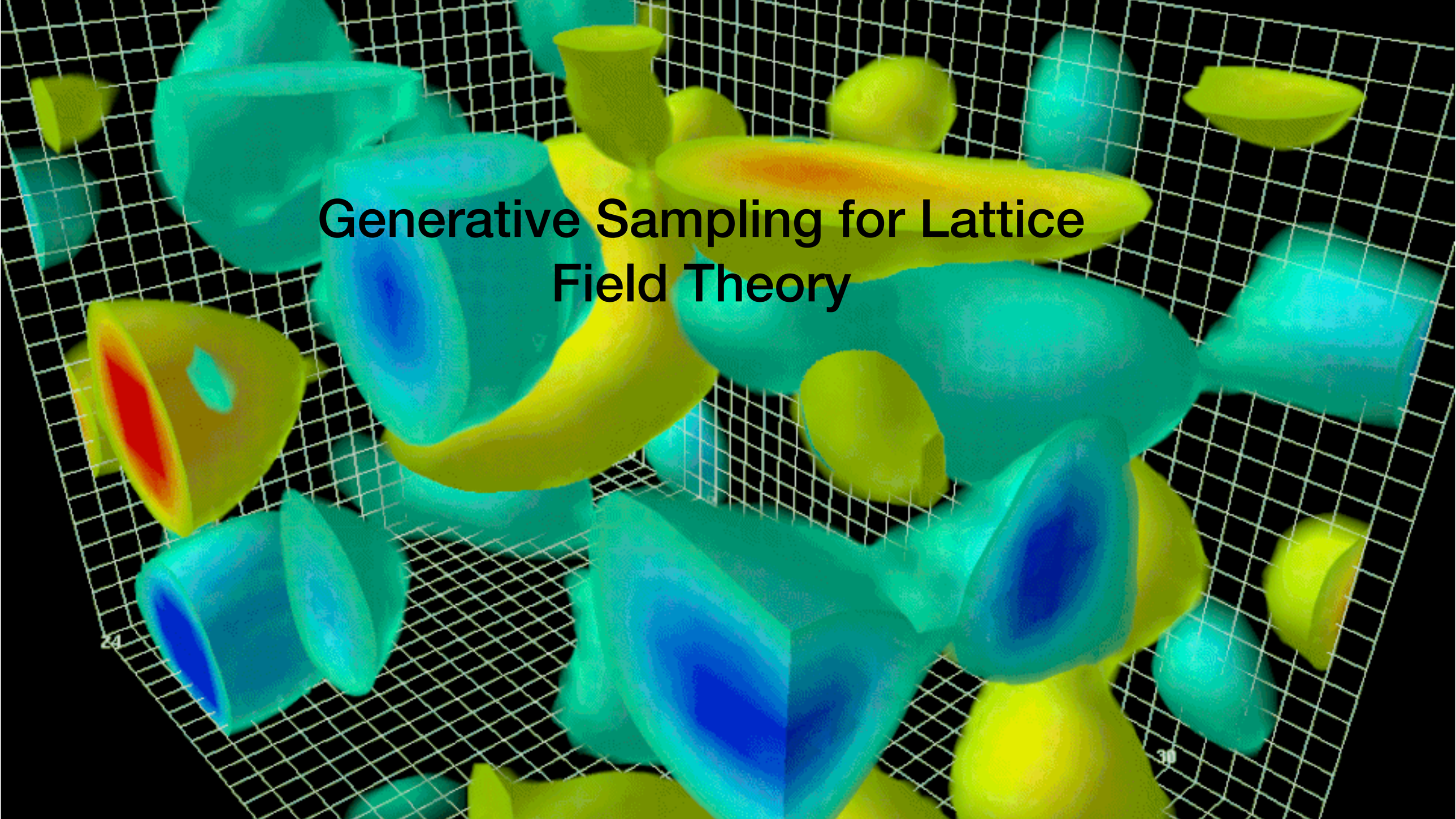
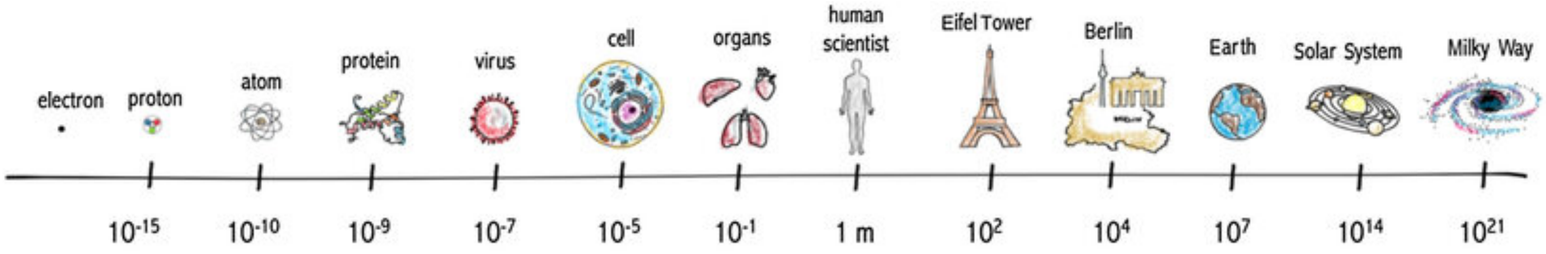
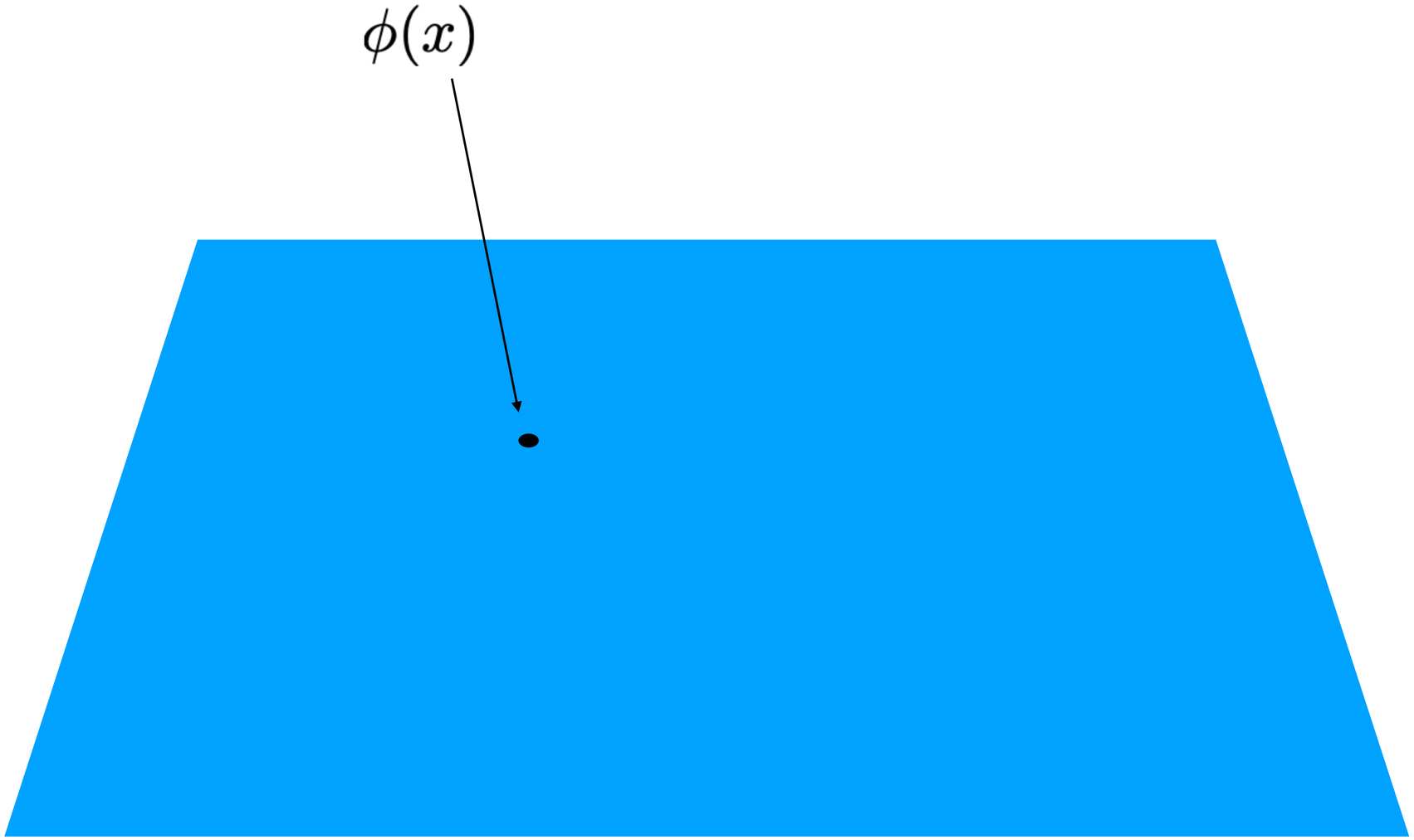
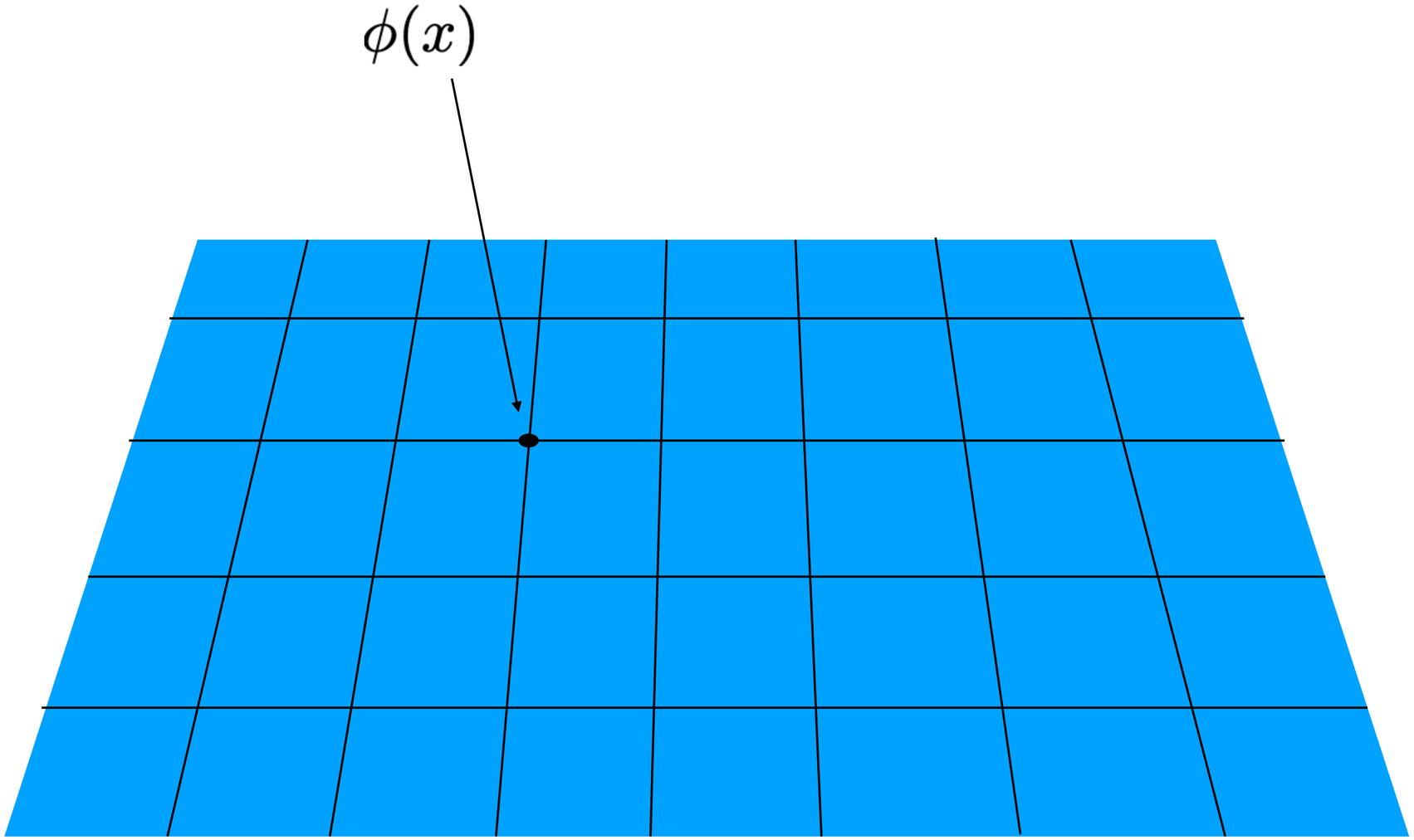
Generative Sampling for Lattice Field Theory

