

(a)

\mathcal{X}_t



\mathcal{X}_{t-1}



$\{\blacksquare, \blacksquare, \dots\} = S_n$
top $p \cdot N$ tokens

$$\downarrow \frac{1}{F} \sum_{f=1}^F E(x_t^{(f)})$$

$$\downarrow \frac{1}{F} \sum_{f=1}^F E(x_{t-1}^{(f)})$$

Z_t

Z_{t-1}

$$\left| \begin{array}{c|ccccc} & 1 & 2 & \dots & N \\ \hline 1 & \text{dark blue} & \text{teal} & \text{yellow} & \text{orange} \\ 2 & \text{orange} & \text{red} & \text{dark blue} & \text{teal} \\ \dots & \text{yellow} & \text{teal} & \text{dark blue} & \text{orange} \\ N & \text{orange} & \text{yellow} & \text{dark blue} & \text{orange} \end{array} \right| - \left| \begin{array}{c|ccccc} & 1 & 2 & \dots & N \\ \hline 1 & \text{teal} & \text{orange} & \text{dark blue} & \text{teal} \\ 2 & \text{yellow} & \text{dark blue} & \text{teal} & \text{orange} \\ \dots & \text{teal} & \text{orange} & \text{dark blue} & \text{yellow} \\ N & \text{dark blue} & \text{teal} & \text{orange} & \text{dark blue} \end{array} \right|_2 = \left| \begin{array}{c|ccccc} & d_1 & d_2 & \dots & d_N \\ \hline & \text{dark blue} & \text{orange} & \text{yellow} & \text{red} \\ & \text{orange} & \text{dark blue} & \text{teal} & \text{teal} \\ & \text{yellow} & \text{teal} & \text{dark blue} & \text{yellow} \\ & \text{red} & \text{teal} & \text{orange} & \text{dark blue} \end{array} \right|$$

$$\|Z_t(i) - Z_{t-1}(i)\|_2 \quad \forall i \in \{1, \dots, N\}$$

(b)

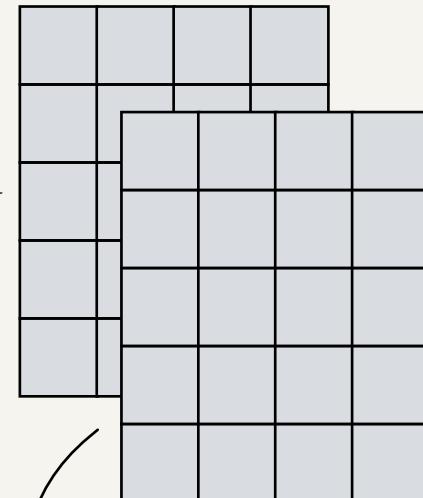
$$D_t \leq \mu_t + \lambda \sigma_t$$



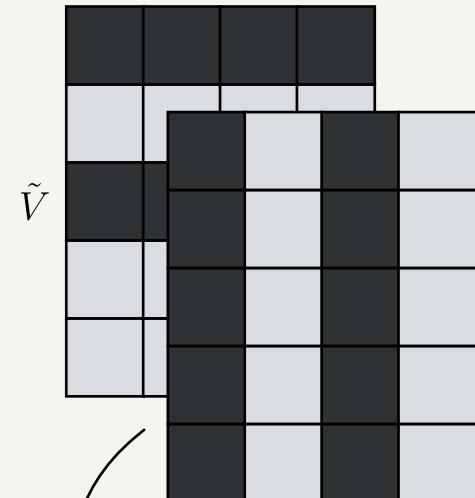
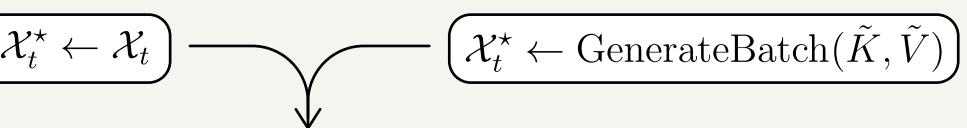
unchanged (K, V)



masked (\tilde{K}, \tilde{V})



$$\mathcal{X}_t^* \leftarrow \mathcal{X}_t$$



$$\mathcal{X}_t^* \leftarrow \text{GenerateBatch}(\tilde{K}, \tilde{V})$$

Update statistics and cache using \mathcal{X}_t^*