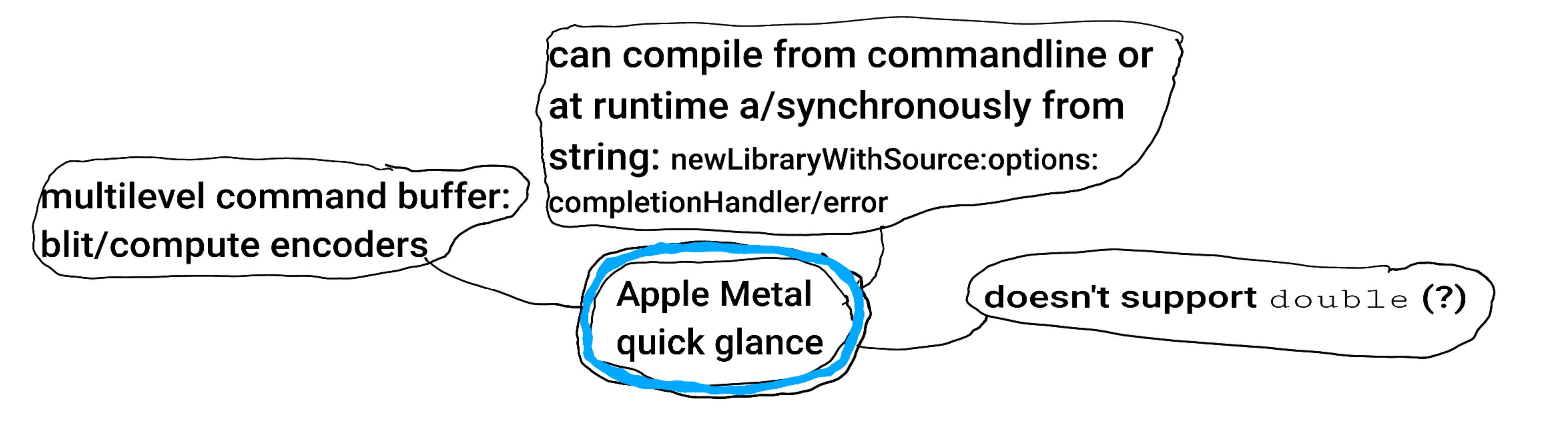
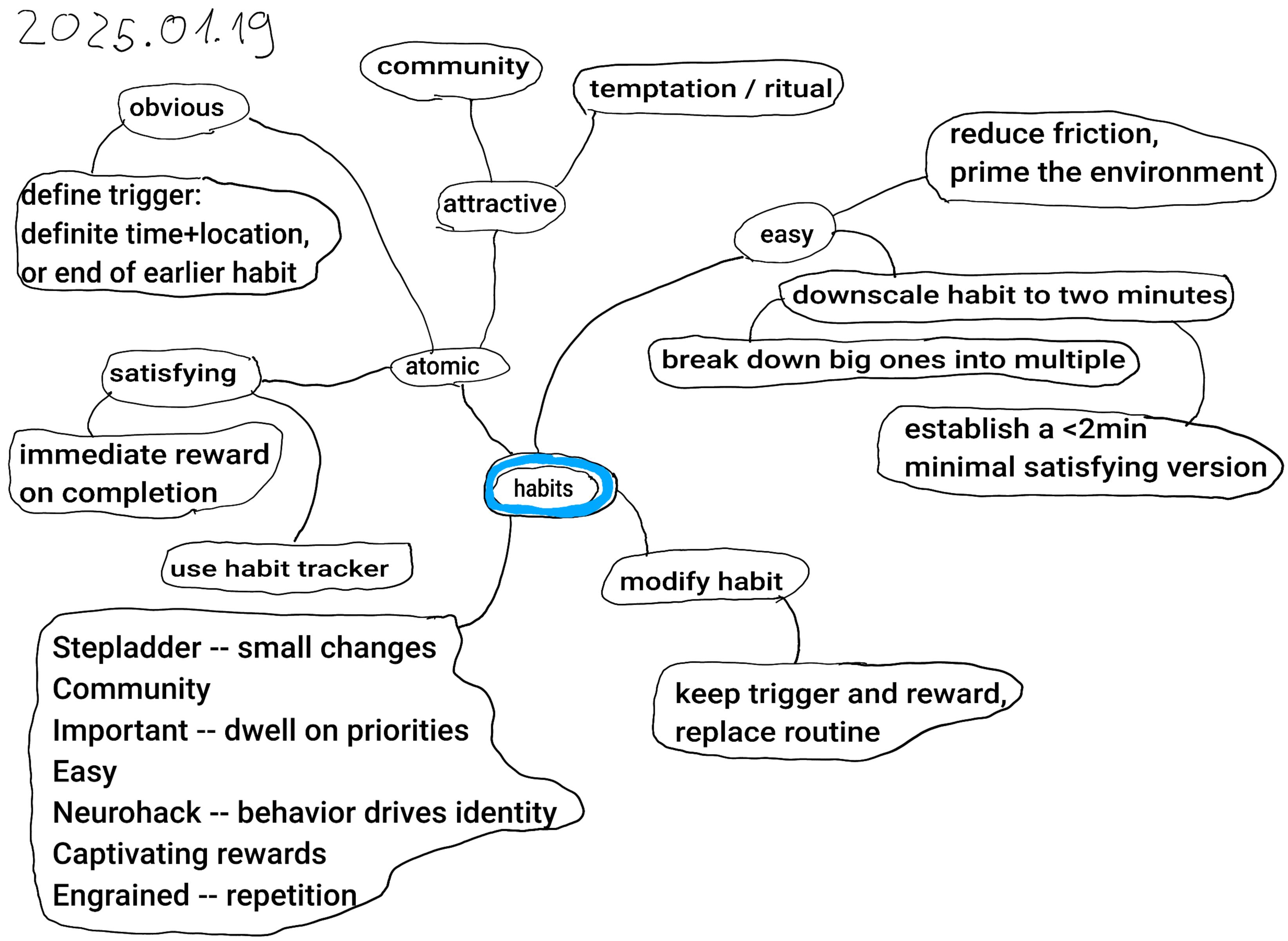
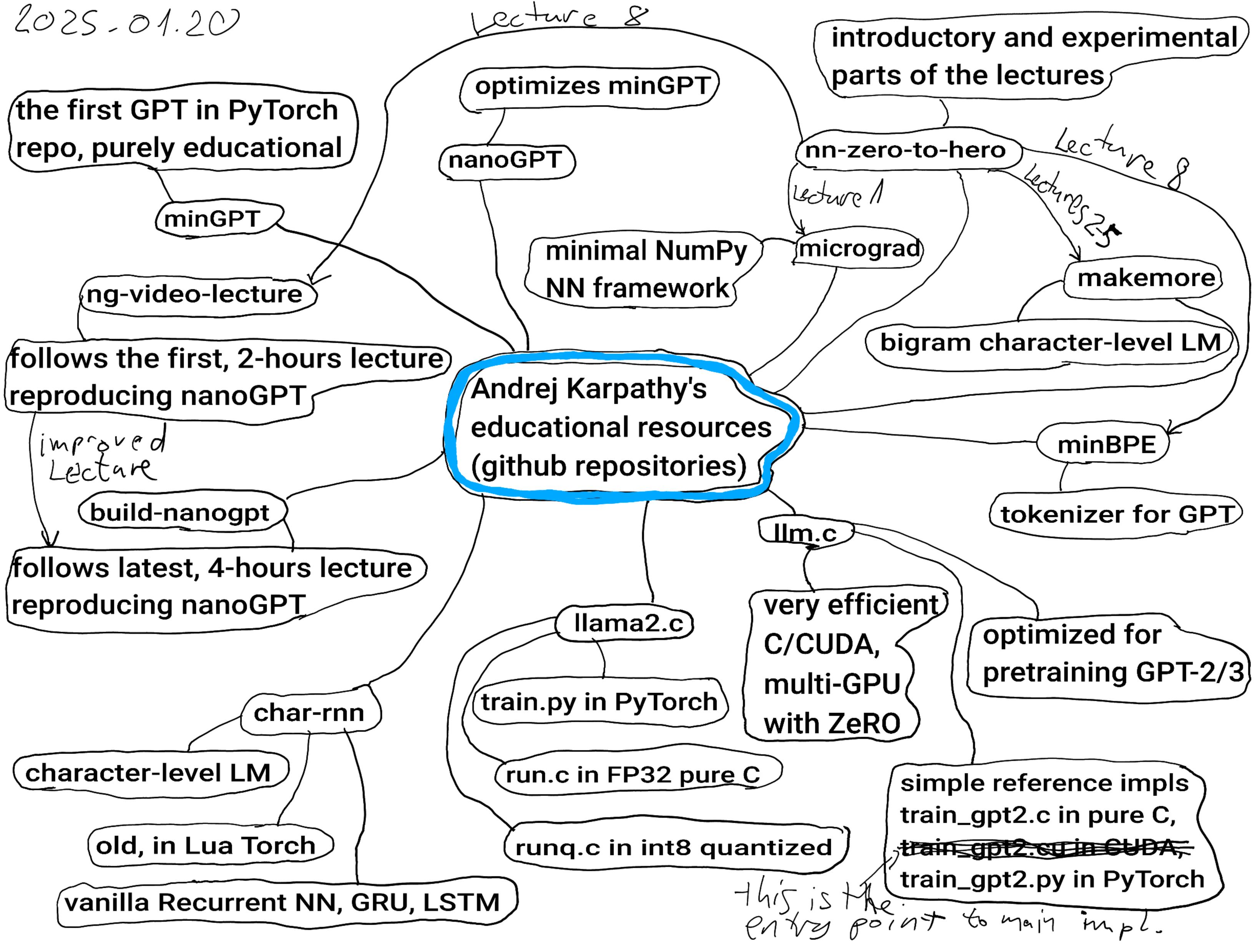


