Sign problem in Monte Carlo simulations and the tempered Lefschetz thimble method

Masafumi Fukuma (Dept Phys, Kyoto Univ)

Sep 2, 2019
熱場の量子論とその応用@YITP

Based on work with

Nobuyuki Matsumoto (Kyoto Univ) & **Naoya Umeda** (PwC)

-- **MF** and **Umeda**, "Parallel tempering algorithm for integration over Lefschetz thimbles" [arXiv:1703.00861, PTEP2017(2017)073B01]
-- **MF**, **Matsumoto** and **Umeda**, "Applying the tempered Lefschetz thimble method to the Hubbard model away from half-filling", [arXiv:1906.04243]

Also, for the geometrical optimization of tempering algorithms and an application to QG:

-- **MF**, **Matsumoto** and **Umeda**,
[arXiv:1705.06097, JHEP1712(2017)001], [arXiv:1806.10915, JHEP1811(2018)060]

→ **Matsumoto's poster (today)**

1. Introduction

Summary

The **numerical sign problem** is one of the major obstacles
when performing numerical calculations in various fields of physics

Typical examples:

① Finite density QCD
② Quantum Monte Carlo simulations of quantum statistical systems
③ Real time QM/QFT

Today, I would like to

-- give a review on various methods towards solving the sign problem

-- argue that
a new algorithm "Tempered Lefschetz thimble method" (TLTM)
is a promising method,
by exemplifying its effectiveness for:

② Quantum Monte Carlo simulations
of strongly correlated electron systems,
especially the Hubbard model away from half-filling

Sign problem

Our main concern is to calculate: $\langle \mathcal{O}(x) \rangle_S \equiv \dfrac{\int dx\, e^{-S(x)} \mathcal{O}(x)}{\int dx\, e^{-S(x)}}$

$$\begin{cases} x = (x^i) \in \mathbb{R}^N \text{: dynamical variable (real-valued)} \\ S(x)\text{: action, } \mathcal{O}(x)\text{: observable} \end{cases}$$

Markov chain Monte Carlo (MCMC) simulation:

probability distribution function

When $S(x) \in \mathbb{R}$, one can regard $p_{\text{eq}}(x) \equiv e^{-S(x)} / \int dx\, e^{-S(x)}$ as a PDF:

$$0 \le p_{\text{eq}}(x) \le 1, \quad \int dx\, p_{\text{eq}}(x) = 1$$

⇨ Generate a sample $\{x^{(k)}\}_{k=1,\ldots,N_{\text{conf}}}$ from $p_{\text{eq}}(x)$

⇨ $\langle \mathcal{O}(x) \rangle \approx \dfrac{1}{N_{\text{conf}}} \displaystyle\sum_{k=1}^{N_{\text{conf}}} \mathcal{O}(x^{(k)})$

Sign problem:

When $S(x) = S_R(x) + i\, S_I(x) \in \mathbb{C}$, one cannot regard $e^{-S(x)} / \int dx\, e^{-S(x)}$ as a PDF

⇨ Reweighting method : treat $e^{-S_R(x)} / \int dx\, e^{-S_R(x)}$ as a PDF

⇨ $\langle \mathcal{O}(x) \rangle_S \equiv \dfrac{\langle e^{-i S_I(x)} \mathcal{O}(x) \rangle_{S_R}}{\langle e^{-i S_I(x)} \rangle_{S_R}} \approx \dfrac{e^{-O(N)} \pm O(1/\sqrt{N_{\text{conf}}})}{e^{-O(N)} \pm O(1/\sqrt{N_{\text{conf}}})}$ $\quad \begin{pmatrix} N\text{ : DOF} \\ N_{\text{conf}}\text{ : sample size} \end{pmatrix}$

⇨ Require $O(1/\sqrt{N_{\text{conf}}}) < e^{-O(N)}$ ⇨ $N_{\text{conf}} \simeq e^{O(N)}$

Sign problem

Our main concern is to calculate: $\langle \mathcal{O}(x) \rangle_S \equiv \dfrac{\int dx\, e^{-S(x)} \mathcal{O}(x)}{\int dx\, e^{-S(x)}} = \dfrac{\int dx\, e^{-S_R(x)} e^{-i S_I(x)} \mathcal{O}(x)}{\int dx\, e^{-S_R(x)} e^{-i S_I(x)}}$

$\begin{cases} x = (x^i) \in \mathbb{R}^N\text{: dynamical variable (real-valued)} \\ S(x)\text{: action, } \mathcal{O}(x)\text{: observable} \end{cases}$

Markov chain Monte Carlo (MCMC) simulation:

probability distribution function

When $S(x) \in \mathbb{R}$, one can regard $p_{\rm eq}(x) \equiv e^{-S(x)} / \int dx\, e^{-S(x)}$ as a PDF:

$$0 \le p_{\rm eq}(x) \le 1, \quad \int dx\, p_{\rm eq}(x) = 1$$

⇨ Generate a sample $\{x^{(k)}\}_{k=1,\ldots,N_{\rm conf}}$ from $p_{\rm eq}(x)$

⇨ $\langle \mathcal{O}(x) \rangle \approx \dfrac{1}{N_{\rm conf}} \sum_{k=1}^{N_{\rm conf}} \mathcal{O}(x^{(k)})$

Sign problem:

When $S(x) = S_R(x) + i S_I(x) \in \mathbb{C}$, one cannot regard $e^{-S(x)} / \int dx\, e^{-S(x)}$ as a PDF

⇨ Reweighting method : treat $e^{-S_R(x)} / \int dx\, e^{-S_R(x)}$ as a PDF

⇨ $\langle \mathcal{O}(x) \rangle_S \equiv \dfrac{\langle e^{-i S_I(x)} \mathcal{O}(x) \rangle_{S_R}}{\langle e^{-i S_I(x)} \rangle_{S_R}} \approx \dfrac{e^{-O(N)} \pm O(1/\sqrt{N_{\rm conf}})}{e^{-O(N)} \pm O(1/\sqrt{N_{\rm conf}})}$ $\begin{pmatrix} N\text{ : DOF} \\ N_{\rm conf}\text{ : sample size} \end{pmatrix}$

⇨ Require $O(1/\sqrt{N_{\rm conf}}) < e^{-O(N)}$ ⇨ $\boxed{N_{\rm conf} \simeq e^{O(N)}}$

Example: Gaussian

Let us consider $\begin{cases} S(x) = \dfrac{\beta}{2}(x-i)^2 \equiv S_R(x) + i\,S_I(x) \\ \mathcal{O}(x) = x^2 \end{cases}$ $\left(\begin{array}{l} S_R(x) = \dfrac{\beta}{2}(x^2-1) \\ S_I(x) = -\beta x \end{array} \right)$

$\boxed{\beta \gg 1}$ ← large β mimics large DOF ($\beta \sim N$)

⇒ $$\langle x^2 \rangle_S = \frac{\left\langle e^{-i S_I(x)} x^2 \right\rangle_{S_R}}{\left\langle e^{-i S_I(x)} \right\rangle_{S_R}} = \frac{\left(\beta^{-1}-1\right) e^{-\beta/2}}{e^{-\beta/2}}$$

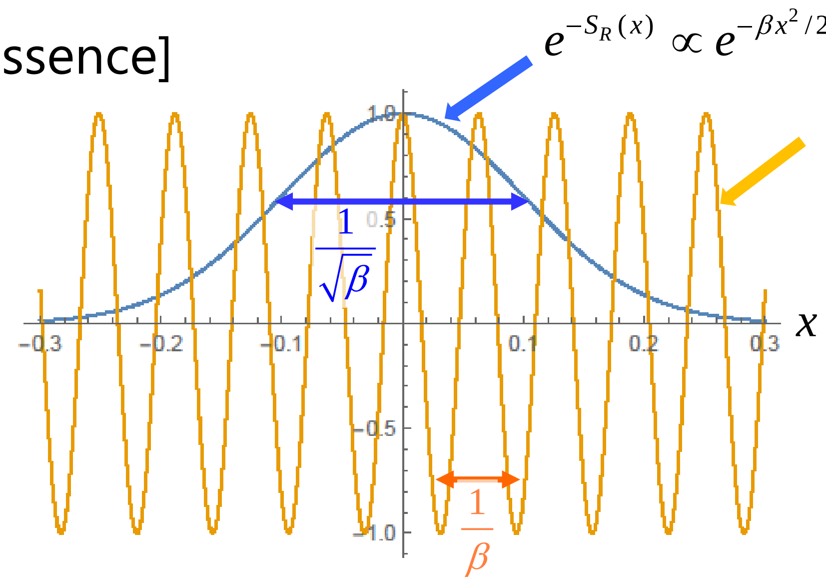
numerically $$\approx \frac{\left(\beta^{-1}-1\right) e^{-\beta/2} \pm O(1/\sqrt{N_{\rm conf}})}{e^{-\beta/2} \pm O(1/\sqrt{N_{\rm conf}})}$$

(NB : The num and the denom are estimated separately.)

⇒ Necessary sample size:

$$1/\sqrt{N_{\rm conf}} < O(e^{-\beta/2}) \quad \Leftrightarrow \quad \boxed{N_{\rm conf} > O(e^{\beta})}$$

[Essence]



In the limit $\beta \to \infty$ $\left(\because 1/\beta \ll 1/\sqrt{\beta}\right)$,

the integration becomes highly oscillatory

Approaches to the sign problem

Various approaches:

(1) Complex Langevin method (CLM) **[Parisi 1983]**

(2) (Generalized) Lefschetz thimble method ((G)LTM) **[Cristoforetti et al. 2012, …] [Alexandru et al. 2015, …]**

(3) … [to be commented later] **[Kashiwa-Mori-Ohnishi 2017, Alexandru et al. 2018] [Di Renzo et al., Ulybyshev et al., …]**

Advantages/disadvantages:

(1) CLM
Pros: fast $\propto O(N)$ (N : DOF)
Cons: "wrong convergence problem" **[Ambjørn-Yang 1985, Aarts et al. 2011, Nagata-Nishimura-Shimasaki 2016]**

(2) LTM
Pros: No wrong convergence problem *iff* only a single thimble is relevant
Cons: Expensive $\propto O(N^3)$ ⇐ Jacobian determinant
Multimodal problem if more than one thimble are relevant (wrong convergence de facto)

⇓

(2′) **TLTM** (Tempered Lefschetz thimble method) **[MF-Umeda 1703.00861, MF-Matsumoto-Umeda 1906.04243]**

"facilitate transitions among thimbles by tempering the system with the flow time"

Pros: Works well even when multi thimbles are relevant
Cons: Expensive $\propto O(N^{3\sim 4})$ ⇐ Jacobian determinant + tempering

Plan

1. Introduction (done)
2. Complex Langevin method (CLM)
3. (Generalized) LTM (GLTM)
4. Tempered LTM (TLTM)
5. Applying the TLTM to the Hubbard model
   - 1D case
   - 2D case
6. Other approaches
7. Conclusion and outlook

Complex Langevin method: basics (1/2)

[Parisi PLB131(1983)393]

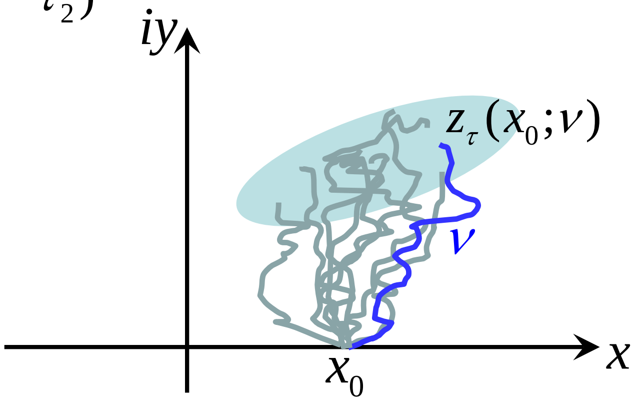
- ν_τ : Gaussian white noise with variance σ^2

$$\langle \nu_{\tau_1}\nu_{\tau_2}\rangle_\nu \left(\equiv \int \prod_\tau \frac{d\nu_\tau}{(2\pi\sigma^2)^{1/2}} e^{-(1/2\sigma^2)\int d\tau\, \nu_\tau^2} \nu_{\tau_1}\nu_{\tau_2} \right) = \sigma^2 \delta(\tau_1 - \tau_2)$$

- Complex Langevin equation:

$$\dot{z}_\tau = \nu_\tau - \frac{\sigma^2}{2} S'(z_\tau) \quad \text{with} \quad z_{\tau=0} = x_0$$

$$\left(z_{n+1} = z_n + \sqrt{\epsilon}\,\nu_n - \frac{\epsilon\sigma^2}{2} S'(z_n) \quad (\tau = n\epsilon) \right)$$



⇨ soln $z_\tau = z_\tau(x_0;\nu)$ for a given $\nu = (\nu_s)$ $(0 \le s \le \tau)$

- Replace x in $\mathcal{O}(x)$ by $z_\tau(x_0;\nu)$ and take the average over ν :

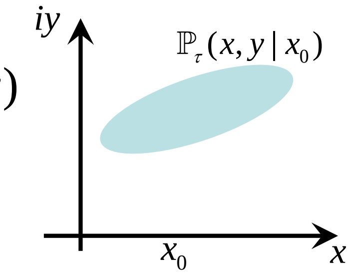
$$\mathcal{O}(x) \to \mathcal{O}\big(z_\tau(x_0;\nu)\big) \to \big\langle \mathcal{O}\big(z_\tau(x_0;\nu)\big)\big\rangle_\nu$$

- The $\tau \to \infty$ limit gives the desired expectation value (under some condition):

$$\boxed{\lim_{\tau\to\infty} \big\langle \mathcal{O}\big(z_\tau(x_0;\nu)\big)\big\rangle_\nu = \big\langle \mathcal{O}(x)\big\rangle_S} \left(= \frac{\int dx\, e^{-S(x)} \mathcal{O}(x)}{\int dx\, e^{-S(x)}} \right) (x_0\text{-independent})$$

Complex Langevin method: basics (2/2)

"proof" **[Aarts-James-Seiler-Stamatescu 1101.3270]**



Introduce a PDF over $\mathbb{C}=\mathbb{R}^2$: $\boxed{\mathbb{P}_\tau(x,y\,|\,x_0)\equiv\langle\delta^2(z-z_\tau(x_0;\nu))\rangle_\nu}$ $(z=x+iy)$

- $\left\langle \mathcal{O}\left(z_\tau(x_0;\nu)\right)\right\rangle_\nu=\int dxdy\;\mathbb{P}_\tau(x,y\,|\,x_0)\,\mathcal{O}(x+iy)$

- $\mathbb{P}_\tau(x,y\,|\,x_0)=e^{-\tau\,\mathbb{H}_{x,y}}\,\delta(x-x_0)\delta(y)$

$$\left(\mathbb{H}_{x,y}\equiv-\frac{\sigma^2}{2}\left[\partial_x\circ\left(\partial_x+S'_R(x,y)\right)+\partial_y\circ S'_I(z)\right]\right)$$

$S'_R(x,y)\equiv\mathrm{Re}\,S'(x+iy),\;\; S'_I(x,y)\equiv\mathrm{Im}\,S'(x+iy)$

$$\therefore\;\left\langle \mathcal{O}\left(z_\tau(x_0;\nu)\right)\right\rangle_\nu\overset{\text{P.I.}}{=}\int dxdy\,\delta(x-x_0)\delta(y)\,e^{-\tau\,\mathbb{H}^T_{x,y}}\,\mathcal{O}(x+iy)$$

Here,
$$\mathbb{H}^T_{x,y}\mathcal{O}(x+iy)=-\frac{\sigma^2}{2}\left[\left(-\partial_x+S'_R(x,y)\right)(-\partial_x)+S'_I(x,y)(-\partial_y)\right]\mathcal{O}(x+iy)$$
$$=-\frac{\sigma^2}{2}\left[\left(-\partial_z+S'_R(x,y)\right)(-\partial_z)+S'_I(x,y)(-i\partial_z)\right]\mathcal{O}(z)$$
$$=-\frac{\sigma^2}{2}\Big[-\partial_z+\underbrace{S'_R(x,y)+iS'_I(x,y)}_{=\,S'(z)}\Big](-\partial_z)\mathcal{O}(z)\equiv H^T_z\,\mathcal{O}(z)$$

$$\left(H_z=-\frac{\sigma^2}{2}\partial_z\circ[\partial_z+S'(z)]\right)$$

$$\therefore\;\left\langle \mathcal{O}\left(z_\tau(x_0;\nu)\right)\right\rangle_\nu=\int dxdy\,\delta(x-x_0)\delta(y)\,e^{-\tau H^T_z}\,\mathcal{O}(x+iy)$$
$$=\int dx\,\delta(x-x_0)\;e^{-\tau H^T_x}\,\mathcal{O}(x)\overset{\text{P.I.}}{=}\int dx\Big[\underbrace{e^{-\tau H_x}\,\delta(x-x_0)}_{\equiv\,P_\tau(x\,|\,x_0)}\Big]\mathcal{O}(x)$$

