A new class of error reduction techniques and its applications

Taku Izubuchi,
with T. Blum E. Shintani,
C. Lehner, T. Kawanai
T. Ishikawa
RBC/UKQCD

BROOKHAVEN NATIONAL LABORATORY

RIKEN BNL Research Center

GGI workshop, New Frontiers in Lattice Gauge Theory, 2012-09-14, Firenze

Hunting the Higgs
Now Playing: CERN live webcast.

live video courtesy of CERN, copyright CERN 2012

Motivation

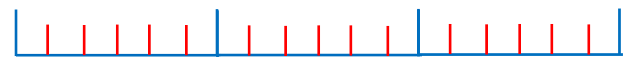
- A new boson is now discovered
- Precise theoretical calculation becomes even more important to confirm or reject the standard model
  Fukaya, Lin, Bernard, Blossier, Golterman, Kronfeld
- More than half of CPU cycles of lattice QCD are for valence calculations (M. Luescher's talk for HMC)
  - Nucleon's structure (e.g. C. Alexandrou's talk….)
  - on physics point QCD simulation
  - higher dimensional bulk operators (cumulants… Ejiri, Gavai, Takeda, Petreczky, Karsch, Allton, Gupta….)
  - multi hadron simulations (S. Aoki…)

It's a shame to be limited by statistical error

Multiple timestep in HMC

- Multiple time steps in MD integrators

- Sexton & Weingarten trick



- Hasenbusch trick : introduce intermediate mass

$$\det[D(m)] = \overbrace{\det[D(m_I)]}^{\text{cheap mode}} \times \overbrace{\det\left[D(m)D(m_I)^{-1}\right]}^{\text{expensive mode}}$$

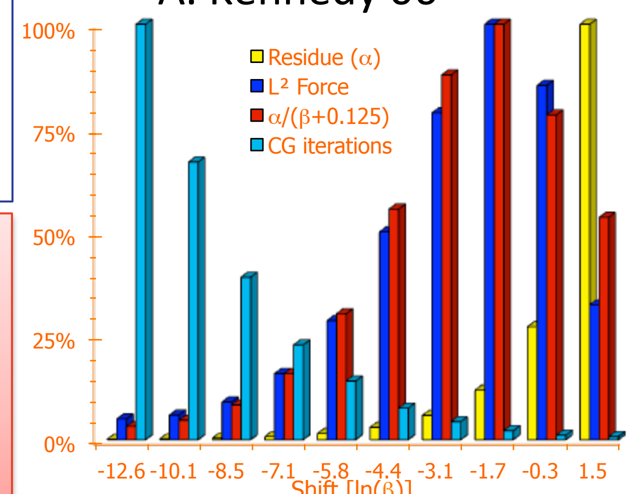
- Clark & Kennedy RHMC (quotient force term)

Berlin Wall was torn down by
Smart Work Sharings
Similar tricks for valence ?

Low Mode Averaging (LMA)

L. Giusti et al. , JHEP 0404, 013 (2004)
see also H. Neff et al Phys. Rev. D 64 (2001) 114509
T. DeGrand and S. Schaefer,
Comput. Phys. Commun. 159 (2004) 185

A. Kennedy 06



State of Obvious

- Many interesting physics are limited by statistical error

$$\mathrm{err} \approx C \times \frac{1}{\sqrt{N_{\mathrm{meas}}}}$$

- Do more number of measurements, N_{meas}

- Change to observable with smaller fluctuation, C

- Covariant Approximation Averaging (CAA)
  Combine the above using
  - symmetries of the lattice action
  - (crude) approximations

Covariant Approximation Averaging (CAA)

- Original observable $\mathcal{O}$
- Covariant approximation of the observable $\mathcal{O}^{(\mathrm{appx})}$ under a lattice symmetry $g \in G$

$$\langle \mathcal{O}^{(\mathrm{appx})} \rangle = \langle \mathcal{O}^{(\mathrm{appx}),g} \rangle$$

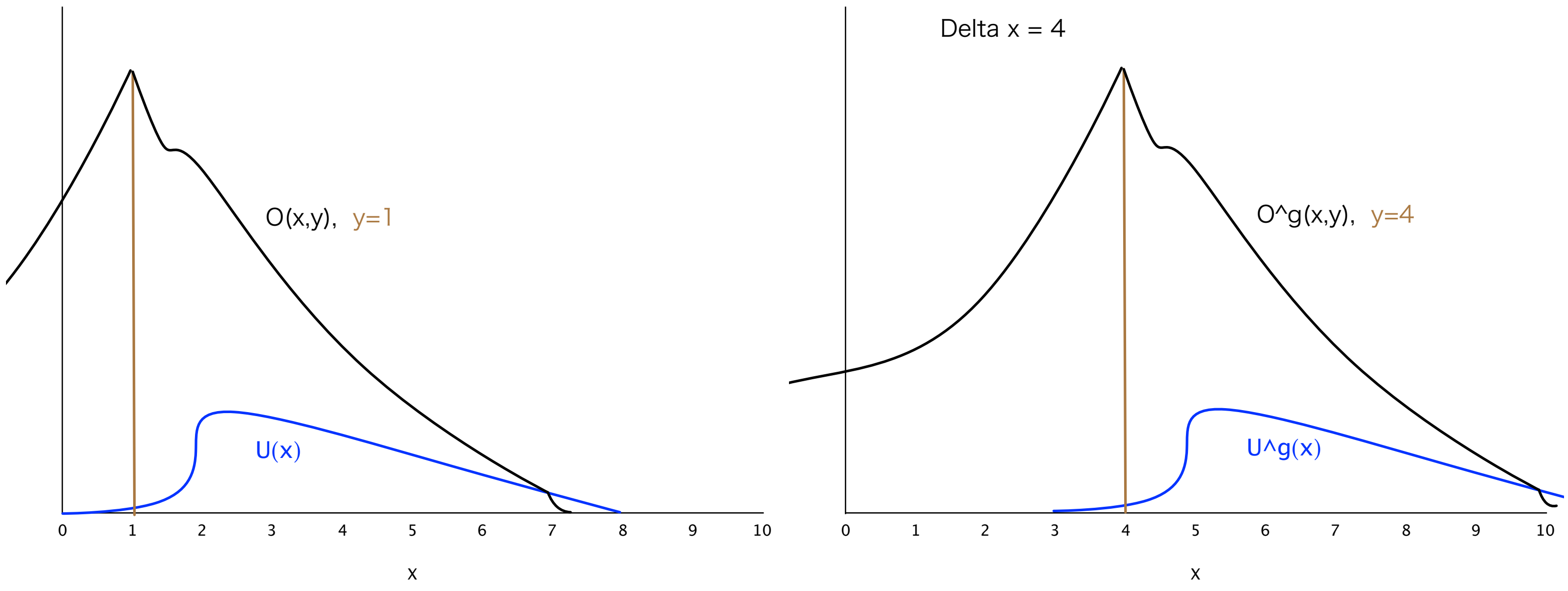
- Unbiased improved estimator

$$\mathcal{O}^{(\mathrm{rest})} = \mathcal{O} - \mathcal{O}^{(\mathrm{appx})}$$

$$\mathcal{O}^{(\mathrm{imp})} = \mathcal{O}^{(\mathrm{rest})} + \frac{1}{N_G} \sum_{g \in G} \mathcal{O}^{(\mathrm{appx}),g}$$

Covariant approximation

- $O^{(appx)}$ needs to be precisely (to the numerical accuracy required) covariant under the symmetry of lattice action to avoid systematic errors.



One should check in the code using explicitly shifted gauge configuration

Unbiasness proof

- Consider a element g of lattice symmetry G e.g. $x_\mu \to x + \Delta x_\mu^{(g)}$
- transformation of fields

$$U_\mu(x) \to U_\mu^g(x) = U_\mu(x - \Delta x^{(g)})$$

$$\begin{aligned}\mathcal{O}[U_\mu] \to & \mathcal{O}^g[U_\mu^g](x_1, x_2, \cdots, x_n) \\ & = \mathcal{O}[U_\mu^g](x_1 - \Delta x^{(g)}x, x_2 - \Delta x^{(g)}x, \cdots, x_n - \Delta x^{(g)}x),\end{aligned}$$

- Observable (and its approximation) is called to have covariance under g iff

$$\mathcal{O}^g[U_\mu^g](x_1, x_2, \cdots, x_n) = \mathcal{O}[U_\mu](x_1, x_2, \cdots, x_n)$$

or, more explicitly,

$$\mathcal{O}[U_\mu^g](x_1 - \Delta x^{(g)}, x_2 - \Delta x^{(g)}, \cdots, x_n - \Delta x^{(g)}) = \mathcal{O}[U_\mu](x_1, x_2, \cdots, x_n) \quad .$$

- When g is a symmetry of lattice, and $O^{(appx)}$ is covariant $\langle \mathcal{O}^g \rangle = \langle \mathcal{O} \rangle$

$$\mathcal{O}^{(\mathrm{rest})} = \mathcal{O} - \mathcal{O}^{(\mathrm{appx})}$$

$$\mathcal{O}^{(\mathrm{imp})} = \mathcal{O}^{(\mathrm{rest})} + \frac{1}{N_G} \sum_{g \in G} \mathcal{O}^{(\mathrm{appx}),g}$$

$$\langle \mathcal{O}^{\mathrm{imp}} \rangle = \langle \mathcal{O} \rangle$$

Why expect improvements ?

$$\mathcal{O}^{(\mathrm{rest})} = \mathcal{O} - \mathcal{O}^{(\mathrm{appx})}$$

$$\mathcal{O}^{(\mathrm{imp})} = \mathcal{O}^{(\mathrm{rest})} + \frac{1}{N_G} \sum_{g \in G} \mathcal{O}^{(\mathrm{appx}),g}$$

- $O^{(\mathrm{imp})}$ has smaller error, smaller C

  <= accuracy of approximation controls error,

  need not to be too accurate (0.1% is good enough)

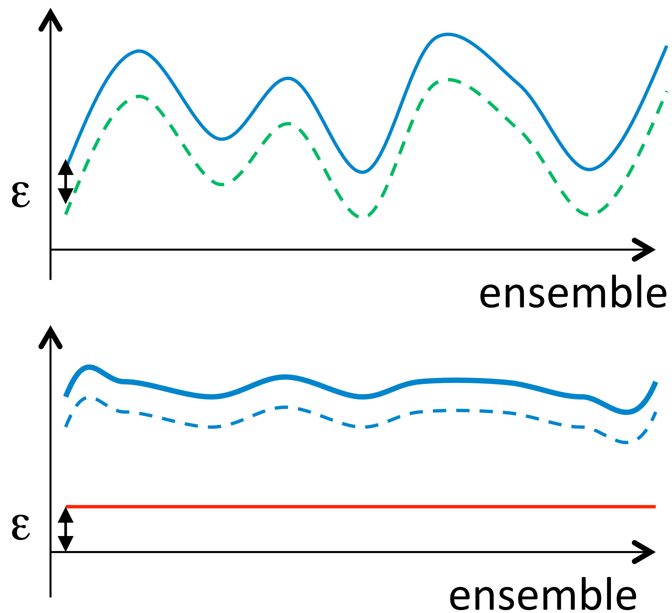
- N_G suppresses the bulk part of noise cheaply

$$\mathrm{err} \approx C \times \frac{1}{\sqrt{N_{\mathrm{meas}}}}$$

Valence version of Hasenbushing in HMC

CMA : a smart work sharing

- Ideal approximation



O

O^{appx}

O^{imp}

$O^{\mathrm{ave\ appx}}$

O^{rest}

O^{appx} is <u>*strongly correlated*</u> with original one.

