格子量子色力学におけるソルバーについて

Yusuke Namekawa (KEK)

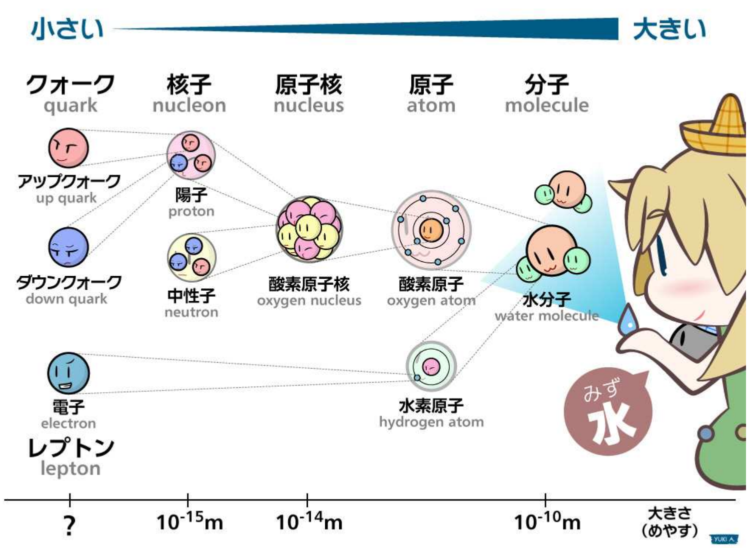
Contents

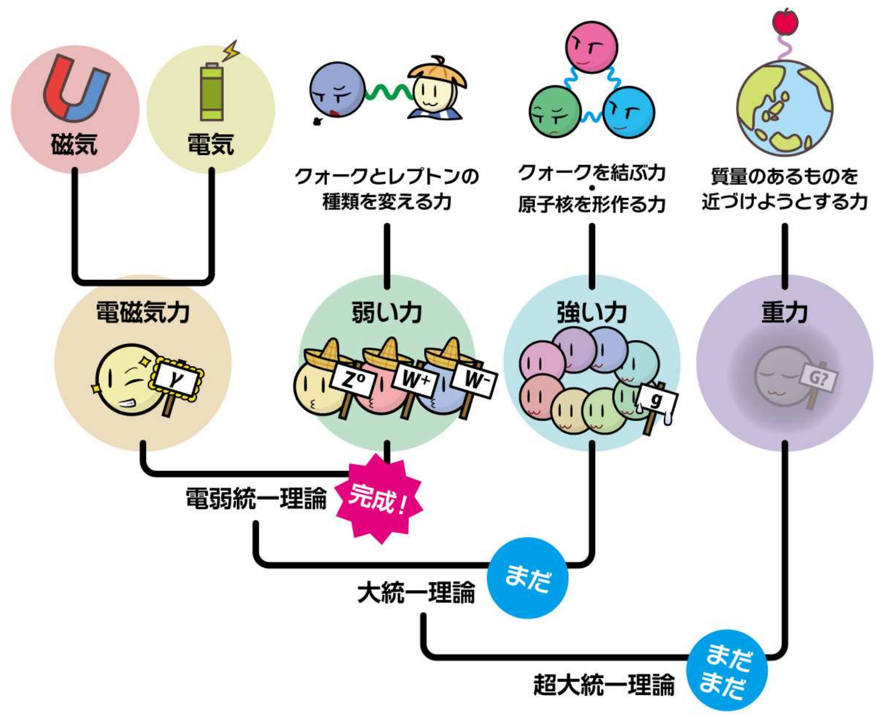
1 Introduction

All materials are made from quarks and leptons

cf. Kanamori-san's talk at the 2nd HPC-Phys meeting

- Theory of the strong interaction among quarks is called "Quantum ChromoDynamics(QCD)"





http://higgstan.com/ ← the designer got PhD on particle physics experiment

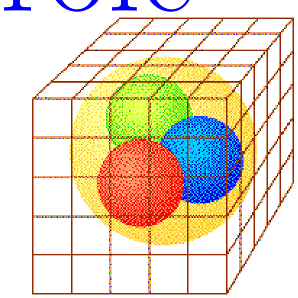
[Quantum ChromoDynamics(QCD)]

- Theory(Lagrangian) is known, but is difficult to be solved analytically

$$\mathcal{L}_{QCD} = \bar{q}(i\not{D} - m)q - \frac{1}{4}G^2$$

  - ♢ One of Millennium Problems http://www.claymath.org/millennium-problems
    → You will win one million USD, if you solve this problem
  - ♢ (cf. one of Millennium Problems on Poincare conjecture has already been solved)

