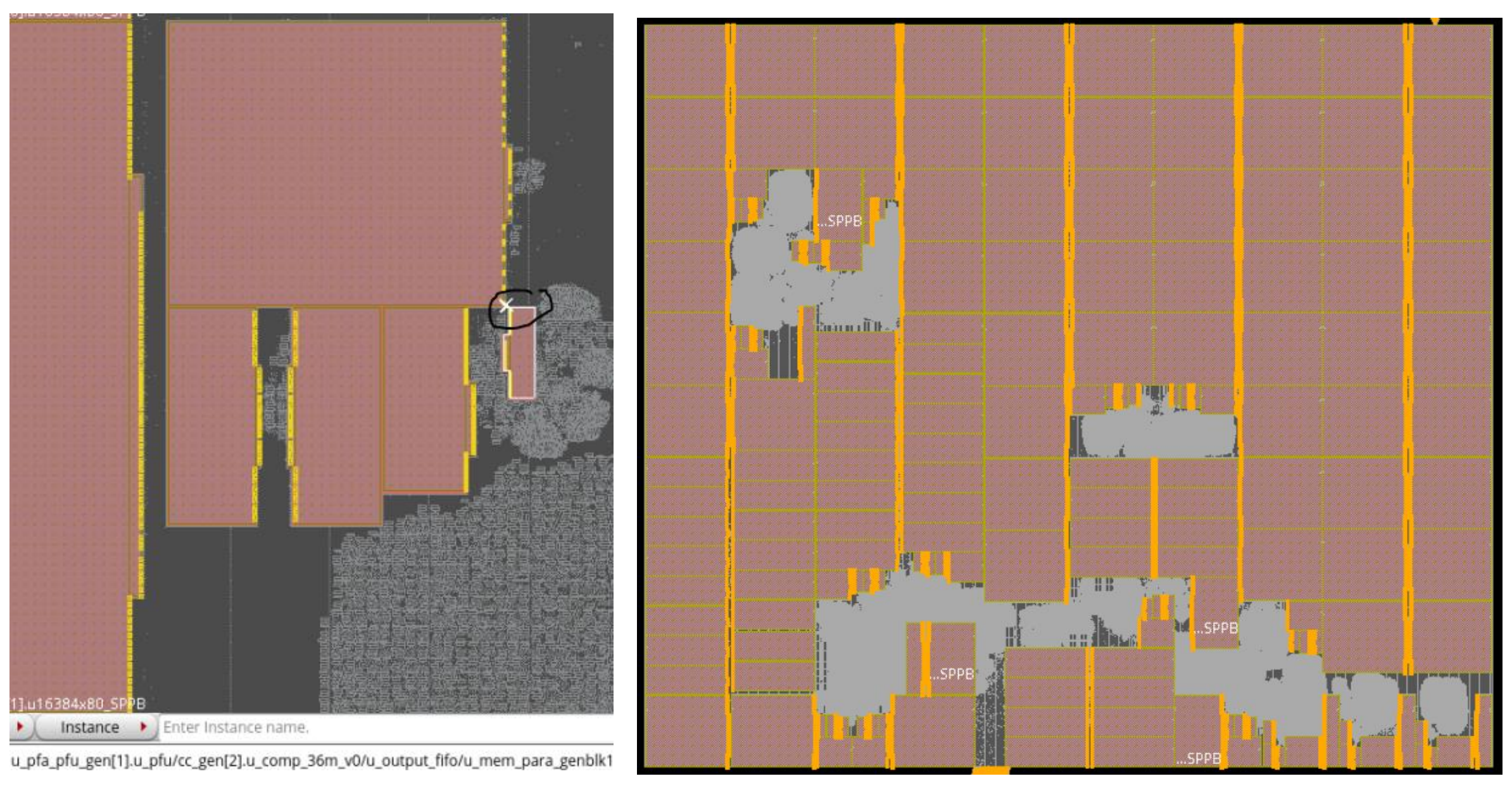
66TH IEEE INTERNATIONAL MIDWEST SYMPOSIUM ON CIRCUITS AND SYSTEMS

Design Space Exploration in the Physical-Design of an AI-Processor at 12nm using Relative-Placement Methodology

Mohit Sharma[1], Pavel Sinha[2], Doni Dattani[2], Mohammed Khalid[1]

[1]Department of Electrical and Computer Engineering, University of Windsor,
401 Sunset Avenue, Windsor, Ontario, N9B 3P4, Canada
[2]Aarish Technologies, 5735 Rue Parizeau, Brossard, Quebec, J4W 1H8, Canada

