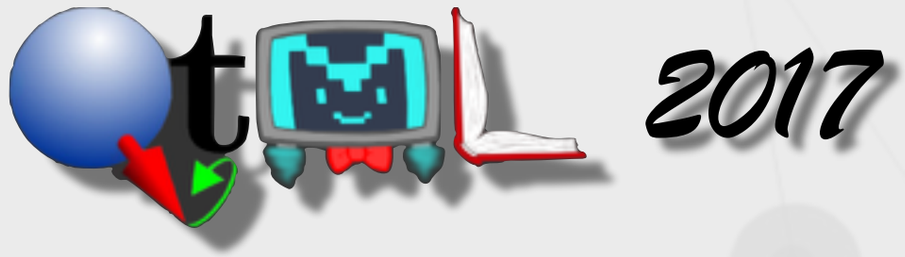
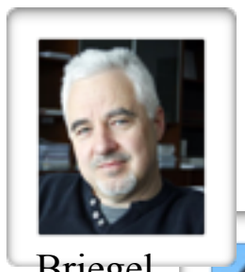
(Advances in) Quantum Reinforcement Learning

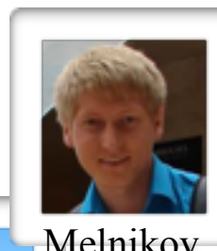
Vedran Dunjko
vedran.dunjko@mpq.mpg.de

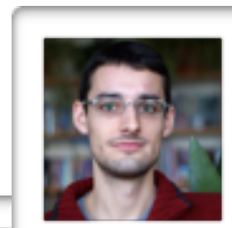


Acknowledgements:

theoretical physics


Briegel


Melnikov


Tiersch


Makmal


Friis


Boyajian

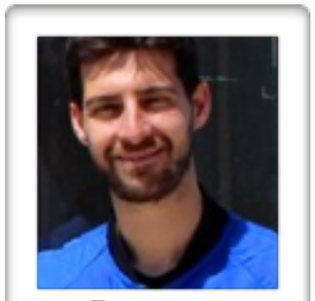

Orsucci


Poulsen Nautrup


Piater


Hangl

intelligent & interactive systems


Paparo


Martin-Delgado


Krenn

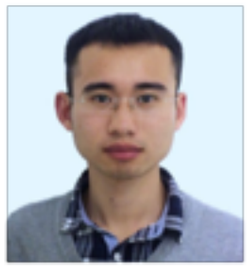

Zeilinger


Taylor

NIST

Joint Quantum Institute


Liu


Wu

Part 1: Intro

- Machine learning, Reinforcement learning (and AI)
- Quantum machine learning

Part 2: (a perspective on) Quantum Reinforcement Learning

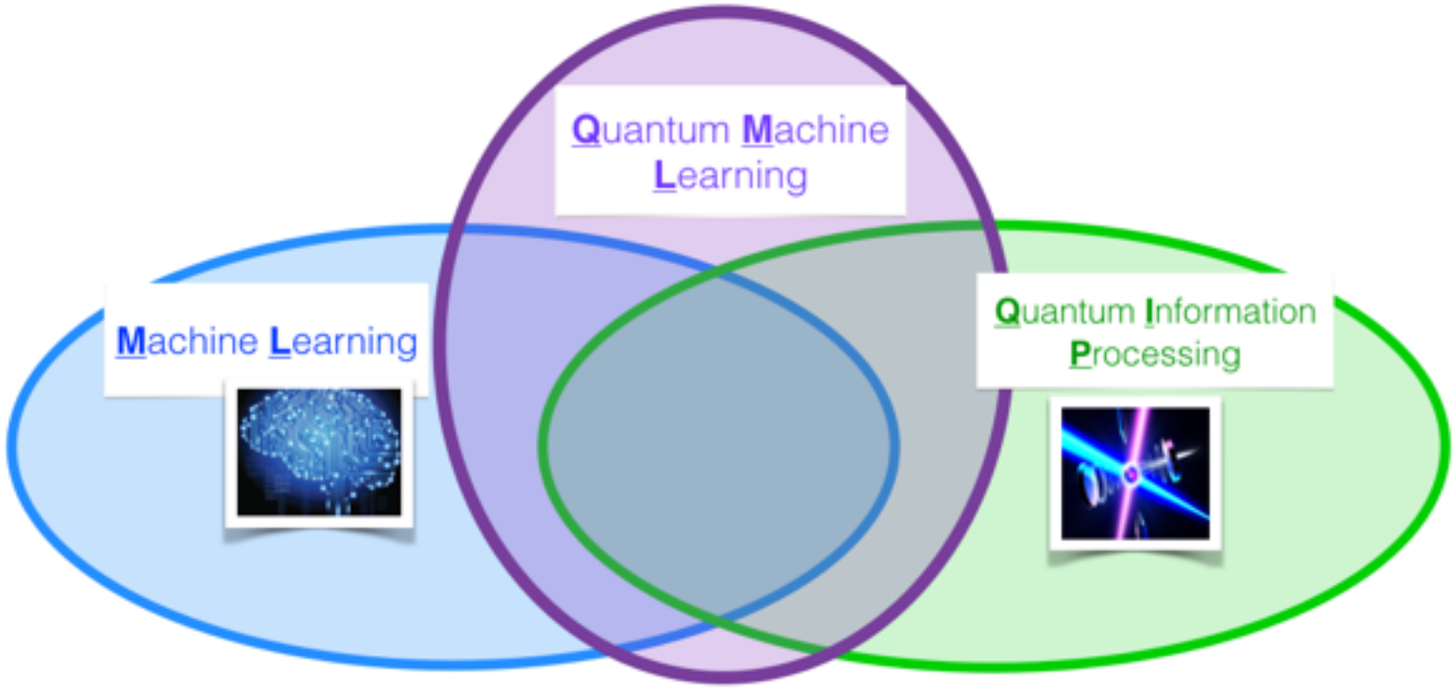
- Quantum-enhanced agents and quantum-accesible environments
- Quantum enhancements of learning agents:
  oraculization, and *three flavours of improvements*

Part 1: Quantum information and machine learning

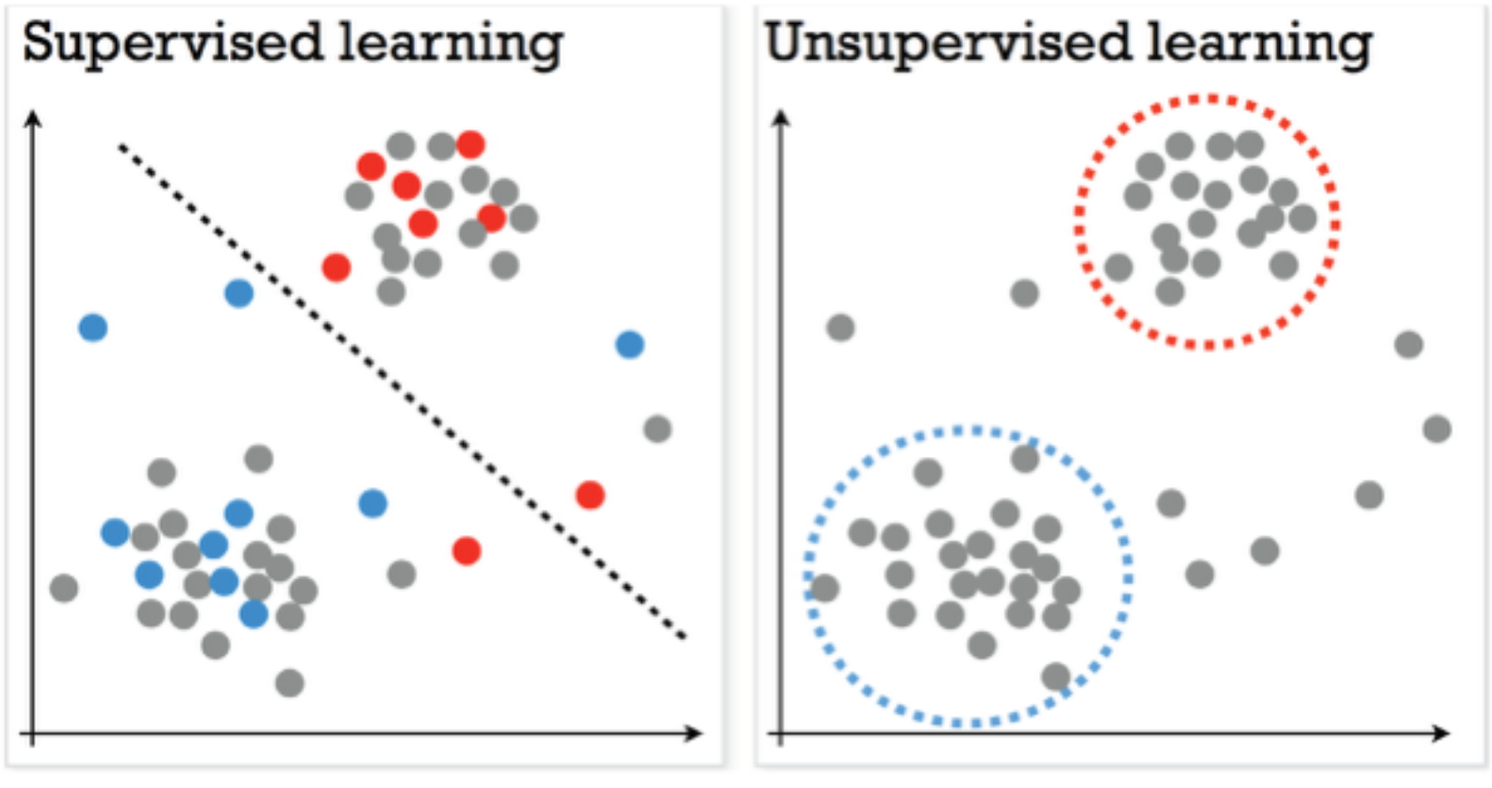
Quantum machine learning *(plus)*

- **QIP→ML** (quantum-enhanced ML) [’94; ‘12]
- **ML→QIP** (ML applied to QIP problems) [70’s? ’10]
- **QIP⇿ML** (generalizations of learning concepts) [’00? ’06]

- ML-insipred QIP
- QIP/QM inspired ML
- *beyond (Q. AI)?*



But… what <u>is</u> Machine Learning?



Learning structure in *P(data)*, give samples from *P(data)*

Learning *P(labels|data)* given samples from *P(data,labels)*

mostly about inferring information from data (about the source)

- *medicine*
- *stock markets*
- *climate/weather*

Who cares?

Unsupervised learning

Supervised learning

ML/Data mining in the spotlight

industry 4.0

“This Is the Year of the Machine Learning Revolution
To predict the future, we look at the data we have about the past.”

give samples from *P(data)*

Learning *P(labels|data)* given samples from *P(data,labels)*

mostly about inferring information from data (about the *source*)
-medicine
-stock-markets
-climate/weather

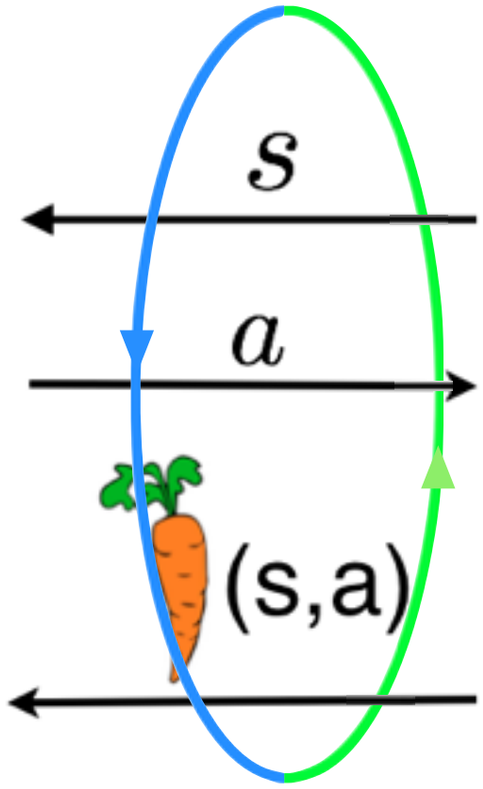
Citation from
https://www.entrepreneur.com/article/287324

Reinforcement learning:

