Path optimization for the sign problem in field theories using neural network

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.

8th Int. Conf. on Quarks and Nuclear Physics
Nov. 13-17, 2018, Tsukuba, Japan

Sign Problem

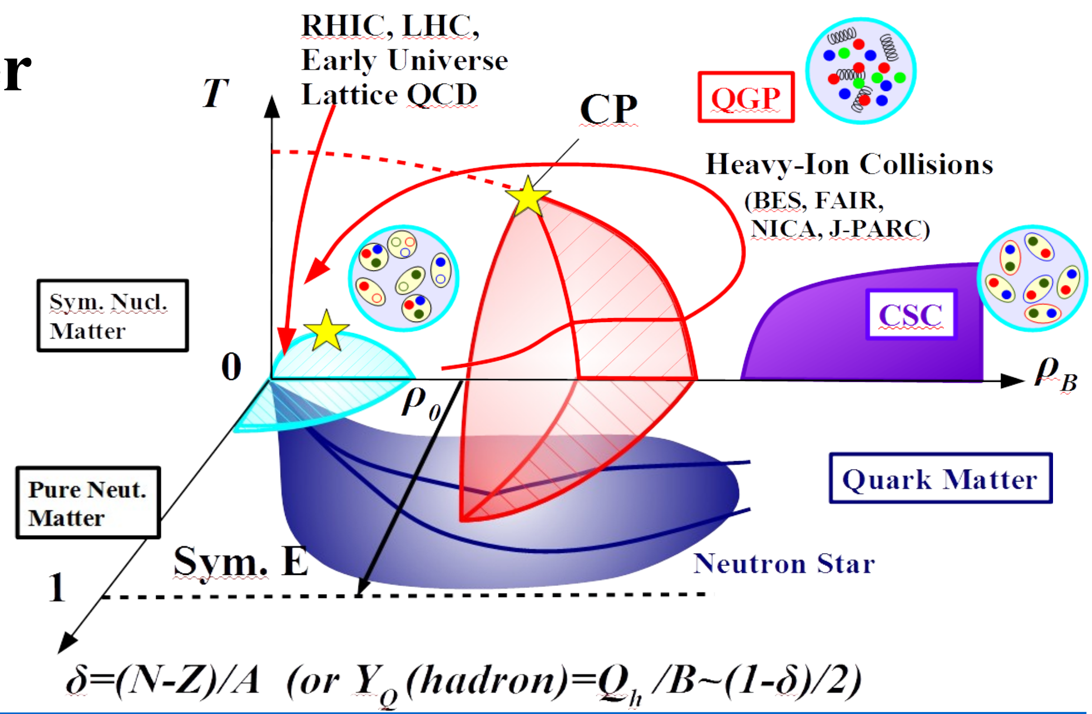
- **Fermion det. is complex at finite density**
  → Strong cancellation of the Boltzmann weight at large volume.

$$\det D(\mu) = (\det D(-\mu^*))^* \to S_{\rm eff} = S - \log\det D \in \mathbb{C}$$

$$\mathcal{Z} = \int \mathcal{D}x\, \exp(-S(x)),\ \mathcal{Z}_{\rm pq} = \int \mathcal{D}x\, |\exp(-S(x))|$$

$$\mathrm{APF} = \langle e^{i\theta} \rangle = \mathcal{Z}/\mathcal{Z}_{\rm pq} \to 0 \quad (V \to \infty)$$

- **Difficult to study dense matter using LQCD**



Sign Problem

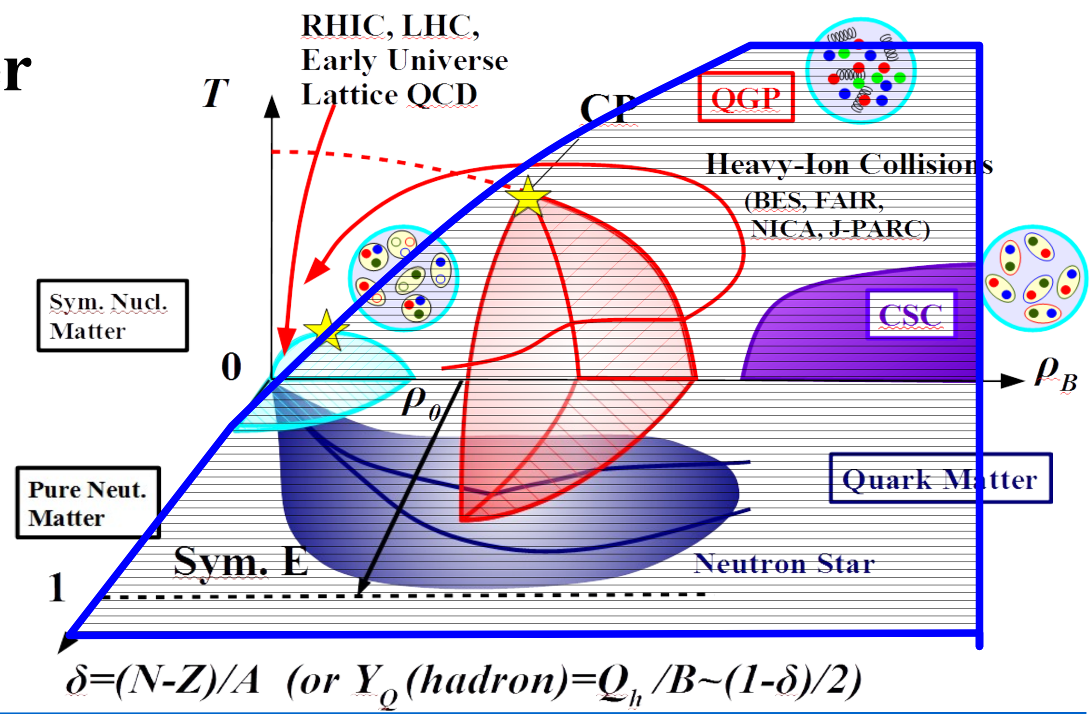
- **Fermion det. is complex at finite density**
  → Strong cancellation of the Boltzmann weight at large volume.

$$\det D(\mu) = (\det D(-\mu^*))^* \to S_{\rm eff} = S - \log\det D \in \mathbb{C}$$

$$\mathcal{Z} = \int \mathcal{D}x \, \exp(-S(x)), \;\; \mathcal{Z}_{\rm pq} = \int \mathcal{D}x \, |\exp(-S(x))|$$

