

AI 與機器學習賦能的 晶片設計與製程協同優化

DTCO.ML™ 與 DTCO.GenAI™ 的未來趨勢
(Draft)

Hock Chen

嶺智科技股份有限公司

DIGWISE TECHNOLOGY CORPORATION, LTD.

目錄

AI 驅動的半導體晶片設計能效與生產力革命	5
第一部分：設計與製程協同優化 (Design Technology Co-Optimization, DTCO).....	6
第 1 章 DTCO 概述與發展.....	6
1.1. 為生產力而設計 (Design for Productivity) 的核心理念	7
1.2. 極致能效設計方法論 (Design for Ultimate Efficiency)	8
1.3. DTCO 的未來發展方向	9
第 2 章 推動 DTCO 的關鍵挑戰與策略	9
2.1. DTCO 實施面臨的主要挑戰	9
2.2. 設計方法創新需求.....	10
2.3. 生產力優化平台建置.....	10
第 3 章 晶片能效與生產力優化實踐.....	11
3.1. 項目啟動前的準備工作.....	11
3.2. 客製化元件與時序策略 (Timing Signoff Strategy).....	12
3.3. 製程優化分析技術.....	13
3.4. 補償機制的設計與實施.....	14
3.5. 近臨界電壓技術的需求與挑戰.....	16
第二部分：DTCO.ML™ - 機器學習驅動的半導體流程優化.....	18
第 4 章 機器學習與 DTCO 的融合 (DTCO.ML™).....	18
4.1. 虛擬晶圓數據建模(Virtual Silicon).....	19
4.2. 回歸模型的構建與推論.....	23
4.3. 數據追蹤與產能優化的應用.....	26

第 5 章 元件庫特徵萃取與分析系統 (Library Metric Extraction, libMetric™)	28
5.1. 元件時序與功耗建模.....	30
5.2. 標準元件特徵萃取 (Cell Feature Extraction)	32
5.3. RO Simulation.....	33
5.4. 標準元件庫批量 PPA 基準評估	34
第 6 章 晶片內感測器設計與整合 (GRO Compiler)	37
6.1. 目標導向的 RO 設計.....	39
6.2. SPICE-Silicon 相關性分析.....	40
6.3. 製程追蹤與調整.....	42
6.4. 晶片內局域電壓分析.....	43
6.4.1. 晶片內局域電壓分布監測.....	43
6.4.2. 補償策略制定.....	46
6.4.3. 動態時序餘量警示與佈局.....	49
6.5. GRO 自動化工具與驗證流程	50
第 7 章 數據分析與機器學習平台 (Copernic™)	52
7.1. 數據標準化與可視化實踐.....	52
7.2. 多維資料的跨域映射.....	54
7.3. 設計流程整合策略.....	55
7.3.1. WAT-aware Timing Re-K	55
7.3.2. WAT-CP 映射與相關性分析	56
7.3.3. OCV Analysis.....	56
7.4. 晶片內變異性(OVC) 分析與設計餘量優化	57
7.5. 後矽製程分析與優化 (Post-Silicon Analysis and Optimization)	59
第 8 章 晶片性能評等策略與優化 (Binning-PG™)	61
8.1. 晶片評等與分箱策略對生產力的影響.....	61

8.2.	晶片體質的分析與挑戰.....	62
8.3.	分箱策略生成 (Binning Policy Generation, Binning-PG™).....	65
8.4.	策略生成自動化與優化.....	67
8.5.	晶片內自分級應用 (On-chip Self-binnig)	68
第三部分：DTCO.GenAI™ - 生成式 AI 驅動的晶片設計創新		70
第 9 章 生成式 AI 與 DTCO 融合 (DTCO.GenAI™)		70
9.1.	傳統建模方法的局限性分析.....	70
9.2.	按圖索驥：Multivariate Normal Distribution	72
9.3.	虛擬晶片數據在 DTCO 中的實際應用 (DTCO.VS).....	74
第 10 章 DTCO.VS 虛擬晶圓數據生成技術 (Virtual Silicon)		77
10.1.	資料集準備.....	77
10.2.	基於生成對抗網路的虛擬晶片數據 (GAN-based Virtual Silicon, GAN-VS)	80
10.2.1.	GAN 模型	80
10.2.2.	GAN 模型性能評估	81
10.3.	基於擴散模型的虛擬晶片數據 (Diffusion Model-based Virtual Silicon, DM-VS)	84
10.3.1.	去噪擴散概率模型 (Denoising Diffusion Probabilistic Model)	85
10.3.2.	Diffusion 模型性能評估	87
第 11 章 Generative AI 驅動的晶片能效優化與建模.....		91
11.1.	WAT 超分辨率 (WAT Super Resolution, WAT-SR)	91
11.2.	高效矽製程偏移建模 (High-Efficiency SPICE-Silicon Bias Modeling, He-SSBM)	96
11.2.1.	One-shot SPICE-Silicon N/P Correlation 的設計原理.....	96
11.2.2.	設計與防守策略優化.....	97
11.3.	高效蒙地卡羅仿真近似 (High-Fidelity Generative Monte Approximation, H Σ -GMA)	101
11.3.1.	傳統 Monte Carlo 方法的限制	101

11.3.2. 生成式神經網路的創新應用.....	101
第 12 章 結論與展望.....	104
12.1. 機器學習與 AI 賦能 DTCO：革新晶片設計與製程優化 (DTCO.ML™)	105
12.2. 生成式 AI 驅動優化的未來趨勢 (DTCO.GenAI™)	105
12.3. 創新 EDA 開發與未來展望.....	105
附錄.....	106
開源資源表列.....	106
參考文獻表列.....	106
專業術語表.....	107

AI 驅動的半導體晶片設計能效與生產力革命

前言

在 AI 生成時代，半導體產業正從傳統的良率提升，邁向以創新思維推動整體生產力優化。為在市場競爭中脫穎而出，晶片設計的焦點已不再局限於良率和成本效益，而是全面提升性能、產能與市場競爭力。本書將深入探討如何利用 AI 技術，全面優化設計餘量、時序簽核、測試策略、數據分析、製程配方調整、分箱策略 (Binning Strategy) 及系統層級補償方案，從而推動半導體設計與生產力的革新。

書籍定位

本書以晶片實體設計流程的視角切入，聚焦於機器學習技術 (ML) 在半導體產業與製程協同優化 (DTCO) 領域的應用。內容涵蓋從傳統機器學習到最前沿的生成式 AI 技術 (GenAI)，全面探討如何有效提升晶片設計效能與生產力。

目標讀者

半導體產業人士：為晶片設計工程師、EDA 開發者及學術研究者，提供實例參考與技術洞見。

相關及跨領域的研究者：幫助加深對 DTCO 技術的理解，並探索 AI 技術的應用潛力。

書籍架構

本書將介紹 DTCO 的基本概念與原理，深入探討機器學習在設計與製程優化中的應用，並分析生成式 AI 如何塑造半導體設計的創新與未來發展。此外，我們將探討創新性 EDA 發展的潛力，探索新技術如何加速設計流程並提升效能。

第一部分：設計與製程協同優化 (Design Technology Co-Optimization, DTCO)

第1章 DTCO 概述與發展

設計與製程協同優化 (DTCO) 方法已廣泛應用於半導體實體設計流程中，以提升晶片生產效率和競爭力。如圖 Fig. 1-1 所示，DTCO 流程可類比為一個大規模神經網路的優化過程。

在前向推論 (Inference) 階段，其核心目標是提升產能 (Productivity)，涵蓋晶片監控、WAT (Wafer Acceptance Test)、CP (Chip Probe) / FT (Final Test)、SLT (System-level Test)、特徵相關性分析及機器學習，最終實現良率與性能的最大化。

在反向傳播優化 (Back-propagation) 階段，則專注於晶片能效 (Efficiency) 的提升，方法包括利用實際測量數據校準晶片模型、調整製程參數、根據 WAT 測量結果校正時序萃取 (Re-K)、定制與優化元件庫、執行 OCV (On-chip Variation) 回歸分析，以及改進設計餘量 (Margin) 與時序簽核策略 (Timing Signoff)。

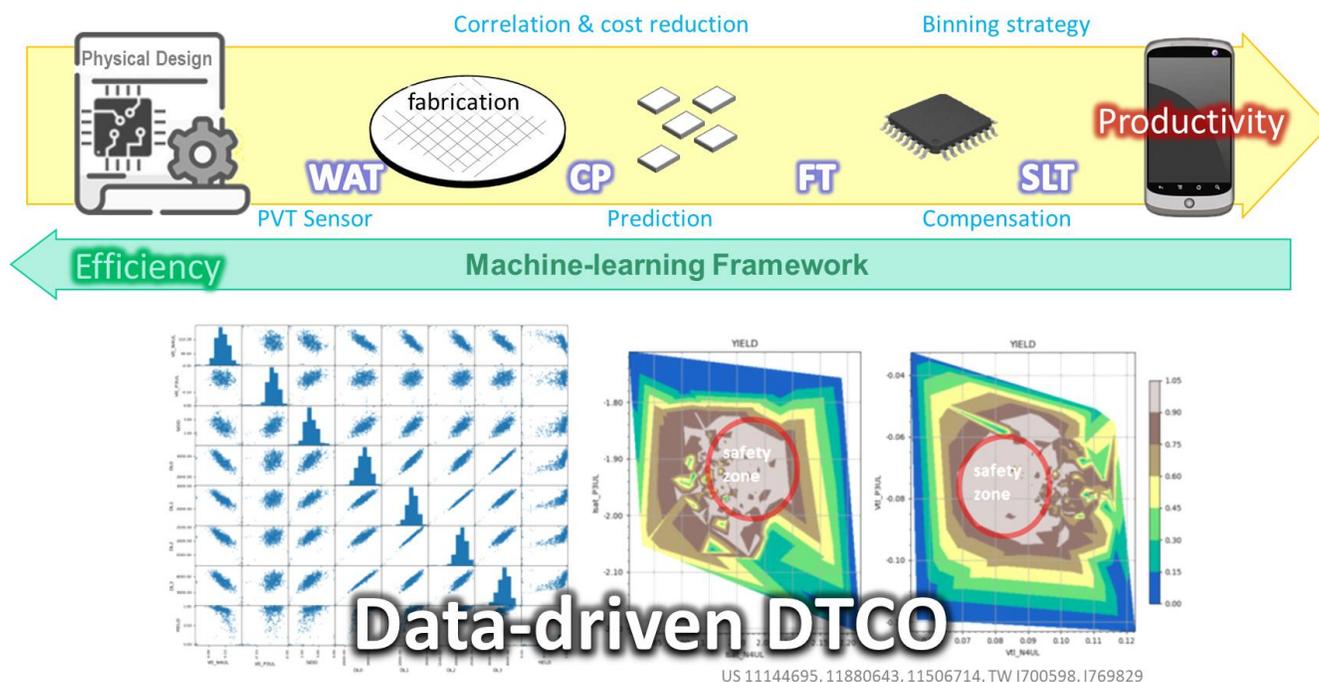


Fig. 1-1 Unleashing Productivity and Efficiency with DTCO

1.1. 為生產力而設計 (Design for Productivity) 的核心理念

晶圓製造過程中常見缺陷包括對稱的共振圖樣（如同心圓波紋/甜甜圈狀波浪）、晶圓拋光過程中的刮痕（Polish Pattern）、氧化層不平坦的傾斜，以及光罩曝光干涉與測試環境不匹配（如 Load Board 和 Probe Card）等引發的系統級缺陷（Test Site Pattern）。

這些缺陷最終與物理隨機變異性相互疊加，對元件行為產生顯著影響。如圖 Fig. 1-2 所示，SIDD（漏電流）與 RO（Ring Oscillator）等關鍵參數均可能受到干擾，特別是類比監測電路（如 V/T Sensor）對測試環境不匹配尤為敏感。業界往往聚焦於微觀參數的調整（如金屬層厚度與寄生電容），卻容易忽略設計餘量受侵蝕的系統性挑戰及其應對策略。

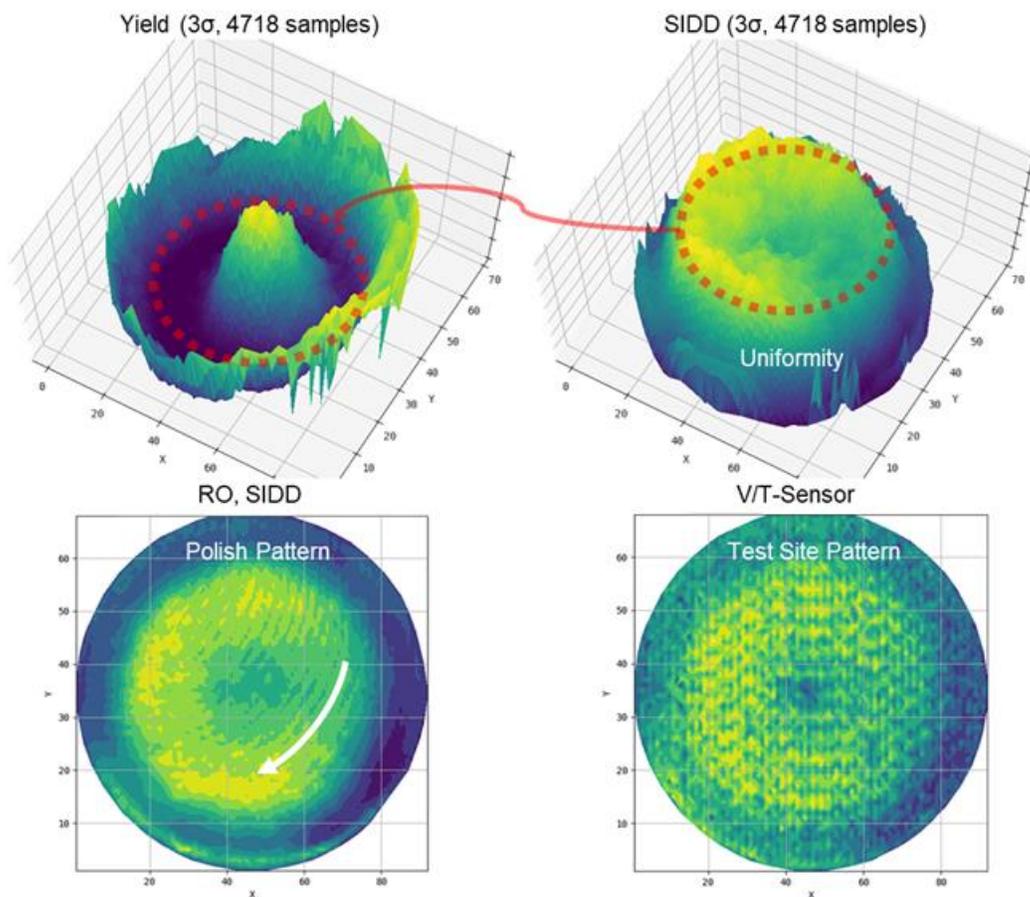


Fig. 1-2 Impact of Wafer Defects on System Performance Metrics

為了有效應對這些系統性挑戰，透過結合 WAT/CP 分析與製程控制管理（PCM），定期追蹤量產期間的製程參數與晶片測試數據，重點關注各參數的分佈、機率密度及 SPICE-to-Silicon（S2S）

分析，以實現系統性的補償與後矽調整。藉由映射與關聯 WAT 和 CP 數據，能夠識別製程優化窗口，進而制定晶片性能分級與分箱策略（Binning Strategy），同時應用補償技術，全面提升生產效率、良率與晶片競爭力。

1.2. 極致能效設計方法論 (Design for Ultimate Efficiency)

在設計階段，透過製程與元件庫分析，優化晶片架構中關鍵元件的驅動力、面積與功耗，以提升整體可靠性。針對關鍵元件進行定制化設計與微架構優化，並運用晶片內監控器（On-chip Sensor）執行防禦策略，有效應對製程變異與動態電壓降（IR-drop）等問題，確保設計的穩健性。進入生產階段後，透過機器學習技術分析數據實際分佈情況，制定時序 Re-K 與簽核策略（Timing Signoff），深入掌握製程均勻性（Uniformity）與 OCV，實現最佳設計餘量。

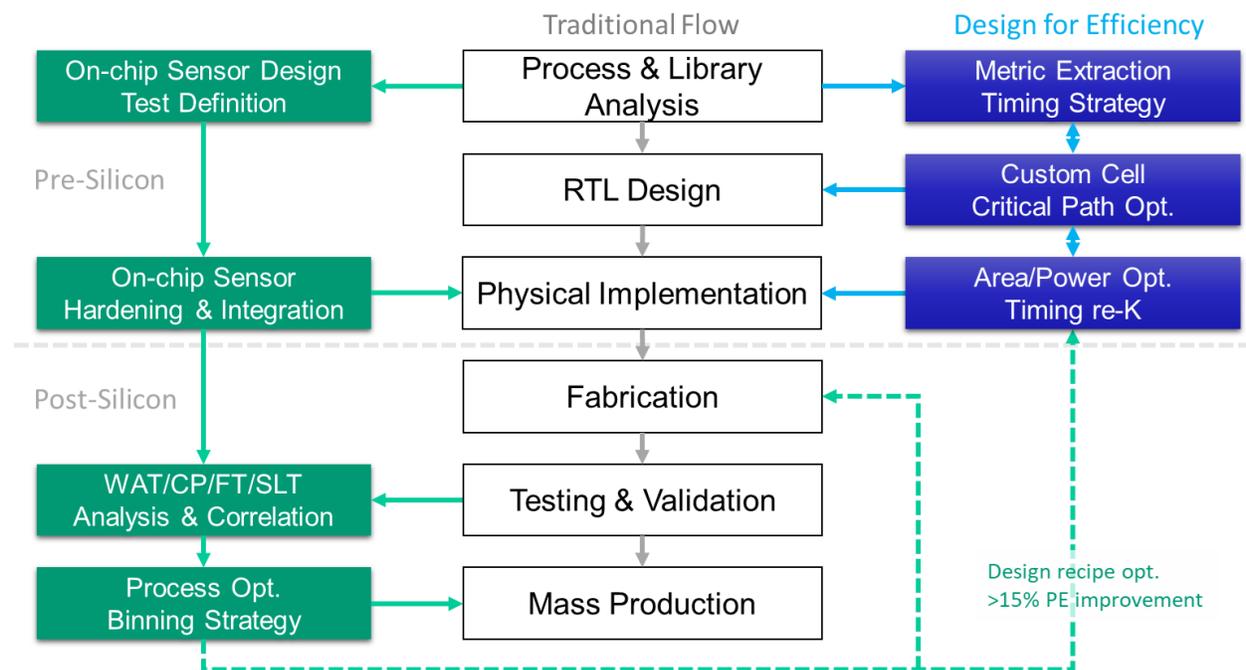


Fig. 1-3 DTCO Physical Design Flow

圖 Fig. 1-3 展示了 DTCO 的實施細節，涵蓋矽前準備與後矽分析與反饋，並以數據與機器學習優化流程為驅動。矽前準備包括製程分析、SPICE 仿真與元件庫分析、特徵提取與建模、關鍵路徑（Critical Path）性能提升，以及面積與功耗優化。後矽分析則涵蓋晶片內監控電路的設計與整合、測試、數據相關性分析、機器學習平台的應用、分箱策略的自動化，以及優化反饋。

1.3. DTCO 的未來發展方向

隨著機器學習 (ML) 與生成式 AI (GenAI) 的引入，DTCO 的潛力正被進一步釋放。從數據驅動的設計分析到生成逼真的虛擬晶片數據 (Virtual Silicon)，DTCO 不僅解決當前挑戰，更有望成為推動未來半導體設計革新的核心動力。後續章節將深入探討如何通過機器學習平台 DTCO.ML™ 與生成式模型平台 DTCO.GenAI™，徹底革新設計效率與生產力，為半導體行業開啟全新可能。

第2章 推動 DTCO 的關鍵挑戰與策略

2.1. DTCO 實施面臨的主要挑戰

雖然 DTCO 具備巨大潛力，但在實現過程中仍面臨諸多挑戰，如圖 Fig. 2-1：

- **數據驅動的瓶頸**：DTCO 的實現依賴大量準確的製程與設計數據，但數據收集與整合經常受到商業機密與技術壁壘的限制。
- **跨領域專業知識的需求**：DTCO 涉及設計工程師與製程專家之間的緊密合作，雙方需要深入了解彼此的技術細節，這對於傳統的分工模式而言是一項重大挑戰。
- **創新方法的支持**：現有 EDA 工具與演算法多聚焦於單一層面，缺乏對設計與製程交互的全局性與跨域整合，難以高效應對 DTCO 的複雜性，從而限制優化潛力的發揮。

