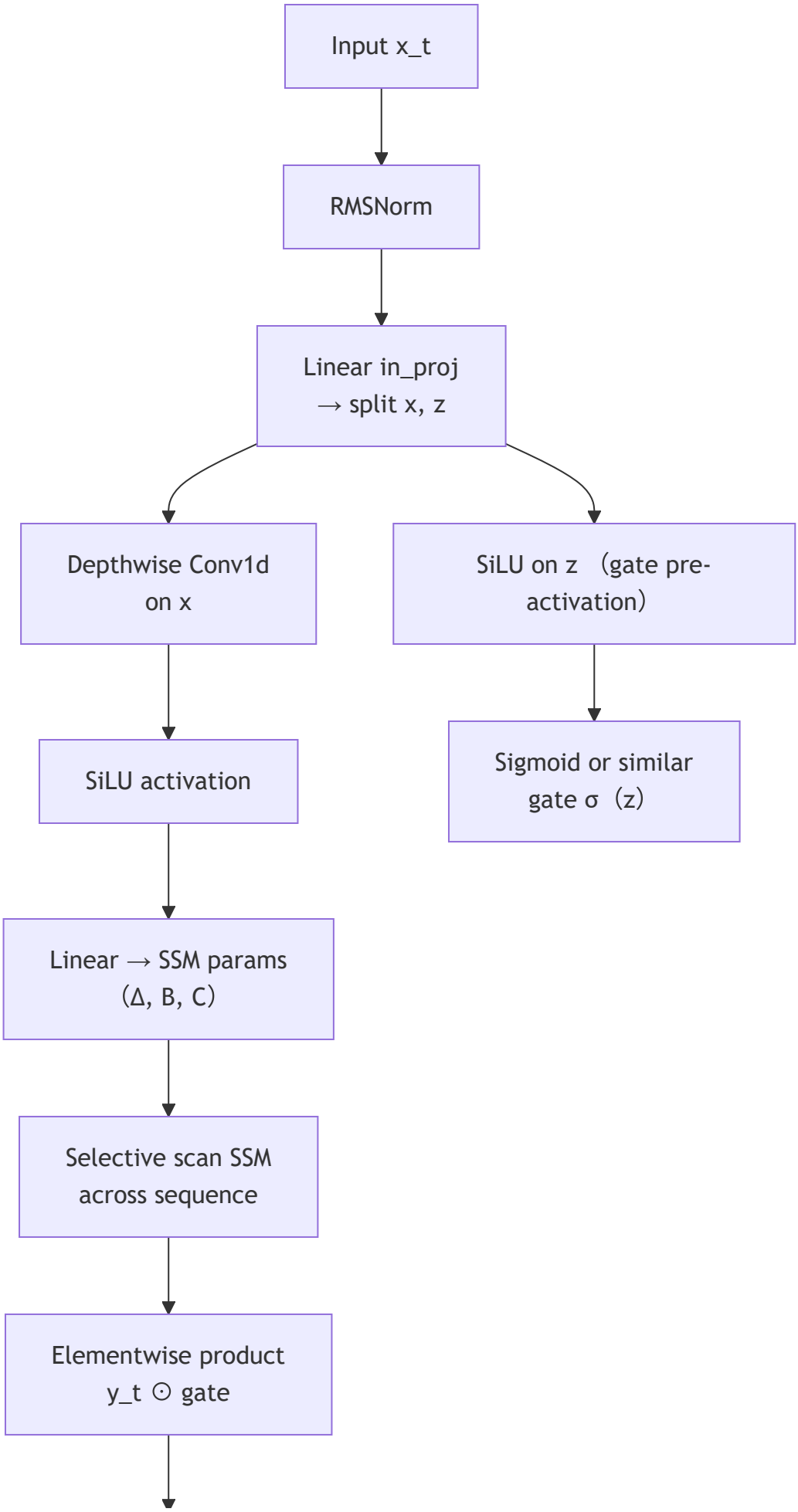
