Towards reduction of autocorrelation in HMC by machine learning

CCNU→ 理研(和光)→ 理研(ブルックヘブン)
富谷 昭夫

共同研究者: 田中章詞 (理化学研究所 革新知能統合研究センター)

Based on arXiv:1712.03893

akio.tomiya@mail.ccnu.edu.cn

自己紹介

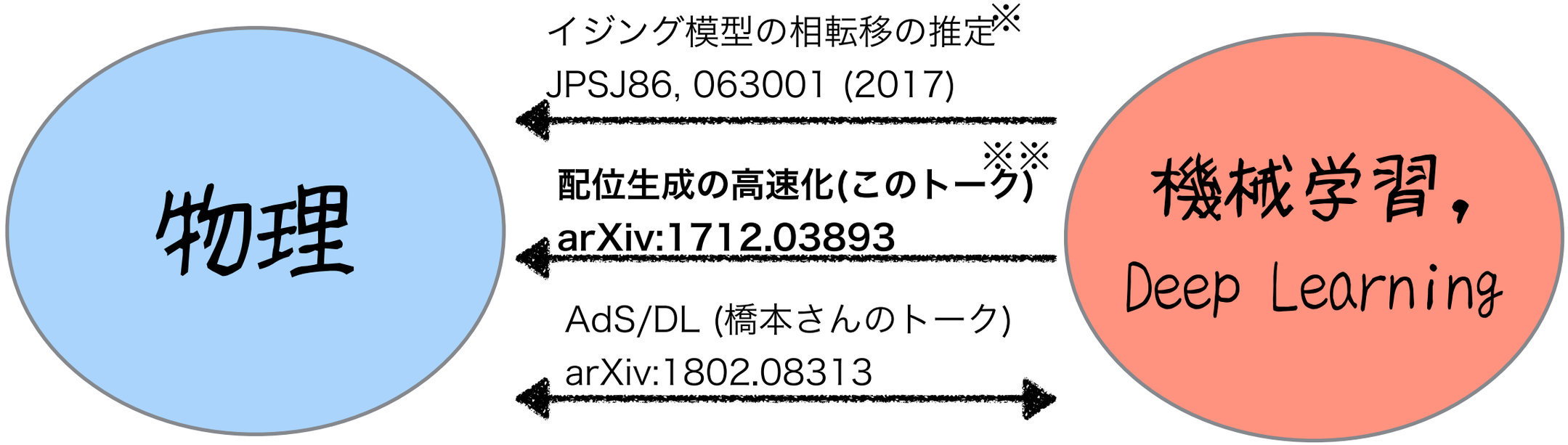
前回は機械学習を用いた測定、今回は機械学習を用いた配位生成

CCNU(中国の真ん中らへん、武漢) のポスドク

2015年 大阪大学 素粒子論研究室でPhD(田中くんと同期)。

専門: 格子QCD(数値計算で原子核などの性質を調べる)

相転移が好きで超伝導→テクニカラー→有限温度QCD、エンタングルメントエントロピー等…



関連するトーク

※柴さん、青木さん、大槻さんのトーク

※※永井さんのトーク

まとめ：問題設定と結果

提案したアルゴリズムで自己相関は短縮。相転移点で不具合

問題: 格子QCDでの配位生成を効率化したい！
効率化できれば、モンテカルロ法(HMC)からくる統計誤差を短時間で減らすことが出来る。

手法: 機械学習をHMCに組み込む
機械学習は、それっぽい画像を作れる
　→ 応用できないか？
性質が良く分かってる3次元φ4理論で試してみた

Radford et al. (2015)

