LA-UR-21-27372

Machine learning regression and error quantification for lattice QCD Monte Carlo observables

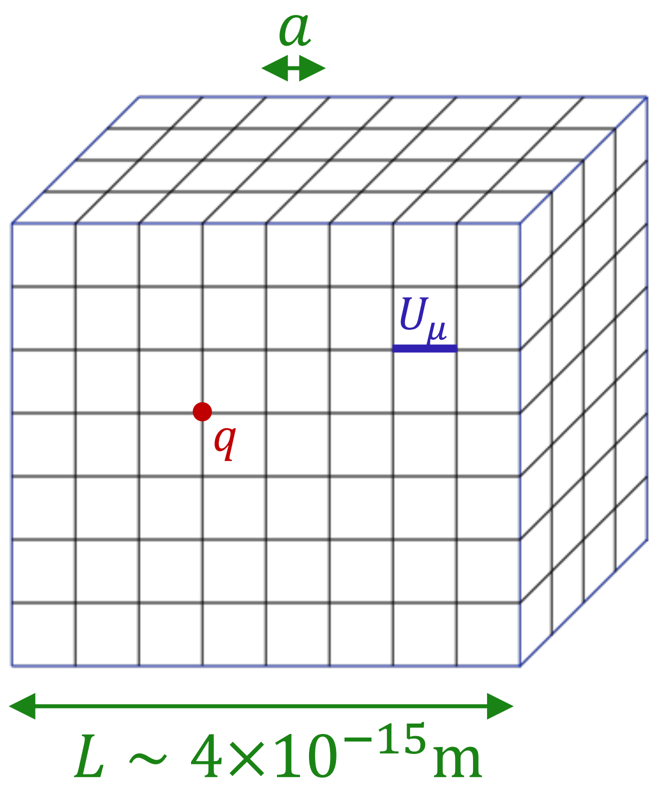
Boram Yoon, Tanmoy Bhattacharya, Rajan Gupta

(Los Alamos National Laboratory)

A Virtual Tribute to Quark Confinement and the Hadron Spectrum 2021, Aug 2-6 July 2021, online

Lattice QCD

- Non-perturbative approach to solving QCD on discretized Euclidean space-time
  - Hypercubic lattice
  - Lattice spacing a
  - Quark fields placed on sites
  - Gluon fields on the links between sites; U_μ
- Numerical lattice QCD calculations using Monte Carlo methods
  - Computationally intensive
  - Use supercomputers
- Continuum results are obtained in $a \to 0$
- Has been successful for many QCD observables
  - Some results are with less than 1% error



Lattice QCD

- Correlation functions

$$\langle O \rangle = Z^{-1} \int dU dq d\bar{q}\, O(U, q, \bar{q}) e^{-S_g - \bar{q}(D+m_q)q}$$

$$= Z^{-1} \int dU \left[O\left(U, (D + m_q)^{-1}\right) e^{-S_g} \det(D + m_q) \right]$$

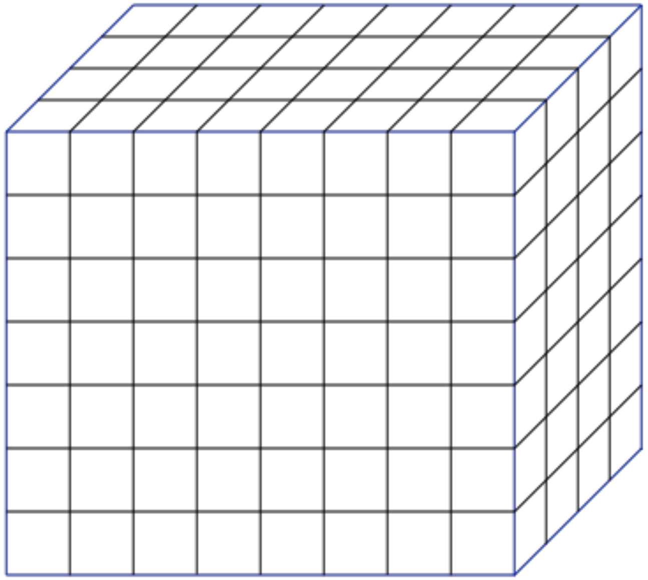
- Monte-Carlo integration
  - Integration variable U is huge

$$N_s^3 \times N_t \times 4 \times 8 \sim 10^9$$

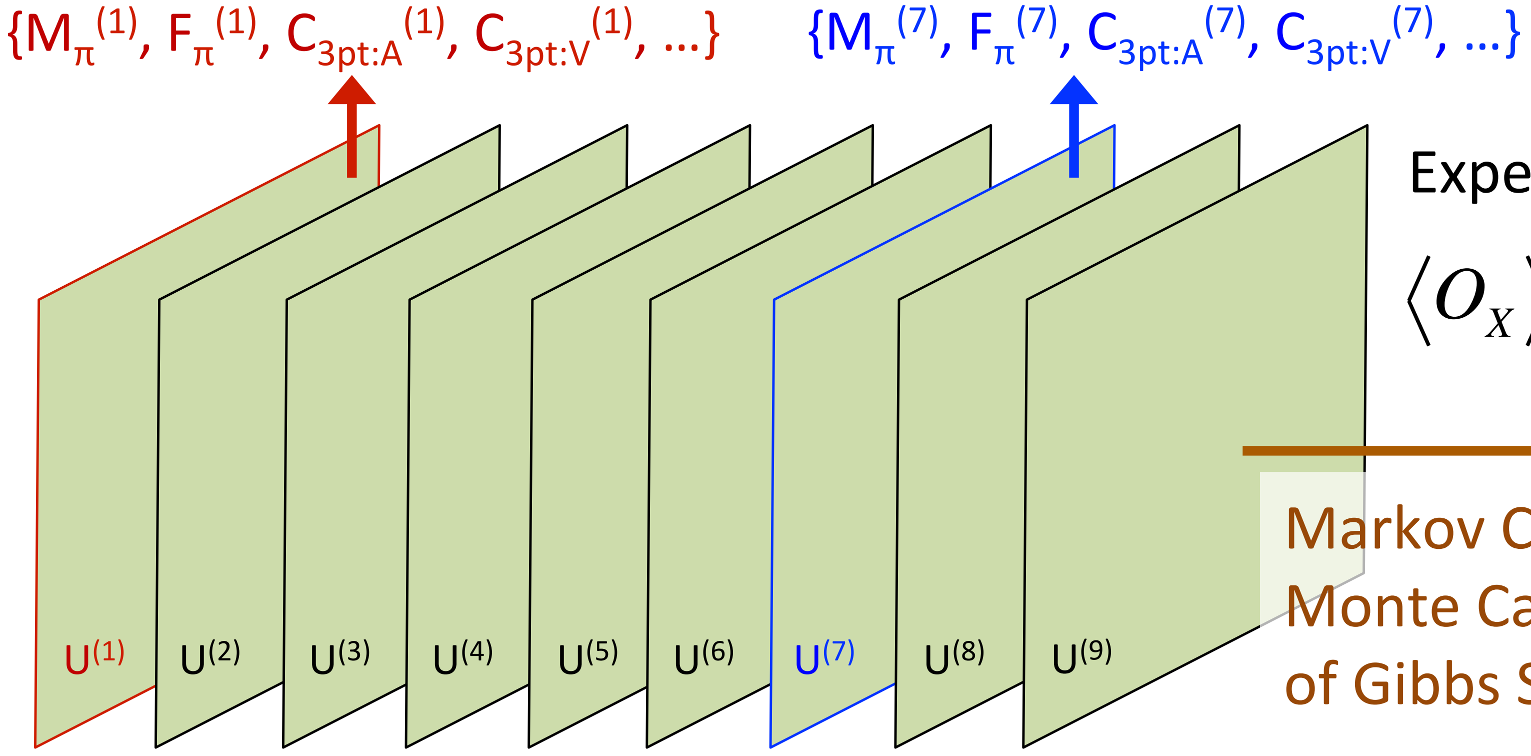
  - Generate Markov chain of gauge configurations U
  - Calculate average as expectation value

$$\langle O \rangle \approx \frac{1}{N} \sum_i O_i\left(U, (D + m_q)^{-1}\right)$$

