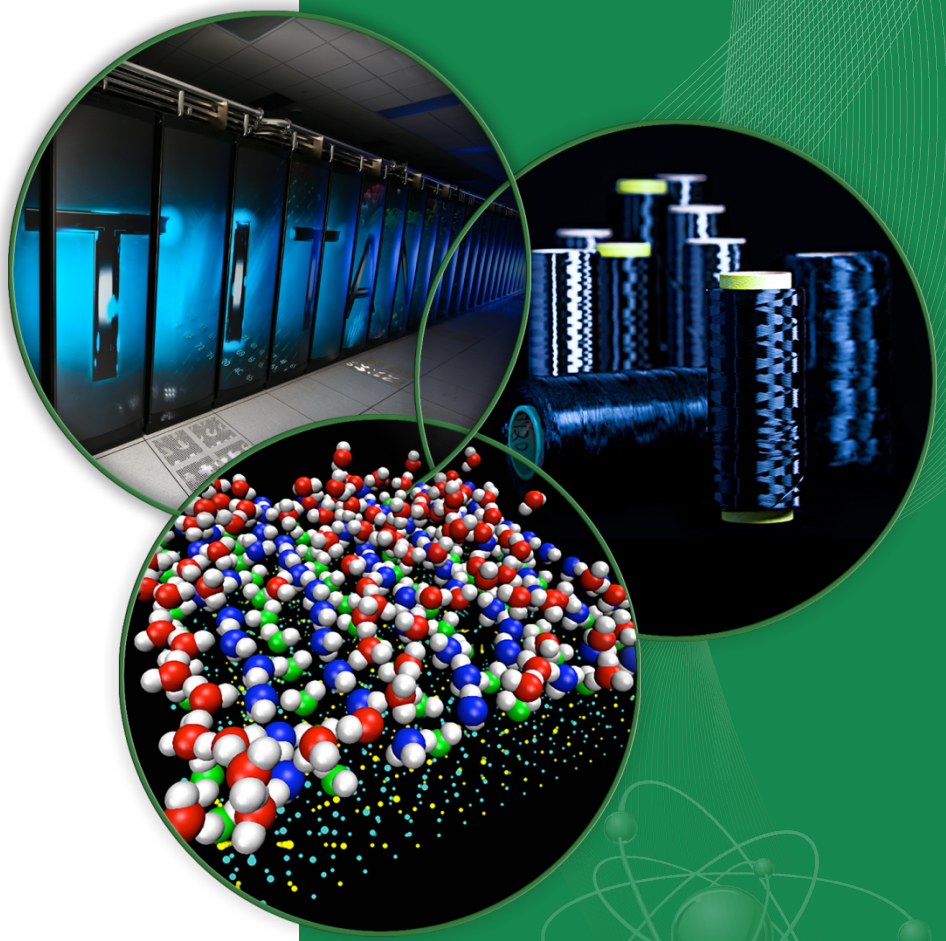
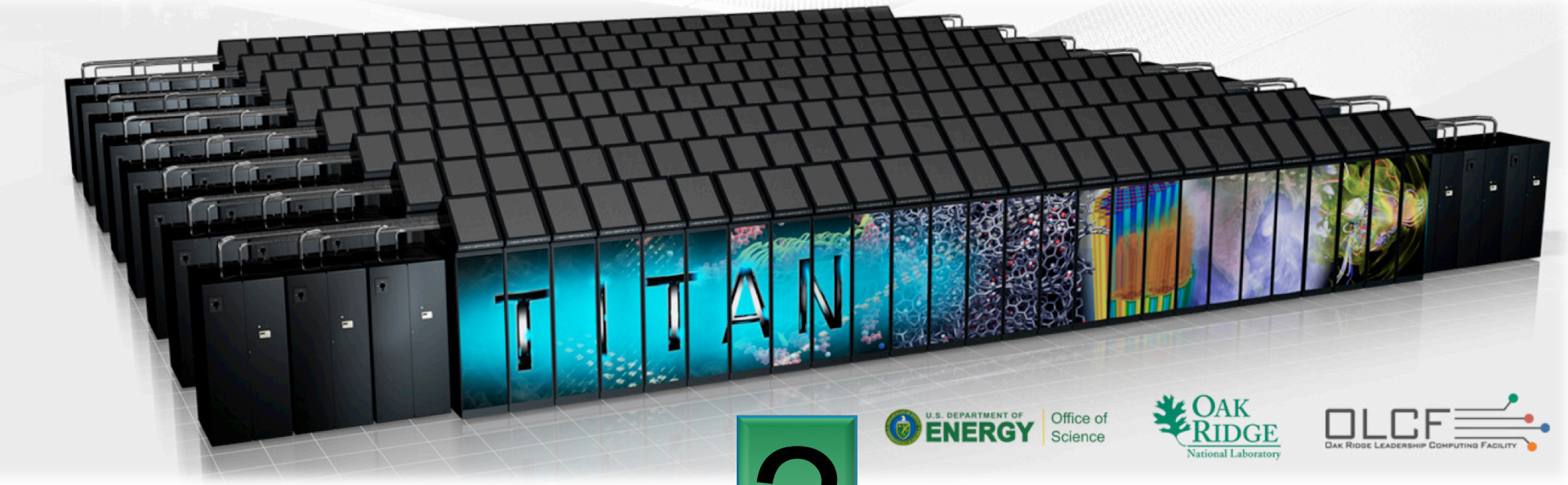
I/O Router Placement and Fine-Grained Routing on Titan to Support Spider II

Matt Ezell (Presenter)
David Dillow
Sarp Oral
Feiyi Wang
Devesh Tiwari
Don Maxwell
Dustin Leverman
Jason Hill

CUG 2014, Lugano Switzerland



The Million Dollar Question

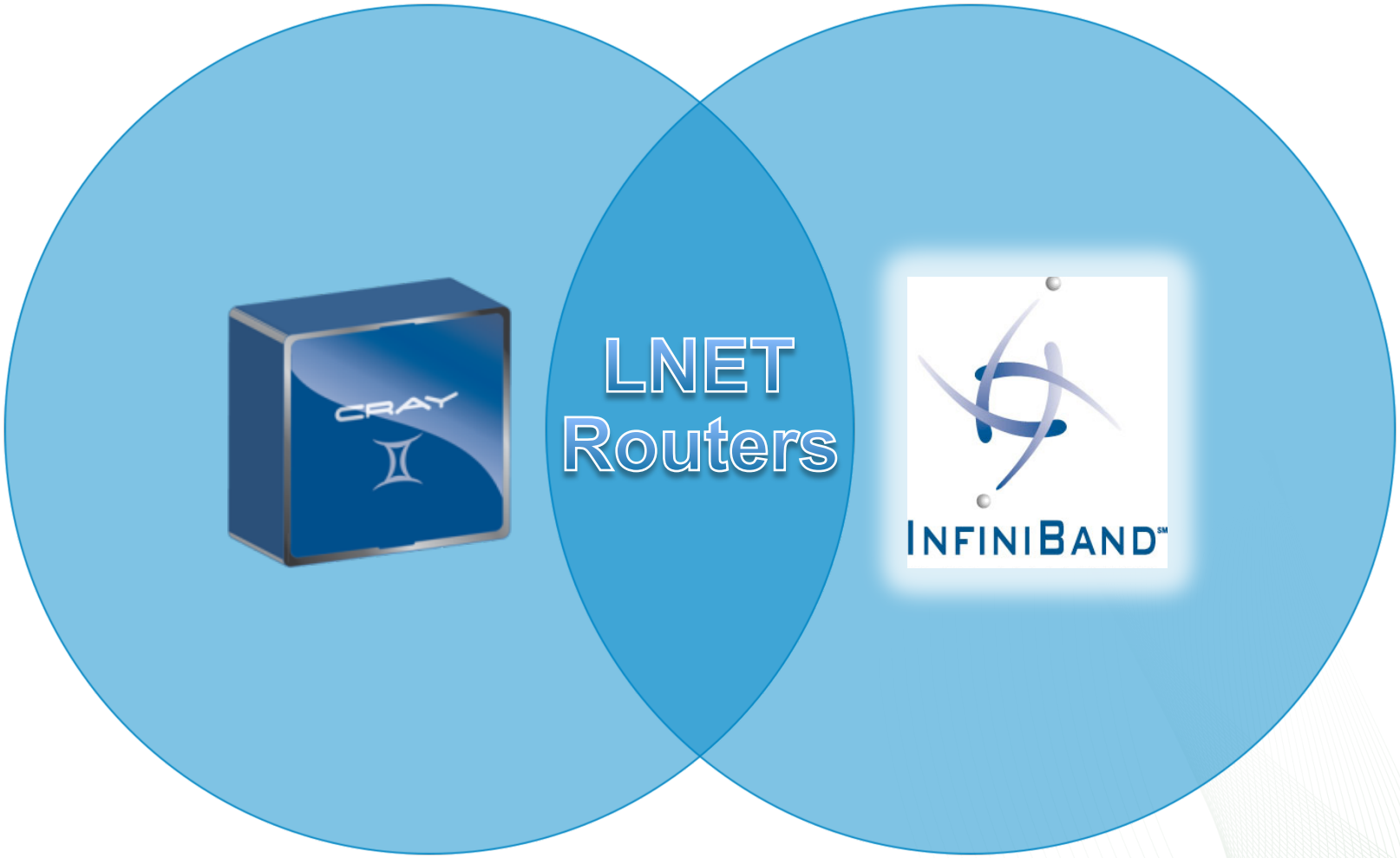


?

