GPU

(Graphics Processing Unit)


CPU
功能
建模
转换
光照
裁剪
设置
2D投影
光栅化
显示
图形控制器
功能

The First Generation(-1999)

GE(Geometry Engine)

The Second Generation(1999-2002)

GeForce 256(1999)

T&L

(Transform and lighting)

GeForce 3(2001)

The Third Generation(2002-)

Using GPU more than videogames and scientific research

Mobile applications rely on GPUs running servers in the cloud.

Stores use GPUs to analyze retail and web data.

Web sites use GPUs to more accurately place ads.

Engineers rely on them in computer-aided engineering applications.

Accelerated computing using GPUs continues to expand at an astounding rate.

CUDA

(Computer Unified Device Architecture)

Standard C Code

```
void saxpy(int n, float a,
           float *x, float *y)
{
  for (int i = 0; i < n; ++i)
    y[i] = a*x[i] + y[i];
}

int N = 1<<20;



// Perform SAXPY on 1M elements
saxpy(N, 2.0, x, y);
```

C with CUDA extensions

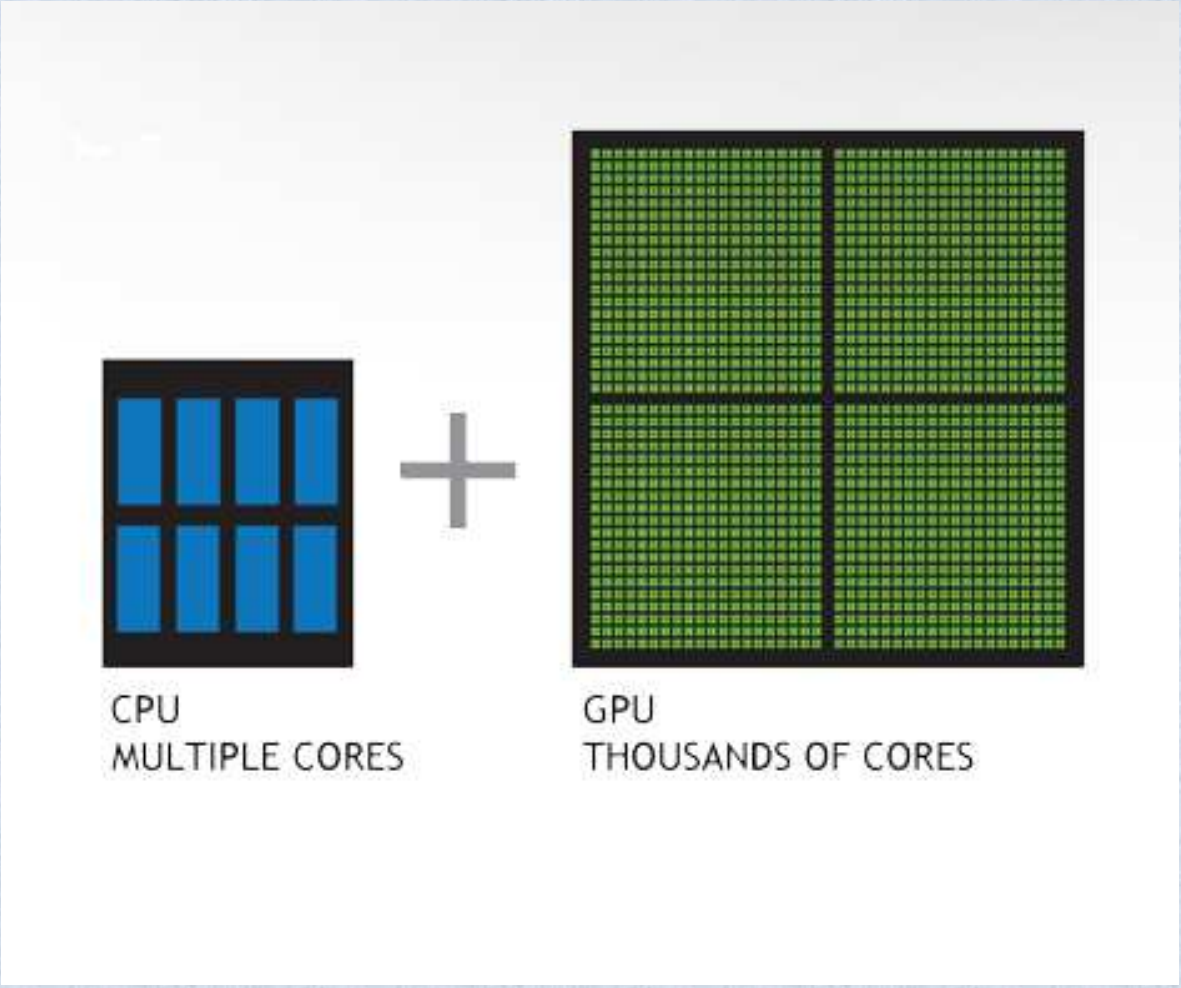
```
__global__
void saxpy(int n, float a,
           float *x, float *y)
{
  int i = blockIdx.x*blockDim.x + threadIdx.x;
  if (i < n) y[i] = a*x[i] + y[i];
}

int N = 1<<20;
cudaMemcpy(x, d_x, N, cudaMemcpyHostToDevice);
cudaMemcpy(y, d_y, N, cudaMemcpyHostToDevice);

// Perform SAXPY on 1M elements
saxpy<<<4096,256>>>(N, 2.0, x, y);

cudaMemcpy(d_y, y, N, cudaMemcpyDeviceToHost);
```

