Two-level MUX Design and Exploration in FPGA Routing Architecture

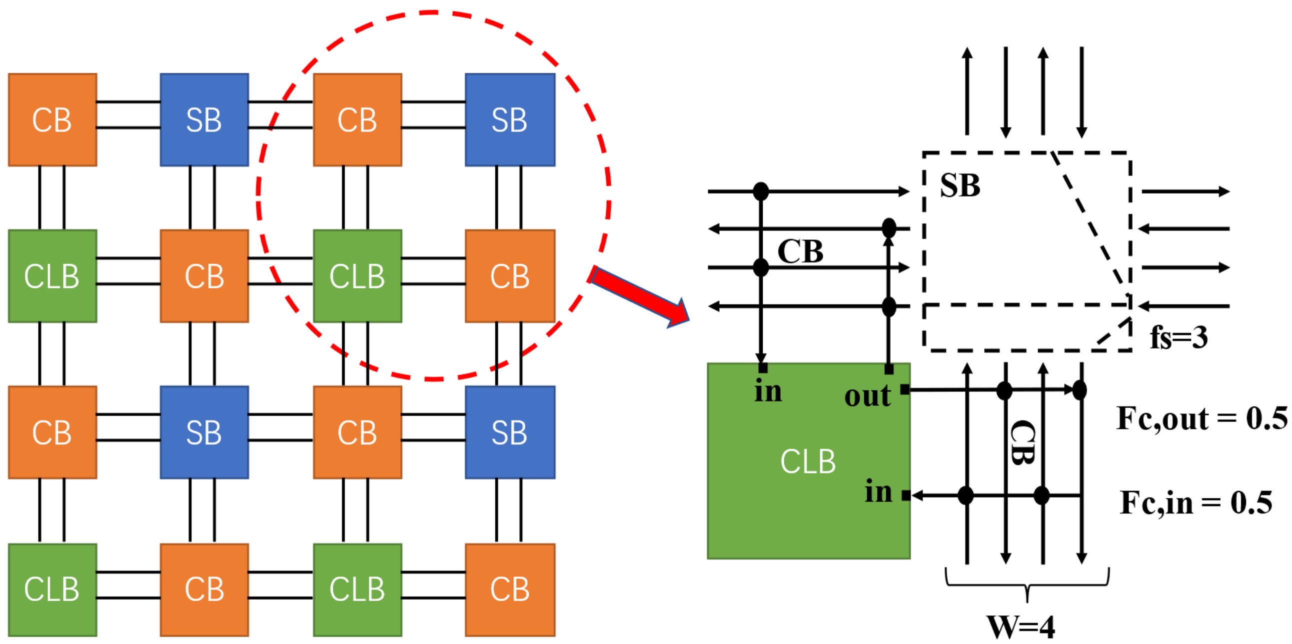
Authors: Yuhang Shen, Jiadong Qian,
Kaichuang Shi, Hao Zhou, Lingli Wang
FPL'21 2021-09-02

State Key Laboratory of ASIC and System
Fudan University, Shanghai, 2021

Outline

- Two-level MUX motivation
- Related Work
- Two-level MUX model
- SB and IB design with two-level MUX model
- Experimental methodology
- Two-level MUX optimization
- Tile comparison with CB-SB architecture
- Benchmark comparison with CB-SB architecture
- Summary and Future work