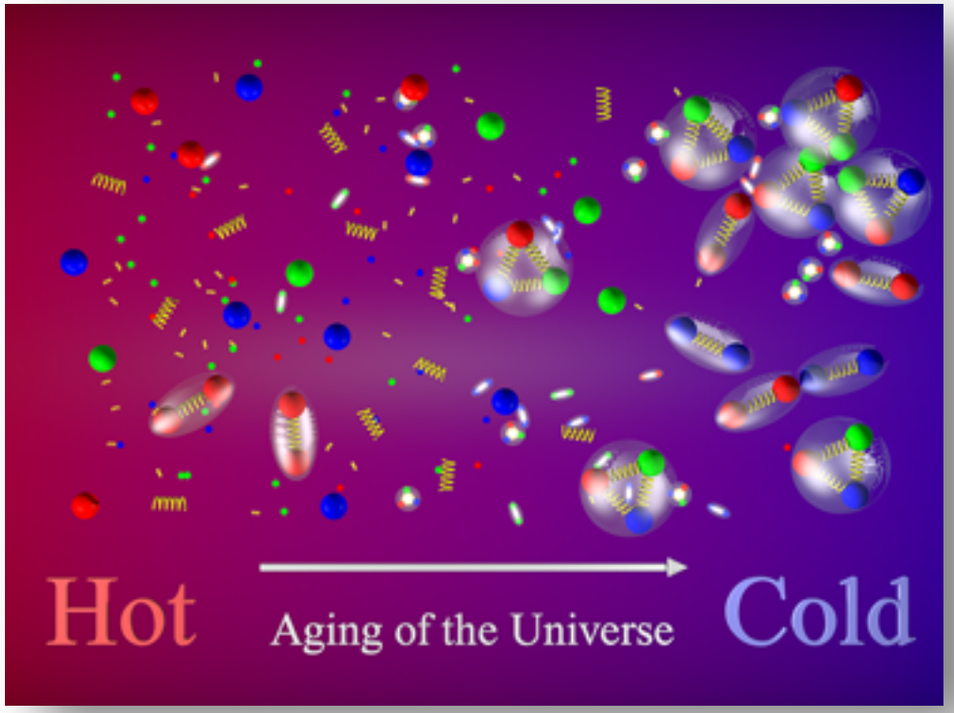
結果: 対称相、破れ相で、1点関数, 2点関数は無矛盾
臨界点で、不具合がある

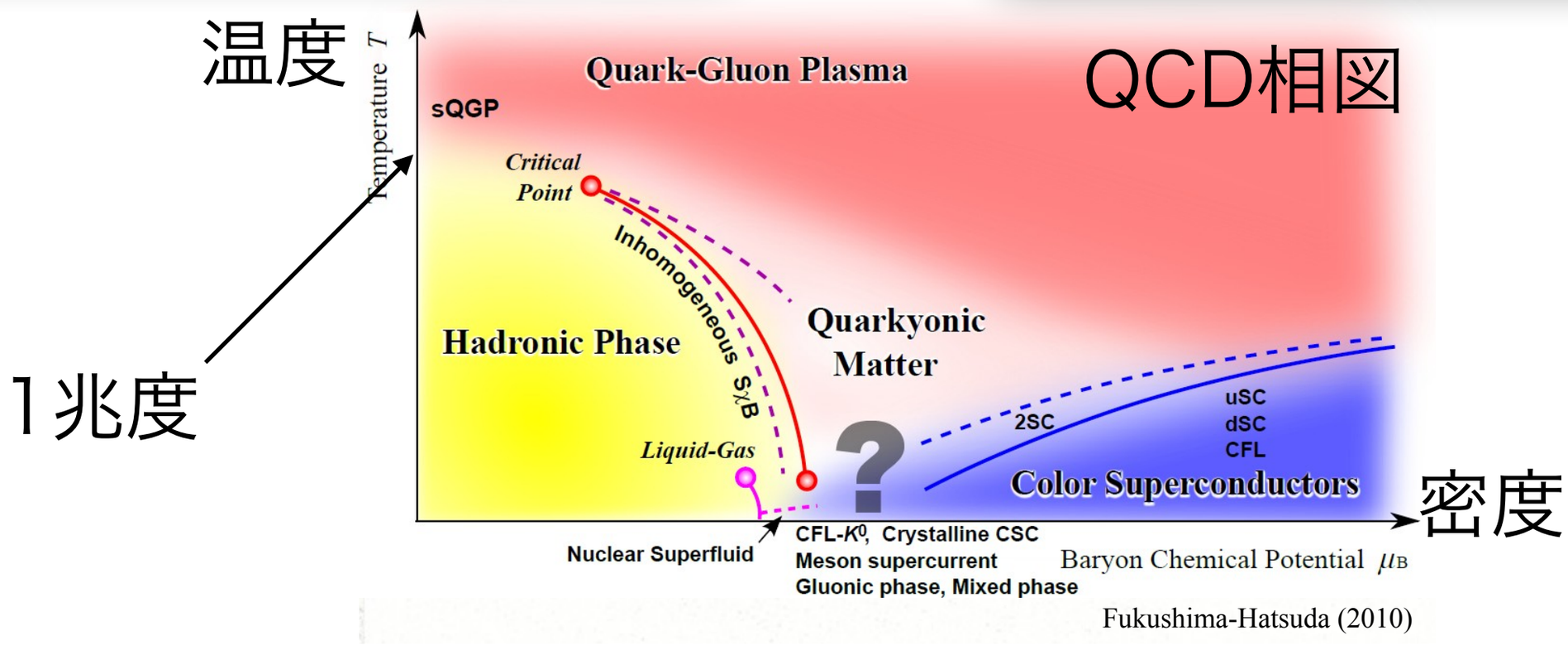
背景・動機

背景: クォークとグルーオンの統計系の相構造を調べたい

A. Tomiya

原子核等は、一兆度でとけてクォークとグルーオンへ(〜相転移)

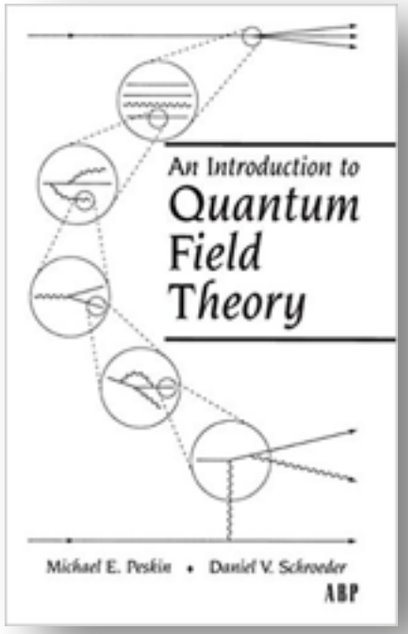




※福嶋さんトーク

背景: クォークとグルーオンの統計系の相構造を調べたい

クォークとグルーオンの量子系は、摂動できない& 非摂動効果が重要



$q(x)$ クォーク場

$\bar{q}(x)$ 反クォーク場

$A_\mu(x)$ グルーオン場

量子色力学(クォークとグルーオンの力学)の経路積分量子化

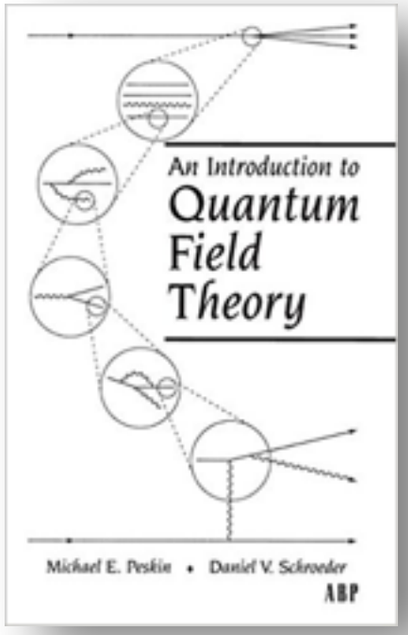
$$Z = \int \prod_x d\bar{q}(x) dq(x) dA(x) \exp\left[-\frac{1}{\hbar} \int_0^{1/(kT)} d\tau \int d^3x \left(\mathcal{L}_{\text{gluons}}(A) + \mathcal{L}_{\text{quarks}}(q,\bar{q}) + \mathcal{L}_{\text{interactions}}(A,q,\bar{q}) \right) \right]$$

- 結合定数が大きいので基本的に摂動できない
- 非摂動効果が重要であることが知られている

↑を見てみると、統計力学の分配関数と似てる…

背景: クォークとグルーオンの統計系の相構造を調べたい

離散時空上のクォークとグルーオンの統計力学は、モンテカルロできる



$q(x)$ クォーク場

$\bar{q}(x)$ 反クォーク場

$A_\mu(x)$ グルーオン場

量子色力学の経路積分量子化(クォークとグルーオンの量子統計力学)

$$Z = \int \prod_x d\bar{q}(x) dq(x) dA(x) \exp\left[-\frac{1}{\hbar} \int_0^{1/(kT)} d\tau \int d^3x \left(\mathcal{L}_{\text{gluons}}(A) + \mathcal{L}_{\text{quarks}}(q,\bar{q}) + \mathcal{L}_{\text{interactions}}(A,q,\bar{q}) \right) \right]$$

