

VIETNAM NATIONAL UNIVERSITY, HANOI  
VNU UNIVERSITY OF SCIENCE  
**FACULTY OF PHYSICS**

**Luan Van TANG**

**BEREZINSKII-KOSTERLITZ-THOULESS  
PHASE TRANSITION IN  
2D XY MODEL**

Submitted in partial fulfillment of the requirements for the degree of  
Bachelor of Science in Physics  
(International Standard Program)

**Hanoi - 2015**

VIETNAM NATIONAL UNIVERSITY, HANOI  
VNU UNIVERSITY OF SCIENCE  
**FACULTY OF PHYSICS**

**Luan Van TANG**

**BEREZINSKII-KOSTERLITZ-THOULESS  
PHASE TRANSITION IN  
2D XY MODEL**

Submitted in partial fulfillment of the requirements for the degree of  
Bachelor of Science in Physics  
(International Standard Program)

**Supervisor: Oanh Hoang NGUYEN**

**Hanoi - 2015**

*to myParents.*

## Acknowledgement

I, Luan Van TANG the son of the former field officer in the special attack force Vietnamese's Army Thanh Ba TANG 1949 who was graduated from the Institute of Advanced Military, Hanoi, fought for the north-west front and his wife Nguyet Thi BUI 1962 a farmer at Lai Vu village, Kim Thanh, Hai Duong, Vietnam.

Chuan Duy BUI 1978 is my first master of mathematics and Dr Tan Ngoc NHAM 1973 is my high school teacher of mathematics.

I have committed in Berezinskii-Kosterlitz-Thouless phase transition at the suggestion of my supervisor Oanh Hoang NGUYEN Ph.D. 1979.

I would like to thank those teachers who taught me and those people who honestly helped, cared or promoted me.

This thesis is officially scheduled to be publicly defended on 17<sup>th</sup> December 2015.<sup>1</sup>

---

<sup>1</sup>[Color online] <http://luantv.github.io>

# Contents

<b>Acknowledgement</b>	i
<b>Table of contents</b>	iii
<b>Abstract</b>	iv
<b>1 Overview</b>	1
<b>2 Getting started</b>	3
2.1 2D XY model . . . . .	3
2.1.1 The Hamiltonian . . . . .	3
2.1.2 Vortex and anti-vortex . . . . .	3
2.2 Spin-wave theory . . . . .	4
2.2.1 Free energy minimization . . . . .	4
2.2.2 Periodic boundary condition . . . . .	7
<b>3 Algorithms</b>	8
3.1 Metropolis . . . . .	8
3.2 Wolff . . . . .	9
<b>4 Observables</b>	10
4.1 Equilibration . . . . .	10
4.1.1 Metastability . . . . .	12
4.1.2 Standard Metropolis . . . . .	12
4.1.3 Standard Wolff . . . . .	15
4.1.4 Parallel Metropolis with CUDA, OpenMP directives . . . . .	15
4.2 Phenomenology . . . . .	18
4.3 Let's measure . . . . .	22
4.3.1 Spin-spin correlation . . . . .	22
4.3.2 Magnetization . . . . .	24
4.3.3 Magnetic susceptibility . . . . .	26
4.3.4 Square of magnetization . . . . .	28
4.3.5 Magnetic reversal . . . . .	29
4.3.6 Empirical evidence of magnetization . . . . .	31
4.3.7 Square of spin phase angle . . . . .	32
4.3.8 Energy . . . . .	33
4.3.9 Specific heat . . . . .	34
4.3.10 Helicity modulus . . . . .	35
4.3.11 Vortex density . . . . .	38
<b>5 Conclusions</b>	40
<b>A Canonical ensemble</b>	41
<b>B Importance Sampling and unbiased estimations</b>	42

<b>C</b>	<b>Markov chain Monte Carlo methods</b>	<b>45</b>
C.1	The theory of finite Markov chain . . . . .	45
C.1.1	Definitions and general properties . . . . .	45
C.1.2	Balance conditions . . . . .	47
C.1.3	Ergodicity in Markov chain . . . . .	48
C.2	Metropolis - Hastings . . . . .	49
C.2.1	Random proposal . . . . .	49
C.2.2	Spread of proposal density . . . . .	51
C.2.3	Typewriter proposal . . . . .	51
C.2.4	Checker board proposal . . . . .	52
C.3	Wolff dynamics . . . . .	55
C.3.1	Generalized spin flip operation . . . . .	55
C.3.2	Markov properties . . . . .	57
<b>D</b>	<b>Data analysis and equilibrium dynamics</b>	<b>58</b>
D.1	Statistics of correlations . . . . .	58
D.2	Equilibrium dynamics . . . . .	59
D.2.1	Autocorrelation function . . . . .	59
D.2.2	Exponential autocorrelation time . . . . .	60
D.2.3	Integrated autocorrelation time . . . . .	61
D.3	Binning analysis . . . . .	62
D.4	Resampling methods: Jackknife estimator . . . . .	64
<b>References</b>		<b>68</b>
<b>Code snippets</b>		<b>71</b>

# Abstract

- × The virtue of Markov chain Monte Carlo methods was partly understood.
- × Metropolis-Hastings random proposal, checker board proposal as well as single cluster Wolff algorithm was studied.
- × Parallel programming based on CUDA API on NVIDIA's GPGPU was investigated and optimized to some level.
- × The correctness of measurements is assured by either studying temporal correlation or using estimation methods for observables of interest.
- × Physical properties of the 2D XY model are demonstrated by measuring some observables, spin-spin spatial correlation, correlation length, energy, specific heat, magnetization, magnetic susceptibility, helicity modulus, and vortex density using periodic boundary condition, consistency check with theoretical approximation as well as simulation and experiment literature.
- × The Nelson-Kosterlitz universal jump of helicity modulus was used to extract the critical temperature  $T_{BKT}$  at which a special kind of phase transition occurs, the Berezinskii-Kosterlitz-Thouless.

**Keywords:** Berezinskii-Kosterlitz-Thouless phase transition, Metropolis, Wolff algorithm, 2D XY model, superconductivity, CUDA.

# 1 Overview

Bloch<sup>2</sup> in 1930 found that the lowest lying excited states of the Heisenberg model were wavelike, called them spin waves, and showed that they reduced the spontaneous magnetization at low temperature.[1, p. 41] XY model is a variant in which the spins are two dimensional while in Heisenberg model they are instead three dimensional. It differs from Ising model with discrete spins  $Z(2)$  that the XY model with continuous spins  $O(2)$  of which thermal fluctuations are strong enough to destroy any long-range order at low temperatures then it was believed for a long time that the system has no phase transition between a phase of order and a disordered phase that occurs in two dimensional Ising model.

On Aug. 16, 1966[2], Stanley and Kaplan published arguments based on a computational method of high-temperature expansions for the possibility of a phase transition in which the susceptibility becomes infinite for the two-dimensional Heisenberg model against the above conclusion of no phase transition in two dimensions from the standard spin-wave argument. On Oct. 17, 1966,<sup>3</sup> Mermin and Wagner published solid proof that there is no spontaneous magnetization in one or two dimensional isotropic Heisenberg models, however, the authors emphasized that their result does not exclude the possibility of other kinds of phase transitions. The validity region of the results of spin wave theory were known to be at very low temperature. On Feb. 27, 1969, P. D. Loly[1] criticized some poor approximations based on this theory since the effective range of the theory was not explicitly understood specifically for large spins, i.e. large systems, however, for 3 dimensions.

On Apr. 11, 1972[3] and Nov. 13, 1972[4], two papers published by Kosterlitz and Thouless used dislocation theory to define a kind of long range order termed topological order that is present in 2D XY model but not in the isotropic Heisenberg model, mentioned Berezinskii's 1971 work, and confirmed the Berezinskii-Kosterlitz-Thouless phase transition in the former case.

The main purpose of this thesis is simply checking consistency with published results of spin wave theory applied to 2D XY model at low temperatures by various algorithms and techniques.

In the following section, first is some introductions of the model, the Hamiltonian, vortex and boundary conditions, followed by the Berezinskii-Kosterlitz-Thouless phase transition is discussed quantitatively theoretically based on spin wave theory, and finally is a phenomenological investigation. In section 3, two Markov chain Monte Carlo methods are introduced, the Metropolis with random proposal as well as checker board proposal, and the single cluster Wolff algorithm which is closely related to the physics of the model. Intuitive arguments for their ergodicity, balance as well as details of their decomposition are given in the appendix C.

In section 4, effort was first given to equilibrate the system. We managed to use the results of equilibration time published in Ref. [5], however, we failed to find a way to realize those proportionalities for our favour of thermal averages. Nevertheless, the authors concluded a remarkable result is that the the relaxation from a hot configuration with vortices is slower<sup>4</sup> than that from a cold configuration with no vortex by some

---

<sup>2</sup>F. Bloch. *Zeitschrift für Physik*, 61(3-4):206–219, 1930.

<sup>3</sup>Wagner H. Mermin, N. D. *Phys. Rev. Lett.*, 17:1133–1136, Nov 1966.

<sup>4</sup>One may underestimate this result saying that it's easy to understand. Then one should note

logarithmic factor. Some authors<sup>5</sup> defined some non-linear relaxation function and its associate time  $\tau_A^{(nl)}$  by which the system is well equilibrated after some amount that is much more greater than that. Although we agree with the authors that the overall equilibration time must be taken from the slowest-relaxing quantity, e.g. the order parameter if any, however, we found that the non-linear time is much more smaller in some large order of magnitude than the amount after which our system can be in exact equilibrium, i.e. we do not understand what do the authors mean by  $t_{M_0} \gg \tau_A^{(nl)}$ . Coming with the above remarkable result, we then push forward to empirical observations and found a practical way to equilibrate the system fully avoiding metastability.

After taking the system to exact equilibrium, the physics of the model is studied by measuring some physical quantities with particular discussions. Some discrepancies come with the susceptibility, namely, in some reference, it is related to the norm of magnetization and some peak near the transition temperature occurs, while in Ref. [6], it was reported to diverge at low temperatures with no peak<sup>6</sup>. We learned the discussion of the use of absolute value of magnetization in Ising model in Ref. [7], and then we discussed for our 2D XY model the phenomenon of magnetic reversal as well as the non-existence of spontaneous magnetization in the thermodynamic limit. Although the same periodic boundary conditions were used, however, in some reference, the net number of vortices was not equal to those of anti-vortices which is contradicted to our result of an equality. Fortunately, with the aid of Ref. [6] we were able to conceive our solid arguments of mathematics as well as physics by which our result holds. Just one more thing, our result of vortex density is not consistent with literature.

The final section includes several conclusions. Beyond these are the appendices that some detailed studies of related concepts and discussions of related problems are given.

---

that the acceptance probability is not symmetric.

<sup>5</sup>Pp. 34, 35, Ref. K. Binder and D. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction*. Graduate Texts in Physics. Springer Berlin Heidelberg, 2010.

<sup>6</sup>Also the prediction from high temperature expansion in Ref. [2].

## 2 Getting started

### 2.1 2D XY model

#### 2.1.1 The Hamiltonian

The 2D XY model is a two-dimensional lattice of linear dimension  $L$ , lattice constant  $a$ , at each lattice site is located a unit vector  $\mathbf{s}_{ij}$  which are magnetic dipoles or spins of unit magnitude that are allowed to rotate in the plane of the lattice. The Hamiltonian of the system for each configuration  $\sigma = \{\mathbf{s}_{ij}, 1 \leq i, j \leq L\}$  is given by [4, p. 1190]

$$\mathcal{H} = - \sum_{\langle ij, lm \rangle} J_{ijlm} \mathbf{s}_{ij}^T \mathbf{s}_{lm} \quad (1)$$

where  $\langle ij, lm \rangle$  denotes a pair of adjacent spins, i.e. either  $l = i + 1, m = j$ , or  $l = i, m = j + 1$ , and  $J_{ijlm}$  is the coupling constant.

Since each spin is characterized by a vector of unit magnitude and an angle  $\theta_{ij}$  made with some arbitrarily fixed axis in the plane[6], the Hamiltonian (1) is equivalent to

$$\mathcal{H} = - \sum_{\langle ij, lm \rangle} J_{ijlm} \cos(\theta_{ij} - \theta_{lm}) \quad (2)$$

A time-independent external magnetic field<sup>7</sup>  $\mathbf{H} = \sum_{ij} \mathbf{h}_{ij}$  applied to the system contributing an additional term  $-\mathbf{HM}$  to the Hamiltonian<sup>8</sup>

$$\mathcal{H} = - \sum_{\langle ij, lm \rangle} J_{ijlm} \mathbf{s}_{ij}^T \mathbf{s}_{lm} - \mathbf{HM} \quad (3)$$

where  $\mathbf{M}$  termed the net magnetization which is equal to  $\sum_{ij} \mathbf{s}_{ij}$ , and  $\mathbf{h}_{ij}$  is the field specifically conjugated to spin  $\mathbf{s}_{ij}$ , i.e.  $\mathbf{h}_{ij} \mathbf{s}_{lm} \sim \delta(ij, lm)$ , and  $\mathbf{HM} = \sum_{ij} \mathbf{h}_{ij} \mathbf{s}_{ij}$ .

#### 2.1.2 Vortex and anti-vortex

Consider a primitive cell, which is a square of size  $a$  with four corners each locates a spin. Some fixed direction is specified to be the positive direction for any closed path in the lattice, e.g., in our simulation the counter-clockwise. Then, the relative spin phase angle difference  $d\theta$  between four pairs of adjacent spins  $d\theta \in [-\pi, \pi]$  in the chosen direction is computed by taking those differences outside the range are shifted a period of  $2\pi$  back to the origin, i.e. if  $d\theta > \pi$ ,  $d\theta - 2\pi$ , and if  $d\theta < -\pi$ ,  $d\theta + 2\pi$ , and their sum<sup>9</sup>  $\oint_P d\theta$  is used to classify the local configurations. These shifts are due to the fact that in simulation, each spin phase angle is generated in the interval  $[0, 2\pi]$ , i.e. an implicit reference axis has been made. An integration used instead of a sum is for convenience. Three possible configurations defined in [Fig. 1]<sup>10</sup>. In our simulation, a vortex is counted when the integral is larger than<sup>11</sup>  $\pi_{\text{def}}$  and for an anti-vortex, it is

---

<sup>7</sup>Each spin can be considered as elementary magnetic dipoles.

<sup>8</sup>See, e.g. [8, p. 164].

<sup>9</sup>The mathematical sum is used with the note of no reference axis, just the relative differences together with a positive direction are concerned. Otherwise, it's simply equal zero.

<sup>10</sup>Adapted, reproduced Ref. Urs Gerber et al. *J. Phys. Conference Series*, 651(1):012010, 2015.

<sup>11</sup>The approximated value of  $\pi$  defined in our program  $\pi_{\text{def}} = 3.141592653589793$ .

smaller than  $-\pi_{\text{def}}$ . Such comparisons is due to the fact that equality can't be realized with floating point precision in computer. The general result for any closed path  $\mathcal{P}$  is

$$\oint_{\mathcal{P}} d\theta = n2\pi, \quad n = 0, \pm 1, \pm 2\dots \quad (4)$$

where  $n$  termed the winding number is the number of vortices or number of anti-vortices or their algebraic sum enclosed by the closed path. This can be proved by simply taking the decomposition of the whole path into those closed one that encircles a single lattice cell.

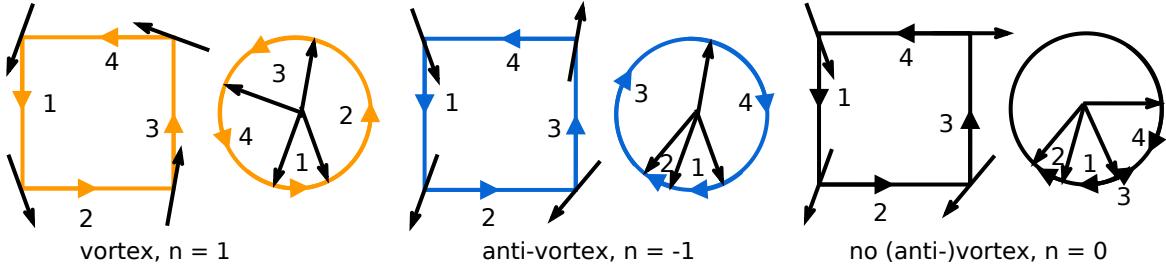


Figure 1: Realization of (4) gives three possibilities. Condition  $d\theta \in [-\pi, \pi]$  is geometrically realized by shortest arcs on the unit circles.

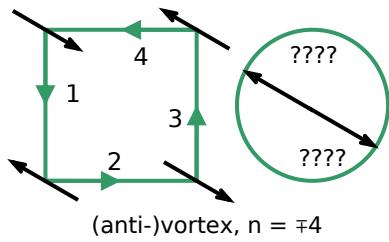


Figure 2: A fourth case.

As mentioned in Ref. [6], there is a possibility in which the integral (4) taken along a closed path connecting four corners of a primitive cell returns some larger than 2 the integral value of  $\pi$ . This case has the directed spin phase angle differences are  $-\pi, \pi, -\pi, \pi$ , respectively [Fig. 2].

However, we realized that this case can only occur due to the problem of limited precision representation on computer, however, the probability of occurrence is significantly small as for other several outputs that may occur related to this problem.

## 2.2 Spin-wave theory

### 2.2.1 Free energy minimization

According to Mermin-Wagner theorem, 2D XY model has no spontaneous magnetization at any finite temperature, i.e. there is no long range order even at low temperatures. Nevertheless, there is another kind of quasi-long range order termed topological order makes it possible to have a special kind of phase transition named Berezinskii-Kosterlitz-Thouless phase transition.

Let's minimize the free energy[9, p. 8]

$$F = -k_B T \ln Z = E - k_B T S \quad (5)$$

For finite temperatures, only configurations [Fig. 3] in which adjacent spins having angles nearly equal give significant contribution to the partition function.[4, p. 1190]

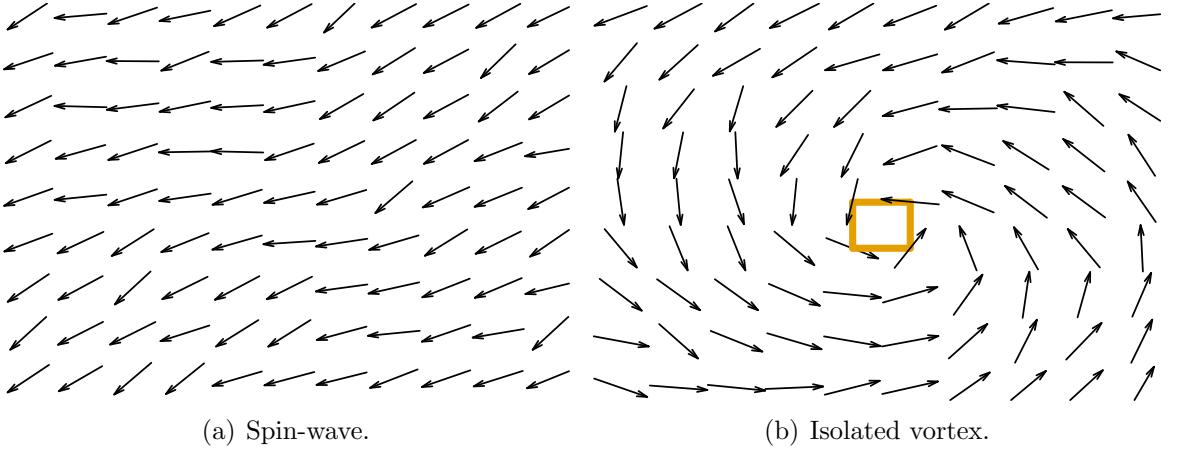


Figure 3:  $L = 32, T = 0.01(J/k_B)$ .

This is easily understood as in such configurations, adjacent spins have small angle differences leading to large value of cosine functions and the Hamiltonian is negatively large and then large-valued terms contributed to the partition function. The Hamiltonian for such configurations with the form of (2) can be expanded up to quadratic terms of angles as<sup>12</sup>

$$\mathcal{H} = - \sum_{\langle ij, lm \rangle} J_{ijlm} \left\{ 1 - 2 \sin^2 [(\theta_{ij} - \theta_{lm}) / 2] \right\} = - \sum_{\langle ij, lm \rangle} J_{ijlm} \left[ 1 - (\theta_{ij} - \theta_{lm})^2 / 2 \right] \quad (6)$$

We're interested in an investigation of vortex or anti-vortex in low temperature phase, so beside the purely spin-wave configurations as [Fig. 3(a)], the configuration [Fig. 3(b)] with an isolated vortex is also such a configuration with adjacent spin phase angles are approximately equal, especially for large system and far from the vortex core. However, there should be a problem of mathematics with the harmonic approximation (6), i.e. near the vortex core, it's definitely that the spin phase angles change significantly between adjacent lattice sites. In order to continue with spin-wave theory, let's denote  $\mathcal{H}_{core}$  the energy of such a core which may be finite or infinite.

A general configuration consists of two parts, the spin-wave part  $\theta_{sp}$  and the vortex part  $\theta_{vor}$ , i.e.

$$\theta(\mathbf{r}) = \theta_{sp}(\mathbf{r}) + \theta_{vor}(\mathbf{r}) \quad (7)$$

As discussed, these parts satisfy the following

$$\oint_{\mathcal{P}} d\theta_{sp}(\mathbf{r}) = 0, \quad \oint_{\mathcal{P}} d\theta_{vor}(\mathbf{r}) = 2n\pi, \quad n = \pm 1, \pm 2, \dots \quad (8)$$

---

<sup>12</sup>Here the approximation is  $\sin(x) \approx x$ , due to Taylor expansion:  $\sin(x) = x - x^3/3! \dots$

Because the system is two-dimensional, the Hamiltonian (6) is equivalent to

$$\mathcal{H} = - \sum_{\langle ij, lm \rangle} J_{ijlm} + \frac{1}{2} \sum_{ij} [(\theta_{ij} - \theta_{lmx})^2 + (\theta_{ij} - \theta_{lmy})^2] \quad (9)$$

$$\approx E_0 + \frac{1}{2} \sum_{ij} \left[ a^2 \left( \frac{\partial \theta_{ij}}{\partial x} \right)^2 + a^2 \left( \frac{\partial \theta_{ij}}{\partial y} \right)^2 \right] \quad (10)$$

$$= E_0 + \frac{1}{2} \sum_{ij} a^2 (\nabla \theta_{ij})^2 \quad (11)$$

$$\xrightarrow{a \rightarrow 0} E_0 + \frac{1}{2} \int dx dy [\nabla \theta(x, y)]^2 \equiv E_0 + \frac{1}{2} \int d^2 r [\nabla \theta(r)]^2 \quad (12)$$

where  $\theta_{lmx}$ ,  $\theta_{lmy}$  are the neighbouring spin angles of spin  $\mathbf{s}_{ij}$  assuring no duplication in the sum.

Now, take into account the decomposition (8), (12) reads

$$\mathcal{H} \approx E_0 + \frac{1}{2} \int d^2 r [\nabla \theta_{sp}(r) + \nabla \theta_{vor}(r)]^2 \quad (13)$$

**Does spin-wave interact with vortex?** The answer is not clear now. Because there should arises also another question of **the interaction between vortices**. Without the former interaction, the cross terms vanish and (13) becomes

$$\mathcal{H} \approx E_0 + \frac{1}{2} \int d^2 r [\nabla \theta_{sp}(r)]^2 + \frac{1}{2} \int d^2 r [\nabla \theta_{vor}(r)]^2 \quad (14)$$

Purposed to determine the gradient  $|\nabla \theta(r)|$ , let's extend the closed path  $\mathcal{P}$  around a lattice cell to cover any square of the same center at the vortex core  $n = 1$  but with radius  $r$ , we have<sup>131415</sup>

$$2n\pi = \oint_{\mathcal{P}} d\theta(r) = \oint_{\mathcal{P}} dl \nabla \theta(r) = \oint_{\mathcal{P}} dl |\nabla \theta(r)| \quad (15)$$

$$= |\nabla \theta(r)| \oint_{\mathcal{P}} dl = |\nabla \theta(r)| r \oint_{\mathcal{P}} d\phi \quad (16)$$

$$\rightarrow |\nabla \theta(r)| = n/r \quad (17)$$

where in polar coordinates[11, pp. 40, 42]  $\hat{\mathbf{r}}, \hat{\phi}$ ,

$$dl = dr \hat{\mathbf{r}} + rd\phi \hat{\phi}, \quad (18)$$

$$\nabla = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\phi} \frac{1}{r} \frac{\partial}{\partial \phi} \quad (19)$$

however, since the radius is fixed, the  $\hat{\mathbf{r}}$ -component vanishes, the line and the gradient have the same direction that leads to the third equality. The fourth equality is because

---

<sup>13</sup>See H. J. Jensen *The Kosterlitz-Thouless transition*, Dept. of Mathematics, Imperial College.

<sup>14</sup>One may see some paradox here, i.e. the first closed path integral is simply zero, however, the theory considers no reference axis, just the relative spin phase angle differences.

<sup>15</sup>An alternative, since  $\oint_{\mathcal{P}} d\theta(r) = n2\pi$ , let  $\theta(r) = n\phi(r) \Rightarrow \nabla \theta(r) = \hat{\phi}n/r$ . See e.g. [10, pp. 116, 122] and P. M. Chaikin and T. C. Lubensky, *Principles of condensed matter physics*, CUP, 1995, p. 527.

of the magnitude of the gradient is simply the difference between adjacent spin phase angles divided by the lattice spacing in the continuum limit  $a \rightarrow 0$  so that it doesn't depend explicitly on  $\phi$ .

The energy due to the presence of a vortex, i.e. the second integral in (14), accounted for a finite region of radius of  $L/2$ , is<sup>16</sup>

$$\mathcal{H}_{vor} = \frac{1}{2} \int d^2r [\nabla \theta_{vor}(\mathbf{r})]^2 = \frac{1}{2} \int d^2r n^2/r^2 = \frac{n^2}{2} \int_{r_0}^{L/2} d(\pi r^2)/r^2 = n^2 \pi \ln(r)|_{r_0}^{L/2}$$

where the limits are the smallest and largest length scales,  $L \gg r_0 \gg a$ ,<sup>17</sup> and the last expression, for large  $L$ , reduces to

$$\mathcal{H}_{vor} = n^2 \pi \ln(L/2r_0) \approx n^2 \pi \ln(L/r_0) = \pi \ln(L/r_0), \quad \text{with } n = 1 \quad (20)$$

Also, the presence of an isolated vortex increases the entropy of the system by an amount related to the number of possible positions that the vortex can be placed,  $\Delta S_{vor} = \ln(L/r_0)^2$ , [10, p. 117]. Then, the presence of an isolated vortex changes the Helmholtz free energy by

$$\Delta F = \mathcal{H}_{vor} - k_B T \Delta S_{vor} = \pi \ln(L/r_0) - 2k_B T \ln(L/r_0) = (\pi - 2k_B T) \ln(L/r_0) \quad (21)$$

Roughly speaking, at temperatures  $T < \pi/2(J/k_B)$ ,  $\Delta F > 0$ , the presence of a free vortex or anti-vortex is not welcomed because it increases the free energy. While in the high temperature region,  $T > \pi/2(J/k_B)$ ,  $\Delta F < 0$ , the free energy can be reduced by introducing more free vortex. So, there are two phases, a low-temperature one in which no free vortex is present and the other of high temperatures in which free vortex is favoured.

Although this is not truly the BKT phase transition, this very rough approximation serves giving a picture in low resolution. Further study comes with a note that the above reasoning was just accounted for a single vortex, i.e. the interaction between vortices in the field of each other hasn't been considered yet. Also, in the BKT theory, the energy of a pair of a vortex and an anti-vortex is finite which is a logarithmic function of the separation between the two.

### 2.2.2 Periodic boundary condition

In our simulation, periodic boundary condition is used, however, the question is **why is this condition applicable and how are the other conditions?** Demonstrated by (20), at low temperatures, the energy of an isolated vortex  $\mathcal{H}_{vor} + \mathcal{H}_{core}$  diverges with system's linear dimension  $L$  then at low temperatures, no isolated vortex may be present in a large system[10, p. 117]. At low temperature, the number of vortices must be equal to that of anti-vortices since if it was not the case, in the thermodynamic limit  $L \rightarrow \infty$ , the energy of either a single vortex or a single anti-vortex would be infinite and the energy of a finite subsystem of our system would be infinite that would be non-physical.

---

<sup>16</sup>Instead of  $d^2r = d(\pi r^2)$ , by Jacobian (determinant),  $d^2r = rdrd\phi$ , where  $0 \leq \phi \leq 2\pi$ ,  $r_0 \leq r \leq L$ .

<sup>17</sup> $r_0 \gg a$  is to make the continuum approximation to be sensible since in which we assumed  $a$  the lattice spacing to be infinitesimally small. See [10, p. 117]

Fortunately, the periodic boundary condition holds this property that can be explained mathematically by taking the closed path connecting the boundary lines and  $\oint_{\mathcal{P}} d\theta$  with the constraint of  $d\theta \in [-\pi, \pi]$  is equal zero. This is simple a consequence of the fact that the two opposite boundary lines are being traversed in opposite directions by the integral.

For other boundary conditions, e.g. the fixed boundary condition must come with some constraint to guarantee the property, because, take a look at some figure with the presence of a vortex at a boundary, e.g. [Fig. 17(a)], it can be seen that fixed boundary can prevent some two spins on this boundary get contact with some other two spins on the opposite boundary to form a vortex though they satisfy the definition, and that's why the property may not trivially hold.

Return to the periodic boundary condition, upto this stage, vortex-vortex interaction has not been considered, so that even though the given expression may not be the energy of an isolated vortex at high temperatures, with the boundary condition, it is neccessary that the number of vortices be equal that of anti-vortices at any finite temperature.

## 3 Algorithms

### 3.1 Metropolis

In a serial program, Metropolis with random proposal in which a spin site is selected randomly to suggest a change was used while in a parallel program, checker board proposal was used instead, i.e. the lattice is divided into two sublattices that are flipped consecutively.

Each update step in Metropolis random proposal consists of the following steps:

- × Choosing a random vector  $\mathbf{r} \in \mathcal{S}_1$ , i.e. an angle  $r \in [0, 2\pi)$  is randomly selected.
- × Select a spin site  $\mathbf{s}_{ij}$  at random by calling 2 uniform random number  $r_1, r_2$ , i.e.  $(r_1, r_2) \in U(0, 1)$  and  $i = r_1 L$ ,  $j = r_2 L$ , [12, p. 633] taking the integral part. The indexes run from 1 to  $L$ , since  $(r_1, r_2) \in [0, 1]^2$ . The program should prevent the case of position outside the lattice, despite that the case of random number equal unity is rare.
- × Compute the energy difference  $\Delta\mathcal{H}$  with the spin flip operation is generalized as in Ref. [13], As later on shown, the trial spin has angle  $\theta'_{ij} = 2r - \theta_{ij} + \pi$ . Note that in simulation,  $\theta'_{ij}$  must be bound in the interval  $[0, 2\pi)$ .
- × If  $\Delta\mathcal{H} \leq 0$  flip the spin, i.e. accept the proposal configuration.
- × If  $\Delta\mathcal{H} > 0$  generate an  $r_3 \in U(0, 1)$ , if  $r_3 < \exp(-\Delta\mathcal{H}/k_B T)$  flip the spin.

The time step  $\delta t$  between adjacent configurations to take measurement can be  $\delta t = 1/(J/k_B)N$ , i.e. every Monte Carlo move (every attempted spin flip) is included in the calculation, however, especially for the case of small scale single spin flip algorithm, adjacent configurations are more correlated as  $N$  is larger, a rough estimate would be  $N$  spin flips to safely escape from correlation. Then it is common to choose  $\delta t = 1$

corresponding to one Monte Carlo step per spin (i.e.  $N$  steps per system of  $N$  spins) [7, p. 187] and  $\delta t$  is termed a Monte Carlo update sweep or sweep for short.

In checker board proposal, each spin site and its four neighbouring spin sites are copied to a thread in charge of flipping that spin with the same rule as in random proposal.

