最適化問題としての符号問題

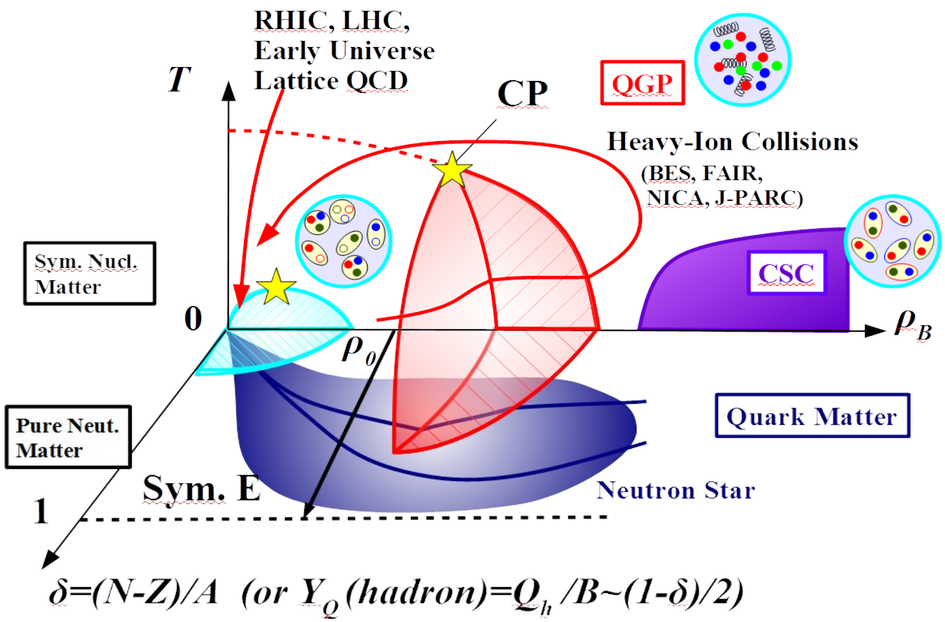
大西　明 (京大基研)
in collaboration with
森　勇登 (M2, 京大理), 柏 浩司 (京大基研)

新潟大学物理教室コロキウム , Dec.12, 2017

Y. Mori, K. Kashiwa, A. Ohnishi, Phys. Rev. D 96 (2017), 111501(R) [arXiv:1705.05605]
Y. Mori, K. Kashiwa, A. Ohnishi, arXiv:1709.03208.

Introduction

- **Sign problem for complex actions**
  - **Grand challenge in theor. phys.**
  - **Largest obstacle to explore QCD phase diagram**

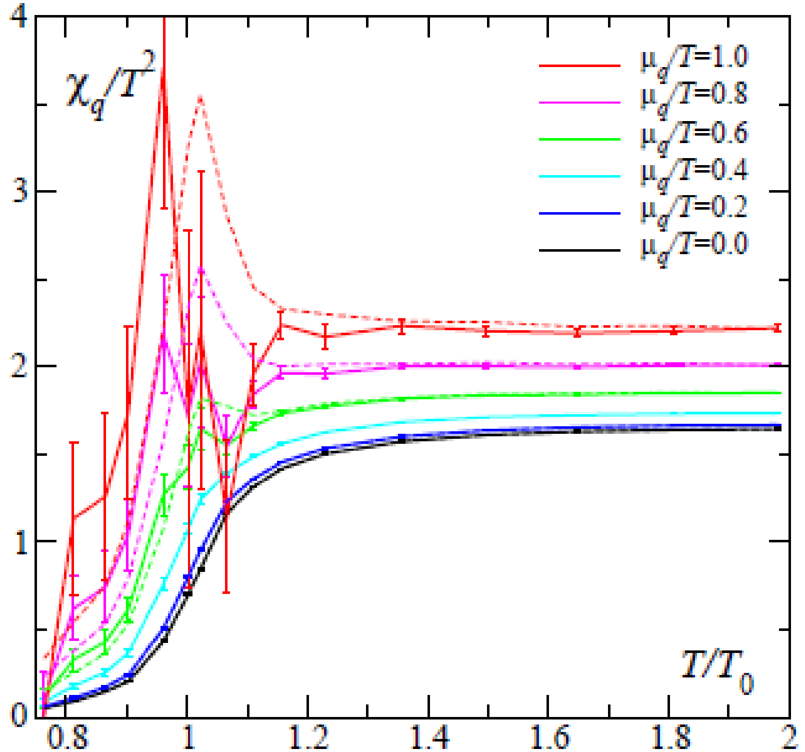


- **Approaches**
  - **Taylor expansion, Analytic cont., Canonical, Strong coupling, …**
  - **Complex Langevin method (CLM)**
    G. Parisi ('83), G. Aarts et al. ('10)
  - **Lefschetz thimble method (LTM)**
    E. Witten ('10), Cristoforetti et al. ('12), Fujii et al. ('13)
  - **Generalized LTM (GLTM)**
    A. Alexandru, et al., ('16)

Complexified variables & Shifting path (area)

Finite Density QCD

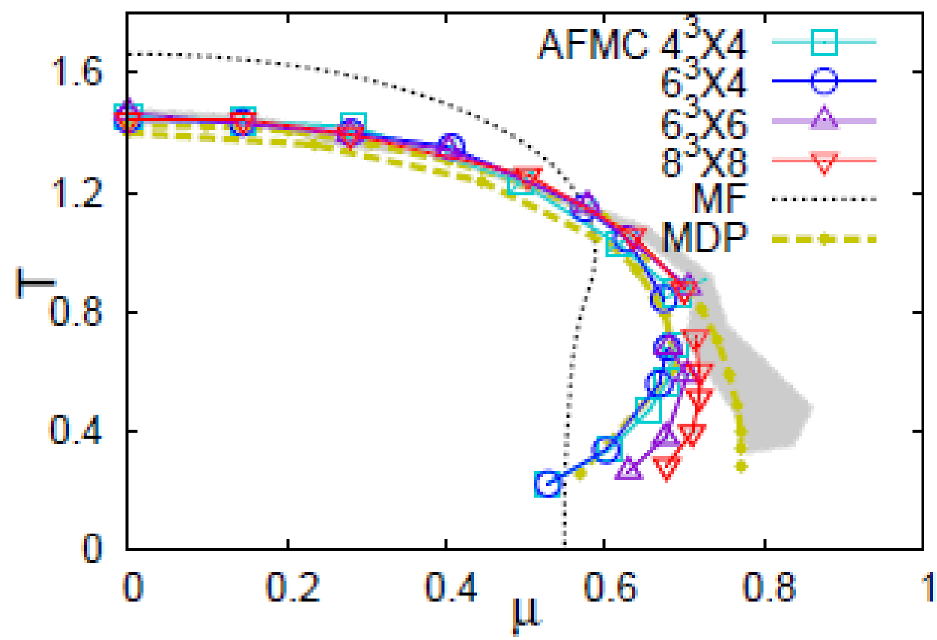
Taylor expansion



(a)

C.R. Allton, M. Doring, S. Ejiri, S.J. Hands, O. Kaczmarek, F. Karsch, E. Laermann, K. Redlich, Phys. Rev. D 71, 054508 (2005), hep-lat/0501030

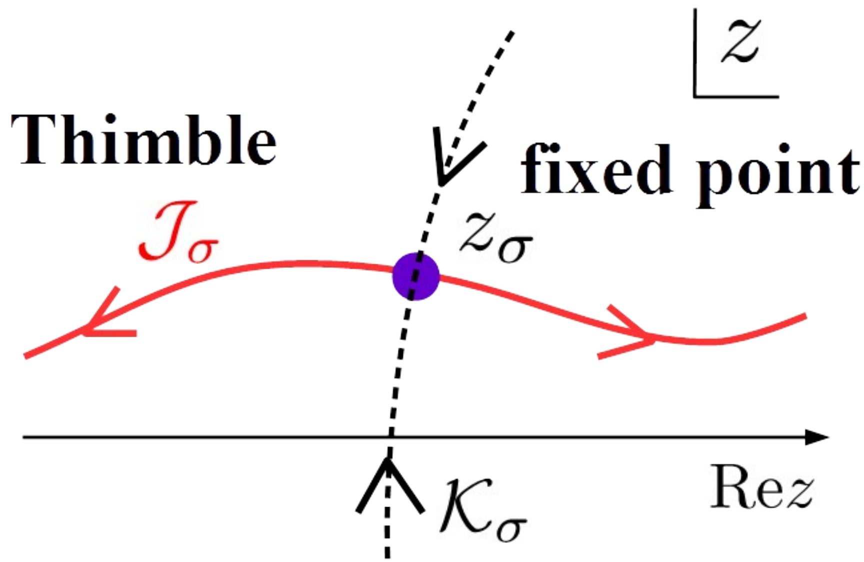
Strong Coupling



T. Ichihara, A. Ohnishi, T.Z. Nakano, PTEP 2014,123D02 [1401.4647].

Lefschetz Thimble Method

- **Integral over thimbles defined by the flow equation for complexified variables → Im S = const. on a thimble**