  - Calculation of $O_i\left(U, (D + m_q)^{-1}\right)$: measurement
  - $(D + m)^{-1}$ is computationally expensive



Lattice QCD Observables are Correlated

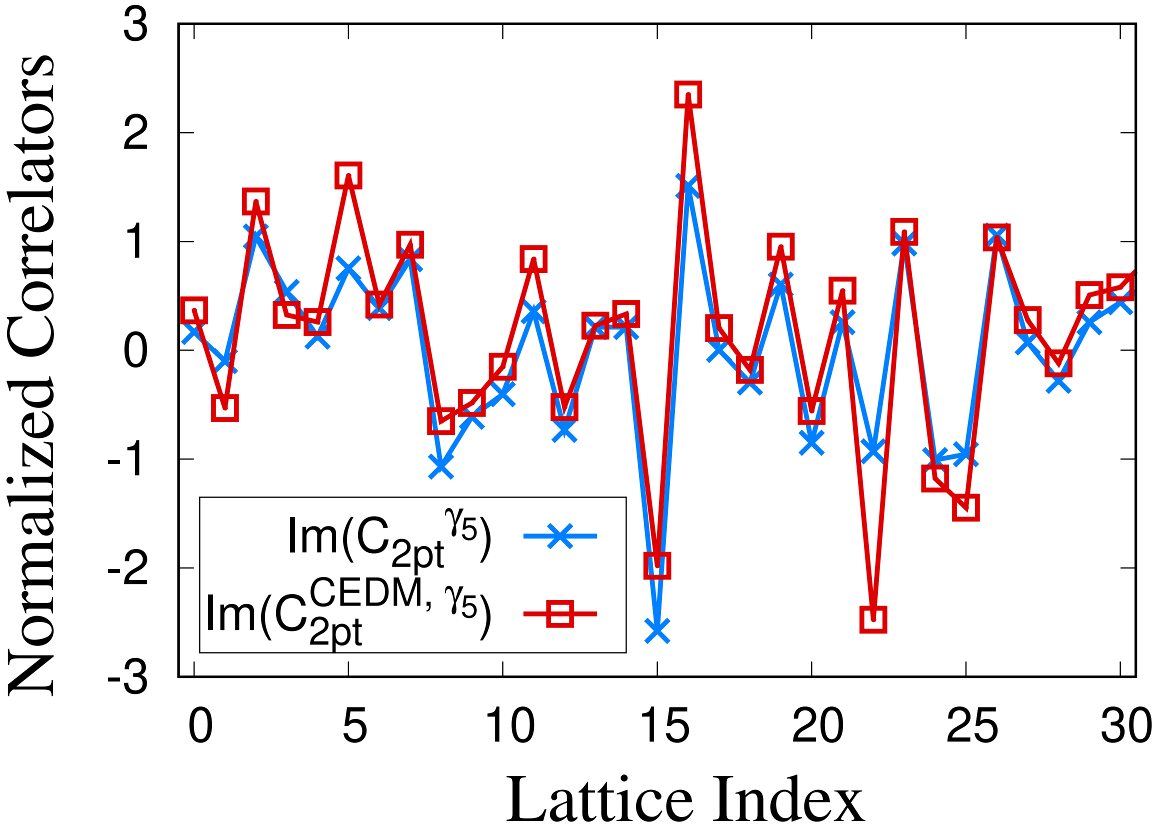


Expectation value

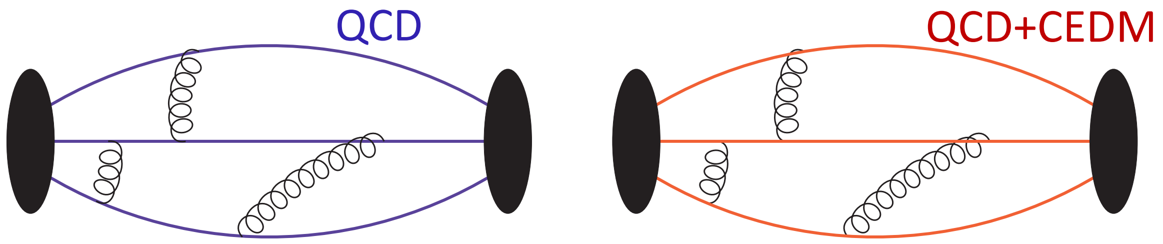
$$\langle O_X \rangle \approx \frac{1}{N}\sum_{n=1}^{N} O_X^{(n)}$$

Markov Chain
Monte Carlo Trajectory
of Gibbs Samples

Correlation Map of Nucleon Observables



- Correlation between proton(uud) 2-pt correlation function and that calculated in presence of CEDM interaction

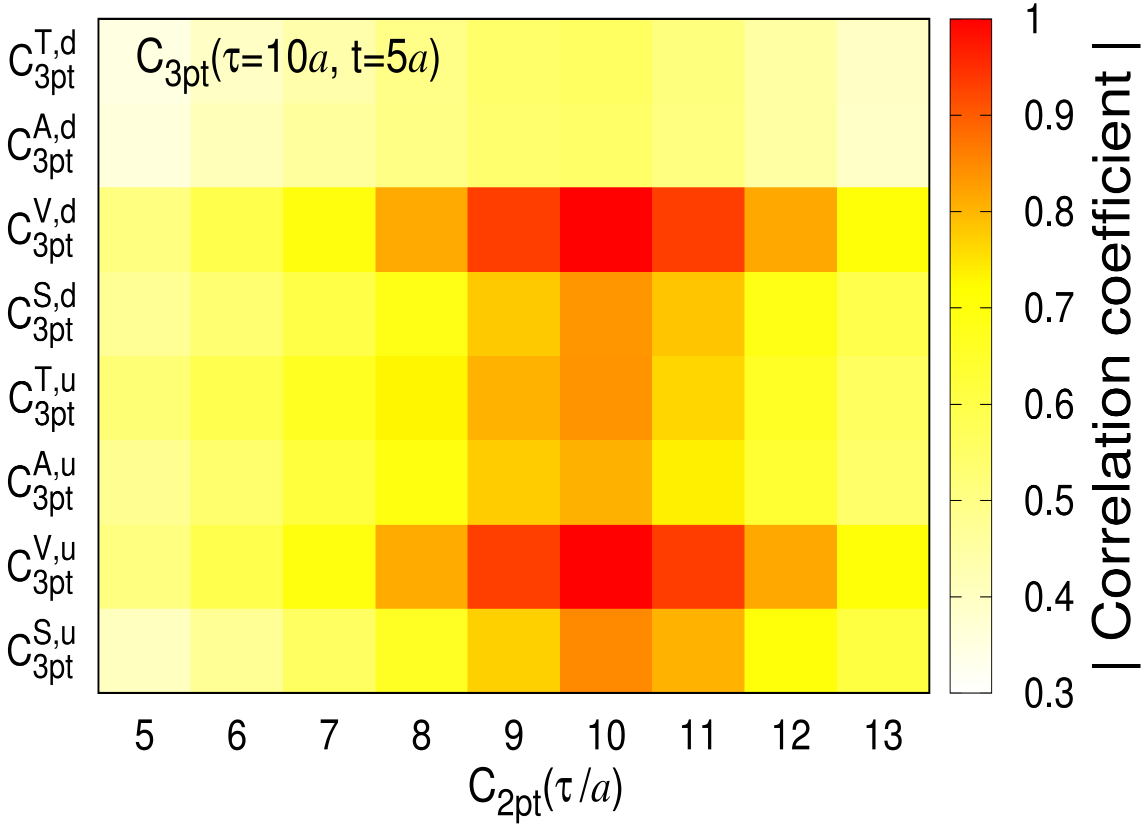


$C_{2pt} \sim \langle N(\tau)N^\dagger(0)\rangle$ $\quad C_{2pt}^{CEDM} \sim \langle N(\tau)N^\dagger(0)\rangle_{CEDM}$

