Machine Learning Topological Sector of SU(3) Yang-Mills Theory

Masakiyo Kitazawa

(Osaka University)

with Takuya Matsumoto, Yasuhiro Kohno

MK, Kohno, Matsumoto, PTEP 2020, accepted for publication
[arXiv:1909.06238 [hep-lat]]

Is machine learning technique meaningfully applicable to theoretical physics?

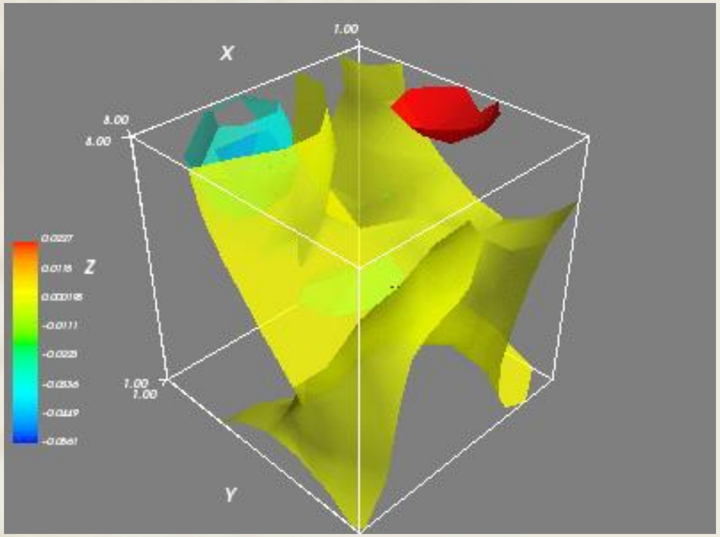
- ML does not provide us with “understanding”.
- Theoretical physicists must have a better solution for problems that ML can deal with.
- 100% accuracy is never reached by the ML.

Topological Charge in YM Theory

$$Q = \int d^4 x q(x) \quad : \text{integer}$$

$$q(x) = -\frac{1}{32\pi^2} \mathrm{tr}[F_{\mu\nu} \tilde{F}_{\mu\nu}]$$

q(x) in SU(3) YM,
β=5.8, 8^4, t/a^2=2.0



- **Interests / applications**
  - Instantons
  - Axial U(1) anomaly
  - Axion cosmology
  - Topological freezing

Topology

- Topology
  - properties of an object that are preserved under continuous deformations

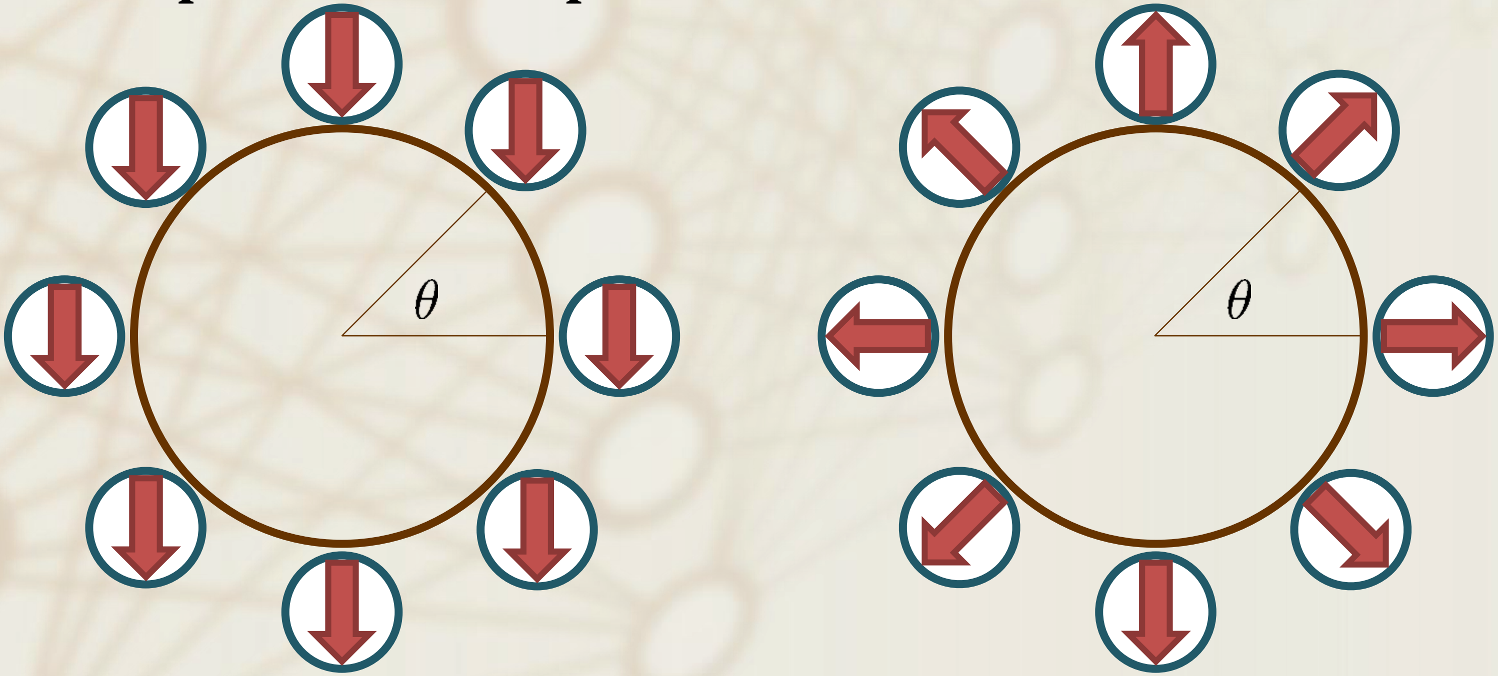
from Wikipedia

Topology

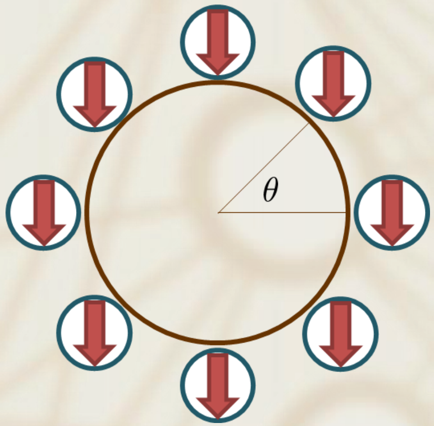
- Topology
  - properties of an object that are preserved under continuous deformations

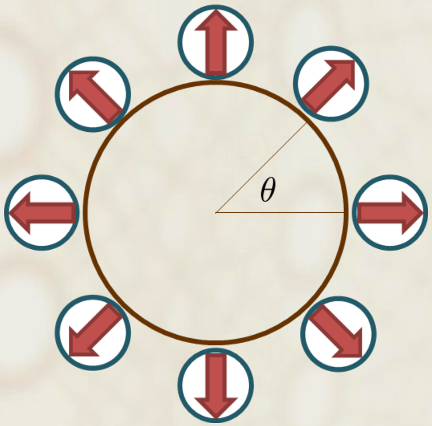
from Wikipedia

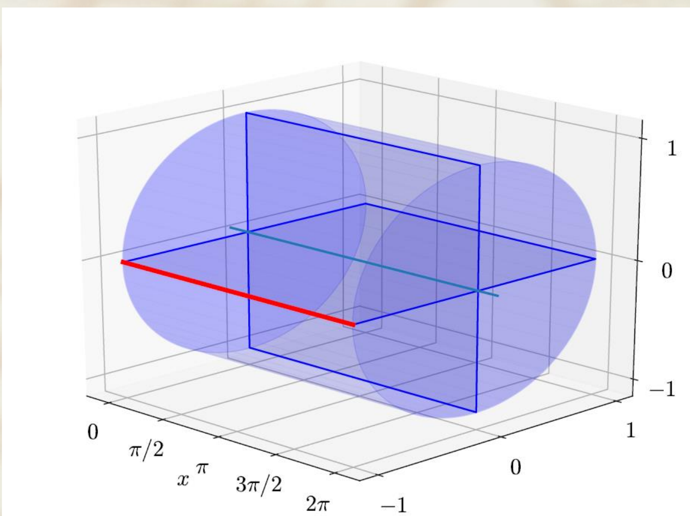
- Example: 1-dim. space



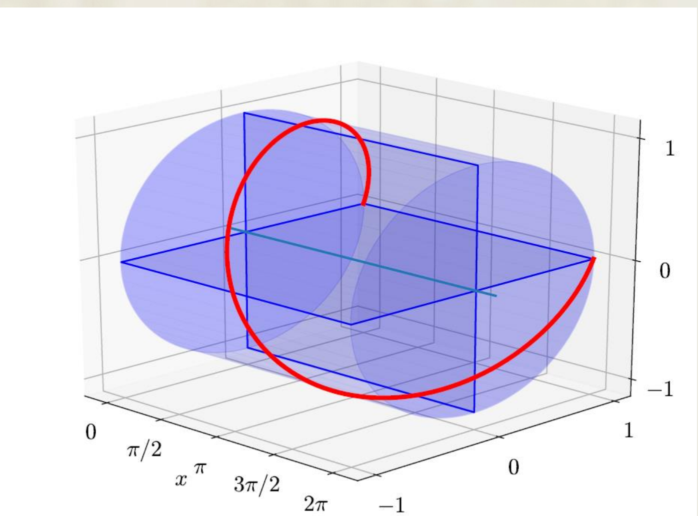
$S_1 \to S_1$







winding number
n=0



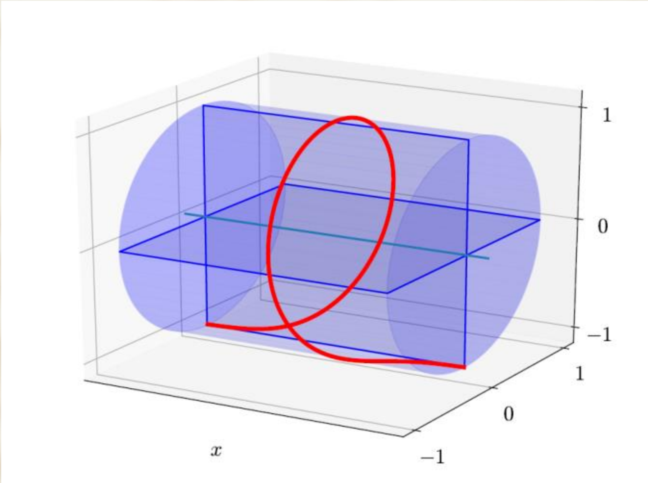
winding number
n=1

Sine-Gordon Model in 1+1D

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - (1 - \cos\phi)$$

- **"kink" solution**

$$\phi(x) = 4\tan^{-1}\exp(x - x_0)$$



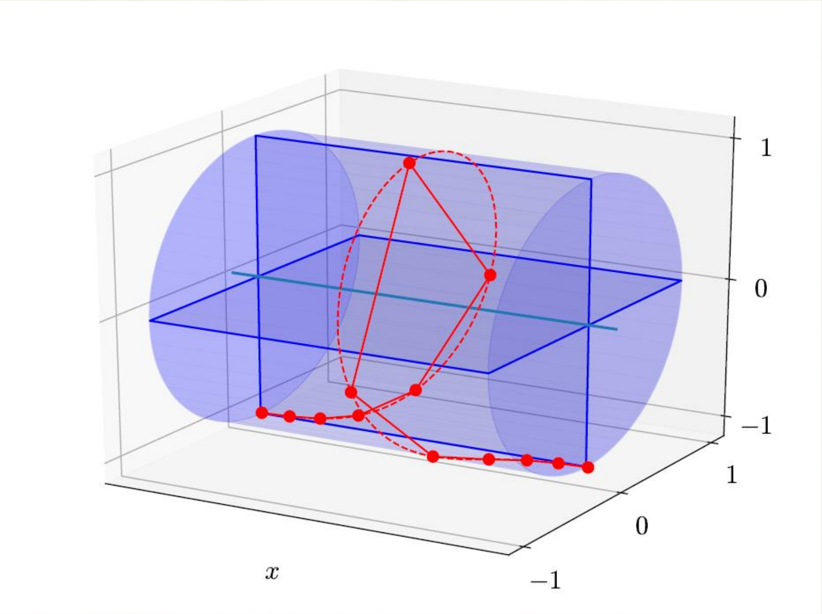
- winding number n=1
- nonzero energy
- topologically stable

$$n = \int dx \partial_x \phi$$

- **multi-"kink"** solution is also possible.

Lattice Theory & Topology

$$\mathcal{L} = \sum_{\mathbf{n}} \sum_{\mu} \frac{1}{2}((\phi_{\mathbf{n}-\hat{\mu}} - 2\phi_{\mathbf{n}} + \phi_{\mathbf{n}+\hat{\mu}}) - \sum_{\mathbf{n}}(1 - \cos\phi_{\mathbf{n}})$$



$$n = \int dx \partial_x \phi$$

$$n = \sum_x (\phi_{x+1} - \phi_x)$$

- Different *n* are connected continuously.
- “Topological sector” becomes obscure on the lattice.
- Topological sectors recover in the continuum limit.

