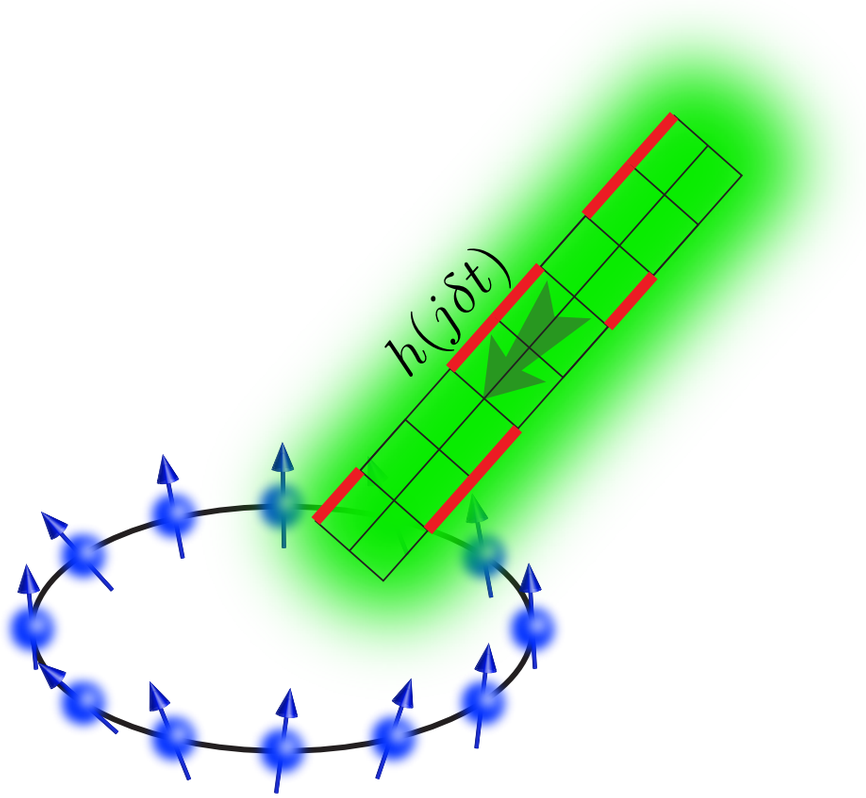
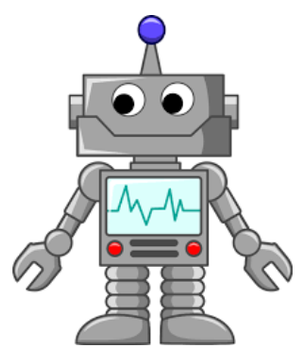
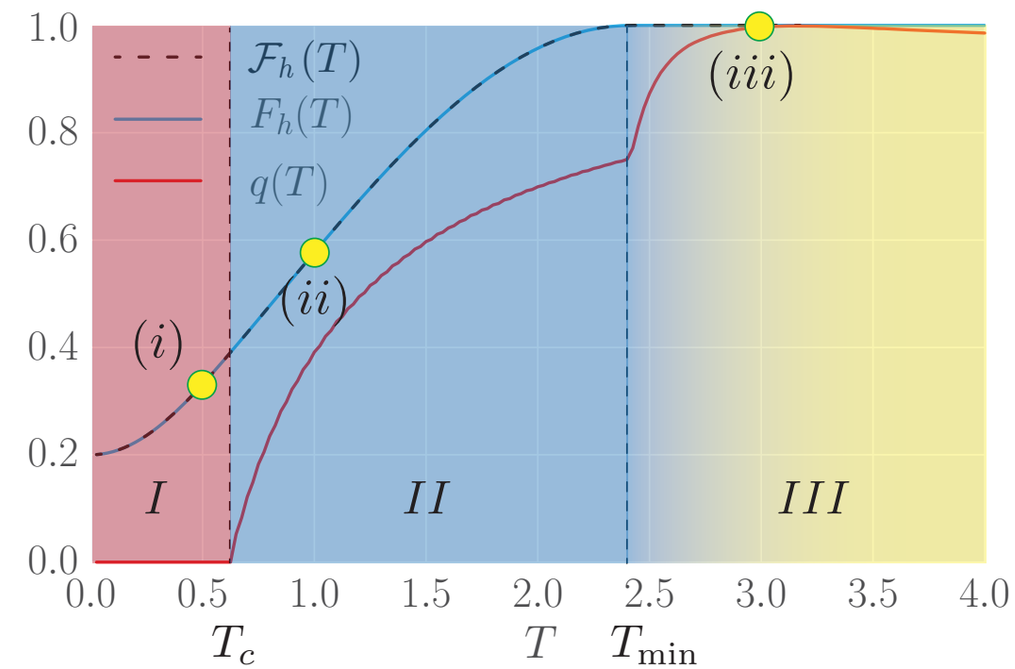
Reinforcement Learning in Different Phases of Quantum Control



nonintegrable
many-body spin chain





phase transitions
in the control landscape

PRX 8 031086 (2018)
PRA 97, 052114 (2018)
PRL 122, 020601 (2019)

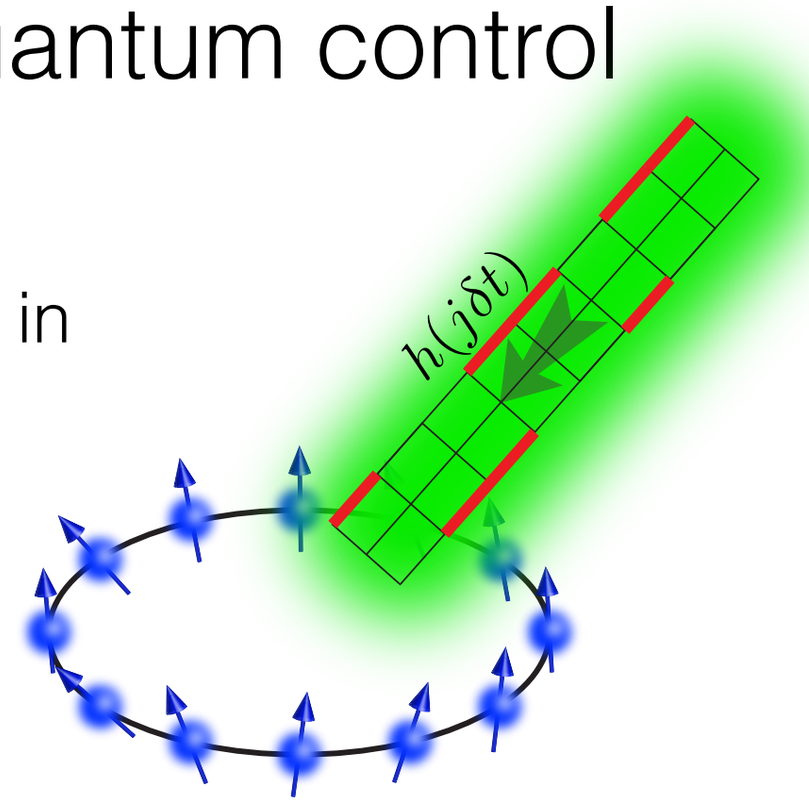
Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

in this talk:
Reinforcement Learning (RL) for quantum control

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

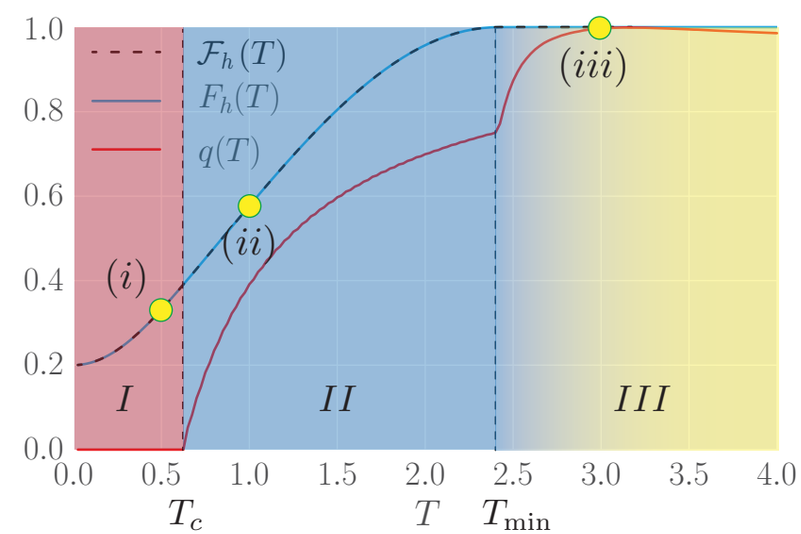
- RL and quantum control
- variational theory for optimal protocols



MB et al, PRX 8 031086 (2018)

→ **Phase transitions in the control landscape:**

- how "hard" is it for the RL agent to Learn?



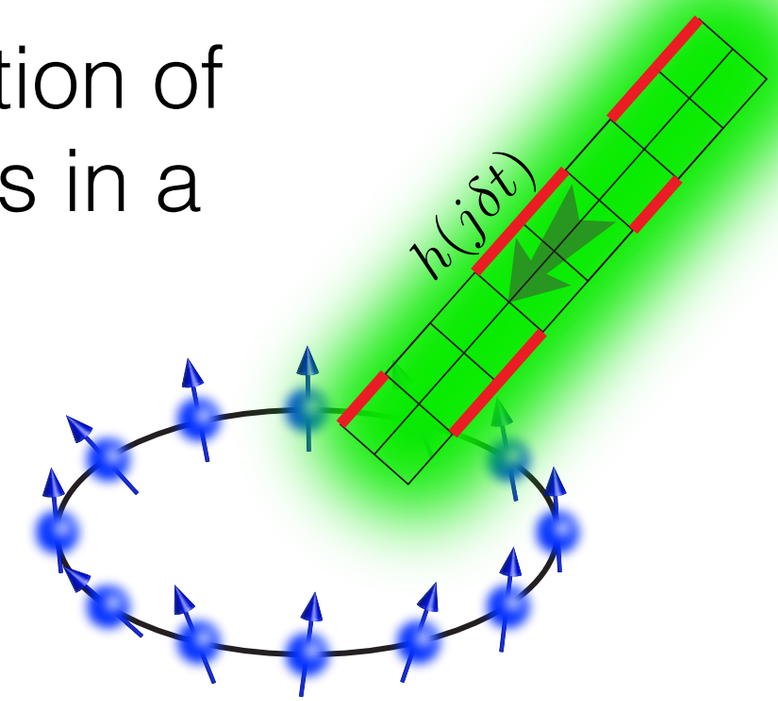
PRA 97, 052114 (2018)
PRL 122, 020601 (2019)

Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

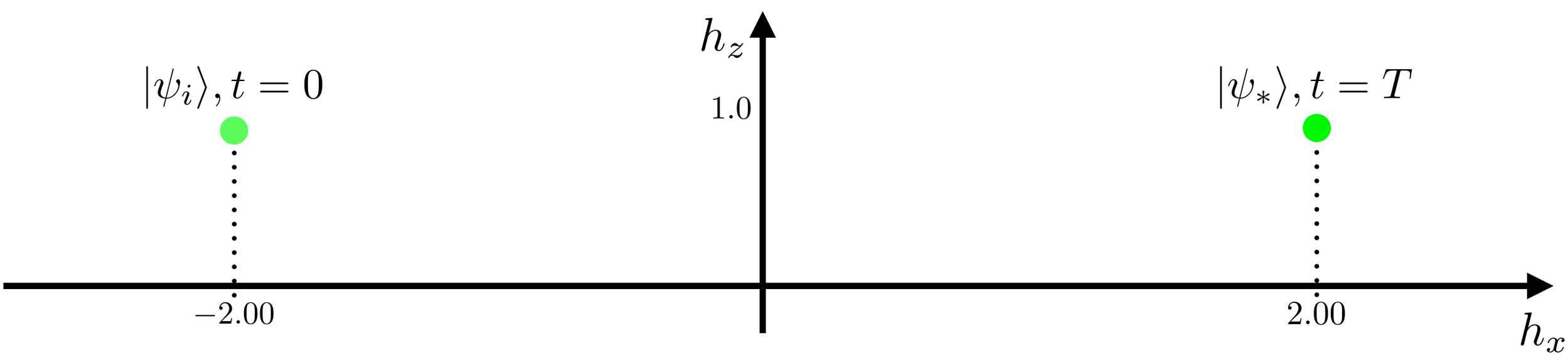
Example:

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:

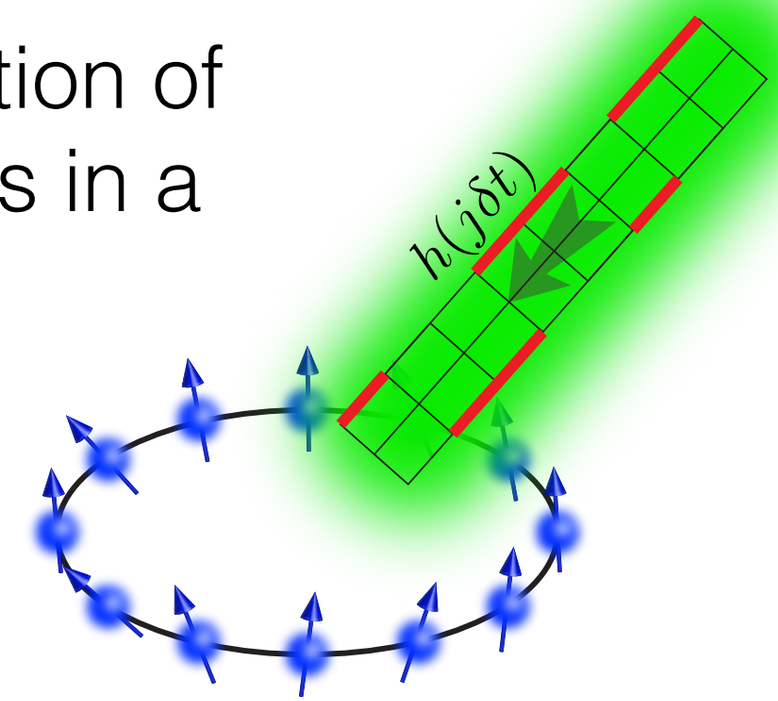


Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

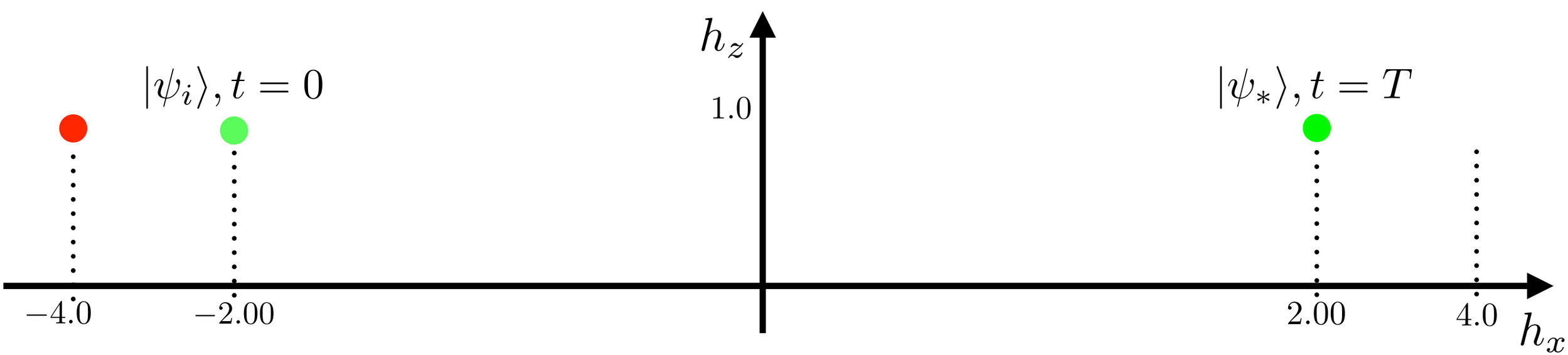
Example:

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:

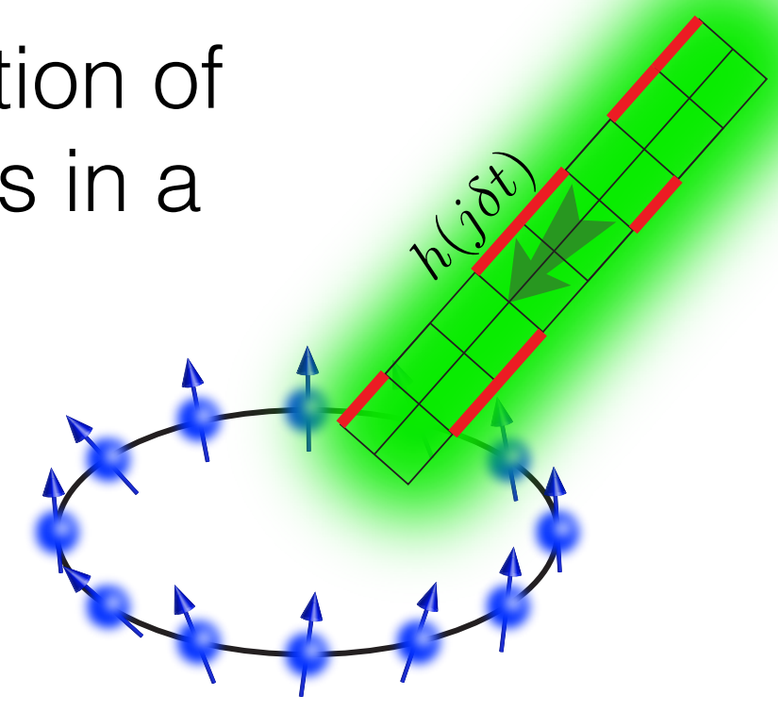


Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

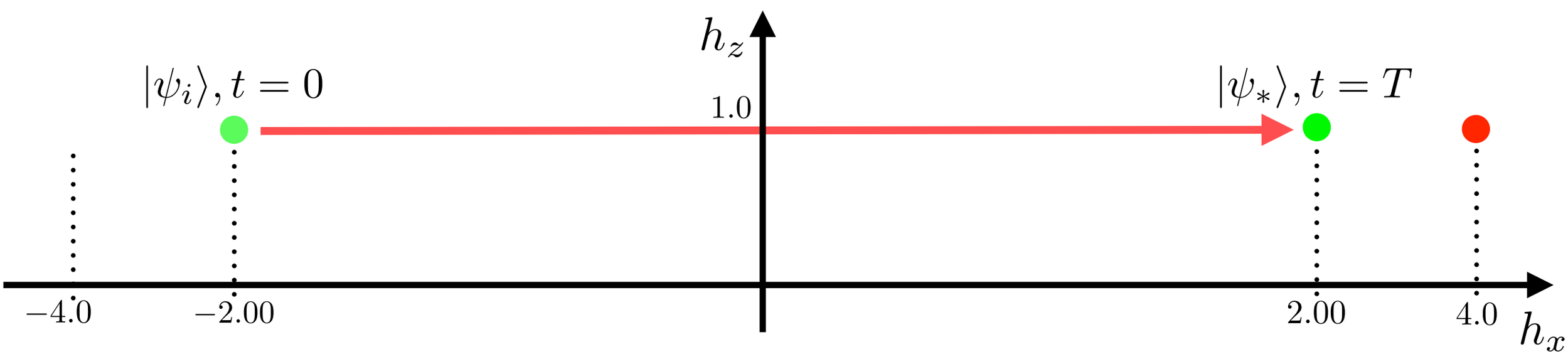
Example:

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:

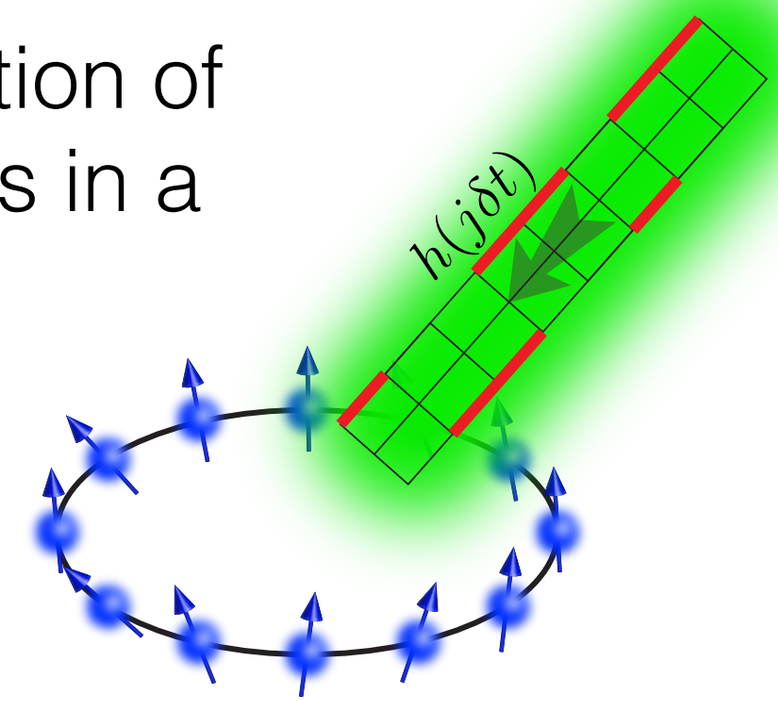


Marin Bukov

Berkeley
UNIVERSITY OF CALIFORNIA

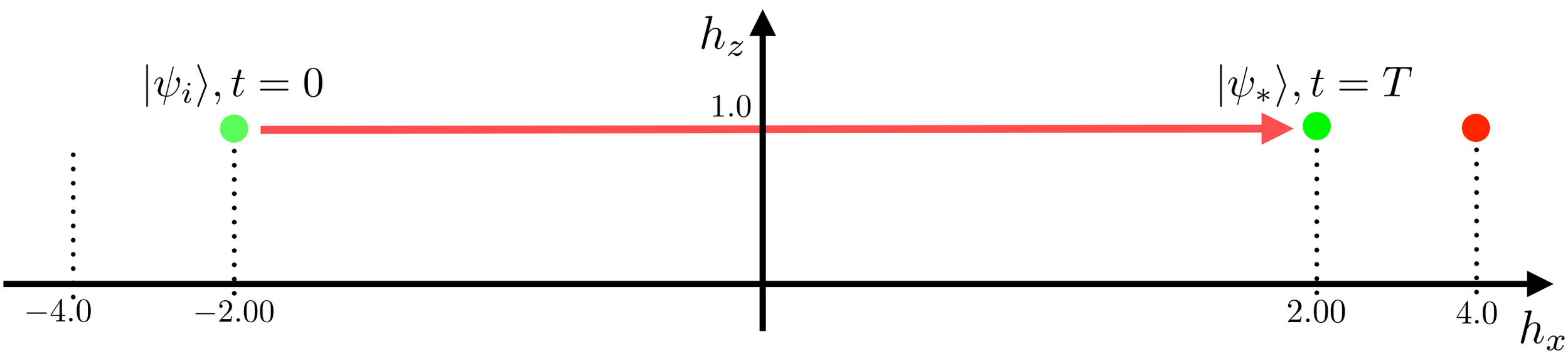
Example:

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:



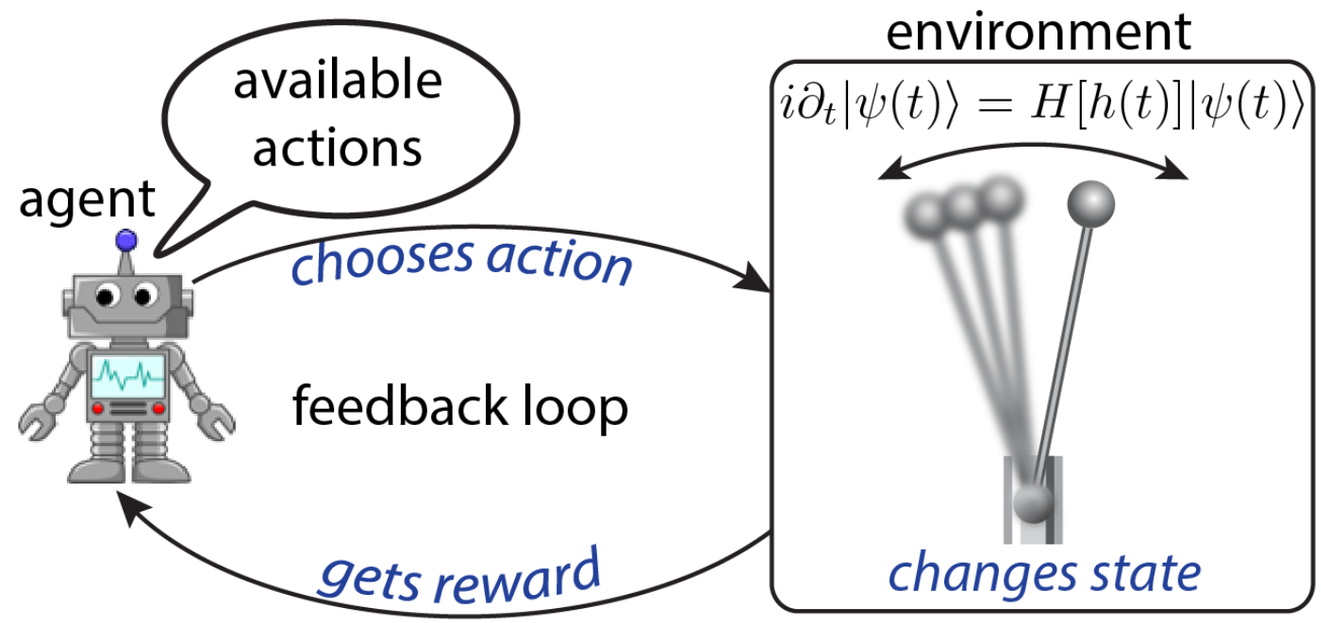
for example: $h_x(t) = [+4, +4, -4, +4, -4, -4, \ldots]$

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

Marin Bukov

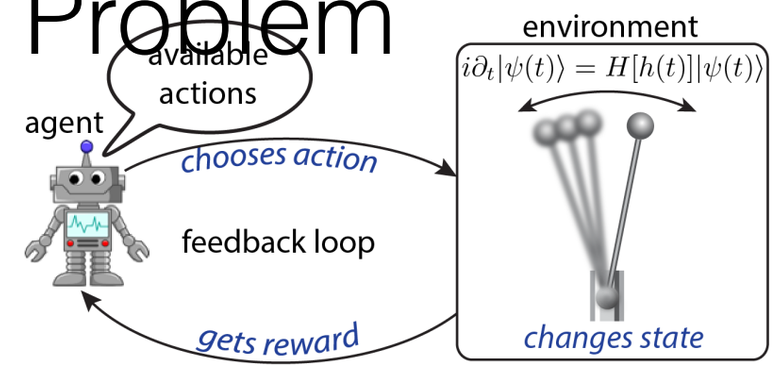
Quantum Control as an RL Problem

→ RL formalism



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

Quantum Control as an RL Problem

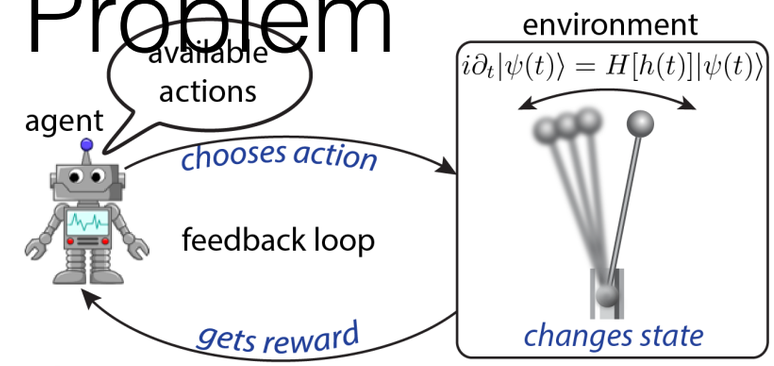


→ RL formalism

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

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

Quantum Control as an RL Problem



→ RL formalism

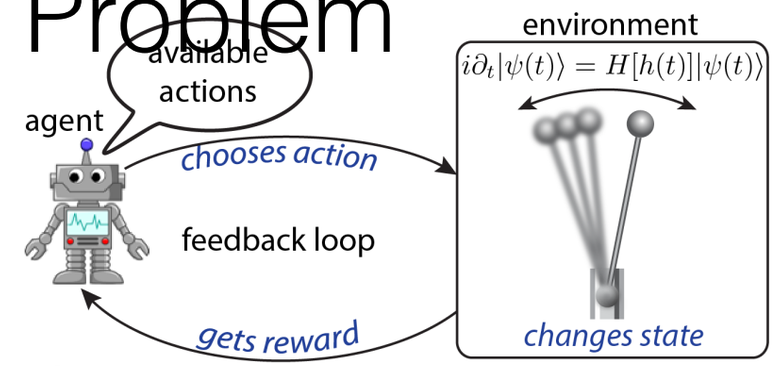
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

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

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

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

Quantum Control as an RL Problem



→ RL formalism

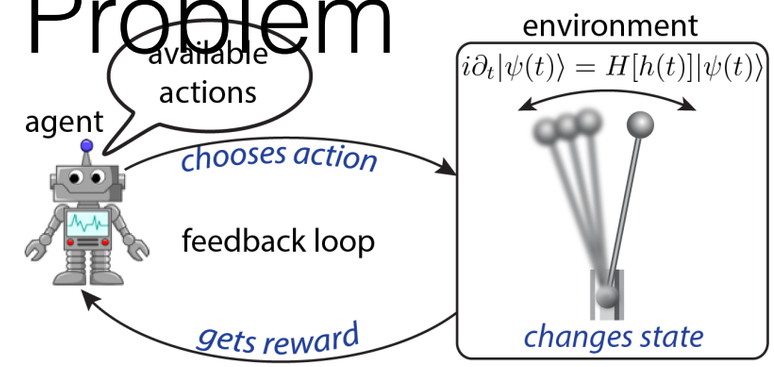
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

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

- 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

Quantum Control as an RL Problem



→ RL formalism

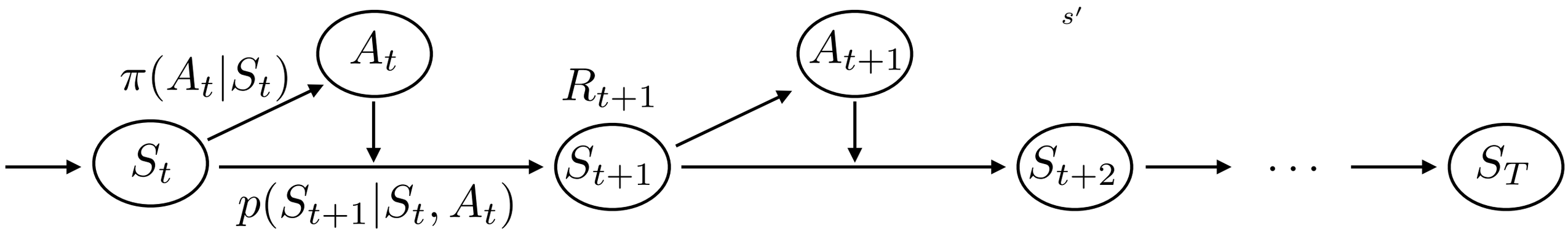
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

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

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

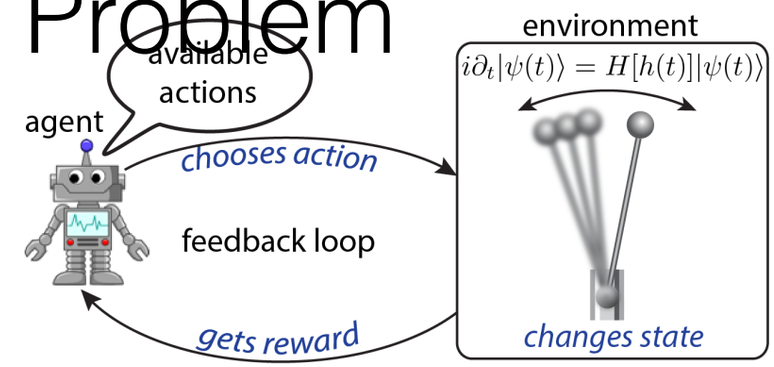
→ RL as Markov decision process

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



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

Quantum Control as an RL Problem



→ RL formalism

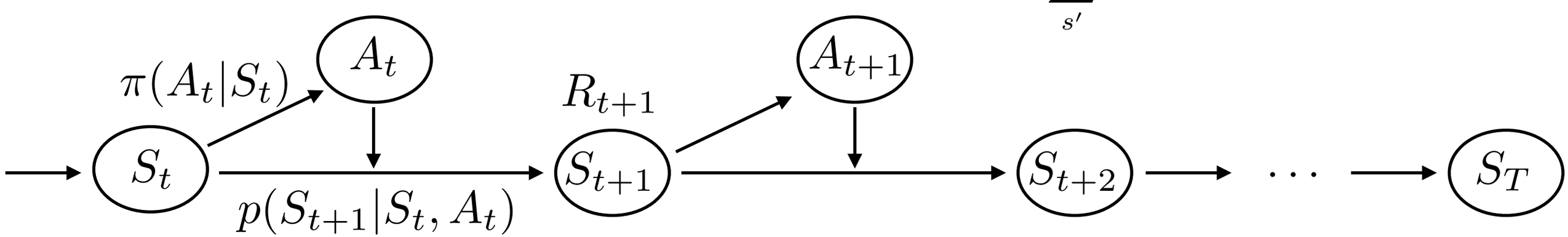
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

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

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

→ RL as Markov decision process

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

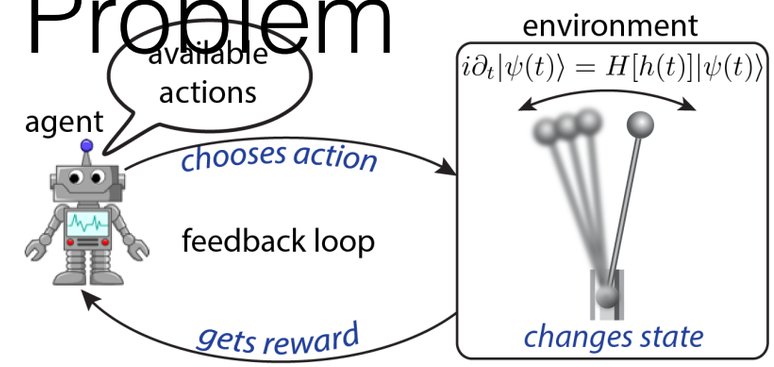


→ RL **objective**: maximize total *expected return* from step t onwards

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

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

Quantum Control as an RL Problem



→ RL formalism

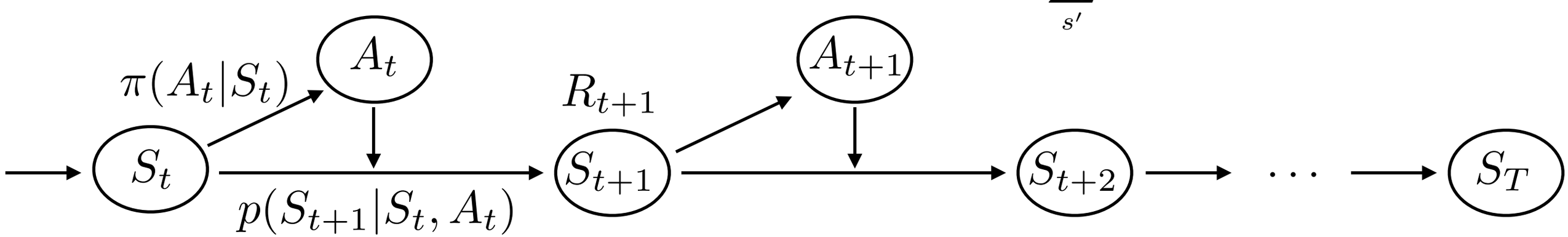
- action space $\mathcal{A} = \{+4, -4\}$
- state space $\mathcal{S}$ all possible strings of $\{+4, -4\}$