- Numerical simulation of QCD on discretized spacetime (lattice QCD) is possible
  - ♢ $Ax = b$ plays the central role
    → Solver is important
    cf. Kanamori-san's and Ishikawa-san's talks at the 2nd, 3rd HPC-Phys meetings



http://www-het.ph.tsukuba.ac.jp

[Concrete form of A for $Ax = b$ in lattice QCD]

- Concrete form of A depends on the fermion formulation

  ◇ One choice is Wilson-type fermion(9-point stencil in 4-dimension, complex non-symmetric large sparse matrix)

  cf. Kanamori-san's and Ishikawa-san's talks at the 2nd, 3rd HPC-Phys meetings

- Condition number $K(A)$ becomes larger for smaller quark mass m_{quark}

  cf. Ishikawa-san's talk at the 3rd HPC-Phys meeting

  ◇ $K(A(m_{\text{ud}})) = O(2700)$, $K(A(m_{\text{s}})) = O(100)$, $m_{\text{s}}/m_{\text{ud}} \sim 27$

$$A(x,y) = \delta_{x,y} - \kappa \sum_{\mu=1}^{4} \left\{ (1-\gamma_\mu) U_\mu(x) \delta_{x+\mu,y} + (1+\gamma_\mu) U_\mu^\dagger(x-\mu) \delta_{x-\mu,y} \right\}$$

: complex $n \times n$ non-symmetric matrix, $n \sim 10^{10}$ for a typical lattice QCD

$$m_{\text{quark}} = \frac{1}{2}\left(\frac{1}{\kappa} - (\text{const})\right)$$

$$K(A) \propto \frac{1}{m_{\text{quark}}}$$

2 Solvers in lattice QCD

Major solvers in lattice QCD are tabulated

- There are many solver algorithms for lattice QCD
  $\to$ Only solvers in the table are explained
- There are many open sources for lattice QCD
  $\to$ Only open sources in the table are explained
- (Preconditioners are not covered in this talk)

Solver	Open source
CG Hestenes,Stiefel(1952)	Bridge++
BiCGStab van der Vorst(1992)	Bridge++, CCSQCDSolverBench
BiCGStab(L) Sleijpen,Fokkema(1993)	Bridge++
BiCGStab(DS-L) Miyauchi et al.(2001)	Bridge++
BiCGStab(IDS-L) Itoh,Namekawa(2003)	Bridge++
GMRES(m) Saad,Schultz(1986)	Bridge++
MultiGrid A.Brandt(1977)	DDalphaAMG

[Lattice QCD code Bridge++ (our open source code)]

- Bridge++ is a code set for numerical simulations of lattice gauge theories including QCD
  $\rightarrow$ Ver.1.5.1 has been released in Aug 2019
- Major solvers(BiCGStab series,CG,GMRES(m)) are covered
- Project members:
  Y.Akahoshi (YITP), S.Aoki (YITP), T.Aoyama (KEK), I.Kanamori (R-CCS), K.Kanaya (Tsukuba), H.Matsufuru (KEK), Y.Namekawa (KEK), H.Nemura (RCNP), Y.Taniguchi (Tsukuba)
  - ◇ I have been the chairperson since 2016

[CCS QCD SolverBench]

- CCS QCD SolverBench is a benchmark BiCGStab program of QCD developed by another CCS(Univ of Tsukuba)
  $\rightarrow$ Ver.0.999(rev.248) has been released in Sep 2017
- BiCGStab with even-odd preconditioning is employed
- Project members:

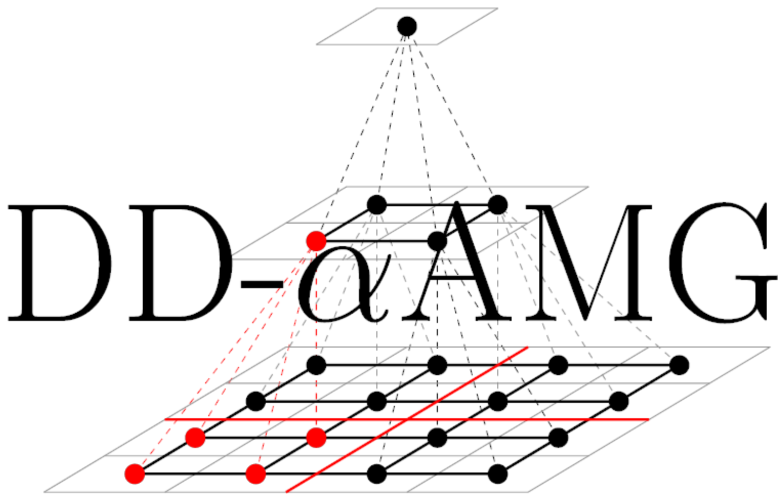
  K-I.Ishikawa (Hiroshima), Y.Kuramashi (Tsukuba), A.Ukawa (Tsukuba), T.Boku (Tsukuba)

https://www.ccs.tsukuba.ac.jp/qcd/

[DDalphaAMG]

