

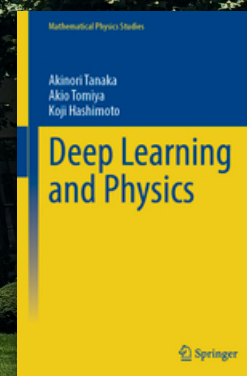
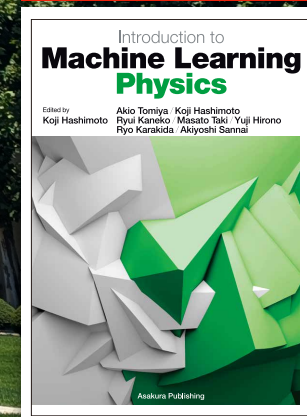
This talk is based on
JPSJ 86, 063001 (2017)
JPSJ 94, 031006 (2025)

From Ising to QCD: Phases, Symmetry, and AI

Akio Tomiya TWCU (Senior Lecturer)
R-CCS (Visiting researcher), Kyoto University
(Program specified associate professor)



My new book
2025 Oct



<http://bit.ly/3JrEjls>

Happy anniversary quantum physics!

Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen.

Von **W. Heisenberg** in Göttingen.

(Eingegangen am 29. Juli 1925.)

In der Arbeit soll versucht werden, Grundlagen zu gewinnen für eine quantentheoretische Mechanik, die ausschließlich auf Beziehungen zwischen prinzipiell beobachtbaren Größen basiert ist.

Bekanntlich läßt sich gegen die formalen Regeln, die allgemein in der Quantentheorie zur Berechnung beobachtbarer Größen (z. B. der Energie im Wasserstoffatom) benutzt werden, der schwerwiegende Einwand erheben, daß jene Rechenregeln als wesentlichen Bestandteil Beziehungen enthalten zwischen Größen, die scheinbar prinzipiell nicht beobachtet werden können (wie z. B. Ort. Umlaufszeit des Elektrons).

3. *Quantisierung als Eigenwertproblem;* *von E. Schrödinger.*

(Erste Mitteilung.)

E. Schrödinger.

§ 1. In dieser Mitteilung möchte ich zunächst an dem einfachsten Fall des (nichtrelativistischen und ungestörten) Wasserstoffatoms zeigen, daß die übliche Quantisierungsvorschrift sich durch eine andere Forderung ersetzen läßt, in der kein Wort von „ganzen Zahlen“ mehr vorkommt. Vielmehr ergibt sich die Ganzzahligkeit auf dieselbe natürliche Art, wie etwa die Ganzzahligkeit der *Knotenzahl* einer schwingenden Saite. Diese neue Auffassung ist verallgemeinerungsfähig und rührt, wie ich

1925 for 1d Ising model

To study phase transition, Wilhelm Lenz proposed “Ising model” in 1920, and it has been solved in 1924-1925 by E. Ising.

Beitrag zur Theorie des Ferromagnetismus ¹⁾.

Von Ernst Ising in Hamburg.

(Eingegangen am 9. Dezember 1924.)

Es wird im wesentlichen das thermische Verhalten eines linearen, aus Elementarmagneten bestehenden Körpers untersucht, wobei im Gegensatz zur Weiss'schen Theorie des Ferromagnetismus kein molekulares Feld, sondern nur eine (nicht magnetische) Wechselwirkung benachbarter Elementarmagnete angenommen wird. Es wird gezeigt, daß ein solches Modell noch keine ferromagnetischen Eigenschaften besitzt und diese Aussage auch auf das dreidimensionale Modell ausgedehnt.

1. Annahmen. Die Erklärung, die P. Weiss ²⁾ für den Ferromagnetismus gegeben hat, ist zwar formal befriedigend, doch läßt sie besonders die Frage nach einer physikalischen Erklärung der Hypothese des molekularen Feldes offen. Nach dieser Theorie wirkt auf jeden

2d version has solved by Lars Onsager in 1944

After these developments, Ising model is a testbed for new method to investigate phase transitions

I used AI to find phase transition (almost) 10 years ago

From Ising model

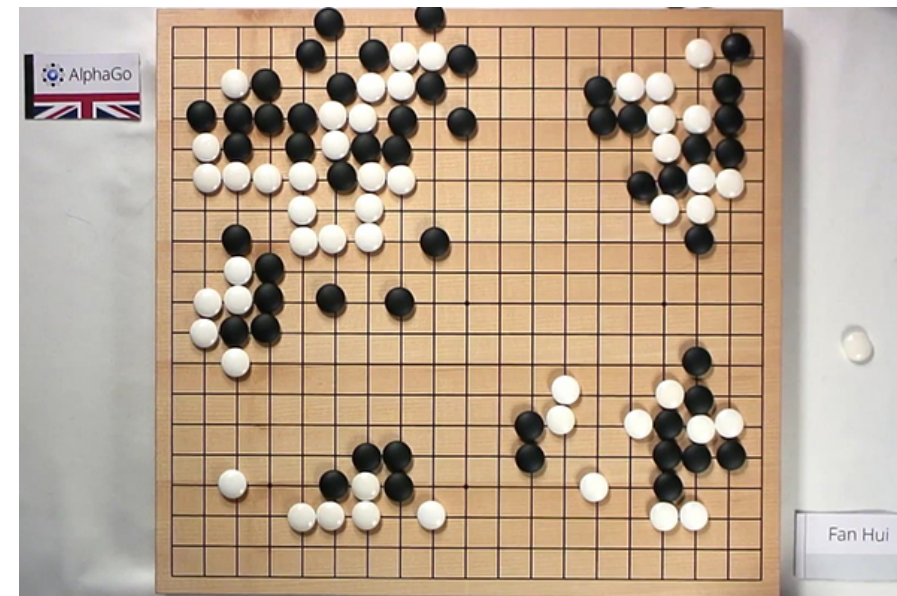
K. Kashiwa, Y. Kikuchi AT 1812.01522
A.Tanaka AT 1609.09087

Neural net as a thermometer (Classification problem)

A.Tanaka AT 1609.09087

K. Kashiwa, Y. Kikuchi AT 1812.01522

In 2016, a professional go-player (Lee Sedol) lose to AI, It make a boom of AI.



DOF



19x19 lattice

Akinori Tanaka and I started to apply AI to theoretical physics
To test ability of “AI”, we decided to use the 2d Ising model
We tried to determine the phase of it without any prior knowledge

Neural net as a thermometer (Classification problem)

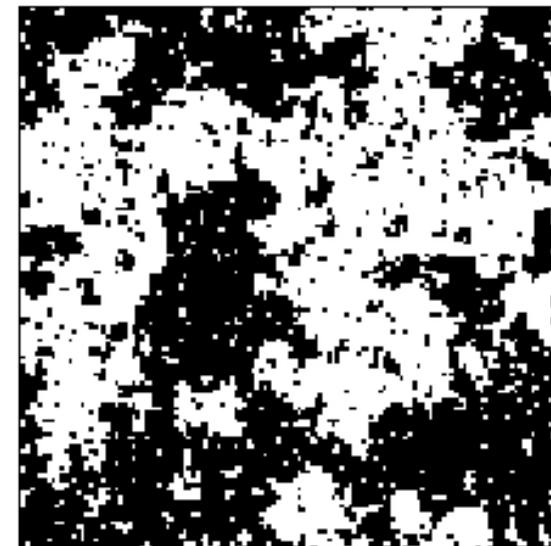
A.Tanaka AT 1609.09087

K. Kashiwa, Y. Kikuchi AT 1812.01522

Ising model = a (classical) toy model of magnets. More than 2d, it has 2 phases.
Hamiltonian has $\mathbb{Z}_2$ symmetry and lowest energy states break it.

$$H(S) = -J \sum_{(i,j):\text{nearest}} S_i S_j$$

$$P(S) = \frac{e^{-\beta H(S)}}{Z}$$



DOF



Using Markov chain Monte Carlo (MCMC), we generate spin configurations and we tried to guess the phase transition temperature from the configurations only
Namely, we do not use input of the critical temperature.

Neural net as a thermometer (Classification problem)

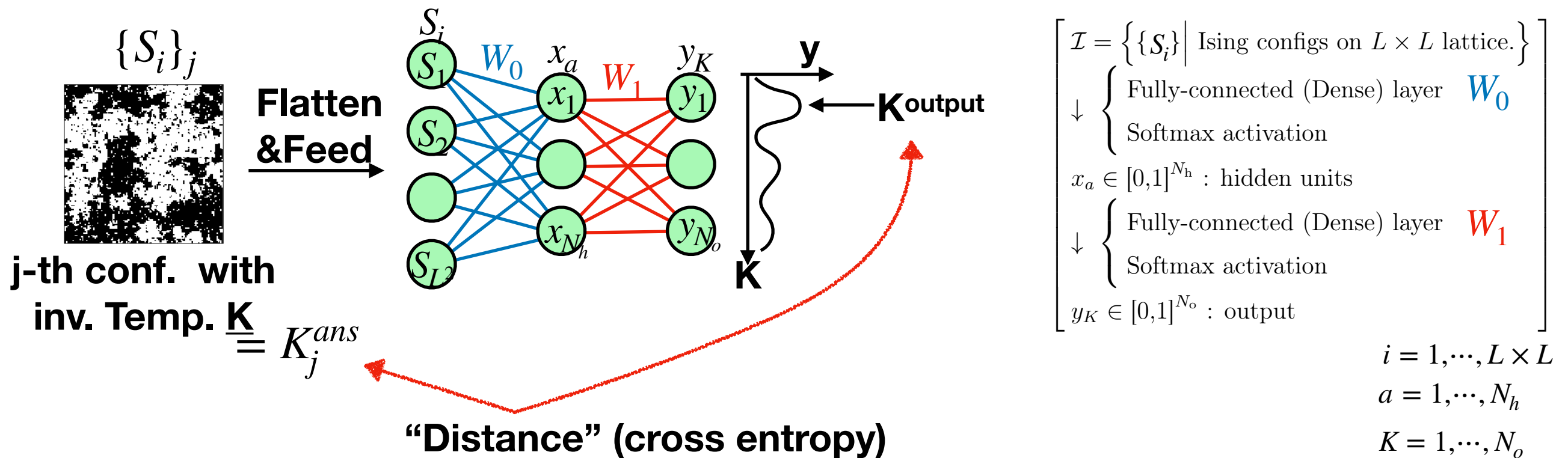
A.Tanaka AT 1609.09087

K. Kashiwa, Y. Kikuchi AT 1812.01522

Input = Configurations of inverse temperatures $\mathbf{K} \in (\mathbf{K}_{\min} < \mathbf{K}_{\text{cr}} < \mathbf{K}_{\max})$

Output = inverse temperature

NN is trained as a “thermometer”



“Distance” (cross entropy)

$$E(y_K) \propto \sum_{j \in \text{Conf}} \delta_{K, K_j^{\text{ans}}} \times (-\log y_K(\{S\}_j)) \quad 0 < y_K < 1$$

