LA-UR-21-27155

Prediction and compression of lattice QCD data using ML algorithms on quantum annealer

Boram Yoon

(Los Alamos National Laboratory)

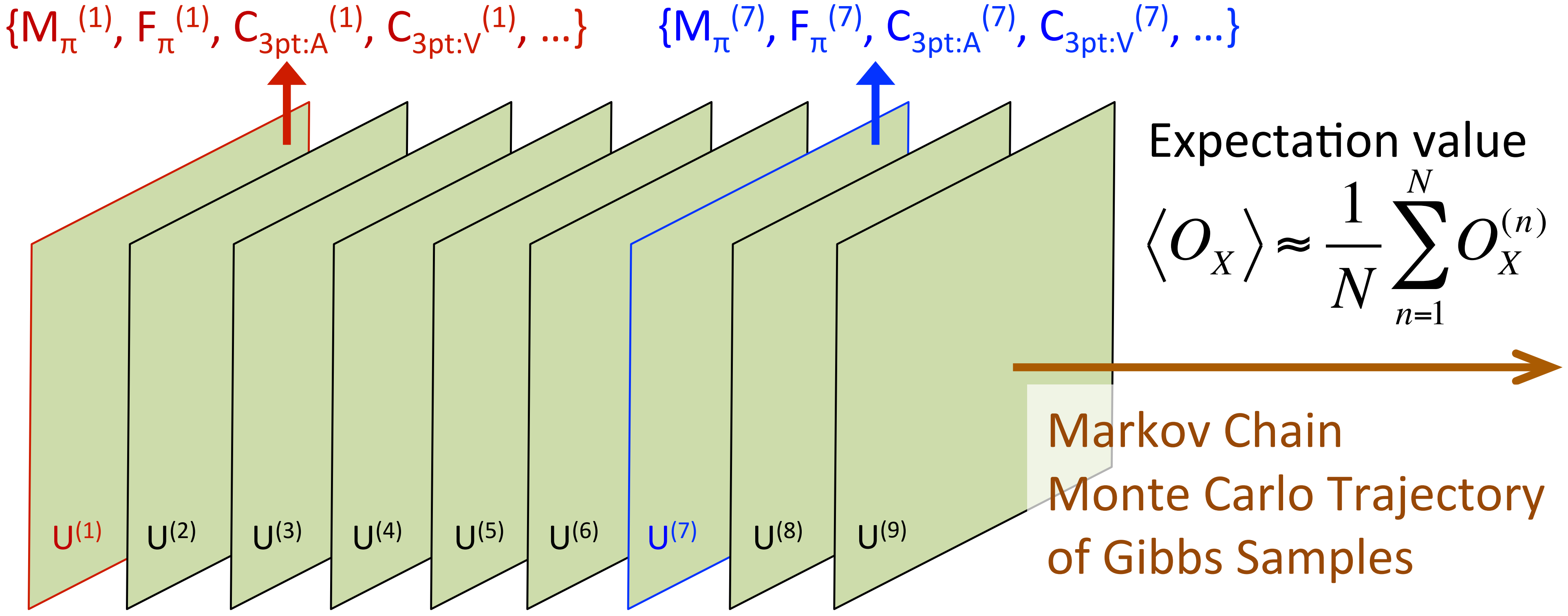
in collaboration with

Nga Nguyen (LANL), Garrett Kenyon (LANL),

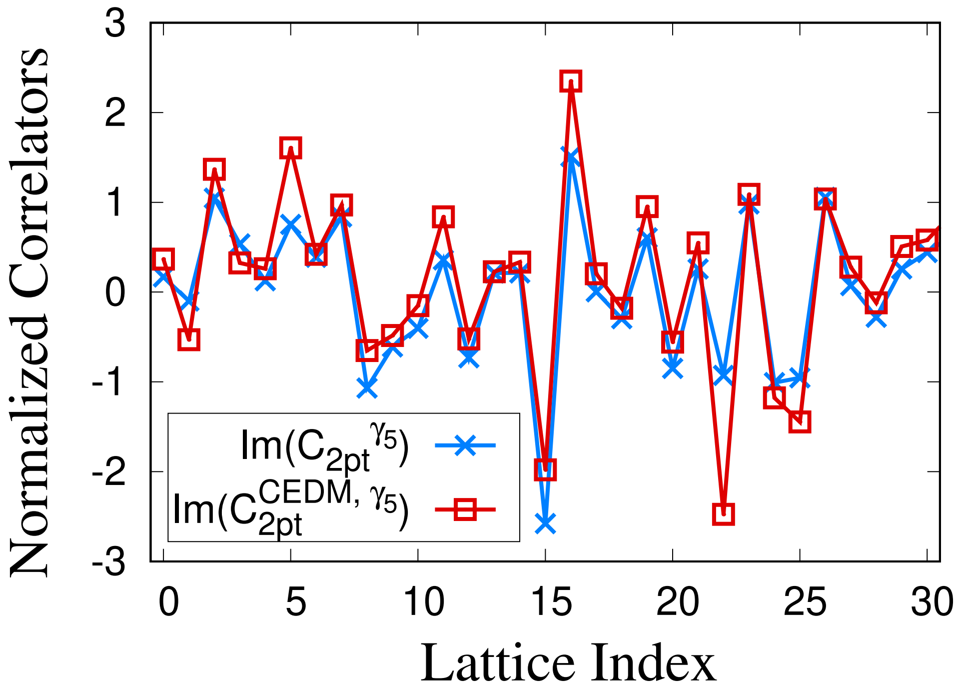
Chia (Jason) Cheng Chang (RIKEN, UC Berkeley, LBNL), Ermal Rrapaj (UC Berkeley)