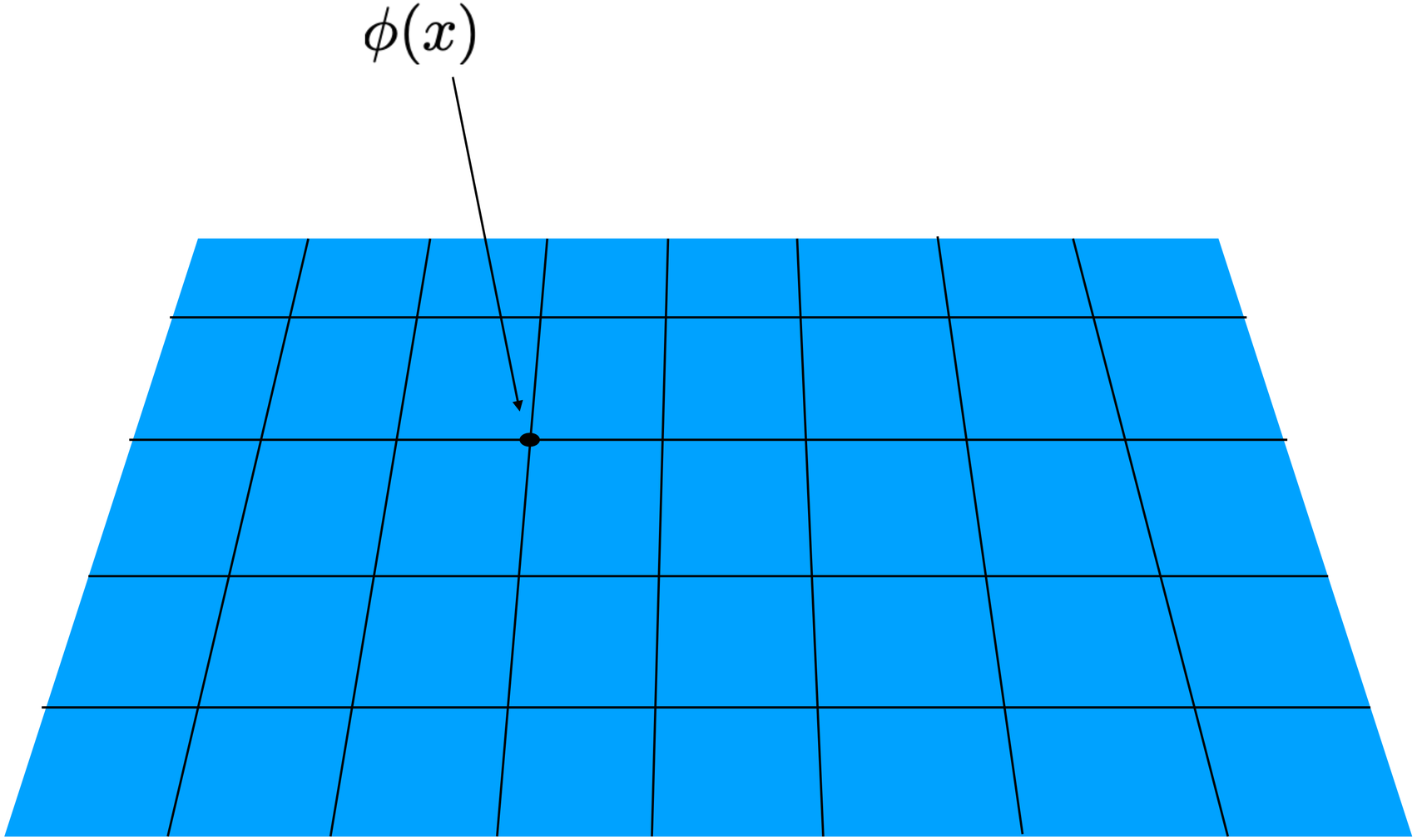



electron
proton
atom
protein
virus
cell
organs
human scientist
Eifel Tower
Berlin
Earth
Solar System
Milky Way
10^{-15}
10^{-10}
10^{-9}
10^{-7}
10^{-5}
10^{-1}
1 m
10^{2}
10^{4}
10^{7}
10^{14}
10^{21}


$\phi(x)$


$\phi(x)$

$$S(\phi) = T(\phi) + V(\phi)$$



Lattice **Quantum** Field Theory (LQFT)

$$p(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

Lattice **Quantum** Field Theory (LQFT)

$$p(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

$$\langle \mathcal{O}(\phi) \rangle$$

For example: $\mathcal{O}(\phi) = |\phi|, \phi^2, \ldots$

Lattice **Quantum** Field Theory (LQFT)

$$p(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

Lattice **Quantum** Field Theory (LQFT)

$$p(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

$$S(\phi) = \sum_{x \in \Lambda} a^2 \sum_{\hat{\mu}=1}^{4} \frac{(\phi(x + a\hat{\mu}) - \phi(x))^2}{a^2} + m^2\phi^2(x) + g\phi^4(x)$$

Lattice **Quantum** Field Theory (LQFT)

$$p(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

$$\langle \mathcal{O}(\phi) \rangle = \int D\phi \, p(\phi) \, \mathcal{O}(\phi)$$

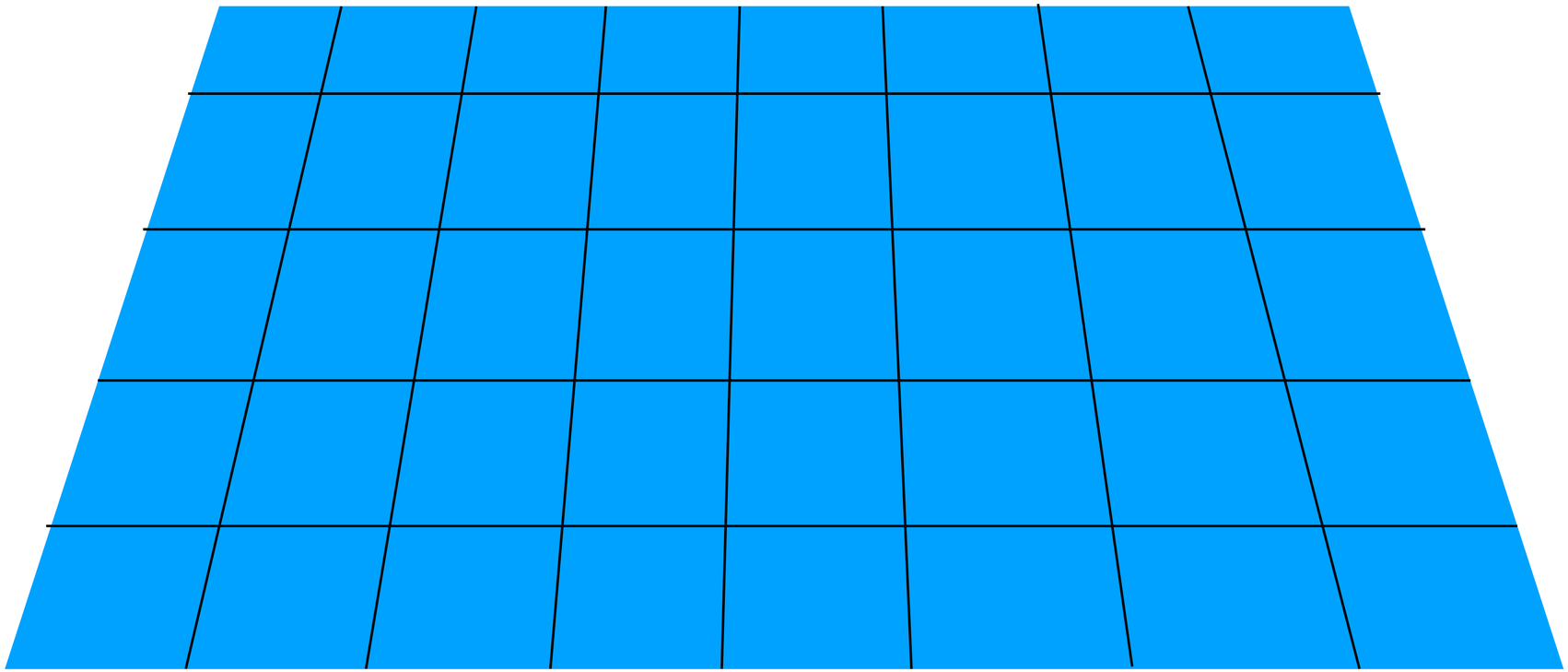
Lattice **Quantum** Field Theory (LQFT)

$$p(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

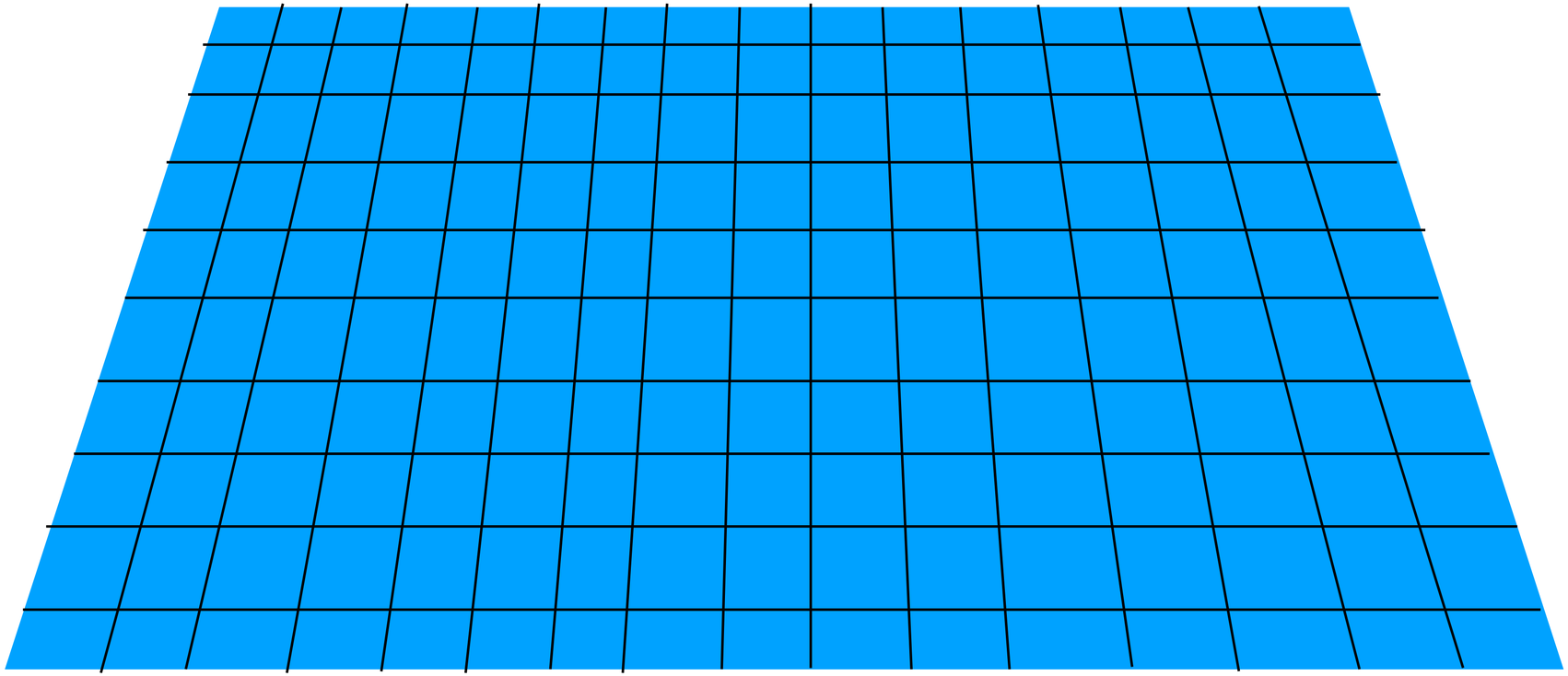
$$\langle \mathcal{O}(\phi) \rangle = \int D\phi \, p(\phi) \, \mathcal{O}(\phi) \quad \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(\phi_i)$$

