

1. Compile-time

1.1. Generate binary CPU code

1.2. Generate LLVM IR code

1.3. Branch loops into separate functions in LLVM IR

1.4. Embed LLVM IR for loops into object file

2. Link-time

2.1. Load LLVM IR form objects

2.2. Extract main entry into separate LLVM IR module

2.3. Resolve (link) dependencies in LLVM IR code

2.4. Embed LLVM IR modules for loops and main entry into executable binary

3. Run-time

3.1. Load LLVM IR from binary

3.2. Load external LLVM IR for math and workflow control

3.3. Optimize, codegen & launch GPU kernel for main entry, from LLVM IR

3.4. Analyze, optimize, codegen & launch GPU kernels for GPU-efficient loops, from LLVM IR

3.5. Handle CPU host calls