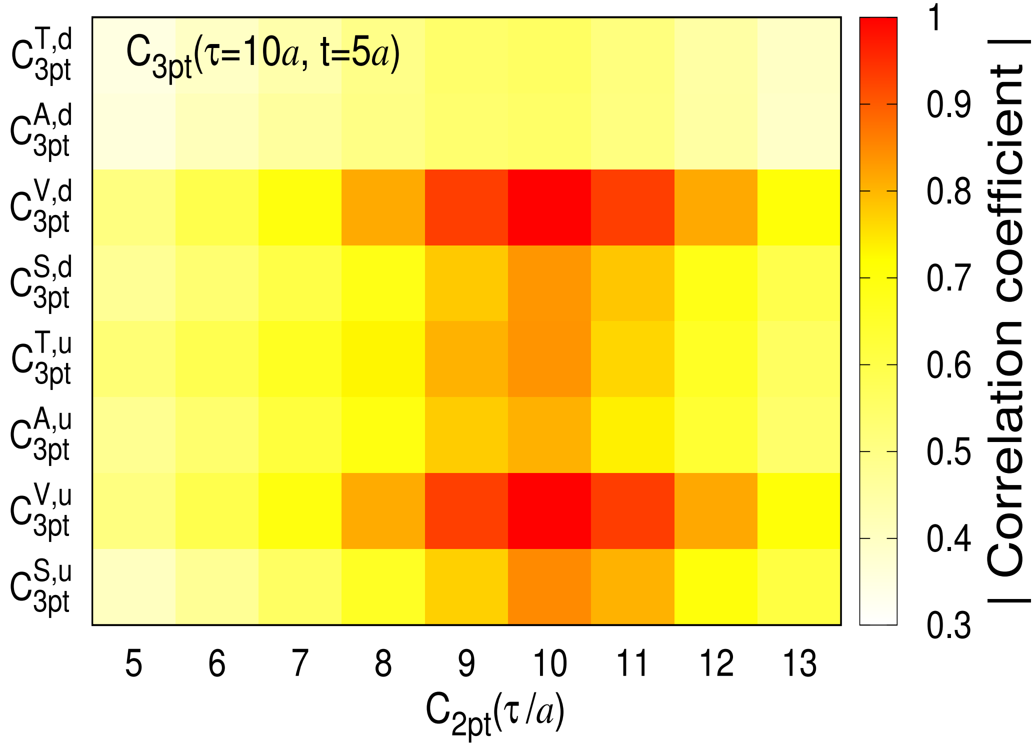
The 38th International Symposium on Lattice Field Theory, July 26-30 July 2021, ZOOM/GATHER@MIT

Lattice QCD Observables are Correlated



Lattice QCD Observables are Correlated



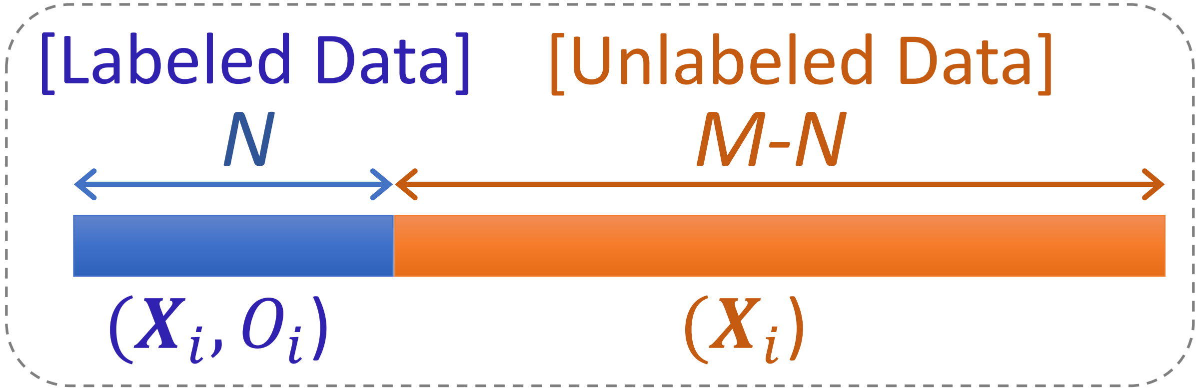


- Correlation between neutron 2-pt correlation function and that calculated in presence of CEDM interaction
- Correlation between proton 3-pt and 2-pt correlation functions

Using the correlation, values of the unmeasured (and expensive) observables can be predicted from the values of the measured (and cheap) observables