Minimize E, NN becomes a thermometer but we focus on W_1

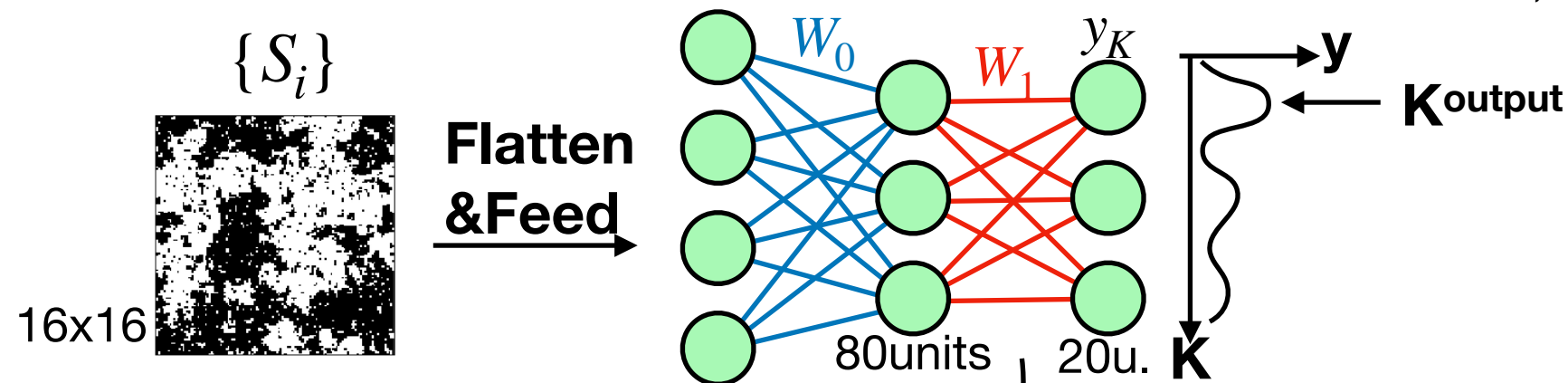
Neural net captures phase transition

Akio Tomiya

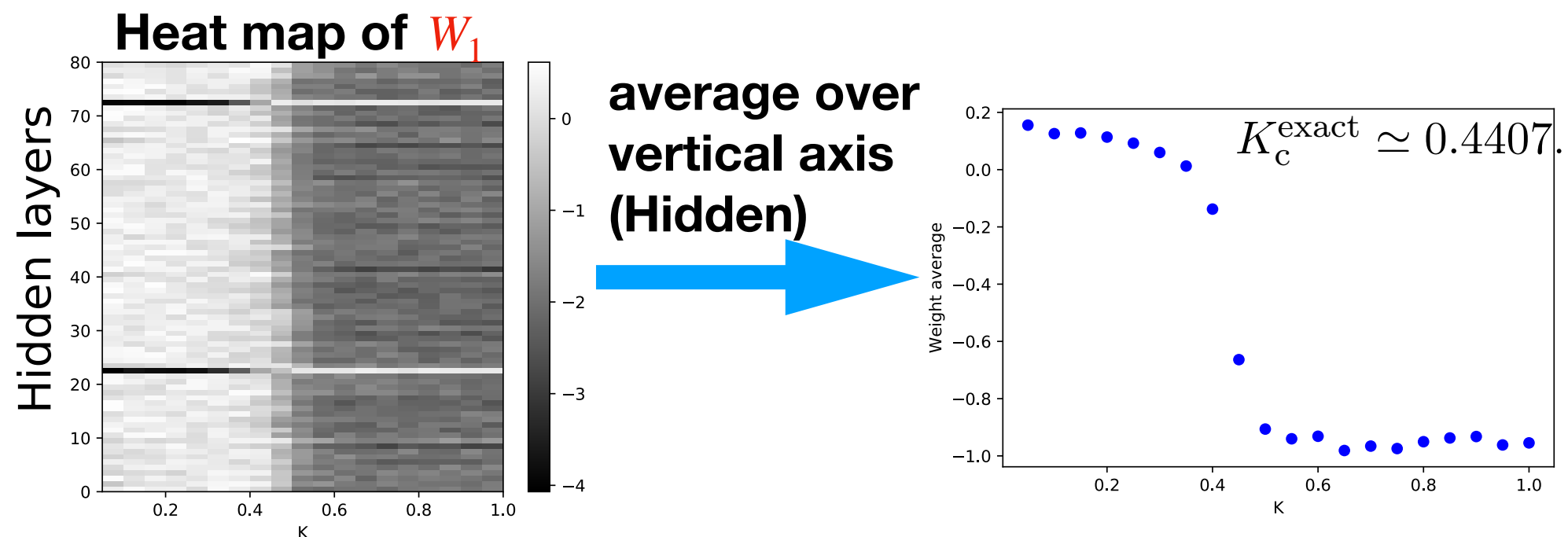
Heat map of weight W in second layer has structure

A.Tanaka AT 1609.09087

K. Kashiwa, Y. Kikuchi AT 1812.01522



After training, W_1 gets some pattern, especially there is a border around critical temp.



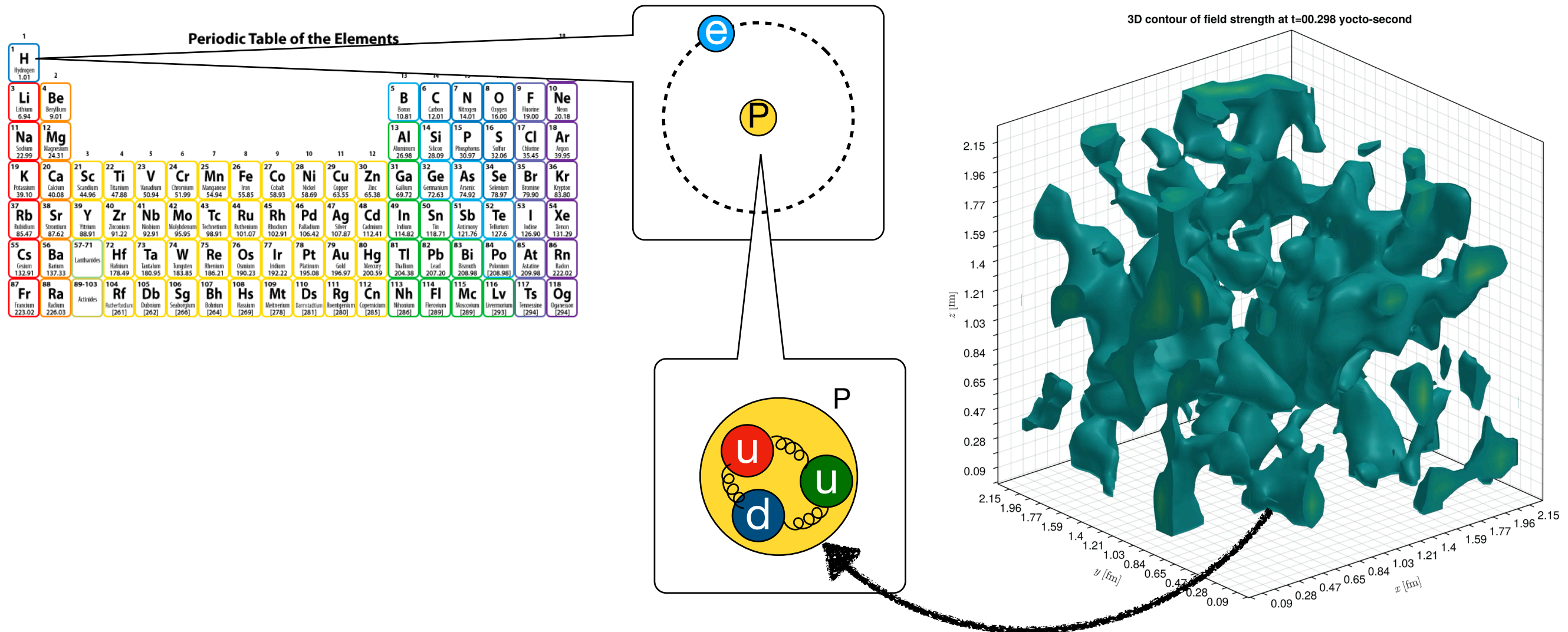
Few percent accuracy of K_{critical} !

Quantum chromodynamics (QCD)

Recognition to Generation

Lattice QCD?

Elementary particle physics



QCD = Quantum Chromo-dynamics

= A fundamental theory for particles inside of nuclei

Quantum many body(fields), relativistic, strongly correlated

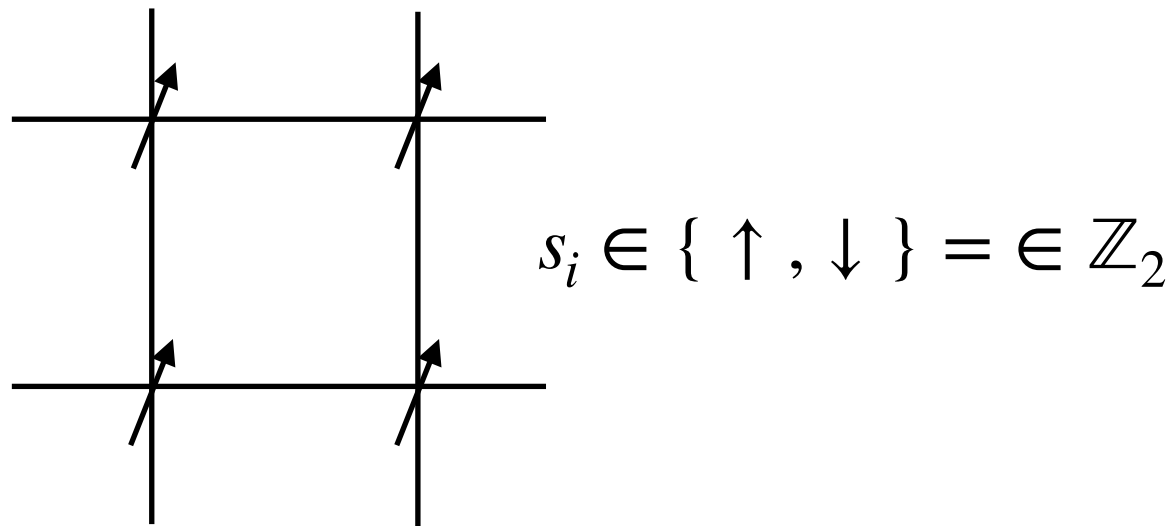
-> One of the hardest problem -> “Lattice” QCD enables to calculate it

Lattice gauge theory is a “generalized” Ising model

Ising model

$$\sum_{\{s\}} e^{-\tilde{H}[s]} O[s]$$

State sum for thermodynamics



Energy function = Hamiltonian

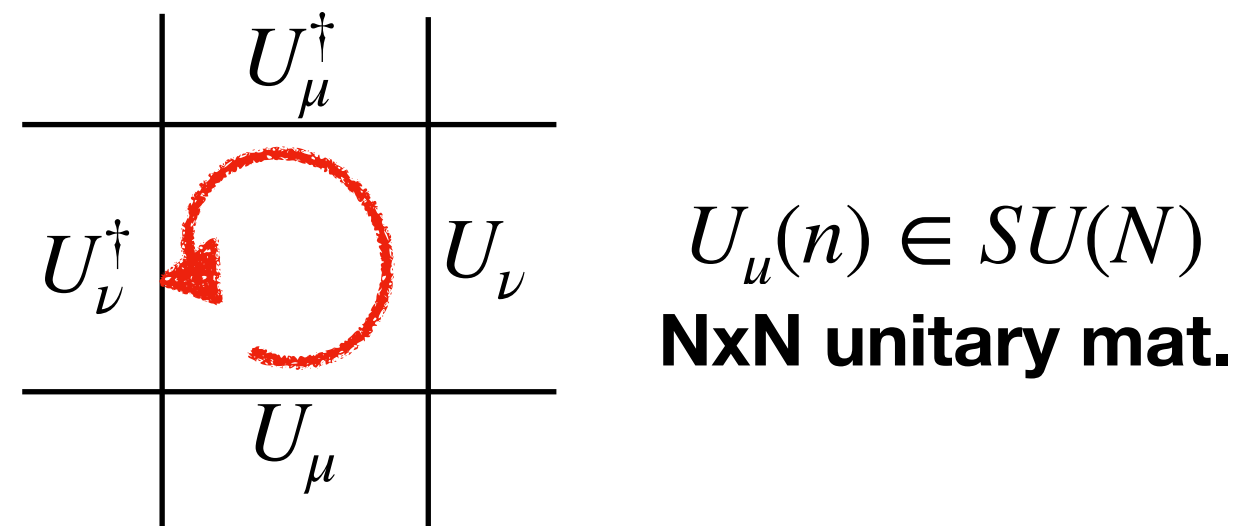
$$\tilde{H} = -\frac{J}{T} \sum_{ij} s_i s_j$$

