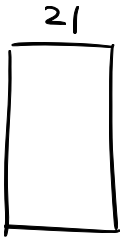
Alphafold 2

N_{res} : # of residues

I ↘ target feature → one-hot (21 d) → N_{res}

(21: 20 AA, 1 unknown)



($N_{res} \times 21$)

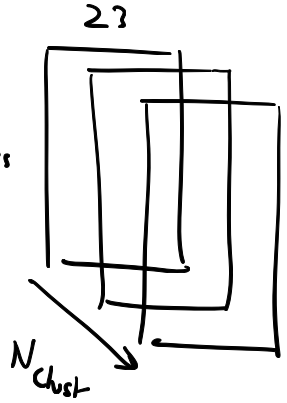
II residue_index → position in the seq → N_{res}

($N_{res} \times 1$)

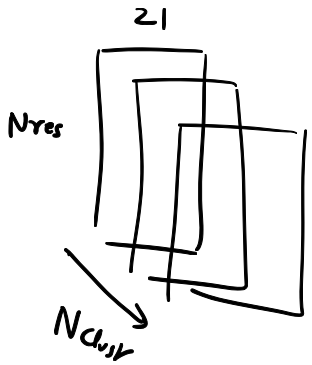
III msa_feature → profile (PSSM) (23 d)

One profile per cluster of sequence

($N_{clust} \times N_{res} \times 49$) actual size in Alphafold2



↳ one-hot of the cluster center sequence



↳ Some more single dim features

IV extra_msa

template-based features (optional!)

V → pair feature → distogram

($N_{templ} \times N_{res} \times N_{res} \times 88$)

N_{templ} ≡ # of templates in PDB to use

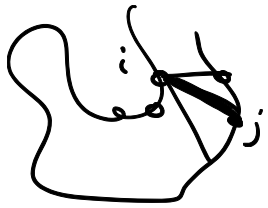
top 20 structure based on seq alone

↓

sample at most 4 structure

a) $N_{templ} \times N_{res} \times N_{res} \times 39$ (one-hot)

pair-wise distance bins



$d_{ij} = dist(r_i, r_j)$

discretize into 38 bins
$[3.25\text{Å}, 50.75\text{Å}]$
+
1 bin for $> 50.75\text{Å}$

b) displacement of C_α in the local ref frame

:

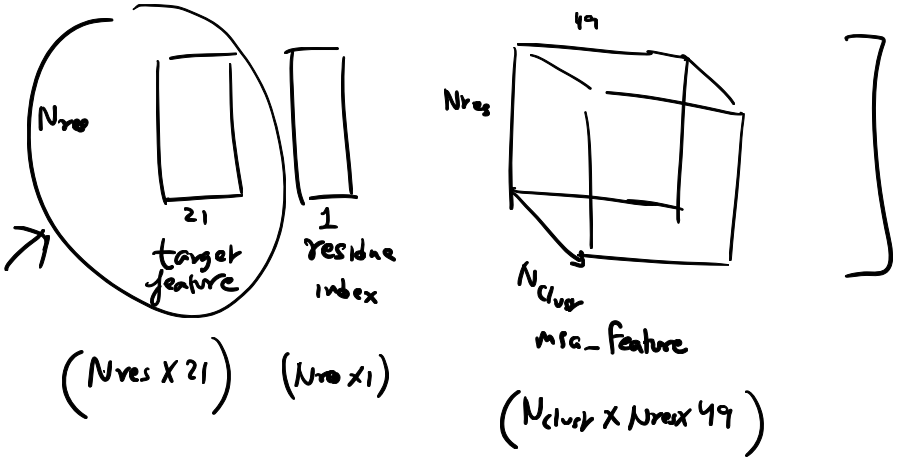
VI template-based angle

↳ ϕ_i, ψ_i, ω_i per residue

↳ χ_i — upto 4 angle

] 7 feature

$N_{templ} \times N_{res} \times 51$



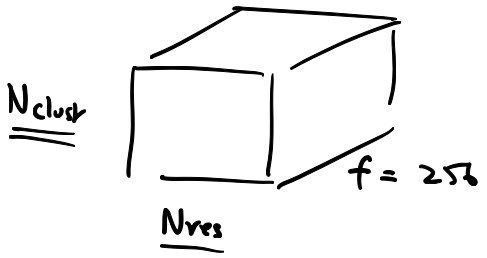
$(N_{res} \times 21)$ $(N_{res} \times 1)$ $(N_{clust} \times N_{res} \times 49)$

