Machine-learning techniques as noise reduction strategies in lattice calculations of the muon $g-2$

Tom Blum, Alessandro Conigli, Lukas Geyer, Simon Kuberski, Alexander Segner, Hartmut Wittig

The 41st International Symposium on Lattice Field Theory — Lattice 2024
The University of Liverpool
28 July to 3 August 2024

HIM
Helmholtz-Institut Mainz

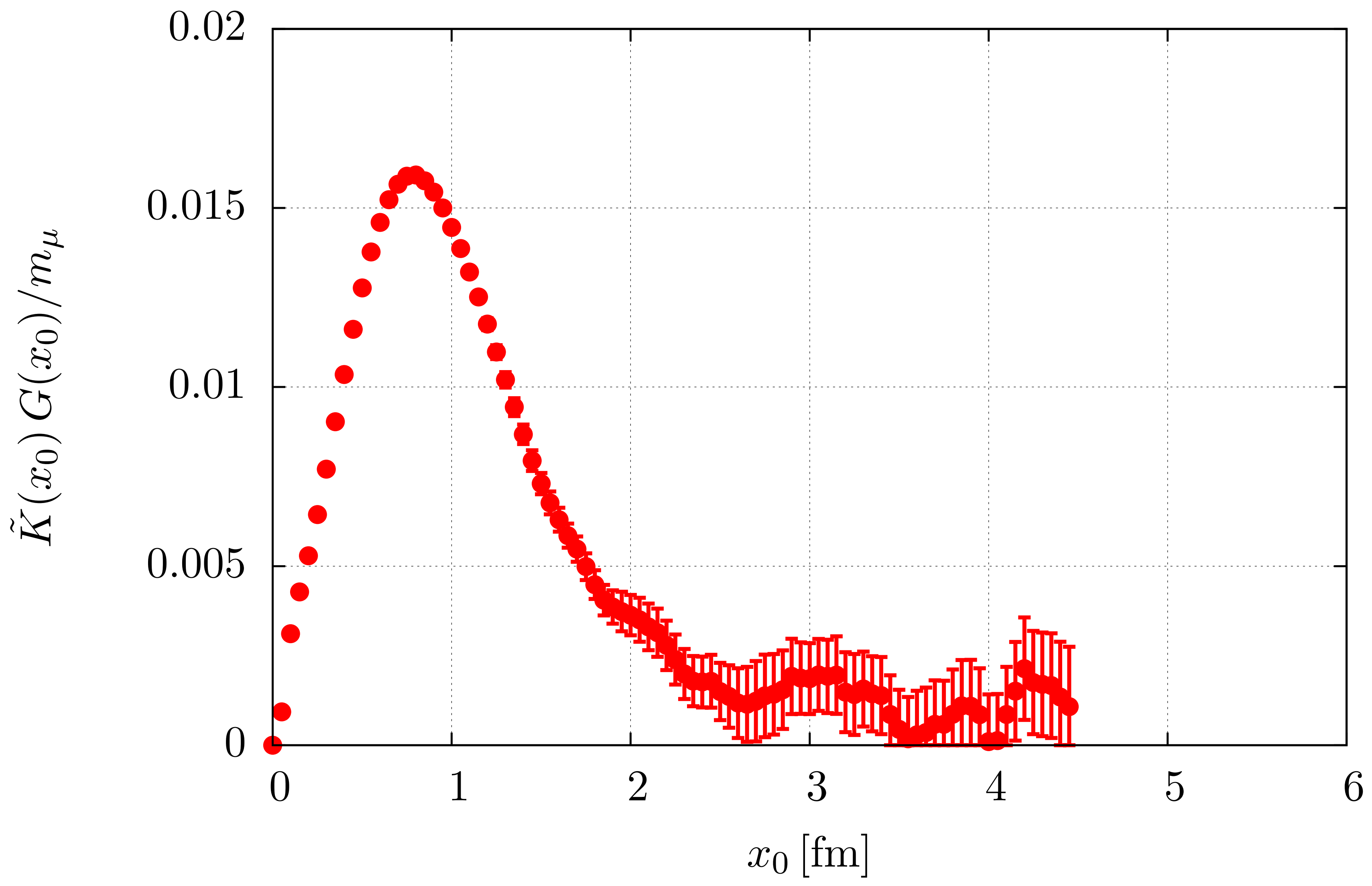
PRiSMA+

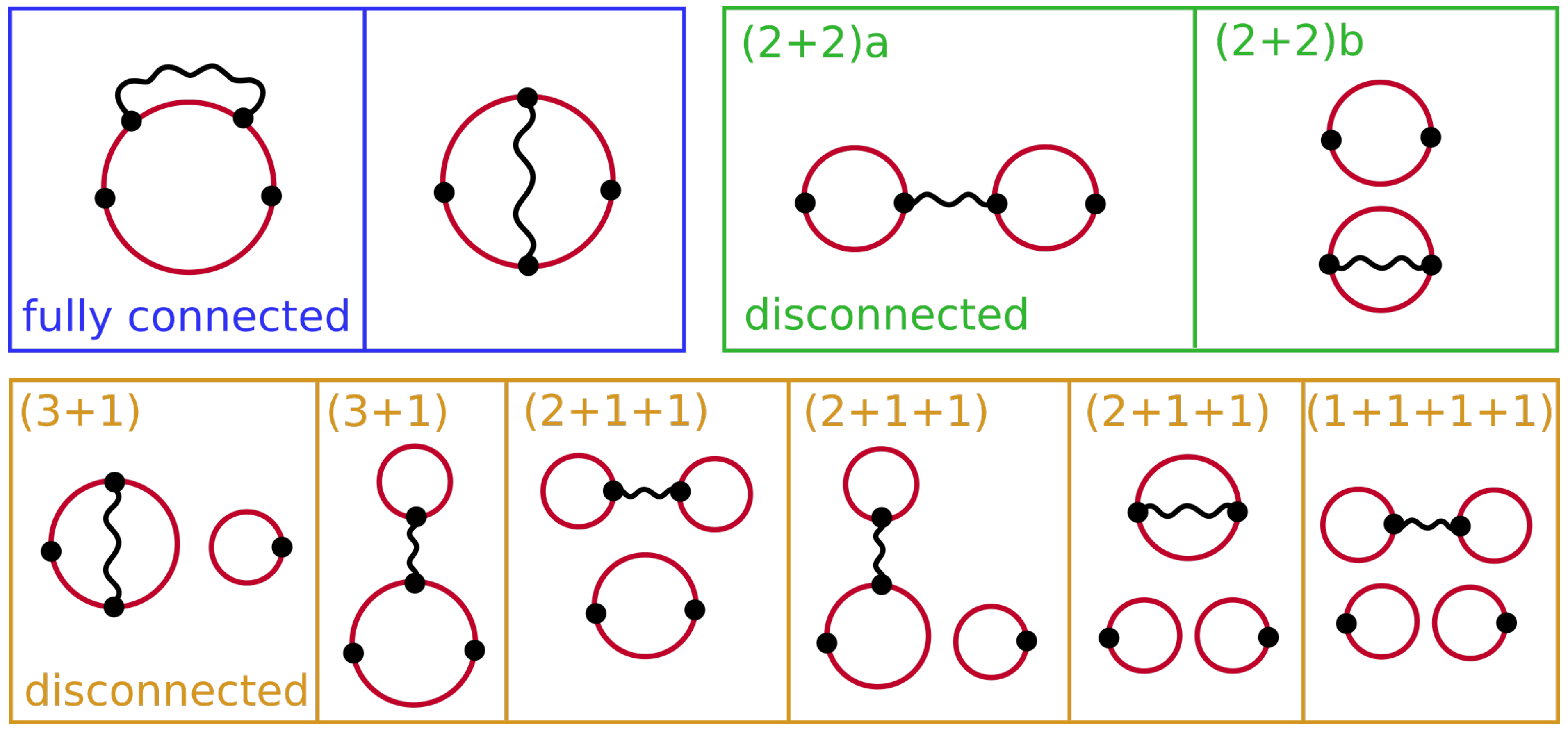
JOHANNES GUTENBERG UNIVERSITÄT MAINZ
JG|U

Motivation

Lattice QCD calculations of precision observables involve high numerical cost:

- Exponential rise of statistical noise at large distances
- Isospin-breaking corrections numerically small but expensive to compute





[©D. Erb, J. Parrino]

Idea: use trained neural net to

- Produce approximate estimates for correlation functions at low numerical cost
- Predict "expensive" contribution using a "cheap" observable

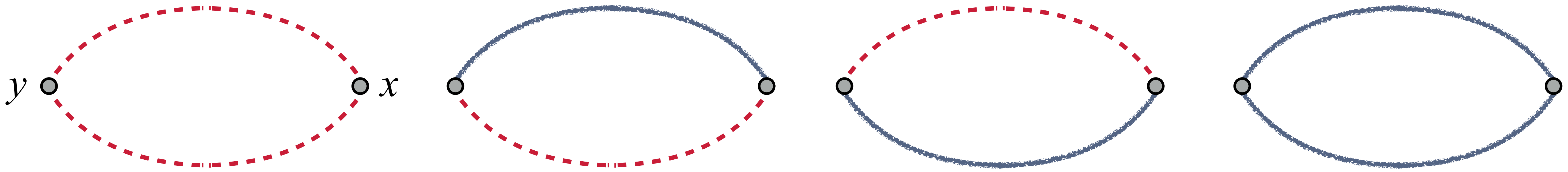
Low-mode averaging

[Giusti, Hernández, Weisz, H.W. 2004; DeGrand and Schaefer 2004]

Split quark propagator into low-mode contribution and that from orthogonal complement

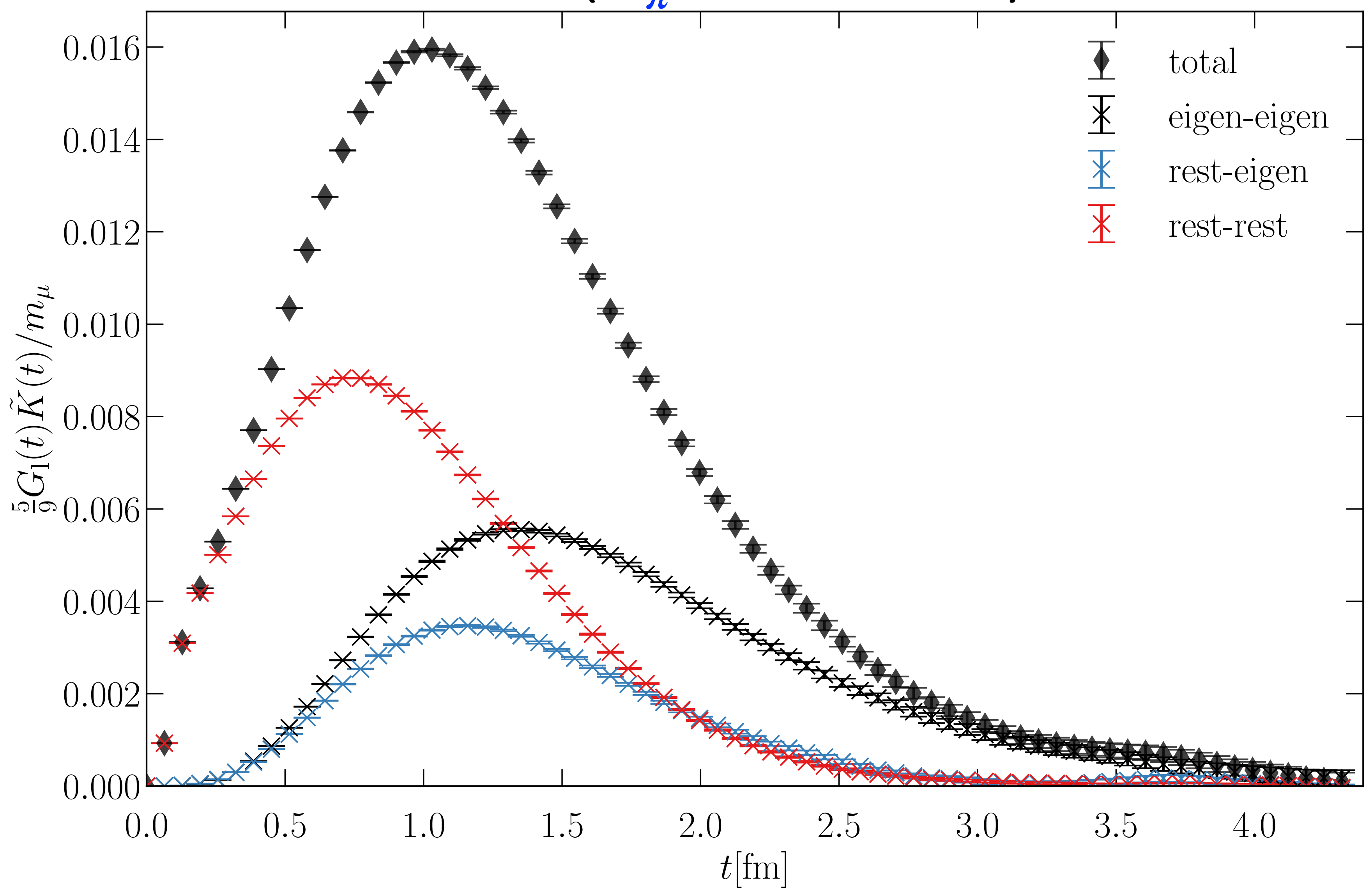
$$S(y,x) = S_{\rm eigen}(y,x) + S_{\rm rest}(y,x) \qquad S_{\rm eigen}(y,x) = \sum_{i=1}^{N_{\rm low}} \lambda_i^{-1}\, v_i(x) \otimes (\gamma_5 v_i(y))^\dagger\,, \qquad (\gamma_5 D_{\rm w})\, v_i(x) = \lambda_i\, v_i(x)$$

