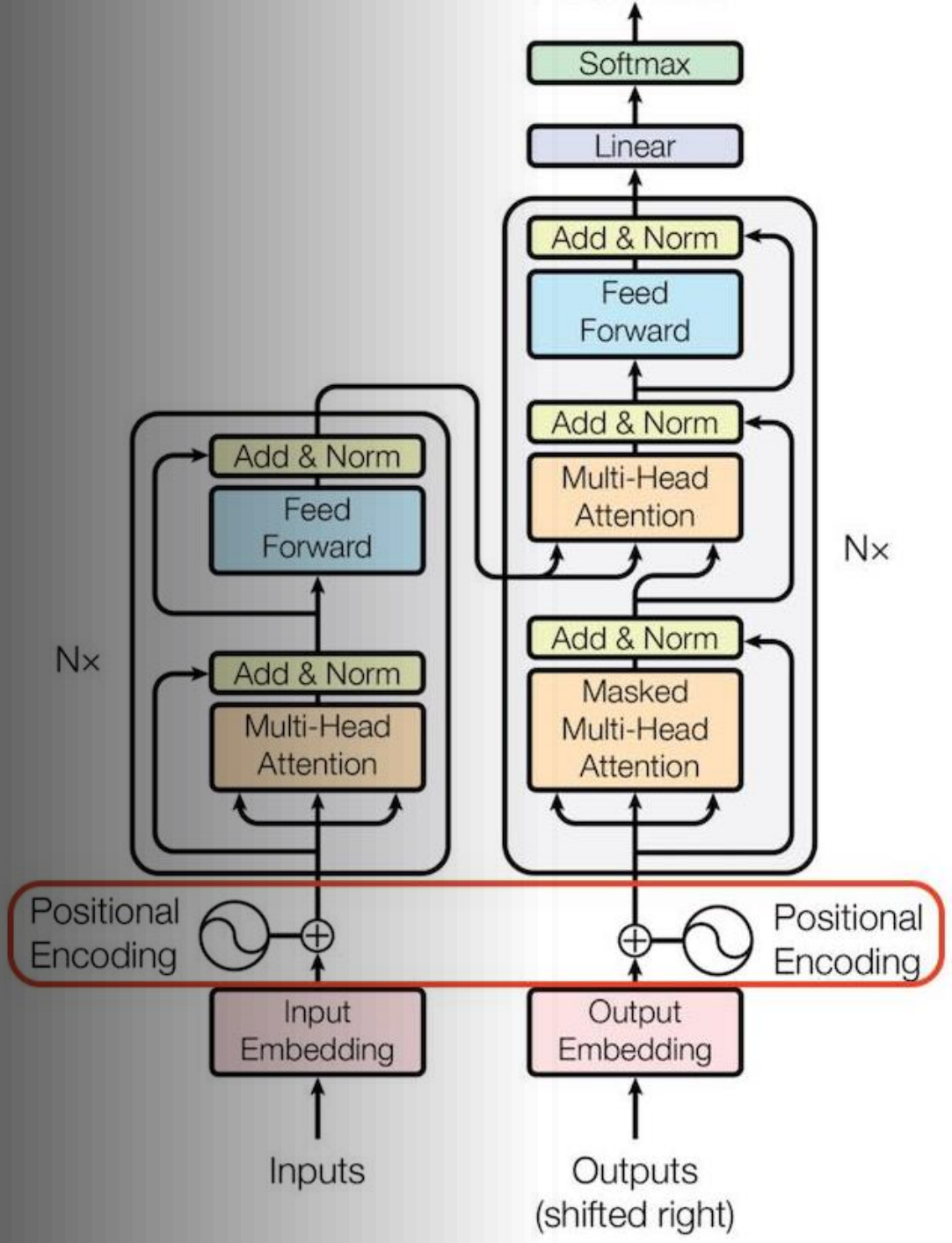
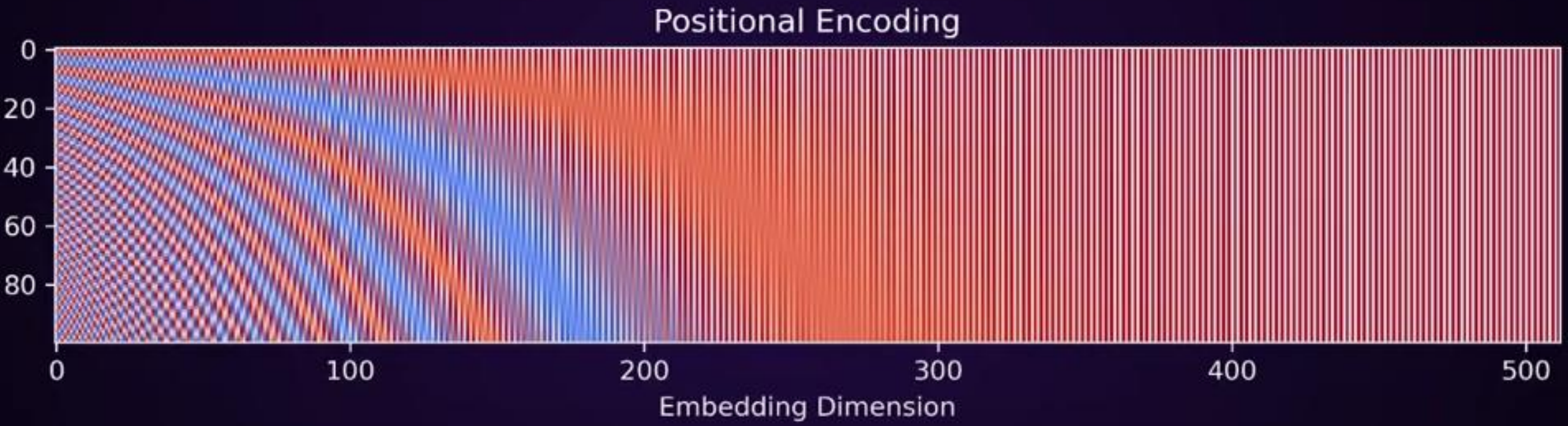