形式的に古典統計力学と似てるがxは、連続なので微妙に違う

→ xを離散化すれば 原子ラベルと同じ→基本的に※統計力学の手法が使える

→ 離散化した時空上の量子色力学(格子QCD)は、モンテカルロで計算できる

K. Wilson(1974)
M. Creutz(1980)

※柏さんのトーク

格子QCD=離散時空上のクォーク・グルーオンの量子多体系

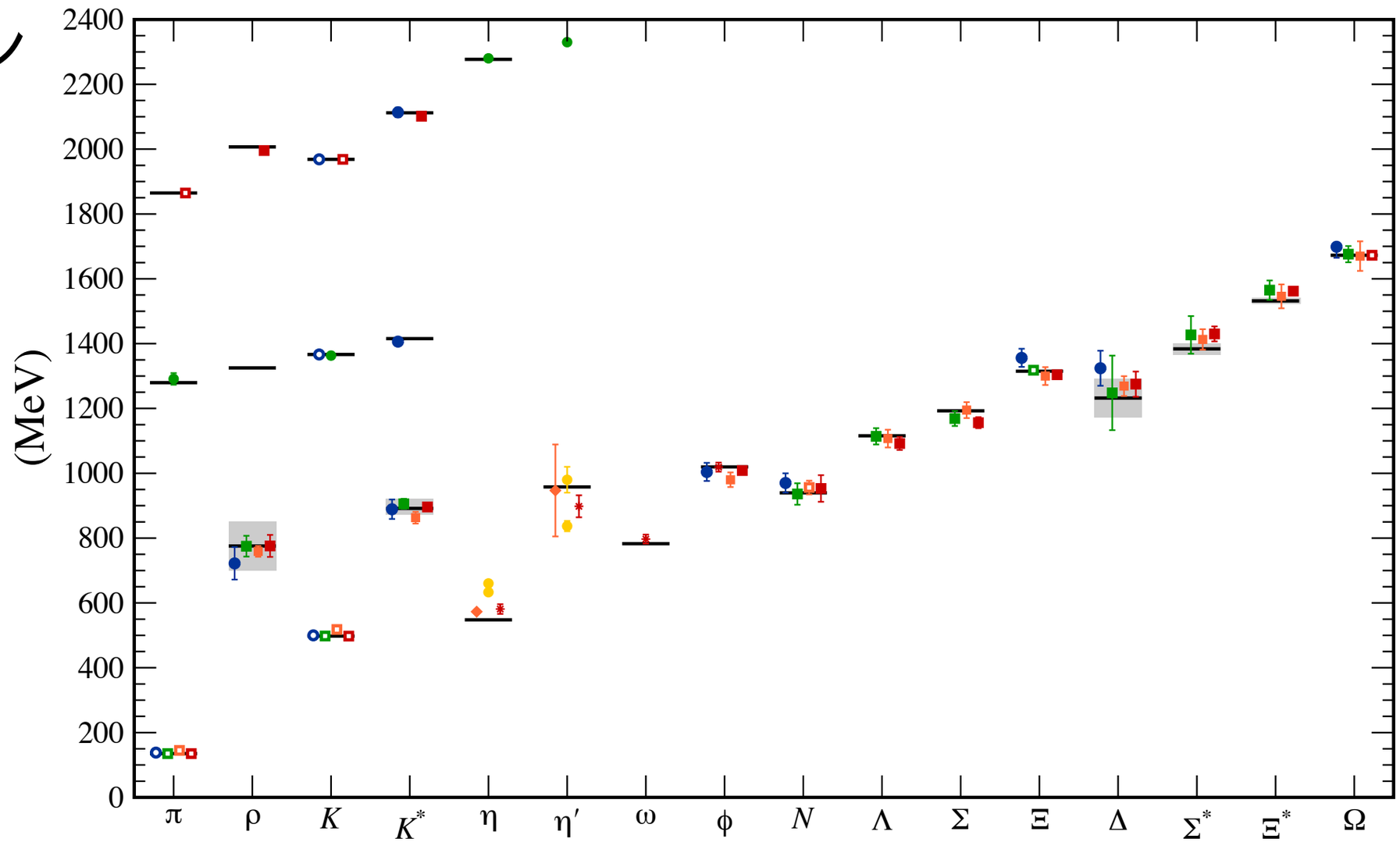
今まで成功してきた。

A. Tomiya

線が実験で決定したハドロンの質量
シンボルが格子QCDで決定したハドロンの質量

格子QCDへ3, 4つのインプット->あう。

ハドロン
質量



ハドロン
の種類

arXiv:1203.1204

ハイブリッドモンテカルロ法(HMC)

A. Tomiya

グルーオン場配位を連鎖的に作り出し、経路積分の結果を推定

経路積分(状態和) = すべての可能な状態の重み付け和(積分)

重み = ボルツマンweight ~ $\exp(-\int L)=\exp(-S)$

→作用S の値が大きいとき、結果への寄与は小さい

作用が小さい所の周りを重点的に計算

= ハイブリッドモンテカルロ法(HMC)

HMCの気持ち:

量子系 ～ <u>運動方程式</u> + <u>ランダム運動量</u>

Sが小さい所　　量子性

HMC(ハイブリッド(ハミルトン)モンテカルロ)とは

ランダム運動量を入れた運動方程式で場を更新し、経路積分をミミックする

HMC(スカラー場版):

0. 場の配位 $\phi^{\mathrm{old}} = \phi(x)$ を一つ選ぶ
1. ランダム運動量(場) $p(x)$ を確率~$\exp(-p^2/2)$ で生成
2. $H = p^2/2 + S[\phi]$ に対するハミルトン方程式を使い
   場 ϕ と運動量pを発展 = 数値的に運動方程式を解く (MD)
3. メトロポリス法で数値誤差を消去、ϕ^{new} を得る

1〜3を繰り返す

Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. Physics Letters B, 195(2):216 – 222, 1987.

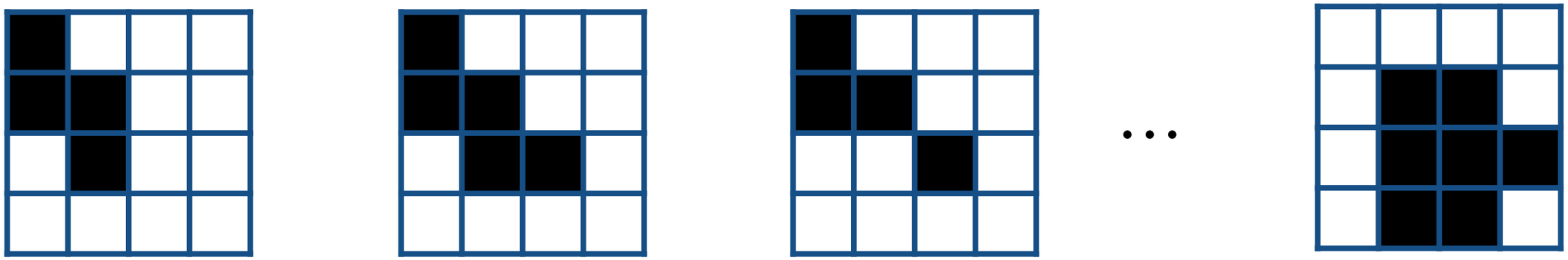
場の配位の列 $\phi_1, \phi_2, \cdots$ の平均 = 期待値

無限回やれば厳密に経路積分に一致することが知られている

自己相関時間

ランダム運動量を入れた運動方程式で場を更新し、経路積分をミミックする

HMC(ハイブリッドモンテカルロ)を使うと、
場の配位の列 ϕ_1, ϕ_2, … が得られる。



場のラベル	ϕ_1	ϕ_2	ϕ_3	…	ϕ_N
黒の数	4	5	4	…	7

黒のセルの数が近いので、独立な配位とみなせない
→ 似ている範囲を自己相関時間と呼ぶ(短い方が良い)

自己相関を短くして、高効率なサンプリングを

A. Tomiya

似たような場の配位のサンプル → 有効的な統計数の減少

HMCの気持ち:

$$\text{量子系} \sim \underset{\text{Sが小さい所}}{\underline{\text{運動方程式}}} + \underset{\text{量子性}}{\underline{\text{ランダム運動量}}}$$

