MLP block

(nodes name can change depending on Module substitution)

mN.w.out_bias
[d] Array

bN.m.p_bias
[L, d] Add

mN.w.proj_out
[d, d_m] Array

mN
[L, d] Einsum ph,oh → po

m.act
[L, d_m] GeneralFunction gelu

m4.w.in_bias
[d_m] Array

m.pre
[L, d_m] Add

m.pre_mul
[L, d_m] Einsum ji,hi → jh

m4.w.proj_in
[d_m, d] Array

m.input
[L, d]

LayerNorm

bN.m.input
[L, d]