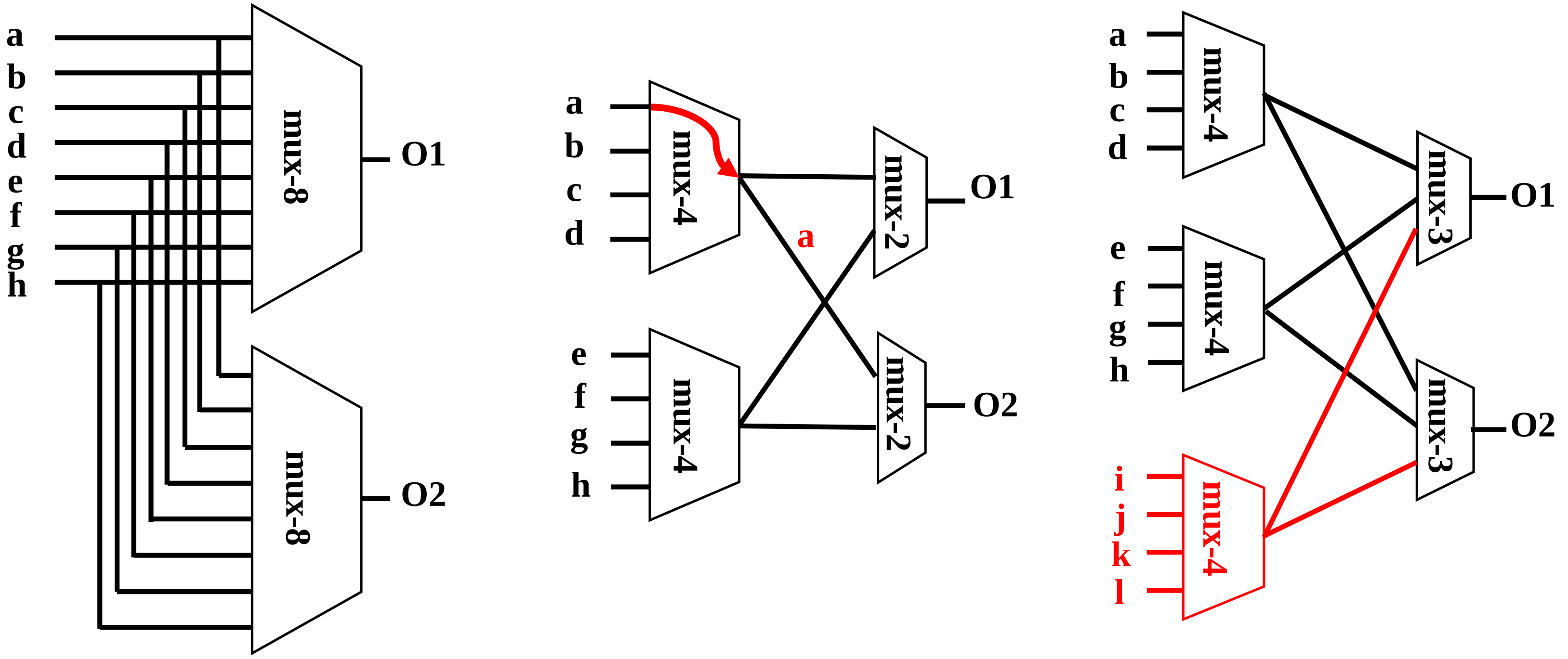
Two-level MUX motivation



Challenges:

- ➢ Occupy much area
- ➢ Load impact on routing wires

Two-level MUX motivation



Two-level MUX provides more design space:

- Decrease load on routing wires
- Change area
- Decrease routability but could use spare wires to compensate

Related work

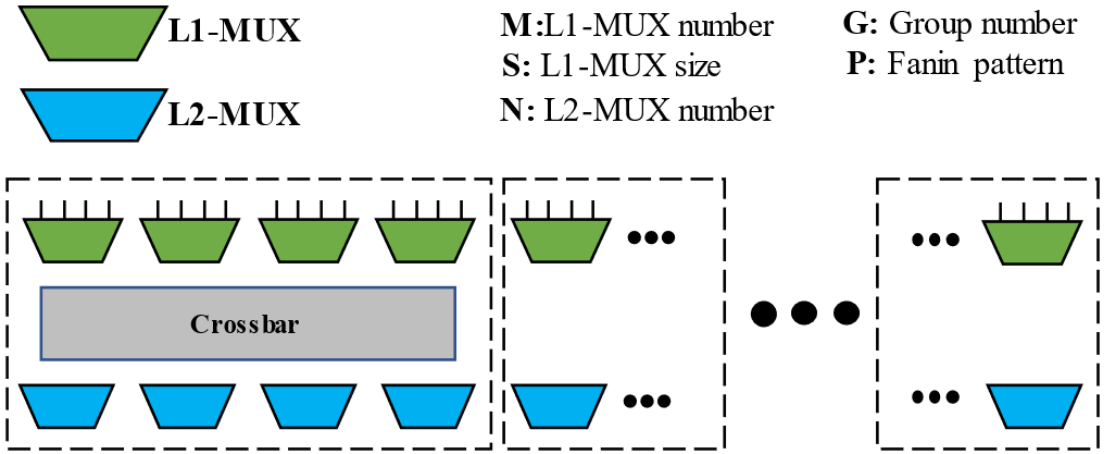
G. Lemieux and D. Lewis, “Using sparse crossbars within LUT clusters,” ACM/SIGDA Int. Symp. F. Program. Gate Arrays - FPGA, pp. 59–68, 2001