→ (仮想)時間での発展が有効的に遅くなるときを考える

= 同じ計算時間でも似た場の配位のみが出てくる

→ ランダムサンプリングとして効率の低下

なんとかしたい！

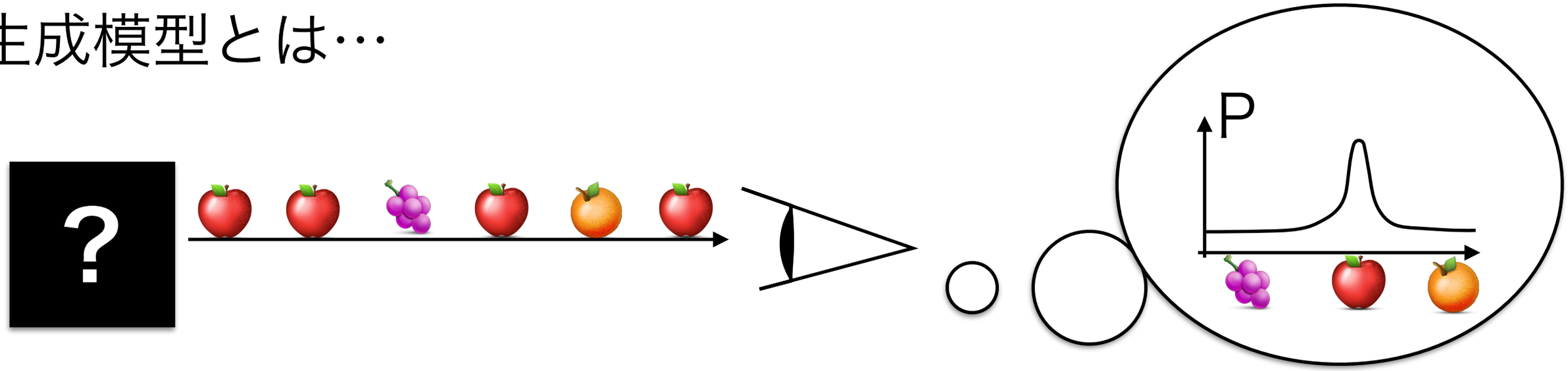
目次

ボルツマンマシン

生成模型、ボルツマンマシン

生成モデルは、経験確率に基づき、模倣したデータを生成できる

生成模型とは…



ブラックボックス(情報源)から情報が出て来るとして、
情報源をモデル化、模倣する機械学習の枠組みの1つ.
情報は、適当な確率分布に従って生成されてると考える

ボルツマンマシンは、生成模型の1つ
情報源から出てくる確率と模型の確率を似せる(最尤法)

制限ボルツマンマシン

A. Tomiya

生成モデルの一種で、経験に基づき、模倣データを生成できる

ボルツマンマシン ≒ イジング模型、2種のスピン自由度v, h(両方2値)：

※大関さん、柴さんのトーク

制限ボルツマンマシン: v同士、h同士には結合のない生成模型

$$E[v,h] = \sum_n b_n v_n + \sum_i c_i h_i + \sum_{nj} v_n w_{nj} h_j$$

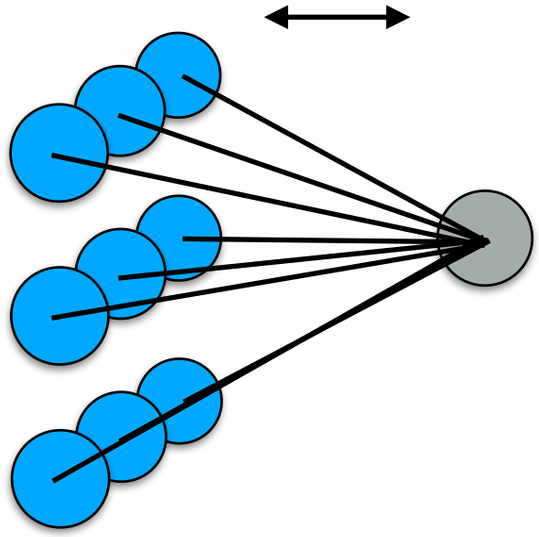
$$\mathcal{P}(v,h) = \frac{1}{Z} e^{-E[v,h]}$$

学習パラメータ: b, c, w

v, h は、スピン自由度(に対応)

v に値を入力→hに関する確率分布

h に値を入力→vに関する確率分布



通常 v(可視層)には、画像データを入力し、画像の分布を模倣する

学習パラメータは、勾配法でPを最大化するように学ばれる

ガウシアン制限ボルツマンマシンGRBM

vのスピン自由度を連続変数にしたもの. hは2値

A. Tomiya

ガウシアン制限ボルツマンマシン **= 可視層の変数を連続なモノにしたversion**

$$E[v,h] = \sum_n b_n v_n + \sum_i c_i h_i + \sum_{nj} v_n w_{nj} h_j$$

$$H_\theta^{\mathrm{GRBM}}[\phi,h] = \sum_n \frac{(\phi_n - a_n)^2}{2\sigma_n^2} - \sum_j b_j h_j - \sum_{n,j} \frac{\phi_n}{\sigma_n} w_{nj} h_j,$$

a, σ, b, w が学習パラメータ

通常 v(可視層)には、画像データを入力し、画像の分布を模倣する
学習パラメータは、勾配法でPを最大化するように学ばれる

先行研究

Self-Learning Monte-Carlo(SLMC)

A. Tomiya

有効模型の係数を実際のシミュレーション結果から決める

永井さんのトーク：

アイデア: 自己相関が長い系の厳密なハミルトニアンが分かってるとする。
係数をパラメータにした有効ハミルトニアンを用意し、シミュレーションと
回帰で係数を決める。有効系でシミュレーションし、自己相関を切る。
メトロポリステストで、厳密なハミルトニアンに対する分布に補正する。

