Classifying Topological Sector via Machine Learning

Masakiyo Kitazawa, Takuya Matsumoto, Yasuhiro Kohno
(Osaka University)

MK, Kohno, Matsumoto, to appear

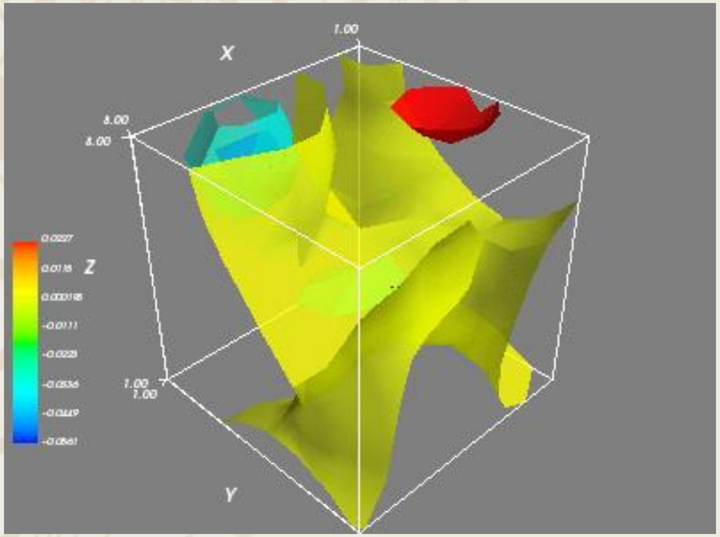
37th international conference on lattice field theory (LATTICE2019), Wuhan, China, June 21, 2019

Topological Charge in YM Theory

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

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

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

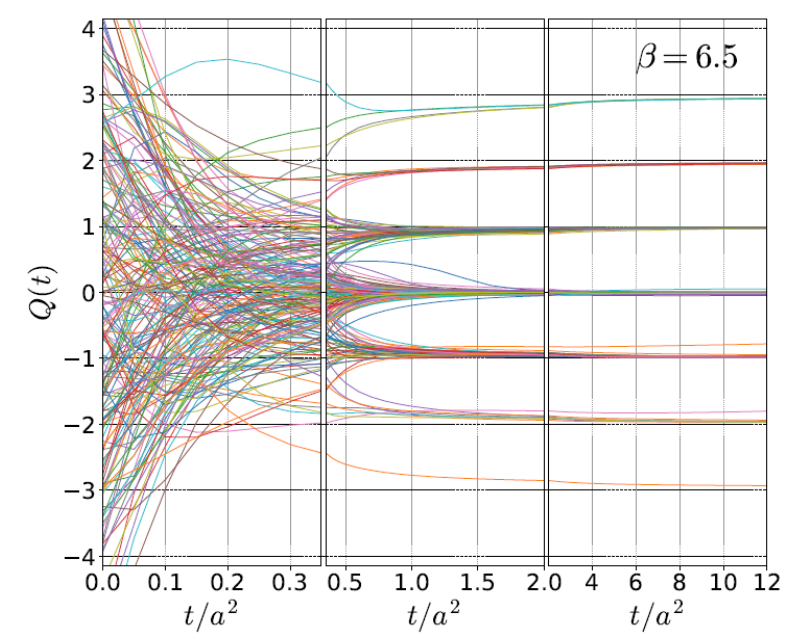
Topology on the Lattice

- Distinct topological sectors on sufficiently fine lattices

Luscher, 1981

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

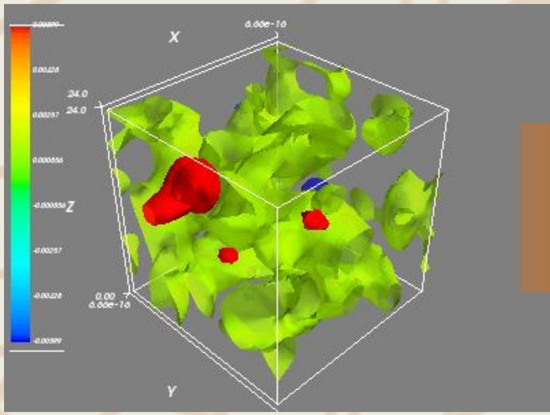
Luscher, Weisz, 2011



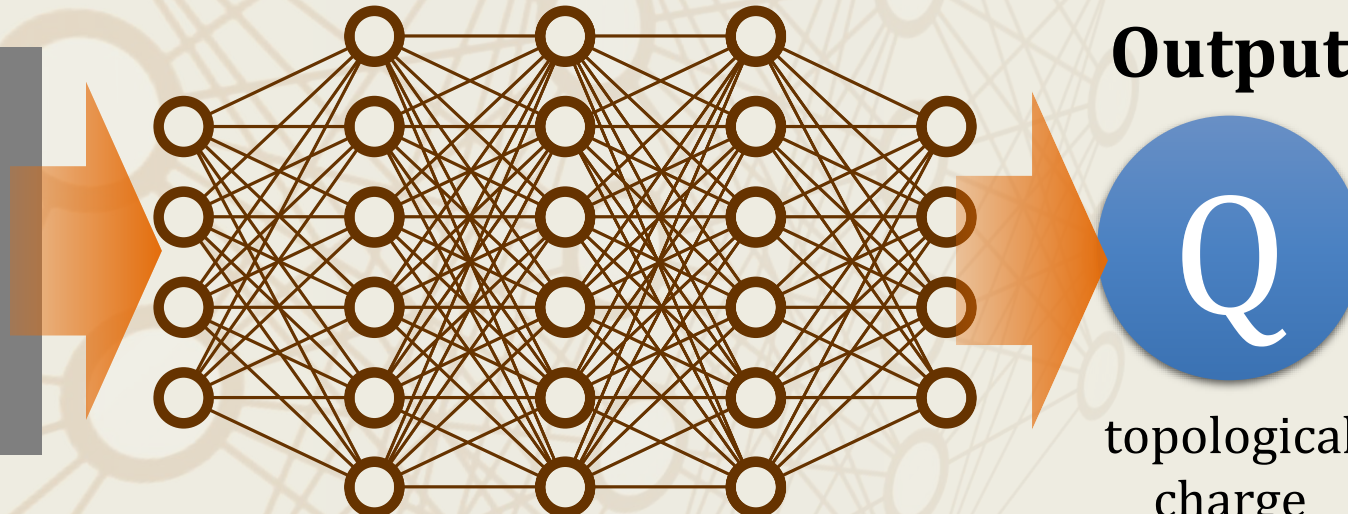
- Good agreement b/w various definitions
- **Faster algorithm is desirable!**

