

Prediction of lattice QCD observables using machine learning

Boram Yoon

Los Alamos National Laboratory

BY, Tanmoy Bhattacharya, Rajan Gupta, *Phys. Rev. D* 100, 014504 (2019)

Rui Zhang, Zhouyou Fan, Ruizi Li, Huey-Wen Lin, BY, *Phys. Rev. D* 101, 034516 (2020)

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

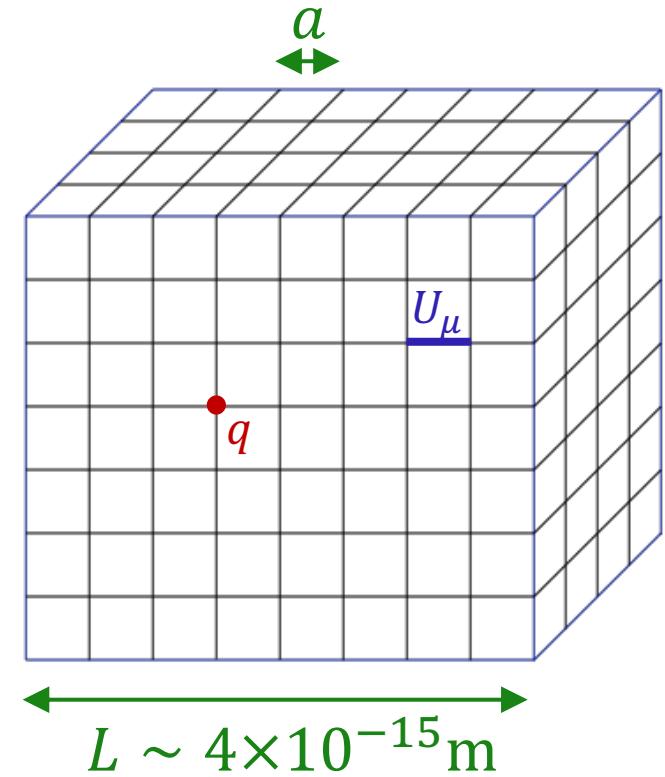
Outline

1. Correlations in lattice QCD observables
2. Machine learning predictions of lattice QCD observables
3. Applications
4. Summary

Correlations in Lattice QCD Observables

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 $a \rightarrow 0$
- Has been successful for many QCD observables
 - Some results are with less than 1% error



Lattice QCD

- Correlation functions

$$\begin{aligned}\langle O \rangle &= Z^{-1} \int dU d\bar{q} d\bar{\bar{q}} O(U, q, \bar{q}) e^{-S_g - \bar{q}(D + m_q)q} \\ &= Z^{-1} \int dU \left[O\left(U, (D + m_q)^{-1}\right) e^{-S_g} \det(D + m_q) \right]\end{aligned}$$

- Monte-Carlo integration

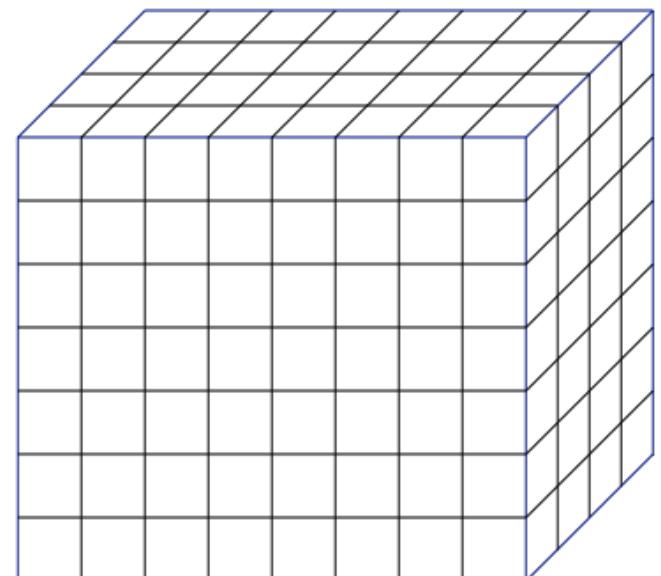
- Integration variable U is huge

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

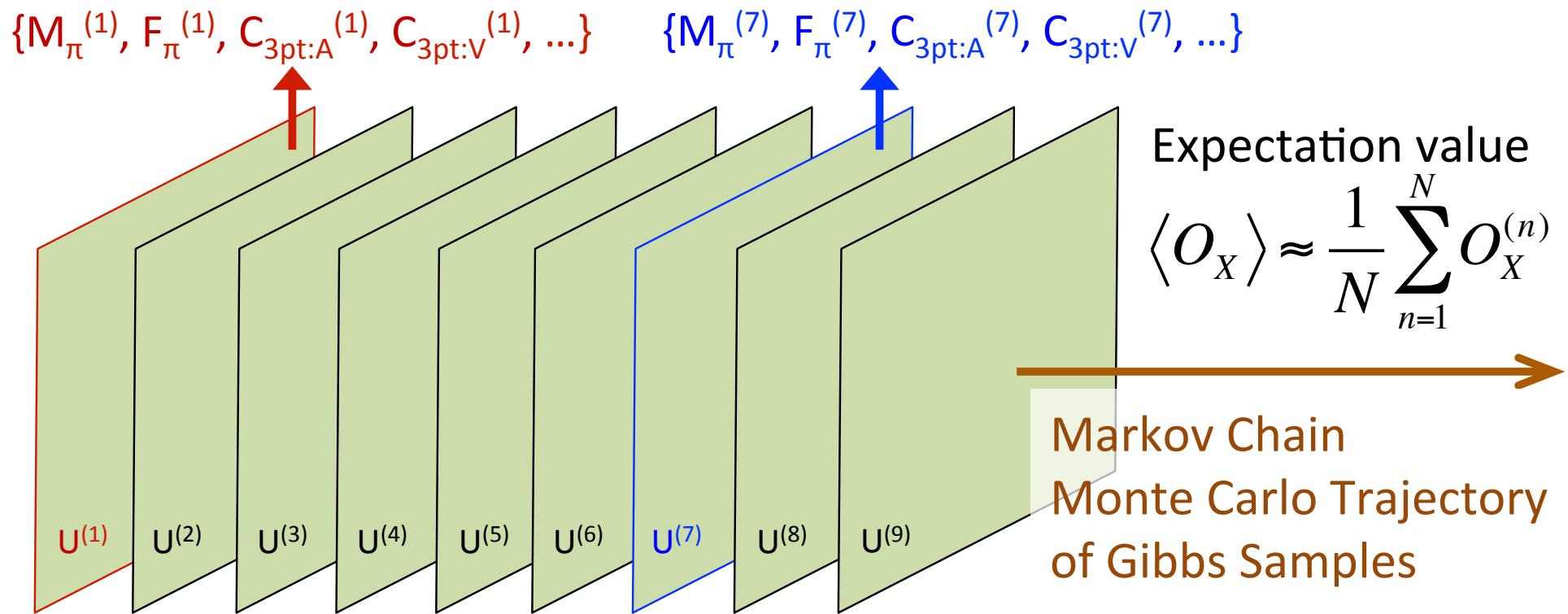
- Generate Markov chain of gauge configurations U
 - Calculate average as expectation value

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

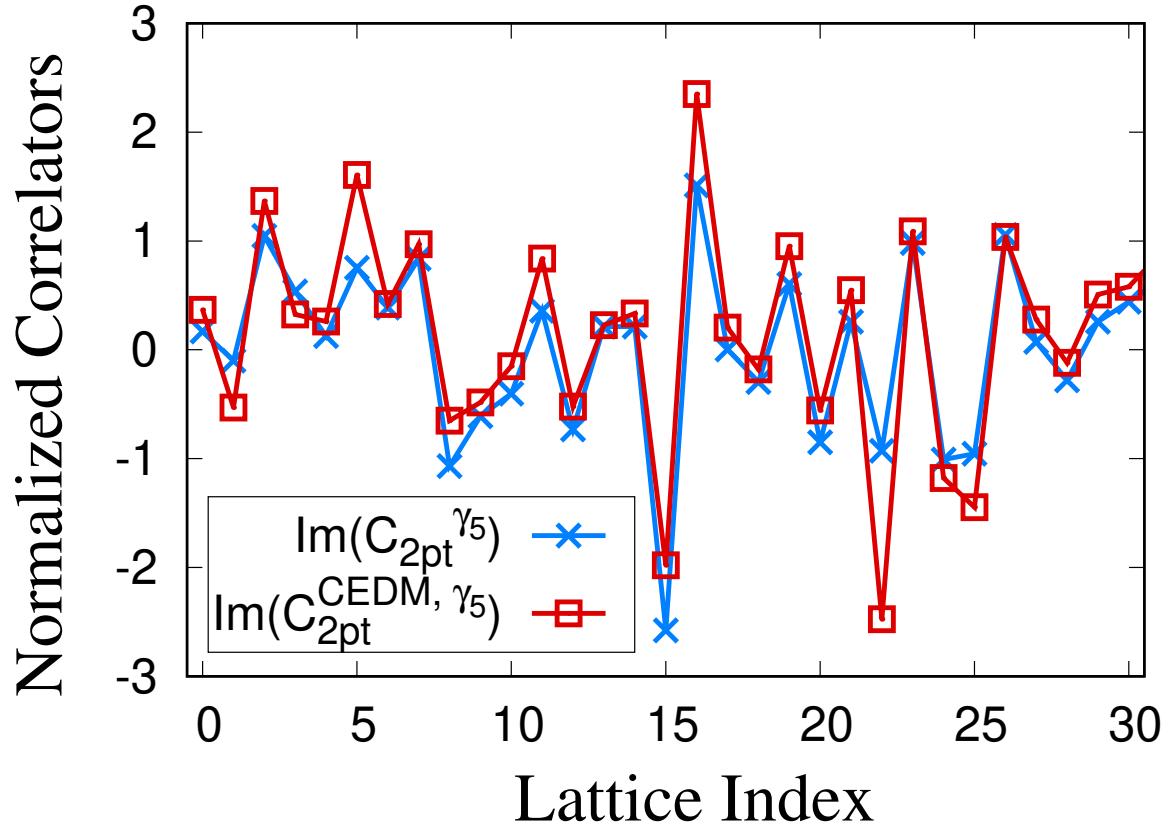
- Calculation of $O_i \left(U, (D + m_q)^{-1} \right)$: measurement
 - $(D + m)^{-1}$ is computationally expensive