上手いこといってる

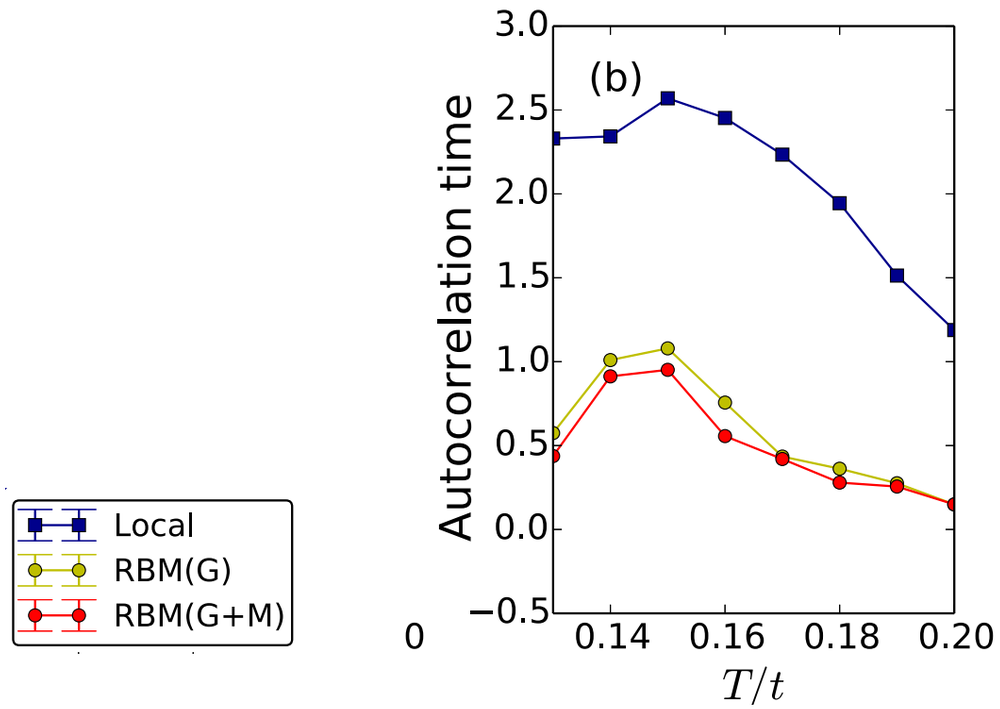
J.Liu et al arXiv: 1610.03137 など

RBMを用いたイジングの配位生成

RBMのみでイジングの配位を生成する

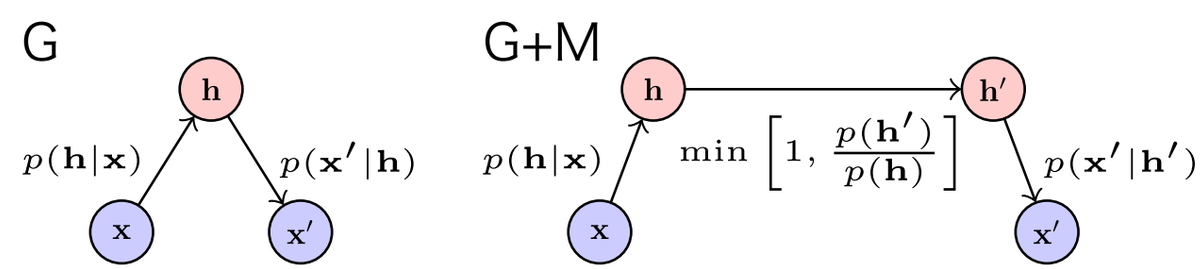
A. Tomiya

中国科学院組



The Falicov-Kimball model

$$\hat{H}_{\mathrm{FK}} = \sum_{i,j} \hat{c}_i^\dagger \mathcal{K}_{ij} \hat{c}_j + U \sum_{i=1}^{N} \left(\hat{n}_i - \frac{1}{2}\right)\left(x_i - \frac{1}{2}\right),$$



有効ハミルトニアンを使わずに、配位を生成

上手いことといってるらしい。

Li Huang and Lei Wang. PRB, 95(3):035105, 2017.

やったこと

機械学習で場の理論の配位を生成した

初の試みなので、多面的にチェックが必要

ボルツマンマシン+HMCで配位生成をし、
自己相関の短縮を試みた。

場の理論の配位生成への初の応用なので
何が起こるかわからない…

格子上の3次元ϕ4でやってみた。
→ よく理解されている模型なので
問題があればすぐにわかる。

HMC(ハイブリッドモンテカルロ)とは

ランダム運動量を入れた運動方程式で場を更新し、経路積分をミミックする

HMC(スカラー場版):

再掲

0. 場の配位 $\phi^{\mathrm{old}} = \phi(x)$ を一つ選ぶ
1. ランダム運動量(場) $p(x)$ を確率~$\exp(-p^2/2)$ で生成
2. $H = p^2/2 + S[\phi]$ に対するハミルトン方程式を使い
   場 ϕ と運動量pを発展 = 数値的に運動方程式を解く
3. メトロポリス法で数値誤差を消去、ϕ^{new} を得る

1〜3を繰り返す

場の配位の列 $\phi_1, \phi_2, \cdots$ が得られるので、期待値を求める

提案したアルゴリズム

配位の更新を運動方程式のみでなく、経験分布もつかう

Boltzmann machine + HMC(スカラー場版):

0. 場の配位 $\phi^{\mathrm{old}} = \phi(x)$ を一つ選ぶ

1. ランダム運動量(場) $p(x)$ を確率~$\exp(-p^2/2)$ で生成

Added 1.5. 学習済のボルツマンマシンに配位 ϕ をいれ、新配位 ϕ を生成

2. $H = p^2/2 + S[\phi]$ に対するハミルトン方程式を使い
   場 ϕ と運動量pを発展 = 数値的に運動方程式を解く

3. メトロポリス法で数値誤差を消去、ϕ^{new} を得る

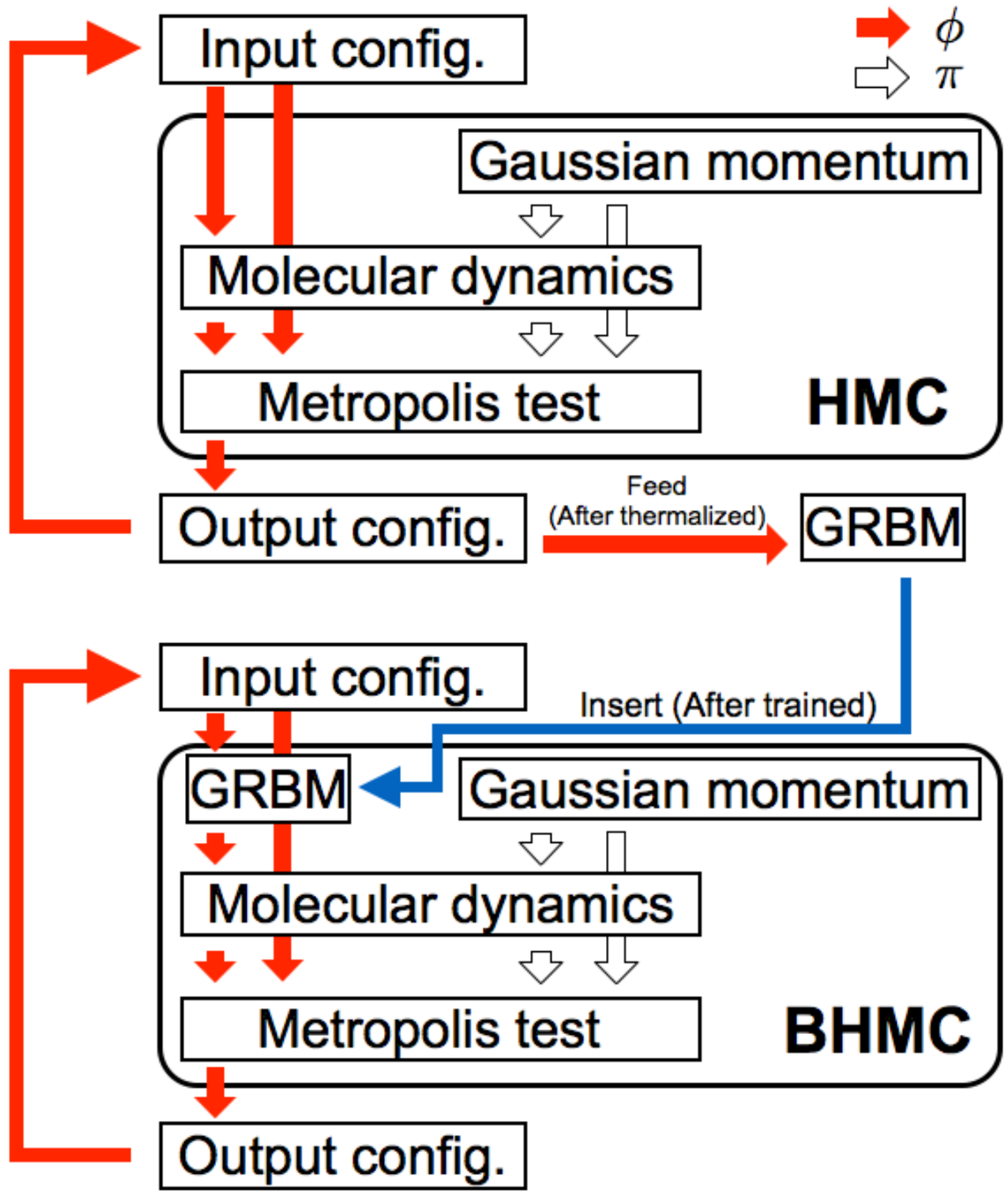
1〜3を繰り返す

場の配位の列 $\phi_1, \phi_2, \cdots$ が得られるので、期待値を求める

実際の運用のながれ

A. Tomiya

最初は、HMCのみで配位を生成。学習が進めば、ボルツマンマシンを挿入。



ϕ4理論

3次元格子上のϕ4理論の計算を行った

$$S[\phi] = \sum_n^N \left[-\frac{1}{2}\phi_n \Delta \phi_n + \frac{m^2}{2}\phi_n^2 + \frac{\lambda}{4!}\phi_n^4 \right].$$