lustre™


LNET ROUTING
US
66

LNET Routing



LNET Routing

identifier@network

10.10.10.101@o2ib0

10.10.10.101@o2ib201

9409@gni0

9409@gni101

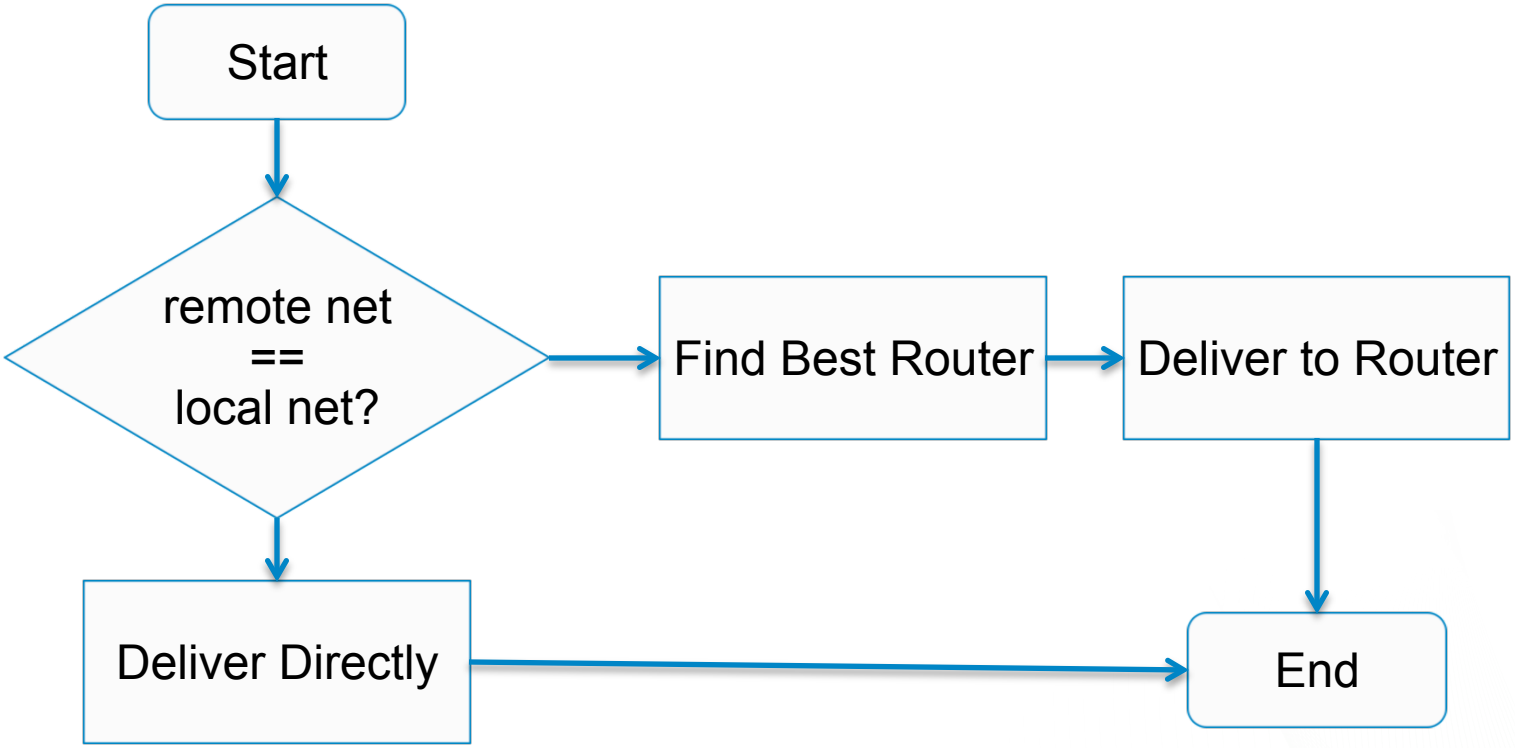
LNET Routing

options lnet routes="remoteNet hops id@localNet"

lctl --net remoteNet add_route id@localNet hops

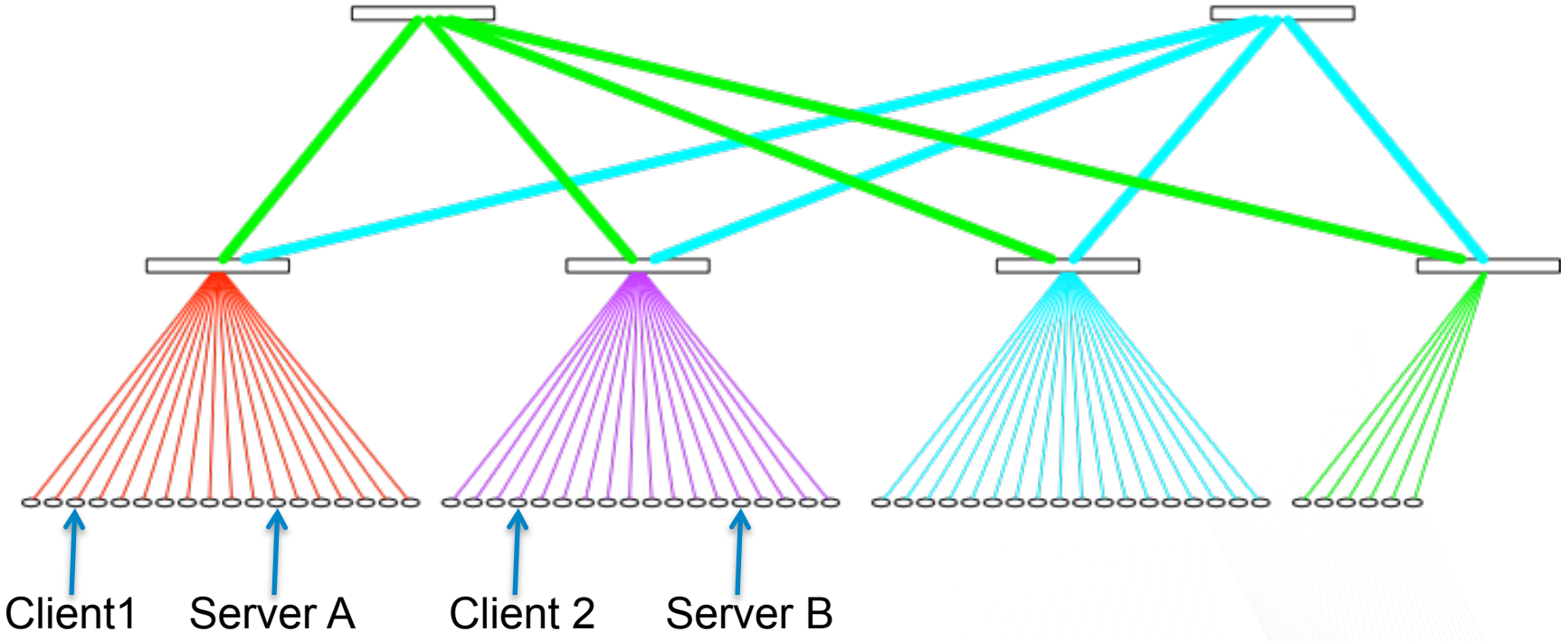
```
net              o2ib225 hops 10 gw          18329@gni102 up
net              o2ib225 hops 1 gw           10871@gni102 up
net              o2ib234 hops 10 gw           5991@gni102 up
net              o2ib234 hops 10 gw          18247@gni102 up
net              o2ib234 hops 1 gw           10921@gni102 up
net              o2ib208 hops 10 gw          15218@gni102 up
net              o2ib208 hops 10 gw           2946@gni102 up
net              o2ib208 hops 1 gw           7788@gni102 up
net              o2ib217 hops 10 gw          15212@gni102 up
net              o2ib217 hops 10 gw           2908@gni102 up
```