- Propose a 50% populated or sparser crossbar inside logic cluster to save area
- Using spare inputs to offset the loss of routability

W. Feng and S. Kaptanoglu, “Designing Efficient Input Interconnect Blocks for LUT Clusters Using Counting and Entropy,” ACM Trans. Reconfigurable Technol. Syst., vol. 1, no. 1, pp. 1–28, 2008

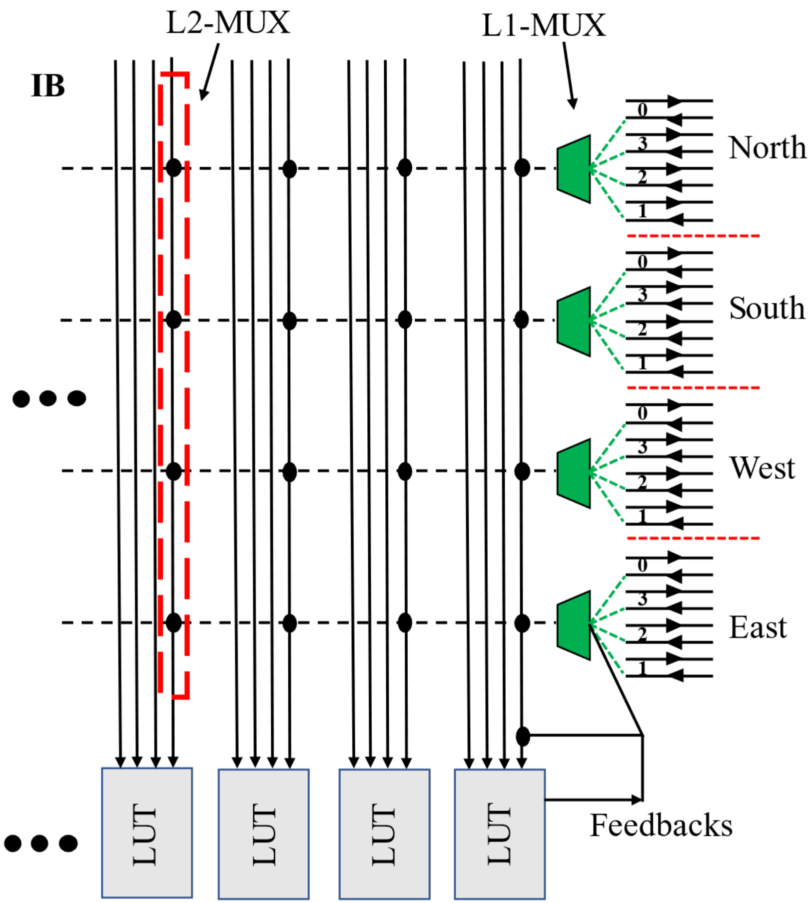
- Propose 2-level IIB(input interconnect block) to reduce area
- Evaluate area and routability but doesn’t consider timing change

