Path optimization method with use of Neural Network for the Sign Problem in Field theories

Akira Ohnishi [1], Yuto Mori [2], Kouji Kashiwa [3]

1. Yukawa Inst. for Theoretical Physics, Kyoto U., 2. Dept. Phys., Kyoto U., 3. Fukuoka Inst. Tech.

The 36th Int. Symp. on Lattice Field Theory, July 22-28, 2018, East Lansing, MI, USA

East Lansing, MI, USA

Collaborators

Akira Ohnishi [1], Yuto Mori [2], Kouji Kashiwa [3]

1. Yukawa Inst. for Theoretical Physics, Kyoto U.,
2. Dept. Phys., Kyoto U., 3. Fukuoka Inst. Tech.

Y. Mori
(grad. stu.)

K. Kashiwa

AO (10 yrs ago)

1D integral: Y. Mori, K. Kashiwa, AO, PRD 96 ('17), 111501(R) [arXiv:1705.05605]
φ4 w/ NN: Y. Mori, K. Kashiwa, AO, PTEP 2018 ('18), 023B04 [arXiv:1709.03208]
Lat 2017: AO, Y. Mori, K. Kashiwa, EPJ Web Conf. 175 ('18), 07043 [arXiv:1712.01088]
NJL thimble: Y. Mori, K. Kashiwa, AO, PLB 781('18),698 [arXiv:1705.03646]
PNJL w/ NN: K. Kashiwa, Y. Mori, AO, arXiv:1805.08940.
0+1D QCD: Y. Mori, K. Kashiwa, AO, in prep.

The Sign Problem

- **When the action is complex, strong cancellation occurs in the Boltzmann weight at large volume. = The Sign Problem**

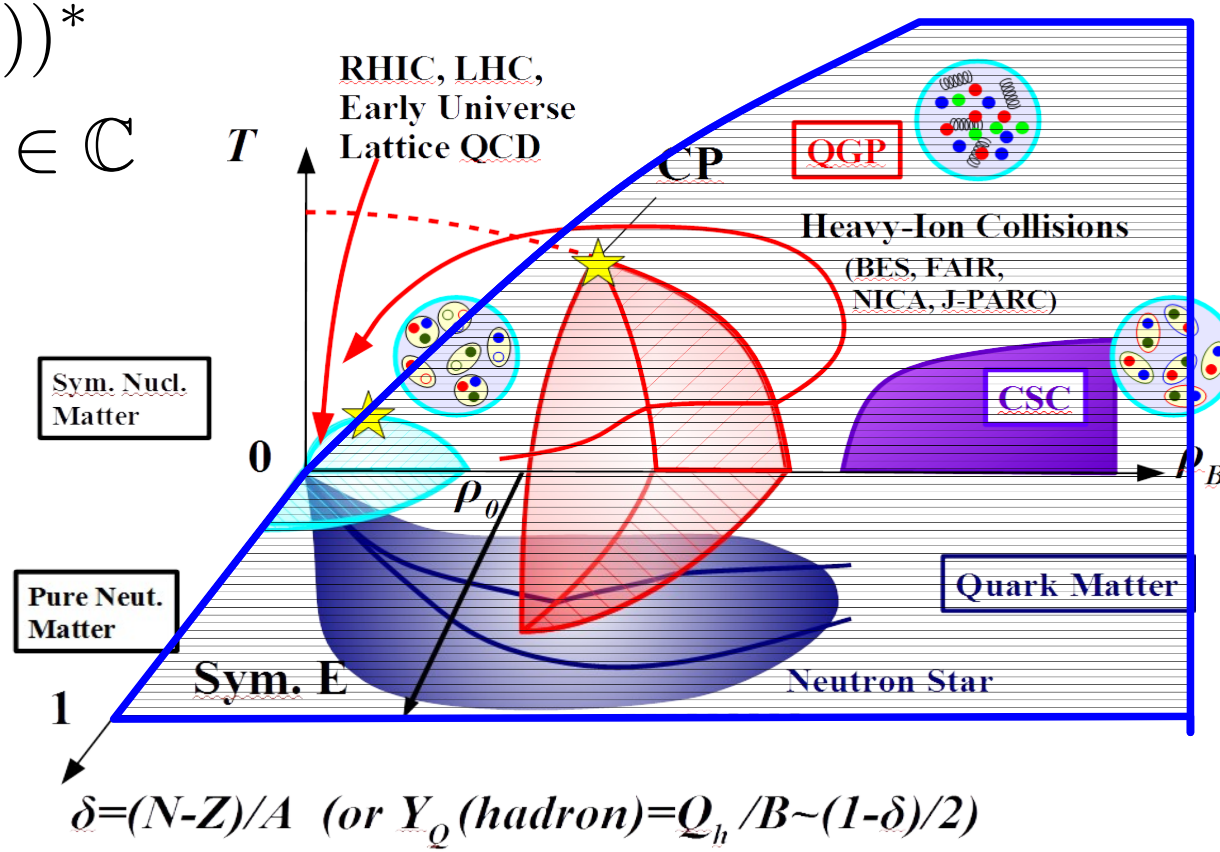
$$\mathcal{Z} = \int \mathcal{D}x \, e^{-S(x)} (S(x) \in \mathbb{C}) \to 0 \ (V \to \infty)$$

- **Fermion det. is complex at finite density**

$$\det D(\mu) = (\det D(-\mu^*))^*$$

$$\to S_{\text{eff}} = S - \log \det D \in \mathbb{C}$$

- **Difficulty in studying finite density in LQCD**
  → Heavy-Ion Collisions, Neutron Star, Binary Neutron Star Mergers, Nuclei, …



Approaches to the Sign Problem in Lattice 2018

- **Standard approaches**
  - **Taylor expansion** *[Ratti(Mon), Mukerjee(Tue), Steinbrecher(Wed)]*
  - **Imaginary μ (Analytic cont. / Canonical)** *[Guenther, Goswami (Wed)]*
  - **Strong coupling** *[Unger, Klegrewe (Fri)]*

  → *Mature, Practically useful, but cannot reach cold dense matter*

- **Integral in Complexified variable space**
  - **Lefschetz thimble method** *[Zambello (Mon)]*
  - **Complex Langevin method**
    [Sinclair, Tsutsui, Attanasio, Ito, Josef (Mon), Wosiek (Fri)]
  - **Path optimization method** *[Lawrence, Warrington, Lamm (Mon), AO (Sat)]*
  - **Action modification** *(e.g. Tsutsui, Doi ('16))*

  → *Premature, but Developing !*

- **Other Approaches** *[Ogilvie (Mon), Jaeger(Fri)]*

Integral in Complexified Variable Space

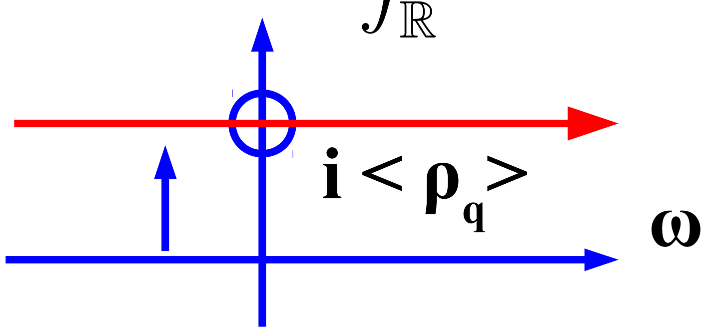
- **Phase fluctuations can be suppressed by shifting the integration path in the complex plain.**

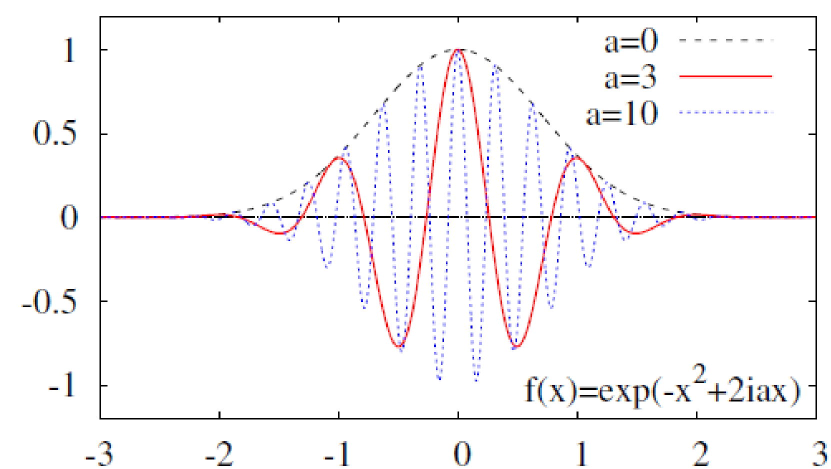
$$\mathcal{Z} = \int_{\mathcal{C}_\mathbb{R}} \mathcal{D}x\, e^{-S(x)} = \int_{\mathcal{C}} \mathcal{D}z\, e^{-S(z)} = \int_{\mathcal{C}_\mathbb{R}} \mathcal{D}x\, J e^{-S(z(x))}$$