LNET Routing



Purpose of FGR

Fat Tree Network



Blue Waters I/O

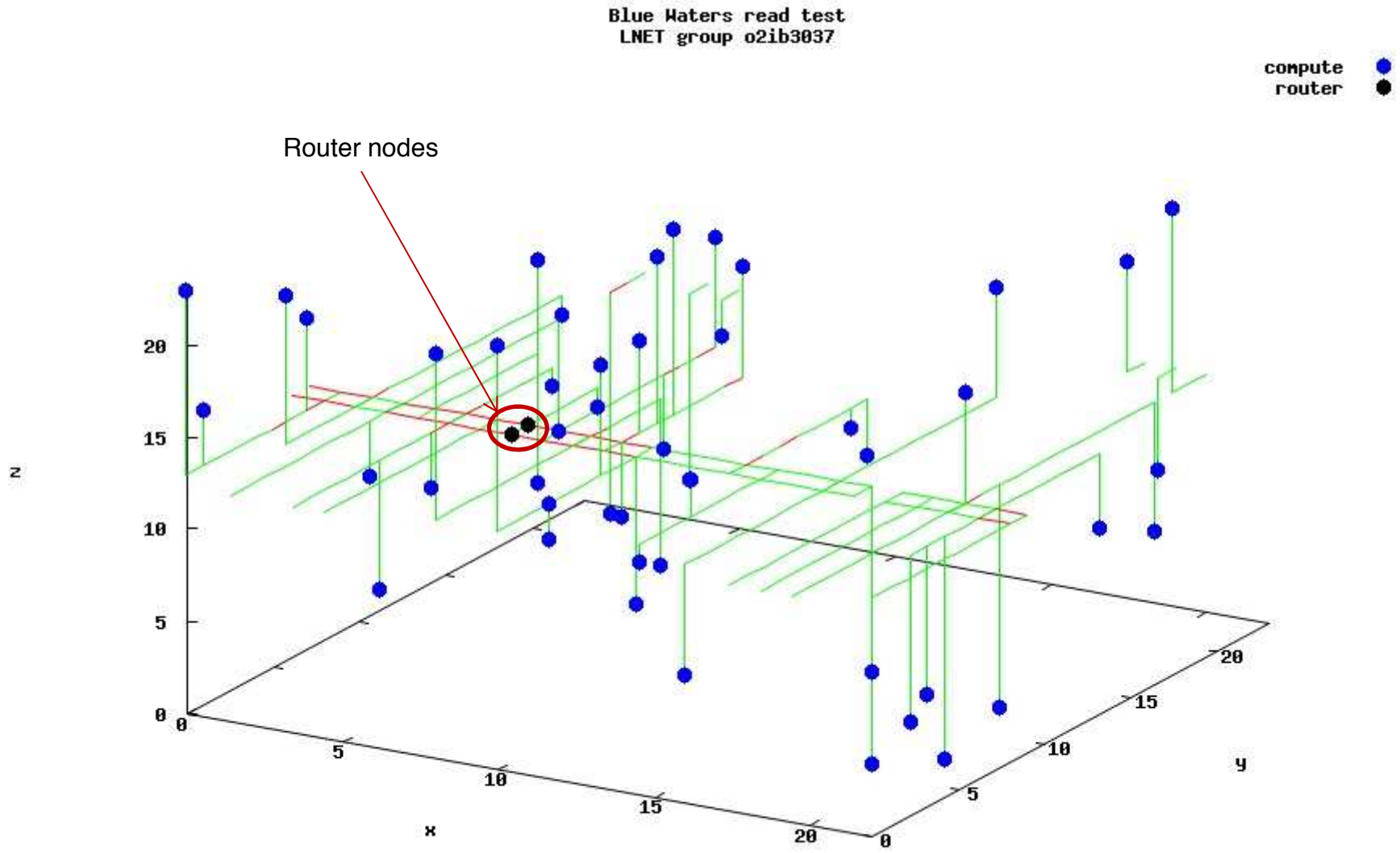
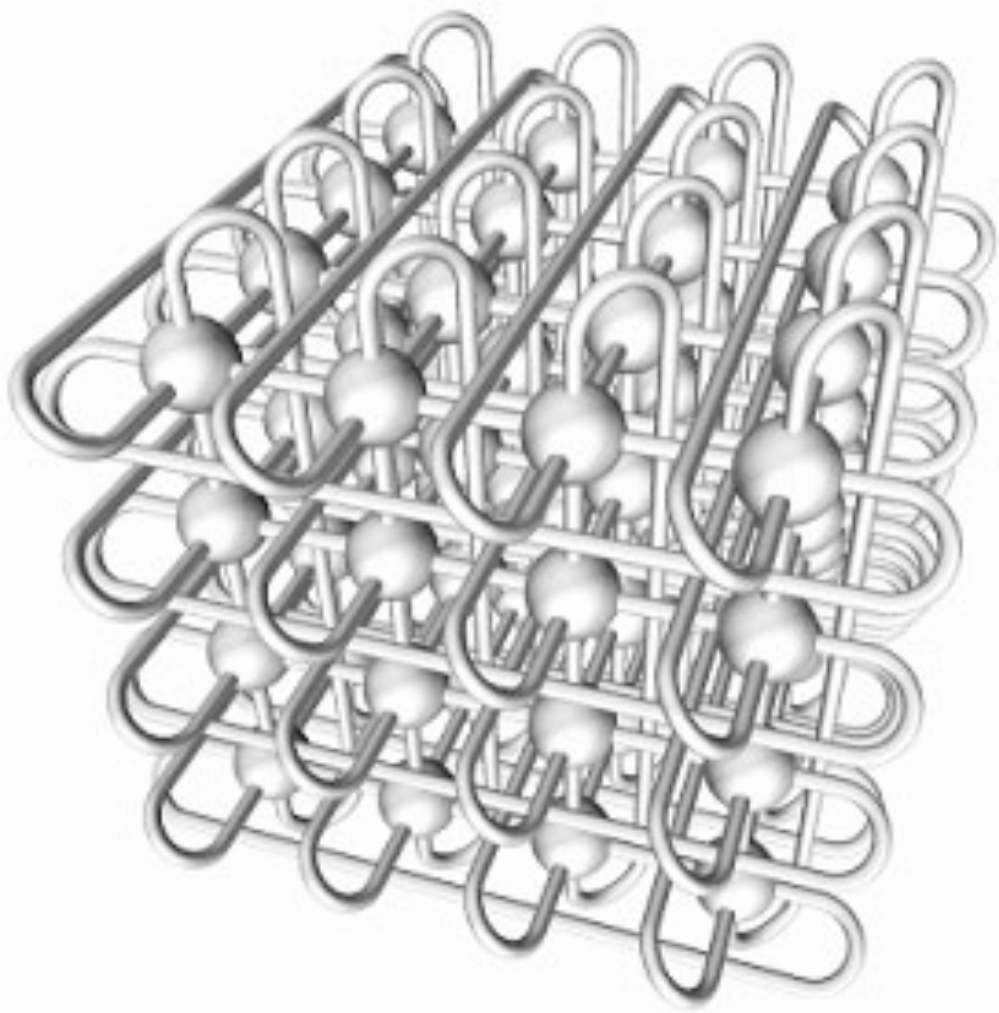


Image taken from the presentation for “Blue Waters I/O Performance” at CUG 2013
https://cug.org/proceedings/cug2013_proceedings/includes/files/pap129-file2.pdf