Topology in 4D YM Theory

- SU(2) gauge field on |r|→∞ sphere in Euclid space
  - Mapping: S_3 (4D sphere)→S_3 (Gauge Tr. U(x))
  - S_3→S_3 has a non-trivial topology

topological charge

$$Q = \int d^4x q(x), \quad q(x) = \frac{1}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[F_{\mu\nu}F_{\rho\sigma}]$$

- **Instanton**

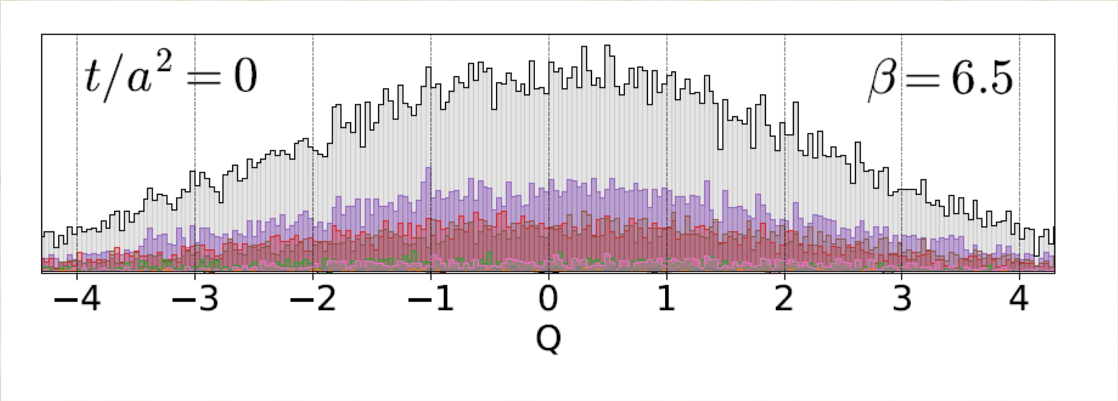
$$q(x) = \frac{6}{\pi^2}\frac{\rho^4}{((x-x_0)^2+\rho^2)^4}$$

- classical solution of YM
- winding number $n = 1$
- nonzero action

Topology on the Lattice

- A naïve definition of Q

$$Q = \int d^4 x q(x), \quad q(x) = \frac{1}{4}\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[F_{\mu\nu}F_{\rho\sigma}]$$



Q is not an integer, but distributes continuously.

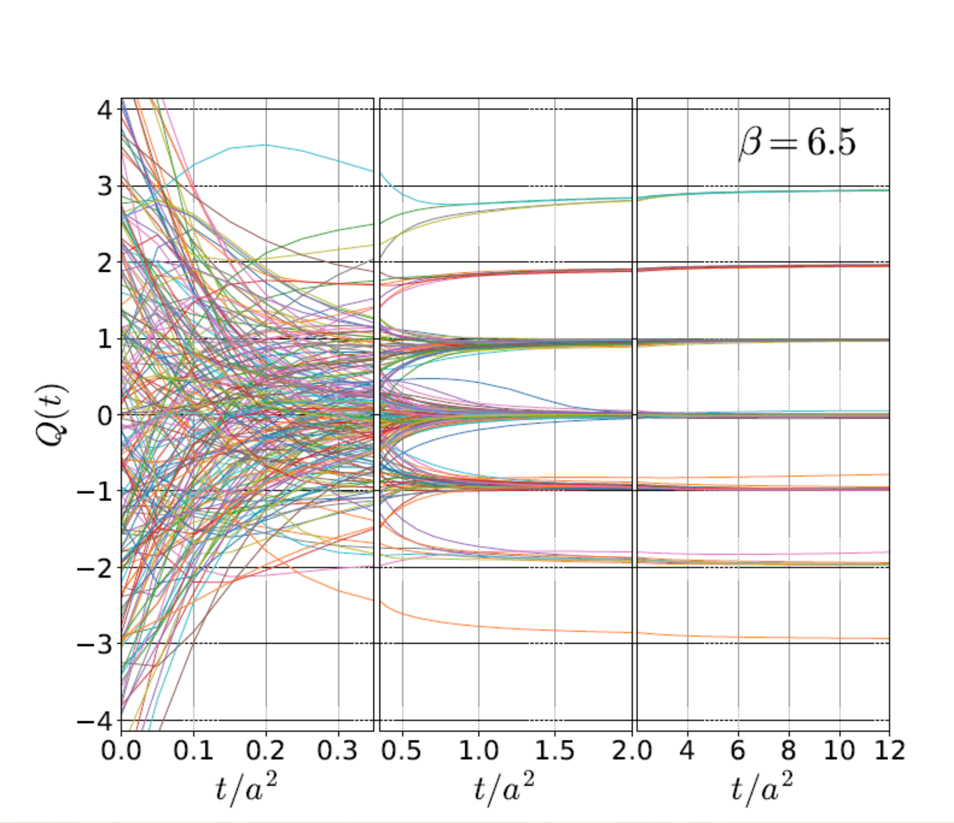
- Distinct topological sectors on sufficiently fine lattices

Luscher, 1981

Topology on the Lattice

- Definitions of Q on the lattice:
  - fermionic: Atiyah-Singer index theorem
  - **gluonic**: $q(x)$ after smoothing
    - cooling, smearing
    - **gradient flow**

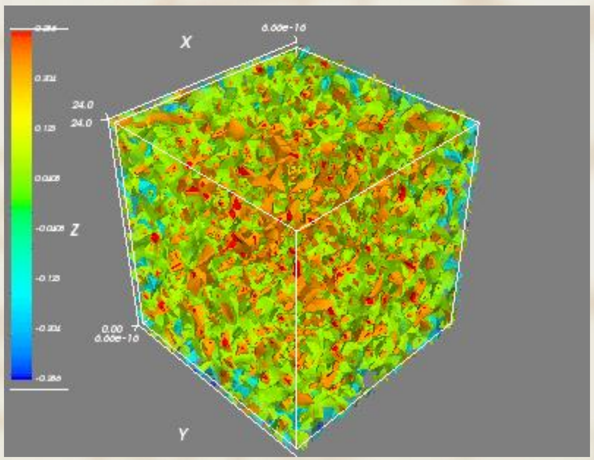
Luscher, Weisz, 2011



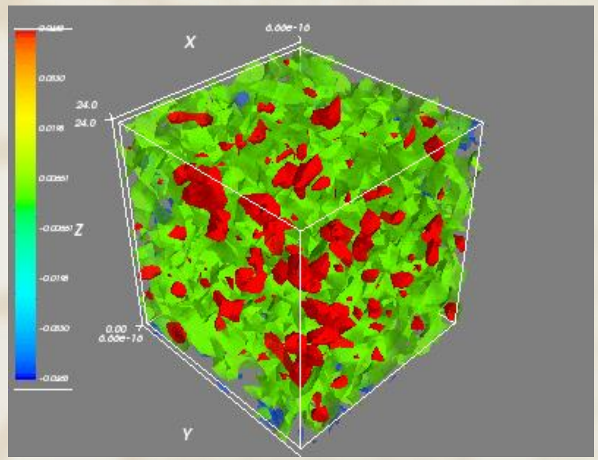
- Good agreement b/w various definitions
- **Faster algorithm is desirable!**

$q(x)$ at Nonzero Flow Time

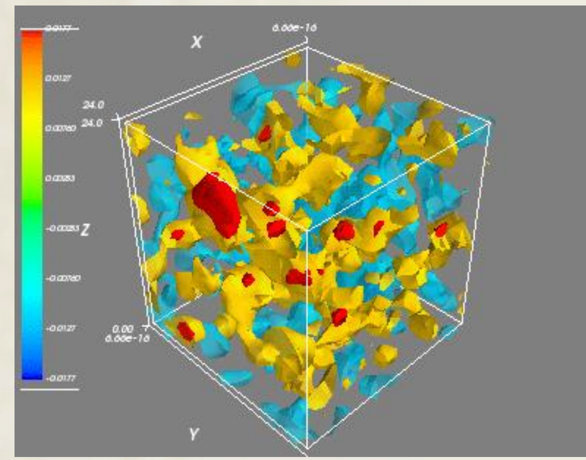
$t/a^2 = 0.1$



$t/a^2 = 0.2$



$t/a^2 = 0.3$



Field becomes smoother for larger t.

Topological Freezing

- Lattice Monte-Carlo simulation → gauge update
- Auto-correlation length of Q becomes longer as lattice spacing becomes finer.

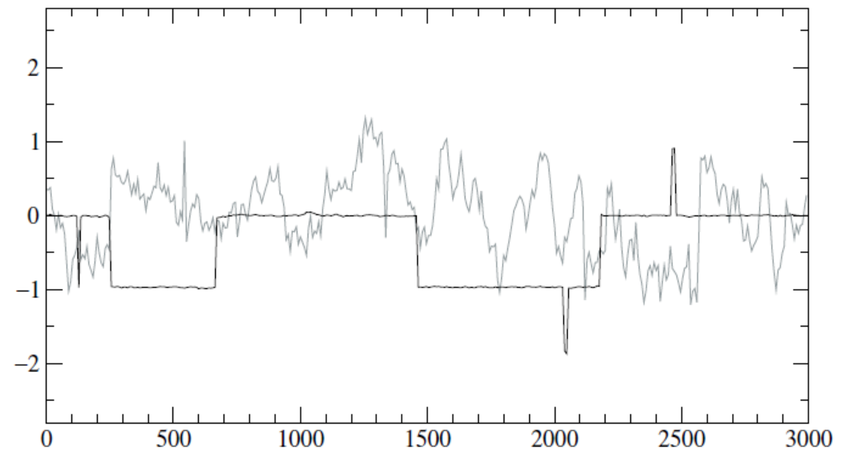
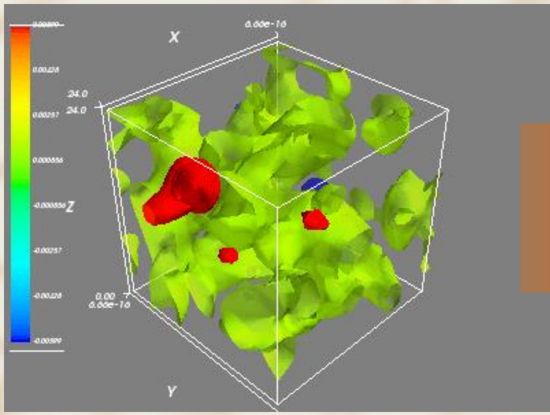


Fig. 3. History of the topological charge in three-flavour QCD on a 36×24^3 lattice with SF (black line) and open-SF (grey line) boundary conditions, plotted as a function of the simulation time in units of molecular-dynamics time (see subsect. 5.2 for further details).

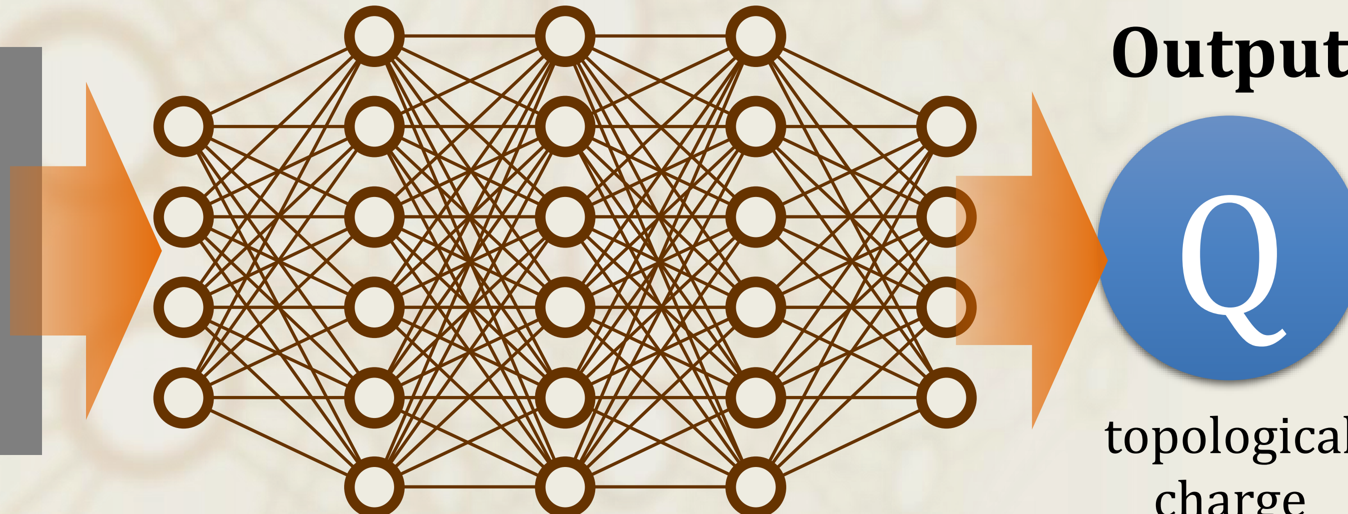
M. Lüscher.(2014)