- DDalphaAMG is a multigrid solver program in lattice QCD
  → Ver.1701 has been released in Jan 2017
  → Ported to K-computer in Apr 2018 Ishikawa,Kanamori(2018)
- Adaptive Algebraic MultiGrid(αAMG) algorithm with Domain Decomposed(DD) smoother is employed
- Project members:

  M.Rottmann, A.Strebel, S.Heybrock, S.Bacchio, B.Leder, I.Kanamori



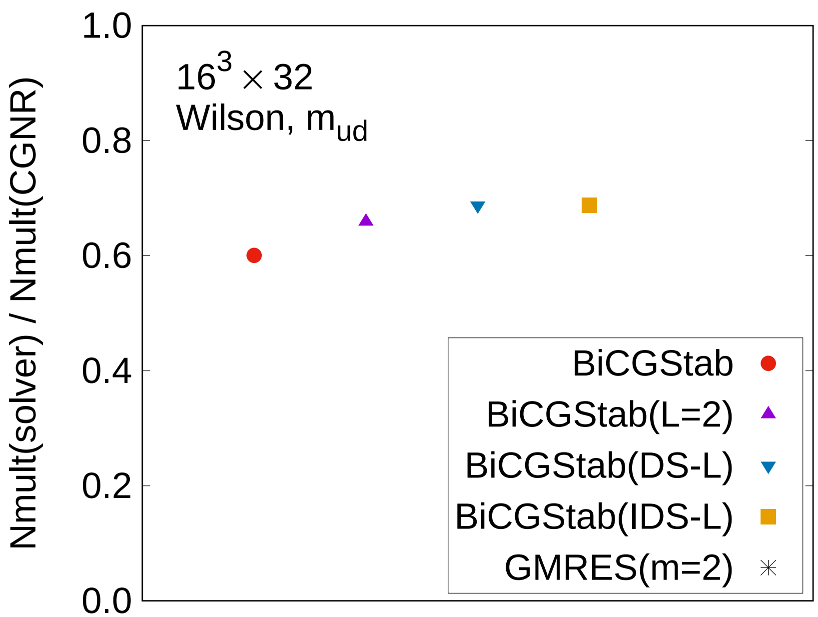
https://github.com/DDalphaAMG

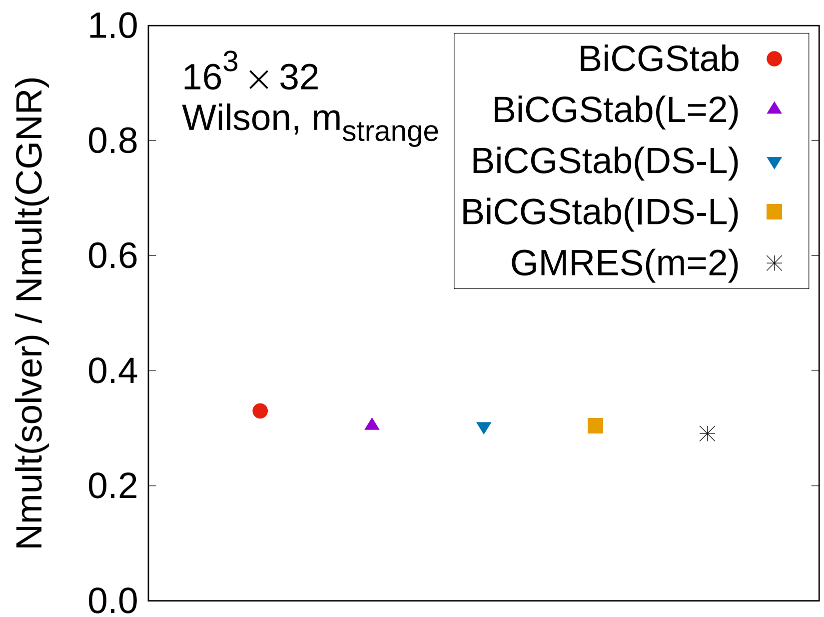
https://github.com/i-kanamori/DDalphaAMG/tree/K

3 Benchmark results

[CG vs BiCGStab series, GMRES(m) by Bridge++]

