

# *Path optimization method to avoid the sign problem*

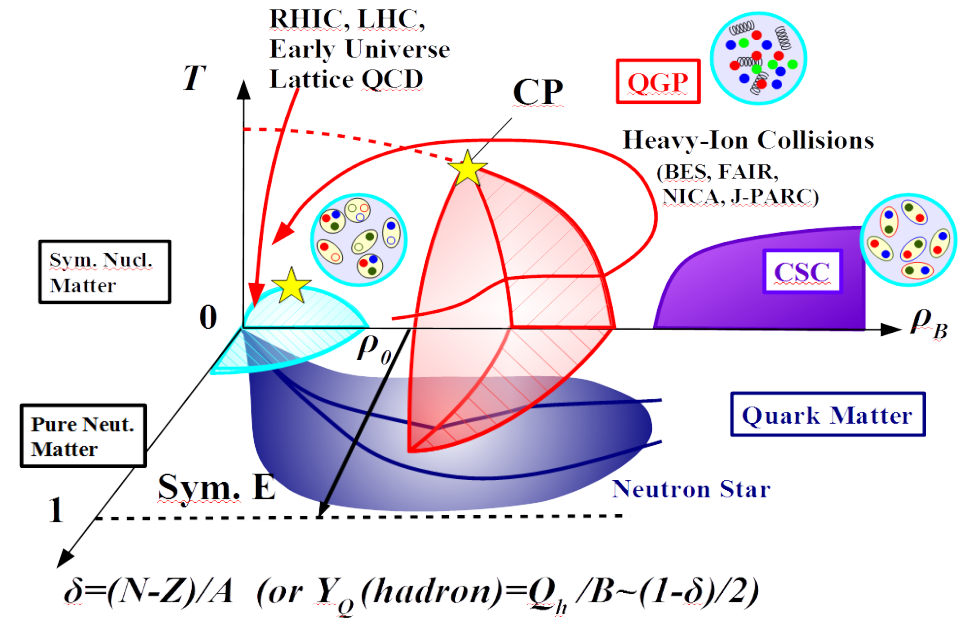
**Akira Ohnishi (YITP)  
in collaboration with  
Yuto Mori (Dept. Phys., Kyoto U.),  
Kouji Kashiwa (Fukuoka Inst. Tech.)**

*Y. Mori, K. Kashiwa, AO, PRD 96 ('17), 111501(R) [arXiv:1705.05605]  
Y. Mori, K. Kashiwa, AO, PTEP 2018 ('18), 023B04 [arXiv:1709.03208]  
AO, Y. Mori, K. Kashiwa, EPJ Web Conf. 175 ('18), 07043  
[arXiv:1712.01088, Lattice 2017 proceedings]  
Y. Mori, K. Kashiwa, AO, PLB, in press [arXiv:1705.03646]  
K. Kashiwa, Y. Mori, AO, arXiv:1805.08940.  
Y. Mori, K. Kashiwa, AO, in prog.*



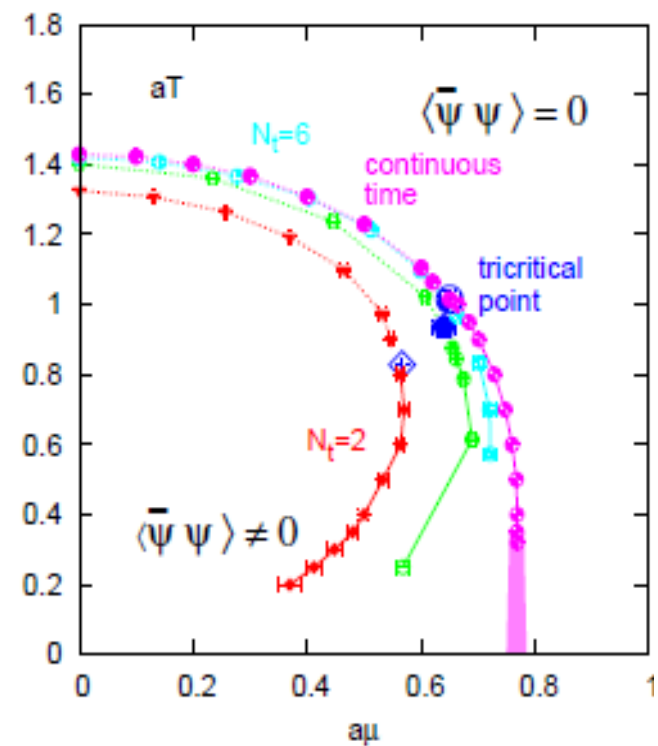
# Introduction

- Sign problem for complex actions
  - Grand challenge in theor. phys.
  - Largest obstacle to explore QCD phase diagram
- Approaches
  - Taylor expansion, Analytic cont., Canonical, Strong coupling, ...

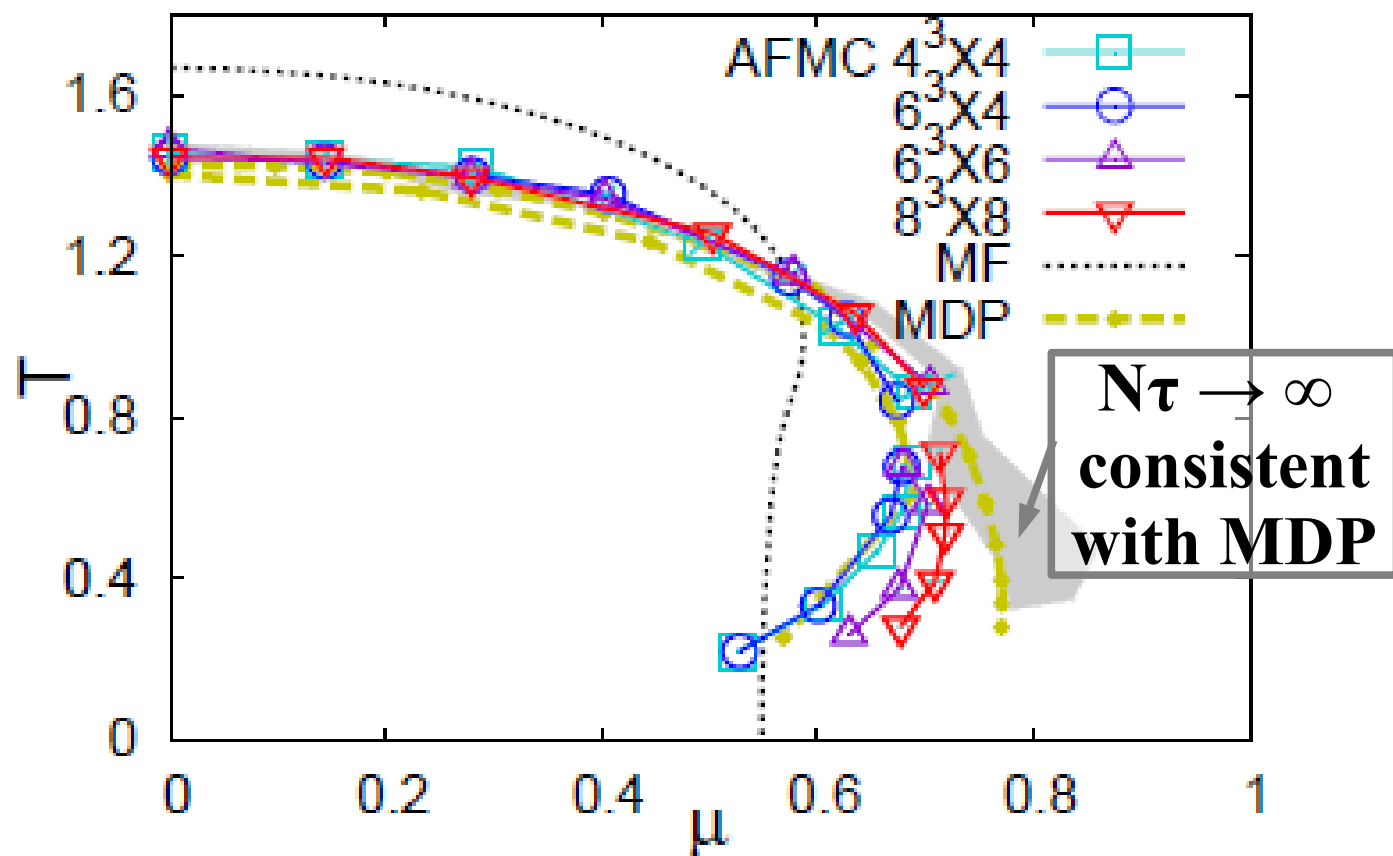


# Phase diagram

- Phase diagrams in two independent methods (MDP & AFMC) agree with each other in the strong coupling limit.  
→ SCL phase diagram is determined !