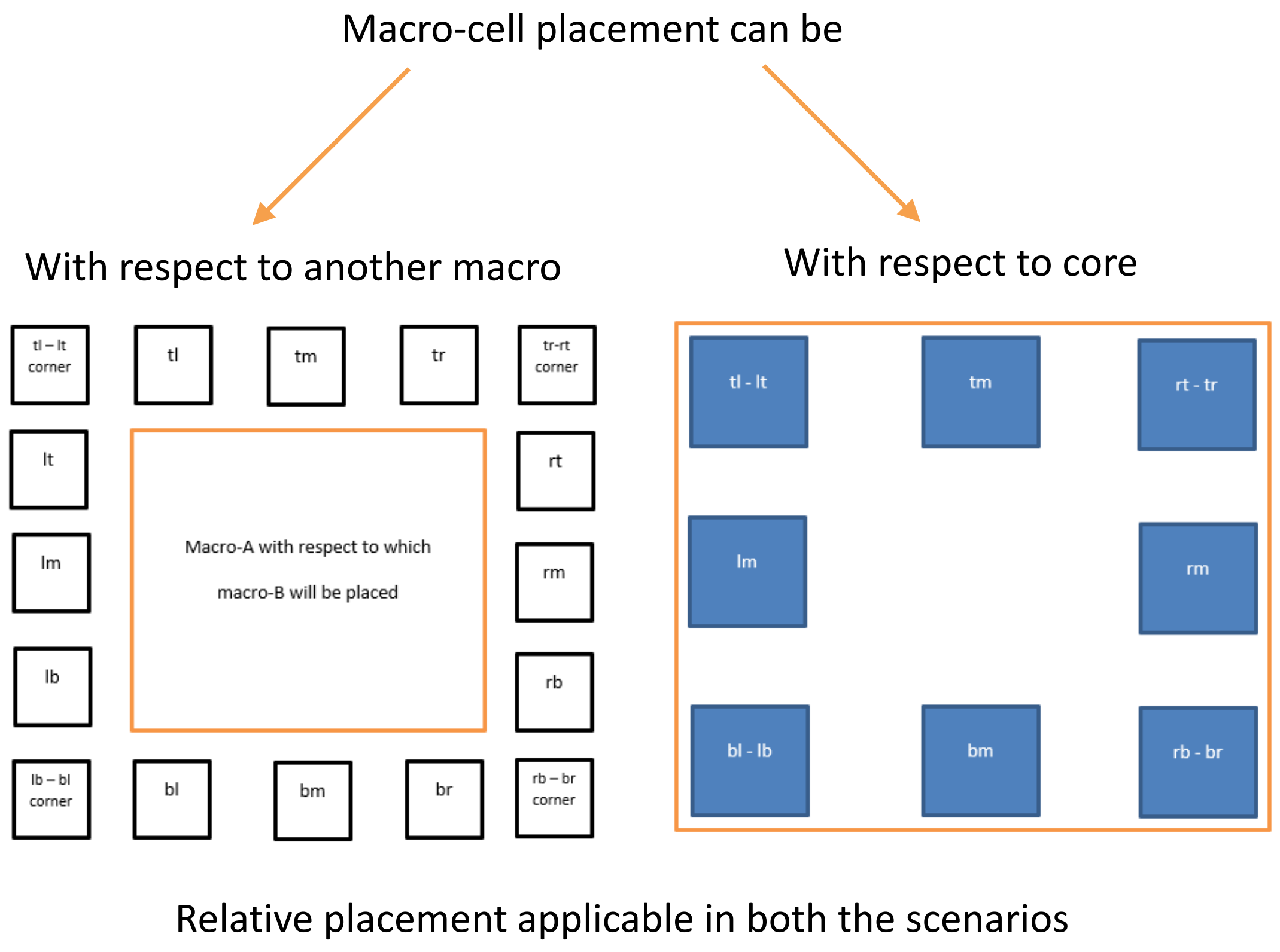
dBackground


...SPPB
u_pfu_pfu_gen[1].u_pfu/cc_gen[2].u_comp_36m_v0/u_output_fifo/u_mem_para_genblk1
Instance
Enter Instance name

- Macro cells are entities that represent certain design logic and can be viewed as black boxes. They are much larger in size as compared to a standard cell.
- Standard Cells (SC) are comparatively smaller in the circuit and have the same height. They need to be placed in specified rows when placing them.
- Macro-cell placement is the problem of placing a given macro cell circuit to optimize a certain design objective e.g., wirelength.
- A macro cell circuit contains a small number of macro cells and a larger number of standard cells.
- No overlap is allowed between any two cells.

