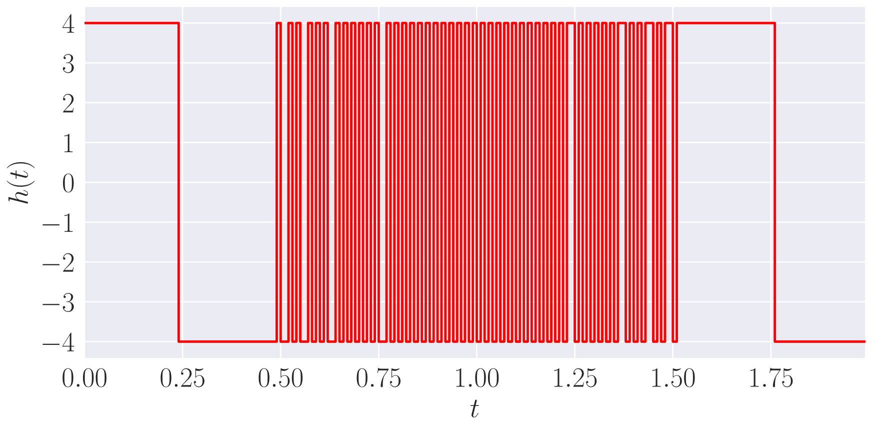
$$-\mathcal{F}_h(T) = \min_{\tau^{(1)}\in[0,T]} \left(-\mathcal{F}_h(T;\tau^{(1)})\right)$$

(a)



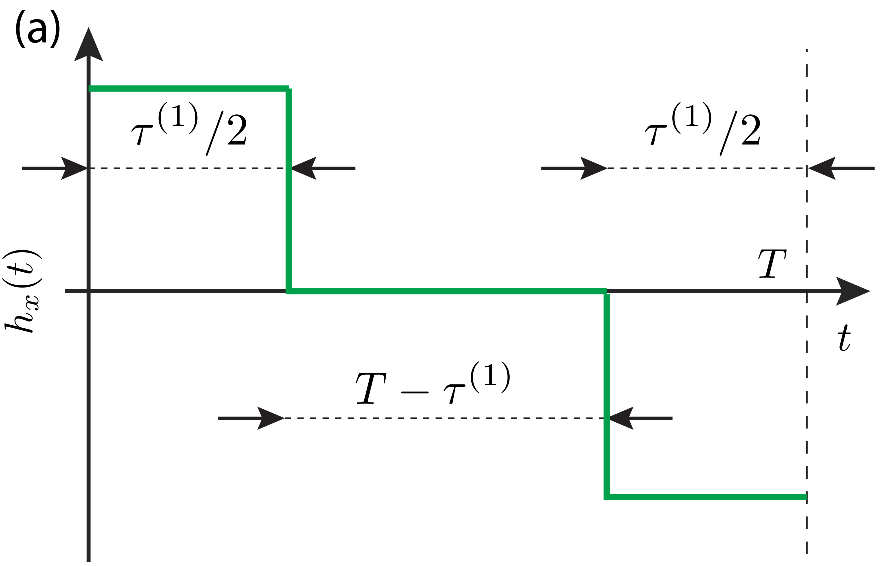


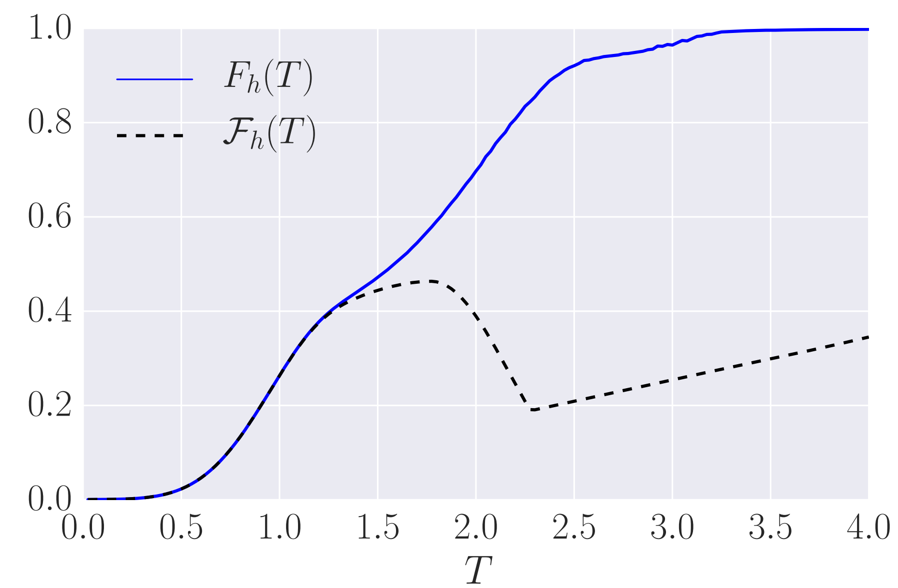
What do we Learn from the RL Agent?

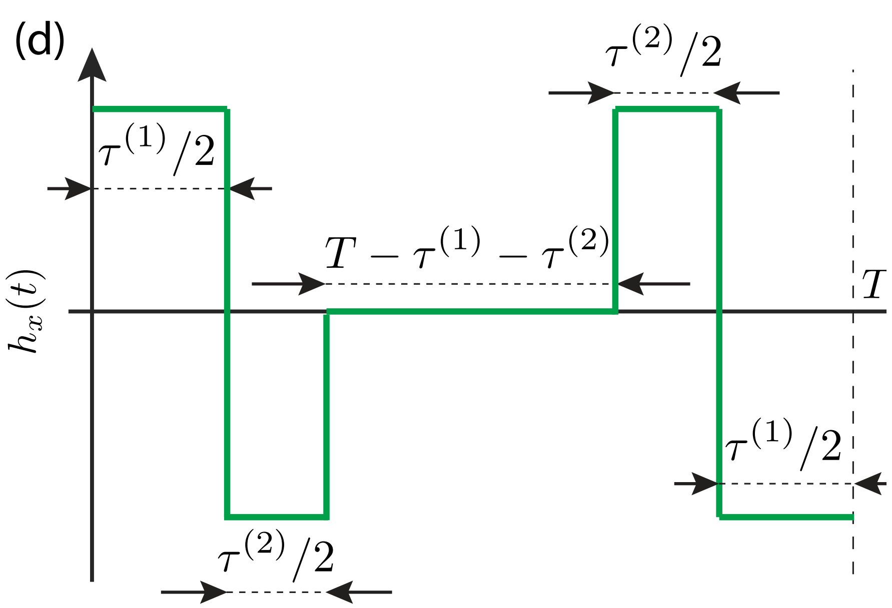


$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

$$-\mathcal{F}_h(T) = \min_{\tau^{(1)}\in[0,T]} \left(-\mathcal{F}_h(T;\tau^{(1)})\right)$$

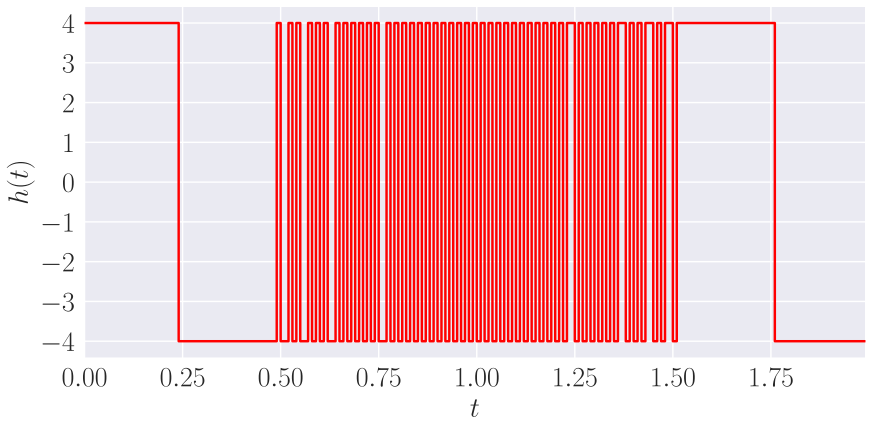






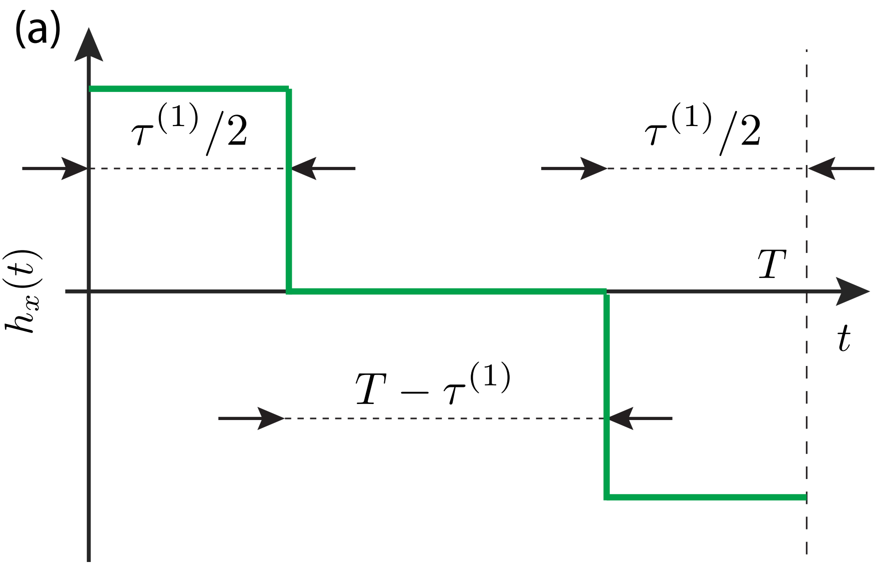
MB et al, PRX 8 031086 (2018)

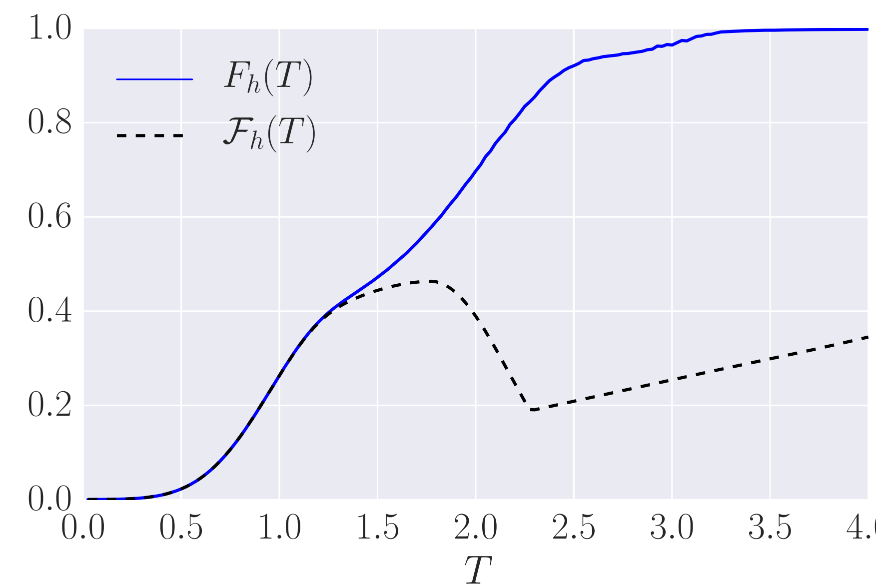
What do we Learn from the RL Agent?

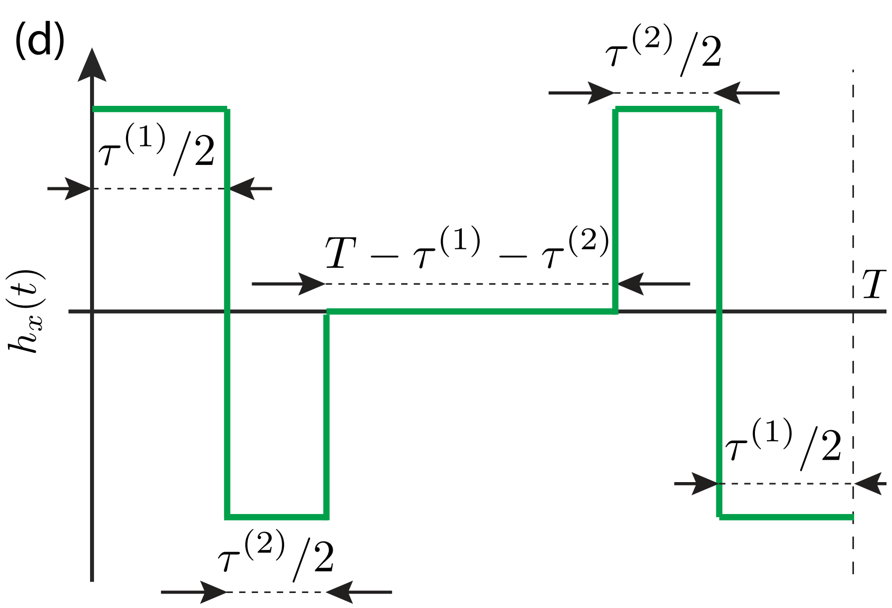


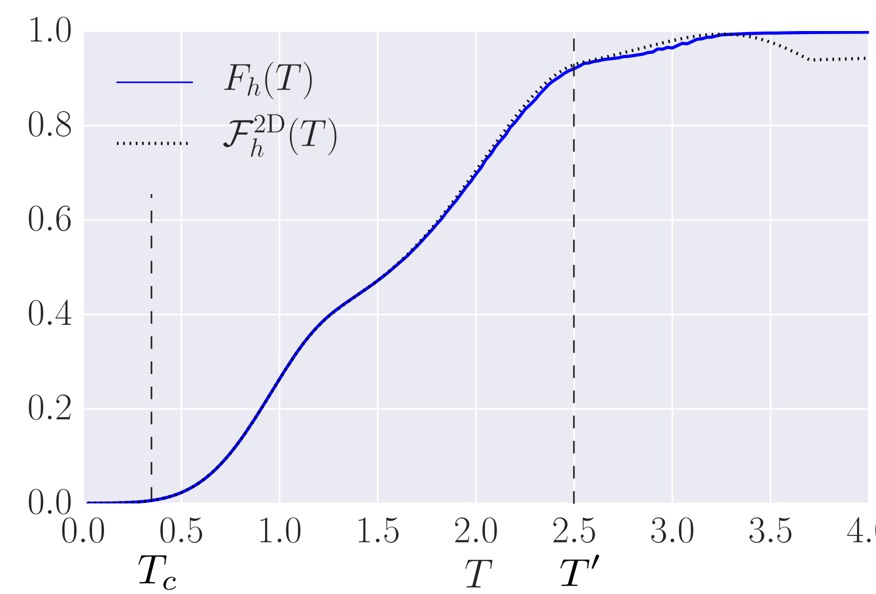
$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

$$-\mathcal{F}_h(T) = \min_{\tau^{(1)}\in[0,T]} \left(-\mathcal{F}_h(T;\tau^{(1)})\right)$$