Machine Learning

Input: q(x)



4-dimensional field



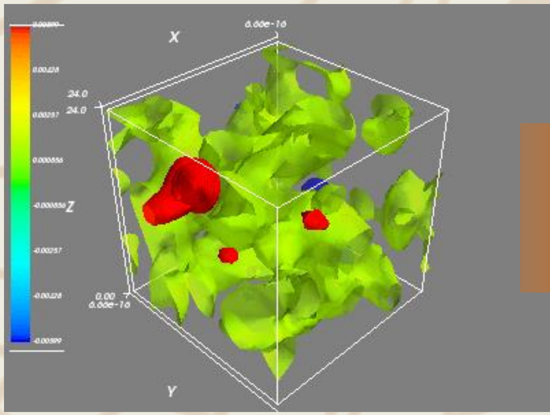
Output

topological
charge

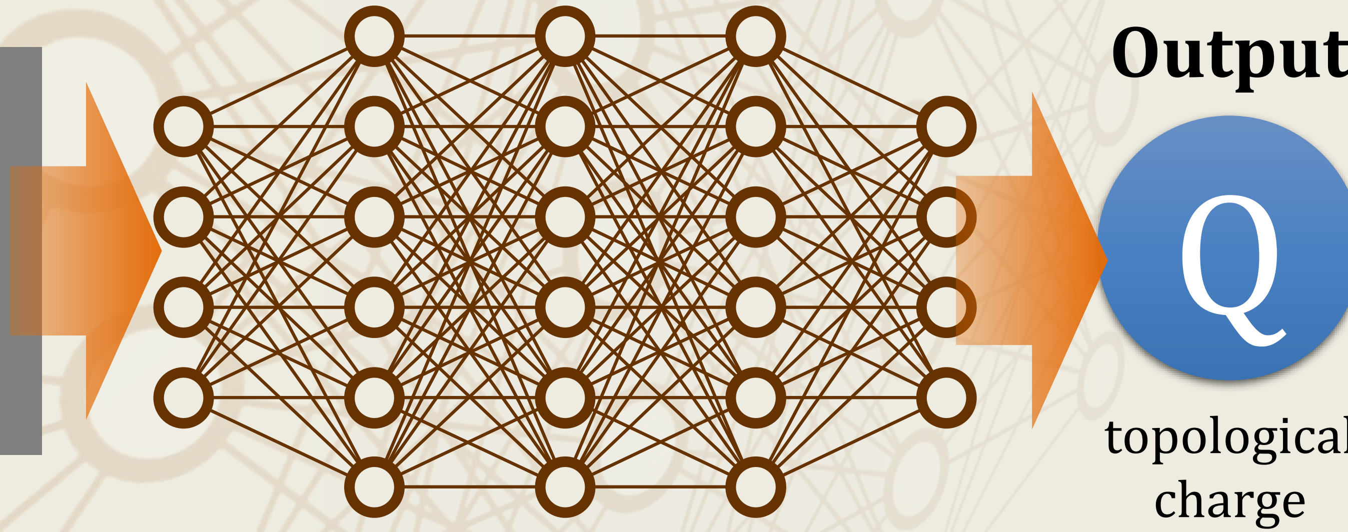
- **Capture "instanton"-like structure?**
- **Acceleration of the analysis of Q?**

Machine Learning

Input: q(x)