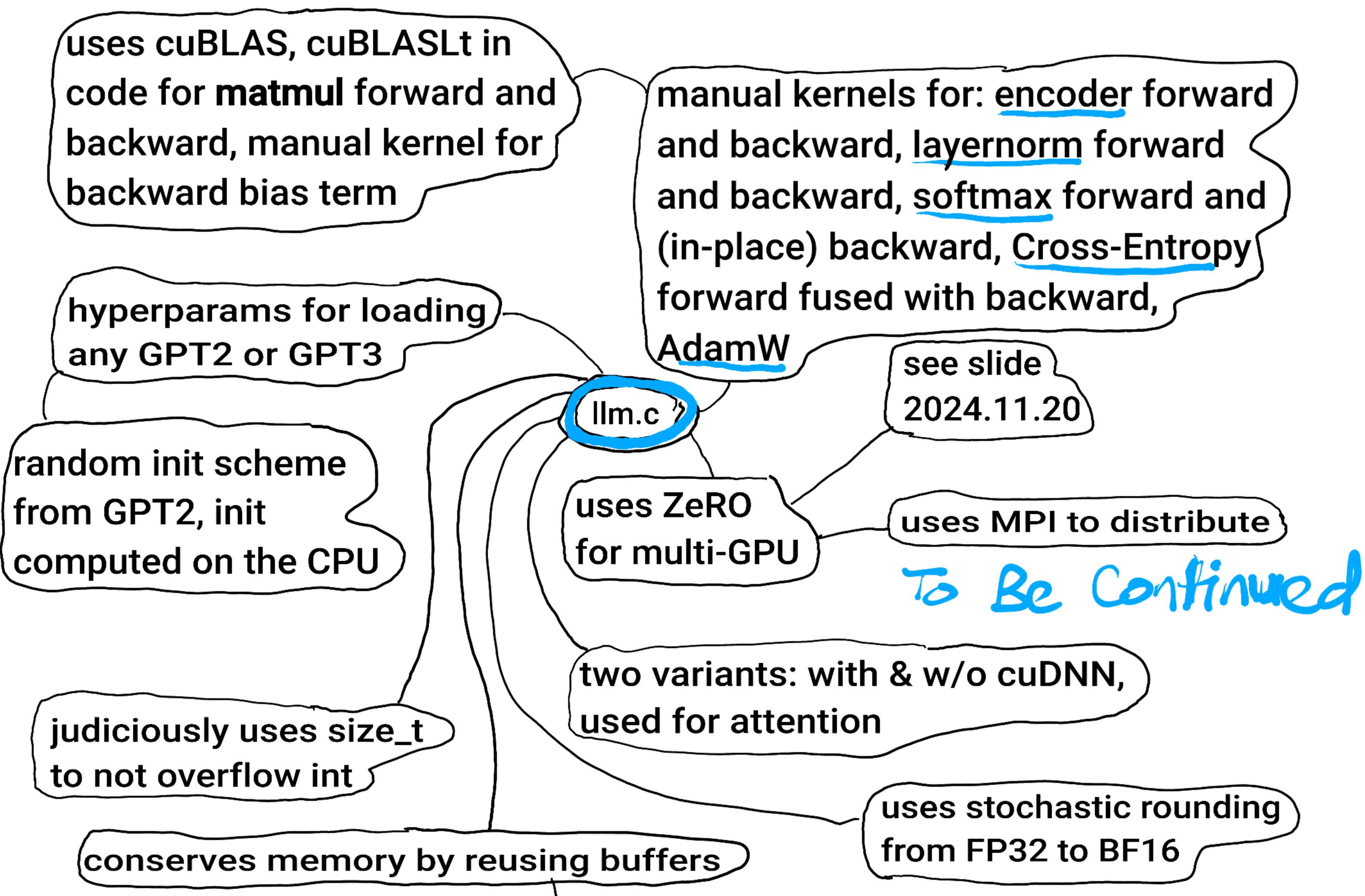
2025.01.15







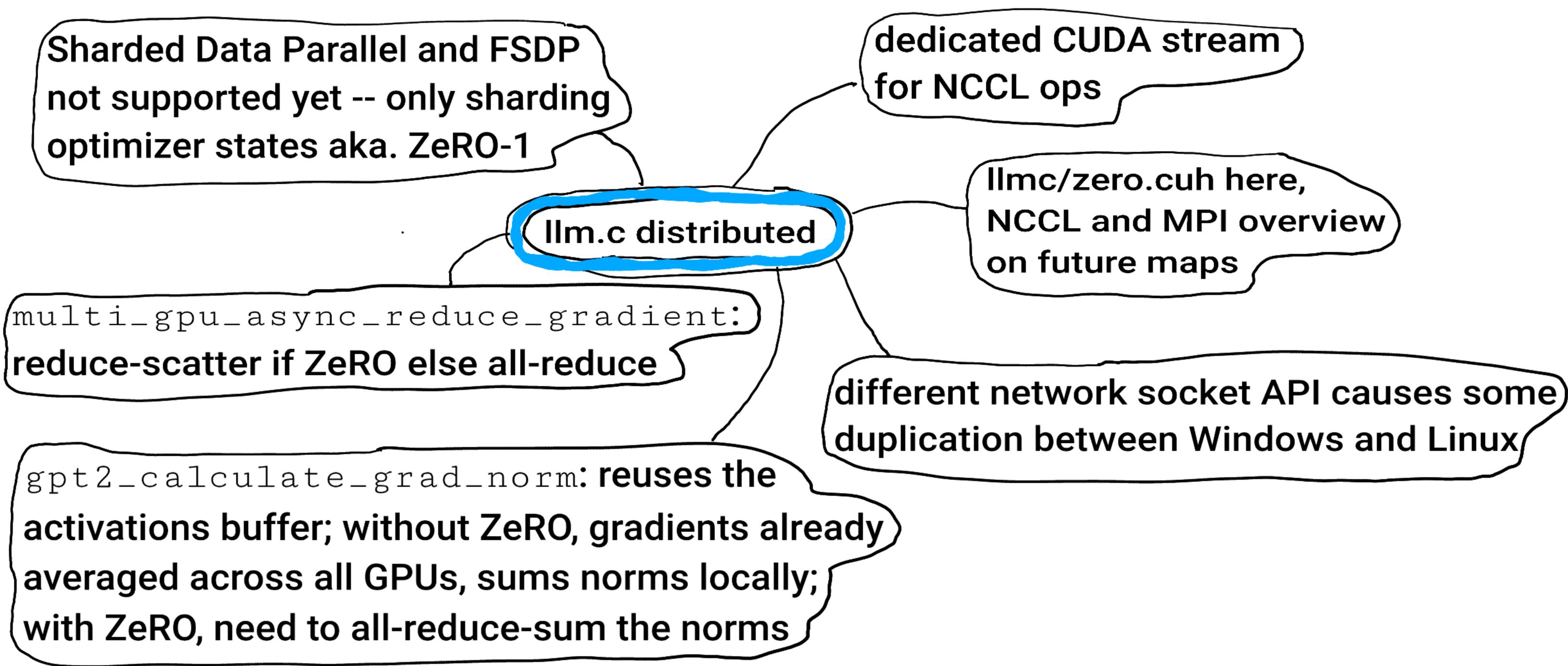