### 3.2 Wolff

Each cluster update step consists of the following operations:

- × Choose a random vector  $\mathbf{r}$  and a random lattice spin  $\mathbf{s}_{ij}$  as the first element of a cluster.
- × Flip the spin  $\mathbf{s}_{ij} \rightarrow \mathbf{R}\mathbf{s}_{ij}$  and mark it.
- × Add all the nearest neighbours of  $\mathbf{s}_{ij}$  to a trial set. Each spin in the set is added to the cluster and marked with probability

$$\mathcal{P}(\mathbf{s}_{ij}, \mathbf{s}_{lm}) = 1 - \exp(\min\{0, -J_{ijlm} [\mathbf{s}_{ij}^T (\mathbf{R}\mathbf{s}_{lm}) - \mathbf{s}_{ij}^T \mathbf{s}_{lm}] / k_B T\}) \quad (22)$$

$$= 1 - \exp(\min\{0, 2J_{ijlm} \cos(\theta_{ij} - r) \cos(r - \theta_{lm}) / k_B T\}) \quad (23)$$

since, from (102),

$$\begin{aligned} \mathbf{s}_{ij}^T (\mathbf{R}\mathbf{s}_{lm}) - \mathbf{s}_{ij}^T \mathbf{s}_{lm} &= \mathbf{s}_{ij}^T (\mathbf{s}_{lm} - 2\mathbf{r}\mathbf{r}^T \mathbf{s}_{lm}) - \mathbf{s}_{ij}^T \mathbf{s}_{lm} \\ &= -2(\mathbf{s}_{ij}^T \mathbf{r}) \mathbf{r}^T \mathbf{s}_{lm} = -2 \cos(\theta_{ij} - r) \cos(r - \theta_{lm}) \end{aligned}$$

(23) is interpreted as that, if the proposal bonding has higher energy than the bonding between the spin in the cluster with the original neighbour, then the bond has zero probability, and if the bonding lowers the energy, then the neighbour is activated with probability  $0 < \mathcal{P}(x, y) < 1$ . Note that, in the expressions of proposal probability, spin  $\mathbf{s}_{ij}$  has been already flipped.

- × If a neighbour of  $\mathbf{s}_{ij}$  is flipped, it is then be treated as the same as the spin in the cluster, i.e. its nearest neighbours are collected to the trial set to be explored and the update sweep stops when no more spin is added and every spin in the trial set is explored.
- × In implementation, it must be made clear that each group of four neighbours is associated with the spin who introduces them to the trial, that spin is termed *referee*, and only unmarked neighbour is referred. The trial probability involves two spins participating in a bond which in this study is of nearest neighbour bond.

A number of  $N/\langle |c| \rangle$  Wolff update sweeps is equivalent to a standard Monte Carlo update sweep. Suppose  $\tilde{\tau}_W$  is an amount of time measured in unit of Wolff sweep, then the equivalent amount in standard update sweep unit is

$$\tau_W = \tilde{\tau}_W \langle |c| \rangle / N \quad (24)$$

where  $\langle |c| \rangle$  is the mean size of flipped clusters which is computed by first identifying all clusters in the system, each cluster  $c_{SW}$  has a size (i.e. the number of spins inside)  $|c_{SW}|$  and an associated probability of finding it  $|c_{SW}|/N$ , and then being given by

$$\langle |c| \rangle = \left\langle \frac{1}{N} \sum_{c_{SW}} |c_{SW}|^2 \right\rangle \quad (25)$$

where the average sign in the right side  $\langle \dots \rangle$  denotes thermal average, and the sum is taken over every cluster identified in the system.

## 4 Observables

### 4.1 Equilibration

In Ref. [5], the authors used heatbath algorithm and produced the following results which hold for  $T \leq T_{BKT}$ . By a cold start, i.e. a ground state configuration with all spins having the same state, the correlation length, which will be studied shortly, is related to simulation time as

$$\xi(t) \sim t^{1/z}, \quad (26)$$

while for a quench from high temperature to the temperatures below criticality, the relation is

$$\xi(t) \sim (t/\ln t)^{1/z} \quad (27)$$

where  $z$  is the usual dynamic exponent characterizing temporal correlations in equilibrium, which is  $z = 2$  for this local algorithm.

Since, in the low temperature phase,  $T \leq T_{BKT}$ , the correlation length is infinite[14], for finite lattices, the maximum value of  $\xi = L$ , the two relations become

$$L \sim t^{1/2} \quad (28)$$

and

$$L \sim (t/\ln t)^{1/2} \quad (29)$$

The following table lists the number of sweeps used for thermalization for different lattice sizes using local update Metropolis algorithm, the seed number is 19880106, and 20112015 first random numbers are discarded. It must be noted that these formulae

Table 1: Theoretical consistency suggests the number of update sweeps should be used for rapid quench from a hot configuration using some local update algorithm.

Type	$L = 4$	$L = 6$	$L = 8$	$L = 16$	$L = 32$	$L = 64$	$L = 128$
Hot start	70	190	380	1950	9500	44000	$2 \times 10^5$

are just show proportionalities, not equalities.

In some Ref.<sup>18</sup>, the thermalization is characterized by a time  $\tau_A^{(nl)}$  associated with a nonlinear relaxation function associated with observable  $A$  as

$$\psi_A^{(nl)}(t) = \frac{\langle A(t) \rangle_{T,\mathbf{H}} - \langle A(\infty) \rangle_{T,\mathbf{H}}}{\langle A(0) \rangle_{T,\mathbf{H}} - \langle A(\infty) \rangle_{T,\mathbf{H}}} \quad (30)$$

where the averages are thermal average and<sup>19</sup>

$$\tau_A^{(nl)} = \int_0^\infty \psi_A^{(nl)}(t) dt \quad (31)$$

Since  $\langle A(t) \rangle_{T,\mathbf{H}} = \langle A(t) \rangle_{T,\mathbf{H}}$ , and then the infinity limit is realized as it is large enough such that either  $\psi_A^{(nl)}(t) = 0$  or (31) saturates somehow. According to this author the equilibration time is much more greater than  $\tau_A^{(nl)}$ .

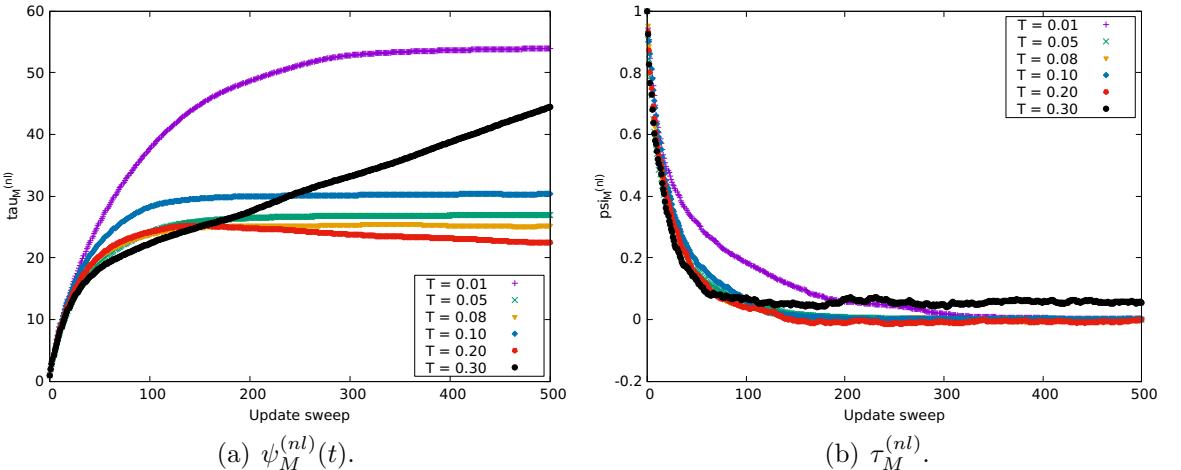


Figure 4:  $L = 8$ . Average taken on hot configurations.

In [Fig. 4(a)], it appears that the decorrelation time is longest for  $T = 0.01(J/k_B)$ , while for this temperature  $\tau_A^{(nl)} \sim 55$ . This is very small compared with the amount of time needed to reach equilibrium for some stream of random numbers. For  $T = 0.30(J/k_B)$ , no saturation is observed. By looking further at [Fig. 4(b)], at this temperature, the nonlinear function never reach zero in this range, and the sum (31) should be truncated at  $\sim 150$  update sweeps and then the nonlinear time is of the same order of magnitude with the some others.

Up to now, it is not easy to understand what does the author mean by much more greater since even of ten times greater than these  $\tau_A^{(nl)}$ , the next section shows that for some stream of random numbers these results as well as the suggestions in [Tab. 1] are still not enough to reach equilibrium in regards of metastability. Also, why the sum (31) though it is similar to the integrated autocorrelation time that will be introduced soon.

<sup>18</sup>Read. Ref. K. Binder and D. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction*. Graduate Texts in Physics. Springer Berlin Heidelberg, 2010.

<sup>19</sup>Trying to understand why the author defined this time. A reasonable explanation may be that, the nonlinear relaxation function is typically expected to decay exponential and so the time is the sum as in the definition of exponential autocorrelation time as will be soon mentioned.

In the following parts we argue for one thing, starting from a cold configuration helps escape from metastability. This is physics. Low temperature phase doesn't prefer vortex since its appearance gives an extra amount of energy beside that contributed by the spin wave. High temperature phase welcomes vortex that dominates spin wave.

#### 4.1.1 Metastability

Experience with strange results was the reason led us to study metastability, since we were to take thermal average [Fig. 5], G stands for gradual, A anneal, C cold, H hot, R random.

Some metastable configurations are demonstrated in [Figs. 6, 7]. In [Fig. 6], two processes Hot 1 and Hot 2, after a transient stage, the system is in the exact equilibrium at which the energies agree. However, the energy given by process Hot 3 at first doesn't agree with these that the system experiences metastable states before attaining the true equilibrium. These metastable states come with non-physical vortex-anti-vortex pairs, as shown in [Fig. 6(a),(b)], although many update sweeps have been passed, however, the system doesn't seem to change much. Similar discussion should be presented for [Fig. 7] in which the metastable states come with neither vortex nor anti-vortex.

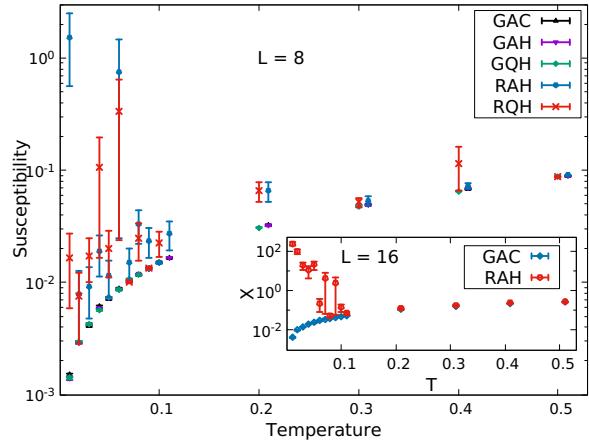
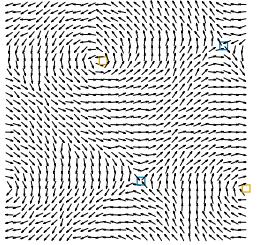
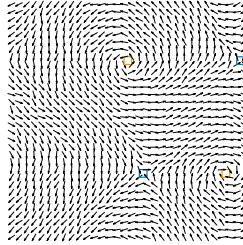


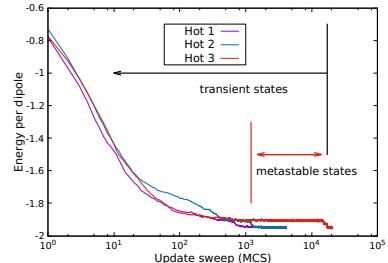
Figure 5: Thermal average of magnetic susceptibility with the norm of magnetization was used rather than the magnetization itself. Different initial conditions as well as different equilibration processes.



(a) At 4096 MCS.



(b) At  $3 \times 4096$  MCS.



(c) Energy per spin.

Figure 6:  $L = 32, T = 0.01(J/k_B)$ . First and middle figure come with the line Hot 3 in the last figure.

#### 4.1.2 Standard Metropolis

As reported in [Fig. 9(a)], there is a case in which the equilibration time is larger than the theoretical expectation (if the relation were equality not proportionality), i.e. *Hot 1* in the figure although they should be consistent on average since most of the others appear starting to converge at the same

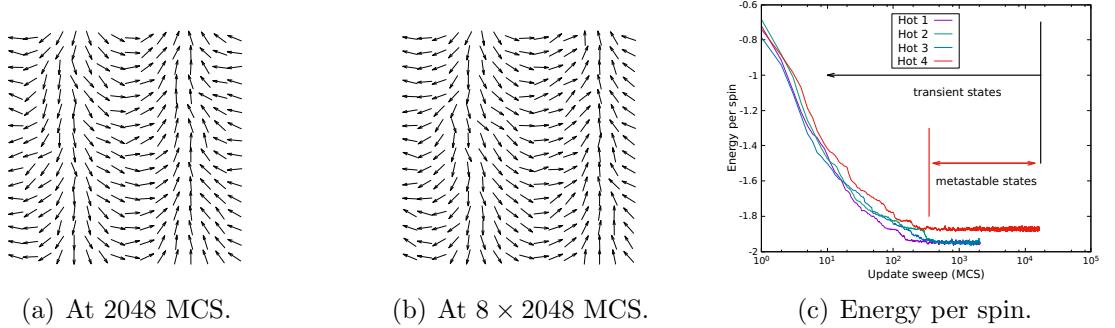


Figure 7:  $L = 16, T = 0.01(J/k_B)$ . First and middle figure come with the line Hot 4 in the last figure.

time that is shorter. And also, there is a case, *Hot 7*, which shows that the evolution meets a meta stable state, i.e. the data are steady but do not converge similarly to the others. Then we investigate other quantities too that shown in [Figs. 9(b), 8(a), 8(b)].

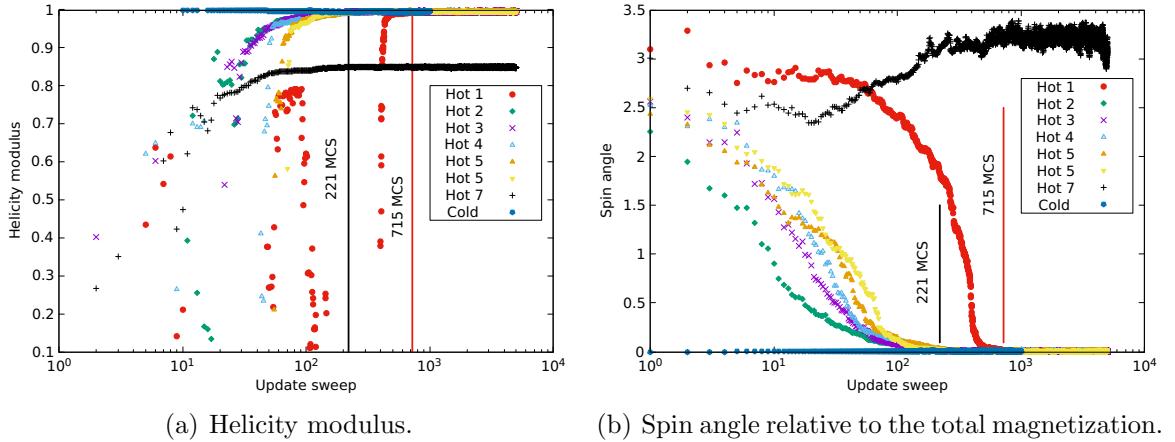


Figure 8:  $L = 8, T = 0.02(J/k_B)$ .

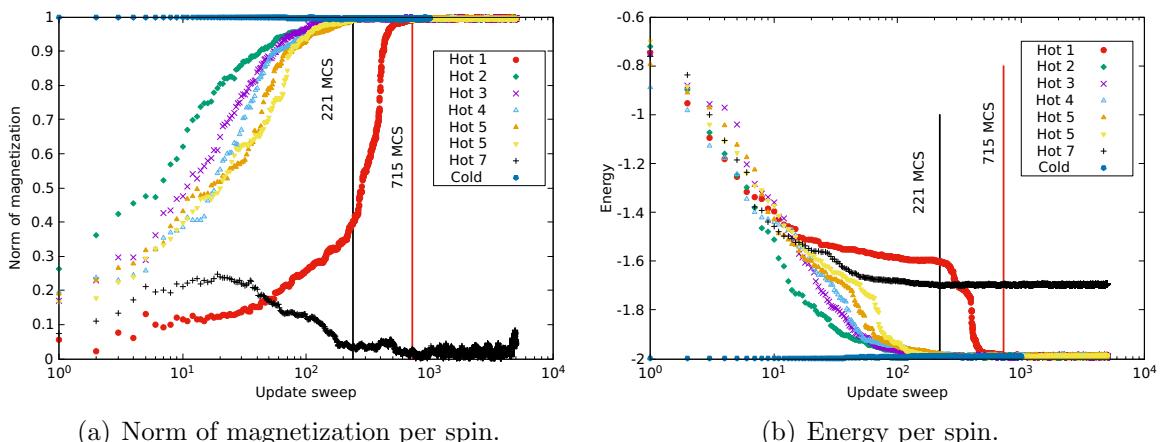


Figure 9:  $L = 8, T = 0.02(J/k_B)$ .

How to avoid metastability? This may be solved by taking a gradual quench or gradual anneal rather than rapid ones, given in [Fig. 10(a)].

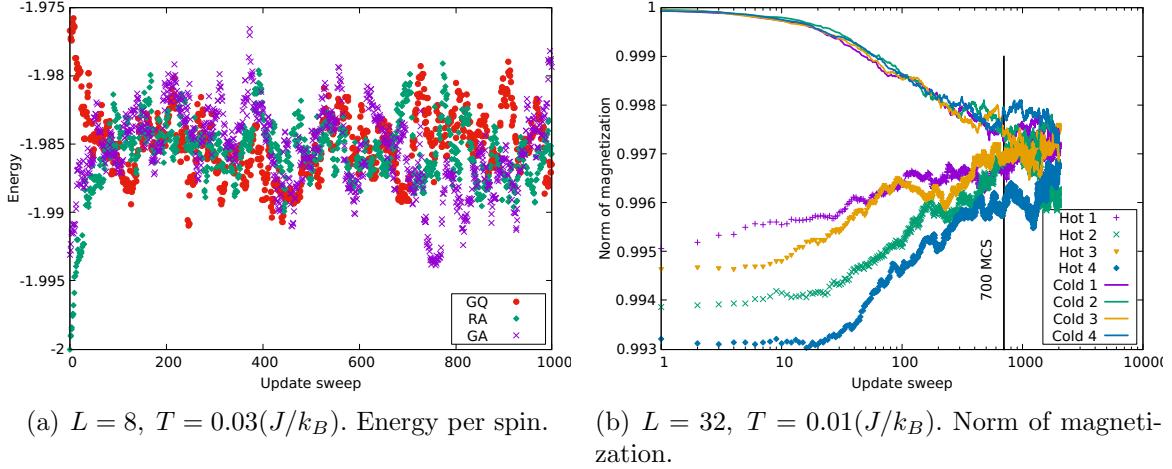


Figure 10: Hot is gradual quench GQ, cold is random anneal RA. GA for gradual anneal.

By [Fig. 10(a)], which is shown for energy per spin, a very good agreement is made for this quantity. It's a bit ugly for magnetization per spin the norm in [Fig. 10(b)], however, every instance still agree to equilibrate after about 700 update sweeps. The result for energy and helicity modulus also agree but the equilibration time is much smaller while that for phase angle per spin agree with norm of magnetization per spin. Note that for the case of gradual quench, all instances are from  $T = 0.02(J/k_B)$ . The initial start of the instances appear that they are from different temperature, but it should be the problem of fluctuation nature. In [Fig. 11(a)], it appears that the

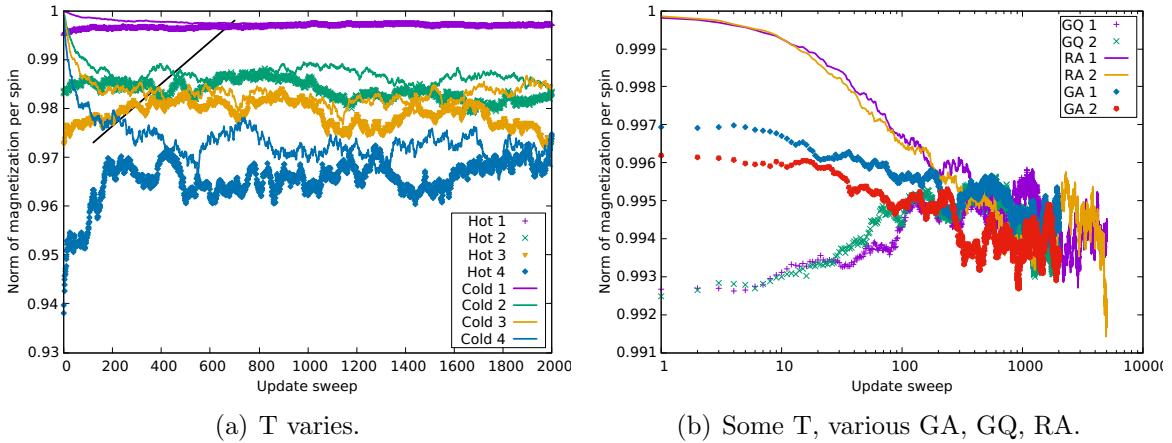


Figure 11:  $L = 32$ . Norm of magnetization per spin. Hot is gradual quench GQ, cold is random anneal RA. GA for gradual anneal.

equilibration time decreases as the temperature decreases though it's not very clear due to statistical fluctuation.

The last consistency check is given in [Fig. 11(b)], which shows agreement between three types of equilibration processes. The data for some process appears not be in

agreement, however, by taking more update sweeps, it is then be able to conclude that they agree to equilibrate at the same time.

In summary, every simulation will be thermalized with a random anneal from a cold configuration. Consistency checked to assure the system has equilibrated correctly with:

- × Gradual quench and gradual anneal.
- × Different part of a single stream of random numbers.
- × Different quantities.

#### 4.1.3 Standard Wolff

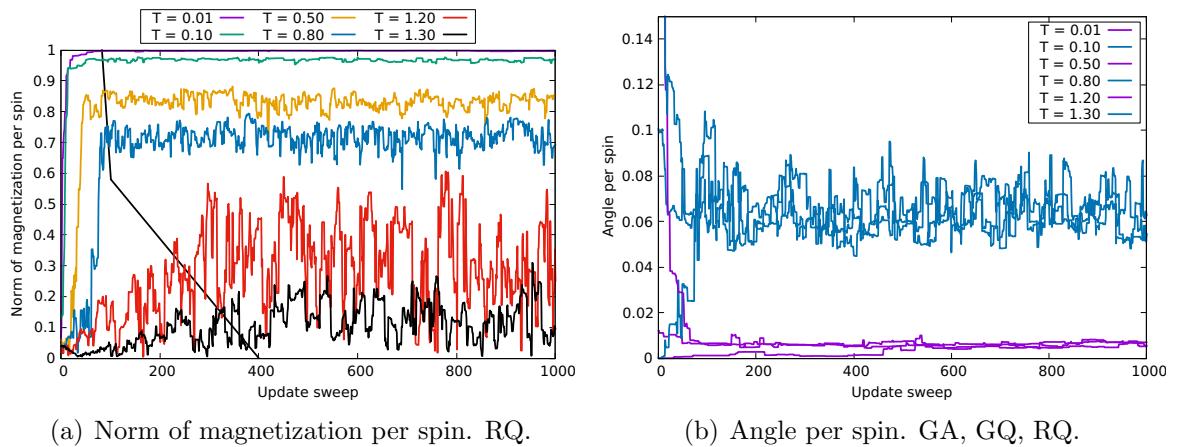


Figure 12:  $L = 32$ . GA gradual anneal, GQ gradual quench, RQ rapid quench.

It appears that the equilibration time increases as the temperature increases, however, this is not true since the temporal unit is different, in each simulation, the unit is the cluster size. The [Fig. 12(a)], the system experiences a rapid quench from a hot configuration, no metastable state is present. This should be the nature of a cluster algorithm in which a cluster of some size is being flipped not a single spin as in local dynamics.

In [Fig. 12(b)], data for rapid quench, gradual anneal and gradual quench agree with each other of equilibrium.

Roughly speaking, measurements taken on configurations generated by Wolff dynamics seem to be less correlated compared with standard Metropolis.

#### 4.1.4 Parallel Metropolis with CUDA, OpenMP directives

Two random number generators were used, one is the built-in XORWow, and the other is XORShift<sup>20</sup>. There are true random numbers that are generated from some physical source like electric current in resistor mainly used in cryptography. However, those are slowly generated that are not suitable for simulation. Then pseudo random numbers are used, those are generated via some deterministic algorithm with a requirement of reproducibility.

---

<sup>20</sup>Ref. Martin Weigel. *J. Comp. Phys.*, 231(8):3064 – 3082, 2012.

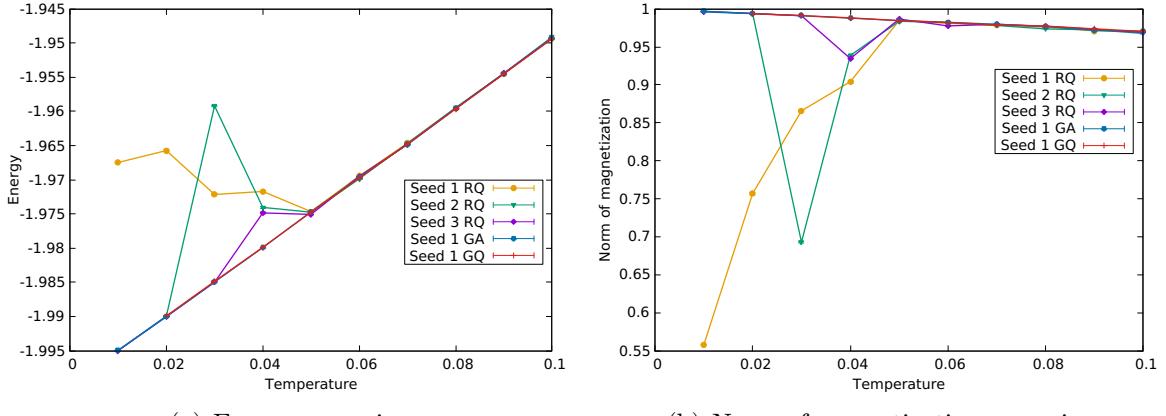
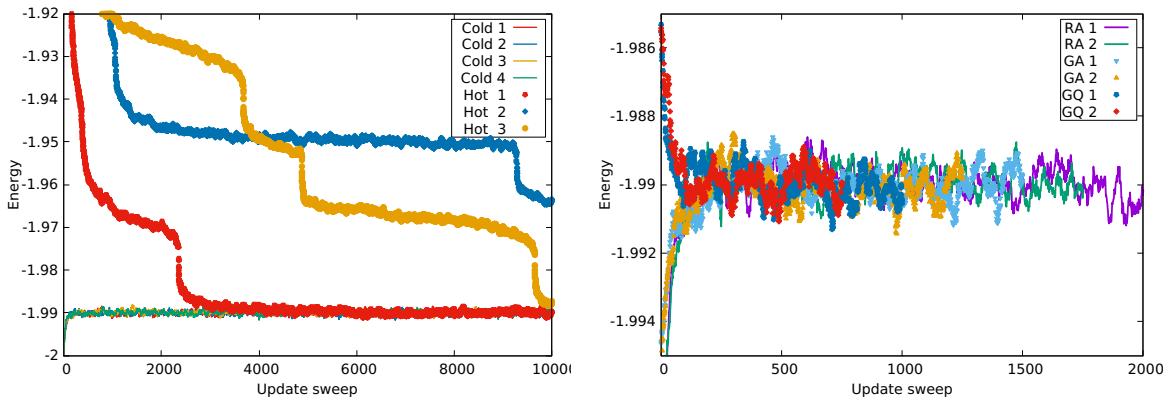


Figure 13: Odd results.  $L = 32$ . The seeds are Seed 1 = 19880106, Seed 2 = 19880601, Seed 3 = 34789453. Abbreviations: RQ rapid quench, GA gradual anneal, GQ gradual quench.

Similar to [Fig. 5], we encountered weird results, in [Fig. 13(a), 13(b)] the phenomenon of metastability also shown in GPGPU programming. At first unexpected results shown for some low temperatures, then changing seed from 19880106 to be 34789453 and these results disappear returning the consistent ones. It was then concluded that the quality of random number generator is bad. However, looking at the kernel which initializes the spin configuration, it was hot meaning that we are not quenching gradually. Up to now Mersenne Twister as well as a version of XOR random number generators have been used, the quality of random numbers should not be the reason for unexpected results.



(a) Different conditions realized by different parts of the same stream of random numbers for each case of either hot or cold. (b) Rapid anneal, gradual anneal and gradual quench have results that coincide.

Figure 14:  $L = 32$ ,  $T = 0.02(J/k_B)$ .

A comparison between annealing and quenching given in [Fig. 14(a)], it takes long time for a rapid quench to reach correct result which is achieved soon by all annealing processes. By trying different streams of random number, it is observed that no metastable state present for an anneal process. We also make a comparison between

these anneals with gradual quenches and to provide more firmly that no metastable state occurs with any anneal process, a comparison of those with gradual anneals given in [Fig. 14(b)].

In these considerations, we have tried to locate metastability for many values of temperatures. Although they are limited, a conclusion may be made is that, no metastability occurs if a random anneal, a gradual anneal or a gradual quench is made. Gradual anneal or gradual quench has its physical meaning. However, for large lattice sizes, it is not possible to run a whole range of temperatures, and also the temperature steps should not be taken uniformly, then a random anneal will be applied, and to be safe, several instances should be tested to assure the exact equilibration if applicable. Nevertheless, starting from a cold configuration, the energy of the system is being raised, the acceptance probability would be low since either Metropolis or Wolff algorithm would prefer the direction of lowering the energy, fortunately we didn't observe this.

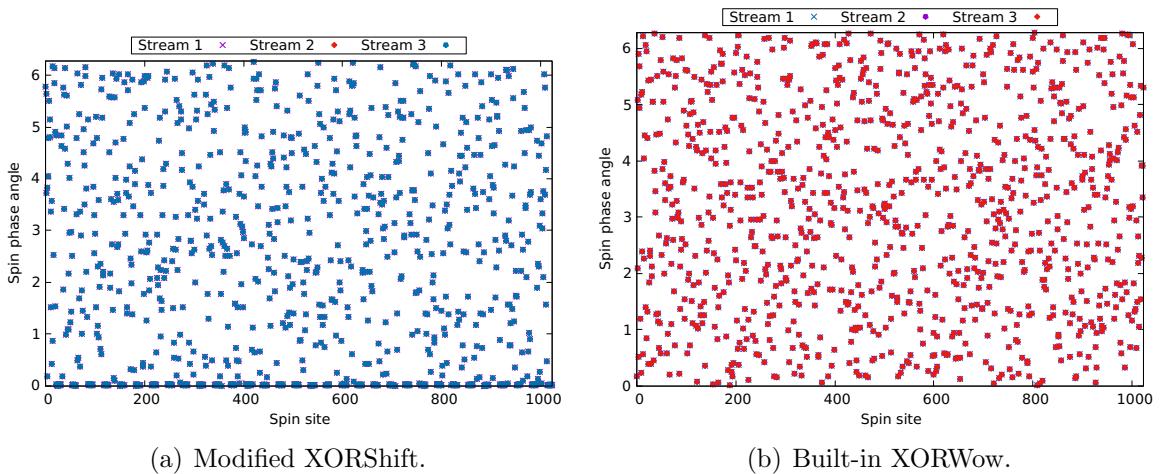


Figure 15:  $L = 32$ . Initialization of spin phase angles.

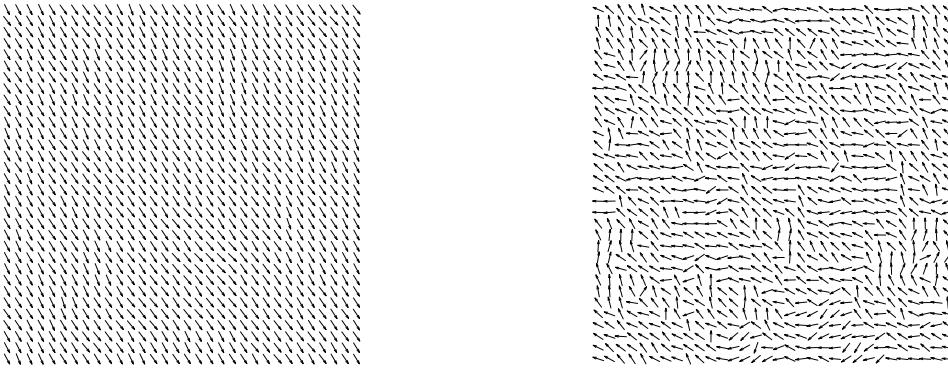
In our work, we encountered a problem of different streams of random numbers generated with the same initial seed that we later found a bug in our source code in the implementation of XORShift. Unfortunately, the quality of the generated random numbers is not good. In [Fig. 15(a)], the random numbers doesn't appear to be uniformly distributed, some strong concentration about zero, while in [Fig. 15(b)], those generated by the built-in XORWow generators seems perfectly uniformly distributed. We haven't figured out the reason or any bug that may present in our modified XORShift yet. Also, we have just figured out a bug in the share kernel of metropolis, since it yields a slightly different results compared with the global kernel.