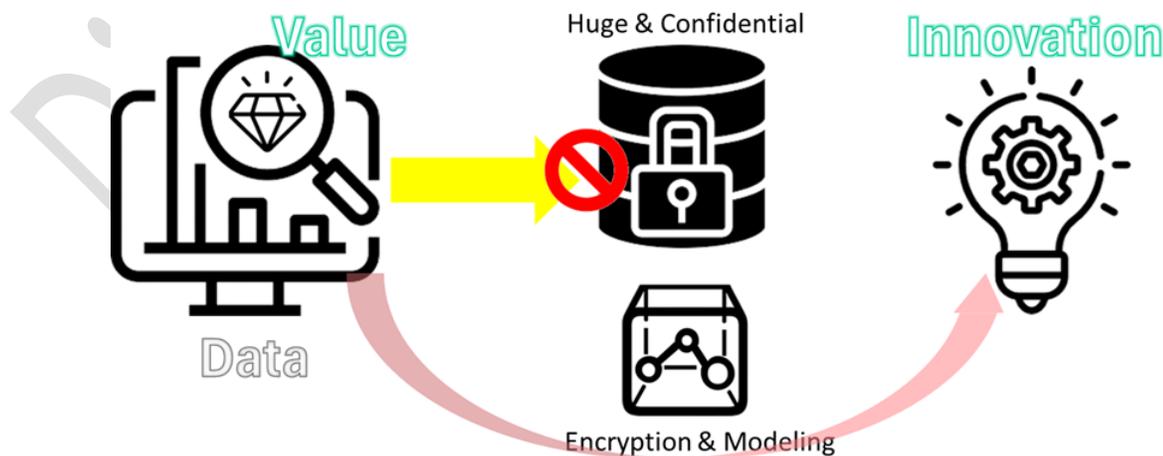


Fig. 2-1 Data Acquisition Barriers

2.2. 設計方法創新需求

晶片高可靠設計與精準監控正為市場帶來全新機遇，特別是在智慧裝置、車載系統、低軌衛星（LEO）、光通訊（CPO）及醫療等領域。包括提供高精度的電源管理系統以延長續航時間，以及結合超低能耗（ULE）與人工智慧（AI）技術進行推論的微控制器（MCU）。

- **可靠性與價值提升**：利用機器學習和人工智能優化性能，增強晶片設計的可靠性與監控能力。同時，結合製程與晶片優化，靈活應對市場需求，提升內部 IP 和 EDA 開發的價值。
- **精準設計與管理**：透過晶片監測與數據分析，深入洞察製程與元件的物理特性，精確管理設計餘量與系統補償方案，探索創新設計理念，確保晶片性能穩定。
- **高效能設計**：結合 AI 和生成模型，重構設計流程與配套之 EDA 工具，實現效率與生產力的雙重提升，推動晶片設計邁向智能化與高效化。

2.3. 生產力優化平台建置

針對 DTCO 實施所面臨的挑戰，生產力優化平台的建置成為關鍵解方，專注提升晶片競爭力與生產力，整合系統、人才、工具與方法：

- **整合數據中心建置**：構建集中化的系統架構，用於整合提取與追蹤設計指標、製程參數及生產數據。這不僅能解決數據分散的瓶頸，還為設計與製程交互優化提供高效數據支持。
- **跨領域專業能力培育**：加強數據科學人才的培養，提升程式設計與數據分析技能，並通過視覺化工具與相關性分析促進設計與製程專家的深度協作，消除技術壁壘。
- **工具創新與標準化**：開發支持 DTCO 全面優化的新型專用 EDA 工具，並推行監控 IP 與數據格式標準化，優化設計與測試方法，打破現有工具在設計與製程交互處理上的局限。
- **AI 與機器學習應用**：深度應用機器學習，通過特徵建模、回歸分析與趨勢預測，實現設計與製程的智能交互優化，涵蓋變異分析、晶片性能分級及動態補償，有效應對複雜的交互挑戰，推動 DTCO 整體效率提升。

第3章 晶片能效與生產力優化實踐

圖 Fig. 3-1 顯示了專案執行的主要階段：設計、製造（封裝與測試）及量產。各階段均提供關鍵支持方案，確保整個流程的生產力與效率達到最優化。

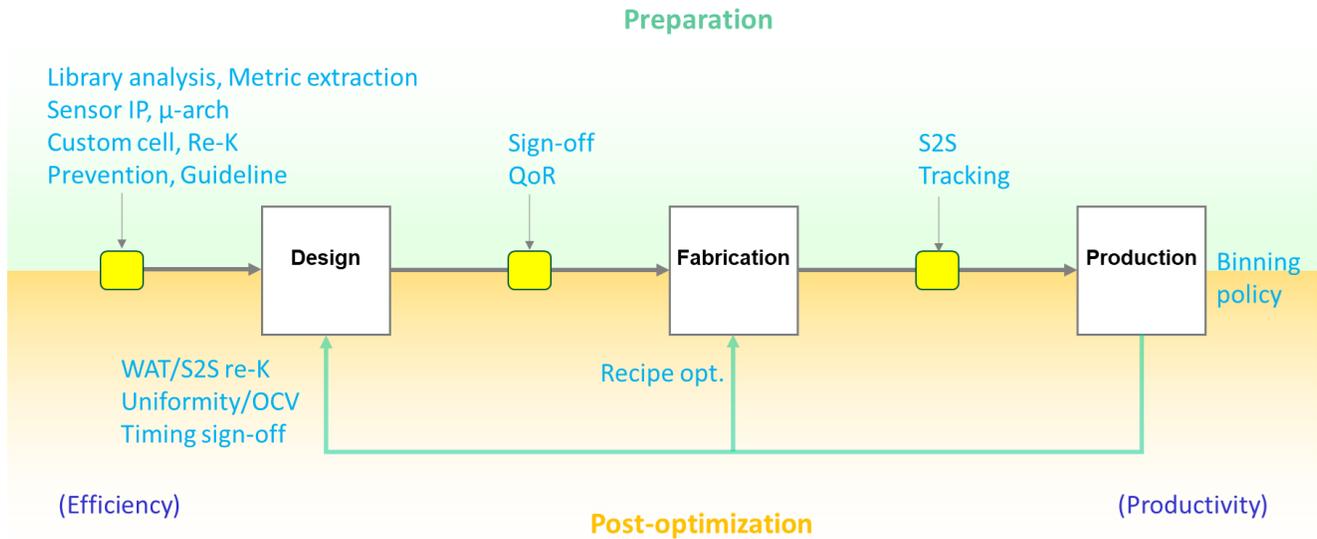


Fig. 3-1 晶片生產力與效率優化實踐

3.1. 項目啟動前的準備工作

- **設計與製程特徵提取**：開發整合工具，將分散的元件庫文件（如面積、時序、功耗數據）匯總至單一數據庫。通過數值分析與特徵量化，快速且精確地提取基本元件的物理參數。
- **製程分析**：包括 WAT 特徵分佈分析（如 Isat N/P、Vsat N/P 等）、SPICE 模型與後矽數據的相關性分析、主要均值偏移的識別，以及設計餘量的建立，旨在確保製程與設計的匹配性。
- **元件特性分析**：對基本元件庫單元進行深入定量分析，檢視其在不同切換速度與負載場景下的趨勢，並研究製程參數、通道長度、閾值電壓以及電壓-溫度條件對元件物理屬性的影響。
- **預防性設計指導**：在設計初期採用去耦電容整合策略，例如將去耦電容融入時鐘網路元件，或採用內建去耦電容的垂直堆疊 Multi-bit DFF 設計，以有效控制局部電流密度。實施分離測試時鐘與掃描重排序等策略，預先降低設計後期 IR-drop 問題風險。

- **配套支持措施**：在晶片內部集成監測 IP，用於後矽數據分析與監控，並制定完整的後矽測量策略。該策略涵蓋 SPICE 模擬、監測器整合、設計測試方法、數據收集與相關驗證流程，確保系統的穩定性與可靠性。

3.2. 客製化元件與時序策略 (Timing Signoff Strategy)

- **最佳晶片能效探索**：分析不同電壓、溫度及設計約束場景，進行時序校正以解決從 WAT 數據中觀察到的主要漂移，並尋找最有效的操作點。
- **設計分析**：優先考慮經常使用的關鍵元件，評估其對面積與功耗的影響。此外，檢視關鍵路徑，包括多位元暫存器 MBDF、反相器和加法器，在不同電壓場景下的表現，並減少低電壓下的信號波形失真。
- **標準元件庫優化**：針對影響繞線的關鍵元件（如 XOR/XNR 和 MUX）以及影響晶片可靠性與功耗的時鐘單元（CKINV/BUF）進行優化，確保驅動能力平衡。同時，提升常用元件的面積與功耗效率，減少不確定性，提升運行頻率並增強可靠性。
- **自訂單元優化**：基於元件庫與關鍵路徑評估進行單元定制，如多位元暫存器和多位元脈衝暫存器（MB Pulse-latch）。確保在不同電壓條件下信號平衡，避免持續時間（Hold-time）問題，並通過 SPICE Monte Carlo 分析評估最小電壓與設計餘量。
- **時序 Re-K**：根據實際 WAT 分佈調整時序 Re-K 和簽核策略，減少設計餘量並提升競爭力。通過物理參數分佈、RO 和 SPICE 相關性、晶片均勻性及回歸分析，推算等效電壓梯度和精確設計餘量。
- **預防性實體設計**：透過自訂電路設計，提升性能、面積利用率、功耗效率與可靠性，實現內建去耦合電容的時鐘驅動平衡 CKINV/CKBUF 和動態 IR-drop 預防等功能。

為降低局部電流密度過高的風險，如圖 Fig. 3-2 所示，將高切換率（Toggle-rate）與高電流密度的元件垂直佈局，並在周圍內置去耦合電容。這樣的設計可以有效避免熱點問題，因為這些熱點若未在設計階段妥善處理，後期即使工具檢測出來也難以解決。

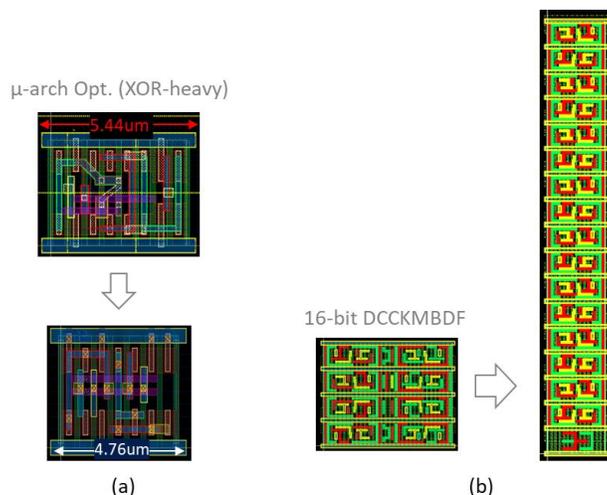


Fig. 3-2 μ-arch Optimization

針對演算法導致的繞線問題，採用自上而下（考慮排線與微架構調整）與自下而上（優化元件或組合元件 MEGA Cell [1]）雙重策略進行全面優化。例如，針對 SHA3 演算法中 XOR-heavy 部分以及交換網絡（Switch）機制的 MUX-heavy 部分進行優化，以提升佈局設計效率，如圖 Fig. 3-3。

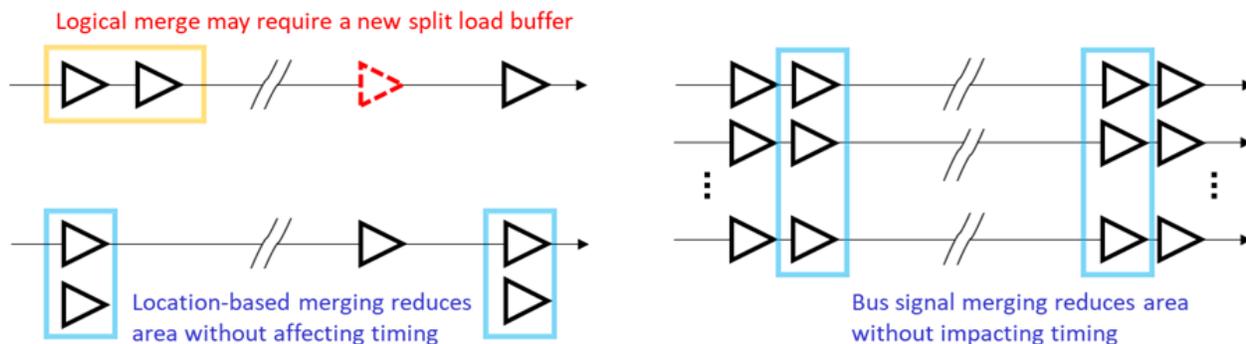


Fig. 3-3 MEGA Cell Exploration

3.3. 製程優化分析技術

- **WAT 與 CP/FT 映射及關聯分析**：分析 WAT 參數分佈（如 I_{sat} N/P、 V_{tl} N/P、 I_{off} N/P 等），結合 CP/FT 測試數據進行多維度關聯性分析，通過數據視覺化揭示製程缺陷並進行優化調整。
- **S2S 相關性與元件庫數據整合**：關聯製程與元件庫數據，對 SPICE 模型與 RO 設計的晶片測量結果進行相關性分析，確保製程與元件庫的一致性與準確性。

- **均勻一致性/OCV 分析**：進行晶圓均勻性回歸分析和晶片內變異分析 (OCV)，並結合 WAT 和 OCV 分析結果，細化時序餘量分佈，提升性能與可靠性。
- **製程配方優化與機器學習應用**：透過關聯與視覺化分析 WAT 和 CP 數據，如圖 Fig. 3-4，快速定位製程窗口與設計配方的最佳優化方向，並應用機器學習與神經網絡技術進行特徵建模與回歸分析，實現智能化生產與設計優化。

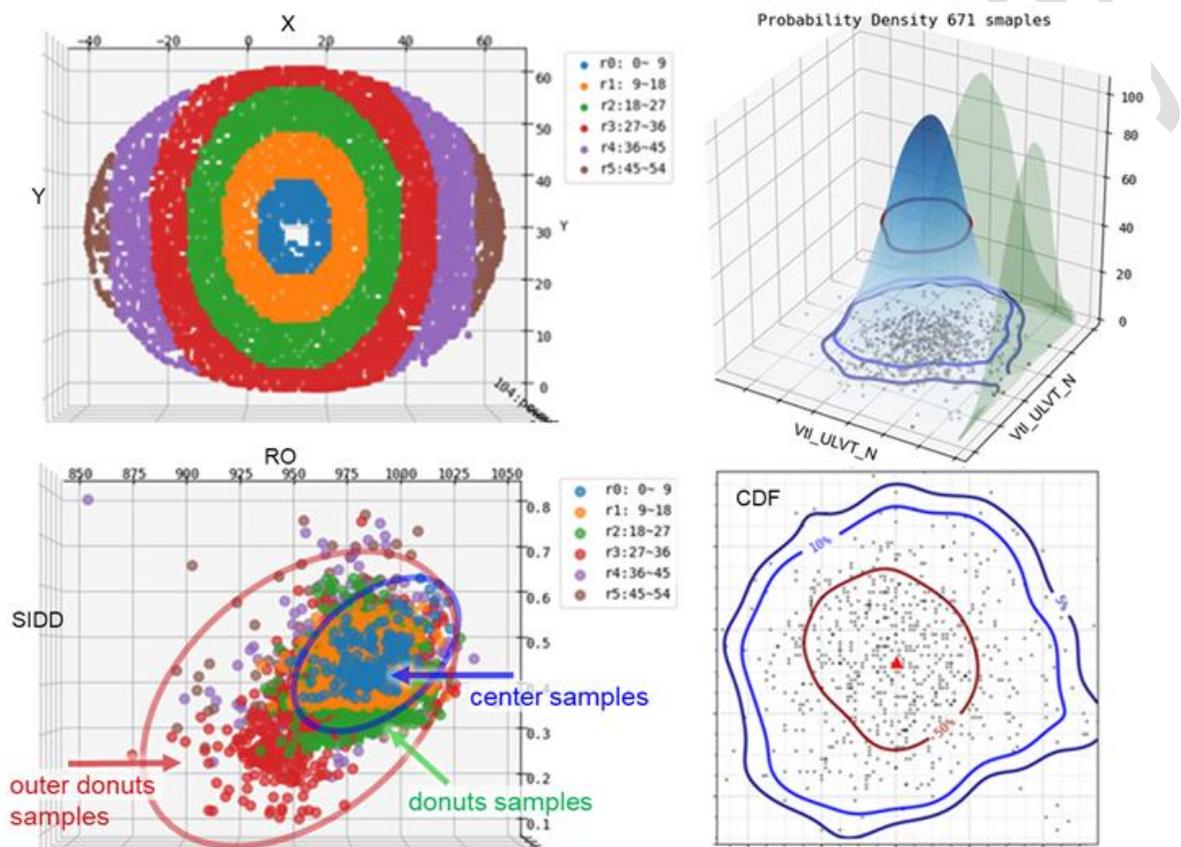


Fig. 3-4 Data Visualization and Feature Correlations

3.4. 補償機制的設計與實施

- **晶片性能評等與分箱策略**：利用機器學習或神經網絡技術，自動化晶片效能評等與分箱策略，預測晶片品質並優化晶片、晶圓及批次排名，以提高產量與生產效率。
- **老化與電壓補償策略**：制定針對晶片老化和電壓變化的補償策略，延長晶片壽命並增強整體可靠性。

- **動態餘量警報與調整策略**：實施動態餘量警報系統 (Dynamic Slack Alert)，根據即時餘量變化動態調整電壓和時鐘策略 (如時鐘頻率、門控、伸縮)，實現即時優化和故障預警，提升性能和抗噪聲能力。
- **IR/變異容忍解決方案**：開發 IR 壓降與製程變異管理策略，包括局部增益、瞬態響應、動態壓降管理及容錯技術，以提升系統穩定性，並有效適應 PVT (製程、電壓、溫度) 變化。

Fig. 3-5 顯示晶圓層級的 RO 頻率分佈，其中 RO 梯度的不均勻性 (Local Variation) 無法透過電壓補償消除。這促使我們重新評估 STA 方法中的 AOCV 或 LVF 是否足夠完善，以及如何更有效納入實際效能梯度。為應對這些挑戰，業界也示範性提出如 LDO Array 區域補償技術及多小晶片 (Chiplet) 封裝替代單一大晶片等策略，以解決製程相關問題。

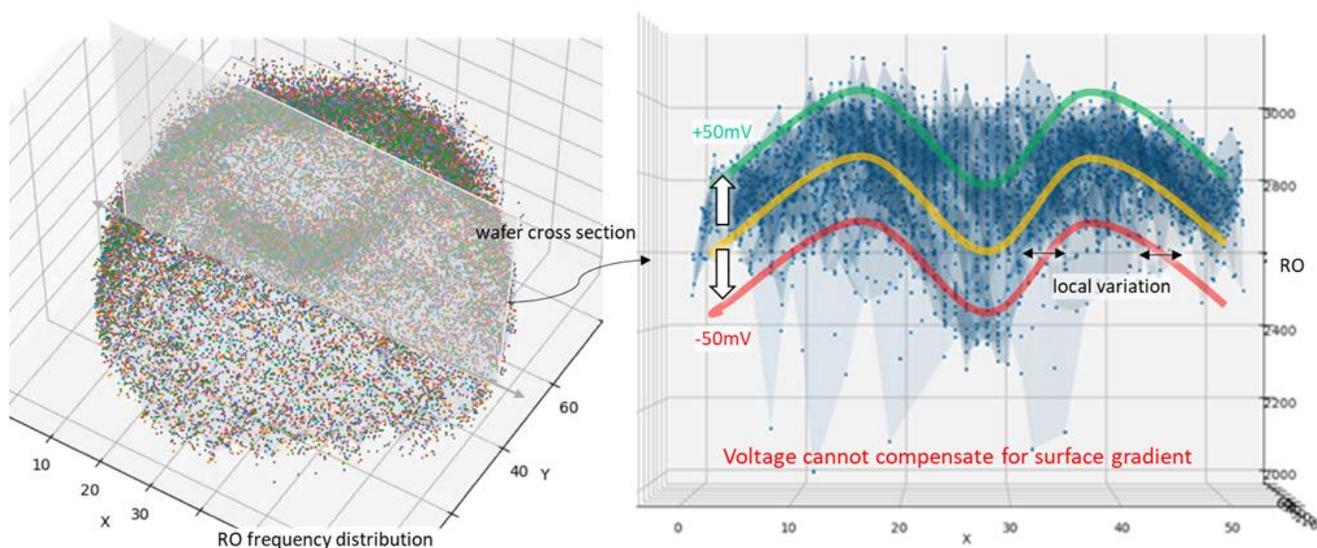


Fig. 3-5 RO Gradient Uniformity

如圖 Fig. 3-6 所示，晶圓的南北半球與內外圈效能梯度可能因製程控制差異而顯著不同。因此，晶片量產與封裝測試時，關鍵路徑往往非由特定元件組成，可能導致 DFT 的 Small-delay Fault 或 MBIST 等性能評等參數與 Fmax 相關性不佳。此外，受晶圓廠製程控制影響，工具與算法選擇的敏感關鍵路徑分屬不同群體，使問題更為複雜。