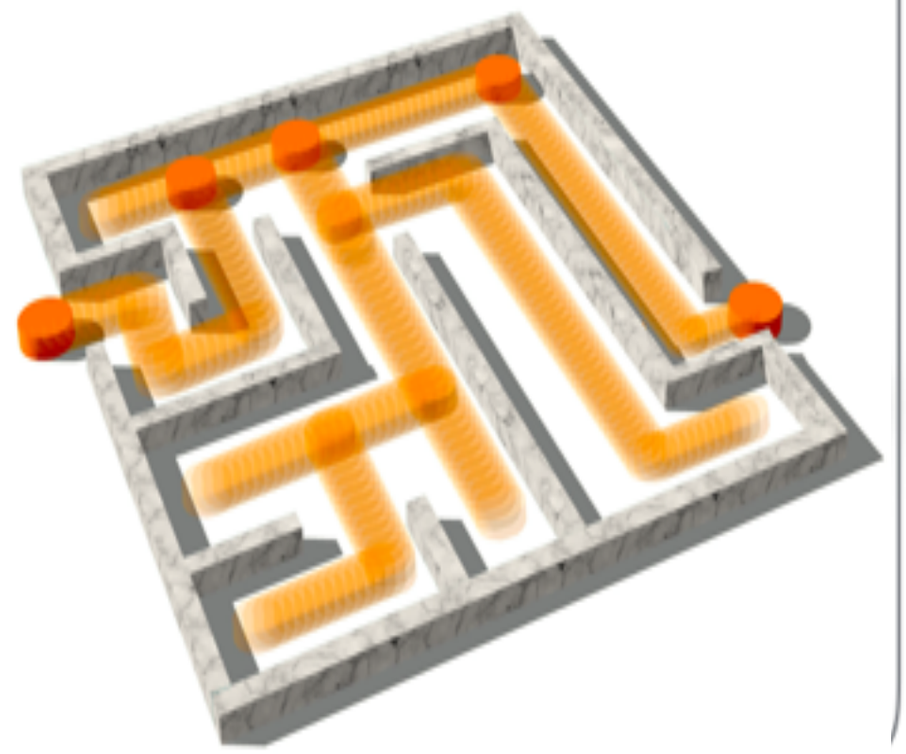
Agent - environment paradigm

Agent

Sony's AIBO



Environment

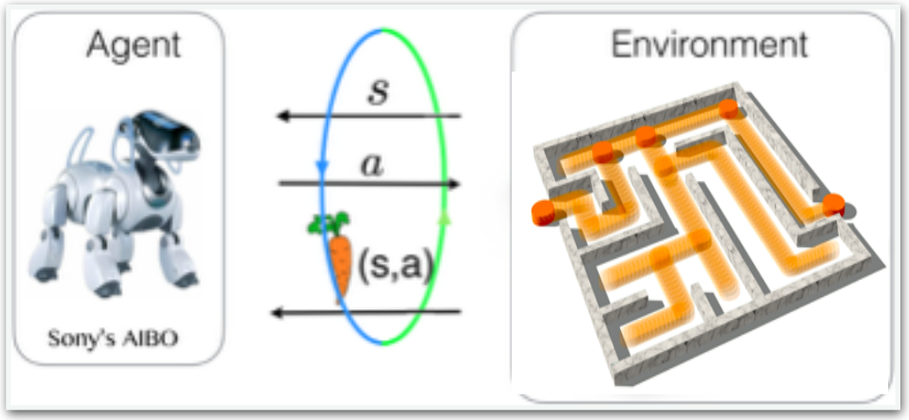


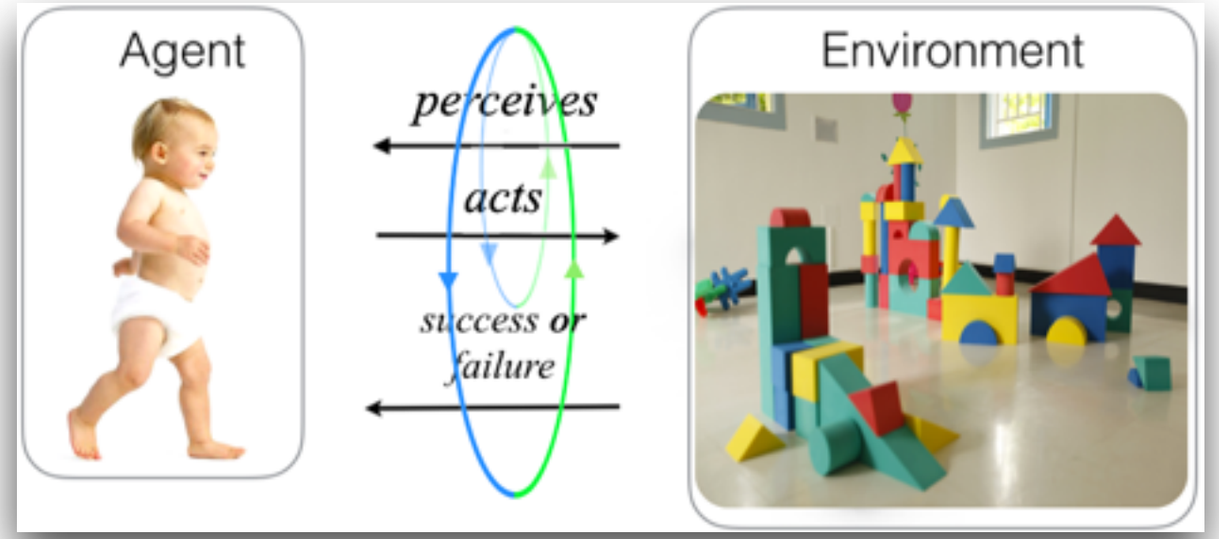
$$\mathcal{S} = \{s_1, s_2, \ldots\}$$

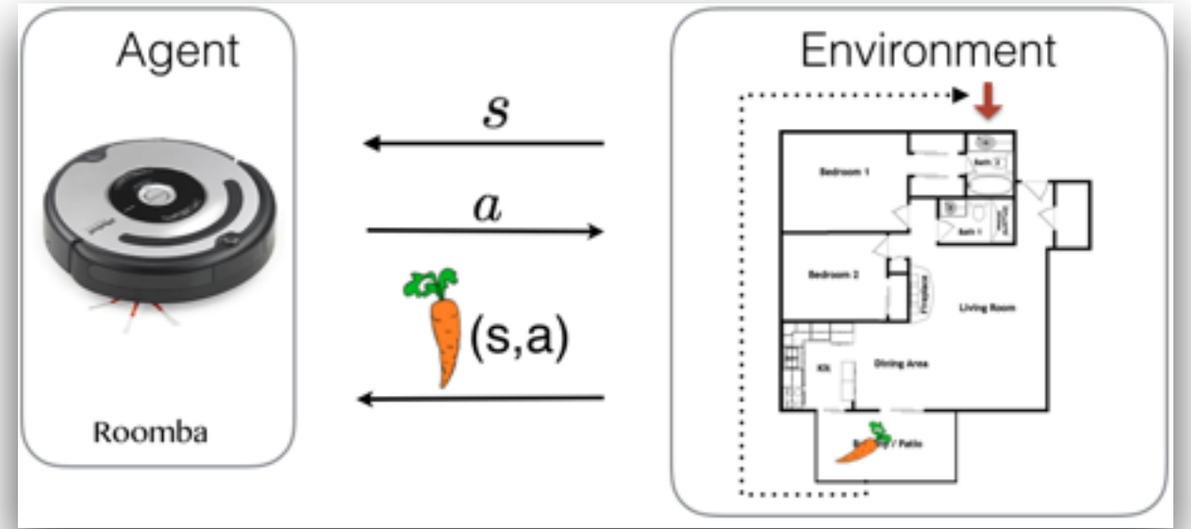
$$\mathcal{A} = \{a_1, a_2, \ldots\}$$

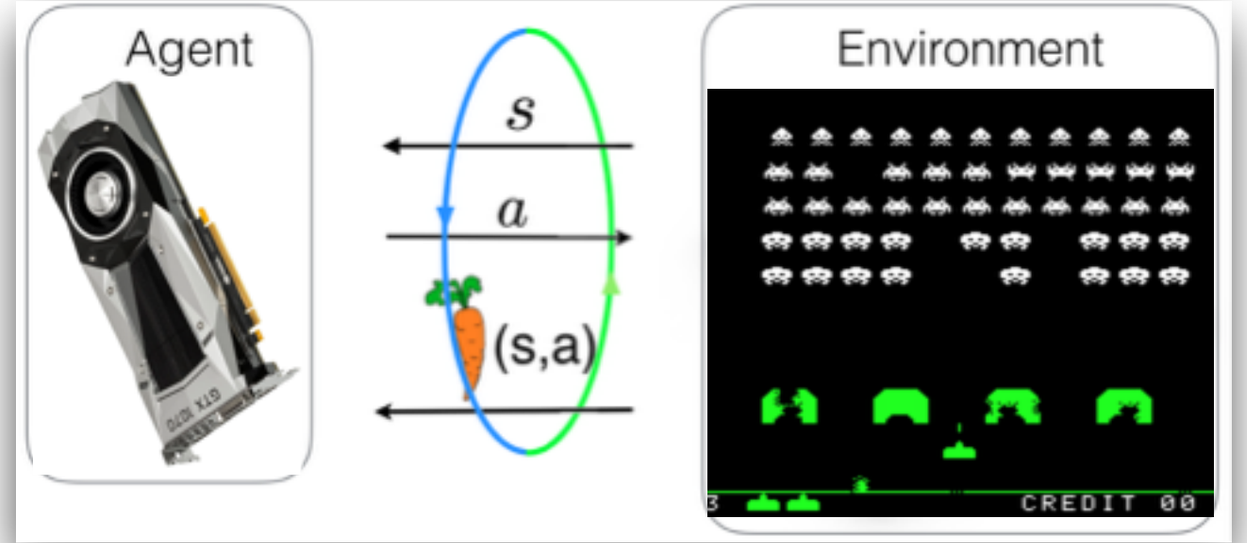
Closer to AI.
There is a body.
Interaction.
Learning.

Figures taken from Wikipedia



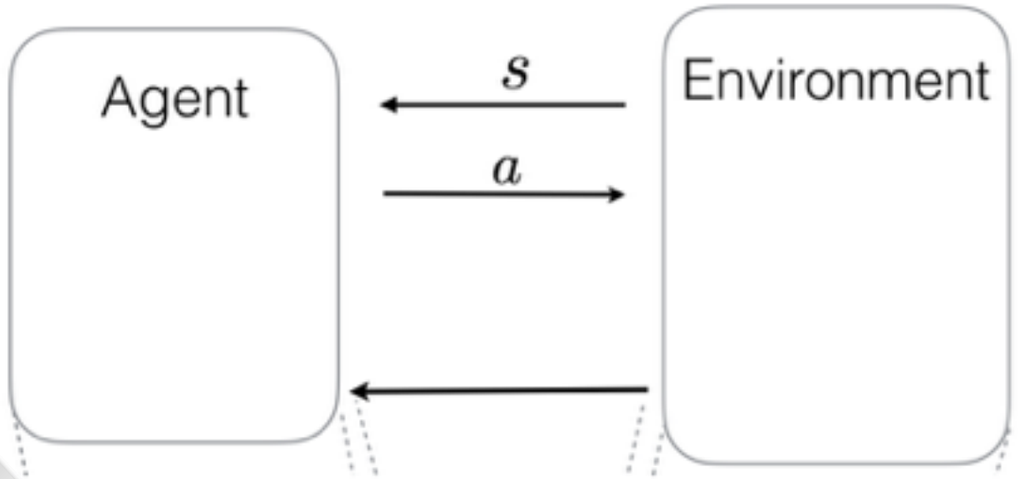


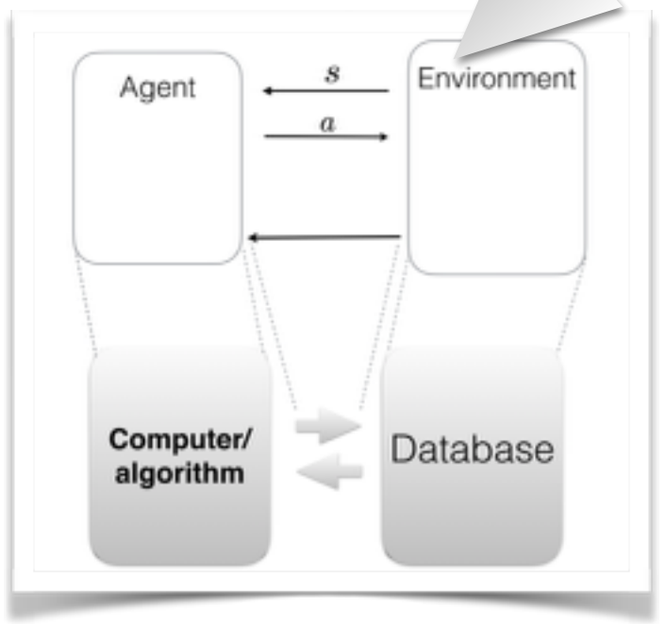


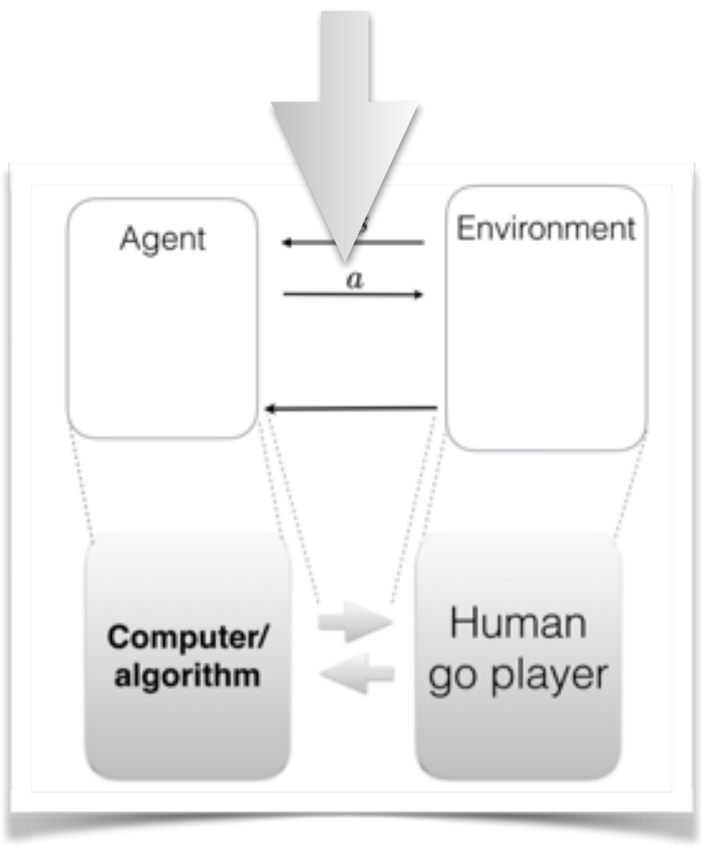


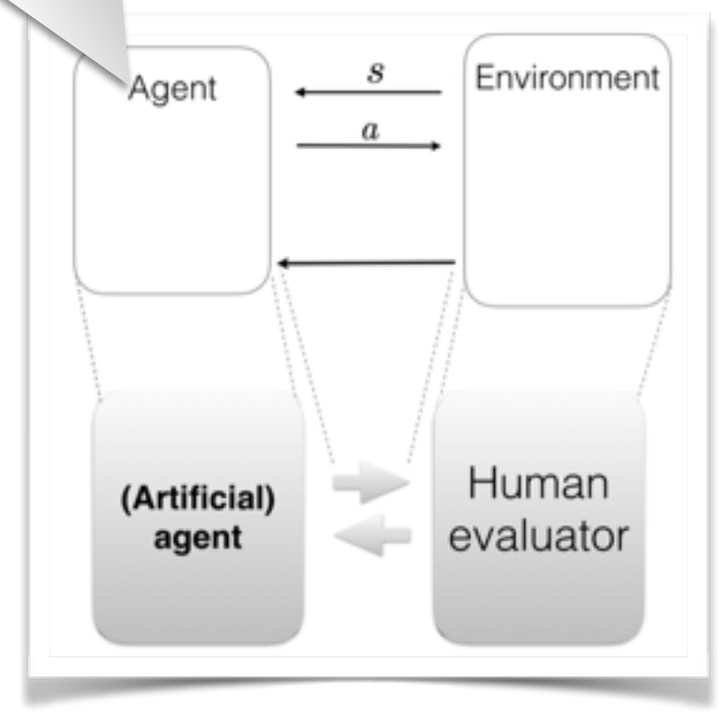
Figures taken from Wikipedia

$$\mathcal{S} = \{s_1, s_2, \ldots\}$$
$$\mathcal{A} = \{a_1, a_2, \ldots\}$$





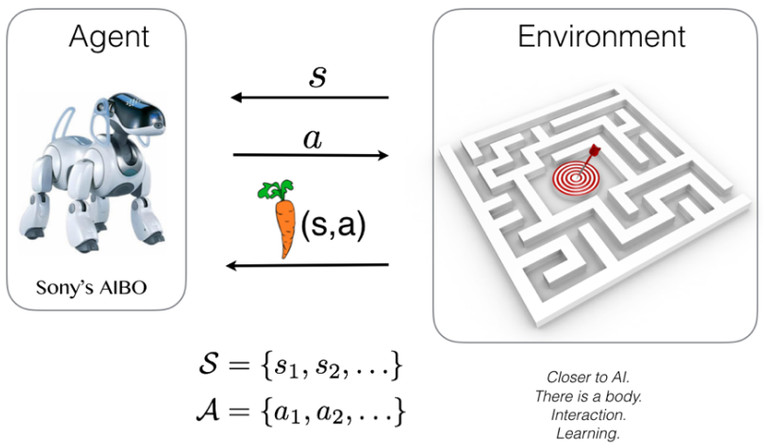




Russel, S. J. and Norvig, P. *Artificial intelligence - A modern approach* (Prentice Hall, New Jersey, 2003), sec-

Reinforcement learning in the spotlight

Reinforcement learning:
Agent - environment paradigm



Closer to AI.
There is a body.
Interaction.
Learning.

MIT technology review:
one of "10 breakthrough technologies in 2017"

The New Stack:
"AI's Next Big Step: Reinforcement learning"

E.g. AlphaGo Zero: superhuman Go performance with no human supervision: pure RL self-play


Mastering the game of Go without human knowledge

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel & Demis Hassabis

Nature **550**, 354–359 (19 October 2017)
doi:10.1038/nature24270

Received: 07 April 2017
Accepted: 13 September 2017

Figures taken from Wikipedia

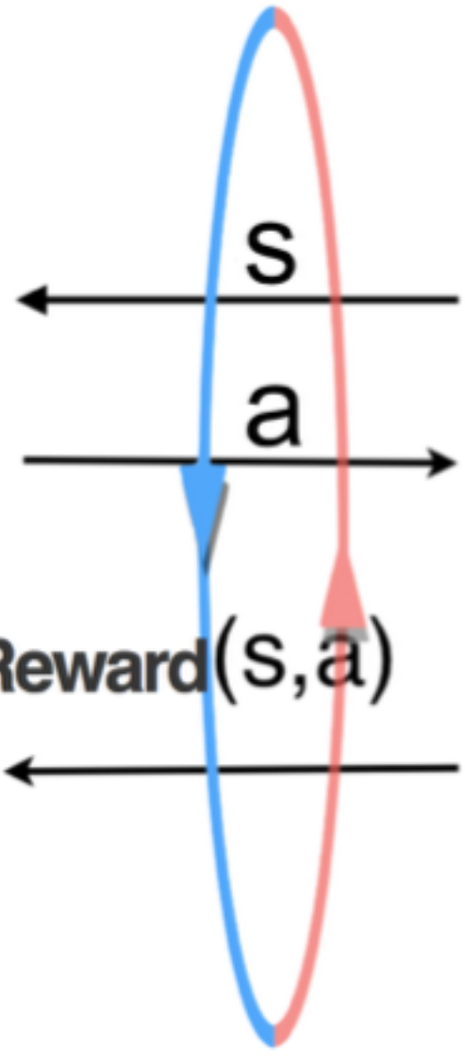
Towards AI?

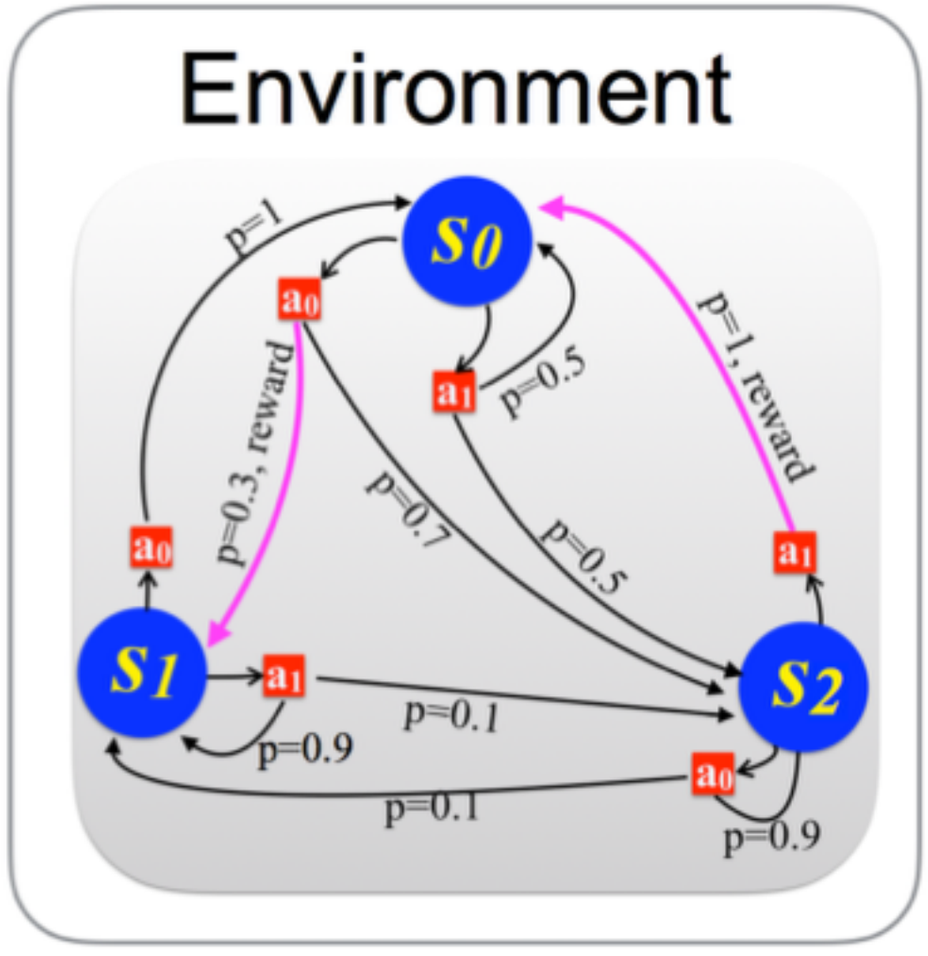

reinforcement
Knowledge representation
OCR&Vision
reasoning & inference

What do agents learn?