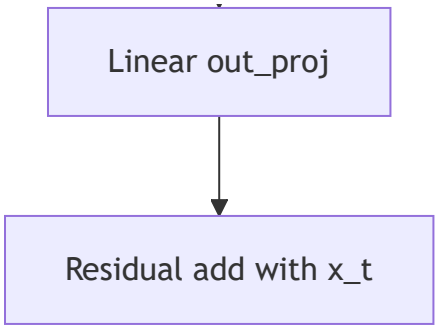
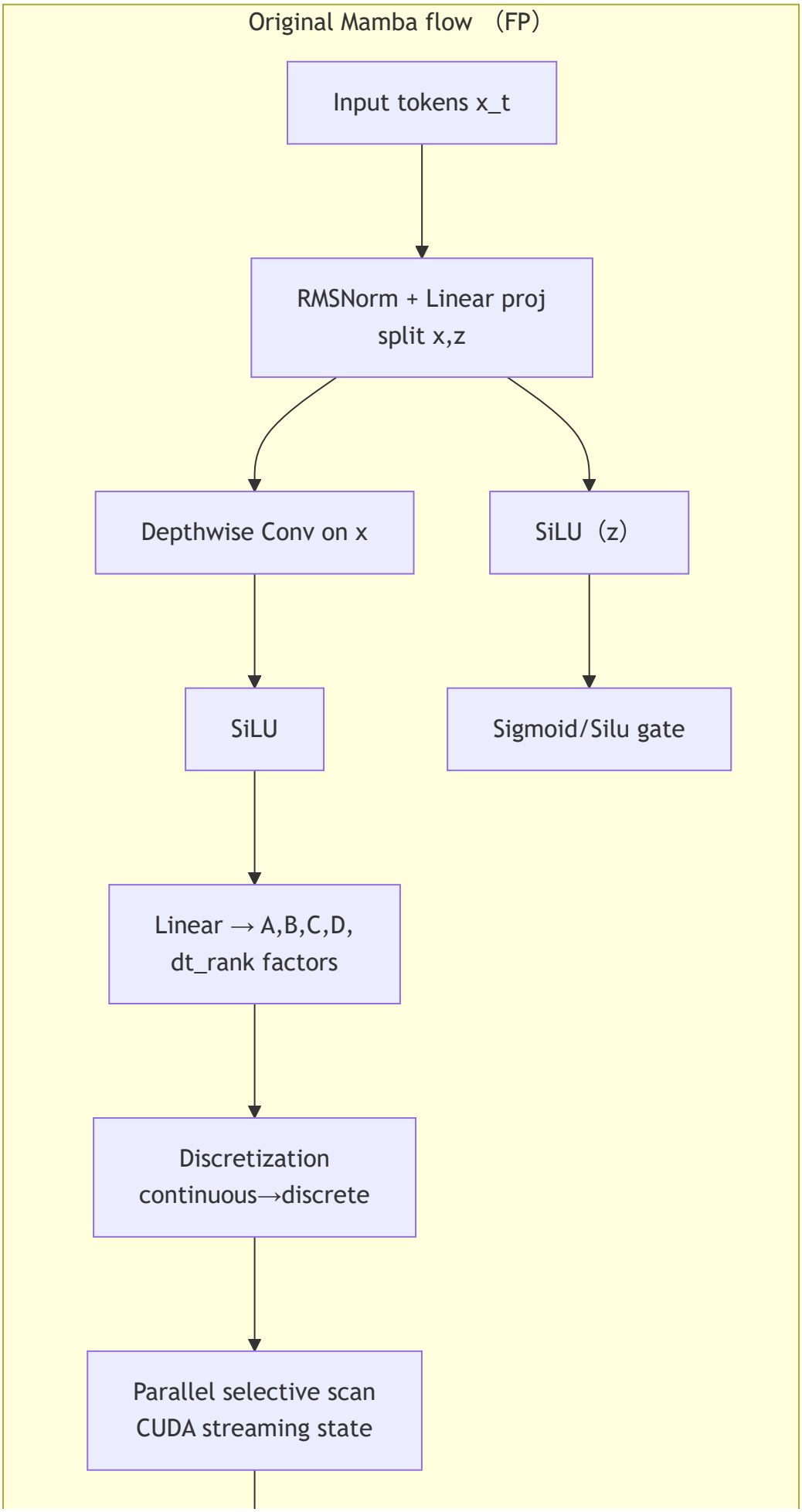