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

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

→ 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

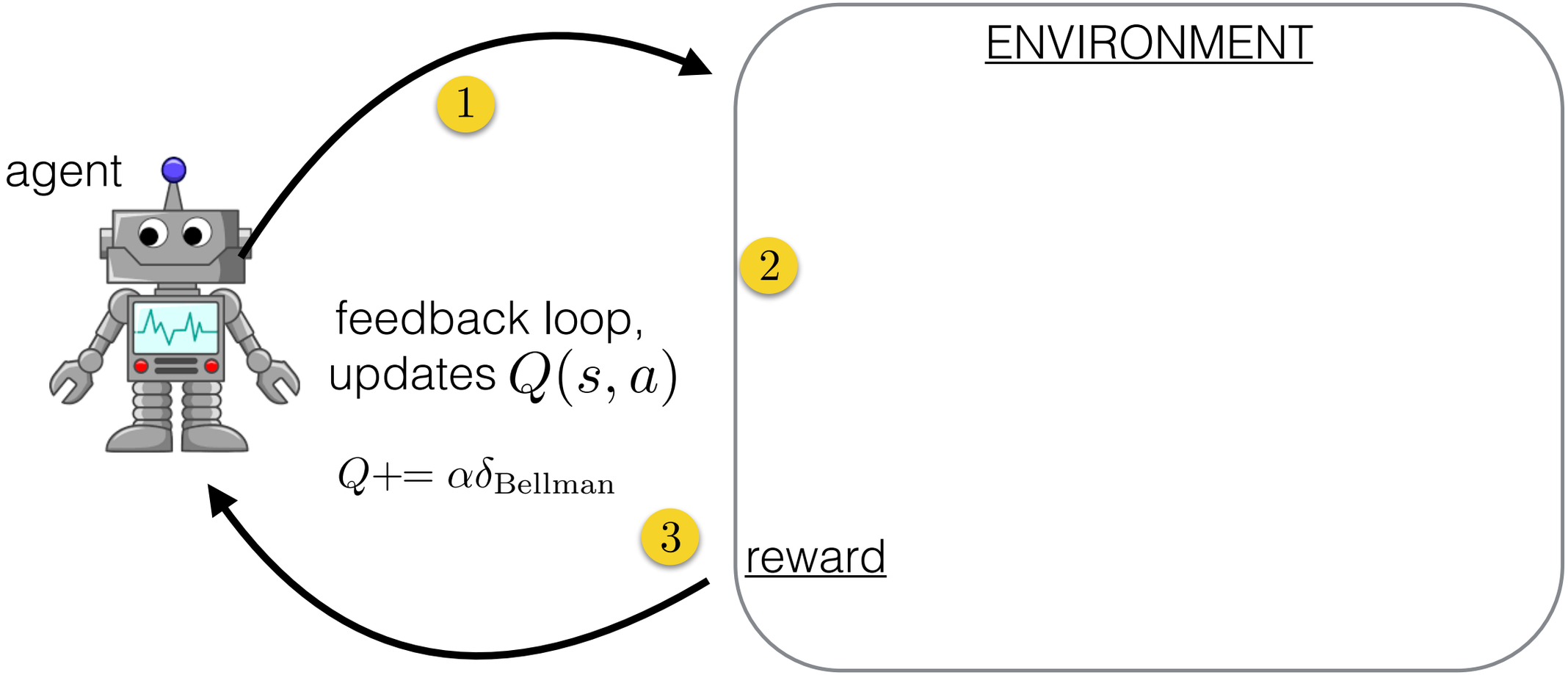
$$Q(s,a) = \mathbb{E}_{a\sim\pi(a|s)}[R_{t+1} + \cdots + R_{t_f}|S_t = s, A_t = a]$$

→ Watkins Q-Learning (basic idea):

$$Q_{\text{new}}(S_t, A_t) = Q_{\text{old}}(S_t, A_t) + \alpha[R_{t+1} + \max_a Q_{\text{old}}(S_{t+1}, a) - Q_{\text{old}}(S_t, A_t)]$$

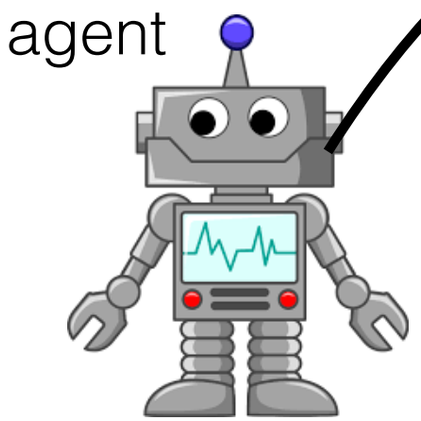
Sutton and Barto, *Reinforcement Learning: an Introduction*, MIT press

RL Applied to Quantum State Preparation



Marin Bukov

RL Applied to Quantum State Preparation



1

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

$Q \mathrel{+}= \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

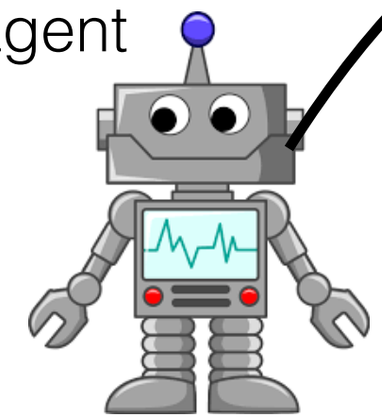
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0, T]$$

reward

Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$$Q += \alpha \delta_{\text{Bellman}}$$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

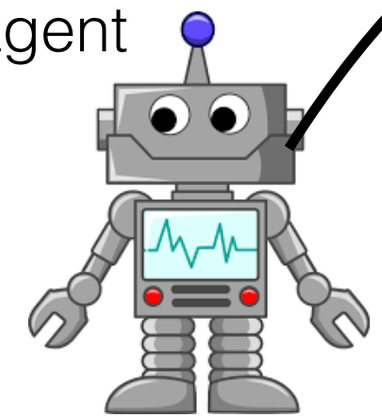
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0,T]$$

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

Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$$Q += \alpha \delta_{\text{Bellman}}$$

3

ENVIRONMENT

2

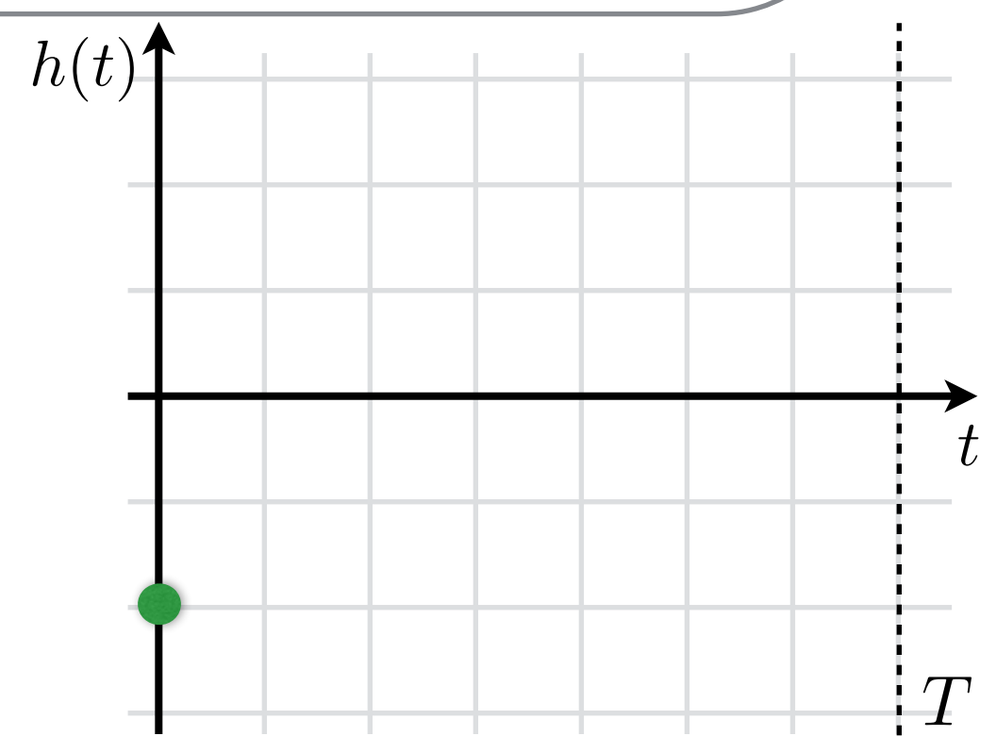
$$H(t) = H_0 + H_{\text{ctrl}}(t)$$

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

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

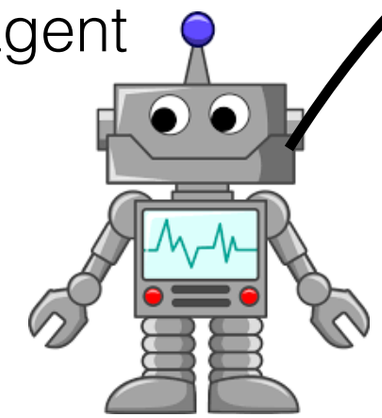
reward $r = \begin{cases} 0 & , 0 \le t < T \\ |\langle \psi_* | \psi(t=T)\rangle|^2, & t = T \end{cases}$



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$Q += \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

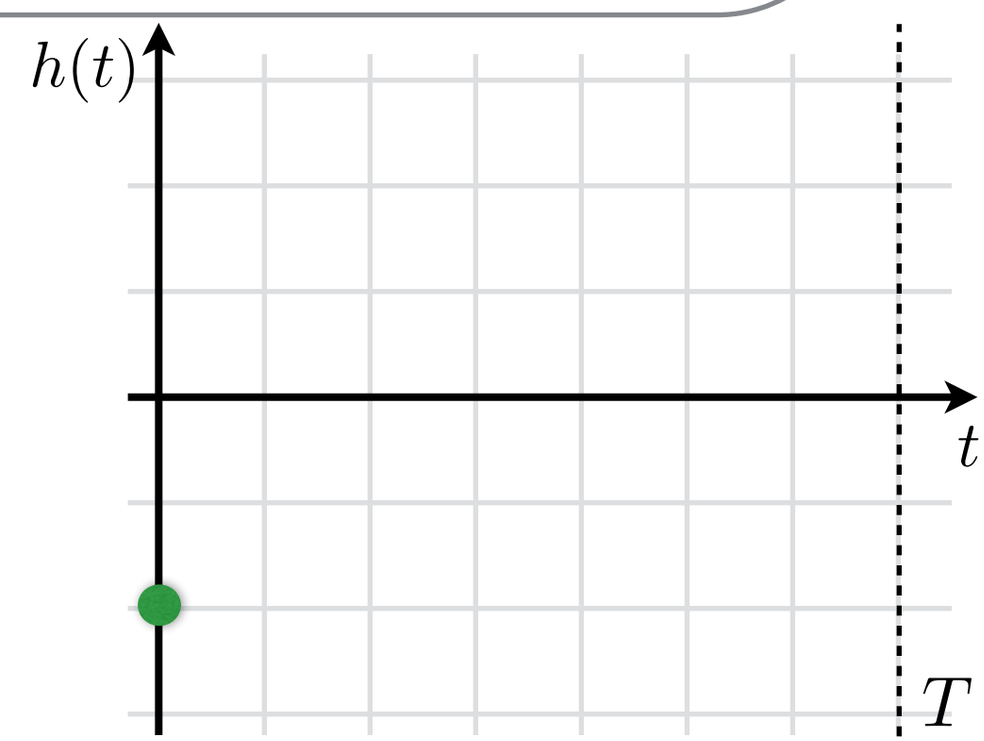
$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

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

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

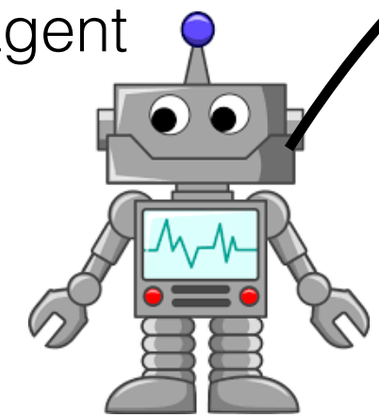
1 start from state $s_0 = [h(0)] = [-4]$



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$Q += \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

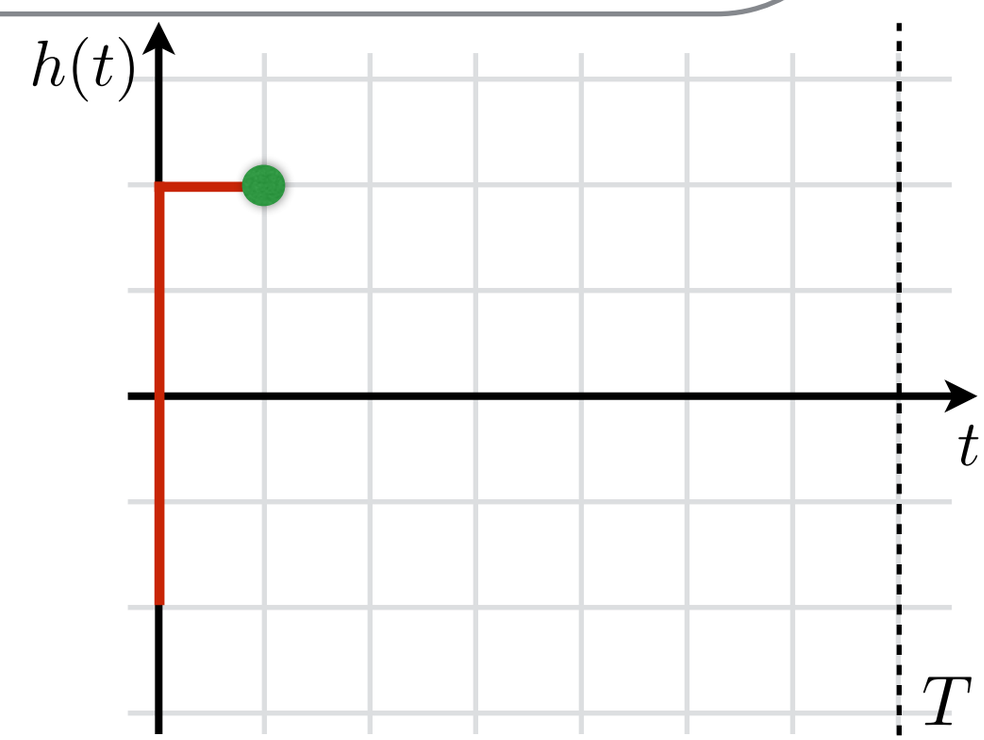
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0,T]$$

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

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

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

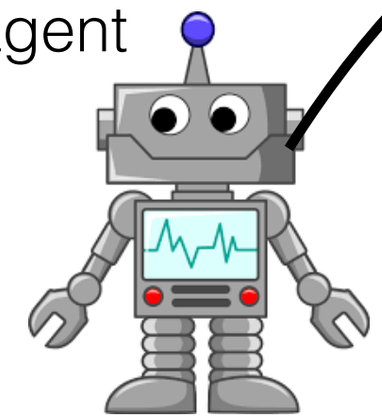
go to state $s_1 = [h(0), h(\delta t)] = [-4, +4]$



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$Q += \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

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

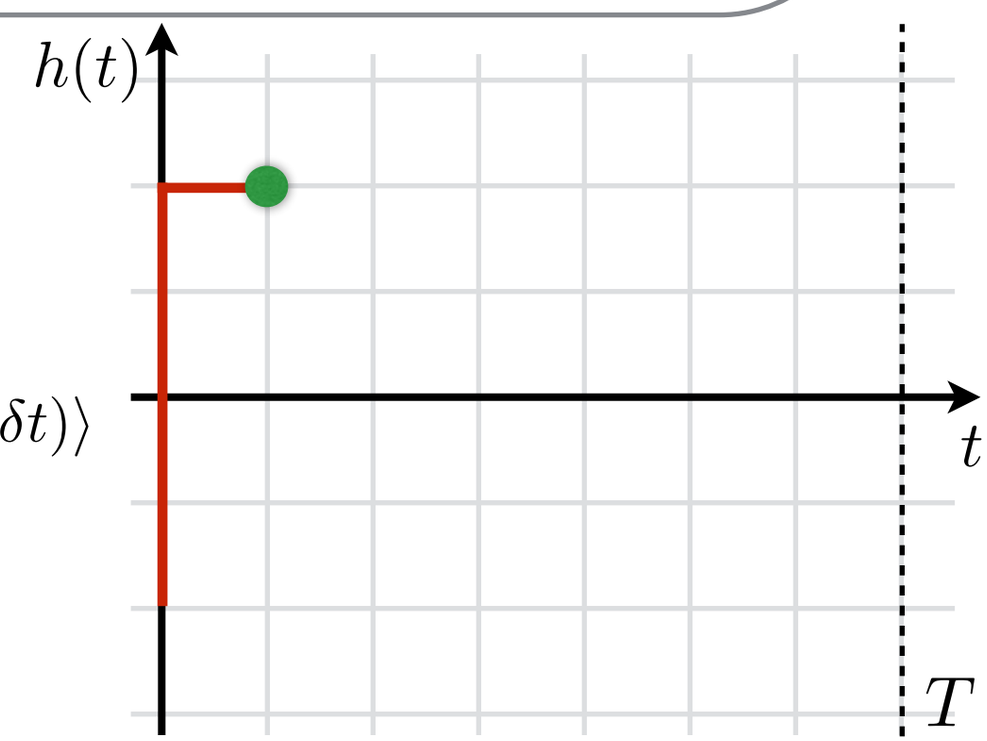
reward $r = \begin{cases} 0 & , 0 \le t < T \\ |\langle \psi_* | \psi(t=T)\rangle|^2, & t = T \end{cases}$

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

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

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

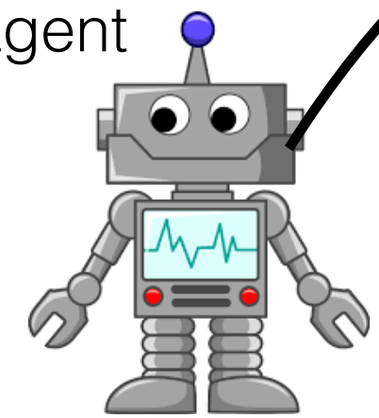
2 solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$Q \mathrel{+}= \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

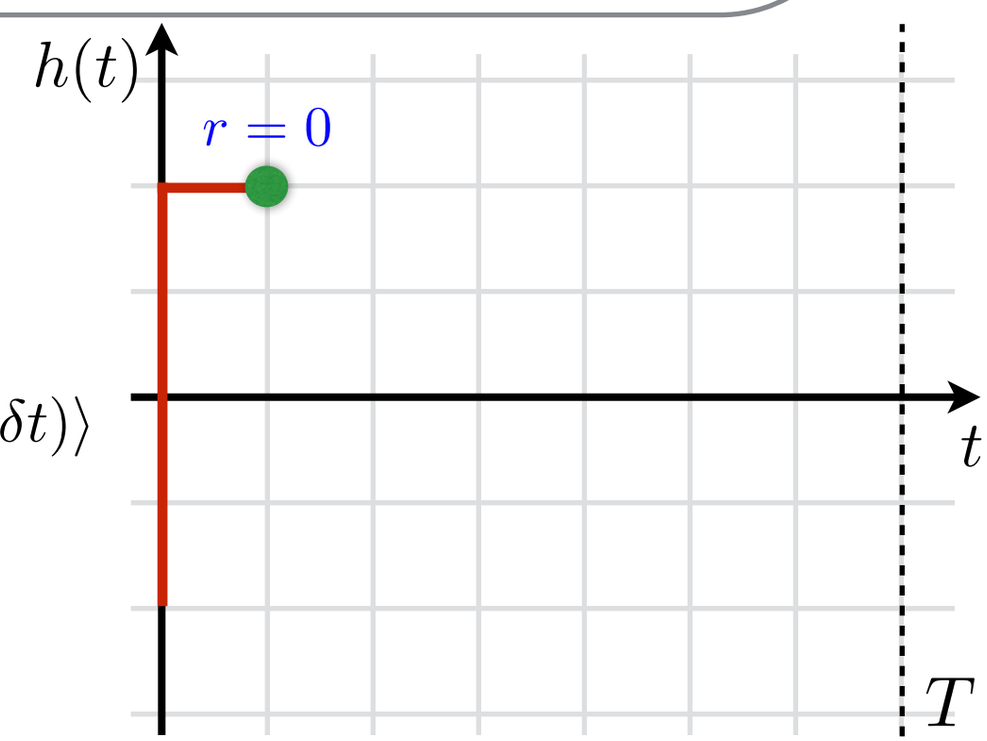
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0,T]$$

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

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 $|\psi(\delta t)\rangle$

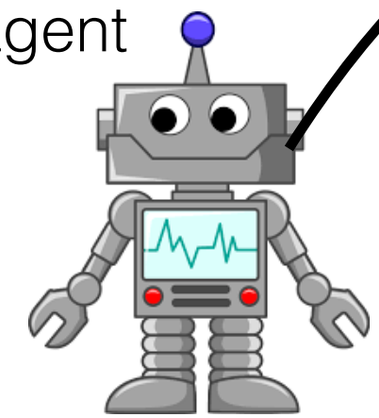
3 calculate reward r
and use it to update $Q(s,a)$
which in turn is used to choose subsequent actions



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$Q \mathrel{+}= \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

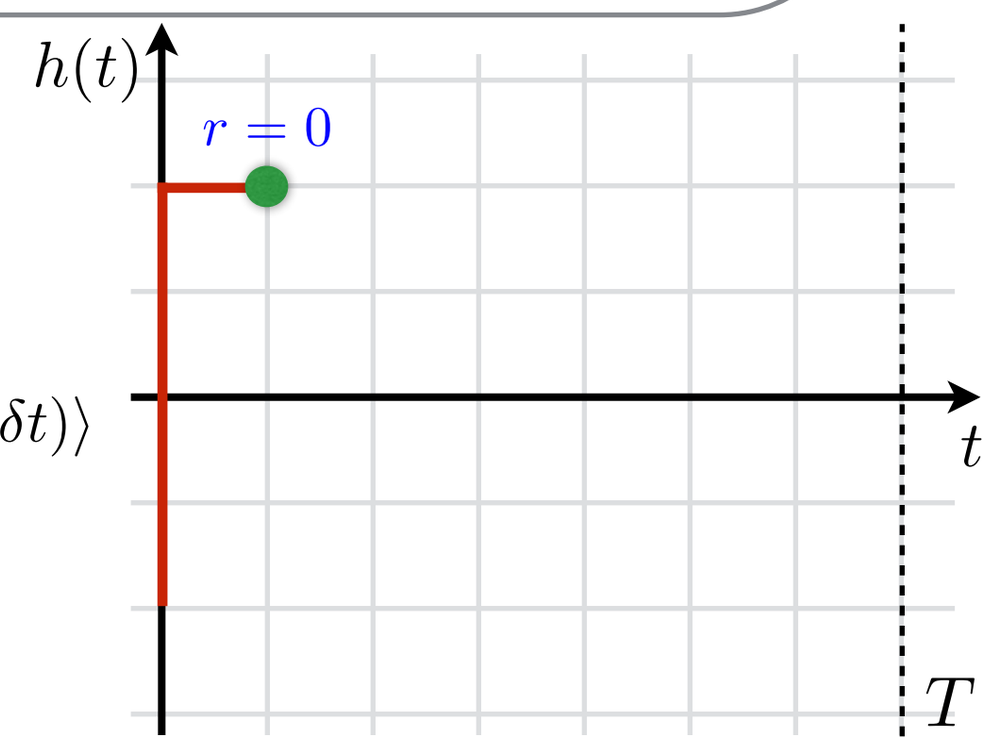
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0,T]$$

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

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 $|\psi(\delta t)\rangle$

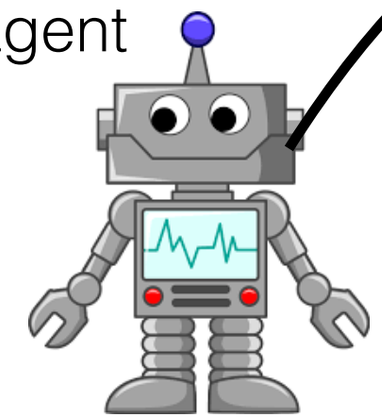
3 calculate reward r
and use it to update $Q(s,a)$
which in turn is used to choose subsequent actions



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$Q += \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

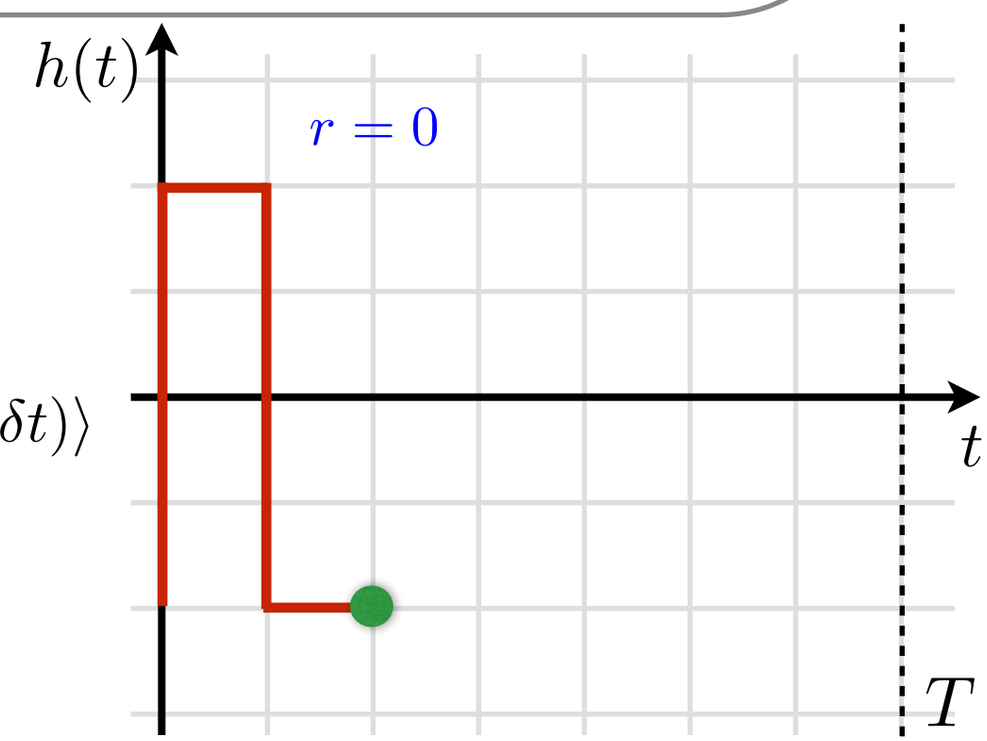
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0,T]$$

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

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 $|\psi(\delta t)\rangle$

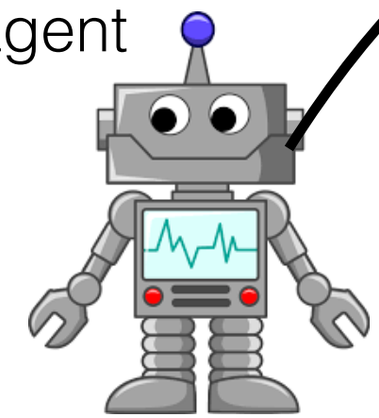
3 calculate reward r
and use it to update $Q(s,a)$
which in turn is used to choose subsequent actions



Marin Bukov

RL Applied to Quantum State Preparation

agent



1

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

$Q \mathrel{+}= \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

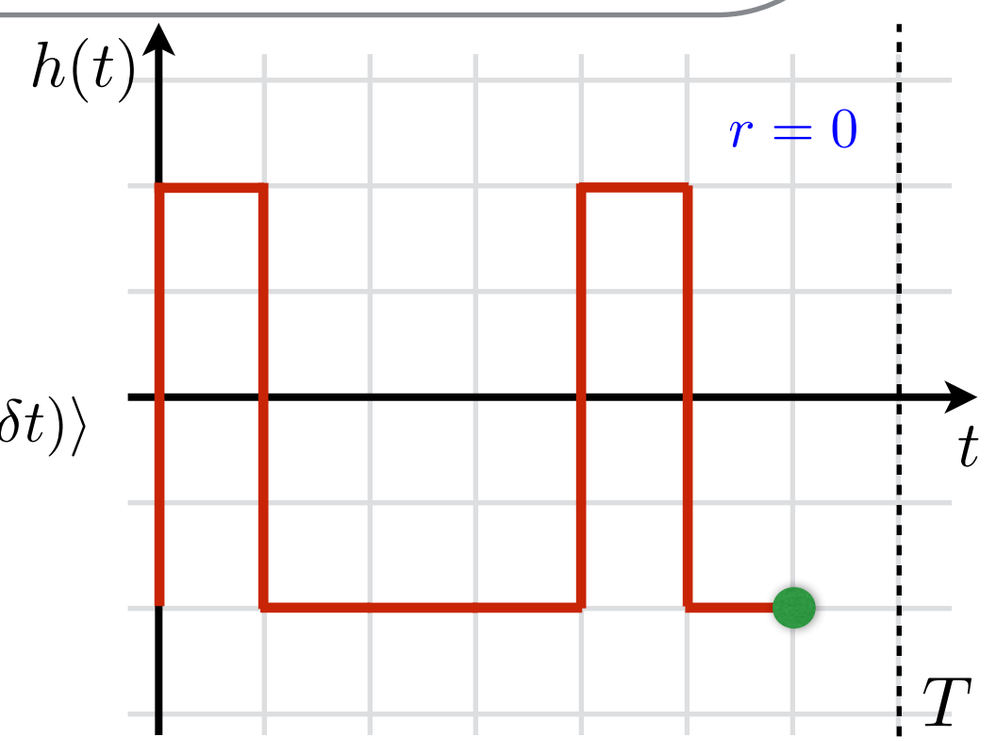
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0,T]$$

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

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

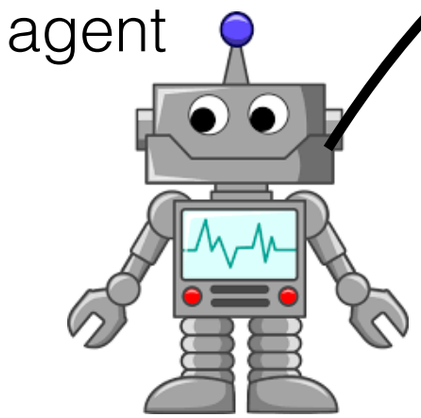
2 solve Schrödinger Eq. and obtain the QM state $|\psi(\delta t)\rangle$

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



Marin Bukov

RL Applied to Quantum State Preparation



1

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

$Q \mathrel{+}= \alpha \delta_{\text{Bellman}}$

3

ENVIRONMENT

2

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

$|\psi_i\rangle$: GS of H_0

$|\psi_*\rangle$: target state

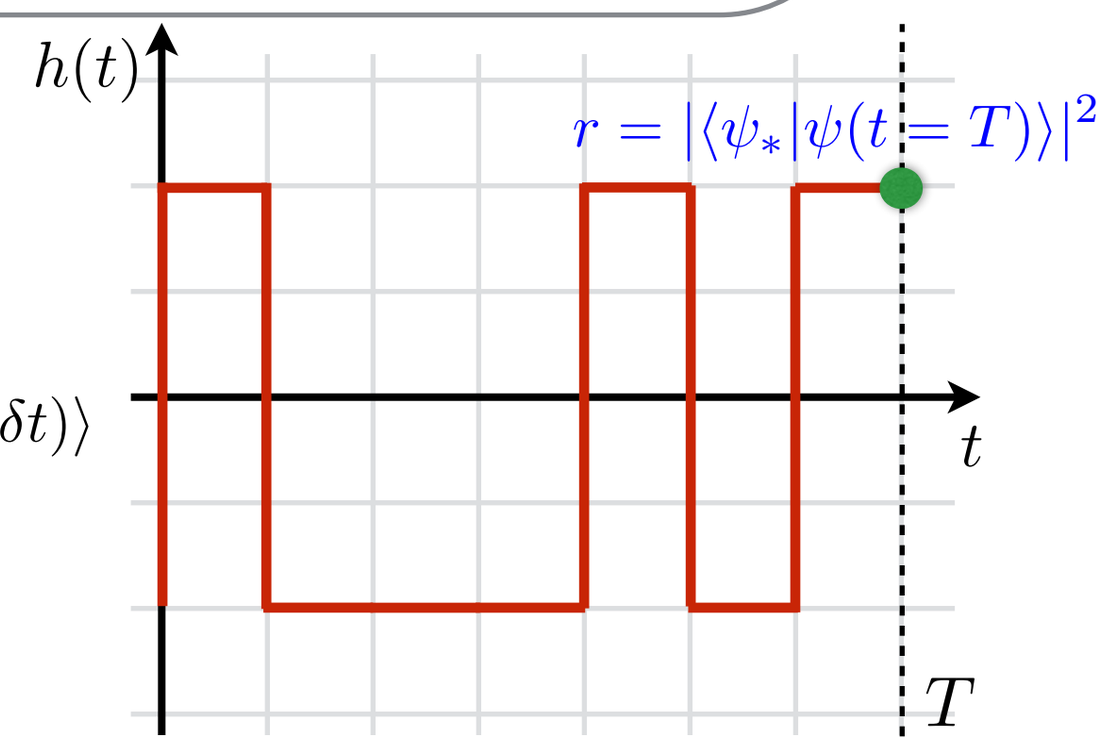
$$i\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle \; t \in [0,T]$$

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

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 interpolation: parametrize the Q-function:

$$Q(s,a) \to Q_\theta(s,a)$$

RL with Function Approximation

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

- can we estimate values of not yet encountered states?

