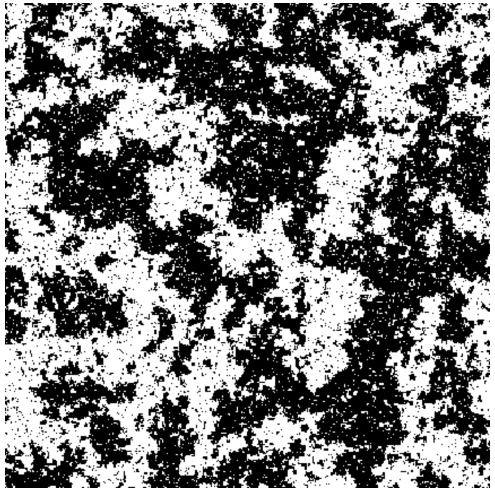
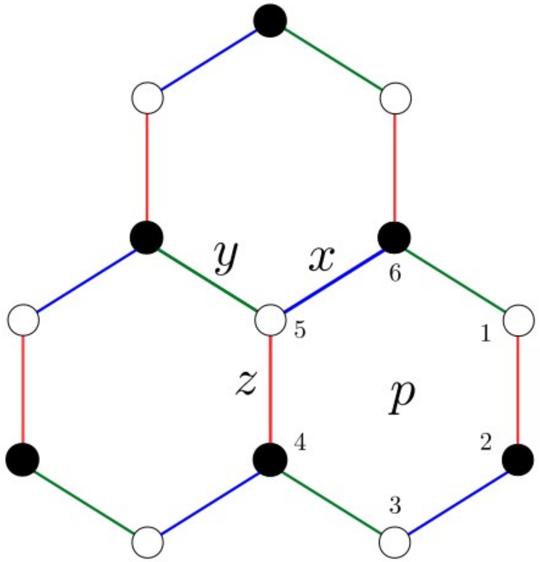
Convolutional restricted Boltzmann machine aided Monte Carlo

Phys. Rev. B 102, 195148



Ising



Kitaev

Daniel Alcalde

Content

- Introduction to Markov Chain Monte Carlo (MCMC)
- Restricted Boltzmann machines (RBM)
  - Problems
  - Convolutional RBM
- Examples:
  - Ising model
  - Honeycomb Kitaev model
- Summary

Li Huang and Lei Wang

Phys. Rev. B 95, 035105

Daniel Alcalde Puente and Ilya M. Eremin

Phys. Rev. B 102, 195148

Content

- Introduction to Markov Chain Monte Carlo (MCMC)
- Restricted Boltzmann machines (RBM)
  - Problems
  - Convolutional RBM
- Examples:
  - Ising model
  - Honeycomb Kitaev model
- Summary

Li Huang and Lei Wang

Phys. Rev. B 95, 035105

Daniel Alcalde Puente and Ilya M. Eremin

Phys. Rev. B 102, 195148

Introduction

Monte Carlo

$$\bar{f} = \sum_{x \in X} f(x)P(x) = \langle f(x) \rangle_{x \in P(x)}$$

$$= \frac{1}{M} \sum_{i=1}^{M} f(x^{(i)})$$

Example

$$P(s) = \frac{e^{-\beta E(s)}}{Z} \qquad s_i = \pm 1$$

$$E_{\text{Ising}}(s) = -\sum_{<i,j>} s_i s_j$$

- Objective:
  - Compute **expectation values** at different T
- How do we **sample** x from $P(x)$?
- Markov Chain $x^{(0)} \xrightarrow{P(x^{(1)}|x^{(0)})} x^{(1)} \xrightarrow{\cdots} x^{(n-1)} \xrightarrow{P(x^{(n)}|x^{(n-1)})} x^{(n)}$
- Detailed Balance $P_{\text{phys}}(x)P(x|x') = P(x'|x)P_{\text{phys}}(x')$
- There are many $P(x|x')$
- Metropolis is a choice $P(x|x') \rightarrow$ universal, slow (local update)

Monte Carlo

$$\bar{f} = \sum_{x \in X} f(x)P(x) = \langle f(x) \rangle_{x \in P(x)}$$

$$= \frac{1}{M} \sum_{i=1}^{M} f(x^{(i)})$$

Example

$$P(s) = \frac{e^{-\beta E(s)}}{Z}$$

$$s_i = \pm 1$$

$$E_{\text{Ising}}(s) = -\sum_{<i,j>} s_i s_j$$

Introduction

RBM

CRBM

Examples

Summary

- Objective:
  - Compute **expectation values** at different $\boldsymbol{T}$
- How do we **sample** x from $P(x)$?
- Markov Chain $x^{(0)} \xrightarrow{P(x^{(1)}|x^{(0)})} x^{(1)} \xrightarrow{\cdots} x^{(n-1)} \xrightarrow{P(x^{(n)}|x^{(n-1)})} x^{(n)}$
- Detailed Balance $P_{\text{phys}}(x)P(x|x') = P(x'|x)P_{\text{phys}}(x')$
- There are many $P(x|x')$
- Metropolis is a choice $P(x|x') \rightarrow$ universal, slow (local update)

Monte Carlo

$$\bar{f} = \sum_{x \in X} f(x)P(x) = \langle f(x) \rangle_{x \in P(x)}$$

$$= \frac{1}{M} \sum_{i=1}^{M} f(x^{(i)})$$

Example

$$P(s) = \frac{e^{-\beta E(s)}}{Z}$$

$$s_i = \pm 1$$

$$E_{\text{Ising}}(s) = -\sum_{<i,j>} s_i s_j$$

Introduction
RBM
CRBM
Examples
Summary

- Objective:
  - Compute **expectation values** at different $\boldsymbol{T}$
- How do we **sample** x from $P(x)$?
- Markov Chain $x^{(0)} \xrightarrow{P(x^{(1)}|x^{(0)})} x^{(1)} \xrightarrow{\cdots} x^{(n-1)} \xrightarrow{P(x^{(n)}|x^{(n-1)})} x^{(n)}$
- Detailed Balance $P_{\text{phys}}(x)P(x|x') = P(x'|x)P_{\text{phys}}(x')$
- There are many $P(x|x')$
- Metropolis is a choice $P(x|x') \rightarrow$ universal, slow (local update)

Monte Carlo

$$\bar{f} = \sum_{x \in X} f(x) P(x) = \langle f(x) \rangle_{x \in P(x)}$$

$$= \frac{1}{M} \sum_{i=1}^{M} f(x^{(i)})$$

Example

$$P(s) = \frac{e^{-\beta E(s)}}{Z} \qquad s_i = \pm 1$$

$$E_{\text{Ising}}(s) = - \sum_{<i,j>} s_i s_j$$

- Objective:
  - Compute **expectation values** at different $\boldsymbol{T}$
- How do we **sample** x from $P(x)$?
- Markov Chain $x^{(0)} \xrightarrow{P(x^{(1)}|x^{(0)})} x^{(1)} \xrightarrow{\cdots} x^{(n-1)} \xrightarrow{P(x^{(n)}|x^{(n-1)})} x^{(n)}$
- Detailed Balance $P_{\text{phys}}(x) P(x|x') = P(x'|x) P_{\text{phys}}(x')$
- There are many $P(x|x')$
- Metropolis is a choice $P(x|x') \rightarrow$ universal, slow (local update)

