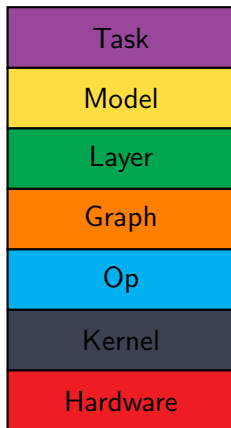


```
>>> c++ deep-learning.cpp
```

```
>>> ./a.out data.csv
```

```
>>> 42
```



## OSI Model of Machine Learning

# Hardware: GPU



**NVIDIA Stock**



**Deep Learning Trend**

# Hardware: GPU

## Hardware: GPU

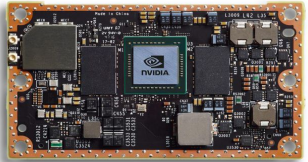


**Titan X**

## Hardware: GPU



**Titan X**

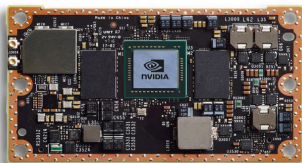


**Jetson TX2**

# Hardware: GPU



**Titan X**



**Jetson TX2**



**DGX-1**

# Hardware: Big Basin

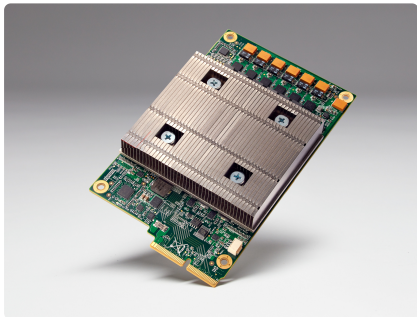


**Big Basin**

- ▶ 8 NVIDIA Tesla P100 GPUs
- ▶ NVLink (12x faster than PCIe)
- ▶ 16 GB RAM
- ▶ 10.6 TFLOPS
- ▶ Reduced Precision Computing



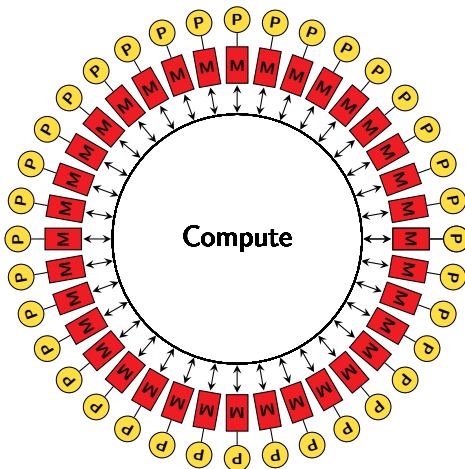
## Hardware: TPU



**TPU**

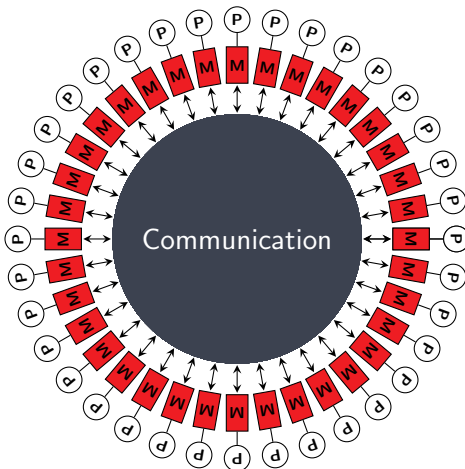
- ▶ *Coprocessor*
- ▶ 92 TOPS for 8-bit int
- ▶ 42 TFLOPS (TPU2)
- ▶ 24 MB Memory

## Hardware: IPU



### **Bulk Synchronous Parallelism**

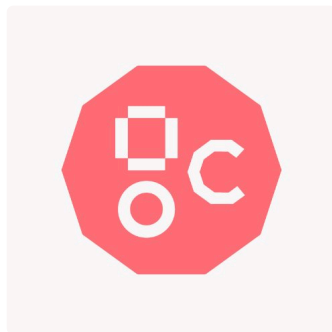
## Hardware: IPU



### **Bulk Synchronous Parallelism**

## Hardware: IPU

- ▶ Startup from Bristol (UK)
- ▶ Graphcore Colossus
- ▶ TBR later this year
- ▶ 1000 Processors/Chip
- ▶ Mixed Precision Arithmetic
- ▶ No DRAM
- ▶ “Thrives on Sparsity”



**Graphcore**