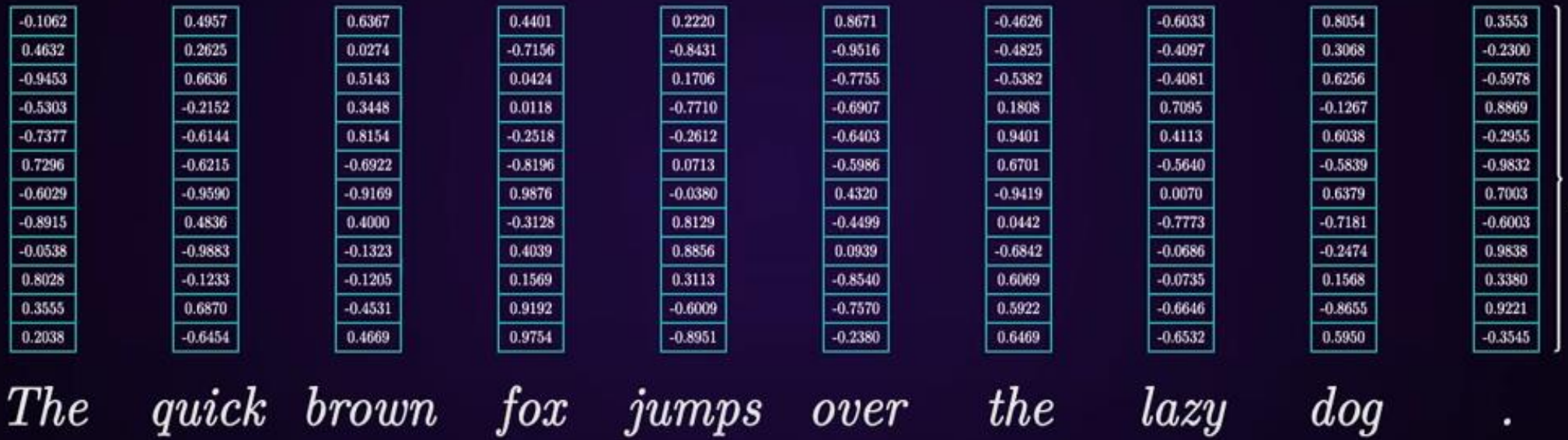
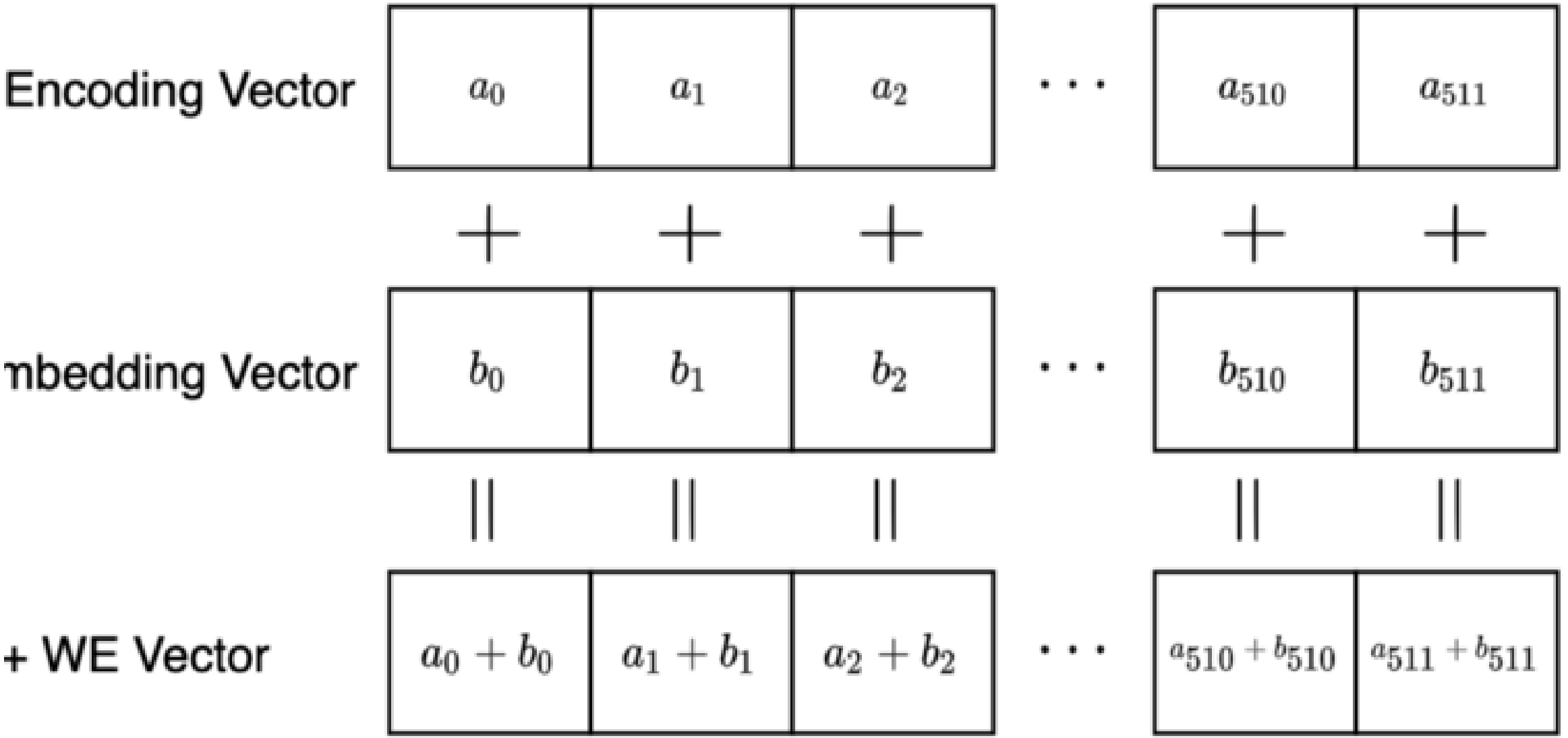
Positional encoding