Monte Carlo

$$\bar{f} = \sum_{x \in X} f(x)P(x) = \langle f(x) \rangle_{x \in P(x)}$$

$$= \frac{1}{M} \sum_{i=1}^{M} f(x^{(i)})$$

Example

$$P(s) = \frac{e^{-\beta E(s)}}{Z}$$

$$s_i = \pm 1$$

$$E_{\text{Ising}}(s) = - \sum_{<i,j>} s_i s_j$$

Introduction
RBM
CRBM
Examples
Summary

- Objective:
  - Compute **expectation values** at different $\boldsymbol{T}$
- How do we **sample** x from $P(x)$?
- Markov Chain $x^{(0)} \xrightarrow{P(x^{(1)}|x^{(0)})} x^{(1)} \xrightarrow{\cdots} x^{(n-1)} \xrightarrow{P(x^{(n)}|x^{(n-1)})} x^{(n)}$
- Detailed Balance $P_{\text{phys}}(x)P(x|x') = P(x'|x)P_{\text{phys}}(x')$
- There are many $P(x|x')$
- Metropolis is a choice $P(x|x')$→universal, slow (local update)

Monte Carlo

$$\bar{f} = \sum_{x \in X} f(x) P(x) = \langle f(x) \rangle_{x \in P(x)}$$

$$= \frac{1}{M} \sum_{i=1}^{M} f(x^{(i)})$$

Example

$$P(s) = \frac{e^{-\beta E(s)}}{Z} \qquad s_i = \pm 1$$

$$E_{\text{Ising}}(s) = - \sum_{<i,j>} s_i s_j$$

Introduction
RBM
CRBM
Examples
Summary

- Objective:
  - Compute **expectation values** at different $\boldsymbol{T}$

- How do we **sample** x from $P(x)$?

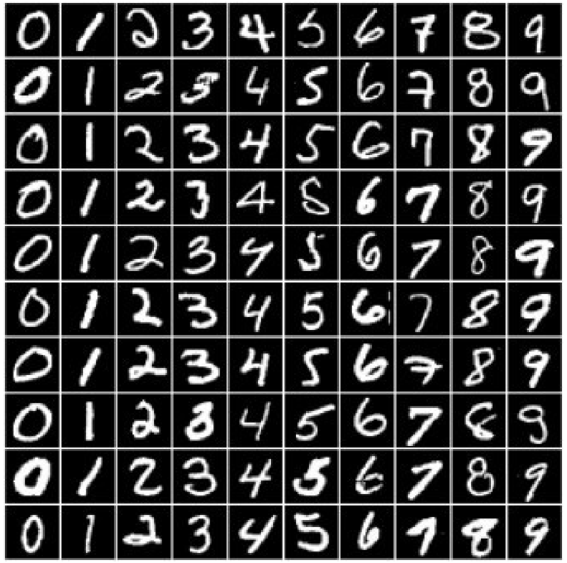
- Markov Chain $x^{(0)} \xrightarrow{P(x^{(1)}|x^{(0)})} x^{(1)} \xrightarrow{\cdots} x^{(n-1)} \xrightarrow{P(x^{(n)}|x^{(n-1)})} x^{(n)}$
- Detailed Balance $P_{\text{phys}}(x) P(x|x') = P(x'|x) P_{\text{phys}}(x')$
- There are many $P(x|x')$

- Metropolis is a choice $P(x|x') \rightarrow$ universal, slow (local update)

Restricted Boltzmann Machine

RBM

- Machine learning model → Learn probability distributions



$P_{\text{RBM}}(x; W) \longleftrightarrow P_{\text{RBM}}(x|x; W)$

faster than Metropolis?

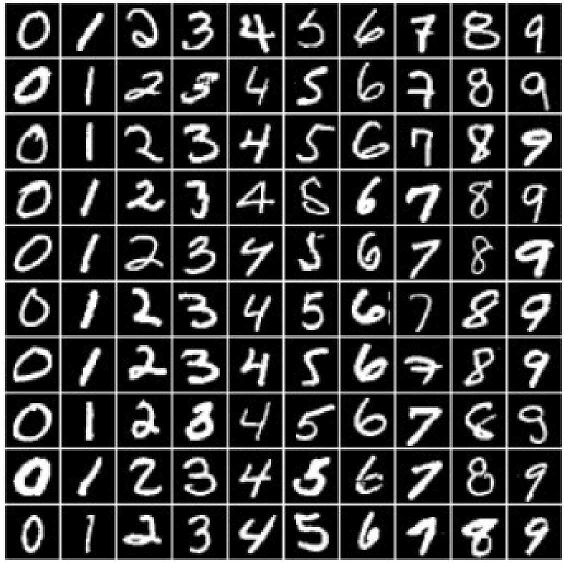
- Different matrix W different $P(x)$

- Objective:
  - Find $W \rightarrow P_{RBM}(x; W) \simeq P_{\text{phys}}(x; T) \quad \forall x$ (training)
  - Sample from $P_{\text{RBM}}(x|x; W)$

Introduction
RBM
CRBM
Examples
Summary

RBM

- Machine learning model $\rightarrow$ Learn probability distributions



$$P_{\mathrm{RBM}}(x;W) \longleftrightarrow P_{\mathrm{RBM}}(x|x;W)$$

faster than Metropolis?

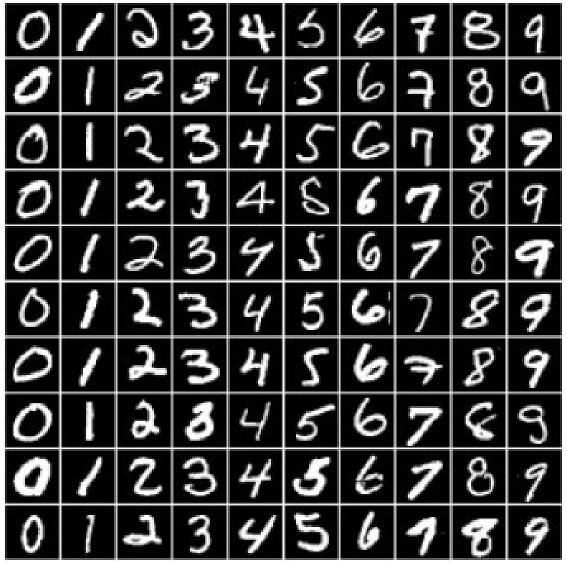
- Different matrix W different $P(x)$

- Objective:
  - Find $W \rightarrow P_{RBM}(x;W) \simeq P_{\mathrm{phys}}(x;T) \quad \forall x$ (training)
  - Sample from $P_{\mathrm{RBM}}(x|x;W)$

Introduction
RBM
CRBM
Examples
Summary

RBM

- Machine learning model $\rightarrow$ Learn probability distributions



$$P_{\text{RBM}}(x;W) \longleftrightarrow P_{\text{RBM}}(x|x;W)$$

faster than Metropolis?

