Digital zero noise extrapolation for quantum error mitigation

Will Zeng

QRE 2020, in Cyberspace

May 30ish, 2020

Joint work with: Tudor Giurgica-Tiron [Stanford], Yousef Hindy [Stanford],
Ryan LaRose [Unitary Fund, UMich], Andrea Mari [Unitary Fund]

Engineering Division

arXiv:2005.10921

Outline

- Why error-mitigation?
- Zero-noise extrapolation (noise-scaling + extrapolation)
- Noise scaling:
    - *Unitary Folding*: a framework for digital noise scaling
    - *Parameter Noise scaling*: a noise model specific framework
- Extrapolation as inference
    - *Non-adaptive Methods*
    - *Adaptive Methods*
- Benchmarks

Error-mitigation is critical for noisy quantum computing

No overhead → Lots of overhead

Today

Cross-your-fingers method

Tomorrow

Error mitigation

- Probabilistic Error Cancellation [1,2]
- Randomized Compiling [3]
- Dynamical Decoupling [4-7]
- Quantum optimal control [8,9]
- **Zero-noise extrapolation**

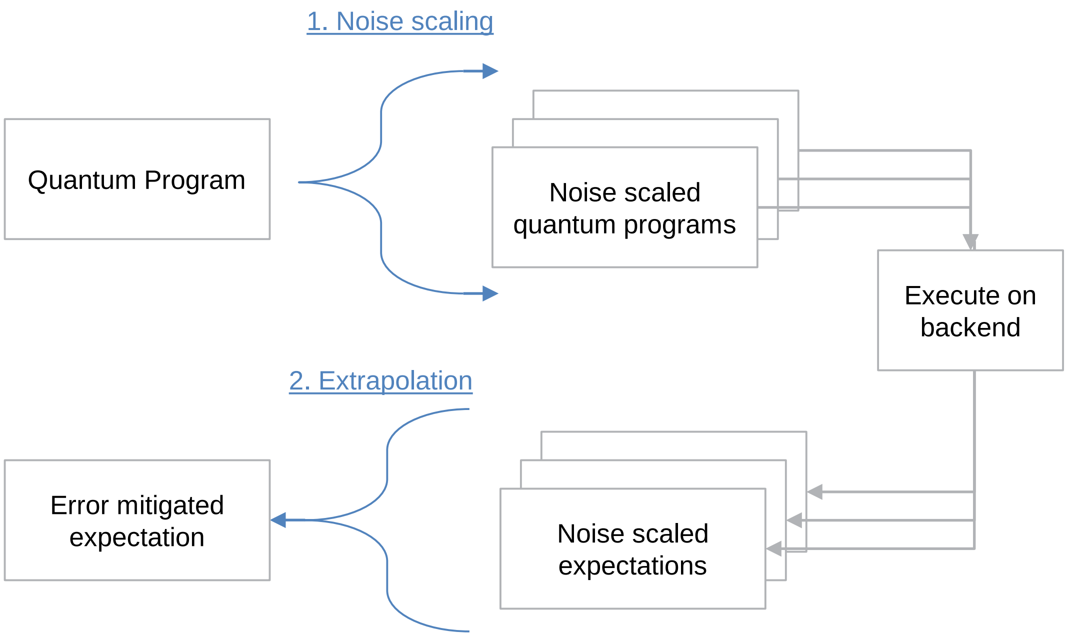
The Future

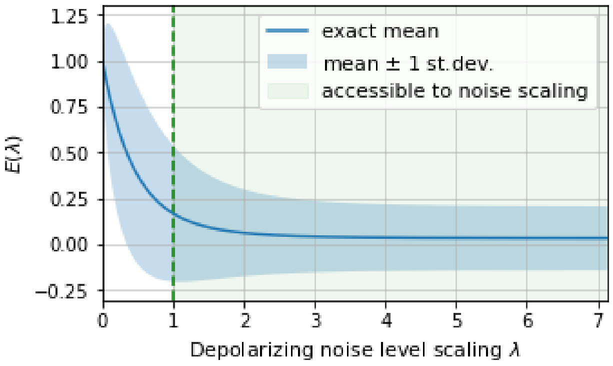
Error correction

- **Uses additional qubits**
- **Requires fast classical control**

[1] K. Temme, S. Bravyi, and J. M. Gambetta, "Error Mitigation for Short Depth Quantum Circuits," Physical Review Letters, vol. 119, p. 180509, 11 2017.
[2] S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications," Physical Review X, vol. 8, no. 3, p. 031027, 2018.
[3] J. J. Wallman and J. Emerson, "Noise tailoring for scalable quantum computation via randomized compiling," Physical Review A, vol. 94, no. 5, p. 052325, 2016.
[4]] E. Knill, "Quantum computing with realistically noisy devices," Nature, vol. 434, no. 7029, pp. 39–44, 2005.
[5] L. Viola and E. Knill, "Random decoupling schemes for quantum dynamical control and error suppression," Physical review letters, vol. 94, no. 6, p. 060502, 2005.
[6] B. Pokharel, N. Anand, B. Fortman, and D. A. Lidar, "Demonstration of fidelity improvement using dynamical decoupling with superconducting qubits," Physical review letters, vol. 121, no. 22, p. 220502, 2018.
[7] P. Sekatski, M. Skotiniotis, and W. Dur, "Dynamical decoupling leads to ¨ improved scaling in noisy quantum metrology," New Journal of Physics, vol. 18, no. 7, p. 073034, 2016.
[8] H. Ball, M. J. Biercuk, A. Carvalho, R. Chakravorty, J. Chen, L. A. de Castro, S. Gore, D. Hover, M. Hush, P. J. Liebermann, et al., "Software tools for quantum control: Improving quantum computer performance through noise and error suppression," arXiv preprint arXiv:2001.04060, 2020.
[9] T. J. Green, J. Sastrawan, H. Uys, and M. J. Biercuk, "Arbitrary quantum control of qubits in the presence of universal noise," New Journal of Physics, vol. 15, no. 9, p. 095004, 2013.

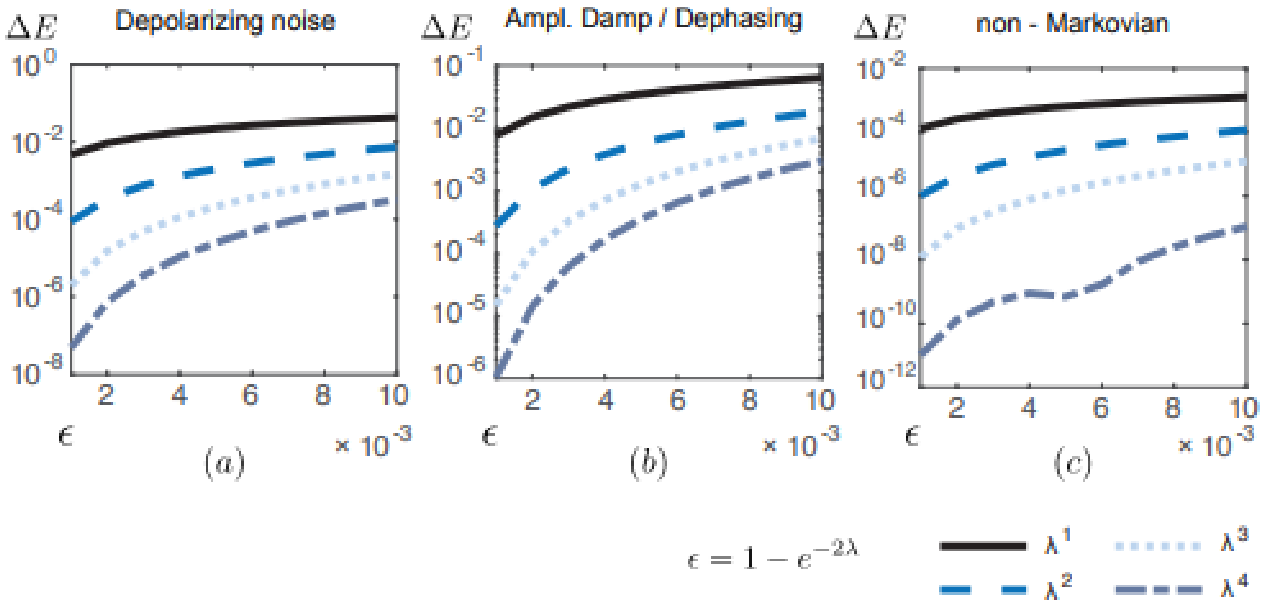
Zero-noise extrapolation





[1] Y. Li and S. C. Benjamin, "Efficient Variational Quantum Simulator Incorporating Active Error Minimization," Physical Review X, vol. 7, 6 2017.
[2] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, ´ and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," Nature, vol. 567, no. 7749, pp. 491–495, 2019.
[3] K. Temme, S. Bravyi, and J. M. Gambetta, "Error Mitigation for Short Depth Quantum Circuits," Physical Review Letters, vol. 119, p. 180509, 11 2017.