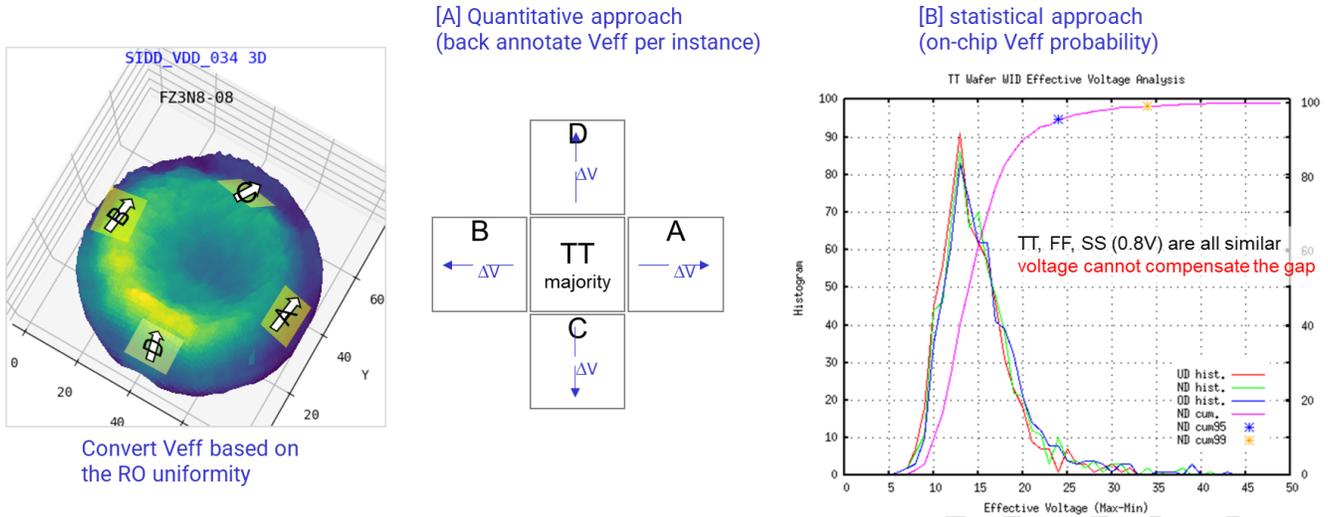


Fig. 3-6 Uniformity and Effective IR-drop

3.5. 近臨界電壓技術的需求與挑戰

在近臨界電壓操作下 (<300mV)，所謂的「逆溫異常現象」已成為常態，如圖 Fig. 3-7 所示。傳統的設計流程中，通常會先取得晶圓廠提供的 SPICE 模型及 Liberty 數據庫，並基於這些數據建立時序簽核的防守邊界。然而，重要的製程參數，如 WAT Isat 或 VtI 等，往往要等到量產驗收前才能從晶圓廠獲得反饋，且這些參數的偏移量實際上反映了邏輯設計端的设计餘量不足，這也突顯了現有設計流程在科學性和實用性上的不足。

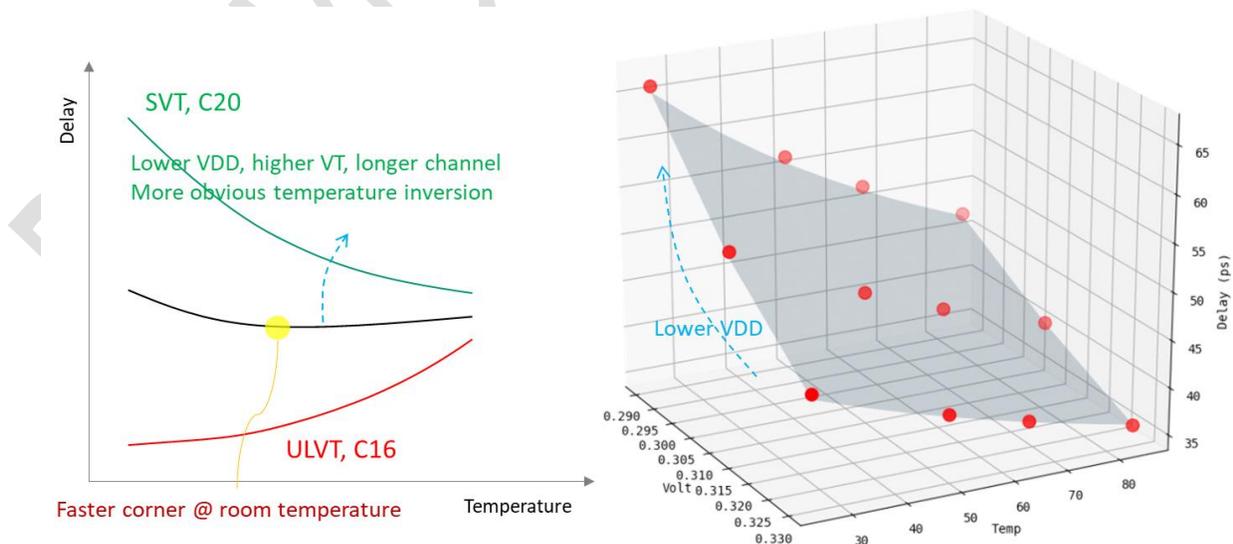


Fig. 3-7 Temperature Inversion

早期晶片的多核心設計中，Package RLC 是電源佈局與分析的關鍵考量之一，因多核平行運算時總電流增加，導致大電流經過元件時放大電壓降噪聲 (Ldi/dt)。為解決此問題，「串接電壓源 (Voltage-Stacking)」設計應運而生。

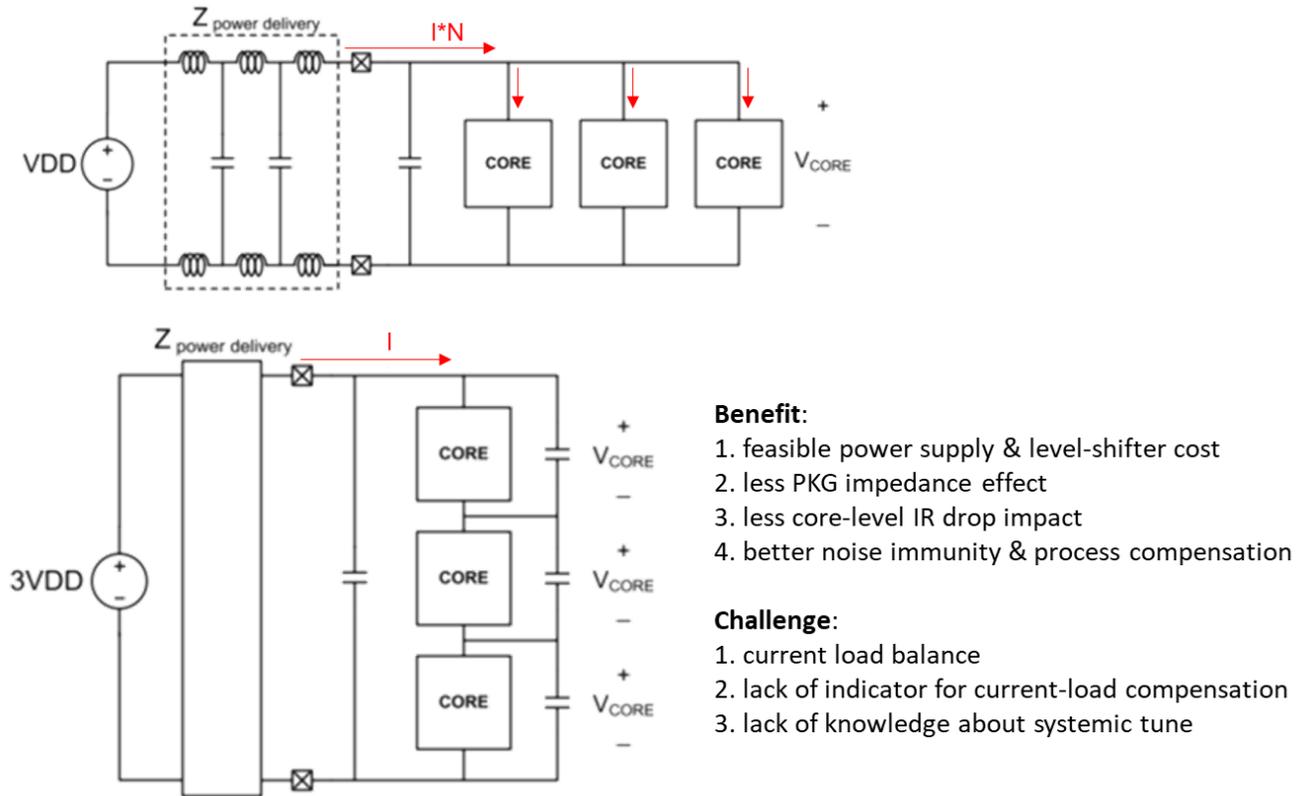


Fig. 3-8 Voltage Stacking

如圖 Fig. 3-8 所示，將晶片串聯並將 V_{min} 降至最低（如低於 300mV）。由於串接結構電流固定，更多晶片能將噪聲分攤，形成具自我補償特性的「電流動態平衡」系統。然而，熱效應或製程偏移引發的漏電流可能使效能下降的晶片分到更低的跨壓，並且非晶片運作產生的電流會對系統造成反效果，使傳統低電壓調節技術（如 CG/DVFS/AVS）不再直覺。這需要針對整體系統進行調整，而非單一晶片，因此是一個值得深入研究的方向。

第二部分：DTCO.ML™ - 機器學習驅動的半導體流程優化

第4章 機器學習與 DTCO 的融合 (DTCO.ML™)

晶片設計與製造深度依賴產業鏈的協同合作。由於傳統設計方法難以全面掌握元件物理與系統行為，時序簽核常導致過大的防守邊界，進一步增加設計餘量控制的挑戰。隨著數據科學的引入，晶片設計流程變得更加科學且合理。如圖 Fig. 4-1 所示，創新設計方法結合機器學習、IP、EDA 工具與設計流程，形成高效的整體解決方案，顯著提升設計效率與生產力。

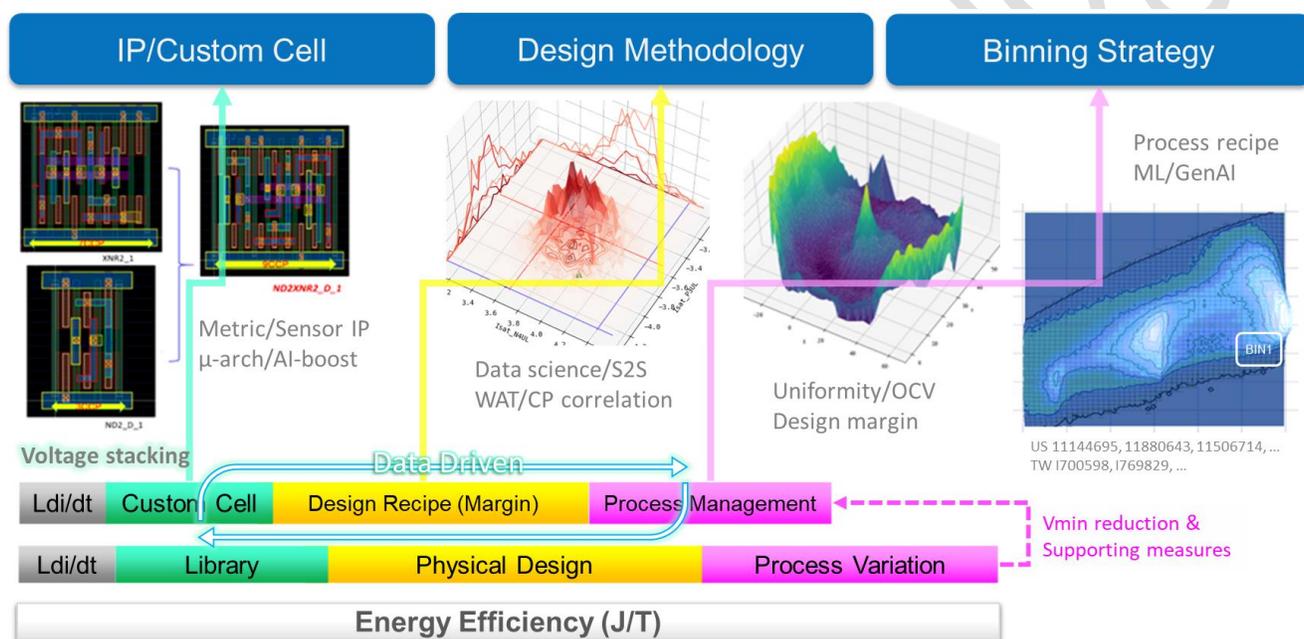


Fig. 4-1 Revolutionizing Efficiency & Productivity through DTCO.ML

- IP 方案與定制單元：**提供晶片內監測、動態電壓與溫度感測、餘量警報及專用計算單元，優化能效、面積與可靠性，並支持異步、自時鐘、PL/PG 等多種時鐘方案。定制標準單元針對時鐘網絡與關鍵路徑元件設計，實現高性能、低功耗與面積效率，並包含動態 IR 預防、微架構 (μ-arch) 和 MEGA 元件優化，以減少繞線問題並提升多邏輯效率。
- EDA 與設計流程開發：**結合機器學習與 AI 開發創新 EDA 工具與設計流程，解決傳統跨領域知識與跨資料維度之間交互分析的不足，從而實現更精確的設計迭代和優化。

- **分箱補償策略**：提供創新工具進行全面的特徵萃取與元件庫分析，捕捉高維度趨勢和見解，並利用 DTCO 機器學習平台協同優化設計與製程，實施分箱與補償策略以提高晶片能效與生產力。

4.1. 虛擬晶圓數據建模(Virtual Silicon)

為清楚說明，我們透過數據觀測、轉換與標準化，結合隨機過程建模，手動構建虛擬晶圓數據。首先，建立虛擬晶圓座標，並僅保留位於晶圓中心半徑內的數據。Example 4-1 示範了虛擬晶圓座標的生成，晶圓尺寸為 65×65 個樣本，如圖 Fig. 4-2 所示。

Example 4-1 Virtual Silicon: Gaussian-Volcano Wafer Data

```
# Virtual Silicon : Gaussian-Volcano Wafer Data
import matplotlib.pyplot as plt
import numpy as np

def genWaferXY(w=65, h=65):
    ix, iy = np.linspace(1, w, w), np.linspace(1, h, h)
    gx, gy = np.meshgrid(ix, iy)
    cx, cy = np.ceil(w/2).astype(int), np.ceil(h/2).astype(int)
    cr = min(cx, cy)
    r = np.sqrt((gx-cx)**2 + (gy-cy)**2)
    mask = (r <= cr)
    x, y, r = gx[mask], gy[mask], r[mask]
    if False:
        plt.figure(figsize=(6,6))
        plt.scatter(x, y, s=15, alpha=0.4, marker='s')
        plt.scatter(cx, cy, s=15, c='r', marker='s')
        plt.xlabel('X')
        plt.ylabel('Y')
        plt.tight_layout()
    return x, y, r

def volcanoSin(r):
    theta = np.interp(r, [1,65], [-np.pi, np.pi])
    z = np.sin(theta - np.pi/2) - 1.5 * np.sin(theta/2 - np.pi/2)
    return np.interp(z, [z.min(), z.max()], [0, 1])

def volcanoGaussian(r, h1=1, s1=15, h2=1, s2=5):
    g1 = h1 * np.exp(-r**2/(2*s1**2))
    g2 = h2 * np.exp(-r**2/(2*s2**2))
    z = g1-0.5*g2
    return np.interp(z, [z.min(), z.max()], [0, 1])

#%% gendataset
x, y, r = genWaferXY()
z1 = volcanoSin(r)
z2 = volcanoGaussian(r)
```

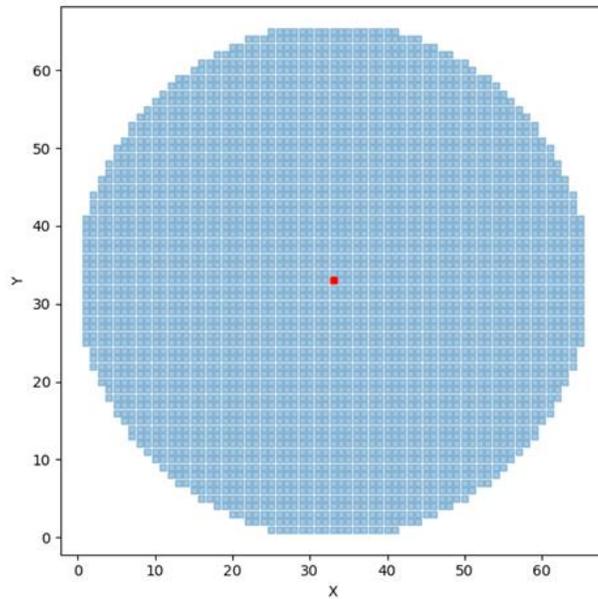


Fig. 4-2 Virtual Wafer Data Mapping

不同晶圓廠的 Uniformity 各具特色，例如某些製程的 RO 或 SIDD 數據在 XY 座標上呈現類似火山錐的系統性缺陷 (systematic defect)。我們可透過疊加不同振幅的 Sin 或 Gaussian 函數來模擬此結構。首先，利用晶圓座標，分別透過 Sin 與 Gaussian 函數生成火山錐結構 (z0)，其機率密度分佈通常呈現 U 型，類似雙曲線正割 (Cosh) 分佈。Sin 與 Gaussian 生成的火山曲面相近，但 Gaussian 產生的機率密度分佈較為平滑，如圖 Fig. 4-3 所示。

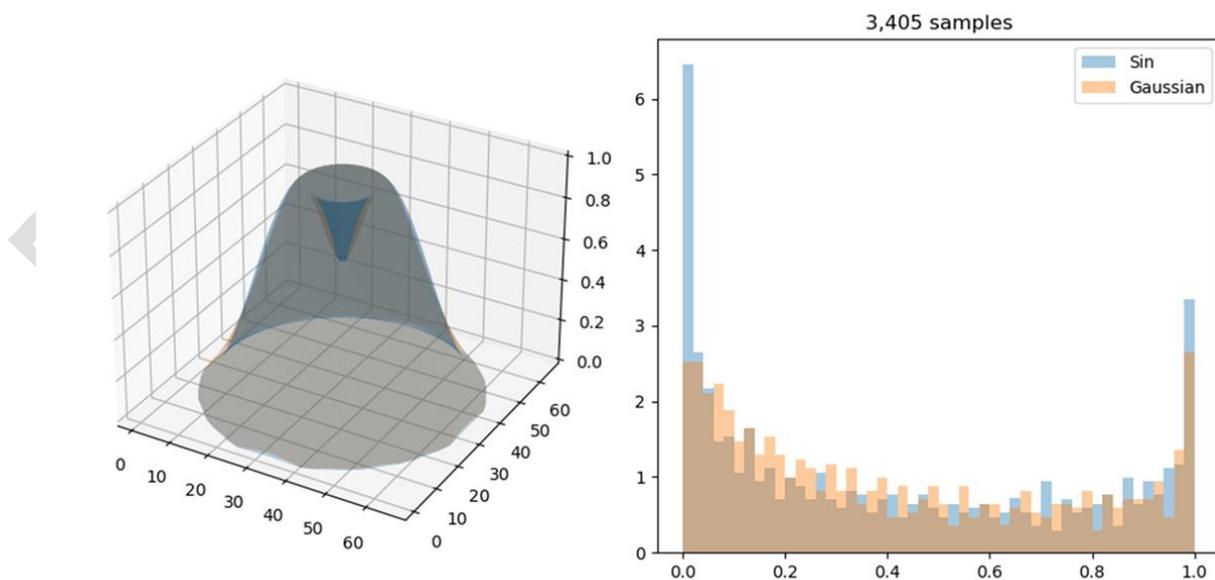


Fig. 4-3 Volcano Surface Distribution and Probability Density

接下來，Example 4-2 基於此曲面生成第一組虛擬晶圓數據 z_1 （例如 RO）。首先，對平滑的 Gaussian-Volcano 曲面 z_0 進行正態分佈變換 ($\epsilon_0 \sim N(0,1)$)，再直接疊加高斯雜訊 ($\epsilon_1 \sim N(0,1)$)，然後轉回原始 z_0 的尺度。為了數據視覺化並比較其效果，我們利用 `np.interp(x, [原分佈範圍], [目標分佈範圍])` 技巧將 z_1 調整至 $[0,1]$ 之間。加入高斯雜訊後 (z_1)，其機率密度呈現偏態高斯分佈，可根據需求調整均值與變異性，以符合實際觀測數據（如 RO），如圖 Fig. 4-4 所示。

Example 4-2 RO Silicon Data Modeling and Standardization

```

%% generate an RO surface with Gaussian noise
x, y, r = genWaferXY()
z0 = volcanoGaussian(r)
e0 = 1/z0.std()*(z0-z0.mean()) # z0 to ~N(0,1)

# combine with noise
e1 = np.random.normal(0, 1, len(z0)) # noise ~N(0,1)
z1 = (e0 + 0.2*e1) # ~N(0,1)
z1 = z0.std()*z1 + z0.mean() # back to z0 scale
z1 = np.interp(z1, [z1.min(), z1.max()], [0, 1]) # to [0, 1] for comparison

%% conditional Gaussian standarization
# generate SIDD with a correlation of 0.9 with z1
rho = 0.9
e1 = 1/z1.std()*(z1-z1.mean()) # z1 to ~N(0,1)
e2 = np.random.normal(0, 1, len(z1)) # noise ~N(0,1)
z2 = rho*e1 + np.sqrt(1-rho**2)*e2 # ~N(0,1)