- Different matrix W different $P(x)$

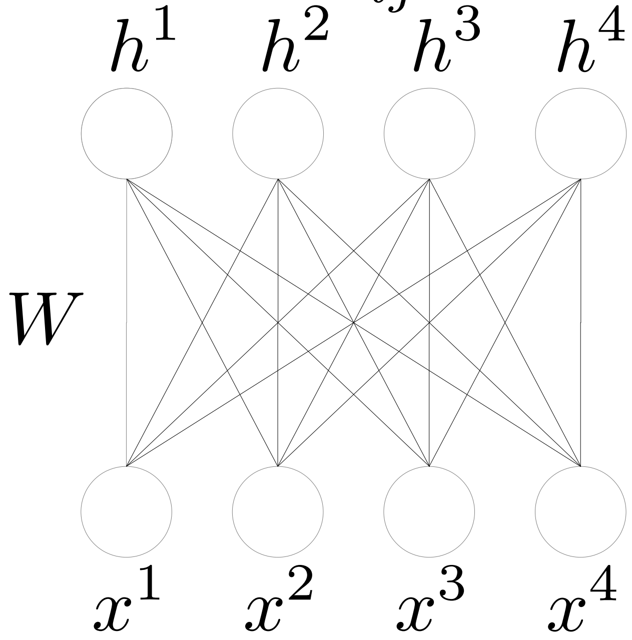
- Objective:
  - Find $W \rightarrow\ P_{RBM}(x;W) \simeq P_{\text{phys}}(x;T) \quad \forall x$ (training)
  - Sample from $P_{\text{RBM}}(x|x;W)$

RBM

- Hidden statistical variables h_i
- h is used for sampling

$$P_{\mathrm{RBM}}(x,h;W) = \frac{e^{-E(x,h;W)}}{Z}$$

$$E(x,h) = -\sum_{ij} h^i W_{ij} x^j$$



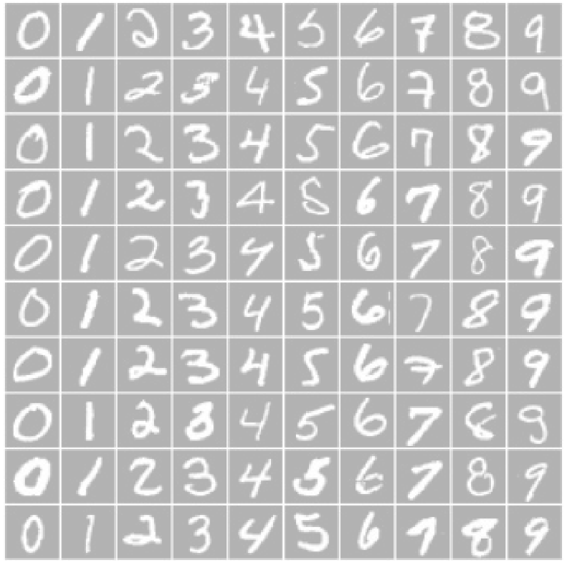
$$P_{\mathrm{RBM}}(x;W) = \sum_{h_i\in\{0,1\}} \frac{e^{-E(x,h;W)}}{Z} = \frac{e^{-F(x;W)}}{Z}$$

$$F(x;W) = -\sum_i \log\left(1 + e^{\sum_j W_{ij}x_j}\right)$$

RBM

Introduction
RBM
CRBM
Examples
Summary

- Machine learning model $\rightarrow$ Learn probability distributions



$P_{\text{RBM}}(x;W) \longleftrightarrow P_{\text{RBM}}(x|x;W)$
Global Update (faster)

- Different matrix W different $P(x)$

- Objective:
  - Find $W \rightarrow P_{RBM}(x;W) \simeq P_{\text{phys}}(x;T) \quad \forall x$ (training)
  - Sample from $P_{\text{RBM}}(x|x;W)$

RBM Training

Introduction

RBM

CRBM

Examples

Summary

- How to train RBM? How to choose W ?
- Measure → How close is P_{RBM} to P_{phys}?

$$\Delta E(x^{(i)};W) = E_{\text{phys}}(x^{(i)})\beta - F_{\text{RBM}}(x^{(i)};W)$$

$$\text{loss}(W) = \frac{1}{M}\sum_{i=1}^{M}\Delta E(x^{(i)};W)^2$$

- If $\text{loss}(W) = 0$ ⟶ $P_{\text{RBM}}(x) = \frac{e^{-F_{\text{RBM}}(x)}}{Z_R} = \frac{e^{-\beta E(x)}}{Z} = P_{\text{phys}}(x)$

Equivalent to minimizing

$$D_{\text{KL}}(P_{\text{RBM}}||P_{\text{phys}})$$

RBM Training

- How to train RBM? How to choose W ?
- Measure $\rightarrow$ How close is P_{RBM} to P_{phys}?

$$\Delta E(x^{(i)}; W) = E_{\text{phys}}(x^{(i)})\beta - F_{\text{RBM}}(x^{(i)}; W)$$

$$\text{loss}(W) = \frac{1}{M}\sum_{i=1}^{M} \Delta E(x^{(i)}; W)^2$$

- If $\text{loss}(W) = 0$ → $P_{\text{RBM}}(x) = \frac{e^{-F_{\text{RBM}}(x)}}{Z_R} = \frac{e^{-\beta E(x)}}{Z} = P_{\text{phys}}(x)$

Equivalent to minimizing

$$D_{\text{KL}}(P_{\text{RBM}} || P_{\text{phys}})$$

RBM Training

- How to train RBM? How to choose W ?
- Measure → How close is P_{RBM} to P_{phys}?

$$\Delta E(x^{(i)}; W) = E_{\text{phys}}(x^{(i)})\beta - F_{\text{RBM}}(x^{(i)}; W)$$

$$\text{loss}(W) = \frac{1}{M}\sum_{i=1}^{M} \Delta E(x^{(i)}; W)^2$$

- If $\text{loss}(W) = 0$ ⇒

$$P_{\text{RBM}}(x) = \frac{e^{-F_{\text{RBM}}(x)}}{Z_R} = \frac{e^{-\beta E(x)}}{Z} = P_{\text{phys}}(x)$$

Equivalent to minimizing

$D_{\text{KL}}(P_{\text{RBM}} || P_{\text{phys}})$

RBM Training

Introduction

RBM

CRBM

Examples

Summary

- How to train RBM? How to choose W ?
- Measure $\rightarrow$ How close is P_{RBM} to P_{phys}?

$$\Delta E(x^{(i)};W) = E_{\text{phys}}(x^{(i)})\beta - F_{\text{RBM}}(x^{(i)};W)$$

$$\text{loss}(W) = \frac{1}{M}\sum_{i=1}^{M}\Delta E(x^{(i)};W)^2$$

- If $\text{loss}(W) = 0$ ⟶

$$P_{\text{RBM}}(x) = \frac{e^{-F_{\text{RBM}}(x)}}{Z_R} = \frac{e^{-\beta E(x)}}{Z} = P_{\text{phys}}(x)$$

Equivalent to minimizing

$$D_{\text{KL}}(P_{\text{RBM}}||P_{\text{phys}})$$

RBM Training

Introduction
RBM
CRBM
Examples
Summary

$$\text{loss}(W) = \frac{1}{M}\sum_{i=1}^{M} \Delta E(x^{(i)}; W)^2$$

- Optimize $W \rightarrow$ Gradient descent: $W_{ij} \rightarrow W_{ij} - \lambda_r \frac{\delta \text{loss}(W)}{\delta W_{ij}}$

- Which states $x^{(i)}$ do we train with?
  - States sampled from P_{phys}

$$\text{sample } P_{\text{phys}} \rightarrow \text{train } P_{RBM} \rightarrow \text{sample } P_{\text{phys}}$$

Problem!!!