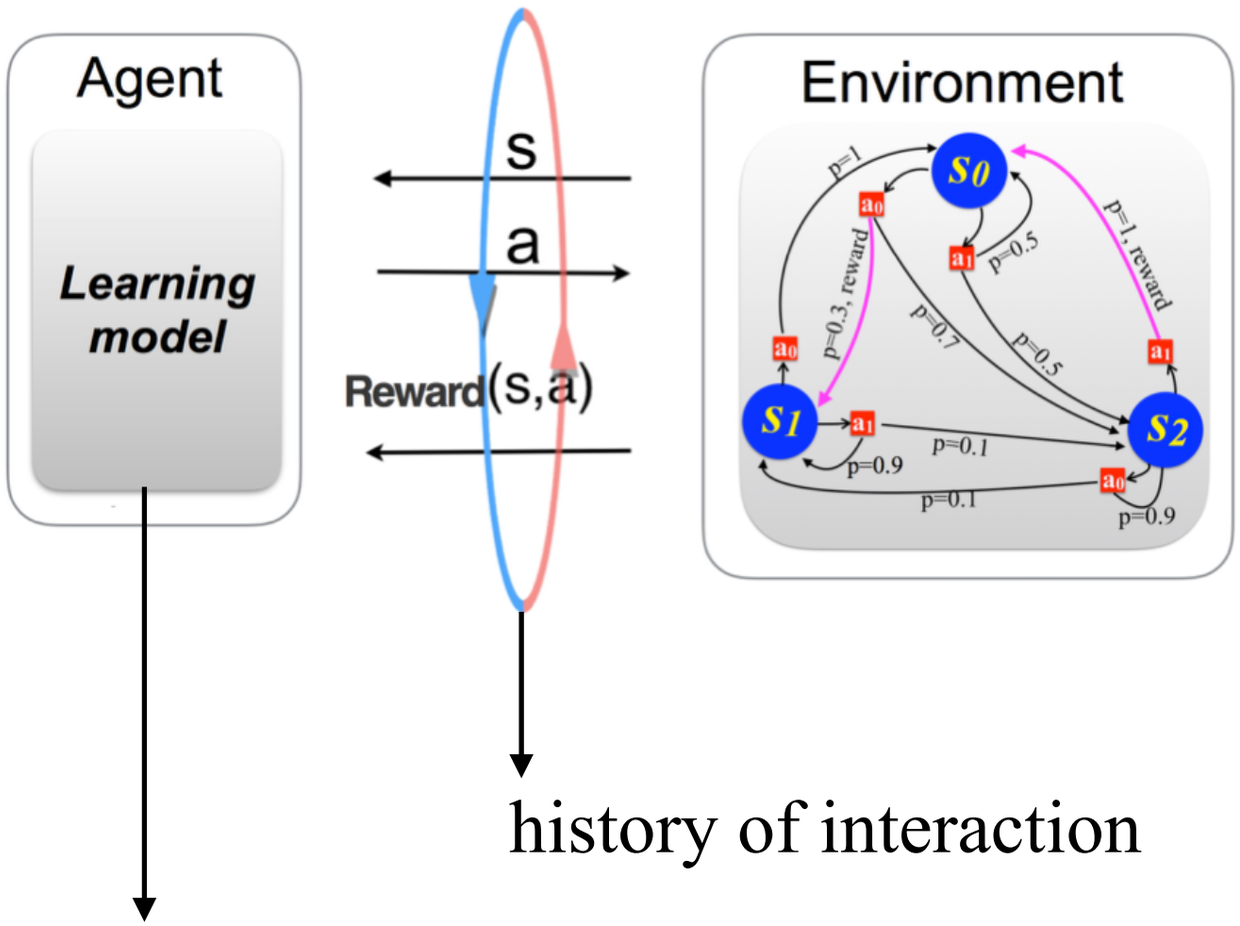
Corner for the serious

(PO) Markov Decision Process





RL glossary



history of interaction

learning model/agent
no-free lunch: metaparameters
computational complexity
"model complexity"

policy: $\pi(a|s)$

optimal: $\pi^*_\gamma(a|s)$

$$\bar{R} = \bar{r}_1 + \bar{r}_2 + \cdots + \bar{r}_k + \cdots$$

Figures of merit:

finite-horizon

$$R_N = \sum_{t \leq N} \bar{r}_t$$

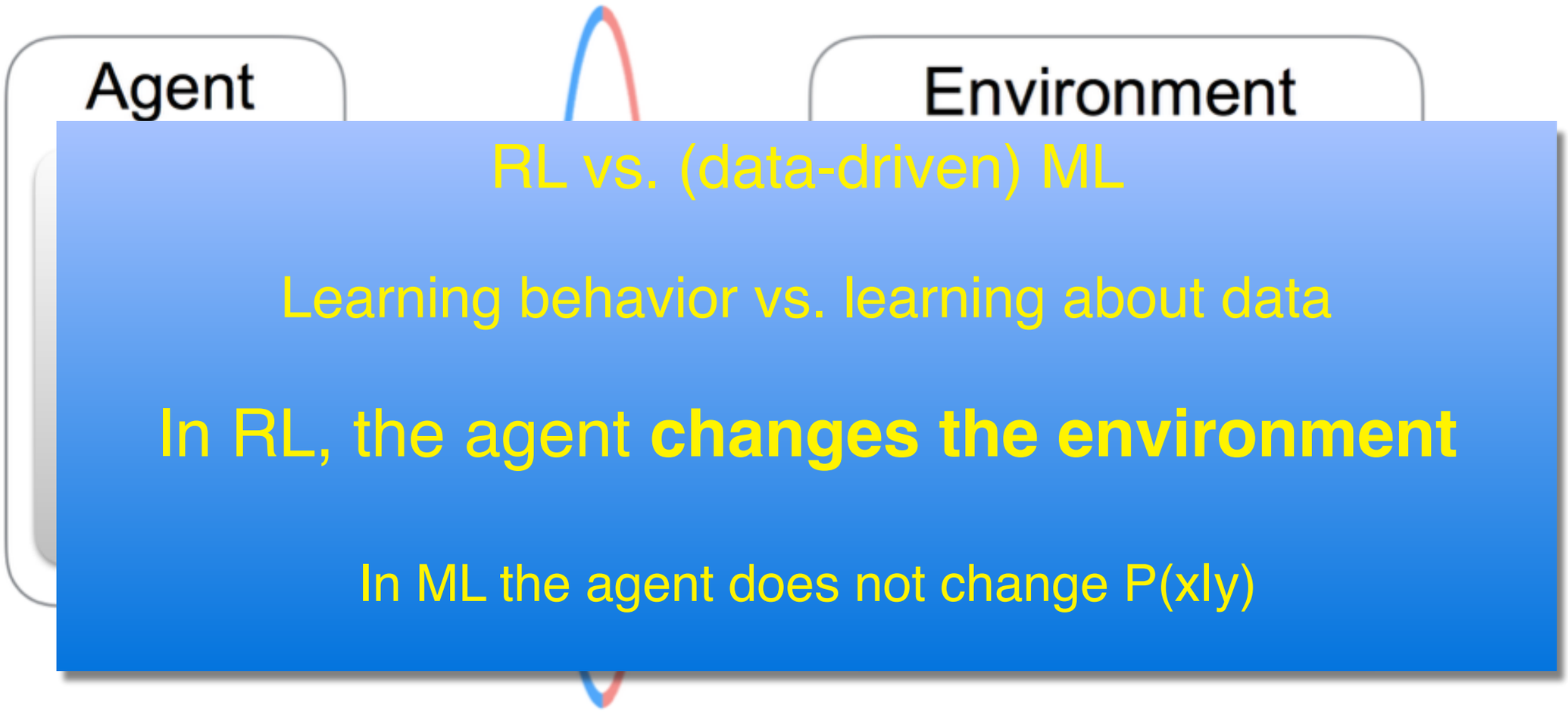
infinite-horizon

$$\bar{R} = \sum_k \gamma^k \bar{r}_k$$

What do agents learn?

Corner for the serious

(PO) Markov Decision Process



RL vs. (data-driven) ML

Learning behavior vs. learning about data

In RL, the agent **changes the environment**

In ML the agent does not change P(x|y)

policy: $\pi(a|s)$

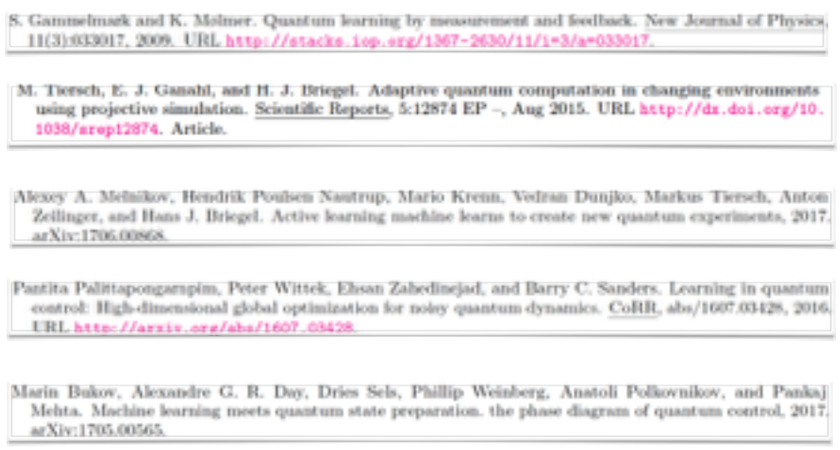
optimal: $\pi^*_\gamma(a|s)$

$$\bar{R} = \bar{r}_1 + \bar{r}_2 + \cdots + \bar{r}_k + \cdots$$

$$\bar{R} = \sum_k \gamma^k \bar{r}_k$$

What […] Quantum Reinforcement Learning

RL in QIP/QM (natural, related to feedback and control)



S. Gammelmark and K. Mølmer. Quantum learning by measurement and feedback. New Journal of Physics, 11(3):033017, 2009. URL http://stacks.iop.org/1367-2630/11/i=3/a=033017.

M. Tiersch, E. J. Ganahl, and H. J. Briegel. Adaptive quantum computation in changing environments using projective simulation. Scientific Reports, 5:12874 EP –, Aug 2015. URL http://dx.doi.org/10.1038/srep12874. Article.

Alexey A. Melnikov, Hendrik Poulsen Nautrup, Mario Krenn, Vedran Dunjko, Markus Tiersch, Anton Zeilinger, and Hans J. Briegel. Active learning machine learns to create new quantum experiments, 2017. arXiv:1706.00868.

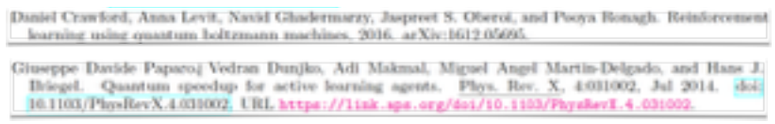
Pantita Palittapongarnpim, Peter Wittek, Ehsan Zahedinejad, and Barry C. Sanders. Learning in quantum control: High-dimensional global optimization for noisy quantum dynamics. CoRR, abs/1607.03428, 2016. URL http://arxiv.org/abs/1607.03428.

Marin Bukov, Alexandre G. R. Day, Dries Sels, Phillip Weinberg, Anatoli Polkovnikov, and Pankaj Mehta. Machine learning meets quantum state preparation. the phase diagram of quantum control, 2017. arXiv:1705.00565.

Learning *for* and *in* quantum experiments
ML building a QC?

Quantum-enhanced ML

- *Q. enhancements for specific algorithms*



Daniel Crawford, Anna Levit, Navid Ghadermarzy, Jaspreet S. Oberoi, and Pooya Ronagh. Reinforcement learning using quantum boltzmann machines, 2016. arXiv:1612.05695.

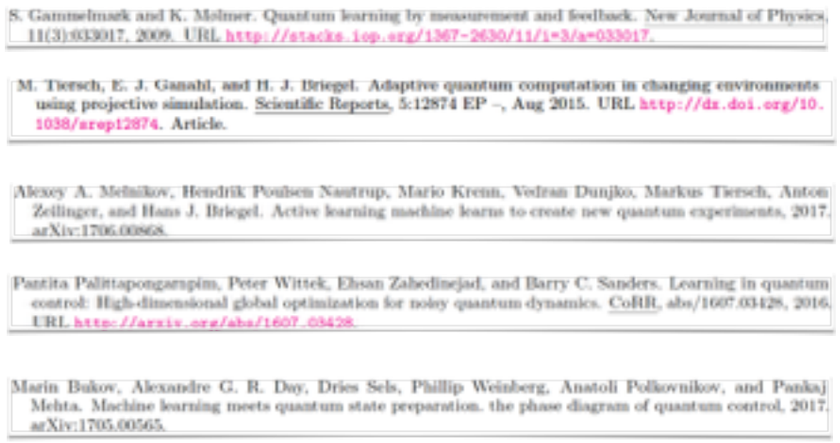
Giuseppe Davide Paparo, Vedran Dunjko, Adi Makmal, Miguel Angel Martin-Delgado, and Hans J. Briegel. Quantum speedup for active learning agents. Phys. Rev. X, 4:031002, Jul 2014. doi: 10.1103/PhysRevX.4.031002. URL https://link.aps.org/doi/10.1103/PhysRevX.4.031002.

- *Q. enhancements via q. interaction (quantum computational learning theory)*
- *Q. generalizations*

QRL in Q. systems enabling QRL?

What […] Quantum Reinforcement Learning

RL in QIP/QM (natural, related to feedback and control)



S. Gammelmark and K. Mølmer. Quantum learning by measurement and feedback. New Journal of Physics, 11(3):033017, 2009. URL http://stacks.iop.org/1367-2630/11/i=3/a=033017.

M. Tiersch, E. J. Ganahl, and H. J. Briegel. Adaptive quantum computation in changing environments using projective simulation. Scientific Reports, 5:12874 EP –, Aug 2015. URL http://dx.doi.org/10.1038/srep12874. Article.

Alexey A. Melnikov, Hendrik Poulsen Nautrup, Mario Krenn, Vedran Dunjko, Markus Tiersch, Anton Zeilinger, and Hans J. Briegel. Active learning machine learns to create new quantum experiments, 2017. arXiv:1706.00868.

Pantita Palittapongarnpim, Peter Wittek, Ehsan Zahedinejad, and Barry C. Sanders. Learning in quantum control: High-dimensional global optimization for noisy quantum dynamics. CoRR, abs/1607.03428, 2016. URL http://arxiv.org/abs/1607.03428.

Marin Bukov, Alexandre G. R. Day, Dries Sels, Phillip Weinberg, Anatoli Polkovnikov, and Pankaj Mehta. Machine learning meets quantum state preparation. the phase diagram of quantum control, 2017. arXiv:1705.00565.

Learning *for* and *in* quantum experiments
ML building a QC?

Quantum-enhanced ML

- *Q. enhancements for specific algorithms*

Daniel Crawford, Anna Levit, Navid Ghadermarzy, Jaspreet S. Oberoi, and Pooya Ronagh. Reinforcement learning using quantum boltzmann machines, 2016. arXiv:1612.05695.

Giuseppe Davide Paparo, Vedran Dunjko, Adi Makmal, Miguel Angel Martin-Delgado, and Hans J. Briegel. Quantum speedup for active learning agents. Phys. Rev. X, 4:031002, Jul 2014. doi: 10.1103/PhysRevX.4.031002. URL https://link.aps.org/doi/10.1103/PhysRevX.4.031002.