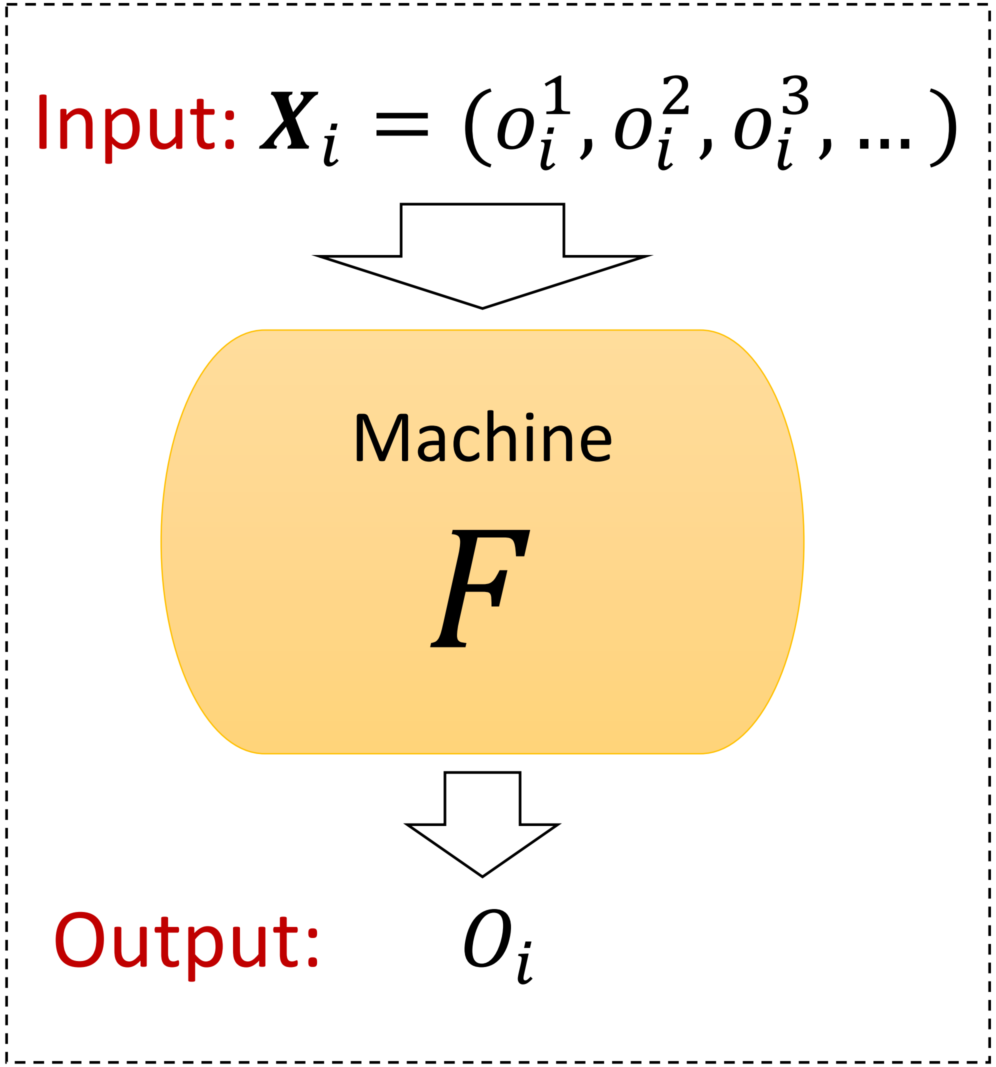
Prediction of Lattice QCD Observables using ML

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



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

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



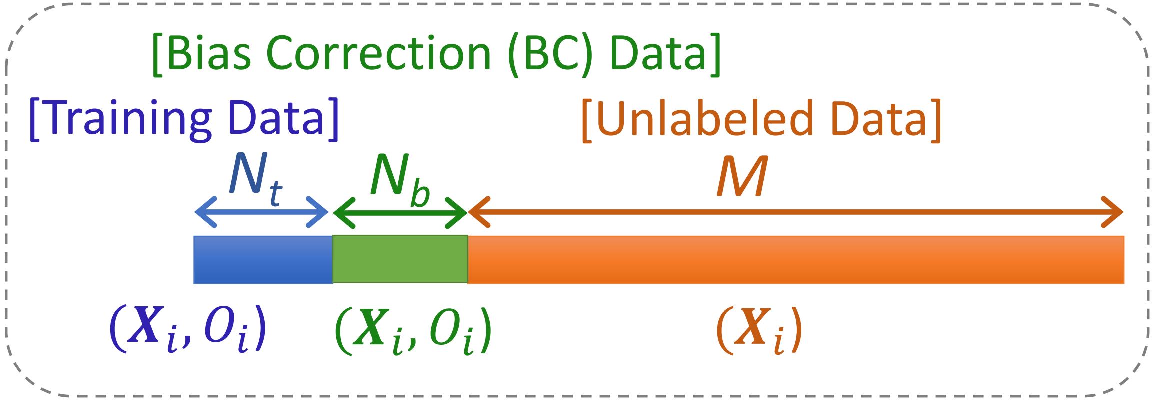
Error Quantification of Inexact ML Predictions

- ML predictions are not exact; introduces bias

$$\langle O^P \rangle = \frac{1}{N}\sum_i O_i^P \neq \langle O \rangle = \frac{1}{N}\sum_i O_i$$

