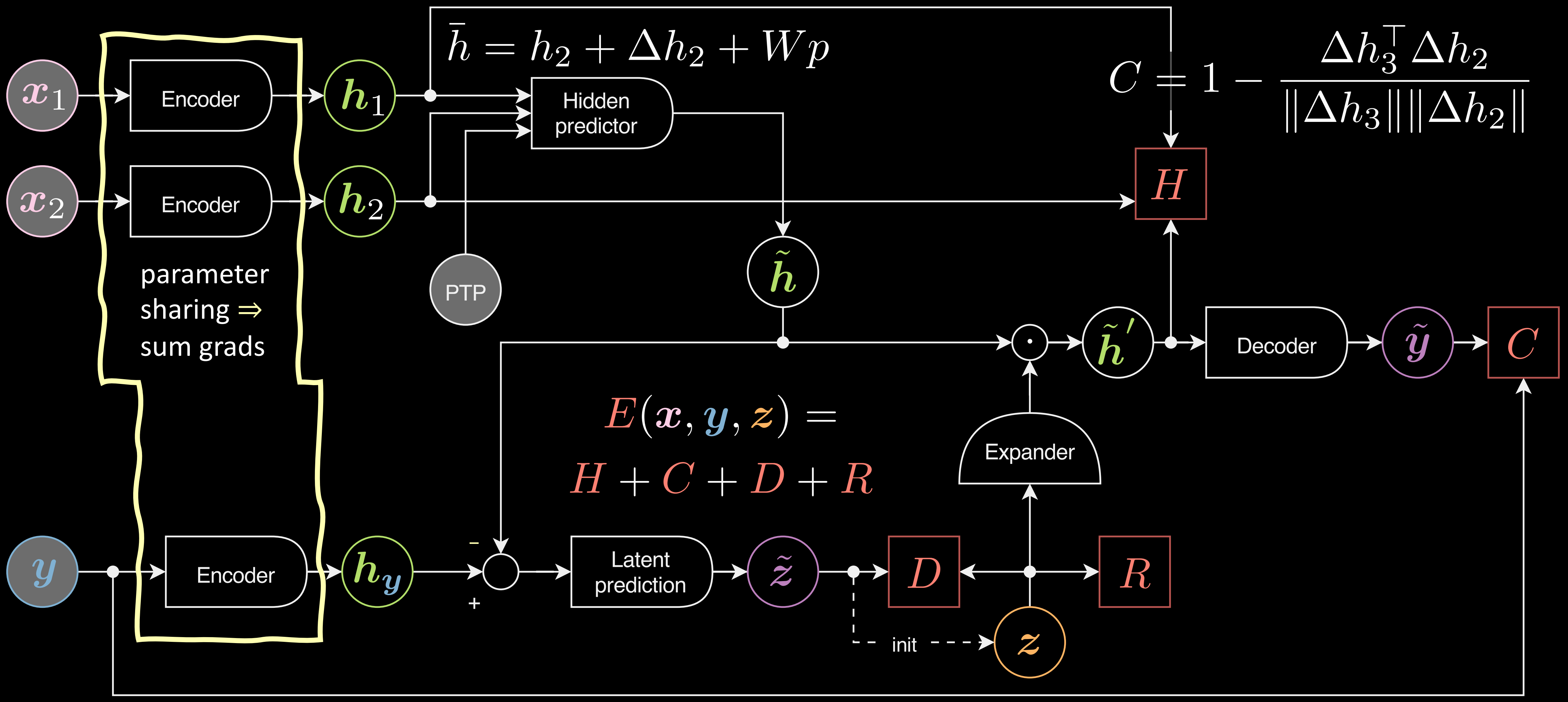
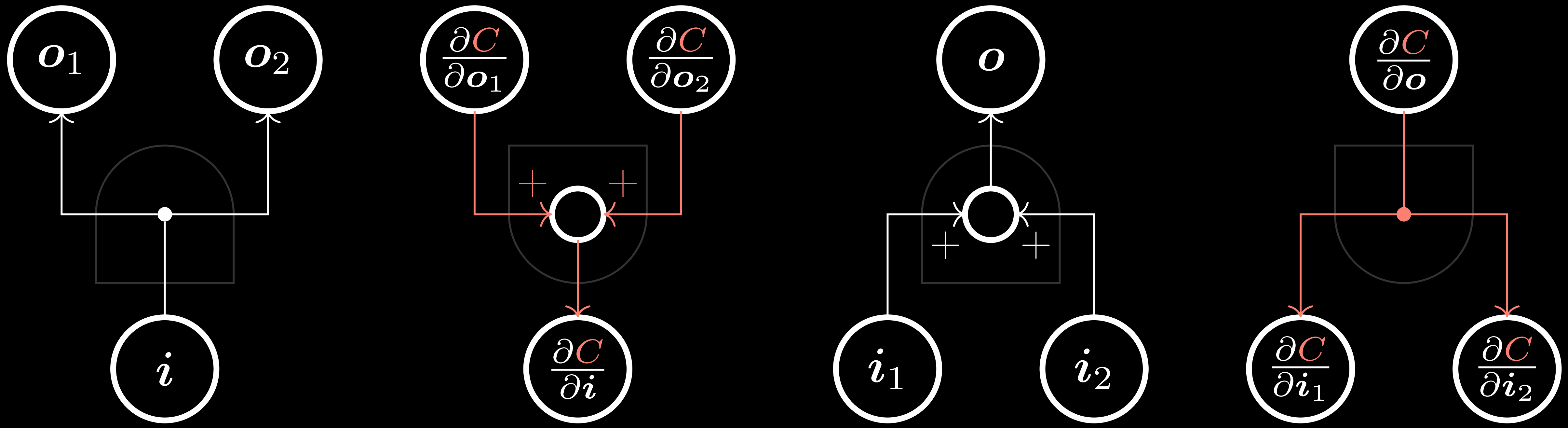