RBM Training

Introduction
RBM
CRBM
Examples
Summary

$$\text{loss}(W) = \frac{1}{M}\sum_{i=1}^{M}\Delta E(x^{(i)};W)^2$$

- Optimize $W \rightarrow$ Gradient descent: $W_{ij} \rightarrow W_{ij} - \lambda_r \frac{\delta \text{loss}(W)}{\delta W_{ij}}$

- Which states $x^{(i)}$ do we train with?
  - States sampled from P_{phys}

$$\text{sample } P_{\text{phys}} \rightarrow \text{train } P_{RBM} \rightarrow \text{sample } P_{\text{phys}}$$

Problem!!!

RBM Training

$$\text{loss}(W) = \frac{1}{M}\sum_{i=1}^{M} \Delta E(x^{(i)}; W)^2$$

- Optimize $W \rightarrow$ Gradient descent: $W_{ij} \rightarrow W_{ij} - \lambda_r \frac{\delta \text{loss}(W)}{\delta W_{ij}}$

- Which states $x^{(i)}$ do we train with?
  - States sampled from P_{phys}

$$\text{sample}\, P_{\text{phys}} \rightarrow \text{train}\, P_{RBM} \rightarrow \text{sample}\, P_{\text{phys}}$$

Problem!!!

Problems RBM

Introduction
RBM
CRBM
Examples
Summary

- Problems:
  - Need to do Metropolis (expensive)
  - More particles → matrix W is larger → longer training
- Solution:
  - Symmetries!!
  - Translational invariance (Convolutional RBM)
  - Number of parameters $\mathcal{O}(L^4)$ → not L dependent
  - CRBM → fully convolutional → same CRBM for different L
  - Train small L → Sample large L

sample P_{phys} with Metropolis at L_{small} → train P_{CRBM} at L_{small} → sample P_{phys} at L_{large}

Problems RBM

Introduction
RBM
CRBM
Examples
Summary

- Problems:
  - Need to do Metropolis (expensive)
  - More particles → matrix W is larger → longer training
- Solution:
  - Symmetries!!
  - Translational invariance (Convolutional RBM)
  - Number of parameters $\mathcal{O}(L^4) \rightarrow$ not L dependent
  - CRBM → fully convolutional → same CRBM for different L
  - Train small L → Sample large L

sample P_{phys} with Metropolis at $L_{\text{small}} \rightarrow$ train P_{CRBM} at $L_{\text{small}} \rightarrow$ sample P_{phys} at L_{large}

Problems RBM

Introduction
RBM
CRBM
Examples
Summary

- Problems:
  - Need to do Metropolis (expensive)
  - More particles $\rightarrow$ matrix W is larger $\rightarrow$ longer training

- Solution:
  - Symmetries!!
  - Translational invariance (Convolutional RBM)
  - Number of parameters $\mathcal{O}(L^4) \rightarrow$ not L dependent
  - CRBM $\rightarrow$ fully convolutional $\rightarrow$ same CRBM for different L
  - Train small L $\rightarrow$ Sample large L

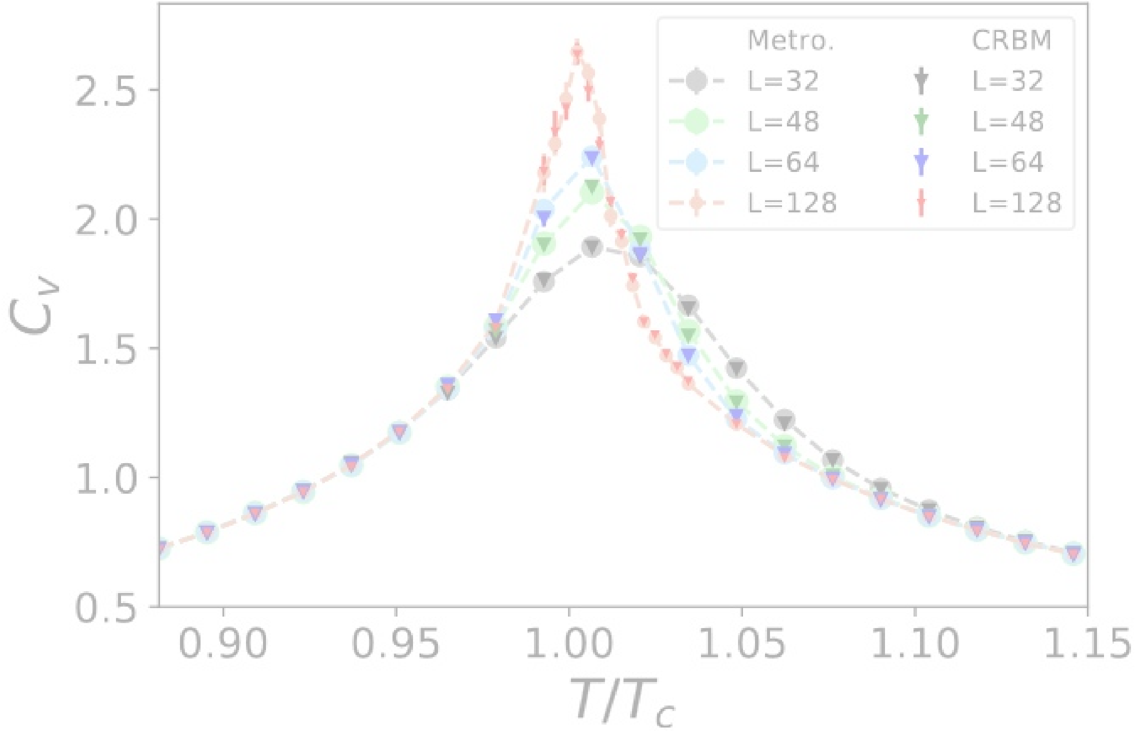
sample P_{phys} with Metropolis at $L_{\text{small}} \rightarrow$ train P_{CRBM} at $L_{\text{small}} \rightarrow$ sample P_{phys} at L_{large}

Examples

2D Ising Model

- Proof of concept
- Nearest neighbour interaction $\rightarrow$ 2x2 kernel (11 parameters vs $\mathcal{O}(L^4)$)
- Sufficient to train at L=3

