Teacher Model: Grokking Dynamics (Modular Division, p=97) 100 80 Accuracy (%) Grokking ~Epoch 70 20 Training Accuracy Validation Accuracy **Grokking Transition** 0 20 40 60 80 100 120 140 **Epoch**