# Cray XD1超级计算系统

## ——国防科大2020年高性能评测与优化课程小组讨论

15-王军华_李岩
指导：龚春叶、甘新标、杨博

# 机构简介

NRL：

美国海军研究实验室，海军研究办公室下属的海军企业实验室。是在爱迪生的建议下于1923年成立的。1946年美国海军研究所设立后，美国海军研究实验室被置于其所长指挥下。

NRL主要利用新型材料、技术、设备、系统，面向海洋应用，进行多学科的科研与技术开发，并为海军提供广泛的专门性科技开发。

# 1. 需求分析

（1）超级计算机多用于国家高科技领域和尖端技术研究，是一个国家科研实力的体现。

（2）超级计算机对国家安全，经济和社会发展具有举足轻重的意义。是国家科技发展水平和综合国力的重要标志。

（3） 我们介绍的Cray XD1替代NRL购买的上一代超级计算机Cray MTA-2，为科学计算提供高性能的计算机资源，继续多核架构和FPGAs提供的多线程评估。

# 2. 动机

（1）Cray MTA-2满足不了科学计算需求。
（2）Cray XD1性能相比Cray MTA-2得到了很大提高。

# 3. 技术方案

XD1硬件套件：（2005）

24个底盘系统、

每个底盘有六个节点（

每个节点有：两个AMD 275双核处理器、

一个Xilinx Vertex II FPGA、

8G的内存、

73G的SATA硬盘）、

15GB字节的光纤通道磁盘系统。

# 3. 技术方案

## MTA-2 vs XD1属性对比

| | MTA-2 | XD1 |
|---|---|---|
| CLOCK RATE | 200 MHZ | 2200 MHZ |
| PROCESSORS | 40 | 576 |
| MEMORY | 160 GB | 1152 GB |
| FEATURES | MULTI-THREADING | DUAL CORES FPGAS |
| MPI ? | NO | YES |

MPI是一个跨语言的通讯协议，用于编写并行计算机。支持点对点和广播。MPI是一个信息传递应用程序接口，包括协议和和语义说明，他们指明其如何在各种实现中发挥其特性。MPI的目标是高性能，大规模性，和可移植性。MPI在今天仍为高性能计算的主要模型。

# 3. 技术方案

XD1软件套件：

Cray-Modified Suse Linux ( kernel 2.75.5)
GNU and Portland group Fortran/C/C++ compilers
MPICH 1.2.6 (MPI support)
AMD Core Math Library（AMD核心数学库）
Cray Scientific Library（ Cray科学图书馆）
PBSPro Batch Queuing System（PBSPro批量排队系统）

