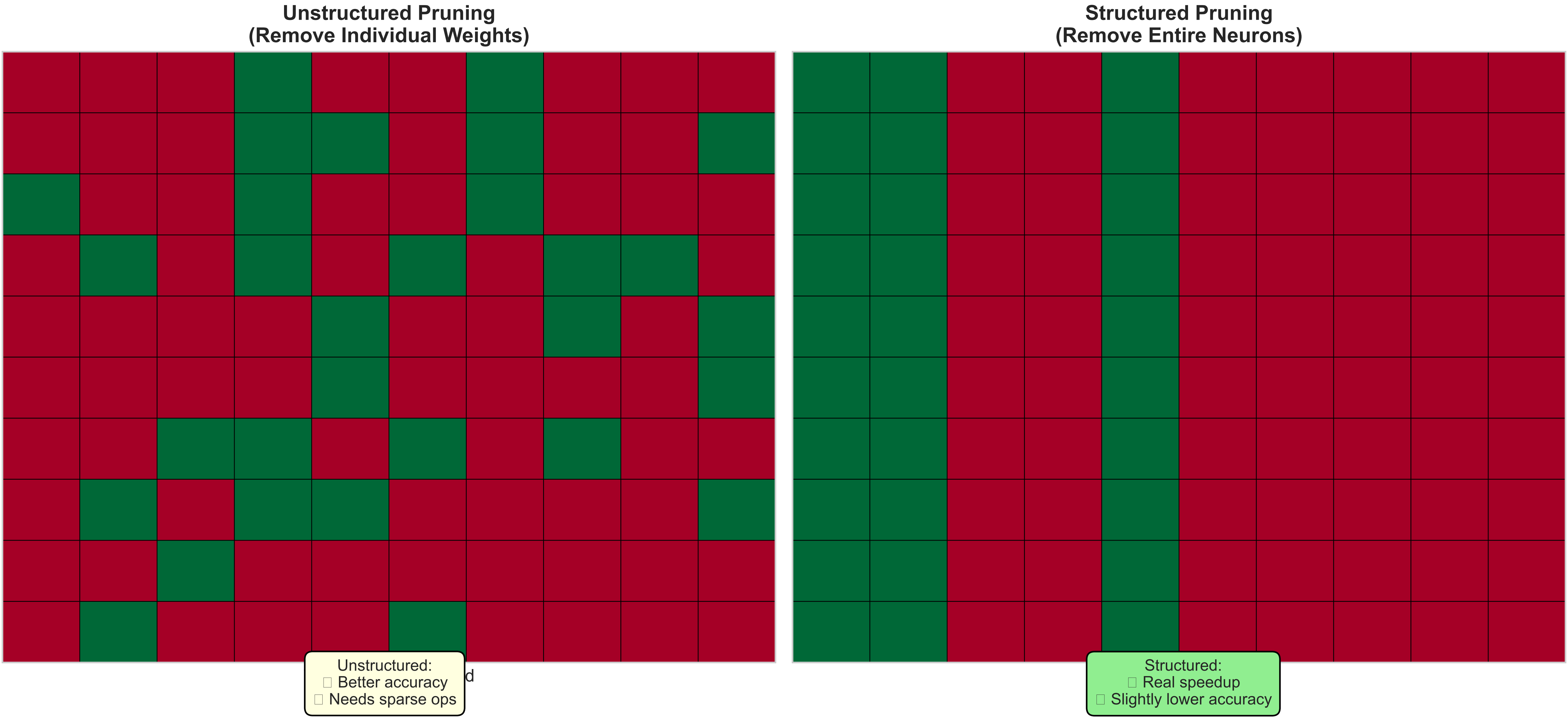
Pruning Strategies: Random vs Structured
Both reduce parameters, but structured is hardware-friendly



Unstructured Pruning
(Remove Individual Weights)

🟥	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥
🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟥	🟥	🟩
🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥
🟥	🟩	🟥	🟩	🟥	🟩	🟥	🟩	🟩	🟥
🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟩
🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩
🟥	🟥	🟩	🟩	🟥	🟩	🟥	🟩	🟥	🟥
🟥	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟩
🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥
🟥	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥

Unstructured:
□ Better accuracy
□ Needs sparse ops

Structured Pruning
(Remove Entire Neurons)

🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥
🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥

Structured:
□ Real speedup
□ Slightly lower accuracy