Zero-noise Extrapolation can work very well

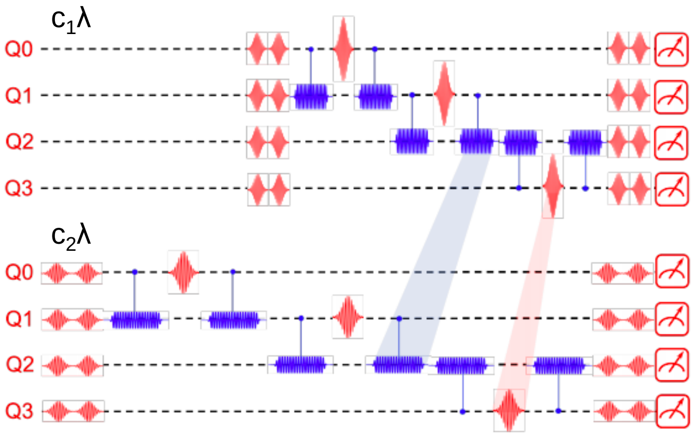


[1] *K. Temme, S. Bravyi, and J. M. Gambetta, "Error Mitigation for Short Depth Quantum Circuits," Physical Review Letters, vol. 119, p. 180509, 11 2017.*

Zero-noise extrapolation

Previously introduced and studied with physical level noise scaling & Richardson extrapolation

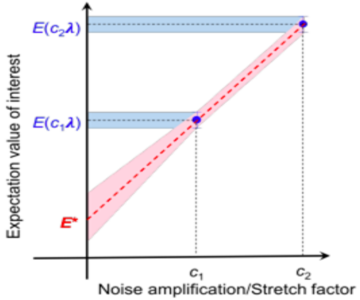
Recalibration Noise-scaling



Richardson Extrapolation

$$E_K(\lambda) = E^* + \sum_{k=1}^{n} a_k \lambda^k + \mathcal{O}(\lambda^{n+1})$$

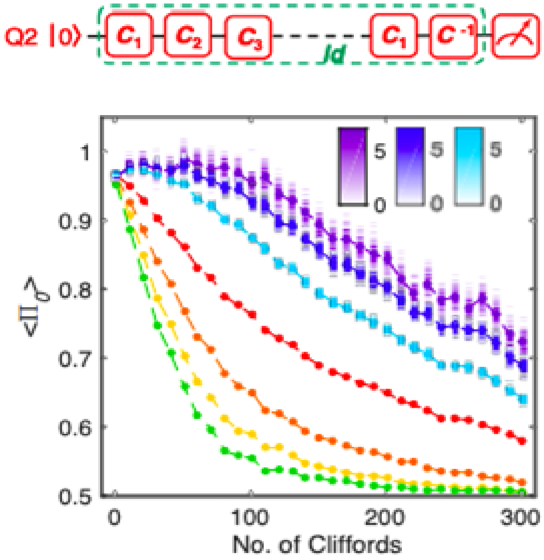
$$E_{\mathrm{Rich}}(\lambda) = E_{\mathrm{poly}}^{(d=m-1)}(\lambda) = c_0 + c_1\lambda + \dots c_{m-1}\lambda^{m-1}$$

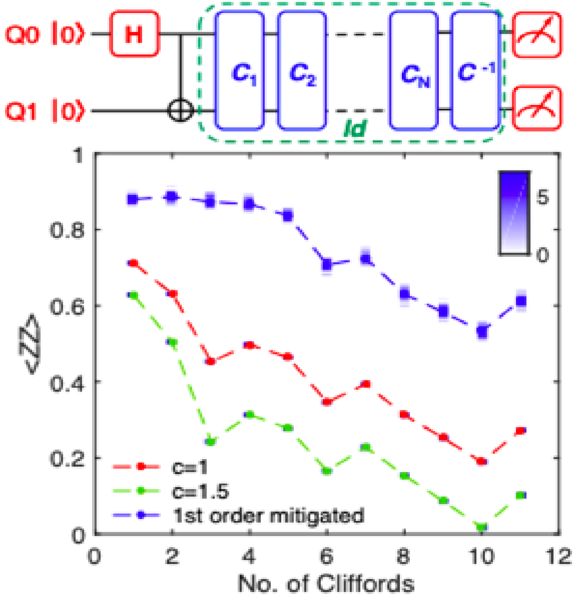


[1] Y. Li and S. C. Benjamin, "Efficient Variational Quantum Simulator Incorporating Active Error Minimization," Physical Review X, vol. 7, 6 2017.
[2] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, ´ and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," Nature, vol. 567, no. 7749, pp. 491–495, 2019.
[3] K. Temme, S. Bravyi, and J. M. Gambetta, "Error Mitigation for Short Depth Quantum Circuits," Physical Review Letters, vol. 119, p. 180509, 11 2017.

Zero-noise extrapolation

Previous benchmarks on one and two qubits





[2] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, ´ and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," Nature, vol. 567, no. 7749, pp. 491–495, 2019.
[3] K. Temme, S. Bravyi, and J. M. Gambetta, "Error Mitigation for ShortDepth Quantum Circuits," Physical Review Letters, vol. 119, p. 180509, 11 2017.

We extend and improve both noise scaling & extrapolation

Recalibration Noise-scaling		Richardson Extrapolation
Unitary Folding		**Extrapolation as inference**
Incorporates CNOT folding [4]	Incorporates random folding [4]	Incorporates Exponential Fits [1]
Parameter Noise Scaling		**Adaptive Extrapolation**

Noise scaling can be performed at the instruction set level only.

14-19X more accurately

[1] Y. Li and S. C. Benjamin, "Efficient Variational Quantum Simulator Incorporating Active Error Minimization," Physical Review X, vol. 7, 6 2017.
[2] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, ´ and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," Nature, vol. 567, no. 7749, pp. 491–495, 2019.
[3] K. Temme, S. Bravyi, and J. M. Gambetta, "Error Mitigation for Short Depth Quantum Circuits," Physical Review Letters, vol. 119, p. 180509, 11 2017.
[4] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, "Resource efficient zero noise extrapolation with identity insertions," arXiv preprint arXiv:2003.04941, 2020.
[5] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski,"Cloud quantum computing of an atomic nucleus," Physical review letters, vol. 120, no. 21, p. 210501, 2018.

We extend and improve both noise scaling & extrapolation

Recalibration Noise-scaling

Unitary Folding

Incorporates CNOT folding [4]

Incorporates random folding [4]

Parameter Noise Scaling

Noise scaling can be performed at the instruction set level only.

Richardson Extrapolation

Extrapolation as inference

Incorporates Exponential Fits [1]

Adaptive Extrapolation

14-19X more accurately

[1] Y. Li and S. C. Benjamin, "Efficient Variational Quantum Simulator Incorporating Active Error Minimization," Physical Review X, vol. 7, 6 2017.

[2] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, ´ and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," Nature, vol. 567, no. 7749, pp. 491–495, 2019.

[3] K. Temme, S. Bravyi, and J. M. Gambetta, "Error

[4] A. He, B. Nachman, W. A. de Jong, and C. W. Bau

[5] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R ... eus," Physical review letters, vol. 120, no. 21, p. 210501, 2018.

And we do larger more systematic benchmarks

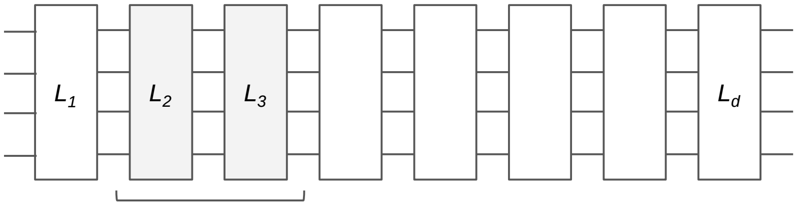
Outline

- Why error-mitigation?
- Zero-noise extrapolation (noise-scaling + extrapolation)
- Noise scaling:
    - *Unitary Folding*: a framework for digital noise scaling
    - *Parameter Noise scaling*: a noise model specific framework
- Extrapolation as inference
    - *Non-adaptive Methods*
    - *Adaptive Methods*
- Benchmarks

Unitary folding

Noise scaling at the instruction set layer

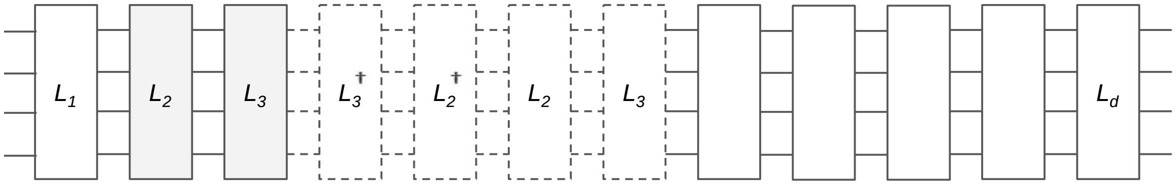
Def: a unitary fold $U \rightarrow U(U^\dagger U)^n$



Fold some subset *n* times

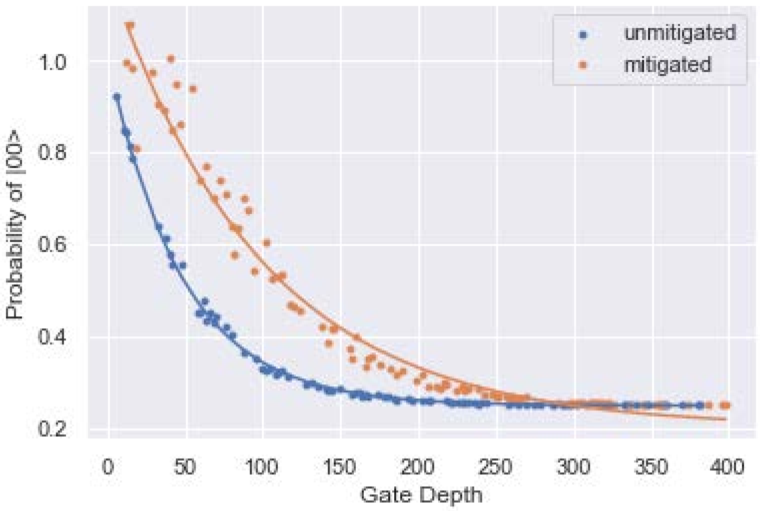
Method	Subset of indices to fold
From left	$S = \{1, 2, \ldots, s\}$
From right	$S = \{d, d-1, \ldots, d-s+1\}$
At random	$S = s$ different indices randomly sampled without replacement from $\{1, 2, \ldots, d\}$.

