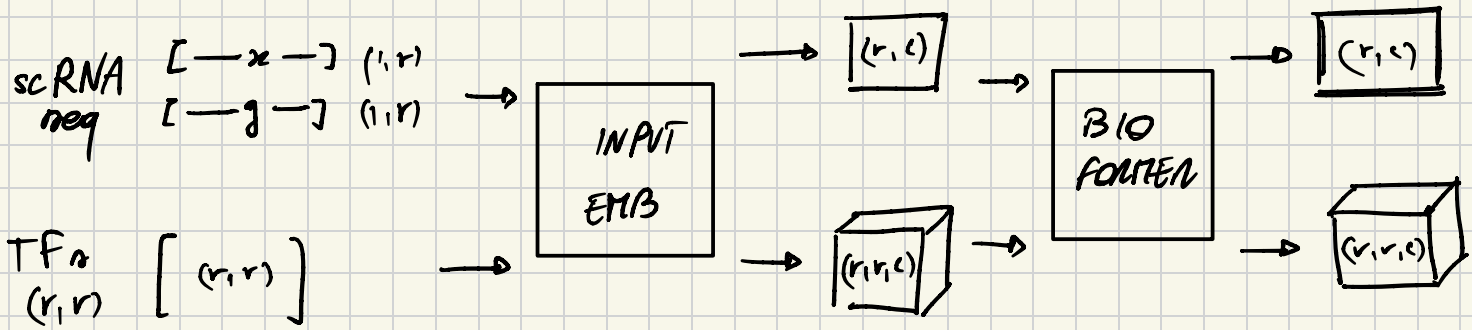
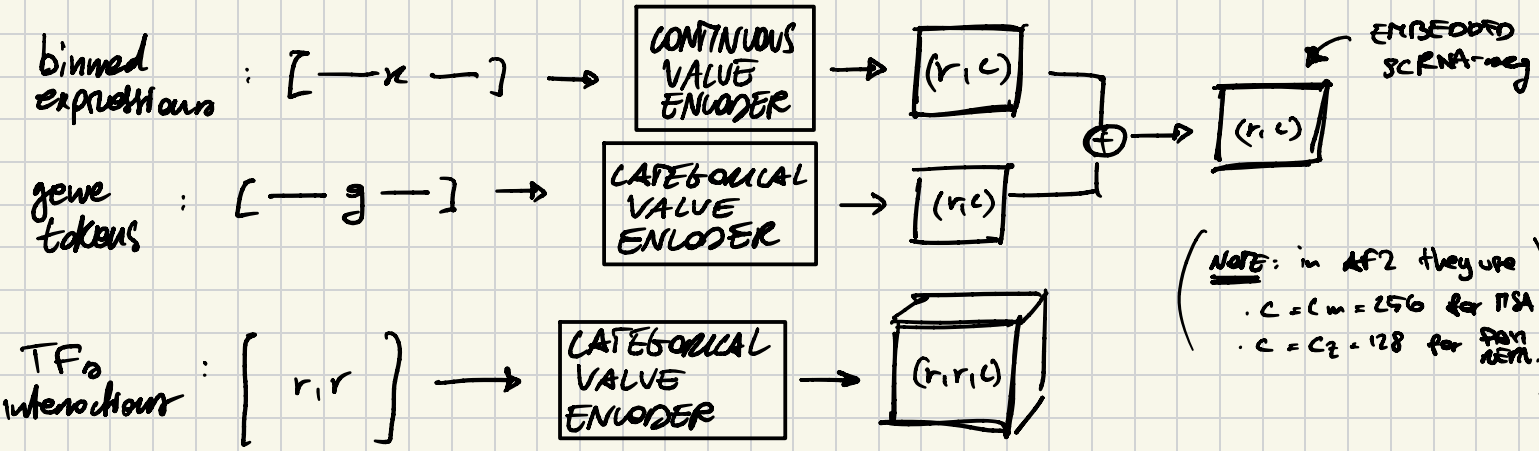
BioLLMs

