# Using Reinforcement Learning for Quantum Control in Magnetic Resonance

Will Kaufman[1] Benjamin Alford[1] Pai Peng[2]
Xiaoyang Huang[2] Linta Joseph[1] Paola Cappellaro[2]
Chandrasekhar Ramanathan[1]

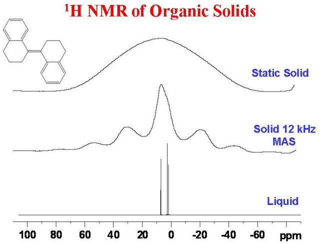
[1]Department of Physics and Astronomy, Dartmouth College
Hanover, NH 03755, USA

[2]Research Laboratory of Electronics, Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, USA

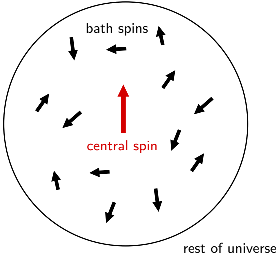
APS March Meeting 2021
Session X32: Quantum Machine Learning III

NSF support under Grants OIA-1921199, PHY1734011, and PHY1915218.

# Magnetic dipolar interactions in solids




From Facey 2008.




Magentic dipolar interactions. . .

- Broaden spectral lines in NMR (Linta Joseph, J33.00003)
- Lead to decay of central spin coherence in bath (Ethan Williams, L29.00010)

$$H_{\mathrm{sys}} = \sum_i \delta_i I_z^i + \sum_{i,j} d_{ij} \left(3 I_z^i I_z^j - \mathbf{I^i} \cdot \mathbf{I^j}\right) = H_{\mathrm{CS}} + H_{\mathrm{D}}$$

Decoupling dipolar interactions would narrow spectral lines and increase coherence times.

# Average Hamiltonian theory

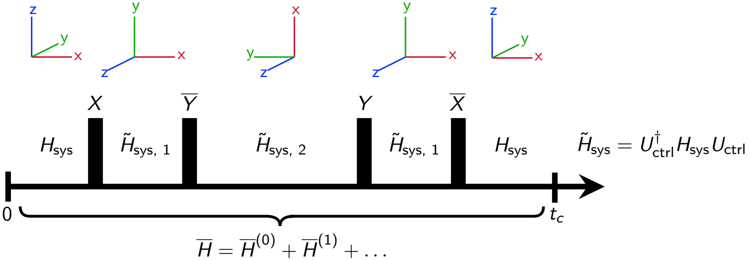
If...

- Consider cyclic and periodic pulse sequences

$$U_{\mathrm{ctrl}}(t_c) = \mathbb{1}, H_{\mathrm{ctrl}}(t) = H_{\mathrm{ctrl}}(t + Nt_c)$$

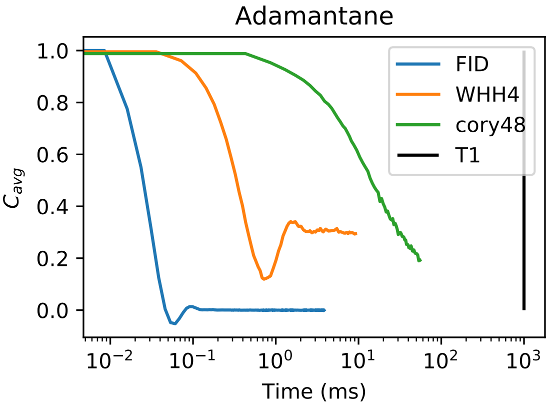
- Observe system stroboscopically ($t = Nt_c$)

...then system appears to evolve under an effective *average* Hamiltonian.