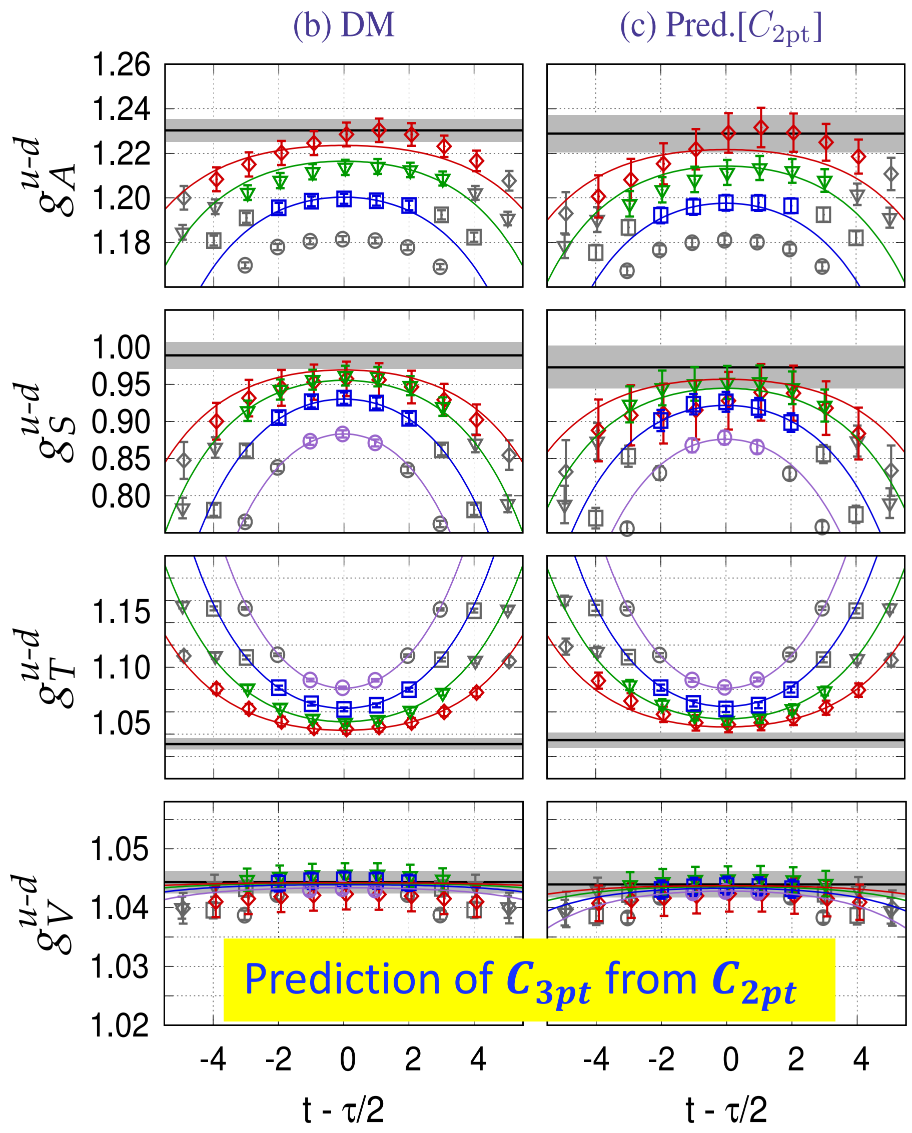
- Unbiased average using small portion of labeled data

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



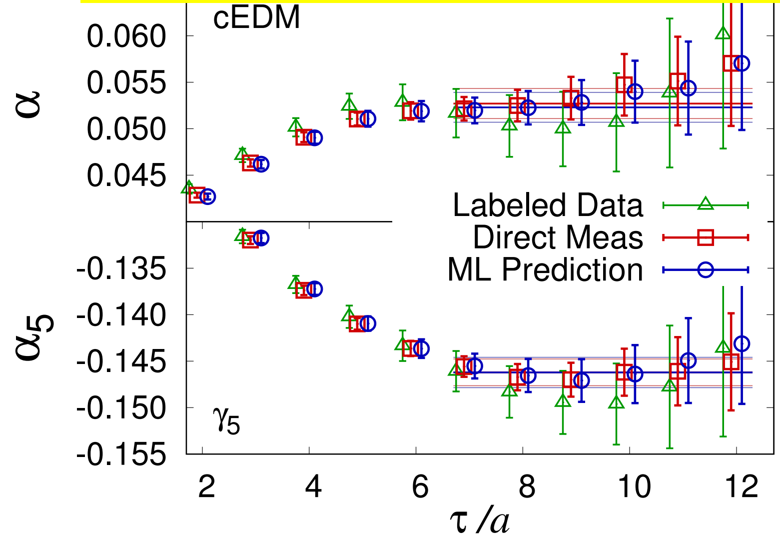
- Similar structure of truncated solver method (Bali, Collins, Schaefer, 2009) or all-mode averaging (Blum, Izubuchi, Shintani, 2012)
- Expectation value, $\langle \bar{O}_{BC} \rangle = \langle O_i^P \rangle + \langle O_i - O_i^P \rangle = \langle O_i \rangle$
- BC term converts systematic error of prediction to statistical uncertainty

