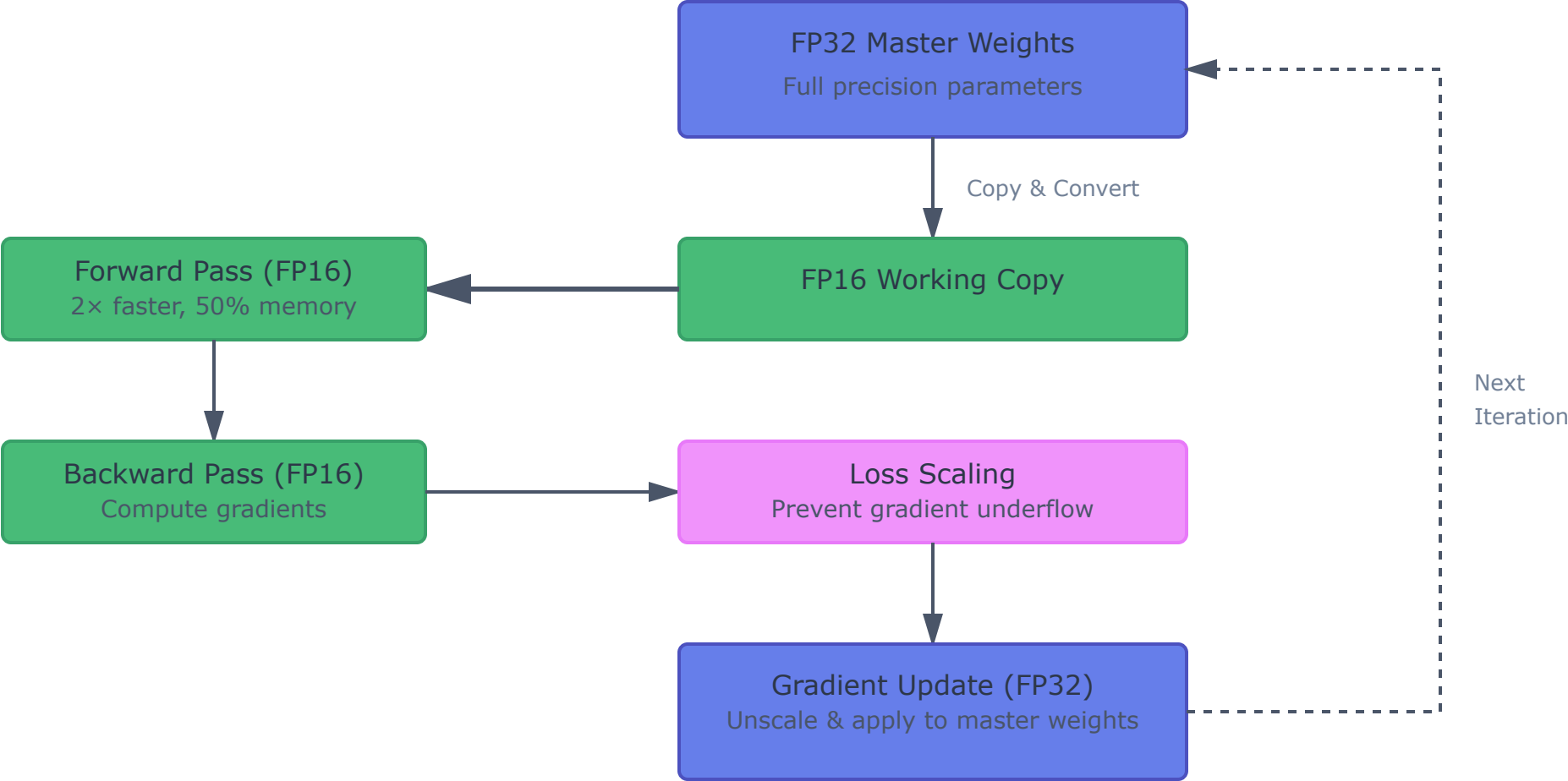


Mixed Precision Training Data Flow



Benefits

- 1.5-2× training speedup on modern GPUs
- 50% memory reduction enables larger models/batches
- Minimal accuracy impact (<0.1% typically)

Requirements

- Modern GPU with Tensor Cores (V100, A100, H100)
- Loss scaling to prevent gradient underflow in FP16 range
- FP32 master weights for numerical stability