Titan 3D Torus



Dimension Ordered Routing

Blue Waters I/O

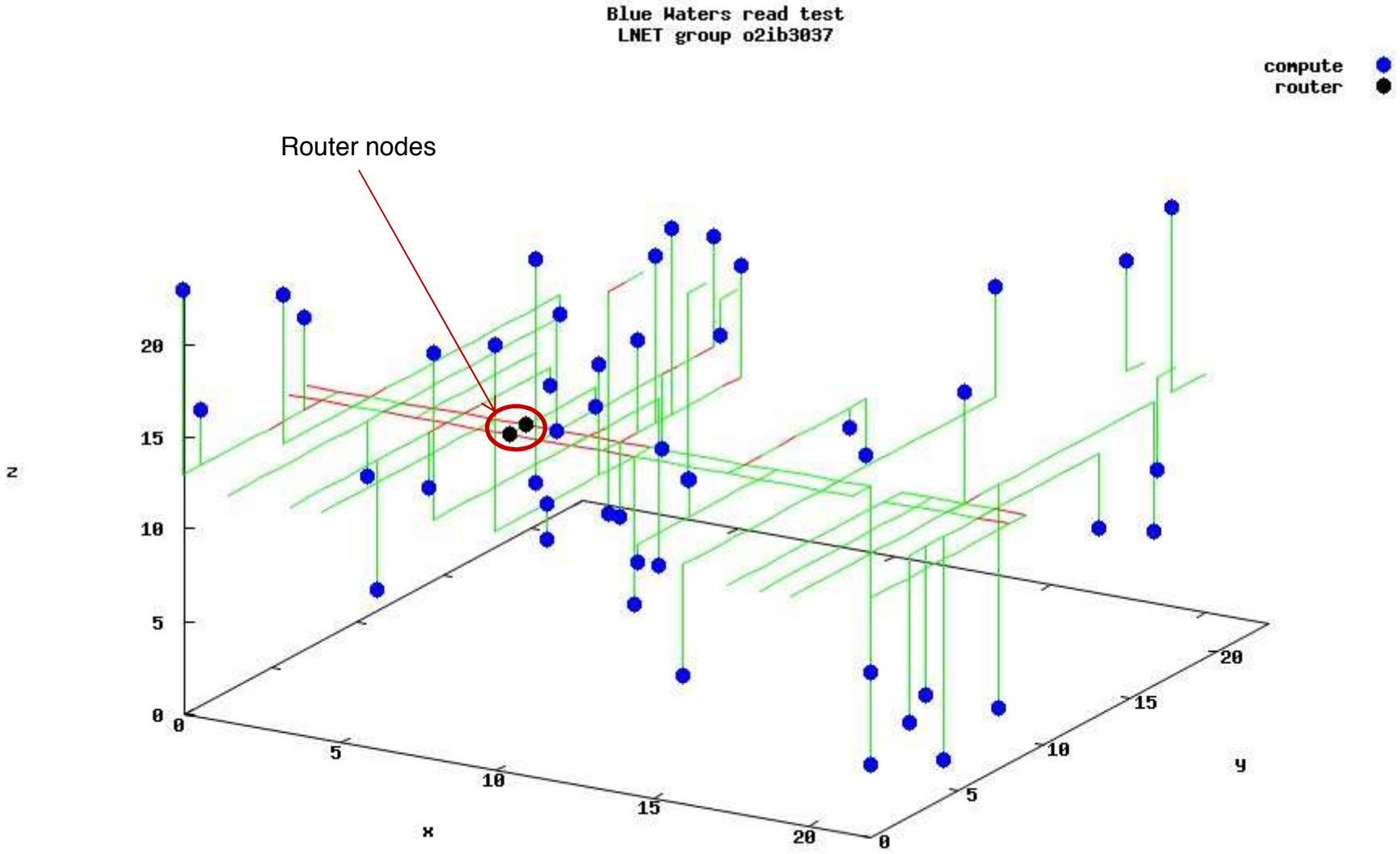


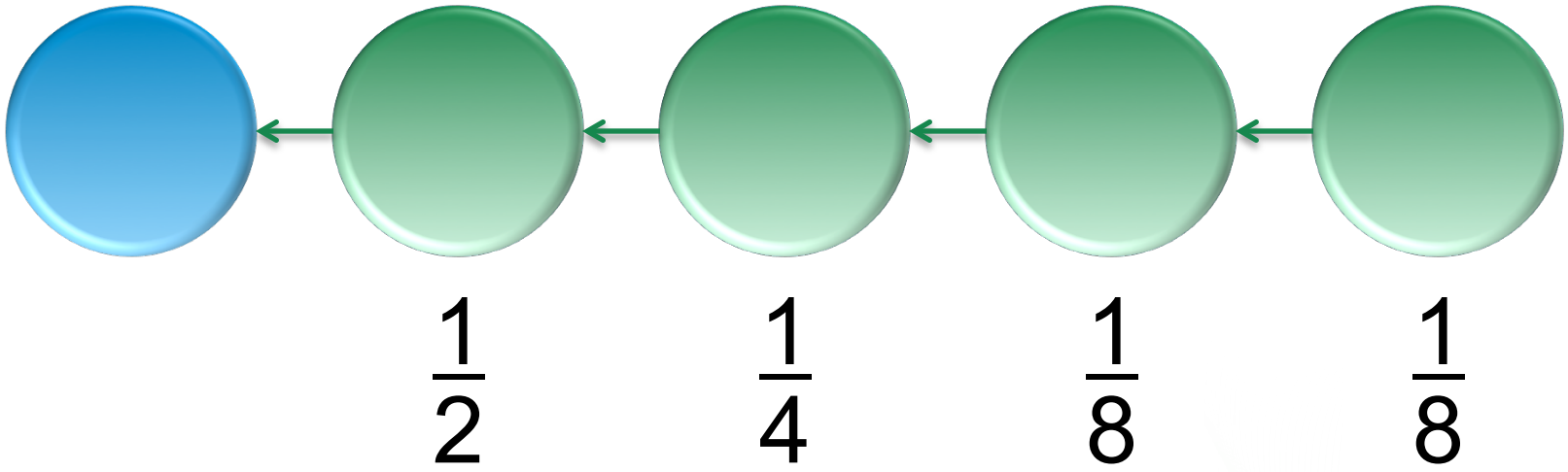
Image taken from the presentation for “Blue Waters I/O Performance” at CUG 2013
https://cug.org/proceedings/cug2013_proceedings/includes/files/pap129-file2.pdf

How fast are Gemini Links?

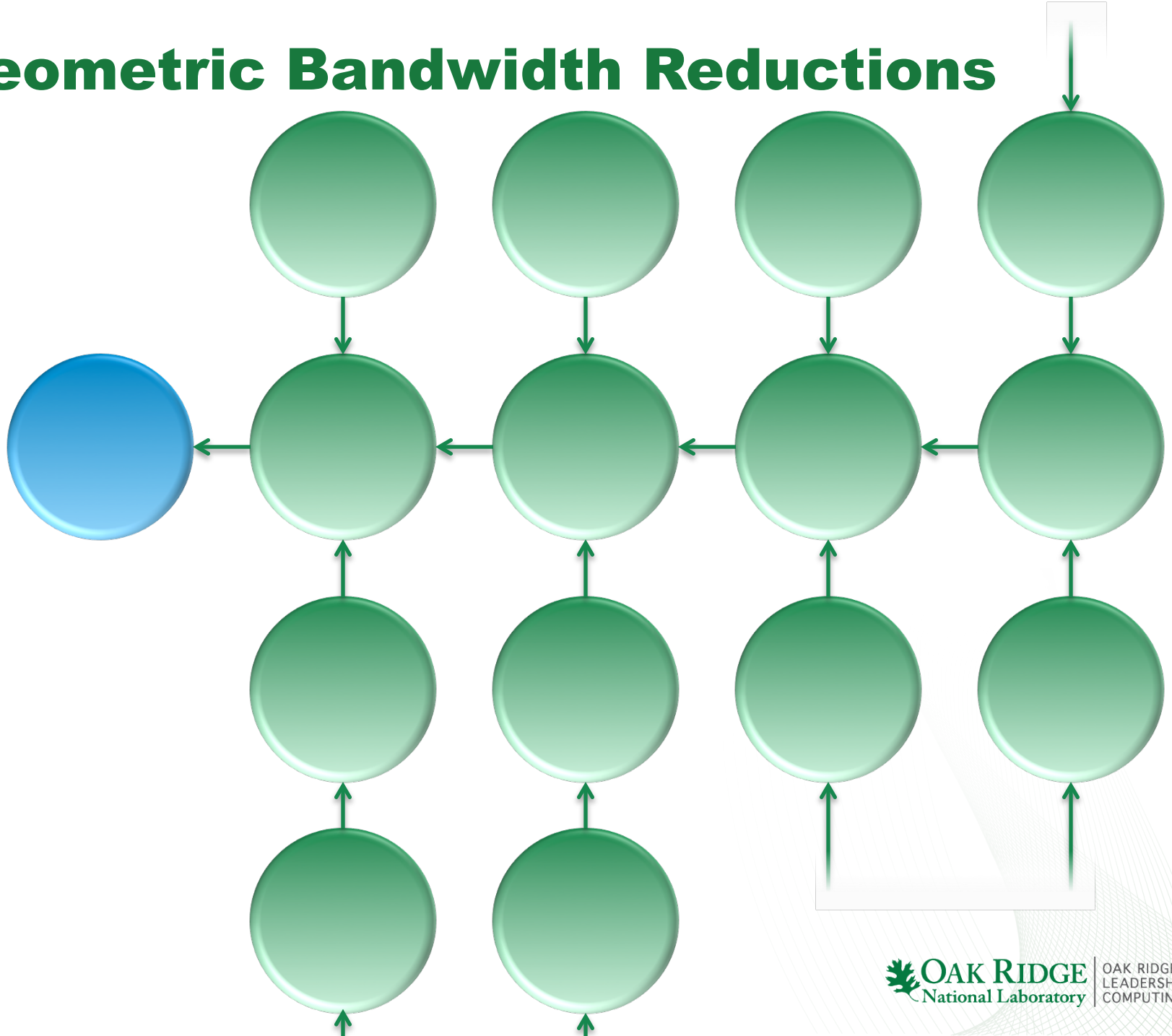
- It depends!
- Link speed depends on link type
- Protocol overhead is around 35% for large messages

Link Type	Link Data Rate	Number Links	Raw Bitrate	Data Rate
Y-Mezzanine	6.25 gbps	12	9.375 GB/s	~ 6 GB/s
Z-Backplane	5.0 gbps	24	15 GB/s	~ 9.75 GB/s
X, Z Cable	3.125 gbps	24	9.375 GB/s	~ 6 GB/s
Y Cable	3.125 gbps	12	4.6875 GB/s	~ 3 GB/s