# Existing approaches to Hamiltonian engineering

- WAHUHA 4-pulse sequence (Waugh *et al.* 1968), decouples dipolar interaction to lowest-order
- CORY 48-pulse sequence (Cory *et al.* 1990) designed analytically using AHT to be robust to experimental imperfections, decouples *all* interactions to second order

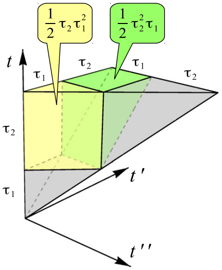


# AHT limitations

$$\overline{H}^{(0)} = \frac{1}{t_c}\int_0^{t_c} \widetilde{H}_{\text{sys}}(t)dt$$

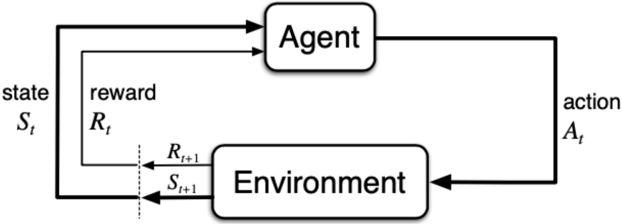
$$\overline{H}^{(1)} = \frac{1}{2it_c}\int_0^{t_c} dt_1 \int_0^{t_1} dt_2 \left[\widetilde{H}_{\text{sys}}(t_1), \widetilde{H}_{\text{sys}}(t_2)\right]$$

$$\overline{H}^{(2)} = -\frac{1}{6t_c}\int_0^{t_c} dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 \left\{\left[\widetilde{H}_{\text{sys}}(t_1), \left[\widetilde{H}_{\text{sys}}(t_2), \widetilde{H}_{\text{sys}}(t_3)\right]\right] + \left[\left[\widetilde{H}_{\text{sys}}(t_1), \widetilde{H}_{\text{sys}}(t_2)\right], \widetilde{H}_{\text{sys}}(t_3)\right]\right\}$$




From Brinkmann 2016.

# Reinforcement learning for Hamiltonian engineering

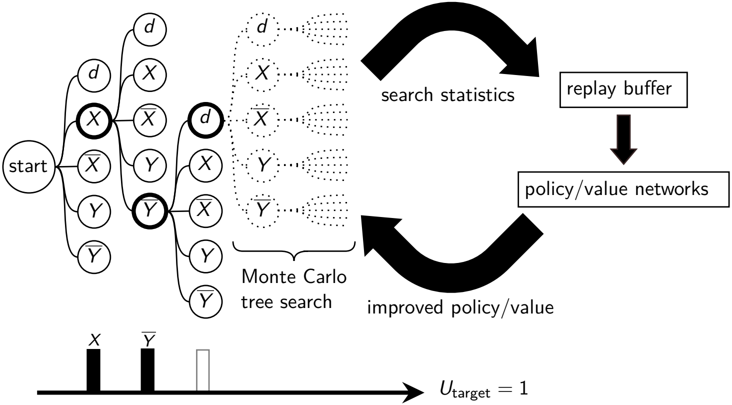



From Sutton & Barto 2018.

- State $\rightarrow$ propagator
- Action $\rightarrow$ control pulses
- Reward $\rightarrow$ propagator fidelity $\left(\mathrm{Re}\,\frac{\mathrm{Tr}\left(U^\dagger_{\mathrm{target}}U(t)\right)}{\mathrm{Tr}(\mathbb{1})}\right)$

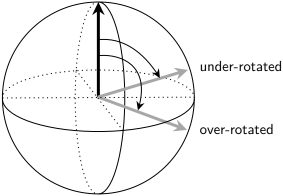
# Constructing pulse sequences using AlphaZero

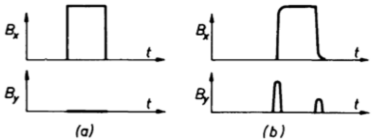
Implemented AlphaZero algorithm (Silver *et al.* 2018, originally for Chess, Shogi, and Go), though there are many different RL approaches (Peng *et al.* 2021, P33.00001).



# Computational results

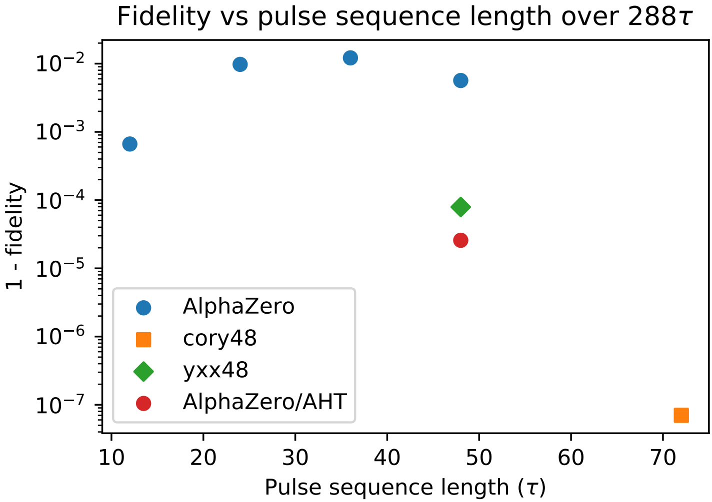
Goal: decouple all interactions ($\overline{H} = 0$) in strongly coupled spin systems with experimental imperfections.