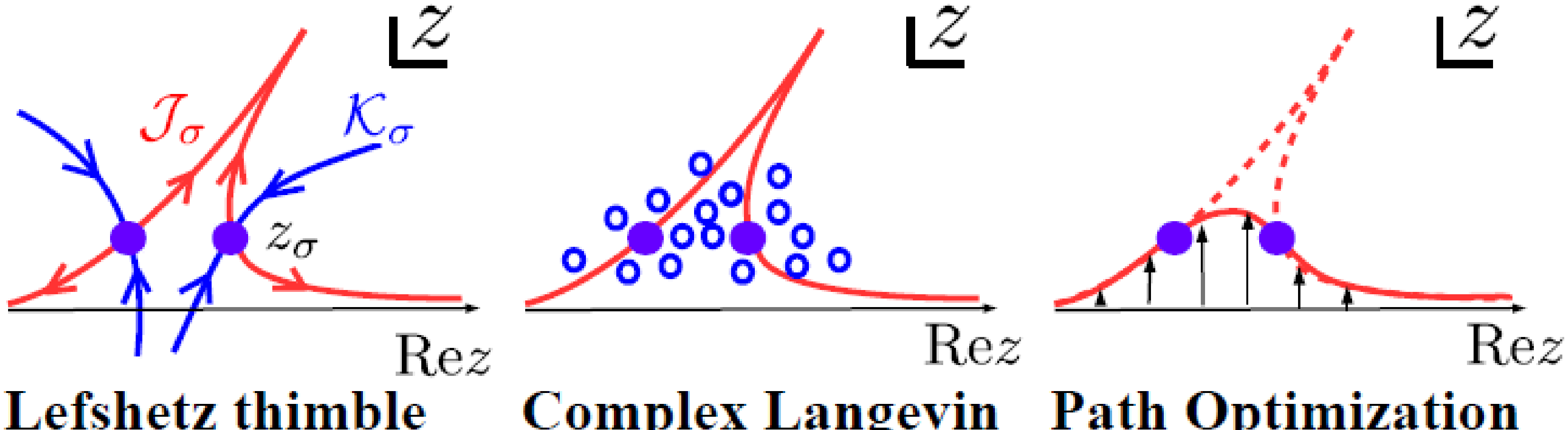
$$\mathrm{APF} = \langle e^{i\theta} \rangle = \mathcal{Z}/\mathcal{Z}_{\rm pq} \to 0 \;\; (V \to \infty)$$

- **Difficult to study dense matter using LQCD**
- **Standard approaches (Taylor exp., Imag. μ, ..)**
  → Useful, but not enough
  to discuss dense matter



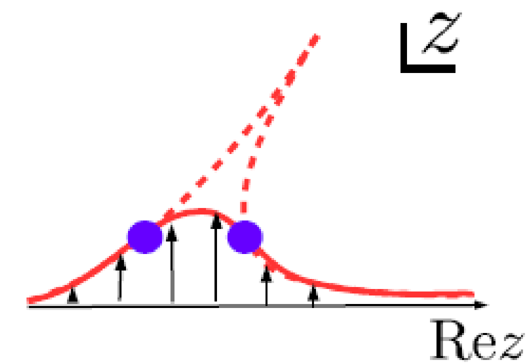
Complexified Variable Methods

- **Lefschetz thimble method**
  E. Witten ('10), Cristoforetti+('12),Fujii+('13), Alexandru+('16).
  - **Flow eq. → Im(S) is constant on thimbles**
  - **Phase from the Jacobian, Diff. phase from diff. thimbles (residual / global sign pr.),**
- **Complex Langevin method (→ Tsutsui's talk)**
  Parisi-Wu('81), Klauder('83), Aarts+('11), Nagata+('16), Seiler+('13), Ito+('16).
  - **Complex Langevin eq. → Expectation value = Ensemble ave.**
  - **Occasional conversion to wrong answers**



Path optimization method

Mori et al. ('17), AO, Mori, Kashiwa (Lattice 2017),
Mori et al. ('18), Kashiwa et al. ('18);
Alexandru et al. ('18 (SOMMe), '18),Bursa, Kroyter ('18)



- **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.**
    (det D=0 → Singular point of Seff, Zero point of exp(-Seff))
- **Integration path is optimized to evade the sign problem.**
  Cost function:

$$\mathcal{F}[z(x)] = \mathcal{Z}_{\mathrm{pq}} - |\mathcal{Z}| = |\mathcal{Z}| \left(\mathrm{APF}^{-1} - 1 \right)$$

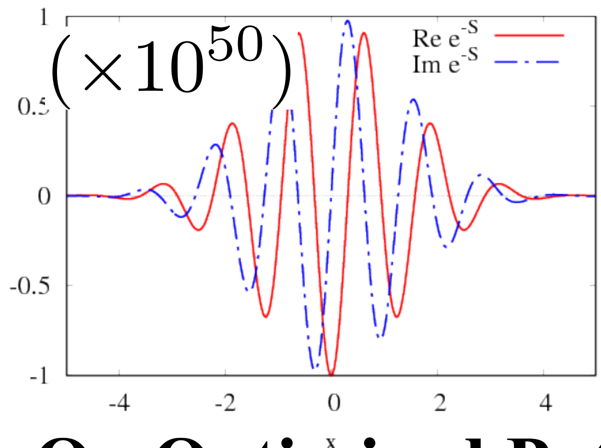
- **Optimization can be performed in various ways.**
  Gradient descent, Stochastic Gradient Descent (SDG),
  Neural network, ….

Sign Problem → Optimization Problem

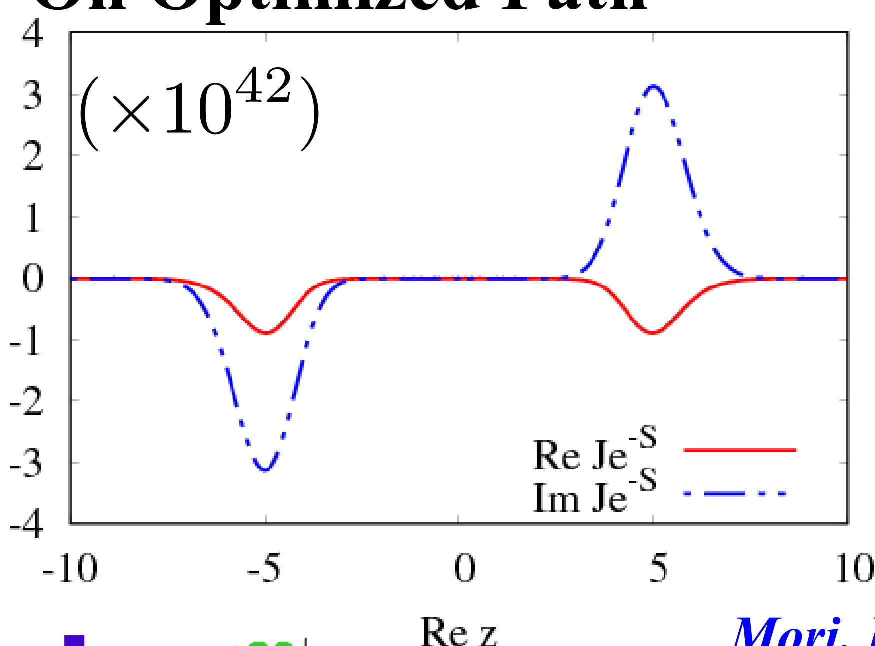
Benchmark test: 1 dim. integral (gradient descent)