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

$$\left.\frac{\partial S}{\partial z_i}\right|_{z_\sigma} = 0 \ , \quad \frac{dz_i(t)}{dt} = \overline{\left(\frac{\partial S[z]}{\partial z_i}\right)} \qquad \mathcal{C} = \sum_\sigma n_\sigma \mathcal{J}_\sigma \ , \quad n_\sigma = \langle \mathcal{C}_\mathbb{R}, \mathcal{K}_\sigma \rangle$$

- **Pons**
  - **Has mathematically solid base.**
- **Cons**
  - **Phase from measure (residual sign prb.)**
  - **Cancellation between thimbles (global sign prb.)**
  - **Flow equation blows up somewhere.**

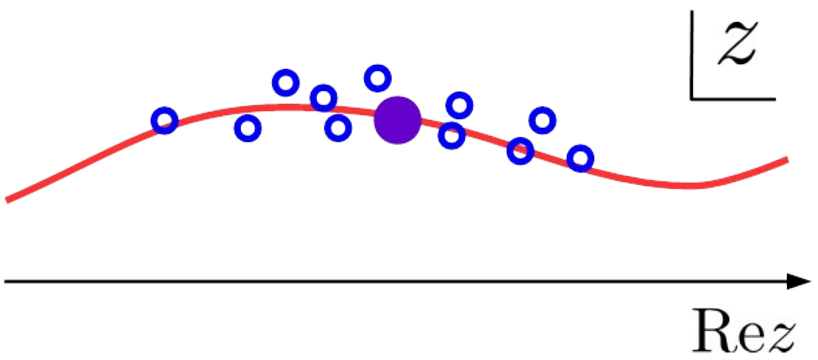
Complex Langevin Method

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

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

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

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



- **Pros**
  - **Easier to apply to large DOF theories**
- **Cons**
  - **Excursion problem → Gauge Cooling** *(Seiler et al. ('13))*
  - **Converged results can be wrong → Criteria** *(Nagata et al. ('16))*
  - **Singular drift problem → Several prescriptions ….**

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

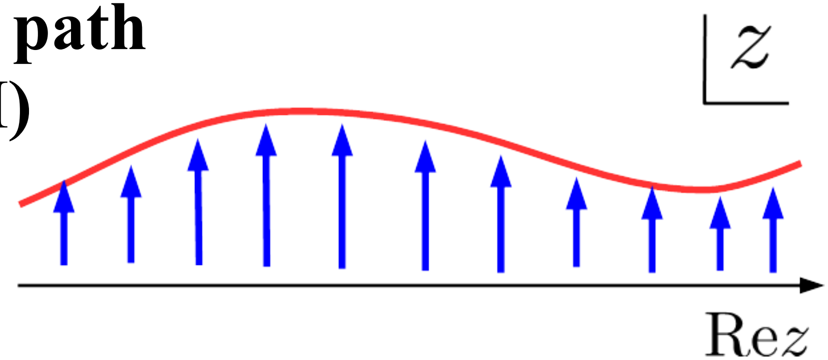
Contents

- **Introduction**
- **Path Optimization Method**
- **Application**
  - **One Variable Toy Model**
  - **Neural Network**
  - **Two dimensional complex $\lambda\varphi^4$ theory**
- **Summary**

Path Optimization Method

Path Optimization Method

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



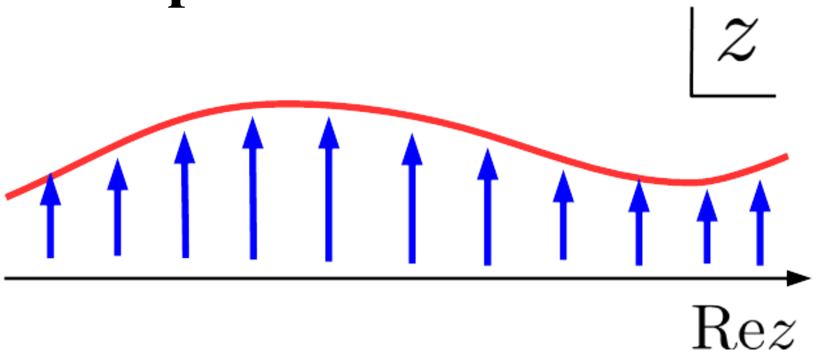
- **POM Procedure**
  - **Parametrize the path appropriately (Trial Function)**
  - **Set a measure of sign problem (Cost Function)**
  - **Tune parameters to minimize the Cost Function (Optimization)**

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**

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



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

- **Set the seriousness of the sign problem (Cost Function)**
  - **How much the phase fluctuate**

$$F[z(t)] = \frac{1}{2\mathcal{Z}} \int dt \left| e^{i\theta(t)} - e^{i\theta_0} \right|^2 \left| J(t) e^{-S[z(t)]} \right| = \left| \langle e^{i\theta} \rangle_{\rm pq} \right|^{-1} - 1 \quad \left[\theta = \arg(Je^{-S}), \ \theta_0 = \arg(\mathcal{Z})\right]$$

- **Optimization: Gradient descent, Neural Network, …**

Merits of using Path Optimization Method

- **Integral on an integral path**

$$\mathcal{Z} = \int dt(dz/dt)\exp(-S[z(t)])\ ,\quad \langle\mathcal{O}\rangle = \frac{1}{\mathcal{Z}}\int dt(dz/dt)\mathcal{O}(z(t))\exp(-S[z(t)])$$

→ Partition function and observable average are independent of the path due to the Cauchy(-Poincare) theorem, as long as,

  - **the path do not go across the singular points of exp(-S),**
  - **and the contribution from Re z → ±∞ is negligible.**

- **We do not have to care the singular points of the action (S), as long as exp(-S) is not singular.**
- **Demerits**
  - **There is no guiding principle to modify the path except for reducing the cost function.**
  - **Then, the number of parameters are large and large CPU power is required.**

Application (1)
One variable toy model

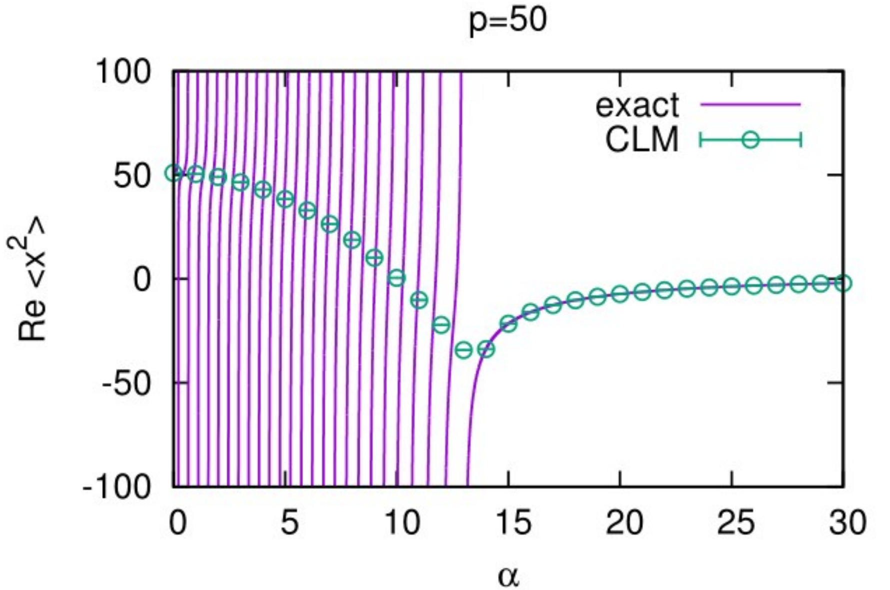
A (Pathological) Toy Model

- **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 *α***
  - **Large p → Strong oscillation of the Boltzmann weight**
  - **Small α → Singular point at *z = -iα* is close to the real axis**



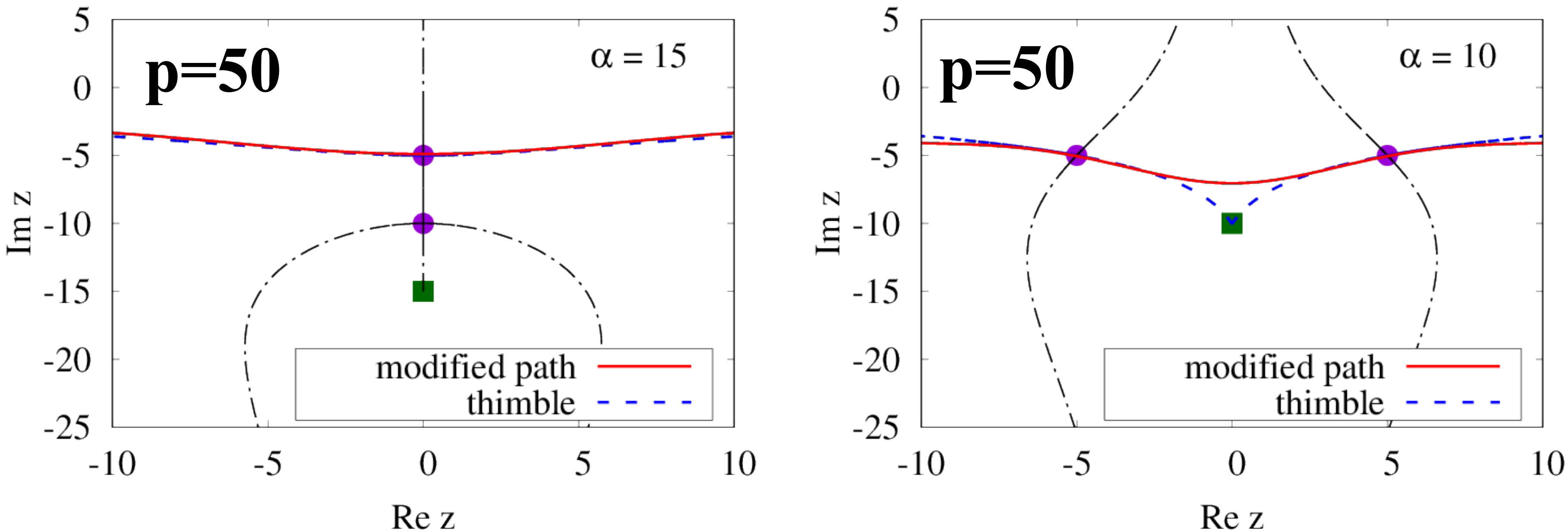
Optimized Path

- **Trial Function**

Mori, Kashiwa, AO ('17)