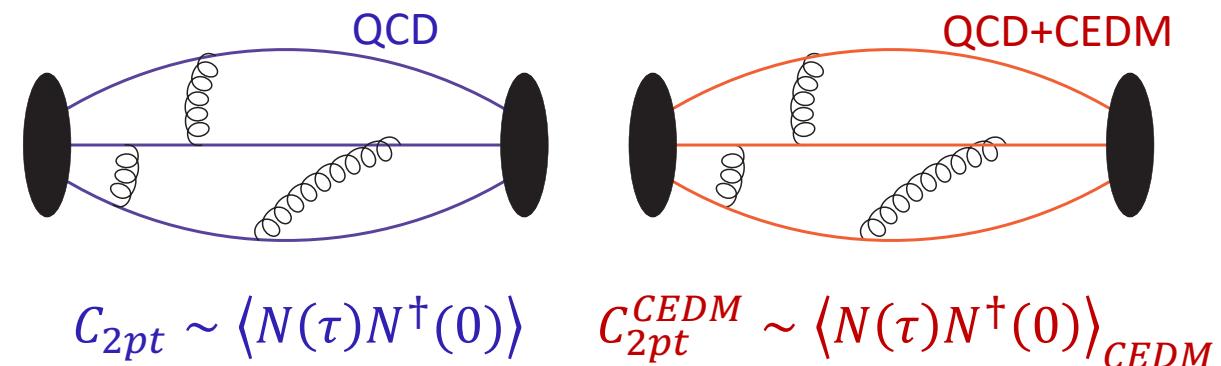
Lattice QCD Observables are Correlated



Correlation Map of Nucleon Observables

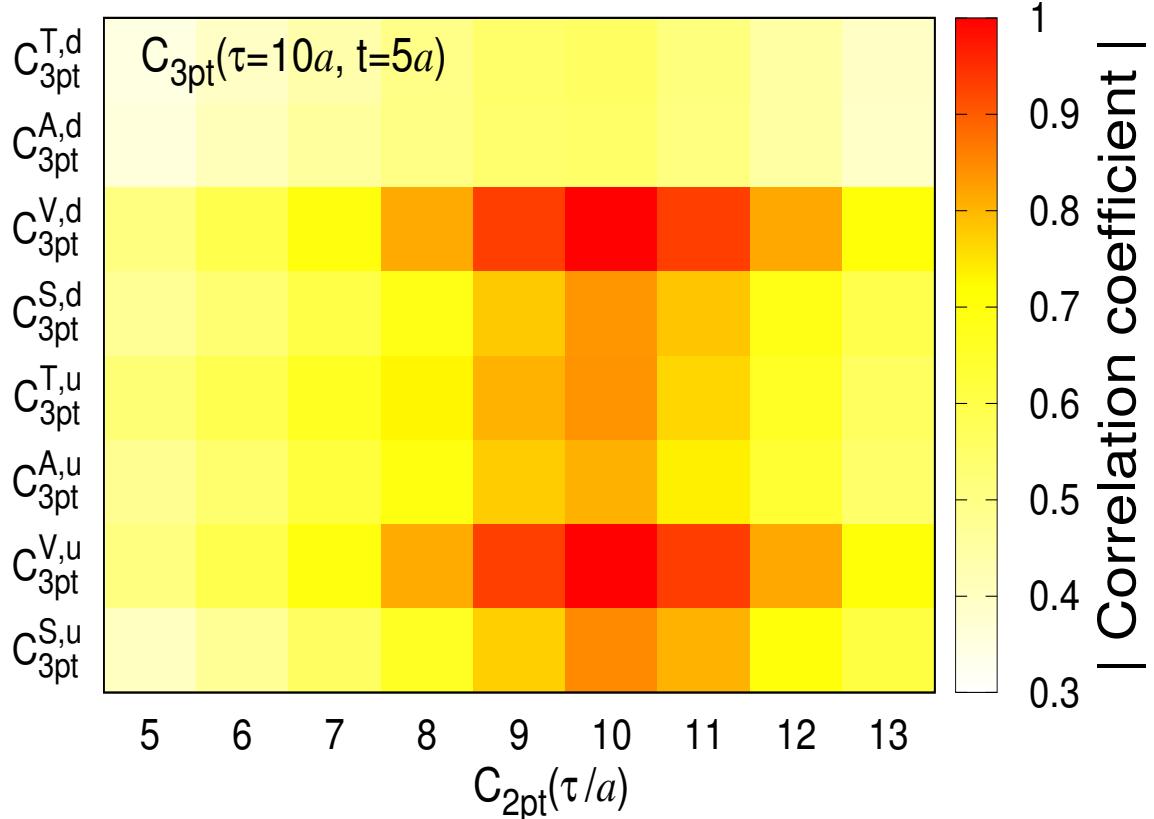


- Correlation between proton(uud) 2-pt correlation function and that calculated in presence of CEDM interaction

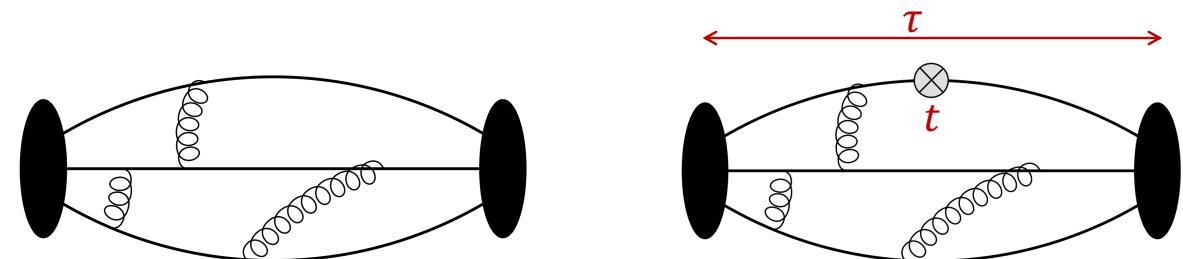


- QCD: D_{clov}
- QCD+CEDM: $D_{\text{clov}} + \frac{i}{2} \epsilon \sigma^{\mu\nu} \gamma_5 G_{\mu\nu}$

Correlation Map of Nucleon Observables



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



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

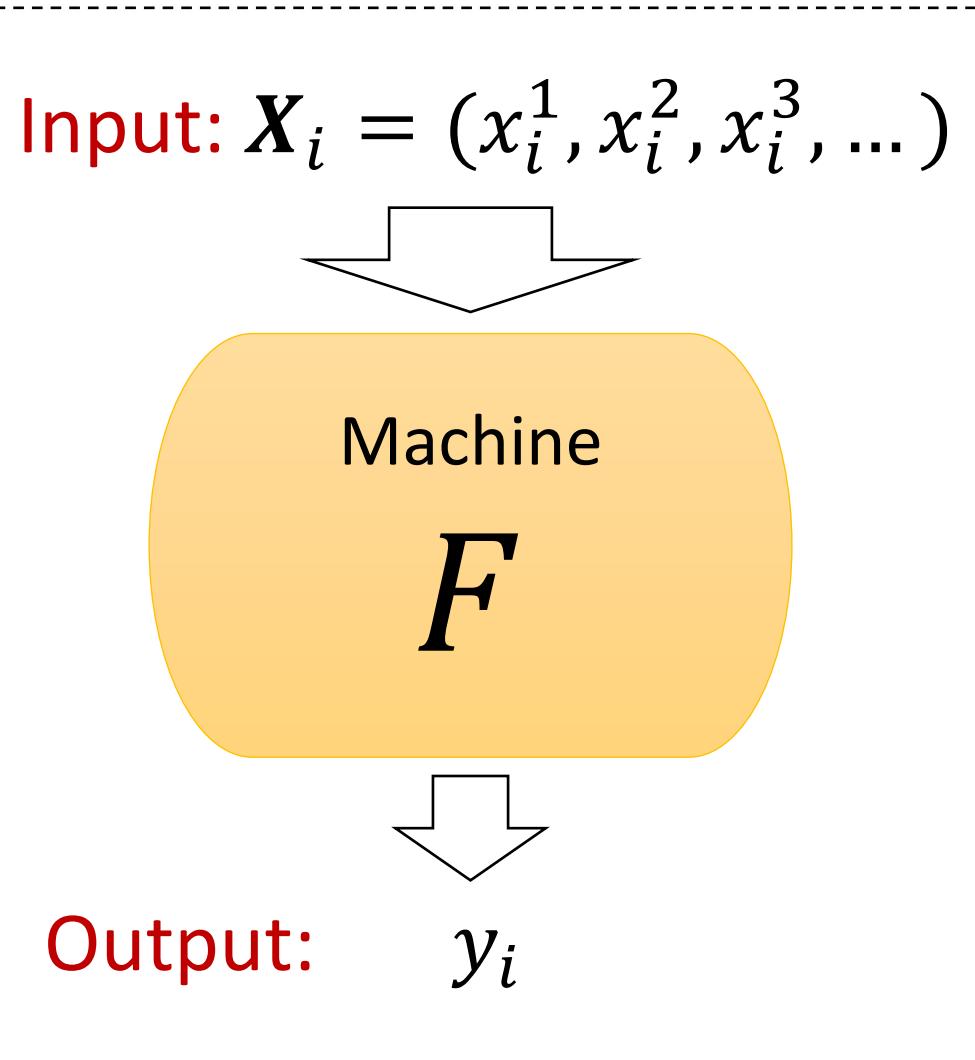
- Using these correlations, C_{3pt} can be estimated from C_{2pt} on each configuration

Machine Learning Predictions on Lattice QCD Observables

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

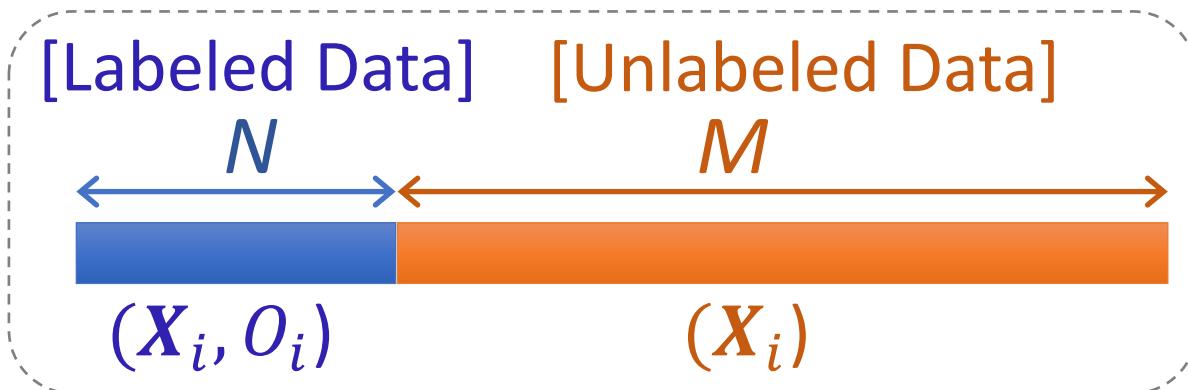
Machine Learning

- One can consider the machine learning (ML) process as a **data fitting**
- The machine F has very **general fitting functional form** with huge number of free parameters
- The free parameters are determined from large number of training data:
$$F(\mathbf{X}_i) \approx y_i$$
- For example,
 \mathbf{X}_i : area, house age, # of rooms, ...
 y_i : price of the house

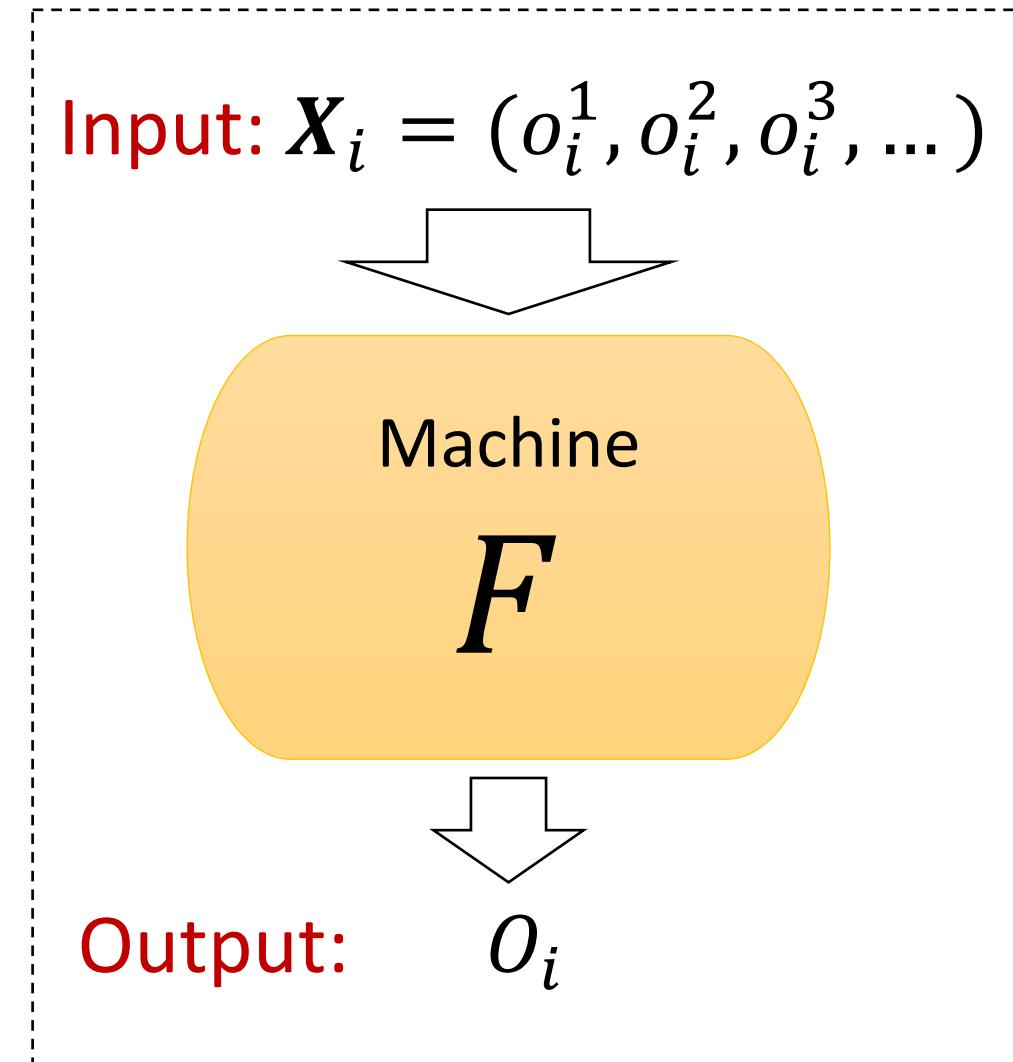


Machine Learning on Lattice QCD Observables

- Assume $N+M$ indep. measurements
- Common observables X_i on all $N+M$
Target observable O_i only on N



