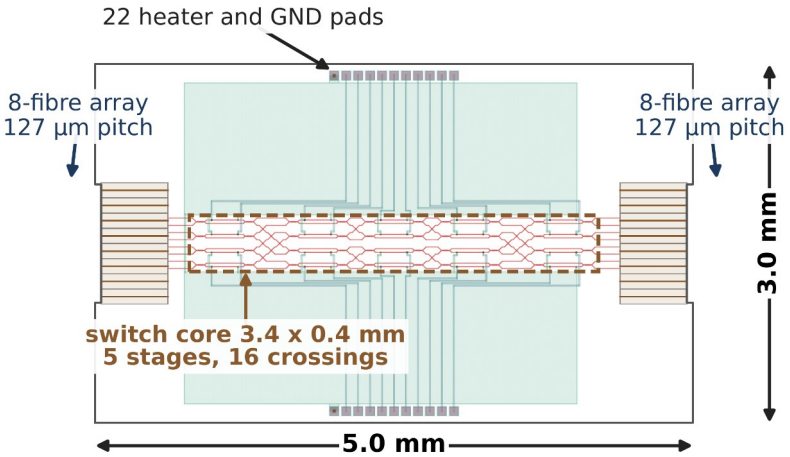
# Design and Evaluation of an 8×8 Silicon Photonic Beneš Switch for GPU Scale-Up Interconnects

## Critical Design Review

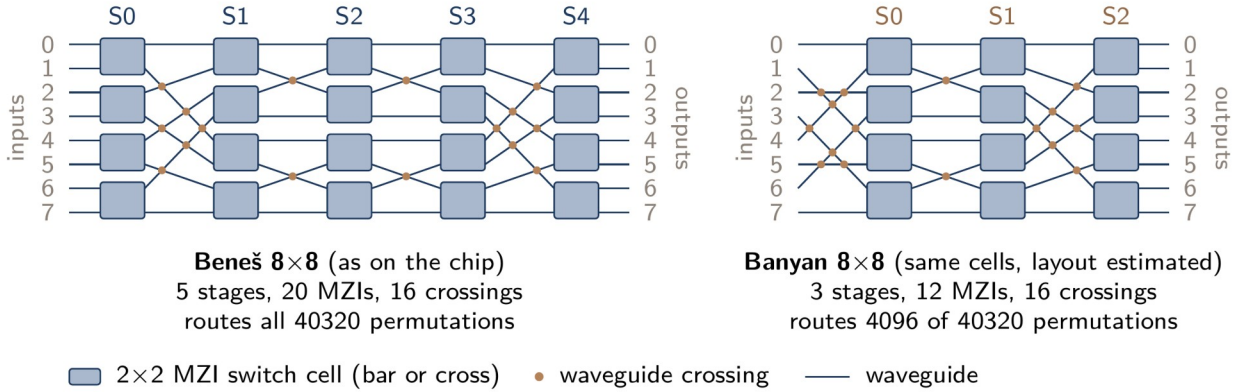
An 8×8 thermo-optic MZI Beneš switch for eight GPUs, compared with an 8×8 Banyan built from the same cells.




ECE 8803 Photonics for AI, Team 4

# Project recap and what changed since the PDR

Since the PDR we simulated the whole chip in INTERCONNECT, designed the heater in HEAT, stressed the design and brought the layout to near final.

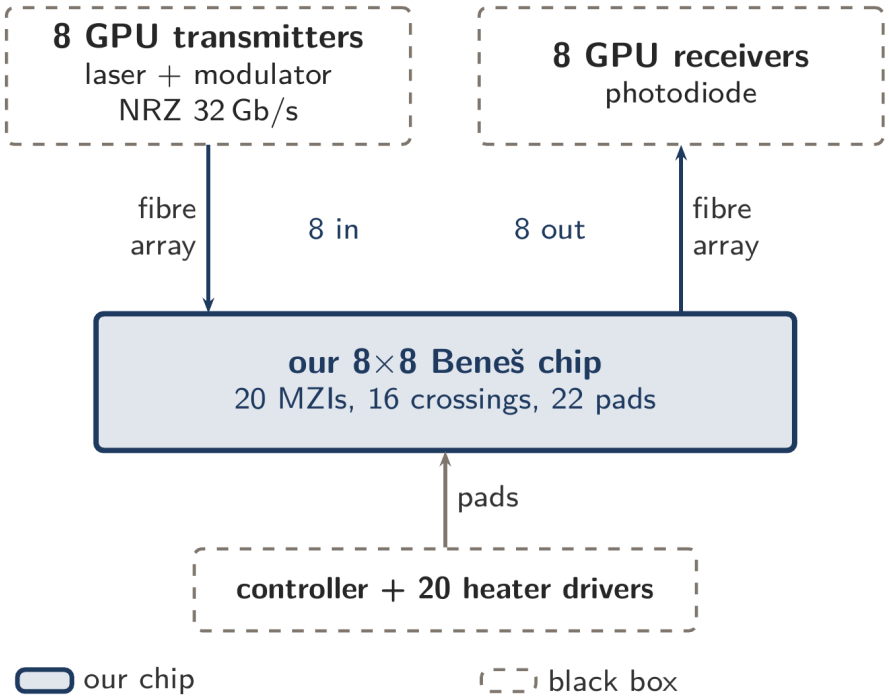



- Beneš 8×8, 5 stages, 20 MZIs, 16 crossings, routes all 40 320 permutations
- Banyan 8×8 from the same cells, 3 stages, 12 MZIs, routes 4096 of them
- C-band, one channel per port at 1550 nm, all inputs active

| At the PDR | At the CDR |
|---|---|
| components only | full chip in INTERCONNECT, wired from the layout |
| heater open | heater in Lumerical HEAT, $P_{\pi}$ 18.9 mW |
| PDR feedback: higher-order modes | mode decomposition, crossing tapers 3 → 6 µm (part C) |
| no stressed results | phase error, process corners, temperature |
| layout picture | near-final layout, checked against the schematic |

# System architecture, black boxes and final targets

We design only the switch chip. Lasers, receivers, fibre arrays and heater drivers are black boxes, and the table holds the chip to its final targets.