d scales to *d + 4*



Unitary folding performs well

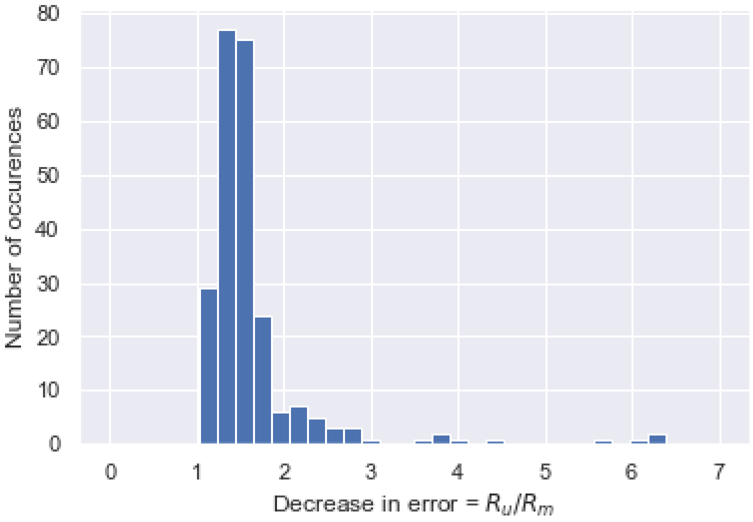
RB Circuits (2-qubits)



97.9% unmitigated to 99.0% mitigated

Exact density matrix simulations with 1% depolarizing as base noise

Random Circuits (6-qubits)



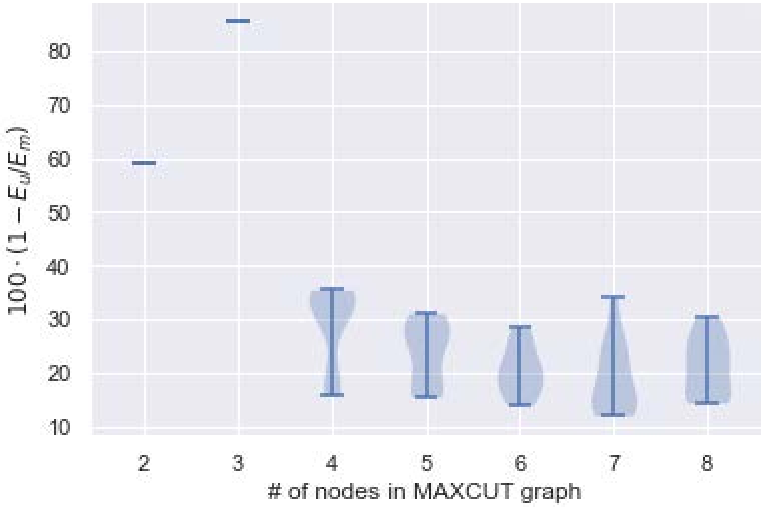
0.114 unmitigated to 0.075 mitigated avg error

250 random circuits of depth 40

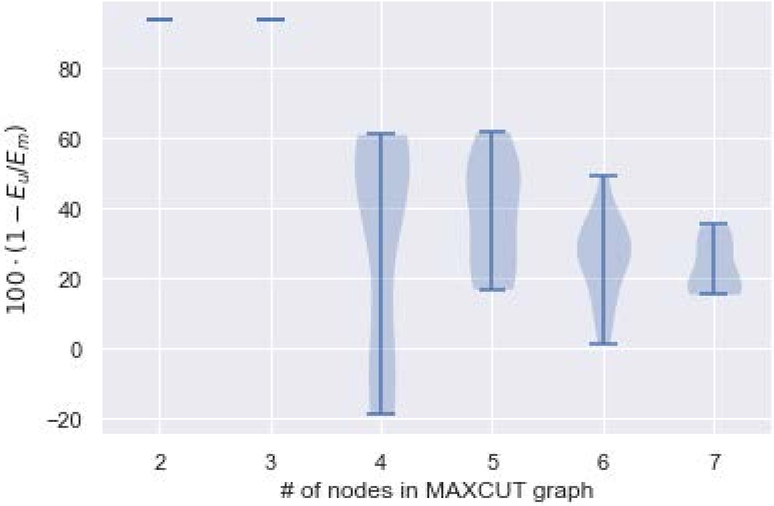
Unitary folding ZNE improve variational algorithms

Study of MAXCUT solved with QAOA

% closer to optimal with ZNE (p=1)



% closer to optimal with ZNE (p=2)



Exact density matrix simulation of 14 Erdor-Renyi random graphs at each size. Solved with Nelder-Mead optimized QAOA under 2% depolarizing base noise. Used global unitary folding and linear extrapolation.

Outline

- Why error-mitigation?
- Zero-noise extrapolation (noise-scaling + extrapolation)
- Noise scaling:
    - *Unitary Folding*: a framework for digital noise scaling
    - *Parameter Noise scaling*: a noise model specific framework
- Extrapolation as inference
    - *Non-adaptive Methods*
    - *Adaptive Methods*
- Benchmarks

Other instruction layer noise scalings are possible

For particular noise model: parameter noise

- Unitary Folding ZNE assumes that noise scales with gate # / circuit depth
- Other noise models can scale with different parameters

Parameter Noise

$$U(\vec{\theta}) \mapsto U(\vec{\theta} + \hat{\vec{\epsilon}})$$

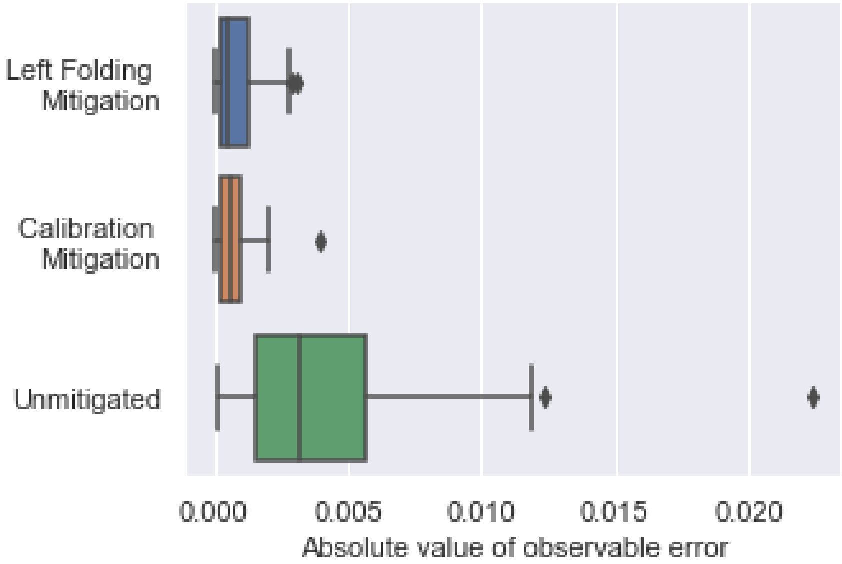
Parameter Noise Scaling

Manually apply sampled noise

$$\hat{\vec{\epsilon}} \sim \mathcal{N}(0, \sigma^2)^m \mapsto \hat{\vec{\epsilon}_\lambda} \sim \mathcal{N}(0, (\lambda\sigma)^2)^m$$

[1] J. P. Barnes, C. J. Trout, D. Lucarelli, and B. D. Clader, "Quantum error correction failure distributions: Comparison of coherent and stochastic error models," Phys. Rev. A, vol. 95, p. 062338, Jun 2017.
[2] J. True Merrill and K. R. Brown, "Progress in compensating pulse sequences for quantum computation," arXiv e-prints, p. arXiv:1203.6392, Mar. 2012
[3] Haller, and V. V. Dobrovitski, "Effect of pulse error accumulation on dynamical decoupling of the electron spins of phosphorus donors in silicon," Phys. Rev. B, vol. 85, p. 085206, Feb 2012
[4] Z.-H. Wang and V. V. Dobrovitski, "Aperiodic dynamical decoupling sequences in the presence of pulse errors," Journal of Physics B Atomic Molecular Physics, vol. 44, p. 154004, Aug. 2011.

Parameter scaled ZNE has similar performance to unitary folding



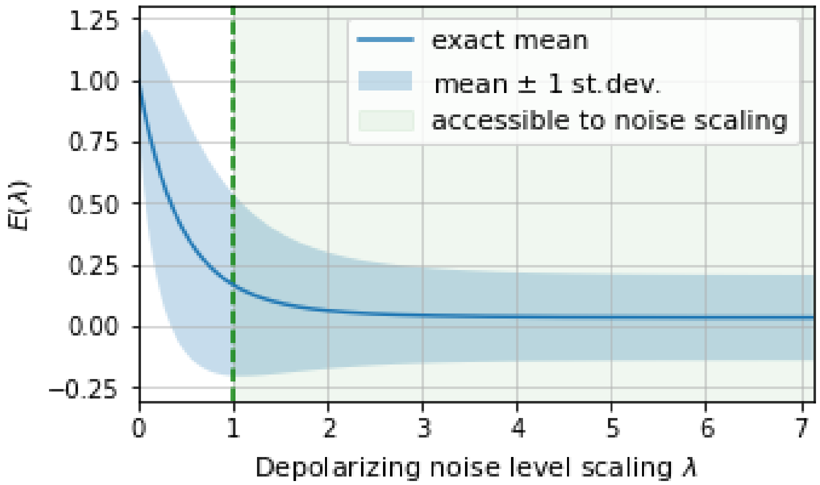
50 random six-qubit circuits. Underlying noise is an angle noise channel at σ 2 = 0.001. ZNE with linear extrapolation with noise scale factors λ = {1, 2, 3}. Results were obtained with exact density matrix simulations.