2025.01.21

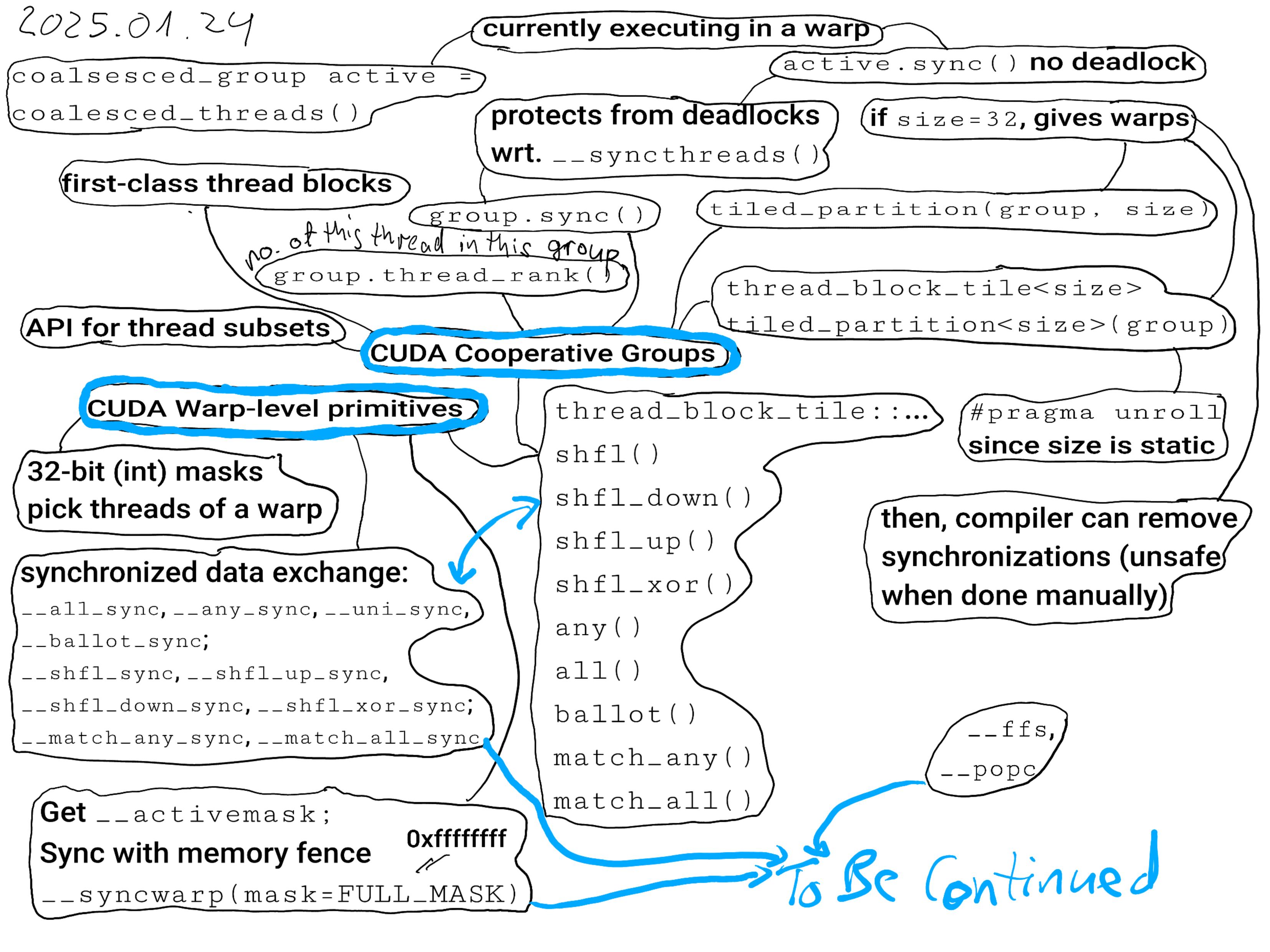


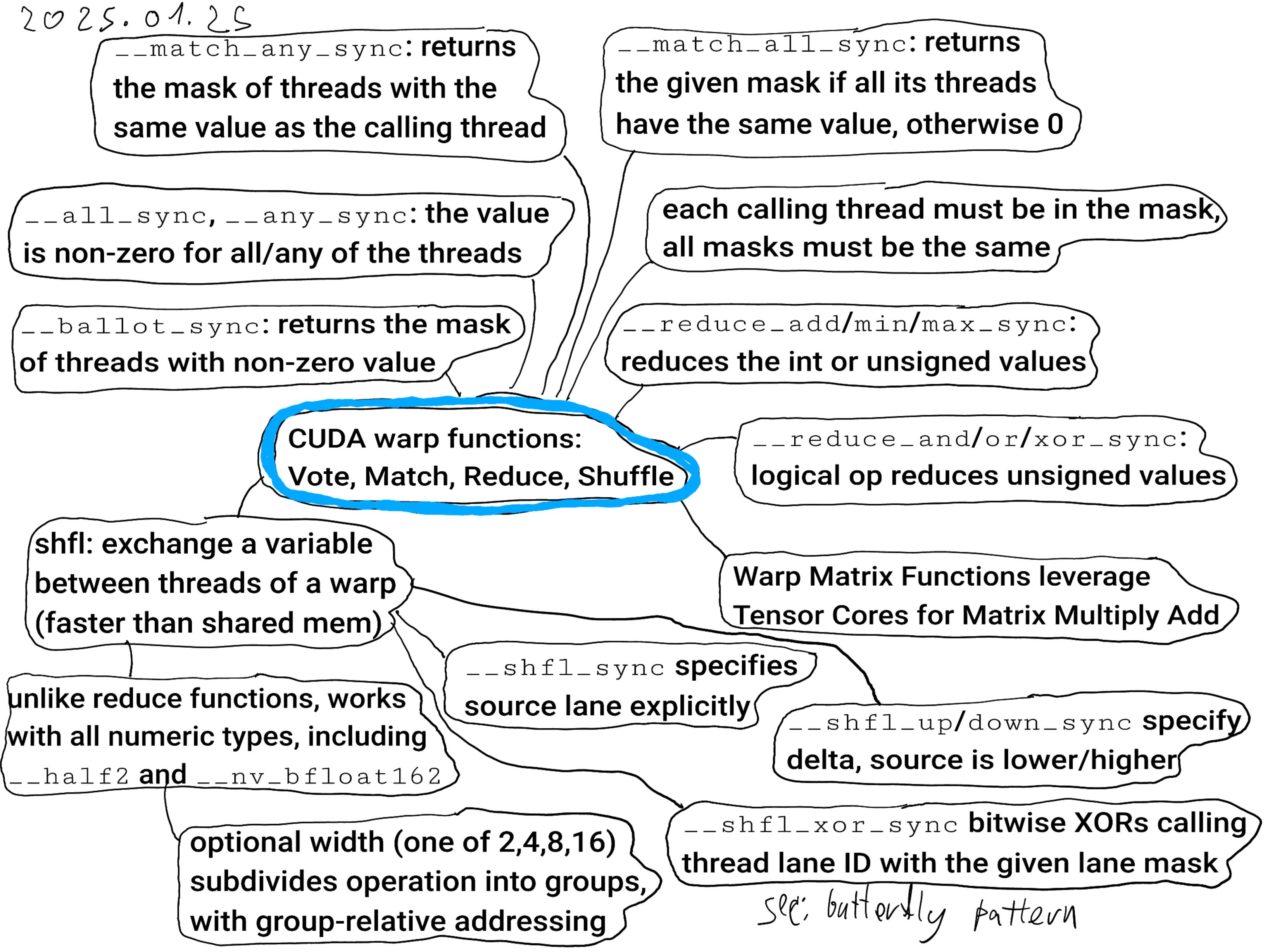