Two-point correlation function:



"eigen-eigen" "rest-eigen" "rest-rest"

Ensemble D452 ($m_\pi = 155\,\mathrm{MeV}$)



- "eigen-eigen" part dominates long-distance regime
- "rest-eigen" part always sub-dominant, but numerical effort scales $\sim N_{\rm low} = \mathrm{O}(1000)$

Machine-learning strategy

All-Mode Averaging / Truncated Solver Method: *[Blum, Izubuchi, Shintani 2012]*

Compute many approximate solutions — obtain exact result after applying bias correction

Idea: Role of "sloppy solves" taken over by a machine-learning algorithm *[Yoon, Bhattacharya, Gupta 2018]*

$$\underbrace{\langle O \rangle}_{\text{exact}} = \underbrace{\left\langle O_{\mathrm{appx}} \right\rangle}_{\text{“cheap”}} + \underbrace{\left\langle (O - O_{\mathrm{appx}}) \right\rangle}_{\text{correction}}$$

Train model or network on the correlation between input and predicted quantities

Ideally O_{appx} should fluctuate closely with O

Machine-learning strategy

All-Mode Averaging / Truncated Solver Method: *[Blum, Izubuchi, Shintani 2012]*

Compute many approximate solutions — obtain exact result after applying bias correction

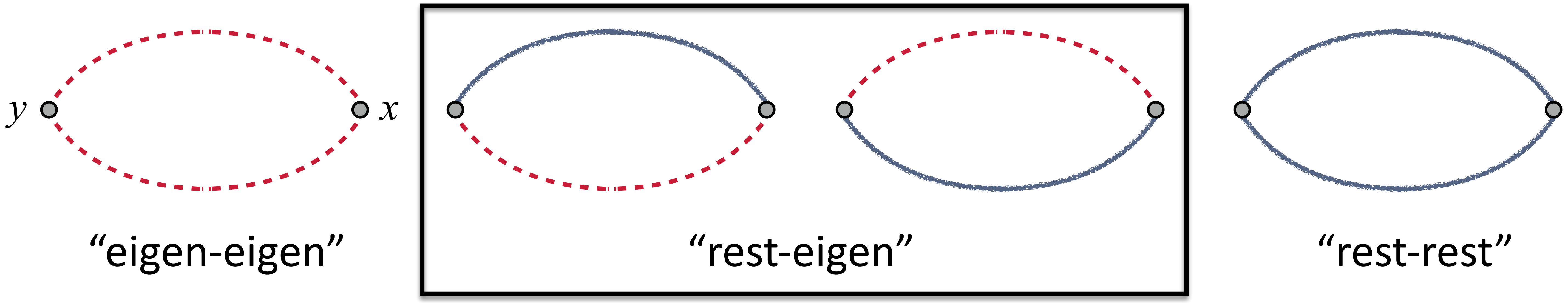
Idea: Role of “sloppy solves” taken over by a machine-learning algorithm *[Yoon, Bhattacharya, Gupta 2018]*

$$\underset{\substack{\uparrow \\ \text{exact}}}{\langle O\rangle} = \underset{\substack{\uparrow \\ \text{“cheap”}}}{\left\langle O_{\mathrm{appx}}\right\rangle} + \left\langle\left(O - \underset{\substack{\uparrow \\ \text{correction}}}{O_{\mathrm{appx}}}\right)\right\rangle$$

Train model or network on the correlation between input and predicted quantities

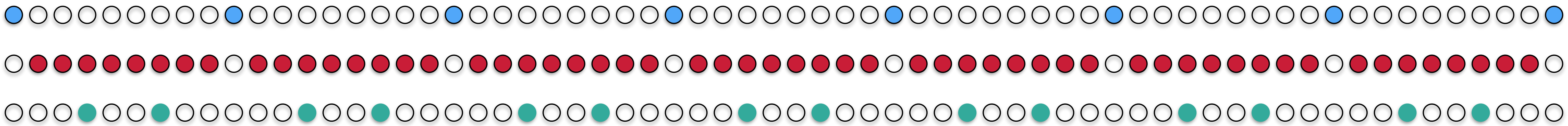
Ideally O_{appx} should fluctuate closely with O

Here: predict “rest-eigen” contribution given the “eigen-eigen” and “rest-rest” as input



Setup

Divide configurations within one ensemble into subsets for training, prediction and bias correction



Fully connected neural network with one hidden layer to predict all timeslices simultaneously:

- ReLU activation functions on hidden layer, linear activation function on output layers
- Dropout layers to help with overfitting

"rest-rest" and "eigen-eigen" contributions as input $\rightarrow$ Input layers of size $2 \cdot (T/a)$

"rest-eigen" contribution entering the loss function $\rightarrow$ Output layer of size (T/a)

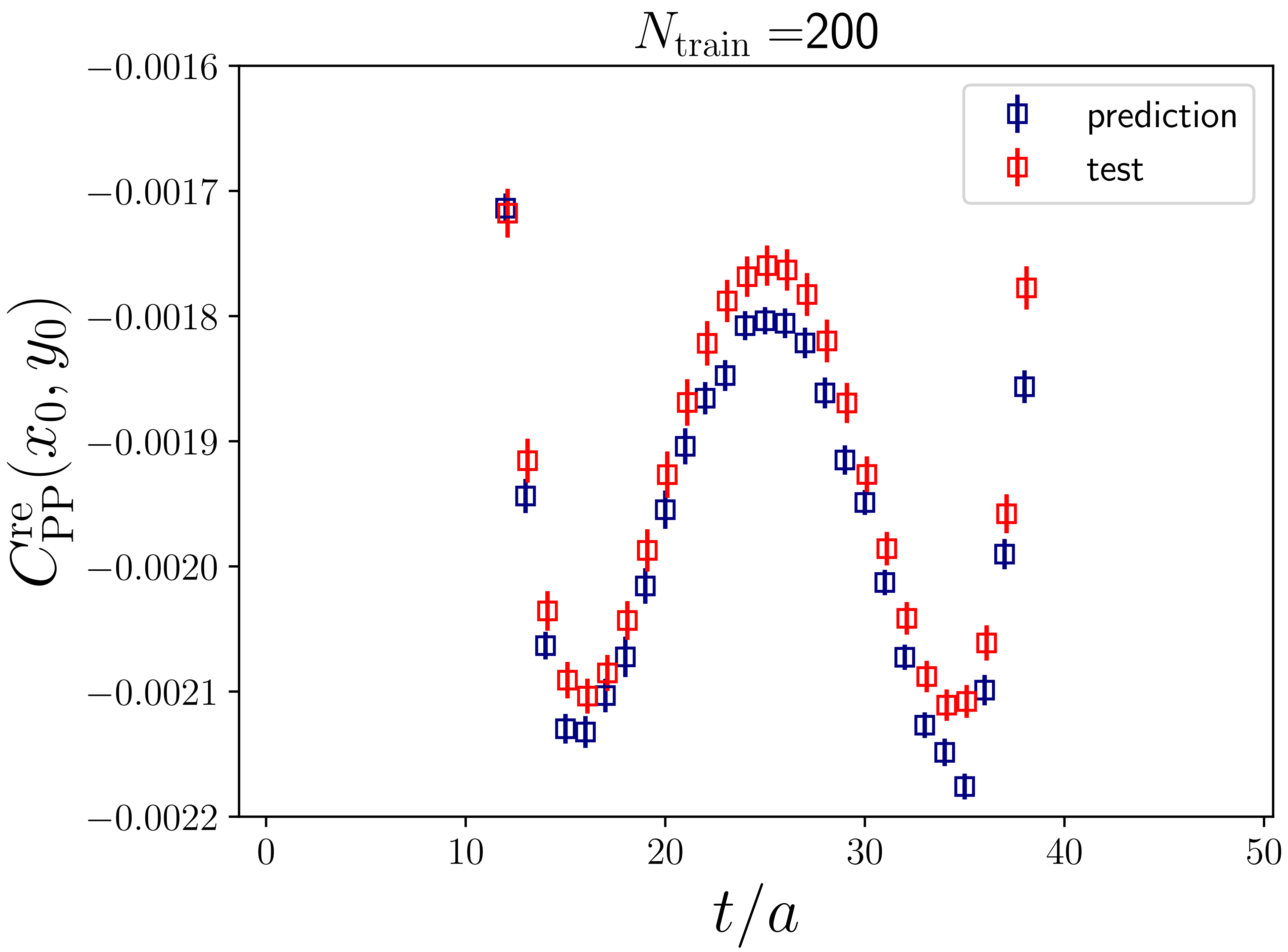
Ensembles:

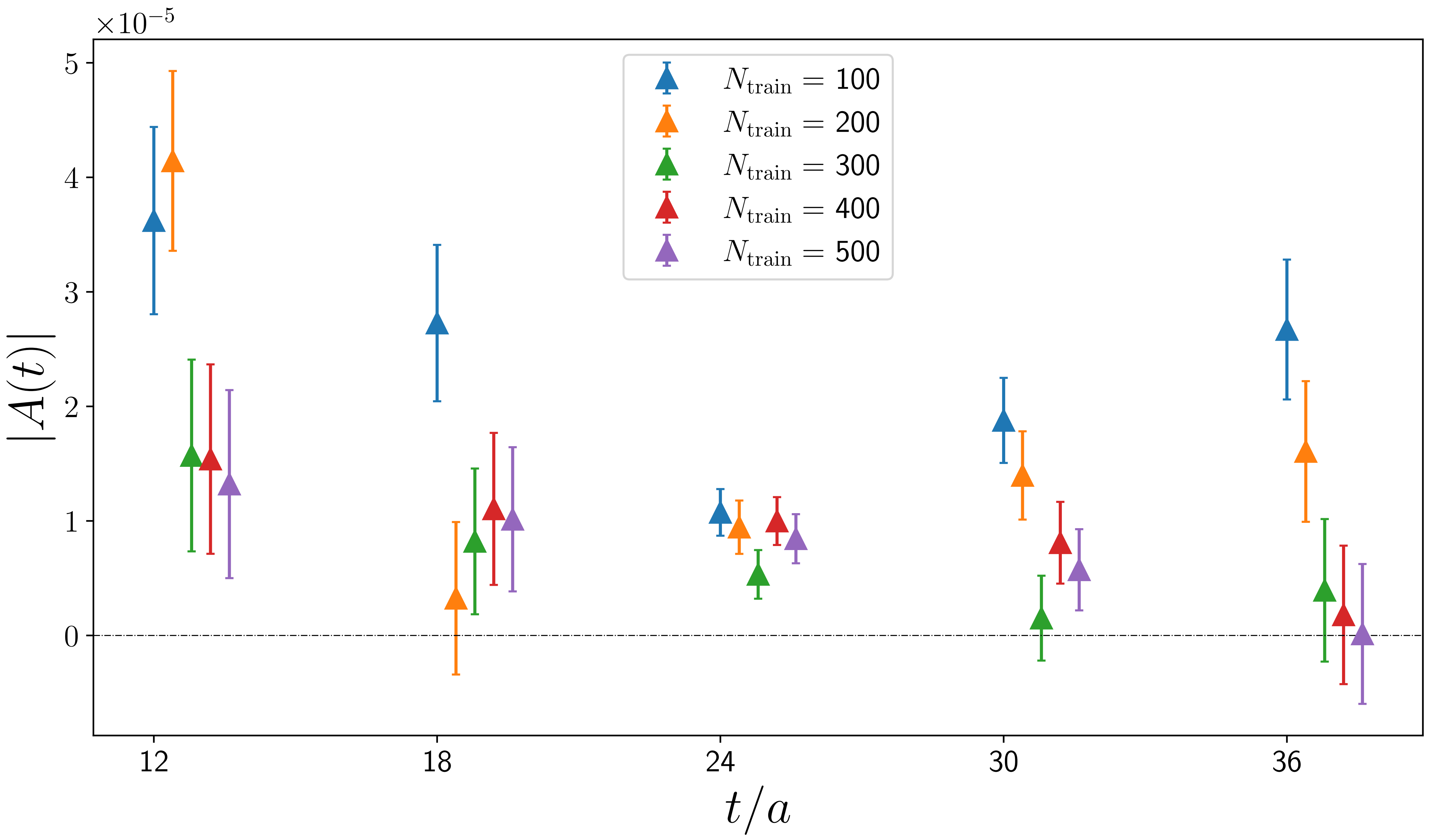
A654: $L/a = 24,\ T/a = 48,\quad m_\pi \simeq 420\,\mathrm{MeV},\quad N_{\mathrm{cfg}} = 2500$

D450: $L/a = 64,\ T/a = 128,\quad m_\pi \simeq 280\,\mathrm{MeV},\quad N_{\mathrm{cfg}} = 500$

Pseudoscalar correlator: rest-eigen contribution

Test quality of prediction for “rest-eigen” contribution — no bias correction (A654 ensemble)

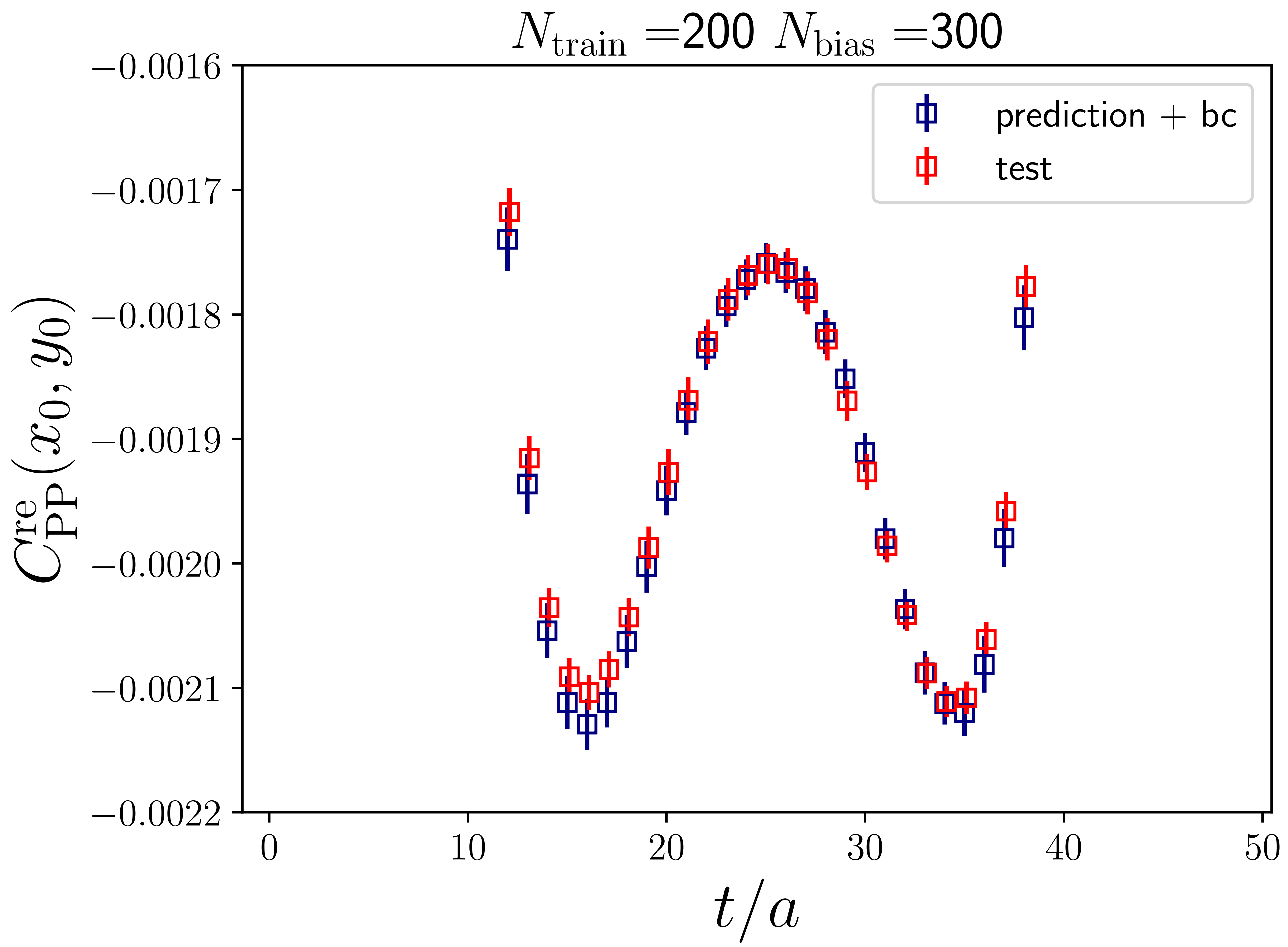


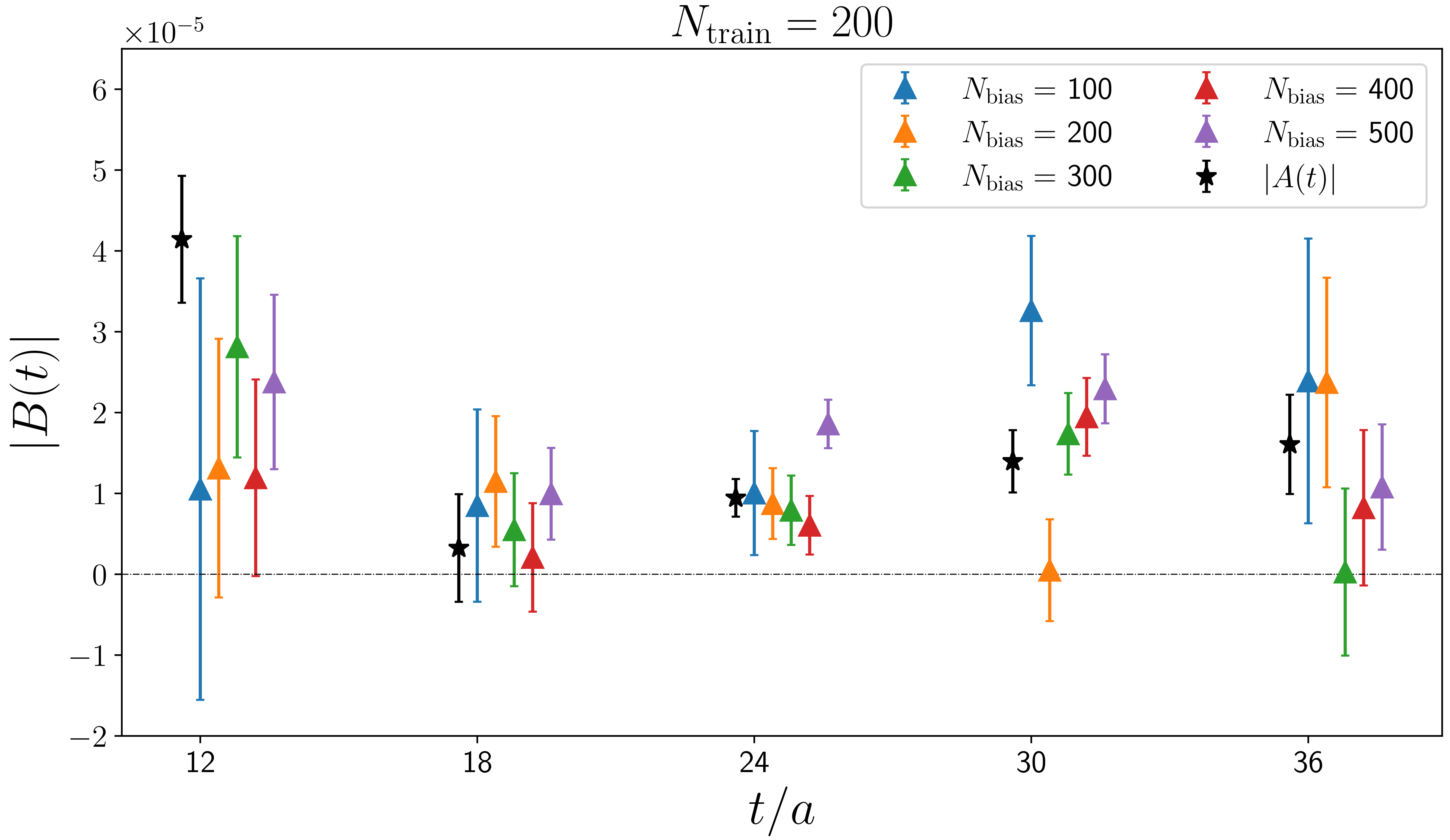


$$A(t) = \left\langle C^{\rm exact}(t) \right\rangle_{\rm test} - \left\langle C^{\rm pred}(t) \right\rangle_{\rm test}$$

Pseudoscalar correlator: rest-eigen contribution

Test quality of prediction for “rest-eigen” contribution — bias-corrected (A654 ensemble)