*de Forcrand, Fromm ('10),  
de Forcrand, Langelage,  
Philipsen, Unger ('14)*



*Ichihara, AO, Nakano ('14)*

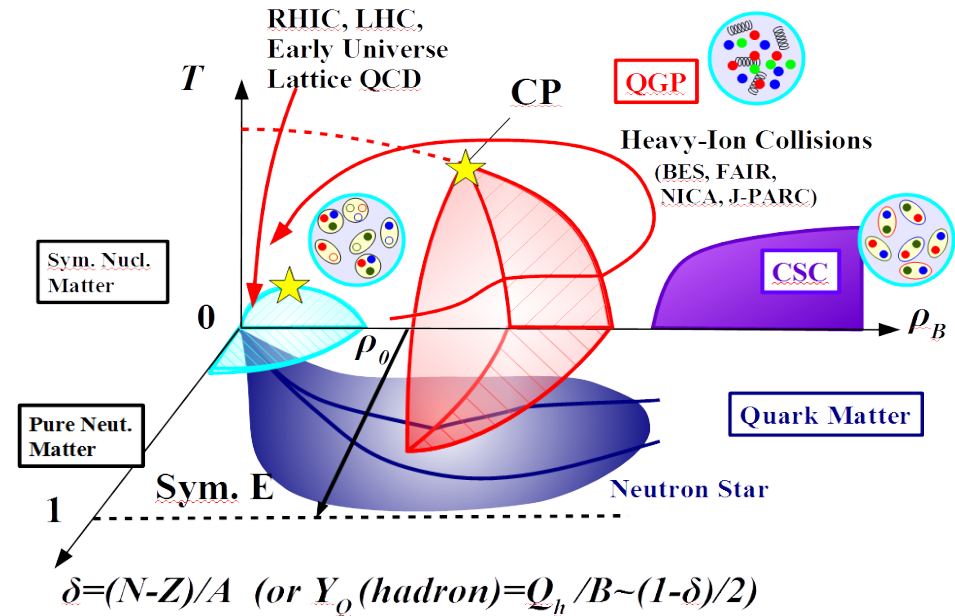
# Introduction

## ■ Sign problem for complex actions

- Grand challenge in theor. phys.
- Largest obstacle to explore QCD phase diagram

## ■ Approaches

- Taylor expansion, Analytic cont., Canonical, Strong coupling, ...



### ● Complex Langevin method (CLM)

*Parisi ('83), Klauder ('83), Aarts et al. ('11), Nagata et al. ('16)*

### ● Lefschetz thimble method (LTM)

*E. Witten ('10), Cristoforetti et al. ('12), Fujii et al. ('13)*

### ● Generalized LTM (GLTM)

*A. Alexandru, et al., ('16)*

*Complexified variables & Shifting path (area)*

# Lefschetz Thimble Method

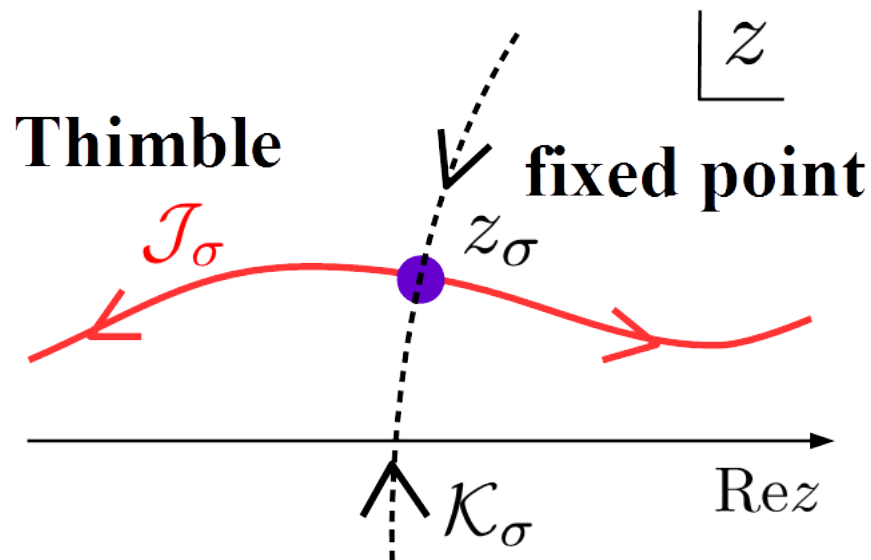
- Integral over thimbles defined by the flow equation for complexified variables  
 →  $\text{Im } S = \text{const. on a thimble}$

*E. Witten ('10), Cristoforetti et al. ('12), Fujii et al. ('13)*

$$\mathcal{Z} = \int_{\mathcal{C}_{\mathbb{R}}} \mathcal{D}x e^{-S[x]} = \int_{\mathcal{C}} \mathcal{D}z e^{-S[z]}$$

$$\left. \frac{\partial S}{\partial z_i} \right|_{z_{\sigma}} = 0, \quad \frac{dz_i(t)}{dt} = \overline{\left( \frac{\partial S[z]}{\partial z_i} \right)}$$

$$\mathcal{C} = \sum n_{\sigma} \mathcal{J}_{\sigma}, \quad n_{\sigma} = \langle \mathcal{C}_{\mathbb{R}}, \mathcal{K}_{\sigma} \rangle$$



## ■ Problems

- How can we find fixed points ? *Alexandru et al., JHEP 1605 ('16) 053*
- Phase from measure (residual sign prb.)
- Cancellation between thimbles (global sign prb.)
- Flow equation blows up somewhere.

# Complex Langevin Method

- **Sample configurations**  
by solving complex Langevin equation  
for complexified variables.

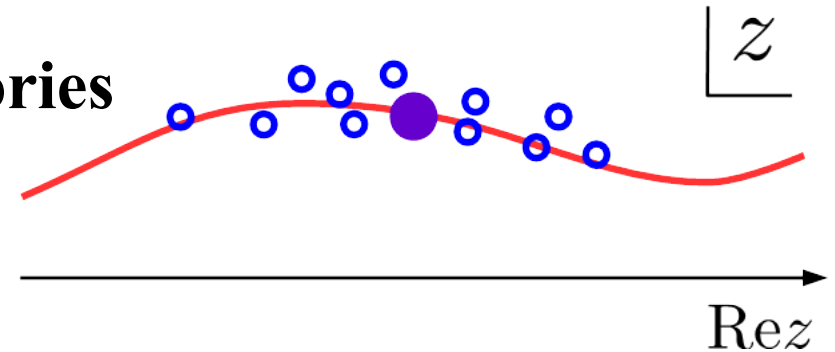
→ **Easier to apply to large DOF theories**

$$\frac{dz_i}{dt} = -\frac{\partial S}{\partial z_i} + \eta_i(t)$$

$$\langle \eta_i(t) \eta_j(t') \rangle = 2\delta_{ij} \delta(t - t')$$

$$\langle \mathcal{O}(x) \rangle = \langle \mathcal{O}(z) \rangle$$

*Parisi ('83), Klauder ('83),  
Aarts et al. ('11), Nagata et al. ('16)*

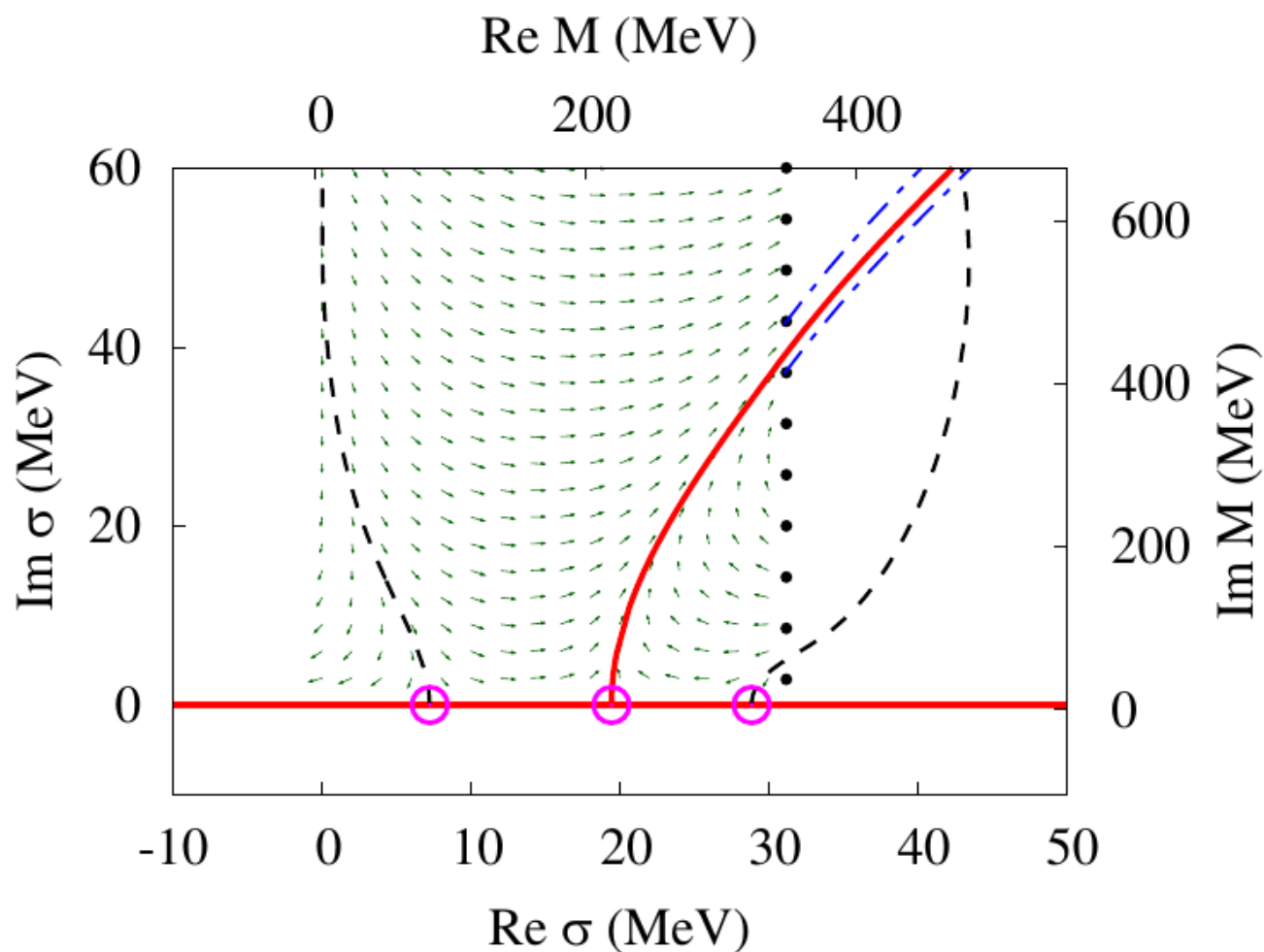


## ■ Problems

- **Excursion problem** → Gauge Cooling (*Seiler et al. ('13)*)
- **Converged results can be wrong**  
→ Criteria (*Nagata, Nishimura, Shimasaki ('16)*)
- **Singular drift problem** → Deformation (*Ito, Nishimura ('16)*)

# Can we avoid singular points of the action ?

- There are many singular points of the action in the complexified variable space, especially around the phase transition boundary.