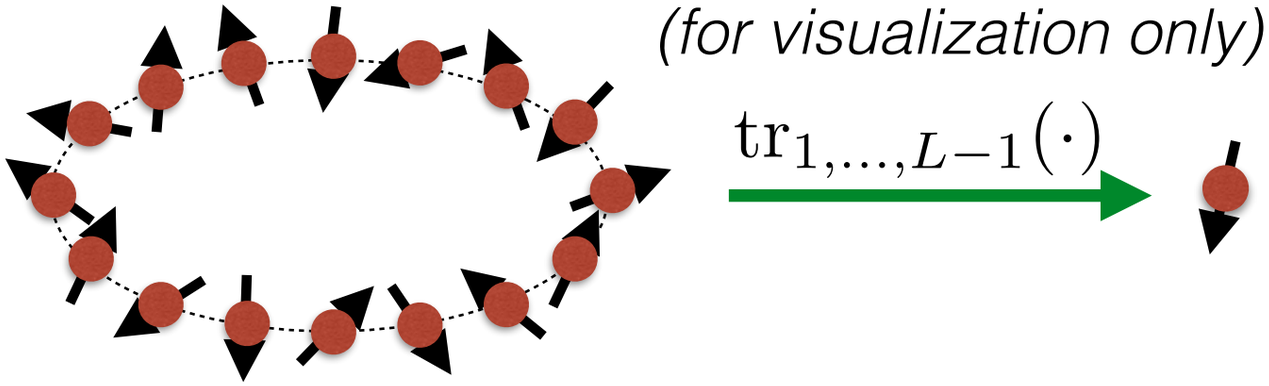
→ YES, via interpolation: parametrize the Q-function:

$$Q(s,a) \to Q_\theta(s,a)$$

- typical approach: use deep neural network (Deep RL)
- caveats:
  1) value-iteration RL algorithms have convergence guarantees only for *linear function approximators*
  2) may learn wrong Q-values (even if convergent)
- lots of empirical tricks to combine Deep Learning and RL

Learning Many-Body Quantum Control

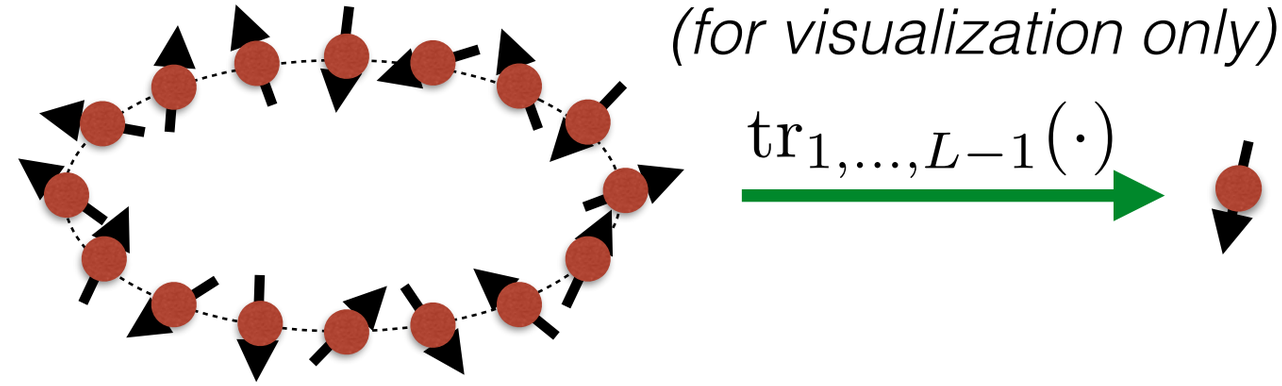
$$H(t) = -\sum_{j=1}^{L} S^z_{j+1} S^z_j + S^z_j + h_x(t) S^x_j$$



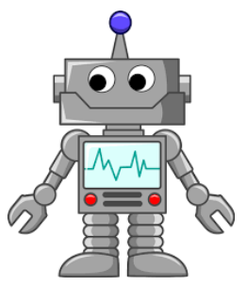
PRX 8 031086 (2018)

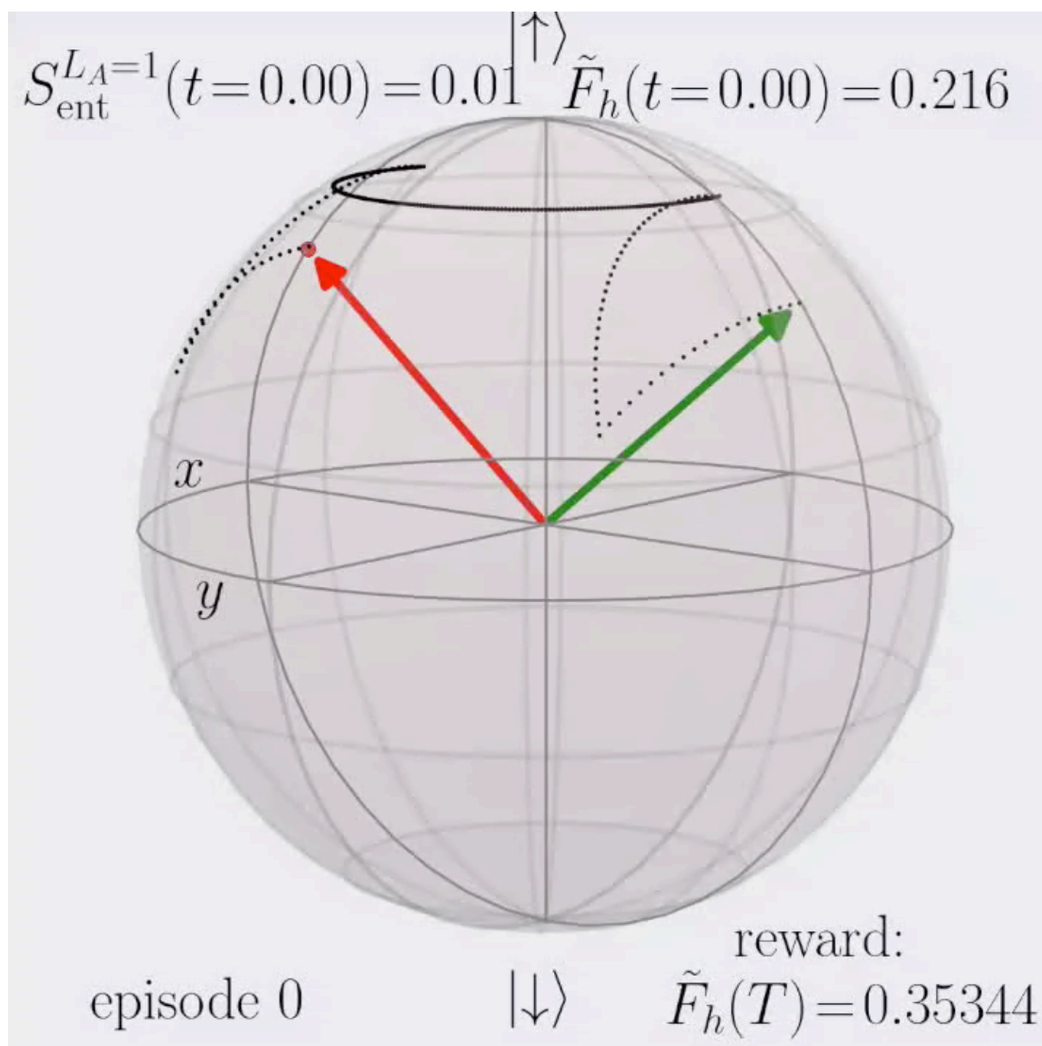
Learning Many-Body Quantum Control

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



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





Bloch sphere

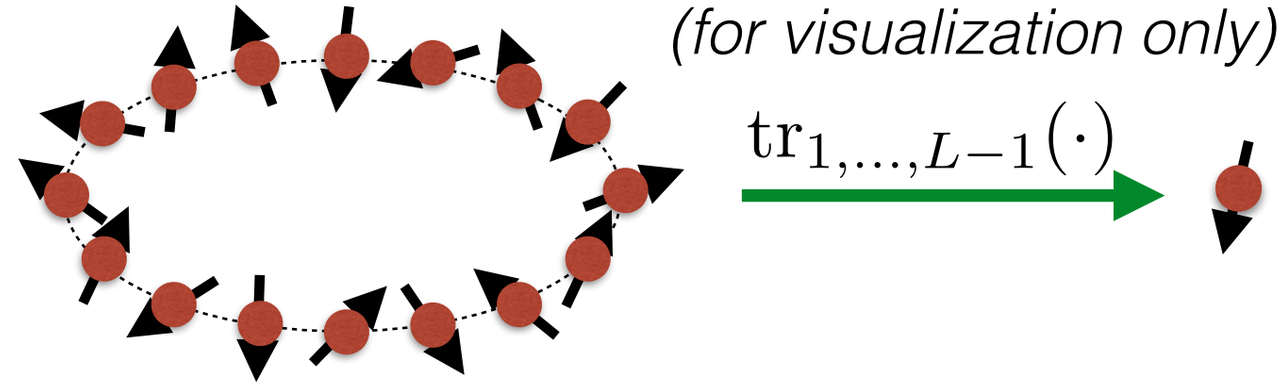
$$\tilde{F}_h = -\frac{1}{L} \log F_h$$

Marin Bukov

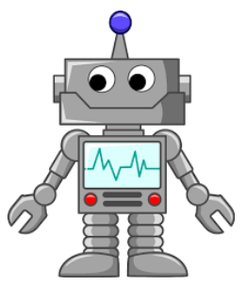
PRX 8 031086 (2018)

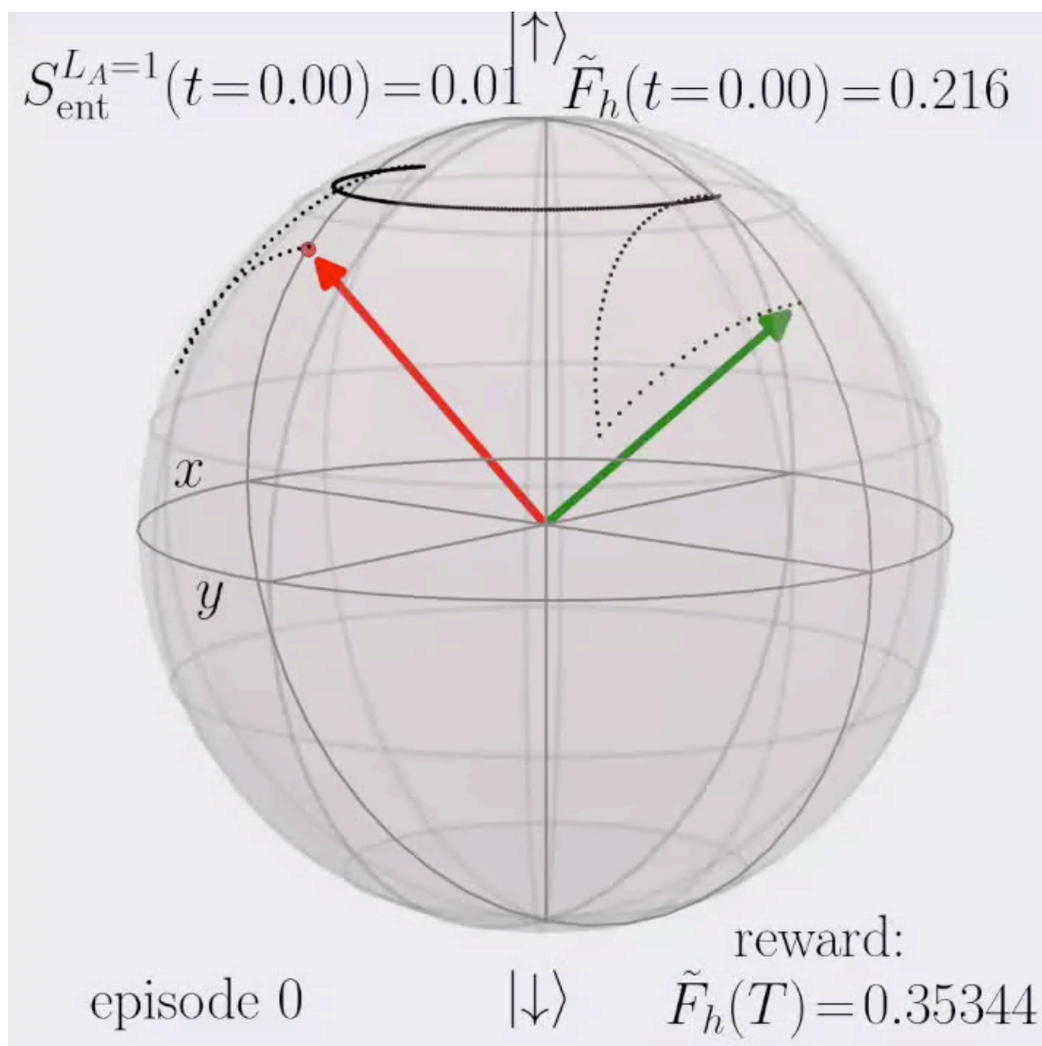
Learning Many-Body Quantum Control

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



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





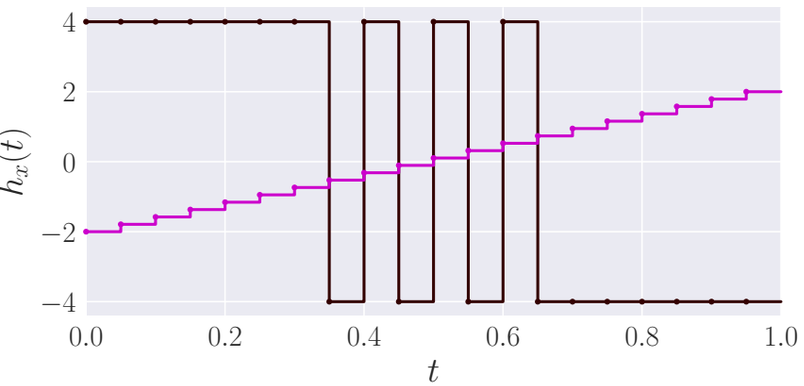
Bloch sphere

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

Marin Bukov

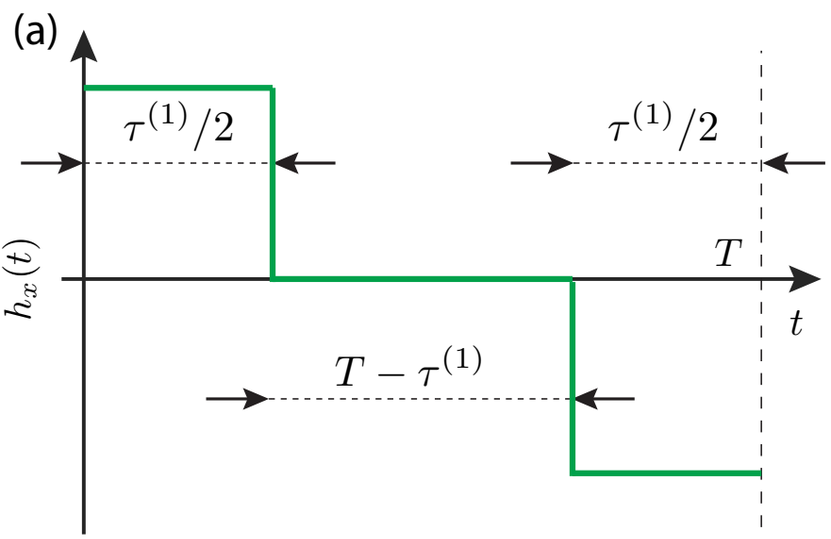
PRX 8 031086 (2018)

What do we Learn from the RL Agent?

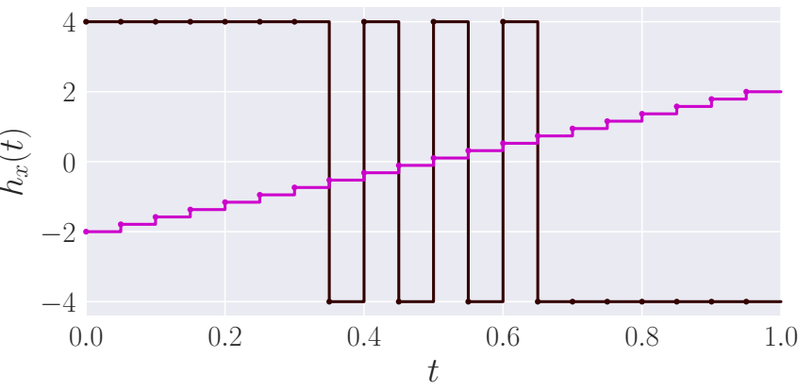


$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

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

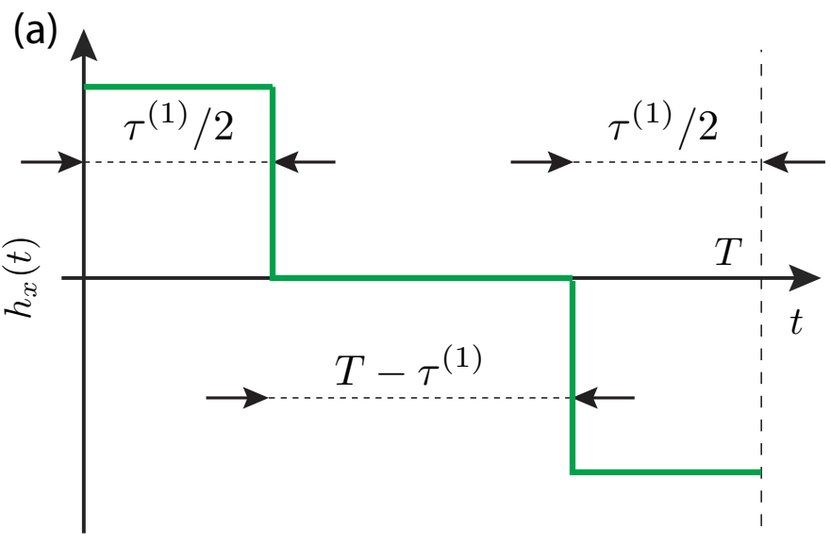


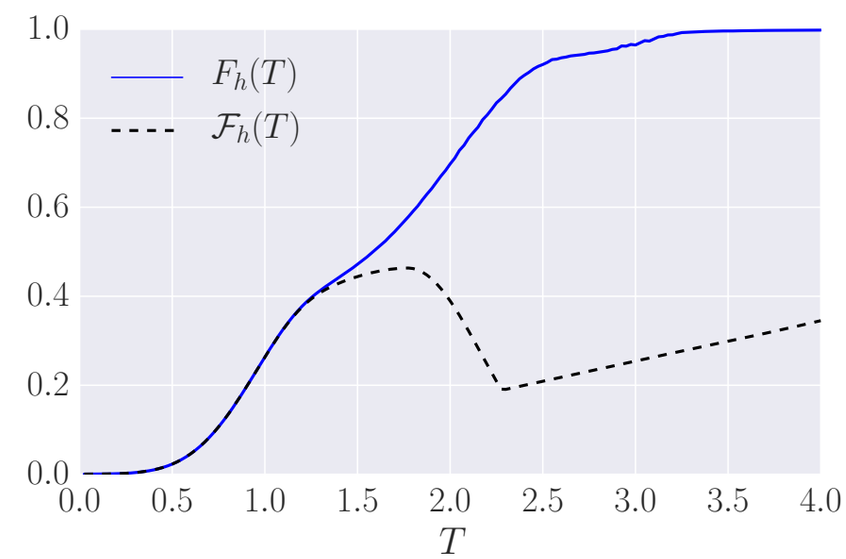
What do we Learn from the RL Agent?



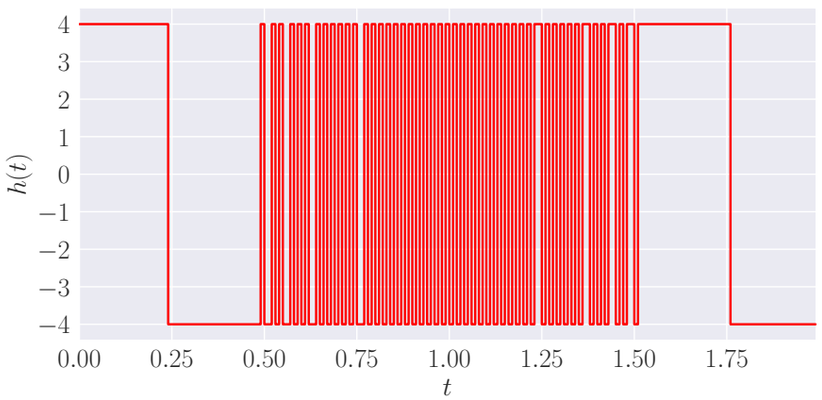
$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

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



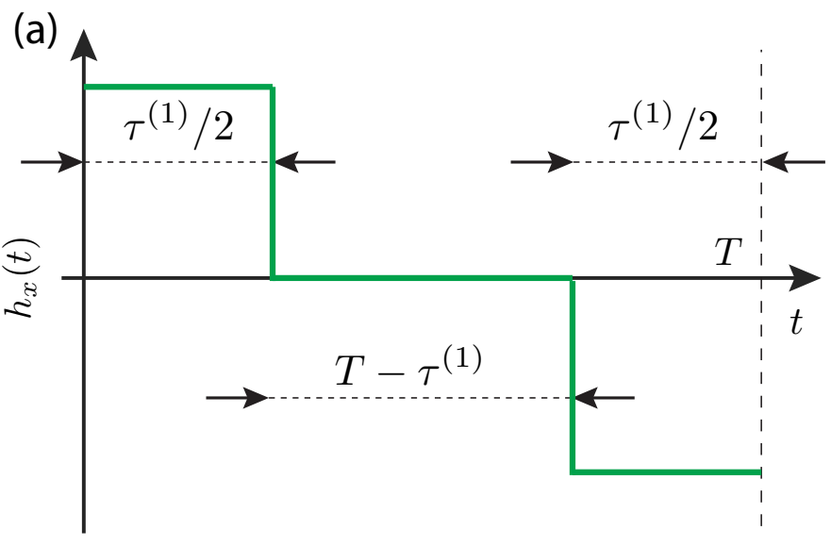


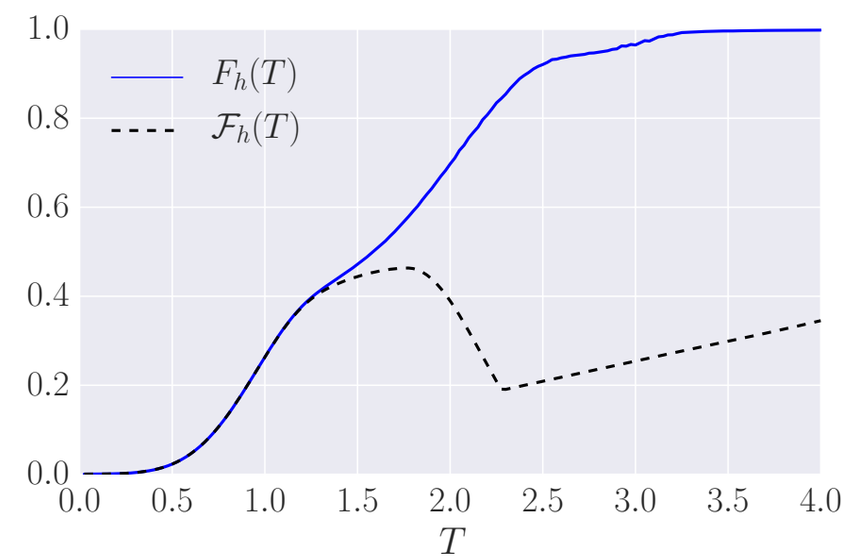
What do we Learn from the RL Agent?

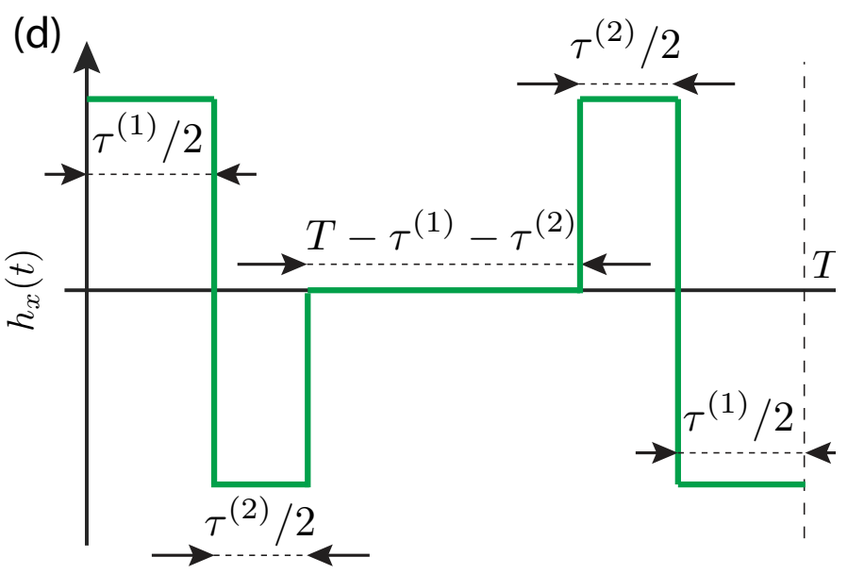


$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

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

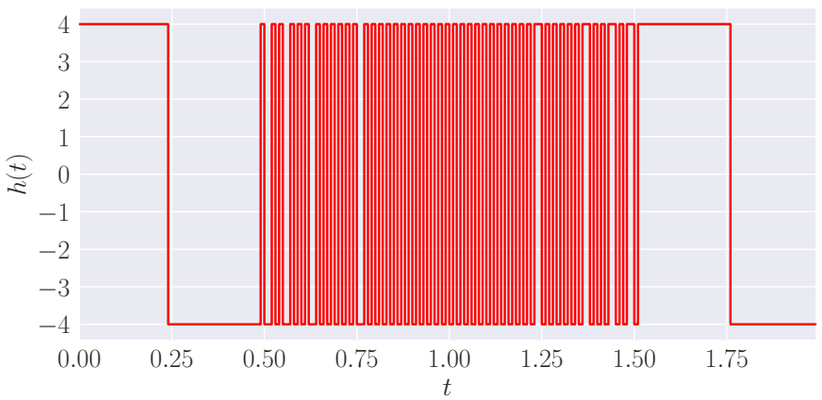






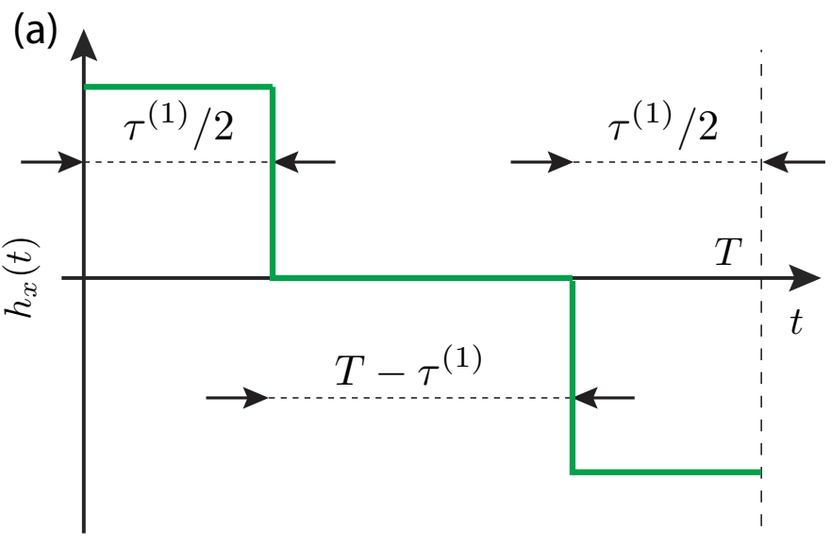
PRX 8 031086 (2018)

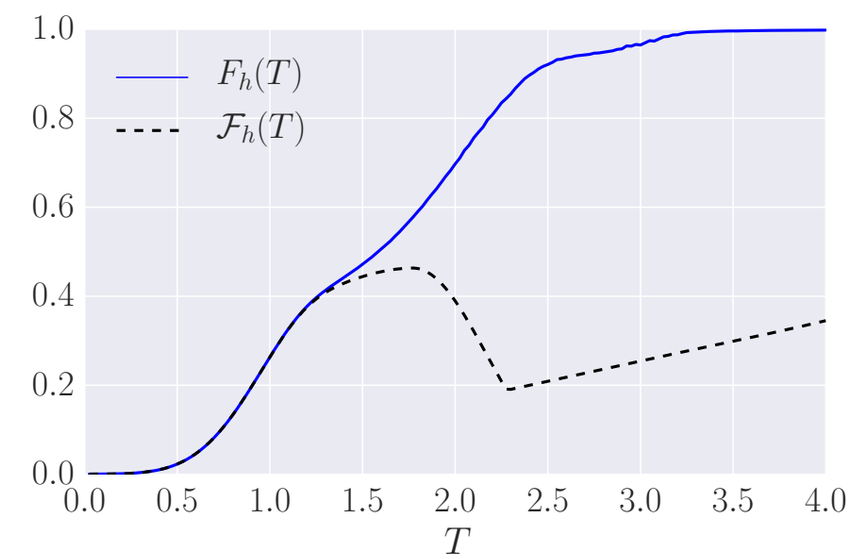
What do we Learn from the RL Agent?

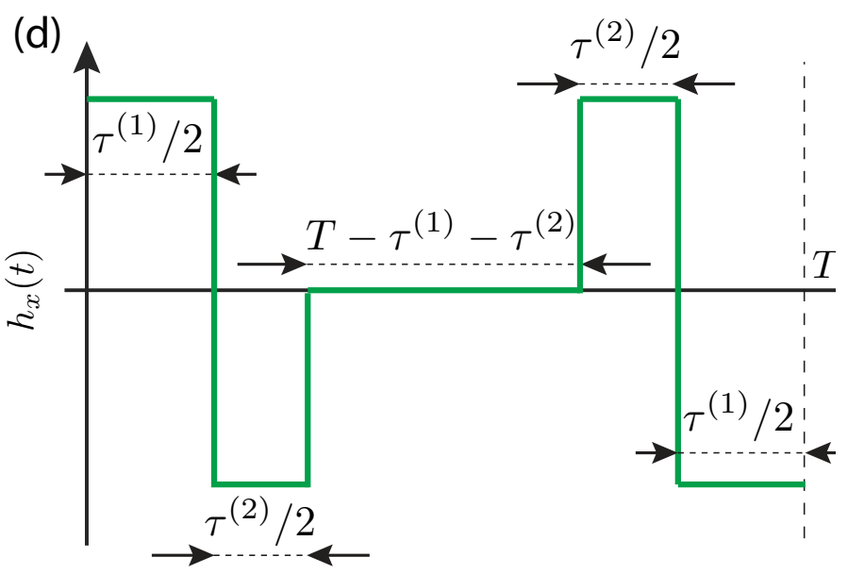


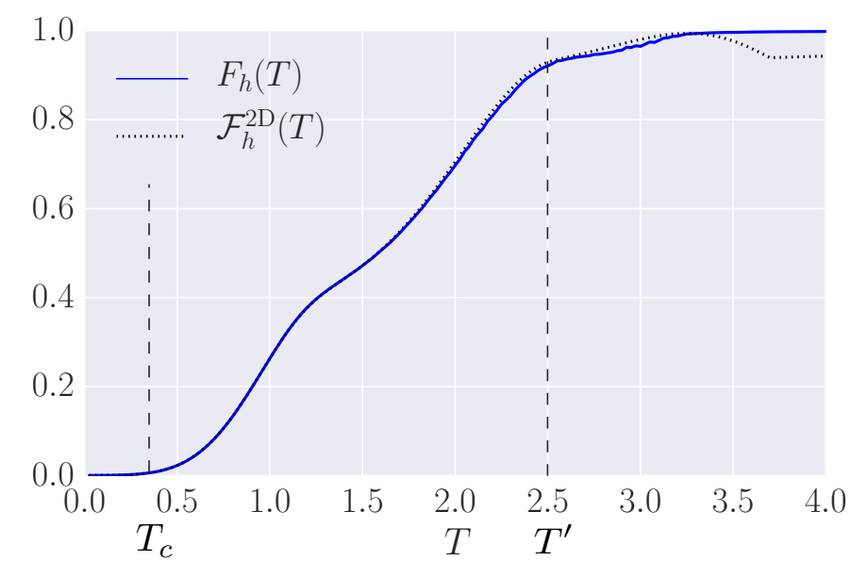
$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

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









Marin Bukov

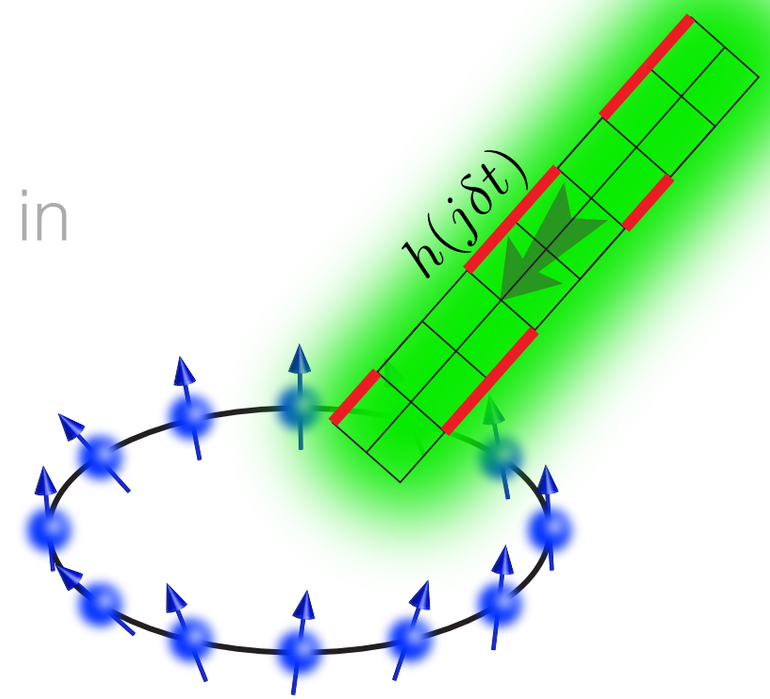
PRX 8 031086 (2018)

Berkeley
UNIVERSITY OF CALIFORNIA

Reinforcement Learning (RL) for quantum control

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

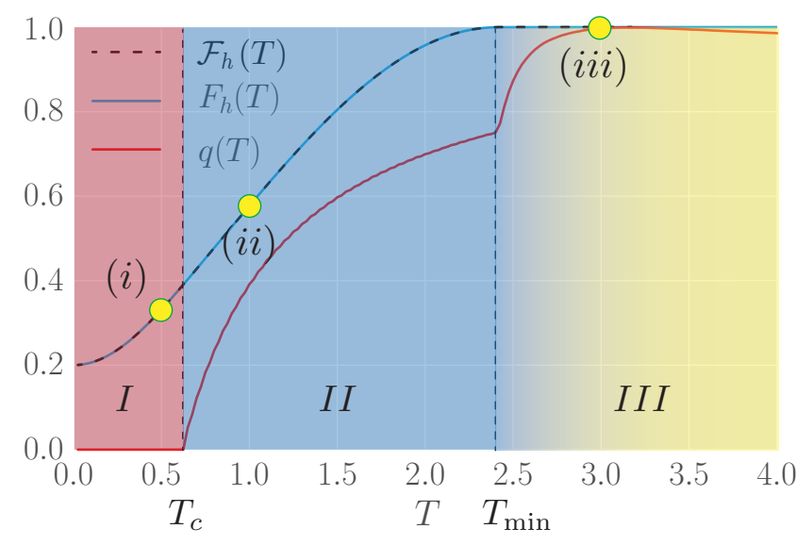
- RL and quantum control
- variational theory for optimal protocols



MB et al, PRX 8 031086 (2018)

→ **Phase transitions in the control landscape:**

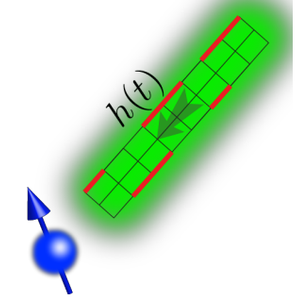
- how "hard" is it for the RL agent to Learn?