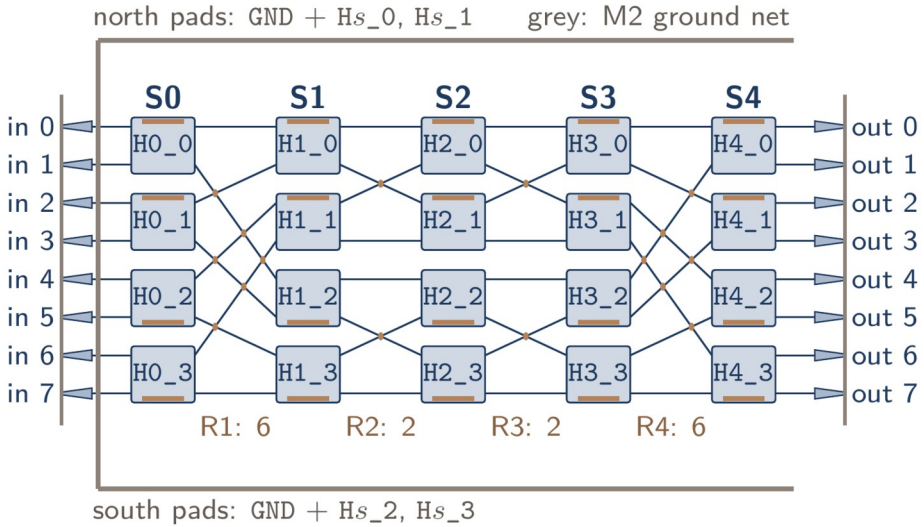



| Metric | Final target |
|---|---|
| ports, channel | 8×8, 1550 nm |
| data rate | 32 Gb/s per port, NRZ |
| crosstalk penalty, all inputs active | 1 dB or less |
| worst-path on-chip loss | 5 dB or less |
| coupling loss per facet | 2 dB or less |
| BER | $10^{-12}$ without FEC |
| heater power | $P_{\pi}$ 30 mW or less |
| switching time | 30 µs or less |

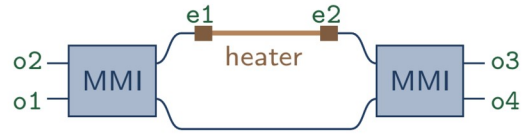
# Final hierarchical schematic

Five stages of four MZI cells, four wiring regions, 16 optical ports and 22 pads. The netlist extracted from the GDS matches this schematic.

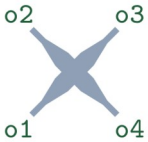
**(a) Switch level: 5 MZI stages, 16 crossings**




**(b) Cell level, with ports**




MMI 6 × 42.70 µm, heater 200 µm arms 300 µm; e1 to its pad, e2 to GND




**MMI crossing**
2.0 × 8.90 µm, 6 µm tapers
20.9 µm port to port



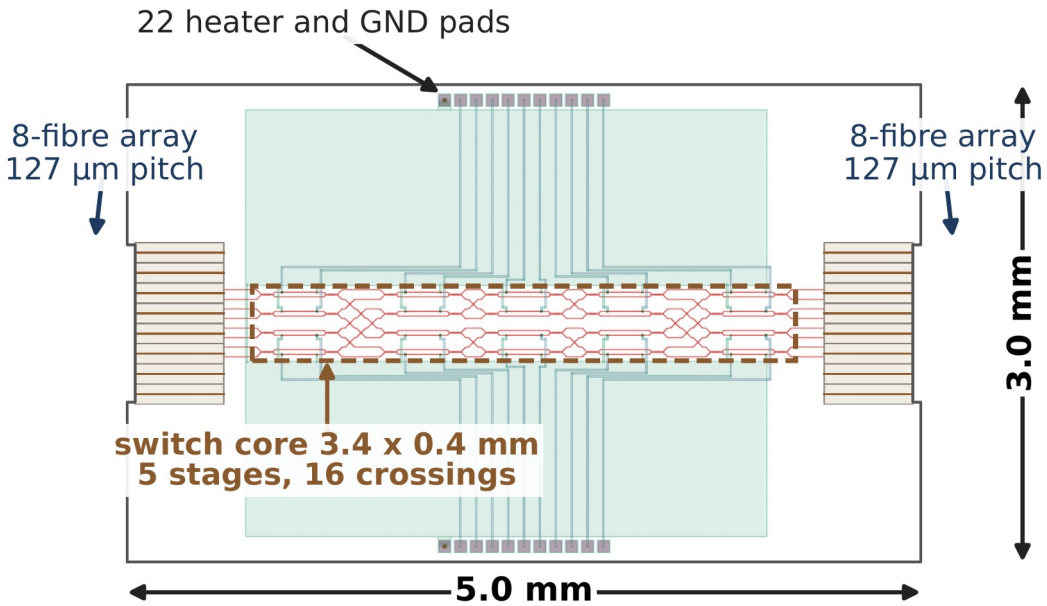

**edge coupler** 100 µm, 0.2 µm tip

**pad** 80 µm, 100 µm pitch

**Since the PDR:** crossing tapers 3 → 6 µm after the mode check, loss 0.33 → 0.16 dB. All our cells registered in the course PDK; layout checked against this schematic.