Neural Network

Input: $q(x)$



4-dimensional field

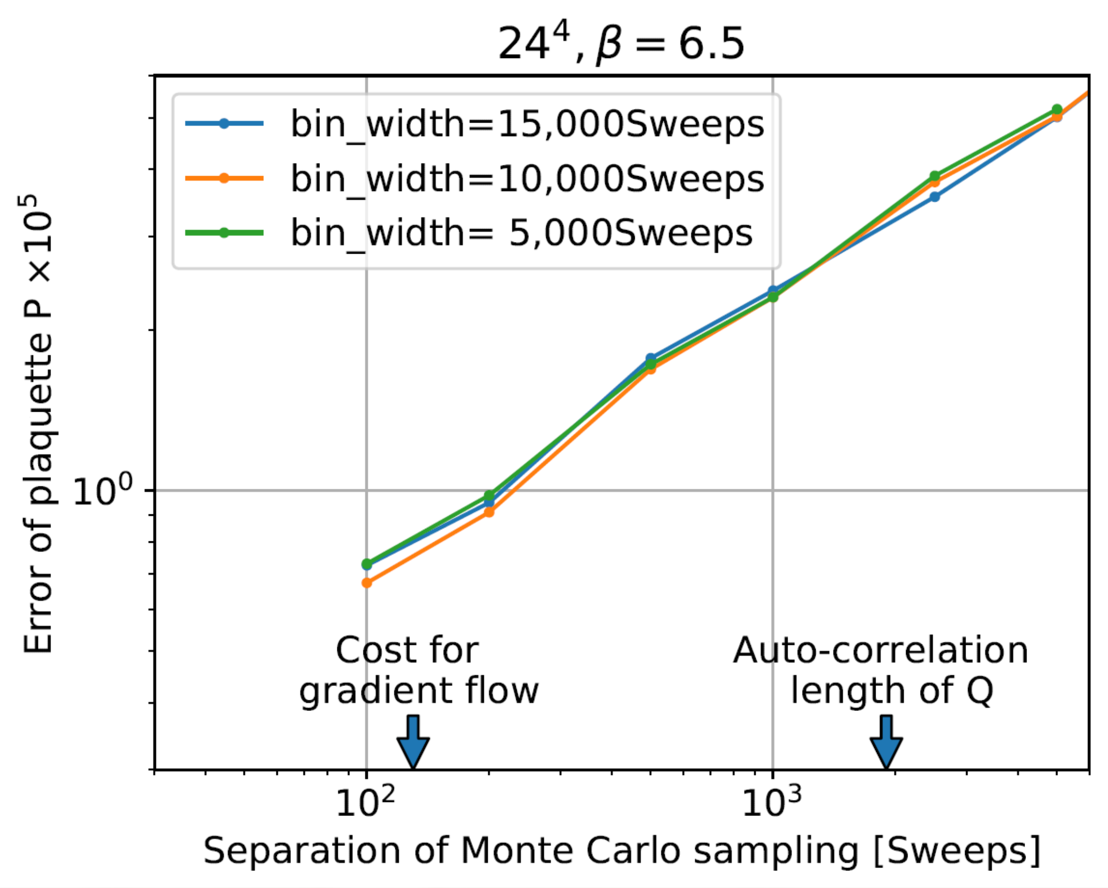


Output

topological charge

- ☐ **Capture "instanton"-like structure?**
- ☐ **Acceleration of the analysis of Q?**

Cost to Measure Q with GF



- Frequent measurements shorter than τ_Q is effective in reducing statistical errors.

slide by
T. Matsumoto

Neural Network

- Approximate arbitrary functions

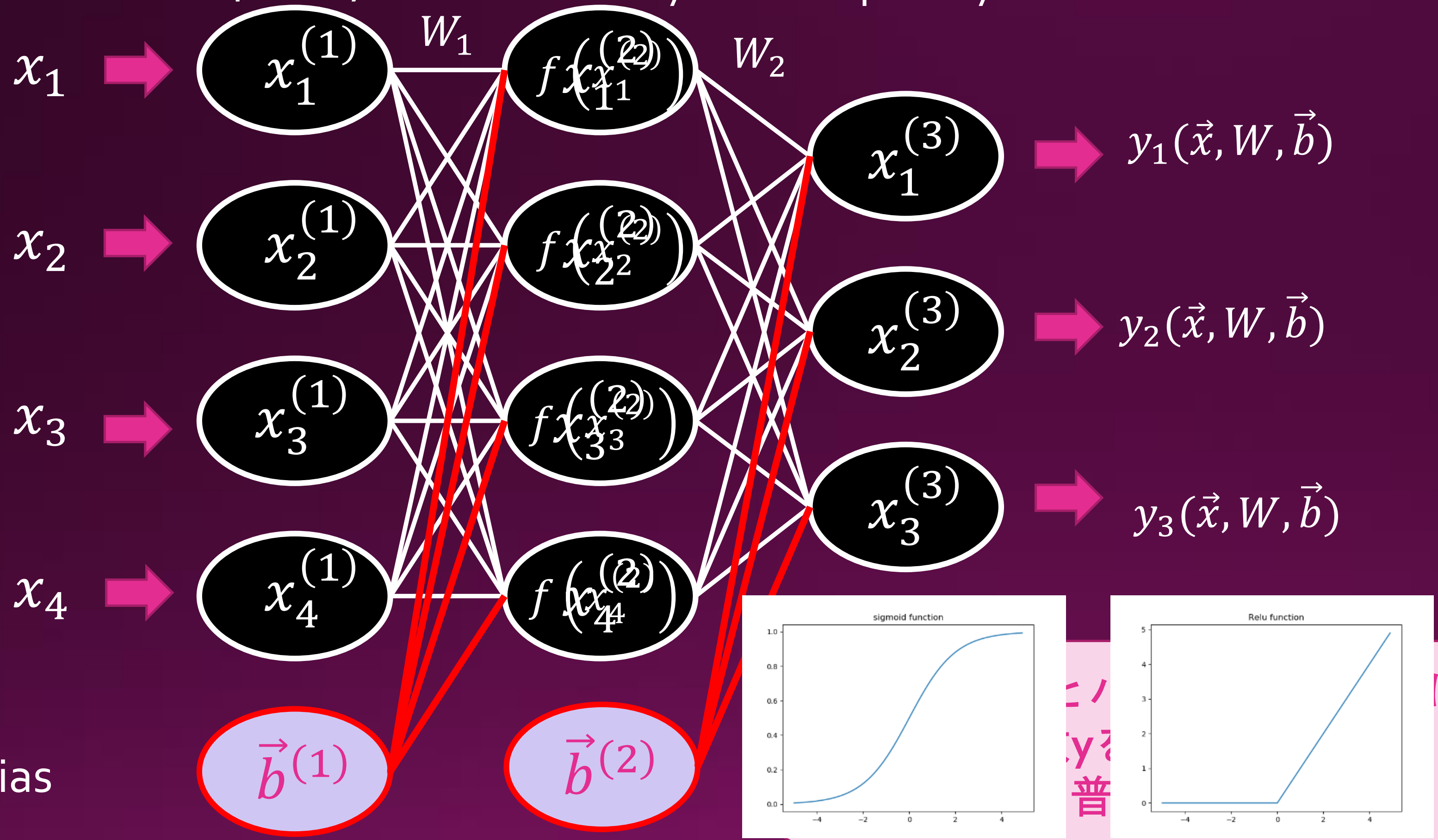
$$\vec{x} \Rightarrow \boxed{\text{NN}} \Rightarrow \vec{y}(x)$$

- Supervised Learning:
  Evaluate errors b/w outputs of NN and y(x)

<u>Tune parameters in the NN to minimize the error</u>
<u>→ "Good" function y(x) is obtained.</u>

Mechanism

slide by
T. Matsumoto



$f(x)$: activation func.

$\vec{x}^{(2)} = f(W_1 \vec{x}^{(1)} + \vec{b}^{(1)})$

$\vec{x}^{(3)} = f(W_2 \vec{x}^{(2)} + \vec{b}^{(2)})$

$f(x) = \frac{1}{1+e^{-x}}$

sigmoid

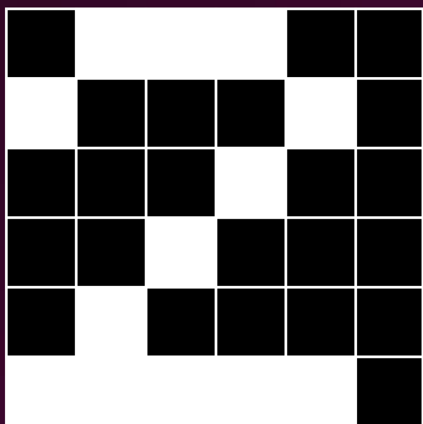
$f(x) = \begin{cases} x & for \quad x > 0 \\ 0 & for \quad x \leq 0 \end{cases}$

ReLU

Convolutional NN (CNN)

slide by
T. Matsumoto

- Example: number 2



0	1	1	1	0	0
1	0	0	0	1	0
0	0	0	1	0	0
0	0	1	0	0	0
0	1	0	0	0	0
1	1	1	1	1	0

0	0	1
0	1	0
1	0	0

filter

畳み込み層の出力

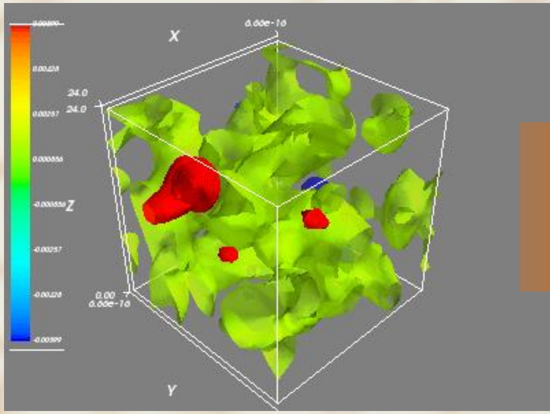
1	1	0	2
0	0	3	0
0	3	0	0
3	1	1	1

フィルターとどれだけ似ているかの度合いを抽出している

CNNではフィルターのパラメータを更新していく

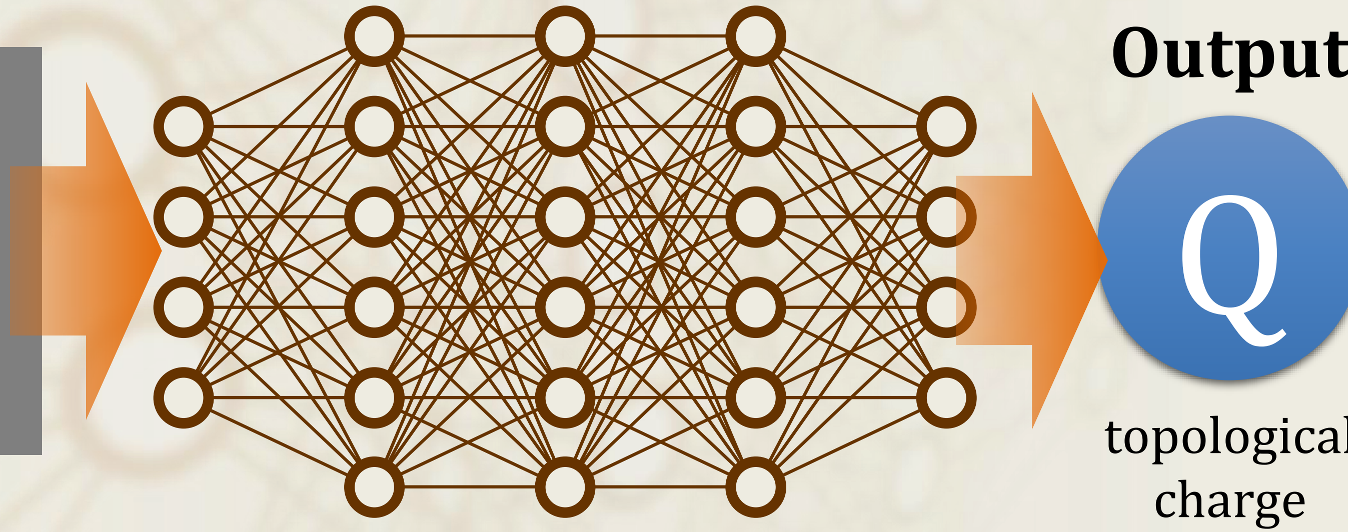
Neural Network

Input: $q(x)$



4-dimensional field

Output

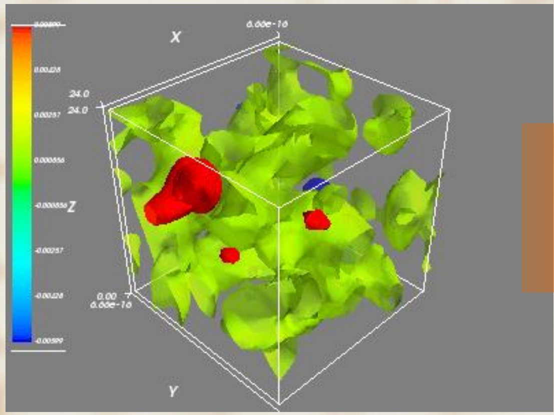


topological charge

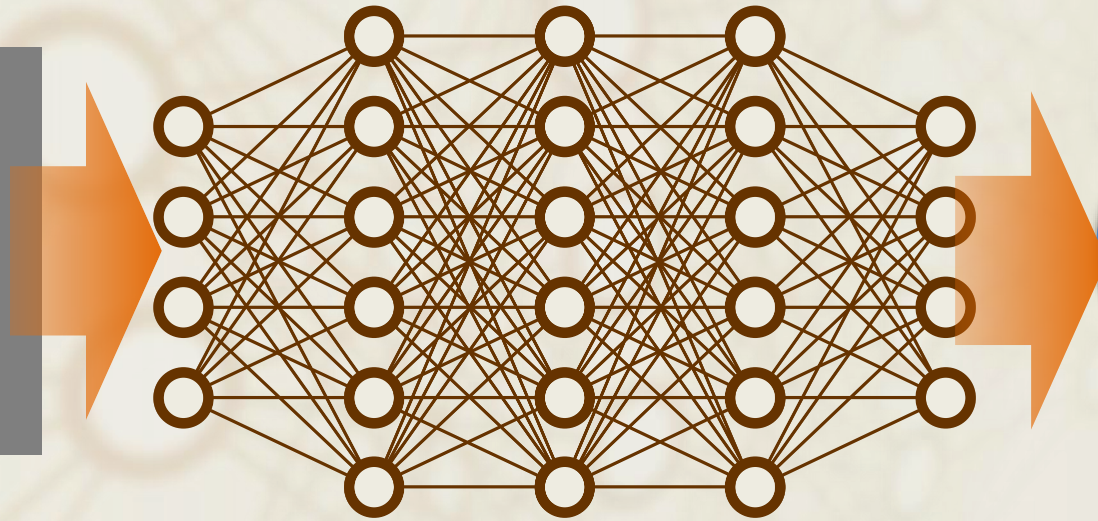
- **Capture “instanton”-like structure?**
- **Acceleration of the analysis of Q?**