Applications

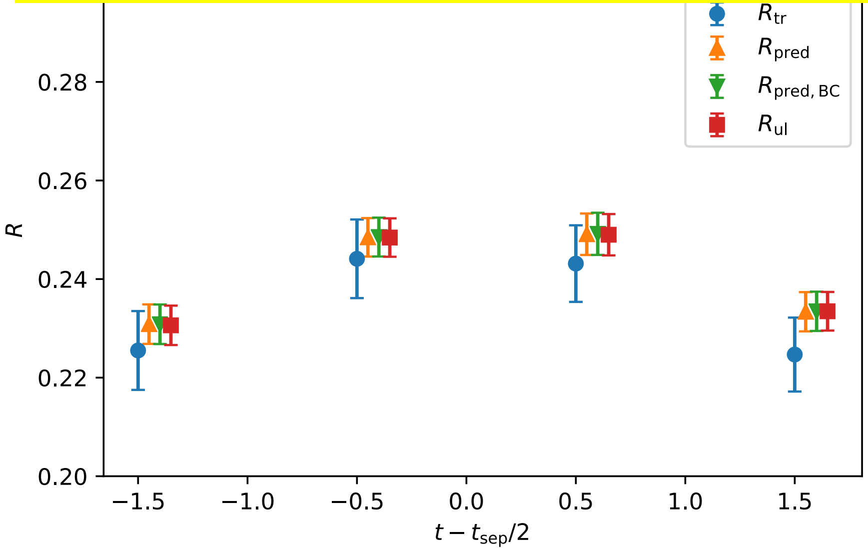


Prediction of C_{3pt} from C_{2pt}

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



Prediction of Kaon quasi-PDF $z = 4$ from $z < 4$



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

ML Regression using D-Wave Quantum Annealer

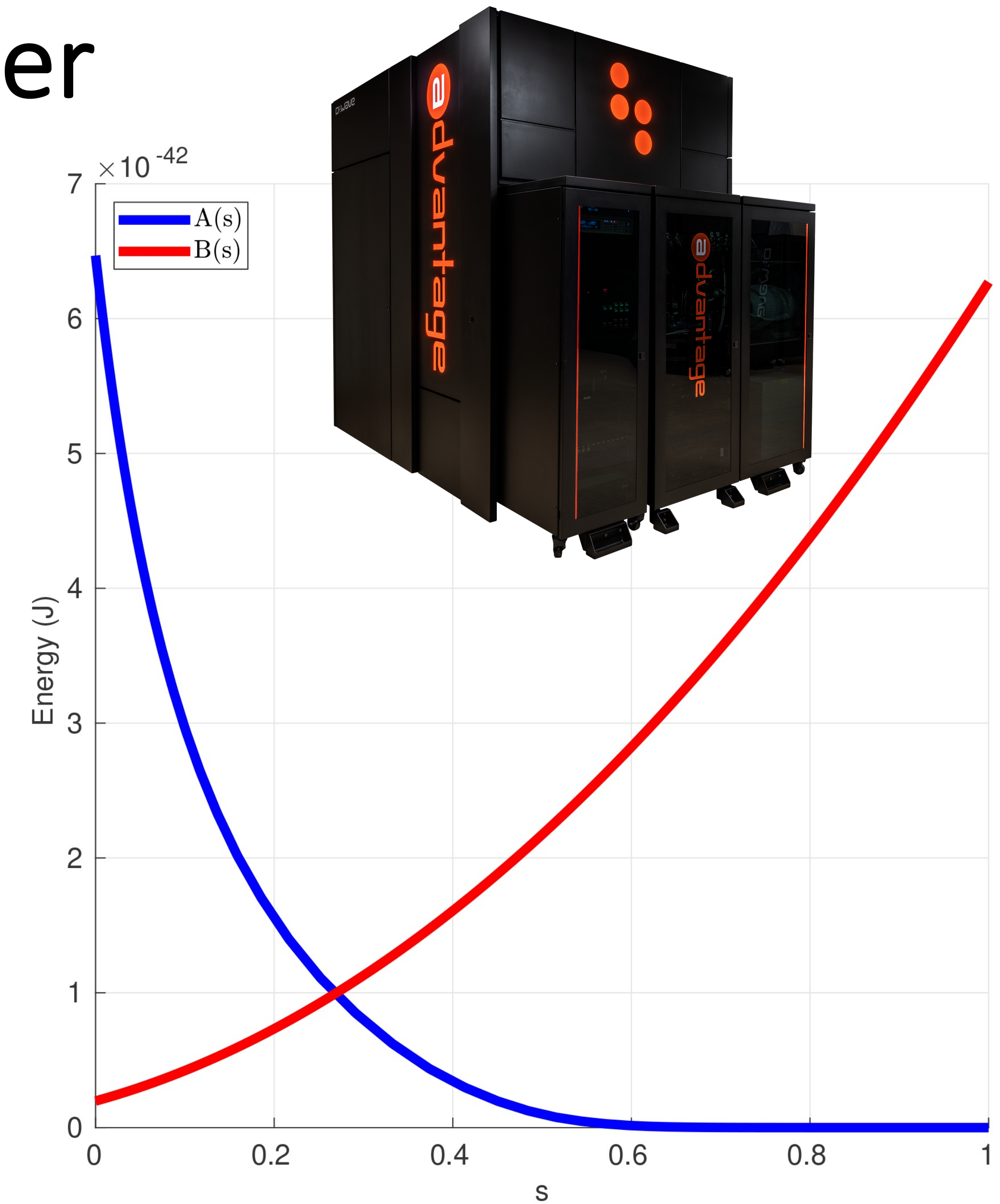
- Most ML algorithms involve optimization problems; many of them rely on stochastic approaches, but expensive for large problems
- D-Wave quantum annealer can be used as a fast and accurate optimizer for ML optimization problems

D-Wave Quantum Annealer

- Hamiltonian

$$H = -\frac{A(s)}{2}\left(\sum_i \hat{\sigma}_x^{(i)}\right) + \frac{B(s)}{2}\left(\sum_i h_i \hat{\sigma}_z^{(i)} + \sum_{i>j} J_{i,j}\hat{\sigma}_z^{(i)}\sigma_z^{(j)}\right)$$

- $h_i, J_{i,j}$: biases and coupling strengths that user can set to their problem parameters
- After annealing at < 15 mK, QPU returns low-energy solution (spin up/down of quantum bits) of the Ising model Hamiltonian
- Large number of reads is required to obtain minimum energy solution for large problems, but each read takes $O(10)\mu s$
- ML typically needs only near-optimal solution



Annealing Time

Sparse Coding

$$\min_{\Phi} \sum_{k=1}^{K} \min_{\vec{a}^{(k)}} \left[\frac{1}{2} \left\| \vec{X}^{(k)} - \boldsymbol{\Phi} \vec{a}^{(k)} \right\|_2 + \lambda \left\| \vec{a}^{(k)} \right\|_0 \right]$$

