

CUDA	CUDA (cont)
Kernel execution	<code>kernelName << <num_blocks, num_threads, sharedMemPerBlock>>> To Symbol (dst, src, sizeIn Bytes) > (args);</code>
Kernel definition	<code>__global__ kernelName(args){...}</code>
Create memory on device	<code>cudaMalloc(memPtr, sizeIn Bytes);</code>
Free device memory	<code>cudaFree(memPtr);</code>
Create pinned memory on host	<code>cudaMallocHost(memPtr, size_in_bytes);</code>
Free pinnen memory	<code>cudaFreeHost(memPtr);</code>
Create stream	<code>cudaStreamCreate(&stream);</code>
Destroy stream	<code>cudaStreamDestroy(stream);</code>
Wait for Device	<code>cudaDeviceSynchronize();</code>
Wait for Stream	<code>cudaStreamSynchronize(stream);</code>
Copy data	<code>cudaMemcpy(dst, src, numElements, cudaMemcpyDefault);</code>
Async mem copy	<code>cudaMemcpyAsync(dst, src, numElements, cudaMemcpyDefault, stream);</code>
Static device variable	<code>__device__ typeName = value;</code>
Static shared mem	<code>__shared__ typeName[size];</code>
Dynamic shared mem	<code>extern __shared__ typeName;</code>
Constant device mem	<code>__constant__ typeName;</code>
	Thrust Universal Vector <code>thrust::universal_vector<type> name(size);</code>