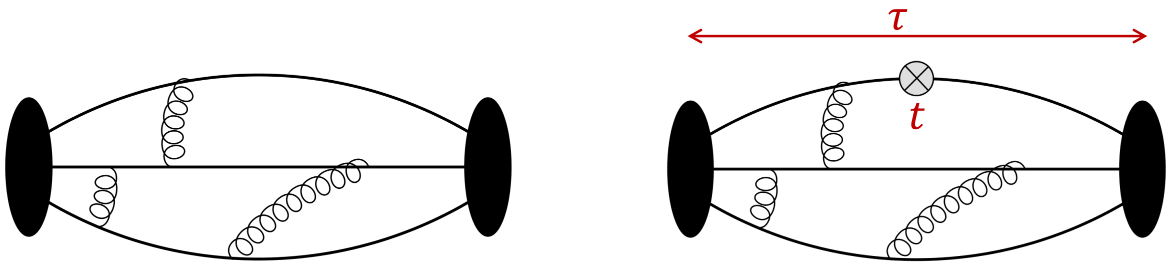
- QCD: D_{clov}

  QCD+CEDM: $D_{clov} + \frac{i}{2}\varepsilon\sigma^{\mu\nu}\gamma_5 G_{\mu\nu}$

Correlation Map of Nucleon Observables



- Correlation between proton(uud) 3-pt and 2-pt correlation functions

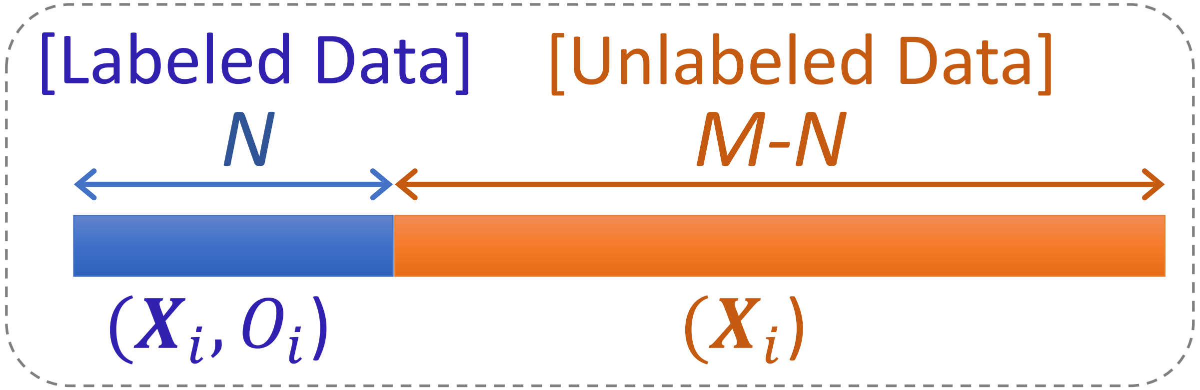


$C_{2pt} \sim \langle N(\tau)N^{\dagger}(0)\rangle$ $\quad$ $C_{3pt}^{A,S,T,V} \sim \langle N(\tau)O(t)N^{\dagger}(0)\rangle$

- Using these correlations, C_{3pt} can be estimated from C_{2pt} on each configuration

Prediction of Lattice QCD Observables using ML

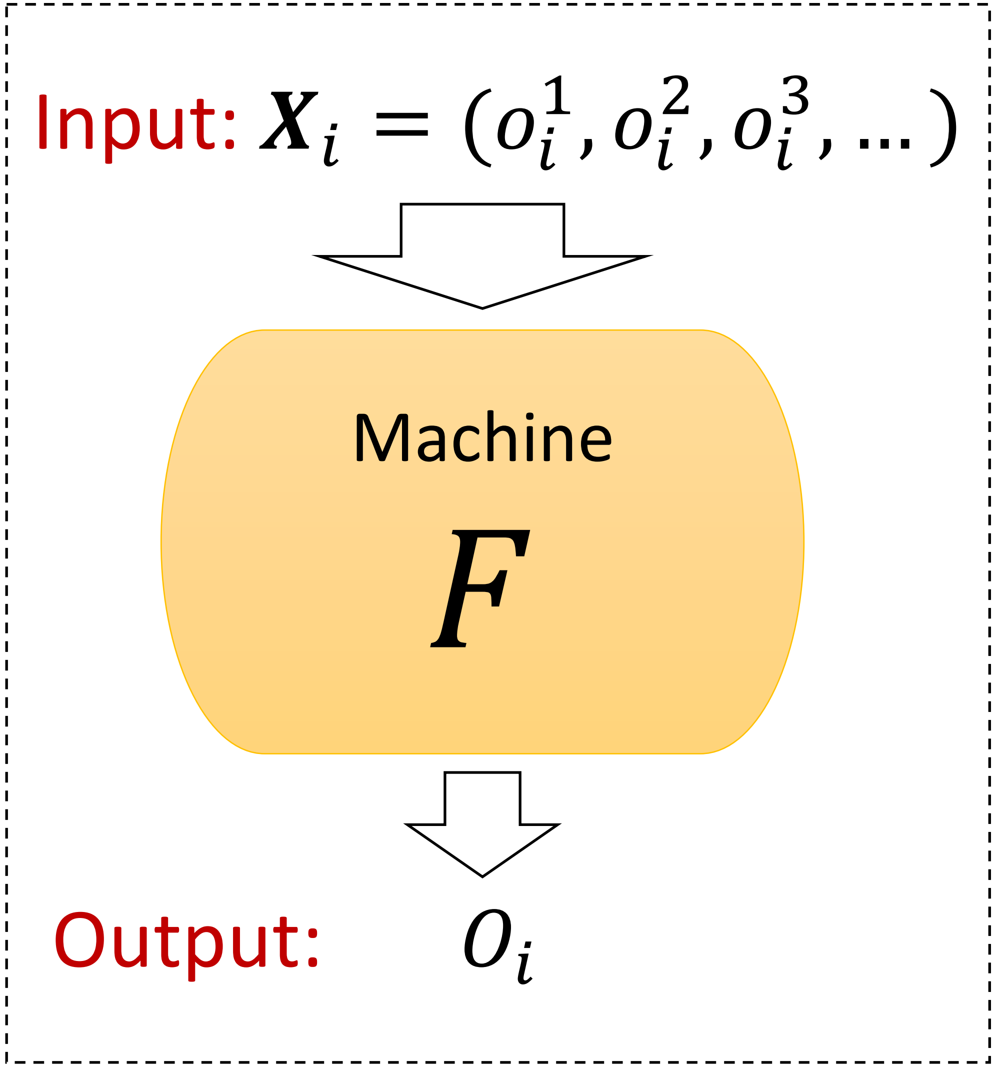
- Assume M indep. measurements
- Common observables $\boldsymbol{X}_i$ on all M
  Target observable O_i on first N



1) **Train machine *F*** to yield O_i from $\boldsymbol{X}_i$ on the Labeled Data