Outline

- Why error-mitigation?
- Zero-noise extrapolation (noise-scaling + extrapolation)
- Noise scaling:
    - *Unitary Folding*: a framework for digital noise scaling
    - *Parameter Noise scaling*: a noise model specific framework
- Extrapolation as inference
    - *Non-adaptive Methods*
    - *Adaptive Methods*
- Benchmarks

Extrapolation as Inference: Non-adaptive

Given the curve above 1, infer the intercept



Extrapolation as Inference: Non-adaptive

Given the curve above 1, infer the intercept

Algorithm 1: Generic non-adaptive extrapolation

Data: A set of increasing noise scale factors $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots \lambda_m\}$, with $\lambda_j \geq 1$ and fixed number of samples N for each λ_j.

Result: A mitigated expectation value

$\boldsymbol{y} \longleftarrow \emptyset$;

begin

  for $\lambda_j \in \boldsymbol{\lambda}$ **do**

    $y_j \longleftarrow \mathit{ComputeExpectation}(\lambda_j, N)$;

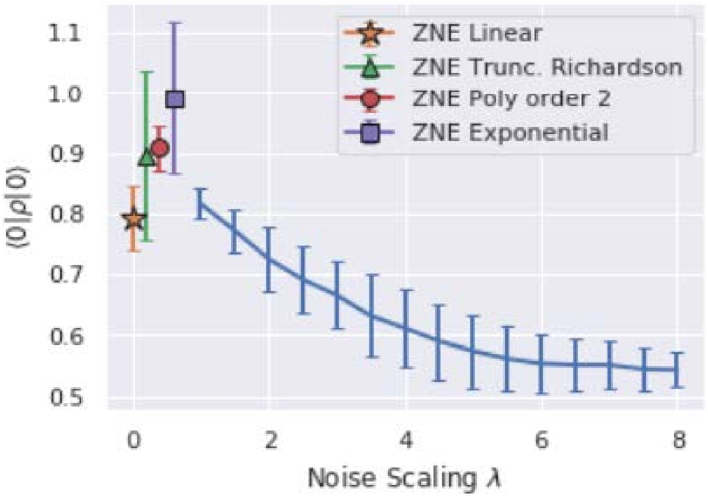
    $\mathit{Append}(\boldsymbol{y}, y_j)$;

  /* Abitrary best fit algorithm (e.g., least squares) */

  $\boldsymbol{\Gamma}^* \longleftarrow \mathit{BestFit}(E_{\text{model}}(\lambda; \boldsymbol{\Gamma}), (\boldsymbol{\lambda}, \boldsymbol{y}))$;

  return $E_{\text{model}}(0; \boldsymbol{\Gamma}^*)$;

ZNE extrapolation comparison on IBMQ Armonk qubit

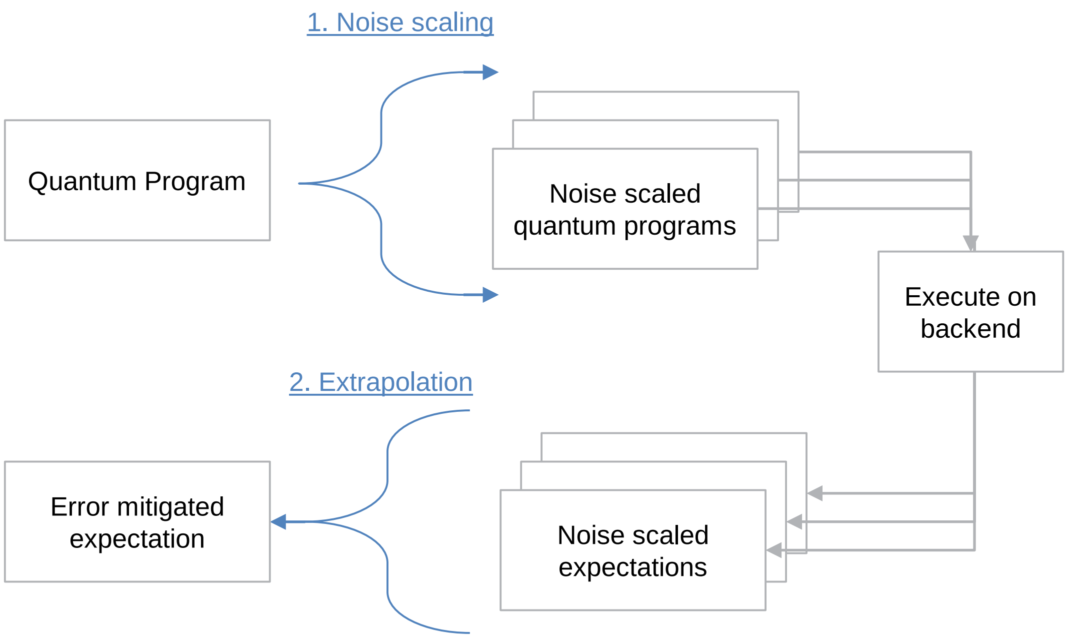


53 1-qubit RB circuits of depth 200

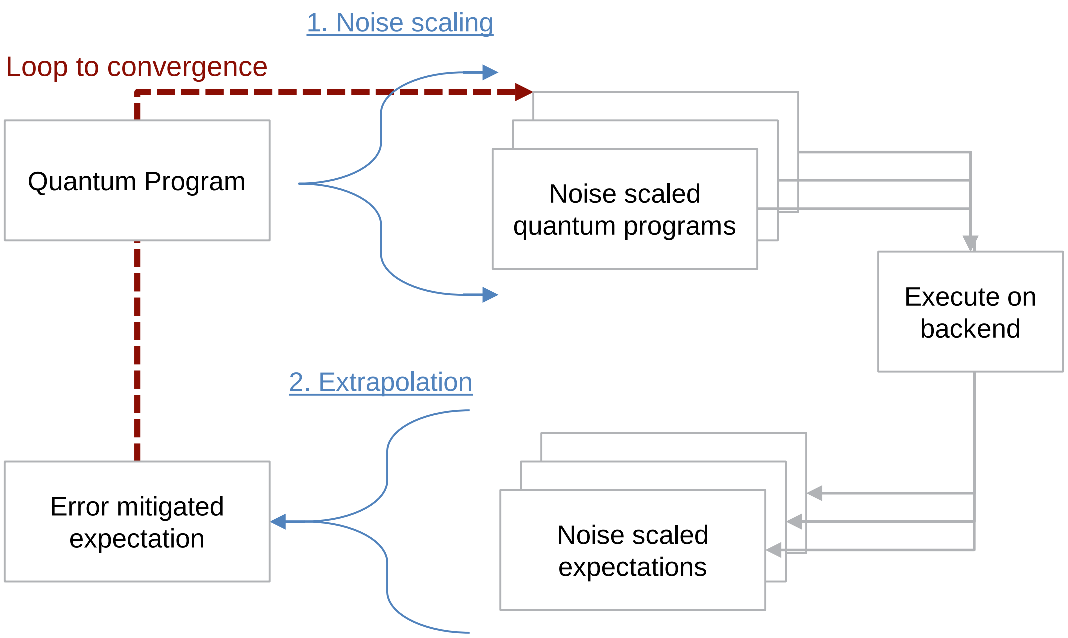
Outline

- Why error-mitigation?
- Zero-noise extrapolation (noise-scaling + extrapolation)
- Noise scaling:
    - *Unitary Folding*: a framework for digital noise scaling
    - *Parameter Noise scaling*: a noise model specific framework
- Extrapolation as inference
    - *Non-adaptive Methods*
    - *Adaptive Methods*
- Benchmarks

Adaptive Zero-noise extrapolation

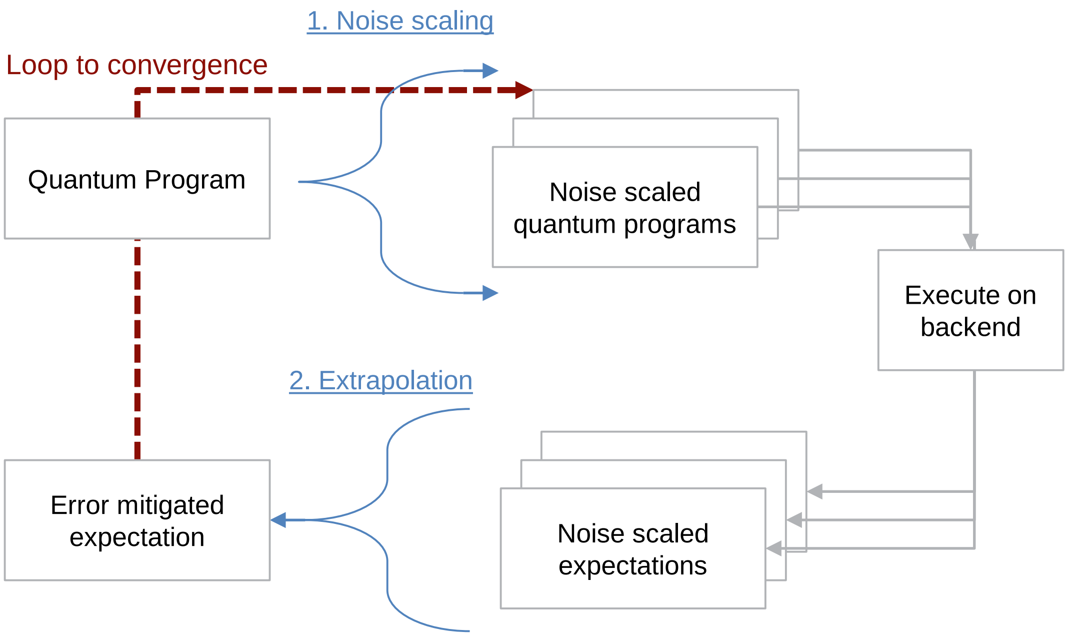


Adaptive Zero-noise extrapolation



Adaptive Zero-noise extrapolation

Optimally choose the next noise scaling (and sample #) based on data seen so far



Algorithm 2: Generic adaptive extrapolation

Data: An initial set of m noise scale factors $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots \lambda_m\}$, with $\lambda_j \geq 1$, m sample numbers $\boldsymbol{N} = (N_1, N_2, \ldots N_m)$ and a maximum number of total samples $N_{\max}$.