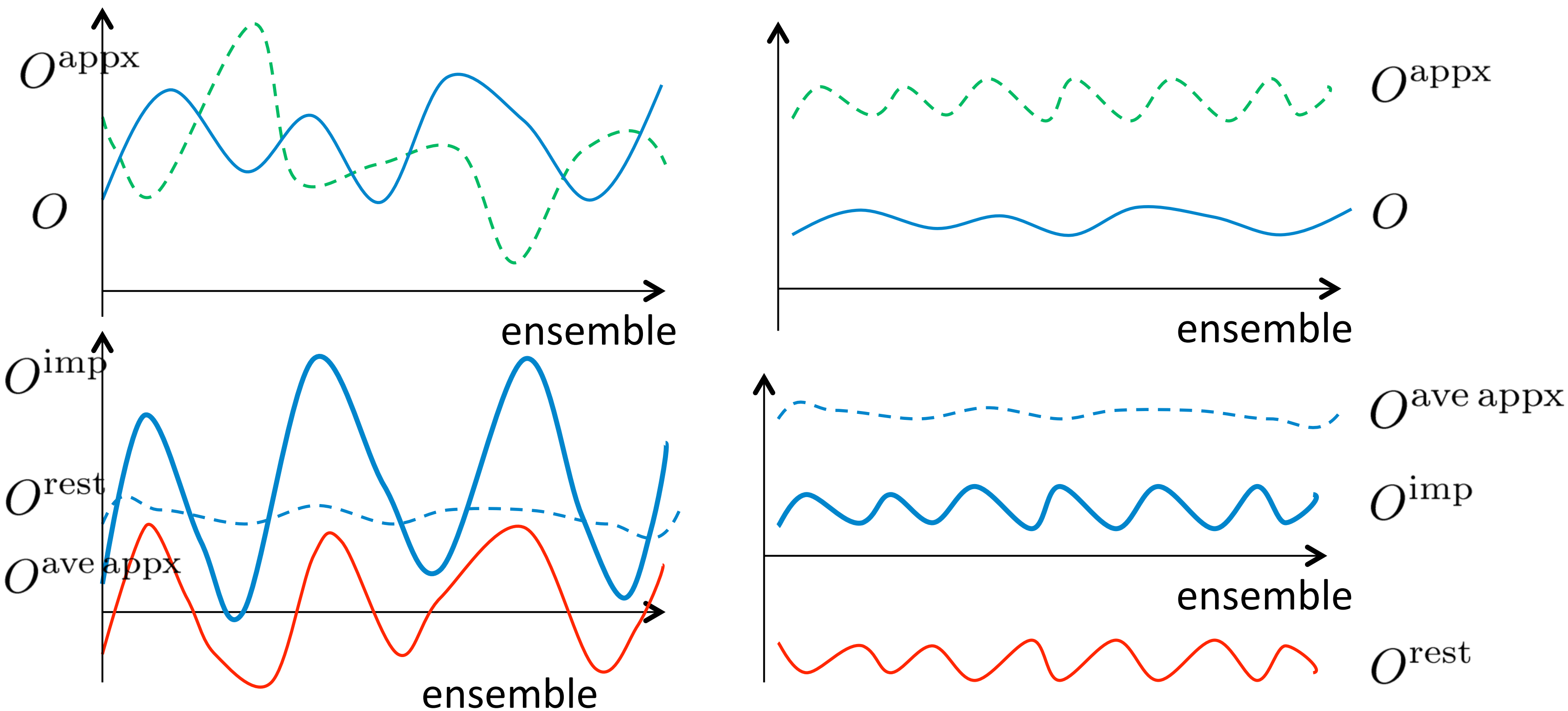
R(corr) b/w O and $O^{(appx)}$ needs to be larger than 0.5 [C. Lehner]

$$\mathrm{err}^{\mathrm{imp}} \simeq \mathrm{err}/\sqrt{N_g}$$

- ε, accuracy of approximation should be smaller than $O^{ave\ appx}$
- ΔO^{rest} which is statistical error of O^{rest} depends on the strength of correlation.
- The computational cost of O^{appx} should be much smaller than original.

CAA : failure senarios

- Approximation is bad for your observables



$$\mathrm{err}^{\mathrm{imp}} \gg \mathrm{err}$$

Summary of CAA

- Three conditions for O$^{(appx)}$:

  1. O$^{(appx)}$ should fluctuate closely with O

$$\langle (\Delta\mathcal{O})^2\rangle \approx \langle (\Delta\mathcal{O}^{(\mathrm{appx})})^2\rangle, \qquad \Delta X = X - \langle X\rangle$$

$$r \equiv \mathrm{Corr}(\mathcal{O}, \mathcal{O}^{(\mathrm{appx})}) = \frac{\langle \Delta\mathcal{O}\Delta\mathcal{O}^{(\mathrm{appx})}\rangle}{\sqrt{\langle(\Delta\mathcal{O})^2\rangle\langle(\Delta\mathcal{O}^{(\mathrm{appx})})^2\rangle}} \approx 1$$

  2. Cost is cheaper (hopefully by a lot)

$$\mathrm{cost}(\mathcal{O}^{(\mathrm{appx})}) < \mathrm{cost}(\mathcal{O})$$

  3. O(appx) is covariant under set of lattice symmetry g in G.
     (should explicitly check this numerically)

$$\mathcal{O}^{(\mathrm{appx})}[U^g] = \mathcal{O}^{(\mathrm{appx}),g}[U], \qquad g \in G$$

$$\mathrm{err}_{(\mathrm{imp})} \approx \mathrm{err}\sqrt{2(1-r) + \frac{1}{N_G}}$$

- Many different Covariant Approximations

  trade off between cost vs accuracy of approximations

- Best way of approximation & optimal accuracy of approximation depends on observables and lattice parameters (lattice spacing, quark mass, volume)

Examples of covariant approximations

- Low mode approximation used in the Low Mode Averaging (**LMA**)

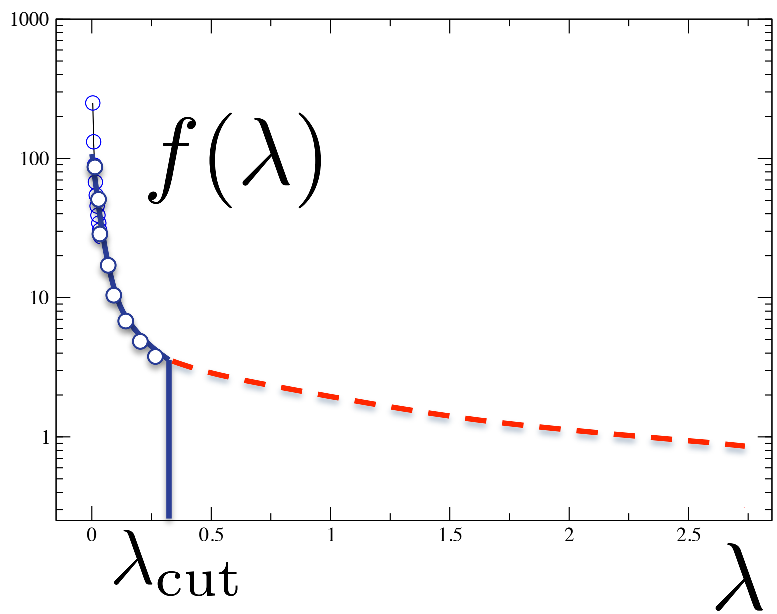
L. Giusti et al (2004), see also T. DeGrand et al. (2004)

accuracy control : # of eigen mode

$$\mathcal{O}^{(\mathrm{appx})} = \mathcal{O}[S_l],$$

$$S_l = \sum_{\lambda} v_\lambda f(\lambda) v_\lambda^\dagger,$$

$$f(\lambda) = \frac{1}{\lambda}\theta(\lambda_{\mathrm{cut}} - |\lambda|)$$



Deflation using low eigenmodes from Lanczos [Neff et al, JLQCD]

- 4D even/odd preconditioning

$$D_{DW} = \begin{pmatrix} M_5 & K(M_4)_{eo} \\ K(M_4)_{oe} & M_5 \end{pmatrix}$$

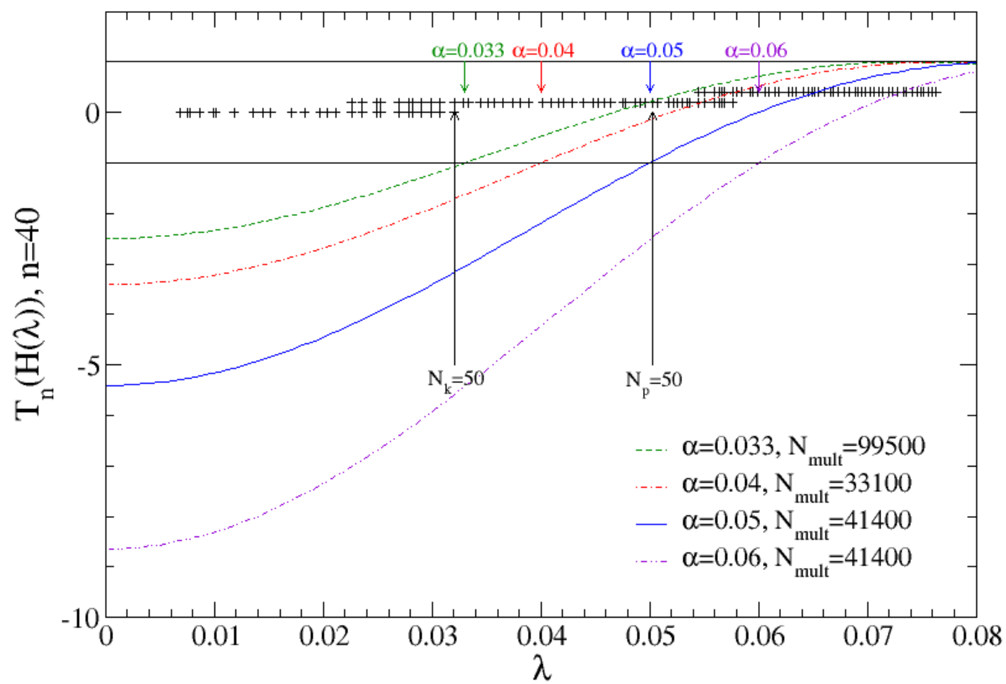
[R. Arthur]

$$D_{DW}^{-1} = \begin{pmatrix} 1 & 0 \\ -KM_5^{-1}(M_4)_{oe} & M_5^{-1} \end{pmatrix} \begin{pmatrix} D_{ee}^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -K(M_4)_{eo}M_5^{-1} \\ 0 & 1 \end{pmatrix}$$

$$D_{ee} = M_5 - K^2(M_4)_{eo}M_5^{-1}(M_4)_{oe}$$

- Polynomial accelerated
  P_n(H_DWF)
- With shift
  H-> H-c
- eigen Compression / decompression

$$\psi = v_1 + v_2$$
$$H(\psi) = \lambda_1 v_1 + \lambda_2 v_2$$



Low-mode decomposition

- 4D even-odd decomposition

$$D_{DW} = \begin{pmatrix} M_{5\,ee} & KM_{4\,eo} \\ KM_{4\,oe} & M_{5\,oo} \end{pmatrix}$$