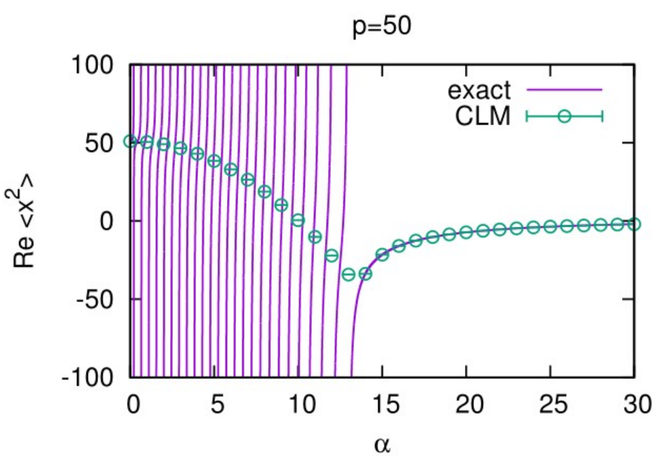
$$z(t) = t + i\left[c_1 \exp(-c_2^2 t^2/2) + c_3)\right]$$

- **Optimization = Gradient descent**
- **Optimized path agrees with thimble(s) around the fixed point(s) !**
  - **Large α → One thimble, Singular point is far away from thimble**
  - **Small α → Go through two FPs.**

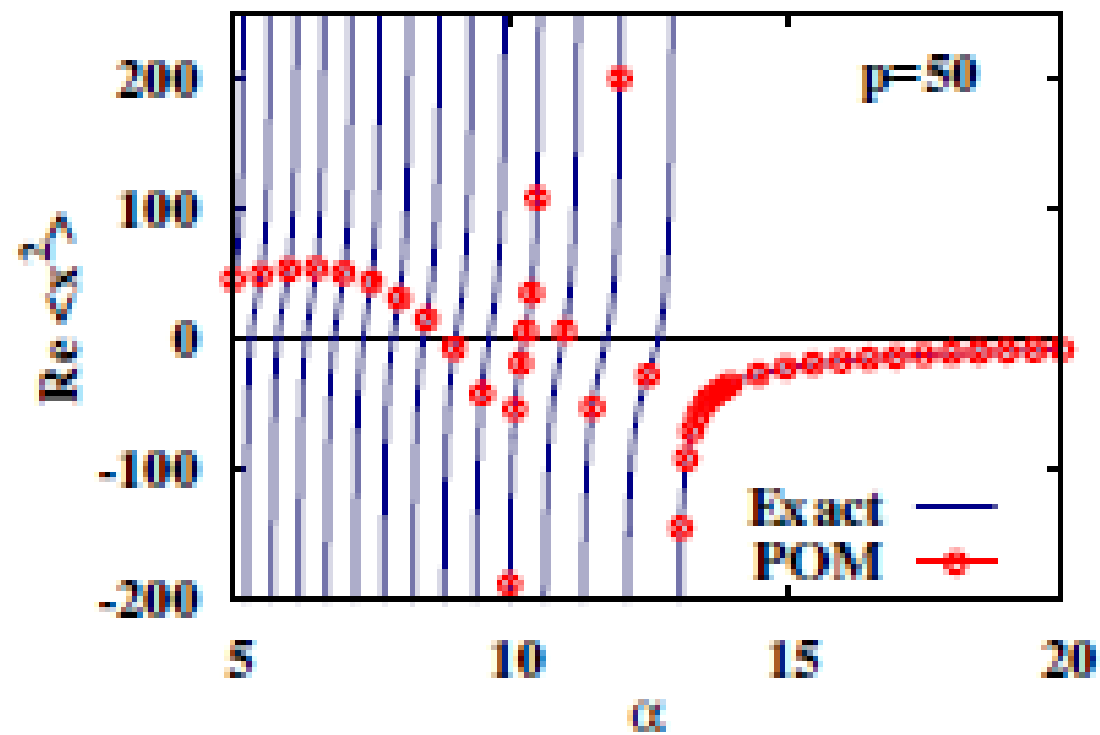


Expectation Value of x^2

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

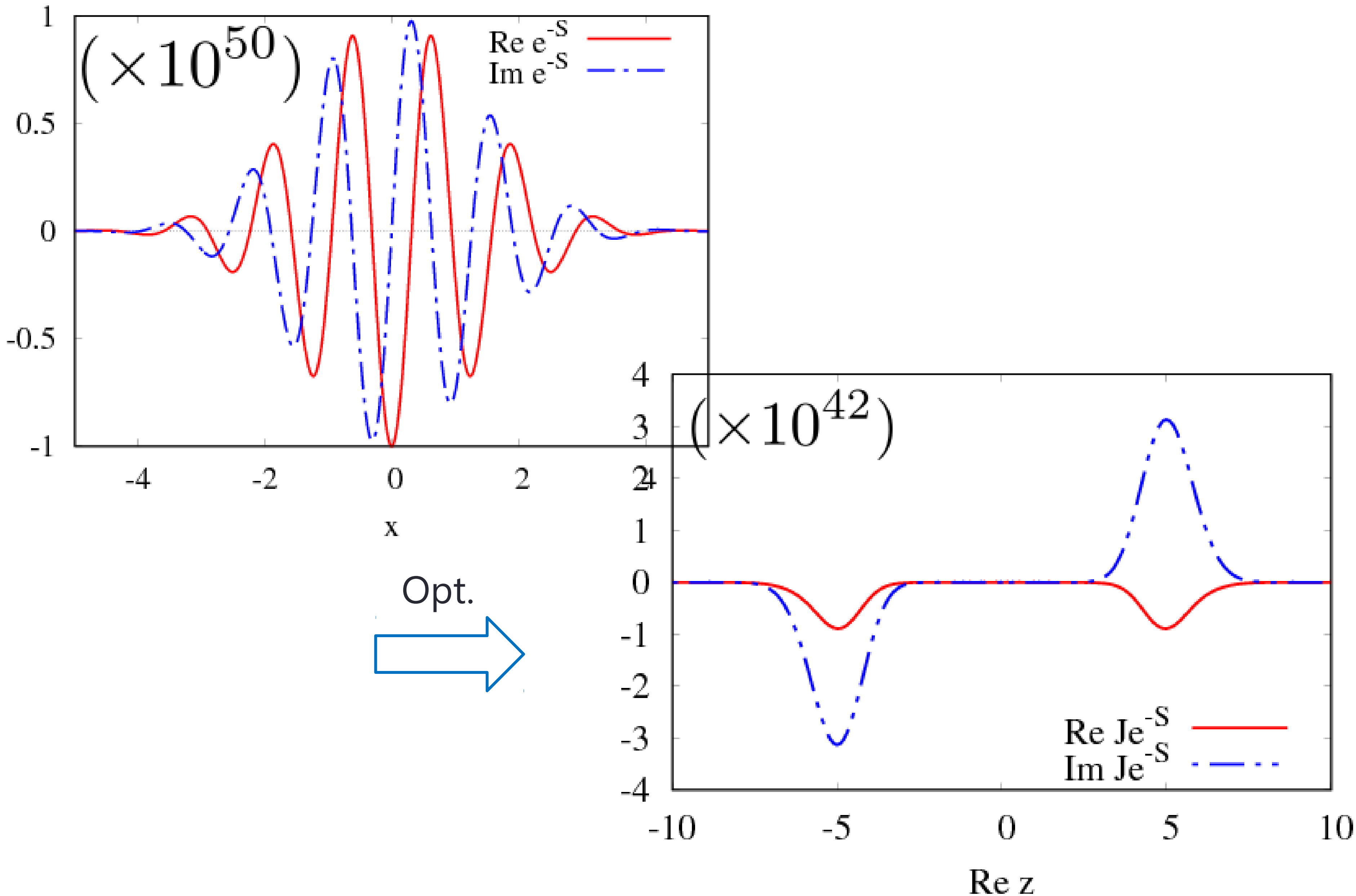


Nishimura, Shimasaki ('15)



Mori, Kashiwa, AO ('17)