generated with MCMC

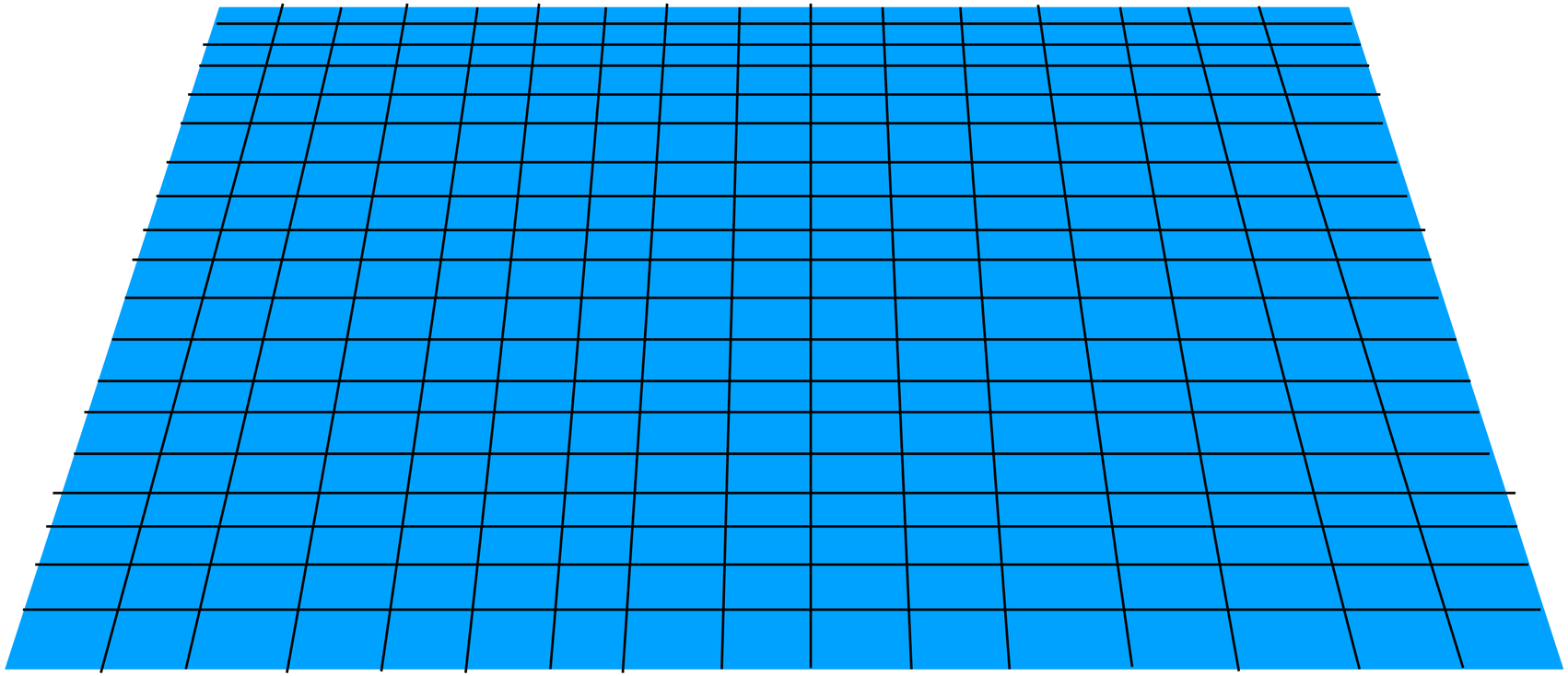
Nature is not a Lattice (as far as we can tell)



Nature is not a Lattice (as far as we can tell)

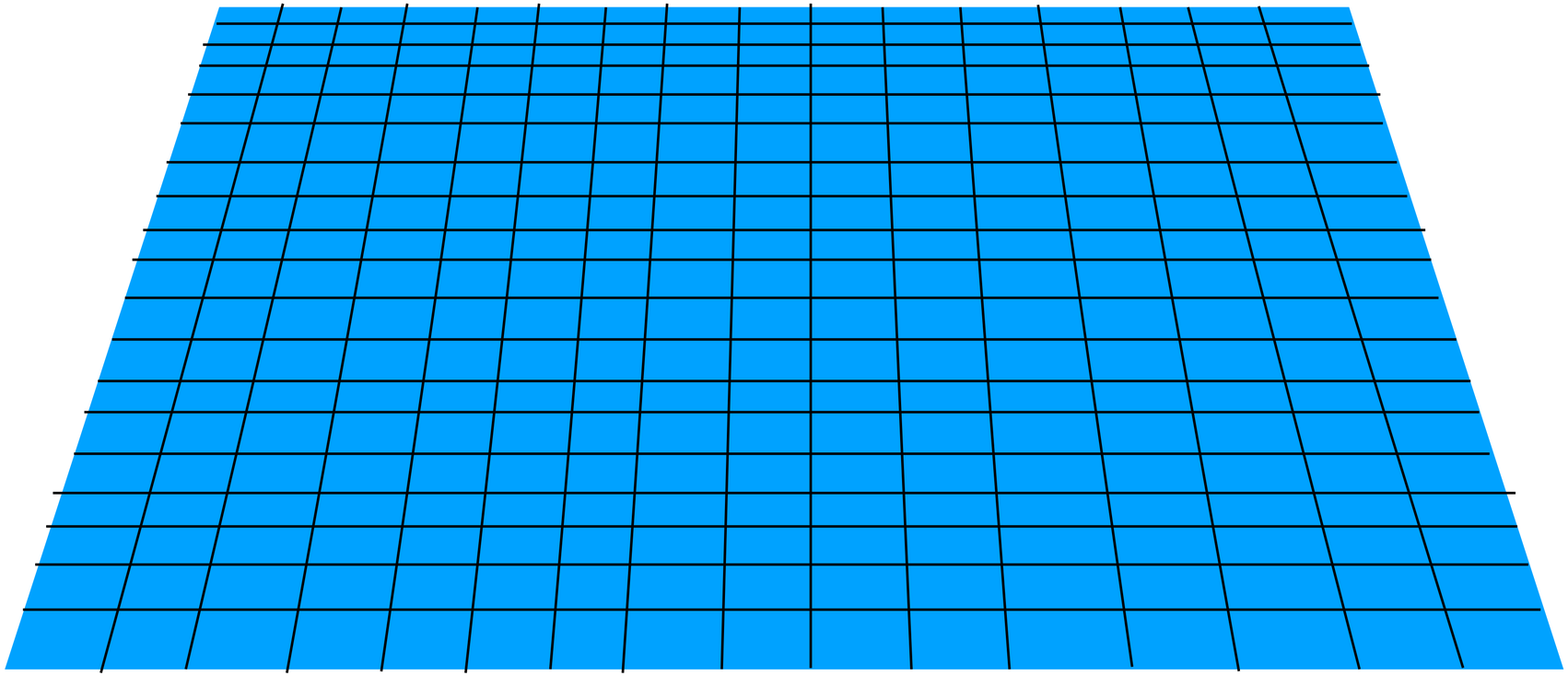


Nature is not a Lattice (as far as we can tell)



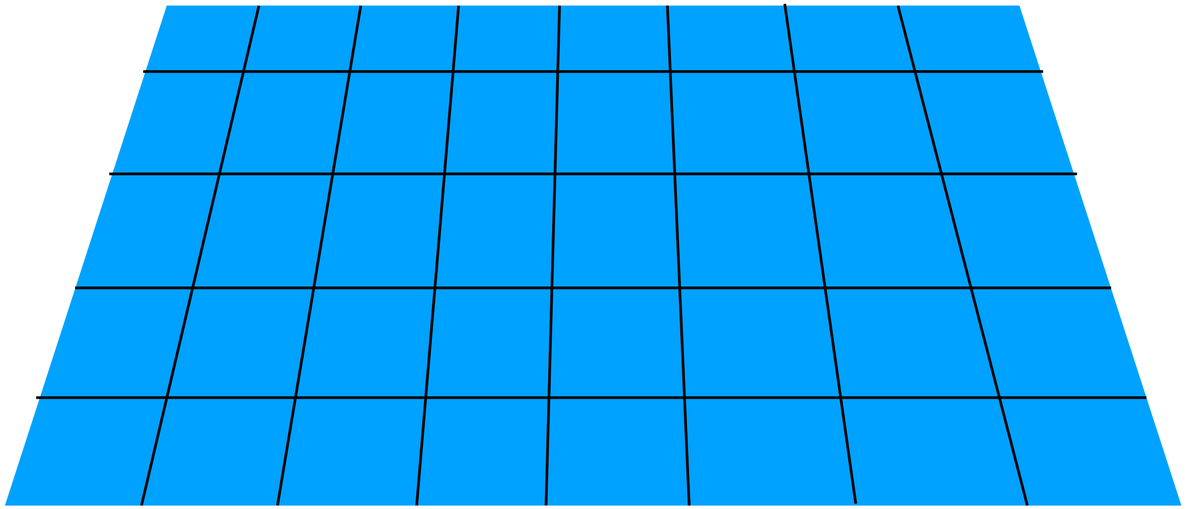
Continuum Limit (intuitive):

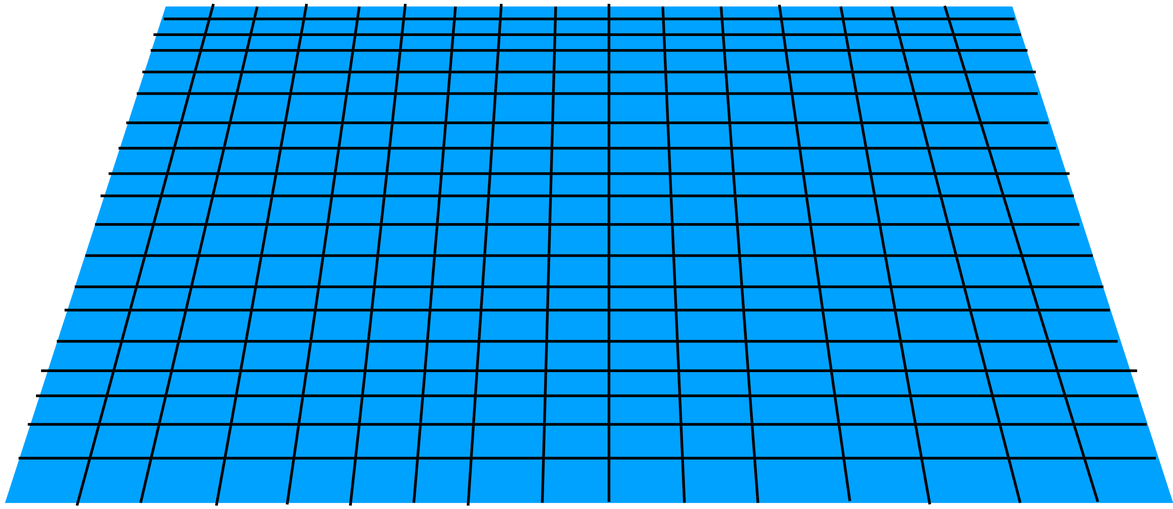
$$a \to 0 \qquad N \to \infty \qquad \text{such that} \qquad a^4 N^4 = V \approx \text{const.}$$



Critical Slowing Down:

$$C(x-y) = \langle \mathcal{O}(x)\mathcal{O}(y) \rangle \sim \exp\left(-\frac{-|x-y|}{\xi_{\mathcal{O}}}\right)$$





Critical Slowing Down:

$$C(x-y) = \langle \mathcal{O}(x)\mathcal{O}(y) \rangle \sim \exp\left(-\frac{-|x-y|}{\xi_{\mathcal{O}}}\right)$$

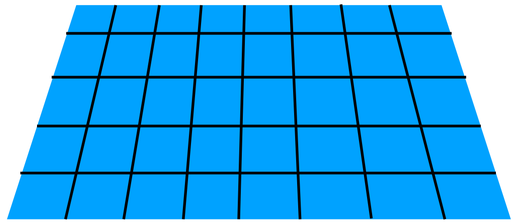
$$Var(\hat{C}) = \frac{\sigma_C}{\frac{N}{2\tau_{\mathcal{O},int}}}$$

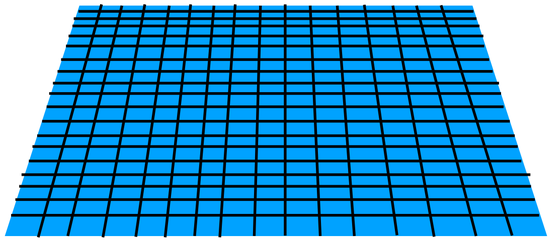
$$\tau_{\mathcal{O},int} \sim \xi_{\mathcal{O}}^{z}$$

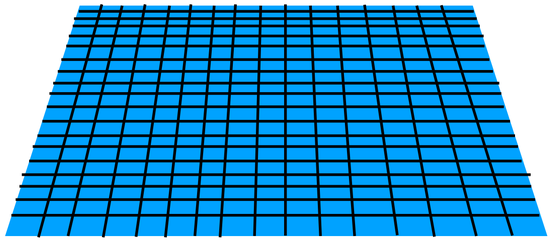
Critical exponent

Intermitent Summary

$$\langle \mathcal{O}(\phi) \rangle = \int D\phi \, p(\phi) \, \mathcal{O}(\phi) \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(\phi_i)$$







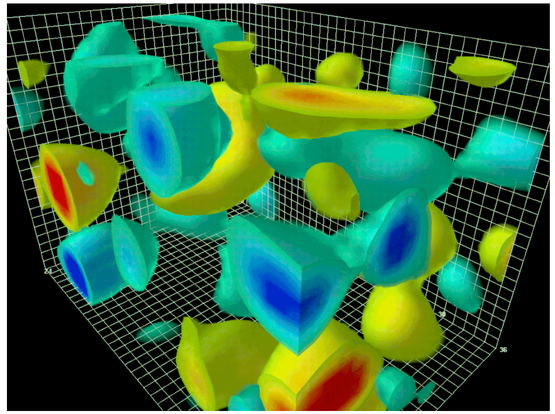
Critical slowing down:

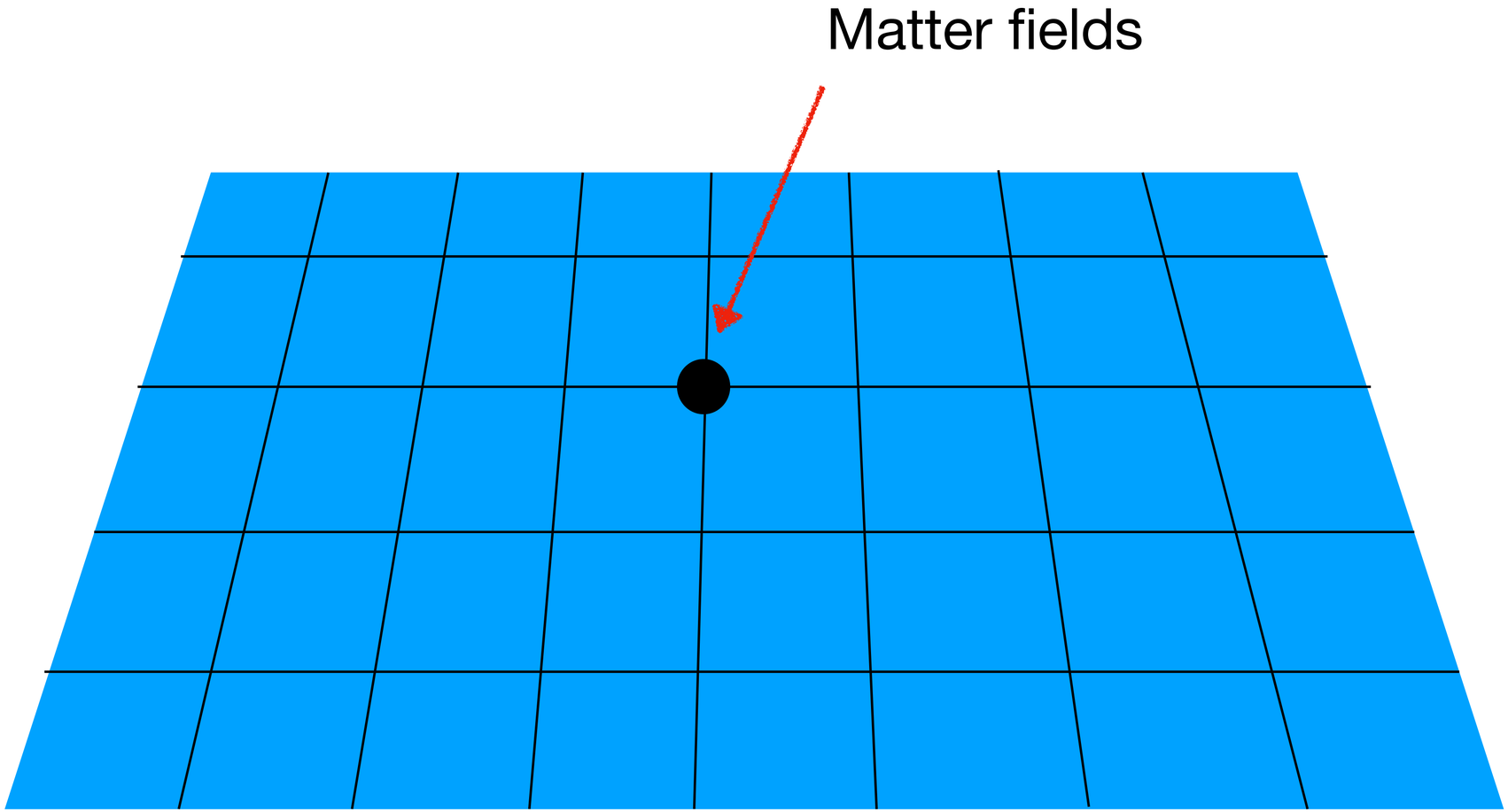
exponentially growing number of samples needed for the same variance as we take the continuum limit.

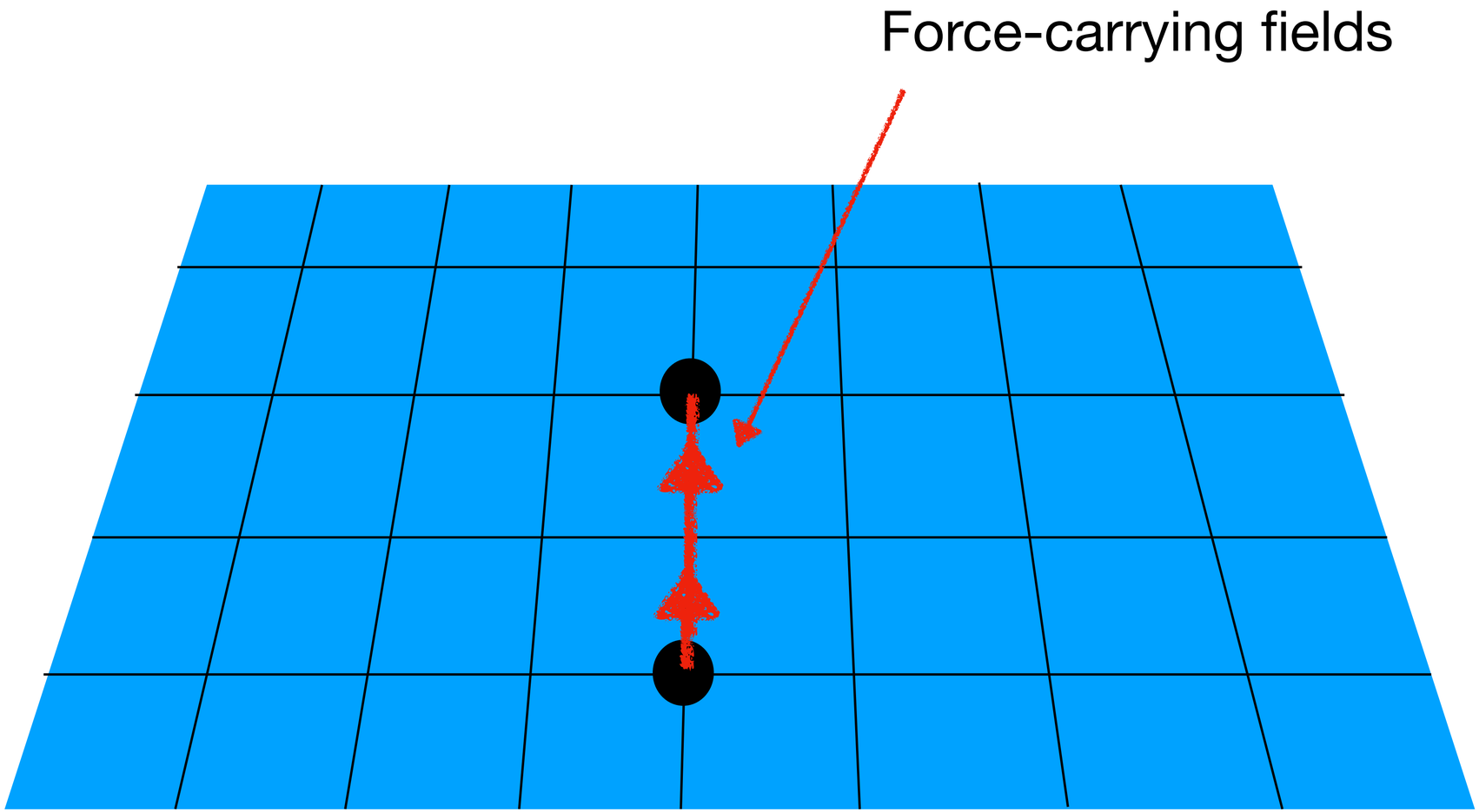
We will now discuss two examples of LQFT:

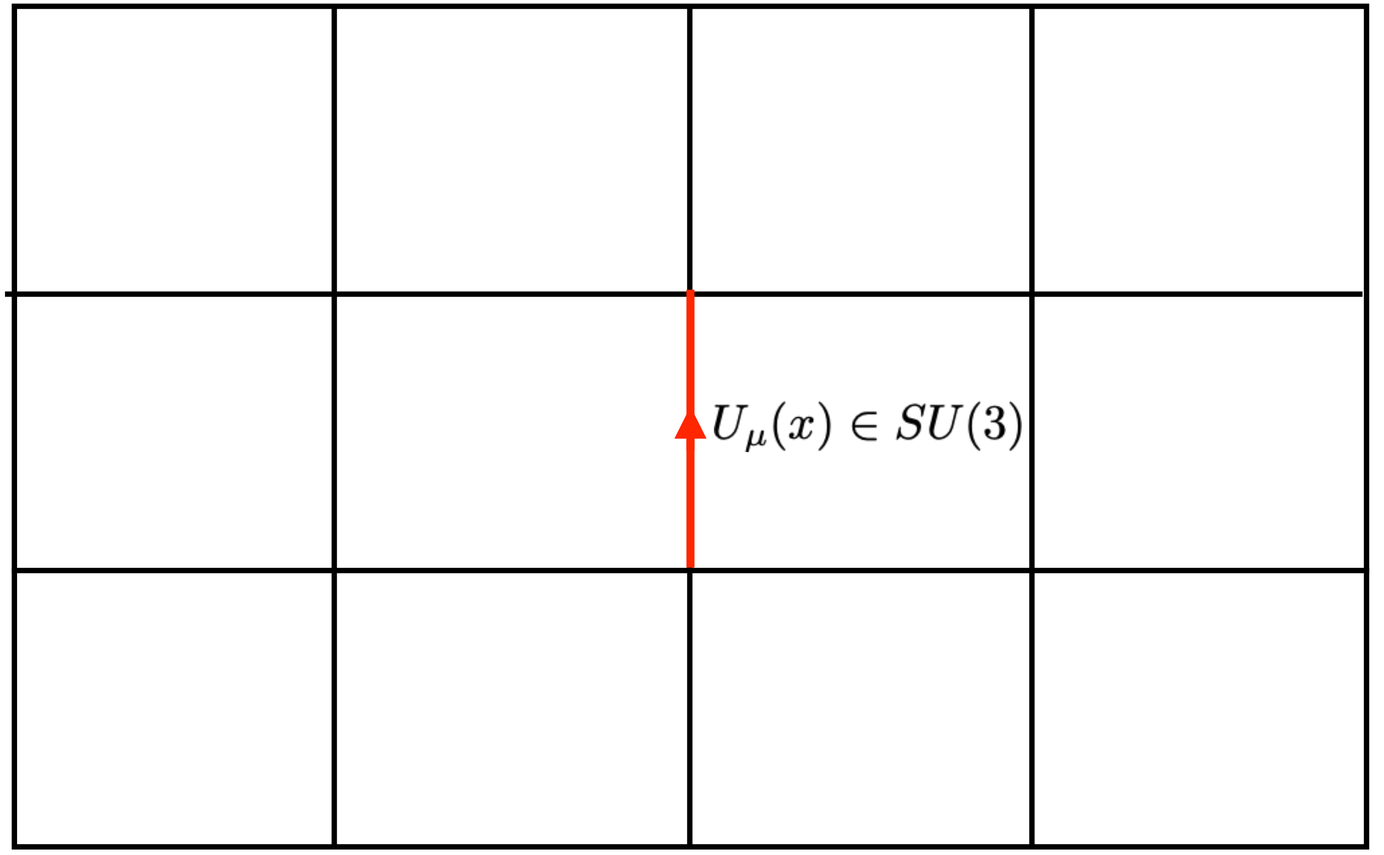
- Lattice Gauge Theory
- Proteins

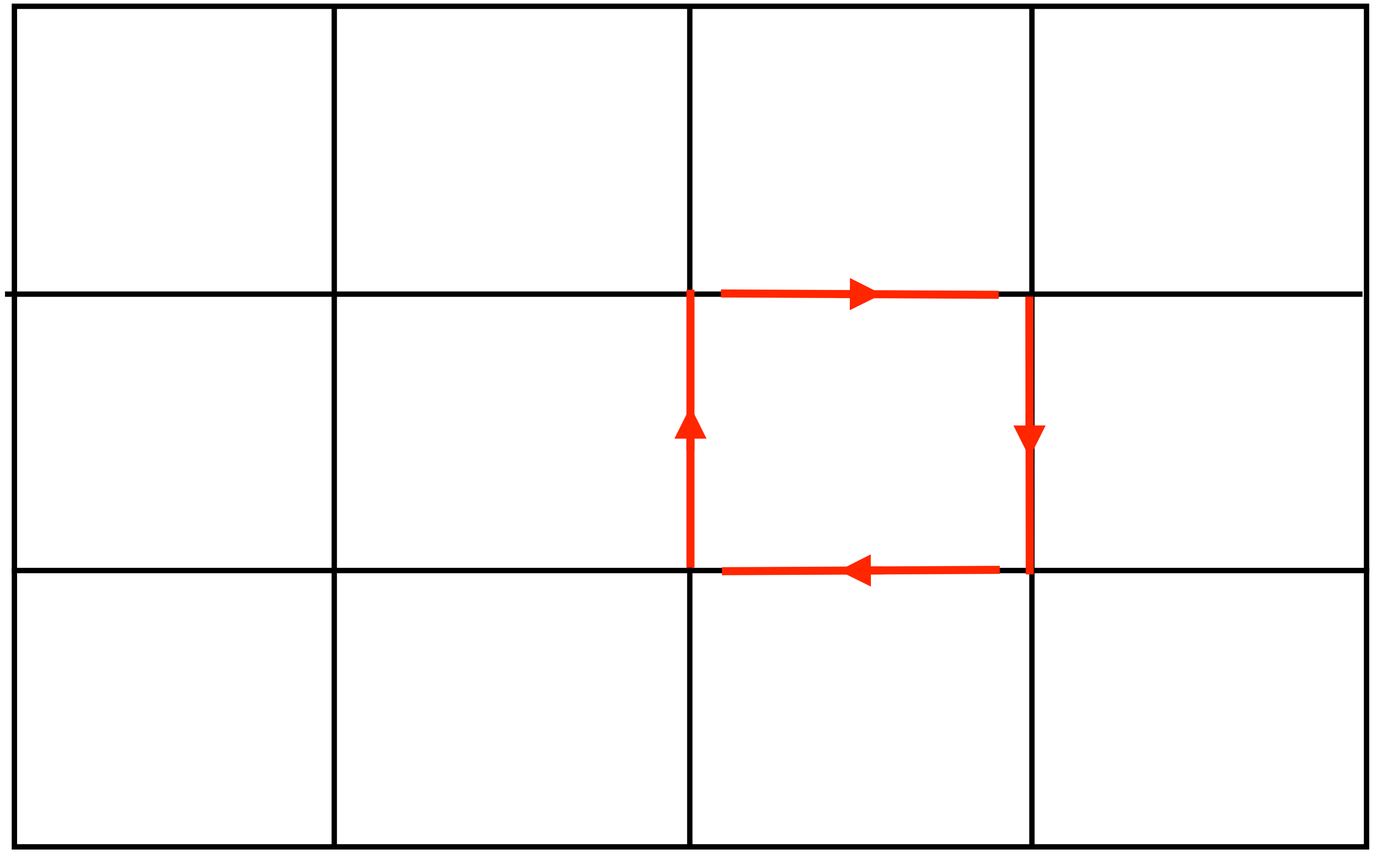
Example I: Lattice Gauge Theory




Matter fields

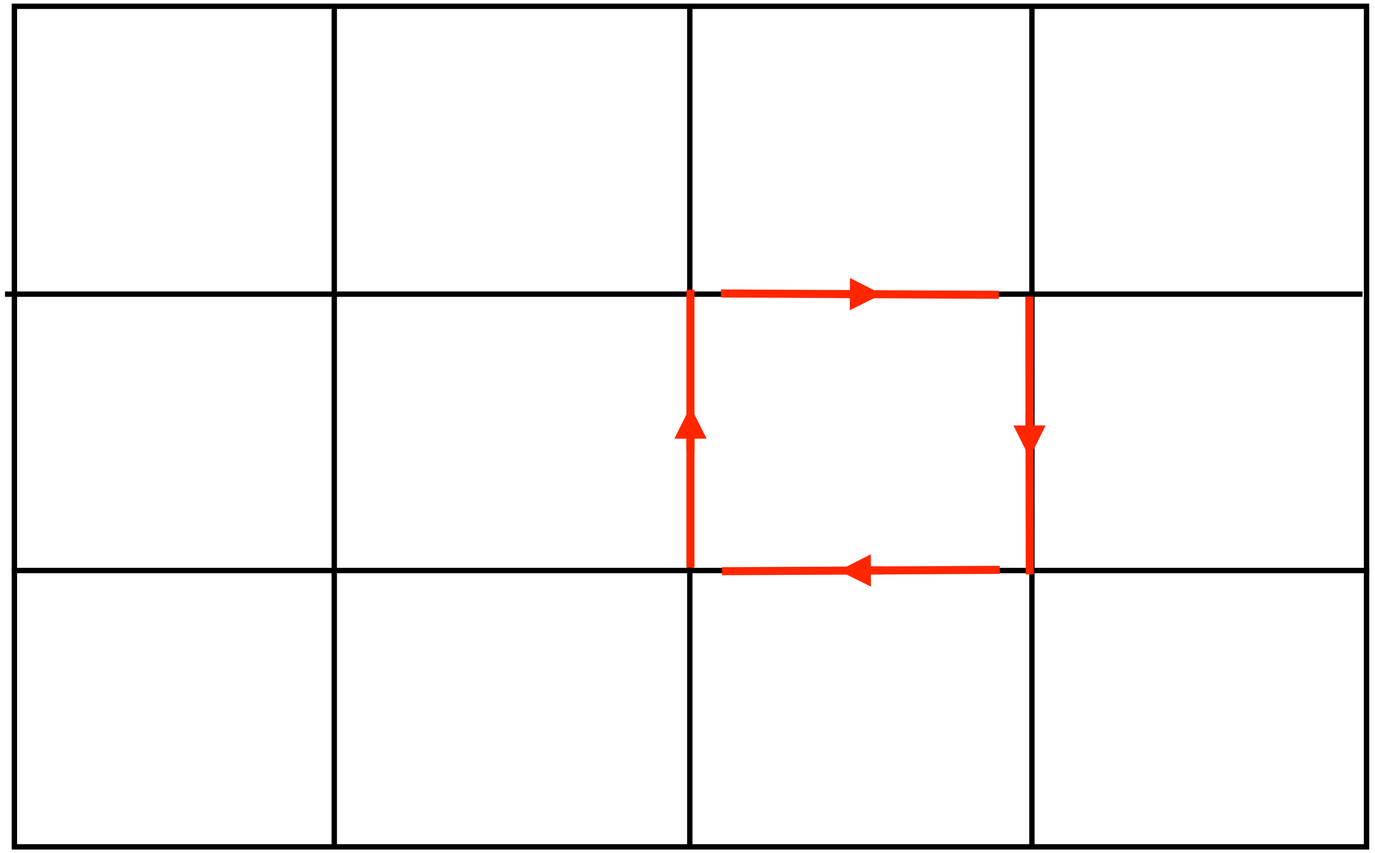

Force-carrying fields


$U_\mu(x) \in SU(3)$



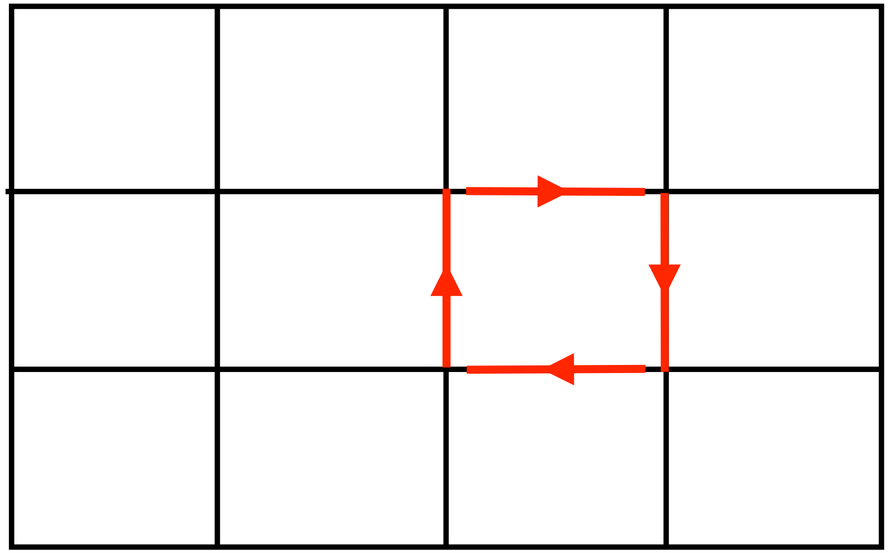
$$U_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu(x+\hat{\nu})^\dagger U_\nu(x)^\dagger$$

$$S(U) = -\frac{\beta}{3} \sum_x \sum_{\mu<\nu} \mathrm{Re}\,\mathrm{Tr}\, U_{\mu\nu}(x)$$



$$U_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu(x+\hat{\nu})^\dagger U_\nu(x)^\dagger$$

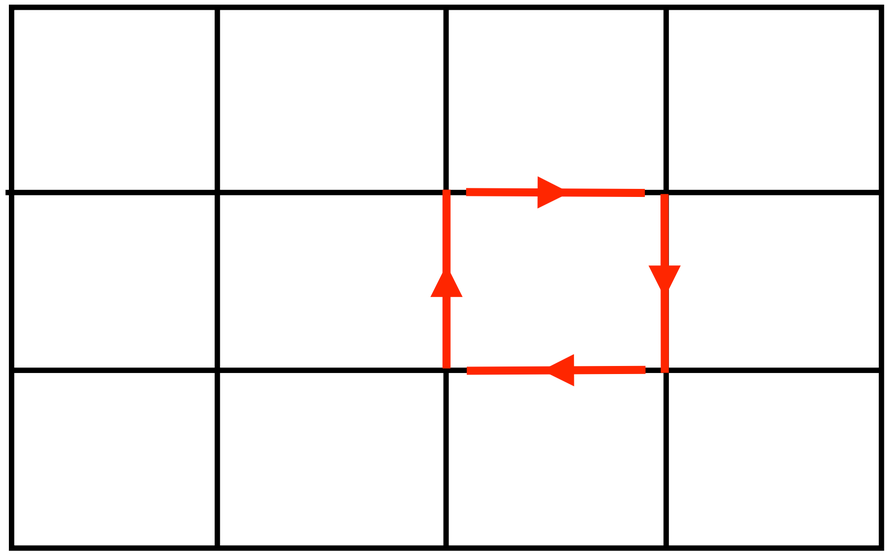
$$U_\mu(x) \to \Omega(x)\, U_\mu(x)\, \Omega(x+\hat{\mu})^\dagger$$



$$U_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu(x+\hat{\nu})^\dagger U_\nu(x)^\dagger$$

$$U_\mu(x) \to \Omega(x)\, U_\mu(x)\, \Omega(x+\hat{\mu})^\dagger$$

$$U_{\mu\nu}(x) \to \Omega(x)\, U_{\mu\nu}(x)\, \Omega(x)^\dagger$$

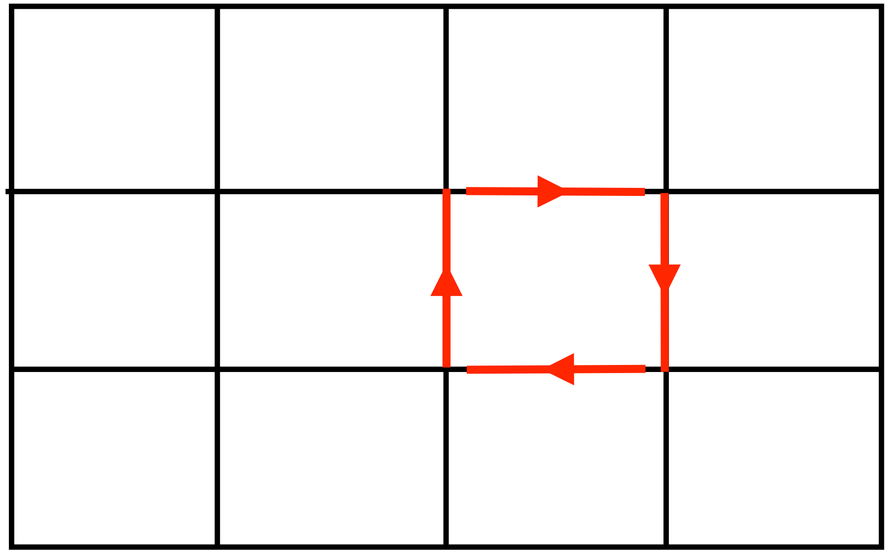


$$U_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu(x+\hat{\nu})^\dagger U_\nu(x)^\dagger$$

$$U_\mu(x) \to \Omega(x)\, U_\mu(x)\, \Omega(x+\hat{\mu})^\dagger$$

$$U_{\mu\nu}(x) \to \Omega(x)\, U_{\mu\nu}(x)\, \Omega(x)^\dagger$$

$$\mathrm{Tr}\, U_{\mu\nu}(x) \to \mathrm{Tr}\, \Omega(x)\, U_{\mu\nu}(x)\, \Omega(x)^\dagger = \mathrm{Tr} U_{\mu\nu}(x)$$



$$U_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu(x+\hat{\nu})^\dagger U_\nu(x)^\dagger$$