$\mathbb{Z}_2$ dofs live on points

Lattice gauge theory (path integral)

$$\int_{\text{all possible config}} DU e^{-S[U]} O[U]$$

Path integral for quantization



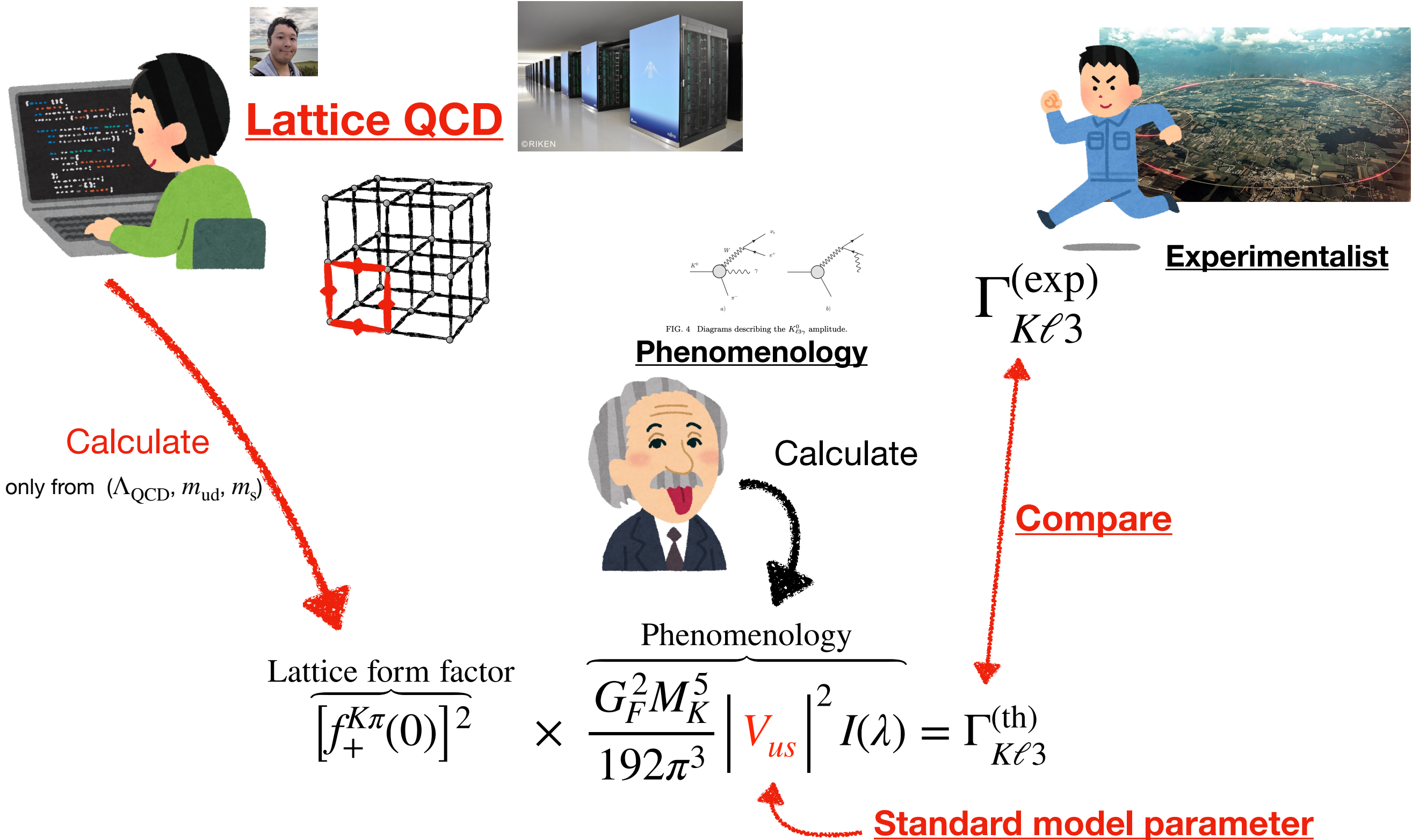
Energy function = Euclidean action

$$S = -\frac{2N}{g^2} \sum_{\square} \text{Re tr} [U_\mu U_\nu U_\mu^\dagger U_\nu^\dagger]$$

$SU(N)$ dofs live on bonds, **N=3 is QCD**

Introduction

Lattice QCD = non-perturbative input for phenomenology

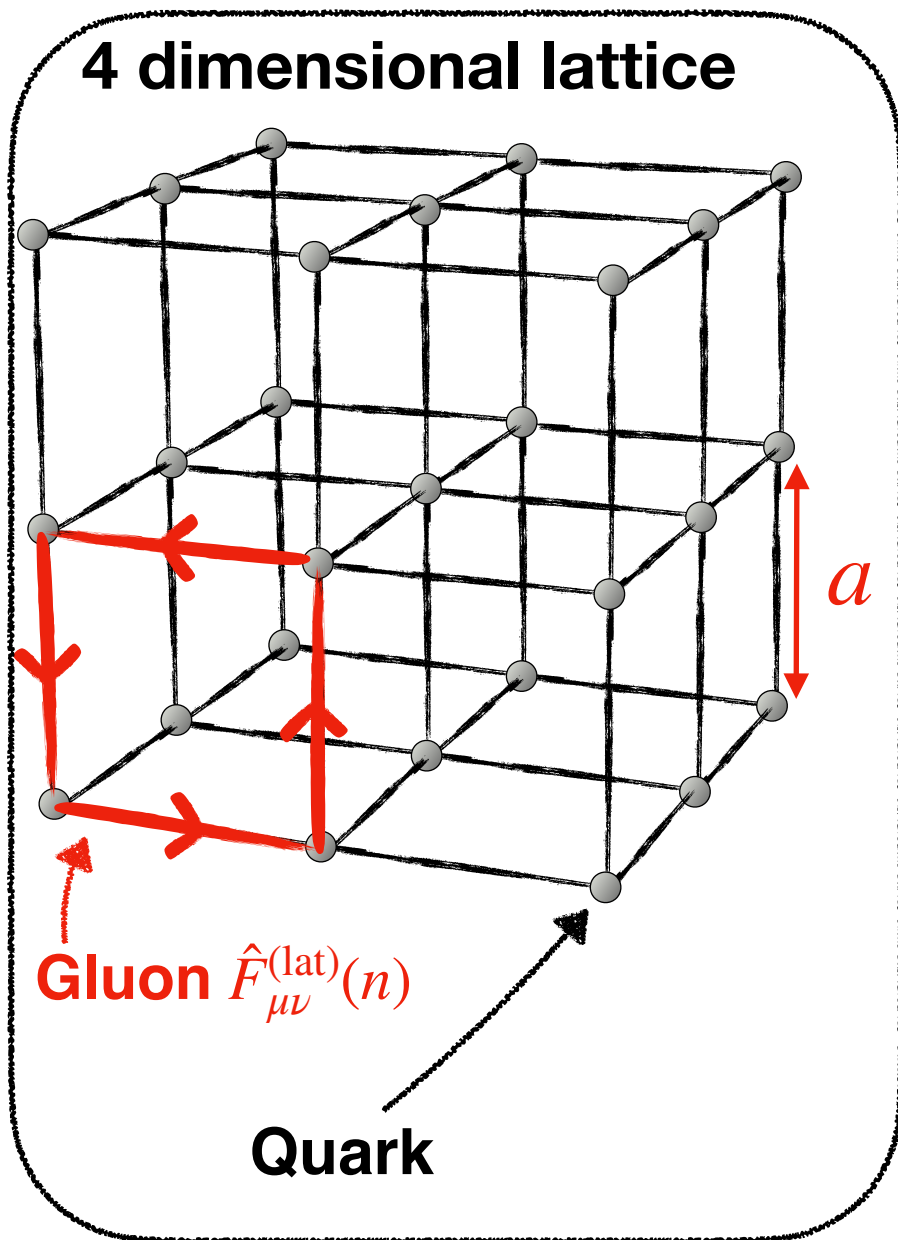


Lattice QCD?

Lattice QCD = QCD on a discretized spacetime

Quantum expectation value $\langle \mathcal{O} \rangle = \frac{1}{Z} \int \frac{DU}{10^{11} \text{ dim. integral}} e^{-S_{\text{QCD}}[U]} \mathcal{O}[U]$ $DU \equiv \prod_{n \in \text{lat}} \prod_{\mu=1}^4 dU_{\mu}(n)$
Finite dim. integral!

4 dimensional lattice



a [fm] is a lattice spacing (unit of discretization) needed to define the theory (as differentiation)

Theory on the lattice spacetime

$$S_{\text{QCD}}[U] = \sum_n \left[-\frac{1}{g^2} \text{Re tr } e^{ia^2 \hat{F}_{\mu\nu}^{(\text{lat})}[U]} \right] + (\text{Quarks})$$

(after all calculations, we take $a \rightarrow 0$ limit with tuning of g)
we can “restore” continuum spacetime with quantum fields

Because of the dimensionality of the integral,
we evaluate this integral using a Markov-Chain
Monte Carlo.

Gauge symmetry is essential



Monte-Carlo integration is available

Simon Duane, Anthony Kennedy, Brian Pendleton and Duncan Roweth 1987

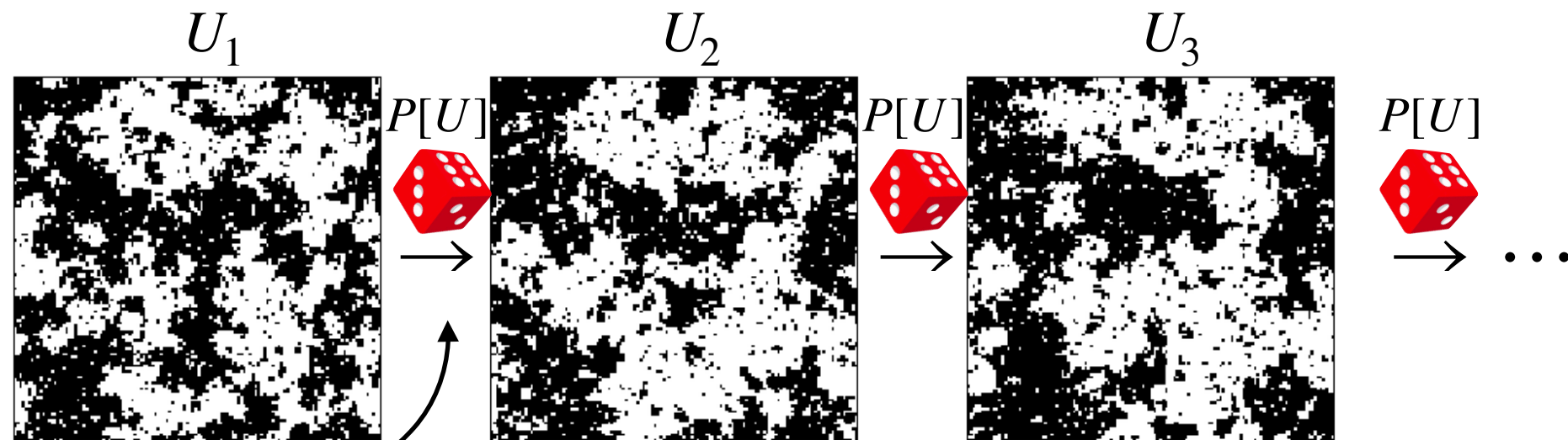
Target integration
= expectation value

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U e^{-S_{\text{eff}}[U]} \mathcal{O}(U)$$

$256^4 \times 4 \times 8$ **dimensional Integral**

$$S_{\text{eff}}[U] = S_{\text{gauge}}[U] - \log \det(\mathcal{D}[U] + m)$$

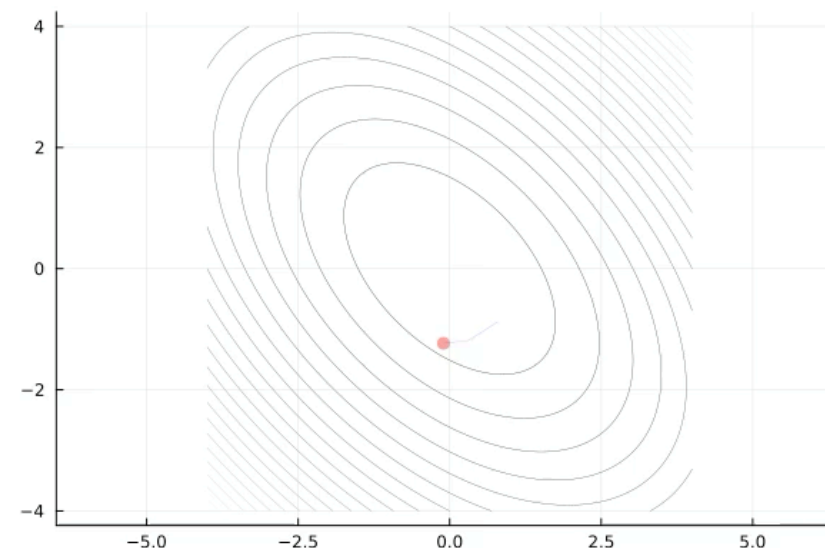
Monte-Carlo: Generate field configurations with “ $P[U] \propto e^{-S_{\text{eff}}[U]}$ ” . It gives expectation value