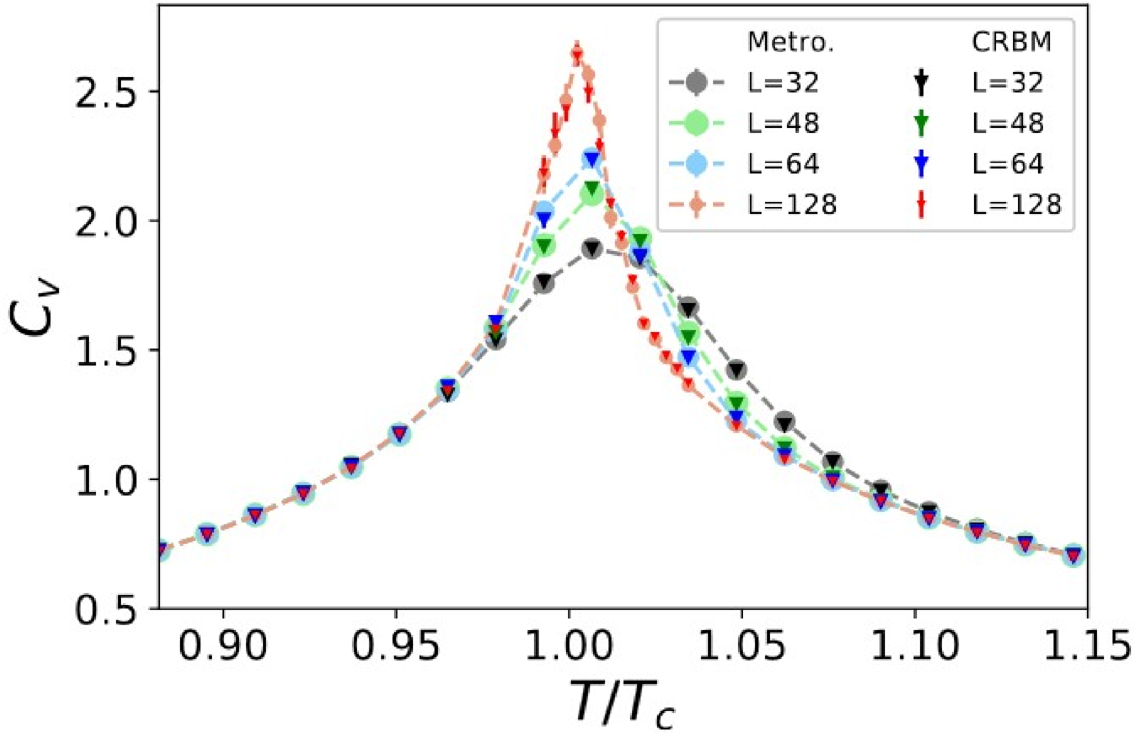
Introduction
RBM
CRBM
Examples
Summary



2D Ising Model

- Proof of concept
- Nearest neighbour interaction → 2x2 kernel (11 parameters vs $\mathcal{O}(L^4)$)
- Sufficient to train at L=3

Introduction
RBM
CRBM
Examples
Summary



Kitaev Model

$$\mathcal{H} = -J_x \sum_{\langle i,j \rangle_x} \sigma_i^x \sigma_j^x - J_y \sum_{\langle i,j \rangle_y} \sigma_i^y \sigma_j^y - J_z \sum_{\langle i,j \rangle_z} \sigma_i^z \sigma_j^z$$

Introduction

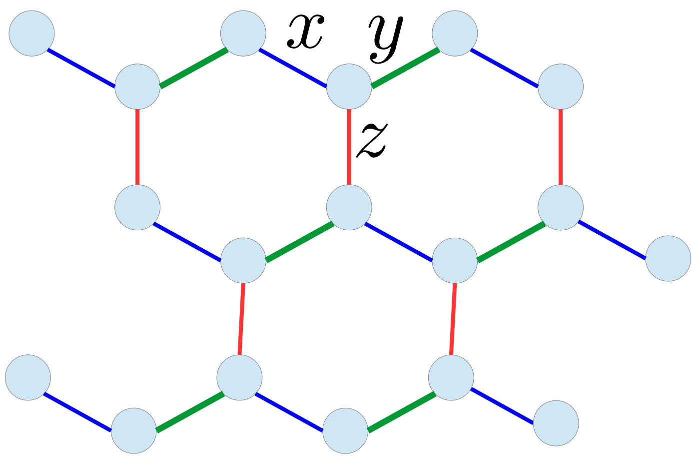
RBM

CRBM

Examples

Summary

- Quantum spin liquid
- Analytical ground state



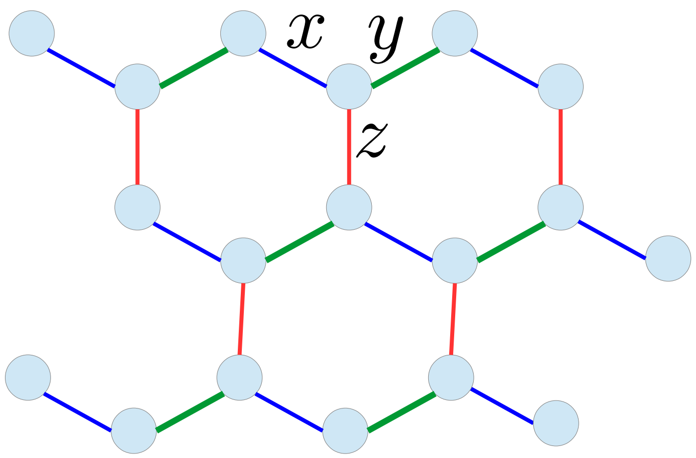
- Fractionalized Quasiparticles:
  - Z_2 gauge fields $u_{ij} = \pm 1$
  - Itinerant Majorana fermions

- Monte Carlo on the gauge field $u_{ij} = \pm 1$
- Each MC step $\rightarrow$ Trace out itinerant Majorana fermions
- Solve eigenvalue problem $O(L^6) \rightarrow$ Only small simulations L<20

Kitaev Model

$$\mathcal{H} = -J_x \sum_{\langle i,j \rangle_x} \sigma_i^x \sigma_j^x - J_y \sum_{\langle i,j \rangle_y} \sigma_i^y \sigma_j^y - J_z \sum_{\langle i,j \rangle_z} \sigma_i^z \sigma_j^z$$

- Quantum spin liquid
- Analytical ground state



- Fractionalized Quasiparticles:
  - Z_2 gauge fields $u_{ij} = \pm 1$
  - Itinerant Majorana fermions

- Monte Carlo on the gauge field $u_{ij} = \pm 1$
- Each MC step→ Trace out itinerant Majorana fermions
- Solve eigenvalue problem $O(L^6)$→ Only small simulations L<20

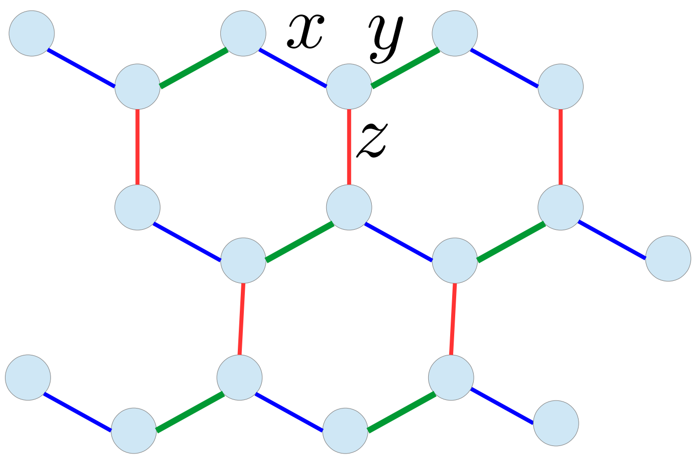
Kitaev Model

$$\mathcal{H} = iJ_x \sum_{x \text{ bonds}} c_w c_b - iJ_y \sum_{y \text{ bonds}} c_b c_w - iJ_z \sum_{z \text{ bonds}} u c_b c_w$$