# 3. 技术方案

升级：（2006）

（1）12个附加底盘（相同的处理器/内存、只有一个底盘有FPGAs (6 个Vertex-4s)）

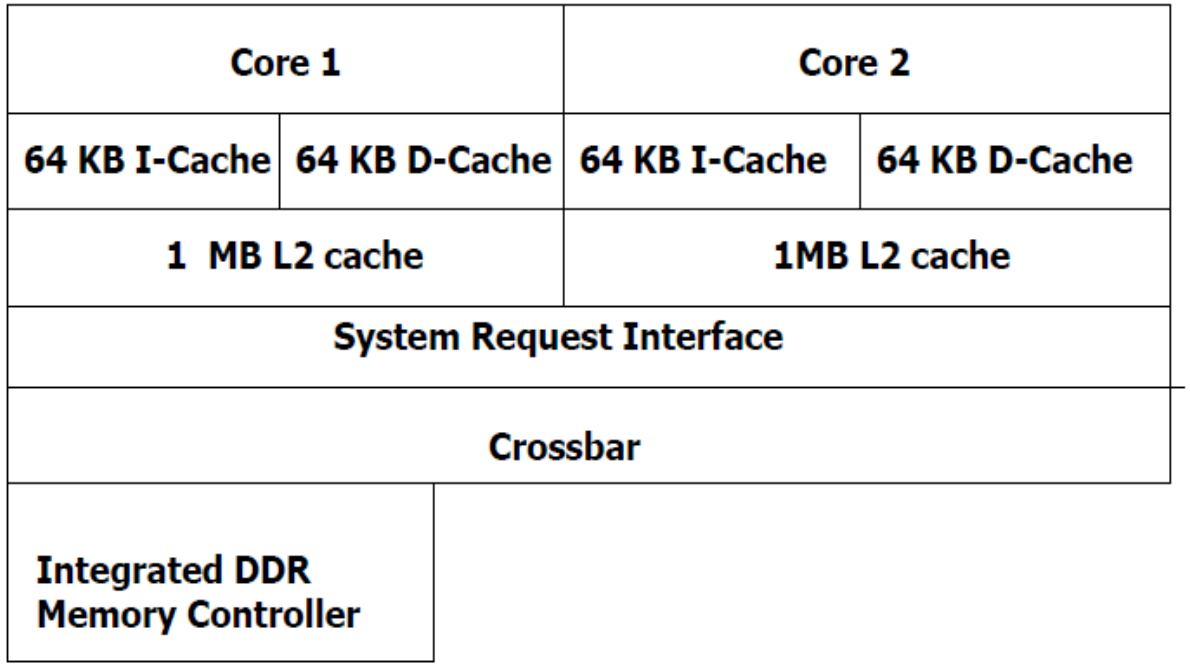
（2）可用磁盘空间扩展到30TBs

（3）FPGA软件

# 4. 改进部分

1、AMD Opteron 275双核

基于0.09微米制造工艺，主频2.2GHz，每个核心分别集成128KB的一级缓存和1MB的二级缓存，两个核心之间通过1.0GHz的HyperTransport总线通讯。Opteron 275（双核2.2G）在2005年刚上市的时候价格是1.3万。



# 4. 改进部分

2、Lustre Disk File System

（1）一个高速并行文件系统，Lustre，可用于XD1上的所有节点

（2）本地低速SATA磁盘驱动器仅可用于本地节点。尽管到本地磁盘的数据速率较慢，但可以避免与系统中其他可能正在向Lustre系统写入数据节点的竞争。

| Number of nodes | Lustre Read | Lustre Write | Local Read | Local Write |
|---|---|---|---|---|
| 1 | 206 | 165 | 40 | 58 |
| 2 | 325 | 324 | 98 | 105 |
| 4 | 629 | 646 | 114 | 209 |
| 8 | 724 | 709 | 273 | 406 |
| 16 | 892 | 862 | 374 | 804 |
| 32 | 859 | 893 | 720 | 1565 |

# 4. 改进部分

3、FPGA

（1）高性能计算FPGA的应用领域：天体物理学、生物信息学、计算化学、计算流体动力学、密码学、电磁学、超光谱成像、网络监测信号处理等。

（2）Cray XD1系列集成了Xilinx Virtex-II fpga，用于应用程序加速。FPGA具有可程序化的功效，因此VIRTEX是一颗可程序化的协同处理器。

# 4. 改进部分

4、High speed interconnect

运算模块内除了有FPGA的协同处理器外，一方面用HyperTransport与XD1的主处理器相连，另一方面使用HyperTransport 实现芯片间的互联，由此实现了高速互联。