M_5 : with 5D differential, 4D diagonal
M_4 : with 4D differential, 5D diagonal

$$= \begin{pmatrix} 1 & KM_{4\,eo}M_{5\,oo}^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} D_{ee} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ KM_{4\,oe} & M_{5\,oo} \end{pmatrix}$$

$$D_{ee} = M_5 - K^2 M_{4\,eo} M_{5\,oo}^{-1} M_{4\,oe}$$

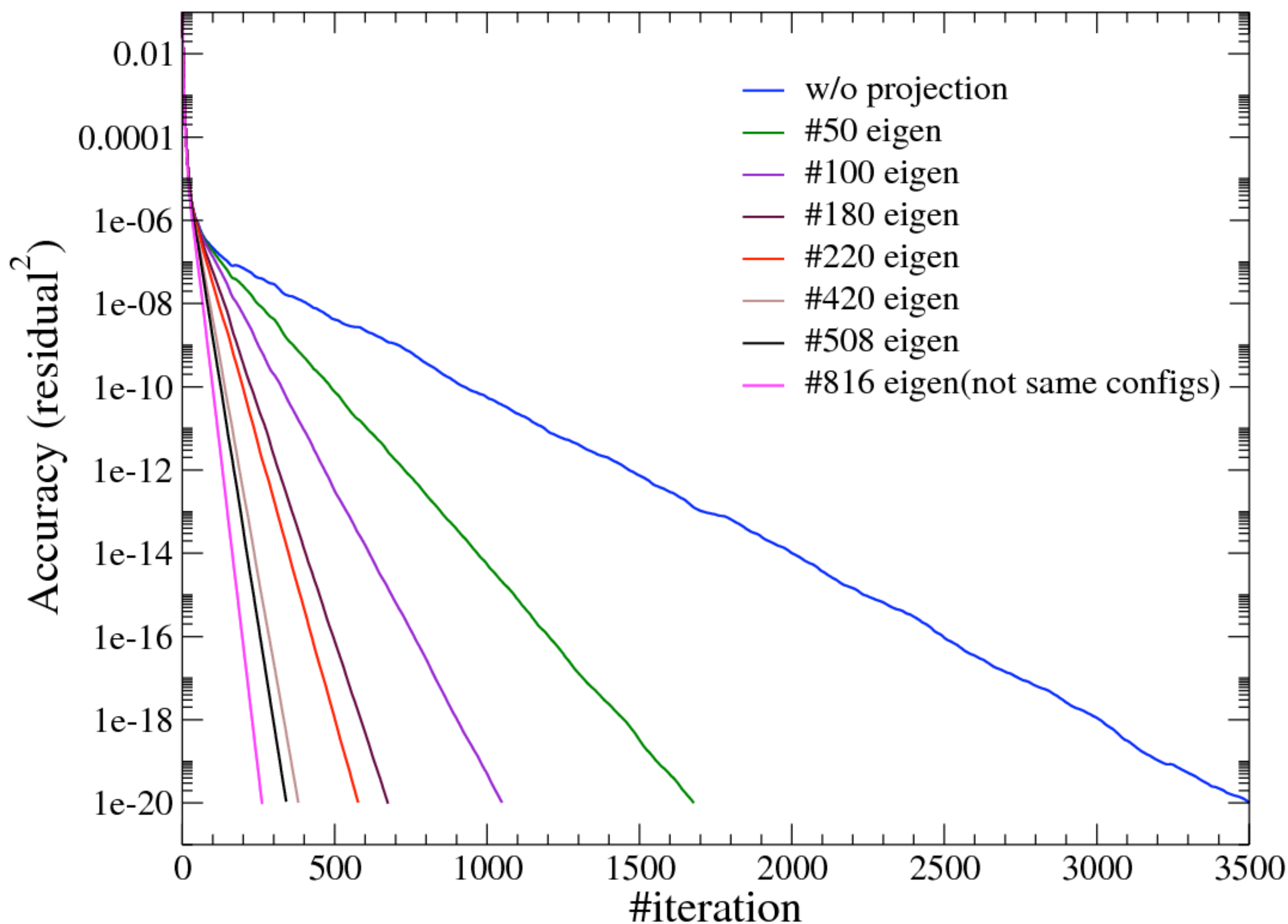
$$D_{DW}^{-1} = \begin{pmatrix} 1 & 0 \\ -KM_{5\,oo}^{-1}M_{4\,oe} & M_{5\,oo}^{-1} \end{pmatrix} \begin{pmatrix} D_{ee}^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -KM_{4\,eo}M_{5\,oo}^{-1} \\ 0 & 1 \end{pmatrix}$$

- Low mode decomposition

$$D_{ee}^{-1} = D_{\mathrm{low}\,ee}^{-1} + D_{\mathrm{high}\,ee}^{-1}$$

$$D_{\mathrm{low}\,ee}^{-1} = H_{\mathrm{low}\,ee}^{-2} D_{ee}^{\dagger} = \sum_k \frac{1}{\lambda_k^2} \psi_k (D_{ee}\psi_k)^{\dagger}, \quad H_{ee}\psi_k = \lambda_k \psi_k, \quad H_{ee} = \Gamma_5 D_{ee}$$

$$D_{\mathrm{low}\,DW}^{-1} = \begin{pmatrix} 1 & 0 \\ -KM_{5\,oo}^{-1}M_{4\,oe} & M_{5\,oo}^{-1} \end{pmatrix} \begin{pmatrix} D_{\mathrm{low}\,ee}^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -KM_{4\,eo}M_{5\,oo}^{-1} \\ 0 & 1 \end{pmatrix}$$



w/o projection
#50 eigen
#100 eigen
#180 eigen
#220 eigen
#420 eigen
#508 eigen
#816 eigen(not same configs)
Accuracy (residual^2)
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
1e-18
1e-20
0
500
1000
1500
2000
2500
3000
3500
#iteration

Examples of Covariant Approximations (contd.)

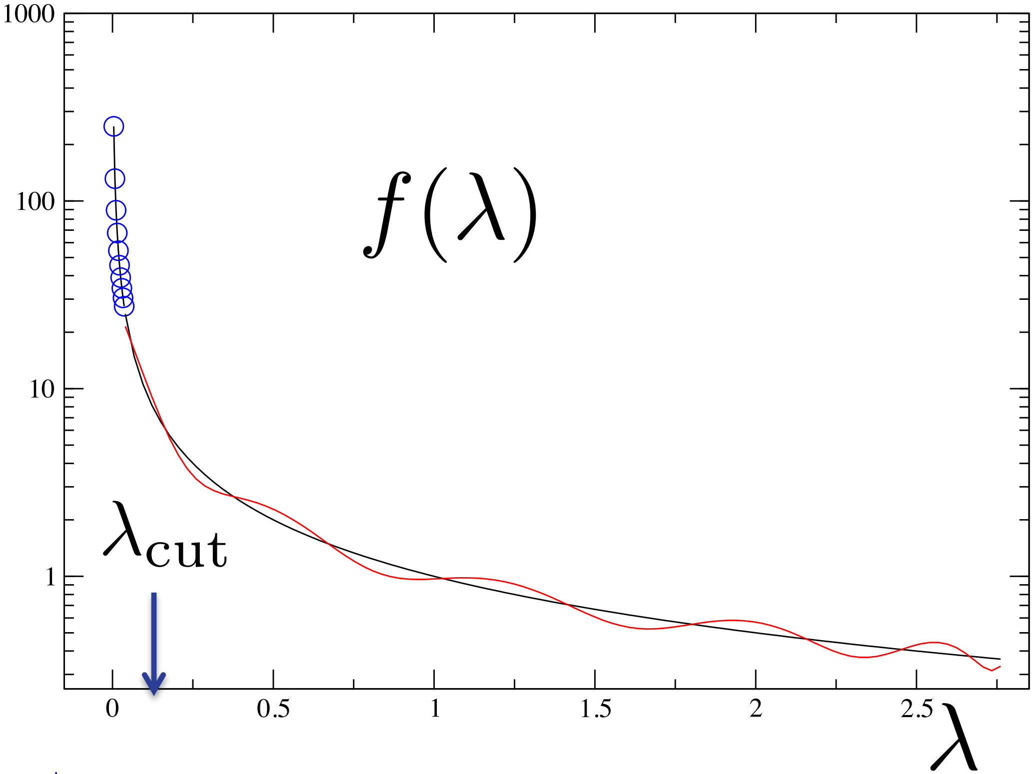
- All Mode Averaging **AMA**
  Sloppy CG or Polynomial approximations

$$\mathcal{O}^{(\mathrm{appx})} = \mathcal{O}[S_l],$$

$$S_l = \sum_\lambda v_\lambda f(\lambda) v_\lambda^\dagger,$$

$$f(\lambda) = \begin{cases} \frac{1}{\lambda}, & |\lambda| < \lambda_{\mathrm{cut}} \\ P_n(\lambda) & |\lambda| > \lambda_{\mathrm{cut}} \end{cases}$$

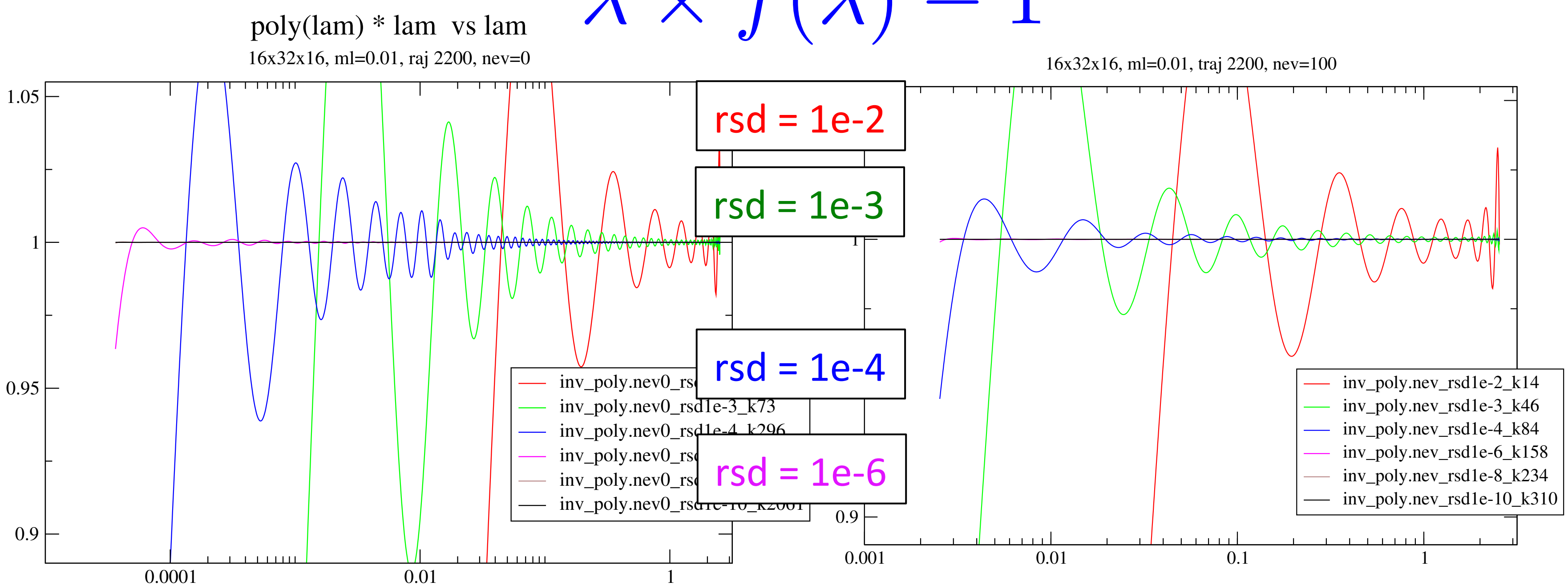
$$P_n(\lambda) \approx \frac{1}{\lambda}$$



accuracy control :
- low mode part : # of eig-mode
- mid-high mode : degree of poly.