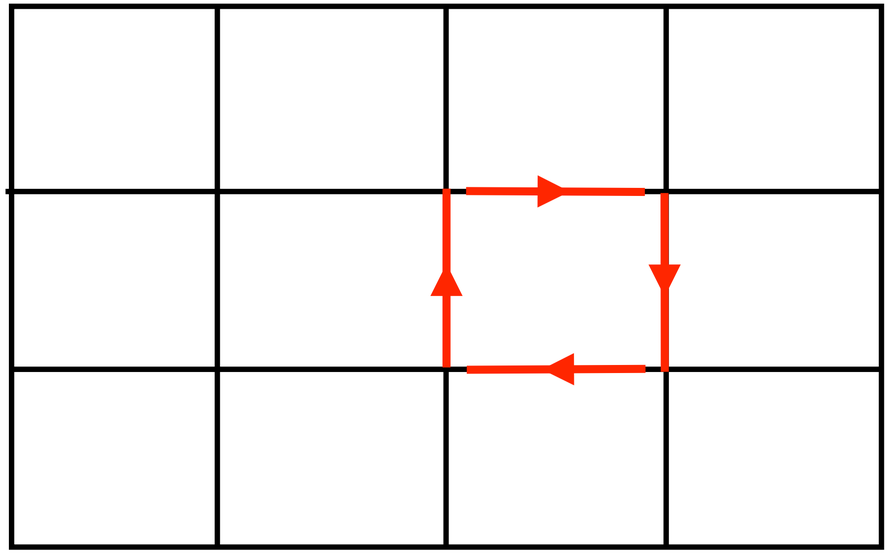
$$U_\mu(x) \to \Omega(x)\, U_\mu(x)\, \Omega(x+\hat{\mu})^\dagger$$

$$U_{\mu\nu}(x) \to \Omega(x)\, U_{\mu\nu}(x)\, \Omega(x)^\dagger$$

$$\mathrm{Tr}\, U_{\mu\nu}(x) \to \mathrm{Tr}\, \Omega(x)\, U_{\mu\nu}(x)\, \Omega(x)^\dagger = \mathrm{Tr} U_{\mu\nu}(x)$$

$$S(U) = -\frac{\beta}{3} \sum_x \sum_{\mu<\nu} \mathrm{Re}\, \mathrm{Tr}\, U_{\mu\nu}(x)$$

is invariant!



$$U_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu(x+\hat{\nu})^\dagger U_\nu(x)^\dagger$$

Action $S(U) = -\frac{\beta}{3} \sum_x \sum_{\mu<\nu} \mathrm{Re}\,\mathrm{Tr}\, U_{\mu\nu}(x)$ is gauge invariant:

$$U_\mu(x) \to \Omega(x)\, U_\mu(x)\, \Omega(x+\hat{\mu})^\dagger$$

Action $S(U) = -\frac{\beta}{3} \sum_x \sum_{\mu<\nu} \mathrm{Re}\,\mathrm{Tr}\, U_{\mu\nu}(x)$ is gauge invariant:

$$U_\mu(x) \to \Omega(x)\, U_\mu(x)\, \Omega(x+\hat{\mu})^\dagger$$

Symmetry group is huge: $8 \times L^4$ free parameters!

Action $S(U) = -\frac{\beta}{3} \sum_x \sum_{\mu<\nu} \mathrm{Re}\,\mathrm{Tr}\, U_{\mu\nu}(x)$ is gauge invariant:

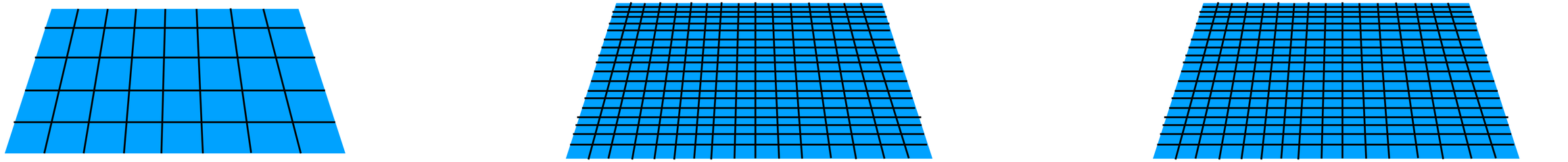
$$U_\mu(x) \to \Omega(x)\, U_\mu(x)\, \Omega(x+\hat{\mu})^\dagger$$

Symmetry group is huge: $8 \times L^4$ free parameters!

$$L = 16 \longrightarrow 500k$$

Intermitent Summary

$$\langle \mathcal{O}[U] \rangle = \int D[U]\, p(U)\, \mathcal{O}(U) \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(U_i)$$

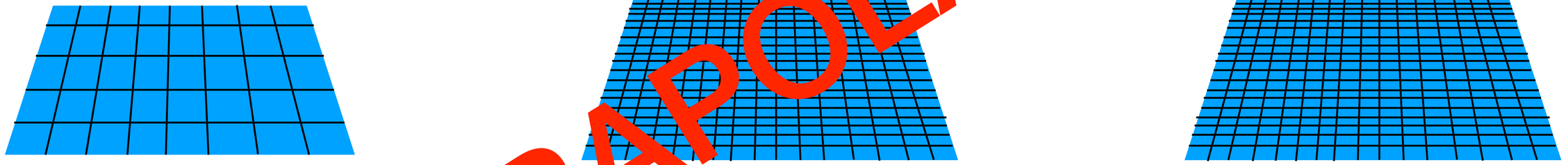


Critical slowing down:

exponentially growing number of samples needed for the same variance as we take the continuum limit.

Intermitent Summary

$$\langle \mathcal{O}[U] \rangle = \int D[U]\, p(U)\, \mathcal{O}(U) \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(U_i)$$

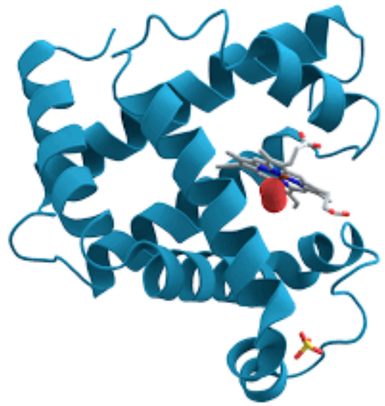


EXTRAPOLATION!

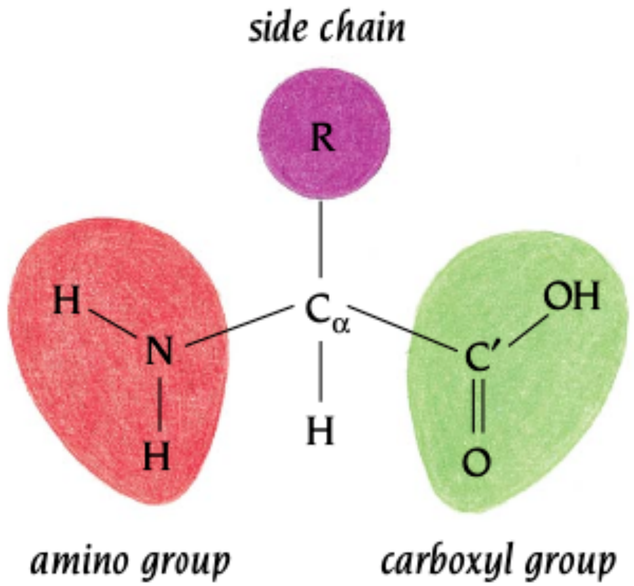
Critical slowing down:

exponentially growing number of samples needed for the same variance as we take the continuum limit.

