100GbE Lambda Switching for Data Center Networks

Nathan Farrington
CEO
Packetcounter, Inc.
nathan.farrington@packetcounter.com

OIDA 100G Lambda Data Center Workshop, San Jose, CA, USA, June 12-13, 2014

How to get to 100G?

Bit rate, $R = x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5$

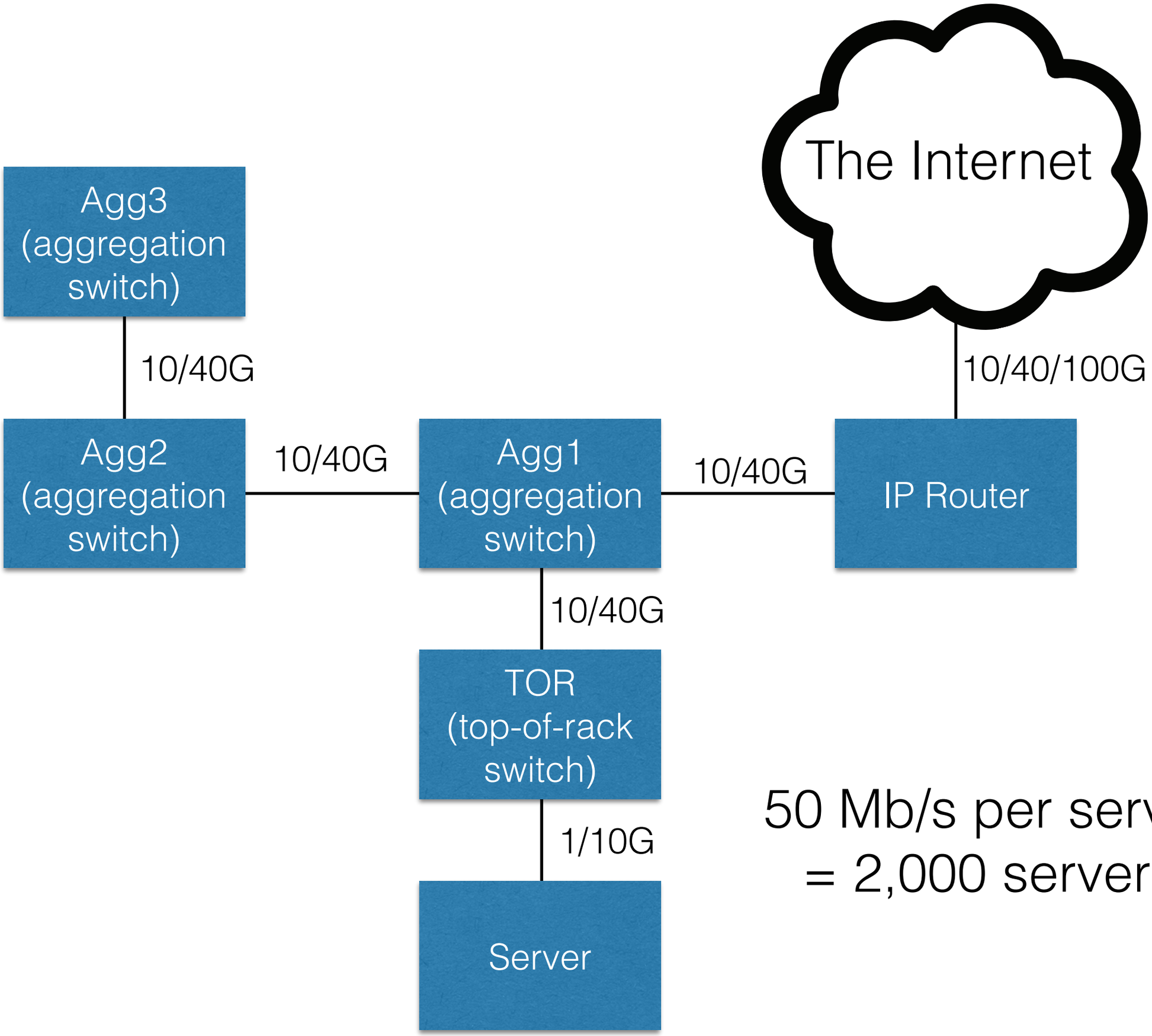
- x_1, number of fiber pairs (e.g. 1, 4, 10)
- x_2, number of wavelengths (WDM, e.g. 1–80)
- x_3, number of polarizations (PDM, e.g. 1, 2)
- x_4, modulation order (bits per symbol, e.g. 1, 2, 4)
- x_5, signaling rate (symbols per second)

Chris Cole, Ilya Lyubomirsky, Ali Ghiasi, Vivek Telang. “Higher-Order Modulation for Client Optics”, In *IEEE Communications Magazine*, March 2013.

