

*13-project*

# Maxwell-FPGA

——国防科大2020年高性能评测与优化课程小组讨论

成员：汪忠洋 乔冠杰

指导：龚春叶、甘新标、杨博



# 目录 CONTENTS

▶ 01 需求分析

▶ 02 设计动机

▶ 03 技术方案

▶ 04 性能分析

▶ 05 未来展望



# 01 需求分析





## 需求分析

在2005年，来自苏格兰的几家公司及学院携手建立了一个FPGA高性能运算联盟(FHPCA)，主要目标是将现场可编程门阵列(FPGA)应用于高性能运算领域。这些公司公布了一个为期两年的合作计划，将设计并构建64节点基于FPGA的超级计算机，能够实现超过1TeraFLOP的处理速度。为验证该系统的能力和灵活性，该联盟从候选应用中选择三种应用，并连接到超级计算机，测试其对真实的HPC应用程序的有效性。

### 目标



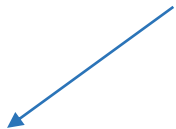
- 1.用FPGA板设计和制造64节点超级计算机。
- 2.验证其在真实HPC应用的有效性。

## 02 设计动机





Maxwell被设计为概念验证通用的FPGA超级计算机。



构建的机器能够尽可能广泛地应用，提供足够的FPGA性能，探索标准HPC应用程序是否可以几乎完全在FPGA硬件、并行通信和所有东西上运行。

FPGA是主要的计算平台，而不是CPU的协处理器。为此，构建了一台能够在大型FPGAs网络上进行并行操作的机器，这些FPGAs直接连接在一起。

# 03 技术方案





### （一）硬件：

Maxwell本质上是FPGA加速的刀片服务器集群，总共包括32个刀片服务器，每个由一个Inter Xeon（至强）处理器和两个Xilinx(赛灵思) Virtex-4 FPGA组成，CPU通过标准的IBM PCI-X扩展模块的方式连接FPGA。

#### 1. 刀片服务器机箱：

Maxwell实际上包括两个19英寸机架和五个IBM 刀片中心，刀片中心中的四个具有七个IBM Intel Xeon刀片服务器，第五个具有四个刀片服务器。每个刀片都是具有2.8 GHz Intel Xeon处理器，同时有1 GB主内存。刀片通过单个48路Netgear交换机通过千兆以太网连接，吞吐量为40 Gb / s。刀片服务器是从头节点，具有4 GB主内存和超过1 TB本地SATA磁盘的普通老式Dell Precision 670通过网络引导的。







### 2.FPGA

Maxwell中的FPGA采用Xilinx Virtex-4器件，有两种形式。Alpha数据卡上的是XC4VFX100部件，而Nallatech卡上的是XC4VLX160。

 **XILINX** 赛灵思半导体(深圳)有限公司



**XILINX**原厂直供

赛灵思-高端FPGA行业领先者



### 3、Nallatech H101

Maxwell一半的卡是Nallatech的现成H101-PCIXM的略微变体。标准卡使用V4100LX设备；Maxwell使用的V4160LX版本要大60%，峰值时钟速度为200 MHz。

H101与V4LX160一起，在四个存储区中具有16 MB的DDR-II SRAM，以及一个512 MB的DDR-II SDRAM。四个SRAM库的峰值带宽为6.4 GB / s，而SDRAM的峰值带宽为3.2 GB / s。

### 4. Alpha Data ADM-XRC-4FX

Maxwell中的其他16个刀片服务器可容纳32个Alpha Data ADM-XRC-4FX卡。与Nallatech卡一样，ADM-XRC-4FX具有16 MB的SRAM，但在四个存储区中具有1,024 MB的DDR-II SDRAM，这为SDRAM提供了8.4GB/s的峰值内存带宽。



## 5、通信网络

到目前为止，Maxwell1有两个独立的通信网络。刀片式CPU通过单个交换机通过标准的千兆以太网联网，因此，CPU具有全部连接。这与FPGA网络形成对比，FPGA网络由相邻FPGA的MGT连接器之间的点对点链接组成，如图所示。单个CPU上托管的FPGA对在网络中形成“东西向”对。