Two-level MUX model

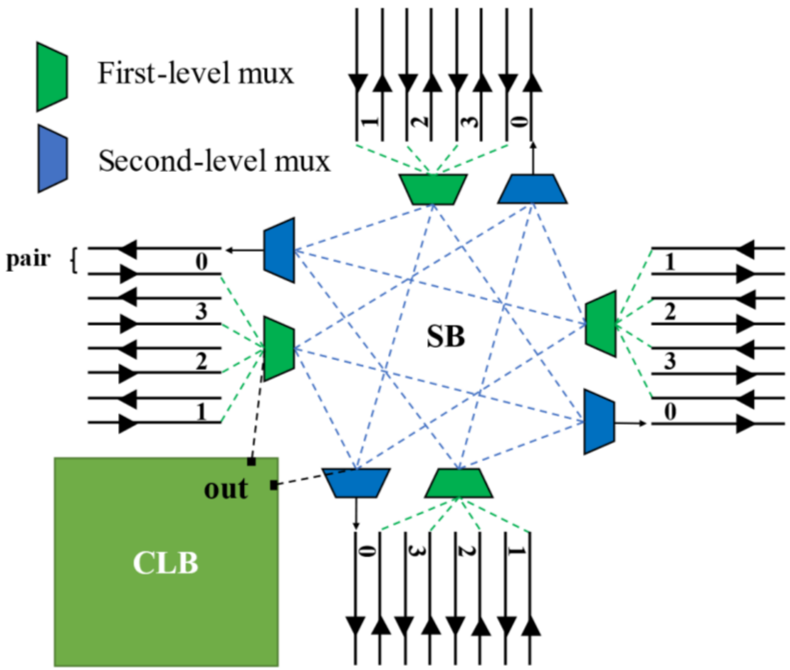


- We apply this model to design switch block(SB) and input block(IB)
- **M:** the number of L1-MUXes, equivalent to input bandwidth
- **S:** fan-in size of L1-MUXes
- **N:** the number of L2-MUXes, determined by the number of sinks
- **G:** the number of sub-groups when partitioning a switch matrix
- **P:** fan-in pattern, determines what signals construct L1-MUX inputs in terms of routing wire direction and switch point of segment length. (same/different Direction same/different Length)