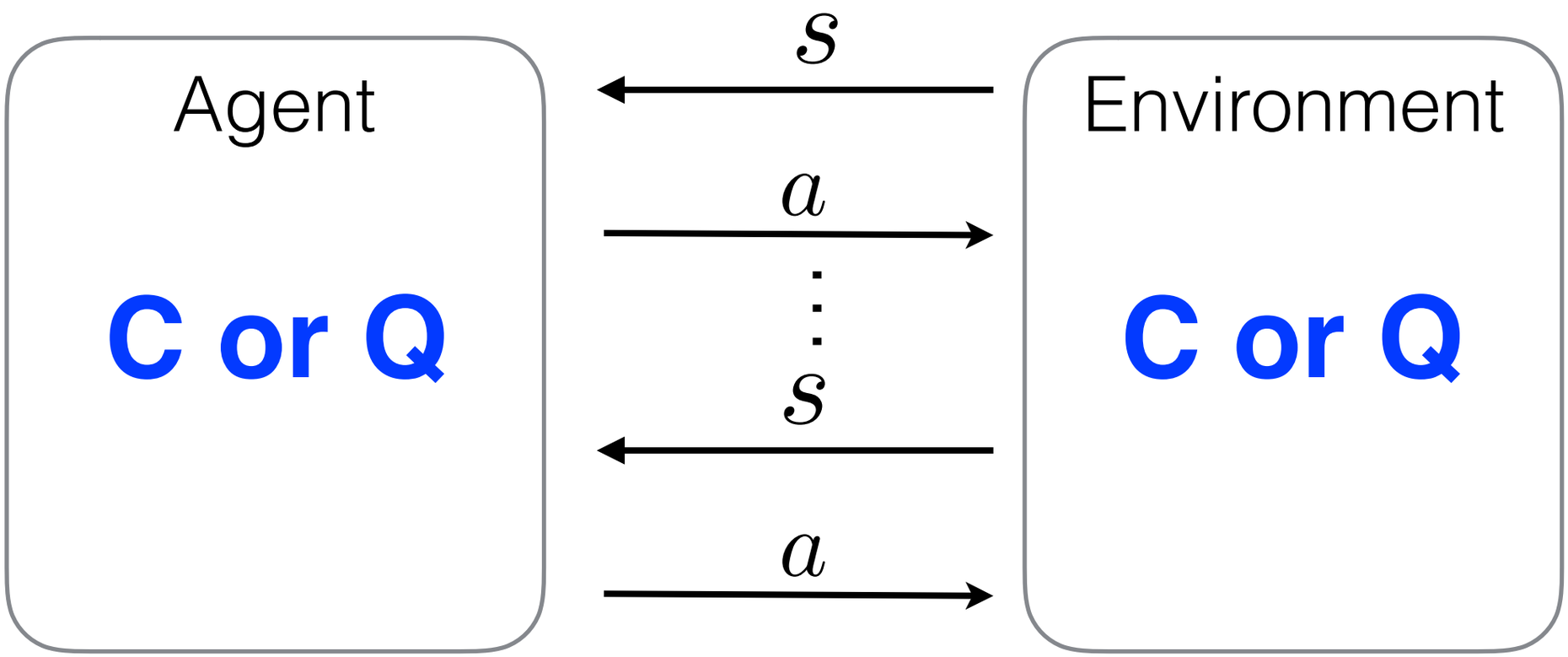
- *Q. enhancements via q. interaction (quantum computational learning theory)*
- *Q. generalizations*

QRL in Q. systems enabling QRL?

Part 2: Quantum information and reinforcement learning

- *Quantum Agent - Environment paradigm?*

Want something like:

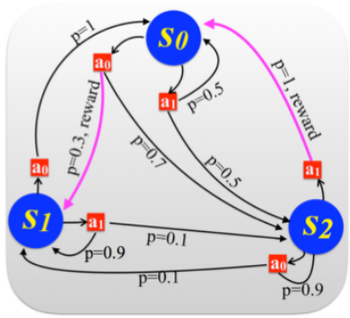


V. Dunjko, J. M. Taylor, H. J. Briegel
Quantum-enhanced learning
In preparation (2016)

V. Dunjko, J. M. Taylor, H. J. Briegel
Framework for learning agents in quantum environments
arXiv:1507.08482 (2015)

Flavors or RL…what are agents, environments?

CS: MDPs, POMDPs, Turing machines



Robotics: Real environments

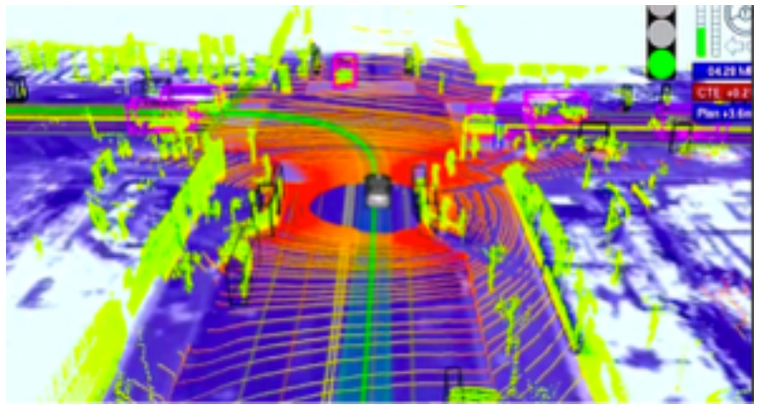
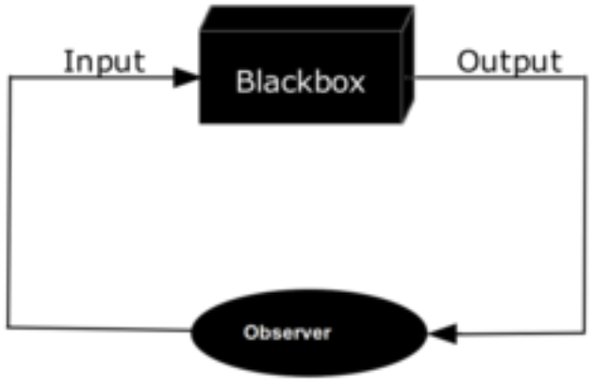
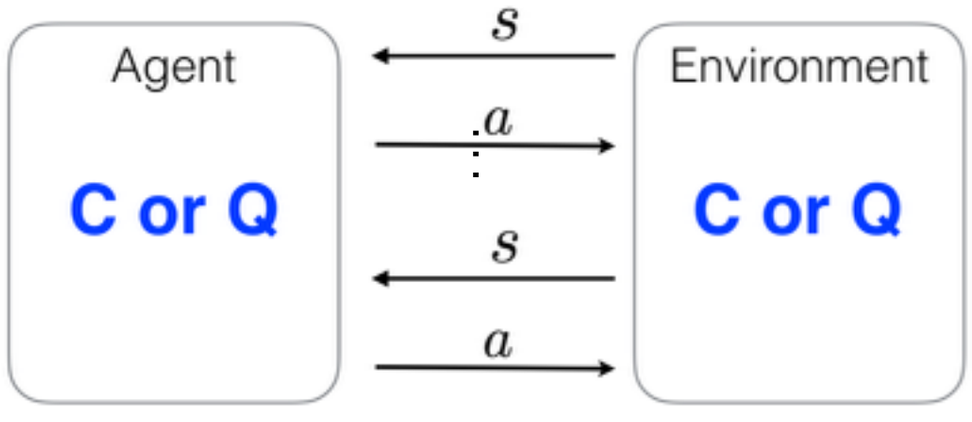


Image: Sebastian Thrun & Chris Urmson/Google

QIP/QML: ?…systems



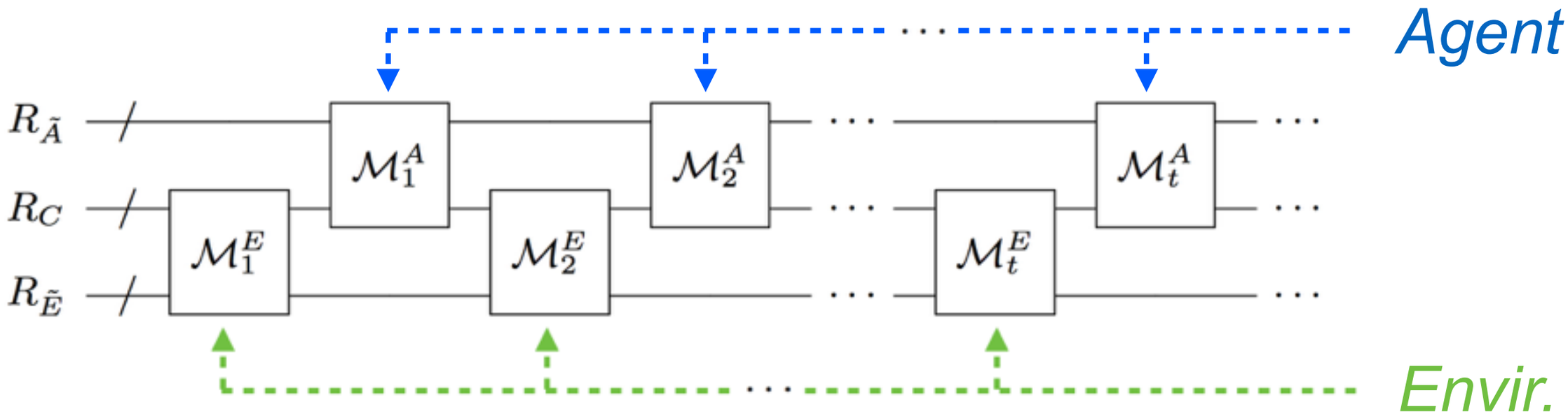
- *Quantum Agent - Environment paradigm*



$$\mathcal{A} = \{a_1, a_2, \ldots\} \qquad \mathcal{S} = \{s_1, s_2, \ldots\}$$

$$\mathcal{H}_A = \mathrm{span}\{|a_i\rangle\}, \; \mathcal{H}_S = \mathrm{span}\{|s_i\rangle\}$$

is equivalent to



- **Agents** (environments) are **sequences of CPTP maps**, acting on a private and a common register - the memory and the interface, respectively.
- Memory channels = combs = quantum strategies

But…. why go there?

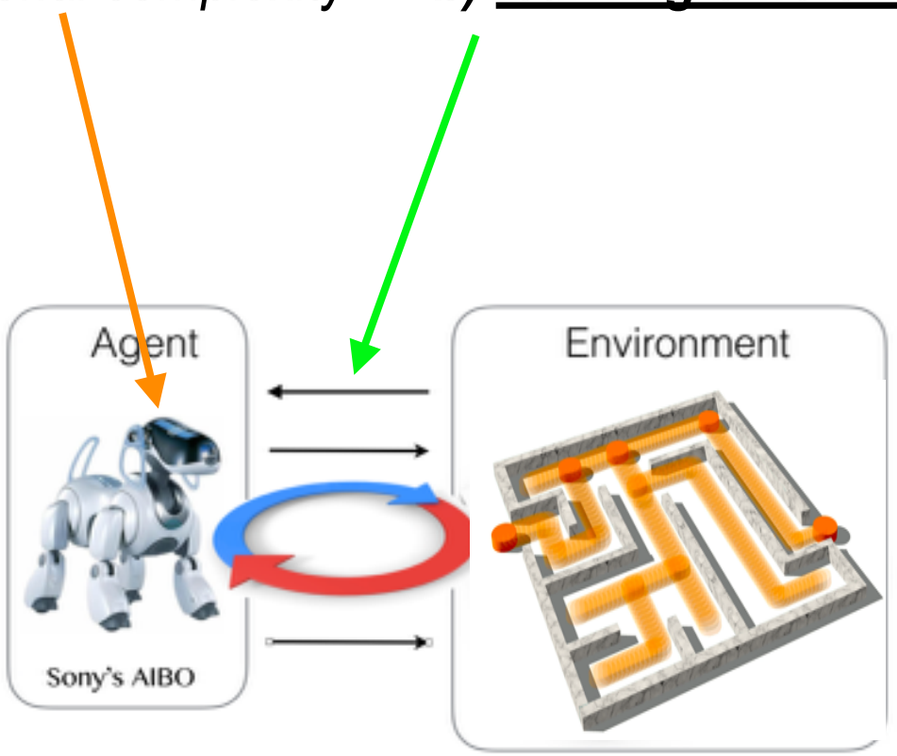
- **Fundamental meaning and role of *learning* in the quantum world**
- **Speed-ups! “faster”, “better” learning**
  - *What can we make better?*

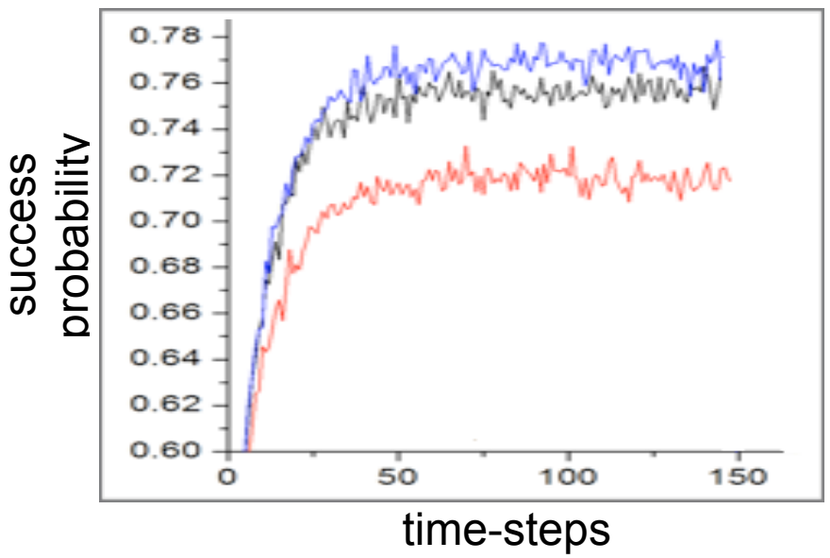
    a) *computational complexity* ***b)*** ***<u>learning efficiency</u>***

But…. why go there?

- **Fundamental meaning and role of *learning* in the quantum world**
- **Speed-ups! “faster”, “better” learning**
  - *What can we make better?*

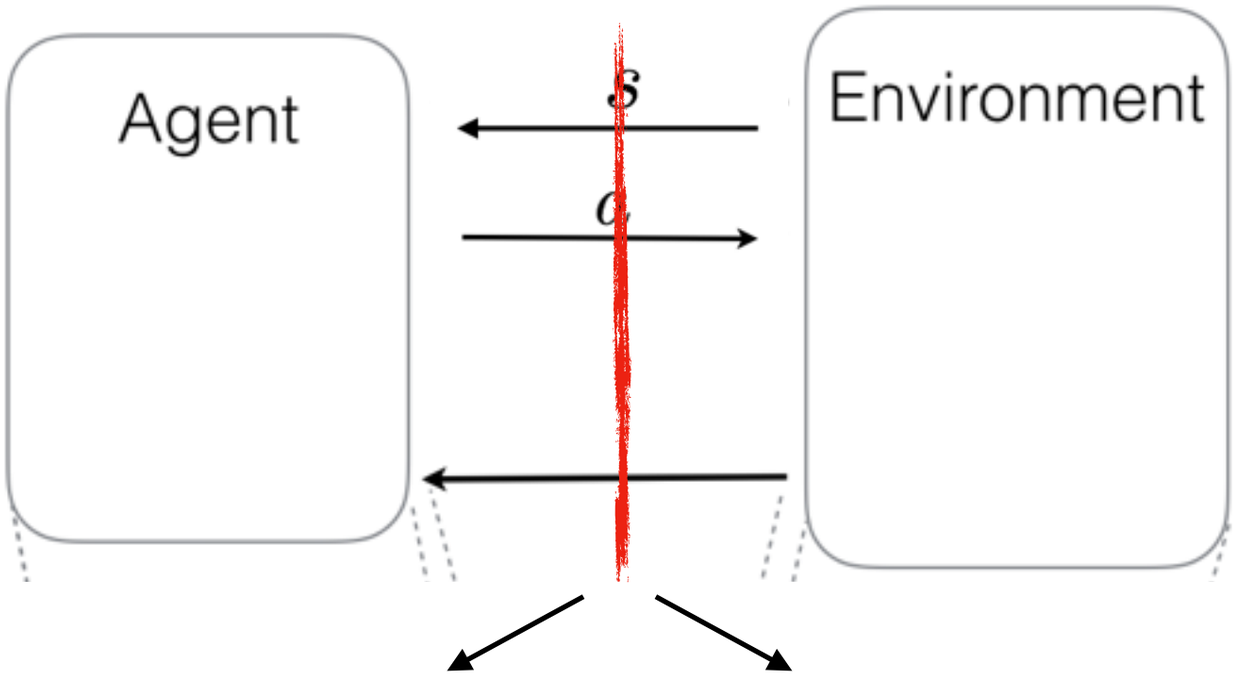
a) *computational complexity* ***b)*** ***<u>learning efficiency</u>*** *(“genuine learning-related figures of merit”)*





related to query complexity

- *Quantum Agent - Environment paradigm*



- All improvements are consequences of computational improvements

★ intuition: “cannot run Grover’s search on a classical phonebook”

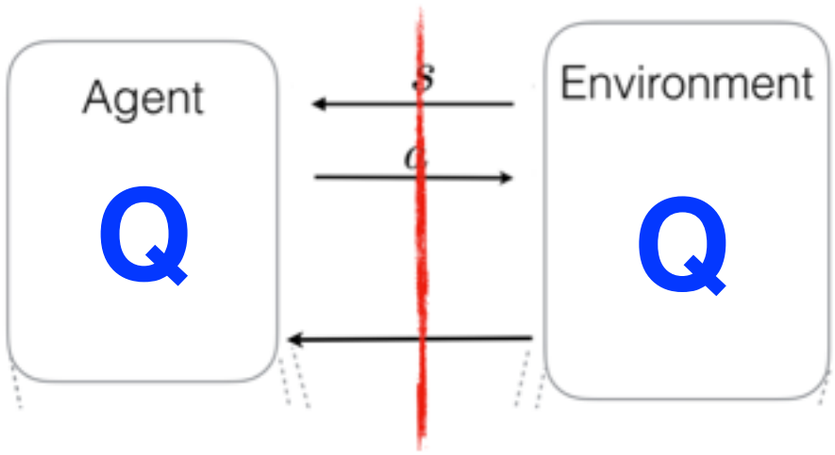
- More radical computational improvements
- More interesting quantum effects may me exploitable

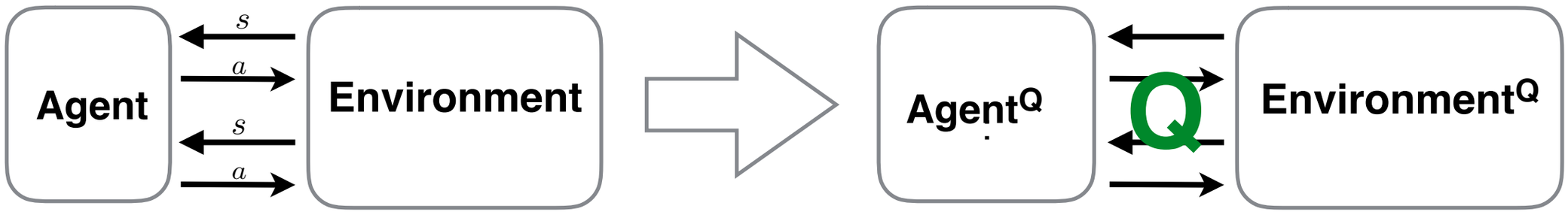
★Oracular quantum computation: exponential separations in principle possible

Paparo, G. D., Dunjko, V., Makmal, A., Martin-Delgado, M. A., Briegel, H. J. Quantum speedup for active learning agents. *Phys. Rev. X* **4**, 031002 (2014). URL http://journals.aps.org/prx/abstract/10.1103/PhysRevX.4.031002.