$P_\tau(x\,|\,x_0)$ satisfies $\dot{P}_\tau(x\,|\,x_0)=-H_xP_\tau(x\,|\,x_0)=\dfrac{\sigma^2}{2}\partial_x\circ[\partial_x+S'(x)]P_\tau(x\,|\,x_0)$

$$\boxed{P_\tau(x\,|\,x_0)\xrightarrow{\tau\to\infty}\frac{1}{Z}\int dx\,e^{-S(x)}}$$

$$\lim_{\tau\to\infty}\left\langle \mathcal{O}\left(z_\tau(x_0;\nu)\right)\right\rangle_\nu=\lim_{\tau\to\infty}\int dxdy\,\mathbb{P}_\tau(x,y\,|\,x_0)\,\mathcal{O}(x+iy)$$
$$=\lim_{\tau\to\infty}\int dx\,P_\tau(x\,|\,x_0)\,\mathcal{O}(x)=\langle\mathcal{O}(x)\rangle_S\;\blacksquare$$

Complex Langevin method: wrong convergence

In order for the partial integration and $e^{-\tau \mathbb{H}_{x,y}^T}$ to be meaningful,

$\mathbb{P}_\tau(x, y \mid x_0)$ should $\begin{cases} \text{not be spread out largely in } |y| \to \infty \text{ direction} \\ \text{not have a significant support around zeros of } e^{-S(z)} \end{cases}$

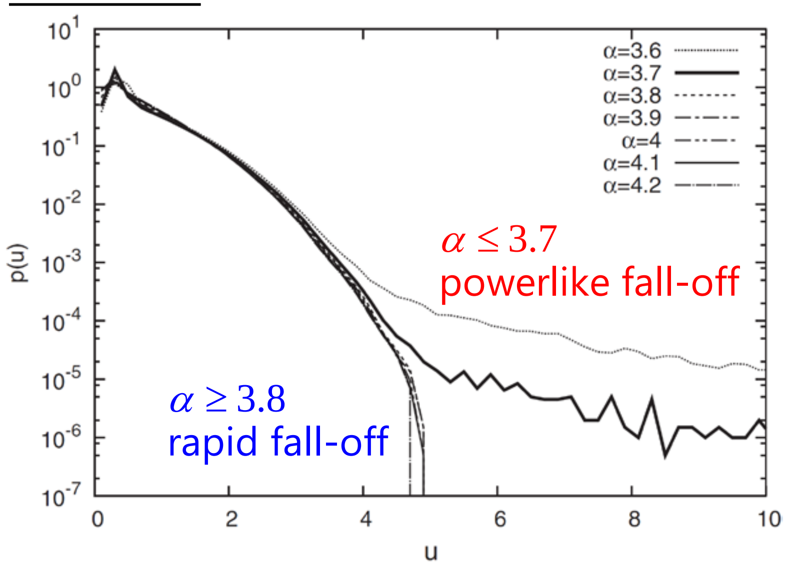
Otherwise, the limit gives a wrong result. **[Aarts-James-Seiler-Stamatescu 1101.3270]**

<u>Criterion</u> **[Nagata-Nishimura-Shimasaki 1606.07627, PRD94 (2016) 114515]**

The histogram of $|S'(x+iy)| = \sqrt{S_R'(x,y)^2 + S_I'(x,y)^2}$

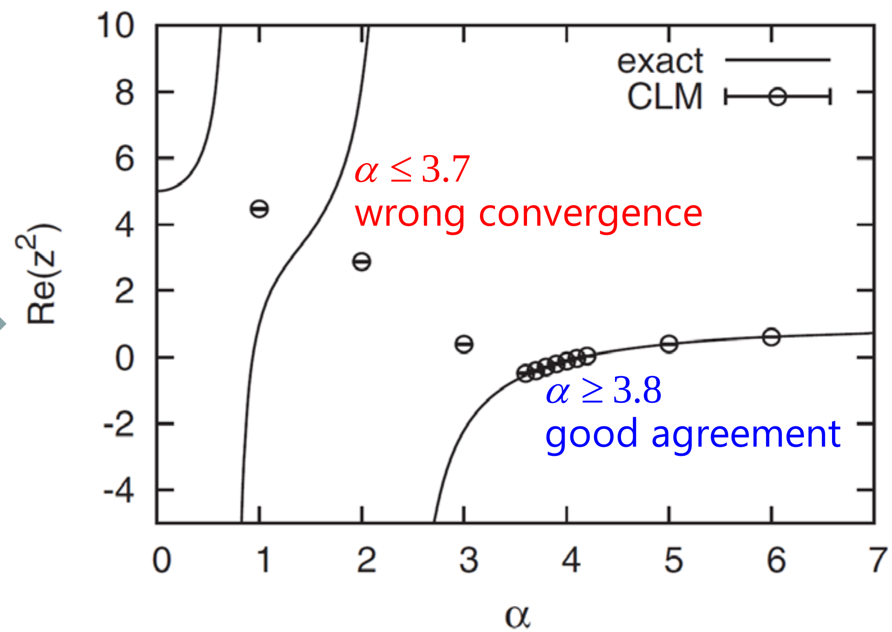
must decrease rapidly (at least exponentially) at large values

<u>Example</u>: $S(x) = (x+i\alpha)^4 e^{-x^2/2}$ **[Nagata-Nishimura-Shimasaki PRD94 (2016) 114515]**



$u \equiv |S'(z)| \quad (z = x+iy)$

$p(u)$: normalized histogram



CLM: attempts to solve the wrong convergence

Aim: reduce the effects from dangerous configurations

(1) configurations far from the original integration region $\mathbb{R}^N$

"excursion problem"

⇒ gauge cooling : **[Seiler-Sexty-Stamatescu 1211.3709]**

repeatedly make "gauge transformations" (if possible)

to send the variables near $\mathbb{R}^N$

(2) configurations close to zeros of $e^{-S(z)}$ "singular drift problem"

⇒ reweighting : **[Bloch 1701.00986 , Bloch et al. 1701.01298]**

Use a parameter with which CLM works

(assuming an enough overlap):

$$e^{-S(x;\alpha)} \rightarrow e^{-S(x;\beta)} \times \underbrace{e^{-S(x;\alpha)+S(x;\beta)}}_{\text{regarded as a part of observable}}$$

deformation : **[Ito-Nishimura 1710.07929]**

Add a parameter s.t. CLM works: $S(x) \rightarrow S(x;\alpha)$,

then take a limit $\alpha \rightarrow 0$

3. (Generalized) Lefschetz thimble method (GLTM)

[Cristoforetti et al. 1205.3996, 1303.7204, 1308.0233]
[Fujii-Honda-Kato-Kikukawa-Komatsu-Sano 1309.4371]
[Alexandru et al. 1512.08764]

Lefschetz thimble method (1/2)

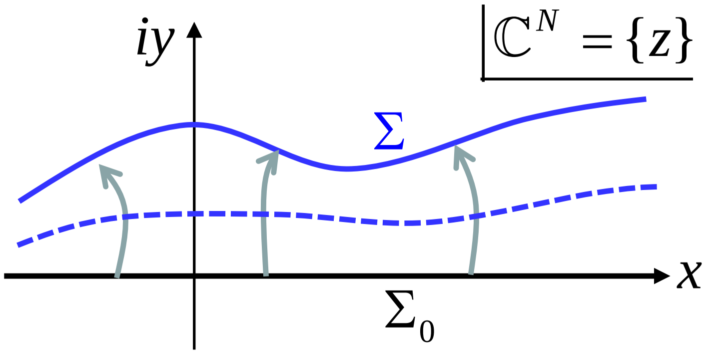
Complexify the variable: $x = (x^i) \in \mathbb{R}^N \;\Rightarrow\; z = (z^i = x^i + iy^i) \in \mathbb{C}^N$

Assumption: $e^{-S(z)}$, $e^{-S(z)}\mathcal{O}(z)$: entire functions over $\mathbb{C}^N$

⇓ Cauchy's theorem

Integral does not change under continuous deformations of the integration region from $\Sigma_0 = \mathbb{R}^N$ to $\Sigma \subset \mathbb{C}^N$ (with the boundary at infinity $|x| \to \infty$ kept fixed) :

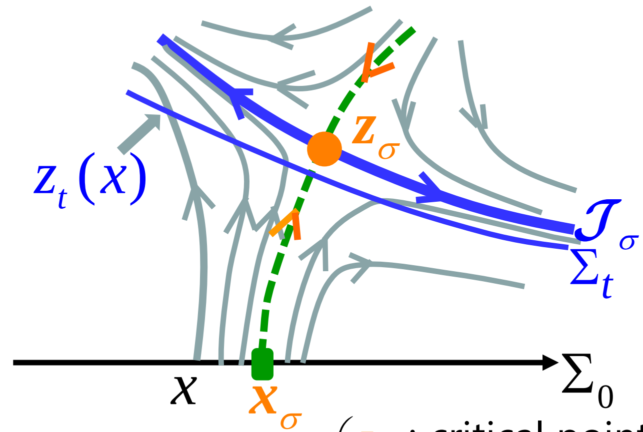
$$\langle \mathcal{O}(x) \rangle_S \equiv \frac{\int_{\Sigma_0} dx\, e^{-S(x)}\,\mathcal{O}(x)}{\int_{\Sigma_0} dx\, e^{-S(x)}} = \frac{\int_{\Sigma} dz\, e^{-S(z)}\,\mathcal{O}(z)}{\int_{\Sigma} dz\, e^{-S(z)}}$$



severe sign problem

sign problem will get much reduced if $\mathrm{Im}\, S(z)$ is almost constant on Σ

Lefschetz thimble method (2/2)



Prescription:

antiholomorphic gradient flow $\dot{z}_t^i = \overline{\partial_i S(z_t)}$ with $z_{t=0}^i = x^i$

(z_σ : critical point ($\partial_i S(z_\sigma) = 0$))

Property: $[S(z_t)]^\cdot = \partial_i S(z_t)\, \dot{z}_t^i = |\partial_i S(z_t)|^2 \geq 0$

⇒ $[\mathrm{Re}\, S(z_t)]^\cdot \geq 0$: real part always increases along the flow

$[\mathrm{Im}\, S(z_t)]^\cdot = 0$: imaginary part is kept fixed ″

⇒ In $t \to \infty$, Σ_t approaches a union of Lefschetz thimbles: $\Sigma_t \to \bigcup_\sigma \mathcal{J}_\sigma$

(on each of which $\mathrm{Im}\, S(z)$ is constant)

Expectation value:

$$\langle \mathcal{O}(x) \rangle_S \equiv \frac{\int_{\Sigma_0} dx\, e^{-S(x)} \mathcal{O}(x)}{\int_{\Sigma_0} dx\, e^{-S(x)}} = \frac{\int_{\Sigma_t} dz_t\, e^{-S(z_t)} \mathcal{O}(z_t)}{\int_{\Sigma_t} dz_t\, e^{-S(z_t)}} = \frac{\int_{\Sigma_0} dx\, \det(\partial z_t^i(x)/\partial x^j)\, e^{-S(z_t(x))}\, \mathcal{O}(z_t(x))}{\int_{\Sigma_0} dx\, \det(\partial z_t^i(x)/\partial x^j)\, e^{-S(z_t(x))}}$$

$$= \frac{\left\langle e^{i\theta_t(x)} \mathcal{O}(z_t(x)) \right\rangle_{S_t^{\mathrm{eff}}}}{\left\langle e^{i\theta_t(x)} \right\rangle_{S_t^{\mathrm{eff}}}}$$

$$\det(\partial z_t^i(x)/\partial x^j)\, e^{-S(z_t(x))} \equiv e^{-S_t^{\mathrm{eff}}(x)} e^{i\theta_t(x)}$$

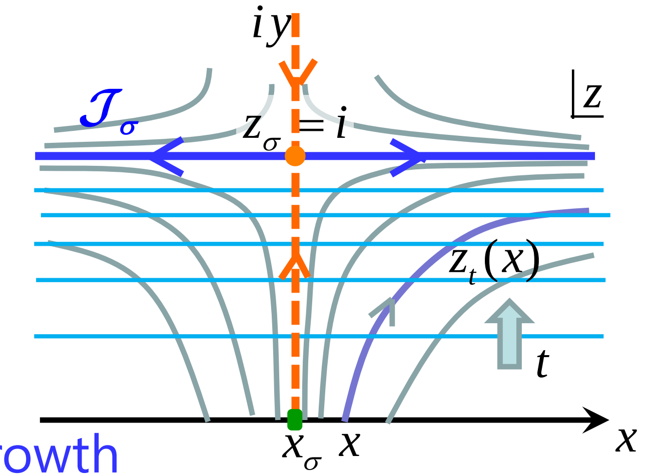
$$e^{-S_t^{\mathrm{eff}}(x)} \equiv e^{-\mathrm{Re}\, S(z_t(x))} \left| \det(\partial z_t^i(x)/\partial x^j) \right|$$

$$e^{i\theta_t(x)} \equiv e^{-i\, \mathrm{Im}\, S(z_t(x)) + i \arg \det(\partial z_t^i(x)/\partial x^j)}$$

Example: Gaussian

Gradient flow: $\left[S(z) = (\beta/2)(z-i)^2 \right]$

$$\dot z_t = \dot x_t + i\,\dot y_t = \overline{S'(z_t)} \Leftrightarrow \begin{cases} \dot x_t = \beta x \\ \dot y = -\beta(y_t - 1) \end{cases} \text{ with } \begin{cases} x_{t=0} = x \\ y_{t=0} = 0 \end{cases}$$



$$\begin{cases} z_t(x) = x e^{\beta t} + i(1 - e^{-\beta t}) \\ J_t(x) = \dfrac{dz_t(x)}{dx} = e^{\beta t} \end{cases} \Rightarrow \begin{cases} S_t^{\text{eff}}(x) = \dfrac{1}{2}\beta e^{2\beta t}(x^2 - e^{-4\beta t}) - \beta t \\ \theta_t(x) = \beta x \end{cases}$$

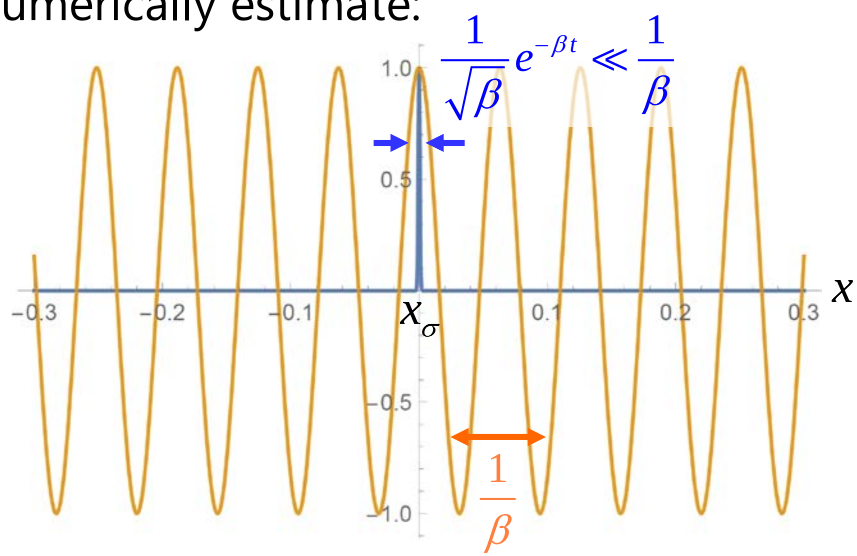
exponential growth of coefficient

$$\Rightarrow \begin{cases} \left\langle e^{i\theta_t(x)} z_t^2(x) \right\rangle_{S_t^{\text{eff}}} = e^{-(\beta/2)e^{-2\beta t}} \left(\beta^{-1} - 1\right) \\ \left\langle e^{i\theta_t(x)} \right\rangle_{S_t^{\text{eff}}} = e^{-(\beta/2)e^{-2\beta t}} \left(= O(1) \text{ if } \beta e^{-2\beta t} \ll 1 \left(\Leftrightarrow e^{-\beta t} \ll \dfrac{1}{\sqrt{\beta}} \right) \right) \end{cases}$$

Taking a large T s.t. $e^{-\beta T} \ll \dfrac{1}{\sqrt{\beta}}$, we can numerically estimate:

$$\left\langle x^2 \right\rangle_S = \frac{\left\langle e^{i\theta_T(x)} z_T^2(x) \right\rangle_{S_T^{\text{eff}}}}{\left\langle e^{i\theta_T(x)} \right\rangle_{S_T^{\text{eff}}}} = \frac{e^{-(\beta/2)e^{-2\beta T}} \left(\beta^{-1} - 1\right)}{e^{-(\beta/2)e^{-2\beta T}}} = \beta^{-1} - 1$$

(no small numbers appears!)