Getting to R=100G

	x	x	x	x	x	
1	10	1	1	1	10G	100GBASE-SR10
2	4	1	1	1	25G	100GBASE-SR4
3	1	4	1	1	25G	100GBASE-LR4
4	4	1	1	1	25G	100G PSM4 MSA
5	1	4	1	1	25G	100G CWDM4 MSA
6	1	1	**1**	**1**	**100G**	**Possible future 1 of 6**
7	1	1	**2**	**1**	**50G**	**Possible future 2 of 6**
8	1	1	**1**	**2**	**50G**	**Possible future 3 of 6**
9	1	1	**4**	**1**	**25G**	**Possible future 4 of 6**
10	1	1	**1**	**4**	**25G**	**Possible future 5 of 6**
11	1	1	**2**	**2**	**25G**	**Possible future 6 of 6**

Where do data centers need 100G?

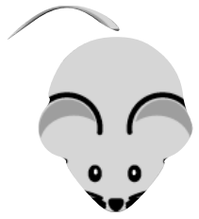


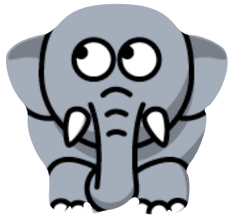
50 Mb/s per server
= 2,000 servers

How could data centers use 100G lambdas?

100G optics is cheap;
400G packet switch ports are expensive

For Mice, 20% of traffic





For Elephants, 80% of traffic

<u>Latency-sensitive traffic</u>

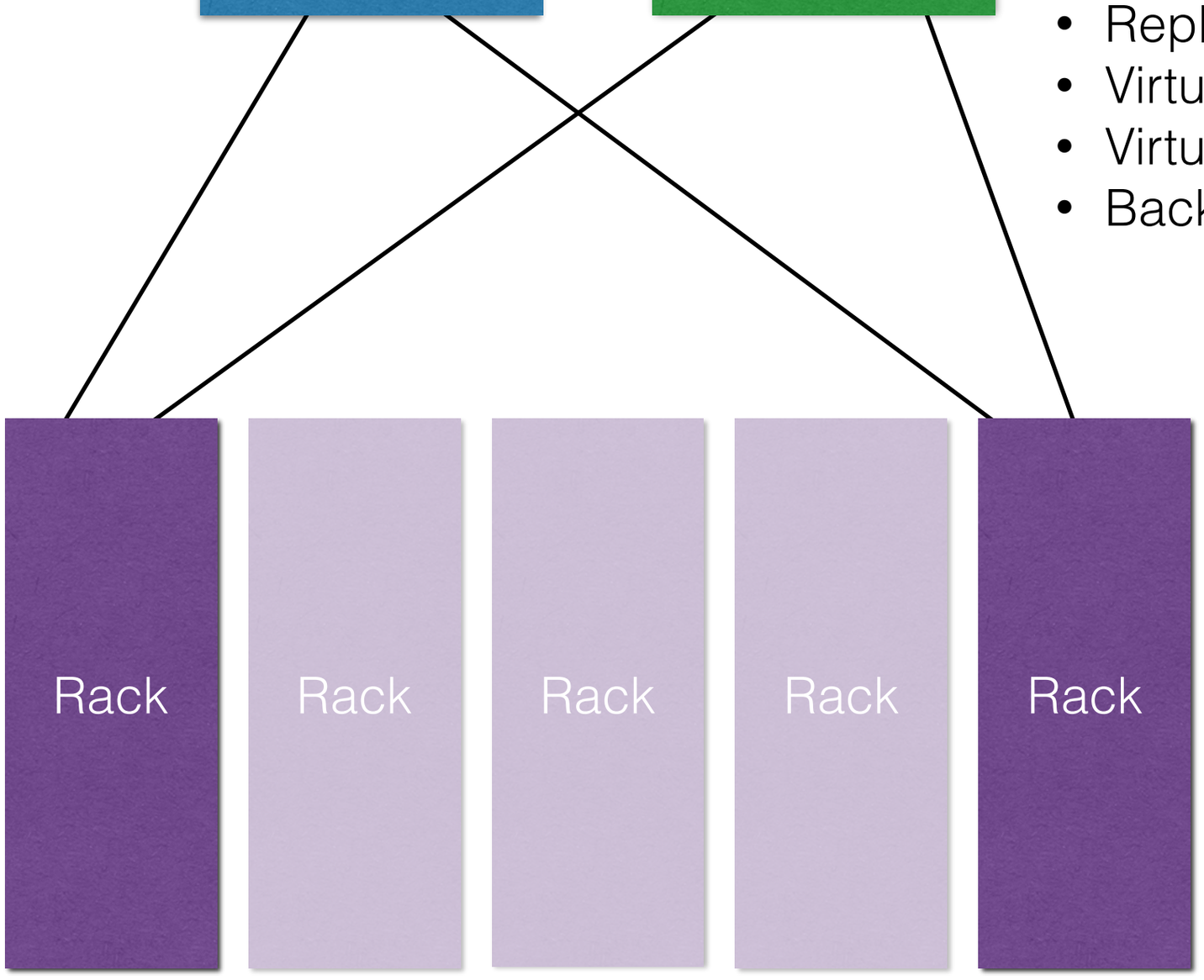
- Remote procedure calls
- Database queries
- Telemetry
- Logging

