All-Mode-Averaging with Approximate Eigenvectors for Twisted-Mass Fermions

A. M. Abdel-Rehim

Computation-based Science and Technology Research Center
The Cyprus Institute

with
C. Alexandrou, G. Koutsou

QCDNA 2014, Yale University, 19-21 June 2014

Covariant Approximate Averaging (CAA)

E. Shintani et. al. arXiv:1402.0244

- Goal:How to reduce the statistical error of lattice correlation functions for a given number of gauge configurations at a low computational cost?
- Approach:
  1. Use symmetrires of the correlation functions to increase statistics.
  2. use a technique called "*All Mode Averaging*" or AMA to reduce the computational cost.
- Hence the name Covariant Approximate Averaging or CAA.
- The method is widely applicable and is tested for quantities of interest such as pion, nucleon and vector meson masses on 2+1 Domain-Wall configurations and is shown to reduce the cost.

Covariant Approximation Averaging (CAA)

- Let $\mathcal{O}[U]$ be some correlator (hadron propagator).
- Let G be a group of symmetry transformations of the action where:
  $g : x \to x^g$
  $g : U(x) \to U^g(x) = U(x^g)$
  $g : \mathcal{O}[U](x,y) \to \mathcal{O}^g[U](x,y) = \mathcal{O}[U](x^g, y^g)$.
- Because G is a symmetry: $\langle \mathcal{O}^g[U] \rangle = \langle \mathcal{O}[U^g] \rangle$
- Since U^g has the same probability weight as U:

$$\langle \mathcal{O}^g[U] \rangle = \langle \mathcal{O}[U] \rangle \tag{1}$$

- If $\mathcal{O}^g[U] = \mathcal{O}[U^g]$ on each configuration then:

$$\sum_{g \in G} \mathcal{O}^g[U] = \sum_{g \in G} \mathcal{O}[U^g] \tag{2}$$

- Define:

$$\mathcal{O}_G[U] \equiv \frac{1}{N_G} \sum_{g \in G} \mathcal{O}^g[U] = \frac{1}{N_G} \sum_{g \in G} \mathcal{O}[U^g] \tag{3}$$

- We have:

$$\langle O_G[U] \rangle = \langle O[U] \rangle \tag{4}$$

- However, statistical error of $\mathcal{O}_G$ decreases by a factor $1/\sqrt{N_G}$.
- Direct evaluation of $\langle O_G[U] \rangle$ requires N_G times extra computational cost.

Reducing the cost of Covariant Averaging

- Replace $\mathcal{O}^g[U]$ by and approximation $\mathcal{O}^{(appx)g}[U]$ such that covariance still holds.
- Similarly, replace $O_G[U]$ by $O_G^{(appx)}[U]$.
- Define an improved estimator for $\mathcal{O}$ by

$$\begin{aligned} \mathcal{O}^{(\mathrm{imp})} &= \mathcal{O} - \mathcal{O}^{(\mathrm{appx})} + \mathcal{O}_G^{(\mathrm{appx})} \\ &\equiv \mathcal{O}^{(\mathrm{rest})} + \mathcal{O}_G^{(\mathrm{appx})}, \qquad (5) \\ \mathcal{O}^{(\mathrm{rest})} &= \mathcal{O} - \mathcal{O}^{(\mathrm{appx})}, \qquad (6) \end{aligned}$$

- Again we can see: $\langle \mathcal{O}^{(\mathrm{imp})} \rangle = \langle \mathcal{O} \rangle$.

Error in $\langle \mathcal{O}^{(\mathrm{imp})} \rangle$

- The standard deviation of the improved estimator is:

$$\sigma^{(\mathrm{imp})} \simeq \sigma \left[2\Delta r + \frac{1}{N_G} - \frac{2}{N_G} \Delta r + R^{\mathrm{corr}} \right]^{1/2}, \tag{7}$$

$$R^{\mathrm{corr}} = \frac{1}{N_G^2} \sum_{g \neq g'} r^{\mathrm{corr}}_{gg'}, \tag{8}$$

- $\sigma^X = \sqrt{\langle (\Delta \mathcal{O}^X)^2 \rangle}$, and $\Delta \mathcal{O}^X = \mathcal{O}^X - \langle O^X \rangle$,
- $r_g = \frac{\langle \Delta \mathcal{O} \Delta \mathcal{O}^{(\mathrm{appx})\, g} \rangle}{\sigma \sigma_g^{(\mathrm{appx})}}$
- $r^{\mathrm{corr}}_{gg'} = \frac{\langle \Delta \mathcal{O}^{(\mathrm{appx})\, g} \Delta \mathcal{O}^{(\mathrm{appx})\, g'} \rangle}{\sigma^{(\mathrm{appx})\, g} \sigma^{(\mathrm{appx})\, g'}}$
- $r = r_{g=I}$, and $\Delta r = 1 - r$.

