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

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

In the sentence "I will go to school", the pos values for "I", "will", "go", "to" and "school" are 0,1,2,3 and 4 respectively. d_{model} is 2 if we use 2 embedding values for each token. Its value is 3 if we use 3 embedding values for each token.

Suppose that: I have the sentence "I will go to school". We are using 4 embedding values for each word/token. Here sequence length is 5. Embedding dimension ($d_{model} = 4$). For the word "go" its pos value is 2.

For the word "go", we calculate the encoding for each dimension:

$$PE_{(2,0)} = \sin \left(\frac{2}{10000^{0/4}} \right) = \sin(2/1) \approx 0.9093$$

$$PE_{(2,1)} = \cos \left(\frac{2}{10000^{0/4}} \right) = \cos(2/1) \approx -0.4161$$

$$PE_{(2,2)} = \sin \left(\frac{2}{10000^{2/4}} \right) = \sin \left(\frac{2}{10000^{1/2}} \right) = \sin \left(\frac{2}{100} \right) \approx 0.019998$$

$$PE_{(2,3)} = \cos \left(\frac{2}{10000^{2/4}} \right) = \cos(0.02) \approx 0.9998$$

These values are then added to the token embeddings, providing positional context.