4-dimensional field



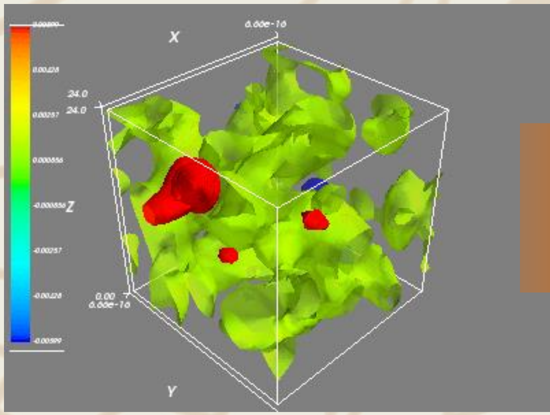
Output

topological charge

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

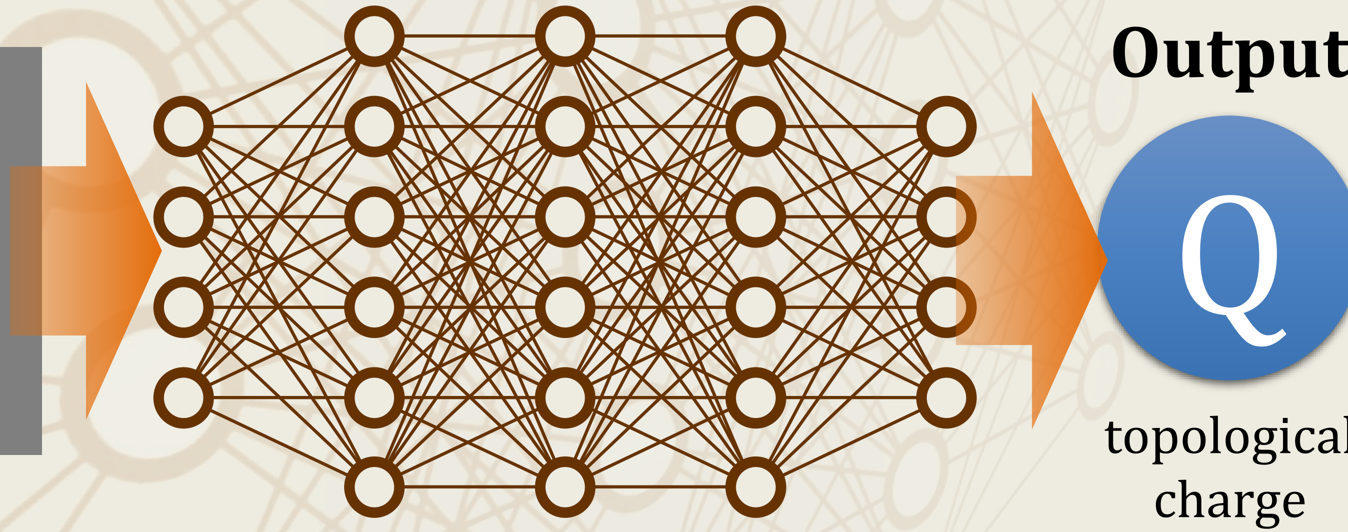
Machine Learning

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

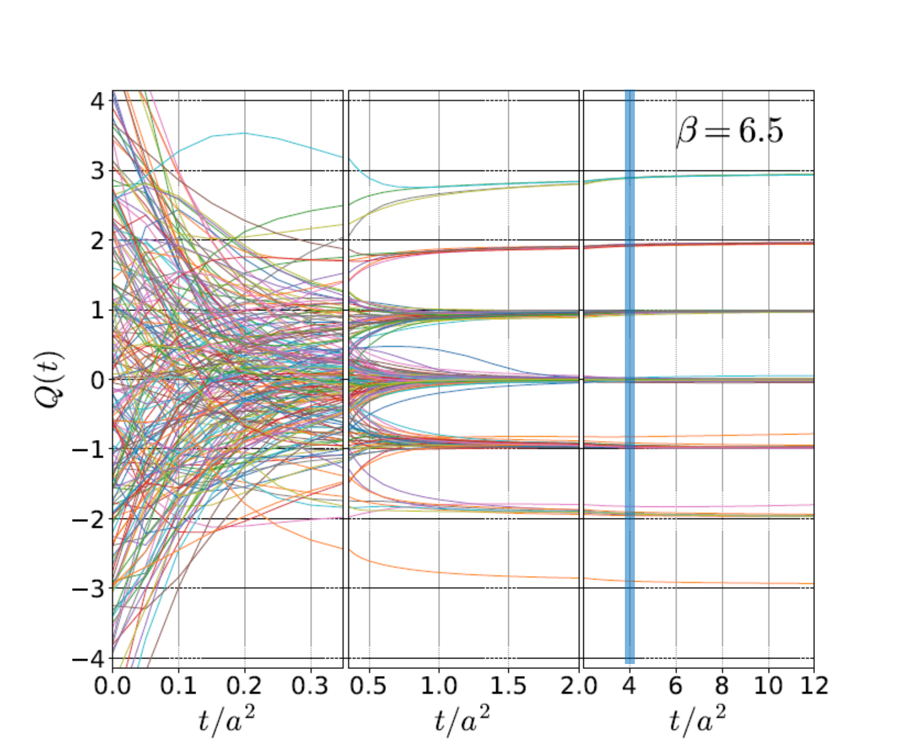
distribution of Q

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

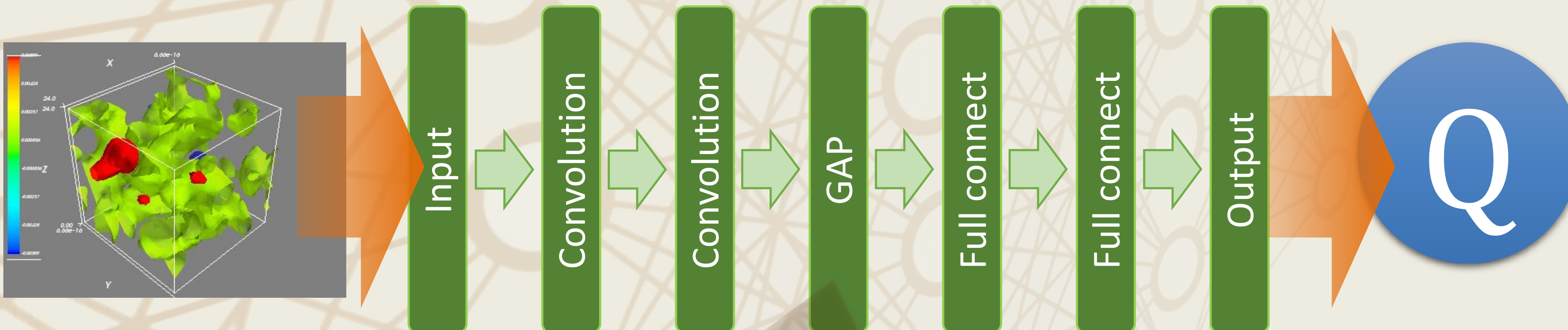
- convolutional neural network by **CHAINER framework**
- supervised learning
- convolutional layer: 4-dim., periodic BC
- regression analysis / round off to obtain integer
- activation: logistic