manually manages memory (no per-tensor alloc/free)

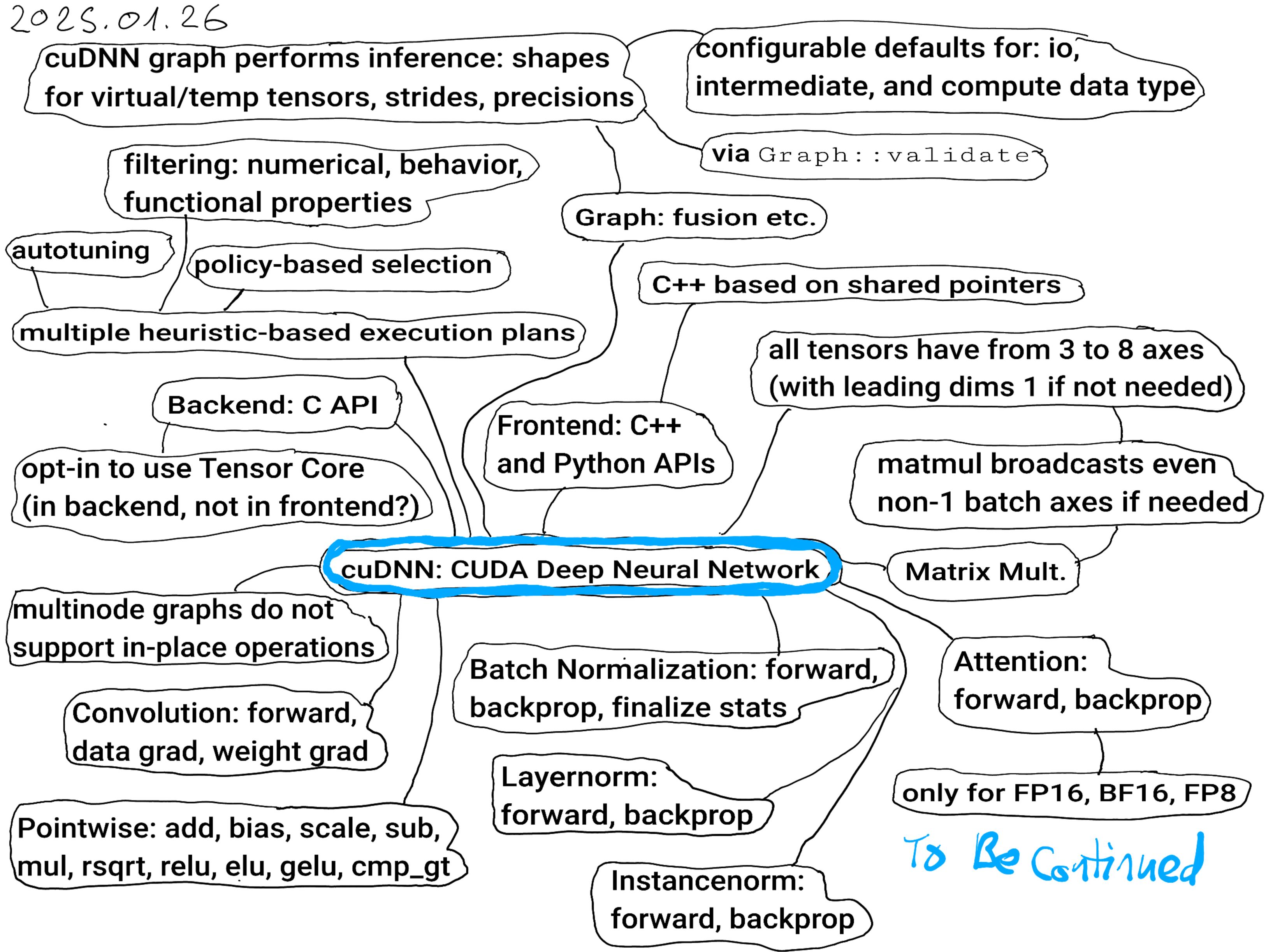
2025.01.23

