

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Novel Computer Aided Design (CAD) Methodology for Emerging Technologies to Fight the Stagnation of Moore's Law

Permalink

<https://escholarship.org/uc/item/2ts172zd>

Author

Ho, Chia-Tung

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

**Novel Computer Aided Design (CAD) Methodology
for Emerging Technologies to Fight the Stagnation of Moore's Law**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical and Computer Engineering

by

Chia-Tung Ho

Committee in charge:

Professor Chung-Kuan Cheng, Chair

Professor Bill Lin, Co-Chair

Professor Sicun Gao

Professor Patrick Mercier

Professor Tajana Simunic Rosing

2022

Copyright
Chia-Tung Ho, 2022
All rights reserved.

The dissertation of Chia-Tung Ho is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my family.

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Table of Contents	v
List of Figures	viii
List of Tables	xii
Acknowledgements	xiv
Vita	xvi
Abstract of the Dissertation	xx
Chapter 1 Introduction and Preliminaries	1
1.1 Complementary-FET (CFET) Technology	3
1.2 Satisfiability Modulo Theories (SMT) for Standard Cell Layout	5
1.3 Preliminaries of CFET Standard Cell (SDC) Synthesis Framework	6
1.3.1 FET Placement Formulations	8
1.3.2 In-Cell Routing Formulations	9
1.3.3 Conditional Design Rule Constraints	12
1.4 Design Technology Co-Optimization (DTCO) and System Technology Co-Optimization (STCO) Flow	18
Chapter 2 Rotuability Driven Complementary-FET (CFET) Standard Cell Synthesis for Block-Level Area Optimization	20
2.1 Introduction	20
2.1.1 Related Works	22
2.1.2 Our Contributions	22
2.2 Routability-Driven Simultaneous Place-and-Route for Complementary-FET (CFET) Standard Cell Synthesis Framework	23
2.2.1 Overall Flow of Framework	24
2.2.2 CFET SDC Synthesis Framework Overview	25
2.2.3 CFET Cell Architecture and Abstract Pin Interface (API)	26
2.2.4 Dynamic Complimentary Pin Allocation (DCPA)	27
2.2.5 Routability-Driven Constraints and Objectives	30
2.2.6 Multi-Objective Optimization (Optimal Priority)	34
2.3 Experiments	37
2.3.1 Experimental Setup	37
2.3.2 CFET vs. Conv. Standard Cell (SDC)	40
2.3.3 Routability-Driven Cell Optimization	43
2.3.4 Block-Level Routability Analysis	47
2.4 Conclusion	51

Chapter 3	Complementary-FET (CFET) Standard Cell Synthesis Framework for Design and System Technology Co-Optimization	53
3.1	Introduction	53
3.1.1	Related Works	54
3.1.2	Our Contributions	55
3.2	Multi-Row CFET Standard Cell Synthesis Framework for DTCO and STCO Explorations	57
3.2.1	Multi-Row CFET SDC Synthesis Framework Overview	57
3.2.2	Multi-Row CFET Cell Architecture	58
3.2.3	Multi-Row Dynamic Complementary Pin Allocation	60
3.2.4	Parametric Conditional Design Rules	65
3.2.5	Multi-Row Cell Area Minimization	67
3.2.6	Multi-Objective Optimization (Optimal Priority)	69
3.3	Experimental Setup	70
3.4	Multi-Row Routability-Driven CFET Cell Optimization	73
3.4.1	Cell Area Minimization with Adaptive Cell Row Number	73
3.4.2	Routability-Driven Cell Optimization for Scaling	77
3.5	DTCO and STCO Exploration for CFET SDC Scaling	80
3.5.1	CFET vs. Conv. SDC	80
3.5.2	DTCO Exploration with CFET SDC Scaling	85
3.5.3	DTCO for Block-Level Area Scaling	93
3.6	Extreme CFET SDC Scaling	96
3.6.1	Scaling to Extreme 2 Routing Tracks (RTs) with Inter-Row Routing Options	96
3.6.2	Block-Level Area Scaling with 2.5T CFET	100
3.7	Conclusion	101
Chapter 4	Machine Learning Prediction for Design and System Technology Co-Optimization Sensitivity Analysis	103
4.1	Introduction	103
4.1.1	Related Works	104
4.1.2	Our Contributions	105
4.2	Design and System Technology Co-Optimization Sensitivity Prediction Framework	107
4.2.1	DTCO and STCO Sensitivity	107
4.2.2	Overall modeling flow	109
4.2.3	Methodology for feature extraction	110
4.2.4	Input features	112
4.2.5	Machine learning techniques	112
4.3	Experimental Results	118
4.3.1	Experiment Setup	118
4.3.2	Prediction Model Accuracy	125
4.3.3	Prediction of New Technologies	130
4.3.4	Prediction of New Power Delivery Network Setting	134
4.3.5	Robustness of New Circuit Prediction	137
4.4	Conclusion	139

Chapter 5	Conclusion and Future Directions	141
	5.1 Conclusion	141
	5.2 Future Directions	143
	5.2.1 CFET Standard Cell Synthesis for Power-Performance-Area (PPA) and Process-Aware Optimization	143
	5.2.2 Design and System Technology Co-Optimization Sensitivity Predic- tion for Block-Level Power-Performance-Area (PPA) Optimization .	144
	5.2.3 Routability-Aware CFET Standard Cell Synthesis using Reinforce- ment Learning	144
Bibliography	152

LIST OF FIGURES

Figure 1.1:	Technology node scaling roadmap [1]	2
Figure 1.2:	Dimensional scaling of standard cell in sub-7nm [2]	2
Figure 1.3:	An illustration of Conventional and Complementary FET (CFET) structure (Top Row). CFET shared and split Gate, Source and Drain structure (Bottom Row) [3–5].	4
Figure 1.4:	Grid-Based placement and routing graph	6
Figure 1.5:	An illustration of diffusion sharing/break and relative positioning constraints. (a) Diffusion sharing and diffusion break illustration. (b) Relative positioning constraints between two FETs.	8
Figure 1.6:	In-Cell routing flow formulation. (a) Flow Formulation (Section 1.3.2). (b) Conditional Design Rules (Section 1.3.3).	10
Figure 1.7:	Supernodes. PIN1 and PIN2 respectively cover four and two vertices on Upper and lower M0A/PC layers (i.e., inner pin S_i for P_{IN}). Outer pins (i.e., P_{EX}) are connected to vertices on M1 of G through Super Outer Node (S_o).	11
Figure 1.8:	An example of geometric variable $g_{d,v}$	13
Figure 1.9:	An example of minimum area rule (MAR). (a) MAR violation. (b) No MAR violation.	14
Figure 1.10:	An example of right-directional end-of-line spacing rule (EOL).	15
Figure 1.11:	An example of via rule (VR) when the minimum distance between vias is 1 grid. . .	16
Figure 1.12:	An example parallel running length (PRL) rule.	17
Figure 1.13:	An example step height rule (SHR).	17
Figure 1.14:	Design Technology Co-Optimization and System Technology Co-Optimization Flow	19
Figure 2.1:	An illustration of Conventional and Complementary FET (CFET) structure (Top Row). CFET shared and split Gate, Source and Drain structure (Bottom Row) [3–5].	21
Figure 2.2:	Overall flow of the proposed routability-driven simultaneous place-and-route CFET standard cell synthesis framework.	24
Figure 2.3:	CFET Standard Cell Synthesis Framework Overview.	25
Figure 2.4:	Grid-Based placement and routing graph with Abstract Pin Interface (API) using 4 RTs P-on-N CFET example.	26
Figure 2.5:	Concept of Dynamic Complementary Pin Allocation (DCPA) for CFET cell structure using 4 RTs P-on-N CFET example. p_1^P =P-FET Gate Pin. p_1^N =N-FET Gate Pin. . .	27
Figure 2.6:	An example of (a) MPL with MAR=2, (b) MPO with EOL/MAR = 1/1.	30
Figure 2.7:	An illustration of Pin Separation (PS) [6] and Edge-Based Pin Separation (EB-PS) [7].	32
Figure 2.8:	An illustration of Pin Separation (PS).	33
Figure 2.9:	An illustration of the Impact of M2 Blockage on the Routing Congestion.	35
Figure 2.10:	Cell Statistics of M0 Core, M1 Core, and AES.	38
Figure 2.11:	An example of transferring the grid-based conditional design rules to the block-level.	38
Figure 2.12:	An example of XOR2x1 Schematic Netlist [8] and CFET SDC Layout.	40
Figure 2.13:	Layouts of 4 routing tracks CFET and Conv. DFFHQN with corrected design constraints. Optimized result of CFET layout: Cell Size (19→16), Metal Length (613→182), #M2 Tracks (2→0). The red dash-line boxes are metal extension for PRL and SHR design constraints.	41
Figure 2.14:	Layout of NAND2x1 cell optimized for Pin Separation objective.	44

Figure 2.15:	Layout of AND3x1 cell optimized generated by Pin Separation [6] and Edge-Based Pin Separation (EB-PS) [7] objectives. The RPA (Worst Case) considers the parallel pin-shape of adjacent cell as described in Figure 2.7.	46
Figure 2.16:	Layouts of FAX1 cell optimized with <i>MinTrack</i> and <i>MinLength</i> objectives. The black dashed rectangles shows the FA I/Os on M1. Note that some signals and I/Os need to be routed with M2 to complete the routing.	48
Figure 2.17:	Block-Level Placement and Route Results of M0 core, M1 core, and AES designs of wPS and woPS under MPO=2 and MPO=3 constraints.	49
Figure 2.18:	Block-Level P&R Results of M0 core, M1 core, and AES designs of <i>MinTrack</i> and <i>MinLength</i>	49
Figure 2.19:	Block-level P&R Results of M0 core, M1 core and AES of proposed routability-driven cell optimization, PS [6], and SP&R [9] objectives CFET SDCs.	50
Figure 2.20:	P&R design views of M0 core at 0.82 util. with proposed routability-driven CFET SDCs versus SP&R [9] objectives CFET SDCs. The white objects represent DRVs.	51
Figure 3.1:	Multi-Row CFET Standard Cell Synthesis Framework Overview	58
Figure 3.2:	Grid-Based placement, routing graph, and pin-shape of P-FET/N-FET using double row 4 RTs P-on-N CFET example.	59
Figure 3.3:	Concept of Multi-Row Dynamic Complementary Pin Allocation (MR-DCPA) for 4 RTs P-on-N CFET cell structure. p_1^P =P-FET Gate Pin. p_1^N =N-FET Gate Pin.	60
Figure 3.4:	An example of M0A/PC routing constraint I: the upper/lower M0A/PC layers in active FET region can only be used for routing by the same net of the corresponding FET pin. Here, the upper M0A of $n(p_2)$ region can only be used for routing $n(p_2)$	65
Figure 3.5:	Examples of M0A/PC routing constraint II: Lower MOA/PC is forbidden for inter-row routing when the upper MOA/PC connects to VDD/VSS.	66
Figure 3.6:	Examples of Parametric Design Rules for routing: (a) MAR, (b) EOL, and (c) VR. All the numbers are in grid.	67
Figure 3.7:	Examples of Parametric Design Rules for multi-patterning: (a) PRL and (b) SHR. All the numbers are in grid.	68
Figure 3.8:	Cell Statistics of M0 Core, M1 Core, and AES.	70
Figure 3.9:	An example of transferring the grid-based conditional design rules to the block-level.	71
Figure 3.10:	An example of XOR2x1 schematic netlist [8] and SDC layouts of (a) Triple-Row and (b) Optimum Cell Row cell structures.	76
Figure 3.11:	Layout of AND3x1 cell optimized generated by Pin Separation [6] and Edge-Based Pin Separation (EB-PS) [7] objectives. The RPA (Worst Case) considers the parallel pin-shape of adjacent cell as described in Figure 2.7.	77
Figure 3.12:	Block-level P&R Results of M0 Core, M1 Core, and AES designs of with “PS [6]” and “EB-PS” under MPO=2 and MPO=3 constraints.	79
Figure 3.13:	P&R design views of M0 core at 0.72 util. with the proposed EB-PS objective [7] 3.5T CFET SDCs versus PS objective [6] 3.5T CFET SDCs. The white objects represent DRVs.	81
Figure 3.14:	An example of XOR2x1 schematic netlist [8] and P-on-N and N-on-P CFET SDC Layout.	82
Figure 3.15:	Layouts of 4.5T and 3.5T CFET and Conv. DFFHQN with corrected design constraints. The metal length is the weighted sum of metal segments and vias.	83

Figure 3.16:	Block-Level P&R Results of M0 Core, M1 Core and AES of Conv. structure, which is generated using [9], and CFET with 4.5T and 3.5T cell height.	85
Figure 3.17:	Min. valid M0 Core and AES block-level areas with 300 #DRVs threshold versus (a) #BEOLs, (b) EOL, and (c) VR.	86
Figure 3.18:	An illustration of NAND2x2 without feasible solution when VR=1.5 in 3.5T P-on-N CFET structure.	89
Figure 3.19:	Block-level P&R Results of M0 Core and AES designs of P-on-N or N-on-P using EOL=3 and VR=1.5 design rules.	92
Figure 3.20:	Design technology co-optimization block-level placement-and-route results of M0 Core, M1 Core and AES. (a) Design Rule Relaxation (EOL=1 and VR=0) and (b) Design Rule Relaxation and increasing the top routing layer from M5 to M7.	94
Figure 3.21:	Cell and block-level area benefits by STCO and track reduction: 4.5T Conv. (black bar) 4.5T CFET (orange), 3.5T CFET (gold), 3.5T CFET with Design Rule Relaxation (DR Relax.) (blue), and DR Relax. plus adding #BEOLs in block-level for 3.5T CFET (purple).	95
Figure 3.22:	An example of XOR2x1 SDC layouts with Upper/Lower M0A/PC for inter-row routing.	98
Figure 3.23:	Layouts of design corrected DFFHQN layouts of (a) Single-Row 3.5T CFET, (b) Double-Row 2.5T CFET, and (c) Double-Row 2.5T CFET with M0A/PC routing. The Metal Length is weighted sum of metal segments and vias.	99
Figure 3.24:	Cell and block-level area benefits by track reduction and M0A/PC routing: (I) 3.5T CFET (black bar), (II) 2.5T CFET (orange), (III) 2.5T M0A/PC-R CFET (blue). (a) Cell Area of Representative 30 SDCs. (b) Block-level P&R results of M0 Core.	100
Figure 4.1:	The illustrations of (a) Traditional DTCO and STCO exploration flow. (b) The proposed DTCO and STCO sensitivity prediction framework. We use [7], and [10] for the automatic SDC synthesis here.	107
Figure 4.2:	An example of DTCO and STCO block-level area sensitivity of (a) #BEOLs and (b) design rules using 4.5T and 3.5T CFET SDC library sets of AES and M0 Core circuits. The number represents the block-level area difference as changing DTCO and STCO parameters from left to right.	108
Figure 4.3:	Overall flow of DTCO and STCO sensitivity prediction: (a) Training flow, and (b) Prediction model for DTCO and STCO exploration flow.	109
Figure 4.4:	An illustration of input features of the proposed DTCO and STCO sensitivity prediction model. Features of the j^{th} and j^{th} technologies are described in Table 4.1.	114
Figure 4.5:	Overview of the developed machine learning model. The model combines bootstrap aggregation and gradient boosting regression tree (GBRT) techniques.	115
Figure 4.6:	An example of generated DFFHQN SDC layouts with variations on three dimensions: (i) Cell Structure (CS), (ii) Design Rule, and (iii) Cell Height (CH).	118
Figure 4.7:	(a) #Fanouts distribution (b) Cell statistics of M0 Core, M1 Core, AES, MPEG, JPEF, and Darkriscv.	120
Figure 4.8:	An example of RPA counts of XOR2x1 in 3.5T CFET EOL=1 VR=0 standard cell library. Pin A and pin B are promoted to M2 for connecting internal FET terminals and satisfy the minimum pin opening constraints.	121
Figure 4.9:	(a) MLP, and (b) RBF neural network structures after Hyperband [11] search on #layers (i.e., 2 - 10), #neurons per layer (i.e., 25 - 500), and dropout rate (i.e., 0.0 - 0.5).	126

Figure 4.10:	Predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of (a) training set and (b) testing set, and (c) error distribution of testing set of the proposed model. The mean of MAE is 3.47×10^{-5} , with standard deviation of 0.0075 for testing set. Hence, 99.7% of predicted $\Delta A_{i,j}$ are within the 3-sigma range of ± 0.023	127
Figure 4.11:	Feature importance (Gain) of the proposed model and XGBoost_DTCO [12] for key feature study.	129
Figure 4.12:	Predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of new SDC library set technology prediction (i.e., orange points) and new BEOL pitch scaling technology prediction (i.e., green points).	131
Figure 4.13:	Minimum block-level area of M0 Core with various front side PDN grid scales (i.e., 32 CPPs, 48 CPPs, and 64 CPPs), and backside PDN architecture using M2 to M6 for signal routing.	134
Figure 4.14:	Predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of new PDN setting (32 CPPs) prediction (green points) and backside PDN prediction (orange points) with (a) Random Forest (implemented with sklearn), (b) XGBoost_DTCO [12], and (c) the proposed model.	135
Figure 4.15:	Accuracy improvement with various ratio of backside PDN data for model update.	138
Figure 5.1:	The illustration of transistor placement canvas for CFET standard cell architecture. There are diffusion sharing between PFET and NFET in the bottom cell row.	145
Figure 5.2:	The illustration of CFET standard cell synthesis using reinforcement learning technique. The standard cell environment is linked to SMT solver for standard cell synthesis.	146
Figure 5.3:	The illustration of the encoded placement grid. PFET/NFET is not allowed to be placed overlapped to another PFET/NFET. The grid coordinates of placed transistors are used to generate a set of relative position constraints in the SMT-based CFET standard cell synthesis framework.	148
Figure 5.4:	The illustration of the encoded placement grid. PFET/NFET is not allowed to be placed overlapped to another PFET/NFET. The grid coordinates of placed transistors are used to generate a set of relative position constraints in the SMT-based CFET standard cell synthesis framework.	148
Figure 5.5:	The reinforcement learning training plots of (a) running average rewards, (b) cell size cost, and (c) HPWL cost using deep Q learning [13].	150
Figure 5.6:	The XOR2x1 layouts from (a) CFET synthesis framework [7], and (b) the best reward of RL agent after the policy converges.	151

LIST OF TABLES

Table 1.1:	Notations for Standard Cell Synthesis	7
Table 2.1:	Experimental Statistics: ML= Metal Length. CPP= Contact Poly Pitch. Cell Width Red. = $\left(\frac{\text{Cell Width of reference} - \text{Cell Width of CFET}}{\text{Cell width of reference}}\right)$. ML Red. = $\left(\frac{\text{ML of reference} - \text{ML of CFET}}{\text{ML of reference}}\right)$	42
Table 2.2:	Experimental results of 30 CFET SDCs without pin-accessibility objective, PS [6] objective and Edge-Based PS (EB-PS) objective under MPO=2 and MPO=3 Constraints.	45
Table 2.3:	Experimental Results of CFET SDCs optimized for <i>MinTrack</i> (i.e., Expression (2.13)) and <i>MinLength</i> (i.e., (a) <i>CellSize</i> , (b) <i>PS</i> , (c) <i>M2 Length</i> , and (d) <i>TotalML</i>). Incr. = $\frac{(\text{MinTrack} - \text{MinLength})}{\text{MinLength}}$, Red. = $\frac{(\text{MinLength} - \text{MinTrack})}{\text{MinLength}}$	47
Table 2.4:	SDC Generation Split Cases for Routability Analysis	48
Table 2.5:	Pin Analysis QoR Report of <i>MinTrack</i> and <i>MinLength</i> from [14]	50
Table 3.1:	Experimental statistics of 2.5T CFET with Triple-Row (TR), Double-Row (DR), Single-Row (SR) [7] and Optimum Row (Opt. CR).	75
Table 3.2:	Experimental results of 30 4.5T and 3.5T CFET SDCs with PS [6] and Edge-Based PS (EB-PS) under MPO=2 and MPO=3 Constraints.	78
Table 3.3:	Experimental statistics of Conv. and CFET of 4.5T and 3.5T structures.	84
Table 3.4:	Design Technology Co-Optimization (DTCO) Experimental Results of 30 4.5T and 3.5T CFET SDCs with various design rule and stacking option using MPO=3 and EB-PS.	86
Table 3.5:	Difference of P-on-N and N-on-P 4.5T and 3.5T CFET SDCs with EOL=3.	90
Table 3.6:	Difference of P-on-N and N-on-P 4.5T CFET SDCs with VR=1.5.	91
Table 3.7:	M0 Core and AES Block Weighted Metric (i.e., $M2Track_d$, $M2ML_d$) of P-on-N (PN) and N-on-P (NP) CFET SDCs with EOL=3 and VR=1.5 design rule. Min. BA = Minimum Valid Block-Level Area (μm^2).	91
Table 3.8:	Experimental statistics of 3.5T CFET, 2.5T CFET, and 2.5T CFET with Upper/Lower M0A/PC routing (2.5T M0A/PC-R).	97
Table 3.9:	Block-level placement and route results of 3.5T CFET, 2.5T CFET, and 2.5T CFET M0A/PC-R.	101
Table 4.1:	Extracted Features Table	113
Table 4.2:	Synthesized block-level circuit table.	118
Table 4.3:	SDC feature values of 19 SDC library sets.	122
Table 4.4:	The breakdown of runtime in each design stage of an automated M0 core block-level P&R implementation using 2.5T CFET EOL=0 VR=0 library, and M2-M7 routing layers.	124
Table 4.5:	Hyperparameter exploration of machine learning algorithms table.	127
Table 4.6:	Prediction accuracy table.	128
Table 4.7:	The $\Delta A_{i,j}$ prediction results of new technologies using utilization model (Util.), random forest, XGBoost_DTCO [12], and the proposed model. MAE=Mean Absolute Error. Gradient ACC=Gradient Accuracy of $\Delta A_{i,j}$. Error Dist.=Error Distribution. Std. Dev.=Standard Deviation.	131

Table 4.8:	The $\Delta A_{i,j}$ prediction results of new front side PDN setting (i.e., 48 CPPs) and back-side PDN architecture using random forest, XGBoost_DTCO [12], and the proposed model. MAE=Mean Absolute Error. Gradient ACC=Gradient Accuracy of $\Delta A_{i,j}$. Error Dist.=Error Distribution. Std. Dev.=Standard Deviation.	135
Table 4.9:	The $\Delta A_{i,j}$ prediction results of selected synthesized block-level circuit. MAE=Mean Absolute Error. Gradient ACC=Gradient Accuracy of $\Delta A_{i,j}$. Error Dist.=Error Distribution. Std. Dev.=Standard Deviation.	137

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, Professor Chung-Kuan Cheng, and co-advisor, Professor Bill Lin for their encouragement, support, and guidance throughout my whole Ph.D. study, as well as the opportunities to collaborate with top-notch companies and researchers. My sincere thanks also go to my thesis committee members Professor Sicun Gao, Professor Patrick Mercier, and Professor Tajana Simunic Rosing for their time, encouragement, feedback and insightful comments.

I would like to thank all my lab colleges in VLSI lab for their active collaboration, help, and all the good memories. I would like to thank Dr. Yu-Min Lee, Dr. Hung-Ming Chen, and Dr. Yi-ming Li for their supports, and advice on the PhD study in the U.S when I was study in National Chiao Tung University in Taiwan. I would like to thank Dr. Victor Moroz, Dr. Deepak Sherlekar, Dr. Eric Chin, Dr. Lars Liebmann, Dr. Xiaoqing Xu, Dr. Joe Jiang, Dr. Jan Schneider, Dr. Dino Ruic, Dr. Cyrus Behroozi, and Dr. Raj B Apte for their guidance and practical training in my internships in CTO office, Synopsys and X, the moonshot factory. I learned many skills that helped me perform well in my research and work through the inspiring and challenging internships.

To whoever is reading this dissertation, I would like to thank you for your time, and I sincerely hope that you enjoy reading this dissertation!

Finally, I appreciate the consistent supports from my father and mother. I also appreciate all the supports from other family members, my girl friend, and my friends.

The material in this dissertation is based on the following publications.

Chapter 2, in part, contains reprints of Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, and Dongwon Park. "A routability-driven complimentary-FET (CFET) standard cell synthesis framework using SMT." In 2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2020; Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, Bill Lin, and Dongwon Park. "Complementary-FET

(CFET) standard cell synthesis framework for design and system technology co-optimization using SMT.” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2021. The dissertation author was the primary researcher and author of these papers.

Chapter 3, in part, contains reprints of Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, Bill Lin, and Dongwon Park. ”Complementary-FET (CFET) standard cell synthesis framework for design and system technology co-optimization using SMT.” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2021; Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, and Bill Lin. ”Multirow Complementary-FET (CFET) Standard Cell Synthesis Framework Using Satisfiability Modulo Theories (SMTs).” IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, 2021. The dissertation author was the primary researcher and author of these papers.

Chapter 4, in part, contains reprints of Chung-Kuan Cheng, Chia-Tung Ho, Chester Holtz, and Bill Lin. ”Design and System Technology Co-Optimization Sensitivity Prediction for VLSI Technology Development using Machine Learning.” In 2021 ACM/IEEE International Workshop on System Level Interconnect Prediction (SLIP), 2021; Chung-Kuan Cheng, Chia-Tung Ho, Chester Holtz, Daeyeal Lee, and Bill Lin. ”Machine Learning Prediction for Design and System Technology Co-Optimization Sensitivity Analysis.” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2022 (to appear). The dissertation author was the primary researcher and author of these papers.

My co-authors (Professor Chung-Kuan Cheng, Mr. Chester Holtz, Professor Bill Lin, Mr. Daeyeal Lee, and Mr. Dongwon Park, listed in alphabetical order) have all kindly approved the inclusion of the aforementioned publications in my dissertation.

VITA

2022	Ph.D., Electrical and Computer Engineering, University of California San Diego
2021-2022	PhD Resident, X, the moonshot factory, US
2019-2021	Technical Intern, Synopsys, Inc., US
2018	Senior Software Development Engineer, Synopsys, Inc., Taiwan
2017-2018	Senior Software Development Engineer, Mentor Graphics, Inc., Taiwan
2013-2016	Computer-Aided Design (CAD) Principal Engineer, Macronix International Co., LTD., Taiwan
2013	M.S., Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan
2011	B.S., Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan

(*) stands for the corresponding author. (+) stands for the co-first author.

- Chung-Kuan Cheng, **Chia-Tung Ho**, Daeyeal Lee, and Bill Lin. "Monolithic 3D Semiconductor Footprint Scaling Exploration based on VFET Standard Cell Layout Methodology, Design Flow, and EDA Platform." IEEE Access, 2022. (under Review) (Alphabetical Order)
- Chung-Kuan Cheng, **Chia-Tung Ho***, Chester Holtz, Daeyeal Lee, and Bill Lin. "Machine Learning Prediction for Design and System Technology Co-Optimization Sensitivity Analysis." IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2022. (to appear) (Alphabetical Order)
- Chung-Kuan Cheng, **Chia-Tung Ho**, and Chester Holtz. "Net Separation-Oriented Printed Circuit Board Placement via Margin Maximization." In 2022 27th Asia and South Pacific Design Automation Conference (ASP-DAC), 2022. (Alphabetical Order)

- Chung-Kuan Cheng, **Chia-Tung Ho**^{*}, Chester Holtz, and Bill Lin. "Design and System Technology Co-Optimization Sensitivity Prediction for VLSI Technology Development using Machine Learning." In 2021 ACM/IEEE International Workshop on System Level Interconnect Prediction (SLIP), pp. 8-15. IEEE, 2021. (Alphabetical Order)
- Chung-Kuan Cheng, **Chia-Tung Ho**^{*}, Daeyeal Lee, and Bill Lin. "Multirow Complementary-FET (CFET) Standard Cell Synthesis Framework Using Satisfiability Modulo Theories (SMTs)." IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, 2021. (Alphabetical Order)
- Lee, Daeyeal, **Chia-Tung Ho**, Ilgweon Kang, Sicun Gao, Bill Lin, and Chung-Kuan Cheng. "Many-tier vertical gate-all-around nanowire fet standard cell synthesis for advanced technology nodes." IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, 2021. (Alphabetical Order)
- Chung-Kuan Cheng, **Chia-Tung Ho**^{*}, Daeyeal Lee, Bill Lin, and Dongwon Park. "Complementary-FET (CFET) standard cell synthesis framework for design and system technology co-optimization using SMT." IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2021 (Alphabetical Order)
- Chung-Kuan Cheng, **Chia-Tung Ho**⁺, and Chester Holtz. "SPICE." Encyclopedia of RF and Microwave Engineering, 2021. (to appear)
- Daeyeal Lee, Dongwon Park, **Chia-Tung Ho**, Ilgweon Kang, Hayoung Kim, Sicun Gao, Bill Lin, and Chung-Kuan Cheng. "SP&R: SMT-based simultaneous Place-and-route for standard cell synthesis of advanced nodes." IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2020

- Chung-Kuan Cheng, **Chia-Tung Ho**^{*}, Daeyeal Lee, and Dongwon Park. "A routability-driven complimentary-FET (CFET) standard cell synthesis framework using SMT." In 2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2020. (Alphabetical Order)
- Chung-Kuan Cheng, **Chia-Tung Ho**⁺, Chao Jia, Xinyuan Wang, Zhiyu Zen, and Xin Zha. "A Parallel-in-Time Circuit Simulator for Power Delivery Networks with Nonlinear Load Models." In 2020 IEEE 29th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2020. (Alphabetical Order)
- Lars Liebmann, Daniel Chanemougame, Peter Churchill, Jonathan Cobb, **Chia-Tung Ho**, Victor Moroz, and Jeffrey Smith. "DTCO acceleration to fight scaling stagnation." In Design-Process-Technology Co-optimization for Manufacturability XIV, vol. 11328, p. 113280C. International Society for Optics and Photonics, 2020.
- **Chia-Tung Ho**^{*}, and Andrew B. Kahng. "IncPIRD: Fast learning-based prediction of incremental IR drop." In 2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2019. (Alphabetical Order)
- Yu-Min Lee, and **Chia-Tung Ho**⁺. "Intrasim: Incremental transient simulation of power grids." IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2017.
- Jiaxing Song, Yu-Min Lee, and **Chia-Tung Ho**. "ThermPL: Thermal-aware placement based on thermal contribution and locality." In 2016 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), 2016.
- **Chia-Tung Ho**^{*}, Yu-Min Lee, Shu-Han Wei, and Liang-Chia Cheng. "Incremental transient simulation of power grid." In Proceedings of the 2014 on International symposium on physical design, 2014.

- Shu-Han Wei, Yu-Min Lee, **Chia-Tung Ho**, Chih-Ting Sun, and Liang-Chia Cheng. "Power delivery network design for wiring and TSV resource minimization in TSV-based 3-D ICs." In 2013 International Symposium on VLSI Design, Automation, and Test (VLSI-DAT), 2013.
- **Chia-Tung Ho***, and Yu-Min Lee, "Efficient Transient Incremental Analysis of On-Chip Power Grid," in Asia-Pacific Radio Science Conference (AP-RASC), 2013.

ABSTRACT OF THE DISSERTATION

**Novel Computer Aided Design (CAD) Methodology
for Emerging Technologies to Fight the Stagnation of Moore's Law**

by

Chia-Tung Ho

Doctor of Philosophy in Electrical and Computer Engineering

University of California San Diego, 2022

Professor Chung-Kuan Cheng, Chair
Professor Bill Lin, Co-Chair

As the Very-Large-Scale Integration (VLSI) technology advances beyond $7nm$, several challenges arise when designers struggle to fulfill the Power-Performance-Area-Cost (PPAC) requirements of modern complex design and continue the scaling trend in the Post Moore's Law era. In order to retain the trend of Moore's Law, Design Technology Co-Optimization (DTCO) and System Technology Co-Optimization (STCO) are introduced together to continue scaling beyond $7nm$ using pitch scaling, patterning, and novel 3D cell structures (i.e., Complementary-FET (CFET)). However, standard cell synthesis for novel 3D cell architectures (i.e., CFET and VFET) demand holistic considerations to maximize the area benefit

of scaling at the block-level due to the extremely limited routability that comes from the stacked structure and reduced cell height. Furthermore, numerous DTCO and STCO iterations are needed to continue block-level area scaling with considerations of physical layout factors: (i) various standard cell (SDC) architectures (i.e., cell heights, Conventional FET, CFET, etc.), (ii) design rules, (iii) back end of line (BEOL) settings, and (iv) power delivery network (PDN) configurations. The growing turnaround time (TAT) among standard cell design, design rule optimization, and block-level area evaluation becomes one of the major bottlenecks in DTCO and STCO explorations. In this dissertation, We aim to develop novel Computer-Aided Design (CAD) methodologies to resolve the design-technology crisis as scaling beyond $7nm$.

We propose an SMT (Satisfiability Modulo theories)-based framework to automate CFET SDC synthesis through a novel Multi-row Dynamic Complimentary Pin Allocation Scheme for standard cell height reduction from 4.5T to the extreme 2.5T for DTCO and STCO explorations on emerging 3D cell architectures. Moreover, we propose pin access and routing resource related objectives/constraints for routability to maximize the block-level area benefits. We demonstrate our CFET SDC synthesis framework with extensive studies on various cell architectures (i.e., cell height, multi-row cells, 3D stacking options, etc.), ground design rules (i.e., tip-to-tip spacing, via rule, and minimum area rule), and BEOL configurations for DTCO and STCO explorations. Moreover, we develop a machine learning modeling approach to improve the performance of holistic DTCO and STCO explorations for block-level metrics (i.e., block-level area), which greatly reduce the TAT among standard cell design, design rule optimization, and block-level area evaluation.

We organized this dissertation as follows. Firstly, we introduce the novel dynamic complementary pin allocation scheme, and pin accessibility constraints/objectives for routability-driven CFET standard cell synthesis. In addition, we present the improvement of proposed pin accessibility con-

straints/objectives in the block-level. Next, we present extensive studies on various cell heights, multi-row cell structures, 3D stacking options, ground design rules, and number of BEOLs in the DTCO and STCO explorations. Finally, we introduce the developed machine learning modeling approach to predict DTCO and STCO sensitivity.

Chapter 1

Introduction and Preliminaries

With the relentless scaling of Very-Large-Scale-Integration (VLSI) technology, the geometric pitch scaling starts to slow down due to process and lithography limitations in sub- $16nm$ as shown in Figure 1.1. The conventional (Conv.) standard cell (SDC) layout scaling is widely used to continue scaling from 9T at $20nm$ node in 2012 to 6T at $7nm$ node in 2019, but it is limited by the saturation of contacted poly pitch (CPP), lateral P-N separation, routing congestion, and performance requirements when the technology node advancing to sub $7nm$ as shown in Figure 1.2. In order to catch up the scaling trend, design technology co-optimization (DTCO) and system technology co-optimization (STCO) are introduced together to continue scaling beyond $7nm$ using pitch scaling, patterning, and novel 3D cell structures (i.e., Complementary-FET (CFET)). However, standard cell synthesis for novel 3D cell architectures (i.e., CFET and VFET) demand holistic considerations to maximize the cell and block-level area benefits due to limited in-cell routing tracks and the extremely limited routability that comes from the stacked structure. In addition, numerous DTCO and STCO iterations are needed to continue block-level area scaling with considerations of physical layout factors: (i) various standard cell (SDC) architectures (i.e., cell heights, multi-row structure, Conventional FET, CFET, etc.), (ii) design rules, (iii) back end of

line (BEOL) settings, and (iv) power delivery network (PDN) configurations. The growing turnaround time (TAT) among standard cell design, design rule optimization, and block-level area evaluation becomes one of the major bottlenecks in DTCO and STCO explorations.

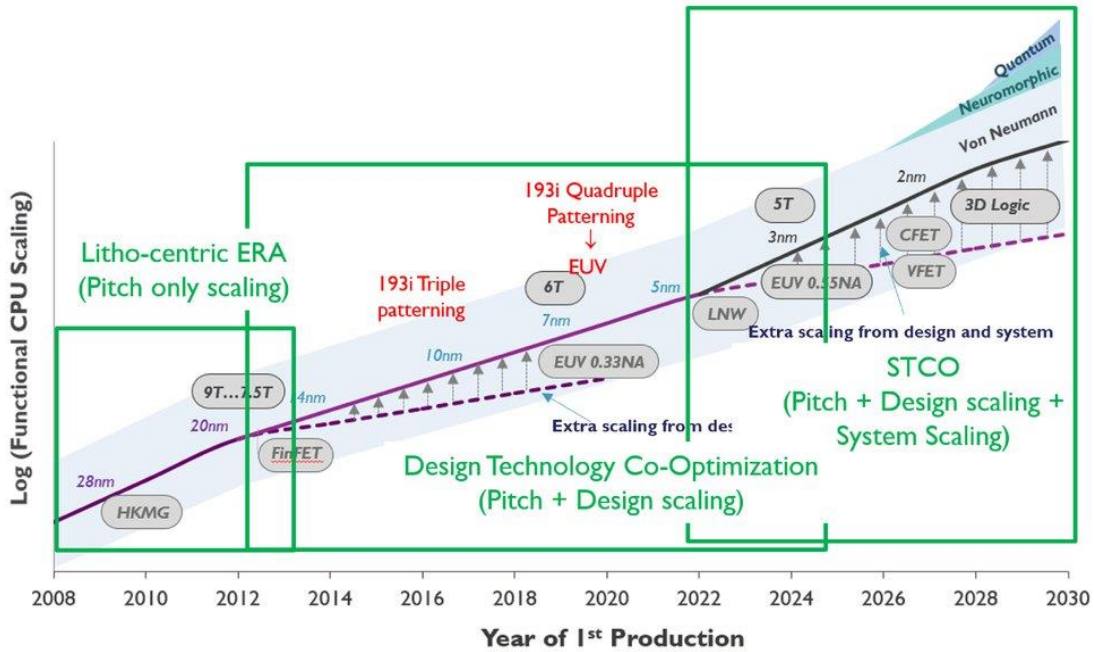


Figure 1.1: Technology node scaling roadmap [1]

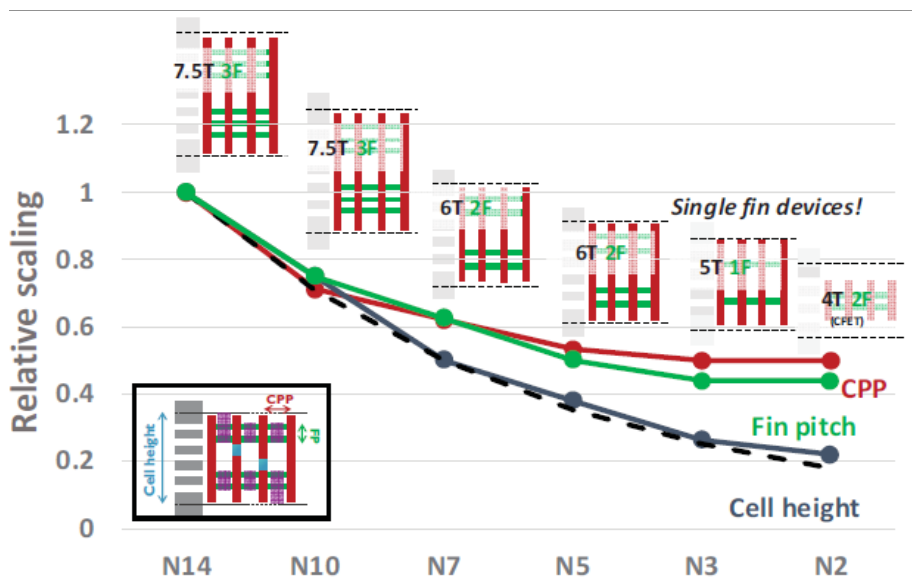


Figure 1.2: Dimensional scaling of standard cell in sub-7nm [2]

This dissertation proposes novel Computer-Aided Design methodologies to enable early design-technology explorations of 3D cell architectures and improve the performance of developing emerging technologies to fight the stagnation of Moore’s Law. In Chapter 2, we introduce the novel simultaneously place-and-route (P&R) CFET standard cell synthesis framework to enable early block-level PPAC evaluation in sub- $7nm$. In Chapter 3, we not only show the capability of the developed CFET standard cell synthesis framework on multi-row and various track heights but also study the impact of cell architectures, ground design rules, and number of back-end-of-lines (BEOLs) on the block-level area in the early design technology co-optimization (DTCO) and system technology co-optimization (STCO) explorations. In Chapter 4, we introduce the developed machine learning modeling approach to greatly reduce the turn-around-time (TAT) in the DTCO and STCO explorations when developing technologies.

The rest of this chapter is organized as follows. Section 1.1 introduces CFET technology. Section 1.2 introduces satisfiability modulo theories (SMT) for standard cell layout generation. Section 1.3 introduces the fundamental FET placement, in-cell routing, and representative design rule constraints for our standard synthesis framework. Section 1.4 introduces the DTCO and STCO exploration flow for technology development.

1.1 Complementary-FET (CFET) Technology

Complementary-FET (CFET) technology, which stacks the P-FET on N-FET or vice versa, can release in-cell routing congestion of P-N connection such that SDC designers can continue cell size reduction in sub- $7nm$. Figure 1.3 shows an illustration of a Conv. cell structure as well as a CFET cell structure that stacks the P-FET on N-FET. Compared to the Conv. cell architecture, the shared or split Gate and Source/Drain (G/S/D) structure provides flexible local interconnect connections. If the G/S/D of P-FET and N-FET share the same net connection, the G/S/D can be merged and connected to the $M0$. On the

contrary, the $G/S/D$ are split and $M0$ drops tall and short vias to connect P-FET and N-FET, respectively. CFET standard cell synthesis demands holistic considerations to maximize the area benefit of scaling at the block-level due to the extremely limited routability that comes from the stacked structure and reduced cell height.

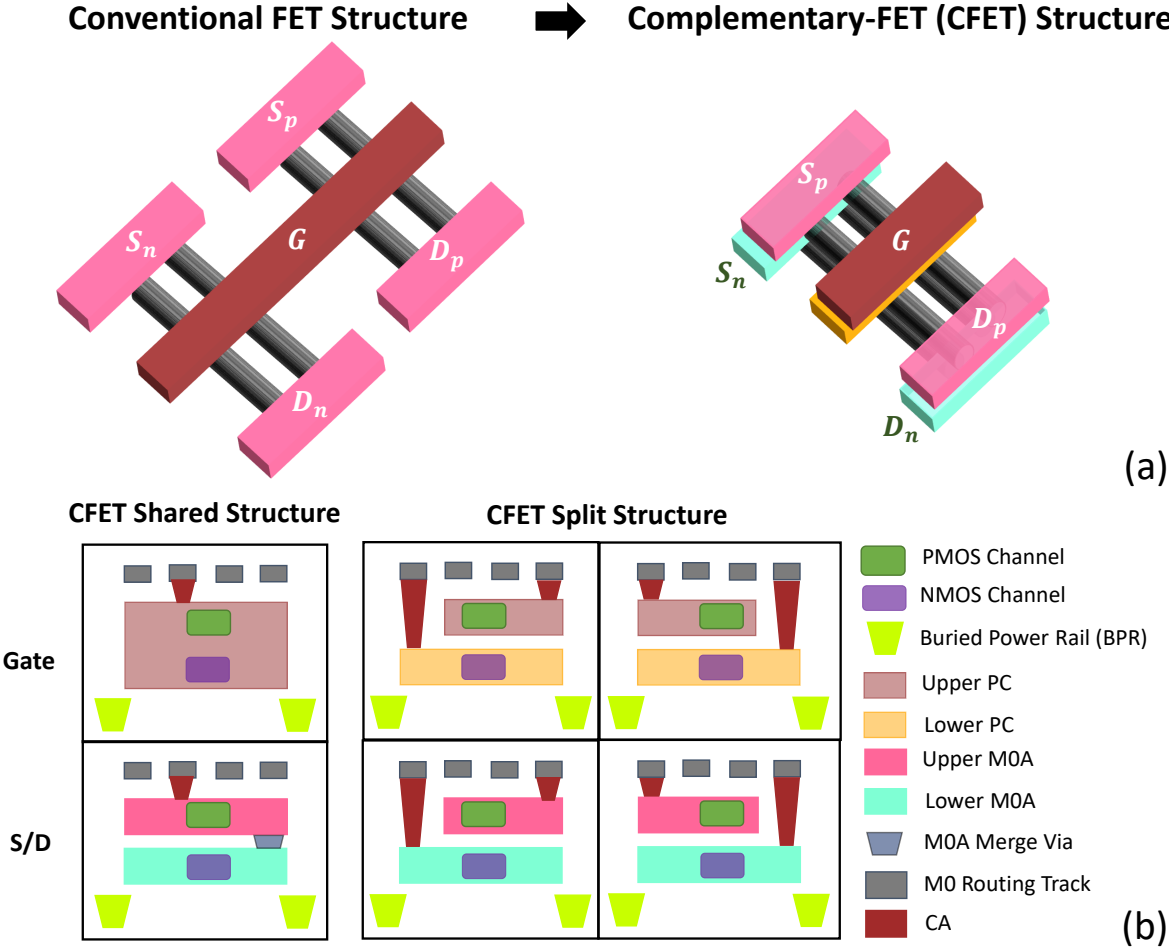


Figure 1.3: An illustration of Conventional and Complementary FET (CFET) structure (Top Row). CFET shared and split Gate, Source and Drain structure (Bottom Row) [3–5].

To fully leverage the CFET standard cell share-and-split $G/S/D$ structure for area scaling, we propose a novel routability driven CFET standard cell synthesis methodology to maximize the area benefits in both cell-level and block-level. Our methodology employs a CFET cell architecture of [3, 5] to generate CFET standard cell layout. The proposed CFET standard cell synthesis methodology is introduced in

Section 2.

1.2 Satisfiability Modulo Theories (SMT) for Standard Cell Layout

Satisfiability Modulo Theories (SMT) [15] includes Boolean Satisfiability (SAT), theories of non-Boolean variables (e.g., integer, bit-vector, etc.), and predicate symbols, which empowers us to represent more expressive language. Recently, several state-of-the-art SMT solvers with the optimization methodology (so-called “OMT”) are released [16, 17], which allow users to pose satisfying assignments, and get optimal assignments with respect to objective functions. In [16], it provides a portfolio of methodologies to solve linear optimization problems with SMT formulas, MaxSMT, and their combinations. The objective functions can be optimized using either Pareto fronts, lexicographical technique, or optimizing each objective independently.

In this dissertation, we formulate an integrated constraint satisfaction problem (CSP) for automating CFET standard cell synthesis and utilize SMT to implement our problem because SMT-based methodologies support a much richer modeling language than SAT or ILP formulas. For example, logical constraints (e.g., “if-then-else” for the “Either-Or” constraint) are able to be easily implemented by “ITE” keyword, meanwhile ILP formula needs to use big M method, which requires additional auxiliary variables, for logical constraints. Furthermore, an state-of-the-art SMT solver [16] include built-in Boolean cardinality functions such as at-most k (i.e., “AM k ”) and at-least k (i.e., “AL k ”).

We introduce the fundamental placement and in-cell routing formulations for our CFET standard cell synthesis in the following Section 1.3.1 and Section 1.3.2. We will introduce novel dynamic complementary pin allocation scheme, which enables simultaneous place-and-route for share-and-split G/S/D in CFET architecture in Chapter 2.

1.3 Preliminaries of CFET Standard Cell (SDC) Synthesis Framework

We introduce fundamental FET placement, in-cell routing, and representative design rule constraints for our CFET standard cell synthesis framework. The novel dynamic complementary pin allocation scheme, which enable simultaneous place-and-route of share-and-split pin shapes in CFET standard cell layout, will be introduced in Chapter 2. Our framework formulates an integrated constraint satisfaction problem (CSP) for automating CFET SDC layout which strictly satisfies transistor placement, in-cell routing, and conditional design rules on a grid-based placement and routing graphs. Table 1.1 shows the notations for the FET placement, in-cell routing, and design rule formulations in this dissertation.

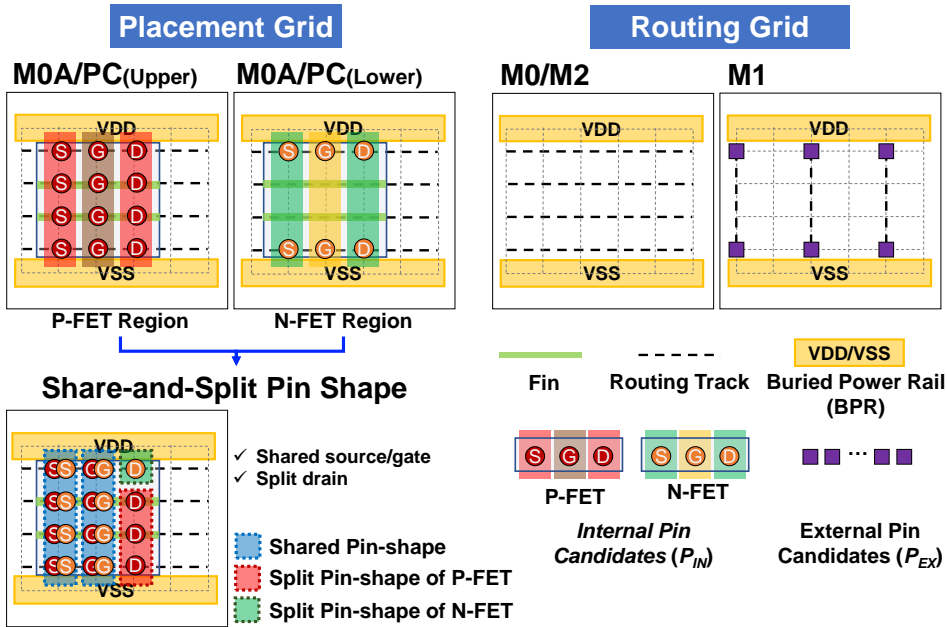


Figure 1.4: Grid-Based placement and routing graph

Figure 1.4 shows the grid-based placement and routing graph (i.e., Upper/Lower M0A/PC, M0, M1, and M2) using 4 routing tracks (RTs) P-on-N CFET example. The routing grid consists of 4 RTs with buried power rails (i.e., 4.5 Track Height) and each layer is defined as unidirectional edges. The P-FET and N-FET regions are stacked up on the upper and the lower M0A/PC layers, respectively. Therefore,

the access to the $M0$ layer from each pin on the N-FET region (i.e. lower $M0A/PC$ layer) is restricted to the top or bottom horizontal routing track unless each source/gate/drain pin in P-FET and N-FET that are overlapped on the same vertical track is shared [3, 5] as shown in Figure 1.3. As a result, there exist three kinds of pin shapes according to the sharing status of each pin in stacked FETs as depicted in Figure 1.4.

Table 1.1: Notations for Standard Cell Synthesis

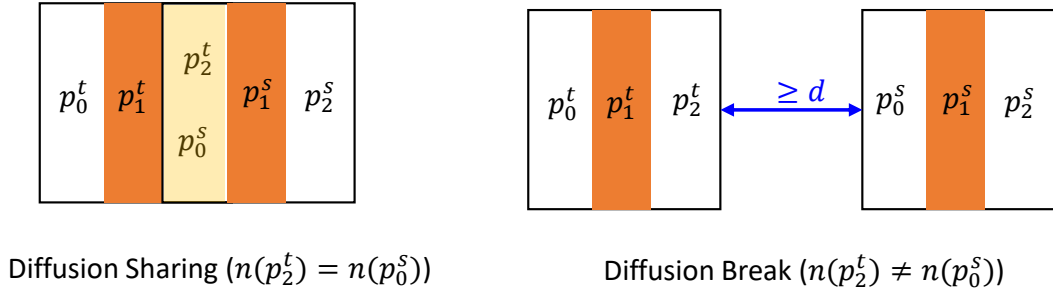
Term	Description
h	Number of M0/M2 horizontal routing tracks.
T	Set of FETs.
x_t^P/x_t^N	X coordinate of lower-left corner of t^{th} P-FET/N-FET.
w_t^P/w_t^N	Width of t^{th} P-FET/N-FET.
P_{EX}	Standard Cell I/O pins.
P_{IN}	Internal pin for FETs.
$p_{i,t}^P/p_{i,t}^N$	i^{th} pin of t^{th} P-FET/N-FET.
$x(p)$	X coordinate of pin p ; (i.e., $x(p_{i,t}^P)=x_t^P+i$).
$n(p)$	Net information of pin p .
d_{int}	Pin interference distance based on design rules.
$d_{int}(k)$	Set of M1 tracks in the pin interference distance of k^{th} M1 track.
$G(V, E)$	Three-dimensional (3D) routing graph.
V	Set of vertices in G .
$V(V_i)$	Set of vertices in (i^{th} metal layer of) the routing graph G
v or $v_{x,y,l}$	A vertex with the coordinate (x, y, l) .
$a(v)$	Set of adjacent vertices of v in G .
$e_{v,u}$	An edge between v and u , $u \in a(v)$.
$w_{v,u}$	Weighted cost for metal segment on $e_{v,u}$.
n	n^{th} multi-pin net.
m	m^{th} sink of n .
v^n	0-1 indicator if v is used for n .
$e_{v,u}^n$	0-1 indicator if $e_{v,u}$ is used for n .
$f_m^n(v, u)$	0-1 indicator if $e_{v,u}$ is used for commodity f_m^n .
$m_{v,u}$	0-1 indicator if there is a metal segment on $e_{v,u}$.
$g_{d,v}$	0-1 indicator if v forms d-side EOL of a metal segment.
E_k^l	Set of l^{th} layer edges in the k^{th} track.
$S(p)$	Column-Based Pin Separation Space of pin p .
$SC(p)$	Column-Based Pin Separation cost of pin p .
$EC(p)$	Edge-Based Pin Separation cost of pin p .

⁰The symbol d is L (Left), R (Right), F (Front), B (Back), U (Up), D (Down), or a combination of these directions, e.g., FL means FrontLeft.

1.3.1 FET Placement Formulations

We consider FET flipping, diffusion sharing and diffusion break in the transistor placement. Diffusion sharing is a common placement technique when the net information are the same between source and drain pins of adjacent FETs. Diffusion break refers to the minimum space d between distinct diffusion regions when they are not shared due to the different net information. We adopt single diffusion break based on the CFET technology [3, 5]. Figure 1.5 (a) shows the illustration of diffusion sharing and diffusion break. FET can be flipped horizontally to leverage more diffusion sharing for smaller cell size.

(a) Diffusion Sharing and Diffusion Break



(b) Relative Positioning Constraints (RPC) between two FETs

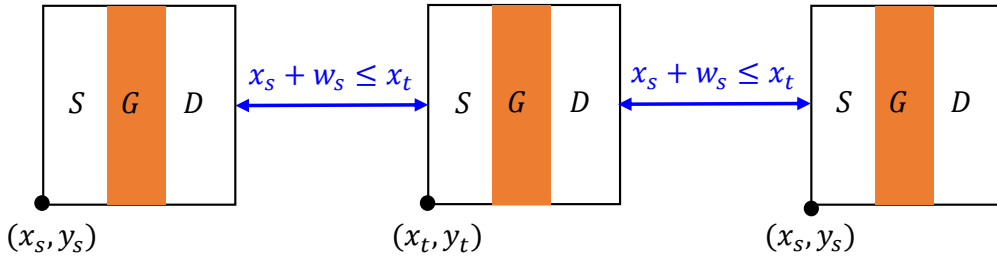


Figure 1.5: An illustration of diffusion sharing/break and relative positioning constraints. (a) Diffusion sharing and diffusion break illustration. (b) Relative positioning constraints between two FETs.

Relative Positioning Constraint (RPC). We adopt the conventional floorplanning design approach (i.e., Relative Positioning Constraint (RPC)) for the FET placement problem [18]. All transistor positions

can be represented by two RPCs as shown in Figure 1.5 (b). In CFET cell structure, the P-FET and N-FET are placed on upper and lower placement grids, respectively, as shown in Figure 1.4. As a result, we construct relatively position constraints with diffusion sharing for P-FET and N-FET individually as shown in Algorithm 1. This SMT geometric constraints ensure only one of the four cases is enabled at once and determines the position and the flip status of FETs in Algorithm 1.