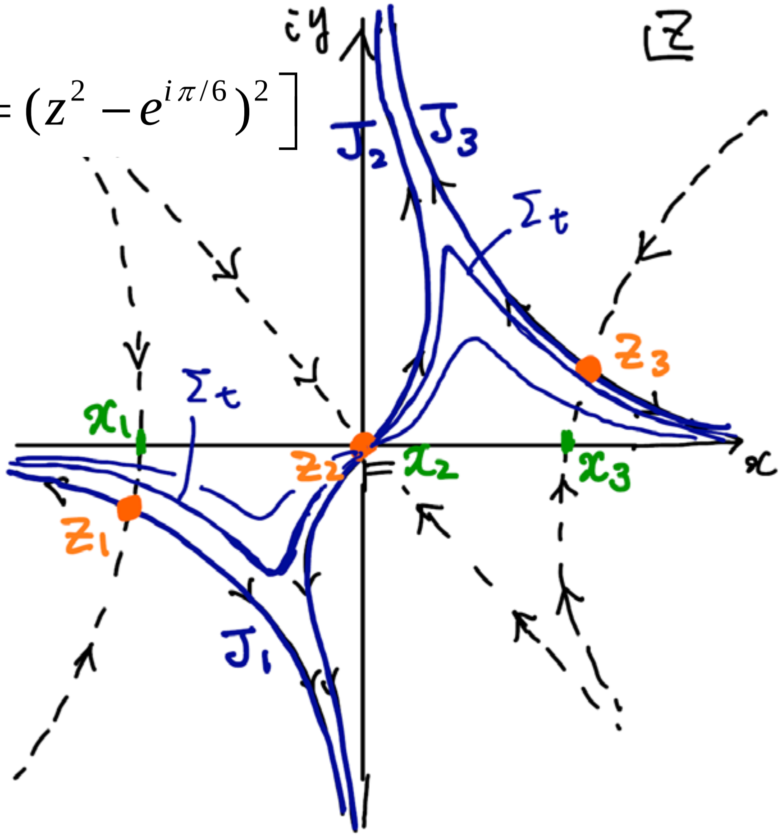


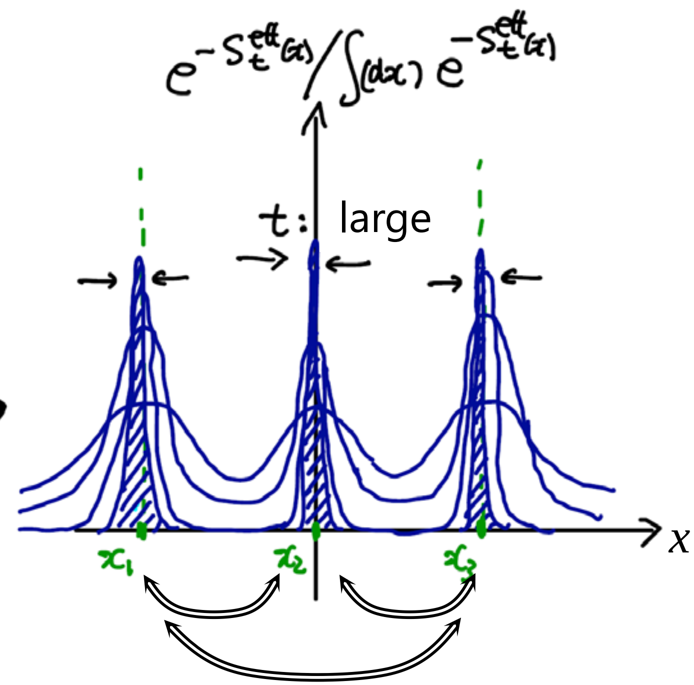
Multimodal problem and Generalized LTM (1/2)

Flow time t needs to be large enough to solve the sign problem

However, this introduces a new problem "multimodal problem"

$$\left[S(z) = (z^2 - e^{i\pi/6})^2 \right]$$





transitions among thimbles become indefinitely difficult as t increases

Dilemma between the sign problem and the multimodal problem

(for small t) (for large t)

Multimodal problem and Generalized LTM (2/2)

Proposal in Generalized LTM: [Alexandru-Basar-Bedaque-Ridgway-Warrington 1512.08764]

Choose a middle value of T s.t. it is large enough for the sign problem but at the same time is not too large for the multimodal problem

flow time $(= T)$	small	medium	large
sign problem	NG	△	**OK**
multimodal problem	**OK**	△	NG

However, the existence of such T is not obvious a priori

Even when it exists, a very fine tuning will be needed

Tempered LTM: [MF-Umeda 1703.00861] (cf. [Alexandru-Basar-Bedaque-Warrington 1703.02414])

Implement a tempering method by using the flow time *t* as a dynamical variable

flow time $(= T)$	small	medium	large
sign problem	NG	**OK**	**OK**
multimodal problem	**OK**	**OK**	**OK**

no fine tuning needed!

4. Tempered Lefschetz thimble method (TLTM)

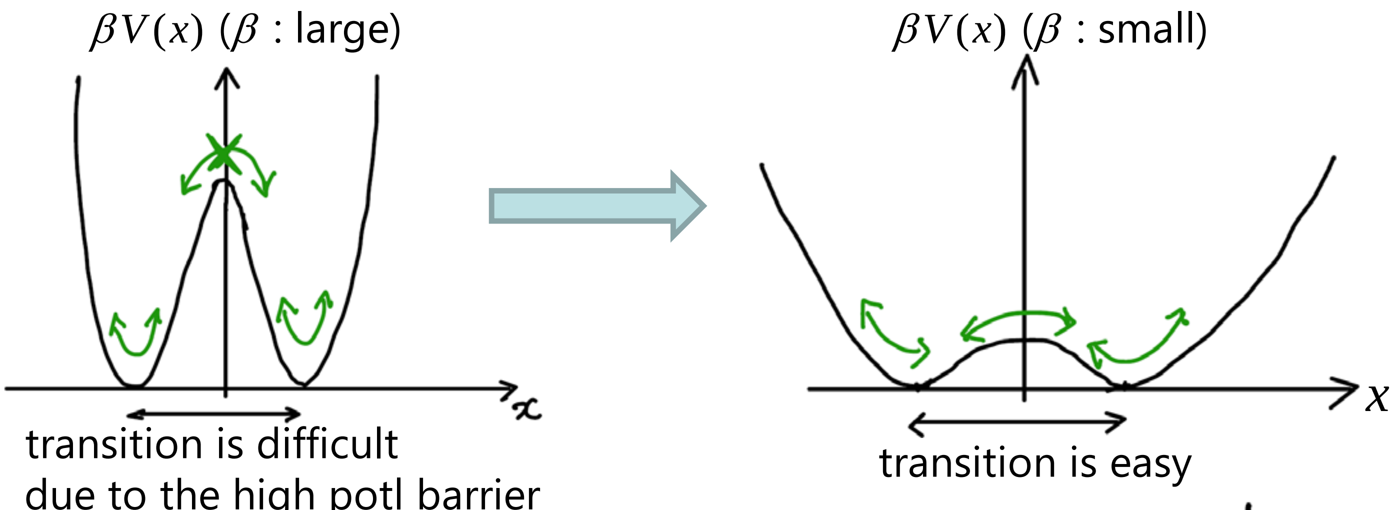
[MF-Umeda 1703.00861]
[MF-Matsumoto-Umeda 1906.04243]

Idea of tempering

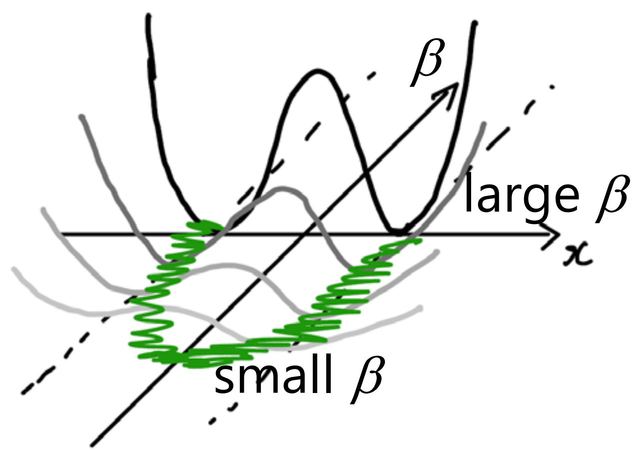
[Marinari-Parisi Europhys.Lett.19(1992)451]

Suppose that the action $S(x;\beta)$ gives a multimodal distribution for the value of β in our main concern (e.g. $S(x;\beta) = \beta V(x)$ with $\beta \gg 1$)

It often happens that multimodality disappears if we take a different value of β (e.g. for $\beta \ll 1$)



In the tempering method,
we extend the config space from $\{x\}$ to $\{(x,\beta)\}$.
Then, transitions between two modes become easy by passing through configs with smaller β



Tempered LTM (1/3)

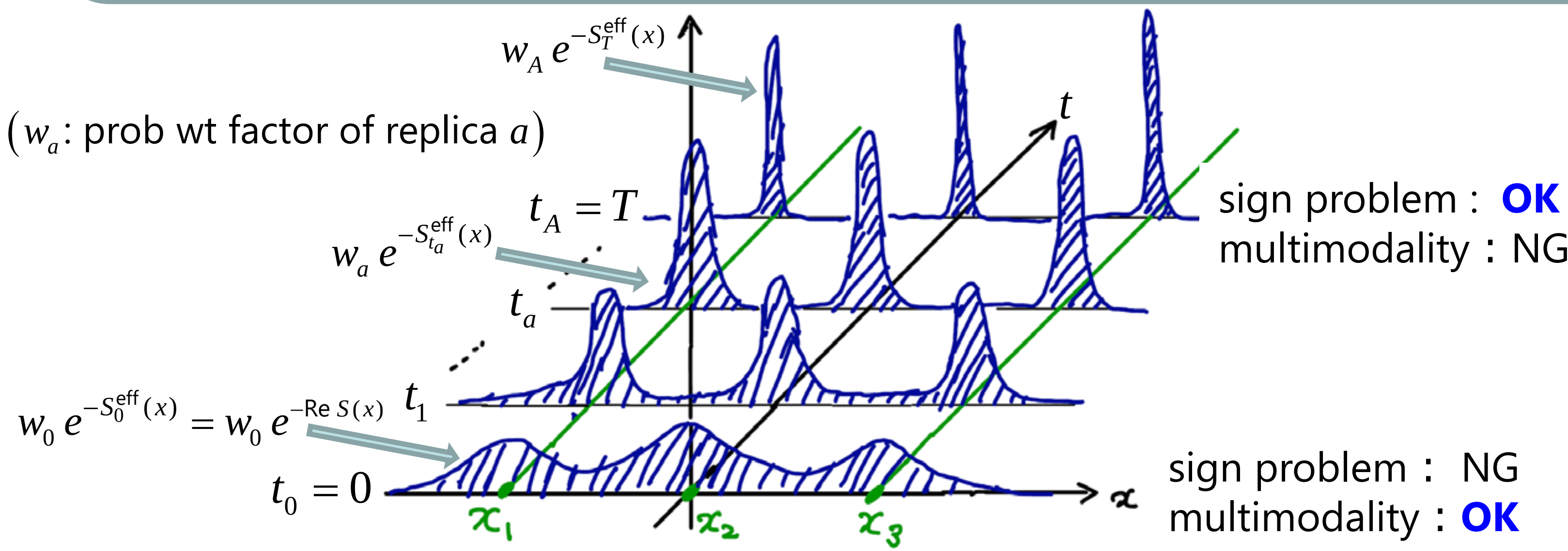
[MF-Umeda 1703.00861]

Algorithm of TLTM

(1) Introduce copies of config space labeled by a finite set of flow times

$$\mathcal{A} = \{t_a\}\ (a = 0, 1, \ldots, A) \quad \left(t_0 = 0 < t_1 < t_2 < \cdots < t_A = T\right),$$

and construct a Markov chain that drives the enlarged system to global equilibrium



Tempered LTM (1/3)

[MF-Umeda 1703.00861]

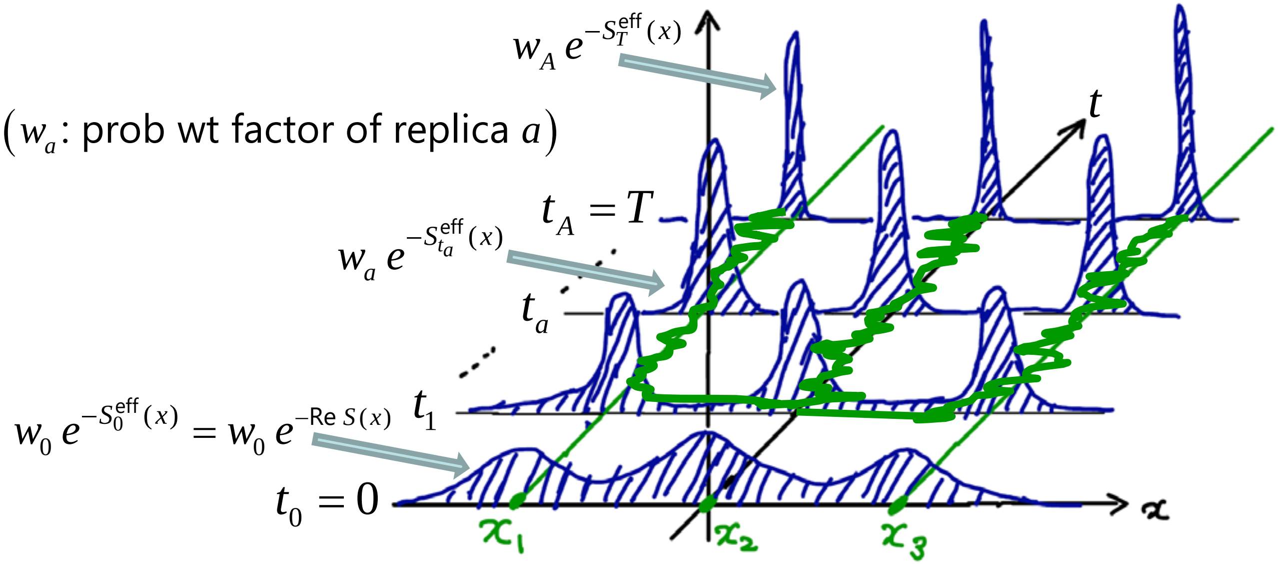
Algorithm of TLTM

(1) Introduce copies of config space labeled by a finite set of flow times

$$\mathcal{A} = \{t_a\}\ (a = 0, 1, \ldots, A) \quad \left(t_0 = 0 < t_1 < t_2 < \cdots < t_A = T\right),$$

and construct a Markov chain that drives the enlarged system to global equilibrium

$\left(w_a\text{: prob wt factor of replica } a\right)$



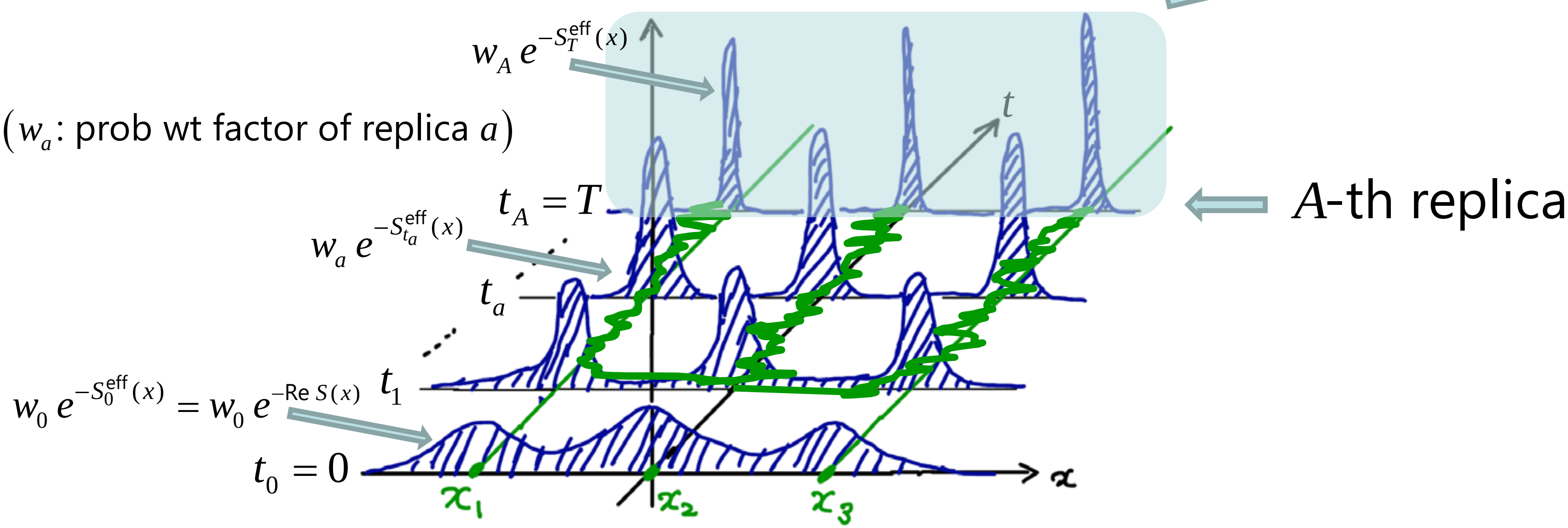
Tempered LTM (2/3)

[MF-Umeda 1703.00861]

Algorithm of TLTM

(2) After the enlarged system is relaxed to global equilibrium, evaluate the expectation value by using the subsample at $t_A = T$ $(a = A)$