- **Simple Example: Gaussian integral (bosonized repulsive int.)**
  Mori, Kashiwa, AO ('18b)

$$\int_{\mathbb{R}} d\omega\, e^{-\omega^2/2 + i\omega\rho_q} = \int_{\mathbb{R}+i\rho_q} d\omega\, e^{-\underbrace{(\omega - i\rho_q)}_{z}{}^2/2 - \rho_q^2/2}$$

$$= \exp(-\rho_q^2/2) \int_{\mathbb{R}} dz\, e^{-z^2/2}$$

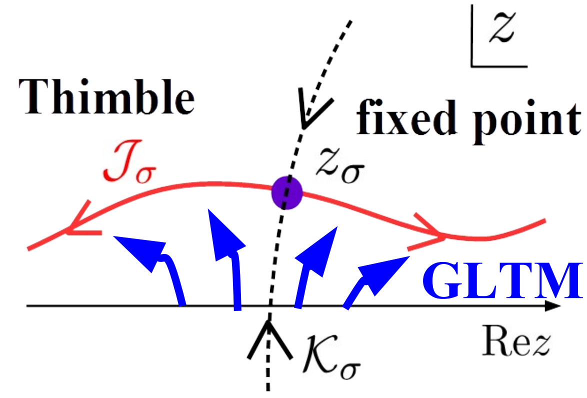




- **Lefschetz thimble / Complex Langevin / Path Optimization**

Lefschetz thimble method

E. Witten ('10), Cristoforetti et al. (Aurora) ('12), Fujii et al. ('13), Alexandru et al. ('16); [Zambello (Mon)]



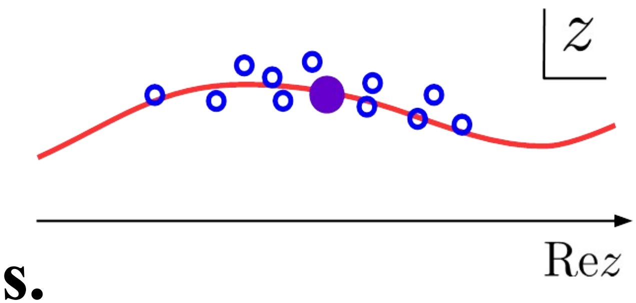
- **Solving the flow eq. from a fixed point σ**
  → Integration path (thimble)
  Note: Im(S) is constant on one thimble

$$\mathcal{J}_\sigma : \frac{dz_i(t)}{dt} = \overline{\left(\frac{\partial S}{\partial z_i}\right)} \to \frac{dS}{dt} = \sum_i \left|\frac{\partial S}{\partial z_i}\right|^2 \in \mathbb{R}, \quad \mathcal{C} = \sum_\sigma n_\sigma \mathcal{J}_\sigma$$

- **Problem:**
  - **Phase from the Jacobian (residual. sign pr.),**
  - **Different Phases of Multi-thimbles (global sign pr.),**
  - **Stokes phenomena, …**

Complex Langevin method

Parisi ('83), Klauder ('83), Aarts et al. ('11),
Nagata et al. ('16); Seiler et al. ('13), Ito et al. ('16);
[Sinclair, Tsutsui, Attanasio, Ito, Joseph (Mon)]



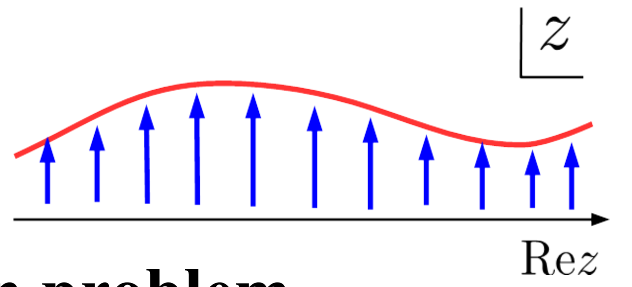
- **Solving the complex Langevin eq.→ Configs.**

$$\frac{dz_i}{dt} = -\frac{\partial S}{\partial z_i} + \eta_i(t) (\eta_i : \text{White noize}),\ \langle \mathcal{O}(x) \rangle = \langle \mathcal{O}(z) \rangle$$

- **No sign problem.**
- **Problem:**
  - **CLM can give converged but wrong results,**
    and we cannot know if it works or not in advance.

Path optimization method

Mori et al. ('17), AO, Mori, Kashiwa (Lattice 2017),
Mori et al. ('18), Kashiwa et al. ('18);
Alexandru et al. ('17 (Learnifold), '18 (SOMMe), '18),
Bursa, Kroyter ('18), [Lawrence, Warrington, Lamm (Mon)]



- **Integration path is optimized to evade the sign problem, i.e. to enhance the average phase factor.**

$$\mathrm{APF} = \langle e^{i\theta} \rangle_{\mathrm{pq}} = \int dx e^{-S} \Big/ \int dx |e^{-S}| = \mathcal{Z}/\mathcal{Z}_{\mathrm{pq}}$$

Sign Problem → Optimization Problem

- **Cauchy(-Poincare) theorem: the partition fn. is invariant if**
  - **the Boltzmann weight W=exp(-S) is holomorphic (analytic),**
  - **and the path does not go across the poles and cuts of W.**
- **At Fermion det.=0, S is singular but W is not singular**
  - **Problem: quarter/square root of Fermion det.**

Cost Function and Optimization

- **Cost function: a measure of the seriousness of the sign problem.**

$$\mathcal{F}[z(x)] = \frac{1}{2}\int dx \left| e^{i\theta(x)} - e^{i\theta_0} \right|^2 \left| J(x) e^{-S(z(x))} \right|$$

$$= |\mathcal{Z}| \left(\left| \langle e^{i\theta} \rangle_{\mathrm{pq}} \right|^{-1} - 1 \right) = \mathcal{Z}_{\mathrm{pq}} - |\mathcal{Z}|$$

$$\left[\theta = \arg(J e^{-S}),\ \theta_0 = \arg(\mathcal{Z}) \right]$$

- **Optimization: the integration path is optimized to minimize the Cost Function. (via Gradient Descent or Machine Learning)**
  - **Example: One-dim. integral → Complete set**

$$z(x) = x + iy(x),\ y(x) = \sum_n c_n H_n(x)$$

$$\mathcal{Z} = \int dx J(x) e^{-S(z(x))},\ J(x) = \frac{dz(t)}{dx}$$

Benchmark test: 1 dim. integral

- **A toy model with a serious sign problem**
  J. Nishimura, S. Shimasaki ('15)

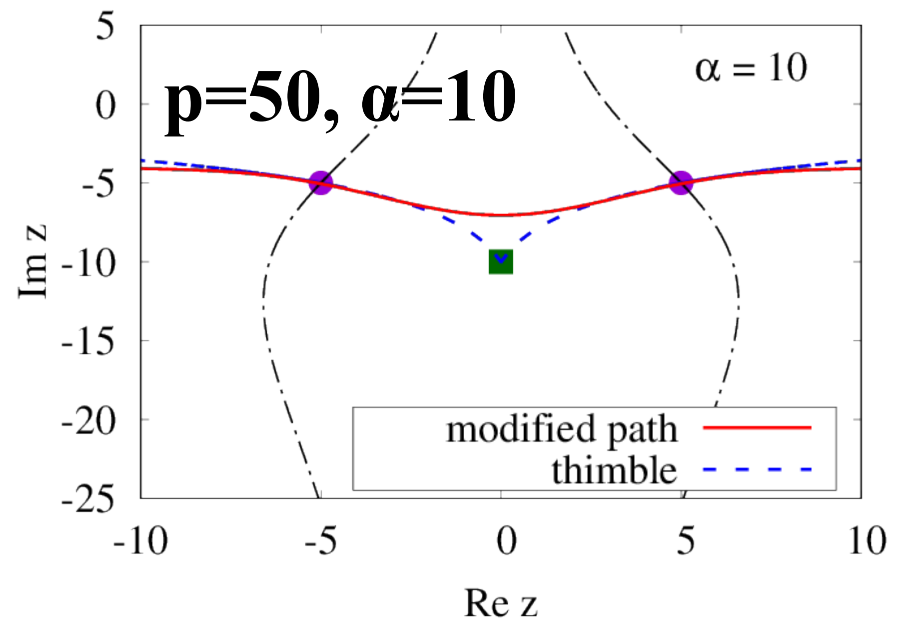
$$\mathcal{Z} = \int dx (x + i\alpha)^p \exp(-x^2/2)$$

  - **Sign prob. is serious with large *p* and small *α* → CLM fails**

- **Path optimization**

$$y(x) = c_1 \exp(-c_2^2 x^2/2) + c_3), \;\; J = 1 + i dy/dx$$

  - **Gradient Descent optimization**
  - **Optimized path ~ Thimble around Fixed Points**

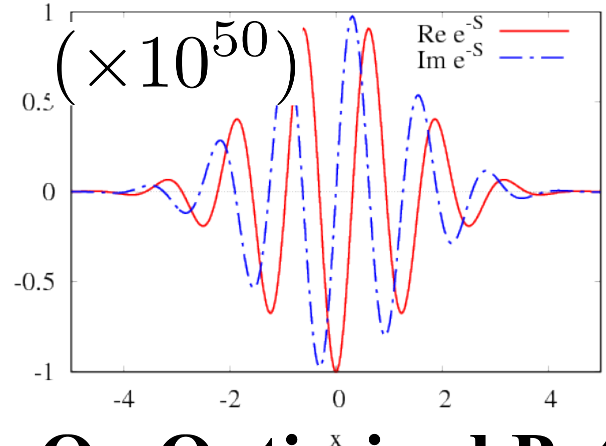


Mori, Kashiwa, AO ('17); AO, Mori, Kashiwa (Lat 2017)

