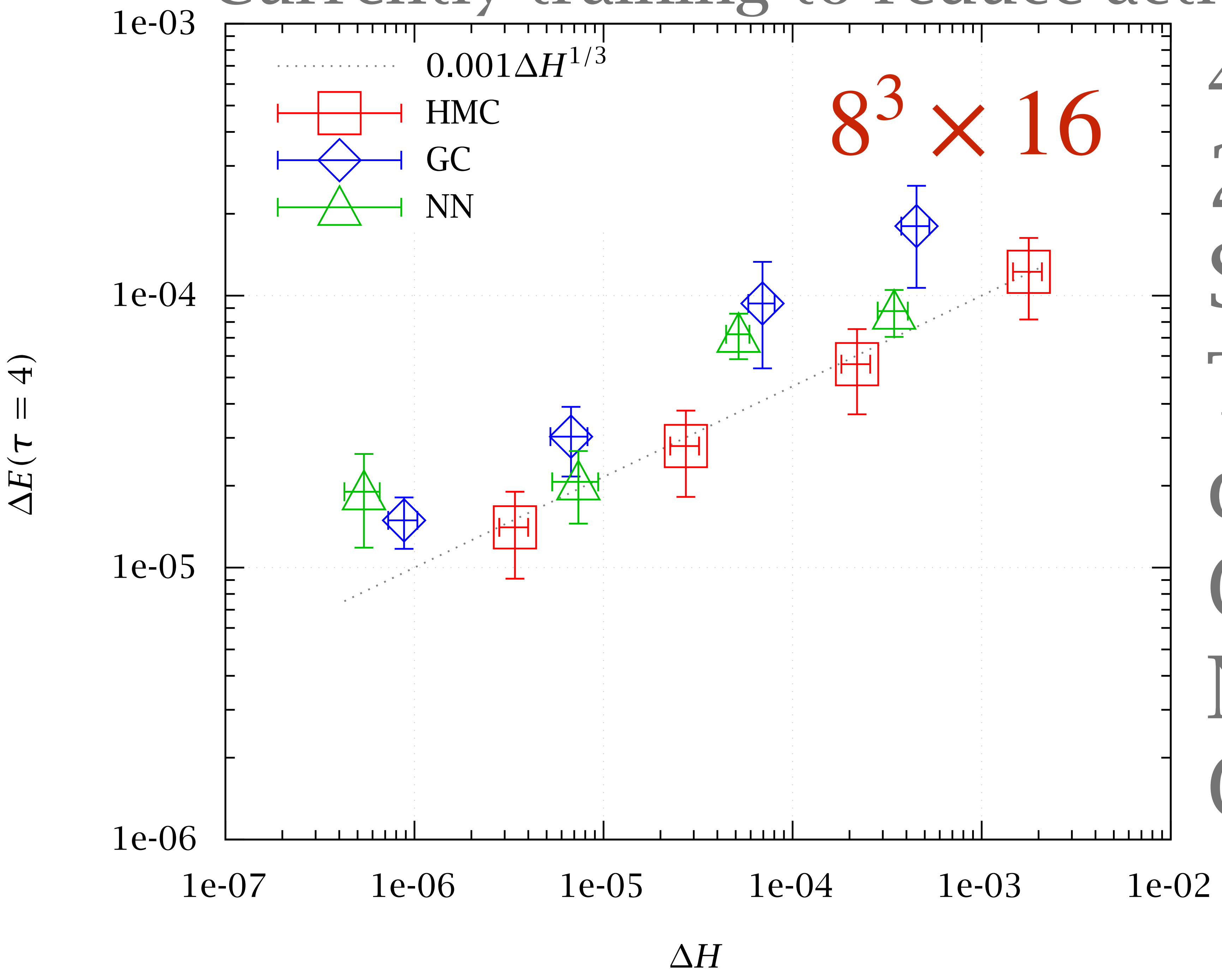
HMC with change of variables using ML and scaling up

