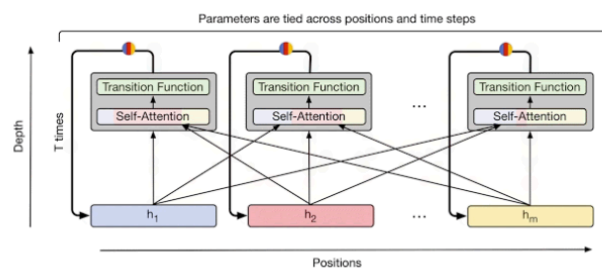


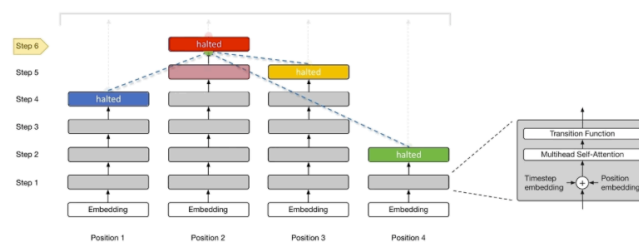
## HW4

- Q1:

- Universal Transformer
- 將同一組共用參數的layer重複呼叫多次，所有的layer都使用同一個參數(並非和普通transformer一樣每層參數可能不同)。



- 此外，藉由讓model藉由機率決定每個字的深度，讓每個字都可以有不同的深度 (其他層數會複製同一個embedding讓較高的層數可以繼續做self-attention)



- Q2:

- Universal Transformer可以藉由使用更多的memory來有效的improve效能，能夠處理更多的task