Modular AWG-based Interconnection for Large-Scale Data Center Networks

Tong Ye, Tony T. Lee, Mao Ge, and Weisheng Hu

yetong@sjtu.edu.cn

State Key Lab of Advanced Optical Communications and Networks
Shanghai Jiao Tong University

Outline

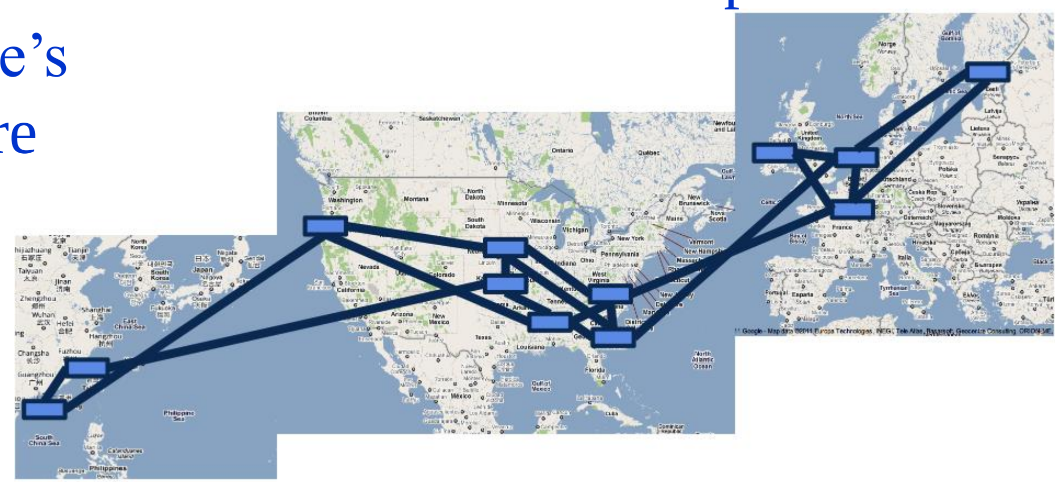
- Background
- AWG-based Interconnection
- Modular AWG-based Interconnection
- Application to Data Center Networks
- Conclusion

Data Centers Play Important Roles

- World-wide information service infrastructure

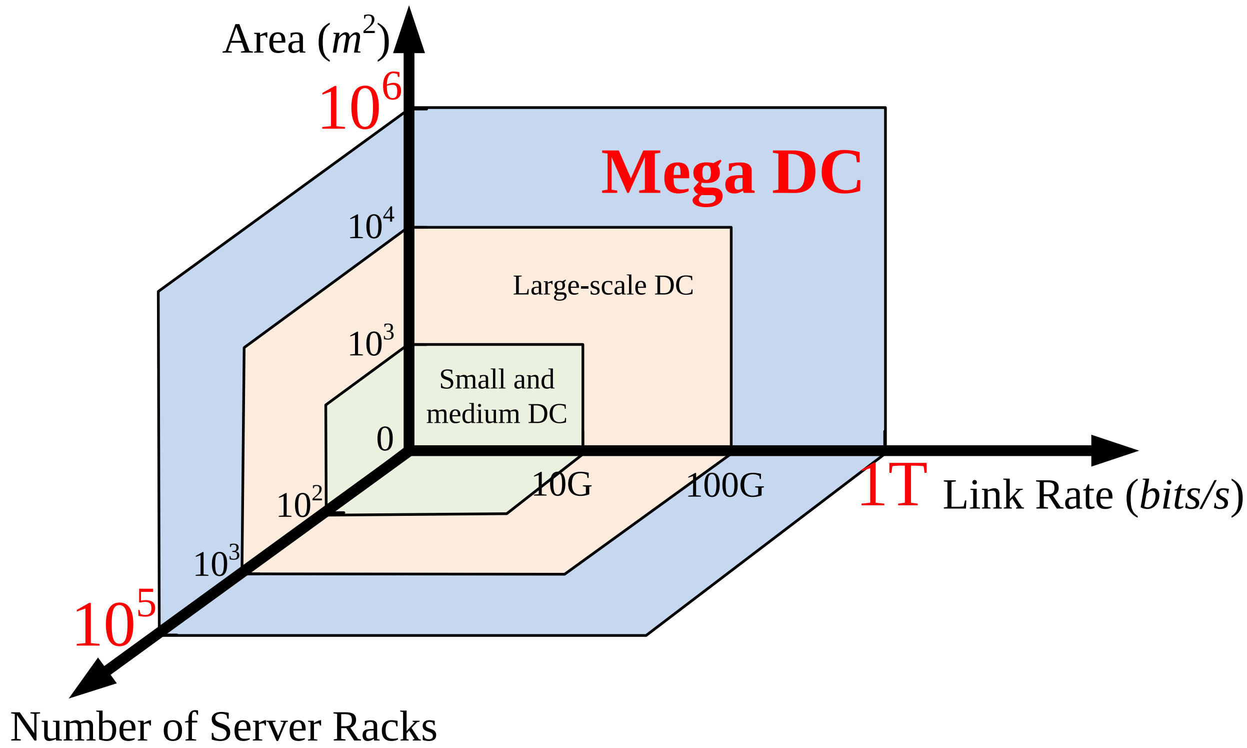
Amazon Web Service's
Global Infrastructure

Google World-wide
Data Center Map



[1] http://datacenterfrontier.com/regional-data-center-clusters-power-amazons-cloud/
[2] Sushant Jain et. al., "B4: Experience with a Globally-deployed Software Defined Wan", ACM SIGCOMM, Oct. 2013, pp. 3-14.

Footprint of Data Centers (DCs)

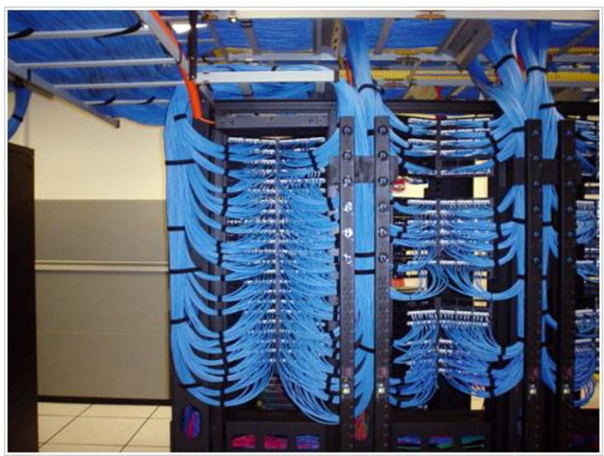


- A mega DC requires a large number of long cables with very high capacity

OIDA/CIAN Data center Workshop, "Quantitative metrics for data centers," in Proc. OFC, 2012.

Cabling Problem

- Cable maintenance is extremely difficult, when
  - network connections change
  - line failures occur

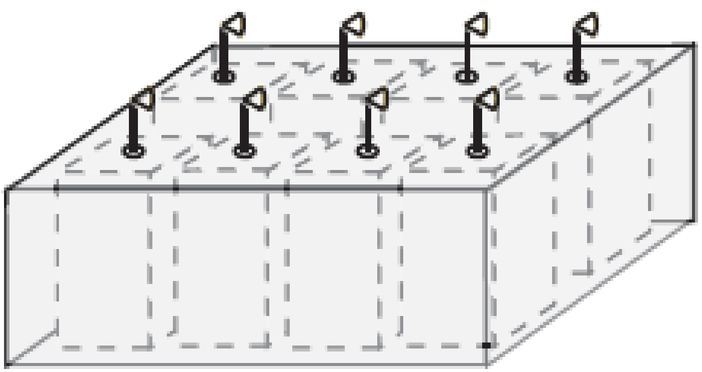


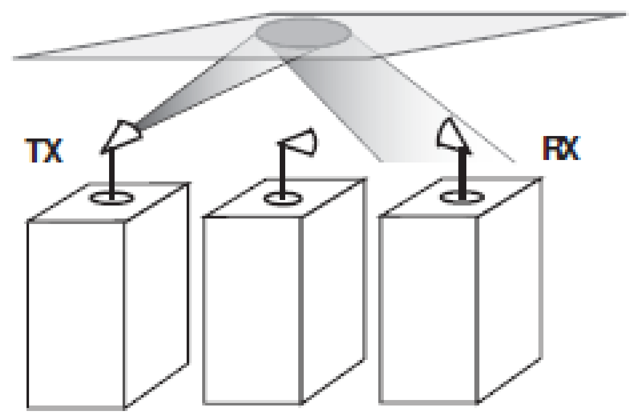
Eventually, cables become a terrible monster…

[1] N. Farrington, E. Rubow, and A. Vahdat, "Data center switch architecture in the age of merchant silicon," in Proc. IEEE HOTI, Aug. 2009.
[2] www.hpl.hp.com/techreports/2015/HPL-2015-8.html
[3] J. Mudigonda, P. Yalagandula, and J. C. Mogul, "Taming the flying cable monster: A topology design and optimization framework for data- center networks," in Proc. ATC, Jun. 2011.

Solution: Wireless Links

- Pros: reduce number of cables
- Cons:
  - Low bandwidth (~Gb/s)
  - Serious radio interference

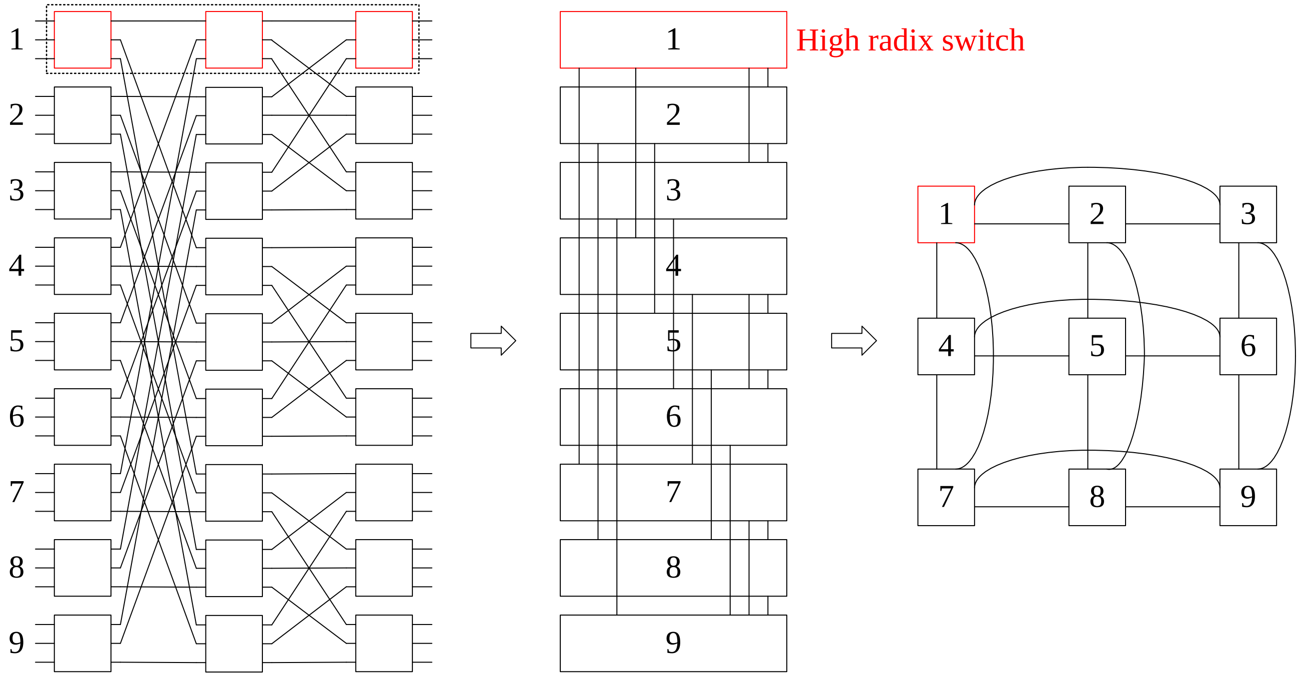




[1] K. Ramachandran, R. Kokku, R. Mahindra, and S. Rangarajan, “60 GHz data-center networking: Wireless => worry less?” Technical Report, NEC, 2008.
[2] Y. Cui, H. Wang, X. Cheng, and B. Chen, “Wireless data center networking,” IEEE Wireless Commun. Mag., vol. 18, no. 6, pp. 46– 53, Dec. 2011.
[3] N. Hamedazimi et al., “Firefly: A reconfigurable wireless data center fabric using free-space optics,” in Proc. ACM SIGCOMM, Oct. 2014.

Solution: Optimal Device Allocation

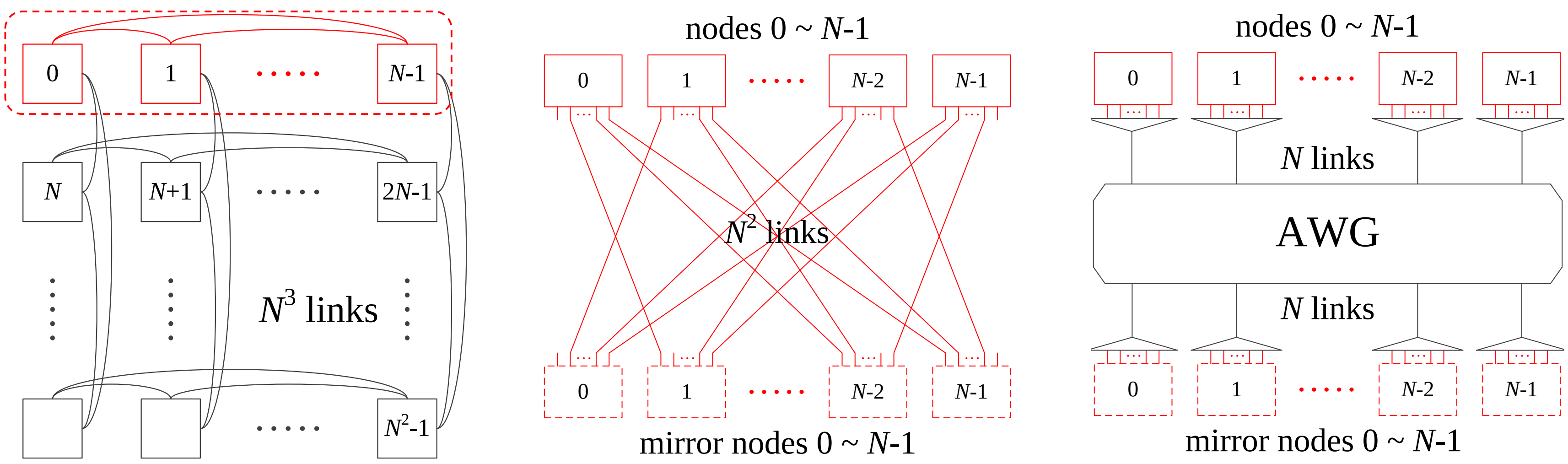
- Idea: combine several switches to form a high radix switch, but
  - specific for Butterfly networks (not universal)
  - reduce the number of cables only by half (not scalable)



J. Kim, W. J. Dally, and D. Abts, "Flattened butterfly: a cost- effective topology for high-radix networks," in Proc. ACM ISCA, Jun. 2007.

