

基于RVV-0.8 SPEC的VPU的设计 与集成

梁展豪

深圳优矽科技有限公司

概述

- 基于RISCV Vector扩展指令集[0.8-release](#)版本
- 使用Chisel设计电路
- 集成到rocket-chip上，commit [26a18fa6](#)
- 具有一定程度的可配置性
- 已开源到Github，仓库名[rocket_chip_vpu](#)

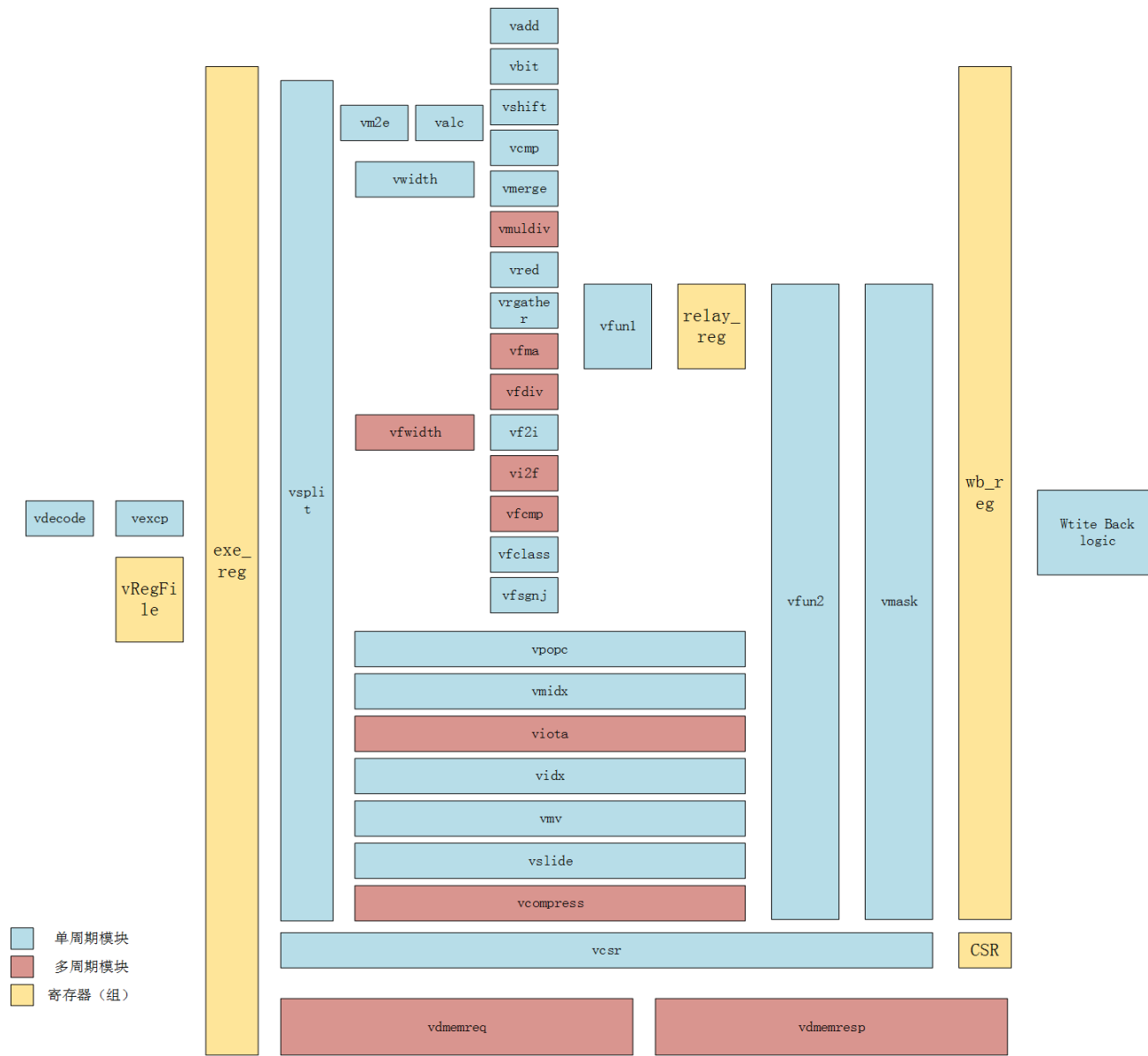
VPU电路设计思路

- 把需要计算的8VLEN数据拆分成不同的位宽，不同深度的多个向量。
- 在模块内部实现各种位宽的元素的操作电路。
- 对于非元素位置操作（即对应索引的元素相互运算），使用ELEN长度划分向量长度，多次运算。
- 对于结果写回周期有要求的指令和对元素位置的操作，取全长度8VLEN向量进行运算。
- 使用标量load/store接口，尽量不改变rocket-chip原有的DCache。