IB and SB design with two-level MUX model



IB design example

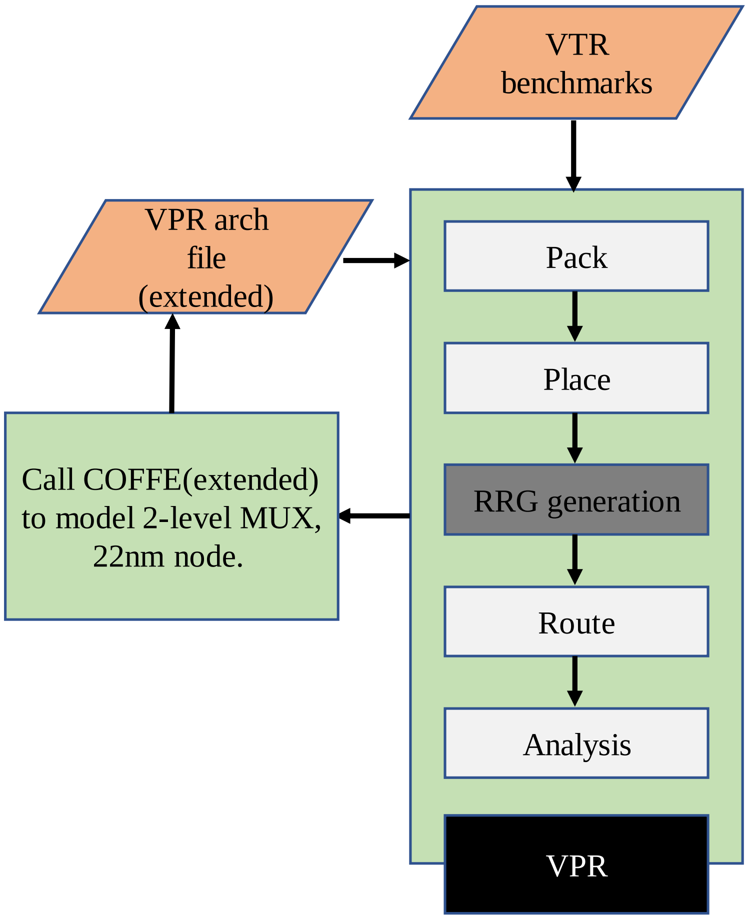


SB design example

Experimental methodology

Baseline architecture parameters

Parameters	Value
CLB Size	Eight 6-input LUTs
Wire Length	4
Channel Width	160
DSP	36×36 Fracturable Multipliers
Memories	32Kb Block RAMs
Output Connections	160
Feedback Connections	80
Fan-in Patterns	SDSL for SB, SDSL for IB
Sub-SB Numbers	20
Sub-IB Numbers	8
SB Input Bandwidth	100
IB Input Bandwidth	80



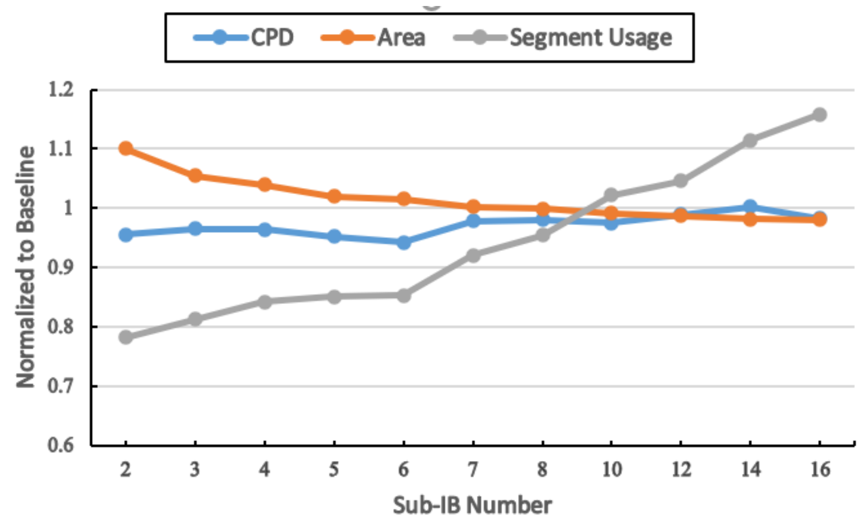
Three optimization objectives

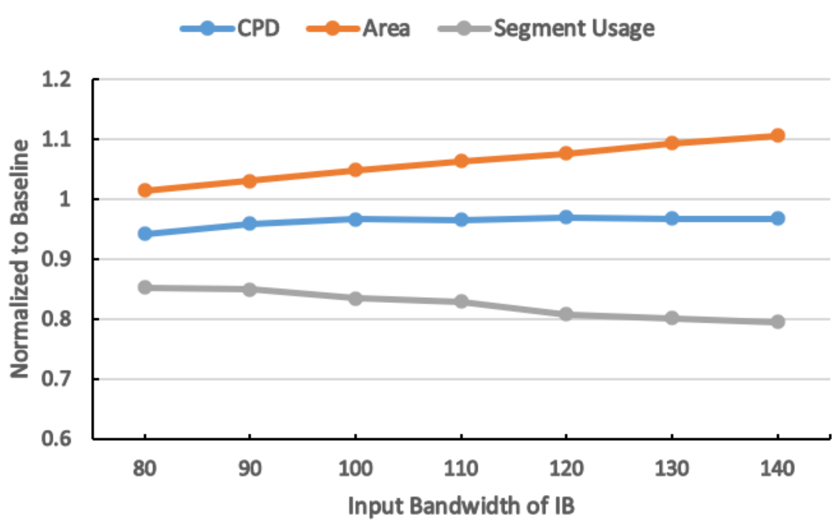
- Avg. critical path delay
- Avg. area
- Avg. segment usage