HMC: Hybrid (Hamiltonian) Monte-Carlo
De-facto standard algorithm (Exact)

Random momentum + EOM
= **Random walk like algorithm**

$$S(x, y) = \frac{1}{2}(x^2 + y^2 + xy)$$



Monte-Carlo integration is available

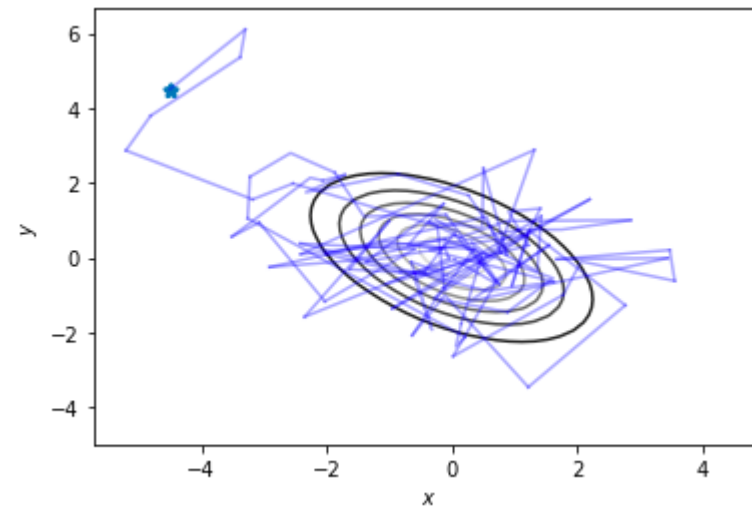
M. Creutz 1980

Target integration
= expectation value

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U e^{-S_{\text{eff}}[U]} \mathcal{O}(U)$$
$$S_{\text{eff}}[U] = S_{\text{gauge}}[U] - \log \det(\mathbb{D}[U] + m)$$

HMC: Hybrid (Hamiltonian) Monte-Carlo
De-facto standard algorithm (Exact)

Random momentum + EOM
= Random walk like algorithm



$$S(x, y) = \frac{1}{2}(x^2 + y^2 + xy)$$

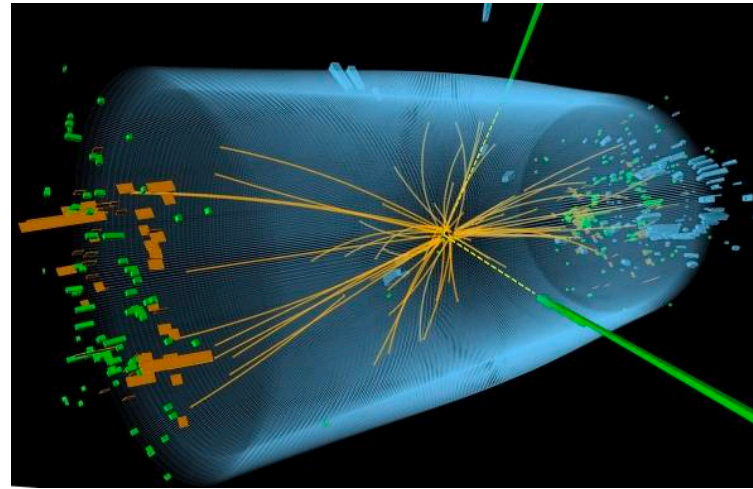
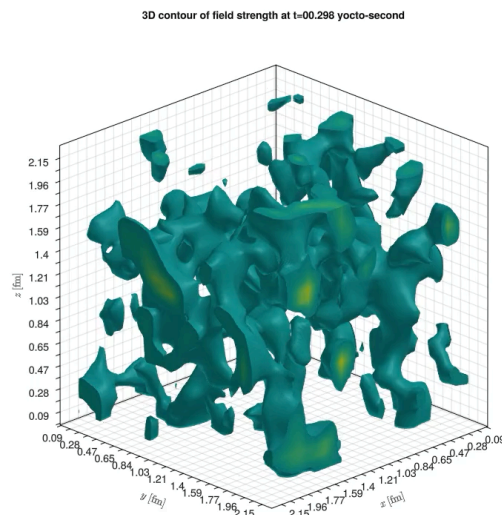
In each time step of “EOM”, we have to solve a linear equation $\mathbb{D}$ (Dslash), namely **quark part, is very expensive**. It dominates 50-90 % of numerical cost.

$$\mathbb{D} \vec{x} = \vec{b}$$

10^9 dimension for $L = 256$

We want to reduce cost of quark part

What is our final goal for our research field?



In short, we simulate of elementary particles in nuclei

Using super computers + Lattice QCD, we can understand...

- melting of protons/neutrons etc. at high temperatures

 - related to the **history of the universe**

- attractive/repulsive forces between atomic nuclei

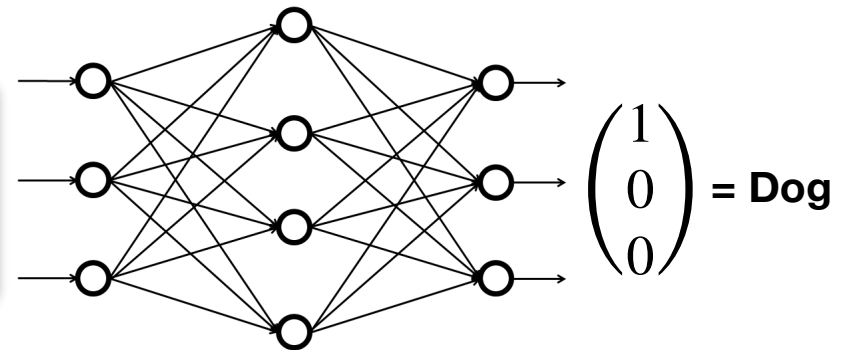
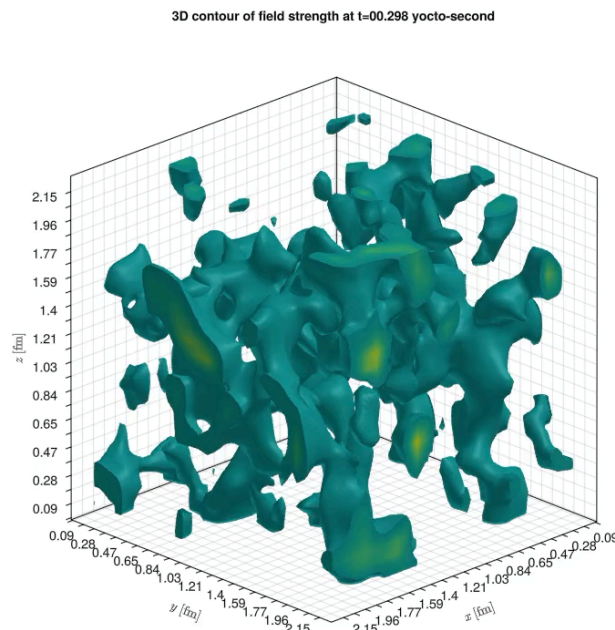
 - to understand **how stars are born and dead**

- properties of a dark matter candidate

etc.

We want to understand our universe from fundamental level!

What is our final goal for our research field?



What we want to solve?

- Improve efficiency of simulations to beyond our current numerical limitations

Restrictions (obstacles) to use ML on physics:

- Exactness & quantitative. Machine learning is an approximator
- **Gauge symmetry**, global symmetry is essential.
- Code. How can we make neural nets w/ HPC?

Gauge symmetric neural networks for MCMC

Configuration generation with machine learning is developing

Year	Group	ML	Dim.	Theory	Gauge sym	Exact?	Fermion?	Reference
2017	AT, Akinori Tanaka	RBM + HMC	2d	Scalar	-	No	No	arXiv: 1712.03893
2018	K. Zhou+	GAN	2d	Scalar	-	No	No	arXiv: 1810.12879
2018	J. Pawłowski +	GAN +HMC	2d	Scalar	-	Yes?	No	arXiv: 1811.03533
2019	MIT+	Flow	2d	Scalar	-	Yes	No	arXiv: 1904.12072
2020	MIT+	Flow	2d	U(1)	Equivariant	Yes	No	arXiv: 2003.06413
2020	MIT+	Flow	2d	SU(N)	Equivariant	Yes	No	arXiv: 2008.05456
2020	AT, Akinori Tanaka +	SLMC	4d	SU(N)	Invariant	Yes	Partially	arXiv: 2010.11900
2021	M. Medvidović+	A-NICE	2d	Scalar	-	No	No	arXiv: 2012.01442
2021	S. Foreman	L2HMC	2d	U(1)	Yes	Yes	No	
2021	AT+	SLHMC	4d	QCD	Covariant	Yes	YES!	
2021	L. Del Debbio+	Flow	2d	Scalar, O(N)	-	Yes	No	
2021	MIT+	Flow	2d	Yukawa	-	Yes	Yes	
2021	S. Foreman, AT+	Flowed HMC	2d	U(1)	Equivariant	Yes	No but compatible	arXiv: 2112.01586
2021	XY Jing	Neural net	2d	U(1)	Equivariant	Yes	No	
2022	J. Finkenrath	Flow	2d	U(1)	Equivariant	Yes	Yes (diagonalization)	arxiv: 2201.02216
2022	MIT+	Flow	2d, 4d	U(1), QCD	Equivariant	Yes	Yes	arXiv:2202.11712 +
2022	AT+	Flow	2d, 3d	Scalar		Yes		
...	...	...	...	...	...	...	...	...

and more +

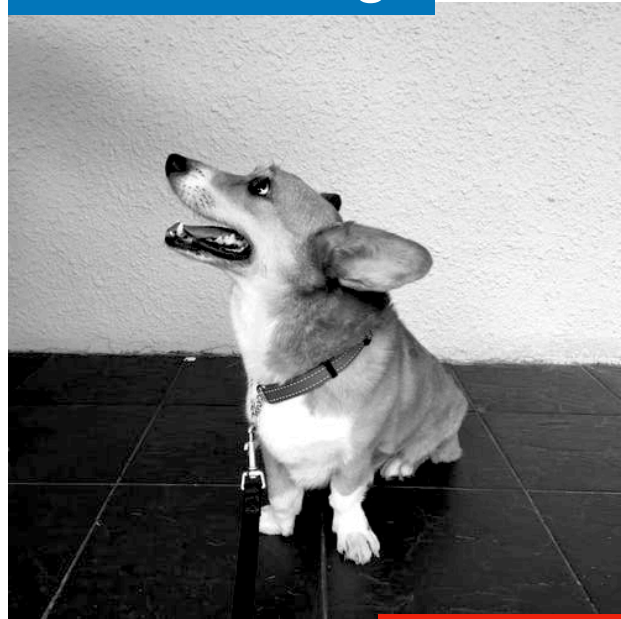
Symmetry is important

- Symmetry is a central concept in theoretical physics
- Lattice QCD fully uses symmetry, not only global symmetry but also gauge symmetry
- To use neural networks in QCD, we need gauge symmetric neural networks

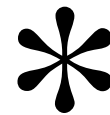
What is conv. neural networks?