Result: A mitigated expectation value

begin

```
/* Initialization */
```

$\boldsymbol{y} \longleftarrow \emptyset$;

for $\lambda_j \in \boldsymbol{\lambda}$ **do**

  $y_j \longleftarrow ComputeExpectation(\lambda_j, N_j)$;

  $Append\,(\boldsymbol{y}, y_j)$;

```
/* Adaptive loop */
```

$N_{\text{used}} \longleftarrow 0$;

while $N_{used} < N_{max}$ **do**

  $\boldsymbol{\Gamma}^* \longleftarrow BestFit(E_{\text{model}}(\lambda; \boldsymbol{\Gamma}), (\boldsymbol{\lambda}, \boldsymbol{y}))$;

  $\lambda_{\text{next}} \longleftarrow NewScale(\boldsymbol{\Gamma}^*, \boldsymbol{\lambda}, \boldsymbol{y})$;

  $N_{\text{next}} \longleftarrow NewNumSamples(\boldsymbol{\Gamma}^*, \boldsymbol{\lambda}, \boldsymbol{y})$;

  $y_{\text{next}} \longleftarrow ComputeExpectation(\lambda_{\text{next}}, N_{\text{next}})$;

  $Append\,(\boldsymbol{\lambda}, \lambda_{\text{next}})$;

  $Append\,(\boldsymbol{y}, y_{\text{next}})$;

  $N_{\text{used}} \longleftarrow N_{\text{used}} + N_{\text{next}}$;

return $E_{\text{model}}(0; \boldsymbol{\Gamma}^*)$;

Optimal adaptive exponential ZNE

Optimally choose the next noise scaling (and sample #) based on data seen so far

Exponential measurement model:

$$y \,|\, \lambda \sim \mathcal{N}\left(a + b\, e^{-c\lambda}, \sigma^2\right)$$

Assumptions:

- Know minimum accessible noise level λ_1
- Know asymptotic value a

Can show that it is best to sample at:

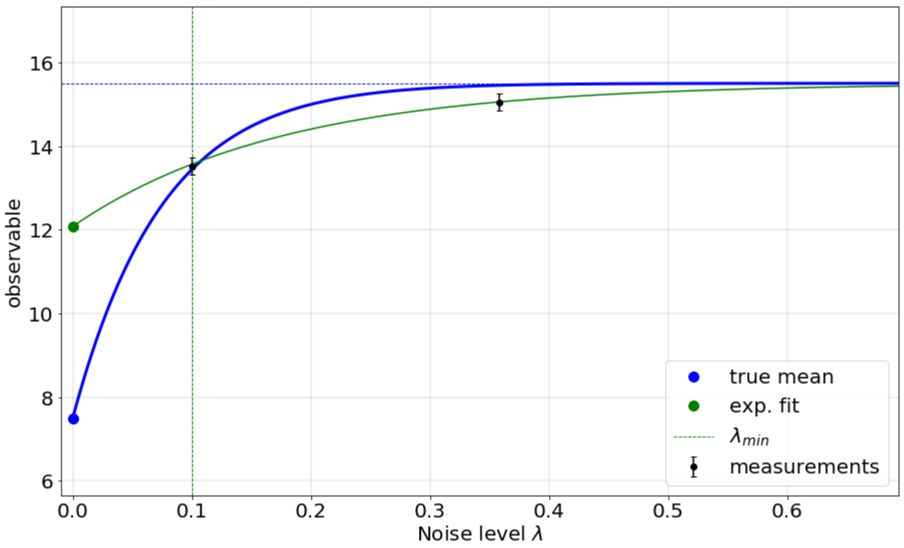
$$\lambda_1 \quad \text{and} \quad \lambda_2 = \lambda_1 + \frac{1.28}{c}$$

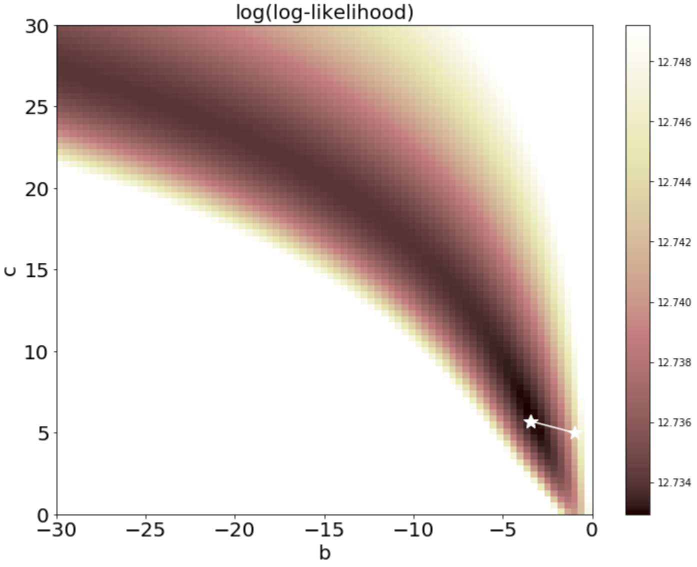
We are interested in the *intercept* $a + b$

We will do inference on b and c

Optimal adaptive exponential ZNE

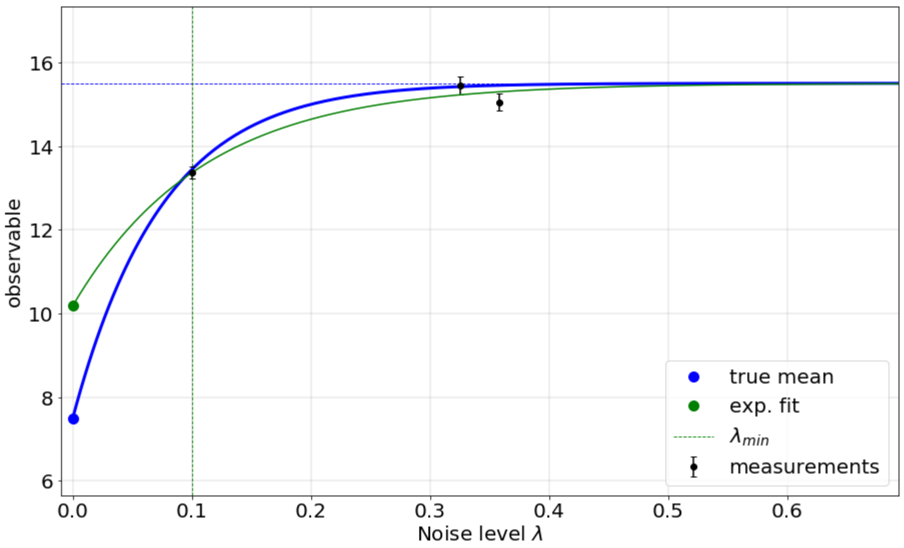
Optimally choose the next noise scaling (and sample #) based on data seen so far