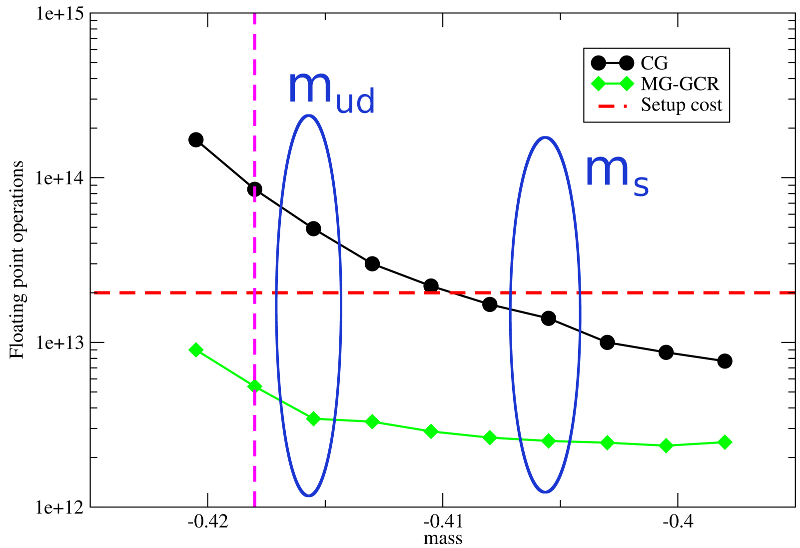
- For m_{ud} (up-down quark mass), which requires a huge Krylov space, BiCGStab series gain 30-40%, while GMRES(m=2–16) shows no gain
- For m_{s} (strange quark mass), which requires not so large Krylov space, BiCGStab series and GMRES(m) gain a factor of 3
  - ◇ Prescription is added to BiCGStab for better stability
    Sleijpen and van der Vorst(1995)





[CG vs MG(MultiGrid)] Babich et al.(2010)

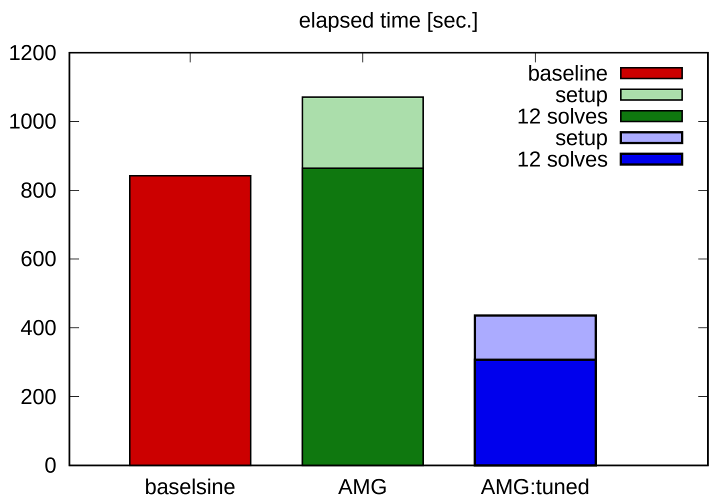
- For $m_{\rm ud}$ (up-down quark mass), which requires a huge Krylov space, multigrid gains a factor of 3
- For $m_{\rm s}$ (strange quark mass), which requires not so large Krylov space, multigrid has no gain due to its overhead
  - ◇ Memory cost of multigrid is larger than that of CG by a factor of 4–5
  - ◇ NB. $m^{\rm phys}_{\rm quark} \propto (m^{\rm bare}_{\rm quark} - m^{\rm critical}_{\rm quark})$ with $m^{\rm critical}_{\rm quark} = -0.4175$



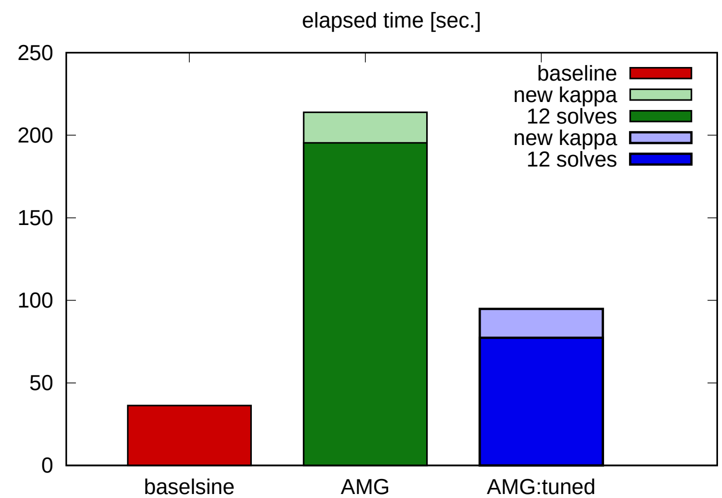
[Nested BiCGStab with precond(SAP + SSOR) vs multigrid] Ishikawa,Kanamori(2018)
Similar results are obtained on K-computer

