

## Discrete Diffusion Generation



Iterative Denoising ↓



Image / Text  
Prompt

Image / Text  
Response

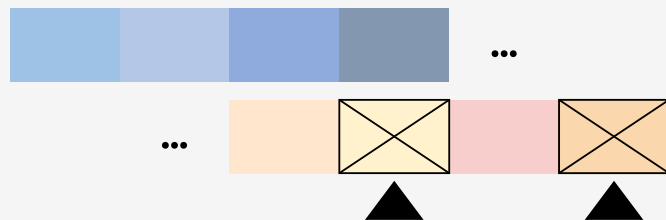
Input Unmask  
Token

Completion  
Token

Mask  
Token

Remasked  
Token

## Remask Response for Masked Log Probs




Masked Token Log Probs

$$\pi'_\theta(o_i | q, p) = \frac{1}{|M|} \sum_{o_{i,t} \in M} \log p_\theta(o_{i,t} | q)$$

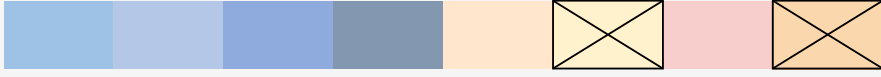
## Uniformly Random Masking in Iterations

Iteration 1  
 $p = 0.25$



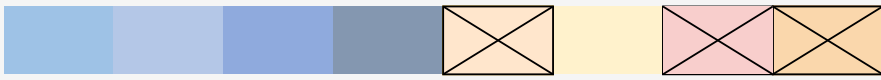
→  $\pi'_\theta(o, p_1)$

Iteration 2  
 $p = 0.50$



→  $\pi'_\theta(o, p_2)$

Iteration 3  
 $p = 0.75$



→  $\pi'_\theta(o, p_3)$

...

...

Sequence Log Probs by Averaging  $\pi'_\theta = \frac{1}{\mu} \sum_{p=1}^{\mu} \pi'_\theta(o, p_i)$