Boltzmann Weight on Optimized Path

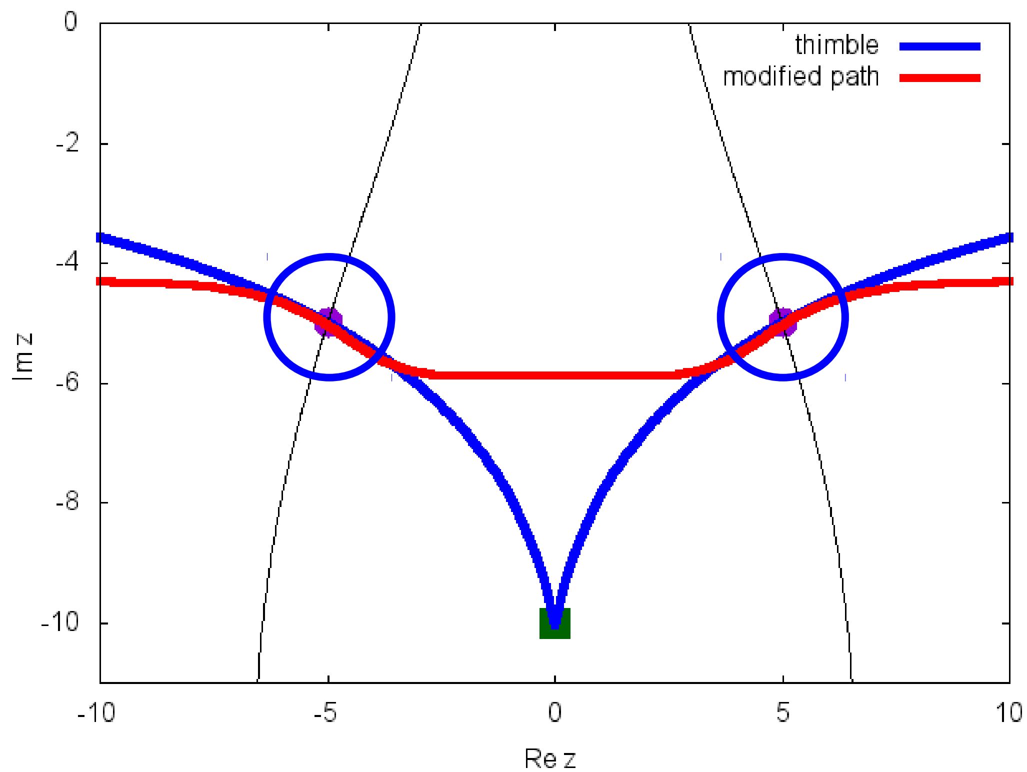


It's just an accident !
POM works only in cases you know thimbles
and you can prepare the trial function
which easily mimic thimbles !

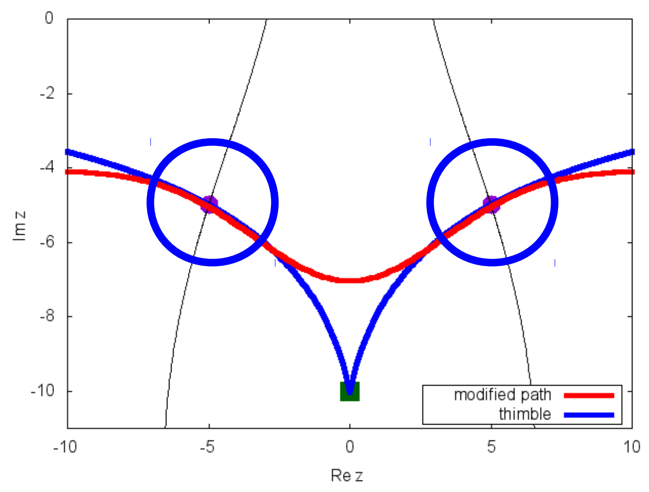
We want to make an objection !
POM works even if we do not specify the function form
but use a general form of functions
provided by a neural network !

Optimized Path by Neural Network

Neural Network



Gaussian +Gradient Descent



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

Mori, Kashiwa, AO (in prep.)

Neural Network

経路最適化法｜ニューラルネットワーク

- 格子上の場の理論・・・多変数
- 積分経路に適切な関数形は非自明

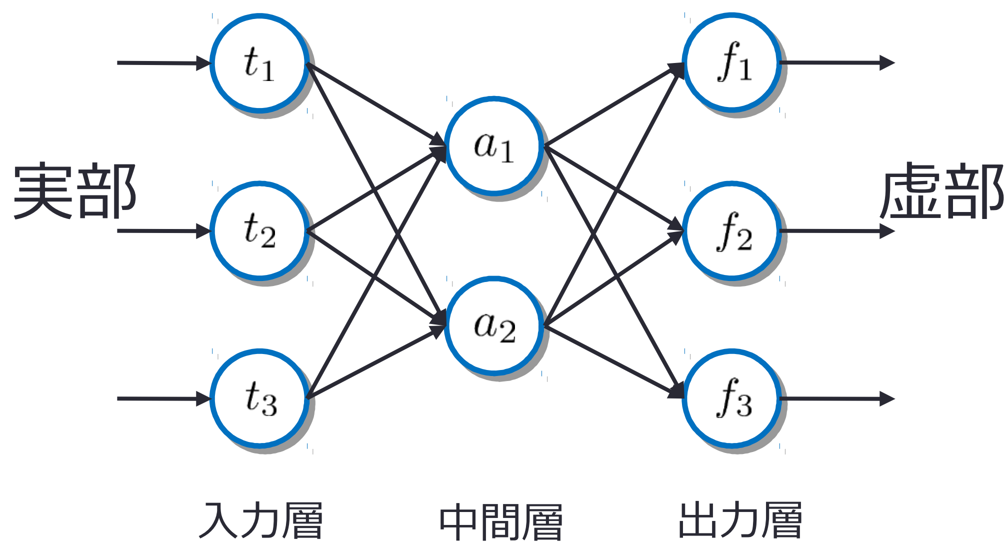
ニューラルネットワーク

$$z_i(t) = t_i + i(\alpha_i f_i(t) + \beta_i)$$

$$\begin{cases} a_i = g(W_{ij}^{(1)} t_j + b_i^{(1)}) \\ f_i = g(W_{ij}^{(2)} a_j + b_i^{(2)}) \end{cases}$$

$g(x)$：活性化関数 (tanh 等)

※ $W,\ b,\ \alpha,\ \beta$ がパラメータ

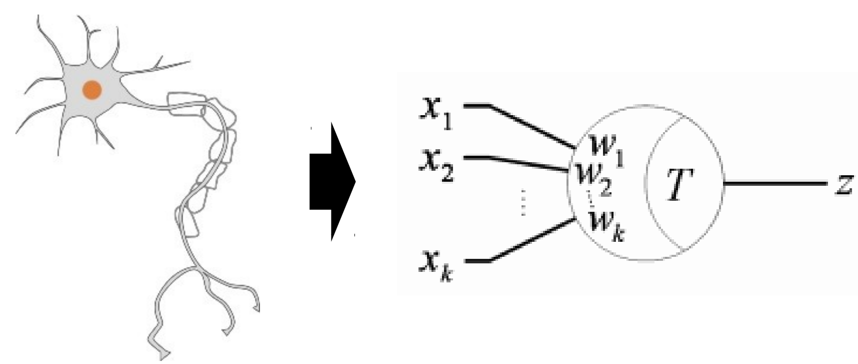


Y. Mori, **K.K.** and A. Ohnishi, arXiv:1705.05605, to be published in PRD
Y. Mori, **K.K.** and A. Ohnishi, arXiv:1709.03208

In the original method,
we prepare functional forms of the trial function by hand