- 1) **Train** machine F to yield O_i from X_i on the Labeled Data
- 2) **Predict** O_i of the Unlabeled data from X_i
$$F(X_i) = O_i^P \approx O_i$$



Prediction Bias

- $F(X_i) = O_i^P \approx O_i$

- Simple average

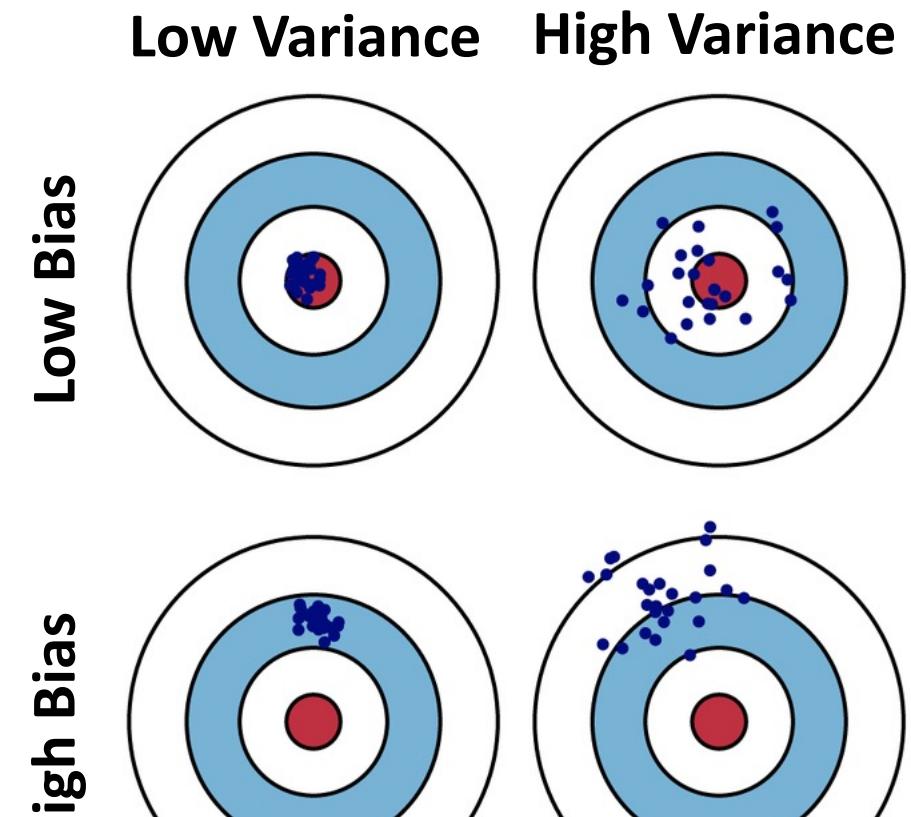
$$\bar{O} = \frac{1}{M} \sum_{i \in \text{Unlabeled}} O_i^P$$

is not correct due to **prediction bias**

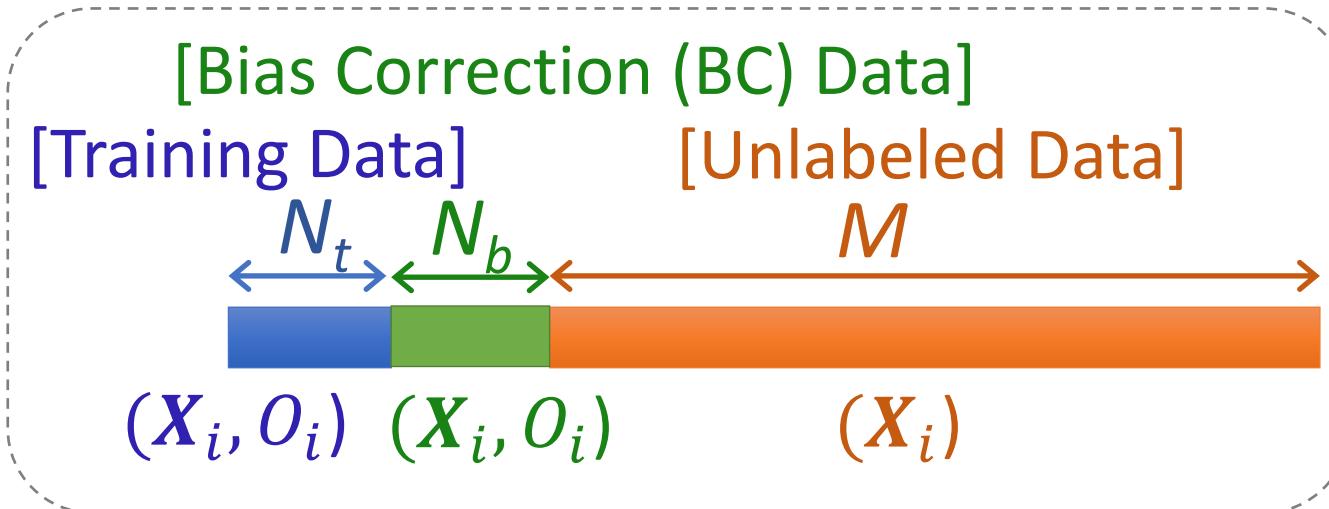
- Prediction = TrueAnswer + Noise + Bias
- ML prediction may have bias

$$\langle O_i^P \rangle \neq \langle O_i \rangle$$

$$\text{Bias} = \langle O_i^P \rangle - \langle O_i \rangle$$



Bias Correction



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

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

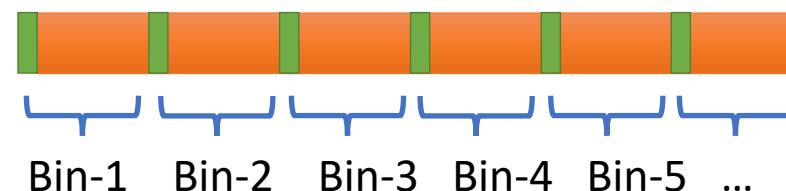
- Expectation value, $\langle \bar{O}_{BC} \rangle = \langle O_i^P \rangle + \langle O_i - O_i^P \rangle = \langle O_i \rangle$
- BC term converts **systematic error of prediction** to **statistical uncertainty**

Incorporating Labeled Data

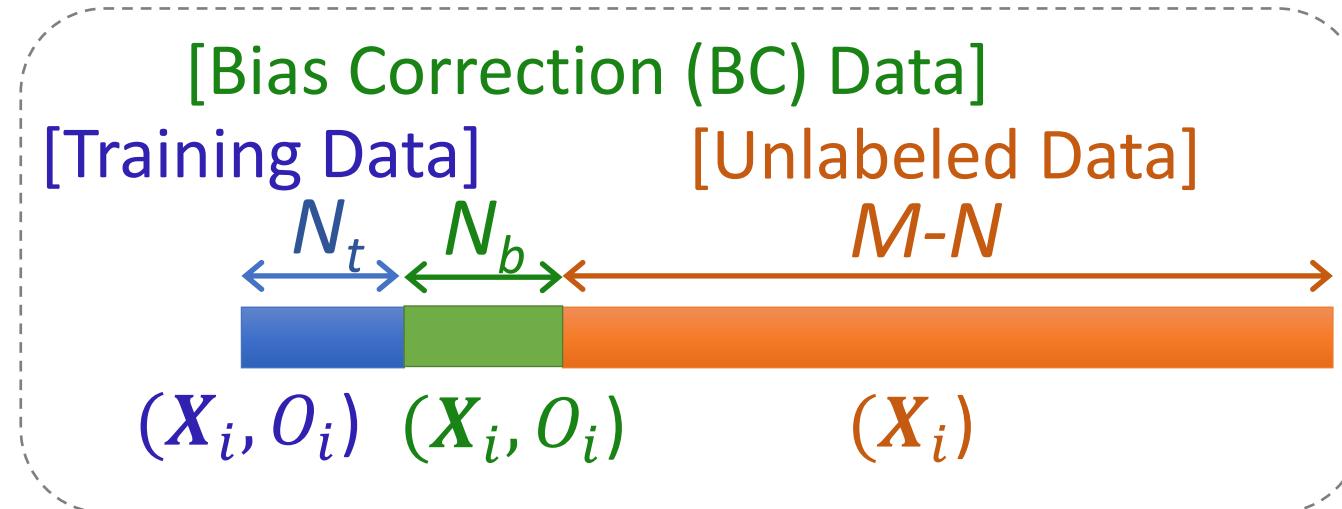
- Include directly measured values O_i from labeled data

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

- 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
- Binning and BC for each bin is another option for complicated data analysis



Error Estimation



- Training, bias correction and predictions are made on **statistical data**
- Final estimates, including the trained ML algorithm, inherit uncertainties
- We use **bootstrap resampling** for error estimation
 - Random choice of bootstrap samples (labeled and unlabeled data set, separately)
 - Split training, BC data in labeled data
 - Train, BC and obtain estimates on each bootstrap sample

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

$$\begin{aligned} \sigma_{\bar{o}_{BC}}^2 &\approx \frac{1}{M} \sigma_{O^P}^2 + \frac{1}{N_{bc}} \sigma_{O-O^P}^2 \\ &\approx \frac{\sigma_o^2}{N} \left(1 + \frac{M}{N_{bc}} \frac{\sigma_{O-O^P}^2}{\sigma_o^2} \right) \equiv \frac{\sigma_o^2}{N} \left(1 + \frac{M}{N_{bc}} Q^2 \right); \quad Q \equiv \frac{\sigma_{O-O^P}}{\sigma_o} \end{aligned}$$

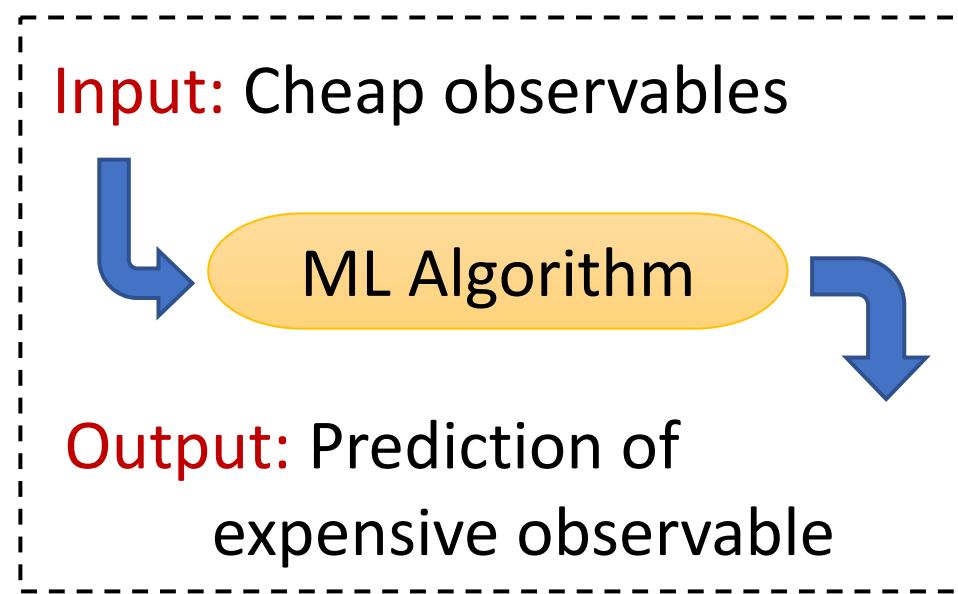
approximations (\approx) for small correlation between the two terms and a good prediction algorithm that gives $\sigma_o^2 \approx \sigma_{O^P}^2$

$$\frac{\sigma_{\bar{o}_{BC}}^2}{\sigma_{\bar{o}}^2} \approx 1 + \frac{M}{N_{bc}} Q^2; \quad \frac{\sigma_{\bar{o}_{BC}}}{\sigma_{\bar{o}}} \approx 1 + \frac{M}{2N_{bc}} Q^2$$

| for $N_{bc}/M = 0.2$ | |
|----------------------|----------------|
| Q | Error Increase |
| 0.5 | 62.5% |
| 0.3 | 22.5% |
| 0.1 | 2.5% |