MY MODEL

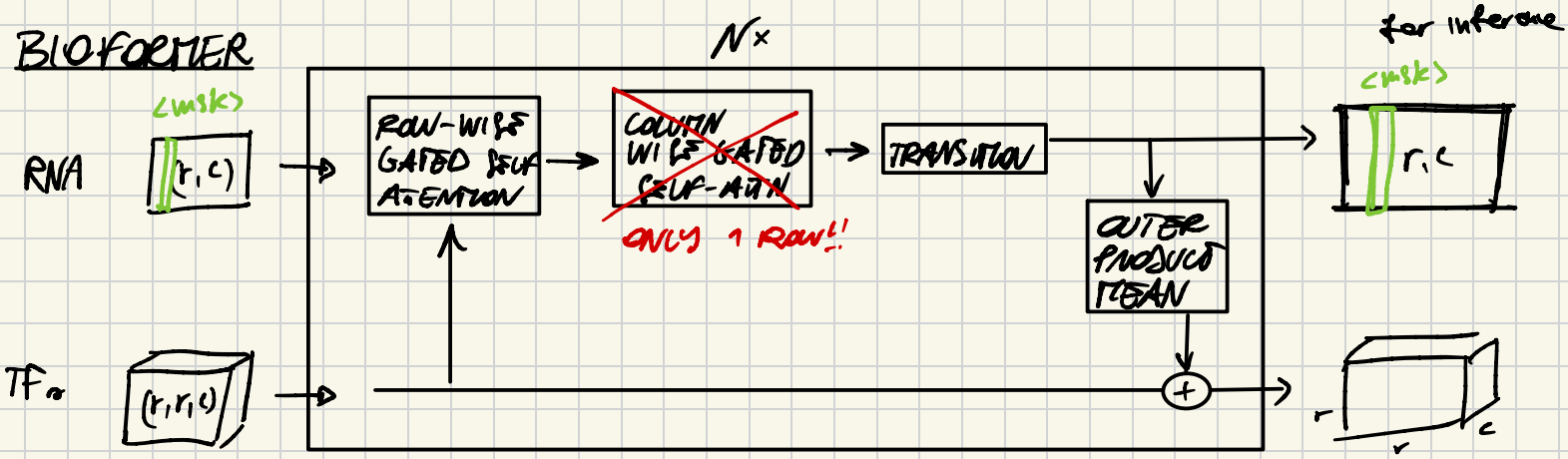


INPUT EMBEDDING



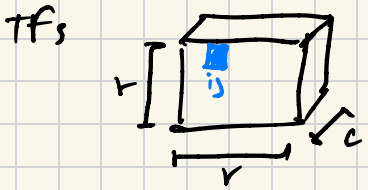
(NOTE: in AF2 they use
- $c = c_m = 256$ for MSA
- $c = c_z = 128$ for pair rem.)

BIOFORMER



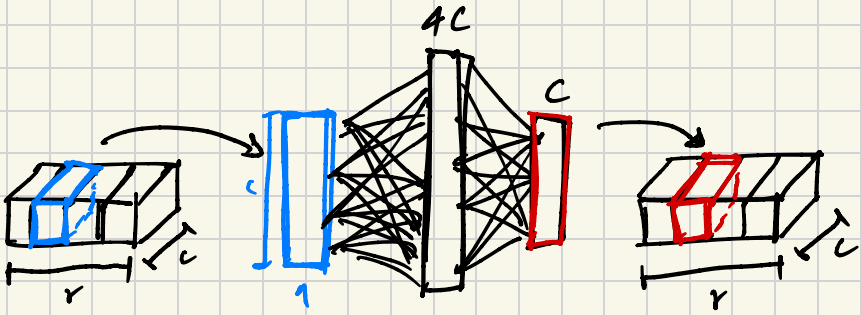
ROW-WISE GATED SELF-ATTN

Attention i,j with pair bias TF_a i,j



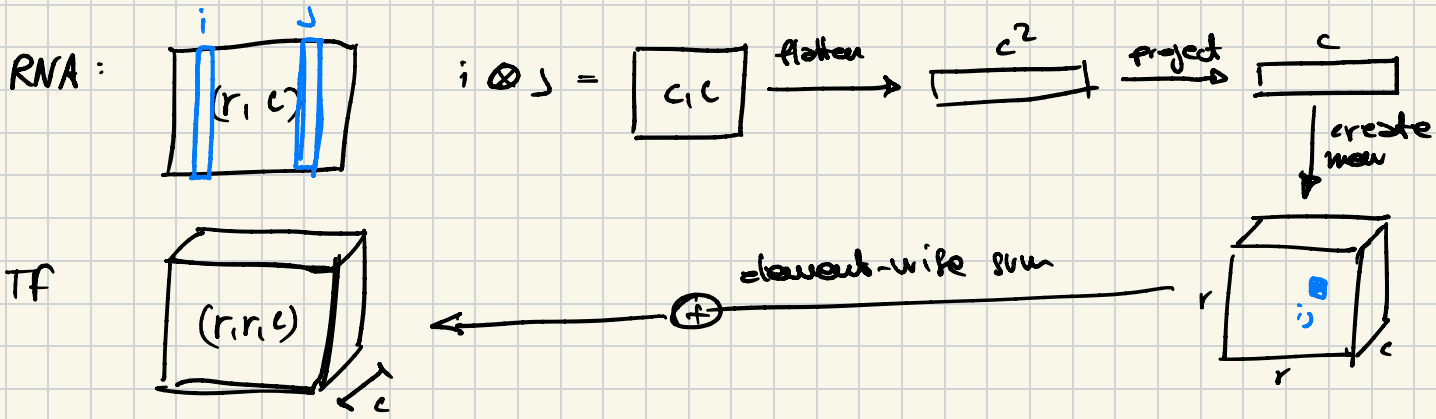
TRANSITION

FFN on each input symbol



OUTER PRODUCT MEAN

→ No need to take the mean, simply take outer product, flatten and project.



CODING

class Model :

def __init__ (args)

INPUT EMB.
- CONTINUOUS ENCODER (x) emb_x
- GENE ENCODER (g) emb_g
- CAT ENCODER (TF_id) emb_TF

- ROW-WISE GATED SELF ATT WITH PAIR BIAS sapb
- TRANSITION ffnn
- OUTER PRODUCT MEAN opm

args:
- rna_emb_dim = c_m
- TF_emb_dim = c_z
- nheads
- N

def forward (x, g)

```
TF = retrieve_tf (g)
h = emb_x(x) + emb_g (g)
TF_emb = emb_tf (tf)
```
embed input

```
h = sapb (h, TF_emb)
h = ffnn (h)
```
update h

```
TF_emb = opm (h, TF_emb)
```
update TF