Introduction
RBM
CRBM
Examples
Summary

- Quantum spin liquid
- Analytical ground state

- Fractionalized Quasiparticles:
  - Z_2 gauge fields $u_{ij} = \pm 1$
  - Itinerant Majorana fermions



- Monte Carlo on the gauge field $u_{ij} = \pm 1$
- Each MC step → Trace out itinerant Majorana fermions
- Solve eigenvalue problem $O(L^6)$ → Only small simulations L<20

Kitaev Model

$$\mathcal{H} = iJ_x \sum_{x \text{ bonds}} c_w c_b - iJ_y \sum_{y \text{ bonds}} c_b c_w - iJ_z \sum_{z \text{ bonds}} u c_b c_w$$

Introduction

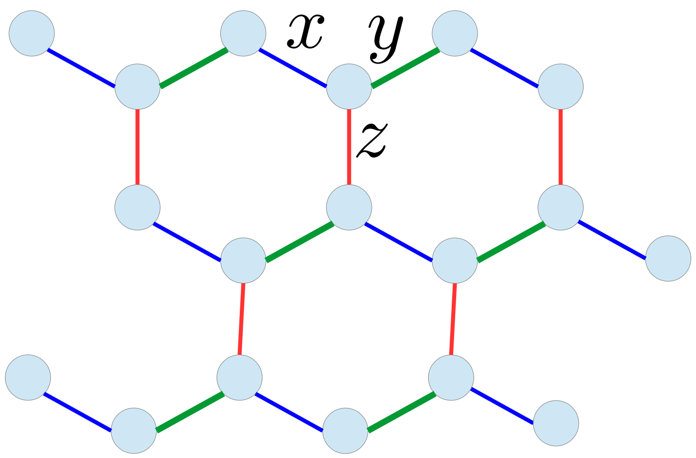
RBM

CRBM

Examples

Summary

- Quantum spin liquid
- Analytical ground state

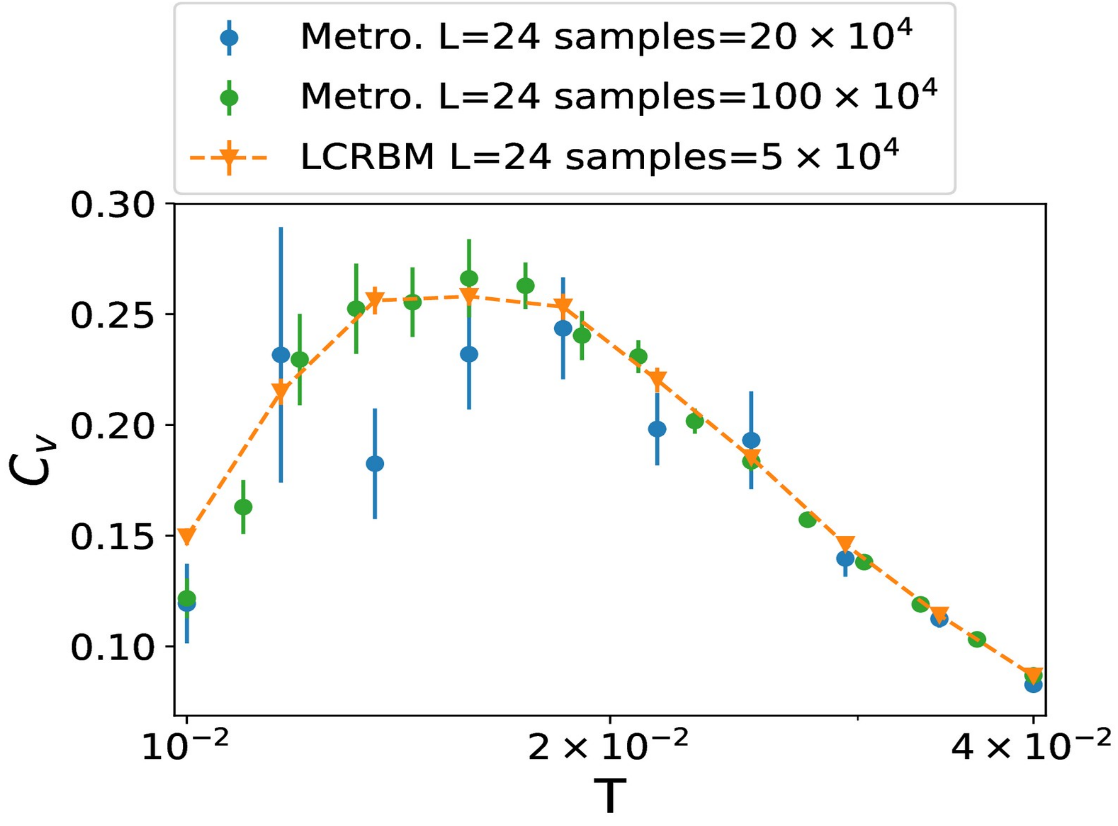


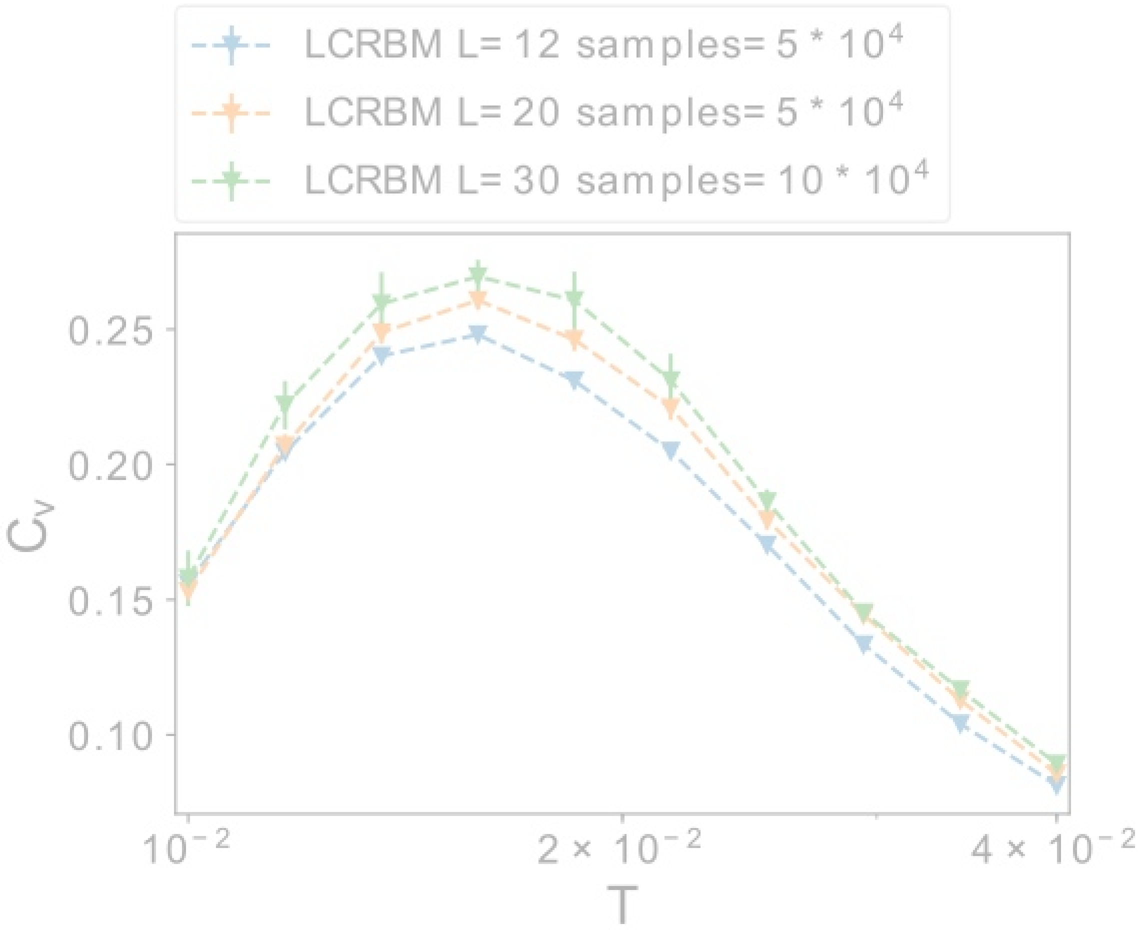
- Fractionalized Quasiparticles:
  - Z_2 gauge fields $u_{ij} = \pm 1$
  - Itinerant Majorana fermions