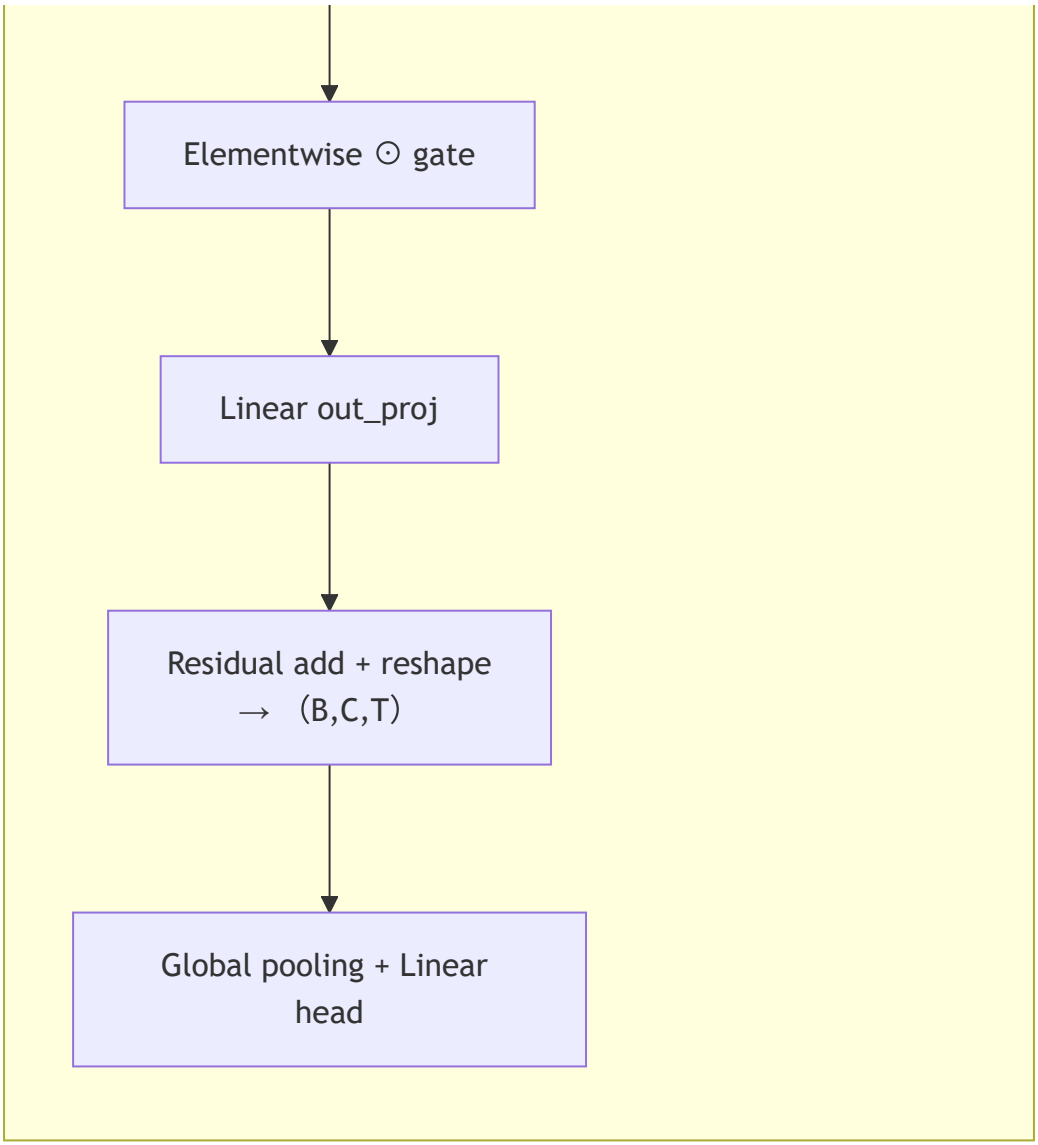