The final remark for this section, the equilibration time is largest for the norm of magnetization per spin, the square of magnetization and the spin phase angle agree with it, that for helicity modulus and energy is much smaller. Our program allows plots of all those quantities be shown automatically. We found the following upper limit of equilibration time used in our simulation. Note that in order to get these results, in the stage of measuring thermalization, we have run our program as long as possible to make our conclusion of short equilibration solid. E.g. deploying  $10^4$  steps for thermalization to make a conclusion of 500 is enough. IMPORTANT! Equilibration time comes with its stream of random numbers, more specifically, the whole program used only one seed.

For reproducibility purpose, in standard Metropolis on CPU, 2000 update sweeps and 4000 update sweeps were used for equilibration with lattice 16, 32 respectively. For XORWow, the maximum number of equilibration steps are correspondingly 2048, 4096, 8192, and 16384 for  $L = 32, 64, 128$  and 256.

## 4.2 Phenomenology

In [Fig. 16(a)], at  $T = 0.01(J/k_B)$  the system is in a spin wave state, no vortex is present. At higher temperatures, e.g.  $T = 0.50(J/k_B)$ , trying different instances of random numbers, no state of vortices [Fig. 16(b)] is observed after the system has reached equilibrium.



(a)  $T = 0.01(J/k_B)$ . Dominated by spin-wave. (b)  $T = 0.50(J/k_B)$ . Excited spin-wave.

Figure 16:  $L = 32$ . Spin configurations.

At  $T = 0.70(J/k_B)$ , after the system is fully equilibrated to a state of no vortex, a state with vortices [Fig. 18(a)] is observed at 5984 update sweeps and also at some larger sweeps. Is this a metastable state? No, a metastable state is the one that differs from the exact equilibrium and last for a long time or slowly being changed while this state doesn't persist for some time but instead is found rarely. See [Figs. 6, 7] for metastable states.

As the temperature is increased further, after the system has been equilibrated, states with vortices occur more frequently even though the system was confirmed to have a equilibrium state of no vortex. E.g.  $T = 0.85(J/k_B)$ , configurations with vortices appear very frequently while those with no vortex hardly occur [Fig. 17(a)] and at higher temperatures the exact equilibrium has vortices [Figs. 17(b), 23(a), 23(b)].

Therefore, it is suggested that, in this region, there is no statistical difference between a state with vortex and a state with no vortex, more rigorously, this is the transition between a state with no vortex to a state with tightly bound vortices. This is also permitted by theoretical prediction that the energy of a pair of vortex and anti-vortex is finite. By observing the states with vortex that come along with sates of no vortex in equilibrium, the agreement is found by noting that the vortices come in pair with anti-vortex and they are tightly bound. See [Figs. 18(a), 18(b)]. So, at high temperatures below the transition, a vortex tightly binds with an anti-vortex because the energy of a vortex-anti-vortex pair is lowest if they are not separated by large distance since[4, p. 1193]  $E_{\text{pair}} \sim \ln |\mathbf{r}_s/r_0|$ .

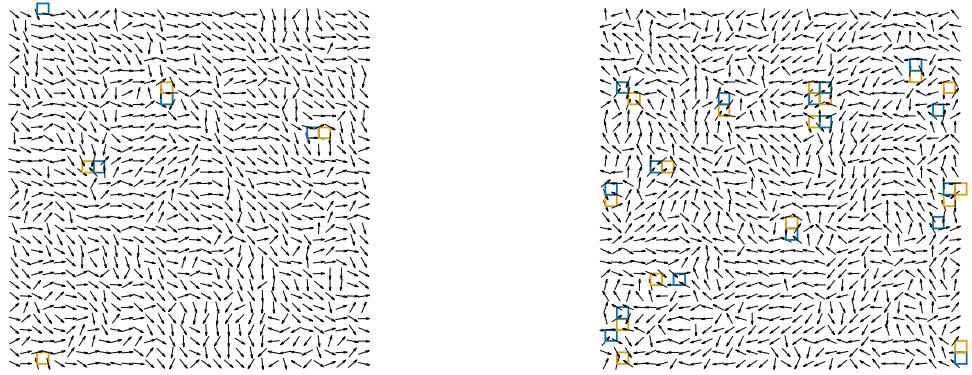


Figure 17:  $L = 32$ . Spin configurations. At  $T = 0.85(J/k_B)$ , vortex and anti-vortex are tightly bound. At  $T = 1.00(J/k_B)$ , 4 pairs unbind.

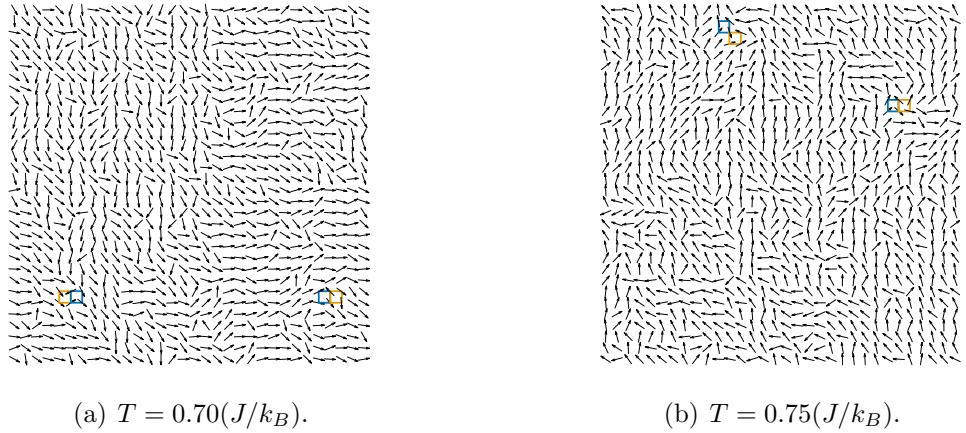


Figure 18:  $L = 32$ . Pairs of vortex and anti-vortex are tightly bound.

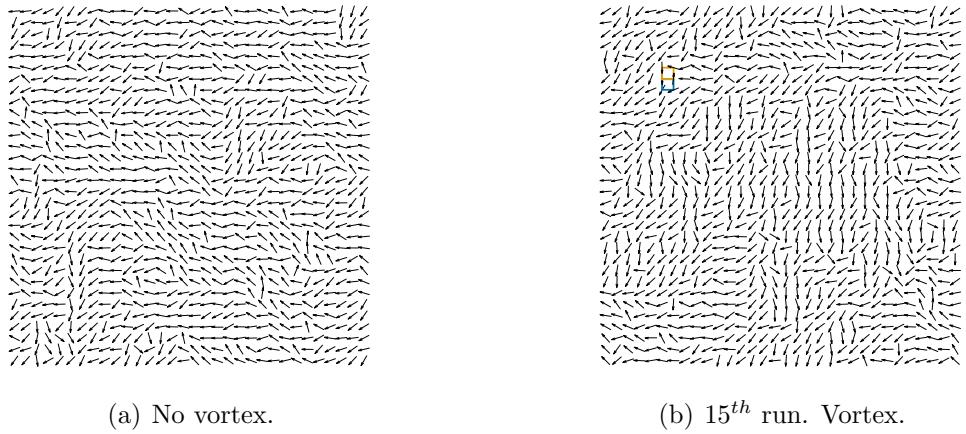
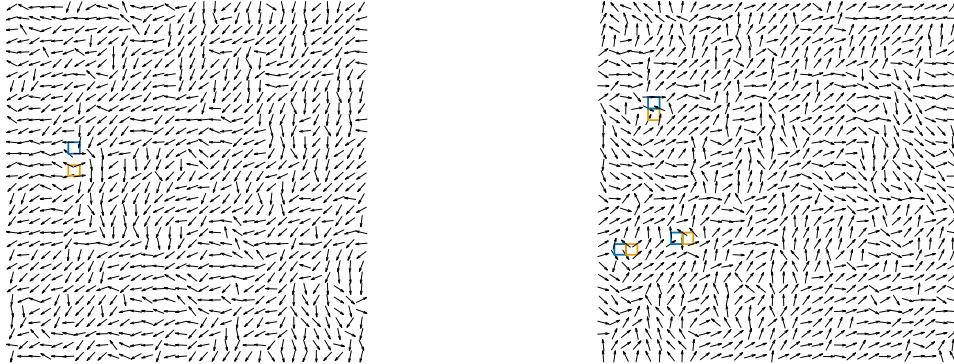


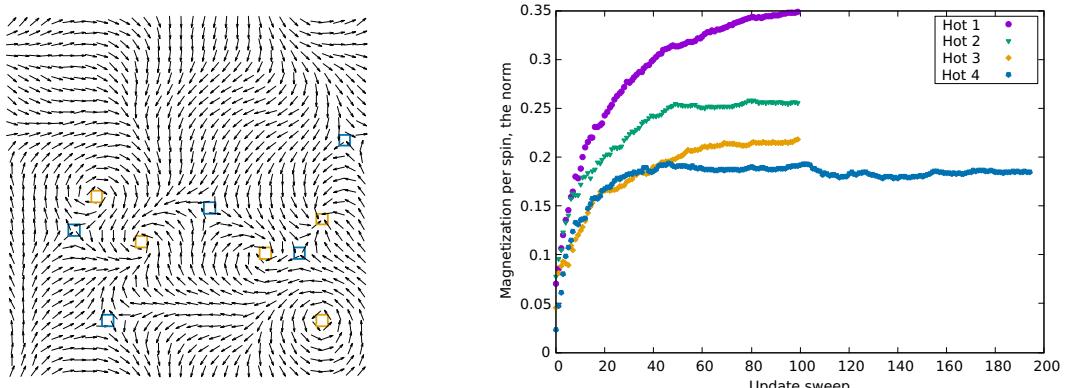
Figure 19:  $L = 32$ ,  $T = 0.60(J/k_B)$ . Spin configurations. In the left, no vortex, purely spinwave. In the right, a vortex pair occurs.

Take a higher resolution to the region in which vortex starts to occur in pairs of vortex and anti-vortex. This is due to the fact that although the energy of a vortex diverges with system size, the energy of a pair of vortex and anti-vortex is finite. Take a look at [Fig. 19(a), 19(b)], this is the region of temperatures at which the power law decay of spin-spin correlation function is modified by the presence of tightly bound vortex-anti-vortex pairs.



(a)  $T = 0.60(J/k_B)$ , free vortex is not physics here. (b)  $T = 0.80(J/k_B)$ , tightly bound vortex pairs.

Figure 20:  $L = 32$ . Spin configurations.



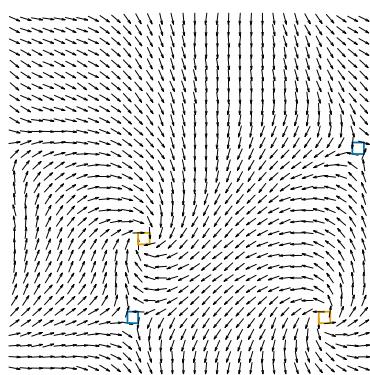
(a) Vortices are present in a transient state when the system experiences a rapid quench. (b) Several parts of a single stream of random numbers are used to seek for metastable state.

Figure 21:  $L = 32$ ,  $T = 0.01(J/k_B)$ . Spin configurations. The exact equilibrium should be the one in which the norm of magnetization (small sized lattice) very close to value of unity.

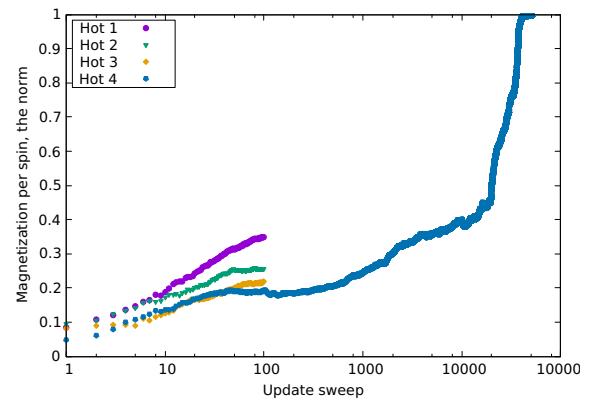
At  $T = 0.60(J/k_B)$ , in the total of 15 runs on a single stream of random numbers, we found at the 15<sup>th</sup> a configuration with a tightly bound vortex pair while all the preceded configurations are purely spinwave, except one configuration at the 4<sup>th</sup> run reported in [Fig. 20(a)] which shows a dissolved pair in which the vortex and the anti-vortex are not tightly bound. This is not physics and must be due to the quality of the random number generator. Let's remind that the seed used is 19880601, warming up for 20112015 random numbers, each run was on 2000 random numbers, this is for the

purpose of reproducibility. The BKT transition can't occur at such a low temperature, this temperature is expected to be in the region at which the power law decay of spin-spin correlation function being modified with the presence of tightly bound vortex pairs.

In order to solve this contradiction, let's investigate the system at a bit higher temperature, e.g.  $T = 0.65(J/k_B)$ , the 5<sup>th</sup> has a vortex pair that is tightly bound, many other runs don't show neither vortex nor unbound vortex pair. A bit higher temperature  $T = 0.70(J/k_B)$ , in the total of 10 runs, half of those was found to have tightly bound vortex pairs, and the other half was purely spinwave, and no instance is found to have vortex pairs that are not tightly bound. At  $T = 0.80(J/k_B)$ , 8 runs have tightly bound vortex pairs in the total of 10 runs, and no sight of free vortex [Fig. 20(b)]. Our point is that if the case of  $T = 0.60(J/k_B)$  with free vortices were physically correct, at neighbouring higher temperatures, such a phenomenon would occur more frequently.

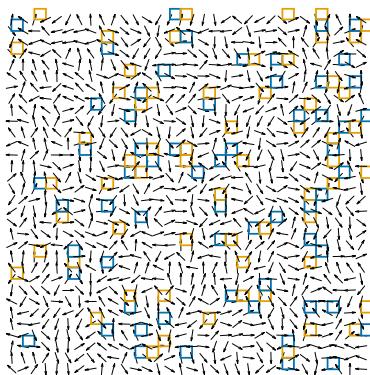


(a) Spin configurations

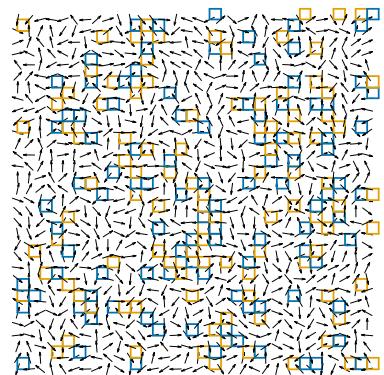


(b) The magnetization per spin at 7000 sweeps.

Figure 22:  $L = 32$ ,  $T = 0.01(J/k_B)$ . The system stays remain in a metastable state even at about 7000 update sweeps, and be fully equilibrated at upto about 41 thousand sweeps.



(a)  $T = 1.50(J/k_B)$ .



(b)  $T = 2.00(J/k_B)$ .

Figure 23:  $L = 32$ . Spin configurations. At  $T = 1.50(J/k_B)$ , many pairs unbind, and more vortex pairs occur but less are tightly bound at  $T = 2.00(J/k_B)$ .

In [Fig. 21(a)], it appears that the system favours vortices even at very low temperature, however, this is only a transient state because, theoretically, by (20), the present of a vortex increases the energy of the system and practically, that is observed in [Fig. 21(b)]. Even after 7000 update sweeps, the system is still in the transient stage shown in [Fig. 22(a)], by equilibrating for longer time, the exact equilibrium is reached after about 41 thousand sweeps [Fig. 22(b)] and the system is dominated by spin wave as predicted by Bloch in [Fig. 16(a)].

At high temperatures, e.g.  $T = 1.50(J/k_B)$  and  $T = 2.00(J/k_B)$  more and more vortex occurs and many of them are free, i.e. are not in tightly bound pairs with anti-vortex.

## Summary

The following remarks are resulted from these phenomenological investigations.

- × In really low temperature region  $T < T_0$ , the system is dominated by spin wave, configurations with vortices are just metastable.
- × At higher temperatures  $T_0 < T < T_1$ , the exact equilibrium can have vortex or not, there is no statistically difference between a state with vortex and a state with no vortex. The crucial point is vortex comes in tightly bound pairs with anti-vortex.
- × At some temperature  $T_1$ , tightly bound vortex pairs start to dissociate.
- × In high temperature phase  $T > T_1$ , free vortex and free anti-vortex dominate.

Are there two phase transitions, one is at  $T_0$  and the other is at  $T_1$ ? Since the phase transition is the phenomenon in which the system changes its response to externally applied field, it will be shown that the susceptibility diverges at the latter temperature signatures the special phase transition.

## 4.3 Let's measure

### 4.3.1 Spin-spin correlation

Consider the equal-time spin-spin correlation function of spin observable  $s$  at time  $t$  which relates two spins  $\mathbf{s}_{ij}(t)$  and  $\mathbf{s}_{lm}(r, t)$  separated by a distance  $r$  as

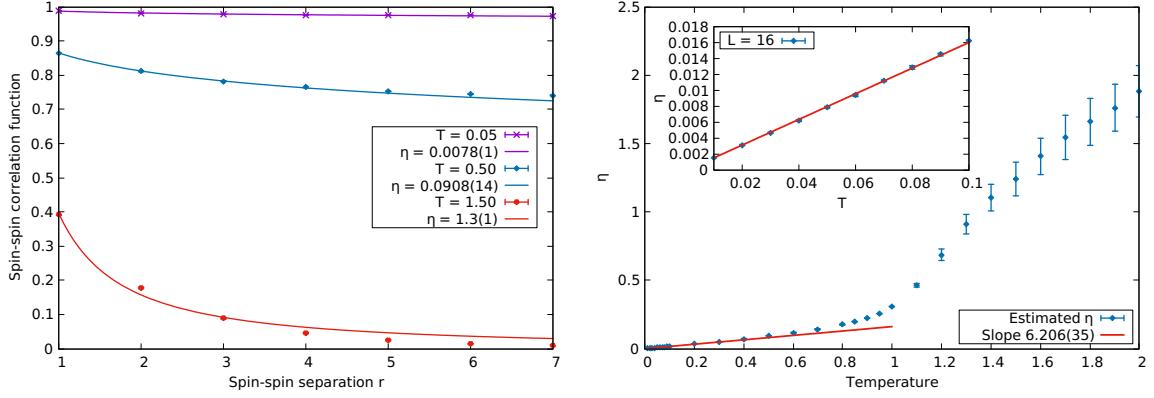
$$\vartheta(r, t) = \langle \mathbf{s}_{ij}(t) \mathbf{s}_{lm}(r, t) \rangle - \langle \mathbf{s}_{ij}(t) \rangle \langle \mathbf{s}_{lm}(t) \rangle \quad (32)$$

The averages are taken over spin sites and samples, then  $\langle \mathbf{s}_{ij}(t) \rangle = \langle \mathbf{s}_{lm}(r, t) \rangle = \langle \mathbf{m}(t) \rangle$  and (32) becomes

$$\vartheta(r, t) = \langle \mathbf{s}_{ij}(t) \mathbf{s}_{lm}(r, t) \rangle - \langle \mathbf{m}(t) \rangle^2 \quad (33)$$

where  $\mathbf{m}(t)$  is the magnetization per spin. Since  $\langle \mathbf{m}(t) \rangle = 0$ , then (33) becomes

$$\vartheta(r, t) = \langle \mathbf{s}_{ij}(t) \mathbf{s}_{lm}(r, t) \rangle = \langle \cos [\theta_{ij}(t) - \theta_{lm}(r, t)] \rangle \quad (34)$$



(a) The function and the power law fit (36).

(b)  $\eta$  vs  $t$  and the fit  $\eta = k_B T / 2\pi J$ .

Figure 24:  $L = 16$ . Spin-spin correlation. Slope  $6.206(35)$  agrees with  $2\pi \approx 6.283\dots$  in three sigma. This low accuracy is due to small lattice size. But note that there's a logarithmic correction mentioned in Ref. [14].

and being estimated by

$$\vartheta(r, t) = \frac{1}{2N} \sum_{ij} \{\cos [\theta_{ij}(t) - \theta_{lmd}(r, t)] + \cos [\theta_{ij}(t) - \theta_{lmr}(r, t)]\} \quad (35)$$

where  $lmd$  and  $lmr$  denotes two spin sites of distance  $r$  apart from site  $ij$  in the two directions that are orthogonal in our two-dimensional lattice, i.e. chosen to be rightward and downward.

An additional thermal average can be applied.  $\vartheta(r, t)$  gives information of how strong a spin influences its distant neighbours. [8, p. 216] According to the prediction of K-T theory[6, p. 19], at low temperature,  $\vartheta(r, t)$  decays algebraically for  $r \gg r_0$ ,

$$\vartheta(r, t) \sim r^{-\eta} \quad (36)$$

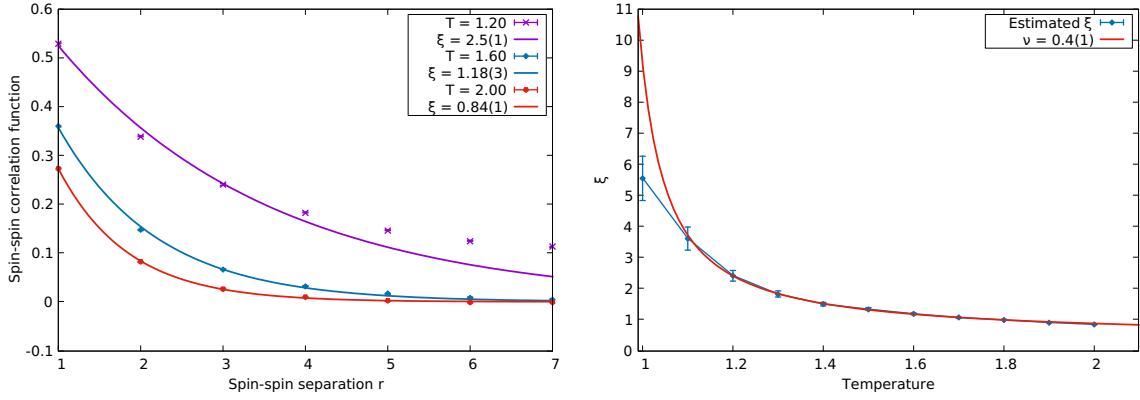
where  $\eta = k_B T / 2\pi J$  by spin wave theory and together with a vortex pair correction given by renormalization group theory.[15]

The power law decay of the equal-time spin-spin correlation function with distance is an indication of quasi-long range order.<sup>21</sup> Typically, for short-range order the correlation function decays exponentially as it is in high temperature phase, and for long-range order it remains constant over some large spin-spin separation as it is in 2D Ising model.

In [Fig. 24(a)] we demonstrate the correlation function at some temperatures and their corresponding power law fit, in the fitting procedure, since the above relation holds for spin-spin separation  $r_0 \ll r \ll L/2$ , where  $r_0$  is the lower limit mentioned in the section of theory of BKT phase transition, and the upper limit is due to periodic boundary condition. We observed that as the temperature decreases, the upper limit of the range in which the relation is preserved reduces from 4 to 3.<sup>22</sup> Note that the limit 3 is kept with the raise of deviation as the temperature is increased since this relation ceases to hold at high temperatures. In the high temperature phase, the function

<sup>21</sup>See page 3744 of Ref. O Kapikranian, et al. *J. Stat. Phys. A*, 40(14):3741, 2007.

<sup>22</sup>Smaller should cause the lack of number of degrees of freedom to fit.



(a) The function and the exponential fit (37).

(b)  $\xi$  vs  $T$  and the fit (38).

Figure 25:  $L = 16$ . Spin-spin correlation. Exponent  $c = 0.416454 \pm 0.1863$  is consistent with  $1/2 = 0.5$ .

decays exponentially

$$\vartheta(r, t) \sim \exp(-r/\xi) \quad (37)$$

where  $\xi$  termed the correlation length.

In [Fig. 25(a)], shown for  $L = 16$ ,  $T = 1.00(J/k_B)$  is the lowest temperature at which an exponential fit of form (37) was made. The upper limit of spin separation for the fit started at  $r = 3$  the smallest, and quickly raised to 6 and 7 with the deviation is reduced significantly. In both case of (36) and (37), it should be more appropriate by comparing the chi square value of each fit with variation of spin-spin separation.

In the neighbourhood of the critical temperature, the correlation length behaves as [14]

$$\xi \sim a \exp(bt^{-1/2}) \quad (38)$$

where  $t = |T - T_{BKT}| / T_{BKT}$  the reduced temperature,  $T_{BKT}$  the critical temperature of BKT phase transition, and  $a, b$  are non universal constants.

Deploying binning analysis or Jackknife estimation for correlation functions would not be feasible due to limited amount of memory. Think about it. Relying on the confidence of the estimation of integrated autocorrelation time, it is possible to approximate some number of configurations must be rejected between adjacent measurements of spin-spin correlation functions. No, it's not the problem of memory but the problem of human labour since each bin size must be determined for each value of spin-spin separation that only small sized lattices will be shown for spin-spin correlation.

#### 4.3.2 Magnetization

The free energy of the system<sup>23</sup>

$$F = -k_B T \ln Z \quad (39)$$

---

<sup>23</sup>The equilibrium condition at constant temperature and fixed volume is that the Helmholtz free energy minimized. [16, p. 37]

where the partition function, recall (67),

$$Z = \prod_{i,j} \left[ \int_{\mathcal{S}_{n-1}} d\mathbf{s}_{ij} \right] \exp \left[ \left( \sum_{\langle ij, lm \rangle} J_{ijlm} \mathbf{s}_{ij}^T \mathbf{s}_{lm} + \mathbf{H} \mathbf{M} \right) / k_B T \right] \quad (40)$$

The magnetization of the system<sup>24</sup>, whose magnitude denoted by  $M$ , is defined to be

$$\mathbf{M} = \sum_{ij} \mathbf{s}_{ij} = \hat{\mathbf{x}} \sum_{ij} \cos \theta_{ij} + \hat{\mathbf{y}} \sum_{ij} \sin \theta_{ij}, M^2 = \left( \sum_{ij} \cos \theta_{ij} \right)^2 + \left( \sum_{ij} \sin \theta_{ij} \right)^2 \quad (41)$$

Table 2: Results of fitting parameters for the relation (44) of form  $f(x) = a - T(\ln N + \ln 2)/b$  at low temperatures from  $T = 0.01(J/k_B)$  to  $T = 0.10(J/k_B)$ .

Type	PRNG	Lattice	$\bar{a}$	$\sigma_a$	$\bar{b}$	$\sigma_b$
Wolff	MT	16	1.00003	0.00001456	25.2058	0.04578
Wolff	MT	32	1.00005	0.00001345	25.1389	0.03407
CUDA	XORWow	32	1.00008	0.00006787	25.1368	0.1839
CUDA	XORWow	64	1.00011	0.00007479	25.1891	0.179
CUDA	XORWow	128	1.00004	0.0001228	25.8772	0.3432

The system may have a spontaneous magnetization, i.e. the present of global order after the applying external magnetic field is removed, is defined as[7, p. 185]

$$\mathbf{M}_{sp} = \lim_{\mathbf{H} \rightarrow 0} \lim_{L \rightarrow \infty} \langle \mathbf{M} \rangle_{T, \mathbf{H}} \quad (42)$$

where the notation  $|_{T, \mathbf{H}}$  means the average is at specific  $T, \mathbf{H}$ , and may be omitted frequently for clarity. Note that the limits are strictly ordered.

The ensemble average of the magnetization is related to the free energy as<sup>25</sup> [8, p. 174]

$$\langle \mathbf{M} \rangle_{T, \mathbf{H}} = \frac{1}{Z} \int_{\mathcal{R}_\sigma} d\boldsymbol{\sigma} \mathbf{M}(\boldsymbol{\sigma}) e^{-\mathcal{H}/k_B T} = - \frac{\partial F}{\partial \mathbf{H}} \Big|_{T, \mathbf{H}} = -\nabla_{\mathbf{H}} F \quad (43)$$

At low tempeature, it was predicted for  $\langle m \rangle = \langle \sqrt{M^2/N} \rangle$ , that

$$\langle m \rangle = 1 - T(\ln N + \ln 2)/8\pi \quad (44)$$

By this relation (44) and the demonstration in [Fig. 26(a)], it is seen that  $\lim_{L \rightarrow \infty} \langle m \rangle = 0$ ,<sup>26</sup> this is the Mermin-Wagner theorem, the absence of spontaneous magnetization. In [Fig. 26(a)], the relation (44) is changed to the form  $a - x(\ln 256 + \ln 2)/b$  to fit for  $\langle m \rangle$  and  $L = 16$ , we earned  $b = 25.20(4)$  as reported in [Tab. 2] is consistent with  $8\pi \approx 25.132\dots$  in two sigma. In the table, the results appear to be less accurate as the system size increases, however, it just reflects the fact of more number of measurements require for larger lattice sizes due to correlation.

<sup>24</sup>Originally, magnetization is defined by the density of magnetic moments, i.e. what here termed magnetization per spin.

<sup>25</sup>In two dimensions,  $\mathbf{H}\mathbf{M} = H_x M_x + H_y M_y$ .

<sup>26</sup>Since  $m$  is non-negative. See also the behaviour of  $\mathbf{M}^2$  reported later which better demonstrates the theorem.

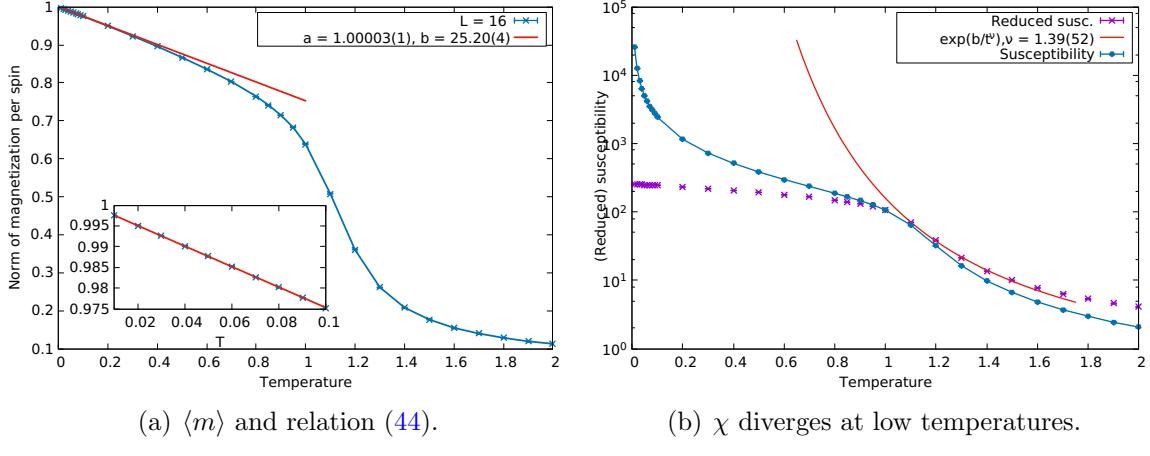


Figure 26:  $L = 16$ . Left figure:  $b = 25.20(4)$  is consistent with  $8\pi \approx 25.132\dots$  in two sigma. Right figure: red (solid) curve is the fit from  $T = 1.10(J/k_B)$  to  $T = 1.50(J/k_B)$  to the reduced susceptibility.