V. Dunjko, J. M. Taylor, H. J. Briegel
Quantum-enhanced machine learning
Phys. Rev. Lett. **117**, 130501 (2016)

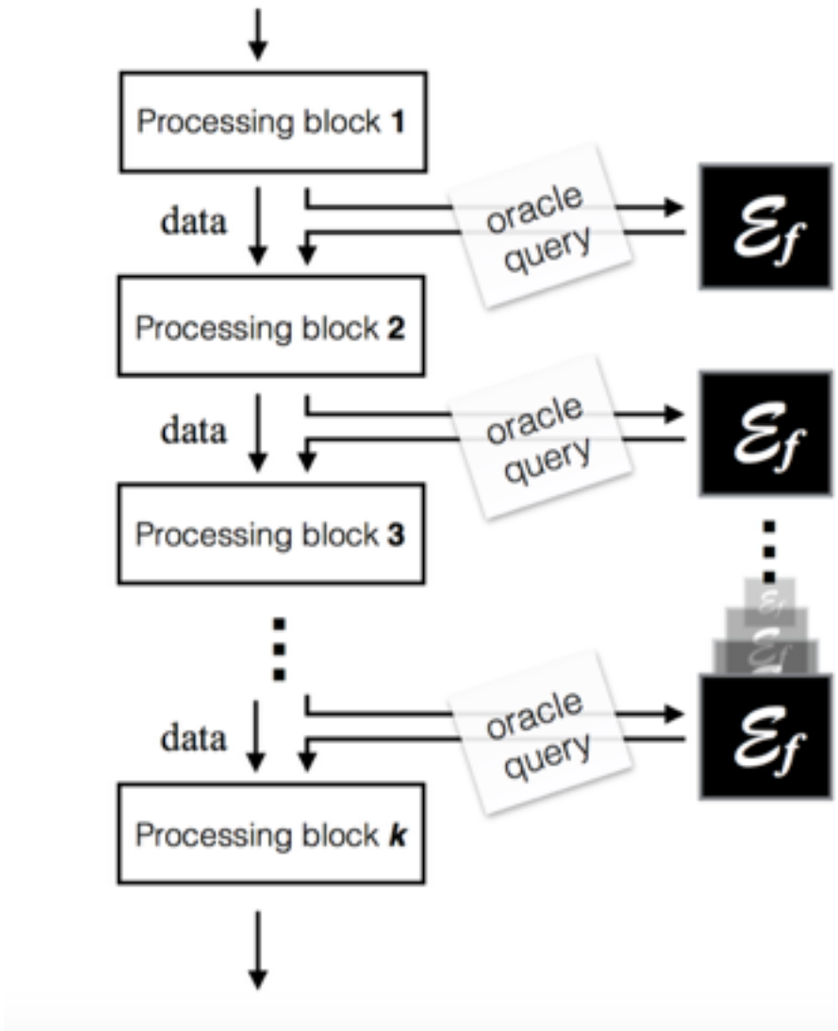
V. Dunjko, J. M. Taylor, H. J. Briegel
Framework for learning agents in quantum environments
arXiv:1507.08482 (2015)



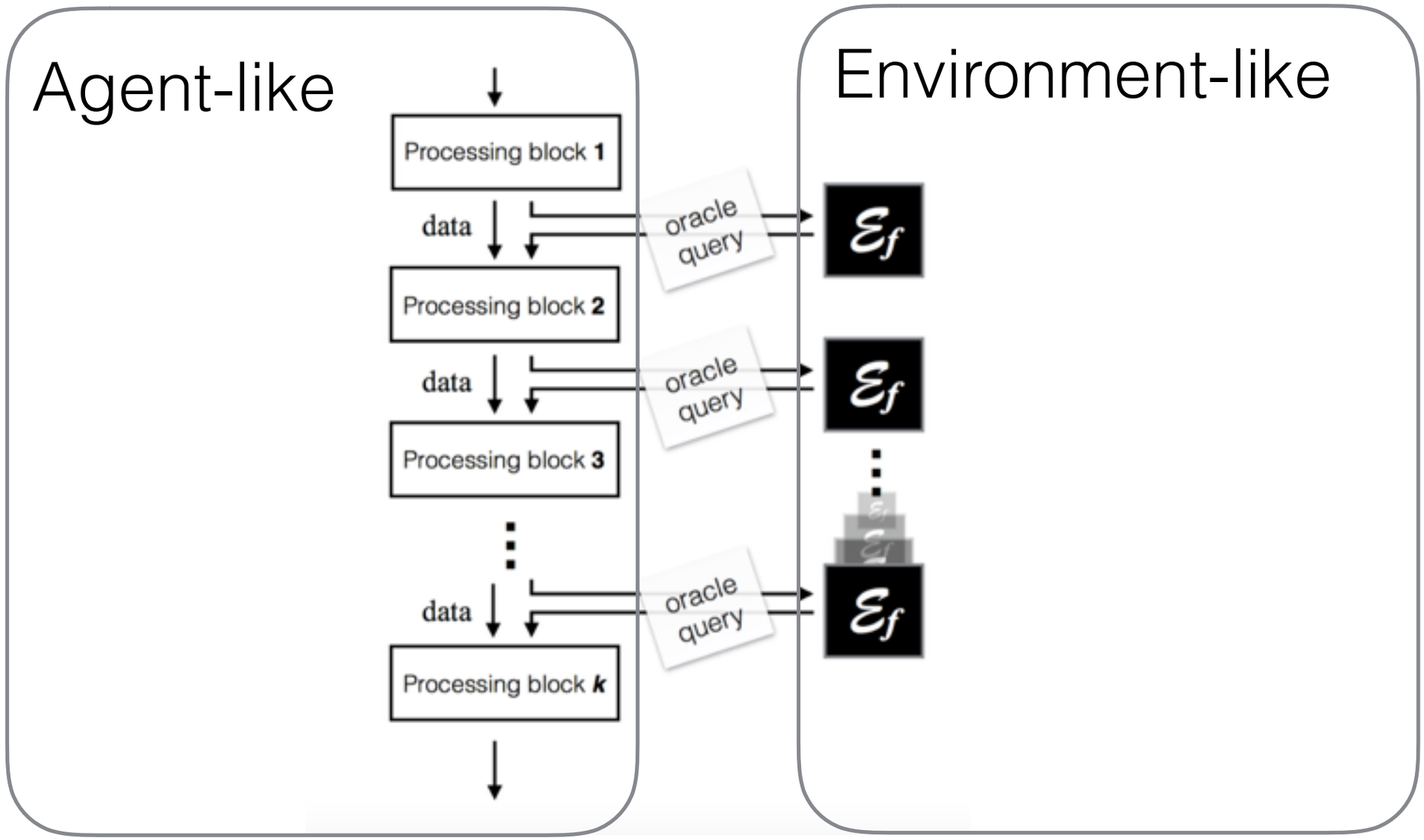


V. Dunjko, J. M. Taylor, H. J. Briegel
Quantum-enhanced machine learning
Phys. Rev. Lett. **117**, 130501 (2016)

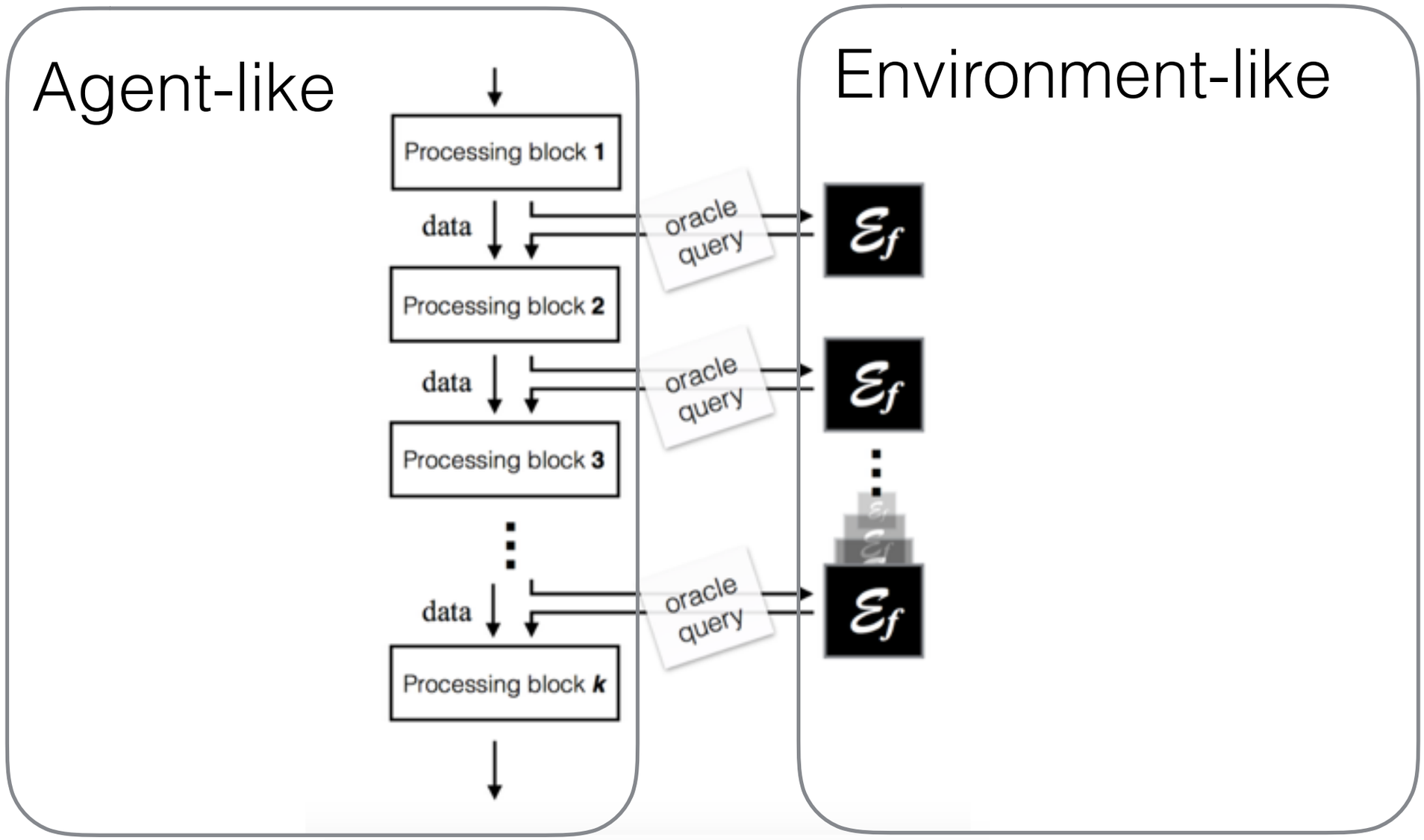
Inspiration from oracular quantum computation…