optimizer states aka. ZeRO-1



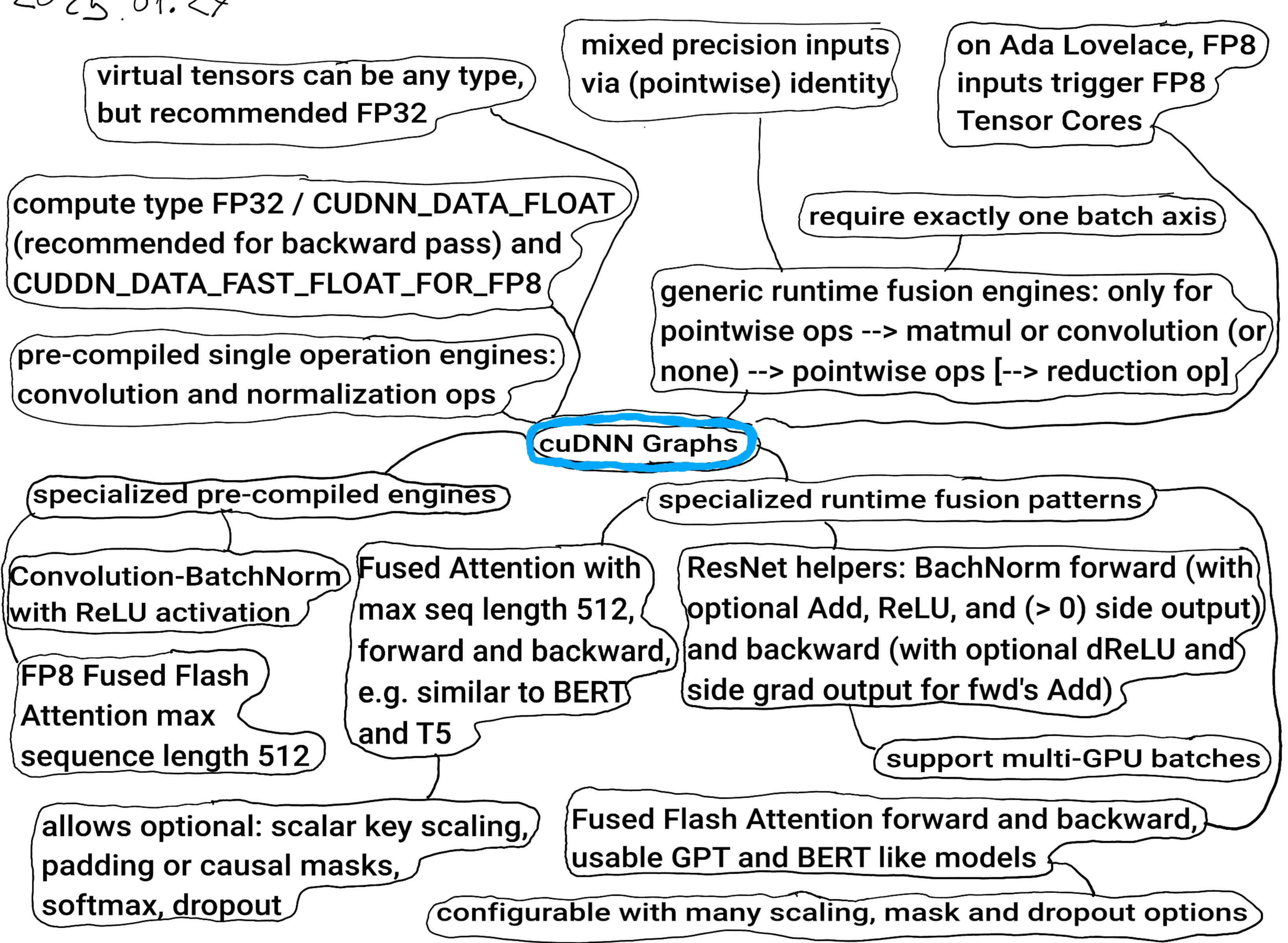


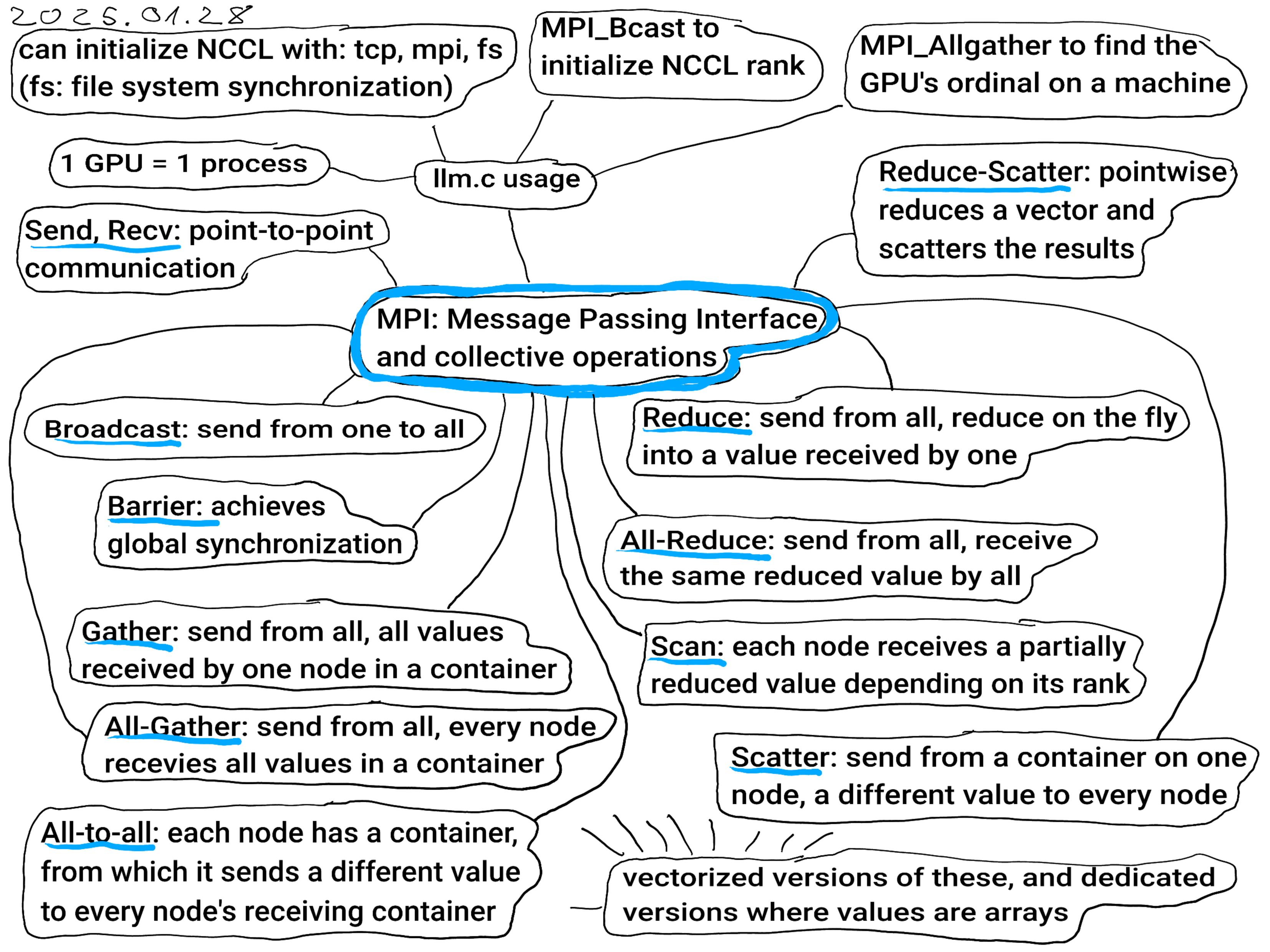