Convolution layer = trainable filter

Filter on image

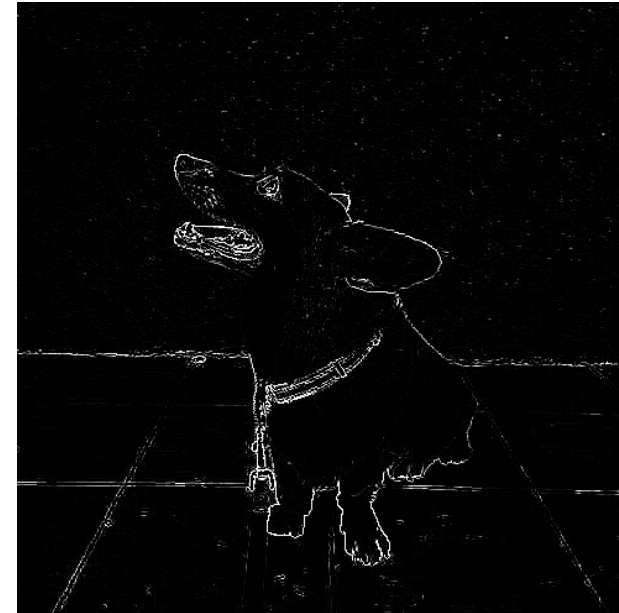


Laplacian filter



0	1	0
1	-2	1
0	1	0

=



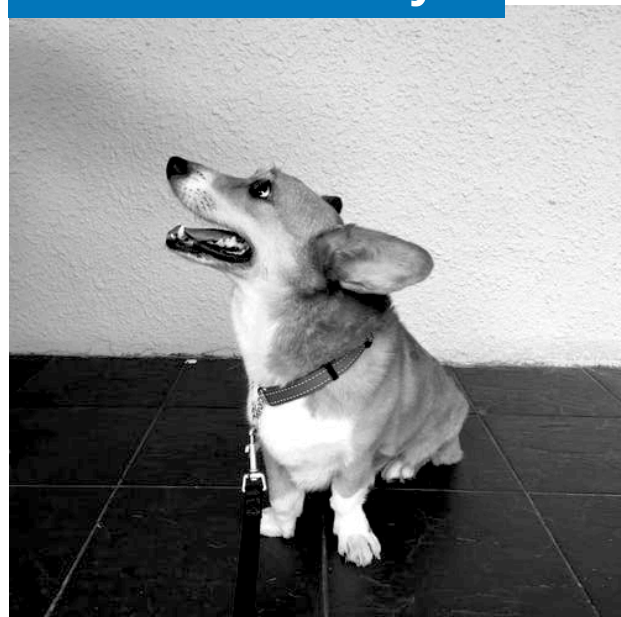
Edge detection

(Discretization of ∂^2)

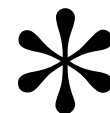
IMPORTANT: If inputs are shifted to right, outputs are shifted to right

= translationally equivariant (similar to covariance, operation just commute)

Convolution layer



Trainable filter



W_{11}	W_{12}	W_{13}
W_{21}	W_{22}	W_{23}
W_{31}	W_{32}	W_{33}



Edge detection

Smoothing
(Gaussian filter)

...

Fukushima, Kunihiro (1980)
Zhang, Wei (1988) + a lot!

Gaussian filter

1	2	1
2	4	2
1	2	1

$\frac{1}{16}$

This can be any filter which helps feature extraction
but still translationally equivariant!

Smearing

Smoothing improves global properties

Eg.

Coarse image



Numerical derivative is unstable

Gaussian filter

$$\frac{1}{16} \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 2 & 4 & 1 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$


Smoothened image



Numerical derivative is stable

We want to smoothen gauge configurations with keeping gauge symmetry

Two types:

APE-type smearing

Stout-type smearing

M. Albanese+ 1987
R. Hoffmann+ 2007
C. Morningster+ 2003

Smoothing with gauge symmetry, APE type

M. Albanese+ 1987
R. Hoffmann+ 2007

APE-type smearing

$$U_\mu(n) \rightarrow U_\mu^{\text{fat}}(n) = \mathcal{N} \left[(1 - \alpha) U_\mu(n) + \frac{\alpha}{6} V_\mu^\dagger[U](n) \right]$$

Covariant sum

Normalization $\mathcal{N}[M] = \frac{M}{\sqrt{M^\dagger M}}$ Or projection

$$V_\mu^\dagger[U](n) = \sum_{\nu \neq \mu} U_\nu(n) U_\mu(n + \hat{\nu}) U_\nu^\dagger(n + \hat{\mu}) + \dots$$

$V_\mu^\dagger[U](n) \& U_\mu(n)$ shows same transformation
 $\rightarrow U_\mu^{\text{fat}}[U](n)$ is as well

Schematically,

$$\Rightarrow = \mathcal{N} \left[(1 - \alpha) \rightarrow + \frac{\alpha}{6} \sum_\nu \left(\uparrow \rightarrow \downarrow + \downarrow \rightarrow \uparrow \right) \right]$$

We can promote this idea to neural network!

$$\Rightarrow = \mathcal{N} \left[w_1 \rightarrow + w_2 \sum_\nu \left(\uparrow \rightarrow \downarrow + \downarrow \rightarrow \uparrow \right) \right]$$

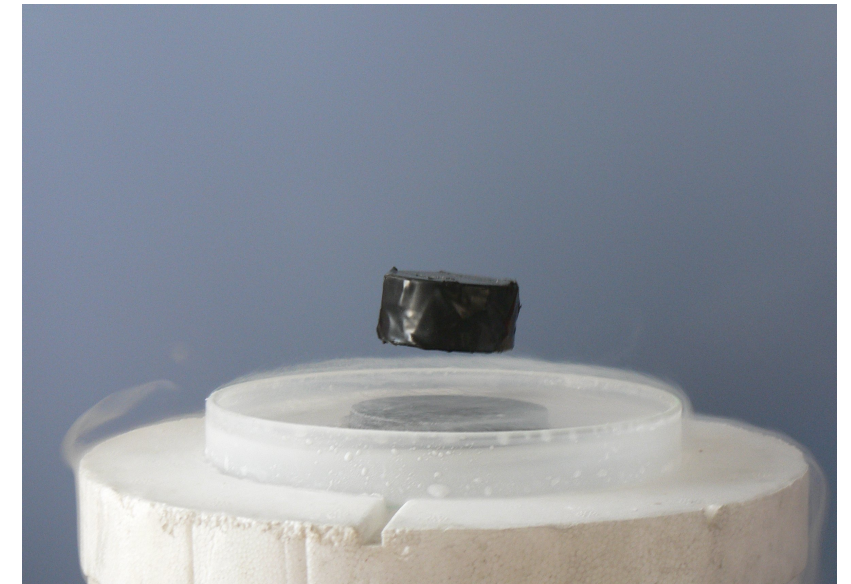
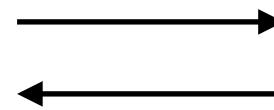
w_1, w_2 are weights, and we can train them!

Intuitive picture of these works

Find a effective action for simulations w/ neural nets

$$F - F_{\text{normal}} = \alpha |\psi|^2 + \frac{\beta}{2} |\psi|^4 + \frac{1}{2m} |(-i\hbar\nabla - 2e\mathbf{A})\psi|^2$$

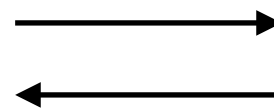
Effective action
~ Fit Ansatz (GL equation)



$$S_{\text{action}} = f_{\text{Neural net}}[U]$$

Effective action
~ Fit Ansatz (Neural net)
for lattice QCD?

More efficient MCMC?



$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \underline{DU} e^{-S_{\text{QCD}}[U]} \mathcal{O}(U)$$

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{sample}}} \sum_{k=1}^{N_{\text{sample}}} \mathcal{O}[U_k]$$

Covariant CNN: convolution type

To make a new effective action, we build it with existing quark action, by shifting mass larger. This would improve the condition number

$$\text{Loss} = \sum_{\text{data}} \left| S^{(\text{quark})}[U^{(\text{cov})}[U]; m = 0.4] - S^{(\text{quark})}[U; m = 0.3] \right|^2 \quad S^{(\text{quark})}[U; m] = \sum_n \phi^\dagger (D[U] + m)^{-1} \phi$$

Quark action

HMC = Molecular-dynamics + Metropolis-test

Molecular-dynamics
(NN effective action)

$$S^{(\text{quark})}[U; m] = \sum_n \phi^\dagger (D[U^{(\text{cov-net})}[U]] + m')^{-1} \phi$$

$m' > m$

Metropolis-test

$$S^{(\text{quark})}[U; m] = \sum_n \phi^\dagger (D[U] + m)^{-1} \phi$$

The gauge action is common

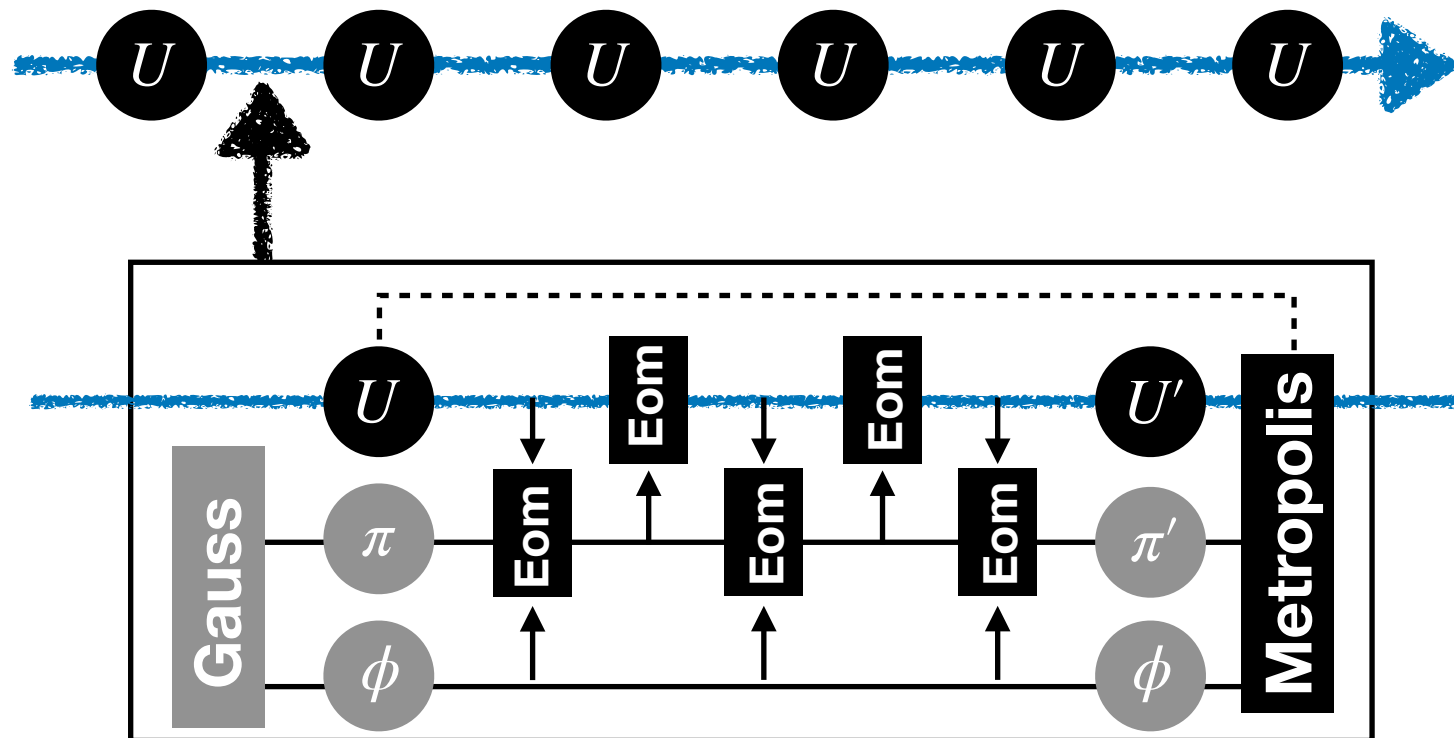
Exact action in Met-test ensures exactness
while we use neural network effective action in MD