#### 4.3.3 Magnetic susceptibility

Representation in 2 dimensions,  $\mathbf{HM} = H_x M_x + H_y M_y$ . In terms of magnetization per spin  $\mathbf{m}(T) = \mathbf{M}(T)/N$ , the magnetic susceptibility characterizes the response of magnetization to the external field  $\mathbf{H}$ , we have<sup>272829</sup>

$$\begin{aligned}\chi &= \nabla_{\mathbf{H}} \langle \mathbf{m} \rangle|_{T,\mathbf{H}} = -\left. \frac{\partial^2 F}{\partial \mathbf{H}^2} \right|_{T,\mathbf{H}} = -\Delta_{\mathbf{H}} F \\ &= -\frac{1}{Z^2} \nabla_{\mathbf{H}} Z \int_{\mathcal{R}_{\sigma}} d\sigma \mathbf{m}(\sigma) e^{-\mathcal{H}/k_B T} + \frac{1}{Z} \int_{\mathcal{R}_{\sigma}} d\sigma \mathbf{m}(\sigma) \nabla_{\mathbf{H}} e^{-\mathcal{H}/k_B T} \\ &= \frac{N}{k_B T} \frac{1}{Z} \int_{\mathcal{R}_{\sigma}} d\sigma \mathbf{m}^2(\sigma) e^{-\mathcal{H}/k_B T} - \frac{N}{k_B T} \left[ \frac{1}{Z} \int_{\mathcal{R}_{\sigma}} d\sigma \mathbf{m}(\sigma) e^{-\mathcal{H}/k_B T} \right]^2\end{aligned}$$

And,<sup>30</sup>

$$\chi = N \left[ \langle \mathbf{m}^2 \rangle_{T,\mathbf{H}} - \langle \mathbf{m} \rangle_{T,\mathbf{H}}^2 \right] / k_B T \quad (45)$$

At low temperatures, the magnetic susceptibility is predicted to scale as[6]

$$k_B T \chi \sim N^{1-\eta/2} \quad (46)$$

<sup>27</sup>Laplacian operator  $\Delta = \nabla^2$ . Properties, see e.g. product rule (iii) [11, p. 21].

<sup>28</sup>Alternatively,  $\nabla_{\mathbf{H}} \langle \mathbf{m} \rangle|_{T,\mathbf{H}} \equiv \frac{\partial \langle m_x \rangle}{\partial H_x} + \frac{\partial \langle m_y \rangle}{\partial H_y}$  would lead to the same result of  $\chi$ . Note that,

$$\langle \mathbf{m} \rangle_{T,\mathbf{H}} = \frac{1}{Z} \int_{\mathcal{R}_{\sigma}} d\sigma \mathbf{m}(\sigma) e^{-\mathcal{H}/k_B T} = \frac{1}{Z} \int_{\mathcal{R}_{\sigma}} d\sigma [\hat{\mathbf{x}} m_x + \hat{\mathbf{y}} m_y] e^{-\mathcal{H}/k_B T} = \hat{\mathbf{x}} \langle m_x \rangle + \hat{\mathbf{y}} \langle m_y \rangle.$$

And  $\langle \mathbf{m}^2 \rangle = \langle m_x^2 \rangle + \langle m_y^2 \rangle$ , and  $\langle \mathbf{m} \rangle^2 = \langle m_x \rangle^2 + \langle m_y \rangle^2$ .

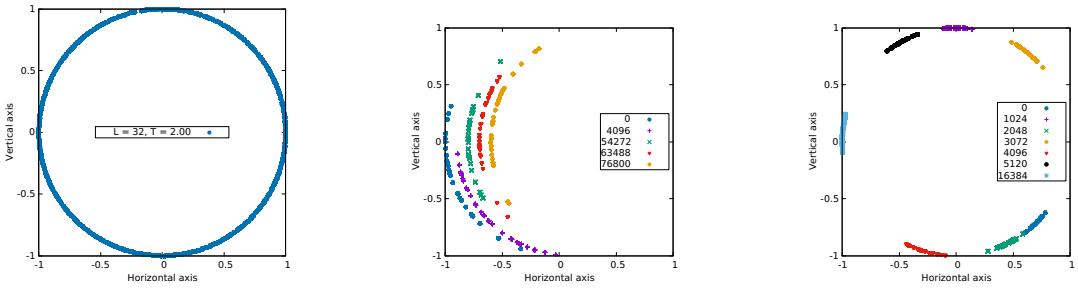
<sup>29</sup>Note that  $\mathbf{m}$  doesn't depend explicitly on  $\mathbf{H}$ , i.e.  $\frac{\partial m_i}{\partial H_j} = 0$ ,  $i \neq j$ ,  $i, j \in \{x, y\}$ .

<sup>30</sup>Note that  $\mathbf{m}^2 = \sum_{ij,lm} \mathbf{s}_{ij} \mathbf{s}_{lm} / N^2$ . Since  $\mathbf{m} \sim 0$ , then  $k_B T \chi = N \langle \mathbf{m}^2 \rangle = \sum_{ij,lm} \langle \mathbf{s}_{ij} \mathbf{s}_{lm} \rangle = \sum_{\mathbf{r}} \langle \mathbf{s}(0) \mathbf{s}(\mathbf{r}) \rangle = \sum_r \langle \mathbf{s}(0) \mathbf{s}(r) \rangle / 2$ . cf. Ref.[2, p. 914], and Ref. [17, p. 760].

This is originated from the divergence of correlation length<sup>31</sup> (see footnote (30), with continuous approximation)

$$\begin{aligned}\chi k_B T &= \int_0^\xi d^2 r \langle \mathbf{s}(0) \mathbf{s}(\mathbf{r}) \rangle = \int_0^\xi r dr \langle \mathbf{s}(0) \mathbf{s}(r) \rangle \\ &= \int_0^{xi} dr r^{1-\eta} = \frac{r^{2-\eta}}{2-\eta} \Big|_0^{xi} \sim \xi^{2-\eta}\end{aligned}$$

with a note that  $\eta$  was omitted due to  $\eta \ll 1$ . In a finite sized system,  $\xi \sim L$ , so  $\chi k_B T \sim N^{1-\eta/2}$ .



(a)  $T = 2.00(J/k_B)$ ,  $\mathbf{M}$  is easily averaged to zero. Metropolis. (b)  $T = 0.20(J/k_B)$ . Zero in time with Metropolis. (c)  $T = 0.01(J/k_B)$ . Zero in no time by Wolff.

Figure 27:  $L = 32$ . Spin phase angle on unit circles. In the middle figure: the horizontal coordinates are shifted for clarity, and the range of spin phase angle fluctuates as the system evolves; the key numbers are the number of MCS deployed.

The susceptibility diverges at criticality[15], [6].

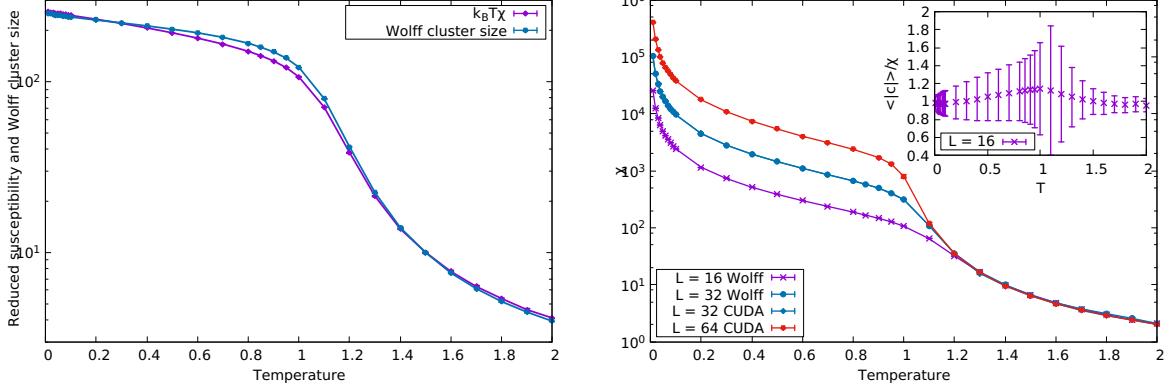
$$k_B T \chi \sim \exp(b/t^\nu) \quad (47)$$

We will follow [6] to set  $\mathbf{M} = 0$  since it would take too long time to average this quantity to zero<sup>32</sup>. See [Figs. 27(a), 27(b)] for local algorithm Metropolis, and [Fig. 27(c)] for cluster algorithm Wolff. In the latter  $\mathbf{M}$  will be averaged to zero in no time, this is the nature of cluster algorithm. A consequence of this reduction is that the binning analysis won't be applicable anymore since  $\mathbf{M}$  is definitely non-zero inside each bin. A suggestion of using the norm of magnetization rather than the magnetization itself can be made. Although in Ising model, the susceptibilities with the two uses then agree at low temperatures and at high temperatures, in thermodynamic limit, they differ by a trivial factor[7, p. 185], but in 2D XY model, the use of norm of magnetization would be underestimated since we don't have a solid theoretical understanding or reference. The latter produces a peak in susceptibility that will only agree with the use of zero magnetization in the thermodynamic limit, i.e. the peak would tend to diminish since at which the norm of magnetization vanishes while the term  $\langle m^2 \rangle = \langle \mathbf{m}^2 \rangle$  dominates.

In Ref. [17] and [13], it is reported that the average cluster size is proportional to the susceptibility by a ratio of 0.81 and our results shown in [Fig. 28(a)] appears to agree with this proportionality. Although our result as shown in [Fig. 28(b)] isn't inconsistent

<sup>31</sup>For  $r \gg \xi$ , correlation function diminishes.

<sup>32</sup>Remember the phenomenon of magnetic reversal in the case of Ising model with discrete symmetry.



(a) The two appears to be proportional.

(b)  $\chi$ . Inset: Ratio  $\langle |c| \rangle / \chi$ , reduced  $\chi$ .

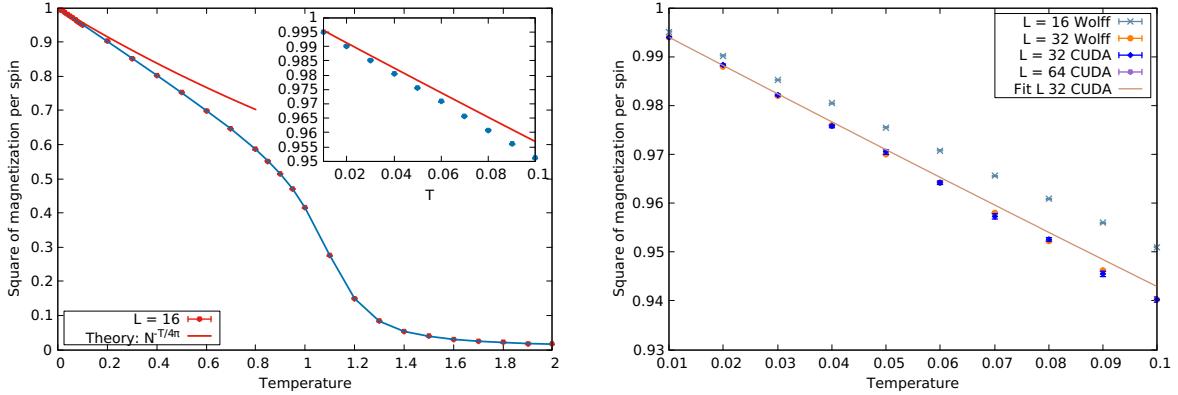
Figure 28: Average Wolff cluster size and reduced susceptibility.  $\chi$  diverges at low temperatures start from  $T_{BKT}$ .

with the ratio of 0.81, we do not completely agree with the author about the ratio, since the average cluster size and the susceptibility are certainly not independent, and estimating of a ratio must come with deviation and which is not trivial. Nevertheless, we agree that the Wolff dynamics is closely related to the physics of the model.

#### 4.3.4 Square of magnetization

In Ref. [6], the square of magnetization per spin at low temperatures obeys the following relation which is resulted from the spin wave approximation

$$k_B T \chi \approx N \langle \mathbf{m}^2 \rangle = N N^{-T/4\pi} = N^{1-T/4\pi} \quad (48)$$



(a)  $L = 16$ . Square of magnetization and relation (48).

(b) Simulation results of square of magnetization per spin and theoretical fit at low temperatures for  $L = 32$  CUDA.

Figure 29: Square of magnetization and also reduced susceptibility.

The result for  $L = 32$ , CUDA case is consistent with  $4\pi \approx 12.566\dots$  within three sigma while the other two cases of Wolff dynamics don't. The reason should be due to the number of samples used in these two cases are not enough for the mean to converge

Table 3: Results of fitting parameters for the relation (48) of form  $f(x) = aN^{-x/b}$  at low temperatures from  $T = 0.01(J/k_B)$  to  $T = 0.03(J/k_B)$ .

Type	$L$	$\bar{a}$	$\sigma_a$	$\bar{b}$	$\sigma_b$
Wolff	16	1.00007	0.00004	11.1669	0.06552
Wolff	32	1.00015	0.00003	11.3197	0.04141
CUDA	32	0.999842	0.0002733	11.8213	0.3167

because the time unit used in Wolff dynamics is not in scale with that in Metropolis so that although we used the same number of update steps is 131072, but for Wolff the effective samples should be smaller than that.

#### 4.3.5 Magnetic reversal

Since the magnetization is a vector, as an alternative to the study of its norm in some previous section, let's imagine a Cartesian coordinate system, [Fig. 30(a)] with Metropolis showing the x-component of the net magnetization per spin, most of the time, the quantity remains at some value close to  $-1$ , and in [Fig. 30(b)] the y-component mostly fluctuates about zero. The average of magnetization taken on some small number of measurements should be definitely non-zero close to minus unity, then  $\langle \mathbf{M} \rangle^2 = M_x^2 + M_y^2 \sim 1$  appears to show the presence of spontaneous magnetization.<sup>33</sup> However, over sufficiently long time, the average of  $M_x$  as well as of  $M_y$  should be zero and there should be no spontaneous magnetization as expected.

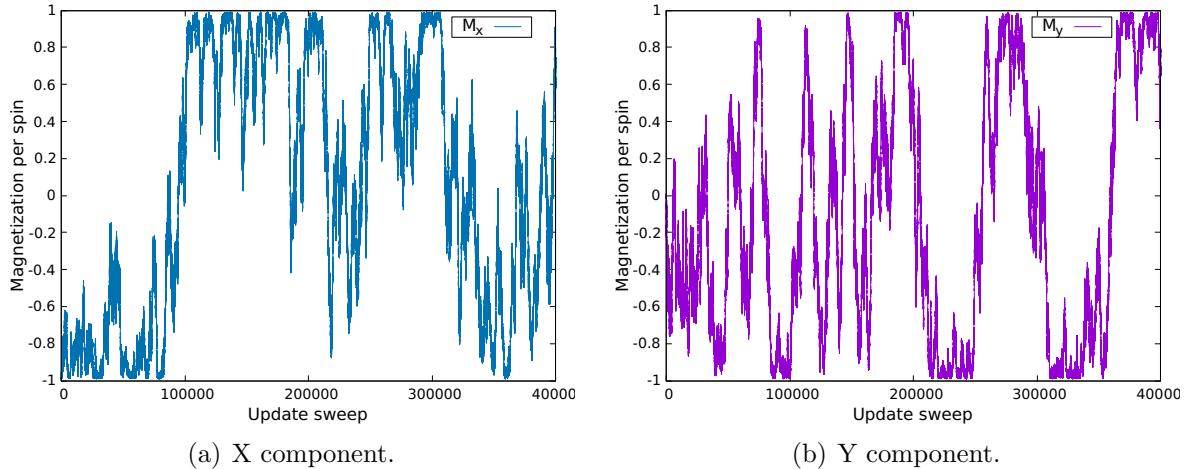


Figure 30:  $L = 8$ ,  $T = 0.01(J/k_B)$ . Magnetization per spin. Metropolis.

We want to argue that it is due to the nature of local update algorithm Metropolis that the system slowly change its state, but in time which may be finite or infinite for a discrete spin system like Ising, but definitely infinite for a continuous spin like XY model, the system will be able to experience all possible states, or finitely long time to pass all possible algebraic values of magnetization from  $-1$  to  $+1$ . To demonstrate this, we would like to introduce [Fig. 31(a)] for Wolff algorithm, even at  $T = 0.01(J/k_B)$ ,

<sup>33</sup>Note that the definition of spontaneous magnetization is given in the thermodynamic limit of infinite volume.

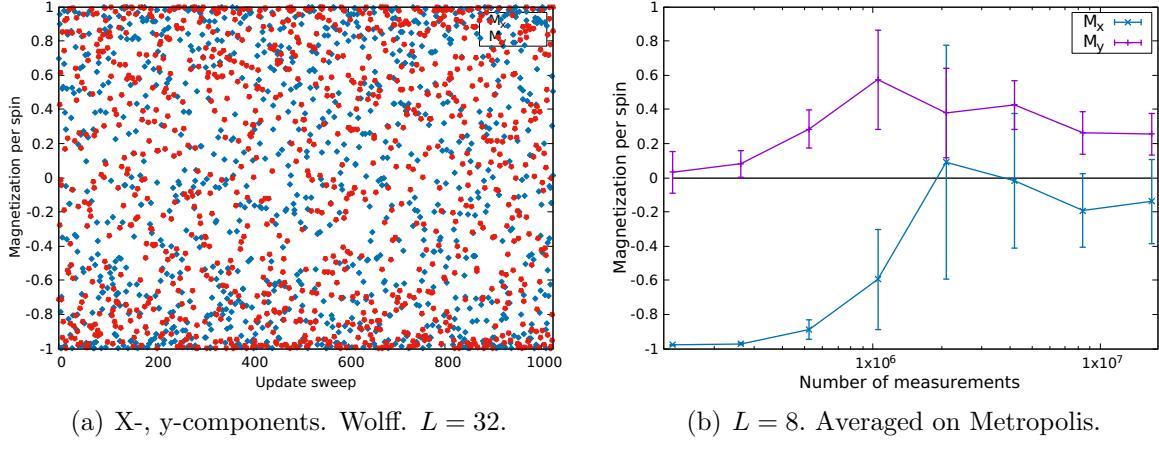


Figure 31:  $T = 0.01(J/k_B)$ . Magnetization per spin.

the change between states is instantly compared with Metropolis in [Figs. 30]. Note that at high temperatures, the two algorithms are approximately the same, since the average cluster size in Wolff should contain more or less than two spin sites. In [Fig. 31(b)], shown for Metropolis, the average of the two components of magnetization converges to zero slowly.

We push forward to [Fig. 32(a)] which shows the histogram of x and y components of magnetization per spin that their values are compact in some bin of values rather than spread symmetrically as in [Fig. 32(b)] which shows for Wolff algorithm. The later is due to the nature of the cluster algorithm, a large number of sites in a single cluster is being flipped to states that are closely related by a common axis making the system a high possibility be swapped into a completely new state. It is seen that by Wolff, the average magnetization is of the order of magnitude of zero. We should convince the readers to a figure demonstrating the average cluster size very soon.

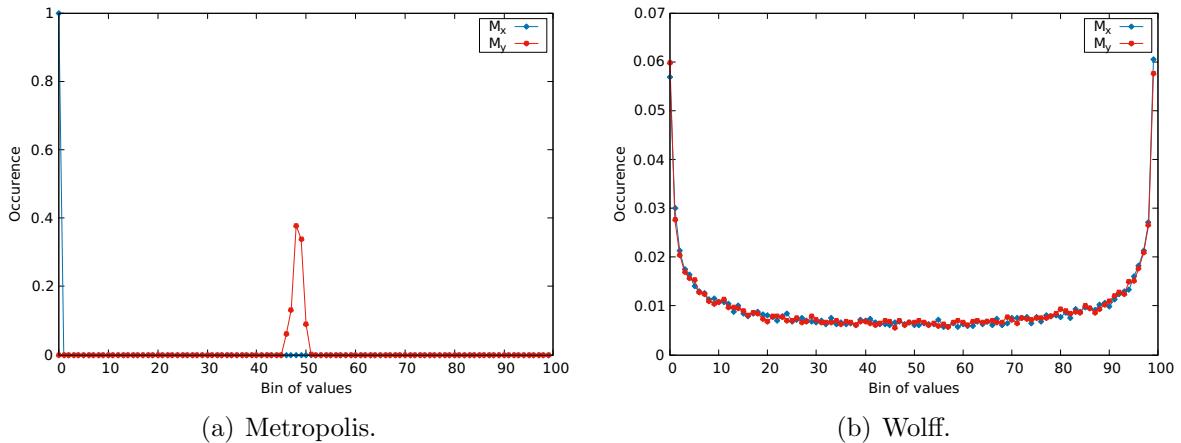


Figure 32:  $L = 32$ ,  $T = 0.01(J/k_B)$ , 32768 samples. Magnetization per spin. Values ranged from  $-1$  to  $+1$ .

The point is that the magnetization is zero at any finite temperature in a finite 2D XY system, with a local algorithm, this average should be very slowly converged while for a cluster algorithm like Wolff, it's very easily converged. This phenomena also occur

in a finite Ising model. However, a distinguish difference must be made. In the latter, there is a spontaneous magnetization at low temperature in the thermodynamic limit, but not in a finite system, since in an infinite system, the time of magnetic reversal is infinite. In the former case, the Mermin-Wagner theorem gives solid confirmation that there is no long range order, i.e. no spontaneous magnetization at any finite temperature in a continuous spin system in lower than 3 dimensional space.

In summary, magnetic reversal is just the nature of local update algorithm which slowly traverses the configuration space and each energy level corresponds to many microscopic configurations that their magnetization differs only by sign.

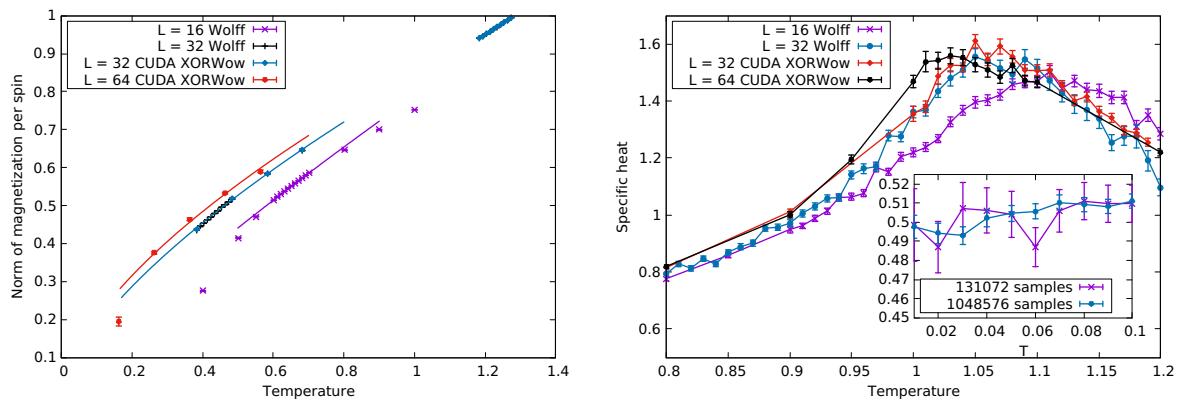
Lastly, either the norm of magnetization or the magnetization itself diminishes only in thermodynamic limit.

#### 4.3.6 Empirical evidence of magnetization

In Ref. [18], the author measured the exponent  $\beta \approx 0.231\dots$  with consistency check on theoretical calculation, simulation and experiments on some compounds, e.g.  $\text{BaNi}_2(\text{PO}_4)_2$ ,  $\text{Rb}_2\text{CrCl}_4$  where  $\beta$  is defined by<sup>34</sup>

$$\beta = \frac{\partial \ln \langle \mathbf{m}^2 \rangle^{1/2}}{\partial \ln(t)}, \quad \langle \mathbf{m}^2 \rangle = t^{2\beta} \quad (49)$$

where  $t = T_{C(J/k_B)} - T$ ,  $T_C(L) = T_{BKT} + \pi^2/c(\ln L)^2$ , and  $c \approx 2.1$  a constant.



(a) Exponent  $\beta$ . Temperature here is  $t$  not  $T$ .

(b)  $L = 16$ .  $C_v$  at low temperature.

Figure 33: Exponent  $\beta$  of magnetization and specific heat. Specific heat increases with temperature noting the statistical fluctuation.

In the same Ref. on page L57, the author mentioned that for finite sized lattices, correction of order  $\beta/\ln(L^2)$  must be introduced and  $L$  must be very large to neglect

<sup>34</sup>Note that the magnetization  $M$  mentioned in Ref. [18] is actually the square root of  $\langle \mathbf{m}^2 \rangle$  because of the relation that the author cited in Ref. [6]. However, the measurement in Ref. [18] is for  $\langle m \rangle$ . How to solve this discrepancy? Let's consider, e.g. 2 measurements only,  $\langle \mathbf{m}^2 \rangle^{1/2} = [(a^2 + b^2)/2]^{1/2}$  while  $\langle m \rangle = (a+b)/2$ . Note the inequality  $(a+b)^2 \leq 2(a^2 + b^2)$ , so  $\langle \mathbf{m}^2 \rangle^{1/2} \geq \langle m \rangle$ , and the equality holds iff  $a = b$ . Since for sufficiently large number of samples, i.e. in the limit of perfect non-correlation, our measurements should be the same, and either average  $\langle \mathbf{m}^2 \rangle^{1/2}$  or  $\langle m \rangle$  is fine. Yes a plot of  $\langle \mathbf{m}^2 \rangle^{1/2}$  and  $\langle m \rangle$  for  $L = 16$  confirms this strongly in the low temperature phase.

Table 4: Results of fitting parameters for the relation (49).  $T_{BKT}$  was chosen to be 0.892.

Type	PRNG	Lattice	$T_C$	Fit range	$\langle \beta \rangle$	$\sigma_\beta$	Correction
Wolff	MT	16	1.503	$0.80 \div 0.90$	0.419824	0.00423	0.075709
Wolff	MT	32	1.283	$0.80 \div 0.90$	0.368365	0.003955	0.053143
CUDA	XORWow	32	1.283	$0.60 \div 0.90$	0.33249	0.01043	0.047968
CUDA	XORWow	64	1.164	$0.60 \div 1.00$	0.309588	0.02828	0.037220
CUDA	XORWow	128	1.092	wait	wait	wait	wait

the correction. That's why an addition column of correction introduced in the table [Tab. 4] to explain our results compared with the empirical ones in the reference and consistency is obvious. In [Fig. 33(b)], as also reported in Ref. [6], the peak of specific heat doesn't seem to be size dependent though more number of measurements should we have to estimate specific heat better in the peak to check this statement.

#### 4.3.7 Square of spin phase angle

In Ref. [6], the average of square of spin phase angle relative to the axis of total instantaneous magnetization is related to the temperature and the number of spin sites as

$$\langle \theta^2 \rangle = T (\ln N + \ln 2) / 4\pi \quad (50)$$

In the same reference, the author mentioned the divergence of the average if the spin

Table 5: Results of fitting parameters for the relation (50) of form  $f(x) = a + x (\ln N + \ln 2) / b$  at low temperatures from  $T = 0.01(J/k_B)$  to  $T = 0.10(J/k_B)$ .

Type	PRNG	Lattice	$\bar{a}$	$\sigma_a$	$\bar{b}$	$\sigma_b$
Wolff	MT	16	-0.000133152	0.0000429	12.4684	0.03331
Wolff	MT	32	-0.000843474	0.001193	12.6518	0.7791
CUDA	XORWow	32	-0.000272221	0.0001512	12.4024	0.1008
CUDA	XORWow	64	-0.000332654	0.000159	12.4101	0.09367

phase angle were measured relative to some fixed axis. Although we understand that the energy of the system is unchanged if a change in the phase angle of every spin is introduced, but we do not get what did the author mean by saying that the divergence is related to this trivial fact. Now we would like to introduce an argument as follow.

Since the Hamiltonian involves the differences of two neighbouring spin phase angles, if the system is rotated by some angle, i.e. every spin is collectively changed by some amount in its phase angle, the Hamiltonian won't be changed. Therefore, it is made physically possible that the collection of spins being distributed over any arc on the unit circle as in [Fig. 27(c)]. Hence, if the spin phase angle is referenced to some fixed axis, by take a look at [Fig. 27(c)], the average  $\langle \theta^2 \rangle$  trivially diverges.

In [Fig. 34], data in the left blue spread in more bins than those in the right blue. This can also be seen in [Fig. 27], in the middle figure, data for each color has its own spread while in the last figure, those for every color appear to have the same range though each located in a distinguished arc. This can be easily argued as, in the random

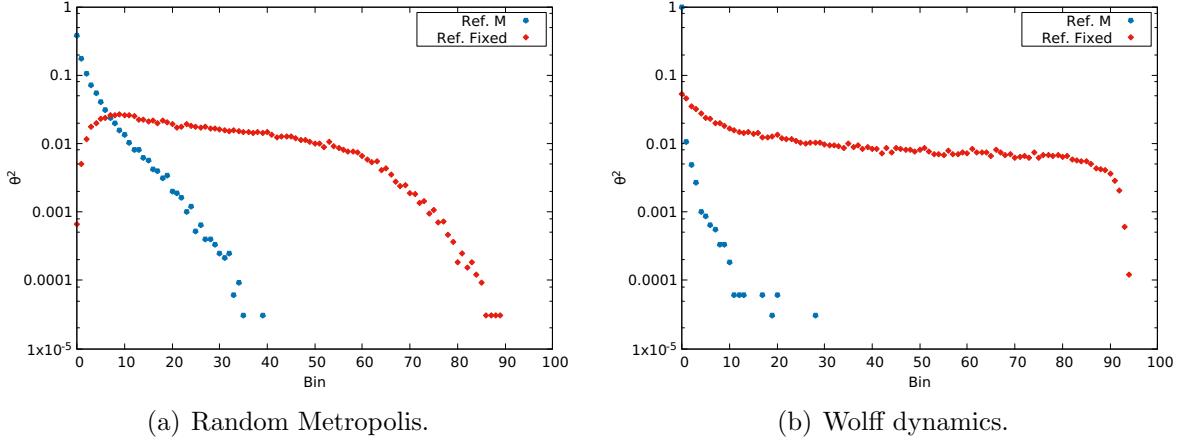


Figure 34:  $L = 32$ ,  $T = 0.01(J/k_B)$ , 32768 samples. Histogram of square of spin phase angle. Blues are referred to the instantaneous magnetization, reds are some fixed axis.

proposal Metropolis, every site has its own axis to flip, while in Wolff, especially at low temperatures, a large number of sites, i.e. those co-exist in a single cluster, have the same flipping axis so that it is highly probable that most spins are in some compact range. The reds show that data present in mostly every bin leading to the divergence of  $\langle \theta^2 \rangle$ .

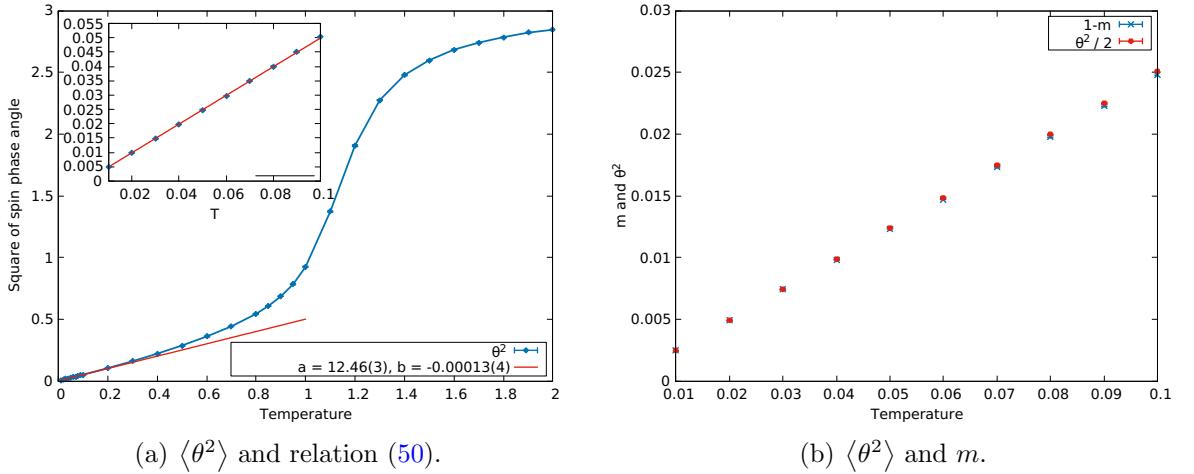


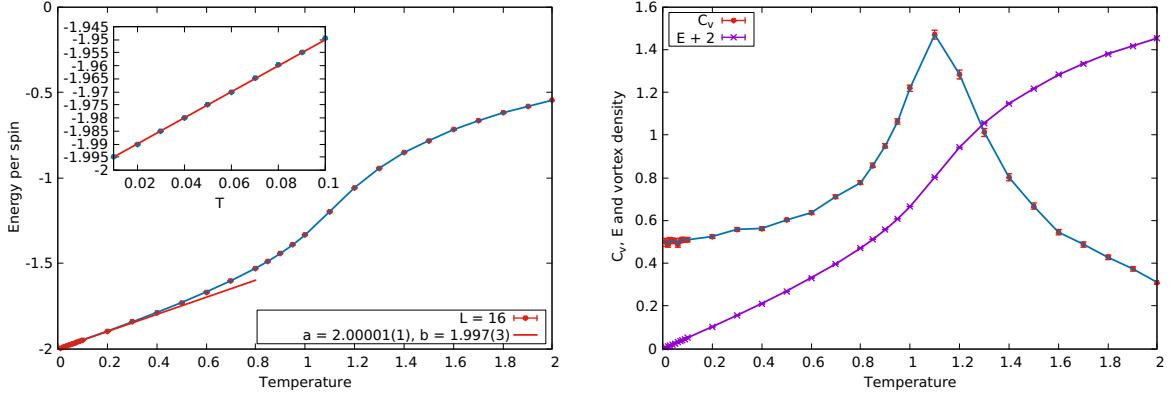
Figure 35:  $L = 16$ . Right figure originated from the similarity somehow between the low temperature relations of  $m$  and  $\langle \theta^2 \rangle$ .