# 5. 效果

验收标准：

48小时硬件测试

使用所有计算节点运行一个作业

应用程序使用FPGA

磁盘I / O带宽（500MB/s）

应用性能

24小时测试

30天承兑期

# 5. 效果

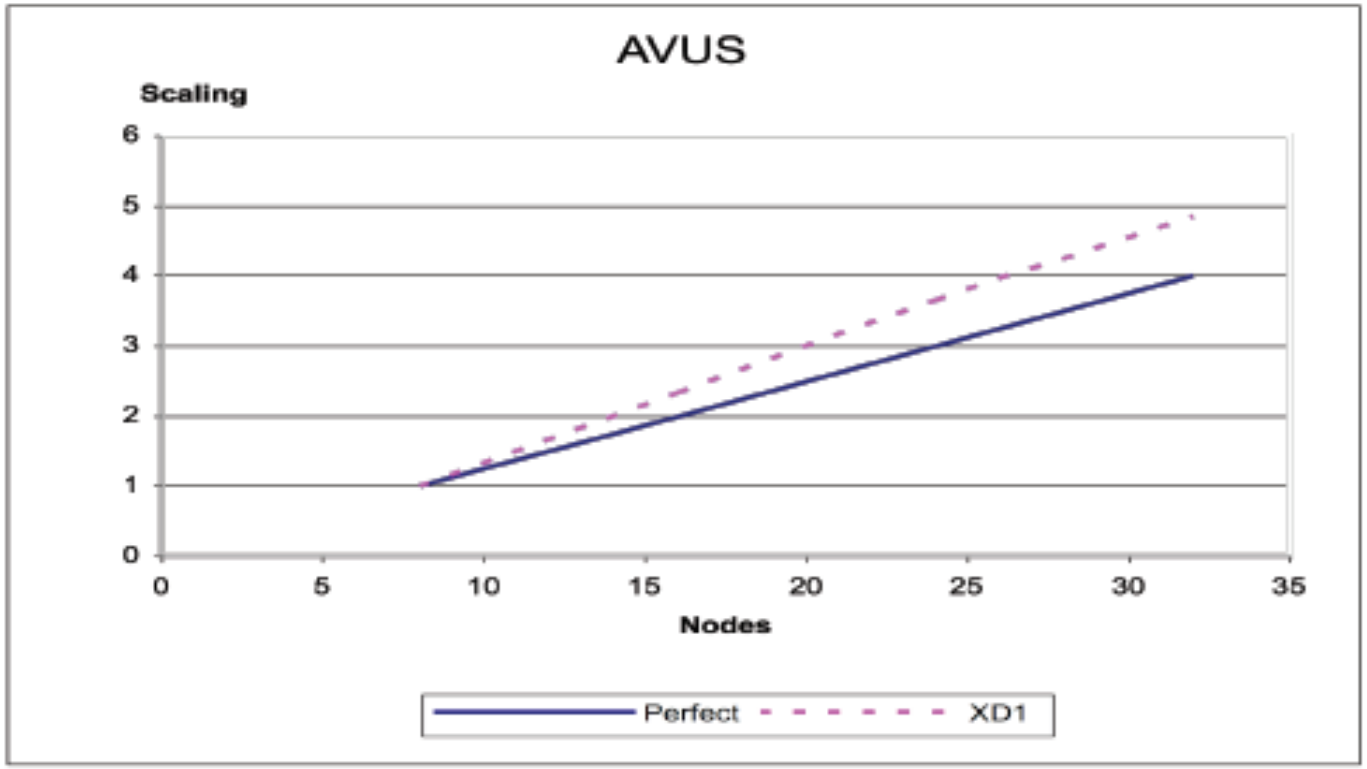
MTA-2应用程序结果

Table 1. MTA vs. XD1 Performance

| Application | MTA (seconds) | XD1 (seconds) | Speedup |
|---|---|---|---|
| STATIC | 9529 | 378 | 25.2 |
| CAUSAL | 936 | 293 | 3.2 |
| LANCZOS | 336 | 284 | 1.2 |