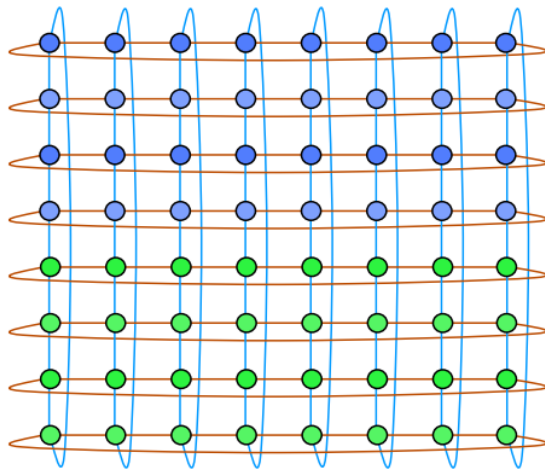


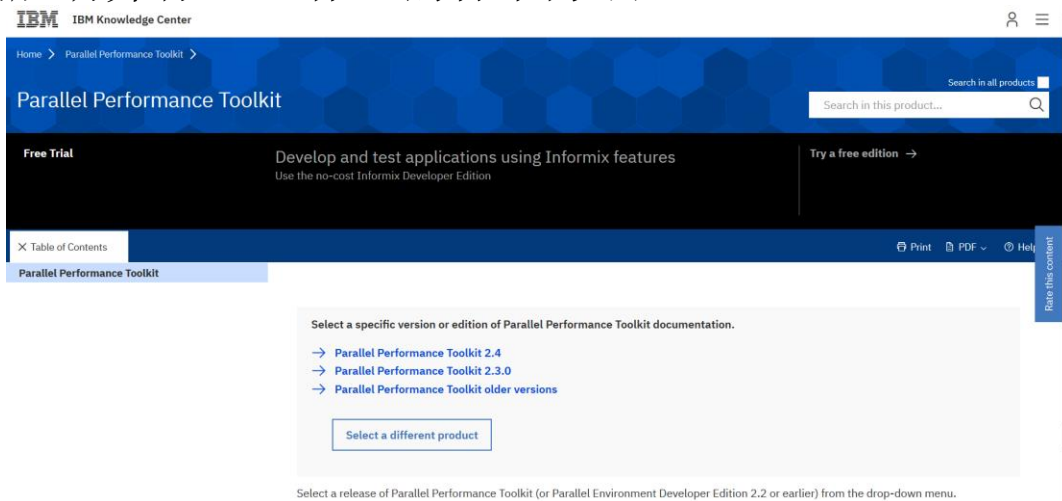
Figure 1. FPGA connectivity in Maxwell



## （二）软件环境：

### 1. HPRC的并行工具包

Parallel Toolkit（PTK）已作为Maxwell总体设计的一部分进行开发，并提供了一种尝试，以在不同的硬件之间对应用程序代码实施自上而下的标准化。PTK包括一个C++类库，该类库提供到FPGA硬件组件的抽象接口，提供可配置任意FPGA硬件的标准方法的类，以及启动并行FPGA作业的标准方法。





## 2.FPGA编程

虽然PTK提供了一种有用的，通过软件配置FPGA的通用方法，但PTK并未解决在不同类型的FPGA硬件上运行应用程序时更深层的可移植性问题。Maxwell仍然要求开发人员离线构建基于Nallatech H101或Alpha Data ADM-XRC-4FX卡的FPGA位流，并将位文件复制到系统中。