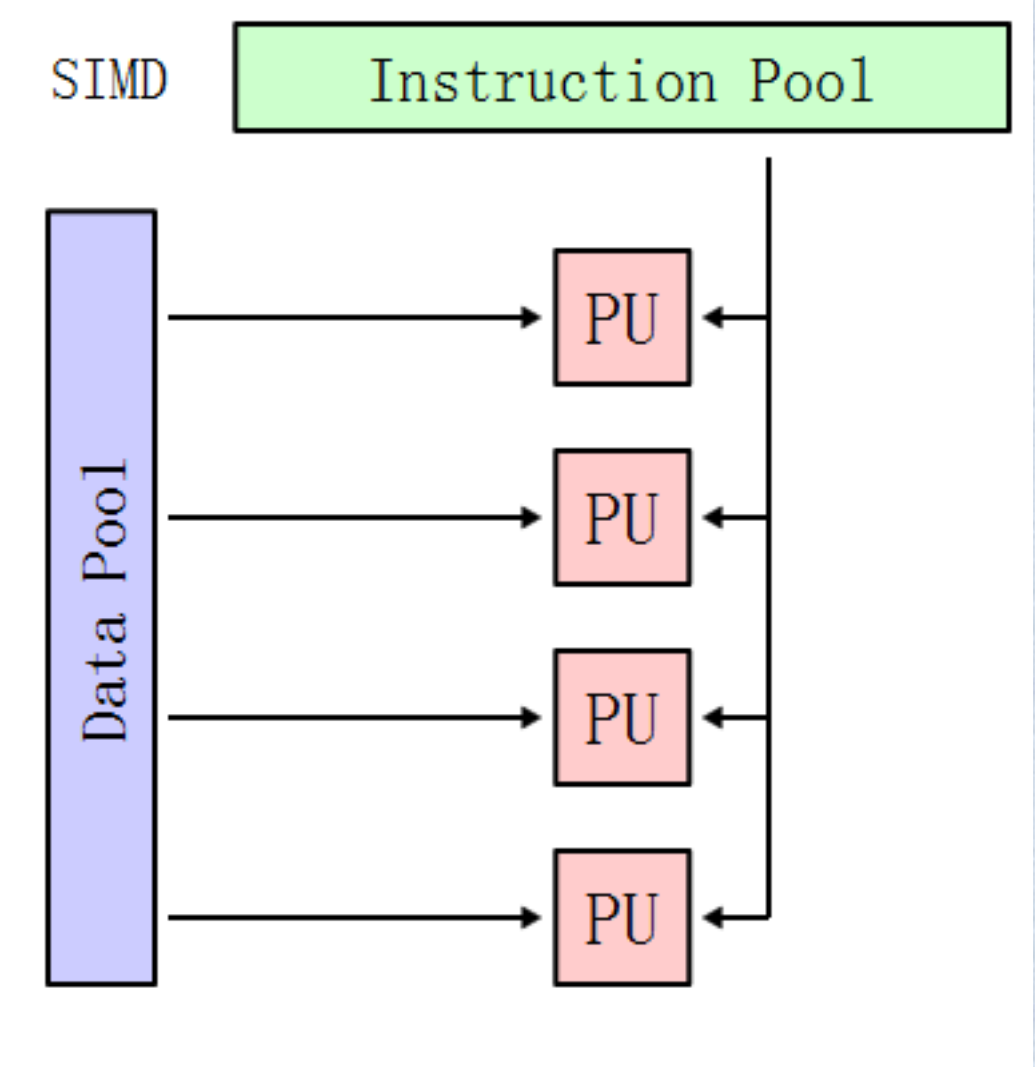
GUP computing



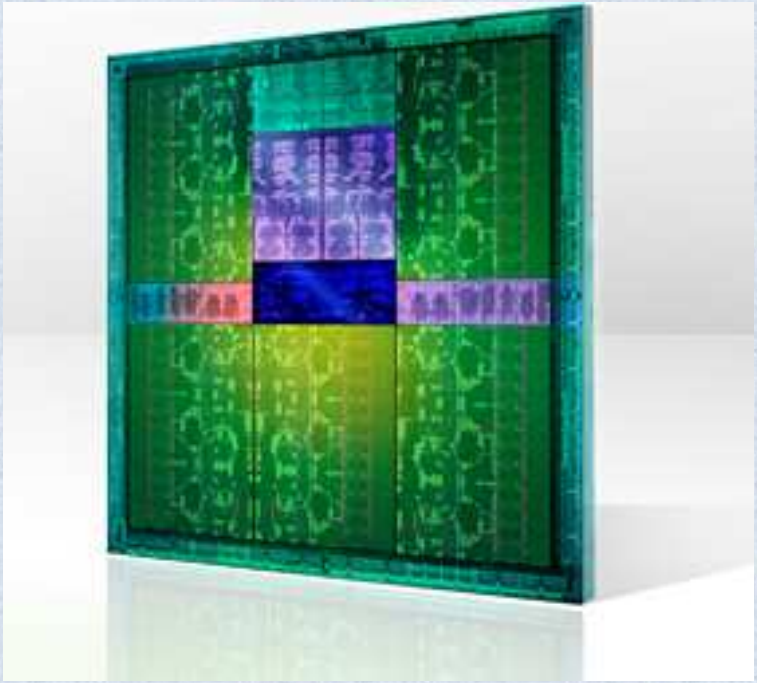
GPGPU

(General Purpose Graphics Processing Unit)