```

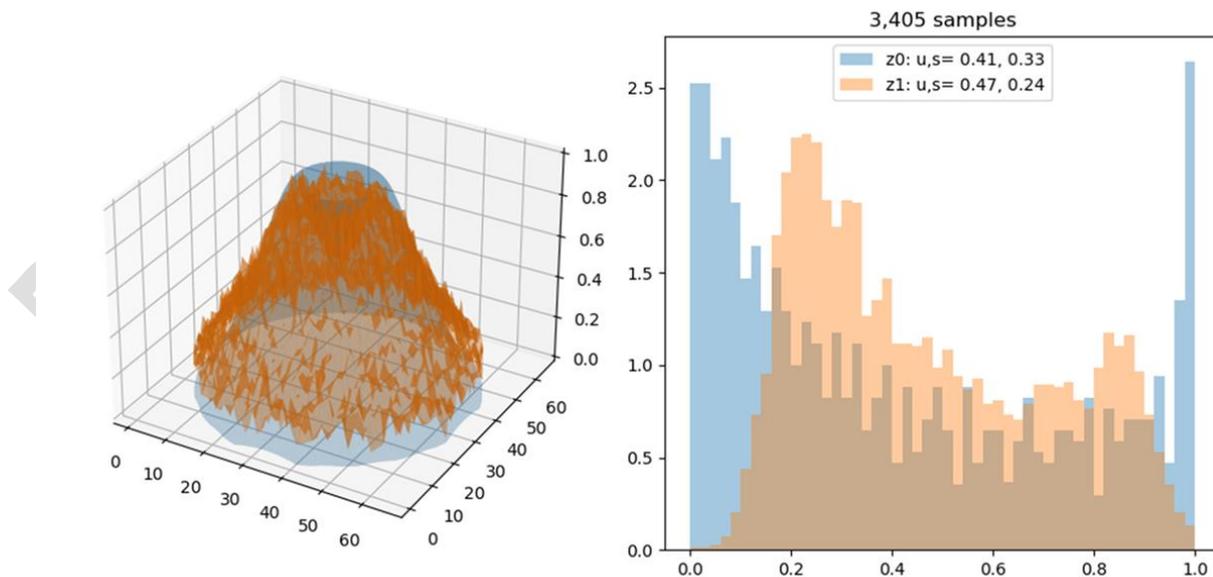


Fig. 4-4 Volcano Surface with Gaussian Noise

接下來，Example 4-3 利用條件高斯 (Conditional Gaussian) 分佈的技巧生成第二組數據 z_2 (例如 SIDD)。根據觀察，RO 與 SIDD 之間具有較高的正相關 ($\rho=0.9$)，因此我們希望 z_2 與 z_1 之間維持相同的相關性。首先，我們將 z_1 標準化為正態分佈 ($\epsilon_1 \sim N(0,1)$)，然後透過公式 $\rho \cdot \epsilon_1 + \sqrt{1-\rho^2} \cdot \epsilon_2$ 生成第二組數據，其中 ϵ_2 是遵守 $N(0,1)$ 的隨機雜訊。

Example 4-3 SIDD Modeling and Standardization

```
def logNormScale(u, s): # normal to log-normal
    log_u = np.log(u**2 / np.sqrt(s**2 + u**2))
    log_s = np.sqrt(np.log(1 + s**2 / u**2))
    return log_u, log_s

mu, sigma = 0.5, 0.1 # target mu and sigma of SIDD
log_u, log_s = logNormScale(mu, sigma) # retarget to log-normal scale
z2 = np.exp(log_s*z2 + log_u) # to z2 scale
z2 = np.interp(z2, [z2.min(), z2.max()], [0, 1]) # to [0, 1] scale

plt.figure(figsize=(10,9))
ax1 = plt.subplot(221, projection='3d', title='z1')
ax1.plot_trisurf(x, y, z0, alpha=0.3)
ax1.plot_trisurf(x, y, z1, alpha=0.4)
ax2 = plt.subplot(222, projection='3d', title='z2')
ax2.plot_trisurf(x, y, z0, alpha=0.3)
ax2.plot_trisurf(x, y, z2, alpha=0.4)
ax3 = plt.subplot(223, title=f'{len(z1):,} samples')
ax3.hist(z0, bins=50, density=True, alpha=0.4, label=f'z0: u,s= {z0.mean():.2f}, {z0.std():.2f}')
ax3.hist(z1, bins=50, density=True, alpha=0.4, label=f'z1: u,s= {z1.mean():.2f}, {z1.std():.2f}')
ax3.hist(z2, bins=50, density=True, alpha=0.4, label=f'z2: u,s= {z2.mean():.2f}, {z2.std():.2f}')
ax3.legend()
ax4 = plt.subplot(224, title=f'{len(z1):,} samples')
ax4.scatter(z1, z2, alpha=0.4)
ax4.set_xlabel('z1: RO')
ax4.set_ylabel('z2: SIDD')
plt.tight_layout()
```

儘管 RO 與 SIDD 之間存在高度正相關，但實際觀察到 SIDD 的機率密度分佈呈現明顯的對數正態 (log-normal) 趨勢。因此，我們重新計算符合對數正態分佈的均值和標準差，並對原分佈進行指數變換。這個方法也可以應用於其中一個維度數據缺失的情況，例如當 x 、 y 和 RO 數據完整，而 WAT 只有稀疏的數據 (例如僅有幾十筆資料)。然而，通過對數據的觀察，我們仍然能夠根據可觀察範圍內的可信資料邊界、統計上的均值與變異性，以及高維資料之間的相關性等信息來生成預測。

如圖 Fig. 4-5 所示，我們獲得兩個維度的虛擬晶圓數據 z_1 (RO) 與 z_2 (SIDD)，其相關性為 $\rho=0.9$ ，且特徵向量在晶圓座標系中保持整體結構趨勢。RO 的機率密度分佈呈現偏態正態 (skew-normal)，而 SIDD 則符合對數正態 (log-normal) 分佈。

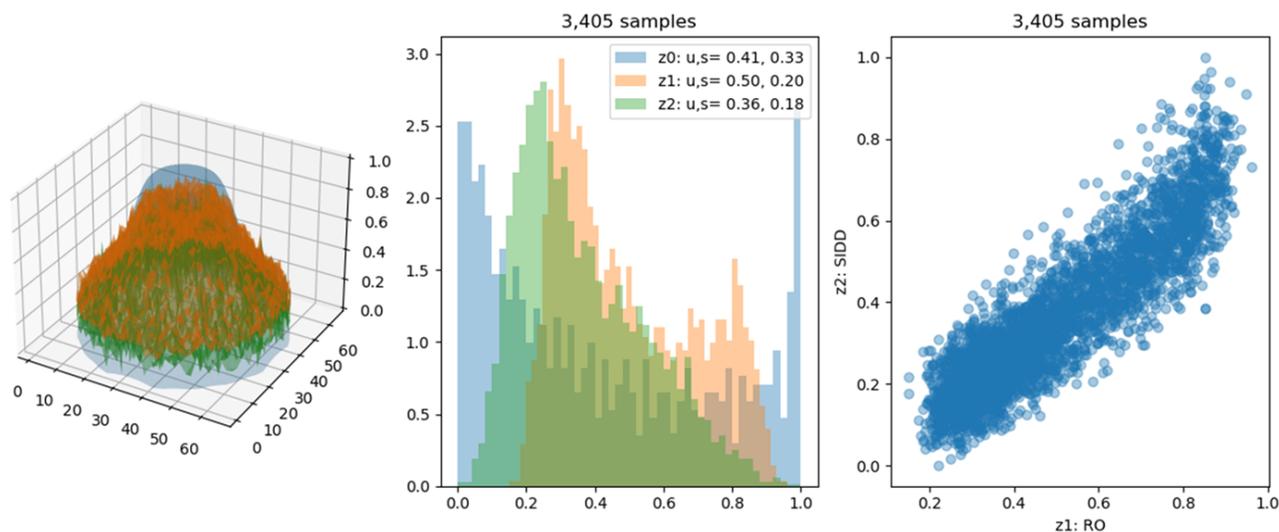


Fig. 4-5 Log-Normal Conditional Gaussian with Proper Covariance

然而，真實的矽晶數據往往不僅僅符合單一偏斜正態或對數正態分佈，特徵向量之間可能存在更複雜的非線性關係，並伴隨 **uniformity** 這類巨型曲面結構。為了有效處理這類傳統機器學習難以建模的數據分佈，後續章節將介紹如何運用生成式 AI 進行更精確的建模與分析。

4.2. 回歸模型的構建與推論

在實體設計工程中，估算時序餘量與功耗通常需要大量時間。以漏電流評估為例，傳統方法依賴 LIB CAD 工程師重建大量 PVT corner 的庫 (Re-K)，並使用 EDA 工具進行分析與報告。然而，對於庫中未定義的 corner，評估結果的可靠性大幅降低，甚至無法得出結論。

為了有效解決這一問題，如圖 Fig. 4-6 (a) 所示，我們可以對晶片運行的主要電壓和頻率點進行 SPICE 仿真與特徵化 (Characterization)，這些藍色點代表離散且稀疏的數據。通過回歸分析，我們能夠輕鬆構建出完美擬合這些數據點的二次回歸曲面 (橙色曲線)，即 $leak = f(V, T)$ 。這使得機器學習能夠高效且可靠地預測在庫未覆蓋條件 (如 0.32V, 70°C) 下的參數值與趨勢。

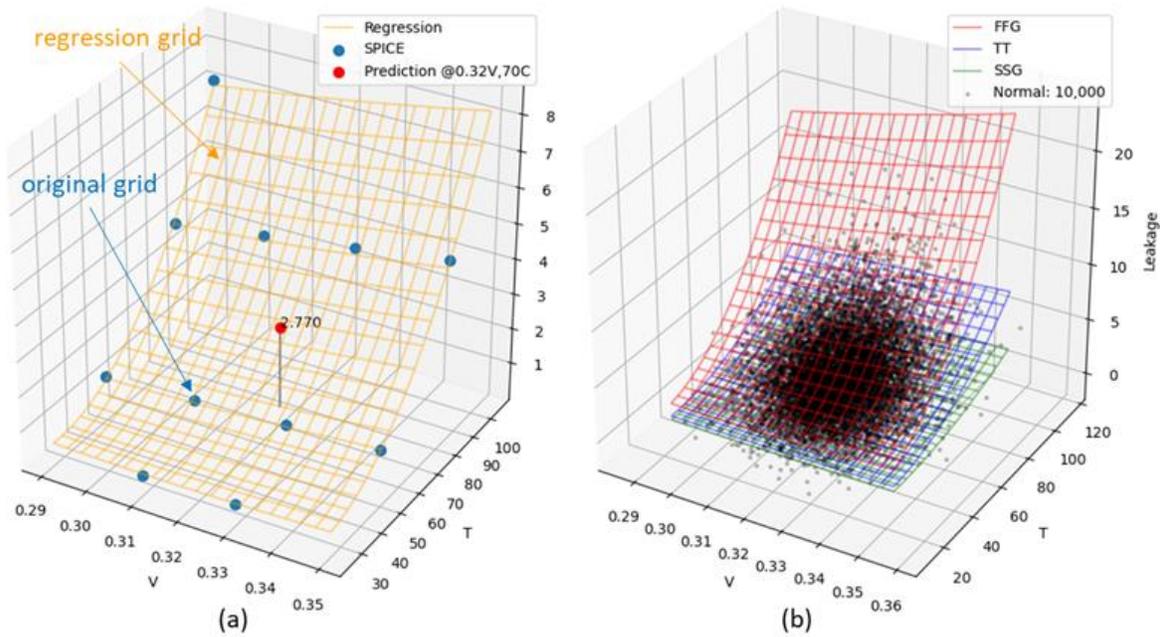


Fig. 4-6 Precise Leakage Prediction

Example 4-4 模擬了 CKINV 在 TT 製程下的 SPICE 仿真數據，涵蓋 11 個離散的電壓與溫度條件。通過最小二乘回歸 (Least Squares Regression, LSR)，我們可精確預測原始 PVT 條件網格中未包含的值，例如 0.32V、70°C 下的 leakage。

Example 4-4 Least Squares Regression

```
# Leakage model regression and prediction.
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import scipy.linalg

# CLKINV1_C10UL @TT
tt = pd.DataFrame(
    [[ 0.29, 50, 1.1656 ], [ 0.29, 85, 3.8962 ], [ 0.29, 105, 7.0997 ],
     [ 0.31, 25, 0.4655 ], [ 0.31, 50, 1.2776 ], [ 0.31, 85, 4.2603 ],
     [ 0.33, 25, 0.5086 ], [ 0.33, 50, 1.3934 ], [ 0.33, 85, 4.6352 ],
     [ 0.35, 50, 1.5129 ], [ 0.35, 85, 5.0208 ]], columns=['V', 'T', 'leak'])

#%% LS Regression
x, y, z = tt.values.T
X = np.array([np.ones_like(x), x, x**2, y, y**2, x*y]).T # training set
C, _, _ = scipy.linalg.lstsq(X, z) # LS regression coefficient

# regression grid
ix, iy = np.linspace(x.min(), x.max(), 20), np.linspace(y.min(), y.max(), 20)
gx, gy = np.meshgrid(ix, iy)
tx, ty = map(np.ravel, (gx, gy))
T = np.array([np.ones_like(tx), tx, tx**2, ty, ty**2, tx*ty]).T
tz = np.dot(T, C)
```

```

# predict condition 0.32V, 70°C not present in the SPICE metric
v, t = 0.32, 70
p = np.dot(np.array([1, v, v**2, t, t**2, v*t]), C)

plt.figure(figsize=(6,6))
ax = plt.subplot(projection='3d')
ax.plot_wireframe(gx, gy, tz.reshape(gx.shape), alpha=0.4, colors='orange', label='Regression')
ax.scatter(x, y, z, s=50, alpha=1, label='SPICE')
ax.scatter(v, t, p, s=50, c='r', alpha=1, label=f'Prediction @{v}V,{t}C')
ax.plot([v,v], [t,t], [0,p], c='gray')
ax.text(v, t, p, f'{p:.3f}')
ax.set_xlabel('V')
ax.set_ylabel('T')
ax.set_zlabel('Leakage')
ax.set_box_aspect((1,1,1), zoom=1.0)
plt.legend()
plt.tight_layout(rect=(-0.1,0.05,1,1))
    
```

此外，我們還可以基於各製程 corner 模型生成大量密集網格數據，並將其提升為三維模型，構建 $leak=f(V,T,P)$ 。通過高斯分佈的概率雲，我們能快速掌握製程偏快（如 $FF +1\sigma$ 或 $+1.5\sigma$ ）的趨勢，從而全面分析並優化設計參數，如圖 Fig. 4-6(b)所示。

延續前述範例，Example 4-5 將 FFG 與 SSG 的離散點構建為 LS 曲面，並生成密集網格的訓練集。假設 FFG 至 SSG 的範圍為 6σ ，並將模型的輸入參數擴展為三維，即 $f(V,T,P)$ ，以構建 leakage 模型。接著，基於高斯正態分佈，以 (V,T,P) 的均值(0.32V,65°C,TT) 為中心，生成 10K 隨機樣本，預測該模型下的完整 leakage 分佈。

Example 4-5 Leakage Modeling and Prediction

```

# model fitting: leak = f(V,T,P)
ss = pd.DataFrame(
    [[ 0.31, 25, 0.1249 ], [ 0.31, 50, 0.3709 ], [ 0.31, 85, 1.3588 ],
     [ 0.31, 105, 2.5932 ], [ 0.33, 25, 0.1361 ], [ 0.33, 50, 0.4033 ],
     [ 0.33, 85, 1.4742 ], [ 0.35, 50, 0.4365 ], [ 0.35, 85, 1.5924 ],
     [ 0.35, 105, 3.0324 ]], columns=['V','T','leak'])

ff = pd.DataFrame(
    [[ 0.29, 50, 3.889 ], [ 0.29, 85, 11.802 ],
     [ 0.31, 25, 1.687 ], [ 0.31, 50, 4.276 ], [ 0.31, 85, 12.939 ],
     [ 0.33, 25, 1.849 ], [ 0.33, 85, 14.115 ]], columns=['V','T','leak'])

# model fitting for each process corner
gnum = 20
px,py = np.linspace(0.29,0.35,gnum),np.linspace(25,105,gnum)
gx,gy = np.meshgrid(px,py)
tx,ty = gx.ravel(),gy.ravel()
tzL = []
T = np.array([np.ones_like(tx), tx, tx**2, ty, ty**2, tx*ty]).T
    
```

```

for d in [tt, ff, ss]:
    x,y,z = d.values.T
    X = np.array([np.ones_like(x), x, x**2, y, y**2, x*y]).T
    C,_,_ = scipy.linalg.lstsq(X,z) # LS regression
    tzL += [np.dot(T,C)]

ptt,pff,pss = tzL

# model fitting for all corners
size = gnum**2
a,b,c = np.array(list(tx)*3),np.array(list(ty)*3),np.array([3]*size+[0]*size+[-3]*size)
z = np.array(list(pff)+list(ptt)+list(pss))
X = np.array([np.ones_like(z), a, a**2, b, b**2, c, c**2, a*b, a*c, b*c]).T
C,_,_ = scipy.linalg.lstsq(X,z) # LS regression, as f(V,T,P)

# Generate 10K random samples from a Gaussian normal distribution
size = 10000
V = np.random.normal(0.32, (0.35-0.29)/6, size)
T = np.random.normal(65, (105-25)/6, size)
P = np.random.normal(0, 1, size)
a,b,c = V,T,P

X = np.array([np.ones_like(a), a, a**2, b, b**2, c, c**2, a*b, a*c, b*c]).T
p = np.dot(X,C) # batch prediction

# predict condition 0.32V, 70°C, TT+1.5σ not present in the SPICE metric
#pa,pb,pc = 0.32, 70, 1.5
#pp = np.dot(np.array([1, pa, pa**2, pb, pb**2, pc, pc**2, pa*pb, pa*pc, pb*pc]).T, C)

plt.figure(figsize=(6,6))
ax = plt.subplot(projection='3d')
ax.plot_wireframe(gx, gy, pff.reshape(gx.shape), colors='r', alpha=0.4, label='FFG')
ax.plot_wireframe(gx, gy, ptt.reshape(gx.shape), colors='b', alpha=0.4, label='TT')
ax.plot_wireframe(gx, gy, pss.reshape(gx.shape), colors='g', alpha=0.4, label='SSG')
ax.scatter(a, b, p, s=5, alpha=0.2, c='k', label=f'Normal: {size:,}')
ax.set_xlabel('V')
ax.set_ylabel('T')
ax.set_zlabel('Leakage')
ax.set_box_aspect((1,1,1), zoom=1.1)
plt.legend()
plt.tight_layout(rect=(-0.1,0.05,1,1))
    
```

4.3. 數據追蹤與產能優化的應用

如 Fig. 4-7 所示，利用特徵迴歸分析預測晶片品質，實現晶片、晶圓與批次的分級排序。藉助跨維度建模與大量數據分析，制定精確的分類策略與良率評估方案，同時提供優化設計配方，顯著提升生產交付效率與流程可靠性。這一數據驅動方法突破傳統設計局限，使晶片設計更具科學性，並有效解決時序簽核防守邊界過大導致的設計餘量挑戰。

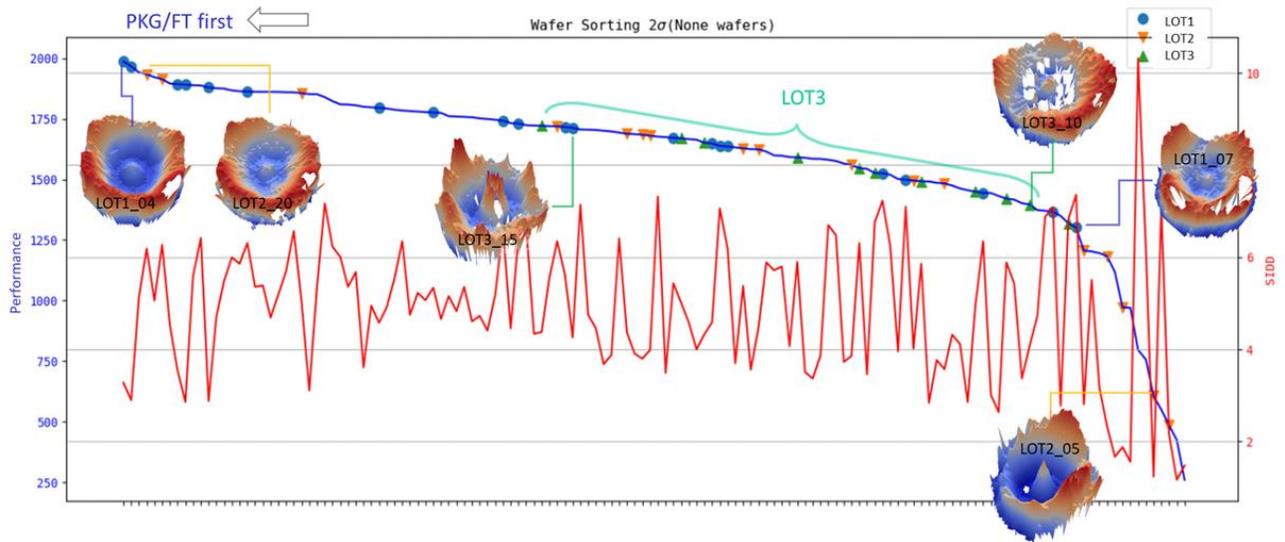


Fig. 4-7 Product Shipment Optimization

結合硬體偵測電路與數據分析，運用機器學習與回歸模型，推導實體設計餘量與時序簽核的最佳建議。基於晶片生產中的測試數據（WAT/CP/FT/SLT），將晶圓廠語言（如 WAT Isat、Vtl）與晶片設計語言（如 CP Performance、Leakage）映射到高維特徵（如良率或能效），確保設計與製造方向的精準與一致性，如圖 Fig. 4-8。