Neural Network

Input: $q(x)$



4-dimensional field



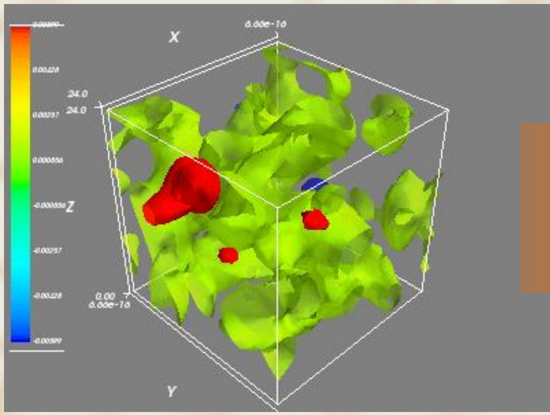
Output

topological charge

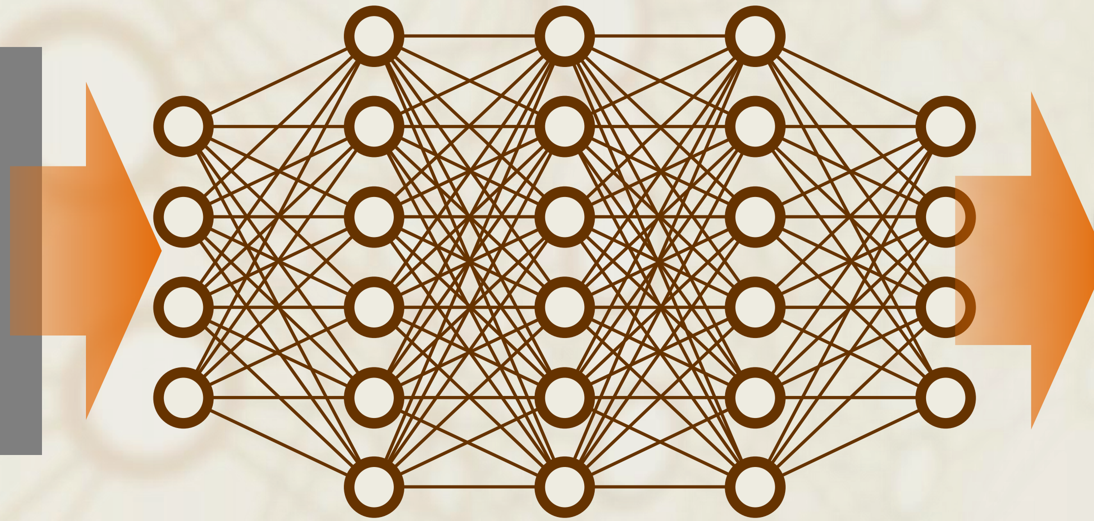
- ☐ **Capture “instanton”-like structure?** 
- ☐ **Acceleration of the analysis of Q?** 

Neural Network

Input: $q(x)$



4-dimensional field



Output

topological charge

- **Why $q(x)$ rather than link variables?**
  - to reduce the input data
  - to skip teaching SU(N) and gauge invariance

Lattice Setting

- **SU(3) Yang-Mills**
- **Wilson gauge action**
- 2 lattice spacings with **same physical volume**
- $LT_c \sim 0.63$
- $\langle Q^2 \rangle \simeq 1.1$

- **Gradient flow** for smoothing

β	N^4	N_{conf}
6.2	16^4	20,000
6.5	24^4	20,000

20,000 confs. in total

Training: 10,000

Validation: 5,000

Test: 5,000

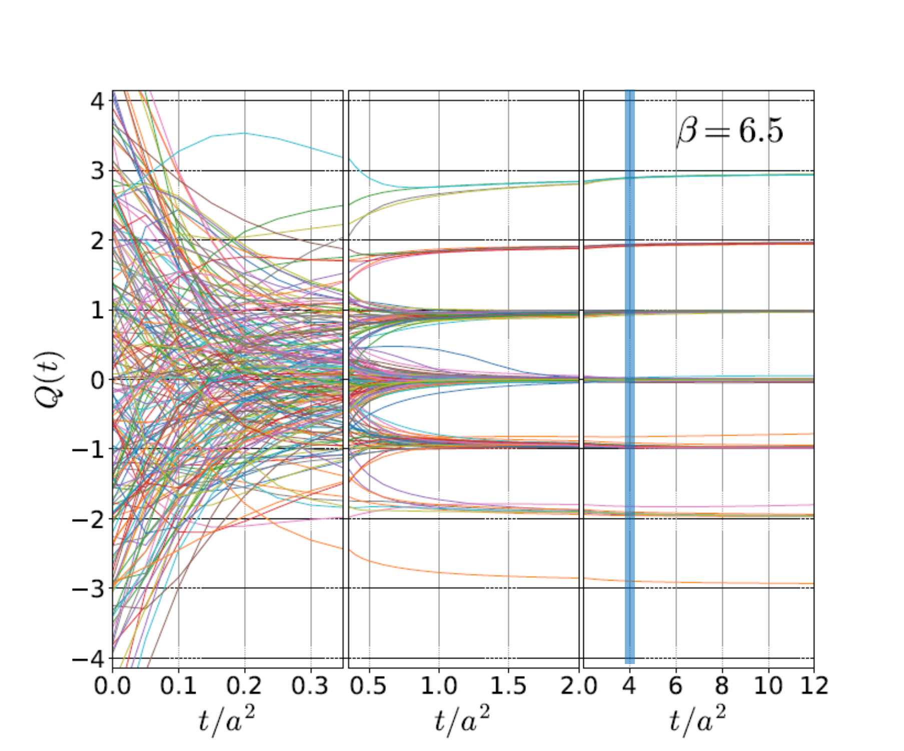
distribution of *Q*

Q	-5	-4	-3	-2	-1	0	1	2	3	4	5
$\beta = 6.2$	2	17	235	1325	4571	7474	4766	1352	240	18	0
$\beta = 6.5$	0	5	105	1080	4639	8296	4621	1039	202	13	0

Neural Network Setting

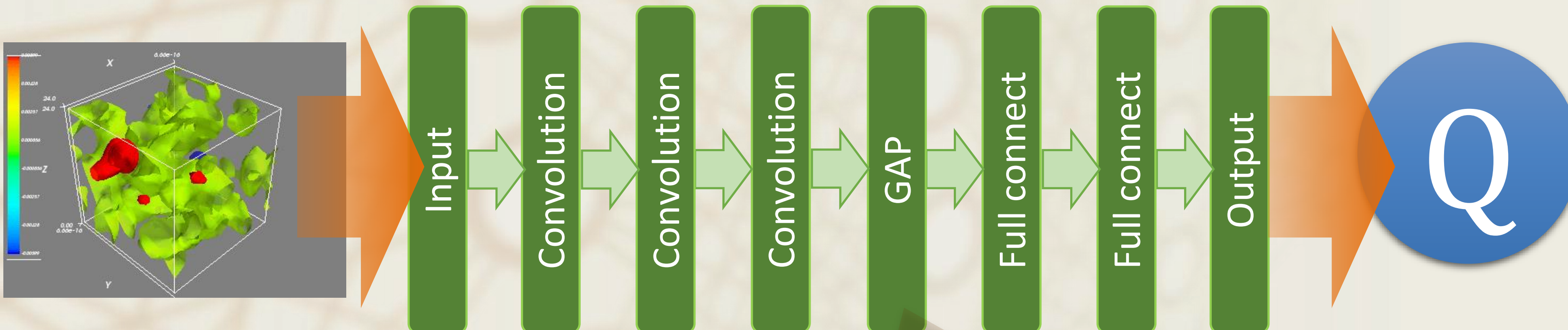
- convolutional neural network by **CHAINER framework**
- supervised learning
- convolutional layer: 4-dim., periodic BC
- regression analysis / round off to obtain integer
- activation: logistic

- answer of Q
  - $Q(t)$ @ $t/a^2 = 4.0$
  - round off



Trial 1: Topol. Charge Density

- Input: $q(x)$ in 4-dim space
- Data reduction to 8^4 (average pooling)

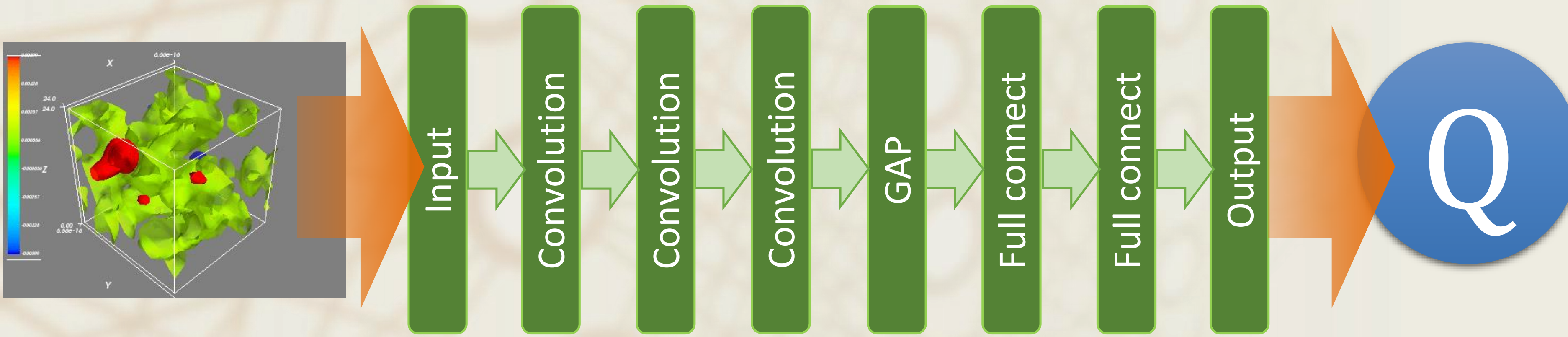


layer	filter size	output size	activation
input	-	$8^d \times N_{\rm ch}$	-
convolution	3^d	$8^d \times 5$	logistic
convolution	3^d	$8^d \times 5$	logistic
convolution	3^d	$8^d \times 5$	logistic
global average pooling	8^d	1×5	-
full connect	-	5	logistic
full connect	-	1	-

GAP=Global Average Pooling
Translational invariance is respected in this NN.

