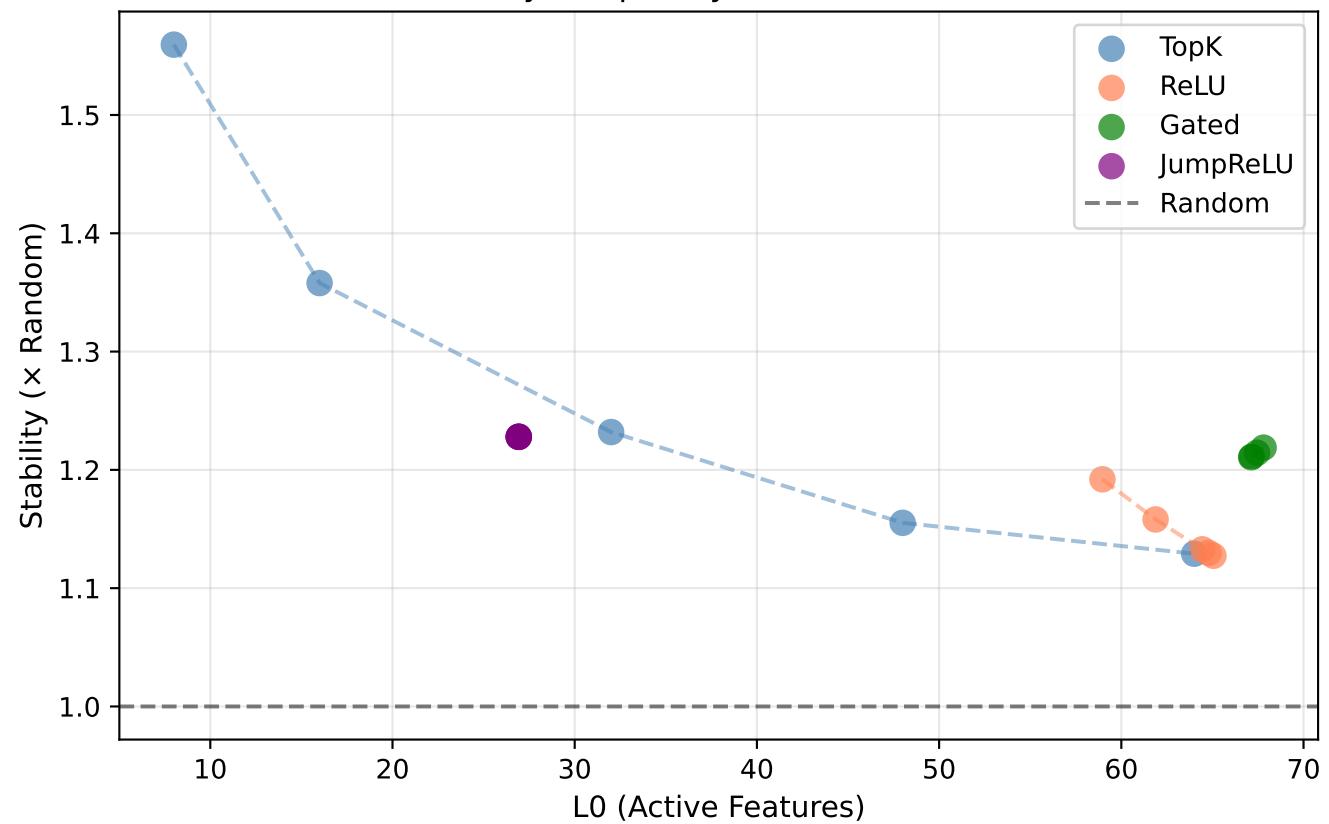