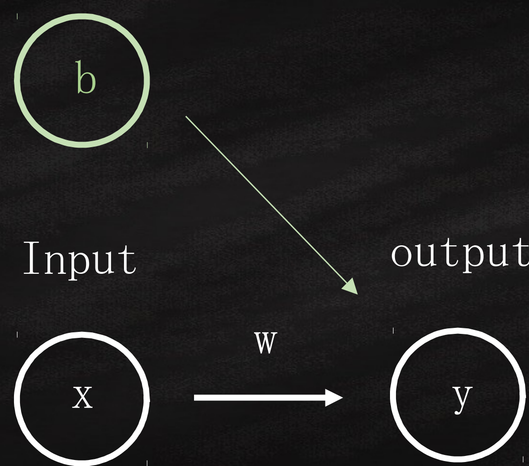
It takes our research time …

Human power ➡ **Neural Network**



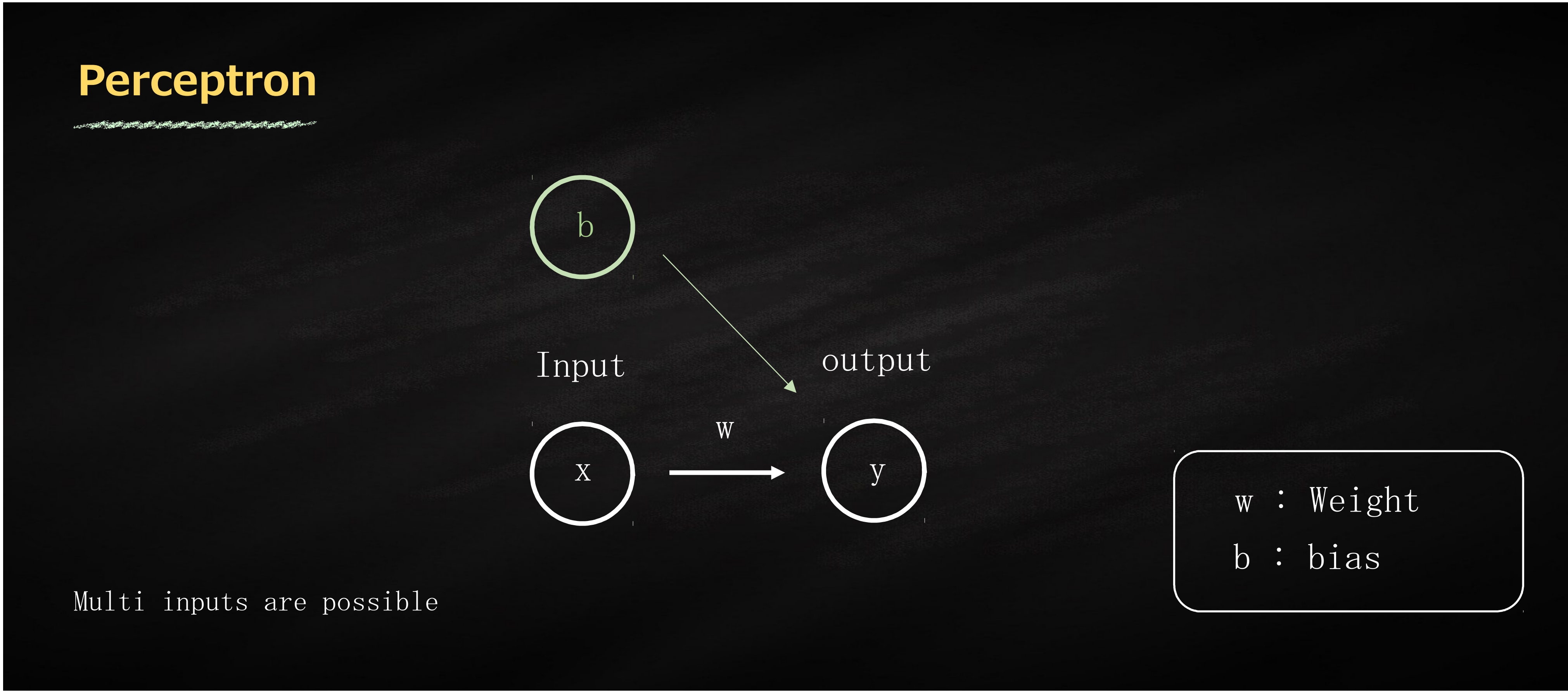
Next: Brief explanation of neural network