2) **Predict** O_i of the Unlabeled data from $\boldsymbol{X}_i$

$$\boldsymbol{F}(\boldsymbol{X}_i) = O_i^P \approx O_i$$



Prediction Bias

- $\boldsymbol{F}(\boldsymbol{X}_i) = O_i^P \approx O_i$
- Simple average

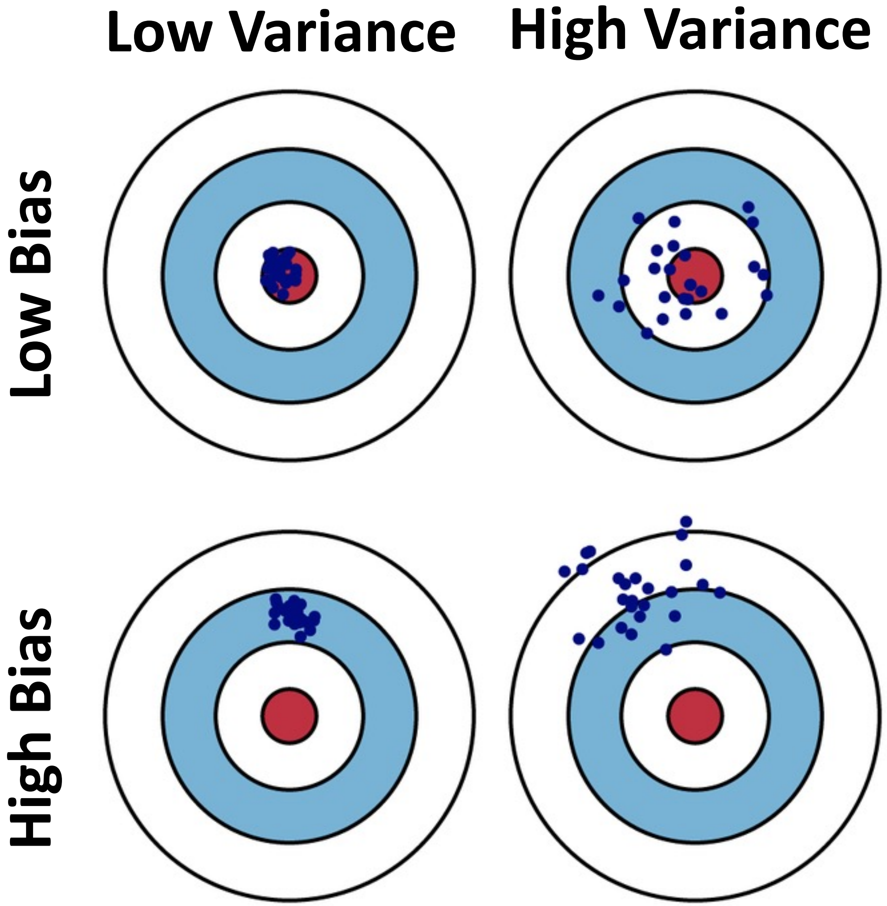
$$\bar{O} = \frac{1}{M} \sum_{i \in \text{Unlabeled}} O_i^P$$

is not correct due to prediction bias

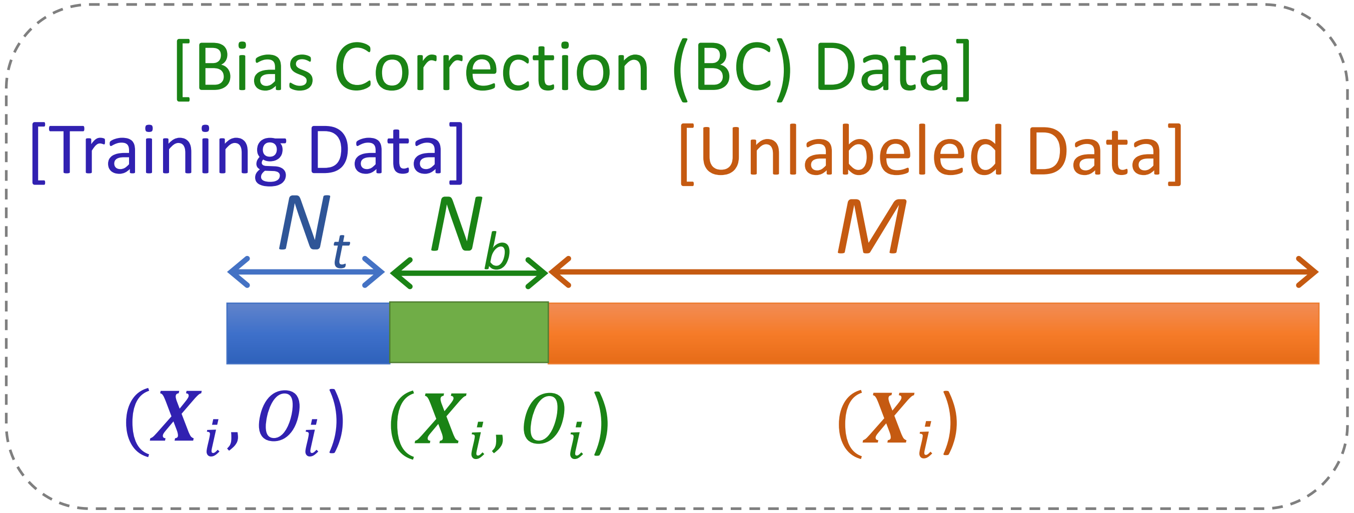
- Prediction = TrueAnswer + Noise + Bias
- ML prediction may have bias

$$\langle O^P \rangle \neq \langle O \rangle$$

Bias $= \langle O^P \rangle - \langle O \rangle$



Bias Correction and Error Quantification



- Split labeled data $N = N_t + N_b$
- Average of predictions on test data with bias correction

$$\boxed{\bar{O}_{BC} = \frac{1}{M} \sum_{i \in \text{Unlabeled}} O_i^P + \frac{1}{N_b} \sum_{i \in BC} (O_i - O_i^P)}$$

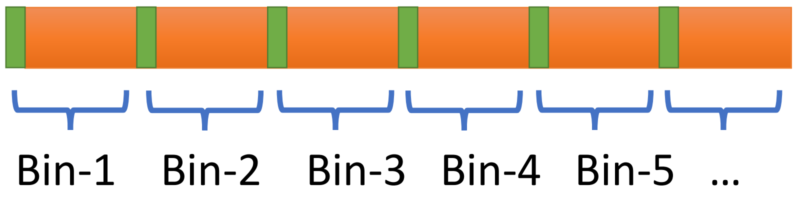
- Expectation value, $\langle \bar{O}_{BC} \rangle = \langle O^P \rangle + \langle O - O^P \rangle = \langle O \rangle$
- BC term converts systematic error of prediction to statistical uncertainty

Incorporating Labeled Data

- Include directly measured values O_i from labeled data

$$\bar{O}_{BC}^{\mathrm{imp}} = w_1 \times \left(\frac{1}{N} \sum_{i \in \mathrm{Labeled}} O_i \right) + w_2 \times \left(\frac{1}{M} \sum_{i \in \mathrm{Unlabeled}} O_i^P + \frac{1}{N_b} \sum_{i \in BC} \left(O_i - O_i^P \right) \right)$$

- w_1, w_2: weights determined based on the (co)variance of two terms
- If you need more than just a simple average in data analysis
  - two different data, O_i on labeled and O_i^P on unlabeled samples
  - simultaneous fit on these two data sets with the same fit parameters
  - O_i and O_i^P have the same mean after BC but may have different variance
- Statistical errors can be estimated using Bootstrap resampling
- Binning and BC for each bin is another option for complicated data analysis



Quality of Prediction

- Bias-corrected average

$$\bar{O}_{BC} = \frac{1}{M}\sum_{i\in\text{Unlabeled}} O_i^P + \frac{1}{N_b}\sum_{i\in BC}(O_i - O_i^P)$$

- Statistical error of the unbiased average

$$\sigma^2_{\bar{O}_{BC}} \approx \frac{1}{M}\sigma^2_{O^P} + \frac{1}{N_{bc}}\sigma^2_{O-O^P}$$
$$\approx \frac{\sigma_O^2}{M}\left(1 + \frac{M}{N_{bc}}\frac{\sigma^2_{O-O^P}}{\sigma^2_O}\right) \equiv \frac{\sigma_O^2}{M}\left(1 + \frac{M}{N_{bc}}Q^2\right); \qquad Q \equiv \frac{\sigma_{O-O^P}}{\sigma_O}$$

approximations ($\approx$) for small correlation between the two terms and a good prediction algorithm that gives $\sigma_O^2 \approx \sigma^2_{O^P}$

$$\frac{\sigma^2_{\bar{O}_{BC}}}{\sigma^2_{\bar{O}}} \approx 1 + \frac{M}{N_{bc}}Q^2; \qquad \frac{\sigma_{\bar{O}_{BC}}}{\sigma_{\bar{O}}} \approx 1 + \frac{M}{2N_{bc}}Q^2$$

