ENLITENED Annual Meeting

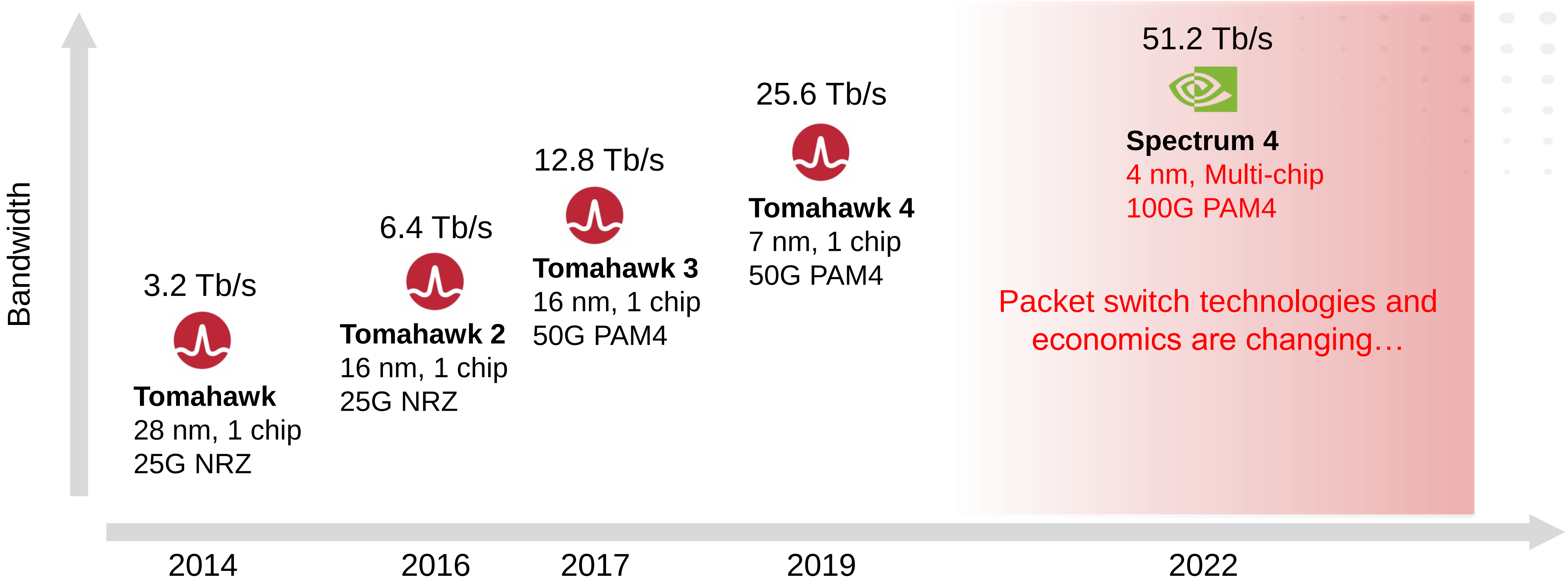
LEED Phase 2

Making Practical Optical Networking a Reality

George Porter/Papen

July 19, 2022

Networking today: Packet switch scaling trends



3.2 Tb/s

Tomahawk
28 nm, 1 chip
25G NRZ

6.4 Tb/s

Tomahawk 2
16 nm, 1 chip
25G NRZ

12.8 Tb/s

Tomahawk 3
16 nm, 1 chip
50G PAM4

25.6 Tb/s

Tomahawk 4
7 nm, 1 chip
50G PAM4

51.2 Tb/s

Spectrum 4
4 nm, Multi-chip
100G PAM4

Packet switch technologies and economics are changing…

https://www.nextplatform.com/2019/12/12/broadcom-launches-another-tomahawk-into-the-datacenter/

https://www.nextplatform.com/2022/04/01/spectrum-4-ethernet-leaps-to-800-gb-sec-with-nvidia-circuits/

LEED: Cost- and energy-efficient network scaling

- *Network scaling* via *packet switch scaling* increasingly no longer "business as usual"

- **Packet switch transistor scaling**
  - Design, fab, test costs increasing
  - Footprint – die size limited
  - Power/thermal envelope – air cooling limited
- *Multi-chip packaging (cost, complexity)*
- *Liquid cooling on the horizon (cost)*

