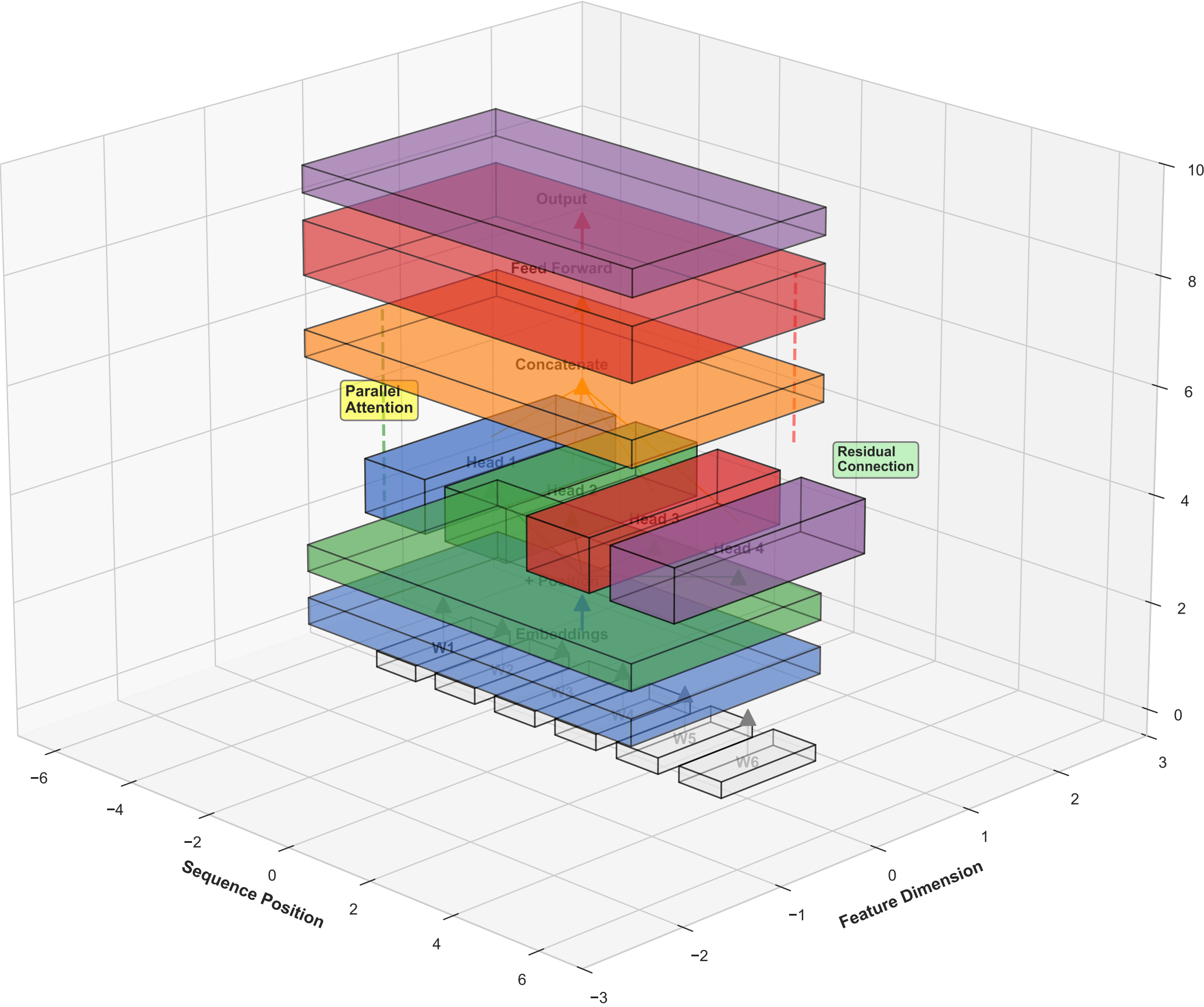


Complete Transformer Architecture in 3D
All Processing Happens in Parallel!

- Embedding Layer
- Positional Encoding
- Multi-Head Attention
- Feed-Forward Network
- Output Layer



Information flows upward through parallel layers - no sequential bottleneck!