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.
It’s worse!
blame it on the FWAlcohol
@nullhound

crypto rush in one image

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

By Joe Miller
BBC News, Davos

23 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
  - Data Center: 626,155

Chart: MIT Technology Review • Source: Strubell et al.
#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

<table>
<thead>
<tr>
<th>TRL 0: Idea</th>
<th>Unproven concept, no testing has been performed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRL 1: Basic Research</td>
<td>Principles postulated and observed but no experimental proof available.</td>
</tr>
<tr>
<td>TRL 2: Technology Formulation</td>
<td>Concept and application have been formulated.</td>
</tr>
<tr>
<td>TRL 3: Applied Research</td>
<td>First laboratory tests completed; proof of concept.</td>
</tr>
<tr>
<td>TRL 4: Small Scale Prototype</td>
<td>Built in a laboratory environment (&quot;ugly&quot; prototype).</td>
</tr>
<tr>
<td>TRL 5: Large Scale Prototype</td>
<td>Tested in intended environment.</td>
</tr>
<tr>
<td>TRL 6: Prototype System</td>
<td>Tested in intended environment close to expected performance.</td>
</tr>
<tr>
<td>TRL 7: Demonstration System</td>
<td>Operating in operational environment at pre-commercial scale.</td>
</tr>
<tr>
<td>TRL 8: First of a Kind Commercial System</td>
<td>Manufacturing issues solved.</td>
</tr>
<tr>
<td>TRL 9: Full Commercial Application</td>
<td>Technology available for consumers.</td>
</tr>
</tbody>
</table>
Unlike NASA/FAA, we don’t have 15+ years

**Technology Readiness Level (TRL) Process**

NASA’s quest to make jet engines quieter led to the development of chevrons, which moved relatively quickly through the TRL process to be deployed into the commercial marketplace.

**TRL 1-2 (1980s)**
- Fundamental investigations of calming devices (chevrons, etc.).
- No specific application, basic research in fluid physics.

**TRL 3 (Early 1990s)**
- Hardware to small-scale test articles.
- Lab tests, concept on paper.

**TRL 4-5 (1995-1997)**
- Model tests for acoustics and aerodynamics.
- Sub-scale model tests.

- Full-scale tests for acoustics and aerodynamics.
- Static engine tests.

**TRL 7 (2001-2005)**
- Validation of concept in flight.
- Flight tests, final design.

**TRL 8-9 (2005-now)**
- Certification by the Federal Aviation Administration.
- Deployed into market.

*Source: NASA*
Surfing on Moore’s Law

Source: TI, Huawei
# A Fast and Large Random Projection

A discrete linear system with a discrete input vector and a discrete output vector (speckle intensity) can be represented as:

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

Where \( H \) is a random transmission matrix.

**Random transmission matrix**

- **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

Data pre-processing

Laser

Spatial Light Modulator

Camera

Data in

01110101010...

Data processing

CPU / GPU

Data out

01001...

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
Why Random Projections?

Johnson-Lindenstrauss Lemma

Lemma For any $0 < \epsilon < 1$ and any 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 $x_i, x_j \in A$

$$(1 - \epsilon)||x_i - x_j||^2 \leq ||f(x_i) - f(x_j)||^2 \leq (1 + \epsilon)||x_i - x_j||^2$$

Source: Volkan Cevher, EPFL
OPU: No need for large memory

```matlab
>> n_input = 1e2;
n_output = 1e2;
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);
```
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.

>>
On a typical Machine Learning training task (transfer learning), the speed is increased by a factor of 8, and the power consumption is reduced by 90%.
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
LightOn.ai