



HyQual 用户使用手册

日期：2025/04/30

智能计算产品事业部

修改记录

日期	修正内容	版本号
2022/04/29	创建 HyQual 用户使用手册文档	v1.0.0
2022/05/06	添加环境配置需求 新增常见问题解析	V1.0.1
2022/06/13	修正部分显示错误 按照版本号规则调整文档版本号	V1.6.0
2022/06/17	更新版本 1.3.5 Thermal 测试可以设置压测时间 修复了 gen4 rocm-band 判断	V1.6.1
2022/09/01	第四节，添加命令行方式执行功能 pcie 带宽测试添加参考基准	V1.9.0
2022/10/27	新增 pice margin test 测试	V1.10.0
2023/07/21	基于 HyQual-2.0.5.tar.gz 版本 新增 DCUburn 测试 兼容 DTK23.04 修复 rocmbandwidth 测试 bug gemmperf 新增 int8、int4 测试 peakperf 新增 fp16/int8 测试 修复 ubuntu OS 上的退出问题	V2.0.0
2024/11/30	合并 gemmperf 和 peakperf 为 gemm_peak 将 runtime 集成到工具中，不再依赖 dtk 移除 anaconda 包， 对 Z100L/K100/K100-AI 进行了兼容性支持 移除 pice margin test 测试	V2.2.6
2025/04/28	优化 runtime，新增特殊硬件支持 修改容器化部署流程	V2.2.7

目录

1	概述.....	1
2	HyQual 安装与配置.....	2
2.1	系统要求.....	2
2.2	HyQual 目录结构介绍.....	2
2.3	安装步骤.....	2
2.3.1	依赖环境安装	2
2.3.2	宿主机部署	3
2.3.3	容器化部署	3
3	使用指南.....	4
3.1	工具参数说明	4
3.2	Thermal Core 测试.....	5
3.2.1	测试项目介绍	5
3.2.2	测试项目运行	5
3.2.3	测试结果查看	5
3.3	Thermal Mem 测试.....	8
3.3.1	测试项目介绍	8
3.3.2	测试项目运行	8
3.3.3	测试结果查看	8
3.4	PCIe 带宽测试.....	11
3.4.1	测试项目介绍	11
3.4.2	测试项目运行	11
3.4.3	测试结果查看	12
3.5	xHMI 带宽测试.....	14
3.5.1	测试项目介绍	14
3.5.2	测试项目运行	14
3.5.3	测试结果查看	14
3.6	显存带宽测试.....	16
3.6.1	测试项目介绍	16
3.6.2	测试项目运行	16
3.6.3	测试结果查看	16
3.7	硬件算力测试.....	18
3.7.1	测试项目介绍	18
3.7.2	测试项目运行	18
3.7.3	测试结果查看	19
3.8	其它.....	21

1 概述

HyQual 是海光公司开发的用于针对 DCU 标卡系列产品的压力、带宽以及算力测试标准工具。

HyQual 可以用作长时间稳定性测试，作为 DCU 的基础环境检查工具，有助于客户检测 DCU 环境是否正常；也可以用作 DCU 相应指标性能的标准化测试，帮助客户检测 DCU 是否符合预期的性能标准。

接下来，本手册将详细介绍 HyQual 的目录结构，测试内容，使用方式以及故障处理等。

2 HyQual 安装与配置

2.1 系统要求

工具名称	HyQual
版本	v2.2.7
硬件要求	支持 DCU 硬件版本-Z100L/K100/K100-AI/K100-AI-ECO 16GB 以上内存 20GB 磁盘空间
软件要求	Python3 驱动 $\geq 6.2.31$ Glibc ≥ 2.17 libstdc++-devel $\geq 4.8.5$
OS 支持	Linux（建议 ubuntu20.04 及以上、Centos7.6 及以上） 推荐 OS: ubuntu20.04 不支持 Win

2.2 HyQual 目录结构介绍

```
[root@master01 hyqual-v2.2.6]# ls
bin lib miniconda3 result rocm_install run
```

文件/目录	说明
bin	存放 HyQual 运行所调用的编译工具
lib	存放 HyQual 运行所需要的动态库
miniconda3	HyQual 工具的依赖环境
result	HyQual 运行结果保存路径
rocm_install	HyQual 运行需要的 runtime 环境
run	HyQual 测试项目调用接口

2.3 安装步骤

2.3.1 依赖环境安装

驱动安装：参考[驱动安装步骤](#)

环境检查：驱动安装完成后，执行下面命令验证下环境：/opt/hyhal/bin/hy-smi

```
core_stress_test_end
[root@master01 hyqual_peak_v2.2.6]# /opt/hyhal/bin/hy-smi

===== System Management Interface =====
=====
DCU      Temp      AvgPwr    Perf      PwrCap    VRAM%     DCU%      Mode
0        36.0C    40.0W     auto      280.0W    0%         0.0%     Normal
1        36.0C    39.0W     auto      280.0W    0%         0.0%     Normal
2        37.0C    38.0W     auto      280.0W    0%         0.0%     Normal
3        38.0C    44.0W     auto      280.0W    0%         0.0%     Normal
=====
===== End of SMI Log =====
```

2.3.2 宿主机部署

```
##### 部署前提 #####
# 需要 root 权限，建议 root 用户下运行 HyQual
##### 部署流程 #####
tar xvf hyqual-v2.2.7.tar.bz2
cd hyQual-v2.2.7
```