VPU模块划分

- 按功能划分
 - 准备运算的元素的模块
 - 拆分元素模块
 - 转换元素位宽模块
 - 分配多次运算的元素模块
 - 元素运算模块
 - 定点运算模块
 - 浮点运算模块
 - 掩码运算模块
 - 元素位置处理模块
 - 处理结果
 - 选择运算模块输出模块
 - 对结果进行掩盖模块
- 按周期划分——见图

VPU模块划分图



VPU代码构建

- 使用Chisel构建，方便参数化配置，以及集成到rocket-chip。
- 使用到rocket-chip的MulDiv模块代码，以及部分辅助性代码。
- 构建浮点模块时，使用到hardfloat库。

VPU代码构建

- 底层构建一个所有模块的超类：VModule
- 尽量一个文件一个模块，一个模块一个功能
- 顶层构建三个文件
 - 电路顶层模块：VPU
 - VPU对外端口定义：VPUIO
 - VPU参数定义：VPUParams

VPU参数

- 默认整型加减，按位操作，移位，整型比较，load/store指令，以及掩码指令。
- 参数分为两种
 - 规模参数，如VLEN，ELEN等。
 - 指令是否支持参数，如MERGE，MULDIV等。

VPU参数解释

- VLEN: 向量寄存器的长度
- ELEN: 最大支持的元素位宽
- SELEN: 最大支持的子元素位宽
- XLEN: 处理器字长
- FLEN: 标量浮点支持的字长
- ZFINX: 是否支持浮点标量来自GPR
- FSEW16: 是否支持浮点元素位宽低至半精度
- FSEWMAX: 最大支持的浮点元素位宽
- LMULMAX: 最大支持的LMUL分组（之后会取消这个参数）
- MERGE: 是否支持合并指令
- MULDIV: 是否支持整型乘除指令
- MULADD: 是否支持整型乘加指令
- QMULADD: 是否支持四倍扩展位宽乘加指令
- mulDiv: 可配置不同位宽的整型乘除子模块的实现方式（可配置包括mulUnroll, divUnroll, mulEarlyOut, divEarlyOut参数）
- RED: 是否支持整型归约指令
- MV: 是否支持整型标量移动指令
- SATADD: 是否支持定点饱和加减指令
- AVERADD: 是否支持定点平均加减指令
- SATMUL: 是否支持定点饱和乘指令
- SCALESR: 是否支持定点缩放右移指令
- NCLIP: 是否支持定点缩减位宽右移指令
- SLIDE: 是否支持vslideup,vslidedown,vslide1up,vslide1down指令
- GATHER: 是否支持vrgather指令
- COMPRESS: 是否支持vcompress指令
- COPY: 是否支持vmv<nf>r指令
- FMA: 是否支持浮点加减乘乘加指令
- FCVT: 是否支持浮点转换指令
- FCMP: 是否支持浮点比较指令
- FSGNJ: 是否支持浮点符号注入指令
- FCLASS: 是否支持浮点分类指令
- FMERGE: 是否支持浮点合并指令
- FRED: 是否支持浮点归约指令
- SEGLS: 是否segment load/store指令
- AMO: 是否之词amo指令
- EDIV: 是否支持拆分成子元素
- DOT: 是否支持整型dot-product指令
- FDOT: 是否支持浮点dot-product指令

VPU对外端口

- VPU与Rocket的交互接口
 - 指令输入
 - 标量数据输入
 - 标量数据输出
 - CSR的输入输出
 - 控制信号
- VPU与FPU的交互接口
 - 标量数据的输入输出
- VPU与DCache的交互接口——定义来自HellaCacheIO的定义

VPU的集成方式

- 电路集成
 - 类似FPU，Rocket的ID阶段接收指令，在MEM阶段返回整型标量。
 - 若无数据冲突，VPU可以在FPU运行的同时运行。
 - VPU访存时，暂时阻塞其他访存请求。
- 工程集成
 - 类似hardfloat，修改build.sbt文件，使得VPU依赖hardfloat，rocket-chip依赖VPU
 - 在rocket-chip顶层构建可配置类

VPU开源

- 地址：
https://github.com/UCTECHIP/rocket_chip_vpu
- 内容包含
 - 对rocket-chip [26a18fa6](#)的补丁
 - VPU的源代码
 - 简单的说明文件README.md
 - 简单的TODO描述

谢谢