Gauge covariant net& SLHMC

SLHMC for gauge system with dynamical fermions

arXiv: 2103.11965 and reference therein

HMC



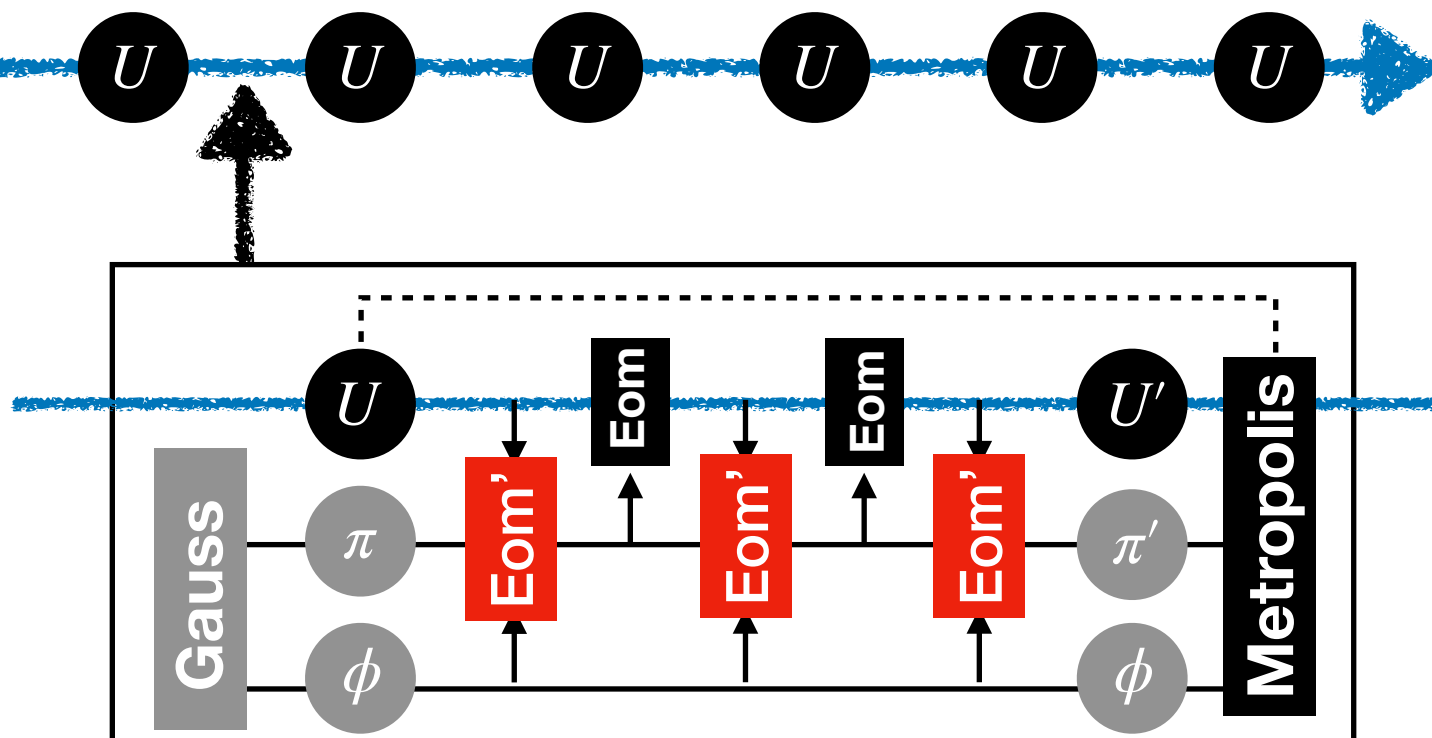
Eom **Metropolis**

Both use

$$H_{\text{HMC}} = \sum \pi^2 + S_g + S_f$$

Non-conservation of H cancels since the molecular dynamics is reversible

SLHMC



Metropolis

$$H = \sum \pi^2 + S_g + S_f[U]$$

Eom'

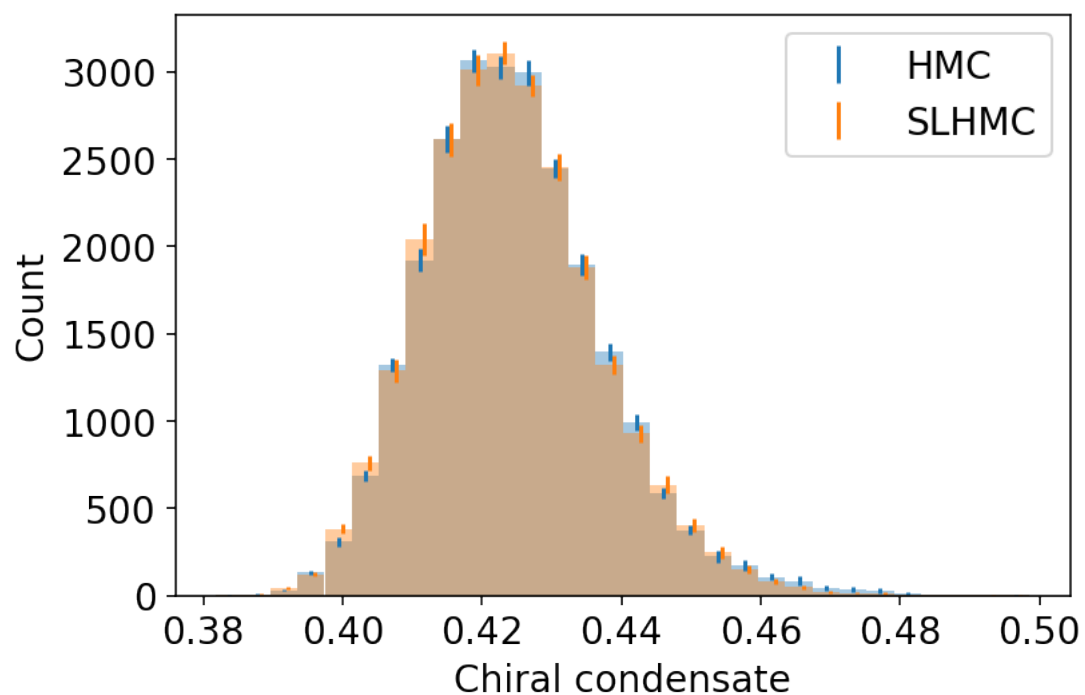
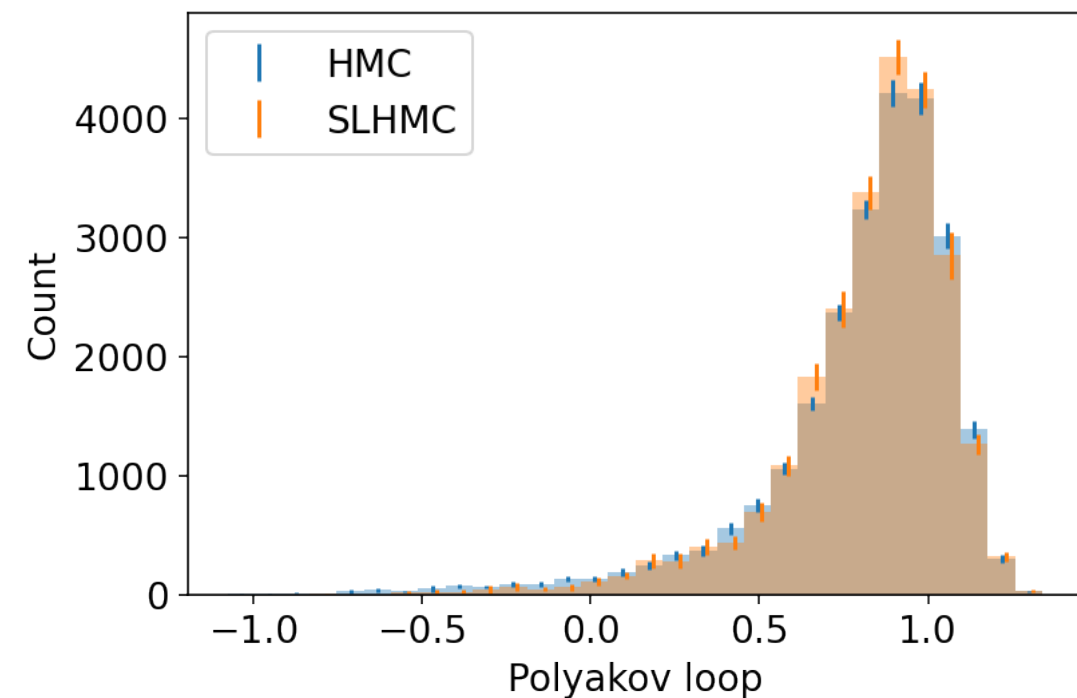
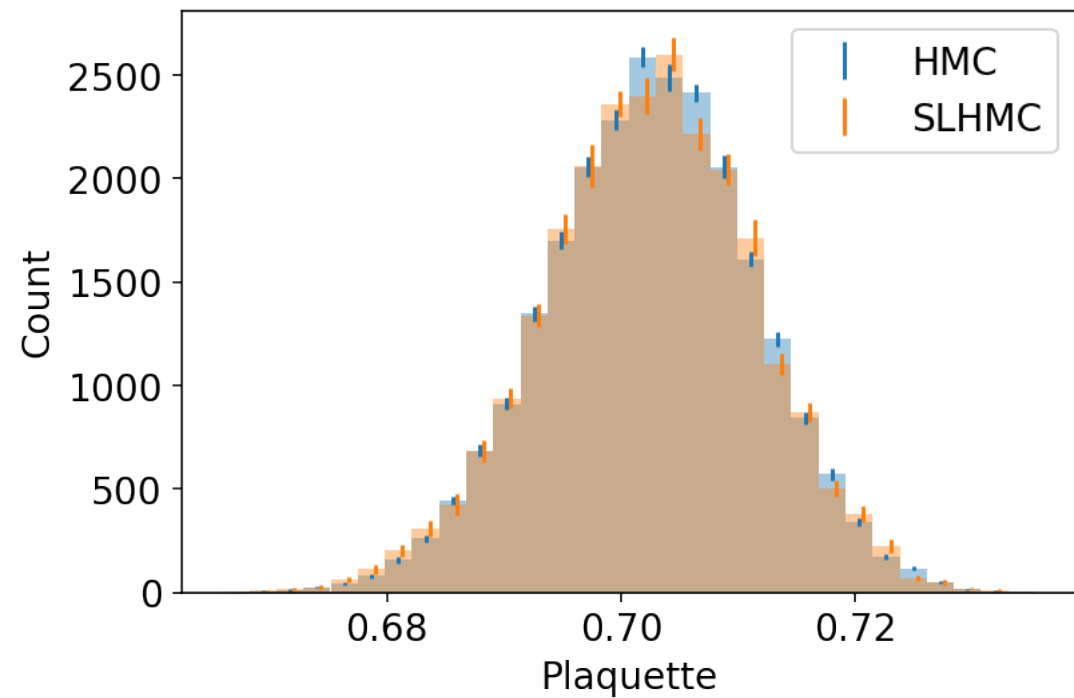
$$H = \sum \pi^2 + S_g + S_f[U^{\text{NN}}[U]]$$

Fermion action is approximated one but exact

Gauge covariant net& SLHMC

Results are consistent with each other

arXiv: 2103.11965 + statistics



Expectation value		
Algorithm	Observable	Value
HMC	Plaquette	0.7025(1)
SLHMC	Plaquette	0.7023(2)
HMC	Polyakov loop	0.82(1)
SLHMC	Polyakov loop	0.83(1)
HMC	Chiral condensate	0.4245(5)
SLHMC	Chiral condensate	0.4241(5)