*Y. Mori, K. Kashiwa, AO, PLB, in press [arXiv:1705.03646]*

---

*Is there any way to obtain the path  
without solving the flow equation  
and without suffering from singular points of the action ?*

*Our proposal: Path Optimization*



- **Introduction**

- **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)*

- **Benchmark Test: 1D integral**

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

- **Application to complex  $\lambda\phi^4$  theory  
with use of neural network**

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

- **Prospects of path optimization**

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

- **Summary**

---

# *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*

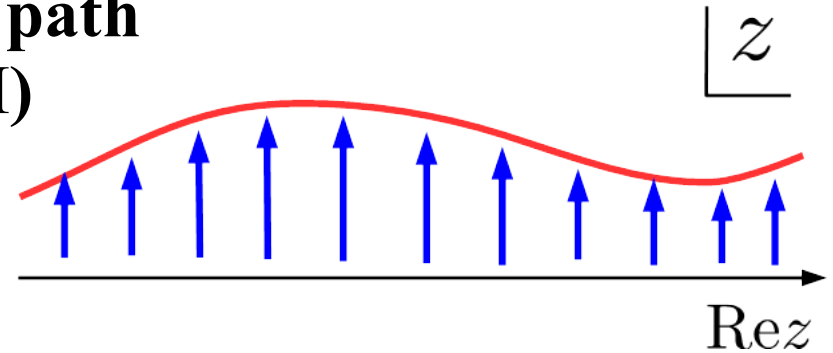
*[arXiv:1712.01088, Lattice 2017 proceedings]*

# Path Optimization Method

- Can we obtain the integration path without solving flow equation ?  
→ Variational shift of the integration path  
(Path Optimization Method: POM)

- POM Procedure

- Parametrize the path appropriately  
(**Trial Function**)
- Set a measure of sign problem  
(**Cost Function**)
- Tune parameters to minimize the Cost Function  
(**Optimization**)



*Sign Problem → Optimization Problem*

# Trial Function, Cost Function, and Optimization

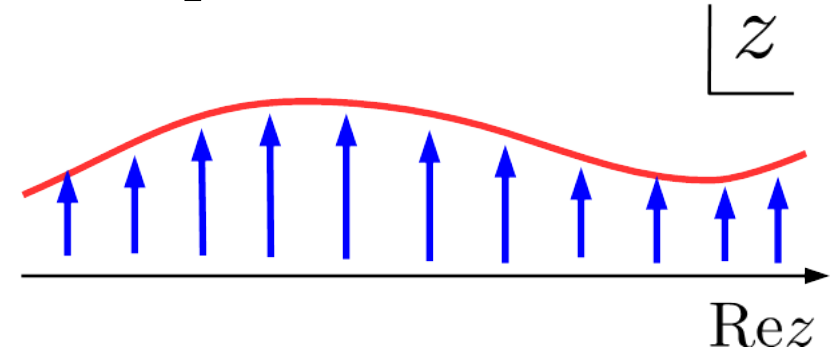
## ■ Parametrize the path in the complex plane (Trial Function)

- Ex. one variable case → Expand in the complete set

$$z(t) = x(t) + iy(t)$$

$$= t + \sum_n (c_n^{(x)} + ic_n^{(y)}) H_n(t)$$

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



## ■ Set the seriousness of the sign problem (Cost Function)

- How much the phase fluctuate

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

$$\frac{F[z(t)]}{|\mathcal{Z}|} = \left| \langle e^{i\theta} \rangle_{\text{pq}} \right|^{-1} - 1 \quad [\theta = \arg(Je^{-S}), \theta_0 = \arg(\mathcal{Z})]$$

## ■ Optimization: Gradient descent, Neural Network, ...

---

# *Benchmark test: 1D integral*

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

# A (Pathological) Toy Model

## ■ A toy model with a serious sign problem

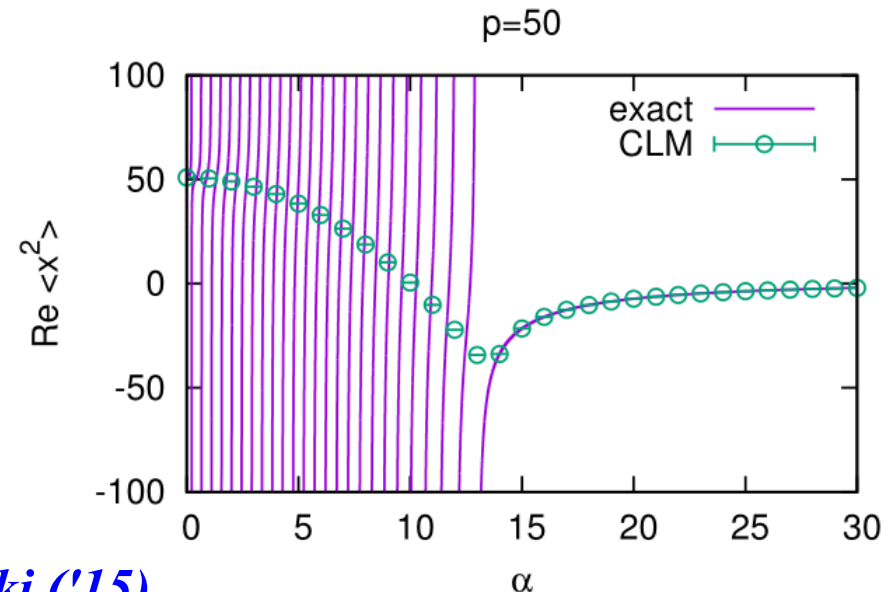
*J. Nishimura, S. Shimasaki ('15)*

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

$$S(x) = x^2/2 - p \log(x + i\alpha)$$

## ■ Complex Langevin Fails at Large $p$ and small $\alpha$

- Large  $p \rightarrow$  Strong oscillation of the Boltzmann weight
- Small  $\alpha \rightarrow$  Singular point at  $z = -i\alpha$  is close to the real axis



*J. Nishimura, S. Shimasaki ('15)*

# Optimized Path

Mori, Kashiwa, AO ('17)

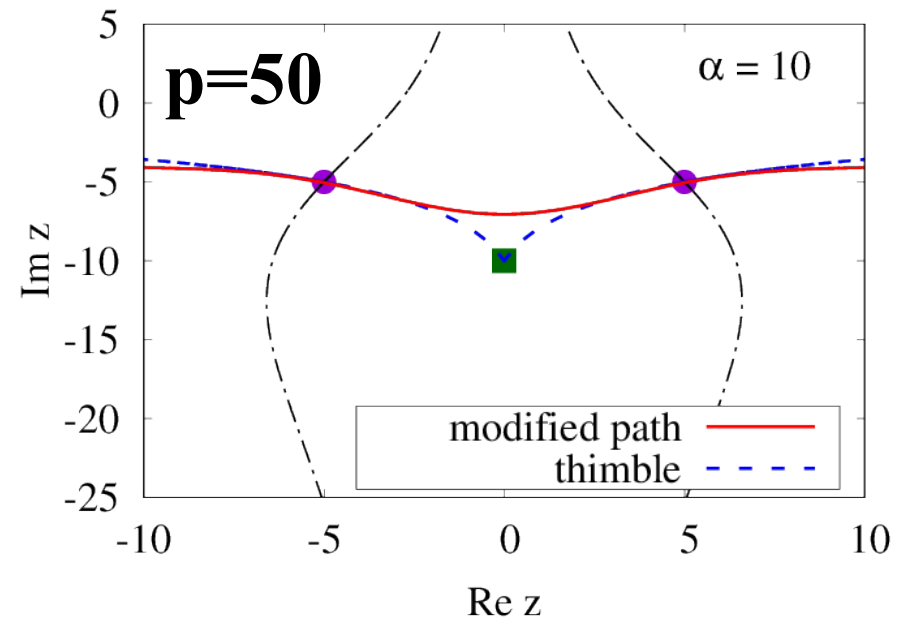
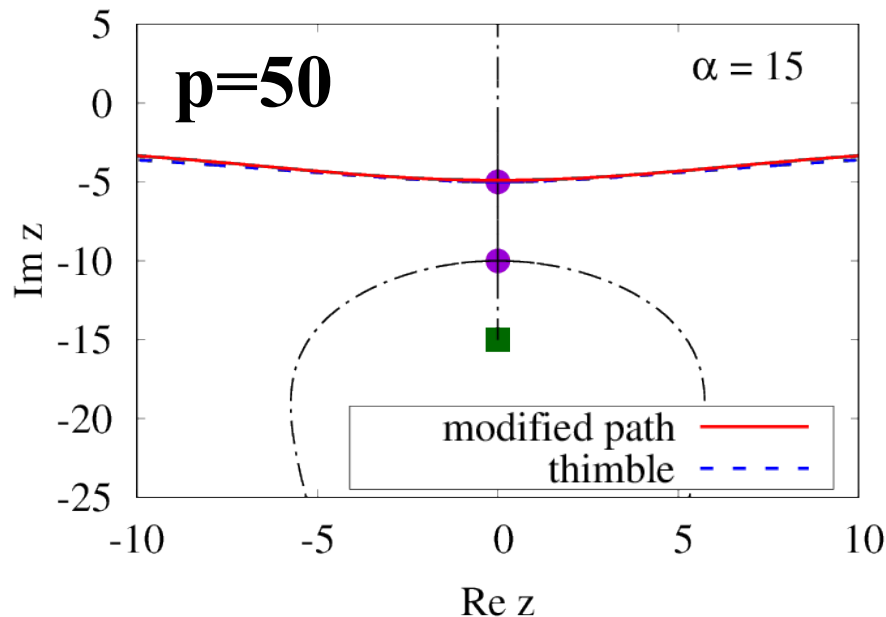
## ■ Trial Function