2.3.3 容器化部署

```
##### 部署前提 #####
# 光源拉取一个 ubuntu 的镜像，比如
docker pull image.sourcefind.cn:5000/dcu/admin/base/dtk:25.04-ubuntu22.04-py3.10-fixpy ###其他镜像也可以
##### 创建容器 #####
docker run -itd --name <your_container_name> \
--privileged \
--shm-size 128g \
--device /dev/dri \
--device /dev/kfd \
--device /dev/mkfd \
-v /opt/hyhal:/opt/hyhal \
-v /sys/kernel/debug:/sys/kernel/debug \
-v /path/to/your_workspace:/your_workspace \
image.sourcefind.cn:5000/dcu/admin/base/dtk:25.04-ubuntu22.04-py3.10-fixpy

##### 复制工具到容器 #####
# 拷贝工具到容器中
cp hyqual_v2.2.7 /path/to/your_workspace
# 进入容器
docker exec -it <your_container_name> bash
cd /your_workspace/hyqual_v2.2.7
```

3 使用指南

HyQual 支持命令行和交互两种运行方式，它可以自动识别服务器上所有的 DCU 卡并进行相同的操作，目前测试项目如下图所示。

```
(base) [root@bogon hyqual_v2.2.7]# ./run
Please select one of the following options:

1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Gemmperf Stress Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
Key-in selection followed by <enter>:
```

3.1 工具参数说明

项目	测试名称	说明
1	Thermal Core Qualification Test	计算核心压力测试，记录 DCU 功耗、温度、频率等性能
2	Thermal Mem Qualification Test	显存压力测试，记录 DCU 功耗、温度、频率等性能
3	Thermal Core and Mem Qualification Test	计算核心和显存同时提供压力，记录 DCU 功耗、温度、频率等性能
4	PCIe Bandwidth Test	测试 DCU 的 PCIe 带宽
5	xHMI Bandwidth Test	测试 DCU 的 XHMI 带宽
6	Graphic Mem Bandwidth Test	测试 DCU 的显存带宽
7	Peak Performance Test	测试 DCU 的不同精度的算力
h	HyQual Version	显示 HyQual 版本信息
t	Generate System Topology	输出 DCU 卡拓扑图
q	Quit HyQual	退出 HyQual 程序

表 2 HyQual 功能参数表格

3.2 Thermal Core 测试

3.2.1 测试项目介绍

该测试项目是对 DCU 的计算核心进行长时间计算任务。

该测试项目会把 DCU 功耗压倒 TDP 功耗，然后计算核心的利用率压到 100%，由于采用的随机的数据，所以显存没有多大占用。

如果散热有问题，该测试项目可能会不通过。

3.2.2 测试项目运行

可以通过两种方式运行，交互方式和命令行方式。

交互方式运行：

```
./run
然后输入 1,sec sec 表示压测时间

(base) [root@bogon hyqual_v2.2.7]# ./run
Please select one of the following options:

1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Gemmperf Stress Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
Key-in selection followed by <enter>:
```