$\left(w_a \text{: prob wt factor of replica } a \right)$

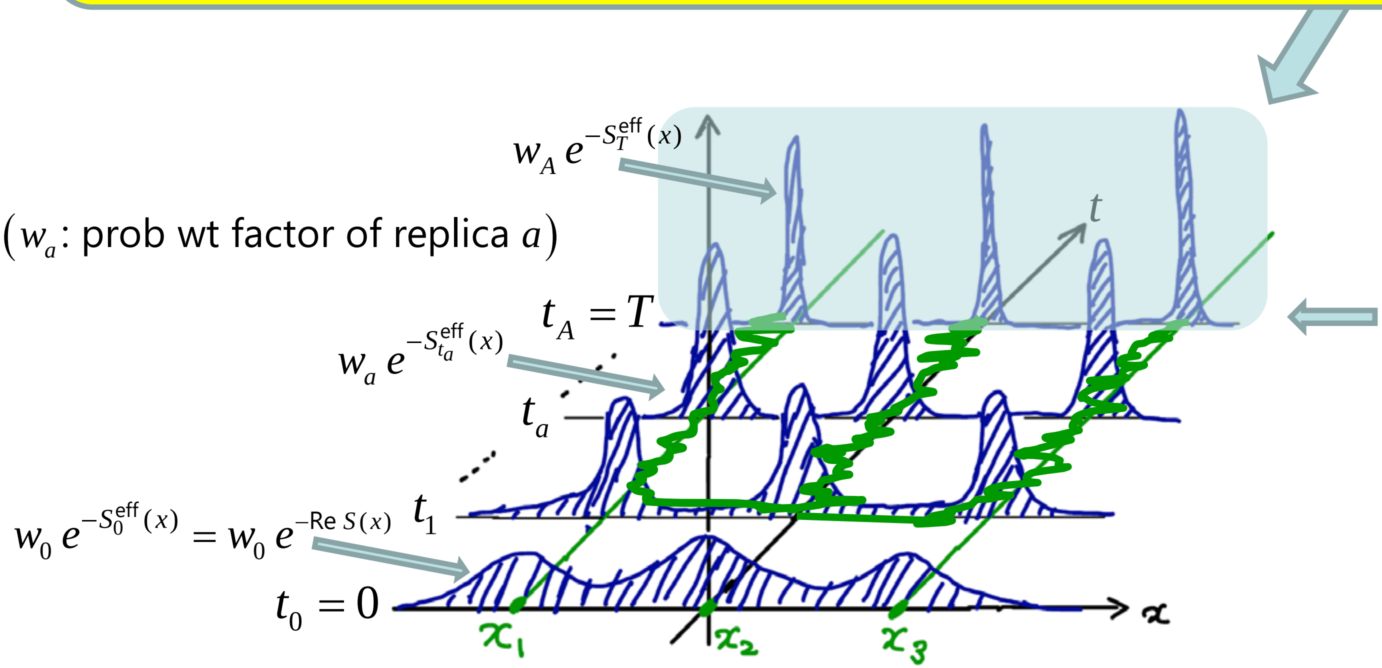


Tempered LTM (2/3)

Algorithm of TLTM

[MF-Umeda 1703.00861]

(2) After the enlarged system is relaxed to global equilibrium, evaluate the expectation value by using the subsample at $t_A = T$ $(a = A)$



NB: various tempering methods $\left(\mathcal{M} \equiv \{x\}\right.$: original config space$\left.\right)$

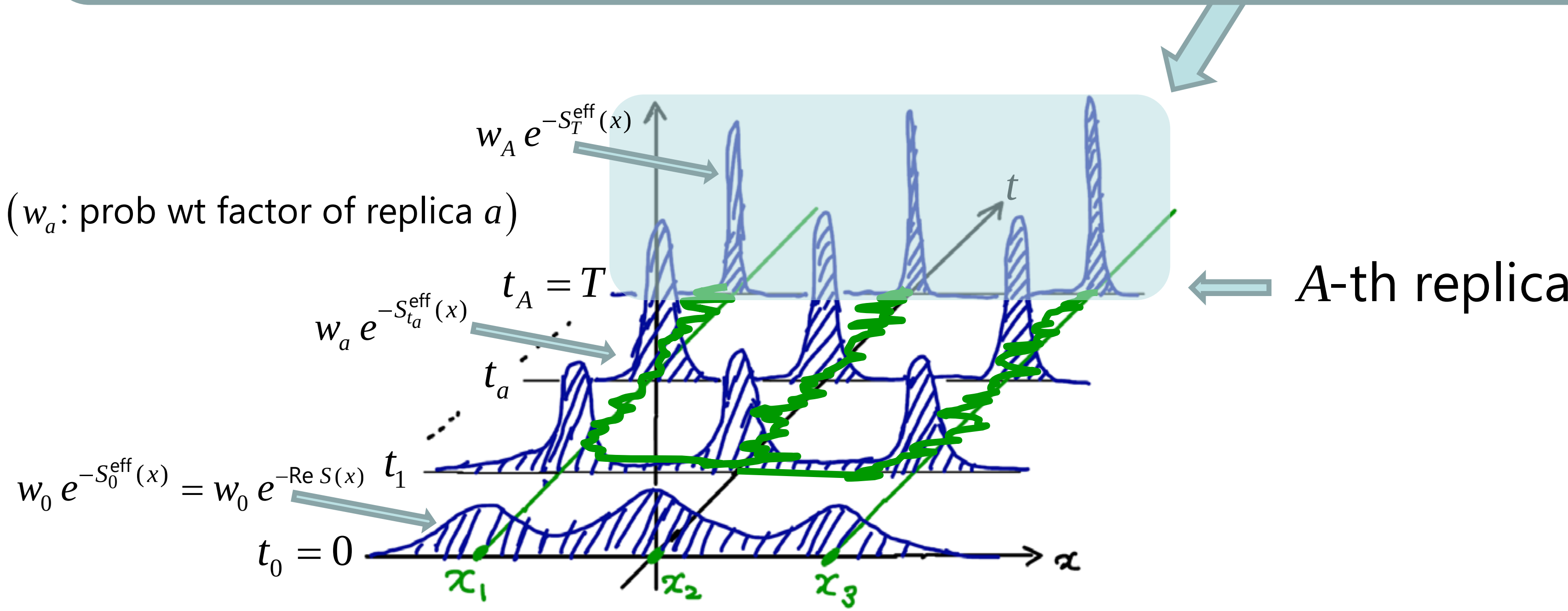
- simulated tempering : enlarged system $\Leftrightarrow$ $\mathcal{M} \times \mathcal{A} = \{(x, t_a)\}$ (△ [tedious task to detemine the weights w_a])
  [Marinari-Parisi 1992]
- parallel tempering (replica exchange MCMC) : enlarged system $\Leftrightarrow$ $\overbrace{\mathcal{M} \times \mathcal{M} \times \cdots \times \mathcal{M}}^{A+1} = \{(x_0, x_1, \ldots, x_A)\}$ (◯)
  most of relevant steps can be done in parallel processes
  [Swendsen-Wang 1986, Geyer 1991, Nemoto-Hukushima 1996]

Tempered LTM (2/3)

Algorithm of TLTM

[MF-Umeda 1703.00861]

(2) After the enlarged system is relaxed to global equilibrium, evaluate the expectation value by using the subsample at $t_A = T$ $(a = A)$



NB: various tempering methods $\left(\mathcal{M} \equiv \{x\}\right.$: original config space$\left.\right)$

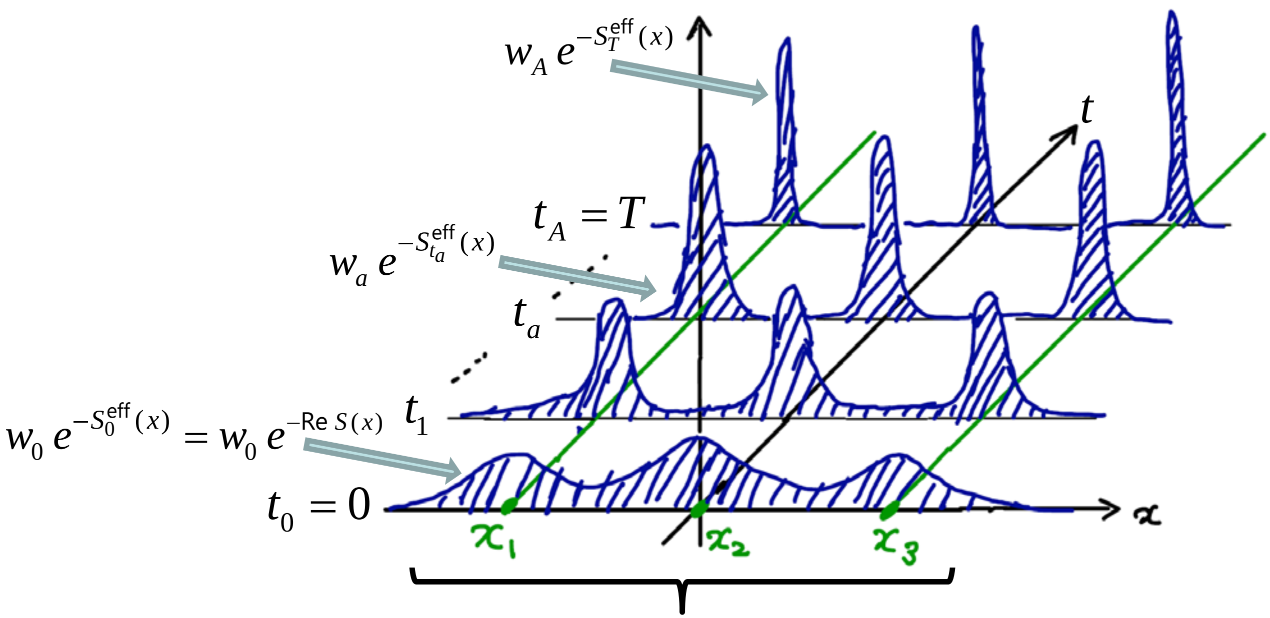
- simulated tempering : enlarged system $\Leftrightarrow$ $\mathcal{M} \times \mathcal{A} = \{(x, t_a)\}$ $\left(\triangle \left[\text{tedious task to detemine the weights } w_a\right]\right)$
  [Marinari-Parisi 1992]
- parallel tempering (replica exchange MCMC) : enlarged system $\Leftrightarrow$ $\overbrace{\mathcal{M} \times \mathcal{M} \times \cdots \times \mathcal{M}}^{A+1} = \{(x_0, x_1, \ldots, x_A)\}$ (☺)
  [Swendsen-Wang 1986, Geyer 1991, Nemoto-Hukushima 1996]
  most of relevant steps can be done in parallel processes

Tempered LTM (3/3)

[MF-Umeda 1703.00861, MF-Matsumoto-Umeda 1906.04243]

Important points in TLTM:

(1) **NO "tiny overlap problem" in TLTM**



Distribution functions have peaks at the same positions x_σ for varying tempering parameter (which is t in our case)

⇨ We can expect significant overlap between adjacent replicas!

(2) **The growth of computational cost due to the tempering can be compensated by the increase of parallel processes**

Example: (0+1)-dim Massive Thirring model (1/3)

Lorentzian action (dim reduction of (1+1)D model):

[Pawlowski-Zielinski 1302.1622, 1402.6042, Fujii-Kamata-Kikukawa 1509.08176]

$$S_M = \int dt \left[i\bar{\psi}\gamma^0\partial_0\psi - m\bar{\psi}\psi - \frac{g^2}{2}(\bar{\psi}\gamma^0\psi)^2 \right] \quad \left((\gamma^0)^2 = 1_2, \quad \gamma^{0\dagger} = \gamma^0 \right)$$

bosonization + discretization

Grand partition function $Z_{\beta,\mu} = \mathrm{tr}\ \mathrm{e}^{-\beta(H-\mu Q)}$:

$$Z_{\beta,\mu} = \int_{\mathrm{PBC}} (d\phi)\, e^{-S(\phi)}$$

with

$$(d\phi) = \prod_{n=1}^{N} \frac{d\phi_n}{2\pi}, \quad e^{-S(\phi)} = \det D(\phi) \exp\left[\frac{-1}{2g^2} \sum_{n=1}^{N} (1-\cos\phi_n) \right]$$

$$D_{nn'}(\phi) = \frac{1}{2}\left(e^{i\phi_n+\mu}\delta_{n+1,n'} - e^{-(i\phi_n+\mu)}\delta_{n-1,n'} - e^{i\phi_N+\mu}\delta_{n,N}\delta_{n',1} + e^{-(i\phi_N+\mu)}\delta_{n,1}\delta_{n',N} \right) + m\,\delta_{n,n'}$$

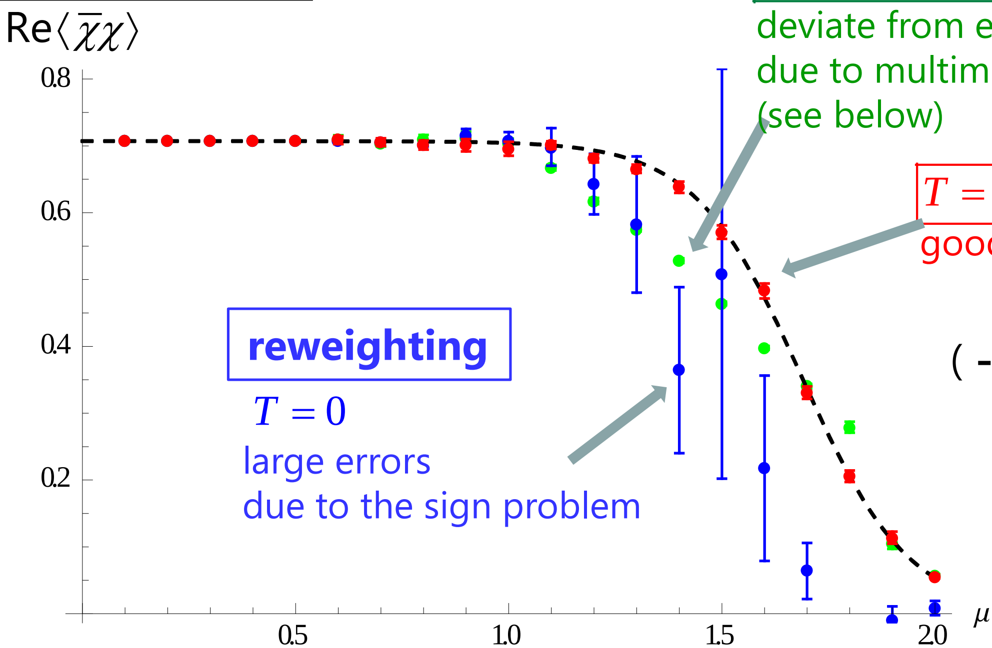
One can show $\boxed{[\det D(\phi;\mu)]^* = \det D(\phi;-\mu)}$ (thus, $\det D \notin \mathbb{R}$ for $\mu \in \mathbb{R}$)

Sign problem will arise when N is very large

Example: (0+1)-dim Massive Thirring model (2/3)

[MF-Umeda 1703.00861]

Chiral condensate $\langle \bar{\chi}\chi \rangle$



$T = 2$ (w/o temp)
deviate from exact values due to multimodality (see below)

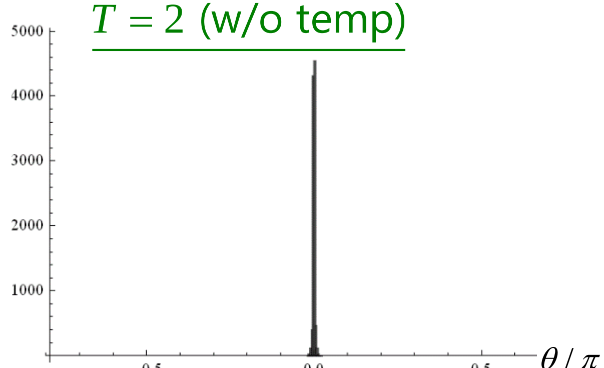
$T = 2$ (w/ temp)
good agreement

reweighting
$T = 0$
large errors due to the sign problem

(------ : exact values)

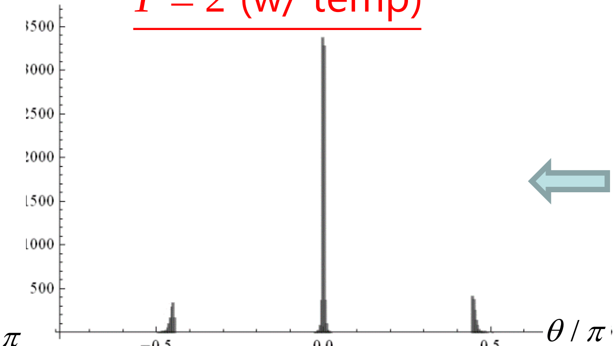
Confirmation of the resolution of multimodality

$T = 2$ (w/o temp)



