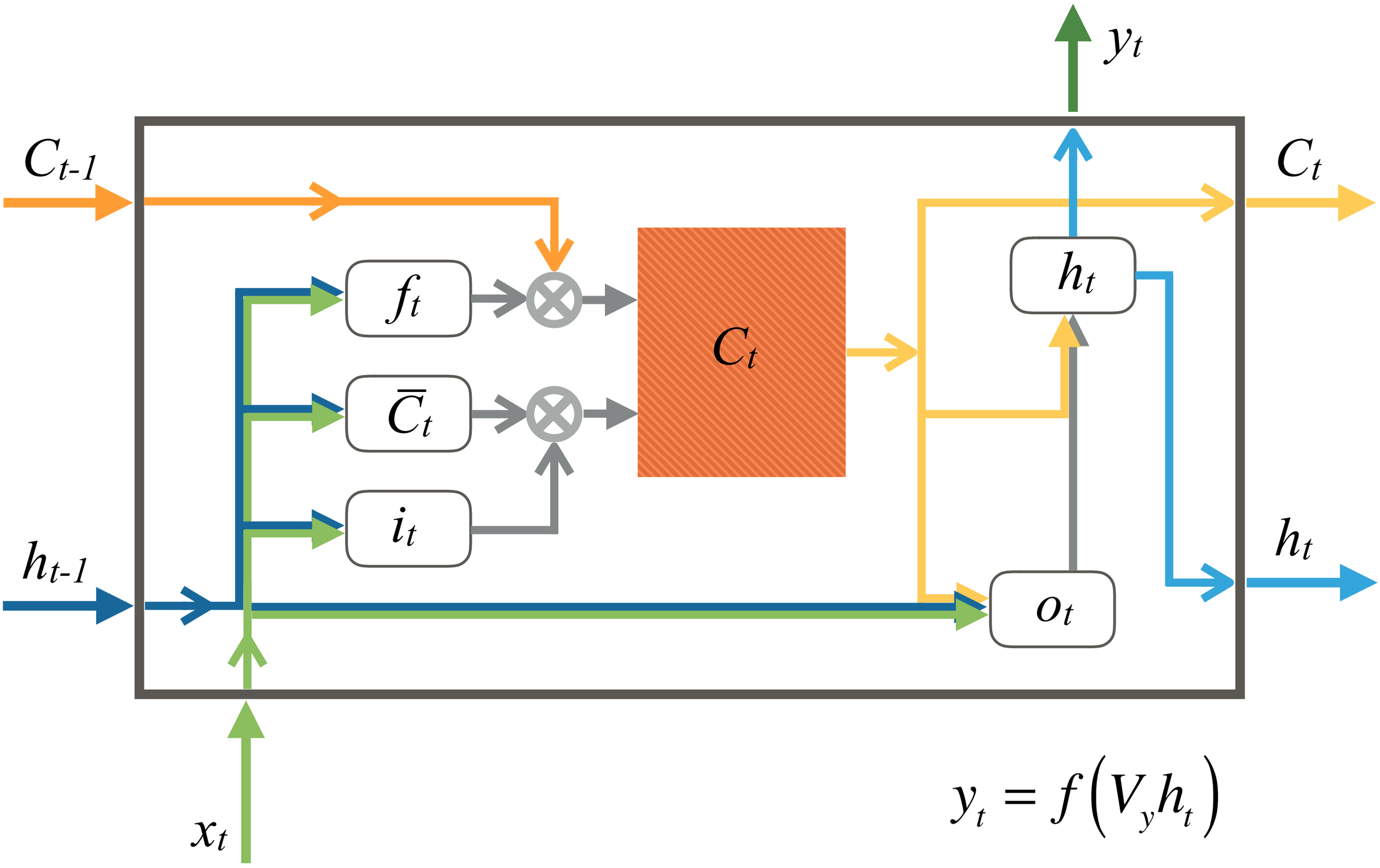
LSTM MEMORY BLOCK: ALL TOGETHER



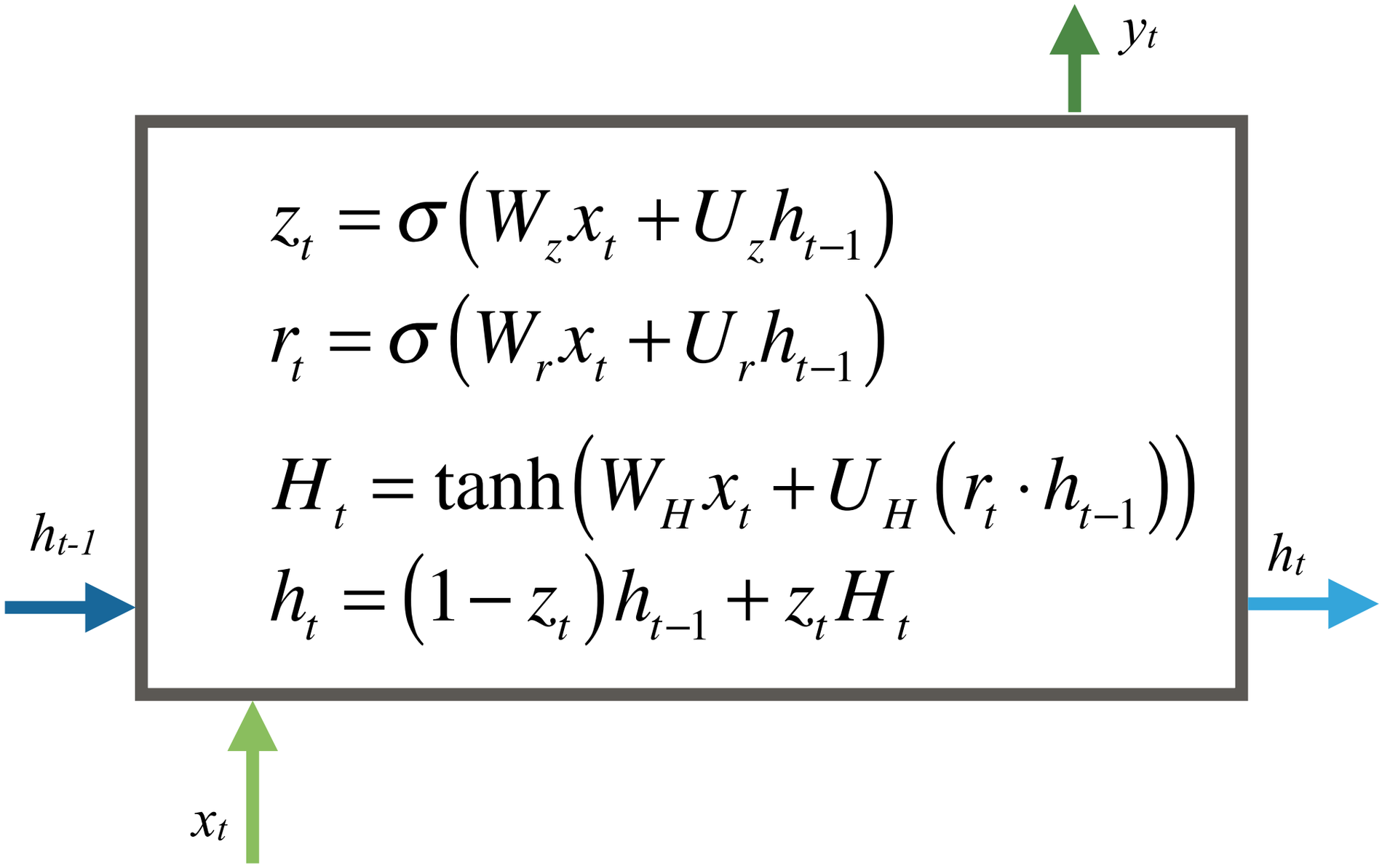
$$y_t = f\left(V_y h_t\right)$$

GATED RECURRENT UNIT

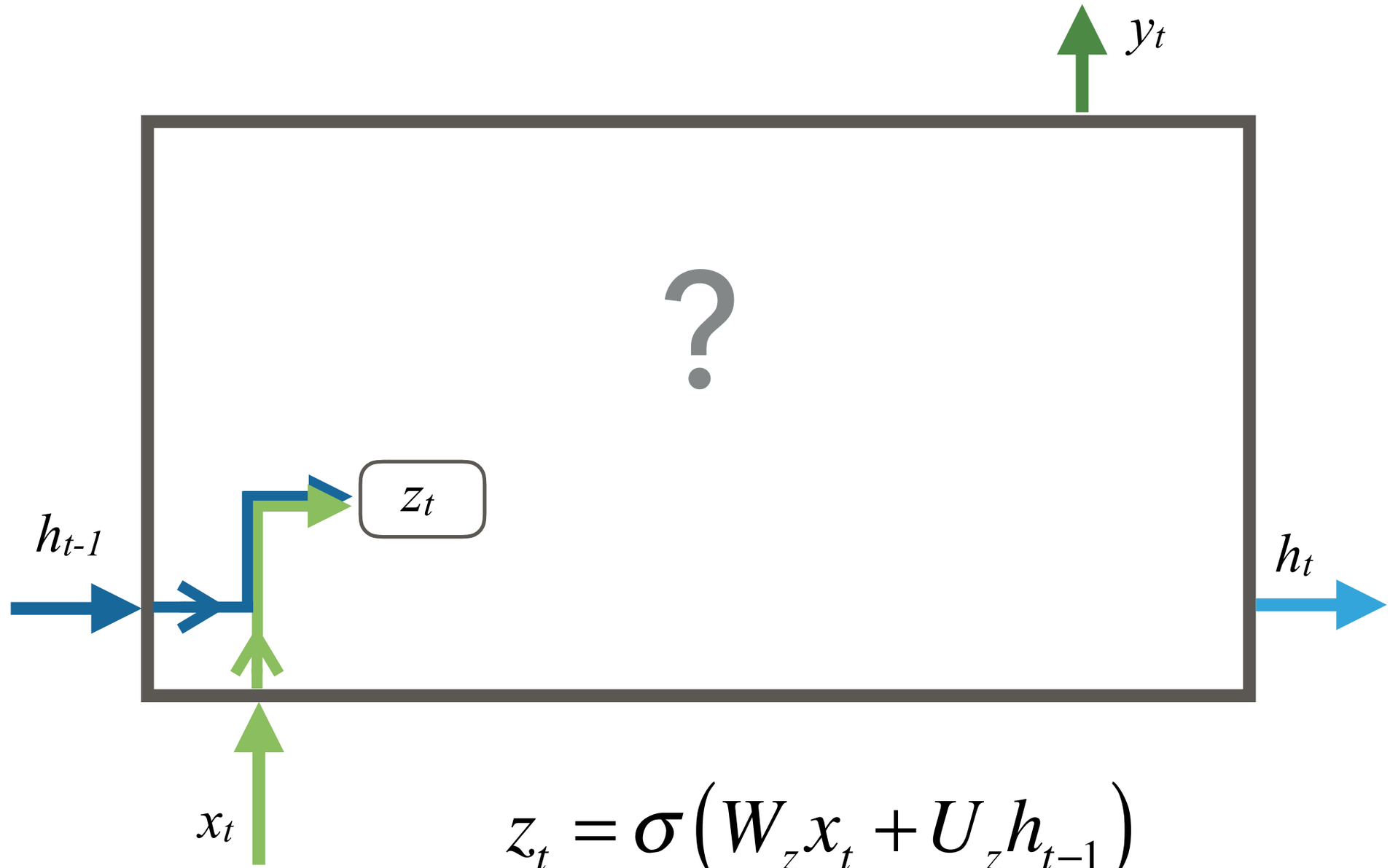
GATED RECURENT UNIT (GRU)

- Proposed by Cho et al. [2014].
- It is similar to LSTM in using gating functions, but differs from LSTM in that it doesn't have a memory cell.
- Each GRU consists of:
  - **update gate**
  - **reset get**
- Model parameters:
  - x_t is the input at time t
  - Weight matrices: $W_z, W_r, W_H, U_z, U_r, U_H$

GRU


y_t
$z_t = \sigma\left(W_z x_t + U_z h_{t-1}\right)$
$r_t = \sigma\left(W_r x_t + U_r h_{t-1}\right)$
$H_t = \tanh\left(W_H x_t + U_H\left(r_t \cdot h_{t-1}\right)\right)$
$h_t = \left(1 - z_t\right) h_{t-1} + z_t H_t$
h_{t-1}
h_t
x_t