Solution: Optical Method

- Replace links of each full mesh by an arrayed waveguide grating (AWG)
  - Pros: reduce cabling complexity + bandwidth guaranteed √
  - Cons: AWG is not scalable if network is very large



M. Csernai, F. Ciucu, R. P. Braun, and A. Gulyas, Towards 48-Fold Cabling Complexity Reduction in Large Flattened Butterfly Networks, in Proc. INFOCOM 2015.

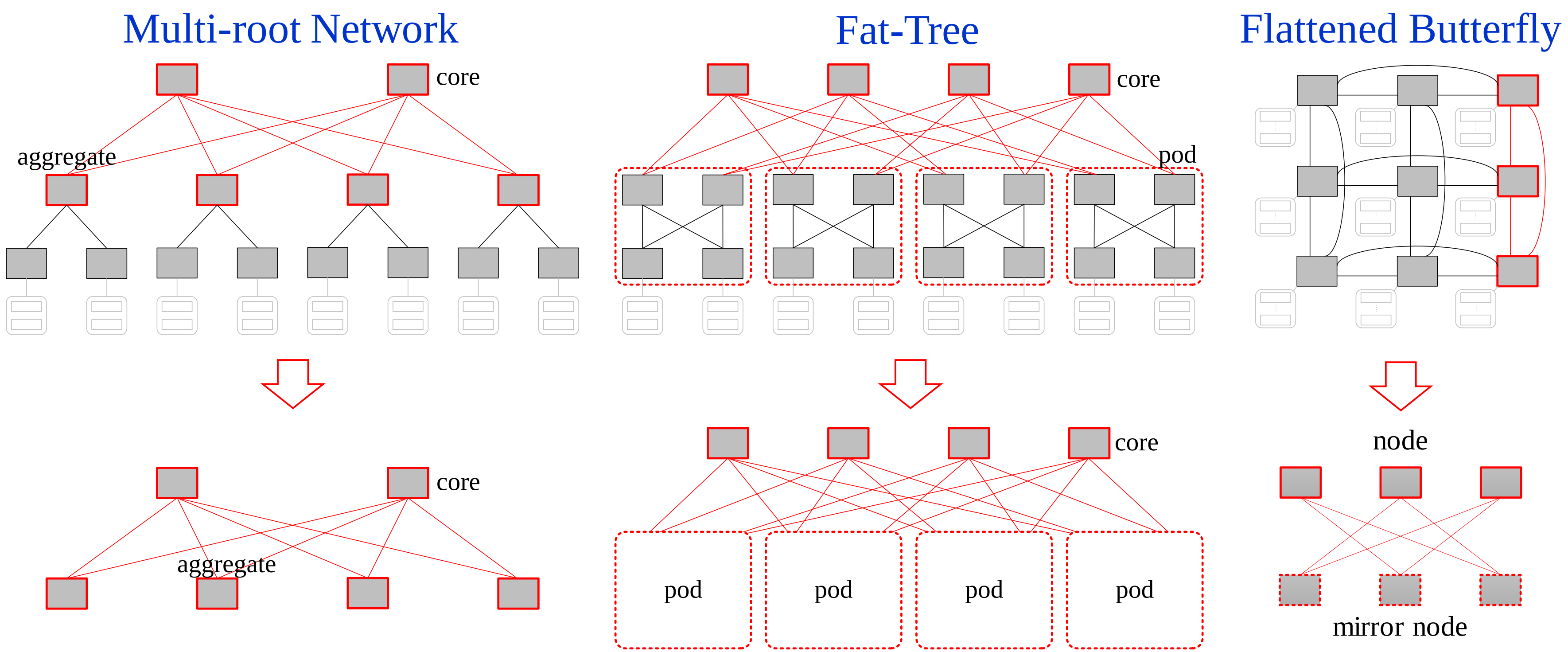
The Goal of Our Work

- Achieve modular AWG-based interconnection:
  - Substantially reduce cabling complexity, while preserving function of original DC networks
  - Scalable even when size of DC networks is very large
  - Can be applied to different DC networks

Outline

- Background
- AWG-based Interconnection
- Modular AWG-based Interconnection
- Application to Data Center Networks
- Conclusion

Topology of Existing Networks

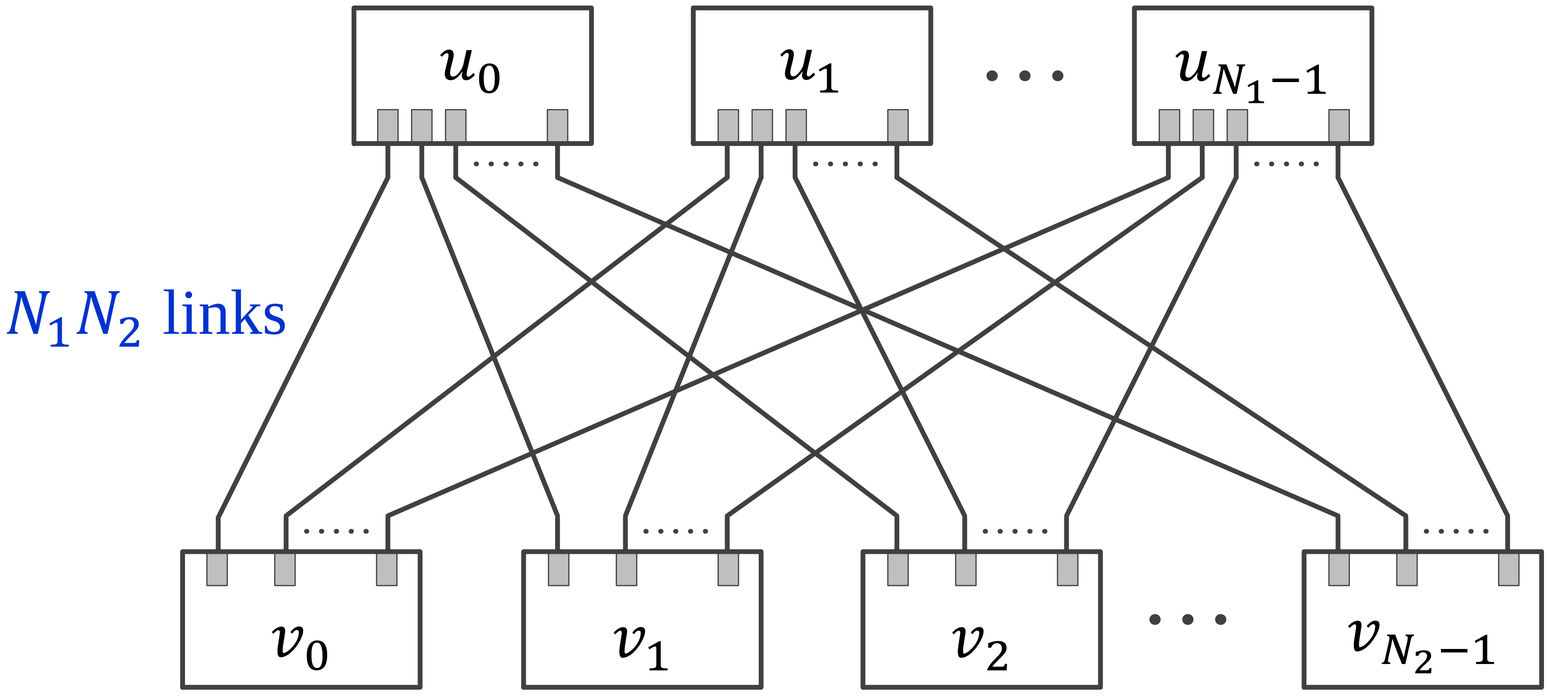


- Different networks have the similar subnetwork

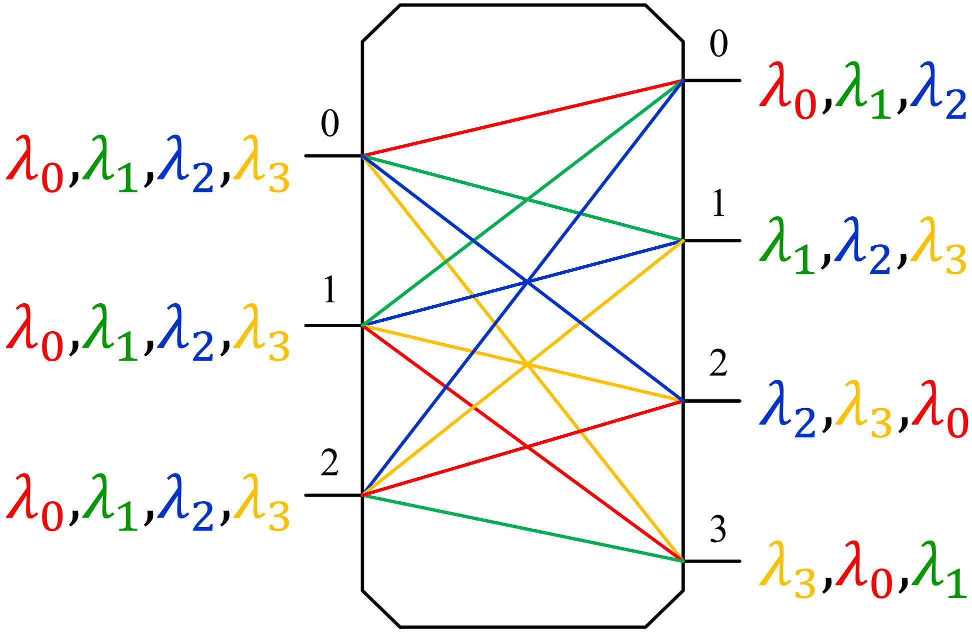
[1] M. F. Bari et al., "Data center network virtualization: a survey," IEEE Commun. Surveys Tuts., vol. 15, no. 2, pp. 909–928, May 2013.
[2] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," in Proc. ACM SIGCOMM, Aug. 2008.
[3] Z. Zhu, S. Zhong, L. Chen, and K. Chen, "Fully programmable and scalable optical switching fabric for petabyte data center", Opt. Express, vol. 32, no. 3, pp. 3563-3580, Feb. 2015.

Banyan-Type Subnetwork: $\mathcal{N}_A$

- Two disjoint node sets
- Exact one fiber link from a node in one set to that in another set



$N_1 \times N_2$ AWG



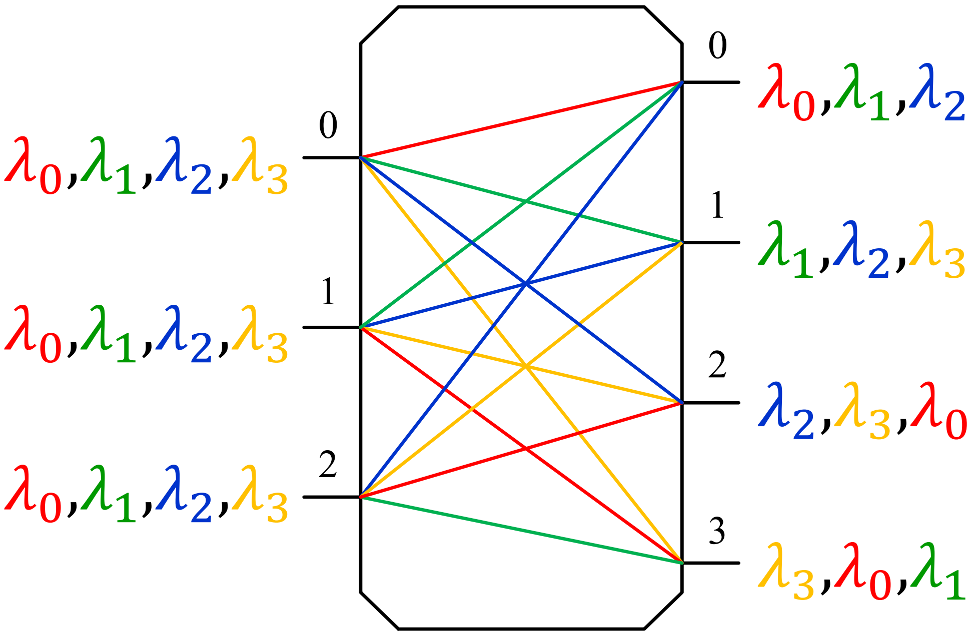
Wavelength (λ) set $\Lambda = \{\lambda_0, \lambda_1, \lambda_2, \lambda_3\}$

- Passive => consume no power
- Provide $N_1 N_2$ links between inputs and outputs

Non-blocking Routing Property

Output# (j) is determined by input# (i) & λ# (k):

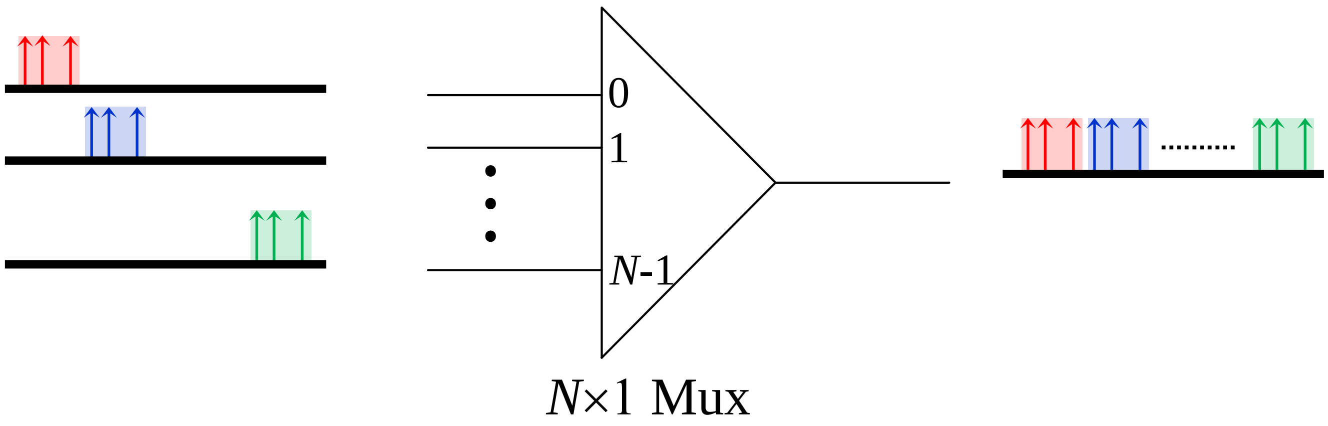
$$j = [k - i]_{|\Lambda|} \overset{\text{def}}{=} (k - i) \bmod |\Lambda|$$