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

On Real Axis

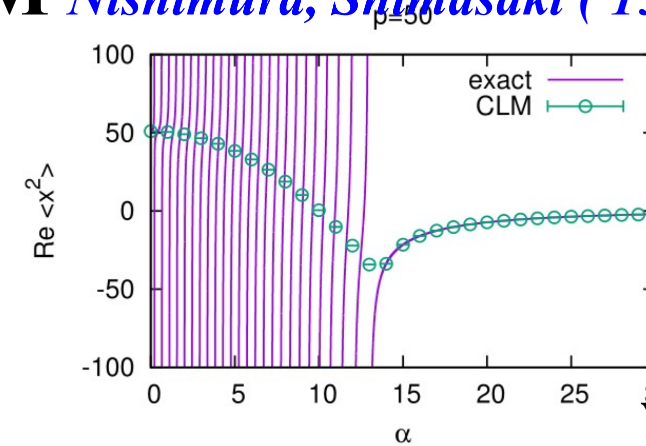


On Optimized Path

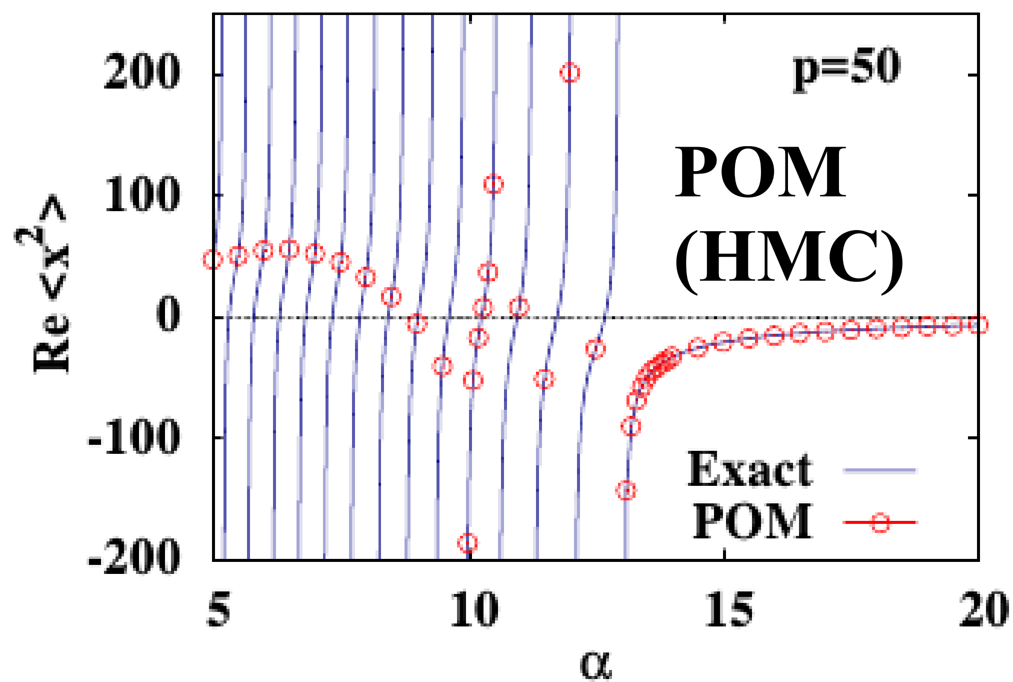


- **Observable**

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 !

Contents

- Introduction
  - Sign problem & Complexified variable methods
  - 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 field theories using neural network
  - Complex φ^4 theory (application to field theory)
    Y. Mori, K. Kashiwa, AO, PTEP 2018 ('18), 023B04 [arXiv:1709.03208]
  - 0+1-dimensional QCD (application to gauge theory)
    AO, Mori, Kashiwa, Lat2018 proc.; Y. Mori, K. Kashiwa, AO, in prep.
  - PNJL model (application to field theory with p.t.)
    K. Kashiwa, Y. Mori, AO, arXiv:1805.08940
- Summary

Path Optimization Method in field theories using neural network (1) Complex φ^4 theory at finite μ

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

Application of POM to Field Theories

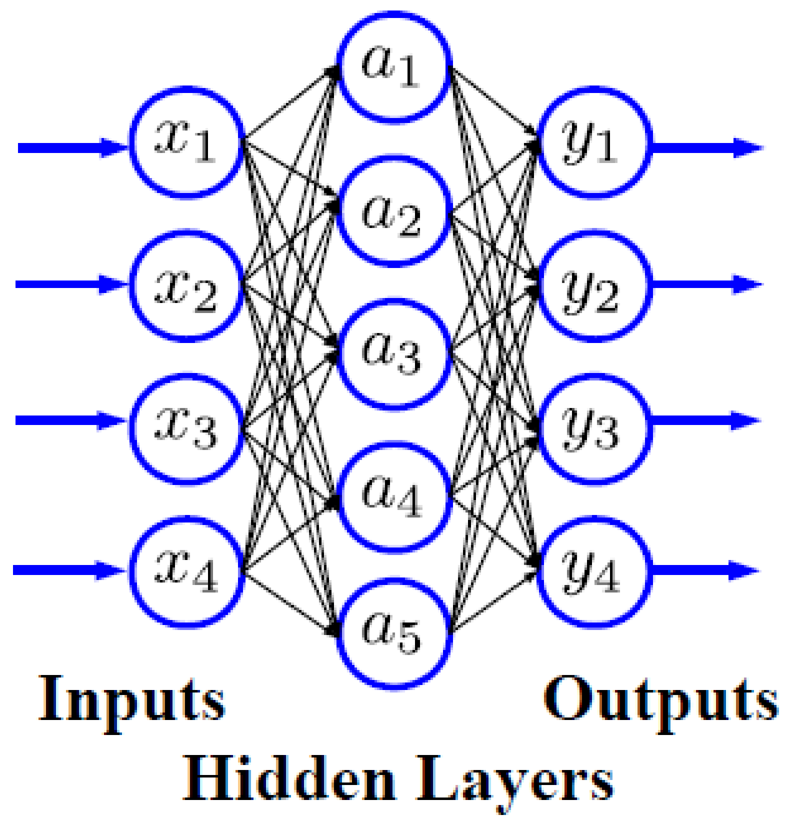
- **Preparation & variation of trial fn. with 1000 variables by hand → Pracitcally impossible**
- **Neural network**
  - **Combination of linear and non-linear transformation.**