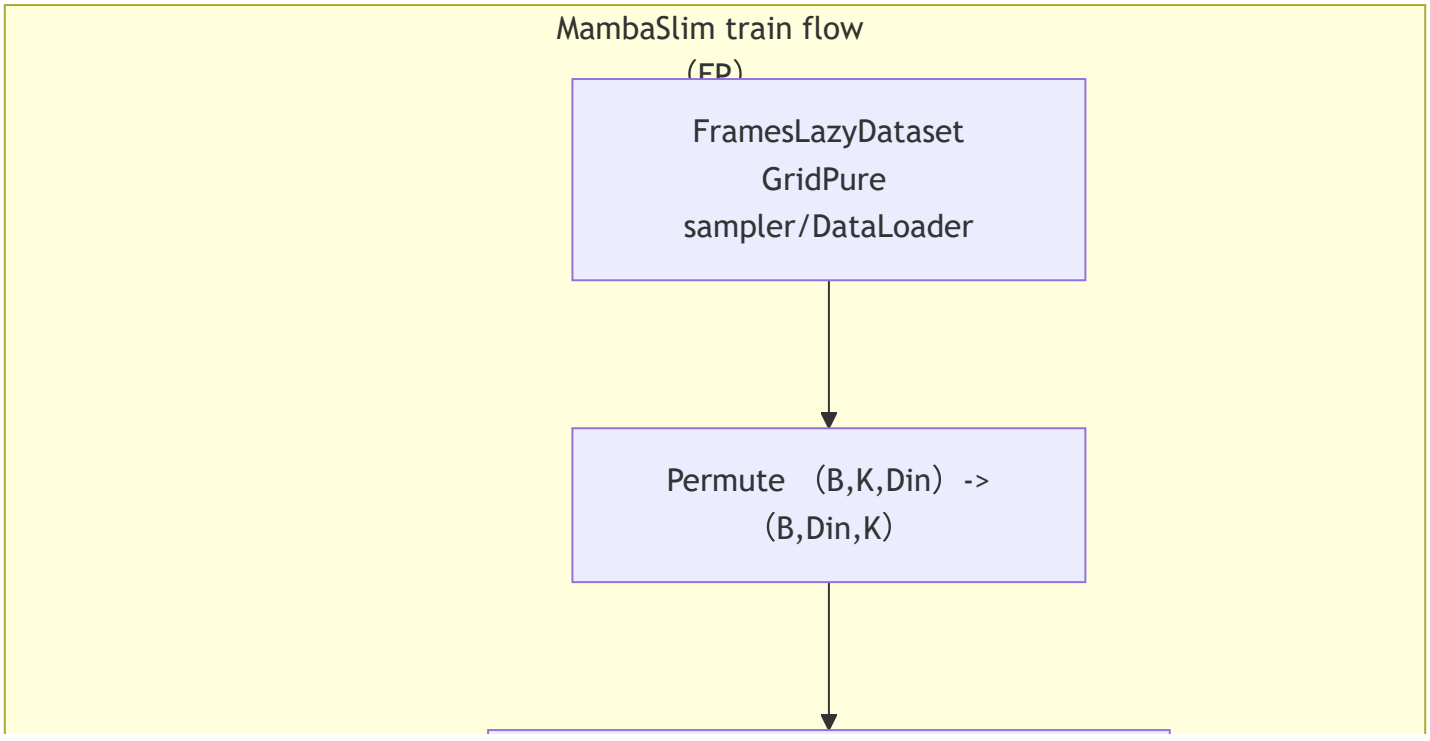
Mamba （Original）