Algorithm 1 Relative Positioning Constraint with Diffusion Sharing (FETs t, s)

Input: t, s : a pair of N-FETs/P-FETs, d_s : distance of a single diffusion break

// Set SMT Constraint

```

1: if  $t$  is on the right side of  $s$  without diffusion sharing then
2:    $x_t \geq x_s + w_s + d_s$ ;
3: else if  $t$  is on the right side of  $s$  with diffusion sharing then
4:    $x_t = x_s + w_s$ ;
5: else if  $t$  is on the left side of  $s$  without diffusion sharing then
6:    $x_t + w_t + d_s \leq x_s$ ;
7: else if  $t$  is on the left side of  $s$  with diffusion sharing then
8:    $x_t + w_t = x_s$ ;
9: else
10:  Unsatisfiable Condition;
11: end if

```

1.3.2 In-Cell Routing Formulations

We adopt conditional design rule-aware multi-commodity network flow theory to formulate the in-cell routing problem as described in [10] [19]. Specifically, to reduce the search space of the routing formulation, the refined constraints for *commodity flow conservation (CFC)* and *vertex exclusiveness (VE)* in uni-directional edges [10] are implemented in our framework. We separate the routing formulation into two parts, *flow formulation* and *conditional design rules* as shown in Figure 1.6. Then, we introduce the flow formulation, which finds the routing path between the source and the sink for each commodity, and Supernode technique in this section. The conditional design rules will be introduced in the Section 1.3.3. The flow formulation secures the routing path between the source and the sink for each commodity without heuristic modeling.

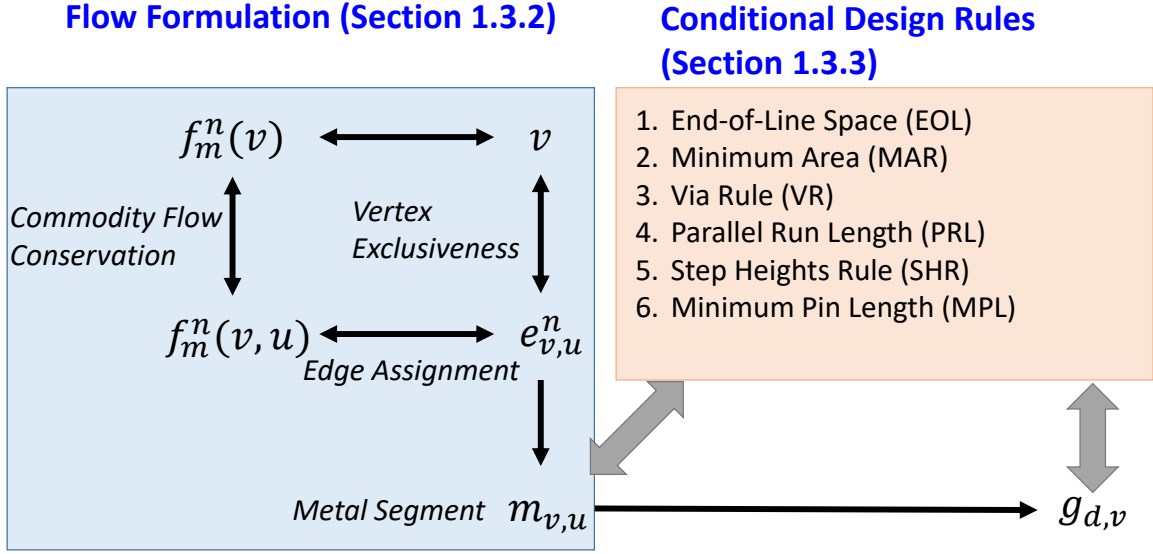


Figure 1.6: In-Cell routing flow formulation. (a) Flow Formulation (Section 1.3.2). (b) Conditional Design Rules (Section 1.3.3).

Commodity Flow Conservation (CFC). Expression (1.1) represents the modified CFC constraint regardless of the flow direction. The number of activated commodity-flow indicators $f_m^n(v, u)$ between a certain vertex v and its adjacent vertices $a(v)$ is 1 (Exactly-1) in case of source s^n or sink d_m^n , and is 0 or 2 in the other cases.

$$\sum_{u \in a(v)} f_m^n(u, v) = \begin{cases} 1, & \text{if } v = s^n, d_m^n \\ 2x, x = \{0, 1\}, & \text{otherwise} \end{cases} \quad (1.1)$$

$$\forall v \in V, \forall n \in N, \forall d_m^n \in D^n$$

Supernode. we group multiple feasible pin locations using Supernode technique as [19, 20]. Figure 1.7 illustrates supernodes for external pins (P_{EX}) on M1 and supernodes for internal pins (P_{IN}) on M0A/PC. The supernode of a pin is connected to vertices covering the pin (red and orange circles on Pin 1 and Pin 2, respectively). A supernode for a pin connected to vertices at the potential standard cell I/O pin candidates

of G (i.e., outer pin) is depicted in Figure 1.7 (purple squares). Each commodity consists of one source and one sink as supernode, respectively. In particular, our supernode for outer pins (i.e., Super Outer Node for P_{EX}) abstracts all supernodes of outer pins so that our framework has only one set of variables for the supernode (i.e, shared supernode). Thus, the complexity of exclusiveness-related constraint is reduced compared to [19]. Note that we denote the pin on M0A/PC layer by Super Inner Node S_i to distinguish from S_o .

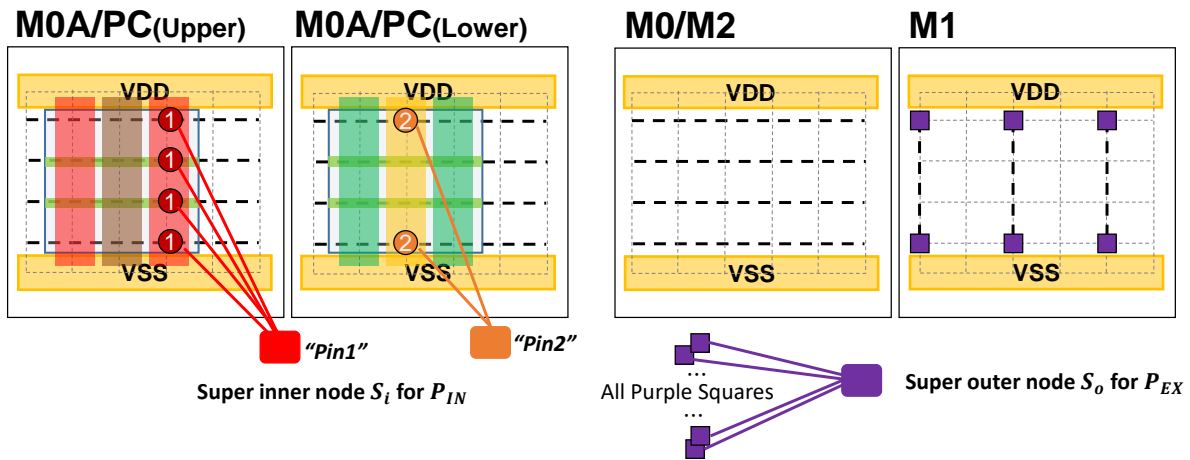


Figure 1.7: Supernodes. PIN1 and PIN2 respectively cover four and two vertices on Upper and lower M0A/PC layers (i.e., inner pin S_i for P_{IN}). Outer pins (i.e., P_{EX}) are connected to vertices on M1 of G through Super Outer Node (S_o).

Vertex Exclusiveness. Expression (1.2) ensures that there are no intersecting routing path of different nets on any vertices except P_{EX} . The supernode of external pins (P_{EX}) should be shared as many as the number of P_{EX} . When $v = P_{IN}$, only one edge indicator must be used. Otherwise, we allow multiple uses of edges against vertex v for a certain net.

$$\begin{cases}
\mathbf{AL1}(E_{IN}(v)) \wedge \mathbf{AM1}(E_{IN}(v)), & \text{if } v = P_{IN} \\
\mathbf{ALk}(E_{EX}(v)) \wedge \mathbf{AMk}(E_{EX}(v)), k = |P_{EX}|, & \text{else if } v = P_{EX} \\
\mathbf{AM1}(\{\bigvee_{u \in a(v)} e_{v,u}^n \mid n \in N\}), & \text{otherwise}
\end{cases}$$

$$E_{IN}(v) = \{e_{v,u}^n \mid u \in a(v)\}, E_{EX}(v) = \{e_{v,u}^n \mid n \in N, u \in a(v)\},$$

$$\forall n \in N, \forall v \in V \tag{1.2}$$

Edge Assignment. To obtain a Steiner tree of net n , we determine the edge indicator $e_{v,u}^n$ by overlapping each commodity flow belonging to the net n . As shown in Expression (1.3), $e_{v,u}^n$ can be either 0 or 1 even the flow indicator $f_m^n(v, u) = 0$, ensuring the multi-commodity flow of net n .

$$\begin{cases}
\text{if } f_m^n(v, u) = 1, e_{v,u}^n = 1 \\
e_{v,u}^n - f_m^n(v, u) \geq 0, \forall n \in N, e_{v,u}^n \in E
\end{cases} \tag{1.3}$$

Metal Segment. We adopt the metal indicator $m_{v,u}$ to determine whether a metal segment is on $e_{v,u}$ or not. Expression (1.4) shows that $m_{v,u}$ is 1 if one of the nets use the $e_{v,u}$.

$$m_{u,v} = \bigvee_{\forall n \in N} e_{u,v}^n, e_{u,v} \in E \tag{1.4}$$

1.3.3 Conditional Design Rule Constraints

The conditional design rule constraints ensure that the routing paths are design-rule violation-free. We use geometric variable $g_{d,v}$ which is determined by the end of line of a metal segment as shown

in the Expression (1.5).

$$\begin{aligned}
 g_{L,v} &= \neg m_{v_L,v} \wedge m_{v,v_R}; g_{R,v} = m_{v_L,v} \wedge \neg m_{v,v_R} \\
 g_{F,v} &= \neg m_{v_F,v} \wedge m_{v,v_B}; g_{B,v} = m_{v_F,v} \wedge \neg m_{v,v_B}
 \end{aligned}
 \tag{1.5}$$

Figure 1.8 shows an example for determining $g_{d,v}$. In Figure 1.8 (a), Vertex v is on the right side of the end of the metal segment. The $g_{R,v}$ and $g_{L,v}$ are 1 and 0, respectively. For the vertical metal segment example in Figure 1.8 (b), the $g_{B,v}$ and $g_{F,v}$ are 1 and 0, respectively. We use representative conditional design rules of [9, 21] for EUV and multi-pattern technologies as shown in Figure 1.6. For routing, we consider minimum area rule (MAR), via rule (VR), and end-of-line spacing rule (EOL). For multi-pattern technologies (i.e., M0 and M2 layers), we use parallel running length (PRL) and step heights rule (SHR) for SADP (Self-aligned double patterning) mask [22]. In this dissertation, all design rules are parameterized by the grid.

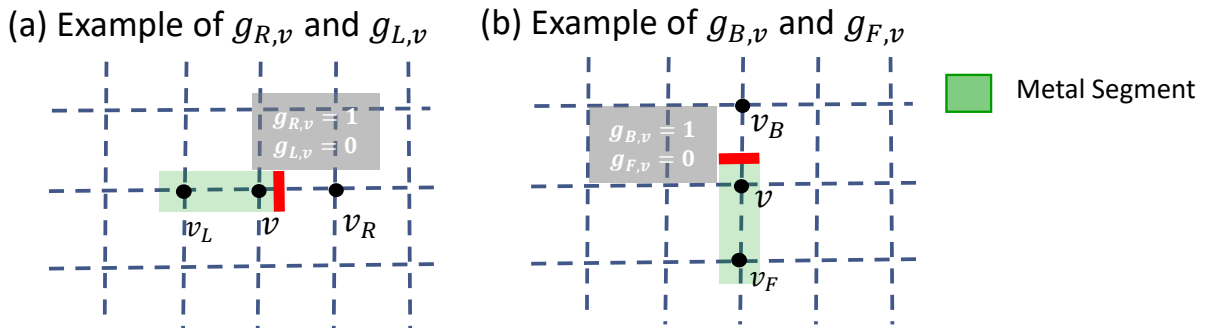


Figure 1.8: An example of geometric variable $g_{d,v}$.

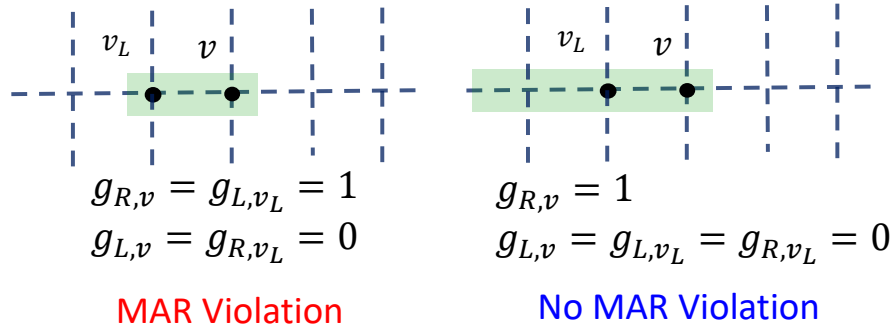


Figure 1.9: An example of minimum area rule (MAR). (a) MAR violation. (b) No MAR violation.

Minimum Area Rule (MAR).

Minimum area rule ensures that each disjoint metal segment is larger than the minimum manufacturable size. Constraint (1.6) ensures that a metal segment must cover at least three vertices for MAR. Figure 1.9 shows the MAR violation case and the metal segment satisfies MAR based on constraint (1.6).

$$g_{L,v} + g_{R,v} + g_{R,v_L} + g_{L,v_L} \leq 1 \quad (1.6)$$

End-of-Line Spacing Rule (EOL).

End-of-Line Spacing Rule describes that the distance between each EOL of two metal segments that are coming from opposite directions should be greater than a minimum spacing value. Constraint (1.7) and Figure 1.10 describe the right-directional EOL when we assume that the minimum distance between any of two opposite EOLs must be larger than L1 norm (i.e., Manhattan distance) of two vertices. The left-, front-, and back-directional EOLs are similarly derived.

$$g_{R,v} + g_{L,v_{FR}} \leq 1$$

$$g_{R,v} + g_{L,v_R} + g_{L,v_{RR}} \leq 1$$

$$g_{R,v} + g_{L,v_{BR}} \leq 1 \tag{1.7}$$

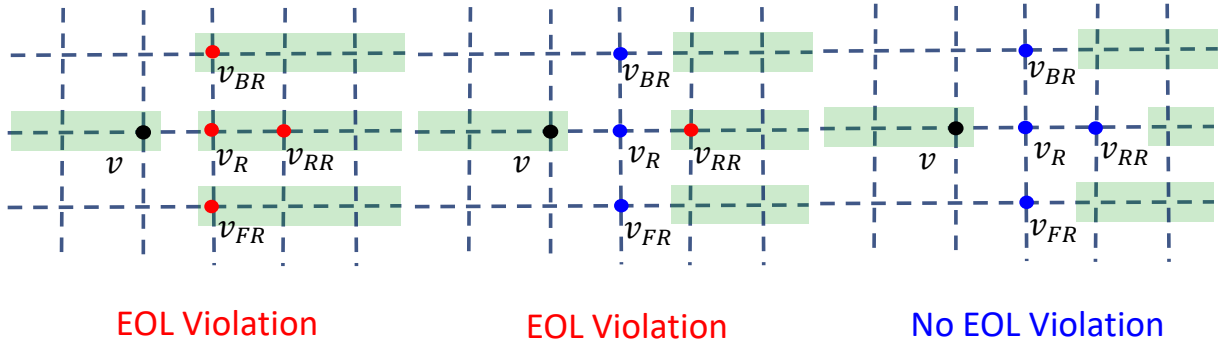


Figure 1.10: An example of right-directional end-of-line spacing rule (EOL).

Via Rule (VR).

Constraint (1.8) represents via rule (VR) related to restriction rules of inter-layer via (i.e., via) locations when the minimum distance (i.e., L2 Norm) between via is 1 grid as shown in Figure 1.11 (i.e., via-to-via spacing rule), which allows diagonal via and disallows adjacent via.

$$m_{v,v_U} + m_{v_R,v_{UR}} + m_{v_B,v_{UB}} \leq 1 \tag{1.8}$$

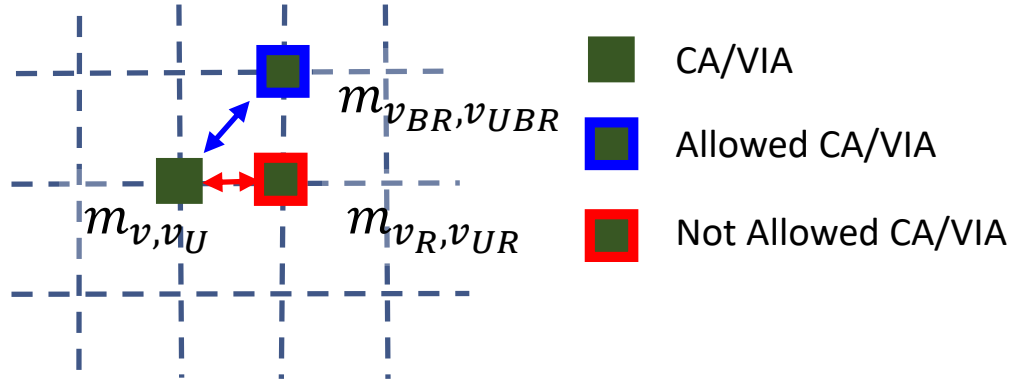


Figure 1.11: An example of via rule (VR) when the minimum distance between vias is 1 grid.

Parallel Running Length Rule (PRL).

PRL rule is one of the important rules to avoid “single-point-contact” in manufacturing SADP mask [22]. Figure 1.12 and Constraint (1.9) represent PRL rule for the right-directional of v and the corresponding formulation when run length is 2 grids. The left-, front-, and back-directional PRLs are similarly derived.

$$\begin{aligned}
 g_{R,v} + g_{L,v_B} + g_{L,v_{BL}} &\leq 1 \\
 g_{R,v} + g_{L,v_F} + g_{L,v_{FL}} &\leq 1
 \end{aligned}
 \tag{1.9}$$

Step Heights Rule (SHR).

SHR is a design rule to avoid “the small step” in manufacturing SADP mask [22]. Figure 1.13 and Constraint (1.10) represents the SHR for the right-directional of v and the corresponding formulation

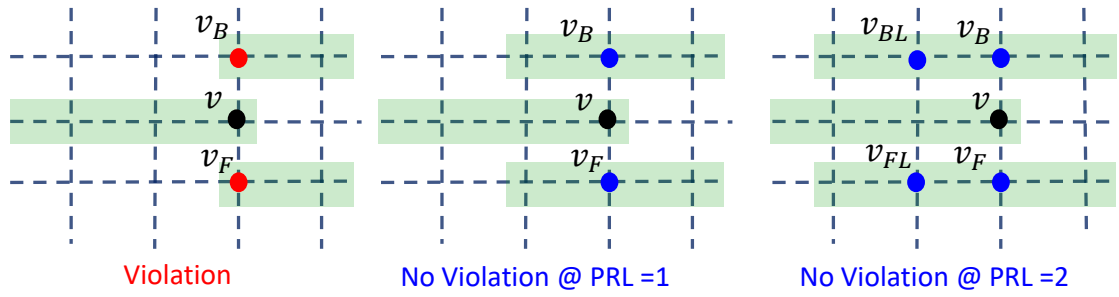


Figure 1.12: An example parallel running length (PRL) rule.

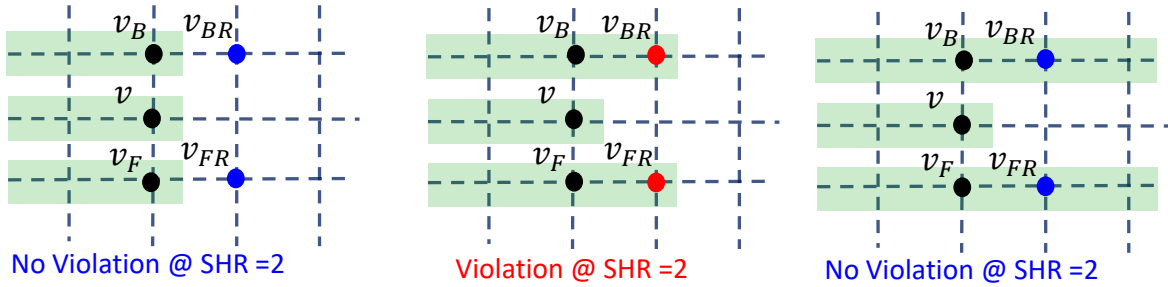


Figure 1.13: An example step height rule (SHR).

when step height is 2. The left-, front-, and back-directional SHRs are similarly derived.

$$g_{R,v} + g_{R,v_{BR}} \leq 1$$

$$g_{R,v} + g_{L,v_{FR}} \leq 1 \tag{1.10}$$

1.4 Design Technology Co-Optimization (DTCO) and System Technology Co-Optimization (STCO) Flow

Design Technology Co-Optimization (DTCO) and System Technology Co-Optimization (STCO) are essential to continue the scaling beyond $7nm$ as shown in Figure 1.1. DTCO/STCO exploration for emerging technology starts with defining technology parameters (i.e., SDC architecture, design rules, BEOL settings). The technology file can be generated using design rules of BEOL and BEOL settings. The SDC libraries can be generated by manual designers or automatic synthesis program. The timing and power of standard cells are characterized using circuit simulators (i.e., HSPICE) and the physical pin shape of standard cells are extracted to construct SDC .lef for block-level P&R implementation. Then, the block-level PPA metrics are extracted to adjust the technology parameters to optimize the block-level PPA.

The DTCO and STCO explorations usually takes weeks or months to optimize the technology for block-level PPA. In this dissertation, we propose automatic cell synthesis [7, 10] and a prediction model to significantly reduce the turn-around-time of DTCO and STCO explorations on various physical layout factors: (i) SDC library sets (i.e., different cell heights, Conv. FET and CFET SDC architectures), (ii) design rules (DR), (iii) BEOL parameters, and (iv) power delivery network (PDN) configurations in Section 3 and Section 4.

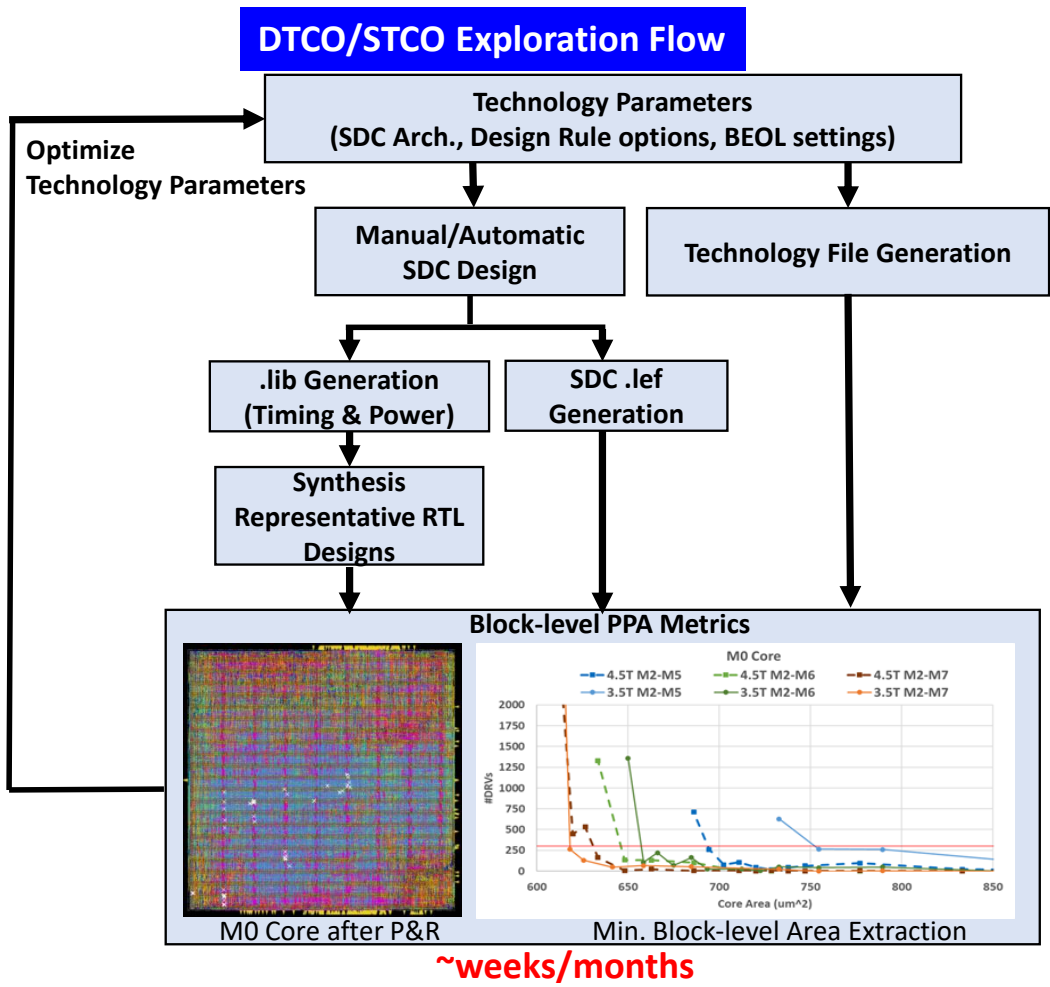


Figure 1.14: Design Technology Co-Optimization and System Technology Co-Optimization Flow

Chapter 2

Rotuability Driven Complementary-FET (CFET) Standard Cell Synthesis for Block-Level Area Optimization

2.1 Introduction

As the technology is scaling beyond $7nm$, cell layout scaling of conventional (Conv.) FET structure is limited due to the routing congestion, the lateral P-N separation, and performance requirements. Complimentary-FET (CFET) technology, which stacks the P-FET on N-FET or vice versa, can release in-cell routing congestion of P-N connection such that SDC designers continue the cell size reduction in sub-7nm. Figure 2.1 shows an illustration of Conv. cell structure and CFET cell structure which stacking P-FET on N-FET. Compared to Conv. cell architecture, the shared or split Gate and S/D (G/S/D) structure

provides flexible local interconnect connection¹.

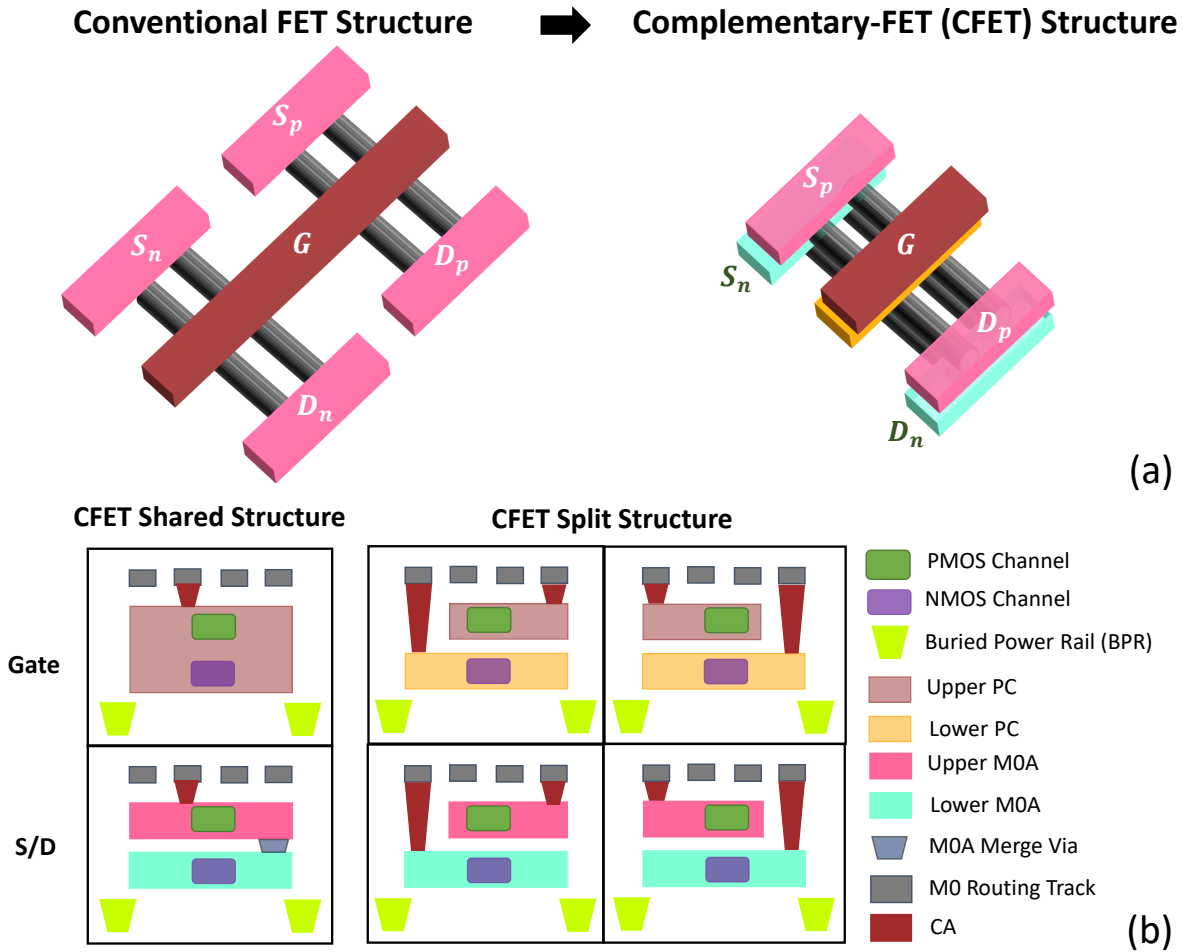


Figure 2.1: An illustration of Conventional and Complementary FET (CFET) structure (Top Row). CFET shared and split Gate, Source and Drain structure (Bottom Row) [3–5].

Recently, feasible CFET-based SDC layouts are successfully proposed [3,5], therefore CFET has been one of the promising cell structure in sub-7nm or beyond. However, the SDC design in sub-7nm demands holistic exploration (including block-level analysis) in terms of pin-accessibility and routing congestion due to the limited routing resource and exploded conditional design rules of later physical design procedures. This exploration for SDC design relies on automatic SDC layout synthesis.

¹If the G/S/D of P-FET and N-FET share the same net connection, the G/S/D can be merged and connect to the M0. On the contrary, the G/S/D are split and M0 drops tall and short vias to connect P-FET and N-FET, respectively.

2.1.1 Related Works

The related works can be categorized into SDC synthesis automation and pin-accessibility-driven cell layout categories.

SDC Synthesis Automation. In [23,24], authors reported full automation of cell layout covering transistor-level placement and in-cell routing together, but these approaches are not applicable in the multi-patterning technologies in sub-5nm. For multi-patterning technology nodes, [25–27] proposed SDC synthesis automation, but the placement and routing are performed in separate operations. Recently, in [9], they integrate the placement and routing with dynamic pin allocation (DPA) interface using Satisfiability Modulo theories (SMT) [16]. However, these works focus on the Conv. FET cell structure optimization thus it is not available in CFET cell structure due to the stack-able P/N-FET structure.

Pin-accessibility-driven Cell Layout. Several approaches have attracted considerable attention to improve the pin-accessibility of SDCs in advanced node technology [28–31]. However, these approaches depend on solving sub-problems therefore it is hard to reach the optimal solution of SDC layout because of the intractable search space partitioning and the intrinsic limitation of heuristic methodology. The [32] suggested strict constraint-based pin-accessibility ensuring methodologies such as Minimum I/O Pin Length (MPL) and Minimum I/O Pin Opening (MPO). However, holistic pin-accessibility between placed SDCs is not proved through block-level analysis.

2.1.2 Our Contributions

We propose a SDC synthesis automation framework for CFET by a dynamic pin shape/allocation scheme, resulting in optimized cell layout considering various design considerations. Moreover, our SDC layout has maximized routability (not only pin-accessibility) through integrated guarantees in our objectives and constraints. Our main contributions are as follows.

- We construct a framework to automate the CFET SDC physical synthesis including concurrent transistor placement and in-cell routing through a novel Dynamic Complimentary Pin Allocation (DCPA) scheme.
- We formulate an integrated constraint satisfaction problem (CSP) for SMT (Satisfiability Modulo theory) solving, including not only place-and-route but also pin-accessibility and design rule related constraints, resulting in the optimized cell layout across whole considerations.
- We propose routability-driven process including strict constraint-based MPL and MPO, objective-based pin separation (pin-accessibility), and objective-based *M2* track use (routing congestion), resulting in a routability-driven SDC layout.
- We validate our framework through the block-level analysis including the #DRV analysis across suggested design features. This empirical results prove that our objectives successfully improve the routability through multi-objective optimization.

The remaining sections are organized as follows. Section 2.2 describes our CFET SDC synthesis framework. Section 2.3 discusses our experiments. Section 2.4 concludes this chapter.

2.2 Routability-Driven Simultaneous Place-and-Route for Complementary-FET (CFET) Standard Cell Synthesis Framework

We utilize an SMT (Satisfiability modulo theories)-based constraints solving methodology for simultaneous place and route of CFET SDCs. In this section, we firstly introduce the overall flow of the proposed routability-driven simultaneous place-and-route CFET SDC synthesis framework. Then, we describe the detailed features of our framework: (i) Overview of CFET SDC Synthesis Framework, (ii)

CFET Cell Architecture and Abstract Pin Interface (API), (iii) Dynamic Complimentary Pin Allocation (DCPA), and (iv) Routability-Driven Cell Optimization. We adopt the FET placement and in-cell routing constraints in Section 1.3.1 and Section 1.3.2 as the basis of our framework and propose a novel dynamic complementary pin allocation scheme to simultaneously place and route CFET SDCs.

2.2.1 Overall Flow of Framework

We show the overall flow of the proposed routability-driven simultaneous place-and-route for CFET Standard Cell Synthesis Framework in Figure 2.2. The inputs of our optimization framework are schematic of cell logic and layout specification (such as design rules, cell heights, etc). Then, our CFET synthesis framework will do concurrent transistor placement and routing to optimize the cell size, routability and total metal length. The output is the optimized CFET cell layout.

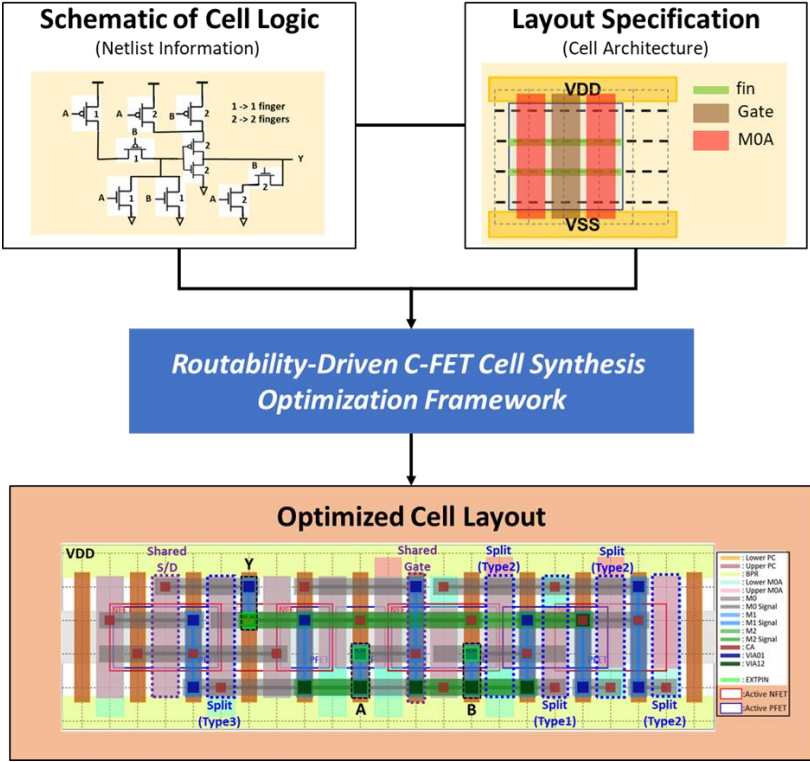


Figure 2.2: Overall flow of the proposed routability-driven simultaneous place-and-route CFET standard cell synthesis framework.

2.2.2 CFET SDC Synthesis Framework Overview

Figure 2.3 shows the overview of the proposed framework. Given cell netlist and layout specification, our framework formulates an integrated constraint satisfaction problem (CSP) for automating CFET SDC layout which strictly satisfies transistor placement, in-cell routing, conditional design rules, and pin-accessibility-driven constraints. Inspired by [9], individual constraints are combined by our novel DCPA constraint. Our framework performs routability-driven lexicographic multiple-objective optimization by implementing (i) Edge-Based Pin Separation and (ii) M2 Track use objectives on top of the cell optimization objectives of [9] (i.e., Cell Size and Metal Length). We utilize five representative conditional design rules, which are *minimum area rule (MAR)*, *end-of-line (EOL)*, *via rule (VR)*, and multi-pattern-aware design rules (i.e., *parallel run-length (PRL)/step height rule (SHR)*) as described in Section 1.3.3. The notations are shown in Table 1.1.

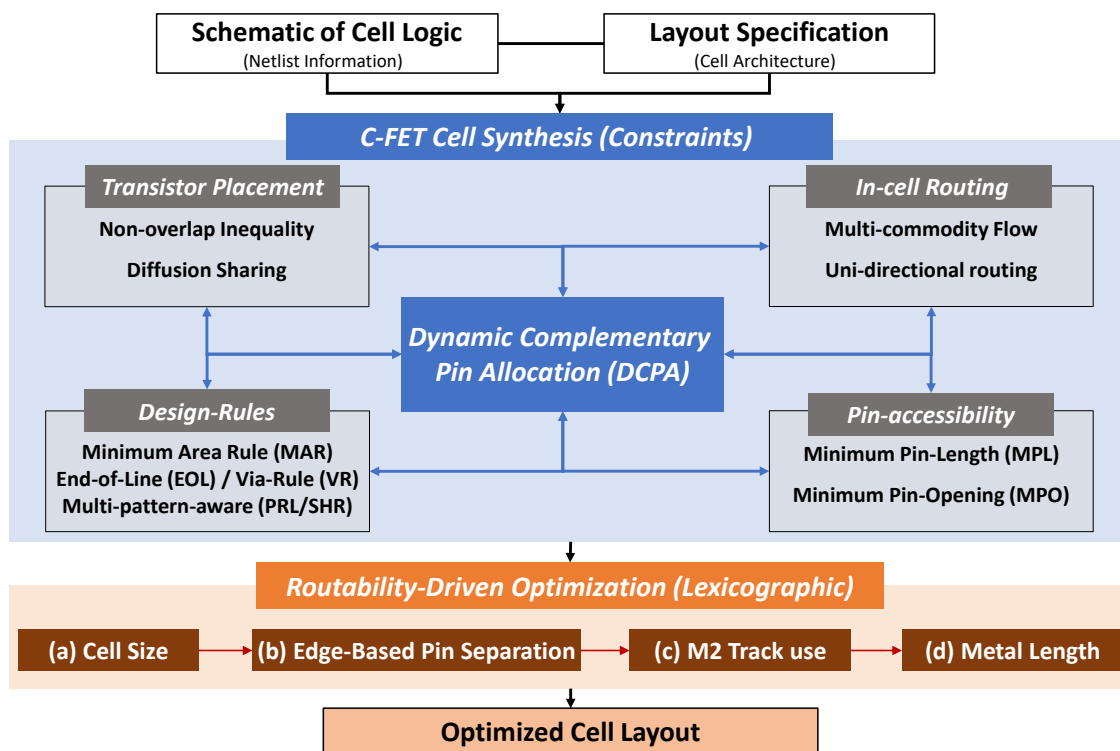


Figure 2.3: CFET Standard Cell Synthesis Framework Overview.

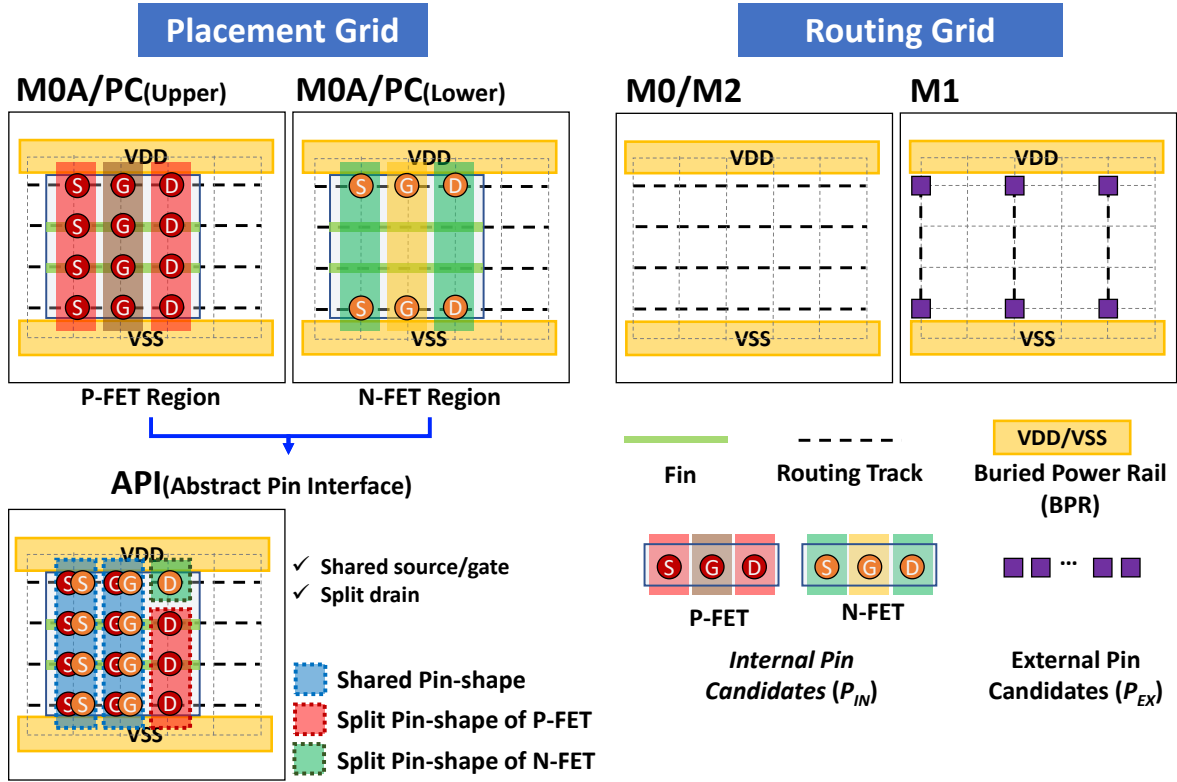


Figure 2.4: Grid-Based placement and routing graph with Abstract Pin Interface (API) using 4 RTs P-on-N CFET example.

2.2.3 CFET Cell Architecture and Abstract Pin Interface (API)

Our framework employs a CFET cell architecture and netlist information of [3, 5] and [8], respectively. Figure 2.4 shows the grid-based placement and routing graph (i.e., Upper/Lower M0A/PC, Abstract Pin Interface (API), M0, M1, and M2) using 4 RTs P-on-N CFET example. The routing grid consists of 4 RTs with buried power rails (i.e., 4.5 Track Height) and each layer is defined as unidirectional edges. We adopt supernodes [19] for the pin of FET (i.e. internal pin, P_{IN}) or the I/O pin of a standard cell (i.e. external pin, P_{EX}). The P-FET and N-FET regions are stacked up on the upper and the lower M0A/PC layers, respectively. Therefore, the access to the M0 layer from each pin on the N-FET region (i.e. lower M0A/PC layer) is restricted to the top or bottom horizontal routing track unless each source/gate/drain pin in P-FET and N-FET that are overlapped on the same vertical track is shared [3, 5]

as shown in Figure 2.1. As a result, there exist three kinds of pin shapes according to the sharing status of each pin in stacked FETs as depicted in Figure 2.1. We propose the Abstract Pin Interface (API) for describing the pin-shapes on upper and lower MOA/PC layer together for in-cell routing, resulting in reduced complexity of our framework.

2.2.4 Dynamic Complimentary Pin Allocation (DCPA)

DCPA dynamically constructs the shared and split pin-shapes of FETs in the API for an optimal in-cell routing exploration of a CFET structure. The DCPA scheme for simultaneous place-and-route follows the same principle as [10] for interconnecting placement and routing formulas using flow capacity variables (i.e., $C_m^n(v, u)$).

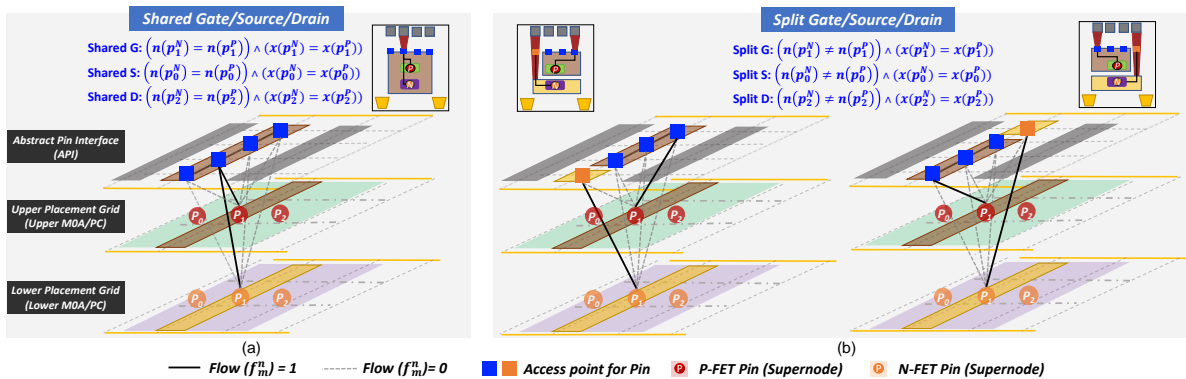


Figure 2.5: Concept of Dynamic Complimentary Pin Allocation (DCPA) for CFET cell structure using 4 RTs P-on-N CFET example. p_1^P =P-FET Gate Pin. p_1^N =N-FET Gate Pin.

Figure 2.5 illustrates the concept of DCPA using 4 RTs P-on-N CFET as an example. When the pins of P-FET and N-FET are located at the same x-coordinate (i.e., $x(p_i^P) = x(p_j^N)$), the pin-shape (i.e., shared or split) at the corresponding column in the API is determined by the net information. For example, in Figure 2.5(a), if both of the gate pins p_1^N and p_1^P belong to the same net (i.e., $n(p_1^N)=n(p_1^P)$), a shared pin-shape is selected and one of the corresponding flow variables (i.e., f_m^n) among four possible access points (i.e., blue squares) is determined by the flow formulation. On the other hand, if each gate

Algorithm 2 Dynamic Complementary Pin Allocation (DCPA)

*/*l is Abstract Pin Interface in $G(V, E)$ */*

```
1: if ( $n(p_{i,t}^P) = n(p_{j,s}^N) \wedge (x(p_{i,t}^P) = x(p_{j,s}^N))$ ) then
2:   {/* Shared Pin-Shape */}
3:   Exp. (2.1) for P-FET and N-FET access.
4: else if ( $n(p_{i,t}^P) \neq n(p_{j,s}^N) \wedge (x(p_{i,t}^P) = x(p_{j,s}^N))$ ) then
5:   {/* Split Pin-Shape of P-on-N CFET Structure */}
6:   if ( $n(p_{i,t}^P) = \text{VDD}$ ) then
7:     Exp. (2.3) for access Lower N-FET.
8:   else if ( $n(p_{j,s}^N) = \text{VSS}$ ) then
9:     Exp. (2.4) for access Upper P-FET.
10:  else
11:    Exp. (2.2)  $\vee$  Exp. (2.3) for access P-FET and N-FET.
12:  end if
13: end if
```

pin belongs to different nets (i.e., $n(p_1^N) \neq n(p_1^P)$), DCPA selects one of two possible split pin-shapes (i.e., top or bottom access point for N-FET) as shown in Figure 2.5(b). Meanwhile, when the upper FET pin has a connection to the power rail (i.e., VDD or VSS), DCPA selects the split pin-shape without blocking the power rail connection of upper FET pin.

Algorithm 2 utilizes SMT's *if-then-else* structure to describe a generation procedure of the constraint for our novel dynamic pin-shape selection scheme. When N-FET and P-FET pins share the same location $x(p)$, the net information $n(p)$ is used to determine the corresponding pin-shape in API. Then, the flows of each N-FET and P-FET pin are set to 1 for the corresponding access points of each pin-shape. The expressions of shared and split pin-shapes are shown as follows.

Shared Pin-Shape:

$$\bigwedge_{y=1, \dots, h-1} (f_m^n(v_{x,y,l}, v_{x,y+1,l}) = 1), \quad n = n(p_{i,t}^P) = n(p_{j,s}^N), x = x_t^P + i \quad (2.1)$$

Split Pin-Shape:

Top Access for Lower FET (Type1):

$$f_m^{n1}(v_{x,h-1,l}, v_{x,h,l}) = 0 \wedge \left(\bigwedge_{y=1, \dots, h-2} (f_m^{n1}(v_{x,y,l}, v_{x,y+1,l}) = 1) \right) \\ \wedge \left(\bigwedge_{y=1, \dots, h-1} (f_m^{n2}(v_{x,y,l}, p_{j,s}^L) = 0) \right) \wedge (f_m^{n1}(v_{x,h,l}, p_{i,t}^U) = 0) \quad (2.2)$$

Bottom Access for Lower FET (Type2):

$$f_m^{n1}(v_{x,1,l}, v_{x,2,l}) = 0 \wedge \left(\bigwedge_{y=2, \dots, h-1} (f_m^{n1}(v_{x,y,l}, v_{x,y+1,l}) = 1) \right) \\ \wedge \left(\bigwedge_{y=2, \dots, h} (f_m^{n2}(v_{x,y,l}, p_{j,s}^L) = 0) \right) \wedge (f_m^{n1}(v_{x,1,l}, p_{i,t}^U) = 0) \quad (2.3)$$

No Access for Lower FET (Type3):

$$\bigwedge_{y=1, \dots, h-1} (f_m^{n1}(v_{x,y,l}, v_{x,y+1,l}) = 1) \wedge \left(\bigwedge_{y=1, \dots, h} (f_m^{n2}(v_{x,y,l}, p_{j,s}^L) = 0) \right) \quad (2.4) \\ n_1 = n(p_{i,t}^U), n_2 = n(p_{j,s}^L), x = x_t^U + i, \quad \begin{cases} U = P, L = N, \text{ if P-on-N} \\ U = N, L = P, \text{ if N-on-P} \end{cases}$$

If N-FET and P-FET pins have the same net information, the shared pin-shape is selected (Lines 1-2). Otherwise, the split pin-shape is selected (Lines 4-12). The split pin-shape consists of three types on API layer. *Type1* and *Type2* represent top ($y=h$) and bottom ($y=1$) accesses for lower FET, respectively. If the net of lower FET pin is VSS, *Type3* is used since there is no connection from M0 to lower FET pin (Lines 9). In P-on-N CFET structure, *Type2* is always selected (Lines 7) when the net of P-FET pin is VDD. Otherwise, *Type1* or *Type2*, which satisfies all the constraints and produces the optimal solution, is selected (Lines 11) for P-on-N or N-on-P CFET structure.

2.2.5 Routability-Driven Constraints and Objectives

For routability, we propose strict constraint-based pin-accessibility improvement methods; (i) minimum I/O pin opening (MPO) and (ii) minimum I/O pin length (MPL). Also, we suggest new objectives, (i) edge-based pin separation (EB-PS) and (ii) M2 track use, to improve the pin-accessibility and the routing resource congestion, respectively.

Constraint-based Pin-Accessibility Improvement

Our framework utilizes MPL and MPO constraints to improve pin accessibility as [32] suggested. The MPL and MPO constraints and examples are shown in the below.

MPL (Minimum I/O Pin Length): MPL rule defines the minimum number of metal segments of the commodity heading to the external pin P_{EX} on the M1 layer as shown in Figure 2.6(a). At-least 1 (**AL1**) metal segment on the M1 layer must be assigned to the commodity whose sink is P_{EX} as expressed in Constraint (2.5). Then, the metal segment on the M1 layer is extended to the minimum length defined by MAR. The vertices on the extended segments are the possible pin access points.

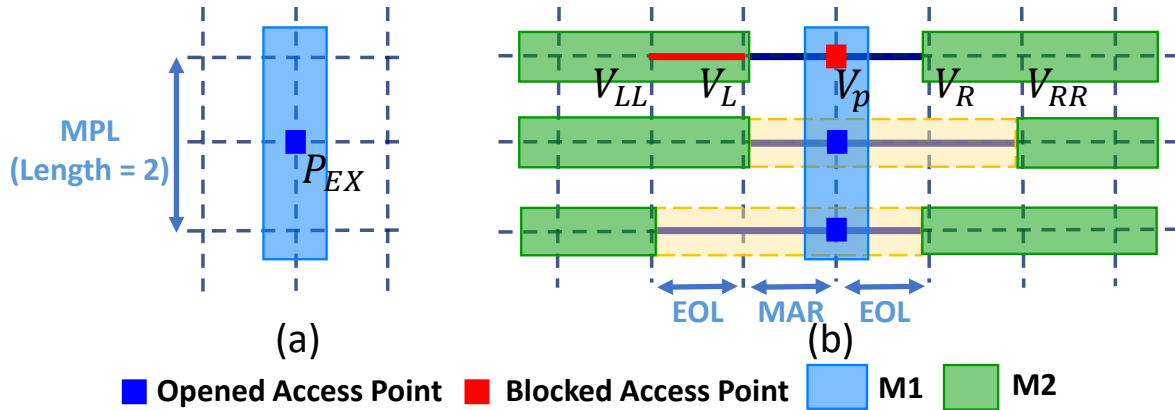


Figure 2.6: An example of (a) MPL with MAR=2, (b) MPO with EOL/MAR = 1/1.

$$\mathbf{AL1}(m_{v,v_F}, m_{v,v_B}), \quad \text{if } f_m^n(v, v_D) = 1, f_m^n(v, v_U) = 1; \quad \forall v \in V_1, m = P_{EX} \quad (2.5)$$

MPO (Minimum I/O Pin Opening): MPO rule ensures the minimum number of unblocked access points (i.e., pin openings) from the *M2* layer for each I/O pin. Figure 2.6(b) illustrates that each pin candidate v_p has to secure enough horizontal space on the *M2* layer so that it can be accessed through the *M2* layer without violating design rules such as the MAR and EOL. MPO considers each v_p as the possible pin opening if there is no routed metal segment in the opening mask (depicted in light yellow rectangles) on the *M2* layer. MPO is a boolean cardinality constraint to ensure at-least-k (**ALk**) true pin opening indicator O_{v_p} among the possible candidates $Q(p)$ as described in Constraint (2.6). If there exist any edges $e_{v,u}^n$ on the *M2* layer, MPO is not applied because the external pin p already has unblocked access points on the *M2* layer.

$$\mathbf{ALk}(\{O_{v_p} \mid v_p \in Q(p)\}), \quad \text{if } \bigvee_{v \in V_2, u \in V_2} e_{v,u}^{n(p)} = 0 \quad (2.6)$$

For the example of Figure 2.6(b), the O_{v_p} is set to 1 (true) if there is no routed metal segment in the opening mask, whose length is the summation of MAR and EOL parameters (i.e., $\text{MAR} + 2 \times \text{EOL}$), as shown in Constraint (2.7).

$$\overline{O}_{v_p} = \sum_{n \in N, n \neq n(p)} \left((e_{v_{LL}, v_L}^n \vee e_{v_L, v_p}^n \vee e_{v_p, v_R}^n) \wedge (e_{v_L, v_p}^n \vee e_{v_p, v_R}^n \vee e_{v_R, v_{RR}}^n) \right),$$

$$\forall v_p \in Q(p), \begin{cases} v \in Q(p), & \text{if } e_{v_D, v_{DF}}^n = 1, e_{v_D, v_{DB}}^n = 1 \\ v \notin Q(p), & \text{otherwise} \end{cases}, \forall v \in V_2 \quad (2.7)$$

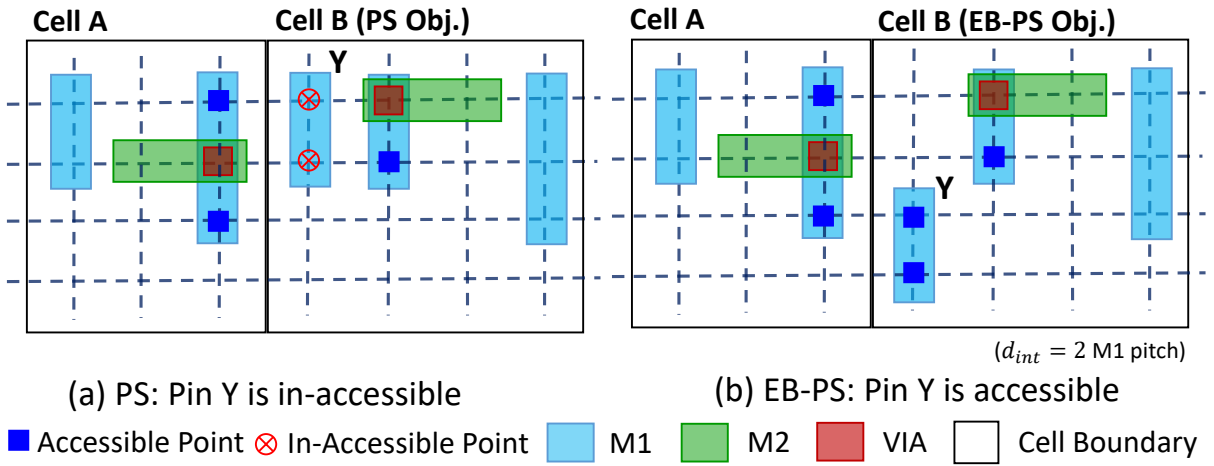


Figure 2.7: An illustration of Pin Separation (PS) [6] and Edge-Based Pin Separation (EB-PS) [7].