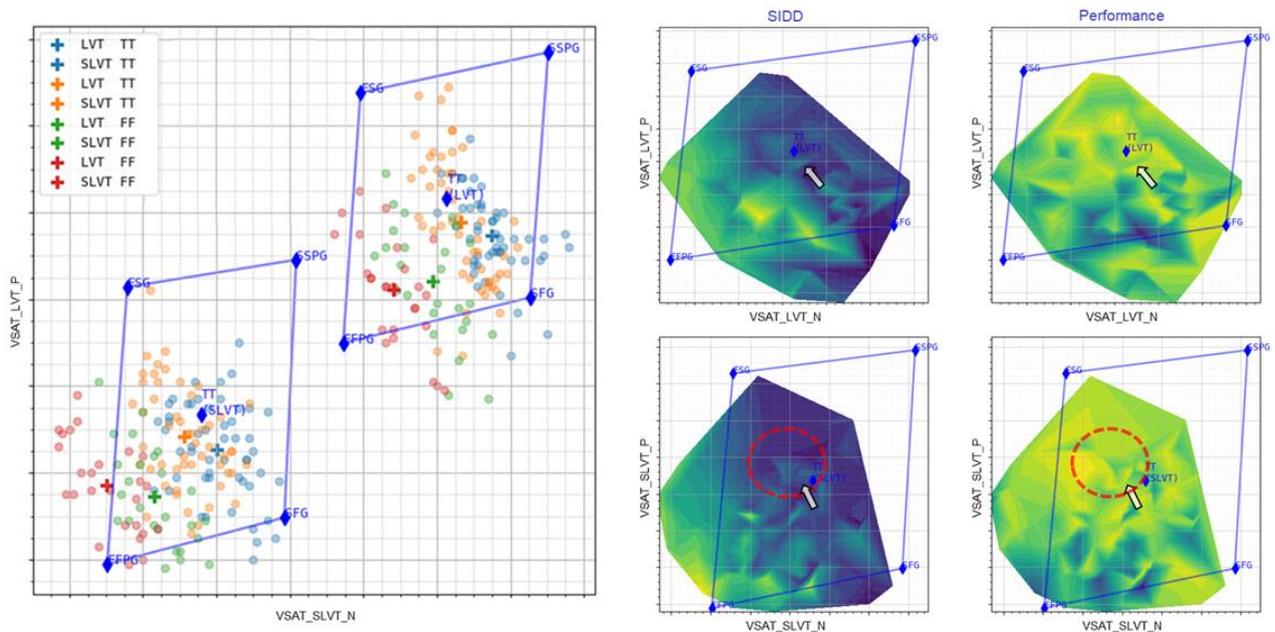


Fig. 4-8 Process Window Aligned with SPICE Target

第5章 元件庫特徵萃取與分析系統 (Library Metric Extraction, libMetric™)

元件基準評比 (Benchmarking) 是一種量化元件性能的有效方法，通過評估驅動能力、延遲、功耗等關鍵指標，為設計優化提供依據。然而，一個優秀的設計並非單純依賴高驅動力的元件，就像一部超大馬力的超級跑車並不適合行駛在坡度不平、彎道連續的山路，如圖 Fig. 5-1 所示。設計過程中，除了駕駛 (designer) 的操作精妙外，還需綜合考量引擎驅動力 (cell driving)、變速箱 (Vth)、底盤 (architecture)、架構 (layout) 等多方面因素的平衡與配合。

透過特徵萃取 (Metric Extraction) 與趨勢分析，可以深入揭示元件參數與實體設計工具或演算法之間的配合不足，這正是數據分析的核心價值，也是開發高效元件庫的重要方向。元件庫在不同 Vth 或通道長度 (Channel-length) 之間可能存在較大的性能間隙，而傳統 EDA 工具在優化過程中，若無法滿足對應的時序需求，往往只能通過犧牲面積 (up-sizing) 或降低功耗效率 (更換 Vth 或通道長度) 來進行折衷，這種妥協可能影響設計的整體性能與效能目標。

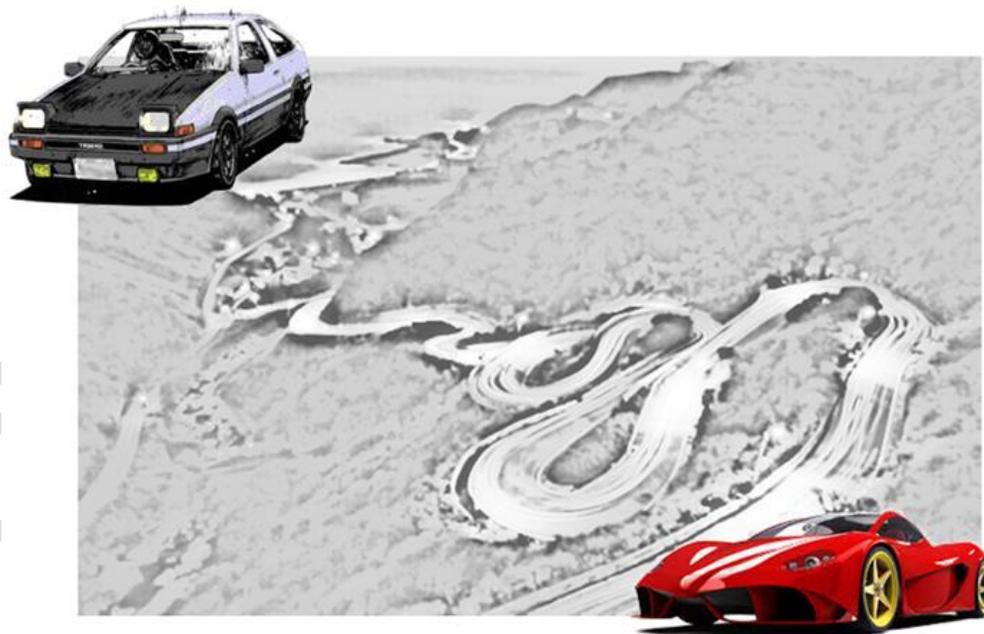


Fig. 5-1 設計效能的平衡之道：像跑車駕馭山路般精妙

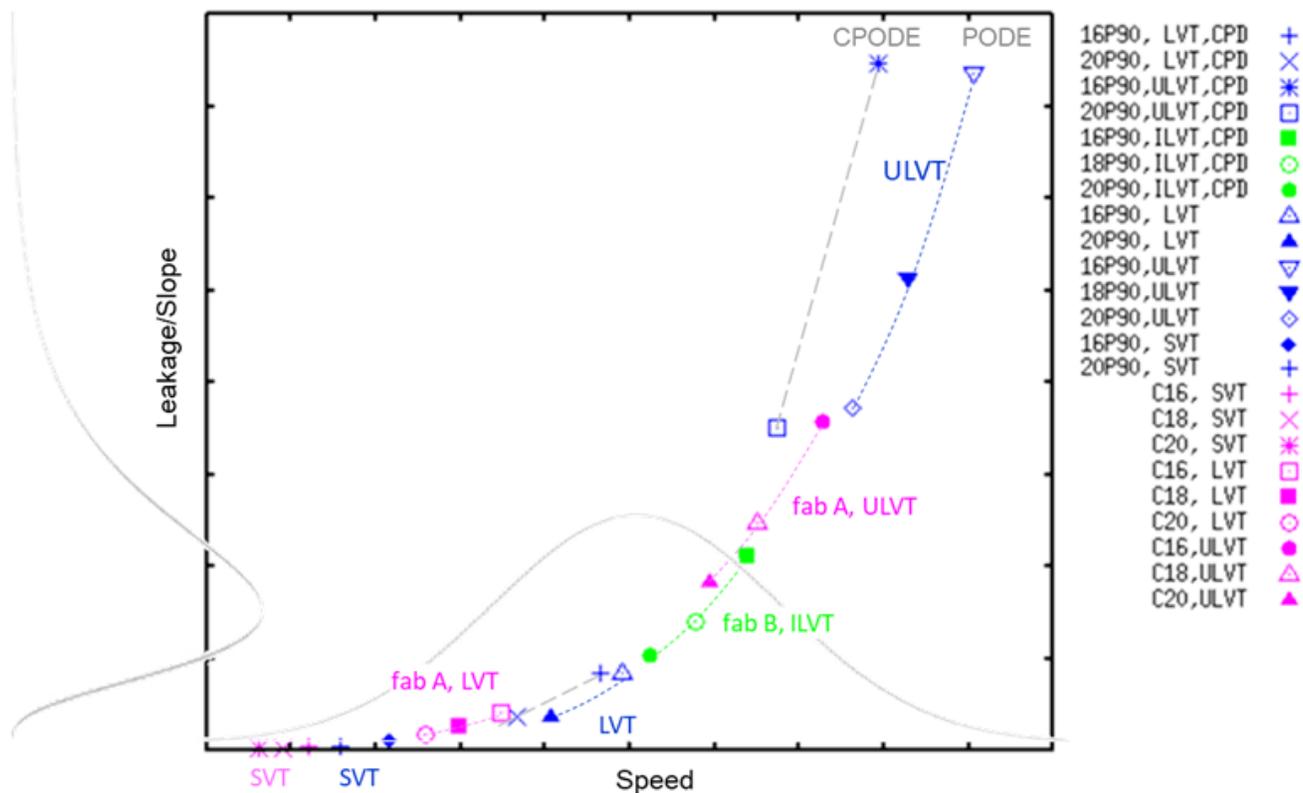


Fig. 5-2 Cell Speed and Leakage Evaluation

如圖 Fig. 5-2 所示，透過資料萃取揭示標準元件庫與 EDA 工具演算法優化的不連續性，進而進行針對性改進，可以真正提升客戶晶片的競爭力。例如，針對電路關鍵路徑中的元件進行電性或驅動力的平衡補償，或針對時脈平衡與穩壓（避免動態電壓降 Dynamic IR）進行時鐘樹元件的調整；在佈局設計中，運用巧思縮小面積與功耗，並通過接腳位置的優化解決 APR 繞線問題。此外，針對重複性與功能性需求，進行元件的整合與高度客製化，重新架構並再設計元件，能顯著提升效能。

實際應用中，可以基於關鍵路徑與瓶頸的分析結果，確定元件優化與開發策略，並結合系統任務與市場需求，通過統計方法對關鍵元件進行重新開發。例如，針對關鍵路徑的性能瓶頸進行排序優化，最終達到性能、功耗與面積的最佳平衡，實現巨大的設計改進效果。

5.1. 元件時序與功耗建模

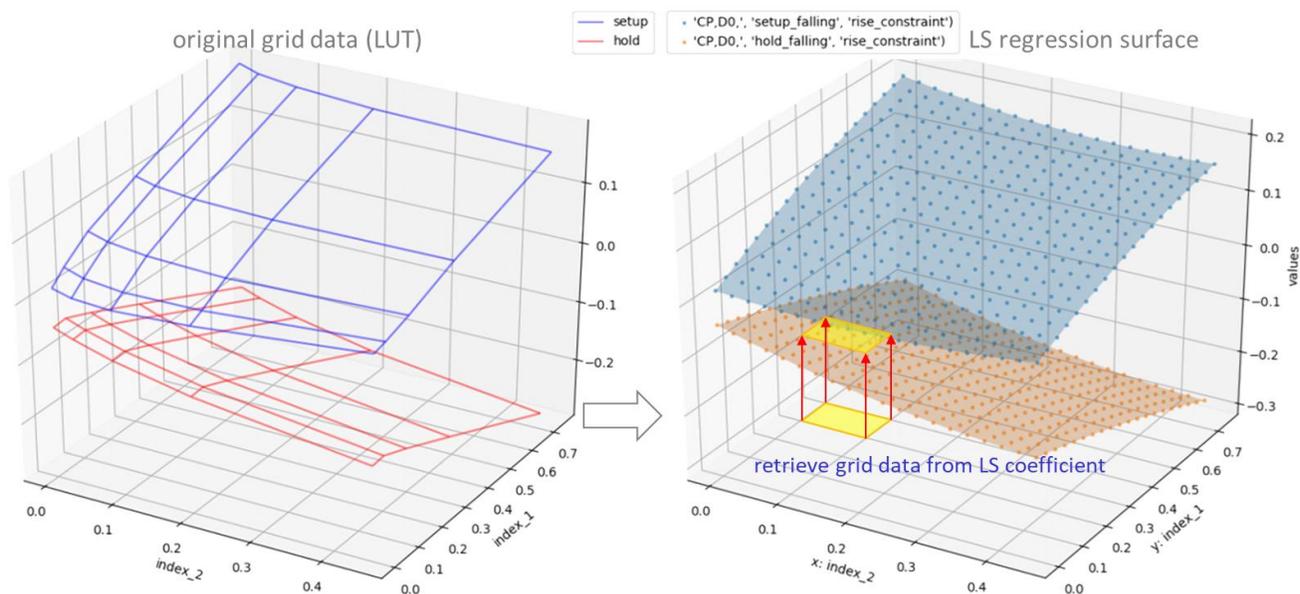


Fig. 5-3 LS Regression

如圖 Fig. 5-3 所示，當前元件庫的時序與功耗數據以離散網格（lookup table）形式存儲，這不僅占用大量記憶體，還帶來許多挑戰，例如難以高效評估元件性能對 PVT 的敏感性或準確比較元件優劣，甚至可能導致誤解。不同晶圓廠對表格索引的單位、解析度及方法缺乏統一標準，且索引刻度通常以 2 的幕次遞增以壓縮數據量，未對齊索引範圍可能導致錯誤推論。此外，為了對齊不同元件間的索引值，工程師需進行大量內插計算，即使是同一元件，由於不同路徑條件（timing arc）和 PVT corner 的變化，也需分別計算，進一步增加工作負擔。

透過回歸分析，我們可以將離散網格數據擬合為二次方程（bias, x , x^2 , y , y^2 , $x*y$ 六個係數），這樣不僅能在統一的索引範圍內進行比較，還能透過內積計算快速重建原始曲面數據。此外，此方法還可擴展至更大範圍，幫助觀察更廣泛的趨勢，並提供更深入的數據分析。

以 INVD1 為例，其延遲 (Delay) 由回歸係數 S_r (cell rise) 與 S_f (cell fall) 給定，這些係數用於計算原始的延遲時間 z_r 與 z_f 。計算公式如下：

$$\text{Delay} = C0 + C1 * x + C2 * x^2 + C3 * y + C4 * y^2 + C5 * x * y,$$

其中， x 表示負載 (Load, 單位：pF)， y 表示轉換時間 (Transition, 單位：ns)。

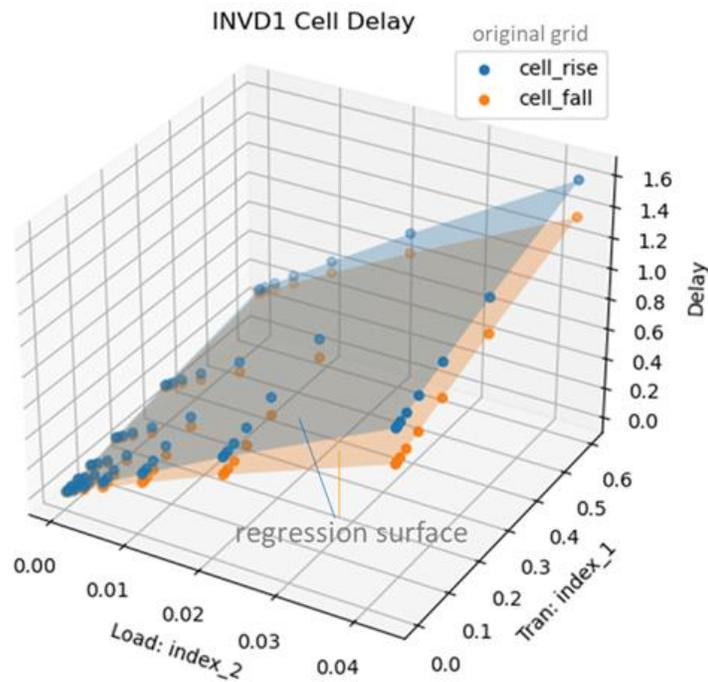


Fig. 5-4 INVD1 Cell Delay

程式碼 Example 5-1 採用最小平方方法回歸 (Least Squares Regression, LS Regression) 擬合數位電路單元 INVD1 的延遲時間 (cell rise 與 cell fall)，並視覺化結果，如圖 Fig. 5-4。

Example 5-1 Cell Timing Modeling

```
# INVD1 Cell Timing Modeling.
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.linalg

x = np.array([0.0002, 0.0005, 0.0012, 0.0026, 0.0053, 0.0107, 0.0215, 0.0432]) # index_2: load (pF)
y = np.array([0.0032, 0.0079, 0.0173, 0.036, 0.0735, 0.1485, 0.2984, 0.5983]) # index_1: tran (ns)
Sr = np.array([0.01, 27.422, -57.761, 0.603, -0.211, 8.85]) # cell rise regression coefficient
Sf = np.array([0.007, 22.323, -56.998, 0.577, -0.223, 9.034]) # cell fall regression coefficient

gx,gy = np.meshgrid(x,y) # mesh grid
tx,ty = map(np.ravel, (gx,gy)) # flattened mesh grid
T = np.array([np.ones_like(tx), tx, tx**2, ty, ty**2, tx*ty]).T # regression grid

zr = np.dot(T, Sr) # original cell rise (ns)
zf = np.dot(T, Sf) # original cell fall (ns)

# LS regression
Cr,_,_ = scipy.linalg.lstsq(T, zr) # LS regression coefficient
Cf,_,_ = scipy.linalg.lstsq(T, zf) # LS regression coefficient
```

```

# visualization
plt.figure(figsize=(7,6))
ax = plt.subplot(111, projection='3d')
ax.scatter(gx,gy,zr.reshape(gx.shape), label='cell_rise')
ax.scatter(gx,gy,zf.reshape(gx.shape), label='cell_fall')
ax.plot_surface(gx,gy,np.dot(T, Cr).reshape(gx.shape), alpha=0.3)
ax.plot_surface(gx,gy,np.dot(T, Cf).reshape(gx.shape), alpha=0.3)
ax.set_xlabel('Load: index_2')
ax.set_ylabel('Tran: index_1')
ax.set_zlabel('Delay')
plt.legend()
plt.tight_layout()
    
```

5.2. 標準元件特徵萃取 (Cell Feature Extraction)

為了執行批量元件特徵萃取，我們需確保每個元件在相同環境條件下操作並進行評比。如圖 Fig. 5-5 (a) 所示，我們將 D1 的 NAND 閘定義為基本單元，並以晶片的主運行工作場景（例如 TT, 0.75V, 25°C）作為採樣條件，負載為驅動 4 個相同基本單元的輸入電容，該結構定義為 BU。如圖 Fig. 5-5 (b) 所示，經過串接 6 級 BU 後，穩定的波形（converged transition）將作為每個待特徵萃取元件（DUE）的輸入，如圖 Fig. 5-5 (c) 所示。

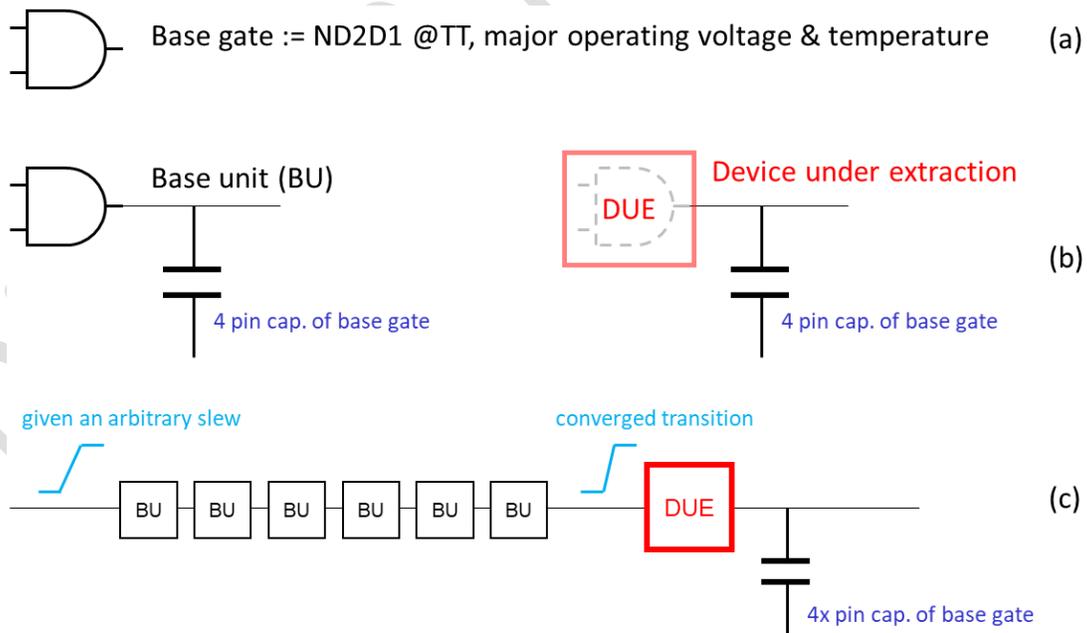


Fig. 5-5 Cell Metric Extraction

目前的標準元件庫使用 Liberty 格式（如 NLDM、CCS、LVF 等）進行表徵，但該格式包含大量數據及多家 EDA 廠商定義的冗餘數據區塊，造成歷史負擔。為加速資料萃取、分析與偵錯，我們將關鍵的時序與功耗信息轉換為輕量化的 JSON 格式（參見開源專案：<https://github.com/dipsci/DTCO/tree/main/libMetric>）。

針對每個元件的時序路徑，進行最小平方法回歸（LS Regression），僅保留回歸係數（LSC），生成元件快照，記錄面積、輸入電容、驅動力、時序與功耗等參數，顯著提升分析效率與精準度。最後，將來自不同 PVT corner 的 Liberty 檔案整合為單一資料庫，建立便於查詢與批量比較的統一平台，優化數據管理與應用流程，如圖 Fig. 5-6 所示。