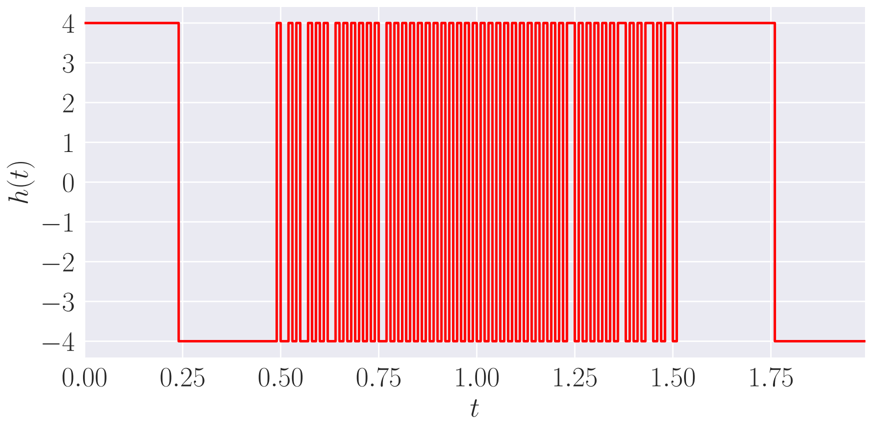




MB et al, PRX 8 031086 (2018)

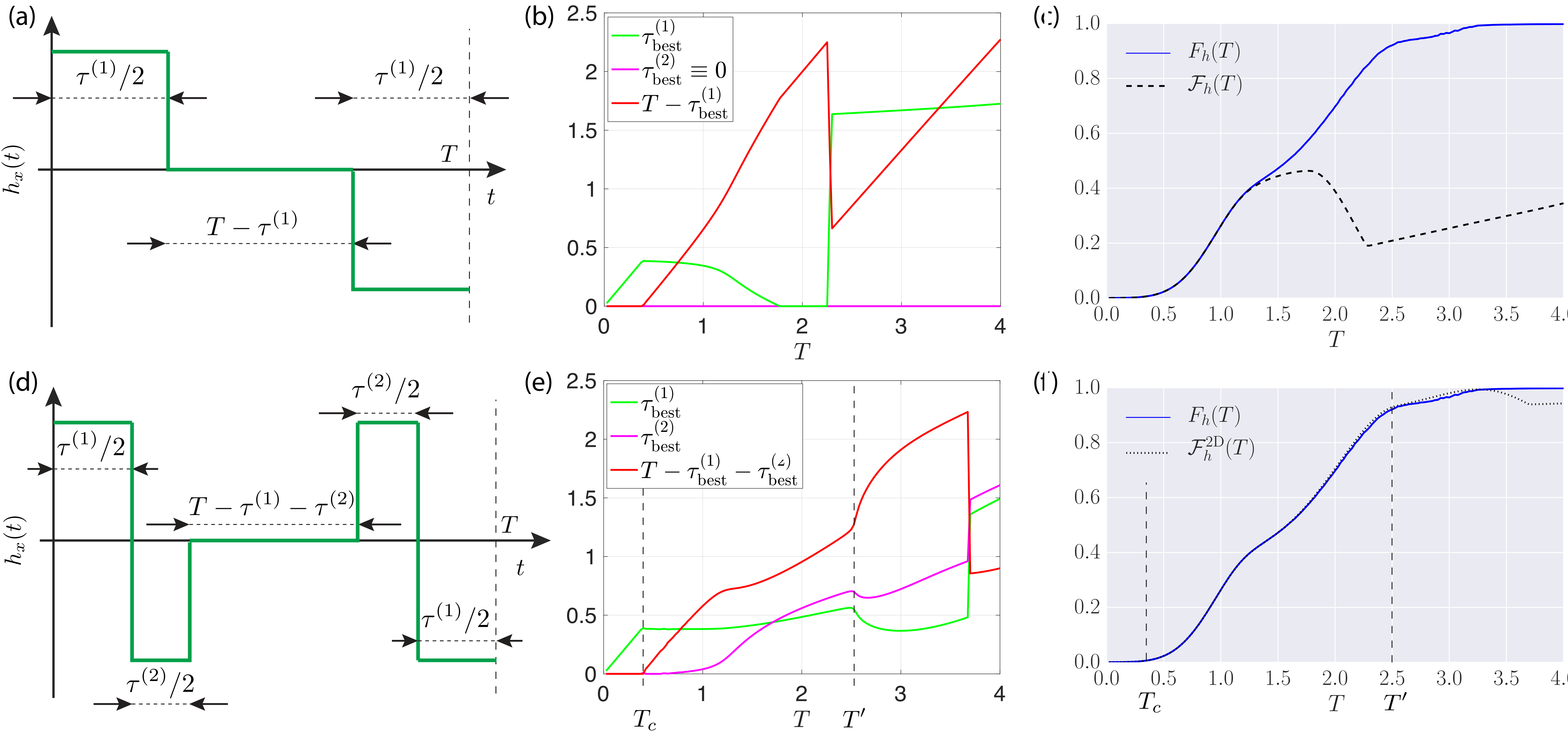
Berkeley
UNIVERSITY OF CALIFORNIA

What do we Learn from the RL Agent?



$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

$$-\mathcal{F}_h(T) = \min_{\tau^{(1)}\in[0,T]} \left(-\mathcal{F}_h(T;\tau^{(1)})\right)$$



Marin Bukov

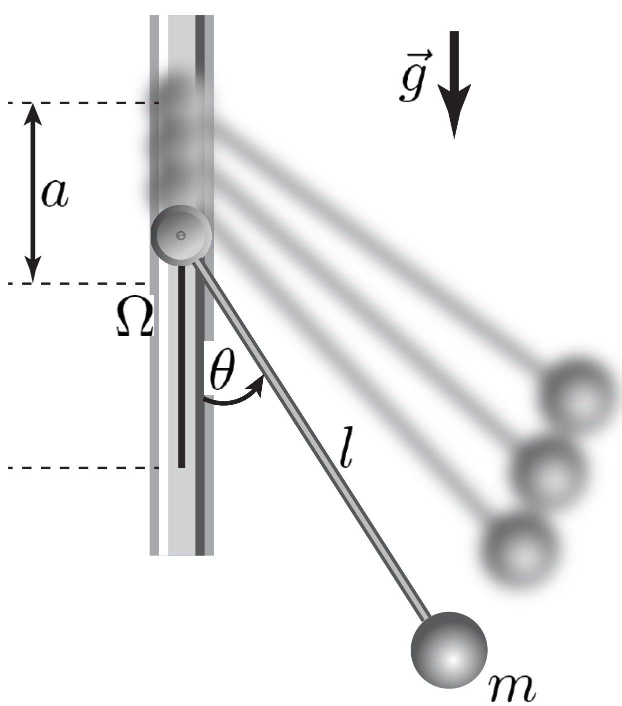
MB et al, PRX 8 031086 (2018)

Berkeley UNIVERSITY OF CALIFORNIA

Example 2:

use RL for autonomous preparation non-equilibrium states in a ***simulation of an "experiment"***

→ Kapitza oscillator



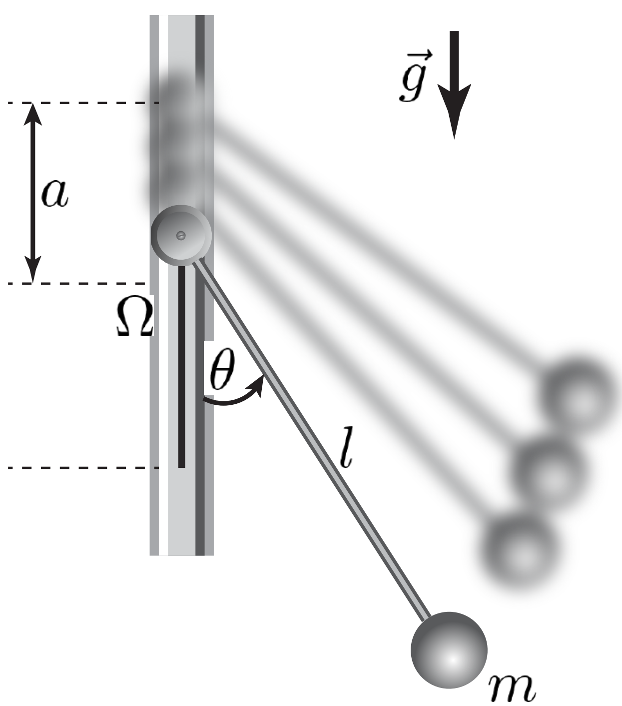
bluedwarf1127: YouTube (2009)

Berkeley
UNIVERSITY OF CALIFORNIA

Example 2:

use RL for autonomous preparation
non-equilibrium states in a ***simulation of an "experiment"***

→ Kapitza oscillator



Marin Bukov

bluedwarf1127: YouTube (2009)

The quantum Kapitza oscillator

→ how do we understand 'dynamical stabilization'?

$$H_{\mathrm{lab}}(t) = \frac{p_\theta^2}{2m} - \left(m\omega_0^2 + A\Omega\cos\Omega t\right)\cos\theta$$

Marin Bukov

The quantum Kapitza oscillator

→ how do we understand 'dynamical stabilization'?

$$H_{\rm lab}(t) = \frac{p_\theta^2}{2m} - \left(m\omega_0^2 + A\Omega\cos\Omega t\right)\cos\theta$$

→ find effective description in high-frequency limit $\Omega \to \infty$

- intuitively: take time average
- problem: drive averages out to zero, yet we've seen the effect!

The quantum Kapitza oscillator

→ how do we understand 'dynamical stabilization'?

$$H_{\text{lab}}(t) = \frac{p_\theta^2}{2m} - \left(m\omega_0^2 + A\Omega\cos\Omega t\right)\cos\theta$$

→ find effective description in high-frequency limit $\Omega \to \infty$

- intuitively: take time average
- problem: drive averages out to zero, yet we've seen the effect!

→ change reference frames to "remove" strong coupling:

$$H_{\text{rot}}(t) = V^\dagger(t) H_{\text{lab}}(t) V(t) - iV^\dagger(t)\partial_t V(t) \qquad V(t) = \mathrm{e}^{iA\sin\Omega t\cos\theta}$$

Marin Bukov

The quantum Kapitza oscillator

→ how do we understand 'dynamical stabilization'?

$$H_{\text{lab}}(t) = \frac{p_\theta^2}{2m} - \left(m\omega_0^2 + A\Omega\cos\Omega t\right)\cos\theta$$

→ find effective description in high-frequency limit $\Omega \to \infty$

- intuitively: take time average
- problem: drive averages out to zero, yet we've seen the effect!

→ change reference frames to "remove" strong coupling:

$$H_{\text{rot}}(t) = V^\dagger(t)H_{\text{lab}}(t)V(t) - iV^\dagger(t)\partial_t V(t) \qquad V(t) = \mathrm{e}^{iA\sin\Omega t\cos\theta}$$

$$H_{\text{rot}}(t) = \frac{p_\theta^2}{2m} - m\omega_0^2\cos\theta - \frac{A}{2m}\sin\Omega t[p,\sin\theta]_+ + \frac{A^2}{2m}\sin^2\Omega t\cos 2\theta$$

The quantum Kapitza oscillator

➞ how do we understand ‘dynamical stabilization’?

$$H_{\text{lab}}(t) = \frac{p_\theta^2}{2m} - \left(m\omega_0^2 + A\Omega\cos\Omega t\right)\cos\theta$$

➞ find effective description in high-frequency limit $\Omega \to \infty$

- intuitively: take time average
- problem: drive averages out to zero, yet we’ve seen the effect!

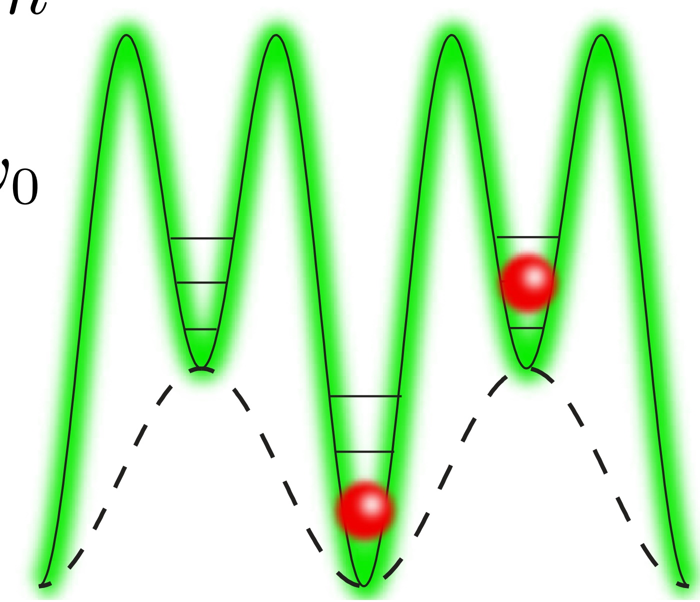
➞ change reference frames to “remove” strong coupling:

$$H_{\text{rot}}(t) = V^\dagger(t)H_{\text{lab}}(t)V(t) - iV^\dagger(t)\partial_t V(t) \qquad V(t) = \mathrm{e}^{iA\sin\Omega t\cos\theta}$$

$$H_{\text{rot}}(t) = \frac{p_\theta^2}{2m} - m\omega_0^2\cos\theta - \frac{A}{2m}\sin\Omega t[p,\sin\theta]_+ + \frac{A^2}{2m}\sin^2\Omega t\cos 2\theta$$

➞ time-average now easy to take

$$A > \sqrt{2}m\omega_0$$



$$H_{\text{ave}} = \frac{p_\theta^2}{2m} - m\omega_0^2\cos\theta + \frac{A^2}{4m}\cos 2\theta$$

Marin Bukov

The quantum Kapitza oscillator

→ how do we understand 'dynamical stabilization'?

$$H_{\text{lab}}(t) = \frac{p_\theta^2}{2m} - \left(m\omega_0^2 + A\Omega\cos\Omega t\right)\cos\theta$$

→ find effective description in high-frequency limit $\Omega \to \infty$

- intuitively: take time average
- problem: drive averages out to zero, yet we've seen the effect!

