Reducing energy and increasing capacity – new directions for integrated optics in handling information

David Miller, Stanford University

For a copy of these slides, please e-mail dabm@ee.stanford.edu

Two new reasons for complex silicon photonics

- Saving energy by eliminating the circuits in interconnect links
  - quantum impedance conversion
  - avoiding time-multiplexing in short links
    10–100 fJ/bit total energy instead of 1–10 pJ/bit
- New architectures and algorithms, allow
  - complex circuits to be set up and stabilized
  - self-correcting circuits
  - self-designing circuits

Potential applications in
  optical communications sensor preprocessing, arbitrary linear transforms, neural networks, linear optical quantum circuits, ...
Both Internet traffic and general purpose computing hardware have grown ~ 60% per year, ~ X 100 in 10 years. Massive challenge for hardware scaling of Energy.

Energy per bit has to reduce. Energy scaling not environmentally sustainable. ~ 4.6 – 9% of electricity in 2012 (Van Heddeghem et al., Computer Comm. 50 64–76 (2014)).

Communication density inside systems already at limits for electrical approaches.

---


MIPS – million instructions per second
~ 3 - 6 instructions = 1 floating point operation (FLOP)
<table>
<thead>
<tr>
<th>Operation</th>
<th>Energy per bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wireless data</td>
<td>10 – 30μJ</td>
</tr>
<tr>
<td>Internet: access</td>
<td>40 – 80nJ</td>
</tr>
<tr>
<td>Internet: routing</td>
<td>20nJ</td>
</tr>
<tr>
<td>Internet: optical WDM links</td>
<td>3nJ</td>
</tr>
<tr>
<td>Reading DRAM</td>
<td>5pJ</td>
</tr>
<tr>
<td>Communicating off chip</td>
<td>1 – 20 pJ</td>
</tr>
<tr>
<td>Data link multiplexing and timing circuits</td>
<td>~ 2 pJ</td>
</tr>
<tr>
<td>Communicating across chip</td>
<td>600 fJ</td>
</tr>
<tr>
<td>Floating point operation</td>
<td>100fJ</td>
</tr>
<tr>
<td>Energy in DRAM cell</td>
<td>10fJ</td>
</tr>
<tr>
<td>Switching CMOS gate</td>
<td>~50aJ – 3fJ</td>
</tr>
<tr>
<td>1 electron at 1V, or 1 photon @1eV</td>
<td>0.16aJ (160zJ)</td>
</tr>
</tbody>
</table>

most energy is used for communications, not logic
Data rates at different length scales

Total long distance internet traffic ~ 280 Tb/s (Cisco)
   Equivalent to everyone talking on the phone at once all the time
Traffic on “rack to rack” network inside one large data center
   ~ 1 Pb/s (Google)
Graphics processor and server chips peak bandwidth on and off chip
   ~ 1.4 Tb/s – 2 Tb/s
Server processor chip on-chip bandwidths
   on-chip network bandwidth ~ 4 Tb/s
   bandwidth in and out of L3 cache ~ 12.8 Tb/s
Energy and information

Though it does take more energy to send a bit over longer distances

there is massively more information sent at shorter distances

so much so that

most energy dissipation is in shorter links and in interconnects inside machines
Capacitance of small structures for fJ operation

<table>
<thead>
<tr>
<th>Structure</th>
<th>Capacitance</th>
</tr>
</thead>
<tbody>
<tr>
<td>100×100μm square conventional photodetector</td>
<td>~1pF</td>
</tr>
<tr>
<td>5×5μm CMOS photodetector</td>
<td>4fF</td>
</tr>
<tr>
<td>Wire capacitance, per μm</td>
<td>~200aF</td>
</tr>
<tr>
<td>FinFET input capacitance</td>
<td>~20 – 200 aF</td>
</tr>
<tr>
<td>1 micron cube of semiconductor</td>
<td>~100aF</td>
</tr>
<tr>
<td>100 nm cube of semiconductor</td>
<td>~10aF</td>
</tr>
<tr>
<td>10 nm cube of semiconductor</td>
<td>~1aF</td>
</tr>
</tbody>
</table>