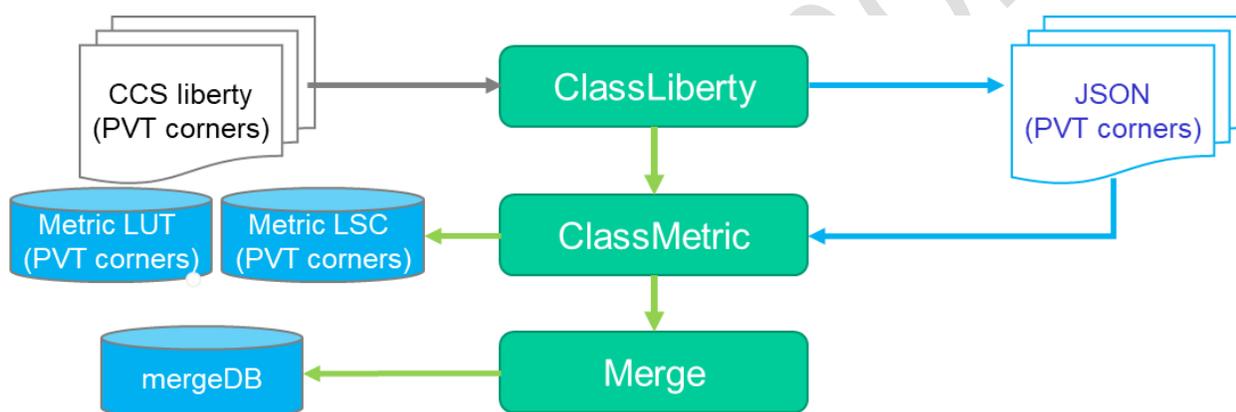


Fig. 5-6 Metric Extraction Flow

透過萃取的物理特徵進行量化和數值分析，我們能高效研究各單元在不同轉換輸入和負載環境下的趨勢，並分析在不同製程參數、通道長度（Channel Length）、臨界電壓（ V_{th} ）、晶片工作電壓和溫度條件下各物理量的變化趨勢與變異。

5.3. RO Simulation

一種常見的製程與元件性能評估方法是設計環形振盪器（RO）並進行 SPICE 仿真，通過 RO 頻率與功耗繪製 2D 散佈圖進行評估。然而，基於先前提到的特徵萃取數據，我們無需建構 RO 或執行 SPICE 模擬，避免了邏輯電路合成、閘級模擬、LVS 佈局驗證及 SPICE 電路提取等繁瑣過程。透過圖 Fig. 5-7 所示的近似計算方法，可快速得出結果。

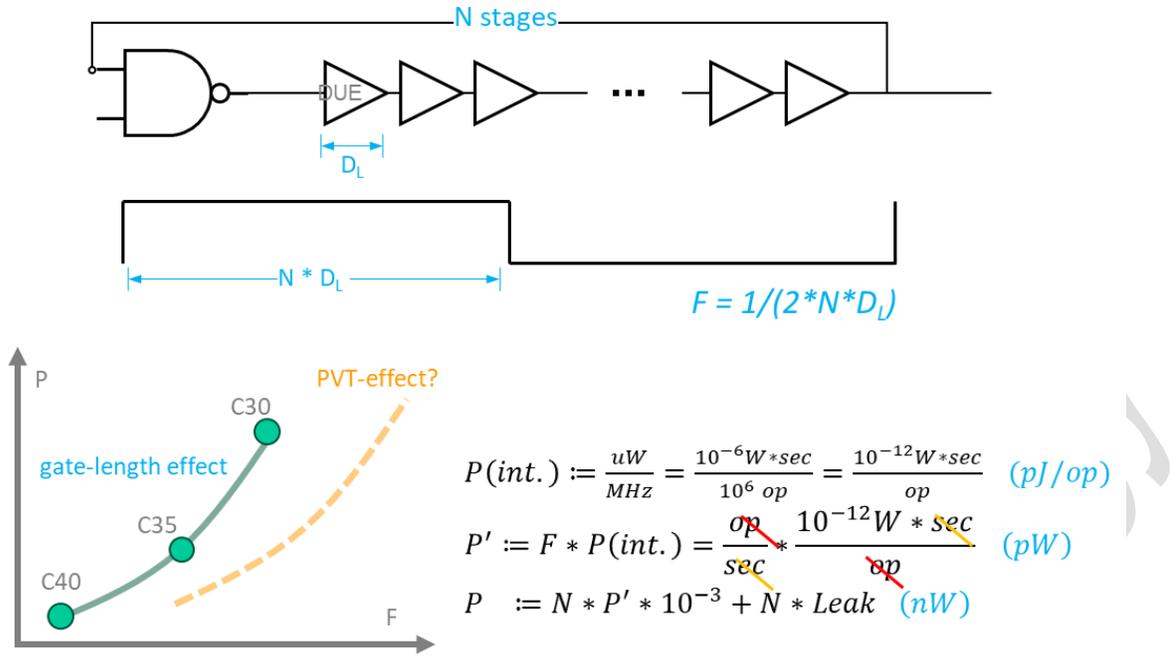


Fig. 5-7 RO Simulation

其中， $P_{int.}$ 表示元件內部功耗，單位為 $\mu W/MHz$ （或 pJ/op ，表示每次電路切換所消耗的能量）。給定待評估的頻率，其功耗為兩者的乘積（單位： pW ）。根據目標頻率推算所需的元件級數 N （忽略第一級 NAND 閘的延遲），RO 的頻率可由延遲周期（ $2N \cdot D_L$ ）的倒數計算得到。其中， D_L 是在特定工作條件下由元件（DUE）萃取的延遲快照。

5.4. 標準元件庫批量 PPA 基準評估

通過前述方法，我們可以高效地對多個不同 PVT corner 的元件庫進行批量比較與評估。例如，在某元件庫 A 中觀察到，從 C20 到 C16 的 SVT 單元，其總功耗（包括動態功耗與漏電功耗）呈現負增益，且隨著工作頻率和切換率的提高，這一增益進一步減少，如圖 Fig. 5-8 所示。

若選擇元件庫 A 在實務應用中，對於低速設計，漏電優化通常需要通過調整不同的閾值電壓（ V_{th} ）與通道長度，但應排除 SVT 和 C20 單元以避免負面影響。對於高速設計（如 CPU 或 GPU），則可能需要暫時禁用漏電優化或替換 C20 單元，以降低總功耗並滿足性能需求。

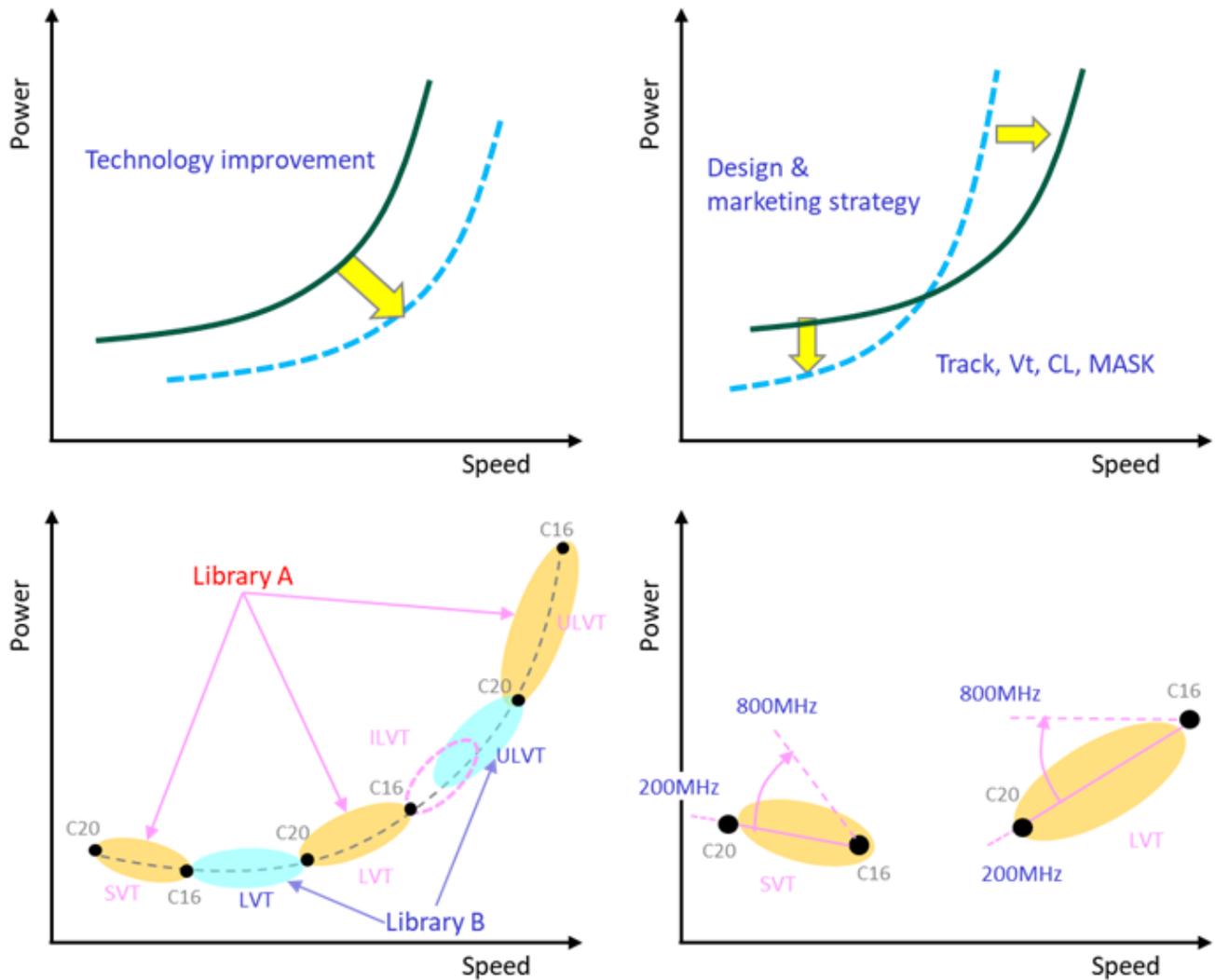


Fig. 5-8 Library Evaluation

Library Metrics 協助工程人員在設計前期評估元件的各種潛在風險，例如在不同電壓場景下的工作需求、針對溫度敏感性變化的分析，以及 timing Re-K cell 的歪斜和 timing window 偏移等問題。如圖 Fig. 5-9 所示，設計者可以先評估所有 DFF 的 setup 和 hold window 要求；若發現 setup 的 skew 問題（例如前端 timing 表現良好但後端 hold-time 無法滿足），可提前將相關元件設置為 don't use，進一步提升設計效率與可靠性。

由於時序特徵以 LS Coefficient 進行記錄，工程人員可以針對特定的 timing arc 調整工作環境條件，或擴展至更廣的工作範圍（透過向量內積），以便進一步探索趨勢，詳細參見圖 Fig. 5-9 右側所示。

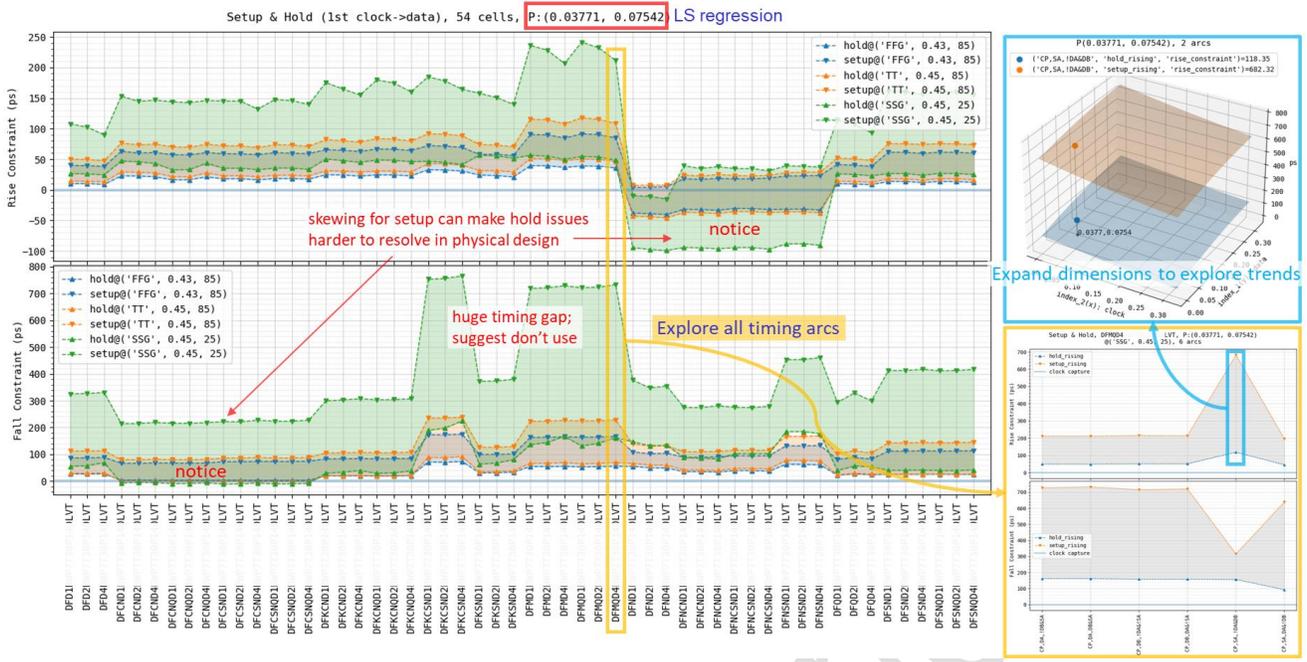


Fig. 5-9 Timing Constraint Batch Comparison

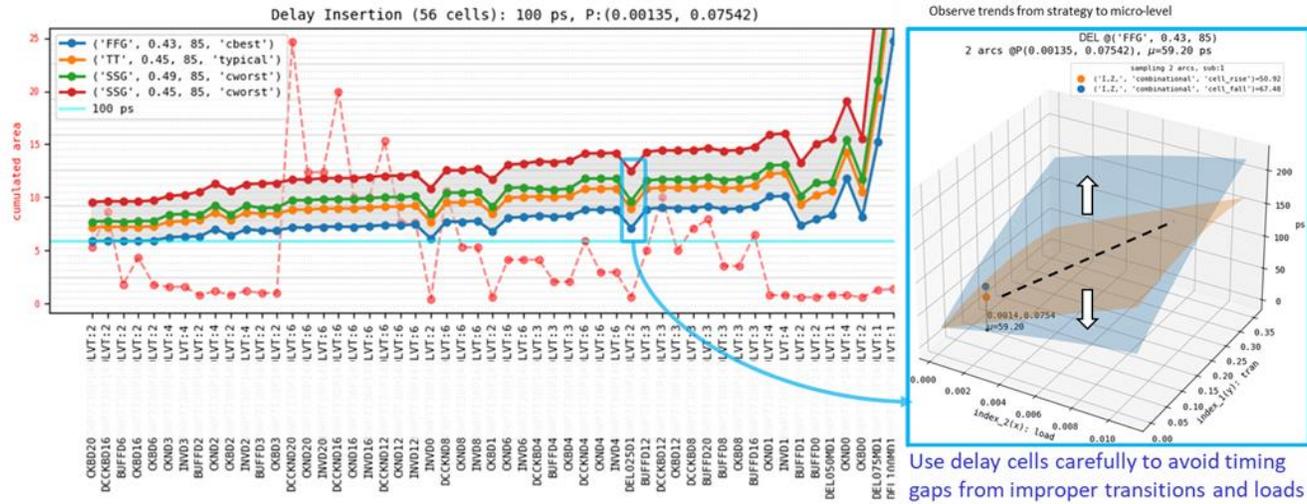


Fig. 5-10 Slew Balance Batch Comparison

如圖 Fig. 5-10 所示，批量特徵萃取可協助識別元件庫潛在風險並提供優化建議。工程人員可根據需求選擇適合的元件進行 clock-tree 優化或 hold buffer 修正。此方法還能及早發現 transition

rise/fall 不平衡，並預測元件在不同 PVT 條件下的表現。根據演算法或應用需求，可在設計前端制定相應規範與策略，例如：

- **初始設計階段禁用**：針對元件庫 A，在工作頻率範圍 200MHz 至 1.2GHz (切換率 10%) 下，建議暫時禁用 SKP、SKN、OPT 單元。
- **驅動力評估**：建議優先合理安排 PLL 和時鐘反向器的位置，確保關鍵路徑的驅動力需求得到滿足，必要時可反饋給元件工程師進行客製化調整。
- **繞線友善評估**：對於具有較多 PIN 腳的複合元件 (如 INR、IND、IINR、IIND、IAO)，因繞線可能出現瓶頸，建議將元件尺寸擴大至 D2 (避免 D1 尺寸)，或反饋給元件工程師進行優化，如圖 Fig. 5-11。

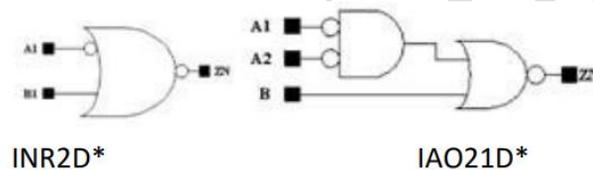


Fig. 5-11 Cell Candidates for Routability Optimization

第6章 晶片內感測器設計與整合 (GRO Compiler)

隨著機器學習的發展，晶片內監測器 (如環振電路，Ring Oscillator, RO) 正逐漸成為必需。如何規劃、簡化架構並最大化效益，仍是工程師關注的挑戰。我們可以從以下幾個角度思考：

- **目標明確化**：我們希望達成什麼目的？例如，掌握矽晶 SPICE 模型與實際性能之間的偏差，優化製程參數與設計簽核配方，或制定晶片功能性補償策略。
- **間接收益**：通過機器學習，我們可以獲得哪些額外資訊？例如，分析晶圓特徵曲面的均勻性 (Uniformity) 以及量化晶片內部變異性 (On-chip Variation, OCV)。
- **先進應用**：如何實現更精細的數據分析和實時監測？例如，動態 IR 感測器、基於週期的 Slack 警示，以及配套的動態調整與系統性補償策略。

透過上述問題的引導，我們能夠明確設計目標，並系統性地規劃 RO 的設計方向，將其整合至晶片中。藉由傳統的實體設計流程和生產測試，我們可以獲得更具針對性的數據，進而基於這些數據優化製程，提升晶片性能。圖 Fig. 6-1 展示了一種網格狀環振電路 (GRO, Grid RO) 的架構。此設計將各類元件整合進可選擇延遲線的 RO 中，並通過 SPI 控制器進行管理。這種配置使多個 RO 能夠策略性地分佈於晶片上，尤其是在容易受到 IR 壓降影響的區域，實現更高效的監測效果。

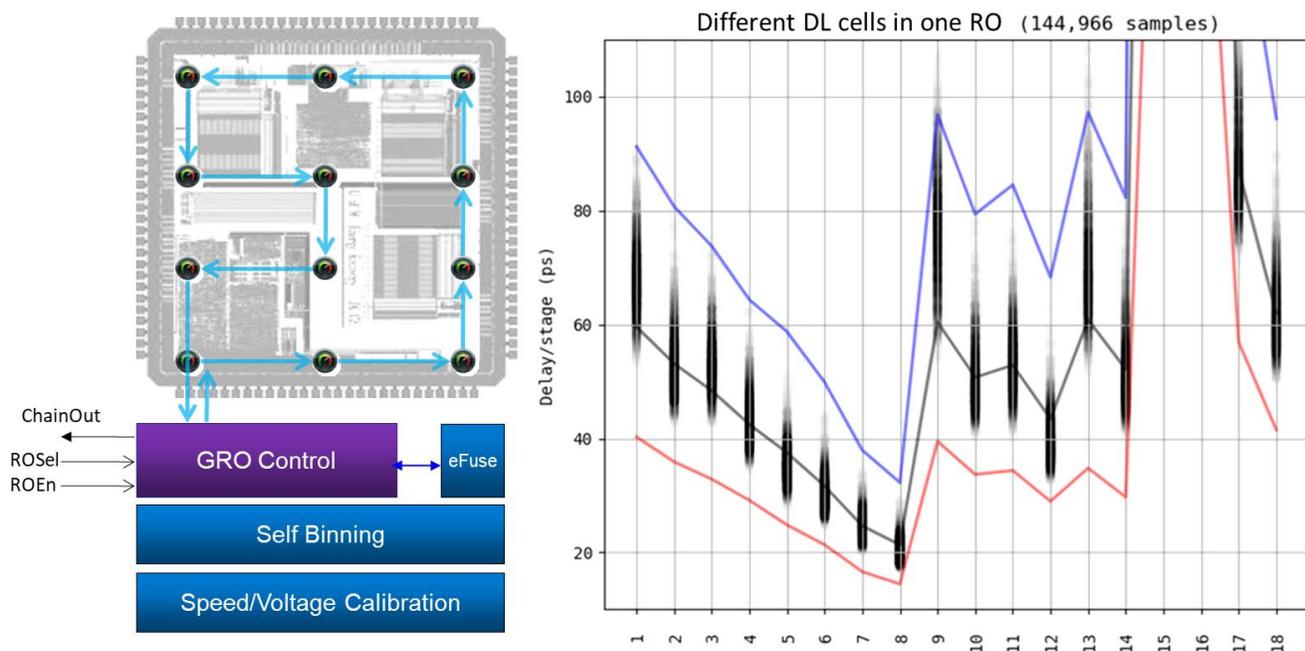


Fig. 6-1 GRO Integration and SPICE-Silicon Correlation

如圖 Fig. 6-2 所示，典型的 RO 環振電路由 NAND 閘控制的 REN 接腳作為開關，串接多級延遲元件 (Delay-Line, DL)，並透過多位元計數器記錄輸出頻率。若 RO 用於對比元件延遲的 SPICE 模型，建議延遲元件級數應佔 RO 週期的 99% 以上，以降低控制閘變異性的影響。實體佈局上，建議採用垂直排列或交錯配置不同 RO 群組，並確保充足的電源分佈，以減少局部動態電壓降對數據分析準確性的影響。