#### 4.3.8 Energy

The ensemble average of the energy of the system is

$$\langle \mathcal{H} \rangle_{T,\mathbf{H}} = \frac{1}{Z} \int_{\mathcal{R}_\sigma} d\sigma \mathcal{H} e^{-\mathcal{H}/k_B T} = F - T \frac{\partial F}{\partial T} \quad (51)$$

The second equality in (51) which relates the average of energy with the free energy



(a) Energy per spin and relation (52).

(b) Specific heat.

Figure 36:  $L = 16$ . Specific heat and energy. Specific heat at low temperatures see the inset of [Fig. 33(a)].

is deduced as the following

$$F - T \frac{\partial F}{\partial T} = k_B T \ln Z + k_B T \ln Z + T \frac{k_B T}{Z} \int_{\mathcal{R}_\sigma} d\sigma \frac{\mathcal{H}}{k_B T^2} e^{-\mathcal{H}/k_B T} = \langle \mathcal{H} \rangle_{T, \mathbf{H}}$$

At low temperature, the relation of energy per spin<sup>35</sup> is[6]

$$\langle E \rangle = -2 + T/2 \quad (52)$$

The results listed in [Tab. 6] mostly consistent in 3 sigma, however, the data for CUDA simulation doesn't appear to be consistent, this is trivially because of the correlation in Wolff dynamics is the smallest, and the case of CUDA simply needs more uncorrelated measurements.

At the transition region, the slop of energy is steepest corresponding to the peak of specific heat that will be argued in the next part.

Table 6: Results of fitting parameters for the relation (52) of form  $f(x) = -a + x/b$  at low temperatures.

Type	PRNG	Lattice	Fit range	$\bar{a}$	$\sigma_a$	$\bar{b}$	$\sigma_b$
Wolff	MT	16	0.01 ÷ 0.06	2.000015	0.000019	1.997105	0.003023
Wolff	MT	32	0.01 ÷ 0.04	2.000015	0.000014	1.992407	0.002845
CUDA	XORWow	32	0.01 ÷ 0.03	2.000014	0.000012	1.993617	0.002768
CUDA	XORWow	64	0.01 ÷ 0.03	2.000025	0.000006	1.989055	0.001497
CUDA	XORWow	128	0.01 ÷ 0.03	2.000023	0.000003	1.989176	0.000684

<sup>35</sup> $E$  that will be defined shortly.

### 4.3.9 Specific heat

The specific heat characterizes the energy required to change the temperature of each degree of freedom of the system by  $1K$  provided the volume of the system is fixed

$$C_v = \frac{1}{N} \left. \frac{\partial \langle \mathcal{H} \rangle}{\partial T} \right|_{T, \mathbf{H}} = \frac{1}{N} \frac{\partial}{\partial T} \left( F - T \frac{\partial F}{\partial T} \right) = -\frac{1}{N} T \frac{\partial^2 F}{\partial T^2} \quad (53)$$

$$= \left[ \langle \mathcal{H}^2 \rangle_{T, \mathbf{H}} - \langle \mathcal{H} \rangle_{T, \mathbf{H}}^2 \right] / N k_B T^2 \quad (54)$$

The last equality in (54) is deduced as

$$\begin{aligned} \frac{\partial \langle \mathcal{H} \rangle}{\partial T} &= -\frac{1}{Z^2} \int_{\mathcal{R}_\sigma} d\sigma \mathcal{H} e^{-\mathcal{H}/k_B T} \frac{\partial Z}{\partial T} + \frac{1}{Z} \int_{\mathcal{R}_\sigma} d\sigma e^{-\mathcal{H}/k_B T} \frac{\partial \mathcal{H}}{\partial T} + \frac{1}{Z} \int_{\mathcal{R}_\sigma} d\sigma \mathcal{H} \frac{\partial}{\partial T} e^{-\mathcal{H}/k_B T} \\ &= -\frac{1}{Z} \frac{\partial}{\partial T} \int_{\mathcal{R}_\sigma} d\sigma e^{-\mathcal{H}/k_B T} \langle \mathcal{H} \rangle_{T, \mathbf{H}} + \frac{1}{Z} \int_{\mathcal{R}_\sigma} d\sigma \frac{\mathcal{H}^2}{k_B T^2} e^{-\mathcal{H}/k_B T} \\ &= \frac{1}{k_B T^2} \left[ \langle \mathcal{H}^2 \rangle_{T, \mathbf{H}} - \langle \mathcal{H} \rangle_{T, \mathbf{H}}^2 \right] \end{aligned}$$

Denote the energy per spin by  $E = \mathcal{H}/N$ , the specific heat is given by

$$C_v = N \left[ \langle E^2 \rangle_{T, \mathbf{H}} - \langle E \rangle_{T, \mathbf{H}}^2 \right] / k_B T^2 \quad (55)$$

The specific heat doesn't diverge at the critical temperature.[15] However, it has a peak at temperatures a bit higher than the transition. Why does the peak appear not to be dependent of lattice size? And will it tend to peak at the transition point of an infinite lattice or will it always be at higher than the transition point? These are not easy to answer and need first a precise determination of vortex density.

### 4.3.10 Helicity modulus

The helicity modulus is said to be a thermodynamic function evidently measuring the "rigidity" of an isotropic system under an imposed phase "twist".<sup>36</sup> Helicity modulus is the second derivative of free energy with respect to an imposed twist phase angle  $\alpha$  on one direction of the boundaries while on the other one remains periodic<sup>37</sup>

$$\Upsilon = \frac{\partial^2}{\partial \alpha^2} F(\alpha)|_{\alpha=0} \quad (56)$$

Helicity modulus is a dimensionless quantity given by[19, p. 598]

$$\Upsilon = -\frac{1}{2} \langle E \rangle - \frac{J}{k_B T N} \left\langle \left[ \sum_{\langle ij, lm \rangle} \sin(\theta_{ij} - \theta_{lm}) \mathbf{r}_{ijlm} \hat{\mathbf{x}} \right]^2 \right\rangle \quad (57)$$

where  $\hat{\mathbf{x}}$  is some fixed direction, here, is taken to be a unit vector lies horizontally, then  $\mathbf{r}_{ijlm}$ , which denotes the vector from site  $ij$  to site  $lm$ , for the case  $\mathbf{s}_{lm}$  are vertical neighbours of  $\mathbf{s}_{ij}$ , is perpendicular to  $\hat{\mathbf{x}}$  leading to null result of the dot product.

---

<sup>36</sup>M. Fisher et al. *Phys. Rev. A*, 8:1111–1124, Aug 1973.

<sup>37</sup>Ref. Urs Gerber et al. *J. Phys. Conference Series*, 651(1):012010, 2015.

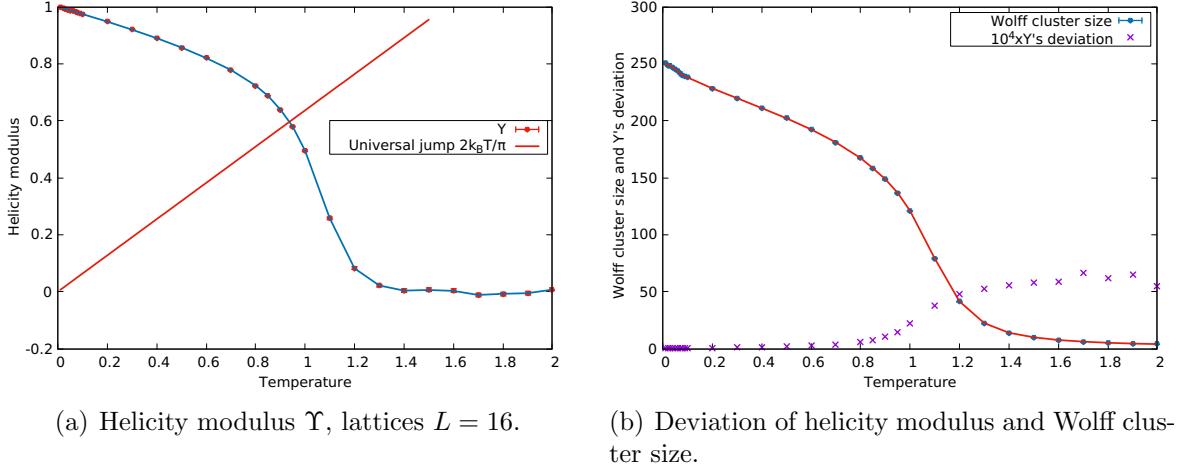


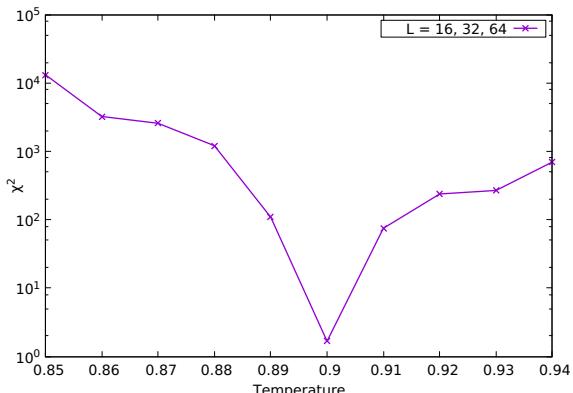
Figure 37:  $L = 16$ . The data of helicity modulus at high temperatures are ugly because each update sweep at high temperature consists of flipping clusters of much more smaller sizes.

Then, the helicity modulus, with horizontally nearest neighbours remain, is calculated according to

$$\Upsilon = -\frac{1}{2} \langle E \rangle - \frac{J}{k_B T N} \left\langle \left\{ \sum_{ij} [\sin(\theta_{ij} - \theta_{lmr}) - \sin(\theta_{ij} - \theta_{lml})] \right\}^2 \right\rangle \quad (58)$$

$$= \left\langle -\frac{1}{2} E - \frac{J}{k_B T N} \left\{ \sum_{ij} [\sin(\theta_{ij} - \theta_{lmr}) - \sin(\theta_{ij} - \theta_{lml})] \right\}^2 \right\rangle \quad (59)$$

where  $lmr$ ,  $lml$  denote the index of spins located at the right and left neighbouring sites of spin  $s_{ij}$ , respectively. The reduction to the expectation of a single term to avoid the correlation of the two former terms since they are estimated from a common configuration. Also, as mentioned, correlation measures only linear relationship, this case would be a good example to illustrate one more times the sake of grouping appropriate expectations into a single one, since, e.g. although  $E(X, Y) = 0$  but  $X, Y$  may still be related to each other and that the variances can't be simply added.



The critical temperature is found by the intersection of the spin stiffness with the line  $\Upsilon = 2T/\pi$  due to the Nelson-Kosterlitz universal jump[19, p. 26]<sup>38</sup> available in the thermodynamic limit  $N \rightarrow \infty$

$$\lim_{T \rightarrow T_{BKT}} \Upsilon(T)/k_B T = \frac{2}{\pi} \quad (60)$$

$T_{BKT} = 0.89294(8)$  was found in Ref. [14].<sup>39</sup>

<sup>38</sup>Should read David R. Nelson and J. M. Kosterlitz. *Phys. Rev. Lett.*, 39:1201–1205, Nov 1977.  
Figure 38: Chi square value resulted from the fit (62) from  $L = 16, 32, 64$ .  
And in Ref. M Hasenbusch and K Henn. *J. Phys. A*, 30(1):63, 1997.

In Ref. [20], the author argued that the helicity modulus for an infinite system can be extrapolated from the following relation resulted from renormalization group

$$\Upsilon(N) = \Upsilon_\infty \left[ 1 + \frac{1}{2} \frac{1}{\ln(N) + C} \right] \quad (61)$$

where  $C$  is some free parameter.

At  $T_{BKT}$ , the spin stiffness of an infinite system is given by the universal jump at  $T_{BKT}$ , then the correct transition temperature is the one that has the smallest chi square from the fit of the data achieved at some temperatures close from below  $T_{BKT}$  to the following function with only one degree of freedom

$$\Upsilon(N) = \frac{2}{\pi} T \left[ 1 + \frac{1}{2} \frac{1}{\ln(N) + C} \right] \quad (62)$$

In our simulation,  $T_{BKT} = 0.90(1)$  that is very close to the most precise value upto now and also consistent with theoretical predictions, e.g. with high-temperature series expansions given somewhere. The temperature estimated in Ref. [20] by the method used here is far from accurate compared with the other two references. The reason might be their use of small sized lattices  $L = 3, 4, 5$  and 6 since they argued that they found it is empirically that the relation (61) is well applicable even to small lattices.

If the transition was estimated from the peak of specific heat, it would be not easy since the peak slowly converges to locate at the transition temperature for an infinite system and the estimation of specific heat in the transition region requires that autocorrelation is overcome and /or large number of measurements is simulated. See [Fig. 33(a)] for specific heat in the region.

Table 7: Results of fitting parameters for the relation (63) of  $\Upsilon$  at low temperatures from  $T = 0.01(J/k_B)$  to  $T = 0.10(J/k_B)$ .

Type	PRNG	Lattice	$\bar{a}$	$\sigma_a$	$\bar{b}$	$\sigma_b$
Wolff	MT	16	1.000044	0.000008	3.951209	0.004068
Wolff	MT	32	1.000053	0.000005	3.933712	0.002263
CUDA	XORWow	32	1.000056	0.000004	3.931145	0.001775
CUDA	XORWow	64	1.000054	0.000002	3.929123	0.000837
CUDA	XORWow	128	1.000018	0.000001	3.965626	0.001386
CUDA	XORWow	256	1.000016	0.000001	3.967394	0.000619

The relation for spin stiffness at low temperature is given by<sup>40</sup>

$$\Upsilon = 1 - T/4 \quad (63)$$

We put a simple argument for (63) based on the relation of energy (52) as follow. By (57), the second term should be insignificantly contributed to the spin stiffness since it is a square of a sum which is on the order of magnitude of zero since at low

---

<sup>40</sup>Read Ref. P. Minnhagen and M. Nylén. *Phys. Rev. B*, 31:1693–1695, February 1985.

temperature, adjacent spins have their phase angles are approximately equal. Then, since  $E = -2 + T/2$ ,

$$\Upsilon \approx -\langle E \rangle /2 = 1 - T/4$$

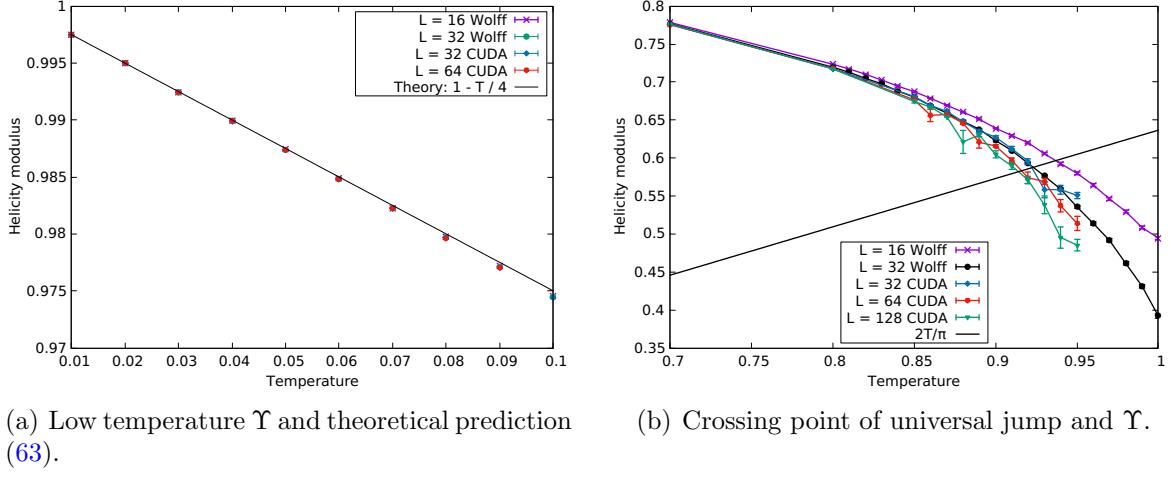


Figure 39:  $L = 16$ . Spin stiffness for different lattice sizes and different simulations.

The fit results to the function (63) in the form  $a - x/b$  are given in table [Tab. 7] and demonstrated in [Fig. 39(a)]. As shown in the table, the results for  $b$  doesn't agree very well with the theoretical prediction  $b = 4$ , though we compared our own fitting program with GNUplot. Even if the chi square is reduced as much as possible, the best case in which we was possible to earn a consistent result with  $b = 4$  is for  $L = 16$ , Wolff dynamics, 3 initial temperatures from  $T = 0.01(J/k_B)$  to  $T = 0.03(J/k_B)$  is  $b = 3.982354 \pm 0.013481$  that agrees within 2 sigma, and for 4 initial temperatures, it was  $b = 3.975851 \pm 0.009752$  that is consistent within 3 sigma. We don't know how to argue these discrepancies, it is possibly due to the fact that, at low temperature, the adjacent configurations are strongly correlated, even though the Wolff dynamics was used, but it still costs some number larger than the order of magnitude of 30 non-scale update sweeps to escape correlation, that more number of update sweeps should be deployed. This possibility is supported by observing that for the Wolff case, lattice  $L = 32$  yields result that is less close to the expected one than  $L = 16$  does, and by going down to the case of CUDA, the results are even less close to the expected value. **We should have more update sweeps here to check this.** Also the fact that in deploying Jackknife as well as binning analysis, the bin size have to be some power of 2, and the deviation certainly be underestimated. The second possibility that doubts the deviation of the estimation methods should be rejected since, it is observed that the deviation even with bin size that is larger than the chosen bin size still be very small that can't be larger than an order of magnitude. Also, we wonder whether some theoretically logarithmic correction should be introduced as in the case of reduced susceptibility or even in the relation of spin-spin correlation function.

#### 4.3.11 Vortex density

Vortex density  $v$  is the number of vortices in the system divided by the number of spins. In order to calculate the density, the rule mentioned previously is applied to all

the primitive cells in the lattice.

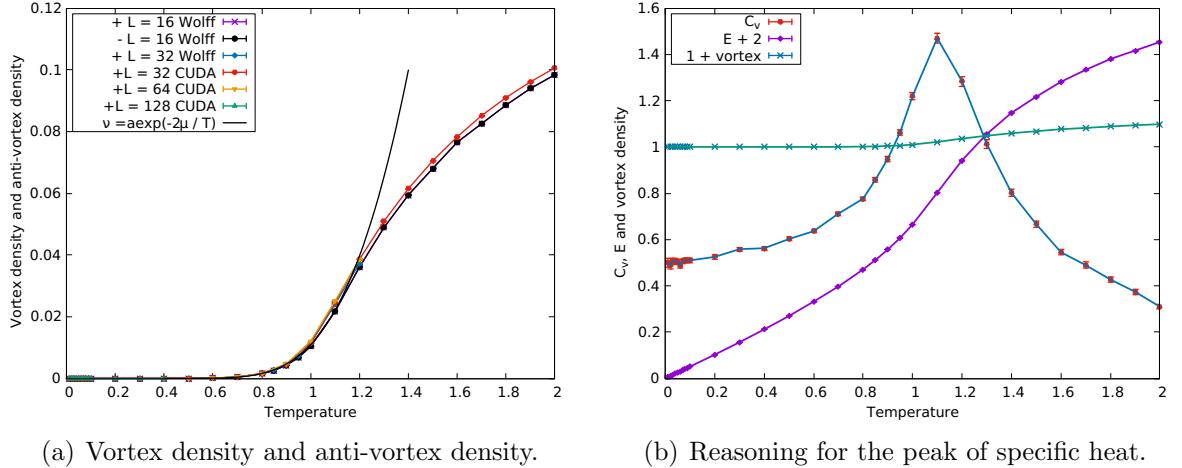


Figure 40:  $L = 16$ . Vortex density, energy and the peak of specific heat.

At low temperatures, vortex density  $\nu$  is related to the temeprature as [6]<sup>41</sup>

$$\nu \sim \exp(-2\mu/T) \quad (64)$$

where  $2\mu = 10.2$  is the energy required for a vortex pair to occur. In [Tab. 8], the

Table 8: Relation (64),  $f(x) = a \exp(-b/x)$ . Not very much consistent with Ref. [6]

Type	PRNG	Lattice	Fit range	$\bar{a}$	$\sigma_a$	$\bar{b}$	$\sigma_b$
Wolff	MT	16	0.40 ÷ 1.10	25.6938	2.731	7.81554	0.1002
Wolff	MT	32	0.40 ÷ 1.10	30.5109	3.829	7.91173	0.1162
CUDA	XORWow	32	0.40 ÷ 1.10	27.7669	4.071	7.78041	0.1303
CUDA	XORWow	64	0.40 ÷ 1.10	29.6065	4.641	7.82637	0.1402

results are not consistent with the theoretical prediciton of BKT theory,  $2\mu = 10.2$ , however our results, e.g. with the case  $L = 32$  with Wolff,  $2\mu = 7.91173 \pm 0.1162$  whose upper bound within three sigma is not very much inconsistent with the lower bound of three sigma also of  $9.4 \pm 0.3$  in Ref. [6]. We should spend more time to study the relation theoretically, since the result in Ref. [6] is accurate compared with the theory but not very precise, i.e. the deviation is large though.

Vortex starts to occur at  $T = 0.40(J/k_B)$  for  $L = 16, 32$ , and  $64$ , but for  $L = 128$  vortex density is non-zero from  $T = 0.30(J/k_B)$ . The temperature at which a vortex starts to occur depends on lattice size and a vortex pair at really low temperatures is met by chance, e.g. more number of update sweeps may raise chance to observe a vortex pair at very low temperatures.

<sup>41</sup>Note that power law that is common in this thesis can be linearized before perform a fit but that comes with a cost of error propagation.

## 5 Conclusions

The transition temperature is found to be  $T_{BKT} = 0.90(1)$  consistent with literature of simulation as well as some theoretical approximation. Here are some exponents

- × The exponent  $\beta$  of magnetization is best found to be  $0.30(6)$  consistent with empirical evidence and theoretical approximation.
- × Exponent  $\eta = k_B T / 2\pi J$ .
- × Exponent  $\nu = 0.4(1)$  for correlation length and  $\nu = 1.39(52)$  for susceptibility are consistent with theoretical prediction of  $\nu = 1/2$ .

In terms of spin separation  $r$ , at any time, the spin-spin correlation function  $\vartheta(r, t)$  decays exponentially at high temperatures meaning strong disorder, while in the really low temperature phase, it decays algebraically instead. This power decay stands for a quasi-long range order due to the presence of Bloch spin wave.

- $\vartheta(r, t) \sim r^{-\eta}$  for  $T \leq T_{BKT}$ .
- $\vartheta(r, t) \sim \exp(-r/\xi)$ .

This spin wave makes occur of no spontaneous magnetization in the thermodynamic limit  $N \rightarrow \infty$  at low temperatures and then the system has no spontaneous magnetization at any finite temperatures.

- Average magnetization  $\langle \mathbf{m} \rangle = 0$  at any finite temperature regardless lattice size.
- Square of magnetization<sup>42</sup> diminishes with the increase of total spins as  $\langle \mathbf{m}^2 \rangle = N^{-\eta/2}$  for  $T < T_{BKT}$ .
- Also, in the vicinity of  $T_{BKT}$ , it obeys  $\langle \mathbf{m}^2 \rangle = t^{2\beta}$ , where  $t$  is a function of temperature.
- The norm of magnetization vanishes in terms of total spins following the expression  $\langle |\mathbf{m}| \rangle = 1 - \eta (\ln N + \ln 2) / 4$ .

Correlation length diverges at  $T_{BKT}$  and vanishes at low temperatures. This is the origin of the divergence of susceptibility at low temperatures.

- \* At low temperature, power law decay of spin-spin correlation function gives no correlation length.
- \* At  $T_{BKT}$ , correlation length diverges as  $\xi \sim a \exp(bt^{-\nu})$  and so does the reduced susceptibility, where  $a, b$  are non-universal constants, and  $t$  reduced temperature.

At some temperature well below  $T_{BKT}$ , the power decay ceases to hold that is being modified by the occurrence of vortices that come in tightly bound pairs with anti-vortices, and at  $T_{BKT}$  some of those pairs unbind raising the energy of the system rapidly and then those vortices quickly cease to happen as the vortex density appears to saturate.

Immediately below the transition temperature, the system is in states of topological order in terms of tightly bound vortex-anti-vortex pairs. With the introduction of topological order, the Berezinskii-Kosterlitz-Thouless phase transition occurs between a disordered phase and a topologically ordered phase.

---

<sup>42</sup>Also,  $\langle |\mathbf{m}| \rangle = t^\beta$  in the vicinity of  $T_{BKT}$ .

## A Canonical ensemble

Each spin in the  $\mathcal{O}(2)$  model can rotate in the plane of the lattice, the phase space  $\mathcal{R}_\sigma$  consists of infinite number of configurations. Dummy variables  $x, y, z$  are used hereafter to index some configuration in the state space unless otherwise stated. Solving a system of infinite number of equations of motion if available, which are differential equations with given initial conditions, is practically impossible [21, pp. 1, 2]. Nevertheless, one can get the value of any physical quantity  $A$  by taking average in the course of time  $t$ , if a function  $A(t)$  can be established [21, p. 4], by

$$\langle A \rangle = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau A(t) dt \quad (65)$$

A problem with (65) is that how long the time interval  $\tau$  would be sufficient for the system to traverse the configuration space to produce meaningful averages [12, p. 285]. Because, we do not know whether the configuration on which our measurement starts is a desired equilibrium configuration or not, this time interval must include the relaxation time by which the system equilibrates and the time that it remains in configurations corresponding to some equilibrium state provided the relaxation time is finite. The later is required to be large enough so that the effect of initial condition is negligible and the contribution of the equilibration stage is insignificant compared to that of equilibrium configurations. And if such a function  $A(t)$  can't be constructed, a problem is that the measuring time resolution is finite.

A statistical ensemble is a set of configurations which is actually the configuration space acquired a probability distribution that gives some additional information about the probability of each configuration. Since the study of phase transition involves the evolution of the phase trajectories of the system, the energy of the system need to be varied, it is convenient to introduce an artificial stochastic dynamics<sup>4344</sup> by which the model can exchange energy with a connected heat reservoir characterized by a temperature  $T$  being kept fixed, while the number of spins remains unchanged<sup>45</sup>. The heatbath and the model constitute a closed system following microcanonical ensemble, while every configuration of the model belongs to the canonical ensemble having the time-independent *probability distribution*<sup>46</sup> [21, p. 77] given by

$$P(x) \equiv P(x, T) \equiv P(\sigma) = \frac{1}{Z} e^{-\mathcal{H}/k_B T} \quad (66)$$

where  $Z$  is the partition function corresponding to the canonical ensemble, which is simply a finite sum running over all possible configurations of the system, for each temperature, for general  $\mathcal{O}(n)$  models,

$$Z \equiv Z(T) = \int_{\mathcal{R}_\sigma} d\sigma e^{-\mathcal{H}/k_B T} = \prod_{i,j} \left[ \int_{\mathcal{S}_{n-1}} ds_{ij} \right] \exp [-\mathcal{H}/k_B T] \quad (67)$$

---

<sup>43</sup>A general  $\mathcal{O}(n)$  model may or may not have an intrinsic dynamics.

<sup>44</sup>Read. Ref. K. Binder and D. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction*. Graduate Texts in Physics. Springer Berlin Heidelberg, 2010.

<sup>45</sup>These two systems form a closed system whose configurations obey the time-independent micro-canonical ensemble satisfying the Liouville's theorem [21, pp. 9, 10].

<sup>46</sup>Read pp. 91, 92, 113 Ref. K. Huang. *Introduction to Statistical Physics*. Taylor & Francis, 2009.

where  $\mathcal{S}_{n-1}$  is the state space for spin vectors, i.e. it is the set of unit vectors whose angles take values in the set of real number,  $n = 2$  for XY model.

The quantity  $A$  can be averaged over the canonical ensemble as

$$\langle A \rangle = \int_{\mathcal{R}_\sigma} d\sigma A(\sigma) P(\sigma) = \frac{1}{Z} \int_{\mathcal{R}_\sigma} d\sigma A(\sigma) e^{-\mathcal{H}/k_B T} \quad (68)$$

Later on, our model is given some dynamics that is required to be ergodic in the sense that the system can transit between any configurations in the phase space<sup>47</sup>. By this property and the virtue that  $P(x)$  is time-independent, the time average (65) can be replaced by averaging over the canonical ensemble (68) [16, p. 65] [21, p. 4]. This can be clearly seen by the following arguments<sup>48</sup>

$$\langle A \rangle_{\text{ens}} = \lim_{\tau \rightarrow \infty} \frac{1}{T} \int_{t=0}^{(J/k_B)\tau} \langle A \rangle_{\text{ens}}(t) dt = \left\langle \lim_{\tau \rightarrow \infty} \frac{1}{T} \int_{t=0}^{(J/k_B)\tau} A(t) dt \right\rangle = \langle \langle A \rangle_\tau \rangle_{\text{ens}} = \langle A \rangle_\tau \quad (69)$$

In (69), the first equality is due to the property that  $P(x)$  is time-independent so that the ensemble average is a constant, the second and the third equalities are simply a swap between time average and ensemble average of  $A$ , and the fourth equality is because that the time average is constant within statistical error.

## B Importance Sampling and unbiased estimations

According to the distribution (66), at a specific temperature, not all configurations have equal probability of occurrence then the contribution to (68) is significant only for some configurations<sup>49</sup> and sampling  $\mathcal{R}_\sigma$  uniformly would waste computational resources since we study some discrete temperatures only<sup>50</sup>. Beside, generating an infinite number of configurations is practically not feasible.

Due to this fact, we can only estimate the unknown finite mean  $\langle A \rangle$  with the aid of the technique of importance sampling [12, pp. 448 - 450] according to the probability (66). Then, the mean (68) can be estimated [22, p. 508] by the sample mean

$$\bar{A} = \frac{1}{q} \sum_{x=1}^q A_x \quad (70)$$

where  $q$  is the number of measurements, i.e. the number of configurations on which the observable is measured.  $A_x$  are measured values of  $A$  on different configurations, which can be considered to be random variables following the same distribution with  $A$ .<sup>51</sup> Since, for some arbitrary random variables, the expectation of their sum is the

<sup>47</sup>This sense is stronger than the ergodic hypothesis stated by Boltzmann in Ref. [16, p. 65]

<sup>48</sup>See [8, p. 67].

<sup>49</sup>The configuration space are infinite.

<sup>50</sup>Even if this is not the case, we don't have storage to store such a huge number of configurations.

<sup>51</sup>The distribution of any observable  $A$  is unspecified though its mean and variance are physically finite, but as later on mentioned its sample mean follows normal distribution.

sum of their expectations [22, p. 219], this is an unbiased estimator of  $\langle A \rangle$  trivially because

$$\langle \bar{A} \rangle = \left\langle \frac{1}{q} \sum_{x=1}^q A_x \right\rangle = \frac{1}{q} \left\langle \sum_{x=1}^q A_x \right\rangle = \frac{1}{q} \sum_{x=1}^q \langle A_x \rangle = \frac{1}{q} q \langle A_x \rangle = \langle A_x \rangle = \langle A \rangle \quad (71)$$

Sampling via distribution  $P(x)$ , the estimator (70) clearly has the minimum variance<sup>52</sup>[22, p. 510] since generated configurations, which have high frequencies of occurrence, i.e. high probabilities, are mostly the exact configurations on which the measured values  $A(\sigma)$  don't vary too strongly from the expectation  $\langle A \rangle$ . A general procedure for importance sampling looks for some realizable probability distribution that mimics the product  $A(\sigma)P(\sigma)$  as much as possible, especially where this product being large. [12, p. 448] Unfortunately, such a finding is practically nontrivial.