$$z(t) = t + i \left[ c_1 \exp(-c_2^2 t^2 / 2) + c_3 \right]$$

## ■ Optimization = Gradient descent

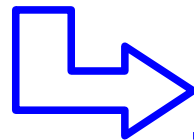
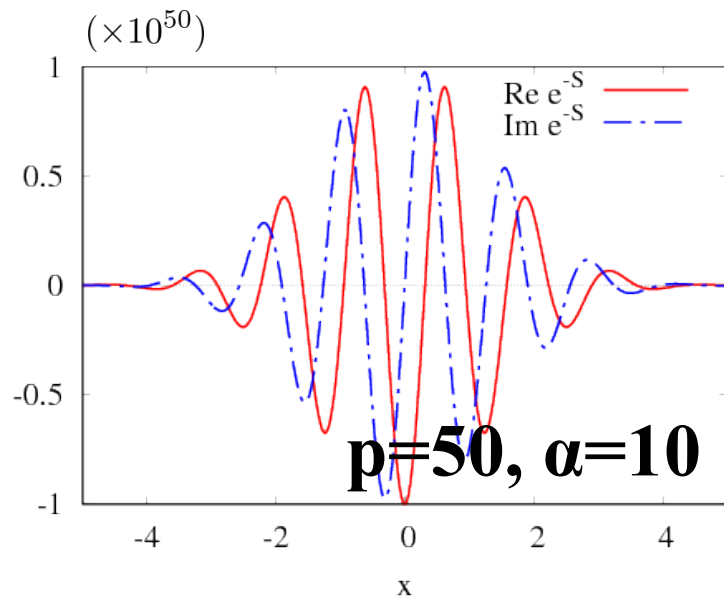
## ■ Optimized path agrees with thimble(s) around the fixed point(s) !

- Large  $\alpha \rightarrow$  One thimble, Singular point is far away from thimble
- Small  $\alpha \rightarrow$  Go through two FPs.

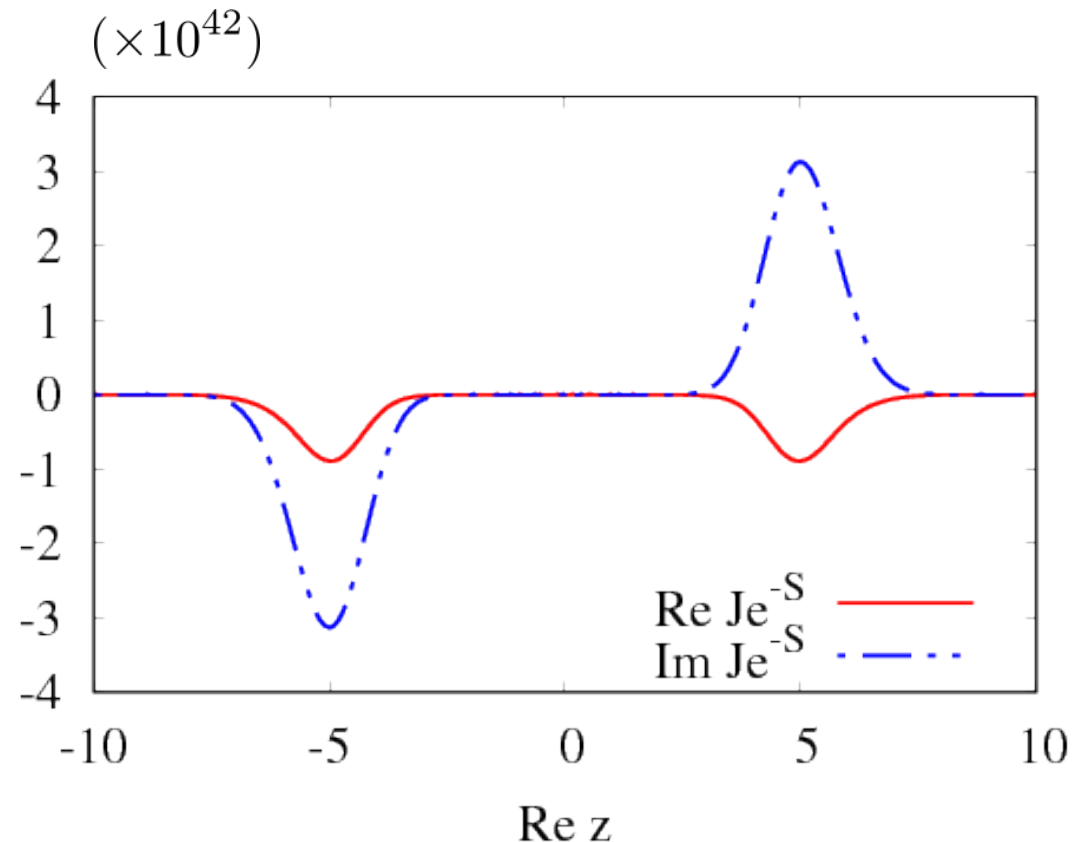


# Boltzmann Weight

- Boltzmann weight ( $Je^{-S}$ ) on the real axis  
→ Large ( $\sim 10^{50}$ ) and rapid oscillation.
- $Je^{-S}$  on the optimized path  
→ Small ( $\sim$  a few  $10^{42}$ ) and slow oscillation.



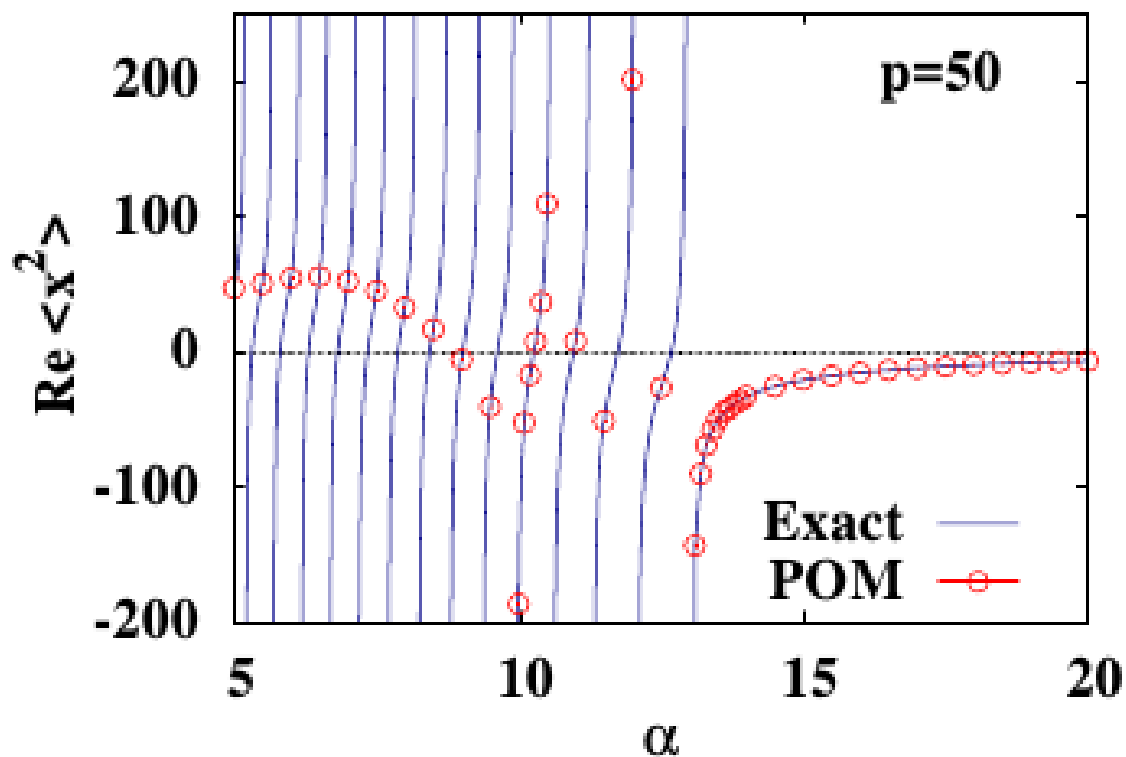
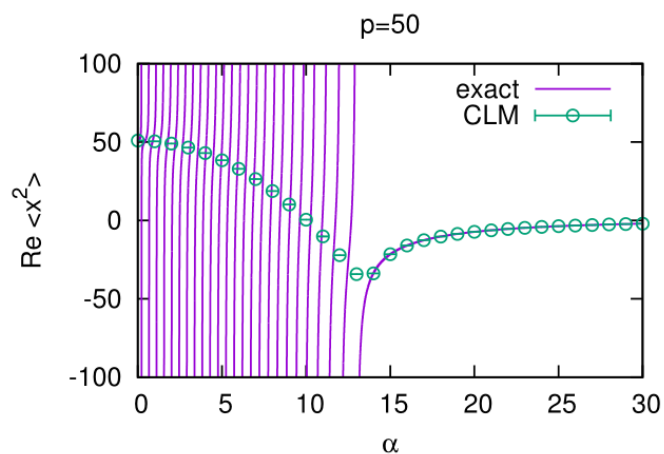
Optimization





# Expectation Value of $x^2$

- Hybrid MC results of  $\langle x^2 \rangle$  on the optimized path well reproduce the exact results.
- Trick:  $\pm x$  ( $=\pm \text{Re}(z)$ ) gives same  $|J e^{-S}|$   
 → Both  $\pm x$  configurations are taken.
- Global sign prob. is not solved (and should not be solved).



