1) Autoregressive self attention where every token attends to its position and all previous positions.
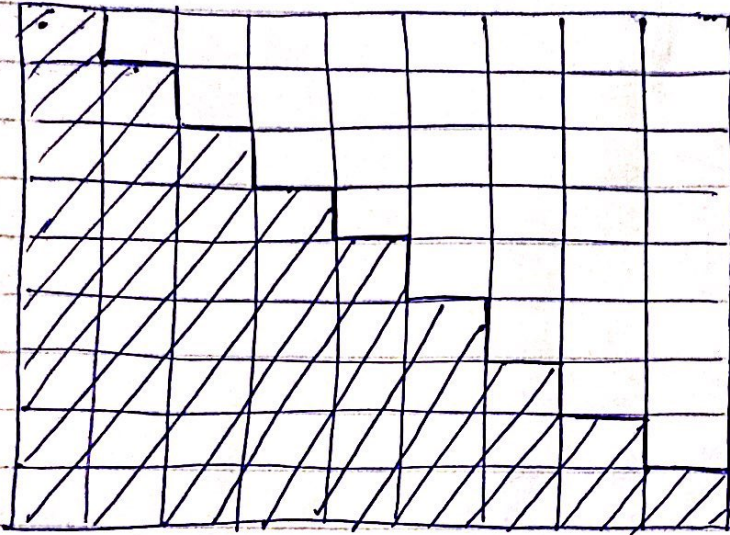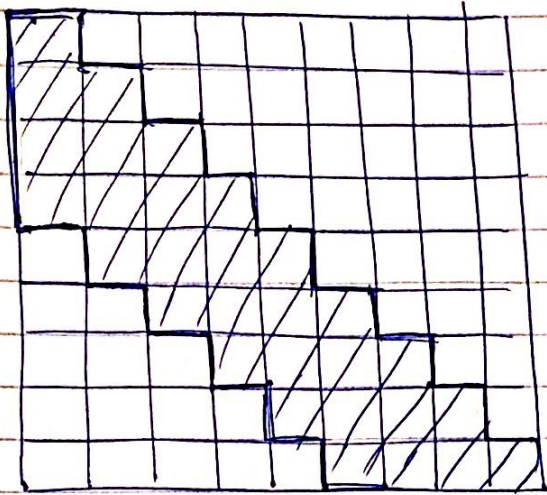

$n = 9$.

No. of dot products required $= \dfrac{n(n+1)}{2}$

② Strided self attention where every token attends to
at most t positions prior to it.

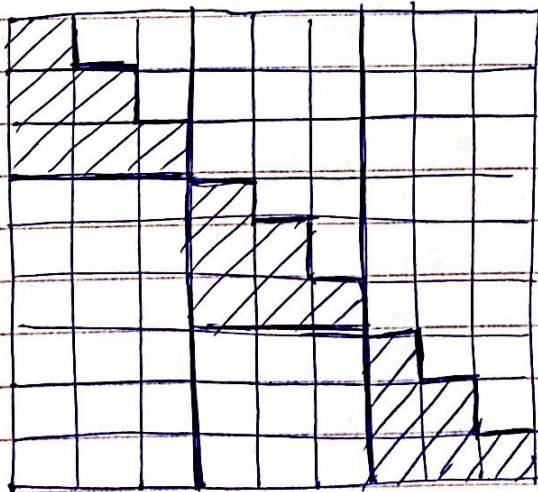$n = 9$ Number of dot products required =
$t = 4$. $$(n-t)(t+1) + \frac{t(t+1)}{2}$$

③ Windowed self-attention where the n tokens are
partitioned into windows of size w, and every token
                              depends on all positions
                              within its window prior to it.

$n = 9$ Number of dot products required
$w = 3$. $$= \frac{n}{w}\left(\frac{w(w+1)}{2}\right)$$