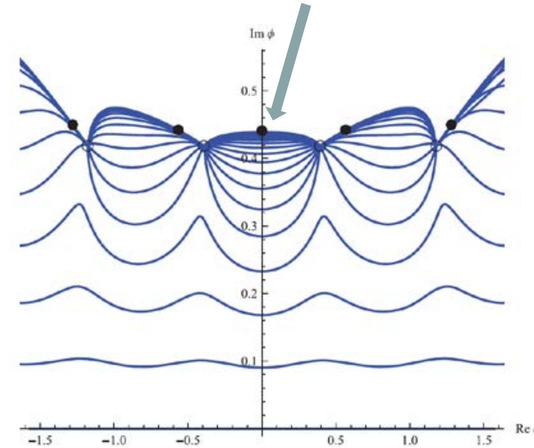
dominated by a single thimble

$T = 2$ (w/ temp)



contributed by multiple thimbles (i.e. config space is well explored)

dominant thimble



Example: (0+1)-dim Massive Thirring model (3/3)

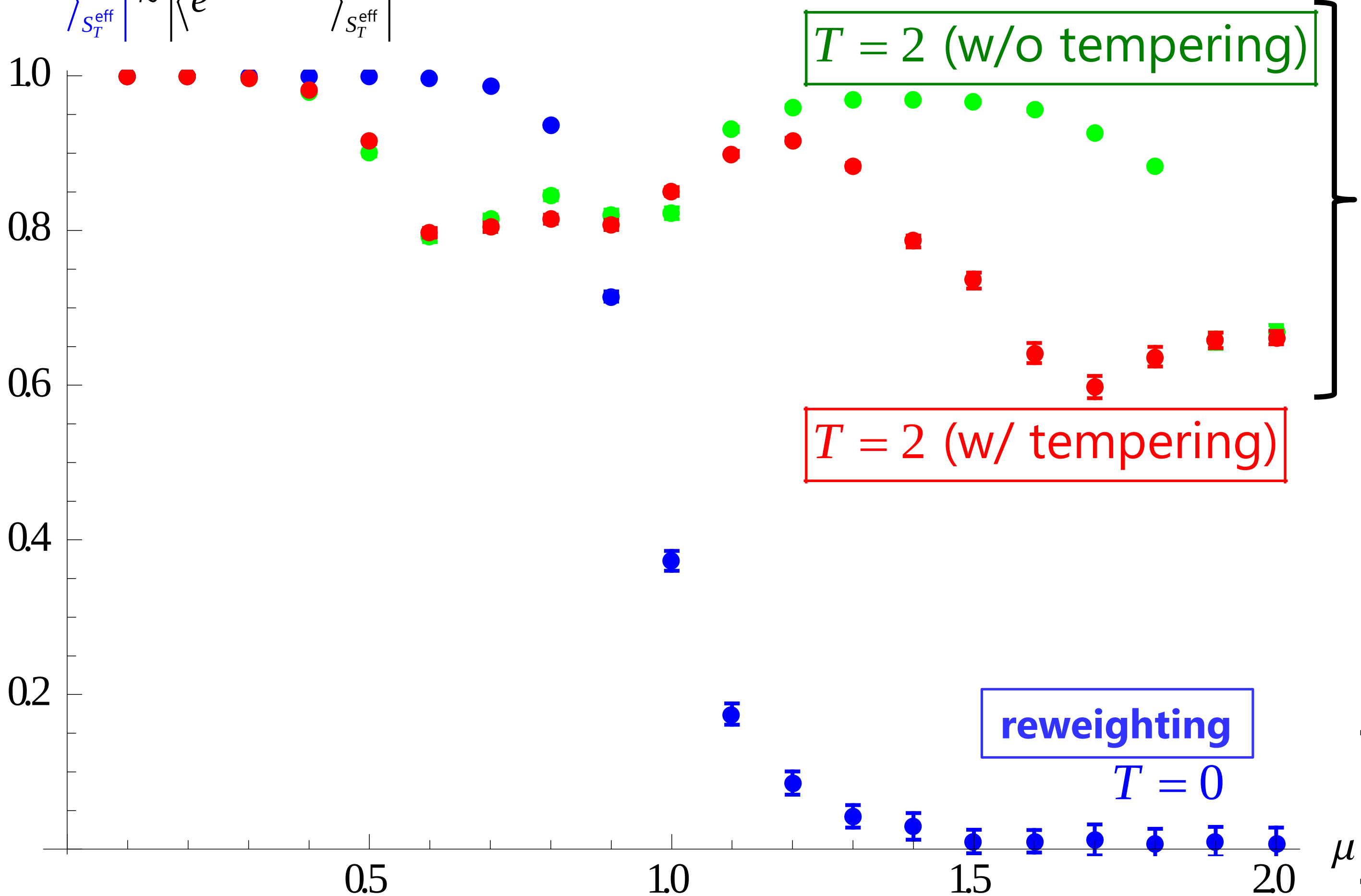
[MF-Umeda 1703.00861]

Confirmation of the resolution of sign problem

$$\left(\langle \mathcal{O}(\phi) \rangle = \frac{\langle e^{i\theta_T(\phi)}\, \mathcal{O}(\phi) \rangle_{S^{\mathrm{eff}}_{t_a}}}{\langle e^{i\theta_T(\phi)} \rangle_{S^{\mathrm{eff}}_T}} \right)$$

sign average

$$\left| \langle e^{i\theta_T(\phi)} \rangle_{S^{\mathrm{eff}}_T} \right| \sim \left| \langle e^{-iS_I(z_T(\phi))} \rangle_{S^{\mathrm{eff}}_T} \right|$$



no sign problem at $T = 2$

(NB: sign avarage $\left| \langle e^{i\theta_T(\phi)} \rangle_{S^{\mathrm{eff}}_T} \right|$ is smaller for the right sampling)

sign problem surely exists for the original action ($T = 0$)

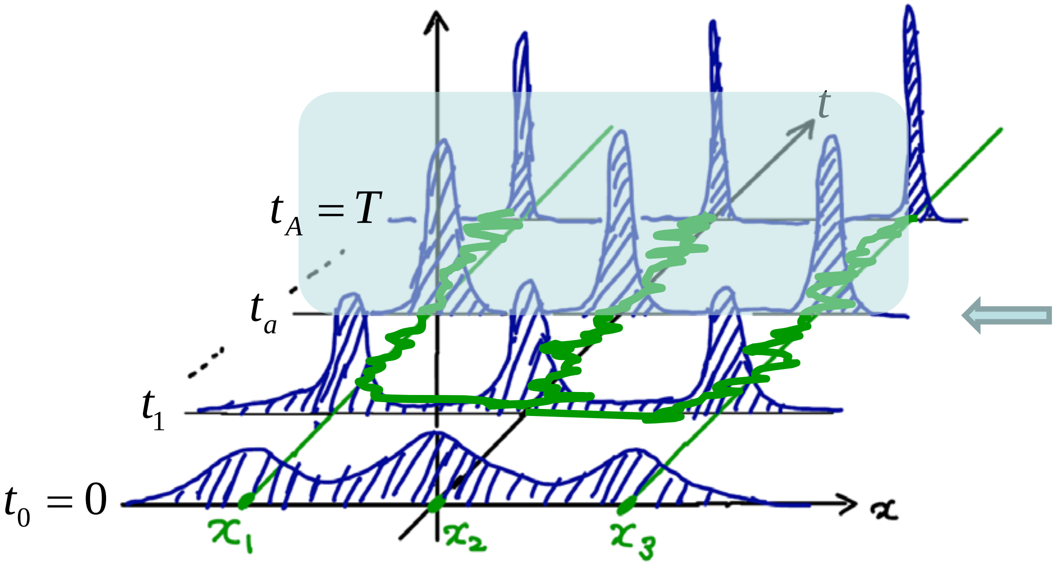
We actually can go further…

[MF-Matsumoto-Umeda 1906.04243]

Consider the estimates of $\langle \mathcal{O} \rangle_S$ at various flow times t_a:

$$\langle \mathcal{O} \rangle_S = \frac{\langle e^{i\theta_{t_a}(x)} \mathcal{O}(z_{t_a}(x)) \rangle_{S^{\mathrm{eff}}_{t_a}}}{\langle e^{i\theta_{t_a}(x)} \rangle_{S^{\mathrm{eff}}_{t_a}}} \approx \frac{\sum_{k=1}^{N_{\mathrm{conf}}} e^{i\theta_{t_a}(x^{(k)})} \mathcal{O}(z_{t_a}(x^{(k)}))}{\sum_{k=1}^{N_{\mathrm{conf}}} e^{i\theta_{t_a}(x^{(k)})}} \equiv \overline{\mathcal{O}}_a \quad (a = 0, 1, \ldots, A)$$

Here the estimation on the RHS is made by using the subsample at t_a:



a-th replica

We actually can go further...

[MF-Matsumoto-Umeda 1906.04243]

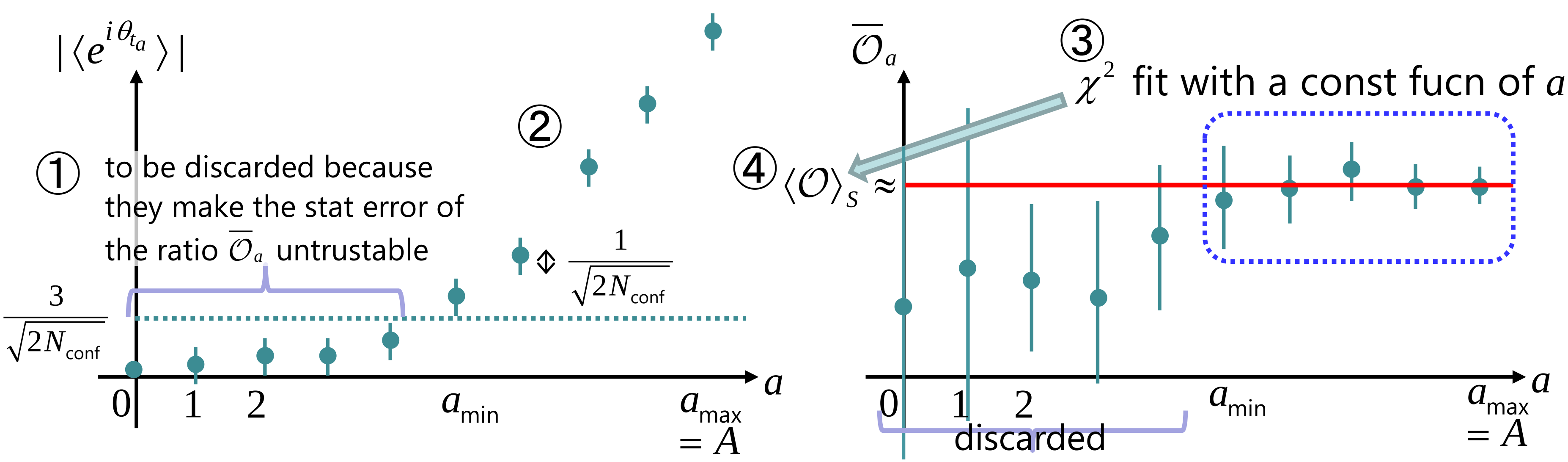
Consider the estimates of $\langle \mathcal{O} \rangle_S$ at various flow times t_a:

$$\langle \mathcal{O} \rangle_S = \frac{\langle e^{i\theta_{t_a}(x)} \mathcal{O}(z_{t_a}(x)) \rangle_{S^{\mathrm{eff}}_{t_a}}}{\langle e^{i\theta_{t_a}(x)} \rangle_{S^{\mathrm{eff}}_{t_a}}} \approx \frac{\sum_{k=1}^{N_{\mathrm{conf}}} e^{i\theta_{t_a}(x^{(k)})} \mathcal{O}(z_{t_a}(x^{(k)}))}{\sum_{k=1}^{N_{\mathrm{conf}}} e^{i\theta_{t_a}(x^{(k)})}} \equiv \overline{\mathcal{O}}_a \quad (a = 0,1,\ldots,A)$$

The LHS must be independent of a due to Cauchy's theorem

The RHS must be the same for all a's within the statistical error margin if the system is in global equilibrium and the sample size is large enough

This gives a method with a criterion for precise estimation in the TLTM!



5. Applying the TLTM to the Hubbard model

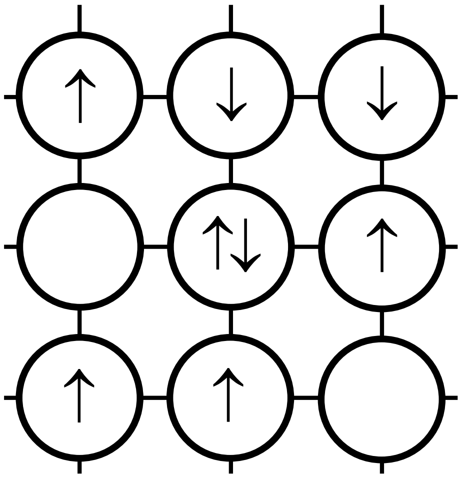
[MF-Matsumoto-Umeda 1906.04243]

Hubbard model (1/2)

Hubbard model [Hubbard 1963]

modeling electrons in a solid

- $c^{\dagger}_{\mathbf{x},\sigma},\ c_{\mathbf{x},\sigma}$: creation/anihilation op of an electron
  on site $\mathbf{x}$ with spin $\sigma\left(=\uparrow,\downarrow\right)$



$\left(N_S\right.$: # of sites$\left.\right)$

- Hamiltonian

$$H = -\kappa \sum_{\langle \mathbf{x},\mathbf{y}\rangle} \sum_{\sigma} c^{\dagger}_{\mathbf{x},\sigma} c_{\mathbf{y},\sigma} - \mu \sum_{\mathbf{x}} \left(n_{\mathbf{x},\uparrow} + n_{\mathbf{x},\downarrow}\right) + U \sum_{\mathbf{x}} n_{\mathbf{x},\uparrow} n_{\mathbf{x},\downarrow}$$

$n_{\mathbf{x},\sigma} \equiv c^{\dagger}_{\mathbf{x},\sigma} c_{\mathbf{x},\sigma}$

$\kappa\left(>0\right)$: hopping parameter

μ : chemical potential

$U\left(>0\right)$: strength of on-site replusive potential

$n_{\mathbf{x},\sigma} \to n_{\mathbf{x},\sigma} - 1/2$ s.t. $\mu = 0 \Leftrightarrow$ half-filling $\sum_{\sigma=\uparrow,\downarrow} \left\langle n_{\mathbf{x},\sigma} - 1/2 \right\rangle = 0$

$$H = \underbrace{-\kappa \sum_{\mathbf{x},\mathbf{y}} \sum_{\sigma} K_{\mathbf{x}\mathbf{y}}\, c^{\dagger}_{\mathbf{x},\sigma} c_{\mathbf{y},\sigma} - \mu \sum_{\mathbf{x}} \left(n_{\mathbf{x},\uparrow} + n_{\mathbf{x},\downarrow} - 1\right)}_{H_1} + \underbrace{U \sum_{\mathbf{x}} \left(n_{\mathbf{x},\uparrow} - \frac{1}{2}\right)\left(n_{\mathbf{x},\downarrow} - \frac{1}{2}\right)}_{H_2}$$

(fermion bilinear) (four fermion)

Hubbard model (2/2)

- Grand partition function (continuous imaginary time) : $Z_{\beta,\mu}^{\text{cont}} = \text{tr}\, e^{-\beta H}$

- Quantum Monte Carlo

$$e^{-\beta H} = e^{-\beta(H_1+H_2)} = \left(e^{-\epsilon(H_1+H_2)}\right)^{N_\tau} \cong \left(e^{-\epsilon H_1} e^{-\epsilon H_2}\right)^{N_\tau} \quad (\beta = N_\tau \epsilon)$$

⇨ Transform $e^{-\epsilon H_2} = \prod_{\mathbf{x}} e^{-\epsilon U\left(n_{\mathbf{x},\uparrow}-1/2\right)\left(n_{\mathbf{x},\downarrow}-1/2\right)}$ to a fermion bilinear using a boson ϕ

⇨
$$Z_{\beta,\mu} = \int [d\phi] e^{-S[\phi_{\ell,\mathbf{x}}]} \equiv \int \prod_{\ell=1}^{N_\tau} \prod_{\mathbf{x}} d\phi_{\ell,\mathbf{x}}\, e^{-(1/2)\sum_{\ell,\mathbf{x}} \phi_{\ell,\mathbf{x}}{}^2} \det M_\uparrow[\phi] \det M_\downarrow[\phi]$$

$$M_{\uparrow/\downarrow}[\phi] \equiv 1_{N_s} + e^{\pm\beta\mu} \prod_\ell \left(e^{\epsilon\kappa K} \,\text{diag}[e^{\pm i\sqrt{\epsilon U}\phi_{\ell,\mathbf{x}}}]\right) \; : N_s \times N_s \text{ matrix}$$

This gives complex actions for non half-filling states $(\mu \neq 0)$

<u>NB</u>: For half filling $(\mu = 0)$

$$\det M_\uparrow[\phi] \det M_\downarrow[\phi] = \left|\det M_\uparrow[\phi]\right|^2 \geq 0$$

$\Rightarrow$ No sign problem

⇨ <u>We apply the Tempered LTM to this system</u>

[MF-Matsumoto-Umeda 1906.04243]