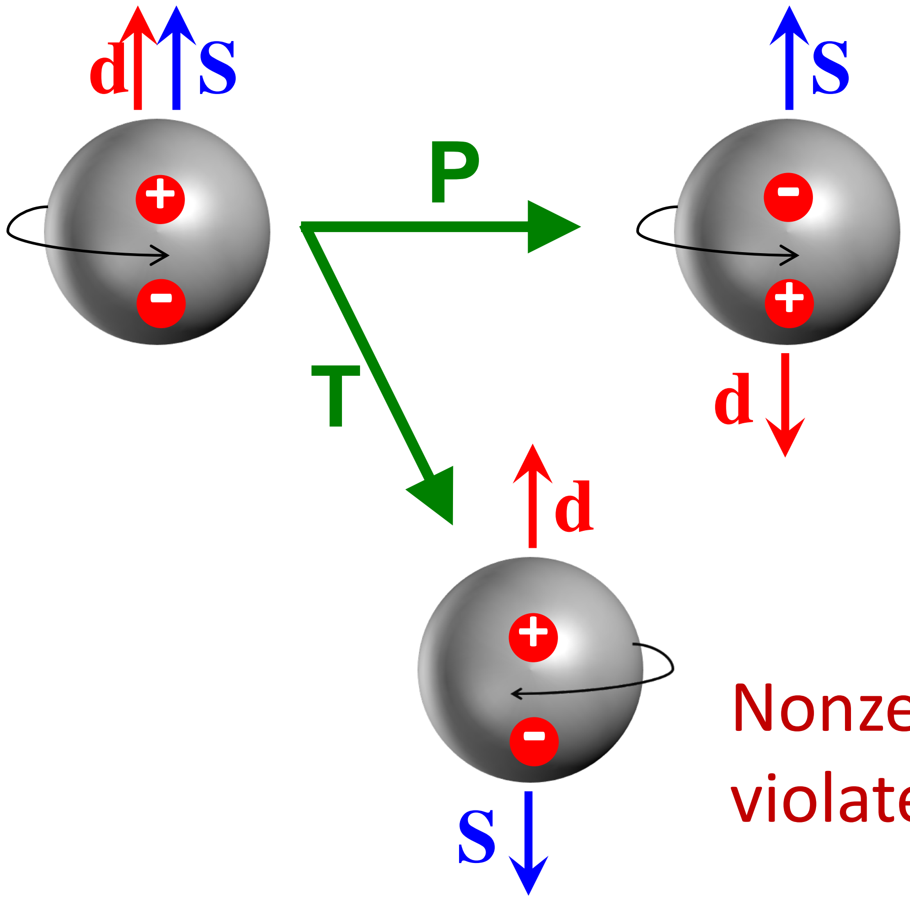
for $N_{bc}/M = 0.2$

Q	Error Increase
0.5	62.5%
0.3	22.5%
0.1	2.5%

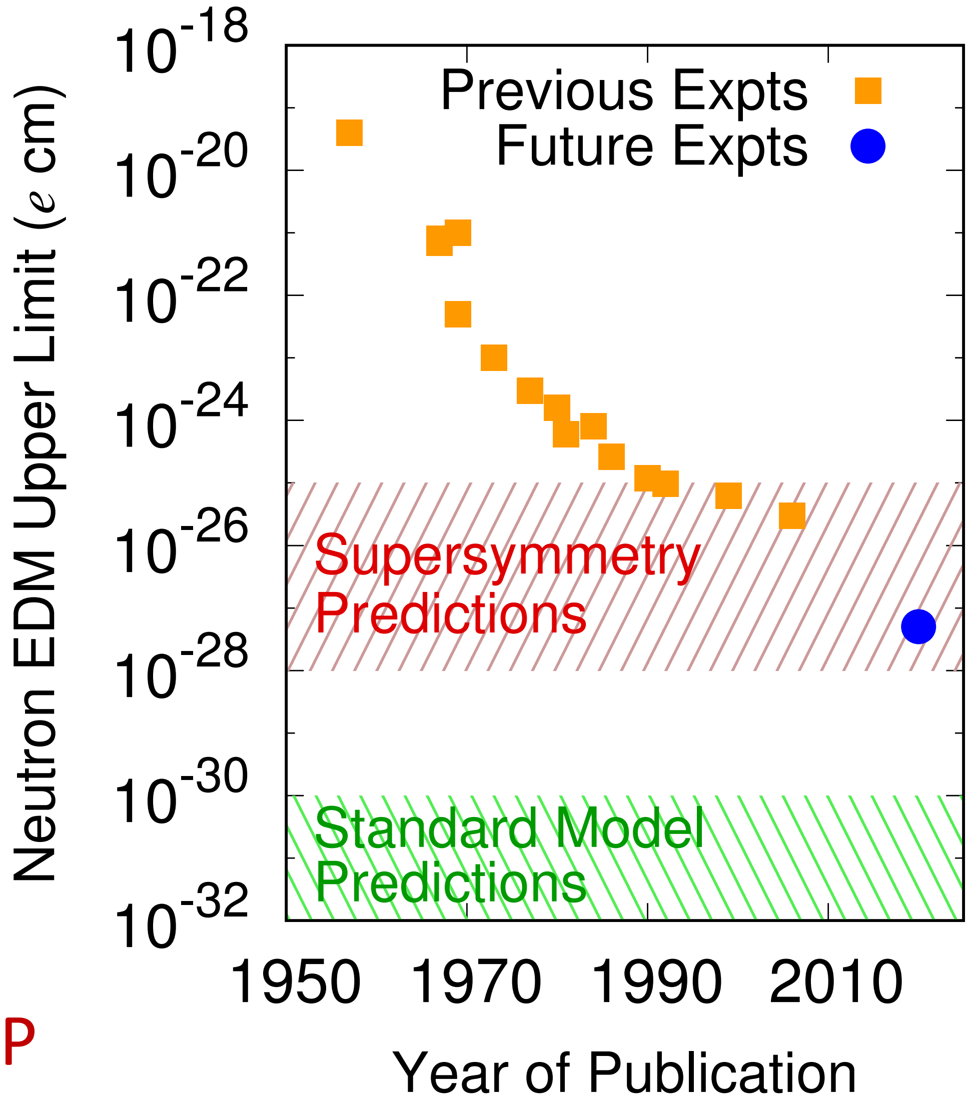
- Q-value shows the expected error-increase due to the ML prediction error
- In practice, BC data have less autocorrelation than full data, because of the many measurements per configuration, so $\sigma_{\bar{O}_{BC}}$ gives smaller error than expected above

Neutron EDM and CP Violation

- Measures separation between centers of (+) and (-) charges



Nonzero nEDM violates P and T, so CP



Effective CPV Lagrangian

$$\mathcal{L}_{\mathrm{CPV}}^{d\leq 6} = -\frac{g_s^2}{32\pi^2}\bar{\theta} G\tilde{G}$$ dim=4 QCD θ-term

$$-\frac{i}{2}\sum_{q=u,d,s} d_q \bar{q}(\sigma\cdot F)\gamma_5 q$$ dim=5 Quark EDM (qEDM)

$$-\frac{i}{2}\sum_{q=u,d,s} \tilde{d}_q g_s \bar{q}(\sigma\cdot G)\gamma_5 q$$ dim=5 Quark Chromo EDM (CEDM)

$$+d_w \frac{g_s}{6} G\tilde{G}G$$ dim=6 Weinberg 3g operator

$$+\sum_i C_i^{(4q)} O_i^{(4q)}$$ dim=6 Four-quark operators

Quark Chromo EDM (cEDM)

- Simulation in presence of CPV cEDM interaction

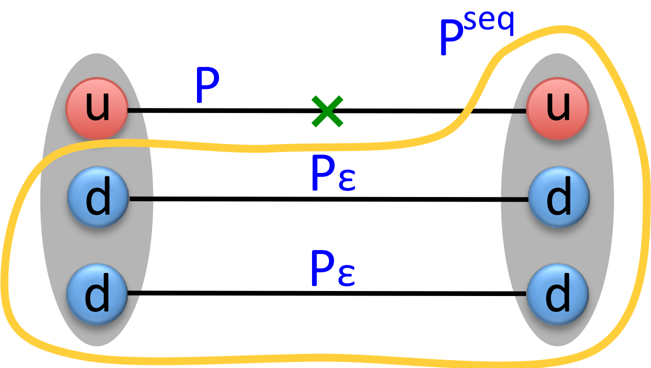
$$S = S_{QCD} + S_{cEDM}$$

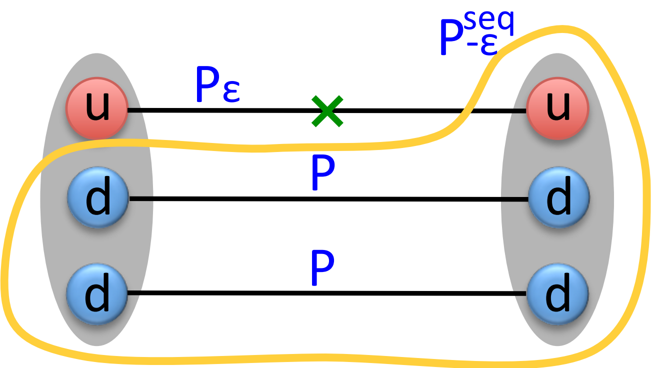
$$S_{cEDM} = -\frac{i}{2}\int d^4x \ \tilde{d}_q g_s \bar{q}(\sigma\cdot G)\gamma_5 q$$