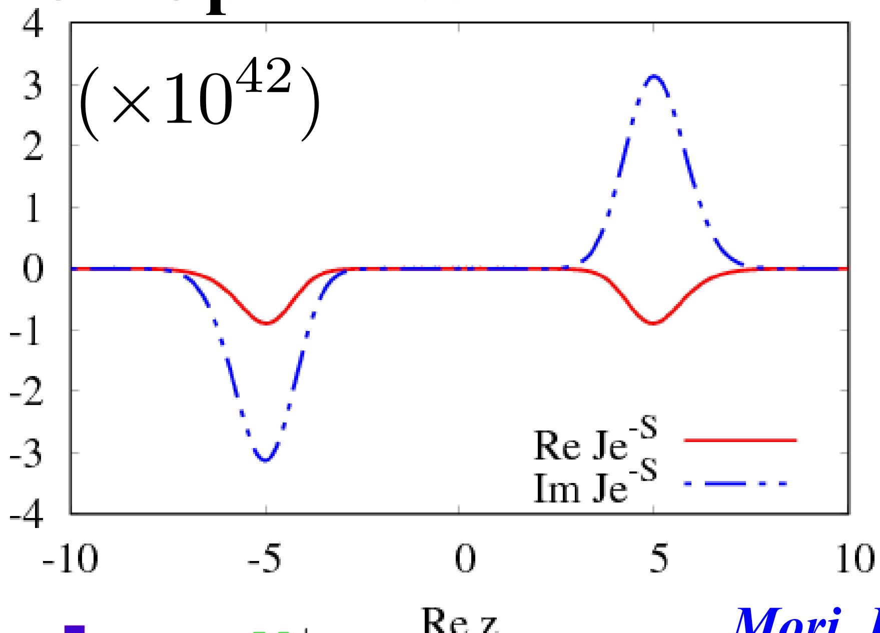
Benchmark test: 1 dim. integral

- **Stat. Weight $J\,e^{-S}$**

On Real Axis

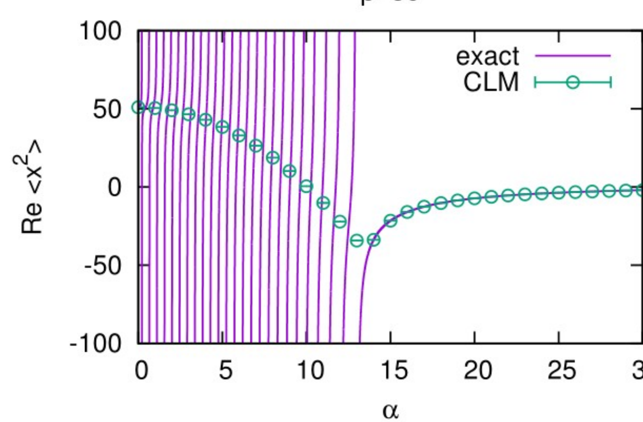


On Optimized Path

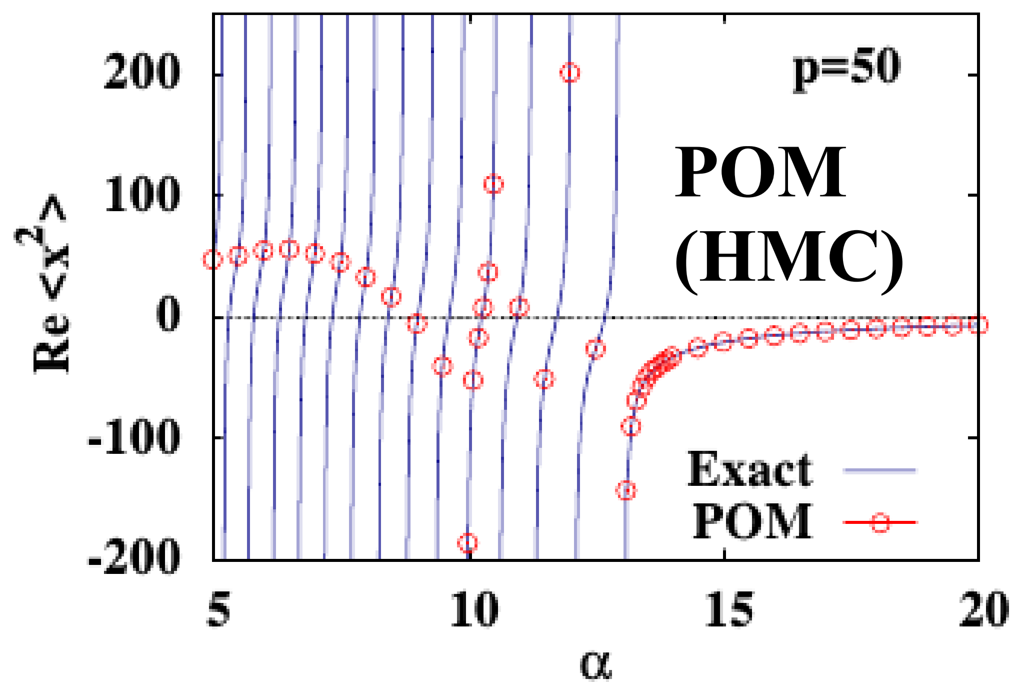


- **Observable**

CLM *Nishimura, Shimasaki ('15)*



vs



Mori, Kashiwa, AO ('17); AO, Mori, Kashiwa (Lat 2017)

Now it's the time to apply POM
to field theories !
Lattice 2017 (Granada) → Lattice 2018 (MSU)

Contents

- **Introduction to Path Optimization Method**

  Y. Mori, K. Kashiwa, AO, PRD 96 ('17), 111501(R) [arXiv:1705.05605]

  AO, Y. Mori, K. Kashiwa, EPJ Web Conf. 175 ('18), 07043 [arXiv:1712.01088] (Lattice 2017 proceedings)

- **Application to complex φ^4 theory using neural network**

  Y. Mori, K. Kashiwa, AO, PTEP 2018 ('18), 023B04 [arXiv:1709.03208]

- **Application to gauge theory: 1-dimensional QCD**

  Y. Mori, K Kashiwa, AO, in prep.

- **Discussions**

- **Summary**

Application to complex φ^4 theory using neural network

Y. Mori, K. Kashiwa, AO, PTEP 2018 (‘18), 023B04 [arXiv:1709.03208]

Application of POM to Field Theory

- **Preparation & variation of trial fn. is tedious in multi-D systems**

$$z_i(x) = x_i + i\sum_{n_1,n_2,\ldots} c_i(n_1 n_2 \ldots) H_{n_1}(x_1) H_{n_2}(x_2) H_{n_3}(x_3) \cdots$$

- **Neural network**
  - **Combination of linear and non-linear transformation.**

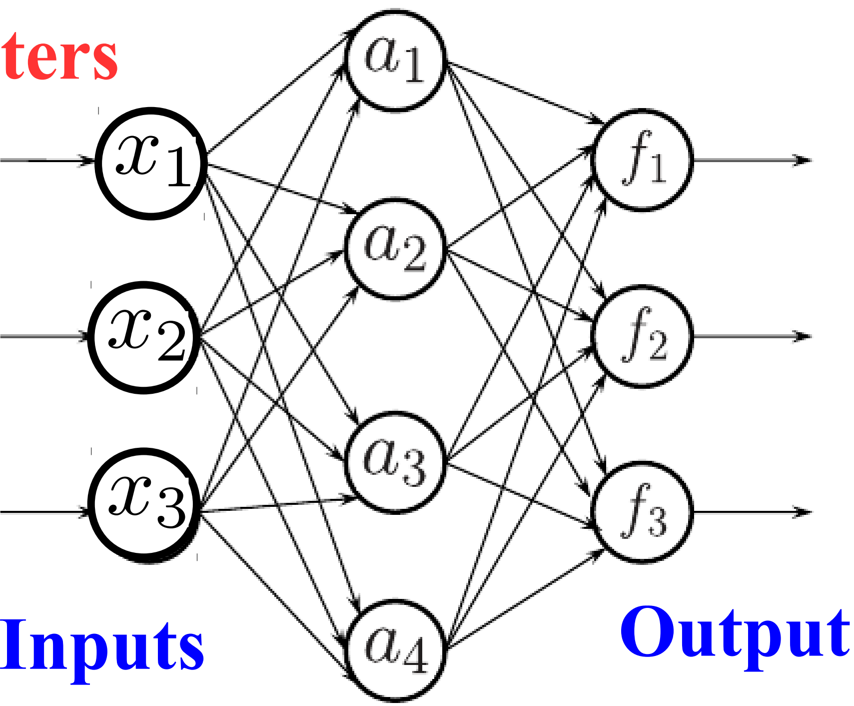
$$a_i = g(\underline{W_{ij}^{(1)}} x_j + \underline{b_i^{(1)}})$$

$$f_i = g(\underline{W_{ij}^{(2)}} a_j + \underline{b_i^{(2)}})$$

$$z_i = x_i + i(\underline{\alpha_i} f_i(x) + \underline{\beta_i})$$

$$g(x) = \tanh x \ \text{(activation fn.)}$$

parameters



  - **Universal approximation theorem**
    Any fn. can be reproduced
    at (hidden layer unit #) → ∞
    G. Cybenko, MCSS 2 ('89) 303
    K. Hornik, Neural networks 4('91) 251

Optimization of many parameters

- **Stochastic Gradient Descent method, E.g. ADADELTA algorithm**
  M. D. Zeiler, arXiv:1212.5701

Grad. Desc. :

$$dc_i/dt = -\partial\mathcal{F}/\partial c_i$$

par. in (j+1)th step → $c_i^{(j+1)}$

Learning rate → η

mean sq. ave. of v → $s_i^{(j)}$

mean sq. ave. of F → $r_i^{(j+1)}$

decay rate → γ

$$c_i^{(j+1)} = c_i^{(j)} - \eta v_i^{(j+1)}$$

$$v_i^{(j+1)} = \frac{\sqrt{s_i^{(j)} + \epsilon}}{\sqrt{r_i^{(j+1)} + \epsilon}} F_i^{(j)}$$

$$r_i^{(j+1)} = \gamma r_i^{(j)} + (1-\gamma)(F_i^{(j)})^2$$

$$s_i^{(j+1)} = \gamma s_i^{(j)} + (1-\gamma)(v_i^{(j+1)})^2$$

$$F_i = \partial\mathcal{F}/\partial c_i$$

gradient evaluated in MC (batch training) → F_i

Cost fn. → $\mathcal{F}$

> ***Machine learning***
> ***~ Educated algorithm***
> ***to generic problems***

Hybrid Monte-Carlo with Neural Network

Initial Config. on Real Axis

HMC $H(x,p) = \frac{p^2}{2} + \mathrm{Re}S(z(x))$

Jacobian
→ via Metropolis judge

Do k = 1, Nepoch
Do j = 1, Nconf/Nbatch

Mini-batch training of Neural Network $c_i^{(j+1)} = c_i^{(j)} - \eta v_i^{(j+1)}$