So that capacitive charging energies do not dominate, we need small devices for low device capacitance very close integration to limit wiring capacitance.
Power dissipation in electrical wires

Wires always have large capacitance per unit length
~ 2 pF/cm, 200 aF/micron

Simple logic-level signaling results in large dissipation
Dissipate at least ~ ¼CV^2 per bit in on-off signaling
E.g., at 2pF/cm and a 2 cm chip, at 1 V on-off signaling
energy per bit communicated at least ~ 1pJ

electrical connection

low impedance and/or
high capacitance / unit length

small, high-impedance
devices
Logic and wiring capacitance

Wiring capacitance even to neighboring gates is comparable to or greater than the transistor capacitance.

Most energy in information processing is in communications, not in logic, even at the gate level.

Most energy dissipation in information processing is in charging and discharging wire capacitance, which is ~200 aF/micron.

Just "touching" a bit typically costs many fJ in CMOS.
Energy and information

The dominant energy dissipation at short distances inside machines

is charging and discharging wire capacitance
Quantum impedance conversion

The photoelectric effect means it is possible to generate a "large" voltage in a detector (e.g., a fraction of a volt), with very little signal power or energy and very little classical voltage in the light beam (< 1mV for 1nW) "quantum impedance conversion"

Optics only has to charge the photodetector and transistor to the logic voltage, not the interconnect line.

How use optics to avoid charging wires?

To exploit quantum impedance conversion, need to reduce energy in optoelectronic devices so the energy to send information optically becomes less than that of wires even for short distances e.g., centimeters or even shorter

Low energy optoelectronic devices

Pushing operating energies into the sub 10fJ or even attojoule range for output devices Modulators, LEDs, lasers including advanced nanophotonic structures Integrating sub-fF photodetectors right beside transistors

First Ge quantum well waveguide-integrated modulator

10 microns long, 0.8 microns wide, 500 nm thick intrinsic region

**On silicon**

**No resonator**

Selective area growth of quantum wells in SOI waveguides

Capacitance ~ 3 fF

3 dB modulation with 4 V bias, 1 V swing, 1460 nm

Dynamic energy per bit ~ **0.75 fJ**

Tested to 7Gb/s (equipment limited)

S. Ren et al., IEEE PTL 24, 461 – 463 (2012)
D. A. B. Miller, Optics Express 20, A293-A308 (2012)

Harris and Miller groups, Stanford
Need to move to optics to save energy

- **New additional conclusion** - Avoid wasting energy in the electrical circuits used to run interconnects
  - low energies in optoelectronic devices themselves cannot be exploited effectively if the dissipation in the associated circuits is large
    - e.g., receiver amplifier circuits dissipating 100’s fJ/bit to pJ’s/bit
    - e.g., time-multiplexing circuitry dissipating pJ’s/bit
      - clock and data recovery (CDR)
      - serialization/deserialization (SERDES)
      - clock distribution

Eliminating receiver energy

Integrate low capacitance photodetectors beside transistor input may eliminate need for voltage amplification altogether

receiverless operation

or limit it to ~ one simple low energy gain stage

“near-receiverless” operation

E.g., with 1 fJ received optical energy

- in 1 pF, generates ~1 mV
- in 30 fF (solder-bumped photodetector), generates ~33 mV
- in 1 fF, generates ~1 V

“Dual rail” schemes also eliminate need for AC coupling and line coding

“single rail” scheme

“dual rail” scheme

Eliminating receiver energy

Integration of optoelectronics right beside transistors
e.g., within < a micron or a few microns at most
allows excess capacitance in the scale of only 100’s of aF
Photodetector elements on scales of 1 micron or less dimensions allow detector capacitance of ~100 aF
Transistors themselves have input capacitances ~ 10’s to 100’s of aF
Hence <1fF total capacitance possible with integration

Eliminating multiplexing and timing energies

Avoid time multiplexing
run interconnects at the logic IC clock rates (e.g., 1 – 2 GHz)
by using massive numbers of optical channels
eliminates the serialization/deserialization (SERDES) circuits