Methodology

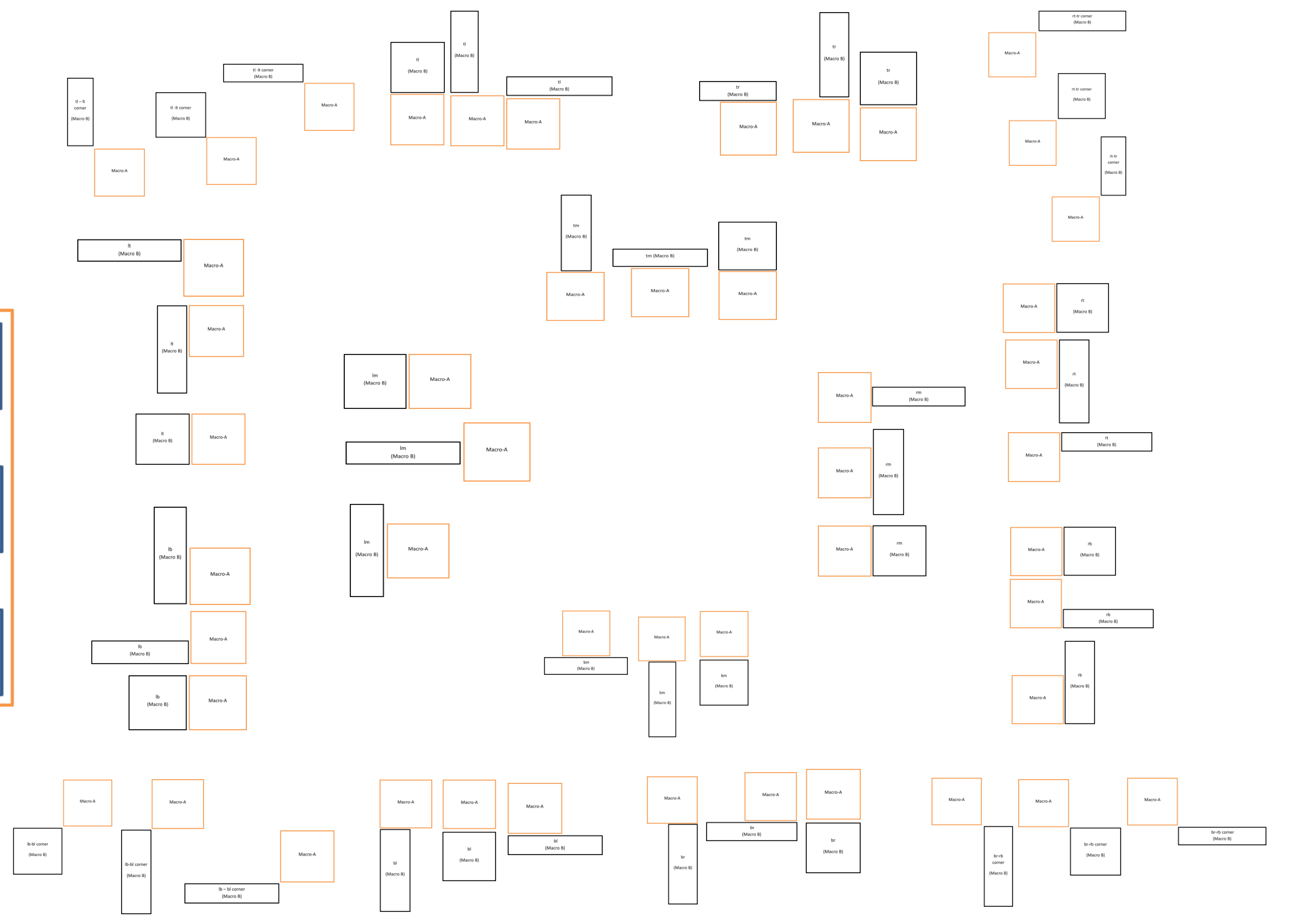
Relative Macro Placement


Macro-cell placement can be

With respect to another macro

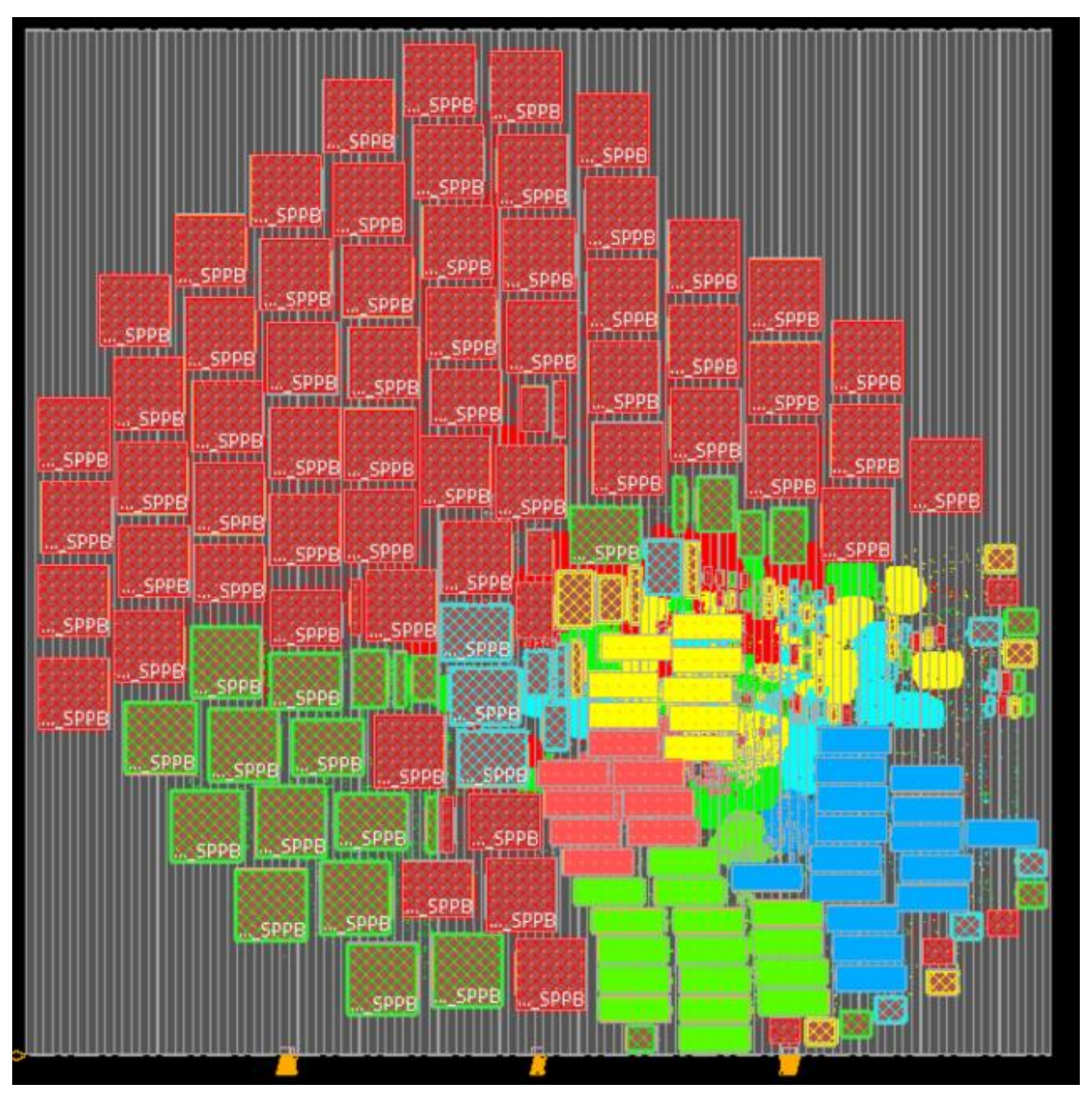
With respect to core

Relative placement applicable in both the scenarios



dMotivation

Challenge: Automated Memory Macro-Placement



- Commercially available physical design CAD tool for automated macro placement
  - Macro placement obtained is unoptimized.
  - It has large Area, Wirelength and Power.

Existing Approaches

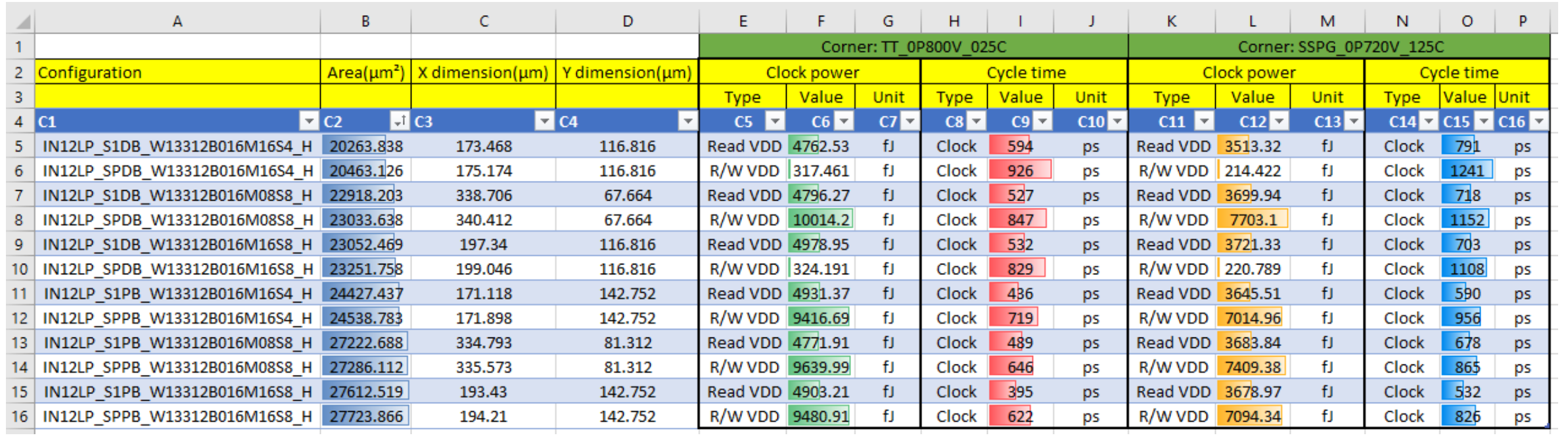
- Current research affirms that automatic macro placement doesn't achieve the most optimized metrics.
- Table on right shows results from a research on an automated macro-placement tool called Hierarchical Dataflow Placement (HiDaP) [2] and compares the placement metrics achieved for eight circuits with Industrial EDA tool (IndEDA) and hand-crafted Floorplan (handFP)
- Hand-crafted Floorplan gave the best placement metrics, among the three methods utilized, for the greatest number of circuits out of the eight circuits placed as highlighted in green.