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



Trial 1: Topol. Charge Density

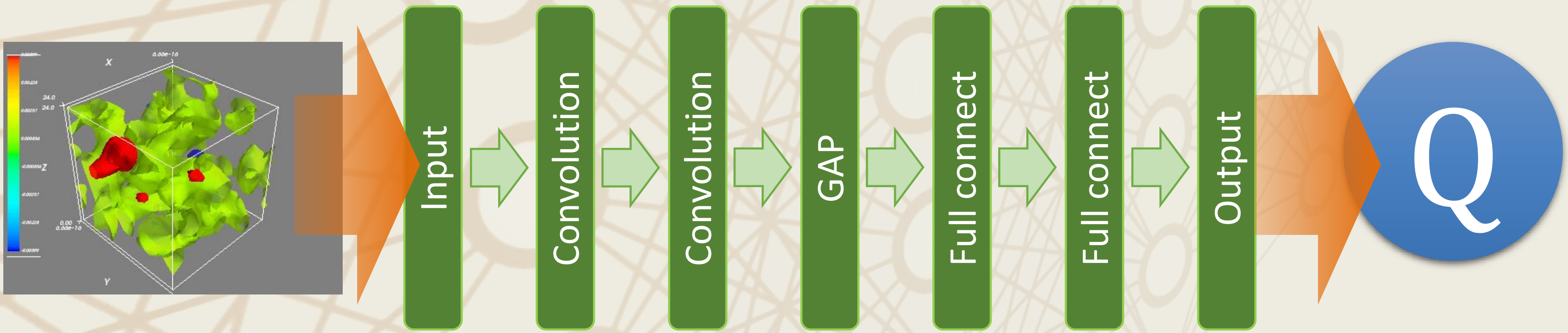
- Input: q(x) in 4-dim space
- Data reduction to 8^4 (average pooling)



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)



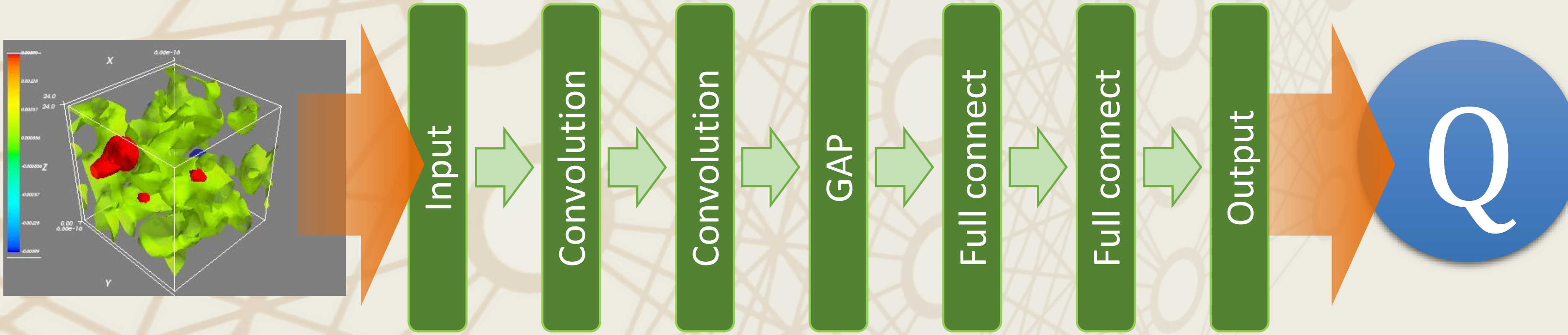
- **Result: best accuracy for β=6.2: 37.0%**

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



Accuracy of each topological sector (%)

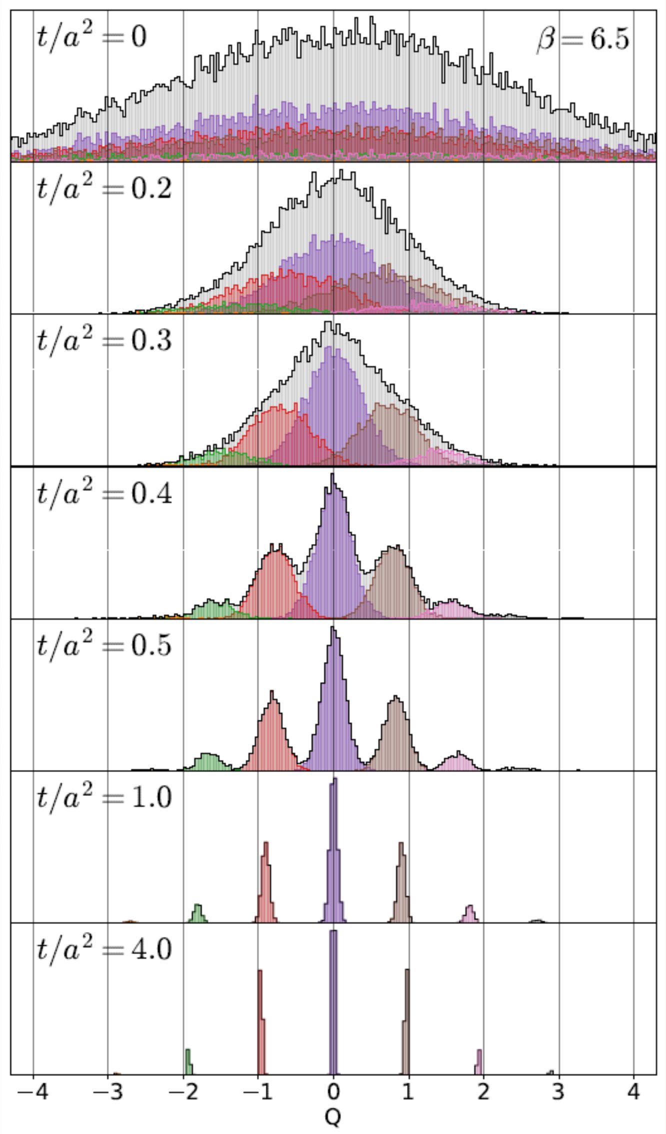
Q	-4	-3	-2	-1	0	1	2	3	4	total
$t/a^2=0$	0	0	0	0	37.2	0	0	0	0	**37.0**
$t/a^2=0.1$	0	0	31.6	39.1	41.4	38.9	19.0	0	0	**40.3**
$t/a^2=0.2$	0	40.0	46.4	53.8	55.9	52.3	48.1	50.0	0	**53.7**
$t/a^2=0.3$	0	91.3	72.9	76.3	79.0	74.8	68.1	70.0	50.0	**76.1**