FPGA编程语言——Verilog HDL是一种用于数字系统设计的语言。用Verilog HDL描述的电路设计就是该电路的Verilog HDL模型也称为模块。Verilog HDL既是一种行为描述的语言也是一种结构描述的语言。这也就是说，无论描述电路功能行为的模块或描述元器件或较大部件互连的模块都可以用Verilog语言来建立电路模型。

```
module test_project_top( //模块名

input clk, // 时钟输入

input resetn, // 复位

input[7:0] a, //信号输入, 信号a 位宽为8 bit

input[7:0] b, //信号输入, 信号b 位宽为8 bit

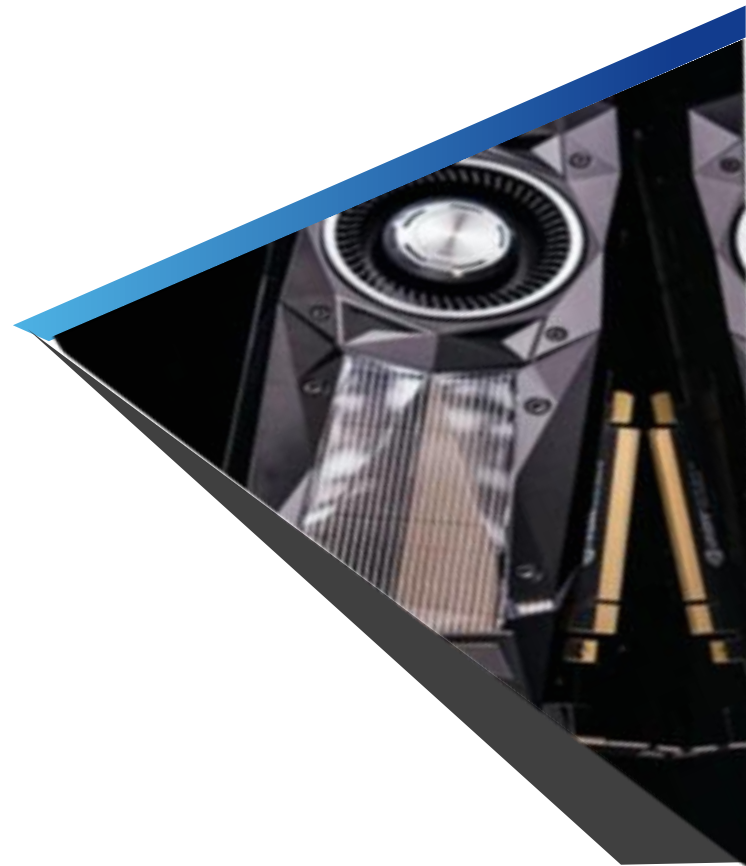
input en, // 信号是能输入 信号en 位宽为1 bit

output reg [8:0] c, // 寄存器类型定义,信号输出 c为9bit

output reg [15:0] d // 寄存器类型定义, 信号输出 d 为16bit

);
```

# 04 性能分析





首先，它们是从金融工程，医学成像以及石油和天然气的应用领域中选择，这三个领域通常被认为硬件加速解决方案中受益最大的领域。其次，选择它们来说明逐渐复杂的并行应用程序功能，从琐碎的并行性和简单的数据需求到全面的分布式内存并行性。

### 1. MCopt-Monte Carlo option pricing（蒙特卡洛期权定价）

$$dS = S r_c dt + S \sigma \varepsilon \sqrt{dt}$$

其中S是当前股票价格，dS是股票价格的变化， $r_c$ 是连续复合的无风险利率， $\sigma$ 是股票的波动率，dt是股票价格变化的时间间隔的长度出现并且 $\varepsilon$ 是从标准高斯概率分布生成的随机数。



## 性能分析

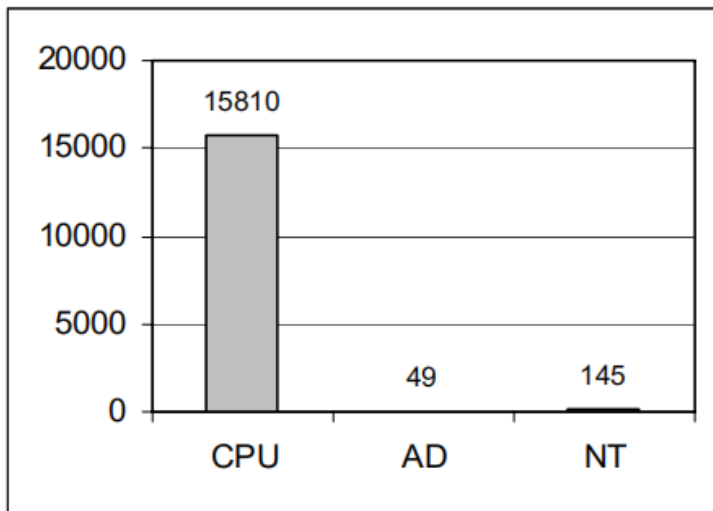


Figure 2. Single-node MCOpt performance (s).