Grad. wrt parameters (Nbatch configs.) $F_i = \frac{1}{N_{\mathrm{batch}}} \sum_n \partial\mathcal{F}(n)/\partial c_i$

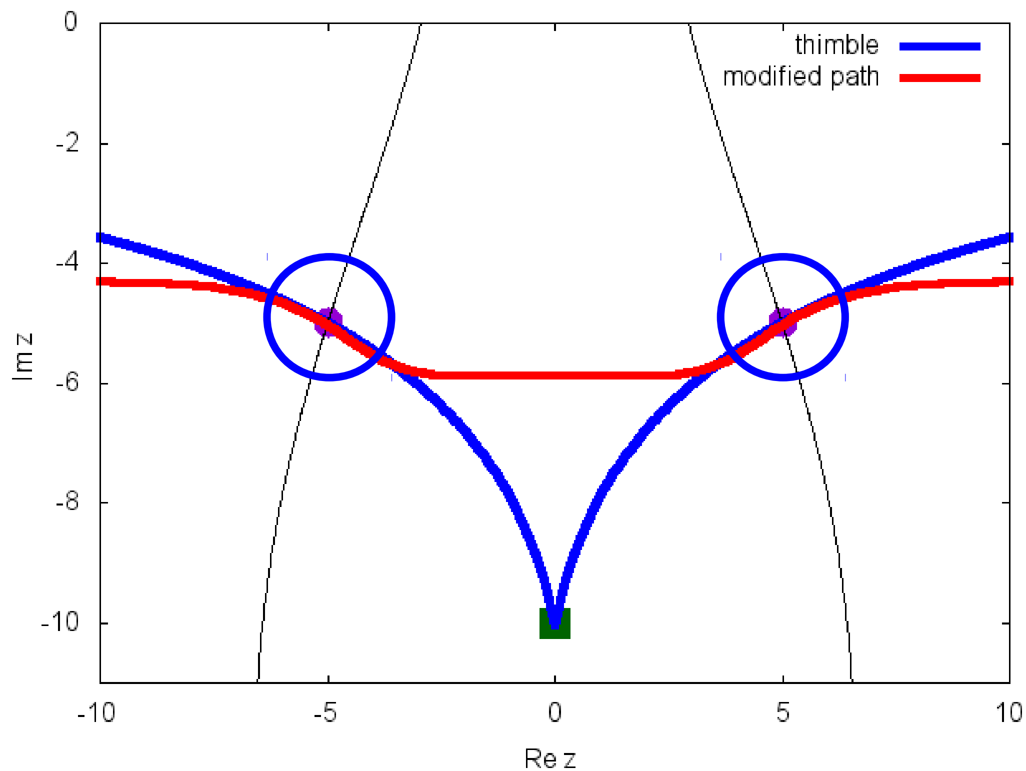
New Nbatch configs. by HMC $H(x,p) = \frac{p^2}{2} + \mathrm{Re}S(z(x))$

Enddo
Enddo

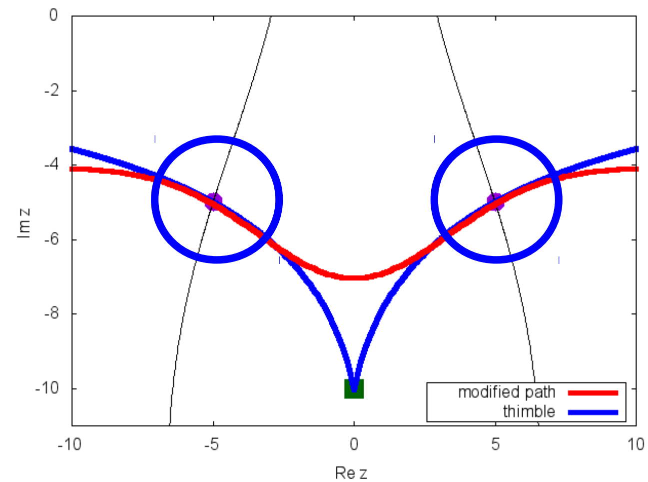
Nbatch ~ 10, Nconfig ~ 10,000, Nepoch ~ (10-20)

Optimized Path by Neural Network

Neural Network



Gaussian +Gradient Descent



Optimized paths are different,
but both reproduce thimbles around the fixed points !

AO, Mori, Kashiwa (Lat 2017)

Complex φ⁴ theory at finite μ

- **Complex φ⁴ theory**

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - m^2 \phi^* \phi - \lambda(\phi^*\phi)^2$$

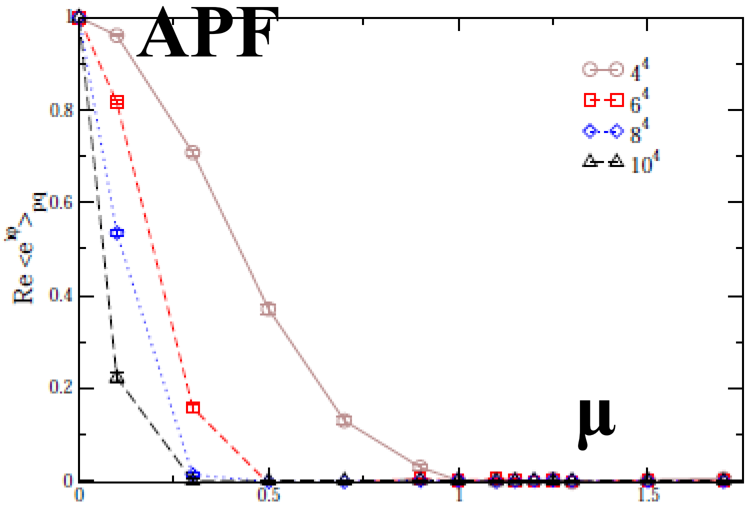
- **Action on Eucledean lattice at finite μ.**

$$S = \sum_x \left[\frac{(4+m^2)}{2} \phi_{a,x}\phi_{a,x} + \frac{\lambda}{4}(\phi_{a,x}\phi_{a,x})^2 - \phi_{a,x}\phi_{a,x+\hat{\imath}} - \cosh\mu\, \phi_{a,x}\phi_{a,x+\hat{0}} \right.$$

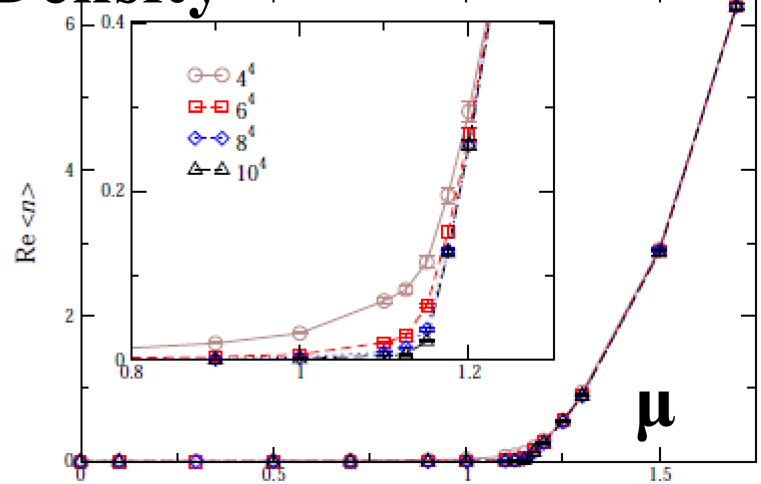
$$\left. + i\epsilon_{ab} \sinh\mu\, \phi_{a,x}\phi_{b,x+\hat{0}} \right] \quad \left(\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \right)$$

complex **Complexify**

APF



Density

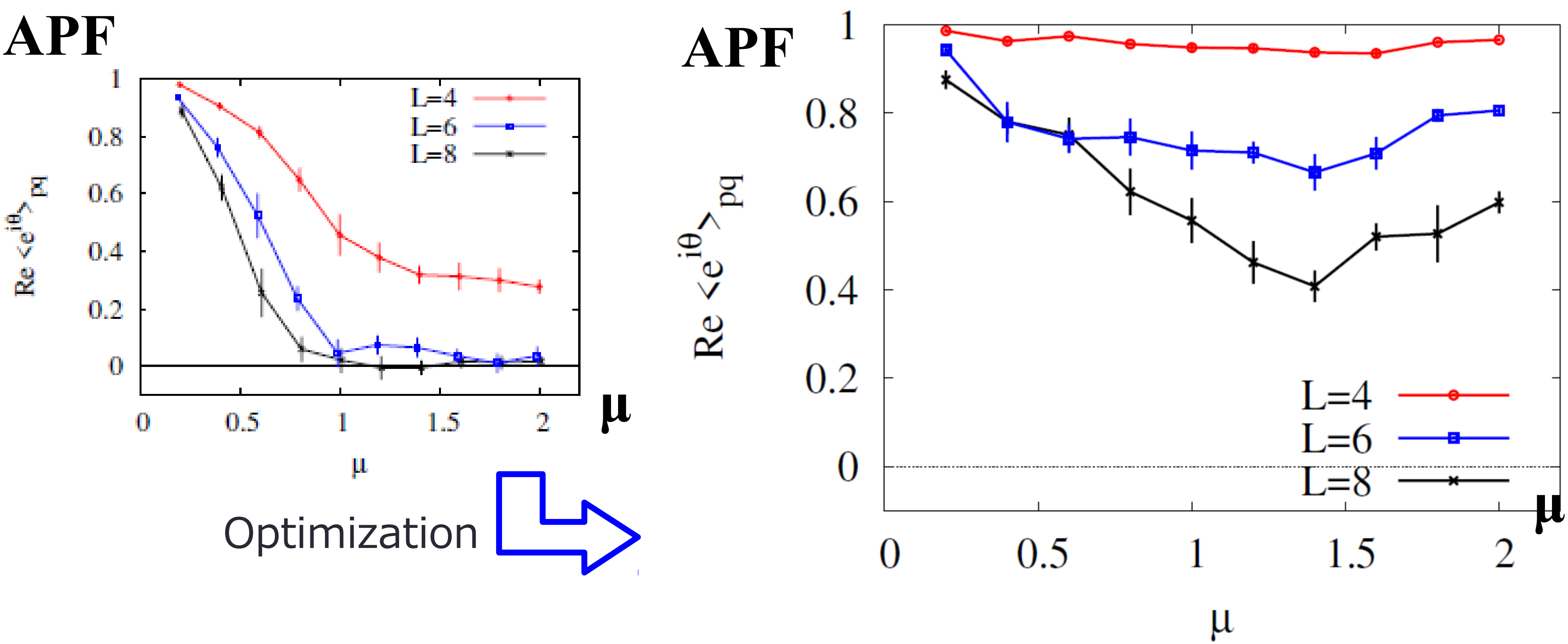


Complex Langevin & Lefschetz thimble work.

