

# sketch2sky

What I Cannot Create, I Do Not Understand —Richard Feynman And I



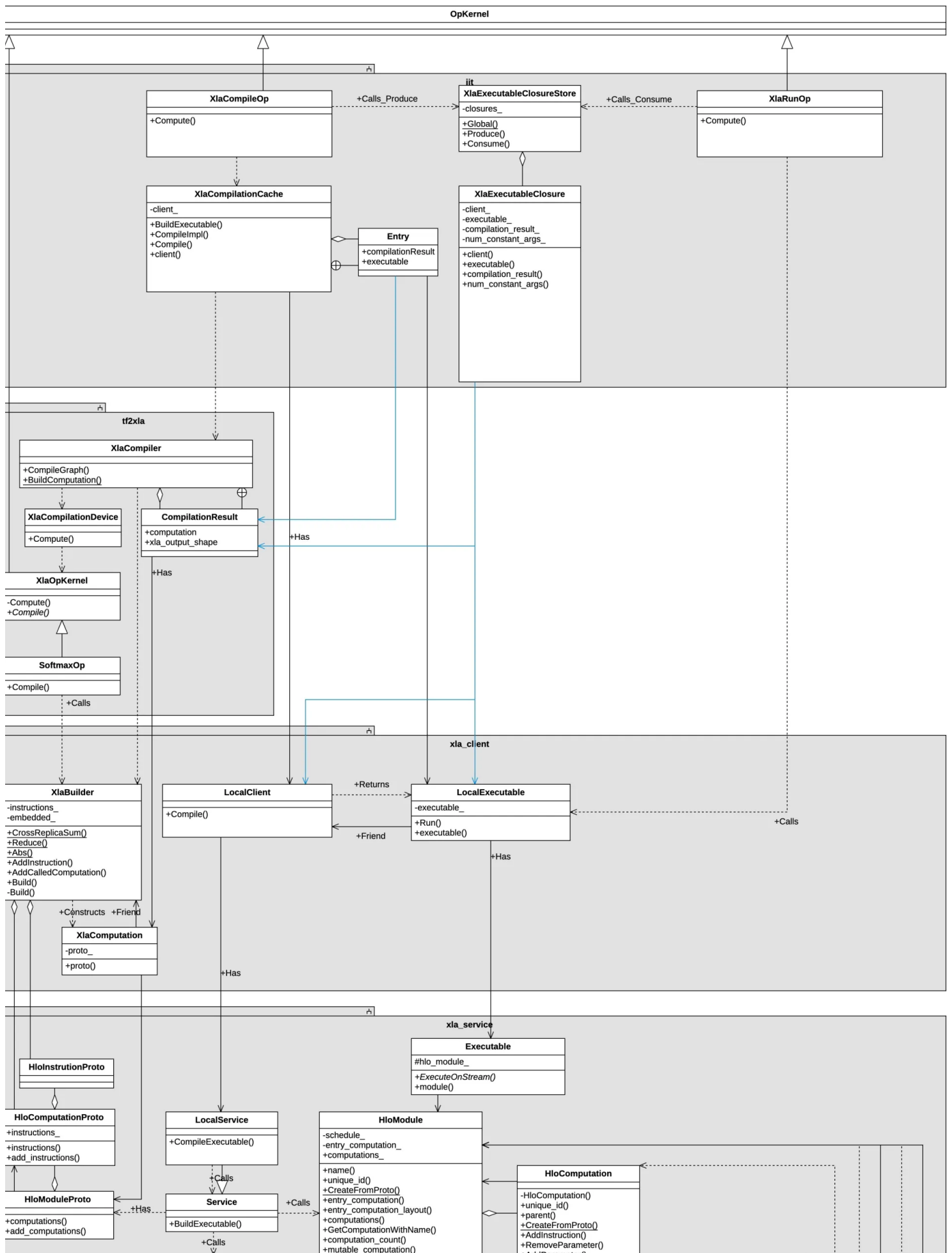
≡ Primary Menu

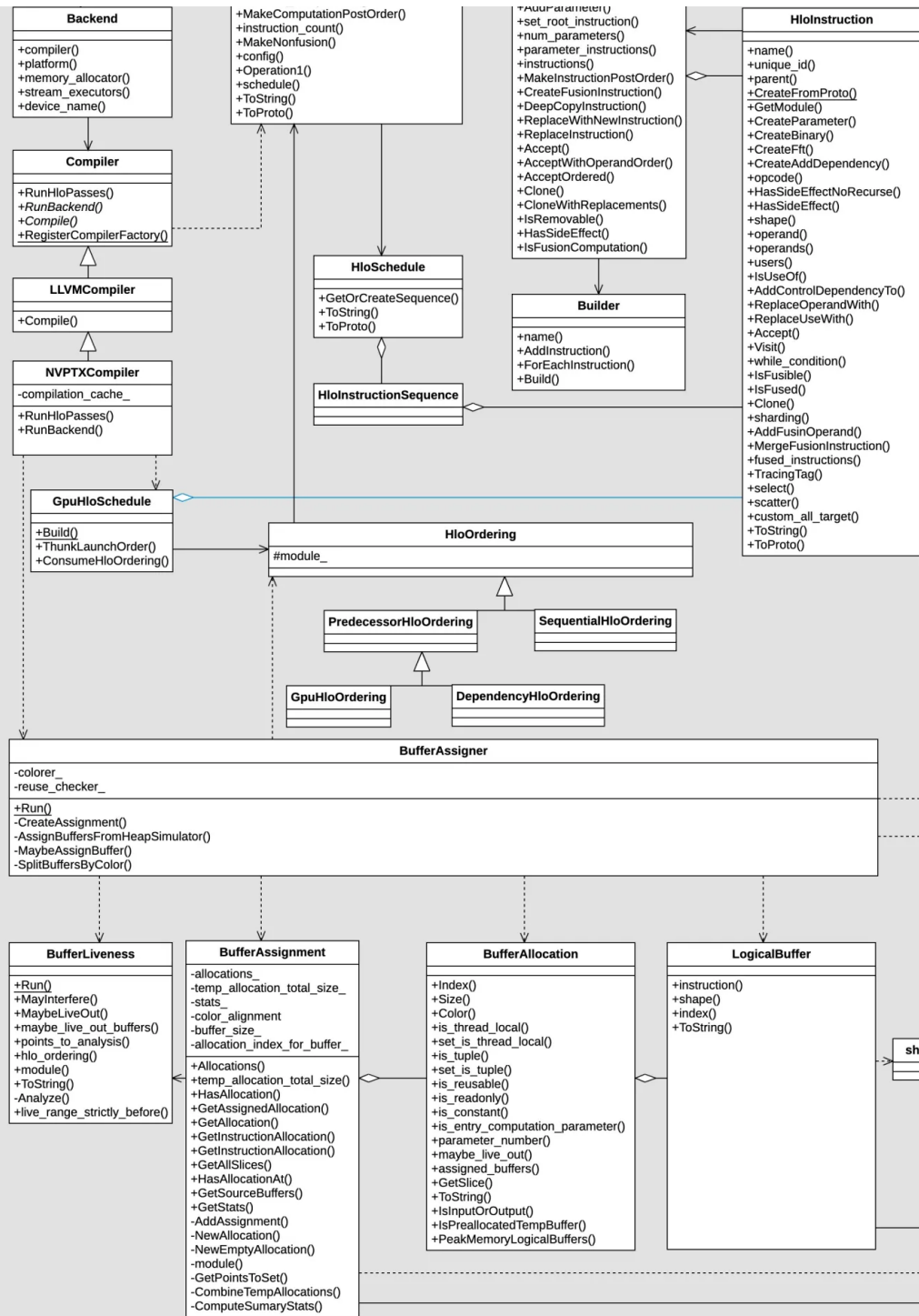
## Tensorflow JIT/XLA UML

🔗 1710 👤 [Jiang XIAO](#)

📅 2019年9月22日 at pm9:00 (last edited 📅 2020年6月20日 at am9:31)

compiler/aot/	以AOT的方式将tf2xla/接入TF引擎
compiler/jit/	以JIT的方式将tf2xla/接入TF引擎，核心是9个优化器和3个tfop，其中XlaCompileOp调用tf2xla的“编译”入口完成功能封装，XlaRunOp调用xla/client完成“运行”功能。
compiler/tf2xla/	对上提供xla_compiler.cc:XlaCompiler::CompileFunction()供jit:compile_fn()使用将cluster转化为XlaComputation。核心是利用xla/client提供的接口，实现对XlaOpKernel的“Symbolic Execution”功能。每个XlaOpKernel子类均做的以下工作: **从XlaOpKernelContext中取出XlaExpression或XlaOp，调用xla/client/xla_buidler.h提供的方法完成计算，将计算结果的XlaOp存入XlaKernelContext.**