$$\langle O \rangle = \frac{1}{\mathcal{Z}_{\rm lat}} \int \mathcal{D}\phi \; O[\phi] \; e^{-S[\phi]},$$

古典的には、

$m^2 < 0$: Z_2対称性が自発的にやぶれた相に。そうでなければ、対称相

対称相、破れ相そして臨界点で計算をおこなった

測定した物理量

作用密度、1点関数、2点関数

代表的な物理量である作用密度、1点関数、2点関数をしらべた。作用密度と1点関数はヒストグラムも計算。1点関数の自己相関を2つのアルゴリズムで比較した

$$\overline{S}/V = \frac{1}{V}\frac{1}{N_{\rm conf}}\sum_{c=1}^{N_{\rm conf}} S[\phi^{(c)}],$$

$$\overline{\phi}/V = \frac{1}{N_{\rm conf}}\sum_{c=1}^{N_{\rm conf}}\frac{1}{V}\sum_{n}^{N}\phi_n^{(c)},$$

$$\overline{G}(t) = \sum_{dx,dy=1}^{N_x,N_y}\frac{1}{N_{\rm conf}}\sum_{c=1}^{N_{\rm conf}}\frac{1}{V}\sum_{n}^{N}\phi^{(c)}_{t+n_t,dx+n_x,dy+n_y}\phi_n^{(c)},$$

(破れ相では、con. part)

結果

統計と結果

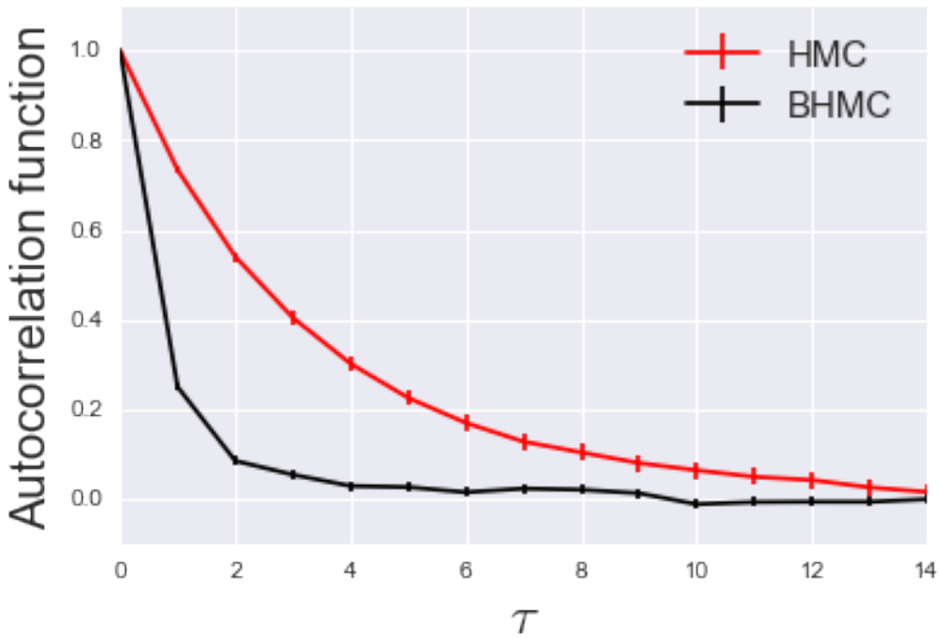
数字だけ見ると良さそうに見える

Alg.	Phase	N_{conf}	$\langle\phi\rangle/V$	$\langle S\rangle/V$	τ_{int}
HMC	Symmetric	10^4	0.00±0.05	0.48±0.03	4.4±0.3
	Broken	10^4	-3.94±0.04	-2.70±0.03	2.8±0.2
BHMC	Symmetric	10^4	0.00±0.04	0.50±0.04	2.0±0.1
	Broken	10^4	-3.95±0.03	-2.73±0.04	2.5±0.2