] raw inputs

Input Embeddings

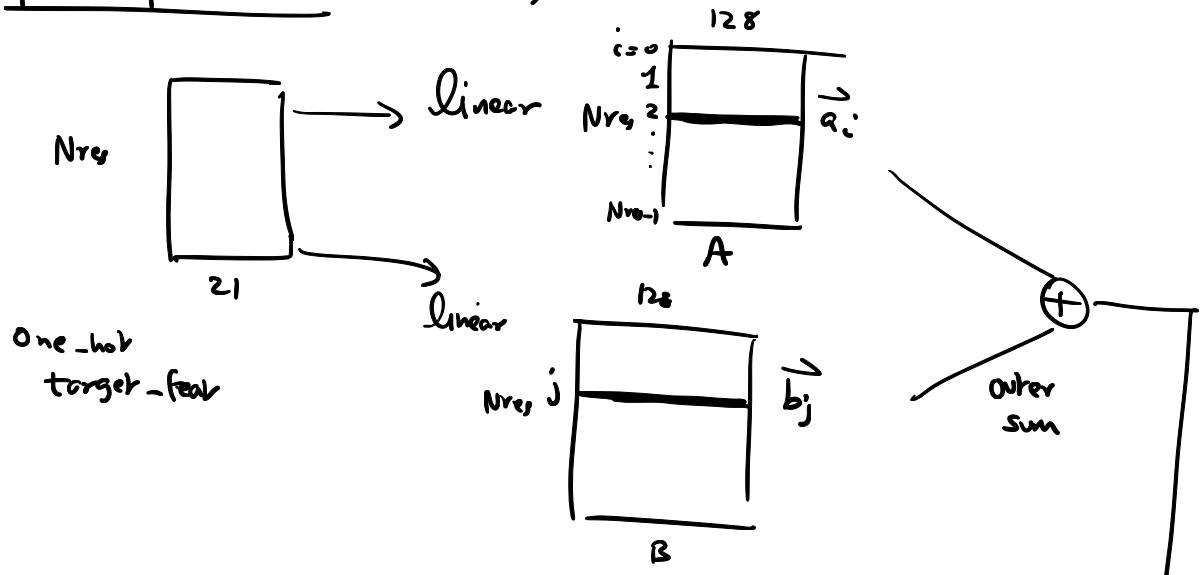
msa_feature	$N_{clust} \times N_{res} \times 49$	linear ⟶	$N_{clust} \times N_{res} \times 256$

$c_m = 256$



Pair

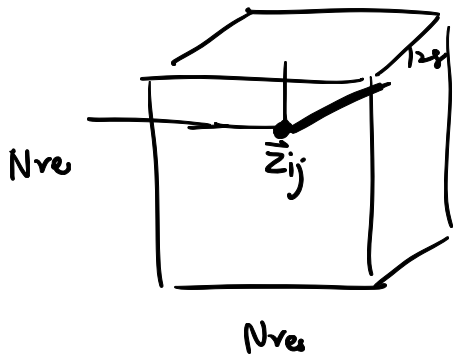
pair_representation ($C_z = 128$)



for $i = 0 \ldots N_{res} - 1$

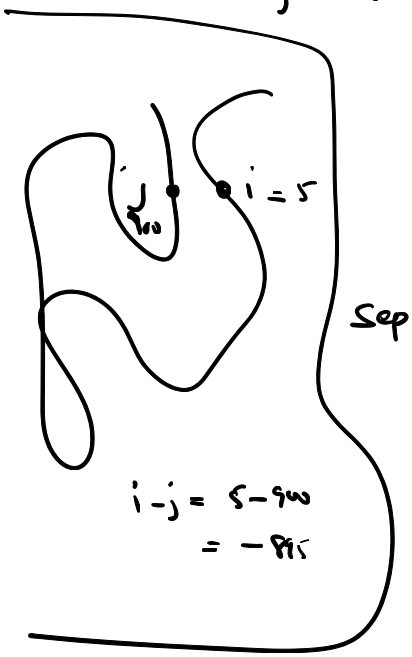
for $j = 0 \ldots N_{res} - 1$

$\vec{z}_{ij} = \vec{a}_i + \vec{b}_j$



1) $\vec{z}_{ij} = \vec{a}_i + \vec{b}_j$

128 128 128



$sep_{ij} = \underbrace{idx_i - idx_j}_{\text{clip to } [-32, -31, \ldots, 31, 32]}$