An unbiased estimator for the finite variance of the observable  $A$  is [22, p. 508]

$$\hat{\sigma}_A^2 := \frac{1}{q-1} \sum_{x=1}^q (A_x - \bar{A}_q)^2 = \frac{q}{q-1} (\bar{A}_q^2 - \bar{A}_q^2) \quad (72)$$

In Ref. [22, p. 474], for a normally distributed random sample  $\{A_i\}$ , the quantity  $\sum_{x=1}^q (A_x - \bar{A}_q)^2 / \sigma_A^2$  has the  $\chi^2$  distribution with  $q-1$  degrees of freedom, then has variance  $2(q-1)$ [22, p. 470]. Although the prior distribution of  $A$  is unspecified, however, since having no choice in literature, provided the measurements are uncorrelated, one may assume that the variance of  $\hat{\sigma}_A^2$  can be given by

$$\text{Var}(\hat{\sigma}_A^2) = 2\sigma_A^4/(q-1) \approx 2\hat{\sigma}_A^4/(q-1) \quad (73)$$

however, the correct measure should be deploying *resampling methods*, no such assumption preserved. Note that as we later mention *binning analysis*, the mean in each bin can be considered to be asymptotically normal, however, all measurements in each bin are mostly correlated, then the central limit theorem may not hold that it would be wrong if the binned variance were treated as if they obey the  $\chi^2$  distribution.

This estimator is unbiased due to the following arguments[22, p. 508],

- × By the requirement that  $A_x$  (with finite means) are either independent [22, p. 230] or at least uncorrelated [22, p. 254], the variance of the sample mean  $\bar{A}$  is

$$\text{Var}(\bar{A}) = \text{Var}\left(\frac{1}{q} \sum_{x=1}^q A_x\right) = \frac{1}{q^2} \text{Var}\left(\sum_{x=1}^q A_x\right) = \frac{1}{q^2} \sum_{x=1}^q \text{Var}(A_x) = \sigma_A^2/q \quad (74)$$

- × Since  $A_x - \langle A \rangle = A_x - \bar{A}_q + \bar{A}_q - \langle A \rangle$ , then

$$\sum_{x=1}^q (A_x - \langle A \rangle)^2 = \sum_{x=1}^q (A_x - \bar{A}_q)^2 + \sum_{x=1}^q (\bar{A}_q - \langle A \rangle)^2 + 2 \sum_{x=1}^q (A_x - \bar{A}_q)(\bar{A}_q - \langle A \rangle)$$

where the last term equal

$$(\bar{A}_q - \langle A \rangle) \sum_{x=1}^q (A_x - \bar{A}_q) = (\bar{A}_q - \langle A \rangle)(q\bar{A}_q - q\bar{A}_q) = 0$$

---

<sup>52</sup>In terms of number of measurements.

× Then,

$$\begin{aligned} \mathbb{E} \left[ \sum_{x=1}^q (A_x - \bar{A}_q)^2 \right] &= \mathbb{E} \left[ \sum_{x=1}^q (A_x - \langle A \rangle)^2 \right] - \mathbb{E} \left[ \sum_{x=1}^q (\bar{A}_q - \langle A \rangle)^2 \right] \\ &= \sum_{x=1}^q \mathbb{E} (A_x - \langle A \rangle)^2 - \sum_{x=1}^q \mathbb{E} (\bar{A}_q - \langle A \rangle)^2 \\ &= q\text{Var}(A) - q\text{Var}(\bar{A}) = (q-1)\text{Var}(A) \end{aligned}$$

And so,

$$\mathbb{E} (\hat{\sigma}_A^2) = \sigma_A^2. \quad (75)$$

□.

This is originated from the estimator of sample variance, since we know not the mean as well as the variance, in the stage of estimating the mean we use all measurements we have, then in the estimation of the sample's variance, we are using the estimated mean also. This is why the sample variance should have  $q-1$  rather than  $q$  since we are having only  $q-1$  degrees of freedom left. This estimator of sample variance in turn has no bias from the exact variance of the quantity and be using as an unbiased estimator for variance. It can be seen that without these comments, no reason is good to persuade one to use this estimator since as we will see this estimator may not be the one with smallest variance with respect to the exact variance, and more importantly, in measurement we are interested in the standard deviation the square root of this variance, and the square root of the unbiased estimator is unfortunately biased.

The quality of an unbiased estimator<sup>53</sup> must be evaluated in terms of its variance or its mean square error[22, p. 511], i.e. how close to the exact variance it is. For the sake of discussion, if  $A$  obeys normal distribution,<sup>54</sup>

$$\mathbb{E} (\hat{\sigma}_A^2 - \sigma_A^2)^2 = 2\sigma_A^4/(q-1), \text{ and } \mathbb{E} [\hat{\sigma}_A^2 - \sigma_A^2]^2 = 2\sigma_A^4/(q+1) \quad (76)$$

where  $\hat{\sigma}_A^2 = \sum_{x=1}^q (A_x - \bar{A}_q)^2 / (q+1)$ .

By (76), the estimator with the smallest variance is not (72) but that with the  $-1$  in the denominator is substituted by  $+1$  becoming a biased one and even this estimator is inadmissibly dominated by other estimators as mentioned in [22, p. 510, 511]. For large  $q$ , there is no significant difference between using  $q$ ,  $q-1$  and  $q+1$ , however, as will be seen later, at criticality the phenomenon of critical slowing down present in which correlation time diverges with system size would lead to the number of uncorrelated measurements being not large compared with unity. We end this little discussion by noting that we have no prior information of the distribution of our measurable  $A$ , and we are trying to study its estimation only.

Combine (74) with (72), we have

$$\sigma_A^2 = \hat{\sigma}_A^2/q = (\bar{A}_q^2 - \bar{A}_q^2) / (q-1) \quad (77)$$

---

<sup>53</sup>There are many problems with the theory of unbiased estimator, e.g. such an estimator may not exist for some quantity.[22, p. 511]

<sup>54</sup>The first is simply the variance of  $\hat{\sigma}_A^2$ , while the second calculation is the variance of the biased estimator added by the square of the bias.[22, p. 511]

By (74), in order to reduce the statistical error of the estimate of  $\langle A \rangle$ , a large number of measurements must be taken. For large  $q$ , the central limit theorem<sup>55</sup> [22, p. 361] provides a more useful information is that the estimator (70) tends to obey a normal distribution with the mean  $\langle A \rangle$  and variance (74), (77).

Since the probability (66) is exponentially non-trivial and the Hamiltonian  $\mathcal{H}$  is a functional of  $N$  variables, i.e. the probability is high dimensional, that we can't use methods like inverse transform [12, pp. 445, 446] which requires the desired probability distribution to be integrable and some equations to be practically solvable, Box - Muller, Fernández - Criado [12, p. 447] for Gaussian distribution, or the von Neumann acceptance - rejection method [12, pp. 458, 459] which may have large number of rejected configurations with generally formal *proposal distribution*. Hence, the technique of importance sampling is then deployed by Markov chain Monte Carlo methods with the key idea [24, p. 386] is Markov chains<sup>56</sup>.

## C Markov chain Monte Carlo methods

This appendix is our understanding of finite state space Markov chain, applied specifically for  $\mathcal{O}(1)$  Ising model with finite state space of  $r = 2^N$  elements , serving as a tool by which intuitive arguments for ergodicity, balances and the decomposition of related methods are given. Although, for general  $\mathcal{O}(n)$ ,  $n \geq 2$  models, the configuration space is  $\mathcal{S}_{n-1}^N$ , where  $\mathcal{S}_{n-1}$  is the spin state space which contains unit vectors of  $n - 1$  real-valued components[13], i.e. the configuration space is infinite, however, the 2D XY model can be discretized in form of a clock model as mentioned in somewhere, e.g. Z(16) in Ref. [13] with a finite state space and all these knowledge can be applied.<sup>57</sup> In the following, all discussions are realized on Ising model with  $Z(2)$  symmetry.

### C.1 The theory of finite Markov chain

This part discusses Markov chain Monte Carlo methods and sketches the connection between the physics and the mathematics, no rigorous proof is made but related references are cited.

#### C.1.1 Definitions and general properties

Since the Hamiltonian of a Ising model doesn't contain the information of how the system changes its configurations, a dynamics, which is represented by an algorithm,

---

<sup>55</sup>The theorem requires the independence in the sample[23, p. 362], [22, p. 361](a random sample collects independent variables[22, p. 158]). Independence leads to no correlation, but no correlation doesn't necessarily require independence, e.g. functional dependence[22, p. 251], however, for Markovian, only two adjacent configurations are explicitly related, then for a single observable, no correlation should be enough for independence.

<sup>56</sup>Andrei Andreyevich Markov (1856 - 1922), Russian, with interest in poetry, originated Markov chain in his 1907 work.[25, p. 687]

<sup>57</sup>For general state space, i.e. the state space is infinite and may be either countable or uncountable, the transition matrix then has infinite size. We can't pretend to treat it as if it was finite, however, the construction requires depth studies of mathematical concepts, e.g. the measure theory, that are beyond the scope of this thesis. Further pursuit should refer to Ref. S. Meyn and R. Tweedie. *Markov chains & stochastic stability*. Springer-Verlag, 1993.

must be introduced separately [12, p. 620] (*cf.* a dynamics using time evolution of an operator for Heisenberg model [12, p. 312]). Markov chain Monte Carlo methods are a class of algorithms [24, p. 383] for sampling from a probability distribution based on constructing a Markov chain that has the desired distribution as its stationary distribution [24, pp. 386, 387]<sup>58</sup>. In the remainder of this thesis, MCMC is abbreviated for Markov chain Monte Carlo.

A *stochastic process* with discrete time parameter defined as a sequence of random variables, here,  $\sigma_1, \sigma_2, \dots$ , is a *Markov chain* if, for each time  $n$ , the conditional distributions of all  $\sigma_{n+k}$  for integral  $k \geq 1$  given  $\sigma_1, \dots, \sigma_n$  depend only on  $\sigma_n$  and not on the earlier states  $\sigma_1, \dots, \sigma_{n-1}$  [22, pp. 188 - 200]. This is called the *Markov property* which means that the process is *memoryless* [25, p. 687] *a.k.a* *Markovian* [8, p. 167]. In symbol,

$$\mathbf{Pr}(\sigma_{n+1} | \sigma_1, \dots, \sigma_n) = \mathbf{Pr}(\sigma_{n+1} | \sigma_n) \quad (78)$$

A *stochastic matrix* is a square matrix whose elements are non-negative and each row sum is equal to unity [25, p. 687] which can be equivalently interpreted in physics as that from any configuration the system must be able to transit to some other in the configuration space and the net probability must be preserved to be equal to unity. By holding the Markov property, an algorithm, which preserves the conservation of probability, can be mathematically represented by a square matrix of nonnegative elements which is [8, p. 168] a *stochastic matrix*  $\mathbf{P}_{r \times r} = [p_{xy}]$  called the *transition matrix*, where  $p_{xy}$  is the transition probability from configuration  $\sigma_x$  to another  $\sigma_y$  in one step. Markovian is interpreted as that the next configuration of the chain can be drawn by specification of which definite element of the matrix that corresponds to the current configuration to determine the transition probability without reference to other configurations. The conservation of probability reads

$$p_{xy} \geq 0, \quad \sum_y p_{xy} = 1 \quad (79)$$

The probabilities of being in various configurations  $\sigma$  at time  $n \geq 1$  is characterized by a *probability distribution vector*  $\boldsymbol{\nu}(n)$  (note that a vector is represented as a row in the whole thesis, i.e.  $\boldsymbol{\nu}(n)_{1 \times r}$ ) which has  $r$  positive components whose sum is unity,  $\nu_x(n)$  is the probability that the system is in configuration  $\sigma_x$  at time  $n$ . The probability distribution vector at time  $n + 1$  is

$$\boldsymbol{\nu}(n + 1) = \boldsymbol{\nu}(n) \cdot \mathbf{P} \quad (80)$$

that is

$$\nu_y(n + 1) = \sum_x \nu_x(n) p_{xy} \quad (81)$$

Consider an arbitrary initial probability distribution vector of the system  $\boldsymbol{\nu}(1)$ , the probability distribution vector at time  $n \geq 1$  is given by

$$\boldsymbol{\nu}(n) = \boldsymbol{\nu}(1) \cdot \mathbf{P}^{n-1} \quad (82)$$

---

<sup>58</sup>Read also Ref. W.H. Press. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

this can be easily seen by the method of mathematical induction.

By (82),  $\mathbf{P}^k$  ( $k \geq 1$ ) is the  $k$ -step transition matrix of the Markov chain, i.e.  $p_{xy}^{(k)}$  is the transition probability of getting from configuration  $\sigma_x$  to  $\sigma_y$  in  $k$  steps.  $\mathbf{P}^k$  ( $k \geq 1$ ) is indeed a stochastic matrix, e.g.

$$\sum_y p_{1y}^{(2)} = \sum_y \sum_x p_{1x}^{(1)} p_{xy}^{(1)} = \sum_x p_{1x}^{(1)} \sum_y p_{xy}^{(1)} = \sum_x p_{1x}^{(1)} = 1 \quad (83)$$

### C.1.2 Balance conditions

A probability distribution vector that satisfies

$$\boldsymbol{\nu} \mathbf{P} = \boldsymbol{\nu} \Leftrightarrow \sum_x \nu_x p_{xy} = \nu_y \quad (84)$$

is called a *stationary distribution* for the Markov chain. [22, p. 198]

The detailed balance condition<sup>59</sup> [8, p. 170] requires any transition related to a current configuration to hold

$$\nu_x p_{xy} = \nu_y p_{yx} \quad (85)$$

which simply asserts the equality and is now substituted to (84), the right equality becomes

$$\nu_y = \sum_x \nu_x p_{xy} = \sum_x \nu_y p_{yx} = \nu_y \sum_x p_{yx} = \nu_y 1 = \nu_y \quad (86)$$

This satisfactory can be interpreted as that the condition of detailed balance (85) is a sufficient condition to guarantee that the Markov chain has a stationary distribution.

By the simple thought of (86), i.e.  $\nu_y = \nu_y \sum_x p_{yx}$ , (81) is equivalent to

$$\nu_y(n+1) - \nu_y(n) = \sum_x [\nu_x(n) p_{xy} - \nu_y(n) p_{yx}] \quad (87)$$

(87) is similar to the well known so-called *the master equation* (see, e.g. [27]) if the Markov chain were termed a stochastic process with continuous time parameter, and the term in the square brackets is put to be equal to zero is exactly the detailed balance condition (85).

Also, the global balance, which relates all transitions from a current configuration<sup>60</sup>, simply sets the sum in the right hand side of (87) rather than every single difference in the square brackets to be zero

$$\sum_x \nu_x(n) p_{xy} = \sum_x \nu_y(n) p_{yx} = \nu_y(n) \quad (88)$$

which  $\nu_y$  simply wants to say that I have an origin, the sum of all probabilities that transit any configuration in the configuration space to my state constitue my existence.

---

<sup>59</sup>Some texts refer this to as reversibility. See e.g. [26].

<sup>60</sup>Ref. Martin Weigel. *J. Comp. Phys.*, 231(8):3064 – 3082, 2012.

### C.1.3 Ergodicity in Markov chain

The *graph*  $\mathcal{G}(\mathbf{A})$  of a square matrix  $\mathbf{A}_{r \times r} = [a_{xy}]$  is defined to be the directed graph on  $r$  nodes  $\{\sigma_1, \sigma_2, \dots, \sigma_r\}$  in which there is a directed edge leading from  $\sigma_x$  to  $\sigma_y$  if and only if  $a_{xy} \neq 0$  [25, p. 671].  $\mathbf{A}$  is said to be *irreducible* [25, p. 671] if and only if  $\mathcal{G}(\mathbf{A})$  is *strongly connected*, i.e. for each pair of nodes  $(\sigma_x, \sigma_y)$  there is a sequence of directed edges leading from  $\sigma_x$  to  $\sigma_y$  [25, pp. 209, 671]. The ergodicity property, which requires that we have non-zero probability of getting between any two states in the chain after some finite time [8, p. 170], is equivalent to the condition of a strongly connected graph<sup>61</sup>. Therefore, a MCMC algorithm, which reserves the ergodicity property, has its transition matrix  $\mathbf{P}$  being irreducible.

A nonnegative irreducible matrix can be either *primitive* or *imprimitive*. A primitive matrix is the square matrix that passes the ***Frobenius's Test for Primitivity*** which asserts that [25, p. 678]

$$\mathbf{A} \geq 0 \text{ is primitive if and only if } \mathbf{A}^k > 0 \text{ for some integer } k > 0. \quad (89)$$

There is also a ***sufficient condition for primitivity*** [25, p. 678] which reads

$$\forall \mathbf{A} \geq 0, \mathbf{A} \text{ is irreducible, if } \exists a_{xx} > 0 \text{ i.e. } \text{trace}(\mathbf{A}) > 0 \text{ then } \mathbf{A} \text{ is primitive.} \quad (90)$$

Provided that the transition matrix  $\mathbf{P}$  is primitive, [25, p. 693]

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \mathbf{e}\boldsymbol{\nu} \quad (91)$$

where  $\mathbf{e}$  is a column of all 1's and  $\boldsymbol{\nu} = \mathbf{p}^T$  is the *unique* stationary probability distribution vector of the chain, i.e. all rows of  $\mathbf{P}$  are  $\boldsymbol{\nu}$ . This *uniqueness* is guaranteed by the ***Perron - Frobenius theorem*** [25, p. 673] (primitivity implies *irreducibility* [25, p. 674]) for the Perron vector  $\mathbf{p}$  of  $\mathbf{P}$ . Once a  $k$  realizes (91), whatever the  $\boldsymbol{\nu}(1)$  is, the probability distribution function from the time beyond  $k$  is always the unique vector  $\boldsymbol{\nu}$  as easily seen by<sup>62</sup>

$$\boldsymbol{\nu}(1) \mathbf{P}^k = \boldsymbol{\nu}(1) \mathbf{e} \boldsymbol{\nu} = \boldsymbol{\nu} \quad (92)$$

A Markov chain, whose transition matrix is primitive, is said to be *regular*, [8, p. 169] [23, p. 433] *a.k.a aperiodic chain* [25, p. 694]. This is the ergodicity property in physics literature [8, p. 169]<sup>63</sup> which is exactly the condition of (89) meaning that one can get from every state  $\sigma_x$  to every other state  $\sigma_y$  in a finite sequence of moves [8, p. 169]. Note that previously an interpretation of ergodicity (we call it non-regular) is already declared by that the similarity between ergodicity and irreducibility (or strongly connected graph) is established, an example is the Ehrenfest urn model given in [23, pp. 411, 433], simply put, some transition is permissible only for odd number of steps, while some others require this number to be even, though it is possible to transit all the states<sup>64</sup>, while this latter interpretation (we call it regular ergodicity) is

<sup>61</sup>Note that Markovian needs not to join ergodicity to give the equivalence with strong connectivity (once momentarily thought) because Markovian quantifies the dependence of transition probabilities.

<sup>62</sup>An independent proof named Doeblin's proof mentioned in [23, pp. 449 - 451] using the test (89).

<sup>63</sup>Read also Ref. Bernd A. Berg and Alain Billoire. *Markov Chain Monte Carlo Simulations*. John Wiley & Sons, Inc., 2007.

<sup>64</sup>See more, a footnotes 4 and 5, in Ref. J.S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer New York, 2001.

more strictly that it requires primitivity, see the example of the weather of the Land of Oz mentioned in [23, pp. 406, 408], for a glance, there exists an integral number  $q$  such that there is positive probability of getting between any state of weather after  $q$  steps. Ergodicity for the Markov chain, which may be either regular or not, is more strictly than the ergodicity defined by Boltzmann for physical system in the sense that the chain can equilibrate to any accessible configuration in the canonical ensemble.

The example of Ehrenfest urn model is a case of ergodicity for non-regular Markov chain whose irreducible transition matrix is imprimitive, *a.k.a periodic chain* [25, p. 694], the period for the model is 2, i.e. a state has the probability to return to itself after some even number of steps only. For  $\mathbf{P}$  imprimitive, [25, p. 693]

$$\lim_{k \rightarrow \infty} \frac{1}{k} [\mathbf{I} + \mathbf{P} + \dots + \mathbf{P}^{k-1}] = \mathbf{e}\boldsymbol{\nu} \quad (93)$$

The subsequence of (93) is that

$$\lim_{k \rightarrow \infty} \frac{1}{k} [\boldsymbol{\nu}(1) + \dots + \boldsymbol{\nu}(k)] = \boldsymbol{\nu}(1)\mathbf{e}\boldsymbol{\nu} = \boldsymbol{\nu} \quad (94)$$

It is easy to see that if the transition matrix is primitive, (93) reduces to the form of (91). Since most MCMC algorithms satisfy the condition of (89), in the remainder, we will consider regular ergodicity property as well mentioned in physics literature. In addition, we haven't known how to realize either the *Cesàro limit* (93) [25, p. 633] or (94) in physical simulation yet since we are interested in a stationary distribution which is predefined to build a transition matrix via some dynamics but not starting from a matrix and specifying its stationary distribution vector [25, p. 635].

In summary, an algorithm in the class of Markov chain Monte Carlo methods is required to 1) be Markovian (78), 2) preserve the conservation of probability (79), 3) be ergodic, i.e. either irreducible or (89) to guarantee the Markov chain converges to some unique stationary probability distribution, this can be the desired one 4) provided that the algorithm holds the sufficient condition of detailed balance (85).

## C.2 Metropolis - Hastings

### C.2.1 Random proposal

Metropolis algorithm first introduced<sup>65</sup> for a square lattice whose every site locates a particle that can move to any other site, the proposal process is carried out by some kind of random walk for successive particles. The authors argued intuitively for the ergodicity in two dimension. The generalization given by Hastings<sup>66</sup>. For the sake of discussion, let  $p_{xy}$  be the transition probability between any configurations  $\sigma_x$  and  $\sigma_y$  in the configuration space. The condition of detailed balance reads

$$P(x)p_{xy} = P(y)p_{yx} \quad (95)$$

where a transition probability  $p_{xy}$  is decomposed into a product of  $\rho(x, y)\alpha(x, y)$ ,  $\rho(x, y)$  termed the proposal probability density and  $\alpha(x, y)$  termed the acceptance probability density.

---

<sup>65</sup>N. Metropolis et al. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

<sup>66</sup>W. K. Hastings. *Biometrika*, 57(1):pp. 97–109, 1970.

Starting from a configuration  $\sigma_x$ , a small change is proposed that results in a configuration  $\sigma_y$  [27]. In Metropolis algorithm, the proposal distribution is taken to be symmetric,  $\rho(x, y) = \rho(y, x) = 1/N$ , i.e. one of the spins is randomly chosen to propose a change, and then (95) becomes,

$$\frac{\alpha(x, y)}{\alpha(y, x)} = \frac{P(y)}{P(x)} \quad (96)$$

It can be seen in (96) that the acceptance probabilities  $\alpha(x, y)$  and  $\alpha(y, x)$  are positively correlated, i.e. if one is large then the other is large too, then one of them should be chosen so that the acceptance rate is as high as possible. Then, the acceptance probability is given by

$$\alpha(x, y) = \min \{1, P(y)/P(x)\} = \min \{1, \exp(-\Delta\mathcal{H}/k_B T)\} \quad (97)$$

where  $\Delta\mathcal{H} = \mathcal{H}_t - \mathcal{H}_c$ , the Hamiltonian  $\mathcal{H}_c$  is for the current configuration and  $\mathcal{H}_t$  for the proposal configuration with the generalized flipping operation (98) which differ by a single spin site  $s_{ij}$ . The interpretation of (97) is that every spin flip which lowers the energy of the system is permitted and for others which may raise the energy is set into a trial with probability  $\exp(-\Delta\mathcal{H}/k_B T)$ .

- × Markovian is obvious because transition probability relates only two configurations, a current one and a proposed one.
- × There are  $N$  configurations have equal proposal  $\rho(x, y) = 1/N$ , the others in the space has zero proposal, the net probability to accept those proposals is  $\sum_y \rho(x, y)\alpha(x, y) + 0$ , the net probability to refuse is  $\sum_y \rho(x, y)[1 - \alpha(x, y)] + 0$ , so the sum is equal to  $\sum_y \rho(x, y) = \sum_y 1/N = 1$ .
- × Detailed balance is from the formation with the crucial point is a current configuration after being transit to a proposed one have chance to be proposed immediately.
- × Irreducibility.
  - ◊ The probability of staying remain in a current configuration  $r(x) = 1 - \sum_{y \neq x} p_{xy}$  is strictly smaller than 1 because acceptance probability is always non-zero and there are  $N$  non-zero proposal.
  - ◊ Each spin has two states, and each spin can be chosen, so all  $2^N$  configurations are connectable.
  - ◊ These two arguments lead to strongly connected.
- × Aperiodicity is guaranteed by the sufficient condition with non-zero trace because there exists, for example, a ground state configuration with the probability to persist is  $r(x) = 1 - \sum_{y \neq x} p_{xy} > 0$ , since there are maximum of  $N$  non-zero proposals differ from a current configuration and all associated acceptance probabilities are absolutely smaller than 1 because every available proposal gives raise to the energy.<sup>67</sup>

---

<sup>67</sup>Note that the sufficient condition applied for irriducibility.

- × Irreducibility and aperiodicity constitute ergodicity[26].

Empirically, the only condition needs check is the detailed balance, global balance or whether the desired distribution can be the stationary distribution of the chain or not. While the ergodicity is checked by consistency since, e.g., as the temperature approaches absolute zero, the acceptance probability is effectively zero and ergodicity may be practically broken, or at criticality, the equilibration stage is slow that the exact configuration may not be achieved, or the present of metastable states, truncation, precision, and so on. Therefore, intuitive reasoning is acceptable since rigorous mathematics may be too technically complicated and also may be not very practical.

### C.2.2 Spread of proposal density

Both proposal as well as acceptance probability are purposed to give high acceptance rate. These are the origin of whether corelations are strongly or weakly correlated. Concern the proposal, if the spread is too large[26], a proposal can differ too much from a current configuration resulting in low acceptance rate, and so strong correlation. If the spread is too small, the configuration space is slowly traversed by the algorithm, e.g. the phenomenon of magnetic reversal in local update Metropolis.

Since the proposal density can be customized, different dynamics can be suggested, e.g. cluster algorithms, for example, the Wolff algorithm used in this text, the proposal is very flexible in which a change can be made from a single spin up to some large cluster of spins. Although the proposal appears to be have high spread, however, spin flip is immediately  $\alpha(x, y) = 1$ , so the acceptance rate remains high.

As in [Fig. 27], in the middle figure, at  $T = 0.20(J/k_B)$ , Metropolis random proposal generates configurations differ by only some parts of spins, while in the last figure, even at  $T = 0.01(J/k_B)$ , the configuration space is traversed quickly that generated configurations are far apart.

The spread of proposal is smallest in Metropolis typewriter, then random proposal, checkerboard proposal and the largest and most flexible spread attributed to Wolff single cluster algorithm.

### C.2.3 Typewriter proposal

In typewriter proposal, spins are selected deterministically in some specific order while acceptance probability is like in the random proposal. Configurations generated with this proposal are less correlated than those in random proposal because in the latter, there is chance by which a spin is selected twice that leads to some configuration to be changed back for low energy is prefered.

- × Markovian is certain.
- × Transition probability is conserved since only one spin or also only one configuration is proposed with  $\rho(x, y) = 1$ , the net probability to accept the only proposal is  $\rho(x, y)\alpha(x, y)$  and that to reject is  $\rho(x, y)[1 - \alpha(x, y)]$ , and so the sum is equal to 1.
- × Detailed balance appears to hold because of the expression  $P(x)\rho(x, y)\alpha(x, y) = P(y)\rho(y, x)\alpha(y, x)$  with  $\rho(x, y) = \rho(y, x) = 1$ . However, the proposals are actually

not simultaneously, i.e. not concurrently symmetric since there is no chance for a spin or a configuration be selected consecutively twice.

- × Fortunately, consecution is unnecessary for the whole flow, i.e.  $\sum_x P(x)p_{xy} = \sum_x P(y)p_{yx}$ , for any  $y$ , so the global balance is preserved. But wait, this equality holds only with  $x \neq y$ .
  - ◊ Global balance appears to be detailed balance that simply wears a sum over  $x$ , however, since  $p_{xy}$  is just defined for  $x \neq y$ , let's take a deeper look at the original expression of global balance, since the sum comes with all configurations including  $x = y$ . Global balance reads

$$\begin{aligned} \sum_x P(x)p_{xy} &= P(y) \Leftrightarrow P(x)p_{xy} + P(y)p_{yy} = P(y) \\ &\Leftrightarrow P(x)p_{xy} + P(y) \left[ 1 - \sum_{x \neq y} p_{yx} \right] = P(y) \\ &\Leftrightarrow P(y) + P(x)p_{xy} = P(y) + P(y)p_{yx}, \end{aligned}$$

for  $x \neq y$ .

- × Irreducibility is not easy since the spread of Metropolis typewriter is too small that involves only one configuration while Metropolis random proposal though comes with single spin too but involves  $N$  configurations.
- × If irreducibility holds, aperiodicity is argued the same as in the random proposal. However aperiodicity known to be failed in one dimension with Ising  $Z(2)$  symmetry, e.g. start with a all-up configuration, and spins are being selected in some order, whenever a spin has chance to be flipped, all subsequent proposals are also accepted and the system is in a state of all-down configuration. That's a pattern of periodicity.

#### C.2.4 Checker board proposal

In the checker board update fashion, that can be implemented concurrently in multicore processor systems, like CPU with OpenMP directives, or GPGPU with CUDA, shown in [Fig. 41] with communication pattern is 2D von Neuman stencil<sup>68</sup>, at each update step, half of the spins, e.g. in a square lattice, all spins with odd sum of coordinates are update in a single step, and the other half is flipped in the second step. The update time step, i.e. a sweep, consists of these two step.

The acceptance probability is the same as in the random proposal with a note is that each spin in a half is unaware of the others in that half, i.e. the acceptance probability corresponds to each spin site in that half relates a common old configuration with a proposal one that differ by only that spin.

The autocorrelation time with this proposal should be smaller than in the two previous proposals because of two reasons, the first is spins are selected deterministically avoiding duplication as in typewriter so it has smaller autocorrelation time than in

---

<sup>68</sup>Course of Introduction to Parallel programming with CUDA at Udacity, a lecture in Lesson 2 given by professor Dave Luebke.

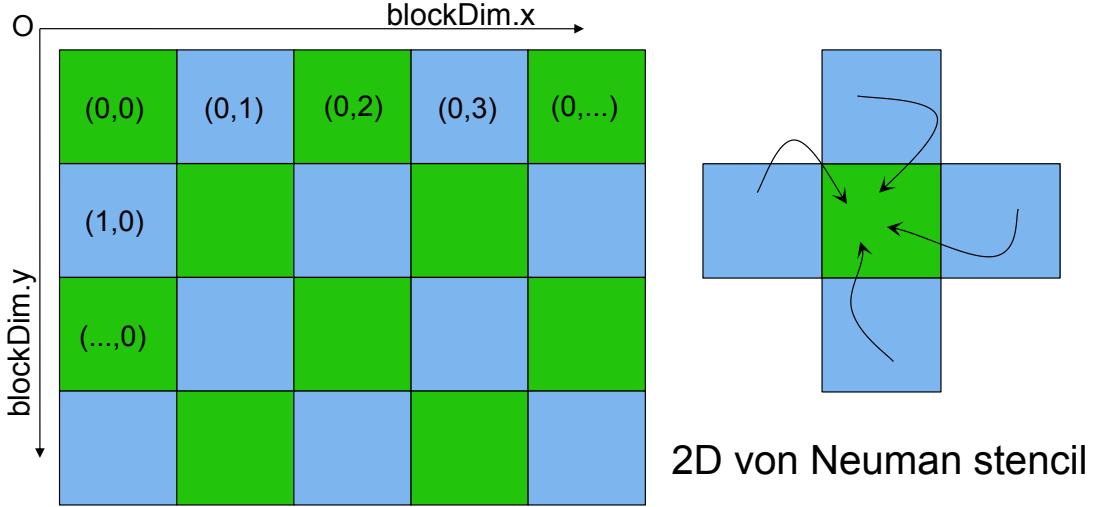


Figure 41: Checker board proposal in Metropolis dynamics.

random proposal, second is the spread of proposal density is larger than in the other two proposals that allows change small as a local spin as well as large groups of spins.

