Modeling

$u_s(Y_s \mid X_t)$

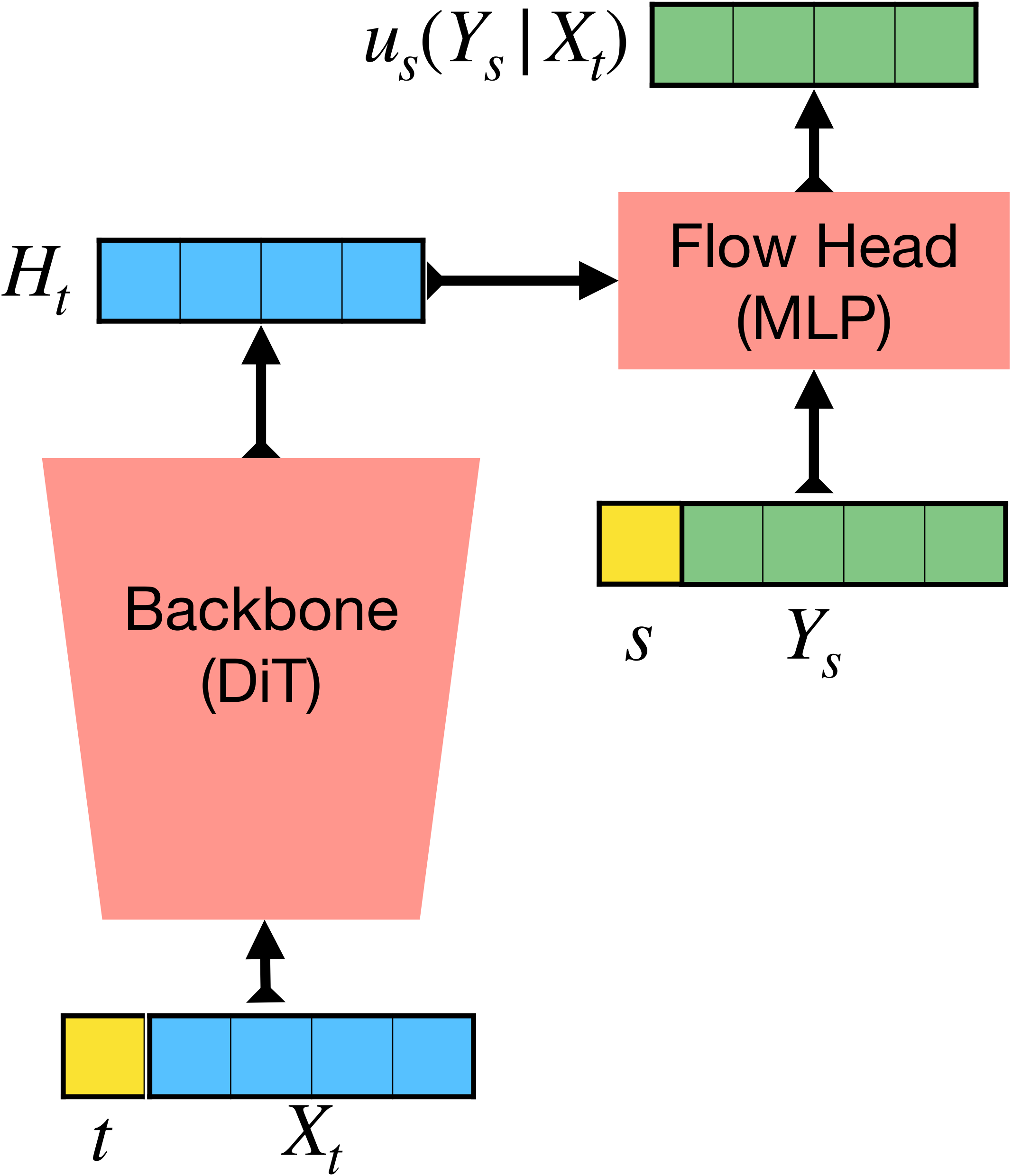H_t

Flow Head (MLP)