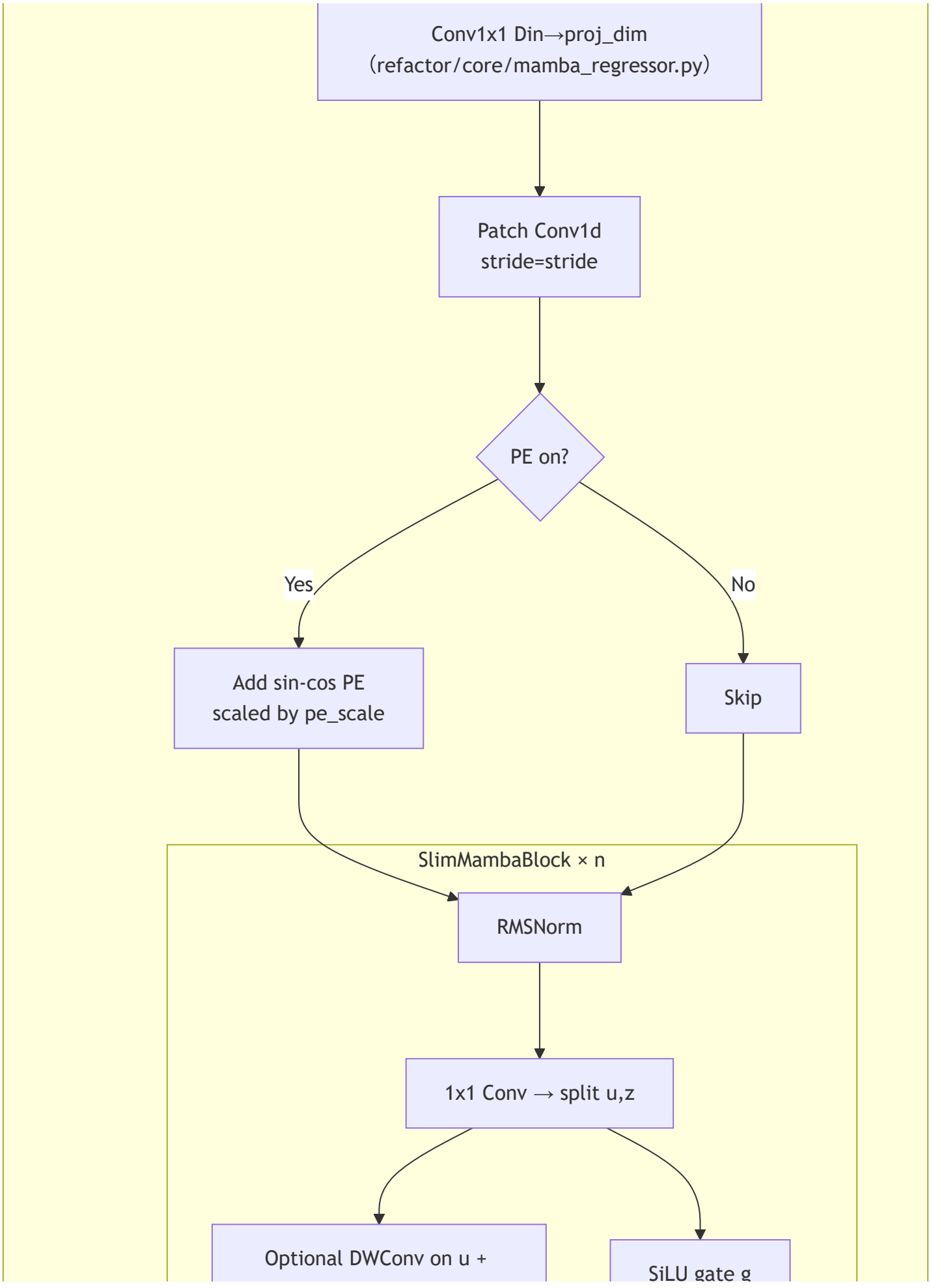

Input x_t
RMSNorm
Linear in_proj
→ split x, z
Depthwise Conv1d
on x
SiLU on z （gate pre-
activation）
SiLU activation
Sigmoid or similar
gate σ（z）
Linear → SSM params
（Δ, B, C）
Selective scan SSM
across sequence
Elementwise product
y_t ⊙ gate