Inspiration from oracular quantum computation…

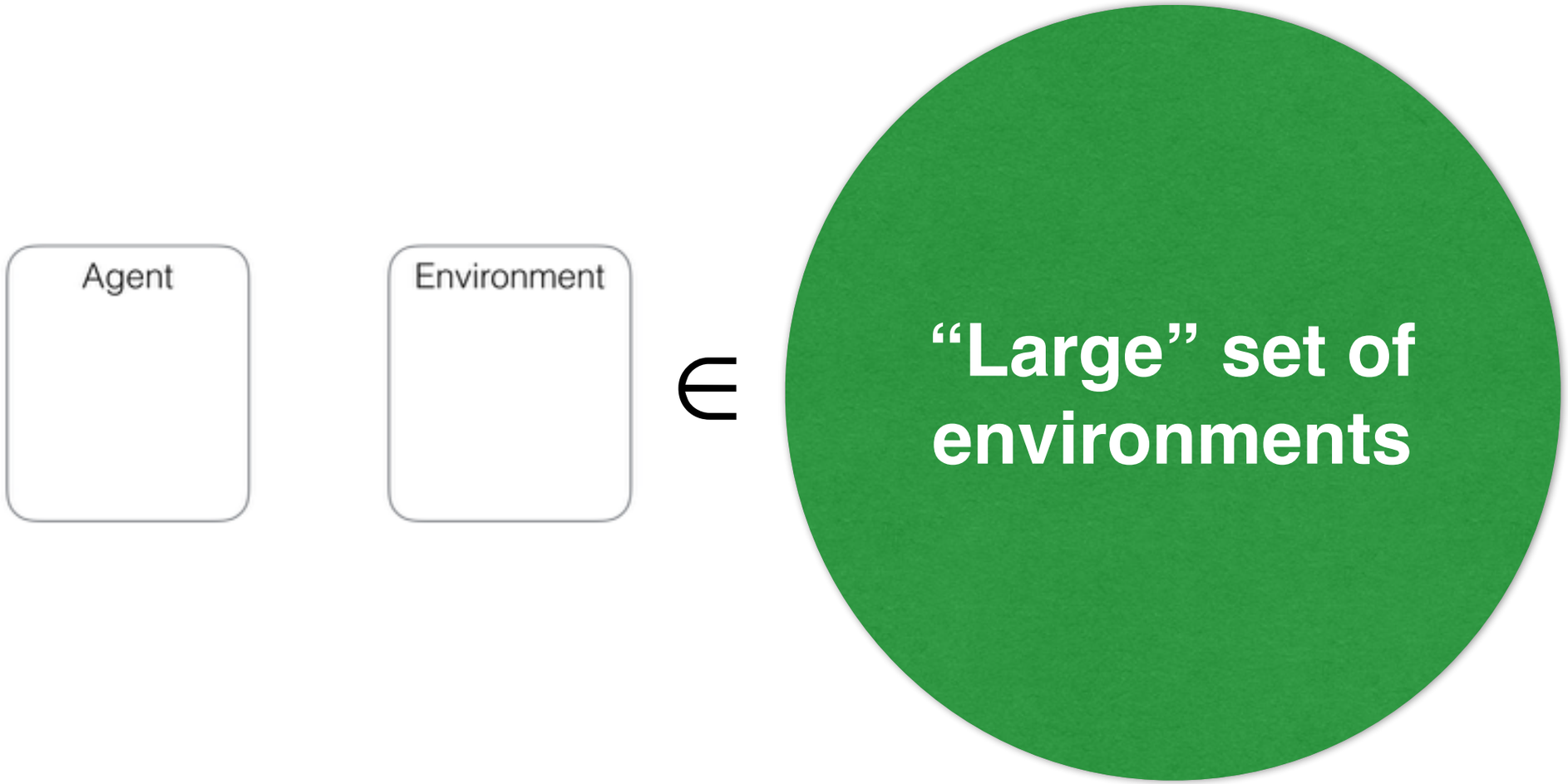


Inspiration from oracular quantum computation…

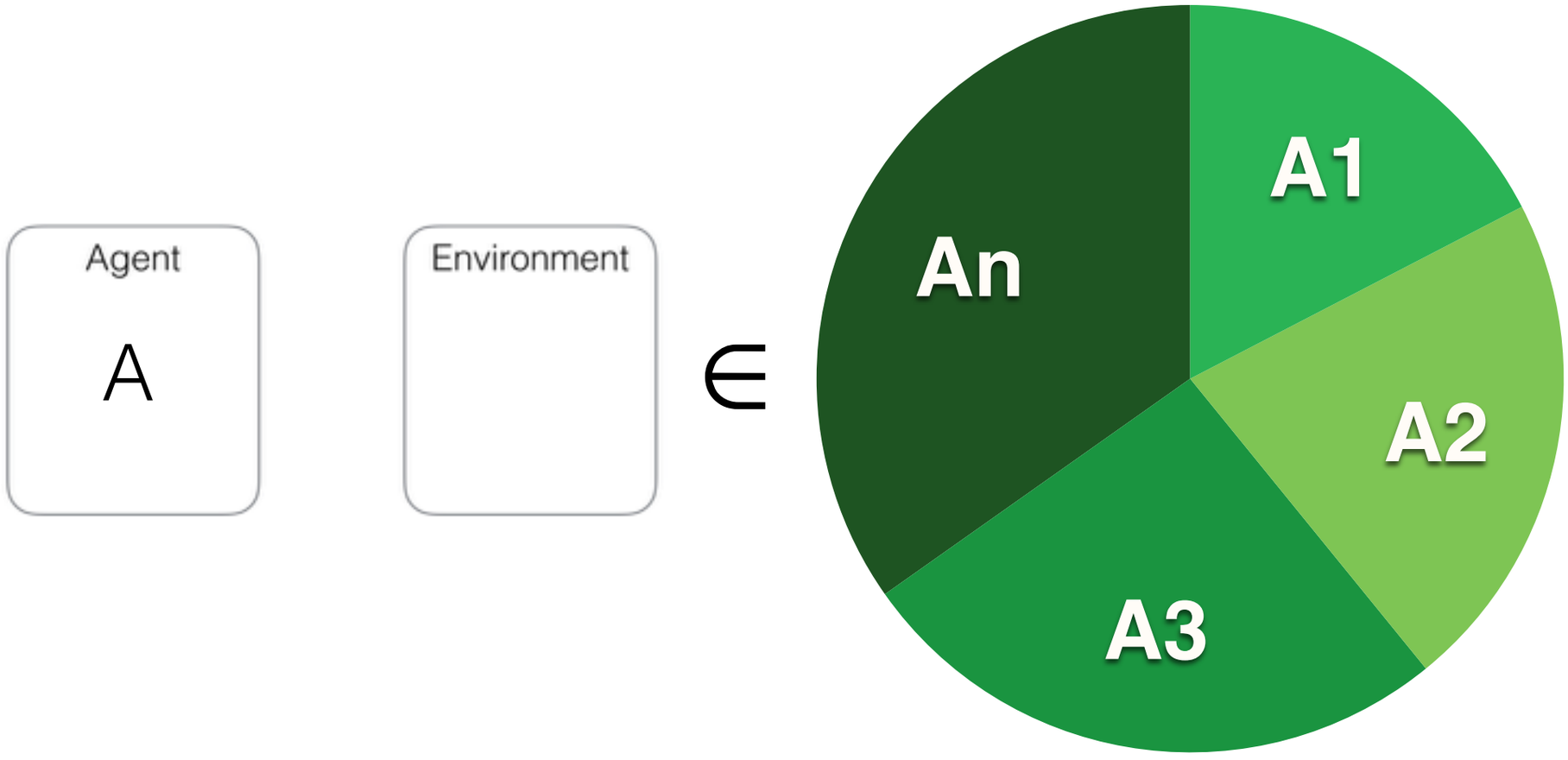


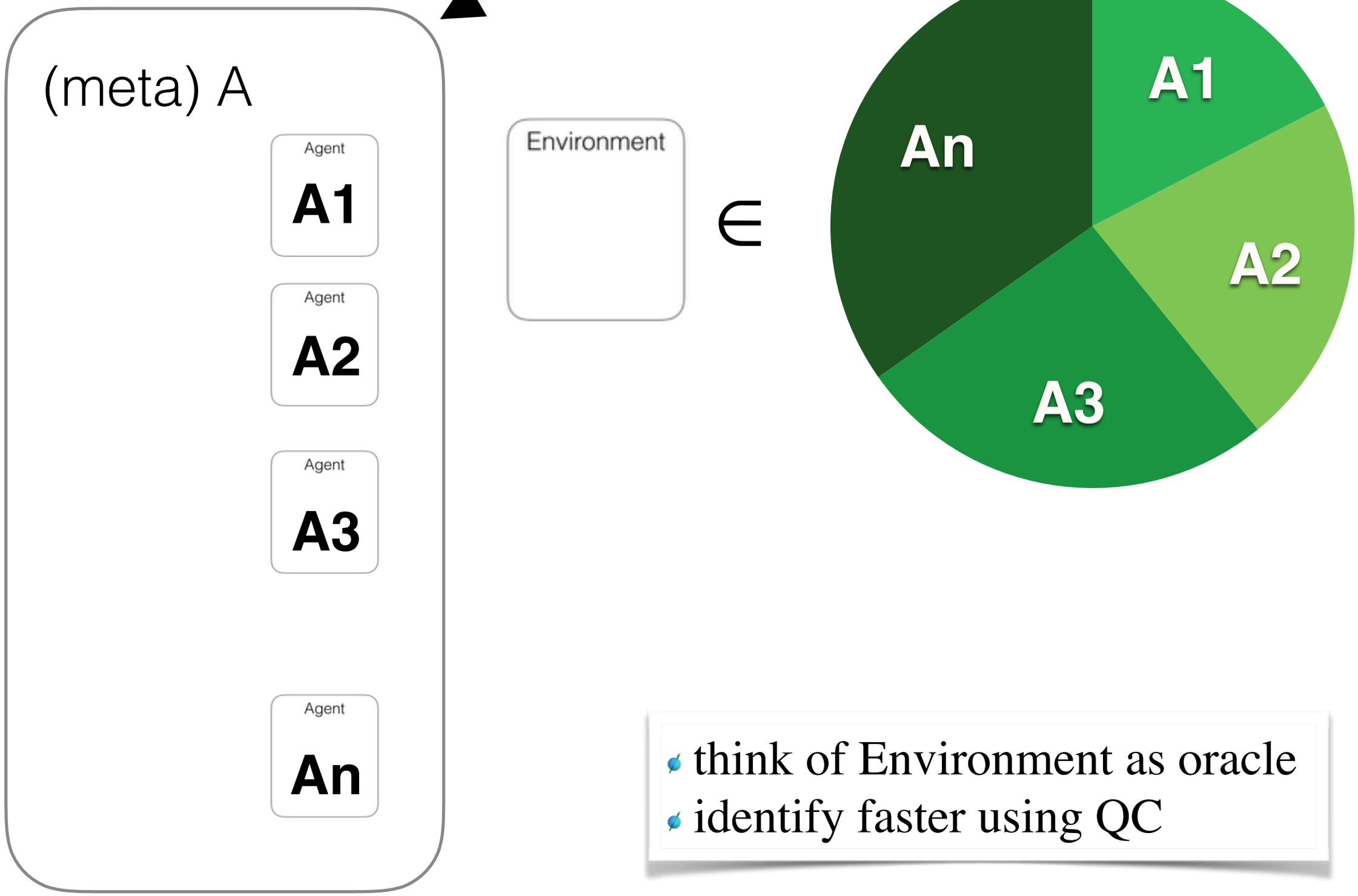
- think of Environment as oracle

How could one use that…
E.g. oracle identification!



If E is chosen at random, on average all agents perform equally *(NFL)*


Agent
A
Environment
∈
An
A1
A2
A3

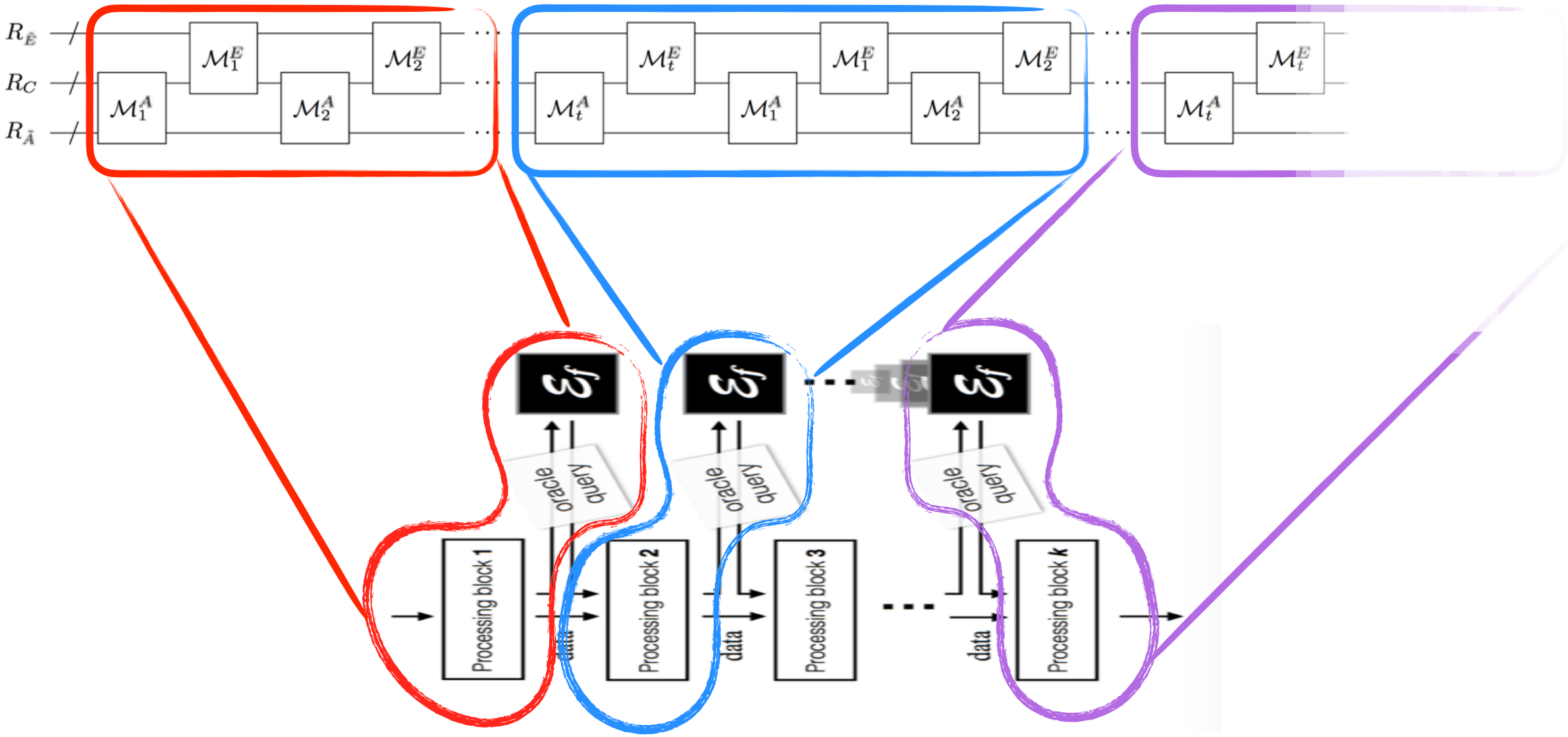


- think of Environment as oracle
- identify faster using QC

But… environments are not like standard oracles…

"Oraculization"
(taming the open environment)

(blocking, accessing purification and recycling)



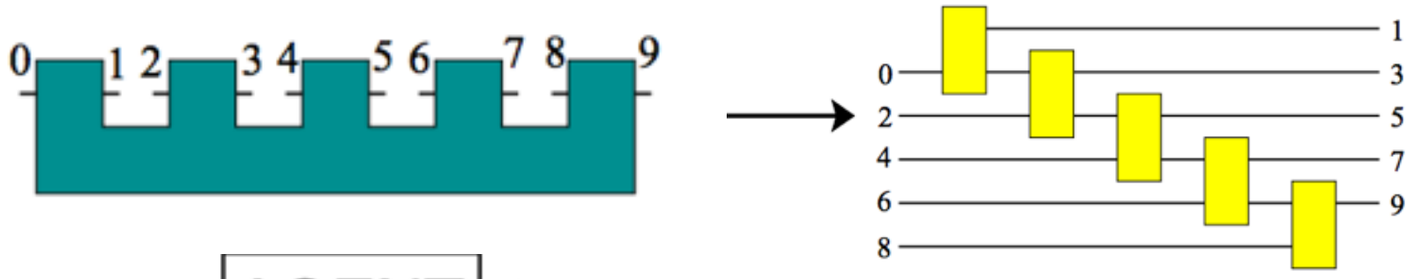
Oraculization (blocking)

(taming the open environment)

1)

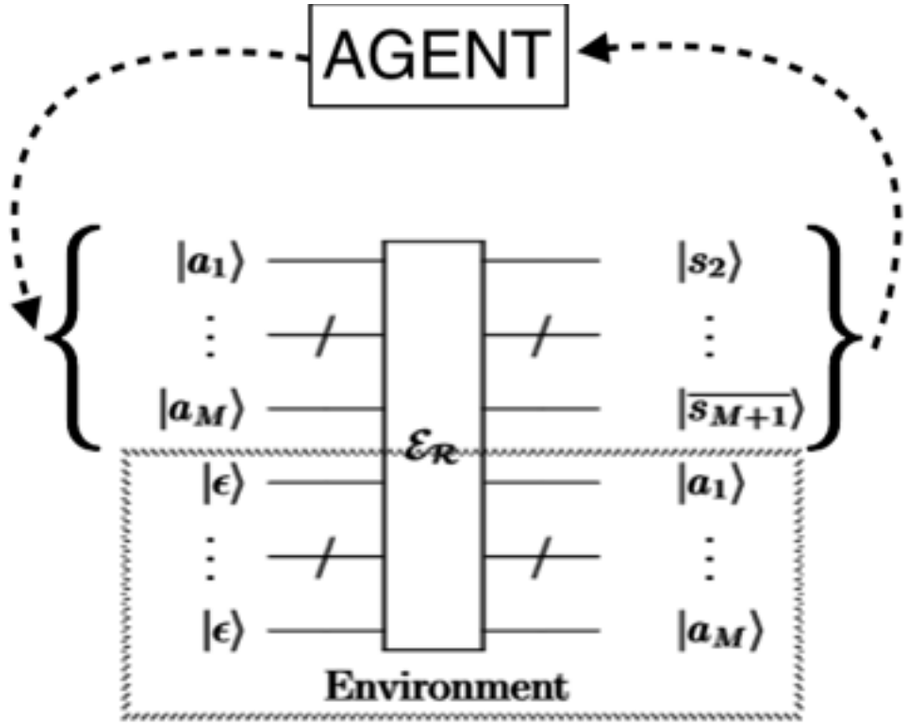
quantum comb

2)



causal network

3)



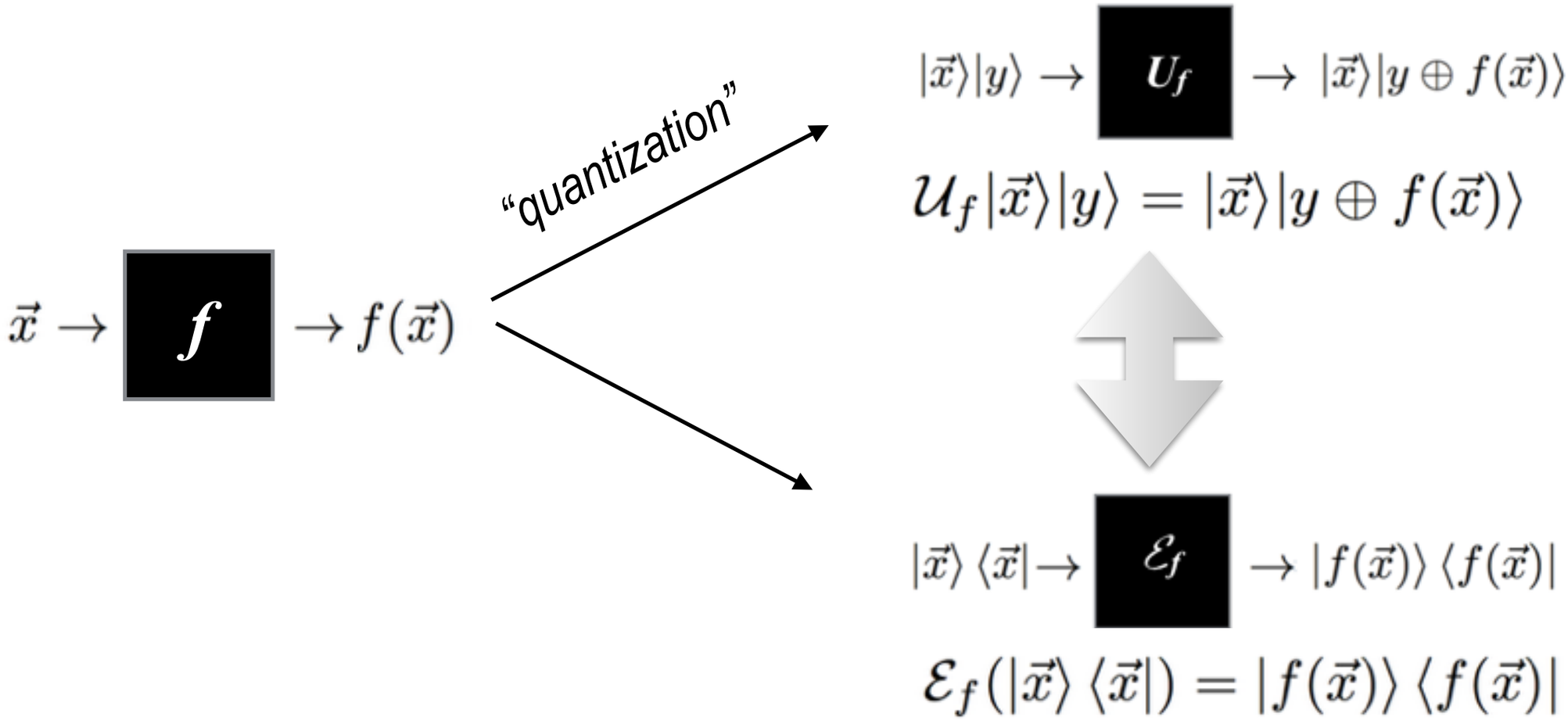
“blocking”

Chiribella, G., D’Ariano, G. M. & Perinotti, P. Quantum Circuit Architecture. *Phys. Rev. Lett.* **101**, 060401 (2008).