Electronic Packet Switch

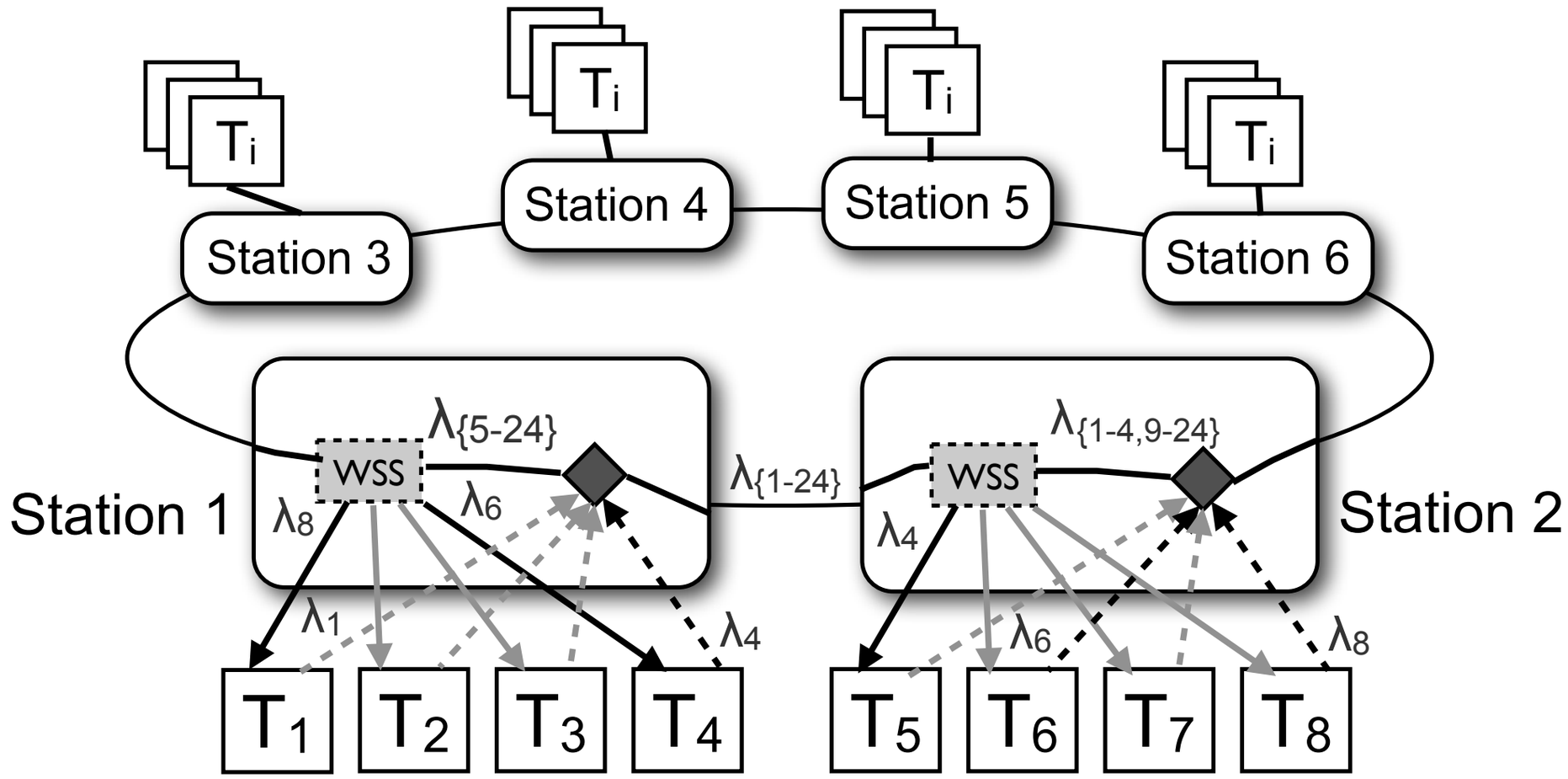
Optical TDMA Switch

<u>Throughput-sensitive traffic</u>

- Video
- Big data
- Replication
- Virtual machine cloning
- Virtual machine migration
- Backup



Optical TDMA Switch



George Porter, Richard Strong, Nathan Farrington, Alex Forencich, Pang-Chen Sun, Tajana Rosing, Yeshaiahu Fainman, George Papen, and Amin Vahdat. “Integrating Microsecond Circuit Switching into the Data Center”. In ACM SIGCOMM 2013.

Nathan Farrington, Alex Forencich, Pang-Chen Sun, Shaya Fainman, Joe Ford, Amin Vahdat, George Porter, and George Papen. “Invited Paper: A 10 μs Hybrid Optical-Circuit/Electrical-Packet Network for Datacenters”. In OFC 2013.

Traffic Matrix Scheduling

Bandwidth Allocation Matrix

0.094	0.286	0.521	0.1
0.144	0.173	0.152	0.531
0.279	0.394	0.027	0.299
0.483	0.147	0.299	0.071

Time Durations | **Assignments**

$$= 0.394 \begin{bmatrix} 0&0&1&0 \\ 0&0&0&1 \\ 0&1&0&0 \\ 1&0&0&0 \end{bmatrix} + 0.144 \begin{bmatrix} 0&1&0&0 \\ 1&0&0&0 \\ 0&0&0&1 \\ 0&0&1&0 \end{bmatrix} + 0.137 \begin{bmatrix} 0&1&0&0 \\ 0&0&0&1 \\ 1&0&0&0 \\ 0&0&1&0 \end{bmatrix}$$