Objective-based Pin-Accessibility Improvement

We firstly introduce pin separation objective [6] for pin-accessibility. Then, we introduce an enhanced objective function, edge-based pin separation [7], to further mitigate the pin interference in the cell layout.

Pin Separation (PS): Pin Separation (PS) objective counts the number of I/O pins that keep the minimum spacing (i.e., d_{int} [29]) from each other. Then we maximize the total number to disperse the pins as many as possible by maximizing the column-based pin separation space (i.e., $S(p)$) as shown in Expression (2.8).

$$\text{Maximize: Pin-accessibility (Pin Separation)} = \sum_{p \in P_{EX}} S(p) \quad (2.8)$$

$$S(p) = \bigwedge_{e_{v,q} \in E_k^{M1}, k \in d_{int}(x(p)), q \in P_{EX}, q \neq p} \neg e_{v,q}^{n(q)}$$

Figure 2.8 illustrates the different pin-accessibility according to the spacing between pins. Figure 2.8(a) and (b) have the I/O pins with spacing of 0 and 1, respectively. When the $M2$ wires are accessing the left and right pins, the access points of the center pin in Figure 2.8(a) become in-accessible. In contrast, the center pin in Figure 2.8(b) maintains its accessibility. This shows that as the more pins are scattered across a cell, the more access points are available.

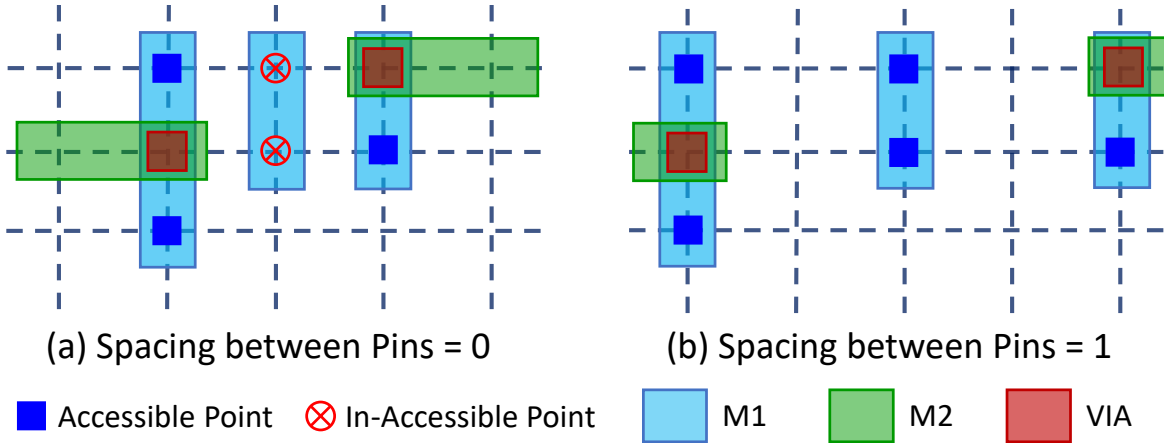


Figure 2.8: An illustration of Pin Separation (PS).

Edge-Based Pin Separation (EB-PS): We propose an Edge-Based Pin Separation (EB-PS) objective, which not only keeps the minimum spacing (i.e., d_{int} [29]) between SDC I/O pins but also reduces the adjacent parallel pin shapes within d_{int} . Then we minimize the total pin cost which caused by adjacent pins and parallel pin shapes within d_{int} . Figure 2.7 illustrates the different pin-accessibility according to Pin Separation (PS) [6], which maximizes the sum of I/O pins that keep d_{int} spacing with other I/O pins, and the proposed EB-PS. In Figure 2.7, Cell A is adjacent to Cell B. Cell B in Figure 2.7 (a) and (b) are

generated by PS [6] and EB-PS, respectively. Although the spacing between pins in cell B is maximized by PS [6] objective, the pin Y is still in-accessible because the $M2$ wires are accessing its left pin in Cell A and right pin in Cell B. Besides considering the spacing between pins, our EB-PS objective also reduces the adjacent parallel pin shapes within d_{int} as shown in Cell B in Figure 2.7 (b), resulting in accessible pin Y. This shows that as the more pins are scattered across a cell with minimum parallel adjacent pin shape within d_{int} , the more access points are available.

Objective-based Routing-Congestion Minimization ($M2$ track use)

We set a # $M2$ track objective, which counts the number of occupied $M2$ tracks in a cell, and minimize it to suppress the routing congestion because the $M2$ layer is jointly used in both front-end and back-end layout design. Figure 2.9 illustrates the impact of the different $M2$ layer blockages on the routing congestion. The I/O pin V_A has a connection to pin V_B . The $M2$ blockages in Figure 2.9(a) has metal segments = 3 and occupies 3 routing tracks. The $M2$ blockages in Figure 2.9(b) has metal segments = 4, but, occupies 2 routing tracks. As a result, the connection in Figure 2.9(b) can be directly routed, while the connection in Figure 2.9(a) must have a detour in the routing path. This demonstrates that the number of occupied $M2$ tracks has more impact on the routing congestion than the length of $M2$ metals.

2.2.6 Multi-Objective Optimization (Optimal Priority)

Our framework has multiple objectives associated with placement and routing problems for standard cell layout design. The first objective is cell size which is defined as the right-most vertical track as shown in Expression (2.9). The second objective is Edge-Based Pin Separation (EB-PS) and it minimizes the summation of column-based and edge-based pin costs (i.e., $SC(p)$ and $EC(p)$) of each SDC I/O pin in Expression (2.10). $SC(p)$ is 1 if there are adjacent pins within an interference distance d_{int} . $EC(p)$

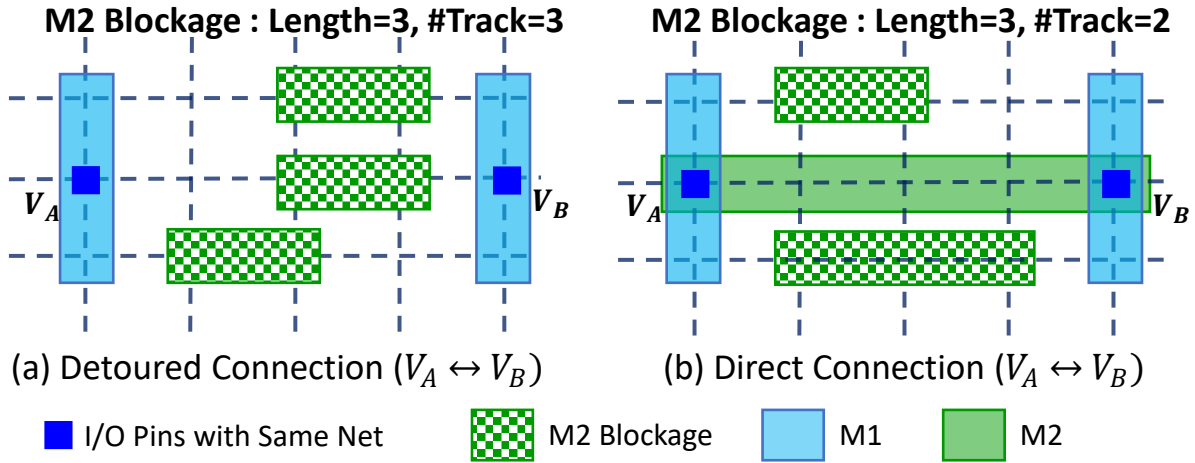


Figure 2.9: An illustration of the Impact of $M2$ Blockage on the Routing Congestion.

is the summation of adjacent parallel pin shapes within d_{int} . The third objective is the number of $M2$ tracks used for in-cell routing (Expression (2.11)). The last objective is the weighted sum of routed metal segments (i.e., Total Metal Length (ML)) as shown in Expression (2.12). In practice, the cell size has the highest priority because it has a direct impact on the area of a whole chip. The PS should be considered as the second objective because the in-accessible pins can not be routed regardless of the routing resources. Then the number of $M2$ tracks has been used as a more important metric than Total ML to maximize the routability by reserving upper routing resources. Therefore, our framework simultaneously optimizes these multiple objectives based on addressed “lexicographic” order (Expression (2.13)) through an optimization feature of OMT [16].

$$\text{Minimize: Placement (Cell Size)} = \max \{x_t + w_t \mid t \in T\} \quad (2.9)$$

$$\text{Minimize: Pin-accessibility (EB-PS)} = \sum_{p \in P_{EX}} SC(p) + EC(p)$$

$$SC(p) = \bigvee_{e_{v,q} \in E_k^{M1}, k \in d_{int}(x(p)), q \in P_{EX}, q \neq p} e_{v,q}^{n(q)} \quad (2.10)$$

$$EC(p) = \sum_{e_{v,u}^{n(p)}} \bigvee_{e_{v,u}^n \in E_k^{M1}, k \in d_{int}(x(p)), n \in N_{EX}, n \neq n(P)} e_{v,u}^n$$

$$N_{EX} = \{n(p) | p \in P_{EX}\}$$

$$\text{Minimize: Routability (\#M2 Track)} = \sum_{k=1}^h \bigvee_{e_{v,u} \in E_k^{M2}} m_{v,u} \quad (2.11)$$

$$\text{Minimize: Total ML} = \sum_{e_{v,u} \in E} (w_{v,u} \times m_{v,u}) \quad (2.12)$$

Lexicographic Optimization:

(a) *CellSize*, (b) *EB-PS*, (c) *\#M2Track*, (d) *TotalML*

(2.13)

2.3 Experiments

Our framework is implemented in Perl/SMT-LIB 2.0 standard-based formula and executed on a workstation with 2.4GHz Intel Xeon E5-2620 CPU and 256GB memory. The single-threaded SMT Solver Z3 [16] (version 4.8.5) is used to produce the optimized solution in the proposed framework.

2.3.1 Experimental Setup

SDC Generation: We use ASAP7 [8] SDC SPICE netlists as inputs of P-on-N CFET SDCs. We adopt the same FET width and number of fingers from [8] for SDC layout generation in the following experiments. Note that 30 representative cells [4], which are specified in Table 2.1, are selected for all experiments. The experiment parameters of conditional design rules [21] are as follows: MAR/EOL/VR/PRL/SHR = 1/1/1/1/2².

Block-level P&R: Three open source RTL designs [33], M0 Core, M1 Core, and AES that respectively have 17K, 20K, and 14K instances are adopted³. The cell statistics of each design are listed in Figure 2.10. We perform the block-level analysis through a Place-and-Route suite [14].

For BEOL, we set the contacted poly pitch (CPP), M0/M2 pitch⁴ and the number of masks for each BEOL layer according to [1]. For M1, VIA12, and M2 layers, the grid-based conditional design rules' parameters are applied at block-level as shown in Figure 2.11. The metals' pitch and width of layers above M2 are set based on reference [34]. The power delivery network consists of top power meshes (M8 and M9), intermediate power stripes (M3), and standard cell rails (BPR). The top power mesh is designed as spaces is allowed. Then, the power is delivered from M3, which is 4× wider than signal wires, to M1 and M1 to BPR using stacked vias and SuperVia models [35], respectively. The

²We assume that the VR of CA layer allows at most two diagonal vias in the experiments.

³The worst negative slacks of M0 Core, M1 Core, and AES are carefully adjusted between 50 and -50ps for a fair comparison in the block-level analysis.

⁴The M0/M2 pitches and widths are 24nm and 12nm with 2 masks. The CPP and M1 pitch are 42nm.

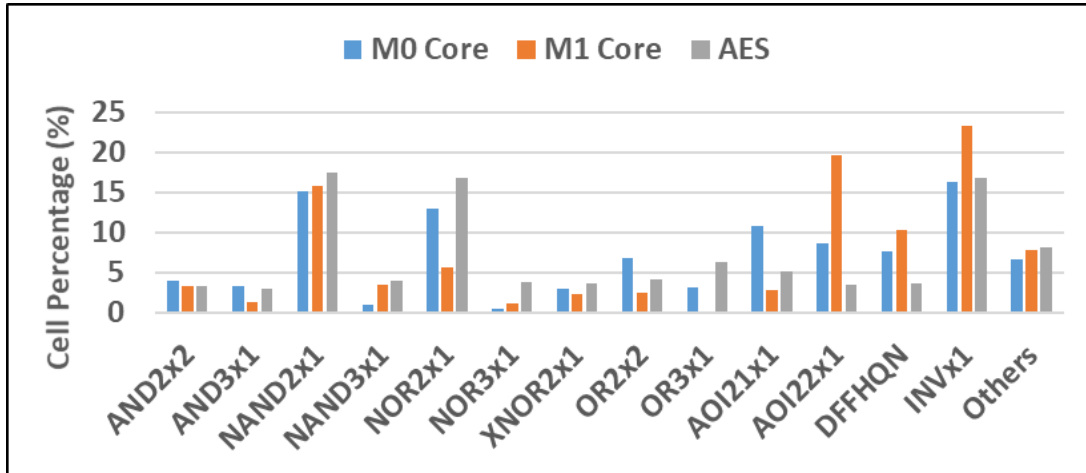


Figure 2.10: Cell Statistics of M0 Core, M1 Core, and AES.

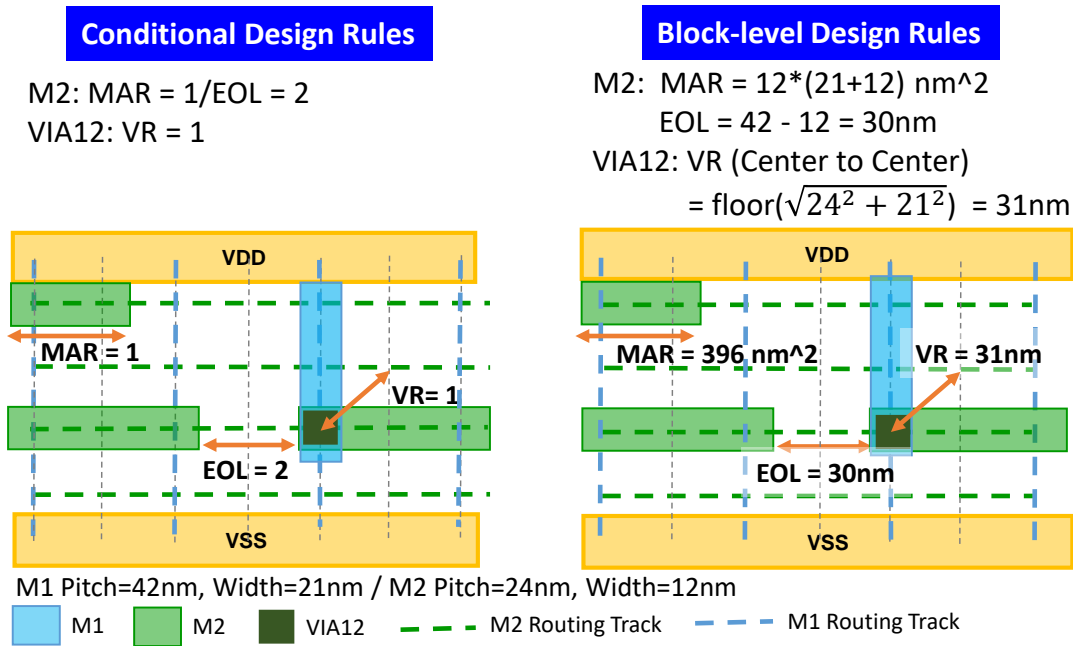


Figure 2.11: An example of transferring the grid-based conditional design rules to the block-level.

M3 power stripes for the BPR (Buried Power Rail) standard cell rail are placed per every 64 CPPs [36].

We use 300 #DRVs threshold⁵, which is depicted in red horizontal line in the figures representing the block-level P&R results, to measure the valid block-level area.

⁵As a common industrial practice, once the number of DRVs increases beyond 300, the block layout is deemed too troublesome to fix with laborious engineering change orders (ECOs).

The experiments are organized as follows:

- **Exp. 2.3.2.** *CFET vs. Conv. SDC*: We study the cell metrics of optimized CFET and Conv. SDC layouts using 4.5T SDC architecture.
- **Exp. 2.3.3.** *Routability-Driven Cell Optimization*: We demonstrate that our routability-driven constraints and objectives improve the routability of CFET SDC layouts using cell metrics.
- **Exp. 2.3.4.** *Block-Level Routability Analysis*: We validate our routability-driven framework using #DRVs analysis at block-level.

2.3.2 CFET vs. Conv. Standard Cell (SDC)

In this section, we demonstrate the optimized XOR2x1 CFET cell layout using our DCPA scheme and compare the cell width, total metal length, and #M2 Tracks of CFET SDCs with Conv. SDCs. The Conv. SDCs are generated using the framework in [9]. For the fair comparison in terms of our metrics, we adopt the same in-cell horizontal routing tracks (i.e., 4 tracks) and push the SDC power rail to BPR layer for Conv. SDC structure [37].

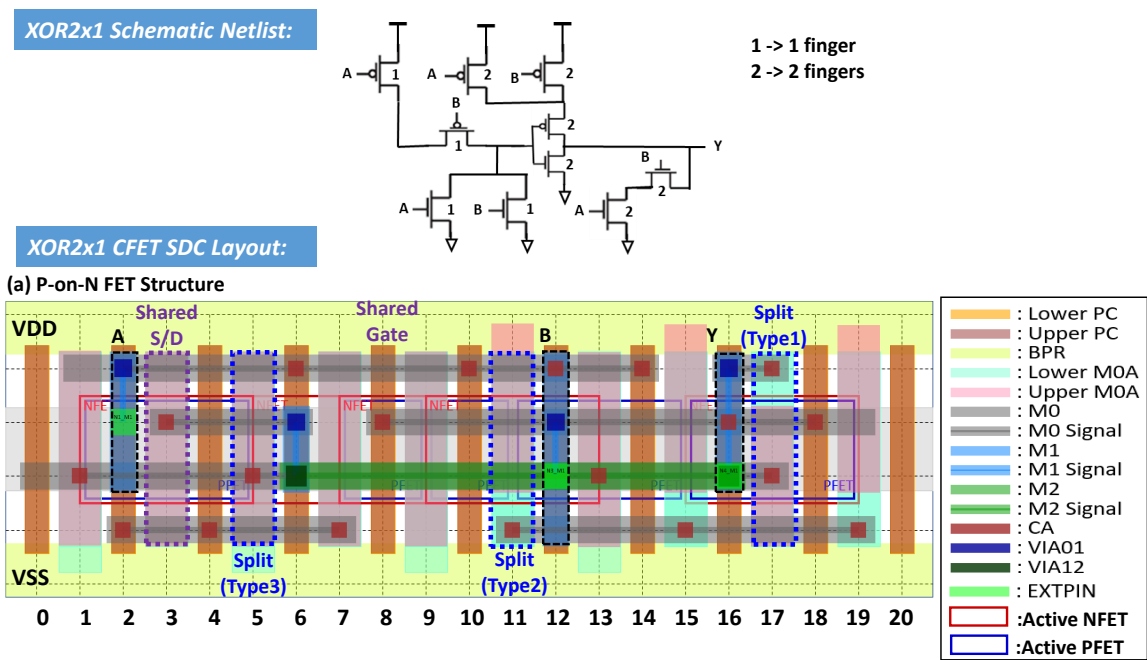


Figure 2.12: An example of XOR2x1 Schematic Netlist [8] and CFET SDC Layout.

Figure 2.12 shows the netlist and the generated CFET cell layout of an XOR2x1 SDC. The shared and split pin-shapes have been successfully selected by our DCPA scheme. When the source and drain of each FET on the same column have the same net information, the shared pin-shapes are selected (depicted in columns 3, 8). If the net information is different, the split pin-shapes are selected and the locations (i.e., top or bottom) of N-FET access points are determined by DCPA. If the source or drain of each FET have no connection to VDD or VSS, the top or bottom (i.e., *Type1*, depicted in columns 17) track is selected as

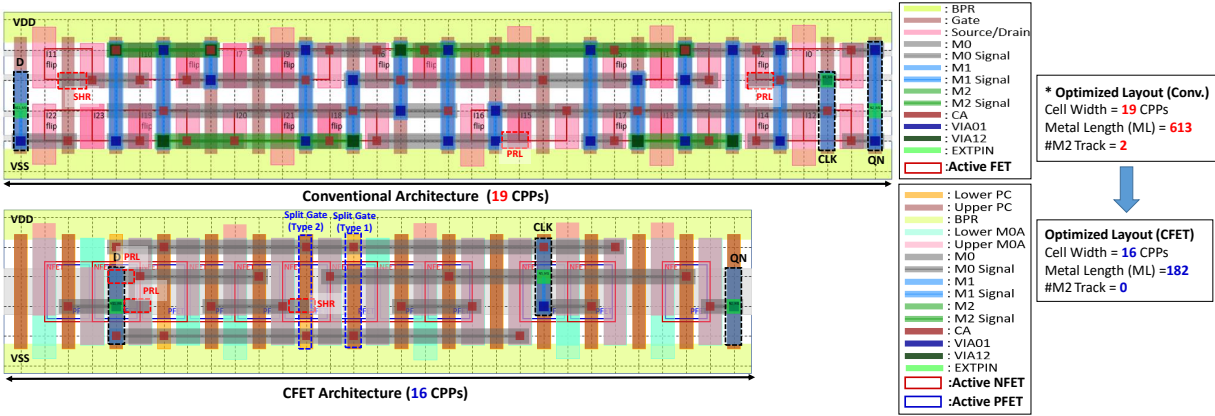


Figure 2.13: Layouts of 4 routing tracks CFET and Conv. DFFHQN with corrected design constraints. Optimized result of CFET layout: Cell Size (19→16), Metal Length (613→182), #M2 Tracks (2→0). The red dash-line boxes are metal extension for PRL and SHR design constraints.

an access point of N-FET. For the P-FET with VDD connection and N-FET with VSS connection, *Type2* (depicted in columns 11) and *Type3* (depicted in column 5) are respectively selected.

Table 2.1 enumerates the comparison results of CFET SDCs and Conv. SDCs. The number of FETs in each cell varies from 2 to 24 and the average runtime per cell is less than 12 minutes. Compared to the Conv. SDCs, the CFET SDCs achieves 3.12%, 22.09%, and 45.9% reduction on the average cell width, metal length, and #M2 tracks, respectively. Figure 2.13 shows a design-rule corrected DFFHQN cell layout for Conv. and CFET architectures. All metal segments that are depicted in red dashed rectangles are successfully extended to satisfy conditional design rules such as PRL and SHR. By virtue of directed P-N connections between stacked FETs, DFFHQN with CFET consumes 70.3%, 2, and 3 less metal length, #M2 tracks, and CPPs of cell width than with Conv. cell structure, respectively.

Table 2.1: Experimental Statistics: ML= Metal Length. CPP= Contact Poly Pitch. Cell Width Red. = ((Cell Width of reference - Cell Width of CFET)/Cell width of reference). ML Red. = ((ML of reference - ML of CFET)/ML of reference).

Cell Specification			Cell Layout Objectives								Runtimes (s)	
Name	#FET	#Net	Cell Width (CPPs)			Metal Length			#M2 Tracks		Conv.	CFET
			Conv.	CFET	Impr (%)	Conv.	CFET	Impr (%)	Conv.	CFET		
AND2x2	6	7	6	6	0.00	75	60	20.00	0	0	8.86	8.12
AND3x1	8	9	6	6	0.00	91	68	25.27	0	0	15.10	30.09
AND3x2	8	9	7	7	0.00	97	76	21.65	0	0	18.36	28.35
AOI21x1	6	8	9	9	0.00	197	142	27.92	1	0	20.66	119.69
AOI22x1	8	10	14	11	21.43	311	255	18.01	1	1	210.91	363.16
BUFx2	4	5	5	5	0.00	61	40	34.43	0	0	3.48	4.92
BUFx3	4	5	6	6	0.00	82	53	35.37	0	0	9.87	11.20
BUFx4	4	5	7	7	0.00	88	59	32.95	0	0	11.47	7.91
BUFx8	4	5	12	12	0.00	149	105	29.53	0	0	34.80	43.65
DFFHQN	24	17	19	16	15.79	613	182	70.31	2	0	3335.88	6831.77
FA	24	17	14	14	0.00	420	379	9.76	3	2	5259.34	6653.07
INVx1	2	4	3	3	0.00	44	23	47.73	0	0	2.94	0.49
INVx2	2	4	4	4	0.00	38	29	23.68	0	0	1.59	1.03
INVx4	2	4	6	6	0.00	65	48	26.15	0	0	5.37	3.46
INVx8	2	4	10	10	0.00	121	92	23.97	0	0	26.19	19.14
NAND2x1	4	6	6	6	0.00	79	74	6.33	0	0	6.80	15.88
NAND2x2	4	6	10	10	0.00	140	131	6.43	0	0	23.56	33.83
NAND3x1	6	8	11	11	0.00	152	149	1.97	0	0	135.14	124.11
NAND3x2	6	8	21	21	0.00	305	286	6.23	0	0	1661.21	2869.53
NOR2x1	4	6	6	6	0.00	79	74	6.33	0	0	6.65	12.89
NOR2x2	4	6	10	10	0.00	140	131	6.43	0	0	162.01	27.94
NOR3x1	6	8	11	11	0.00	152	148	2.63	0	0	35.45	52.33
NOR3x2	6	8	21	21	0.00	304	283	6.91	0	0	2503.73	1897.53
OAI21x1	6	8	11	9	18.18	247	146	40.89	1	0	168.64	52.52
OAI22x1	8	10	14	11	21.43	311	240	22.83	1	1	235.15	612.60
OR2x2	6	8	6	6	0.00	75	60	20.00	0	0	7.20	12.99
OR3x1	8	9	6	6	0.00	91	68	25.27	0	0	13.33	76.77
OR3x2	8	9	7	7	0.00	97	76	21.65	0	0	19.16	89.22
XNOR2x1	10	9	12	11	8.33	274	220	19.71	1	1	702.89	977.00
XOR2x1	10	9	12	11	8.33	276	214	22.46	1	1	1107.30	134.86
Avg.	6.80	7.70	9.73	9.30	3.12	172.47	130.37	22.09	0.37	0.20	525.10	703.87

2.3.3 Routability-Driven Cell Optimization

In this section, we validate our routability-driven constraints and objectives using the statistics of cell metrics, pin-accessibility metrics, #M2 tracks, and M2 metal length with multiple CFET SDC sets generated by our framework.

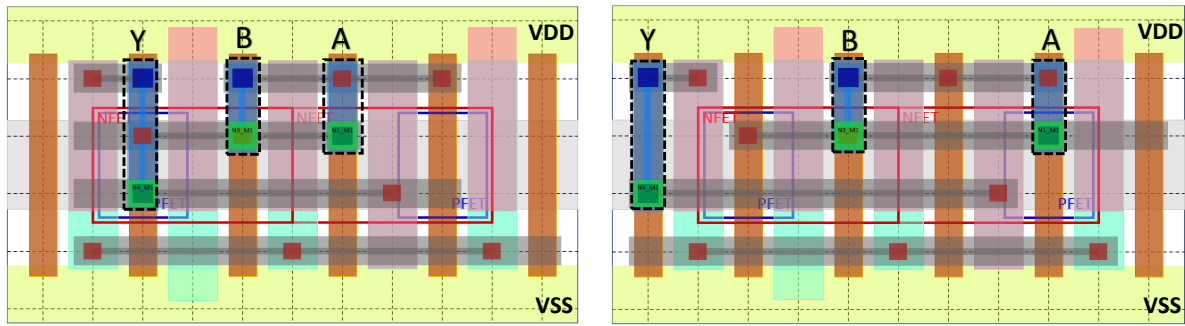
Optimization for Pin-Accessibility

For pin-accessibility, we validate the proposed Pin Separation (PS) and Edge-Based Pin Separation (EB-PS) objectives using the cell-level metrics.

Pin Separation (PS): Figure 2.14 shows the different I/O pin distributions by PS objective for a NAND2x1 cell with design parameters, $MPO = 2$ and $d_{int} = 1.5$ M1 pitch. The MPL and MPO constraints respectively ensure at-least one $M1$ metal segment and at-least 2 pin-openings for each I/O pin (depicted in blue dashed rectangles). Though the number of pin-openings is the same for each I/O pin, the PS cost of each cell is different. While the pins of the “MPO Only” cell layout are concentrated in a certain region (Figure 2.14(a)), the pins of the “PS with MPO” cell layout are distributed keeping the minimum distance from each other (Figure 2.14(b)). The RPA⁶ value of each pin in “MPO Only” cell layout is smaller (i.e., worse) than that of the cell layout with PS. In particular, the RPA value of the pin B without PS is 0.33. This means that the pin B is not likely to be accessed successfully, because we need at least one access point. On the contrary, all the pins with PS have the same RPA value with the number of pin-openings. This demonstrates that the PS optimization efficiently improves the pin-accessibility by taking the full advantages of the MPO constraint.

Table 2.2 shows the comparison of key metrics for the split cases of MPO constraint and PS objective. MPO=2/MPO=3 denotes that the SDCs are generated with the MPO parameter = 2/3. “wo PS/EB-

⁶The RPA of [29] indicates how many access points of a pin remain after the accesses of its neighboring pins.



(a) MPO Constraint Only

(b) Pin Separation with MPO Constraint

I/O Pin	MPO Only			PS with MPO		
	Y	B	A	Y	B	A
Pin Separation		0			3	
#Pin Opening	3	2	2	3	2	2
RPA	2	0.33	1	3	2	2

($MPO = 2, d_{int} = 1.5 \text{ M1 pitch}$)

Figure 2.14: Layout of NAND2x1 cell optimized for Pin Separation objective.

“PS” denotes that the SDCs are generated without pin-access objective (i.e., (a)*CellSize*, (b)*#M2Track*, and (c)*TotalML*); “PS” denotes that the SDCs are generated using PS [6] objective for pin-accessibility (i.e., (a)*CellSize*, (b)*PS* (i.e., Expression (2.8)), (c)*#M2Track*, and (d)*TotalML*) and “EB-PS” refers to the objectives with the proposed EB-PS (i.e., the same objectives with Expression (2.13)).

The Min.#PO values, which refers to the minimum of pin-openings in each SDC, shows that our cell layout is successfully ensured #PO by suggested MPO constraint. The average PS-objective values of “PS” cases are $2.43\times$ and $2.15\times$ larger than “wo PS/EB-PS” cases for $MPO = 2$ and 3 , respectively. This shows that our PS objective effectively dispersed the I/O pins for each SDC. The Min.RPA value represents the minimum accessible pin-openings of generated CFET SDCs. “PS” cases have more accessible pin-openings than “wo PS/EB-PS” cases by 13.0% and 18.5% for “MPO=2” and “MPO=3”, respectively. This demonstrates that the MPO constraint successfully ensured the minimum number of pin-openings and the PS objective contributed to maximize the effective pin-openings. As the MPO increases and the PS is maximized, the average ML and *#M2Track* are increased due to the enlarged and scattered I/O pins.

Edge-Based Pin Separation (EB-PS): Here, we demonstrate that our EB-PS objective efficiently im-

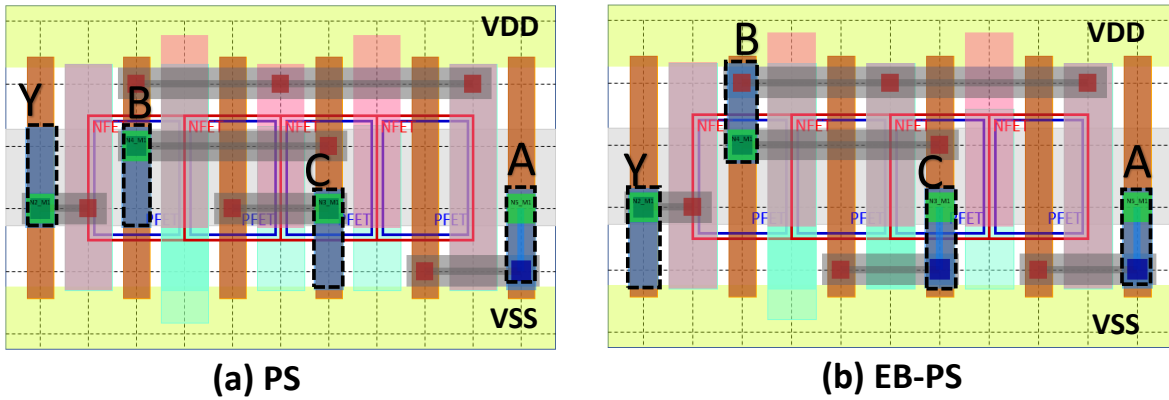
Table 2.2: Experimental results of 30 CFET SDCs without pin-accessibility objective, PS [6] objective and Edge-Based PS (EB-PS) objective under MPO=2 and MPO=3 Constraints: All values are averages, CW = Cell Width, ML = Total Metal Length, #M2Track = the number of used M2 tracks, Min.#PO = the minimum of pin-openings in a cell, Min.RPA = minimum remaining pin access [29], RPA impr. = improvement ratio (EB-PS–PS)/PS, interference distance $d_{int}=2$ M1 pitch $((2MAR + EOL)/2)$, opening mask for MPO = (2EOL + 1MAR)

Constraint	Pin-Access Objective	Cell Metrics				Pin-Accessibility				
		CellWidth	ML	M2 ML	#M2Track	PS []	EB-PS	Min. #PO	Min. RPA	RPA impr. (%)
MPO=2	wo PS/EB-PS	9.30	121.80	1.73	0.20	1.16	4.03	2.00	1.62	22.84%
	PS	9.30	129.47	2.93	0.23	2.83	1.67	2.00	1.83	8.74%
	EB-PS	9.30	129.90	2.60	0.23	3.00	0.77	2.00	1.99	-
MPO=3	wo PS/EB-PS	9.30	123.53	1.80	0.20	1.30	7.27	3.00	2.33	21.46%
	PS	9.30	125.13	2.20	0.20	2.80	2.60	3.00	2.76	2.54%
	EB-PS	9.30	125.27	2.20	0.20	2.93	1.93	3.00	2.83	-

proves and ensures the pin-accessibility by maximizing the advantages of the MPO constraint. Figure 2.15 shows the different I/O pin distributions by PS [6] and the proposed EB-PS objective for an AND3x1 cell with design parameters, MPO=2 and $d_{int}=2$ M1 pitch. The MPL and MPO constraints respectively ensure at-least one M1 metal segment and at-least 2 pin-openings for each I/O pin (depicted in black dashed rectangles). Note that for the RPA value⁷ of a pin less than one, the pin is not likely to be accessed successfully, because we need at least one access point. In Figure 2.15 (a), the pin Y will be in-accessible when there is a parallel pin-shape of adjacent cell as illustrated in Figure 2.7. On the contrary, the EB-PS objective ensures that the pins of SDC can be accessed because EB-PS considers not only the space between pins but also the physical pin shapes within d_{int} .

In Table 2.2, “EB-PS” cases significantly improve the number of accessible pin-openings than “wo PS/EB-PS” by 22.84% and 21.46% with “MPO=2” and “MPO=3”. Furthermore, “EB-PS” cases also increase the number of accessible pin-openings than “PS [6]” by 8.74% and 2.54% using “MPO=2” and “MPO=3”, respectively. The pin-accessibility metrics demonstrate that the MPO constraint successfully ensures the minimum number of pin-openings and the EB-PS objective contributes to maximize the

⁷The RPA of [29] indicates how many access points of a pin remain after the accesses of its neighboring pins.



	PS				EB-PS			
I/O Pin	Y	C	B	A	Y	C	B	A
PS [4] Obj. (Maximize)	0		1	1	0		1	1
EB-PS Obj. (Minimize)	2	0	2	0	1	0	1	0
#Pin Opening	2				2			
RPA	1	2	1	2	2	2	2	2
RPA (Worst Case)	0	2	1	1	1	2	2	1

($MPO = 2, d_{int} = 2$ M1 pitch)

Figure 2.15: Layout of AND3x1 cell optimized generated by Pin Separation [6] and Edge-Based Pin Separation (EB-PS) [7] objectives. The RPA (Worst Case) considers the parallel pin-shape of adjacent cell as described in Figure 2.7.

effective accessible pin-openings. As the MPO increases and the pin-accessibility (i.e., “EB-PS”) is maximized, the average ML and #M2Track are increased around 2% compared to PS [6] due to the enlarged and scattered I/O pin shapes.

Optimization for Routing-Congestion Minimization

We compare the $M2$ routing-resource related metrics between our proposed #M2Track and $M2$ Length objectives, which are discussed in Section. 2.2.5, as described in Table 2.3. “MinTrack” denotes that the SDCs are generated with the objectives including #M2Track (i.e., Expression (2.13)). “MinLength” is $M2$ Length-oriented objectives (i.e., (a)CellSize, (b)PS, (c) $M2$ Length, and (d)TotalML). Compared to “MinLength”, “MinTrack” reduces the $M2$ track usage by 46.67% with 28.00% of increment

in $M2$ length. Figure 2.16 shows the layouts of FA cell that are optimized using $MinTrack$ and $MinLength$ objectives. While both layouts have the same $M2$ metal length, the used $M2$ tracks of $MinTrack$ is 1 less than $MinLength$. As discussed in Section 2.2.5, we expect that $MinTrack$ reduces the routing congestion in block-level more effective than $MinLength$ due to the reduced $M2$ tracks in spite of the increased $M2$ length. We validate this in Section. 2.3.4.

Table 2.3: Experimental Results of CFET SDCs optimized for $MinTrack$ (i.e., Expression (2.13)) and $MinLength$ (i.e., (a) $CellSize$, (b) PS , (c) $M2$ Length, and (d) $TotalML$). $Incr. = (MinTrack - MinLength)/MinLength$, $Red. = (MinLength - MinTrack)/MinLength$.

Cell Name	Cell Layout Objectives					
	M2 Metal Length			#M2 Tracks		
	$MinLength$	$MinTrack$	Incr. (%)	$MinLength$	$MinTrack$	Red. (%)
AOI22x1	10	14	40.00%	2	1	50.00%
OAI22x1	8	12	50.00%	2	1	50.00%
XOR2x1	10	10	0.00%	2	1	50.00%
XNOR2x1	12	18	50.00%	2	1	50.00%
FA	24	24	33.33%	3	2	50.00%
Avg.	12.80	15.60	28.00%	2.20	1.20	46.67%

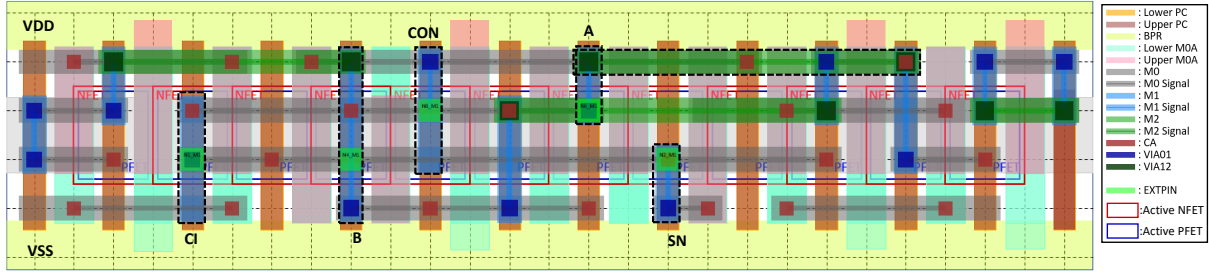
2.3.4 Block-Level Routability Analysis

We validate our framework through a block-level analysis including the #DRV analysis across suggested design features. For BEOL, we use $M2 - M5$ for detailed routing. The block-level analysis setup is as described in Section 2.3.1.

We analyze the routability of three RTL designs using multiple CFET SDC sets that are generated under different split cases of pin-accessibility and routability related constraints and objectives as described in Table 2.4.

Analysis 1 (Pin-accessibility). Figure 2.17 shows the #DRVs trends of wPS and woPS under $MPO=2$ and

(a) *MinTrack Objective*



(b) *MinLength Objective*

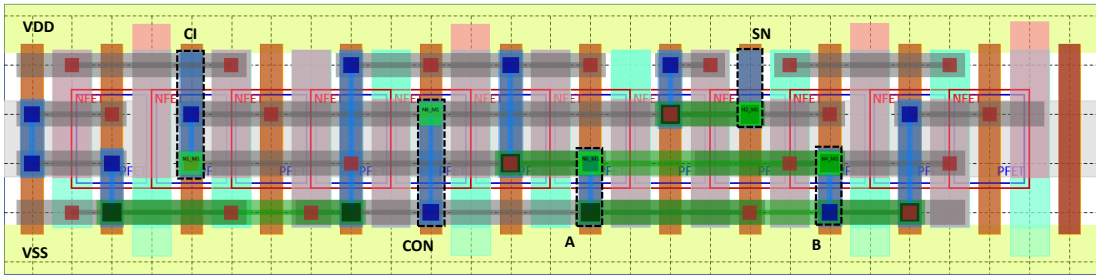


Figure 2.16: Layouts of FAX1 cell optimized with *MinTrack* and *MinLength* objectives. The black dashed rectangles shows the FA I/Os on M1. Note that some signals and I/Os need to be routed with M2 to complete the routing.

Table 2.4: SDC Generation Split Cases for Routability Analysis

Cases	Objectives	Constraints
wPS MPO=2 (resp. 3)	(a)CellSize (b)PS (c)#M2Track (d)TotalML	MPO=2 (resp. 3)
woPS MPO=2 (resp. 3)	(a)CellSize (b)#M2Track (c)TotalML	MPO=2 (resp. 3)
<i>MinLength</i>	(a)CellSize (b)PS (c)M2Min (d)TotalML	MPO=2
<i>MinTrack</i>	(a)CellSize (b)PS (c)#M2Track (d)TotalML	MPO=2
PS [6]	(a)CellSize (b)PS (c)#M2Track (d)TotalML	MPO=3
Proposed (Best setting)	(a)CellSize (b)EB-PS (c)#M2Track (d)TotalML	MPO=3
SP&R [9]	(a)CellSize (b)TotalML	MPO=N/A

3. The #DRVs of MPO=3 cases increases slower than MPO=2 cases as the design utilization increases, because the number of pin-openings are secured as much as the MPO parameter and, also, the Min. RPAs of MPO=3 cases are both 40% larger than MPO=2 cases (Table 2.2). The #DRVs of wPS cases consistently grows slower than woPS cases under the same MPO for all three designs. In particular, the

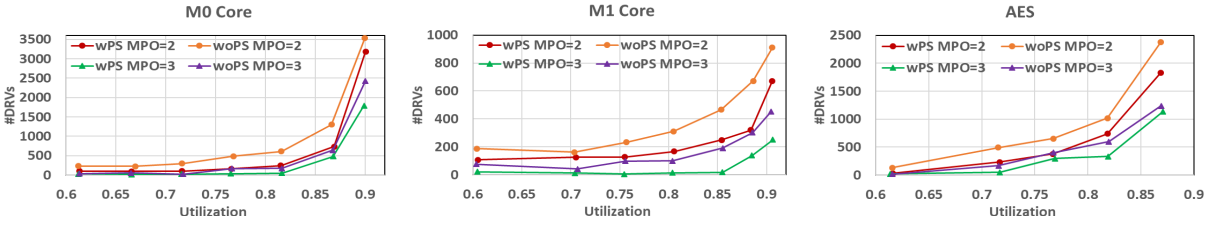


Figure 2.17: Block-Level Placement and Route Results of M0 core, M1 core, and AES designs of wPS and woPS under MPO=2 and MPO=3 constraints.

#DRVs of wPS with MPO=2 and MPO=3 are 44/47/23% and 25/91/8% smaller than woPS with MPO=2 and MPO=3 at the 0.87/0.85/0.87 utilization in M0 Core/M1 Core/AES, respectively. This demonstrates that our MPO constraint and PS objective successfully maximize the effective accessible pin-openings, resulting in the improvement of the routability.

Analysis 2 (Routing Congestion). Figure 2.18 shows the #DRVs trends for *MinTrack* and *MinLength* objectives. The #DRVs of *MinTrack* increases slower than *MinLength* objective for all designs. Specifically, the #DRVs of *MinTrack* are 33/19/12% smaller than #DRVs of *MinLength* at the 0.87/0.85/0.87 utilization in M0 Core/M1 Core/AES, respectively. Table 2.5 shows the pin analysis QoR reports of the commercial place-and-route (P&R) tool [14]. The horizontal congestion of *MinLength* are larger than *MinTrack* from 0.1% up to 2.0% at each design utilization showing obvious #DRVs differences. This validates that *MinTrack* is more effective objective for reducing the routing congestion than *MinLength* objective.

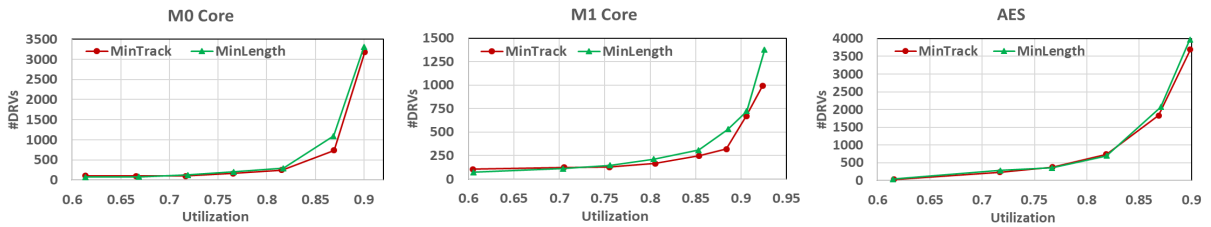


Figure 2.18: Block-Level P&R Results of M0 core, M1 core, and AES designs of *MinTrack* and *MinLength*.

Table 2.5: Pin Analysis QoR Report of *MinTrack* and *MinLength* from [14]

		M0 Core		M1 Core		AES	
Utilization		0.87	0.90	0.91	0.92	0.87	0.90
Horizontal	<i>MinTrack</i>	15.5	35.8	21.9	30.8	28.3	33.5
Congestion (%)	<i>MinLength</i>	16.9	36.3	22.4	32.2	30.3	33.6
Impr.(<i>MinLength</i> - <i>MinTrack</i>)(%)		1.4	0.5	0.5	1.4	2.0	0.1

Analysis 3 (Block-level Routability). Figure 2.19 shows the block-level P&R results of proposed, PS [7], and SP&R [9]. Compared to SP&R, our Proposed case shows 4.2%, 8.1%, and 10.3% improvement (depicted in blue arrows) for the design utilization at 300 #DRVs threshold of M0 Core, M1 Core, and AES. In addition, the Proposed case reduces 84%, 98%, and 58% #DRVs (depicted in blue arrows) of M0 Core, M1 Core, and AES at the utilization when the #DRVs of SP&R starts to exceed 300⁸ threshold line of each design. Moreover, the #DRVs of the proposed cases consistently grows slower than “PS [6]” cases under the same MPO for all three designs. In particular, the proposed case reduces the #DRVs up to 28% compared to PS with MPO=3 at 0.95 utilization in M1 Core.

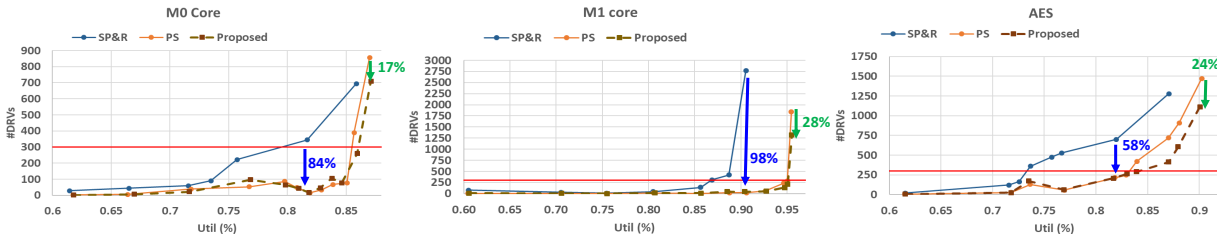


Figure 2.19: Block-level P&R Results of M0 core, M1 core and AES of proposed routability-driven cell optimization, PS [6], and SP&R [9] objectives CFET SDCs.

This validates that our routability-driven constraints and objectives not only reduce the #DRVs but also improve the block-level area scaling. Figure 2.20 shows block-level placement-and-route snapshots and a #DRVs report of M0 Core at 0.82 utilization. The #DRVs of SP&R case is $6.4\times$ larger than our Proposed case. Most of DRVs (depicted in white objects), which are caused by heavy routing congestion

⁸From the industrial guidance, designs with #DRVs smaller than 300 usually can be fix in Engineering Change Order (ECO) stage.

on $M2$ layer and near the $M3$ power stripes, have been successfully reduced in our Proposed case by improved pin-accessibility and optimized routing resources.

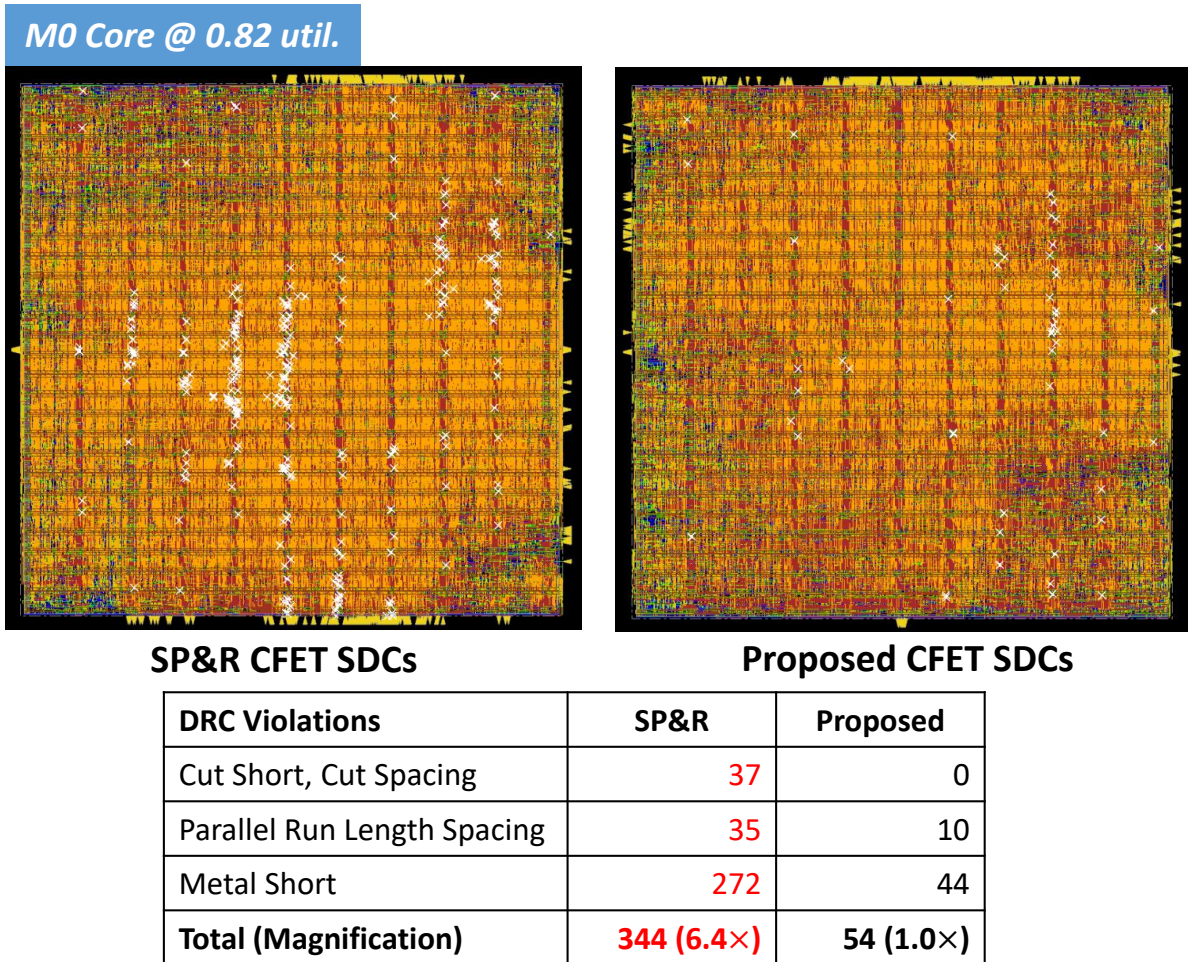


Figure 2.20: P&R design views of M0 core at 0.82 util. with proposed routability-driven CFET SDCs versus SP&R [9] objectives CFET SDCs. The white objects represent DRVs.

2.4 Conclusion

In this chapter, we have introduced a routability-driven CFET standard cell framework using novel Dynamic Complementary Pin Allocation scheme to generate optimum cell layout in terms of cell area, pin-accessibility, routing congestion, and total metal length. For routability, the novel Edge-based

Pin Separation and #M2 Track objectives with Minimum Pin Length and Minimum Pin Opening constraints are implemented and validated with the statistics of cell-level metrics and block-level routability analysis in multiple designs. We demonstrate that CFET cell structure provides 10.1 and 22.2% on average reduced cell width and metal length, respectively, maintaining the scaling advantage of CFET structure, compared to conventional FET structure with 4 in-cell horizontal routing tracks. The block-level routability analysis shows that our routability-driven framework improves 4.2% utilization and reduces 83% routing errors on average over the previous work [9] with 300 #DRVs threshold.

This chapter contains materials from "A routability-driven complimentary-FET (CFET) standard cell synthesis framework using SMT", by Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, and Dongwon Park, which appears in 2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2020; "Complementary-FET (CFET) standard cell synthesis framework for design and system technology co-optimization using SMT", by Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, Bill Lin, and Dongwon Park, which appears in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2021. The dissertation author was the primary investigator and author of these papers.

Chapter 3

Complementary-FET (CFET) Standard Cell Synthesis Framework for Design and System Technology Co-Optimization

3.1 Introduction

As technology continues to scale beyond $7nm$, cell layout scaling of conventional (Conv.) FET structure is limited due to routing congestions, lateral P-N separations, and performance requirements. In addition, design technology co-optimization (DTCO) based on pitch scaling and patterning is starting to reach its limitations in mitigating the cost in 2D IC technology. System technology co-optimization (STCO) is introduced to assist DTCO scaling and bridge the 2D IC technology to novel Complementary-FET (CFET) and 3D integrated logic [4, 38, 39]. CFET technology, which stacks the P-FET on N-FET or vice versa, can release in-cell routing congestion of P-N connection such that SDC designers can continue

cell size reduction in sub- $7nm$ as described in Section 1.1.