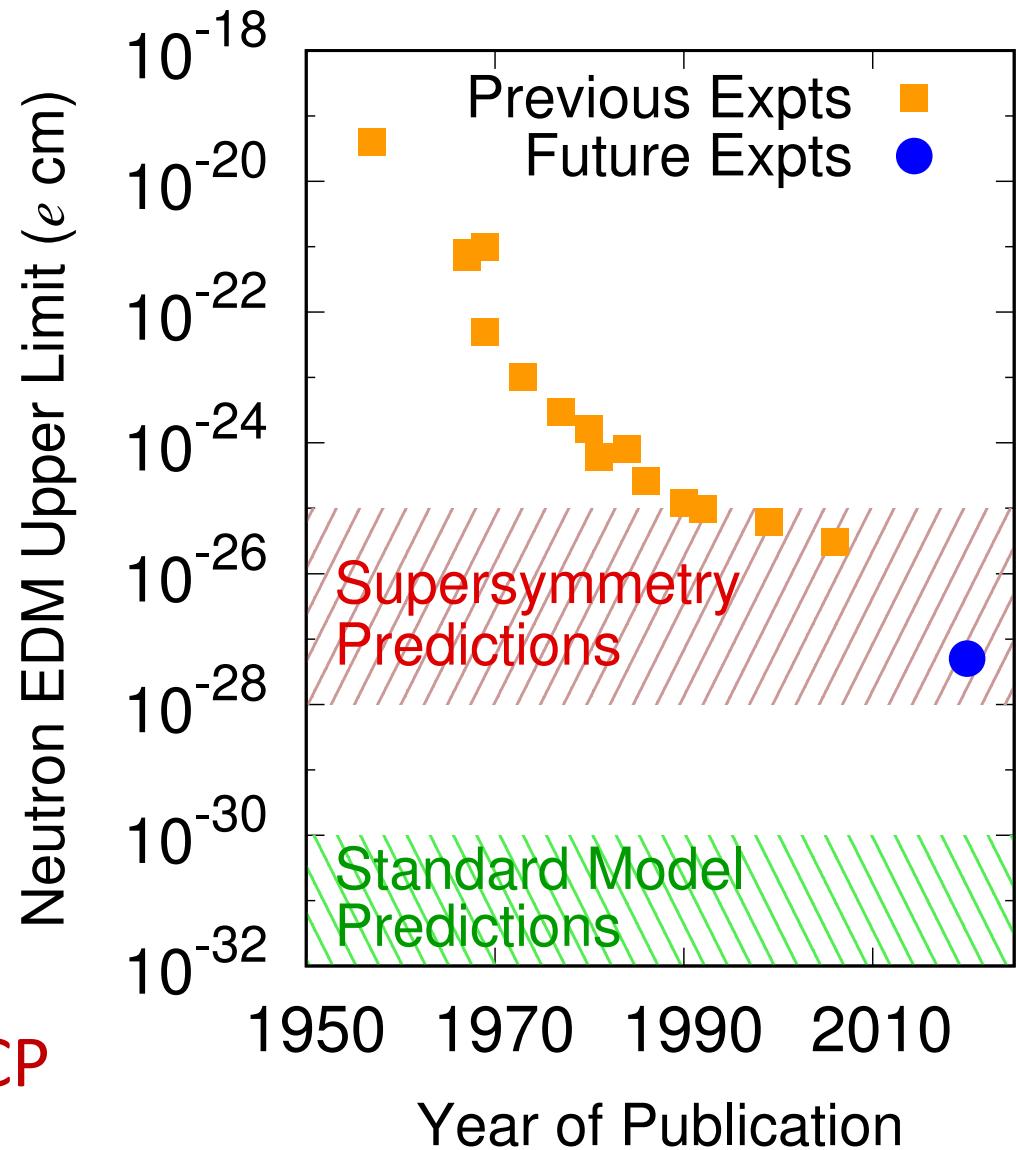
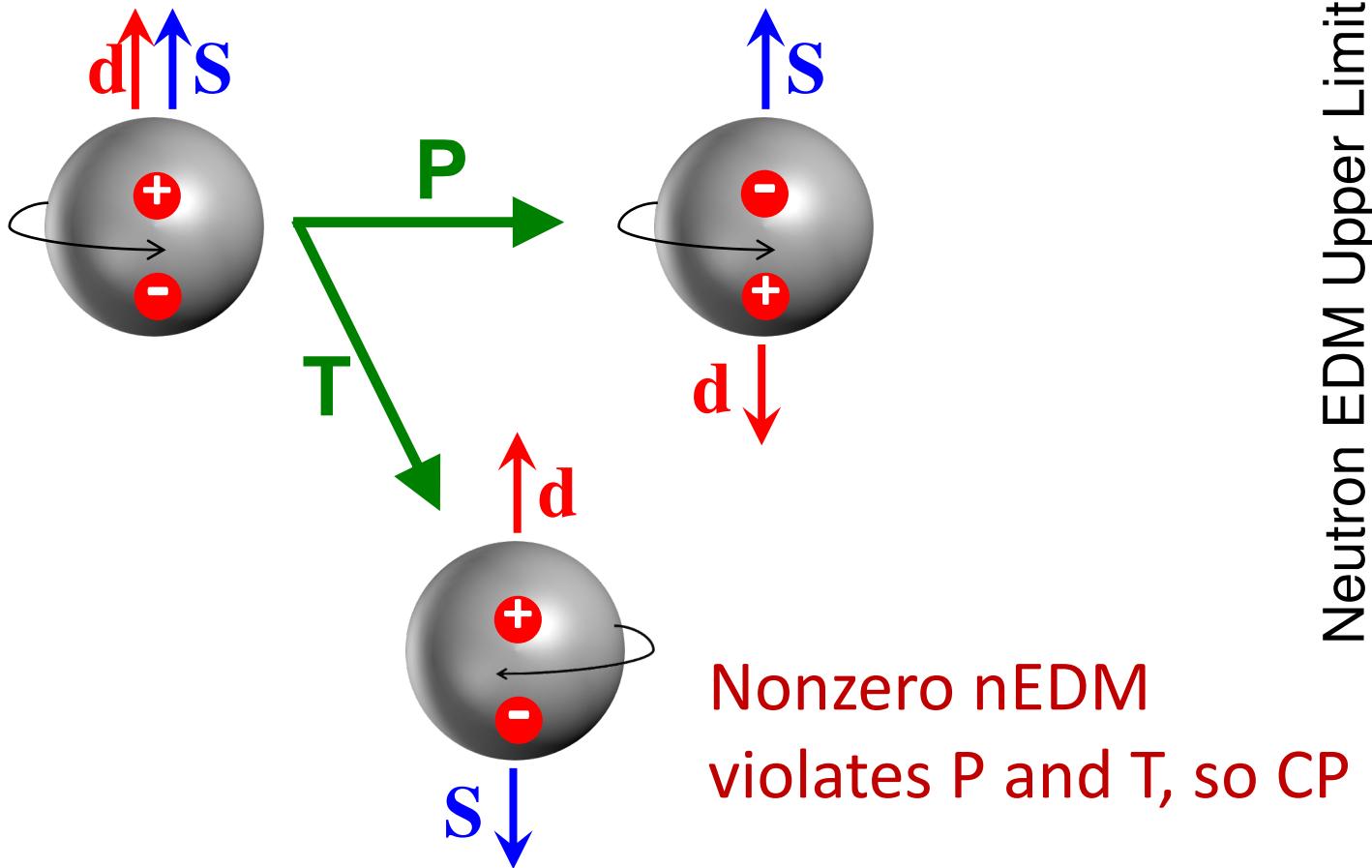
- Q-value shows the expected error-increase due to the ML prediction error
- In practice, BC data have less autocorrelation than full data, because of the many measurements per configuration, so $\sigma_{\bar{o}_{BC}}$ gives smaller error than expected above

Applications



Neutron EDM and CP Violation

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



Effective CPV Lagrangian

$$\mathcal{L}_{\text{CPV}}^{d \leq 6} = -\frac{g_s^2}{32\pi^2} \bar{\theta} G \tilde{G}$$

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

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

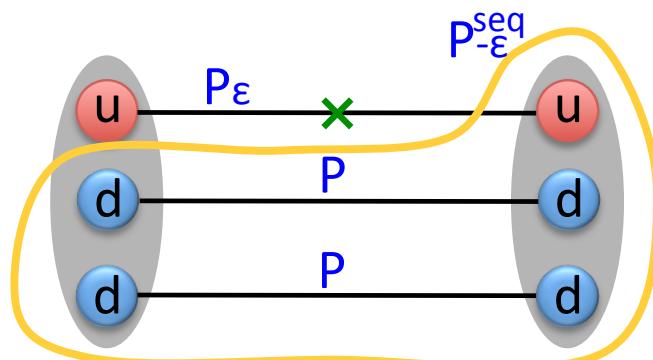
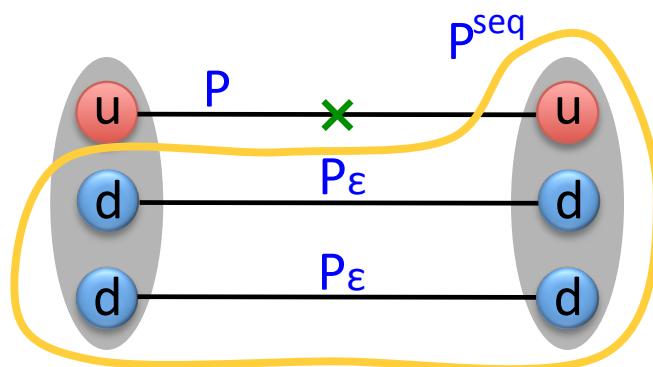
$$S_{cEDM} = -\frac{i}{2} \int d^4x \ \tilde{d}_q g_s \bar{q} (\sigma \cdot G) \gamma_5 q$$