Perceptron



w : Weight
b : bias

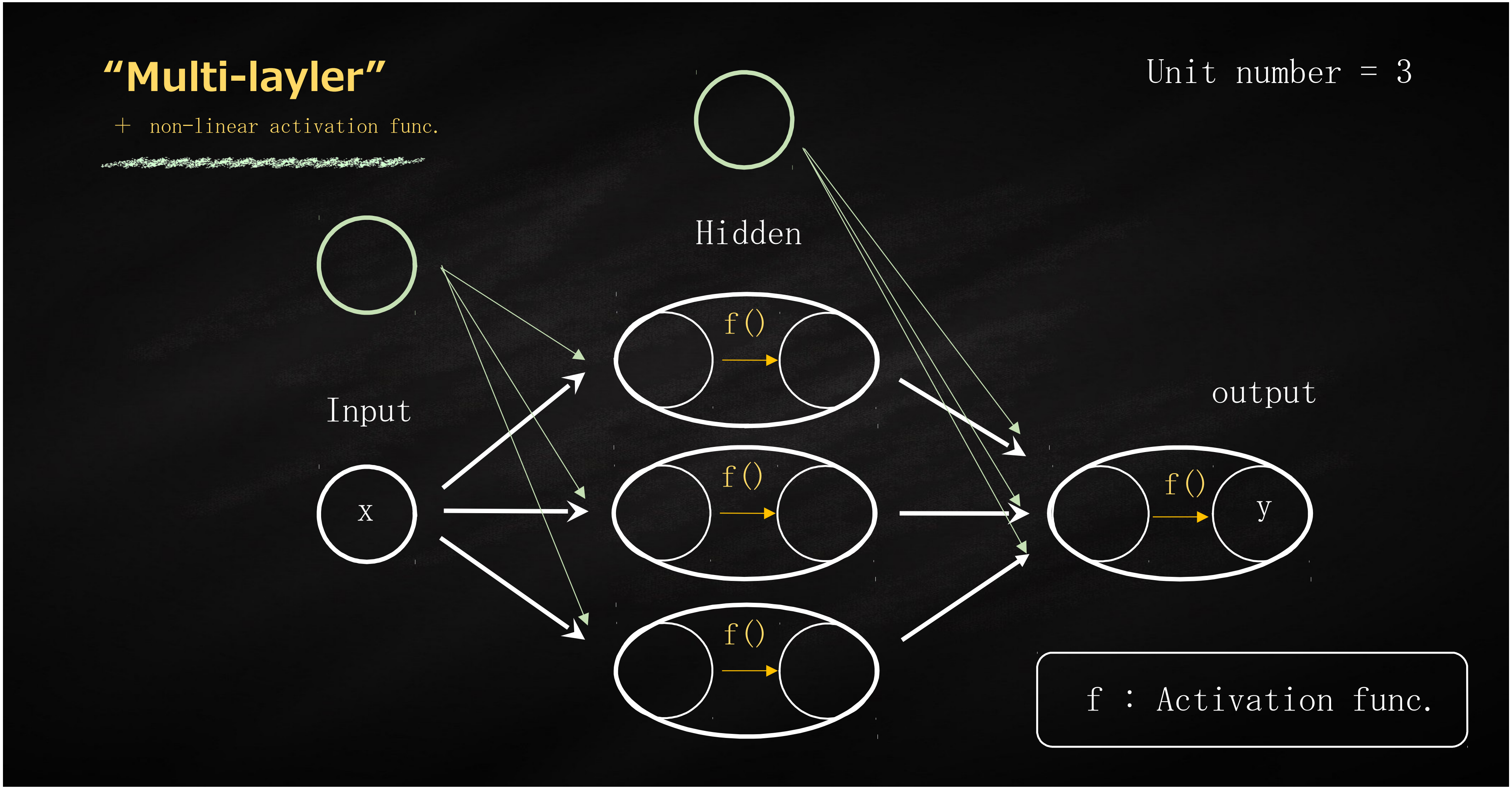
Multi inputs are possible

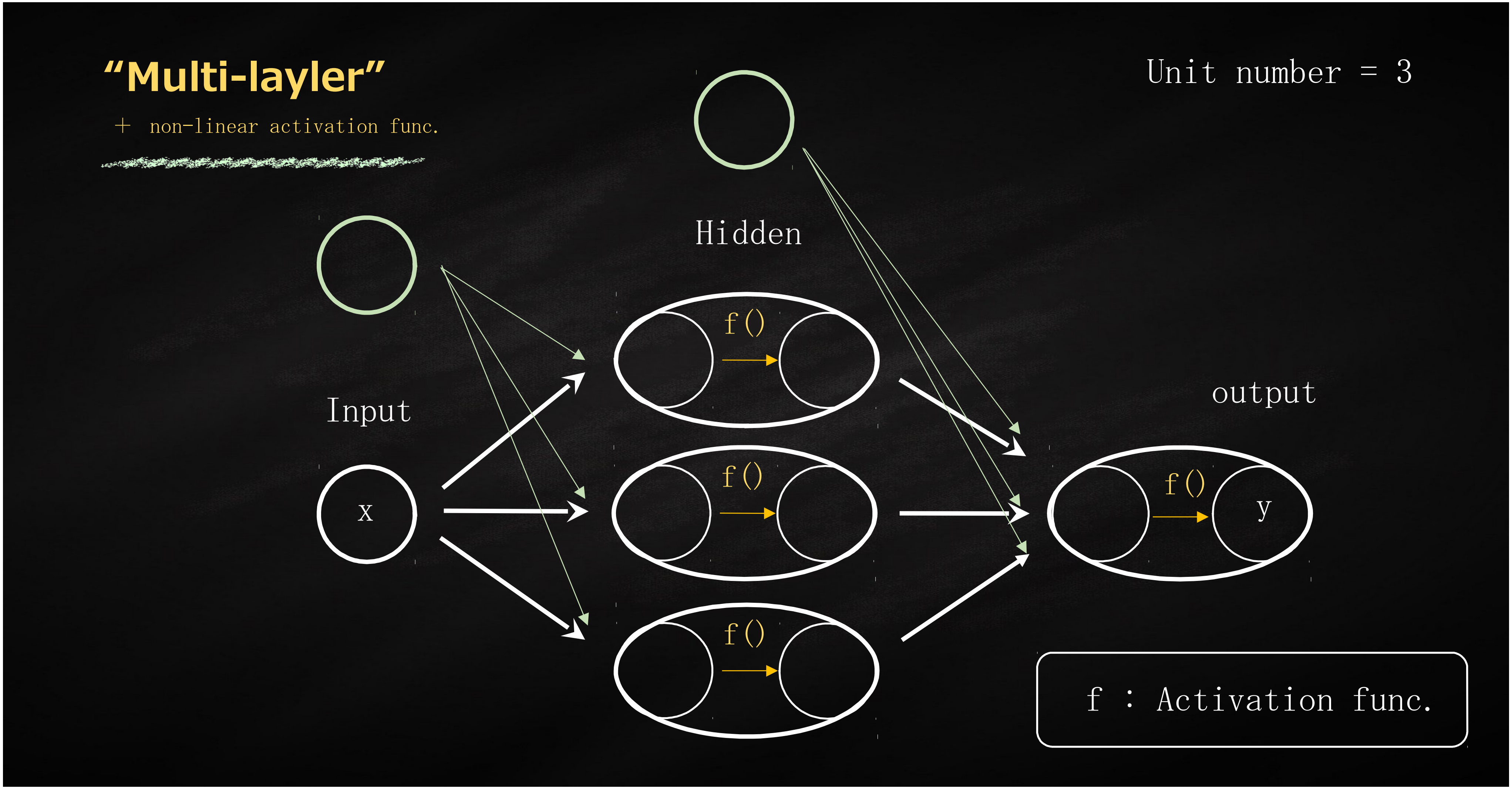


In this case, we will obtain the result; $y = wx + b$

It only contains simple addition and multiplication processes

（Usually、 we put the step function in the perceptron; the output becomes 0 or 1）

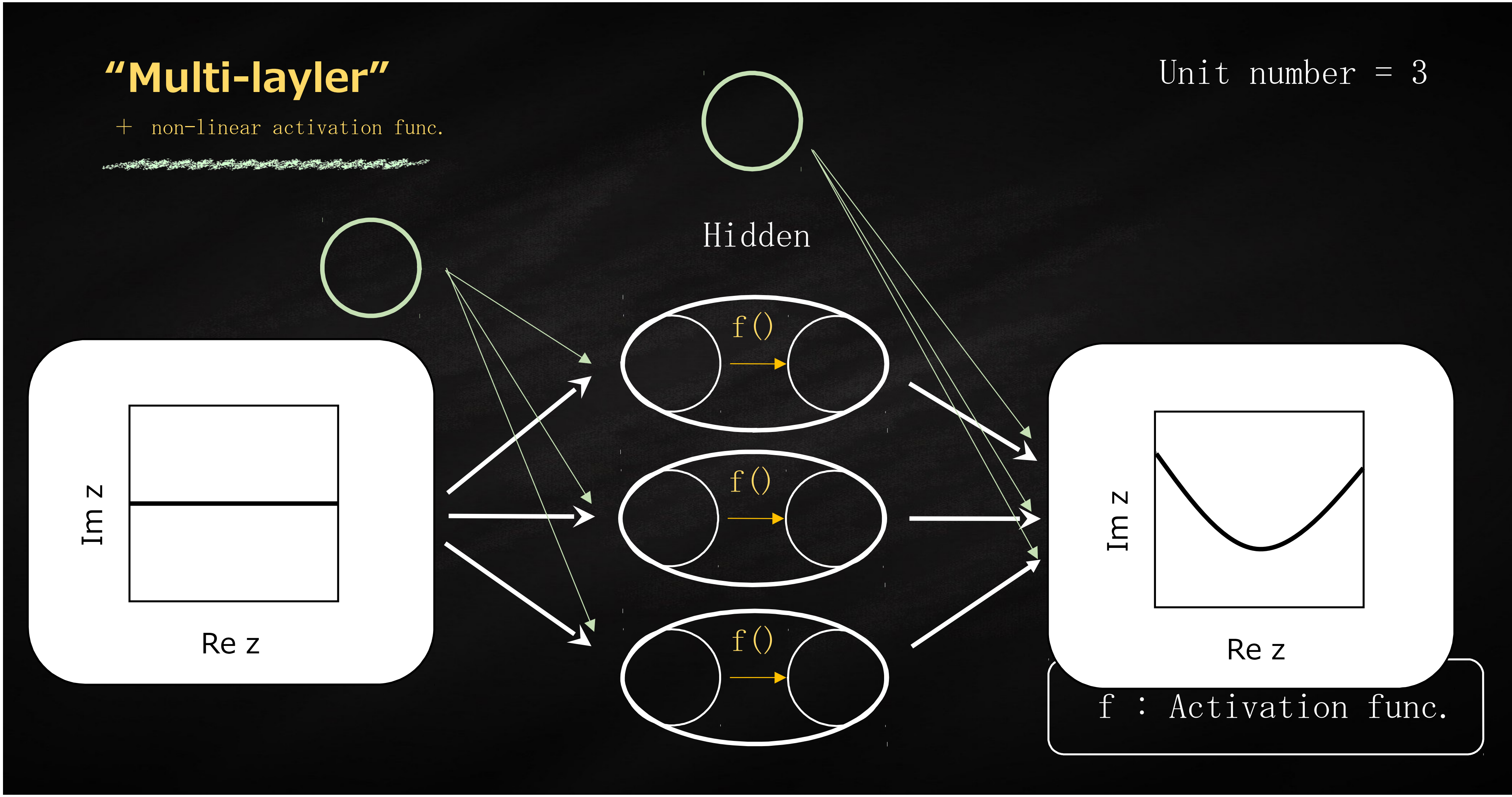

"Multi-layler"
+ non-linear activation func.
Unit number = 3
Hidden
Input
x
f()
f()
f()
output
f()
y
f : Activation func.



To perform nonlinear calculation,
we should add the hidden layer and the activation function

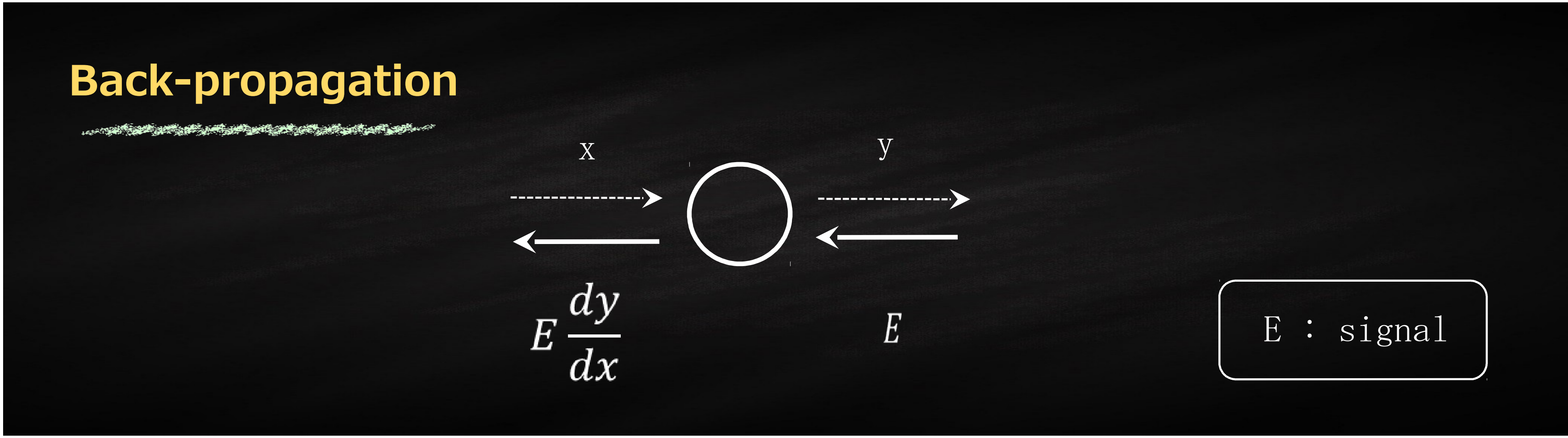
Parameters are determined via the back-propagation