$i \rightarrow idx_i$, $j \rightarrow idx_j$

⇓

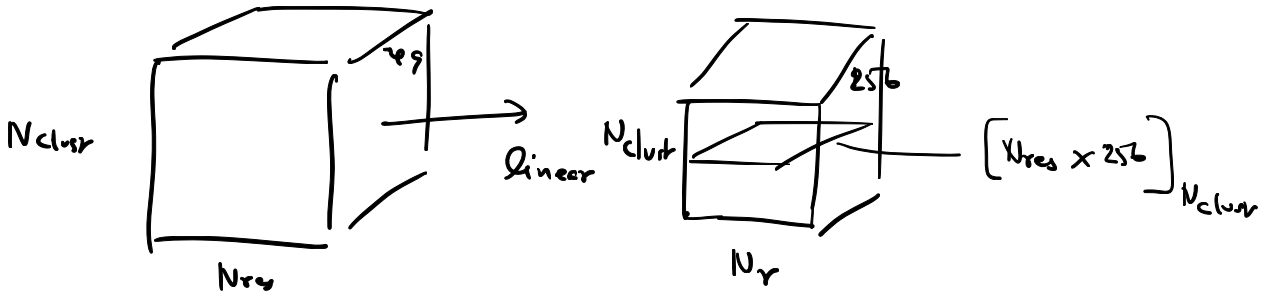
one-hot bin for sep_{ij}

2) relative position$_{ij}$ (128)

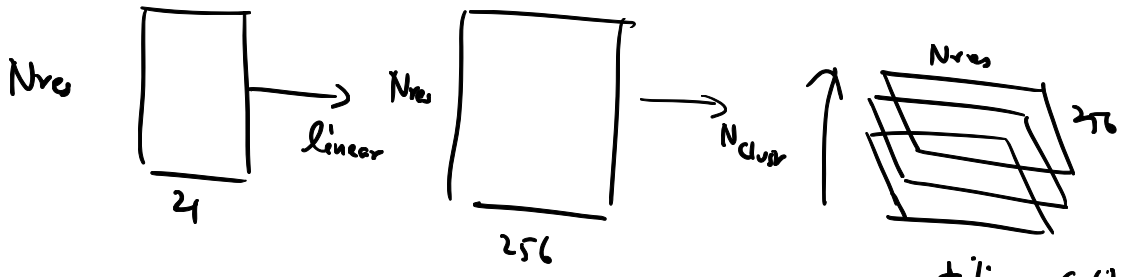
$\vec{z}_{ij} = \vec{z}_{ij} + \overrightarrow{relpos}_{ij}$

$relpos_{ij} = \underset{128}{linear}(one\text{-}hot(sep_{ij}))$

msa_feature revisited ($c_m = 256$)



+

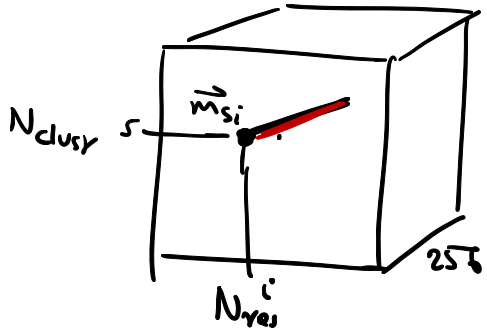


tiling (tile N_{clust} times)