Recently, feasible CFET-based SDC layouts have been successfully proposed [3–5]; therefore, CFET has been one of the promising cell structures in sub- $7nm$ and beyond. However, SDC scaling in sub- $7nm$ demands holistic STCO and DTCO explorations on multi-row cell architectures, various cell heights, 3D FET stacking, pin-accessibility, routing congestion, and block-level area due to the limited routing resources and the exploding conditional design rules of later physical design procedures. These explorations for SDC design relies on an automatic multi-row SDC layout synthesis scaling framework, which supports track number reduction, multi-row CFET standard cell architectures, 3D FET stacking, design rule changes, etc.

3.1.1 Related Works

The related works can be categorized into Conv. SDC synthesis automation, and CFET SDC synthesis automation categories.

Conv. SDC Synthesis Automation. For single-row SDC synthesis, several works have reported full automation of cell layout covering transistor-level placement and in-cell routing together [23, 24], but these approaches are not applicable in the multi-patterning technologies in sub- $5nm$. Also, several SDC synthesis automation works have been proposed for multi-patterning technology [25–27], but the placement and routing are performed in separate operations. Recently, in [9], they proposed an approach that integrates the placement and routing with dynamic pin allocation interface using Satisfiability Modulo theories (SMT) [16]. For multi-row SDC synthesis, a minimum width transistor placement method for multi-row structure using SAT has been proposed in [40], but this approach does not guarantee the optimal solution after routing due to the lack of considerations of multi-patterning and design rules. Recently, Y.L. Li et al. [41] developed an entire placement and routing flow for synthesizing multi-row SDCs, but the

placement and routing are performed sequentially and the number of cell rows is not optimized in terms of cell area. These works focus on the Conv. cell structure optimizations, and thus they are not available for CFET cell structure that have stackable P/N-FET.

CFET SDC Synthesis Automation. CFET SDC synthesis framework that performs FET place-and-route concurrently with novel dynamic complementary pin allocation (DCPA) approach have been proposed in [6]. However, these works focus on single-row CFET SDC synthesis, and thus they are not available for multi-row cell area optimization which consider single-row and multi-row structure together and various inter-row routing options (i.e., M0A/PC layers) in multi-row CFET SDC structure.

3.1.2 Our Contributions

In this paper, we develop a Multi-Row CFET SDC synthesis automation scaling framework that supports track number reduction, design rule changes, FET stacking alternatives, and M0A/PC for inter-row routing option for holistic STCO and DTCO explorations using concurrent FET placement and route through a multi-row dynamic pin shape/allocation scheme, resulting in optimized cell layout with optimum number of cell rows, various CFET SDC architectures and design rule selections. Our optimized SDC layout has maximized pin-accessibility and routability through the proposed routability-driven objectives and constraints. Our main contributions are as follows.

- We develop the CFET SDC synthesis scaling framework including concurrent transistor placement and in-cell routing through a novel Dynamic Complementary Pin Allocation (DCPA) scheme to explore CFET SDC scaling with track number reduction, multi-row cell architectures, stacking options (i.e., P-on-N/N-on-P FET), and design rule selections.
- We formulate an integrated constraint satisfaction problem (CSP) for SMT (Satisfiability Modulo theory) solving, including not only place-and-route but also pin-accessibility and design rule related

constraints, resulting in the optimized cell layout across single-row, multi-row and various track number cell architectures.

- We propose a novel multi-row cell area objective to minimize the cell area considering single-row and multi-row structures together.
- We develop Multi-Row Dynamic Complementary Pin Allocation (MR-DCPA) scheme to enable explorations of upper/lower M0A/PC for inter-row routing.
- We demonstrate that our routability-driven objectives and constraints successfully improve the routability through the block-level analysis including the #DRV analysis across suggested design features with various cell track number.
- In STCO studies on 3D stacking, we explore the CFET architecture with P-on-N and N-on-P stacking from 4 routing tracks (RTs) to 2 RTs and compare the results of CFET in the cell and block-level to Conv. structures.
- For DTCO in the cutting-edge technology node, we study the impacts of design rule changes, interaction between design rules and CFET stacking (i.e., P-on-N/N-on-P FET), and the number of back end of lines (#BEOLs) on various CFET architectures and block-level area. In addition, we explore the cell-level metrics and block-level area benefits as reducing 3 routing tracks (RTs) to 2 RTs with/without upper/lower M0A/PC for inter-row routing.

The remaining sections are organized as follows. Section 3.2 describes our Multi-Row CFET SDC synthesis framework for DTCO and STCO explorations. Section 3.3 shows the experimental setup for the CFET standard cell synthesis and block-level analysis for the following experiments. Section 3.4 validates the proposed novel multi-row cell area objective and studies the pin-accessibility constraints

and objectives as SDC scaling from 4.5T to 3.5T. Section 3.5 presents our main experiments for DTCO and STCO explorations. Section 3.6 shows the results of scaling CFET to the extreme 2 RTs CFET architecture. Section 3.7 concludes the paper.

3.2 Multi-Row CFET Standard Cell Synthesis Framework for DTCO and STCO Explorations

We utilize an SMT (Satisfiability modulo theories)-based constraints solving methodology for simultaneous place and route of CFET SDCs. In this section, we describe the detailed features of our framework: (i) Overview of Multi-Row CFET SDC Synthesis Framework, (ii) Multi-Row CFET Cell Architecture, (iii) Multi-Row Dynamic Complementary Pin Allocation, (iv) Parametric Conditional Design Rules, (v) Multi-Row Cell Area Minimization, and (vi) Multi-Objective Optimization.

3.2.1 Multi-Row CFET SDC Synthesis Framework Overview

Figure 3.1 shows the overview of our framework. Given cell netlist and layout specification, our framework generates an integrated constraint satisfaction problem (CSP) for automating CFET SDC layout which strictly satisfies transistor placement, in-cell routing, conditional design rules, and pin-accessibility-driven constraints. Inspired by [6,9], individual constraints are combined by our novel MR-DCPA constraint. Our framework performs routability-driven lexicographic multiple-objective optimization by implementing (i) Multi-Row Cell Area Minimization, (ii) Edge-Based Pin Separation and (iii) M2 Track use objectives, and (iv) Metal Length. We utilize five representative conditional design rules as described in Section 1.3.3, which are *minimum area rule (MAR)*, *end-of-line (EOL)*, *via rule (VR)*, and multi-pattern-aware design rules (i.e., *parallel run-length (PRL)/step height rule (SHR)*). The notations are shown in Table 1.1.

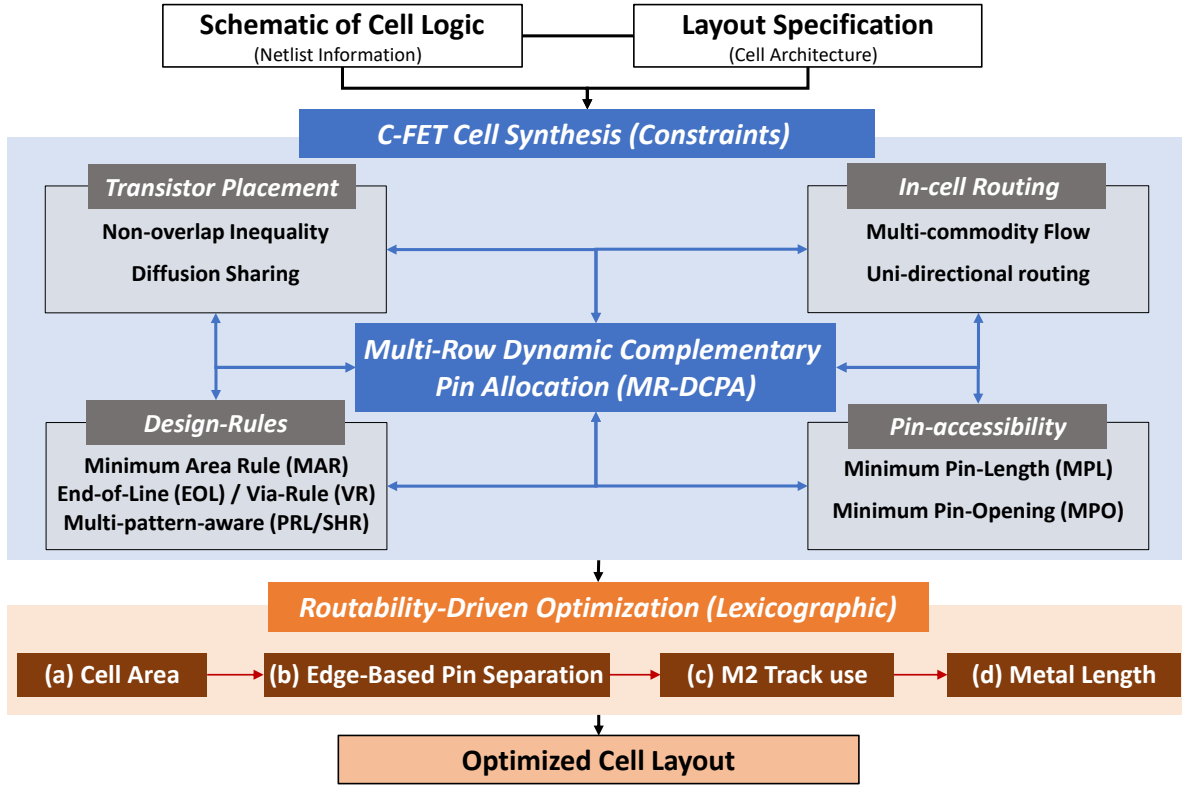


Figure 3.1: Multi-Row CFET Standard Cell Synthesis Framework Overview

3.2.2 Multi-Row CFET Cell Architecture

Our framework employs a CFET cell architecture and netlist information of [3, 5] and [8], respectively. Figure 3.2 shows the grid-based placement and routing graph (i.e., Upper/Lower M0A/PC (DTCO), M0, M1, and M2) using double-row 4 RTs P-on-N CFET example. The routing grid consists of 4 RTs with buried power rails for each cell row and each layer is defined as unidirectional edges. We adopt supernodes [19] for the pin of FET (i.e. internal pin, P_{IN}) or the I/O pin of a standard cell (i.e. external pin, P_{EX}). The P-FET and N-FET regions are stacked up on the upper and lower M0A/PC layers, respectively. Therefore, the access to the M0 layer from each pin on the N-FET region (i.e. lower M0A/PC layer) is restricted to the top or bottom horizontal routing track unless each source/gate/drain pin in P-FET and N-FET that are overlapped on the same vertical track is shared [3, 5] as shown in Figure 2.1. As a

result, there are three kinds of pin shapes according to the sharing status of each pin in stacked FETs as depicted in Figure 2.1. Our framework supports stacking N-FET on P-FET by swapping the FET-related variables, different number of RTs by adjusting h variable, and multi-row CFET SDC structures with R variable as described in (3.1), (3.2), (3.3), and (3.4). For inter-row routing, our framework also supports Upper/Lower M0A/PC routing which is introduced in Section 3.2.3.

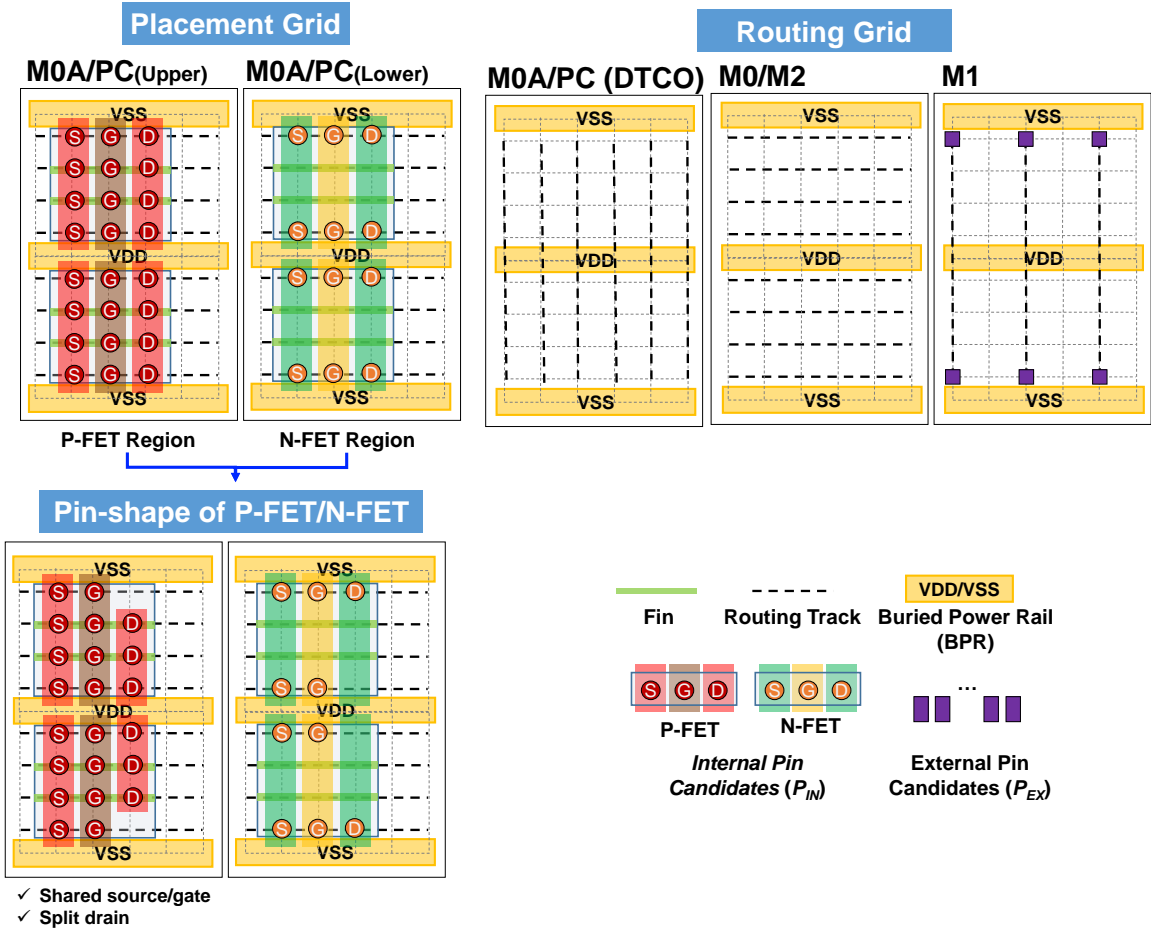


Figure 3.2: Grid-Based placement, routing graph, and pin-shape of P-FET/N-FET using double row 4 RTs P-on-N CFET example.

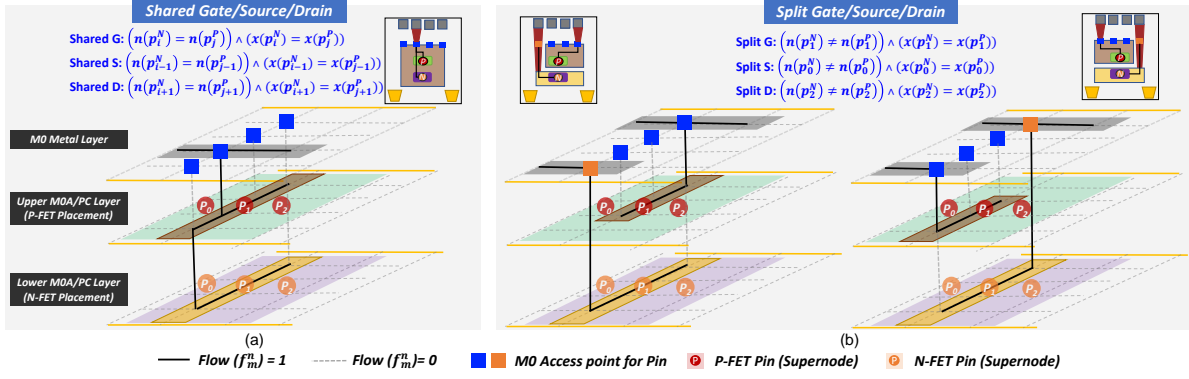


Figure 3.3: Concept of Multi-Row Dynamic Complementary Pin Allocation (MR-DCPA) for 4 RTs P-on-N CFET cell structure. p_1^P =P-FET Gate Pin. p_1^N =N-FET Gate Pin.

3.2.3 Multi-Row Dynamic Complementary Pin Allocation

Multi-Row Dynamic Complementary Pin Allocation (MR-DCPA) dynamically constructs the shared and split pin-shapes of FETs for optimal in-cell and inter-row routing exploration of multi-row CFET structure. The MR-DCPA scheme for simultaneous place-and-route follows the same principle as [9] for interconnecting placement and routing formulas using flow capacity variables (i.e., $C_m^n(v, u)$). Here, We introduce (i) the constraints for shared and split pin-shapes of FETs, and (ii) constraints for Upper/Lower M0A/PC inter-row routing.

Shared and split pin-shapes of FETs

Figure 3.3 illustrates the concept of MR-DCPA using 4 RTs P-on-N CFET as an example. When the pins of P-FET and N-FET are located at the same x-coordinate (i.e., $x(p_i^P) = x(p_i^N)$), the pin-shapes (i.e., shared or split) at the corresponding column in the Upper/Lower M0A/PC layers are determined by the net information. For example, in Figure 3.3 (a), if both of the gate pins p_1^N and p_1^P belong to the same net (i.e., $n(p_1^N) = n(p_1^P)$), a shared pin-shape on Upper/Lower PC layers is selected and one of the corresponding flow variables (i.e., f_m^n) among four possible M0 access points (i.e., blue squares) is determined by the flow formulation. On the other hand, if each gate pin belongs a different net (i.e.,

$n(p_1^N) \neq n(p_1^P)$, MR-DCPA selects one of two possible split pin-shapes (i.e., top or bottom M0 access point for N-FET) as shown in Figure 3.3 (b). Meanwhile, when the upper FET pin has a connection to the power rail (i.e., VDD or VSS), MR-DCPA selects the split pin-shape without blocking the power rail connection of upper FET pin. The expressions of shared and split pin-shapes are shown as follows.

Shared Pin-Shape Expressions: $\bigwedge_{y=y_i^f, \dots, y_i^l} (f_m^n(v_{x,y,l}, v_{x,y+1,l})=1)$

$$l = \{PC^U/M0A^U, PC^L/M0A^L\}, n = n(p_{i,t}^P) = n(p_{j,s}^N), x = x_t^P + i \quad (3.1)$$

Split Pin-Shape Expressions:

Top Access for Lower FET (Type1):

$$\begin{aligned} f_m^{n_1}(v_{x,y_i^l-1,l_1}, v_{x,y_i^l,l_1})=0 \wedge \left(\bigwedge_{y=y_i^f, \dots, y_i^l-2} (f_m^{n_1}(v_{x,y,l_1}, v_{x,y+1,l_1})=1) \right) \\ \wedge \left(\bigwedge_{y=y_i^f, \dots, y_i^l-1} (f_m^{n_2}(v_{x,y,l_1}, p_{j,s}^L)=0) \right) \wedge (f_m^{n_1}(v_{x,y_i^l,l_0}, p_{i,t}^U)=0) \end{aligned} \quad (3.2)$$

Bottom Access for Lower FET (Type2):

$$\begin{aligned} f_m^{n_1}(v_{x,y_i^f,l_1}, v_{x,y_i^f+1,l_1})=0 \wedge \left(\bigwedge_{y=y_i^f+1, \dots, y_i^l-1} (f_m^{n_1}(v_{x,y,l_1}, v_{x,y+1,l_1})=1) \right) \\ \wedge \left(\bigwedge_{y=y_i^f+1, \dots, y_i^l} (f_m^{n_2}(v_{x,y,l_0}, p_{j,s}^L)=0) \right) \wedge f_m^{n_1}(v_{x,y_i^f,l_1}, p_{i,t}^U)=0 \end{aligned} \quad (3.3)$$

No Access for Lower FET (Type3):

$$\bigwedge_{y=y_i^f, \dots, y_i^l-1} (f_m^{n_1}(v_{x,y,l_1}, v_{x,y+1,l_1})=1) \wedge \left(\bigwedge_{y=y_i^f, \dots, y_i^l} (f_m^{n_2}(v_{x,y,l_0}, p_{j,s}^L)=0) \right) \quad (3.4)$$

$$l_0 = PC^L/M0A^L, l_1 = PC^U/M0A^U, n_1 = n(p_{i,t}^U), n_2 = n(p_{j,s}^L), x = x_t^U + i, \quad \begin{cases} U = P, L = N, \text{ if P-on-N} \\ U = N, L = P, \text{ if N-on-P} \end{cases}$$

Algorithm 3 Shared and Split Pin-Shapes Selection

*/*Input: Given $G(V,E)$; Output: MR-DCPA constraints; StackFlag=P-on-N/N-on-P-*/*

```
1: for  $r = 1, 2, \dots, R$  do
2:   Set  $y_i^f = (r-1)h + 1, y_i^l = rh$ ;
3:   if  $(n(p_{i,t}^P) = n(p_{j,s}^N)) \wedge (x(p_{i,t}^P) = x(p_{j,s}^N))$  then
4:     /*Shared Pin-Shape*/
5:     Exp. (3.1) for P-FET and N-FET access.
6:   else if  $(n(p_{i,t}^P) \neq n(p_{j,s}^N)) \wedge (x(p_{i,t}^P) = x(p_{j,s}^N))$  then
7:     /*Split Pin-Shape*/
8:     if  $(StackFlag = P\text{-on-N})$  then
9:       /*P-on-N CFET*/
10:      if  $(n(p_{i,t}^P) = VDD)$  then
11:        /*VDD net at Upper FET pin*/
12:        if  $r\%2=1$  then
13:          Exp. (3.3) for access Lower N-FET.
14:        else if  $r\%2=0$  then
15:          Exp. (3.2) for access Lower N-FET.
16:        end if
17:      else if  $(n(p_{j,s}^N) = VSS)$  then
18:        /*VSS net at Lower FET pin*/
19:        Exp. (3.4) for access Upper P-FET.
20:      else
21:        Exp. (3.2)  $\vee$  Exp. (3.3) for access P-FET and N-FET.
22:      end if
23:    else if  $(StackFlag = N\text{-on-P})$  then
24:      { /*N-on-P CFET*/ }
25:      if  $(n(p_{i,t}^P) = VDD)$  then
26:        /*VDD net at Lower FET pin*/
27:        Exp. (3.4) for access Upper N-FET.
28:      else if  $(n(p_{j,s}^N) = VSS)$  then
29:        /*VSS net at Upper FET pin*/
30:        if  $r\%2=1$  then
31:          Exp. (3.2) for access Lower P-FET.
32:        else if  $r\%2=0$  then
33:          Exp. (3.3) for access Lower P-FET.
34:        end if
35:      else
36:        Exp. (3.2)  $\vee$  Exp. (3.3) for access P-FET and N-FET.
37:      end if
38:    end if
39:  end if
40: end for
```

Algorithm 3 utilizes SMT's *if-then-else* structure to describe a generation procedure of the constraint for shared and split pin-shapes of FETs selection scheme for multi-row structures. For each cell row, the y_i^f and y_i^l are set for corresponding shared and split pin-shapes selection (Lines 1-2). If N-FET and P-FET pins have the same net information, the shared pin-shape is selected (Lines 3-5). Otherwise, the split pin-shape is selected (Lines 6-39). The split pin-shape consists of three types on Upper/Lower M0A/PC layers. *Type1* and *Type2* represent top ($y=h$) and bottom ($y=1$) accesses for lower FET, respectively. If the net of lower FET pin is VSS or VDD, *Type3* is used since there is no connection from M0

to lower FET pin (Lines 19 and Lines 27). When the net of upper FET pin is VDD or VSS, *Type2* is always selected in the odd cell row for P-on-N stacking and even cell row for N-on-P stacking (Lines 13 and Lines 33); *Type1* is always selected in the even cell row for P-on-N stacking and odd cell row for N-on-P stacking (Lines 15 and Lines 31). Otherwise, *Type1* or *Type2*, which satisfies all the constraints and produces the optimal solution, is selected (Lines 21 and Lines 36).

M0A/PC routing constraints

The routing grid is extended to Upper/Lower M0A/PC layers for simultaneous place-and-route using flow capacity variables (i.e., $C_m^n(v, u)$). We consider the interaction between FET pin connection and FET stacking when using Upper/Lower M0A/PC for routing and formulate the following constraints.

Routing constraint I. The Upper/Lower M0A/PC layers at the column in active FET can only be used for routing by the same net of the corresponding FET pin as described in (3.5). Figure 3.4 shows an example of the M0A/PC layers in the active region can only be used for routing the same pin net in the active FET region.

$$\bigwedge_{n \neq n(p^F)} \left(\bigwedge_{y=y_i^f, \dots, y_i^l-1} (f_m^n(v_{x,y,l}, v_{x,y+1,l}) = 0) \right),$$

$$x = x(p^F), \begin{cases} l = PC^U / M0A^U, \text{ if } ((F=P \wedge P\text{-on-N}) \vee (F=N \wedge N\text{-on-P})) \\ l = PC^L / M0A^L, \text{ if } ((F=N \wedge P\text{-on-N}) \vee (F=P \wedge N\text{-on-P})) \end{cases} \quad (3.5)$$

Routing constraint II. If the upper FET pin connects to power rail (i.e., VDD or VSS), the lower layers (i.e., M0A/PC) at the same column can not be used for inter-row routing as described in (3.6) and shown in Figure 3.5.

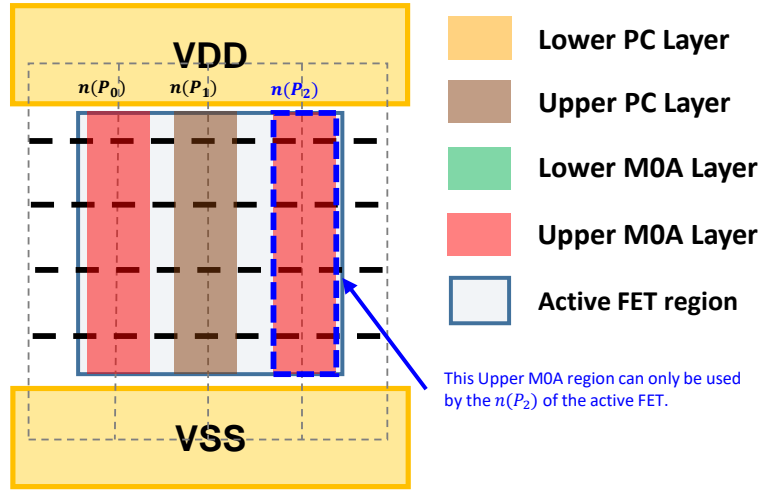


Figure 3.4: An example of M0A/PC routing constraint I: the upper/lower M0A/PC layers in active FET region can only be used for routing by the same net of the corresponding FET pin. Here, the upper M0A of $n(p_2)$ region can only be used for routing $n(p_2)$.

$$\begin{aligned}
 & \bigwedge_{\forall n \in N, n \neq n(p^U)} f_m^n(v_{x,y_i^l,l}, v_{x,y_{i+1}^f,l}) = 0, l = PC^L/M0A^L, \\
 & \text{if } (n(p^U) = n(PR_i)) \wedge (x = x(p^U)) \wedge (y_i^f \leq y(p^U) \leq y_i^l) \quad (3.6)
 \end{aligned}$$

3.2.4 Parametric Conditional Design Rules

We use representative conditional design rules of [9, 21] for EUV and multi-pattern technologies as described in Section 1.3.3. For routing, we consider MAR, VR, and EOL. For multi-pattern technologies (i.e., M0 and M2 layers), we use PRL and SHR for SADP (Self-aligned double patterning) mask [22]. In our framework, all design rules are parameterized by the grid.

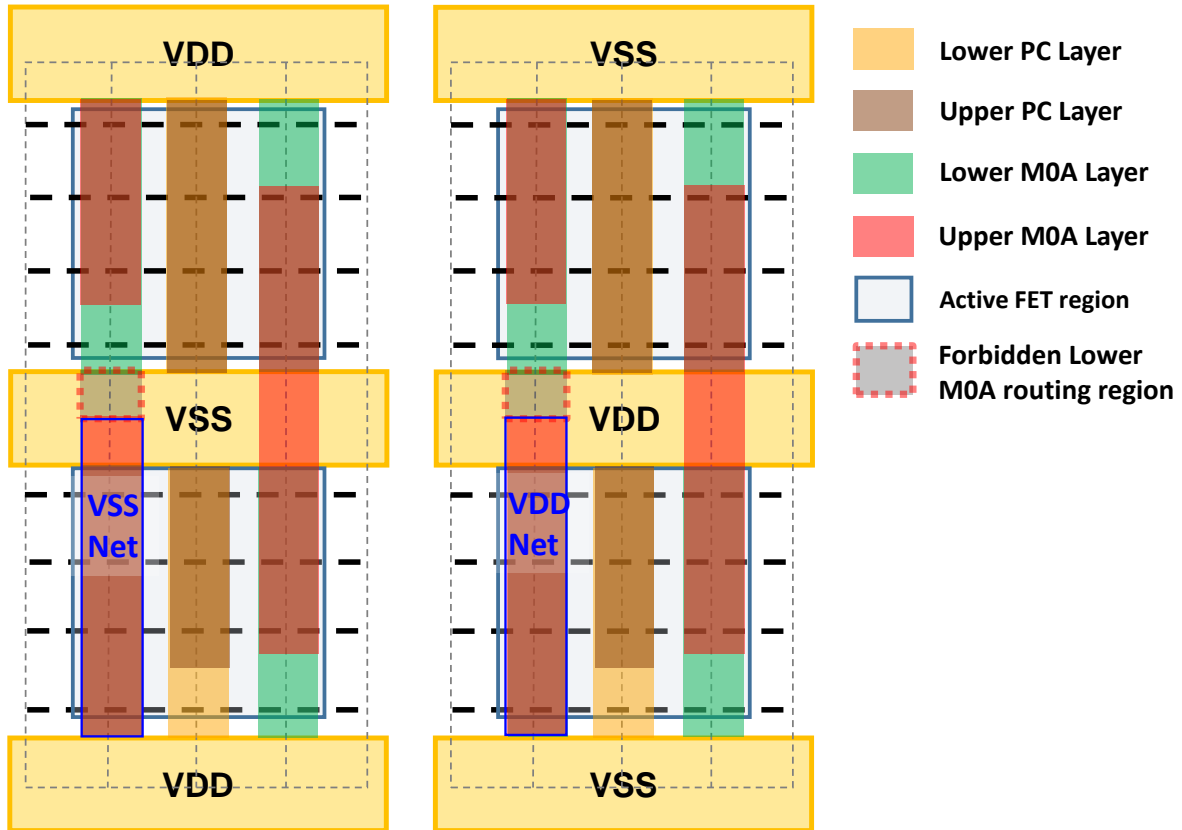


Figure 3.5: Examples of M0A/PC routing constraint II: Lower MOA/PC is forbidden for inter-row routing when the upper M0A/PC connects to VDD/VSS.

Routing Design Rule

The MAR, EOL, and VR examples are shown in Figure 3.6. All the parametric design rule numbers are in grid. The EOL/MAR number defines at least the number of grids need to be satisfied for EOL spacing/metal length. For example, EOL=1 defines that the EOL spacing between two metal segments needs to be at least one grid. The VR number defines that the number of grid between vias needs to be larger than the VR number. As a result, VR=1 allows diagonal via but forbids adjacent via.

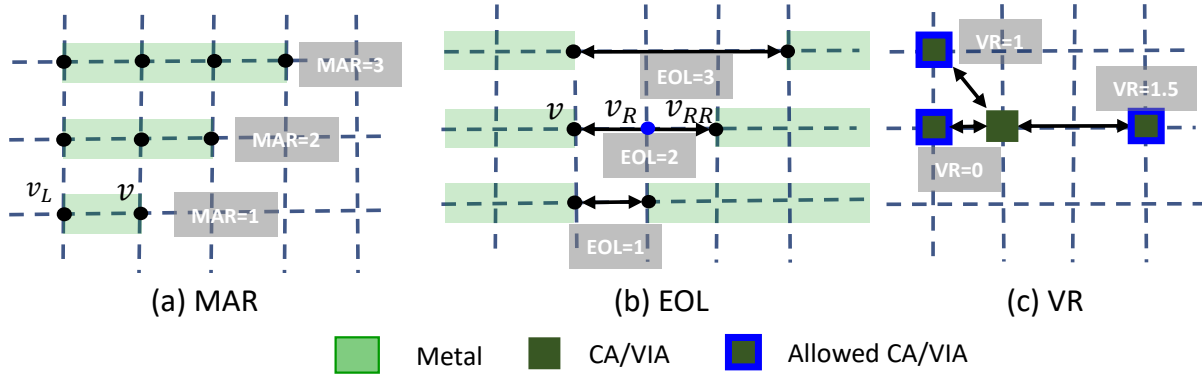


Figure 3.6: Examples of Parametric Design Rules for routing: (a) MAR, (b) EOL, and (c) VR. All the numbers are in grid.

Multi-Patterning Design Rule

We consider the PRL and SHR rules on metal layers (i.e., M0 and M2) by referring to [1]. PRL rule is one of the important rules to avoid “single-point-contact” in manufacturing SADP mask [22]. Figure 3.7 (a) shows an example of parametric PRL rule. SHR is a design rule to avoid “the small step” in manufacturing SADP mask [22]. Figure 3.7 (b) illustrates the SHR when step height is 2.

3.2.5 Multi-Row Cell Area Minimization

We introduce the novel Multi-Row Cell Area Minimization objective, which considers the solutions of single-row and multi-row structures simultaneously and generates the minimum cell area layouts with optimum cell row (Opt. CR). The maximum cell width is defined as the right-most vertical track occupied by the FET among all cell rows as shown in (3.7). Then, if there is any FET be placed in i^{th} cell row or the cell row larger than i , the W_i is set to W_{max} . Otherwise, the W_i is 0 as described in (3.8). With (3.8), we can minimize the cell area with the considerations of single-row and multi-row structures simultaneously.

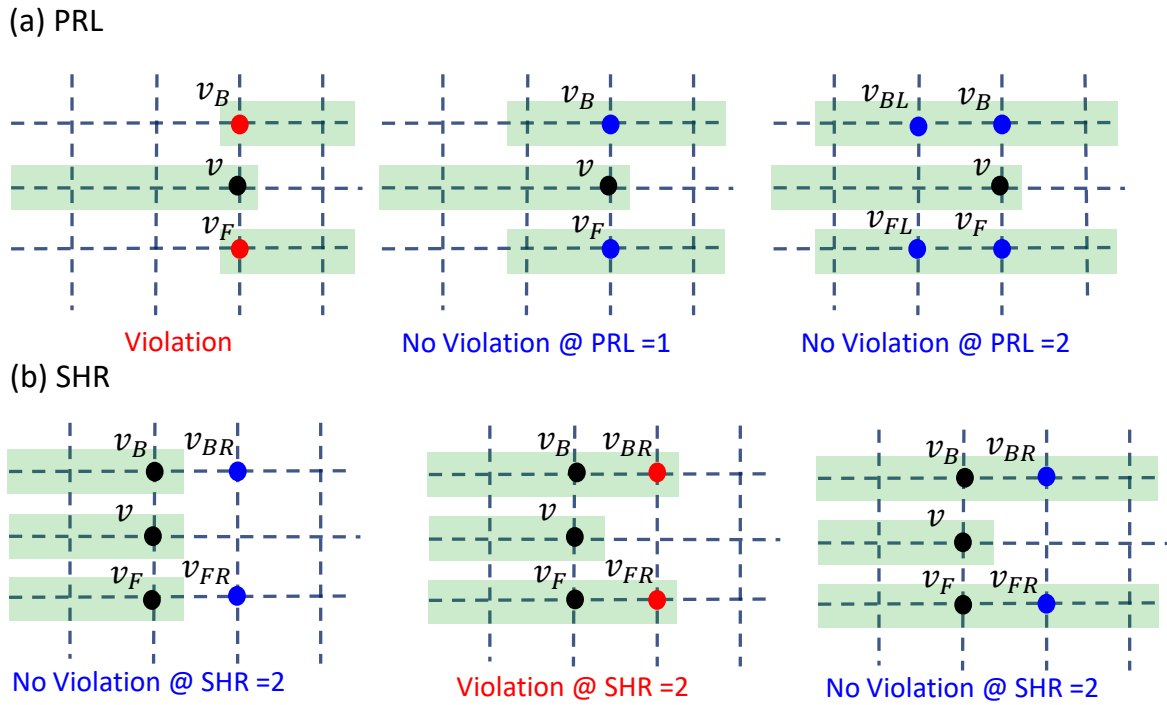


Figure 3.7: Examples of Parametric Design Rules for multi-patterning: (a) PRL and (b) SHR. All the numbers are in grid.

$$W_{max} = \max \{x_t + w_t \mid t \in T\} \quad (3.7)$$

$$W_i = \begin{cases} W_{max}, \text{ if } i = 1 \\ W_{max}, \text{ if } (y_i^f \leq y_t \leq y_i^l), \forall t \in T \\ W_{max}, \text{ if } W_j = W_{max}, \forall j > i \\ 0, \text{ otherwise} \end{cases} \quad (3.8)$$

3.2.6 Multi-Objective Optimization (Optimal Priority)

Our framework has multiple objectives associated with placement and routing problems for standard cell layout design. The first objective is cell area which is defined as the sum of W_i of each cell row as shown in (3.9). The second objective is Edge-Based Pin Separation (EB-PS) [7] and it minimizes the summation of column-based and edge-based pin costs (i.e., $SC(p)$ and $EC(p)$) of each SDC I/O pin in (3.10). $SC(p)$ is 1 if there are adjacent pins within an interference distance d_{int} . $EC(p)$ is the summation of adjacent parallel pin shapes within d_{int} . The third objective is the number of $M2$ tracks used for in-cell routing in (3.11) [6]. The last objective is the weighted sum of routed metal segments and vias (i.e., Total Metal Length (ML)) as shown in (3.12). In practice, the cell size has the highest priority because it has a direct impact on the area of a whole chip. The EB-PS should be considered as the second objective because the in-accessible pins can not be routed regardless of the routing resources [6]. Then the number of $M2$ tracks has been used as a more important metric than Total ML to maximize the routability by reserving upper routing resources. Therefore, our framework simultaneously optimizes these multiple objectives based on addressed “lexicographic” order in (3.13) through an optimization feature of OMT [16].

$$\text{Minimize: Multi-Row Placement (Cell Area)} = \sum_{i=1, \dots, R} W_i \quad (3.9)$$

$$\text{Minimize: Pin-accessibility (EB-PS)} = \sum_{p \in P_{EX}} SC(p) + EC(p)$$

$$SC(p) = \bigvee_{e_{v,q} \in E_k^{M1}, k \in d_{int}(x(p)), q \in P_{EX}, q \neq p} e_{v,q}^{n(q)} \quad (3.10)$$

$$EC(p) = \sum_{e_{v,u}^{n(p)} e_{v,u}^n \in E_k^{M1}, k \in d_{int}(x(p)), n \in N_{EX}, n \neq n(P)}$$

$$N_{EX} = \{n(p) | p \in P_{EX}\}$$

$$\text{Minimize: Routability (\#M2 Track)} = \sum_{k=1}^h \bigvee_{e_{v,u} \in E_k^{M2}} m_{v,u} \quad (3.11)$$

$$\text{Minimize: Total Metal Length} = \sum_{e_{v,u} \in E} (w_{v,u} \times m_{v,u}) \quad (3.12)$$

Lexicographic Optimization: (3.13)

(a) *CellSize*, (b) *EB-PS*, (c) *#M2Track*, (d) *TotalML*

3.3 Experimental Setup

Our framework is implemented in Perl/SMT-LIB 2.0 standard-based formula and executed on a workstation with 2.4GHz Intel Xeon E5-2620 CPU and 256GB memory. The single-threaded SMT Solver Z3 [16] (version 4.8.5) is used to produce the optimized solution in the proposed framework.

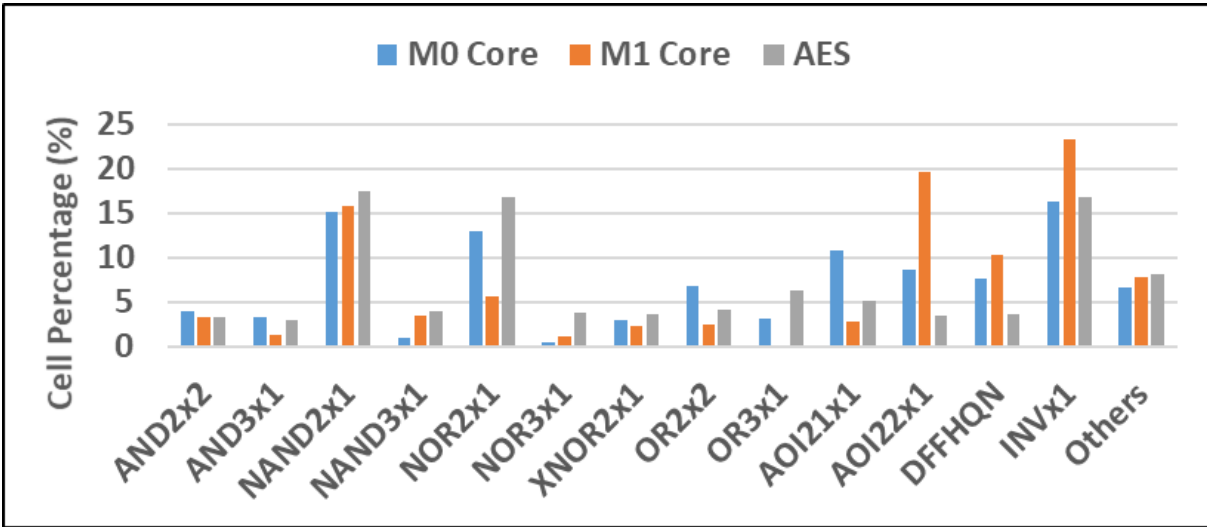


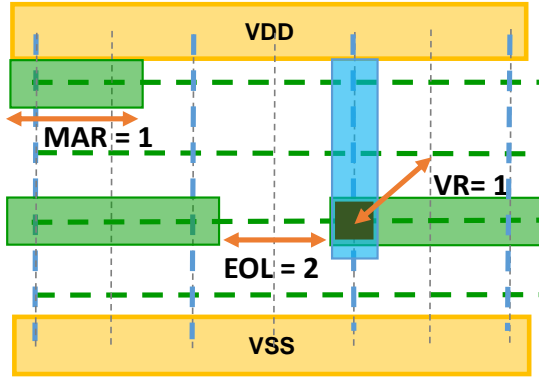
Figure 3.8: Cell Statistics of M0 Core, M1 Core, and AES.

SDC Generation: We use ASAP7 [8] SDC SPICE netlists as inputs of CFET SDCs. We adopt the same number of fingers from [8] for SDC layout generation in the following experiments. To evaluate the

Conditional Design Rules

M2: $MAR = 1/EOL = 2$

VIA12: $VR = 1$



M1 Pitch=42nm, Width=21nm / M2 Pitch=24nm, Width=12nm

■ M1 ■ M2 ■ VIA12 - - - M2 Routing Track - - - M1 Routing Track

Block-level Design Rules

M2: $MAR = 12 \cdot (21 + 12) \text{ nm}^2$

$EOL = 42 - 12 = 30 \text{ nm}$

VIA12: VR (Center to Center)

$= \text{floor}(\sqrt{24^2 + 21^2}) = 31 \text{ nm}$

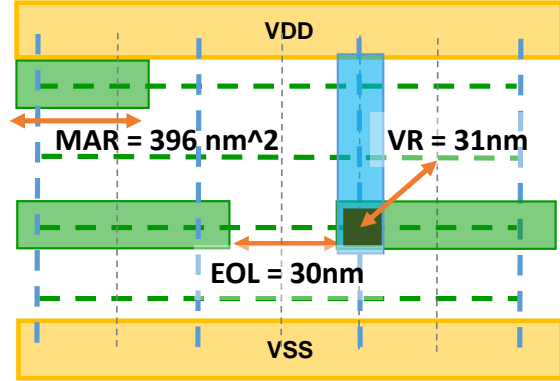


Figure 3.9: An example of transferring the grid-based conditional design rules to the block-level.

block-level PPA in early DTCO exploration, we select 30 representative SDCs [4], which are specified in Table 3.3, for all experiments. The number of FETs in each cell varies from 2 to 24. For standard cell architecture in the experiments, we generate 4.5T, 3.5T, and 2.5T CFET SDCs with 4, 3, and 2 RTs through our framework, respectively. Here, the Metal Length (ML) is calculated by the weighted sum of the via and metal grid as shown in Expression (3.12). The weightings of via are $4 \times$ metal grid considering the parasitic resistance [42]. The baseline parameters of conditional design rules [21], which are described in Section 1.3.3, are as follows: $MAR/EOL/VR/PRL/SHR = 1/2/1/1/2$.

Block-level P&R: Three open source RTL designs [33], M0 Core, M1 Core, and AES that respectively have 17K, 20K, and 14K instances are adopted⁹. The cell statistics of each design are listed in Figure 3.8.

We perform the block-level analysis through a Place-and-Route suite [14].

⁹The worst negative slacks of M0 Core, M1 Core, and AES are carefully adjusted between 50 and -50ps for a fair comparison in the block-level analysis.

For BEOL, we set the contacted poly pitch (CPP), M0/M2 pitch¹⁰ and the number of masks for each BEOL layer according to [1]. For M1, VIA12, and M2 layers, the grid-based conditional design rules' parameters are applied at block-level as shown in Figure 3.9. The metals' pitch and width of layers above M2 are set based on reference [34]. The power delivery network consists of top power meshes (M8 and M9), intermediate power stripes (M3), and standard cell rails (BPR). The top power mesh is designed as spaces is allowed. Then, the power is delivered from M3, which is 4× wider than signal wires, to M1 and M1 to BPR using stacked vias and SuperVia models [35], respectively. The M3 power stripes for the BPR (Buried Power Rail) standard cell rail are placed per every 64 CPPs [36]. We use 300 #DRVs threshold¹¹, which is depicted in red horizontal line in the figures representing the block-level P&R results, to measure the valid block-level area.

The experiments are organized as follows:

- **Exp. 3.4.** *Multi-Row Routability-Driven CFET Cell Optimization:* We firstly demonstrate SDC design with adaptive cell row number can reduce the cell area compared to SDC design with a fixed cell row number. Then, we discuss that edge-based pin separation (EB-PS) objective can further improve the routability of CFET SDC layouts when scaling 4.5T to 3.5T cell height using cell metrics and block-level analysis with baseline design rules.
- **Exp. 3.5.** *DTCO and STCO Exploration for CFET SDC Scaling:* We firstly explore CFET and Conv. SDC architectures using baseline design rules for system technology co-optimization (STCO). Secondly, we vary #BEOLs and design rule parameters, which are perturbed from the baseline, for exploring their impacts on SDCs and block-level area. Then, we exploit the design rules and #BEOLs to maximize the area benefits of cell height reduction in the block-level.

¹⁰The M0/M2 pitches and widths are 24nm and 12nm with 2 masks. The CPP and M1 pitch are 42nm.

¹¹As a common industrial practice, once the number of DRVs increases beyond 300, the block layout is deemed too troublesome to fix with laborious engineering change orders (ECOs).

- **Exp. 3.6. Extreme CFET SDC Scaling:** We compare the cell area, metal length, #Vias, and #M2 Track with/without Upper/Lower M0A/PC for inter-row routing as scaling 3.5T to 2.5T CFET structure using adaptive cell row number for cell area minimization. Then, we explore the minimum valid block-level areas of M0 Core, M1 Core, and AES with 300 #DRVs threshold for 3.5T CFET, and 2.5T CFET with/without Upper/Lower M0A/PC routing SDCs.

3.4 Multi-Row Routability-Driven CFET Cell Optimization

We demonstrate SDC design with adaptive cell row number can reduce the cell area compared to SDC design with a fixed cell row number using the novel multi-row cell area minimization objective in Section 3.2.5. Then, we discuss that the proposed EB-PS can further improve the routability as scaling 4.5T to 3.5T cell height using cell metrics and block-level analysis with baseline design rule.

3.4.1 Cell Area Minimization with Adaptive Cell Row Number

In this section, we compare the SDC areas of adaptive cell row number, triple-row (TR), double-row (DR), and single-row (SR) [7] in 2.5T CFET cell structure. For demonstrating the cell area benefit of adaptive cell row SDC structure while synthesizing each SDC, we use objective (3.9) and equation (3.8) to generate the minimum cell area with optimum cell row (Opt. CR) in our framework. The TR, DR, and SR only considers TR, DR and SR architectures, respectively, during synthesizing the SDC layout. Note that TR and DR SDCs are generated by adding $W_3=W_{max}$ and $W_2=W_{max}$ constraints, respectively.

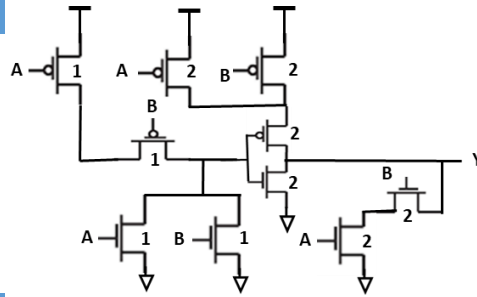
Table 3.1 depicts the SDC comparison results of triple-row (TR), double-row (DR), single-row (SR) [7] and Opt. CR in 2.5T CFET cell structure. Compared to TR, DR and SR [7] cell structures, the average SDC cell areas are reduced by 20.69%, 8.37% and 3.33%, respectively, with the proposed multi-row cell area minimization objective. The FA, XOR2x1, and XNOR2x1 can not be generated by

SR due to the severe in-cell routing congestion from [7]. Figure 3.10 shows XOR2x1 layouts of Opt. CR achieve 22% smaller cell area than TR structure. From the results, the proposed multi-row cell area objective successfully generates the minimum cell area with adaptive cell row number compared to fixed cell row number (i.e., TR, DR, and SR [7]).

Table 3.1: Experimental statistics of 2.5T CFET with Triple-Row (TR), Double-Row (DR), Single-Row (SR) [7] and Optimum Row (Opt. CR). CR=Cell Row; CW=Cell Width; Cell Area Impr. = $((CW \times TR/DR/SR - CW \times \text{Opt. CR})/CW \times TR/DR/SR)$; FA, XOR2x1 and XNOR2x1 can not be generated by SR [7] in 2.5T cell structure due to the severe in-cell routing congestion.

Cell Specification			Cell Layout Objectives							
Name	#FET	#Net	Cell Width (CPPs)				CR	Cell Area Impr (%)		
			TR	DR	SR [7]	Opt. CR	Opt. CR	TR	DR	SR [7]
AND2x2	6	7	4	4	6	6	1	50.00	25.00	0.00
AND3x1	8	9	4	4	7	7	1	41.67	12.50	0.00
AND3x2	8	9	4	5	8	8	1	33.33	20.00	0.00
AOI21x1	6	8	5	7	15	7	2	6.67	0.00	6.67
AOI22x1	8	10	7	9	20	9	2	14.29	0.00	10.00
DFFHQN	24	17	7	10	24	10	2	4.76	0.00	16.67
FA	24	17	7	9	N/A	9	2	14.29	0.00	N/A
NAND3x1	6	8	5	8	14	14	1	6.67	12.50	0.00
NAND3x2	6	8	9	14	27	27	1	0.00	3.57	0.00
NOR3x1	6	8	5	8	14	14	1	6.67	12.50	0.00
NOR3x2	6	8	9	14	26	26	1	3.70	7.14	0.00
OAI21x1	6	8	5	7	15	7	2	6.67	0.00	6.67
OAI22x1	8	10	7	9	20	9	2	14.29	0.00	10.00
OR2x2	6	8	4	4	6	6	1	50.00	25.00	0.00
OR3x1	8	9	4	4	7	7	1	41.67	12.50	0.00
OR3x2	8	9	4	5	8	8	1	33.33	20.00	0.00
XNOR2x1	10	9	6	7	N/A	7	2	22.22	0.00	N/A
XOR2x1	10	9	6	7	N/A	7	2	22.22	0.00	N/A
Avg.	9.11	9.50	5.67	7.50	12.06	10.44	1.44	20.69	8.37	3.33

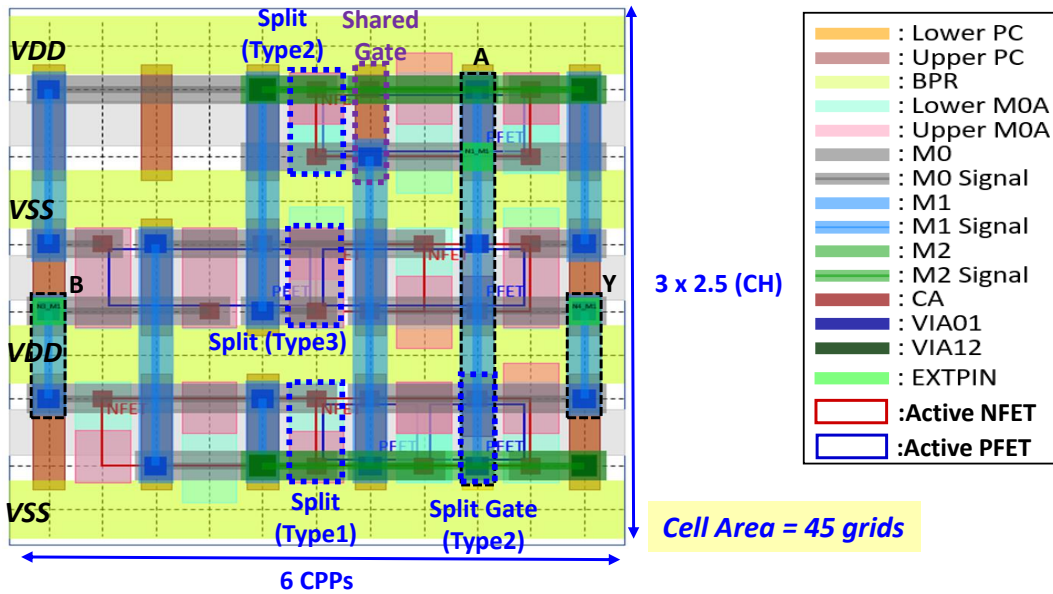
XOR2x1 Schematic Netlist:



1 -> 1 finger
2 -> 2 fingers

XOR2x1 CFET SDC Layout:

(a) Triple-Row cell structure



(b) Optimum Cell Row cell structure -> 22% Smaller Cell Area

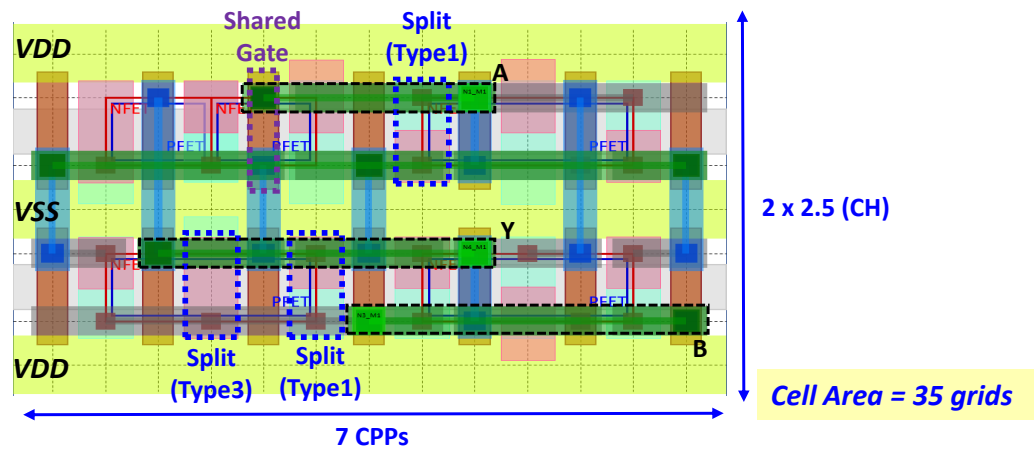
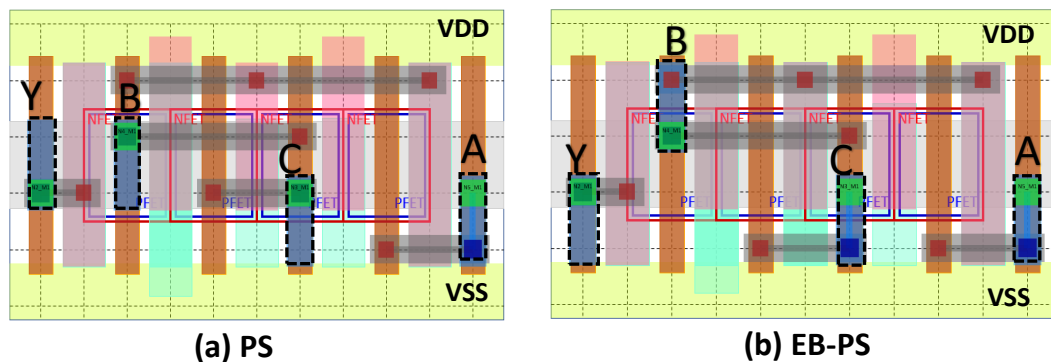


Figure 3.10: An example of XOR2x1 schematic netlist [8] and SDC layouts of (a) Triple-Row and (b) Optimum Cell Row cell structures.