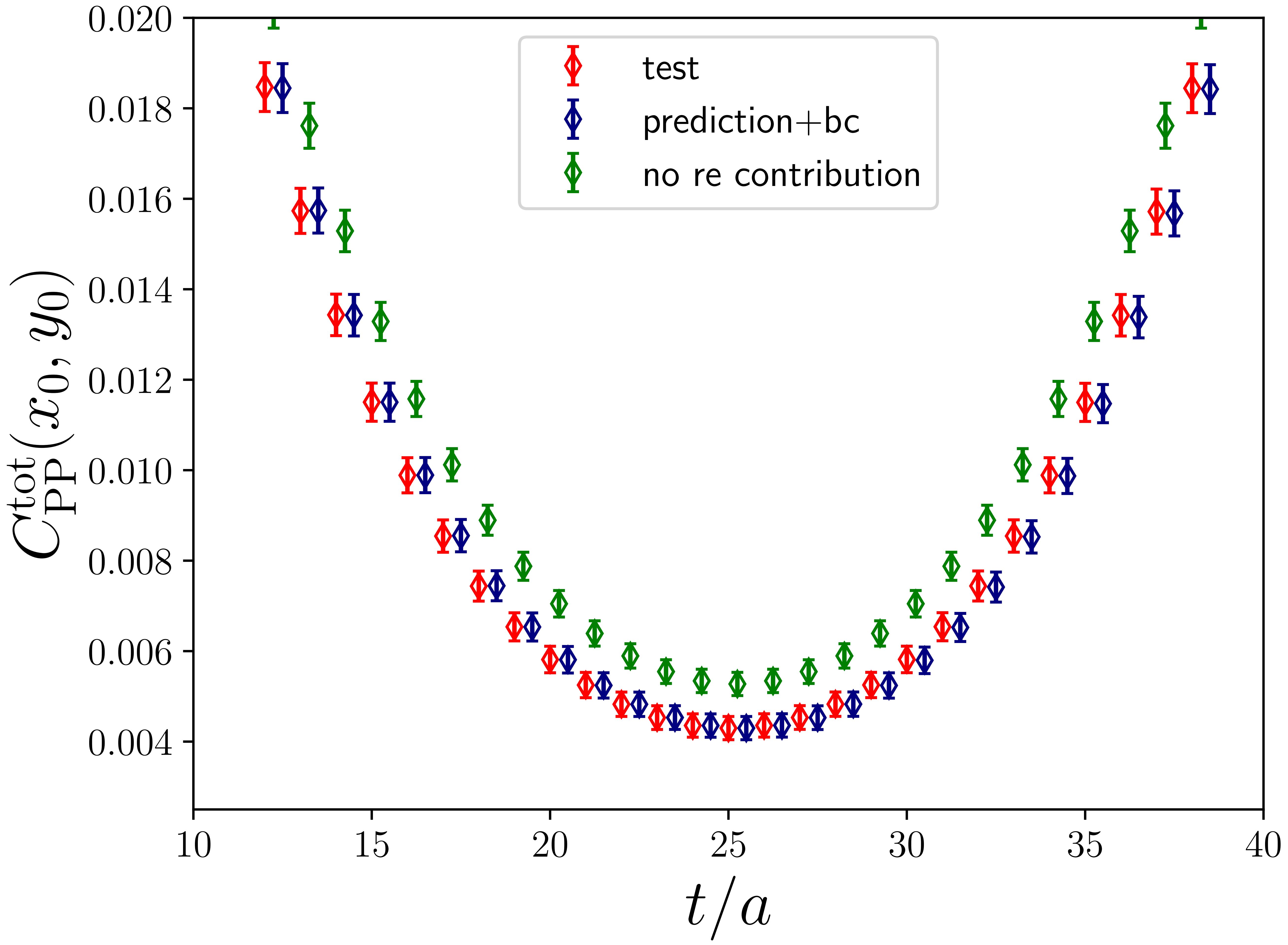


$$A(t) = \left\langle C^{\mathrm{exact}}(t) \right\rangle_{\mathrm{test}} - \left\langle C^{\mathrm{pred}}(t) \right\rangle_{\mathrm{test}}$$

$$B(t) = \left\langle C^{\mathrm{exact}}(t) - C^{\mathrm{pred}}(t) \right\rangle_{\mathrm{bias}}$$

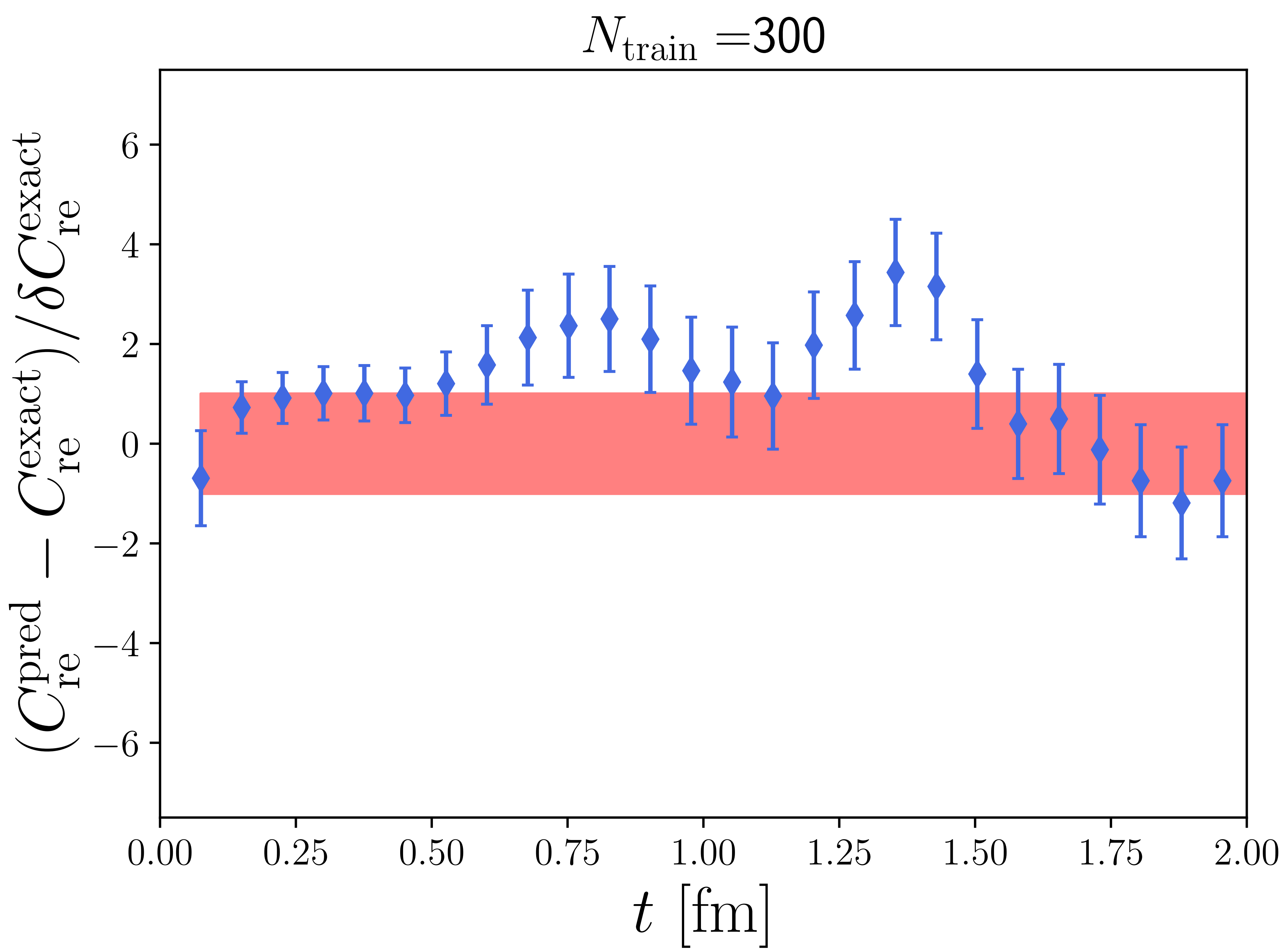
Pseudoscalar correlator: total contribution

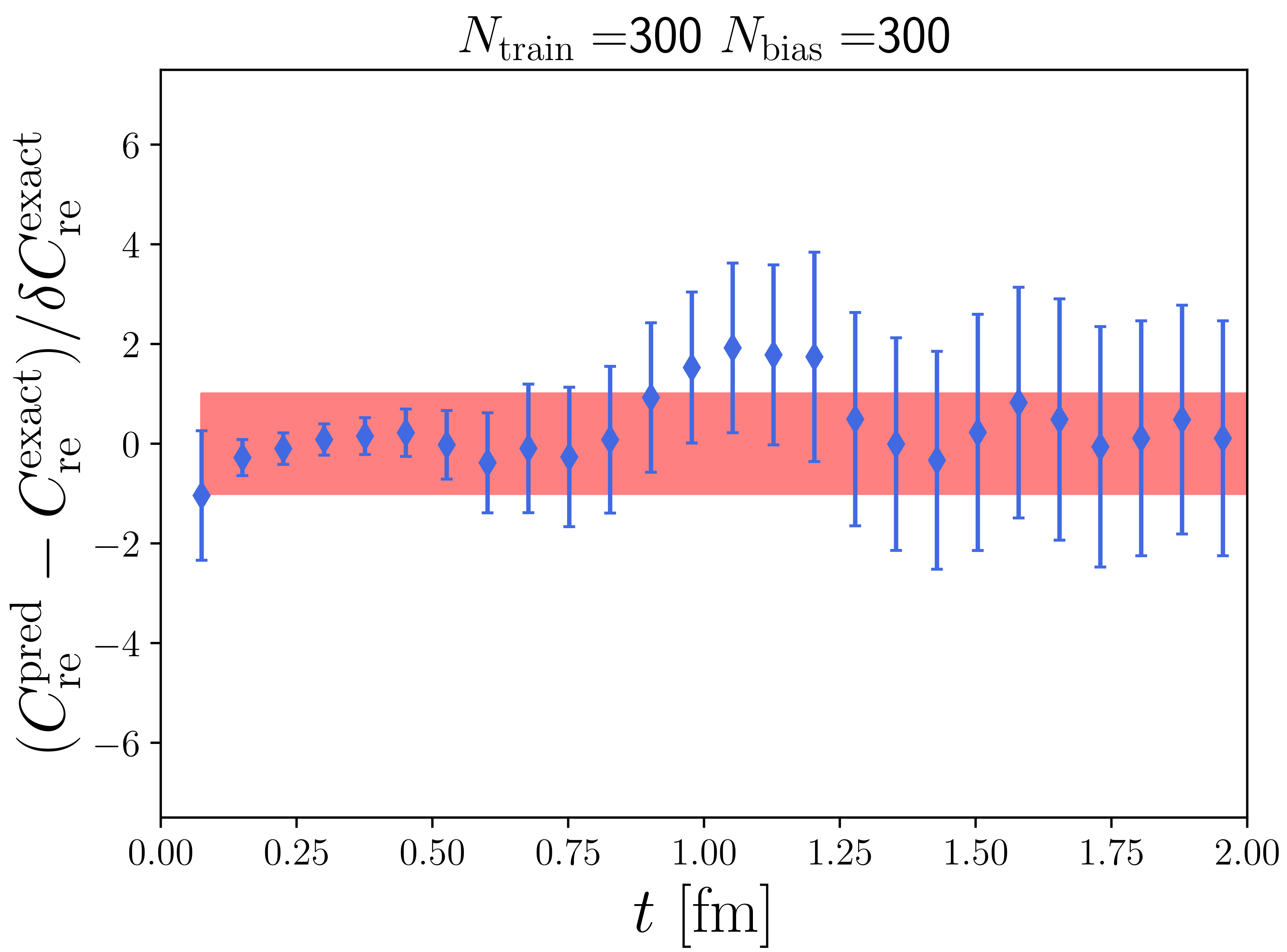
Total contribution (A654 ensemble)



Vector correlator

Test quality of prediction of rest-eigen part — with and without bias correction (A654 ensemble)