All mode approximation via sloppy CG

$$\lambda \times f(\lambda) - 1$$



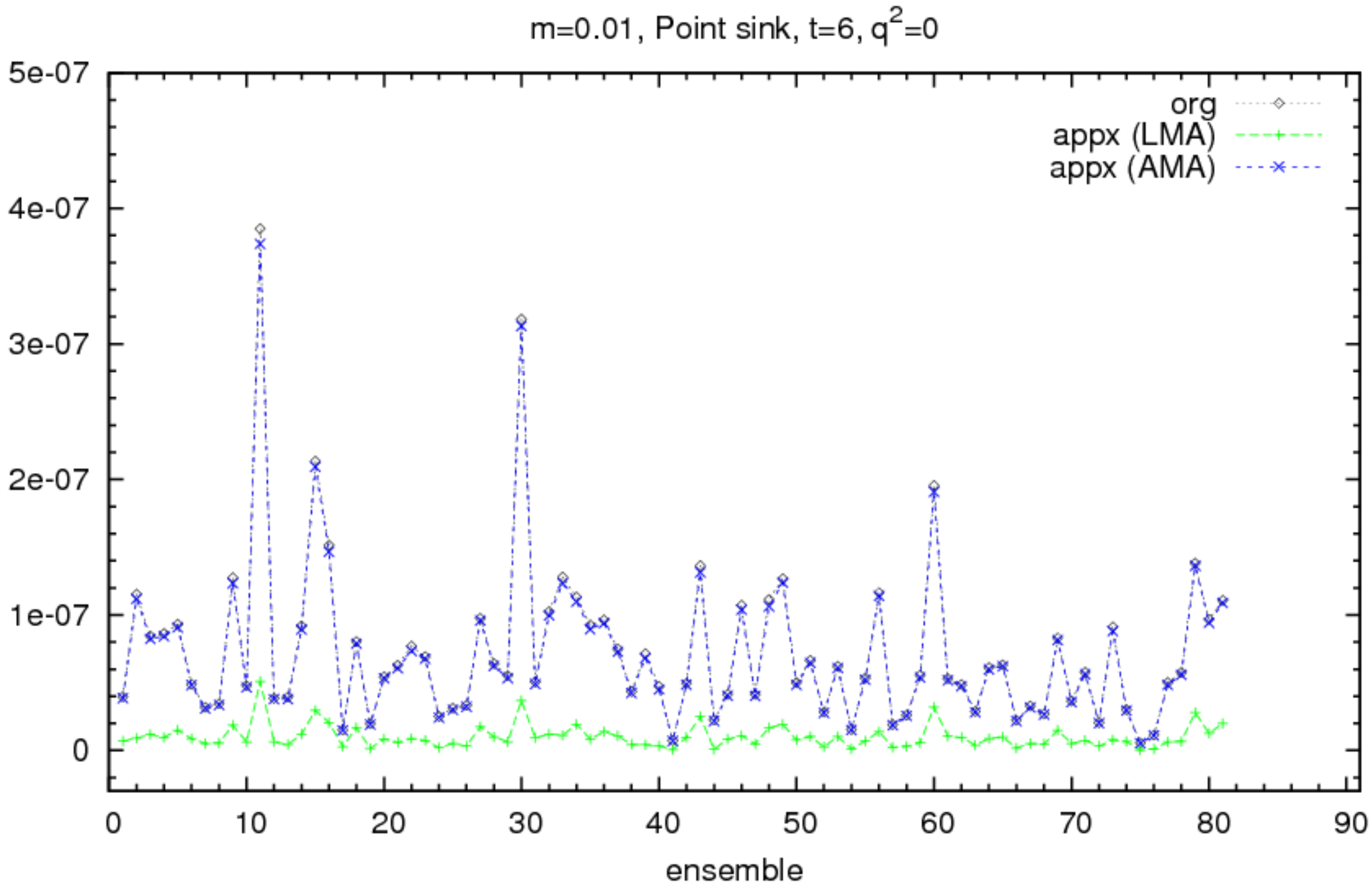
no eigenvector assists

100 eigenvector assists

- Conjugate residual with sloppy convergence criteria, which is equivalent to construct a polynomial approximating 1/λ
- The starting vector needs to be translation invariant to be a covariant approx.
- low eigenvectors reduces the size of the dynamic range of 1/λ
  - → Better approximation with smaller polynomial degrees
- low λ region has larger relative errors
- One could employ other construction of polynomial approximation for 1/λ, such as min-max, conjugate rasidual

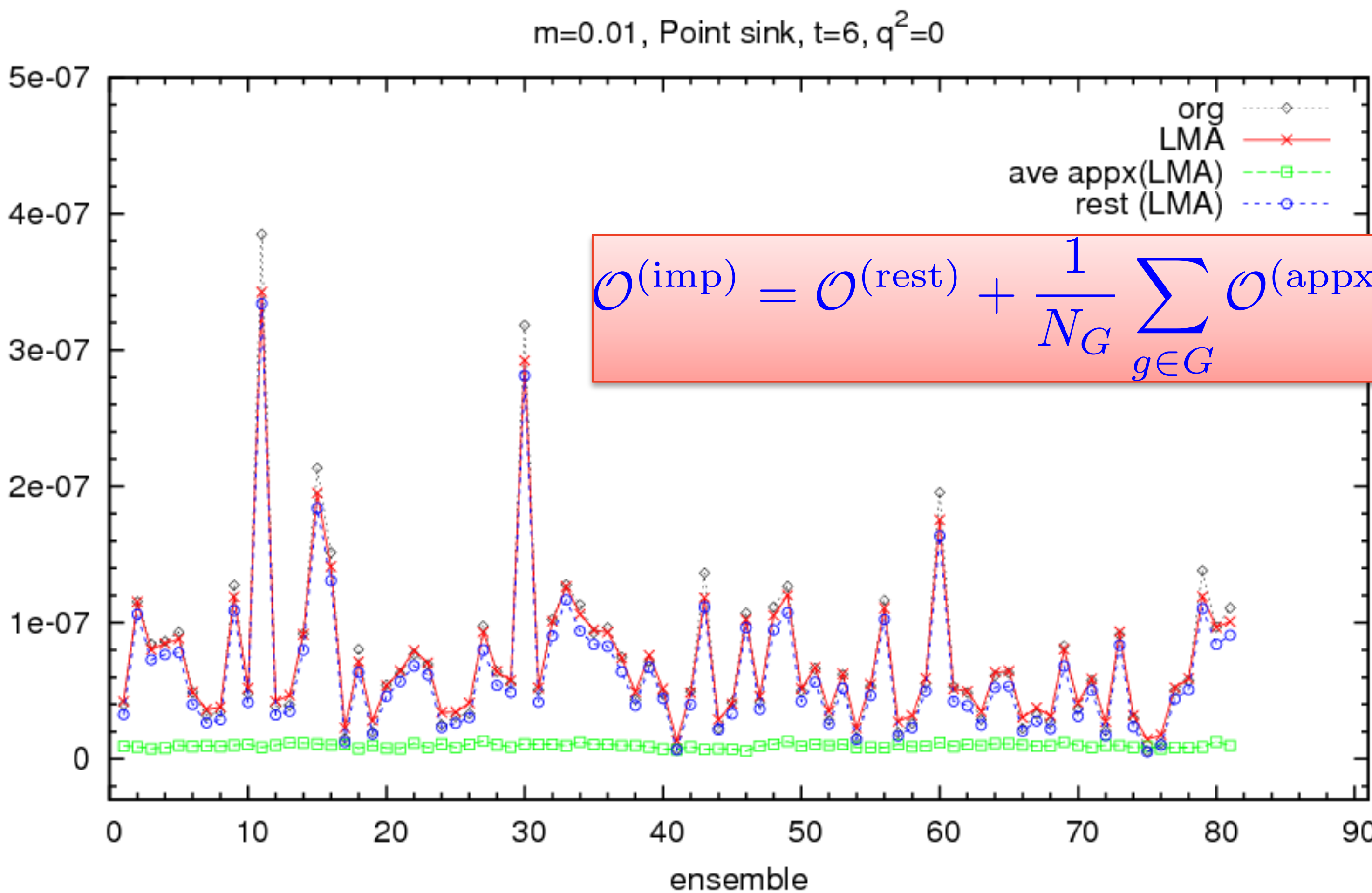
Correlation

- NN propagator at short time-slice



Correlation

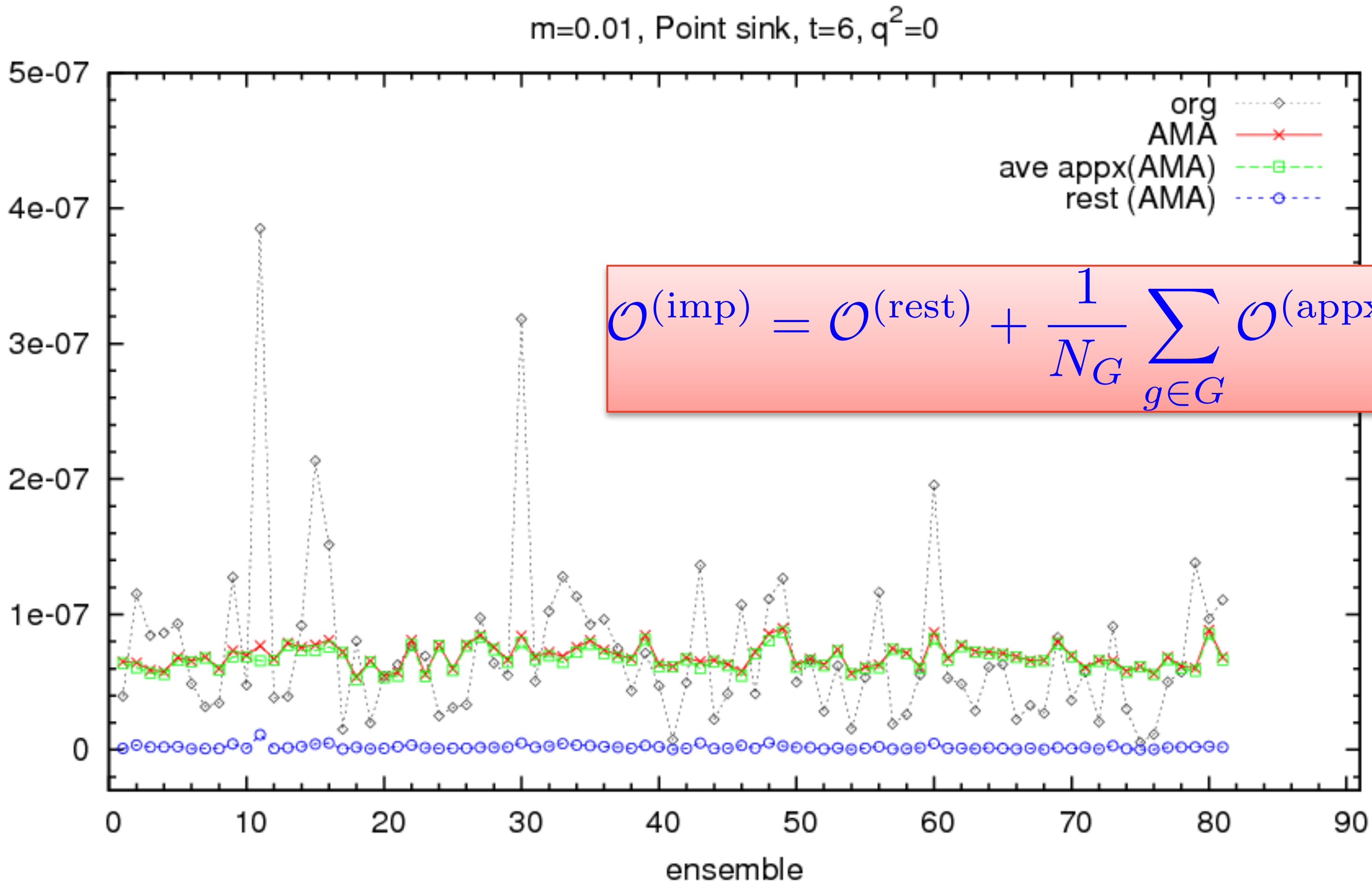
- NN propagator (LMA) at short time-slice



$$\mathcal{O}^{(\mathrm{imp})} = \mathcal{O}^{(\mathrm{rest})} + \frac{1}{N_G}\sum_{g\in G}\mathcal{O}^{(\mathrm{appx}),g}$$