# Near-final layout

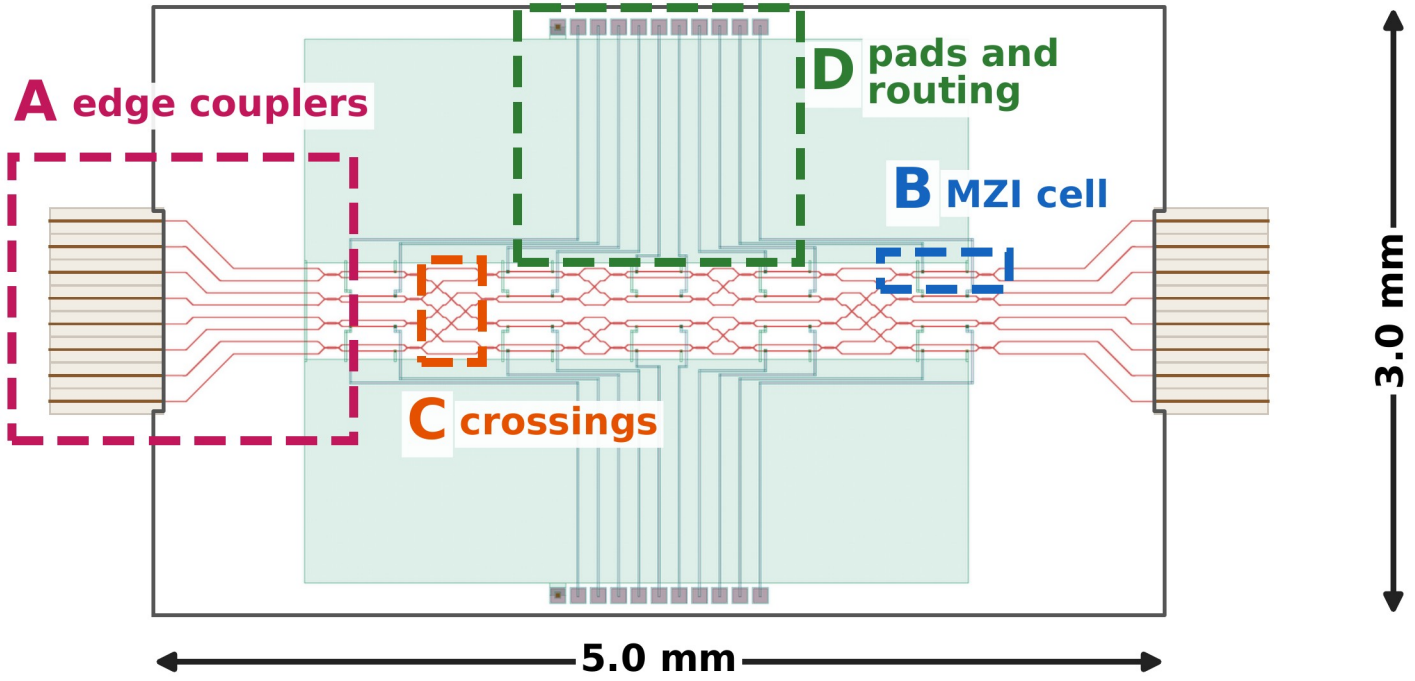
The 8×8 Beneš chip is drawn on the course PDK in 5.0×3.0 mm, with the fibre arrays at the facets and the pads in two rows.




1. Die 5.0×3.0 mm. Switch core 3.37×0.42 mm with 20 MZIs and 16 crossings (6, 2, 2, 6 per region)
2. 8 edge couplers per facet at 127 µm, one 8-fibre array each
3. 22 pads in two rows: 20 heaters and 2 ground
4. Optical paths 5.13 to 5.45 mm with 0 to 8 crossings
5. GDS checked against the schematic, 0 DRC markers with our deck

# What we look at in the layout

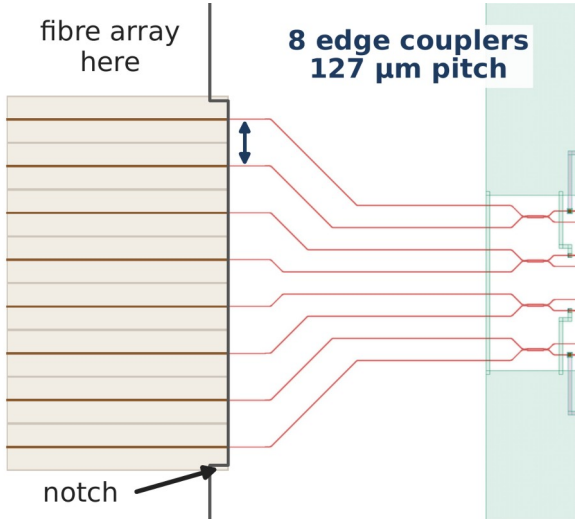
We unpack the chip in four parts and show how we designed each: A edge couplers, B MZI cell, C crossings, D pads.