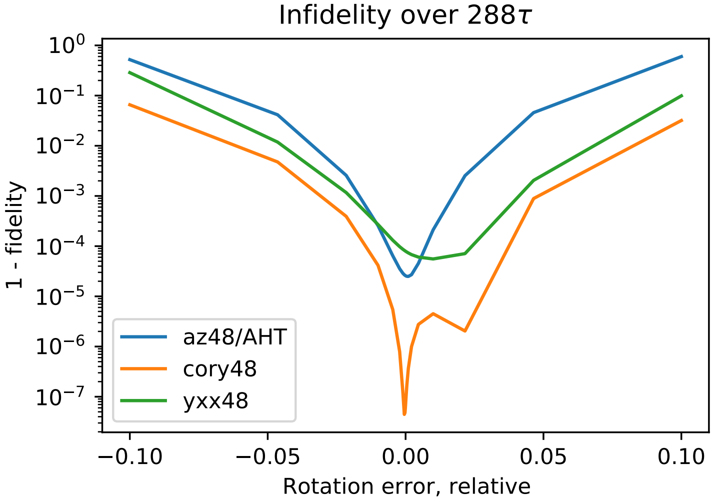




- Unconstrained search (*tabula rasa*, no AHT knowledge), 1% pulse rotation error, different pulse sequence lengths $(12\tau, 24\tau, 36\tau, 48\tau)$
- AHT-constrained search, 1% pulse rotation error, $48\tau$ sequence length

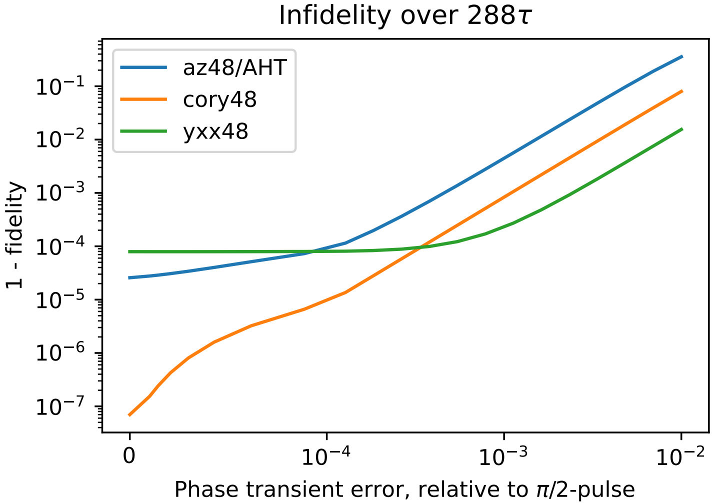
# Fidelity vs. pulse sequence length



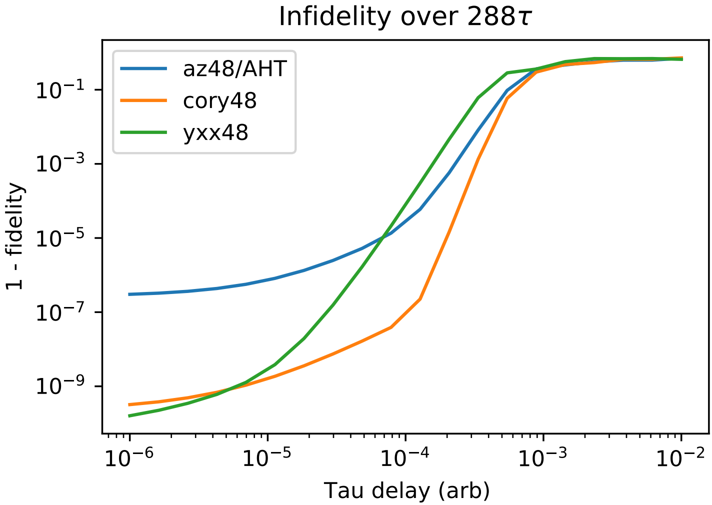
# Robustness to pulse rotation error: AHT constraints