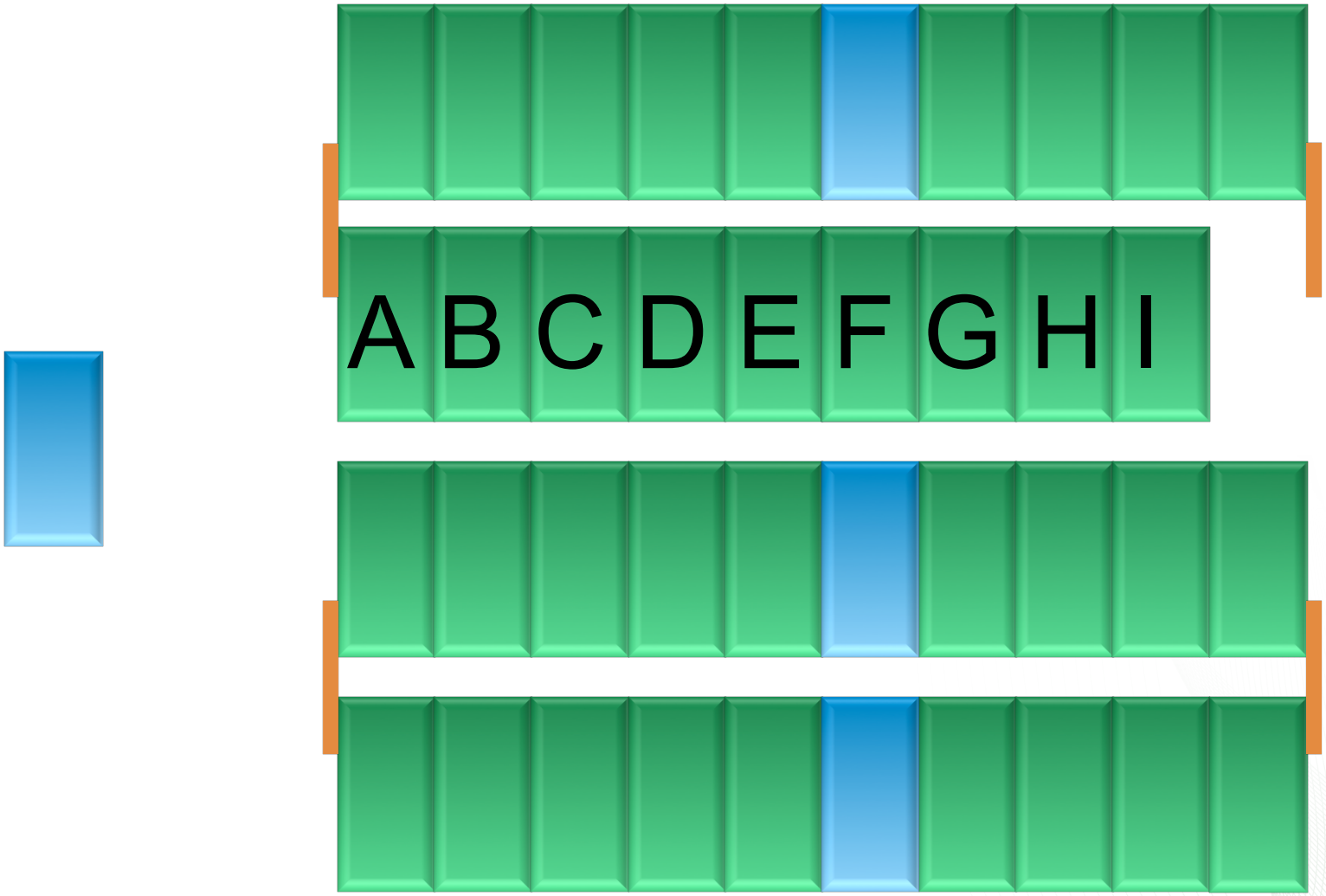
Geometric Bandwidth Reductions



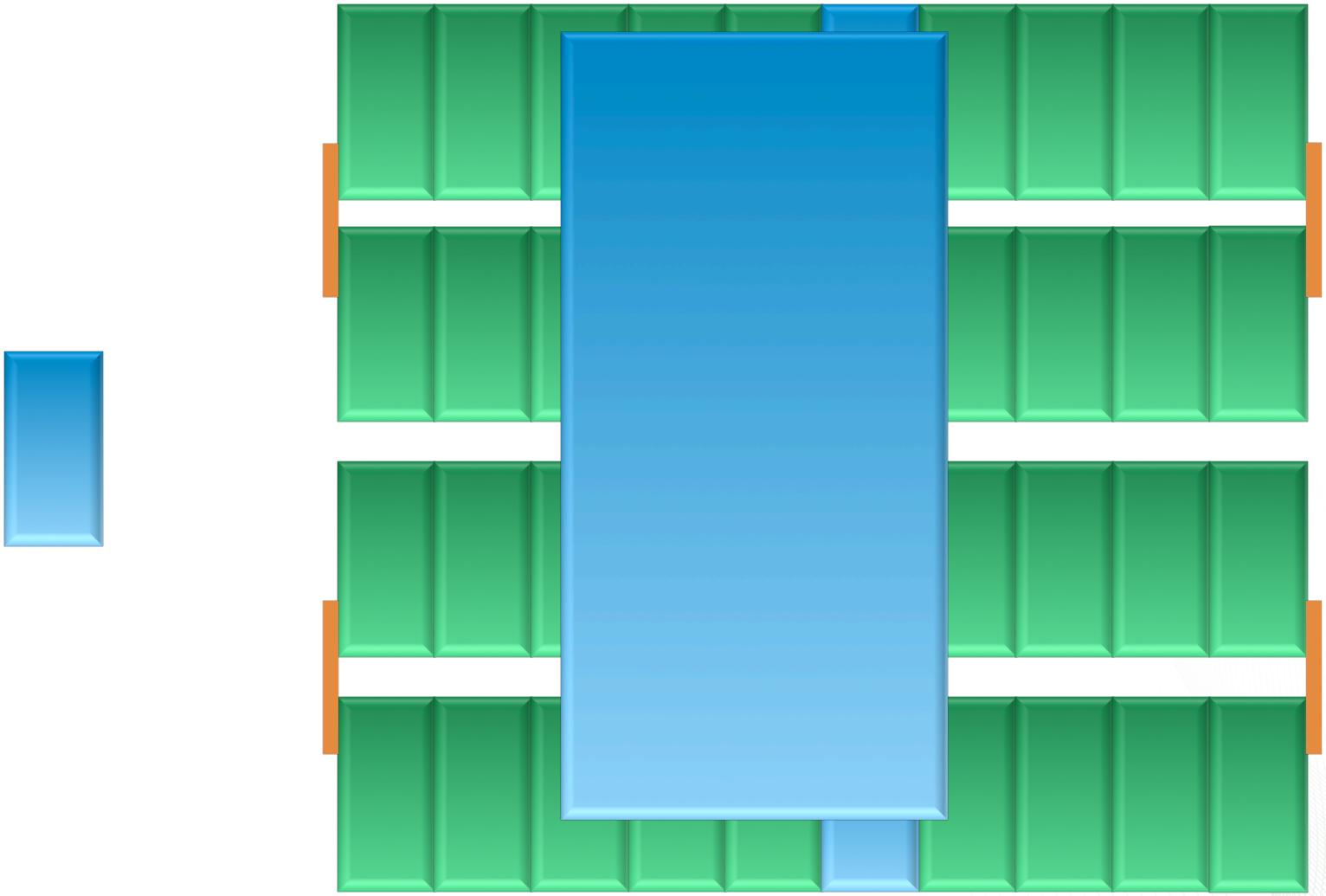
Geometric Bandwidth Reductions



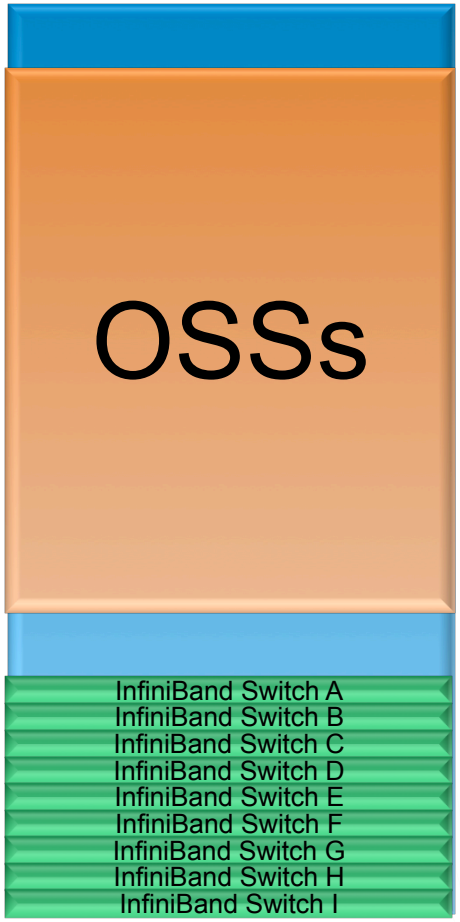
Atlas Layout



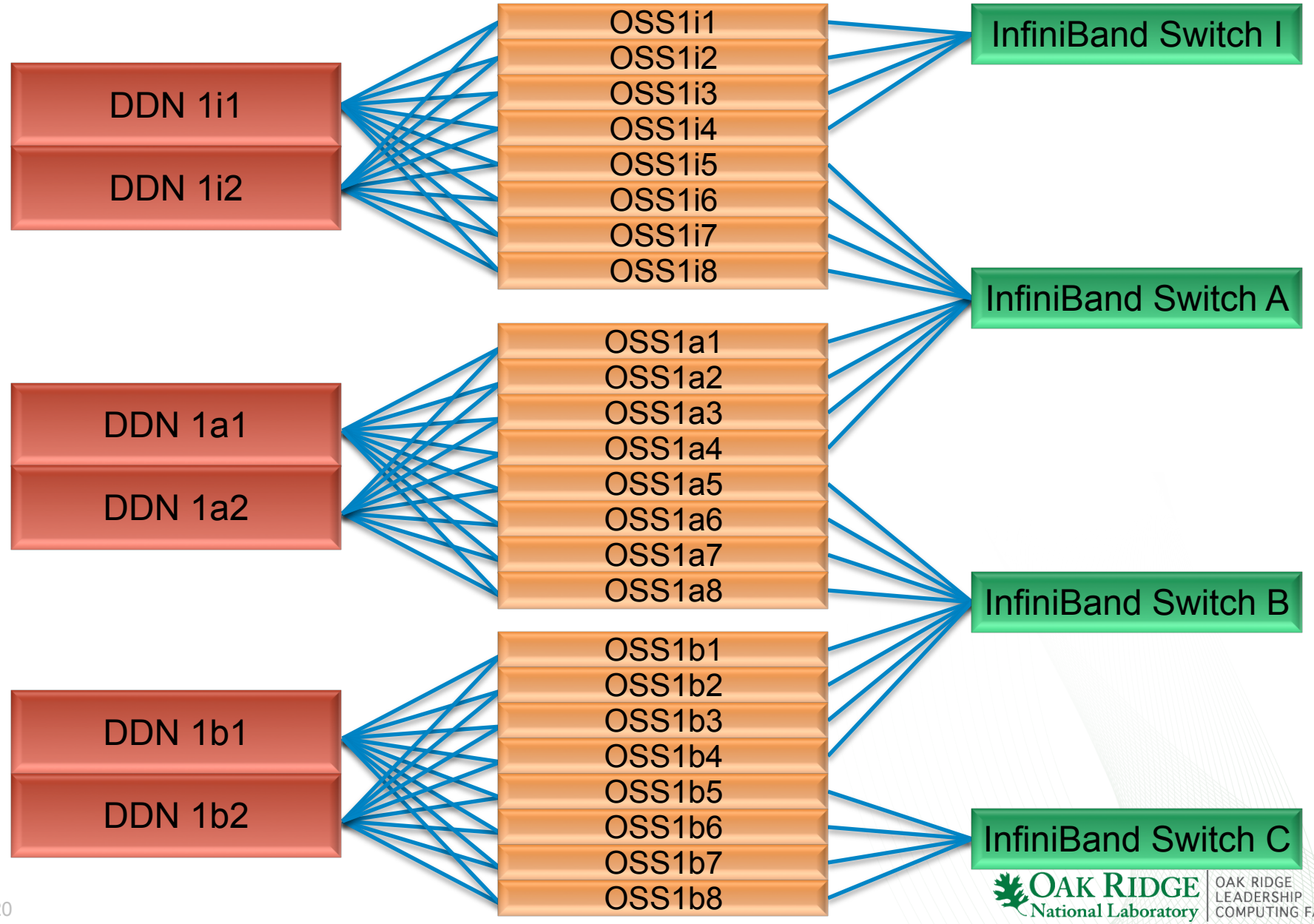
Atlas Layout



Atlas Layout



OSS Connections



How many routers?