PRA 97, 052114 (2018)
PRL 122, 020601 (2019)

Marin Bukov

RL-inspired Discovery: Phase Diagram of Quantum Control

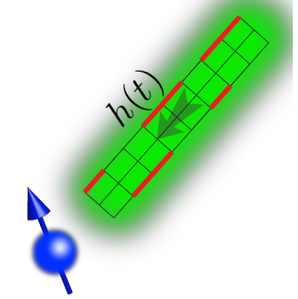


$$H(t) = -S^z - h_x(t) S^x$$

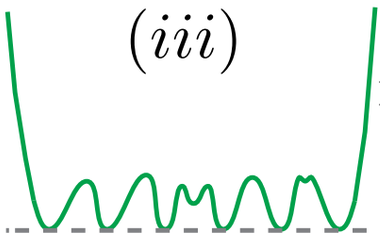
bang-bang protocols

$$h \mapsto 1 - F_h(T)$$

RL-inspired Discovery: Phase Diagram of Quantum Control



infidelity landscape (schematic)



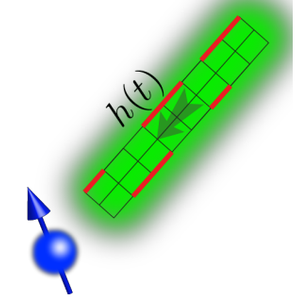
$$H(t) = -S^z - h_x(t)S^x$$

bang-bang protocols

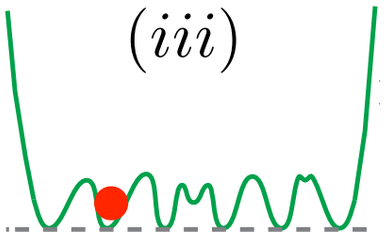
$$h \mapsto 1 - F_h(T)$$

infidelity landscape minima: $\{h^\alpha\}$

RL-inspired Discovery: Phase Diagram of Quantum Control



infidelity landscape (schematic)



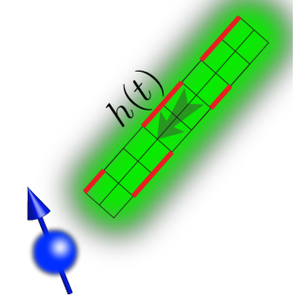
$$H(t) = -S^z - h_x(t)S^x$$

bang-bang protocols

$$h \mapsto 1 - F_h(T)$$

infidelity landscape minima: $\{h^\alpha\}$

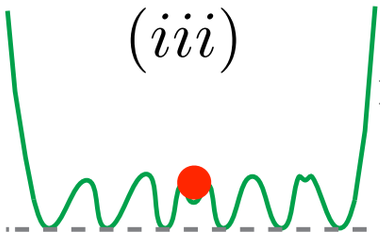
RL-inspired Discovery: Phase Diagram of Quantum Control



infidelity landscape (schematic)

$$H(t) = -S^z - h_x(t)S^x$$

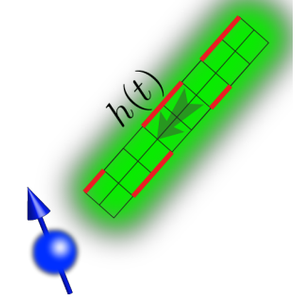
bang-bang protocols



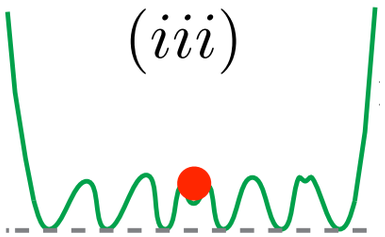
$$h \mapsto 1 - F_h(T)$$

infidelity landscape minima: $\{h^\alpha\}$

RL-inspired Discovery: Phase Diagram of Quantum Control



infidelity landscape (schematic)



$$H(t) = -S^z - h_x(t)S^x$$

bang-bang protocols

$$h \mapsto 1 - F_h(T)$$

infidelity landscape minima: $\{h^\alpha\}$

$$\bar{h}(t) = \frac{1}{\#_{\text{real}}} \sum_\alpha h^\alpha(t)$$

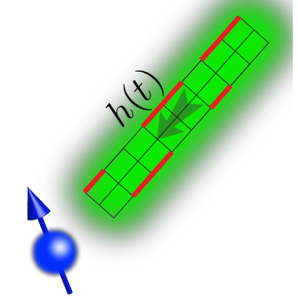
Edwards-Anderson-like order parameter:

$$q(T) \sim \sum_{j=1}^{N_T} \overline{\{h(j\delta t) - \overline{h}(j\delta t)\}^2}$$

Marin Bukov

PRX 8 031086 (2018)

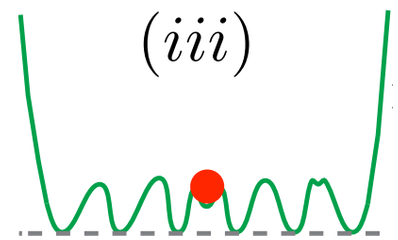
RL-inspired Discovery: Phase Diagram of Quantum Control



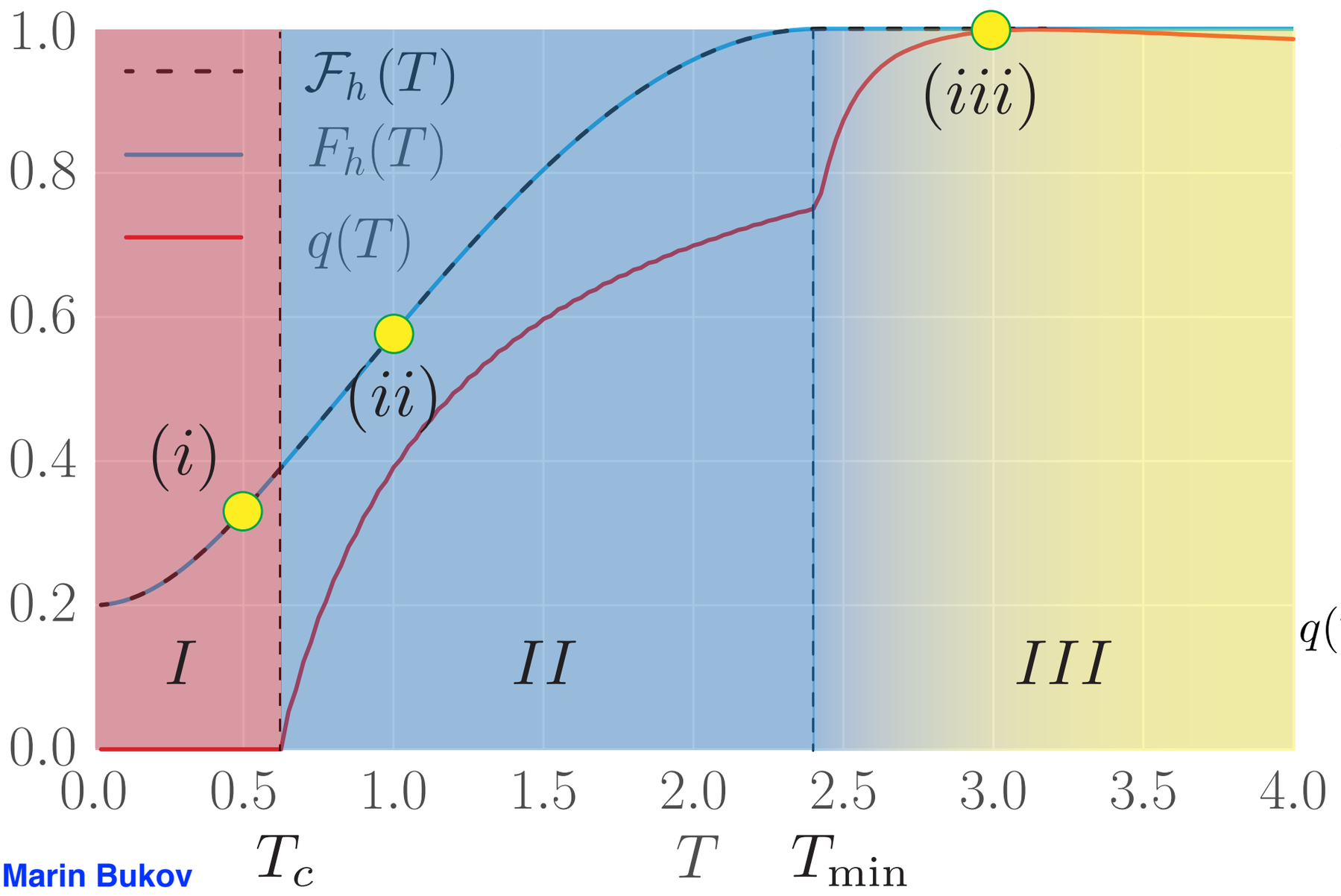
infidelity landscape (schematic)

$$H(t) = -S^z - h_x(t) S^x$$

bang-bang protocols



$$h \mapsto 1 - F_h(T)$$



infidelity landscape minima: $\{h^\alpha\}$

$$\bar{h}(t) = \frac{1}{\#_{\text{real}}} \sum_\alpha h^\alpha(t)$$

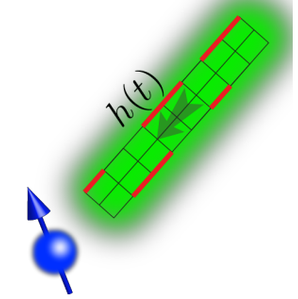
Edwards-Anderson-like order parameter:

$$q(T) \sim \sum_{j=1}^{N_T} \overline{\{h(j\delta t) - \overline{h}(j\delta t)\}^2}$$

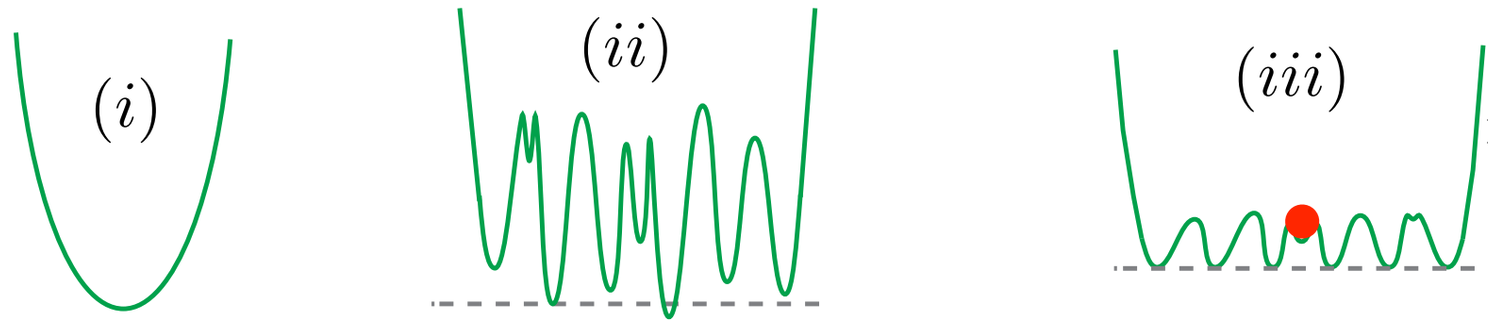
Marin Bukov

PRX 8 031086 (2018)

RL-inspired Discovery: Phase Diagram of Quantum Control



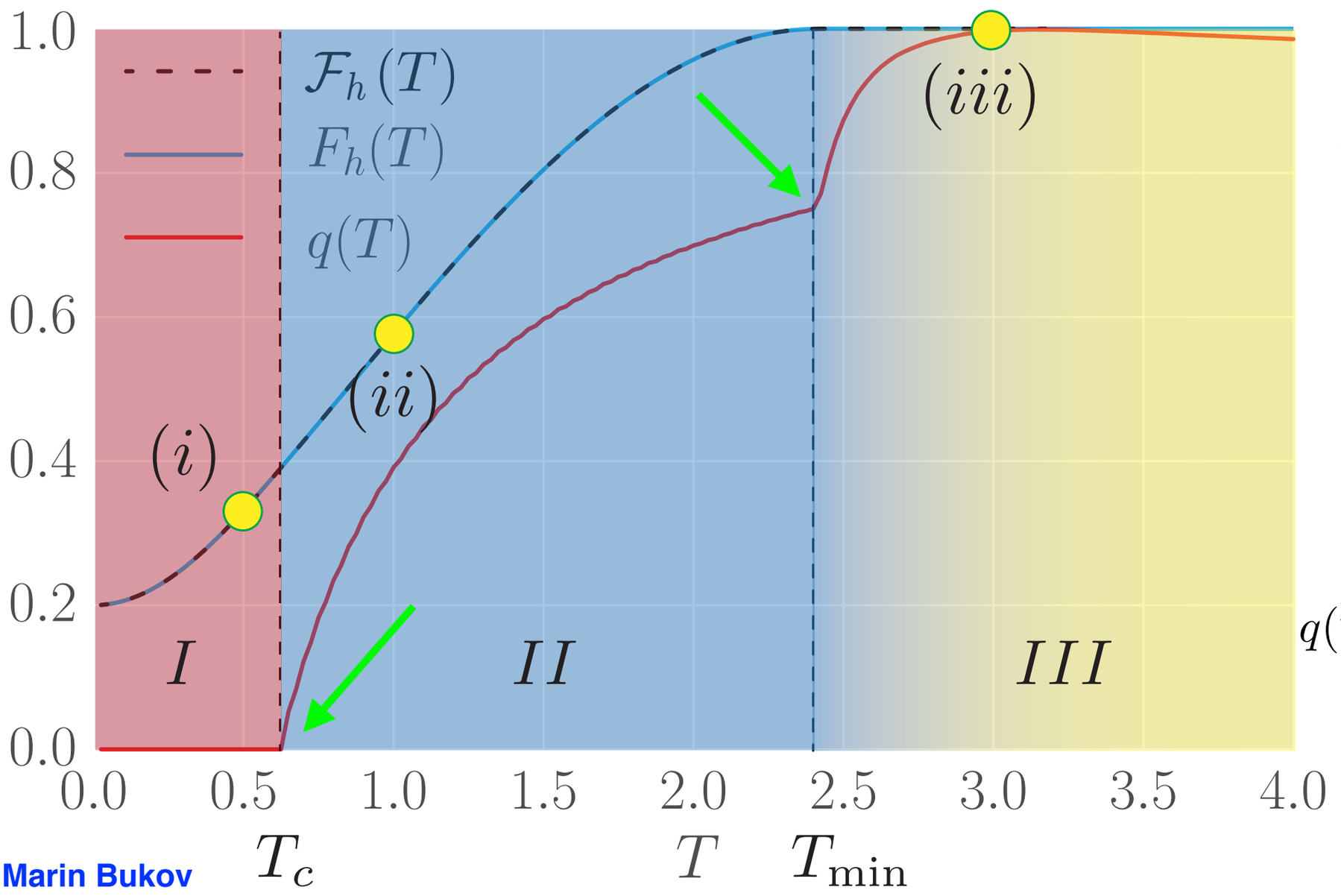
infidelity landscape (schematic)



$$H(t) = -S^z - h_x(t)S^x$$

bang-bang protocols

$$h \mapsto 1 - F_h(T)$$



infidelity landscape minima: $\{h^\alpha\}$

$$\bar{h}(t) = \frac{1}{\#_{\text{real}}} \sum_\alpha h^\alpha(t)$$

Edwards-Anderson-like order parameter:

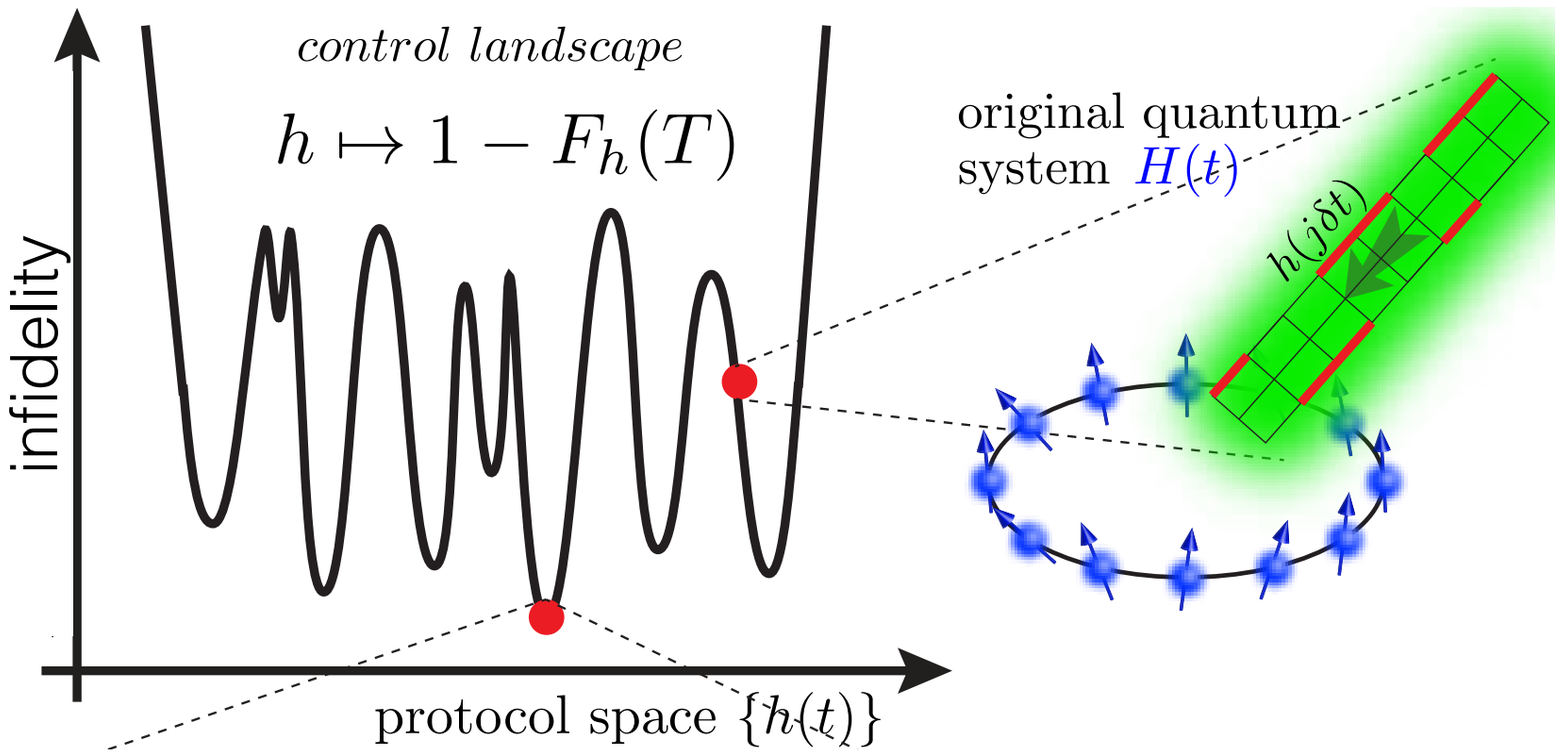
$$q(T) \sim \sum_{j=1}^{N_T} \overline{\{h(j\delta t) - \overline{h}(j\delta t)\}^2}$$

Marin Bukov

PRX 8 031086 (2018)

Effective Classical Energy Model

→ one-to-one correspondence:

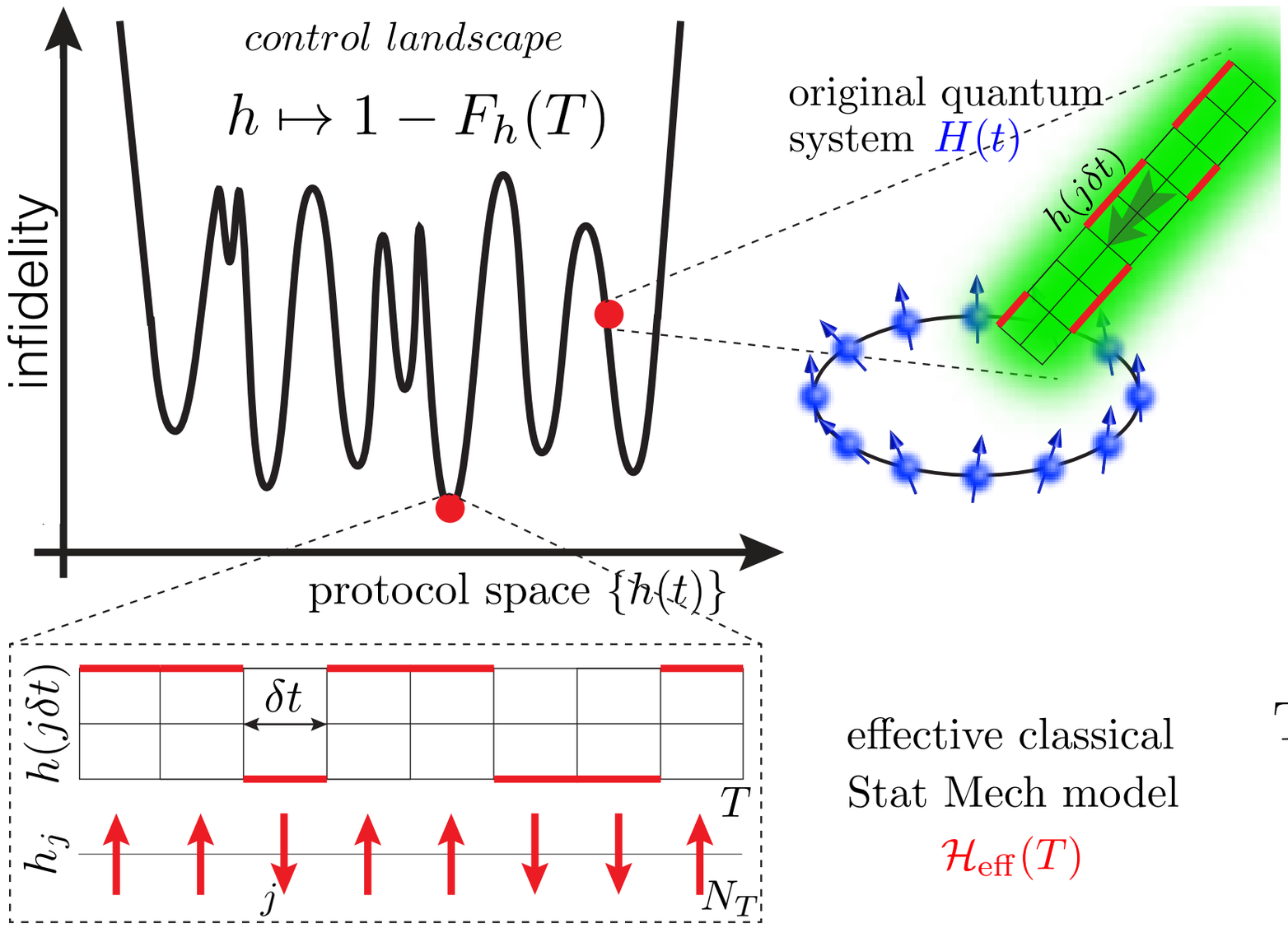


$$H(t) = -S^z - h_x(t)S^x$$

PRL 122, 020601 (2019)

Effective Classical Energy Model

→ one-to-one correspondence:



$$H(t) = -S^z - h_x(t) S^x$$

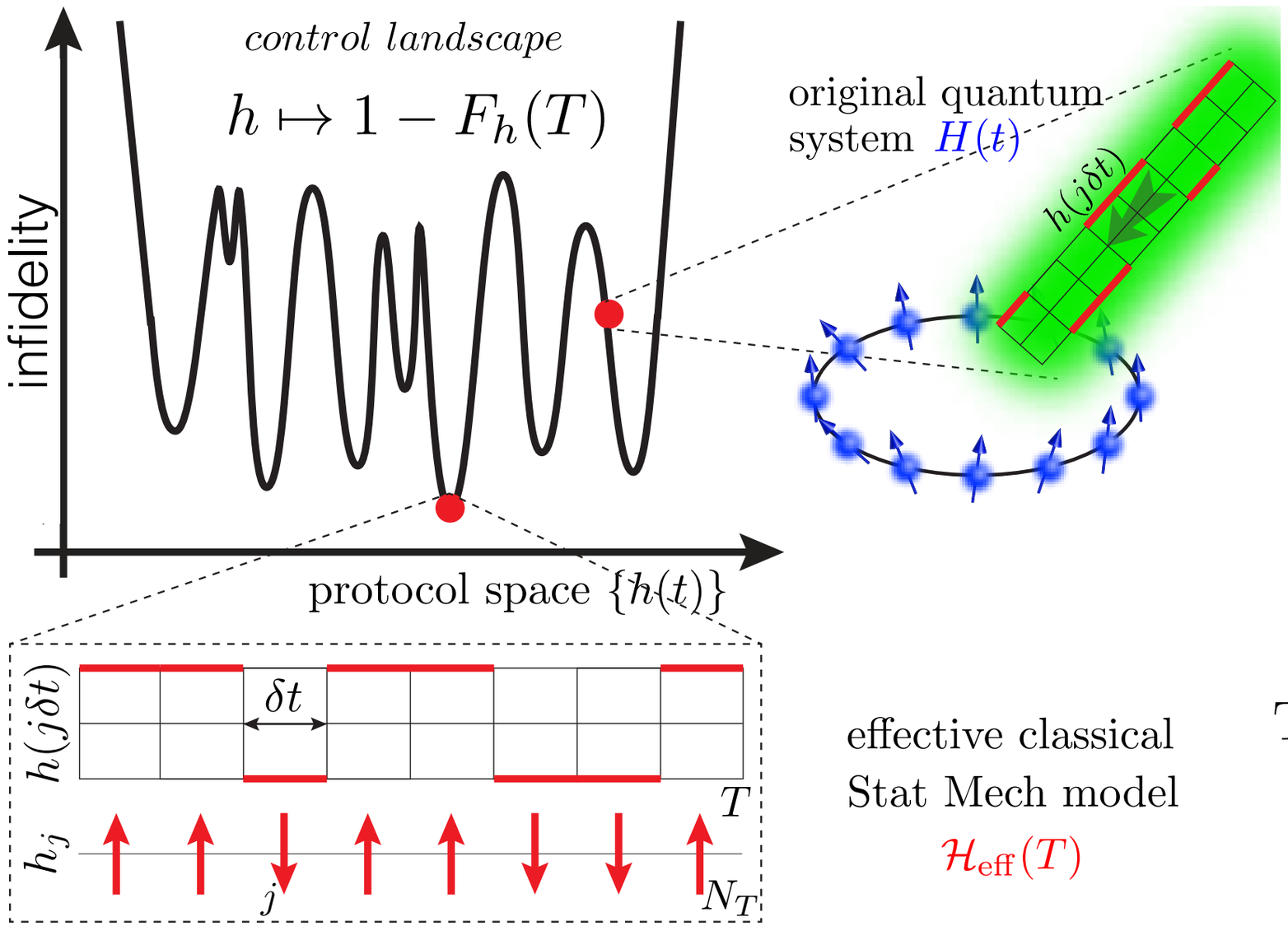
effective classical Stat Mech model $\mathcal{H}_{\text{eff}}(T)$

$$\text{TD limit}: \; N_T \to \infty, \delta t \to 0$$
$$N_T \delta t = const.$$

PRL 122, 020601 (2019)

Effective Classical Energy Model

→ one-to-one correspondence:



$$H(t) = -S^z - h_x(t)S^x$$

effective classical Stat Mech model $\mathcal{H}_{\text{eff}}(T)$

TD limit : $N_T \to \infty, \delta t \to 0$

$N_T \delta t = const.$

→ effective *classical* spin energy describes control landscape

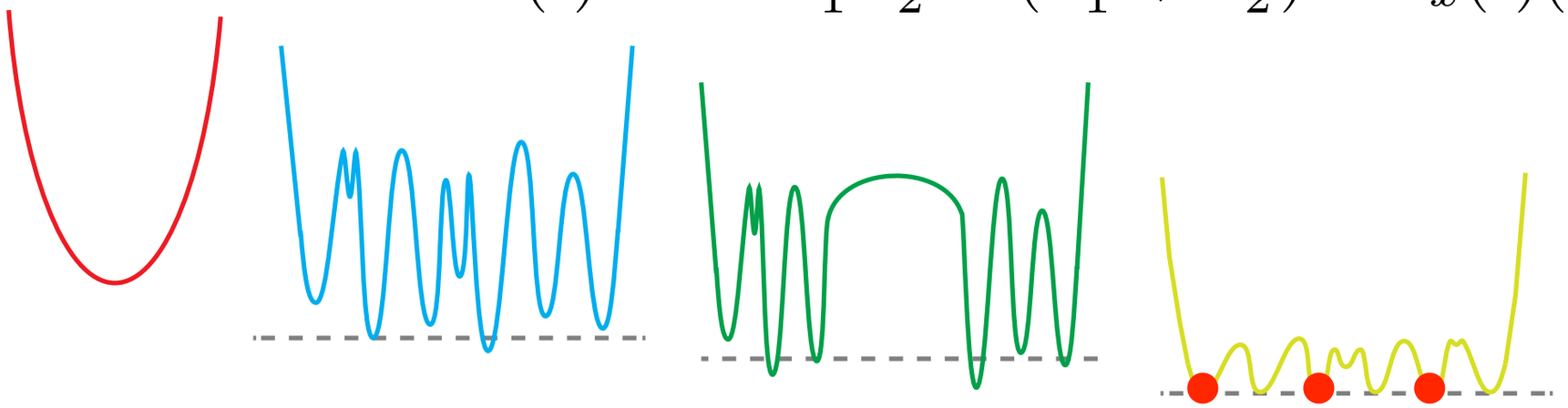
$$\mathcal{H}_{\text{eff}}(T) = I(T) + \sum_j G_j(T)h_j + \sum_{ij} J_{ij}(T)h_i h_j + \sum_{ijk} K_{ijk}(T)h_i h_j h_k + \ldots$$

j : sites on time lattice

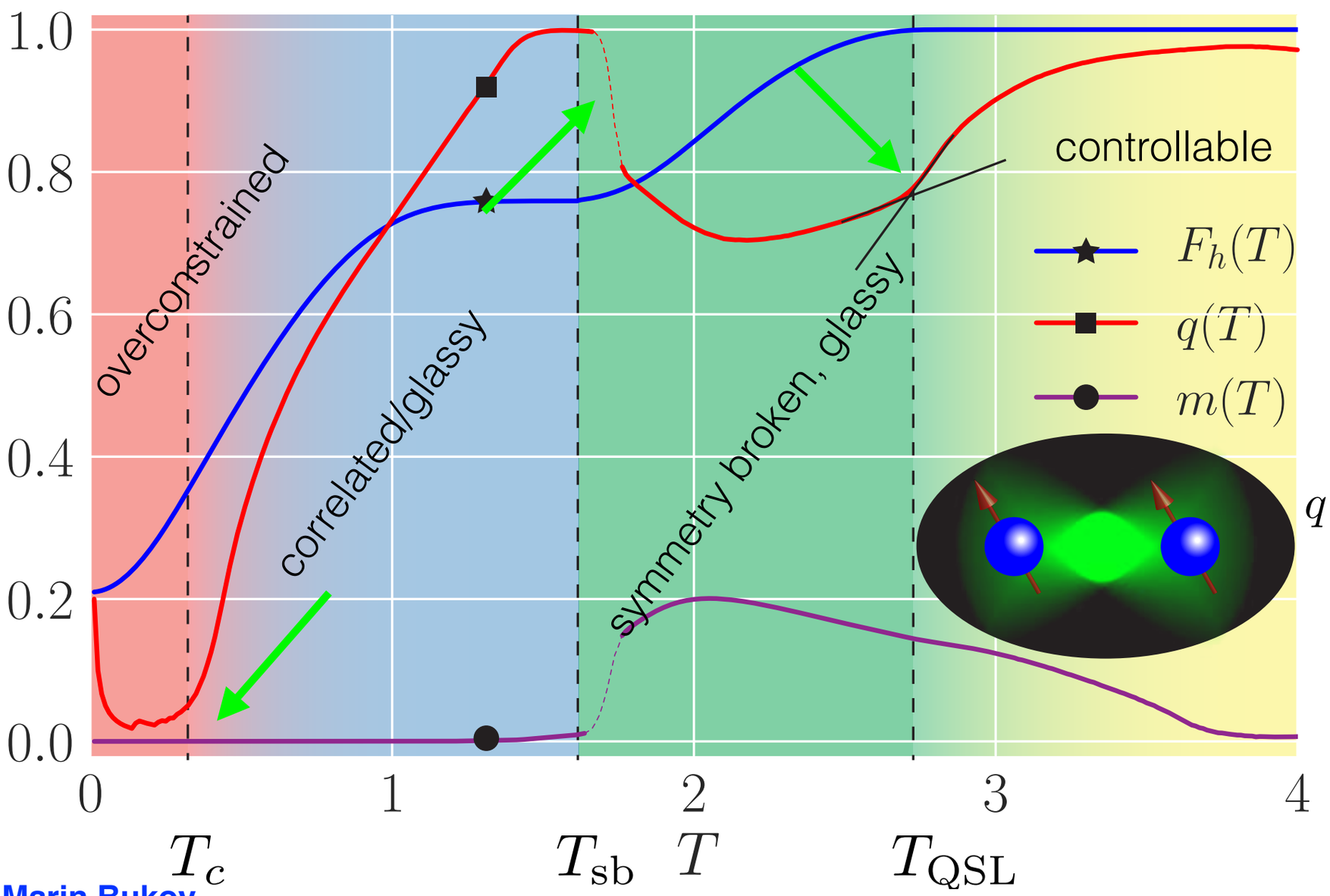
PRL 122, 020601 (2019)

Symmetry Breaking in the Control Landscape

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$



infidelity landscape minima: $\{h^\alpha\}$



$$\bar{h}(t) = \frac{1}{\#_{\rm real}} \sum_\alpha h^\alpha(t)$$

Edwards-Anderson-like order parameter:

$$q(T) \sim \sum_{j=1}^{N_T} \overline{\{h(j\delta t) - \overline{h}(j\delta t)\}^2}$$

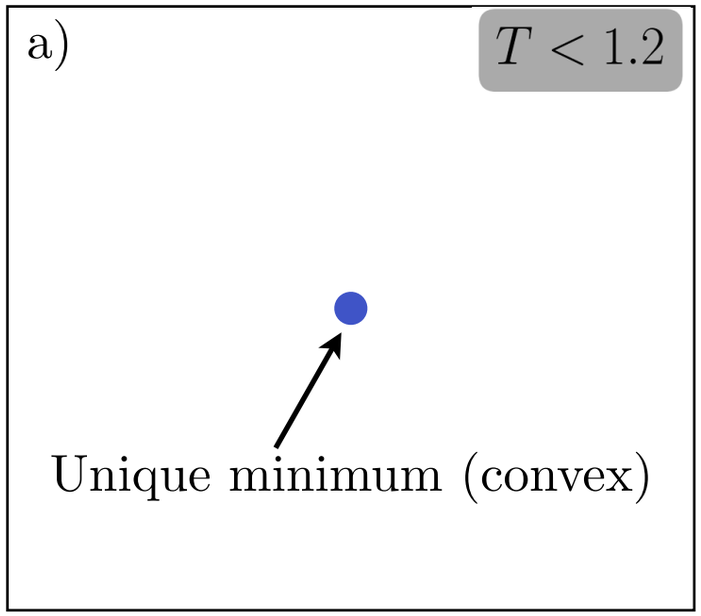
Marin Bukov

PRL 95, 200201 (2005)
PRA 97, 052114 (2018)