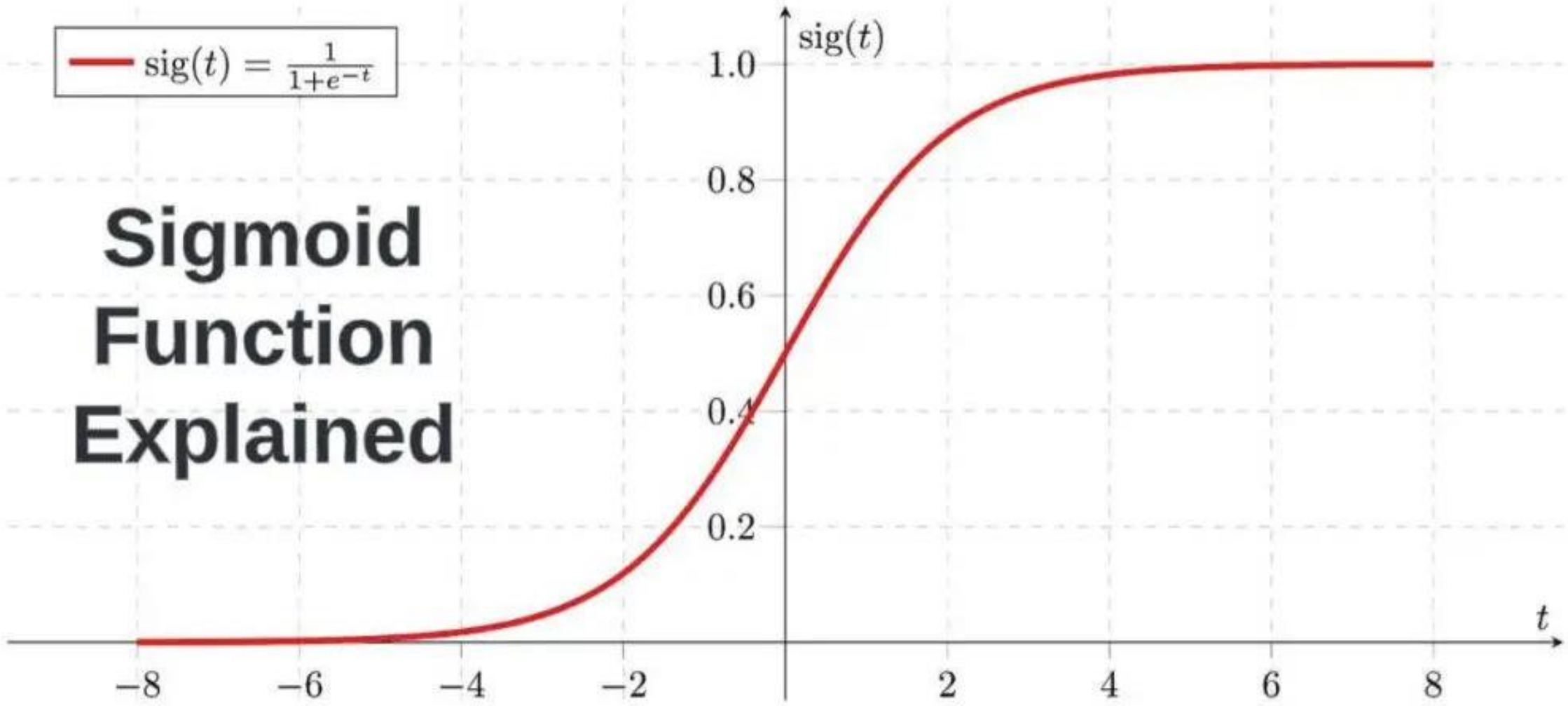
Ahmed ibrahim

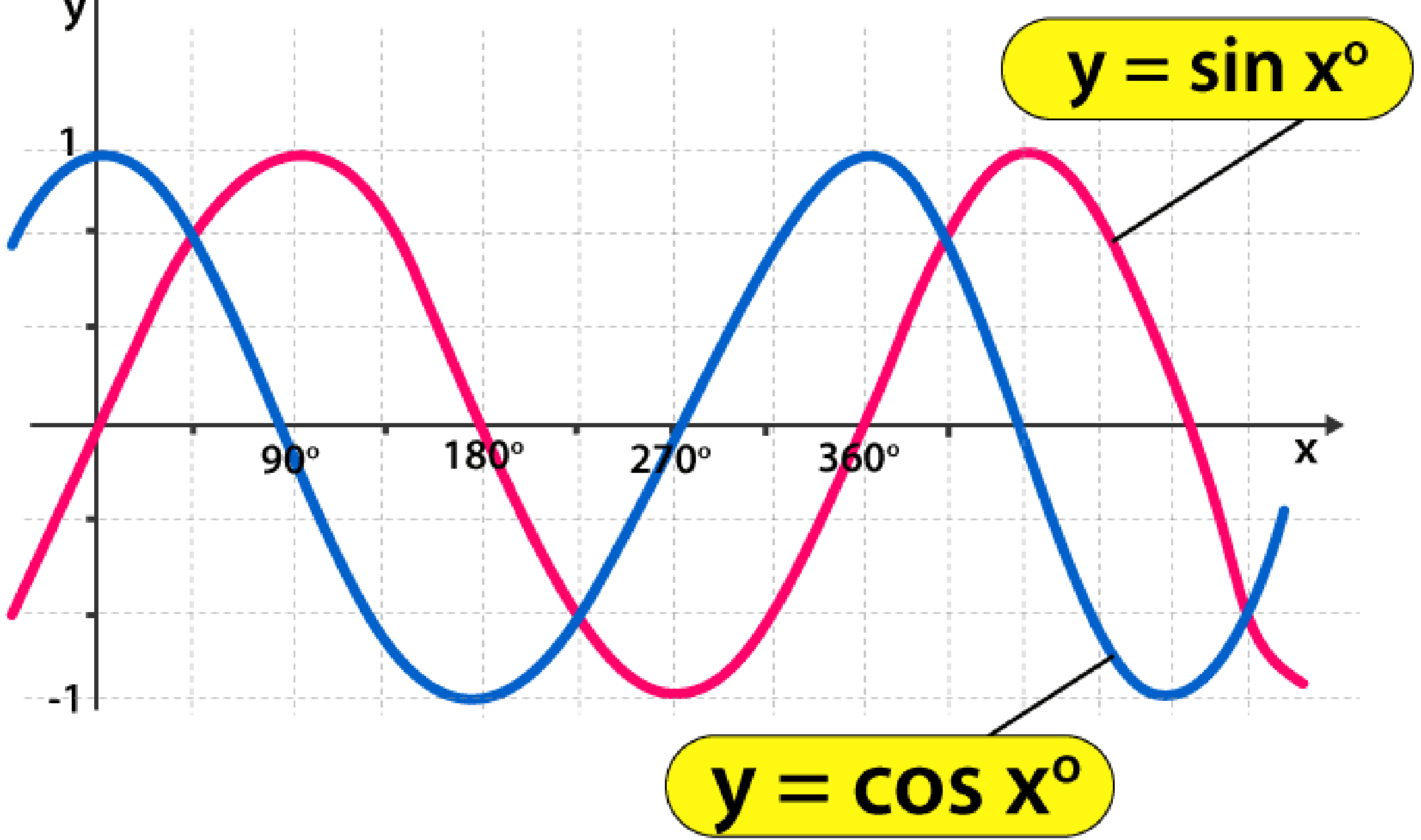



Positional Encoding
0
20
40
60
80
0
100
200
300
400
500
Embedding Dimension


-0.1062 0.4632 -0.9453 -0.5303 -0.7377 0.7296 -0.6029 -0.8915 -0.0538 0.8028 0.3555 0.2038
0.4957 0.2625 0.6636 -0.2152 -0.6144 -0.6215 -0.9590 0.4836 -0.9883 -0.1233 0.6870 -0.6454
0.6367 0.0274 0.5143 0.3448 0.8154 -0.6922 -0.9169 0.4000 -0.1323 -0.1205 -0.4531 0.4669
0.4401 -0.7156 0.0424 0.0118 -0.2518 -0.8196 0.9876 -0.3128 0.4039 0.1569 0.9192 0.9754
0.2220 -0.8431 0.1706 -0.7710 -0.2612 0.0713 -0.0380 0.8129 0.8856 0.3113 -0.6009 -0.8951
0.8671 -0.9516 -0.7755 -0.6907 -0.6403 -0.5986 0.4320 -0.4499 0.0939 -0.8540 -0.7570 -0.2380
-0.4626 -0.4825 -0.5382 0.1808 0.9401 0.6701 -0.9419 0.0442 -0.6842 0.6069 0.5922 0.6469
-0.6033 -0.4097 -0.4081 0.7095 0.4113 -0.5640 0.0070 -0.7773 -0.0686 -0.0735 -0.6646 -0.6532
0.8054 0.3068 0.6256 -0.1267 0.6038 -0.5839 0.6379 -0.7181 -0.2474 0.1568 -0.8655 0.5950
0.3553 -0.2300 -0.5978 0.8869 -0.2955 -0.9832 0.7003 -0.6003 0.9838 0.3380 0.9221 -0.3545
d_{model}
The quick brown fox jumps over the lazy dog .