*Nishimura, Shimasaki ('15)*

*Mori, Kashiwa, AO ('17)*

---

*Application to complex  $\lambda\phi^4$  theory  
with use of 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

$$z_i(t) = t_i + \sum_{n_1, n_2, \dots} (c_{i, n_1 n_2 \dots}^{(x)} + i c_{i, n_1 n_2 \dots}^{(y)}) H_{n_1}(t_1) H_{n_2}(t_2) H_{n_3}(t_3) \cdots$$

- Neural network

- Combination of linear and non-linear transformation.

$$a_i = g(\underline{W}_{ij}^{(1)} t_j + \underline{b}_i^{(1)}) \quad \text{parameters}$$

$$f_i = g(\underline{W}_{ij}^{(2)} a_j + \underline{b}_i^{(2)})$$

$$z_i = t_i + i(\underline{\alpha}_i f_i(t) + \underline{\beta}_i)$$

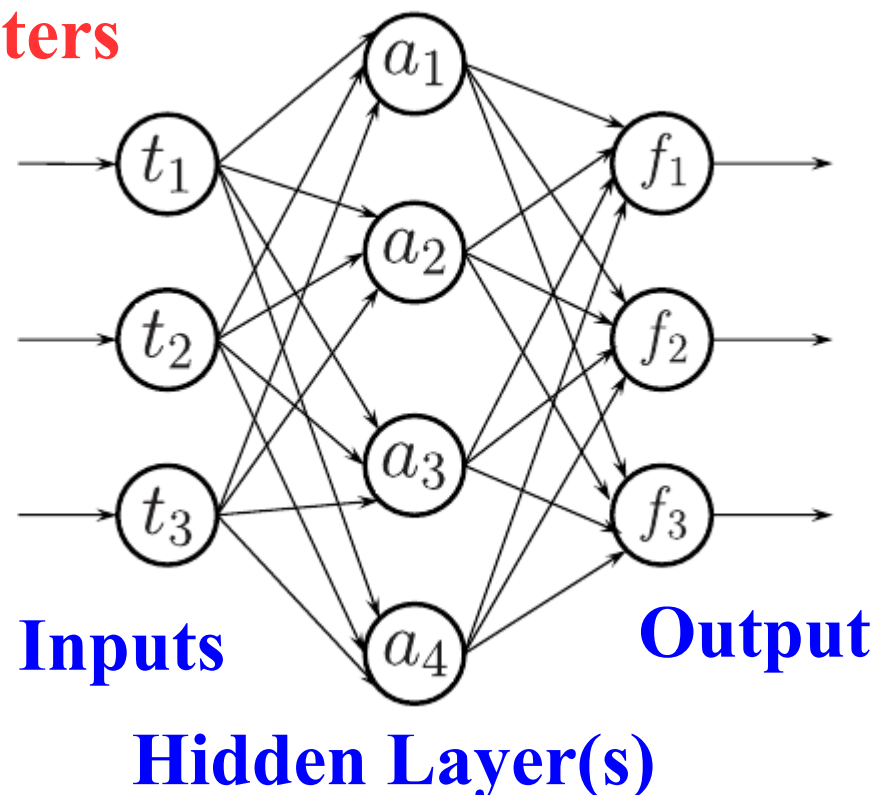
$$g(x) = \tanh x \quad (\text{activation fn.})$$

- Universal approximation theorem

Any fn. can be reproduced  
at (hidden layer unit #)  $\rightarrow \infty$

*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*

par. in (j+1)th step

Learning rate

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

mean sq. ave. of v

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

decay rate

mean sq. ave. of F

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

gradient evaluated in MC (batch training)

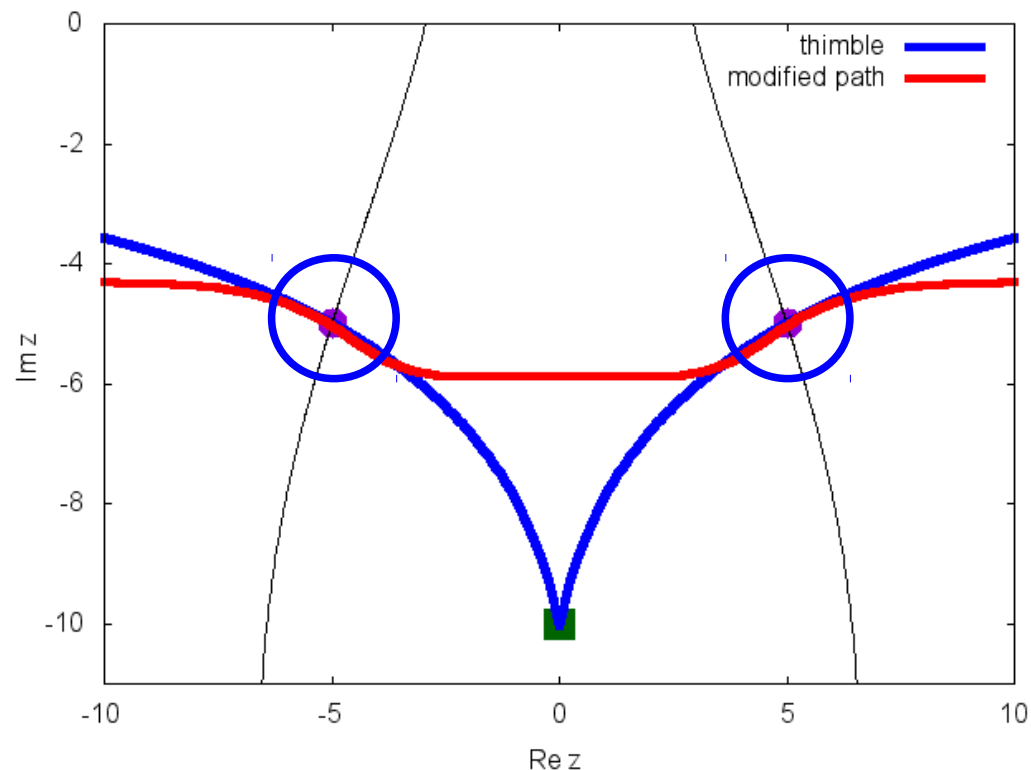
$$F_i = \partial \mathcal{F} / \partial c_i$$

Cost fn.

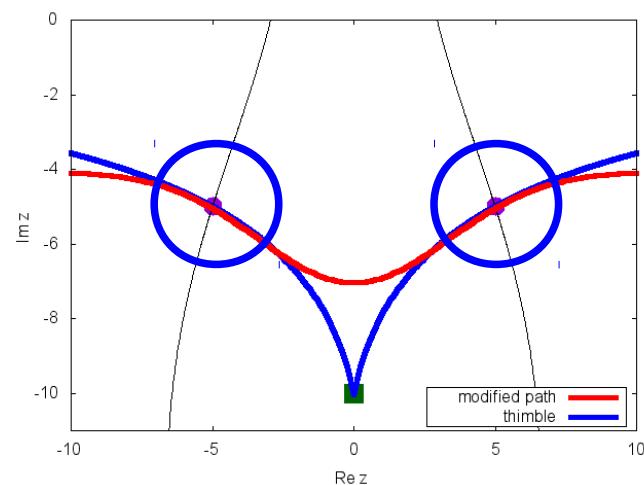
*Machine learning  
~ Educated algorithm  
to generic problems*