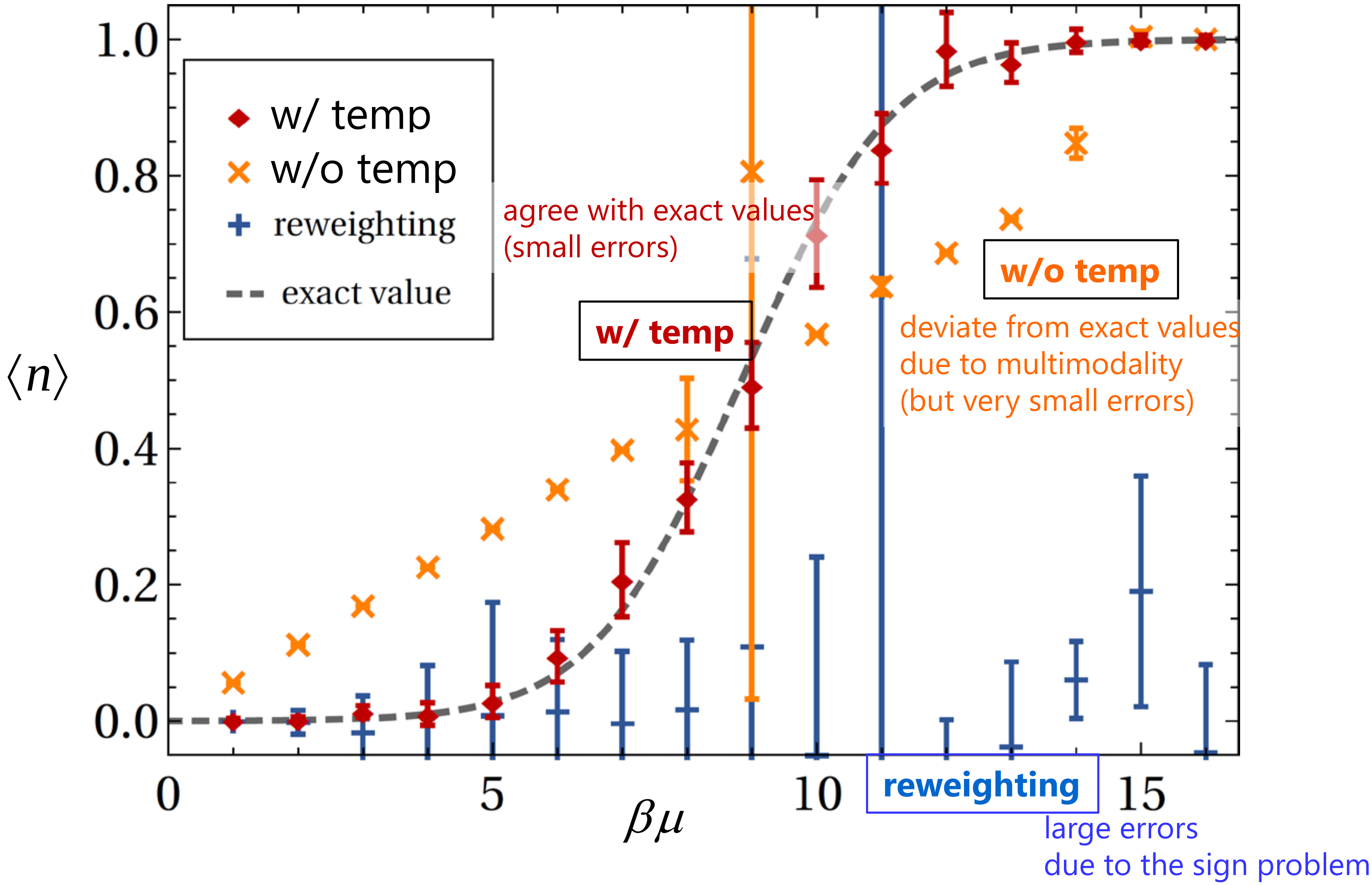
$$\begin{pmatrix} x = (x^i) = (\phi_{\ell,\mathbf{x}}) \in \mathbb{R}^N \\ i = 1,\ldots,N \;\; (N = N_\tau N_s) \end{pmatrix}$$

Results for 1D lattice (1/3)

[MF-Matsumoto-Umeda 2019]

imaginary time : 2 steps ($N_\tau = 2$)
spatial lattice: 1D periodic lattice with $N_s = 2$
$\beta\kappa = 1$, $\beta U = 16$, max flow time $T = 0.4$
sample size: 5,000

number density $n = \frac{1}{N_s}\sum_x (n_{x,\uparrow} + n_{x,\downarrow} - 1)$

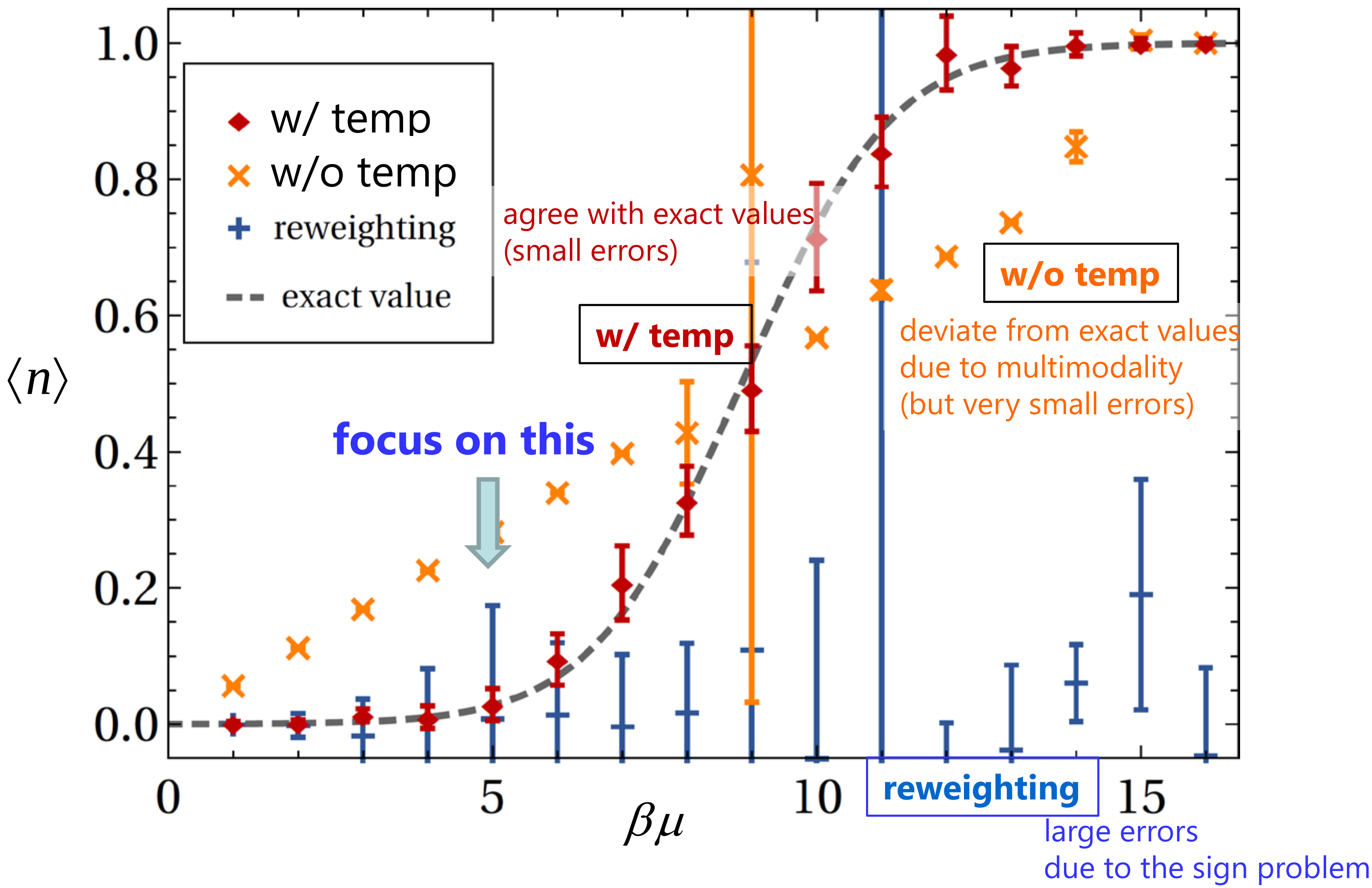


Results for 1D lattice (1/3)

[MF-Matsumoto-Umeda 2019]

imaginary time : 2 steps $(N_\tau = 2)$
spatial lattice: 1D periodic lattice with $N_s = 2$
$\beta\kappa = 1, \quad \beta U = 16,$ max flow time $T = 0.4$
sample size: 5,000

number density $n = \frac{1}{N_s}\sum_x \left(n_{x,\uparrow} + n_{x,\downarrow} - 1\right)$

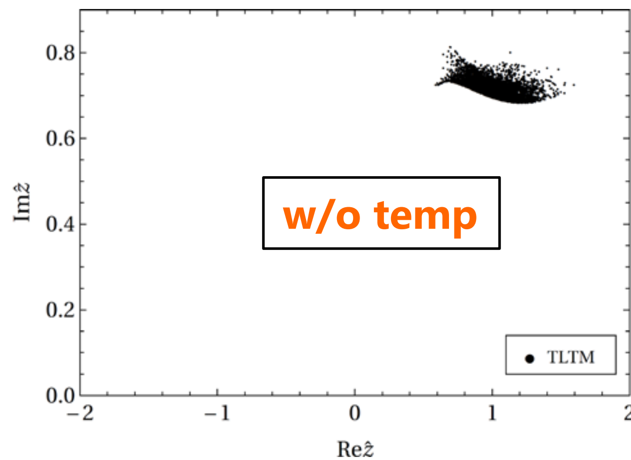


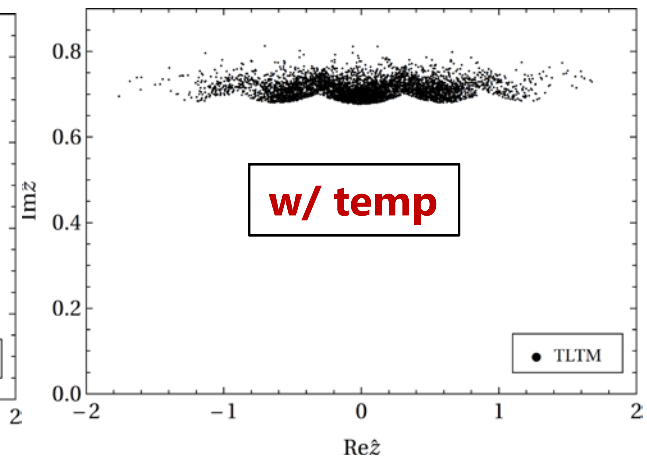
Results for 1D lattice (2/3)

[MF-Matsumoto-Umeda 2019]

Distribution of flowed configs at flow time $T = 0.4$

(projected on a plane)



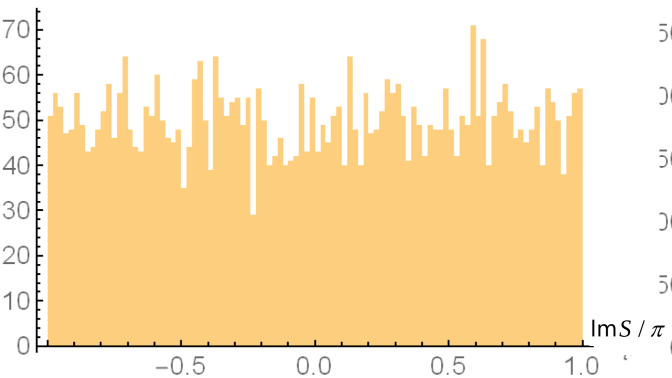


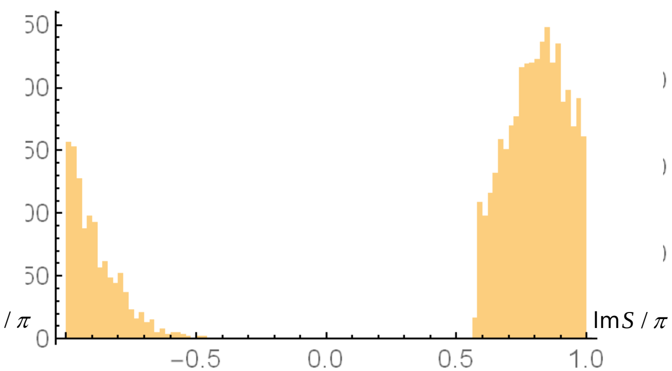
Histogram of ImS(z)/π

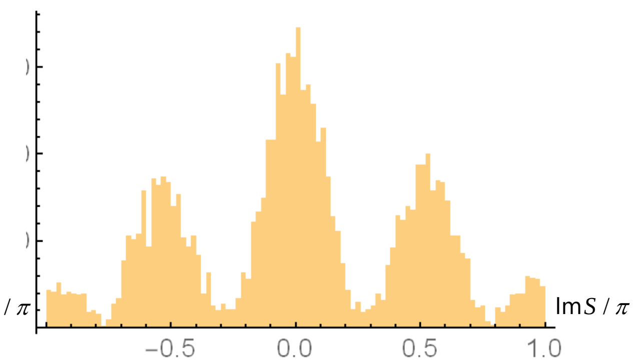
reweighting

w/o temp

w/ temp







distributing uniformly
from −π to +π

⇨ severe sign problem

peaked at a single angle ~0.8 π
due to the trap to a single thimble
(errors become small
because the thimble is well sampled)

peaked at several angles
because of sufficient transitions
among thimbles
(errors become a bit larger
due to the small size of sampling)

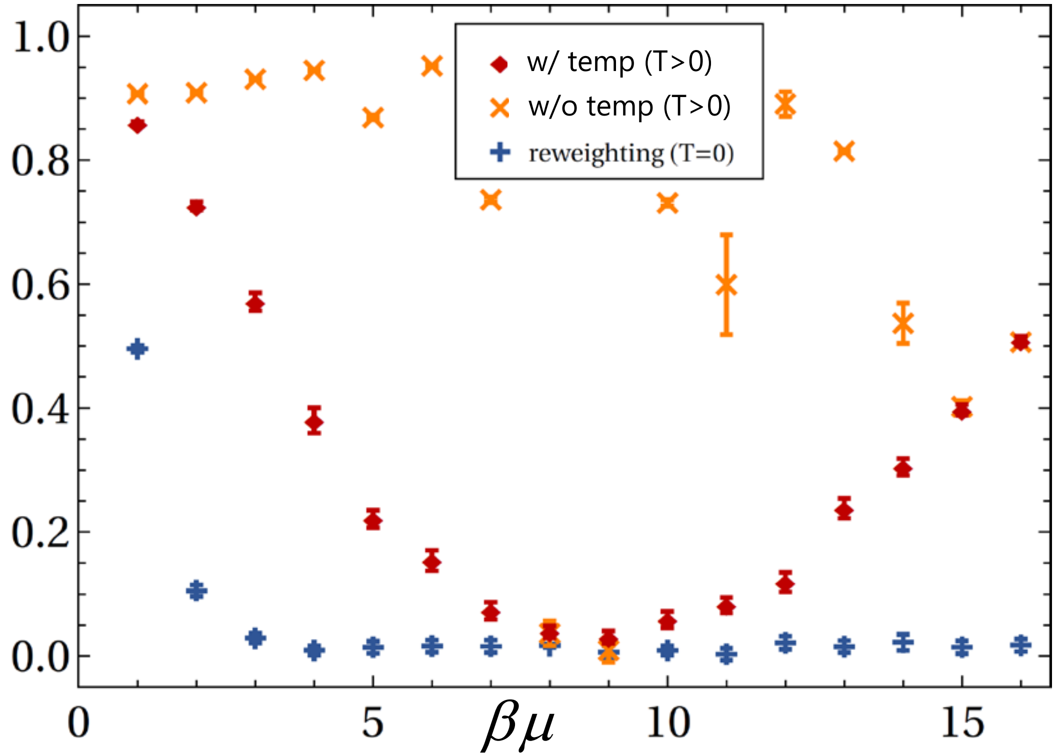
Results for 1D lattice (3/3)

[MF-Matsumoto-Umeda 2019]

sign average

$$\left(\langle \mathcal{O}(x) \rangle = \frac{\langle e^{i\theta_T(x)}\, \mathcal{O}(z_T(x)) \rangle_{S_T^{\mathrm{eff}}}}{\langle e^{i\theta_T(x)} \rangle_{S_T^{\mathrm{eff}}}} \right)$$

$$\left| \langle e^{i\theta_T(x)} \rangle_{S_T^{\mathrm{eff}}} \right|$$



When only a single (or very few) thimble(s) is sampled, the sign average can become larger than the correct sampling due to the absence of phase mixtures among thimbles

⇒ It is generally dangerous to regard the sign average as an index of the "resolution of the sign problem"