- Schwinger source method

Include cEDM term in valence quark propagators
by modifying Dirac operator

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

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

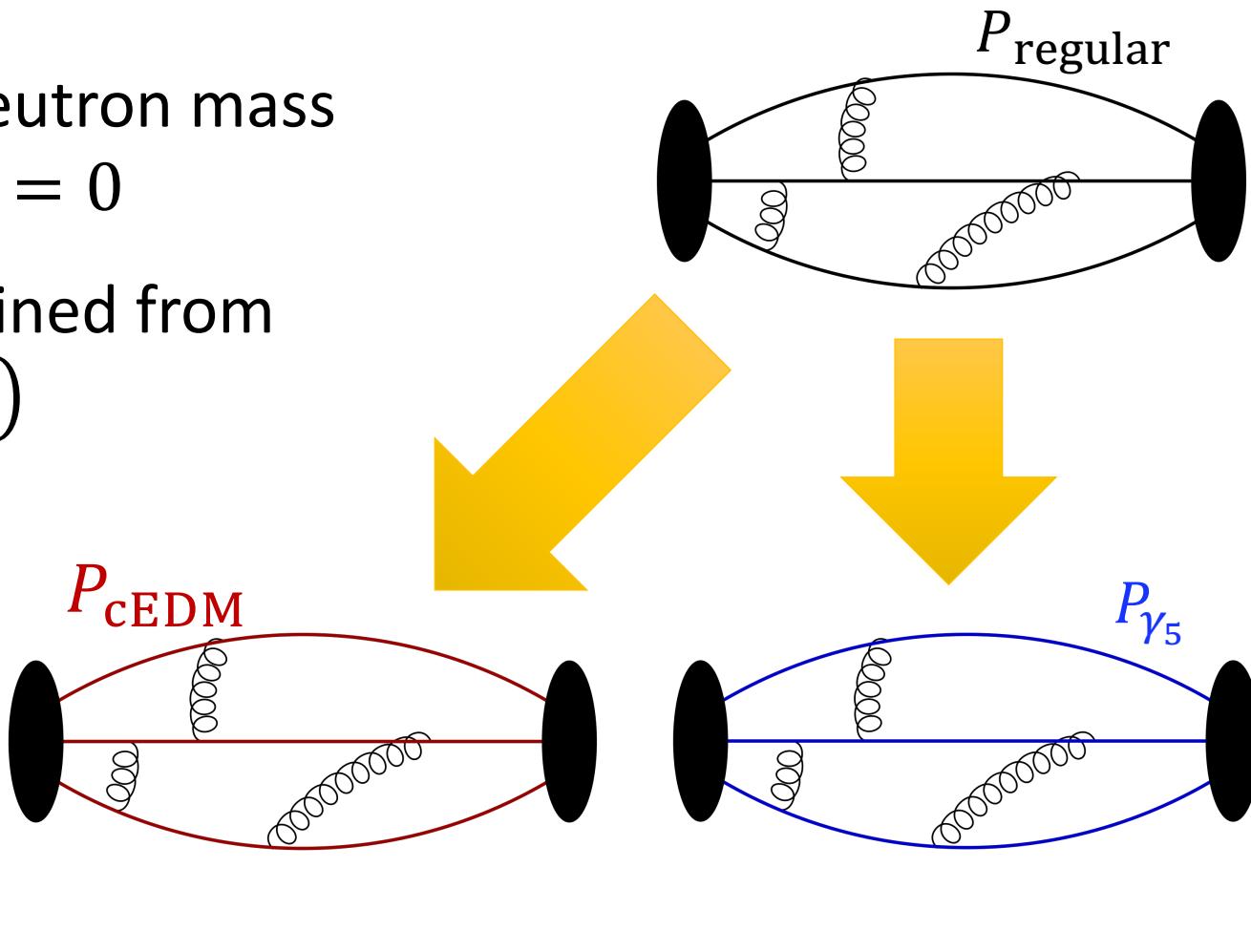
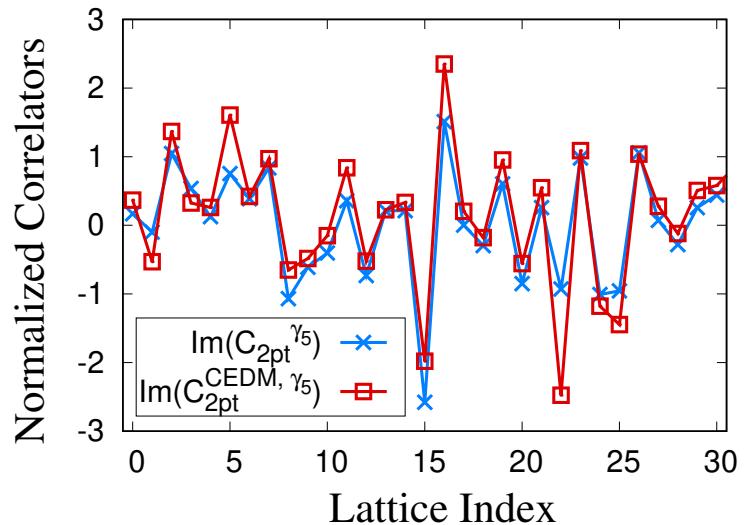


Prediction of C_{2pt}^{CPV} from C_{2pt}

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

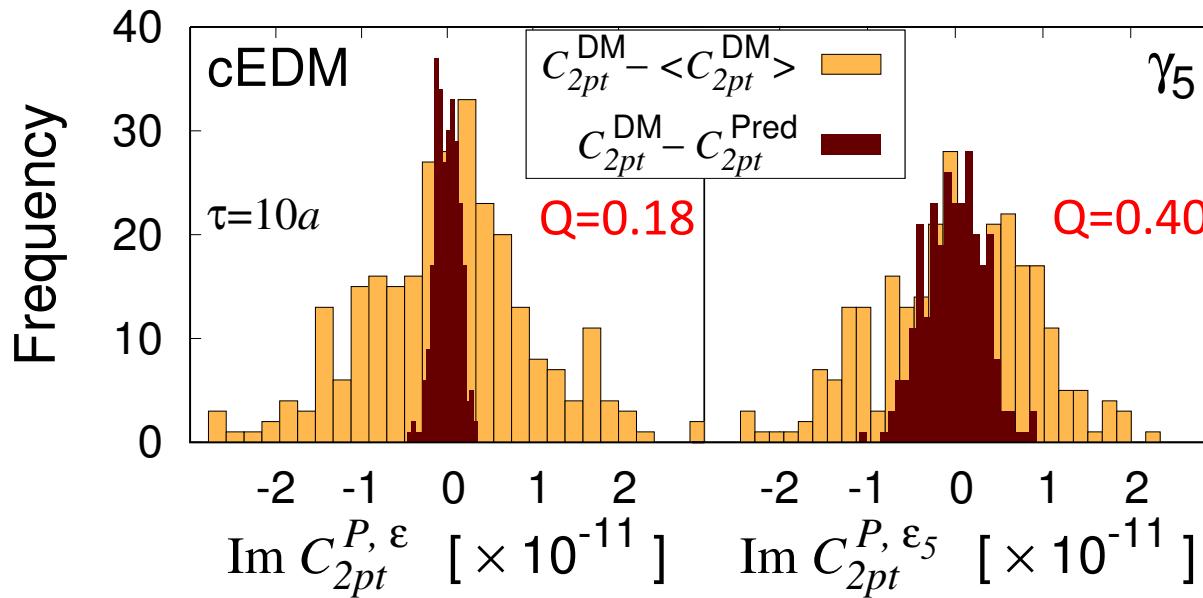
$$(ip_\mu\gamma_\mu + me^{-2i\alpha\gamma_5})u_N = 0$$
- At leading order, α can be obtained from

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



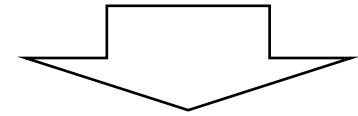
Prediction of C_{2pt}^{CPV} from C_{2pt}

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



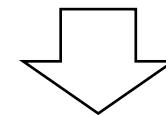
Input:

$$X_i = \{\text{Re}, \text{Im}[C_{2pt}^{S,P}(0 \leq \tau/a \leq 16)]\}$$

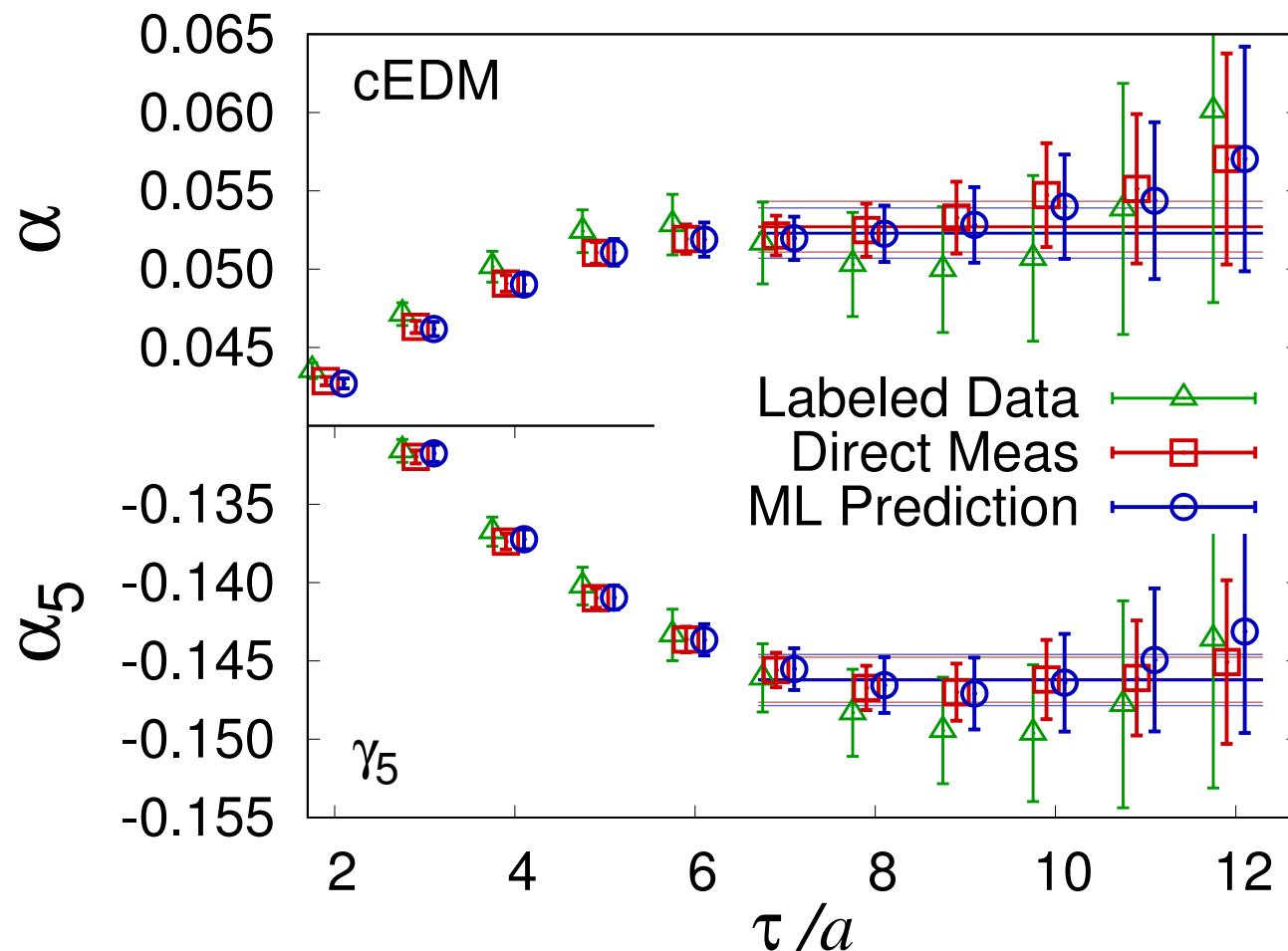


Boosted
Decision Tree
Regression

Output: $\text{Im } [C_{2pt}^P(c\text{EDM}, \gamma_5)(\tau)]$

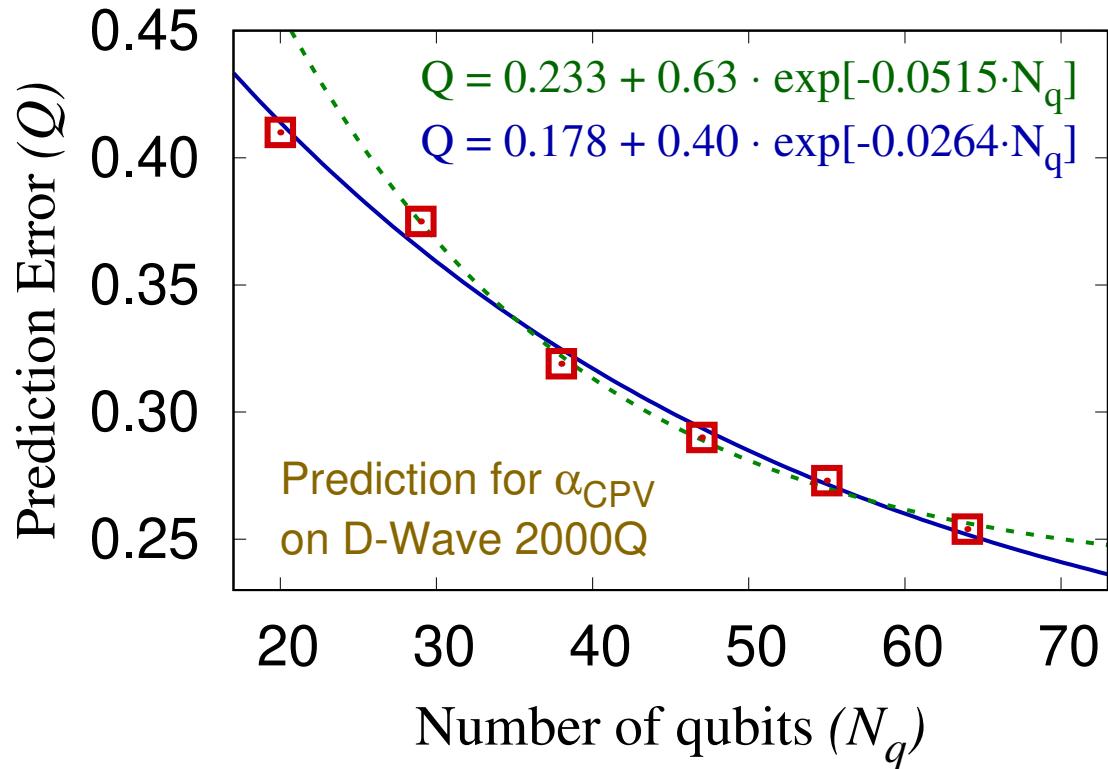


Prediction of C_{2pt}^{CPV} from C_{2pt}



- α (cEDM)
 - DM: 0.0527(17)
 - Prediction: 0.0525(18)
- α_5 (γ_5)
 - DM: -0.1463(14)
 - Prediction: -0.1460(17)
- DM: DM on 400 confs
- Prediction: DM on 120 confs + ML prediction on 280 confs