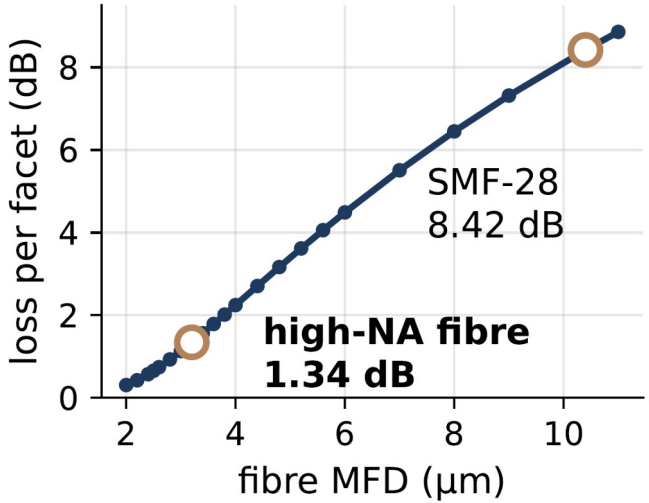


# A. Edge couplers and fibre arrays

A

The PDK edge coupler loses 1.34 dB per facet with high-NA fibre against 8.42 dB with SMF-28, so each facet takes a high-NA fibre array.



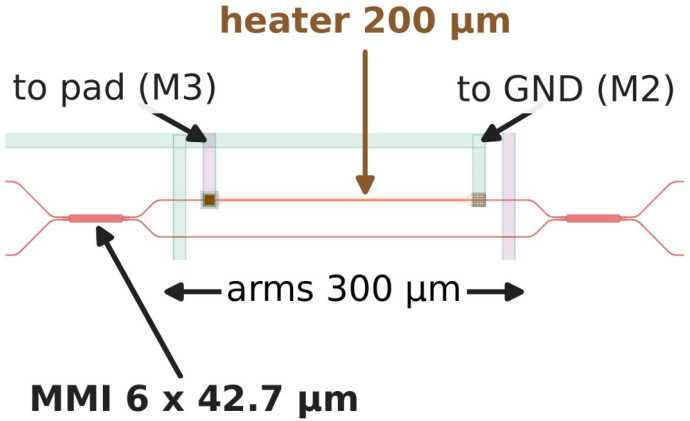



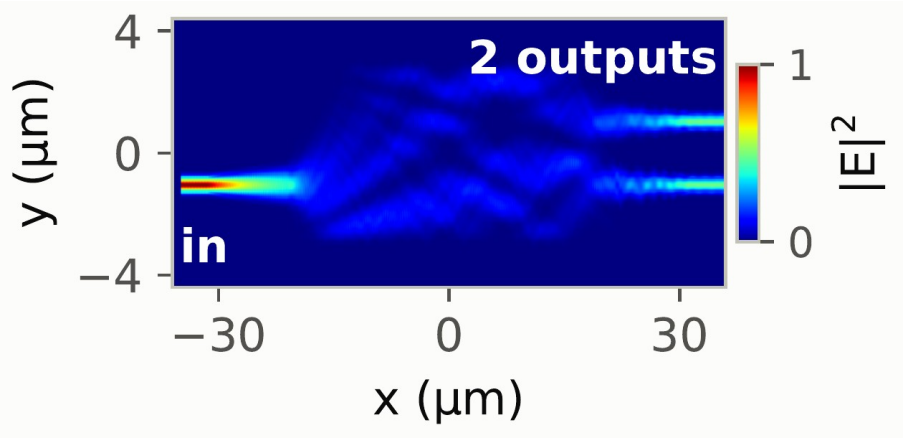

1. 100 µm taper to a 0.2 µm tip, 3D FDTD: 1.34 dB at 1550 nm, 1.19 to 1.58 dB over the C-band
2. 8 couplers per facet at 127 µm face one 8-fibre array. A 0.46 µm offset adds 0.5 dB
3. The claddings span 1014 µm against the 989 µm PDK notch, so we widen the notch by 75 µm

# B. MZI switch cell

B

Each MZI cell is two 6×42.70 µm MMIs and a 200 µm heater. It loses 0.34 dB and leaks at most −29.9 dB.