Visualizing the Glassy Transition with t-SNE

Berkeley
UNIVERSITY OF CALIFORNIA

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

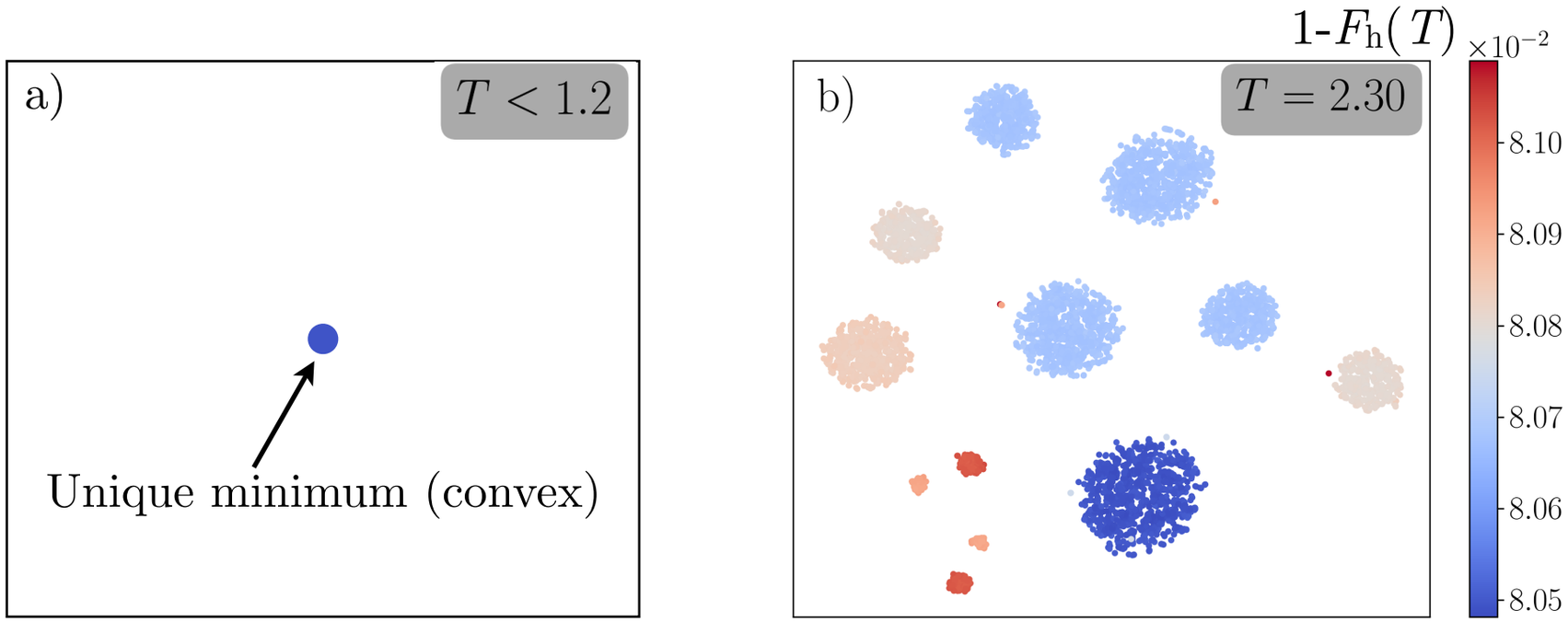


PRL 122, 020601 (2019)

Visualizing the Glassy Transition with t-SNE

Berkeley
UNIVERSITY OF CALIFORNIA

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

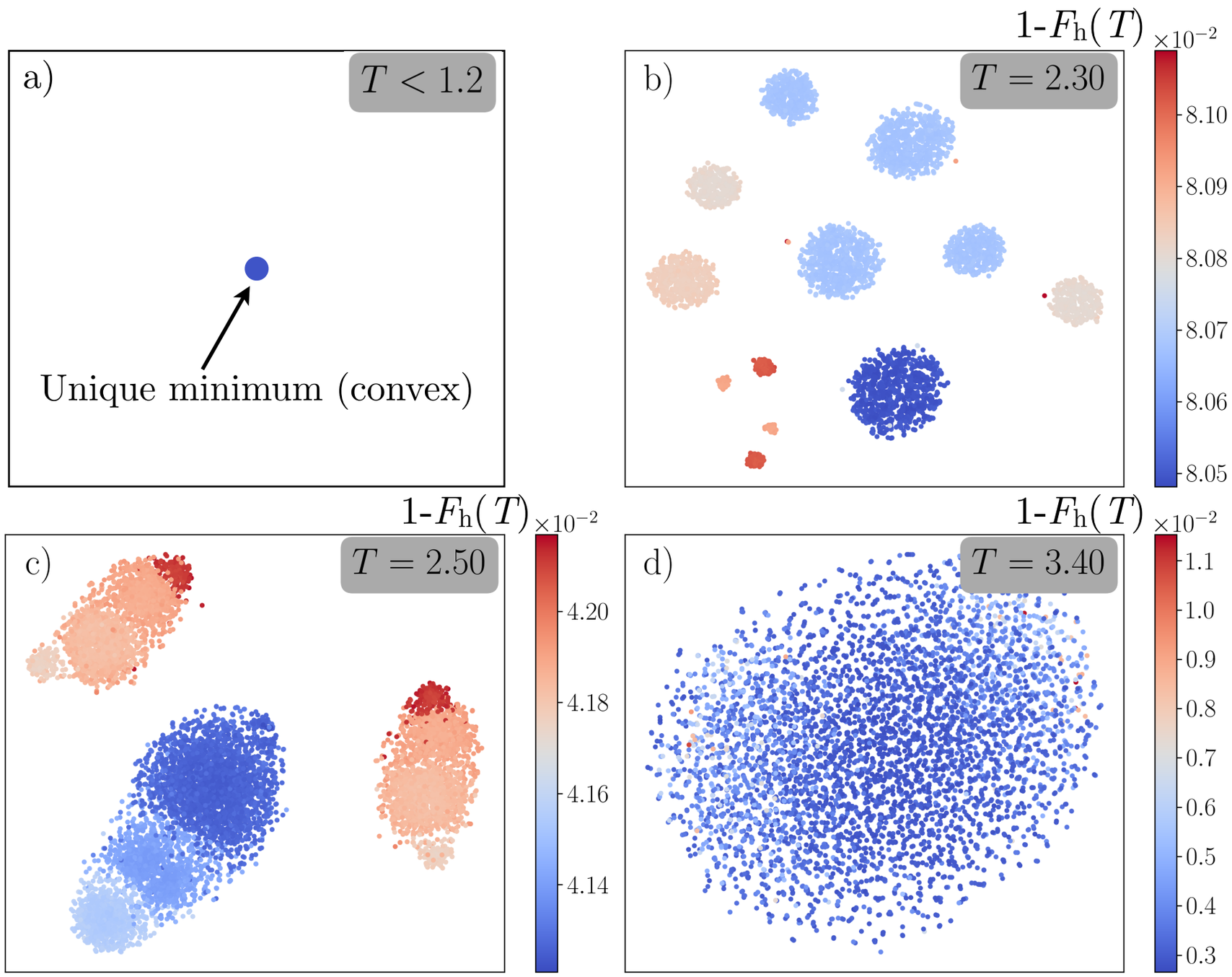


PRL 122, 020601 (2019)

Visualizing the Glassy Transition with t-SNE

Berkeley
UNIVERSITY OF CALIFORNIA

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



PRL 122, 020601 (2019)

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 exactly
- use (deep) RL to find guiding principles *away from equilibrium*?
- solid-state materials: we don't know the Hamiltonian

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

→ **autonomous:** does not require supervision

Marin Bukov

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 exactly
- use (deep) RL to find guiding principles *away from equilibrium*?
- solid-state materials:
  we don't know the Hamiltonian

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

- RL agent gathers 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

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 exactly
- use (deep) RL to find guiding principles *away from equilibrium*?
- solid-state materials: we don't know the Hamiltonian

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

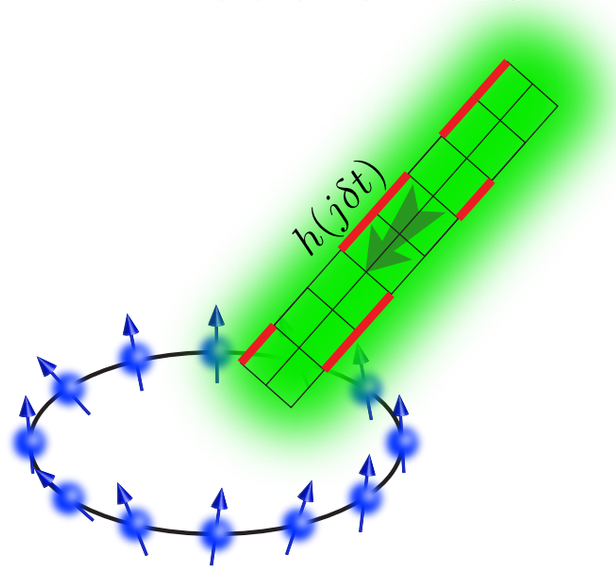
- RL agent gathers knowledge about the environment which *can be used after training*
- can RL reveal similarities between at first sight unrelated problems?

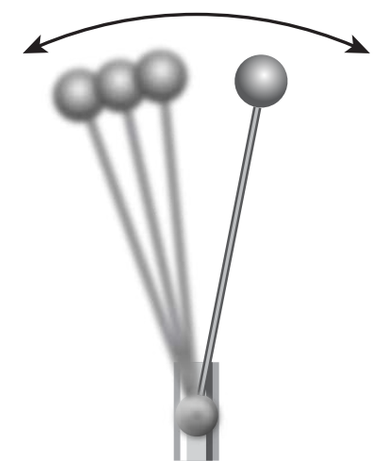
→ **autonomous:** does not require supervision

- can RL automate experimental setups?

Outlook

- 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?





spin chain: PRX 8 031086 (2018)

control phases: PRL 122, 020601 (2019)

PRA 97, 052114 (2018)

Kapitza oscillator: PRB 98, 224305 (2018)

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

open source python package for **many-body** lattice systems

(with P. Weinberg, BU)

RL and Optimal Control (OC)

→ different sides of the same medal

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

RL and Optimal Control (OC)

→ different sides of the same medal

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

→ 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

RL and Optimal Control (OC)

→ different sides of the same medal

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

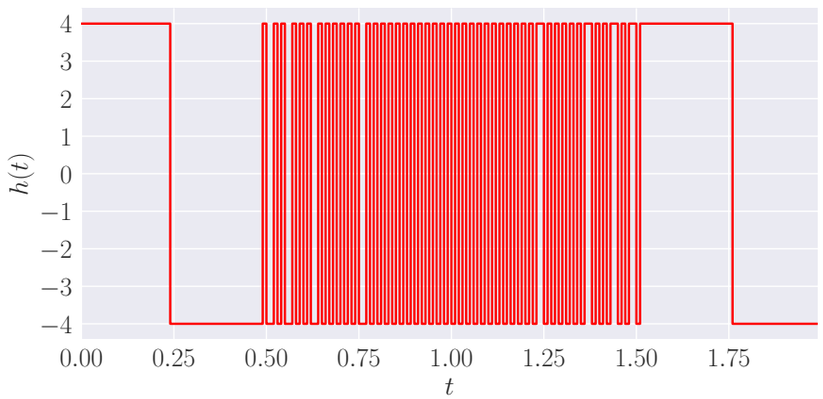
→ modern Control Theory: both RL and OC under same hood

→ currently in physics: preferred approach is OC

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

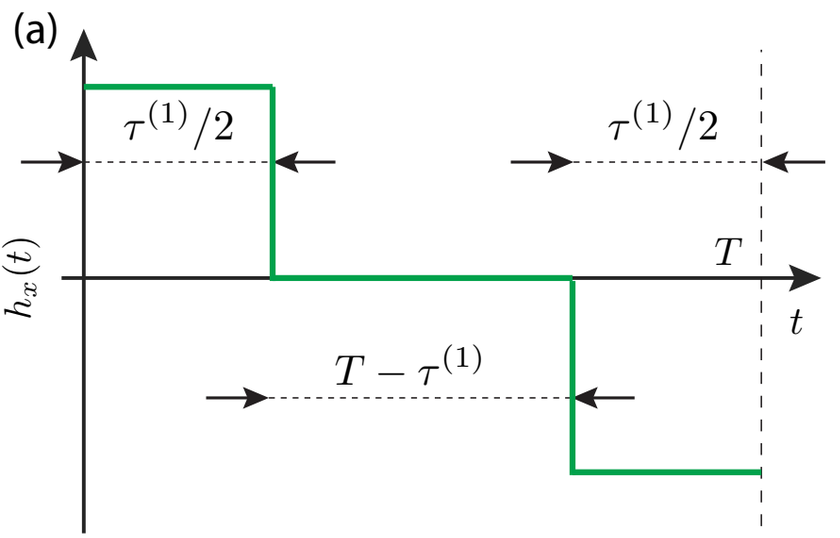
OC ← *closely related* → **RL**	
based on: ***variational calculus***	***Markov decision processes***
• needs model for environment to express cost function in.	• no model of controlled system, adaptive, autonomous.
• best suited for deterministic environments.	• stochastic/uncertain environments.
• 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.

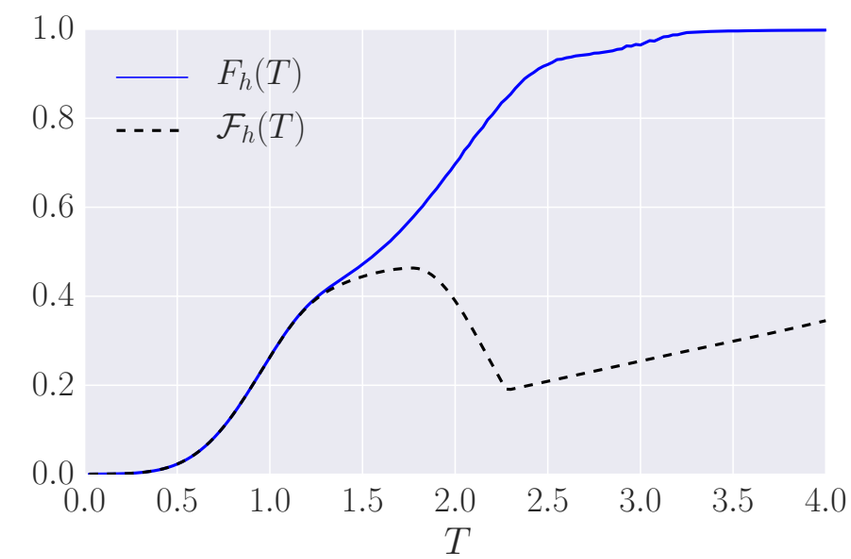
What do we Learn from the RL Agent?

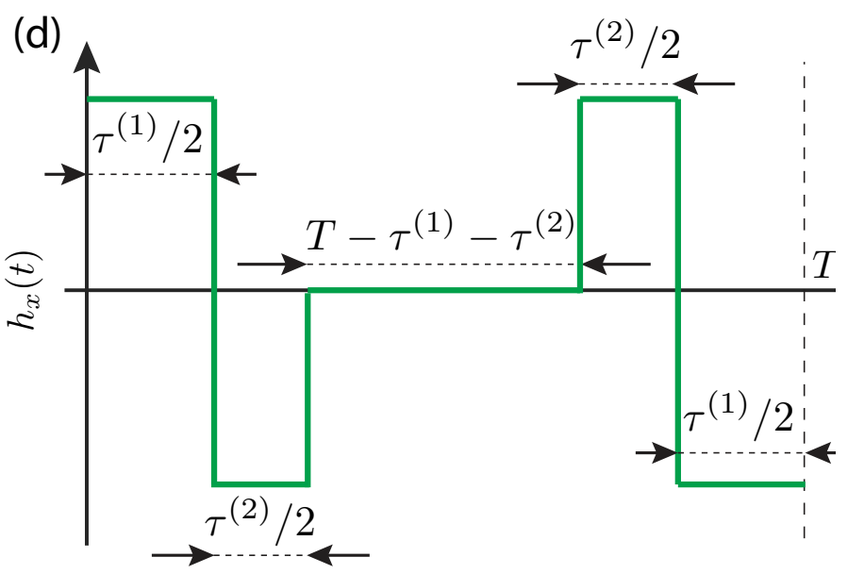


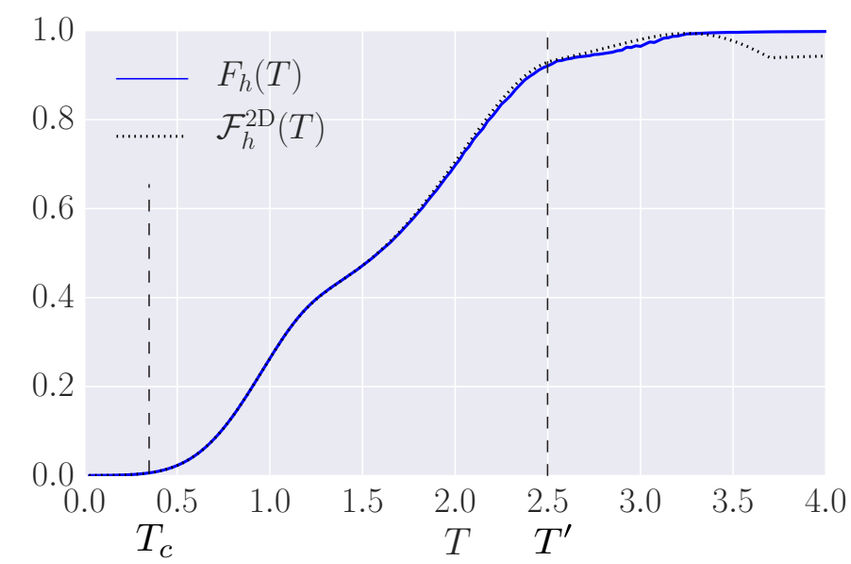
$$H = \sum_j -S^z_{j+1}S^z_j - h_z S^z_j - h_x(t) S^x_j$$

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





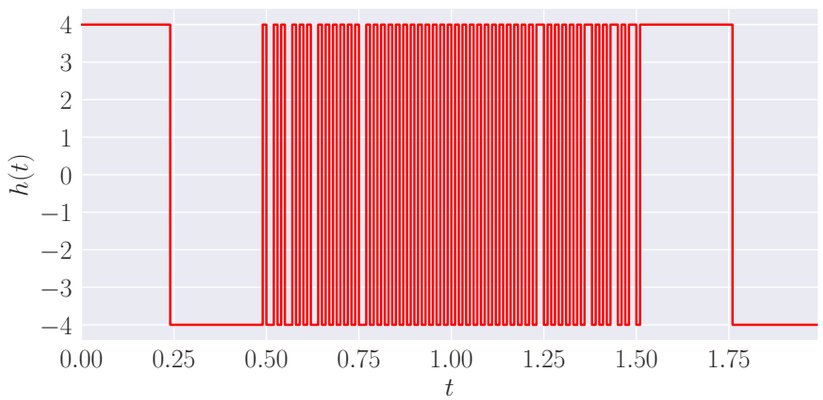




Marin Bukov

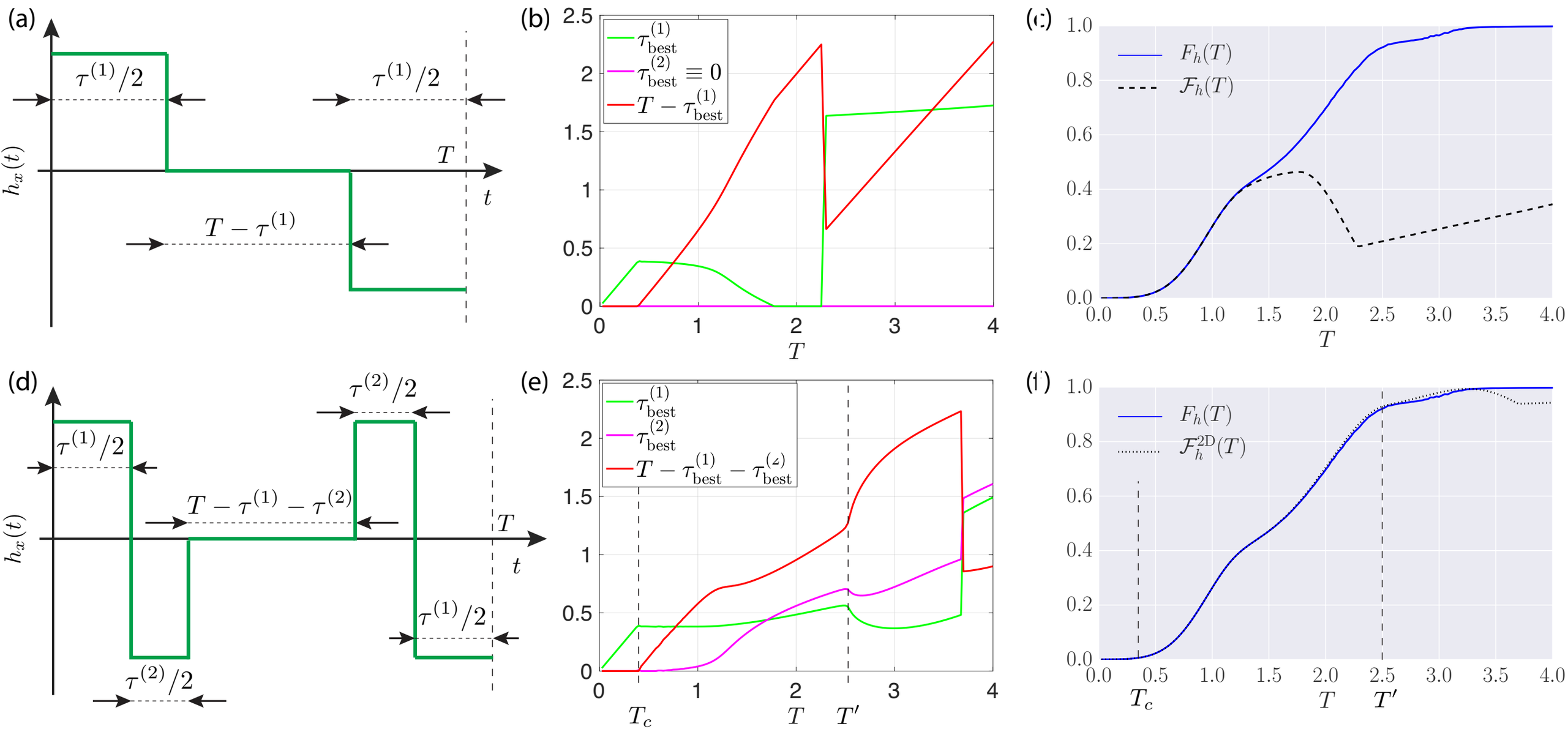
PRX 8 031086 (2018)

What do we Learn from the RL Agent?



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

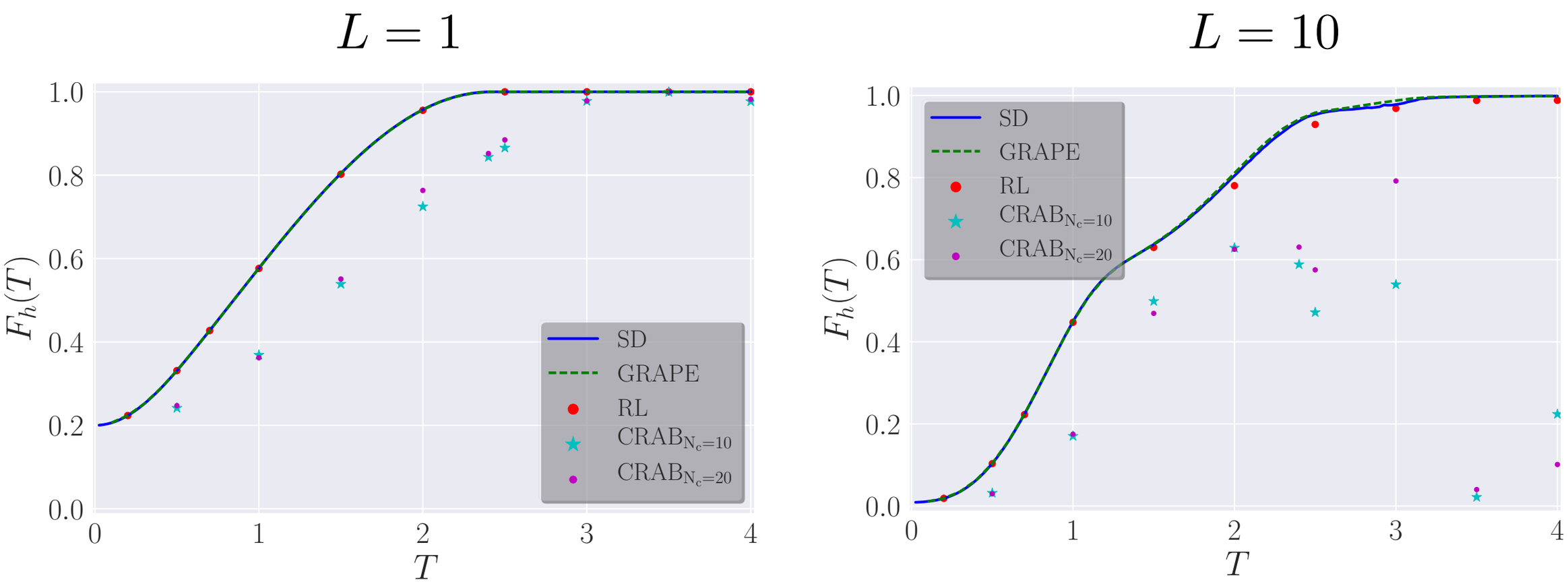
$$-\mathcal{F}_h(T) = \min_{\tau^{(1)}\in[0,T]} \left(-\mathcal{F}_h(T;\tau^{(1)})\right)$$



Marin Bukov

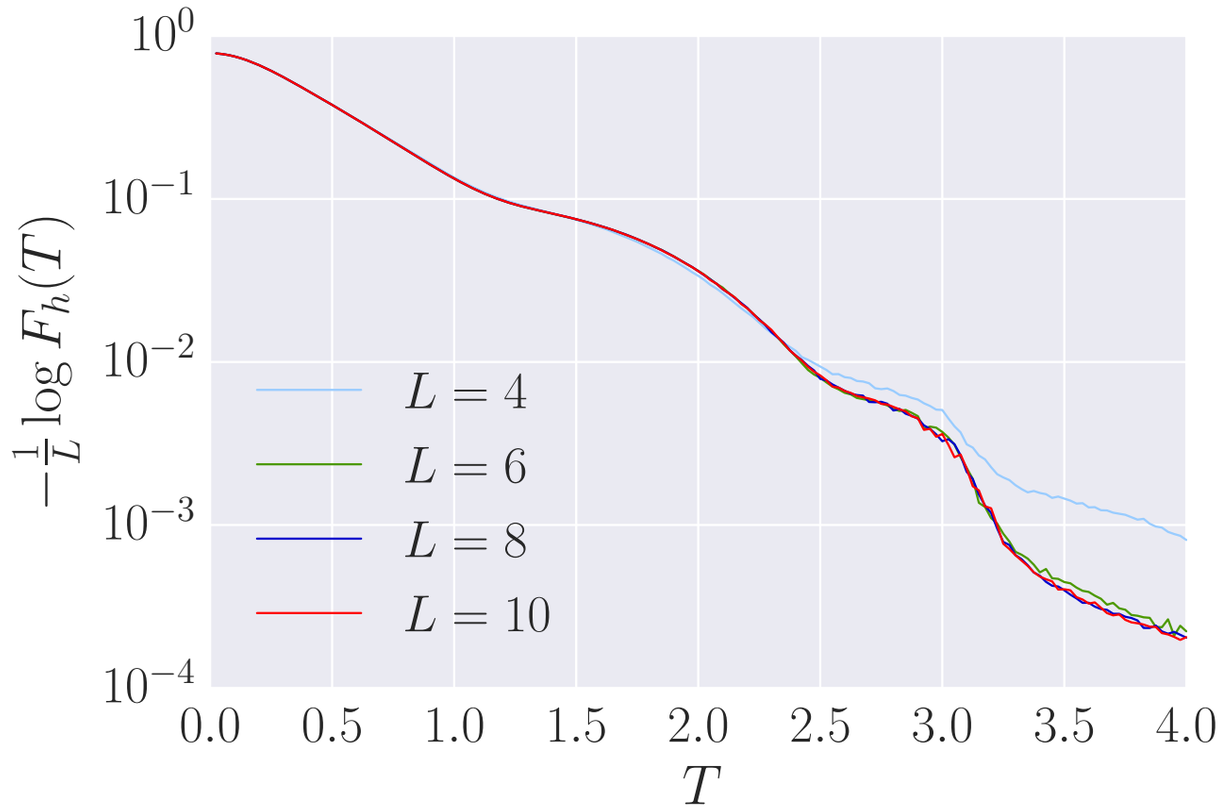
PRX 8 031086 (2018)

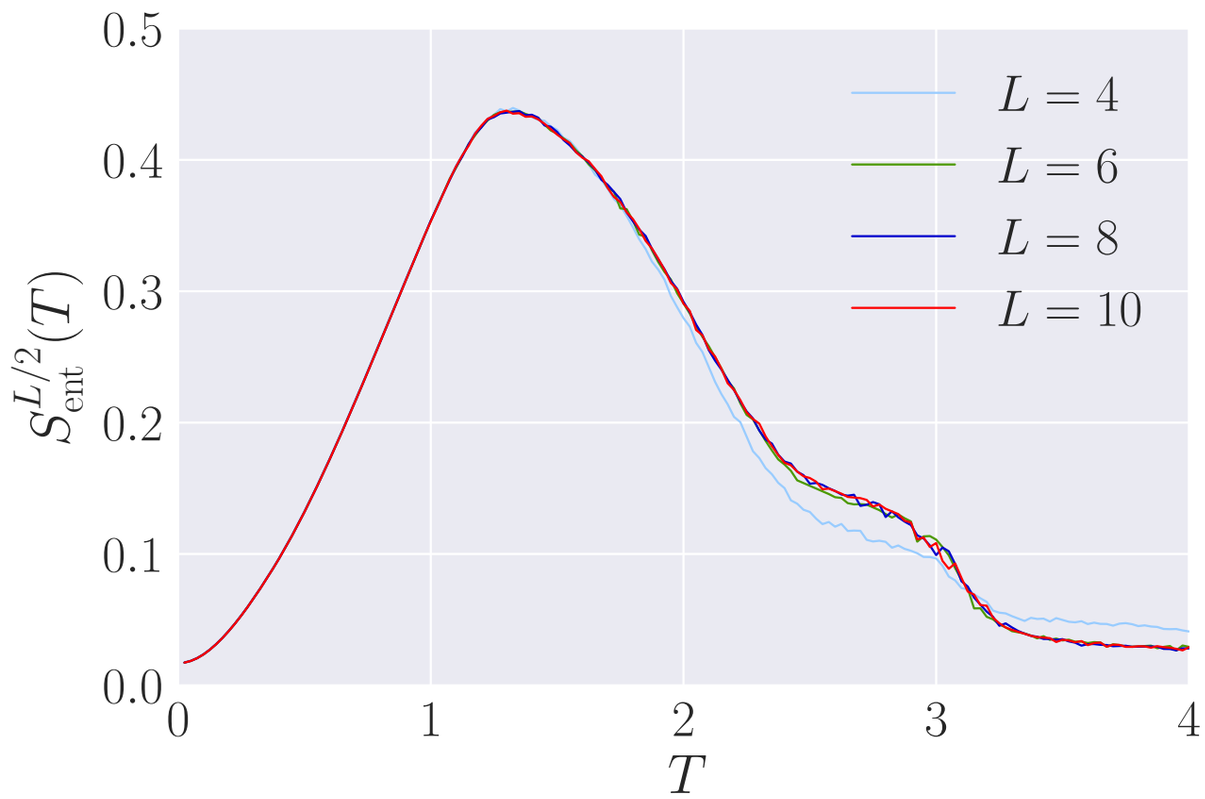
RL vs OC

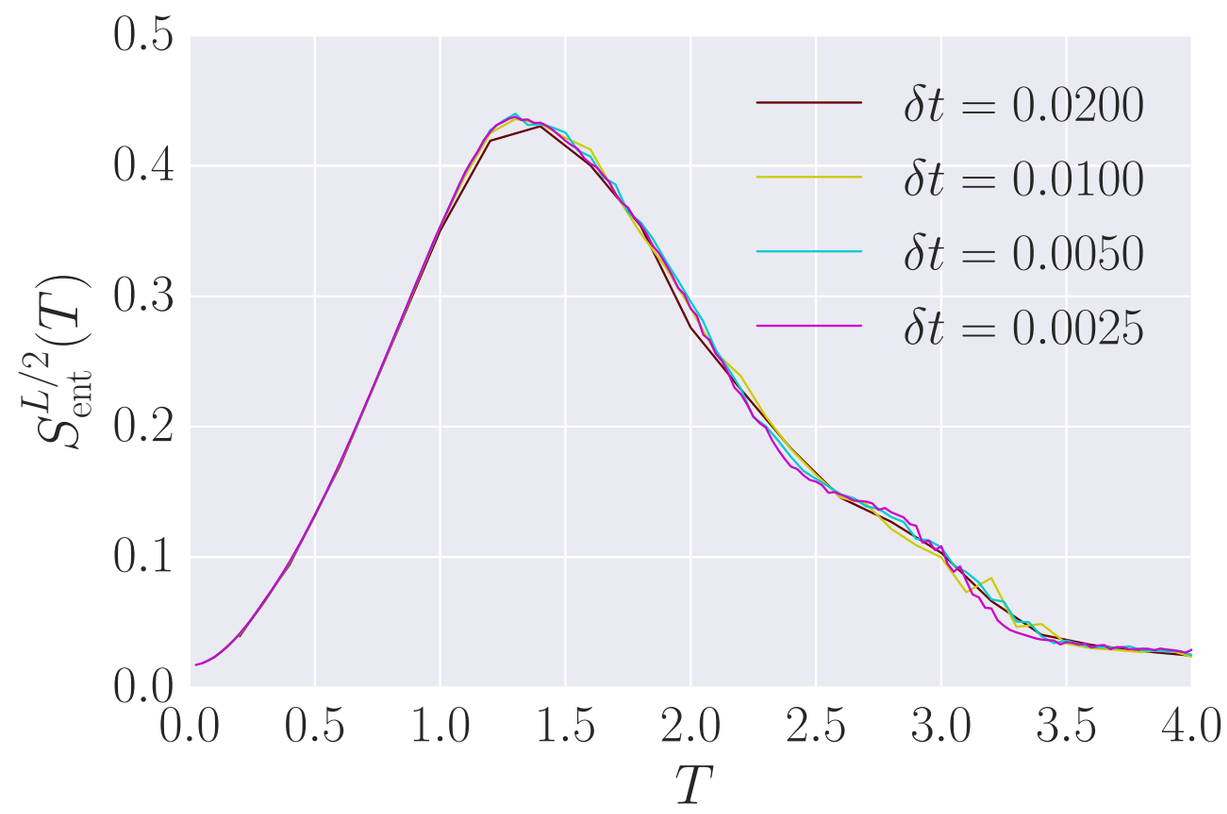


PRX 8 031086 (2018)