Trial 1: Topol. Charge Density

- Input: $q(x)$ in 4-dim space
- Data reduction to 8^4 (average pooling)



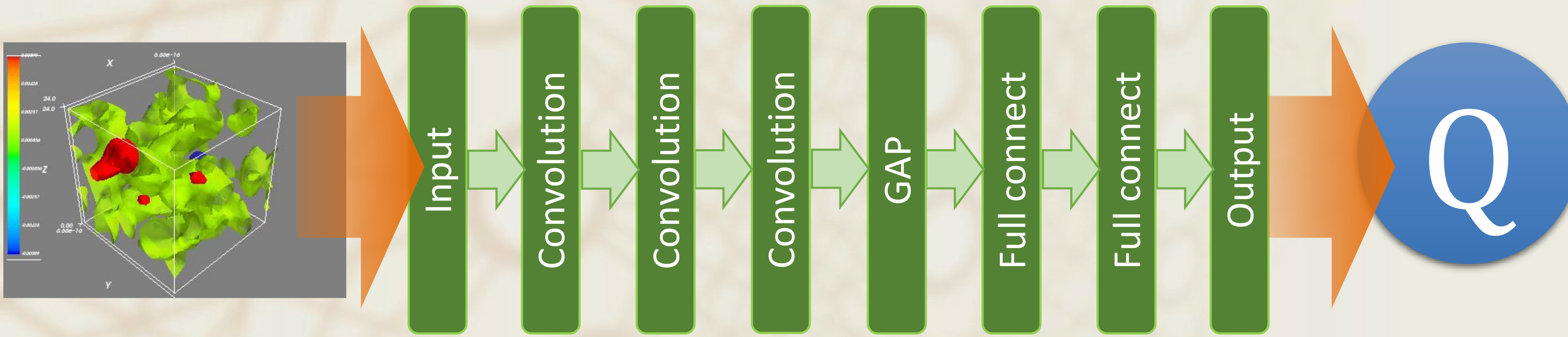
- **Result: best accuracy for β=6.2: 37.0%**

Accuracy of each topological sector (%)

Q	-4	-3	-2	-1	0	1	2	3	4	total
$t/a^2=0$	0	0	0	0	37.2	0	0	0	0	37.0

Trial 2: Topol. Density @ $t > 0$

- Input: $q(x, \boldsymbol{t})$ in 4-dim space at nonzero flow time
- Data reduction to 8^4 (average pooling)



Accuracy of each topological sector (%)

Q	-4	-3	-2	-1	0	1	2	3	4	total
$t/a^2=0$	0	0	0	0	37.2	0	0	0	0	**37.0**
$t/a^2=0.1$	0	0	31.6	39.1	41.4	38.9	19.0	0	0	**40.1**
$t/a^2=0.2$	0	40.0	46.4	53.8	55.9	52.3	48.1	50.0	0	**55.2**
$t/a^2=0.3$	0	91.3	72.9	76.3	79.0	74.8	68.1	70.0	50.0	**77.6**

Benchmark

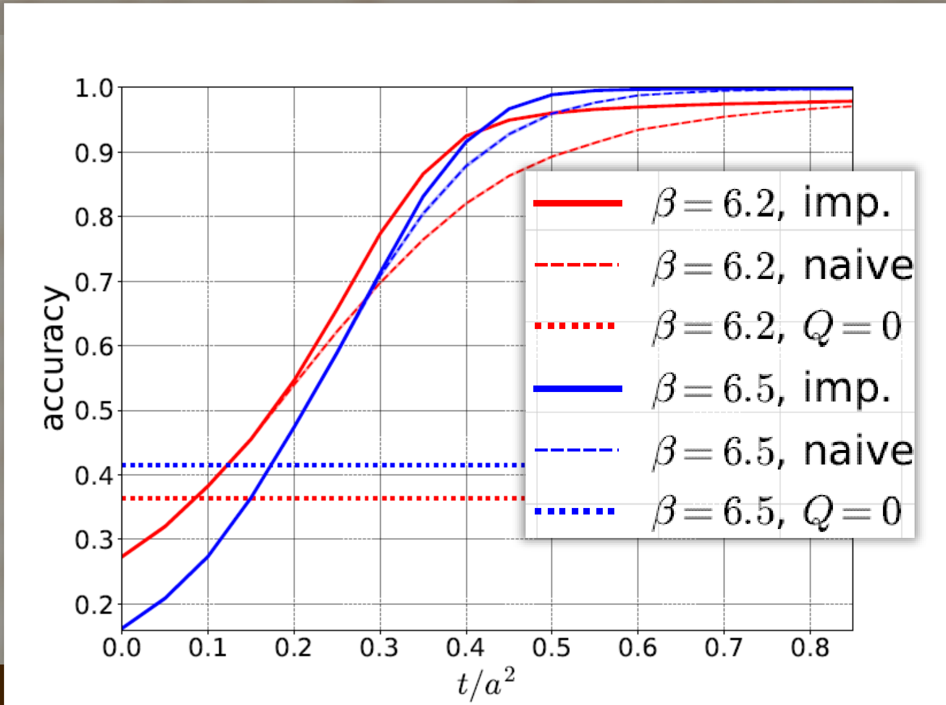
Simple estimator from $Q(t)$

1) Naïve: $Q = \mathrm{round}[Q(t)]$

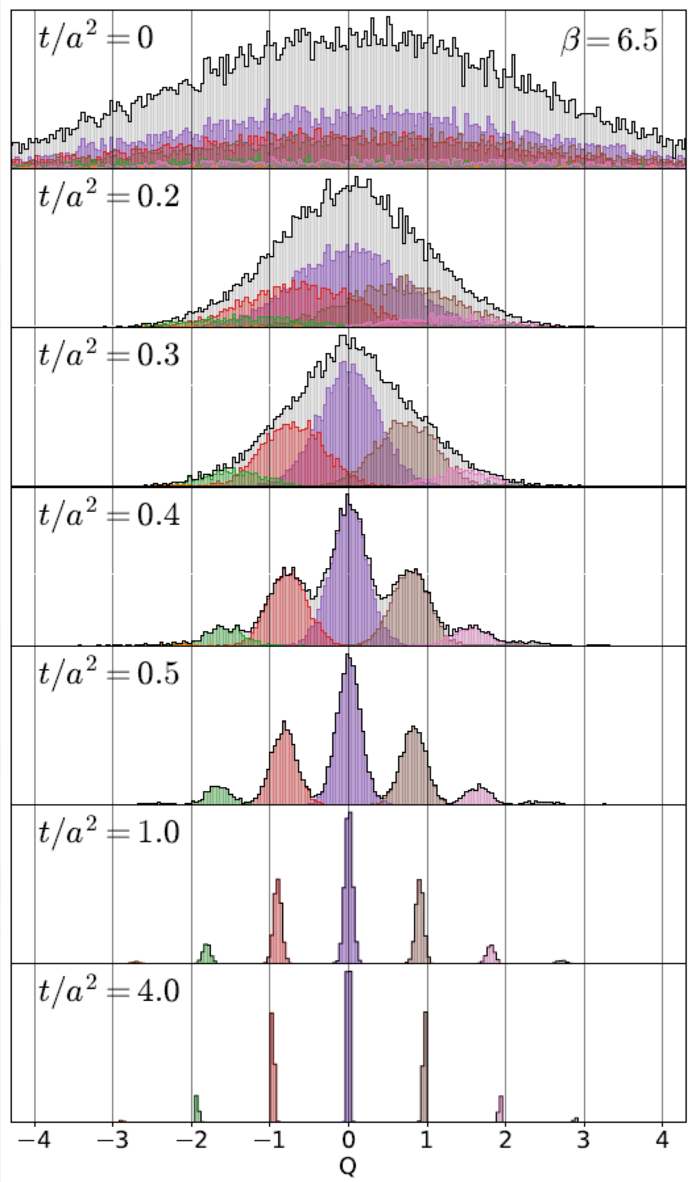
2) Improved: $Q = \mathrm{round}[cQ(t)]$

c>1: optimization param.

3) zero: $Q = 0$



Distribution of $Q(t)$



Comparison: NN vs Benchmark

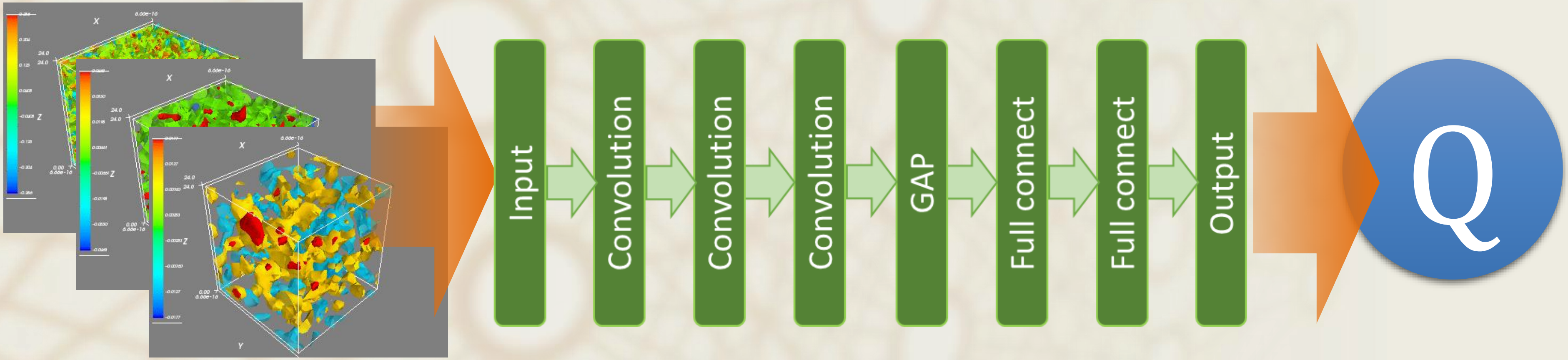
accuracy at β=6.2

	ML (Trial 2)	naïve	improved
$t/a^2=0$	37.0	27.3	27.3
$t/a^2=0.1$	40.1	38.3	38.3
$t/a^2=0.2$	55.2	54.0	54.6
$t/a^2=0.3$	77.6	69.8	77.3

- Machine learning cannot exceed the benchmark value.
- NN would be trained to answer the “improved” value.
- **No useful local structures found by the NN.**