Xiao-Yong Jin
CPS

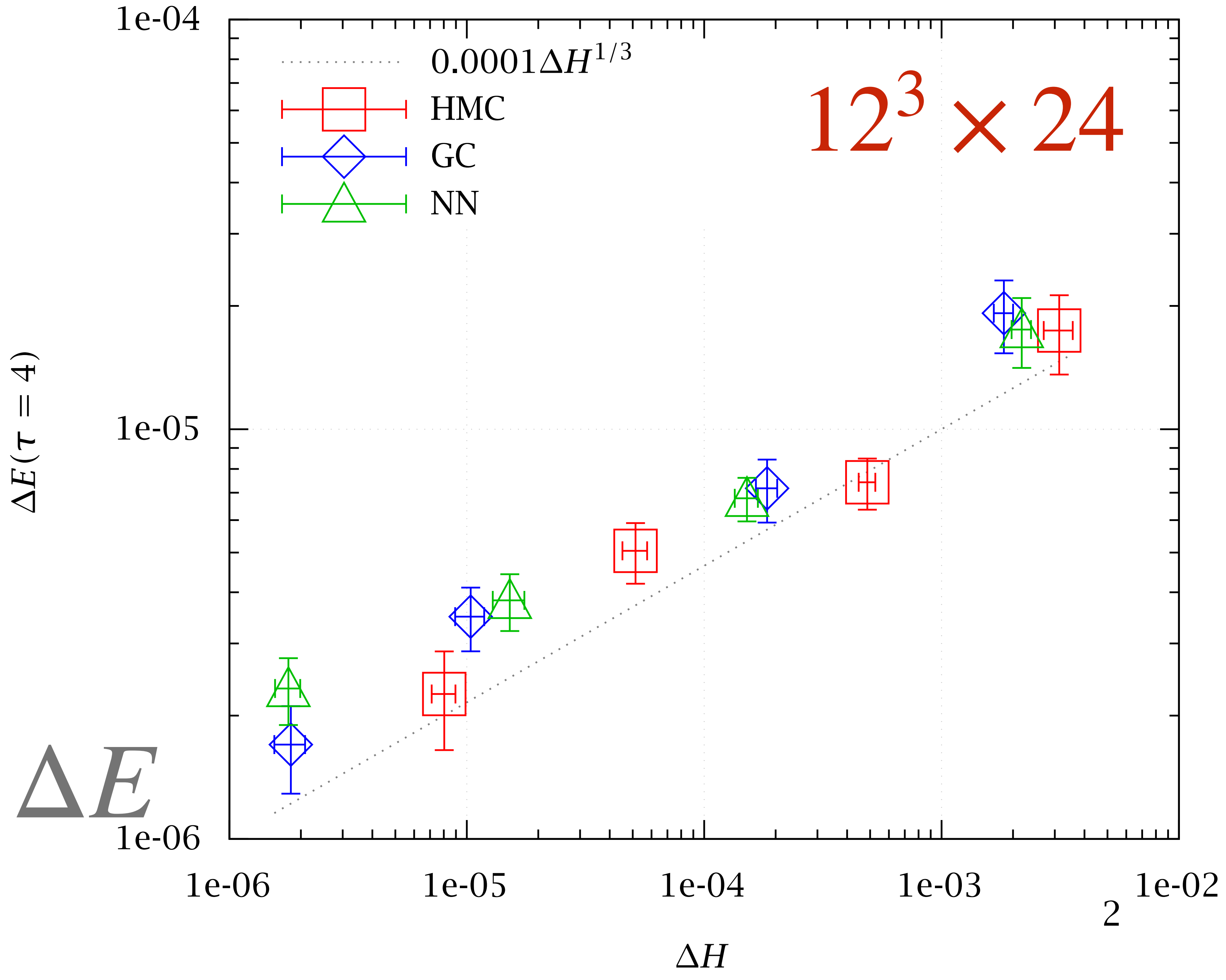
August 17, 2023
Data-Intensive Computing and AI/ML Applications at Scale

Change of variables

- Continuously differentiable bijective map $\mathscr{F}^{-1}$ from target field U to the mapped field $V = \mathscr{F}^{-1}(U)$
  $\langle \mathscr{O} \rangle = \frac{1}{Z}\int \mathscr{D}U\mathscr{O}\big(U\big)e^{-S(U)} = \frac{1}{Z}\int \mathscr{D}V\mathscr{O}\big(\mathscr{F}(V)\big)e^{-S(\mathscr{F}(V))+\ln|\mathscr{F}_*|}$ where $\mathscr{F}_* = \frac{\partial \mathscr{F}(V)}{\partial V}$
- Sample V with HMC according to the new action: $S_{\mathrm{FT}}\big(V\big) = S\big(\mathscr{F}(V)\big) - \ln|\mathscr{F}_*(V)|$
- Want the effective action to have lower potential barriers, more uniform dynamics
- The Jacobian determinant and its derivative must remain simple
- Currently training to reduce action derivatives by a fraction



4D SU(3) lattice QCD
2 different lattice sizes
Same models: GC, NN
Trained with $8^3 \times 16$
on a single A100
GC: 3 hours training
NN: 11 hours training
Goal: small ΔH while large ΔE



- **Scaling up**
  - $8^3 \times 16$, model: < 1 MB, gauge field: 4.5 MB, train: 30 GB, inference: 10 GB
  - $12^3 \times 24$, above times 5
  - Production volumes: $\sim O(10^4)$ times larger
  - Objectives
    - Better memory use when taking derivatives
    - More efficient 4D shifts and convs
    - Distribute the model
    - More effective network architecture
    - More effective loss function
    - Target subvolume of the whole lattice

- **Possible applications or extensions**
  - Other systems on a grid that would benefit from a change of variables?
  - Neural network architecture that are easier to compute Jacobian and derivatives of Jacobian

- **Code and more**
  - https://github.com/nftqcd/nthmc using TensorFlow
  - Previous toy model, arXiv:2201.01862
  - Recent talk, NN gauge field transformation @lattice2023