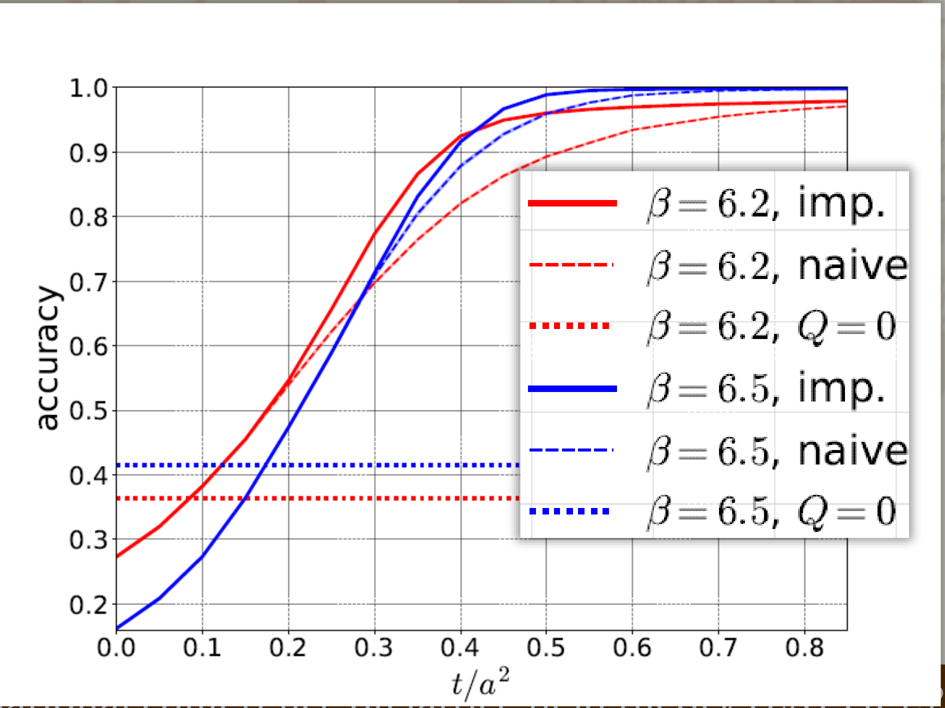
Benchmark

□ 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

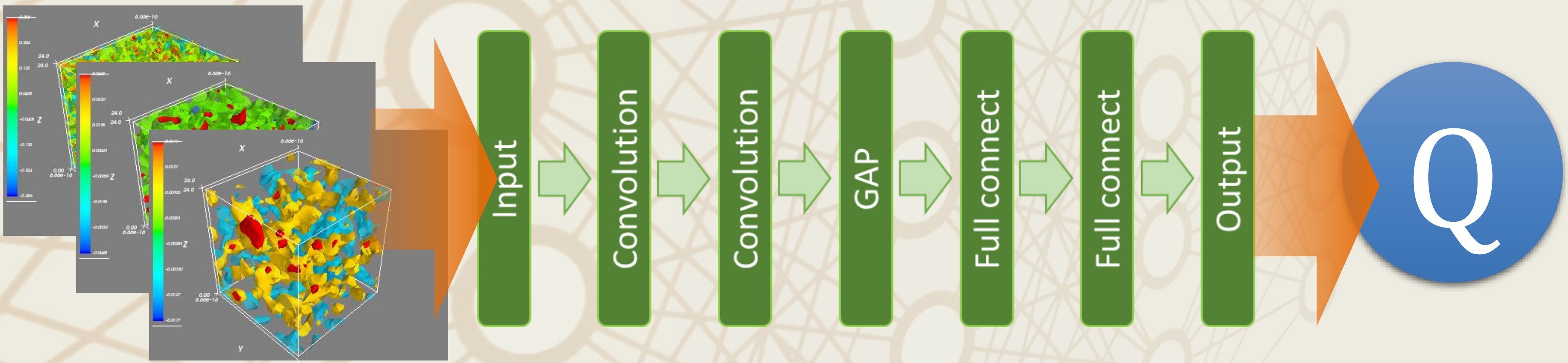
accuracy at β=6.2

	ML (Trial 2)	naïve	improved
$t/a^2=0$	37.0	27.3	27.3
$t/a^2=0.1$	40.3	38.3	38.3
$t/a^2=0.2$	53.7	54.0	54.6
$t/a^2=0.3$	76.1	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.**