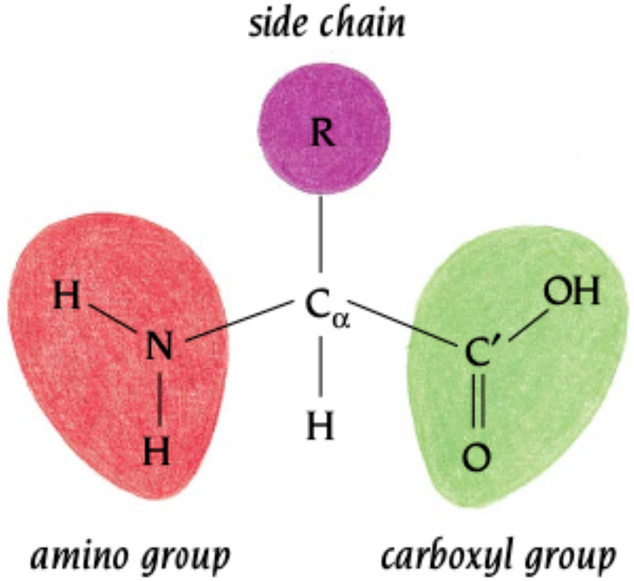
Example II:
Proteins

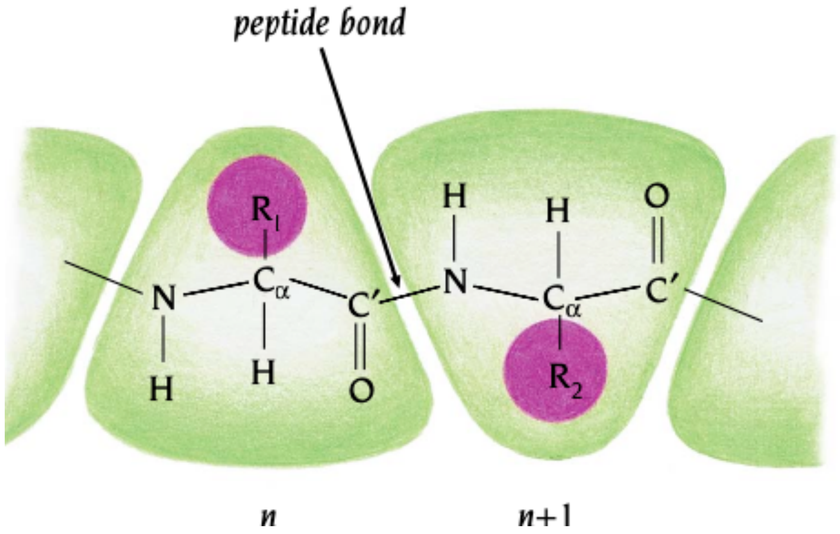


Aminoacid

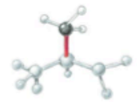


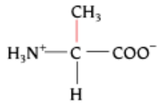
Aminoacid


side chain
R
H
N
H
C_α
H
C′
OH
O
amino group
carboxyl group

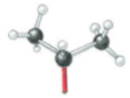

peptide bond
R_1
N
H
C_α
H
C′
O
H
N
H
C_α
R_2
O
C′
n
n+1

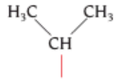
(a) Hydrophobic amino acids



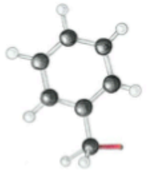


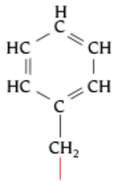
A Ala, Alanine



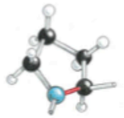


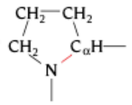
V Val, Valine



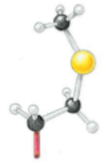


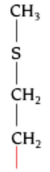
F Phe, Phenylalanine



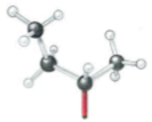


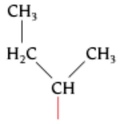
P Pro, Proline



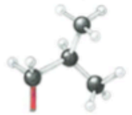


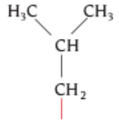
M Met, Methionine





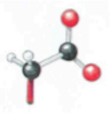
I Ile, Isoleucine

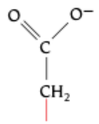




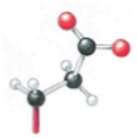
L Leu, Leucine

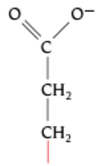
(b) Charged amino acids



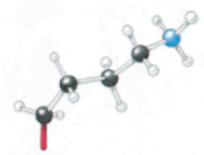


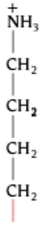
D Asp, Aspartic acid



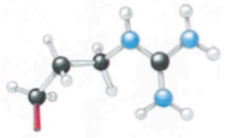


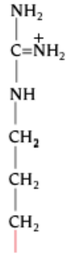
E Glu, Glutamic acid





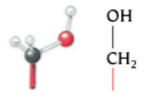
K Lys, Lysine



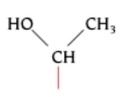


R Arg, Arginine

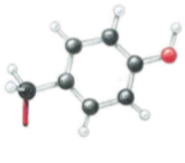
(c) Polar amino acids

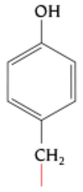


S Ser, Serine

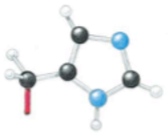


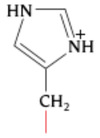
T Thr, Threonine





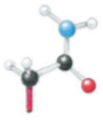
Y Tyr, Tyrosine

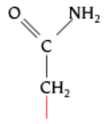




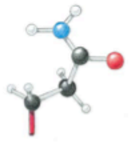
H His, Histidine

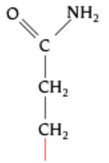
C Cys, Cysteine



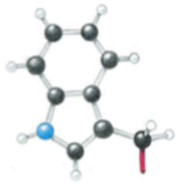


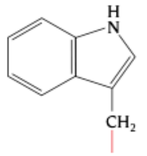
N Asn, Asparagine





Q Gln, Glutamine

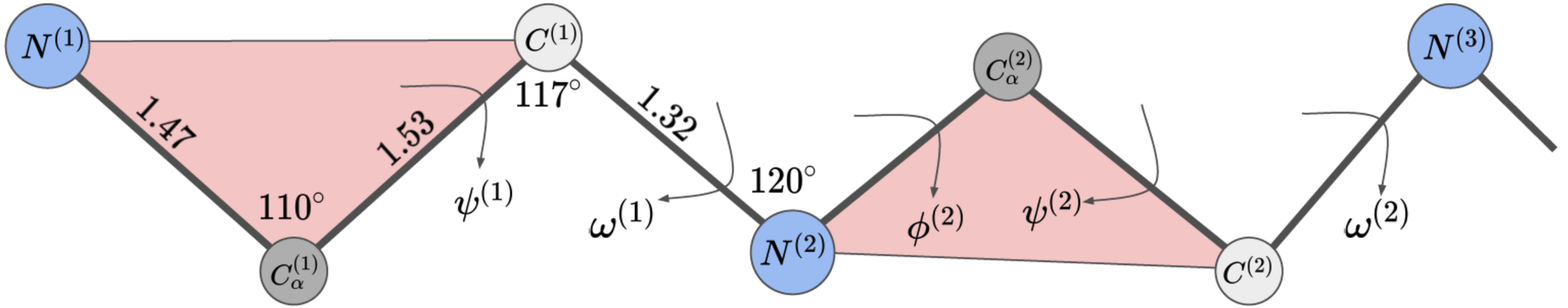


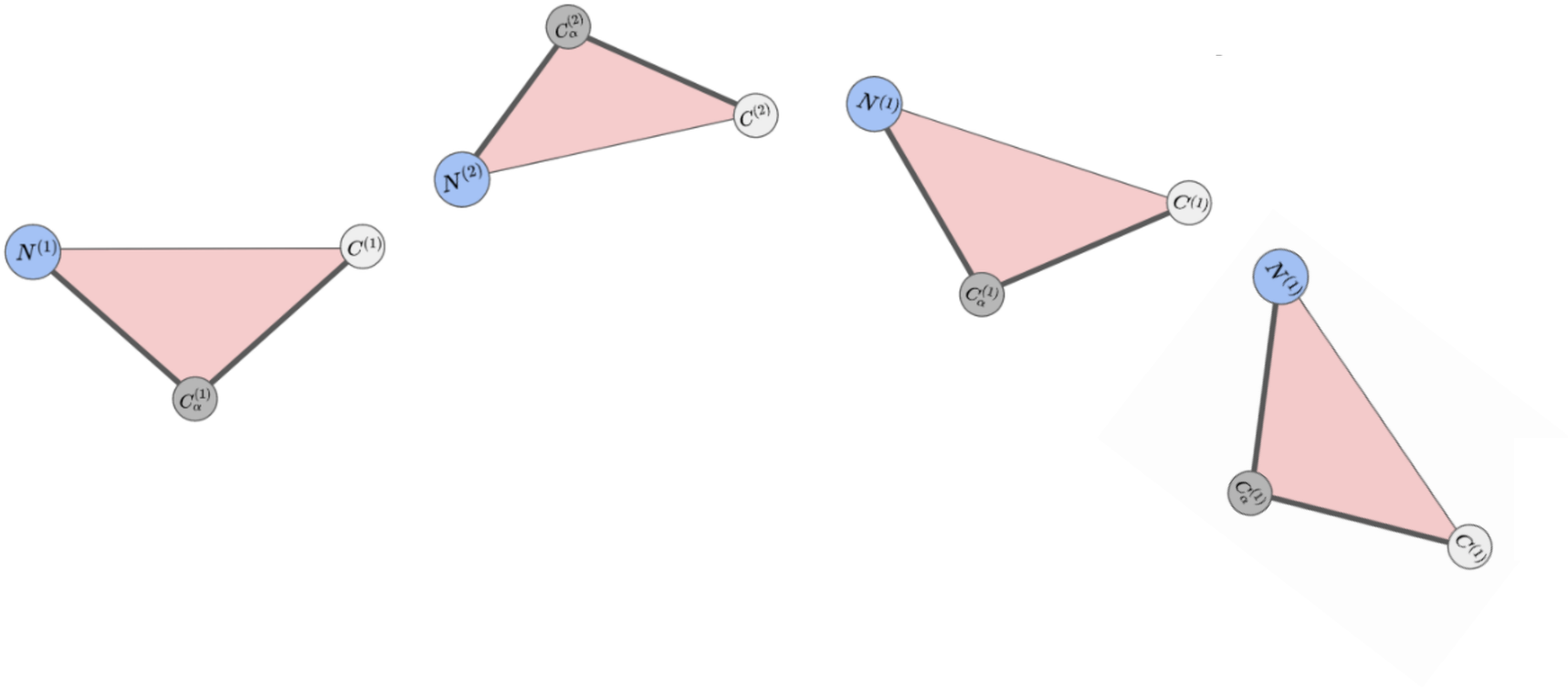


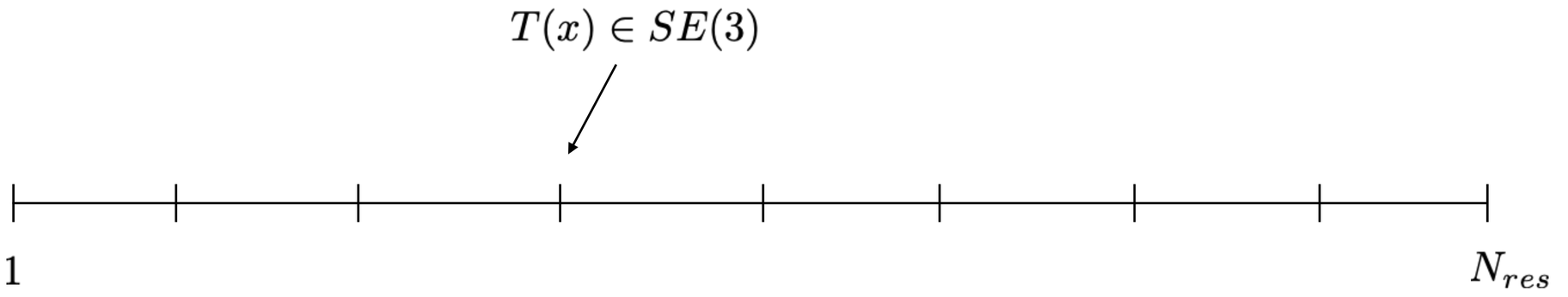
W Trp, Tryptophan

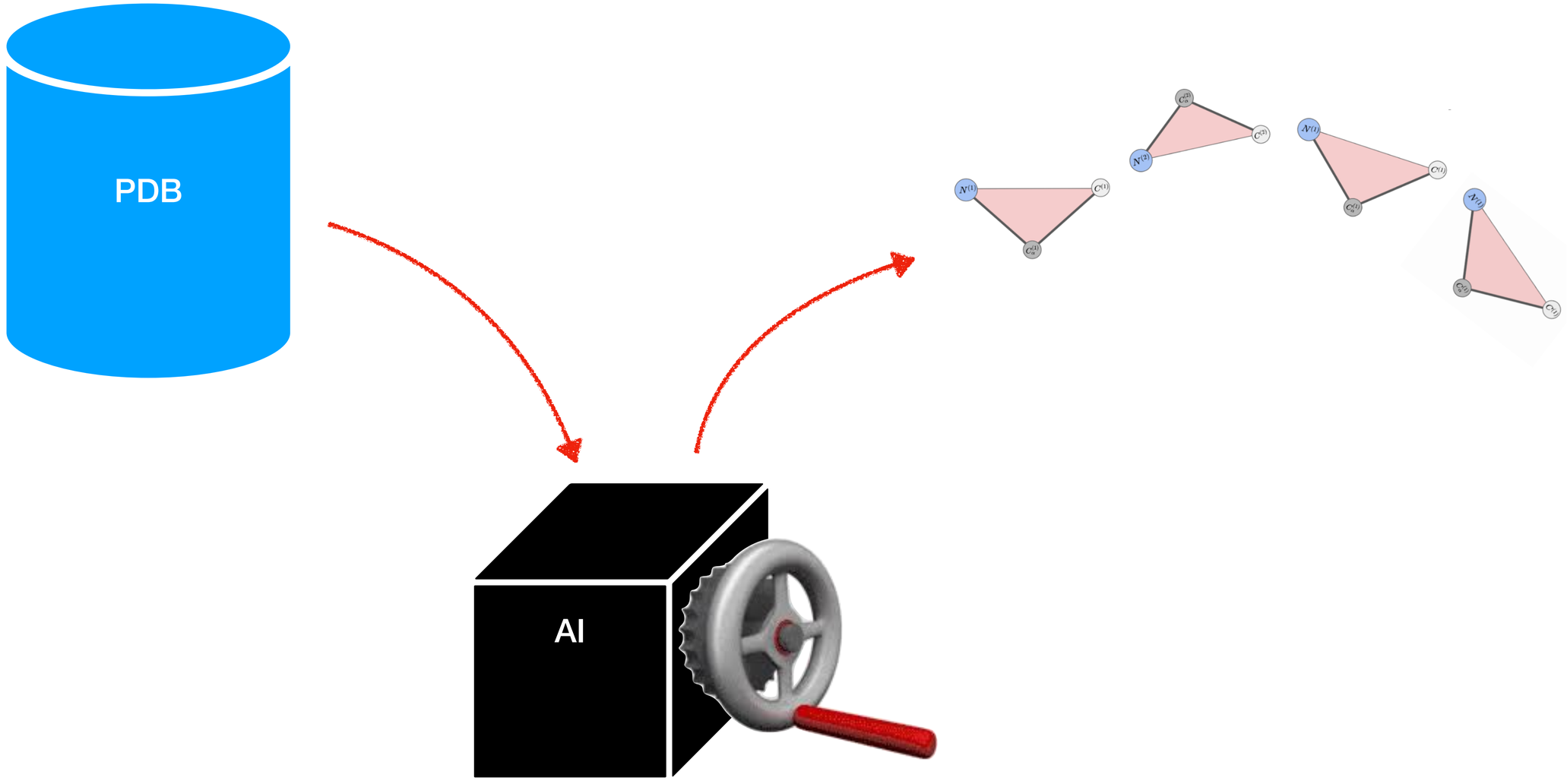
(d) Glycine

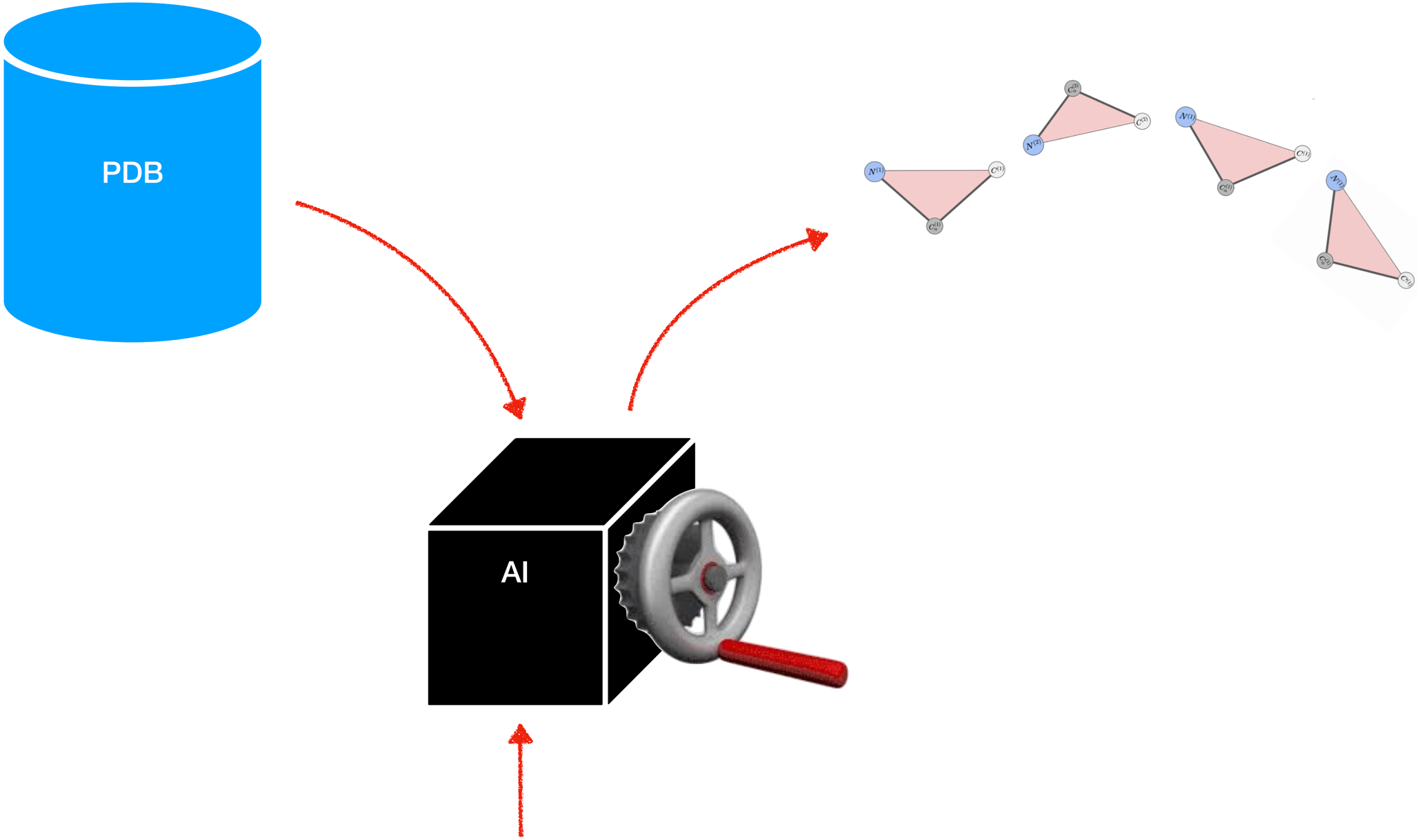
G Gly, Glycine

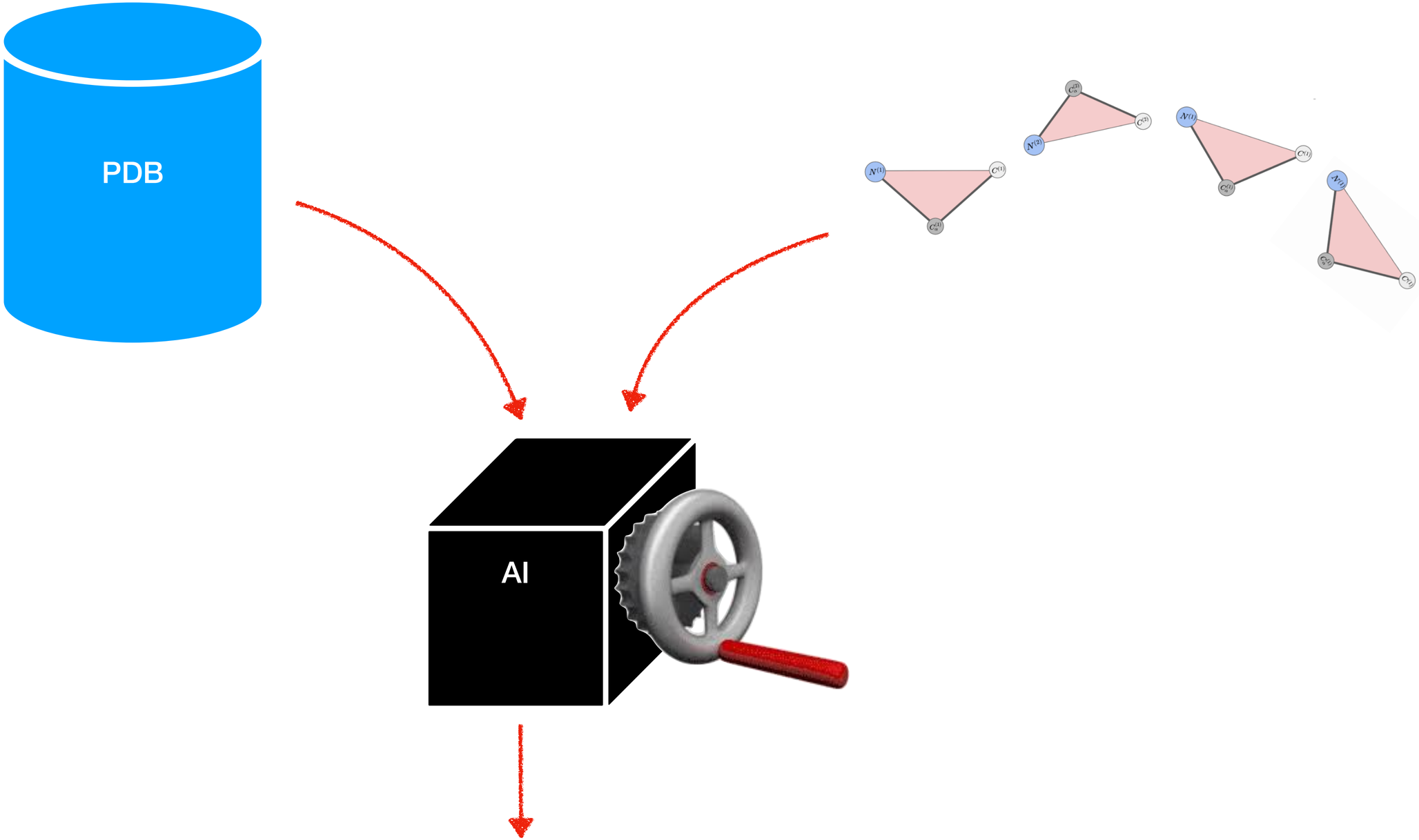

$N^{(1)}$
$C^{(1)}$
$C_\alpha^{(2)}$
$N^{(3)}$
1.47
1.53
117°
1.32
120°
110°
$\psi^{(1)}$
$\omega^{(1)}$
$\phi^{(2)}$
$\psi^{(2)}$
$N^{(2)}$
$C^{(2)}$
$\omega^{(2)}$
$C_\alpha^{(1)}$


$N^{(1)}$
$C^{(1)}$
$C_\alpha^{(1)}$
$C_\alpha^{(2)}$
$N^{(2)}$
$C^{(2)}$
$N^{(1)}$
$C^{(1)}$
$C_\alpha^{(1)}$
$N^{(1)}$
$C_\alpha^{(1)}$
$C^{(1)}$


$T(x) \in SE(3)$
1
N_{res}


PDB
AI


PDB
AI
ERTTIMDPTLGIEGQAYHAGTYSFYFCKMKRANNKDNILFFFEEVTWGHWYCLFVSGYVN
WIEEHCPENNNTHCVAVHSNDVPLGHIDHEQLGDSFYRQHILWKSFVMCKARDKAVRHEQ
VAVIDPKMYKVMAEDNSMTFQIMCPDWRWQARDKKMKVELGKYAIFLHHHVWNIYSSFFW
RPHMRIPLHLLRTAWFYFPLKLLEANKSKDYLYKHYEMTPWIACEDEGRHPWTGRSATTM


PDB
AI
ERTTIMDPTLGIEGQAYHAGTYSFYFCKMKRANNKDNILFFFEEVTWGHWYCLFVSGYVN
WIEEHCPENNNTHCVAVHSNDVPLGHIDHEQLGDSFYRQHILWKSFVMCKARDKAVRHEQ
VAVIDPKMYKVMAEDNSMTFQIMCPDWRWQARDKKMKVELGKYAIFLHHHVWNIYSSFFW
RPHMRIPLHLLRTAWFYFPLKLLEANKSKDYLYKHYEMTPWIACEDEGRHPWTGRSATTM

$$p(T) = \frac{1}{Z} e^{-\beta S(T)} \longrightarrow \langle \mathcal{O}(T) \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}_i$$

Differences to Lattice Gauge Theory:

- Exact action is unknown. We only have crude approximations.
- No gauge symmetry
- No need to necessarily extrapolate, e.g. structure prediction
- Joint challenge: landscape rugged and thus large autocorrelation.

Training

Reverse KL: $KL(q,p) = \langle \log q(U) + S(U) \rangle_q + \text{const.}$

Forward KL: $KL(q,p) = \langle \log q(U) \rangle_p + \text{const.}$

Training

Reverse KL: $KL(q,p) = \langle \log q(U) + S(U) \rangle_q + \text{const.}$

