Nsight 性能分析使用方法说明

1. Nsight安装

在172.20.158.26服务器 /data/xfzhu5/tools/nsight-cuda/Nsight.msi，可以直接安装在虚拟云桌面上，安装完成后可能会提醒 msvcp140\_1.dll 和 vcruntime140\_1.dll找不到，在相同路径下，已经准备好，放置在C:\Windows\System32路径下即可使用，具体可以在百度搜索相关的解决办法。

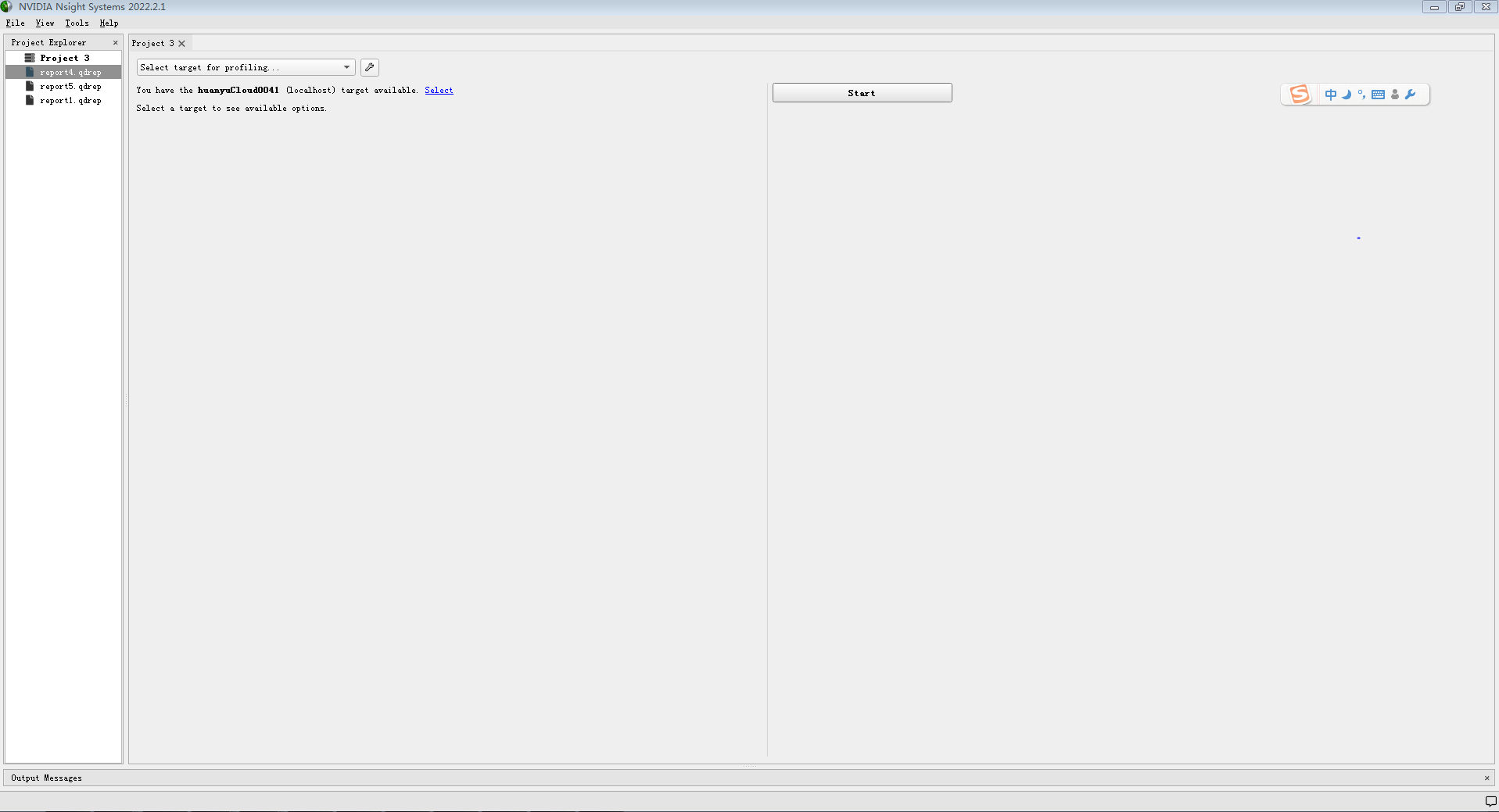


图1 安装完后打开软件的界面示意图

1. 使用Nsight分析工具

在使用Nsight分析工具之前，你需要使用cuda环境提供的套件对你的目标分析程序进行一次性运行数据的采样，采样完成后会工具会输出一份数据报表，Nsight基于这份数据报表生成可视化的数据。具体使用的方式如下：