G. Aarts, PRL102('09)131601; H. Fujii, et al., JHEP 1310 (2013) 147

POM result (1): Average phase factor

- **POM for 1+1D φ^4 theory**
  - **4^2, 6^2, 8^2 lattices, λ=m=1**
  - **$\mu_c \sim 0.96$ in the mean field approximation**
  - **Enhancement of the average phase factor after optimization.**

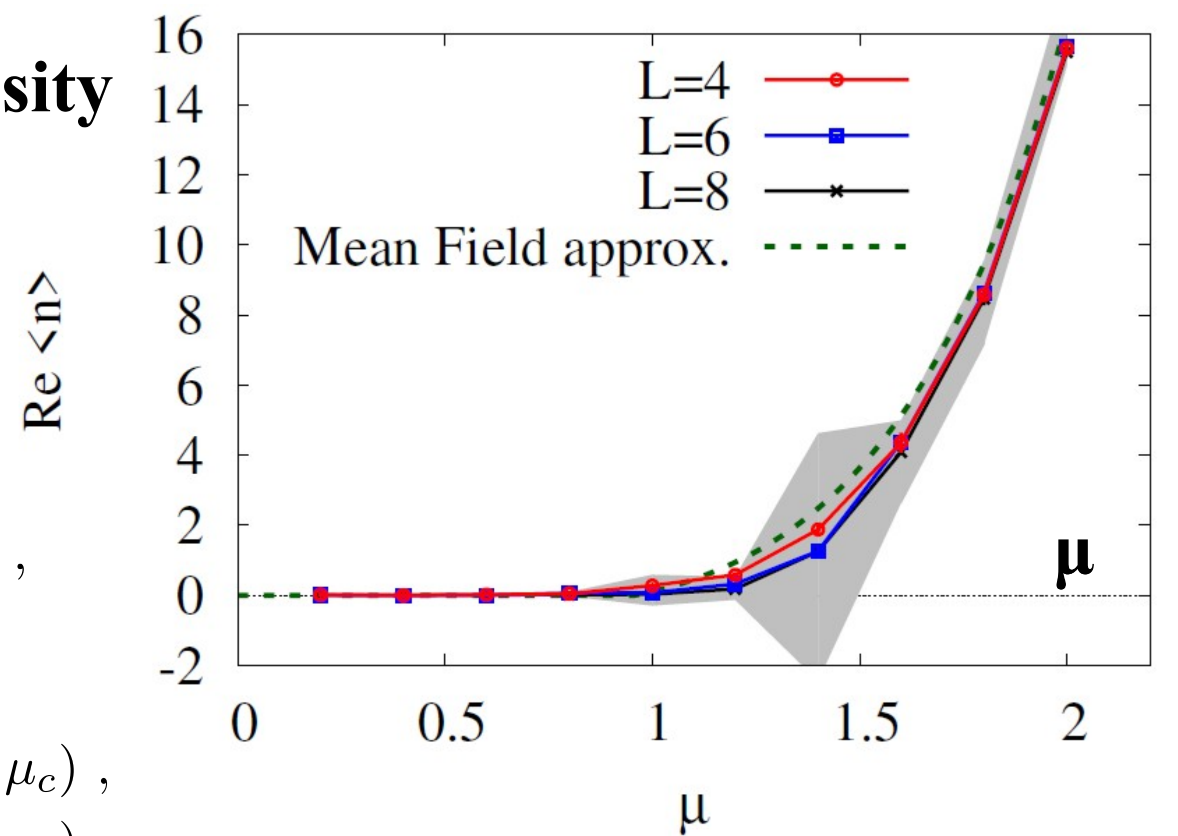


Y. Mori, K. Kashiwa, AO, PTEP 2018 ('18), 023B04 [arXiv:1709.03208]

POM result (2): Density

- **Results on the real axis**
  Small average phase factor, Large errors of density
- **On the optimized path**
  Finite average phase factor, Small errors

Density



Mean Field App.

$$\frac{S}{V} = \left(1 + \frac{m^2}{2} - \cosh\mu\right)\phi^2 + \frac{\lambda}{4}\phi^4 \ ,$$

$$n = \phi^2 \sinh\mu \ ,$$

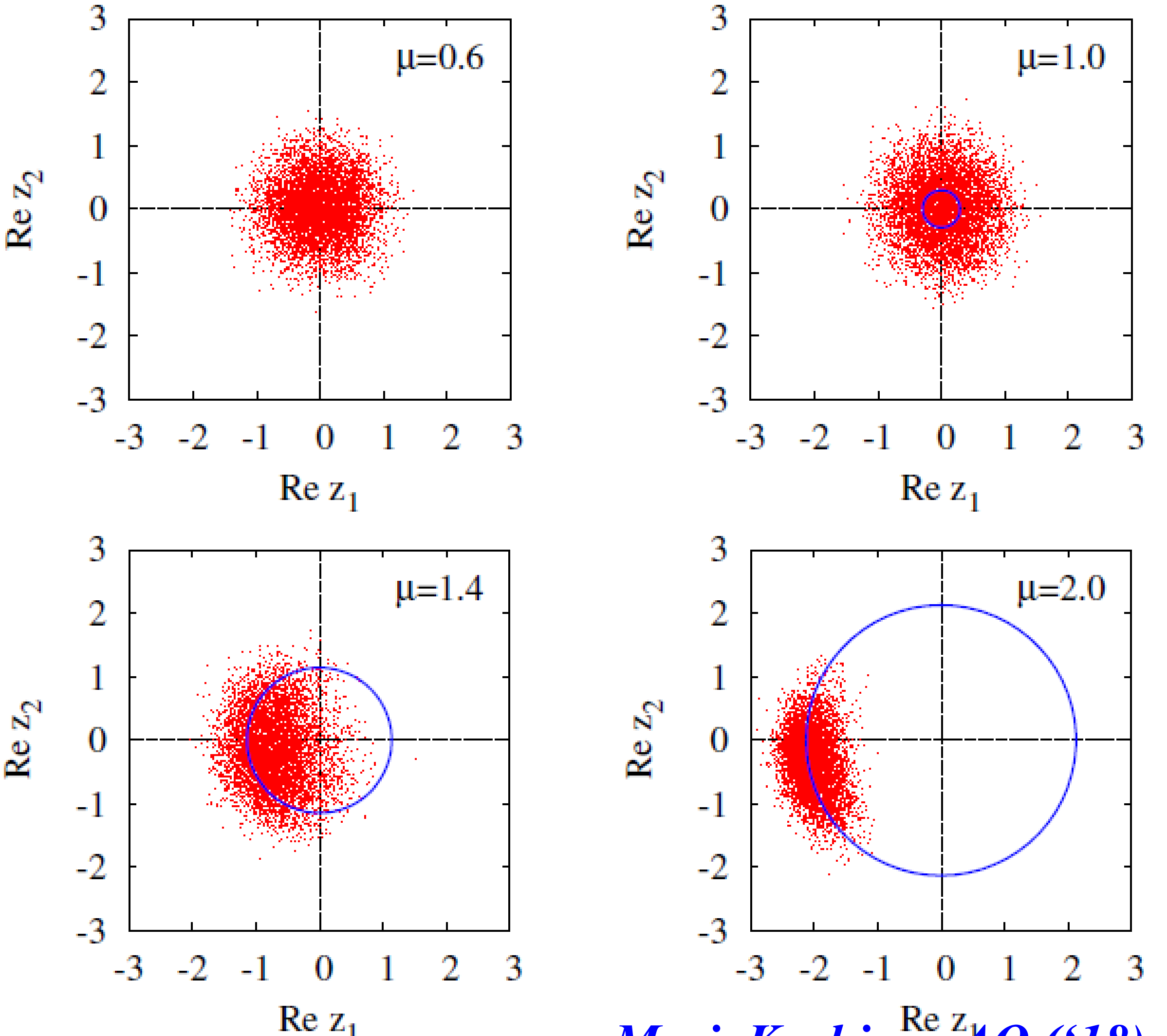
$$\phi^2_{\mathrm{stat.}} = \begin{cases} 0 & (|\mu| < \mu_c) \ , \\ \frac{2}{\lambda}\left(\cosh\mu - 1 - \frac{m^2}{2}\right) & (|\mu| \geq \mu_c) \ , \end{cases}$$

Mori, Kashiwa, AO ('18)

POM result (3): Configurations

- **Updated configurations after optimization**
  → sampled around the mean field results

- **Global U(1) symmetry in (φ_1, φ_2) is broken(*) by the optimization or by the sampling.**