Correlation

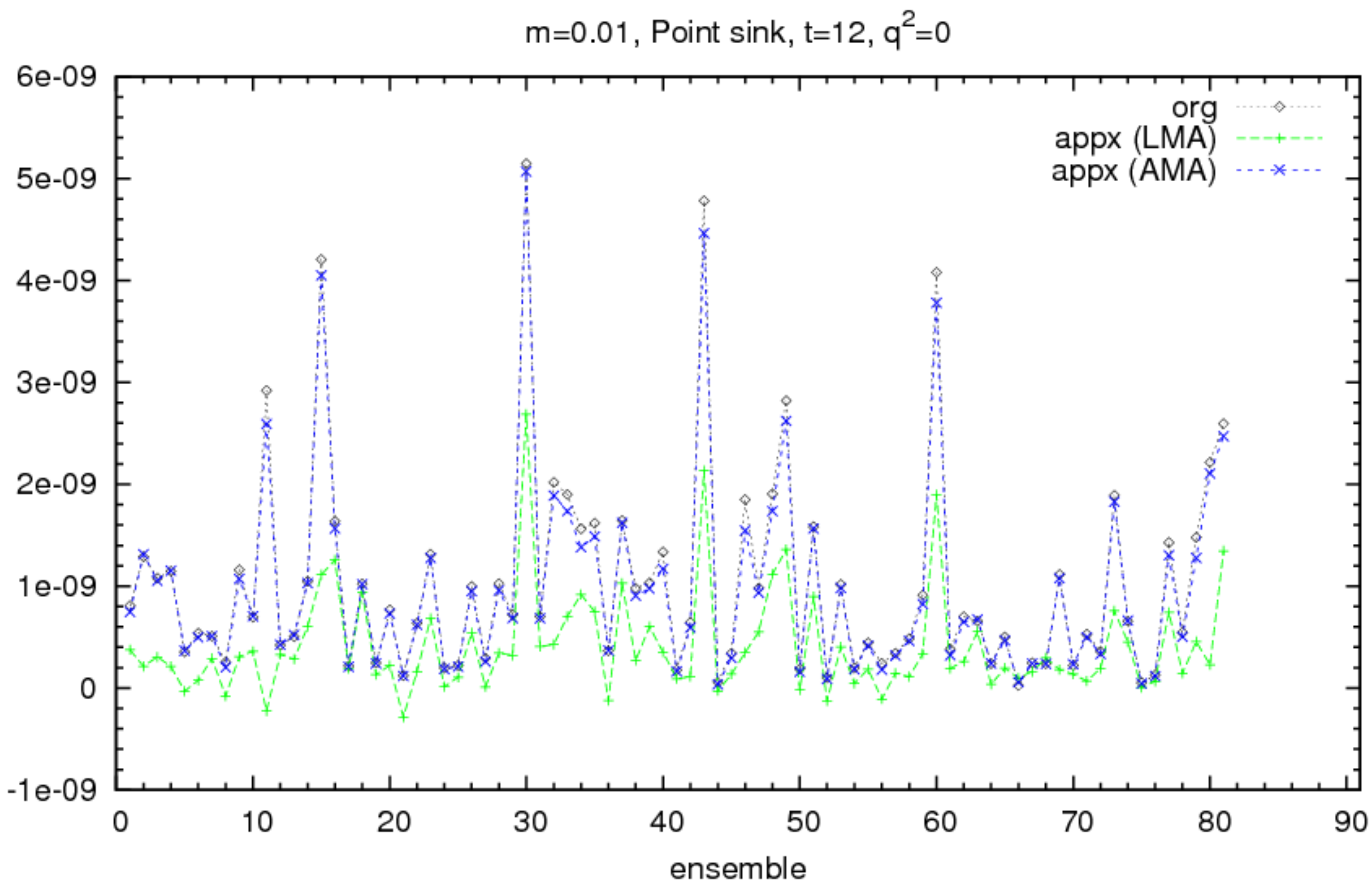
- NN propagator (AMA) at short time-slice



$$\mathcal{O}^{(\mathrm{imp})} = \mathcal{O}^{(\mathrm{rest})} + \frac{1}{N_G} \sum_{g \in G} \mathcal{O}^{(\mathrm{appx}),g}$$