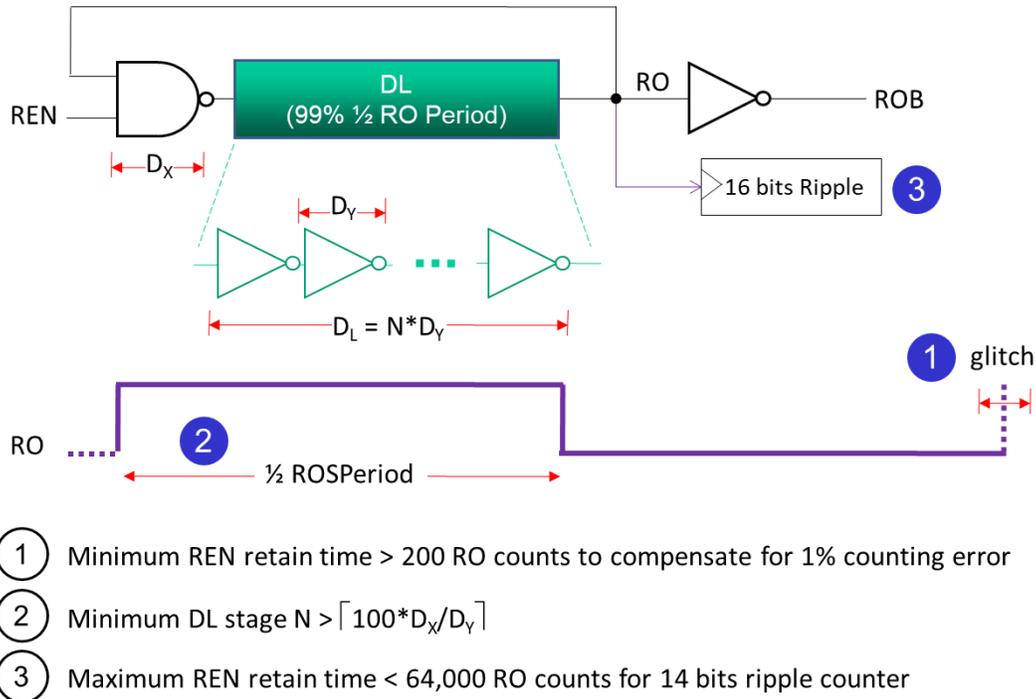


Fig. 6-2 RO Design Guideline

我們也可利用晶片唯一識別碼 UID 重新調整性能晶片評等策略與篩選方法，並探索新的 RO 結構。例如，將兩個相距較遠的 RO (功能性 RO) 串接，以形成更大尺度的 gross-effect RO。此外，可進一步探討 RC 變異效應的量化方式，儘管其與元件驅動能力相關，仍可透過數據分析「間接」推導出 RC 效應的回歸模型。

6.1. 目標導向的 RO 設計

另一個關鍵課題是，在控制面積開銷的同時，如何最大化設計效益。RO 可由多種不同標準元件組成延遲元件 (Delay-Line, DL)，但如何選擇這些元件來實現最優化的 RO 設計？

如圖 Fig. 6-3 所示，實體設計過程可分為多個階段，包括前端的設計約束 (Design Constraint)、F/V Shmoo 評估，以及後端晶片佈局的 pre-CTS、post-CTS、post-route 等階段。透過分析 cell usage 報表，我們發現，面積與功耗的前 80% 主要來自於前 20% 的核心單元與時鐘樹 (Clock-Tree) 相關元件，這些元件可作為 S2S (SPICE-Silicon) 優化的重點。例如，SHA-3 演算法廣泛使用 XNR、XOR 元件，而 Switch 模組則常依賴多位元的 MUX 元件。

此外，若關注 N/P 製程參數的平衡，可引入特定結構元件，如 P-MOS stacking 的 NOR 和 N-MOS stacking 的 NAND，作為輔助指標來協助後矽測量，評估製程對 N/P 偏移的影響。這些元件有助於在滿足性能需求的同時，提供關鍵的製程優化依據，從而進一步提升整體設計效益。

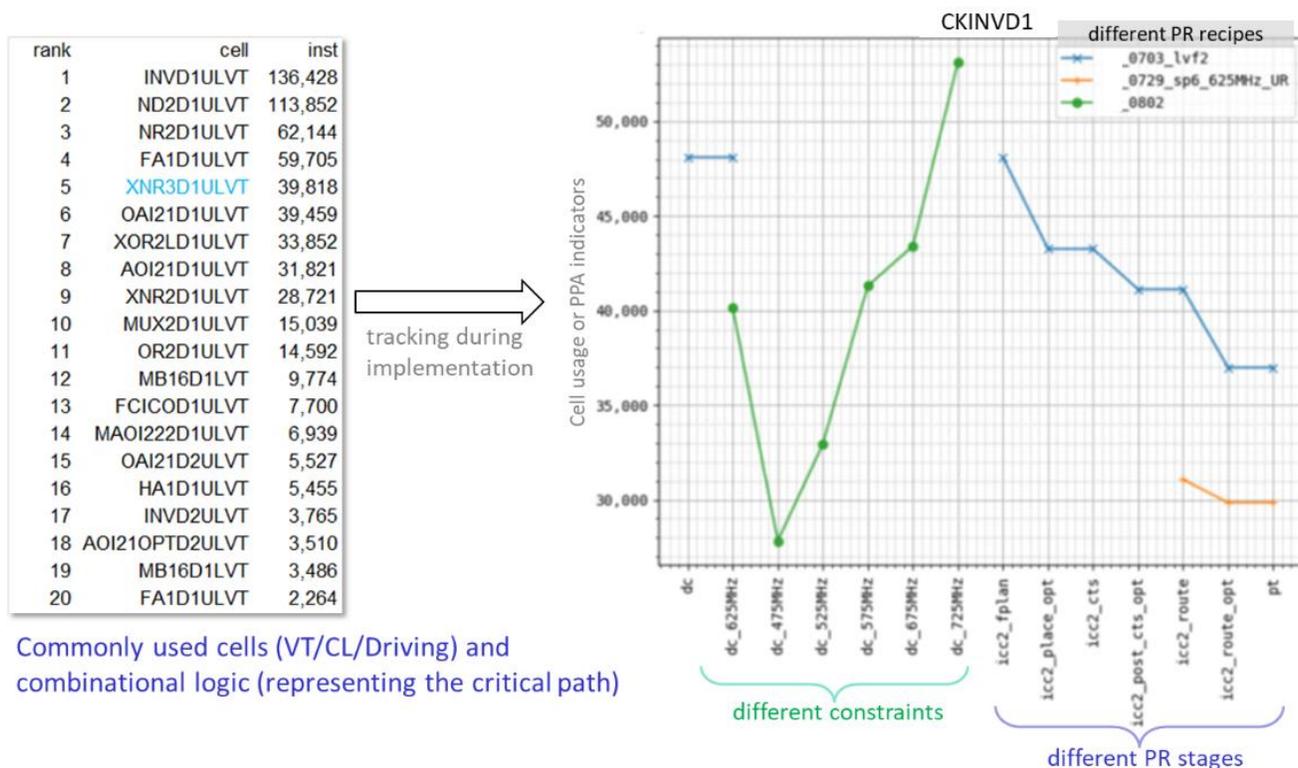


Fig. 6-3 Cell Usage and PPA Evaluation

6.2. SPICE-Silicon 相關性分析

如圖 Fig. 6-4，後矽 RO 測量結果可精確比對 SPICE 模擬在不同製程條件 (SS、TT、FF) 下的目標值，從而確定當前製程條件下的實際分佈及其與 SPICE 預測的偏差，精度可達皮秒 (pico-second) 級。若進一步引入 Lot-Wafer ID 或時間軸參數，則可追蹤製程參數調整對晶片性能分佈的影響，並評估製程控制的能力與穩定性。

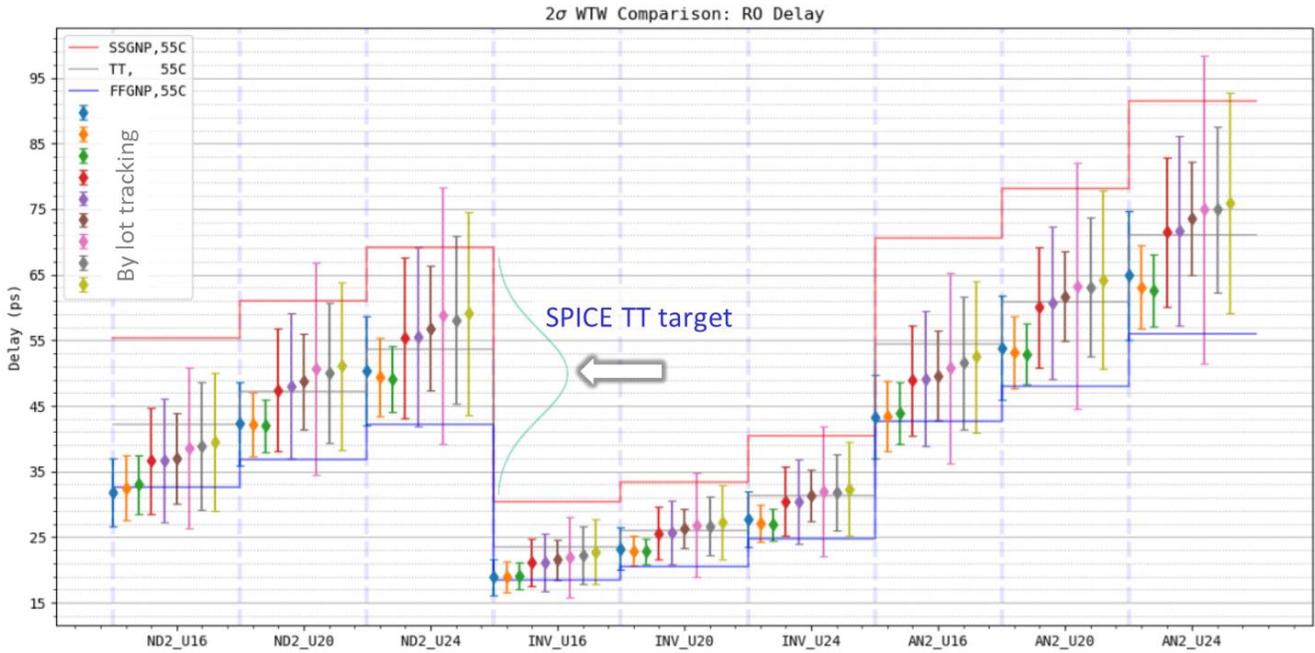


Fig. 6-4 RO SPICE-Silicon Correlation

受成本限制，後矽晶片的量測環境自由度通常較低。此外，測試環境中的 IR-drop 與溫度變化等因素往往難以完全控制，而 SPICE 模擬條件設定亦無法涵蓋所有可能場景。在此情況下，可透過少量或稀疏的模擬數據構建回歸模型，以預測非模擬條件下的行為，從而輔助後矽量測數據與模擬結果的對應校準，提升分析的準確性與效率，如圖 Fig. 6-5 所示。

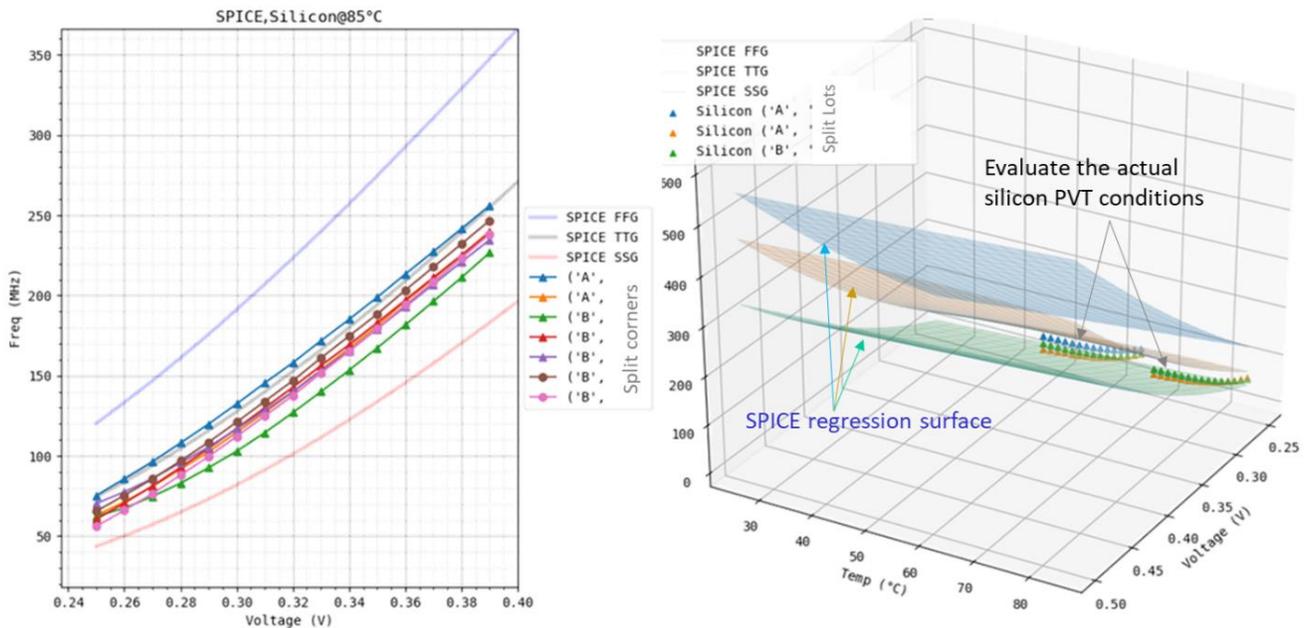


Fig. 6-5 V-F Shmoo and SPICE Correlation

6.3. 製程追蹤與調整

製程追蹤與調整是確保晶片性能穩定與提升良率的關鍵環節。透過晶圓廠提供的 WAT 與 CP 資料，結合晶片內環振電路 (RO) 數據，可持續監測製程參數變化，例如臨界電壓漂移 (Vth shift)、漏電流 (SIDD) 變化等，進而掌握製程變異對電路性能的影響，如圖 Fig. 6-6 所示。此外，RO 提供的高靈敏度時序數據，可有效反映製程的均勻性與局部變異性。基於這些監測數據，工程師可採取以下策略進行優化調整：

- **製程偏移校正**：根據監測結果優化關鍵製程參數，修正時序萃取參數 (S2S/Re-K) 並校準時序簽核目標，以降低製程偏移對電路的影響。
- **功能性補償**：運用機器學習演算法進行系統級晶片性能評估 (Binning)，並透過動態調整晶片內部參數 (如動態電壓與頻率調整，DVFS) 來補償製程不一致性。
- **長期趨勢分析**：累積不同批次晶片的 RO 數據，識別製程變化的長期趨勢，為製程改良與新設計提供數據支持。

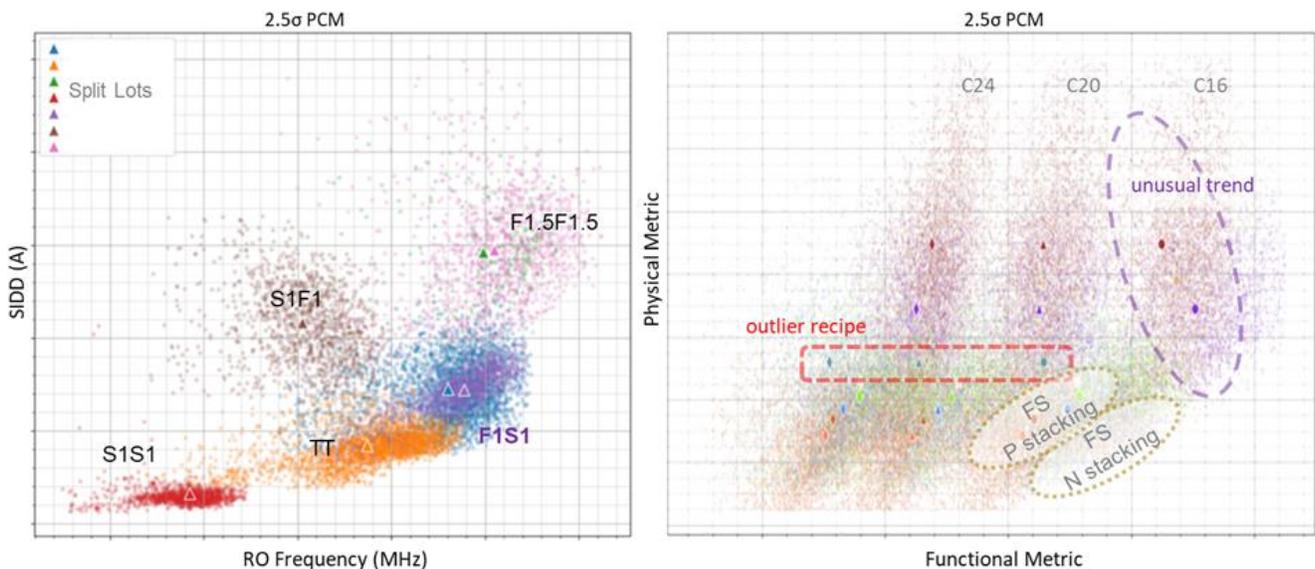


Fig. 6-6 Process Tuning

6.4. 晶片內局域電壓分析

6.4.1. 晶片內局域電壓分布監測

若面積允許，可在晶片內均勻布設高密度 RO，並透過 V-F sweep 測量 RO 頻率，如圖 Fig. 6-7 所示。RO 頻率與電壓大多呈線性關係，製程變異影響主要反映在截距變化，斜率影響較小。其中，每種顏色代表單片晶片內 24 個不同位置的 RO 頻率分布。

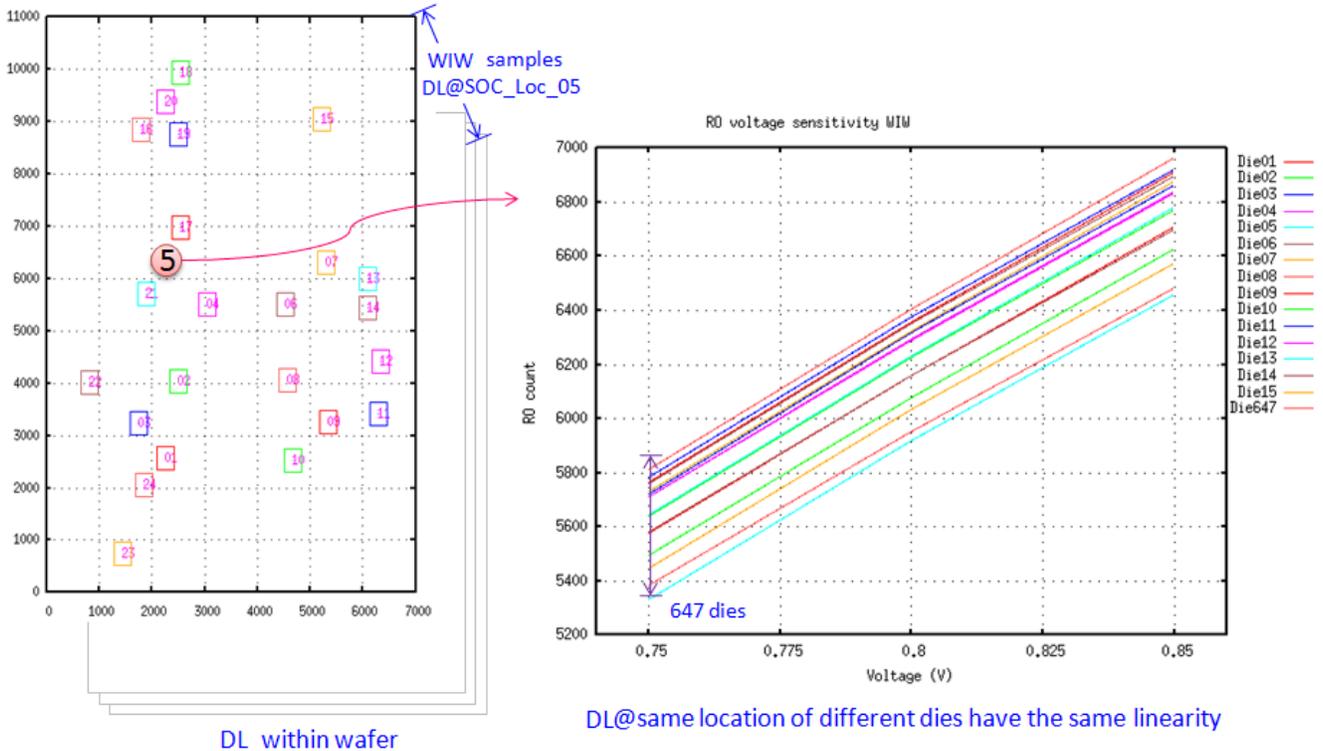


Fig. 6-7 On-chip Ring Oscillators (ROs) and V-F Curve

利用此特性，可將 RO 頻率差異轉換為相對有效電壓變化，以反映電壓分布，但僅適用於相對電壓差，無法推算絕對電壓值。如圖 Fig. 6-8 所示，不同 RO 數據可反推出各偵測點的等效電壓，而單片晶片內的區域等效電壓分布看似隨機，且每片晶片的起伏趨勢不同。

然而，如圖 Fig. 6-9 所示，若統計同一片 wafer 上約 700 顆晶片的數據，並取相同座標 RO 的頻率均值，則可發現頻率與位置呈現強烈相關性（粗黑線），甚至超越製程變異影響。值得注意的是，此趨勢在 SS、TT、FF 各製程條件下皆保持一致。