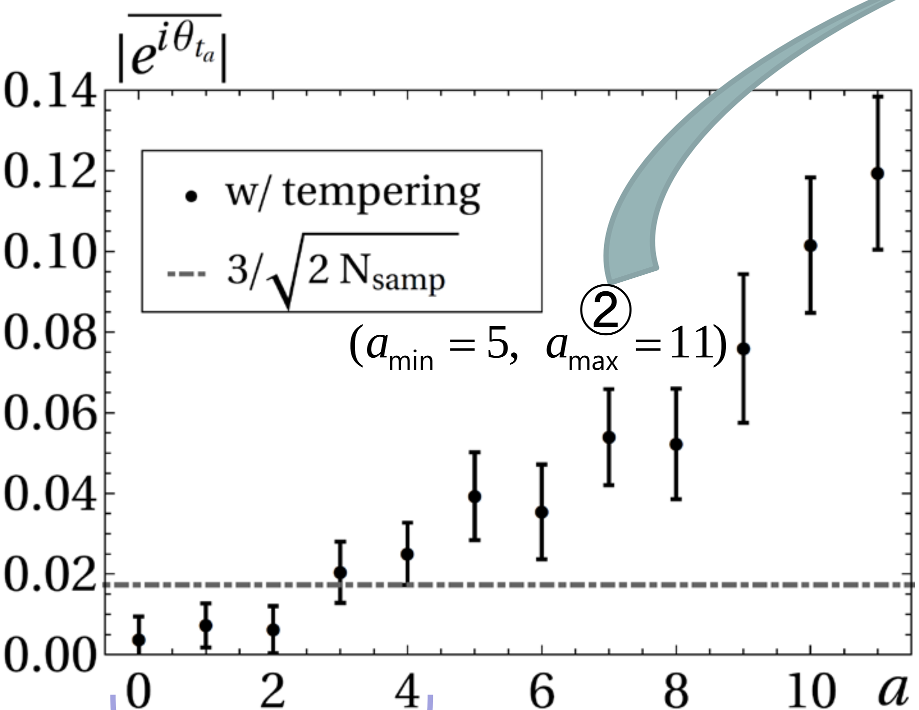
Results for 2D lattice (1/5)

[MF-Matsumoto-Umeda 1906.04243]

- imaginary time : 5 steps ($N_\tau = 5$)
- spatial lattice: 2D periodic lattice with $N_s = 2 \times 2$
- $\beta\kappa = 3 \quad \beta U = 13$, max flow time $T = 0.5$
- sample size: 5,000~25,000 depending on $\beta\mu$

$$\left(\langle n \rangle = \frac{\langle e^{i\theta_{t_a}(x)}\, n(z_{t_a}(x)) \rangle_{S^{\mathrm{eff}}_{t_a}}}{\langle e^{i\theta_{t_a}(x)} \rangle_{S^{\mathrm{eff}}_{t_a}}} \approx \overline{n}_a \right)$$

Example: $\beta\mu = 5$



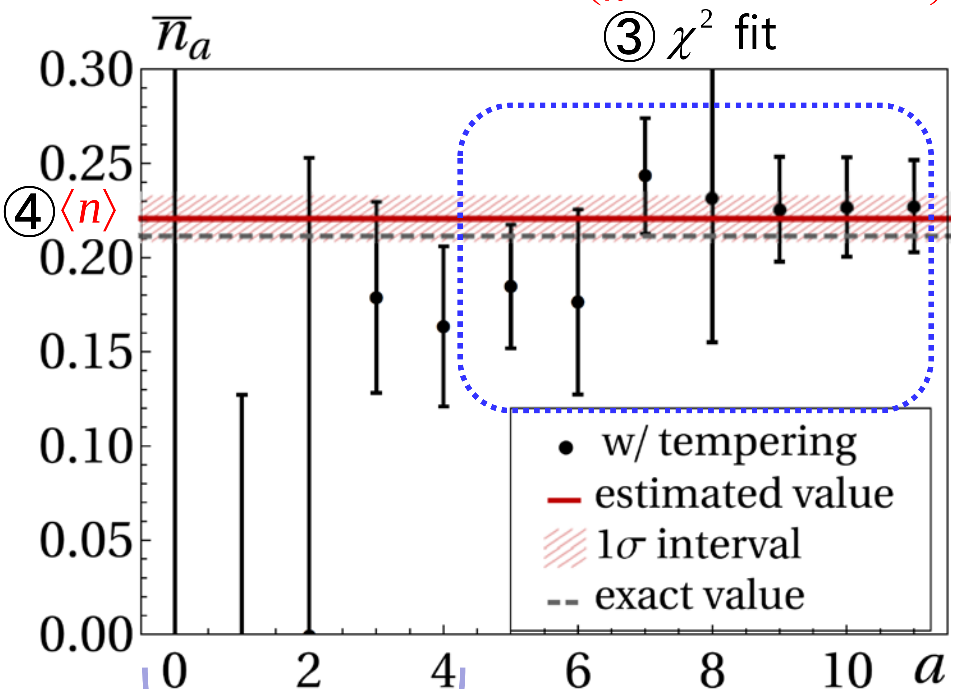
② ($a_{\min} = 5,\ a_{\max} = 11$)

① discarded

$\left(\begin{array}{l} \langle n \rangle \approx 0.221 \pm 0.012 \\ \chi^2 / \mathrm{DOF} = 0.45 \end{array} \right)$

③ χ^2 fit

④ $\langle n \rangle$



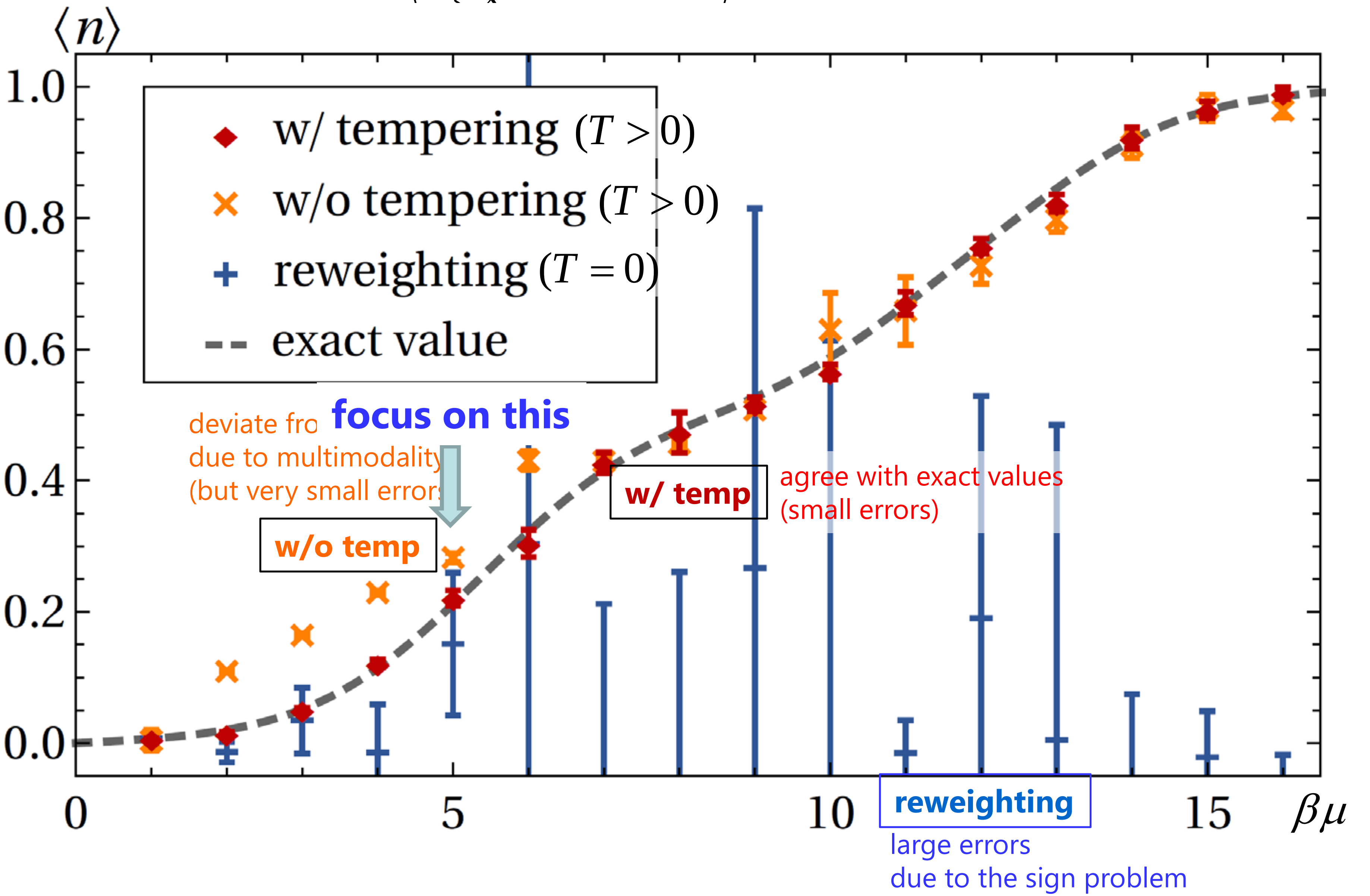
discarded

Results for 2D lattice (2/5)

$\begin{bmatrix} N_\tau = 5, N_s = 2\times 2 \\ \beta\kappa = 3, \quad \beta U = 13 \end{bmatrix}$

$$\langle n \rangle = \left\langle \frac{1}{N_s} \sum_{\mathbf{x}} \left(n_{\mathbf{x},\uparrow} + n_{\mathbf{x},\downarrow} - 1 \right) \right\rangle$$

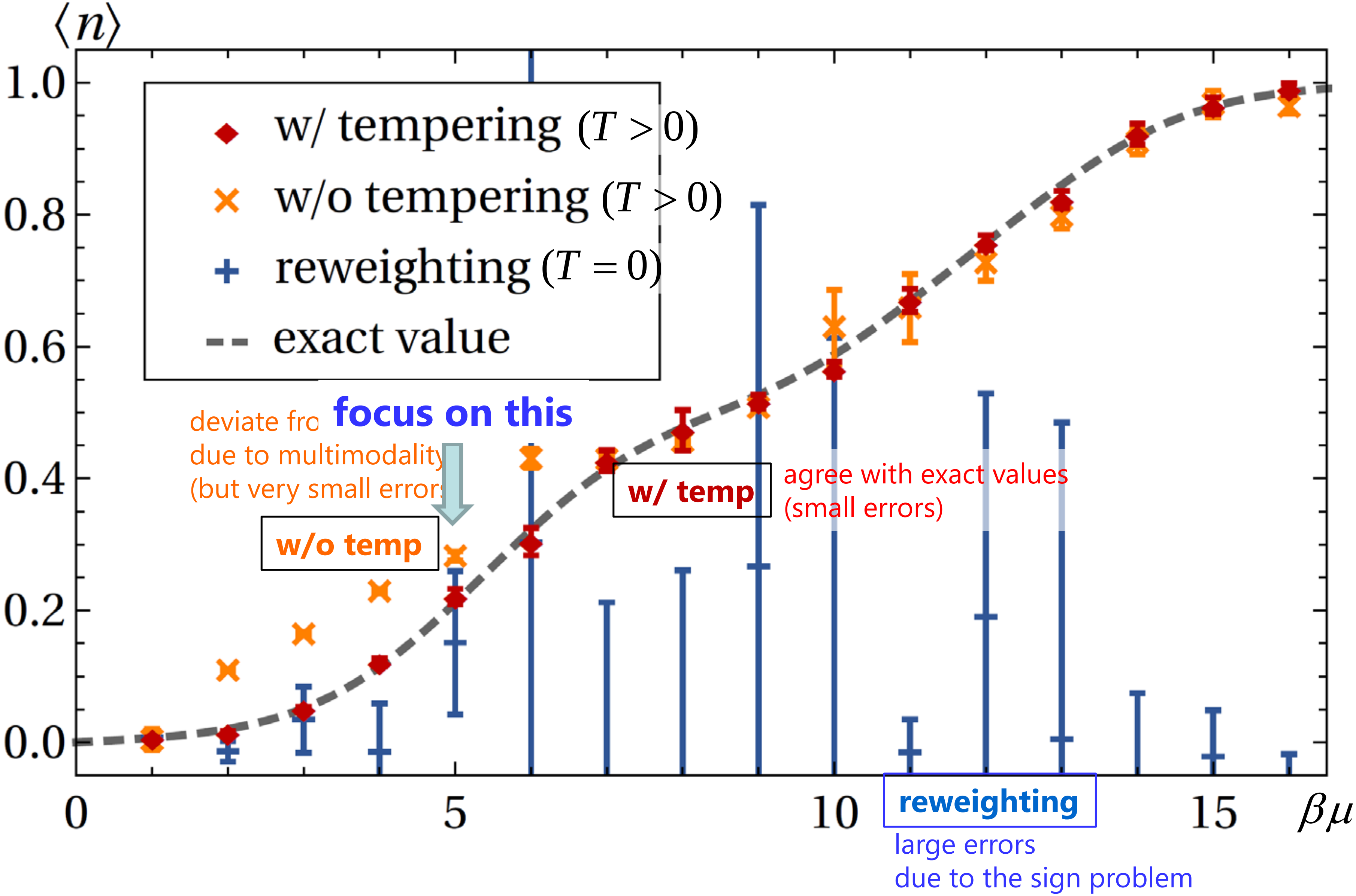
[MF-Matsumoto-Umeda 1906.04243]



Results for 2D lattice (2/5)

$$\begin{bmatrix} N_\tau = 5, N_s = 2\times 2 \\ \beta\kappa = 3, \quad \beta U = 13 \end{bmatrix} \qquad \langle n \rangle = \left\langle \frac{1}{N_s} \sum_{\mathbf{x}} \left(n_{\mathbf{x},\uparrow} + n_{\mathbf{x},\downarrow} - 1 \right) \right\rangle$$

[MF-Matsumoto-Umeda 1906.04243]

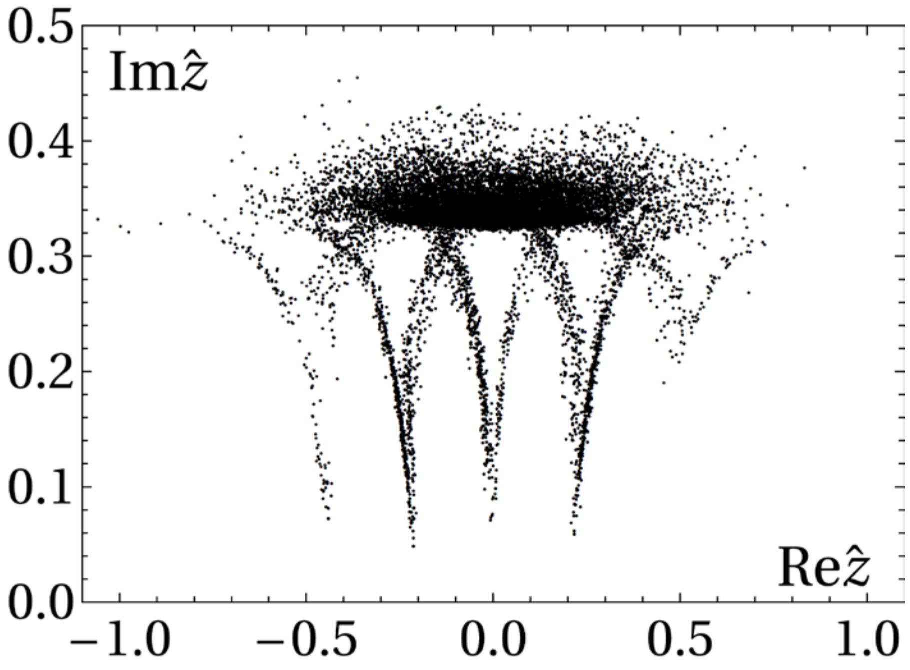


Results for 2D lattice (3/5)

[MF-Matsumoto-Umeda 1906.04243]