命令行方式运行：

```
./run 1,sec
```

3.2.3 测试结果查看

交互界面会呈现压测的最终结果，如下图所示


```

[root@anolis hyqual-v2.2.6]# ./run
Please select one of the following options:

1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Peak Performance Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
Key-in selection followed by <enter>:1,200
core_stress begin,will continue 200 sec
[Info] Device count: 1
[Test Start] 2024-12-9 10:35:13
[Info] ===== Device 0 =====
[Info] kernel Launch Success

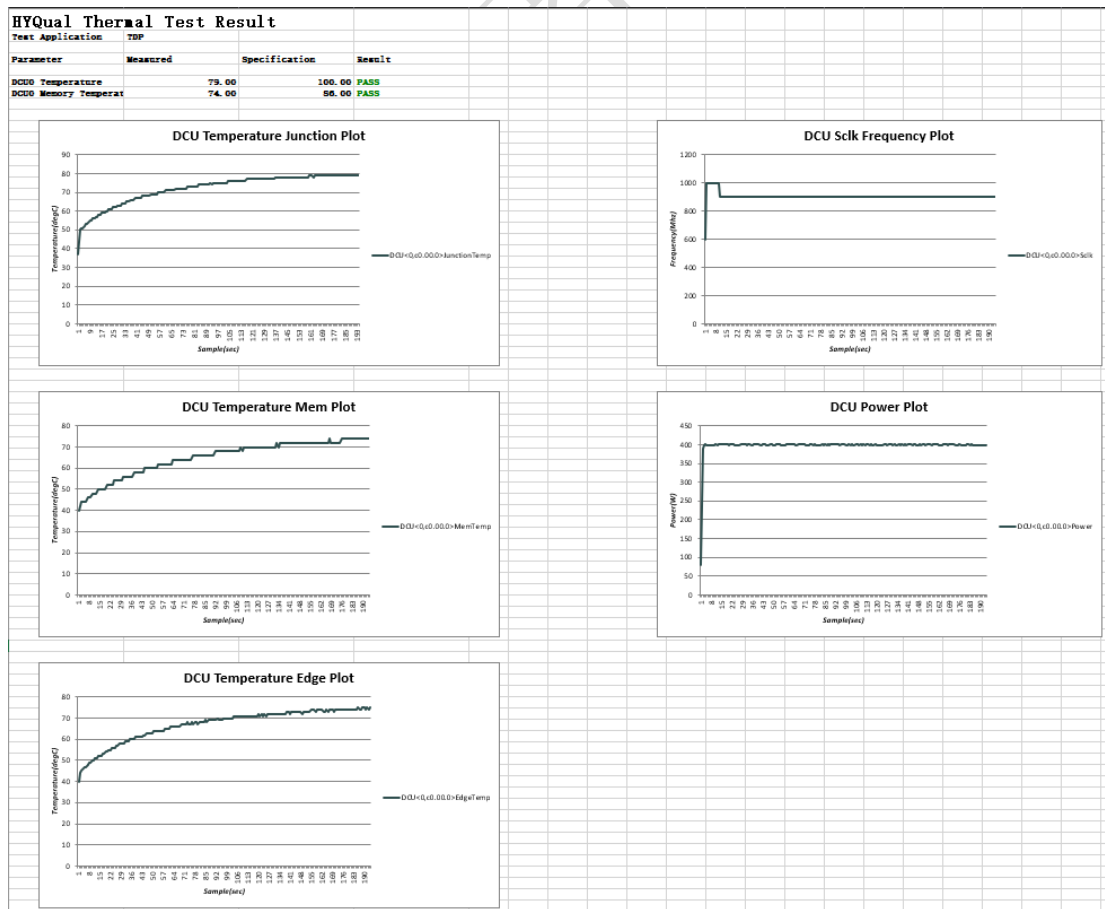
Thermal Core Result: PASS

Thermal Core Result: PASS

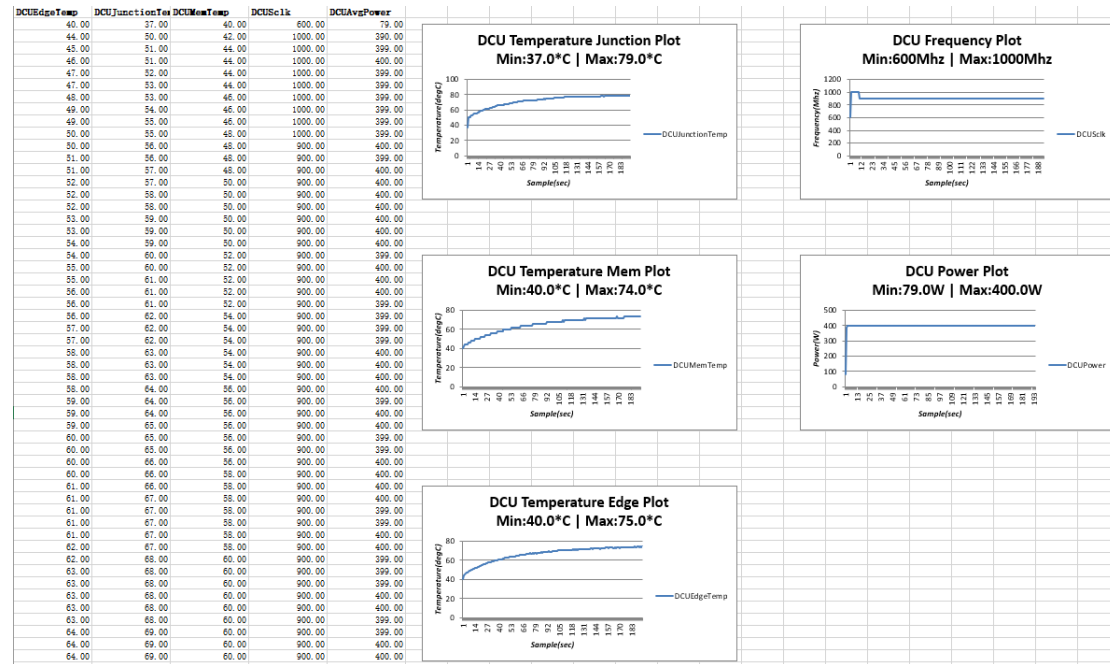
```

./result/thermal 中会记录运行数据，命名为：thermal_test_core_*.xlsx，里面有详细的压测信息。

首页记载了所有设备的压测总结，包含了计算核心与显存单元压测阶段的最大值，以及它们的门限值，如果温度超过了门限值的话，该测试失败，需要查看服务器的散热有没有问题。除此之外，首页还记载了所有 DCU 设备的监控信息的概览图。



其它 sheet 则记录了每一张 DCU 设备压测过程中的监控信息，包括每一项测试信息的记录以及总的信息概览图。



3.3 Thermal Mem 测试

3.3.1 测试项目介绍

该测试项目是对 DCU 的显存进行长时间读写任务，然后对读到的数据进行加和等基础操作。

该测试项目会把 DCU 功耗压倒 TDP 功耗，显存占用在 4%左右，计算核心的利用率也会压倒最大。

该项目也跟服务器的散热息息相关，如果散热有问题，该测试项目可能会不通过。

3.3.2 测试项目运行

可以通过两种方式运行，交互方式和命令行方式。

交互方式运行：

```
./run
然后输入 1,sec sec 表示压测时间

[root@anolis hyqual-v2.2.6]# ./run
Please select one of the following options:

  1. Thermal Core Qualification Test
  2. Thermal Mem Qualification Test
  3. Thermal Core and Mem Qualification Test
  4. PCIe Bandwidth Test
  5. xHMI Bandwidth Test
  6. Graphic Mem Bandwidth Test
  7. Peak Performance Test

  h. HYQual Version
  t. Generate System Topology
  q. Quit HYQual
key-in selection followed by <enter>:2,200
mem_stress begin,will continue 200 sec
Device count: 1
Start time: 2024-12-09_14:43:13
===== Device 0 =====
Device name   : K100_AI
CU count      : 120
Global memory: 68702699520 bytes
Execute time  : 200 s
Is verify     : Yes
Is cached     : No
```