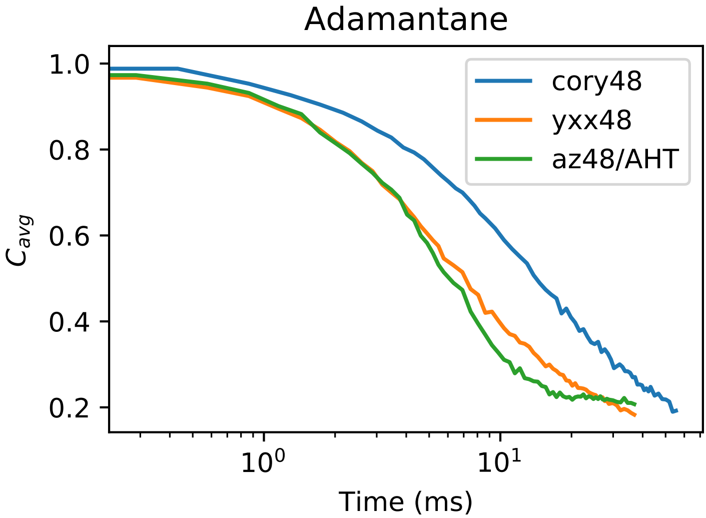
# Robustness to phase transient error



# Fidelity vs. tau spacing



# Experimental results



# Summary

- Decoupling dipolar interactions is important for narrowing linewidths, increasing coherence times
- RL is promising new tool to design new pulse sequences
  - Tailored control for specific system characteristics and errors
  - Best-performing approach likely is a mix of RL and knowledge from AHT

# Summary

- Decoupling dipolar interactions is important for narrowing linewidths, increasing coherence times
- RL is promising new tool to design new pulse sequences
  - Tailored control for specific system characteristics and errors
  - Best-performing approach likely is a mix of RL and knowledge from AHT

Thanks for listening!

# References I


1. Brinkmann, A. Introduction to average Hamiltonian theory. I. Basics. *Concepts in Magnetic Resonance Part A* **45A,** e21414. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cmr.a.21414. https://onlinelibrary.wiley.com/doi/abs/10.1002/cmr.a.21414 (2016).
2. Cory, D., Miller, J. & Garroway, A. Time-suspension multiple-pulse sequences: applications to solid-state imaging. *Journal of Magnetic Resonance (1969)* **90,** 205–213. ISSN: 0022-2364. https://www.sciencedirect.com/science/article/pii/002223649090380R (1990).
3. Facey, G. *1H NMR Spectra of Solids.*
4. Gerstein, B. & Dybowski, C. *Transient Techniques in NMR of Solids: An Introduction to Theory and Practice.* 1st ed. (Academic Press, 1985).

# References II


5. Peng, P. *et al.* *Deep reinforcement learning for quantum Hamiltonian engineering.* 2021. arXiv: 2102.13161 [quant-ph].

6. Silver, D. *et al.* A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362,** 1140–1144. ISSN: 0036-8075. eprint: https://science.sciencemag.org/content/362/6419/1140.full.pdf. https://science.sciencemag.org/content/362/6419/1140 (2018).

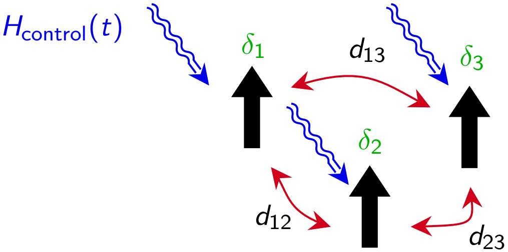
7. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction.* (MIT press, 2018).

8. Waugh, J. S., Huber, L. M. & Haeberlen, U. Approach to High-Resolution nmr in Solids. *Phys. Rev. Lett.* **20,** 180–182. https://link.aps.org/doi/10.1103/PhysRevLett.20.180 (5 Jan. 1968).