1. 找到nsys可执行程序，一般在cuda的bin目录下，可能有些cuda版本在安装的时候没有带，所以在使用之前一定要先找到该可执行程序（举例： 172.20.158.26服务器 在/usr/local/cuda-10.2/bin 可以找到，如下图2所示）

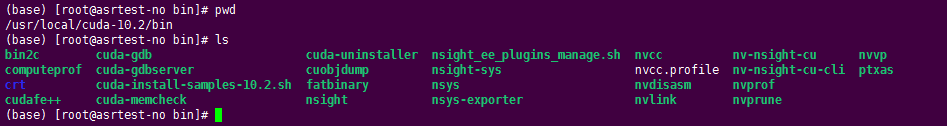


图2 nsys可执行程序

1. 运行程序（C程序和python程序均可进行分析）

例如对AI字幕的引擎进行分析，在引擎目录下运行

/usr/local/cuda-10.2/bin/nsys profile ./mtrec\_plus effect\_chinese\_debug.cfg

在运行结束后在引擎运行的目录下产生一个类似report\*.qdrep，\*代表一个数字，但是在执行的过程中，会看到后缀名字是qdstrm的文件，类似report\*.qdstrm，在程序执行结束后，会有工具自动转换。

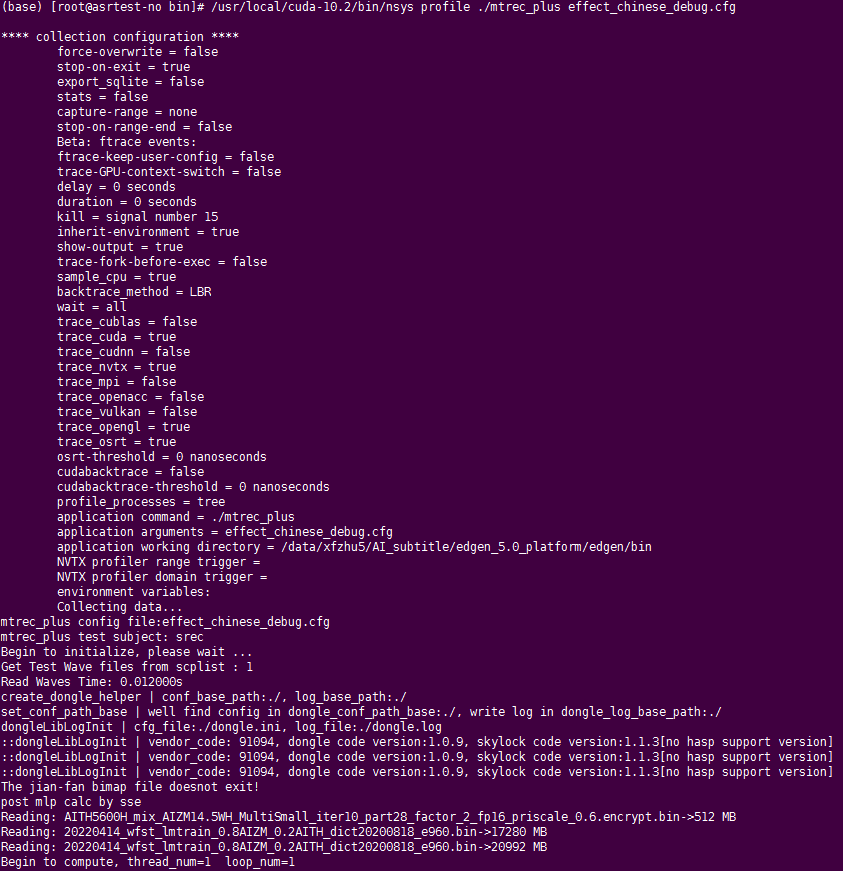


图3 nsys运行过程示意图

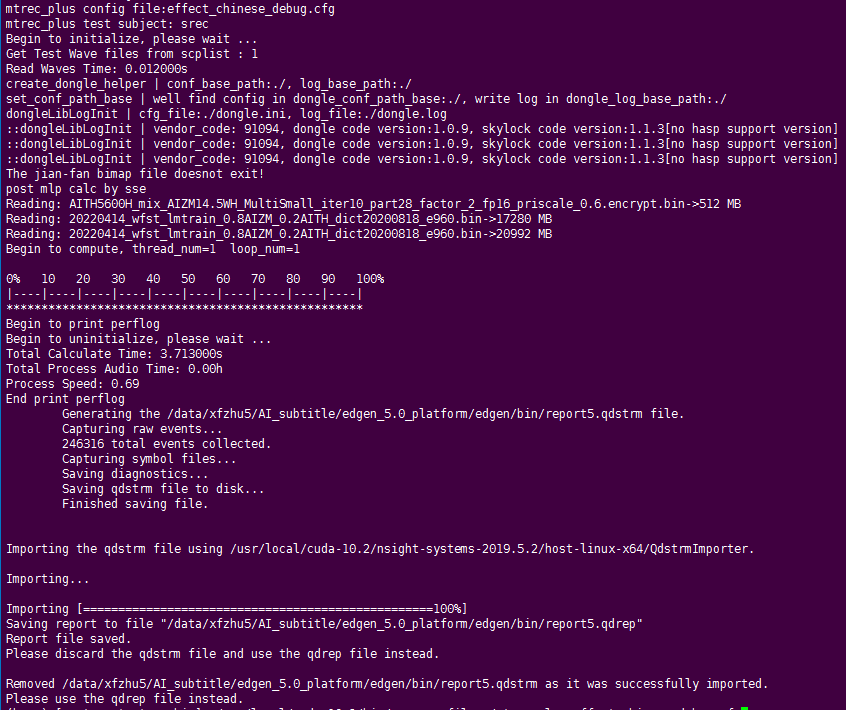


图4 nsys运行示意图

1. 使用Nsight分析数据报表
2. 使用Nsight打开性能数据文件report\*.qdrep，既可以得到如下所示图5所示界面
3. 从该分析界面中可以查看GPU stream的执行情况，包括各类执行核函数、cuda内存相关的处理函数memset memcpy等, 如图6所示。