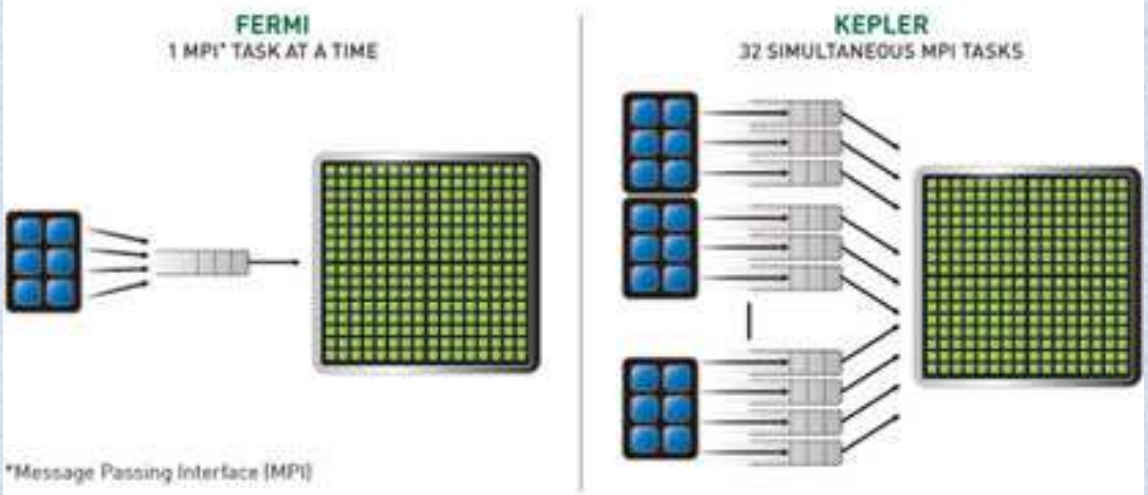
Single Instruction Multiple Data



Kepler GPU

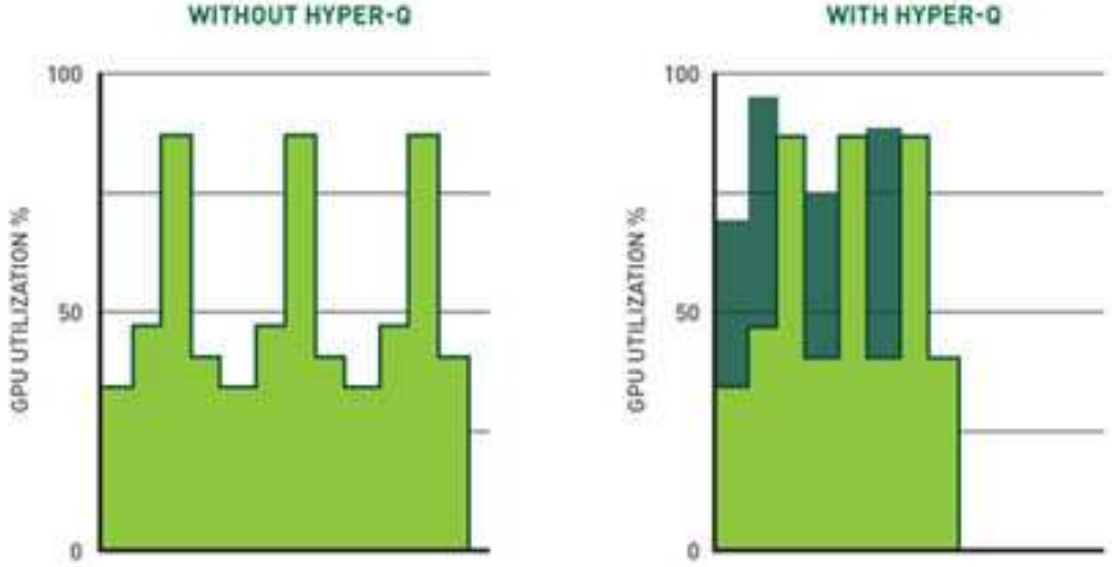


NVIDIA HYPER-Q



*Message Passing Interface (MPI)

Kepler*s Hyper-Q increases GPU utilization by providing streams access to 32 independent hardware work queues or MPI ranks leading to advanced programmability and efficiency.



Hyper-Q enables multiple CPU cores to launch work on a single GPU simultaneously, thereby dramatically increasing GPU utilization and slashing CPU idle times.

THE END