System Size Dependence





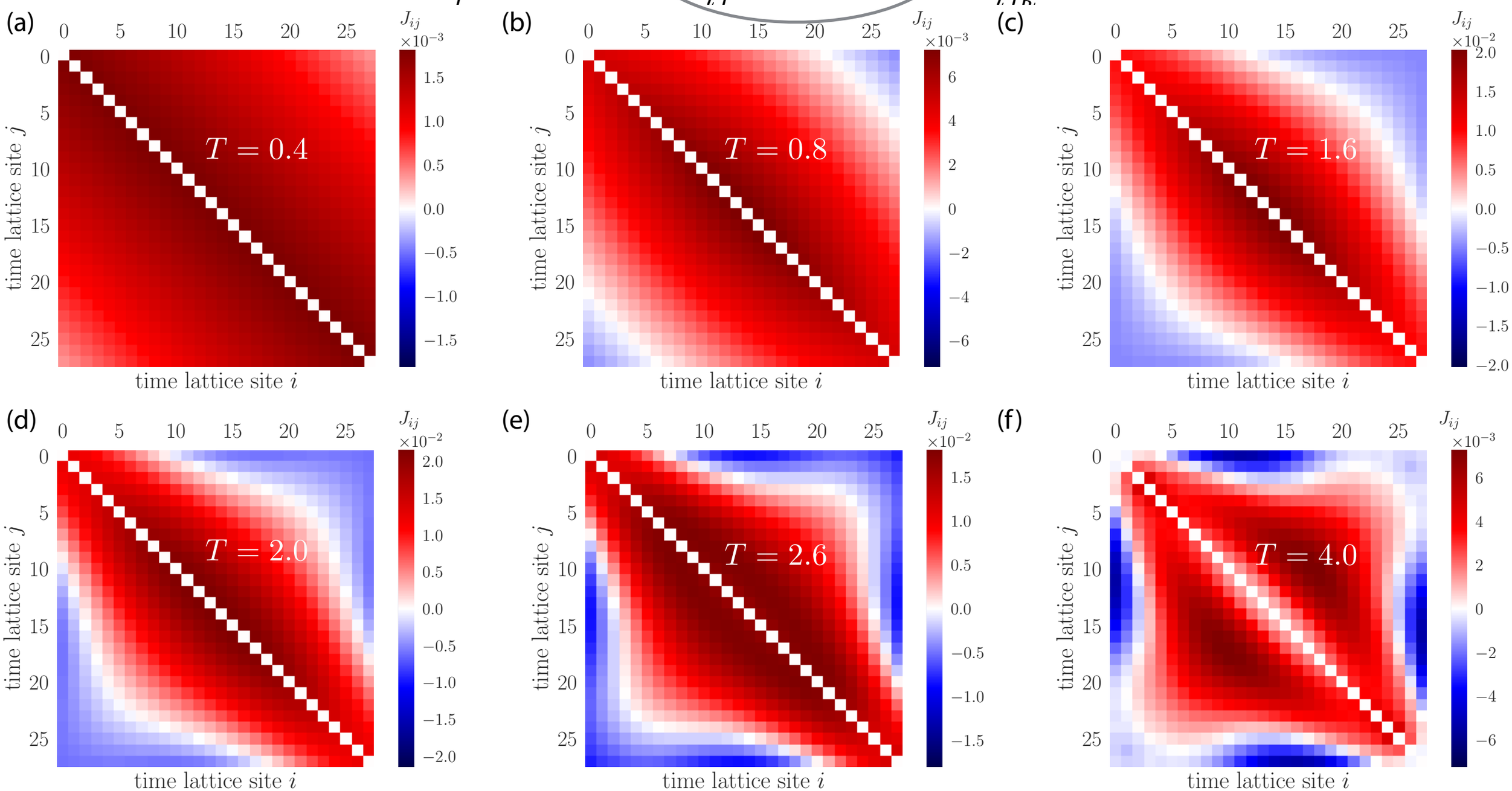


PRX 8 031086 (2018)

Properties of Effective Model $\mathcal{H}_{\text{eff}}(T)$

→ properties of effective coupling strengths: GS in frustrated!

$$\mathcal{H}_{\text{eff}}(T) = I(T) + \sum_{i} G_j(T) h_j + \sum_{ij} J_{ij}(T) h_i h_j + \sum_{ijk} K_{ijk}(T) h_i h_j h_k + \ldots$$



PRL 122, 020601 (2019)

Spontaneous Symmetry Breaking

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$

$\mathbb{Z}_2$ *symmetry of protocols:*

$$|\psi_*\rangle = \mathrm{e}^{-i\pi(S_1^z + S_2^z)}|\psi_i\rangle$$

$$F_{h(t)}(T) = F_{-h(T-t)}(T)$$

PRA 97, 052114 (2018)

Spontaneous Symmetry Breaking

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$

$\mathbb{Z}_2$ *symmetry of protocols:*

$$|\psi_*\rangle = \mathrm{e}^{-i\pi(S_1^z + S_2^z)}|\psi_i\rangle$$

$$F_{h(t)}(T) = F_{-h(T-t)}(T)$$

optimal protocol:

→ *either:* symmetric & unique

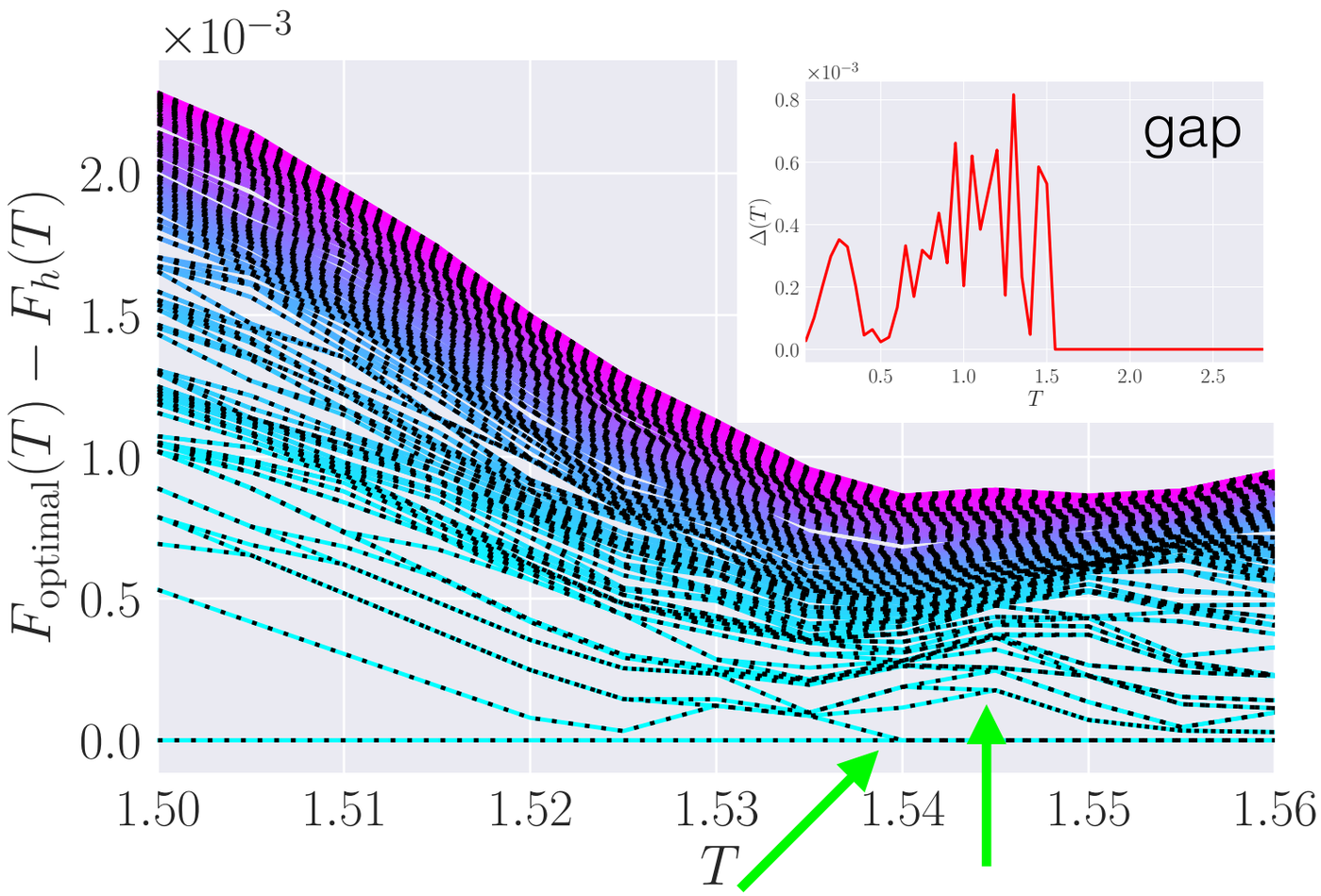
→ *or:* symmetry broken & doubly degenerate

PRA 97, 052114 (2018)

Spontaneous Symmetry Breaking

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$

'spectrum' of effective classical model $\mathcal{H}_{\text{eff}}(T)$



$\mathbb{Z}_2$ *symmetry of protocols:*

$$|\psi_*\rangle = \mathrm{e}^{-i\pi(S_1^z + S_2^z)}|\psi_i\rangle$$

$$F_{h(t)}(T) = F_{-h(T-t)}(T)$$

optimal protocol:

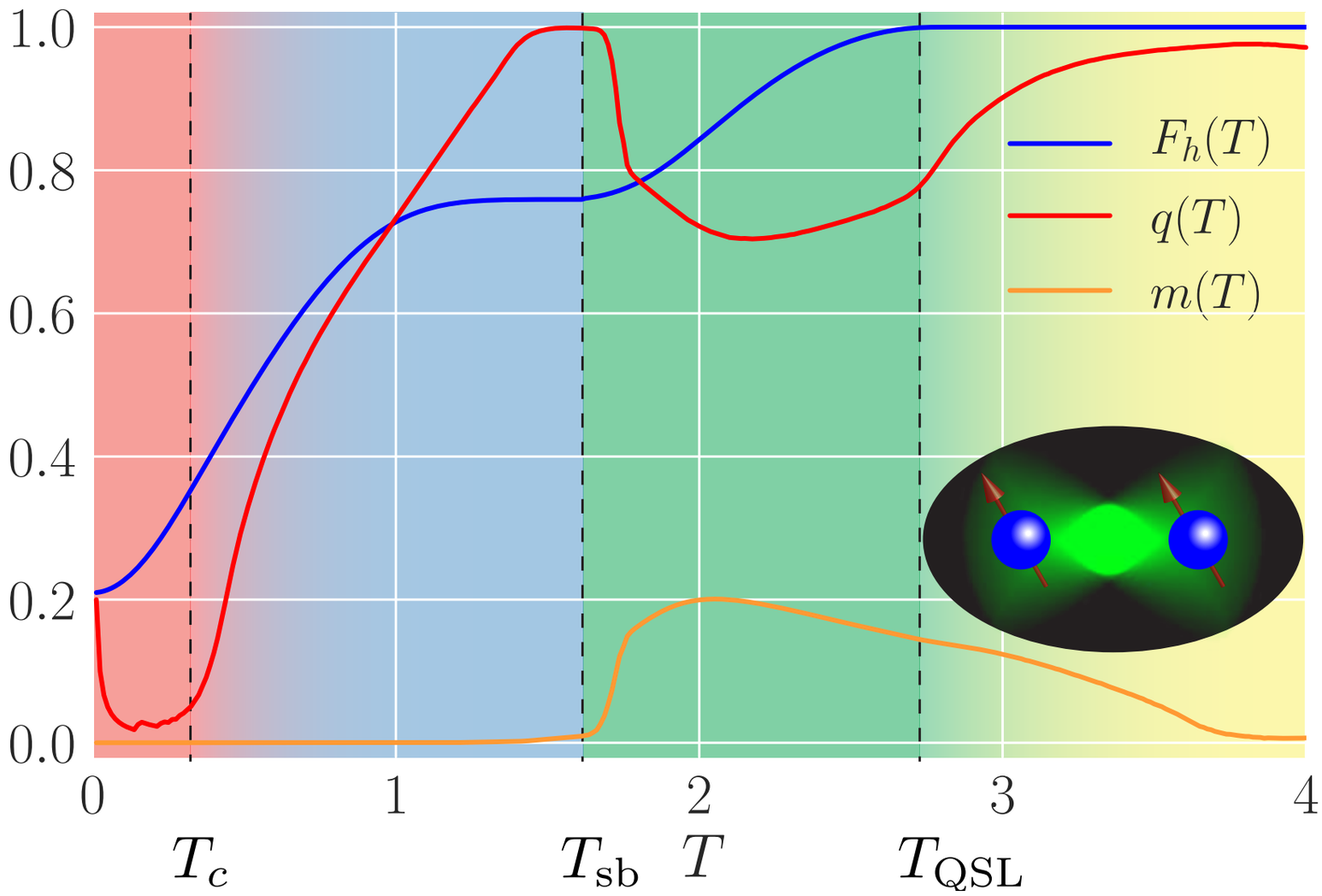
→ *either:* symmetric & unique

→ *or:* symmetry broken & doubly degenerate

PRA 97, 052114 (2018)

Spontaneous Symmetry Breaking

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$



$\mathbb{Z}_2$ *symmetry of protocols:*

$$|\psi_*\rangle = \mathrm{e}^{-i\pi(S_1^z+S_2^z)}|\psi_i\rangle$$

$$F_{h(t)}(T) = F_{-h(T-t)}(T)$$

optimal protocol:

→ *either:* symmetric & unique

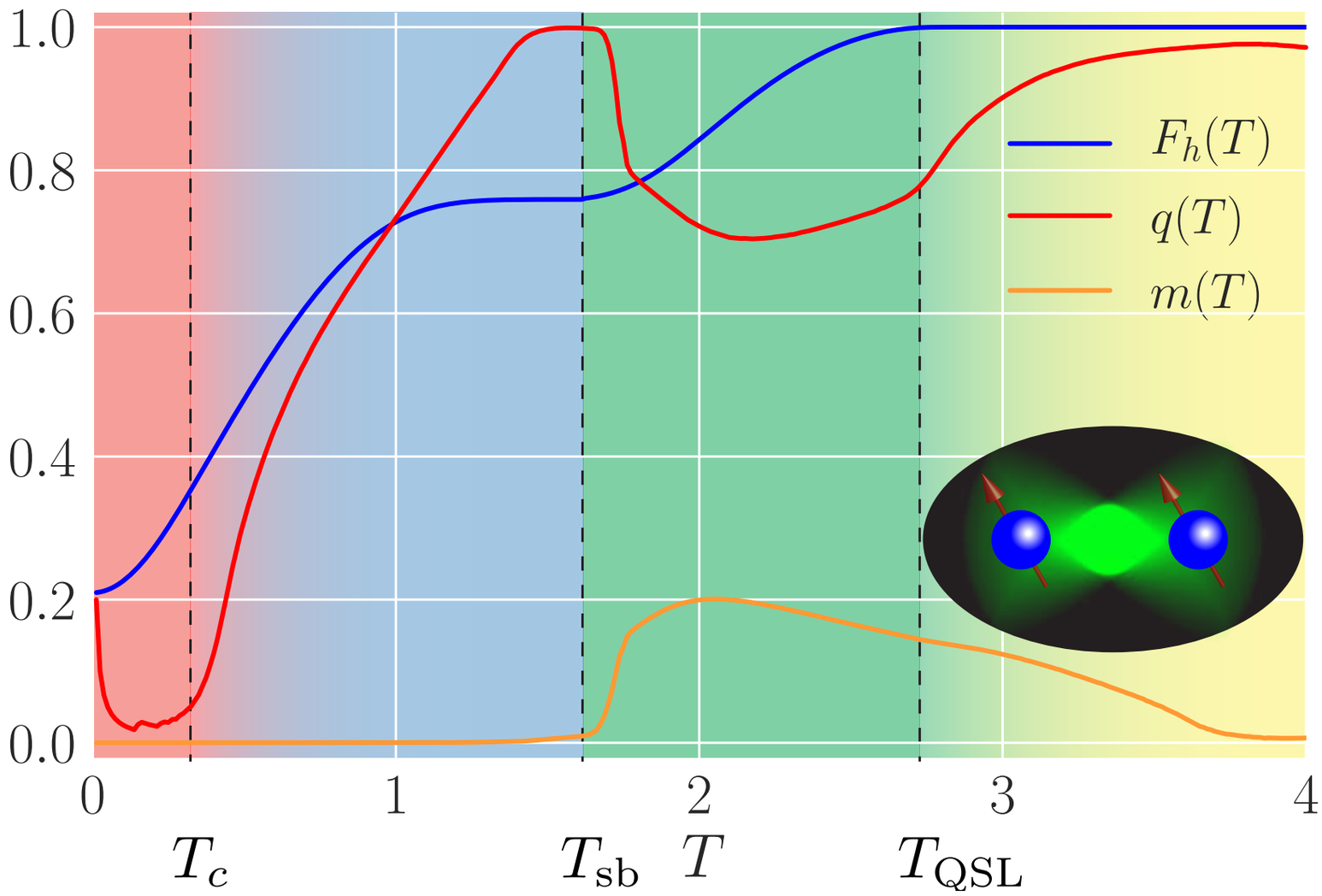
→ *or:* symmetry broken & doubly degenerate

→ "magnetisation" of a single protocol: $m_h(T) = N_T^{-1} \sum_{n=1}^{N_T} h_x(n\delta t)$

PRA 97, 052114 (2018)

Spontaneous Symmetry Breaking

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$



$\mathbb{Z}_2$ *symmetry of protocols:*

$$|\psi_*\rangle = \mathrm{e}^{-i\pi(S_1^z+S_2^z)}|\psi_i\rangle$$

$$F_{h(t)}(T) = F_{-h(T-t)}(T)$$

optimal protocol:

→ *either:* symmetric & unique

→ *or:* symmetry broken & doubly degenerate

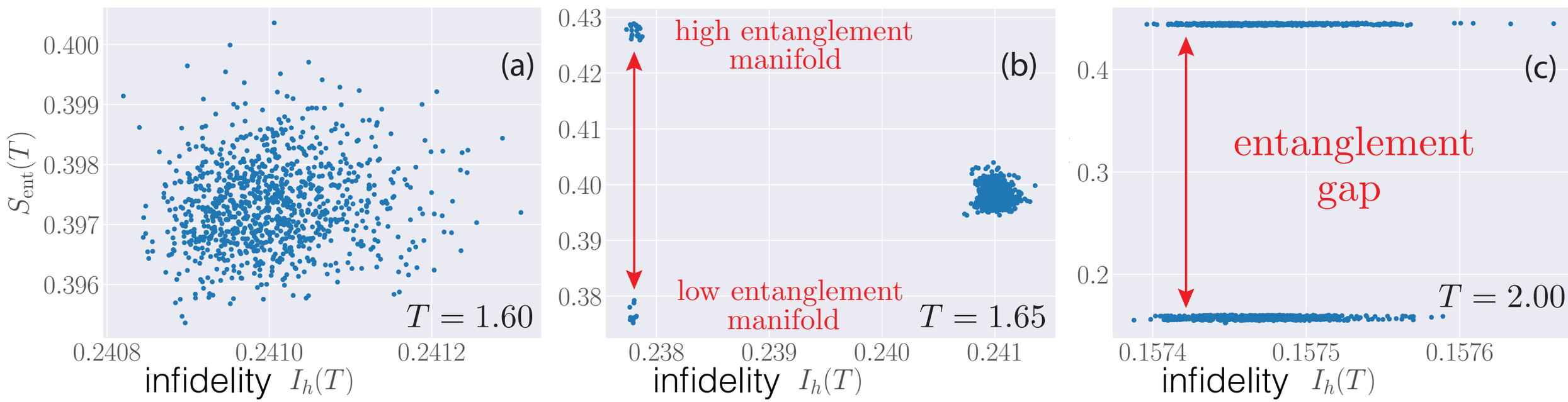
→ "magnetisation" of a single protocol: $m_h(T) = N_T^{-1} \sum_{n=1}^{N_T} h_x(n\delta t)$

→ "magnetisation" order parameter: $m(T) = \frac{1}{N_{\rm real}} \sum_{\alpha=1}^{N_{\rm real}} |m_{h^\alpha}(T)|$

PRA 97, 052114 (2018)

Implications for Physics

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$

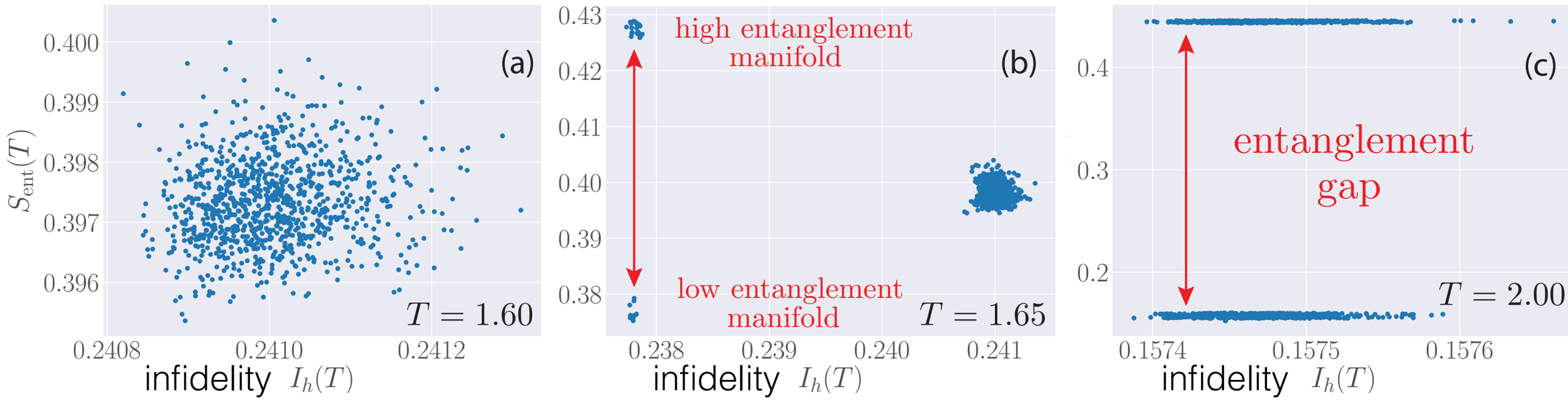


PRA 97, 052114 (2018)

Implications for Physics

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$

→ symmetry breaking ***on top of glassy*** low-infidelity manifold

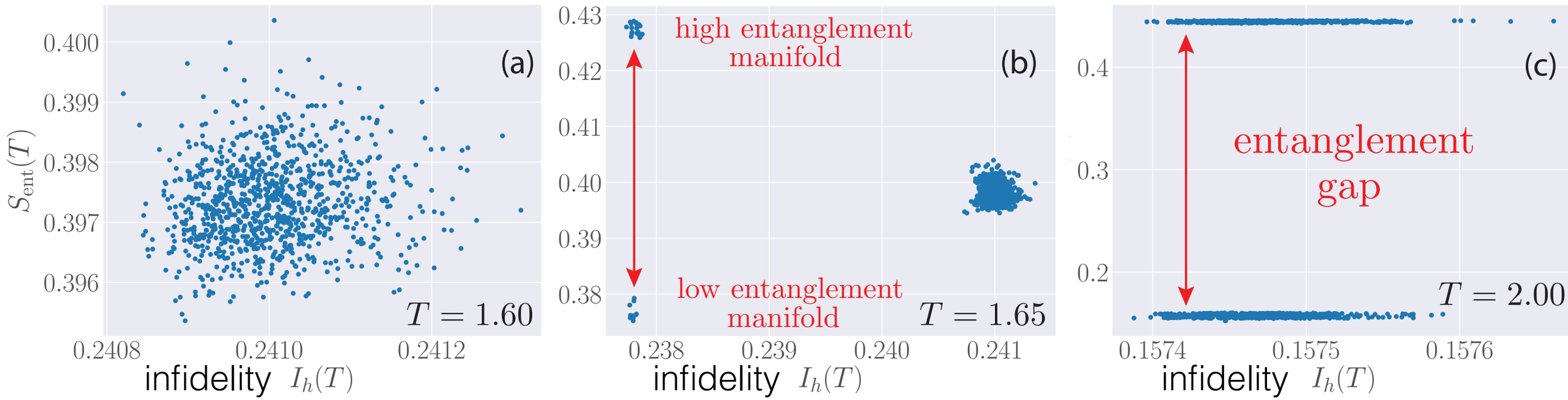


PRA 97, 052114 (2018)

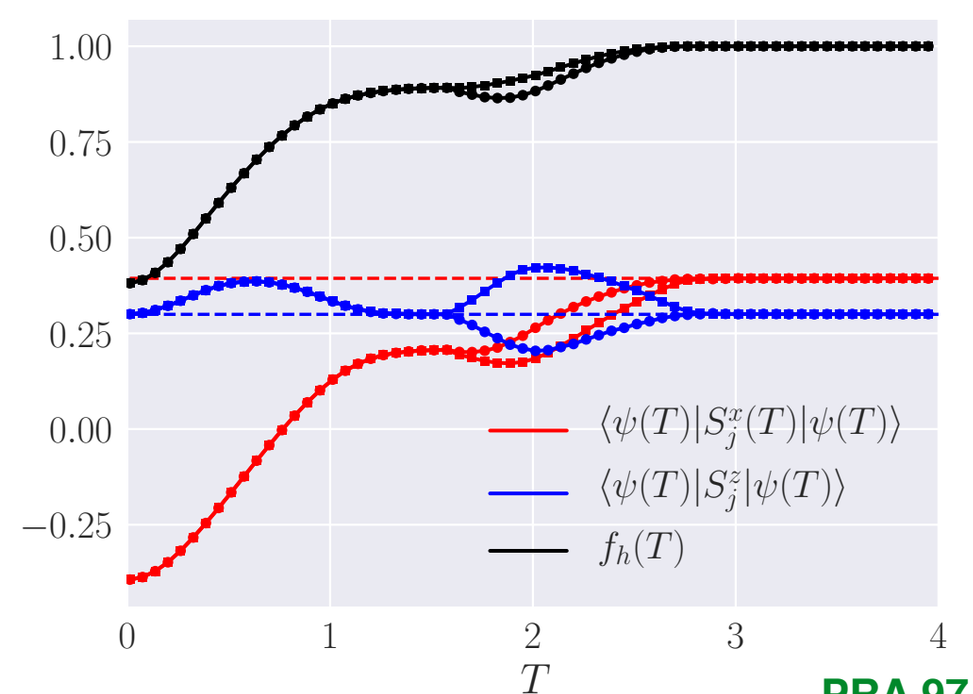
Implications for Physics

$$H(t) = -2S_1^z S_2^z - (S_1^z + S_2^z) - h_x(t)(S_1^x + S_2^x)$$

→ symmetry breaking ***on top of glassy*** low-infidelity manifold



→ can be detected in expectations of local observables



Marin Bukov

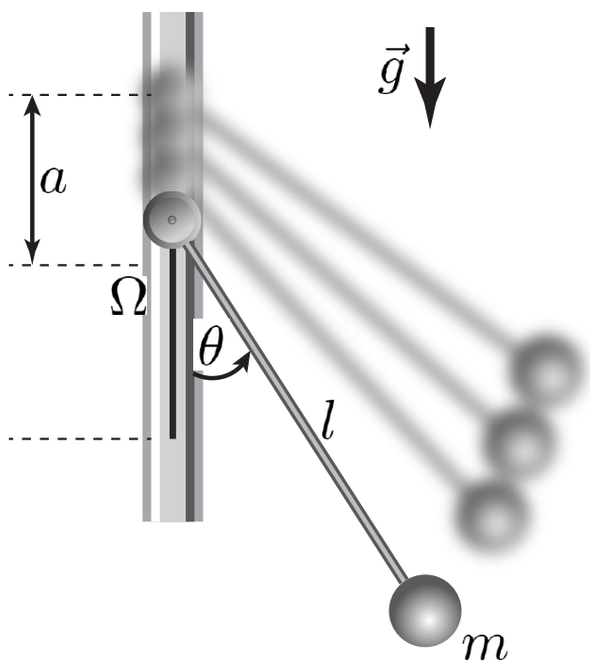
PRA 97, 052114 (2018)

Berkeley
UNIVERSITY OF CALIFORNIA

Example:

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

→ Kapitza oscillator



Marin Bukov

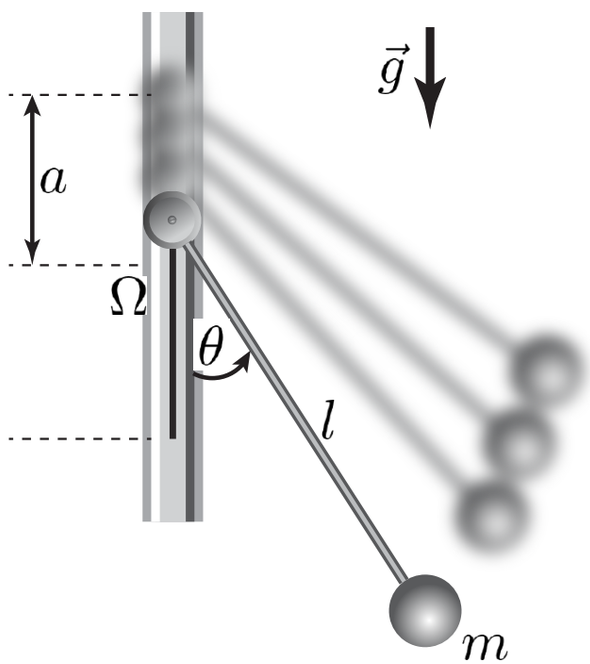
bluedwarf1127: YouTube (2009)

Berkeley
UNIVERSITY OF CALIFORNIA

Example:

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’?

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

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'?

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

→ 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!

→ change reference frames to "remove" strong coupling:

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

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!

→ change reference frames to "remove" strong coupling:

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

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

The quantum Kapitza oscillator

→ how do we understand 'dynamical stabilization'?

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

→ 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!

→ change reference frames to "remove" strong coupling:

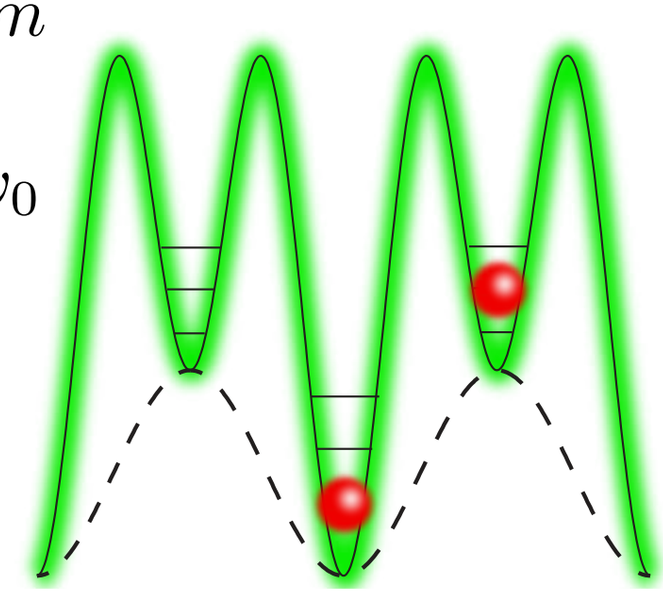
$$H_{\rm rot}(t) = V^\dagger(t) H_{\rm lab}(t) V(t) - iV^\dagger(t)\partial_t V(t) \qquad V(t) = {\rm e}^{iA\sin\Omega t\cos\theta}$$

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

→ time-average now easy to take

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

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



Marin Bukov

The quantum Kapitza oscillator

→ how do we understand 'dynamical stabilization'?

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

→ 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!