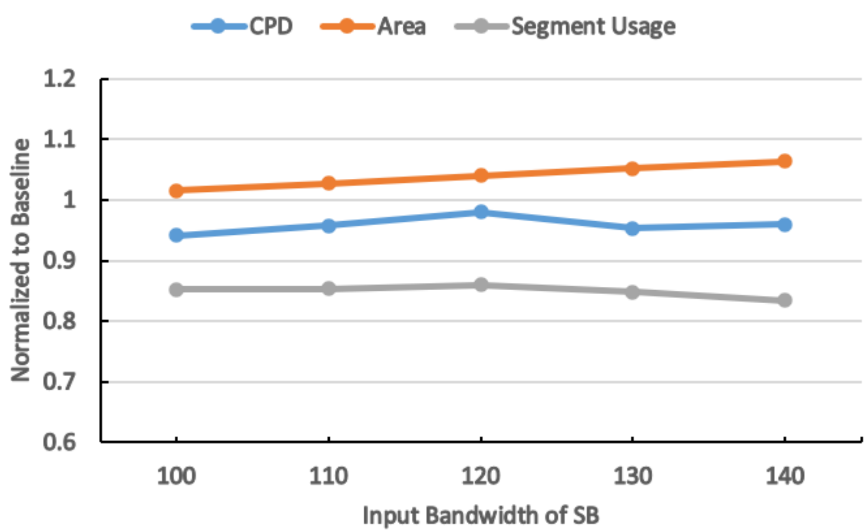
Two-level MUX optimization

Optimization for fan-in pattern, sub-IB number, input bandwidth of IB and SB.

Fan-in Pattern	Route Fails	CPD(ns)	Area(e+6)	Segment Usage
(DDDL,SDSL)	0	11.28	133.94	18.60%
(DDSL,SDSL)	0	11.07	133.94	17.89%
(SDDL,SDSL)	0	11.23	133.94	17.88%
(SDSL,SDSL)	**0**	**11.33**	**133.94**	**18.76%**
(DDDL,SDDL)	2	8.05	61.8	17.06%
(DDSL,SDDL)	7	5.20	25.17	11.19%
(SDDL,SDDL)	5	5.53	32.05	13.84%
(SDSL,SDDL)	0	11.18	133.94	18.65%