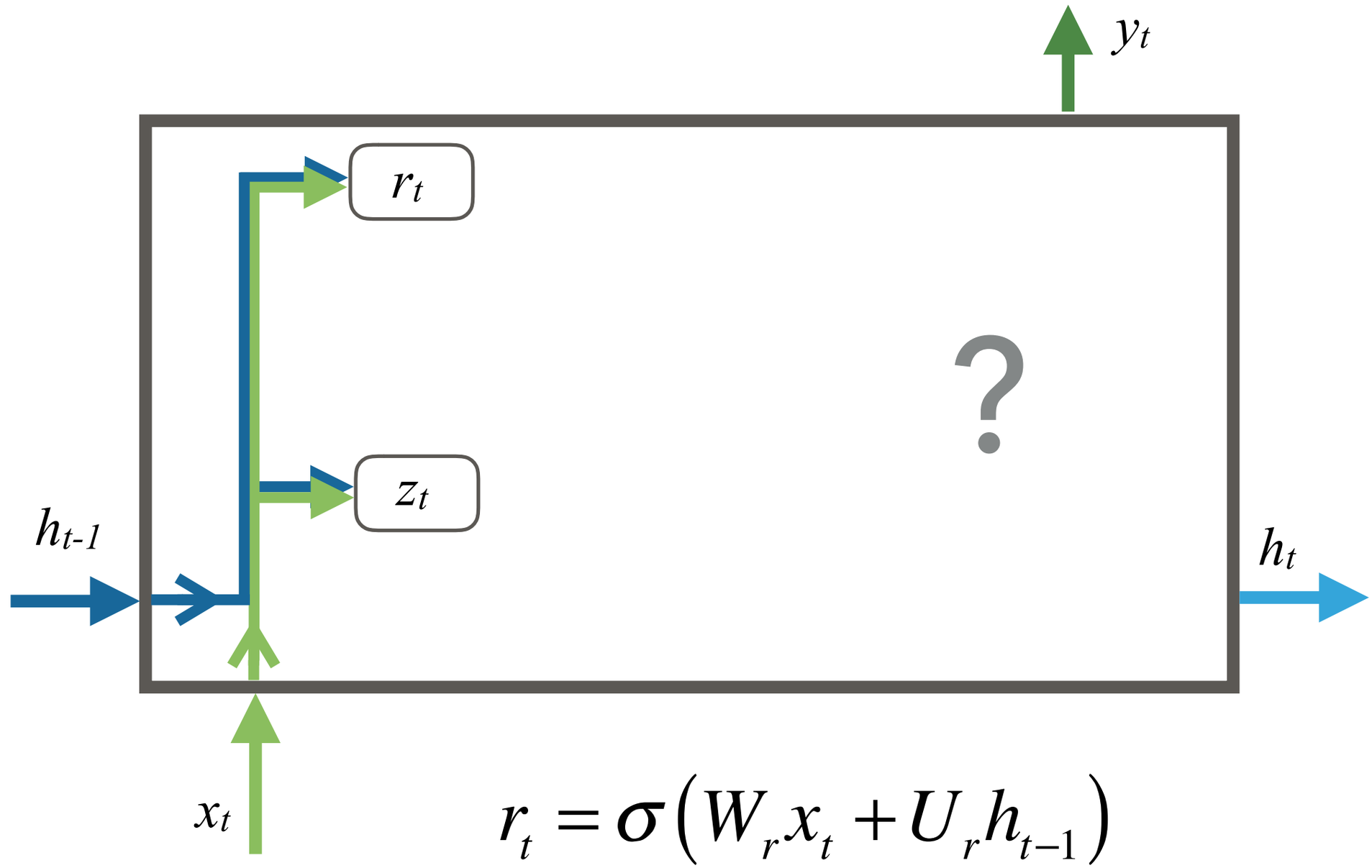
GRU: UPDATE GATE



$$z_t = \sigma\left(W_z x_t + U_z h_{t-1}\right)$$

- Update gate z_t decides how much unit update its activation or content

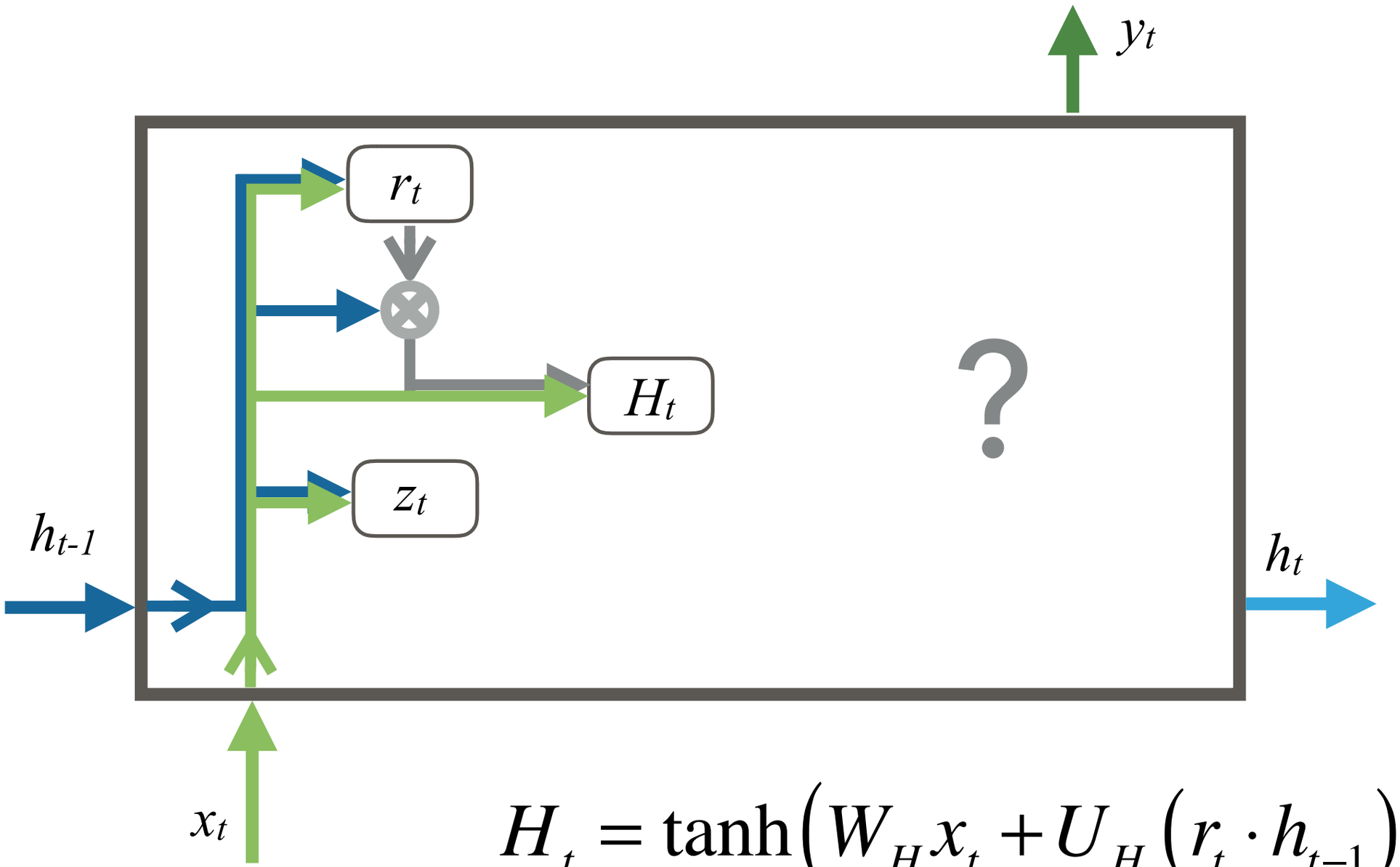
GRU: RESET GATE



$$r_t = \sigma\left(W_r x_t + U_r h_{t-1}\right)$$

- When r_t close to 0 (gate off), it makes the unit act as it's reading the first symbol from the input sequence, allowing it to forget previously computed states