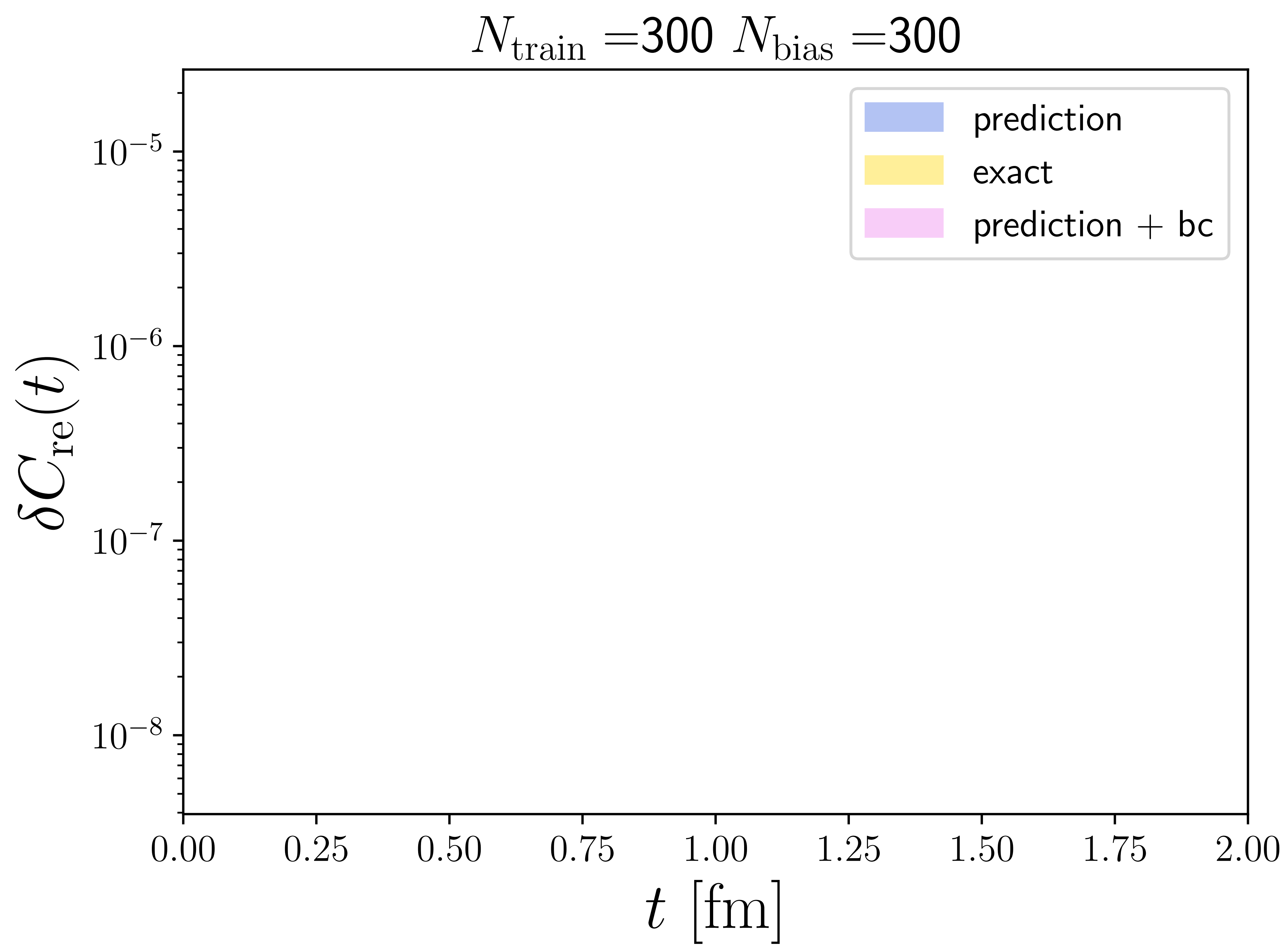


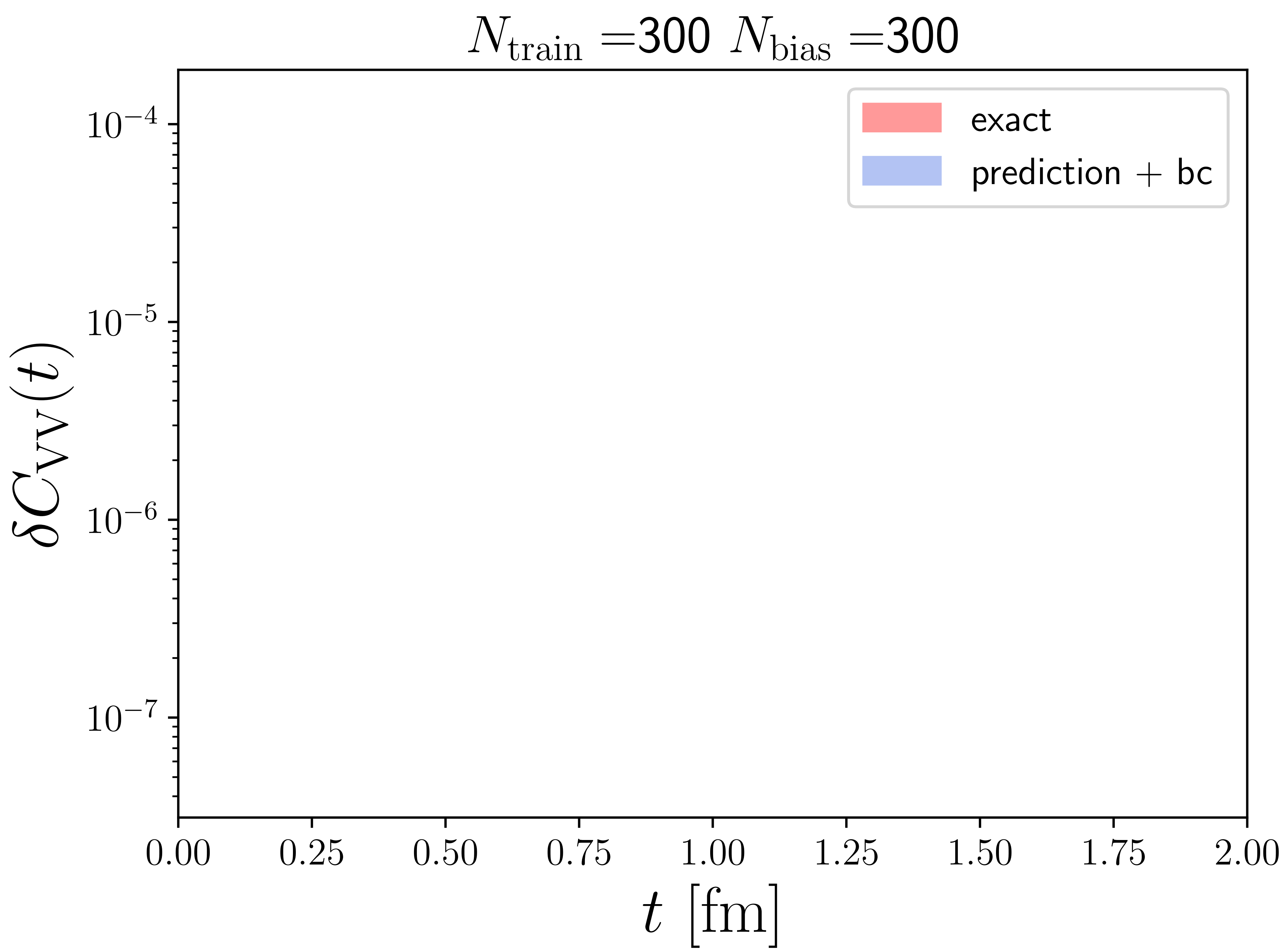


Bias-corrected “re”-correlator consistent with exact calculation, but errors increase at large t

Vector correlator

Absolute error of rest-eigen part and total contribution (A654 ensemble)



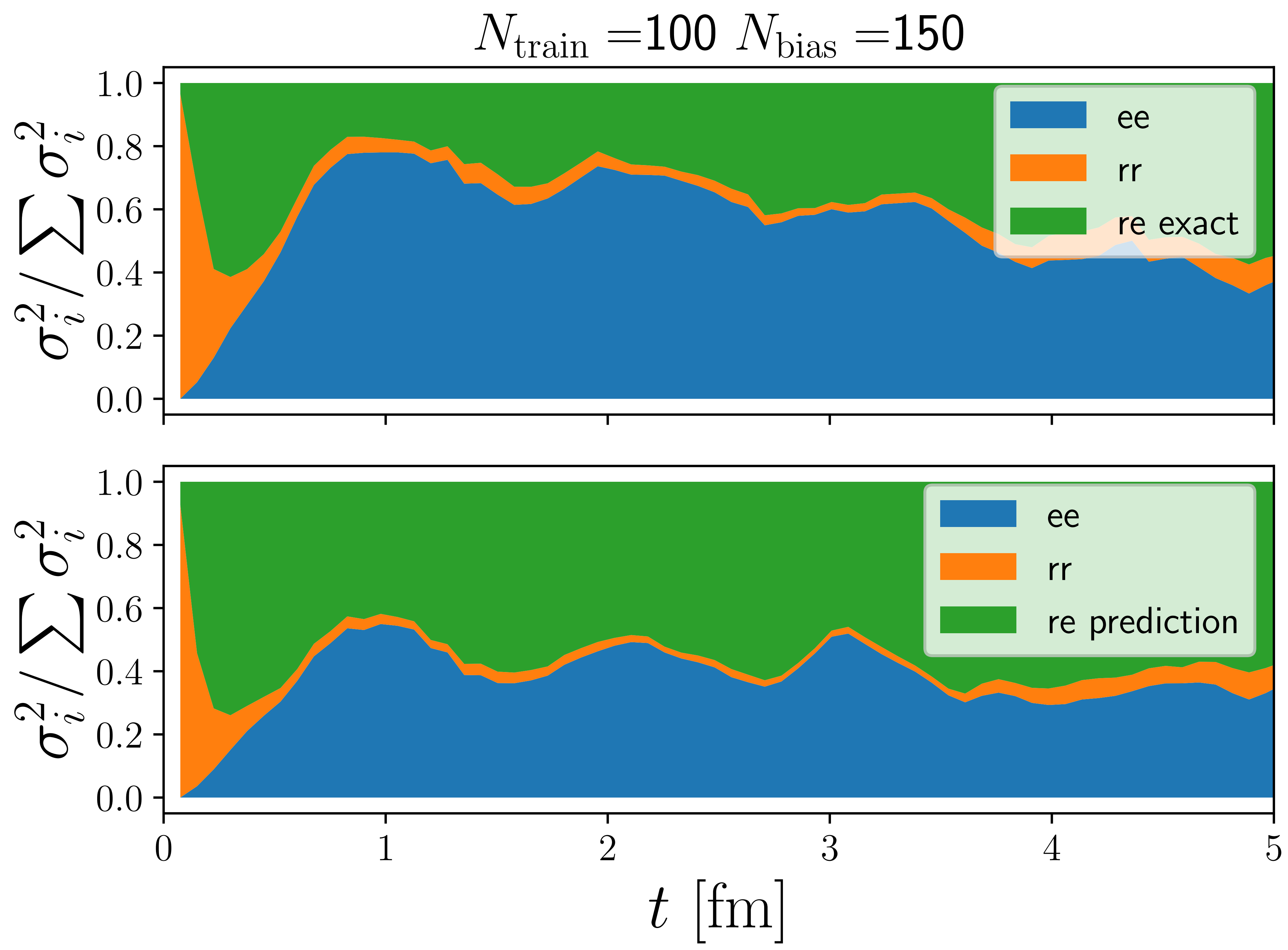


Bias-corrected “re”-correlator consistent with exact calculation, but errors increase at large t

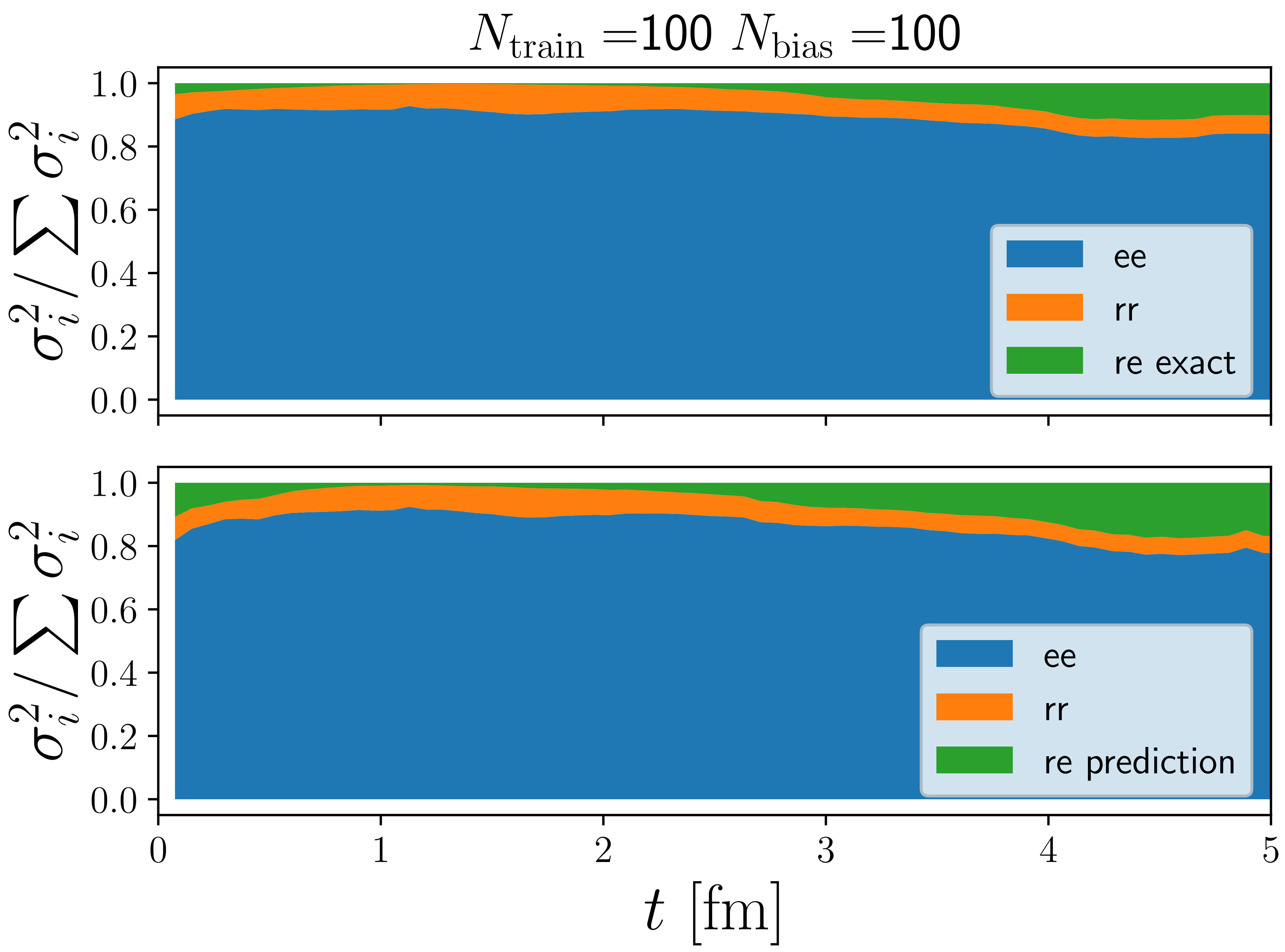
No gain in statistical precision after summing all contributions, unless N_{bias} is increased further