## Quantum control in spin systems

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

Can use $H_{\text{ctrl}}(t)$ to achieve a unitary transformation $U$ given by an effective Hamiltonian $H_{\text{eff}}$.




$$H_{\text{system}} = \sum_i \delta_i I_z^i + \sum_{i,j} d_{ij} \left( 3 I_z^i I_z^j - \mathbf{I^i} \cdot \mathbf{I^j} \right) = H_{CS} + H_D$$

# Quantum control in spin systems

$$
\begin{aligned}
H_{\mathrm{sys}} &= \sum_i \delta_i I_z^i + \sum_{i,j} d_{ij} \left( 3 I_z^i I_z^j - \mathbf{I^i} \cdot \mathbf{I^j} \right) \\
&= H_{\mathrm{CS}} + H_{\mathrm{D}}
\end{aligned}
$$

$$
H_{\mathrm{ctrl}}(t) = -B_1(t) \sum_i \gamma_n^i I_x^i - B_2(t) \sum_i \gamma_n^i I_y^i
$$

# Average Hamiltonian Theory (AHT)

The time-evolution operator (or propagator) follows the differential equation

$$i\frac{d}{dt}U(t) = H(t)U(t)$$

$$U(0) = \mathbb{1}$$

The Magnus Expansion gives an exponential solution for the propagator via an average Hamiltonian $\overline{H}$ at time $t$

$$U(t) = \exp\left(-i\overline{H}t\right)$$

with $\overline{H} = \overline{H}^0 + \overline{H}^1 + \dots$.
The series converges rapidly when $t||H|| \ll 1$.

# AHT (cont.)

We often work in the interaction frame of the control Hamiltonian, with transformation operator

$$\frac{d}{dt}U_{\text{ctrl}}(t) = -iH_{\text{ctrl}}(t)U_{\text{ctrl}}(t)$$
$$U_{\text{ctrl}}(0) = \mathbb{1}$$

So the Hamiltonian in the interaction frame becomes

$$\widetilde{H}(t) = \widetilde{H}_{\text{sys}}(t) = U_{\text{ctrl}}(t)^{\dagger}H_{\text{sys}}U_{\text{ctrl}}(t)$$

Brinkmann 2016.

# AHT: Pulse sequences

If a pulse sequence is both cyclic and periodic Gerstein & Dybowski 1985

$$U_{\mathrm{ctrl}}(t_c) = T \exp\left(-i \int_0^{t_c} H_{\mathrm{ctrl}}(t) dt\right) = \pm\mathbb{1} \text{ (cyclic)}$$
$$H_{\mathrm{ctrl}}(t) = H_{\mathrm{ctrl}}(t + Nt_c) \text{ (periodic)}$$

then the interaction frame and the lab frame coincide at multiples of the cycle time, and the propagator can be given by

$$U(t_c) = \exp\left(-it_c(\overline{H}^{(0)} + \overline{H}^{(1)} + \dots)\right)$$

Higher-order terms for average Hamiltonian become nasty...

# AHT: Special Cases

- Symmetric pulse sequences ($H(\tau) = H(t_c - \tau)$): all odd-order terms in average Hamiltonian are zero
- Antisymmetric pulse sequences ($H(\tau) = -H(t_c - \tau)$): all terms in average Hamiltonian are zero

# Simulation/RL parameters

- $N = 3$ spin-1/2 system, $\delta_i \sim \mathcal{N}(0, 1)$, $d_{ij} \sim \mathcal{N}(0, 100)$
- Delay $\tau = 10^{-4}$, pulse length $t_p = 10^{-5}$
- Ensemble of 50 spin systems with different chemical shifts and dipolar interactions
- Replay buffer size: $10^6$ "experiences" ($(s, a, r)$)
- Batch size: 2048
- Training duration: $10^4$ training steps