# 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 $\lambda\phi^4$ theory at finite $\mu$

## ■ Complex $\lambda\phi^4$ theory

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

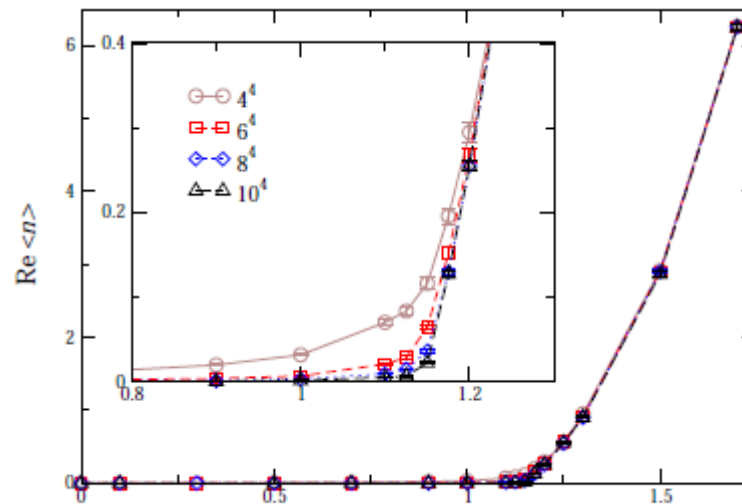
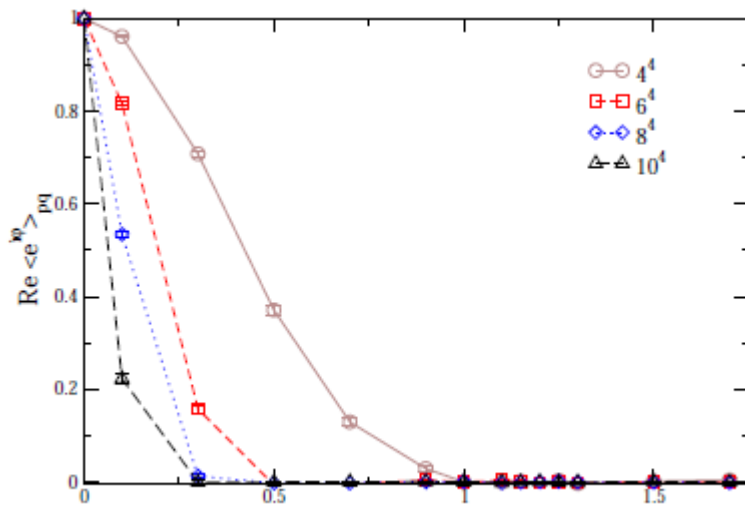
## ■ Action on Euclidean lattice at finite $\mu$ .

$$S = \sum_x \left[ \frac{(4 + m^2)}{2} \phi_{a,x} \phi_{a,x} + \frac{\lambda}{4} (\phi_{a,x} \phi_{a,x})^2 - \phi_{a,x} \phi_{a,x+\hat{1}} \right.$$

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

complex

Complexify



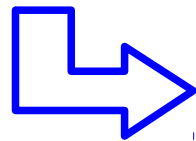
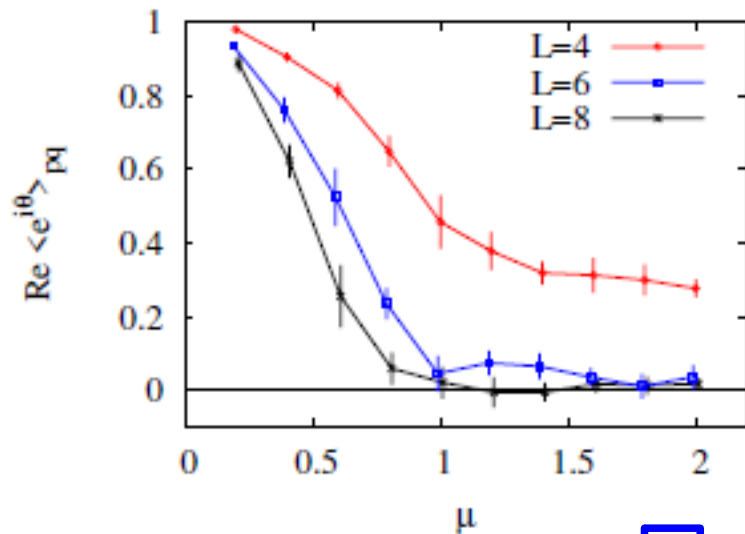
Complex  
Langevin  
& Lefschetz  
thimble  
works.

<sup>\mu</sup>G. Aarts, PRL102('09)131601; H. Fujii, et al., JHEP 1310 (2013) 147

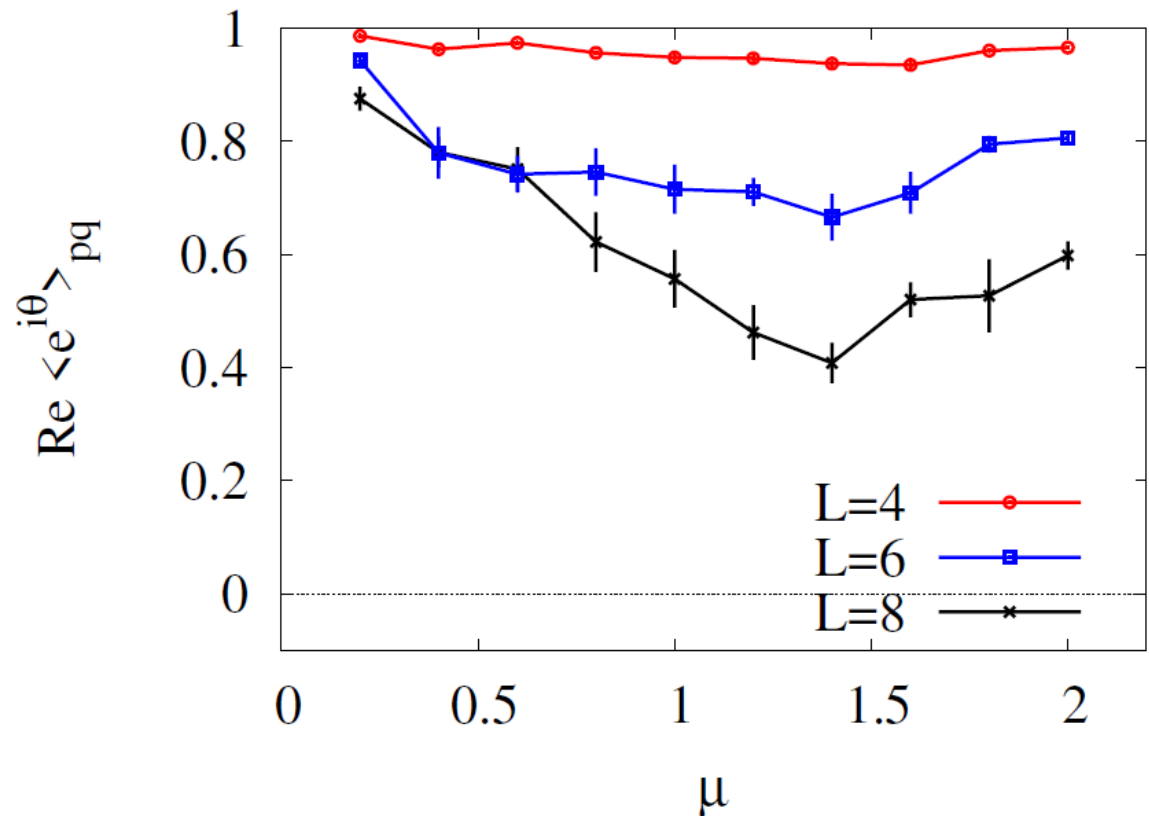
# POM result (1): Average phase factor

## ■ POM for 1+1D $\lambda\phi^4$ theory

- $4^2, 6^2, 8^2$  lattices,  $\lambda=m=1$
- $\mu_c \sim 0.96$  in the mean field approximation
- Enhancement of the average phase factor after optimization.



Optimization



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

*Ohnishi @ NFQCD 2018, June 5, 2018*

23 / 31

# POM result (2): Density

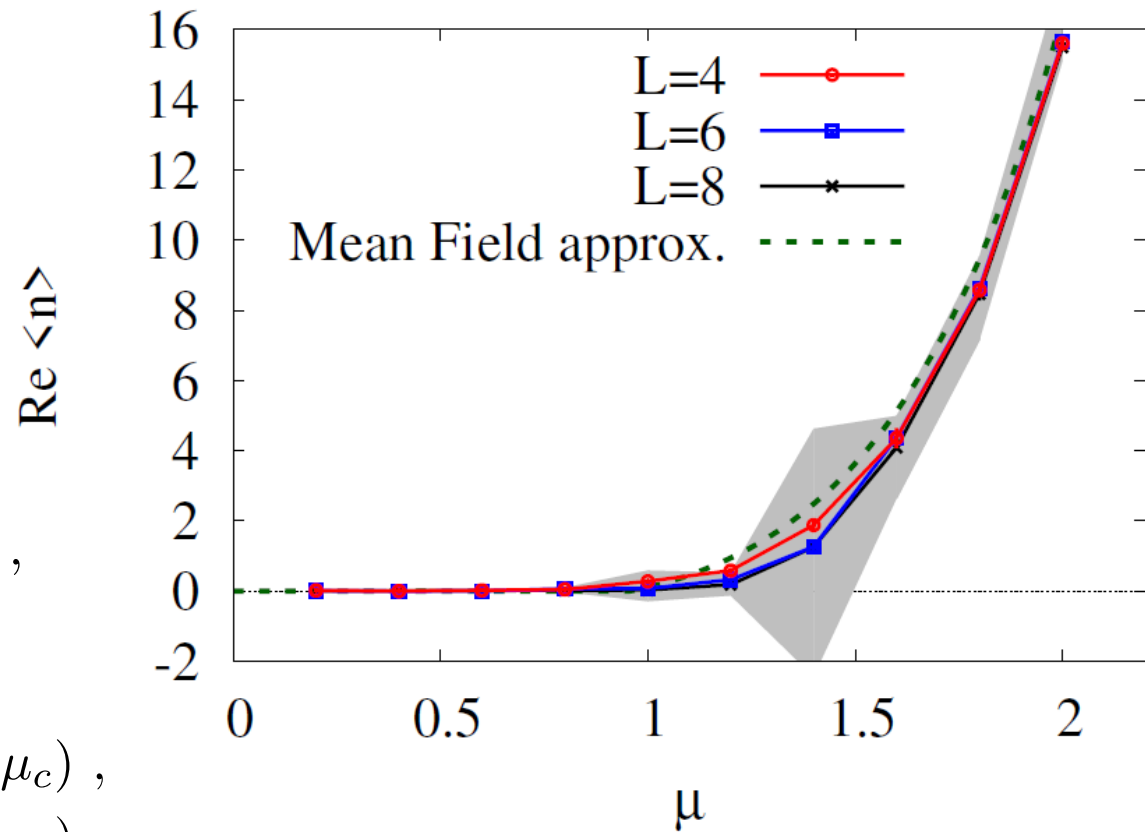
- Results on the real axis  
Small average phase factor, Large errors of density
- On the optimized path  
Finite average phase factor, Small errors