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



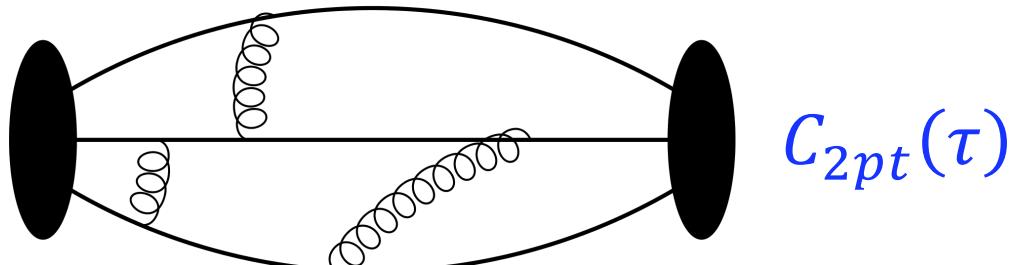
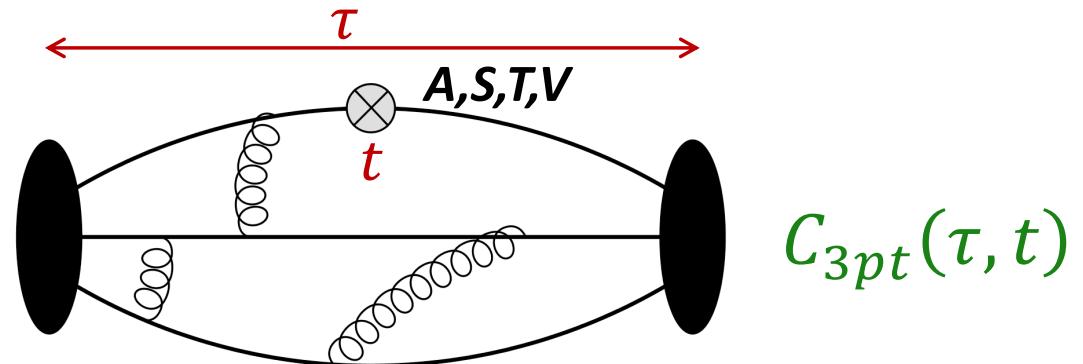
- D-Wave quantum processor realizes the quantum Ising spin system and finds the lowest (or the near-lowest) energy states
- We developed a new ML regression algorithm utilizing D-Wave as an efficient optimizer for ML loss function
- After encoding correlations between the cheap and expensive lattice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

Nucleon Isovector Charges $g_{A,S,T,V}$

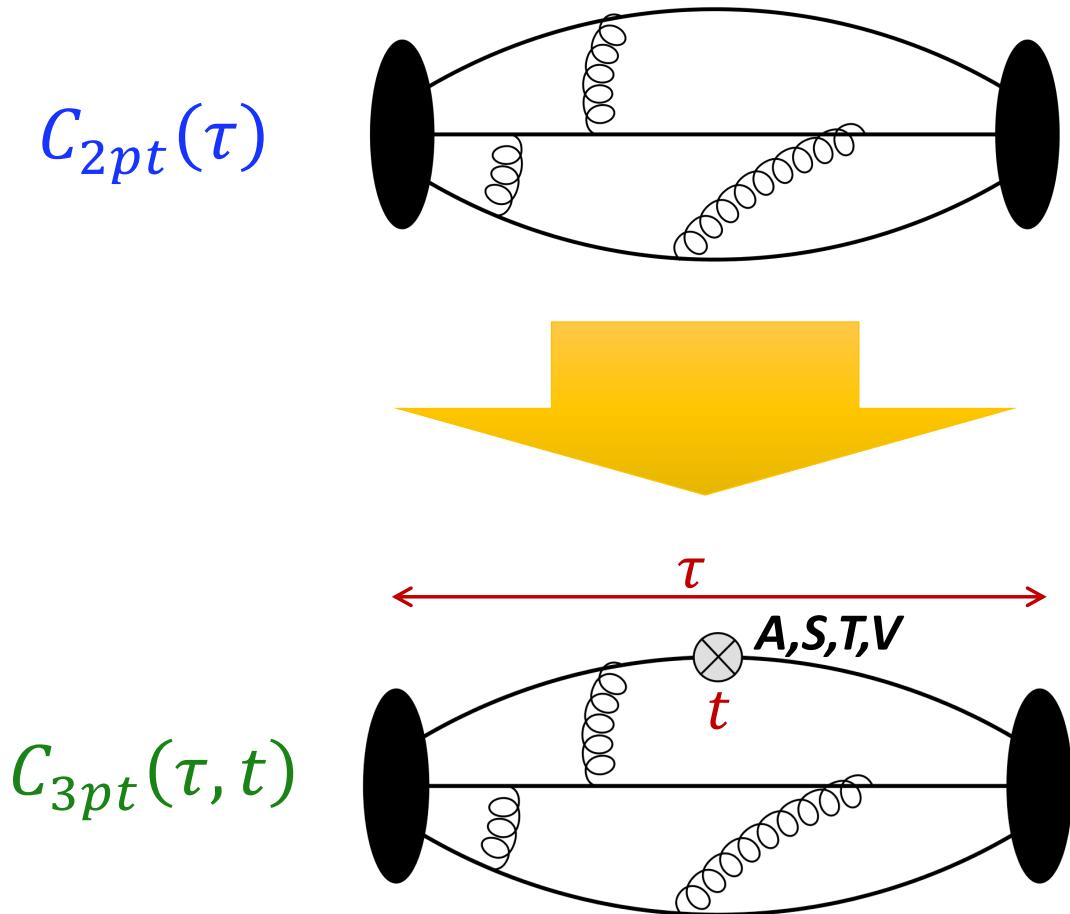
$$\langle p | \bar{u} \Gamma d | n \rangle = g_\Gamma \bar{\psi}_p \Gamma \psi_n$$

- $g_A = 1.2772(20)$ [Expt]
- g_T : quark EDM
- On lattice,

$$\frac{C_{3pt}}{C_{2pt}} \rightarrow g_\Gamma$$



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



Input: $X_i = \{C_{2pt}(0 \leq \tau/a \leq T_{\max})\}$



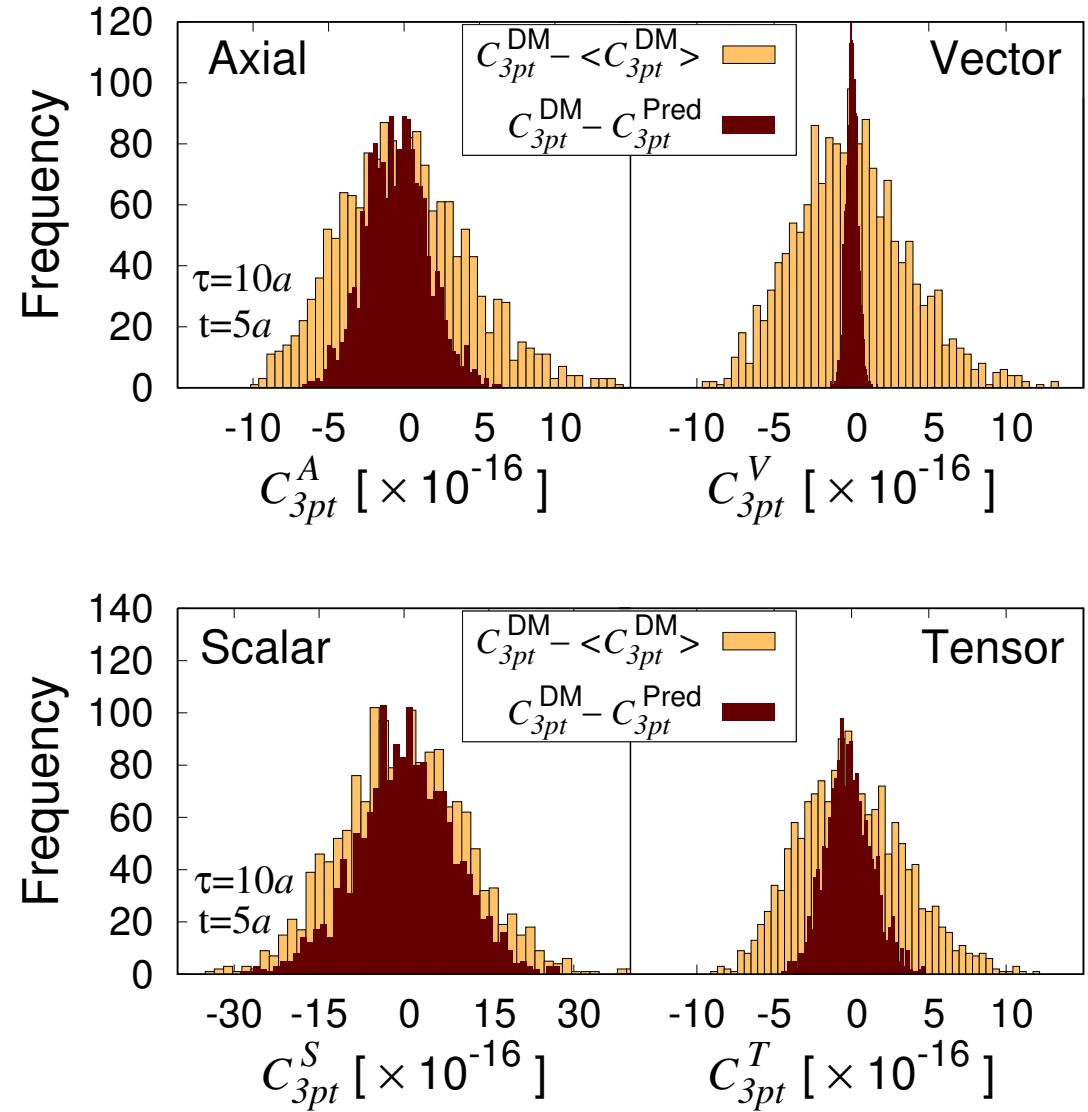
Output: $C_{3pt}^{A,S,T,V}(\tau, t)$

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

- Training and Test performed for
 - Clover-on-HISQ
 - $a = 0.089\text{fm}$, $M_\pi = 313 \text{ MeV}$
 - Measurements: 2263 confs \times 64 srcts
- # of training data: 60 confs
- # of BC data: 620 confs
- # of unlabeled data: 1583 confs
- Prediction error