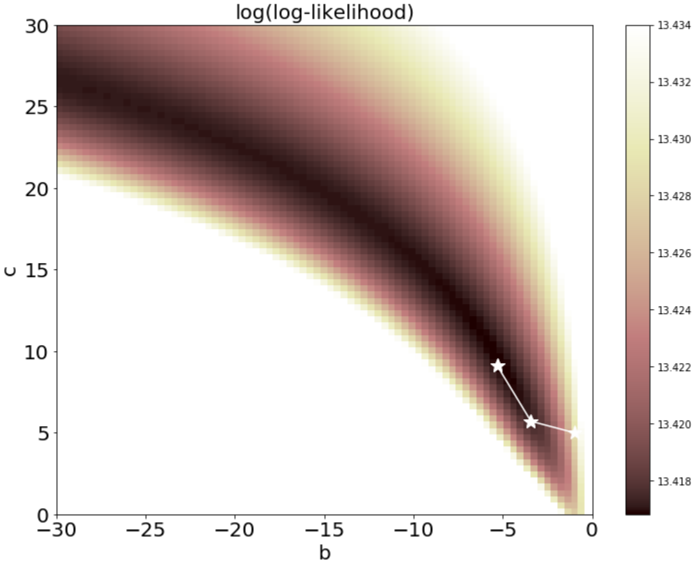




Optimal adaptive exponential ZNE

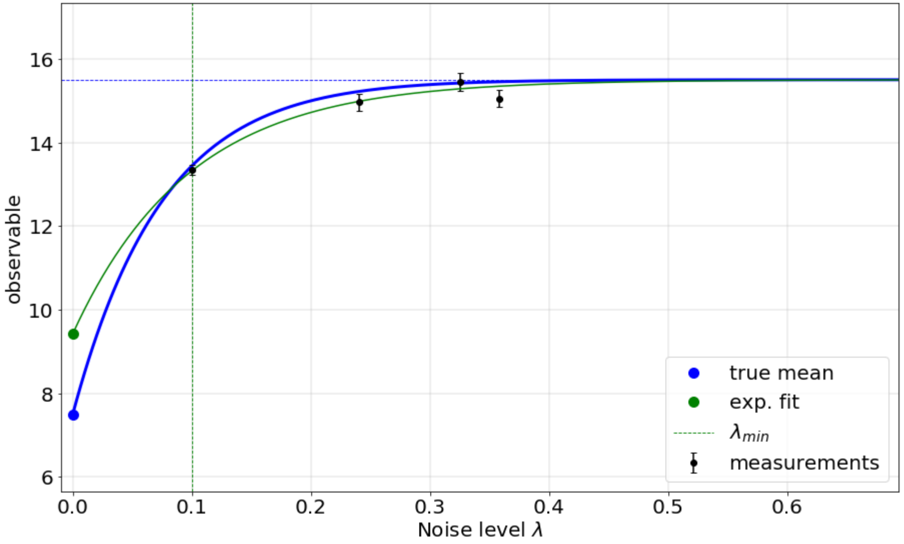
Optimally choose the next noise scaling (and sample #) based on data seen so far

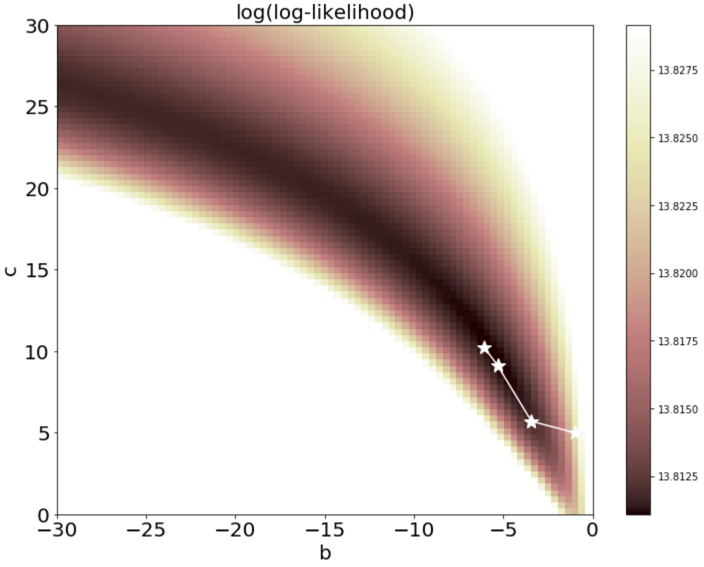




Optimal adaptive exponential ZNE

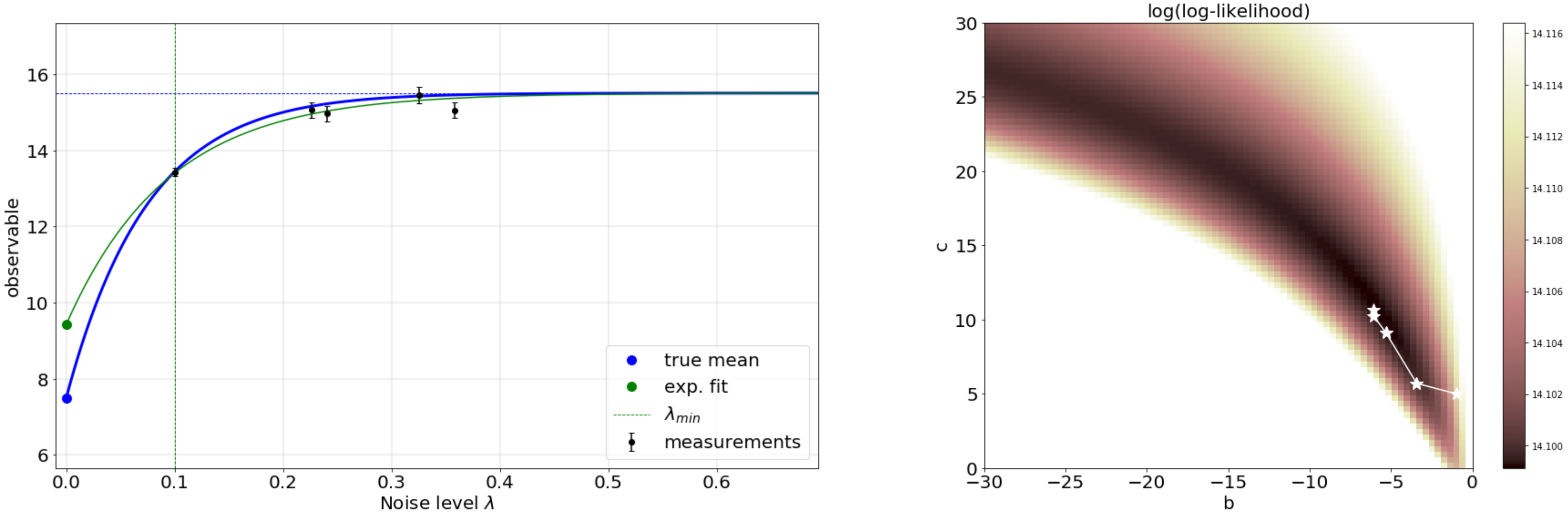
Optimally choose the next noise scaling (and sample #) based on data seen so far





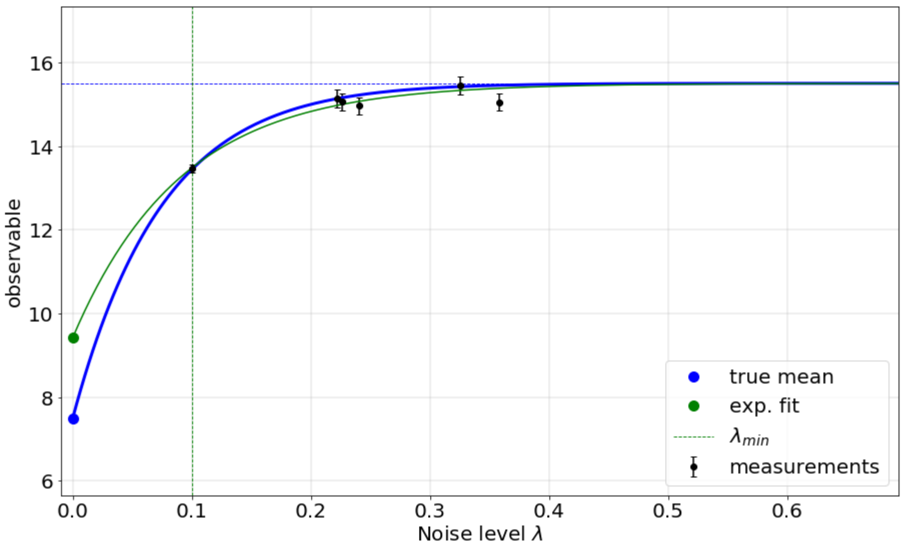
Optimal adaptive exponential ZNE

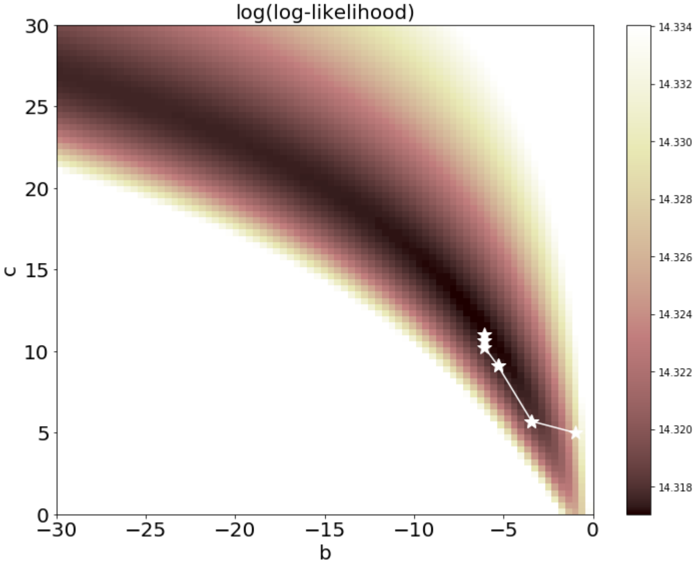
Optimally choose the next noise scaling (and sample #) based on data seen so far