Vector vs. pseudoscalar correlator

Fraction of the total variance of “ee”, “re” and “rr” contributions to (D450 ensemble)



Vector correlator



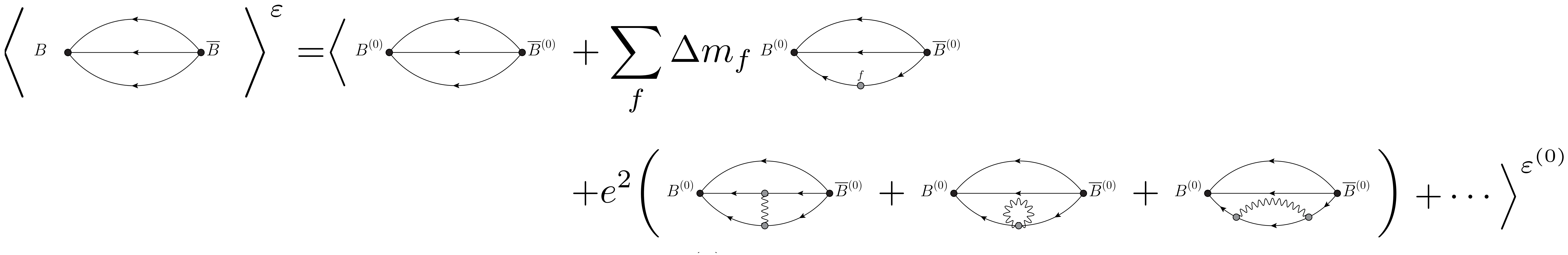
Pseudoscalar correlator

“Eigen-eigen” contribution dominates error in long-distance regime of the pseudoscalar correlator

Isospin-breaking corrections in baryon masses

Precision scale setting for $(g-2)_\mu$ HVP calculations

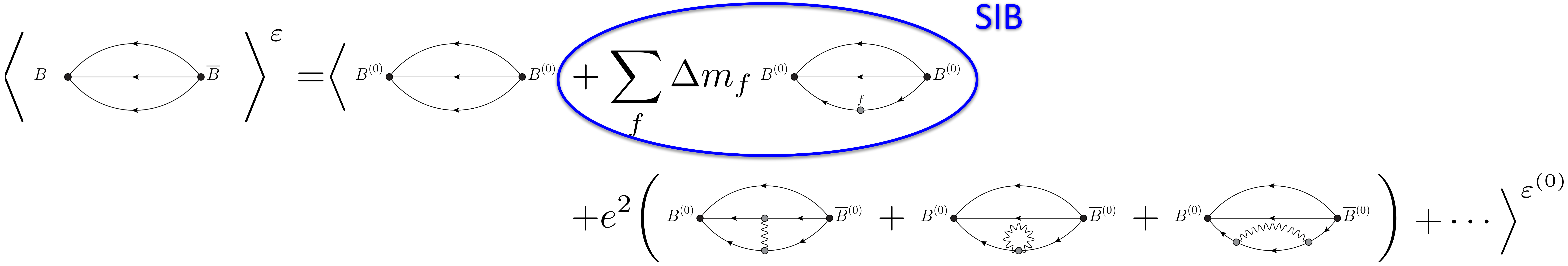
RM123 approach: expansion about iso-symmetric QCD



Isospin-breaking corrections in baryon masses

Precision scale setting for $(g-2)_\mu$ HVP calculations

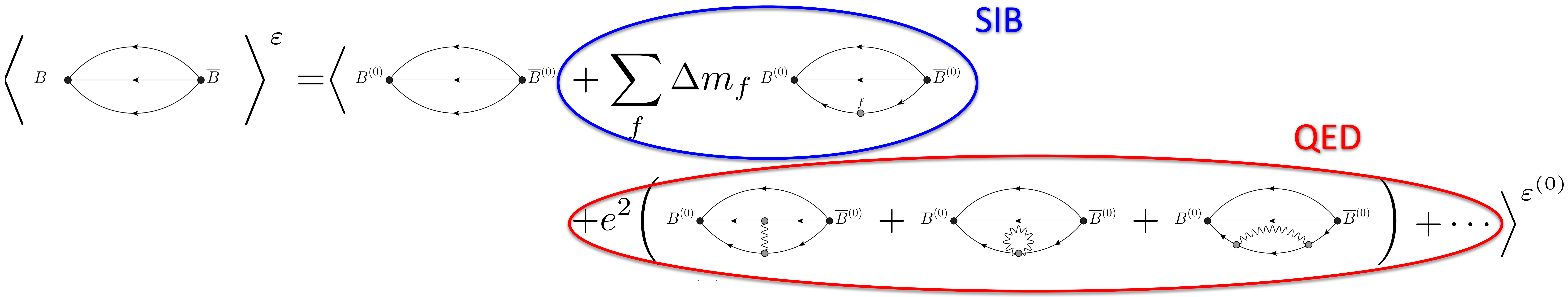
RM123 approach: expansion about iso-symmetric QCD



Isospin-breaking corrections in baryon masses

Precision scale setting for $(g-2)_\mu$ HVP calculations

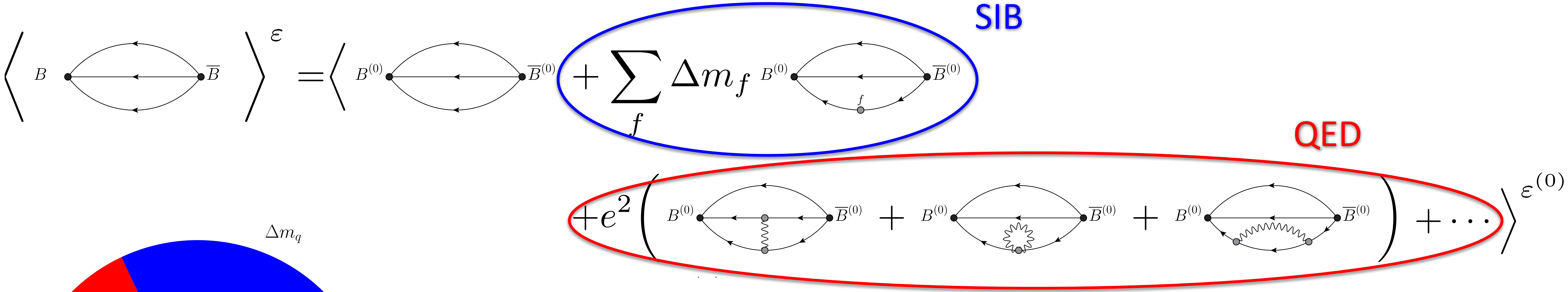
RM123 approach: expansion about iso-symmetric QCD