命令行方式运行：

```
./run 1,sec
```

3.3.3 测试结果查看

交互界面会呈现压测的最终结果，如下图所示

```
Key-in selection followed by <enter>:2,200
mem_stress begin,will continue 200 sec
Device count: 1
Start time: 2024-12-09_11:05:09
===== Device 0 =====
Device name   : K100_AI
CU count      : 120
Global memory: 68702699520 bytes
Execute time  : 200 s
Is verify     : Yes
Is cached     : No

Copy data to device 0 done.
Device 0 memory stress test in progress...

Thermal Core Result: PASS

Thermal Core Result: PASS

=====System Management Interface =====
=====End of SMI Log=====
Device 0 test done.
mem_stress test end
```

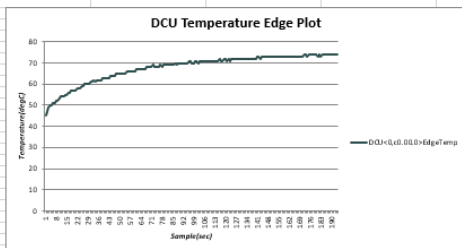
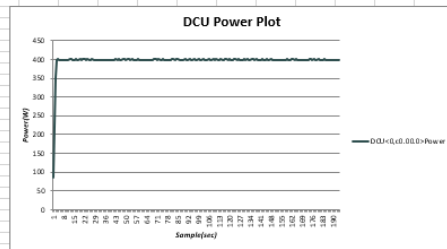
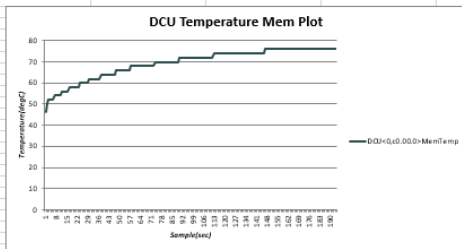
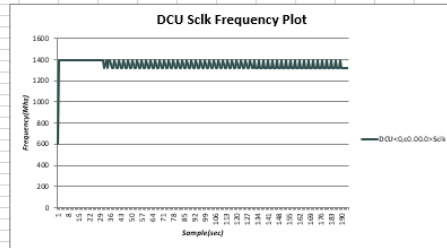
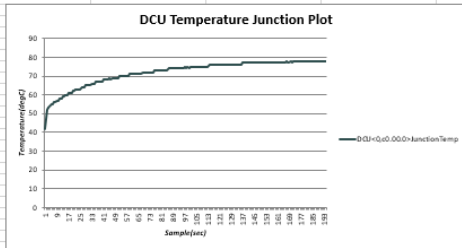
./result/thermal 中会记录运行数据，命名为：thermal_test_mem_*.xlsx，里面有详细的压测信息。

首页记载了所有设备的压测总结，包含了计算核心与显存单元压测阶段的最大值，以及它们的门限值，如果温度超过了门限值的话，该测试失败，需要查看服务器的散热有没有问题。除此之外，首页还记载了所有 DCU 设备的监控信息的概览图。

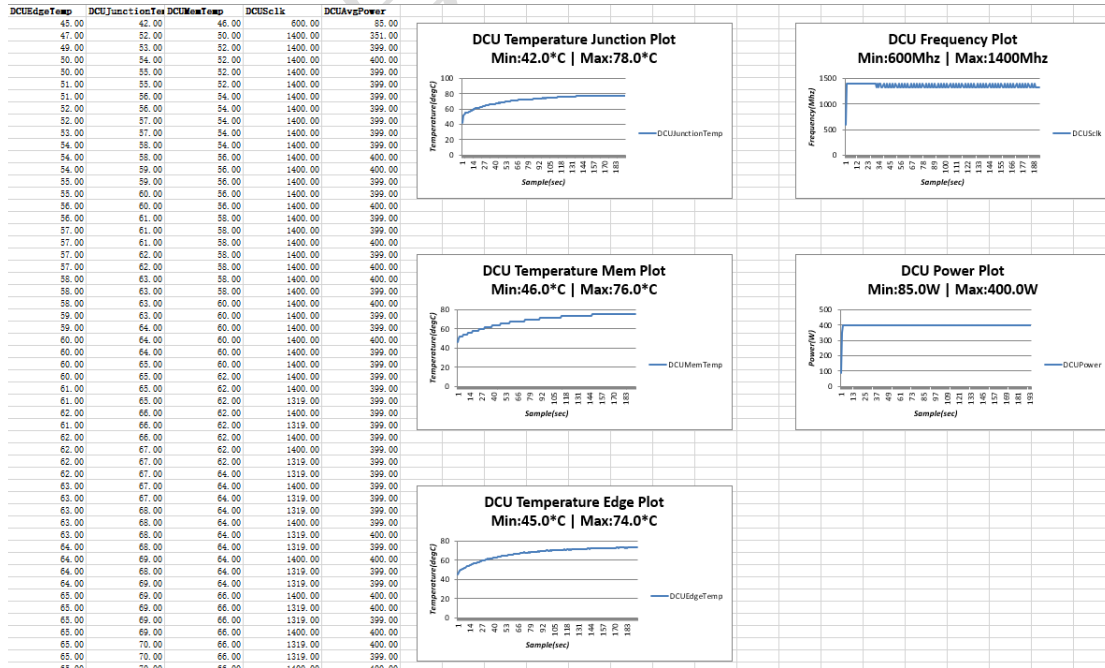
HYQual Thermal Test Result

Test Application TSP

Parameter	Measured	Specification	Result
DCUJ Temperature	78.00	100.00	PASS
DCUJ Memory Temperature	78.00	86.00	PASS