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

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

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

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



  - **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

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

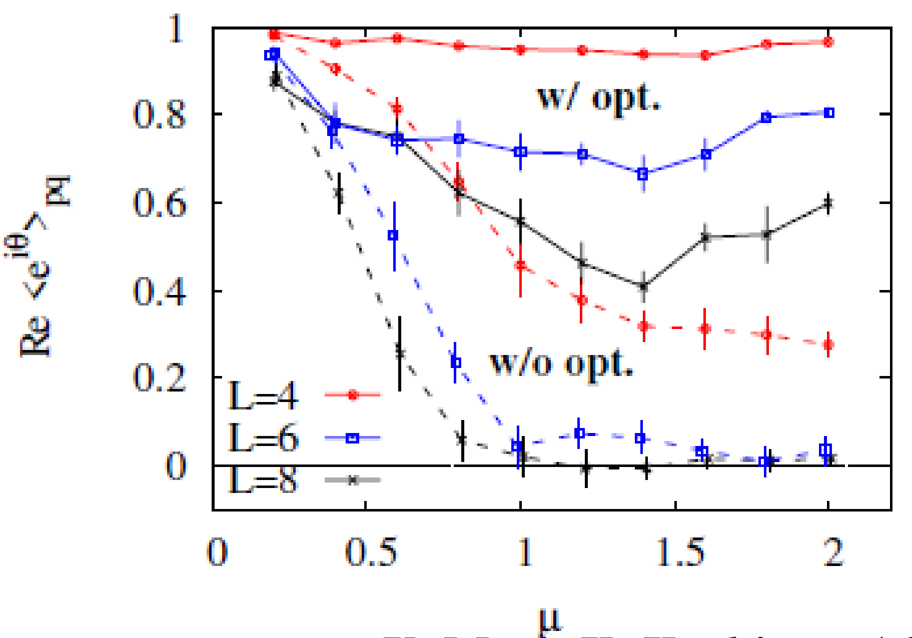
1+1 dim. Complex φ⁴ theory at finite μ

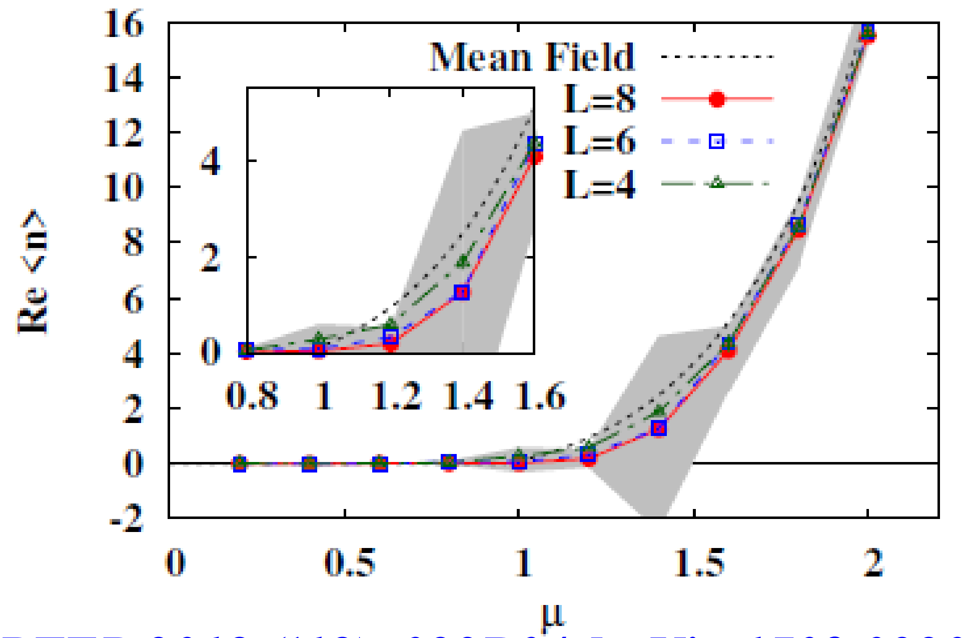
- **Complex φ⁴ theory** $\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - m^2 \phi^* \phi - \lambda(\phi^* \phi)^2$
- **Action on Eucledean lattice at finite μ.**
  G. Aarts, PRL102('09)131601; H. Fujii, et al., JHEP 1310 (2013) 147.

$$S = \sum_x \left[\frac{(4+m^2)}{2} \phi_{a,x}\phi_{a,x} + \frac{\lambda}{4}(\phi_{a,x}\phi_{a,x})^2 - \phi_{a,x}\phi_{a,x+\hat{1}} - \cosh\mu\, \phi_{a,x}\phi_{a,x+\hat{0}} \right.$$

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

complex   Complexify





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

Path Optimization Method in field theories using neural network (2) 0+1 dimensional QCD (Application to Gauge Theory)

AO, Mori, Kashiwa, Lat2018 proc.; Y. Mori, K. Kashiwa, AO, in prep.

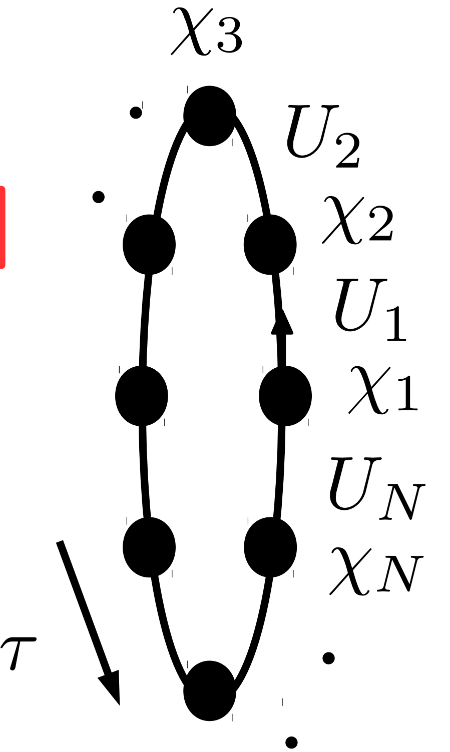
- **0+1 dimensional QCD (1 dim. QCD) with one species of staggered fermion on a 1xN lattice**

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

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

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

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.**
- **Reduced to be a one-link problem.**
  → Analytic results are known.
- **Studied well in the context of strong coupling LQCD**
  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)**

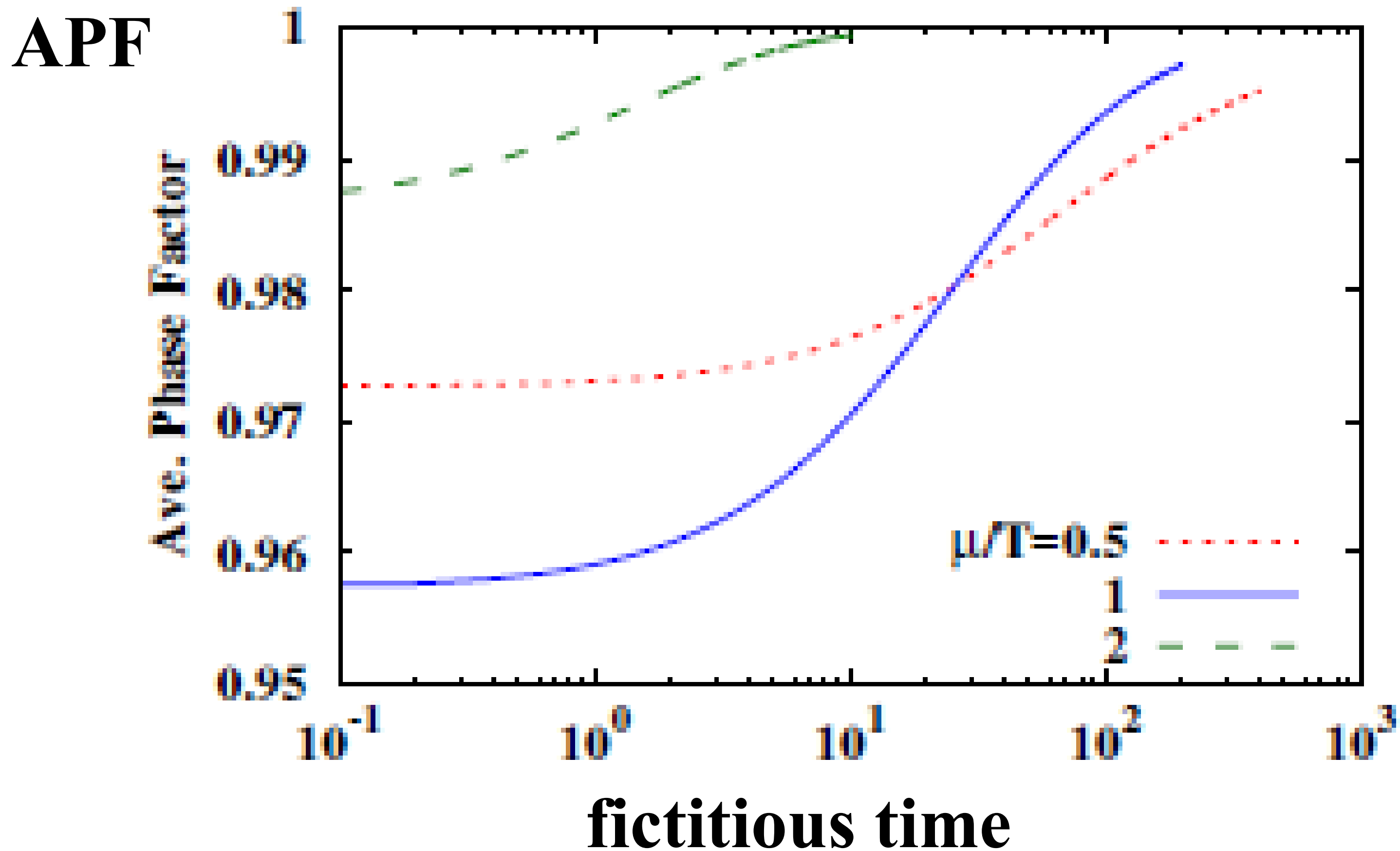
- **Path optimization (t: ficticious time)**
  → y(x_1,x_2) itself is the parameter on the (x_1,x_2) mesh point