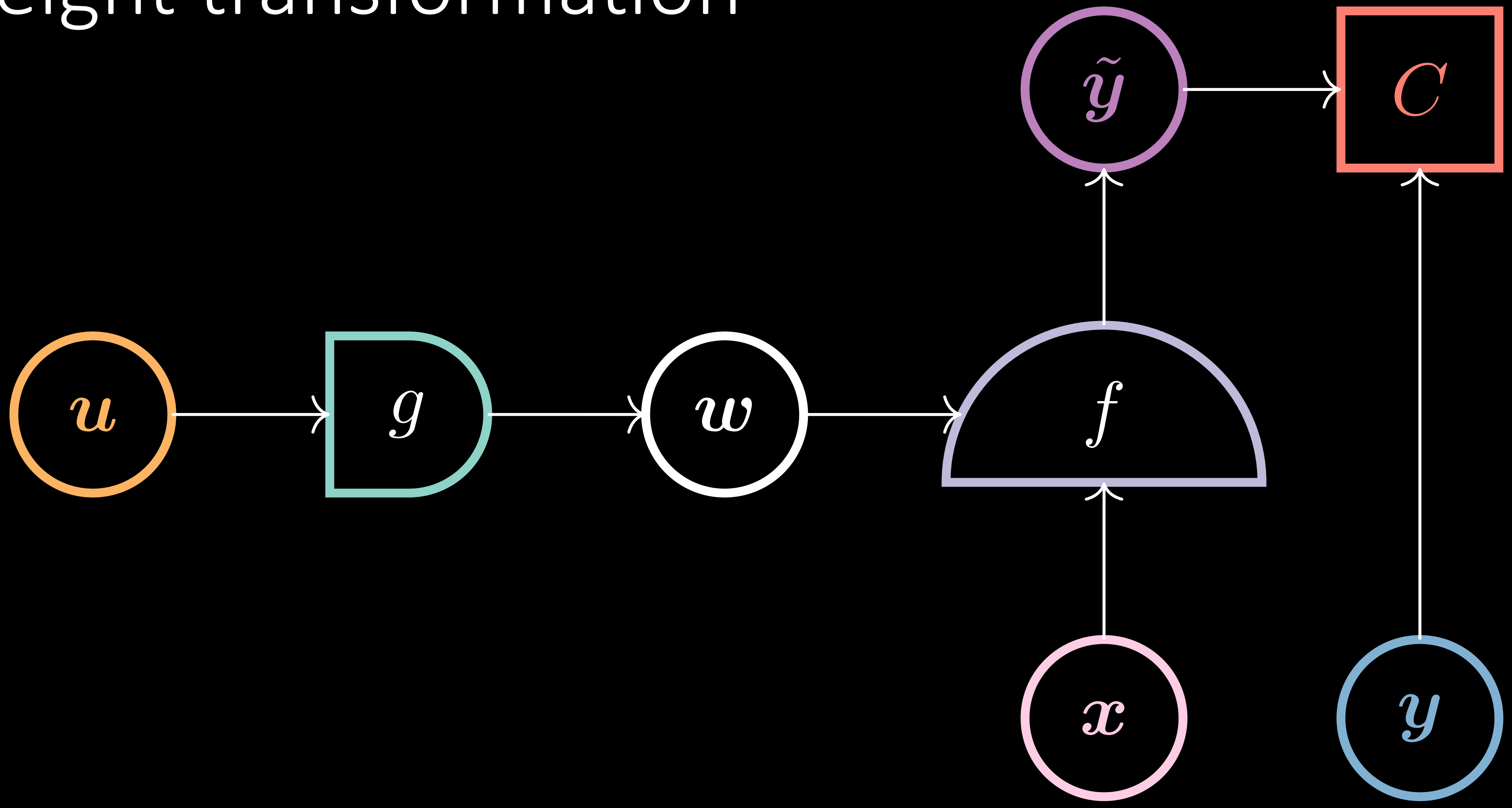
Why backward accumulates ∇params (II)