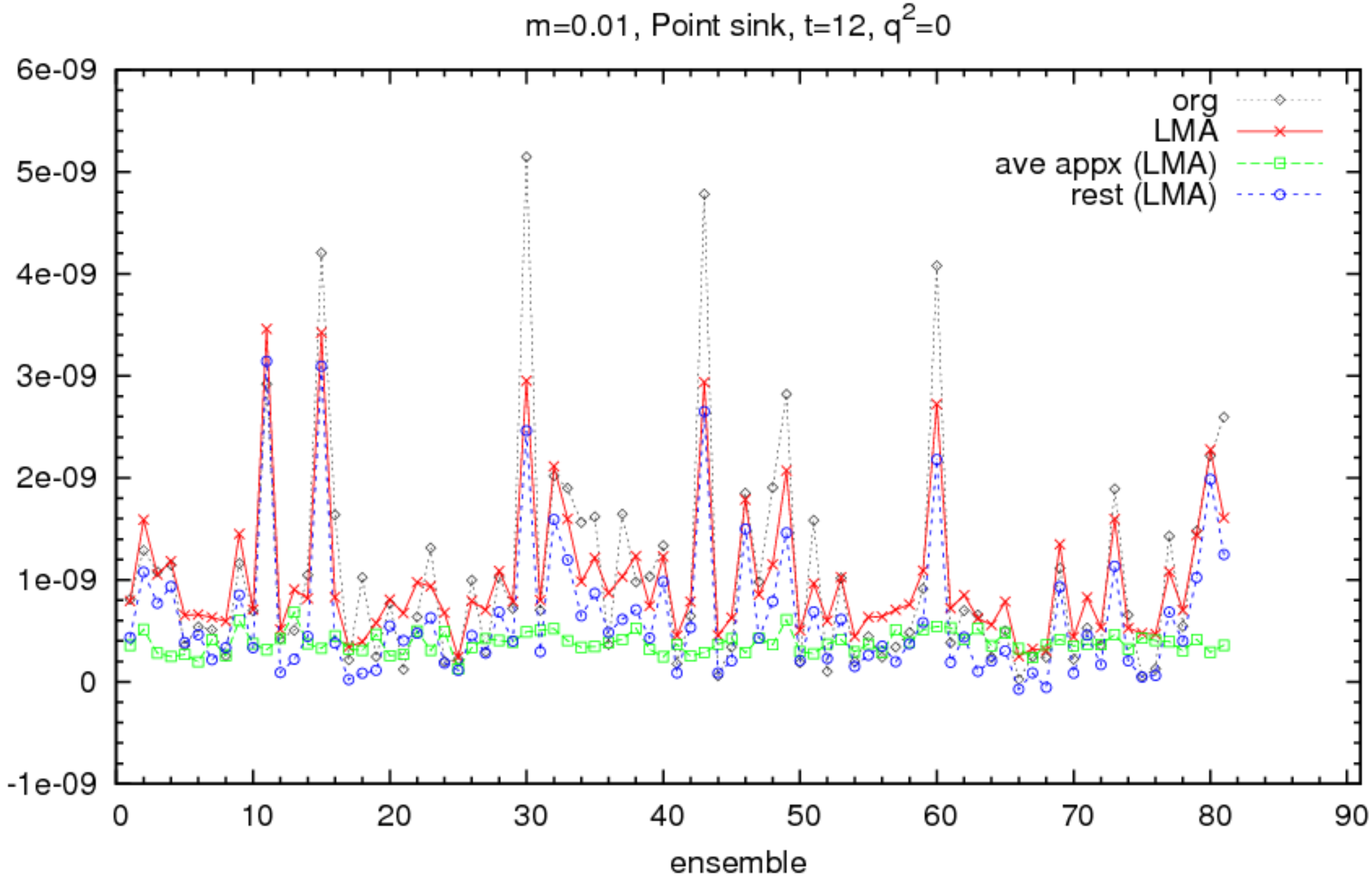
Correlation

- NN propagator at long time-slice



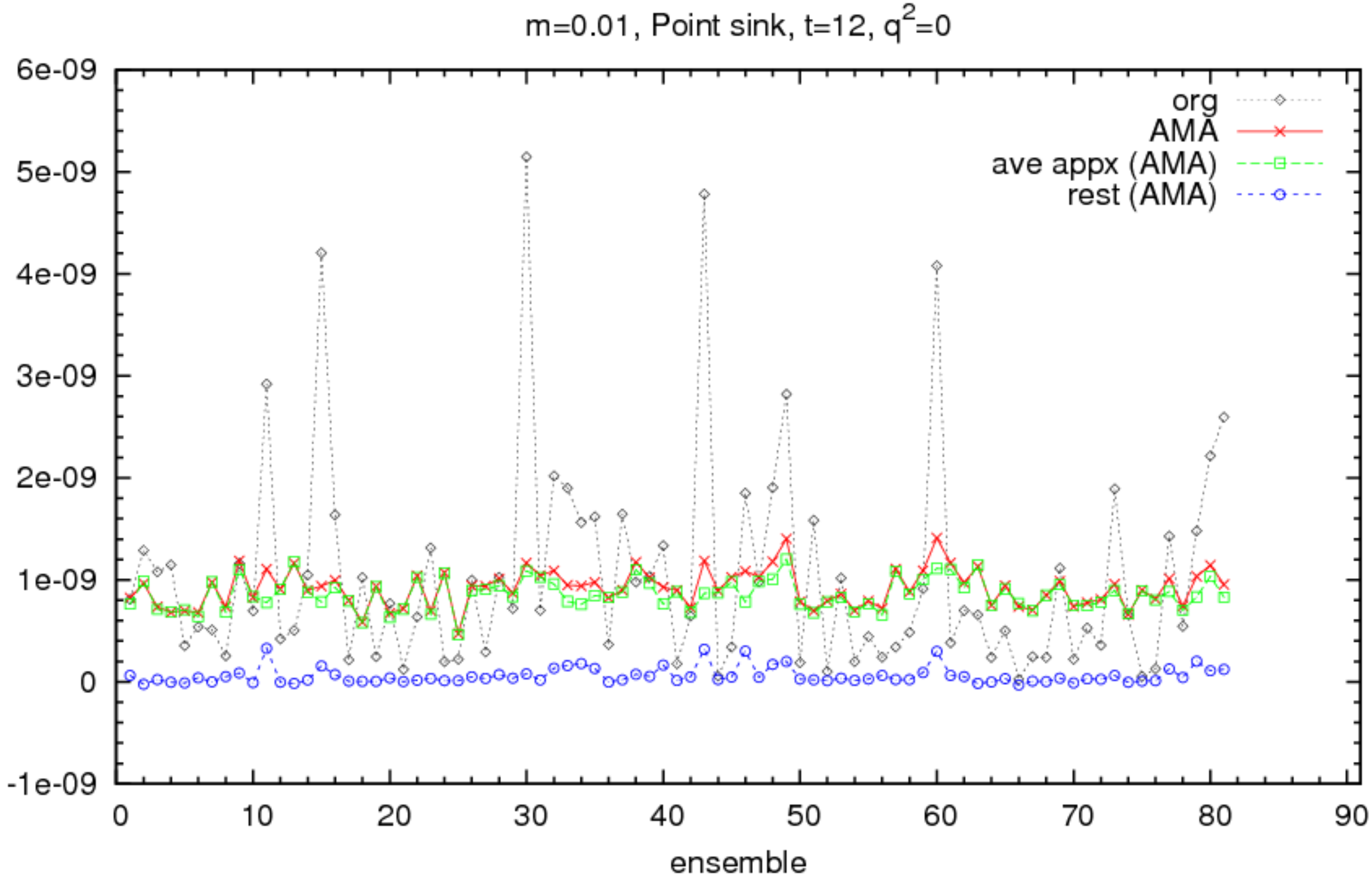
Correlation

- NN propagator (LMA) at long time-slice

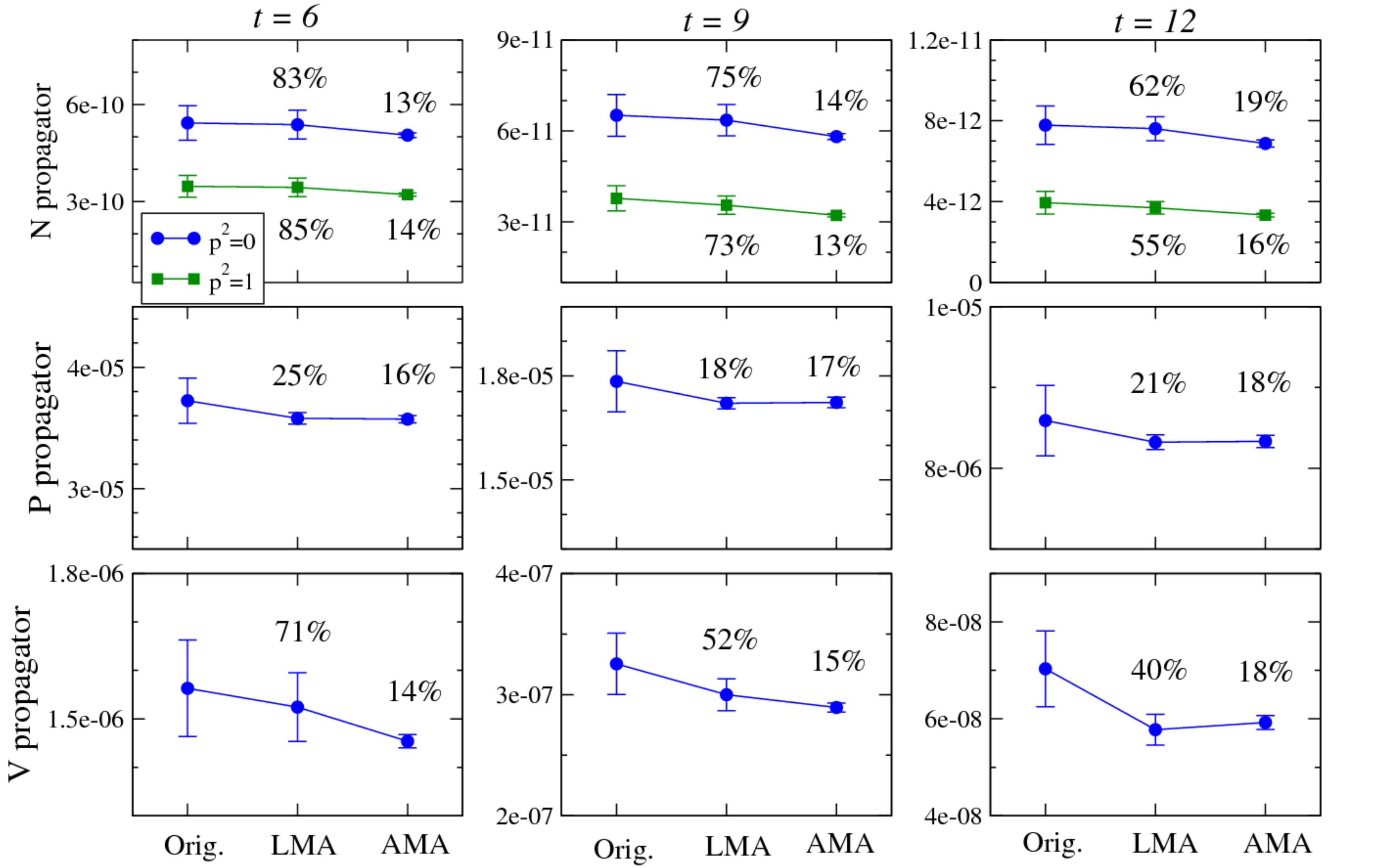


Correlation

- NN propagator (AMA) at long time-slice

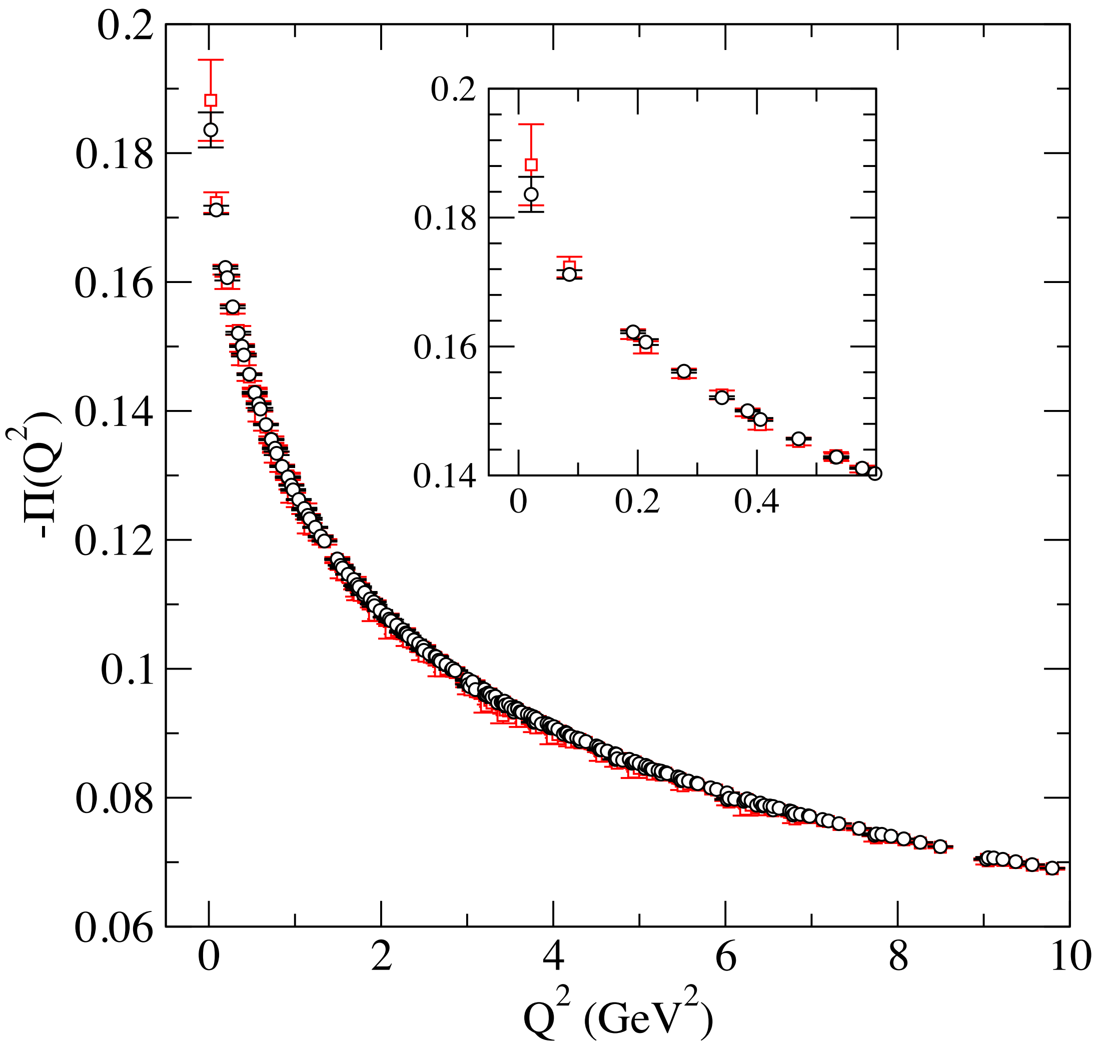


AMA results for hadron 2pt functions [E. Shintani] (NG=32)



Nucleon 2pt, m=0.005 DWF, 24cube, Gauss src, Gauss / point sink

HVP (c.f. M. Golterman's talk)



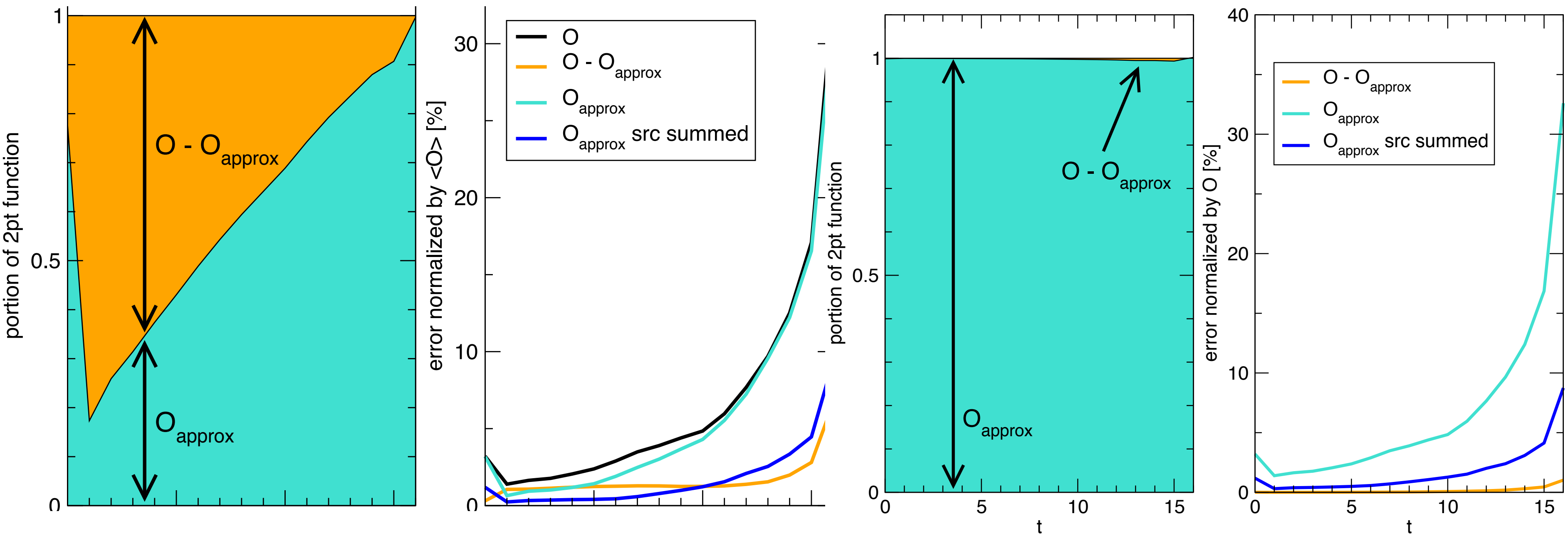
Cost (in the case of 24cube m=0.01)

	$N_{\rm conf}$	$N_{\rm meas}$	LM	$\mathcal{O}$	$\mathcal{O}_G^{\rm (appx)}$	Tot.	scaled cost	
m_N			$m = 0.005$, 400 LM				gauss	pt
AMA	110	1	213	18	91+23	350	0.063	0.065
LMA	110	1	213	18	23	254	0.279	0.265
Ref. [2]	932	4	-	3728	-	3728[a]	1	1
			$m = 0.01$, 180 LM					
AMA	158	1	297	74	300+22	693	0.203	0.214
LMA	158	1	297	74	22	393	0.699	0.937
Ref. [2]	356	4	-	1424	-	1424	1	1
HVP			$m = 0.0036$, 1400 LM				max	min
AMA	20	1	96	11	504+420	1031	0.387	0.050
LMA	20	1	96	11	420	527	10.3	3.56
Ref. [27]	292	2	-	584	-	584	1	1

[2] Yamazaki et al., PRD79, 114505 (2009)

AMA in USQCD Static-light
[PI Tomomi Ishikawa]

16^3x64x16, 20 conf, 100 eigenvectors



LMA

AMA

3pt function [E. Shintani]

- Application to the form factor measurement
  - CP-even and CP-odd nucleon EM form factor

$$\langle n(P_1)|J_\mu^{\rm EM}|n(P_2)\rangle_\theta = \bar{u}_N^\theta \Big[\underbrace{\frac{F_3^\theta(Q^2)}{2m_N}\gamma_5\sigma_{\mu\nu}Q_\nu}_{\text{P,T-odd}} + \underbrace{F_1\gamma_\mu + \frac{F_2}{2m_N}\sigma_{\mu\nu}Q_\nu}_{\text{P,T-even}} + \cdots \Big] u_N^\theta$$

  - Complicated structure in the ratio method

Cf. Yamazaki et al., PRD79, 114505 (2009)