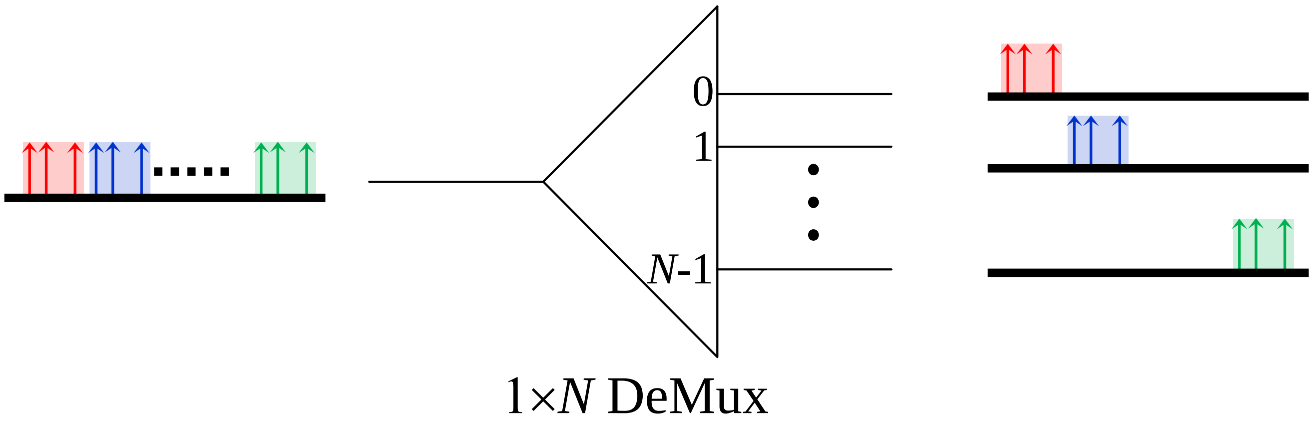


	OUT 0	OUT 1	OUT 2	OUT 3
IN 0	λ_0	λ_1	λ_2	λ_3
IN 1	λ_1	λ_2	λ_3	λ_0
IN 2	λ_2	λ_3	λ_0	λ_1

Cyclic Latin Square if $N_1 = N_2$

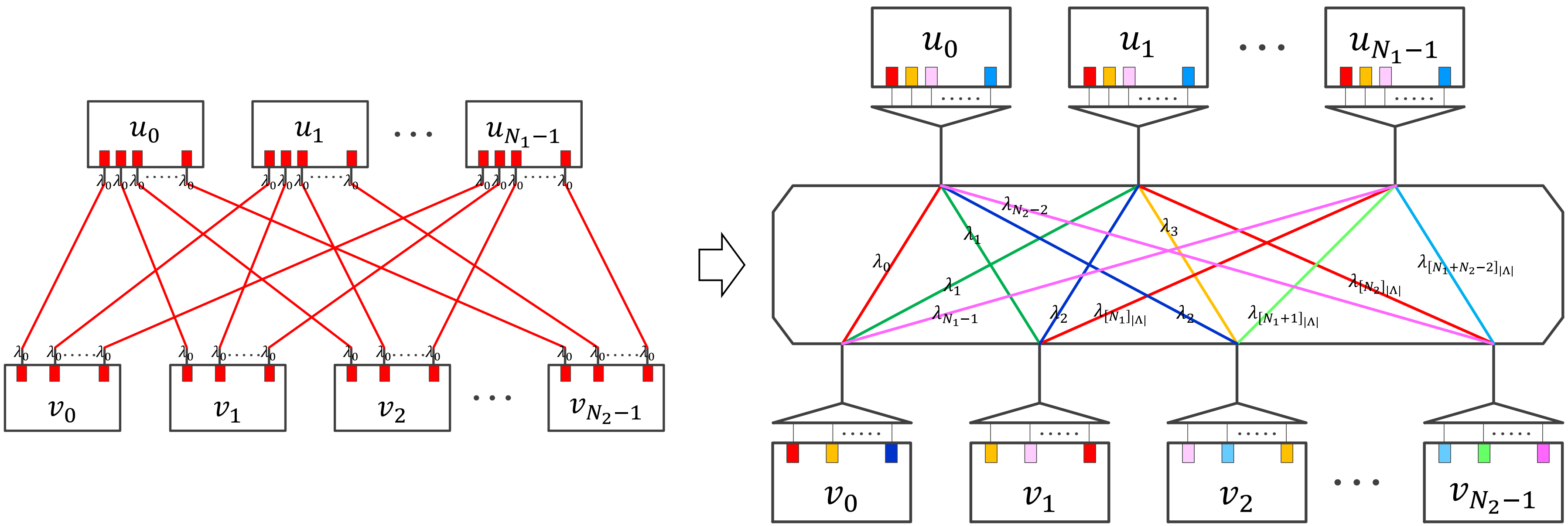
$N \times 1$ AWG: λ Mux/DeMux





AWG-based Interconnection

- Replacing fiber links in $\mathcal{N}_A$ by an AWG yields a network $\mathcal{N}_B$



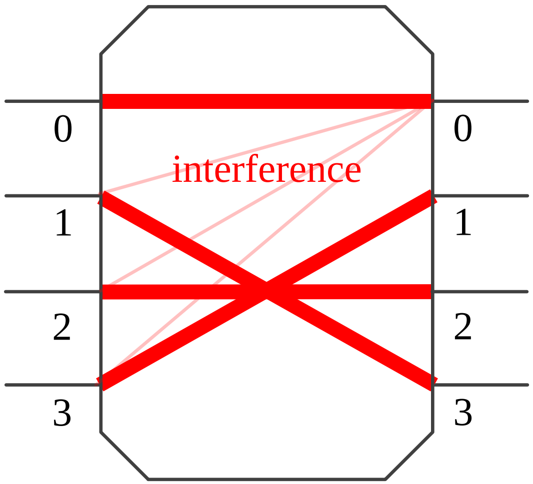
$\mathcal{N}_A$: $N_1 N_2$ fiber links

$\mathcal{N}_B$: $N_1 + N_2$ fiber links

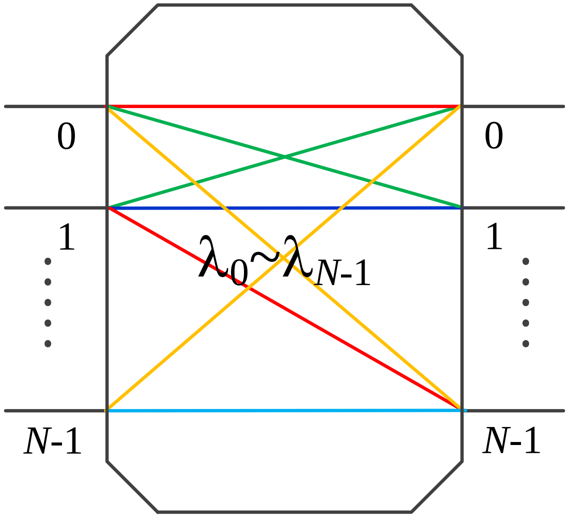
M. Csernai, F. Ciucu, R. P. Braun, and A. Gulyas, Towards 48-Fold Cabling Complexity Reduction in Large Flattened Butterfly Networks, in Proc. INFOCOM 2015.

Limitations of $N \times N$ AWG

- If N is very large:
  - In-band crosstalk is prominent (bad physical-layer performance)
  - Synthesis is very difficult
  - A large number of λs are required



signals at the same λ interferes with each other



R. Gaudino, G. G. Castillo, F. Neri, and J. Finochietto, in *Proc. IEEE ICC*, May 2008, pp. 5331–5337.

Modular AWG-based Interconnection

- Phase 1: AWG decomposition
  - Suppress in-band crosstalk
  - Cut down synthesis difficulty

- Phase 2: Wavelength reuse
  - Reduce number of required wavelengths

Outline

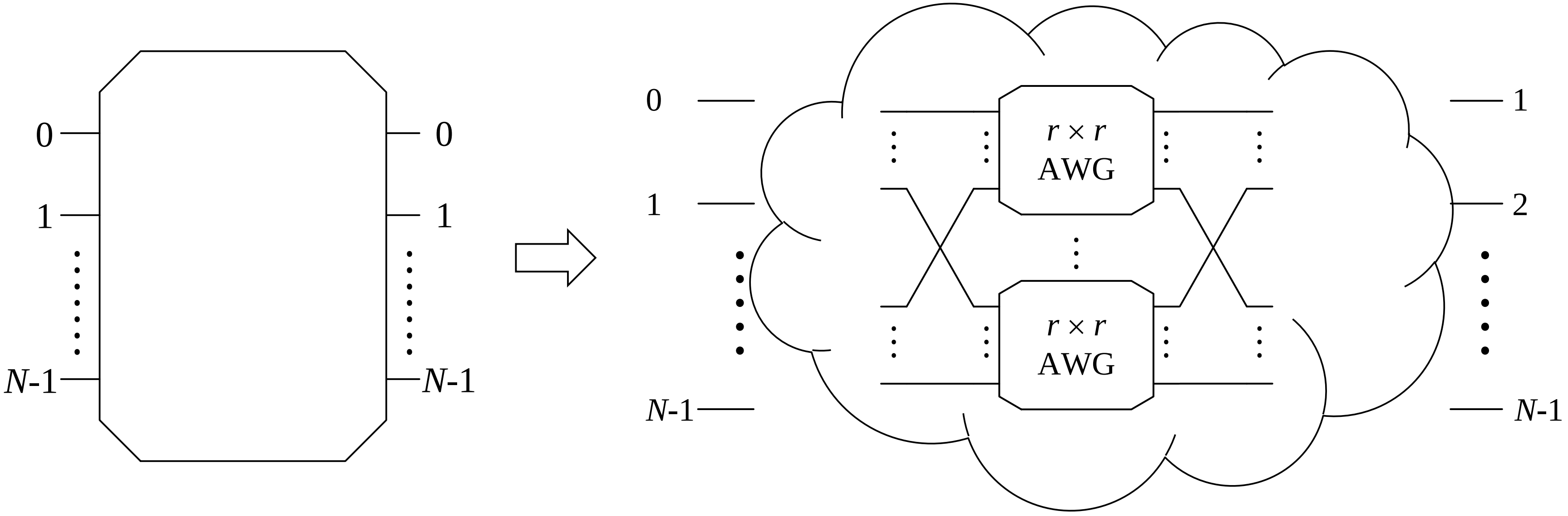
- Background
- AWG-based Interconnection
- Modular AWG-based Interconnection
  - Phase 1: AWG Decomposition
  - Phase 2: Wavelength Reuse
- Application to Data Center Networks
- Conclusion

AWG Decomposition

- $N \times N$ AWG $\Rightarrow$ $N \times N$ network of AWGs:
  - same routing property

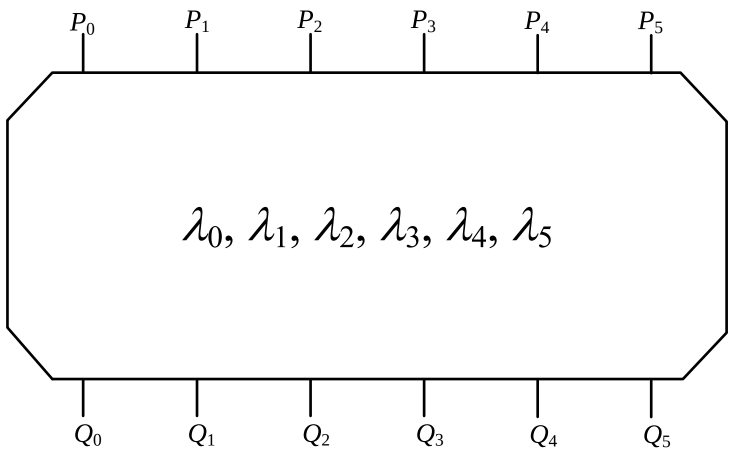
    <=> output# is uniquely determined by input# and λ#

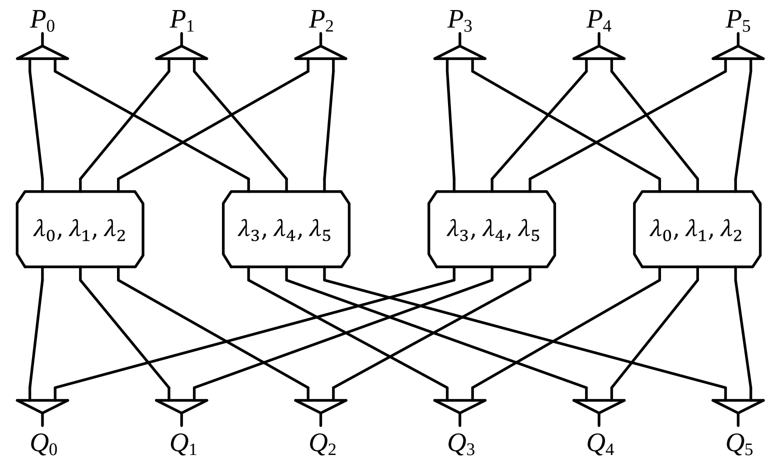
$r < N$



Tong Ye, T. T. Lee, and Weisheng Hu, IEEE Journal of Lightwave Technology, vol. 30, no. 13, pp. 2125-2133, Jul. 2012.