$$+ 0.094 \begin{bmatrix} 1&0&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 0&1&0&0 \end{bmatrix} + 0.071 \begin{bmatrix} 0&0&1&0 \\ 0&1&0&0 \\ 1&0&0&0 \\ 0&0&0&1 \end{bmatrix} + 0.056 \begin{bmatrix} 0&0&1&0 \\ 0&1&0&0 \\ 0&0&0&1 \\ 1&0&0&0 \end{bmatrix}$$

$$+ 0.053 \begin{bmatrix} 0&0&0&1 \\ 0&0&1&0 \\ 1&0&0&0 \\ 0&1&0&0 \end{bmatrix} + 0.027 \begin{bmatrix} 0&0&0&1 \\ 0&1&0&0 \\ 0&0&1&0 \\ 1&0&0&0 \end{bmatrix} + 0.018 \begin{bmatrix} 0&0&0&1 \\ 0&1&0&0 \\ 1&0&0&0 \\ 0&0&1&0 \end{bmatrix} + 0.045 \begin{bmatrix} 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 1&0&0&0 \end{bmatrix}$$

Nathan Farrington, George Porter, Yeshaiahu Fainman, George Papen, and Amin Vahdat. "Hunting Mice with Microsecond Circuit Switches". In ACM HotNets 2012.

Summary

- Six possible futures for single-wavelength 100G
- 100G single-wavelength transceivers will be available before 400G four-wavelength transceivers
- 100G single-wavelength transceivers are best used with an optical TDMA switch