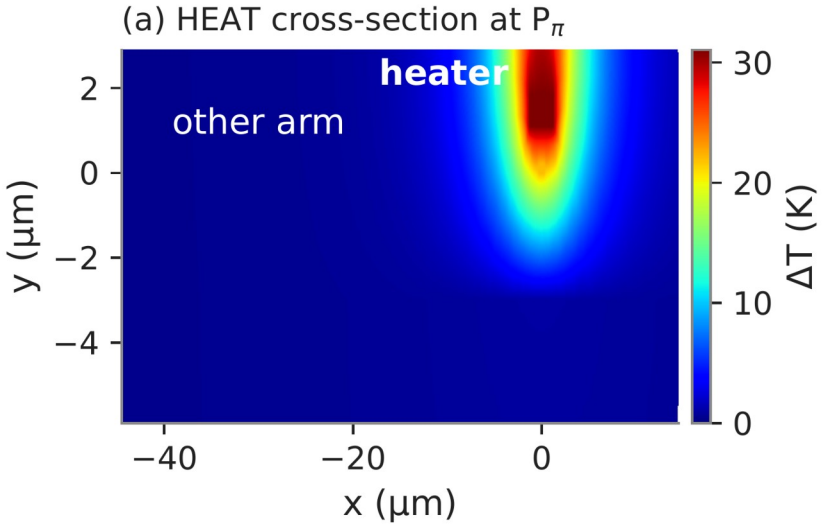



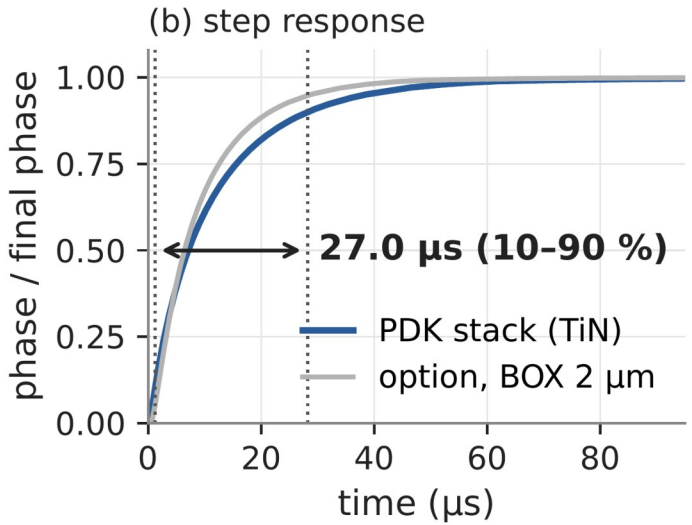

1. MMI length swept from 42.5 to 43.0 µm in 3D FDTD: 42.70 µm gives 0.14 dB excess loss. The 2D estimate of 47.1 µm leaked −16.9 dB
2. Arms 300 µm long and 30 µm apart. One MZI phase setting per state covers the C-band
3. Cell in INTERCONNECT from the FDTD S-parameters: 0.34 dB at 1550 nm, leak at most −29.9 dB over the C-band

# B. MZI cell: the heater

B

The PDK heater switches an MZI with 18.9 mW, and its 27 µs 10–90 % response meets the 30 µs target.



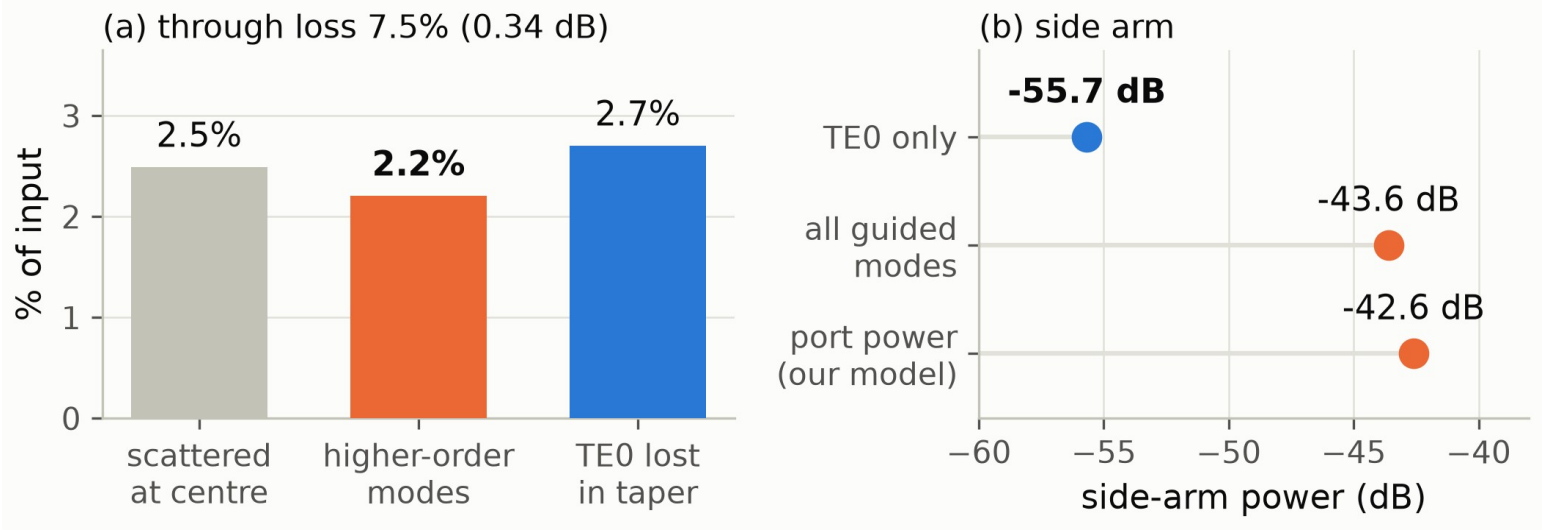




- $P_\pi$ 18.9 mW, 4.87 mA and 3.89 V per heater (800 Ω from an assumed 10 Ω/sq). 13 to 16 heaters are on per step, 246 to 303 mW
- 10–90 % in 27.0 µs, but settling to 1° takes 78 µs. Pre-emphasis brings the rising edge to 22 µs
- 30 µm arm spacing and 120 µm MZI pitch: the neighbouring MZI picks up only 0.47° at its π drive

# C. Crossings, PDR feedback: where the crossing loses light

C

The mode decomposition shows that two thirds of the crossing loss leaves in its short output taper, and that our crosstalk already counted every mode.