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

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

Path Opt. of 0+1 dim. QCD in diagonal temporal gauge

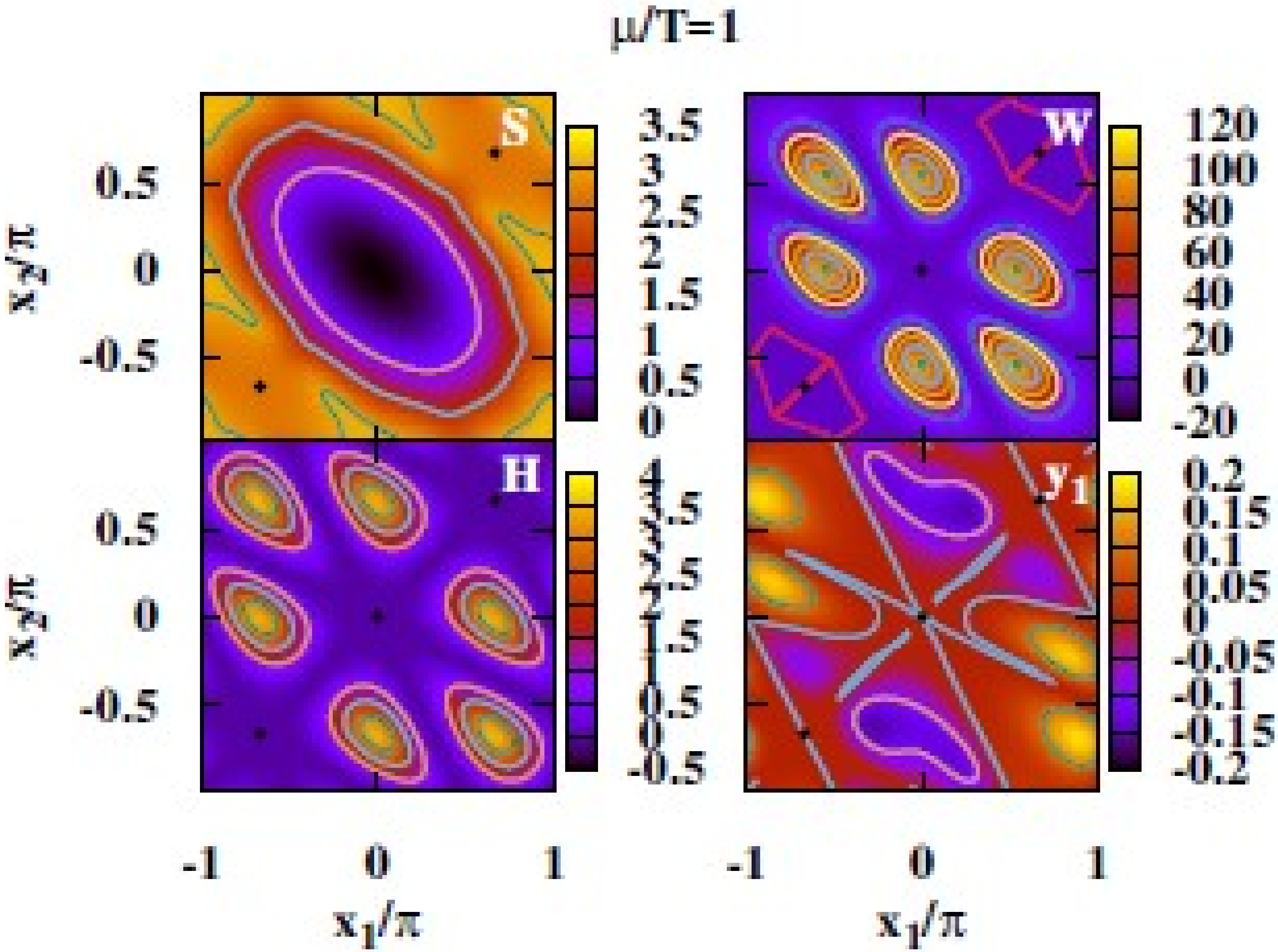
- **Path optimization → APF > 0.99 → Easily achieved**
  - **3+1 dim. QCD (L^3(=V) × Nt lattice) → $APF_{3+1} \approx (APF_{0+1})^V$**
    APF_{0+1}=0.95 → APF_{3+1} =4×10^{-12}, APF_{0+1} = 0.995 → APF_{3+1} = 0.08 (8^3×Nt lattice)



AO, Mori, Kashiwa, Lat2018 proc.; Y. Mori, K. Kashiwa, AO, in prep.

Path Opt. of 0+1 dim. QCD

- **exp(-S) and Haar Mesure→ Six separated regions** *Schmidt+('16, LTM)*
  - **Problematic in MC simulations to overcome Statistical pot. barrier**



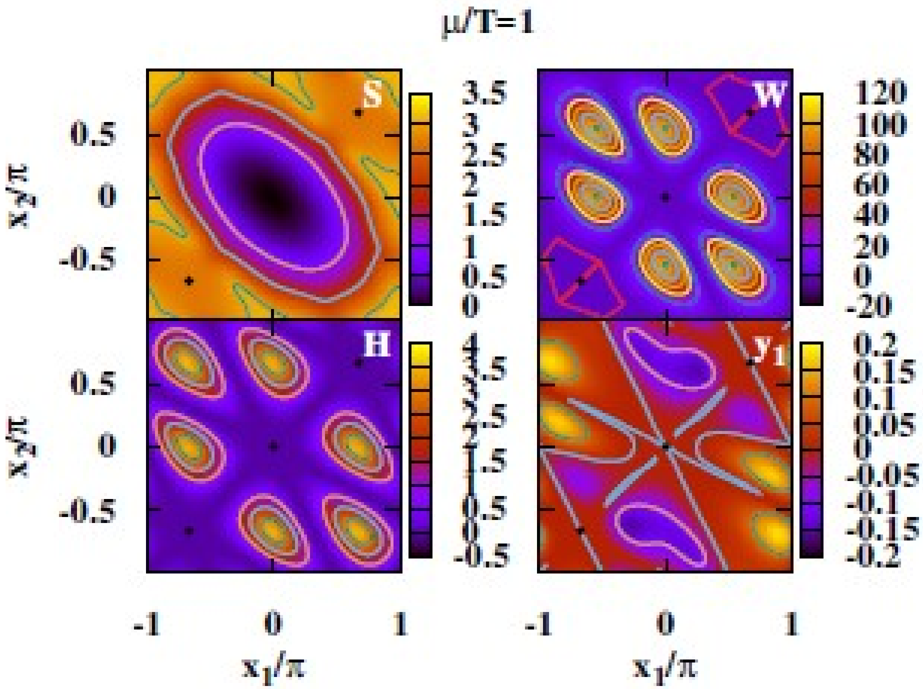
AO, Mori, Kashiwa, Lat2018 proc.; Y. Mori, K. Kashiwa, AO, in prep.

