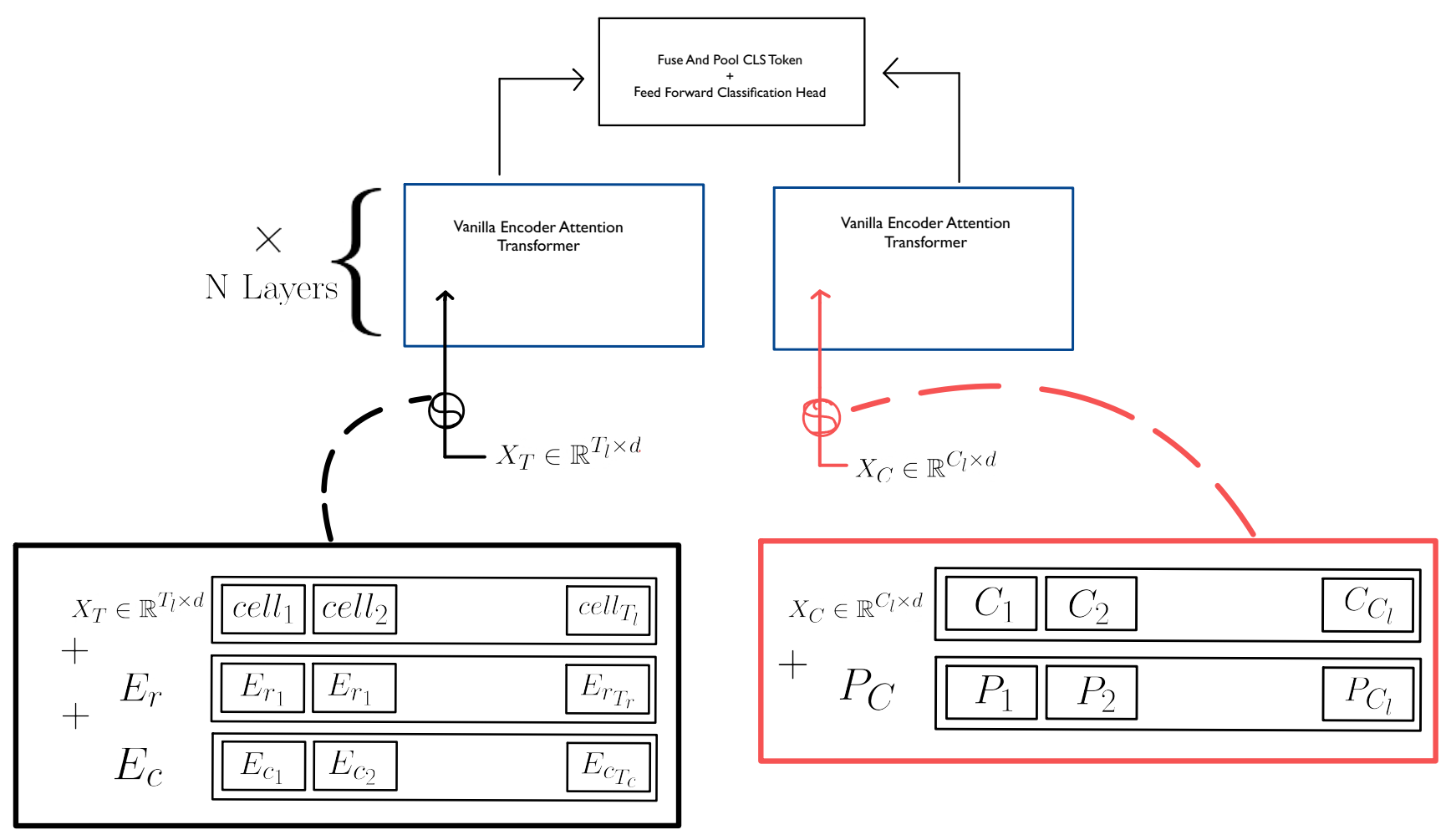
Fuse And Pool CLS Token
+
Feed Forward Classification Head
Vanilla Encoder Attention Transformer
Vanilla Encoder Attention Transformer
N Layers
$X_T \in \mathbb{R}^{T_l \times d}$
$X_C \in \mathbb{R}^{C_l \times d}$
$X_T \in \mathbb{R}^{T_l \times d}$
$cell_1$ $cell_2$ $cell_{T_l}$
E_r
E_{r_1} E_{r_1} $E_{r_{T_r}}$
E_c
E_{c_1} E_{c_2} $E_{c_{T_c}}$
$X_C \in \mathbb{R}^{C_l \times d}$
C_1 C_2 C_{C_l}
P_C
P_1 P_2 P_{C_l}