- For m_{ud} (up-down quark mass), which requires a huge Krylov space, multigrid gains a factor of 2 over the baseline BiCGStab
- For m_{s} (strange quark mass), which requires not so large Krylov space, multigrid has no gain due to its overhead

$\diamondsuit$ The best solver depends on the target system



Up-down quark case



Strange quark case

4 Additional hot topics with multiple right hand side

$$
\begin{aligned}
Ax_{n_{\mathrm{rhs}}} &= b_{n_{\mathrm{rhs}}} \\
\text{where} & \\
A &:= n \times n \text{ matrix} \\
& \quad n \sim 10^{10} \text{ for a typical lattice QCD} \\
& \quad \forall n_{\mathrm{rhs}} = 1, 2, \ldots
\end{aligned}
$$

- Block solver(multiple right hand side solver) O'Leary(1980)
- Truncated solver Collins,Bali,Schäfer(2007)
- Deflation de Forcrand(1996),Lüscher(2007)

[Block solver(multiple right hand side solver)] O'Leary(1980)

$$AX = B \quad \text{instead of } Ax = b$$

where

$$\begin{aligned} A &:= n \times n \text{ matrix}, \\ X, B &:= n \times n_{\rm rhs} \text{ matrix} \\ & \quad n \sim 10^{10} \text{ for a typical lattice QCD} \\ & \quad \forall n_{\rm rhs} = 1, 2, \ldots \end{aligned}$$

- The philosophy is sharing Krylov space for multiple right hand sides
  - ◇ Practical advantage is better use of cache, which increase the sustained speed by a factor of 2-5
- Two problems are known
  $\rightarrow$ Next page

[Block solver(continued)]

- There are some attempts in lattice QCD

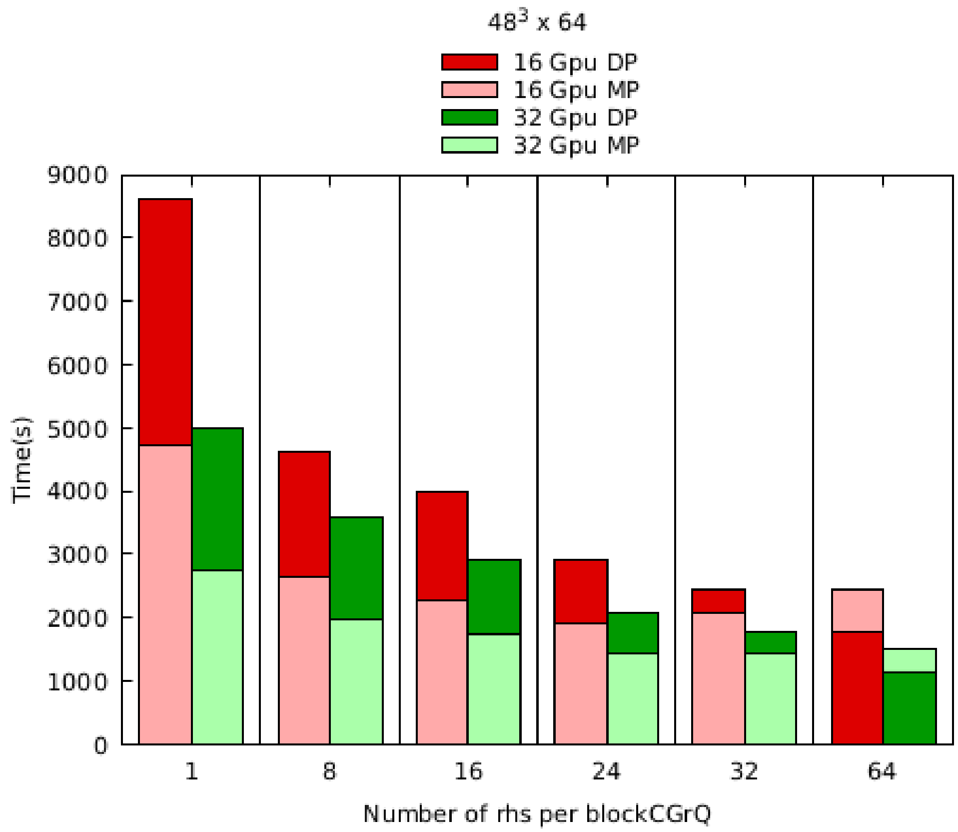
  de Forcrand(1996), Sakurai et al.(2010), Tadano et al.(2010), Nakamura et al.(2011), Birk and Frommer(2012,2014), Clark et al.(2018), de Forcrand and Keegan(2018)

  ◇ Problem 1 : naive Block solver has a gap between true and recursion residuals
  $\rightarrow$ Improved versions are proposed Dubrulle(2001), Tadano et al.(2009), ...

