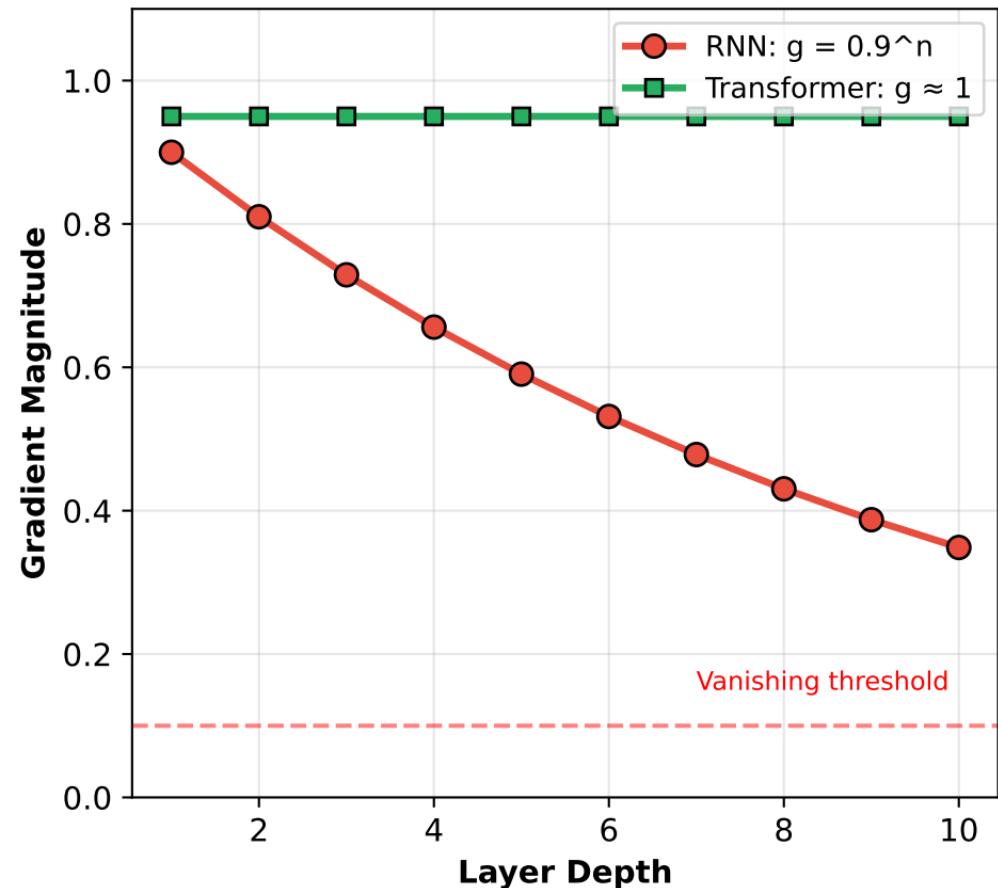


# Gradient Flow: Why Transformers Train Better

## Gradient Flow Through Layers



## Signal Strength Visualization