其它 sheet 则记录了每一张 DCU 设备压测过程中的监控信息，包括每一项测试信息的记录以及总的信息概览图。



3.4 PCIe 带宽测试

3.4.1 测试项目介绍

该测试项目是测试 CPU 到 DCU 的数据传输速度，以及 DCU 到 CPU 的数据传输速度。目前 DCU 支持的是 PCIe4.0 以及 X16 的链路带宽。表 为 PCIe 的传输速率。

PCIe	5.0	4.0	3.0	2.0	1.0
速率	32GT/s	16GT/s	8GT/s	5GT/s	2.5GT/s
X16	64GB/s	32GB/s	16GB/s	8GB/s	4GB/s
X8	32GB/s	16GB/s	8GB/s	4GB/s	2GB/s
X4	16GB/s	8GB/s	4GB/s	2GB/s	1GB/s
X1	4GB/s	2GB/s	1GB/s	500MB/s	250MB/s

DCU 实际上运行的 PCIe 版本既跟 CPU 有关，也跟服务器有关：CPU 的影响如下，其它条件满足的条件下，如果采用 HG 7285 CPU，它最高支持 PCIe3.0 的协议，则相关的 PCIe 带宽受 PCIe3.0 的限制，单向理论带宽是 16GB/s；如果是 HG 7380 CPU，支持 PCIe4.0 的协议，单向理论带宽是 32GB/s。CPU 满足 PCIe4.0 的条件，如果服务器仅支持到 PCIe3.0 的话，则 DCU 单向理论带宽是 16GB/s；如果服务器支持到 PCIe4.0 的话，则 DCU 单向理论带宽能到 32GB/s。

3.4.2 测试项目运行

可以通过两种方式运行，交互方式和命令行方式。

交互方式运行：

```
./run
然后输入 4

[root@anolis hyqual-v2.2.6]# ./run
Please select one of the following options:

1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Peak Performance Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
Key-in selection followed by <enter>:4
-----
Theoretical raw bandwidth = 32   GB/s
Min average raw bandwidth = 22.4 GB/s
```

命令行方式运行：

```
./run 4
[root@anolis hyqual-v2.2.6]# ./run 4
-----
Theoretical raw bandwidth = 32   GB/s
Min average raw bandwidth = 22.4 GB/s
```

3.4.3 测试结果查看

一般 PCIe 的实际性能在理论性能的 80%左右，这属于正常的性能损失。

测试结果中“Fb to Sys”表示 D2H（device to host）的带宽，“Sys to Fb”表示 H2D（host to device）带宽。

本次测试结果在理论性能的 70%-100%之间，都认为是满足需求，受不同 OS、不同平台、不同硬件设备的影响，测试结果会略有差异。

```
[root@bogon hyqual_peak_v2.2.6]# ./run 4
-----
Theoretical raw bandwidth = 32   GB/s
Min average raw bandwidth = 22.4 GB/s
Could not open /var/log/hylog/.
DCU0 PCIe Fb to Sys:  28.174 GB/s PASS
DCU0 PCIe Sys to Fb:  26.381 GB/s PASS
```

1. HG 3 号平台 PCIe4.0 测试结果

```
[root@master01 hyqual-v2.2.6]# ./run 4
-----
Theoretical raw bandwidth = 16   GB/s
Min average raw bandwidth = 11.2 GB/s
DCU0 PCIe Fb to Sys:  14.132 GB/s PASS
DCU0 PCIe Sys to Fb:  13.061 GB/s PASS
DCU1 PCIe Fb to Sys:  14.132 GB/s PASS
DCU1 PCIe Sys to Fb:  13.076 GB/s PASS
DCU2 PCIe Fb to Sys:  14.132 GB/s PASS
DCU2 PCIe Sys to Fb:  13.056 GB/s PASS
DCU3 PCIe Fb to Sys:  14.132 GB/s PASS
DCU3 PCIe Sys to Fb:  13.056 GB/s PASS
```

(1) HG 2 号平台 PCIe3.0 测试结果

测试详细结果可以在 `results/bandwidth/ pcie_bandwidth_*.log` 中查看，该测试项目主要看单向带宽，如下图所示，0-7 表示 CPU 的 8 个 numa，8-15 表示 8 个 DCU，图中结果表示纵坐标设备往横坐标设备数据传输结果，选取每个设备的最大性能，比如查看 DCU0 的 H2D 性能，图中结果是[0-7]->8 的性能，最大性能是 3->8 的性能，所以 DCU0 的 H2D 性能是 25.971GB/s；DCU0 的 D2H 性能对应 8->[0-7]，最大性能是 8->3 的性能，所以 DCU0 的 D2H 性能是 28.184GB/s