Error in $\langle \mathcal{O}^{(\mathrm{imp})} \rangle$

- The standard deviation of the improved estimator is:

$$\sigma^{(\mathrm{imp})} \simeq \sigma \left[2\Delta r + \frac{1}{N_G} - \frac{2}{N_G} \Delta r + R^{\mathrm{corr}} \right]^{1/2}, \tag{9}$$

$$R^{\mathrm{corr}} = \frac{1}{N_G^2} \sum_{g \neq g'} r_{gg'}^{\mathrm{corr}}, \tag{10}$$

- To get a reduction in the error we need:
  - $r \simeq 1$: $\mathcal{O}$ and $\mathcal{O}^{(appx)}$ positively correlated.
  - $r_{gg'}^{\mathrm{corr}}$ small and positive: very little correlation between $\mathcal{O}^{(appx)g}$ and $\mathcal{O}^{(appx)g'}$

- Extreme cases:
  - $r = 1$, $r_{gg'}^{\mathrm{corr}} = 0$, then $\sigma^{(\mathrm{imp})} = \frac{\sigma}{\sqrt{N_G}}$.
  - $r = 0$, $r_{gg'}^{\mathrm{corr}} = 1$, then $\sigma^{(\mathrm{imp})} \simeq \sigma\sqrt{2}$.

Conditions of CAA to work

- **CAA-1**: $\mathcal{O}^{(\mathrm{appx})}$ is covariant under G.
- **CAA-2**: $\mathcal{O}^{(\mathrm{appx})}$ is strongly correlated with $\mathcal{O}$, *i.e.* $\Delta r \ll 1$.
- **CAA-3**: The computational cost of $\mathcal{O}^{(\mathrm{appx})}$ is much smaller than $\mathcal{O}$.
- **CAA-4**: The transformation $g \in G$ is chosen to give small (compared to $1/N_G$) positive correlations among $\{\mathcal{O}^{(\mathrm{appx})\, g}\}_{g \in G}$, *i.e.* $R^{\mathrm{corr}} \ll 1/N_G$.
- The question now is: How to construct $\mathcal{O}^{(\mathrm{appx})}$?.
- Two approaches:
  - Low Mode Averaging (LMA).
  - All Mode Averaging (AMA).

Low Mode Averaging

- $\mathcal{O}^{(\mathrm{appx})} = \mathcal{O}^{(\mathrm{LMA})}$,
  The inverse of the Dirac operator $S[U] \approx S^{(\mathrm{low})}$.

$$
\begin{aligned}
\mathcal{O}^{(\mathrm{LMA})} &= \mathcal{O}[S^{(\mathrm{low})}], \\
\mathcal{O}_G^{(\mathrm{LMA})} &= \frac{1}{N_G} \sum_{g \in G} \mathcal{O}[S^{(\mathrm{low})\,\mathrm{g}}], \\
S^{(\mathrm{low})}(x, y) &= \sum_{k=1}^{N_\lambda} \lambda_k^{-1} \psi_k(x) \psi_k^\dagger(y),
\end{aligned}
$$

- λ_k and ψ_k are eigenvalues and eigenvectors of the Hermetian Dirac operator $H(x, y)$.

All Mode Averaging (AMA)

-
$$
\begin{aligned}
\mathcal{O}^{(\mathrm{AMA})} &= \mathcal{O}[S^{(\mathrm{all})}], \\
\mathcal{O}_G^{(\mathrm{AMA})} &= \frac{1}{N_G}\sum_{g\in G}\mathcal{O}[S^{(\mathrm{all})\,\mathrm{g}}], \\
S^{(\mathrm{all})}b &= \sum_{k=1}^{N_\lambda}\lambda_k^{-1}(\psi_k^\dagger b)\psi_k + f_\varepsilon(H)b, \\
f_\varepsilon(H)b &= \sum_{i=1}^{N_{\mathrm{CG}}}(H)^i c_i,
\end{aligned}
\tag{11}
$$

- $f_\varepsilon b$ is a polynomial of H with vector “coefficients” c_i.
- In practice this combination is obtained from the CG, depending on the source vector b and initial guess x_0.
- The subscript ε indicates the norm of the residual vector after N_{CG} iterations, or steps, of the CG.