→ change reference frames to "remove" strong coupling:

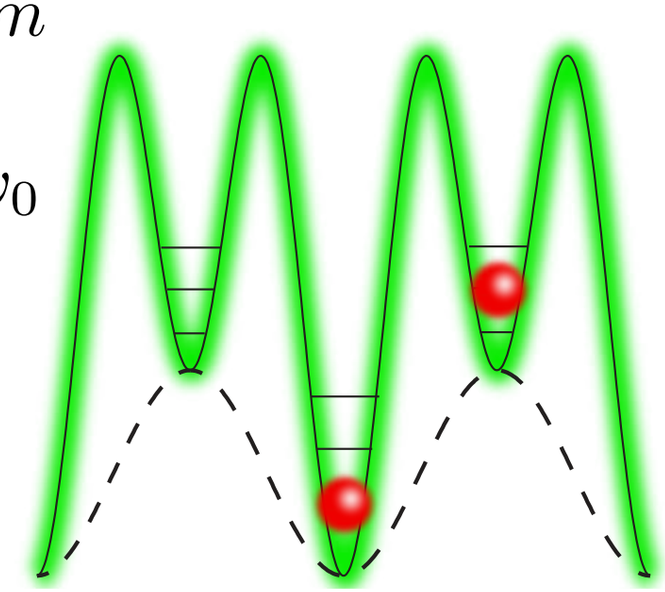
$$H_{\rm rot}(t) = V^\dagger(t) H_{\rm lab}(t) V(t) - iV^\dagger(t)\partial_t V(t) \qquad V(t) = {\rm e}^{iA\sin\Omega t\cos\theta}$$

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

→ time-average now easy to take

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

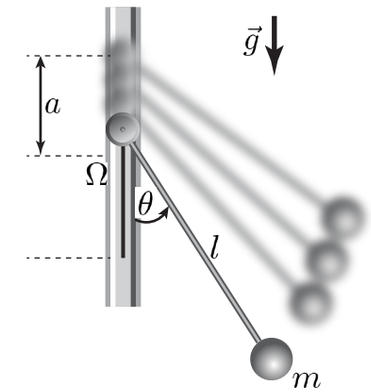
$$H_{\rm ave} = \frac{p_\theta^2}{2m} - m\omega_0^2\cos\theta + \frac{A^2}{4m}\cos 2\theta$$



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

Marin Bukov

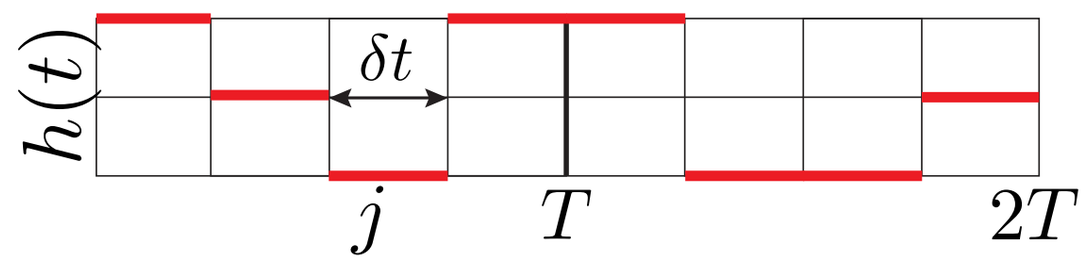
Floquet Control Problem



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

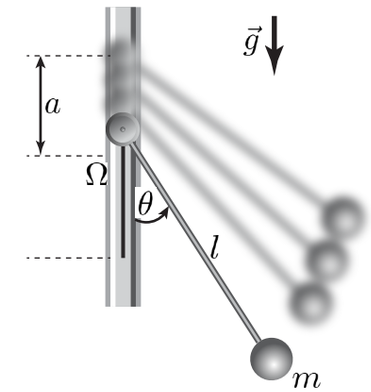
$$H_{\rm rot}(t) = H_0 + H_{\rm drive}(t) + H_{\rm control}(t)$$

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



PRB 98, 224305 (2018)

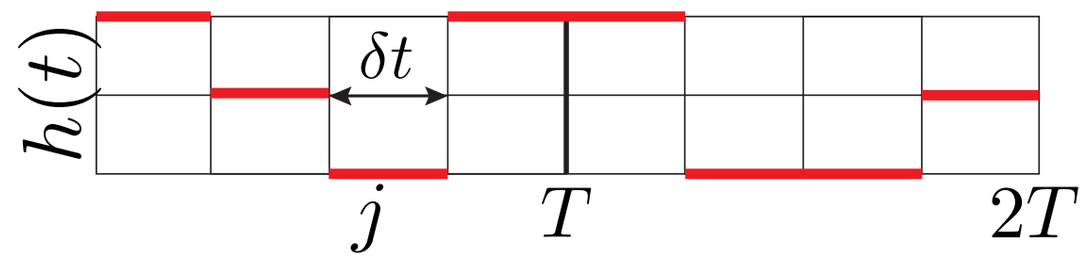
Floquet Control Problem



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

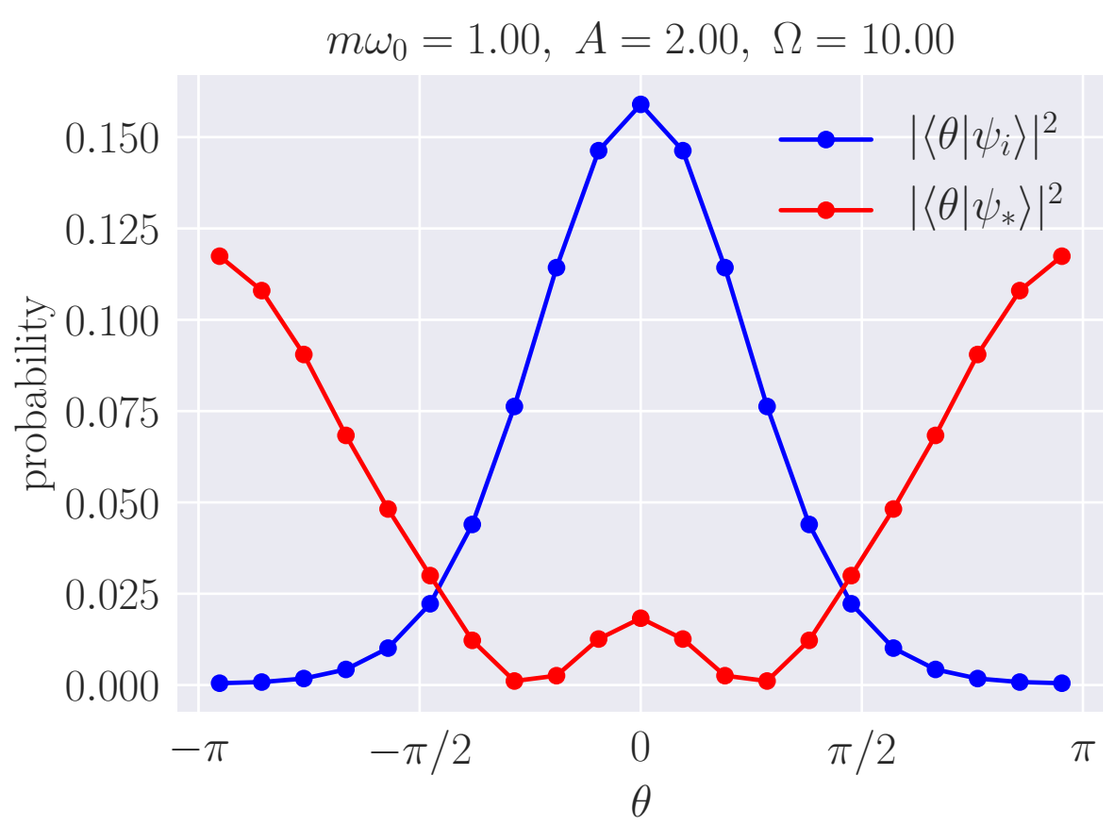
$$H_{\rm rot}(t) = H_0 + H_{\rm drive}(t) + H_{\rm control}(t)$$

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



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

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



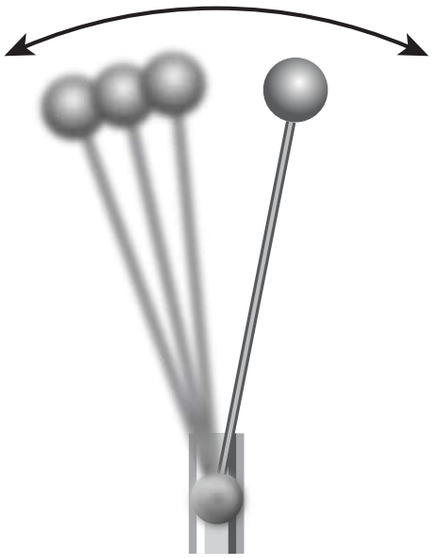
Marin Bukov

PRB 98, 224305 (2018)

Simulation of a Quantum Experiment

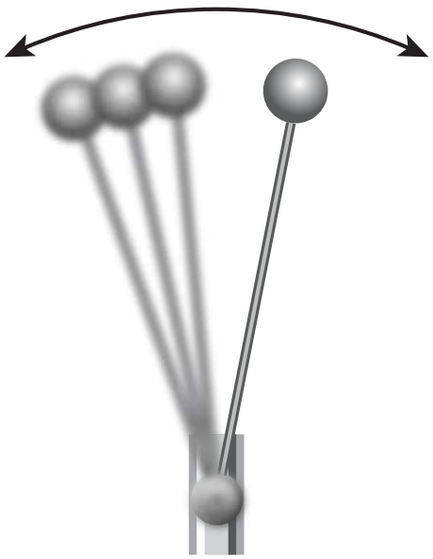
→ **no** direct **access** to quantum state:
"play game w/o looking at screen" (only know score)

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



PRB 98, 224305 (2018)

Simulation of a Quantum Experiment



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

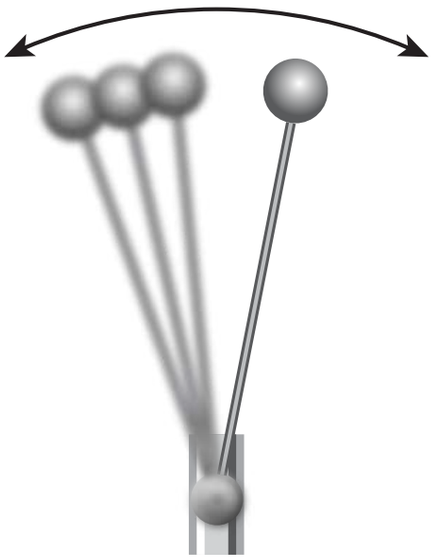
$$\{|\psi(t)\rangle : |\psi(t)\rangle = U_h(t,0)|\psi_i\rangle\} \hat{=} \{h(t) : |\psi_i\rangle\}$$

→ **probabilistic** quantum measurements

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

PRB 98, 224305 (2018)

Simulation of a Quantum Experiment



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

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

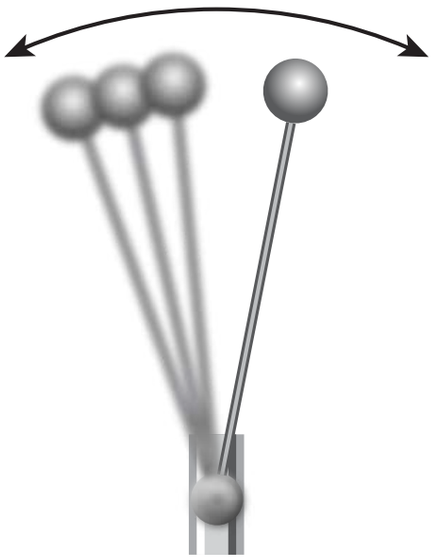
→ **probabilistic** quantum measurements

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

→ **uncertainty** in preparing initial state

PRB 98, 224305 (2018)

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

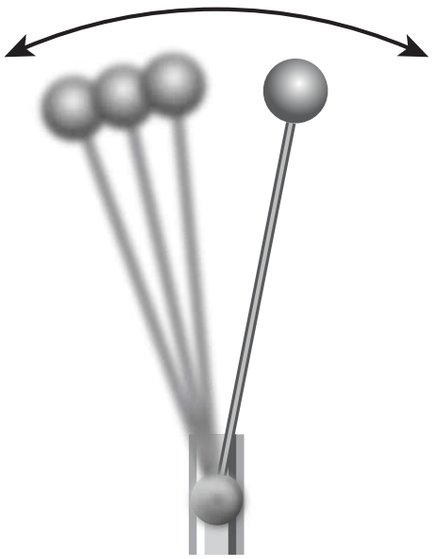
$+1$ with probability $F_h(t_f) = |\langle\psi(t_f)|\psi_*\rangle|^2$
-1 otherwise

→ **uncertainty** in preparing initial state

→ occasional **failure** of control apparatus

PRB 98, 224305 (2018)

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

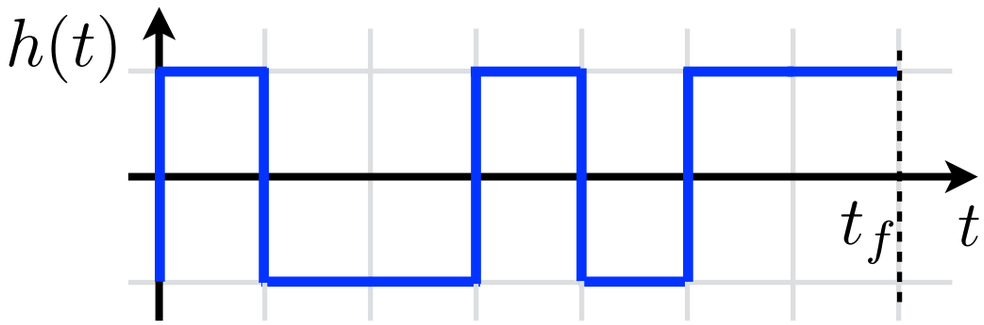
→ **uncertainty** in preparing initial state

→ 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

PRB 98, 224305 (2018)

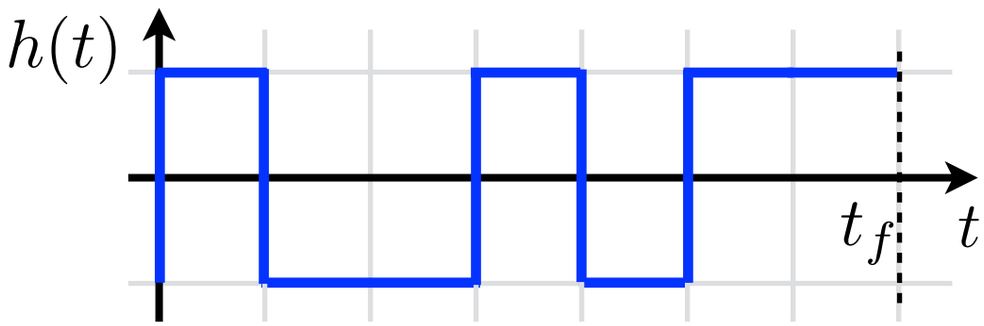
Let's give this "game" a try!



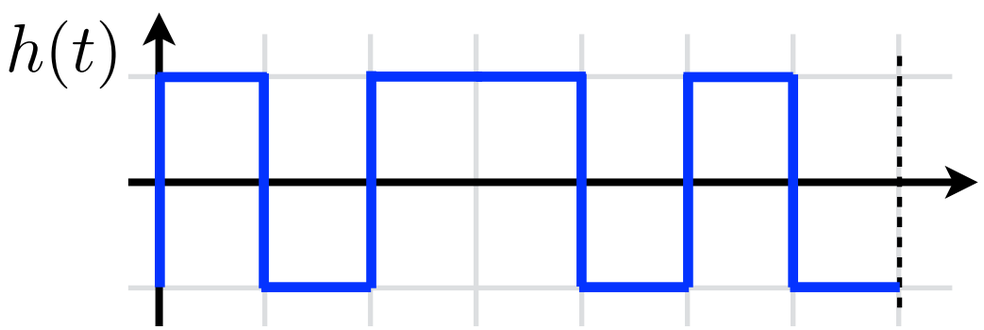
measurement: -1

PRB 98, 224305 (2018)

Let's give this "game" a try!



measurement: -1

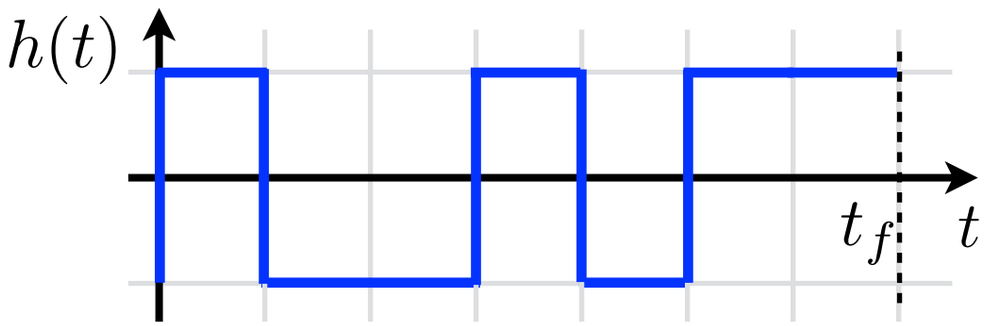


measurement: $+1$

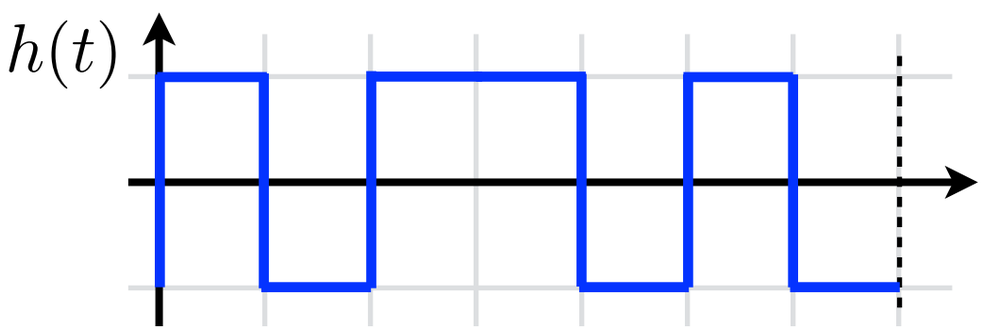
(different final state: different probability to be in the target state)

PRB 98, 224305 (2018)

Let's give this "game" a try!



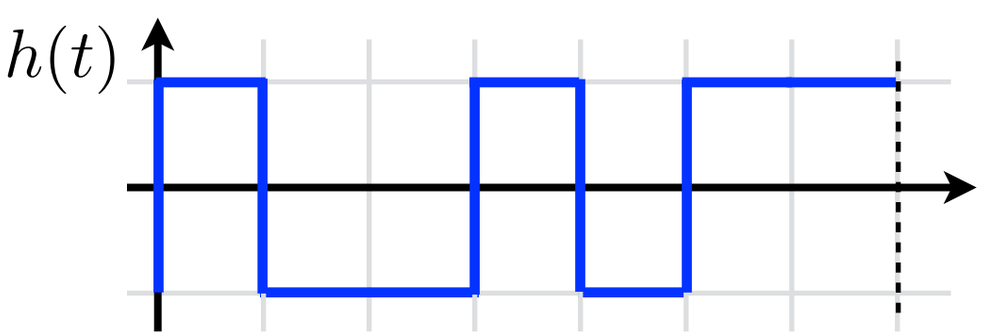
measurement: -1



measurement: $+1$

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

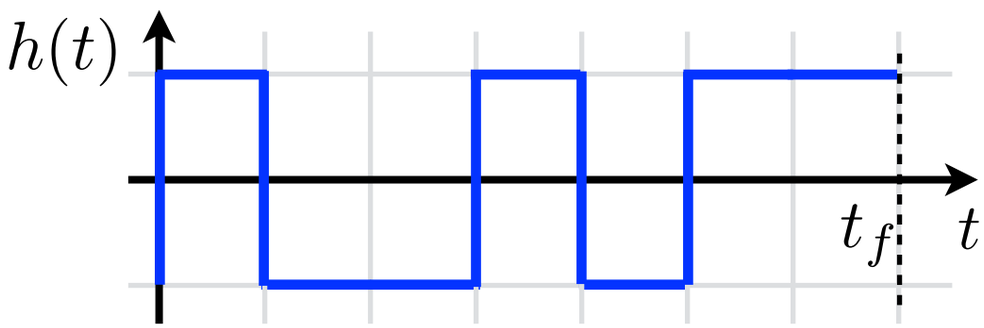
→ repeat protocol!



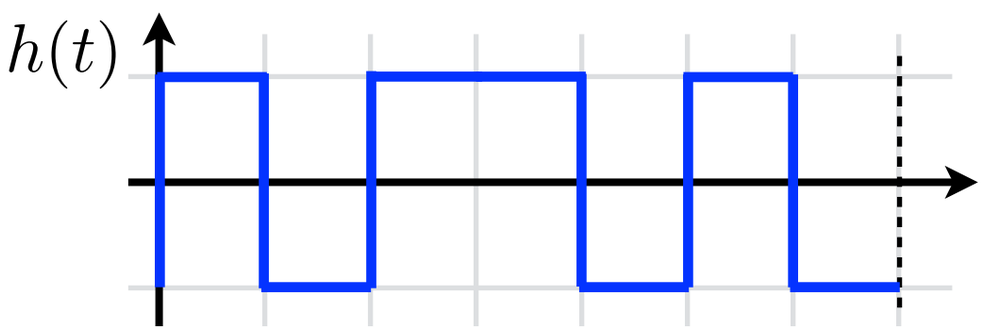
measurement: $+1$

PRB 98, 224305 (2018)

Let's give this "game" a try!



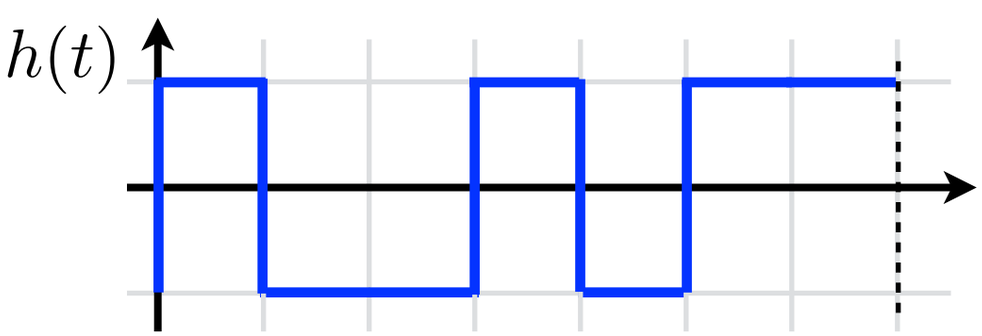
measurement: -1



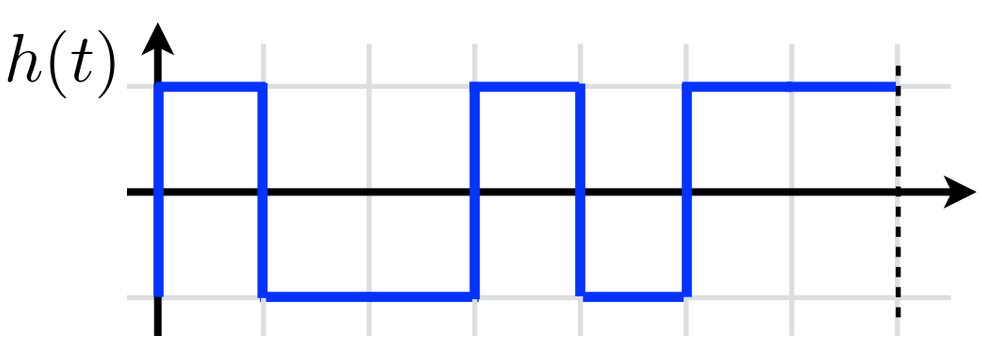
measurement: $+1$

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

→ repeat protocol!



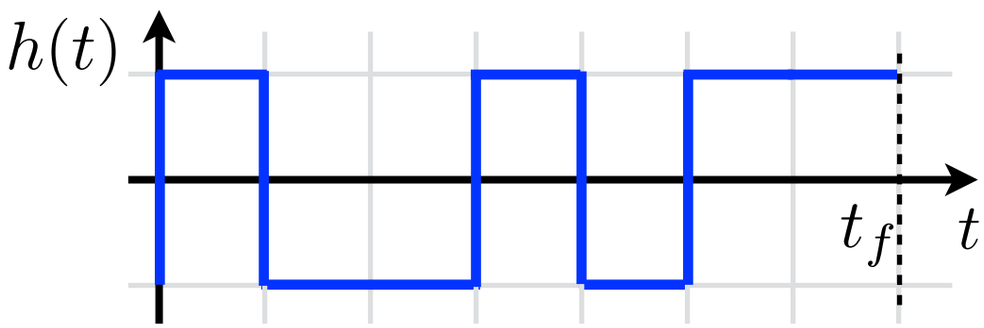
measurement: $+1$



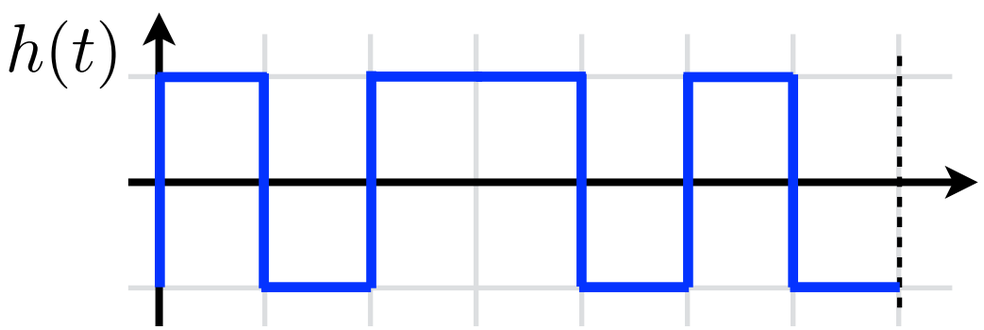
measurement: -1

PRB 98, 224305 (2018)

Let's give this "game" a try!



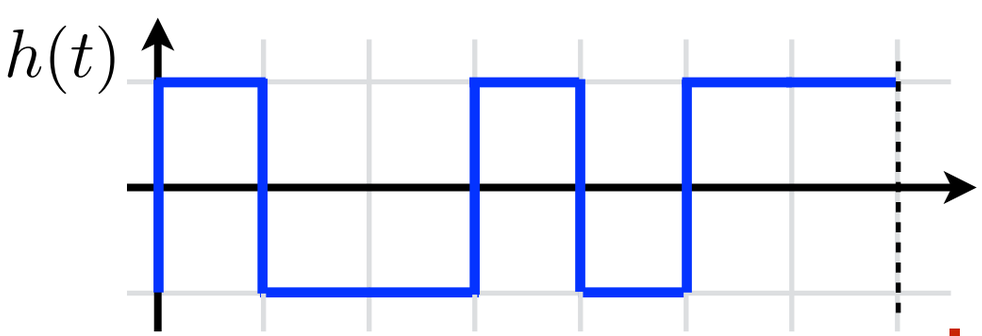
measurement: -1



measurement: $+1$

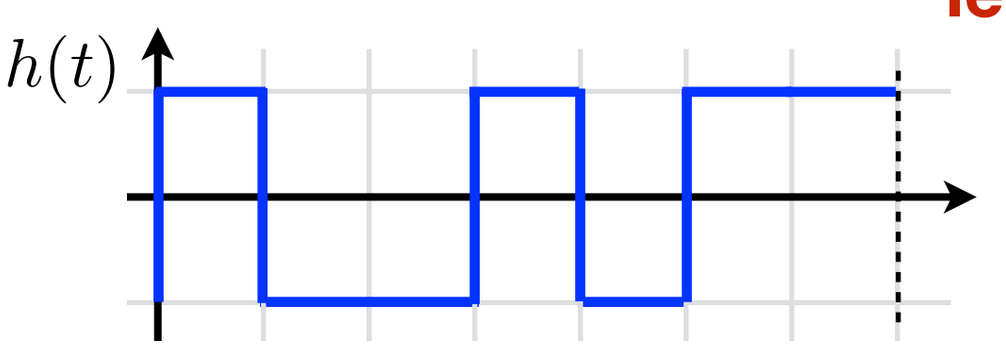
(different final state: different probability to be in the target state)

→ repeat protocol!



measurement: $+1$

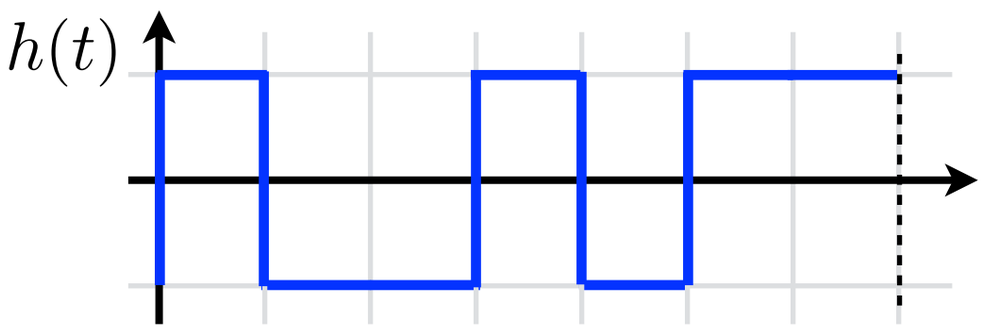
learn from noisy, nondeterministic rewards



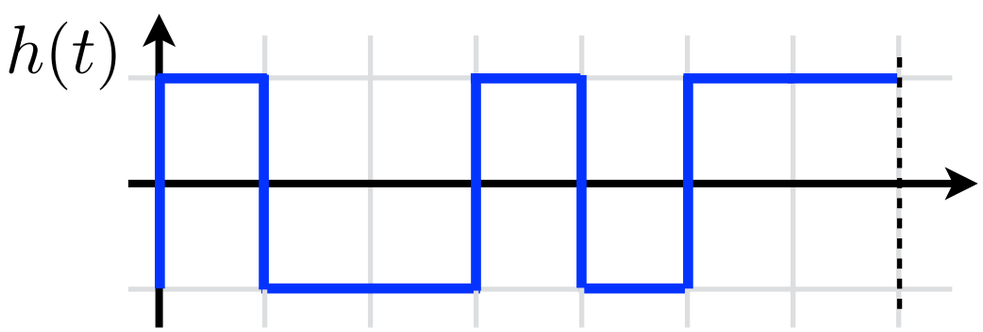
measurement: -1

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

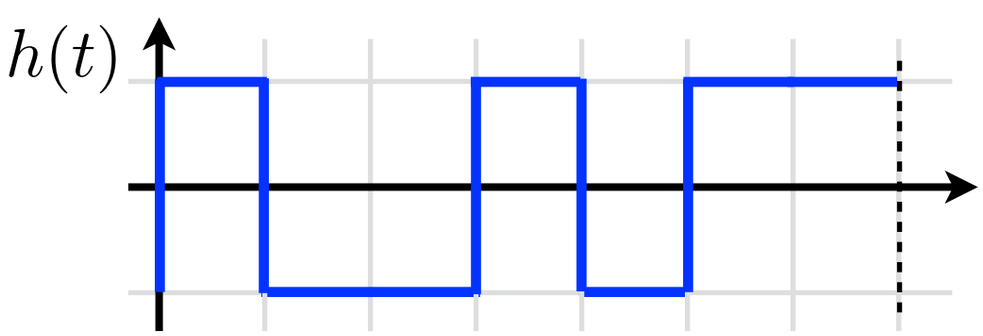
→ repeat protocol again!



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



measurement: $F_h = 0.592$

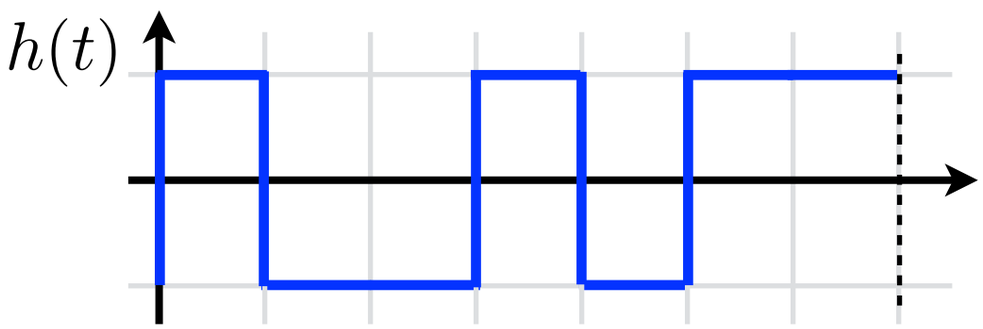


measurement: $F_h = 0.668$

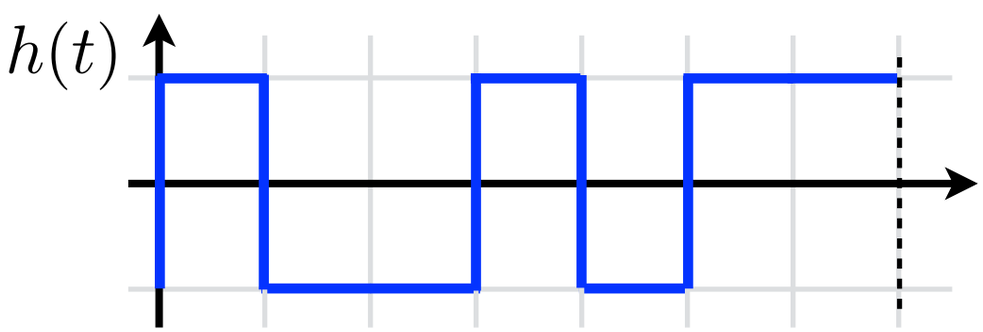
PRB 98, 224305 (2018)

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

→ repeat protocol again!

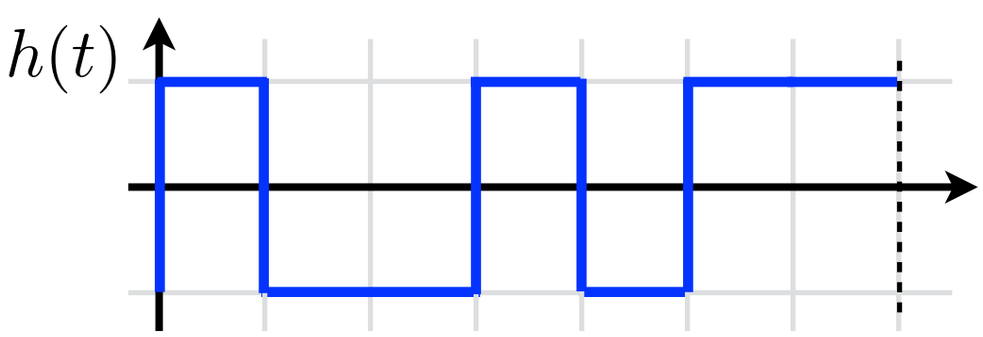


measurement:
$F_h = |\langle\psi(T)|\psi_*\rangle|^2 = 0.632$



measurement: $F_h = 0.592$

initial state could not be prepared perfectly: more headache!

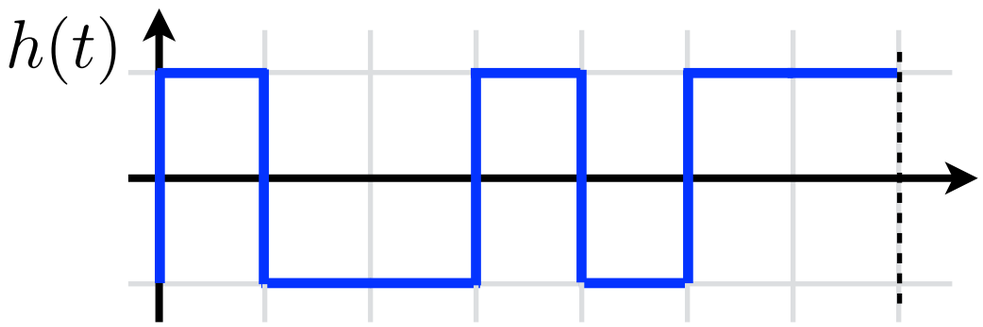


measurement: $F_h = 0.668$

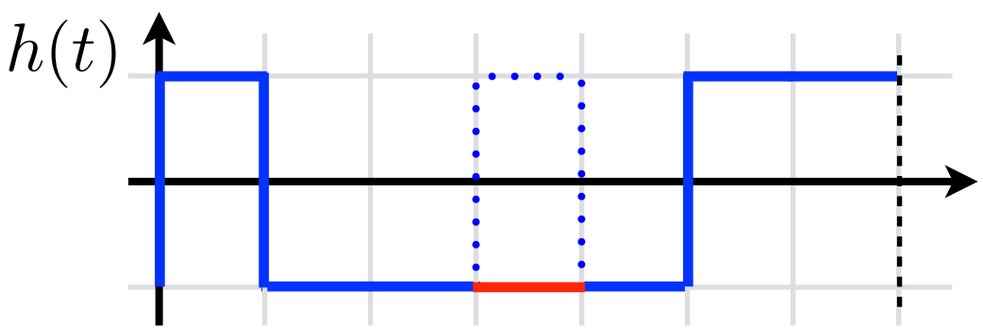
PRB 98, 224305 (2018)

what if we fix the initial state:

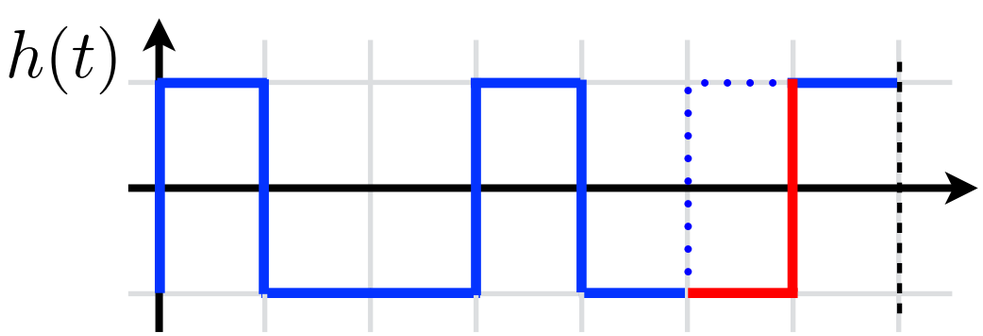
→ repeat protocol again!



measurement: $F_h = 0.627$



measurement: $F_h = 0.572$

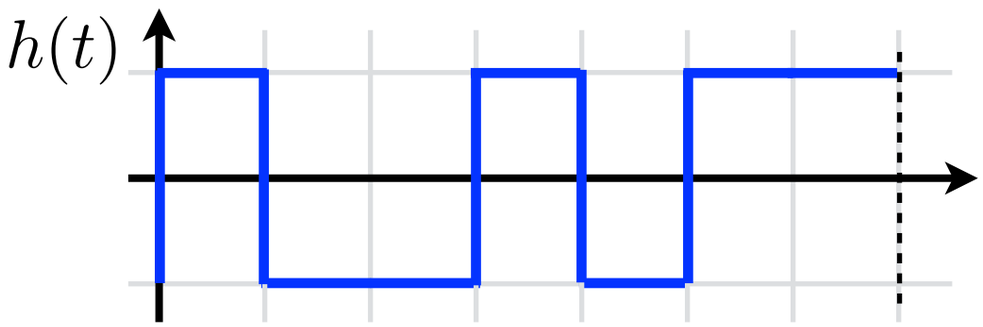


measurement: $F_h = 0.657$

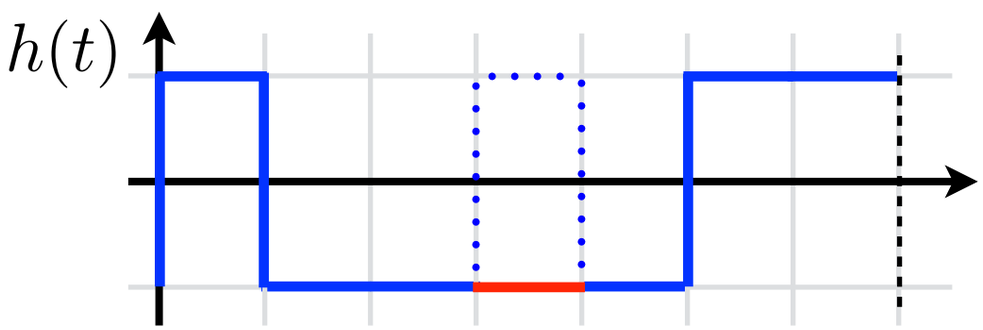
PRB 98, 224305 (2018)

what if we fix the initial state:

→ repeat protocol again!