- Schwinger source method
  Include cEDM term in valence quark propagators by modifying Dirac operator

$$\boxed{D_{\text{clov}} \to D_{\text{clov}} + i\varepsilon\sigma^{\mu\nu}\gamma_5 G_{\mu\nu}}$$

- cEDM contribution to nEDM can be obtained by calculating vector form-factor F_3 with propagators including cEDM & $O_{\gamma_5} = \bar{q}\gamma_5 q$





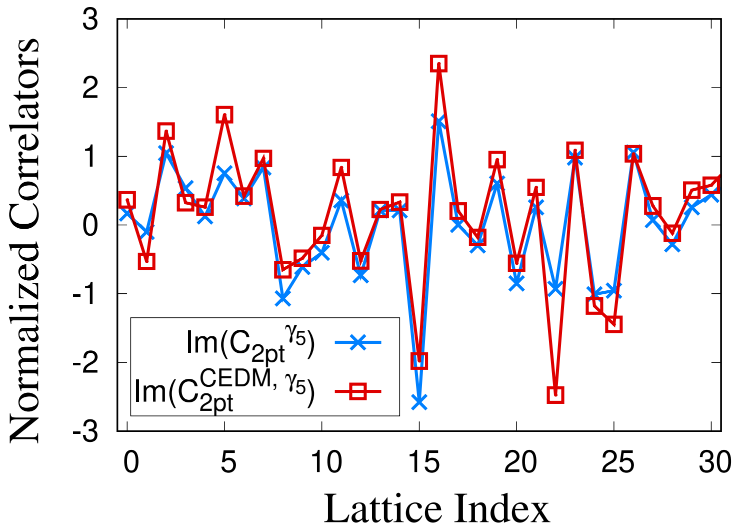
Prediction of $C_{2pt}^{\mathbf{CPV}}$ from C_{2pt}

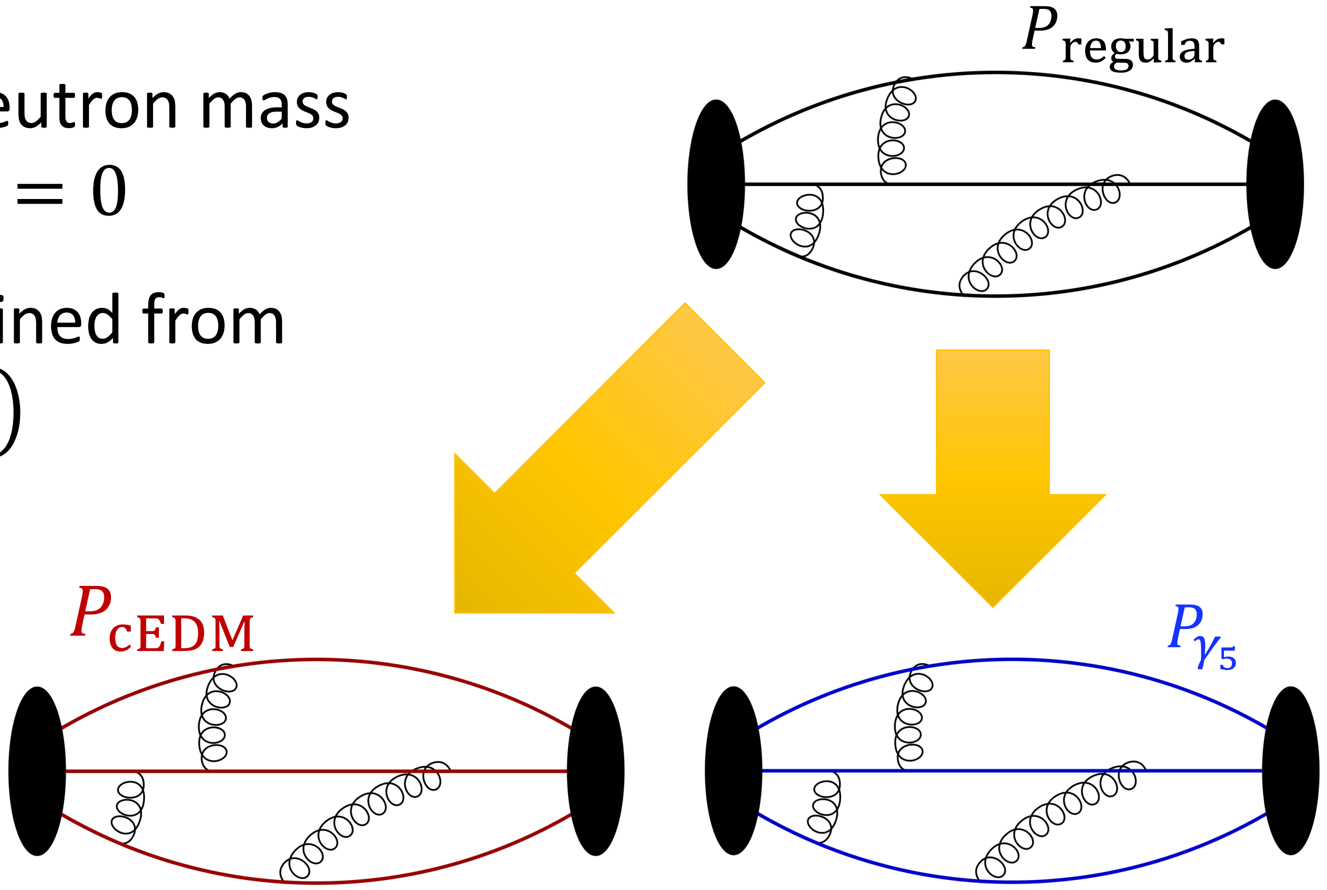
- Predict C_{2pt} for cEDM and γ_5 insertions from C_{2pt} without CPV
- CPV interactions ➔ phase in neutron mass

$$\left(ip_\mu\gamma_\mu + me^{-2i\alpha\gamma_5}\right)u_N = 0$$

- At leading order, α can be obtained from

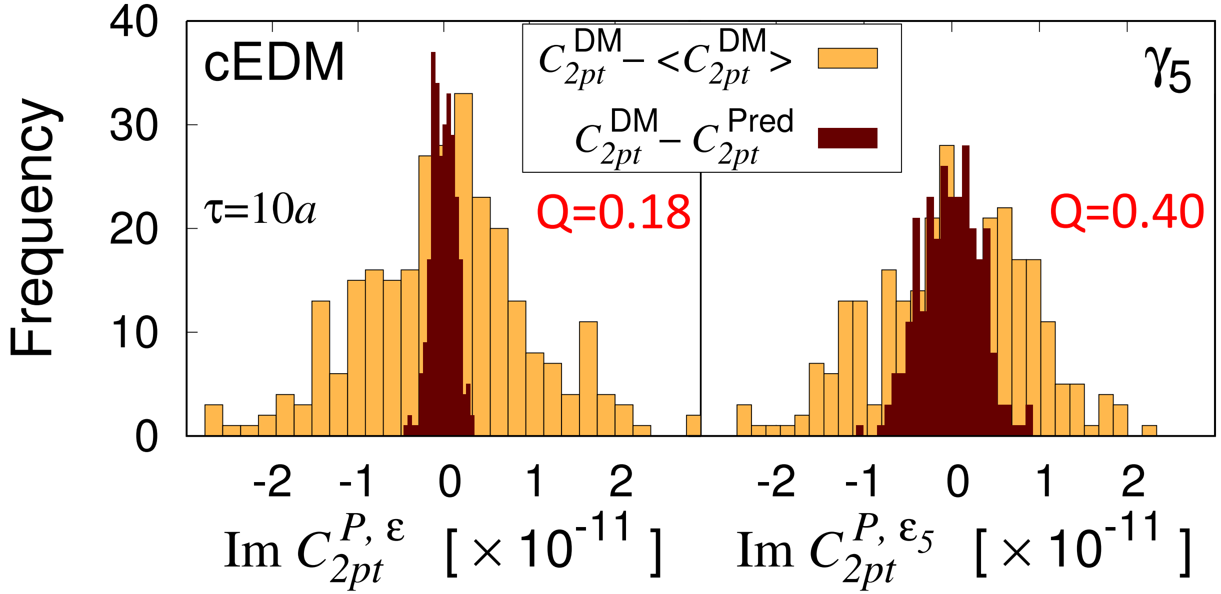
$$C_{2pt}^{P} \equiv \mathrm{Tr}\left(\gamma_5\langle NN^\dagger\rangle\right)$$