Avoid need for data timing realignment
Optics has well-defined propagation time, so we
can run optical links with deterministic “integer” lengths
i.e., integer numbers of clock cycles
viable up to, e.g., 10 m lengths even over 0 – 100 C temperature ranges
eliminates the clock and data recovery (CDR) circuits
The number of possible optical channels (per polarization) between two surfaces of areas $A_1$ and $A_2$ separated by a distance $L$ at a wavelength $\lambda$ as limited by diffraction, is

$$N_C \approx \frac{\Omega_T A_R}{\lambda^2} = \frac{A_T A_R}{L^2 \lambda^2}$$

e.g., at 1 $\mu$m wavelength for 10 cm x 10 cm surfaces separated by 10 m

$$N_C \approx 10^6$$

for 2mm x 2mm surfaces separated by 2 cm

$$N_C \approx 4 \times 10^4$$

Number of possible free-space channels

2D arrays of 1024 free space channels

E.g., 10 x 10 micron optical "pads"

either packed closely

or spaced out, and using lenslet arrays

Free-space optical system approaches

free-space optics in large arrays
1000’s or 10,000’s or channels
even running at energy
efficient clock-rates
e.g., 2 GHz
can allow multiple Tb/s on
and off chip
even in only square
millimeters of chip area
channels can be to
neighboring chips on a board
or different boards or racks

see also J. M. Kahn and D. A. B. Miller,
“Communications expands its space,”
doi:10.1038/nphoton.2016.256
An example low energy system approach

Key additional technology
use a silicon photonics optical “interposer” layer especially with additional materials
e.g. III-Vs, germanium optical couplers, including optical vias waveguide arrays free-space couplers

Key desirable advance beam and mode couplers with %’s of loss, not dB’s of loss major opportunity for nanophotonics

Goal – 10 fJ/bit up to 10 m distance

“Straw man” system concept exploiting
• tightly integrated optoelectronics • efficient beam couplers • free-space communications with 1000’s to 10,000’s of channels

Two new reasons for complex silicon photonics

- Saving energy by eliminating the circuits in interconnect links especially eliminating time-multiplexing in short links 10–100 fJ/bit total energy instead of 1–10 pJ/bit

- New architectures and algorithms, allow
  - complex circuits to be set up and stabilized
  - self-correcting circuits
  - self-designing circuits

Potential applications in
  - optical communications, sensor preprocessing, arbitrary linear transforms, neural networks, linear optical quantum circuits, ...
Can we disentangle light beams?

Yes!

There is a way of undoing complicated scattering to separate out channels automatically, without calculations or calibrations based on self-configuring meshes of interferometers e.g., in silicon photonics circuits.

Applications

- demultiplexing of signals in multimode fibers and waveguides
  - losslessly separating overlapping light beams of different shapes
- circuits for coherent communications, tracking polarization
- sensor preprocessing, arbitrary linear transforms, neural networks
- self-aligning beam couplers and beam combiners
- tracking a source in three dimensions, undoing optical scattering
- linear optical quantum circuits, quantum computing gates
- microwave photonics

Politecnico di Milano
Recent work on mesh optics

Self-configuring and self-correcting optics demonstrations


Other recent mesh demonstrations


Theory of universal and self-configuring/correcting optics


Simple optical components – a mirror

We “design” a plane mirror
by choosing its angle
so it takes a beam of one angle
and changes it into a beam
of another angle

For another beam at another angle
the mirror changes it to a beam of
yet another angle
but we have no independent control
of what happens for the second beam
Simple optical components – a lens

We design a lens

by choosing its index and curvatures

so it takes a plane wave in one direction

and focuses it to a spot

For another plane wave in another direction

the lens focuses it to another spot

but we have no independent control

of what happens for the second beam
Simple “thin” optical components

This kind of behavior is general for “thin” optical components
  e.g., thin holograms, diffractive optical elements
  spatial light modulators, adaptive optics metasurfaces
We design them to perform some useful function for one input beam
  but we have no independent control of what happens for other beams
So these are not “arbitrary” optical components

"How complicated must an optical component be?"
An example problem

Separating overlapping beams

Politecnico di Milano, May 16, 2018
Example - Separating overlapping beams

Suppose we have two different (orthogonal) beams, e.g., from an optical fiber such as a “single bump” beam and a “two bump” beam.
Example - Separating overlapping beams

Mathematically, two (non-zero) beams are “orthogonal” if

\[ \iint E_1^* (x, y) \cdot E_2 (x, y) \, dx \, dy = 0 \]

Here, the product of the single-bump beam and the two-bump beam would be negative in the top half but positive in the bottom half, so the resulting integral would be zero.