Circuit	Flow	Wirelength		Congestion	Timing	
		WL	Norm.	GRC%	WNS%	TNS
c1	IndEDA	13.19	1.029	6.51	0.0	0
520k cells	HiDaP	13.40	1.046	7.83	0.3	0
32 macros	**handFP**	12.81	1.000	7.36	-0.2	0
c2	IndEDA	46.01	1.180	12.99	-44.5	-931
3.95M cells	HiDaP	40.72	1.045	13.00	-19.0	-329
100 macros	**handFP**	38.97	1.000	9.33	-11.2	-213
c3	IndEDA	44.83	1.175	10.09	-75.5	-553
3.78M cells	HiDaP	35.02	0.918	8.29	-17.5	-260
94 macros	**handFP**	38.16	1.000	9.15	-17.8	-317
c4	IndEDA	45.03	1.174	7.24	-54.4	-2167
4.81M cells	HiDaP	40.43	1.054	4.94	-31.2	-2686
122 macros	**handFP**	38.35	1.000	3.33	-22.8	-1736
c5	IndEDA	44.25	1.162	2.02	-30.8	-1940
1.39M cells	HiDaP	39.51	1.038	4.72	-25.1	-1149
133 macros	**handFP**	38.06	1.000	3.42	-39.8	-1017
c6	IndEDA	96.42	1.288	9.95	-70.0	-15341
2.87M cells	HiDaP	79.20	1.058	2.22	-37.0	-5051
90 macros	**handFP**	74.87	1.000	1.63	-27.3	-3688
c7	IndEDA	41.44	1.174	38.56	-34.9	-1060
1.67M cells	HiDaP	35.52	1.007	6.47	-29.9	-1059
108 macros	**handFP**	35.29	1.000	4.61	-20.4	-774
C8	IndEDA	24.85	0.987	1.02	-3.4	-44
2.20M cells	HiDaP	23.75	0.944	1.37	0.0	0
37 macros	**handFP**	25.17	1.000	0.93	-3.9	-24

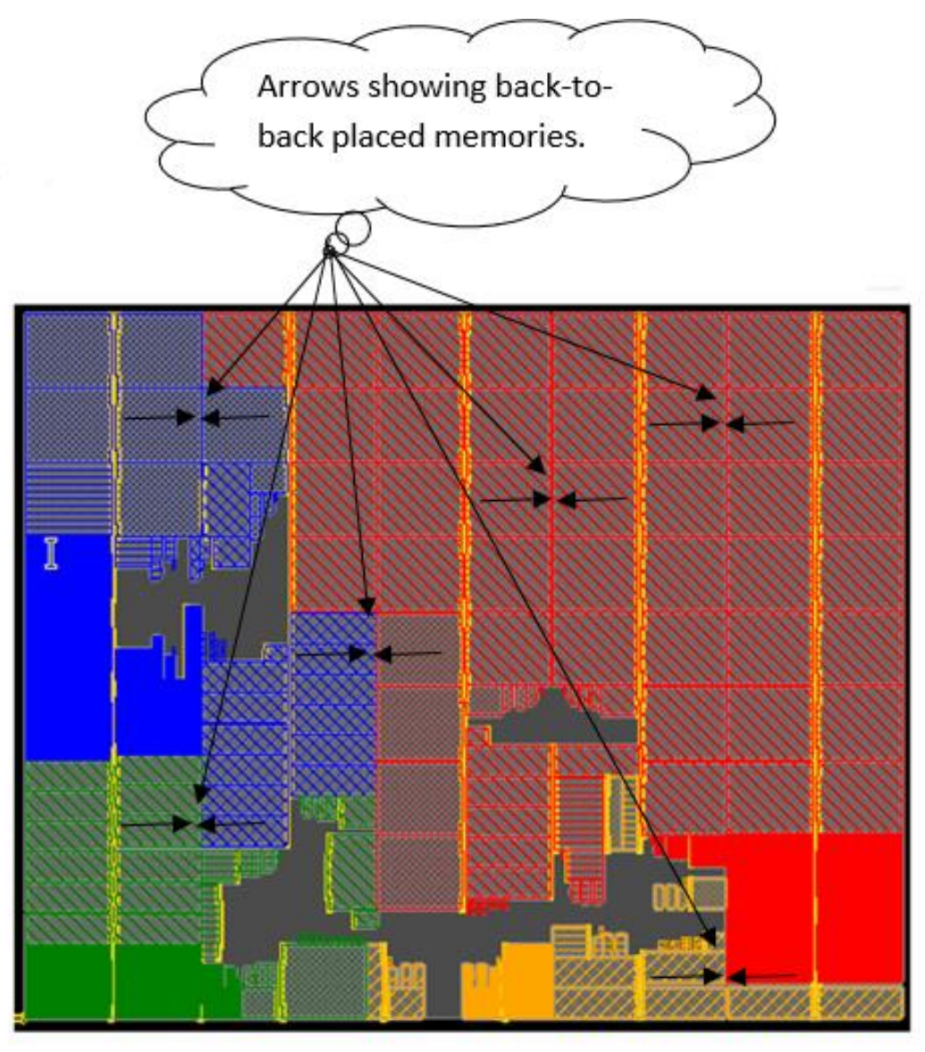
Designs	Std. Cell count	Macro count
swerv_wrapper	78K	28
ariane	114K	37
simd	207K	46
coyote	208K	15
bp_single	323K	49
ca53	445K	25

- Another research titled RTL-MP [3] presents improved results but it only considers designs with small number of macro cells. Table on the left sourced from [3] shows the macro cell count for the designs used for the research. The maximum macro-cell count used was about 50.

