Encoder-Decoder with Attention

h1 h2 h3

Attention Mechanism

Enc 1 → **Enc 2** → **Enc 3**

Dec 1 **Dec 2** **Dec 3**

Input Sequence

Output Sequence

Transformer Processing Pipeline

Input Embedding — **Positional Encoding** — **Multi-Head Attention** — **Feed Forward** — **Output Projection**

[B, L, D]   *[B, L, D]*   *[B, L, D]*   *[B, L, D]*   *[B, L, V]*
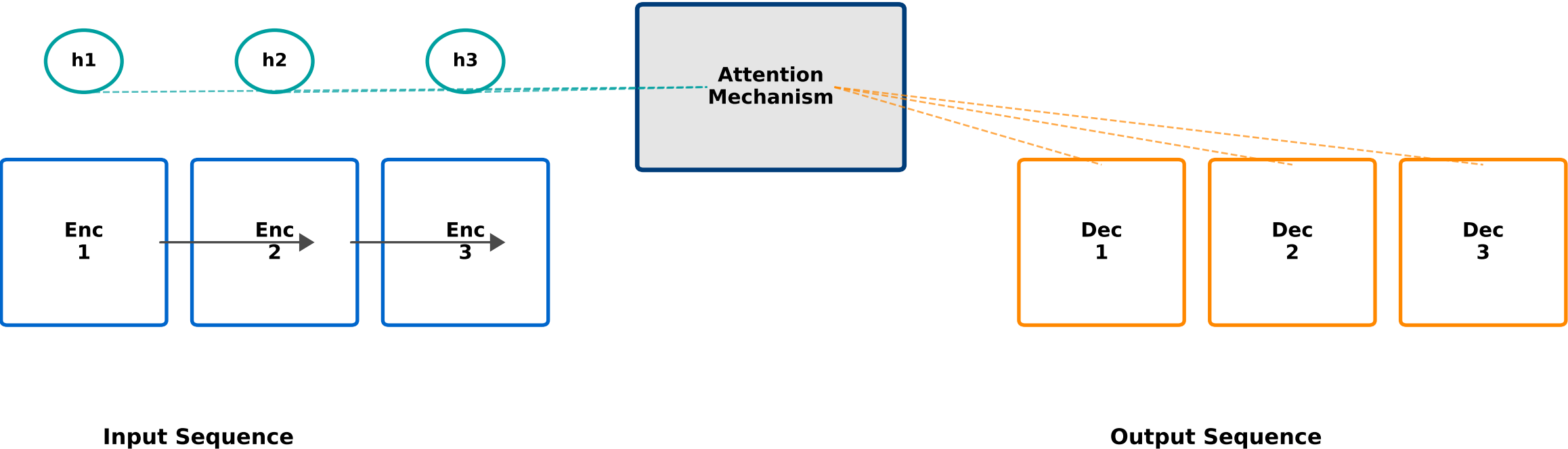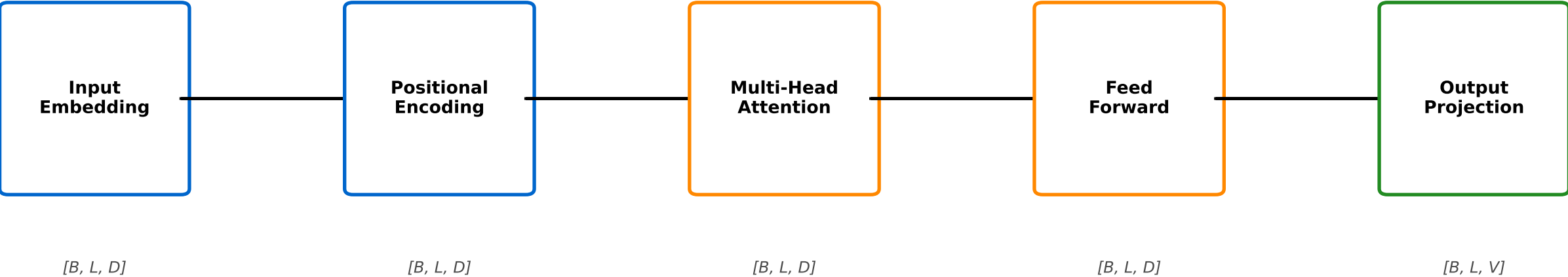
B: Batch Size | L: Sequence Length | D: Model Dimension | V: Vocabulary Size