Politecnico di Milano, May 16, 2018
Example - Separating overlapping beams

If both of these beams emerge simultaneously from the fiber, how can we separate them for example to different fibers without loss?

“loss-less” mode separator
Example - Separating overlapping beams

If both of these beams emerge simultaneously from the fiber, how can we separate them for example to different fibers without loss?

“loss-less” mode separator
If both of these beams emerge simultaneously from the fiber, how can we separate them for example to different fibers without loss?

In situations with fixed highly symmetric beams, good specific low-loss separation solutions are known.
Example - Separating overlapping beams

But for general cases of lower symmetry and/or higher complexity or where the beams change in time general solutions have not been known.

“loss-less” mode separator

Politecnico di Milano, May 16, 2018
Dividing the beam into “patches”

We can approach the beam-separation problem
by presuming it will be good enough
to imagine that we can divide the beam up
into a finite number of “patches”
Dividing the beam into “patches”

We can approach the beam-separation problem by presuming it will be good enough to imagine that we can divide the beam up into a finite number of “patches.”

We treat each of these patches as if it was approximately uniform in intensity and in phase.

At least with a sufficiently large number of patches this could be a good enough approximation and “sampling loss” may be small.
Dividing the beam into “patches”

Even relatively small numbers of patches are sufficient to distinguish beams of low or moderate complexity.
Dividing the beam into “patches”

Even relatively small numbers of patches are sufficient to distinguish beams of low or moderate complexity.
Dividing the beam into “patches”

Even relatively small numbers of patches are sufficient to distinguish beams of low or moderate complexity.
A self-aligning universal beam coupler

The first step ...

Politecnico di Milano, May 16, 2018
Coupling an arbitrary input beam

For illustration
suppose, for simplicity, that
an arbitrary input beam can be adequately described by splitting it into 4 sections

Coupling an arbitrary input beam

For illustration, suppose, for simplicity, that an arbitrary input beam can be adequately described by splitting it into 4 sections, each approximately uniform in intensity and “flat” in phase. For simplicity, neglect diffraction for the moment, assuming each of these sections will propagate as a “square” section of the beam.

Self-aligning beam coupler

Phase shifters

Beamsplitter blocks

Controllable reflectors

Detectors
Self-aligning beam coupler

Adjust phase shifter in first block to minimize power in first detector
Adjust reflectivity in first block to minimize power again in first detector
Repeat for each block
Leaves no power in detectors, all input power in output beam

Politecnico di Milano, May 16, 2018
Self-aligning the beam coupler

Now all the input beam power is coupled to the output beam
Regardless of the form of the input beam
And without any calculation or detailed calibration of devices

Politecnico di Milano, May 16, 2018
Mach-Zehnder interferometer as controllable reflector and phase shifter

A Mach-Zehnder interferometer functions both as a controllable "reflector"

Using differential drive $\Delta \theta$ of the two phase shifter arms

And as a controllable phase shifter

Using common mode drive $\theta_{av}$ of the two phase shifter arms

To control amplitude and phase of the outputs
Mach-Zehnder self-aligning implementations

Adjust phase in device 4 to minimize power on D3
Adjust split ratio in device 3 to minimize power on D3
Adjust phase in device 3 to minimize power on D2
Adjust split ratio in device 2 to minimize power on D2
Adjust phase in device 2 to minimize power on D1
Adjust split ratio in device 1 to minimize power on D1
All power from the input waveguides now in output beam

Mach-Zehnder self-aligning implementations

Grating couplers would allow us to couple a free-space beam to a Mach-Zehnder implementation of the device.
The second step ...