Performance Goal: > 1TB/s

÷

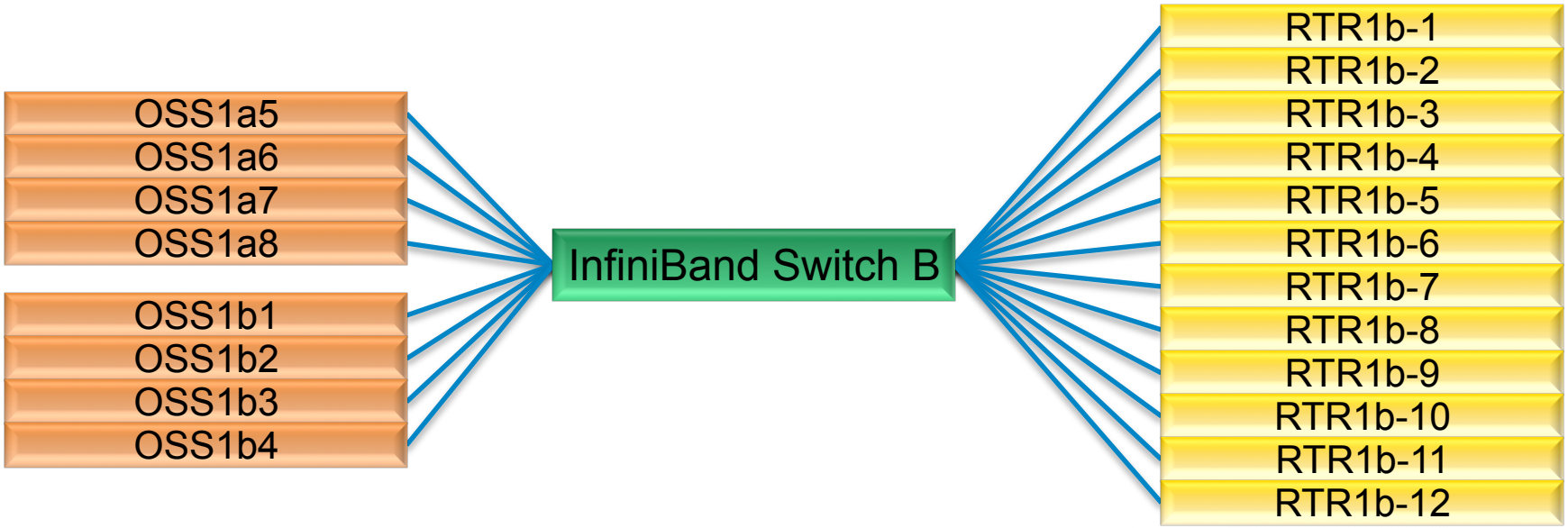
Router Performance: 2.6GB/s

=

Required Router Count: > 385

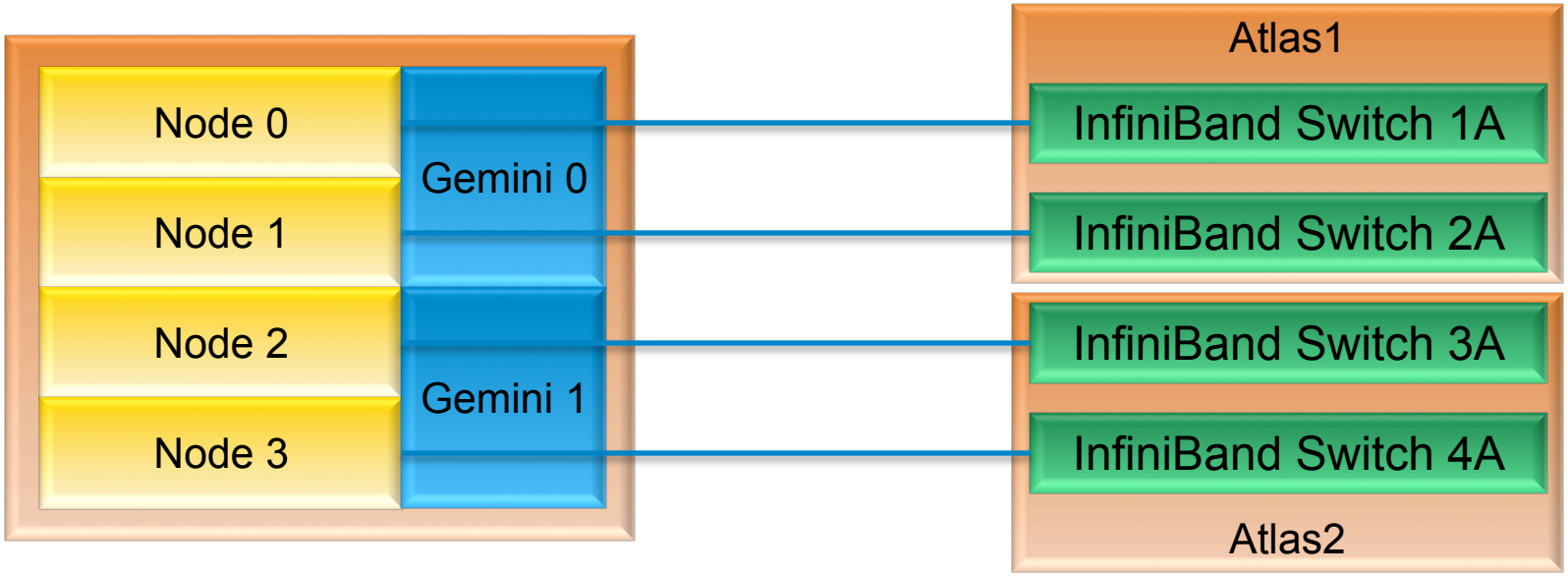
Deployed Router Count: > 432

FGR Group