3.4.2 Routability-Driven Cell Optimization for Scaling

In this section, we validate our routability-driven constraints and objectives using the statistics of cell metrics, pin-accessibility metrics, and #M2 track with multiple CFET SDC sets generated by our framework using baseline design rules. Moreover, we further validate our routability-driven constraints and objectives in block-level with M0 Core, M1 Core, and AES [33] designs.



	PS				EB-PS			
I/O Pin	Y	C	B	A	Y	C	B	A
PS [4] Obj. (Maximize)	0		1	1	0		1	1
EB-PS Obj. (Minimize)	2	0	2	0	1	0	1	0
#Pin Opening	2				2			
RPA	1	2	1	2	2	2	2	2
RPA (Worst Case)	0	2	1	1	1	2	2	1

($MPO = 2, d_{int} = 2$ M1 pitch)

Figure 3.11: Layout of AND3x1 cell optimized generated by Pin Separation [6] and Edge-Based Pin Separation (EB-PS) [7] objectives. The RPA (Worst Case) considers the parallel pin-shape of adjacent cell as described in Figure 2.7.

CFET SDC Pin-Accessibility Optimization

We demonstrate that our EB-PS objective efficiently improves and ensures the pin-accessibility by maximizing the advantages of the MPO constraint. Figure 3.11 shows the different I/O pin distributions by PS [6] and the proposed EB-PS objective for an AND3x1 cell with design parameters, $MPO=2$ and $d_{int}=2$ M1 pitch. The MPL and MPO constraints respectively ensure at-least one M1 metal segment and

at-least 2 pin-openings for each I/O pin (depicted in black dashed rectangles). Note that for the RPA value¹² of a pin less than one, the pin is not likely to be accessed successfully, because we need at least one access point. In Figure 3.11 (a), the pin Y will be in-accessible when there is a parallel pin-shape of adjacent cell as illustrated in Figure 2.7. On the contrary, the EB-PS objective ensures that the pins of SDC can be accessed because EB-PS considers not only the space between pins but also the physical pin shapes within d_{int} .

Table 3.2: Experimental results of 30 4.5T and 3.5T CFET SDCs with PS [6] and Edge-Based PS (EB-PS) under MPO=2 and MPO=3 Constraints: All values are averages, CW = Cell Width, ML = Total Metal Length, #M2Track = the number of used M2 tracks, Min.#PO = the minimum of pin-openings in a cell, Min.RPA = minimum remaining pin access [29], RPA impr. = improvement ratio (EB-PS – PS)/PS, interference distance $d_{int}=2$ M1 pitch $((2MAR + EOL)/2)$, opening mask for MPO = $(2EOL + 1MAR)$

Cell Height	Settings	Cell Metrics			Pin Accessibility		
		CW	ML	#M2Track	Min. #PO	Min. RPA	RPA impr. (%)
4.5T	MPO=2 PS	9.30	126.97	0.20	2.00	1.83	8.74%
	MPO=2 EB-PS	9.30	127.60	0.20	2.00	1.99	
	MPO=3 PS	9.30	130.26	0.20	3.00	2.76	2.54%
	MPO=3 EB-PS	9.30	130.37	0.20	3.00	2.83	
3.5T	MPO=2 PS	10.30	204.47	1.03	2.07	1.75	13.14%
	MPO=2 EB-PS	10.30	208.93	1.03	2.07	1.98	
	MPO=3 PS	10.30	212.23	1.27	3.00	2.73	3.66%
	MPO=3 EB-PS	10.30	215.73	1.27	3.00	2.83	

We show the comparison of key metrics for the split cases of MPO constraint and pin-accessibility objective in Table 3.2. “MPO=2 PS” denotes that the SDCs are generated using PS [6] objective for pin-accessibility (i.e., (a)CellSize, (b)PS [6], (c)#M2Track, and (d)TotalML) and “MPO=2 EB-PS” refers to the objectives with the proposed EB-PS (i.e., the same objectives with Expression (3.13)). The Min.#PO values, which refers to the minimum pin-openings in each SDC, shows that our MPO constraint successfully ensures required #PO. The Min.RPA value represents the minimum accessible pin-openings of generated CFET SDC layouts. “EB-PS” cases improve the number of accessible pin-openings than

¹²The RPA of [29] indicates how many access points of a pin remain after the accesses of its neighboring pins.

“PS [6]” by 8.74% and 2.54% for 4.5T CFET SDCs and by 13.14% and 3.66% for 3.5T CFET SDCs using “MPO=2” and “MPO=3”, respectively. The pin-accessibility metrics demonstrate that the MPO constraint successfully ensures the minimum number of pin-openings and the EB-PS objective contributes to maximize the effective accessible pin-openings. As the MPO increases and the pin-accessibility (i.e., “EB-PS”) is maximized, the average ML and #M2Track are increased around 2% compared to PS [6] due to the enlarged and scattered I/O pin shapes.

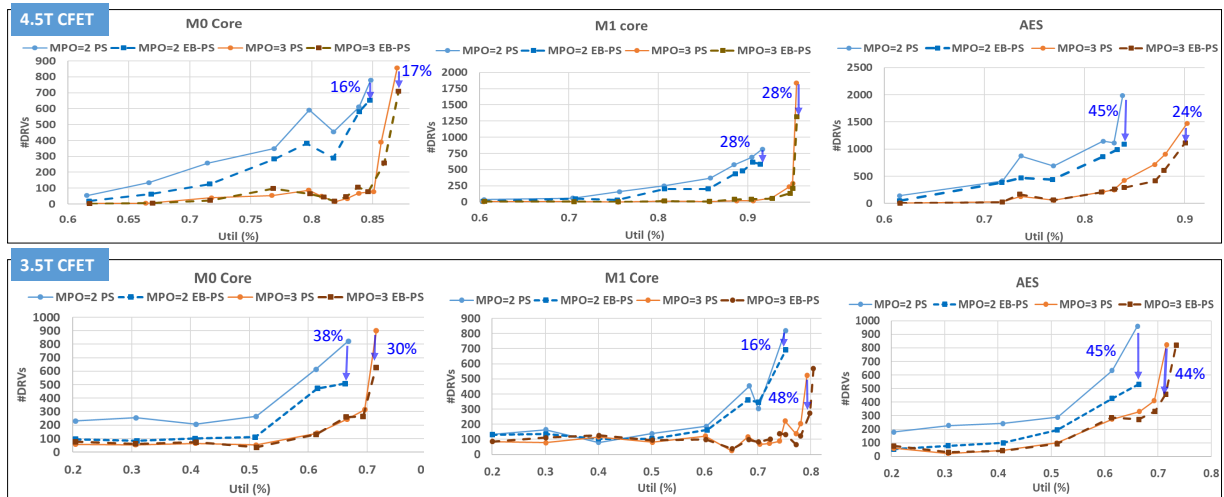


Figure 3.12: Block-level P&R Results of M0 Core, M1 Core, and AES designs of with “PS [6]” and “EB-PS” under MPO=2 and MPO=3 constraints.

Block-level Validation

We analyze the routability of CFET SDC sets in Table 3.2 with block-level designs using M2 to M5 routing layers. Figure 3.12 shows the #DRVs trends of “PS [6]” and “EB-PS” under MPO=2 and MPO=3. The #DRVs of MPO=3 cases increase slower than MPO=2 cases as increasing the design utilization because the number of pin-openings are secured as much as the MPO parameter and the Min. RPAs of MPO=3 cases are both 40% larger than MPO=2 cases (Table 3.2). The #DRVs of “EB-PS” cases consistently grows slower than “PS [6]” cases under the same MPO for all three designs across 4.5T

and 3.5T CFET SDC structures. In particular, EB-PS reduces the #DRVs up to 48%¹³ compared to PS with MPO=3 at 0.79 utilization in M1 Core using 3.5T CFET SDCs. The block-level results show that our MPO constraint and “EB-PS” objective successfully maximize the effective accessible pin-openings, resulting in the improvement of the routability.

Figure 3.13 shows P&R snapshots and a #DRVs report of M0 Core at 0.72 utilization with MPO=3 with 3.5T CFET cell architecture. The #DRVs of PS [6] objective is increased by 44% compared to our proposed EB-PS objective [7]. Most of the DRVs (depicted in white objects), which are caused by heavy routing congestion on M2 layer and near the M3 power stripes, have been successfully reduced with “EB-PS” by further mitigating the interference of physical pin shapes.

3.5 DTCO and STCO Exploration for CFET SDC Scaling

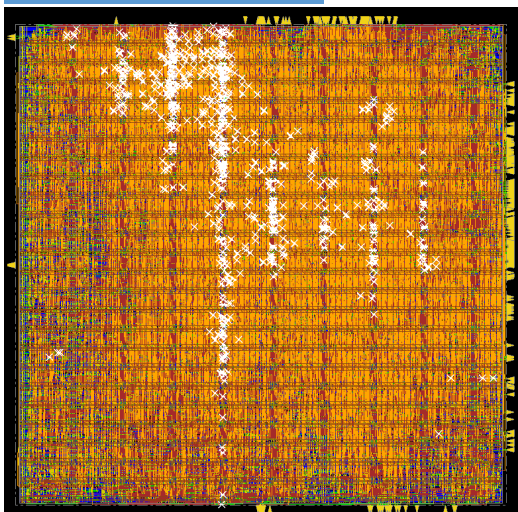
In this section, we firstly explore CFET and Conv. SDC architectures using baseline design rules for system technology co-optimization (STCO). Secondly, we vary #BEOLs and design rule parameters, which are perturbed from the baseline, for exploring their impacts on SDCs and block-level area. Lastly, we exploit the design rules and #BEOLs to maximize the area benefits of cell height reduction in the block-level.

3.5.1 CFET vs. Conv. SDC

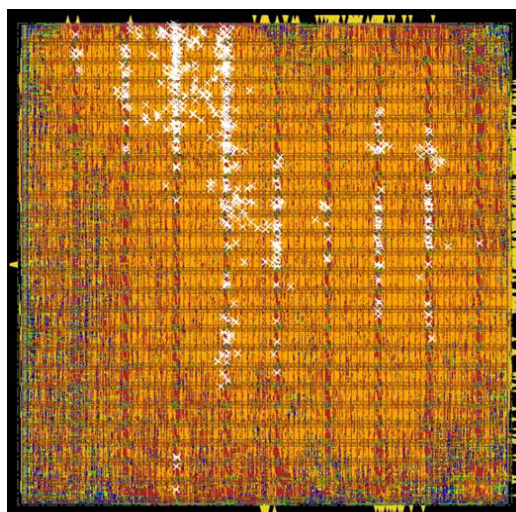
In this section, we firstly compare CFET SDC layouts with the Conv. SDC layouts [9] for the 4.5T and 3.5T cell structures. Then, we explore the block-level area benefits of the CFET cell architecture. For a fair comparison, we adopt the same in-cell horizontal RTs (i.e., 4 and 3 RTs), baseline design rules, and MPO=3 constrain. Also, we push the SDC power rail to BPR layer for Conv. SDC structure [37].

¹³We pick the maximum utilization in Figure 3.12 for comparison to show the major differences in the trend between PS [6] and the proposed EB-PS.

M0 Core @ 0.72 util.



3.5T CFET SDCs with PS



3.5T CFET SDCs with EB-PS

DRC Violations	PS	EB-PS
Cut Short, Cut Spacing	57	17
Parallel Run Length Spacing	35	10
Metal EOL Spacing	213	127
Metal Short	632	498
Total (Magnification)	937 (44% ↑)	652 (1.0X)

Figure 3.13: P&R design views of M0 core at 0.72 util. with the proposed EB-PS objective [7] 3.5T CFET SDCs versus PS objective [6] 3.5T CFET SDCs. The white objects represent DRVs.

Cell-Level Comparison: Table 3.3 depicts the comparison results of CFET and Conv. SDCs in 4.5T and 3.5T cell structures. The average runtime per cell is less than 12 minutes. Compared to Conv. SDCs, CFET achieves 10.94%, 21.27%, and 16% reduction on the average cell width, metal length (ML), and #M2 track, respectively, as scaling to 3.5T cell structure. Figure 3.14 shows the netlist and the generated P-on-N and N-on-P CFET cell layouts of XOR2x1. The shared and split pin-shapes have been successfully selected by our DCPA scheme. Figure 3.15 shows design rule corrected DFFHQN cell layouts for Conv. and CFET architectures in 4.5T and 3.5T. All metal segments that are depicted in red dashed rectangles are successfully extended to satisfy conditional design rules such as PRL and SHR. By virtue of the direct

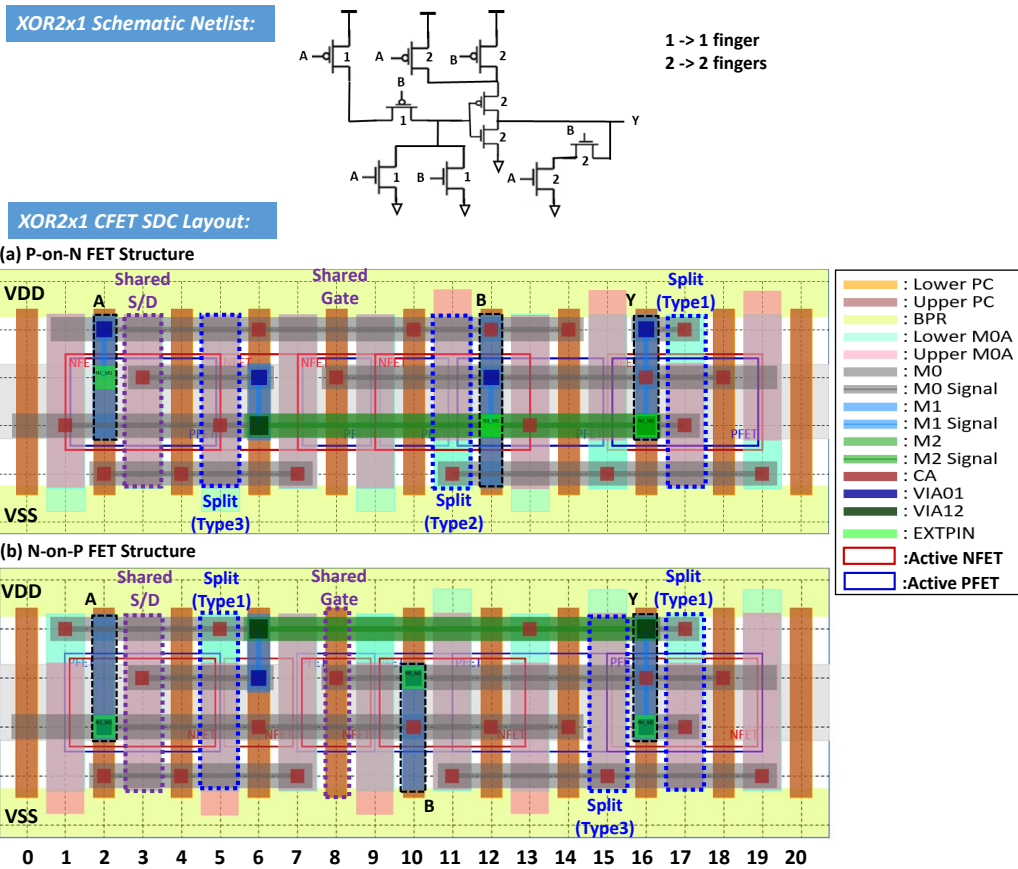


Figure 3.14: An example of XOR2x1 schematic netlist [8] and P-on-N and N-on-P CFET SDC Layout.

P-N connection and more FET terminal access points, the CFET consumes less routing resources (up to 71% less metal length and 2 less #M2Track) and achieves up to 6 CPPs smaller cell width than Conv. structure.

When the SDC cell height is reduced from 4.5T to 3.5T, the average cell area is reduced from 43.79 grids (i.e., Cell Width×Cell Height) to 41.65 for Conv. cell structure and reduced from 41.99 to 36.05 for CFET cell structure. CFET provides around 9% more on average reduced cell area. Figure 3.15 shows that CFET reduces 11% more cell area than Conv. cell structure as scaling 4.5T to 3.5T for DFFHQN SDC.

Comparing P-on-N with N-on-P CFET structures, the cell width and #M2 Track are the same.

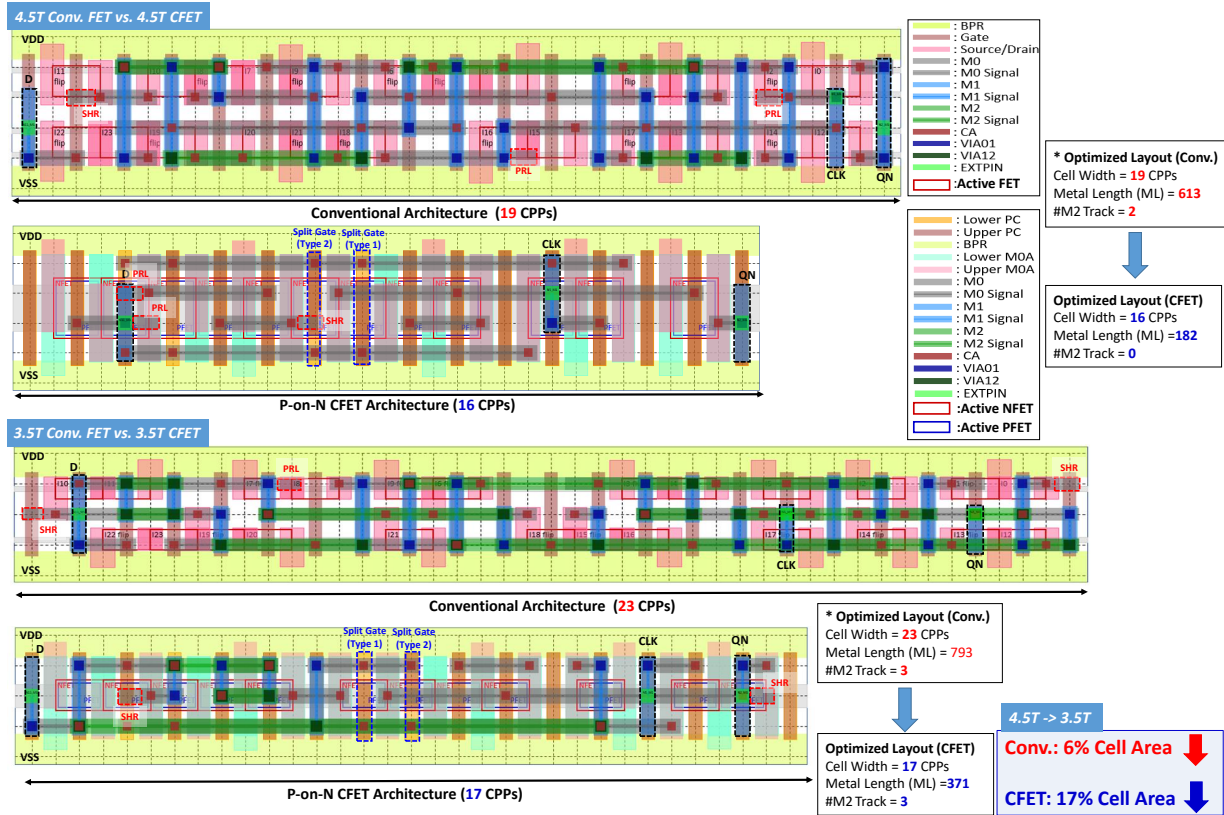


Figure 3.15: Layouts of 4.5T and 3.5T CFET and Conv. DFFHQN with corrected design constraints. The metal length is the weighted sum of metal segments and vias. Optimized result of 4.5T CFET layout: Cell Size (19→16), Metal Length (613→182), #M2Track (2→0); Optimized result of 3.5T CFET layout: Cell Size (23→17), Metal Length (793→371), #M2Track (3→3); When scaling from 4.5T to 3.5T, CFET provide 11% more cell area reduction than Conv. structure in DFFHQN SDC. The red dash-line boxes are metal extension for PRL and SHR design constraints.

The different net connections and number of fingers of pull-up (i.e., P-FET) and pull-down (i.e., N-FET) networks in SDC netlist cause approximately 1% average metal length variation. As a result, we use P-on-N CFET SDCs in the following experiments when using baseline design rules. However, when the design rules (i.e., EOL and VR) become stricter than baseline, the different number of available access points of upper and lower FETs leads to considerable differences in SDC layouts. This will be discussed in Exp. 3.5.2.

Block-Level Comparison: We use the 4.5T and 3.5T CFET and Conv. SDCs [9] in block-level P&R

Table 3.3: Experimental statistics of Conv. and CFET of 4.5T and 3.5T structures: ML= Metal Length (each via and M2 grid costs 4 grids), #M2 Track=number of used M2 tracks, CPP= Contact Poly Pitch, Cell Width Impr. = ((Cell Width of Conv. - Cell Width of CFET)/Cell width of Conv.), ML Impr. = ((ML of Conv. - ML of CFET)/ML of Conv.), PN/NP CFET=P-on-N/N-on-P CFET structure. Note that the PN and NP CFET cell width and #M2 Track are the same. Runtime=PN CFET SDC generation time (difference of avg. runtime of PN and NP CFET SDCs is less than 60s).

Cell Specification			4.5T CFET Cell Layout Objectives									3.5T CFET Layout Objectives									Runtime (s)	
Name	#FET	#Net	Cell Width (CPPs)			Metal Length				#M2 Tracks		Cell Width (CPPs)			Metal Length				#M2 Tracks		4.5T	3.5T
			Conv.	CFET	Impr (%)	Conv.	PN CFET	NP CFET	Impr (%)	Conv.	CFET	Conv.	CFET	Impr (%)	Conv.	PN CFET	NP CFET	Impr (%)	Conv.	CFET		
AND2x2	6	7	6	6	0.00	75	60	60	20.00	0	0	6	6	0.00	65	56	56	13.85	0	0	8.12	7.79
AND3x1	8	9	6	6	0.00	91	68	68	25.27	0	0	7	6	14.29	78	63	63	19.23	0	0	30.09	11.69
AND3x2	8	9	7	7	0.00	97	76	76	21.65	0	0	8	7	12.50	84	71	71	15.48	0	0	28.35	16.87
AOI21x1	6	8	9	9	0.00	197	142	142	27.92	1	0	15	12	20.00	285	285	285	0.00	2	2	119.69	100.71
AOI22x1	8	10	14	11	21.43	311	255	240	18.01	1	1	19	17	10.53	581	499	495	14.11	2	3	363.16	2735.42
BUFx2	4	5	5	5	0.00	61	40	40	34.43	0	0	6	5	16.67	54	38	38	29.63	0	0	4.92	1.85
BUFx3	4	5	6	6	0.00	82	53	53	35.37	0	0	6	6	0.00	176	51	51	71.02	2	0	11.20	3.95
BUFx4	4	5	7	7	0.00	88	59	59	32.95	0	0	7	7	0.00	190	57	57	70.00	2	0	7.91	4.72
BUFx8	4	5	12	12	0.00	149	105	105	29.53	0	0	12	12	0.00	274	102	102	62.77	2	0	43.65	29.62
DIFFHQN	24	17	19	16	15.79	613	182	182	70.31	2	0	23	17	26.09	793	371	371	53.22	3	3	6831.77	242.98
FA	24	17	14	14	0.00	420	379	379	9.76	3	2	21	17	19.05	857	676	663	21.12	3	3	6653.07	8417.49
INVx1	2	4	3	3	0.00	44	23	23	47.73	0	0	3	3	0.00	26	20	20	23.08	0	0	0.49	0.20
INVx2	2	4	4	4	0.00	38	29	29	23.68	0	0	4	4	0.00	36	27	27	25.00	0	0	1.03	0.91
INVx4	2	4	6	6	0.00	65	48	48	26.15	0	0	6	6	0.00	62	46	46	25.81	0	0	3.46	2.08
INVx8	2	4	10	10	0.00	121	92	92	23.97	0	0	10	10	0.00	118	86	86	27.12	0	0	19.14	6.11
NAND2x1	4	6	6	6	0.00	79	74	74	6.33	0	0	7	6	14.29	84	71	71	15.48	0	0	15.88	4.13
NAND2x2	4	6	10	10	0.00	140	131	131	6.43	0	0	13	10	23.08	154	123	123	20.13	0	0	33.83	30.75
NAND3x1	6	8	11	11	0.00	152	149	146	1.97	0	0	13	12	7.69	254	318	232	-25.20	2	2	124.11	78.78
NAND3x2	6	8	21	21	0.00	305	286	283	6.23	0	0	25	21	16.00	650	502	534	22.77	3	2	2869.53	199.38
NOR2x1	4	6	6	6	0.00	79	74	74	6.33	0	0	7	6	14.29	84	71	71	15.48	0	0	12.89	7.64
NOR2x2	4	6	10	10	0.00	140	131	131	6.43	0	0	13	10	23.08	154	123	123	20.13	0	0	27.94	285.32
NOR3x1	6	8	11	11	0.00	152	148	156	2.63	0	0	13	12	7.69	248	232	318	6.45	2	2	52.33	201.91
NOR3x2	6	8	21	21	0.00	304	283	286	6.91	0	0	25	21	16.00	642	534	502	16.82	3	3	1897.53	665.28
OAI21x1	6	8	11	9	18.18	247	146	149	40.89	1	0	14	12	14.29	416	428	428	-2.88	3	3	52.52	133.19
OAI22x1	8	10	14	11	21.43	311	240	255	22.83	1	1	19	17	10.53	581	495	499	14.80	2	2	612.60	559.16
OR2x2	6	8	6	6	0.00	75	60	60	20.00	0	0	6	6	0.00	65	56	56	13.85	0	0	12.99	4.80
OR3x1	8	9	6	6	0.00	91	68	68	25.27	0	0	7	6	14.29	78	62	62	20.51	0	0	76.77	11.46
OR3x2	8	9	7	7	0.00	97	76	76	21.65	0	0	8	7	12.50	84	71	71	15.48	0	0	89.22	10.97
XNOR2x1	10	9	12	11	8.33	274	220	223	19.71	1	1	17	14	17.65	573	492	491	14.14	3	3	977.00	574.68
XOR2x1	10	9	12	11	8.33	276	214	213	22.46	1	1	17	14	17.65	441	446	446	-1.13	3	3	134.86	848.13
Avg.	6.80	7.70	9.73	9.30	3.12	172.47	130.37	130.70	22.09	0.37	0.20	11.90	10.30	10.94	272.90	215.73	215.27	21.27	1.23	1.03	703.87	506.60

using M2-M5 routing layers. In Figure 3.16, CFET SDCs achieve 7.04%, 11.97%, and 4.12% (depicted in blue dashed arrows) smaller minimum valid block-level areas than Conv. SDCs for 4.5T structure and 7.78%, 11.38%, and 15.10% (depicted in orange arrows) for 3.5T structure in M0 Core, M1 Core and AES, respectively. The area improvement is obtained by the cell area shrinkage¹⁴ and M2 track usage reduction from CFET architecture conversion.

¹⁴The block-level SDC area of CFET SDCs are 5.6% and 14.0% smaller than Conv. SDCs on average at the same design utilization for 4.5T and 3.5T, respectively.

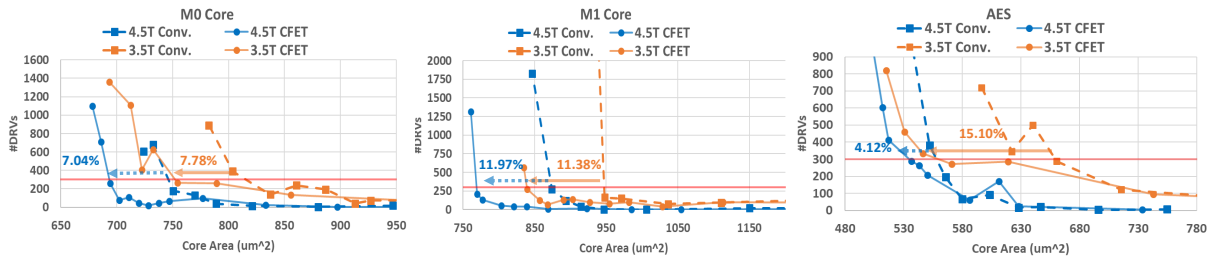


Figure 3.16: Block-Level P&R Results of M0 Core, M1 Core and AES of Conv. structure, which is generated using [9], and CFET with 4.5T and 3.5T cell height.

3.5.2 DTCO Exploration with CFET SDC Scaling

We explore the impacts of DTCO on cell metrics and block-level area using M0 Core and AES designs (i.e., CPU core and signal processing block). In Table 3.4, we generate each case by perturbing each EOL and VR design rule from the baseline with two CFET stacking options (i.e., P-on-N and N-on-P) for 4.5T and 3.5T cell heights. The M2 EOL Spacing and VIA12 Spacing rules of each case for the block-level routing are derived from the conditional design rules as shown in Figure 3.9. The pin-accessibility of all the SDCs are ensured (i.e., Avg. Min. RPA>2) by applying “MPO=3” and “EB-PS” objective. In Exp. 3.5.2, 3.5.2, and 3.5.2, we only consider P-on-N stacking cases for explorations of #BEOLs, EOL, and VR because the Max. ΔCW and $\Delta M2Track$, which are significantly related to the block-level area and routability, are 0 for most of the cases. Except for Exp. 3.5.2, M2-M5 layers are used in block-level routing. The DTCO explorations are organized as follows.

- Exp. 3.5.2: *Explorations of #BEOLs:* We vary the #BEOLs to explore the block-level area variation with baseline.
- Exp. 3.5.2: *Explorations of EOL Rule:* We vary the EOL and study its impact on cell metrics and block-level areas.
- Exp. 3.5.2: *Explorations of VR Rule:* We tune VR and study the variations in both cell-level and

block-level.

- Exp. 3.5.2: *Explorations of CFET Stacking Option*: We explore the impact of the CFET stacking options on cell metrics and block-level area with EOL=3 and VR=1.5.

Table 3.4: Design Technology Co-Optimization (DTCO) Experimental Results of 30 4.5T and 3.5T CFET SDCs with various design rule and stacking option using MPO=3 and EB-PS. CH = Cell Height, CW = Cell Width, ML = Total Metal Length, M2ML = the metal length of M2, #M2Track = the number of used M2 tracks, PN/NP = P-on-N/N-on-P, Baseline: EOL/VR = 2/1, CW_{PN}/CW_{NP} = Cell Width of PN/NP CFET, $M2ML_{PN}/M2ML_{NP}$ = M2 Metal Length of PN/NP CFET, $\#M2Track_{PN}/\#M2Track_{NP}$ = #M2 Track of PN/NP CFET, Max ΔCW = maximum value of $\frac{\|CW_{PN}-CW_{NP}\|}{CW_{PN}}$ (%), Max $\Delta M2ML$ = maximum value of $\|M2ML_{PN} - M2ML_{NP}\|$, Max $\Delta M2Track$ = maximum value of $\|\#M2Track_{PN} - \#M2Track_{NP}\|$, C-to-C = Center-to-Center.

CH	DTCO		Cell Metrics									Block-Level Design Rule	
	Design Rules	Stacking	Avg. CW	Max. ΔCW	Avg. ML	Max. ΔML	Avg. M2ML	Max. $\Delta M2ML$	Avg. #M2Track	Max. $\Delta M2Track$	Avg. Min. RPA	M2 EOL Spacing (nm)	VIA12 C-to-C Spacing (nm)
4.5T	Baseline	PN	9.30	0.00%	130.37	6.25%	2.40	6.00	0.20	0.00	2.81	30.00	31.00
		NP	9.30		130.70		2.20		0.20		2.83		
	EOL=1	PN	9.30	0.00%	123.87	10.54%	2.20	4.00	0.13	0.00	2.94	12.00	31.00
		NP	9.30		124.47		2.40		0.13		2.99		
	EOL=3	PN	9.50	11.11%	154.30	40.94%	5.87	14.00	0.50	1.00	2.48	51.00	31.00
		NP	9.50		150.50		5.33		0.43		2.50		
	VR=0	PN	9.30	0.00%	124.77	11.45%	2.07	8.00	0.13	0.00	2.75	30.00	24.00
		NP	9.30		121.70		1.67		0.13		2.73		
	VR=1.5	PN	10.20	9.09%	224.47	183.59%	13.90	90.00	1.03	3.00	2.98	30.00	40.00
		NP	10.10		223.77		13.73		1.10		2.90		
3.5T	Baseline	PN	10.30	0.00%	215.73	37.07%	16.93	10.00	1.03	0.00	2.83	30.00	31.00
		NP	10.30		215.27		16.23		1.03		2.79		
	EOL=1	PN	9.70	0.00%	199.27	14.08%	11.77	6.00	0.97	0.00	2.94	12.00	31.00
		NP	9.70		198.90		11.87		0.97		2.92		
	EOL=3	PN	11.43	16.67%	264.43	265.10%	23.67	16.00	1.40	2.00	2.59	51.00	31.00
		NP	11.43		265.10		23.47		1.43		2.59		
	VR=0	PN	9.63	0.00%	187.90	22.47%	11.33	11.00	0.93	0.00	2.83	30.00	24.00
		NP	9.63		184.80		11.73		0.93		2.82		

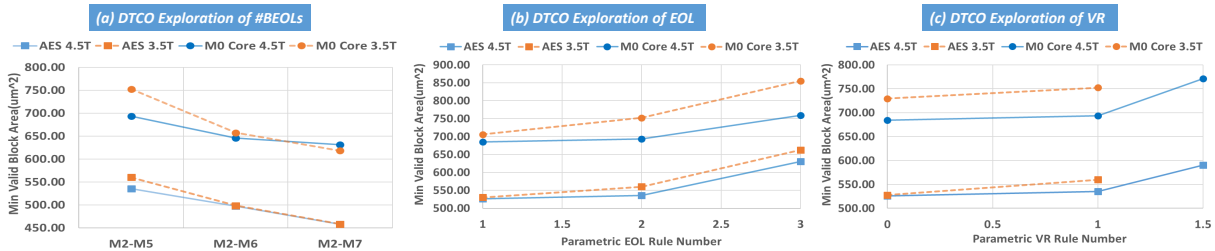


Figure 3.17: Min. valid M0 Core and AES block-level areas with 300 #DRVs threshold versus (a) #BEOLs, (b) EOL, and (c) VR.

Exploration of #BEOLs

We vary the #BEOLs (i.e., routing resource) to explore the block-level area variation using 4.5T and 3.5T CFET SDCs with baseline in Table 3.4. Figure 3.17 (a) shows the minimum valid block-level area with M2-M5, M2-M6, and M2-M7 routing layers¹⁵ using 300 #DRVs as the threshold. When using the M2-M5 as routing layers, the minimum valid M0 Core and AES block-level areas for 4.5T CFET SDCs are 7.80 and 4.42% smaller than 3.5T CFET SDCs, respectively, because of 29% more #M2 track usage per cell on average¹⁶, and 75% more Avg. M2 Metal Length in 3.5T CFET SDCs. When adding two more routing layers, the minimum valid M0 Core and AES block-level areas of 3.5T CFET SDCs are reduced significantly (18.0% on average) and even can achieve smaller valid block-level area (i.e., M0 Core) than 4.5T CFET SDCs. The block area reduction comes from more routing resource that alleviates the routing congestion caused by the #M2 track usage and M2 ML in 3.5T CFET SDCs.

Exploration of EOL Rule

We explore the impact of the EOL spacing rule in both cell-level and block level with 4.5T and 3.5T CFET SDCs. In Table 3.4, when adjusting EOL=2 (baseline) to EOL=1, the Avg. CW, ML, M2ML, and #M2Track are reduced by 0.00, 6.50, 0.20, and 0.07 for 4.5T CFET SDCs and 0.60, 16.46, 5.16, and 0.06 for 3.5T CFET SDCs, respectively. On the other hand, when changing EOL=2 to EOL=1, the Avg. CW, ML, M2ML, and #M2Track are increased by 0.20, 23.93, 3.47, and 0.30 for 4.5T CFET SDCs and 1.13, 48.70, 6.74, and 0.37 for 3.5T CFET SDCs, respectively.

For block-level area study, we extract the minimum valid M0 Core and AES block-level areas with 300 #DRVs threshold for EOL=1, EOL=2, and EOL=3 as shown in Figure 3.17 (b). Compared to

¹⁵Using the top routing layer below M5 is expected to be limited by insufficient routing resource because M2 is also used in SDC. M7 is the maximum routing layer here because M8 and M9 are used by top power mesh as stated in Section 3.3.

¹⁶Avg. #M2 track usage per cell = (Avg. #M2Track)/(#M2 RTs in Cell).

baseline, the minimum valid M0 Core and AES block-level areas are increased by 8.67% and 15.16% for 4.5T CFET SDCs and 12.05% and 15.51% for 3.5T CFET SDCs with EOL=3. On the other hand, the minimum valid M0 Core and AES block-level areas with EOL=1 are decreased by 1.23% and 1.70% for 4.5T CFET SDCs and 6.14% and 5.28% for 3.5T CFET SDCs, compared to the baseline. From the results, we observe that 3.5T CFET SDCs have larger variation on cell metrics and block-level area than 4.5T CFET SDCs by the perturbation of the EOL due to the less in-cell routing resource which results in more occupied M2 resources (i.e., #M2Tracks and M2 ML) in 3.5T CFET SDCs.

Exploration of VR Rule

We tune the VR from the baseline (i.e., VR=1, allowing diagonal via) to VR=0 (i.e., allowing all adjacent via) and VR=1.5 (i.e., not allowing diagonal via) to study the impact on SDC metrics and block area. In Table 3.4, when changing VR=1 (baseline) to VR=0, the Avg. CW, ML, M2ML, and #M2Track are reduced by 0.00, 5.60, 0.33, and 0.07 for 4.5T CFET SDCs and 0.67, 27.83, 5.60, and 0.10 for 3.5T CFET SDCs, respectively. On the contrary, the Avg. CW, ML, M2ML, and #M2Track are increased by 0.90, 94.1, 11.50, and 0.83 for 4.5T CFET SDCs with VR=1.5. Here, many 3.5T CFET SDCs (i.e., NAND2x2, NAND3x1, etc.) don't have feasible solutions when applying VR=1.5 because the VR blocks the access of adjacent G/S/D FET terminals if any CA (depicted in red square shape) is placed at the middle horizontal track as shown in Figure 3.18.

In block-level analysis, the minimum valid M0 Core and AES block-level areas are increased by 10.04% and 9.36% in 4.5T CFET SDCs, respectively, as shown in Figure 3.17 (c) when changing VR=1 to VR=1.5. The increased block-level areas are caused by more occupied #M2Track and M2 ML in SDC layouts and larger via spacing rule. On the other hand, when adjusting VR=1 to VR=0, the minimum valid M0 Core and AES block-level areas are reduced by 1.36% and 1.74% for 4.5T CFET SDCs and 3.02%

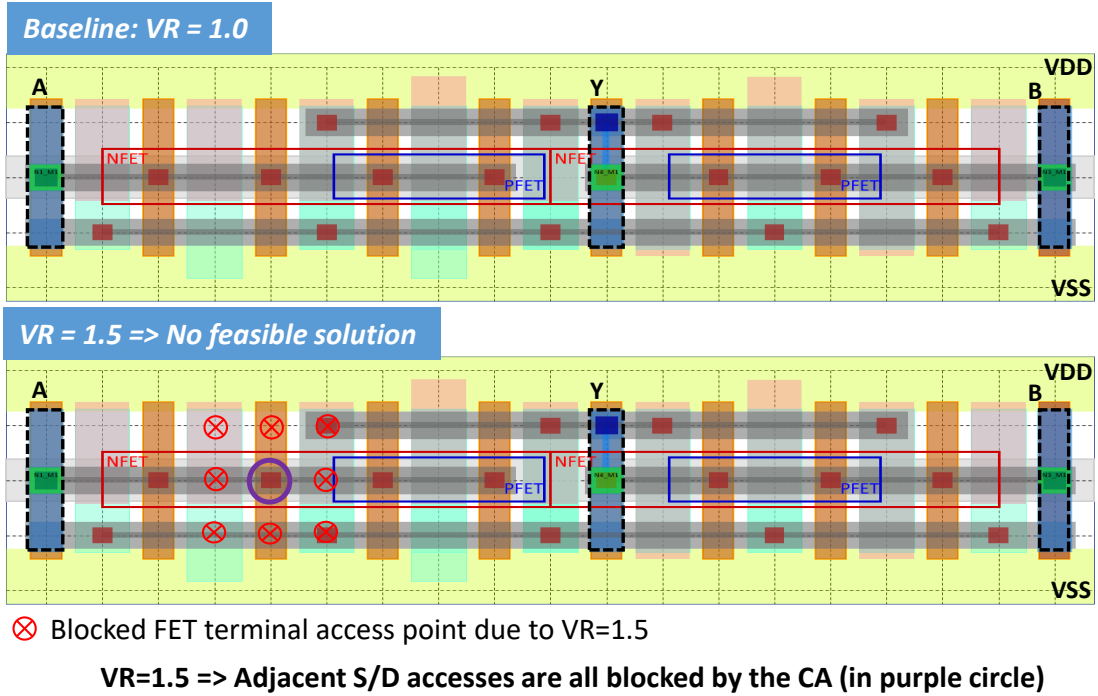


Figure 3.18: An illustration of NAND2x2 without feasible solution when VR=1.5 in 3.5T P-on-N CFET structure.

and 5.76% for 3.5T CFET SDCs, respectively. Here, we also observe that 3.5T CFET SDCs have larger variation on cell metrics and block-level area than 4.5T CFET SDCs when varying the VR rule because of less in-cell routing resource and more occupied M2 resources (M2 tracks and M2 metal length) in 3.5T CFET SDCs.

Exploration of CFET Stacking Option

We explore the impact of the CFET stacking options that lead to nontrivial difference in CFET SDCs with EOL=3 and VR=1.5 due to the different number of access points of upper and lower FETs and different net connections in pull-up (i.e., P-FET) and pull-down (i.e., N-FET) networks in SDC. Also, these differences in SDC layouts results in the block-level area variations.

EOL=3 Case: In Table 3.4, the Max. ΔCW , ΔML , $\Delta M2ML$, and $\Delta M2Track$ of P-on-N and N-on-P

Table 3.5: Difference of P-on-N and N-on-P 4.5T and 3.5T CFET SDCs with EOL=3. CH=CellHeight, CW=Cell Width, CW_{PN}/CW_{NP} =Cell Width of PN/NP CFET, $M2ML_{PN}/M2ML_{NP}$ =M2 Metal Length of PN/NP CFET, $\#M2Track_{PN}/\#M2Track_{NP}$ =#M2 Track of PN/NP CFET, $\Delta CW = \frac{\|CW_{PN} - CW_{NP}\|}{CW_{PN}}$ (%), $\Delta M2ML = \|M2ML_{PN} - M2ML_{NP}\|$, $\Delta M2Track = \|\#M2Track_{PN} - \#M2Track_{NP}\|$.

CH	Cell	ΔCW	$\Delta M2ML$	$\Delta M2Track$
4.5T	AOI21x1	10.00%	2.00	1.00
	AOI22x1	0.00%	2.00	0.00
	NAND3x1	0.00%	14.00	1.00
	NAND3x2	0.00%	6.00	1.00
	NOR3x1	0.00%	8.00	1.00
	NOR3x2	0.00%	6.00	1.00
	OAI21x1	11.11%	6.00	1.00
	OAI22x1	0.00%	4.00	0.00
	Avg.	2.64%	6.00	0.75
3.5T	AOI22x1	0.00%	7.00	0.00
	NAND2x1	14.29%	10.00	2.00
	NAND2x2	9.09%	10.00	1.00
	NAND3x1	7.69%	12.00	0.00
	NAND3x2	4.35%	16.00	0.00
	NOR2x1	16.67%	10.00	2.00
	NOR2x2	10.00%	10.00	1.00
	NOR3x1	8.33%	4.00	0.00
	NOR3x2	4.55%	16.00	0.00
	OAI22x1	0.00%	11.00	0.00
	XNOR2x1	0.00%	12.00	0.00
	XOR2x1	0.00%	3.00	0.00
	Avg.	6.25%	10.08	0.50

stacking with EOL=3 are up to 16.67%, 265.10%, 16, and 2 in 4.5T and 3.5T structures. In addition, the variations of cell metrics of 3.5T CFET SDCs are larger than 4.5T CFET SDCs due to less in-cell routing resource and less FET G/S/D access points. The pin-accessibility of P-on-N and N-on-P CFET SDCs are secured by the proposed EB-PS objective (i.e., Avg. Min. RPA>2). Table 3.5 depicts the SDCs with

Table 3.6: Difference of P-on-N and N-on-P 4.5T CFET SDCs with VR=1.5.

CH	Cell	ΔCW	$\Delta M2ML$	$\Delta M2Track$
4.5T	AOI21x1	9.09%	16.00	2.00
	AOI22x1	0.00%	18.00	1.00
	FAx1	0.00%	12.00	0.00
	NAND3x1	8.33%	6.00	1.00
	NAND3x2	8.33%	90.00	3.00
	NOR3x1	9.09%	2.00	0.00
	NOR3x2	9.09%	90.00	3.00
	OAI21x1	8.33%	6.00	0.00
	OAI22x1	0.00%	29.00	0.00
	Avg.	5.81%	29.89	1.11

Table 3.7: M0 Core and AES Block Weighted Metric (i.e., $M2Track_d$, $M2ML_d$) of P-on-N (PN) and N-on-P (NP) CFET SDCs with EOL=3 and VR=1.5 design rule. Min. BA = Minimum Valid Block-Level Area (μm^2).

CH	Settings	Stacking	M0 Core			AES		
			$M2Track_d$	$M2ML_d$	Min. BA (μm^2)	$M2Track_d$	$M2ML_d$	Min. BA (μm^2)
4.5T	EOL=3	PN	0.36	3.13	759.47	0.21	2.18	630.85
		NP	0.25	2.93	716.74	0.16	1.78	618.55
		Diff. (%)	44.00%	6.82%	5.63%	8.33%	22.47%	1.95%
	VR=1.5	PN	0.84	8.80	771.10	0.47	4.85	590.46
		NP	0.70	8.56	758.65	0.37	4.54	553.91
		Diff. (%)	20.00%	2.80%	1.61%	27.02%	6.61%	6.19%
3.5T	EOL=3	PN	1.01	12.34	855.45	1.21	11.83	662.70
		NP	1.06	12.25	889.56	1.24	11.74	669.50
		Diff. (%)	5.00%	0.73%	3.83%	2.48%	0.76%	1.02%

nontrivial difference between P-on-N and N-on-P stacking for EOL=3.

To quantify the M2 resource occupied by SDC layouts in block-level, we calculate the block weighted $M2Track_d$ and $M2ML_d$ based on the corresponding cell metric and cell percentage in the design d using Equation (3.14) and the SDCs in Table 3.5 as presented in Table 3.7.

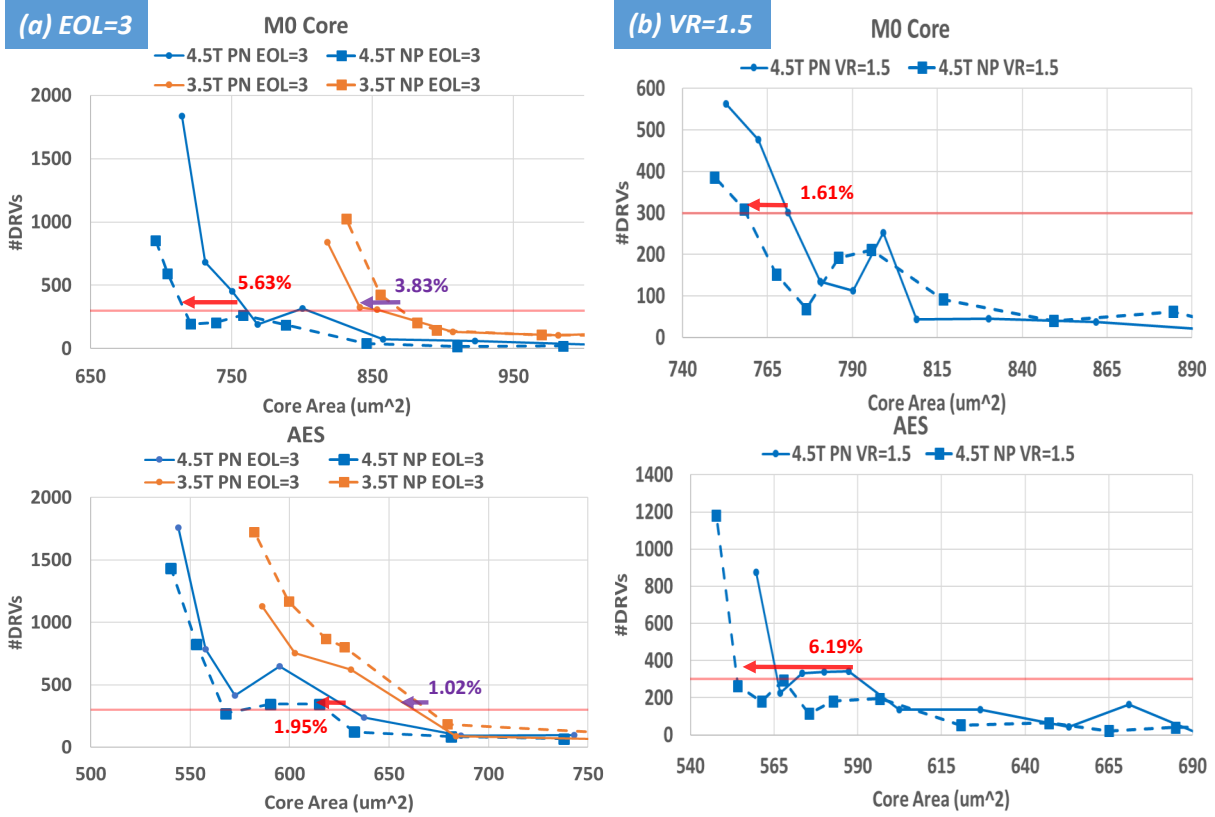


Figure 3.19: Block-level P&R Results of M0 Core and AES designs of P-on-N or N-on-P using EOL=3 and VR=1.5 design rules.

$$Metric_d = \sum_c Metric_c * CP_{d,c} \quad (3.14)$$

Where $Metric_d$ denotes the block weighted metric of design d , $Metric_c$ is the cell level metric of cell c , and the $CP_{d,c}$ is the percentage of cell c in design d .

The block-level P&R results of P-on-N and N-on-P CFET SDCs with EOL=3 are presented in Figure 3.19 (a). In Table 3.7, in 4.5T CFET SDCs, the $M2Track_d$ and $M2ML_d$ of P-on-N stacking are both larger than N-on-P stacking in M0 Core and AES designs. As a result, the minimum valid M0 Core and AES areas using N-on-P CFET are 5.63% and 1.95% (depicted in red arrows) smaller than P-on-N CFET for 4.5T, respectively. For 3.5T CFET SDCs, the $M2Track_d$ of P-on-N stacking are 5% and 2.48% smaller with 0.73% and 0.76% increment of $M2ML_d$ than N-on-P stacking in M0 Core and AES designs,

respectively. Hence, the minimum valid M0 Core and AES block-level areas using 3.5T P-on-N CFET are 3.83% and 1.02% (depicted in purple arrows) smaller than using 3.5T N-on-P CFET because of larger impact of #M2 Track usage on routability [6].

VR=1.5 Case: In Table 3.4, the Max. ΔCW , ΔML , $\Delta M2ML$, and $\Delta M2Track$ of P-on-N and N-on-P stacking are 9.09%, 183.59%, 90, and 3 in 4.5T CFET structure. Note that 3.5T CFET SDCs are not discussed because many SDCs have no feasible solutions as illustrated in Figure 3.18. Table 3.6 depicts the SDCs with considerable difference of P-on-N and N-on-P stacking for VR=1.5.

For block-level analysis, the $M2Track_d$ and $M2ML_d$ are calculated using SDCs in Table 3.6 as presented in Table 3.7. With 4.5T CFET SDCs, the $M2Track_d$ and $M2ML_d$ of P-on-N stacking are both larger than N-on-P stacking in M0 Core and AES designs. Therefore, the minimum valid M0 Core and AES block-level areas of N-on-P stacking are 1.61% and 6.19% (depicted in red arrows) smaller than P-on-N stacking, respectively.

In summary, cell metrics (i.e., CW, ML, M2 ML, and #M2Track) and minimum valid block-level area of 3.5T CFET SDCs are more sensitive to the variations of #BEOLs, EOL, and VR than 4.5T CFET SDCs since less in-cell routing resources and more M2 resource usage (i.e., M2 ML and #M2 Track usage) in 3.5T CFET SDC. When applying tighter design rules, the stacking options of CFET need to be considered because of they can impact the block-level area up to 6.19%. These studies suggest that we can potentially achieve smaller block-level area with 3.5T CFET SDCs than 4.5T CFET SDCs by loosening the design rules and increasing the #BEOLs.

3.5.3 DTCO for Block-Level Area Scaling

We exploit the CFET SDC cell height reduction and drive block-level area scaling of 3.5T CFET SDCs through design technology co-optimization. We adopt design parameters (EOL=1 and VR=0) and increase the #BEOLs from M2-M5 to M2-M7, that are observed to be more beneficial for the 3.5T CFET

SDCs in terms of the block-level area through the extensive DTCO explorations of CFET SDCs.

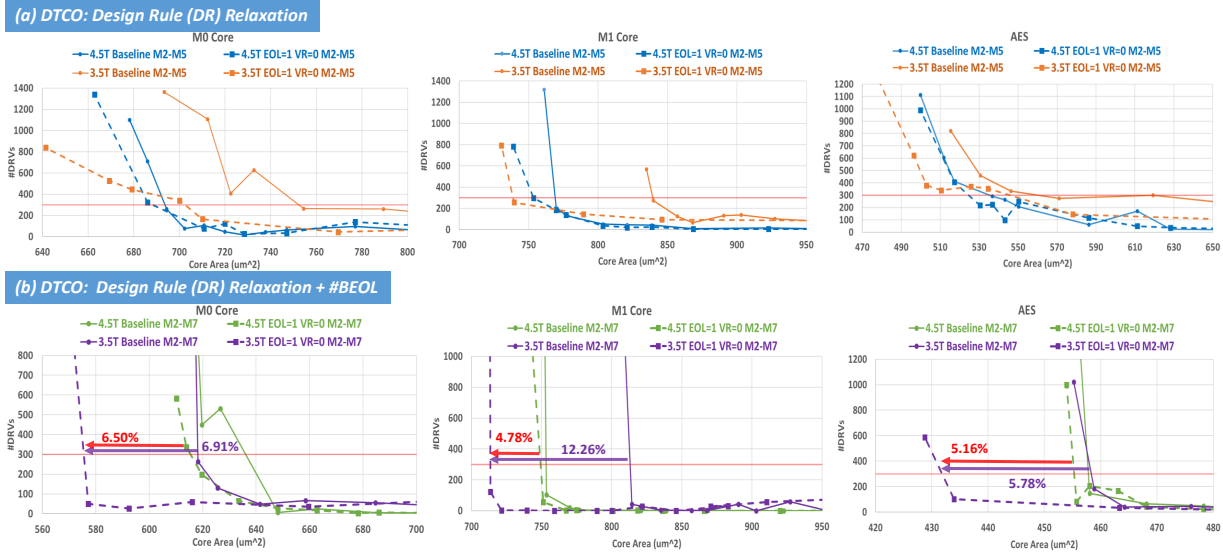


Figure 3.20: Design technology co-optimization block-level placement-and-route results of M0 Core, M1 Core and AES. (a) Design Rule Relaxation (EOL=1 and VR=0) and (b) Design Rule Relaxation and increasing the top routing layer from M5 to M7.