Separating multiple orthogonal beams

Politecnico di Milano, May 16, 2018
Separating multiple orthogonal beams

Once we have aligned beam 1 using detectors D11 – D13 an orthogonal input beam 2 passes through the nearly transparent detectors to the second row where we can self-align it using detectors D21 – D22 separating two overlapping orthogonal beams to separate outputs.

D11, D12, D13 are mostly-transparent detectors
Since alignment and re-alignment need not be performed at data-rate speeds
only need small signals from the detectors
Many ways of making such mostly-transparent detectors
Separating multiple orthogonal beams

Adding more rows and self-alignments separates a number of orthogonal beams equal to the number of beam “segments” here, 4
Extensions – tracking a source

Optimize reception of a wave from a source

An adaptive front end for a sensing system

e.g., finding and/or tracking a source

Extensions – tracking a source

Optimize reception of a wave from a source

An adaptive front end for a sensing system
e.g., finding and/or tracking a source

Creating an optimum backwards channel to a source
for communications or powering

Extensions – self-stabilizing optical systems

Keep “training beams” on all the time in the background
“key” each training beam, e.g., with an amplitude modulation
at a different frequency for each beam
photodetector rows respond only to “keyed” signals
With the local electronic feedback loops enabled in sequence
the system will stabilize itself
to continue implementing the trained function
E.g., arbitrary transforms for optical and/or quantum processing

Extensions – optimal channels

Iterating back and forward between the two sides
Finds the optimal orthogonal channels through any medium
Physically, performs the singular value decomposition of the optics from the waveguides on the left to the waveguides on the right

A self-configuring universal spatial device

The third step ...

A self-configuring universal spatial device
Arbitrary mode converter

Exploit two back-to-back self-aligning universal beam couplers

Self-align input coupler by shining in beam of interest and adjusting using local feedback loops as before

Arbitrary mode converter

Self-align the output coupler

shining desired output beam backwards into the output coupler

technically, the phase conjugate of the desired output beam

And adjusting using local feedback loops as before

But now in the output coupler
Self-configuring linear spatial device

Now any single desired input beam is converted to any single desired output beam with no calculations by training the device with the desired beams. Can also adjust modulator to give desired coupling strength.
General multiple mode converter

Can extend to all the possible orthogonal beams the device can support
Mach-Zehnder implementation

Same concept can be implemented in a planar Mach-Zehnder form
No crossing waveguides required
Example here has the output order flipped for compactness

Describing an arbitrary linear optical component

Any linear optical component can be described by a linear “device” or scattering operator $D$ and we can perform the singular value decomposition (SVD)

$$D = V D_{\text{diag}} U^\dagger$$

$U$ and $V$ are unitary operators and $D_{\text{diag}}$ is a diagonal operator.

If we can emulate any SVD for given input and output spaces then we can make any linear optical device for those spaces.

These configurations implement the singular value decomposition with full generality so we can design and make any linear optical component for these input and output spaces. Proves any linear optical component is possible.

Conclusions

Novel low-energy optoelectronics, together with photodetectors right beside transistors allow optoelectronics at 10 fJ or even 1 fJ levels and can eliminate 100fJ/bit – pJ/bit receiver power dissipation.

Adding free-space optical arrays and silicon photonics as an interposer allow us to eliminate the time-multiplexing circuitry so we can go from 1 – 10 pJ/bit to 10 – 100 fJ/bit.

New algorithms and designs for interferometer meshes allow us to exploit the silicon photonics complexity.

Applications in communications, sensing signal processing and computing linear optical quantum circuits.

New classes of self-configuring and self-optimizing optics adaptable, manufacturable, complex optics.
Conclusions

If you can think of an linear optical device
at least for one wavelength
there is a way of designing it
and making it (if it is not too complicated)
with interferometer meshes
which can also design and stabilize themselves

Now is the time to have fun with this!
Something new for silicon photonics
get funded
make these devices
get rich!

Politecnico di Milano, May 16, 2018


“Designing Linear Optical Components,” Optics in 2013 Special Issue, Optics and Photonics News, December 2013, p. 38


For an overview, including all these links, see http://www-ee.stanford.edu/~dabm/Selfalign.html