## Mean Field App.

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

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

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

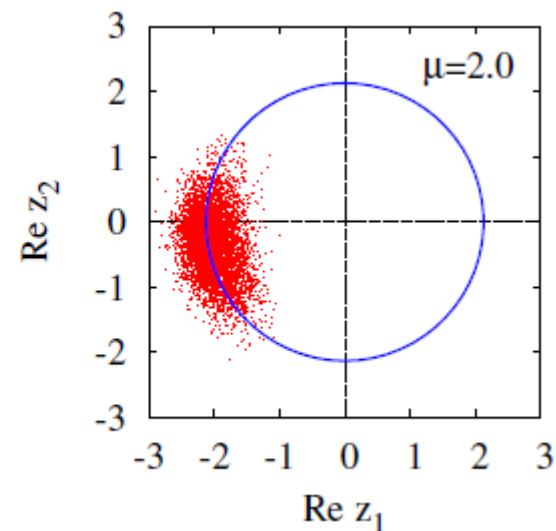
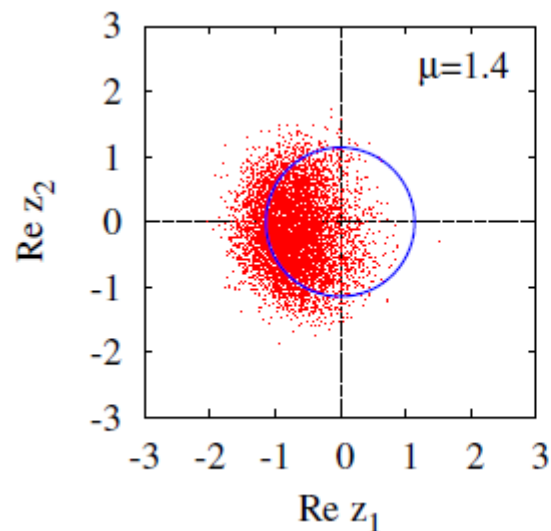
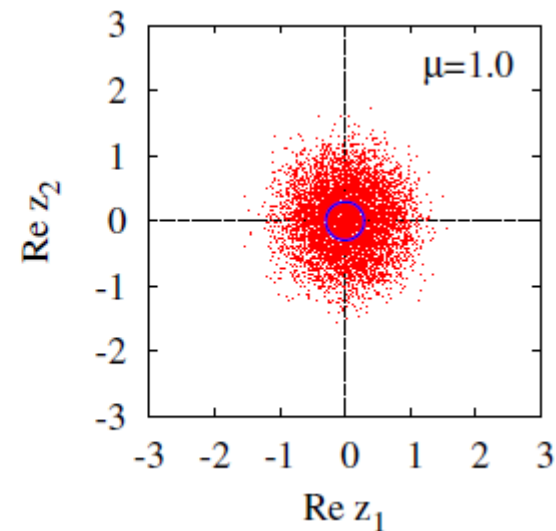
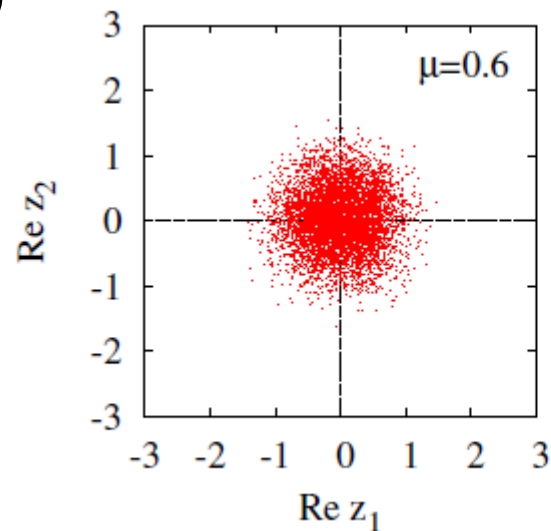


Mori, Kashiwa, AO ('18)



# POM result (3): Configurations

- Updated configurations after optimization  
→ sampled around the mean field results
- U(1) symmetry in  $(\varphi_1, \varphi_2)$   
is broken spontaneously  
by the optimization ?  
(or by the sampling)



---

# *Prospects of path optimization*

# Problems in Path Optimization

## ■ Merits

- Optimization of the path including Jacobian
- Singular points of the “Action” do not matter in many cases(\*).  
(\* Singularity coming from zeros of Fermion determinant does not harm path optimization.)

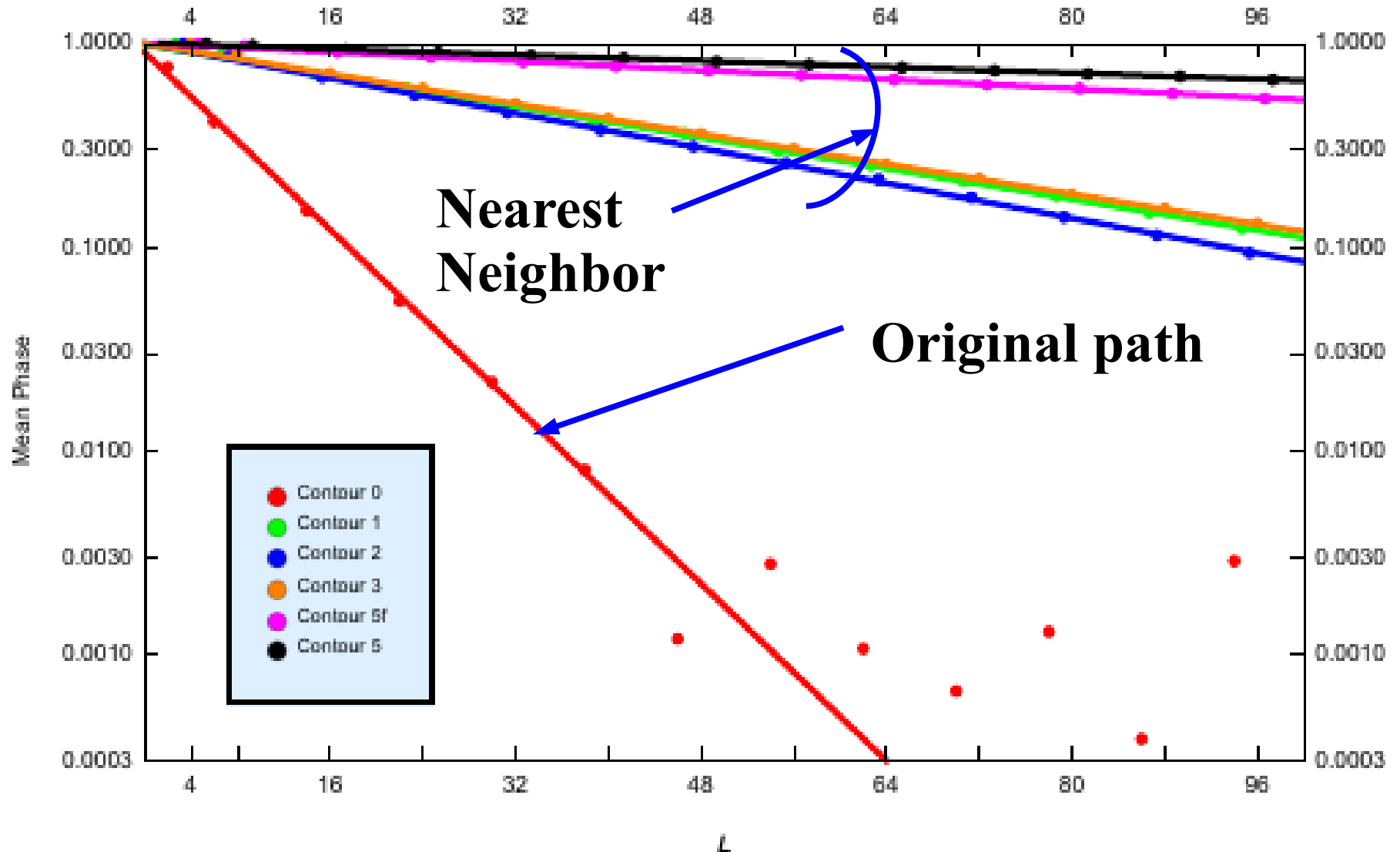
## ■ Problem

- Optimization is not trivial even with neural network.
- Large number of parameters to be optimized,  $O((N_{\text{dof}})^2)$ .
- Huge CPU times would be necessary,  $O((N_{\text{dof}})^3)$  for Jacobians.
  - Reduced pars. assuming nearest neighbor correlation  
*F. Bursa, M. Kroyter, arXiv:1805.04941*
  - Jacobian only from the same site ? (We doubt it !)  
*A. Alexandru et al., PRD96 ('17), 094505 [arXiv:1709.01971];*  
*A. Alexandru et al., PRD97('18), 094510 [arXiv:1804.00697].*

# Optimization with Nearest Neighbor Site Correlation

- Economical optimization with sparse Jacobian matrix.