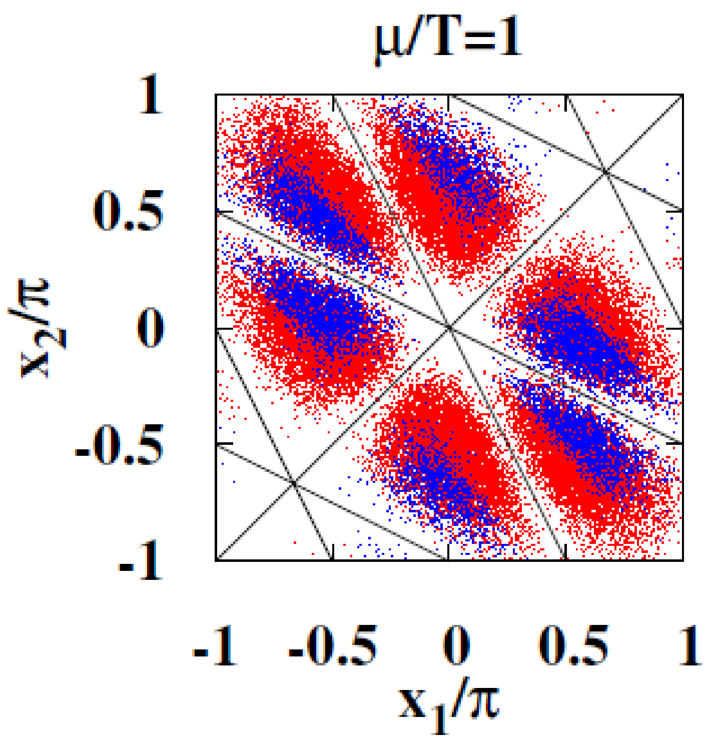
Path Opt. of 0+1 dim. QCD

- **exp(-S) and Haar Mesure→ Six separated regions** *Schmidt+('16, LTM)*
  - **Problematic in MC simulations to overcome Statistical pot. barrier**
- **Hybrid Monte-Carlo in 1 dim. QCD w/o gauge fixing using NN**

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

SL(3)

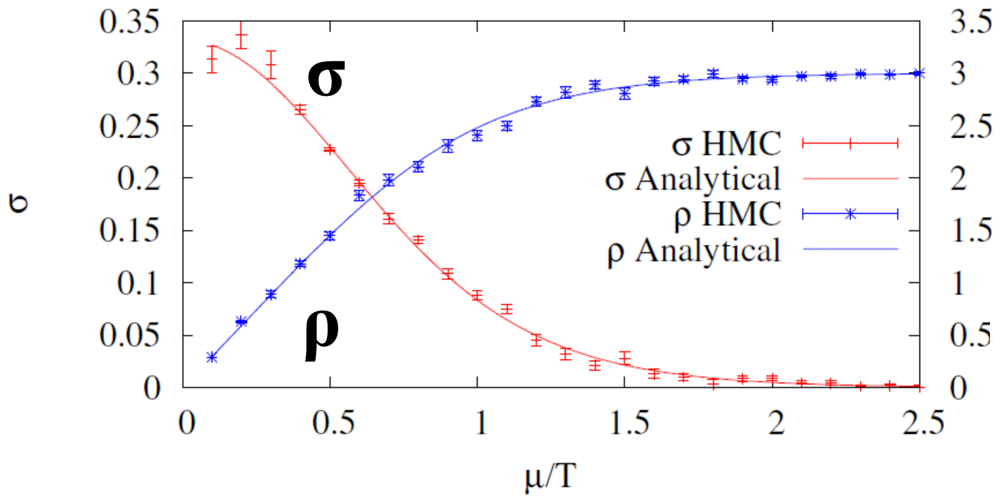




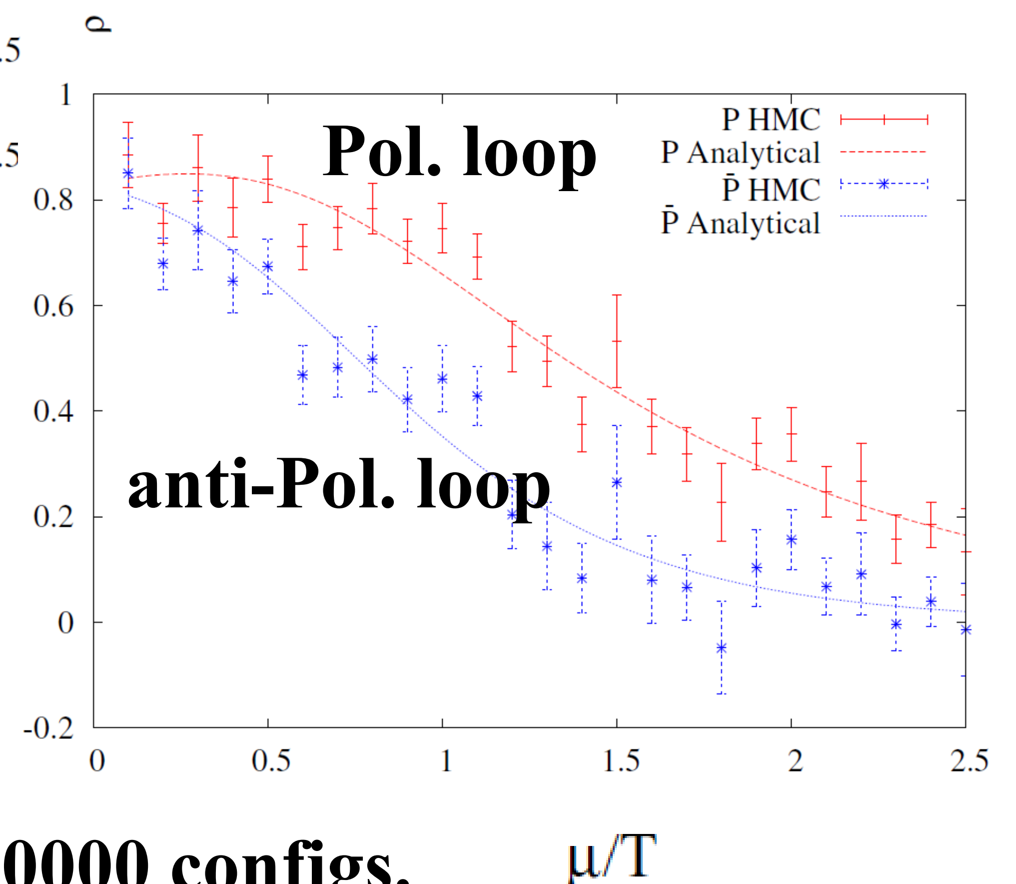
AO, Mori, Kashiwa, Lat2018 proc.; Y. Mori, K. Kashiwa, AO, in prep.

Path Opt. of 0+1 dim. QCD

- **Hybrid Monte-Carlo in 1 dim. QCD w/o gauge fixing using NN → reproduces exact results, as expected.**



1000 configs.



10000 configs.