- **Packet switch I/O scaling**
  - Serdes footprint & power increasing
  - Board trace losses increasing
- *Multi-chip package and/or Copackaging (cost, complexity)*

Optical switching provides an additional / alternative path for scaling networks

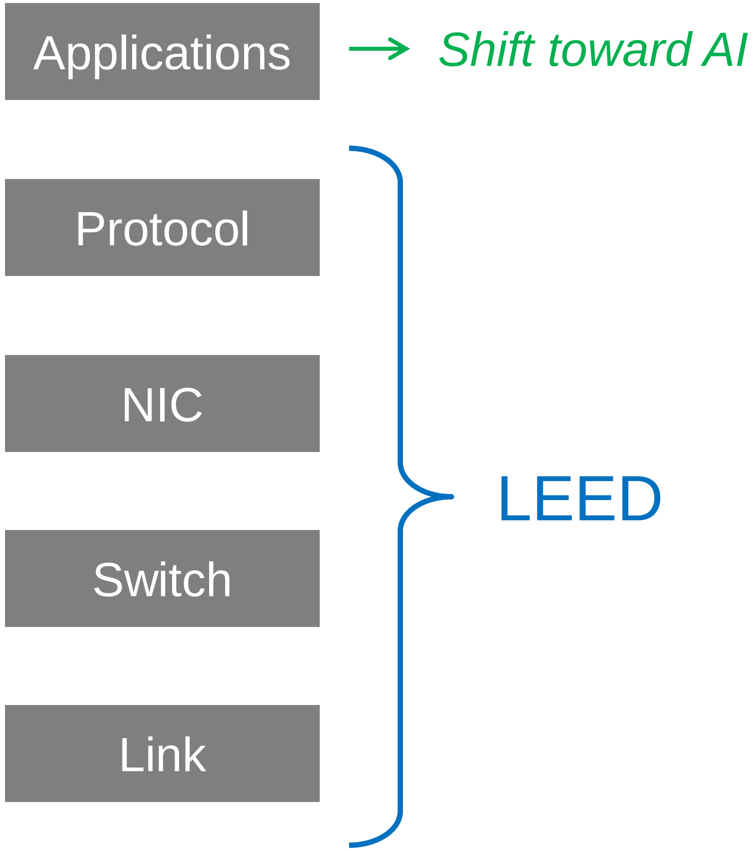
CHANGING WHAT'S POSSIBLE

Opportunities and challenges for optical switching

- **Opportunities**
  - Reduced CAPEX
    - Hardware cost savings: fewer optical transceivers + lower cost per switch port
  - Reduced OPEX
    - Lower networking power consumption
    - More networking bandwidth/CAPEX → More power efficient compute

- **Challenges**
  - Overcoming customer hesitancy around new technology
  - Supporting applications designed for a packet-switched world
  - Introducing new functionality at multiple points in the networking stack…