$\longrightarrow$ Only action needed. $\longrightarrow$ Relies on self-sampling.

Forward KL: $KL(q,p) = \langle \log q(U) \rangle_p + \text{const.}$

Training

Reverse KL: $KL(q,p) = \langle \log q(U) + S(U) \rangle_q + \text{const.}$

$\longrightarrow$ Only action needed. $\longrightarrow$ Relies on self-sampling.

Forward KL: $KL(q,p) = \langle \log q(U) \rangle_p + \text{const.}$

$\longrightarrow$ No action but samples needed.

Training [edit]

AlphaZero was made & trained by it simply playing against itself multiple(many) times, using 5,000 first-generation TPUs to generate the games and 64 second-generation TPUs to train the neural networks. Training took several days, totaling about 41 TPU-years. In parallel, the in-training AlphaZero was periodically matched against its benchmark (Stockfish, Elmo, or AlphaGo Zero) in brief one-second-per-move games to determine how well the training was progressing. DeepMind judged that AlphaZero's performance exceeded the benchmark after around four hours of training for Stockfish, two hours for Elmo, and eight hours for AlphaGo Zero.[2]

Architectures

Attention is all we need?

Architectures

Proteins

Algorithm 22 Invariant point attention (IPA)

def InvariantPointAttention($\{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{T_i\}, N_{\text{head}} = 12, c = 16, N_{\text{query points}} = 4, N_{\text{point values}} = 8$) :

$\mathbf{q}_i^h, \mathbf{k}_i^h, \mathbf{v}_i^h = \text{LinearNoBias}(\mathbf{s}_i)$ $\quad \mathbf{q}_i^h, \mathbf{k}_i^h, \mathbf{v}_i^h \in \mathbb{R}^c,\ h \in \{1, \ldots, N_{\text{head}}\}$

$\vec{\mathbf{q}}_i^{hp}, \vec{\mathbf{k}}_i^{hp} = \text{LinearNoBias}(\mathbf{s}_i)$ $\quad \vec{\mathbf{q}}_i^{hp}, \vec{\mathbf{k}}_i^{hp}, \in \mathbb{R}^3,\ p \in \{1, \ldots, N_{\text{query points}}\},$ units: nanometres

$\vec{\mathbf{v}}_i^{hp} = \text{LinearNoBias}(\mathbf{s}_i)$ $\quad \vec{\mathbf{v}}_i^{hp} \in \mathbb{R}^3,\ p \in \{1, \ldots, N_{\text{point values}}\},$ units: nanometres

$b_{ij}^h = \text{LinearNoBias}(\mathbf{z}_{ij})$

$w_C = \sqrt{\frac{2}{9N_{\text{query points}}}},$

$w_L = \sqrt{\frac{1}{3}}$

$a_{ij}^h = \text{softmax}_j\left(w_L\left(\frac{1}{\sqrt{c}}\,\mathbf{q}_i^{h\top}\mathbf{k}_j^h + b_{ij}^h - \frac{\gamma^h w_C}{2}\sum_p \left\|T_i \circ \vec{\mathbf{q}}_i^{hp} - T_j \circ \vec{\mathbf{k}}_j^{hp}\right\|^2\right)\right)$

$\tilde{\mathbf{o}}_i^h = \sum_j a_{ij}^h \mathbf{z}_{ij}$

$\mathbf{o}_i^h = \sum_j a_{ij}^h \mathbf{v}_j^h$

$\vec{\mathbf{o}}_i^{hp} = T_i^{-1} \circ \sum_j a_{ij}^h \left(T_j \circ \vec{\mathbf{v}}_j^{hp}\right)$

$\tilde{\mathbf{s}}_i = \text{Linear}\left(\text{concat}_{h,p}(\tilde{\mathbf{o}}_i^h, \mathbf{o}_i^h, \vec{\mathbf{o}}_i^{hp}, \left\|\vec{\mathbf{o}}_i^{hp}\right\|)\right)$

return $\{\tilde{\mathbf{s}}_i\}$

Lattice Gauge Theory

inputs for another transformation. While the attention mechanism is at the heart of some of the most powerful artificial intelligence applications to date [75], we find that in our experiments, convolution architectures generally train faster and result in higher performances.

Abbot et al (2305.02402)

Ongoing work by A. Tomiya et al

ABSTRACT

We propose a unifying approach that starts from the perturbative construction of trivializing maps by Lüscher and then improves on it by learning. The resulting continuous normalizing flow model can be implemented using common tools of lattice field theory and requires several orders of magnitude fewer parameters than any existing machine learning approach. Specifically, our model can achieve competitive performance with as few as 14 parameters while existing deep-learning models have around 1 million parameters for $SU(3)$ Yang-Mills theory on a 16^2 lattice. This has obvious consequences for training speed and interpretability. It also provides a plausible path for scaling machine-learning approaches toward realistic theories.

Bacchio et al, 2023

Can we use ML in other ways?