→ change reference frames to "remove" strong coupling:

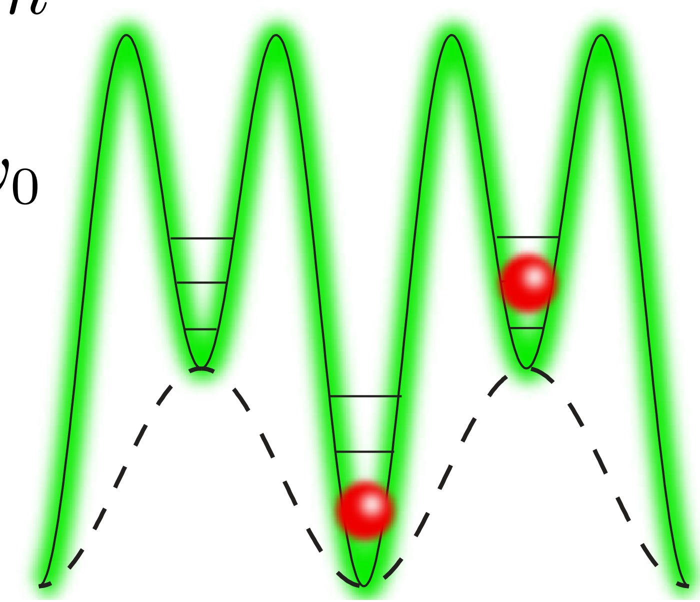
$$H_{\text{rot}}(t) = V^\dagger(t)H_{\text{lab}}(t)V(t) - iV^\dagger(t)\partial_t V(t) \qquad V(t) = \mathrm{e}^{iA\sin\Omega t\cos\theta}$$

$$H_{\text{rot}}(t) = \frac{p_\theta^2}{2m} - m\omega_0^2\cos\theta - \frac{A}{2m}\sin\Omega t[p,\sin\theta]_+ + \frac{A^2}{2m}\sin^2\Omega t\cos 2\theta$$

→ time-average now easy to take

$$A > \sqrt{2}m\omega_0$$

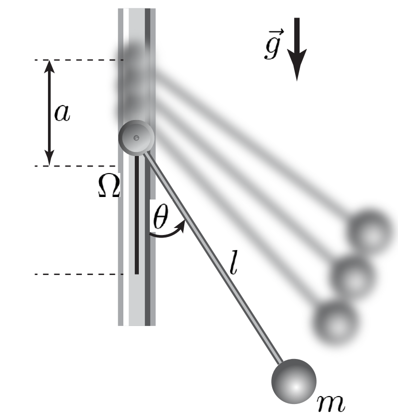
$$H_{\text{ave}} = \frac{p_\theta^2}{2m} - m\omega_0^2\cos\theta + \frac{A^2}{4m}\cos 2\theta$$



→ finite frequencies: Floquet Hamiltonian $H_F(\Omega)$

Marin Bukov

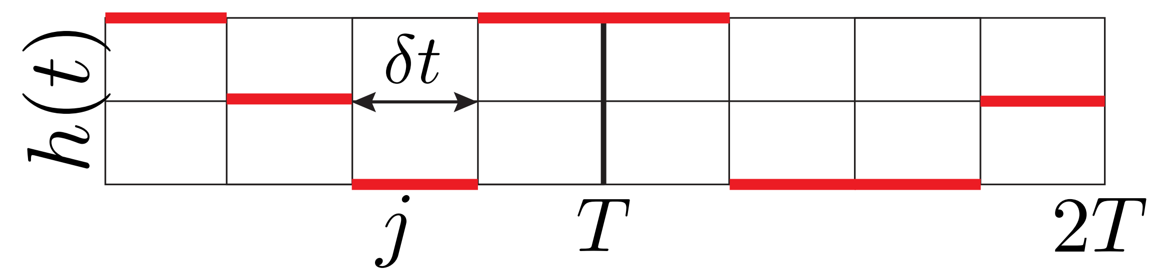
Floquet Control Problem



→ find optimal control field ***on top of periodic drive***

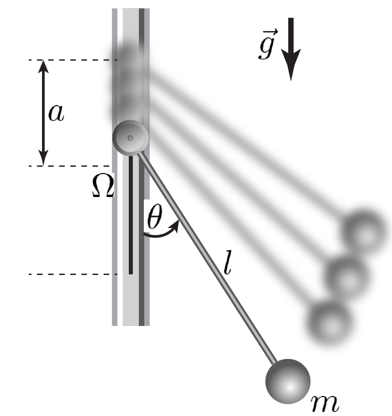
$$H_{\rm rot}(t) = H_0 + H_{\rm drive}(t) + H_{\rm control}(t)$$

$$H_{\rm control}(t) = h(t)\sin\theta$$ horizontal kicks



MB, arXiv:1808:08910

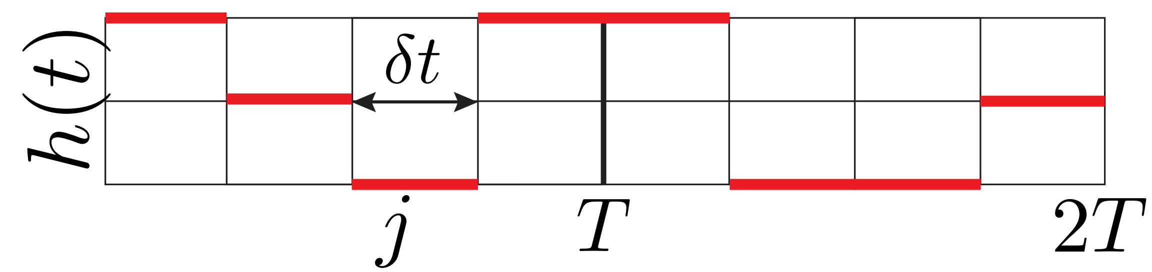
Floquet Control Problem



→ find optimal control field ***on top of periodic drive***

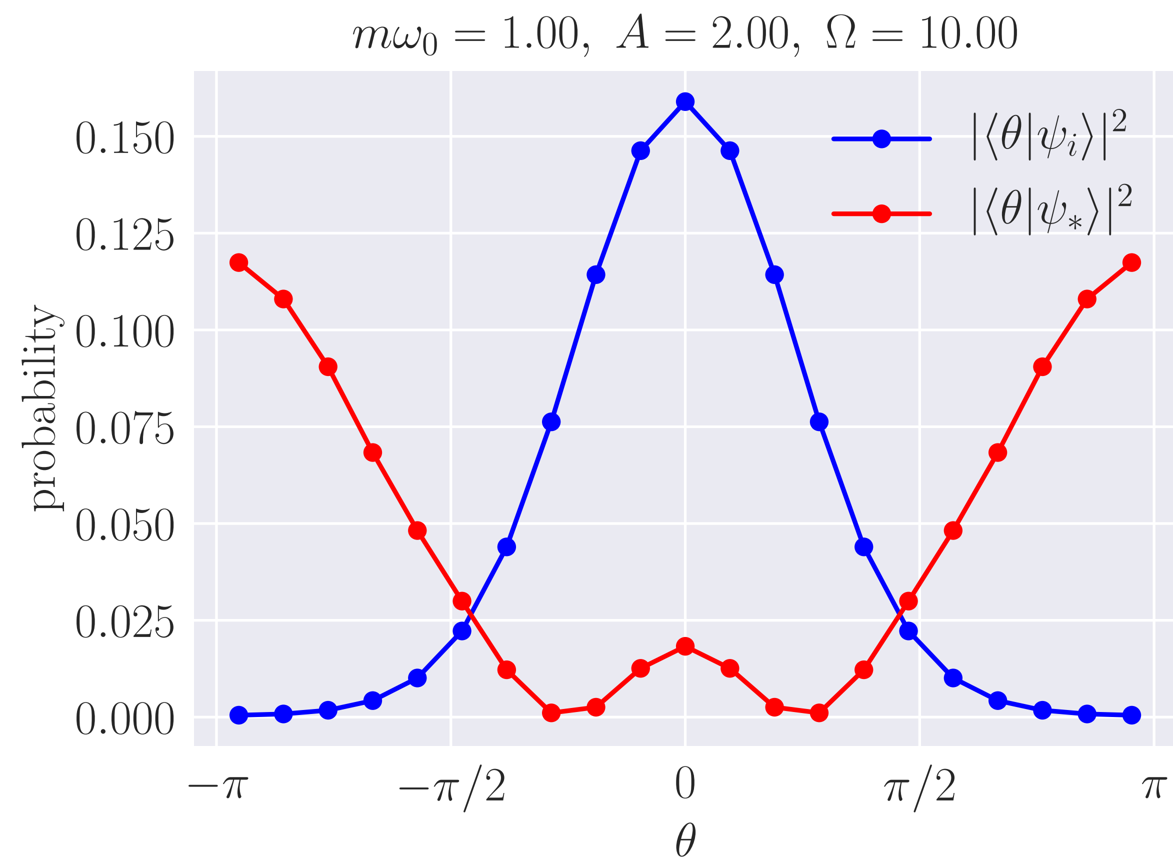
$$H_{\text{rot}}(t) = H_0 + H_{\text{drive}}(t) + H_{\text{control}}(t)$$

$$H_{\text{control}}(t) = h(t)\sin\theta$$ horizontal kicks



initial state: $|\psi_i\rangle$: GS of H_0

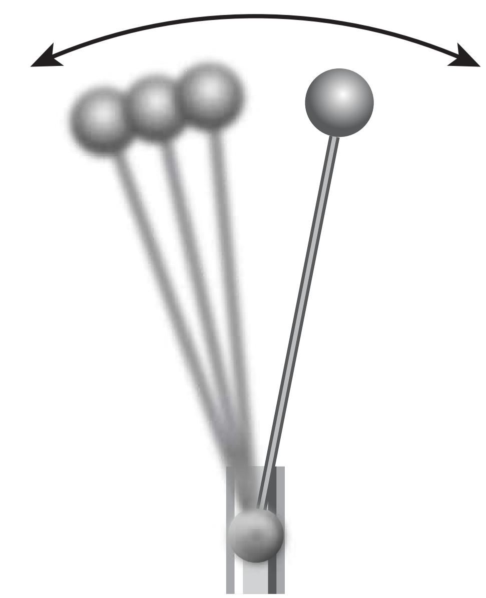
target state: $|\psi_*\rangle$ inverted position eigenstate of $H_F(\Omega)$



Marin Bukov

MB, arXiv:1808:08910

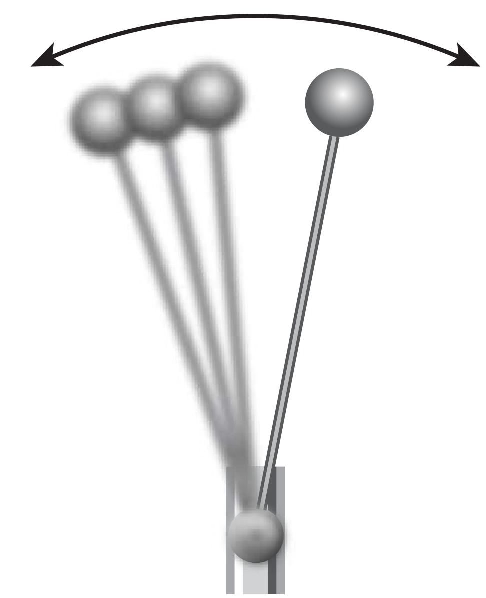
Simulation of a Quantum Experiment



→ **no** direct **access** to quantum state:
"play game w/o looking at screen" (only know score)

$$\{|\psi(t)\rangle : |\psi(t)\rangle = U_h(t,0)|\psi_i\rangle\} \hat{=} \{h(t) : |\psi_i\rangle\}$$

Simulation of a Quantum Experiment



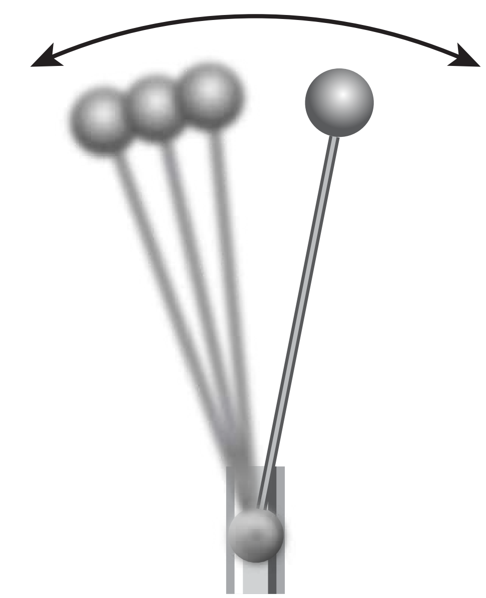
→ **no** direct **access** to quantum state:
"play game w/o looking at screen" (only know score)

$$\{|\psi(t)\rangle : |\psi(t)\rangle = U_h(t,0)|\psi_i\rangle\} \hat{=} \{h(t) : |\psi_i\rangle\}$$

→ **probabilistic** quantum measurements

$+1$ with probability $F_h(t_f) = |\langle\psi(t_f)|\psi_*\rangle|^2$
-1 otherwise

Simulation of a Quantum Experiment



→ **no** direct **access** to quantum state:
"play game w/o looking at screen" (only know score)

$$\{|\psi(t)\rangle : |\psi(t)\rangle = U_h(t,0)|\psi_i\rangle\} \hat{=} \{h(t) : |\psi_i\rangle\}$$

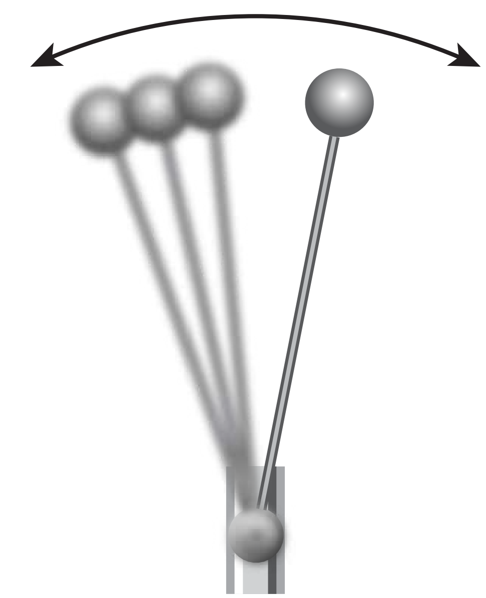
→ **probabilistic** quantum measurements

$+1$ with probability $F_h(t_f) = |\langle\psi(t_f)|\psi_*\rangle|^2$

-1 otherwise

→ **uncertainty** in preparing initial state

Simulation of a Quantum Experiment



→ **no** direct **access** to quantum state:
"play game w/o looking at screen" (only know score)

$$\{|\psi(t)\rangle : |\psi(t)\rangle = U_h(t,0)|\psi_i\rangle\} \hat{=} \{h(t) : |\psi_i\rangle\}$$