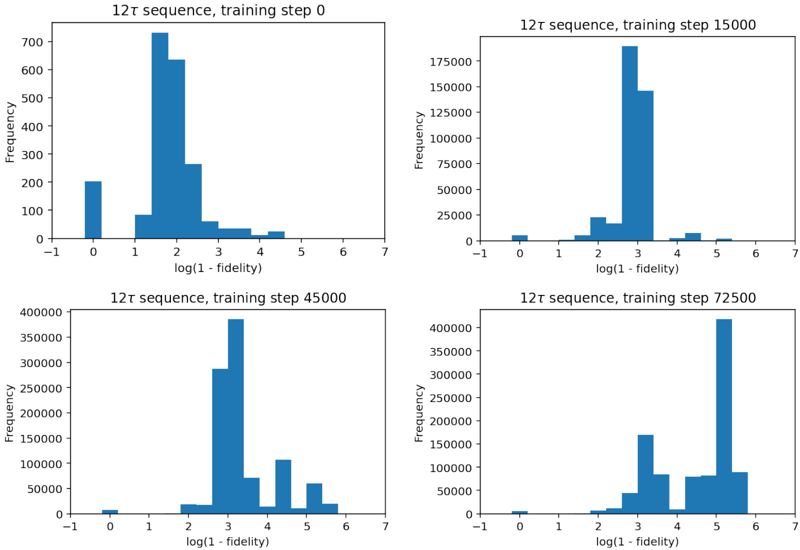
$$\text{fidelity}(U, U_{\text{target}}) = \text{Re}\,\frac{\text{Tr}\left(U_{\text{target}}^{\dagger} U(t)\right)}{\text{Tr}\left(\mathbb{1}\right)}$$

For RL algorithm performance, use log infidelity as "reward"

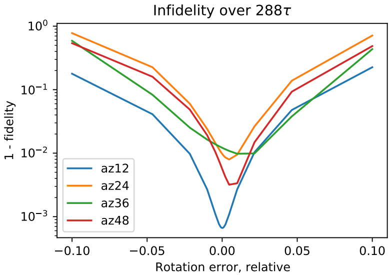
$$r = -\log\left(1 - \text{fidelity}\right)$$

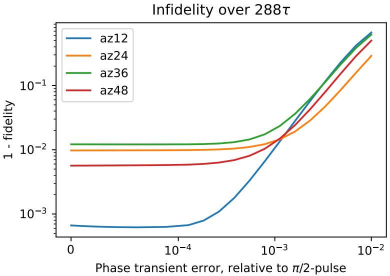
$$r = 4 \iff \text{fidelity} = 0.\underline{9999}$$

# Computational results: AlphaZero algorithm learns



# Robustness to errors: unconstrained search



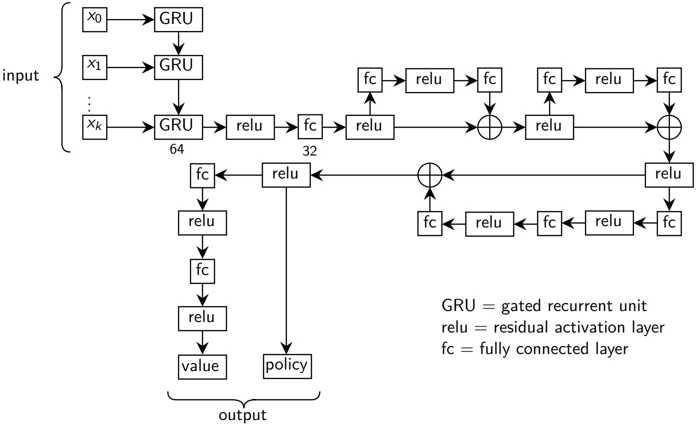



# Comparison between RL approaches

Different RL algorithm used by our collaborators (Peng *et al.* 2021).

| Characteristic | Evolutionary Reinforcement Learning | AlphaZero |
|---|---|---|
| State representation | Sequence of previous pulses | Same |
| Action space | Delay or $\pi/2$-pulse along $\pm$X, $\pm$Y | Same |
| Learning method | Evolutionary algorithms (gradient-free) | Tree search and experience replay (gradient based) |
| Prior knowledge | Builds longer sequences from shorter ones | Uses AHT to prune tree search |
| Pulse sequences ($H_{\text{eff}} = 0$) | yxx48 | az48 |

# Neural network structure




GRU = gated recurrent unit
relu = residual activation layer
fc = fully connected layer