LEED Project Overview



Applications → *Shift toward AI*

Protocol

NIC

Switch

Link

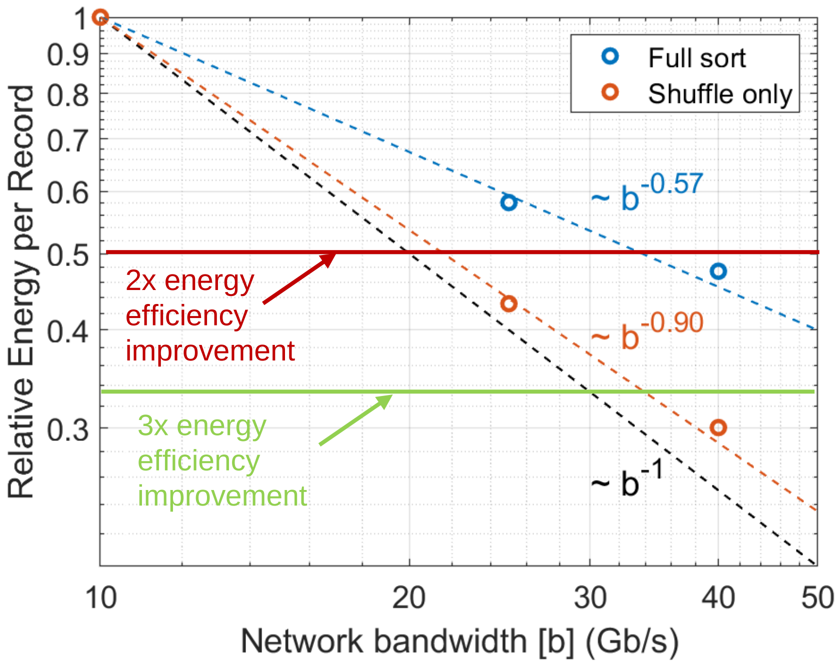
LEED

System level goal: 2x increase in transactions/Joule

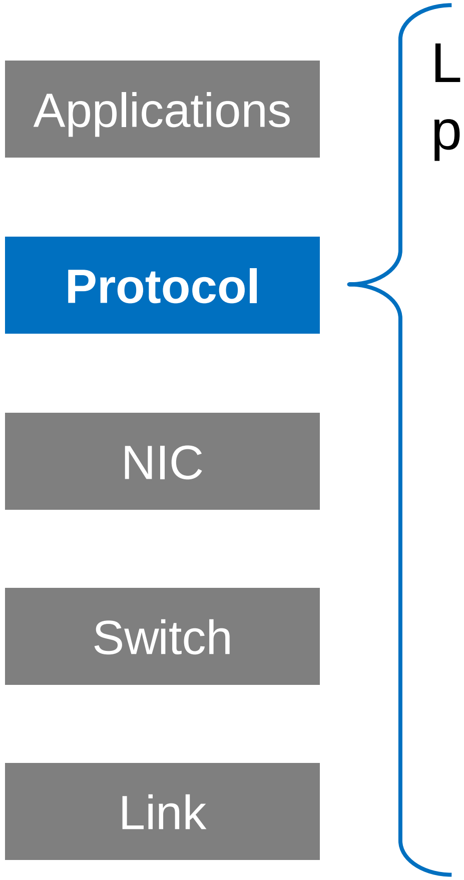
Initial experiments confirm LEED premise:

Higher bandwidth leads directly to higher energy efficiency

- For shuffle-bound apps: *~2x bandwidth → 2x energy efficiency*
- For sort application: *~3x bandwidth → 2x energy efficiency*
- "Bandwidth per buck" determines actual operating point



Protocols for sending packets over circuits



LEED Phase II: developing protocols that allow applications designed for a packet-switched world to work with an underlying circuit-switched network

Bandwidth efficiency via store-and-forward (RotorLB)

- Network modeling indicates 2-3x cost-normalized bandwidth improvement ✓

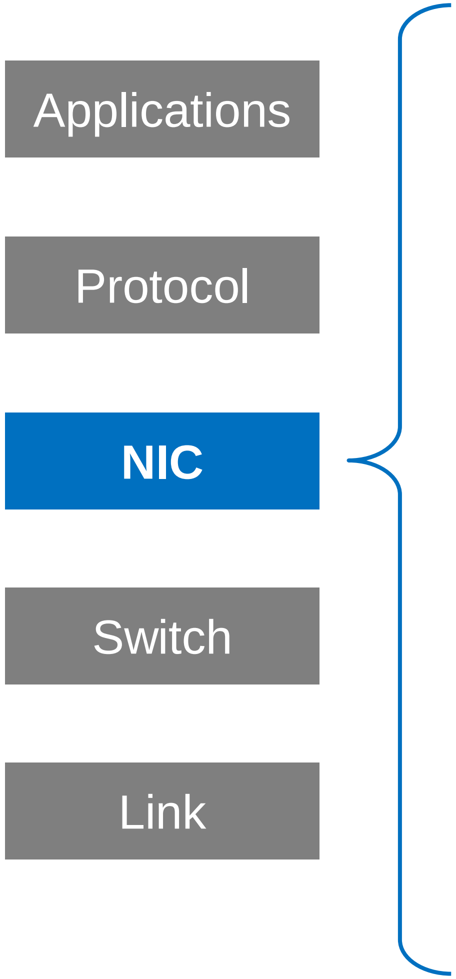
Low latency via cut-through-forwarding (Opera)

- Network modeling indicates comparable latencies to a packet switched network ✓

Minimal buffering & congestion control

- Initial designs show ~ 50 kB buffers possible for a modest throughput reduction ✓
- Optimize tradeoff between buffering and maximum throughput [ongoing]