Macro Placement Rules



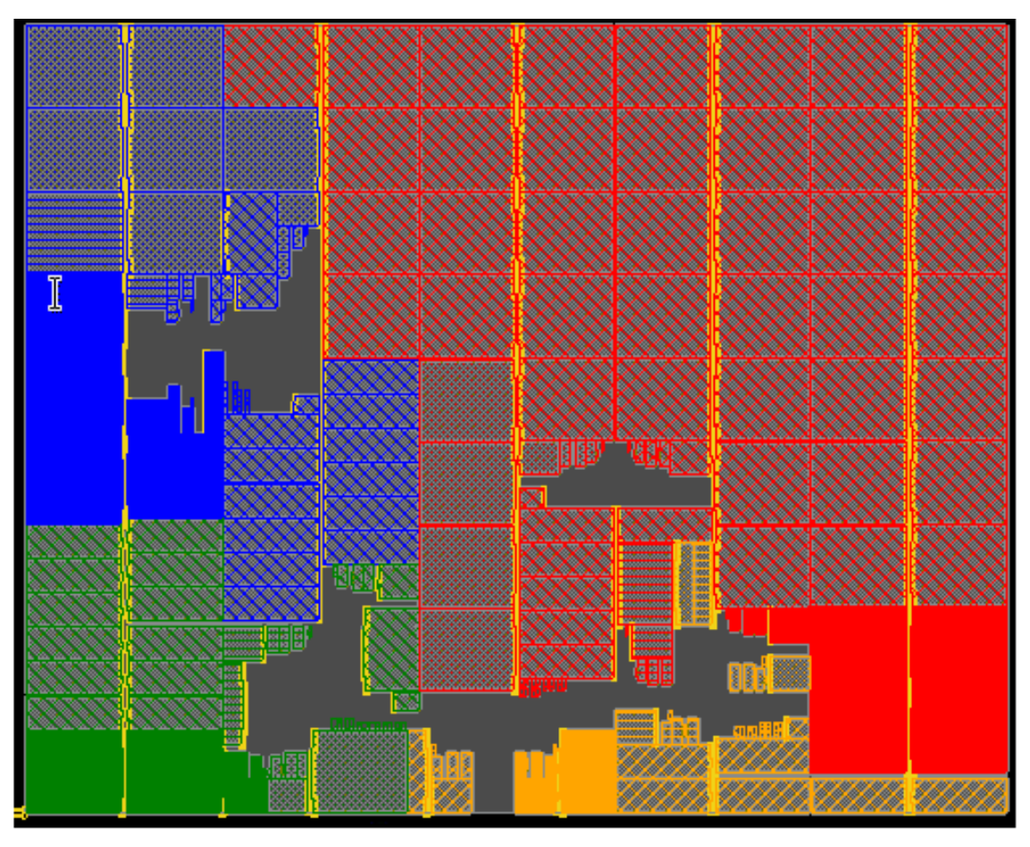
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1					Corner: TT_0P800V_025C						Corner: SSPG_0P720V_125C					
2	Configuration	Area(µm²)	X dimension(µm)	Y dimension(µm)	Clock power			Cycle time			Clock power			Cycle time		
3					Type	Value	Unit	Type	Value	Unit	Type	Value	Unit	Type	Value	Unit
4	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
5	IN12LP_S1DB_W13312B016M16S4_H	20263.838	173.468	116.816	Read VDD	4762.53	fJ	Clock	594	ps	Read VDD	3513.32	fJ	Clock	791	ps
6	IN12LP_SPDB_W13312B016M16S4_H	20463.126	175.174	116.816	R/W VDD	317.461	fJ	Clock	926	ps	R/W VDD	214.422	fJ	Clock	1241	ps
7	IN12LP_S1DB_W13312B016M08S8_H	22918.203	338.706	67.664	Read VDD	4796.27	fJ	Clock	527	ps	Read VDD	3699.94	fJ	Clock	718	ps
8	IN12LP_SPDB_W13312B016M08S8_H	23033.638	340.412	67.664	R/W VDD	10014.2	fJ	Clock	847	ps	R/W VDD	7703.1	fJ	Clock	1152	ps
9	IN12LP_S1DB_W13312B016M16S8_H	23052.469	197.34	116.816	Read VDD	4978.95	fJ	Clock	532	ps	Read VDD	3721.33	fJ	Clock	703	ps
10	IN12LP_SPDB_W13312B016M16S8_H	23251.758	199.046	116.816	R/W VDD	324.191	fJ	Clock	829	ps	R/W VDD	220.789	fJ	Clock	1108	ps
11	IN12LP_S1PB_W13312B016M16S4_H	24427.437	171.118	142.752	Read VDD	4931.37	fJ	Clock	436	ps	Read VDD	3645.51	fJ	Clock	590	ps
12	IN12LP_SPPB_W13312B016M16S4_H	24538.783	171.898	142.752	R/W VDD	9416.69	fJ	Clock	719	ps	R/W VDD	7014.96	fJ	Clock	956	ps
13	IN12LP_S1PB_W13312B016M08S8_H	27222.688	334.793	81.312	Read VDD	4771.91	fJ	Clock	489	ps	Read VDD	3683.84	fJ	Clock	678	ps
14	IN12LP_SPPB_W13312B016M08S8_H	27286.112	335.573	81.312	R/W VDD	9639.99	fJ	Clock	646	ps	R/W VDD	7409.38	fJ	Clock	865	ps
15	IN12LP_S1PB_W13312B016M16S8_H	27612.519	193.43	142.752	Read VDD	4903.21	fJ	Clock	395	ps	Read VDD	3678.97	fJ	Clock	532	ps
16	IN12LP_SPPB_W13312B016M16S8_H	27723.866	194.21	142.752	R/W VDD	9480.91	fJ	Clock	622	ps	R/W VDD	7094.34	fJ	Clock	826	ps



Back-to-Back facing memory macros

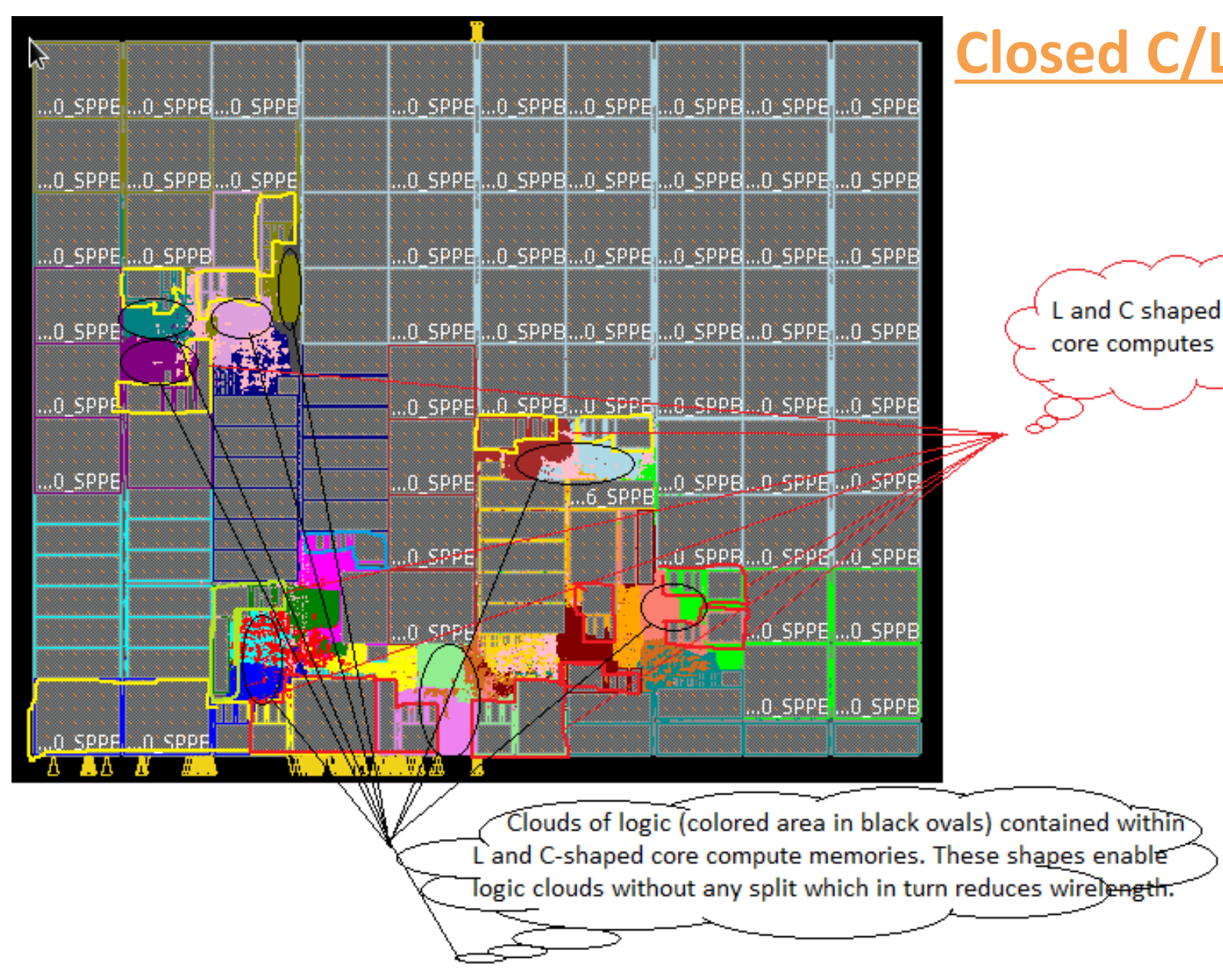
- Memory-macros were placed back-to-back with their non-pin facing sides touching each other.
- This ensured that standard-cells were placed close to the pin side of the memory macros. This also helped improve utilization.