Example of Decomposition





	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_3	λ_4	λ_5	λ_0
P_2	λ_2	λ_3	λ_4	λ_5	λ_0	λ_1
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_0	λ_1	λ_2	λ_3
P_5	λ_5	λ_0	λ_1	λ_2	λ_3	λ_4

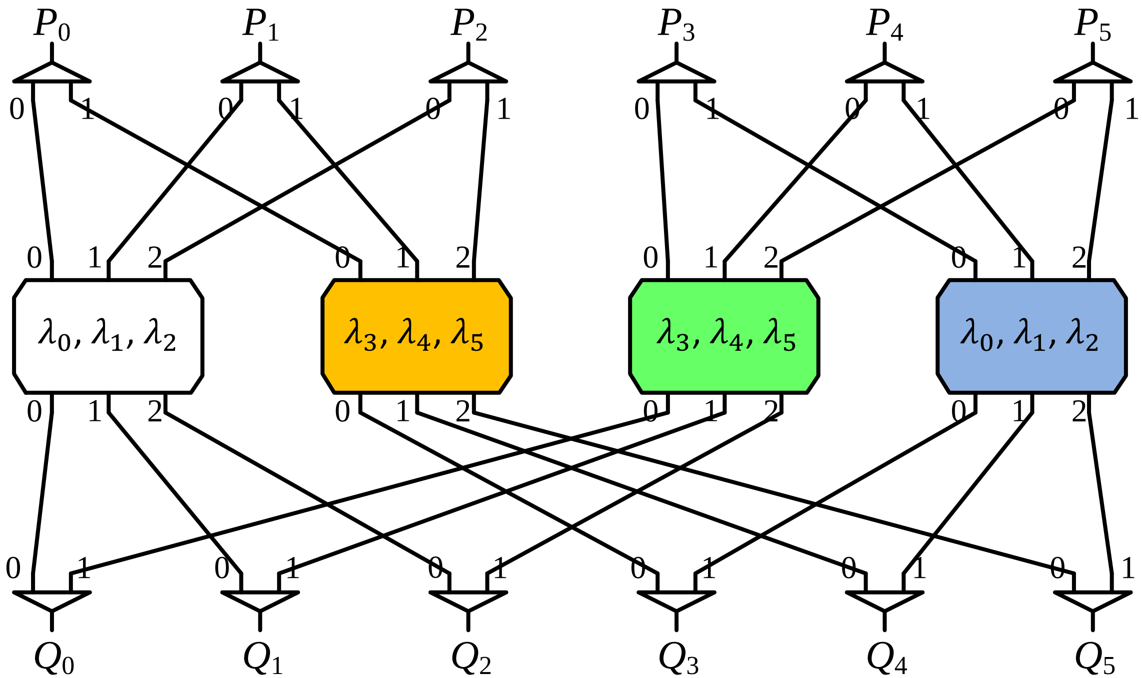
6×6 cyclic Latin square

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

A

6×6 Latin square

Example: Observation 1



	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$ $\mathbf{A}_{01}$ $\mathbf{A}_{10}$ $\mathbf{A}_{11}$

- $\mathbf{A}$ consists of 2^2 3×3 cyclic Latin squares
- Each square is associated with a 3×3 AWG

Example: Observation 2

$\mathbf{M}_0$

λ_0	λ_1	λ_2
λ_1	λ_2	λ_0
λ_2	λ_0	λ_1

$\mathbf{M}_1$

λ_3	λ_4	λ_5
λ_4	λ_5	λ_3
λ_5	λ_4	λ_3

$\mathbf{A}$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

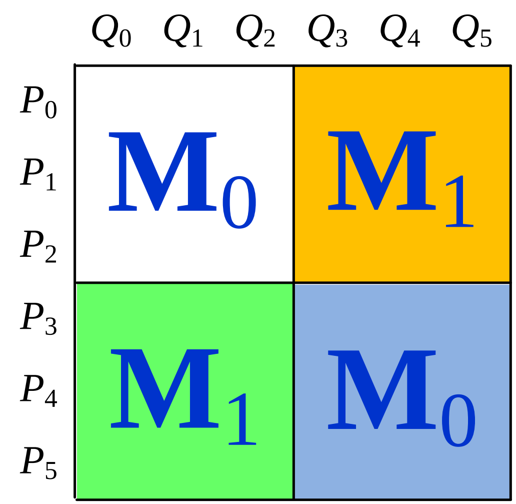
$\mathbf{A}_{00}$, $\mathbf{A}_{01}$, $\mathbf{A}_{10}$, $\mathbf{A}_{11}$

$\mathbf{M}_1$

λ_3	λ_4	λ_5
λ_4	λ_5	λ_3
λ_5	λ_4	λ_3

$\mathbf{M}_0$

λ_0	λ_1	λ_2
λ_1	λ_2	λ_0
λ_2	λ_0	λ_1



recursively cyclic

- $\mathbf{M}_0$ is defined on λ-set $\{\lambda_0, \lambda_1, \lambda_2\}$
- $\mathbf{M}_1$ is defined on λ-set $\{\lambda_3, \lambda_4, \lambda_5\}$

Matrix-based AWG Decomposition

- Initialization:
  - Define n $r \times r$ cyclic Latin squares ($nr = N$)

$$\mathbf{M}_0, \mathbf{M}_1, \cdots, \mathbf{M}_{n-1}$$

  - Specify an $N \times N$ Latin square $\mathbf{A}$ with n^2 $r \times r$ blocks

$$\mathbf{A}_{ab} = \mathbf{M}_{[a+b]_n}$$

- Construct an AWG network according to **A**:
  - S1. Central stage construction
  - S2. Upper-layer stage construction
  - S3. Lower-layer stage construction

Initialization: $\mathbf{M}_0, \mathbf{M}_1, \cdots, \mathbf{M}_{n-1}$

- Specify n $r \times r$ cyclic Latin squares, where $N = 6, n = 2, r = 3$
  - $\mathbf{M}_0$ is defined on λ-set $\Lambda_0 = \{\lambda_0, \lambda_1, \lambda_2\}$
  - $\mathbf{M}_1$ is defined on λ-set $\Lambda_1 = \{\lambda_3, \lambda_4, \lambda_5\}$

$\mathbf{M}_0$

λ_0	λ_1	λ_2
λ_1	λ_2	λ_0
λ_2	λ_0	λ_1

$\{\lambda_0, \lambda_1, \lambda_2\}$

$\mathbf{M}_1$

λ_3	λ_4	λ_5
λ_4	λ_5	λ_3
λ_5	λ_3	λ_4

$\{\lambda_3, \lambda_4, \lambda_5\}$

Initialization: Specify $\mathbf{A}$

- $\mathbf{A}_{ab} = \mathbf{M}_{[a+b]_n}$

$\mathbf{M}_0$

λ_0	λ_1	λ_2
λ_1	λ_2	λ_0
λ_2	λ_0	λ_1

$\mathbf{M}_1$

λ_3	λ_4	λ_5
λ_4	λ_5	λ_3
λ_5	λ_3	λ_4

$\mathbf{A}$

	Q_0 Q_1 Q_2	Q_3 Q_4 Q_5
P_0 P_1 P_2	$\mathbf{M}_0$	$\mathbf{M}_1$
P_3 P_4 P_5	$\mathbf{M}_1$	$\mathbf{M}_0$

$\mathbf{M}_1$

λ_3	λ_4	λ_5
λ_4	λ_5	λ_3
λ_5	λ_3	λ_4

$\mathbf{M}_0$

λ_0	λ_1	λ_2
λ_1	λ_2	λ_0
λ_2	λ_0	λ_1

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$ $\mathbf{A}_{01}$ $\mathbf{A}_{10}$ $\mathbf{A}_{11}$

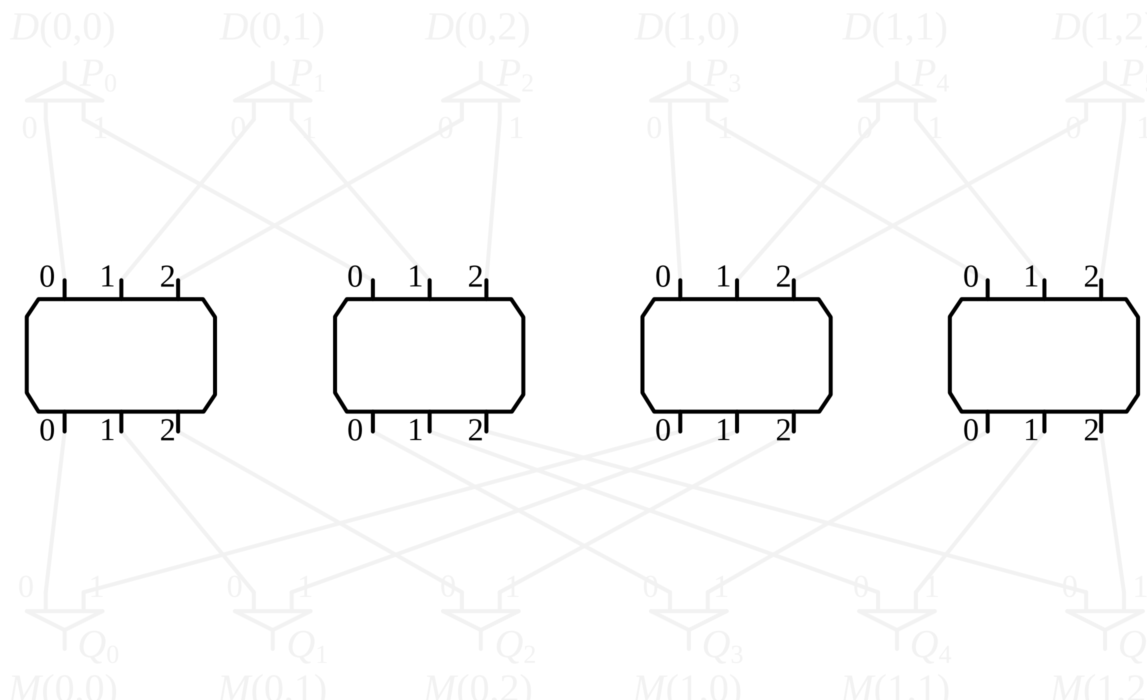
S1. Central Stage Construction

- Layout n^2 $r \times r$ AWGs from left to right
- Label kth AWG by $A(a, b)$ and associate it with $\mathbf{A}_{ab}$, where $a = \lfloor k/n \rfloor$, $b = [k]_n$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

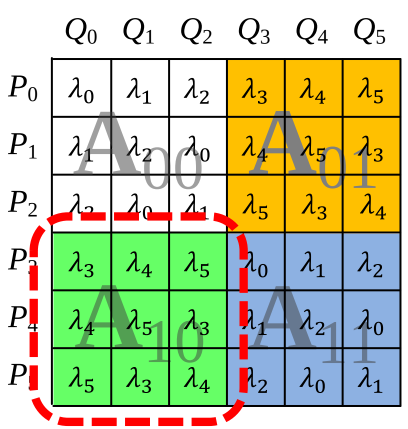
$\mathbf{A}_{00}$ $\mathbf{A}_{01}$ $\mathbf{A}_{10}$ $\mathbf{A}_{11}$

$r = 3, n = 2$

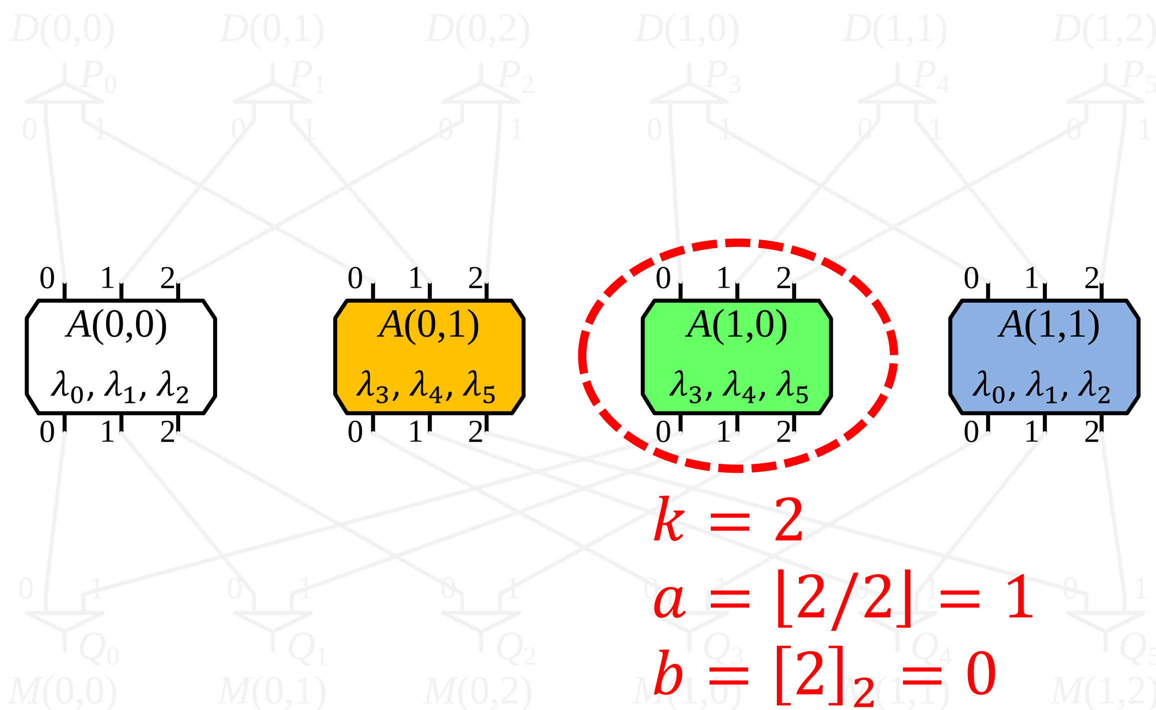


S1. Central Stage Construction

- Layout n^2 $r \times r$ AWGs from left to right
- Label kth AWG by $A(a, b)$ and associate it with $\mathbf{A}_{ab}$, where $a = \lfloor k/n \rfloor$, $b = [k]_n$



$r = 3, n = 2$



$k = 2$

$a = \lfloor 2/2 \rfloor = 1$

$b = [2]_2 = 0$

S2. Upper-layer Stage Construction

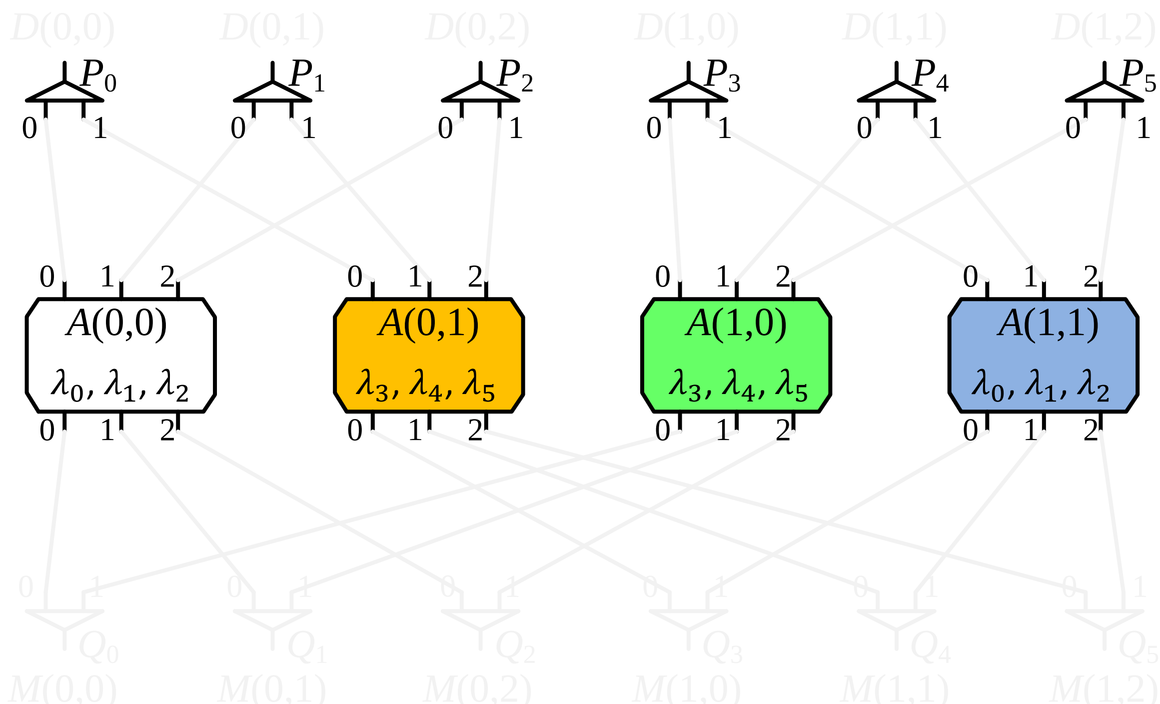
- Layout N DeMuxs at upper layer
- If ith row of $\mathbf{A}$ is αth row of $\mathbf{A}_{ab}$

output b of DeMux $i \leftrightarrow$ upper port α of $A(a,b)$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$, $\mathbf{A}_{01}$, $\mathbf{A}_{10}$, $\mathbf{A}_{11}$