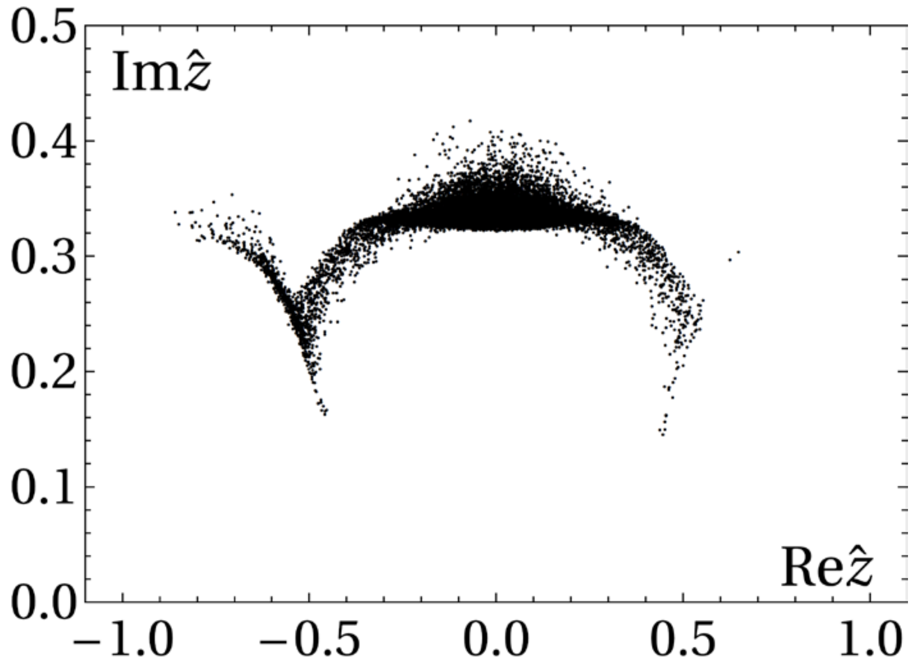
Distribution of flowed configs at flow time $T = 0.5$ $(\beta\mu = 5)$
(projected on a plane)

w/ temp



distributed widely
over many thimbles

w/o temp



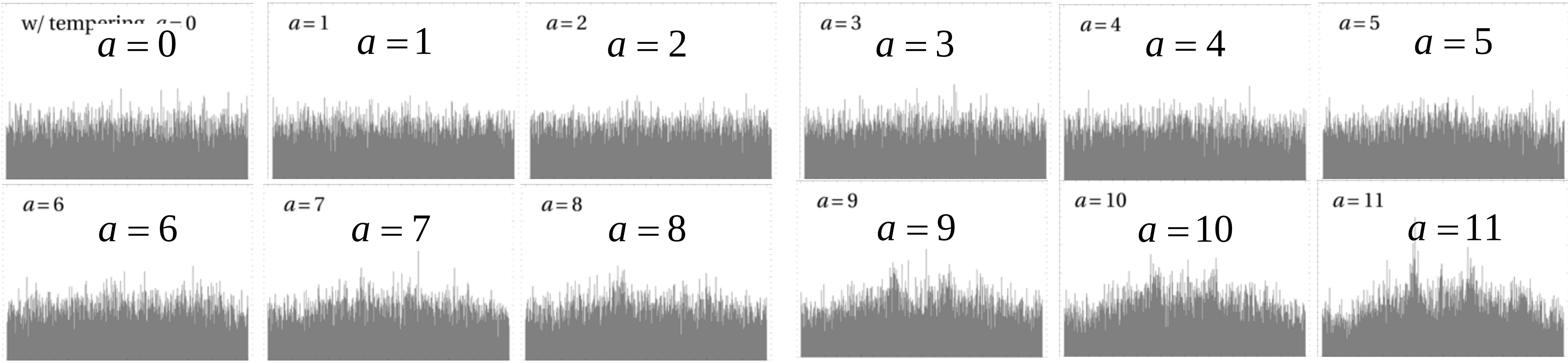
distributed over only
a small number of thimbles

Results for 2D lattice (4/5)

[MF-Matsumoto-Umeda 2019]

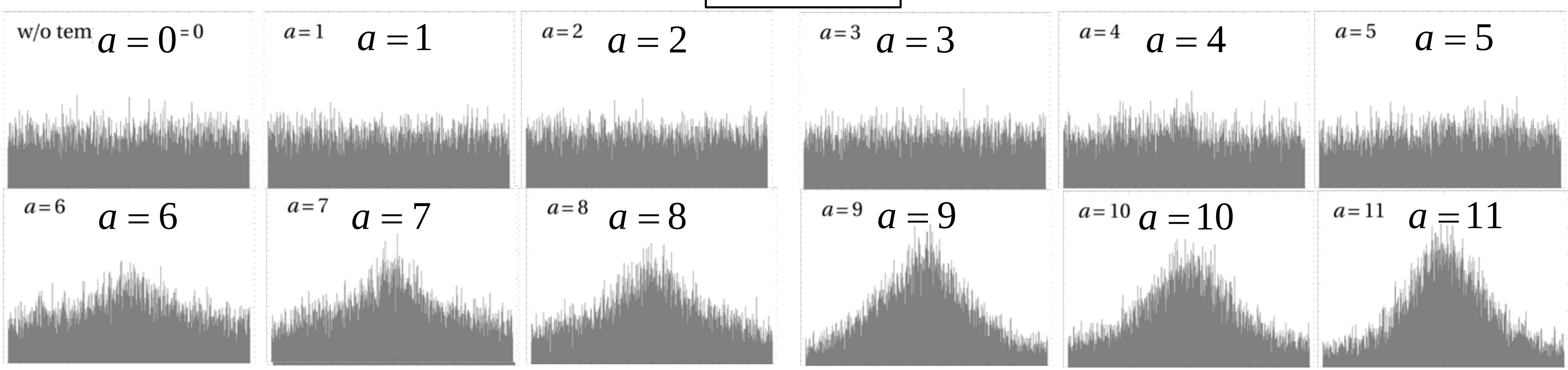
Histogram of $\theta_{t_a} \in [-\pi, \pi]$

w/ temp



many peaks (may not be so obvious because there are so many peaks and the peaks are broadened by Jacobian)

w/o temp



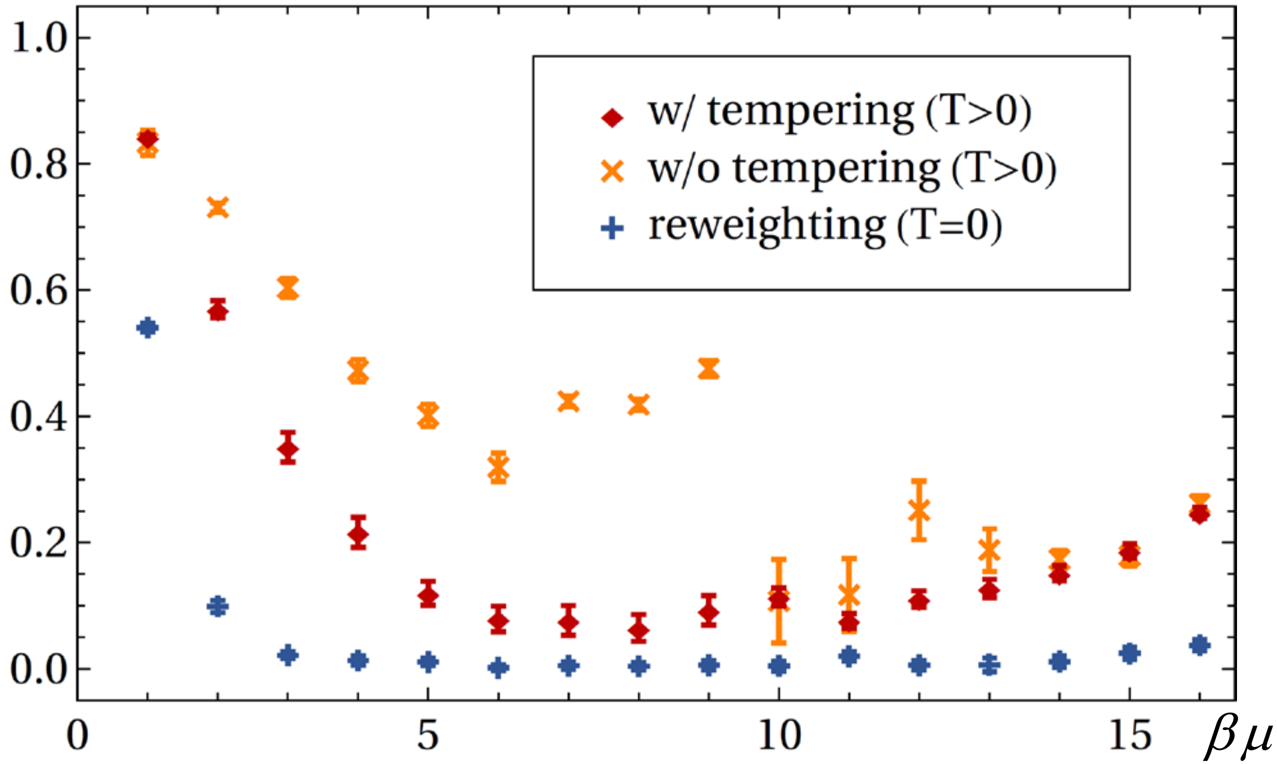
unimodal distribution

Results for 2D lattice (5/5)

[MF-Matsumoto-Umeda 1906.04243]

<u>sign average</u>

$$\left|\langle e^{i\theta_T(x)}\rangle_{S_T^{\mathrm{eff}}}\right| \quad \left(\langle \mathcal{O}(x)\rangle = \frac{\langle e^{i\theta_T(x)}\,\mathcal{O}(z_T(x))\rangle_{S_T^{\mathrm{eff}}}}{\langle e^{i\theta_T(x)}\rangle_{S_T^{\mathrm{eff}}}} \right)$$



When only a single (or very few) thimble(s) is sampled,
the sign average can become larger than that in the correct sampling
due to the absence of phase mixtures among thimbles

⇒ It is generally dangerous to regard the sign average
as an index of the "resolution of the sign problem"

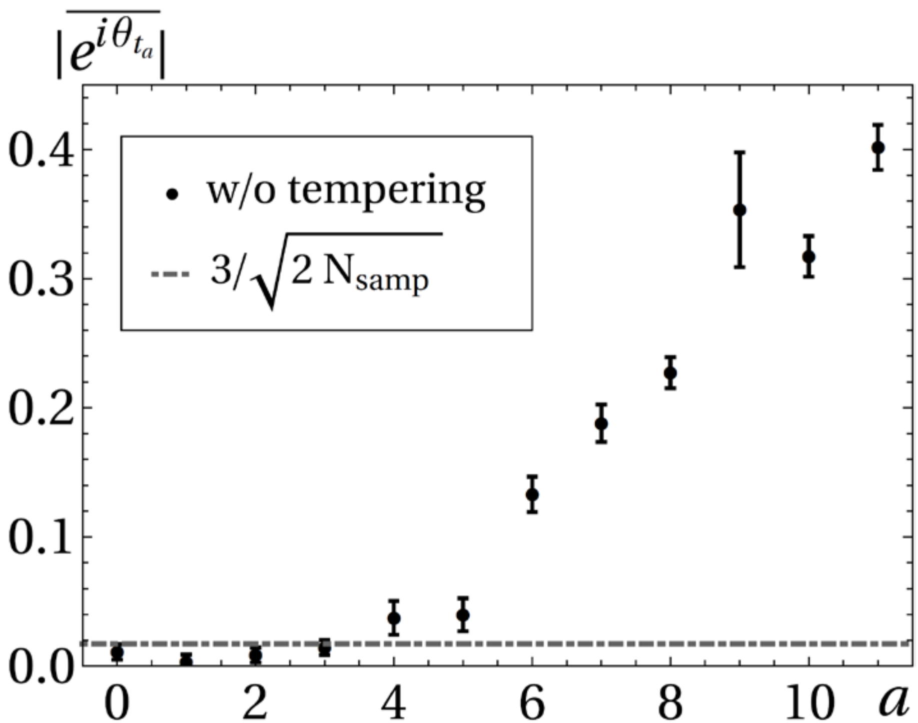
Comment on the Generalized LTM

[MF-Matsumoto-Umeda 1906.04243]

- imaginary time : 5 steps $(N_\tau = 5)$
- spatial lattice: 2D periodic lattice with $N_s = 2\times 2$
- $\beta\kappa = 3,\quad \beta U = 13,\quad 0 \le T \le 0.4 \left(\Leftrightarrow 0 \le a \le 10\right)$
- sample size: 5,000~25,000 depending on $\beta\mu$

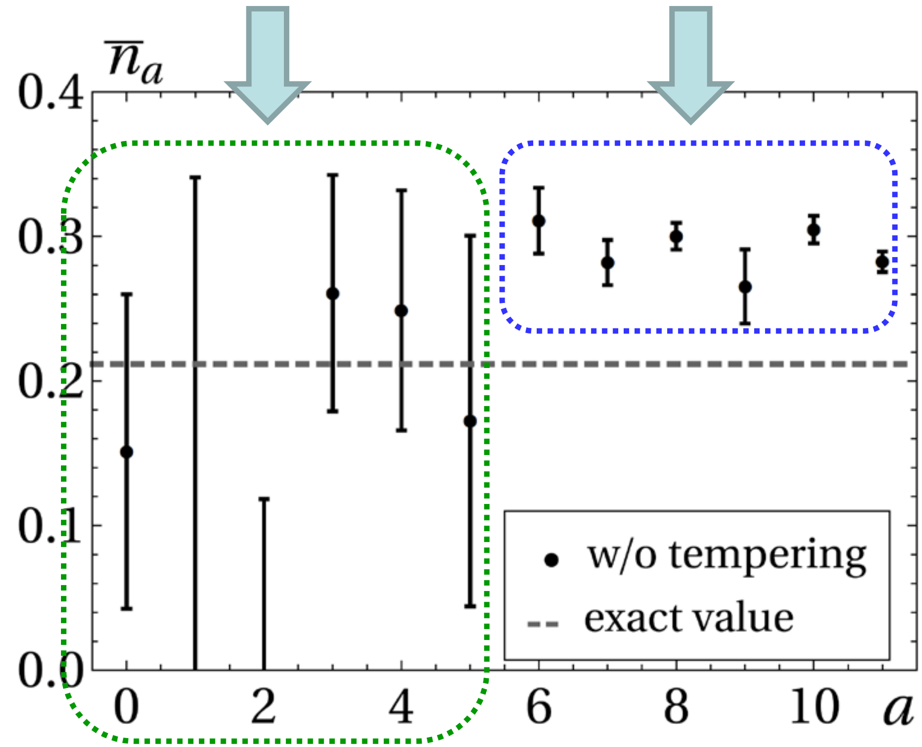
$$\left(\langle n \rangle = \frac{\langle e^{i\theta_{t_a}(x)}\, n(z_{t_a}(x)) \rangle_{S^{\mathrm{eff}}_{t_a}}}{\langle e^{i\theta_{t_a}(x)} \rangle_{S^{\mathrm{eff}}_{t_a}}} \approx \overline{n}_a \right)$$

Example: $\beta\mu = 5$



large stat errors (due to sign problem)

wrong value (due to multimodality)



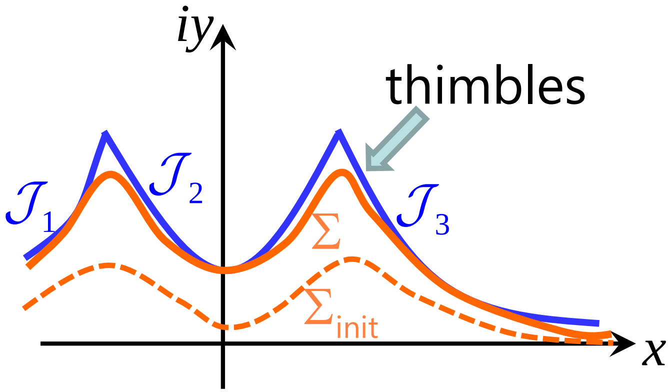
It is a hard task to find an intermediate flow time that solves both sign problem and multimodality

6. Other approaches

Path optimization (sign maximization) method

[Kashiwa-Mori-Ohnishi 1705.05605]
[Alexandru et al. 1804.00697]

Find a sign-optimized manifold Σ
where $|\langle e^{i\theta(z)}\rangle|$ takes a maximal value



NB

$|\langle e^{i\theta(z)}\rangle|$ may take larger values
when only a small number of thimbles are taken into account

⇒ Care must be paid not to miss good surfaces
when multi thimbles are relevant

This may also be used as a complementary method to TLTM
for improving the precision after one obtains
a rough shape of thimble and the corresponding sign average

Single-thimble dominance

[History]

There had been an expectation **[Cristoforetti et al. 1205.3996, 1303.7204, 1308.0233]** that only a single thimble dominates at criticality.

⇨ First counterexample: (0+1)-dim Thirring model **[Fujii-Kamata-Kikukawa 1509.08176]**

⇩

Multi thimbles are taken care of in Generalized LTM and Tempered LTM

Other approach: sticking to the single-thimble dominance

Develop a machinary so that
the problem can be reduced to caluculations over a single thimble

- Change of dynamical variables **[Di Renzo-Zambello, Ulybyshev et al. ,...]**
  - Works for the Hubbard model in some parameter region **[Ulybyshev et al. 1906.07678]**
  - May not be a versatile method ...
  - May be combined with TLTM to further improve the precision
- ⋯

7. Conclusion and outlook

Conclusion and outlook

What we have done:

- We proposed the tempered Lefschetz thimble method (TLTM) as a versatile method to solve the numerical sign problem
- We further developed it and found an algorithm to estimate expec. values with a criterion ensuring global equilibrium and the sample size (the key: $\overline{\mathcal{O}}_a$ should not depend on replica a due to Cauchy's theorem)
- GLTM can easily give incorrect results or large ambiguities
- TLTM works for the Hubbard model and gives correct results, avoiding both the sign and multimodal problems simultaneously

Outlook: **[MF-Matsumoto, work in progress]**

- Investigate the Hubbard model of larger temporal and spatial sizes to understand the phase structure [computational cost: $O(N^{3\sim 4})$]
- More generally, apply the TLTM to the following three typical subjects:

  ① Finite density QCD
  ② Quantum Monte Carlo (incl. the Hubbard model)
  ③ Real time QM/QFT

- Develop a more efficient algorithm with less computational cost

Thank you.