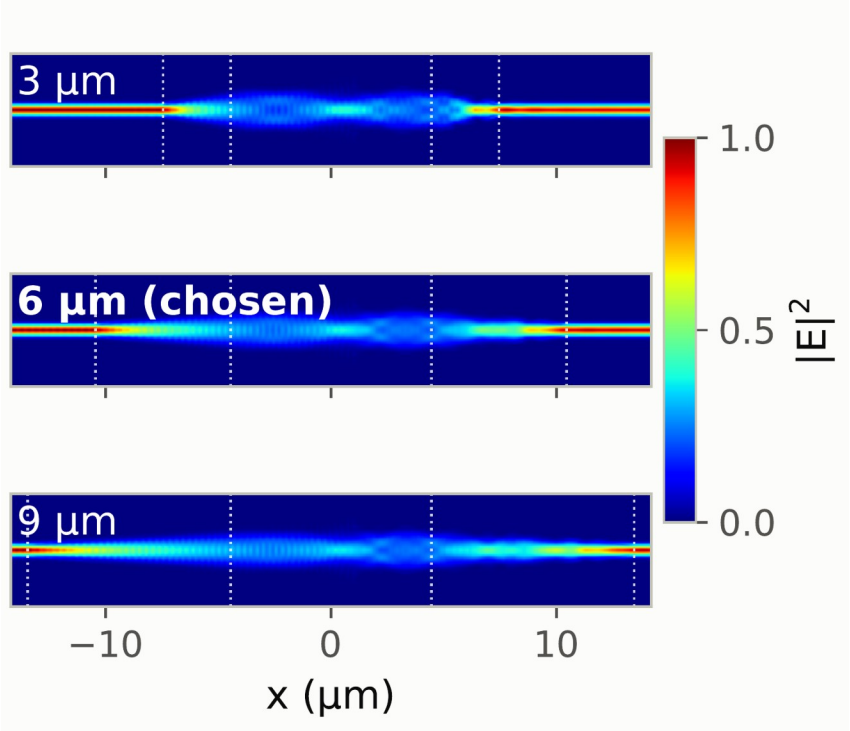



- Mode expansion monitors at the end of the 2.0 µm multimode section and at the taper end, 3D FDTD at mesh 4
- Of the 7.5 % (0.34 dB) through loss, 2.2 % is TE2 and TE4 left at the end of the section and 2.7 % is TE0 lost in the 3 µm taper; 2.5 % scatters at the centre
- Side arm: our −42.6 dB was the port power over all modes. TE0 alone is −55.7 dB, so the crosstalk was not underestimated

# C. Crossing taper: three options compared, 6 µm chosen

C

A 6 µm taper delivers the multimode field back into the guide. It halves the crossing loss and lowers the crosstalk by 4 dB.



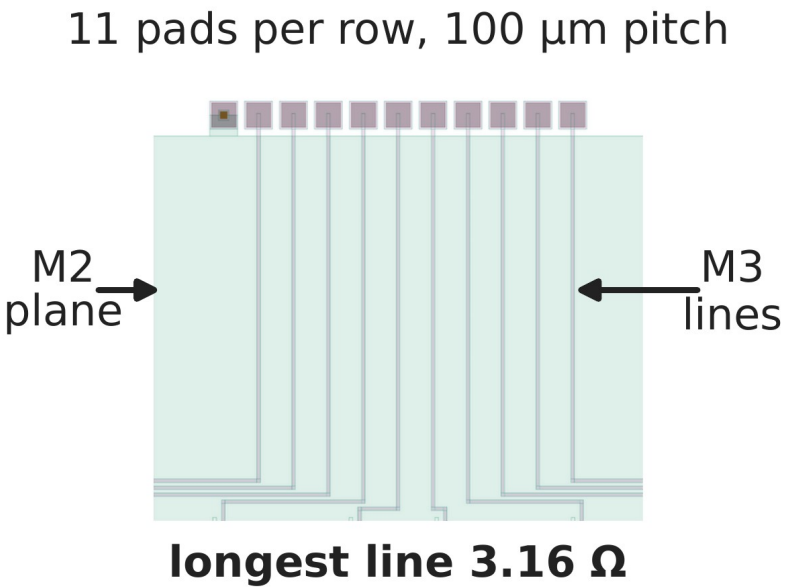

| Taper | Cell | Through loss | Crosstalk |
|---|---|---|---|
| 3 µm | 14.9 µm | 0.33 dB | −43.1 dB |
| **6 µm** | **20.9 µm** | **0.16 dB** | **−47.0 dB** |
| 9 µm | 26.9 µm | 0.17 dB | −42.8 dB |

- 9 µm gains no more loss and leaks more, so we chose 6 µm (1550 nm, 3D FDTD; C-band worst 0.17 dB, −45.0 dB)
- It stays at 0.161 to 0.169 dB in every process corner (3 µm: 0.33 to 0.42 dB)
- It fits the existing wiring regions with 0 DRC markers. The chip's worst path drops from 4.77 to 3.55 dB