Tips：当你将鼠标放置在核函数执行的矩形块内时，会显示该核函数的具体信息，例如grid和block的维度，执行时间，使用GPU显存，如图7所示

1. 除了GPU stream外，也可以查看cpu线程的执行情况
2. 各种的性能分析方式和Nsight提供的功能需要各位小伙伴自行深入探索

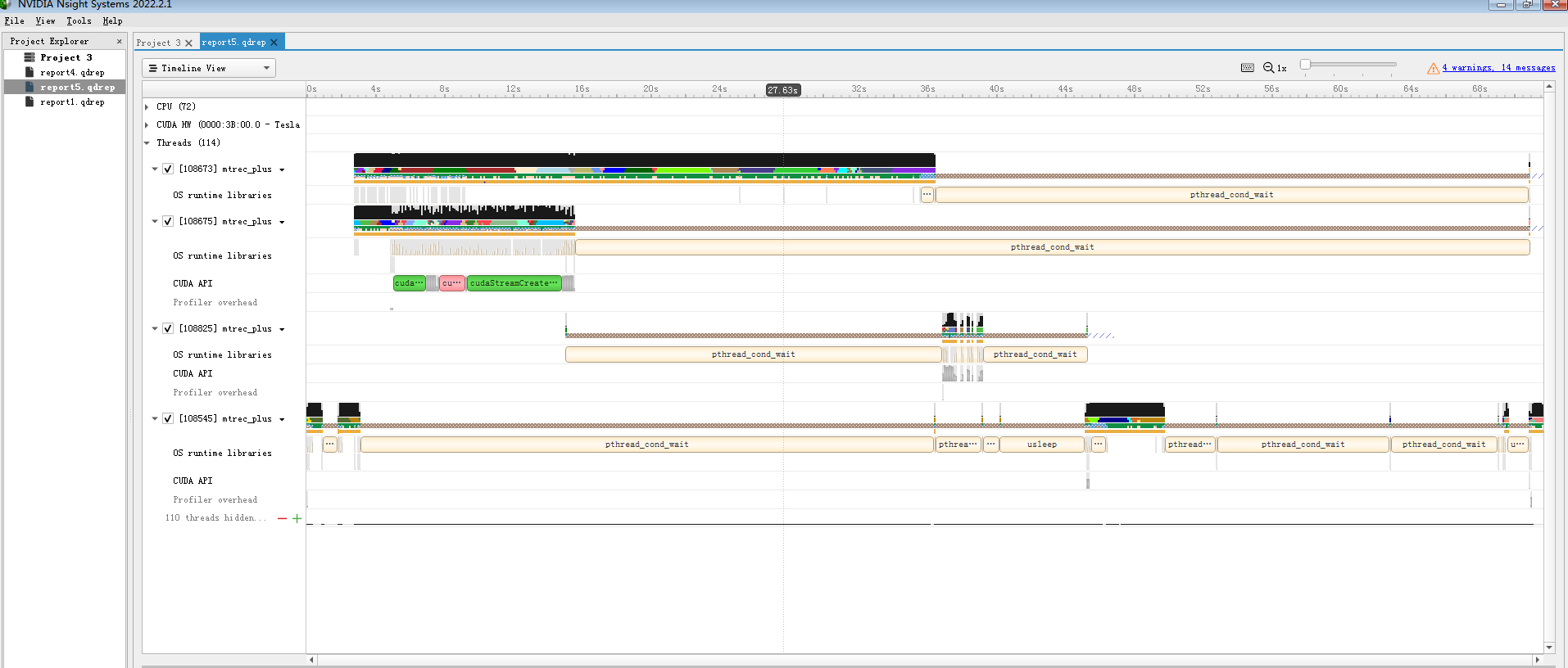


图5 nsight分析界面

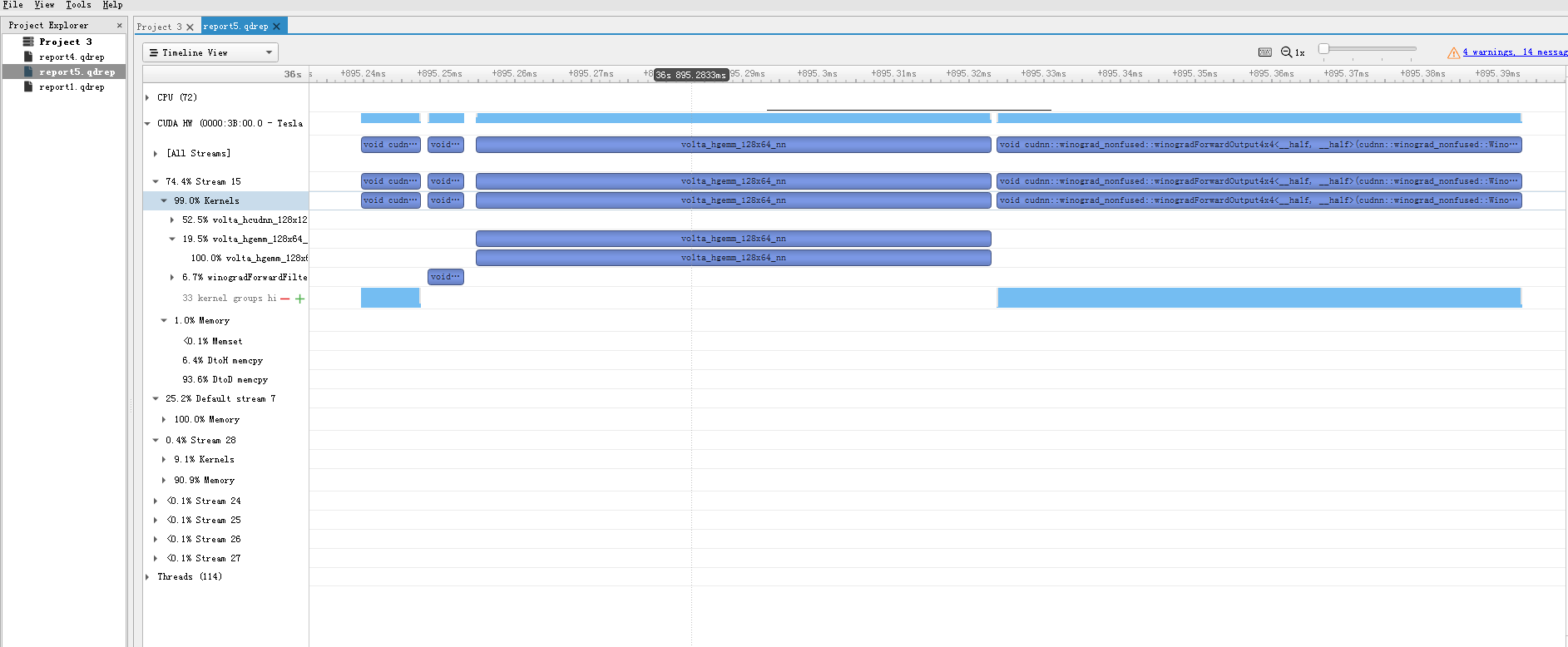