The block-level exploration results of M0 Core, M1 Core, and AES are demonstrated in Figure 3.20. The minimum block-level areas of M0 Core, M1 Core, and AES for 4.5T and 3.5T CFET SDCs are smaller than baseline with EOL=1 and VR=0 using M2-M5 routing layers as shown in Figure 3.20 (a). However, the minimum valid M0 Core and AES block areas for 3.5T CFET SDCs are still larger than 4.5T CFET SDCs with EOL=1 and VR=0 using M2-M5 routing layers. After increasing the top routing layer to M7, the 3.5T CFET SDCs achieve 6.50%, 4.78%, and 5.16% (depicted in red arrows) smaller minimum valid block-level area than 4.5T CFET SDCs with EOL=1 and VR=0 in M0 Core, M1 Core, and AES, respectively, as shown Figure 3.20 (b). In addition, compared to baseline, the minimum block-level areas of M0 Core, M1 Core, and AES are reduced by 6.91%, 12.26%, and 5.78% (depicted in purple arrows), respectively, with EOL=1 and VR=0 for 3.5T CFET SDCs as shown in Figure 3.20 (b). From the results, the cell area benefit of 3.5T CFET SDCs is maximized for further block-level area scaling through DTCO.

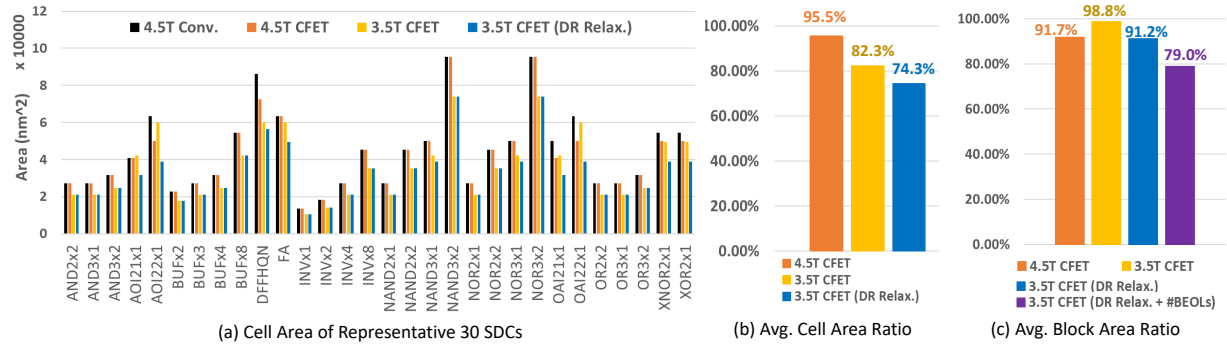


Figure 3.21: Cell and block-level area benefits by STCO and track reduction: 4.5T Conv. (black bar) 4.5T CFET (orange), 3.5T CFET (gold), 3.5T CFET with Design Rule Relaxation (DR Relax.) (blue), and DR Relax. plus adding #BEOLs in block-level for 3.5T CFET (purple). (a) Cell Area of Representative 30 SDCs. (b) Avg. Cell Area Ratio (Avg. Cell Area/Avg. Cell Area of 4.5T Conv.) (c) Avg. Block Area Ratio (Avg. Block Area/Avg. Block Area of 4.5T Conv.). CellArea = $CW \times CPP \times CH \times M2Pitch$, $CPP=42nm$, $M2Pitch=24nm$. Avg. Block Area=Avg. of min. valid block areas of M0 Core, M1 Core, and AES (Figure 3.16 in Exp. 3.5.1 and Figure 3.20 (b) in Exp. 3.5.3).

Here, we summarize the cell and block-level area benefits by STCO, cell height reduction, and DTCO (EOL=1 and VR=0) in Figure 3.21. In Figure 3.21 (a), we plot the cell area of representative 30 SDCs of 4.5T Conv. (black bar), 4.5T CFET (orange), 3.5T CFET (gold), and 3.5T CFET with design rule relaxation (DR Relax.) through DTCO (blue). Note that some 3.5T CFET (gold) SDC layouts (e.g. AOI21X1, AOI22x1, OAI21x1, and OAI22x1) are larger than 4.5T CFET (orange) SDC layouts. The growth of cell area is caused by the extra columns needed to maintain the routability when reducing 4 RTs to 3 RTs in SDC. For these cells, the design rule relaxation derives more area reduction due to the column reduction. Figure 3.21 (b) lists the average cell area reduction from 4.5T Conv. are 4.5%, 17.7% and 25.7% for 4.5T CFET, 3.5T CFET, and 3.5T CFET with DR Relax., respectively.

Figure 3.21 (c) shows that the ratio of average block area of 4.5T CFET (orange), 3.5T CFET (gold), 3.5T CFET with DR Relax. (blue), and 3.5T CFET with DR Relax. plus using M2-M7 BEOLs (purple) to average block area of 4.5T Conv. are 91.7%, 98.8%, 91.2% and 79.0%, respectively. Note that the average block-level area grows 7.1% for CFET when scaling 4.5T to 3.5T without DTCO. The growth of block-level area is caused by more M2 resource usage and less M2 horizontal tracks for accessing SDC

in 3.5T structure. With the assistance of DTCO for 3.5T CFET, the average block area is reduced by 21% compared to 4.5T Conv. SDCs.

3.6 Extreme CFET SDC Scaling

We explore the limit of CFET SDC layouts by scaling the cell height to 2.5T.¹⁷ VR relaxation are required for split structure in 2 RTs structure. The conditional design rules [21, 44] are as follows: MAR/EOL/VR/PRL/SHR = 1/1/0/1/2. The minimum I/O pin opening (MPO) constraint [6] is set to 3 for pin-accessibility. The experiments are organized as follows:

- **Exp. 3.6.1.** *Scaling to Extreme 2 RTs with Inter-Row Routing Options:* We compare the cell area, metal length, #Vias, and #M2 Track with/without Upper/Lower M0A/PC for inter-row routing as scaling 3.5T to 2.5T CFET structure using adaptive cell row number for cell area minimization.
- **Exp. 3.6.2.** *Block-Level Area Scaling with 2.5T CFET:* We explore the minimum valid block-level areas of M0 Core, M1 Core, and AES with 300 #DRVs threshold for 3.5T CFET, and 2.5T CFET with/without Upper/Lower M0A/PC routing SDCs.

3.6.1 Scaling to Extreme 2 Routing Tracks (RTs) with Inter-Row Routing Options

We explore the CFET SDC cell area benefits as reducing the number of tracks using the proposed Multi-Row CFET SDC synthesis framework with/without Upper/Lower M0A/PC for inter-row routing options.

Inter-Row routing with metal layers only: We compare the cell area, #M2 Tracks, ML, and #Vias of 3.5T CFET and 2.5T CFET with metal layers (i.e., M1) for inter-row routing in Table 3.8. The average

¹⁷2.5T is the limit for CFET SDC structure since the split structure needs at least 2 access points from M0 as shown in Figure 2.1. From [43], 2 RTs Conv. cell structure can not be implemented due to the limitation of P-N separation.

Table 3.8: Experimental statistics of 3.5T CFET, 2.5T CFET, and 2.5T CFET with Upper/Lower M0A/PC routing (2.5T M0A/PC-R): CW= Cell Width (CPP), Opt. CR= Optimum Cell Row, ML=Metal Length (Not including Vias), #Vias=Number of Vias, #M2 Track=number of used M2 tracks, CPP= Contact Poly Pitch, Cell Area Impr. = $((3.5T\ CW \times 3.5T\ Opt.\ CR - 2.5T/(2.5T\ M0A/PC-R)\ CW \times 2.5T/(2.5T\ M0A/PC-R)\ Opt.CR)/(3.5T\ CW \times 3.5T\ Opt.\ CR))$.

Cell Specification			Cell Layout Objectives															Runtime (s)																	
Name	#FET	#Net	CW			Opt. CR			Cell Area Impr. (%)		Interconnect						#M2Track			3.5T	2.5T	2.5T M0A/PC-R													
			3.5T	2.5T	2.5T M0A/PC-R	3.5T	2.5T	2.5T M0A/PC-R	2.5T	2.5T M0A/PC-R	3.5T		2.5T		2.5T M0A/PC-R		3.5T	2.5T	2.5T M0A/PC-R																
										ML	#Vias	ML	#Vias	ML	#Vias	ML	#Vias																		
AND2x2	6	7	6	6	6	1	1	1	28.57	28.57	16	10	21	16	21	16	0	1	1	22.72	14.70	8.65													
AND3x1	8	9	6	7	7	1	1	1	16.67	16.67	20	11	20	22	20	22	0	2	2	34.65	28.89	50.44													
AND3x2	8	9	7	8	8	1	1	1	18.37	18.37	22	12	26	20	26	20	0	1	1	44.91	69.37	50.52													
AOI21x1	6	8	9	7	6	1	2	2	-11.11	4.76	52	28	65	45	50	28	2	3	2	259.22	3243.52	3575.33													
AOI22x1	8	10	11	9	7	1	2	2	-16.88	9.09	71	41	89	69	66	40	3	4	2	1285.34	6710.78	6808.18													
BUFx2	4	5	5	5	5	1	1	1	28.57	28.57	10	7	16	11	16	11	0	1	1	12.59	6.21	4.22													
BUFx3	4	5	6	6	6	1	1	1	28.57	28.57	18	9	18	13	18	13	0	1	1	19.61	9.96	5.49													
BUFx4	4	5	7	7	7	1	1	1	28.57	28.57	17	10	21	14	21	14	0	1	1	24.46	13.86	8.75													
BUFx8	4	5	12	12	12	1	1	1	28.57	28.57	36	17	39	21	39	21	0	1	1	79.89	49.81	48.64													
DFFHQN	24	17	16	10	9	1	2	2	10.71	19.64	76	34	85	56	66	35	1	3	0	12982.42	23423.95	21071.77													
FA	24	17	14	9	8	1	2	2	8.16	18.37	126	61	96	70	88	50	3	4	4	15071.87	26287.88	24394.16													
INVx1	2	4	3	3	3	1	1	1	28.57	28.57	5	4	12	8	12	8	0	2	2	1.01	3.48	1.48													
INVx2	2	4	4	4	4	1	1	1	28.57	28.57	9	5	12	10	12	10	0	1	1	7.69	3.06	1.94													
INVx4	2	4	6	6	6	1	1	1	28.57	28.57	16	8	18	13	18	13	0	1	1	12.06	6.29	457.00													
INVx8	2	4	10	10	10	1	1	1	28.57	28.57	30	14	32	19	32	19	0	1	1	32.29	14.94	15.08													
NAND2x1	4	6	6	6	6	1	1	1	28.57	28.57	23	12	23	20	23	20	0	2	2	18.92	13.98	12.76													
NAND2x2	4	6	10	11	11	1	1	1	21.43	21.43	45	20	51	37	35	16	0	2	2	60.52	78.31	41.26													
NAND3x1	6	8	11	14	14	1	1	1	9.09	9.09	70	38	85	50	83	50	1	2	2	167.83	383.02	887.56													
NAND3x2	6	8	21	26	26	1	1	1	11.56	11.56	135	55	129	78	129	78	2	2	2	607.30	957.39	909.03													
NOR2x1	4	6	6	6	6	1	1	1	28.57	28.57	24	12	23	20	23	20	0	2	2	23.71	21.23	16.96													
NOR2x2	4	6	10	11	11	1	1	1	21.43	21.43	46	20	51	37	51	37	0	2	2	86.80	78.96	44.37													
NOR3x1	6	8	11	14	14	1	1	1	9.09	9.09	55	25	83	50	83	50	1	2	2	596.39	894.79	914.35													
NOR3x2	6	8	21	26	26	1	1	1	11.56	11.56	131	51	174	84	174	84	1	2	2	340.20	1102.90	1027.98													
OAI21x1	6	8	9	7	6	1	2	2	-11.11	4.76	72	37	74	49	46	26	3	4	2	174.99	2183.95	2122.94													
OAI22x1	8	10	11	9	7	1	2	2	-16.88	9.09	79	50	89	69	69	40	3	4	2	890.15	6313.42	7043.85													
OR2x2	6	8	6	6	6	1	1	1	28.57	28.57	16	10	21	16	21	16	0	1	1	21.56	15.11	14.22													
OR3x1	8	9	6	7	6	1	1	1	16.67	28.57	20	11	20	22	16	14	0	2	2	42.01	28.96	73.58													
OR3x2	8	9	7	8	7	1	1	1	18.37	28.57	22	12	26	20	26	20	0	1	2	48.66	68.54	95.94													
XNOR2x1	10	9	12	7	7	1	2	2	16.67	16.67	93	57	69	51	68	50	3	4	2	6899.62	5941.67	5766.47													
XOR2x1	10	9	12	7	7	1	2	2	16.67	16.67	84	36	72	48	68	47	3	4	2	3108.52	2168.67	2122.94													
Avg.	5.93	7.26	9.37	9.13	8.80	1.00	1.27	1.27	16.44	20.61	47.97	23.90	52.00	35.27	47.33	29.60	0.87	2.10	1.67	1432.60	2671.25	2586.53													

runtime per cell is around 45 minutes for 2.5T CFET and 24 minutes for 3.5T CFET. As scaling from 3.5T to 2.5T CFET cell architecture, the average cell area is reduced by 16.44% with 8.40%, 47.57% and 1.23 increment on average ML, #Vias, and #M2 Track, respectively. The increase of ML, #Vias, and #M2 Track is caused by less in-cell routing resources and the constraints of design rules and pin-accessibility in 2.5T CFET cell structure.

Figure 3.23 (a) and (b) shows the design corrected DFFHQN layouts of 3.5T and 2.5T CFET cell

structures, respectively. The double-row 2.5T CFET cell structure achieves 10.7% smaller cell area than 3.5T CFET cell structure. The reduced cell area comes from leveraging direct M1 inter-row connection of the shared and split structures as shown in the red dash box in Figure 3.23 (a).

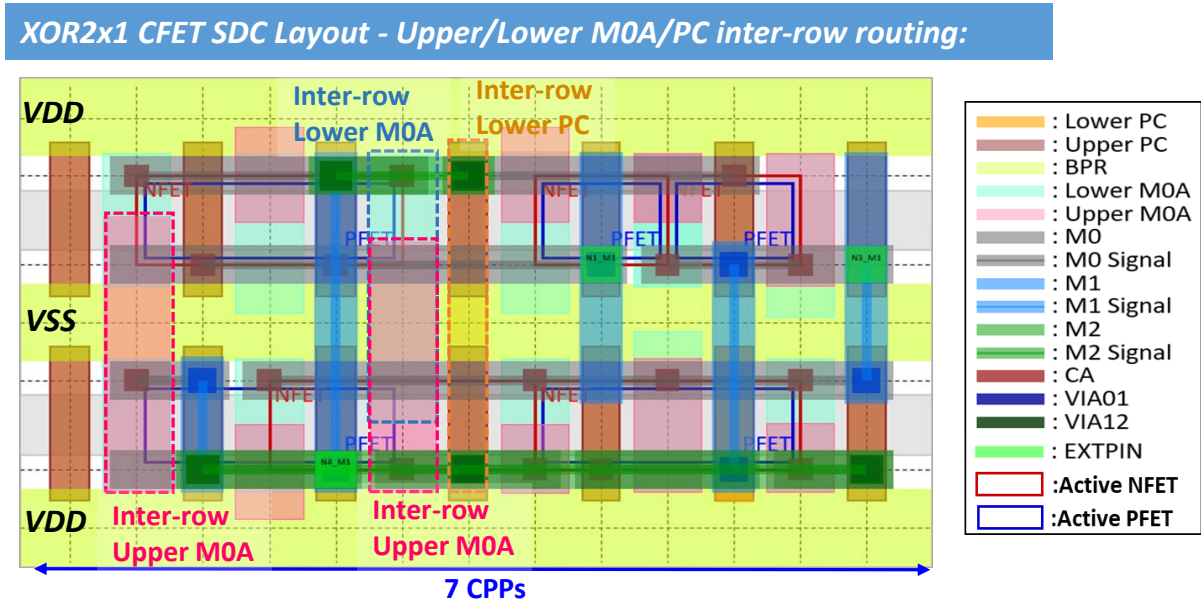


Figure 3.22: An example of XOR2x1 SDC layouts with Upper/Lower M0A/PC for inter-row routing.

Enable Upper/Lower M0A/PC for inter-row routing: We enable the inter-row routing with Upper/Lower M0A/PC in 2.5T CFET structure (2.5T M0A/PC-R CFET) and compare the cell area, #M2 Tracks, ML, and #Via of 2.5T M0A/PC-R CFET with 3.5T CFET in Table 3.8. The average runtime per cell is around 43 minutes for 2.5T M0A/PC-R CFET. Compared to 3.5T CFET, 2.5T M0A/PC-R CFET achieves 20.61% and 1.33% smaller cell area and ML on average with 23.85% and 0.80 increment on average #Vias and #M2 Track, respectively. Compared to 2.5T CFET, 2.5T M0A/PC-R CFET provides 4.03%, 8.98%, 16.08%, and 20.48% smaller cell area, ML, #Vias, and #M2 Track on average, respectively. This shows that enabling M0A/PC for routing can reduce not only cell size but also parasitic resistance in SDC.

Figure 3.22 shows the shared-and-split structures across cell rows through M0A/PC layers with

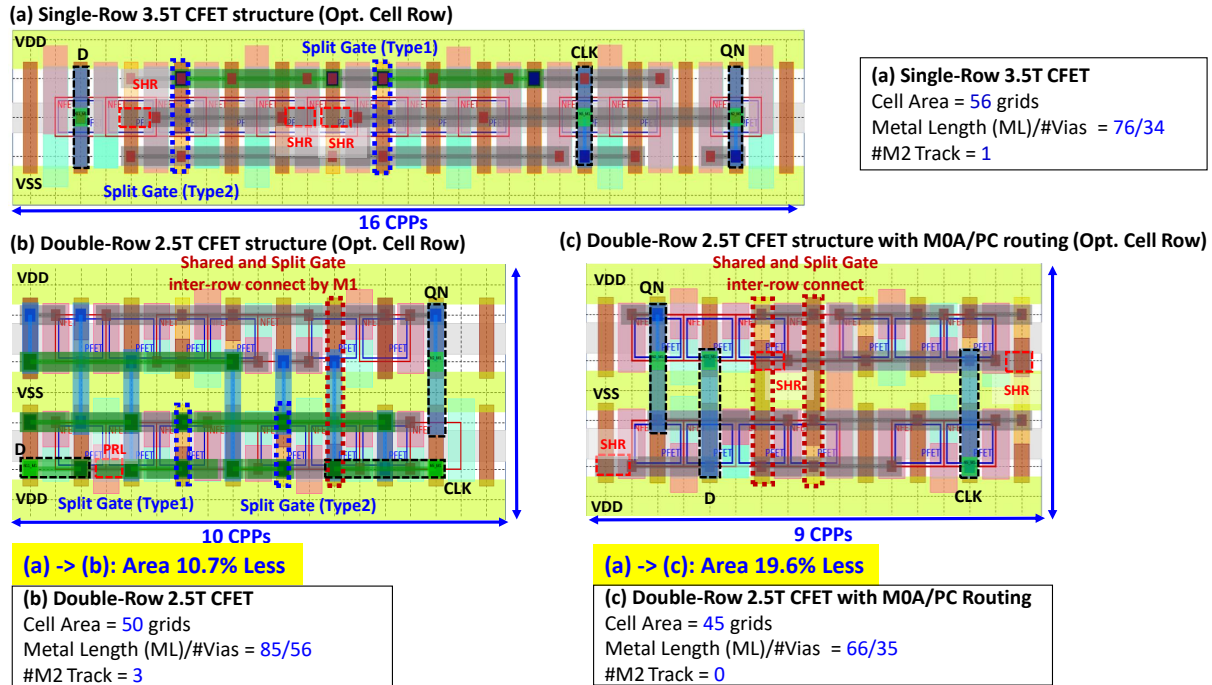


Figure 3.23: Layouts of design corrected DFFHQN layouts of (a) Single-Row 3.5T CFET, (b) Double-Row 2.5T CFET, and (c) Double-Row 2.5T CFET with M0A/PC routing. The Metal Length is weighted sum of metal segments and vias. Optimized results of scaling (a) Single-Row 3.5T CFET layout to (b) Double-Row 2.5T CFET layout: Cell Area (56→50), Metal Length (76→85), #Vias (34→56), and #M2Track (1→3); Optimized results of scaling (a) Single-Row 3.5T CFET layout to (c) Double-Row 2.5T CFET with M0A/PC routing layout: Cell Area (56→45), Metal Length (76→66), #Vias (34→35) and #M2Track (1→0).

optimized XOR2x1 SDC layout. Figure 3.23 (a) and (c) shows the design corrected DFFHQN layouts of 3.5T CFET and 2.5T M0A/PC-R CFET, respectively. The double-row 2.5T M0A/PC-R CFET cell structure achieves 19.6% and 10.0% smaller cell area than 3.5T CFET and double-Row 2.5T CFET cell architecture.

Last, Figure 3.24 (a) summarizes the average cell area benefit of the representative 30 SDCs by track number reduction (i.e., 3.5T to 2.5T) and M0A/PC routing option. Note that the cell areas of AOI21x1, AOI22x1, OAI21x1, and OAI22x1 with 2.5T CFET are still larger than 3.5T CFET due to the severe in-cell routing congestion. With enabling M0A/PC layers routing (i.e., 2.5T M0A/PC-R CFET) for maximizing the area benefit of track number reduction, all SDC areas are smaller than 3.5T CFET.

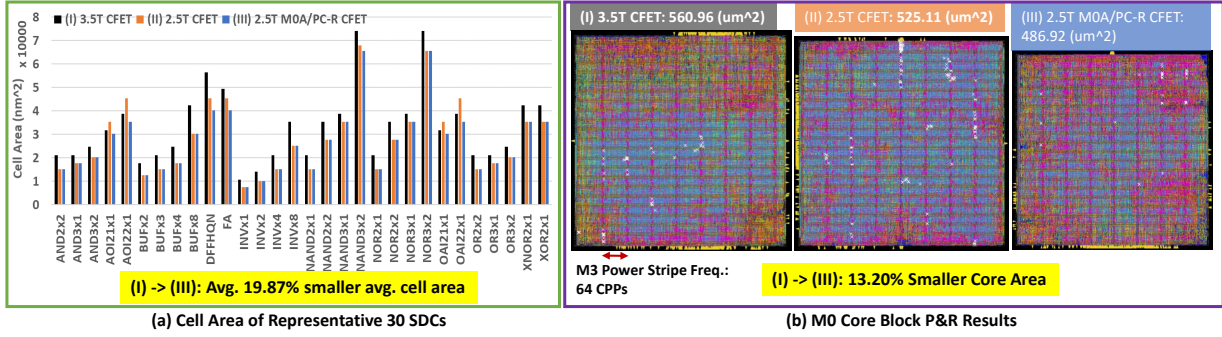


Figure 3.24: Cell and block-level area benefits by track reduction and M0A/PC routing: (I) 3.5T CFET (black bar), (II) 2.5T CFET (orange), (III) 2.5T M0A/PC-R CFET (blue). (a) Cell Area of Representative 30 SDCs. (b) Block-level P&R results of M0 Core. The core area is improved by 13.20% by track number reduction and using M0A/PC for routing. The red arrow shows the 64 CPPs M3 power stripe grid. CellArea = $CW \times CPP \times CH \times M2Pitch$, $CPP=42nm$, $M2Pitch=24nm$.

3.6.2 Block-Level Area Scaling with 2.5T CFET

We compare the block-level areas of 3.5T CFET SDCs, 2.5T CFET SDCs, and 2.5T M0A/PC-R CFET SDCs from Exp. 3.6.1 using three open source RTL designs [33]: M0 Core, M1 Core, and AES¹⁸. For BEOLs, the design rule are set as described in section 3.3 and M2-M7 are used for block-level routing. For power delivery network, we set up top power mesh and intermediate power stripes as described in section 3.3. In addition, to avoid dropping the SuperVia [35] on the Upper/Lower M0A/PC layers, which are used by inter-row routing, for connecting the BPR in the block-level, we extract Upper/Lower M0A/PC layers as blockages in the block-level.

The block-level P&R results of 3.5T CFET, 2.5T CFET, and 2.5T M0A/PC-R CFET are shown in Table 3.9. The valid minimum block-level area is obtained using 300 #DRVs threshold [7]. Compared to 3.5T CFET, the average minimum block-level area of M0 Core, M1 Core, and AES are reduced by 6.29% for 2.5T CFET and 13.43% for 2.5T M0A/PC-R CFET; the average total wirelength is also reduced by 7.65% for 2.5T CFET and 14.40% for 2.5T M0A/PC-R CFET. Figure 3.24 (b) shows that 2.5T M0A/PC-

¹⁸The worst negative slacks of M0 Core, M1 Core, and AES are carefully adjusted between 50 and -50ps for a fair comparison in the block-level analysis.

Table 3.9: Block-level placement and route results of 3.5T CFET, 2.5T CFET, and 2.5T CFET M0A/PC-R: #Inst=Number of Instances, SDC Area=Standard Cell Area, Total WL=Total Wirelength, Min. Area=Minimum Valid Block-Level Area, Area Impr.=(Min. Area of 3.5T CFET - Min. Area of 2.5T CFET)/(2.5T M0A/PC-R CFET)/(Min. Area of 3.5T CFET).

Design	#Inst	3.5T		2.5T		2.5T M0A/PC-R		Core Area Impr. (%)	
		Total WL (um)	Min. Area (um ²)	Total WL (um)	Min. Area (um ²)	Total WL (um)	Min. Area (um ²)	2.5T	2.5T M0A/PC-R
M0 Core	17K	44242.24	560.96	41121.82	525.11	36411.02	486.92	6.39	13.20
M1 Core	20K	47072.22	687.94	41638.28	642.08	38983.24	574.03	6.67	16.56
AES	14K	30094.68	416.67	29365.58	392.48	28531.86	372.75	5.81	10.54
Avg.	17K	40469.71	555.19	37375.22	519.89	34642.04	477.90	6.29	13.43

R CFET achieves 13.20% smaller core area than 3.5T CFET for M0 Core design. This area benefit comes from further cell area reduction by connecting shared-and-split structure across cell rows through M0A/PC layers.

In summary, we show that 2.5T M0A/PC-R CFET can not only achieve 20.61% smaller cell area on average but also provide 13.43% and 14.40% less block-level area and total wirelength on average, respectively, compared to 3.5T CFET SDCs. Leveraging the direct connection of shared-and-split structures between cell rows with M0A/PC layers can maximize the cell and block-level area benefits of reducing cell height to 2.5T.

3.7 Conclusion

We propose an SMT-based Multi-Row CFET SDC synthesis framework, which supports track number reduction, design rule selections, multi-row architectures, and different stacking options, for fast and holistic STCO and DTCO explorations on cell area and block-level area scaling. The novel Multi-Row Dynamic Complementary Pin Allocation scheme enables the exploration of using Upper/Lower M0A/PC for inter-row routing to maximize the advantage of CFET shared and split structure across cell

rows. In addition, the novel multi-row cell area objective explores single-row and multi-row structures together and generates the minimum cell area with optimum cell row. We firstly demonstrate that the proposed novel cell area objective achieves 20.69%, 8.37%, and 3.33% smaller SDC cell areas on average compared to triple-row, double-row, and single-row [7] structures, respectively. For routability, the proposed routability-driven objectives/constraints successfully reduce up to 48% #DRVs at the block-level compared with [6] as scaling 4.5T CFET to 3.5T CFET architecture. Then, through extensive DTCO explorations on ground design rules and #BEOLs, 3.5T CFET SDCs achieve up to 6.50% smaller block-level areas than 4.5T CFET SDCs. With the assistance of STCO and DTCO, 3.5T CFET SDCs achieve 21.0% on average in reduced block-level areas compared to 4.5T Conv. SDCs. Lastly, in the extreme CFET SDC scaling studies, we firstly demonstrate that enabling Upper/Lower M0A/PC for inter-row routing can achieve 20.61% smaller cell area on average when scaling 3.5T to 2.5T cell structure. Then, we show that the 2.5T CFET with M0A/PC layers for inter-row routing achieves 13.43% and 14.40% less block-level area and total wirelength on average compared to 3.5T CFET, respectively.

This chapter contains materials from "Complementary-FET (CFET) standard cell synthesis framework for design and system technology co-optimization using SMT", by Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, Bill Lin, and Dongwon Park, which appears in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2021; "Multirow Complementary-FET (CFET) Standard Cell Synthesis Framework Using Satisfiability Modulo Theories (SMTs)", by Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, and Bill Lin, which appears in IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, 2021. The dissertation author was the primary investigator and author of these papers.

Chapter 4

Machine Learning Prediction for Design and System Technology Co-Optimization

Sensitivity Analysis

4.1 Introduction

As VLSI technology continues to advance relentlessly beyond $5nm$, geometric pitch scaling starts to slow down. Moreover, design technology co-optimization (DTCO) [45] based on pitch scaling and patterning is unable to continue the cost scaling in 2D IC technology. In order to keep the trend of Moore's Law, system technology co-optimization (STCO) has been introduced to assist DTCO scaling with 3D integrated logic and novel 3D cell structure (i.e., Complementary-FET (CFET)) [4, 38] beyond $5nm$. However, technology development beyond sub $5nm$ demands enormous engineering effort for identifying the optimal technology options (i.e., evaluation of cost and determination of standard cell

(SDC) heights, 2D/3D SDC architectures, design rules, power delivery networks (PDNs), and back end of line (BEOL) settings). Furthermore, process architects must be aware of the impact of the technology transition on the power, performance, area, and cost (PPAC) for further optimization.

Therefore, finding the optimal technology option necessitates numerous DTCO and STCO iterations among SDC optimization, design rule optimization, and block level area evaluation. This results in exploding turnaround-time (TAT) in DTCO and STCO explorations. There is a high demand for a holistic, fast, and robust prediction methodology that provides information on the potentially optimal technology options and the impact on PPAC from the technology transition.

4.1.1 Related Works

The related works can be categorized into DTCO and STCO frameworks, and machine learning (ML)-based DTCO and STCO approaches.

DTCO and STCO frameworks. In [46], Song et al. proposed a unified technology platform using integration analysis for DTCO and STCO at sub $7nm$ node. Kahng et al. [47] proposed a routability metric k_{th} to evaluate the routing capacity of BEOL stacks, but this work lacks of explorations on various cell structures, and does not provide the change of block-level metrics (i.e., area) from the technology transition. Recently, in [27], a novel design rule evaluation technique using automatic cell layout generation for DTCO exploration is proposed, but the focus is limited to conventional FET (Conv. FET) structures (i.e., FinFET). Cheng et al. [7] also proposed a novel automatic CFET cell layout synthesis framework for DTCO and STCO explorations. However, these works use conventional block-level placement-and-route (P&R) to evaluate block area, and thus result in longer TAT for DTCO and STCO technology development.

Machine learning (ML)-based DTCO and STCO approaches. Recently, many ML-based DTCO and

STCO approaches have been proposed to shorten the DTCO and STCO exploration time. In [48], an ML-based modeling framework is developed to generate compact models of novel devices, but this work does not consider block-level evaluations. Ceyhan et al. [49] used ML techniques to search and find optimal combinations of design, technology, and flow recipes for high-performance CPU designs in the enormous solution space, but its performance on 3D SDC architectures (i.e., CFET) has not been explored, and the methodology requires 4 to 6 weeks of TAT. Recently, Cheng et al. [50] extended routability metric, k_{th} , to cell-level and block-level, and applied ML-assisted prediction on k_{th} for various technology options. However, we focus on exploring technology options with Conv. FET SDC structure and does not predict/provide the changes in block-level metrics (i.e., area) induced by the technology transition from one option to another in this work. In [12], a modeling approach for DTCO and STCO sensitivity prediction have been proposed, but they performed limited exploration on machine learning models.

4.1.2 Our Contributions

In this paper, we propose a novel DTCO and STCO sensitivity prediction framework, which provides information on the change/gradient of block-level metrics from the technology transition. Also, we develop a machine learning model that combines bootstrap aggregation and gradient boosting techniques to improve the prediction accuracy. Figure 4.1 shows the difference between (a) the traditional DTCO and STCO exploration flow and (b) the proposed DTCO and STCO sensitivity prediction framework. The proposed prediction model and automatic cell synthesis [7, 10] significantly reduce the TAT of DTCO and STCO explorations on various physical layout factors: (i) SDC library sets (i.e., different cell heights, Conv. FET and CFET SDC architectures), (ii) design rules (DR), (iii) back end of line (BEOL) parameters, and (iv) power delivery network (PDN) configurations. In this work, we focus on the sensitivity of block-level area variations according to different technology features and demon-

strate the feasibility of machine learning techniques in DTCO and STCO exploration flows¹⁹. Our main contributions are as follows.

- We propose a novel DTCO and STCO sensitivity prediction framework that improves the efficiency of explorations by orchestrating the proposed machine learning model and automatic cell synthesis [7, 10].
- We develop a machine learning model using bootstrap aggregation and gradient boosting techniques to predict the change/gradient of block-level metrics from the technology transition.
- We perform extensive studies on various machine learning algorithms for block-level area sensitivity prediction, and demonstrate that the developed machine learning model outperforms other machine learning algorithms on DTCO and STCO sensitivity prediction.
- We identify key features of each SDC and extract cell and block-level features for prediction. We validate the extracted features via feature importance analysis in Exp. 4.3.2.
- We perform extensive studies on model accuracy for new technologies and model robustness for new designs across Conv. FET and CFET SDC architectures, various cell heights, design rules, power delivery networks (PDNs) and BEOL settings.

The remaining sections are organized as follows. Section 4.2 describes our DTCO and STCO sensitivity prediction approach. Section 4.3 presents our main experiments. Section 4.4 concludes the paper.

¹⁹DTCO and STCO sensitivity prediction for incorporating block-level power and performance are one of the future works as discussed in Section 5.2

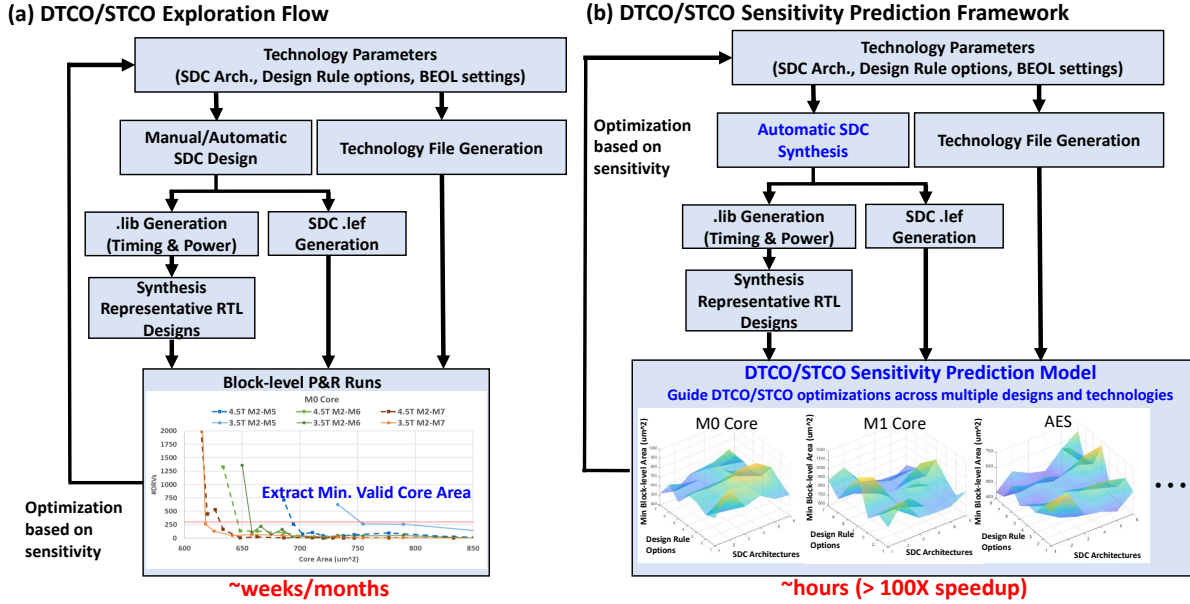


Figure 4.1: The illustrations of (a) Traditional DTCO and STCO exploration flow. (b) The proposed DTCO and STCO sensitivity prediction framework. We use [7], and [10] for the automatic SDC synthesis here.

4.2 Design and System Technology Co-Optimization Sensitivity Prediction Framework

We apply machine learning techniques to predict the sensitivity of DTCO and STCO explorations on block-level areas considering various physical layout factors: (i) SDC architectures (e.g., cell height, multi-row/single-row, CFET, and Conv. FET.), (ii) design rules (DRs), (iii) BEOL parameters, and (iv) power delivery network configurations. In this section, we describe the specifics of our prediction methodology: (i) DTCO and STCO Sensitivity, (ii) overall modeling flow, (iii) methodology for feature extraction, (iv) input features, and (v) machine learning techniques.

4.2.1 DTCO and STCO Sensitivity

The DTCO and STCO sensitivity for block-level area of two technologies of a block-level circuit is the percentage of the block-level area difference of these two technologies, $\Delta A_{i,j}$, as shown in

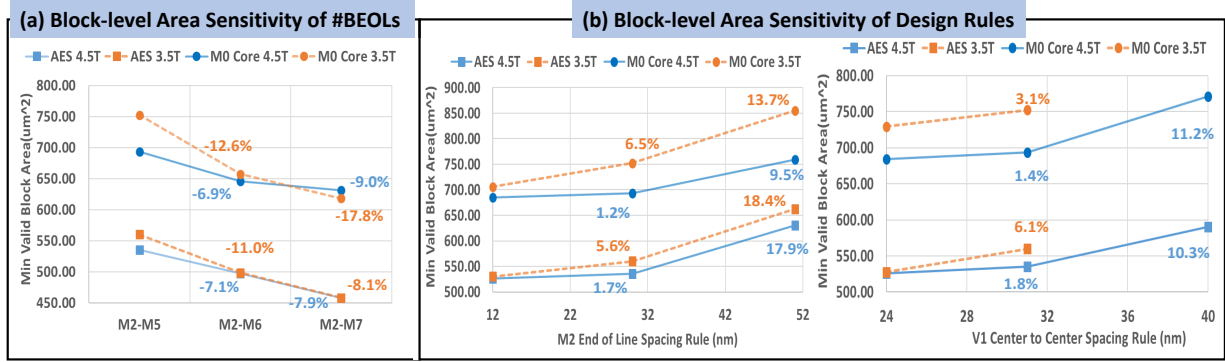


Figure 4.2: An example of DTCO and STCO block-level area sensitivity of (a) #BEOLs and (b) design rules using 4.5T and 3.5T CFET SDC library sets of AES and M0 Core circuits. The number represents the block-level area difference as changing DTCO and STCO parameters from left to right. Many 3.5T CFET SDCs (i.e., NAND2x2, NAND3x1, etc.) don't have feasible solutions when V1 center to center spacing is 40 nm [7]. As a result, there is no data points of the block-level area of 3.5T CFET using 40 nm V1 center to center spacing rule.

Equation (4.1).

$$\Delta A_{i,j} = (A_i - A_j)/A_i \quad (4.1)$$

where A_i and A_j are the minimum valid block-level areas of the i^{th} technology and the j^{th} technology. Here, a technology is the combination of SDC library set, design rules, BEOL parameters, and PDN configuration. Figure 4.2 shows an example of DTCO and STCO block-level area sensitivity of (a) #BEOLs and (b) design rules using 4.5T and 3.5T CFET SDC library sets of AES and M0 Core block-level circuits. Different SDC library sets, #BEOLs, and design rules can potentially impact the block-level area up to 17.9%. The importance of knowing the information of change/gradient of block-level area from the technology transition is needed for holistic technology development.

In this work, we focus on studying the minimum block-level area of various cell heights, cell architectures (i.e., CFET and Conv. FET), cell pin-accessibility, design rules, BEOL parameters, and power delivery network structures and develop a model to predict the $\Delta A_{i,j}$ for reducing the TAT of

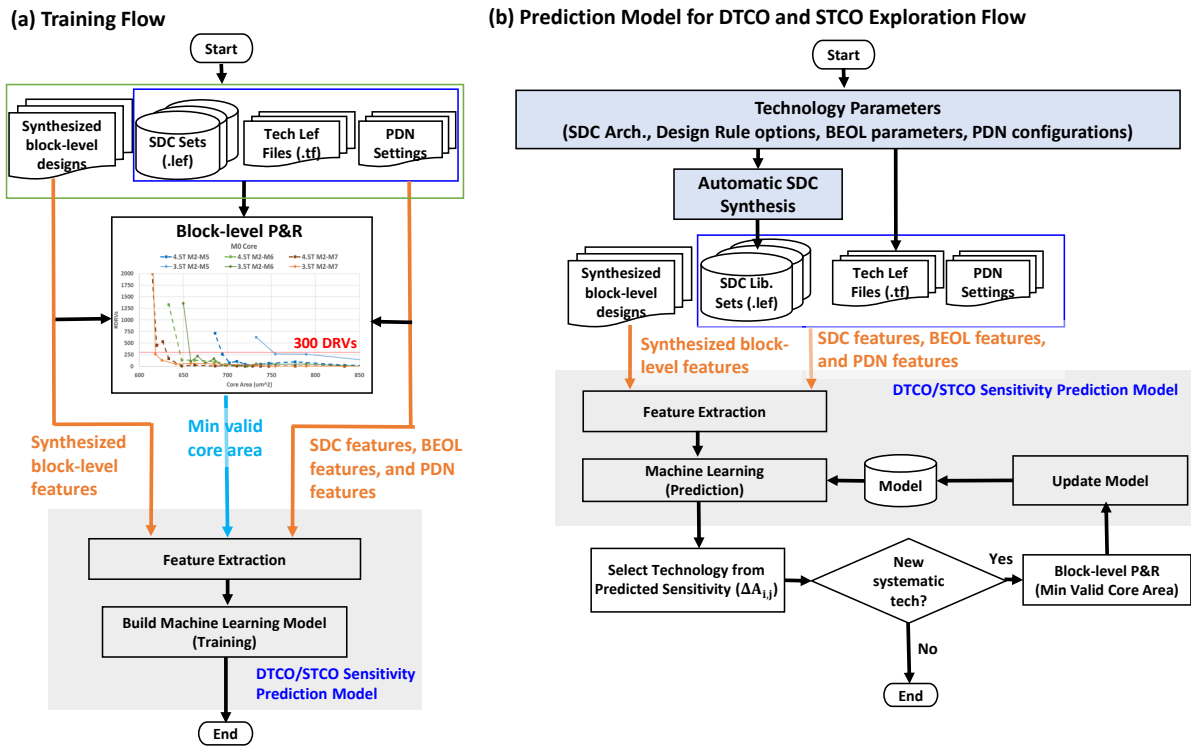


Figure 4.3: Overall flow of DTCO and STCO sensitivity prediction: (a) Training flow, and (b) Prediction model for DTCO and STCO exploration flow. Technology developers can select the optimal technology candidate which provides the largest improvement of block-level area metric compared to baseline technology from the predicted $\Delta A_{i,j}$ in (b). If the selected technology is new technology, which brings systematic physical layout change at block-level, the model can be updated with the block-level P&R data of the new technology. Here, We use [7] , and [10] for the automatic SDC synthesis.

DTCO and STCO explorations.

4.2.2 Overall modeling flow

Figure 4.3 shows the proposed training flow and prediction model for DTCO and STCO exploration flow. In the training phase as shown in Figure 4.3(a), we generate multiple SDC library sets, BEOL parameters, and power delivery configurations to perform multiple block-level P&R runs with the synthesized block-level circuits through a commercial P&R suite [14]. The minimum valid block-level area of a technology combination (i.e., SDC library set, design rules and BEOL combination), is extracted

with 300 design rule violations (#DRVs)²⁰ with multiple P&R runs as shown in Figure 4.1(a). Then, the percentage of the block-level area difference of two technologies (i.e., $\Delta A_{i,j}$) are extracted in the feature extraction stage for the training.

We show the prediction model for DTCO and STCO exploration flow in Figure 4.3(b). Prediction flow utilizes the same input types with new technology parameters to explore. The proposed DTCO and STCO sensitivity prediction model outputs the predicted $\Delta A_{i,j}$.

In our envisioned usage scenario, technology developers define and generate multiple circuit designs, SDC library sets, tech lef files (.tf), and PDN configurations for DTCO and STCO explorations. The proposed framework assists and guides the technology tuning process to find one of the optimal technology candidates by predicting the gradient of the block-level area, $\Delta A_{i,j}$, for block-level area cost evaluation. With the predicted $\Delta A_{i,j}$ of all the technology pairs, technology developers can find the technology which provides the largest improvement on block-level area metric compared to the baseline technology. If the selected technology combination is a new technology, which involves systematic physical layout change (i.e., backside PDN technology), to the prediction model, the block-level P&R is launched to extract the minimum valid block-level area and the data is used to update the prediction model. Otherwise, technology developers adopt the selected technology for the next phase in technology development.

4.2.3 Methodology for feature extraction

We describe the feature extraction component of our framework. Table 4.1 summarizes four categories of input features: (i) synthesized block-level circuit statistics, (ii) SDC architectures, (iii) BEOL parameters, and (iv) power delivery network configurations.

Synthesized block-level circuit statistics. We extract the statistics of the block-level circuit, which is

²⁰As a common industrial practice, once the number of DRVs increases beyond 300, the block layout is deemed too troublesome to fix with laborious engineering change orders (ECOs).

derived after logic synthesis and before physical layout. The data includes circuit structures, instance numbers, and standard cell area from the synthesized block-level circuit. For circuit structures, we consider the distribution of fanout counts ($\#fanouts$), and the Rent’s multiplier k and exponent p component of Rent’s Rule [51]. These terms define an empirical power-law relationship between number of gates “ N ” and number of terminals “ T ” as shown in Equation (4.2).

$$T = kN^p \quad (4.2)$$

For each circuit, we extract the (T, N) pairs, and perform linear regression to obtain the k and p for each design. In addition, we extract the number of fanouts per net ($\#fanouts$), number of sequential cells ($\#Seq$), number of Combinational cells ($\#Comb$), and number of buffers ($\#Buf$), and SDC area from the report of synthesis tool.

SDC architectures. We extract key metrics which impact routability at the block-level and lead to larger minimum valid block-level area, such as average Remaining Pin Access (RPA) value [6, 29] of I/O pins, number of M2 Track usage ($\#M2Track$) [7], and M2 metal length ($M2ML$) [7] in the cell level. Then, we calculate the block weighted RPA_d , $M2Track_d$, and $M2ML_d$ with the corresponding SDC metrics and cell percentage of the synthesized block-level circuit d using Equation (4.3) [7].

$$Metric_d = \sum_c Metric_c * CP_{d,c} \quad (4.3)$$

where $Metric_d$ denotes the block weighted metric of design d . $Metric_c$ is the cell level metric of cell c , such as average RPA value, $\#M2Track$, and $M2ML$. The $CP_{d,c}$ is the percentage of cell c in the synthesized block-level circuit d . In addition, we use cell height as one of the features since the cell height limits the horizontal routing tracks/resources, which shows greater impacts on SDC less than $5T$ [6], for accessing

M1 pin in SDC.

BEOL parameters. We introduce BEOL parameters related to the design rule and BEOL settings. We use representative design rules such as min spacing rule, end-of-line spacing rule (EOL), via rule (VR), same net VR, and fat metal spacing rule for metal and via layers as the input features. For BEOL settings, the pitch of each routing metal layer and the total number of routing metal layers (#BEOLs) are selected as the input features of our model.

Power delivery network (PDN) configurations. We categorized the PDN into front side PDN and backside PDN categories [36]. For front side PDN, we mainly study the M3 power strap period, which is critical to the power integrity and signal routing. With a denser M3 power strap, the IR drop will be improved, but it will result in the poor routability and a larger core area because it takes more metal resources for signal routing. On the other hand, a sparser M3 power strap may lead to a power integrity issue and causes functional failure. For backside PDN, we set power strap period feature to a large number (i.e., $1e^6$), since there are no power straps on the front side at block-level.

4.2.4 Input features

We describe the input features used to predict the $\Delta A_{i,j}$ in Equation (4.1). Figure 4.4 shows an illustration of the input features of the proposed DTCO and STCO sensitivity prediction model. The input features consist of the extracted features, which are shown in Table 4.1, of i^{th} and j^{th} technologies.

4.2.5 Machine learning techniques

We develop our machine learning model with bootstrap aggregation and gradient boosting regression tree techniques to achieve state-of-the-art results on DTCO and STCO sensitivity prediction. We introduce the overview of the proposed model, the feature selection technique, and the modeling approach

Table 4.1: Extracted Features Table

Feature Scope	Feature Types	Feature Name
Synthesized block-level circuit design statistics	Net complexity	#Fanouts
		Rent's multiplier
		Rent's exponent
	Instance	#Seq
		#Comb
		#Buf
	Synthesized Design Area	SDC area
SDC features	Block weighted SDC metric	RPA_d
		M2Track_d
		M2ML_d
	Horizontal Routing Resource	Cell Height
Design Rule & BEOL settings	Design Rules for Metal and Via layers	Min spacing
		EOL
		VR
		Same net VR
		Fat metal spacing
	BEOL settings	BEOL Pitches
		#BEOL
Power Delivery Network (PDN) features	PDN settings	M3 power strap period

below.

Model overview. Figure 4.5 shows the developed machine learning model, which combines bootstrap aggregation and gradient boosting regression tree techniques. In the bootstrap aggregation technique, the bootstrap sampling is used to estimate statistics on a population by sampling a data set with replacement, and can be used to create meaningful simulated data sets to control the variance of a model. Then, the simulated data sets are used to train a set of gradient boosting regression tree (GBRT) models. Lastly, the outputs of GBRT models are aggregated for predicting DTCO and STCO sensitivity. We

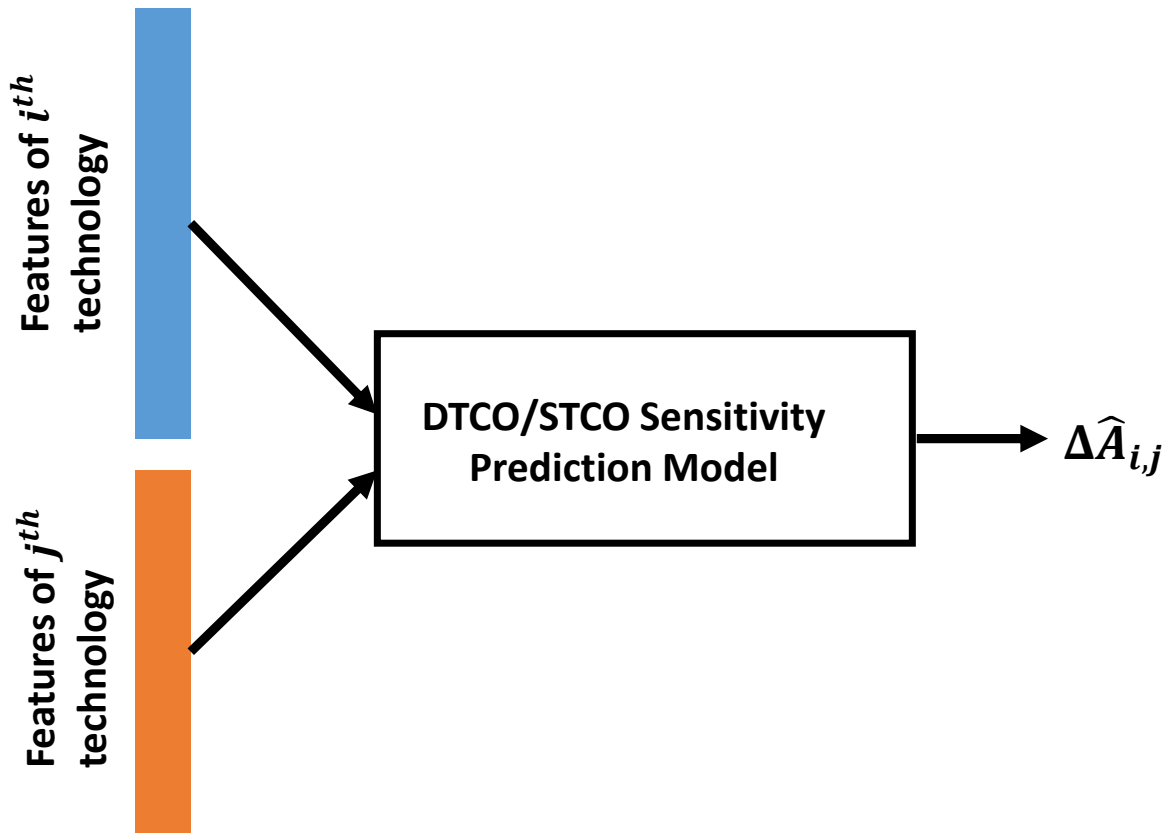


Figure 4.4: An illustration of input features of the proposed DT/STCO sensitivity prediction model. Features of the i^{th} and j^{th} technologies are described in Table 4.1.

use XGBoost [52] for implementing the GBRT models in the proposed model. XGBoost implements machine learning algorithms using a GBRT, which achieves state-of-the-art results on tabular data prediction. To avoid structural similarity of GBRT trees and have a high correlation of their predictions, we set `colsample_bytree`²¹ to 0.7 for each GBRT model. Finally, the final predicted $\Delta \hat{A}_{i,j}$ are calculated by averaging the prediction of all GBRT models.

Feature selection technique. We describe the feature selection technique here. Firstly, we extract the feature importance of a trained GBRT model. Then, we use the Variance Inflation Factor (VIF) [53] to

²¹`colsample_bytree` is the fraction of features (randomly selected) that will be used to train each tree in the XGBoost library [52].