How does it work?

- Compute some low modes N_λ exactly!
- Project these out from the source, as you do in deflation: $b_{\mathrm{proj}} \equiv \left(1 - \sum_{k=1}^{N_\lambda} \psi_k \psi_k^\dagger\right) b$.
- Solve the projected system with CG to get a solution x_{CG}.
- $x_{\mathrm{CG}} + \sum_{k=1}^{N_\lambda} \lambda_k^{-1} (\psi_k^\dagger b) \psi_k = S^{(\mathrm{all})} b$.
- High modes are included approximately in x_{CG}.
- Two possible stopping criteria for CG:
  - The norm of the residual is smaller than ϵ.
  - Do a fixed number of iterations.
- In the first approach, it might happen that the covaraince condition which leads to $\langle \mathcal{O}^{(\mathrm{AMA})} \rangle = \langle \mathcal{O}_G^{(\mathrm{AMA})} \rangle$ will be violated by round off errors. Although highly unlikely as the authors mention.

Nucleon Electromagnetic Form Factors (3pt functions)

- $N_f = 2+1$ DWF configurations from RBC/UKQCD on $24^3 x 64$ lattice.
- Quark mass parameter $m = 0.005, 0.01$ corresponding to $m_\pi = 330, 420$ MeV.

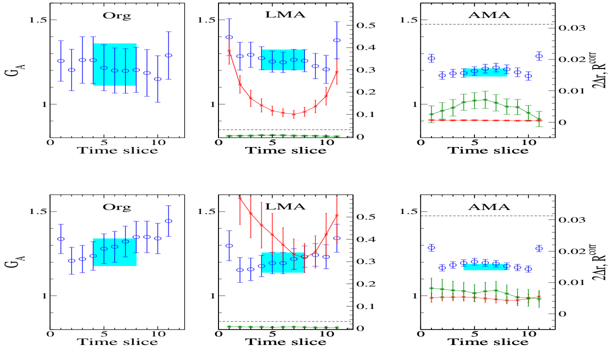


Figure: Time-slice dependence of axial-charge G_A in $m = 0.005$ (top) and $m = 0.01$ (bottom) with standard method (left), LMA (middle) and AMA (right). The cross symbols and star symbols denote $2\Delta r$ and R^{corr} for three-point function. The colored band is the constant fitting result in this range.

CAA and AMA with inexact eigenvectors

- Two main questions:
  - Does using approximate eigenvectors introduce bias?
  - Whether we combine computation of the eigenvectors with computation of the correlation function?
- Approaches envisioned:
  - Incremental EigCG.
  - Inexact deflation as used in DD by Luscher.
  - AMG or DD-αAMG.

The issue of bias

- In the case of exact deflation, we assume that there is no bias because the eigenvectors are exact.
- In a sense, exact here will mean that the accuracy of the eigenvectors is higher than the required accuracy of the linear system.
- Also, the eigenvectors were computed in a separate calculation that is independent of the sources that is used in the improved estimator.
- DD or Multigrid:
  - Approximate eigenvectors obtained from a set of random fields.
  - This setup phase is separate from the solution phase.
  - This probably ensures no bias in the solution.
- Incremental EigCG:
  - Usually approximate eigenvectors are computed simultaneously while solving the linear systems.
  - This combination speeds up the whole calculation.
  - However, approximate eigenvectors will depend on the sources which could lead to a bias.