Communicating CC memories

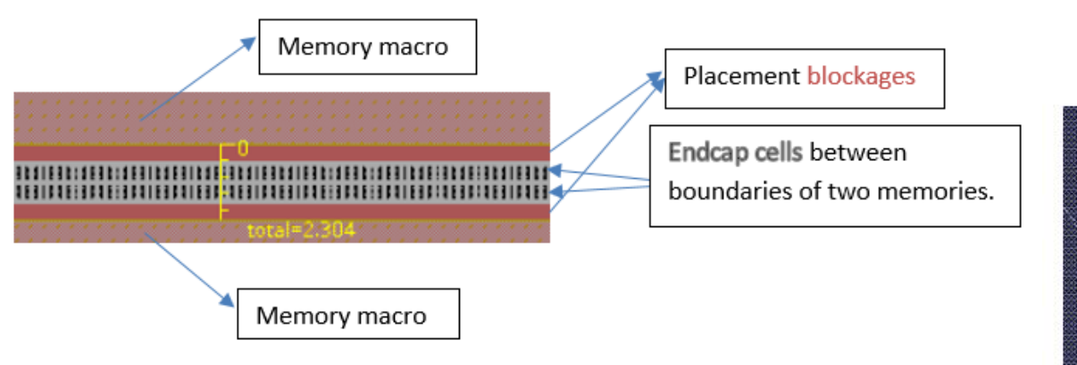


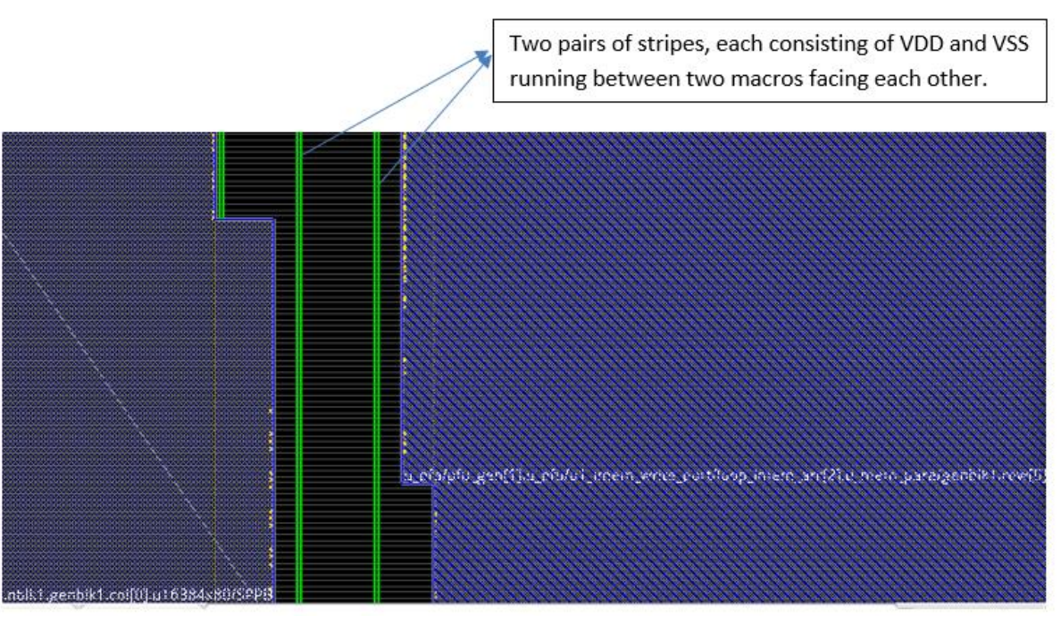
- Communicating memories placed close to each other to minimize wirelength.
- Figure on right shows memory macros that communicate with each other more often placed close to each other. PFU's are highlighted in four colors: Red, Green, Blue and Orange. CC memories are distinguished by pattern: Solid, Striped, Dotted, Crossed.

Closed C/L – shaped CC memories


...0_SPPE
...0_SPPB
...6_SPPB
L and C shaped core computes
Clouds of logic (colored area in black ovals) contained within L and C-shaped core compute memories. These shapes enable logic clouds without any split which in turn reduces wirelength.

- Core compute memories were placed in closed C-shape or L-shape
- Ensured that standard cell logic clouds were in spaces enclosed by corresponding CC memory macros.
- Helped reduce wirelength and improve timing.




Two pairs of stripes, each consisting of VDD and VSS running between two macros facing each other.