- Unsupervised ML algorithm
- Find dictionary $\boldsymbol{\Phi} \in \mathbb{R}^{D \times N_q}$ and sparse representation $\vec{a}^{(k)} \in \mathbb{R}^{N_q}$ from which input data $\vec{X}^{(k)} \in \mathbb{R}^D$ can be reconstructed by

$$\boxed{\vec{X}^{(k)} \approx \boldsymbol{\Phi} \vec{a}^{(k)}} = a_1^{(k)} \vec{v}_1 + a_2^{(k)} \vec{v}_2 + \cdots + a_1^{(k)} \vec{v}_1$$

- The representation is sparse because the λ-term enforces a minimal set of dictionary elements for the reconstruction of a given input data
- Optimization in $\vec{a}^{(k)}$ of l^0-norm function is a highly non-convex problem

Sparse Coding on D-Wave quantum annealer

$$\min_{\Phi} \sum_{k=1}^{K} \min_{\vec{a}^{(k)}} \left[\frac{1}{2} \left\| \vec{X}^{(k)} - \mathbf{\Phi} \vec{a}^{(k)} \right\|_2 + \lambda \left\| \vec{a}^{(k)} \right\|_0 \right]$$

- The sparse coding problem can be mapped onto D-Wave by

$$H(\vec{h}, \boldsymbol{Q}, \vec{a}) = \sum_i a_i h_i + \sum_{i<j} Q_{ij} a_i a_j$$

$$\vec{h} = -\mathbf{\Phi}^T \vec{X} + \left(\lambda + \frac{1}{2} \right), \qquad \boldsymbol{Q} = \frac{1}{2} \Phi^T \Phi$$

- On D-Wave, a_i is restricted to binary: $\vec{a}^{(k)} \in \{0,1\}^{N_q}$
- D-Wave finds $\vec{a}^{(k)}$ minimizing H
- Optimization for Φ is performed offline (on classical computers)

Inpainting

Nvidia AI Playground - Inpainting

Ground Truth

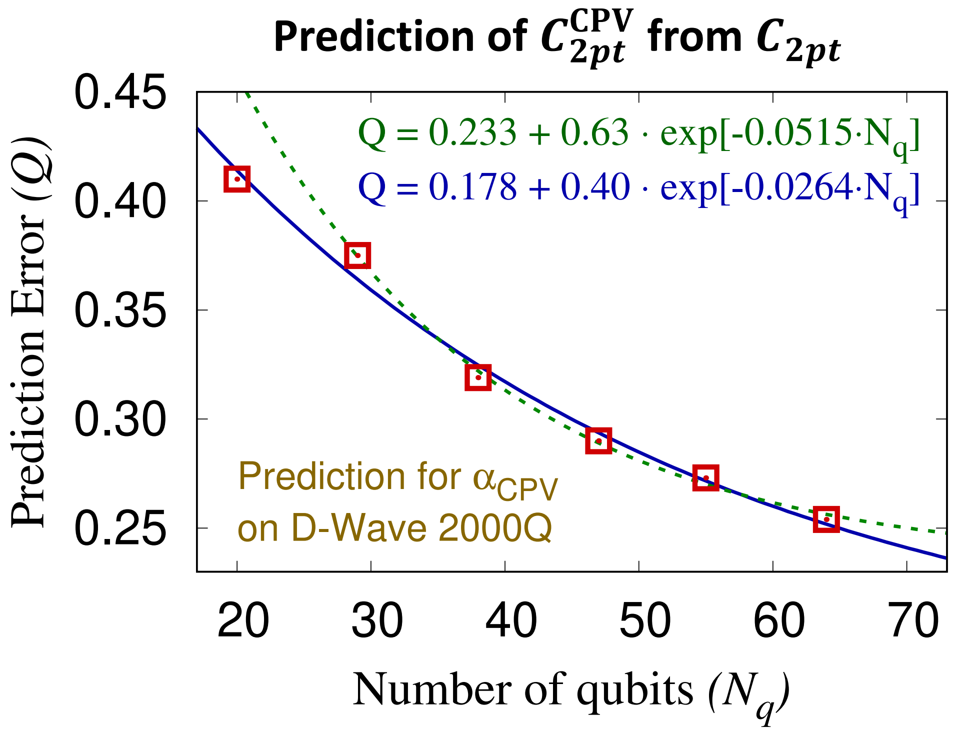
Data with Missing Pixels

Inpainted Results

- Inpainting: restorative conservation where damaged, deteriorating, or missing parts of an artwork are reconstructed as it was originally created
- Sparse coding works as an inpainting algorithm because the reconstruction $\vec{X}^{(k)} \approx \mathbf{\Phi}\vec{a}^{(k)}$ fills up the missing pixels based on the correlation pattern $\mathbf{\Phi}$ learned

Sparse Coding Regression on D-Wave

- Goal: prediction of y from $\vec{x} = \{x_1, x_2, \dots, x_D\}$
- Procedure:
  1) Obtain $\mathbf{\Phi_0} \in \mathbb{R}^{D \times N_q}$ of $\vec{x}$ from unlabeled data
  2) Extend $\mathbf{\Phi_0}$ to $\mathbf{\Phi} \in \mathbb{R}^{(D+1) \times N_q}$ and encode correlation between $\vec{x}$ and y in Φ using augmented vector $\{\vec{x}, y\}$
  3) For unknown y, reconstruct new vector $\{\vec{x}, \bar{y}\}$ using Φ; reconstruction replaces $\bar{y}$ with its prediction
- This approach is a semi-supervised learning as it utilizes unlabeled data to improve prediction
- D-Wave is used for optimization in $\vec{a}^{(k)}$



- Currently, the performance is limited by the maximum number of qubits available on D-Wave, but the predictions applied on lattice QCD data look promising