图6 Nsight显示的stream执行的核函数情况

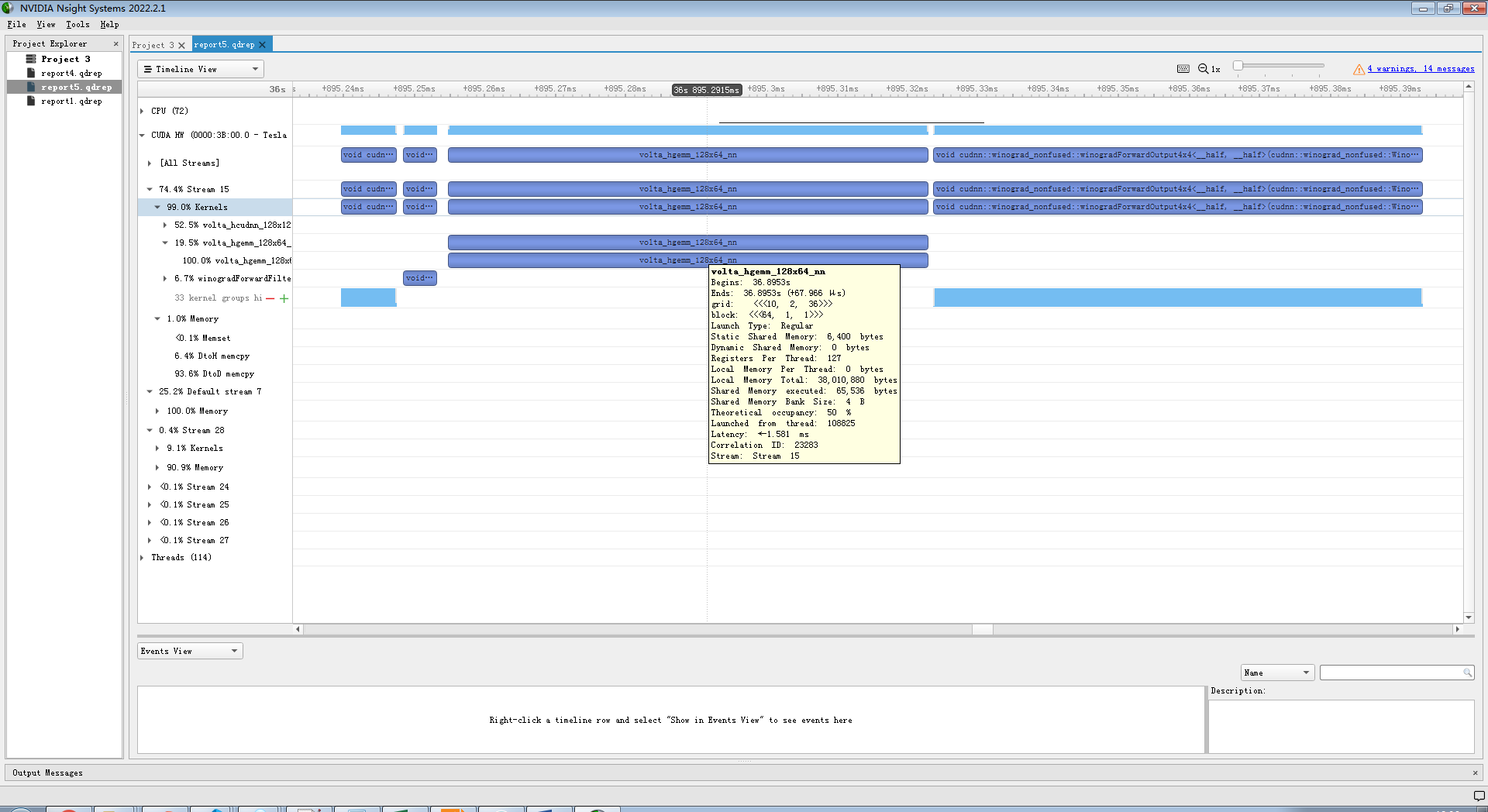


图7 kernal核函数信息

欢迎交流学习（联系邮箱 xfzhu5@iflytek.com）