# Positional Encoding Methods

## Sinusoidal
*Original Transformer*

```
PE(pos, 2i) =
sin(pos / 10000^(2i/d_model))
```

```
PE(pos, 2i+1) =
cos(pos / 10000^(2i/d_model))
```

✓ **Advantages**

🎯 Works for **unseen sequence lengths**

📐 Deterministic pattern

♾️ Generalizes to any length

## Learned
*BERT Approach*

```
Trainable embedding
matrix:

PE ∈ ℝ^(max_len × d_model)
```

✓ **Advantages**

🎯 **Better for fixed-length** tasks

📊 Task-specific optimization

🔧 Learned during training

## Relative
*T5, Transformer-XL*

```
Encodes relative
distances between
positions:

distance = pos_i - pos_j
```

✓ **Advantages**

🎯 **Captures relationships** better

📏 Distance-based encoding

🌐 Position-invariant patterns

| Sinusoidal | Learned | Relative |
|---|---|---|
| ✓ Any length | ✓ Task-optimized | ✓ Best relationships |