一見良さそうに見えるが…？

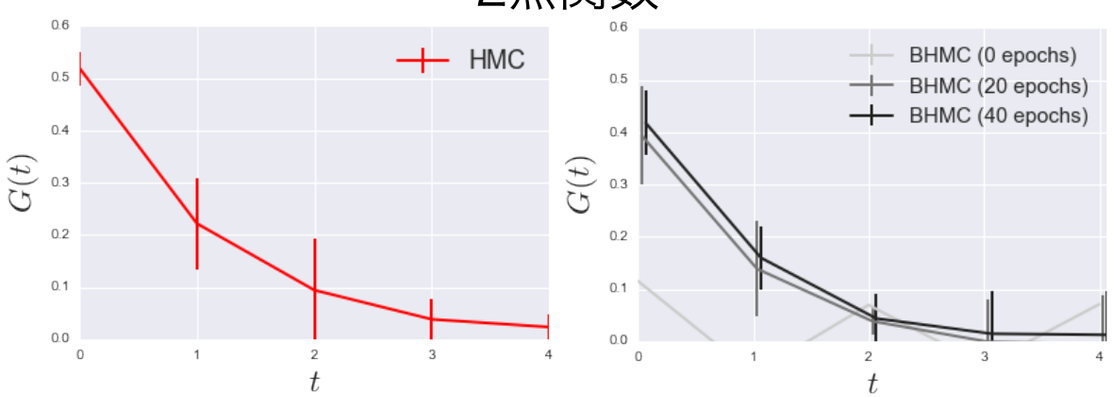
結果

対称相での結果

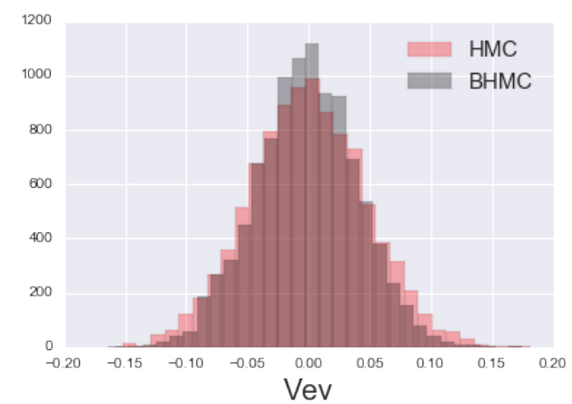


- 自己相関は、短くなる
- 1点関数、2点関数はコンシステント
- 作用密度が少しズレている

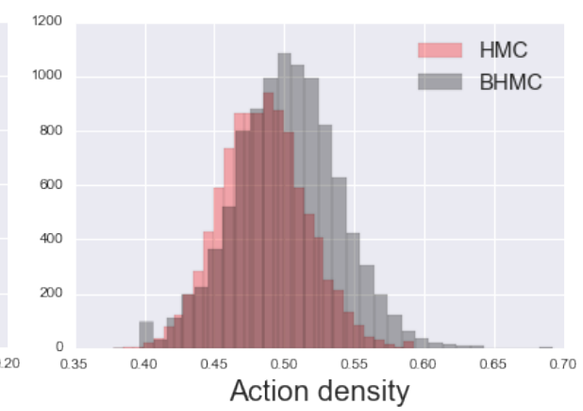
2点関数



1点関数

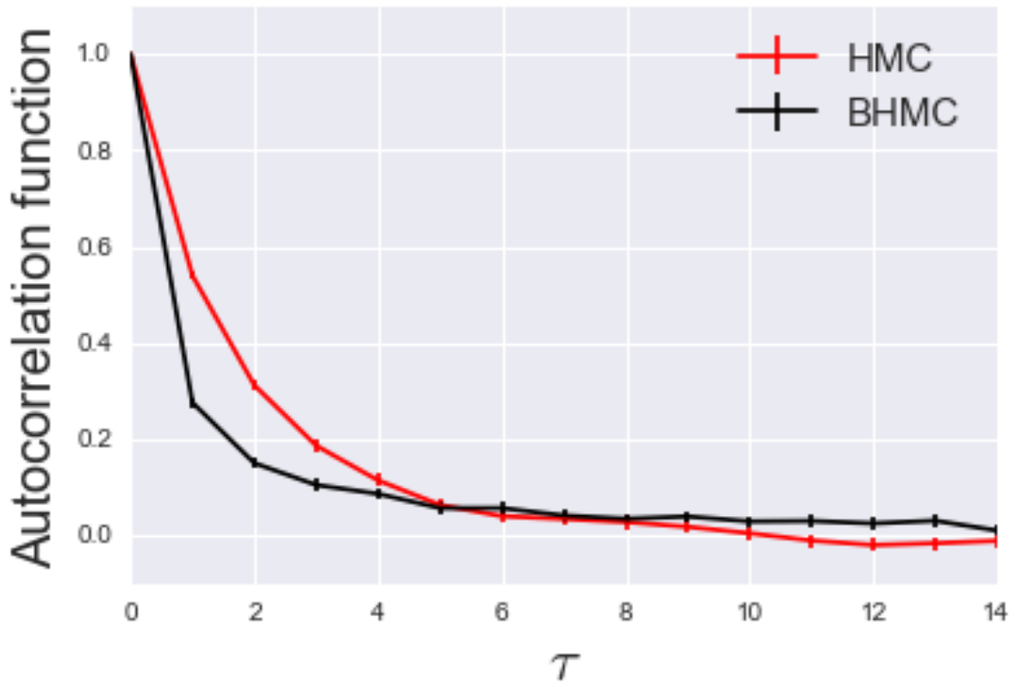


作用密度



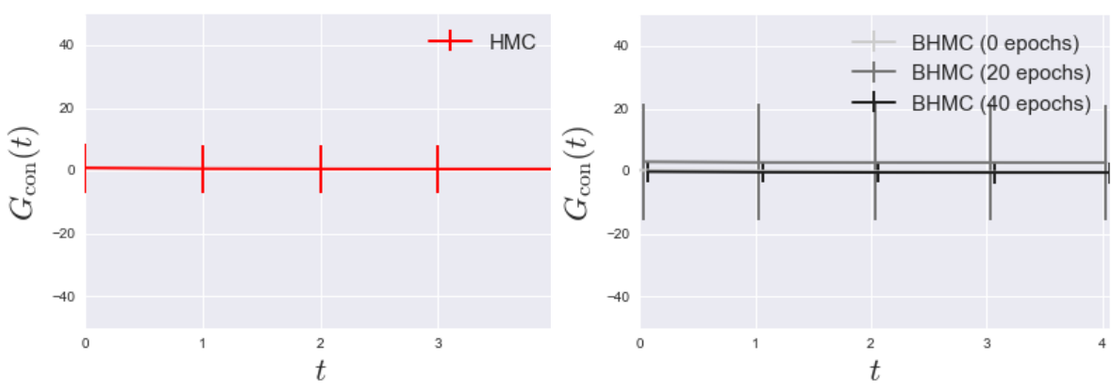
結果

破れ相での結果

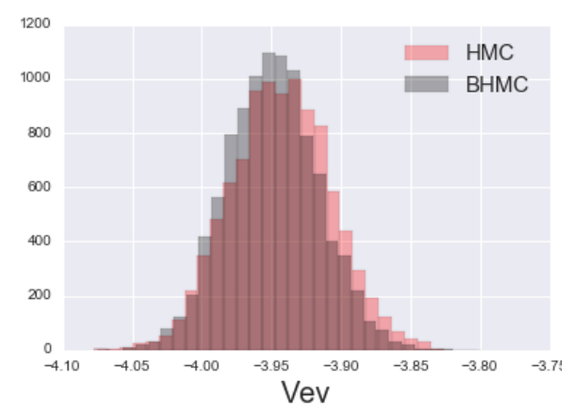


- 自己相関は、(少し)短くなる
- 1点関数、2点関数はコンシステント
- 作用密度が少しズレている

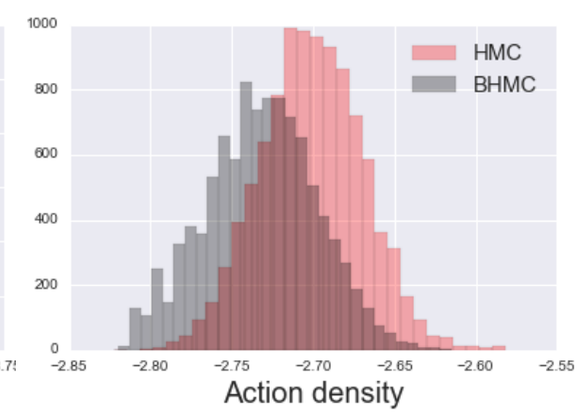
2点関数



1点関数

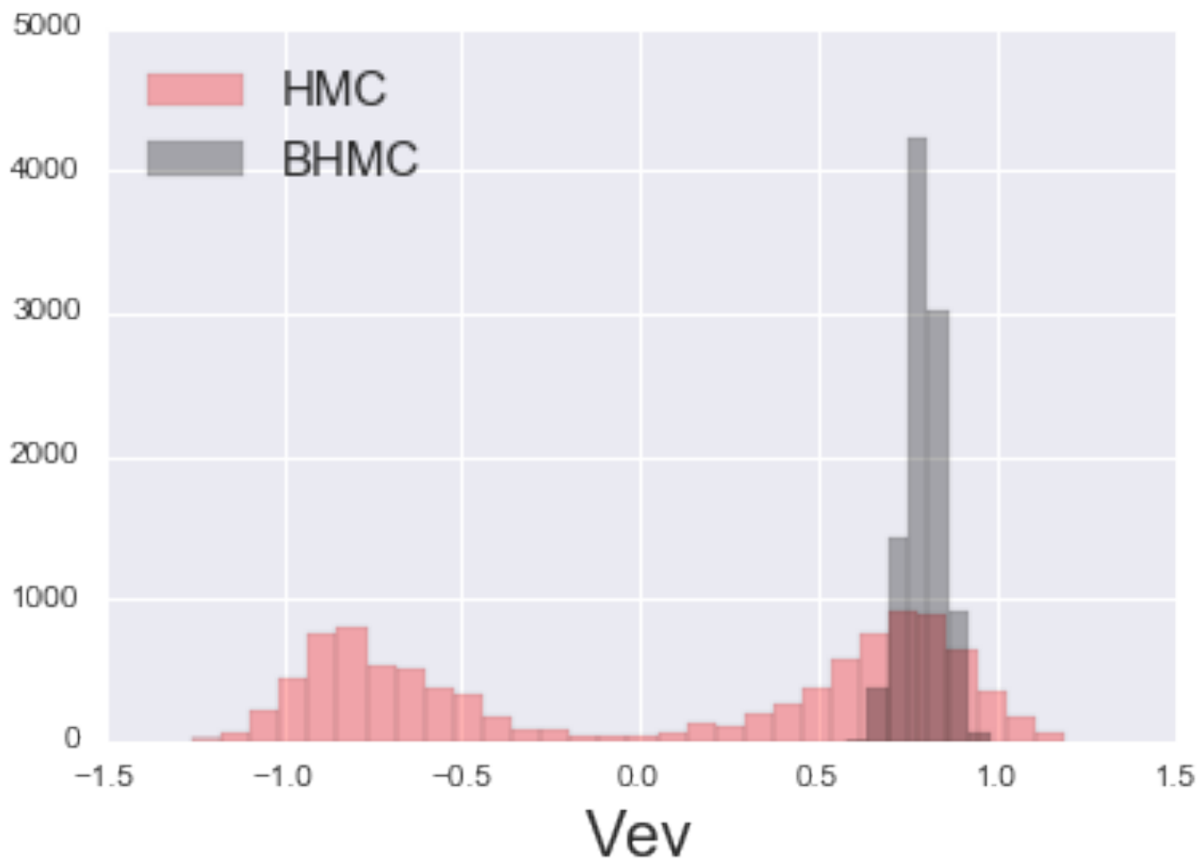


作用密度



結果

臨界点での結果



- １点関数すら合わない

ボルツマンマシンの限界？

機械学習のモデルを変えればうまくいく？

結果を省みる

モデルの表現力？

機械学習のモデルを変えればうまくいく？

$$H_\theta^{\mathrm{GRBM}}[\phi, h] = \sum_n \frac{(\phi_n - a_n)^2}{2\sigma_n^2} - \sum_j b_j h_j - \sum_{n,j} \frac{\phi_n}{\sigma_n} w_{nj} h_j,$$