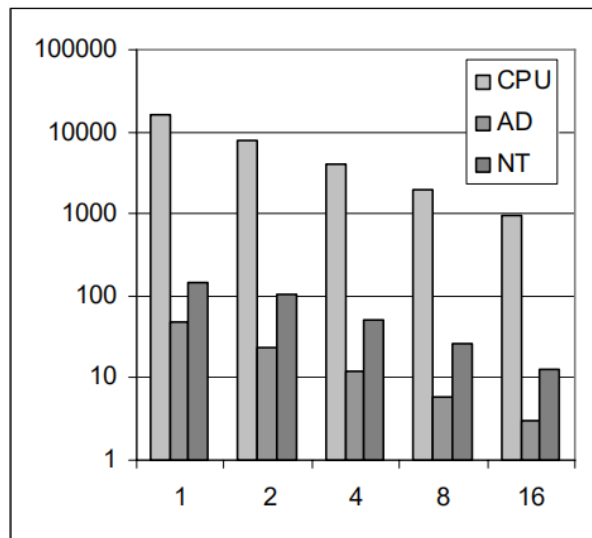


Figure 3. MCOpt scaling performance on Maxwell (times in s). Note the scale is logarithmic.



### 2、DI3D – 面部成像

使用数码相机捕获成对的静止（3D）图像或视频（4D）图像。然后将每个对组合在一起，以生成一个包含完整三维信息的深度图，并用于创建3D软件模型或4D捕获情况下的3D视频。我们这里的测试涉及批量处理32对视频静止图像，每对图像的大小约为150 kB。

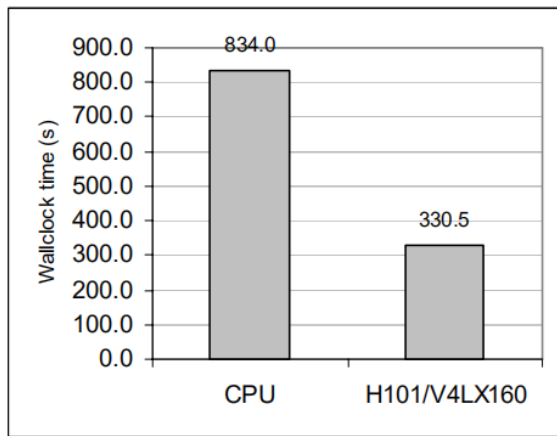


Figure 4. Two-node DI3D facial imaging performance (times in s) comparison between software and Nallatech hardware versions.

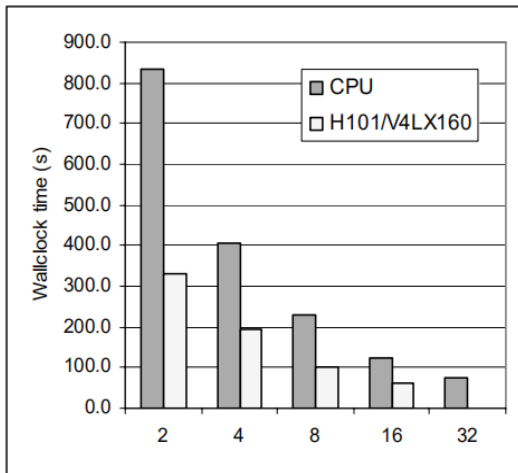
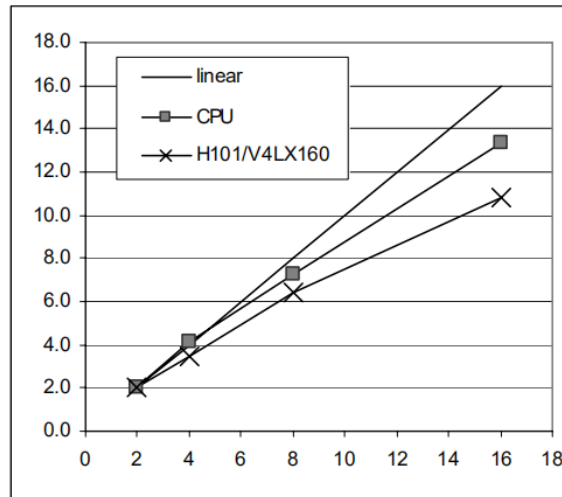


Figure 5. Facial imaging scaling performance on Maxwell (times in s) comparison of software and Nallatech H101 hardware versions.