---

set of feature for Evoformer

msa_feature



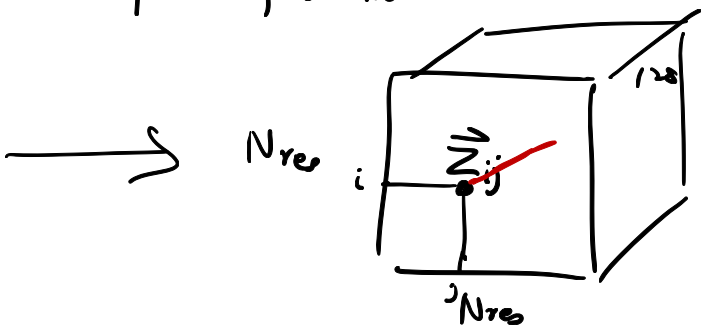
$c_m = 256$

s: seq index in N_{clust}

i: pos index in N_{res}

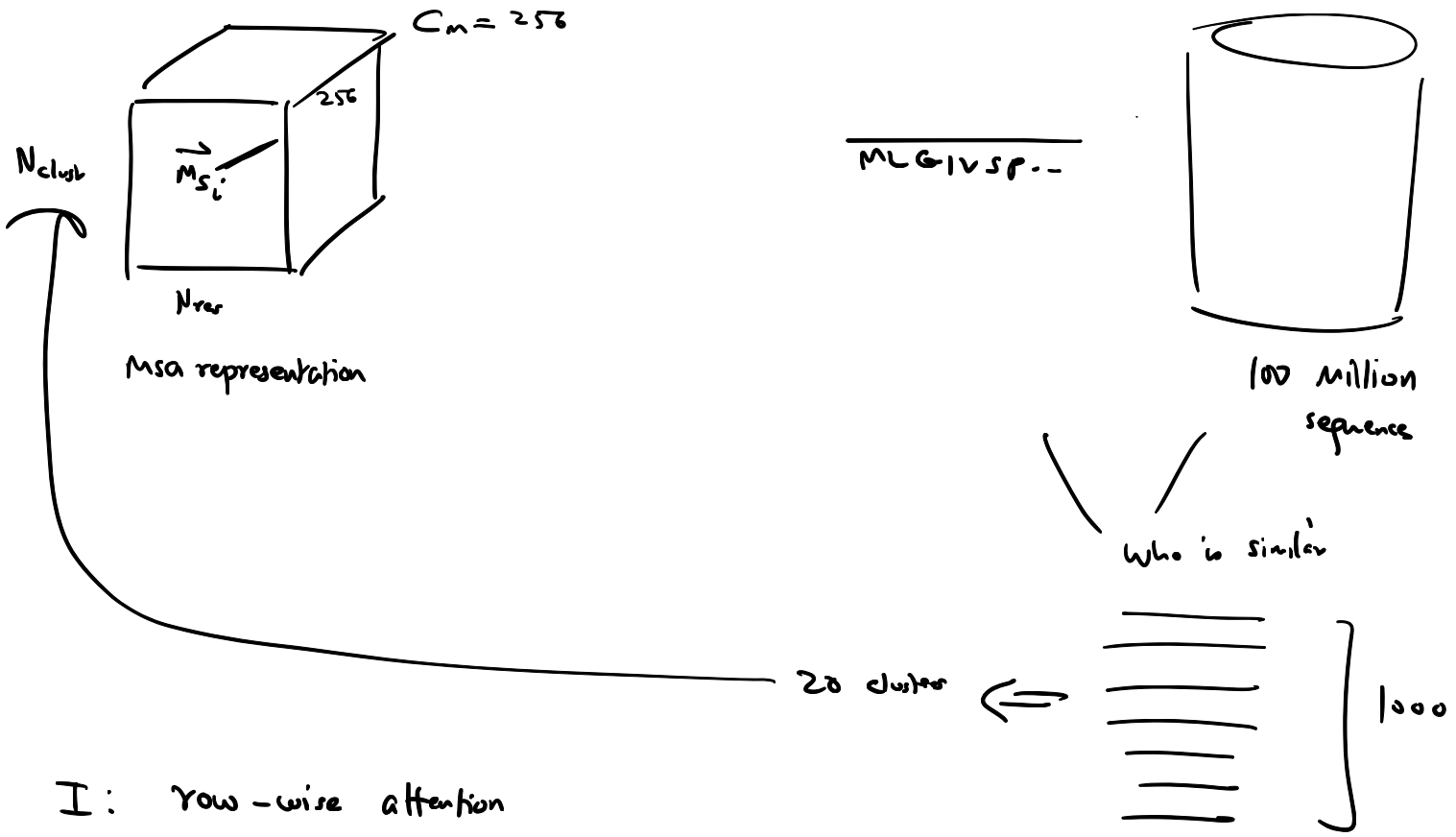
$\vec{m}_{si} \in \mathbb{R}^{256} \equiv \mathbb{R}^{c_m}$