# AlphaZero algorithm

Explore new pulse sequences

1. Start with a zero-length pulse sequence as the root node
2. With the given root node, perform Monte Carlo Tree Search (MCTS) to explore potential pulses
   MCTS uses a neural network to estimate the prior probabilities for selecting each pulse and the value (fidelity) for the final pulse sequence
3. Sample the next pulse from the root node's children weighted by their visit counts
4. Repeat steps 2-4 until a complete pulse sequence is determined
5. Record the child nodes' visit counts and final pulse sequence fidelity to a data buffer for training

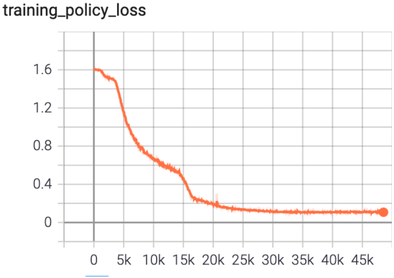
Parameters for MCTS, training, etc.

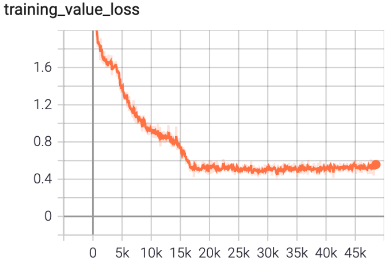
# AlphaZero algorithm (cont.)

Train neural networks on collected data

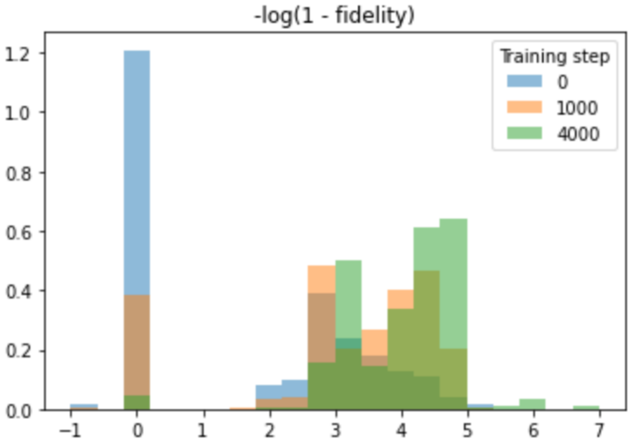
- Policy loss: want to minimize the difference between MCTS visit counts $\mathbf{p}$ and learned policy $\pi_\theta$
- Value loss: want to minimize the difference between calculated fidelity from pulse sequence $z$ and predicted fidelity from neural network $v$
- L2 regularization: prevent overfitting to data
- $l(\theta) = -\mathbf{p} \cdot \log \pi_\theta + (z - v)^2 + c||\theta||^2$

# Neural network training



# Training performance



# Pulse sequences identified using AlphaZero

az48 pulse sequence (decouple all interactions):

$-X, \tau, Y, \tau, Y, \tau, X, \tau, Y, \tau, Y, \tau$
$-Y, \tau, X, \tau, X, \tau, -Y, \tau, X, \tau, X, \tau$
$Y, \tau, X, \tau, X, \tau, -Y, \tau, X, \tau, X, \tau$
$-Y, \tau, X, \tau, -Y, \tau, X, \tau, X, \tau, -Y, \tau$
$-X, \tau, -X, \tau, Y, \tau, Y, \tau, -X, \tau, Y, \tau$
$Y, \tau, -Y, \tau, X, \tau, -Y, \tau, -Y, \tau, X, \tau$
$-Y, \tau, X, \tau, X, \tau, -Y, \tau, X, \tau, X, \tau$
$-Y, \tau, -X, \tau, -X, \tau, -Y, \tau, -X, \tau, -X, \tau$

# RL advantages and disadvantages

- Generalized approach to learning problem: no assumed prior knowledge
- Can tailor problem to specific system of interest (e.g. strongly coupled system, timing precision constraints)
- Robustness against known errors by including them in simulation of spin system

# RL advantages and disadvantages

- Generalized approach to learning problem: no assumed prior knowledge
- Can tailor problem to specific system of interest (e.g. strongly coupled system, timing precision constraints)
- Robustness against known errors by including them in simulation of spin system
- Computationally expensive
- Poor accuracy of many-body spin simulations
- No guarantees for convergence to optimal (or good) solution