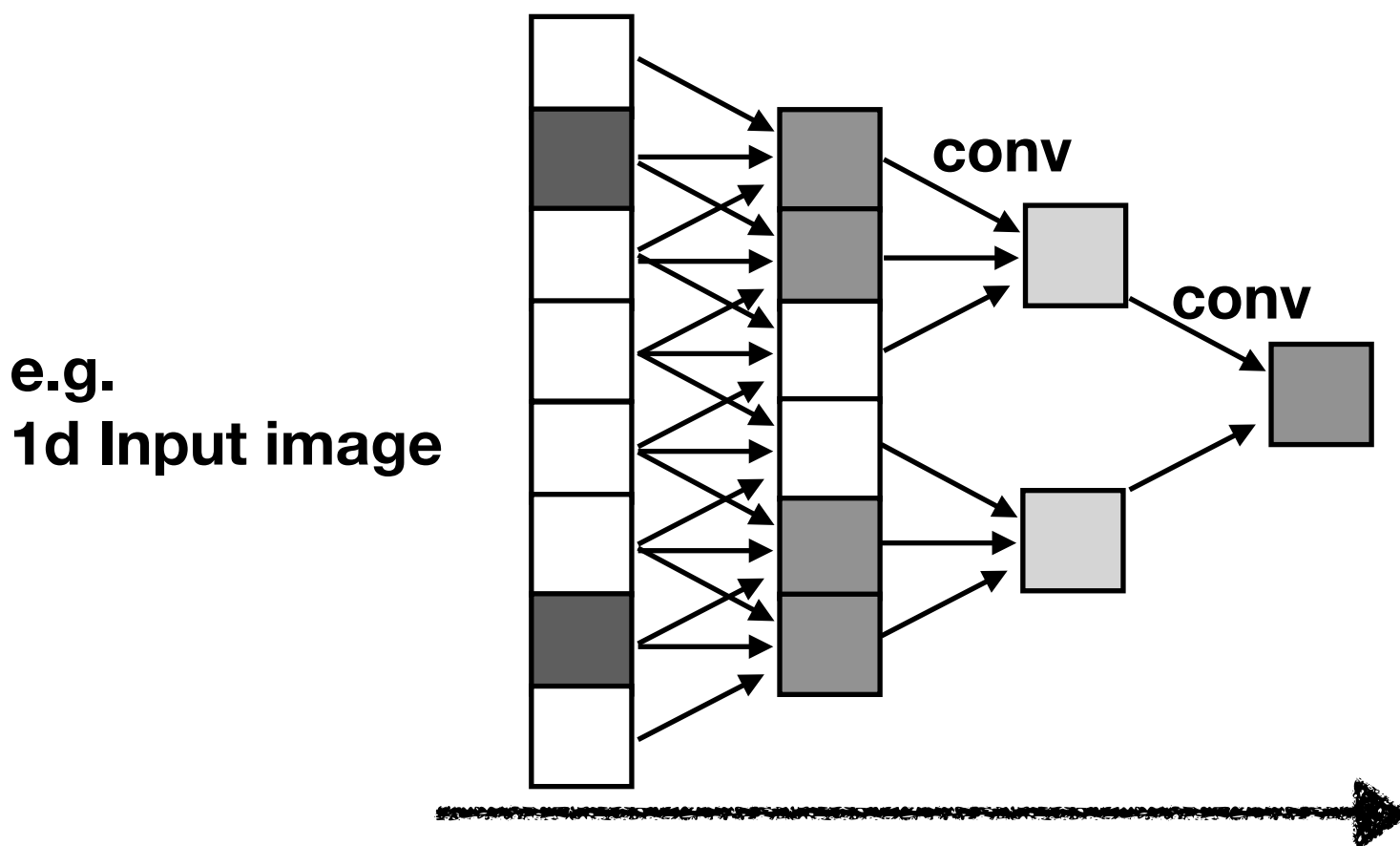
Gauge symmetric Transformer for LQCD/MCMC

Equivariance and convolution

Convolutional Neural network have been good job but local

2106.04554

conv ~ neural net with n-th nearest neighbor connections (local)



Long range correlation in input is captured by deep layers since operation is local

However, 1 step of **convolutional layer can pick up only local correlation** and representability of neural networks is limited. Global correlations are sometimes important.

How can we overcome these difficulties?

Configuration generation in LQCD

Akio Tomiya

Attention layer used in Transformers (GPT etc)

arXiv:1706.03762

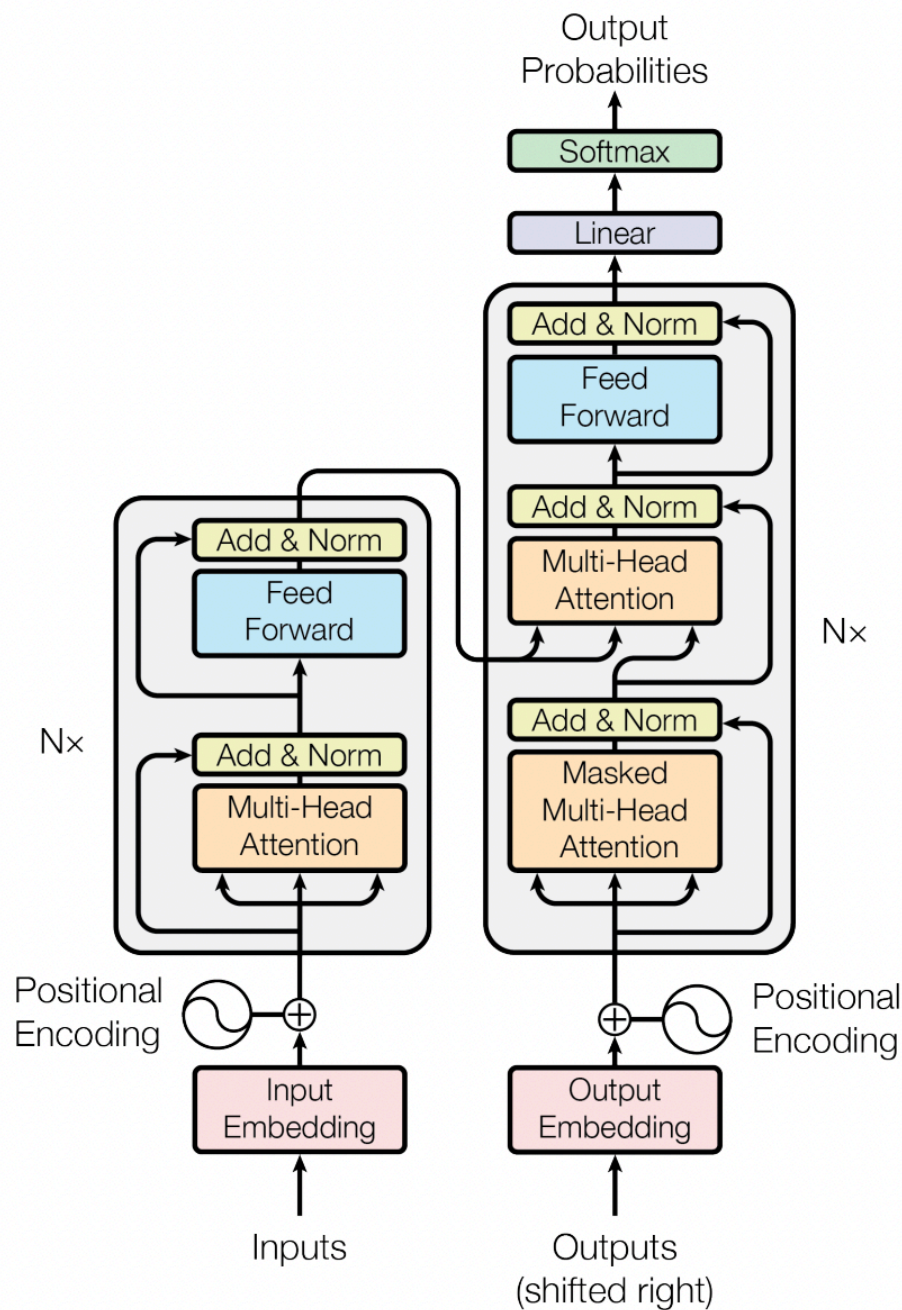


Figure 1: The Transformer - model architecture.



Attention layer (in transformer model) has been introduced in a paper titled **“Attention is all you need”** (1706.03762) State of the art architecture of language processing.

Attention layer is essential.

Configuration generation in LQCD

Akio Tomiya

Attention layer can capture non-local correlations

arXiv:1706.03762

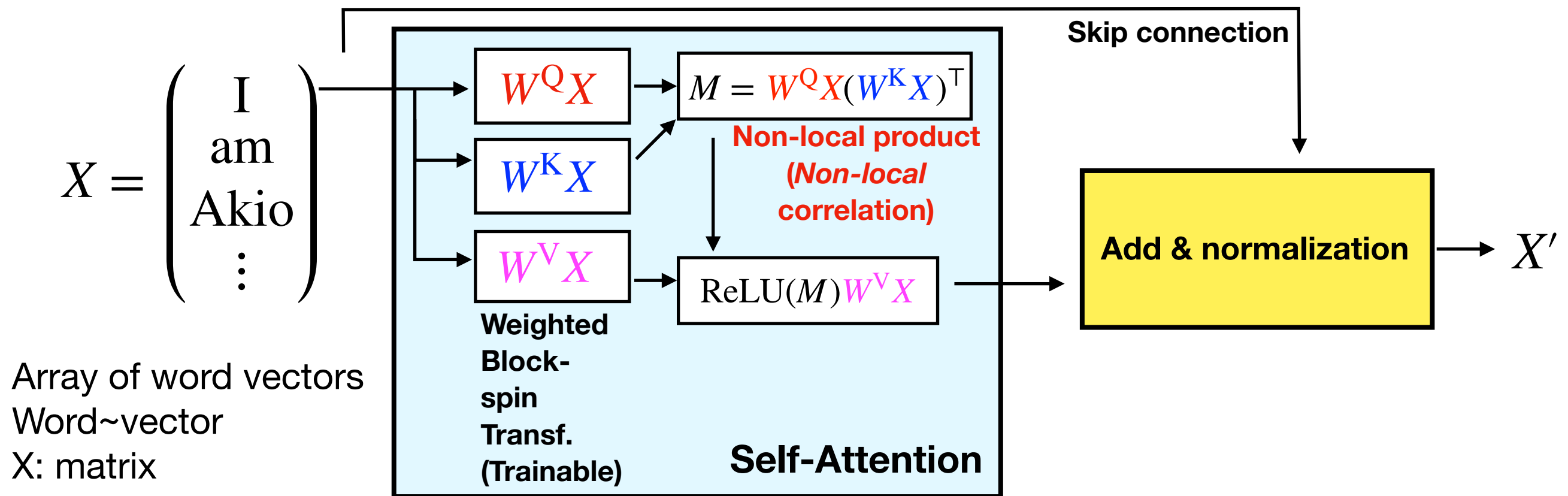
Modifier in language can be non-local

Eg. I am **Akio Tomiya** living in Japan, **who** studies machine learning and physics

In physics terminology, this is **non local correlation**.

The attention layer enables us to treat non-local correlation

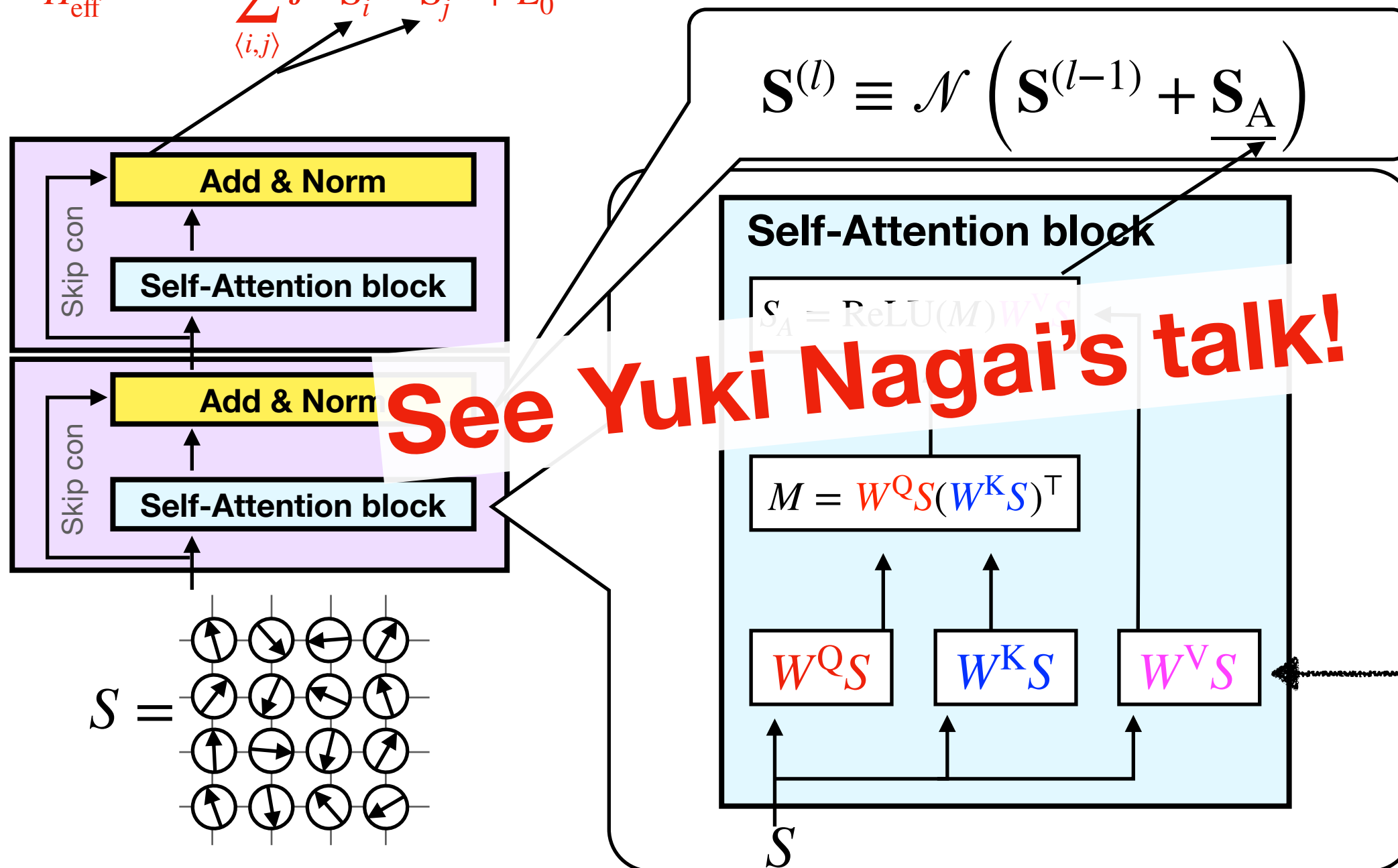
Simplified version of Attention/Transformer



These can be repeated

We can construct effective hamiltonian with output of Attention layer
because **“output of Attention = smeared fields with non-local correlation”**

$$H_{\text{eff}}^{\text{Linear}} = - \sum_{\langle i,j \rangle} J^{\text{eff}} \mathbf{S}_i^{\text{eff}} \cdot \mathbf{S}_j^{\text{eff}} + E_0$$



Smeared fields
Rot. equivariant
Trsl. equivariant
trainable!

