Forward pass

$\mathbf{X}$ → [LN] → [⊙] → [+] → [GL] → [DO] → [⊙] → [+] → [DO] → $\mathbf{Y}$

- $\mathbf{x}$ $[B, S, D]$
- $\mathbf{H}$ $[B, S, D]$
- $\mathbf{Z}_{\text{up}}$ $[B, S, D_{ff}]$
- $\mathbf{A}_{\text{up}}$ $[B, S, D_{ff}]$
- $\mathbf{H}_{\text{inter}}$ $[B, S, D_{ff}]$
- $\mathbf{U}_{\text{in}}$ $[B, S, D_{ff}]$
- $\mathbf{z}_{\text{down}}$ $[B, S, D]$
- $\mathbf{A}_{\text{out}}$ $[B, S, D]$
- $\mathbf{Y}$ $[B, S, D]$

Inputs:
- $\tilde{\mathbf{W}}_{\text{up}}$: $\tilde{\mathbf{w}}_{\text{up}}$ $[D, D_{ff}]$
- $\tilde{\mathbf{b}}_{\text{up}}$: $\text{BC}_{B,S}(\tilde{\mathbf{b}}_{\text{up}})$ $[D_{ff}]$
- $\tilde{\mathbf{W}}_{\text{down}}$: $\tilde{\mathbf{w}}_{\text{down}}$ $[D_{ff}, D]$
- $\tilde{\mathbf{b}}_{\text{down}}$: $\text{BC}_{B,S}(\tilde{\mathbf{b}}_{\text{down}})$ $[D]$

Note: $D_{ff} \gg D$ (e.g., 4×)

Backward pass

- $\mathbf{H}$ $[B, S, D]$ → [T] → $\mathbf{H}^T$ $[B, D, S]$ → [⊙] → $\mathrm{d}\tilde{\mathbf{W}}_{\text{up}}$ $[D, D_{ff}]$
- $\mathrm{d}\tilde{\mathbf{b}}_{\text{up}}$ $[D_{ff}]$ ← $\Sigma_{B,S}$
- $\mathbf{U}_{\text{in}}$ $[B, S, D_{ff}]$ → [T] → $\mathbf{U}_{\text{in}}^T$ $[B, D_{ff}, S]$ → [⊙] → $\mathrm{d}\tilde{\mathbf{W}}_{\text{down}}$ $[D_{ff}, D]$
- $\mathrm{d}\tilde{\mathbf{b}}_{\text{down}}$ $[D]$ ← $\Sigma_{B,S}$

$\mathrm{d}\mathbf{X}$ ← [dLN] ← [⊙] ← [dGL] ← [dDO] ← [⊙] ← [dDO] ← $\mathrm{d}\mathbf{Y}$

- $\mathrm{d}\mathbf{X}$ $[B, S, D]$
- $\mathrm{d}\mathbf{H}$ $[B, S, D]$
- $\mathrm{d}\mathbf{Z}_{\mathbf{up}} = \mathrm{d}\mathbf{A}_{\mathbf{up}}$ $[B, S, D_{ff}]$
- $\mathrm{d}\mathbf{H}_{\text{inter}}$ $[B, S, D_{ff}]$
- $\mathrm{d}\mathbf{U}_{\text{in}}$ $[B, S, D_{ff}]$
- $\mathrm{d}\mathbf{Z}_{\mathbf{down}} = \mathrm{d}\mathbf{A}_{\mathbf{out}}$ $[B, S, D]$
- $\mathrm{d}\mathbf{Y}$ $[B, S, D]$

- $\tilde{\mathbf{W}}_{\text{up}}^T$ $[D_{ff}, D]$
- $\tilde{\mathbf{W}}_{\text{down}}^T$ $[D, D_{ff}]$