Combining Deflation and CAA with Incremental EigCG

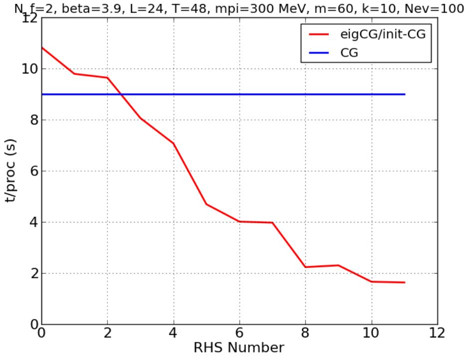
- Two approaches:
  1. Take a set of random fields ϕ_l and solve them with eigCG to obtain a set of approximate eigenvectors.
  2. Combine eigenvector computation with linear system solution and take the point of view that the final solution is what matters (a solution is a solution regardless of how you got it).
- Here we test the second option first as it is most cost effective.

Wilson Twisted-Mass Action at Maximal Twist

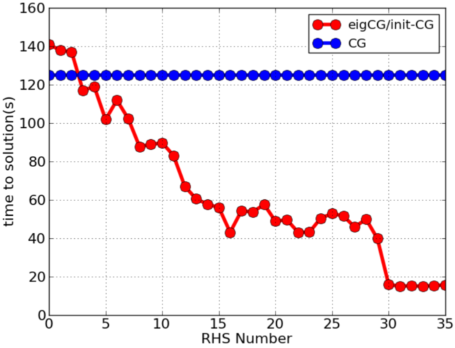
- doublet of light quarks: $\psi = \begin{pmatrix} u \\ d \end{pmatrix}$
- cut-off effects are automatically $\mathcal{O}(a)$ improved
- *BiCGStab* doesn't work for Twisted-Mass.
- Implemented EigCG in tmLQCD software.

EigCG for Twisted-Mass

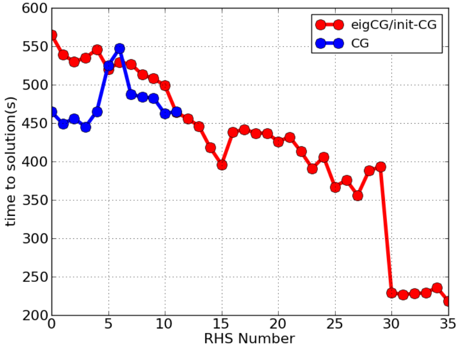
- 2-flavor TM configuration, $24^3 \times 48$, $m_\pi = 300$ MeV, $m = 60$, $k = 10$, $nev = 100$.



- 2+1+1 TM configuration, $48^3 \times 96$, $m_\pi = 230$ MeV, $m = 60$, $k = 10$, $nev = 300$.



- 2-flavor, TM+clover, $48^3 \times 96$, $m_\pi = 140$ MeV, $m = 240$, $k = 5$, $nev = 150$



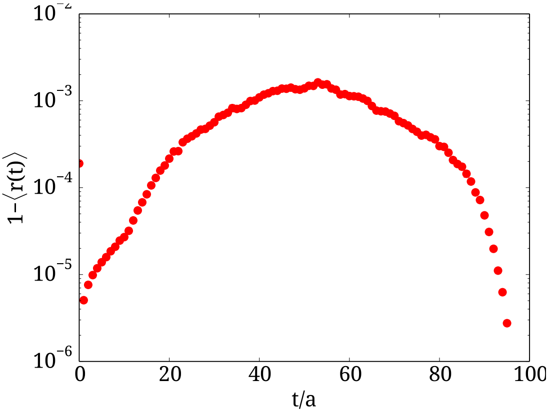
Results

- D15.48 ensemble: 2+1+1, $\beta = 2.1$, $48^3 \times 96$, $m_\pi \approx 200$ MeV, 232 configurations.
- Twelve inversions for the source at twelve source positions were carried out requiring the residual to be $|r^2| < 10^{-18}$ - high precision (hp) inversions.
- For the same 12 source positions the residual was required at $|r^2| < 10^{-4}$ - low precision (lp) inversions.
- An additional 72, the residual was required at $|r^2| < 10^{-4}$ - low precision (lp) inversions.
- Although the spatial components of the source position were chosen to be random, the time component of the source position is not entirely random.