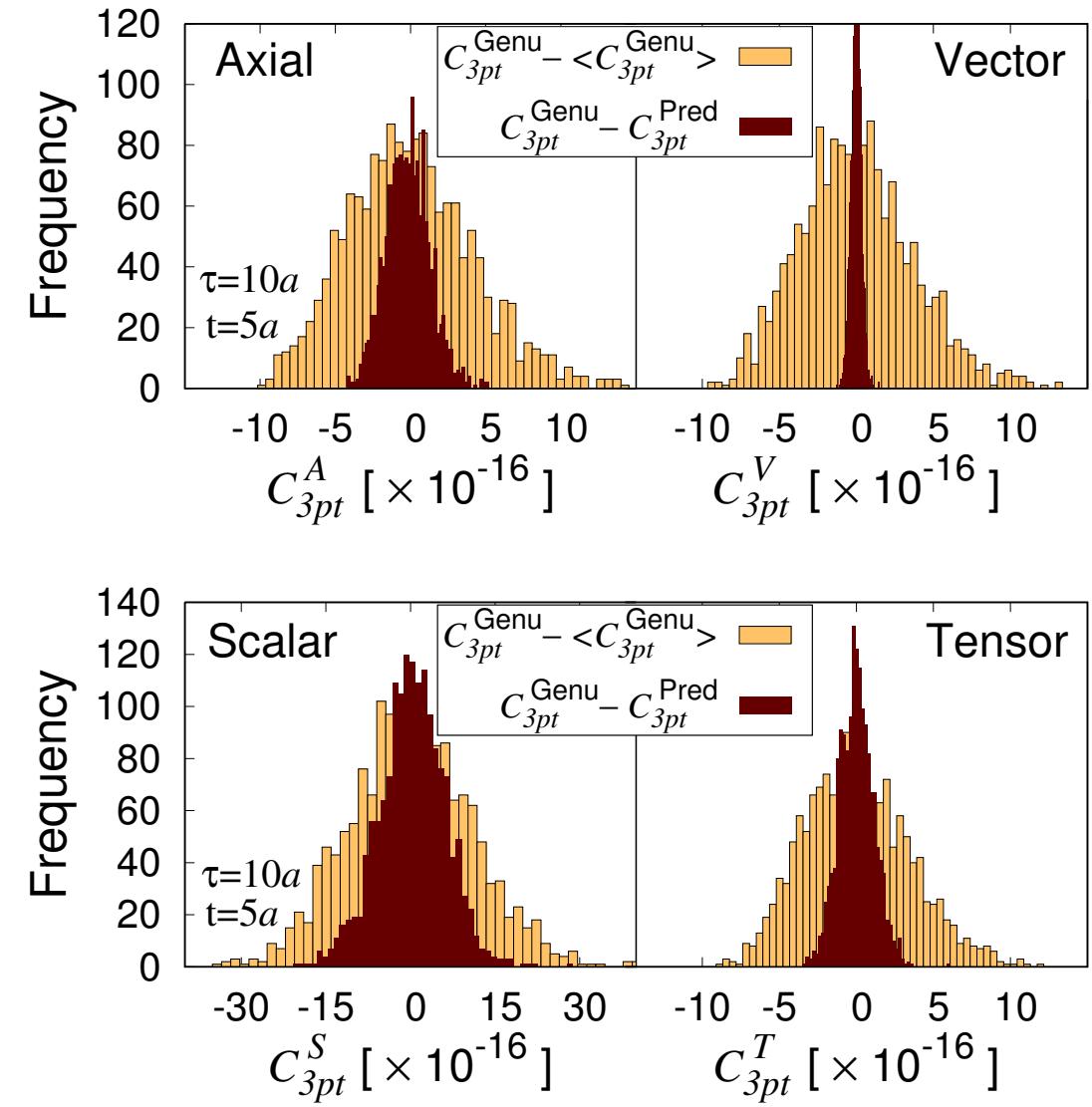
$$\text{PE} = C_{3pt}^{\text{DirectMeas.}} - C_{3pt}^{\text{Pred}}$$
- Prediction quality for $C_{3pt}^\Gamma(10,5)$

$$Q = 0.79 \text{ (S)}, 0.49 \text{ (A)}, 0.44 \text{ (T)}, 0.12 \text{ (V)}$$

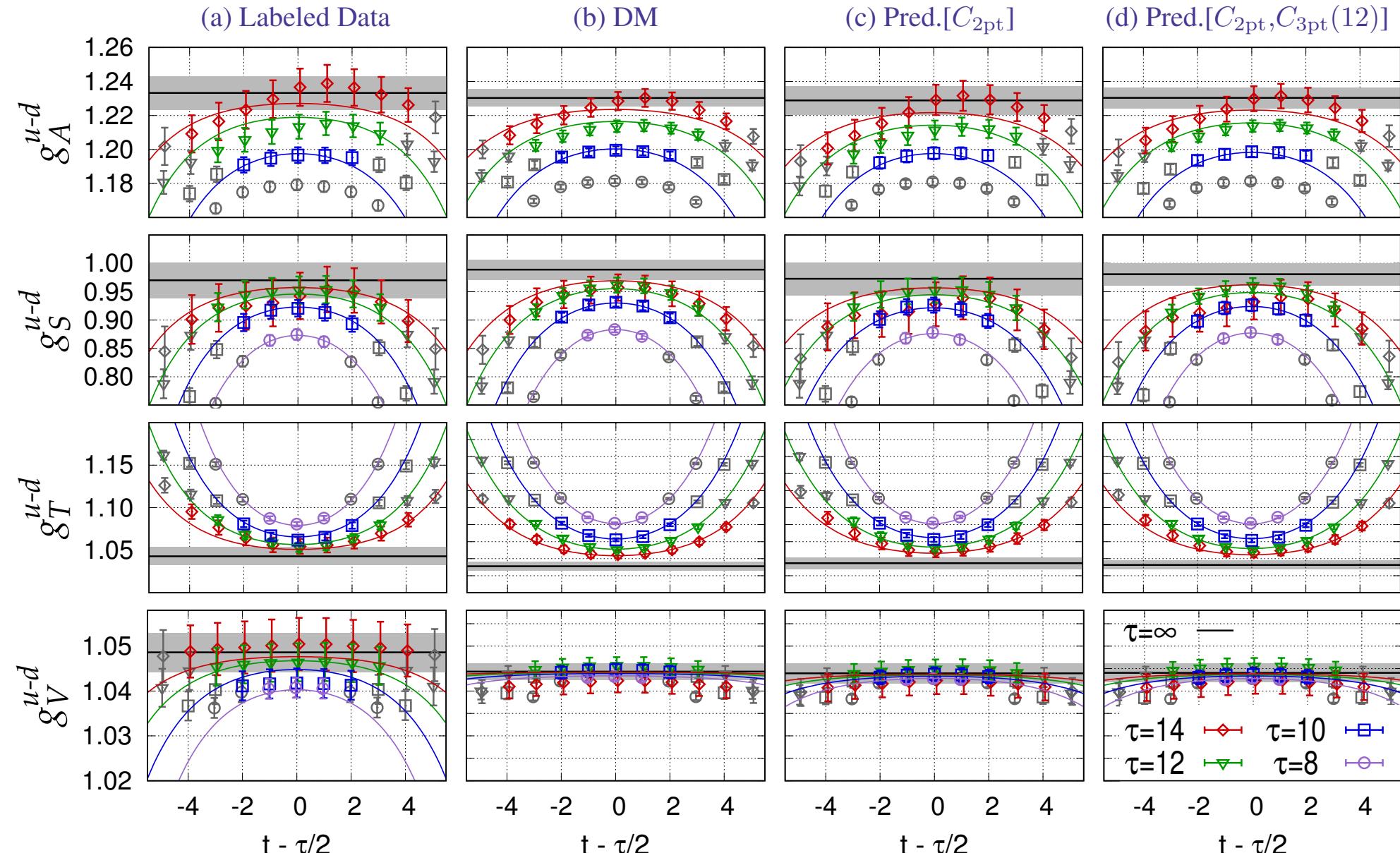


Prediction of $C_{3pt}(8, 10, 14)$ from C_{2pt} and $C_{3pt}(12)$

- We need C_{3pt} at four values of source-sink separations ($\tau = 8, 10, 12, 14$) for analysis
- $C_{3pt}(8, 10, 14)$ can be predicted from C_{2pt} and $C_{3pt}(12)$ data
- Prediction quality $C_{3pt}^{\Gamma}(10, 5)$
 $Q = 0.53 (S), 0.35 (A), 0.33 (T), 0.10 (V)$
(c.f.) only from C_{2pt} : $0.79 (S), 0.49 (A), 0.44 (T), 0.12 (V)$



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



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

- Results extrapolated to $\tau \rightarrow \infty$

| | DM | Pred. $[C_{2pt}]$ | Pred. $[C_{2pt}, C_{3pt}(12)]$ |
|-------|------------|-------------------|--------------------------------|
| g_S | 0.989(18) | 0.973(29) | 0.981(20) |
| g_A | 1.2303(51) | 1.2289(83) | 1.2304(61) |
| g_T | 1.0311(51) | 1.0347(68) | 1.0326(54) |
| g_V | 1.0443(19) | 1.0439(22) | 1.0440(21) |

2263 DM 680 DM 680 DM
(Direct Meas.) + 1583 Pred. + 1583 Pred.

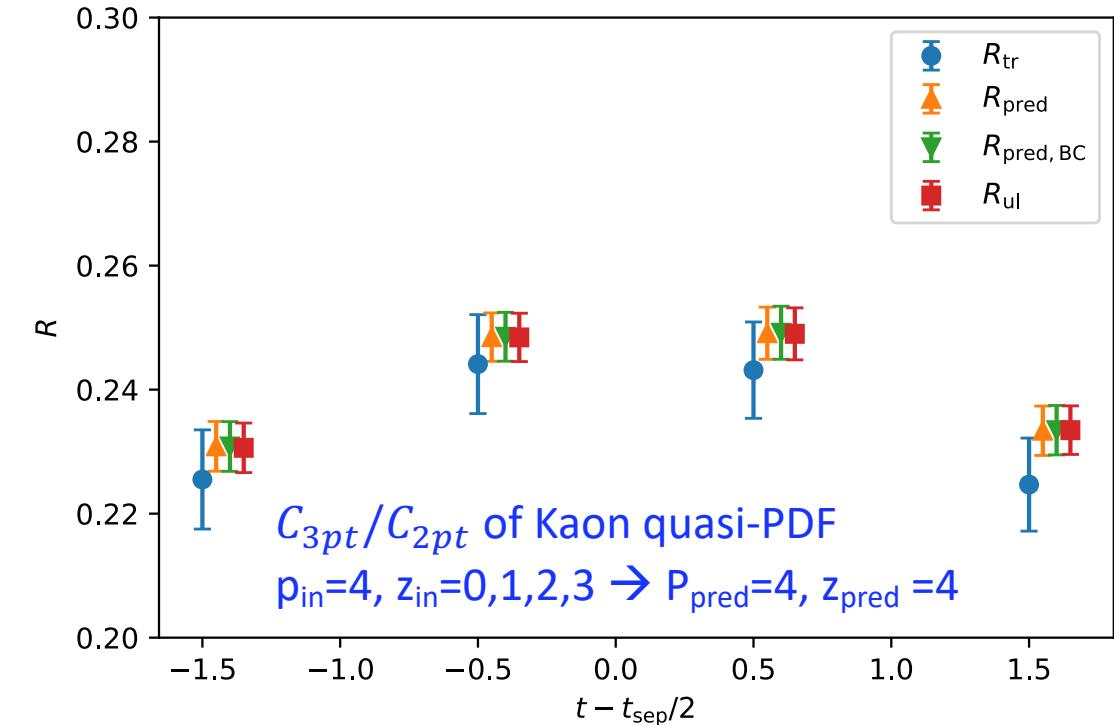
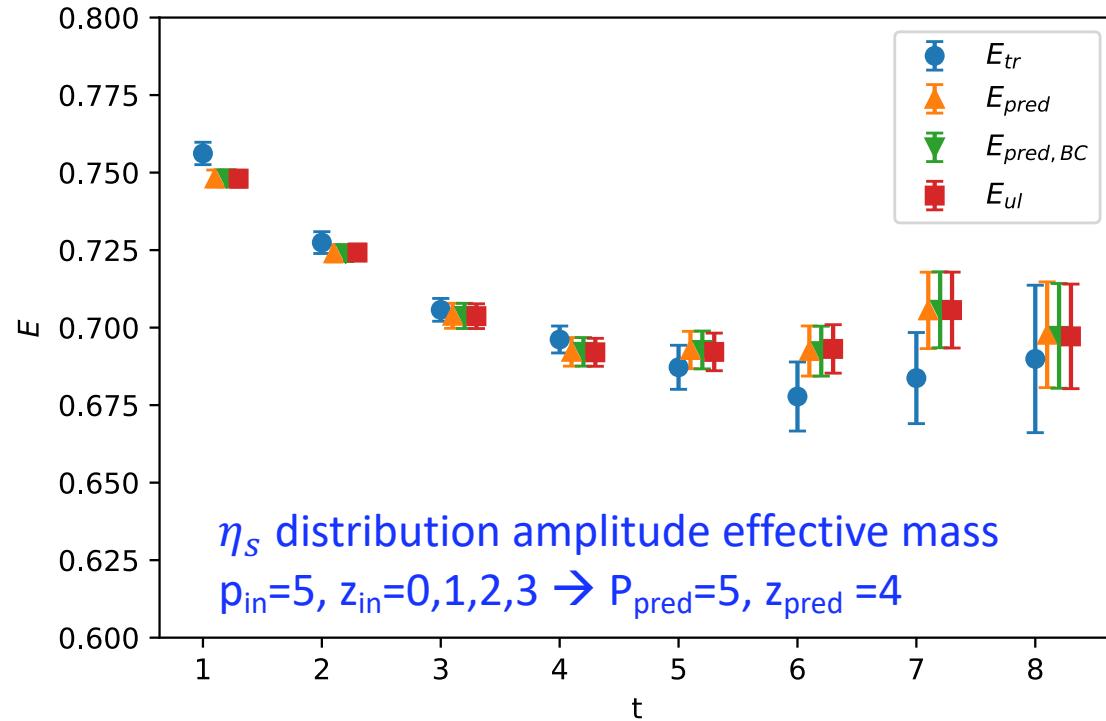
Prediction for Quasi-PDF Matrix Elements

