Architecture Comparison: Validation Loss SDT MoD 6 × 10⁰ STT Validation Loss (Log Scale) 0.6 0.2 0.4 8.0 0.0 1.0 **Proportion of Training**