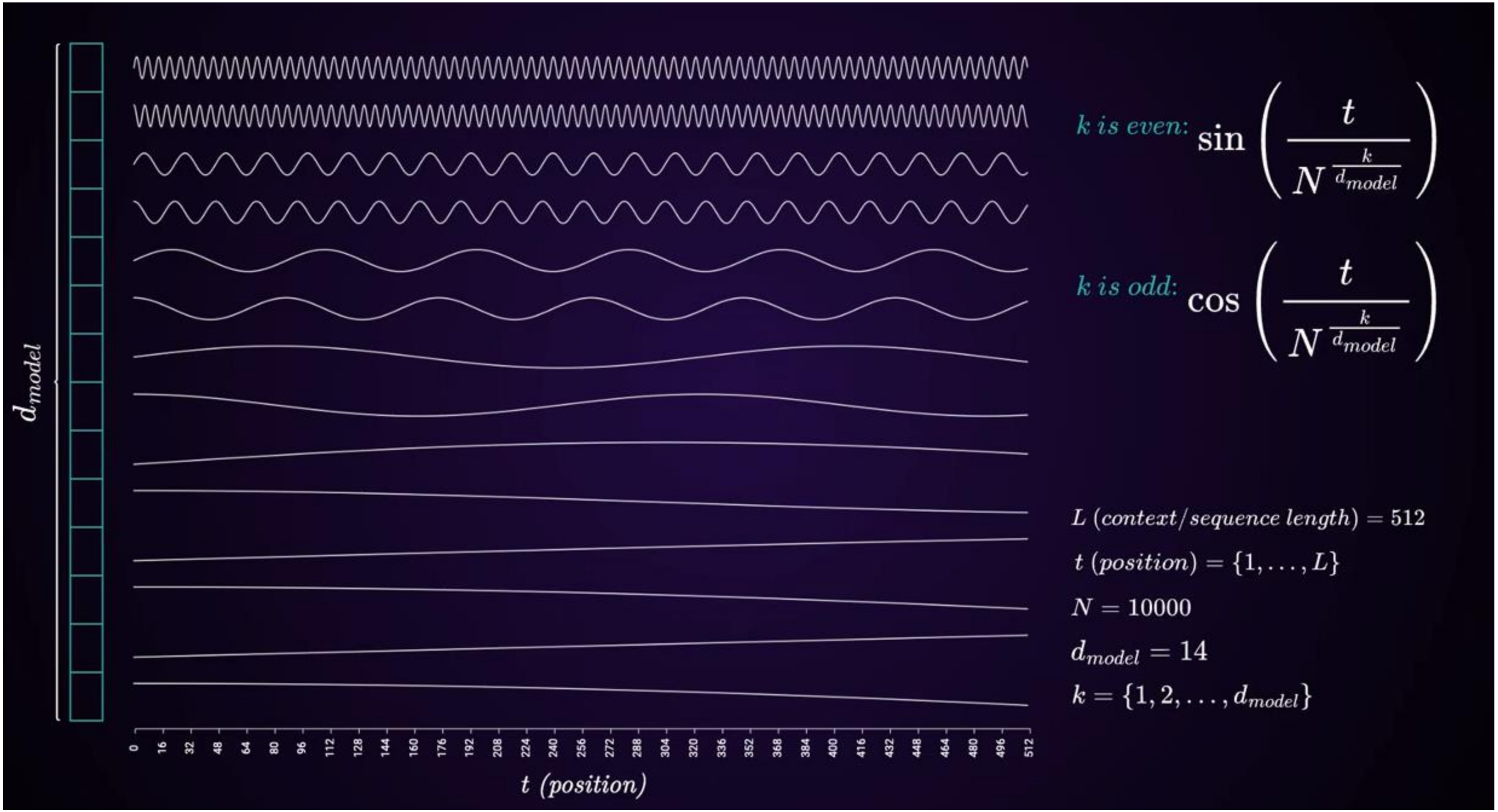

Encoding Vector
a_0 a_1 a_2 ... a_{510} a_{511}
+ + + + +
nbedding Vector
b_0 b_1 b_2 ... b_{510} b_{511}
|| || || || ||
+ WE Vector
$a_0 + b_0$ $a_1 + b_1$ $a_2 + b_2$ ... $a_{510} + b_{510}$ $a_{511} + b_{511}$

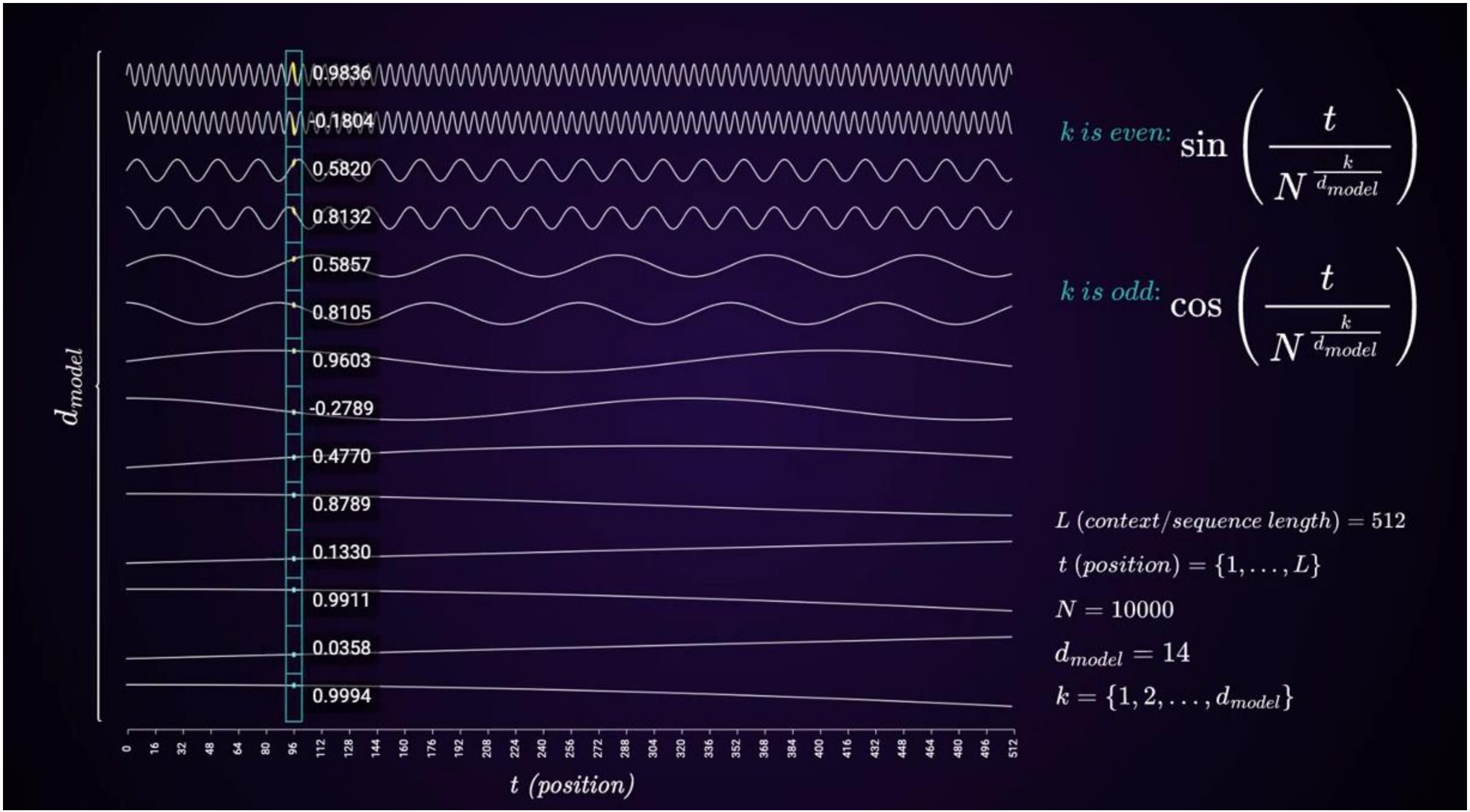

$\text{sig}(t) = \frac{1}{1+e^{-t}}$
Sigmoid Function Explained
sig(t)
1.0
0.8
0.6
0.4
0.2
t
−8
−6
−4
−2
2
4
6
8