- Large-momentum effective theory (LaMET) allows us to calculate Bjorken-x dependence of hadron structure on the lattice
- **Target observables**
 - Meson quasi-DAs (distribution amplitudes)

$$C_{2\text{pt}}(z, P, t) = \langle 0 | \int d^3y e^{i\vec{P}\cdot\vec{y}} \bar{\psi}_1(\vec{y}, t) \gamma_z \gamma_5 \prod_{x=0}^{z-1} U_z(y + x\hat{z}, t) \psi_2(\vec{y} + z\hat{z}, t) \bar{\psi}_2(0, 0) \gamma_5 \psi_1(0, 0) | 0 \rangle$$

- Meson quasi-PDFs (parton distribution functions)
- Gluon quasi-PDF
- **Predictions**
 - Prediction of higher boost momentum from the lower ones (*p*-prediction)
 - Prediction of longer link length from the lower ones (*z*-prediction)

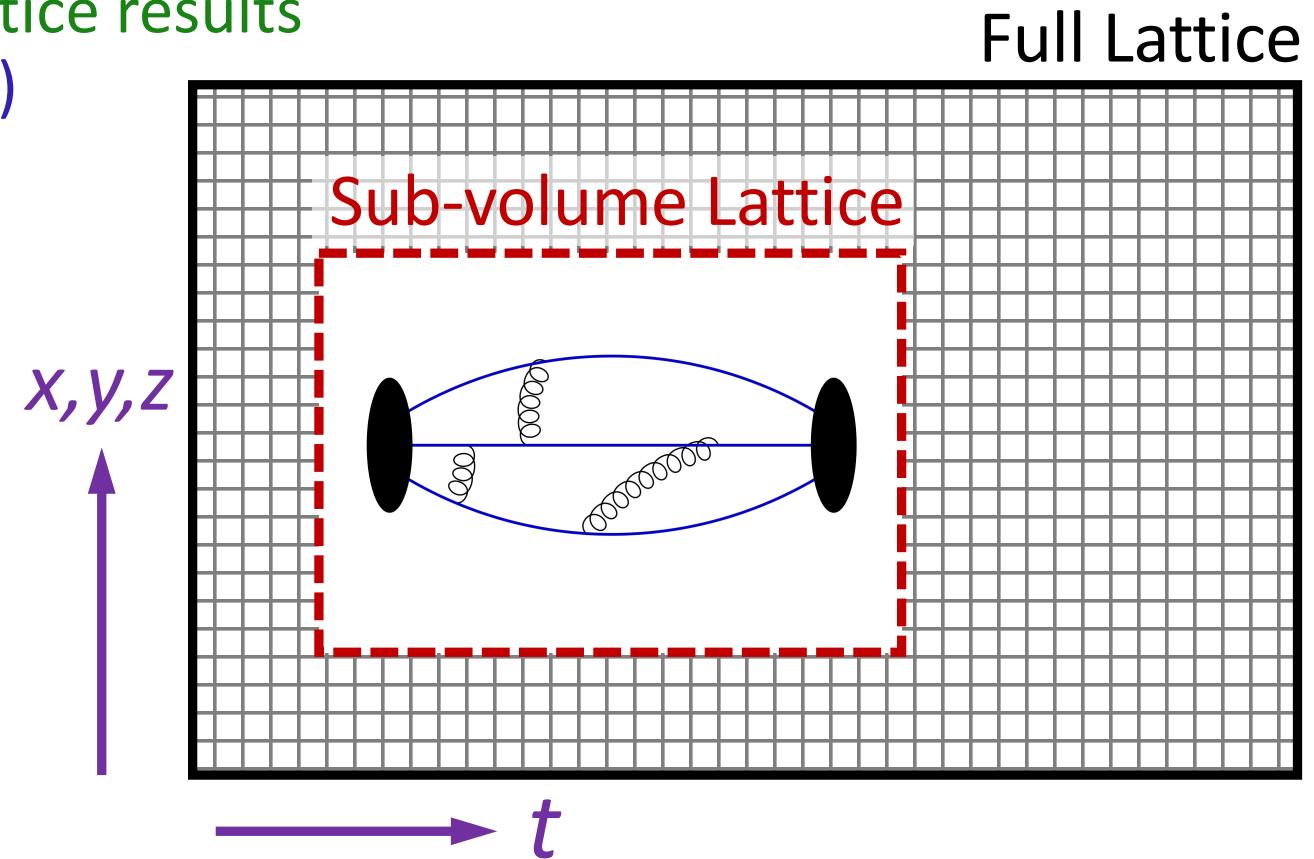
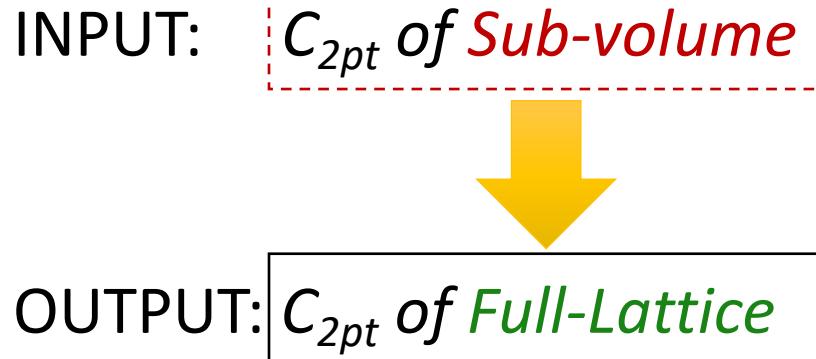
Prediction for Quasi-PDF Matrix Elements



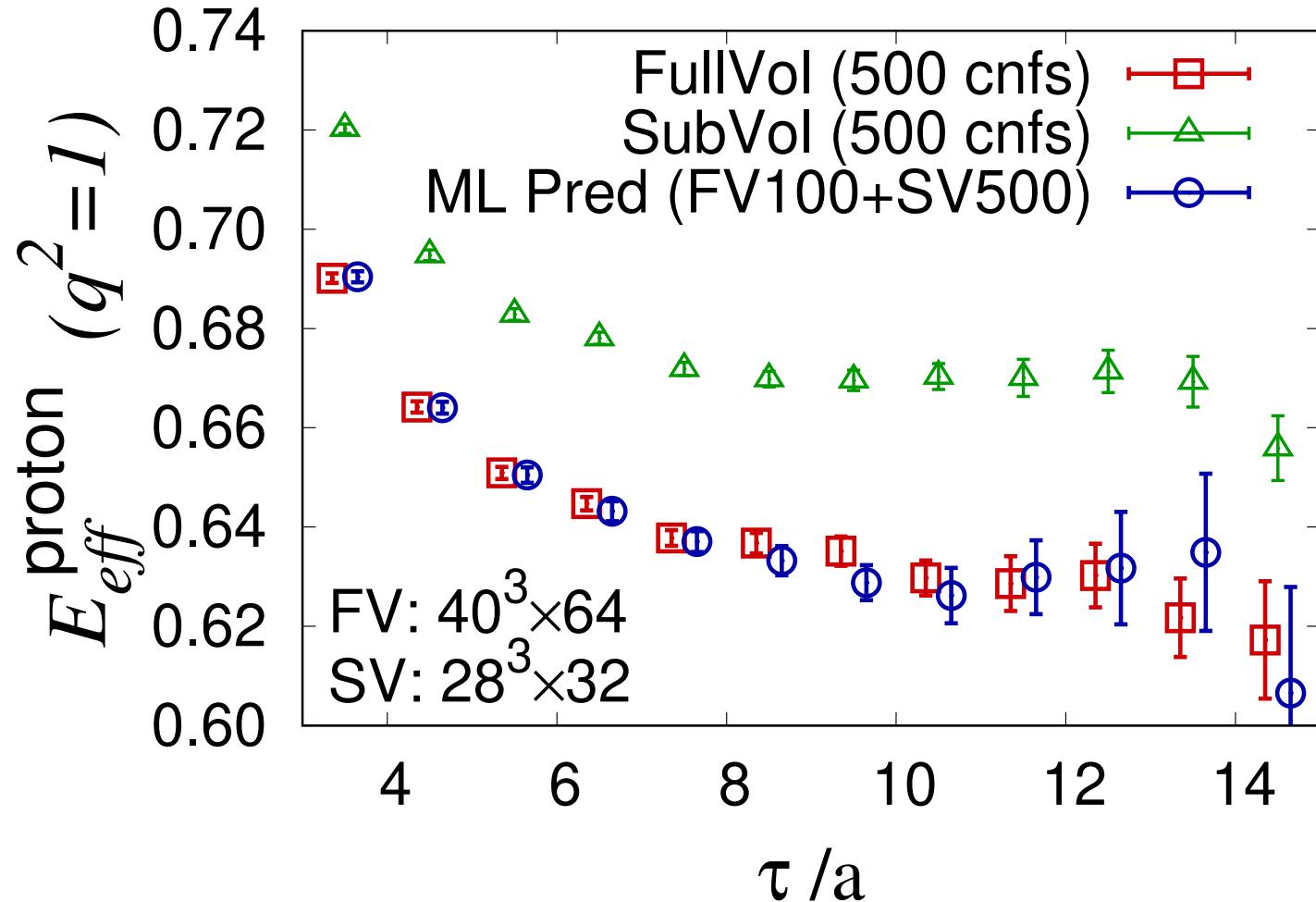
- **z -predictions** work much better than p -predictions
- Predictions for the **meson distribution amplitudes** and **quasi-PDFs** work better than those for gluon quasi-pdf

Calculation on Sub-volume Lattice

- 1) Take **sub-volume (SV)** lattice around C_{2pt} source and sink
- 2) Carry out all calculations (src/snk smearing, inversion, ...) on SV
- 3) Convert **SV** results to **full-lattice** results
using ML (Linear Regression)



Preliminary Study with Nucleon 2pt



- **a12m220L** ($40^3 \times 64$)
- **Reduced volume**
 $40^3 \times 64 \rightarrow 28^3 \times 32$
(reduced to $1/6$)
- **Better condition number**
(due to broken low-modes of
broken boundary condition)
BiCG iterations for LP (10^{-4})
 $613 \rightarrow 159$ (reduced to $1/4$)
- **Bias correction is essential**
Sub Vol results
→ Bias corrected results

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
 - 4) Prediction of full-volume from sub-volume calculation
- Developed a **new regression algorithm utilizing D-Wave quantum annealer** and showed promising prediction ability