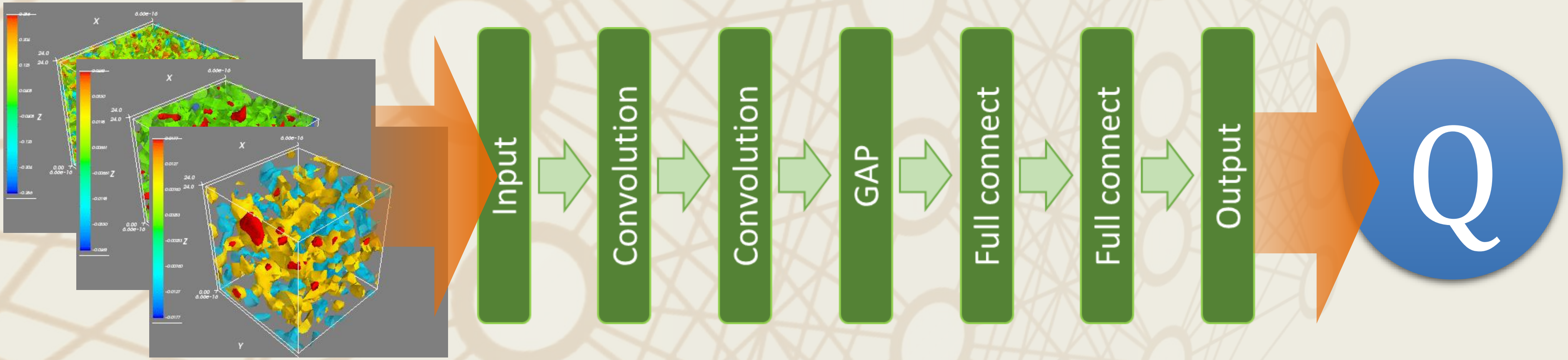
Trial 3: Multi-Channel Analysis

- Input: q(x,t) in four-dimensional space **at t/a^2=0.1, 0.2, 0.3**



Trial 3: Multi-Channel Analysis

- Input: q(x,t) in four-dimensional space **at t/a^2=0.1, 0.2, 0.3**



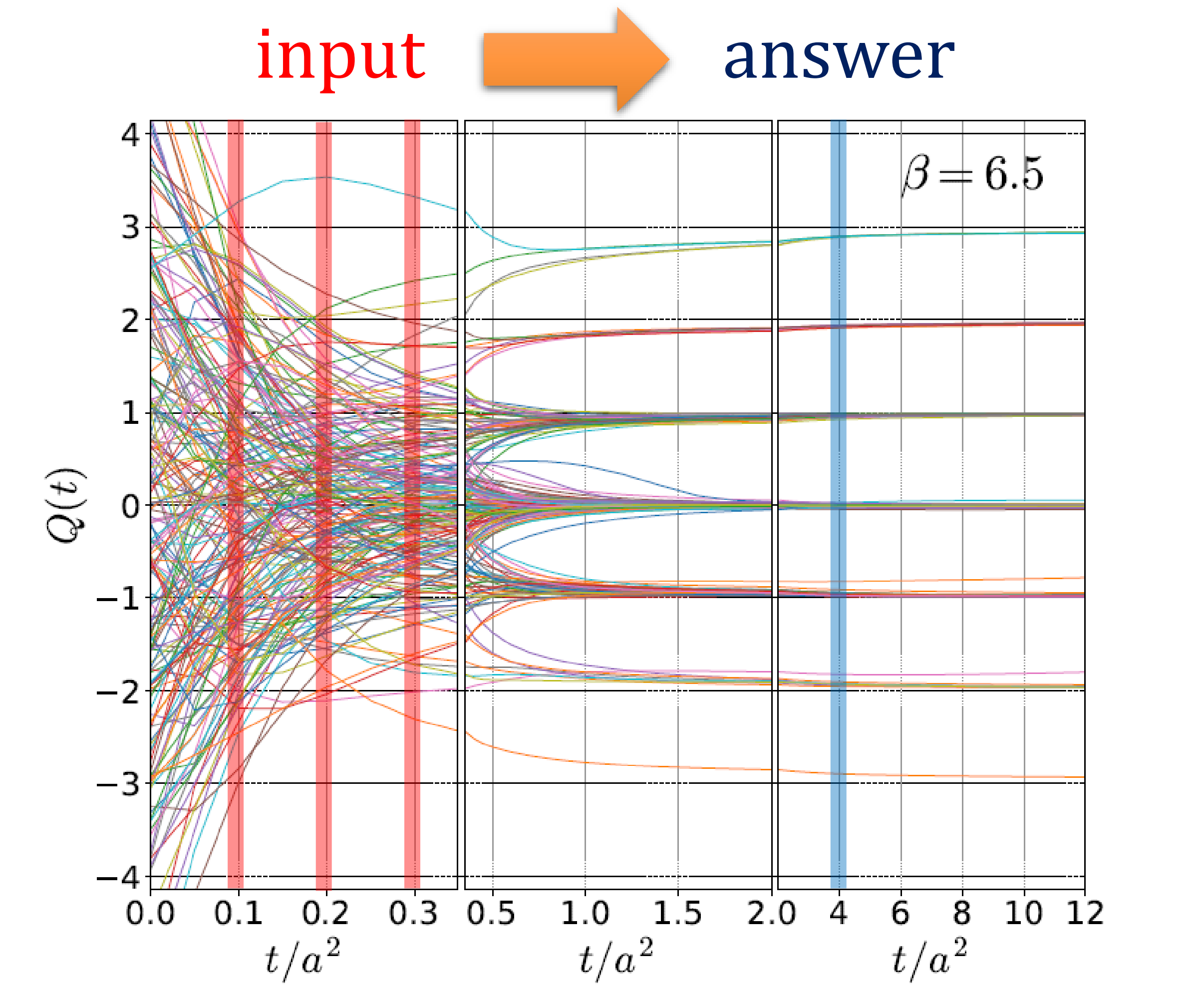
- **Result**

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!!**

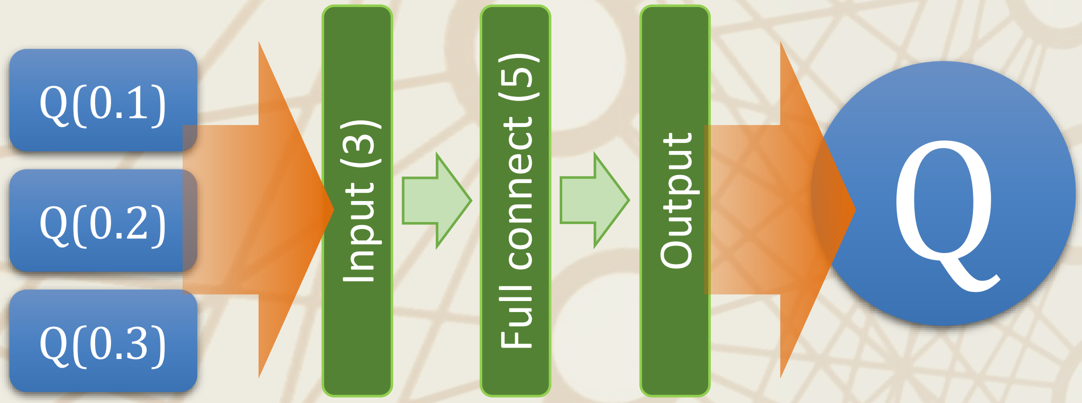
Is this a non-trivial result?