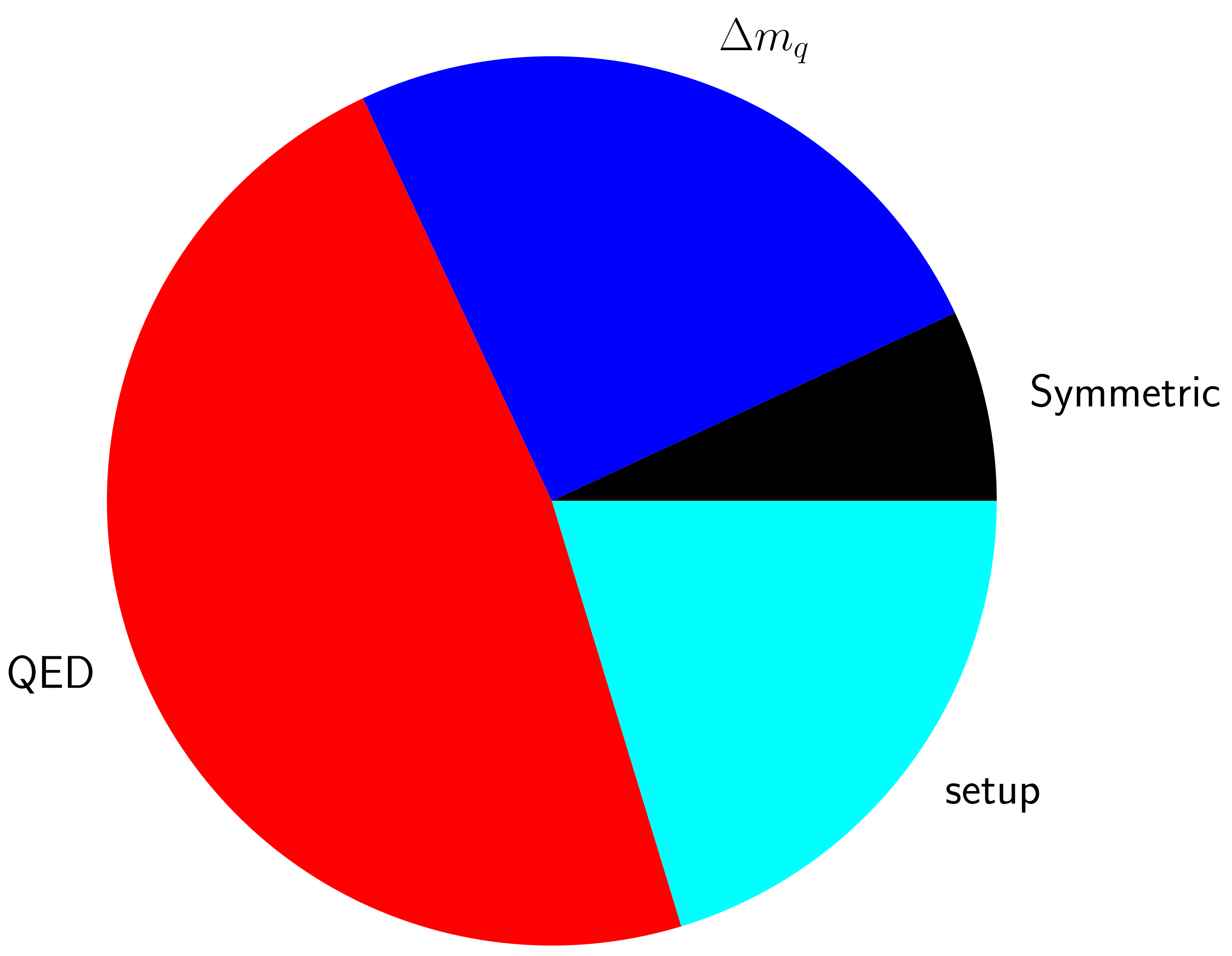


Isospin-breaking corrections in baryon masses

Precision scale setting for $(g-2)_\mu$ HVP calculations

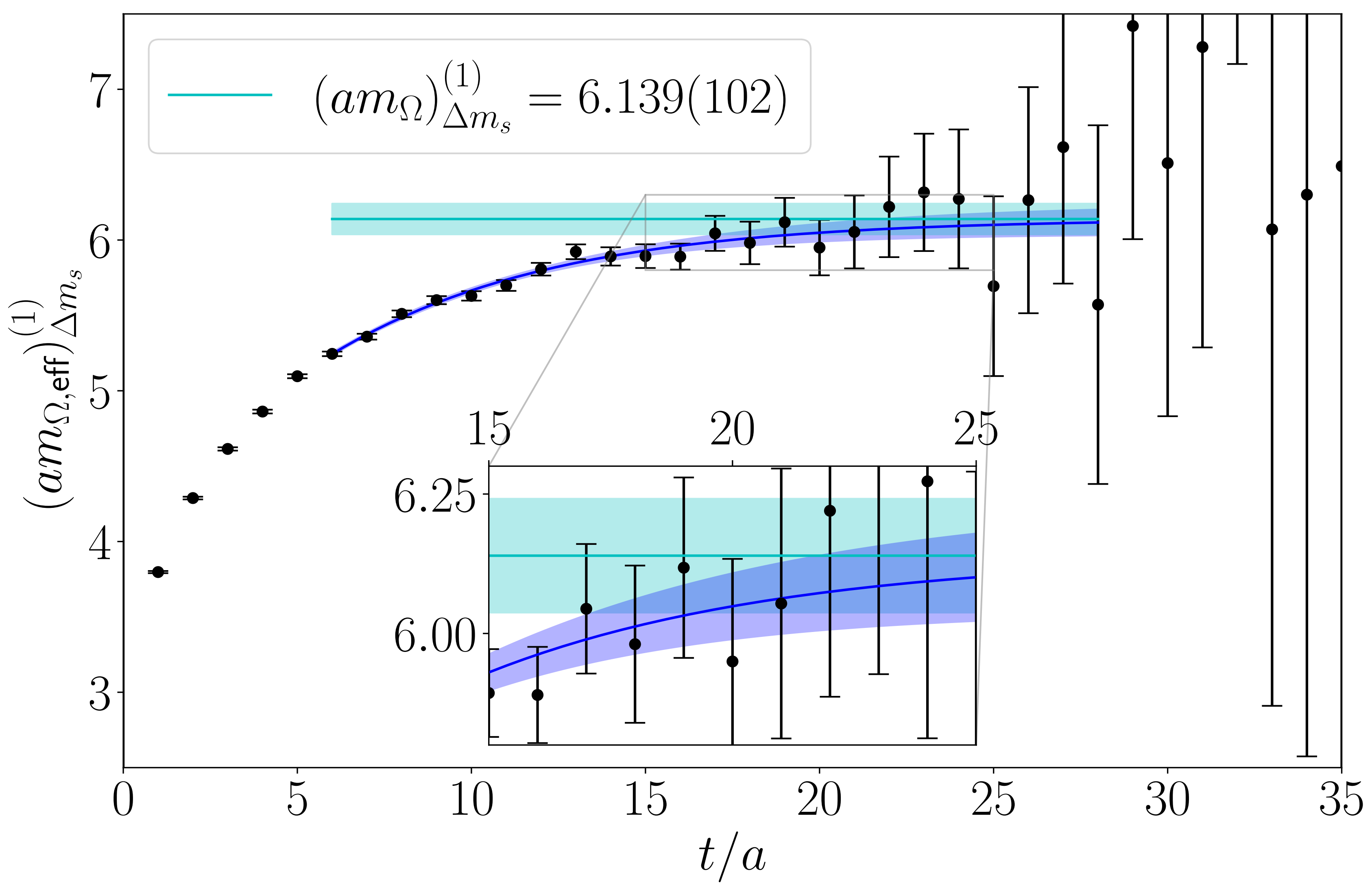
RM123 approach: expansion about iso-symmetric QCD

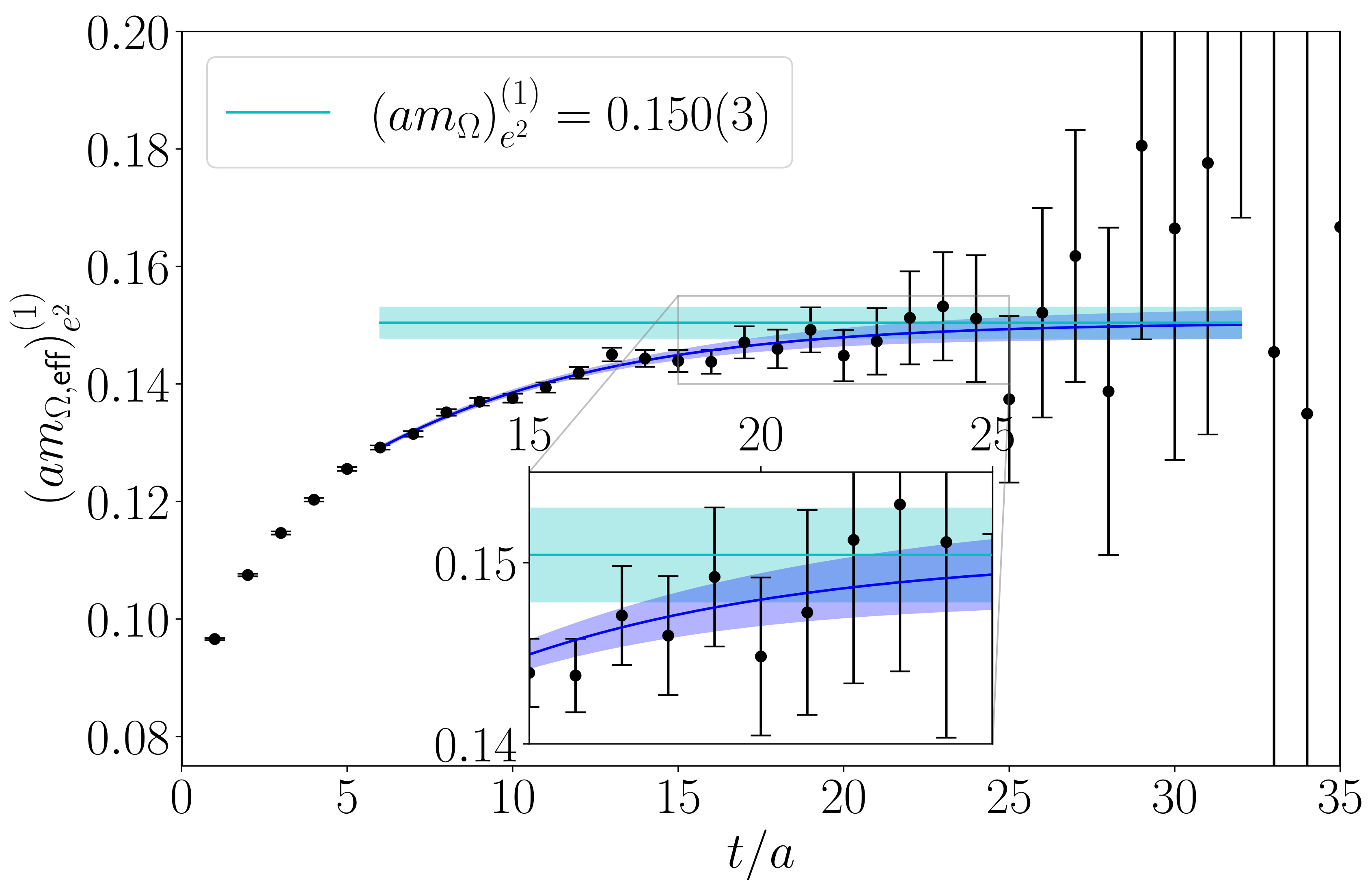




- QED contribution by far the most expensive part
- Observe strong correlation between the correlators describing QED and strong isospin-breaking contributions

Machine-learning model





- Train model $M(C^{(0)}, C^{(1)}_{\Delta m_u}, C^{(1)}_{\Delta m_d}, C^{(1)}_{\Delta m_s})$ to predict the QED contribution $C^{(1)}_{e^2}$

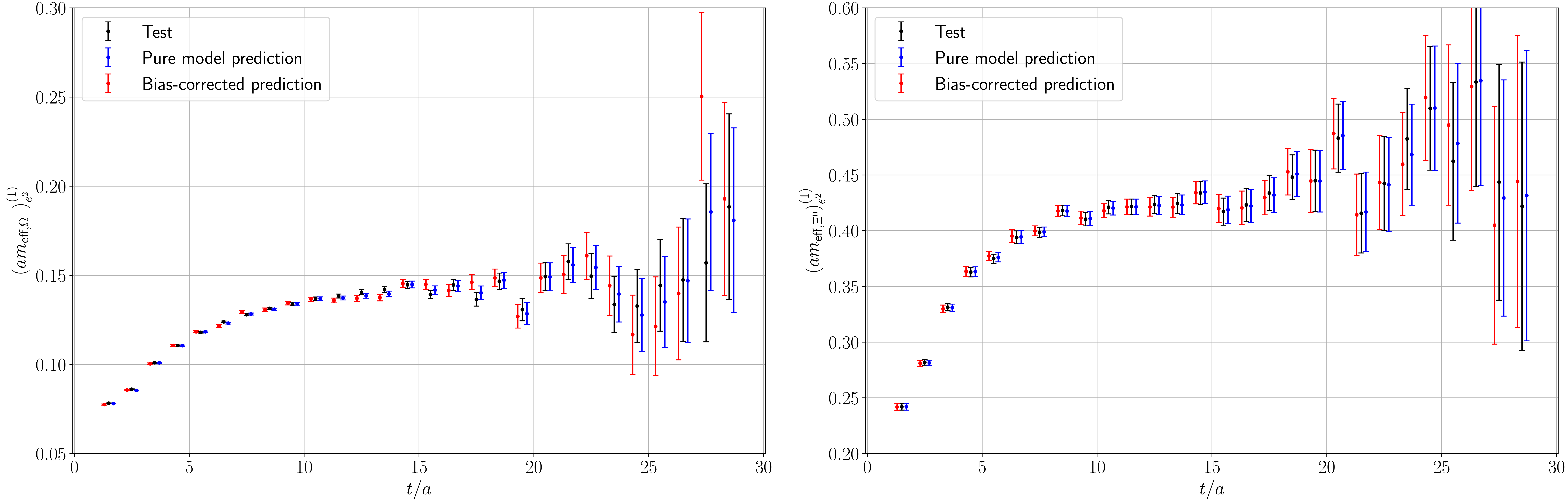
$$M(t) = \alpha\, C^{(0)}(t) + \beta\, C^{(1)}_{\Delta m_u}(t) + \gamma\, C^{(1)}_{\Delta m_d}(t) + \delta\, C^{(1)}_{\Delta m_s}(t) + \epsilon$$

- Correct for bias by using a small number of sources on each configuration: $N_{\mathrm{src,\,bias}} \ll N_{\mathrm{src}}$

- Ensemble N451: $48^3 \cdot 96, \quad m_\pi \simeq 280\,\mathrm{MeV}, \quad N_{\mathrm{cfg}} = 1011, \;\; N_{\mathrm{train}} = 20, \quad N_{\mathrm{src}} = 32, \quad N_{\mathrm{src,\,bias}} = 1$