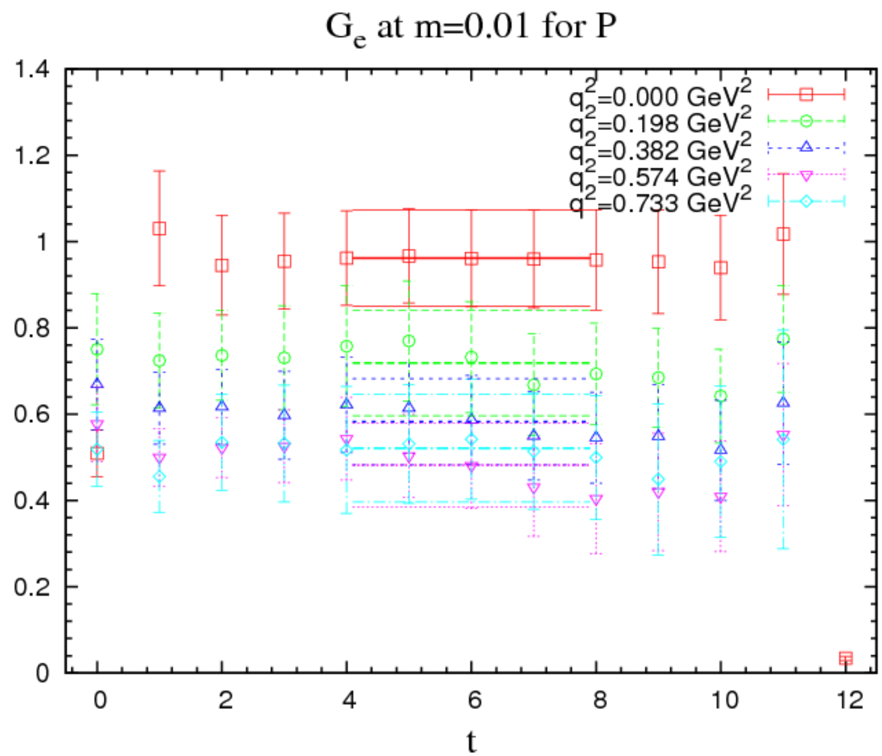
$$R_{J_\mu}(t,\vec{q}) = \sqrt{\frac{m_N}{2(E_N+m_N)}}\frac{\langle \eta_N^g J_\mu \bar{\eta}_N^g\rangle(t,\vec{q})}{\langle \eta_N^l \bar{\eta}_N^g\rangle(t_{\rm snk}-t_{\rm src},0)} R(t,\vec{q}),$$

$$R(t,\vec{q}) = \Big[\frac{\langle \eta_N^l \bar{\eta}_N^g\rangle(t_{\rm snk}-t,\vec{q})\langle \eta_N^g \bar{\eta}_N^g\rangle(t-t_{\rm src},0)\langle \eta_N^l \bar{\eta}_N^g\rangle(t_{\rm snk}-t_{\rm src},0)}{\langle \eta_N^l \bar{\eta}_N^g\rangle(t_{\rm snk}-t,0)\langle \eta_N^g \bar{\eta}_N^g\rangle(t-t_{\rm src},\vec{q})\langle \eta_N^l \bar{\eta}_N^g\rangle(t_{\rm snk}-t_{\rm src},\vec{q})}\Big]^{1/2}$$

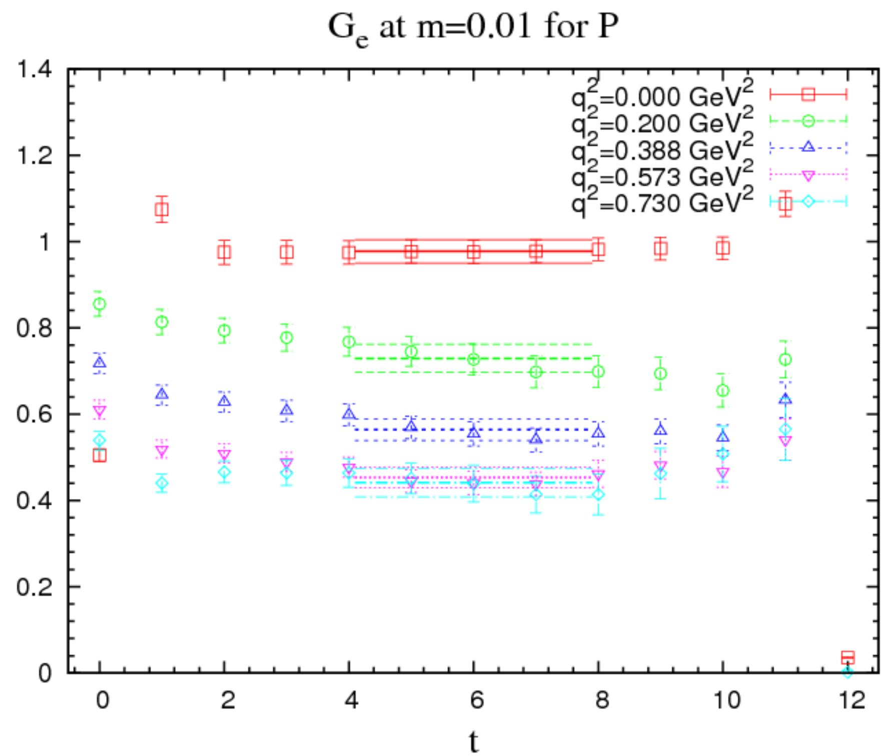
Ratio has complicated combination of both low and high mode,

so AMA has more advantage than LMA even if AMA need larger cost.

LMA



AMA

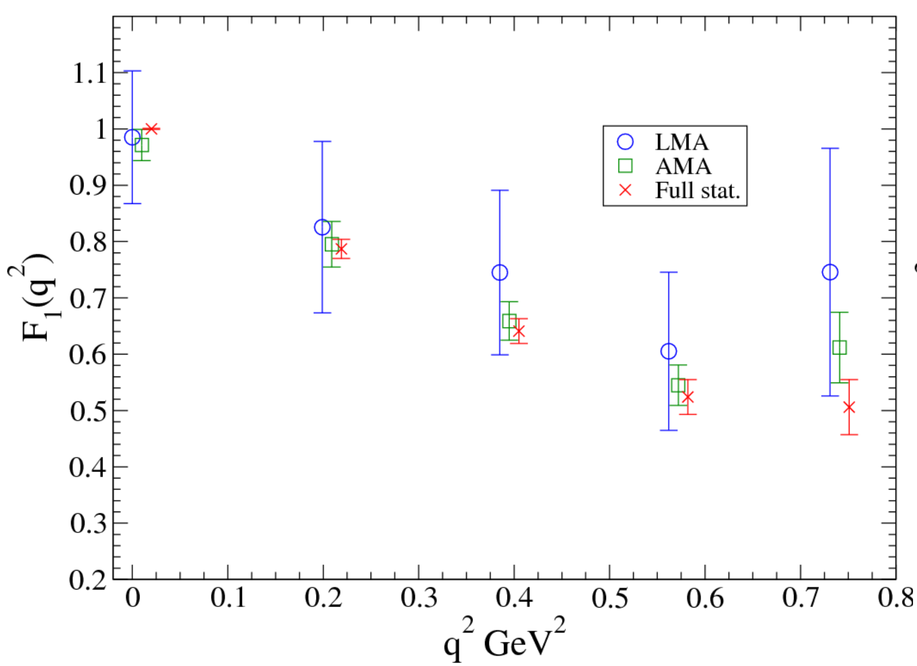


q^2 GeV^2	G_e (LMA)	G_e (AMA)
0.0	0.96(11)	0.98(3)
0.198	0.72(12)	0.73(3)
0.382	0.58(10)	0.56(3)
0.574	0.48(10)	0.45(2)
0.733	0.52(12)	0.44(3)

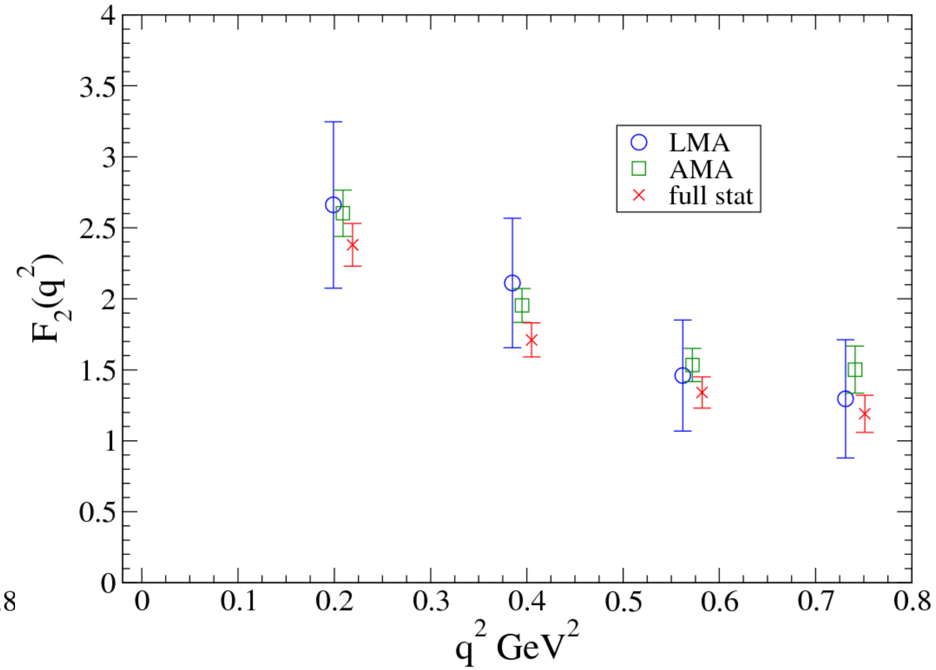
Statistical error of AMA is about 3--5 times smaller than LMA.

Comparison of isovector $F_{1,2}$ [E. Shintani]

m=0.01



m=0.01



- Results are well consistent with full statistics.
- Statistical error is much reduced in AMA rather than LMA.
- Compared to full statistics, AMA results (m=0.01) have still 1.2 -- 1.5 times larger statistical error (except for $F_1(0)$).
- This may be due to correlation between different source points.

CP-odd part

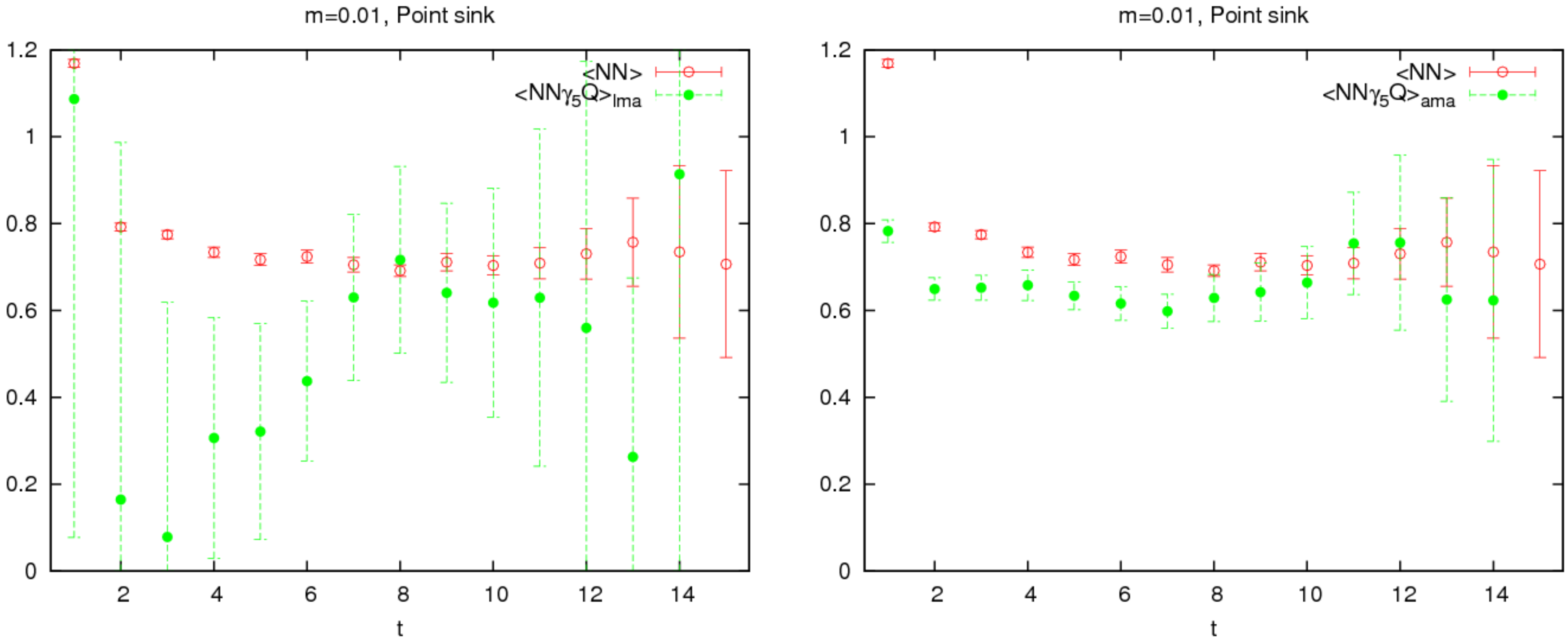
- Nucleon 2pt function with θ reweighting

$$\langle \eta_N \bar{\eta}_N \rangle_\theta(\vec{p}) = Z_N^2 \frac{ip \cdot \gamma + m_N e^{i\alpha(\theta)\gamma_5}}{2E_N}$$

$$\mathrm{tr}\left[\gamma_5 \langle Q \eta_N \bar{\eta}_N \rangle(\vec{p})\right] \simeq Z_N^2 \frac{2m_N}{E_N} \alpha e^{-E_N t}$$