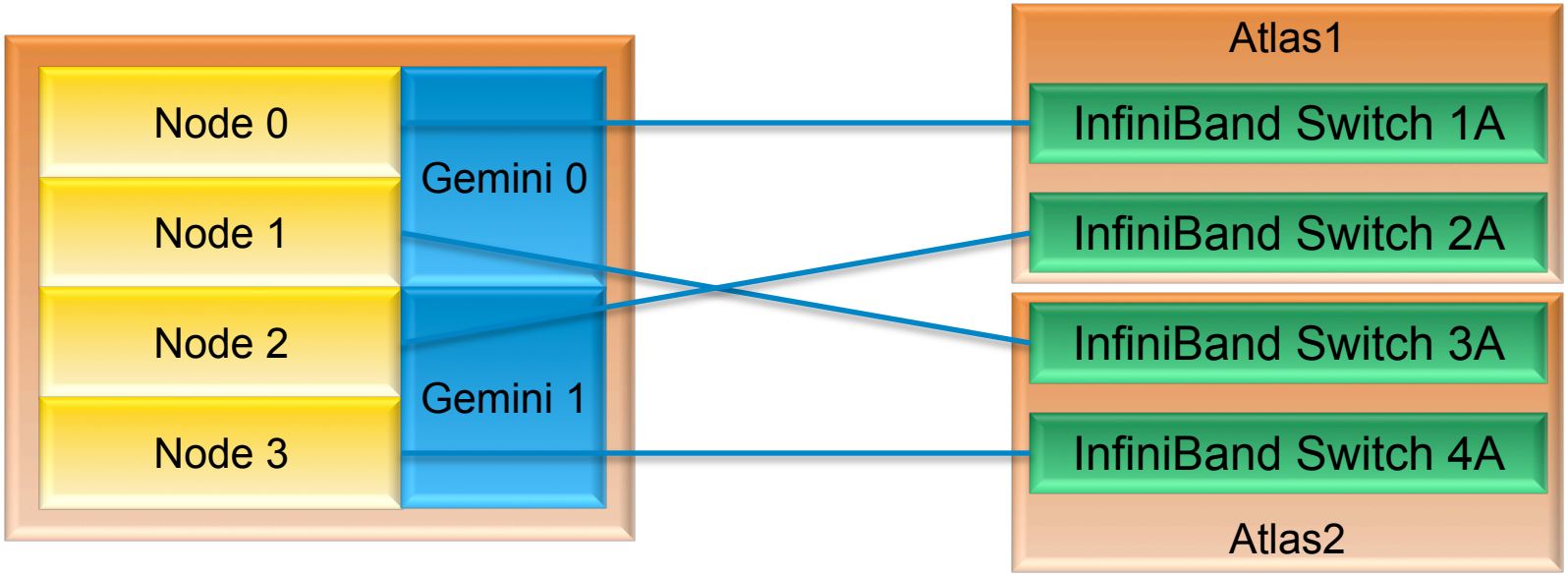


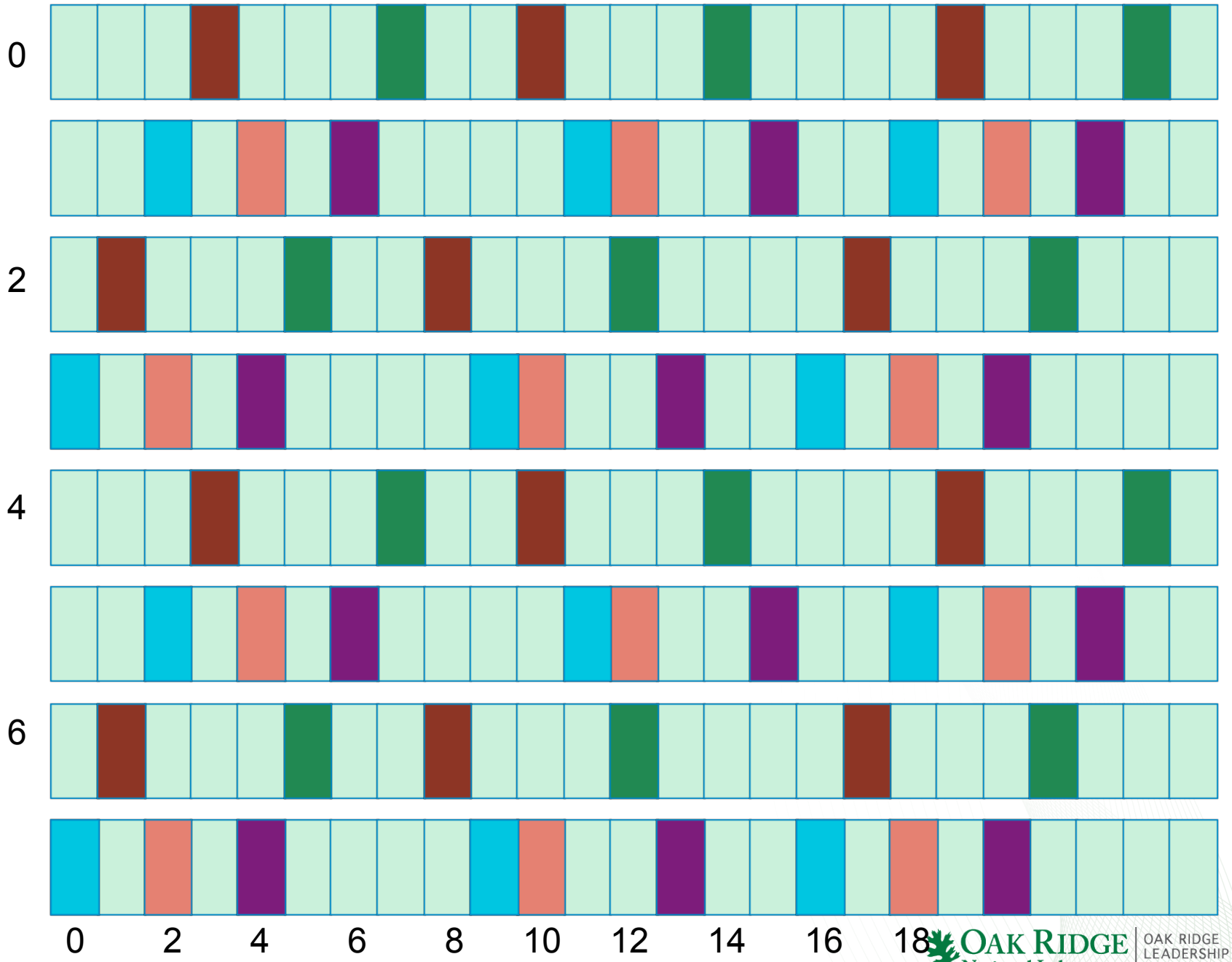
8:12 Ratio

Router Module

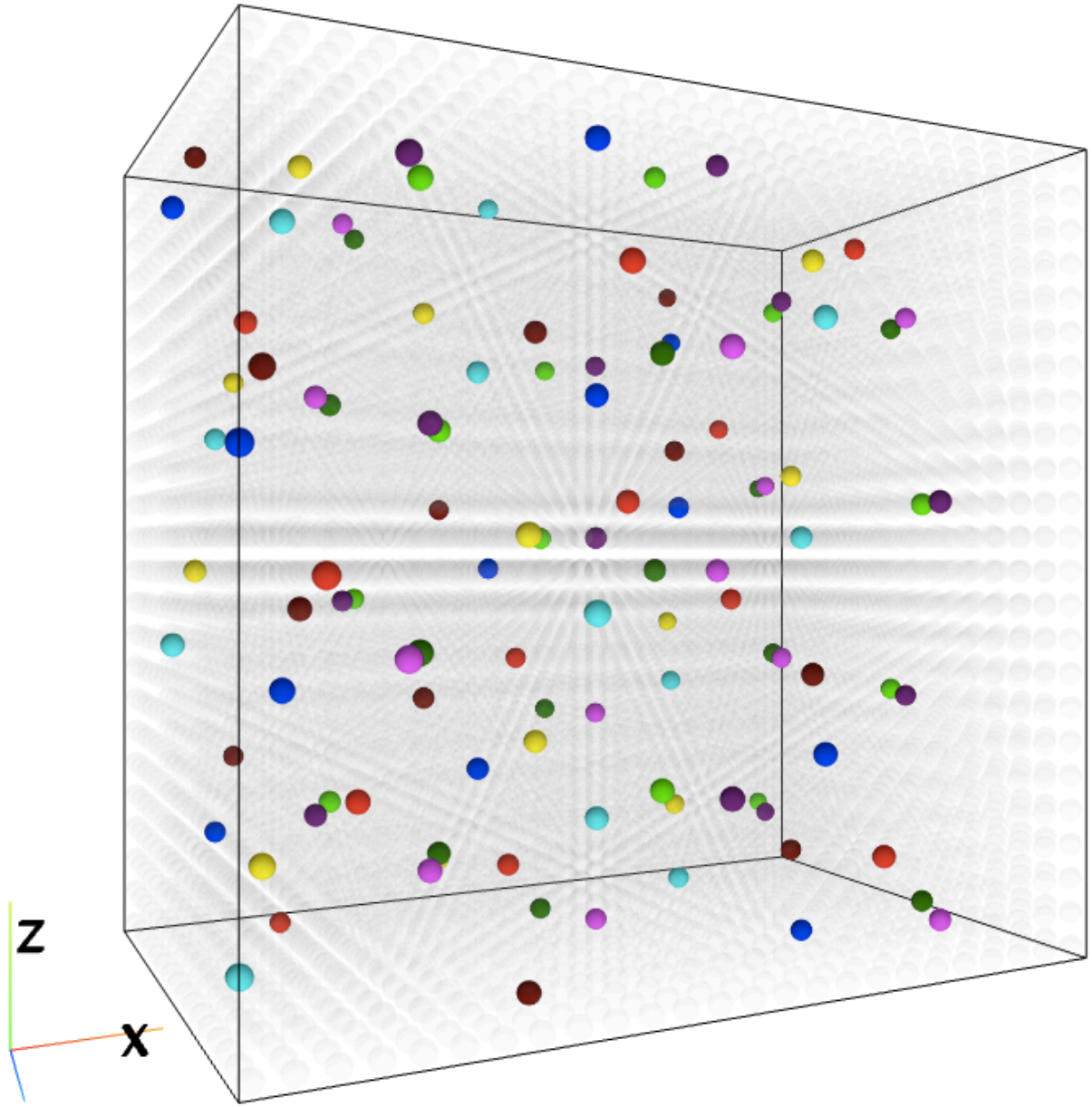


Router Module





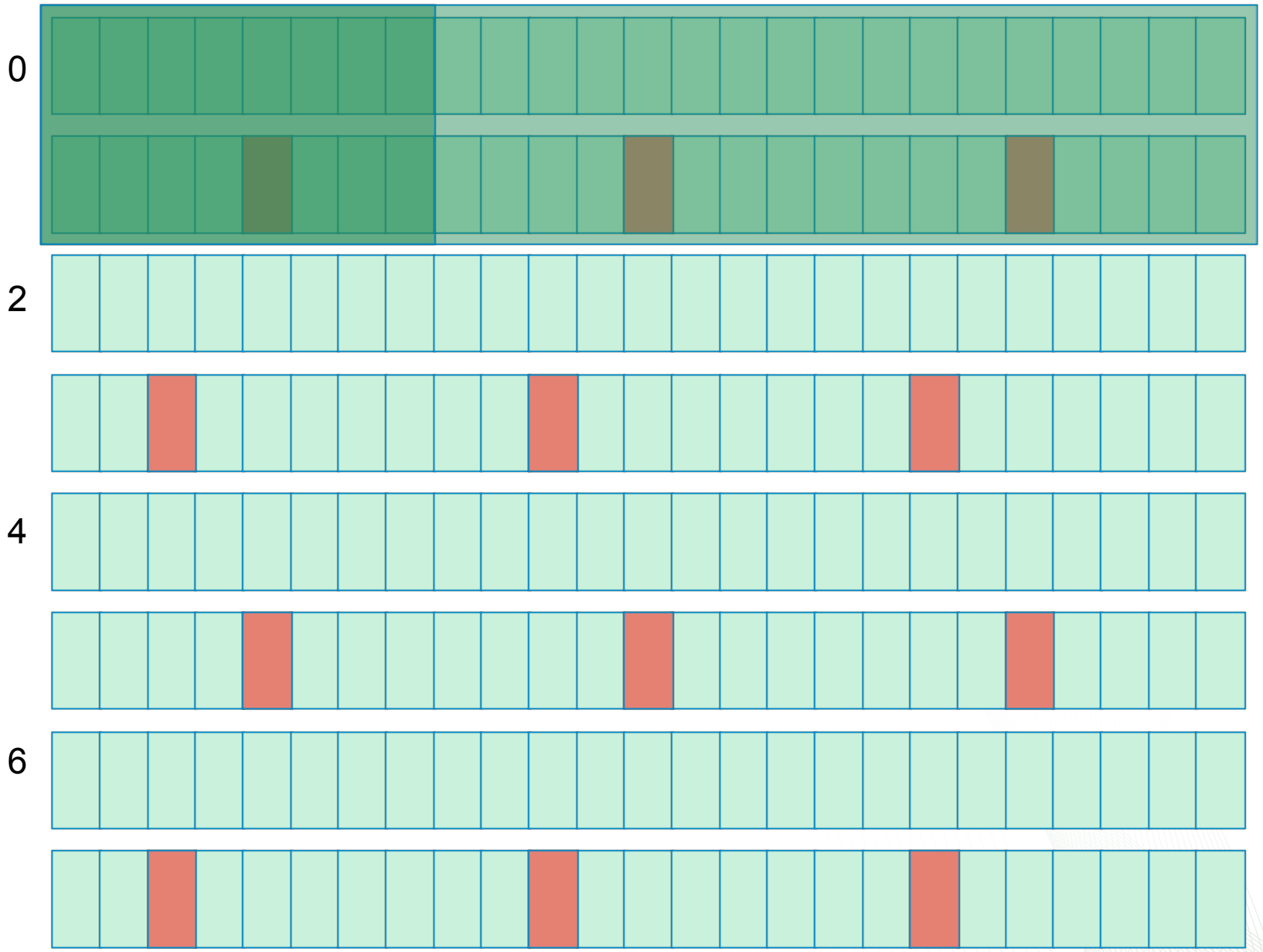
Router Placement