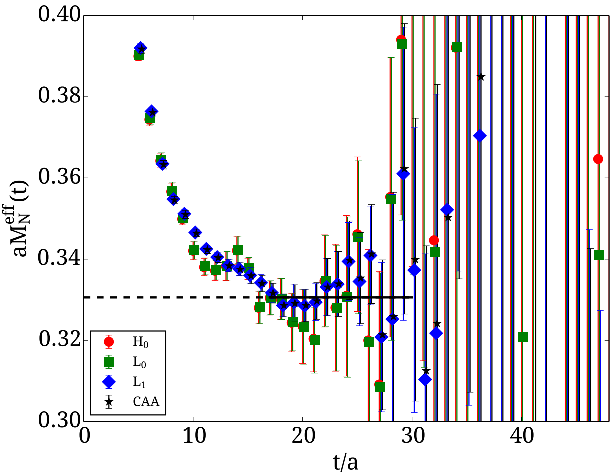
- The CAA method prescribes the improved observable:

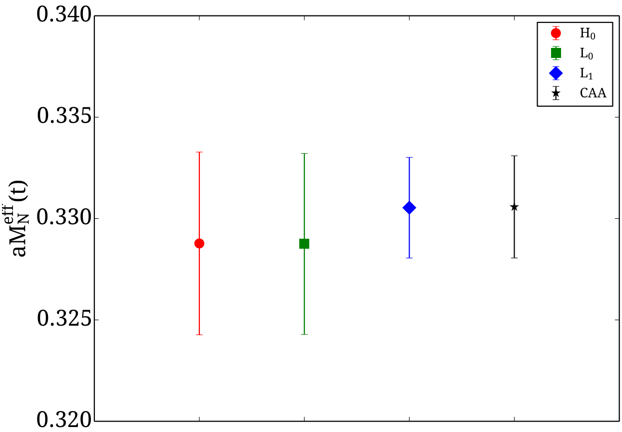
$$O^{\mathrm{imp}} = \underbrace{\frac{1}{N_{\mathrm{hp}}} \sum_{i=0}^{N_{\mathrm{hp}}-1} O_i^{(\mathrm{hp})}}_{H_0} - \underbrace{\frac{1}{N_{\mathrm{hp}}} \sum_{i=0}^{N_{\mathrm{hp}}-1} O_i^{(\mathrm{lp})}}_{L_0} + \underbrace{\frac{1}{N_{\mathrm{lp}}} \sum_{i=N_{\mathrm{hp}}}^{N_{\mathrm{hp}}+N_{\mathrm{lp}}-1} O_i^{(\mathrm{lp})}}_{L_1} .$$

- There are $N_{\mathrm{hp}} + N_{\mathrm{lp}} = 12 + 72$ source positions, of which the first N_{hp} have been inverted to high precision and low precision, while the remaining N_{lp} have been inverted only to low precision.
- If there is a bias due to the approximation made to obtain the low precision inversions, but translational invariance holds, then $\langle L_1 \rangle = \langle L_0 \rangle$, such that $\langle O^{\mathrm{imp}} \rangle = \langle H_0 \rangle$, i.e. the approximation cancels in the mean value.
- As regards the error, if the correlators in the H_0 sum are highly correlated with those of L_0 then the error should scale as $\frac{1}{\sqrt{N_{\mathrm{lp}}}}$.
- In fact, to first approximation the error should scale as $\sqrt{2(1-r) + \frac{1}{N_{\mathrm{lp}}}}$, where $r \in [0, 1]$ is the normalized correlation between H_0 and L_0.

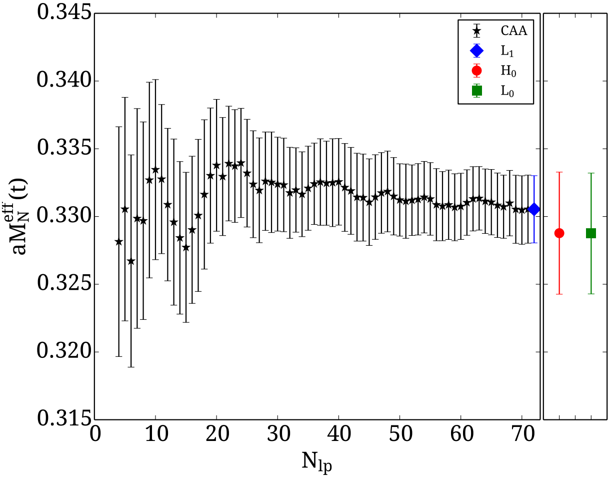


10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
$1-\langle r(t) \rangle$
0
20
40
60
80
100
t/a

