Estimated Multi-GPU Scaling Efficiency (Based on Single-GPU Performance Data)

Estimated scaling based on single-GPU performance and typical neural network communication patterns

100±0%

96±3%

90±5%

81±8%

Methodology: Based on actual hpfracc performance data with realistic communication overhead modeling

- hpfracc Multi-GPU Efficiency (Estimated)
- Ideal Linear Scaling
- Efficiency Loss

Number of GPUs

Efficiency (%)