Unidirectional copy peak bandwidth GB/s																	
D/D	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	20.061	19.992	20.213	20.003	17.415	16.752	16.793	16.741
1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	20.007	20.299	20.254	20.525	17.936	17.906	17.984	18.254
2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	20.026	20.018	20.008	20.002	17.681	17.717	17.706	17.678
3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	25.971	25.971	25.971	25.969	18.152	17.694	17.828	17.727
4	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	16.817	17.220	17.176	17.141	20.522	20.667	20.468	20.500
5	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	18.327	18.227	18.236	18.315	20.609	20.290	20.342	20.414
6	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	17.753	17.741	17.758	17.982	20.648	20.192	20.237	20.182
7	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	17.714	17.850	17.726	18.373	26.214	26.211	26.211	26.206
8	18.191	18.207	18.204	28.184	14.136	14.133	14.132	14.137	276.214	28.182	28.184	28.180	14.131	14.130	14.135	14.127	
9	18.192	18.210	18.204	28.182	14.138	14.132	14.143	14.141	28.182	276.032	28.184	28.178	14.129	14.125	14.131	14.134	
10	18.193	18.208	18.205	28.182	14.134	14.132	14.133	14.134	28.186	28.184	500.814	28.188	14.147	14.109	14.127	14.129	
11	18.194	18.206	18.204	28.182	14.137	14.133	14.182	14.139	28.182	28.182	28.186	276.305	14.149	14.159	14.132	14.158	
12	14.122	14.116	14.094	14.093	18.208	18.210	18.189	28.184	14.083	14.083	14.083	14.084	276.578	28.182	28.182	28.180	
13	14.084	14.080	14.081	14.091	18.207	18.198	18.200	28.186	14.077	14.081	14.084	14.093	28.180	276.669	28.182	28.180	
14	14.095	14.091	14.090	14.092	18.207	18.211	18.196	28.182	14.089	14.093	14.081	14.087	28.184	28.184	276.761	28.184	
15	14.093	14.087	14.091	14.090	18.207	18.183	18.204	28.180	14.090	14.099	14.086	14.152	28.180	28.180	28.184	276.852	

3.5 xHMI 带宽测试

3.5.1 测试项目介绍

该测试项目是针对 DCU 互联的测试项目，主要是针对硬件是 DCU 的 xHMI 桥接卡，以及后续的 HYSwitch 等互联设备。

目前支持该测试项目的卡只有 Z100L，像 K100、K100-AI 暂时取消了该接口。执行测试前，需要先安装好 xHMI 桥接卡，并且冲刷 vbios，使能桥接卡。

由于 Z100L 上的 XHMI 桥接卡已经停产，该测试项目作为保留项目，用于后续卡的互联测试项目。

3.5.2 测试项目运行

可以通过两种方式运行，交互方式和命令行方式。

交互方式运行：

```
./run
然后输入 5
(base) [root@bogon hyqual-v2.2.6]# ./run
Please select one of the following options:
1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Peak Performance Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
Key-in selection followed by <enter>:5
```

命令行方式运行：

```
./run 5
```

3.5.3 测试结果查看

本测试主要测试 DCU XHMI 的带宽值，并给出测试结果，如果当前 DCU 不支持 XHMI 将提示 “DCU* XHMI not support !!”。

对于已经有 XHMI 连接的设备，刷完相应的 VBIOS 后，默认卡间数据传输走的是 XHMI。测试结果会在交互界面直接给出。

```

DCU0 XHMI not support !!
DCU1 XHMI not support !!
DCU2 XHMI not support !!
DCU3 XHMI bandwidth test: PASS
DCU4 XHMI bandwidth test: PASS

```

详细的测试数据会在 `result/bandwidth/xhmi_bandwidth_*.log` 记录，相应的结果跟 PCIe 的测试内容一致，不过这部分关注的是 D2D 的性能，也就是卡间互联的性能，如下图，关注的是 DCU[8-15]->DCU[8-15] 的性能（图中演示的是 4 卡互联的桥接卡的性能）

21.786	21.803	21.788	21.804	30.681	30.681	30.681	30.681
N/A	57.521	56.697	57.733	28.014	28.012	28.015	28.013
57.521	N/A	57.735	56.931	28.009	28.009	28.029	28.009
56.697	57.735	N/A	57.505	28.015	28.006	28.010	28.016
57.733	56.931	57.505	N/A	28.002	28.013	28.013	28.026
28.014	28.009	28.015	28.002	N/A	57.483	56.872	57.936
28.012	28.009	28.006	28.013	57.483	N/A	57.744	57.188
28.015	28.029	28.010	28.013	56.872	57.744	N/A	57.715

3.6 显存带宽测试

3.6.1 测试项目介绍

此项测试是基于 BabelStream 测试构建的，是在传统 CPU stream 测试的基础上，新增了几项跟 GPU 相关测。主要是在之前四种数组运算（数组的复制-Copy、数组的尺度变换-Scale、数组的矢量求和-ADD、数组的复合矢量求和-Triad）的基础上加了数组点乘-Dot 等操作。

3.6.2 测试项目运行

可以通过两种方式运行，交互方式和命令行方式。

交互方式运行：

```
./run
```

然后输入 6

```
Please select one of the following options:
1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Peak Performance Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
key-in selection followed by <enter>:6
graphic mem bandwidth test begin
```

命令行方式运行：

```
./run 6
```

3.6.3 测试结果查看

该测试主要是基于 BabelStream 测试，关注点主要是在 Copy 测试上，要求 Copy 带宽要在理论带宽的 70%-100%以内，表示环境正常。通常显存带宽的实测结果在理论带宽的 80+%左右。

```

q: quit HyQual
Key-in selection followed by <enter>:6
graphic mem bandwidth test begin
BabelStream
Version: 3.4
Implementation: HIP
Running kernels 800 times
Precision: long
Array size: 2516.6 MB (=2.5 GB)
Total size: 7549.7 MB (=7.5 GB)
Using HIP device K100_AI
Driver: 50731921
initA:1, initB:2, initC:0, initScalar:4
Function      MBytes/sec   Min (sec)   Max         Average      Rate(0.0)
Copy          727696.817   0.00687     0.00952     0.00692
Mul           730481.251   0.00687     0.00697     0.00689
Add           661083.425   0.01135     0.01429     0.01142
Triad         662560.338   0.01134     0.01427     0.01139
Dot           189341.095   0.02607     0.02935     0.02658
Write         719709.165   0.01031     0.01308     0.01049
Read          836855.770   0.00599     0.00843     0.00601
Theoretical raw bandwidth = 896 GB/s
Min average raw bandwidth = 627.2 GB/s
DCU0 graphic mem bandwidth 727.697GB/s PASS
graphic mem bandwidth test end