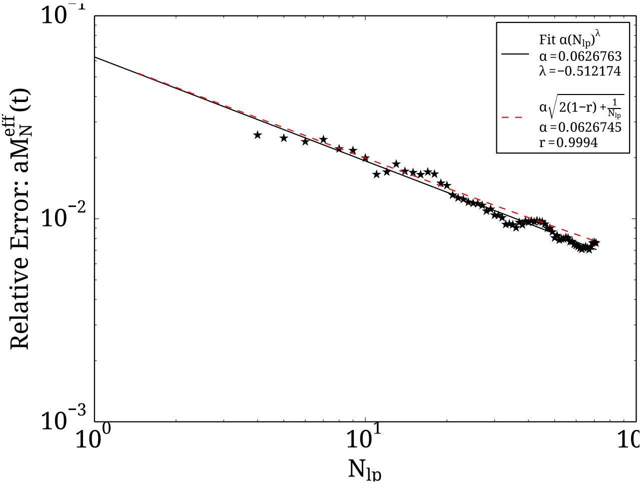

0.40
0.38
0.36
0.34
0.32
0.30
0
10
20
30
40
$aM_N^{eff}(t)$
t/a
H_0
L_0
L_1
CAA



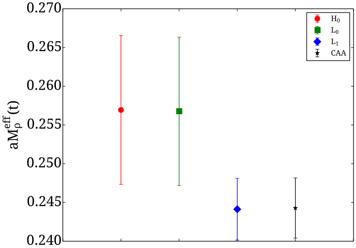
0.340
0.335
0.330
0.325
0.320
$aM_N^{eff}(t)$
H_0
L_0
L_1
CAA

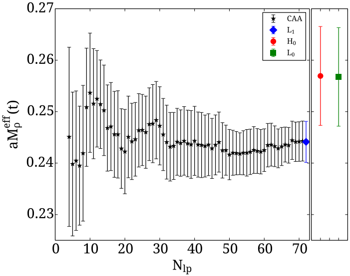


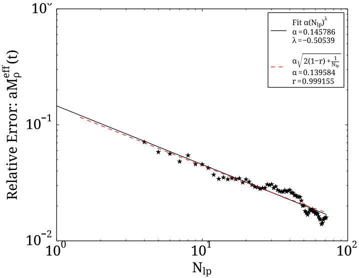
0.345
0.340
0.335
0.330
0.325
0.320
0.315
$aM_N^{eff}(t)$
0
10
20
30
40
50
60
70
N_{lp}
CAA
L_1
H_0
L_0


Fit $\alpha(N_{lp})^{\lambda}$
$\alpha = 0.0626763$
$\lambda = -0.512174$
$\alpha\sqrt{2(1-r)+\frac{1}{N_{lp}}}$
$\alpha = 0.0626745$
$r = 0.9994$
Relative Error: $aM_N^{eff}(t)$
N_{lp}
10^{-1}
10^{-2}
10^{-3}
10^{0}
10^{1}

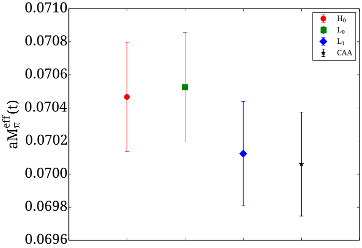
Results for the ρ

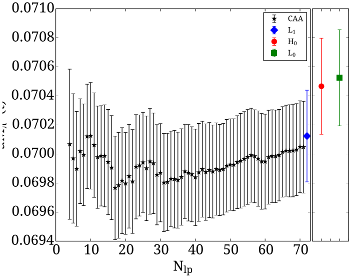


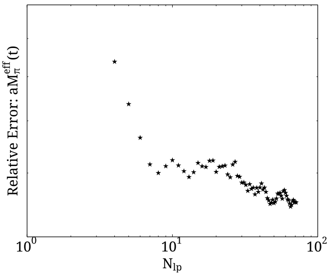




Results for the π







Conclusions & Outlook

- CAA+AMA methods are efficient for reducing errors in hadronic observables.
- EigCG is an efficient solver for Twisted-Mass fermions.
- Tested combining deflation and CAA+AMA and found that there were no bias introduced by the use of approximate eigenvectors.
- Outlook: more testing is planned to see for example how the results will be affected if one solves for the eigenvectors first using randm sources.