$r = 3, n = 2$



S2. Upper-layer Stage Construction

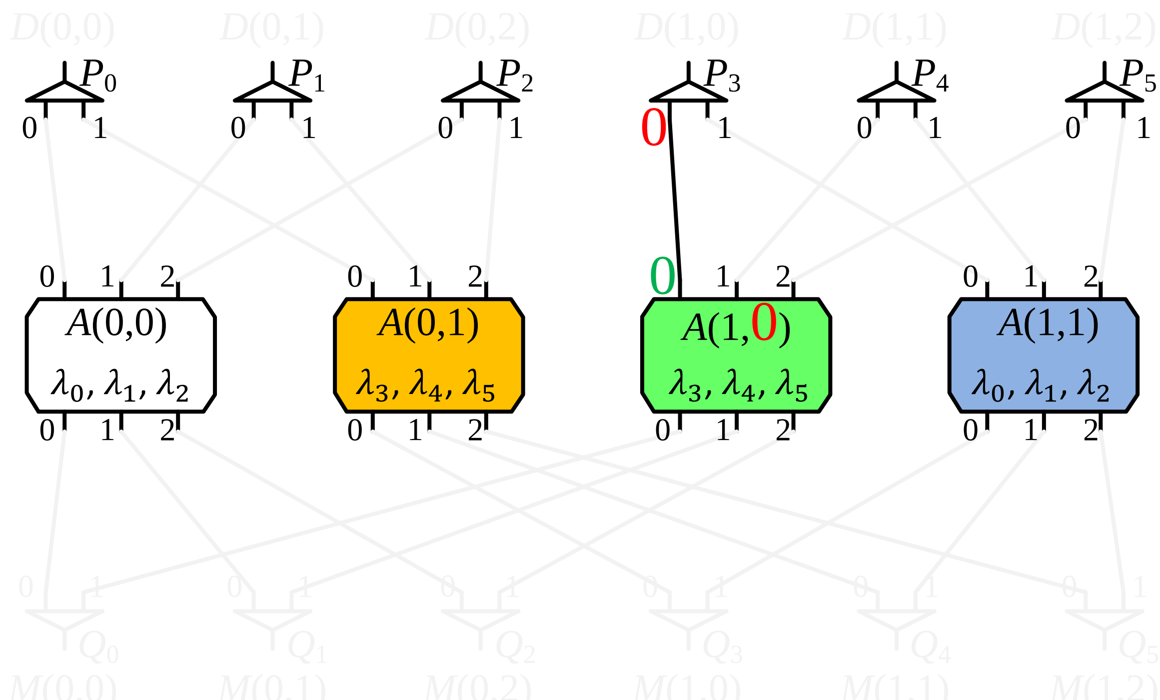
- Layout N DeMuxs at upper layer
- If ith row of $\mathbf{A}$ is αth row of $\mathbf{A}_{ab}$ ($b = 0 \sim n-1$)

$$\text{output } b \text{ of DeMux } i \leftrightarrow \text{upper port } \alpha \text{ of } A(a, b)$$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$, $\mathbf{A}_{01}$, $\mathbf{A}_{10}$, $\mathbf{A}_{11}$

$r = 3, n = 2$



S2. Upper-layer Stage Construction

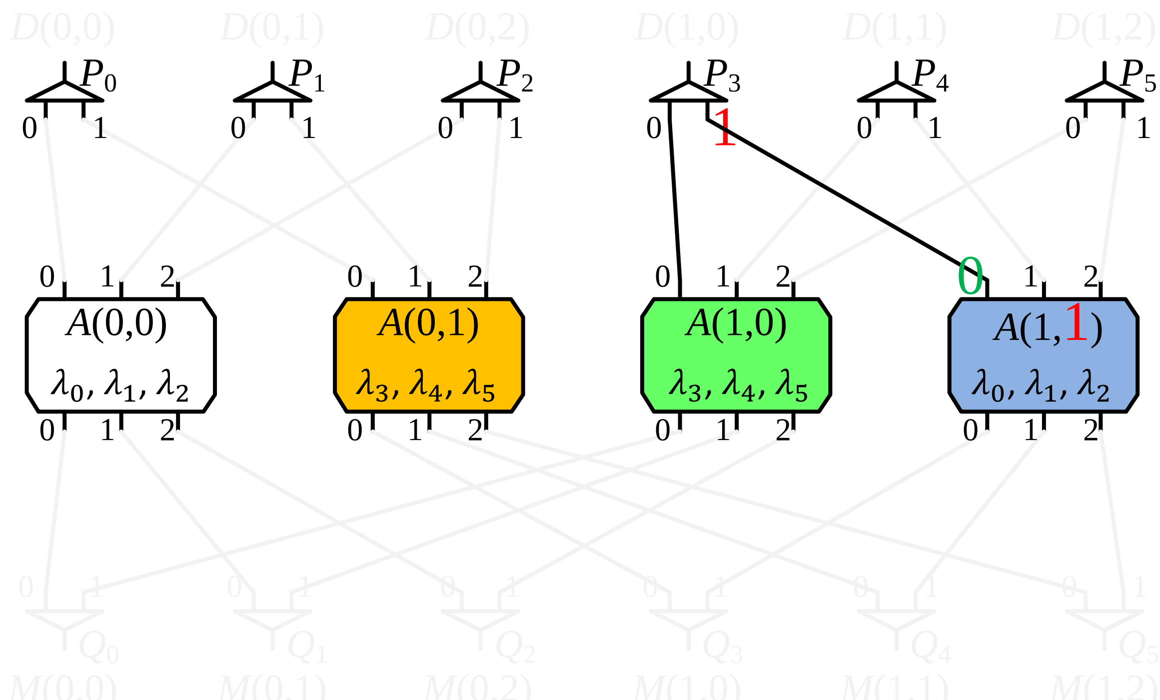
- Layout N DeMuxs at upper layer
- If ith row of $\mathbf{A}$ is αth row of $\mathbf{A}_{ab}$ ($b = 0 \sim n-1$)

$$\text{output } b \text{ of DeMux } i \leftrightarrow \text{upper port } \alpha \text{ of } A(a,b)$$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$, $\mathbf{A}_{01}$, $\mathbf{A}_{10}$, $\mathbf{A}_{11}$

$r = 3, n = 2$



S2. Upper-layer Stage Construction

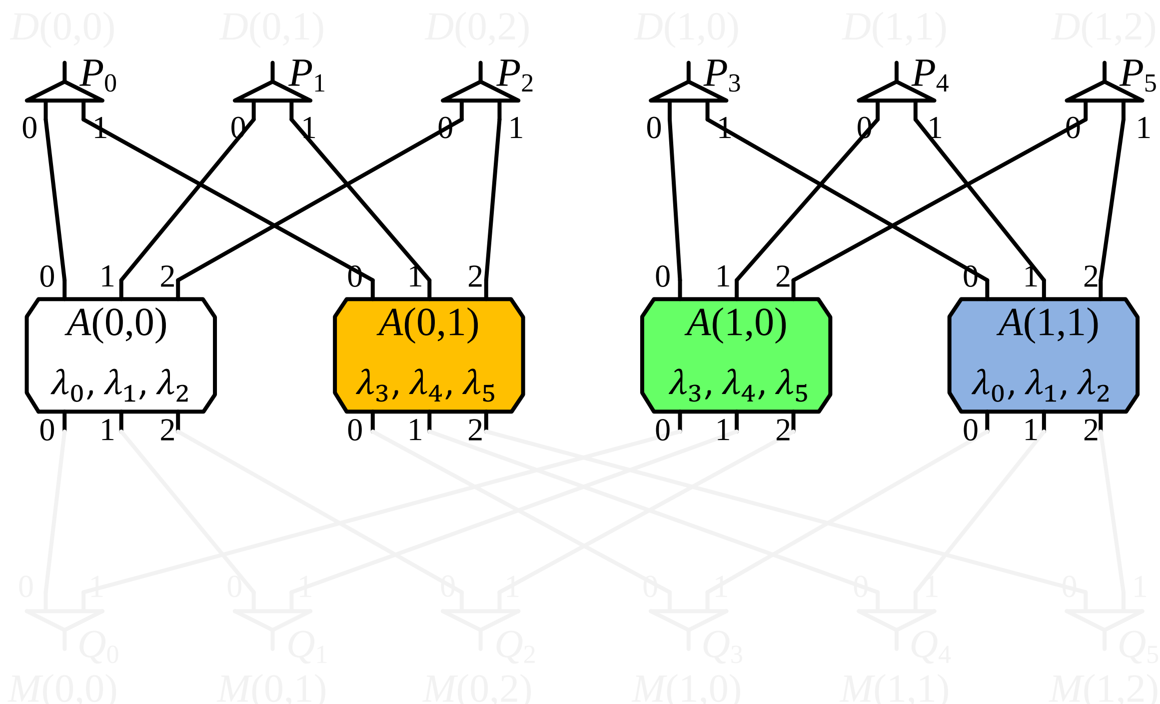
- Layout N DeMuxs at upper layer
- If ith row of $\mathbf{A}$ is αth row of $\mathbf{A}_{ab}$ ($b = 0 \sim n-1$)

$$\text{output } b \text{ of DeMux } i \leftrightarrow \text{upper port } \alpha \text{ of } A(a,b)$$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$ $\mathbf{A}_{01}$ $\mathbf{A}_{10}$ $\mathbf{A}_{11}$

$r = 3, n = 2$



S3. Lower Stage Construction

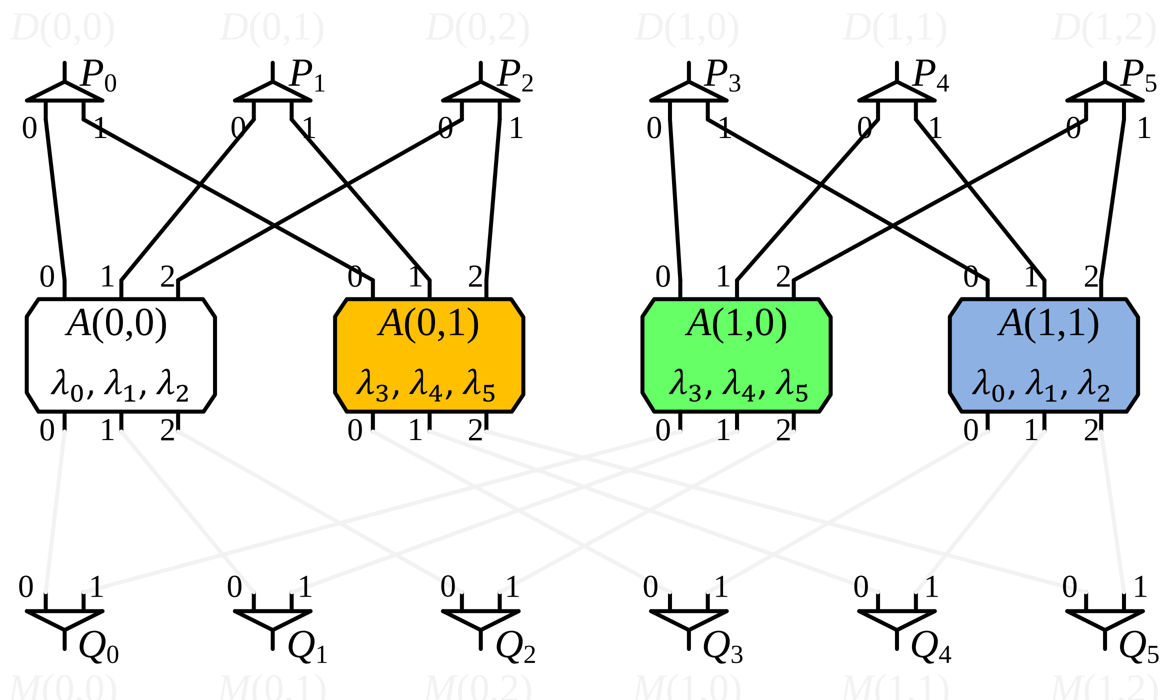
- Layout N Muxs at lower layer
- If jth col of $\mathbf{A}$ is βth col of $\mathbf{A}_{ab}$

$$\text{input } a \text{ of Mux } j \leftrightarrow \text{lower port } \beta \text{ of } A(a,b)$$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$ $\mathbf{A}_{01}$ $\mathbf{A}_{10}$ $\mathbf{A}_{11}$

$$r = 3, n = 2$$



S3. Lower Stage Construction

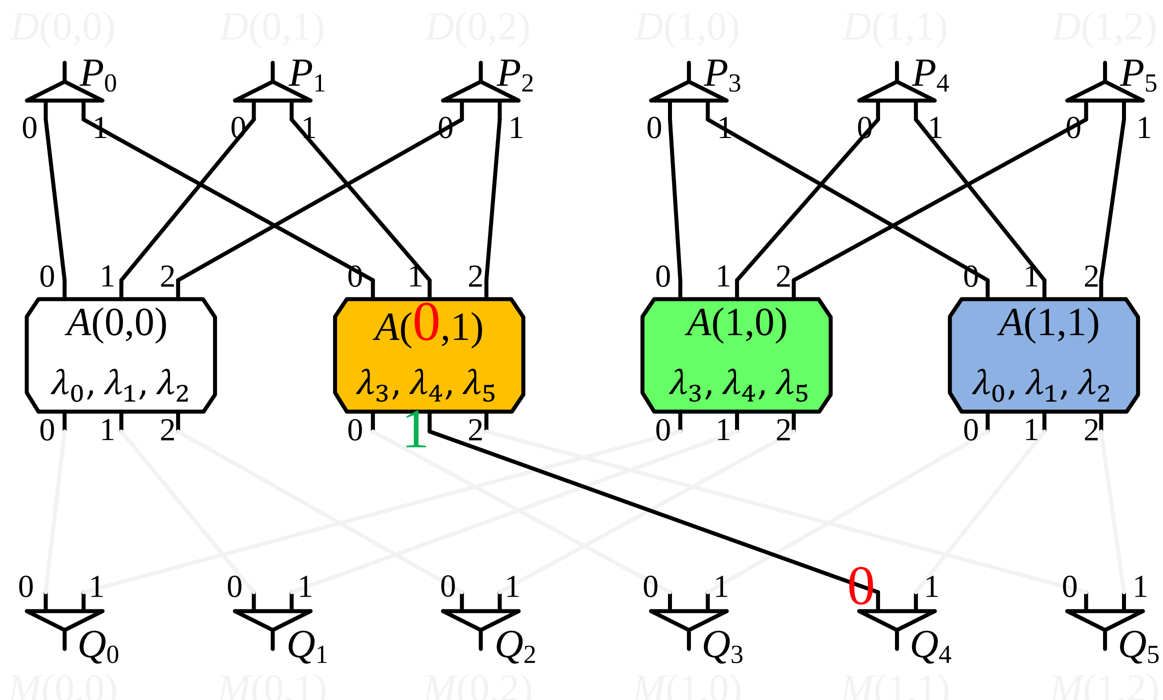
- Layout N Muxs at lower layer
- If jth col of $\mathbf{A}$ is βth col of $\mathbf{A}_{ab}$ $(a = 0 \sim n - 1)$

$$\text{input } a \text{ of Mux } j \leftrightarrow \text{lower port } \beta \text{ of } A(a, b)$$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$, $\mathbf{A}_{01}$, $\mathbf{A}_{10}$, $\mathbf{A}_{11}$