Backbone (DiT)

s Y_s

t X_t

Loss

$$\mathcal{L}(\theta) = \mathbb{E}\left[\|u_s^\theta(Y_s \mid X_t) - \dot{Y}_s\|^2\right],$$

where

$$s \in [0,1],$$

$$Y_s = (1-s)Y_0 + sY,$$

$$Y_0 \sim \mathcal{N}(0, I),$$

X_t, Y - Supervising process

Modeling

Loss

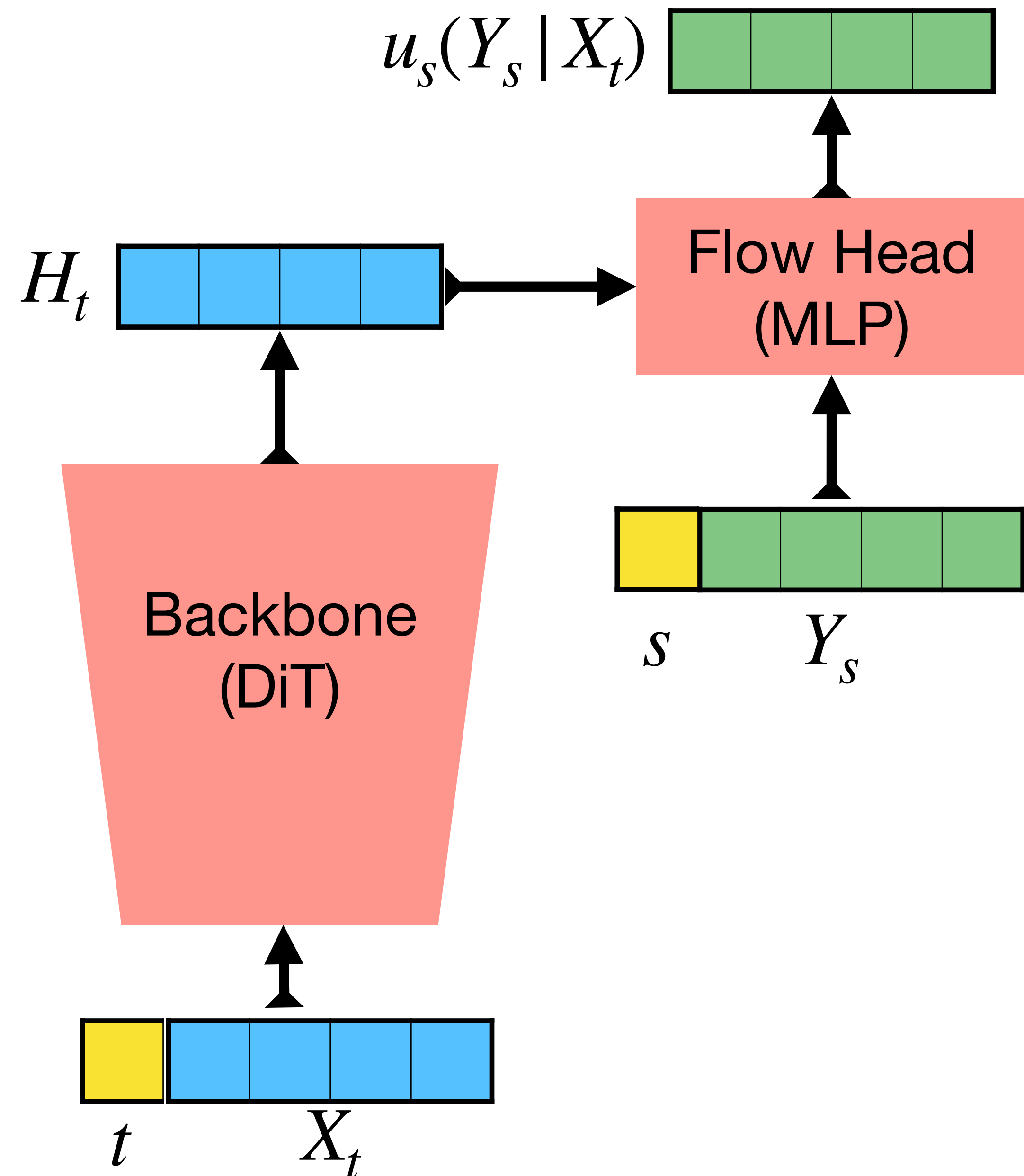$$\mathcal{L}(\theta) = \mathbb{E}\left[\|u_s^\theta(Y_s \mid X_t) - \dot{Y}_s\|^2\right],$$

where

$s \in [0,1],$

$Y_s = (1-s)Y_0 + sY,$

$Y_0 \sim \mathcal{N}(0, I),$

X_t, Y - Supervising process

$u_s(Y_s \mid X_t)$

H_t

Flow Head (MLP)

s $\quad$ Y_s

Backbone (DiT)

t $\quad$ X_t

DTM vs. FM