- Monte Carlo on the gauge field $u_{ij} = \pm 1$
- Each MC step → Trace out itinerant Majorana fermions
- Solve eigenvalue problem $O(L^6)$ → Only small simulations L<20

Introduction
RBM
CRBM
Examples
Summary

Kitaev Model

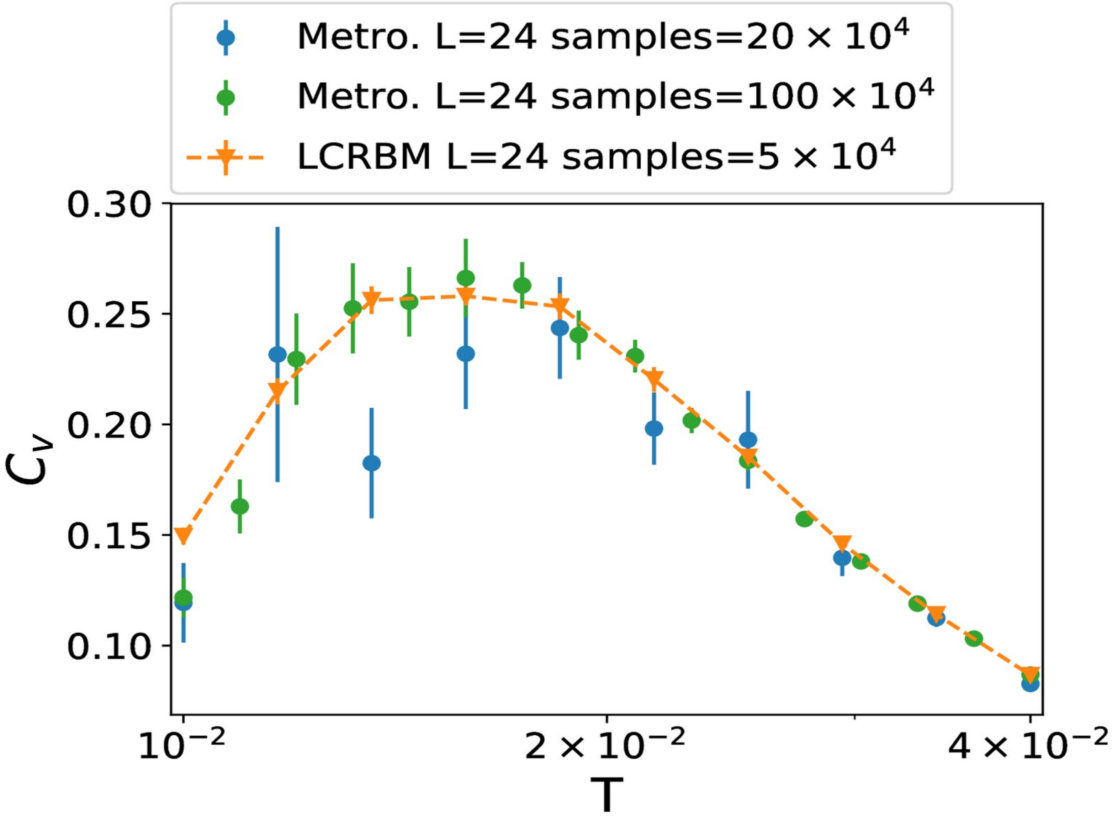


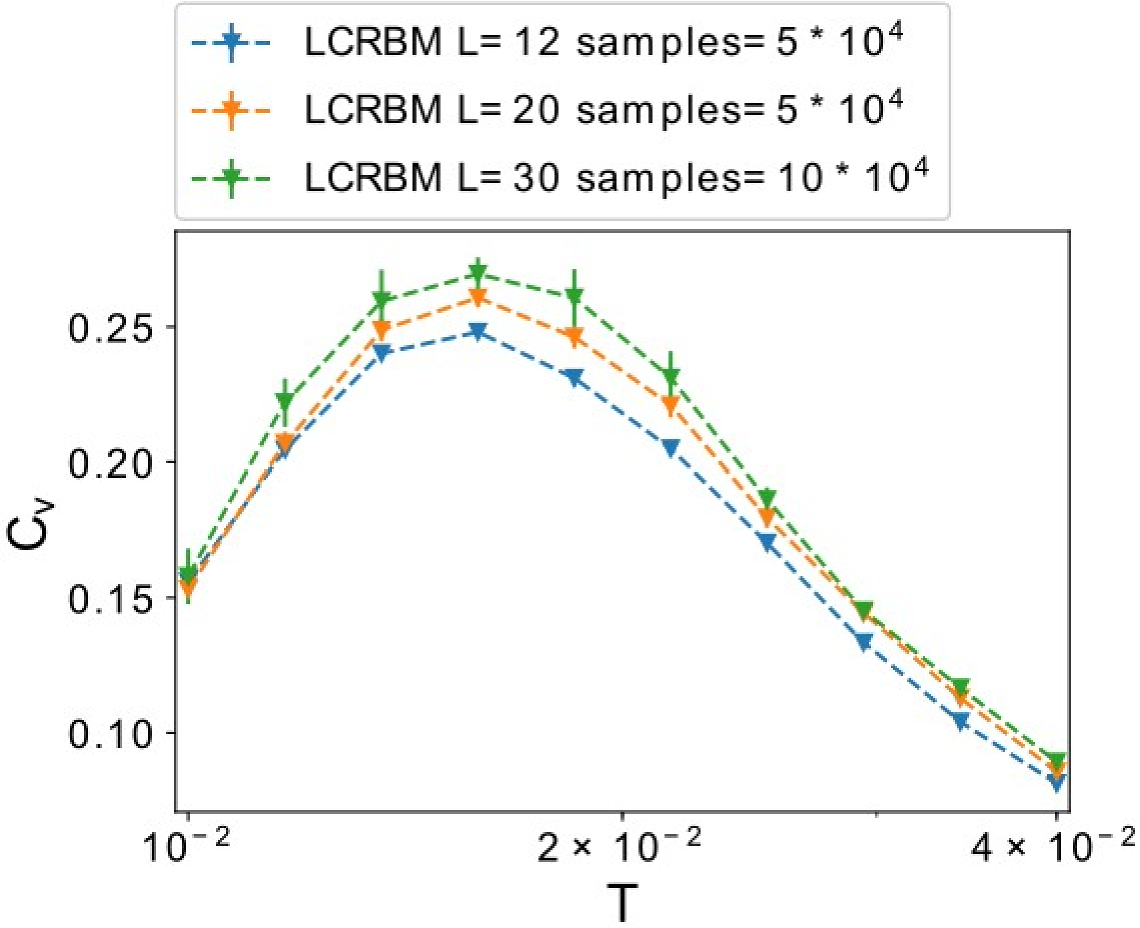


- Metropolis (1 week)
- CRBM (8 hours)

