

# SCC经验分享

黄可钊

# GPU@SCC

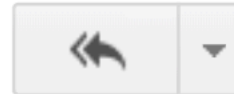
- benchmark + deep learning: GPU数量是永远的谜

(万一神秘应用是GPU应用呢?)



Manuel Burger ETHZ

19/11/20



[将帖子翻译为中文](#)

Hi, thank you for your reply.

I can extract from this, that you really want to force us to use 6 GPUs. Our team brought 12 GPUs here and we are paying with that with idle power consumption and heat etc. We were really happy to hear, that your application allows us to use the GPUs. I can easily make all of the inputs run on all of our GPUs and the results are perfectly fine and still correct. So this is neither a limitation of the software nor of the hardware, but artificially imposed. Bringing an accelerator with us in the cluster is a design and tradeoff decision, as the objective here is anyway to design and run the best cluster within the power limit of 3000W.

Could you please elaborate on why you forbid the usage of more GPUs?

Thank you  
Manuel Burger, ETH Zürich

# GPU@SCC

- 是否要把应用上GPU?

应用并行性，计算func种类数和粒度，数据传输频率，性能分是总时间还是计算阶段的flop

- 对于Benchmark程序的快速上G:

一核有难，万核围观

实际时间多很多，但是只算计算部分的时间

# GPU@SCC

- GPU profiling

Nvprof, nvvp

- Nvidia CUDA samples, NCCL-Test

各种micro benchmark, 更好了解GPU各种性能特性: GPU-GPU, GPU-CPU通信, GPU的一些基本算子的FLOP, GPU中各种memory(shared mem, texture mem, global mem)的bw&lat, GPU里提供的fancy的功能

# InfiniBand@SCC

- Host chaining or 对插:

(在某些比赛) 减少IB交换机功耗 (~100W)

- IB的温度同样需要关心

# 据官方文档Operational Temperature是0°C - 55°C Non-operational 是-40°C to 70°C

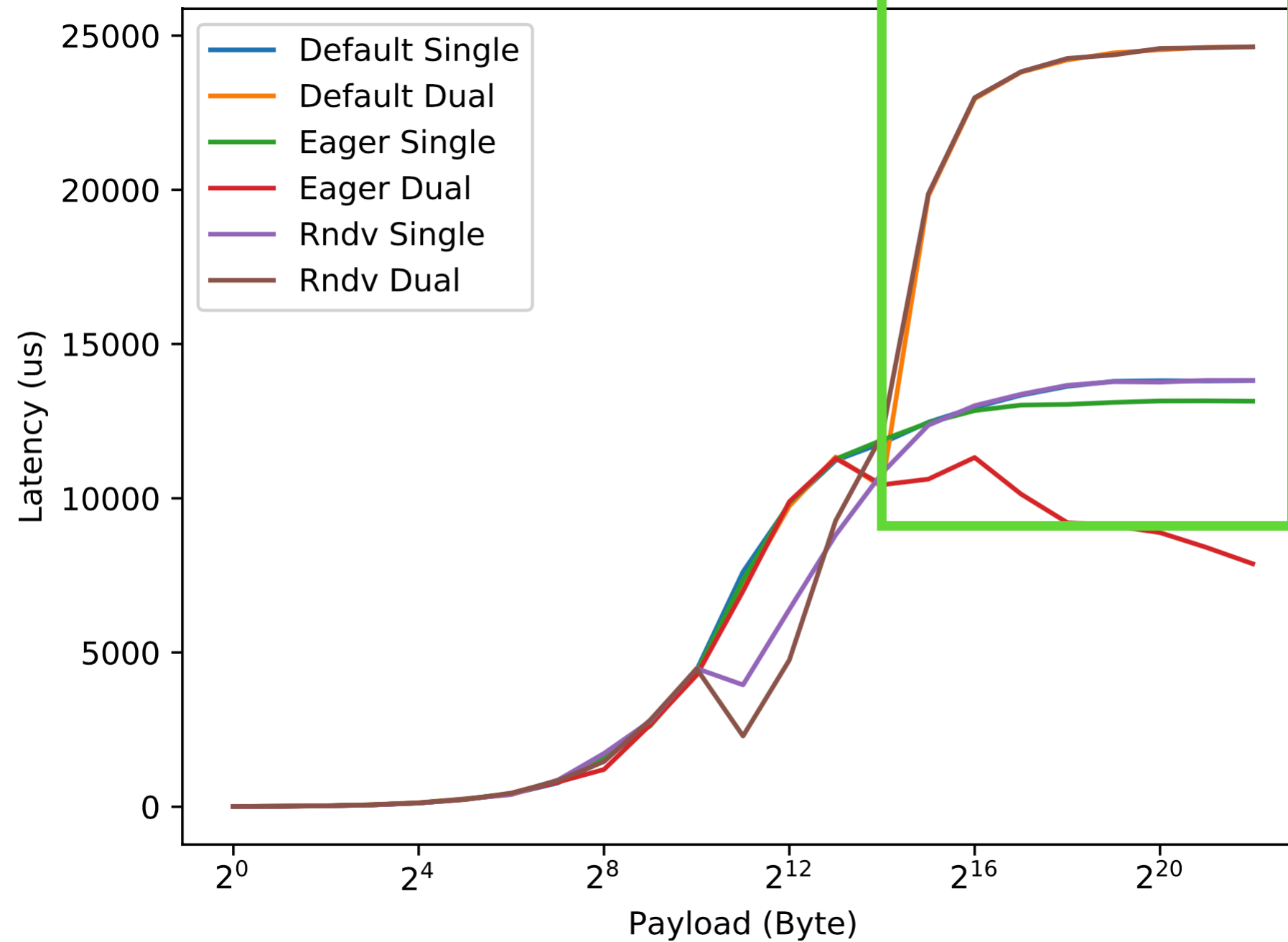
```
sudo mget_temp -d $devname
```

# InfiniBand@SCC

- Dual-IB:
  - 更高的带宽，还可弥补因PCIE宽度不够带来的EDR->FDR
  - 通信没有NUMA effect
  - 甚至有加速节点内NUMA间通信的奇效
  - 但是真正用上Dual-IB了吗？用OSU-Benchmark看看
  - 应用适合使用双IB吗

# InfiniBand@SCC

要跑的应用的消息传递在这个位置吗？



*MPI Protocol*, 消息大小对双IB带宽的影响

# MPI@SCC

- MPI的选择：IMPI / Openmpi / HPCX
- 还是需要profiling一个：

IPM, HPC-Toolkit (hpcrun, hpcstruct, hpcprof, hpcviewer, hpctraceviewer) , vTune

- 优化点常常来自于load imbalance



# 程序优化@SCC

- 非代码相关：

试编译器，MPI，tmpfs，omp/pthread，进程绑定，rank数量选择

- 代码相关：

可能有优化点？

以不变应万变

# 其他@SCC

- 现场改代码：快**糙**猛

HPCC：现场上GPU

VPIC：现场改打checkpoint方式

- 多问组委会：

问清楚永远不是坏事，赛场上最准确的分类器

# 其他@SCC

- 搞清楚出了什么问题，而不是直接重启
- 组委会究竟需要提交什么东西（对于一些要dump很多数据的科学应用）



# 收获的东西@SCC

- 做性能优化相关研究的一些基本功
- 可以和硬件亲密接触，建立感情（毕业就靠它们了）
- 发现高性能计算中真实的问题
- 好玩