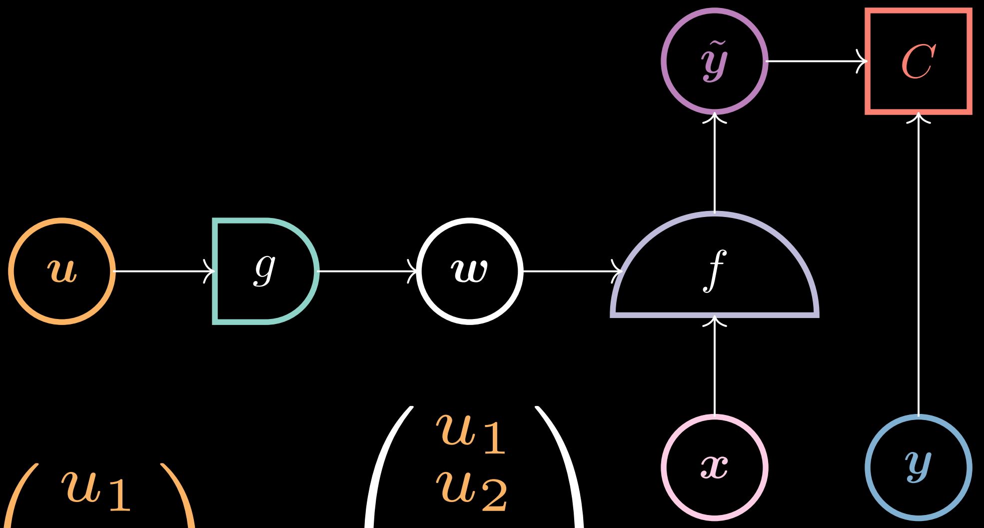
Node and sum modules



Weight transformation



Weight sharing (I)



$$\boldsymbol{w} = \boldsymbol{G}\boldsymbol{u} \quad \Rightarrow \quad \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_3 \\ u_3 \end{pmatrix}$$

$$\frac{\partial C}{\partial \boldsymbol{u}} = \frac{\partial C}{\partial \boldsymbol{w}} \frac{\mathrm{d}\boldsymbol{w}}{\mathrm{d}\boldsymbol{u}} = \frac{\partial C}{\partial \boldsymbol{w}} \frac{\mathrm{d}g}{\mathrm{d}\boldsymbol{u}} = \frac{\partial C}{\partial \boldsymbol{w}} \boldsymbol{G}$$

$$\frac{\mathrm{d}g}{\mathrm{d}\boldsymbol{u}} = \boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$\boldsymbol{\nabla}_{\boldsymbol{u}} C = \boldsymbol{G}^\top \boldsymbol{\nabla}_{\boldsymbol{w}} C$$