→ **probabilistic** quantum measurements

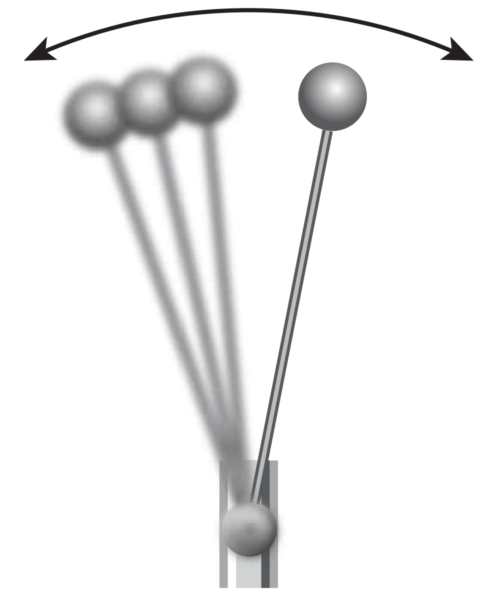
$+1$ with probability $F_h(t_f) = |\langle\psi(t_f)|\psi_*\rangle|^2$
-1 otherwise

→ **uncertainty** in preparing initial state

→ occasional **failure** of control apparatus

Berkeley UNIVERSITY OF CALIFORNIA

Simulation of a Quantum Experiment



→ **no** direct **access** to quantum state:
"play game w/o looking at screen" (only know score)

$$\{|\psi(t)\rangle : |\psi(t)\rangle = U_h(t,0)|\psi_i\rangle\} \hat{=} \{h(t) : |\psi_i\rangle\}$$

→ **probabilistic** quantum measurements

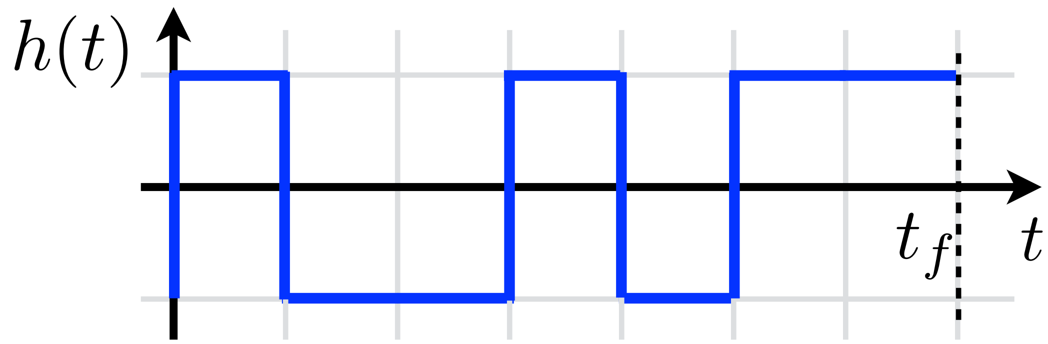
$+1$ with probability $F_h(t_f) = |\langle\psi(t_f)|\psi_*\rangle|^2$
-1 otherwise

→ **uncertainty** in preparing initial state

→ occasional **failure** of control apparatus

→ *additionally:* all other problems of how to actually prepare the state if the above were absent and no analytic solution is known

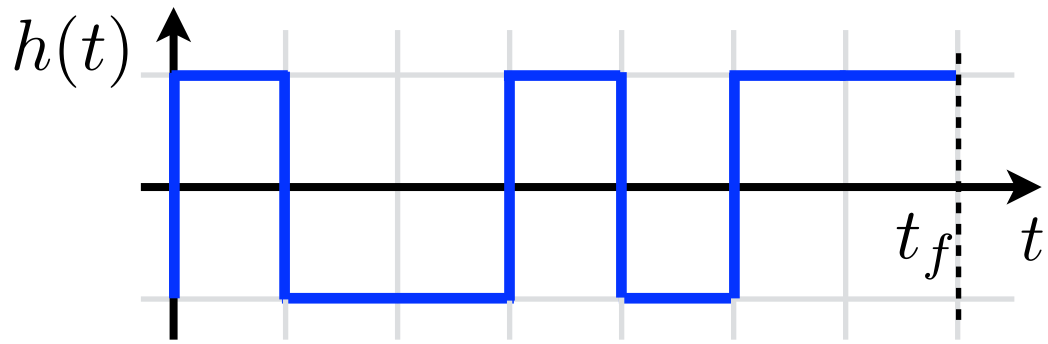
Let's give this "game" a try!



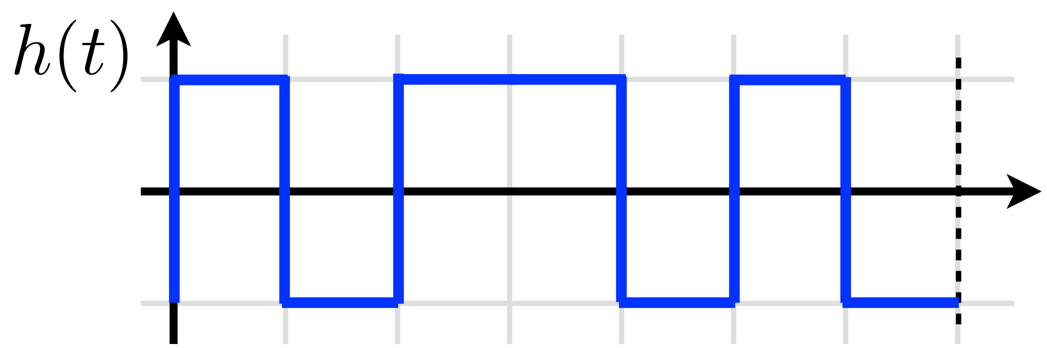
measurement: -1

Berkeley UNIVERSITY OF CALIFORNIA

Let's give this "game" a try!



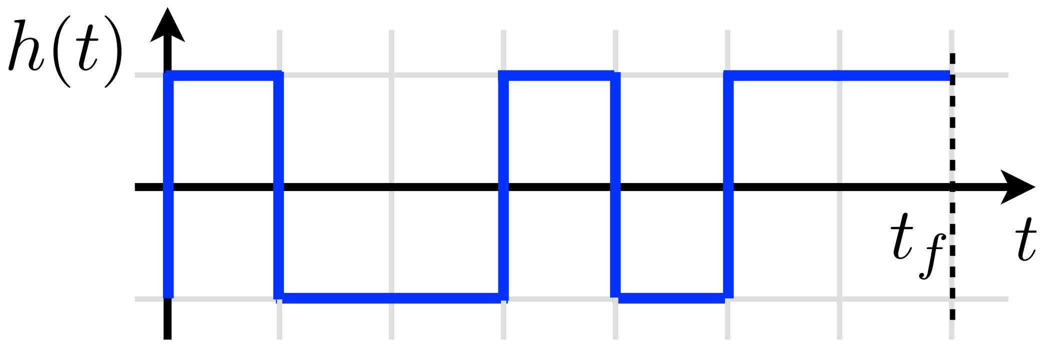
measurement: -1



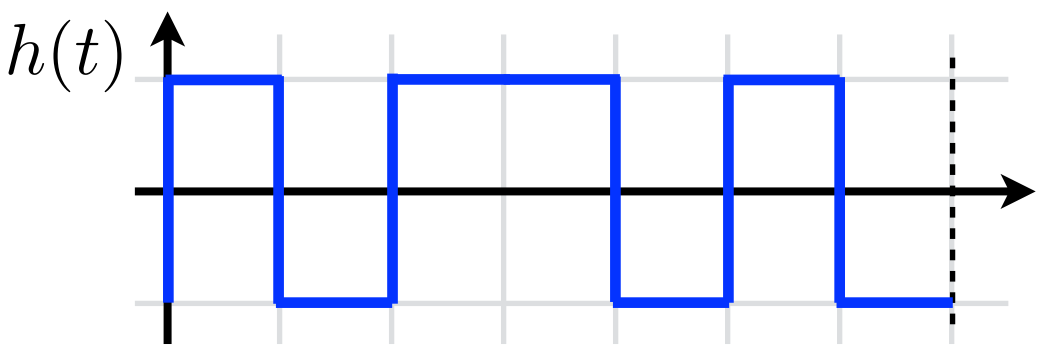
measurement: $+1$

(different final state: different probability to be in the target state)

Let's give this "game" a try!



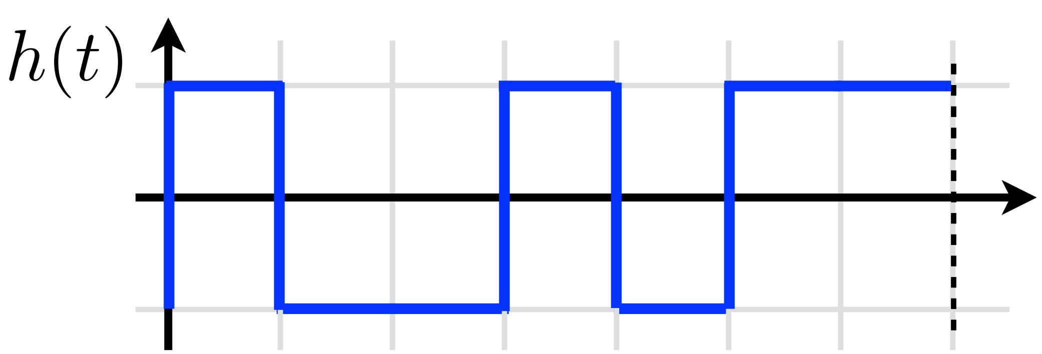
measurement: -1



measurement: $+1$

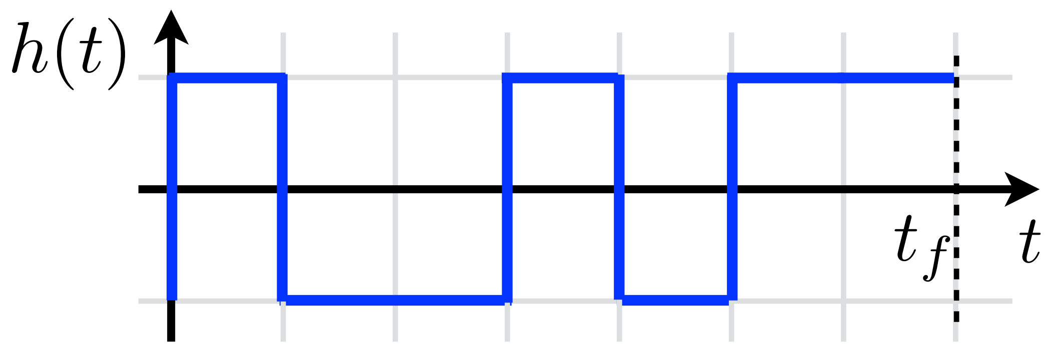
(different final state: different probability to be in the target state)

→ repeat protocol!

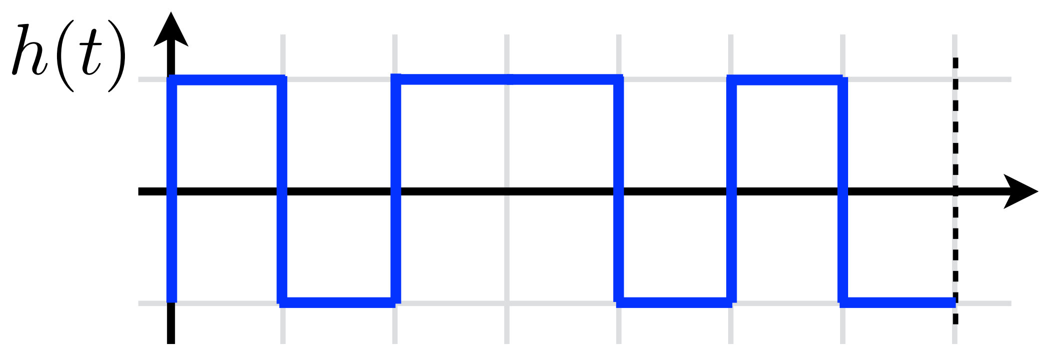


measurement: $+1$

Let's give this "game" a try!



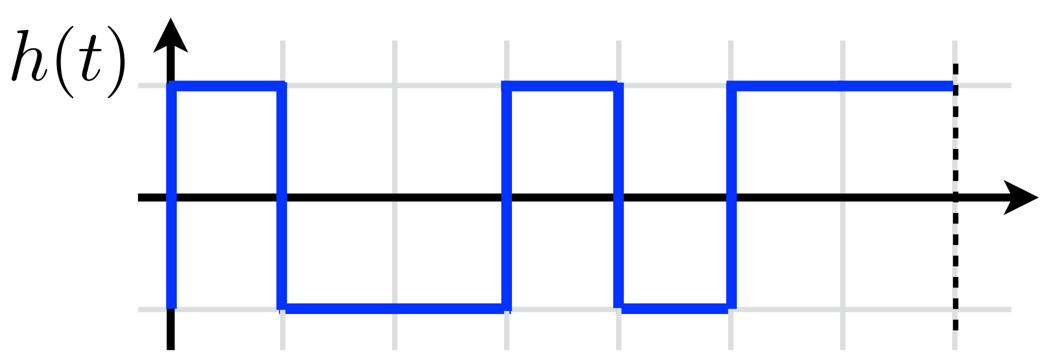
measurement: -1



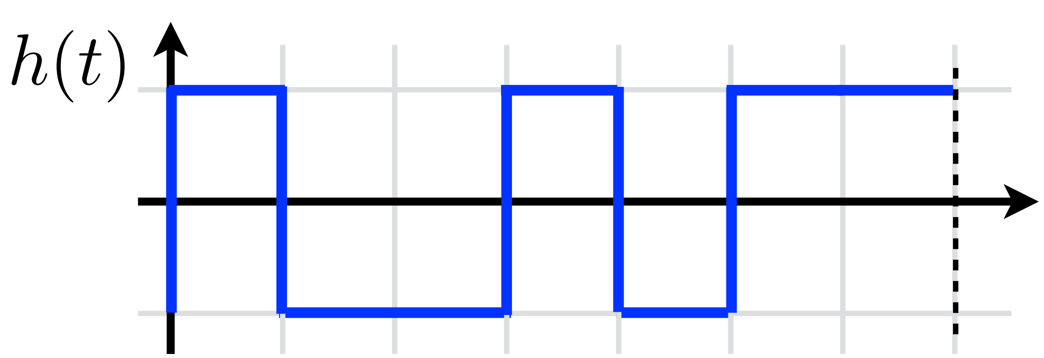
measurement: $+1$

(different final state: different probability to be in the target state)

→ repeat protocol!