*** This does not contradict the Elitzur's theorem.**

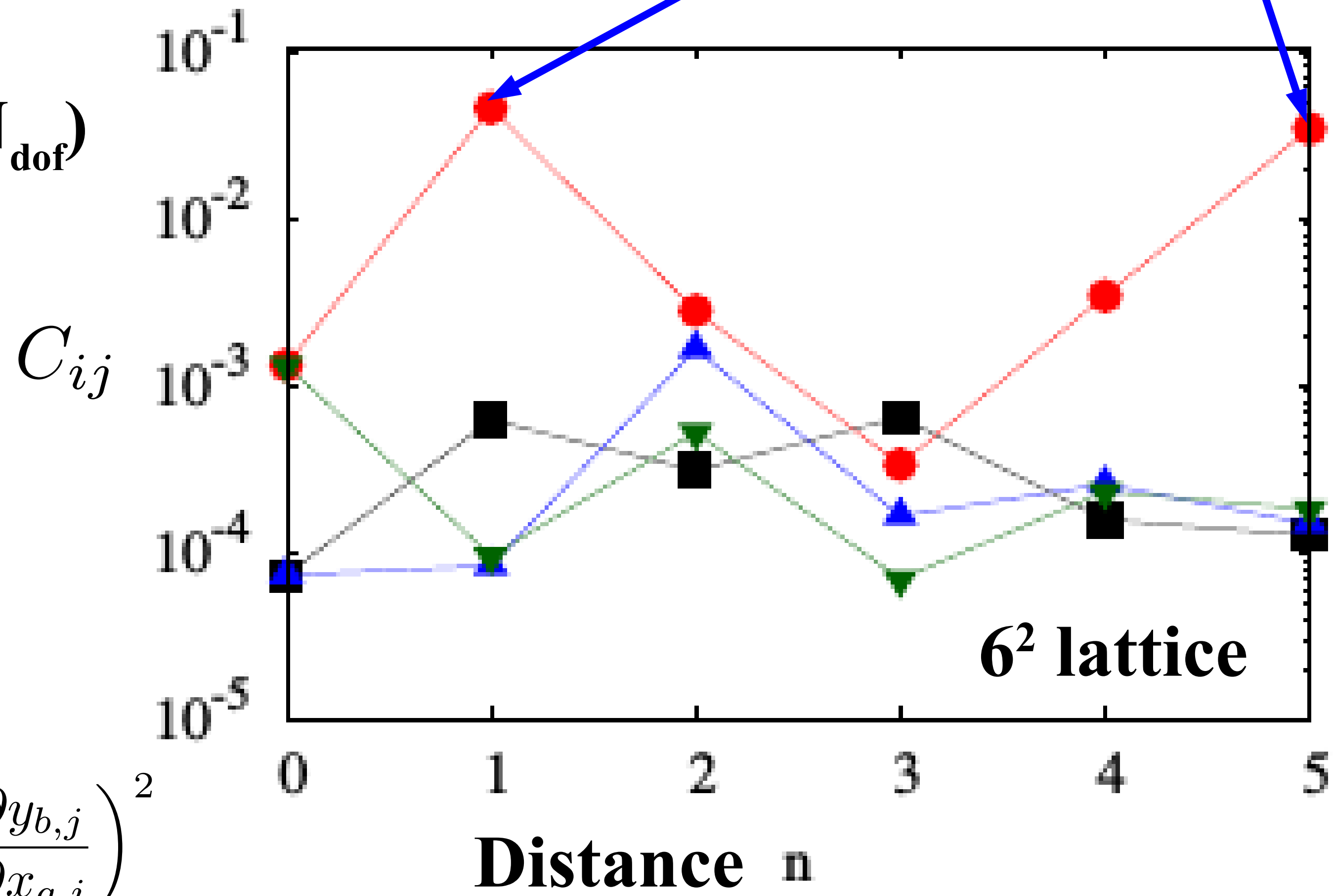
Mori, Kashiwa, AO ('18)

Which y's should be optimized ?

- **Correlation btw (z_1,z_2) of temporal nearest neighbor sites are strong. Other correlations ~ 10^{-2} times smaller**

$$\mathrm{Im}(S) = \sum_x \epsilon_{ab} \sinh \mu \, \phi_{a,x} \phi_{b,x+\hat{0}}$$

- **Hope to reduce the cost to be O(N_{dof})**



$$C_{ij} \equiv \left(\frac{\partial y_{a,i}}{\partial x_{b,j*}}\right)^2 + \left(\frac{\partial y_{b,j}}{\partial x_{a,i}}\right)^2$$

Y. Mori, Master thesis

Application to Gauge Theory: 1 dimensional QCD

0+1 dimensional QCD

- **0+1 dimensional QCD (1 dim. QCD) with one species of staggered fermion on a 1xN lattice**

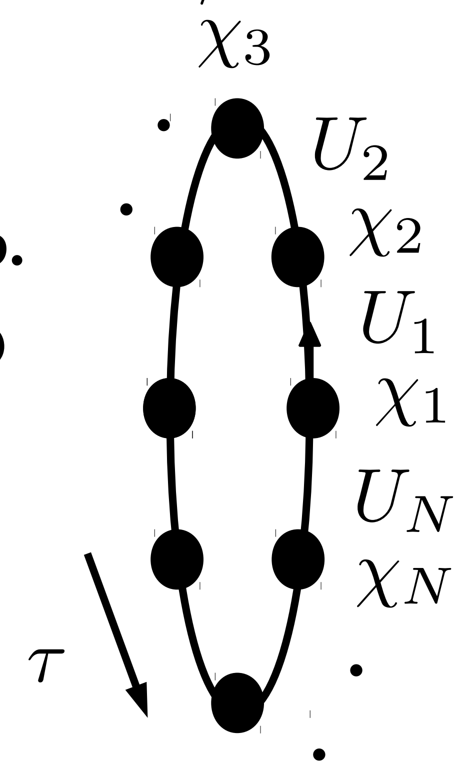
$$S = \frac{1}{2}\sum_\tau \left(\bar{\chi}_\tau e^{\mu} U_\tau \chi_{\tau+\hat{0}} - \bar{\chi}_{\tau+\hat{0}} e^{-\mu} U_\tau^{-1} \chi_\tau\right) + m \sum_\tau \bar{\chi}_\tau \chi_\tau = \frac{1}{2}\bar{\chi} D \chi$$

$$\mathcal{Z} = \int \mathcal{D}U \det D[U] = \int dU \det\left[X_N + (-1)^N e^{\mu/T} U + e^{-\mu/T} U^{-1}\right]$$

$$X_N = 2\cosh(E/T)\ ,\ E = \operatorname{arcsinh} m\ ,\ U = U_1 U_2 \cdots U_N\ ,\ T = 1/N$$

Bilic+('88), Ravagli+('07), Aarts+('10, CLM), Bloch+('13, subset), Schmidt+('16, LTM), Di Renzo+('17, LTM)

  - **A toy model, but the actual source of QCD sign prob.**
  - **Studied well in the context of strong coupling LQCD**
    E.g. Miura, Nakano, AO, Kawamoto('09,'09,'17), de Forcrand, Langelage, Philipsen, Unger ('14)



1 dim. QCD in diagonal gauge

- **Diagonal gauge**

$$U = (e^{iz_1}, e^{iz_2}, e^{iz_3}) \quad (z_1 + z_2 + z_3 = 0)$$

$$\mathcal{Z} = \int dU e^{-S} = \int dx_1 dx_2 J H e^{-S}$$

$$= \int dx_1 dx_2 \det\left(\frac{\partial z_a}{\partial x_b}\right) \left[\frac{8}{3\pi^2} \prod_{a<b} \sin^2\left(\frac{z_a - z_b}{2}\right)\right] \left[\prod_a (X_N + 2\cos(z_a - i\mu))\right]$$

Jacobian **Haar measure** **exp(-S)**

- **Path optimization (t: ficticious time)**
  → y(x_1,x_2) itself is the parameter on the (x_1,x_2) mesh point

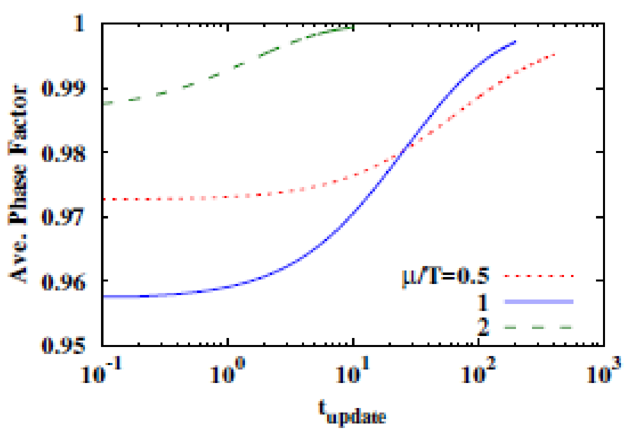
$$z_i = x_i + iy_1,\ y_i = y_i(x_1, x_2)$$

$$\frac{dy_i}{dt} = -\frac{\partial \mathcal{Z}_{\mathrm{pq}}}{\partial y_i},\ \mathcal{Z}_{\mathrm{pq}} = \int dx_1 dx_2 |J H\, e^{-S}|$$