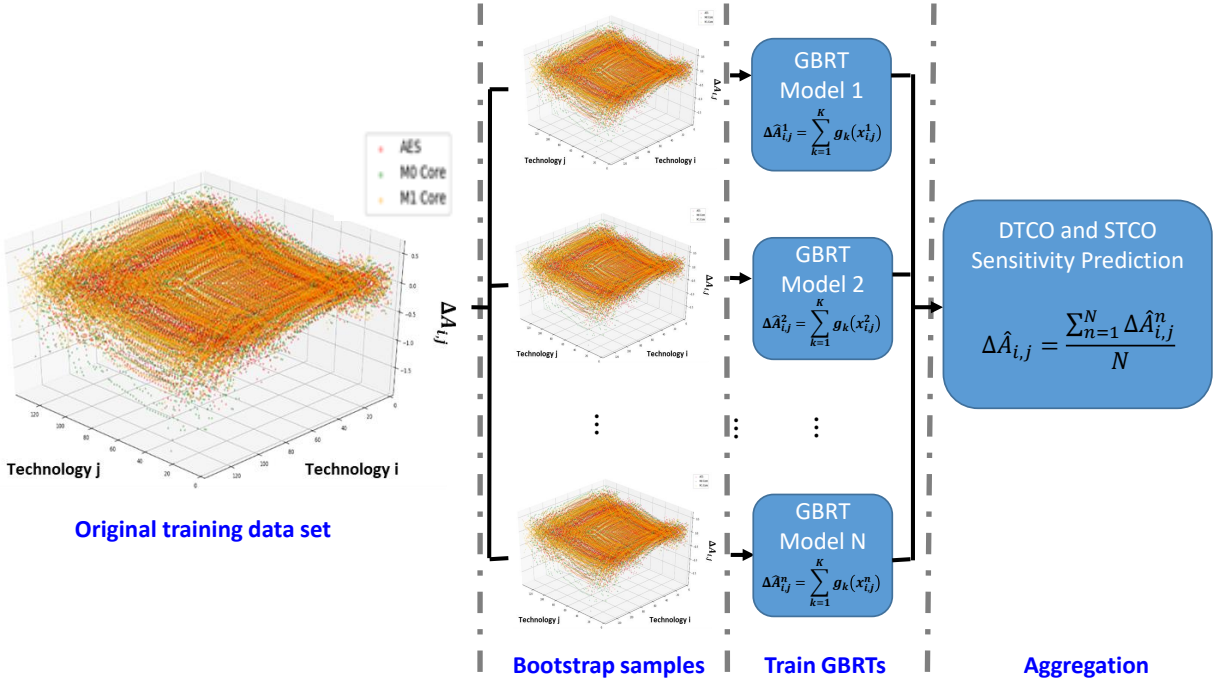


Figure 4.5: Overview of the developed machine learning model. The model combines bootstrap aggregation and gradient boosting regression tree (GBRT) techniques.

detect instances of multicollinearity, which result in the high sensitivity to small changes in correlated features. Finally, we perform feature selection as described in Algorithm 4. Here, we use the “gain” for feature importance. The gain of a leaf node is the difference of metric before and after splitting at the leaf node [52]. The “gain” of the feature is the total gain of using the feature to split nodes divided by the number of times the feature used to split a node. The feature selection technique reduces average MAE by 0.02 (i.e., 25%) for new design prediction in Exp. 4.3.5.

In Algorithm 4, firstly, we split the data set D into training set, T , and validation set, V (Line 1). We train a Gradient Boosting Regression Tree (GBRT) model with training and validation sets (Line 2). Then, we sort the features based on the gain of the GBRT model in descending order (Line 7) After that, we sequentially add features to F_{sub} and train a GBRT model with selected features F_{sub} (Line 9-12). Then, we calculate VIF of each feature in data set T_{sub} and extract the VIF, $f_{i,vif}$, of f_i (Line 13). If the

validation error, E_{val} , is larger than the minimum validation error, E_{val}^{min} , and the VIF of f_i is larger than a VIF threshold, we remove f_i from F_{sub} (Line 14-16). If the validation error is smaller than the minimum validation error, we record F_{sub} as \hat{F} and update the minimum validation error (Line 17-20). Lastly, we return the feature subset, \hat{F} , which has minimum validation error (Line 22).

Algorithm 4 Feature Selection

```

/*Input: Data set  $D$ , and Feature set  $F$ ; Output: Feature subset  $\hat{F}$ .*/
1: Split data set  $D$  into 80% training set,  $T$ , and 20% validation set,  $V$ ;
2: Train a model with  $T$  and  $V$  with  $F$  using GBRT with early stopping;
3: Get the validation error,  $E_{val}$ ;
4: Set  $E_{val}^{min} = E_{val}$ ;
5: Set  $\hat{F} = F$ ;
6: Set  $F_{sub} = \{\}$ ;
7: Set  $F = \text{Sort } F \text{ based on the gain of features in descending order}$ ;
8: Set  $m = |F|$ ;
9: for  $i = 1, 2, \dots, m$  do
10:   Set  $F_{sub} = F_{sub} + f_i$ ;
11:   Extract  $F_{sub}$  from  $T$  and  $V$  to  $T_{sub}$  and  $V_{sub}$ , respectively;
12:   Train a model with  $T_{sub}$  and  $V_{sub}$  with  $F_{sub}$  using GBRT with early stopping;
13:   Get validation error,  $E_{val}$ ;
14:   Calculate the VIF of each feature in data set  $T_{sub}$  and get  $f_{i,vif}$  value;
15:   if  $E_{val} > E_{val}^{min}$  &&  $f_{i,vif} \geq VIF_{th}$  then
16:     Remove  $f_i$  from  $F_{sub}$ ;
17:   end if
18:   if  $E_{val} \leq E_{val}^{min}$  then
19:     Set  $\hat{F} = F_{sub}$ ;
20:     Set  $E_{val}^{min} = E_{val}$ ;
21:   end if
22: end for
23: Return  $\hat{F}$ 

```

Modeling approach. We extract the input features from all the technologies as shown in Section 4.2.4, compose all the technologies into pairs, and perform feature selection to compose a data set $D = (x_{i,j}, \Delta A_{i,j})$, where $x_{i,j} \in \mathbb{R}^m$ corresponds to the m input features after feature selection, and $\Delta A_{i,j} \in \mathbb{R}$ is the percentage of the block-level area difference of i^{th} and j^{th} technologies. We aim to predict the $\Delta A_{i,j}$ using the developed machine learning model. The D is resampled to generate N data sets, \hat{D}^n . We increase the number of samples until each bootstrap sample (i.e., \hat{D}^n) contains approximately 63.2% of the data points in the

training set [54].

For each GBRT model, XGBoost sequentially builds an ensemble of K regressors. Predictions, $\Delta\hat{A}_{i,j}^n$, are made by taking the weighted sum of predictions made by the individual members of the ensemble as shown in Equation (4.4).

$$\Delta\hat{A}_{i,j}^n = \sum_{k=1}^K g_k(x_{i,j}^n), g_k \in G \quad (4.4)$$

where G is the space of regression trees, and n represents the n^{th} GBRT model. The goal is to minimize $L(\Delta A_{i,j}^n, \Delta\hat{A}_{i,j}^n)$ in Equation (4.5).

$$L(\Delta A_{i,j}^n, \Delta\hat{A}_{i,j}^n) = \sum_i l(\Delta A_{i,j}^n, \Delta\hat{A}_{i,j}^n) + \sum_k \Omega(f_k) \quad (4.5)$$

where $\Omega(f) = \gamma T + \frac{1}{2} \lambda ||w||$

where each $l(\Delta A_{i,j}^n, \Delta\hat{A}_{i,j}^n)$ is a differentiable convex function that measures the difference of $\Delta A_{i,j}^n$ and $\Delta\hat{A}_{i,j}^n$. We use mean absolute error (MAE) as the evaluation metric. Ω is a function that penalizes the complexity of the model. T is the number of leaves in the tree and w is the leaf weight. We use 10-fold cross-validation [55] to perform hyperparameter tuning (i.e., *min_child_weight*, *eta*, etc.) to train our model. Then, to predict $\Delta A_{i,j}$, the $\Delta\hat{A}_{i,j}$ is obtained using the average of the prediction results of N GBRT model, $\Delta\hat{A}_{i,j}^n$, as shown in Equation (4.6).

$$\Delta\hat{A}_{i,j} = \frac{\sum_{n=1}^N \Delta\hat{A}_{i,j}^n}{N} \quad (4.6)$$

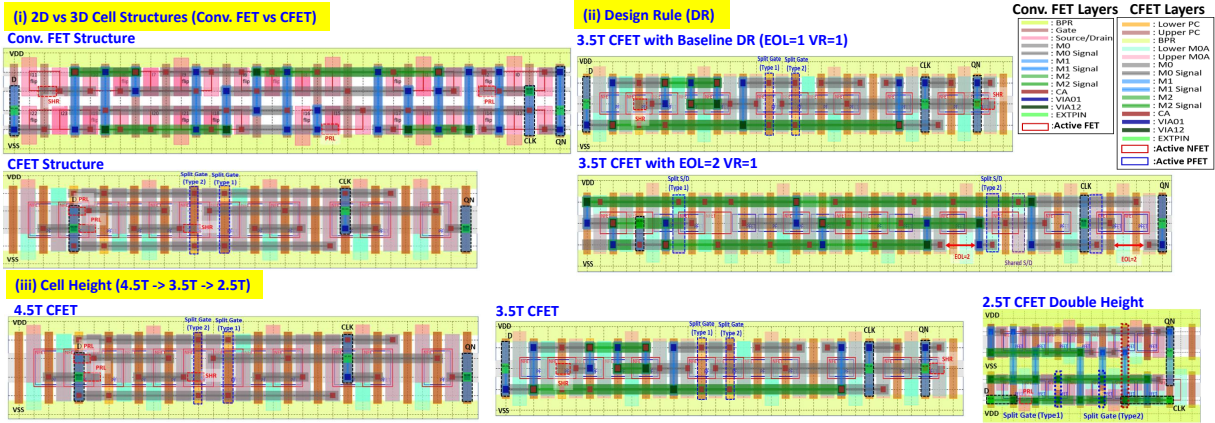


Figure 4.6: An example of generated DFFHQ SDC layouts with variations on three dimensions: (i) Cell Structure (CS), (ii) Design Rule, and (iii) Cell Height (CH).

Table 4.2: Synthesized block-level circuit table.

Design Name	#Instance	Rent's multipliers	
		(k)	(p)
M0 Core	17k	2.69	0.73
M1 Core	20k	2.72	0.71
AES	14k	2.62	0.70
MPEG	18k	3.58	0.61
JPEG	45k	2.71	0.78
Darkiscv	7k	5.78	0.25

4.3 Experimental Results

Our framework is implemented in Python and is executed on a workstation with 2.4GHz Intel Xeon E5-2620 CPU and 256GB memory. For the proposed model in Figure 4.5, we implement the bootstrap sampling technique with sklearn library [56], and GBRT tree models with XGBoost library [52].

4.3.1 Experiment Setup

We use the synthesized block-level circuits, SDC library sets generated from [7, 10], design rules, BEOL settings, and power delivery network configurations to generate the data for our experiments. We

run multiple block-level P&R runs through a commercial test suite [14] and use a 300 #DRV threshold to measure the minimum valid block-level area of each synthesized block-level circuit for each technology combination.

Synthesized block-level circuits. For synthesized block-level circuits, 6 open source RTL designs [33], M0 Core, M1 Core, AES, MPEG, JPEG, and DarkRiscV that respectively have 17K, 20K, 14K, 18K, 45K, and 7K instances using 30 representative SDCs [7]. The worst negative slack (WNS) of each synthesized block-level circuit is carefully adjusted between +/- 50ps for a fair comparison to study the change of minimum block-level area of various cell heights, cell architectures (i.e., CFET and Conv. FET), cell pin-accessibility, design rules, BEOL parameters, and power delivery network structures.

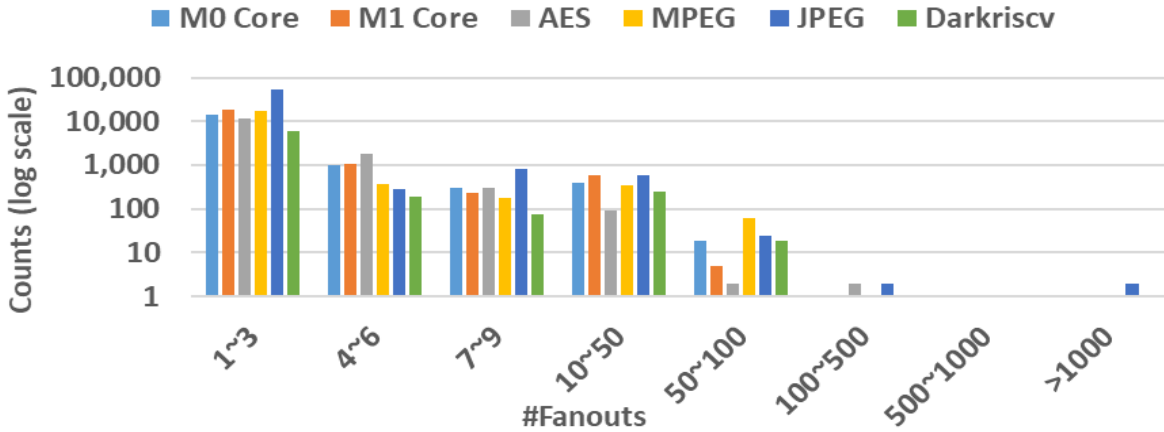
The Rent's multipliers, k , and the Rent's exponent, p , of each design are listed in Table 4.2. For the number of fanouts per net (#fanouts), we categorize the number of fanouts per net into 8 bins, which are 1-3 #fanout nets, 4-6 #fanout nets, 7-9 #fanout nets, 10-50 #fanout nets, 50-100 #fanout nets, 100-500 #fanout nets, 500-1000 #fanout nets, and more than 1000 #fanout nets. Figure 4.7 shows the (a) #Fanouts distribution, and (b) cell statistics of these 6 block-level circuits.

SDC library sets generation. To evaluate the block-level PPA during early DTCCO exploration, we select 30 representative SDCs [7]. We generate 19 SDC library sets with 4.5T, 3.5T, and 2.5T cell heights, different EOL and VR design rule parameters, and two cell architectures (i.e., CFET and Conv. FET) using [7, 10]. The top layer is M2 for SDC generation. We generate SDC library sets with variations on three dimensions as follows.

1. **Cell structures:** we generate Conv. FET and CFET SDC layouts for explorations on 2D and 3D cell structure in the experiments.
2. **Cell Height:** The CFET SDC cell height is scaling from 4.5T to the extreme 2.5T cell height [7,57].

For Conv. FET SDC, we generate 4.5T and 3.5T cell height because using 2 horizontal routing

(a) #Fanouts distribution



(b) Cell statistics

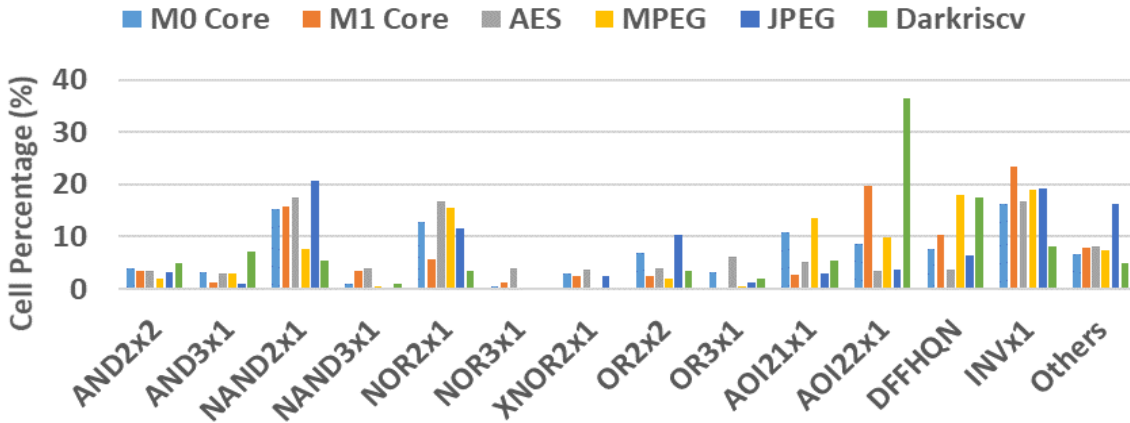


Figure 4.7: (a) #Fanouts distribution (b) Cell statistics of M0 Core, M1 Core, AES, MPEG, JPEF, and Darkriscv.

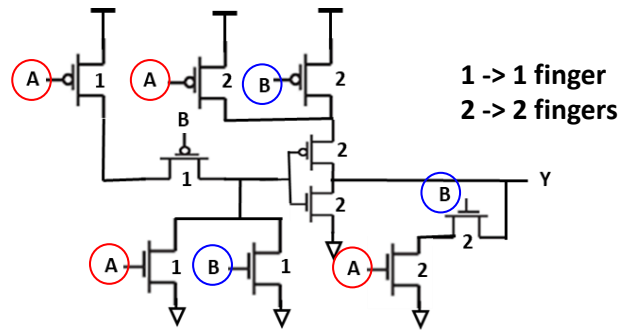
tracks for Conv. cell structure cannot be implemented due to the limitation of P-N separation [43].

3. **Design rules:** We use grid-based DR parameters to generate SDC layouts for layers up to M2, and they are applied to block-level using the corresponding metal pitch values [7]. Here, the baseline DR parameters are EOL=1 and VR=1.

Table 4.3 shows the average cell area, average RPA [29], average $M2Track$, and average $M2ML$, which are extracted for predicting $\Delta A_{i,j}$ as described in Section 4.2.4, of each SDC library set. Note that the

XOR2x1 Schematic Netlist

- (A) : Pin A has to connect 6 gate terminals and a I/O pin shape, which satisfy minimum pin opening constraint.
- (B) : Pin B has to connect 5 gate terminals and a I/O pin shape, which satisfy minimum pin opening constraint.



XOR2x1 3.5T CFET Layout (EOL=1 VR=0)

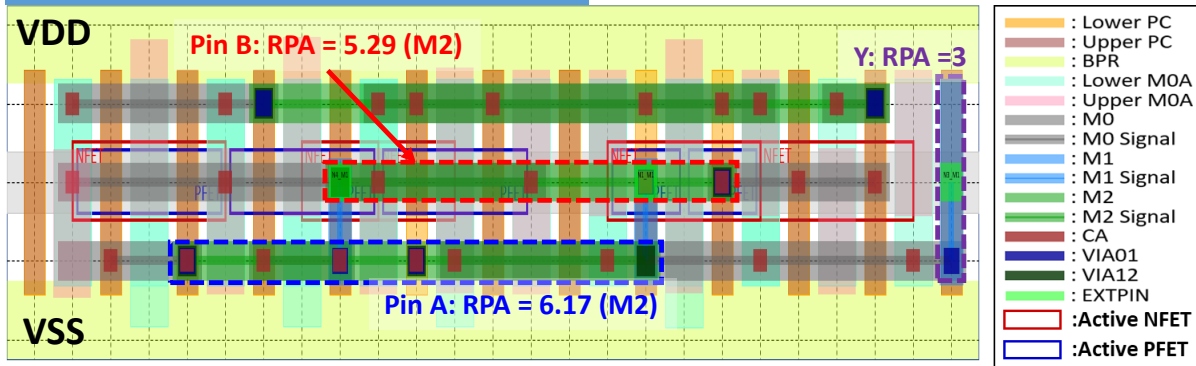


Figure 4.8: An example of RPA counts of XOR2x1 in 3.5T CFET EOL=1 VR=0 standard cell library. Pin A and pin B are promoted to M2 for connecting internal FET terminals and satisfy the minimum pin opening constraints.

M0/M2 pitches are 24nm, and contacted poly pitch (CPP) is 42nm for all the SDC library sets in Table 4.3. Figure 4.6 shows an example of generated DFFHQN SDC layouts with variations on these three dimensions (i) cell structures, (ii) design rule, and (iii) cell height. Notice that the *AvgRPA* metric of a cell library might be larger than the cell height, because the limited horizontal M0 routing resource and the connection of standard cell external pins need to be promoted to M2 for connecting FET terminals and satisfy the minimum pin opening constraint [7] for medium or large cell (i.e., XOR2x1, FAx1, etc.). Figure 4.8 shows the RPA value of each pin of XOR2x1 in 3.5T CFET EOL=1 VR=0 standard cell set. Pin A and pin B are promoted to M2 for connecting internal FET terminals and satisfy the minimum pin opening constraints.

Considering the coverage of CFET and Conv. FET cell structures, 4.5T, 3.5T, and 2.5T cell

heights, and various DRs, we select 15 SDC library sets as listed in the Train column of Table 4.3 to build our prediction model for Exp. 4.3.2, Exp. 4.3.3, and Exp. 4.3.4. Then, to test the accuracy of the proposed prediction model on new SDC library sets, we use the remaining 4 SDC library sets in Exp. 4.3.3.

Table 4.3: SDC feature values of 19 SDC library sets. CH=Cell Height. CS=Cell Structure. Conv.=Conv. FET. The baseline DR parameters are EOL=1 and VR=1. The DR parameters are grid-based parameter and is applied to block-level using the corresponding M1 and M2 metal pitch values [7]. The 2.5T CFET with EOL=0, VR=0, and use PC and M0A layers for routing (PC/M0A-R) is generated using [57]. Train=Used for training the proposed prediction model in Exp. 4.3.2 and Exp. 4.3.3.

CH	CS	DR parameters	Avg Cell Area (μm^2)	Avg RPA (access point)	Avg M2Track (track)	Avg M2ML (segment)	Train
4.5T	Conv.	Baseline	0.04415	3.290	0.433	4.900	V
		EOL=2 VR=1	0.04551	2.830	0.600	9.533	-
		EOL=0 VR=1	0.04309	2.805	0.267	2.867	V
		EOL=1 VR=1.5	0.05232	4.119	1.067	13.900	-
		EOL=1 VR=0	0.04355	2.831	0.500	5.667	V
	CFET	Baseline	0.04249	3.204	0.200	2.200	V
		EOL=2 VR=1	0.04324	3.100	0.500	5.867	V
		EOL=0 VR=1	0.04234	3.266	0.133	1.800	V
		EOL=1 VR=1.5	0.04581	3.813	0.933	12.533	V
		EOL=1 VR=0	0.04234	3.198	0.267	2.067	V
		EOL=0 VR=0	0.04229	3.058	0.167	1.333	V
3.5T	Conv.	Baseline	0.04151	2.839	1.233	19.233	V
	CFET	Baseline	0.03657	2.784	1.033	14.400	V
		EOL=2 VR=1	0.04057	3.692	1.367	23.500	-
		EOL=0 VR=1	0.03422	3.734	1.033	12.300	V
		EOL=1 VR=0	0.03410	3.516	0.933	11.333	V
		EOL=0 VR=0	0.03375	3.496	0.833	9.267	V
2.5T	CFET	EOL=0 VR=0	0.02915	3.010	2.000	19.167	-
		EOL=0 VR=0 PC/M0A-R	0.02764	2.874	1.700	18.100	V

BEOL parameters. We adjust DRs in the block-level based on the DR parameters used in the SDC library set generation [7] for M1, VIA12, and M2 layers. Then, the metals' pitch and width of layers

above M2 are set based on LEF/DEF guide [34]. For via layers above M2, the via spacing is set to allow diagonal via, and same net via spacing is set to allow adjacent via.

For the BEOL settings, we generate various M4-M7 metal pitches by varying the baseline metal pitches from $0.5\times$ to $1.5\times$. If the metal pitch is smaller/larger than the smallest pitch/largest pitch after scaling, its metal pitch is set to the smallest pitch/largest pitch. Here, the smallest vertical/horizontal metal pitch is M1/M2 metal pitch; the largest horizontal/vertical metal pitch is M8/M9 metal pitch. For the BEOL routing layers, we use M2-M5, M2-M6, and M2-M7 options for block-level routing.

In total, there are 45 various design rules and BEOL pitches technologies. For each BEOL technology, there are 3 BEOL routing options. As a result, there are 135 BEOL settings in the experiment.

Power delivery network (PDN) configurations. We study front side PDN and backside PDN in the following experiments. For front side PDN structure, The power delivery network is constructed with top power mesh on M8 and M9, and they are designed as spaces are allowed. Then, the power is delivered through M3 power straps to standard cells. Here, we vary the M3 power strap period with 24 contacted poly pitches (CPPs), 32 CPPs, 48 CPPs, and 64 CPPs based on the power delivery network studies in [35,36] for early DTCCO exploration. For backside PDN architecture, there is no power delivery network in the front side at block-level.

Minimum valid block-level area extraction. Multiple block-level P&R runs are launched for minimum valid block-level area extraction as shown in Figure 4.1 (a). In each block-level P&R run, the floorplan (i.e., including PDN generation), placement (i.e., including placement optimization), clock tree synthesis (CTS), and routing (i.e., including global routing and detail routing) stages are performed. Table 4.4 shows the breakdown of the runtime in each stage of an automated M0 core block-level P&R implementation using 2.5T CFET EOL=0 VR=0 library, and M2-M7 routing layers. The routing stage takes 94% of the total runtime because fixing DRC violations in detail routing stage is time-consuming and usually needs

many iterations (i.e., 69 iterations in this example). As a result, it takes more than 8 hours to extract minimum valid block-level area of a technology combination.

Table 4.4: The breakdown of runtime in each design stage of an automated M0 core block-level P&R implementation using 2.5T CFET EOL=0 VR=0 library, and M2-M7 routing layers. The core area, and utilization are $577.58 \text{ } \mu\text{m}^2$, and 0.73, respectively. The WNS, and TNS are -0.068ns, and -5.731ns, respectively. The final #DRVs is 1977.

Runtime	Design Stages					
	Floorplan	Placement	CTS	Routing	others	Total
wall time (s)	3	294	92	8412	118	8919

We generate the data using the synthesized block-level circuit, SDC library sets, BEOL parameters, and power delivery network configurations for our experiments. The total runtime to extract input features and to train the proposed model is around 15 hours. However, it takes us 2 months to generate all the block-level P&R data from 19 SDC library sets, 5 PDN configurations, 135 DRs and BEOL settings, and 6 block-level circuits for the experiments²². The experiments are organized as follows:

- **Exp. 4.3.2:** We explore various machine learning algorithms and demonstrate our prediction model accuracy on training, validation, and testing data.
- **Exp. 4.3.3:** We show the accuracy of our prediction model on prediction of new SDC library sets and BEOL parameters.
- **Exp. 4.3.4:** We show the accuracy of the proposed model on prediction of various power delivery network configurations.
- **Exp. 4.3.5:** We study the robustness of our prediction model on new block-level circuit prediction.

For Exp. 4.3.3, Exp. 4.3.4, and Exp. 4.3.5, we introduce gradient accuracy (Gradient ACC) metric to measure the accuracy of the direction of $\Delta A_{i,j}$. If the signs of actual $\Delta A_{i,j}$ and predicted $\Delta A_{i,j}$ are the

²²We use 8 CPU cores for each block-level P&R job, and run multiple block-level P&R jobs simultaneously.

same, we consider that the prediction is accurate for Gradient ACC metric.

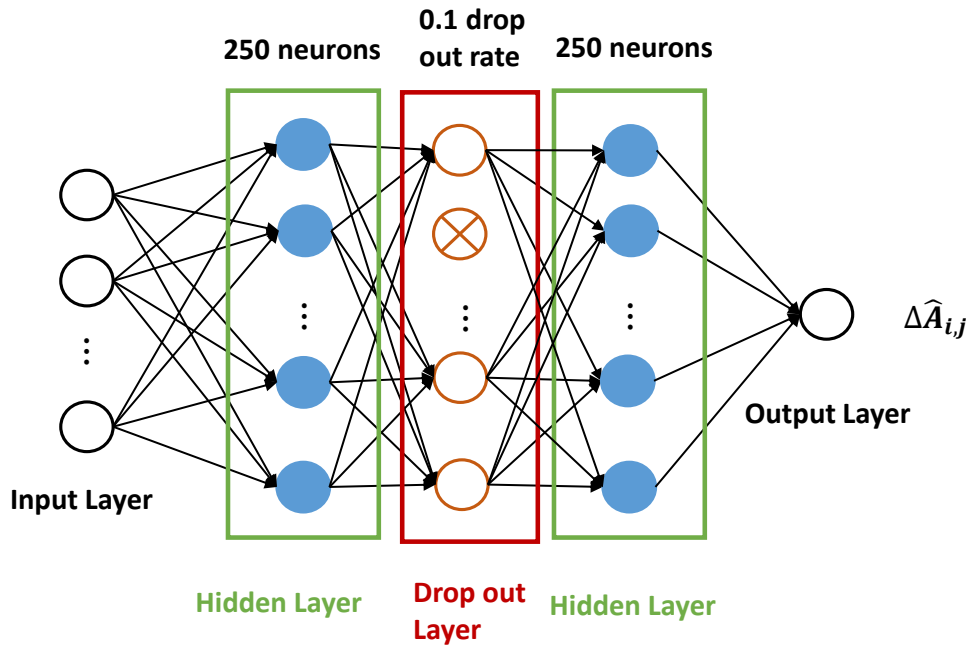
4.3.2 Prediction Model Accuracy

We study various machine learning algorithms and demonstrate prediction model accuracy with training, validation, and testing data sets. We firstly split the generated data of 15 SDC library sets (i.e., Table 4.3), 6 block-level synthesized designs, 102 DRs and BEOL settings, and 3 power delivery network settings (i.e., 24 CPPs, 48 CPPs, and 64 CPPs power strap periods), using 80% as training data and 20% as testing data based on the empirical study in [58]. Then, we split the 80% training data after bootstrap sampling such that 80% is used for model training and 20% is used for model validation. The validation data set is used to avoid overfitting with the early stopping technique in the training phase of each GBRT model. In the following experiments, XGBoost_DTICO is a GBRT tree model used in [12].

Hyperparameter tuning: We explore multi-layer perceptron (MLP) neural network, radial basis function (RBF) neural network, random forest (i.e., implemented with sklearn library), XGBoost_DTICO [12], and the proposed machine learning algorithm, which integrates bootstrap aggregation and gradient boosting regression tree techniques. We tune the hyperparameters of each machine learning modeling algorithms for our DTICO and STCO sensitivity prediction. For optimizing neural network structure for our regression problem, we adopt Hyperband [11] to set the number of layers, number of neurons of each layer, dropout rate, batch size, and learning rate of MLP and RBF neural networks. Figure 4.9 shows the selected MLP and RBF neural network structures with Hyperband [11] algorithm. For the XGBoost_DTICO [12], Random Forest, and the proposed machine learning model, we use 10-fold cross-validation [55] to set the hyperparameters of our prediction model. Table 4.5 shows the range of each hyperparameter in the explored machine learning algorithms. In the proposed model, the max_depth, sub_sample, min_child_weight, and learning rate of each GBRT model are 9, 1.0, 7, and 0.05, respec-

tively. We select 100 #GBRTs for the proposed model after hyperparameter tuning.

(a) Selected MLP structure from HyperBand



(b) Selected RBF neural network structure from HyperBand

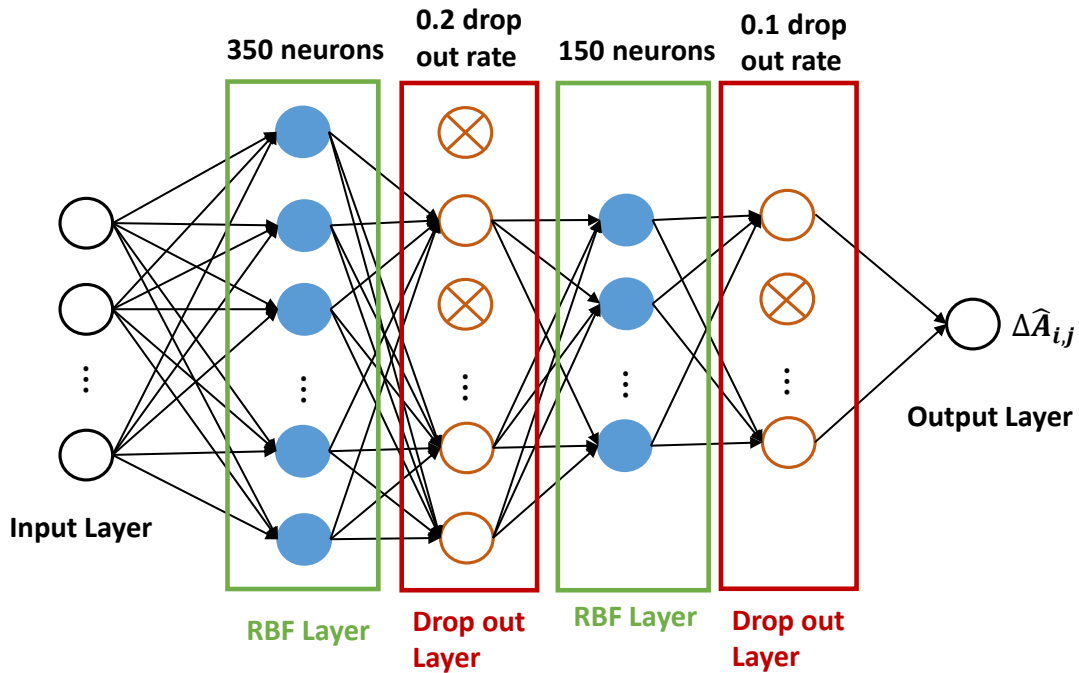


Figure 4.9: (a) MLP, and (b) RBF neural network structures after Hyperband [11] search on #layers (i.e., 2 - 10), #neurons per layer (i.e., 25 - 500), and dropout rate (i.e., 0.0 - 0.5).

Table 4.5: Hyperparameter exploration of machine learning algorithms table.

Machine Learning Alg.	Hyperparameter	Value Range
MLP/ RBF neural network	#layers	2 - 10
	#neurons per layer	25 - 500
	drop out rate	0.0 - 0.5
	learning rate	{1e-2, 5e-3, 1e-3, 5e-4, 1e-4}
	batch size	{128, 256, 512}
Random Forest (sklearn) [56]	#estimators	50 - 500
	max_depth	{10 - 100, None}
	min_samples_leaf	1 - 4
	min_samples_split	2 - 10
XGBoost_DTCO [12]	max_depth	6 - 15
	min_child_weight	5 - 11
	sub_sample	0.7 - 1.0
	learning rate	{5e-2, 1e-2, 5e-3, 1e-3, 5e-4, 1e-4}
Proposed Method	#GBRTs	50 -500
	GBRT parameters	Same as XGBoost_DTCO [12]

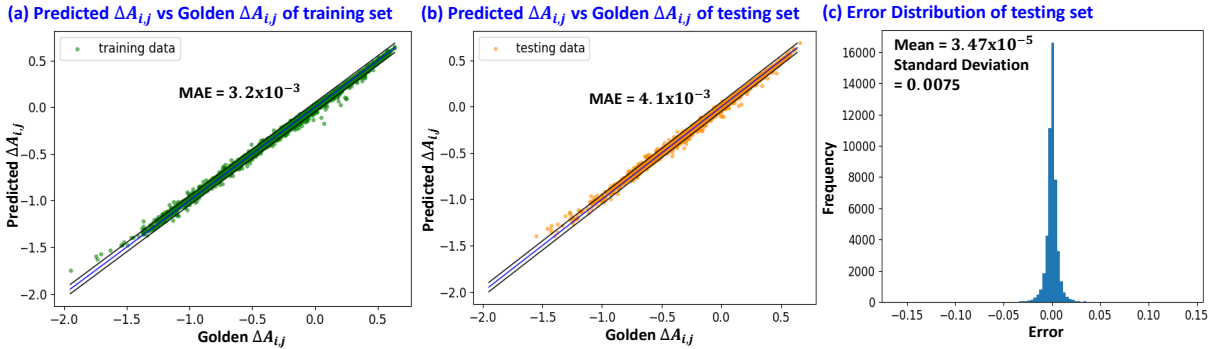


Figure 4.10: Predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of (a) training set and (b) testing set, and (c) error distribution of testing set of the proposed model. The mean of MAE is 3.47×10^{-5} , with standard deviation of 0.0075 for testing set. Hence, 99.7% of predicted $\Delta A_{i,j}$ are within the 3-sigma range of ± 0.023 .

Prediction accuracy: Table 4.6 shows the prediction accuracy results of MLP, RBF neural network, XGBoost_DTCO [12], Random Forest, and the proposed Method. The MAE of the proposed model is 4.1×10^{-3} on testing set. Compared to MLP and RBF neural networks, the proposed model achieves

Table 4.6: Prediction accuracy table. Impr. MAE= $(MAE_{MLAlg} - MAE_{proposed})/MAE_{MLAlg} \times 100$. Here, MAE_{MLAlg} represents the MAE error of MLP/RBF neural network/XGBoost_DTTCO [12]/Random Forest [56]

Machine Learning Alg.	MAE		Impr. MAE (%)	
	Training set	Testing set	Training set	Testing set
MLP	0.0570	0.0578	94.3	92.9
RBF Neural Network	0.0155	0.0159	79.4	74.2
Random Forest (sklearn) [56]	0.0066	0.0120	51.5	65.8
XGBoost_DTTCO [12]	0.0034	0.0049	5.9	16.3
Proposed	0.0032	0.0041	-	-

92.9% and 74.2% less MAE on the testing set, respectively. Moreover, the proposed model provides 65.8% and 16.3% less MAE on the testing set than random forest, and XGBoost_DTTCO [12], respectively.

Figure 4.10(a) and 4.10(b) show predicted $\Delta A_{i,j}$ values versus golden $\Delta A_{i,j}$ values for training and testing sets of the proposed model. The solid blue line in the middle indicates a perfect correlation between golden $\Delta A_{i,j}$ and predicted $\Delta A_{i,j}$. The upper and lower black solid lines are 5% away from the solid blue line, respectively. We can observe that most of the error of predicted $\Delta A_{i,j}$ are within 5% in the training and testing sets. The mean absolute errors (MAE) are 3.2×10^{-3} for the training set and 4.1×10^{-3} for the testing set. Figure 4.10(c) shows that the error distribution of testing set. The mean is 3.47×10^{-5} , with standard deviation of 0.0075 (hence, 99.7% of predicted $\Delta A_{i,j}$ values are within the 3-sigma range of +/-0.023). Furthermore, compared to XGBoost_DTTCO [12], the proposed model reduces the standard deviation of error distribution by 0.0011 (i.e., 12.8%) for the testing set. This shows the proposed model is more robust than XGBoost_DTTCO [12] on the model accuracy.

Key features study: To further study the key features, we combined the gain of the same features of the first and second technologies of each technology pair in XGBoost_DTTCO [12] model and the proposed

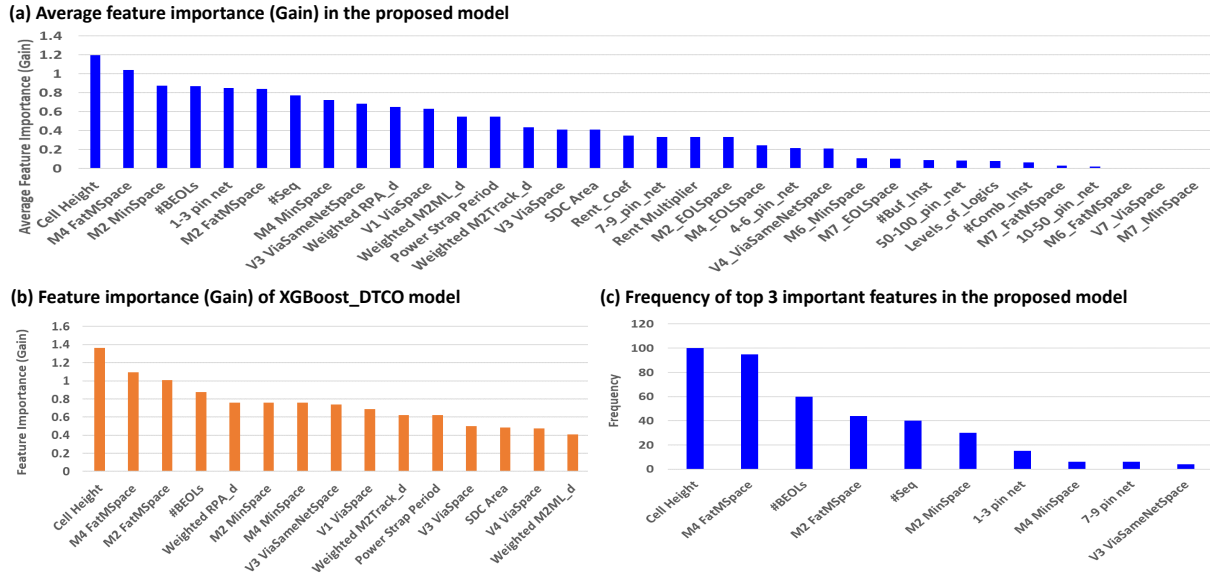


Figure 4.11: Feature importance (Gain) of the proposed model and XGBoost_DTCO [12] for key feature study. (a) Average combined feature importance (gain) of GBRTs in the proposed model. (b) Top 15 Combined feature importance (gain) in the trained XGBoost_DTCO model [12]. (c) #Counts of important features, which are extracted from top 3 gain of 100 GBRT models, in the proposed model.

model. Figure 4.11 (a), and (b) show the average combined important features of 100 GBRTs in the proposed model, and top 15 combined important features in XGBoost_DTCO [12] after feature selection (Section 4.2.5), respectively.

The most important feature in the proposed model and XGBoost_DTCO [12] is cell height. Cell height is highly related to the block-level area because it determines the size of each cell row in the block-level. For the pin accessibility and routing congestion metrics in SDCs, the proposed weighted RPA_d , weighted $M2Track_d$, and weighted $M2ML_d$ are also very important for $\Delta A_{i,j}$ prediction in both XGBoost_DTCO [12] model and the proposed model in the block-level. For the synthesized design feature, the 1-3 fanouts and the number of sequential cells (#Seq) features are recognized as top 15 average important features in the proposed model.

For design rule feature, we can observe that the V1 spacing, M2 minimum spacing, V3 spacing, and V3 same net spacing all have large gains in both XGBoost_DTCO [12] model, and the proposed

model because these layers are mainly used for accessing the SDC pins on M1/M2. For the design rule features of layers above M4, their gains are smaller since these layers are mainly used to connect above and below metal layers instead of accessing SDC pins. Note that the M2 and M4 fat metal spacing rules (FatMSpace), which are usually related to the wider metal used for power straps, are recognized as important features in the proposed model and XGBoost_DTCO [12] model. In addition to the design rules related to power strap, the power strap period feature is also in the top 15 important features in the both model, because its impact on the block-level is nontrivial as shown in Figure 4.13. Here, although via spacing and same net via spacing has high correlation, they could be remaining in the input features after feature selection stage since we remove the feature only if the validation error, E_{val} , is larger and the VIF of the feature is larger than a VIF threshold in Algorithm 4. The "V3 via same net space" and "V3 via space" are both used as input features in the Figure 4.11 (a).

With more simulated data sets generated by bootstrap aggregation, we observe various important features of each GBRT model in the proposed model as shown in Figure 4.11 (c). Figure 4.11 (c) shows the #Counts of the important features, which are extracted from the top 3 gain of each GBRT model, in the proposed model. The top 9 important feature with larger average gain in Figure 4.11 (b) are also in the top 3 important features of 100 GBRTs frequently in the proposed model. Here, the 7-9 fanouts net feature is also frequently appears in the top 3 important features of 100 GBRTs in the proposed model though its average gain across 100 GBRT models is not in the Figure 4.11 (b).

4.3.3 Prediction of New Technologies

We apply the trained model from Exp. 4.3.2 to predict the $\Delta A_{i,j}$ of new SDC library sets, and new BEOL parameters. Here, we implement a benchmark utilization prediction model (Util. model) with XGBoost algorithm, which takes the features of a technology (i.e., Table 4.1) and predicts the utilization

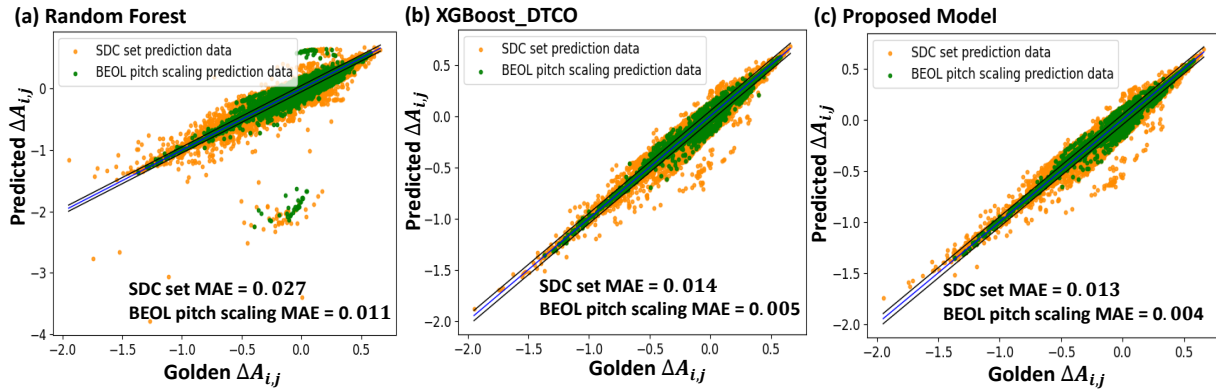


Figure 4.12: Predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of new SDC library set technology prediction (i.e., orange points) and new BEOL pitch scaling technology prediction (i.e., green points) with (a) Random Forest (i.e., implemented with sklearn), (b) XGBoost_DTDCO [12], and (c) the proposed model.

Table 4.7: The $\Delta A_{i,j}$ prediction results of new technologies using utilization model (Util.), random forest, XGBoost_DTDCO [12], and the proposed model. MAE=Mean Absolute Error. Gradient ACC=Gradient Accuracy of $\Delta A_{i,j}$. Error Dist.=Error Distribution. Std. Dev.=Standard Deviation.

Prediction Type	Model	MAE	Gradient ACC (%)	Error Dist.	
				Mean	Std. Dev.
New SDC lib. set	Util.	0.150	77.3%	-0.012	0.233
	Random Forest [56]	0.027	94.8%	0.002	0.069
	XGBoost_DTDCO [12]	0.014	97.2%	0.001	0.031
	Proposed	0.013	97.3%	0.001	0.031
New BEOL pitch scaling tech.	Util.	0.147	88.8%	-0.025	0.210
	Random Forest [56]	0.011	96.9%	0.001	0.049
	XGBoost_DTDCO [12]	0.005	96.9%	0.000	0.013
	Proposed	0.004	97.1%	0.000	0.012

after block-level P&R. Then, we calculate the block-level area after P&R from the output of Util. model and obtain the $\Delta A_{i,j}$ of every technology pairs for comparison. In this experiment, we compare the accuracy of the proposed model, random forest, XGBoost_DTDCO [12], and Util. model on DTDCO and STCO sensitivity prediction of new SDC library sets and new BEOL parameters²³.

For new SDC library sets, we study the accuracy of the proposed prediction model to predict

²³We mainly study the tree based machine learning models (i.e., the proposed model, random forest, and XGBoost_DTDCO [12]) since the MAEs of tree based machine learning models on testing set are better than neural network models in Exp. 4.3.2.

$\Delta A_{i,j}$ of 20% of 19 SDC library sets in Table 4.3. The 4 new SDC library sets are carefully selected to include different cell heights (i.e., 4.5T, 3.5T, and 2.5T), different cell structures (i.e., Conv. and CFET), and design rules including strict and loose DR parameters (i.e., EOL=2 VR=1, and EOL=0 VR=0) to demonstrate the prediction of new SDC library sets. The BEOL routing layer options for these 4 testing SDC library sets are M2-M5, M2-M6, and M2-M7. Table 4.7 shows the prediction results of $\Delta A_{i,j}$ using Util. model, and the proposed DTCO and STCO sensitivity prediction approach with random forest, XGBoost_DTCO [12], and the proposed model. The proposed model provides 0.013 MAE and 97.3% gradient accuracy on new SDC library set prediction. Compared to Util. model, our proposed model achieves 91.3% less MAE error and 20.0% better gradient accuracy. Compared to random forest, and XGBoost_DTCO [12], the proposed model still maintains 51.9%, and 7.1% less MAE error, respectively. Figure 4.12 shows the predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ for SDC library set prediction (i.e., orange point) with (a) random forest, (b) XGBoost_DTCO [12], and (c) the proposed model. There are clearly more data points of random forest prediction outside of the black solid line, which represents 5% away from the perfect correlation line in the middle. This matches the larger standard deviation and MAE in Table 4.7.

For new BEOL pitch scaling settings, we study the accuracy of the proposed model on prediction of 11 BEOL pitch scaling technologies. Combing these 11 BEOL pitch scaling technologies with 15 SDC library sets, 3 #BEOL layer options (i.e., M2-M5, M2-M6, and M2-M7), and 3 PDN settings, there are 1485 technology combinations for prediction in this experiment. In addition, the BEOL pitch scaling also affects DRs, such as minimum spacing, end-of-line spacing, via spacing, and same net via spacing. In Table 4.7, the MAE and gradient accuracy of the proposed model are 0.004 and 97.1%, respectively. Compared to Util. model, the proposed model achieves 97.2% less MAE error and 8.3% better gradient accuracy. Moreover, the MAEs of the proposed model are 63.6% and 20.0% smaller than random forest

and XGBoost_DTTCO [12], respectively. Figure 4.12 shows the predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ for BEOL pitch scaling prediction (i.e., green points) with (a) random forest, (b) XGBoost_DTTCO [12], and (c) the proposed model. Here, we can observe that there are obviously more data points of random forest prediction outside of the black solid line.

To summarize, the proposed DTTCO and STTCO sensitivity prediction modeling approach achieves better accuracy than the Util. model because it directly minimizes the MAE of $\Delta A_{i,j}$ and $\Delta \hat{A}_{i,j}$ during the training phase. On the other hand, there are utilization prediction error from the Util. model and inherent differences between synthesized block-level circuit area and block-level area after P&R in Util. model.

$$E_{sen} = \frac{(\hat{A}_i - \hat{A}_j)}{\hat{A}_i} - \frac{(A_i - A_j)}{A_i} = \frac{E_i A_j - E_j A_i}{A_i(A_i + E_i)} \quad (4.7)$$

Equation (4.7) shows the DTTCO and STTCO sensitivity error when we use the predicted minimum block-level area from Util. model. Here, E_{sen} and E_i are the error of DTTCO/STTCO sensitivity and predicted minimum block-level area, respectively. $\hat{A}_i = A_i + E_i$. When E_i is very small and $E_j > A_i$, the predicted block-level error (i.e., E_j) leads to large E_{sen} on DTTCO and STTCO sensitivity prediction. For example, from the one of the data points in new SDC library set technologies prediction, A_i , A_j , \hat{A}_i , and \hat{A}_j are 276.652, 1138.511, 280.911, and 814.554, respectively. The E_{sen} is 1.19, which is larger than 99.7% (i.e., 3-sigma range 0.093+/-0.001) of the error of the proposed model. As a result, we observe that Util. model has larger standard deviation of error than the proposed model in the Table 4.7. Moreover, compared to random forest and XGBoost_DTTCO [12], the proposed model provides smaller MAE and better gradient accuracy for DTTCO and STTCO sensitivity prediction with bootstrap aggregation and gradient boosting regression tree techniques.

4.3.4 Prediction of New Power Delivery Network Setting

We study the model accuracy on predicting $\Delta A_{i,j}$ of new power delivery network grid scales and architectures (i.e., backside PDN). Here, we select front side PDN with 24 CPPs, 48 CPPs, and 64 CPPs power strap period to train our model, and use the trained model to predict $\Delta A_{i,j}$ of (a) new PDN setting with 32 CPPs power strap period, and (b) backside power delivery network architecture.

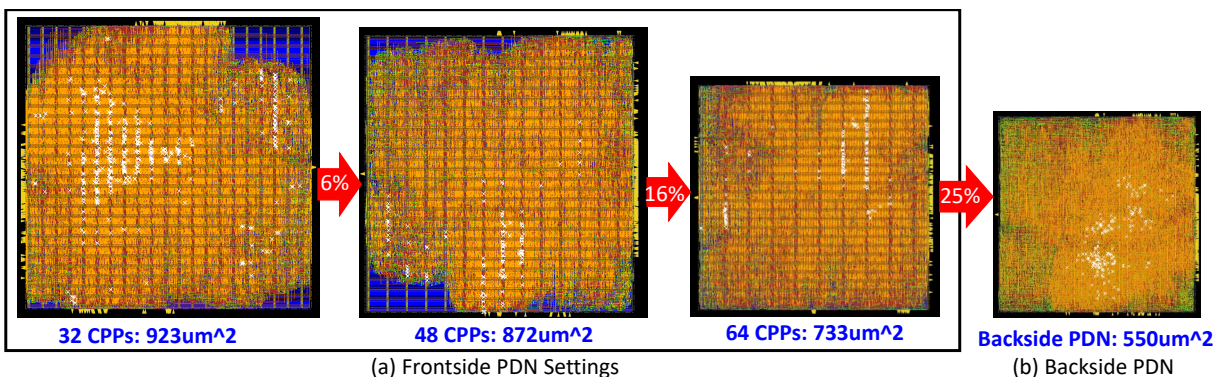


Figure 4.13: Minimum block-level area of M0 Core with various front side PDN grid scales (i.e., 32 CPPs, 48 CPPs, and 64 CPPs), and backside PDN architecture using M2 to M6 for signal routing. Compared to 32 CPPs front side PDN setting, the core area of backside PDN is 40% smaller. The standard cell library is 3.5T CFET with Baseline DR parameters in Table 4.3.

Figure 4.13 (a) shows the snapshots of M0 Core design with various M3 power strap period. For backside PDN architecture, there is no power strap in the front side at block-level as shown in Figure 4.13 (b). We can observe that the power strap period and power delivery network architecture (i.e., backside power delivery) can potentially impact the block-level area from 6% to 25% from Figure 4.13. Since the Util. model performs poorly in Exp. 4.3.3, we mainly study the accuracy of random forest, XGBoost_DTICO [12], and the proposed model in this experiment.

Prediction of new PDN setting: Table 4.8 shows the prediction results of new front side PDN setting prediction (i.e., 32 CPPs). For new front side PDN setting prediction, the proposed model achieves 0.027 MAE and 94.4% gradient accuracy, which are 27.0% less MAE and 1.9% better gradient accuracy than

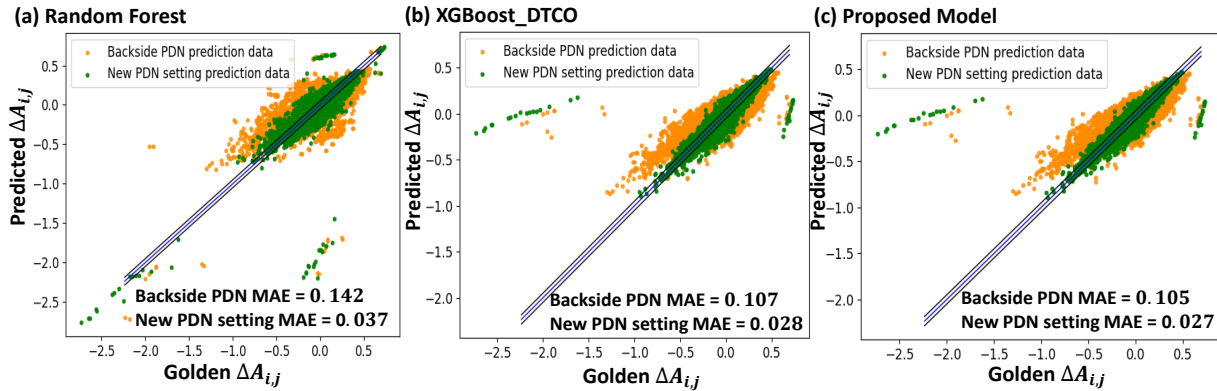


Figure 4.14: Predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of new PDN setting (32 CPPs) prediction (green points) and backside PDN prediction (orange points) with (a) Random Forest (implemented with sklearn), (b) XGBoost_DTDCO [12], and (c) the proposed model.

random forest. Compared to XGBoost_DTDCO [12], the proposed model achieves 3.6% less MAE and 0.3% better gradient accuracy. Figure 4.14 shows the predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of new PDN setting prediction (i.e., green points). Although there are few green points located far away from the perfect center line in the proposed model, the gradient accuracy is 94.4%. Therefore, the accuracy of the proposed model can be calibrated along the gradient of $\Delta A_{i,j}$ from one technology to another technology as shown in Figure 4.3 (b).

Prediction of backside PDN: Here, to further study the robustness of the proposed model on new PDN architecture, we firstly use the trained model, which is trained using front side PDN with various power

Table 4.8: The $\Delta A_{i,j}$ prediction results of new front side PDN setting (i.e., 48 CPPs) and backside PDN architecture using random forest, XGBoost_DTDCO [12], and the proposed model. MAE=Mean Absolute Error. Gradient ACC=Gradient Accuracy of $\Delta A_{i,j}$. Error Dist.=Error Distribution. Std. Dev.=Standard Deviation.