Results

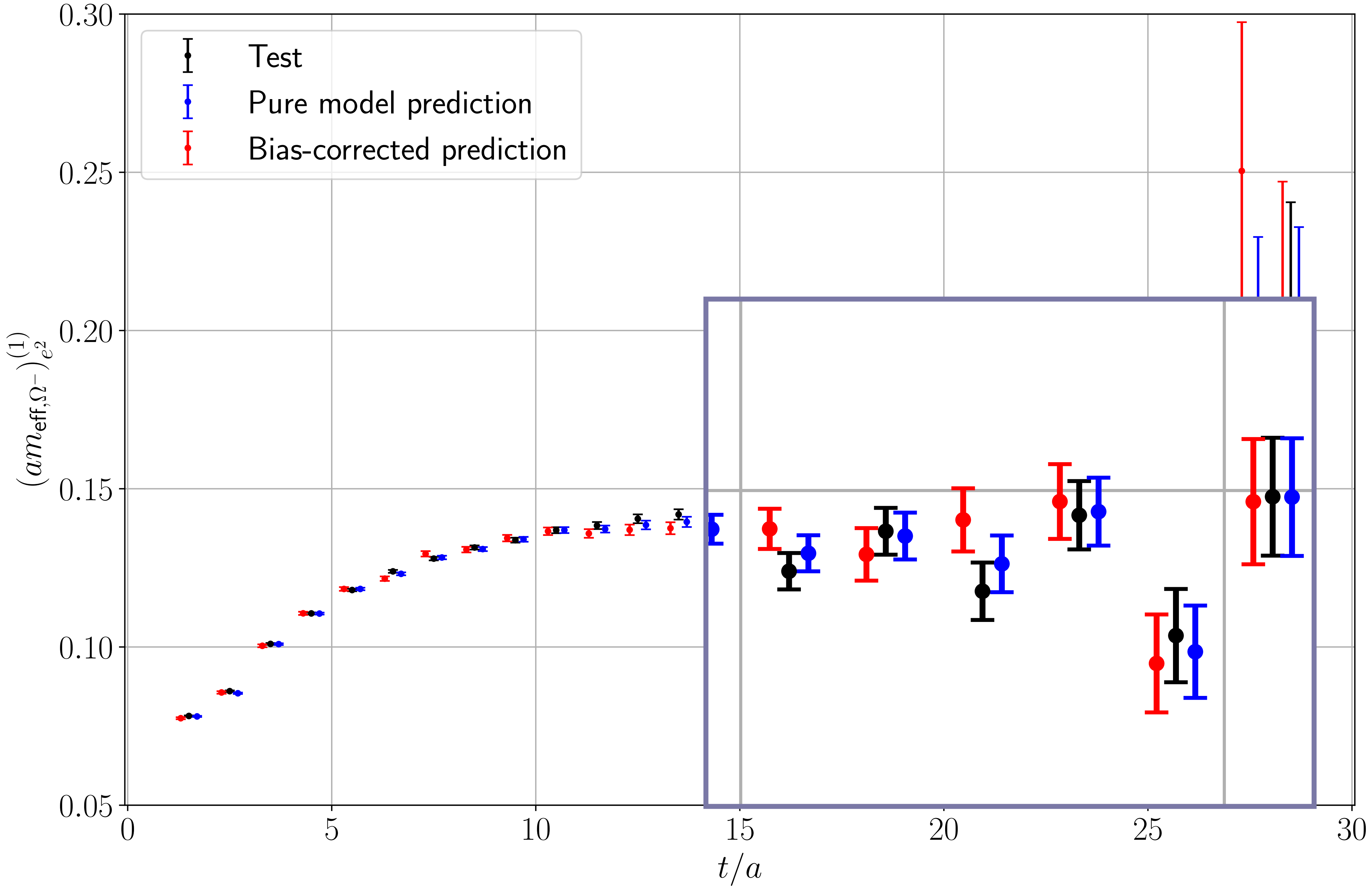
QED correction to Ω^- and Ξ^- masses on N451 ensemble

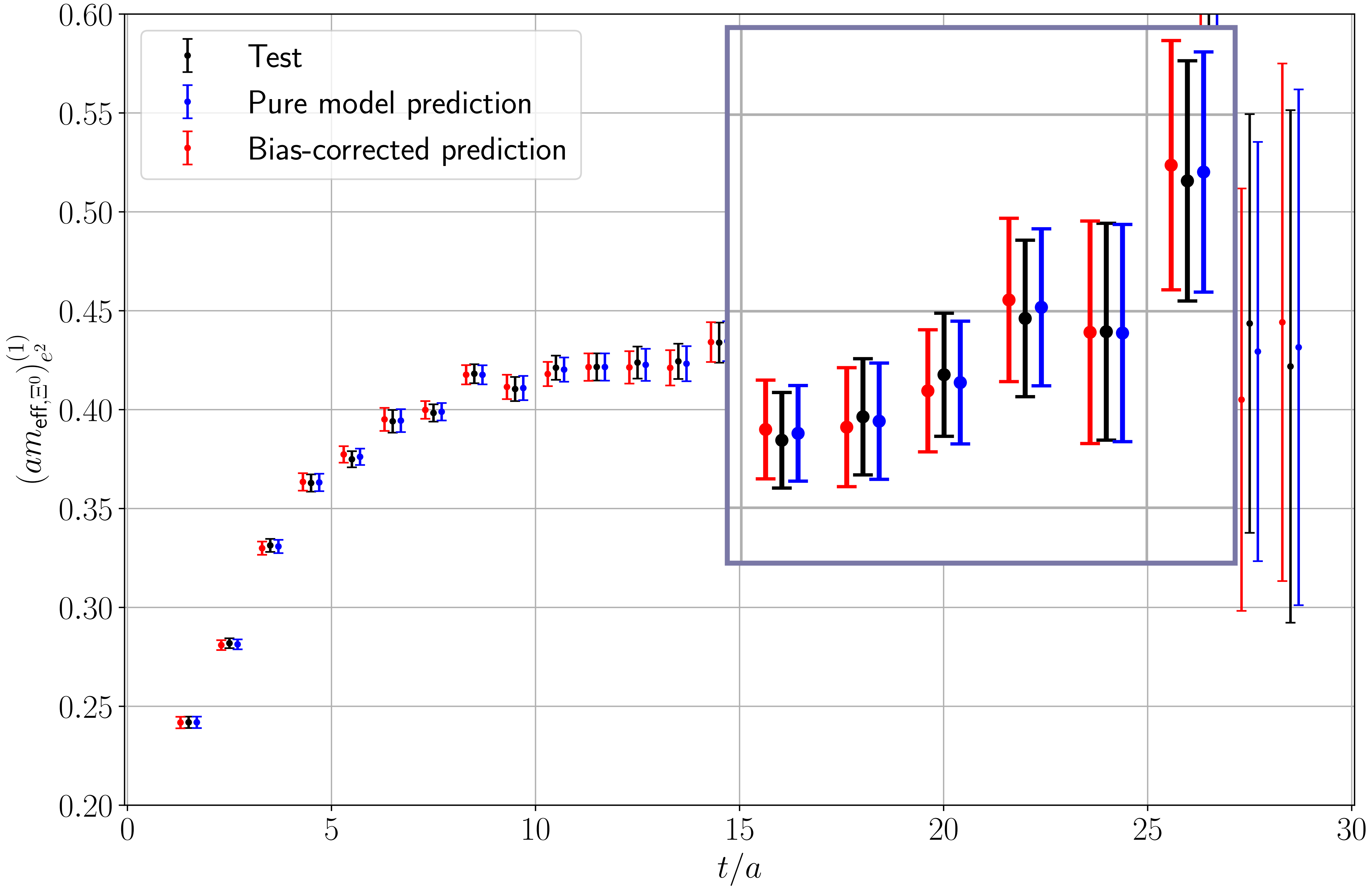


- Increasing $N_{\text{src, bias}}$ has no effect on the uncertainty in the bias-corrected result
- Training time negligible; reduction of numerical cost by 50%

Results

QED correction to Ω^- and Ξ^- masses on N451 ensemble





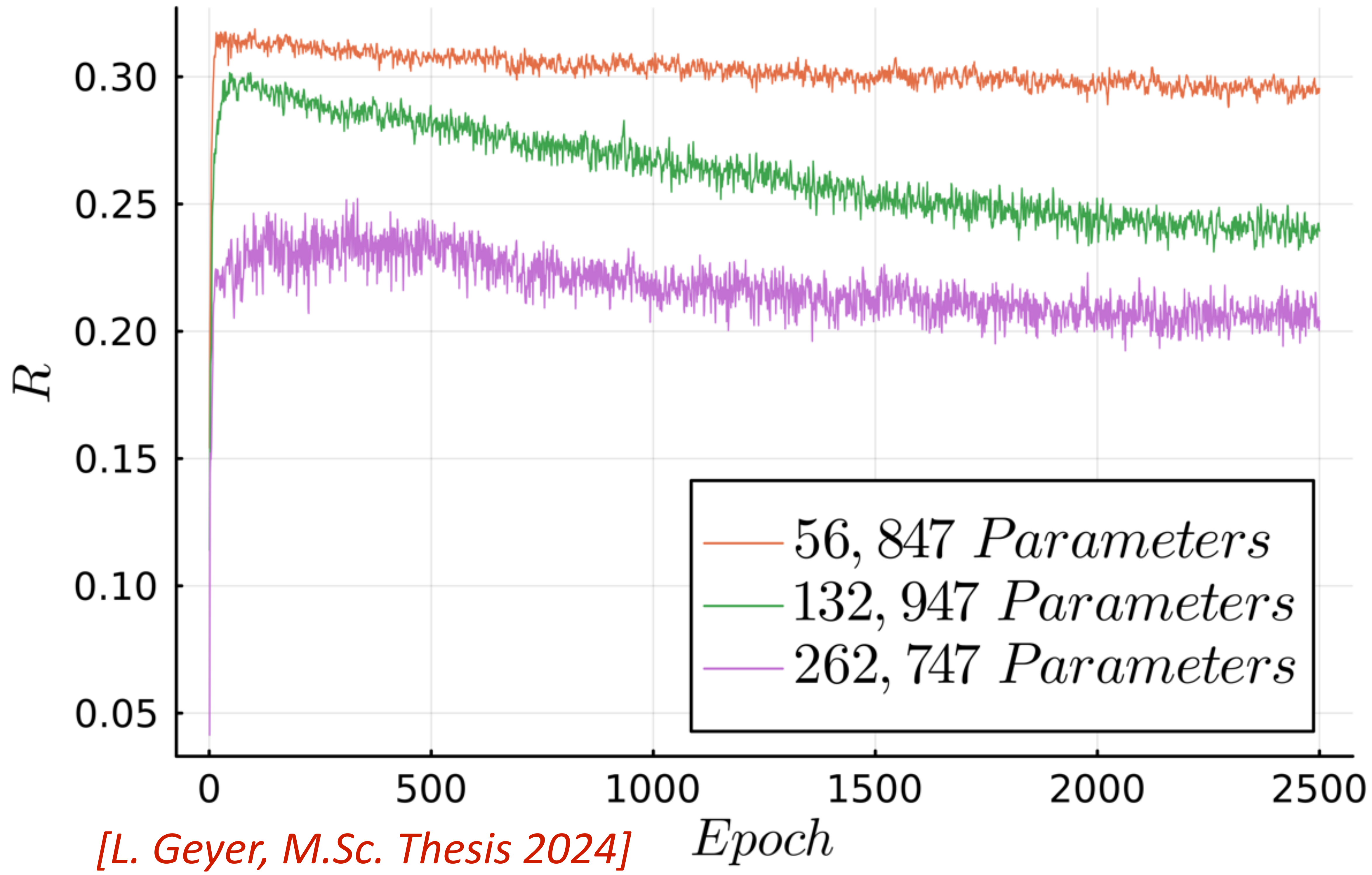
- Increasing $N_{\text{src, bias}}$ has no effect on the uncertainty in the bias-corrected result
- Training time negligible; reduction of numerical cost by 50%

Summary and conclusions

- Machine-learning models with bias correction are able to reproduce exact calculations with comparable statistical precision
- No “noise reduction” observed in all models studied so far
- Saving in computer time can be substantial, but depends strongly on the observable
- Using ML for QED part leads to 50% reduction in numerical effort for computing baryon masses including isospin-breaking corrections
- Rest-eigen part of vector correlator:
  bias correction dominates the total uncertainty; less CPU time produces a larger error
- Outlook: optimise setup to increase correlations between O_{appx} and O

Backup

Grid search and R-score



[L. Geyer, M.Sc. Thesis 2024]

Models with increasing complexity lead to overfitting

$$R_k = 1 - \frac{\sum_{i=1}^{47}(v_{\mathrm{t},k,i}^{\mathrm{ER}} - v_{\mathrm{p},k,i}^{\mathrm{ER}})^2}{\sum_{i=1}^{47}(v_{\mathrm{t},k,i}^{\mathrm{ER}} - \hat{v}_{\mathrm{t},i}^{\mathrm{ER,tr}})^2}, \quad R_k \in (-\infty, 1].$$