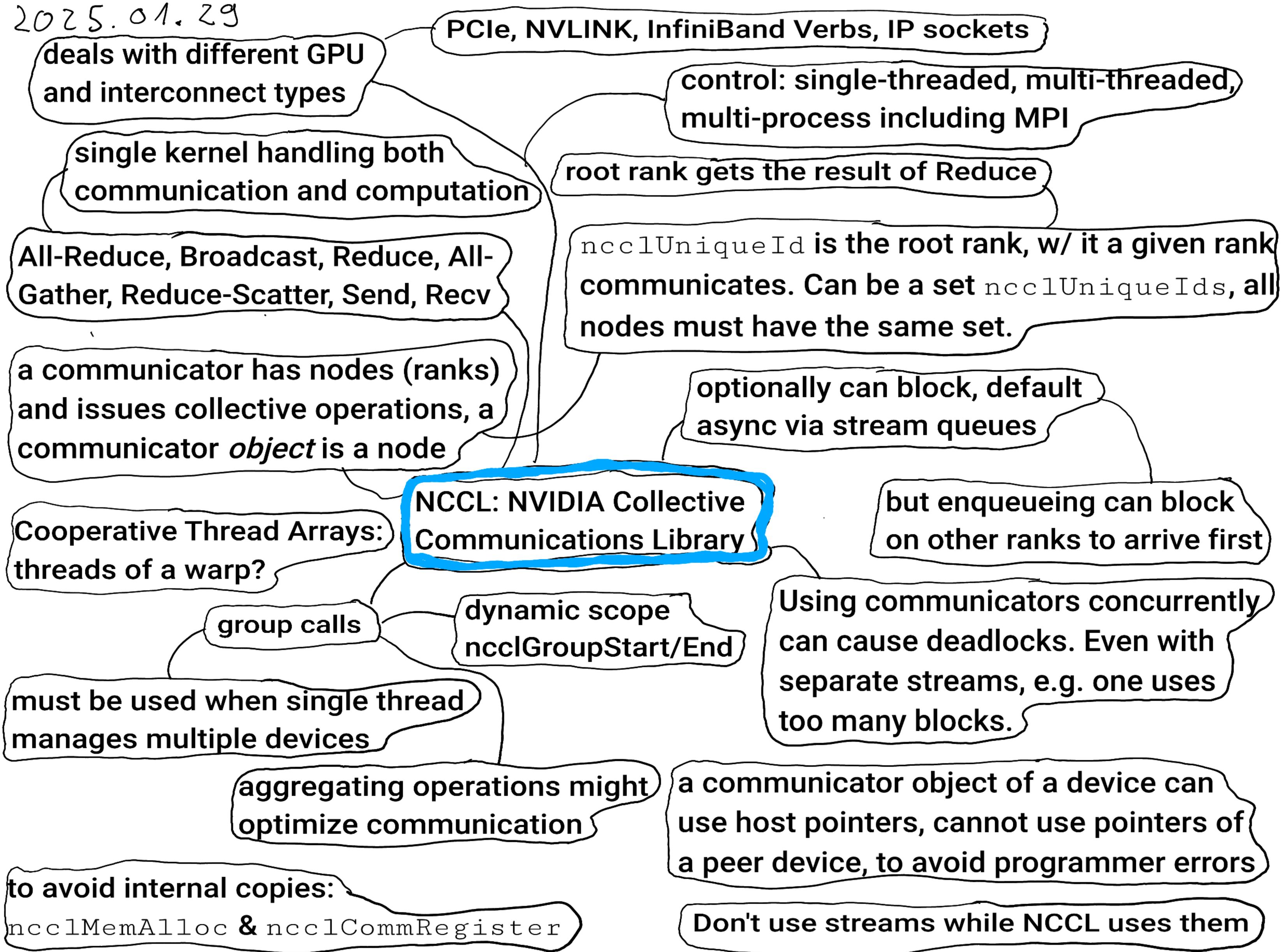


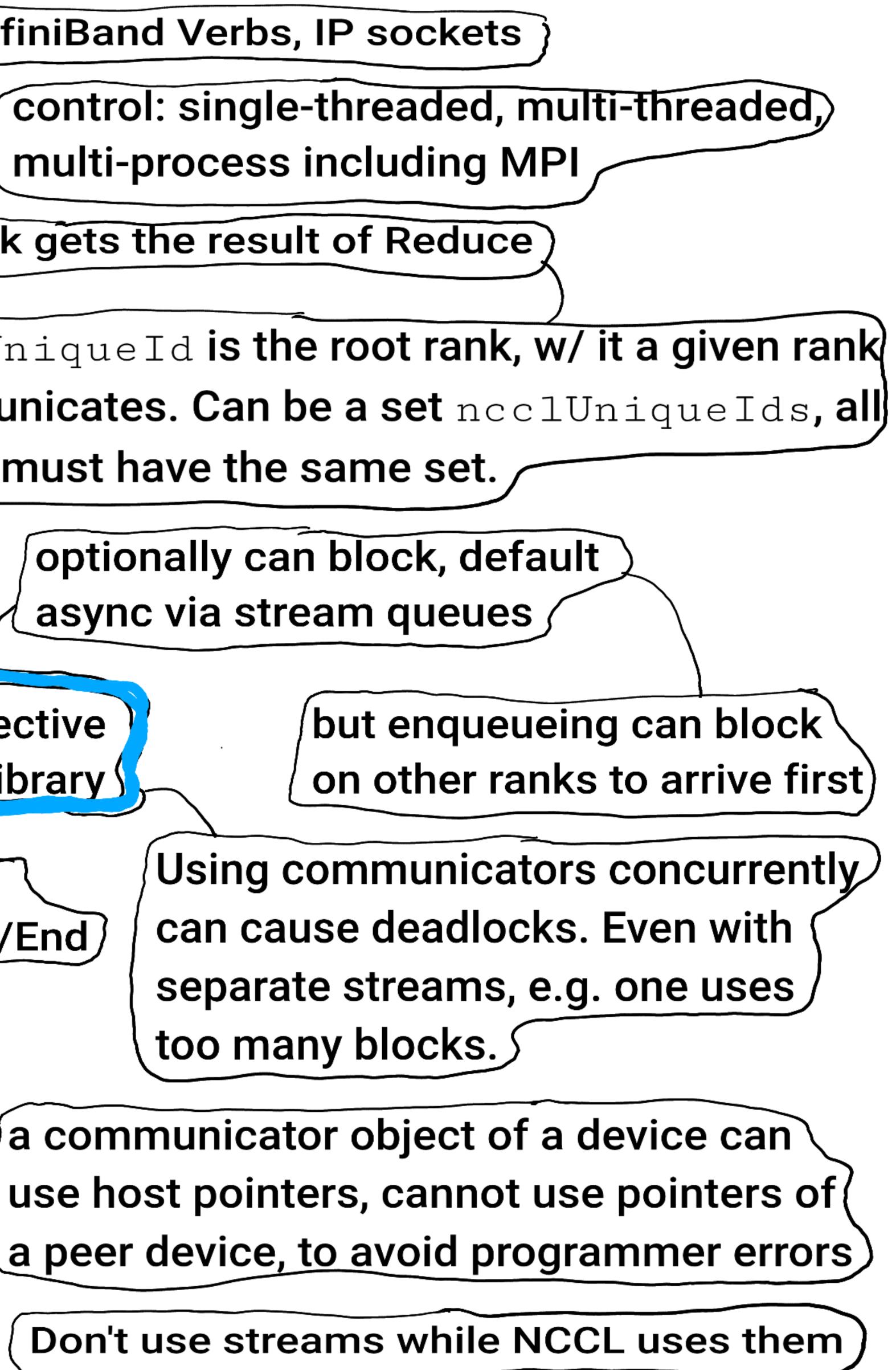
2025 01.27

