Building Innovative Hardware in an Era of Artificial Intelligence

Igor Carron, CEO and Co-Founder

LightOn.ai
Why Build Innovative Hardware?
Fig. A8. Total energy of computing.

Source: Rebooting the IT Revolution: A Call to Action, SIA, SRC, 2015
Source: Mike P. Frank, SNL
It’s worse!
crypto rush in one image

Source: Twitter, Coinbase
Microsoft boss: World needs more computing power

By Joe Miller
BBC News, Davos

03 January 2018

The world is rapidly "running out of computing capacity", the head of tech giant Microsoft has warned.

Source: BBC
It’s much worse!
The Footprint of Machine Learning

CO2 eq.

- Airplane: 1,984
- Car: 126,000
- Machine Learning: 626,155

Chart: MIT Technology Review • Source: Strubell et al.
NeurIPS Conference
@NeurIPSConf

#NIPS2018 The main conference sold out in 11 minutes 38 seconds

9:17 AM - 4 Sep 2018

687 Retweets 1,079 Likes
NeurIPS Conference
@NeurIPSConf

#NIPS2018 The main conference sold out in 11 minutes 38 seconds

9:17 AM - 4 Sep 2018

687 Retweets 1,079 Likes
Scaling up a new technology from the ground up....fast
Technology Readiness Levels

**TRL 0:** Idea. Unproven concept, no testing has been performed.

**TRL 1:** Basic research. Principles postulated and observed but no experimental proof available.

**TRL 2:** Technology formulation. Concept and application have been formulated.

**TRL 3:** Applied research. First laboratory tests completed; proof of concept.

**TRL 4:** Small scale prototype built in a laboratory environment ("ugly" prototype).

**TRL 5:** Large scale prototype tested in intended environment.

**TRL 6:** Prototype system tested in intended environment close to expected performance.

**TRL 7:** Demonstration system operating in operational environment at pre-commercial scale.

**TRL 8:** First of a kind commercial system. Manufacturing issues solved.

**TRL 9:** Full commercial application, technology available for consumers.
Unlike NASA/FAA, we don’t have 15+ years
Surfing on Moore’s Law

Source: TI, Huawei
Surfing Moore’s Law and Using Nature
A Fast and Large Random Projection

Discrete input vector

\[
\begin{pmatrix}
  h_{1,1} & h_{1,2} & \ldots & h_{1,N} \\
  h_{2,1} & h_{2,2} & \ldots & h_{2,N} \\
  \vdots & \vdots & \ddots & \vdots \\
  h_{M,1} & h_{M,2} & \ldots & h_{M,N}
\end{pmatrix}
\]

Random transmission matrix

\[
x \cdot H
\]

Discrete output vector (speckle intensity)

\[
y = |H \cdot x|^2
\]

EXTRA-LARGE
H of size higher than
\(10^6 \times 10^6\)
(TBs of memory)

&

SUPER-FAST
kHz operation
\(\rightarrow 10^3\) such
multiplies / s
How it works

"Big & Fat" Data

Results

01110010101010...
Data in

010001...
Data out

Data processing
CPU / GPU

Laser

Camera

Spatial Light Modulator

DAC

ADC

Data pre-processing

OPU

"Big & Fat" Data

Copyright LightOn
Climbing up the TRL scale fast

Fast prototyping among a diverse set of expertise

TRL 3-4 (2016)  
TRL 5 (2017)  
TRL 6-7 (2018-2019)
Optical Processing Unit
Johnson-Lindenstrauss Lemma

Lemma For any \(0 < \epsilon < 1\) and any positive integer \(n\) let \(k\) be a positive integer such that

\[
k \geq \frac{24}{3\epsilon^2 - 2\epsilon^3} \log n
\]

then for any set \(A\) of \(n\) points \(\in \mathbb{R}^d\) there exists a map \(f : \mathbb{R}^d \rightarrow \mathbb{R}^k\) such that for all \(x_i, x_j \in A\)

\[
(1 - \epsilon)||x_i - x_j||^2 \leq ||f(x_i) - f(x_j)||^2 \leq (1 + \epsilon)||x_i - x_j||^2
\]
OPU: No need for large memory
n_input = 1e6;
n_output = 1e6;
x = rand(n_input,1) > 0.5;
C = randn(n_output,n_input) + 1i*randn(n_output,n_input);
y = C*x;
figure; plot(abs(y).^2);

Error using randn
Requested 1000000x1000000 (7450.6GB) array exceeds maximum array size preference. Creation of arrays greater than this limit may take a long time and cause MATLAB to become unresponsive. See array size limit or preference panel for more information.

f >> |
On a typical Machine Learning training task (transfer learning), the combination of NVIDIA and LightOn shows an x8 increase in speed compared to NVIDIA alone. Additionally, there is a 90% reduction in power consumption when using LightOn in conjunction with NVIDIA.
Other Uses Already Investigated

- Image classification
- Recommender systems
- Graph analysis
- Anomaly detection
- NLP
- Video classification
Recommender Systems shape our lives at scale!
Randomized Matrix Decompositions using R, Aug 2016, N. Benjamin Erichson, Sergey Voronin, Steven L. Brunton, J. Nathan Kutz

Copyright LightOn