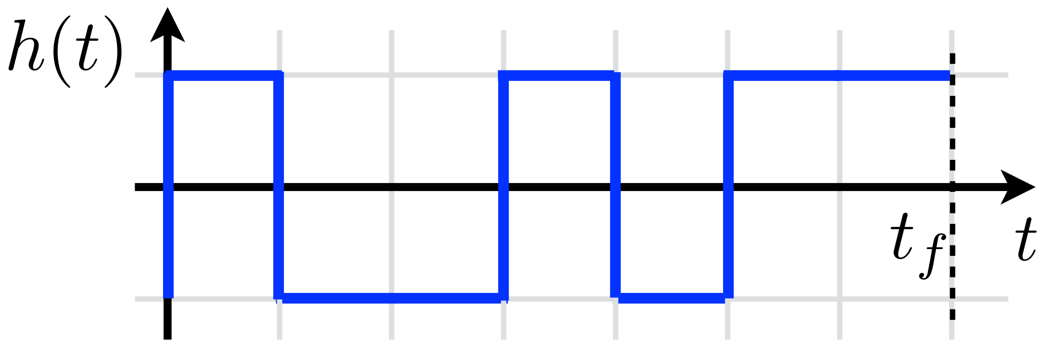


measurement: $+1$

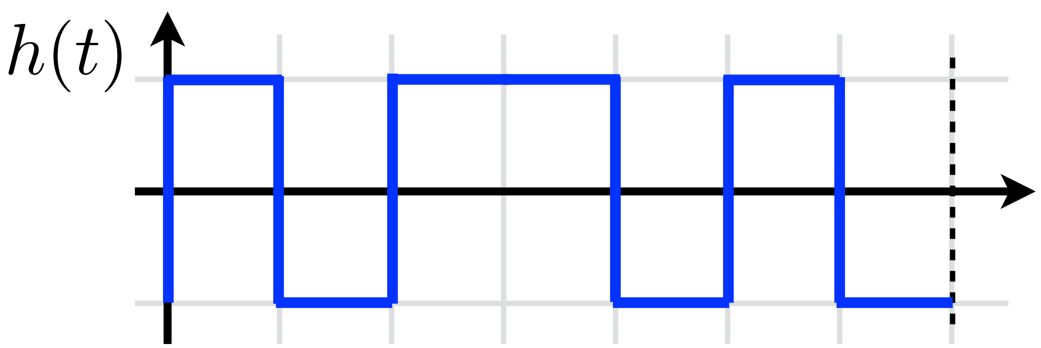


measurement: -1

Let’s give this “game” a try!



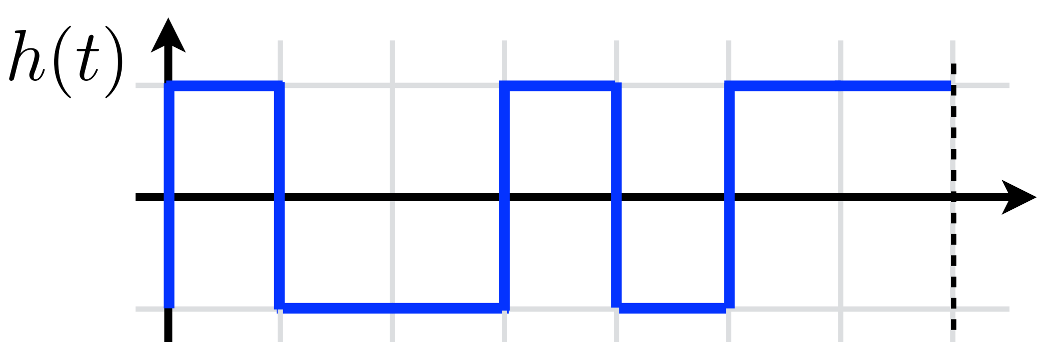
measurement: -1



measurement: $+1$

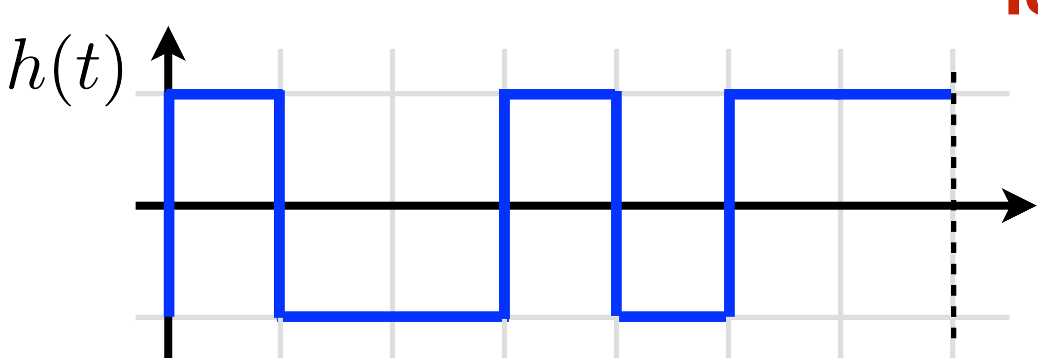
(different final state: different probability to be in the target state)

→ repeat protocol!



measurement: $+1$

learn from noisy, nondeterministic rewards

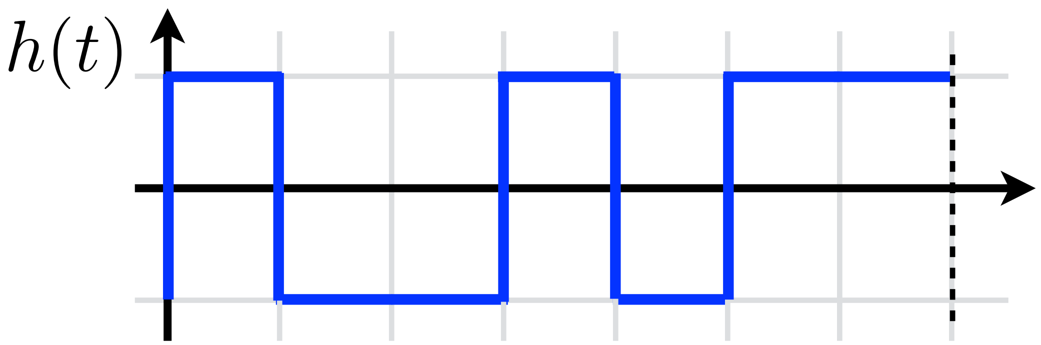


measurement: -1

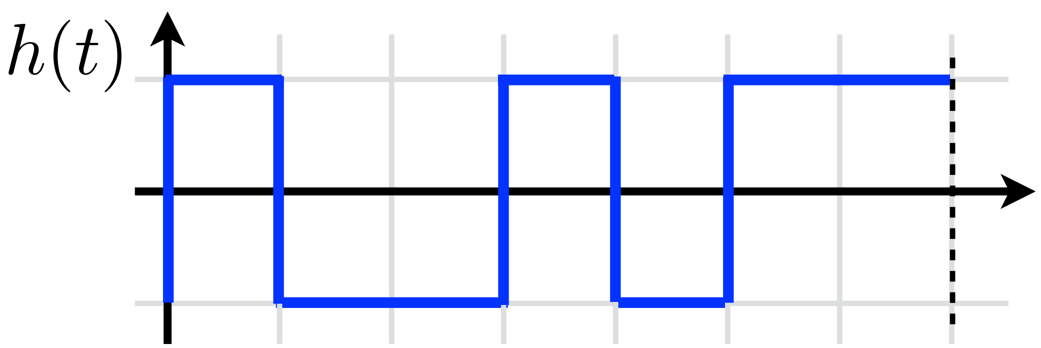
Berkeley
UNIVERSITY OF CALIFORNIA

Let's get rid of this 'quantumness' for a sec

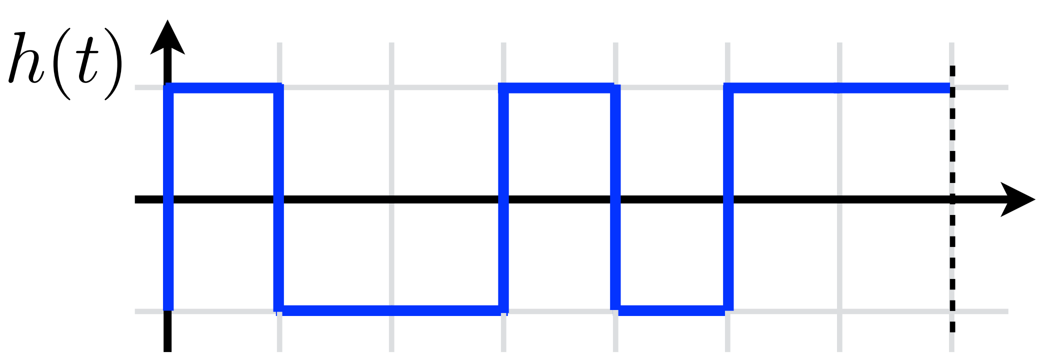
→ repeat protocol again!



measurement:
$F_h = |\langle\psi(T)|\psi_*\rangle|^2 = 0.632$



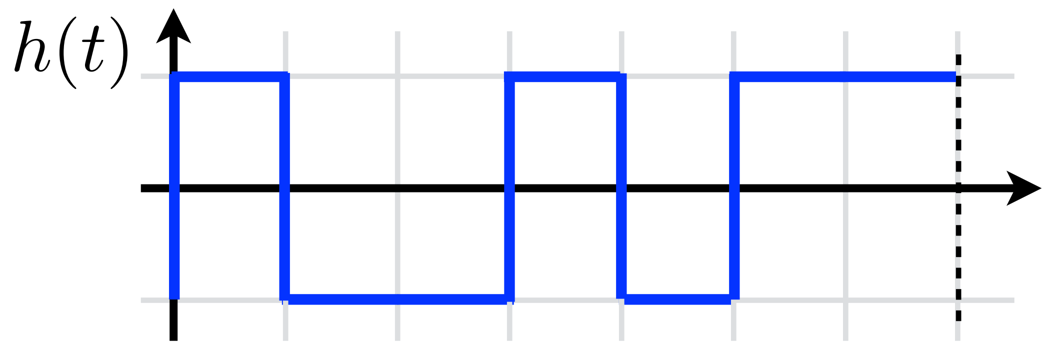
measurement: $F_h = 0.592$



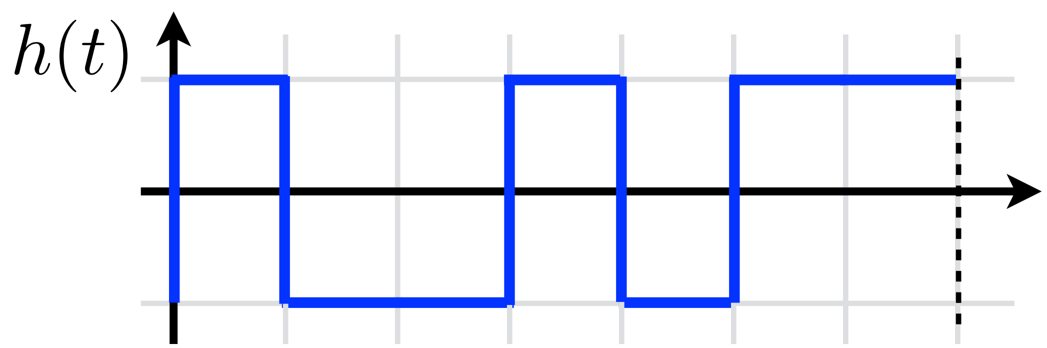
measurement: $F_h = 0.668$

Let's get rid of this 'quantumness' for a sec

→ repeat protocol again!

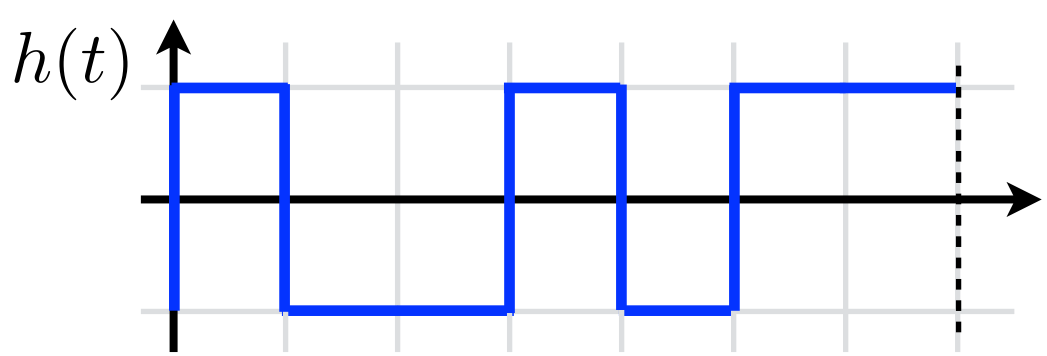


measurement:
$F_h = |\langle\psi(T)|\psi_*\rangle|^2 = 0.632$



measurement: $F_h = 0.592$

initial state could not be prepared perfectly: more headache!



measurement: $F_h = 0.668$

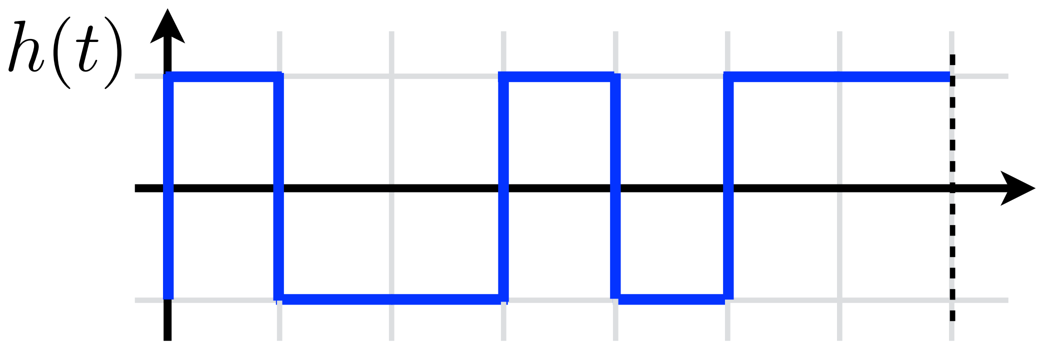
Marin Bukov

MB PRB 98, 224305 (2018)

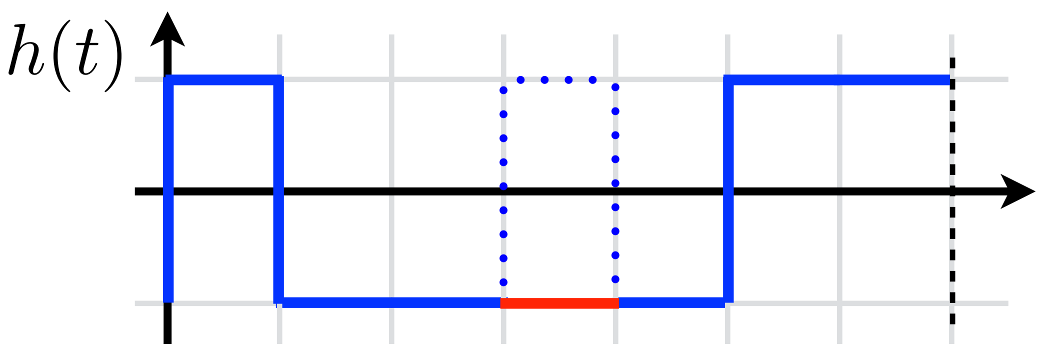
Berkeley UNIVERSITY OF CALIFORNIA

what if we fix the initial state:

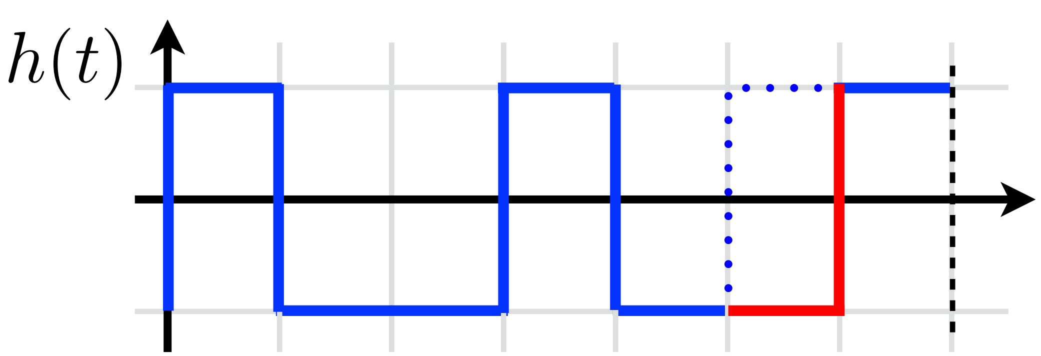
→ repeat protocol again!



measurement: $F_h = 0.627$



measurement: $F_h = 0.572$

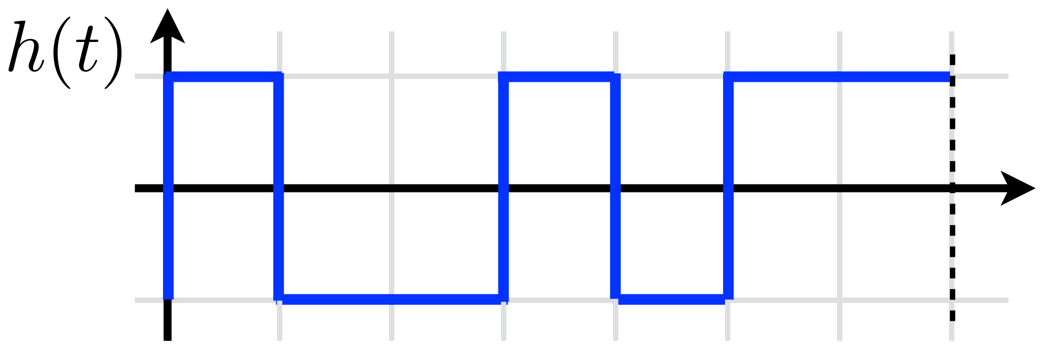


measurement: $F_h = 0.657$

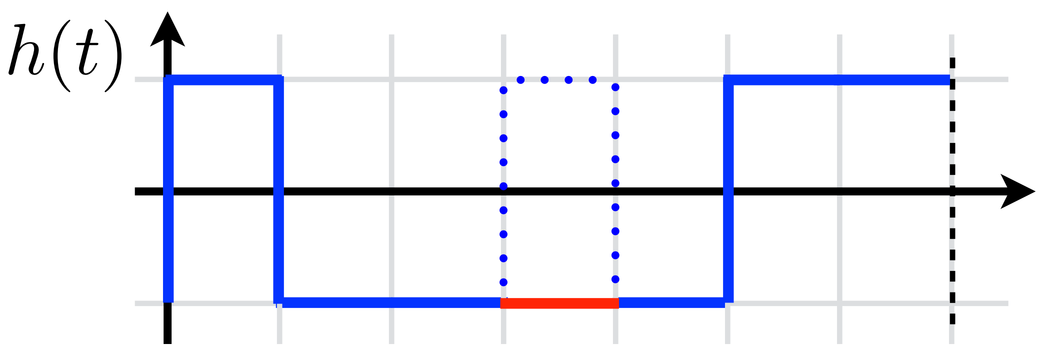
Berkeley
UNIVERSITY OF CALIFORNIA

what if we fix the initial state:

→ repeat protocol again!

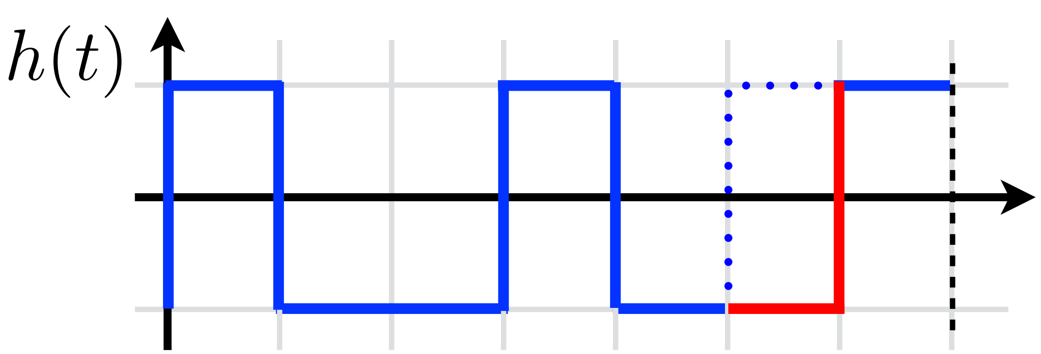


measurement: $F_h = 0.627$



measurement: $F_h = 0.572$

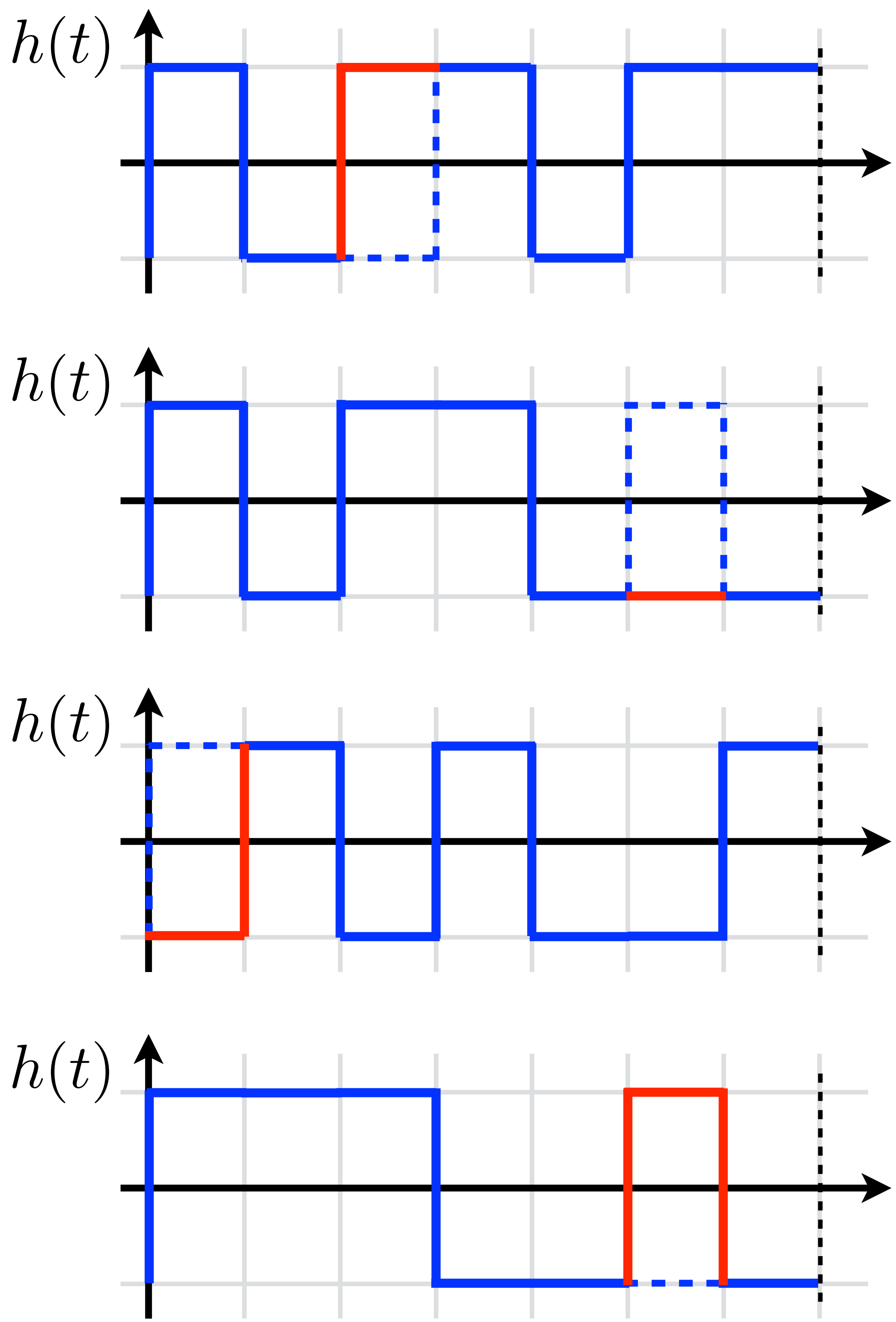
control apparatus failed: it can't be!



measurement: $F_h = 0.657$

Berkeley UNIVERSITY OF CALIFORNIA

The Cruel Reality: all together (and probably much more!)



measurement: -1

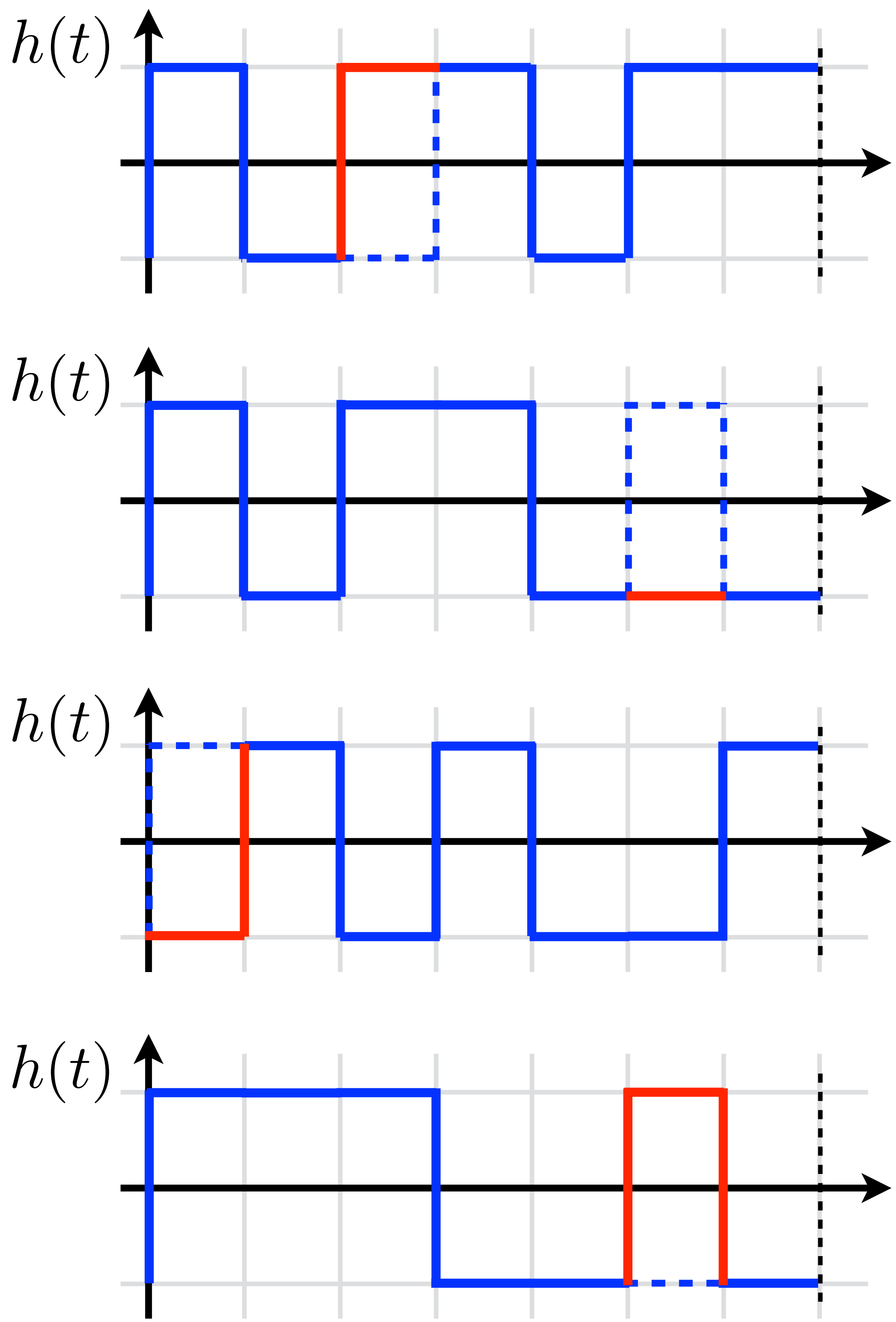
measurement: $+1$

measurement: -1

measurement: -1

Berkeley UNIVERSITY OF CALIFORNIA

The Cruel Reality: all together (and probably much more!)



measurement: -1

measurement: $+1$

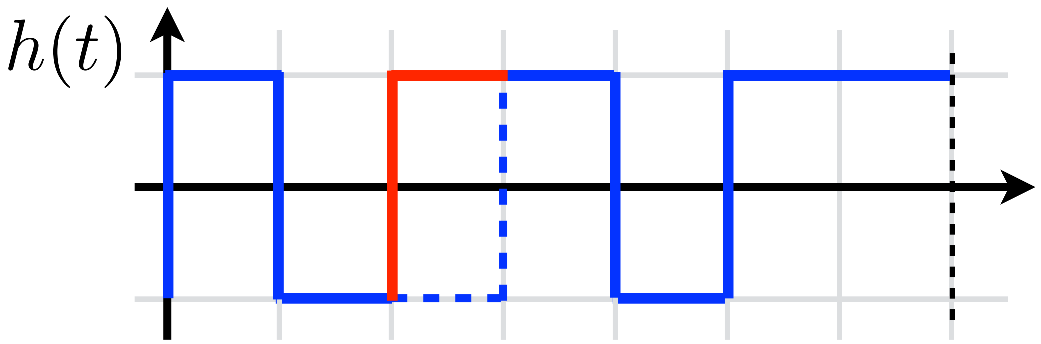
extremely tedious task!

measurement: -1

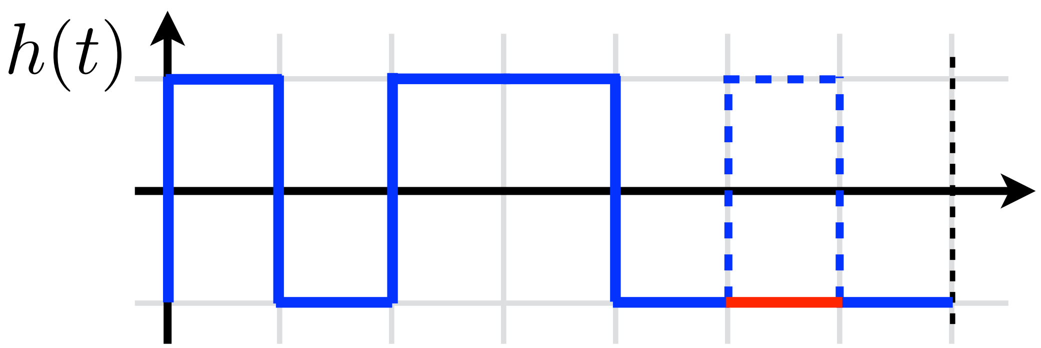
measurement: -1

Berkeley
UNIVERSITY OF CALIFORNIA

The Cruel Reality:
all together (and probably much more!)