compiler/xla/client/	对上提供xla_builder.cc:Builder等供CompileFunction()使用，将Graph由Op表达转化为HloModuleProto:HloComputationProto:HloInstructionProto表达并保存在XlaComputation中。 对上提供local_client.cc:LocalClient::Compile()，作为编译入口供jit: BuildExecutable()使用，将已经得到的XlaComputation交给service并进一步编译为二进制。 对上提供local_client.cc:LocalExecutable::Run()，作为运行入口供jit/kernels/xla_ops.cc:XlaRunOp使用，通过Key找到相应的二进制交给service层处理
compiler/xla/service/	对上提供local_service.cc:LocalService::BuildExecutable()供LocalClient::Compile()使用实现真正的编译，承接XlaComputation封装的HloProto，将其转化为HloModule:HloComputation:HloInstruction表达，对其进行优化之后，使用LLVM后端将其编译为相应Executable后端的二进制代码 对上提供executable.cc:Executable::ExecuteOnStream()供LocalExecutable::Run()使用实现真正的执行二进制。





## Related:

[Tensorflow XLA Service Buffer优化详解](#)

[Tensorflow XLA Service 详解 II](#)

[Tensorflow XLA Service 详解 I](#)

[Tensorflow XLA Client | HloModuleProto 详解](#)

[Tensorflow XlaOpKernel | tf2xla 机制详解](#)

[Tensorflow JIT 技术详解](#)

[Tensorflow JIT/XLA UML](#)