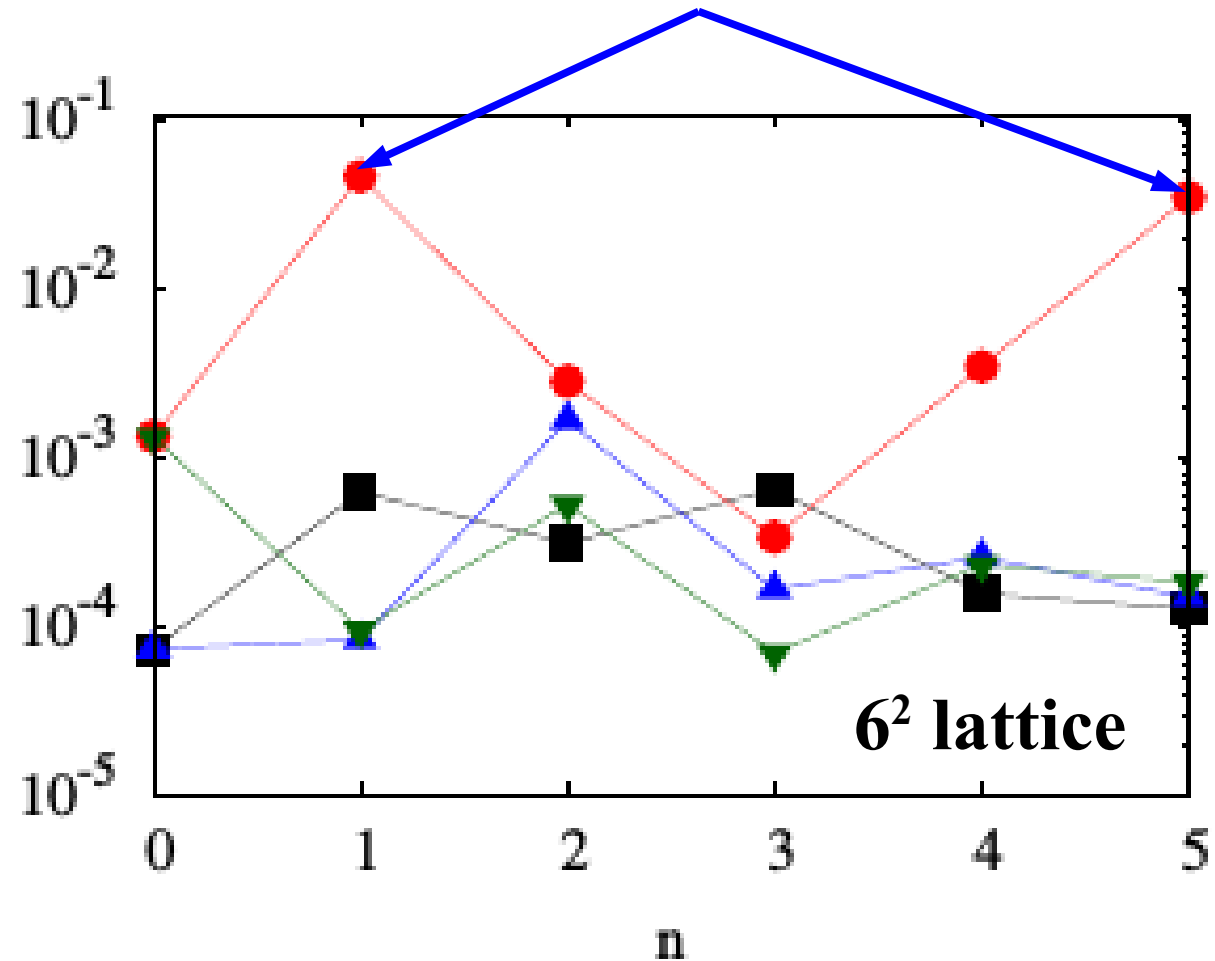
*F. Bursa, M. Kroyter, arXiv:1805.04941*



# Nearest Neighbor Site Correlation Dominates ?

- Yes, in complex  $\lambda\phi^4$ .  
Other correlations  
 $\sim 10^{-1}$ - $10^{-2}$  times smaller
- Hope to reduce  
the cost to be  $O(N_{\text{dof}})$

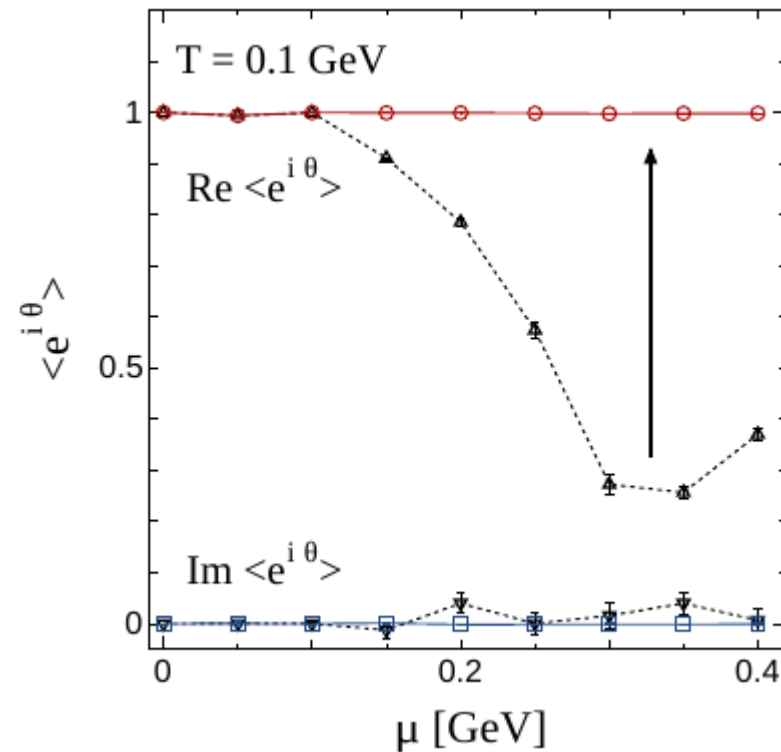
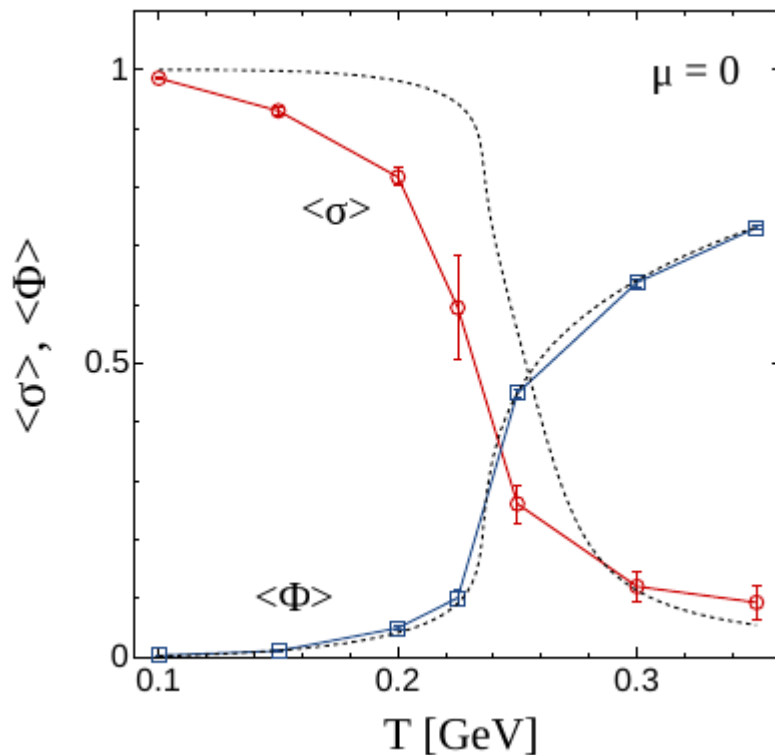
$(\phi_1, \phi_2)$  nearest temporal  
neighbor site corr.



*Y. Mori, Master thesis*

# Application to PNJL

- PNJL model with homogeneous condensates,  $(\sigma, \pi, \Phi, \bar{\Phi})$ .
  - 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

---

- **Path Optimization Method is proposed to attack the sign problem.**
  - Path is parametrized by the **Trial Function**.
  - Seriousness of the sign problem is given by the **Cost Function**.
  - Sign problem is regarded as the **Optimization Problem**.
- **Usefulness of POM is demonstrated in several models.**
  - Optimized path reproduces the thimble(s) around the fixed point(s).
  - Many of singular points of the action do not matter, as long as they are not **the singular points of the Boltzmann weight**.
  - Global sign problem is unsolved (and should not be solved).
- **POM is applicable to field theories with use of neural network.**
- **Simplification based on sparse Jacobian matrix seems to be promising.**

# Application to Gauge Theories

## ■ 2D SU(2) Yang-Mills theory with complex coupling

*K. Nagata, J. Nishimura and S. Shimasaki, arXiv:1802.01876 [hep-lat],*

*H. Makino, H. Suzuki and D. Takeda, PRD92 ('15)085020.*

→ Struggling ....