$r = 3, n = 2$



S3. Lower Stage Construction

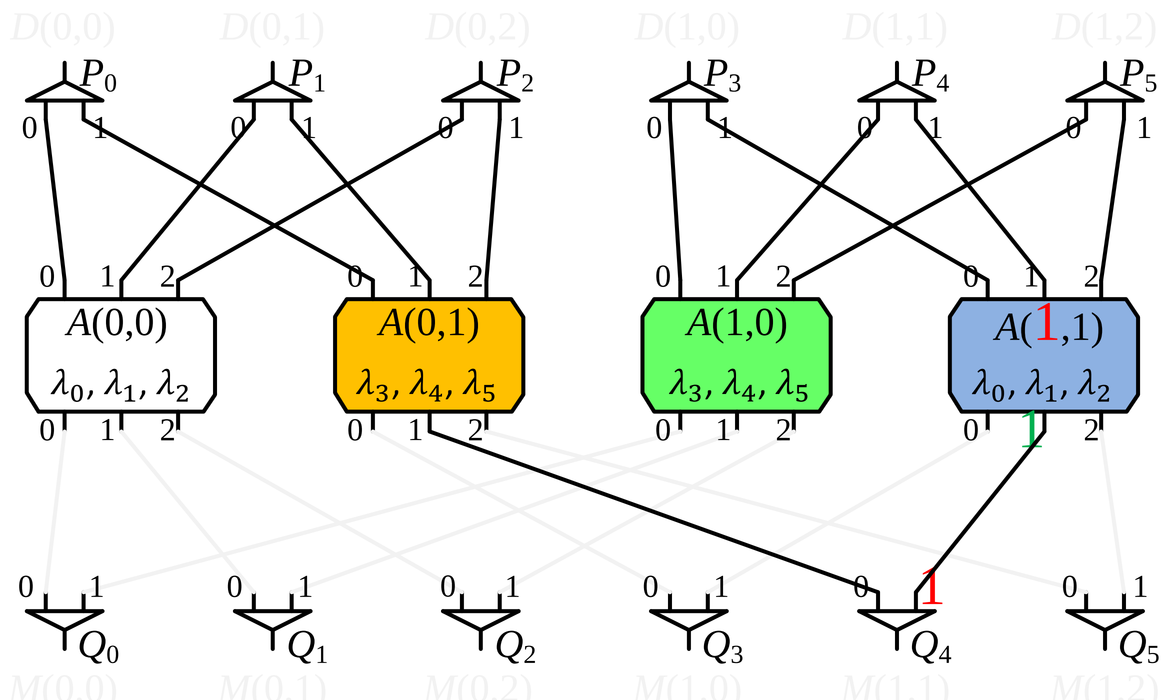
- Layout N Muxs at lower layer
- If jth col of $\mathbf{A}$ is βth col of $\mathbf{A}_{ab}$ $(a = 0 \sim n-1)$

$$\text{input } a \text{ of Mux } j \leftrightarrow \text{lower port } \beta \text{ of } A(a, b)$$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

$\mathbf{A}_{00}$, $\mathbf{A}_{01}$, $\mathbf{A}_{10}$, $\mathbf{A}_{11}$

$r = 3, n = 2$



S3. Lower Stage Construction

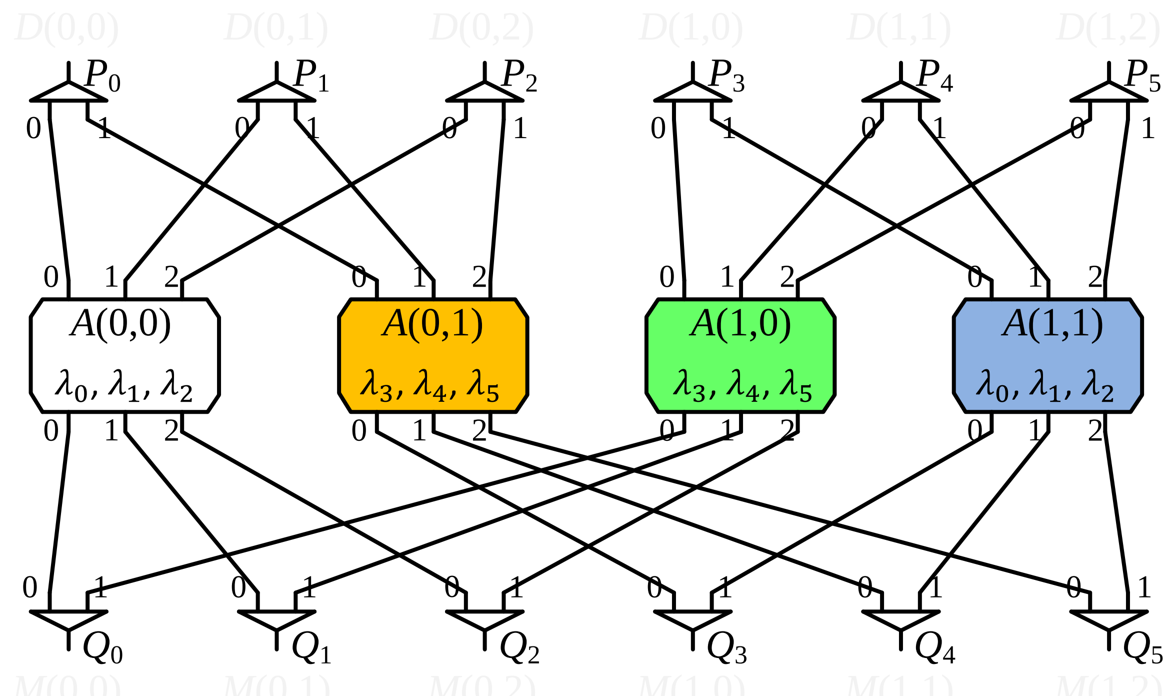
- Layout N Muxs at lower layer
- If jth col of $\mathbf{A}$ is βth col of $\mathbf{A}_{ab}$ ($a = 0 \sim n-1$)

$$\text{input } a \text{ of Mux } j \leftrightarrow \text{lower port } \beta \text{ of } A(a,b)$$

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

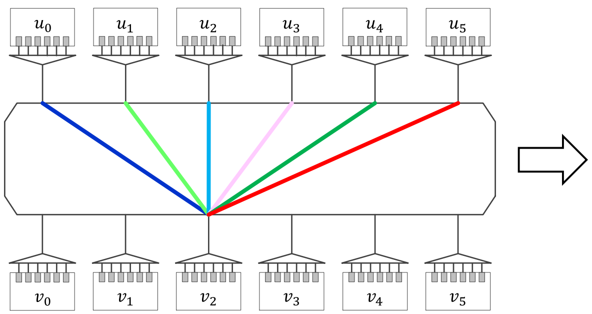
$\mathbf{A}_{00}$ $\mathbf{A}_{01}$ $\mathbf{A}_{10}$ $\mathbf{A}_{11}$

$r = 3, n = 2$

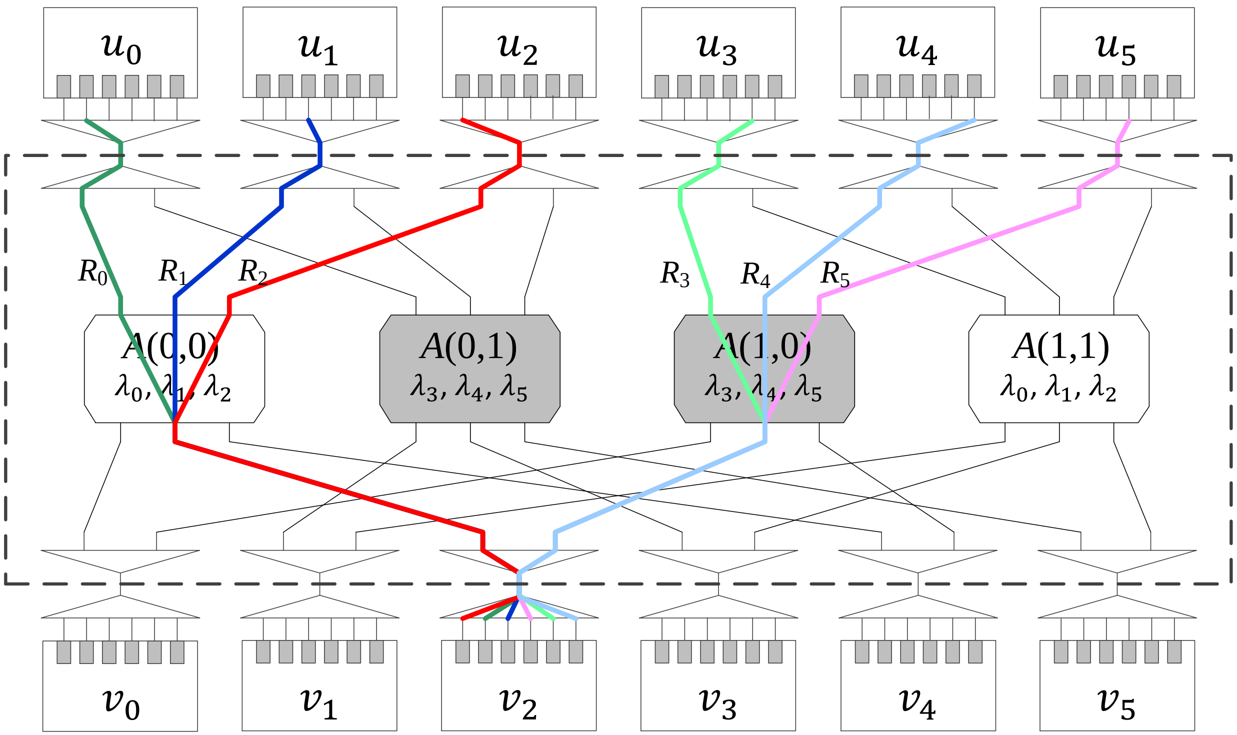


Network After AWG Decomposition

- $N \times N\ \mathcal{N}_B \Rightarrow N \times N\ \mathcal{N}_C(n, r)$, where $nr = N$



$\mathcal{N}_B$

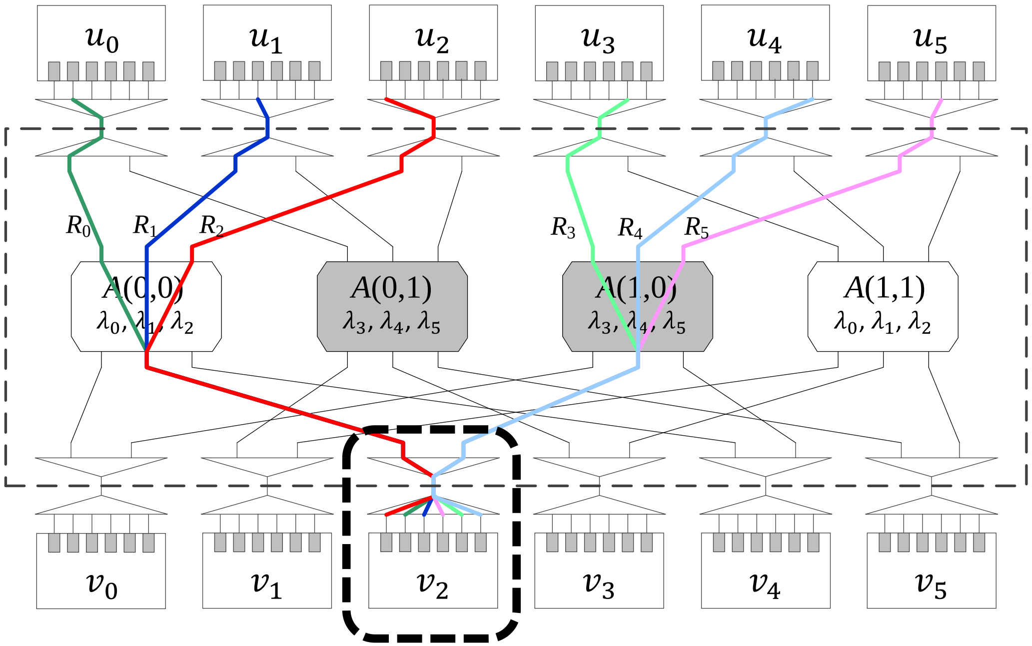


$\mathcal{N}_C(3,2)$

Outline

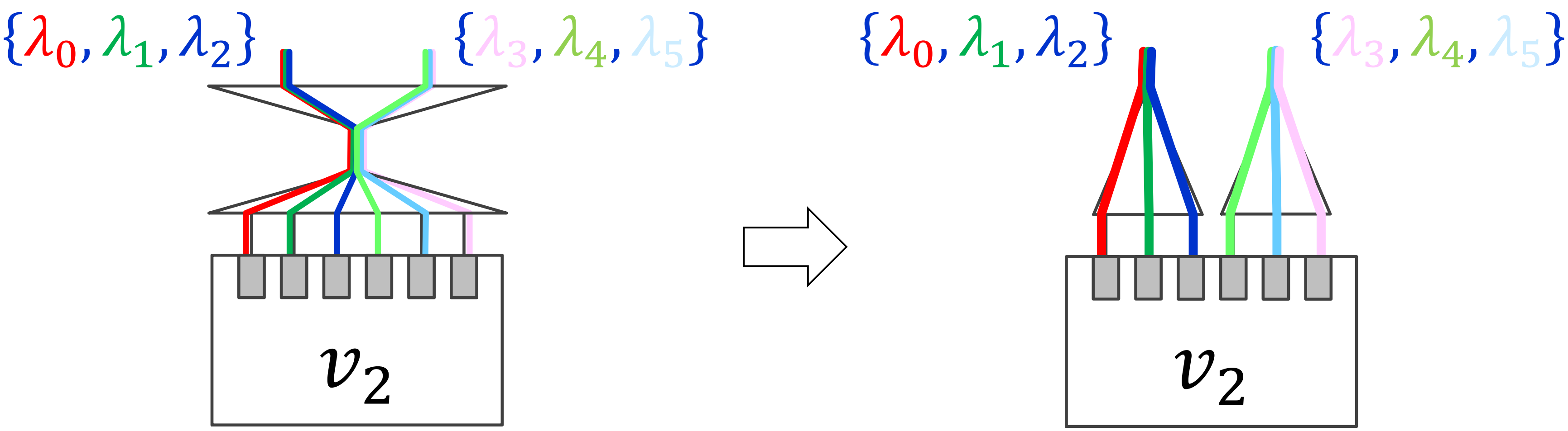
- Background
- Preliminaries
- Modular AWG-based Interconnection
  - Phase 1: AWG Decomposition
  - Phase 2: Wavelength Reuse
- Application to Data Center Networks
- Conclusion