Network interface hardware



Applications

Protocol

NIC

Switch

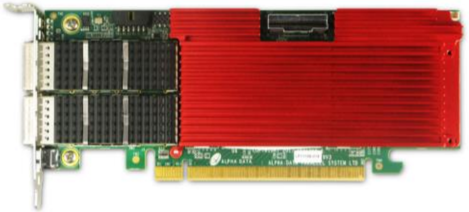
Link

LEED Phase II: developing a network interface controller capable of implementing time-synchronized protocols and interfacing with an optical switch at high bandwidth

Corundum – open-source FPGA NIC design providing:

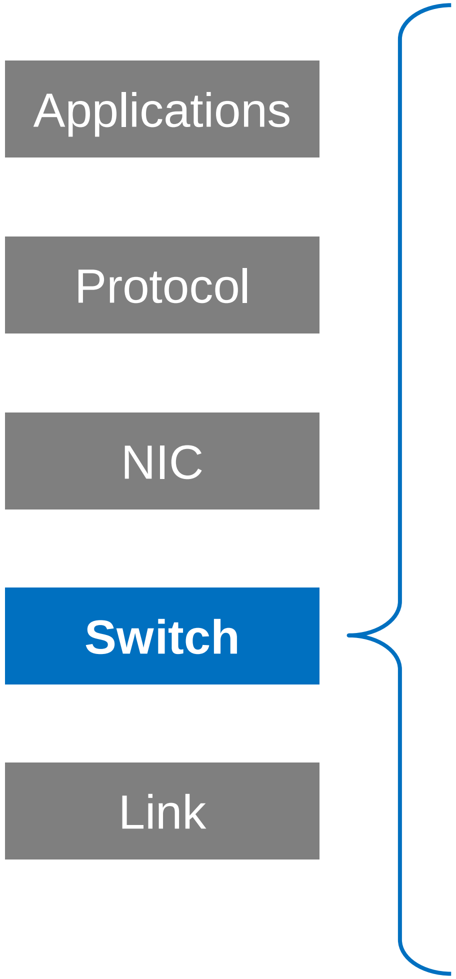
- High-precision time sync with PTP IEEE 1588 ✓
- Timed packet transmission ✓
- 1,000s of independent hardware queues ✓
- High-bandwidth PCIe interface with DMA ✓
- Protocol implementation [ongoing]
- Software driver (currently sockets based, RDMA planned) [ongoing]

FPGA NIC
200 Gb/s network I/O



Industry support from Xilinx, Intel, Cisco and Silicom – all code is open source

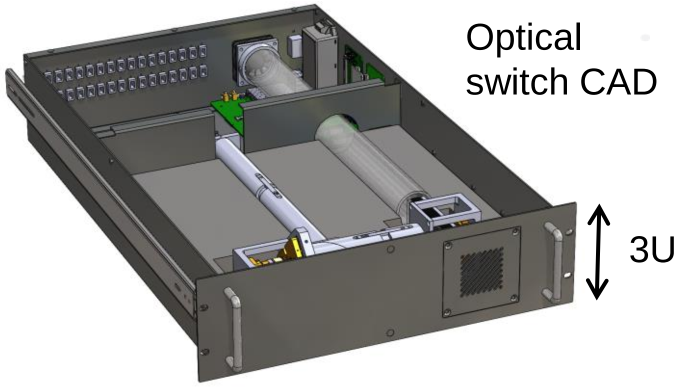
Optical switching hardware



LEED Phase II: fabricating a high-radix, high-speed, low-loss optical rotor switch using technologies compatible with low-cost manufacture

Optical switch design ✓

- 128 in x 128 out single-mode fibers
- 8 µs switching time
- 5 dB insertion loss
- Lower-cost mfg. methods (vs. Phase I)



Optical switch CAD

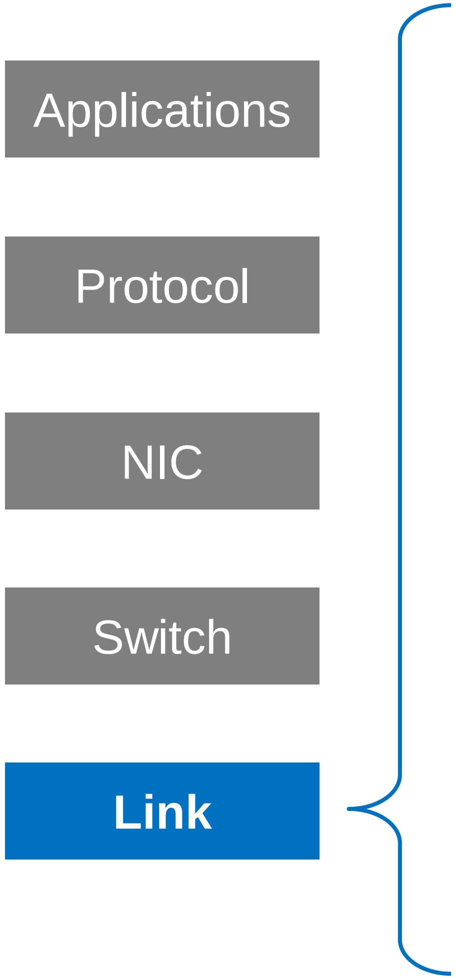
Optical switch fabrication:

- 3U enclosure, electronics, mechanics designed ✓
- Full assembly waiting on subsystem supplier delays [ongoing]

Motor control:

- Two rotors synchronized to ± 6 µs phase error at 7,200 rpm ✓
- Improving controller to achieve similar performance at 15,000 rpm [ongoing]

Optical link hardware



LEED Phase II: demonstrating link-level components needed to acquire and close a link via burst-mode without optical amplification in the presence of power transients

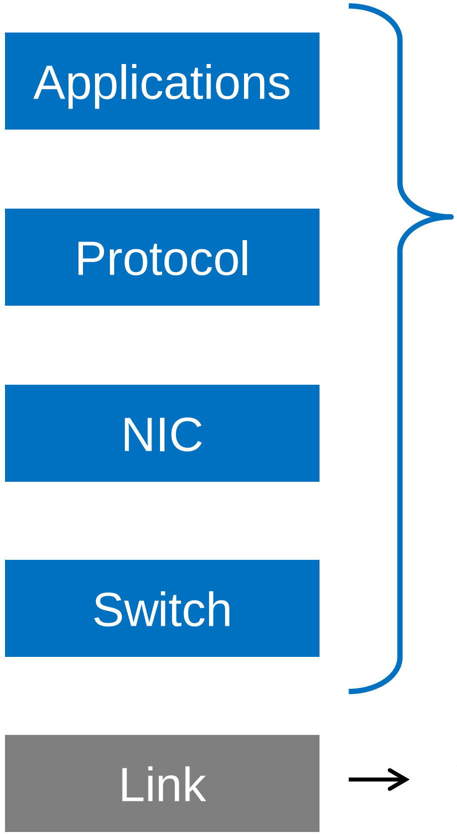
Larger link margin using ADPs (Sandia/UCSD)

- Compensate for signal attenuation through optical switch
- Operation at 25 Gb/s demonstrated with sufficient responsivity for switch loss
- Potential new LEED research in compensating burst-mode nonlinearities

Larger link margin via high-OMA modulator (Axalume)

- Compensate for switch loss at the transmitter
- Integrated 25 Gb/s per-lane modulator array compatible with WDM source

Technology Transition: Testbed Deployment


Applications

Protocol

NIC

Switch

Link →

Collaboration with Sandia National Labs

- 32 compute nodes with 200 Gb/s optical I/O per node
- Integrates: applications, driver, protocol, NIC, and optical switch with COTS links
- Can be configured as an equal-bandwidth packet switched network for comparison

Outcomes:

- *Demonstrate interoperation of key pieces of LEED network stack*
- *Quantify value propositions of the LEED network for customer-specified applications*

We will initially use commercially available links

- Custom links not required for network operation
- Custom links will enable maximum performance

Technology Transition: Pathways

- **Axalume** (est. March 2017)
  - Pathway for commercializing link technology
  - Products can support optically-switched networks and/or other market segments
  - Secured external funding/investment (gov. and industry)
- **inFocus Networks** (est. March 2018)
  - Pathway for commercializing switch and networking tech.
  - Focus is optically-switched datacenter/HPC networks, but initial traction in other markets as well
  - Secured external funding (NSF)

Ongoing work

- Phase II technical:
  - Complete switch fabrication + NIC & protocol implementation
  - Complete testbed build-out
  - Analyze application performance on testbed & quantify value propositions

- Phase II tech transfer:
  - Use testbed results to advance partner/customer relationships
  - Confirm initial target market segment and engage customers

- Follow on (SCALEUP / private investment):
  - Optimizations for initial use-cases/customers
  - Hardware manufacture
  - Pilot deployment with customer