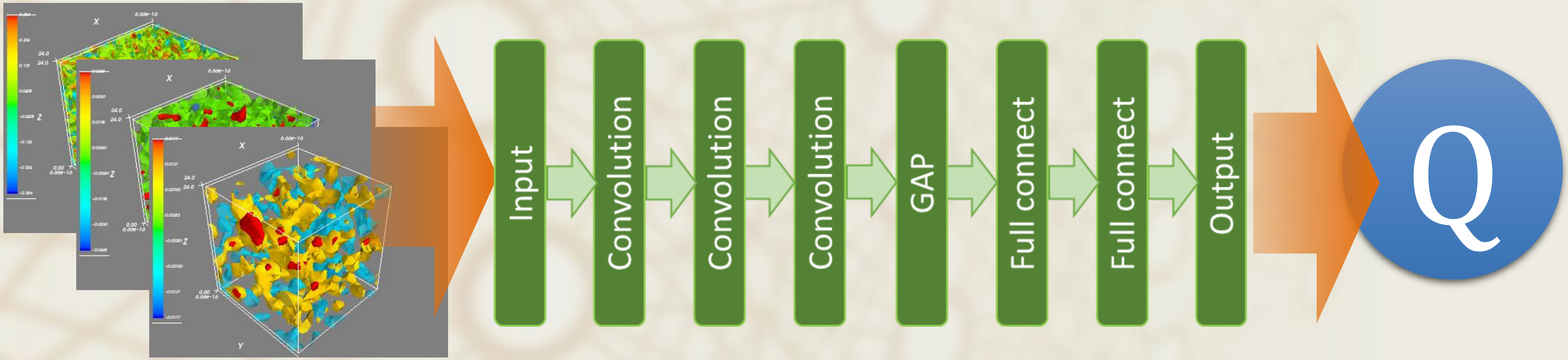
Trial 3: Multi-Channel Analysis

- Input: $q(x,t)$ in 4-dim. space **at t/a^2=0.1, 0.2, 0.3**



Trial 3: Multi-Channel Analysis

- Input: $q(x,t)$ in 4-dim. space **at t/a^2=0.1, 0.2, 0.3**



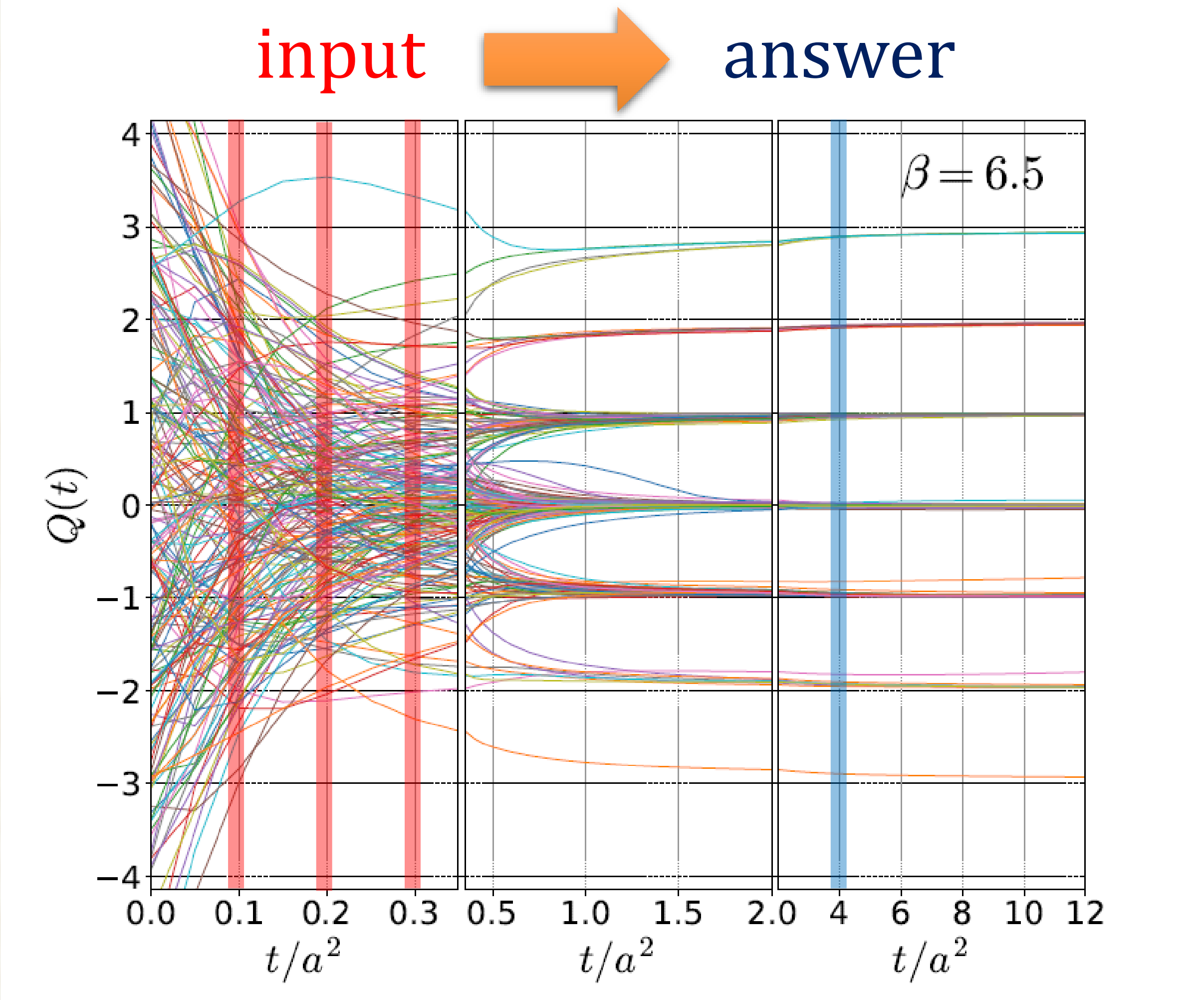
- **Result**

machine learning / benchmark @ t/a^2=0.3

	machine learning	benchmark @ t/a^2=0.3
β=6.2	**93.8**	77.3
β=6.5	**94.1**	71.3

- **non-trivial improvement from the benchmark!!**

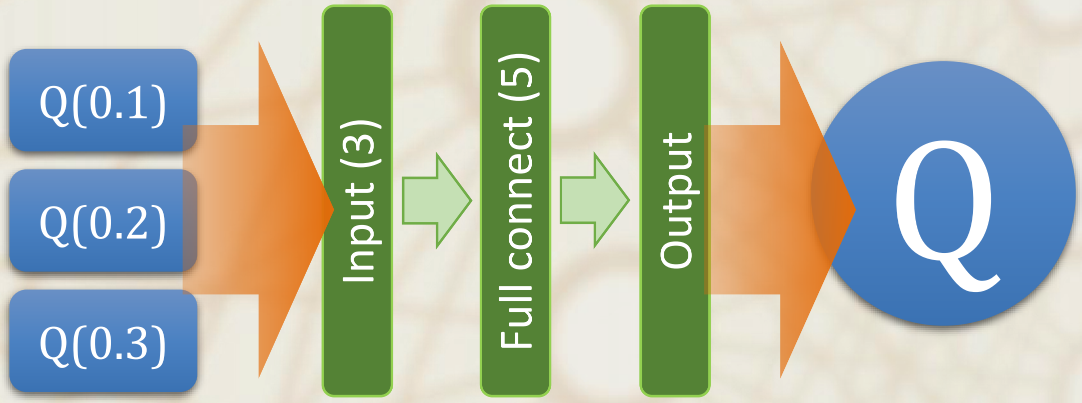
Is this a non-trivial result?

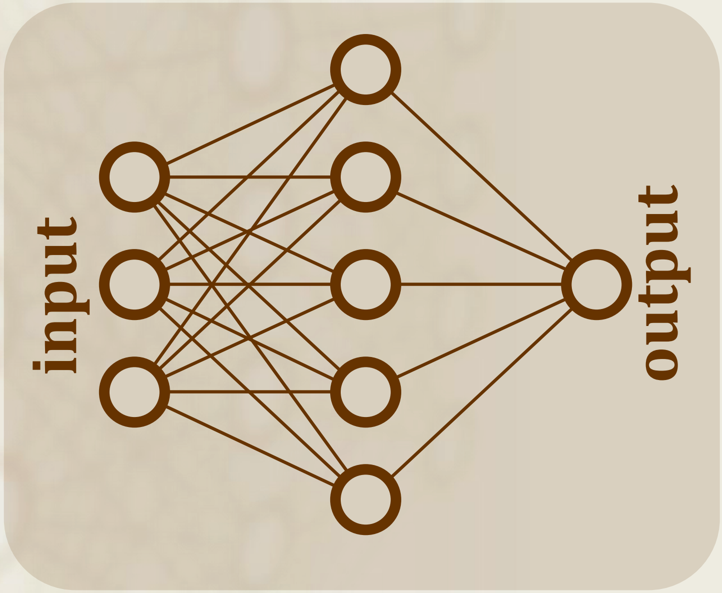


We can estimate the answer from $Q(t)$ by our eyes…

Trial 4: Feed $Q(t)$ [0-dim]

- Input: $Q(t)$ at $t/a^2 = 0.1, 0.2, 0.3$





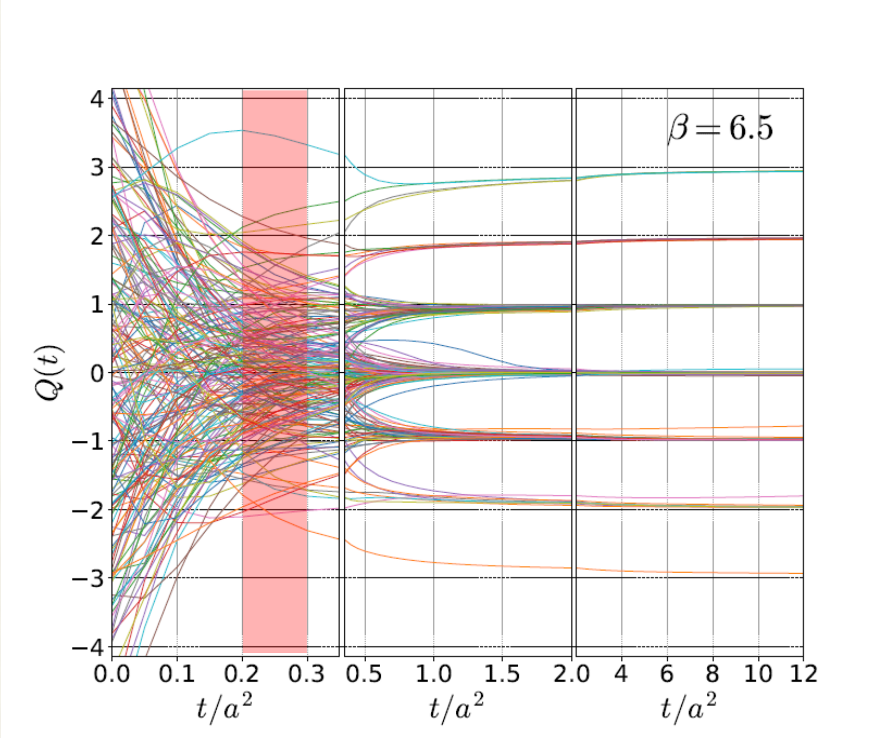
Result

	$Q(t)$	Trial 3 (4dim)	benchmark
β=6.2	**94.1**	93.8	77.3
β=6.5	**95.7**	94.1	71.3

- High accuracy can be obtained only from $Q(t)$

Using different flow times

t/a^2	β=6.2	β=6.5
0.3, 0.25, 0.2	**95.9(2)**	**99.0(2)**
0.3, 0.2, 0.1	94.1(2)	95.7(2)
0.25, 0.2, 0.15	93.9(3)	95.0(2)
0.2, 0.15, 0.1	86.4(3)	83.1(4)
0.2, 0.1, 0	74.1(5)	68.2(4)
0.15, 0.1, 0.05	69.2(4)	64.7(8)
0.1, 0.05, 0	53.8(5)	49.9(3)



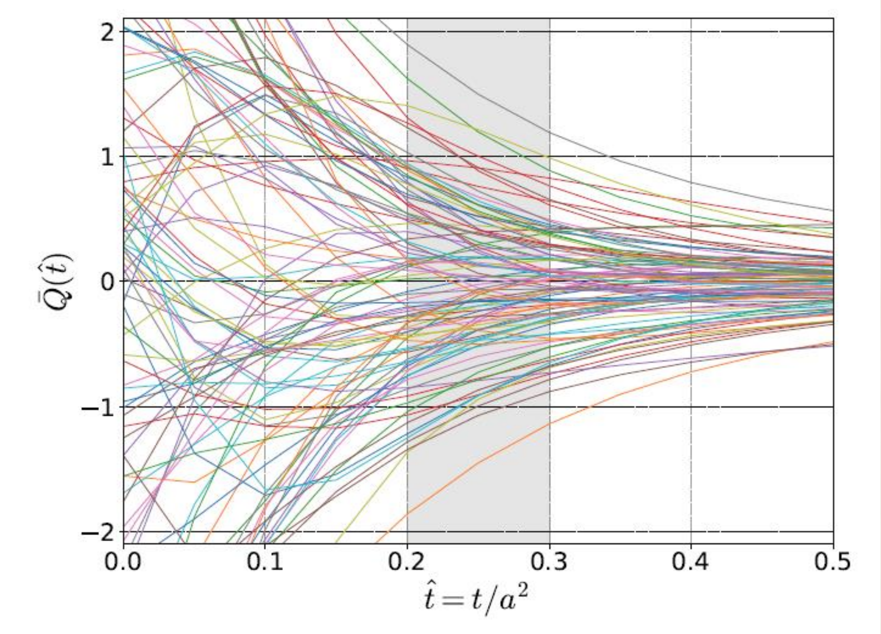
- t/a^2=0.3, 0.25, 0.2 gives the best accuracy.
- Better accuracy on the finer lattice.
- More than three input data do not improve accuracy.

  - error: variance in 10 independent trainings

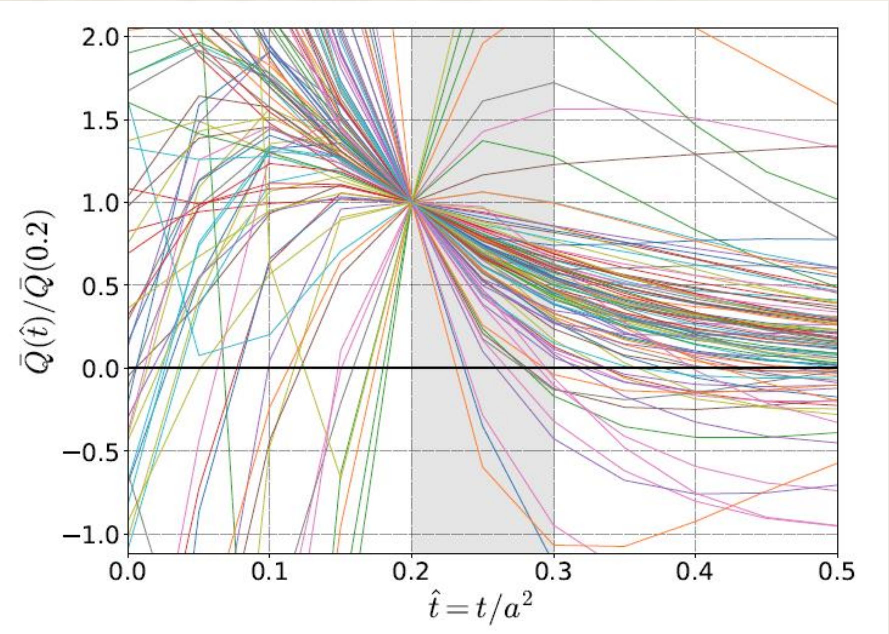
Trivial Check

- beta=6.5
- 100 samples

$\bar{Q}(t) = Q(t) - \mathcal{Q}$



$\bar{Q}(t)/\bar{Q}(0.2)$



- 99% accuracy is difficult by simple prescriptions.

Reducing the Training Data

- Smaller training data will reduce numerical cost for the training.

Training data	10,000	5,000	1,000	500	100
β=6.2	**95.9(2)**	**95.9(2)**	**95.9(2)**	95.5(3)	90.3(7)
β=6.5	**99.0(2)**	**99.0(2)**	**98.9(2)**	98.9(1)	90.2(8)

- 1000 configurations are enough to train the NN successfully!
- Numerical cost for the training is small.