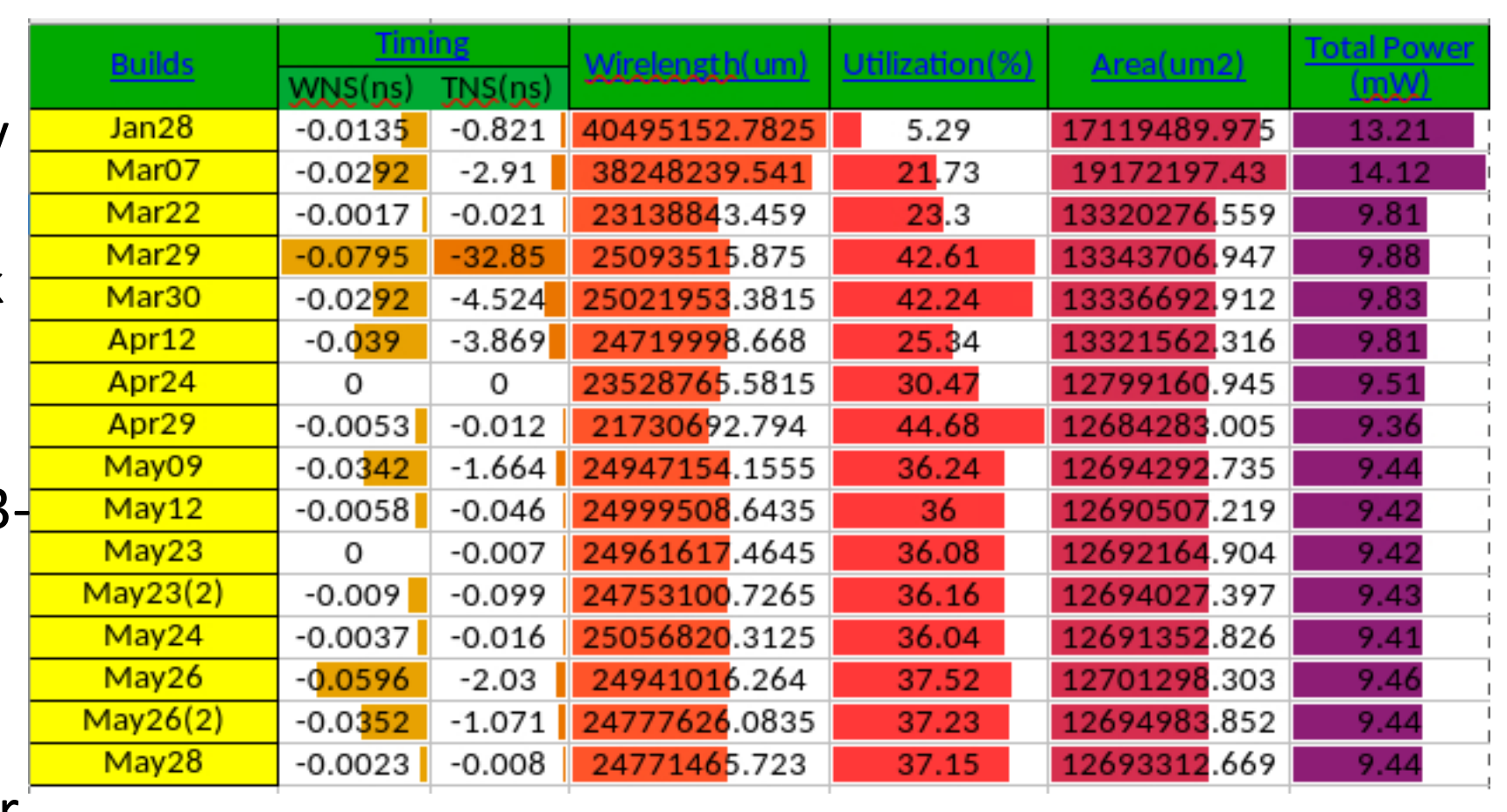
Space for placement blockages and endcap cells as well as power lines.

- Macros were separated by space enough for placement blockages and endcap cells.
- Pin facing sides of macros had enough space for two stripes of VDD and VSS each.

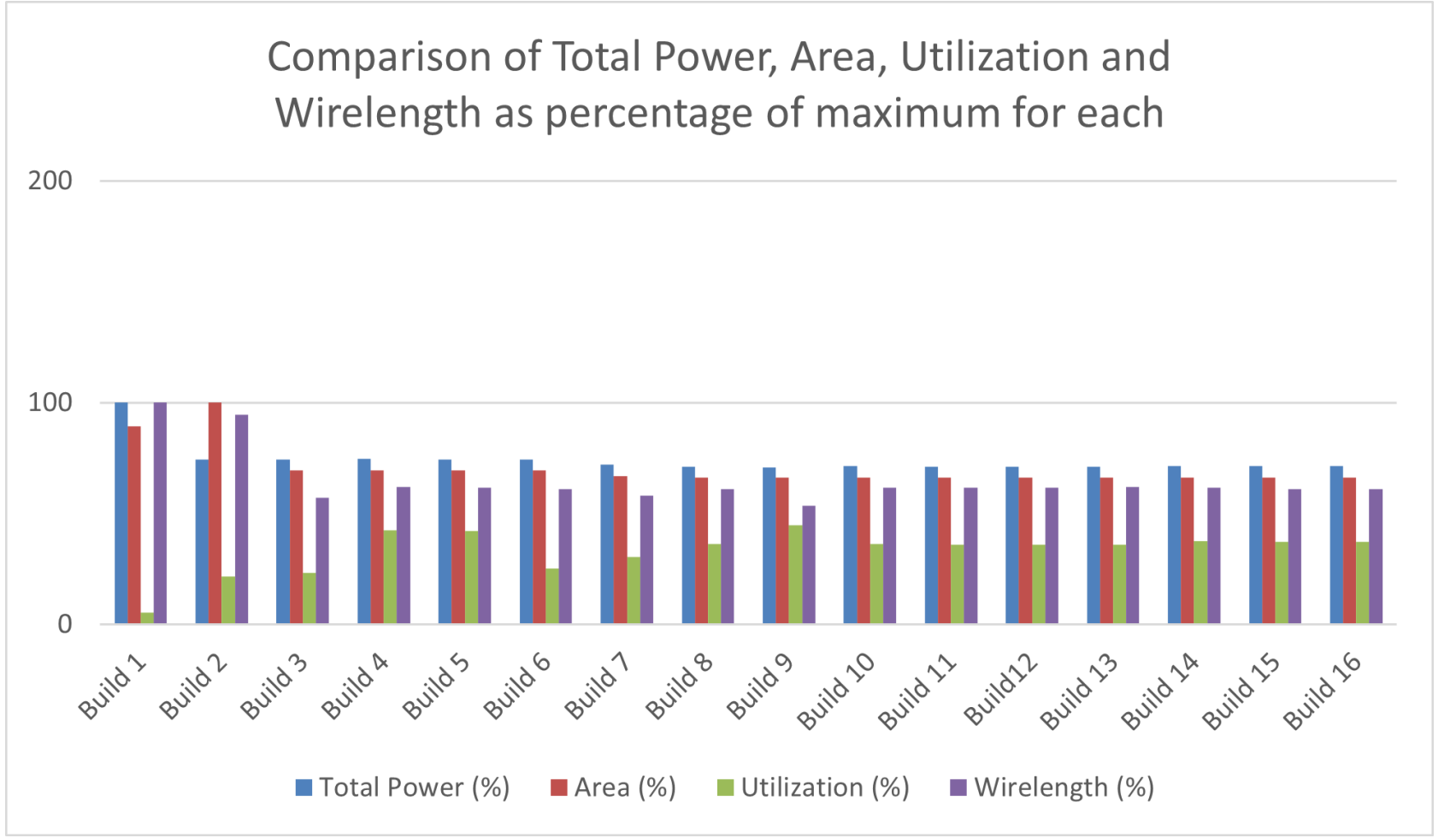
Experiments: Design Space Exploration

The script developed based on the Relative Placement methodology improved the turn-around time for the task of placement from multiple weeks for a subset of the design to 3-4 days for the complete design provided by our industry partner [6]. Optimized placement for the DSE was achieved in a span of two months.

