GNN GPU Optimization: Speedup Contributions

Each optimization phase's marginal speedup factor

Speedup Factor

Label	Speedup	Time
Baseline (CPU, sequential)	1.0x	33.0 min
Batch-parallel forward	12.0x	2.8 min
Mixed precision (FP16)	2.5x	1.1 min
Batch size 128	1.4x	0.8 min
Async GPU transfers	1.1x	0.7 min
9x more data (2367 games)	1.8x	0.4 min