We use tanh as the activation function



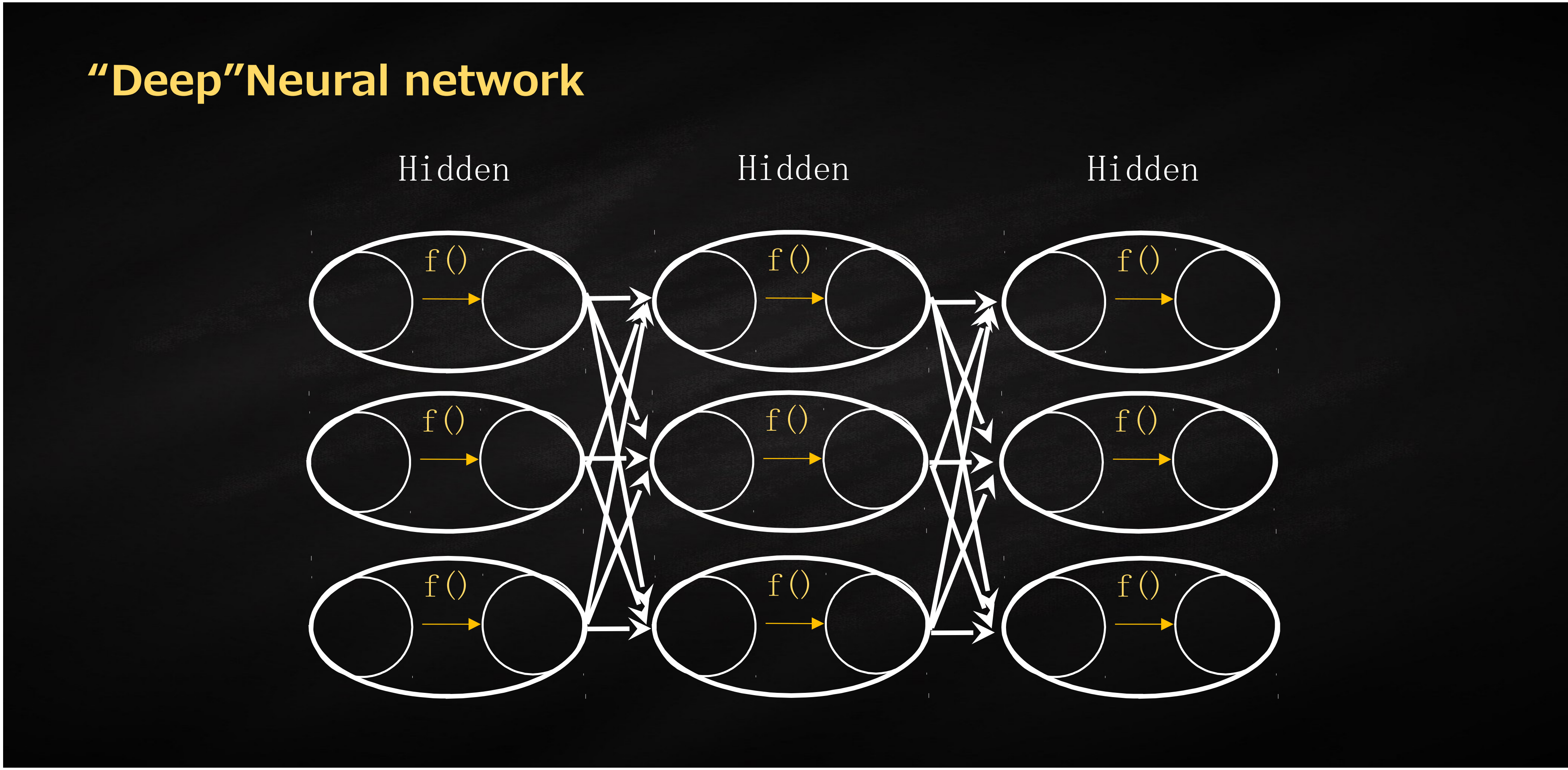
Real part of the integral path is input and then the imaginary part becomes output

We develop the numerical code by Fortran



Rough flow-chart of the path optimization method

1. Input the real part of the integral path (Hybrid Monte-Calro + mini-batch training)
2. Estimate gradients of the cost function from upstream (Back-propagation)
3. Update parameters (Stochastic gradient decent 、 AdaGrad, Adadelta etc…)
4. Output the imaginary part of the integral path (Exponential moving average)



Even in the simple NN, we can approximate **arbitral continuous functions**

Universal approximation theorem

G. Cybenko, MCSS 2, 303 (1989)
K. Hornik, Neural networks 4, 251 (1991)

It works well
in the path optimization method

Two dimensional complex $\lambda\varphi^4$ theory

複素 ϕ^4 理論

- 2D Lattice

$$S = \sum_x \left[(4+m^2)\phi_x^*\phi_x + \lambda(\phi_x^*\phi_x)^2 - \sum_{\nu=0}^{1} \underline{(\phi_x^* e^{-\mu\delta_{\nu,0}}\phi_{x+\hat{\nu}} + \phi_{x+\hat{\nu}}^* e^{+\mu\delta_{\nu,0}}\phi_x)} \right]$$