Prediction Type	Model	MAE	Gradient ACC (%)	Error Dist.	
				Mean	Std. Dev.
New front side PDN setting (32 CPPs)	Random Forest [56]	0.037	92.5%	0.004	0.102
	XGBoost_DTDCO [12]	0.028	94.1%	0.000	0.107
	Proposed	0.027	94.4%	0.000	0.106
Backside PDN Architecture	Random Forest [56]	0.142	77.8%	-0.014	0.198
	XGBoost_DTDCO [12]	0.107	84.8%	-0.017	0.147
	Proposed	0.105	86.9%	-0.015	0.155

strap period, to predict the $\Delta A_{i,j}$ of backside PDN architecture. Then, we further study the improvement of prediction accuracy of random forest, XGBoost_DTICO [12], and the proposed model using various ratio (i.e., 10% to 80%) of backside PDN data points to update the models.

Firstly, for predicting backside PDN without any backside PDN data for training, the MAE and gradient accuracy of the proposed model are 0.105 and 86.9%, respectively. The proposed model achieves 35.2% and 1.8% less MAE than random forest and XGBoost_DTICO [12], respectively. In addition, compared to random forest and XGBoost_DTICO [12], the proposed model provides 9.1% and 2.1% better gradient accuracy, respectively. Figure 4.14 shows the predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of backside PDN prediction (i.e., orange points). The block-level area difference of backside PDN technology, which brings the systematic physical layout change at block-level, can not be fully captured (i.e., MAE is larger than 0.1) with only front side PDN training data using random forest, XGBoost_DTICO [12], and the proposed model. As a result, we further study the accuracy improvement of prediction models using various ratios (i.e., 10% to 80%) of backside PDN data points to update the models, which is the outer loop in Figure 4.3 (b).

Figure 4.15 (a) shows the MAE of backside PDN prediction of XGBoost_DTICO [12] and the proposed model with various ratio (i.e., 10% to 80%) of backside PDN data points for updating the models. The proposed model provides larger accuracy improvement than XGBoost_DTICO [12] when giving a ratio of backside PDN data for model update. Moreover, the proposed model achieves up to 60.8% MAE reduction when updating the model with 20% backside PDN data. Figure 4.15 (b) shows that predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of 0%, 10%, and 20% backside PDN data for model update. From Figure 4.15 (b), the proposed model can efficiently capture the block-level area difference ($\Delta A_{i,j}$) of backside PDN with 10% to 20% backside PDN data for model update. This shows that the proposed model can be updated efficiently and robustly with small amount of data of new technologies, which lead

to the systematic physical layout change at block-level.

To summarize, the bootstrap aggregation technique creates meaningful simulated data sets from the given training data set which can reduce model variance while avoiding overfitting. For each GBRT model in the proposed model in Figure 4.5, the gradient boosting tree technique improves the accuracy by sequentially building an ensemble of K regressors to minimize the prediction error. Therefore, the proposed model can provide better accuracy and robustness than random forest and XGBoost_DTICO [12] model on predicting new PDN setting and backside PDN architecture.

4.3.5 Robustness of New Circuit Prediction

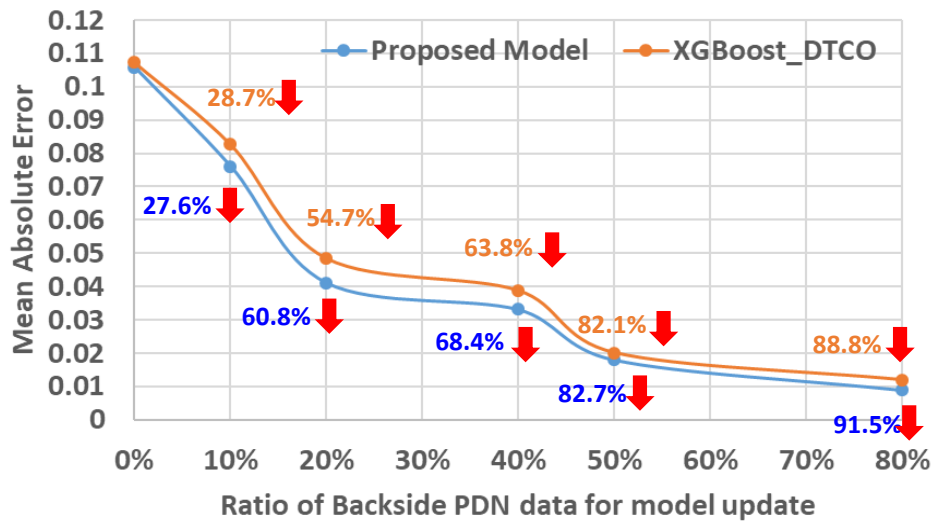
We study the robustness of the proposed modeling approach for predictions on new block-level circuits in this experiment. Here, we iteratively select one synthesized block-level circuit out of the 6 synthesized block-level circuits (i.e., Table 4.2) for testing the robustness of model prediction on new designs. Then, train the model with the rest of the 5 synthesized block-level circuits with all SDC library sets, DRs, and BEOL settings, and apply the trained prediction model to predict the $\Delta A_{i,j}$ of the selected synthesized block-level circuit with all the SDC library sets, DRs, and BEOL settings.

Table 4.9: The $\Delta A_{i,j}$ prediction results of selected synthesized block-level circuit. MAE=Mean Absolute Error. Gradient ACC=Gradient Accuracy of $\Delta A_{i,j}$. Error Dist.=Error Distribution. Std. Dev.=Standard Deviation. Improvement= $(Metric_{MLAlg} - Metric_{proposed})/Metric_{MLAlg}$. Metric=MAE/Gradient ACC.

Selected Design for Prediction	Random Forest [56]				XGBoost_DTICO [12]				Proposed Model				Improvement			
	MAE	Gradient ACC	Error Dist.		MAE	Gradient ACC	Error Dist.		MAE	Gradient ACC	Error Dist.		Random Forest [56]		XGBoost_DTICO [12]	
			Mean	Std. Dev.			Mean	Std. Dev.			Mean	Std. Dev.	MAE	Gradient ACC	MAE	Gradient ACC
MO	0.0929	80.68%	-0.0114	0.1195	0.0742	87.19%	-0.0082	0.0846	0.0725	87.86%	-0.0067	0.0846	21.96%	8.17%	2.29%	0.77%
M1	0.0487	87.00%	0.0070	0.0680	0.0445	88.10%	0.0117	0.0622	0.0446	88.13%	0.0122	0.0623	8.42%	1.28%	-0.22%	0.03%
AES	0.0988	77.75%	0.0339	0.1223	0.0637	87.47%	0.0098	0.0840	0.0608	88.17%	0.0085	0.0802	38.46%	11.82%	4.55%	0.80%
MPEG	0.058	73.86%	0.0047	0.0770	0.0516	87.89%	-0.0025	0.0700	0.0492	87.97%	-0.0042	0.0674	15.17%	16.04%	4.65%	0.09%
JPEG	0.0706	78.26%	0.0256	0.0818	0.0562	87.65%	0.0049	0.0730	0.0524	87.64%	0.0130	0.0680	25.78%	10.70%	6.76%	-0.01%
Darkrisc	0.0831	78.01%	0.0075	0.1107	0.0549	85.36%	0.0012	0.0718	0.0536	87.67%	0.0013	0.0692	35.50%	11.02%	2.37%	2.71%
Avg	0.0754	79.26%	0.0112	0.0965	0.0575	87.28%	0.0028	0.0743	0.0555	87.91%	0.0040	0.0720	24.22%	9.84%	3.40%	0.73%

Table 4.9 shows the robustness of random forest, XGBoost_DTICO [12], and the proposed model to make predictions on designs unseen in the training set. The average MAE and average gradient accu-

(a) MAE versus Ratio of backside PDN data for model update



(b) Predicted $\Delta A_{i,j}$ versus Golden $\Delta A_{i,j}$ of different ratio of backside PDN data for model update (proposed model)

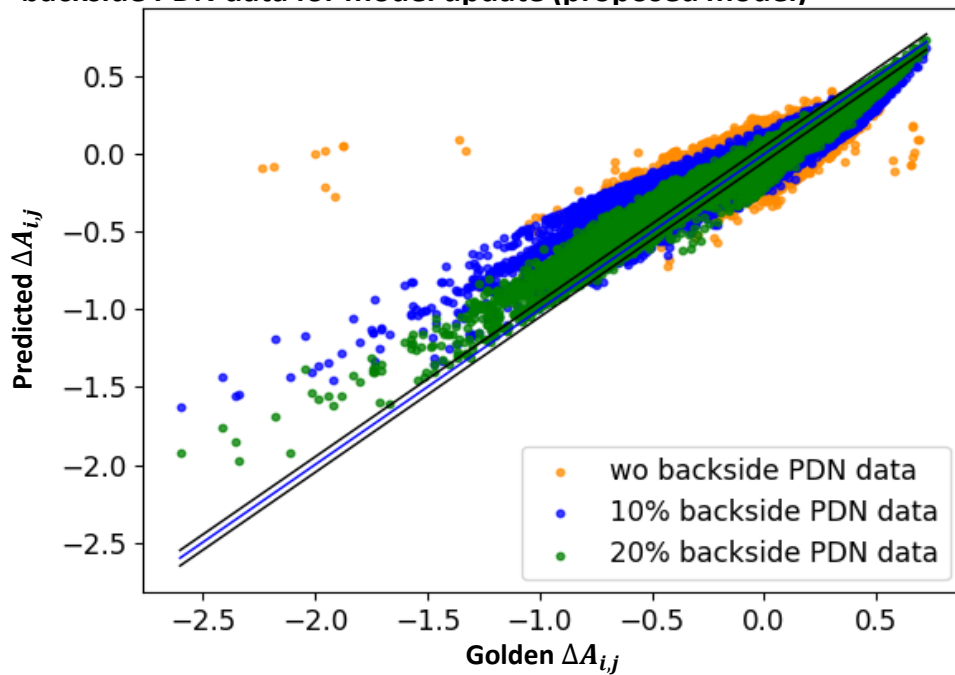


Figure 4.15: Accuracy improvement with various ratio of backside PDN data for model update. (a) MAE versus ratio of backside PDN data for model update. Orange/Blue number is the reduced MAE percentage of XGBoost_DTCO [12]/proposed model compared to 0% backside PDN data for model update. (b) Predicted $\Delta A_{i,j}$ versus golden $\Delta A_{i,j}$ of 0%, 10%, and 20% backside PDN data for model update. 10% to 20% backside PDN data for model update greatly reduce up to 60.8% MAE for the proposed model.

accuracy are 0.0555 and 87.91% for DTTCO and STTCO sensitivity prediction on new designs using the proposed model. Moreover, the proposed modeling approach achieves 24.22% and 3.40% smaller average MAEs than random forest and XGBoost_DTTCO [12], respectively. Also, the proposed model provides 9.84% and 0.73% better gradient accuracy on average than random forest and XGBoost_DTTCO [12], respectively. This shows the proposed model is able to robustly guide DTTCO optimization on designs unseen during training.

Regarding runtime performance, it takes less than one minute to predict 10k block-level area sensitivities of one technology to another technology. On the other hand, it takes more than 8 hours to extract minimum valid block-level area of a new technology combination for block-level metric comparison (i.e., $\Delta A_{i,j}$) as described in Section 4.3.1. The proposed prediction model achieves more than $100\times$ speedup on finding the optimal technology candidate in the potential technology list compared to running the block-level P&R runs for multiple potential technology candidates, extracting the minimum valid block-level area, and found the optimal technology candidate. In summary, we show that our modeling approach not only captures the block-level area difference on new SDC library sets, BEOL parameters, and various power delivery network configurations, but is also capable of robustly predicting $\Delta A_{i,j}$ of various technology options for new circuit designs.

4.4 Conclusion

We propose an overall framework along with the proposed DTTCO and STTCO sensitivity prediction model, and automatic SDC synthesis [7, 10] to significantly reduce the TAT of DTTCO and STTCO explorations. In addition, we develop a machine learning model using bootstrap aggregation and gradient boosting techniques to predict the difference of block-level area between two different technology options for reducing the runtime of block-level P&R in DTTCO and STTCO explorations.

We firstly demonstrate that the MAEs of the proposed DTCO and STCO sensitivity prediction model are 3.2×10^{-3} for training set and 4.1×10^{-3} for testing set. In addition, 99.7% of prediction errors are within ± 0.023 . Then, we validate the importance of the proposed block-level SDC metrics (i.e., weighted *RPA*, *M2Track*, and *M2ML*) through the feature importance in the proposed model. For prediction on new technologies, we showed that our machine learning model not only achieves 7.1% less MAE on predicting new SDC library sets across different designs, but also provides 20.0% less MAE on predicting new BEOL settings than XGBoost_DTCO [12]. For the studies on predicting $\Delta A_{i,j}$ of new PDN setting and backside PDN structure, we not only show that the proposed model achieves 0.027 MAE for new front side PDN configuration, but also demonstrate that the MAE of the proposed model is reduced up to 60.8% with only 10% to 20% backside PDN data for model update. Lastly, we demonstrate that the proposed modeling approach achieves 5.55×10^{-2} MAE and 87.91% gradient accuracy on average in the robustness experiment of new design prediction. For the performance, it takes less than one minute to predict 10k block-level area sensitivities of one technology to another technology, and provide more than 100 \times speedups compared to running block-level P&R for technologies and extracting minimum valid block-level area.

This chapter contains materials from "Design and System Technology Co-Optimization Sensitivity Prediction for VLSI Technology Development using Machine Learning." by Chung-Kuan Cheng, Chia-Tung Ho, Chester Holtz, and Bill Lin, which appears in 2021 ACM/IEEE International Workshop on System Level Interconnect Prediction (SLIP), 2021; "Machine Learning Prediction for Design and System Technology Co-Optimization Sensitivity Analysis." by Chung-Kuan Cheng, Chia-Tung Ho, Chester Holtz, Daeyeal Lee, and Bill Lin, which will appear in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2022. The dissertation author was the primary researcher and author of these papers.

Chapter 5

Conclusion and Future Directions

5.1 Conclusion

This thesis describes novel computer-aided design (CAD) methodologies and its automated frameworks for emerging Complementary-FET (CFET) technology in sub-7 *nm* in three topics: (1) routability-driven Complementary-FET (CFET) standard cell synthesis for block-level area optimization, (2) CFET standard cell synthesis framework for design and system technology co-optimization, and (3) Machine learning prediction for design and system technology co-optimization sensitivity analysis.

Chapter 2 introduces a SMT (Satisfiability Modulo theories)-based framework to automate CFET SDC synthesis through the simultaneous place-and-route optimization methodology with a novel Dynamic Complimentary Pin Allocation scheme. Moreover, our framework generates optimized CFET SDC layouts in terms of routability through our novel pin access and routing resource related objectives/constraints, while the scaling advantage of CFET structure is maintained compared to conventional FET structure. We demonstrate that CFET cell structure provides 10.1% and 22.2% on average reduced cell width and metal length, respectively, compared to conventional FET structure. Moreover, we validate

that our routability-driven features successfully improve routability in practical circuit designs through block-level analysis. Compared to the previous work, our routability-driven framework improves 4.2% utilization and reduces 83% routing errors on average in block-level designs.

Chapter 3 shows the proposed Multi-Row CFET SDC synthesis framework that simultaneously solves place-and-route to minimize the cell area by considering single-row and multi-row placement together for design technology and system technology co-optimization explorations. In addition, we enable explorations on Upper/Lower M0A/PC routing to leverage the shared-and-split structure across cell rows with the proposed multi-row dynamic complementary pin allocation scheme. In the system technology co-optimization experiment on 2D and 3D standard cell architecture, CFET structure achieves 10.94% and 21.27% reduction on average cell area and metal length, respectively, and 15.10% smaller block-level area compared to 2D conventional cell structure as scaling down to 3.5T architecture. Then, through extensive DTCO explorations on ground design rules and #BEOLs, 3.5T CFET SDCs achieve up to 6.50% smaller block-level areas than 4.5T CFET SDCs. Lastly, with the assistance of STCO and DTCO, 3.5T CFET SDCs achieve 21.0% on average reduced block-level areas compared to 4.5T Conv. SDCs. Lastly, in the extreme CFET cell architecture scaling experiment, multi-row 2.5T CFET without and with Upper/Lower M0A/PC routing achieve 16.44% and 20.61% on the average reduced cell areas, respectively, compared to 3.5T CFET. Moreover, multi-row 2.5T CFET SDCs achieve 13.43% and 14.40% less block-level area and total wirelength on average compared to 3.5T CFET SDCs.

Chapter 4 introduces the developed machine learning model which combines bootstrap aggregation and gradient boosting techniques to predict the sensitivity of minimum valid block-level area of various physical layout factors. We firstly demonstrate that the proposed model achieves 16.3% less mean absolute error (MAE) than the previous work for testing sets. Then, we show that the proposed model successfully captures the block-level area sensitivity of new SDC library sets, new BEOL settings, and

new PDN settings with 0.013, 0.004, and 0.027 MAE, respectively. Lastly, compared to the previous work, the proposed approach improves the robustness of predicting new circuit designs by up to 6.76%. The proposed framework provides more than $100\times$ speedup compared to conventional design and system technology co-optimization exploration flows.

5.2 Future Directions

There are mainly three major topics for the future works: (i) CFET standard cell synthesis for power-performance-area (PPA) and process-aware optimization; (ii) Design and system technology co-optimization sensitivity prediction with power, performance, and area block-level metrics across 3D standard cell/transistor architectures; (iii) Improvement of the scalability of CFET standard cell synthesis framework using reinforcement learning technique.

5.2.1 CFET Standard Cell Synthesis for Power-Performance-Area (PPA) and Process-Aware Optimization

The important directions for future researches here include (i) in-corporating timing and power information of CFET for further Power-Performance-Area (PPA) explorations in both cell-level and block-level, (ii) developing CFET SDC synthesis framework considering emerging Monolithic 3D integration [59], [60], and (iii) developing process variation aware CFET SDC synthesis framework for both FET and interconnect level. To reduce the process variation, Design for Manufacturing (DFM) and Design for reliability (DFR) are applied in SDC design [61]. For DFM, the proposed Multi-Row CFET SDC synthesis framework considers not only EOL, VR, and MAR [7] but also PRL and SHR for multi-patterning technology [22]. To deal with the DFR, adding the objectives/constraints related to reliability (i.e., layout-dependent aging effect [62], double via for EM, etc.) are essential for process variation aware

CFET SDC synthesis.

5.2.2 Design and System Technology Co-Optimization Sensitivity Prediction for Block-Level Power-Performance-Area (PPA) Optimization

The future research directions of DTCO and STCO sensitivity prediction include (i) conducting an extensive study on multiple 3D SDC architectures, such as many-tier VFET SDC [63], (ii) incorporating more circuit designs in the study (i.e., deep learning accelerators [64]), (iii) extending the DTCO and STCO area sensitivity prediction model for power and performance metrics. Here, the timing and power information (i.e., SPICE model card and transistor-level parasitic extraction) of emerging cell architecture (i.e., CFET and VFET) are essential to conduct comprehensive Power-Performance-Area (PPA) studies in both cell-level and block-level in DTCO and STCO explorations. This enables the prediction of DTCO and STCO sensitivity on block-level power and performance metrics with the consideration of power information of cells (i.e., dynamic power, internal power, and leakage power), timing information of cells (i.e., cell delay, slew), coupling capacitance, parasitic resistance, parasitic capacitance, etc.

5.2.3 Routability-Aware CFET Standard Cell Synthesis using Reinforcement Learning

The CFET standard cell synthesis framework places the transistors on a 2D transistor placement canvas on Upper/Lower M0A/PC placement grids as shown in Figure 5.1. To improve the scalability of standard cell synthesis using SMT solver, the partition technique and relative position constraints are introduced for FA and DFFHQN cells to perform simultaneous placement-and-route [10]. However, the standard cell synthesis framework still takes more than 2 hours for FA and DFFHQN standard cell in CFET structure even with the partition technique. In our study, using the placement order of each transistor can achieve $87.56\times$ and $121.45\times$ less runtime than partitioning the cell into groups of transistors

for 4.5T CFET FA and DFFHQN cells, respectively²⁴.

Upper/Lower M0A/PC Placement Canvas

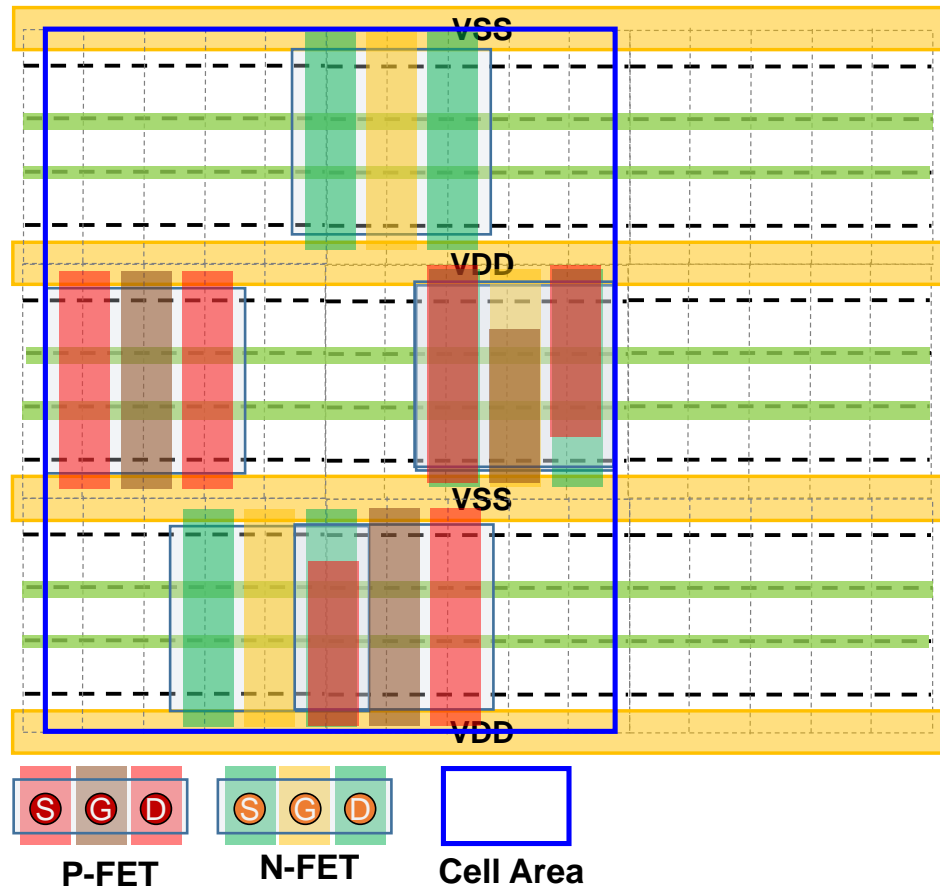


Figure 5.1: The illustration of transistor placement canvas for CFET standard cell architecture. There are diffusion sharing between PFET and NFET in the bottom cell row.

As a result, the reinforcement learning (RL) technique can be leveraged to improve the scalability of CFET standard cell synthesis framework by providing the placement order of transistors for the standard cells which have more than 30 transistors (i.e., multi-bits full adders). Given the placement order of transistors from reinforcement learning agent, SMT-based transistor placement and route framework provide the reward, which includes routability, cell area, and wire length, to the reinforcement learning

²⁴The runtime of 4.5T CFET FA cell is reduced from 6653.07s to 75.98s; The runtime of 4.5T CFET DFFHQN cell is reduced from 6831.77s to 56.25s.

agent for training as shown in Figure. 5.2.

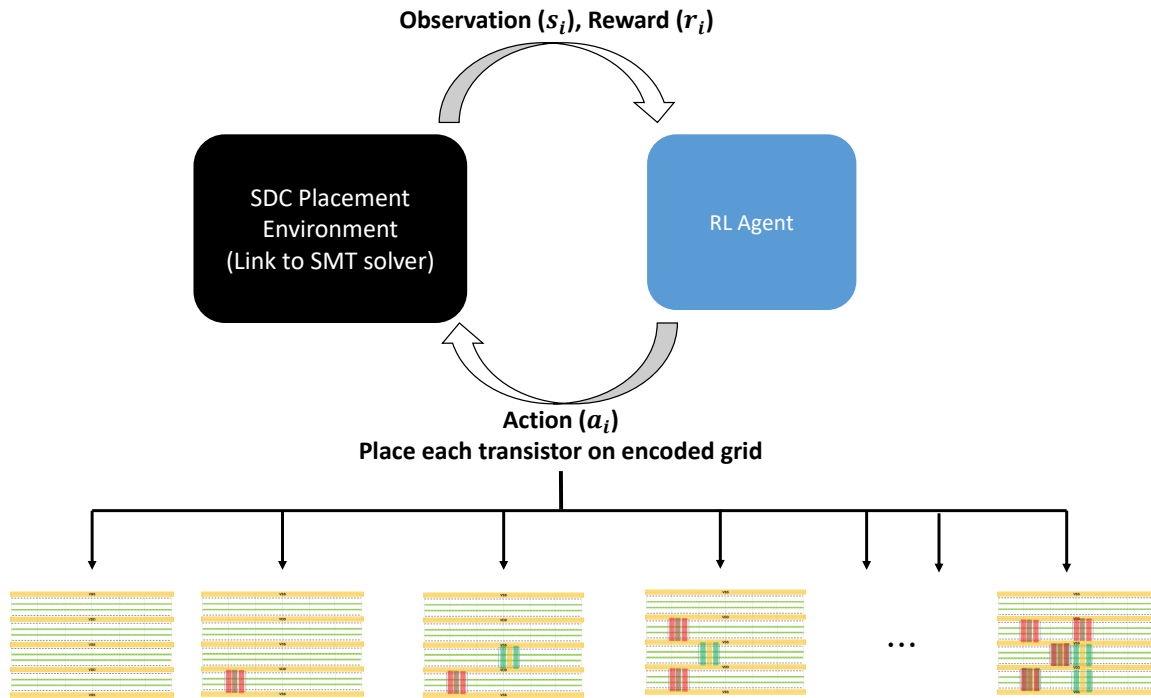


Figure 5.2: The illustration of CFET standard cell synthesis using reinforcement learning technique. The standard cell environment is linked to SMT solver for standard cell synthesis. Given the observation (s_i) and reward (r_i), the reinforcement learning agent places each transistor sequentially across multiple cell row and considering the stacking of FETs.

In the following subsections, we firstly show the action space of RL agent. Then, the observation (s_i) is introduced in the RL framework. Finally, we introduce the reward system of RL framework.

Observation and Feature Extraction

The observation in the RL framework includes the connection of transistors in the standard cells, type of transistor (i.e., NFET or PFET), stacking option (i.e., P-on-N or N-on-P), cell height, locations of placed transistors, and current transistor to be placed. We extract the static and dynamic features from the observation (s_i) to train the RL model. The static features include the adjacent matrix of transistors, cell height, type of transistor, stacking option, and cell height. The dynamic features are locations of placed

transistors, and current transistor to be placed.

Action Space of RL Agent on the Encoded Placement Grid

We introduce the encoded placement grid for the RL agent to consider the placement order of transistors in each cell row. Figure 5.3 shows the encoded placement grid. The PFET/NFET can not be placed overlapped to another PFET/NFET on the encoded grid. The grid coordinates of placed transistors are used to generate a set of relative position constraints as shown in Figure 5.3 to provide the placement order of transistors information to SMT-based CFET standard cell synthesis framework. Expression (5.1) show the relative position constraints of a pair of placed transistor (i.e., t^{th} and s^{th} FETs) on the encoded grid.

$$\mathbf{X\ direction:} \quad x_t \leq x_s + w_s + b, b = \begin{cases} 2(x_2 - x_1 - 1), \text{if } sqe = \text{false} \\ 0, \text{if } sqe = \text{true} \end{cases} \quad (5.1)$$

$$\mathbf{Y\ direction:} \quad y_t = y_s = y_1 = y_2$$

Where the (x_1, y_1) and (x_2, y_2) are the grid locations of s^{th} and t^{th} FETs on the encoded grid, respectively. If the *sqe* flag is set to true, the SMT-based transistor synthesis framework ignores the empty grid points between s^{th} and t^{th} FETs on the encoded grid. In the RL framework, the RL agent places transistors on the encoded grid sequentially as mentioned in Figure 5.2. As a result, the action space of RL agent is the number of grid points on the encoded grid.

Reward Function

In the RL framework, the routability can not be evaluated until all the transistors are placed. As a result, the reward is zero for the intermediate actions, and the final reward is a weighted sum of cell

Encoded upper/lower MOA/PC placement grid - Determine the order and cell row of transistors

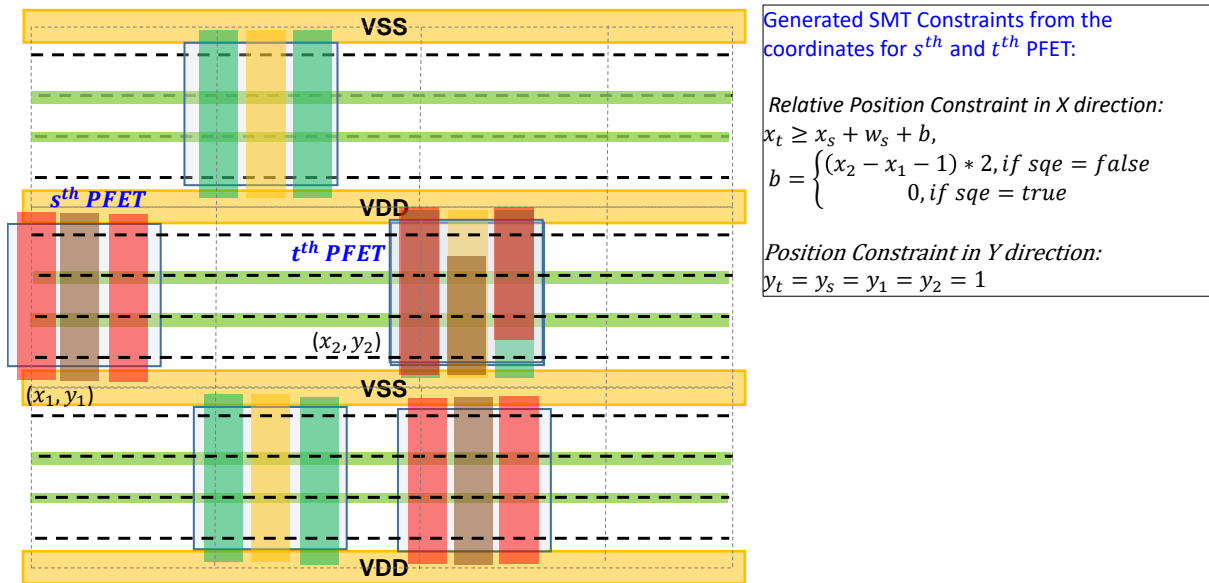


Figure 5.3: The illustration of the encoded placement grid. PFET/NFET is not allowed to be placed overlapped to another PFET/NFET. The grid coordinates of placed transistors are used to generate a set of relative position constraints in the SMT-based CFET standard cell synthesis framework.

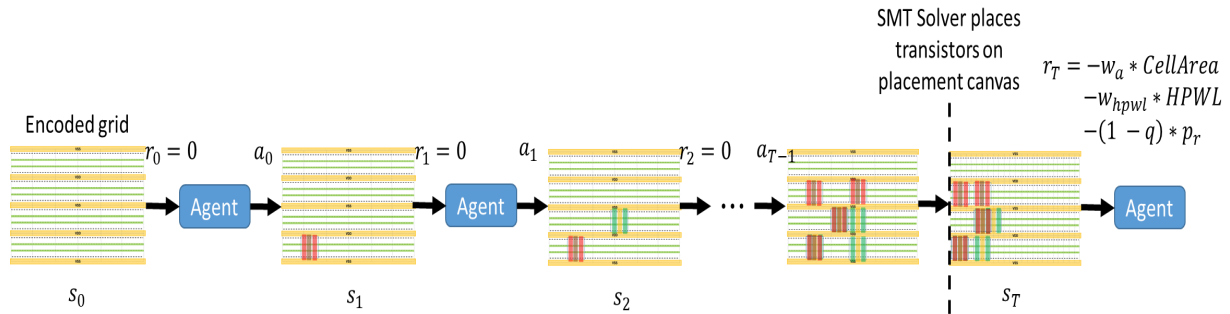


Figure 5.4: The illustration of the encoded placement grid. PFET/NFET is not allowed to be placed overlapped to another PFET/NFET. The grid coordinates of placed transistors are used to generate a set of relative position constraints in the SMT-based CFET standard cell synthesis framework.

size, half perimeter wire length (HPWL) of the nets of transistor terminals, and routability as shown in

Figure 5.4. Equation (5.2) shows the reward function for training a RL agent.

$$r_T = -w_a * CellArea - w_{hpwl} * HPWL - (1 - q) * p_r, q = \begin{cases} 1, \text{if routable} \\ 0, \text{if unroutable} \end{cases} \quad (5.2)$$

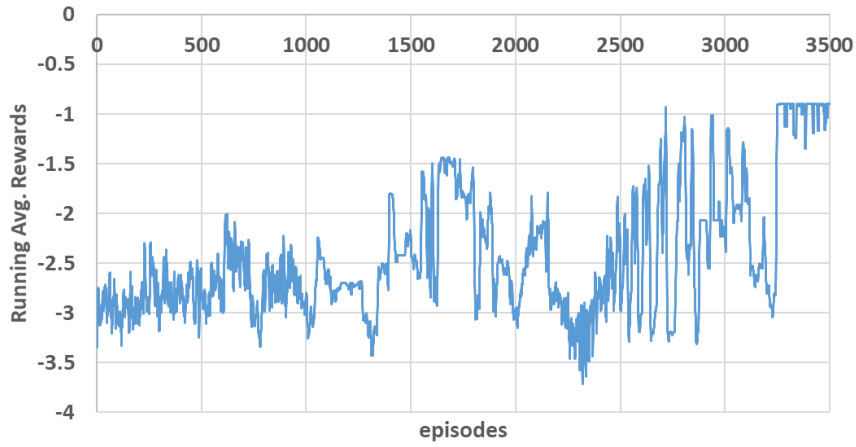
Where r_T is the final reward after placing all transistors on the encoded grid, and p_r is the penalty (i.e., a large number) when the placement of the cell is unroutable. w_a and w_{hpwl} are the weights of cell area and HPWL of the nets of transistor terminals, respectively.

Preliminary Experiments

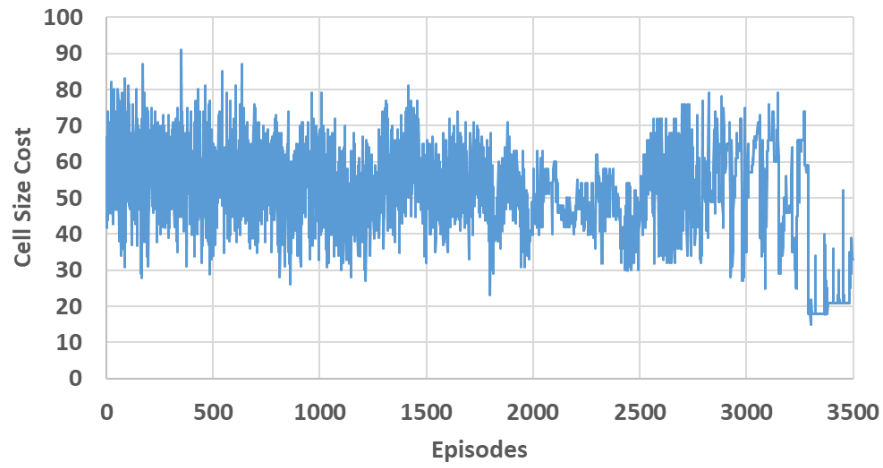
We implement the reinforcement learning framework in Python and link the SDC placement environment, which is implemented in Perl/SMT-LIB 2.0 standard-based formula. The experiments are executed on a workstation with 2.4GHz Intel Xeon E5-2620 CPU and 256GB memory. Here, the deep Q-learning algorithm [13] is implemented to train the RL agent. We demonstrate the reinforcement framework with XOR2x1 circuit design on a 2 by 16 encoded grid.

Figure 5.5 shows the reinforcement learning training plots of (a) running average rewards, (b) cell size cost, and (c) HPWL cost using deep Q-learning algorithm [13]. The average rewards, cell size, and hpwl converge after training RL agent with 3500 episodes. Figure 5.6 shows the XOR2x1 layouts from CFET synthesis framework [7] and the best reward of RL agent after policy converges. The trained RL agent successfully achieves the optimal cell size (11 CPPs) and total metal length with the reward system proposed in Figure 5.4. Moreover, with the information of FETs on the encoded grid from trained RL agent, the runtime of the proposed standard cell synthesis using RL methodology is $73.8\times$ faster than CFET synthesis framework [7].

(a) Running Avg. Reward



(b) Cell Size Cost



(c) HPWL Cost

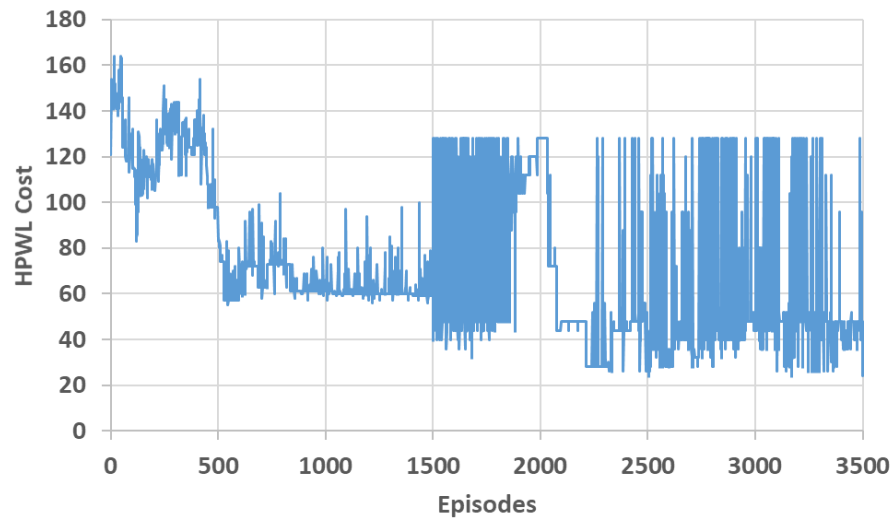
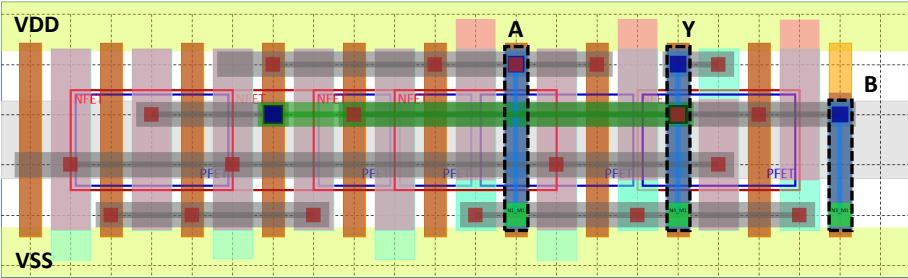


Figure 5.5: The reinforcement learning training plots of (a) running average rewards, (b) cell size cost, and (c) HPWL cost using deep Q learning [13].

(a) Optimal XOR2x1 Layout from SMT (748.70s)



(b) XOR2x1 Layout from trained RL agent (10.14s)

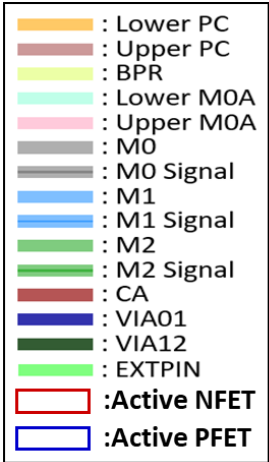
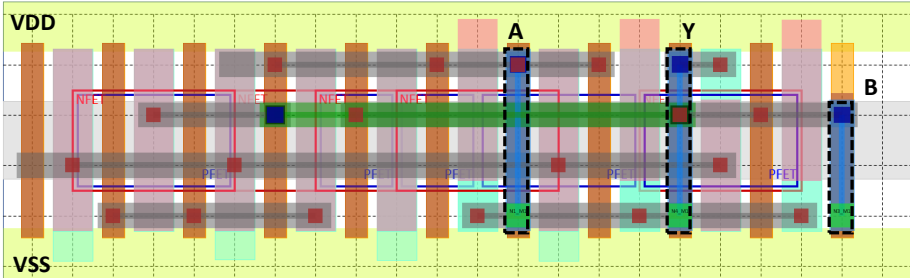


Figure 5.6: The XOR2x1 layouts from (a) CFET synthesis framework [7], and (b) the best reward of RL agent after the policy converges.

Bibliography

- [1] SM Yasser Sherazi, Miroslav Cupak, Pieter Weckx, Odysseas Zografos, Doyoung Jang, Peter Debacker, Diederik Verkest, Anda Mocuta, RH Kim, A Spessot, et al. Standard-cell design architecture options below 5nm node: The ultimate scaling of finfet and nanosheet. In *Design-Process-Technology Co-optimization for Manufacturability XIII*, volume 10962, page 1096202. SPIE, 2019.
- [2] Scotten Jones. Spie advanced lithography conference – imec design papers. <https://semiwiki.com/semiconductor-services/ic-knowledge/273127-spie-advanced-lithography-conference-imec-design-papers/>, 2019.
- [3] Jeffrey Smith. Design Technology Co-Optimization Approaches for Integration and Migration to CFET and 3D Logic. In *THE SURFACE PREPARATION AND CLEANING CONFERENCE (SPCC)*. Linx, 2019.
- [4] Lars Liebmann, Daniel Chanemougame, Peter Churchill, Jonathan Cobb, Chia-Tung Ho, Victor Moroz, and Jeffrey Smith. Dtco acceleration to fight scaling stagnation. In *Design-Process-Technology Co-optimization for Manufacturability XIV*, volume 11328, page 113280C. International Society for Optics and Photonics, 2020.
- [5] SMY Sherazi, JK Chae, P Debacker, L Matti, D Verkest, A Mocuta, RH Kim, A Spessot, A Dounde, and J Ryckaert. Cfet standard-cell design down to 3track height for node 3nm and below. In *Design-Process-Technology Co-optimization for Manufacturability XIII*, volume 10962, pages 16–27. SPIE, 2019.
- [6] Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, and Dongwon Park. A routability-driven complimentary-fet (cfet) standard cell synthesis framework using smt. In *2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pages 1–8. IEEE, 2020.
- [7] Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, Bill Lin, and Dongwon Park. Complementary-fet (cfet) standard cell synthesis framework for design and system technology co-optimization using smt. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 29(6):1178–1191, 2021.
- [8] Lawrence T Clark, Vinay Vashishtha, Lucian Shifren, Aditya Gujja, Saurabh Sinha, Brian Cline, Chandarasekaran Ramamurthy, and Greg Yeric. Asap7: A 7-nm finfet predictive process design kit. *Microelectronics Journal*, 53:105–115, 2016.

- [9] Dongwon Park, Daeyeal Lee, Ilgweon Kang, Sicun Gao, Bill Lin, and Chung-Kuan Cheng. Sp&r: Simultaneous placement and routing framework for standard cell synthesis in sub-7nm. In *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 345–350. IEEE, 2020.
- [10] Daeyeal Lee, Dongwon Park, Chia-Tung Ho, Ilgweon Kang, Hayoung Kim, Sicun Gao, Bill Lin, and Chung-Kuan Cheng. Sp&r: Smt-based simultaneous place-and-route for standard cell synthesis of advanced nodes. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 40(10):2142–2155, 2020.
- [11] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816, 2017.
- [12] Chung-Kuan Cheng, Chia-Tung Ho, Chester Holtz, and Bill Lin. Design and system technology co-optimization sensitivity prediction for vlsi technology development using machine learning. In *2021 ACM/IEEE International Workshop on System Level Interconnect Prediction (SLIP)*, pages 8–15. IEEE, 2021.
- [13] Hado Van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.
- [14] Cadence Innovus User Guide. <http://www.cadence.com>, 2020.
- [15] Clark Barrett and Cesare Tinelli. Satisfiability modulo theories. In *Handbook of model checking*, pages 305–343. Springer, 2018.
- [16] Nikolaj Bjørner, Anh-Dung Phan, and Lars Fleckenstein. vz-an optimizing smt solver. In *International conference on Tools and Algorithms for the Construction and Analysis of Systems*, pages 194–199. Springer, 2015.
- [17] Roberto Sebastiani and Patrick Trentin. Optimathsat: A tool for optimization modulo theories. *Journal of Automated Reasoning*, 64(3):423–460, 2020.
- [18] Suphachai Sutanthavibul, Eugene Shragowitz, and J Ben Rosen. An analytical approach to floorplan design and optimization. In *Proc. DAC*, pages 187–192, 1991.
- [19] Ilgweon Kang, Dongwon Park, Changho Han, and Chung-Kuan Cheng. Fast and precise routability analysis with conditional design rules. In *Proceedings of the 20th System Level Interconnect Prediction Workshop*, pages 1–8, 2018.
- [20] Xiaotao Jia, Yici Cai, Qiang Zhou, and Bei Yu. A multicommodity flow-based detailed router with efficient acceleration techniques. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(1):217–230, 2017.
- [21] Dongwon Park, Ilgweon Kang, Yeseong Kim, Sicun Gao, Bill Lin, and Chung-Kuan Cheng. Road: Routability analysis and diagnosis framework based on sat techniques. In *Proceedings of the 2019 International Symposium on Physical Design*, pages 65–72, 2019.
- [22] Yuangsheng Ma, Jason Sweis, Hidekazu Yoshida, Yan Wang, Jongwook Kye, and Harry J Levinson. Self-aligned double patterning (sadb) compliant design flow. In *Design for Manufacturability through Design-Process Integration VI*, volume 8327, pages 49–61. SPIE, 2012.

- [23] Mohan Guruswamy, Robert L Maziasz, Daniel Dulitz, Srilata Raman, Venkat Chiluvuri, Andrea Fernandez, and Larry G Jones. Cellularity: A fully automatic layout synthesis system for standard cell libraries. In *Proceedings of the 34th annual Design Automation Conference*, pages 327–332, 1997.
- [24] Adriel Mota Ziesemer and Ricardo Augusto da Luz Reis. Simultaneous two-dimensional cell layout compaction using milp with astran. In *2014 IEEE Computer Society Annual Symposium on VLSI*, pages 350–355. IEEE, 2014.
- [25] Pascal Cremer, Stefan Hougardy, Jan Schneider, and Jannik Silvanus. Automatic cell layout in the 7nm era. In *Proceedings of the 2017 ACM on International Symposium on Physical Design*, pages 99–106, 2017.
- [26] Yih-Lang Li, Shih-Ting Lin, Shinichi Nishizawa, Hong-Yan Su, Ming-Jie Fong, Oscar Chen, and Hidetoshi Onodera. Nctucell: A dda-aware cell library generator for finfet structure with implicitly adjustable grid map. In *2019 56th ACM/IEEE Design Automation Conference (DAC)*, pages 1–6. IEEE, 2019.
- [27] Kyeongrok Jo, Seyong Ahn, Jungho Do, Taejoong Song, Taewhan Kim, and Kyumyung Choi. Design rule evaluation framework using automatic cell layout generator for design technology co-optimization. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 27(8):1933–1946, 2019.
- [28] Wei Ye, Bei Yu, David ZO Pan, Yong-Chan Ban, and Lars Liebmann. Standard cell layout regularity and pin access optimization considering middle-of-line. In *Proceedings of the 25th edition on Great Lakes Symposium on VLSI*, pages 289–294, 2015.
- [29] Jaewoo Seo, Jinwook Jung, Sangmin Kim, and Youngsoo Shin. Pin accessibility-driven cell layout redesign and placement optimization. In *2017 54th ACM/EDAC/IEEE Design Automation Conference (DAC)*, pages 1–6. IEEE, 2017.
- [30] Xiaoqing Xu, Bei Yu, Jih-Rong Gao, Che-Lun Hsu, and David Z Pan. Parr: Pin-access planning and regular routing for self-aligned double patterning. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 21(3):1–21, 2016.
- [31] Nikolay Ryzhenko, Steven Burns, Anton Sorokin, and Mikhail Talalay. Pin access-driven design rule clean and dfm optimized routing of standard cells under boolean constraints. In *Proceedings of the 2019 International Symposium on Physical Design*, pages 41–47, 2019.
- [32] Chung-Kuan Cheng, Daeyeal Lee, and Dongwon Park. Standard-cell scaling framework with guaranteed pin-accessibility. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2020.
- [33] OpenCores: Open-Source IP Cores. <https://opencores.org/>, 2020.
- [34] LEF/DEF Language Reference. <http://www.ispd.cc/contests/18/lefdefref.pdf>, 2020.

- [35] Anshul Gupta, Shreya Kundu, Lieve Teugels, Jürgen Bommels, Christoph Adelman, Nancy Heylen, Geraldine Jamieson, Olalla Varela Pedreira, Ivan Ciofi, Bharani Chava, et al. High-aspect-ratio ruthenium lines for buried power rail. In *2018 IEEE International Interconnect Technology Conference (IITC)*, pages 4–6. IEEE, 2018.
- [36] Bharani Chava, Julien Ryckaert, Luca Mattii, Syed Muhammad Yasser Sherazi, Peter Debacker, Alessio Spessot, and Diederik Verkest. Dtco exploration for efficient standard cell power rails. In *Design-Process-Technology Co-optimization for Manufacturability XII*, volume 10588, page 105880B. International Society for Optics and Photonics, 2018.
- [37] J Ryckaert, A Gupta, A Jourdain, B Chava, G Van der Plas, D Verkest, and E Beyne. Extending the roadmap beyond 3nm through system scaling boosters: A case study on buried power rail and backside power delivery. In *2019 Electron Devices Technology and Manufacturing Conference (EDTM)*, pages 50–52. IEEE, 2019.
- [38] Lars Liebmann. Design Technology Co-Optimization for 3nm and Beyond. In *IEDM*. IEEE, 2019.
- [39] Ryoung-han Kim, Yasser Sherazi, Peter Debacker, Praveen Raghavan, Julien Ryckaert, Arindam Malik, Diederik Verkest, Jae Uk Lee, Werner Gillijns, Ling Ee Tan, et al. Imec n7, n5 and beyond: Dtco, stco and euv insertion strategy to maintain affordable scaling trend. In *Design-Process-Technology Co-optimization for Manufacturability XII*, volume 10588, page 105880N. International Society for Optics and Photonics, 2018.
- [40] Tetsuya Iizuka, Makoto Ikeda, and Kunihiro Asada. Exact minimum-width multi-row transistor placement for dual and non-dual cmos cells. In *2006 IEEE International Symposium on Circuits and Systems*, pages 4–pp. IEEE, 2006.
- [41] Yih-Lang Li, Shih-Ting Lin, Shinichi Nishizawa, and Hidetoshi Onodera. Mcell: multi-row cell layout synthesis with resource constrained max-sat based detailed routing. In *Proceedings of the 39th International Conference on Computer-Aided Design*, pages 1–8, 2020.
- [42] Lawrence T Clark, Vinay Vashishtha, David M Harris, Samuel Dietrich, and Zunyan Wang. Design flows and collateral for the asap7 7nm finfet predictive process design kit. In *2017 IEEE International Conference on Microelectronic Systems Education (MSE)*, pages 1–4. IEEE, 2017.
- [43] Pieter Weckx, Julien Ryckaert, E Dentoni Litta, Dmitry Yakimets, Philippe Matagne, Pieter Schud-dinck, Doyoung Jang, Bilal Chehab, Rogier Baert, Mohit Gupta, et al. Novel forksheet device architecture as ultimate logic scaling device towards 2nm. In *2019 IEEE International Electron Devices Meeting (IEDM)*, pages 36–5. IEEE, 2019.
- [44] Dongwon Park, Daeyeal Lee, Ilgweon Kang, Chester Holtz, Sicun Gao, Bill Lin, and Chung-Kuan Cheng. Grid-based framework for routability analysis and diagnosis with conditional design rules. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 39(12):5097–5110, 2020.
- [45] Lars W Liebmann and Rasit O Topaloglu. Design and technology co-optimization near single-digit nodes. In *Proc. ICCAD*, pages 582–585. IEEE, 2014.

- [46] SC Song, J Xu, D Yang, K Rim, P Feng, Jerry Bao, J Zhu, J Wang, G Nallapati, Mustafa Badaroglu, et al. Unified technology optimization platform using integrated analysis (utopia) for holistic technology, design and system co-optimization at ≤ 7 nm nodes. In *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, pages 1–2. IEEE, 2016.
- [47] Alex Kahng, Andrew B Kahng, Hyein Lee, and Jiajia Li. Probe: A placement, routing, back-end-of-line measurement utility. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(7):1459–1472, 2017.
- [48] Zhe Zhang, Runsheng Wang, Cheng Chen, Qianqian Huang, Yangyuan Wang, Cheng Hu, Dehuang Wu, Joddy Wang, and Ru Huang. New-generation design-technology co-optimization (dtco): Machine-learning assisted modeling framework. In *2019 Silicon Nanoelectronics Workshop (SNW)*, pages 1–2. IEEE, 2019.
- [49] A Ceyhan, J Quijas, S Jain, H-Y Liu, WE Gifford, and S Chakravarty. Machine learning-enhanced multi-dimensional co-optimization of sub-10nm technology node options. In *2019 IEEE International Electron Devices Meeting (IEDM)*, pages 36–6. IEEE, 2019.
- [50] Chung-Kuan Cheng, Andrew B Kahng, Hayoung Kim, Minsoo Kim, Daeyeal Lee, Dongwon Park, and Mingyu Woo. Probe2. 0: A systematic framework for routability assessment from technology to design in advanced nodes. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2021.
- [51] Phillip Christie and Dirk Stroobandt. The interpretation and application of rent’s rule. *IEEE Trans. on VLSI*, 8(6):639–648, 2000.
- [52] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [53] Roman Salmerón Gómez, José García Pérez, María Del Mar López Martín, and Catalina García García. Collinearity diagnostic applied in ridge estimation through the variance inflation factor. *Journal of Applied Statistics*, 43(10):1831–1849, 2016.
- [54] Bradley Efron and Robert Tibshirani. Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.
- [55] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.
- [56] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [57] Chung-Kuan Cheng, Chia-Tung Ho, Daeyeal Lee, and Bill Lin. Multirow complementary-fet (cfet) standard cell synthesis framework using satisfiability modulo theories (smts). *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, 7(1):43–51, 2021.
- [58] Afshin Gholamy, Vladik Kreinovich, and Olga Kosheleva. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. 2018.

- [59] Kyungwook Chang, Abhishek Koneru, Krishnendu Chakrabarty, and Sung Kyu Lim. Design automation and testing of monolithic 3d ics: Opportunities, challenges, and solutions. In *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 805–810. IEEE, 2017.
- [60] Bilal Chehab, Julien Ryckaert, Pieter Schuddinck, Pieter Weckx, Naoto Horiguchi, Gioele Mirabelli, Alessio Spessot, and Myunghee Na. Design-technology co-optimization of sequential and monolithic cfet as enabler of technology node beyond 2nm. In *Design-Process-Technology Co-optimization XV*, volume 11614, page 116140D. International Society for Optics and Photonics, 2021.
- [61] Bei Yu, Xiaoqing Xu, Subhendu Roy, Yibo Lin, Jiaojiao Ou, and David Z Pan. Design for manufacturability and reliability in extreme-scaling vlsi. *Science China Information Sciences*, 59(6):1–23, 2016.
- [62] Pengpeng Ren, Xiaoqing Xu, Peng Hao, Junyao Wang, Runsheng Wang, Ming Li, Jianping Wang, Weihai Bu, Jingang Wu, Waisum Wong, et al. Adding the missing time-dependent layout dependency into device-circuit-layout co-optimization-new findings on the layout dependent aging effects. In *2015 IEEE International Electron Devices Meeting (IEDM)*, pages 11–7. IEEE, 2015.
- [63] Daeyeal Lee, Chia-Tung Ho, Ilgweon Kang, Sicun Gao, Bill Lin, and Chung-Kuan Cheng. Many-tier vertical gate-all-around nanowire fet standard cell synthesis for advanced technology nodes. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, 7(1):52–60, 2021.
- [64] Nvidia deep learning accelerator (nvdla). <https://github.com/nvdla/hw>, 2018.