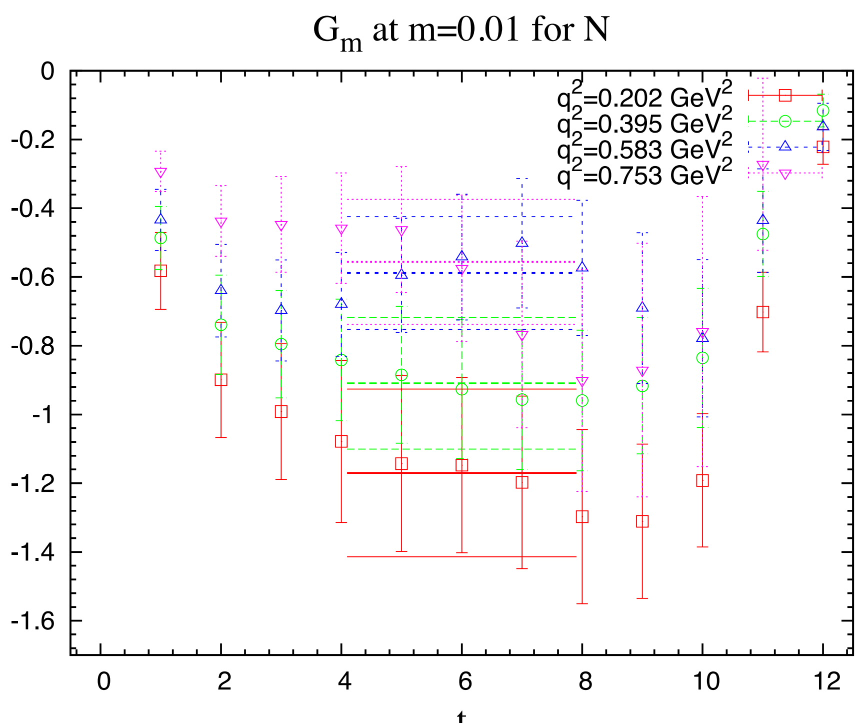
  - Q is topological charge.
  - α which is CP-odd phase is necessary to extract EDM form factor.
  - It is good check of applicability of LMA/AMA to CP-odd sector.
  - Effective mass plot shows the consistency of the above formula

CP-odd part [E. Shintani]

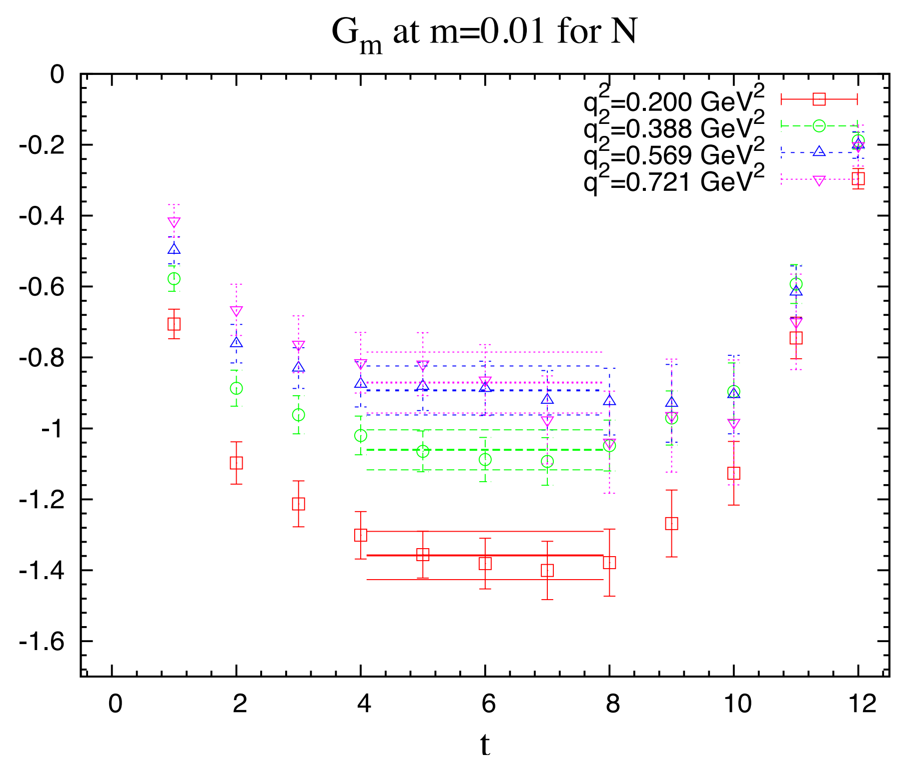


- There is good plateau in AMA, and this figure actually shows CP-odd part has consistent exponent with CP-even(nucleon mass) part as expected.
- CP-odd part has both contribution from high and low lying mode.
- AMA works well even in CP-odd sector !

Nucleon Magnetic formfactor



Original CG



AMA

Variants of CAA

- CAA (Covariant Approximation Averaging)
  - Name
    approximation,
    approximation accuracy control
  - LMA (Low Mode Averaging)
    low mode approx of propagator,
    # of eigen vectors
  - AMA (All Mode Averaging),
    low mode (optional)+Polynomial approx,
    (# of eigenV) Polynomial degree
    (also other type of minimization)
  - Heavy quark averaging [T. Kawanai]
    heavier mass quark prop as an approx of light prop
    quark mass
  - ?????

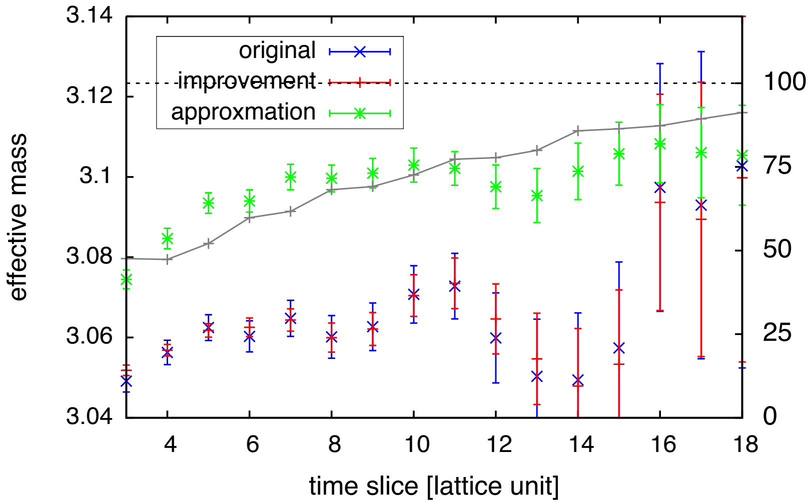
Other Examples of Covariant Approximations

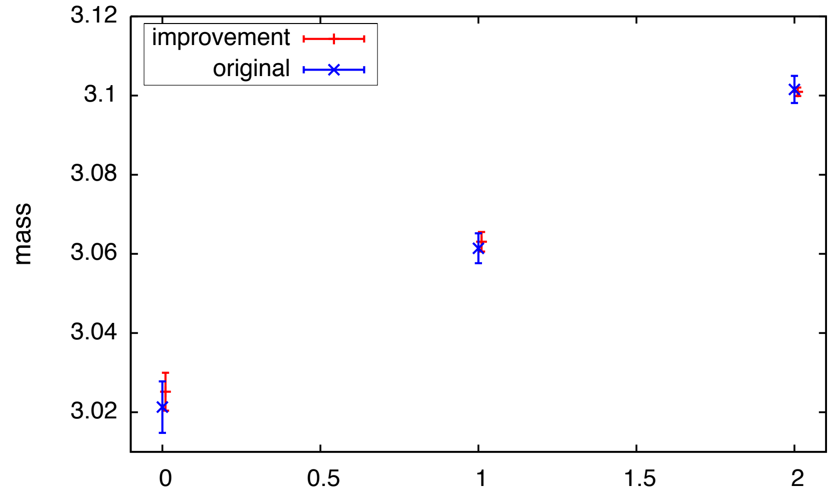
- Less expensive (parameters of) fermions :
  - Different polynomial approximation than that of CG
  - Larger mf
  - Smaller Ls DWF
  - Different boundary conditions
  - Mobius
  - even staggered or Wilson …..
- Use other part of configuration only for $O^{(appx)}$ may be useful for disconnected loops / bulk observables
- More than one kinds of approximation (c.f. multi mass Hasenbushing)

Strongly depends on Observables / Physics (YMMV)
Would work better for EXPENSIVE observables and/or fermion, potentially a game changer ?

Larger mass as CAA
[Taichi Kawanai]

24^3x64x16, 20 config ,
mf=0.01 (target) mf=0.04 “approximation”





Further optimized approximation

- Further optimized approximation based on data could be considered. For example,

$$\mathcal{O}^{(\mathrm{appx})} \to \mathcal{O}^{(\mathrm{appx})'} = C\mathcal{O}^{(\mathrm{appx})}$$

$$C = \sqrt{\frac{\langle(\Delta\mathcal{O}^{(\mathrm{appx})})^2\rangle}{\langle(\Delta\mathcal{O})^2\rangle}}$$

which suppresses fluctuation of $O^{(\mathrm{rest})}$ further.

- One needs to make sure $O^{(\mathrm{appx})'}$ is covariant to avoid bias from the refinement.
- One could use different set of ensemble for tuning (computing C above) and the actual computation

Other related/similar techniques

- LMA

  L. Giusti, P. Hernandez, M. Laine, P. Weisz and H. Wittig, JHEP 0404, 013 (2004)
  see also H. Neff, N. Eicker, T. Lippert, J. W. Negele and K. Schilling, Phys. Rev. D 64 (2001) 114509 and T. DeGrand and S. Schaefer, Comput. Phys. Commun. 159 (2004) 185

  works well for low mode dominant quantities

- Truncated Solver Method (TSM)

  G. Bali, S. Collins, A. Schaefer, Comput. Phys. Commun. 181 (2010) 1570

  uses stochastic noise to avoid systematic error

- All-to-all propagator

  J.Foley, K.Juge, A. O'Cais, M. Peardon, S. Ryan, J-I. Skullerud, Comput.Phys.Commun. 172 (2005) 145

  uses stochastic noise
  could use CAA as a part of A2A

Summary

- CAA , LMA, AMA, ... :
  A new Class of Statistical error reduction technique

  - AMA is a valence version of the Hasenbush trick
  - AMA could improve existing data easily

  - 16 times less cost for DWF nucleon mass (3fm, 330 MeV pion)

  - 2.6—20 times less cost for HVP on AsqTadsta (4-5 fm 315 MeV pion)

- YMMV, find a good / cheap / funny approximations

Other technical details

- Implicitly Restarted Lanczos with Polynomial acceleration and spectrum shifts for DWF and staggered in CPS++ [E. Shintani, T. Blum, TI].
- Eigen Vector compression / decompression

- Sea Electric Charge is now controlled by QED reweighting
  [T. Ishikawa et. al. arXiv:1202.6018]
- Aslash-SeqSrc method