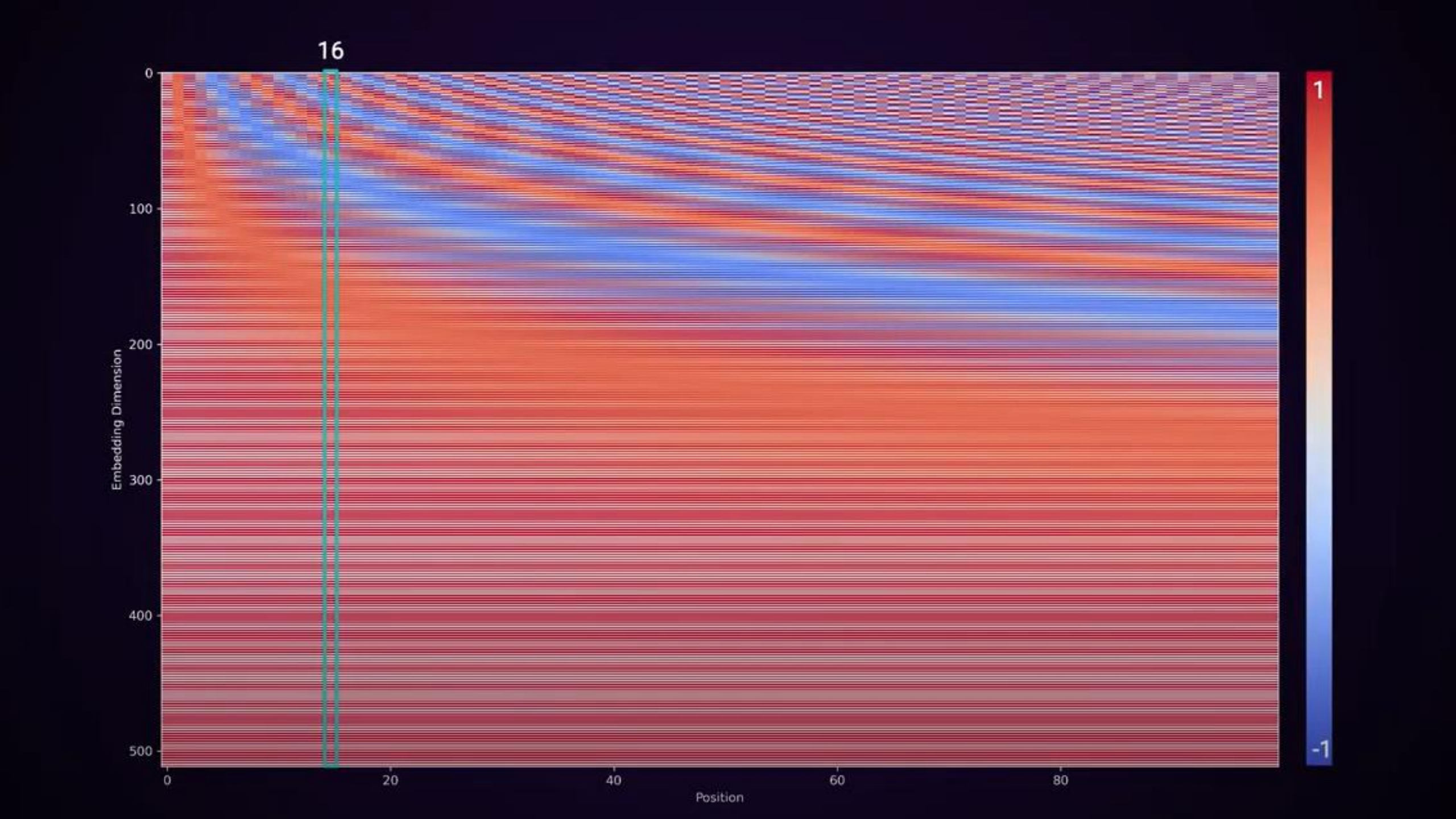

y
1
-1
90°
180°
270°
360°
x
y = sin x°
y = cos x°

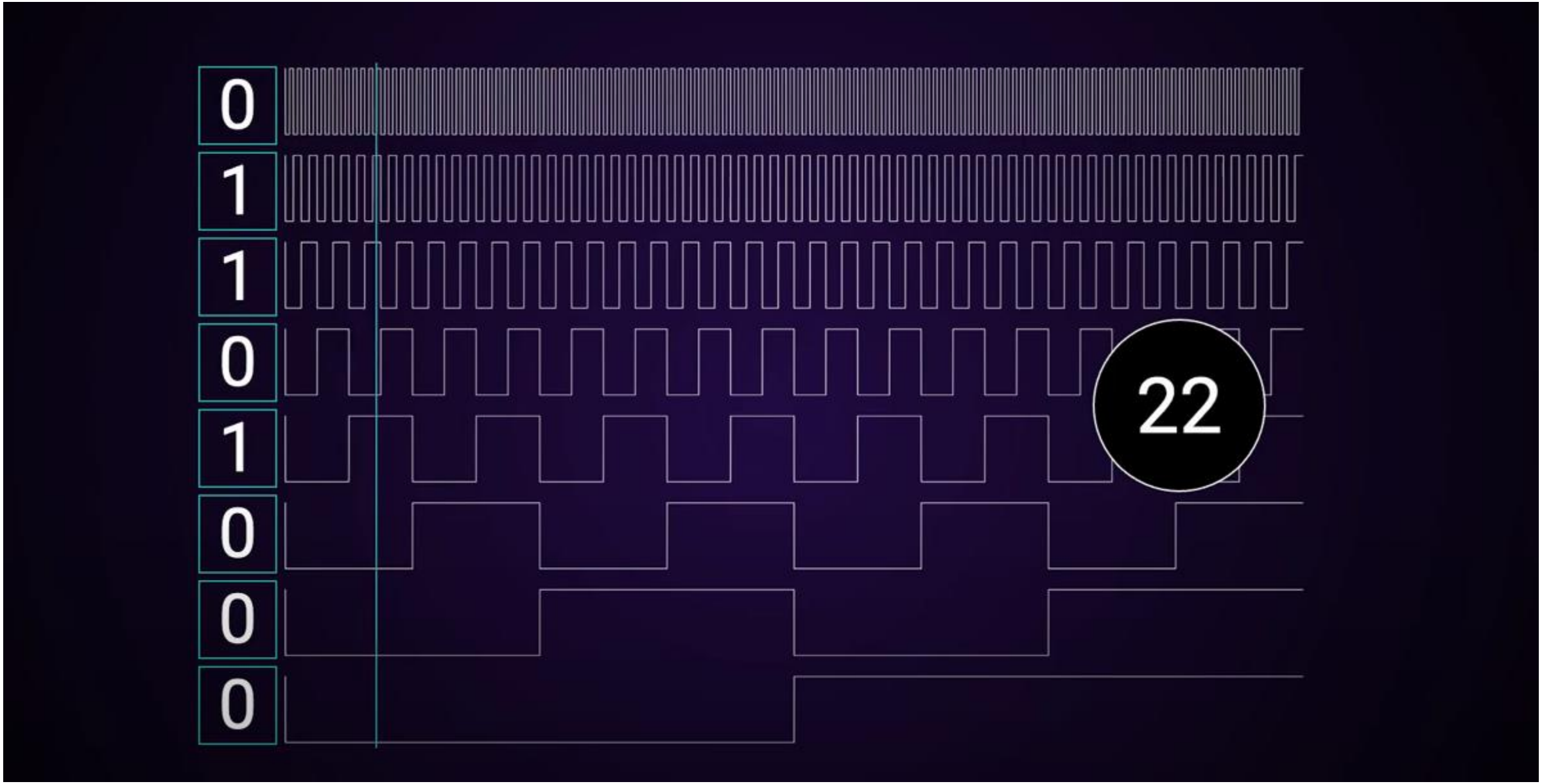
$$\mathrm{PE}(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right)$$

$$\mathrm{PE}(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right)$$


d_{model}
0
16
32
48
64
80
96
112
128
144
160
176
192
208
224
240
256
272
288
304
320
336
352
368
384
400
416
432
448
464
480
496
512
t (position)
k is even: $\sin\left(\frac{t}{N^{\frac{k}{d_{model}}}}\right)$
k is odd: $\cos\left(\frac{t}{N^{\frac{k}{d_{model}}}}\right)$
L (context/sequence length) $= 512$
t (position) $= \{1, \ldots, L\}$
$N = 10000$
$d_{model} = 14$
$k = \{1, 2, \ldots, d_{model}\}$


0.9836
-0.1804
0.5820
0.8132
0.5857
0.8105
0.9603
-0.2789
0.4770
0.8789
0.1330
0.9911
0.0358
0.9994
d_{model}
0 16 32 48 64 80 96 112 128 144 160 176 192 208 224 240 256 272 288 304 320 336 352 368 384 400 416 432 448 464 480 496 512
t (position)
k is even: $\sin\left(\frac{t}{N^{\frac{k}{d_{model}}}}\right)$
k is odd: $\cos\left(\frac{t}{N^{\frac{k}{d_{model}}}}\right)$
L (context/sequence length) $= 512$
t (position) $= \{1, \ldots, L\}$
$N = 10000$
$d_{model} = 14$
$k = \{1, 2, \ldots, d_{model}\}$


16
0
100
200
300
400
500
Embedding Dimension
0
20
40
60
80
Position
1
-1


0
1
1
0
1
0
0
0
22

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$