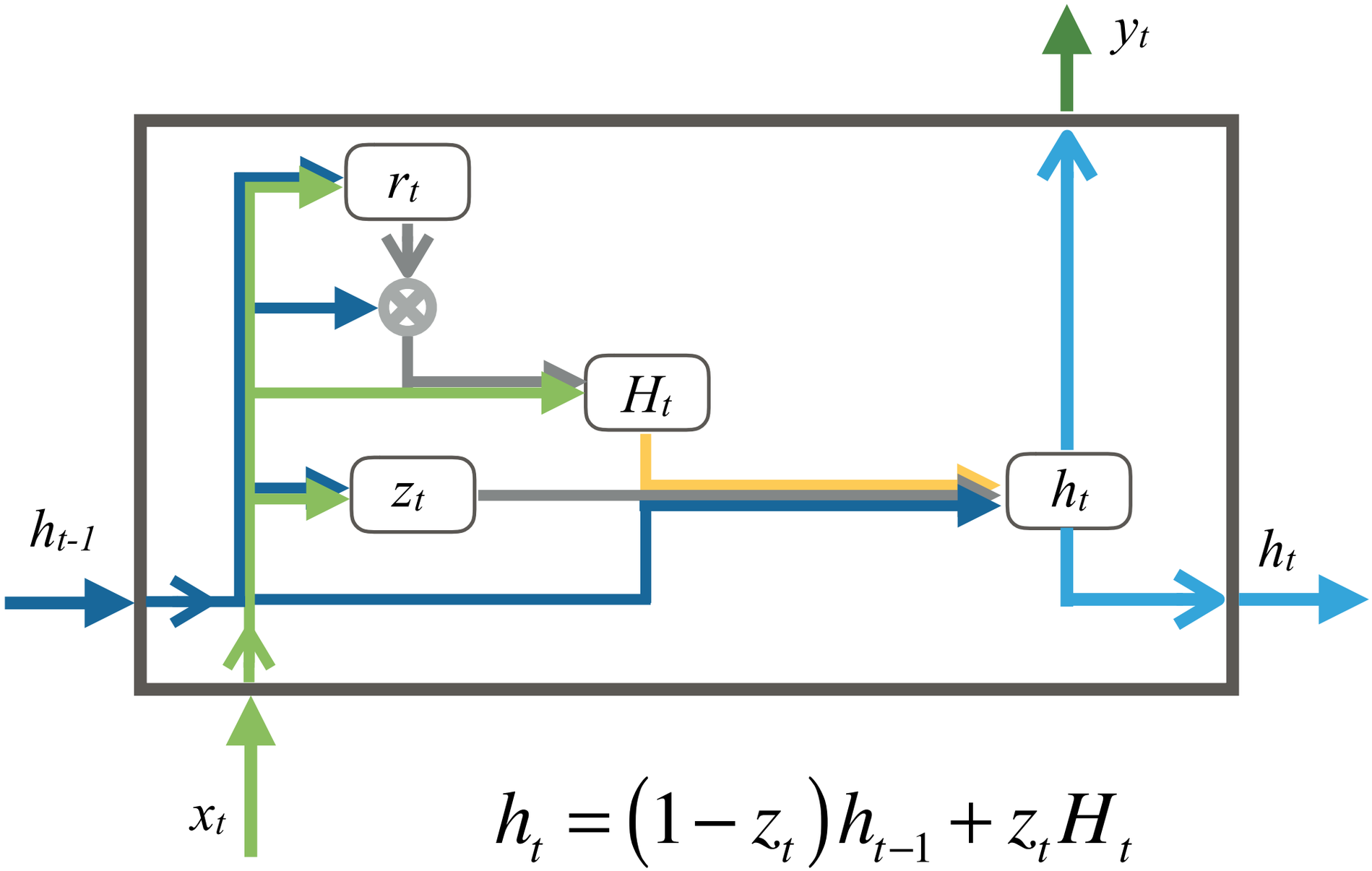
GRU: CANDIDATE ACTIVTION



$$H_t = \tanh\left(W_H x_t + U_H\left(r_t \cdot h_{t-1}\right)\right)$$

- Update gate z_t decides how much unit update its activation or content.

GRU: HIDDEN STATE



$$h_t = (1 - z_t) h_{t-1} + z_t H_t$$

- Activation at time t is the linear interpolation between previous activations h_{t-1} and candidate activation H_t

GRU: ALL TOGETHER