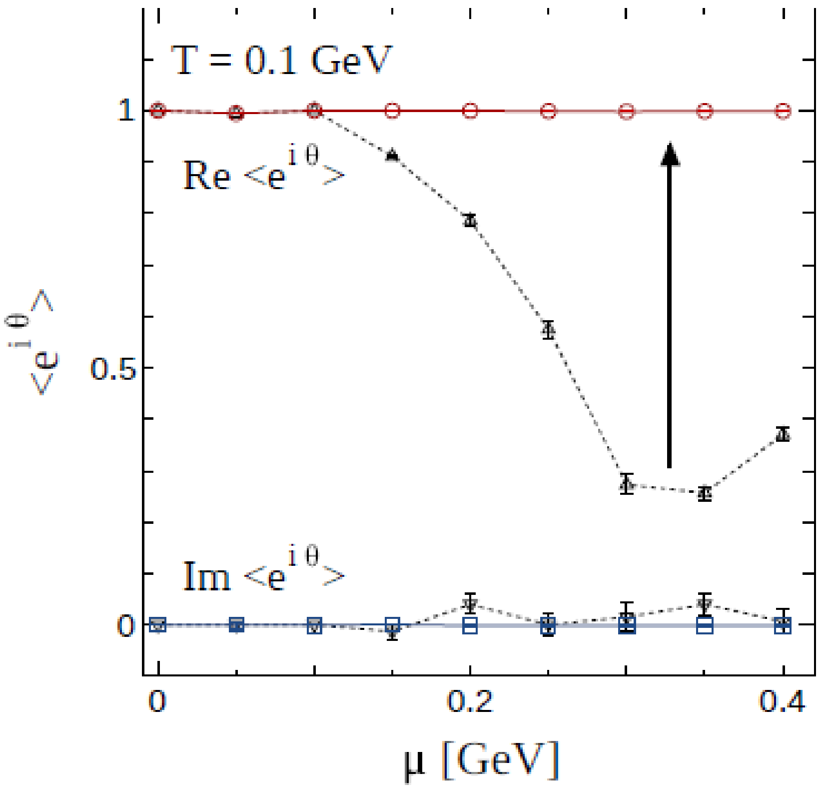
Y. Mori, K. Kashiwa, AO, in prep.

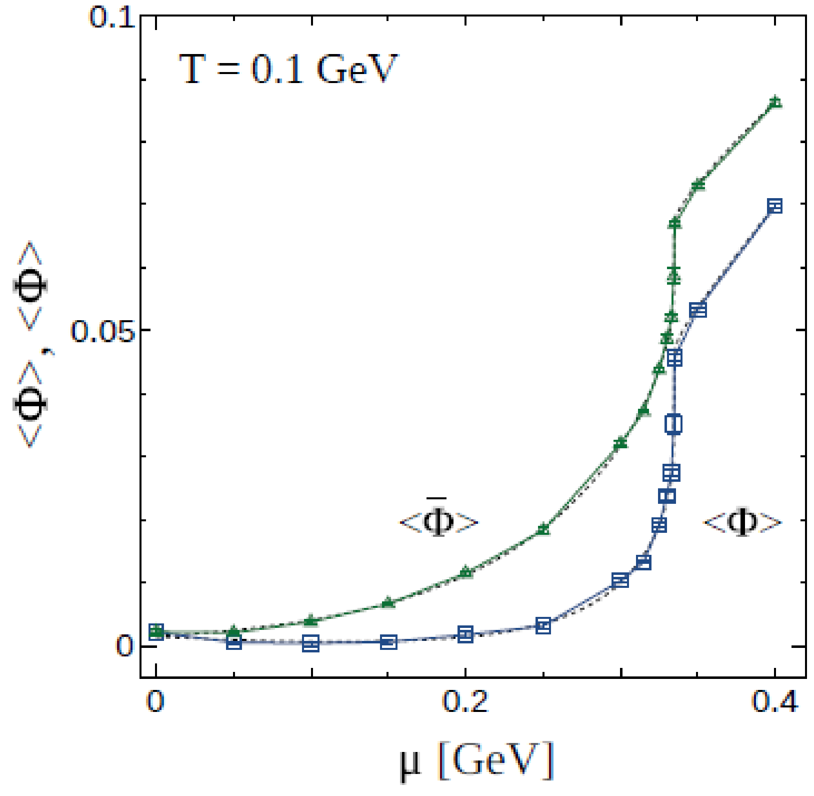
Path Optimization Method in field theories using neural network (3) PNJL model (Application to Field Theory w/ p.t.)

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

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

- **Path optimization** with the use of the **neural network** is demonstrated to work in **field theories with the sign problem** having many variables.
  - 1+1D φ^4 theory at finite μ (neural network)
  - 0+1D QCD w/ fermions (2D mesh, neural network)
  - 3+1D homogeneous PNJL (neural network)
- **Neural network (single hidden layer) is the basic device of machine learning, and it helps us to generate and optimize generic multi-variable functions, $y_i=y_i(\{x\})$.**
- **It would be possible to reduce the numerical cost and to apply POM to 3+1 dim. QCD by using the simplified ansatz.**
  E.g. Alexandru, Bedaque, Lamm, Lawrence, PRD97('18)094510,
  F. Bursa, M. Kroyter, arXiv:1805.04941

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.)

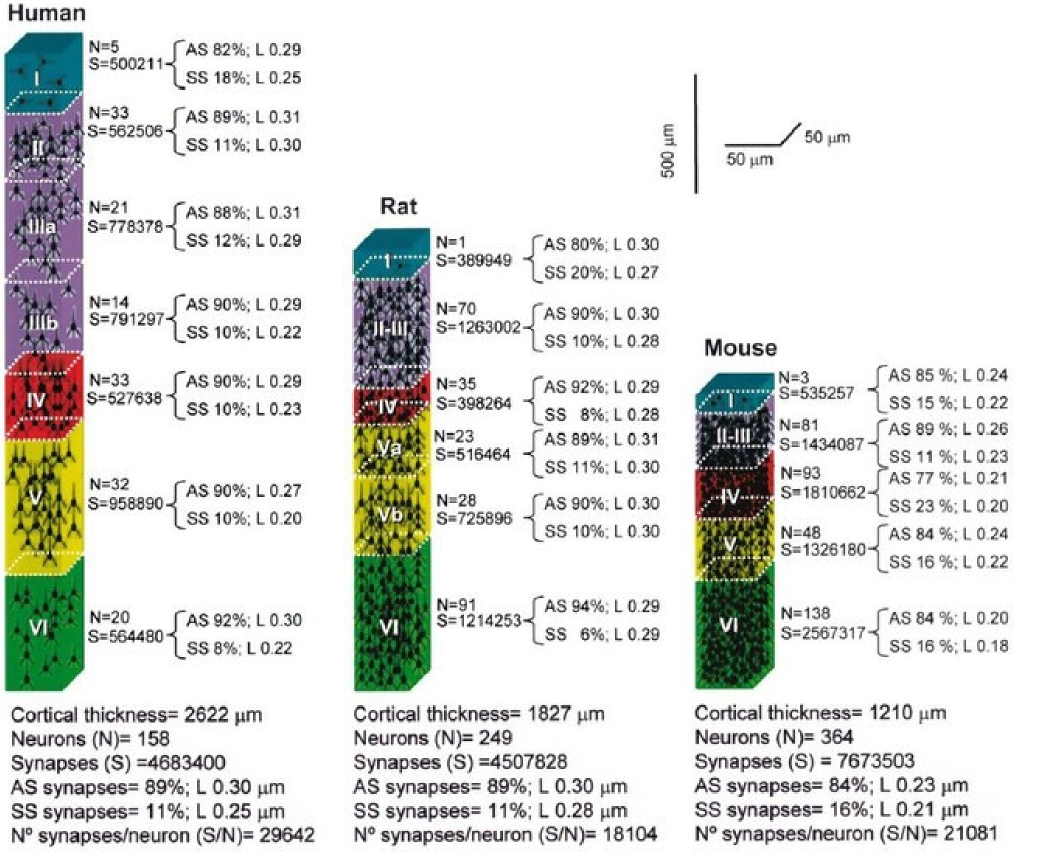
K. Kashiwa

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: AO, Mori, Kashiwa, Lat2018 proc.; Y. Mori, K. Kashiwa, AO, in prep.
PNJL with vector int. using NN: K. Kashiwa, Y. Mori, AO, in prep.