Oraculization (recovery and recycling)

(taming the open environment)

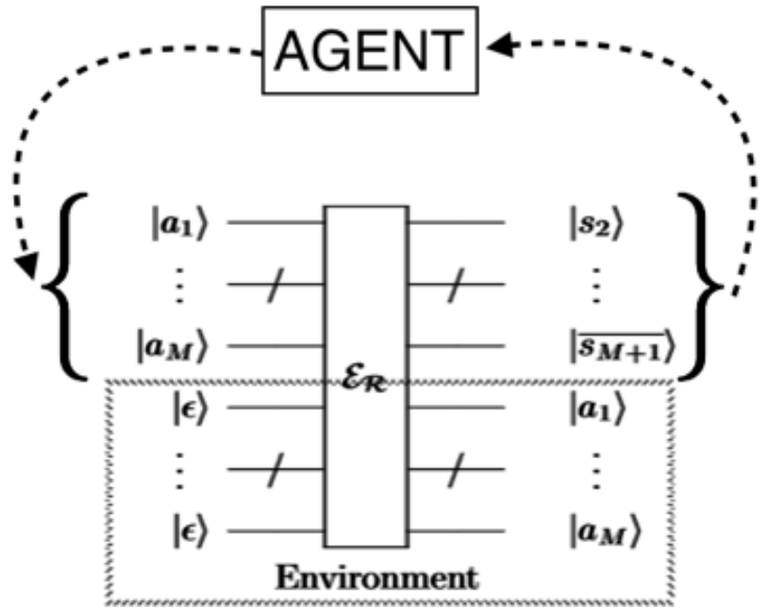
Classically specified oracle $\vec{x} = (x_1, x_2, \ldots, x_n) \rightarrow f(\vec{x}) \in \{0, 1\}$

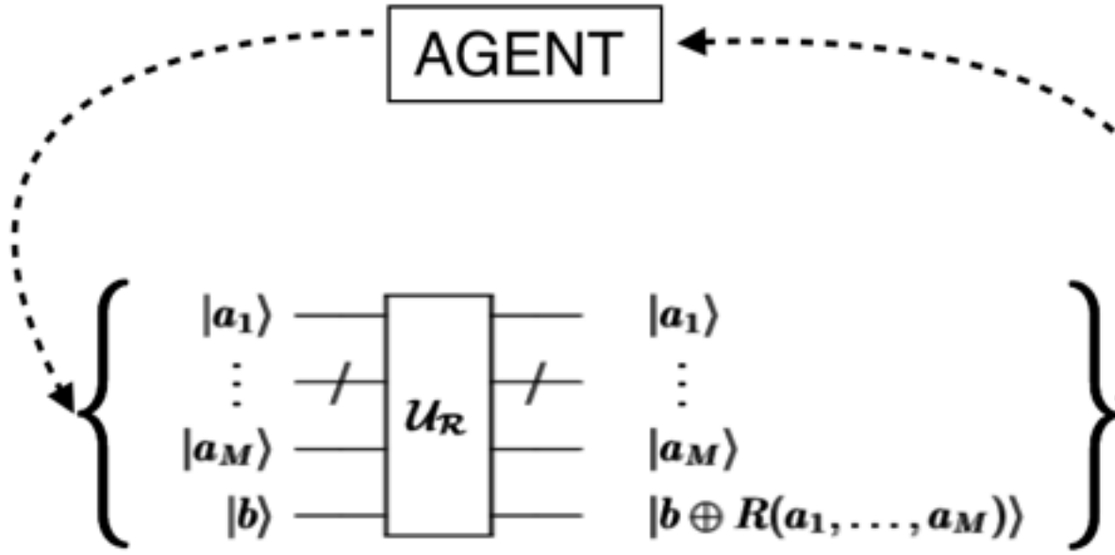


Oraculization (continued)

(taming the open environment)

4)



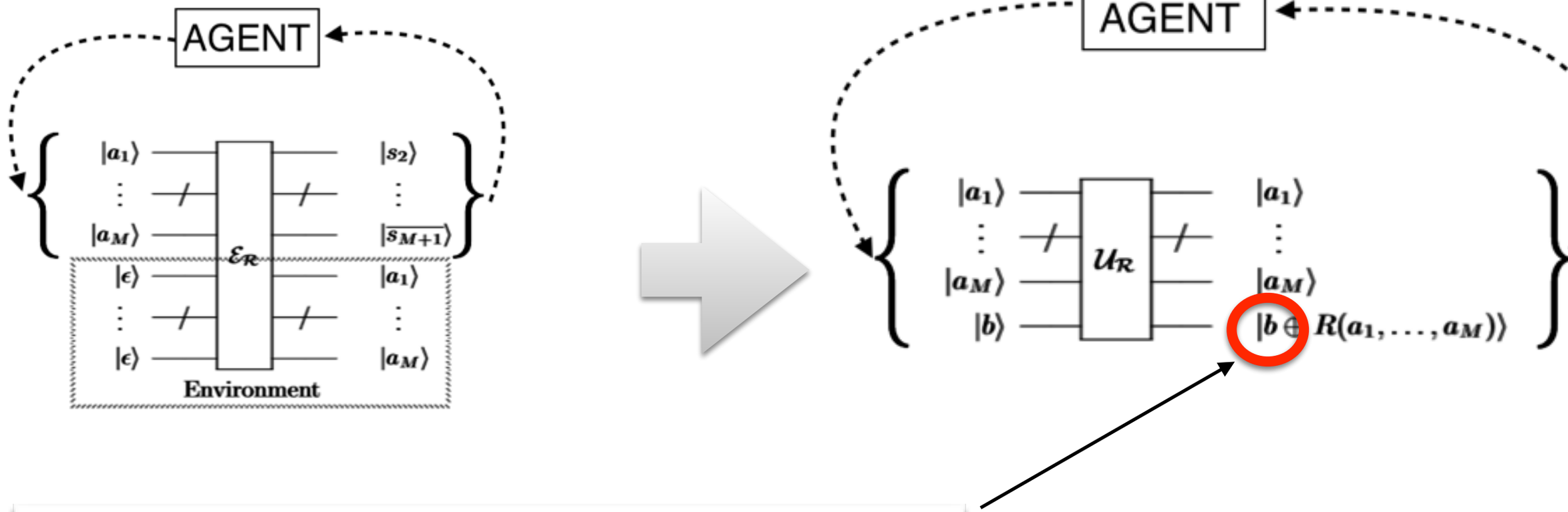


Chiribella, G., D'Ariano, G. M. & Perinotti, P. Quantum Circuit Architecture. *Phys. Rev. Lett.* **101**, 060401 (2008).

Oraculization (continued)

(taming the open environment)

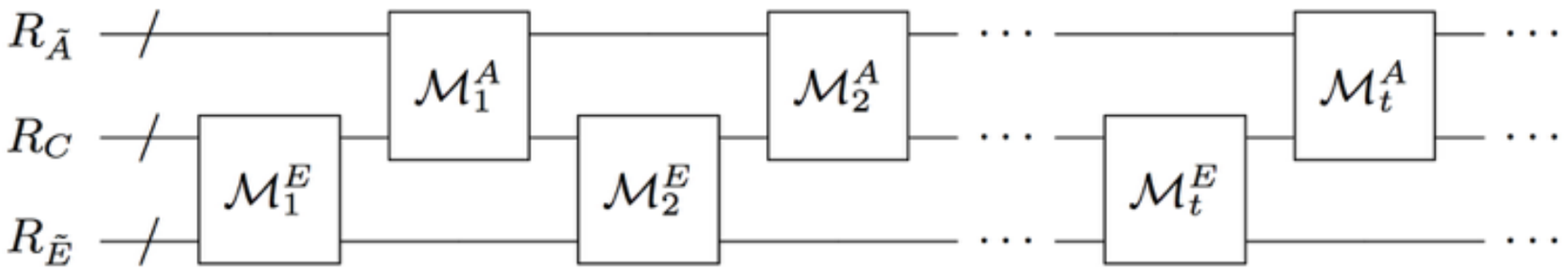
4)



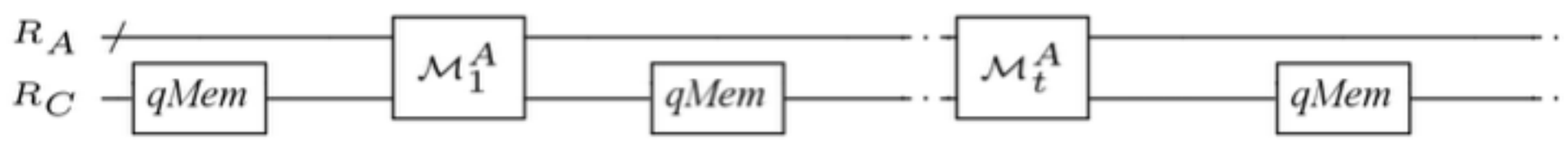
Relevant information about (aspect of) environment accessible encoded in *effective quantum oracle*

Oraculization in data-driven ML?

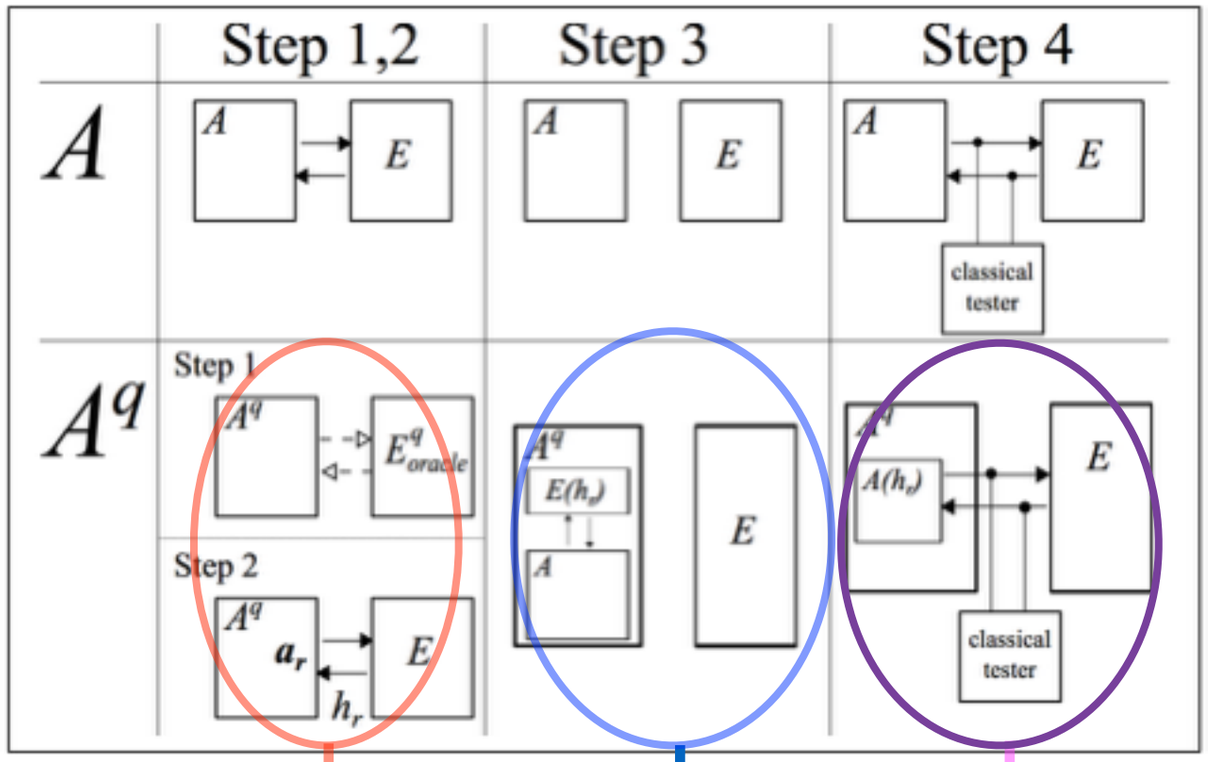
RL:



DL:



Result schemata: *for every A we construct A^q s.t. $A^q > A$…*



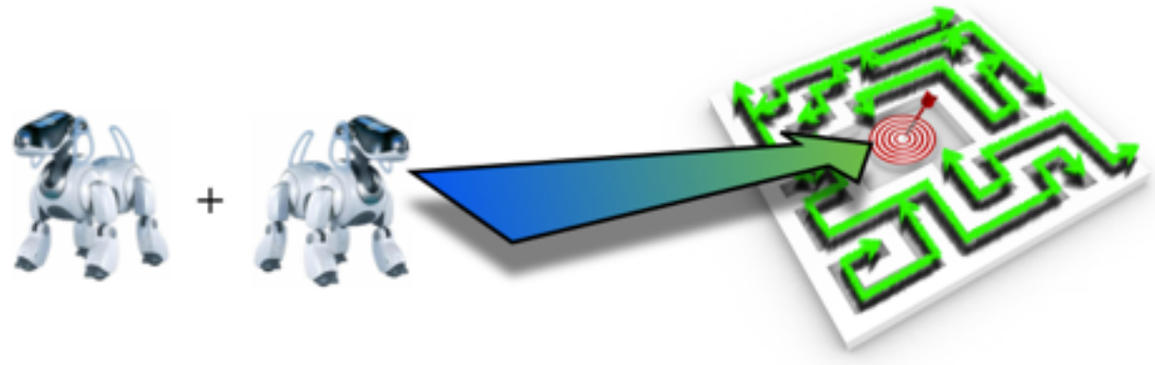
find useful info using quantum access

optimize a (given) learning mechanism

compare to given (best) classical agent

First result: **quantum upgrade** of a learning model

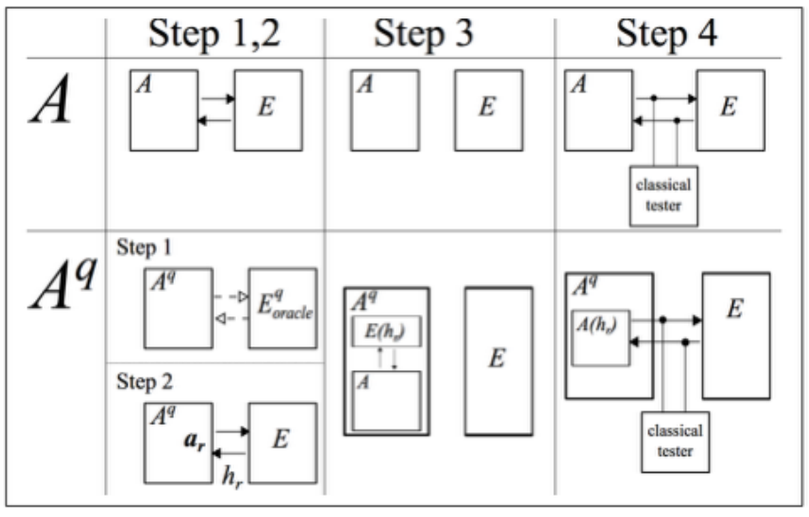
polynomial improvements
for a broad class of task environments
*by using **Grover-like amplification***