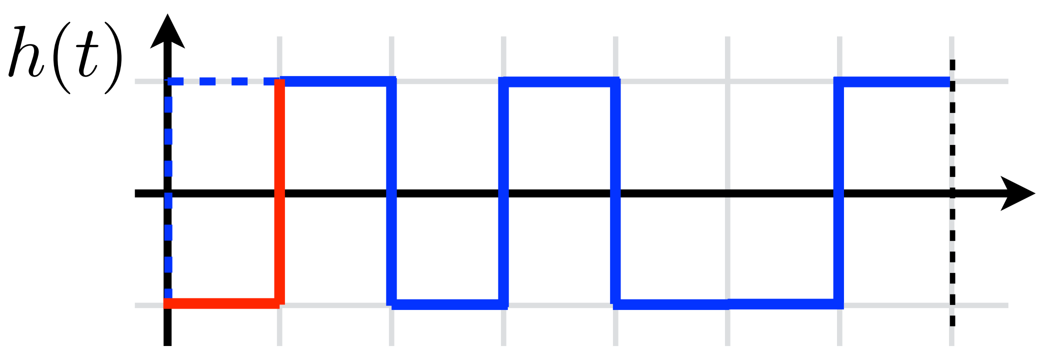


measurement: -1



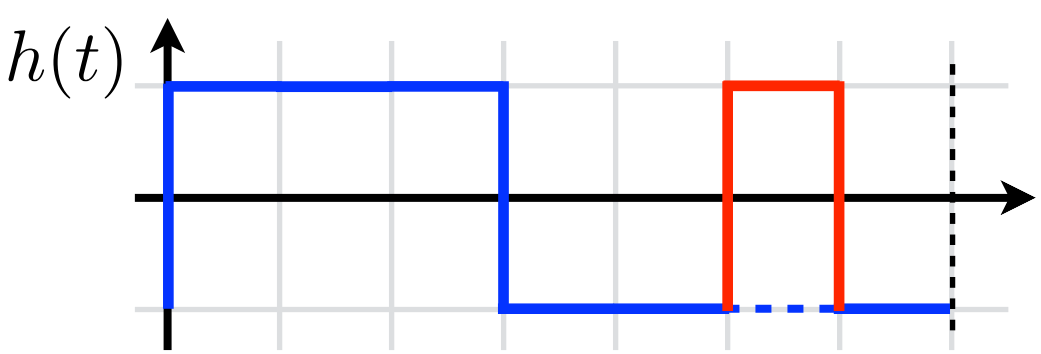
measurement: $+1$

extremely tedious task!



measurement: -1

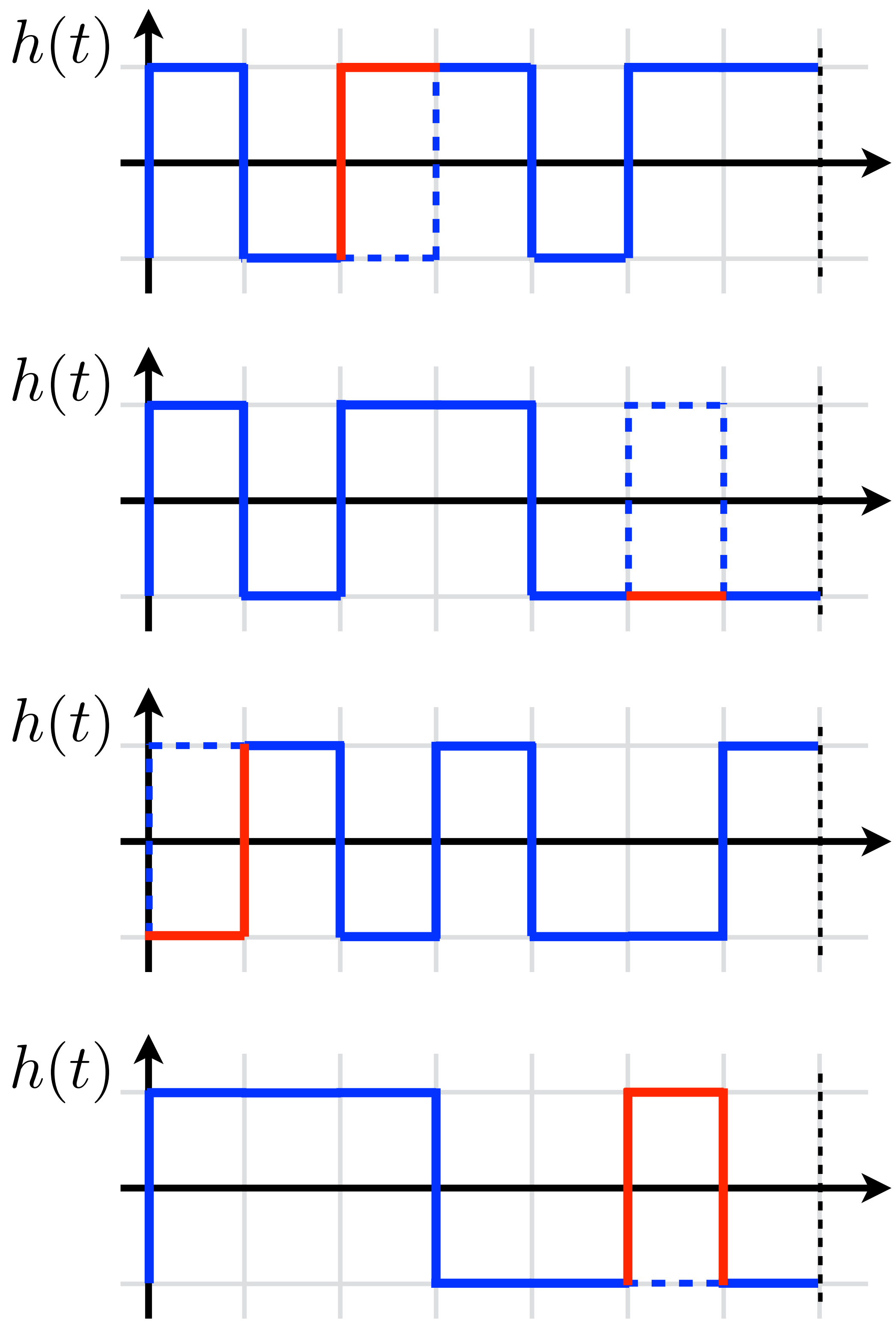
how do we solve it efficiently?



measurement: -1

Berkeley UNIVERSITY OF CALIFORNIA

The Cruel Reality: all together (and probably much more!)



measurement: -1

measurement: $+1$

extremely tedious task!

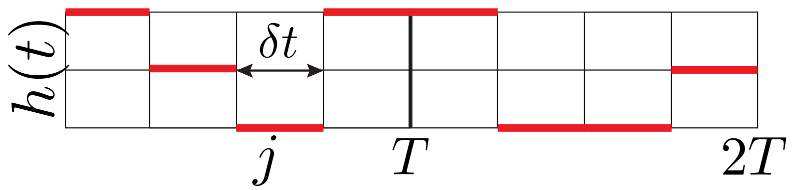
measurement: -1

how do we solve it efficiently?

measurement: -1

can we automate it?

Reinforcement Learning to Prepare the Inverted Position Floquet State

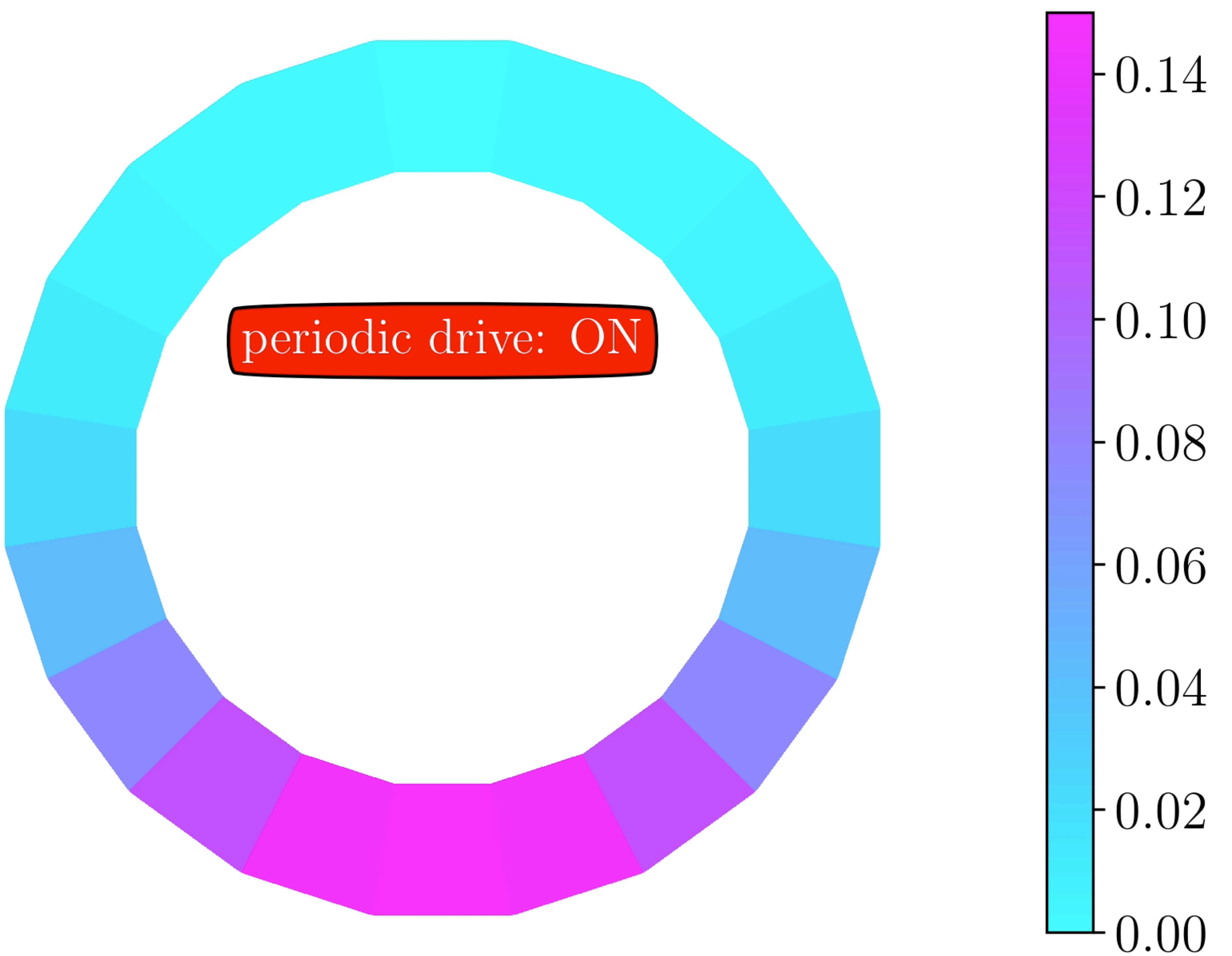


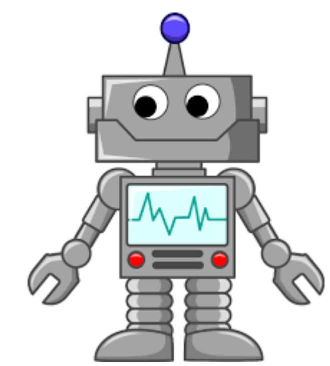
15 driving cycles (periods), 120 steps (8 per period)

$$|\mathcal{A}|^{N_T} = 3^{120} \approx 10^{57}$$

quantum Kapitza oscillator

$t/T = 0.00$ $F_h(t_f) = 0.00689$ $|\langle\theta|\psi(t)\rangle|^2$



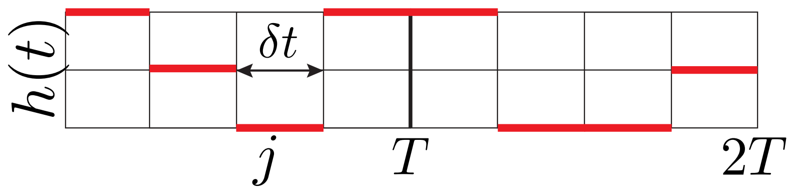


$h_{\max}/(m\omega_0) = 4.0$ $\Omega/(m\omega_0) = 10.0$ $A/(m\omega_0) = 2.0$

Marin Bukov

MB PRB 98, 224305 (2018)