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

Lossy Data Compression for Lattice QCD

- Modern lattice QCD simulations produce O(PetaBytes) of data that need to be stored for future analysis
- Exploiting correlation between the data components can reduce storage requirement → **Machine learning**
- Reconstruction error sufficiently smaller than the observables statistical fluctuation is good enough for most of the analysis → **Lossy compression**

Lossy Data Compression Algorithm

- **Goal:** find $\mathbf{\Phi} \in \mathbb{R}^{D\times N_q}$ and $\vec{a}^{(k)} \in \{0,1\}^{N_q}$ precisely reconstructing input vectors $\vec{X}^{(k)} \in \mathbb{R}^D$ such that $\vec{X}^{(k)} \approx \mathbf{\Phi}\vec{a}^{(k)} \equiv \vec{X}'^{(k)}$
  - ➢ $\mathbf{\Phi}$ is common for all $k = 1,2,3, \ldots, N$, so memory usage is small
  - ➢ Each vector $\vec{a}^{(k)}$ can be stored in N_q bits
  - ➢ Storing $\left(\{\vec{a}^{(k)}\}_{k=1}^{N}, \mathbf{\Phi}\right)$ for $\{\vec{X}^{(k)}\}_{k=1}^{N}$: compression of D floating-point numbers into N_q bits
  - ➢ Correlation between X_i, encoded in $\mathbf{\Phi}$, allows precise reconstruction with $N_q \ll 32D$

- Such solutions of $\mathbf{\Phi}$ and $\vec{a}^{(k)}$ can be obtained by solving

$$\min_{\mathbf{\Phi}} \sum_{k=1}^{N} \min_{\vec{a}^{(k)}} \left[\left(\vec{X}^{(k)} - \mathbf{\Phi}\vec{a}^{(k)}\right)^2\right]$$

  - ➢ **Finding binary solution of $\vec{a}^{(k)}$ is an NP-hard problem but can be solved using D-Wave**
  - ➢ Finding $\mathbf{\Phi}$ is done on classical computers with stochastic optimizer
  - ➢ Iterate $\vec{a}^{(k)}$- and $\mathbf{\Phi}$-optimizations until it reaches the minimum reconstruction error
  - ➢ Need standardization of $\vec{X}^{(k)}$ beforehand if the data exhibits heteroskedasticity

Bias Correction of Lossy Reconstruction

- Lossy reconstruction introduces error $\vec{X}^{(k)} \neq \boldsymbol{\Phi}\vec{a}^{(k)} \equiv \vec{X}'^{(k)}$
  Simple average is a biased estimator $\langle f(\vec{X}) \rangle \neq \frac{1}{N}\sum_k f\left(\vec{X}'^{(k)}\right)$
- Unbiased estimator of $\langle f(\vec{X}) \rangle$ can be defined using small portion of original data

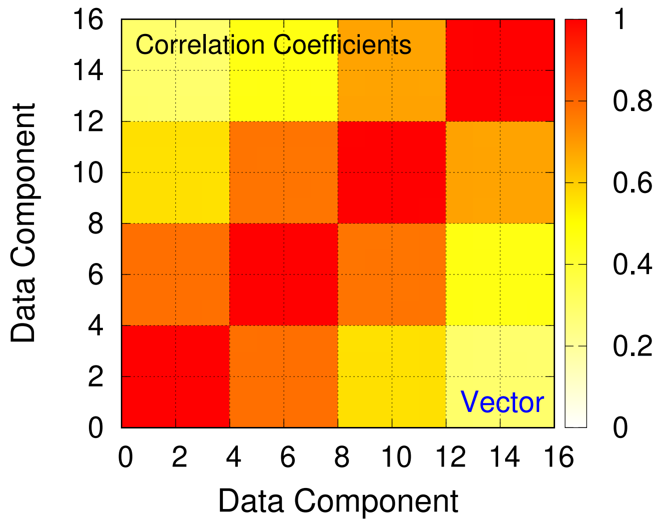
$$\bar{O} = \frac{1}{N}\sum_{k=1}^{N} f\left(\vec{X}'^{(k)}\right) + \frac{1}{N_{bc}}\sum_{k=1}^{N_{bc}}\left(f\left(\vec{X}^{(k)}\right) - f\left(\vec{X}'^{(k)}\right)\right)$$

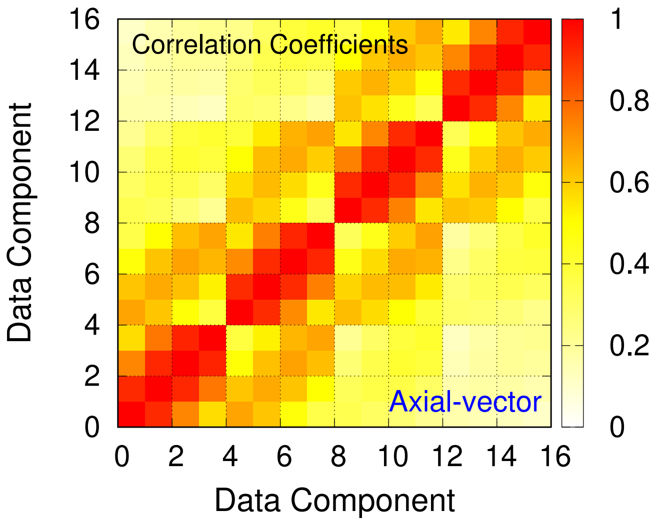
- Quality of lossy-compression on statistical data