# D. Pads and heater routing

D

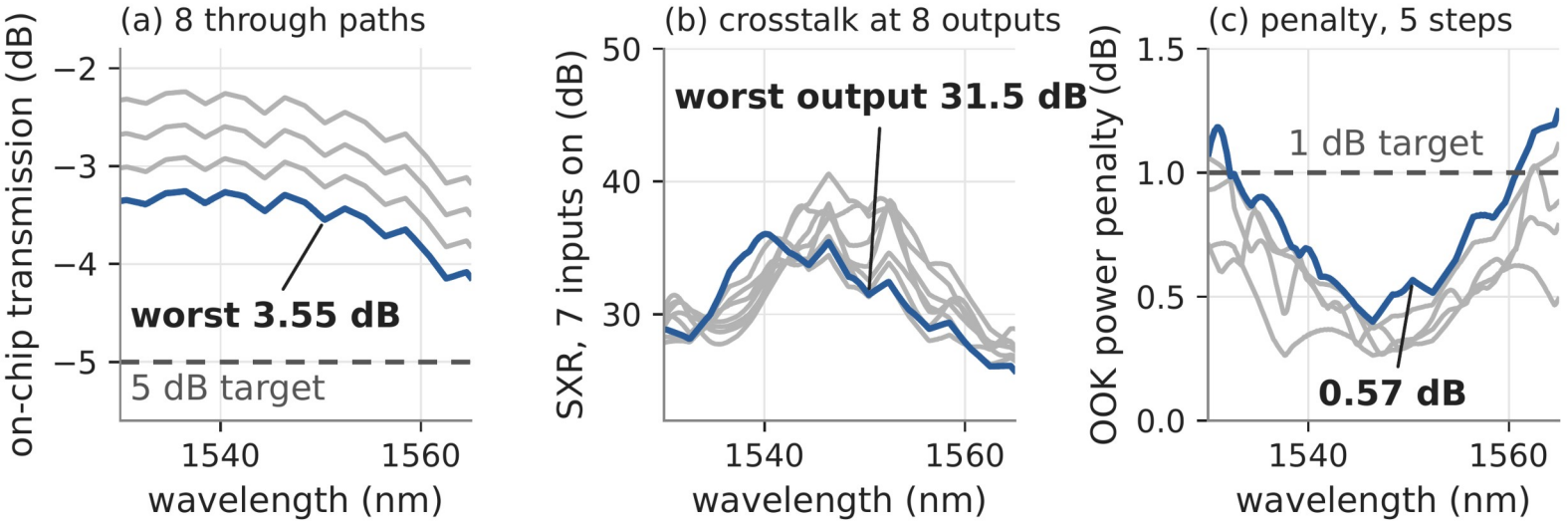
Two pad rows and metal-2 ground planes keep every heater line short and its ground quiet.




1. 22 pads, 80 µm at 100 µm pitch, in a north and a south row. Each heater line stays on its half with no metal crossing
2. Heater lines on M3, 10 µm wide. The longest is 2.34 mm and 3.16 Ω: 17.6 mV at the π current, 0.4 % of the heater voltage
3. Returns on M2 ground planes: 4.7 mV ground rise against 335 mV with a 10 µm bus
4. 60 µm lane pitch keeps each heater 50 µm from the next MZI. 42 µm gave 40 spacing violations

# Full chip in INTERCONNECT

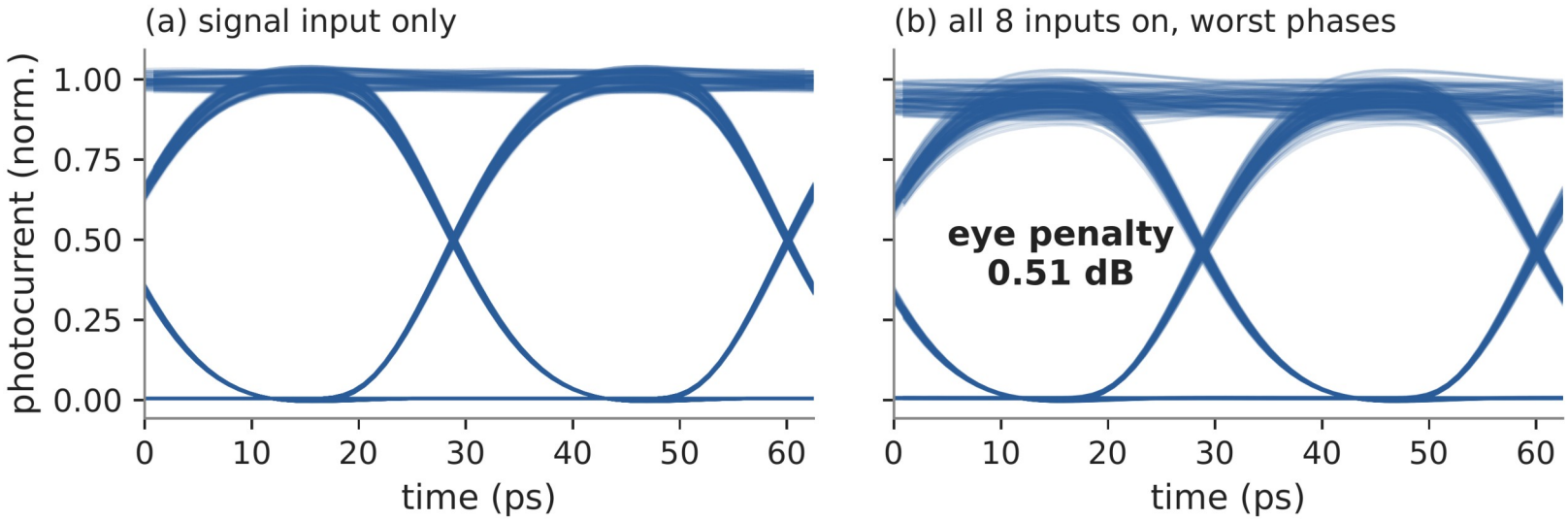
The full chip, wired as drawn, loses 3.55 dB on its worst path and keeps the all-inputs penalty at 0.57 dB, inside both targets.