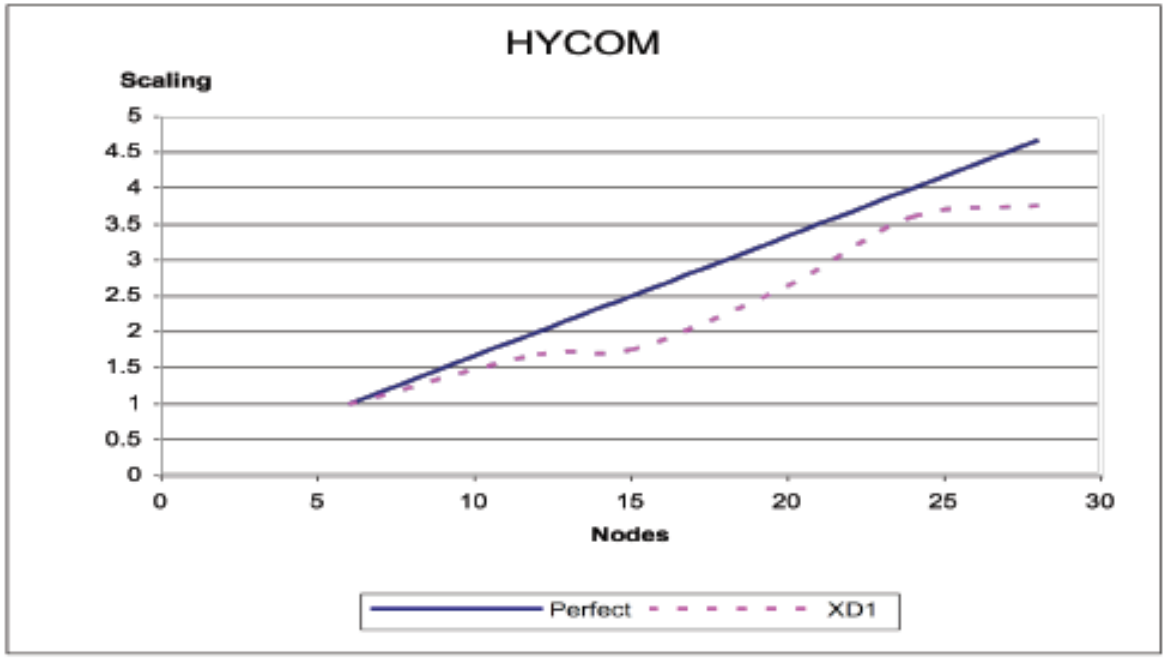
# 5. 效果

扩展应用程序



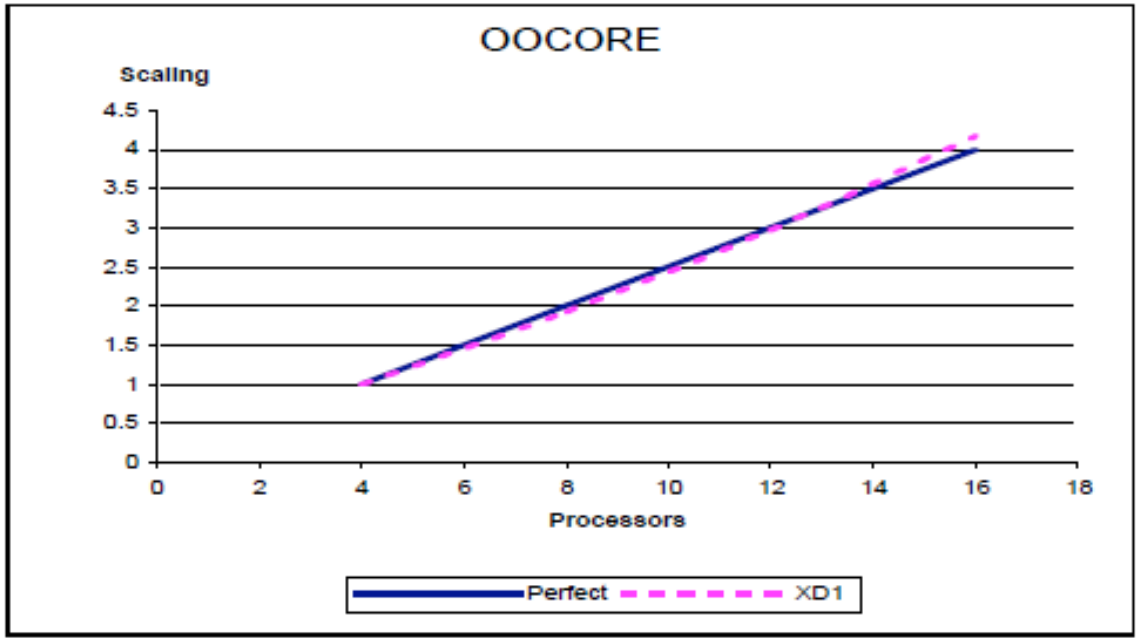
# 5. 效果

扩展应用程序


HYCOM
Scaling
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
5
10
15
20
25
30
Nodes
Perfect
XD1