  ◇ Problem 2 : Block solver often fails to converge (breakdown and stagnation), though it can be tamed in part by QR decomposition Dubrulle(2001), Nakamura et al.(2011), ...
  $\rightarrow$ We do not employ the block solver in a large scale simulation

[Block solver(continued)]

- Block solver(blockCGrQ) gains a factor of 2-5, if it converged



Clark et al.(2018)

DP := Double Precision
MP := Mixed Precision

- Mixed precision is usually faster, but it is not for a larger number of rhs, probably due to less stability

[Truncated solver] Collins,Bali,Schäfer(2007)

- Truncated solver := many approximate solver results corrected by exact solver result
- (cf. all mode averaging := truncated solver + low-mode averaging) Blum et al.(2012)
  - ◇ Truncated solver leads to a factor of 10 speed up for an expectation value constructed from the solution x

$$\left\langle O^{\mathrm{exact}}[x]\right\rangle = \left\langle O^{\mathrm{improved}}[x]\right\rangle, \quad \langle O\rangle := \frac{1}{N_{\mathrm{sample}}}\sum_{i=1}^{N_{\mathrm{sample}}} O_i$$

where

$$O^{\mathrm{improved}}[x] = (O[x_1^{\mathrm{exact}}] - O[x_1^{\mathrm{approx}}]) + \frac{1}{N_{\mathrm{approx}}}\sum_{n'_{\mathrm{rhs}}=2}^{N_{\mathrm{approx}}} O[x_{n'_{\mathrm{rhs}}}^{\mathrm{approx}}]$$

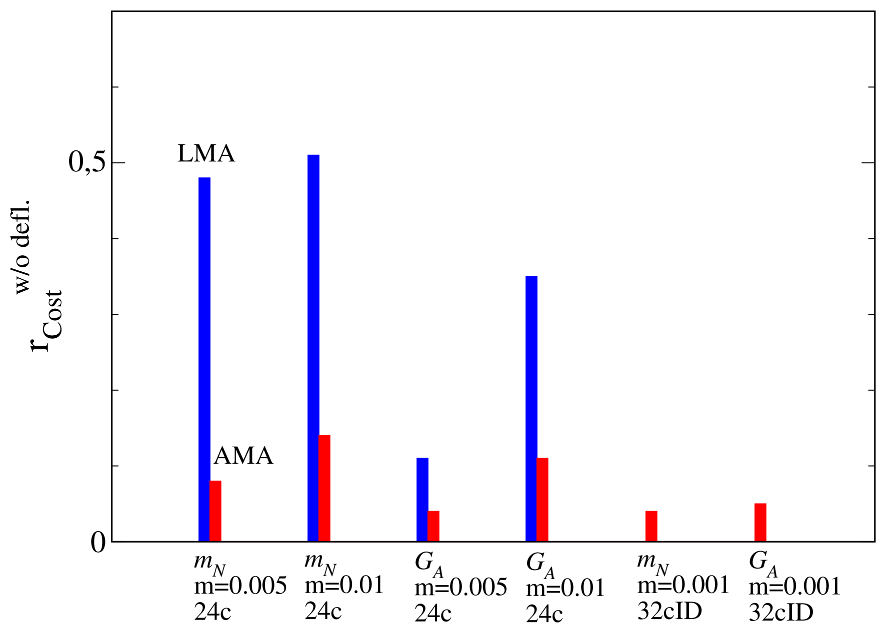
$$Ax_{n_{\mathrm{rhs}}}^{\mathrm{exact}} = b_{n_{\mathrm{rhs}}}, \quad \text{strict stopping condition (ex. } 10^{-16})$$

$$Ax_{n'_{\mathrm{rhs}}}^{\mathrm{approx}} = b_{n'_{\mathrm{rhs}}}, \quad \text{loose stopping condition (ex. truncated at } N_{\mathrm{iter}} = 50)$$

$$\forall n_{\mathrm{rhs}}, \forall n'_{\mathrm{rhs}} = 1, 2, \ldots \qquad \text{larger gain for } n_{\mathrm{rhs}} < n'_{\mathrm{rhs}}$$

[Truncated solver(continued)]

- Truncated solver (+ low mode averaging) leads to $O(10)$ speed up
  - ◇ NB. care is needed for the choice of the truncation (ex. $N_{\rm iter} = 50$). Too aggressive choice gives a wrong result.