Robustness Test

- Analyze configurations with a different parameter set

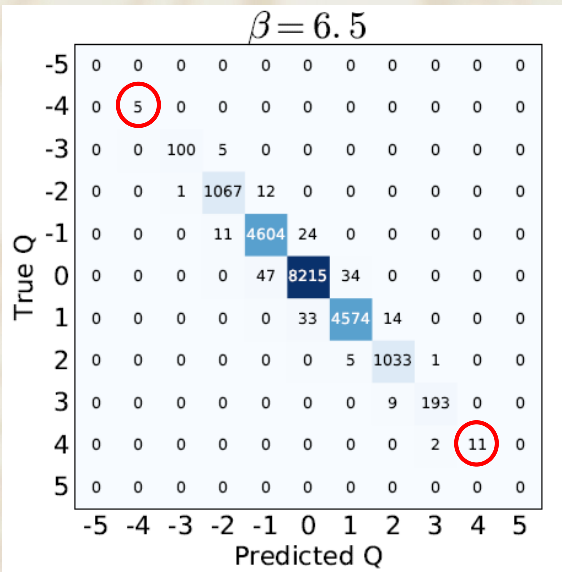
analyzed data

training data	β=6.2	β=6.5
β=6.2	95.9(2)	98.6(2)
β=6.5	95.6(2)	99.0(2)
both	**95.8(1)**	**98.9(2)**

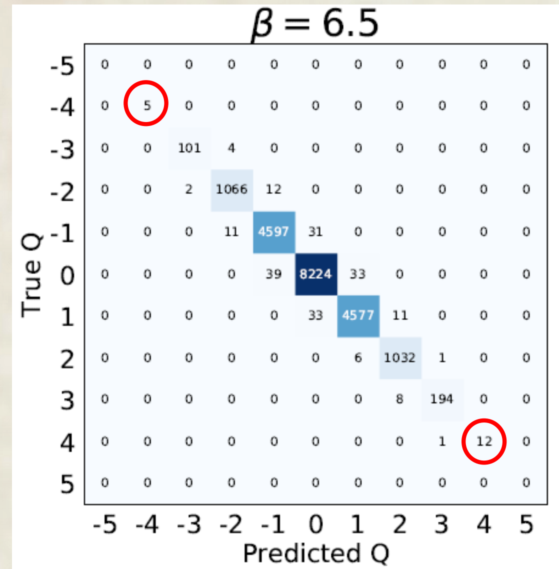
- NNs trained for β=6.2 and 6.5 can be used for another parameter successfully.
- **Universal NN would be developed!**
- Note: same physical volume

Untrained Answers

standard training



training w/o |Q|=4,5



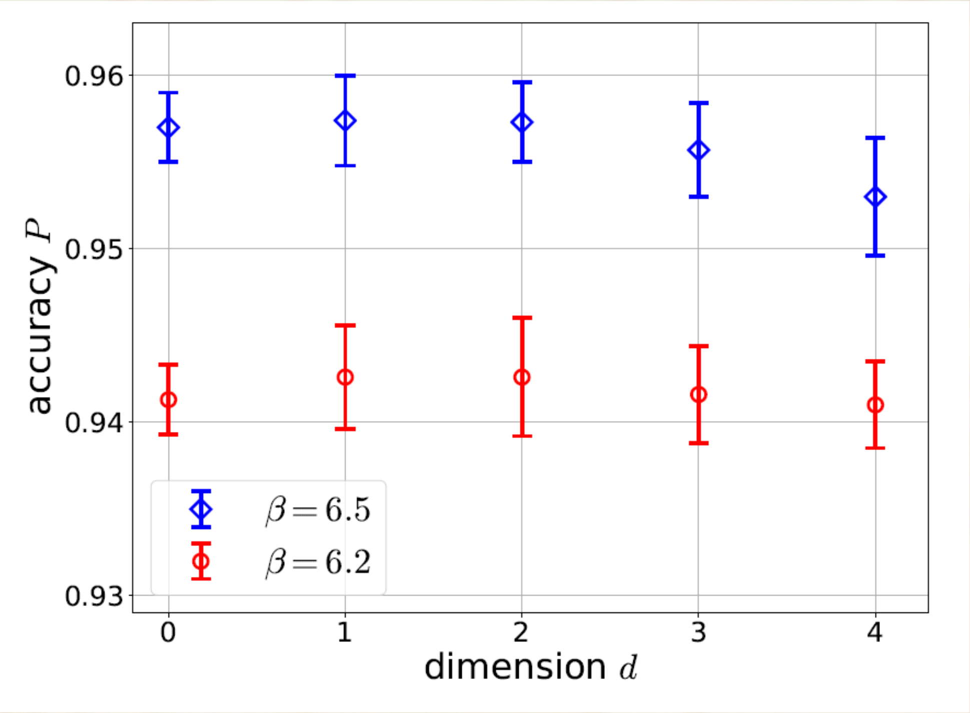
- CNN can make accurate predictions even for untrained values of Q.

Trial 5: Dimensional Reduction

- Optimal dimension between d=0 and 4?
- d-dimensional CNN
- Input: $q_d(x)$ after dimensional reduction
- 3-channel analysis: t/a^2=0.1, 0.2, 0.3

$$q_3(x, y, z) = \int d\tau q(x)$$

$$q_2(x, y) = \int d\tau dz q(x)$$



- No *d* dependence
- Failed in finding features in multi-dim. space.
- No instanton-like local structure in QCD vacuum?

Summary and Outlook

- **Topological charge can be estimated with high accuracy from $Q(t)$ at $0.2 < t/a^2 < 0.3$ with the aid of the machine learning technique.**
- On the finer lattices, the better accuracy.
- Applications: checking topological freezing, etc.

- No local structures in multi-dim. space captured by NN
- No "Instanton"-like structure? Or too noisy data?

- Future Study
  - Continuum limit / volume dependence
  - High T configurations where DIGA is valid

Is machine learning technique meaningfully applicable to theoretical physics?

- ML does not provide us with “understanding”.
- Theoretical physicists must have a better solution for problems that ML can deal with.
- 100% accuracy is never reached by the ML.

Search for 4-dim Features

- **Subtracting the average (normalization)**

- Large flow time
- Nonzero T configurations → DIGA

- New numerical analysis @ $T = 1.43T_c$

β	N^4	N_{conf}
6.3	36^4x8	2,000

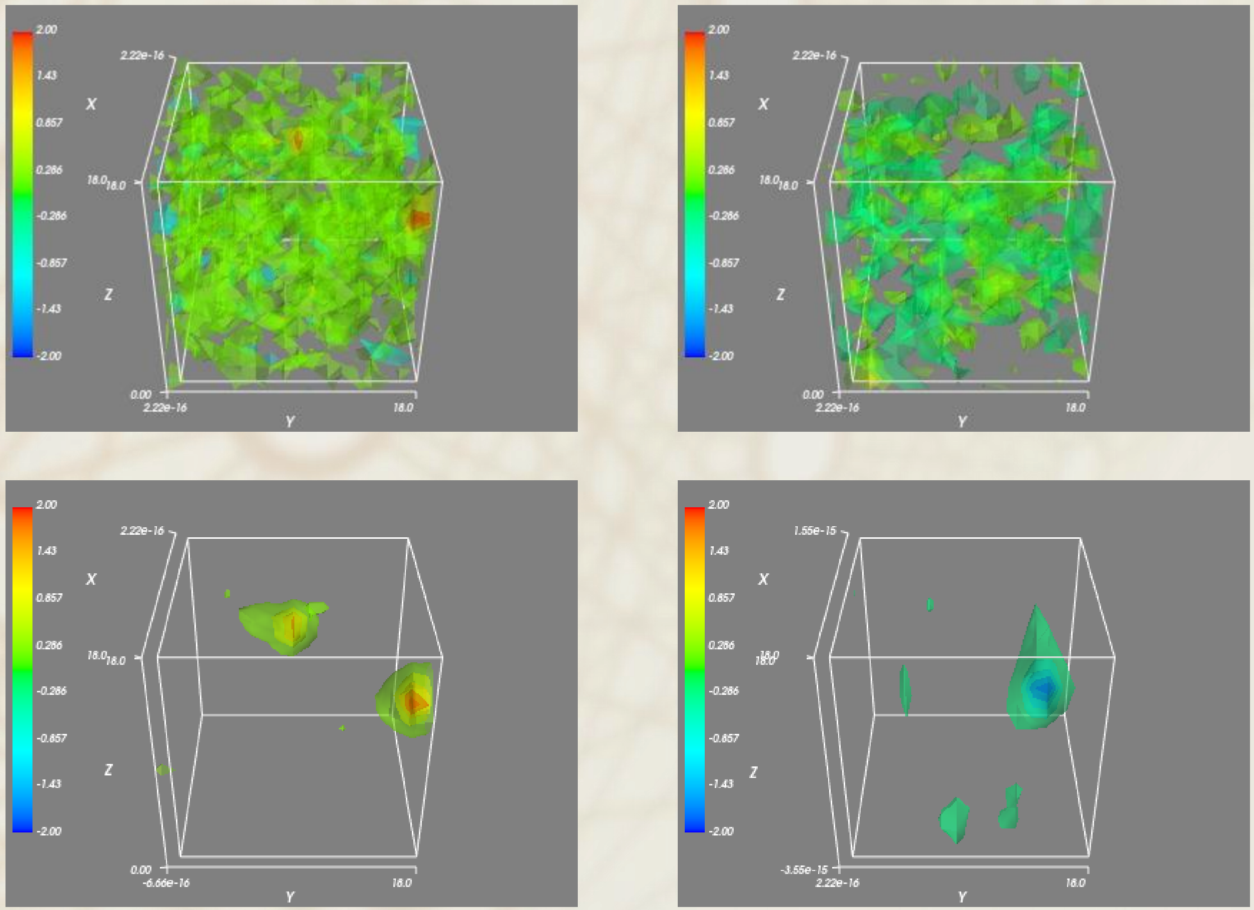
$q(\boldsymbol{x}, t)$

$$q_t(\boldsymbol{x}) = \int d\tau q_t(\boldsymbol{x}, \tau)$$

$Q = 2$ $\qquad$ $Q = -1$

$t/a^2 = 1.0$

$t/a^2 = 4.0$



- ☐ Clear local objects at large t

Results

- t/a^2=4.0 (3D CNN, normalized)

Q	-2	-1	0	1	2	total
accuracy	1.000	1.000	0.991	1.000	0.981	0.995

- t/a^2=1.0 (normalized)
  - 3D-CNN

Q	-2	-1	0	1	2	total
accuracy	0.556	0.638	0.697	0.700	0.698	0.658

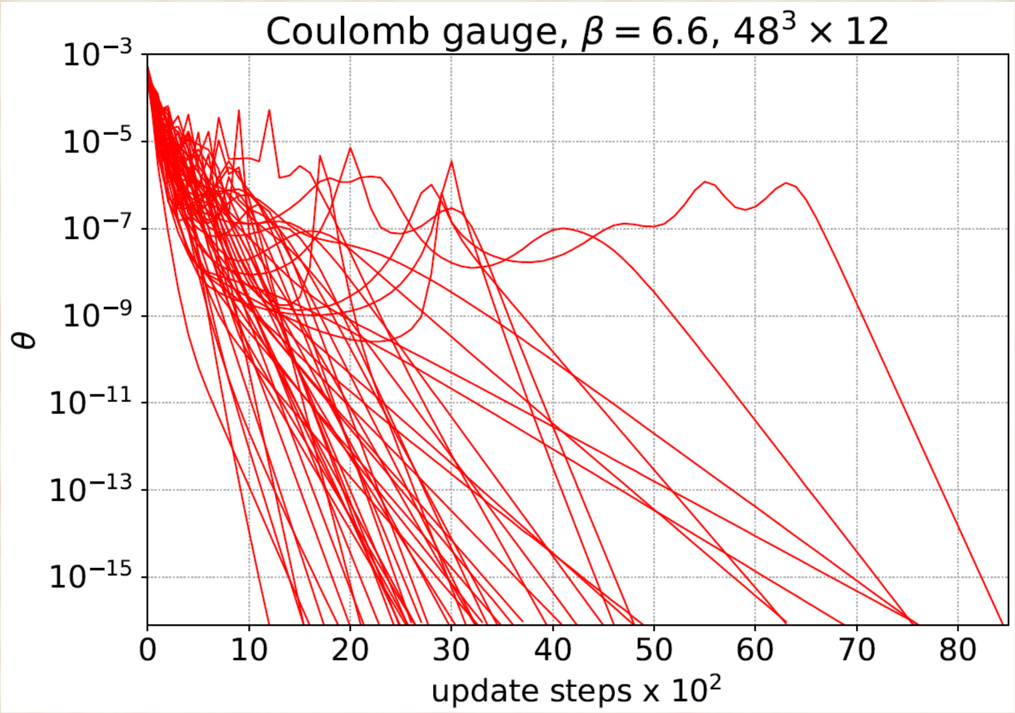
  - 4D-CNN

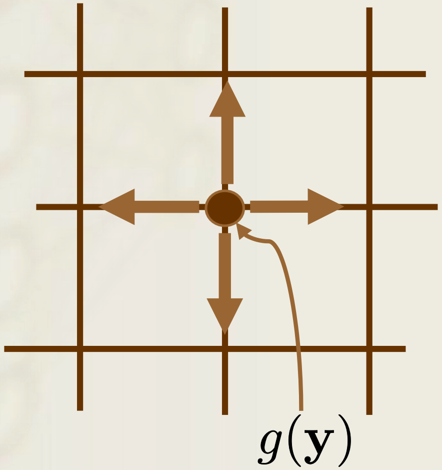
Q	-2	-1	0	1	2	total
accuracy	0.667	0.670	0.770	0.663	0.679	0.693

- Multi-dim. features can be captured at sufficiently large *t*.
- Non-trivial feature in 4D space captured?