- Cv converges
- Crossover (no phase transition)

Kitaev Model





- Metropolis (1 week)
- CRBM (8 hours)

- Cv converges
- Crossover (no phase transition)

Kitaev Model - Extended

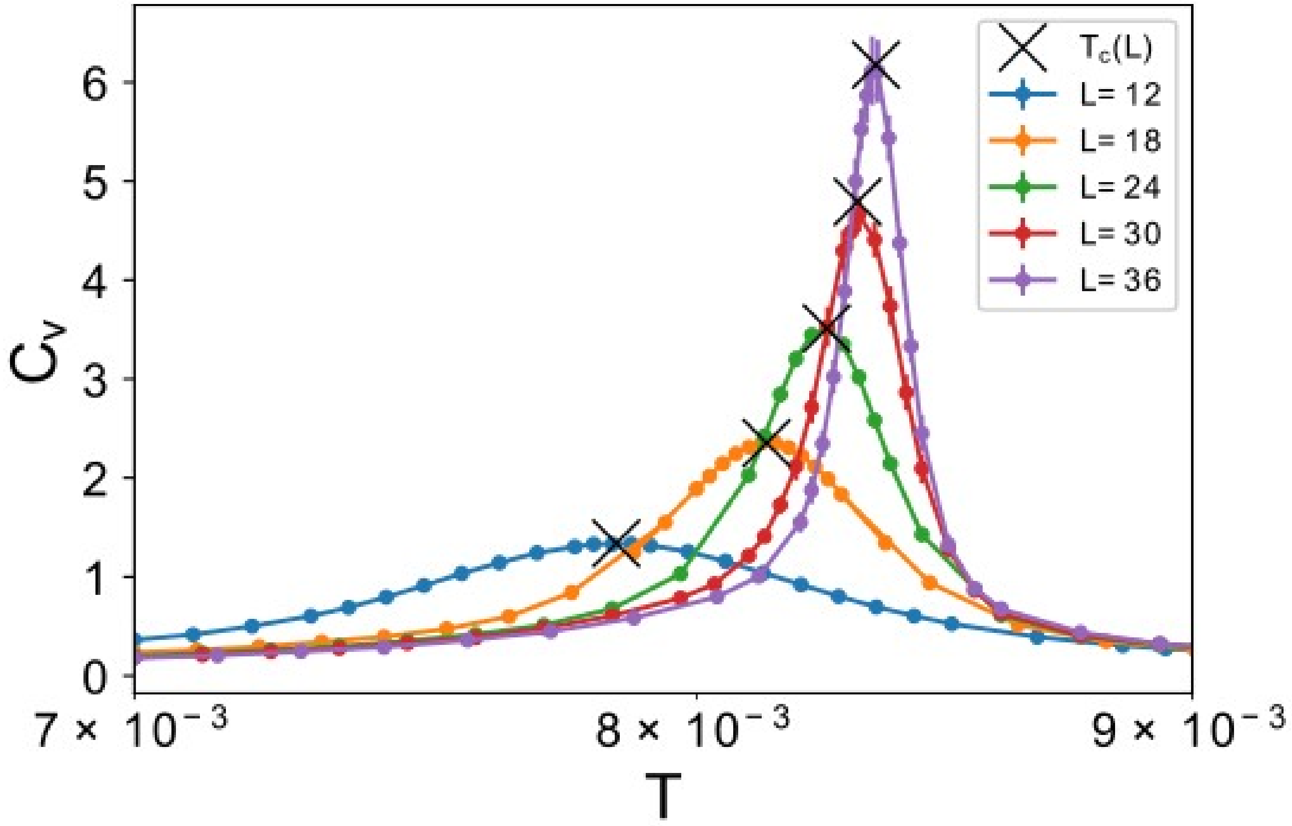
Introduction

RBM

CRBM

Examples

Summary



Critical exponents

$$T_c = (8.458 \pm 0.015) \times 10^{-3}$$

$$\nu = 0.646 \pm 0.035$$

$$\alpha = 0.90 \pm 0.05$$

- Phase transition
- Previous work $\rightarrow$ no critical exponents

Summary

- Sampling $P(x) \rightarrow$ difficult
- Method $\rightarrow$ Markov chain
- Objective:
  - Markov Chain $\rightarrow$ moves quickly

- RBM has a $P(x;W)$ and a Markov Chain $P(x'|x;W)$
  - Learn $P_{RBM}(x;W) \simeq P_{\text{phys}}(x)$
  - Sample from RBM

- Translational invariance $\rightarrow$ CRBM
- Train at small $L \rightarrow$ Sample at large L

- Ising model $\rightarrow$ Proof of concept
- Kitaev model $\rightarrow$ MC larger lattices

Introduction
RBM
CRBM
Examples
Summary

Summary

- Sampling $P(x) \rightarrow$ difficult
- Method $\rightarrow$ Markov chain
- Objective:
  - Markov Chain $\rightarrow$ moves quickly

- RBM has a $P(x;W)$ and a Markov Chain $P(x'|x;W)$
  - Learn $P_{RBM}(x;W) \simeq P_{\text{phys}}(x)$
  - Sample from RBM

- Translational invariance $\rightarrow$ CRBM
- Train at small $L \rightarrow$ Sample at large L

- Ising model $\rightarrow$ Proof of concept
- Kitaev model $\rightarrow$ MC larger lattices

Summary

- Sampling $P(x) \rightarrow$ difficult
- Method $\rightarrow$ Markov chain
- Objective:
  - Markov Chain $\rightarrow$ moves quickly

- RBM has a $P(x; W)$ and a Markov Chain $P(x'|x; W)$
  - Learn $P_{RBM}(x; W) \simeq P_{\text{phys}}(x)$
  - Sample from RBM

- Translational invariance $\rightarrow$ CRBM
- Train at small L $\rightarrow$ Sample at large L

- Ising model $\rightarrow$ Proof of concept
- Kitaev model $\rightarrow$ MC larger lattices

Summary

- Sampling $P(x) \to$ difficult
- Method $\to$ Markov chain
- Objective:
  - Markov Chain $\to$ moves quickly

- RBM has a $P(x;W)$ and a Markov Chain $P(x'|x;W)$
  - Learn $P_{RBM}(x;W) \simeq P_{\mathrm{phys}}(x)$
  - Sample from RBM

- Translational invariance $\to$ CRBM
- Train at small $L \to$ Sample at large L

- Ising model $\to$ Proof of concept
- Kitaev model $\to$ MC larger lattices