1. Grover finds "rewarding sequences"

2. Algorithm is "pre-trained" to focus more on rewarding space of policies

3. Works whenever algorithm makes sense, and environment s.t. it prefers winners

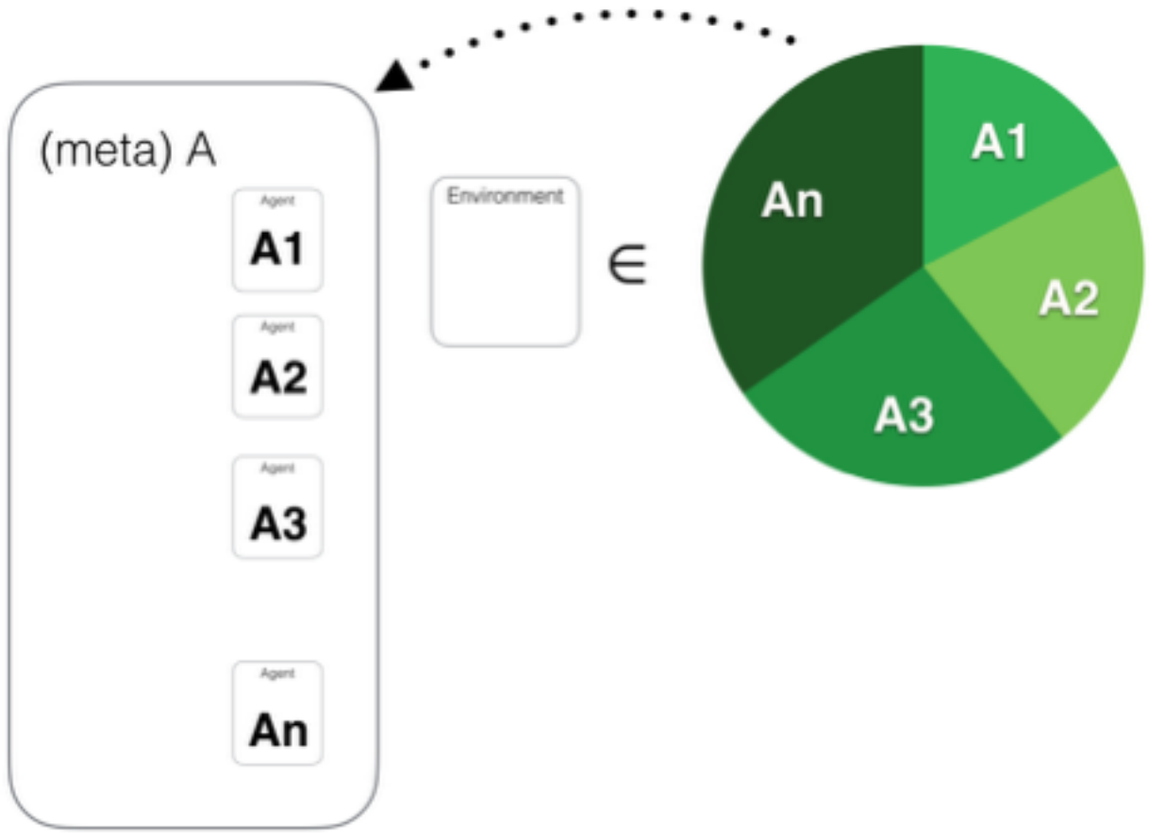


Theorem 1. *Let E be a controllable environment, over action space $\mathcal{A}$, thus it is, on the agent's demand, accessible in the form $E^q_{control}$. Moreover, let E correspond to a deterministic, fixed-time M, single-win game, with a unique winning sequence of length M, for the period of $O(|\mathcal{A}|^M)$ time-steps (after which it no longer needs to be controllable, nor deterministic, fixed-time, single win). Let A be a learning agent such that (E, A) are luck-favoring for all histories, relative to some figure of merit $Rate(\cdot)$, which is increasing in the number of rewards in the history, and which only depends on the rewards. Then there exists a quantum learning agent A^q based on A which outperforms A in terms of $Rate(\cdot)$ and relative to a chosen sporadic classical tester.*

V. Dunjko, J. M. Taylor, H. J. Briegel
Quantum-enhanced machine learning
Phys. Rev. Lett. **117**, 130501 (2016)

Second result: **quantum speedup** of *meta-learning*

polynomial improvements
for meta-learning settings by using
***quantum optimization** methods*



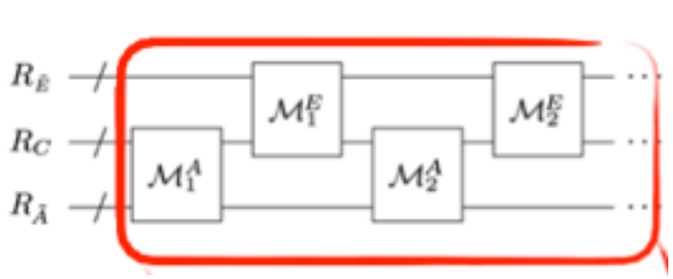
- given a fixed model and environment, the joint AE interaction can be oracularized
- model parameters can be “quantum-optimized” in (any) given quantum-accessible environment

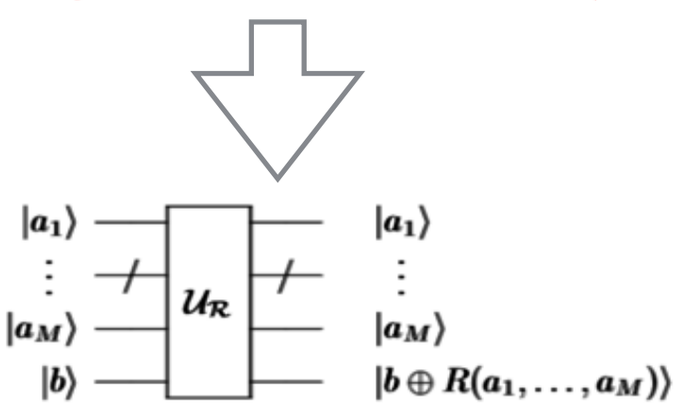
V. Dunjko, J. M. Taylor, H. J. Briegel
Advances in Quantum Reinforcement Learning
IEEE SMC 2017

quantum upgrade of a learning model

- underlying Quant. alg.: Grover-based optimization
- oraculization of Environment

$$|a_1,\dots,a_M\rangle|\epsilon,\dots,\epsilon\rangle\Big|\epsilon,\overset{0}{\psi}\Big\rangle \to |a_1,\dots,a_M\rangle|s_2,\dots,s_M\rangle\Big|s_{M+1},\overset{r}{\psi}\Big\rangle.$$



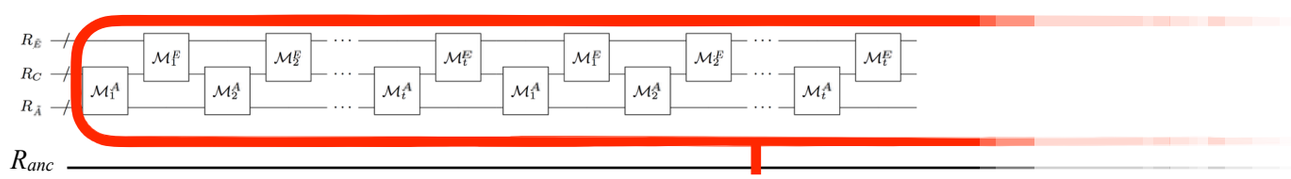


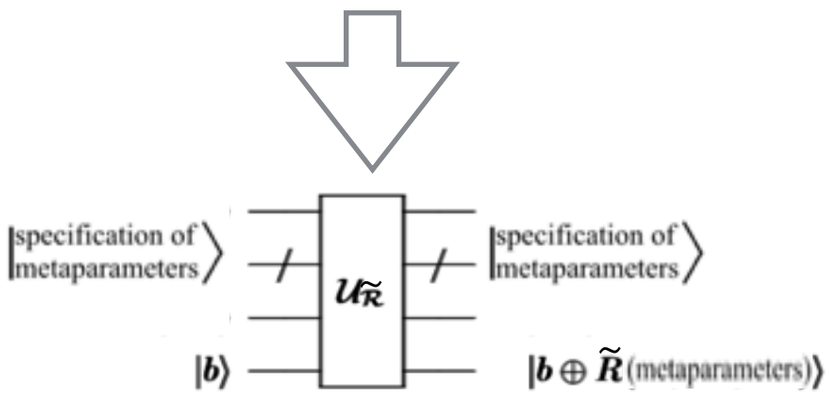
oracle encoding
properties of environment

quantum speedup of *meta-learning*

- underlying Quant. alg.: Grover-based optimization
- oraculization of Agent-Env. interaction

$$|\alpha\rangle_{A_{anc}}|0\rangle_{A_{stats}}|0\rangle_{A_{work}}\otimes|0\rangle_E \xrightarrow{\text{oraculized AE}} |\alpha\rangle_{A_{anc}}|\mathbf{stats}\rangle_{A_{stats}}|0\rangle_{A_{work}}\otimes|0\rangle_E$$





oracle encoding performance of
Agent-Environment interaction

Quantum folklore:

quadratic improvements a-plenty,
super-polynomial hard to come by

What about reinforcement learning?

Third result: **absolute separation** of classical & quantum learning

RL is *VERY BIG. Trivial answers aplenty*

Given oracular problem, construct MDPs which

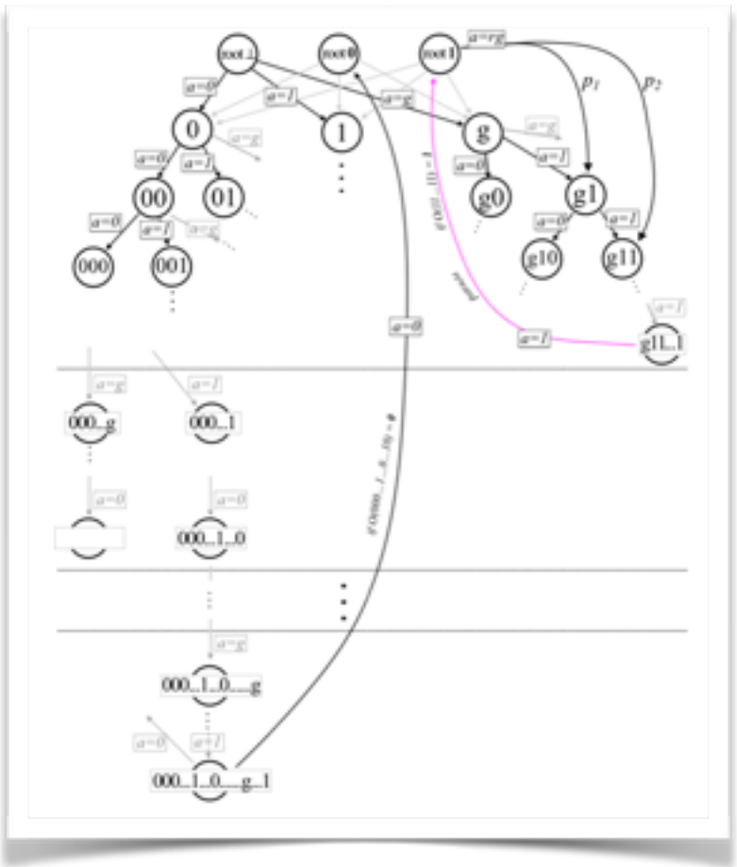
1) have “generic RL properties”:
   delayed rewards, genuine actions,
   controlled amount degeneracy in optimal policies (stochasticity)

2) Are hard to learn for **any** classical agent

3) Oraculization possible, and reduces to oracles which allow superpolynomial speed-ups

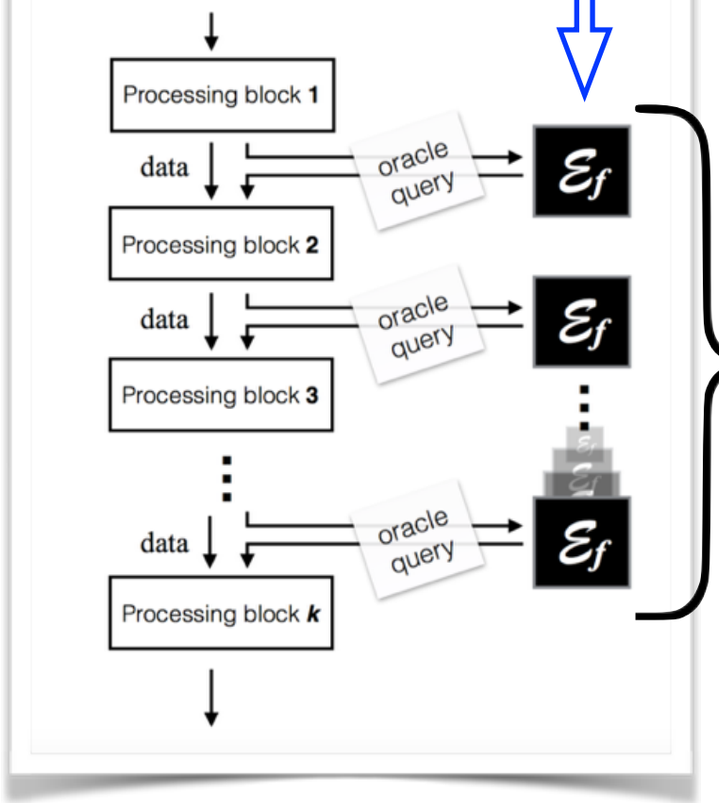
***super-polynomial* improvements possible**
for a class of ***genuine RL*** *task environments*
via reduction to ***(generalized) Recursive Fourier Sampling***

Complicated, and
genuine RL problem



oraculization
process

RFS-type oracle
hiding a necessary “key”

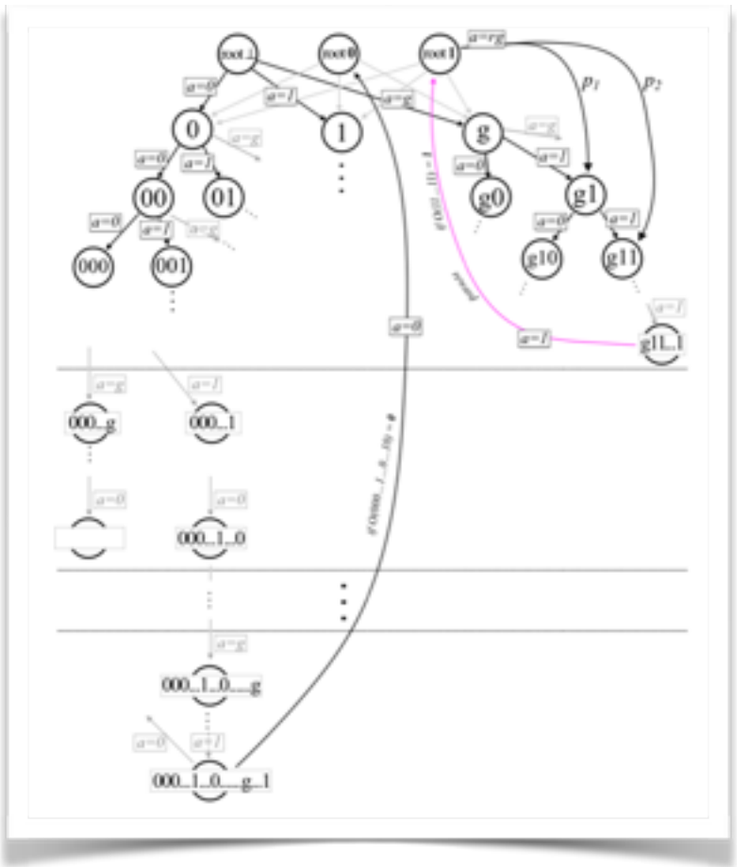


super-polynomial
separations known

V. Dunjko, Y. Liu, X. Wu, J. M. Taylor
Super-polynomial separations for quantum reinforcement learning (in preparation)

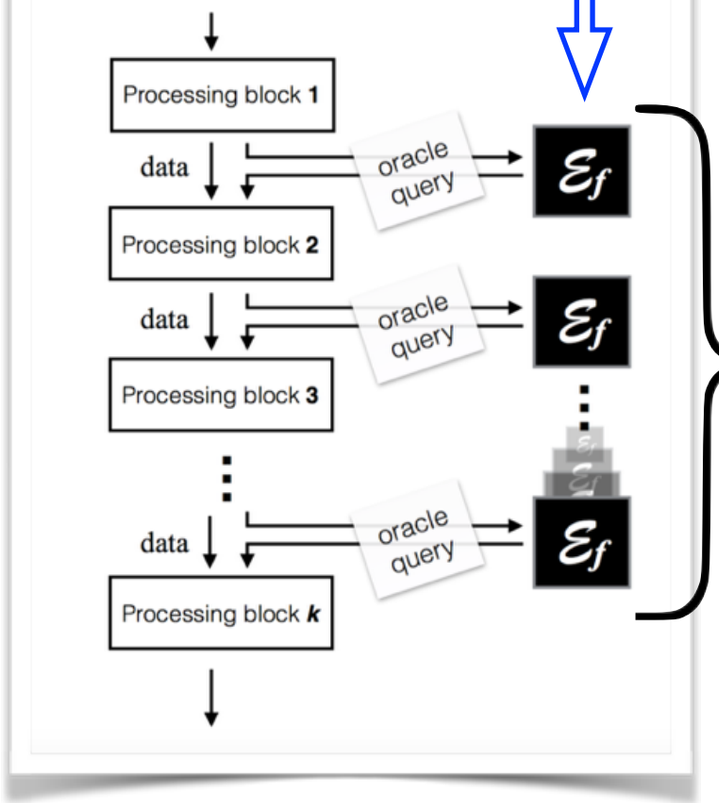
***super-polynomial* improvements possible**
for a class of ***genuine RL*** *task environments*
via reduction to ***(generalized) Recursive Fourier Sampling***

Complicated, and
genuine RL problem



oraculization
process

RFS-type oracle
hiding a necessary “key”



super-polynomial
separations known

can be adapted to e.g. Simon’s problem:
strict exponential separation possible!

V. Dunjko, Y. Liu, X. Wu, J. M. Taylor
Super-polynomial separations for quantum reinforcement learning (in preparation)

quantum upgrade of a learning model	**absolute separation** c & q learning
“natural” environments: mazes	Start with solution; oracle problems with separation
oraculization to most natural info: rewarding sequences	reverse-enigneered matching environments
Grover is (almost always) applicable: large “natural” class of settings where improvements possible	identify generalizations/deformations which a) maintain classical hardness b) allow for meaningful oraculization and application of known q. algorithms
	(shhh: call those above “the generic ones”!)

- Perspectives of (*this type of*) Quantum Reinforcement learning

  - *QAE: <u>fundamental</u> questions of learnability*
    Quantum AI and quantum Turing test (behavioral intelligence)

  - *QRL-enhancements (theory): connections to oracular models.*
    Elitzur-Vaidman & counterfactual computation. Q. Metrology, q. illumination.

  - *Making it concrete: connection to QML!*

  - *QIP overlaps: generalization of standard oracular models*

  - *QRL-enhancements:*
    Minimal oraculizations.
    Quantum agents in quantum systems (quantum experiments, QQ ML).

    Model-based agents. Quantum behavioral databases.

The end