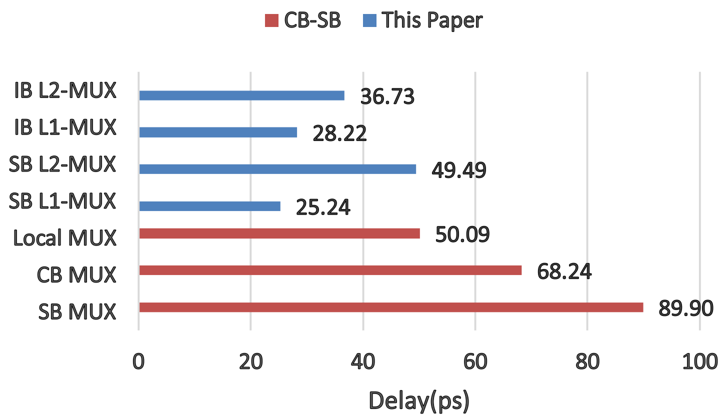


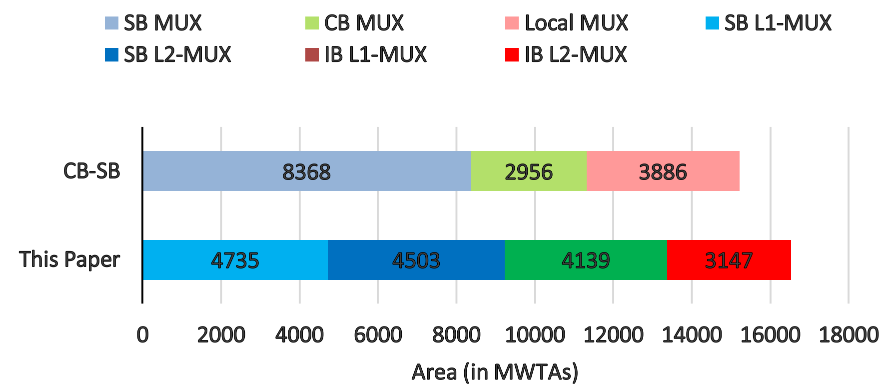




Tile comparison with CB-SB architecture

Architecture	MUX Type	MUX Num	MUX Size	Total Switch Count
CB-SB	CB MUX	32	16	2432
	Local MUX	48	20	
	SB MUX	80	12	
This Paper	IB L1-MUX	80	5	1820 (-25%)
	IB L2-MUX	48	13/14	
	SB L1-MUX	100	4	
	SB L2-MUX	80	4/5	





- The results are dependent on technology node and optimization constraints

Benchmark comparison with CB-SB architecture

Benchmark	CPD(ns)			Area(e+6)			Segment Usage		
	CB-SB	This Paper	Ratio	CB-SB	This Paper	Ratio	CB-SB	This Paper	Ratio
arm_core	11.24	8.80	78.3%	99.30	97.90	98.6%	39.7%	36.0%	90.7%
bgm	11.08	8.35	75.4%	189.90	184.53	97.2%	33.6%	30.4%	90.5%
blob_merge	6.10	4.23	69.3%	48.64	48.87	100.5%	30.5%	23.8%	78.0%
boundtop	1.49	0.94	63.2%	4.92	5.56	113.0%	3.4%	2.7%	79.4%
ch_intrinsics	1.80	1.66	92.1%	4.36	4.94	113.3%	3.5%	2.8%	80.6%
diffeq1	17.69	16.24	91.8%	9.17	10.89	118.8%	14.4%	9.1%	63.3%
diffeq2	13.54	11.96	88.3%	9.17	10.89	118.8%	10.2%	6.5%	64.1%
LU8PEEng	50.03	41.67	83.3%	205.38	207.06	100.8%	32.0%	27.3%	85.3%
LU32PEEng	51.69	38.12	73.7%	690.24	688.70	99.8%	41.3%	35.2%	85.2%
mcml	46.30	36.17	78.1%	644.21	634.20	98.4%	20.3%	19.9%	98.0%
mkDelayWorker32B	4.66	4.74	101.7%	106.52	105.93	99.4%	2.3%	2.0%	86.7%
mkPktMerge	3.41	3.51	102.8%	31.30	32.42	103.6%	5.0%	4.1%	82.1%
mkSMAdapter4B	3.81	3.20	83.9%	15.50	16.11	103.9%	14.9%	10.5%	70.5%
or1200	9.78	7.75	79.3%	28.14	26.69	94.8%	27.4%	22.0%	80.3%
raygentop	3.96	3.95	99.6%	18.17	22.67	124.8%	15.6%	10.7%	68.6%
sha	8.28	6.04	73.0%	19.25	20.10	104.4%	22.9%	16.8%	73.4%
stereovision0	2.20	1.72	78.3%	99.31	100.45	101.1%	13.4%	10.8%	80.6%
stereovision1	4.32	4.09	94.6%	90.80	97.90	107.8%	25.3%	19.6%	77.5%
stereovision2	10.64	9.01	84.6%	324.64	402.13	123.9%	30.0%	25.5%	85.0%
stereovision3	1.54	1.19	77.0%	0.75	1.00	133.3%	6.9%	4.2%	60.8%
Avg.		**-19.01%**		**3.00%**				**-3.63%**	

- 19% decrement in average critical path delay
- 3% increment in more area
- 3.6% decrement in average segment usage

Summary and future work

Summary:

- Introduce two-level MUX motivation and model
- Describe IB and SB design with two-level MUX model
- Perform P&R experiments with VPR and COFFE to search design space and compare with CB-SB counterpart in tile and benchmark level.

Future work:

- Better EDA support for modeling two-level MUX
- Apply and optimize two-level MUX towards different application domains
- Perform more detailed explorations on two-level MUX design parameters and different technology nodes