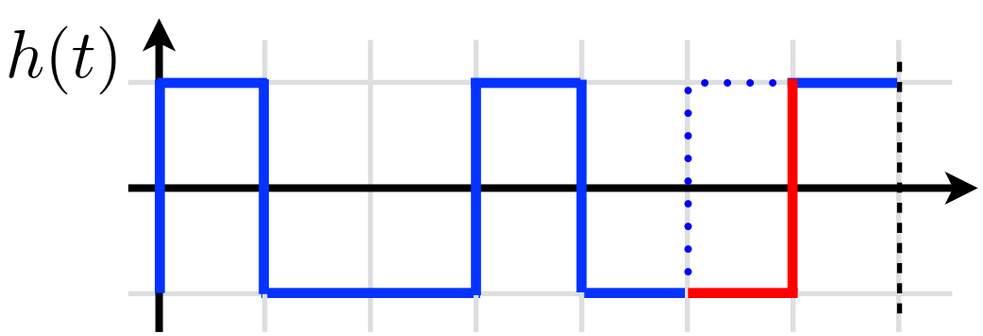


measurement: $F_h = 0.627$



measurement: $F_h = 0.572$

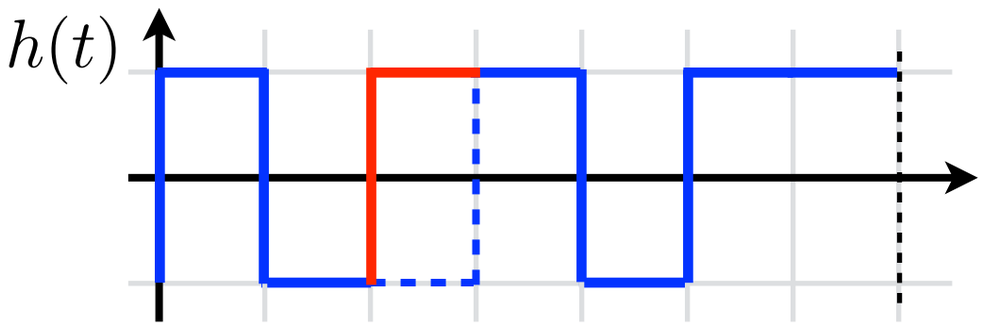
control apparatus failed: it can't be!



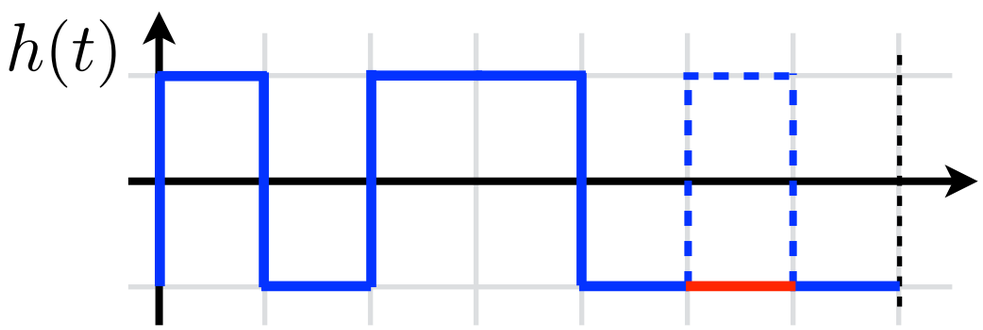
measurement: $F_h = 0.657$

PRB 98, 224305 (2018)

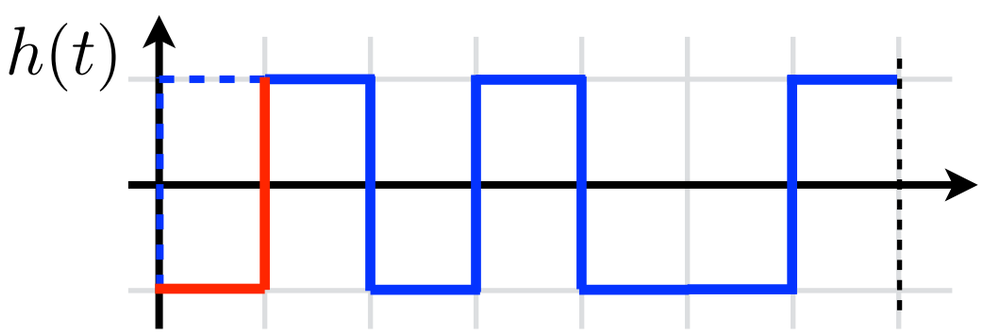
The Cruel Reality: all together (and probably much more!)



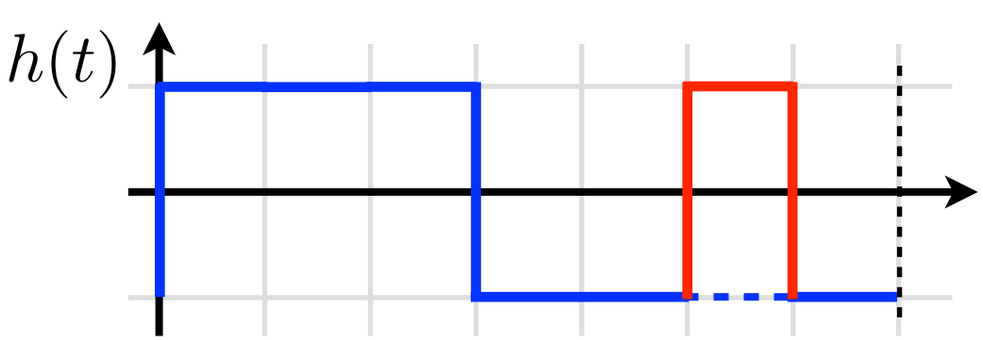
measurement: -1



measurement: $+1$



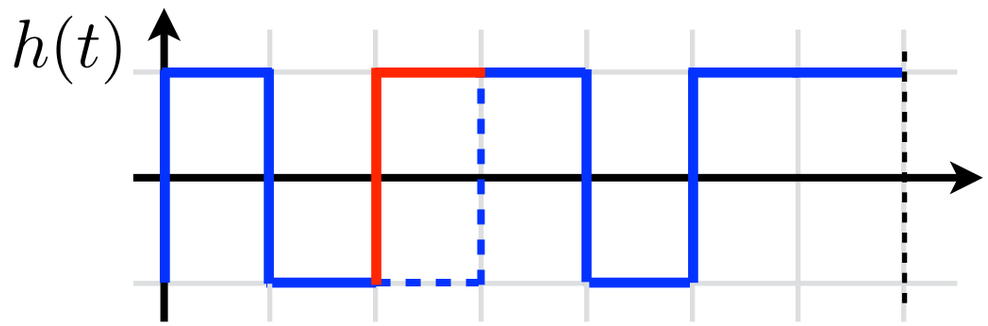
measurement: -1



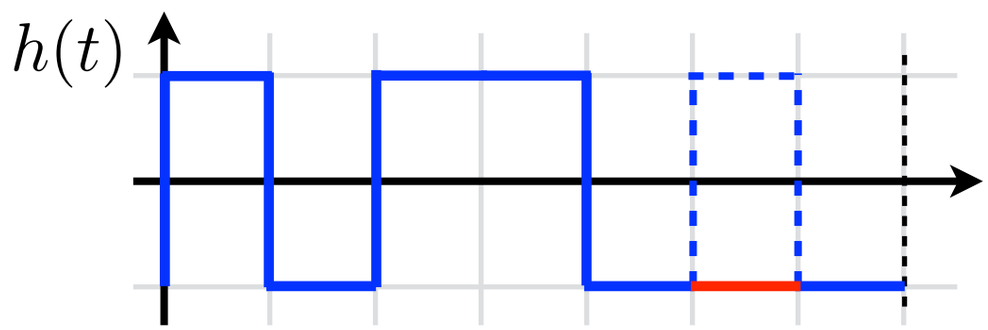
measurement: -1

PRB 98, 224305 (2018)

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

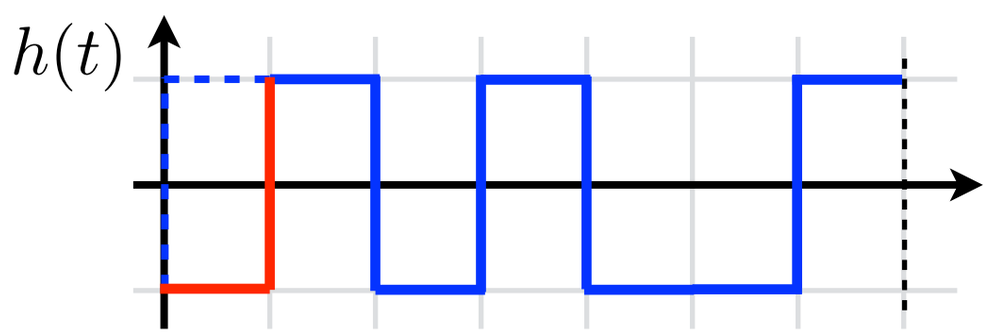


measurement: -1

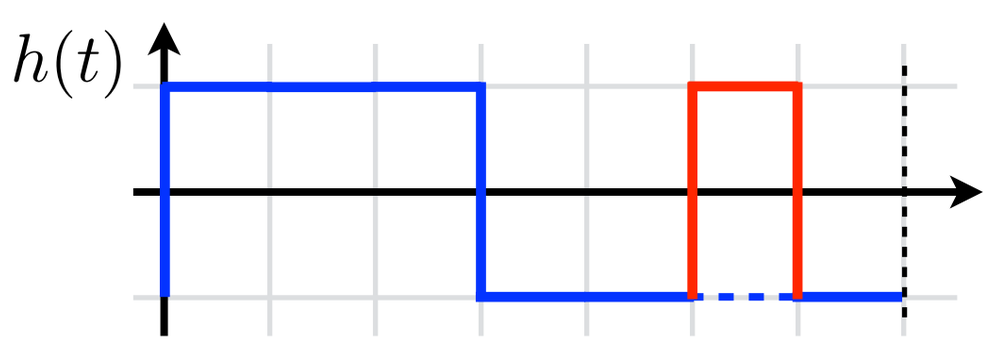


measurement: $+1$

extremely tedious task!



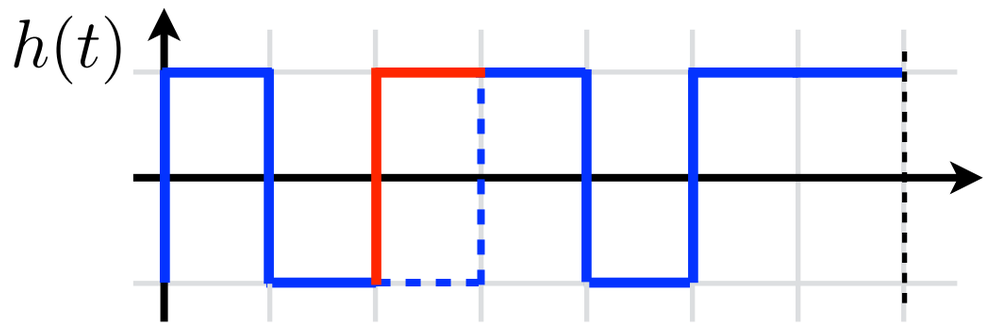
measurement: -1



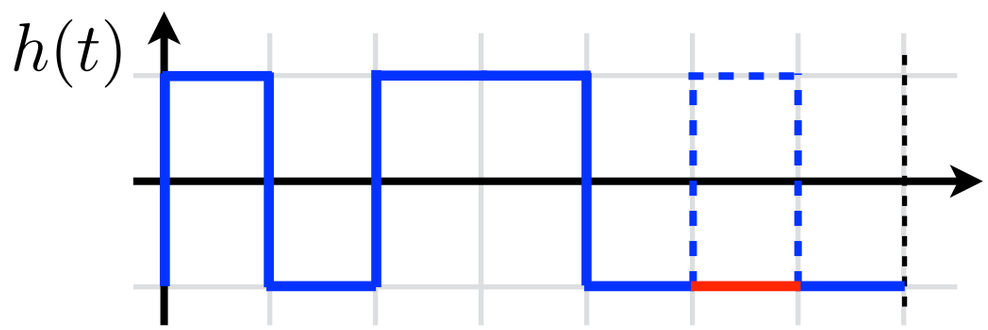
measurement: -1

PRB 98, 224305 (2018)

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

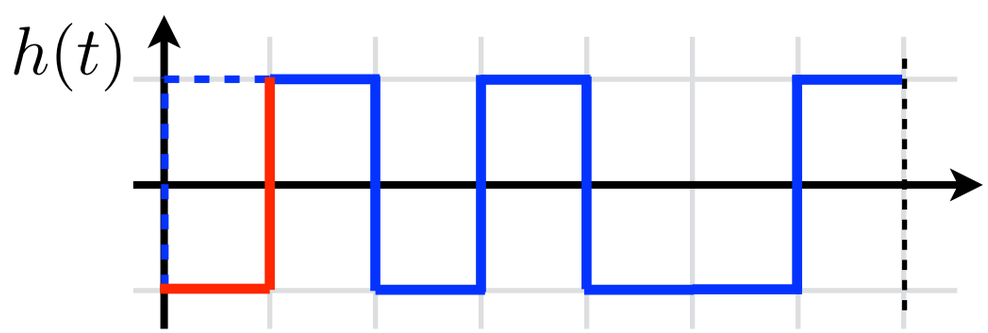


measurement: -1



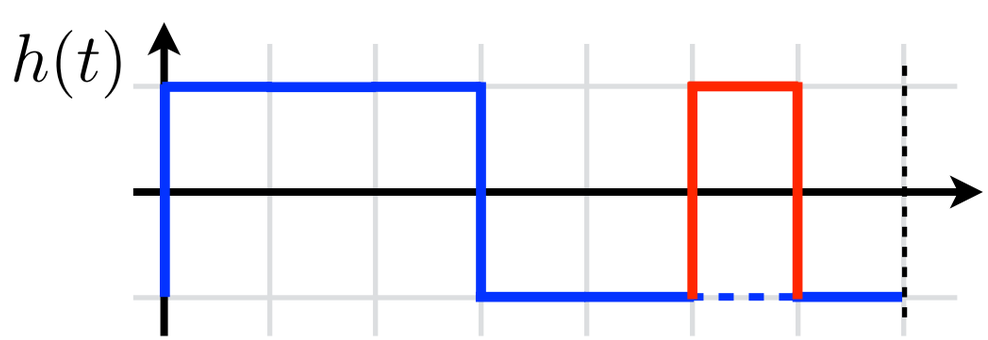
measurement: $+1$

extremely tedious task!



measurement: -1

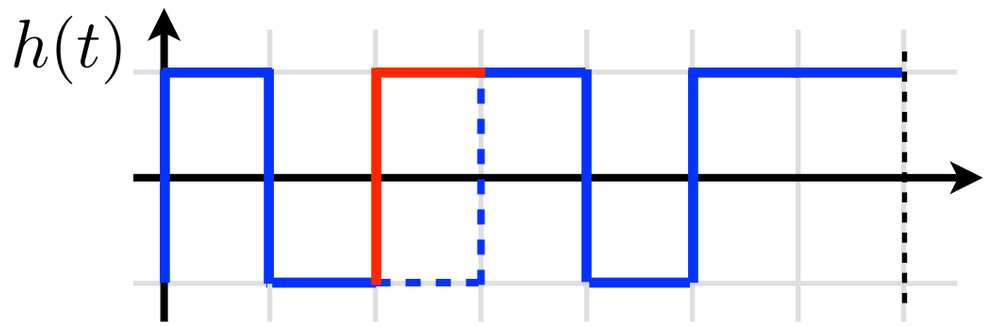
how do we solve it efficiently?



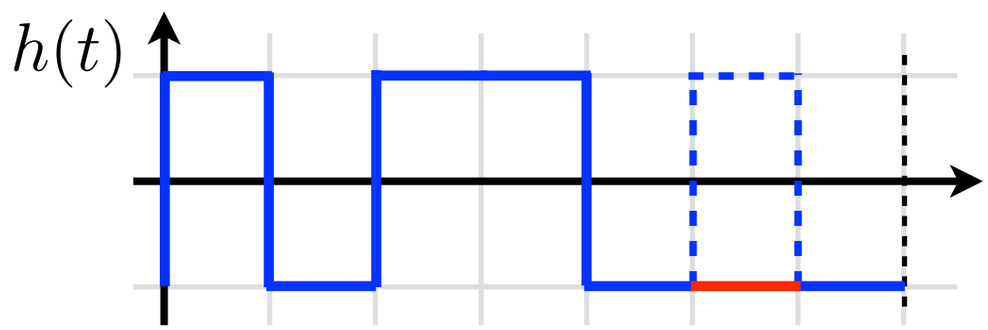
measurement: -1

PRB 98, 224305 (2018)

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

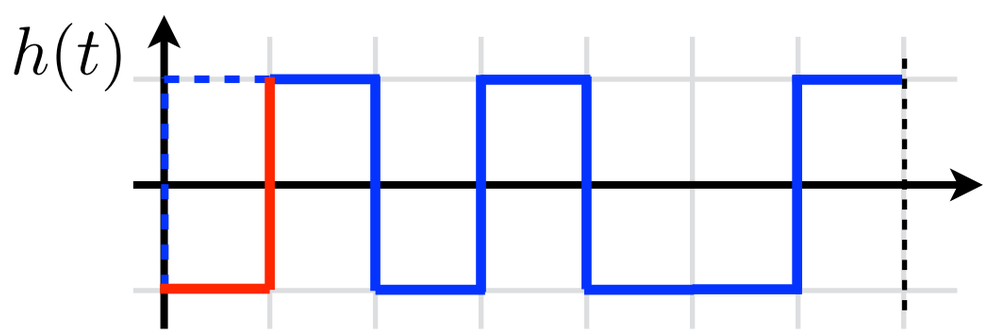


measurement: -1



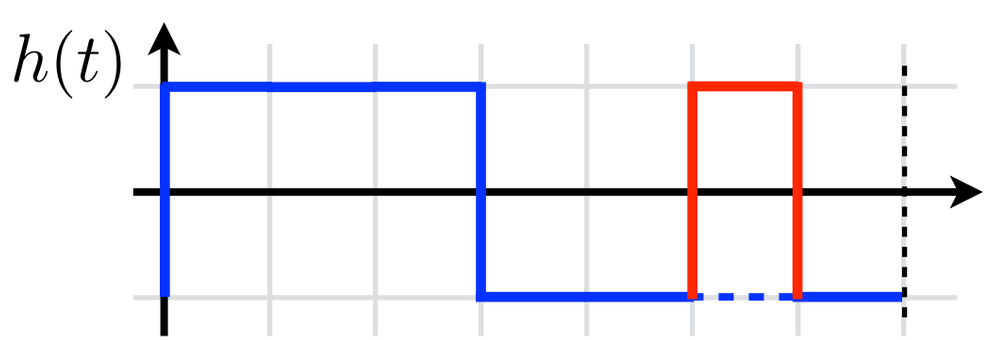
measurement: $+1$

extremely tedious task!



measurement: -1

how do we solve it efficiently?

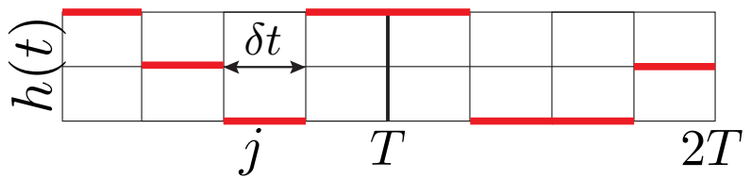


measurement: -1

can we automate it?

PRB 98, 224305 (2018)

Reinforcement Learning to Prepare the Inverted Position Floquet State

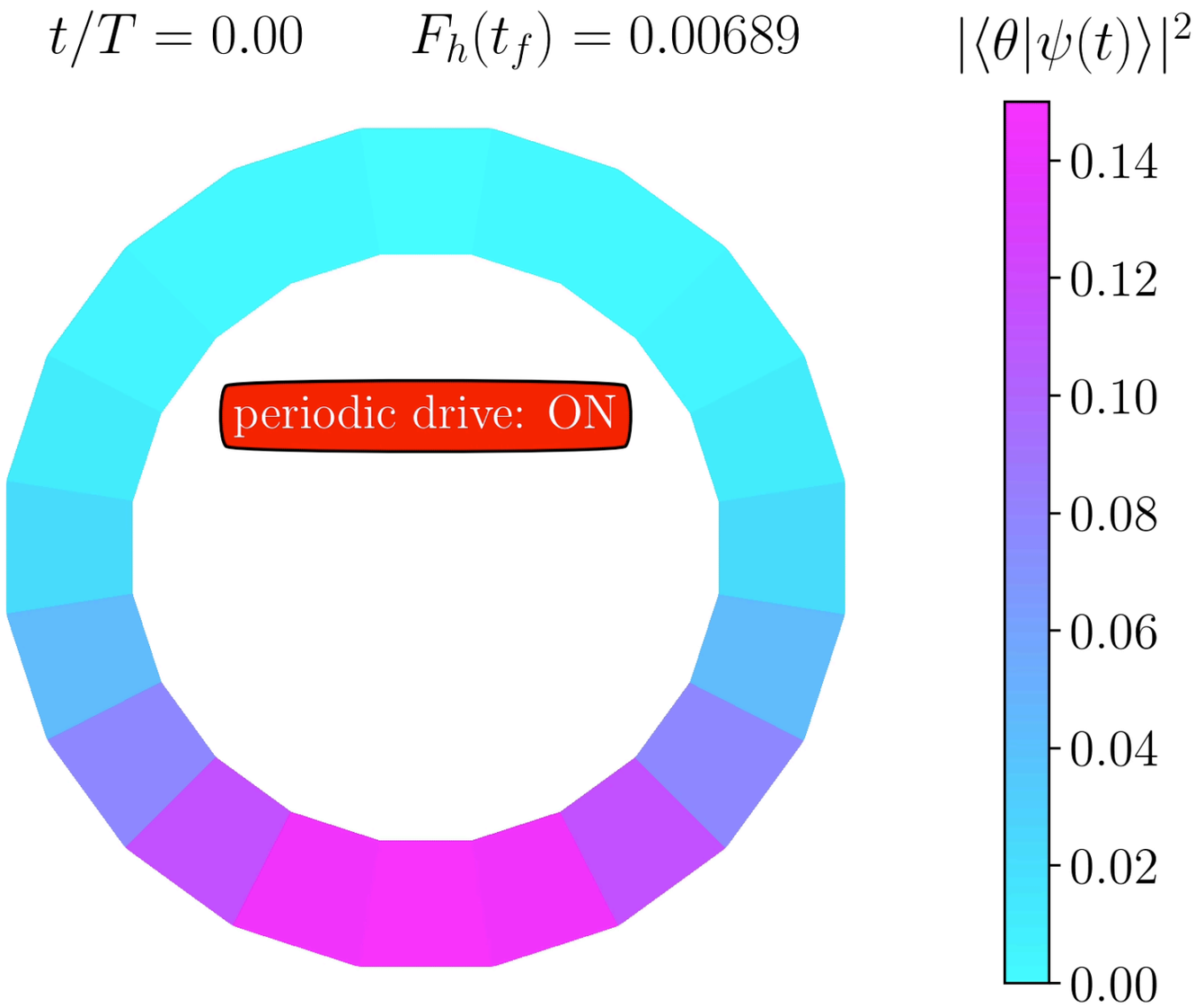


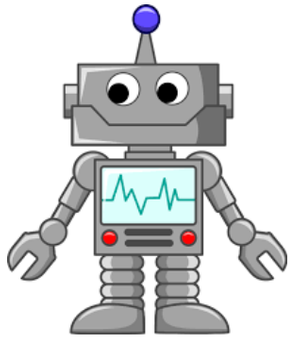
15 driving cycles (periods), 120 steps (8 per period)

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

quantum Kapitza oscillator

$t/T = 0.00$ $\quad F_h(t_f) = 0.00689$



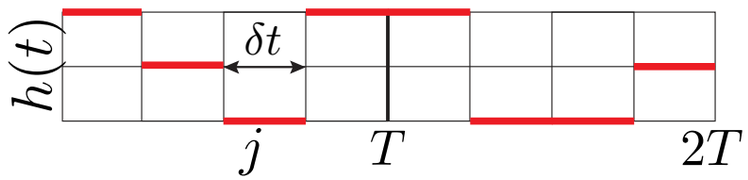


$h_{\max}/(m\omega_0) = 4.0$ $\quad \Omega/(m\omega_0) = 10.0$ $\quad A/(m\omega_0) = 2.0$

Marin Bukov

PRB 98, 224305 (2018)

Reinforcement Learning to Prepare the Inverted Position Floquet State

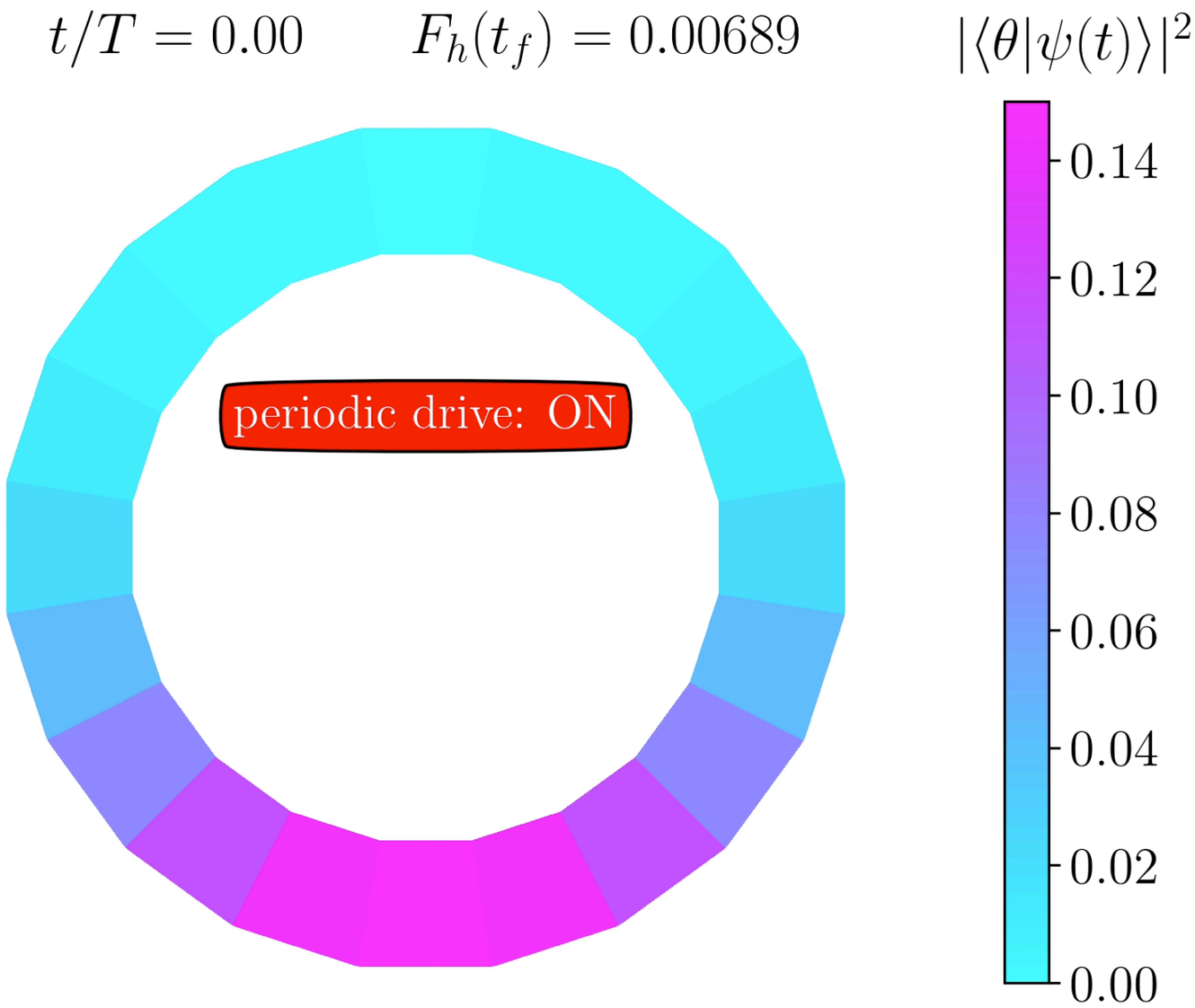


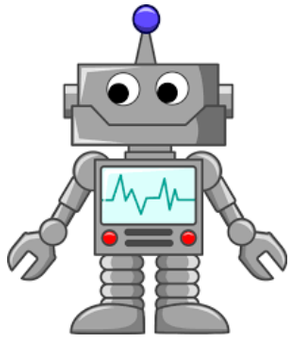
15 driving cycles (periods), 120 steps (8 per period)

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

quantum Kapitza oscillator

$t/T = 0.00$ $\quad F_h(t_f) = 0.00689$



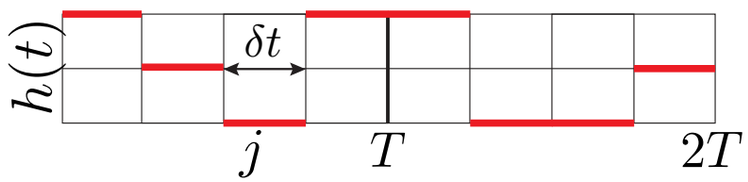


$h_{\max}/(m\omega_0) = 4.0$ $\quad \Omega/(m\omega_0) = 10.0$ $\quad A/(m\omega_0) = 2.0$

Marin Bukov

PRB 98, 224305 (2018)

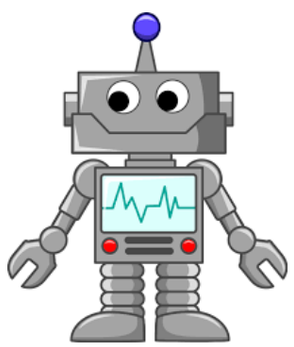
Reinforcement Learning
Classical Kapitza Pendulum



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

Kapitza pendulum

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

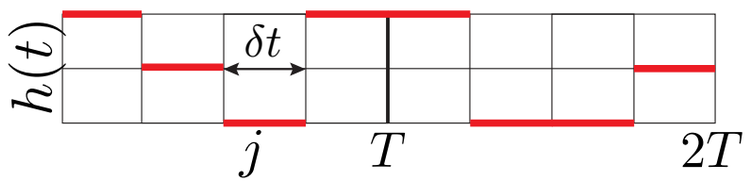


periodic drive: ON

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

Marin Bukov

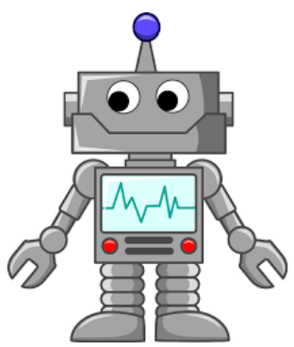
Reinforcement Learning
Classical Kapitza Pendulum



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

Kapitza pendulum

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

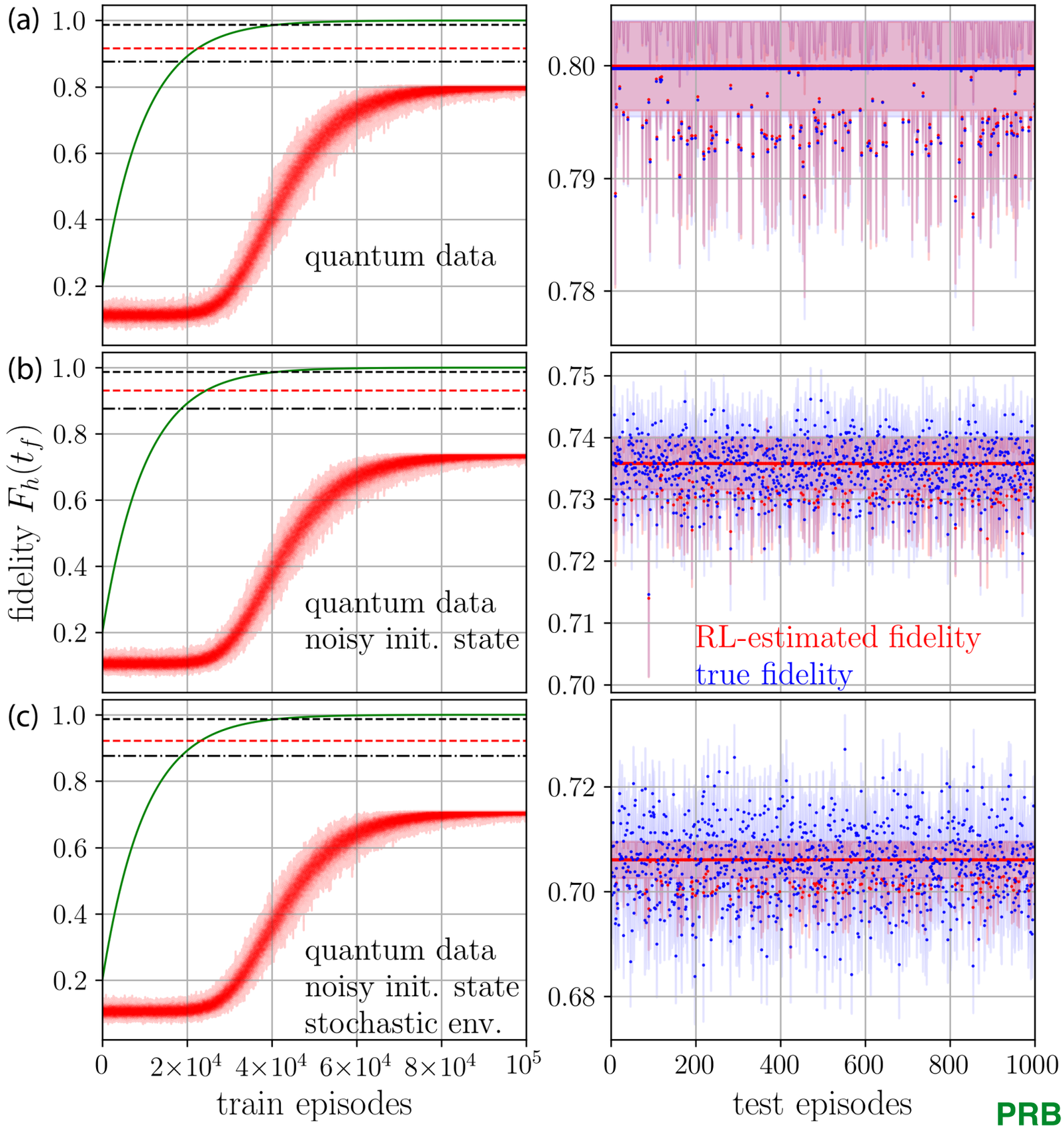


periodic drive: ON

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

Marin Bukov

Kapitza Learning Curves



Marin Bukov

PRB 98, 224305 (2018)