Prediction of C_{2pt}^{CPV} from C_{2pt}

- Training and Test performed for
  - $a = 0.12$ fm, $M_\pi = 305$ MeV
  - Measurements: 400 confs × 64 srcs
- \# of training data: 70 confs
  \# of BC data: 50 confs
  \# of unlabeled data: 280 confs



Input:

$$\boldsymbol{X}_i = \{\mathrm{Re}, \mathrm{Im}[C_{2pt}^{S,P}(0 \le \tau/a \le 16)]\}$$

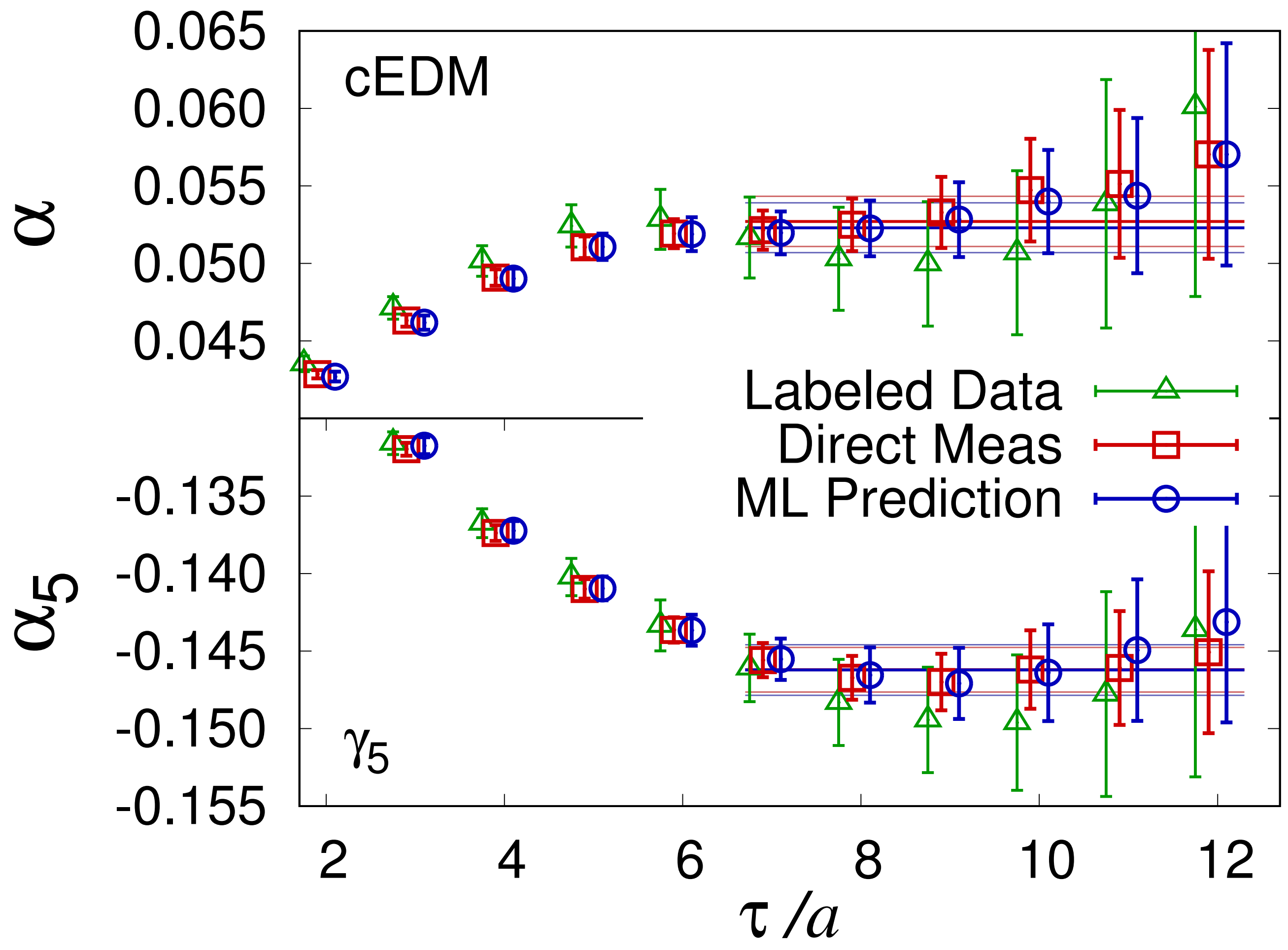
⇩

Boosted Decision Tree Regression

⇩

Output: $\mathrm{Im}\left[C_{2pt}^{P\,(\mathrm{cEDM},\gamma_5)}(\tau)\right]$

Prediction of C_{2pt}^{CPV} from C_{2pt}



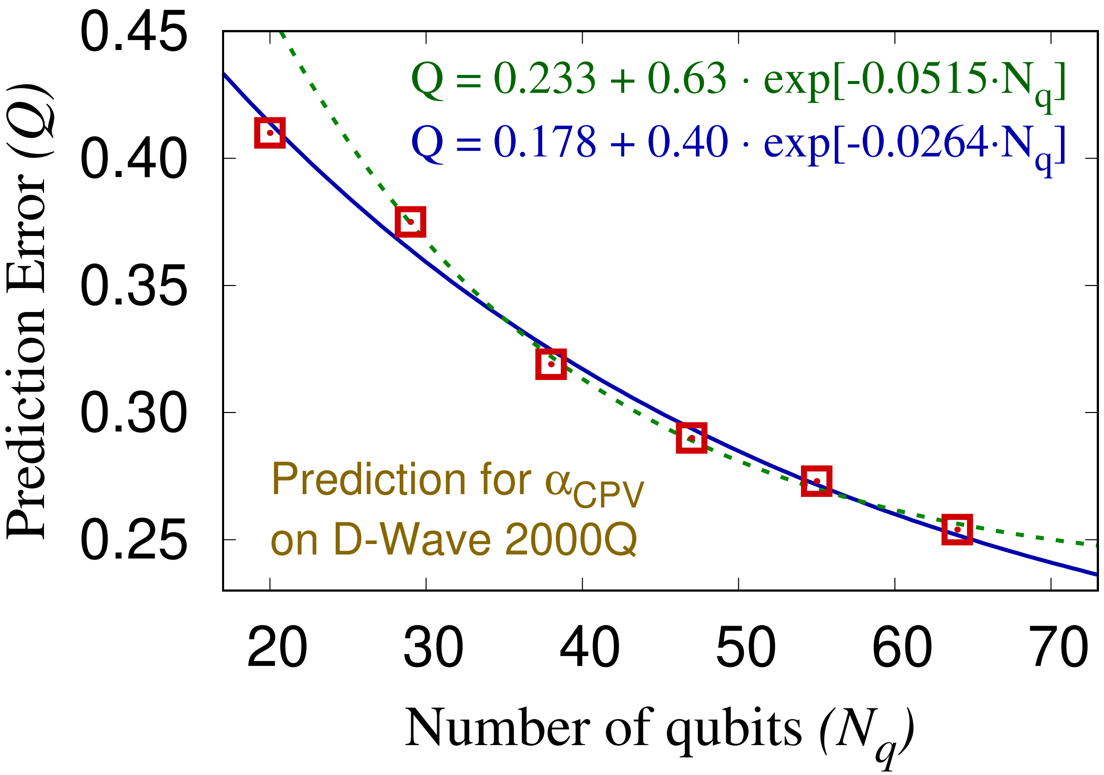
- $\boldsymbol{\alpha}$ (cEDM)
  - DM: 0.0527(17)
  - Prediction: 0.0525(18)
- $\boldsymbol{\alpha_5}$ $(\boldsymbol{\gamma_5})$
  - DM: -0.1463(14)
  - Prediction: -0.1460(17)