Reinforcement Learning to Prepare the Inverted Position Floquet State

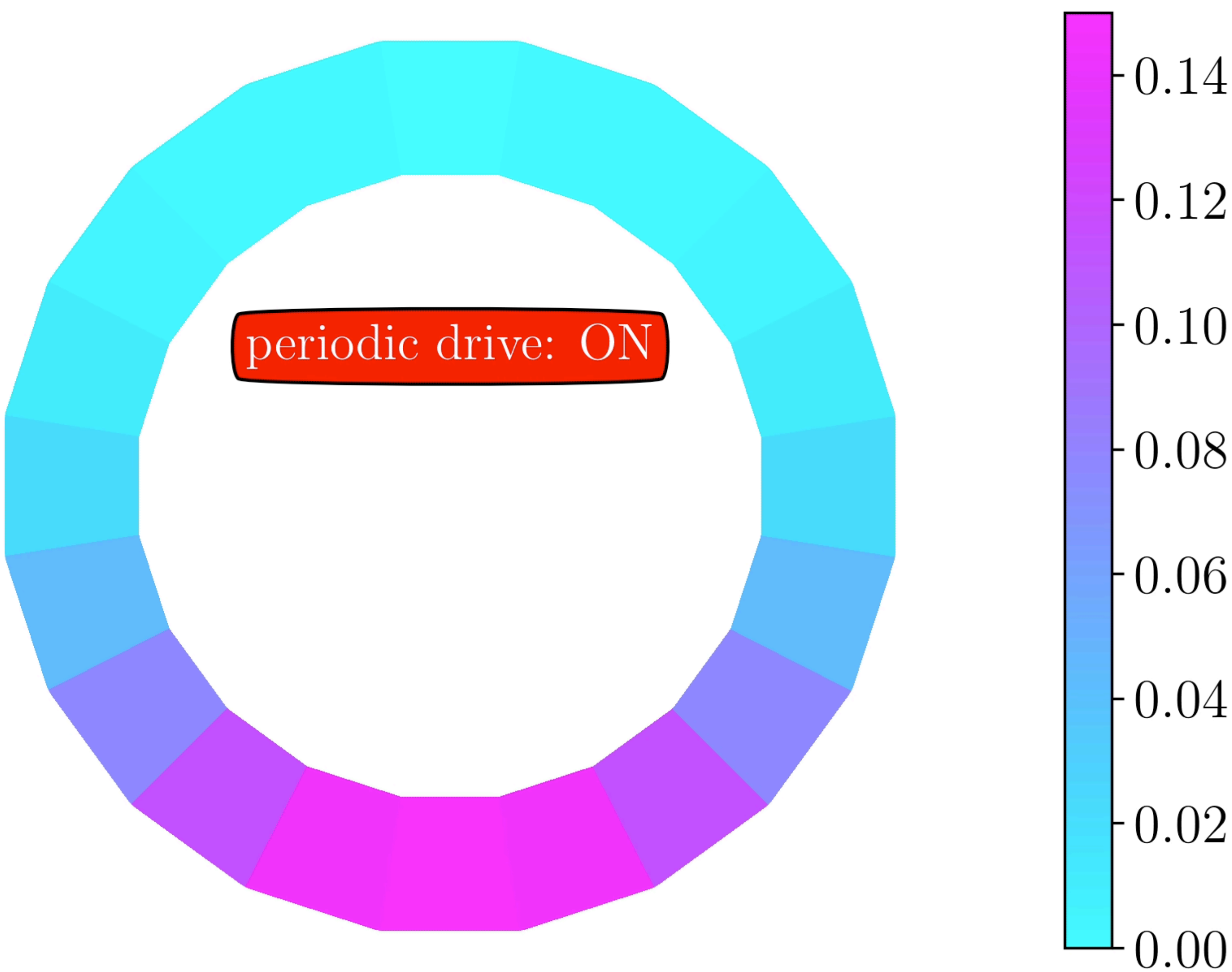


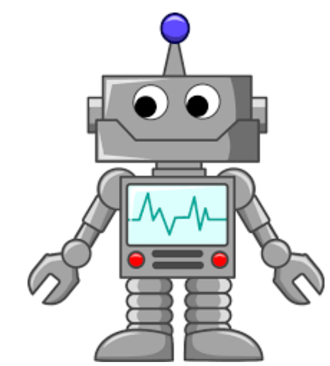
15 driving cycles (periods), 120 steps (8 per period)

$$|\mathcal{A}|^{N_T} = 3^{120} \approx 10^{57}$$

quantum Kapitza oscillator

$t/T = 0.00$ $\quad F_h(t_f) = 0.00689$ $\quad |\langle\theta|\psi(t)\rangle|^2$





$h_{\max}/(m\omega_0) = 4.0$ $\quad \Omega/(m\omega_0) = 10.0$ $\quad A/(m\omega_0) = 2.0$

Marin Bukov

MB PRB 98, 224305 (2018)

RL and Optimal Control (OC)

→ different sides of the same medal

- RL: appeared first in behavioral psychology: decision making
- OC: appeared in optimization problem solving: variational calculus

RL and Optimal Control (OC)

- different sides of the same medal
  - RL: appeared first in behavioral psychology: decision making
  - OC: appeared in optimization problem solving: variational calculus
- modern Control Theory: both RL and OC under same hood
- currently in physics: preferred approach is OC
  - for technical reasons: RL required large computational power, big data

Berkeley UNIVERSITY OF CALIFORNIA

RL and Optimal Control (OC)

→ different sides of the same medal

- RL: appeared first in behavioral psychology: decision making
- OC: appeared in optimization problem solving: variational calculus

→ modern Control Theory: both RL and OC under same hood

→ currently in physics: preferred approach is OC

- for technical reasons: RL required large computational power, big data

OC ⟵ *closely related*	⟶ **RL**
based on: ***variational calculus***	***Markov decision processes***
• needs model for environment to express cost function in.	• no model of controlled system, adaptive, autonomous.
• best suited for deterministic environments.	• stochastic/uncertain environments.
• differentiable cost function C_h uses gradient descent.	• reward function can be discontinuous, noisy.
• advantage: if we can compute analytically derivative of C_h.	• figures out effective degrees of freedom without a model.

Marin Bukov

Berkeley UNIVERSITY OF CALIFORNIA

Outlook

UNIVERSITY OF CALIFORNIA BERKELEY

→ Which problems can we study with RL that we can't do otherwise?

→ How do we use RL to discover new physics?

→ What are RL's most natural/appropriate applications in physics?

GORDON AND BETTY MOORE FOUNDATION

spin chain: PRX 8 031086 (2018)

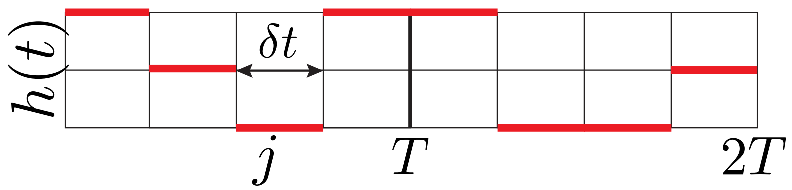
Kapitza oscillator: PRB 98, 224305 (2018)

QuSpin: *http://weinbe58.github.io/QuSpin*

python package for ED & many-body dynamics (with P. Weinberg, BU)

Berkeley UNIVERSITY OF CALIFORNIA

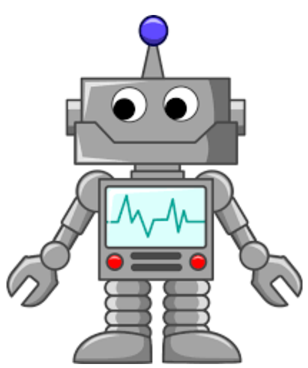
Reinforcement Learning to Prepare the Inverted Position Floquet State

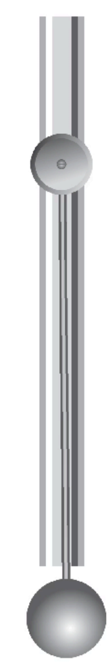


4 driving cycles (periods), 32 steps (8 per period)

Kapitza pendulum

$t/T = 0.00,\ \theta(t) = 0.00\pi,\ p_\theta(t) = 0.00,\ r(t) = 0.00$





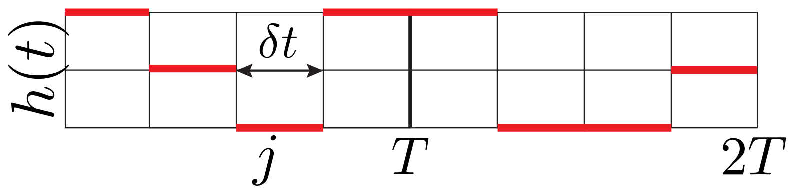
periodic drive: ON

$h_{\max}/(m\omega_0) = 4.0$ $\Omega/(m\omega_0) = 10.0$ $A/(m\omega_0) = 2.0$

Marin Bukov

Berkeley UNIVERSITY OF CALIFORNIA

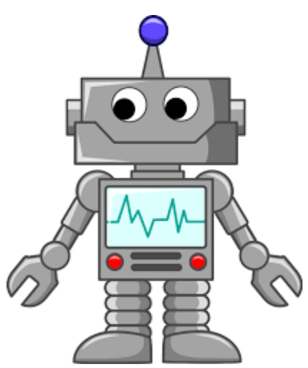
Reinforcement Learning to Prepare the Inverted Position Floquet State

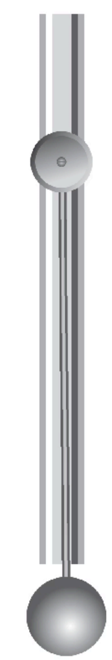


4 driving cycles (periods), 32 steps (8 per period)

Kapitza pendulum

$t/T = 0.00,\ \theta(t) = 0.00\pi,\ p_\theta(t) = 0.00,\ r(t) = 0.00$



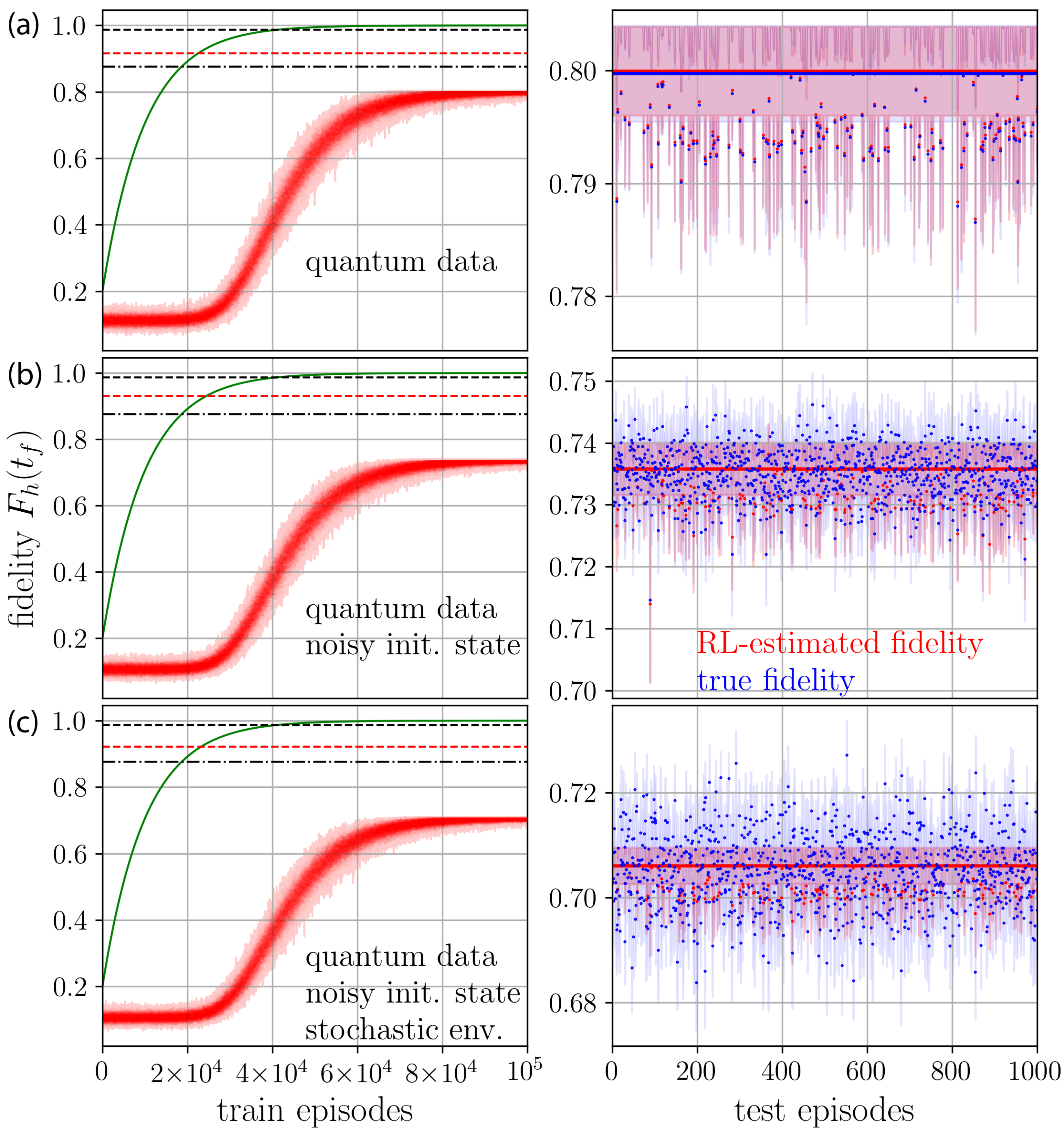


periodic drive: ON

$h_{\max}/(m\omega_0) = 4.0$ $\Omega/(m\omega_0) = 10.0$ $A/(m\omega_0) = 2.0$

Marin Bukov

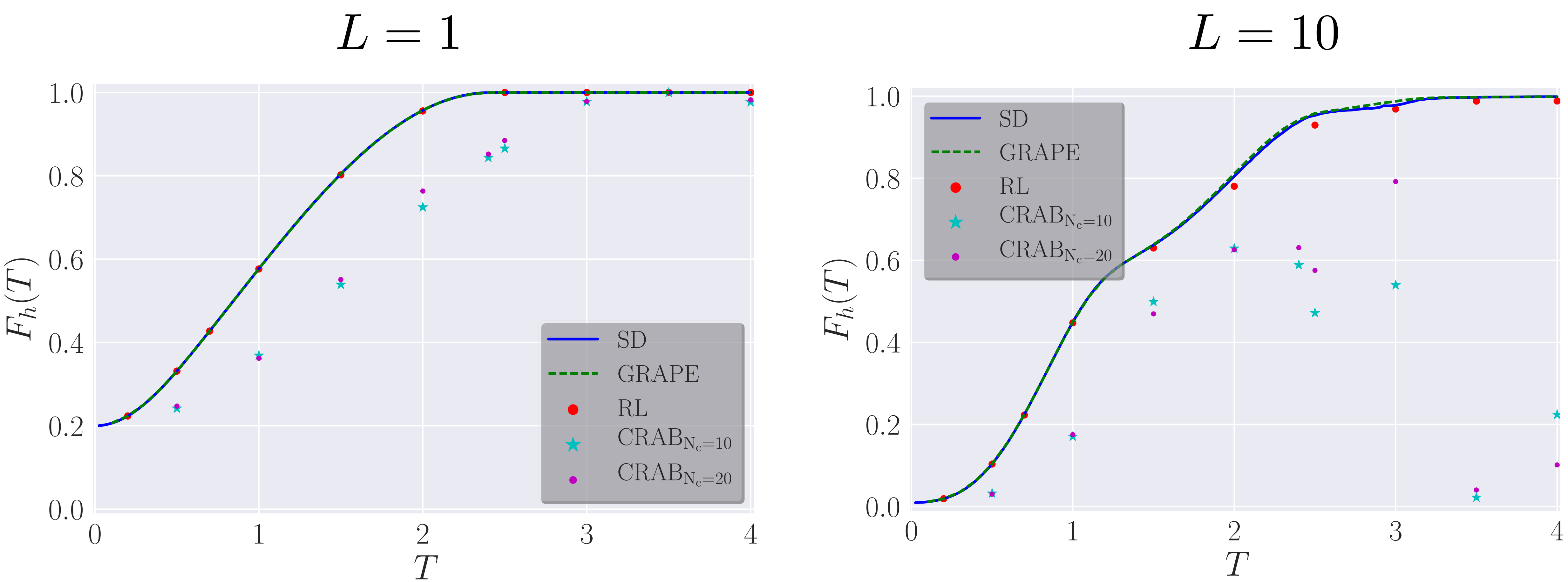
Kapitza Learning Curves



Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

RL vs OC



Marin Bukov