$$p_\theta[\phi, h] = \frac{e^{-H_\theta^{\mathrm{GRBM}}[\phi,h]}}{\mathcal{Z}_\theta},$$

ボルツマンマシンは、本質的にガウス分布。
混合ガウス分布にする？
GAN(Generative Adversarial Networks) などより表現力の高い模型？
理論的に経路積分に収束する模型？(あるのか？)

まとめ：問題設定と結果

提案したアルゴリズムで自己相関は短縮。相転移点で不具合

問題: 格子QCDでの配位生成を効率化したい！
効率化できれば、モンテカルロ法(HMC)からくる統計誤差を短時間で減らすことが出来る。

手法: 機械学習をHMCに組み込む
機械学習は、それっぽい画像を作れる
　→ 応用できないか？
性質が良く分かってる3次元φ4理論で試してみた

Radford et al. (2015)

結果: 対称相、破れ相で、1点関数, 2点関数は無矛盾
臨界点で、不具合がある

バックアップ

自己相関

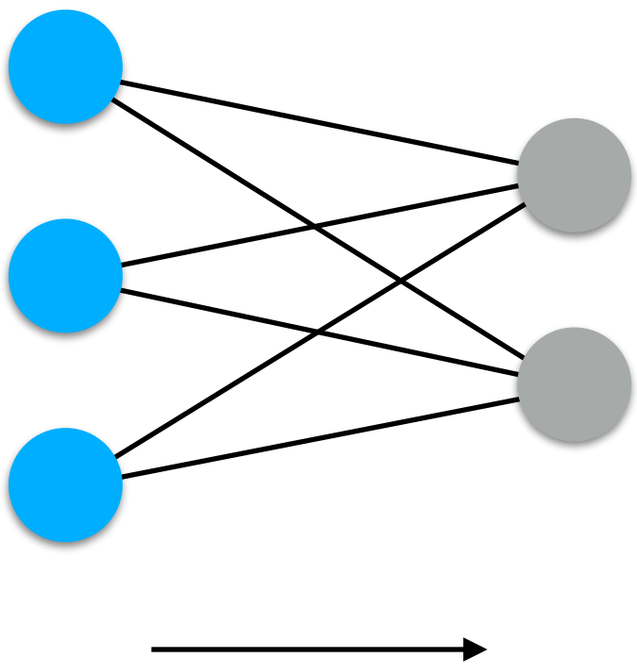
MCMCでつくられる配位間の類似性

$$\overline{\Gamma}(\tau) = \frac{1}{N_{\mathrm{conf}} - \tau} \sum_{c}^{N_{\mathrm{conf}}} (O_c - \overline{O})(O_{c+\tau} - \overline{O}),$$

ニューラルネットワークとボルツマンマシン

ニューラルネットワーク=一方向, ボルツマンマシン=双方向

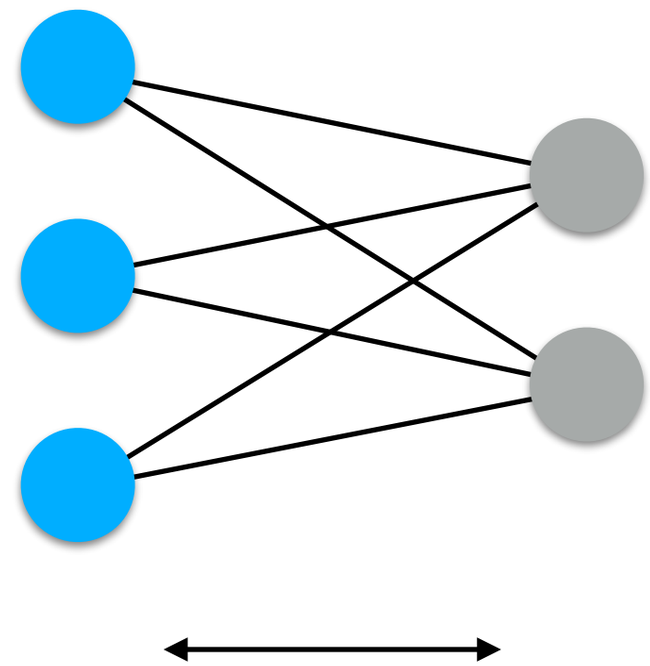
ニューラルネット



決定論的

正解と出力を合わせる

ボルツマンマシン



確率論的

分布を近づける