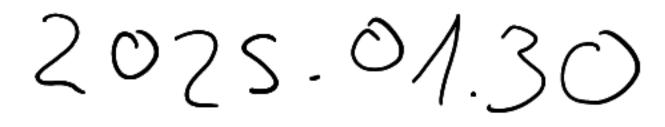
but recommended FP32

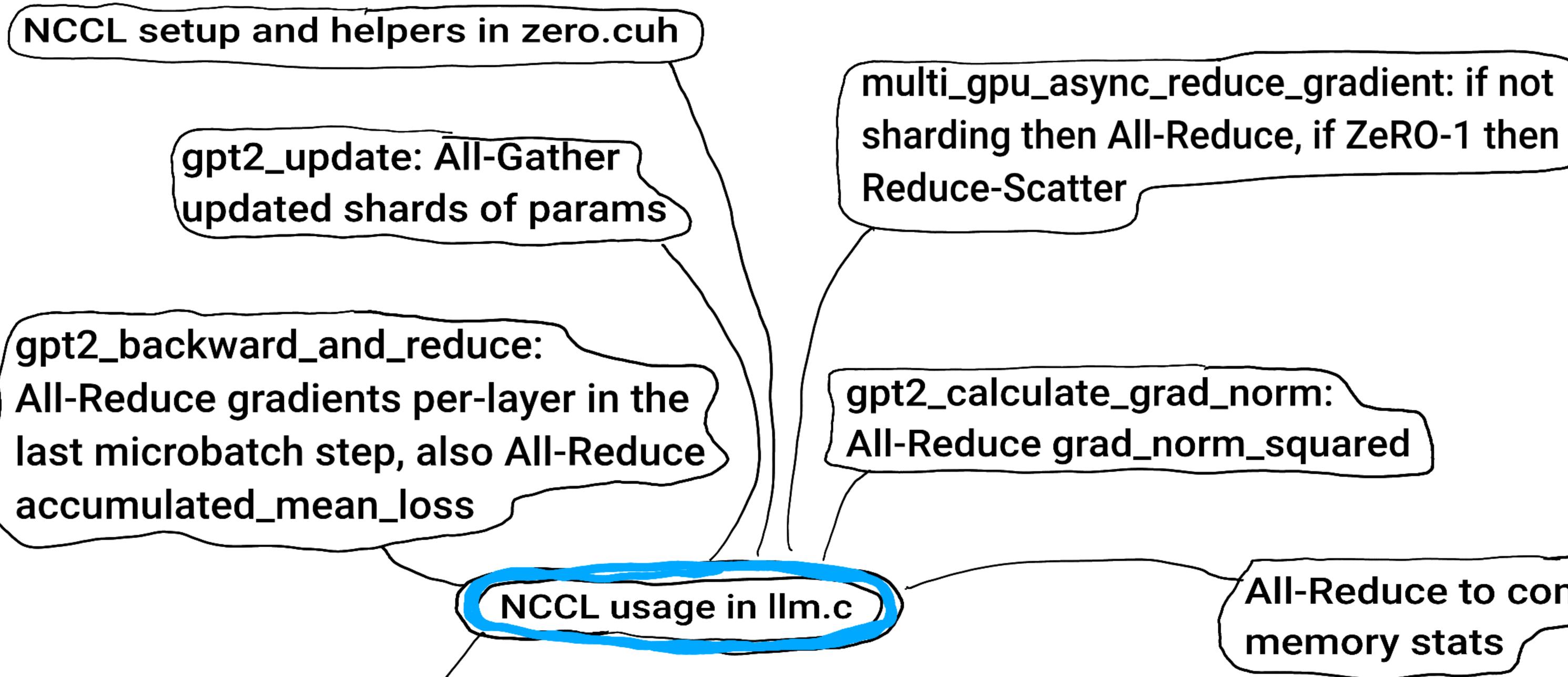














main: Periodically compute and All-Reduce: validation loss, HellaSwag accuracy



All-Reduce to compute memory stats