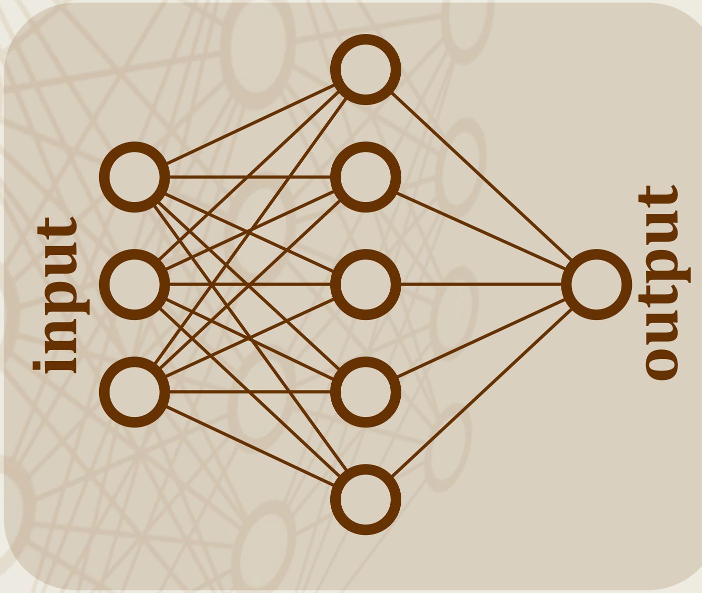


We can estimate the answer from Q(t) by our eyes…

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

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





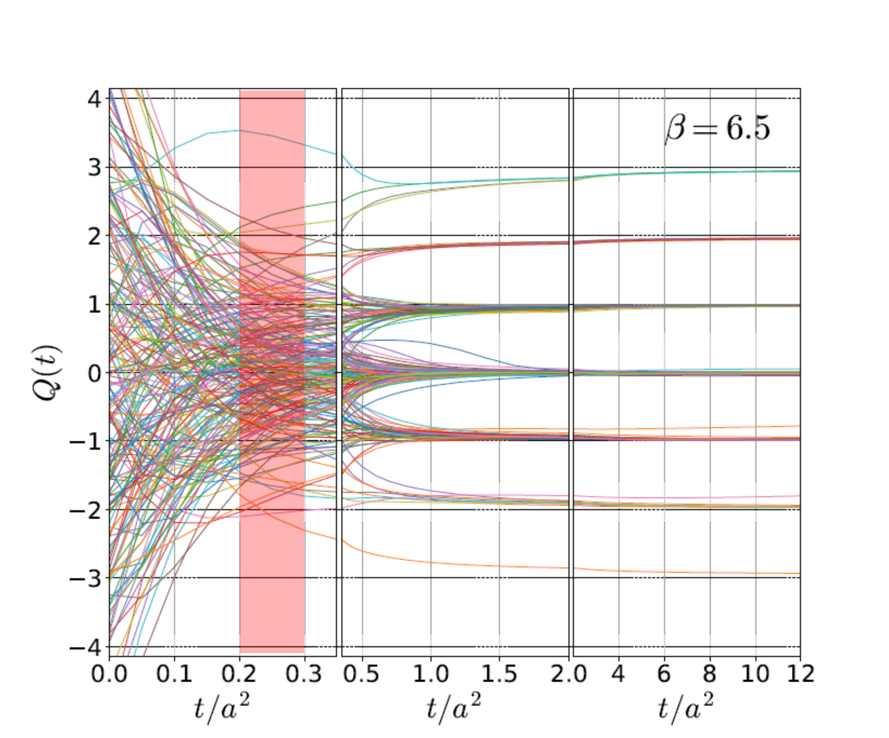
- **Result**

	Q(t)	Trial 3 (4dim)	benchmark
β=6.2	**95.5**	93.8	77.3
β=6.5	**95.7**	94.1	71.3

- Good accuracy is 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	95.5(2)	95.7(2)
0.25, 0.2, 0.15	95.1(3)	95.0(2)
0.2, 0.15, 0.1	86.9(3)	83.1(4)
0.2, 0.1, 0	75.6(5)	68.2(4)
0.15, 0.1, 0.05	71.8(4)	65.2(4)
0.1, 0.05, 0	54.8(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 t values do not improve accuracy.
  - error: variance in 10 independent trainings

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.

Versatility

- Analyze configurations with a different parameter set

training data	analyzed data: β=6.2	analyzed data: β=6.5
β=6.2	95.9(2)	98.6(2)
β=6.5	95.6(2)	99.0(2)