- Worst path 3.55 dB on chip and 6.23 dB fibre to fibre (5 MZIs, 7 crossings), 4.15 dB at the C-band edge
- All 8 inputs on: penalty 0.57 dB against 1 dB, and 1 dB holds from 1532 to 1561 nm with the heaters set at 1550 nm
- Checks: every path equals a hand budget from the layout within 0.004 dB, our Python model agrees within 0.02 dB, and the Banyan from the same cells loses 2.77 dB at the same 0.57 dB

# Full chip at 32 Gb/s: eye diagram

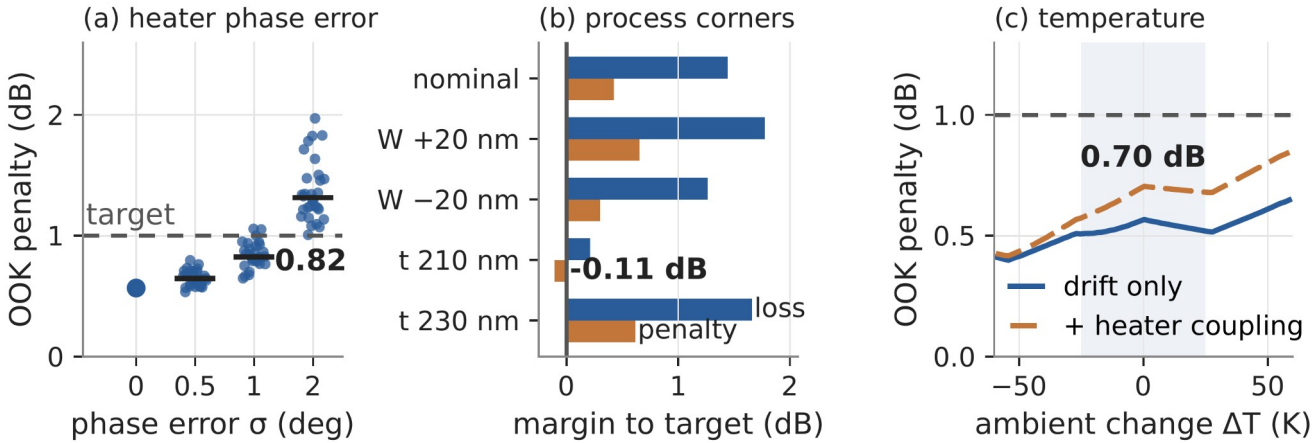
At 32 Gb/s NRZ with all eight inputs on, the worst output keeps an open eye with a 0.51 dB penalty.




1. INTERCONNECT time-domain run of the full chip: worst output, all 8 inputs on with worst-case laser phases
2. Eye penalty 0.51 dB, in line with the 0.57 dB power-sum penalty
3. BER without FEC: Q 23.3, below $10^{-15}$, with 0 dBm launch and −12 dBm receiver sensitivity (assumed)

# Stressed results

Temperature barely matters. MZI phase accuracy and a thin silicon layer set the margin.




| (a) MZI phase error | (b) Process corners | (c) Temperature |
|---|---|---|
| **How:** a random error added to all 20 MZI phases in INTERCONNECT, 30 random chips for each σ | **How:** MMI and crossing re-run in 3D FDTD at width ±20 nm and Si 210 or 230 nm, then put into the chip | **How:** ambient change shifts every index by $2\times10^{-4}$/K (0.07 nm/K); heater coupling from HEAT, 0.47° per neighbour |
| **Result:** σ 0.5°, every chip within 1 dB; σ 1°, 90 % | **Result:** all pass except 210 nm Si, 1.11 dB | **Result:** ±25 K with heater coupling, 0.70 dB |

# System-level performance against targets

Every target is met at nominal. Settling time and the thin-silicon corner fall short.

| Metric | Target | Result | Margin |
|---|---|---|---|
| ports, channel | 8×8, 1550 nm | 8×8, 1550 nm | met |
| data rate | 32 Gb/s NRZ | eye penalty 0.51 dB, all inputs on | met |
| worst-path loss, on chip | 5 dB | 3.55 dB (C-band edge 4.15 dB) | +1.45 dB (+0.85) |
| fibre to fibre | 9 dB | 6.23 dB (6.56 dB) | +2.77 dB |
| coupling loss per facet | 2 dB | 1.34 dB (1.58 dB) | +0.66 dB |
| crosstalk penalty, all inputs | 1 dB | 0.57 dB | +0.43 dB |
| BER without FEC | $10^{-12}$ | Q 23.3, below $10^{-15}$ | met |
| heater $P_{\pi}$ | 30 mW | 18.9 mW | +11.1 mW |
| heater power, chip | 600 mW | 246 to 303 mW per step, 379 mW all on | +221 mW |
| switching time | 30 µs | 27.0 µs (10–90 %) | +3.0 µs |
| settling to 1° | 30 µs | 78 µs | −48 µs |
| penalty at 210 nm Si | 1 dB | 1.11 dB | −0.11 dB |

# Conclusions, remaining risks and next steps

The chip works with margin at nominal and its layout is near final, from the provided pai_pdk.
Settling time and the thin-silicon corner are the open items.

## What works

- Full chip in INTERCONNECT, wired from the layout: 3.55 dB worst path, 0.57 dB all-inputs penalty
- PDR feedback answered: the mode check led to 6 µm crossing tapers, 1.2 dB off the worst path
- Heater 18.9 mW per π; layout matches the schematic and passes our DRC deck

## Still open and our plan

- Settling 78 µs: pre-emphasis and a guard time, or a second heater per MZI
- 210 nm Si corner: re-centre the MMI length, then rerun the corners