Mux/DeMux Replacement



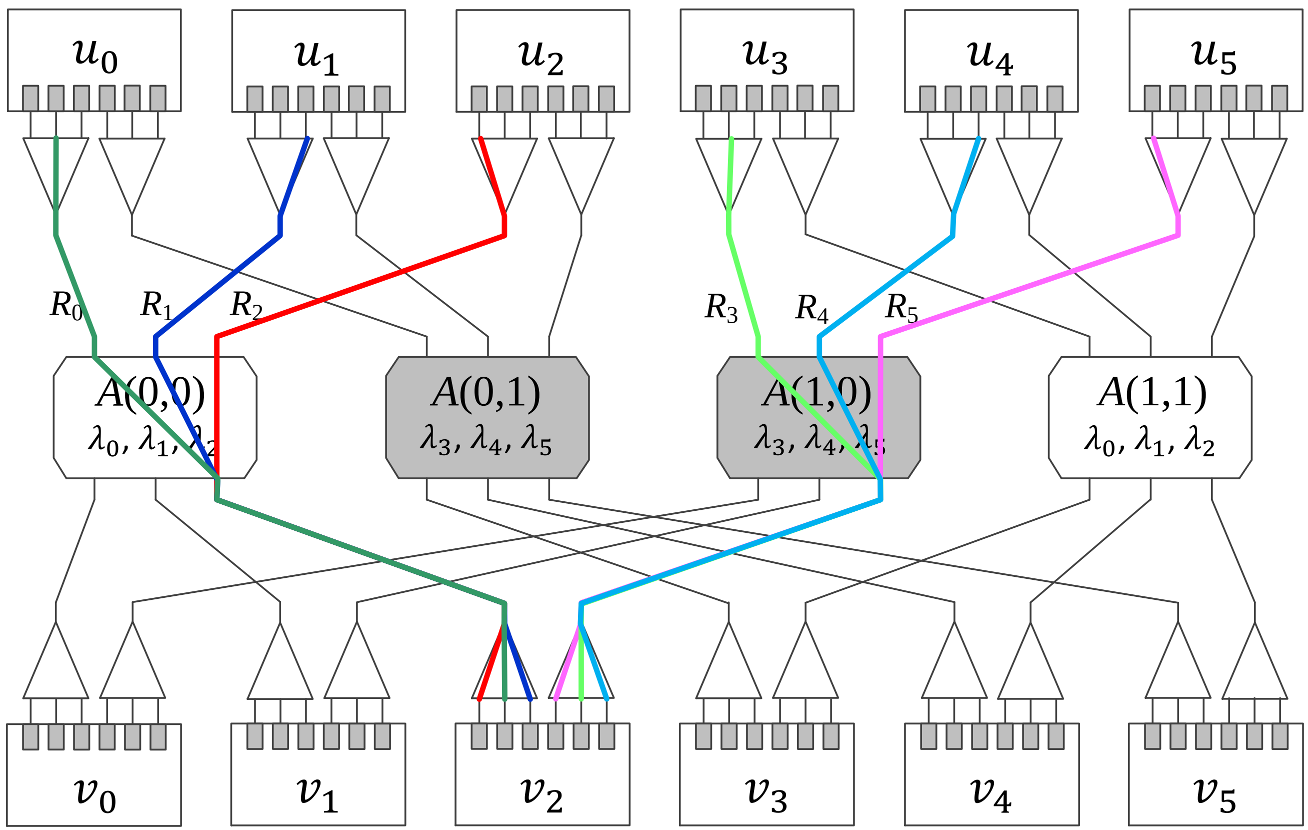
	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

A_{00} A_{01} A_{10} A_{11}



Network After Mux Replacement

- Connections passing through different AWGs are link-disjoint => reuse the same λ-set

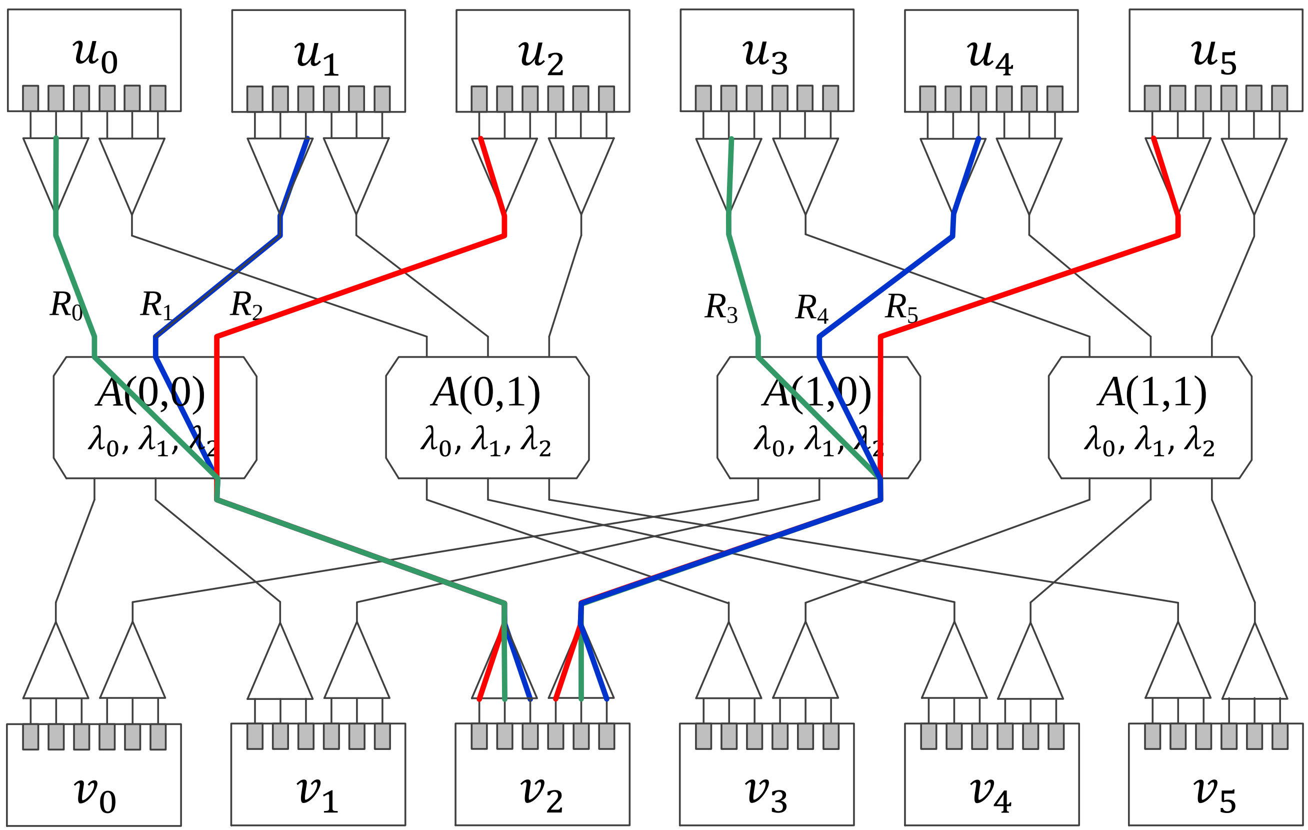


	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
P_1	λ_1	λ_2	λ_0	λ_4	λ_5	λ_3
P_2	λ_2	λ_0	λ_1	λ_5	λ_3	λ_4
P_3	λ_3	λ_4	λ_5	λ_0	λ_1	λ_2
P_4	λ_4	λ_5	λ_3	λ_1	λ_2	λ_0
P_5	λ_5	λ_3	λ_4	λ_2	λ_0	λ_1

A_{00} A_{01} A_{10} A_{11}

Network After Wavelength Reuse

- $\mathcal{N}_D(n, r)$ where $n = 2$ and $r = 3$



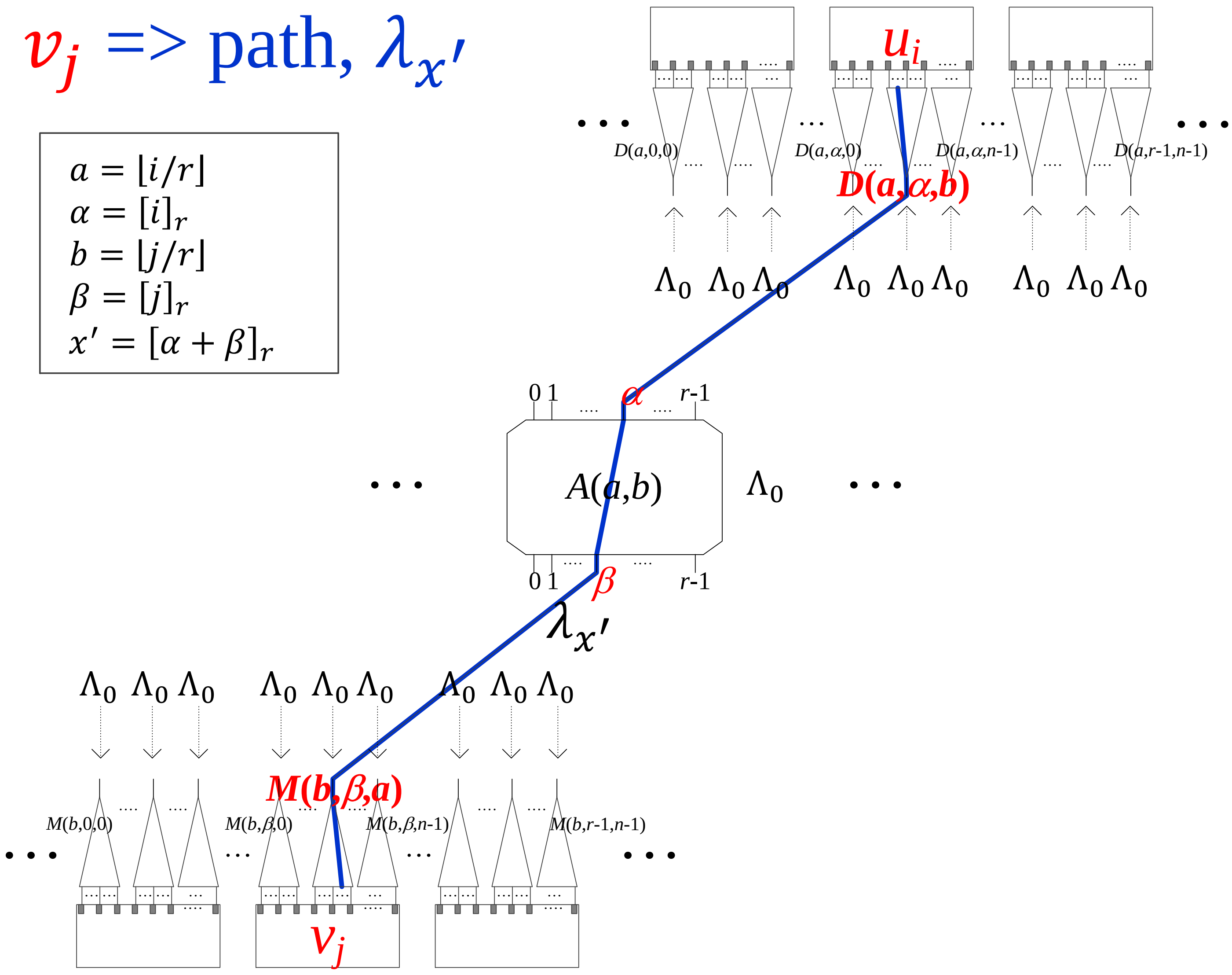
	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
P_0	λ_0	λ_1	λ_2	λ_0	λ_1	λ_2
P_1	λ_1	λ_2	λ_0	λ_1	λ_2	λ_0
P_2	λ_2	λ_0	λ_1	λ_2	λ_0	λ_1
P_3	λ_0	λ_1	λ_2	λ_0	λ_1	λ_2
P_4	λ_1	λ_2	λ_0	λ_1	λ_2	λ_0
P_5	λ_2	λ_0	λ_1	λ_2	λ_0	λ_1

A_{00} A_{01} A_{10} A_{11}

$\mathcal{N}_D(n,r)$ in General

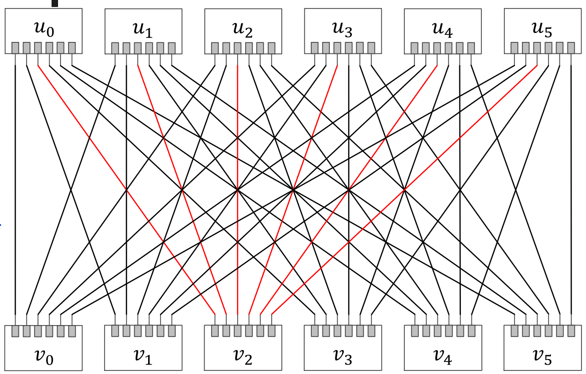
- u_i, v_j => path, $\lambda_{x'}$

$a = \lfloor i/r \rfloor$
$\alpha = [i]_r$
$b = \lfloor j/r \rfloor$
$\beta = [j]_r$
$x' = [\alpha + \beta]_r$