Optimal adaptive exponential ZNE

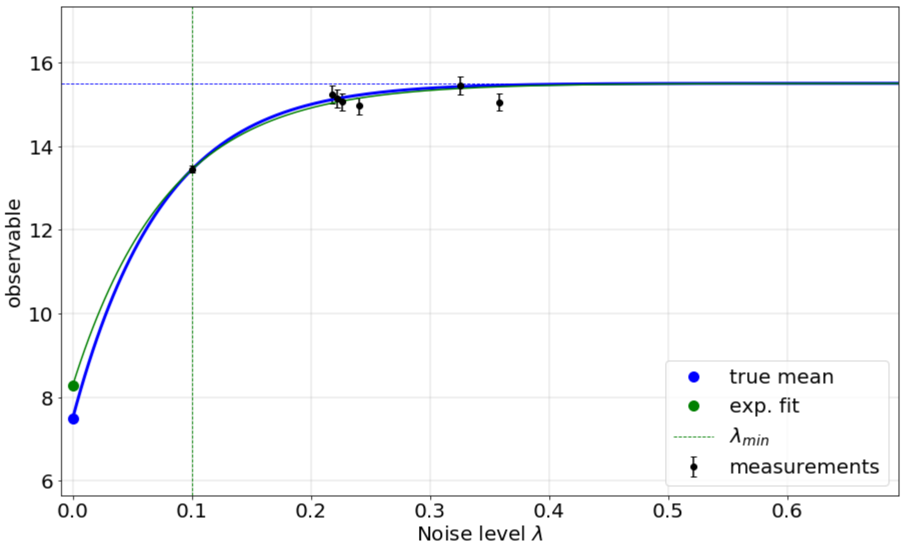
Optimally choose the next noise scaling (and sample #) based on data seen so far

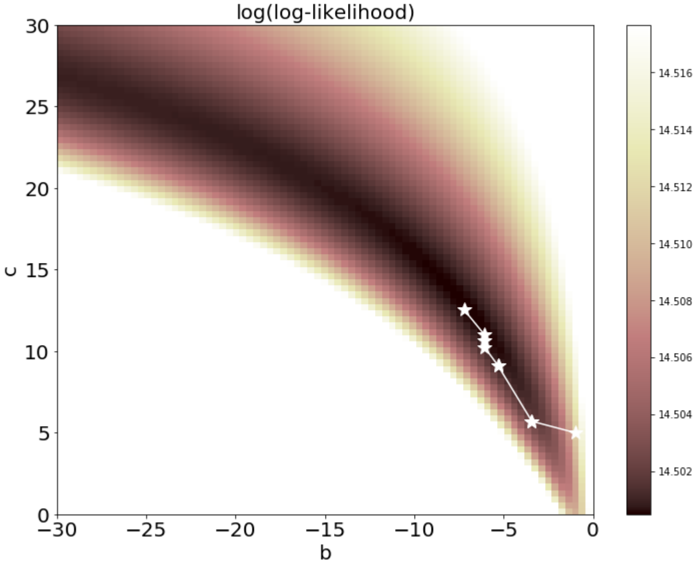




Optimal adaptive exponential ZNE

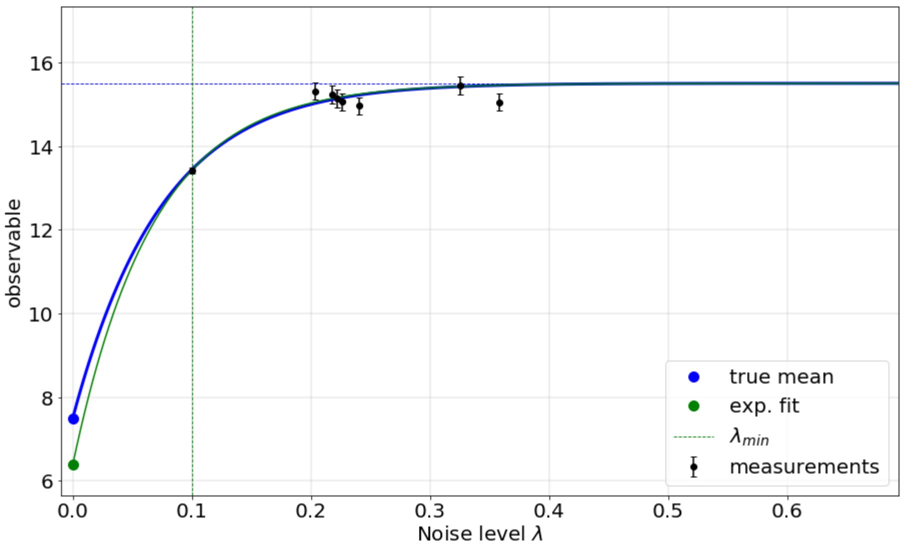
Optimally choose the next noise scaling (and sample #) based on data seen so far

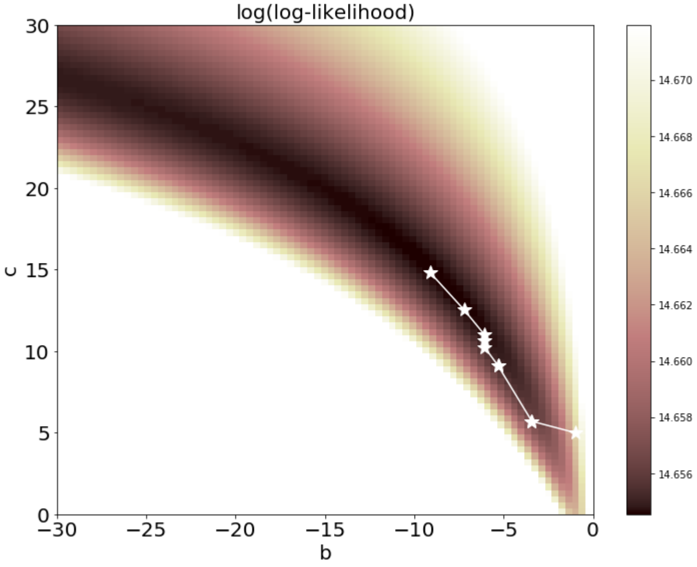




Optimal adaptive exponential ZNE

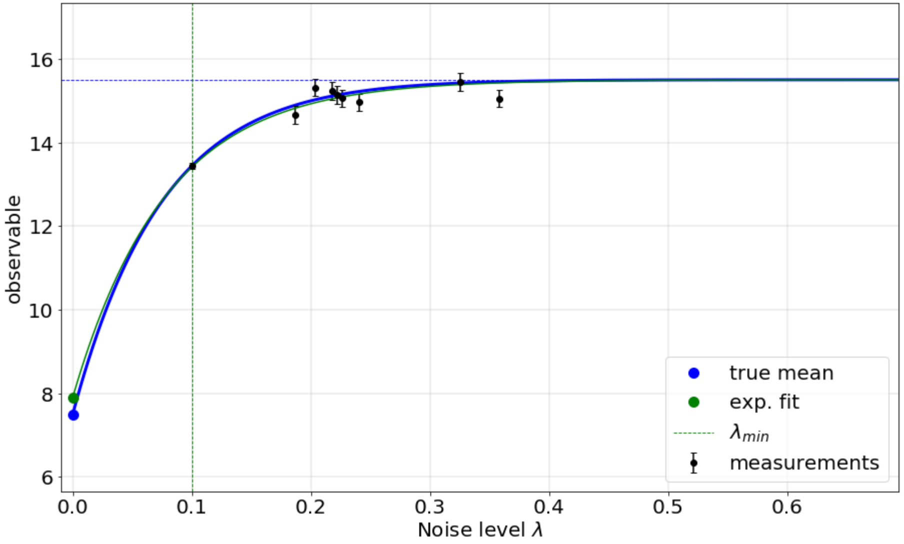
Optimally choose the next noise scaling (and sample #) based on data seen so far

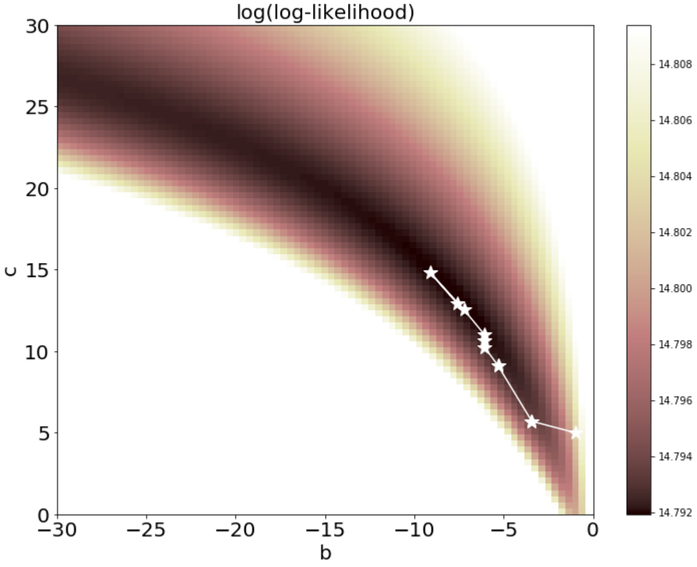




Optimal adaptive exponential ZNE

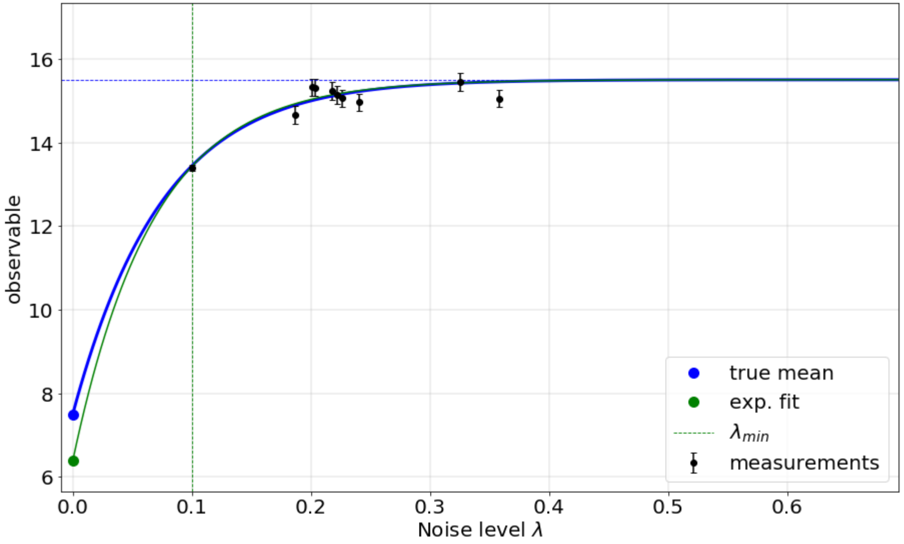
Optimally choose the next noise scaling (and sample #) based on data seen so far