```

测试结果也会保存到./result/bandwidth/graphicmem_bandwidth_*.log，显示跟交互界面一致。

3.7 硬件算力测试

3.7.1 测试项目介绍

本测试项基于通用矩阵乘（GEMM, General Matrix Multiplication）运算，根据不同精度的 GEMM 运算，来计算 DCU 的实际算力值。基本原理是通过计算固定 size 的 gemm 运算需要的计算量，以及收集完成 gemm 运算的时间，来计算 DCU 的实际工作算力。

根据不同的硬件支持的数据类型，该测试项目会测试不同的 GEMM 结果。常见的 GEMM 如下：

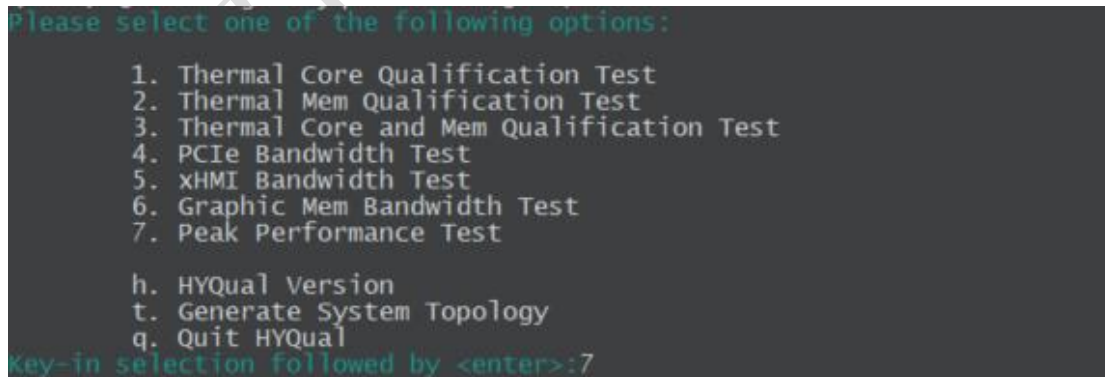
- dgemm: 双精（double）gemm 运算，数据精度 fp64
- sgemm: 单精（single）gemm 运算，数据精度 fp32
- hgemm: 半精（half）gemm 运算，数据精度 fp16
- bf16gemm: bf16（Brain Float 16）gemm 运算，数据精度 bf16
- tf32gemm: tf32(Tensor Float 32) gemm 运算，数据精度 tf32
- i8gemm: int8 gemm 运算，数据精度 int8
- u8gemm: unsigned int8 gemm 运算，数据精度 unsigned int8

3.7.2 测试项目运行

可以通过两种方式运行，交互方式和命令行方式。

交互方式运行：

```
./run
然后输入 7
```



```
Please select one of the following options:
1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Peak Performance Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
Key-in selection followed by <enter>:7
```

命令行方式运行：

```
./run 7
```

3.7.3 测试结果查看

该测试会根据硬件，自动选择相应支持的 Gemm 运算，依次进行不同类型的 gemm 运算，每一项 gemm 运算会指定所有的卡运行。所有的 gemm 运算会运行三次，取三次测试的最优结果。下图为 K100-AI 的测试内容：

```
key in selection followed by <enter>:7
DCU DEVICE NUMBER IS 1
sgemm[NT]: m, n, k, lda, ldb, ldc, stridea, strideb, stridec = 5760, 2048, 50000, 12000, 12000, 12000, 600000000, 600000000, 24576000
iterator 1:
DCU0 25.182058 ms, 46.844781 TFLOPS
iterator 2:
DCU0 25.262857 ms, 46.694955 TFLOPS
iterator 3:
DCU0 25.210218 ms, 46.792455 TFLOPS
-----
DCU0: min: 46.694955 max: 46.844781 mean: 46.777397 stdev: 0.062086
DCU DEVICE NUMBER IS 1
hgemm[NT]: m, n, k, lda, ldb, ldc, stridea, strideb, stridec = 5760, 2048, 50000, 12000, 12000, 12000, 600000000, 600000000, 24576000
iterator 1:
DCU0 6.230874 ms, 189.323037 TFLOPS
iterator 2:
DCU0 6.204154 ms, 190.138413 TFLOPS
iterator 3:
DCU0 6.200794 ms, 190.241437 TFLOPS
-----
DCU0: min: 189.323037 max: 190.241437 mean: 189.900962 stdev: 0.410814
DCU DEVICE NUMBER IS 1
bf16gemm[NT]: m, n, k, lda, ldb, ldc, stridea, strideb, stridec = 5760, 2048, 50000, 12000, 12000, 12000, 600000000, 600000000, 24576000
iterator 1:
DCU0 6.248634 ms, 188.784945 TFLOPS
iterator 2:
DCU0 6.216634 ms, 189.756714 TFLOPS
iterator 3:
DCU0 6.209274 ms, 189.981636 TFLOPS
-----
DCU0: min: 188.784945 max: 189.981636 mean: 189.507765 stdev: 0.519294
DCU DEVICE NUMBER IS 1
tf32gemm[NT]: m, n, k, lda, ldb, ldc, stridea, strideb, stridec = 5760, 2048, 50000, 12000, 12000, 12000, 600000000, 600000000, 24576000
iterator 1:
DCU0 12.399029 ms, 95.140355 TFLOPS
iterator 2:
DCU0 12.358069 ms, 95.455686 TFLOPS
iterator 3:
DCU0 12.379509 ms, 95.290371 TFLOPS
-----
DCU0: min: 95.140355 max: 95.455686 mean: 95.295471 stdev: 0.128784
DCU DEVICE NUMBER IS 1
DCU CU NUMBER IS 120
i8gemm[NT]: m, n, k, lda, ldb, ldc, stridea, strideb, stridec = 384, 320, 32, 384, 320, 384, 12288, 10240, 122880
iterator 1:
DCU0 0.000206 ms, 382.156080 TOPS
iterator 2:
DCU0 0.000207 ms, 381.452220 TOPS
iterator 3:
DCU0 0.000206 ms, 382.174039 TOPS
-----
DCU0: min: 381.452220 max: 382.174039 mean: 381.927446 stdev: 0.336116
DCU DEVICE NUMBER IS 1
DCU CU NUMBER IS 120
u8gemm[NT]: m, n, k, lda, ldb, ldc, stridea, strideb, stridec = 384, 320, 32, 384, 320, 384, 12288, 10240, 122880
iterator 1:
DCU0 0.000206 ms, 382.161687 TOPS
iterator 2:
DCU0 0.000207 ms, 381.092905 TOPS
iterator 3:
DCU0 0.000206 ms, 382.140753 TOPS
-----
DCU0: min: 381.092905 max: 382.161687 mean: 381.798448 stdev: 0.498968
-----
```

