GPU性能分析

聚焦于GPU计算

如何评测

| GPU调试分析 |

| CPU/GPU 追踪 |

| 应用程序追踪 |

GPU利用率低

| SM效率低 | Achieved Occupancy低 | 显存瓶颈 | 指令瓶颈 |

- GPU线程太少
- 寄存器限制
- Shared memory太大
- ...

- Cache misses
- 带宽限制
- 访存类型
- ...

- 算输指令
- 控制流
- ...

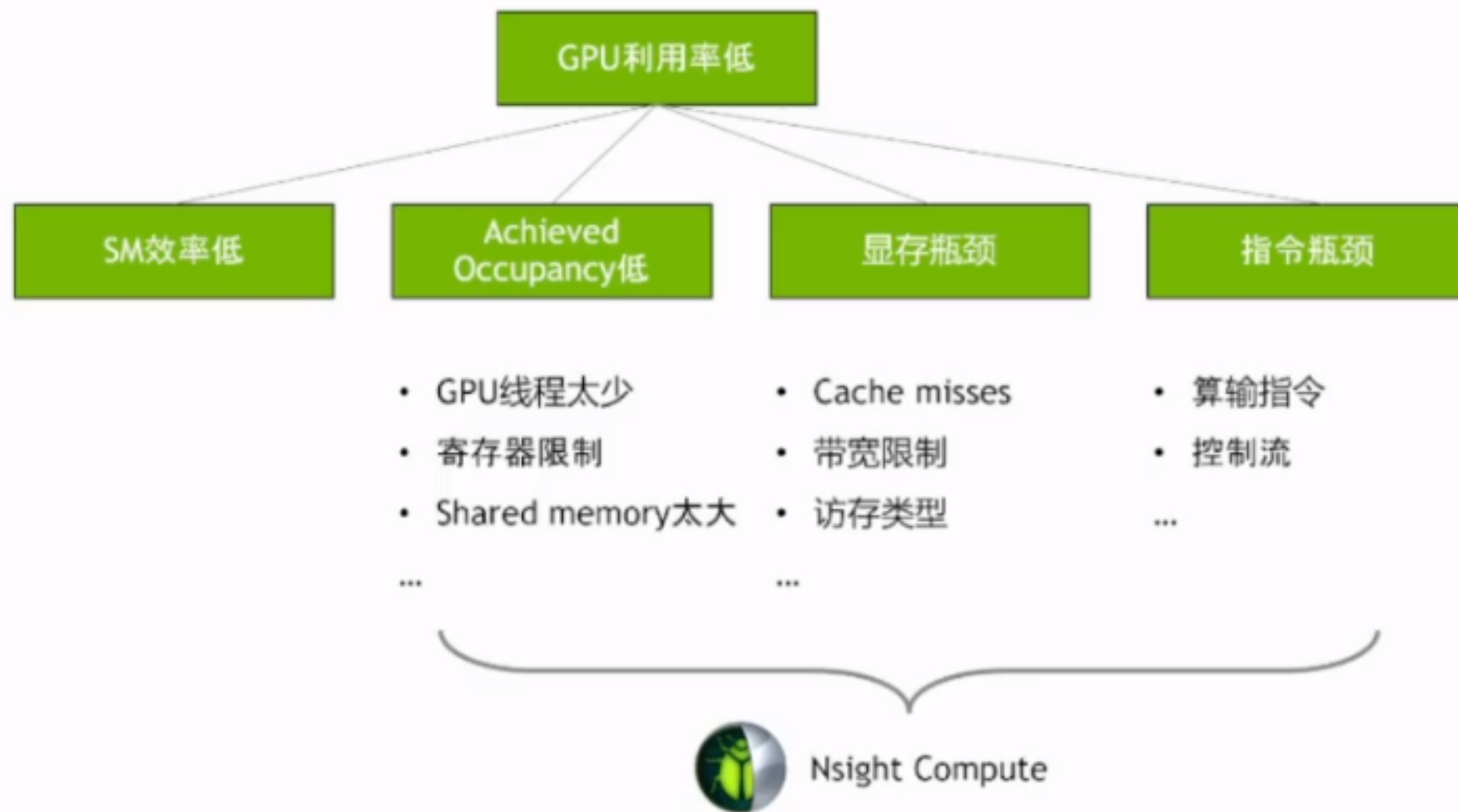Nsight Compute

Nsight System

系统级调试器

- 核心特点：
  - 系统级应用算法调试
    - 支持多进程树
  - 找到优化机会
    - 以GUI时间线的方式可视化数百万个事件(event)
    - 未使用的CPU和GPU事件间隔显示
  - 在多CPU和GPU之间平衡工作负载
    - CPU算法，利用率，线程状态
    - GPU stream，kernel，内显存传输等
  - 命令行、独立方式或IDE集成

OS：Linux (x86, Power, Arm SBSA, Tegra), Windows, MacOSX(host)

GPUs: Pascal架构及以上

文档/产品页：https://developer.nvidia.com/nsight-systems

Nsight Systems 界面展示

支持追踪的事件、输出格式、平台架构

Trace:
- TensorRT
- Direct3D11,12,DXR
- Vulkan
- OpenGL
- OpenACC
- MPI
- OpenMP
- Ftrace
- ETW
- WDDM
- GPU Context Switch

输出:
- SQLite
- HDF5
- JSON

架构:
- X86_64
- Power
- Arm SBSA
- Tegra