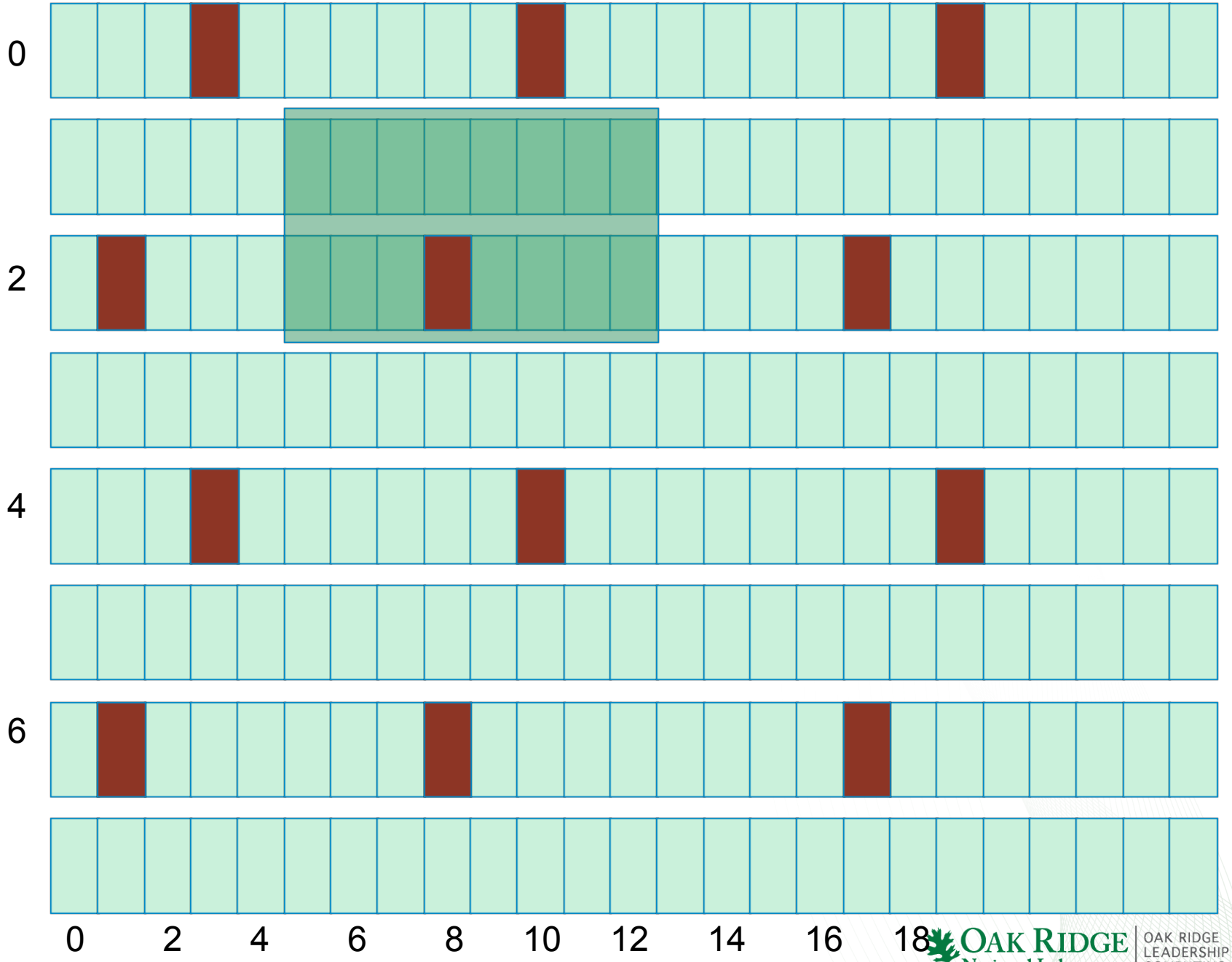


Fine-Grained Routing Configuration

36

12





FGR Best Practices

- Create large FGR groups
- Create FGR groups in both directions
- Verify all settings and physical connections

Questions?

ezellma@ornl.gov