pair_representation

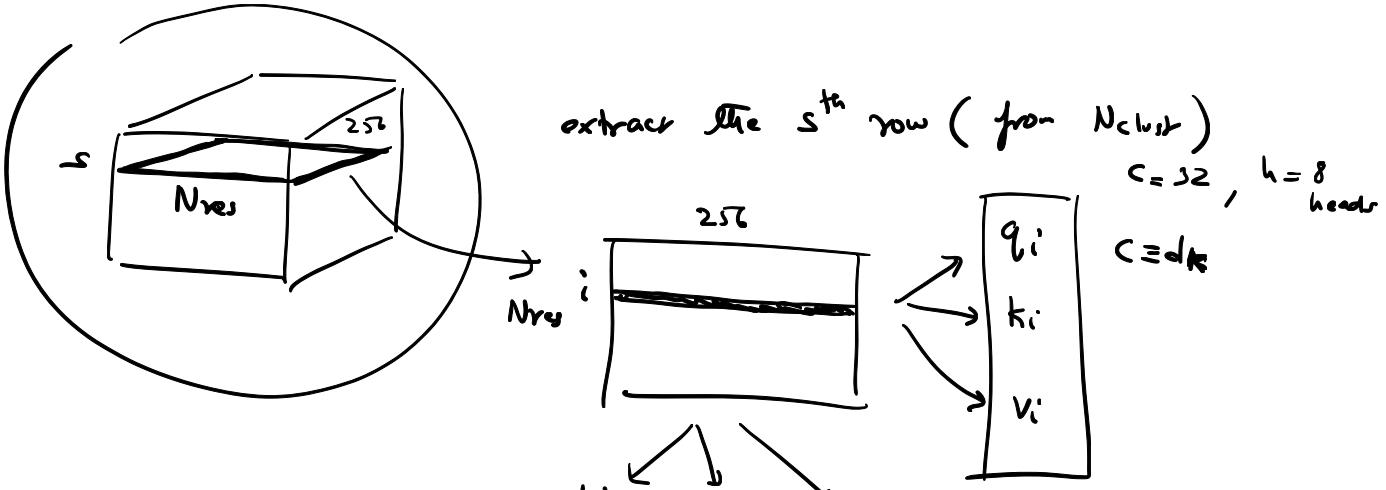


$c_z = 128$

$\vec{z}_{ij} \in \mathbb{R}^{128} \equiv \mathbb{R}^{c_z}$



I: row-wise attention



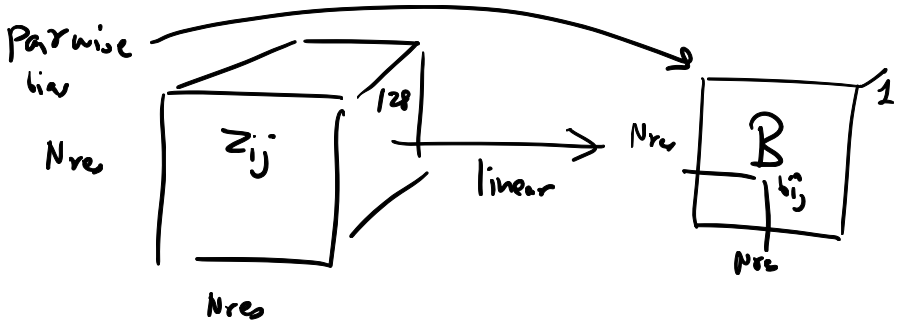
i [gated attention
ii add a pair bias]

a_{ij} = attn between q_i & k_j

Pair bias

$$\text{Softmax}\left(\frac{1}{\sqrt{c}} q_i^T k_j + b_{ij}\right)$$

$$A = \text{Softmax}\left(\frac{1}{\sqrt{c}} QK^T + B\right)$$



$$O_i = g_i \odot \left(\sum_j a_{ij} \vec{V}_j \right)$$

$$G_i \odot A_i V_i$$

sigmoid gate that controls the extent of the value update

$[0,1]$

$$G_{\forall s \forall i} = \boxed{g_{si} = \text{Sigmoid}\left(m_{si}\right)}$$