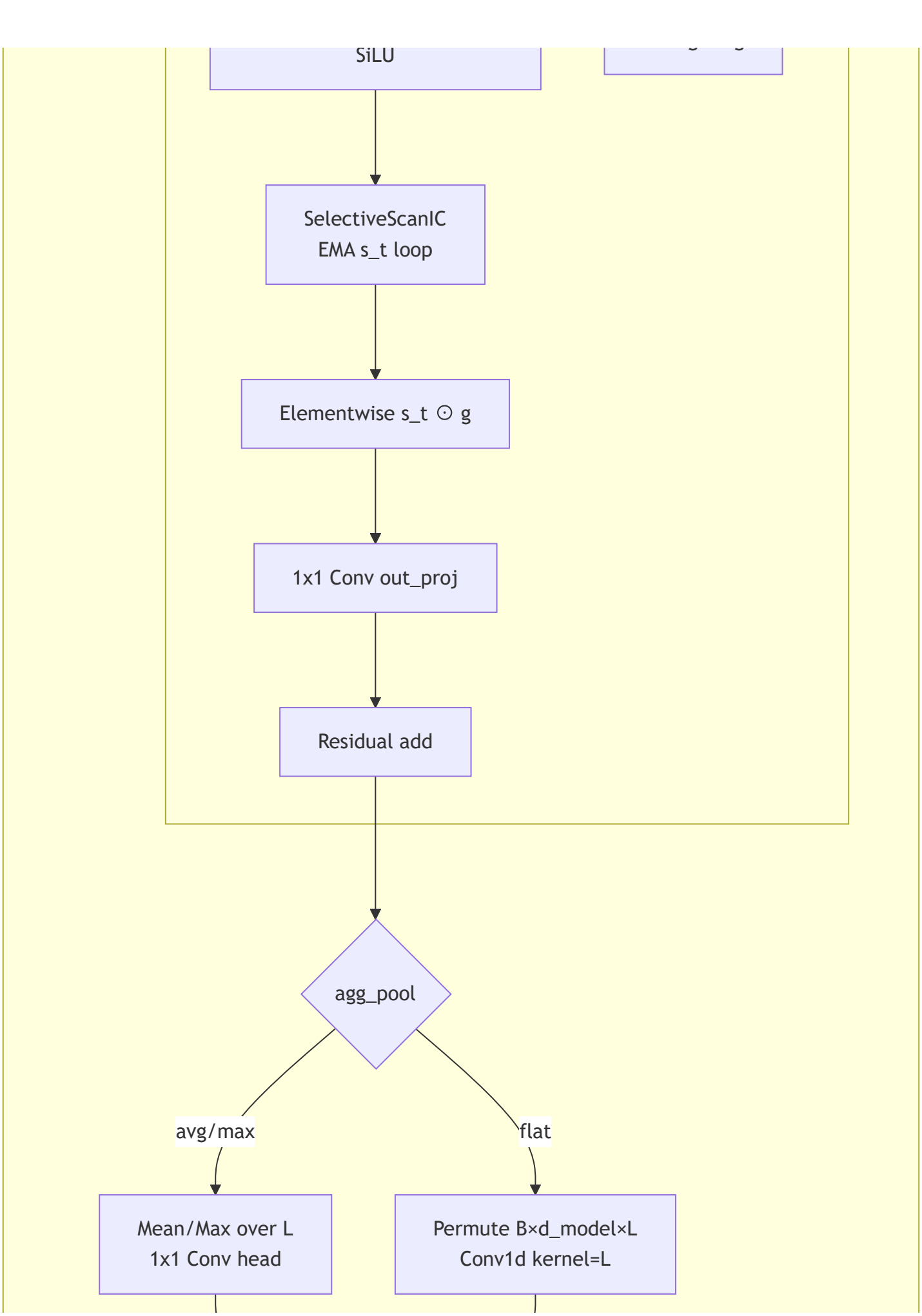

Linear out_proj
Residual add with x_t

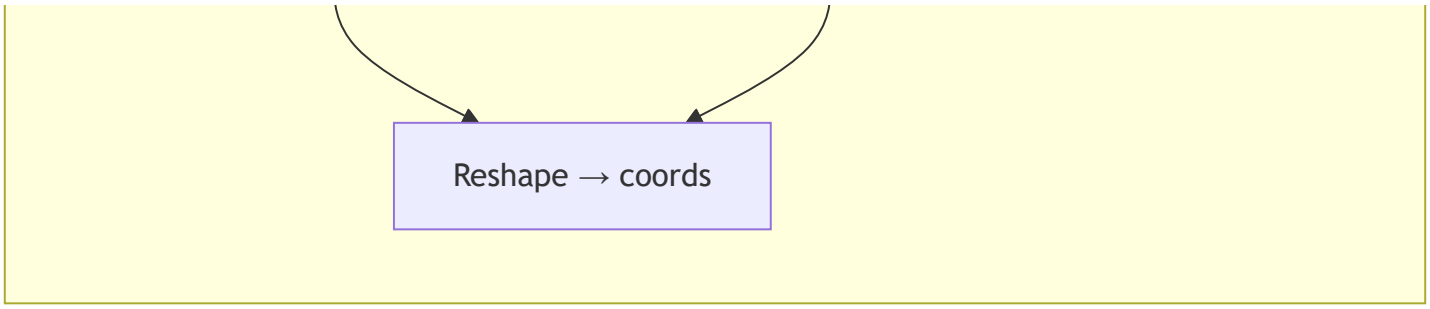

Original Mamba flow （FP）
Input tokens x_t
RMSNorm + Linear proj
split x,z
Depthwise Conv on x
SiLU（z）
SiLU
Sigmoid/Silu gate
Linear → A,B,C,D,
dt_rank factors
Discretization
continuous→discrete
Parallel selective scan
CUDA streaming state


Elementwise ⊙ gate
Linear out_proj
Residual add + reshape
→ (B,C,T)
Global pooling + Linear
head


MambaSlim train flow
(FP)
FramesLazyDataset
GridPure
sampler/DataLoader
Permute (B,K,Din) ->
(B,Din,K)


Conv1x1 Din→proj_dim
(refactor/core/mamba_regressor.py)
Patch Conv1d
stride=stride
PE on?
Yes
No
Add sin-cos PE
scaled by pe_scale
Skip
SlimMambaBlock × n
RMSNorm
1x1 Conv → split u,z
Optional DWConv on u +
SiLU gate g


SiLU
SelectiveScanIC
EMA s_t loop
Elementwise s_t ⊙ g
1x1 Conv out_proj
Residual add
agg_pool
avg/max
flat
Mean/Max over L
1x1 Conv head
Permute B×d_model×L
Conv1d kernel=L


Reshape → coords