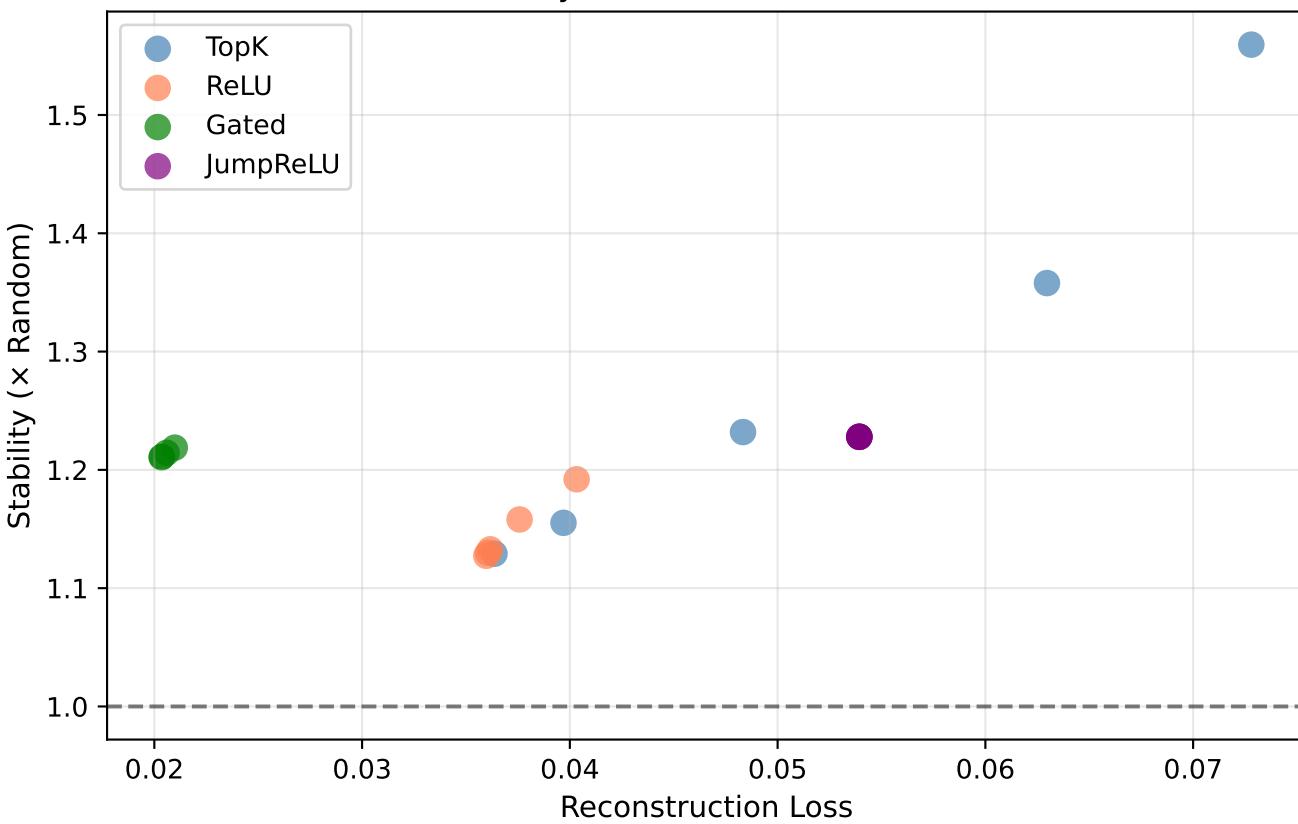