测试结果如下：算力测试结果一般都满足标称算力的 90+%

```
Theoretical peak sgemm perf = 49.15 TFLOPS
Peak min sgemm perf = 34.41 TFLOPS
Theoretical peak hgemm perf = 196.61 TFLOPS
Peak min hgemm perf = 137.63 TFLOPS
Theoretical peak bf16gemm perf = 196.61 TFLOPS
Peak min bf16gemm perf = 137.63 TFLOPS
Theoretical peak tf32gemm perf = 98.30 TFLOPS
Peak min tf32gemm perf = 68.81 TFLOPS
Theoretical peak i8gemm perf = 393.22 TOPS
Peak min i8gemm perf = 275.25 TOPS
Theoretical peak ui8gemm perf = 393.22 TOPS
Peak min ui8gemm perf = 275.25 TOPS

DCU0 Peak Performance:
peak sgemm: 46.84 TFLOPS PASS
peak hgemm: 190.24 TFLOPS PASS
peak bf16gemm: 189.98 TFLOPS PASS
peak tf32gemm: 95.46 TFLOPS PASS
peak i8gemm: 382.17 TOPS PASS
peak ui8gemm: 382.16 TOPS PASS
```

测试结果也会保存到 ./result/bandwidth/ gemmpeak_*_DCU.log，显示跟交互

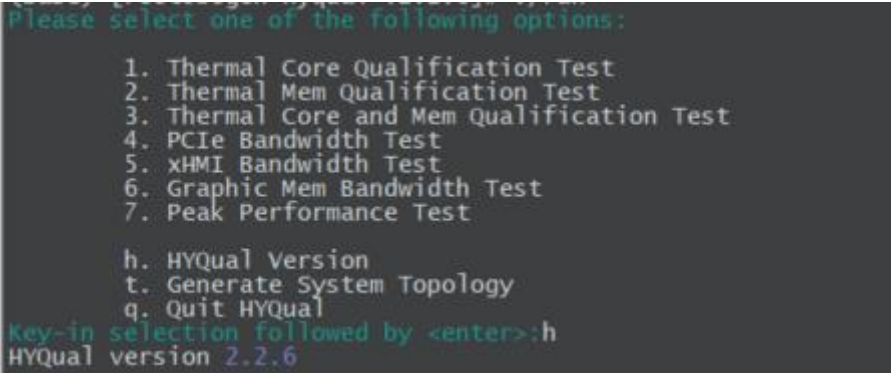
界面一致。

开文基社

3.8 其它

运行 ./HyQual,

- 输入 h 将提示 HyQual 的版本信息。
- 输入 t 将输基于 numa 的拓扑图，在 result/目录生成 topology.txt 文件。
- 输入 q 将退出 HyQual 程序。



```
Please select one of the following options:
1. Thermal Core Qualification Test
2. Thermal Mem Qualification Test
3. Thermal Core and Mem Qualification Test
4. PCIe Bandwidth Test
5. xHMI Bandwidth Test
6. Graphic Mem Bandwidth Test
7. Peak Performance Test

h. HYQual Version
t. Generate System Topology
q. Quit HYQual
Key-in selection followed by <enter>:h
HYQual version 2.2.6
```

图 21 查看版本信息

保密声明

在此声明，该文档展示的全部技术信息及其相关内容， 版权皆属于开发者社区

(<https://developer.hpccube.com/>) 所有。未经允许，
严禁截屏、大规模传播及转发。另外，对使用该技术文档
而导致任何侵犯第三方专利或其它权利的行为，开发者
社区不承担任何责任。

感谢您的理解与支持。