Shintani et al.(2014)

[Deflation] de Forcrand(1996),Lüscher(2007),...

- Deflation := eigenvectors + solver for the remaining part
  - ◇ Deflation is independent of $n_{\rm rhs}$ i.e. larger $n_{\rm rhs}$ gives larger gain
  - ◇ The gain is a factor of 2-8, though deflation needs overhead and large memory consumption of eigenvector estimation

$$
\begin{aligned}
Ax_{n_{\rm rhs}} &= b_{n_{\rm rhs}} \\
A\phi_i &= \lambda_i \phi_i, \quad i = 1, \ldots, N_{\rm deflation}
\end{aligned}
$$

Then

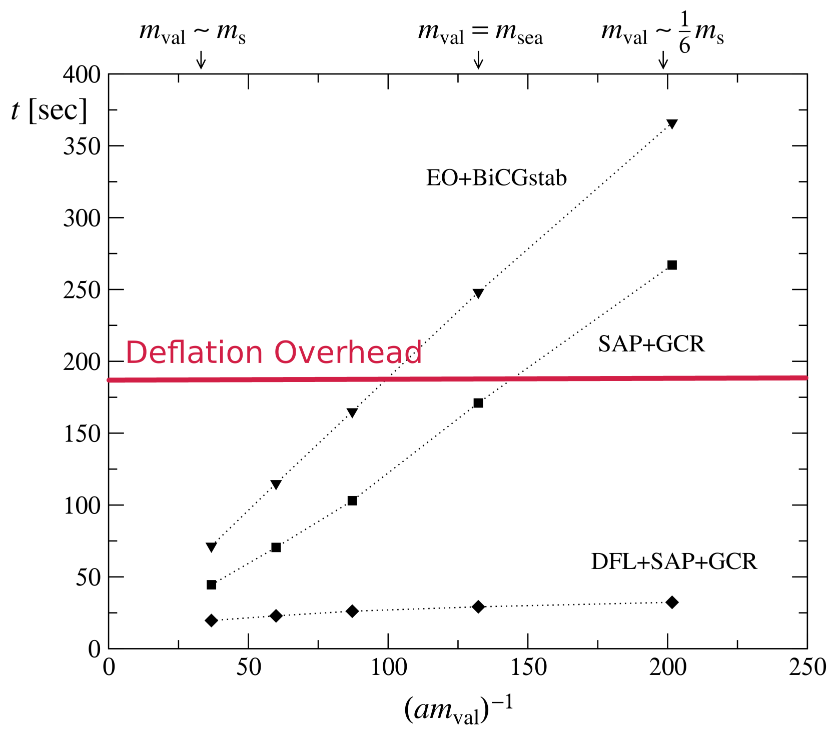
$$
x_{n_{\rm rhs}} = x^{\rm solver}_{n_{\rm rhs}} + \sum_{i,j=1}^{N_{\rm deflation}} \phi_i A^{-1}_{ij} (\phi_j, b_{n_{\rm rhs}})
$$

where

$$
\begin{aligned}
P_{\rm deflation} A x^{\rm solver}_{n_{\rm rhs}} &= P_{\rm deflation} b_{n_{\rm rhs}} \\
P_{\rm deflation} x_{n_{\rm rhs}} &= x_{n_{\rm rhs}} - \sum_{i,j=1}^{N_{\rm deflation}} A\phi_i A^{-1}_{ij} (\phi_j, b_{n_{\rm rhs}})
\end{aligned}
$$

[Deflation(continued)]

- The gain is a factor of 2-8, though deflation gives overhead
  - ◇ NB. the best choice of $N_{\text{deflation}}$ depends on the system



Lüscher(2007)