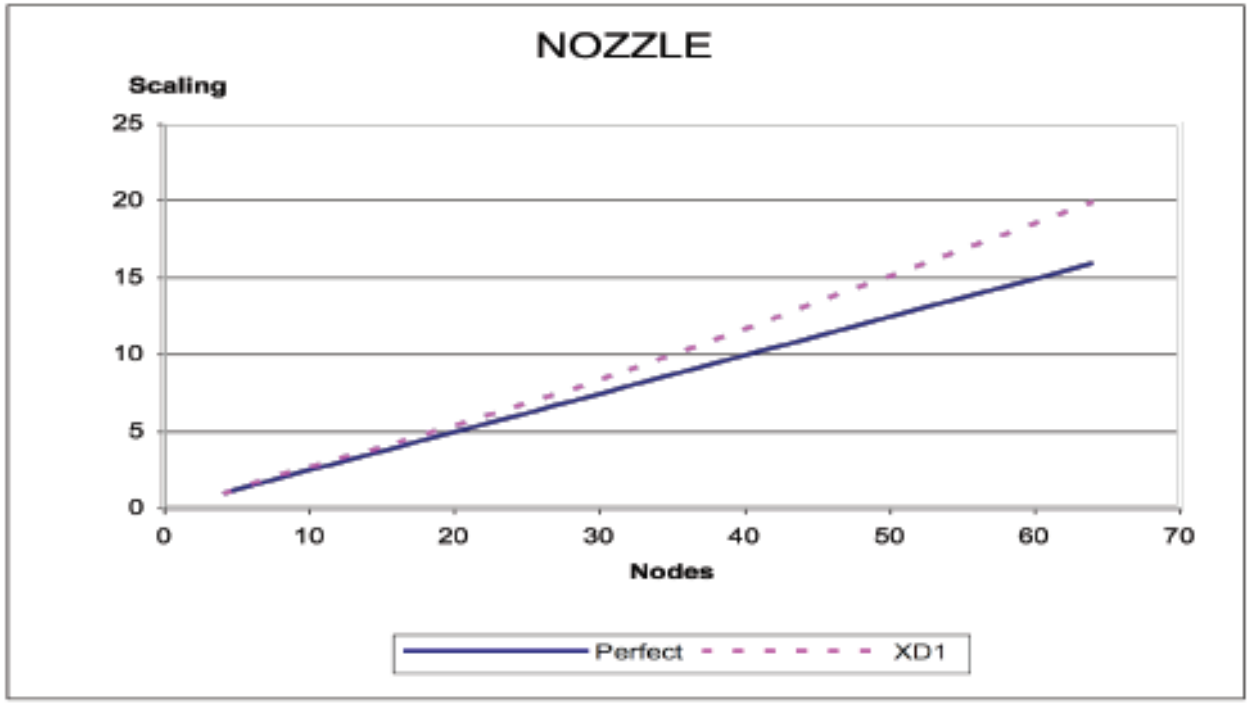
# 5. 效果

扩展应用程序


OOCORE
Scaling
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
2
4
6
8
10
12
14
16
18
Processors
Perfect
XD1

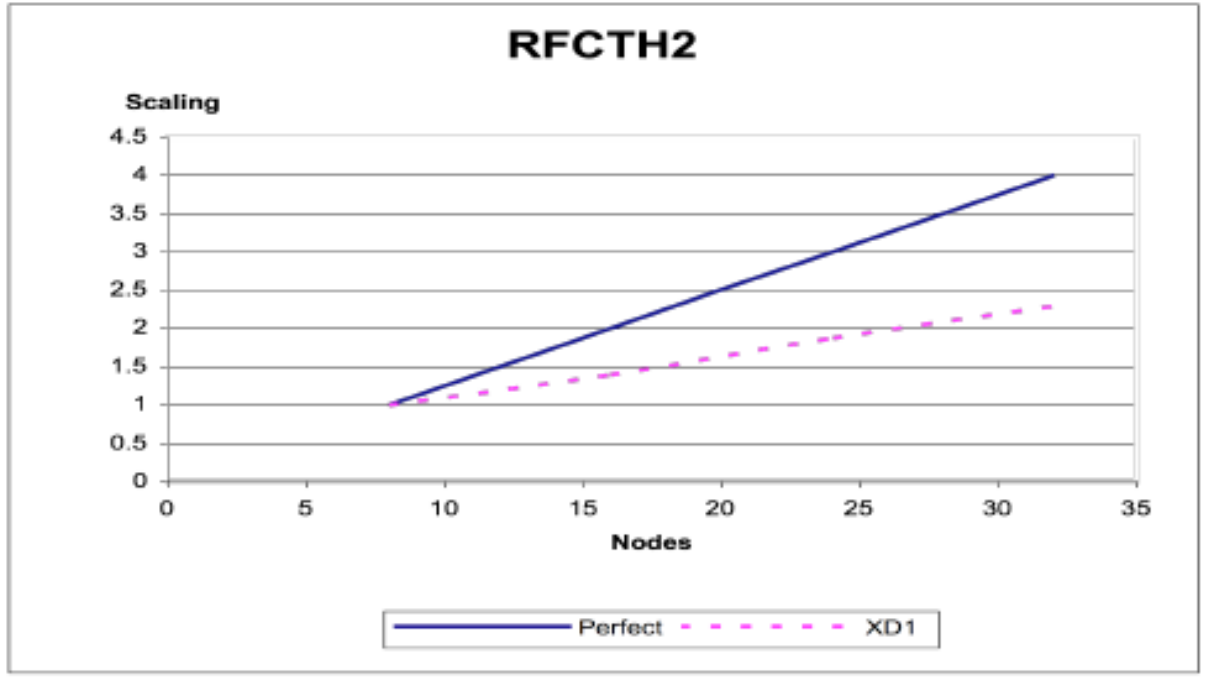
# 5. 效果

扩展应用程序


NOZZLE
Scaling
25
20
15
10
5
0
0
10
20
30
40
50
60
70
Nodes
Perfect
XD1

# 5. 效果

扩展应用程序


RFCTH2
Scaling
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
5
10
15
20
25
30
35
Nodes
Perfect
XD1

# 6. 分析总结

XD1主要通过以下四点提升了系统性能：

Opteron 275 Dual Cores
Lustre Disk File System
High speed interconnect
FPGA

# 6. 分析总结

哪些可以接着改进：

（1） 处理器性能可进一步提升

（2） 通信延迟

（3）FPGA改进

问题：

# 谢谢大家！