$\in \mathbb{C}$

- 積分変数

$\phi = \frac{1}{\sqrt{2}}(x_1 + ix_2)$ 積分変数 $x_1,\ x_2$ について作用 は解析的

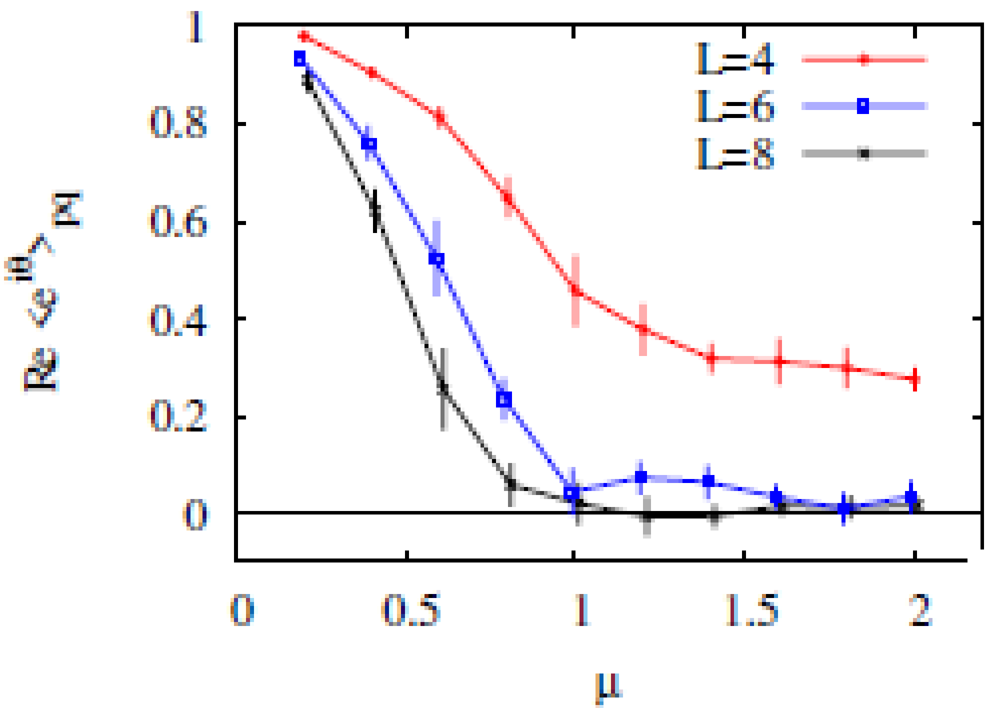
→ $x_1,\ x_2$ それぞれを複素化

$$z_i(t) = t_i + i(\alpha_i f_i(t) + \beta_i)$$

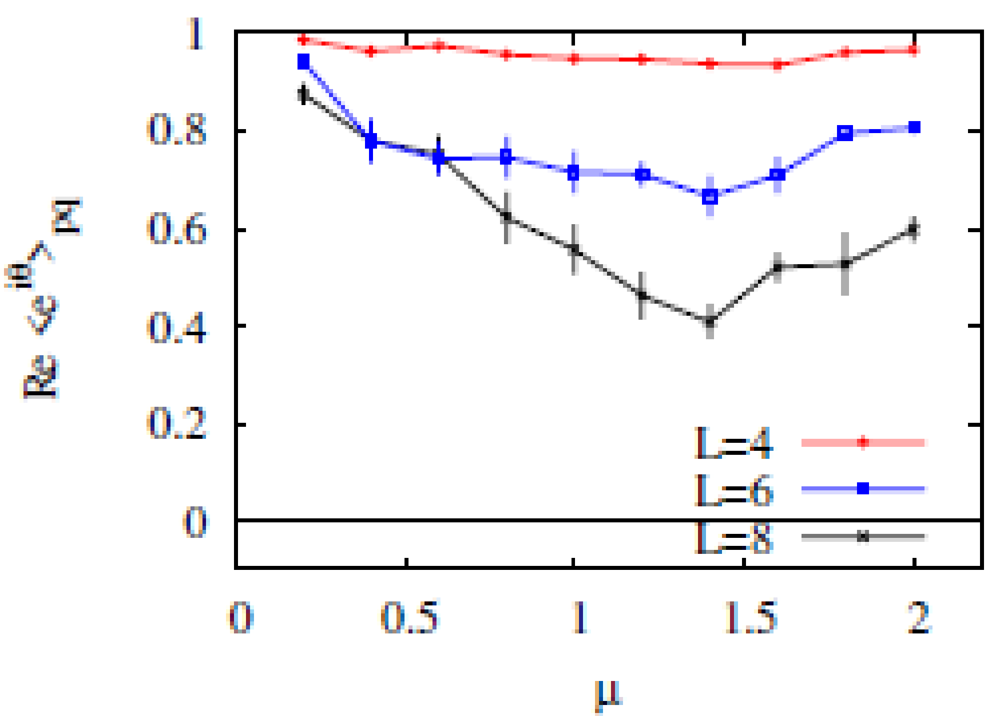
以下の計算では $m=1,\ \lambda=1$ の場合を考える

Average Phase Factor

Y. Mori, K. Kashiwa, AO, 1709.03208

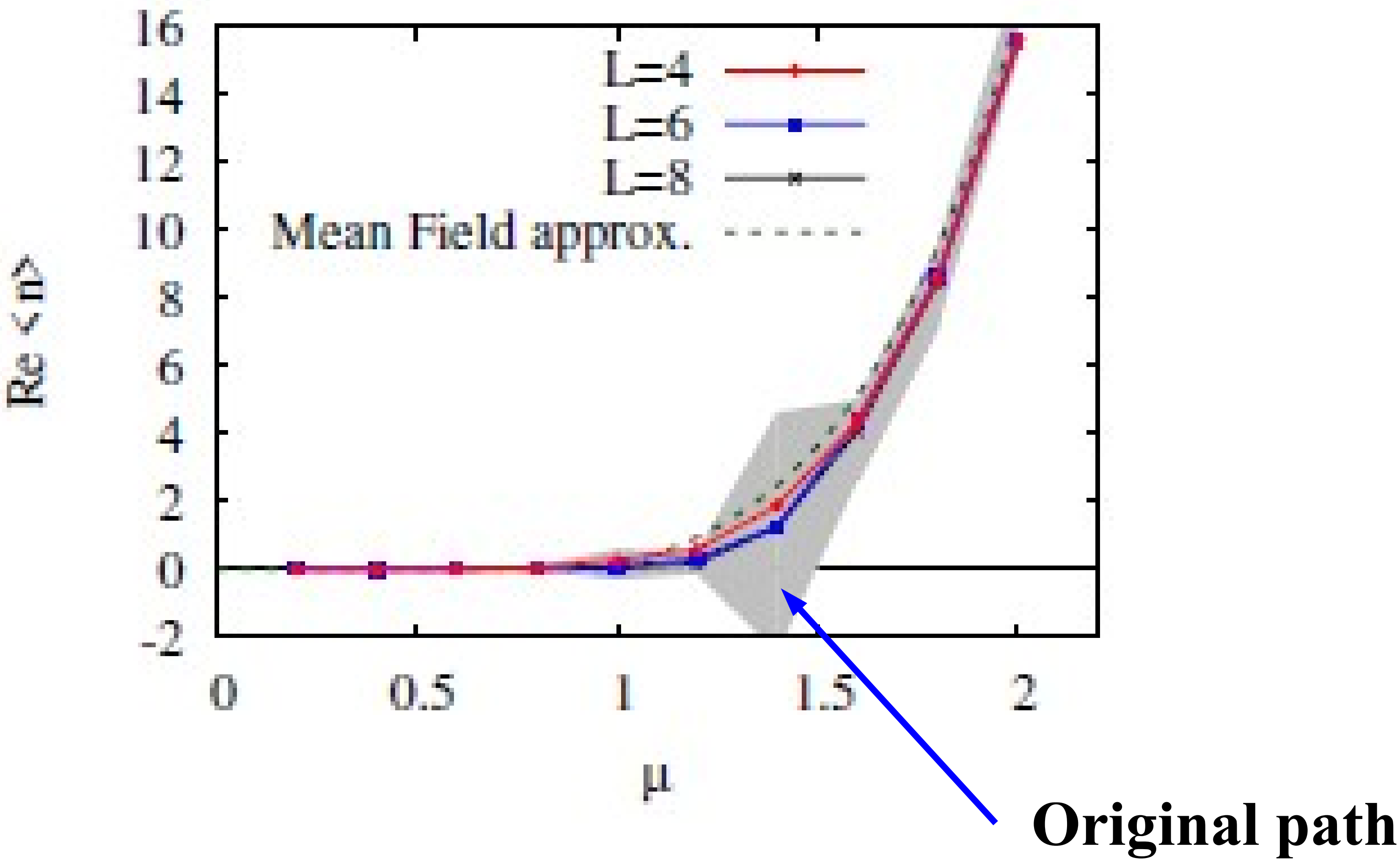


Path Optimization



Number Density

Y. Mori, K. Kashiwa, AO, 1709.03208



Complex λφ⁴ Theory

Mean Field Approximation

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

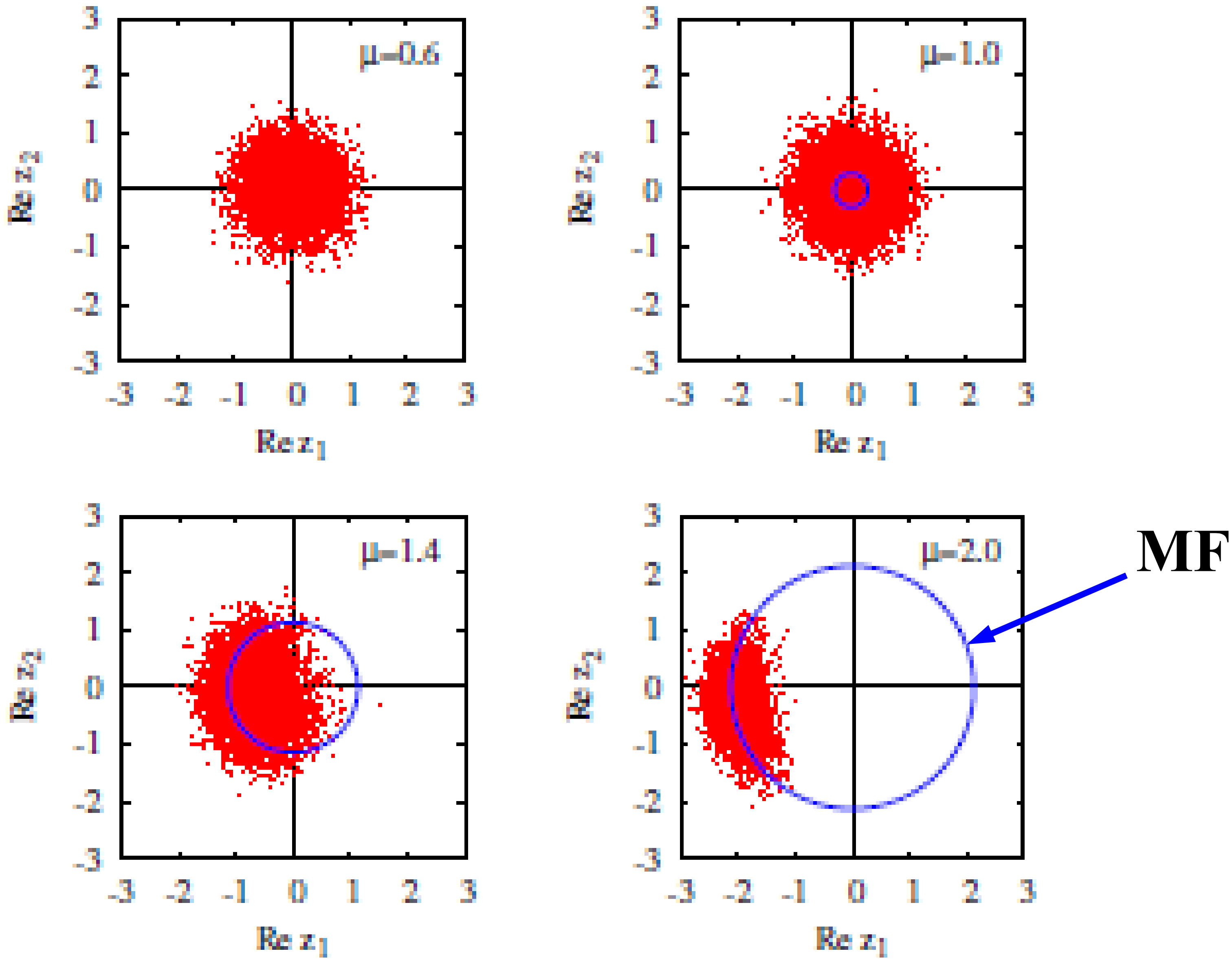
$$n = \phi^2 \sinh\mu \ ,$$

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

Spontaneous symmetry breaking takes place at $\mu > \mu_c = 0.962...$

Field Configuration

Y. Mori, K. Kashiwa, AO, 1709.03208



Summary

- **Path Optimization Method for the sign problem is proposed.**
  → Sign problem can be regarded as an optimization problem.
- **Usefulness of POM is demonstrated in a toy model.**
  - **Optimized path reproduces the thimble(s) around the fixed point(s).**
  - **Singular points with zero Boltzmann weight do not matter, since they do not contribute to the integral.**
- **POM is applied to field theory with optimization using neural network.**
  - **Optimization with small human power (large computer power).**
  - **Two dim. φ^4 theory seems to be solved well.**
  - **Spontaneous symmetry breaking takes place.**
- **Stay tuned.**

Parameter update

- **ADADELTA algorithm: Sophisticated gradient descent method**

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

$$v_i^{(j+1)} = \frac{\sqrt{s_i^{(j)} + \epsilon}}{\sqrt{r_i^{(j+1)} + \epsilon}} F_i^{(j)} \quad (F_i = \partial \mathcal{F}_{\rm cost} / \partial c_i)$$

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

η= learning rate, γ=decay rate (survival rate)

- **Batch training (average of the derivative with several conigs.)**

$$F_i \to \frac{1}{N_{\rm batch}} \sum_{k=1}^{N_{\rm batch}} F_i(t^{(k)}, c).$$