➢ DM: DM on 400 confs

➢ Prediction: DM on 120 confs + ML prediction on 280 confs

BY, Tanmoy Bhattacharya, Rajan Gupta, PRD 100, 014504 (2019)

Prediction of $C_{2pt}^{\mathbf{CPV}}$ from C_{2pt} on D-Wave Quantum Annealer

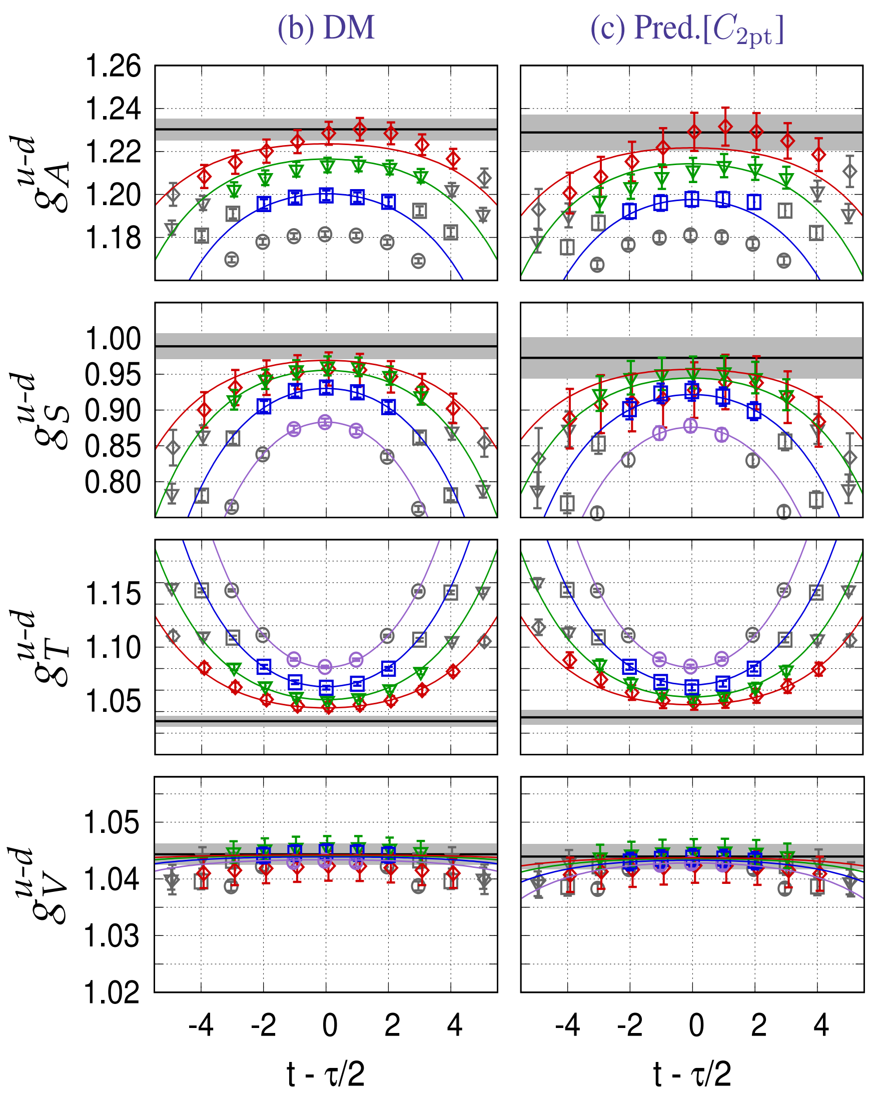


- D-Wave quantum processor realizes the quantum Ising spin system and finds the lowest (or the near-lowest) energy states
- We developed a new ML regression algorithm utilizing D-Wave as an efficient optimizer for ML loss function
- After encoding correlations between the cheap and expensive lattice QCD observables in the sparse coding dictionary $\boldsymbol{\phi}$, the dictionary is used to predict expensive observables
- Current prediction performance is limited by the available number of qubits on D-Wave

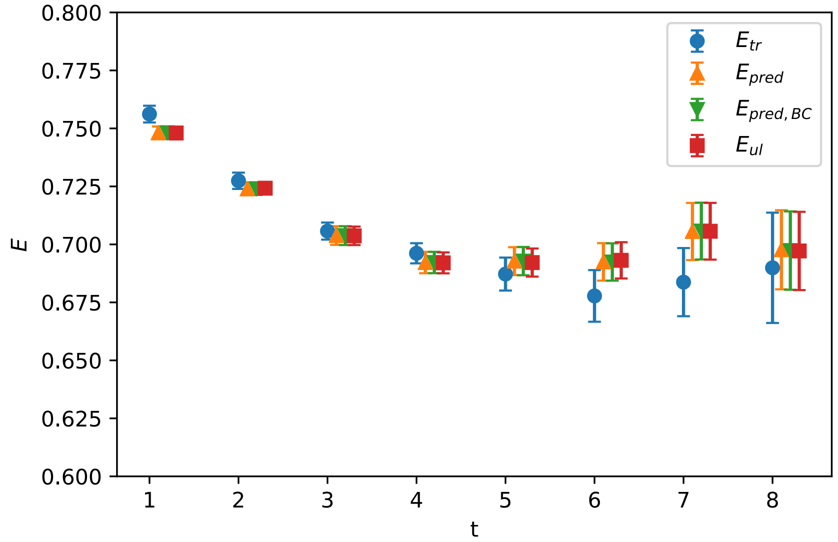
Nga Nguyen, Garrett Kenyon, BY, Sci. Rep. 10, 10915 (2020)

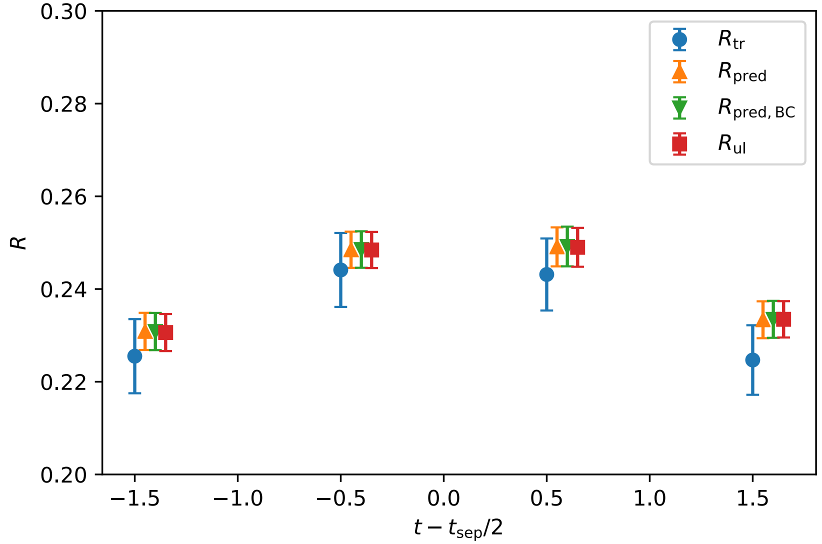
Other Applications

BY, Tanmoy Bhattacharya, Rajan Gupta, PRD 100, 014504 (2019)
Rui Zhang, Zhouyou Fan, Ruizi Li, Huey-Wen Lin, BY, PRD 101, 034516 (2020)



Prediction of C_{3pt} from C_{2pt}





Prediction of η_s distribution amplitude (upper) and Kaon quasi-PDF (lower) $z = 4$ from $z < 4$

Summary

- Machine learning (ML) is employed to predict unmeasured observables from measured observables
  (Expensive lattice QCD calculation → Cheap ML estimators)

- Bias correction is used to quantify the ML prediction error

- Demonstrated in lattice QCD calculations
  1) Prediction of C_{2pt}^{CPV} from C_{2pt}
  2) Prediction of C_{3pt} from C_{2pt}
  3) Prediction of Quasi-PDF Matrix Elements from lower z or lower p

- Developed a new regression algorithm utilizing D-Wave quantum annealer and showed promising prediction ability