5 Summary

Overview of solvers in lattice QCD was presented

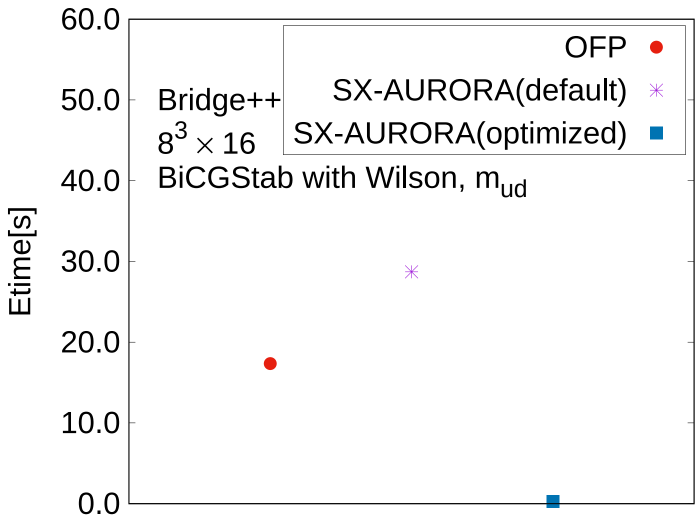
- Major solvers are covered by open sources(Bridge++, CCSQCDSolver-Bench, DDalphaAMG, ...)
- Benchmark results show the best solver depends on the physics
    - ◇ multigrid is best for m_{ud} (requiring a huge Krylov space)
    - ◇ BiCGStab series and GMRES(m) is faster for m_{s} (requiring not so large Krylov space)
- Additional hot topics with multiple right hand side are explained
    - ◇ Block solver(multiple right hand side solver) gains a factor of 2-5, though it often fails to converge
    - ◇ Truncated solver leads to $O(10)$ speed up, though too aggressive truncation gives a wrong result
    - ◇ Deflation gains a factor of 2-8, though it needs overhead and large memory consumption of eigenvector estimation

[Not covered in this talk]

- Preconditioner
  - ◇ Even-odd(red/black), SAP(Schwarz Alternating Procedure), ILU, SSOR, ...

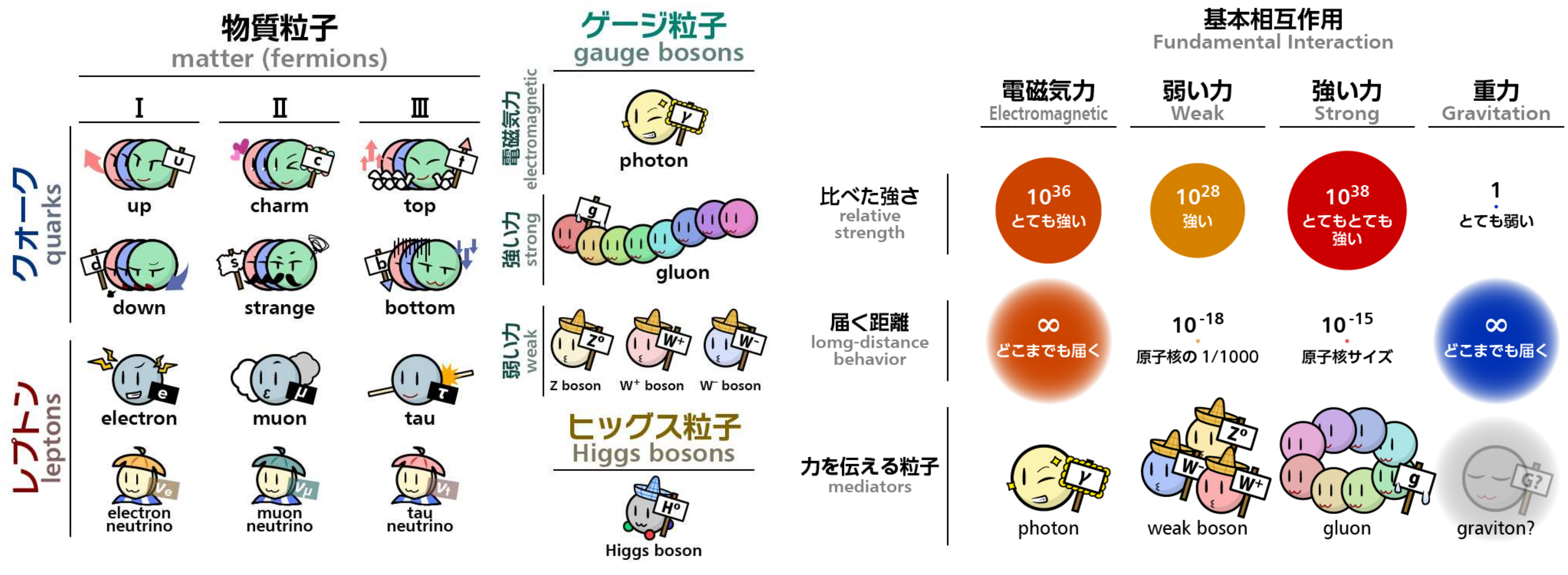
[Advertise new supercomputer at KEK(SX-AURORA,156.8 TFlop)]

- Unfortunately KEK supercomputer had been terminated since 2017, but is renewal in 2019 http://scwww.kek.jp/
- Tuning for discrete vector accelerator leads to O(100) speed up



Appendix

[Table of elementary particles and interactions]



http://higgstan.com/ ← the designer got PhD on particle physics experiment