Smearing
Rot. equivariant
Trsl. equivariant
trainable!

See Yuki Nagai's talk!

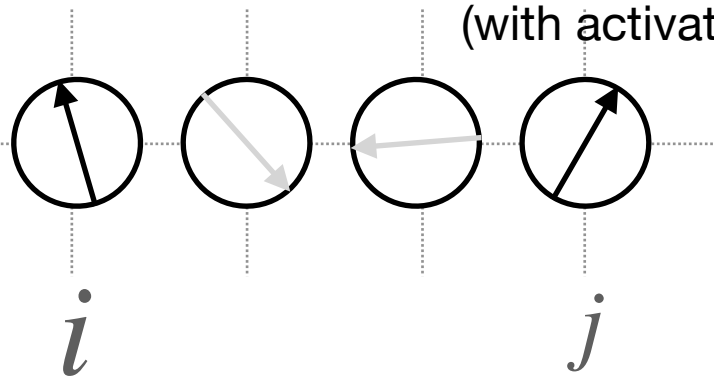
Attention matrix in transformer ~ correlation function (with block-spin transformed spin)

-> we replace it with “correlation function of links” in a **covariant** way

Attention for spins

$$a_{ij} \sim \vec{S}_i \cdot \vec{S}_j$$

(with activation)

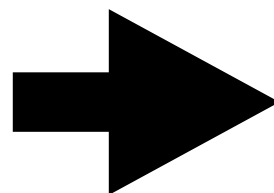


invariant
under **global** O(3)

$$a_{ij} \sim (R\vec{S})_i^\top R\vec{S}_j = \vec{S}_i^\top \vec{S}_j$$

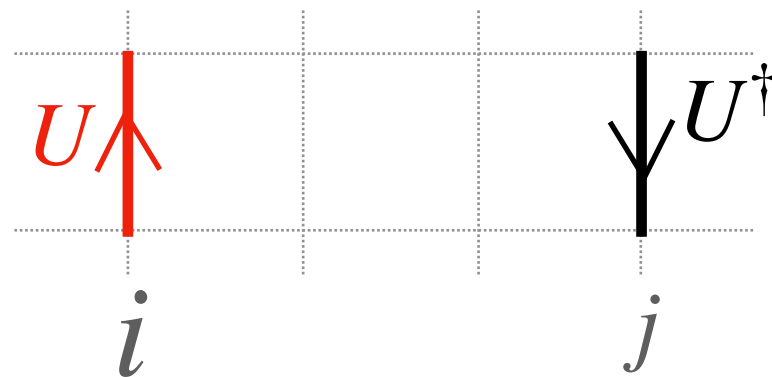
In total, output is covariant

2310.13222 AT+, 2306.11527 AT+



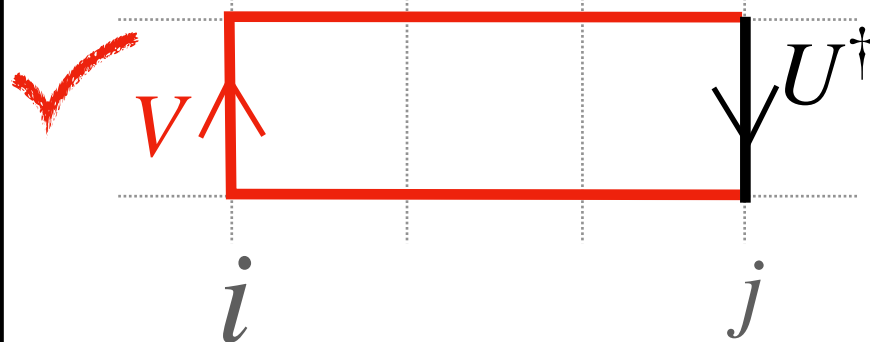
Gauging

X $a_{i\mu,j\mu} \sim \text{Re tr } U_\mu(i)U_\mu^\dagger(j)$



not invariant
(cannot be used)

Lattice gauge covariant attention



invariant under
local SU(N)

$$a_{i\mu,j\mu} \sim \text{Re tr } V_\mu(i)U_\mu^\dagger(j) \quad (\text{with activation})$$

In total, output is covariant

CASK: Covariant Transformer

CASK for SU(2), SU(3) gauge theory + fermions

Akio Tomiya



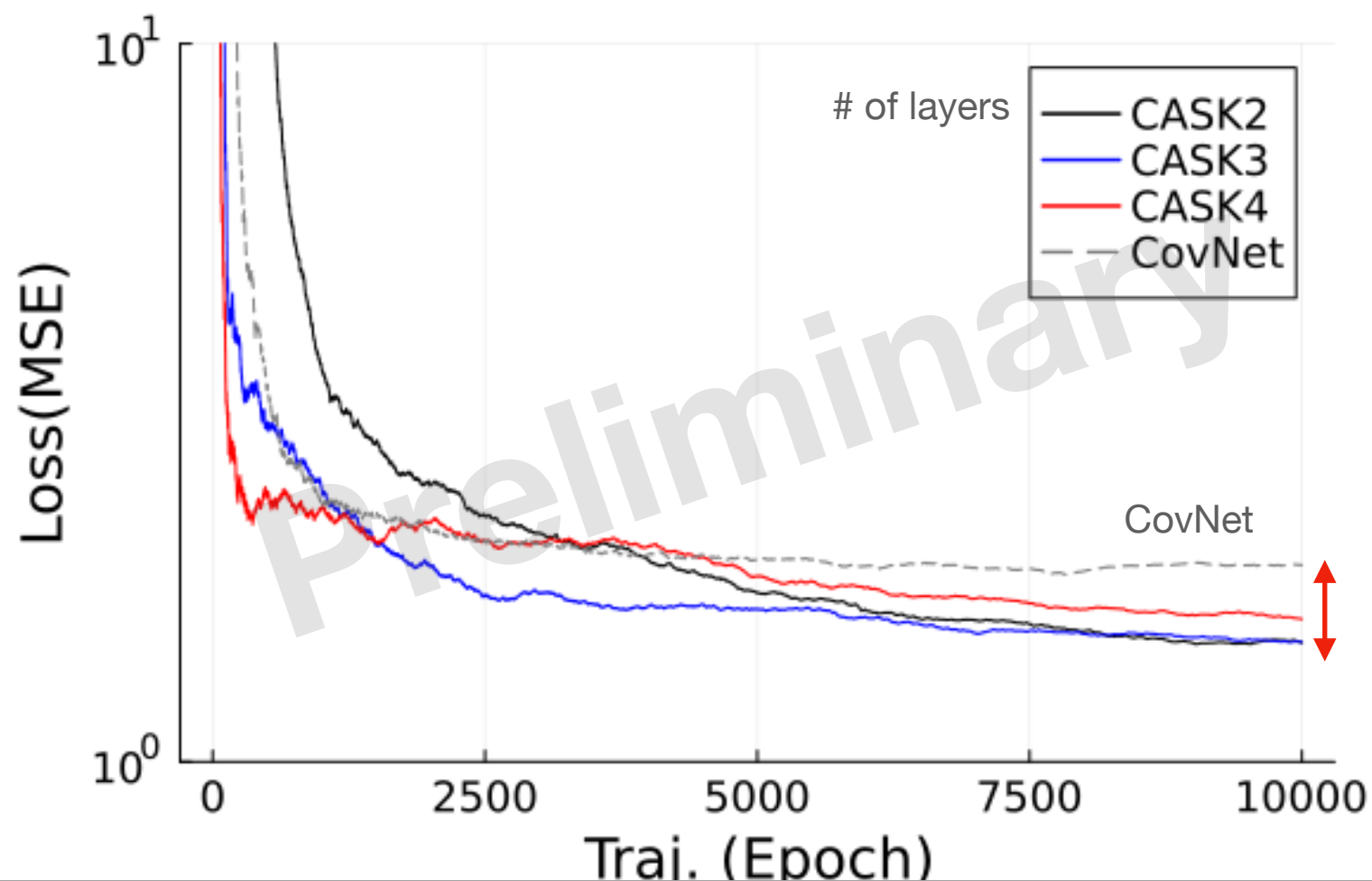
arXiv: 2501.16955

Comparison Covariant convolution (CovNet) and Covariant transformer (CASK)

$$U^{(\text{NN-out})} = \begin{cases} U_{\mu}^{(\text{CovNet})} = g_{\theta}^{(\text{CovNet})} U_{\mu}^{(\text{in})} & \text{CovNet (convolution based, baseline)} \\ U_{\mu}^{(\text{CASK})} = g_{\theta}^{(\text{CASK})} U_{\mu}^{(\text{in})} & \text{CASK (transformer based)} \end{cases}$$

$$\text{Loss} = \sum_{\text{data}} \left| S^{(\text{quark})}[U^{(\text{NN-out})}; m = 0.4] - S^{(\text{quark})}[U; m = 0.3] \right|^2 \quad S^{(\text{quark})}[U; m] = \sum_n \phi^{\dagger} (D[U] + m)^{-1} \phi$$

Energy function



- Dynamical simulation
- $L^4 = 4^4$, SU(2), $ma = 0.3$
33% larger mass in MD
- CASK has better expressibility than CovNet (Covariant CNN)
- SU(3) works as well

Performant code with NN

Other codes, motivations

Julia is the best for LQCD + ML

- Big success of lattice QCD. Quantitative and non-perturbative
 - **Machine learning? We need test applicability of it!**
- A code for “machine learning + lattice QCD”?
 - Lattice QCD is expensive. Fortran/C++ with parallelization
 - e.g. $O(10^9) \times O(10^9)$ matrix operation for $L^4 = 100^4$
 - Python is used for machine learning. But *SLOW*!
We can use libraries written in C++, but debugging is hard
- **Julia language supports both of performance and AutoGrad. No stress from “two languages”**



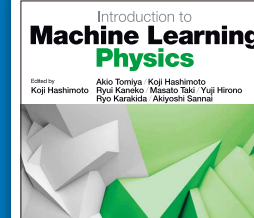
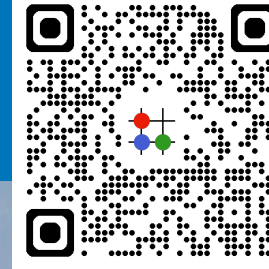
AT & Y. Nagai

37

Summary

AI/code for quantum fields

Code



<http://bit.ly/3JrEjls>

Akio Tomiya

This talk is based on
JPSJ 86, 063001 (2017)
JPSJ 94, 031006 (2025)

- **Phase transition in Ising model**
 - It can be seen from weights of neural net. It can be determined without prior knowledge
- **Gauge covariant convolution for gauge theory**
 - It works but might be inefficient if there are long range correlations induced by fermions
- **Gauge covariant transformer for gauge theory**
 - It works well. It shows scaling law? (need more systematic study)
- **Code: Julia works both for HPC and ML, so best choice for this field**