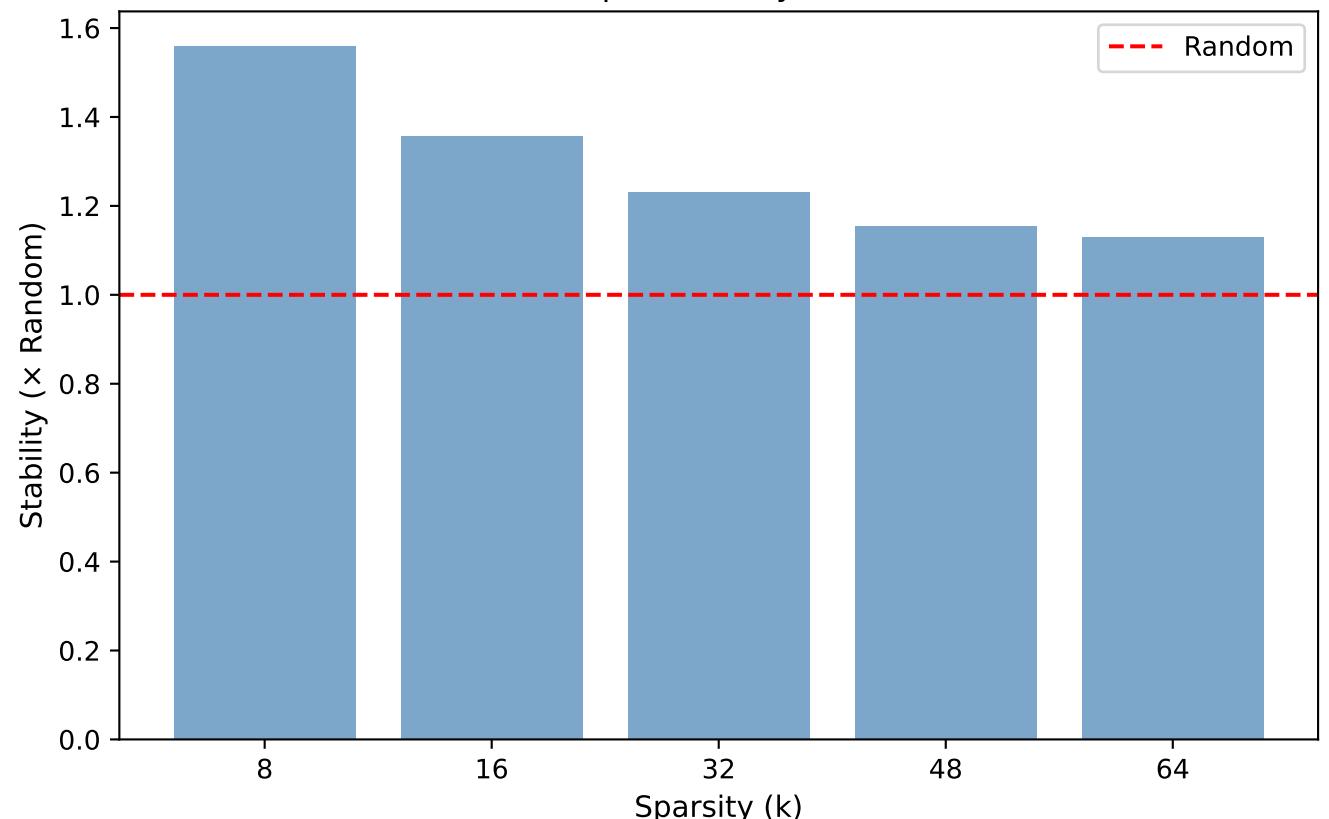
Stability vs Sparsity (All Architectures)



Stability-Reconstruction Tradeoff



TopK: Stability vs k

**VERIFICATION SUMMARY**

**TopK:**  
L0 range: 8.0 - 64.0  
Stability range: 1.13x - 1.56x  
Correlation(L0, Stability): -0.917  
→ CONFIRMS: Higher L0 = Lower Stability

**ReLU:**  
L0 range: 59.0 - 65.1  
Stability range: 1.13x - 1.19x  
Correlation(L0, Stability): -0.999  
→ CONFIRMS: Higher L0 = Lower Stability

**Gated:**  
L0 range: 67.1 - 67.8  
Stability range: 1.21x - 1.22x  
Correlation(L0, Stability): 1.000  
→ CONTRADICTS: Higher L0 = Higher Stability

**JumpReLU:**  
L0 range: 26.9 - 26.9  
Stability range: 1.23x - 1.23x  
Correlation(L0, Stability): nan  
→ WEAK relationship

**CONCLUSION:**

Overall correlation: -0.725  
✓ VERIFIED: Stability DECREASES with L0 across ALL architectures tested