- × Memoryless is obvious because of the proposal density is deterministic and the acceptance probability is the same as in the random proposal.
- × For conservation of transition probability, half of spins is selected and so  $N/2$  changes are proposed.
  - ◊ There is a set of  $2^{N/2} - 1$  independent configurations proposed with proposal  $\rho(x, y) = 1$ , each configuration consists of a sequence of acceptance or rejection a proposal made at each of the half of spin sites with acceptance probability is of form  $\prod_{i=1}^k [1 - \alpha_i(x, y)] \prod_{j=1}^{N/2-k} \alpha_j(x, y)$ , for  $k = 0, 1, \dots, N/2 - 1$ .
  - ◊ The other configuration that is excluded from  $2^{N/2}$  is the one in which all  $N/2$  proposed spins are rejected and the associated probability is the probability for a current configuration to persist is  $[1 - \alpha(x, y)]^{N/2}$ .
  - ◊ The net probability to accept is  $\rho(x, y) \sum_{k=0}^{N/2-1} \prod_{i=1}^k [1 - \alpha_i(x, y)] \prod_{j=1}^{N/2-k} \alpha_j(x, y) + 0$ , and the net probability for all proposals to be rejected is equal to  $\rho(x, y) [1 - \alpha(x, y)]^{N/2}$ , and the sum of these two can be proved to be equal 1 by the method of mathematical deduction.
  - ◊ Consider case  $N = 4$ , the sum should be of the form  $[ab + (1 - a)b + (1 - b)a] + (1 - a)(1 - b)$ , and be equal to 1.
- × Detailed balance failed because a half of spins that is selected and past the trial doesn't have a second chance to be selected immediately.
- × For global balance, let's realize  $k = \overline{0, N/2 - 1}$  permutations of transition probabilities, one of which is  $p(x, y) = \rho(x, y) \prod_{i=1}^k [1 - \alpha_i(x, y)] \prod_{j=1}^{N/2-k} \alpha_j(x, y)$ .  $\rho(x, y)$

comes here is just for mathematical convenience, not that a whole set of configurations is proposed, not a single one. Suppose that  $x$  has been changed to  $y$  which differs from  $x$  by  $N/2 - k$  spins.

$$\diamond \text{ Coming in correspondence with a single mathematical expression, } P(x)\rho(x,y) \\ \prod_{i=1}^k [1 - \alpha_i(x,y)] \prod_{j=1}^{N/2-k} \alpha_j(x,y) = P(y)\rho(y,x) \prod_{i=1}^k [1 - \alpha_i(y,x)] \prod_{j=1}^{N/2-k} \alpha_j(y,x) \\ \alpha_j(y,x) \Leftarrow P(x) \prod_{j=1}^{N/2-k} \alpha_j(x,y) = P(y) \prod_{j=1}^{N/2-k} \alpha_j(y,x) \Leftarrow P(x)\alpha(x,y) = P(y)\alpha(y,x).$$

- The first arrow is because of  $k$  spins were rejected to flip and their corresponding four neighbours are also unchanged that gives their acceptance probabilities symmetric, i.e.  $\alpha_i(y,x) = \alpha_i(x,y)$ ,  $i = \overline{1,k}$ .
- The second arrow is related to the other  $N/2 - k$  independent spins belong to one half of the total spins have been flipped while their four associated neighbours remain independently unchanged, and be argued as following:

† Since energy is additive in terms of short range coupling energy, the product of those  $(N/2 - k)$   $\alpha_j(x,y)$  or  $\alpha_j(y,x)$  factors can be reduced to a single exponential term of energy difference between the two configurations  $x$  and  $y$ .

† Specifically,  $\prod_{j=1}^{N/2-k} \alpha_j(x,y) \equiv \alpha(x,y)$ , and  $\prod_{j=1}^{N/2-k} \alpha_j(y,x) \equiv \alpha(y,x)$ .

† First demonstration, given for the case in which all min functions give the exponential terms in the left product  $[\mathcal{H}_x^{(1)} - \mathcal{H}_y^{(1)}] + [\mathcal{H}_x^{(2)} - \mathcal{H}_y^{(2)}] + \dots + [\mathcal{H}_x^{(N/2-k)} - \mathcal{H}_y^{(N/2-k)}] = \mathcal{H}_x - \mathcal{H}_y$ .

† Second demonstration is for the case in which min functions give value of 1's along with exponential terms in both products, e.g.  $P(x) \times f(\mathcal{H}_x^{(1)}, \mathcal{H}_y^{(1)}) \times 1 \times f(\mathcal{H}_x^{(k)}, \mathcal{H}_y^{(k)}) \times \dots = P(y) \times 1 \times f(\mathcal{H}_y^{(2)}, \mathcal{H}_x^{(2)}) \times 1 \times \dots$ , where  $f(u,v) = \exp[(u-v)/k_B T]$ . This is obviously equivalent to  $P(x) \times f(\mathcal{H}_x^{(1)}, \mathcal{H}_y^{(1)}) \times f(\mathcal{H}_x^{(2)}, \mathcal{H}_y^{(2)}) \times f(\mathcal{H}_x^{(k)}, \mathcal{H}_y^{(k)}) \times \dots = P(y) \times 1 \times 1 \times 1 \times \dots$ , and this returns to the first demonstration.

- ◊ And the same treatment with the case of typewriter proposal can be used to prove the global balance.

- × Irreducibility is possible. It is stronger than in typewriter proposal but weaker than in random proposal. However, it has chance to be failed because spin selection is deterministic.

- ◊ Each configuration is connected to at least a set of  $2^{N/2} - 1$  configurations that is  $\ll 2^N$  in total.

- ◊ Persistent probability is  $r(x)$  is strictly smaller than unity.

- × If irreducibility holds, aperiodicity can be argued since, e.g., a ground state has  $r(x) = [1 - \alpha(x,y)]^{N/2}$  definitely be greater than zero.

In parallel programming and in particular GPGPU programming, we will also implement parallel reduction [Fig. 42]<sup>69</sup> very frequently. Note that in general the reduce primitive comes with a binary operation that is required to be associative, however, in our case, it is just a simple addition in the set of real numbers. An important note is that *race condition* is a general pitfall in deploying parallelism.

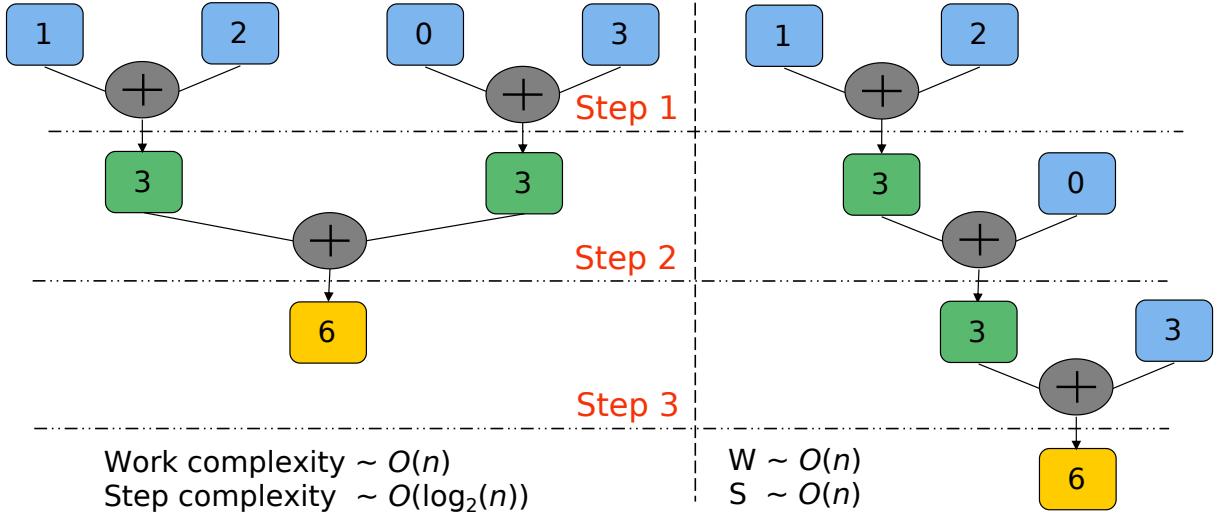


Figure 42: Diagrams of parallel reduce vs serial reduce.

A final note is that, there is also a typewriter proposal in which every spin in the lattice is consecutively selected in some order. This proposal is known to break ergodicity in one dimension, while it can be used in two dimensions. Although it should be faster than the random proposal, no solid confirmation on its ergodicity in two dimensions, it won't be reported here.

### C.3 Wolff dynamics

#### C.3.1 Generalized spin flip operation

The Wolff algorithm is a cluster algorithm introduced in Ref. [13] that the author considered the general  $\mathcal{O}(n)$  model in which the spin flip operation  $\mathbf{s}_{ij} \rightarrow -\mathbf{s}_{ij}$  in the Ising model is generalized to be as the reflection of a spin  $\mathbf{s}_{ij}$  with respect to the hyperplane orthogonal to any chosen vector  $\mathbf{r} \in \mathcal{S}_{n-1}$ ,  $n \geq 2$ , i.e.

$$\mathbf{R}(\mathbf{r})\mathbf{s}_{ij} = \mathbf{s}_{ij} - 2(\mathbf{s}_{ij} \cdot \mathbf{r})\mathbf{r} \quad (98)$$

As seen in (98), a zero vector  $\mathbf{r}$  means no change proposal, fortunately every spin state vector is nonzero in the sense that their magnitude is equal to unity. (98) can be interpreted as first find the perpendicular projection of  $\mathbf{s}_{ij}$  on  $\mathbf{r}$  and then some trivial vector operations. The angle

$$\theta'_{ij} = 2r - \theta_{ij} + \pi \quad (99)$$

---

<sup>69</sup>Adapted, reproduced a lecture in Lesson 3 in Parallel programming with CUDA course at Udacity given by professor John Owens.

can be deduced by the following elementary trigonometric transformation

$$\begin{aligned}\cos \theta'_{ij} &= \cos \theta_{ij} - 2(\cos \theta_{ij} \cos r + \sin \theta_{ij} \sin r) \cos r \\ &= \cos \theta_{ij} (1 - 2 \cos^2 r) - 2 \sin \theta_{ij} \sin r \cos r \\ &= -\cos \theta_{ij} \cos 2r - \sin \theta_{ij} \sin 2r \\ &= \cos(2r - \theta_{ij} + \pi) \\ \sin \theta'_{ij} &= \sin(\theta_{ij} - 2r) = \sin(2r - \theta_{ij} + \pi)\end{aligned}$$

$\mathcal{S}_{n-1}$ , which is a set of unit vectors whose components represent angles of the spins taking values in the field of real number (with two binary operators are the typical vector addition and scalar - vector multiplication), can be proved in some trivial task to be a vector space over the field [25, pp. 159, 160] (note that the magnitude of unity is fixed and takes no part in the set). This vector space can be associated with the vector dot product as an inner product [25, p. 286] on  $\mathcal{S}_{n-1}$ . The Wolff-said hyperplane orthogonal to  $\mathbf{r}$  is actually the *orthogonal complement* [25, p. 403] of the subset  $\{\mathbf{r}\}$  (contains only  $\mathbf{r}$ ) of  $\mathcal{S}_{n-1}$  which is a set  $\mathbf{r}^\perp$  of all vectors in the inner-product space  $\mathcal{S}_{n-1}$  that are orthogonal to every vector in the subset  $\{\mathbf{r}\}$ , i.e.

$$\mathbf{r}^\perp = \{\mathbf{s} \in \mathcal{S}_{n-1} \mid \langle \mathbf{s} | \mathbf{r} \rangle = \mathbf{s}^* \mathbf{r} = 0 \forall \mathbf{r} \in \{\mathbf{r}\}\} \quad (100)$$

For any nonzero  $\mathbf{r}_{1 \times (n-1)} \in \mathcal{S}_{n-1}$ , the elementary reflector [25, pp. 324, 325] about  $\mathbf{r}^\perp$  is defined as

$$\mathbf{R} = \mathbf{I} - 2 \frac{\mathbf{r} \mathbf{r}^*}{\mathbf{r}^* \mathbf{r}}, \text{ or simply } \mathbf{R} = \mathbf{I} - 2 \mathbf{r} \mathbf{r}^T \quad (101)$$

since  $\|\mathbf{r}\| = \mathbf{r}^* \mathbf{r} = \mathbf{r}^T \mathbf{r} = 1$  and  $\mathbf{r}$  is real valued.

The reflector (101) acting on  $\mathbf{s}_{ij}$ , by the distributive and associative laws of matrix multiplication [25, p. 105], gives

$$\mathbf{R} \mathbf{s}_{ij} = (\mathbf{I} - 2 \mathbf{r} \mathbf{r}^T) \mathbf{s}_{ij} = \mathbf{s}_{ij} - 2 \mathbf{r} \mathbf{r}^T \mathbf{s}_{ij} = \mathbf{s}_{ij} - 2 \mathbf{r} (\mathbf{r}^T \mathbf{s}_{ij}) \quad (102)$$

(102) is equivalent to (101), i.e. the reflector mentioned in [13] is actually an elementary reflector, *a.k.a* Householder transformation<sup>70</sup>, which is involutory ( $\mathbf{R}^2 = \mathbf{I}$ , which agrees with  $\mathbf{R}(\mathbf{r})^2 = 1$  in Ref. [13]) [25, p. 325], but not idempotent as given in [13] since idempotent operator satisfies  $\mathbf{R}^2 = \mathbf{R}$  [25, p. 386], idempotent is the case if the constant 2 in (101) is omitted and the reflector becomes a projector.

From (102), the term  $2(\mathbf{r}^T \mathbf{s}_{ij})$  is simply a *real* scalar, then the reflection  $\mathbf{R} \mathbf{s}_{ij}$  of  $\mathbf{s}_{ij}$  is obvious a spin vector in  $\mathcal{S}_{n-1}$  due to the properties of a vector space. In addition, since the reflector is involutory, applying it a second times will return to the original spin vector.

The reason why the spin flip operation must be generalized for spins with continuous spin state space  $\mathcal{S}_{n-1}$  is for the invariance of the partition function (67) under global  $\mathbf{R}$  transformation[13]

$$Z \equiv Z(T) = \prod_{i,j} \int_{\mathcal{S}_{n-1}} d\mathbf{s}_{ij} \exp \left[ \sum_{\langle ij, lm \rangle} J_{ijlm} \mathbf{s}_{ij}^T \mathbf{s}_{lm} / k_B T \right], \quad (103)$$

---

<sup>70</sup>Alston Scott Householder (1904 - 1994), one of the first people to appreciate and promote the use of elementary reflectors for numerical applications, whose passion was mathematical biology. [25, p. 324]

The invariance is simply due to the existence of a one to one correspondence, for every pair of spins<sup>71</sup>  $(\mathbf{s}_{ij}, \mathbf{s}_{lm})$ ,

$$[\mathbf{R}\mathbf{s}_{ij}]^T \mathbf{R}\mathbf{s}_{lm} = \mathbf{s}_{ij}^T \mathbf{R}^T \mathbf{R}\mathbf{s}_{lm} = \mathbf{s}_{ij}^T \mathbf{s}_{lm} \quad (104)$$

It can be seen that, the flipping fashion in which a spin is directly proposed by a random value in  $[0, 2\pi)$ , violates this invariance since this is not appropriate to be a geometrical mirror in the sense of a global flip. Also, the generalized spin flip would be appropriate for the concept of flipping operation, since it is involutory while the random flip is indeterministic. However, in practical simulation, there would be no recognizable discrepancy between this fashion and the generalized fashion for Metropolis algorithm, since single spin site is updated in a step, there would be no difference with taking a spin angle at random provided the random number generator, e.g. Mersenne Twister has long period.

### C.3.2 Markov properties

The transition probability  $p_{xy} = \rho(x, y)\beta(x, y)\alpha(x, y)$ , where the proposal density is symmetric, i.e.  $\rho(x, y) = \rho(y, x) = 1/N$ , the acceptance probability is symmetric,  $\alpha(x, y) = \alpha(y, x) = 1$ , and  $\beta(x, y)$  is the probability of forming a cluster, i.e. a product of probabilities adding a spin to the cluster  $\mathcal{P}(x, y)$  as well as those of spins that are refused to join the cluster  $1 - \mathcal{P}(x, y)$ , where the bonding activation probability  $\mathcal{P}(x, y)$  is also symmetric for any pair of spins in a cluster as can be seen in the argument given for the invariance of the partition function in (104) while it is not for a spin in a cluster and a spin outside that cluster<sup>72</sup>.

- × For detailed balance, the formation of a cluster requires the existence of non-flipping spins at the boundary of the cluster, then

$$\begin{aligned} \frac{p_{xy}}{p_{yx}} &= \prod_{\langle ij, lm \rangle \in \partial c} \frac{1 - \mathcal{P}(\mathbf{R}\mathbf{s}_{ij}, \mathbf{s}_{lm})}{1 - \mathcal{P}(\mathbf{R}\mathbf{s}'_{ij}, \mathbf{s}'_{lm})} \\ &= \prod_{\langle ij, lm \rangle \in \partial c} \frac{\exp \left( \min \left\{ 0, -J_{ijlm} \left[ (\mathbf{R}\mathbf{s}_{ij})^T (\mathbf{R}\mathbf{s}_{lm}) - (\mathbf{R}\mathbf{s}_{ij})^T \mathbf{s}_{lm} \right] / k_B T \right\} \right)}{\exp \left( \min \left\{ 0, -J_{ijlm} \left[ \mathbf{s}'_{ij}^T (\mathbf{R}\mathbf{s}_{lm}) - \mathbf{s}'_{ij}^T \mathbf{s}_{lm} \right] / k_B T \right\} \right)} \\ &= \prod_{\langle ij, lm \rangle \in \partial c} \exp \left\{ J_{ijlm} \left[ (\mathbf{R}\mathbf{s}_{ij})^T \mathbf{s}_{lm} - \mathbf{s}'_{ij}^T \mathbf{s}_{lm} \right] / k_B T \right\} \\ &= \prod_{\langle ij, lm \rangle \in \partial c} \exp \left[ J_{ijlm} \left( \mathbf{s}'_{ij}^T \mathbf{s}'_{lm} - \mathbf{s}_{ij}^T \mathbf{s}_{lm} \right) / k_B T \right] \\ &= \exp \left[ \sum_{\langle ij, lm \rangle \in \partial c} J_{ijlm} \left( \mathbf{s}'_{ij}^T \mathbf{s}'_{lm} - \mathbf{s}_{ij}^T \mathbf{s}_{lm} \right) / k_B T \right] = \exp [-(\mathcal{H}_t - \mathcal{H}_c) / k_B T] \end{aligned}$$

---

<sup>71</sup>From this, if we set  $\mathbf{s}_{lm} = \mathbf{s}_{ij}$ , it is trivial that the flipped spin has the same norm and then be an element of the spin state space. This trivial insight learned recently from Chen, Chen, *Ph.D. dissertation*, The University of Hongkong, 2011. However, our approach to involutory operation makes things trivial and helps geometrical visualization. Also, our derivation of  $\theta_{ij} = 2r - \theta_{ij} + \pi$  is sufficient to say the proposal vector is satisfied provided the angle is put in the range  $[0, 2\pi)$ .

<sup>72</sup>The second argument in the minimum function is anti-symmetric then.

where the last expression is actually the ratio  $P(y)/P(x)$ , the present of the operator  $\mathbf{R}$  in the first arguments of probability  $\mathcal{P}$  is because of when a trial  $\langle ij, lm \rangle \in \partial c$  is launched, either spins  $\mathbf{s}_{ij}$  or  $\mathbf{s}'_{ij}$  has already been flipped. It is due to the involutory of the reflection,  $\mathbf{R}\mathbf{s}'_{ij} = \mathbf{s}_{ij}$ , and spins outside the cluster stay remain,  $\mathbf{s}'_{lm} = \mathbf{s}_{lm}$ , and the last equality is argued as the difference in energy of a current configuration and its proposal configuration is the difference in the bonding energy between spins in the cluster and the spins just outside the cluster.

- × Proposal, bond activation probability relates only one current configuration with many proposal configuration that differ by a spin at different sites. Acceptance probability is always equal to 1.
- × Since whenever a spin is chosen it is then flipped immediately, there is only two kinds of flow, accepted with cluster formation with net probability  $\sum_N \rho(x, y)\beta(x, y)$   $\alpha(x, y) + 0 = \sum_N \beta(x, y)/N$ , and accepted with no formation of cluster  $\sum_N \rho(x, y)[1 - \beta(x, y)]$   $\alpha(x, y) + 0 = \sum_N [1 - \beta(x, y)]/N$ . The sum of these two flows is equal to 1. Transition probability is conserved.
- × It is hard better not to argue for irreducibility for the failure of aperiodicity since all diagonal elements of the transition matrix is equal to zero. This is because every non-zero proposal is always accepted regardless cluster formation.
- × Primitivity holds by the Frobenius test which is passed easily after not too many updates of the transition matrix since clusters can be formed in any form from a single spin site to some large collection of spins, even of all spins, so that any configuration can be proposed and the transition matrix should have all elements except diagonal ones are non-zero.

## D Data analysis and equilibrium dynamics

### D.1 Statistics of correlations

For any random variables  $X$  and  $Y$  having finite means  $\mu_X$ ,  $\mu_Y$ , the *covariance*  $\text{Cov}(X, Y)$  of  $X$  and  $Y$  is defined as [22, pp. 248, 249]

$$\text{Cov}(X, Y) = \text{E}[(X - \mu_X)(Y - \mu_Y)] \quad (105)$$

Provided the variances are finite, (105) is rewritten as

$$\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) \quad (106)$$

If the covariance (105) is positive, then the two random variables have values smaller or larger than their corresponding mean together. If the covariance is negative, the values of the two variables are related to each other as that if one is larger than its mean, then the other is smaller than its mean and vice versa. In these two cases,  $X$  and  $Y$  are said to be correlated, the former is termed as positively correlated and the latter is negatively correlated. The case  $\text{Cov}(X, Y) = 0$ , the value of  $X$  can be arbitrarily smaller or larger than its corresponding mean regardless the behaviour of  $Y$ , they are said to be uncorrelated. [22, pp. 249, 251]

The *correlation* of  $X$  and  $Y$  is defined as [22, p. 250]

$$\varrho(X, Y) = \text{Cov}(X, Y) / (\sigma_X \sigma_Y) \quad (107)$$

According to the *Cauchy - Schwarz Inequality* [22, p. 251]

$$-1 \leq \varrho(X, Y) \leq 1 \quad (108)$$

(108) means that (107) actually plays the role of normalizing the magnitude of the covariance.

Finally, correlation measures only linear relationship[22, p. 253], in this Ref. an example, for brevity, of  $V = \Lambda^2$ , with  $E(\Lambda) = 0$ , then  $\text{Cov}(V, \Lambda) = 0$  but  $\Lambda, V$  are functionally related. Fortunately, since our concern is the relation of different measurements of a single quantity that act as random variables which are not some explicit function of each other so non-linear relation would not be expected.

## D.2 Equilibrium dynamics

### D.2.1 Autocorrelation function

As mentioned, the spread of proposal density and the choice of acceptance probability of an algorithm can lead to strongly or weakly correlated configurations. Then let's consider the autocorrelation function  $\psi_A(t)$  of an observable  $A$  which relates the measured values  $A(0)$  at some time origin and  $A(t)$  at some time  $t$  is given by

$$\psi_A(t) = [\langle A(0)A(t) \rangle - \langle A(0) \rangle \langle A(t) \rangle] / [\sigma_{A(0)}\sigma_{A(t)}] \quad (109)$$

When the system is in equilibrium, the averages in(109) are taken over every time origin since in equilibrium there is an invariance with respect to the choice of the time origin [7, p. 187] [12, p. 635], then  $\langle A(0) \rangle = \langle A(t) \rangle = \langle A \rangle$  and  $\sigma_{A(0)} = \sigma_{A(t)} = \sigma_A$ , and (109) reduced to

$$\psi_A(t) = [\langle A(0)A(t) \rangle - \langle A \rangle^2] / \sigma_A^2 \quad (110)$$

where  $\sigma_A^2$  is estimated by  $\hat{\sigma}_A^2$  in (72). The estimation of (110) is quite ambiguous that needs to be made clear. If we estimate expectation of the pair by its arithmetic mean and combine with the square of that of the observable  $A$ , we would suffer discrepancy immediately at  $t = 0$ , since  $\sigma_A^2$  is unbiased estimated by  $\hat{\sigma}_A^2$ , not simply the sample deviation<sup>73</sup>. Also the estimation of error of the numerator would be not easy, even when correlation hasn't been considered, i.e. correlation in measurements<sup>74</sup> since the variance of a sum is not simply an addition of variance of constituents, i.e. the two quantities being estimated from the same series would not be expected to be uncorrelated. Then, for consistency in statistics, (110) is changed to the following form

$$\psi_A(t) = \langle A(0)A(t) - A^2 \rangle / \sigma_A^2 + 1 \quad (111)$$

---

<sup>73</sup>Which has the fraction  $1/q$  instead of  $1/(q-1)$  in (72).

<sup>74</sup>Being computed from a large number of uncorrelated measurements,  $\bar{A}$  obeys normal distribution, then  $\bar{A}^2$  obeys the  $\chi^2$  distribution with one degree of freedom.[22, p. 471] Therefore,  $\sigma_{\bar{A}^2}^2 = 2$ . Let's emphasize that we can estimate  $\sigma_{\bar{A}^2}^2$  but the combining estimate of the means would still be incorrect.

An unbiased estimator of the numerator is

$$\bar{n}_{\psi_A}(t) = \frac{1}{t_A - t} \sum_{x=1}^{t_A-t} n_{\psi_A}(x, t) = \frac{1}{t_A - t} \sum_{x=1}^{t_A-t} [A(x)A(x+t) - A^2(x)] \quad (112)$$

where  $t_A$  is the observation time associated with quantity  $A$ .

It would be more accurate if  $\langle A^2 \rangle$  is estimated separately, since for large time  $t$ , the amount  $t_A - t$  decreases significantly. Unfortunately, the pair as well as  $\langle A^2 \rangle$  are estimated from the same time series, we do not expect they are uncorrelated and then the variance on the numerator can't be a simple addition. We later favoured the latter when we are deploying resampling methods.

This estimator (112) for large  $t_A - t$  would obey normal distribution, unfortunately, the measurements are mostly correlated, so practically one should not expect such an asymptotical approximation.

We have been able to give the estimators for the enumerator and the denominator, both are estimated unbiasedly, however, how to estimate the ratio. This is not trivial as one first may thought, since the expectation of a ratio is in general not equal to the ratio of expectations and though we may agree that the estimation based on the ratio of these expectation should be a biased one, how do we estimate the variance since we do not know the distribution of the estimator of the mean of the ratio. And by taking the expectation of a ratio to be the ratio of expectations, the estimator would be biased. How to reduce the bias? We then come to the introduction of *resampling methods*. And then 112 will not be used anymore but the two terms are being estimated separately. However, once again, these may not be very important, since not all data in the series are uncorrelated and  $t \ll t_A$ .

It can be argued that  $\psi_A(t)$  are non-negative, since in a Markov chain, adjacent configurations are correlated in the sense that they differ from each other only by some small part or even no difference at all<sup>75</sup> and then the statistical properties are quite similar and the values of a quantity measured on these adjacent configurations should be positively correlated at most, i.e. they are equally large or equally small relative to the mean. Hence, for large time,  $\psi_A(t)$  tends to zero.

### D.2.2 Exponential autocorrelation time

$\psi_A(t)$  is expected to decay exponentially at *large time*.<sup>76</sup> as

$$\psi_A(t) \sim \exp(-t/\tau_A^{(exp)}) \quad (113)$$

$\tau_A^{(exp)}$  termed as the exponential autocorrelation time by which  $\psi_A(t)$  falls off a factor of  $1/e$ . It is obvious that correlation still presents with this amount of time. Nevertheless, it is non-trivial to specify the time window in which  $\psi_A(t)$  truly decays exponentially and there is not necessary a unique exponential term in (113), though only the largest  $\tau_A^{(exp)}$  is considered to be the exponential autocorrelation time.

---

<sup>75</sup>For the probabilistic nature of evolution, there's chance that the chain stays remain at a current configuration.

<sup>76</sup>Read pp. 60, 65 Ref. M. Newman and G. Barkema. *Monte Carlo Methods in Statistical Physics*. OUP, 2001. Roughly speaking,  $\tau^{(exp)}$  is related to the second largest eigenvalue of the transition matrix in a finite state space Markov chain, i.e. the Ising model.

Another correlation time termed integrated autocorrelation time  $\tau_{(int)}$ , which will be derived shortly, associated with a specific quantity, is given by

$$\tau_A^{(int)} = \sum_{t=0}^{\infty} \psi_A(t) \quad (114)$$

In general,  $\tau_A^{(int)} \neq \tau_A^{(exp)}$ , except the case (113) holds ideally,  $\tau_A^{(int)} \sim \tau_A^{(exp)}$ , since

$$\sum_{t=0}^{\infty} \psi_A(t) \approx \int_0^{\infty} \psi_A(t) dt \sim \int_0^{\infty} \exp\left(-t/\tau_A^{(exp)}\right) dt = -\tau_A^{(exp)} \exp\left(-t/\tau_A^{(exp)}\right)\Big|_0^{\infty} = \tau_A^{(exp)}$$

### D.2.3 Integrated autocorrelation time

Previously, standard error of the mean was discussed. However, the situation here is more complicated, since the adjacent measurements are correlated depending on the associated dynamics. Consider the variance of the mean from the exact expectation of quantity  $A$ ,<sup>77</sup>

$$\sigma_{Ac}^2 = E[\bar{A} - \langle A \rangle]^2 \quad (115)$$

$$= E\left[\frac{1}{q} \sum_{x=1}^q A_x - \langle A \rangle\right]^2 \quad (116)$$

$$= E\left[\frac{1}{q} \sum_{x=1}^q A_x\right]^2 - E\left[\frac{2}{q} \langle A \rangle \sum_{x=1}^q A_x\right] + \langle A \rangle^2 \quad (117)$$

$$= \frac{1}{q^2} E\left[\sum_{x=1}^q A_x \sum_{y=1}^q A_y\right] - \langle A \rangle^2 \quad (118)$$

$$= \frac{1}{q^2} E\left[\sum_{x=1}^q A_x^2 + \sum_{x=1}^q \sum_{\substack{y=1 \\ y \neq x}}^q A_x A_y\right] - \langle A \rangle^2 \quad (119)$$

$$= \frac{1}{q} (\langle A^2 \rangle - \langle A \rangle^2) + \frac{1}{q^2} \sum_{x=1}^q \sum_{\substack{y=1 \\ y \neq x}}^q \langle A_x A_y \rangle - \frac{q(q-1)}{q^2} \langle A \rangle^2 \quad (120)$$

Since the double sum has  $q(q-1)$  products of  $\langle A_x A_y \rangle$ ,  $y \neq x$ , then (120) is equivalent to

$$\sigma_{Ac}^2 = \frac{1}{q} (\langle A^2 \rangle - \langle A \rangle^2) + \frac{1}{q^2} \sum_{x=1}^q \sum_{\substack{y=1 \\ y \neq x}}^q [\langle A_x A_y \rangle - \langle A \rangle^2] \quad (121)$$

With the note that in the double sum of (121), each term  $\langle A_x A_y \rangle$  appears twice, and by concerning the time separation  $t = |y - x|$  between two measurements,  $\langle A_x A_y \rangle :=$

---

<sup>77</sup>Read page 33 Ref. K. Binder and D. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction*. Graduate Texts in Physics. Springer Berlin Heidelberg, 2010.

$\langle A(0)A(t) \rangle$ , and it can be seen that each such a pair appears  $q - t$  times, then (121) is rewritten as<sup>78</sup>

$$\sigma_{Ac}^2 = \frac{1}{q} (\langle A^2 \rangle - \langle A \rangle^2) + \frac{2}{q^2} \sum_{t=0}^q (q-t) [\langle A(0)A(t) \rangle - \langle A \rangle^2] \quad (122)$$

$$= \frac{1}{q} (\langle A^2 \rangle - \langle A \rangle^2) \left[ 1 + 2 \sum_{t=0}^q \left( 1 - \frac{t}{q} \right) \frac{\langle A(0)A(t) \rangle - \langle A \rangle^2}{\langle A^2 \rangle - \langle A \rangle^2} \right] \quad (123)$$

$$= \sigma_A^2 \left[ 1 + 2 \sum_{t=0}^q \left( 1 - \frac{t}{q} \right) \psi_A(t) \right] \quad (124)$$

where  $\sigma_A^2$  is given in (77) by treating all measurements as if they were uncorrelated and  $\psi_A(t)$  is given by (110).