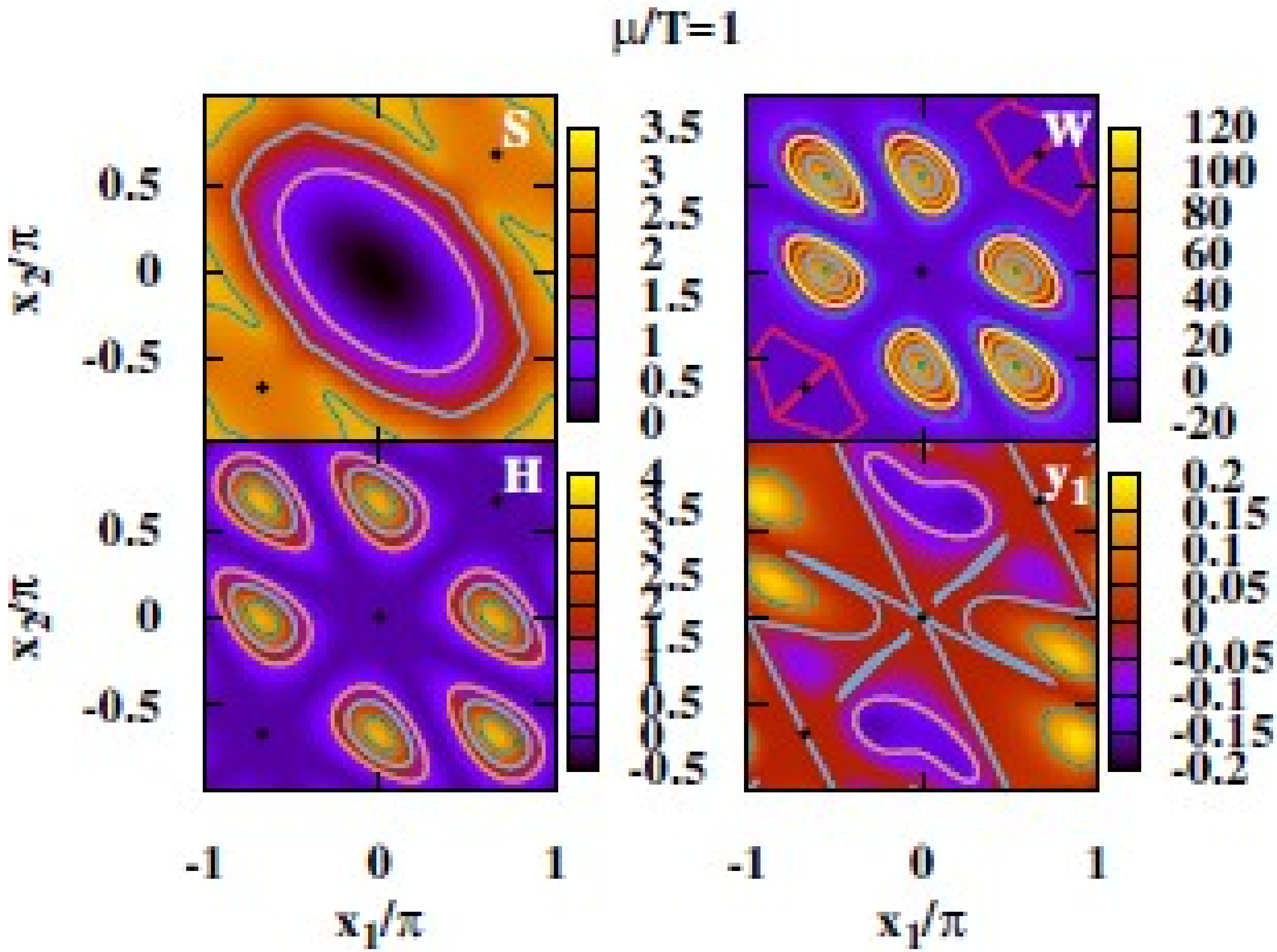
Path Opt. of 1 dim. QCD in diagonal temporal gauge

- Path optimization
  - Average phase factor > 0.99 → Easily achieved
  - exp(-S) and Haar Mesure → “six pads” *Schmidt+('16, LTM)*

APF



fictitious time



Mori, Kashiwa, AO, in prep.

1 dim. QCD with Hybrid MC

- **Concern…**
  - **Six pads are separated by the Haar measure barrier.**

    $$\text{Symmetry}: \quad S(-z) = (S(z^*))^*, z_i \leftrightarrow z_j (i,j = 1,2,3)$$

    Do we need exchange MC or different tempering ?
    E.g. Fukuma, Matsumoto, Umeda ('17)

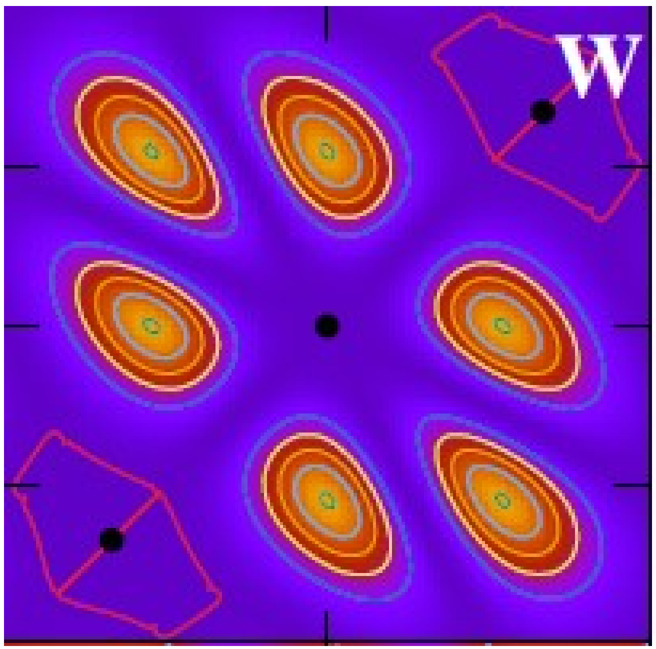
- **Hybrid Monte-Carlo in 1 dim. QCD**

$$U \to \mathcal{U}(U) = U \underbrace{\prod_{a=1}^{N_c^2-1} e^{-y_i \lambda_i/2}}_{\text{SL(3)}}, \quad H = \frac{P^2}{2} + \mathrm{Re}(S(\mathcal{U}(U)))$$

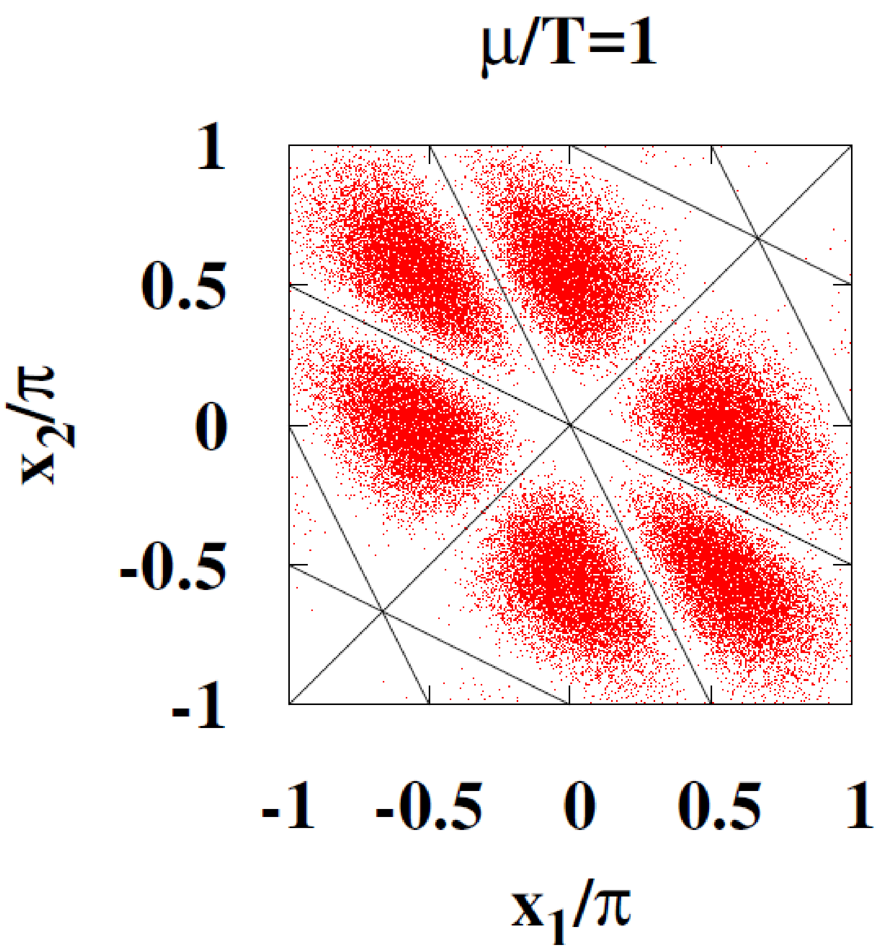
  - **8 variables → path optimization using Neural Network**

1 dim. QCD with Hybrid MC

- HMC + diagonalization of the link
  → All six pads are visited, and no Ex. MC needed.



Mesh point + Grad. Desc.



HMC + NN

Mori, Kashiwa, AO, in prep.

Discussions

Frequently Asked Questions

- **How many parameters do you have ?**
  → Many ;) For generic trial funciton (V= # of variables)

$$y_i = y_i(x_1, x_2, \ldots x_V)$$

$$N_{\rm par} = (N_{\rm layer} + 1) \times V \times (N_{\rm unit} + 1) + 2V$$

- **How about the numerical cost ?**
  → A lot ;) Derivative of J with respect to parameters cost most.

$$\frac{\partial J}{\partial c_i} = J \frac{\partial J^{-1}_{jk}}{\partial z_l} \frac{\partial z_l}{\partial c_i} \to \mathcal{O}(V^3)$$

- **It is still polynomial.**
  Does the sign problem becomes "P" problem?
  → No. The average phase factor is still exp(-# V).
  If extrapolation is possible from finite V, we have a hope.

- **How can we reduce the cost ? → Next page**

How can we reduce the numerical cost ?