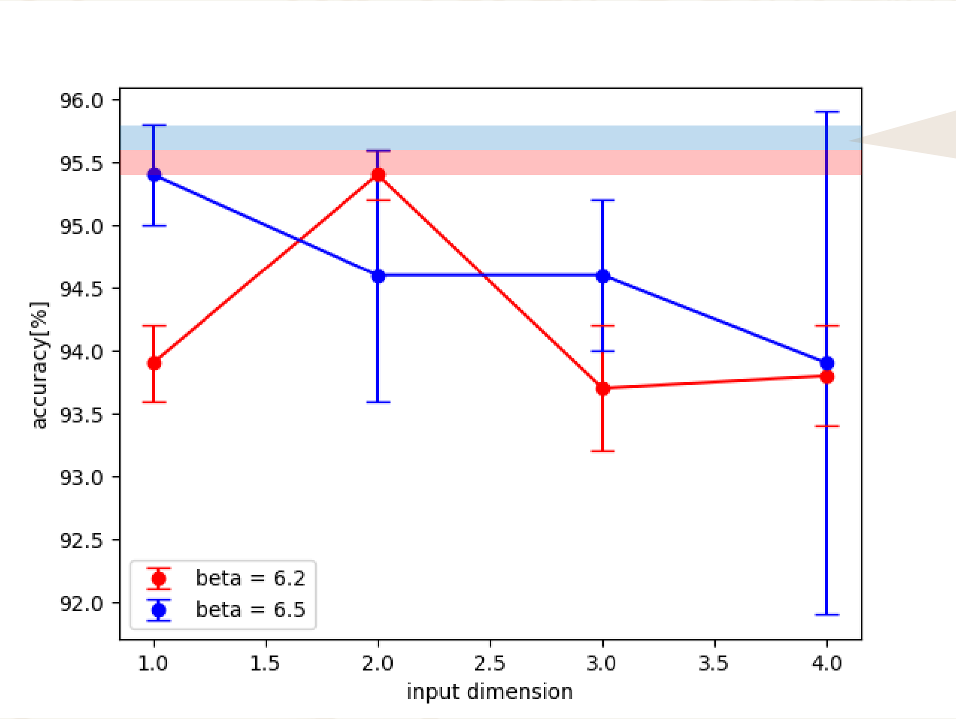
- NNs trained for β=6.2 and 6.5 can be used for another parameter successfully.
- **Universal NN would be developed!**
- Note: same physical volume

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

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

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



Accuracy of Trial 4

Summary and Outlook

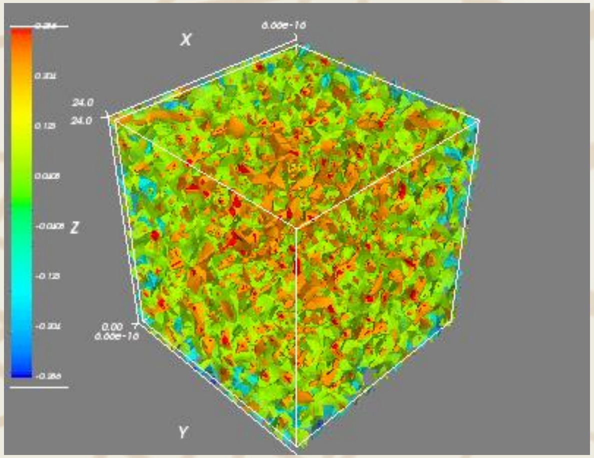
- **Topological charge can be estimated with high accuracy from Q(t) at $0.2<t/a^2<0.3$ with the aid of the machine learning technique.**
- On the finer lattices, the better accuracy.
- Applications: checking topological freezing, etc.

- No local structure captured by NN
- No "Instanton"-like structure? Or too noisy data?

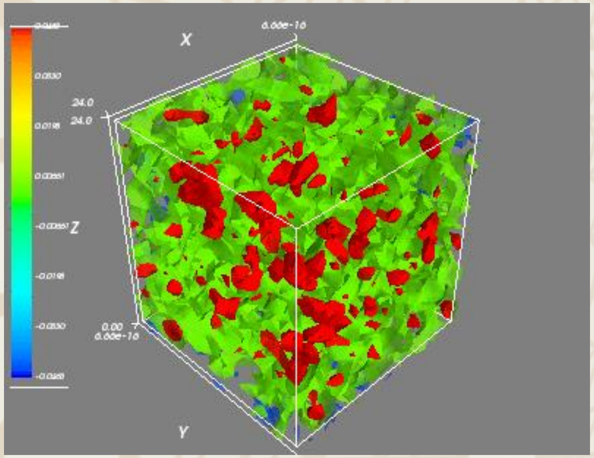
- Future Study
  - Continuum limit / volume dependence
  - High T configurations where DIGA is valid

Topological Charge Density

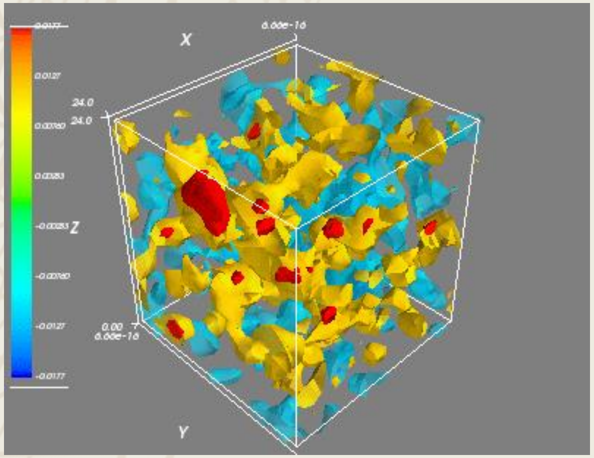
$t/a^2 = 0.1$



$t/a^2 = 0.2$



$t/a^2 = 0.3$



No isolated instanton structure…

backup

Details of NN

Trial 1~3

Layer	Filter size	Output size	Activation
input		$8^4 \times 1$	
convolution	$3^4 \times 5$	$8^4 \times 5$	Logistic
convolution	$3^4 \times 5$	$8^4 \times 5$	Logistic
average pooling	$8^4 \times 1$	5×1	
full connect		5	Logistic
full connect		1	

Trial 4

Layer	Output size	Activation
input	3	
full connect	5	Logistic
full connect	1	

Trial 5

Layer	Filter size	Output size	Activation
input		$12^d \mathrm{or} 8^d \times 3$	
convolution	$3^d \times 5$	$12^d \mathrm{or} 8^d \times 5$	Logistic
convolution	$3^d \times 5$	$12^d \mathrm{or} 8^d \times 5$	Logistic
convolution	$3^d \times 5$	$12^d \mathrm{or} 8^d \times 5$	Logistic
average pooling	$12^d \mathrm{or} 8^d \times 1$	5×1	
full connect		5	Logistic
full connect		1