Comparison

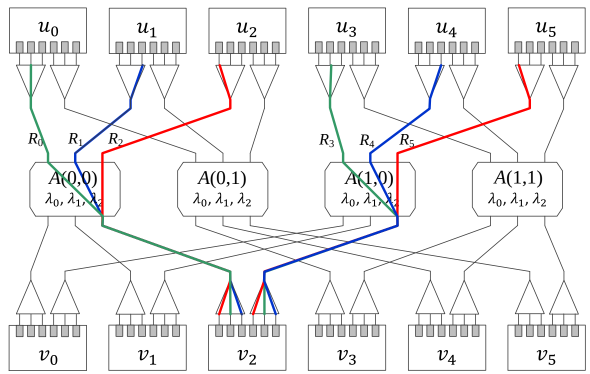
$\mathcal{N}_A$



Cabling complexity: $O(N^2)$

Number of required wavelengths: $O(1)$

$\mathcal{N}_D$

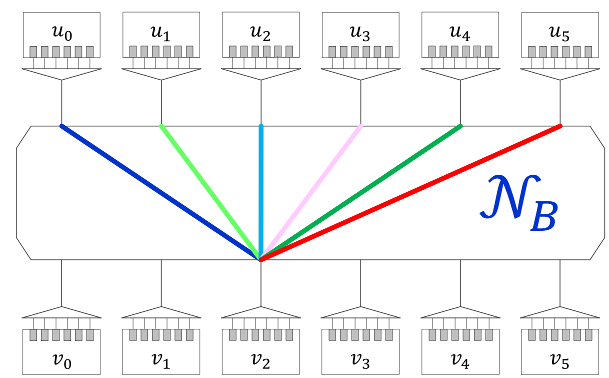


$O\left(\frac{N^2}{r}\right)$

$O(r)$

$O(N)$

$O(N)$

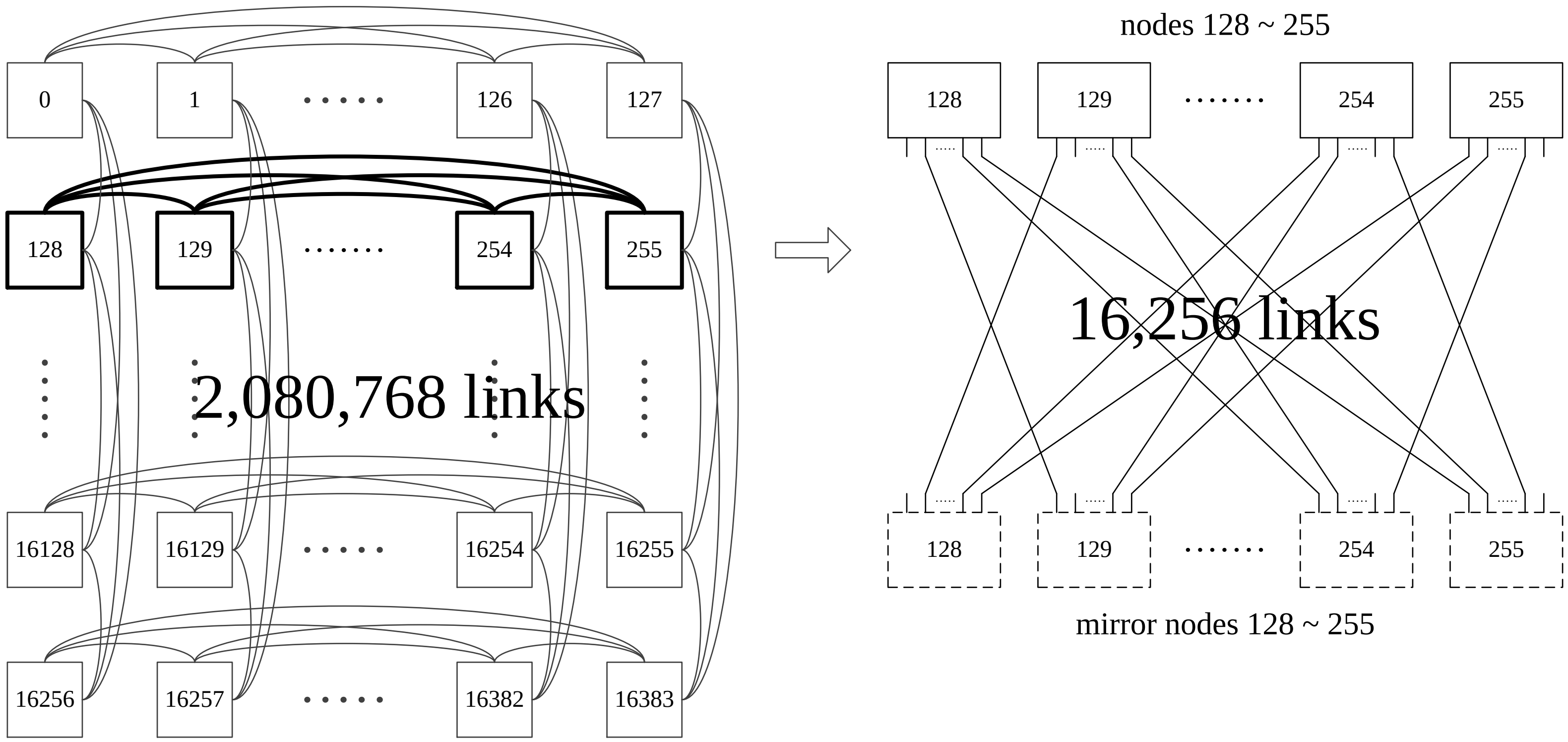


Outline

- Background
- Preliminaries
- Modular AWG-based Interconnection
  - AWG Decomposition
  - Wavelength Reuse
- Application to Data Center Networks
- Conclusion

2-D FB Network

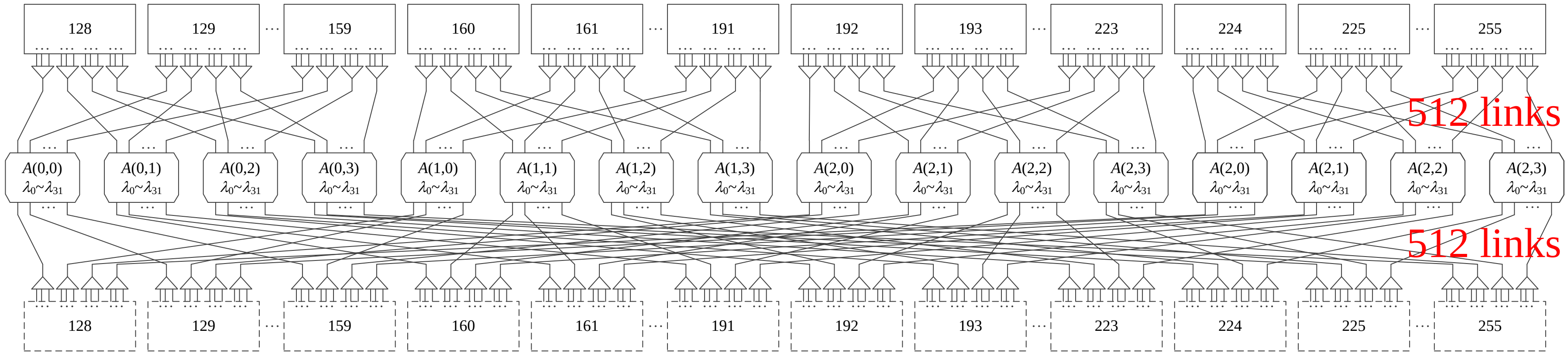
- A 2-D 16,384-node DC network

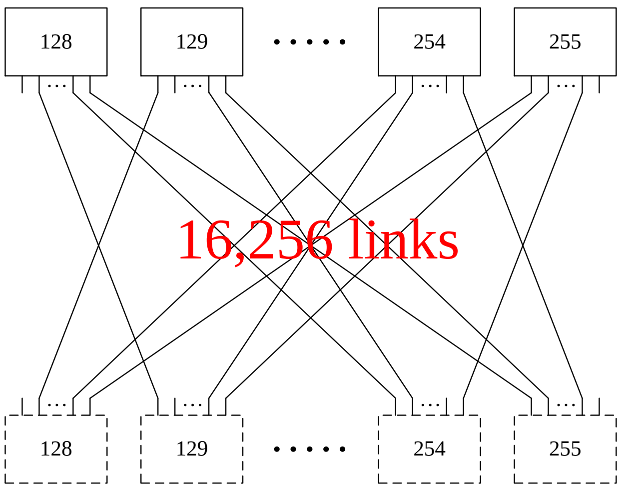


Z. Zhu, S. Zhong, L. Chen, and K. Chen, "Fully programmable and scalable optical switching fabric for petabyte data center", Opt. Express, vol. 32, no. 3, pp. 3563-3580, Feb. 2015.

AWG-based Interconnection Scheme

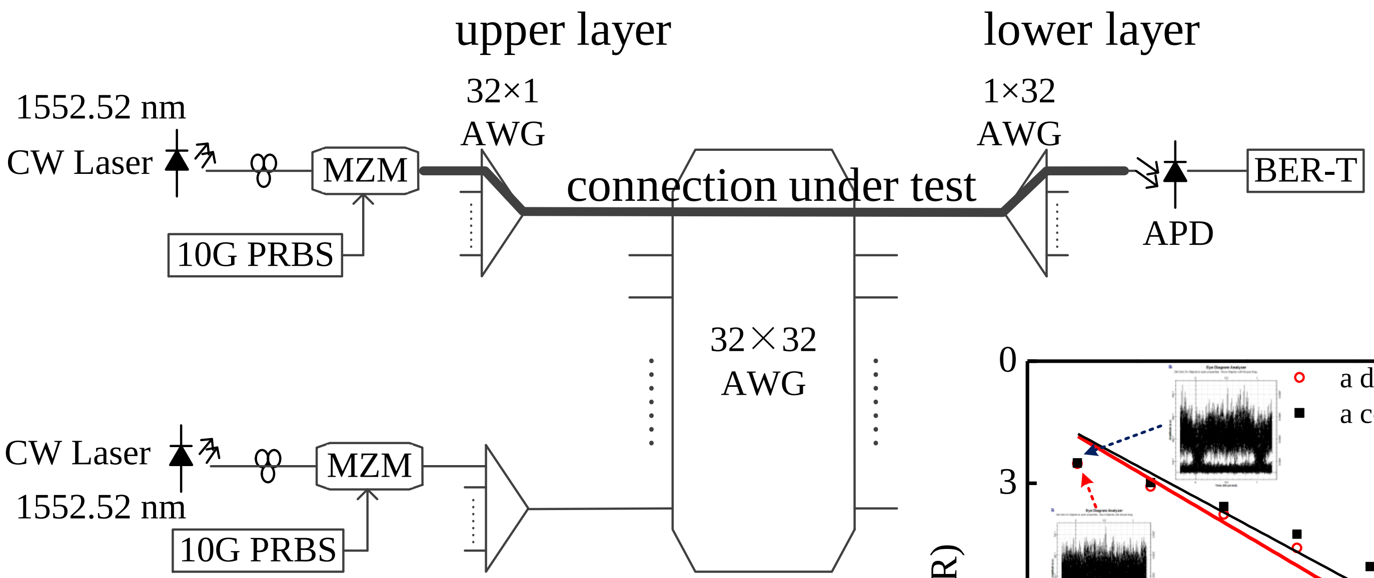
- $\mathcal{N}_D(4,32)$: 32×32 AWGs in the central stage

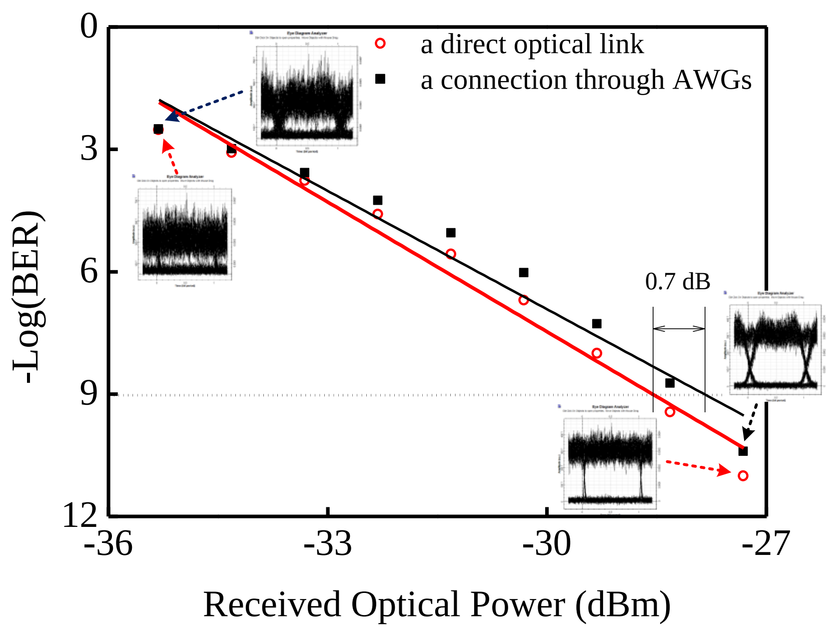




Physical Layer Performance

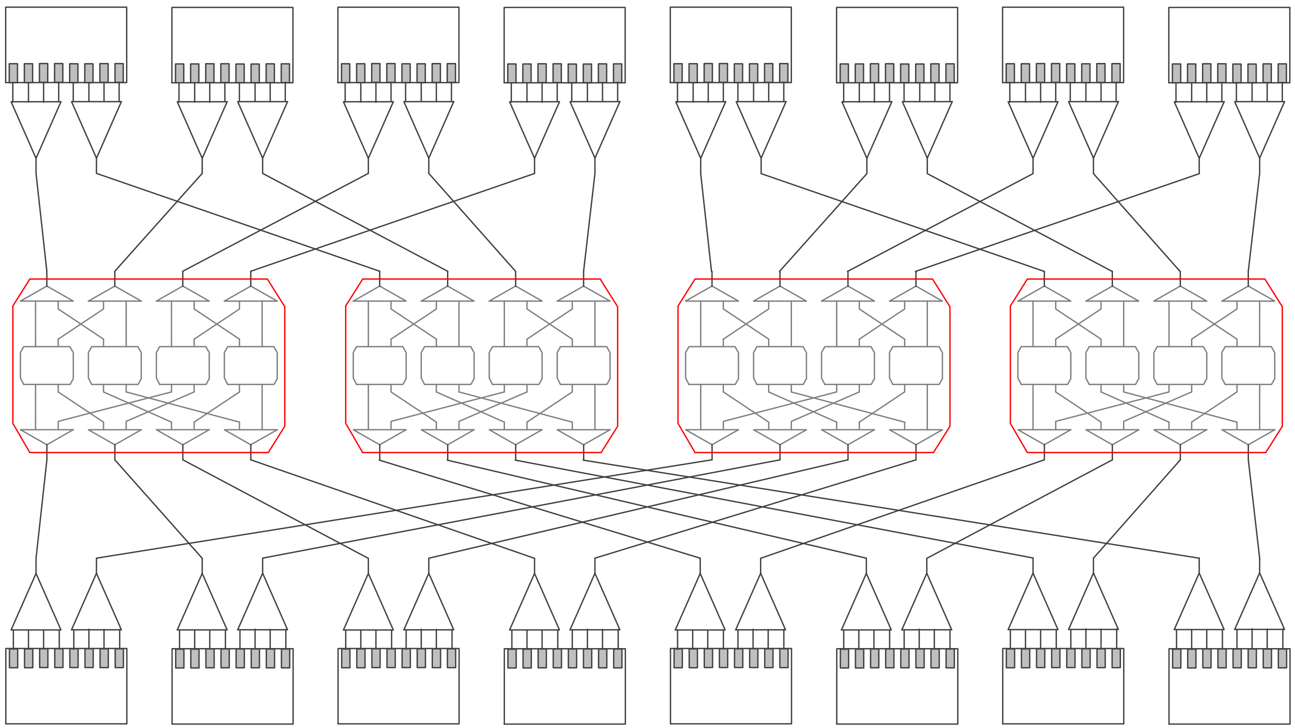
- Power penalty (~0.7dB) is very small





If Network is very Large …

- Each central AWG is replaced by an integrated AWG-network module



Conclusions

- AWG-based interconnection networks is proposed for DC networks
  - Substantially reduce cabling complexity
  - Only employ small-size AWG modules to avoid
    - serious in-band crosstalk
    - difficult synthesis technology
  - Reuse same wavelength set, such that
    - number of required wavelengths is small
- Feasibility is confirmed by Physical-layer performance evaluations

Tong Ye, Tony T. Lee, Mao Ge, and Weisheng Hu, Modular AWG-based Interconnection for Large-Scale Data Center Networks, IEEE Trans. on Cloud Computing, Accepted.

Q & A

Thank you for your attention!