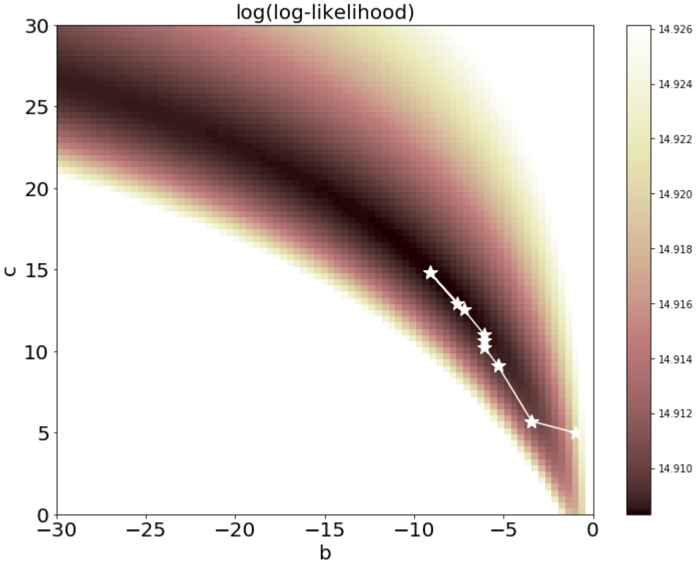




Optimal adaptive exponential ZNE

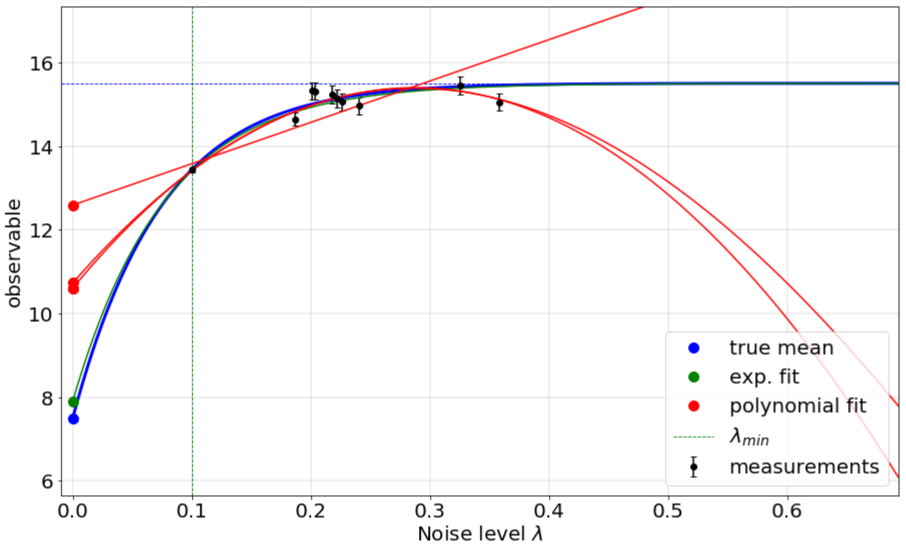
Optimally choose the next noise scaling (and sample #) based on data seen so far

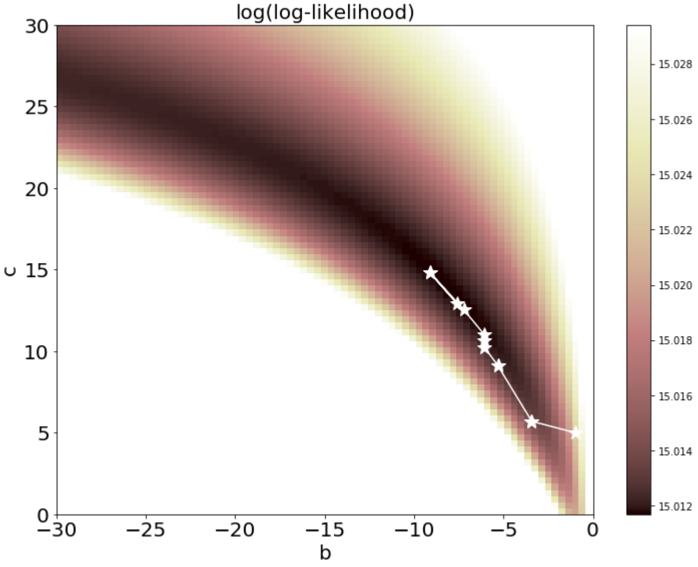




Optimal adaptive exponential ZNE

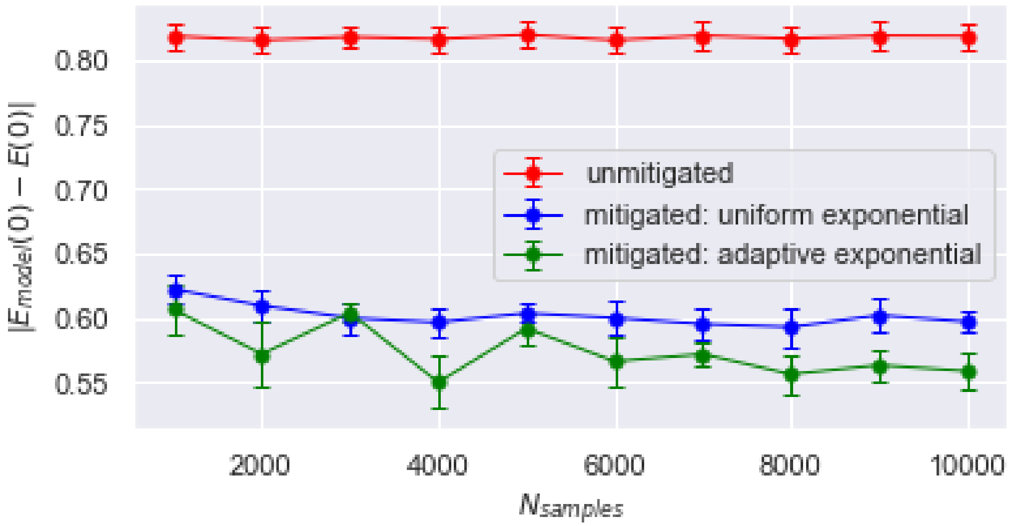
Optimally choose the next noise scaling (and sample #) based on data seen so far





Adaptive ZNE can give a 5X speedup

Error by number of total samples taken (proportional to runtime)



5 qubit RB circuits of depth 10 under 5% simulated depolarizing noise

Outline

- Why error-mitigation?
- Zero-noise extrapolation (noise-scaling + extrapolation)
- Noise scaling:
    - *Unitary Folding*: a framework for digital noise scaling
    - *Parameter Noise scaling*: a noise model specific framework
- Extrapolation as inference
    - *Non-adaptive Methods*
    - *Adaptive Methods*
- Benchmarks

Exponential extrapolation performs best

Scaling	Extrapolation	Error % (dep.)	Error % (amp. damp.)
none	unmitigated	29.9 ± 5.1	16.7 ± 4.0
circuit	linear $(d = 1)$	14.6 ± 4.6	5.40 ± 2.3
circuit	quadratic $(d = 2)$	6.35 ± 3.6	3.53 ± 3.4
circuit	Richardson $(d = 3)$	17.6 ± 11	17.9 ± 16
circuit	exponential $(a = 0.25)$	2.73 ± 1.9	2.06 ± 1.6
circuit	adapt. exp. $(a = 0.25)$	$\mathbf{1.27 \pm 1.1}$	2.69 ± 2.8
at random	linear $(d = 1)$	15.6 ± 5.3	5.20 ± 2.4
at random	quadratic $(d = 2)$	5.54 ± 4.4	8.00 ± 8.1
at random	Richardson $(d = 3)$	30.0 ± 24	24.0 ± 18
at random	exponential $(a = 0.25)$	2.84 ± 1.8	$\mathbf{0.95 \pm 1.0}$
at random	adapt. exp. $(a = 0.25)$	1.77 ± 1.4	2.18 ± 1.2
from left	linear $(d = 1)$	14.4 ± 4.5	5.16 ± 2.3
from left	quadratic $(d = 2)$	6.73 ± 3.7	3.88 ± 3.7
from left	Richardson $(d = 3)$	18.4 ± 12	16.1 ± 13
from left	exponential $(a = 0.25)$	3.17 ± 2.1	2.19 ± 2.0
from left	adapt. exp. $(a = 0.25)$	1.43 ± 1.1	3.08 ± 3.6

Average of 20 different two-qubit randomized benchmarking circuits with mean depth 27. 1% Depolarizing noise. Amplitude damping channel with $\gamma = 0.01$. For all non-adaptive methods we used $\lambda = \{1, 1.5, 2, 2.5\}$. Adaptive extrapolation was iterated up to 4 scale factors.

We can now do zero-noise extrapolation:

with only gate level access

and

14-19X more accurately

Upcoming: ***mitiq***