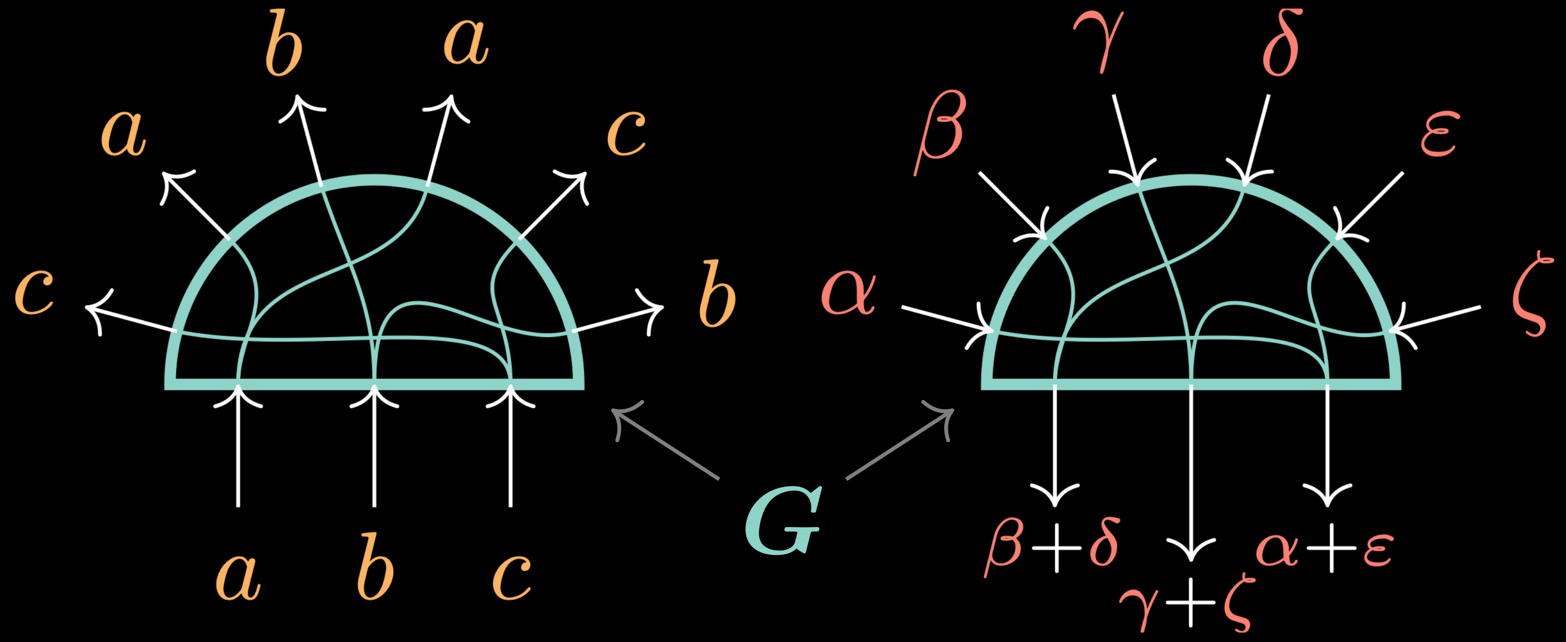
Weight sharing (II)

$$\nabla_{\boldsymbol{u}} C = G^\top \nabla_{\boldsymbol{w}} C$$

$$\nabla_{\boldsymbol{u}} C = \begin{pmatrix} \partial C/\partial u_1 \\ \partial C/\partial u_2 \\ \partial C/\partial u_3 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} \partial C/\partial w_1 \\ \partial C/\partial w_2 \\ \partial C/\partial w_3 \\ \partial C/\partial w_4 \\ \partial C/\partial w_5 \end{pmatrix} \Rightarrow$$

$$\Rightarrow \begin{cases} \dfrac{\partial C}{\partial u_1} = \dfrac{\partial C}{\partial w_1} \\ \dfrac{\partial C}{\partial u_2} = \dfrac{\partial C}{\partial w_2} \\ \dfrac{\partial C}{\partial u_3} = \dfrac{\partial C}{\partial w_3} + \dfrac{\partial C}{\partial w_4} + \dfrac{\partial C}{\partial w_5} \end{cases}.$$

The routing and branching matrix



$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} c \\ a \\ b \\ a \\ c \\ b \end{pmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \varepsilon \\ \zeta \end{pmatrix} = \begin{pmatrix} \beta+\delta \\ \gamma+\zeta \\ \alpha+\varepsilon \end{pmatrix}$$