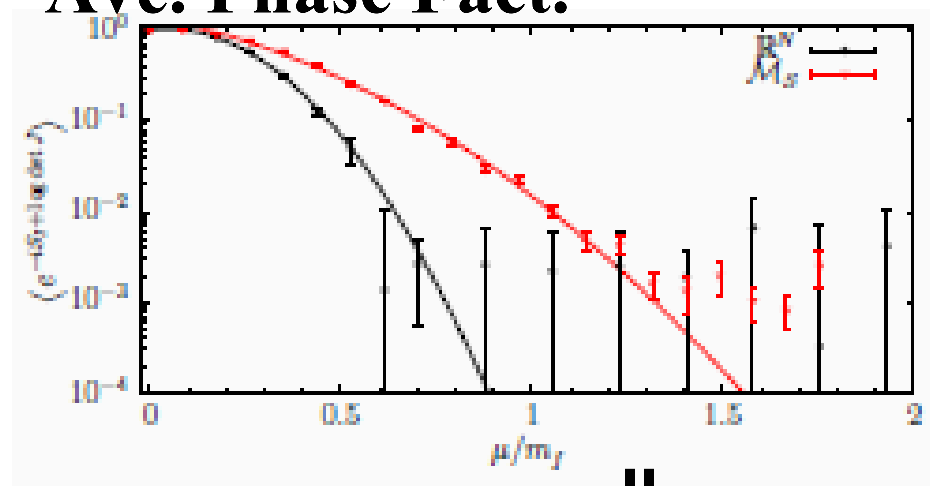
- **Restrict the function form of y(x).**
  - **Imaginary part is a function of its real part.**
    E.g. Alexandru, Bedaque, Lamm, Lawrence, PRD97('18)094510
    [Lawrence, Warrington, Lamm (Mon)]
    Thirring model, 1+1D QED

$$y_i = f(x_i), f(x) = \lambda_0 + \lambda_1 \cos x$$

  - **Nearest neighbor site**
    F. Bursa, M. Kroyter, arXiv:1805.04941
    0+1 D φ^4 theory
    Translational inv. + U(1) sym.

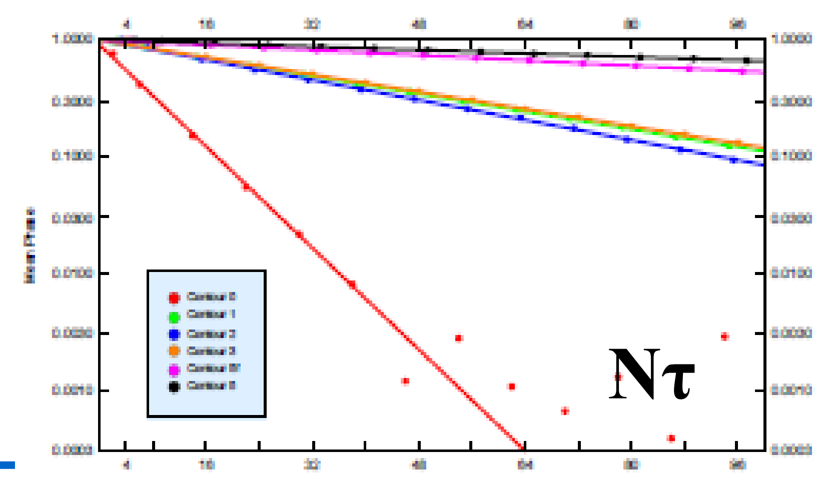
$$y_{a,i} = \frac{\varepsilon_{ab} x_{a,i+1}}{1 + x_{1,i}^2 + x_{2,i}^2}$$

Ave. Phase Fact.



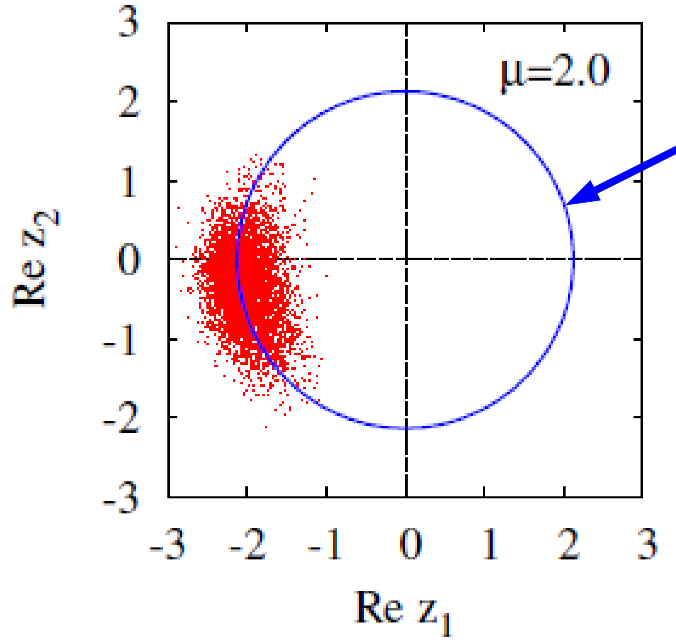
μ

Ave. Phase Fact.



Frequently Asked Questions (cont.)

- **What happens when we have 10^{10} fixed points ?**
  → In that case we should give up. (My answer @ Lattice 2017)
  → If those fixed points are connected by the symmetry,
  we may be able to perform path optimization.

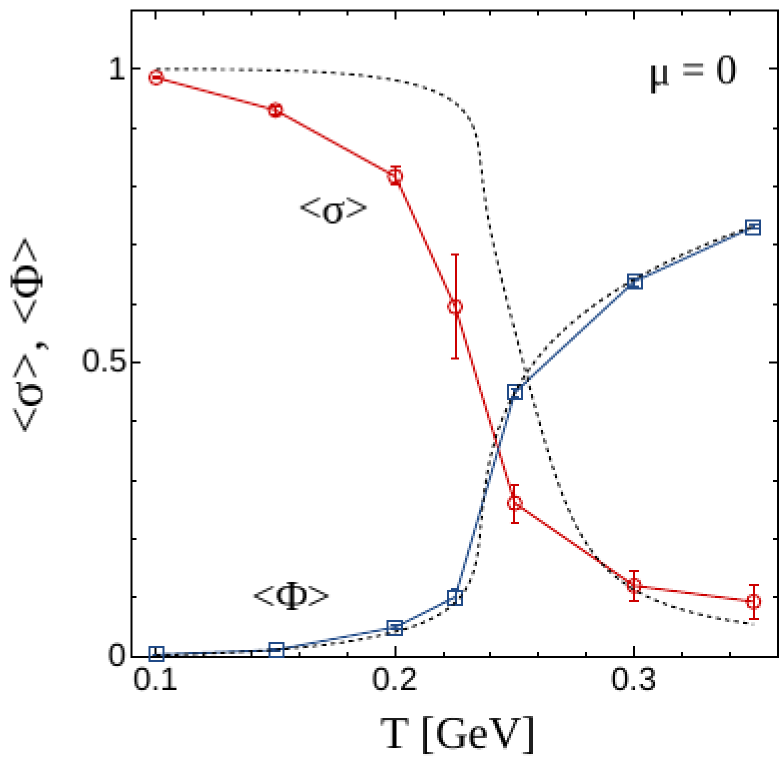


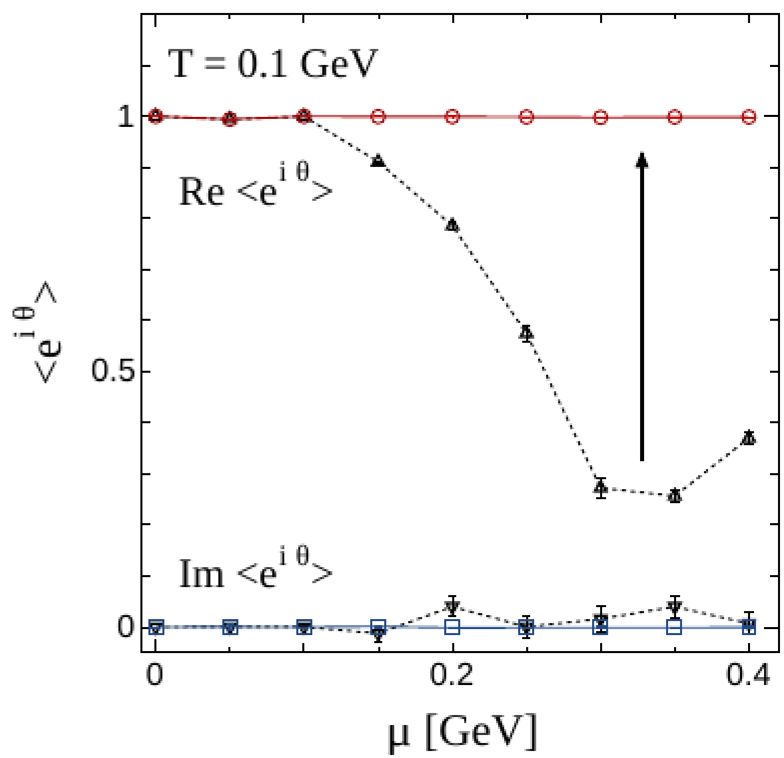
Mean field results
= Degenerate fixed points
(All have the same θ.)

If they have different complex phases, the global sign problem emerges and the partition function would be almost zero.
E.g. H. Fujii, S. Kamata, Y. Kikukawa, arXiv:1710.08524

Application to PNJL

- **PNJL model with homogeneous condensates, (σ, π, Φ, $\bar{\Phi}$).**
  - **Has Sign problem in finite volume**
  - **Converges to mean field results in the large volume limit**





K. Kashiwa, Y. Mori, AO, arXiv:1805.08940.

Summary

- The sign problem is a grand challenge in theoretical physics, and appears in many fields of physics,
  - finite density QCD, real time evolution, Hubbard model off half-filling, other quantum MC with fermions, …

  and complexified variable methods (LTM, CLM, POM) would be promising to evade the sign problem.
- Path optimization with the use of the neural network is demonstrated to work in field theories having many variables.
  - 1+1D φ^4 theory at finite μ (neural network)
  - 0+1D QCD w/ fermions (grad. descent, neural network)
  - 3+1D homogeneous PNJL (neural network)
- Neural network (single hidden layer) is the simplest device of machine learning, and it helps us to generate and optimize generic multi-variable functions, $y_i = y_i(\{x\})$.

Prospect

- **Path optimization in 3+1 D field theories would require reduction of numerical cost.**
  Imaginary part
  = f (real parts of same point and nearest neighbor points)
  may be a good guess.
- **Deep learning (# of hidden layers > 3) may be helpful to explore complex path, which human beings (~ 7 layers) cannot imagine, while "Understanding" the results of machine learning need to be done by human beings (at present).**

Thank you for your attention !