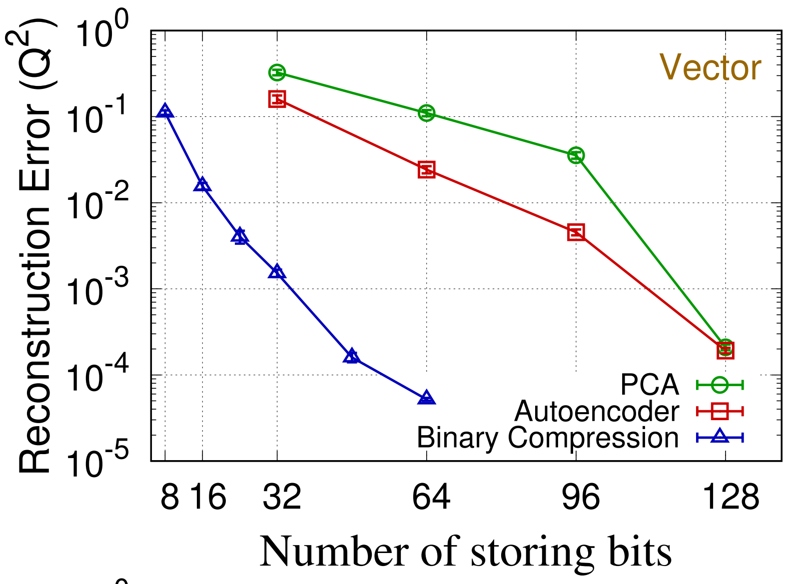
$$Q^2 \equiv \frac{1}{D}\sum_{i=1}^{D}\frac{\sigma^2_{X_i - X'_i}}{\sigma^2_{X_i}}$$

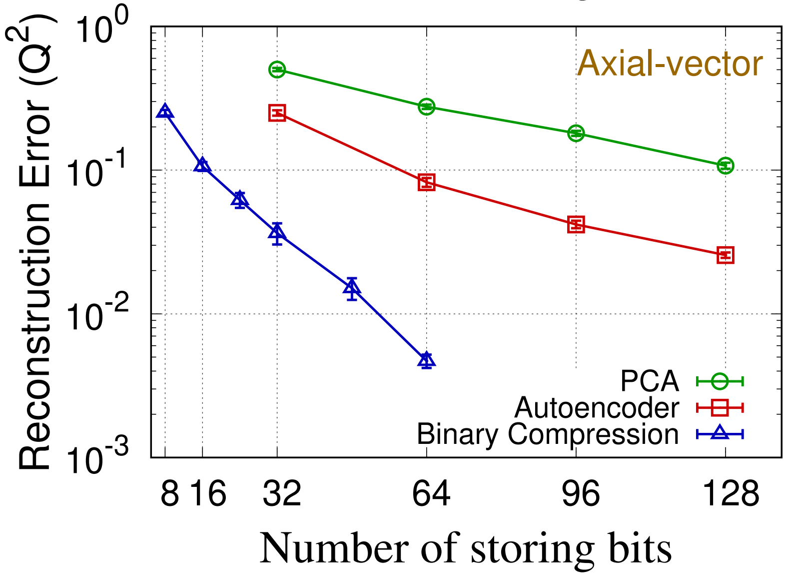
  - Smaller Q^2 indicates the better compression
  - Increase of statistical error due to bias correction is proportional to $\frac{N}{2N_{bc}}Q^2$
  - eg) With 10% of bias correction data (N_{bc}/N=0.1) and compression of $Q^2 = 0.01$, original data is typically reconstructed within 5% statical error increase

Compression of Lattice QCD data









- Compression of "*4 timeslices ×4 src-sink separations*" of vector and axial-vector nucleon 3pt correlators
- Compression performance of the new algorithm outperforms those based on principal component analysis (PCA) or neural-network autoencoder
- Results from D-Wave simulated annealing; real QPU gives worse performance due to noise in *h* and *J* parameters
- PCA and NN-Autoencoder with single-precision (32bits) codes

BY, Nga Nguyen, Jason Chang, Chia Cheng Chang, Ermal Rrapaj, will appear on arXiv soon

More Use of Binary Compression Algorithm

- **Outlier detection**
  - An input data with large reconstruction error can be marked anomalous
  - Could find events of new physics or data corruption
- **Cheaper operations in $\vec{a}$-space ($\vec{X}^{(k)} \approx \mathbf{\Phi}\vec{a}^{(k)}$)**
  - Operations on floating-point numbers $\vec{X}^{(k)}$ can be replaced by those on single-bit coefficients $\vec{a}^{(k)}$ with much cheaper computational cost
  - eg 1) sum of vectors $$\sum_{k=1}^{N} \mathbf{X}^{(k)} \approx \sum_{k=1}^{N} \boldsymbol{\phi} \boldsymbol{a}^{(k)} = \boldsymbol{\phi} \left(\sum_{k=1}^{N} \boldsymbol{a}^{(k)} \right)$$
  - eg 2) sum of l^2-norm squares

$$\sum_{k=1}^{N} ||\mathbf{X}^{(k)}||^2 \approx \sum_{k=1}^{N} \sum_{i=1}^{D} \left(\sum_{j=1}^{N_q} \phi_{ij} a_j^{(k)} \right)^2$$

$$= \sum_{i=1}^{D} \left[\sum_{j=1}^{N_q} \phi_{ij}^2 \left(\sum_{k=1}^{N} a_j^{(k)} \right) + 2 \sum_{l<m} \left(\sum_{k=1}^{N} a_l^{(k)} a_m^{(k)} \right) \phi_{il} \phi_{im} \right]$$

Summary

- Developed a new regression algorithm utilizing D-Wave quantum annealer and showed promising results in predicting unmeasured lattice QCD observables
- Developed a new data compression algorithm utilizing D-Wave quantum annealer and showed a good performance in compressing lattice QCD observables