Table 1: Taxonomy of decay mechanisms in various linear attention variants. Different implementations exhibit unique parametrization strategies and structural characteristics. Here, j denotes the head index (out of h total heads), and l represents the layer index (out of L total layers). $\mathbf{A}^j, \Delta_t^j, \tau \in \mathbb{R}$, and $\mathbf{f}_t^j \in \mathbb{R}^{d/h}$ for vector decay or $\mathbf{f}_t^j \in \mathbb{R}$ for scalar decay. lse represents the logsumexp operator, i.e., $\text{lse}(\mathbf{x}) = \log \sum \exp(x_i)$, and $\text{sigmoid}(\mathbf{x}) = 1/(1+\exp(-\mathbf{x}))$.

Method	**Parameterization Strategy**	**Parameter Sharing**	**Scalar**	**Recurrence Formula**
Mamba2	$\lambda_t^j = \text{sigmoid}\left(-\mathbf{f}_t^j - \Delta^j\right)^{\exp(\mathbf{A}^j)}$	55	✓	$\mathbf{s}_t^j = \lambda_t^j \mathbf{s}_{t-1}^j + \mathbf{k}_t^j (\mathbf{v}_t^j)^\top$
Mamba2 wo $\mathbf{A}$	$\lambda_t^j = \text{sigmoid}\left(-\mathbf{f}_t^j - \Delta^j\right)$	55	✓	
Mamba2 wo Δ	$\lambda_t^j = \text{sigmoid}\left(-\mathbf{f}_t^j\right)^{\exp(\mathbf{A}^j)}$	55	✓	
Mamba2 wo $\mathbf{A}$ & Δ	$\lambda_t^j = \text{sigmoid}\left(-\mathbf{f}_t^j\right)$	55	✓	
GLA	$\lambda_t^j = \text{sigmoid}(\mathbf{f}_t^j)^{1/\tau}$	55	55	$\mathbf{s}_t^j = \text{diag}(\lambda_t^j)\mathbf{s}_{t-1}^j + \mathbf{k}_t^j (\mathbf{v}_t^j)^\top$
Hgrn2	$\lambda_t^j = \lambda^j + (1-\lambda^j)\text{sigmoid}(\mathbf{f}_t^j)$	✓	55	$\mathbf{s}_t^j = \text{diag}(\lambda_t^j)\mathbf{s}_{t-1}^j + (1-\lambda_t^j)(\mathbf{v}_t^j)^\top$
Lightnet	$\lambda_t^j = \exp(\text{lse}(\mathbf{f}_{<t-1}^j) - \text{lse}(\mathbf{f}_{<t}^j))$	✓	55	$\mathbf{s}_t^j = \text{diag}(\lambda_t^j)\mathbf{s}_{t-1}^j + (1-\lambda_t^j)(\mathbf{v}_t^j)^\top$
TNL	$\lambda_t^j = \exp\left(-8j/h \times (1-l/L)\right)$	55	✓	$\mathbf{s}_t^j = \lambda^j \mathbf{s}_{t-1}^j + \mathbf{k}_t^j (\mathbf{v}_t^j)^\top$
Simple Decay	$\lambda_t^j = \text{sigmoid}\left(\mathbf{f}_t^j + \Delta^j\right)$	55	both	$\mathbf{s}_t^j = \text{diag}(\lambda_t^j)\mathbf{s}_{t-1}^j + \mathbf{k}_t^j (\mathbf{v}_t^j)^\top$