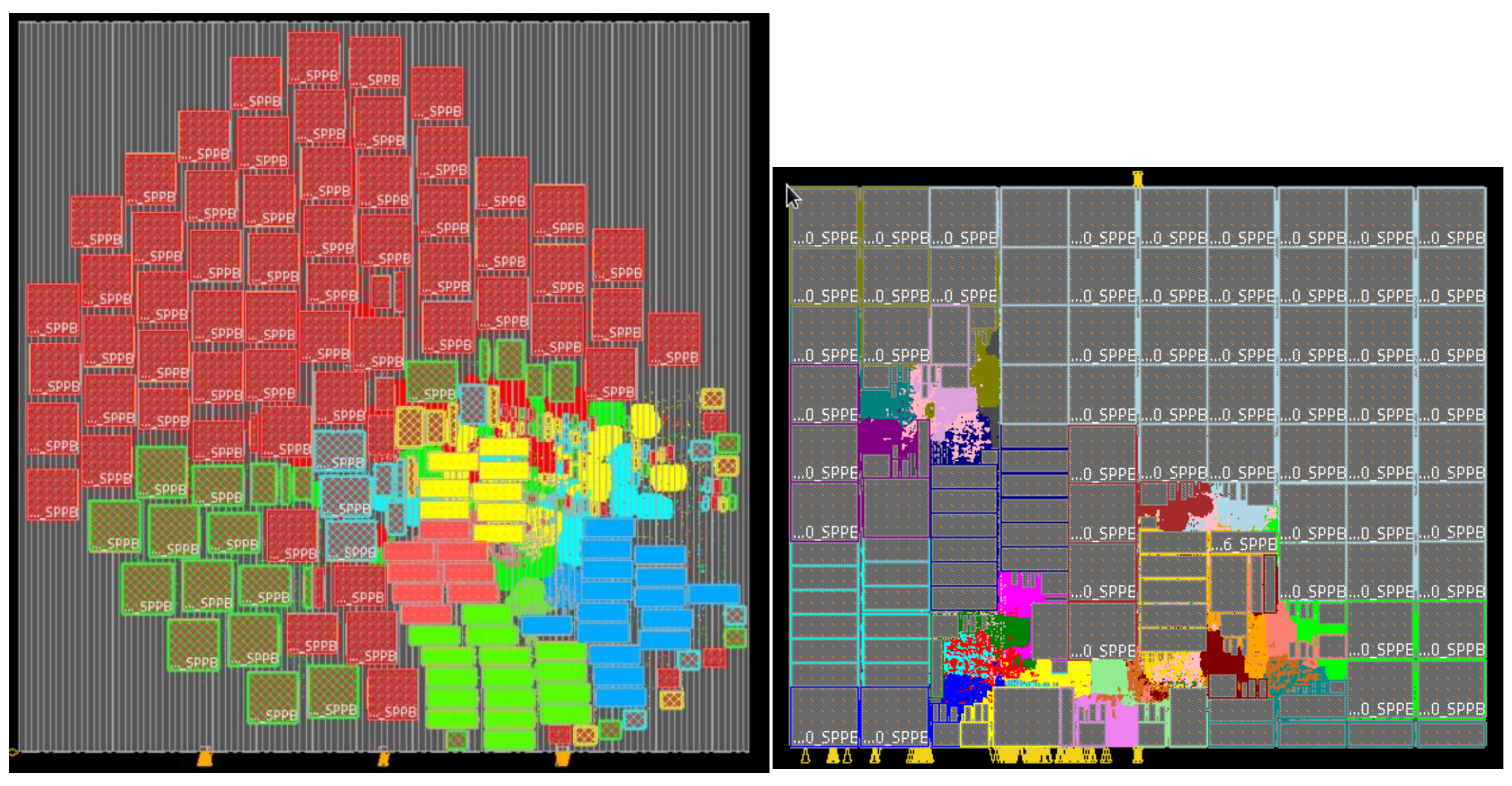
Figure presents the comparison of Total Power, Area, Utilization and Wirelength as a percentage of the maximum value of each metric.



Builds	Timing		Wirelength(um)	Utilization(%)	Area(um2)	Total Power (mW)
	WNS(ns)	TNS(ns)				
Jan28	-0.0135	-0.821	40495152.7825	5.29	17119489.975	13.21
Mar07	-0.0292	-2.91	38248239.541	21.73	19172197.43	14.12
Mar22	-0.0017	-0.021	23138843.459	23.3	13320276.559	9.81
Mar29	-0.0795	-32.85	25093515.875	42.61	13343706.947	9.88
Mar30	-0.0292	-4.524	25021953.3815	42.24	13336692.912	9.83
Apr12	-0.039	-3.869	24719998.668	25.34	13321562.316	9.81
Apr24	0	0	23528765.5815	30.47	12799160.945	9.51
Apr29	-0.0053	-0.012	21730692.794	44.68	12684283.005	9.36
May09	-0.0342	-1.664	24947154.1555	36.24	12694292.735	9.44
May12	-0.0058	-0.046	24999508.6435	36	12690507.219	9.42
May23	0	-0.007	24961617.4645	36.08	12692164.904	9.42
May23(2)	-0.009	-0.099	24753100.7265	36.16	12694027.397	9.43
May24	-0.0037	-0.016	25056820.3125	36.04	12691352.826	9.41
May26	-0.0596	-2.03	24941016.264	37.52	12701298.303	9.46
May26(2)	-0.0352	-1.071	24777626.0835	37.23	12694983.852	9.44
May28	-0.0023	-0.008	24771465.723	37.15	12693312.669	9.44



Design Space Exploration Results



Conclusion and Future Work

- DSE using relative placement methodology led to a 25% area improvement. The methodology eased the process of macro placement and made placement iterations faster.
- Manual placement is a viable alternative to automated macro-placement until automated macro-placement research advances enough to give reliable placements for large number of macros.
- As a future task, the proposed relative placement script can be automatically generated by the tool based on GUI (Graphical User Interface) macro placements done by the user.
- Also, an RTL-Aware Dataflow-Driven Macro Placement tool that can achieve the physical placement of macros in an automated manner.

References

[1] I. L. Markov, J. Hu and M. -C. Kim, "Progress and Challenges in VLSI Placement Research," in Proceedings of the IEEE, vol. 103, no. 11, pp. 1985-2003, Nov. 2015, doi: 10.1109/JPROC.2015.2478963.

[2] A. Vidal-Obiols, J. Cortadella, J. Petit, M. Galceran-Oms and F. Martorell, "RTL-Aware Dataflow-Driven Macro Placement," 2019 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2019, pp. 186-191, doi: 10.23919/DATE.2019.8714812.

[3] A. B. Kahng, R. Varadarajan, and Z. Wang, 'RTL-MP: Toward Practical, Human-Quality Chip Planning and Macro Placement', in Proceedings of the 2022 International Symposium on Physical Design, Virtual Event, Canada, 2022, pp. 3–11.

[4] Ghimire, D.; Kil, D.; Kim, S.-h. A Survey on Efficient Convolutional Neural Networks and Hardware Acceleration. Electronics 2022, 11, 945. https://doi.org/10.3390/electronics11060945

[5] A. Reuther, P. Michaleas, M. Jones, V. Gadepally, S. Samsi, and J. Kepner, "Survey of Machine Learning Accelerators," 2020 IEEE High Performance Extreme Computing Conference (HPEC), 2020, pp. 1-12, doi: 10.1109/HPEC43674.2020.9286149.

[6] Configurable Processor for implementing convolution neural networks, by P. Sinha (2021, Jan 28) Patent PCT/IB2020/000609 [Online]. Available: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2021014215