Gauge Fixing (Coulomb or Landau)

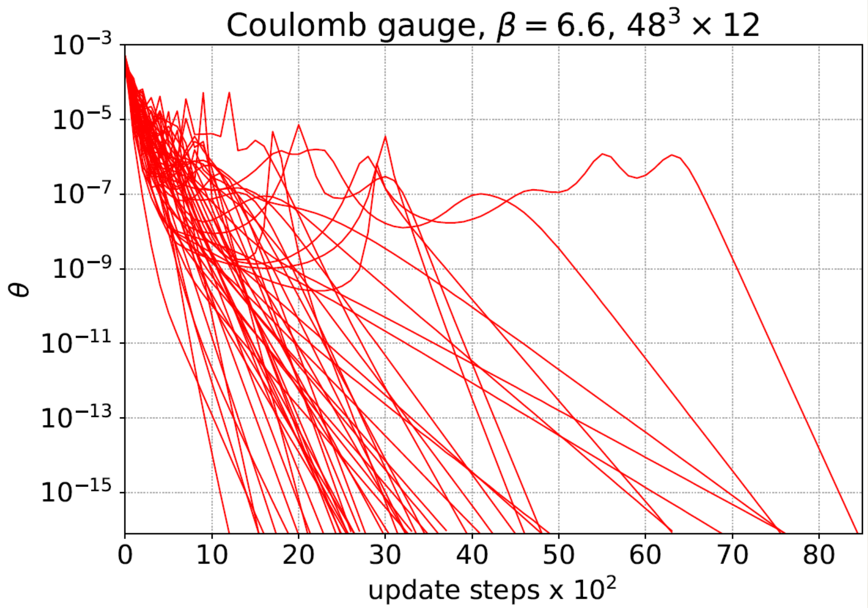
- Standard algorithm: local updates with over relaxation
- Slow convergence for several initial conditions
- Convergence is extremely slow sometimes



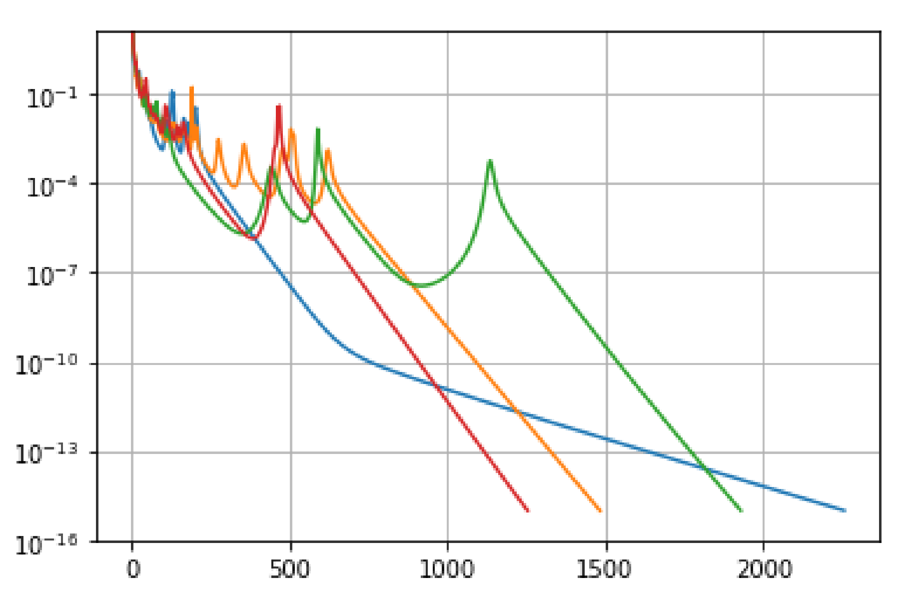


U(1) Gauge Fixing

SU(3) YM



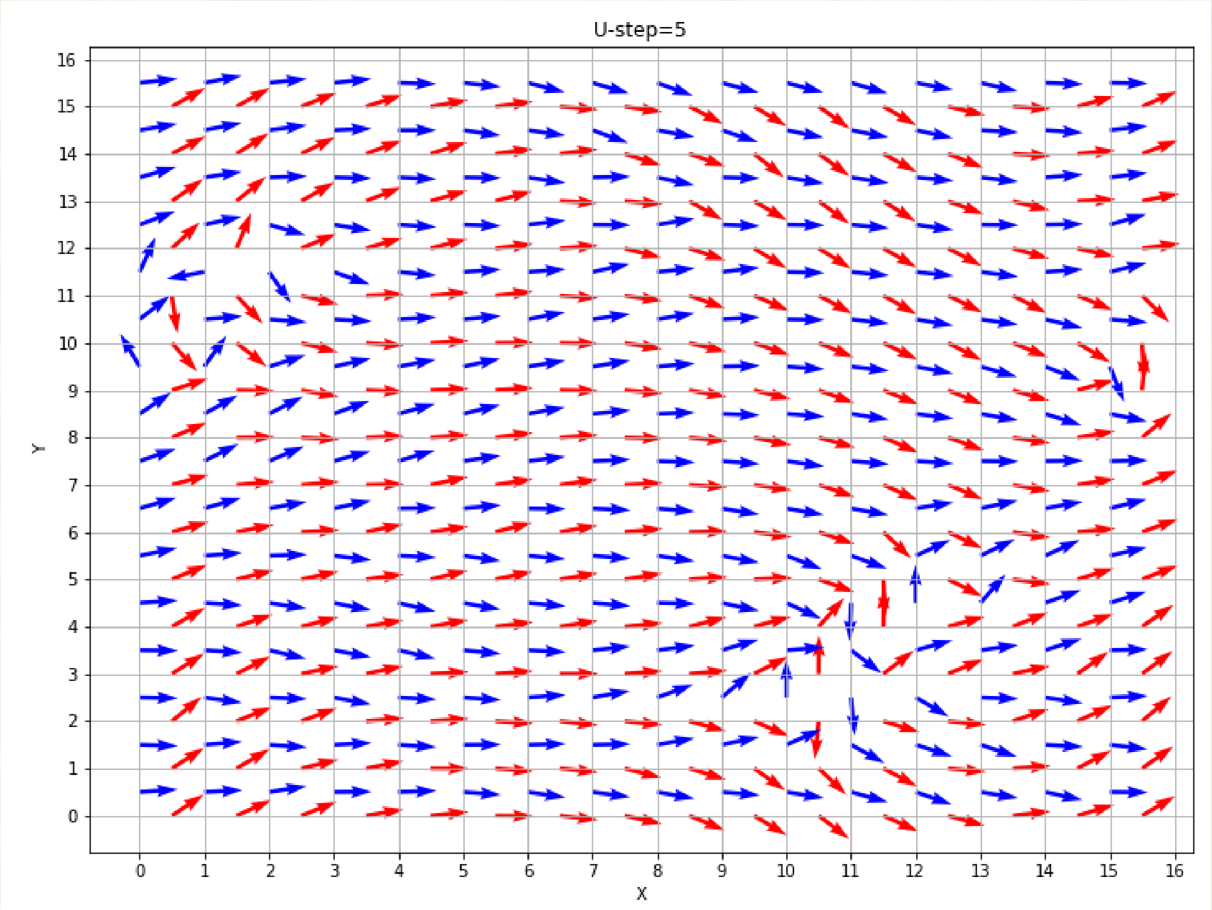
U(1) gauge theory



- U(1) gauge fixing has similar behavior as SU(3) YM

Numerical Results: 2D-Coulomb

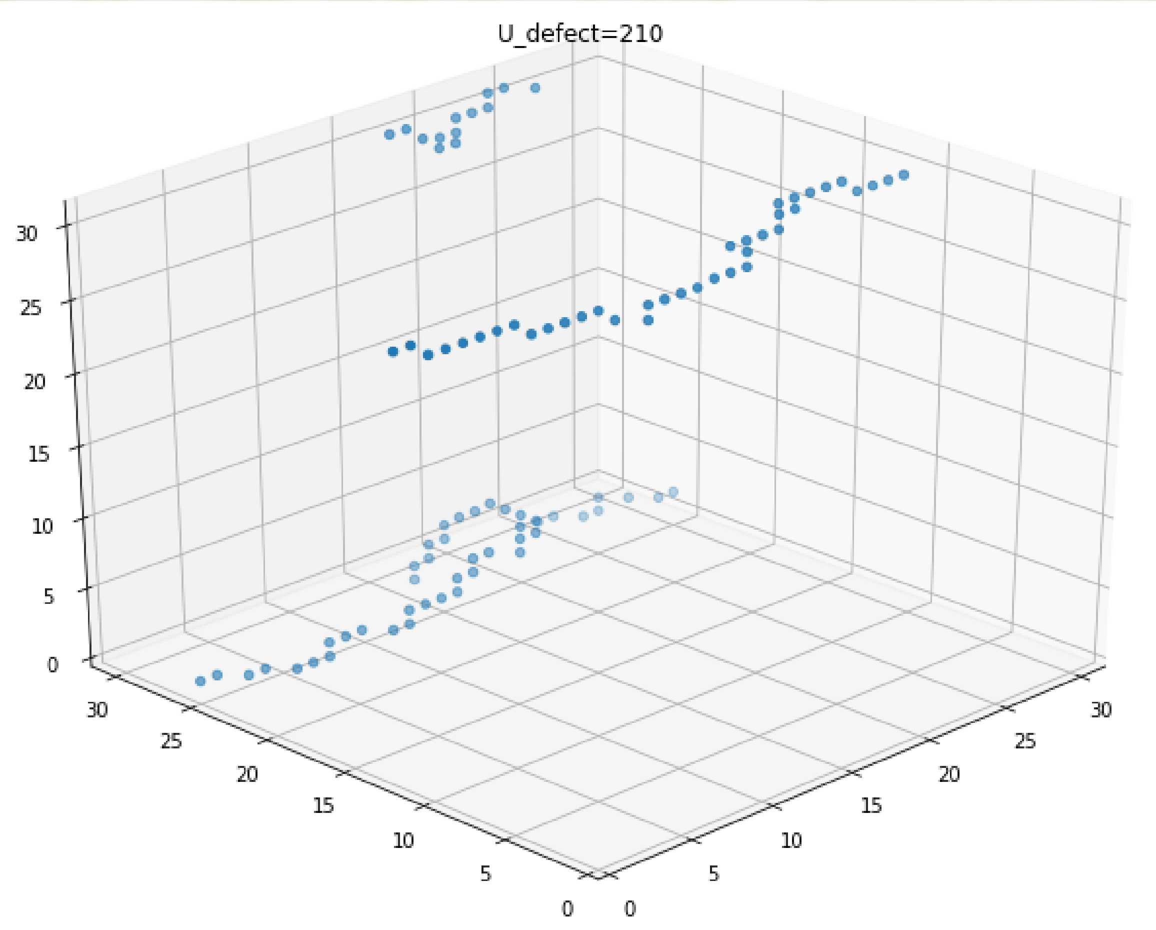
- Initial: trivial configuration + random gauge rotations



- Manifestation of "Dirac string" → prevent convergence

Numerical Results: 3D-Coulomb

- Initial: trivial configuration + random gauge rotations

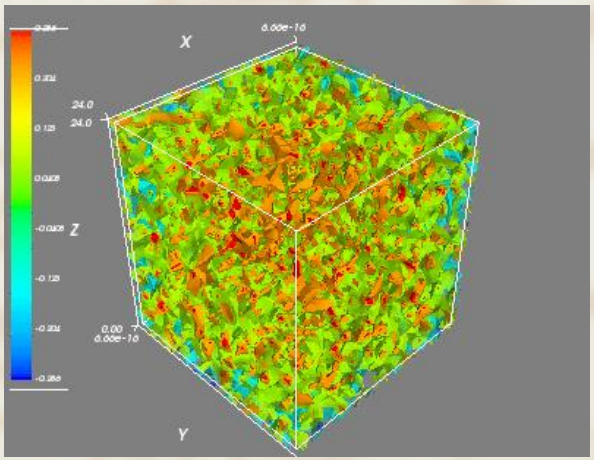


- Manifestation of “Dirac string” → prevent convergence

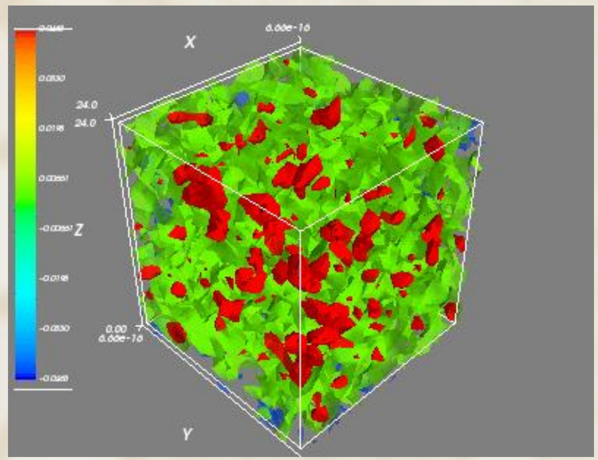
backup

Topological Charge Density

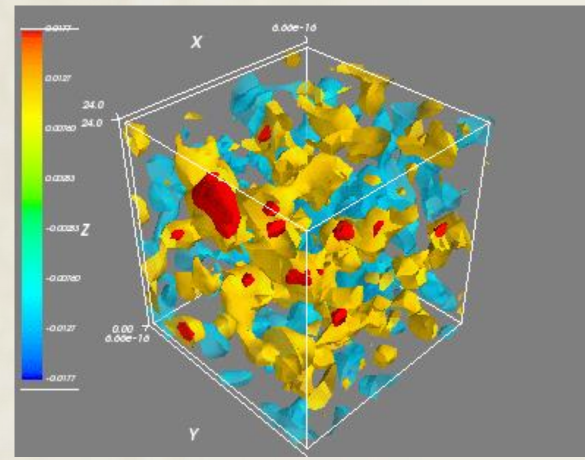
$t/a^2 = 0.1$



$t/a^2 = 0.2$



$t/a^2 = 0.3$



No isolated instanton structure…