## 3.OHM3D – CSEM建模

OHM专门从事一种称为受控源电磁建模的模拟形式，该技术利用材料的导电特性来分析海床碎片，以寻找石油或天然气储量。我们使用了一个大小为 $50 \times 45 \times 200$ （450,000点）的样本数据集，并运行求解器，直到其收敛于一个解决方案为止（在CPU和FPGA情况下均为1000次迭代）

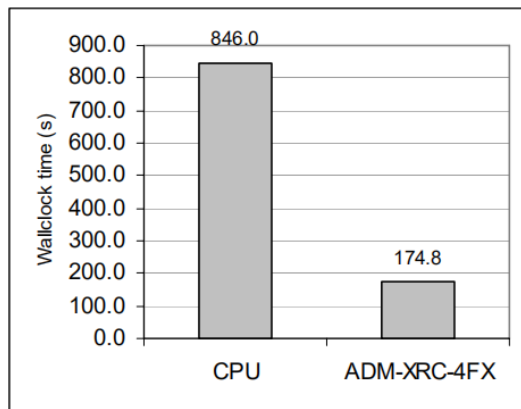


Figure 7. OHM3D 8-node performance in seconds, comparison of software and Alpha Data ADM-XRC-4FX hardware versions.

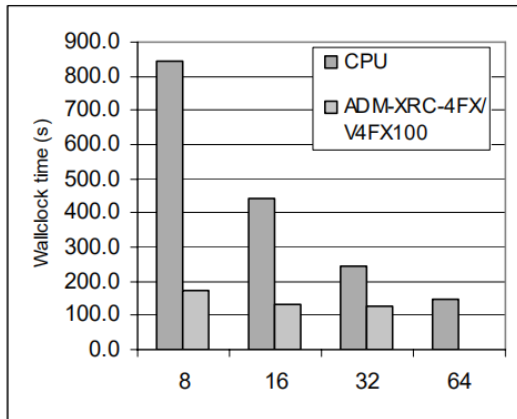


Figure 8. OHM3D scaling performance on Maxwell (times in s), comparison of software and Alpha Data ADM-XRC-4FX hardware versions

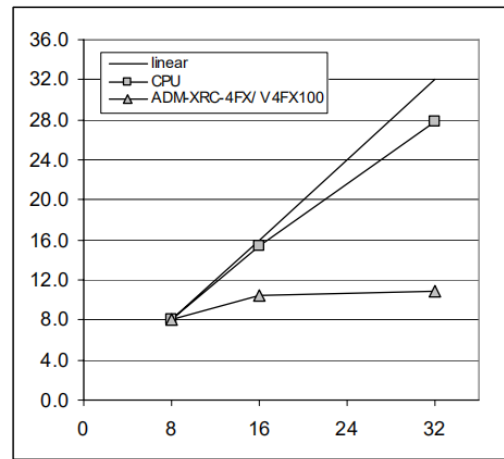


Figure 9. Parallel speedup of the OHM3D CSEM code on Maxwell, comparison of software and Alpha Data ADM-XRC-4FX hardware versions. This graph is rebased to linear at 8 nodes.

## 05 未来展望





Maxwell是世界领先的HPRC研究平台，测试的结果表明，它既有“通用”的特性，而且拥有良好的计算性能。

- 1、探索提高Maxwell和基于FPGA的系统可编程性的方法。
- 2、加强软件和硬件工程师之间的合作。
- 3、HPRC社区制定标准，将FPGAs固定到高性能计算的新架构中。

*13-project*

**汇报完毕  
谢谢欣赏**



问题：

Maxwell在哪种架构基础上进行改进的？

答案：

Maxwell在哪种架构基础上进行改进的？

Maxwell可以算Kepler(开普勒)的改进版架构。两个架构最明显的变化是在SMX单元和GPC单元上。Maxwell的SMM（之前叫SMX）单元从之前Kepler的包含192个CUDA Core下降到128个，但发射器从之前的每SMX一个变为了每SMM四个，目的是降低每个SMM单元的运算压力提升效率。增加了两个寄存器，然后L1缓存翻倍，GPC单元的L2缓存增加到了2M。