For large  $q$ ,  $\psi_A(t) = 0$  upto some upper bound of  $t$ ,  $t_{sup} \ll q$ , i.e. time  $t$  contributed in (124) is small compared with  $q$ , then  $t/q \ll 1$ , in terms of continuous time, (124) is replaced by

$$\sigma_{Ac}^2 = \sigma_A^2 \left[ 1 + 2 \int_0^\infty \psi_A(t) dt \right] \quad (125)$$

And the integrated autocorrelation time is defined to be

$$\tau_A^{(int)} = \int_0^\infty \psi_A(t) dt \quad (126)$$

where the discrete version given in (114).

$\tau_A^{(int)}$  is empirically computed by contributing as many terms to the sum of  $\psi_A(t)$  as possible and looking for some saturation signal. The sum is cut where  $\psi_A(t)$  changes from positive value to negative value for the first time, or at the first significant minimum<sup>79</sup>.

Finally, since having the mathematical expression (125), the estimated  $\tau_A^{(int)}$  can be substituted in that and combining with the naive error  $\sigma_A$ , we then get the exact error.

### D.3 Binning analysis

As a result from (125), the integrated autocorrelation time can be computed by the ratio of the two kinds of errors [Fig. 43(a)], namely

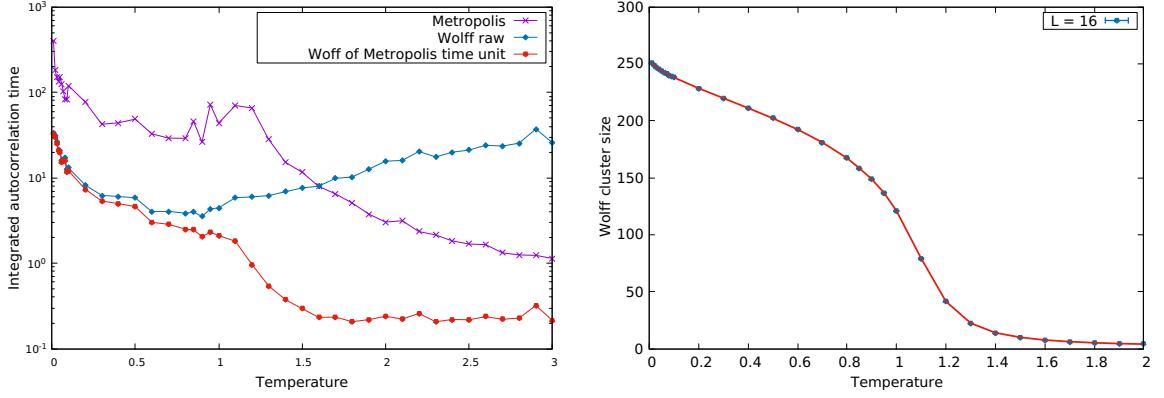
$$\tau_A^{(int)} = (\sigma_{Ac}^2 / \sigma_A^2 - 1) / 2 \approx \sigma_{Ac}^2 / 2\sigma_A^2 \quad (127)$$

Hence, we can have a set of uncorrelated measurements, the standard error analysis can be applied. Then by forming  $q_b$  number of bins of measurements, each bin with the bin size  $q$  is large so that all the measurements in a bin are correlated to each other but not to those outside the bin, and then the arithmetic mean computed in each bin

---

<sup>78</sup>For  $t = q$ , term  $\langle A(0)A(t) \rangle$  appears no times, and for  $t = 0$ ,  $\langle A(0)A(0) \rangle - \langle A \rangle^2 = 0$ .

<sup>79</sup>By significant, we mean either the depth of or the range in which the minimum is preserved.



(a) Norm of magnetization per spin for Wolff and Metropolis measured by estimating the ratio of variances.

(b) Average cluster size with Wolff dynamics.

Figure 43:  $L = 16$ . Wolff is very efficient even at high temperatures due to the presence of cluster size containing more than a single spin. Note that in our simulation the raw results of Wolff used rather than the scaled ones.

together with the those in the other bins can be treated as statistically uncorrelated. This is the so-called *binning analysis*, see e.g. [Fig. 44(a)]. The act of computing arithmetic mean in each bin instead of choosing some value of some particular measurement is to exploit the law of large number which states that the arithmetic mean converges in probability to the exact mean of the corresponding observable.[22, p. 352]

Furthermore, by (125) and (126), by taking adjacent measurements separated by some amount  $\tau_A^{(int)}$ , the standard error analysis for uncorrelated measurements can be applied since if  $\tau_A^{(int)} = 0, \sigma_{Ac}^2 = \sigma_A^2$ . Rejecting correlated configurations would be feasible for later shown GPGPU parallel simulation since flipping time would be reduced significantly and without transaction between memories through the slow PCI express card, the simulating time doesn't increase much. However, since the estimate of expectation doesn't require measurements to be uncorrelated, rejecting large number of configurations would waste computer resource. Also, in practical simulation,  $\tau^{(int)}$  gives an approximation of bin size by which correct statistical errors can be acquired. Once again, (127) involves estimation of a ratio, and we can't refuse to meet *resampling methods* if we are going to estimate the autocorrelation time via this analysis. In the estimation of expectations by binning analysis, the appropriate bin size should be the one at which the binned variance stops its initial increasing trend to show some strange or wild behaviour. See [Figs. 44(a), (b)].

Concerning the binning analysis, we disagree with some argument<sup>80</sup> in which the author argued that the sample means in the bins are approximately Gaussian due to the central limit theorem and then concluded that  $\sigma_A^2$  is analytically known from the  $\chi^2$  distribution. This is nonsense since the the central limit theorem requires the sample to be independent[22, pp. 158, 361] while all configurations in a bin are mostly correlated and the bin size is practically just of the order of magnitude of  $\tau^{(int)}$  by which adjacent bins would be uncorrelated, i.e.  $\tau^{(int)}$  may not be very large. Therefore we

<sup>80</sup>See p. 203 Ref. B.A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis: With Web-based Fortran Code*. World Scientific, 2004.

can't say that binned means are normally distributed and in order to estimate the error of quantities that are being derived from some variance, we again rely on *resampling methods*.

Return to (127),  $\sigma_{Ac}^2$  is estimated by

$$\sigma_{Ab}^2 = \left( \overline{A_b^2} - \overline{A_b}^2 \right) / (q_b - 1) \quad (128)$$

where  $A_b$  is the sample mean in each bin, i.e.  $A_b = \sum_{x=1}^q A_x / q$ .

While,

$$\sigma_A^2 = \left( \overline{A^2} - \overline{A}^2 \right) / (t_A - 1) \quad (129)$$

where  $t_A$  is the observation time measured in number of update sweeps.

Historically, it was reported in Ref. [17, pp. 765, 766],<sup>81</sup><sup>82</sup> which reported for Ising model that the binning method is not very accurate even compared with the direct computation of integrated autocorrelation time which is itself already not very accurate due to the growing of statistical errors dominating dying exact statistics. As we have seen, if  $\tau^{(int)} = 0$  the two variances are equal, this is the decorrelation time, then the bin size should be of the same order with it. Unfortunately, we later observed that the bin size is much larger than  $\tau^{(int)}$ . We argue this by the fact that configurations generated by Markov chain are correlated, but this is not deterministic but probabilistic, therefore, the decorrelation time is different at different moments that we can't take a unique bin size for every bin, at some region of low correlation, the bin may cover not only a range of correlated measurements but also other measurements that are not correlated to these correlated range too while at some other region of high correlation, a bin may not cover enough the range of correlation, one shoule take a look at some equilibrium states given in [Fig. 11(a)], different degrees of correlation shown, e.g. Cold 4. This is why the bin size is actually larger than the decorrelation time (which is taken on average) since more number of uncorrelated measurements needed to reduce the effect of correlated ones in each bin, the optimal bin size should be the largest decorrelation time of all moments.

The binning analysis can't be used to estimate quantities like specific heat and susceptibility since they are fluctuating quantites that relate to the variances which are computed on the whole time series of measured observables.

## D.4 Resampling methods: Jackknife estimator

There are two methods in the class of resampling methods, one is the Jackknife and the other named bootstrap. The former, which was introduced in 1949 by Maurice Quenouille and generalized in 1956 by the author, is a standard in literature and rather familiar that will be implemented throughout of this text.

Jackknife is based on the observation that, in terms of large number of measurements  $n$ , the inverse of which is small compared with unity and the bias of the estimator

---

<sup>81</sup>Which cites Ref. Neal Madras and Alan D. Sokal. *J. Stat. Phys.*, 50(1-2):109–186, 1988 that needs to be tracked down.

<sup>82</sup>And also in p. 207 Ref. B.A. Berg. *Markov Chain Monte Carlo Simulations and Their Statistical Analysis: With Web-based Fortran Code*. World Scientific, 2004.

can be expanded in power series in term of this inverse. By dividing the samples into two parts, the bias can be reduced to the order of  $O(n^2)$  instead of  $O(n)$  in the original treatment. See, e.g. [?] In this thesis,  $d$ -Jackknife is implemented, i.e. divide the

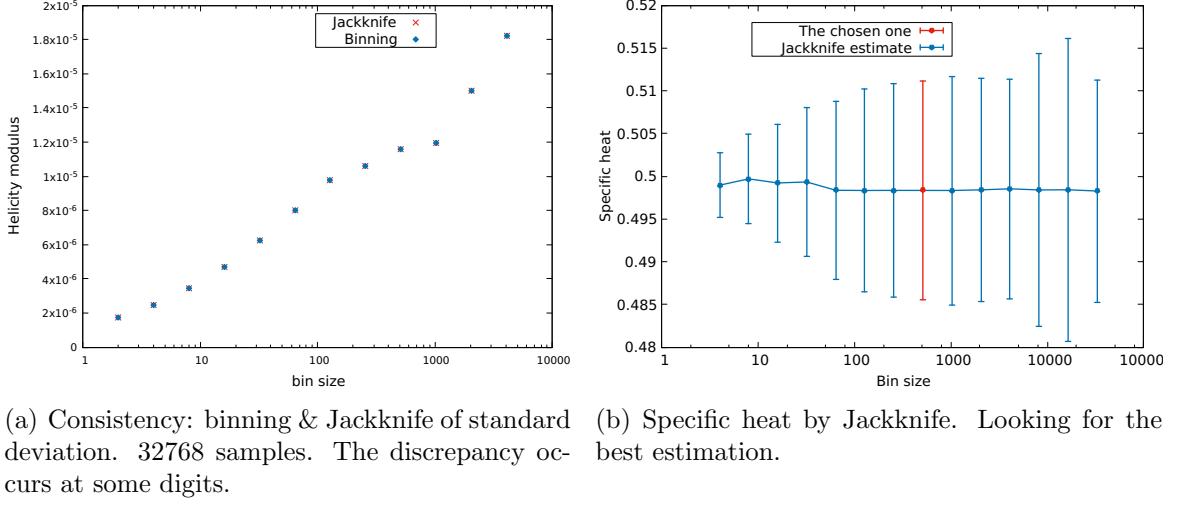


Figure 44:  $L = 16, T = 0.01(J/k_B)$ .

original sample into some number of  $d$  bins, the new estimator is the difference of two quantities, first is the total number of bins multiplied by the original estimator taken on the original sample, the second is the net number of bins minus 1 multiplied by the original estimator taken on the sample of all bins except the one which is being deleted from the original sample. Each of new estimators is then being treated as if they are uncorrelated and their mean which gives the final estimation obeying the normal distribution.

A quick introduction of Jackknife is now given. Consider a sample  $\mathbf{X} = \{X_1, X_2, \dots, X_r\}$ , which is being divided into  $n$  bins of size  $q = r/n$ . Let's  $\theta(\mathbf{X})$  be the parameter that needs to be estimated,  $\hat{\theta}(\mathbf{X})$  be the estimator taken on the original sample,  $\hat{\theta}(\mathbf{X}_i)$  be that taken on the sample of all bins except  $i$ -th bin. The Jackknife estimator of  $\theta(\mathbf{X})$  corresponding to  $i$ -th bin is given by

$$\hat{\theta}_J(\mathbf{X}_i) = n\hat{\theta}(\mathbf{X}) - (n-1)\hat{\theta}(\mathbf{X}_i) \quad (130)$$

(130) can be rewritten as

$$\hat{\theta}_J(\mathbf{X}_i) = \hat{\theta}(\mathbf{X}) + (n-1) [\hat{\theta}(\mathbf{X}) - \hat{\theta}(\mathbf{X}_i)] \quad (131)$$

From (131), the Jackknife estimator can be roughly interpreted as bias-correction. The theory of Jackknife estimator or bootstrap should be studied if time permits. One important note is that using Jackknife estimator needs to avoid quantities that are not very continuous which would yield undetermined results. E.g. in estimation of ratio of variances, or the autocorrelation function, it's clear the possibility that the ratio is not analytically smooth, however, by large number of degrees of freedom and the nature of Markov chain which generate random configurations of a distribution, it is expected to be applicable. Let's see.

[Fig. 44(a)] shows agreement in the standard deviations of the mean of the norm of magnetization per spin by two methods, one is the Jackknife and the other is binning analysis. This is simply due to the fact that the means are being estimated unbiasedly.

The bin size is varied and the best selected bin size is the one at which the deviation changes its behaviour to be unusual because by the virtue of correlation in adjacent measurements, as the degree of correlation is reduced, the variance will increase in magnitude.

For example, as mentioned, the unbiased estimator of the variance of an observable is given by (72). The interesting point with Jackknife resampling method is that, whatever the ratio in (72) coming with denominator of either  $n-1$ ,  $n$  or  $n+1$ , the Jackknife estimator will definitely return the unbiased one. Results for the estimation of the

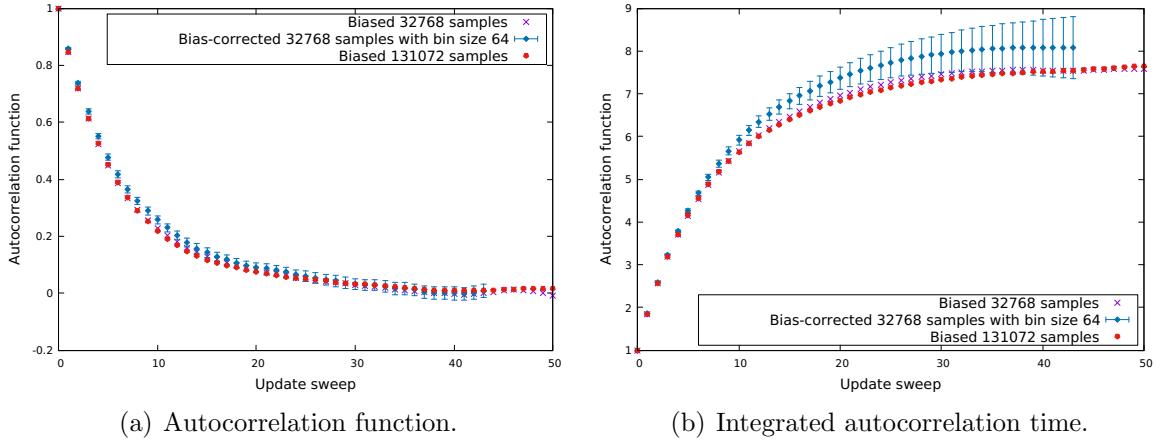


Figure 45:  $L = 16, T = 1.50(J/k_B)$  Norm of magnetization per spin.

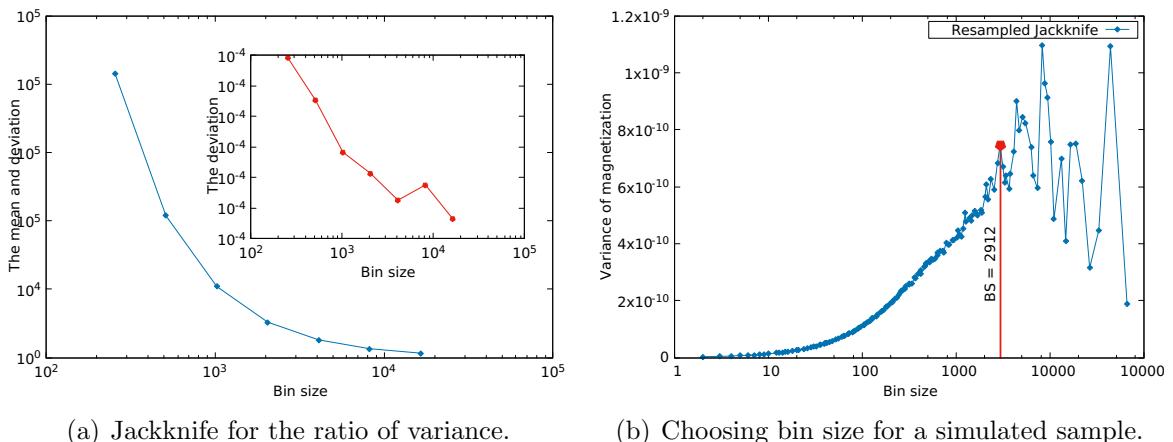


Figure 46:  $L = 16, T = 0.01(J/k_B)$ . Norm of magnetization per spin.

means agree between Jackknife and binning analysis since the original estimators are already unbiased. Let's note our failure in applying Jackknife for the estimation of a ratio, e.g. in the estimation of integrated time via the ratio of naive variance and binned variance, or in the estimation of autocorrelation function. The latter case is practically impossible, since, for example, at low temperatures, the autocorrelation

functions may require a large number of updates to reach zero or stop its initial decay. In [Fig. 45(b)], the integrated autocorrelation time is estimated by Jackknife to be

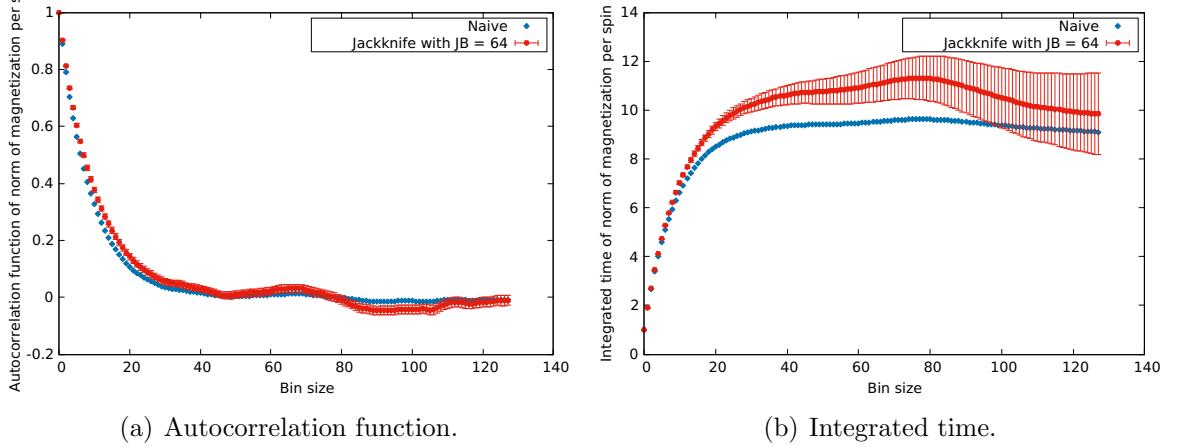


Figure 47:  $L = 32, T = 1.50(J/k_B)$ . Norm of magnetization per spin.

8.07(63) that is consistent with 7.59 estimated by the biased estimation of the ratio of variances. In [Fig. 45(a)], no significant difference shown for different number of update sweeps, however, the bias-corrected version is clearly different from the biased one.

The integrated autocorrelation time of norm of magnetization per spin simulated on GPGPU was found by by the estimation of the ratio of the variances to be 8.04 that is smaller than 11.5 on CPU with random Metropolis. Note that the former is for  $L = 32$ , while the latter is for  $L = 16$ . In [Fig. 47(a)], a comparison between the naive estimation of the autocorrelation function and the Jackknife version with jack bin 64 is shown, and [Fig. 47(b)] shows the corresponding integrated autocorrelation time for the same quantity, with the Jackknife version, the time is 10.7(4), this is higher than the previous estimation of about 8.04 and fairly smaller than 11.5 of random Metropolis, but the bias-corrected version for random Metropolis should yield some result that is larger than this 11.5. A final note is that  $JB = 64$  is just a suggestion that might not be the optimal one.

In [Fig. 46(a)], Jackknife is used to estimate the ratio of the variance of norm of magnetization per spin, the mean is on its way to converge, however, the deviation doesn't show our expectation related to correlation, i.e. less correlated means the variance or the deviation increase somehow. Note that there was 15 steps like [Fig. 46(b)] of choosing bin size including the first naive estimation over the original sample.

## References

- [1] P.D Loly. *Annals of Physics*, 56(1):40 – 57, 1970.
- [2] H. E. Stanley and T. A. Kaplan. *Phys. Rev. Lett.*, 17:913–915, Oct 1966.
- [3] J. M. Kosterlitz and D. J. Thouless. *J. Phys. C*, 5:L124–L126, 1972.
- [4] J. M. Kosterlitz and D. J. Thouless. *J. Phys. C*, 6:1181–1203, 1973.
- [5] A. J. Bray et al. *Phys. Rev. Lett.*, 84:1503–1506, Feb 2000.
- [6] Jan Tobochnik and G. V. Chester. *Phys. Rev. B*, 20:3761–3769, Nov 1979.
- [7] Kurt Binder and Erik Luijten. *Physics Reports*, 344(4–6):179 – 253, 2001.
- [8] J. Sethna. *Statistical Mechanics*. OUP Oxford, 2006.
- [9] Hagen Kleinert. *Gauge Fields in Condensed Matter: Disorder Fields and Applications to Superfluid Phase Transition and Crystal Melting*. World Scientific Publishing, 1990.
- [10] Igor Herbut. *A Modern Approach to Critical Phenomena*. CUP, 2007.
- [11] David J. Griffiths. *Introduction to Electrodynamics*. Addison-Wesley, 2012.
- [12] H. Gould, J. Tobochnik, and W. Christian. *An Introduction to Computer Simulation Methods: Applications to Physical Systems*. Pearson Addison Wesley, 2007.
- [13] Ulli Wolff. *Phys. Rev. Lett.*, 62:361–364, Jan 1989.
- [14] Martin Hasenbusch. *J. Phys. A*, 38(26):5869, 2005.
- [15] S. Miyashita et al. *Progress of Theoretical Physics*, 60(6):1669–1685, 1978.
- [16] K. Huang. *Introduction to Statistical Physics*. Taylor & Francis, 2001.
- [17] Ulli Wolff. *Nucl. Phys. B*, 322:759–774, August 1989.
- [18] P. C. W. Holdsworth S. T. Bramwell. *J. Phys: Cond. Mat.*, 5(4):L53, 1993.
- [19] S. Teitel and C. Jayaprakash. *Phys. Rev. B*, 27:598–601, Jan 1983.
- [20] Hans Weber and Petter Minnhagen. *Phys. Rev. B*, 37:5986–5989, Apr 1988.
- [21] L. D. Landau and E. M. Lifshitz. *Statistical Physics*. Pergamon Press, 1970.
- [22] M.H. DeGroot and M.J. Schervish. *Probability & Statistics*. Addison-Wesley, 2012.
- [23] C Grinstead and J. Snell. *Introduction to Probability*. AMS, 2012.
- [24] Matthew Richey. *The American Mathematical Monthly*, 117(5):pp. 383, 386, 2010.
- [25] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.
- [26] S. Chib and E. Greenberg. *The American Statistician*, 49(4):pp. 327–335, 1995.
- [27] J.-C. Walter and G.T. Barkema. *Physica A*, 418:78 – 87, 2015.

# Code snippets

Only important libraries listed. Online: <https://github.com/luantv>

## Wolff dynamics.

```
wolff.h
1  /*
2   * author: Julia
3   * Succeeded on 5th Nov. 2015.
4   */
5   void wolff(double * lattice){
6       int i, j, k;
7       double r;
8       memset(lattice_id, 0, idsize);
9       r = 2 * pi * mersenne(); // A random vector in [0, 2pi].
10      /* A random site */
11      j = (int)L * mersenne();
12      k = (int)L * mersenne();
13      /* Flip selected site (j,k). */
14      i = k + L * j;
15      cluster_growth(lattice, i, r, 1);
16  }
17
18  /*
19   * author: Julia
20   * Succeeded on December 4, 2015.
21   */
22  double cluster_size(double *lattice){
23      int i, j, k, hersize, mysize;
24      double r;
25      memset(lattice_id, 0, idsize);
26      k = 1;
27      // We are not playing with the original lattice at all. Do not disturb me.
28      for(i = 0; i < L * L; i++) latticex[i] = lattice[i];
29      /* Identify all clusters. */
30      for(i = 0; i < L * L; i++){
31          if(lattice_id[i] == 0){
32              r = 2 * pi * mersenne();
33              cluster_growth(latticex, i, r, k);
34              k++;
35          }
36      }
37
38      // Return the expectation of cluster size (in a single Wolff update step)
39      mysize = 0;
40      for(i = 1; i < k; i++){
41          hersize = 0;
42          for(j = 0; j < L * L; j++) if(lattice_id[j] == i) hersize ++;
```

```

43     mysize += hersize * hersize;
44 }
45     return mysize / (double)(L * L);
46 }
47 /*
48 author: Julia
49 */
50 void cluster_growth(double *lattice, int i, double r, int k){
51     int visitme, myc, current, myref, q, addme, index;
52     int neighbour[4] = {-1, 1, -L, L};
53     double prob, patrol;
54     lattice[i] = flip(r, lattice, i);
55     /* Grow a cluster of site (j,k). */
56     visitme = i;
57     myc = 0;
58     current = 0;
59     myref = i;
60     do{
61         if(lattice_id[visitme] == 0){
62             if(current == 0) lattice_id[i] = 1; // Mark the generating site.
63             prob = cos(lattice[myref] - r) * cos(r - lattice[visitme]);
64             prob = 1 - exp(2 * prob / temp);
65             patrol = mersenne();
66             if(visitme == i || patrol < prob){
67                 if(visitme != i) {
68                     lattice[visitme] = flip(r, lattice, visitme);
69                     lattice_id[visitme] = k; // Mark it.
70                 }
71                 else if(visitme == i) lattice_id[i] = k;
72                 for(q = 0; q < 4; q++){
73                     /* Periodic boundary condition */
74                     if(visitme % L == 0 && q == 0) addme = L - 1;// Left boundary, left neig
75                     else if(visitme % L == L - 1 && q == 1) addme = -(L - 1); // Right boundai
76                     else if(visitme < L && q == 2) addme = L * (L - 1); // Top boundary, upper
77                     else if(visitme >= L * (L - 1) && q == 3) addme = -L * (L - 1); // Bottom
78                     else addme = neighbour[q]; // Bulk neighbours
79                     index = visitme + addme;
80                     if(lattice_id[index] == 0){ // Add unmarked neighbours only.
81                         myadd[myc] = index;
82                         referee[myc] = visitme; // Who is my referee?
83                         myc++;
84                     }
85                 }
86             }
87         }
88         visitme = myadd[current];
89         myref = referee[current];
90         current++;
91     } while(current <= myc);
92 }
```

## Metropolis with CUDA API. The update kernel.

```
kernel.h
1 __global__ void global(state_t * state, double * d_in, double temp, int L){
2     int k, left, right, up, down;
3     int i = threadIdx.x + Kx * blockIdx.x;
4     int j = threadIdx.y + Ky * blockIdx.y;
5     int idx = i + L * j;
6     int tid = threadIdx.x + Kx * threadIdx.y;
7     double r, theta, pb;
8     __shared__ state_t stateblock[(WARPSIZE + WORDSHIFT + 1) * (Kx * Ky / WARPSIZE) - 1];
9     state_t l_state = state[idx];
10    set_neighbours(&left, &right, &up, &down, i, j, L);
11    for(k = 0; k <= 1; k++){
12        if((threadIdx.x + threadIdx.y) % 2 == k){
13            //generate a trial state
14            r = 2 * pi * prng_update(&l_state, tid, stateblock);
15            theta = flip(d_in, idx, r);
16            pb = de(d_in, idx, left, right, up, down, theta);
17            pb = exp(-pb / temp);
18            if(prng_update(&l_state, tid, stateblock) < pb) d_in[idx] = theta;
19        }
20        __syncthreads();
21    }
22    state[idx] = l_state;
23 }

25 __global__ void share(state_t * state, double * d_in, double temp, int L){
26     int k, left, right, up, down;
27     int i = threadIdx.x + Kx * blockIdx.x;
28     int j = threadIdx.y + Ky * blockIdx.y;
29     int idx = i + L * j;
30     int tid = threadIdx.x + Kx * threadIdx.y;
31     double r, theta, pb;
32     state_t l_state;
33     __shared__ state_t stateblock[(WARPSIZE + WORDSHIFT + 1) * (Kx * Ky / WARPSIZE) - 1];
34     __shared__ double tile[Kx * Ky];
35     tile[tid] = d_in[idx];
36     __syncthreads();
37     for(k = 0; k <= 1; k++){
38         if((threadIdx.x + threadIdx.y) % 2 == k){
39             //generate a trial state
40             r = 2 * pi * prng_update(&l_state, tid, stateblock);
41             theta = flip(d_in, idx, r);
42             if(threadIdx.x % (Kx - 1) == k || threadIdx.y % (Ky - 1) == k){
43                 set_neighbours(&left, &right, &up, &down, i, j, L);
44                 pb = de(d_in, idx, left, right, up, down, theta);
45             }
46             else{
47                 set_neighbours(&left, &right, &up, &down, threadIdx.x, threadIdx.y, Kx);
48                 pb = de(tile, tid, left, right, up, down, theta);
49             }
50             if(prng_update(&l_state, tid, stateblock) < pb) d_in[idx] = theta;
51         }
52         __syncthreads();
53     }
54     state[idx] = l_state;
55 }
```

```

49     }
50     pb = exp(-pb / temp);
51     if(prng_update(&l_state, tid, stateblock) < pb) d_in[idx] = tile[tid] = theta;
52   }
53   __syncthreads();
54 }
55 state[idx] = l_state;
56 }
57 __global__ void init(state_t * state, double * d_in, int L){
58   int i = threadIdx.x + Kx * blockIdx.x;
59   int j = threadIdx.y + Ky * blockIdx.y;
60   int idx = i + L * j;
61   int tid = threadIdx.x + Ky * threadIdx.y;
62   __shared__ state_t stateblock[(WARPSIZE + WORDSHIFT + 1) * (Kx * Ky / WARPSIZE +
63   d_in[idx] = pi; // 2 * pi * prng_update(&state[idx], tid, stateblock);
64 }
65 void algorithm(int CHOICE, state_t * d_state, double * d_in, double temp, int L, d_
66   if(CHOICE == 1) global<<<blocks, threads>>>(d_state, d_in, temp, L);
67   else if (CHOICE == 2) share<<<blocks, threads>>>(d_state, d_in, temp, L);
68 }
69 __device__ void set_neighbours(int *left, int *right, int *up, int *down, int i, i_
70   *left = (i - 1 + L) % L + L * j;
71   *right = (i + 1) % L + L * j;
72   *up = i + L * ((j - 1 + L) % L);
73   *down = i + L * ((j + 1) % L);
74 }
75
76 __device__ double flip(double *d_in, int idx, double r){
77   int rat;
78   double theta;
79   theta = 2 * r - d_in[idx] + pi;
80   rat = theta / (2 * pi);
81   if(theta < 0 && theta > - 2 * pi) theta = theta + 2 * pi;
82   else if(theta > 2 * pi || theta < - 2 * pi) theta = theta - rat * 2 * pi;
83   return theta;
84 }
85
86 __device__ double de(double *d_in, int idx, int left, int right, int up, int down,
87   double h, pb;
88   h = 0;
89   h += -cos(d_in[idx] - d_in[left]);
90   h += -cos(d_in[idx] - d_in[right]);
91   h += -cos(d_in[idx] - d_in[up]);
92   h += -cos(d_in[idx] - d_in[down]);
93   pb = 0;
94   pb += -cos(theta - d_in[left]);
95   pb += -cos(theta - d_in[right]);
96   pb += -cos(theta - d_in[up]);
97   pb += -cos(theta - d_in[down]);
98   pb -= h;
99   return pb;
100 }
```