Prospect

- **Deep learning (# of hidden layers > 3) may be helpful to explore complex paths, which human beings (~ 7 layers) cannot imagine, while "Understanding" the results of machine learning need to be done by human beings (at present).**



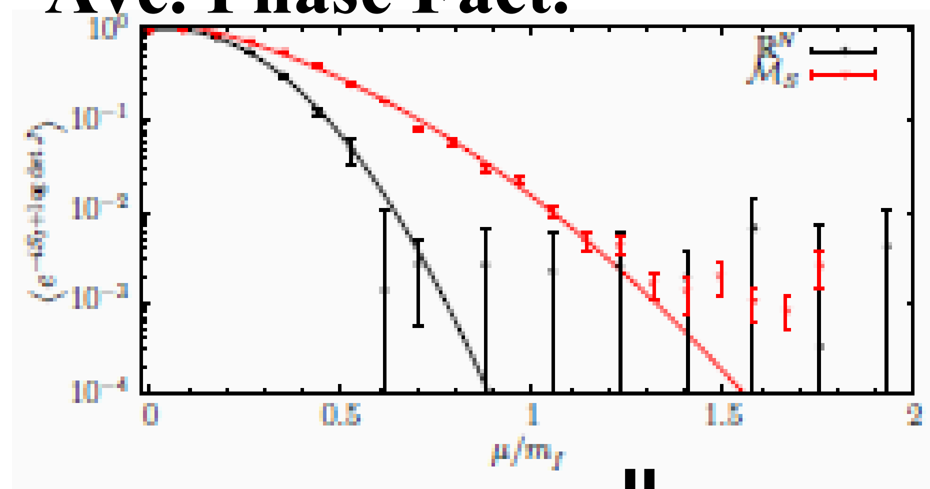
Defelipe, Front Neuroanat 5 (2011), 29.

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
    Thirring model, 1+1D QED

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

Ave. Phase Fact.

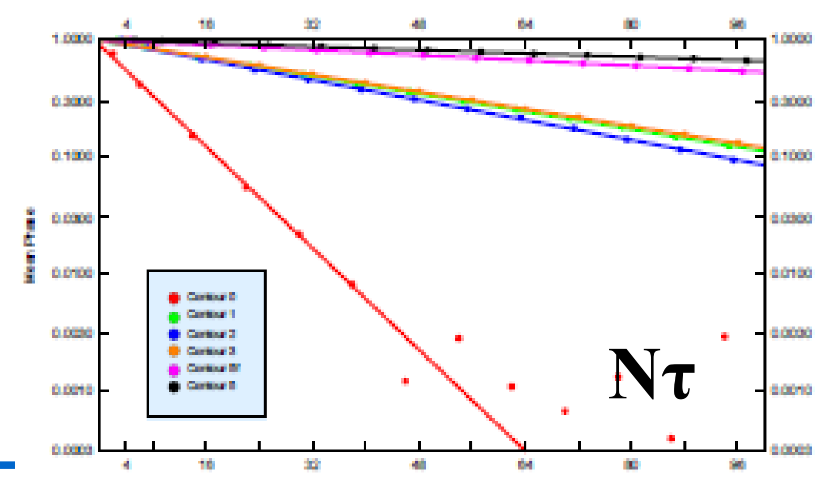


μ

  - **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.

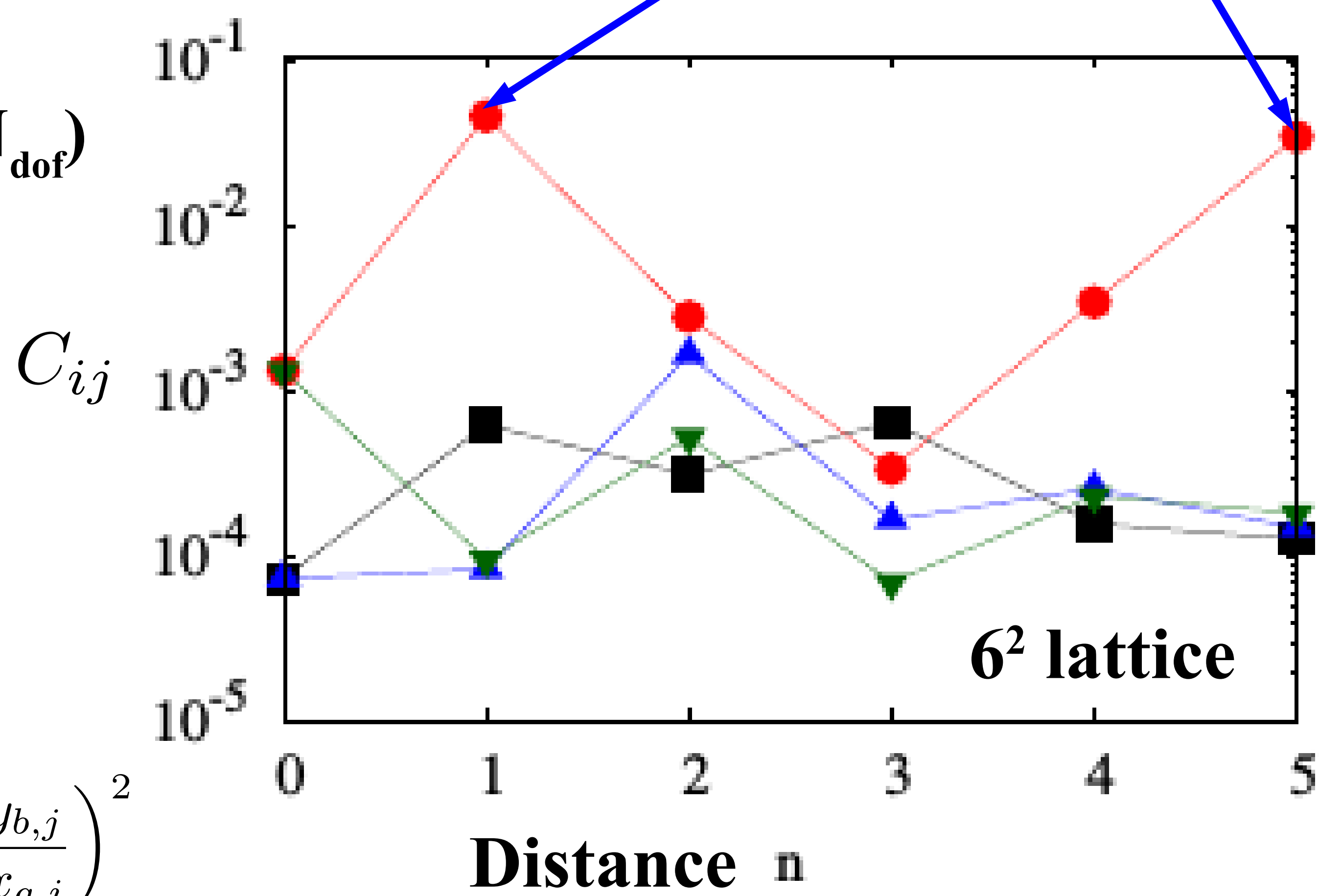


Which y's should be optimized ?

- **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