Built for Big Data: Accelerated Deep Learning with POWER

Ruchir Puri
IBM Fellow
**POWER8 Core:** Back bone of big data computing system

- Enhanced Micro-Architecture
- Increased Execution Bandwidth
- SMT 8
- Transactional Memory

- Vector/Scalar Unit
  - High-performance Integer & FP Vector Processor
  - Optimized for Data Rich Applications
Putting it all together with the memory links, on- and off-node SMP links as well as PCIe, at 7.6Tb/s of chip I/O bandwidth
General-Purpose CPU Design

- Many competing requirements
  - Branchy control-flow dominated code
  - Code with unpredictable data access patterns
  - Operating system code
  - Multiple separate applications
  - Multiple virtual machines concurrently

- Result in low efficiency for any one metric
  - Flops / area
  - Integer ops / area
  - …
Power GPU Acceleration

- Support for up to 18 GPUs or more
- Tested with up to 24 devices
- Exploit IBM design for big data
- Large address space enables rich accelerator configurations
- 1 TB of address space per PCI host interfaces
- Standard LE Linux Nvidia drivers
- CUDA 7.5
- Available now
Power Machine and Deep Learning Software Stack

Frameworks

CAFFE (Berkeley)
Torch (NYU, Facebook)
Theano (Montreal)
TensorFlow (Google)

CNTK (Microsoft)
DL4J (Spark)

Library layer

DLSL

BLAS, LAPACK
cuBLAS, cuDNN
Nervana Auviz lib
Xilinx blks
Power GPU acceleration

- CUDA programming environment supported under LE Linux
  - GPU as compute accelerator
  - Offload dominated compute-intensive application portions to GPU

- Advances in GPU Performance and Programmability
  - UVA – Universal Virtual Addressing
  - UM – Unified Memory

- Ongoing collaboration to co-optimize systems
  - Next generation hardware enhancements
Programming Heterogeneous Systems

C++?
Java?
OpenCL?
SystemC
?VHDL?
CUDA?
Portability and Optimization in Heterogeneous Systems

Application

Cognitive Middleware

Application

Application

Library Layer

CPU enablement

GPU enablement

FPGA interface & configuration

Accelerator X Enablement

Accelerator X
First Power Machine and Deep Learning Distro available

<table>
<thead>
<tr>
<th>DL Framework</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAFFE</td>
<td>Ported. Included in First Release of Power MLDL Distro.</td>
</tr>
<tr>
<td>Torch</td>
<td>Ported. Included in First Release of Power MLDL Distro.</td>
</tr>
<tr>
<td>Theano</td>
<td>Ported. Included in First Release of Power MLDL Distro.</td>
</tr>
<tr>
<td>DIGITS-3</td>
<td>Ported. Included in First Release of Power MLDL Distro.</td>
</tr>
<tr>
<td>TensorFlow</td>
<td>Ported.</td>
</tr>
<tr>
<td>DL4J</td>
<td>Ported.</td>
</tr>
<tr>
<td>CNTK</td>
<td>Ported.</td>
</tr>
</tbody>
</table>
Accelerated Deep Learning on Power

Designed for Compute-Intensive Cognitive Workloads
Personalized Medicine, Adverse Drug Reaction Prediction w/ ML

100s of TBs of data

Ingest

Chemical Similarity 1 to ...

Drug1 Drug2 Sim
Salsalate Aspirin .9
Dicoumarol Warfarin .76
Salsalate Aspirin .7
Dicoumarol Warfarin .6

Known Interactions of type 1 to ...

Drug1 Drug2
Aspirin Gliclazide
Aspirin Dicoumarol
Aspirin Probenecid
Aspirin Azilsartan

Candidate Interactions of type i

Features Drug2 Best Sim1*Sim1 Best SimN*SimN
Salsalate Gliclazide .9*.1 .7*.1
Salsalate Warfarin .9*.76 .7*.6

50 million patients, 2000 drugs, 2000 features

Learn

Machine Learning Model

Predict

Interactions of type 1 Prediction

Drug1 Drug2 Prediction
Salsalate Gliclazide 0.85
Salsalate Warfarin 0.7

Interactions of type M Prediction

Drug1 Drug2 Prediction
Salsalate Gliclazide 0.53
Salsalate Warfarin 0.32

30X improvement in Learning performance
Personalized Medicine – Adverse Drug Reaction Workload

Personalization will result in massive increase in computation complexity

Real time prediction requirements for operational needs (< 1 minute for emergency situations)

- Computational pattern:
  - Sparse cube to dense cube with patient as additional dimension

- Training:
  - Number of patients above 50 Million
  - Number of features around 1800
  - Additional samples for training $O(\#\text{patients})$
  - Number of cross-validation stages and #models per stage increases dramatically
  - **100X increase in training complexity with ~100 TBs of Data**

- Prediction:
  - Input Model (#features) and dataset (# patients in the hospital)
Thank You!