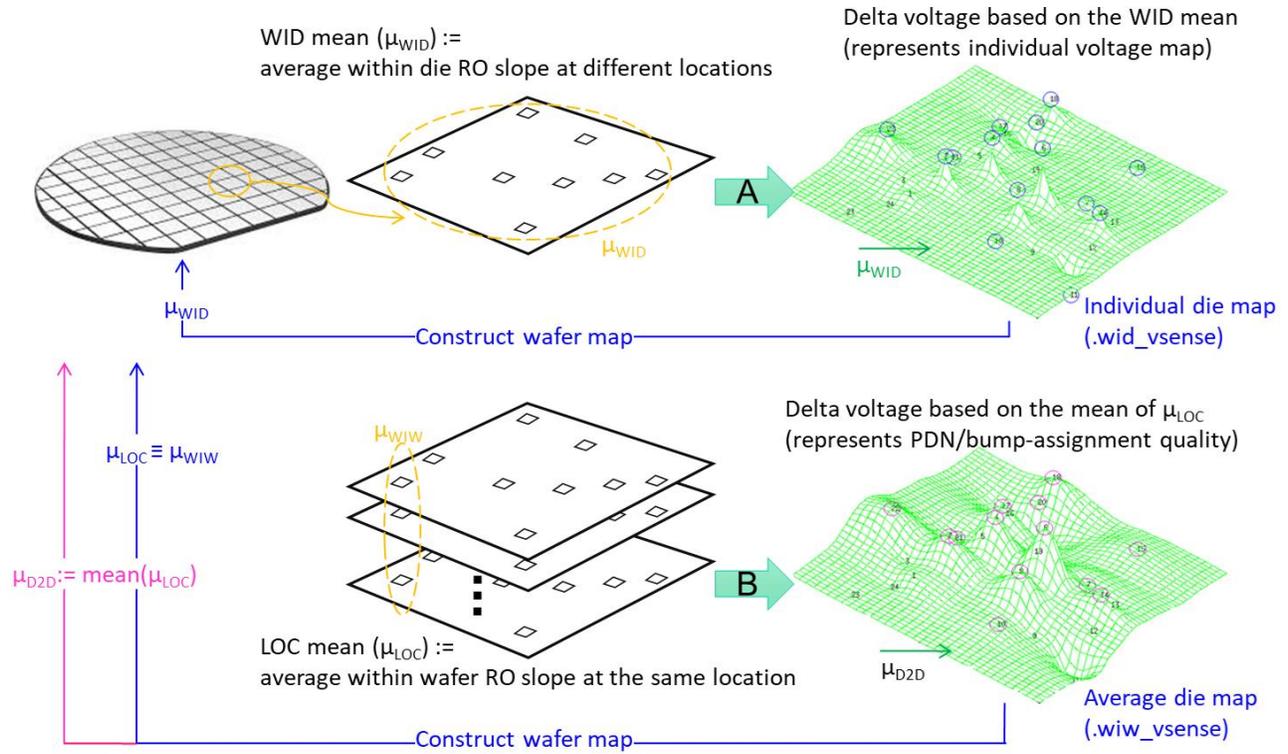


Fig. 6-8 On-chip Effective Voltage Regression

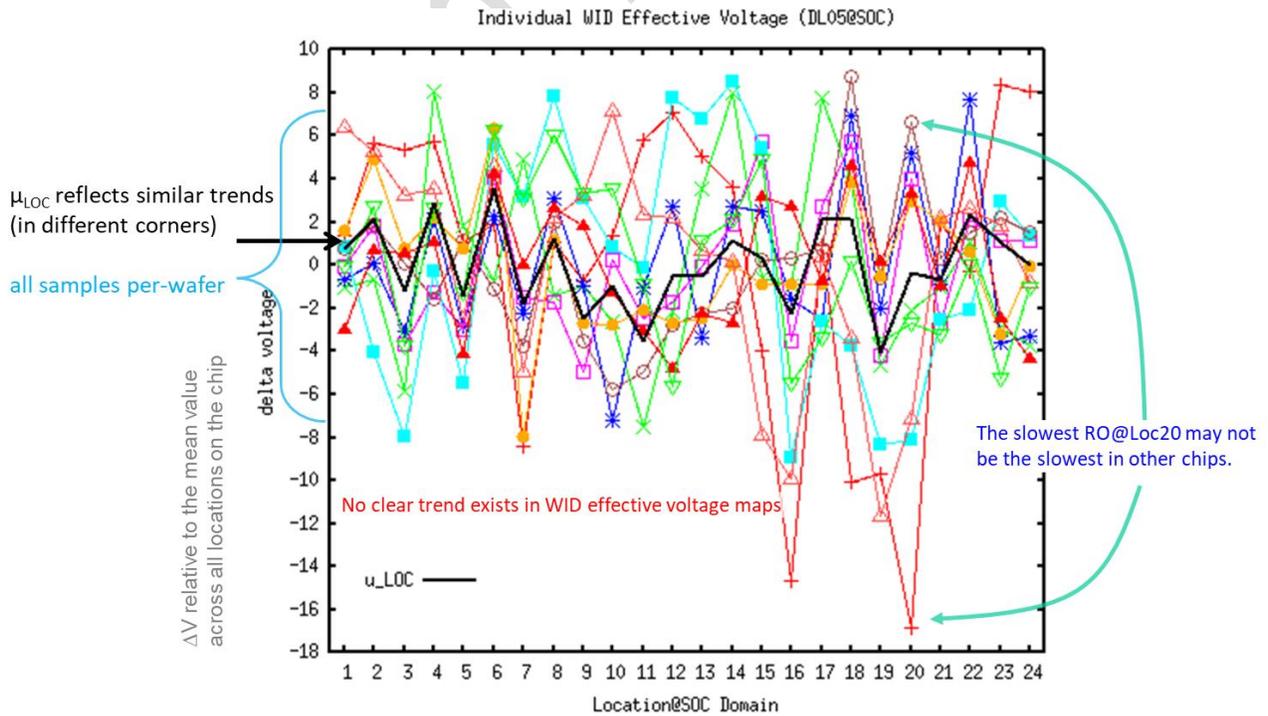


Fig. 6-9 On-chip RO Effective Voltage Analysis

這反映出一種非隨機的系統性問題或結構性缺陷，可能與金屬密度、佈局效應或 PDN/Bump 分配等因素相關，導致整體電壓分佈的不平坦性，必須在設計過程中加以考量。此外，若針對同一片 wafer 上的所有晶片，依照 RO 的座標位置分別取均值後，再分析其 WID (Within-Die) 等效電壓差的機率密度分佈，則可發現即使在不同製程條件 (如 SS、TT、FF 或偏態製程 SF/FS) 下，該分佈趨勢仍保持一致。值得注意的是，無論提升或降低晶片整體電壓，這種等效壓差趨勢皆無法被消除，如圖 Fig. 6-10 所示。

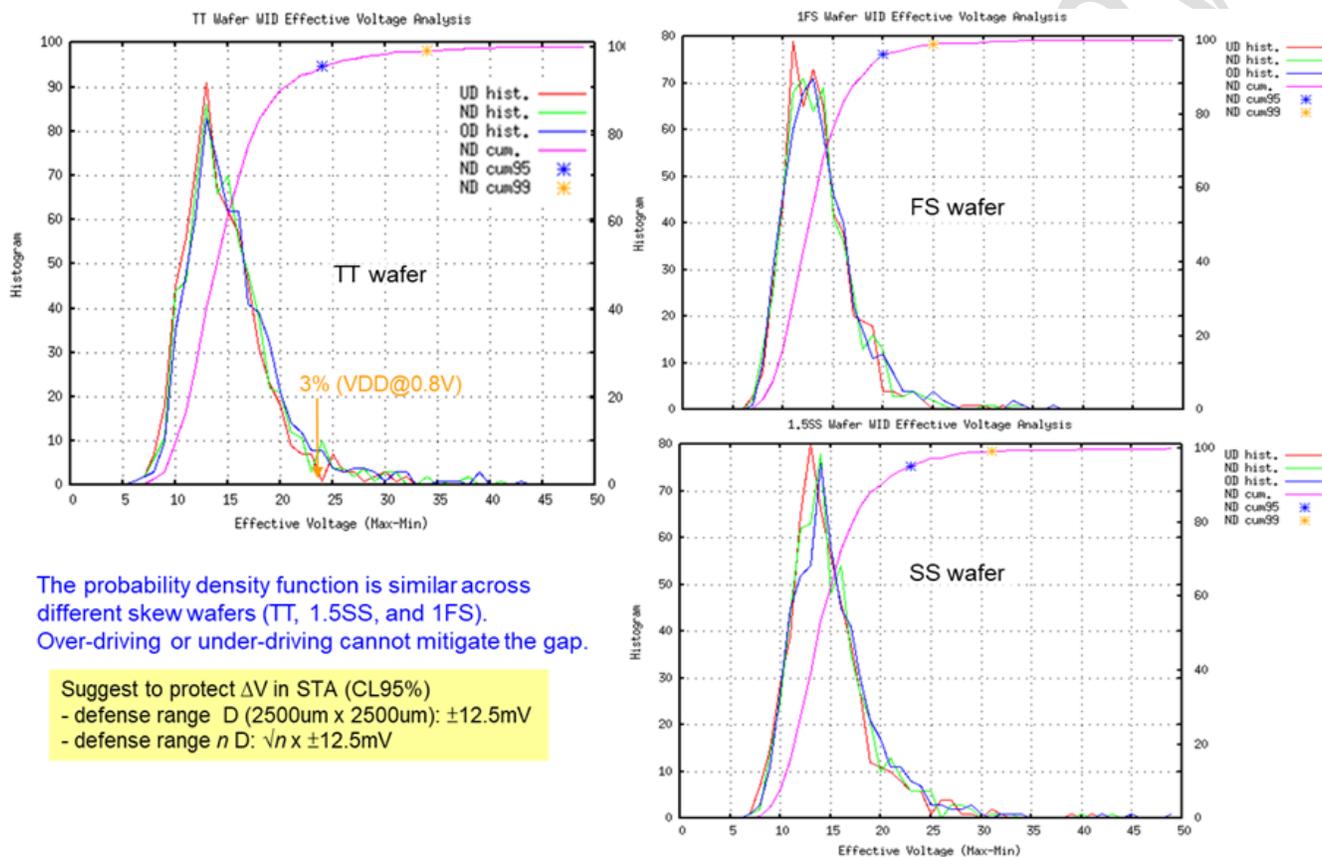


Fig. 6-10 Voltage-drop Probability Density

進一步對 WID RO 均值離散數據點進行三次多項式曲面回歸後，可得出有效電壓等高線的熱圖，如圖 Fig. 6-11 所示。分析結果顯示，電壓差與 power bump 分配密度高度正相關，並觀察到先天電壓差範圍達 -12mV 至 $+6\text{mV}$ (晶片已上電但未運行)。透過簡單的第三方 ERA (Early Rail Analysis) 分析工具，可均勻貼上 tap-current 進行預測，提前識別潛在問題，並在 ERA 分析中定位熱點，這些熱點可能進一步演變為晶片上的低壓區塊。

值得注意的是，此靜態效應與製程無關，且不具隨機性，對良率的影響應受到更多關注。因此，應在實體設計階段預先進行規劃與優化，以避免後期才依賴傳統的動態 IR 分析。由於動態 IR 分析通常發生在設計後期，不僅耗時，且一旦問題發生，往往難以有效解決。因此，在前期即考量電壓分佈問題，能顯著降低設計風險，提高晶片的整體可靠性與良率。

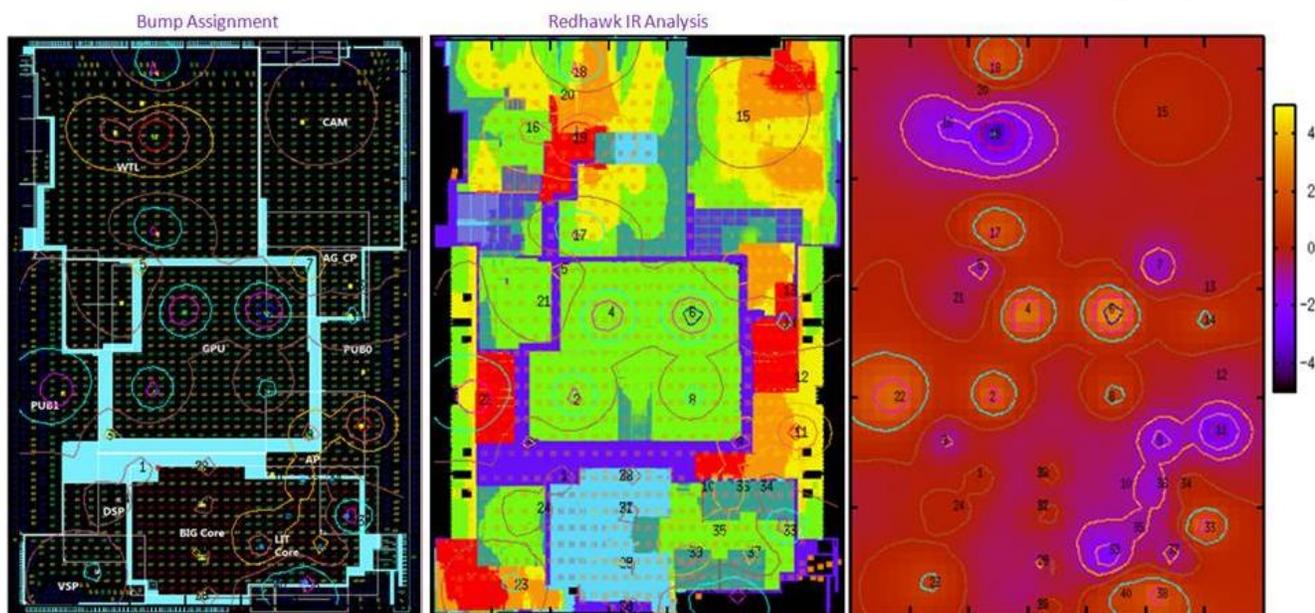


Fig. 6-11 Effective Voltage and ERA Correlation

6.4.2. 補償策略制定

在晶片量產過程中，由於多種因素，經常會出現非預期的結果。例如，SPICE 模型的固有誤差、實體設計流程的不完善、過於激進或保守的時序簽核策略、製程控制不當（包括量產製程參數漂移和晶圓均勻性不足）等問題。如圖 Fig. 6-12 所示，若選擇大幅調整製程參數進行補償，例如將目標從 TT 製程參數轉向 FF 製程參數，雖然可能達到速度要求，但卻可能引發漏電流及能耗指數性增長，並使晶片性能對溫度變化變得難以控制。因此，電壓補償作為一種較為可預測且成本較低的方式，無論是靜態還是動態補償，都能有效平衡性能與能耗。

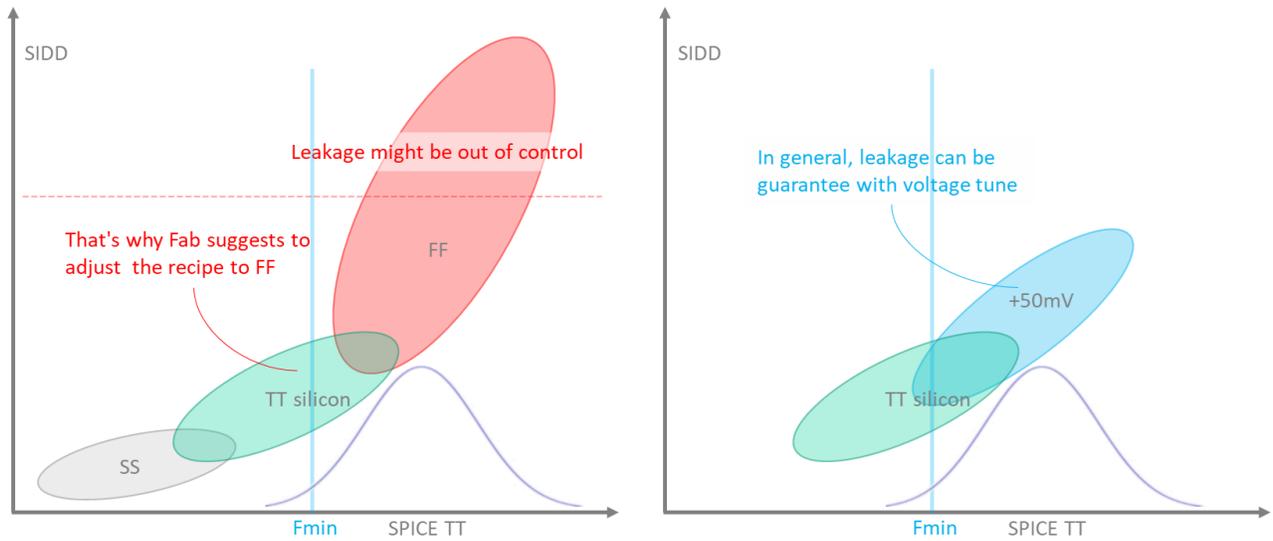


Fig. 6-12 Process and Voltage Tuning

然而，現實情況往往並非如此直觀。製程參數的偏移並非隨機或呈高斯分佈，而通常伴隨明顯的機械式諧波（harmonic spin），這成為設計餘量誤差的主要障礙。這些非隨機性與製程及測試環境的偏差疊加後，最終才引入隨機性成分。這種非隨機性使得原本被認為穩妥的防守策略實際上變得不可靠。圖 Fig. 6-13 顯示了晶圓的各種物理特性，如 RO 頻率或 SIDD，隨電壓或溫度變化的趨勢。可見，局部的不平坦性無法僅依賴電壓補償來解決，晶片性能的梯度也無法通過電壓補償消除。

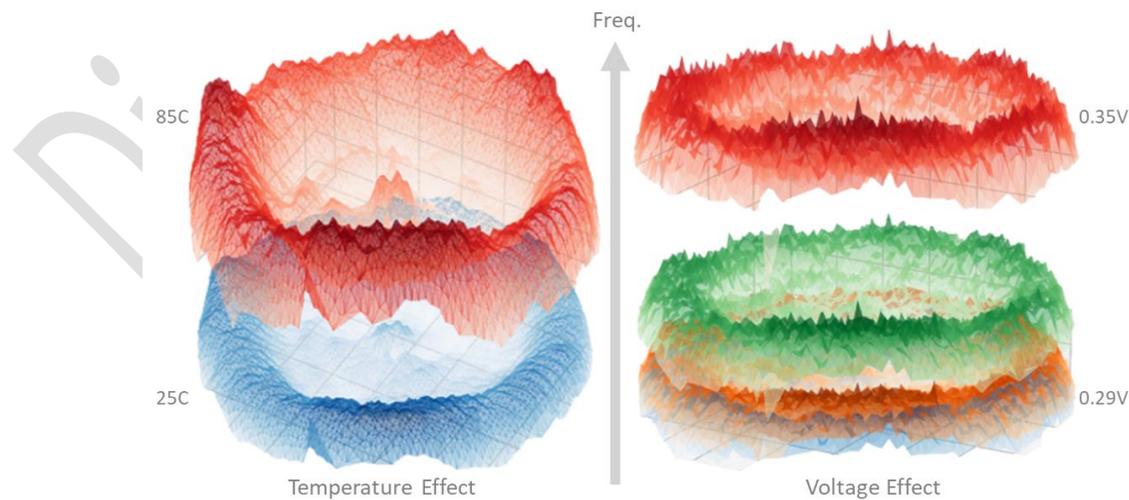


Fig. 6-13 RO Frequency vs. Temperature and Voltage Effects

在先進晶片設計中，結合晶片內監測器 (On-chip Monitors, OCM) 與機器學習補償機制，已成為提升性能與可靠性的關鍵手段。晶片內監測器能實時收集電壓、頻率、溫度等運行參數，提供高精度狀態信息，為靜態與動態補償提供基礎。圖 Fig. 6-14 展示了一個涵蓋系統級設計到設計架構各層面的多層次性能補償策略。圖示中建議採用多層次補償策略，包括：製程參數優化、晶片分級評估、電壓頻率配置，以及實時電壓與頻率的動態調整，全面提升晶片效能並確保穩定性。

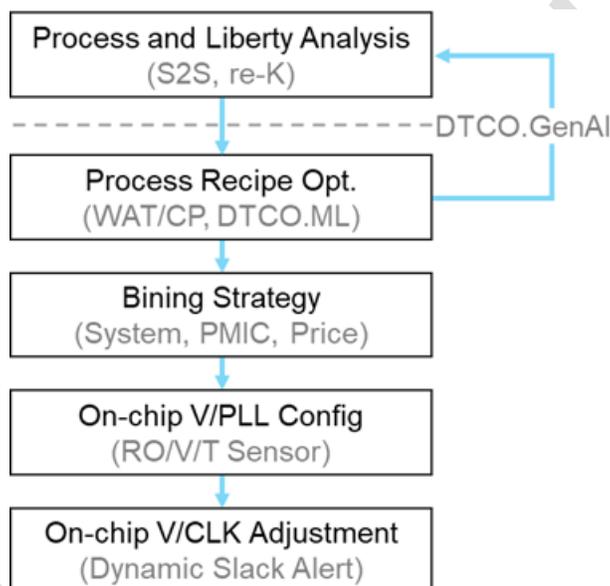


Fig. 6-14 多層次補償策略

透過簡單的靜態 RO 並結合機器學習技術，設計者得以解析多維數據間的複雜關聯性，優化製程參數，進而實現晶片分級評估 (Binning Strategy) 與系統級電壓補償，達到平衡定價與產能的最佳化，如圖 Fig. 6-15。此外，基於動態時序餘量警示 (Dynamic Slack-alert) 電路的即時反饋，進行電壓與時脈的調整，有效平衡性能、功耗與可靠性。該設計方法不僅適用於高效能計算晶片，還能顯著提升物聯網與嵌入式系統的設計效率與能源利用率，展現廣泛的應用價值。

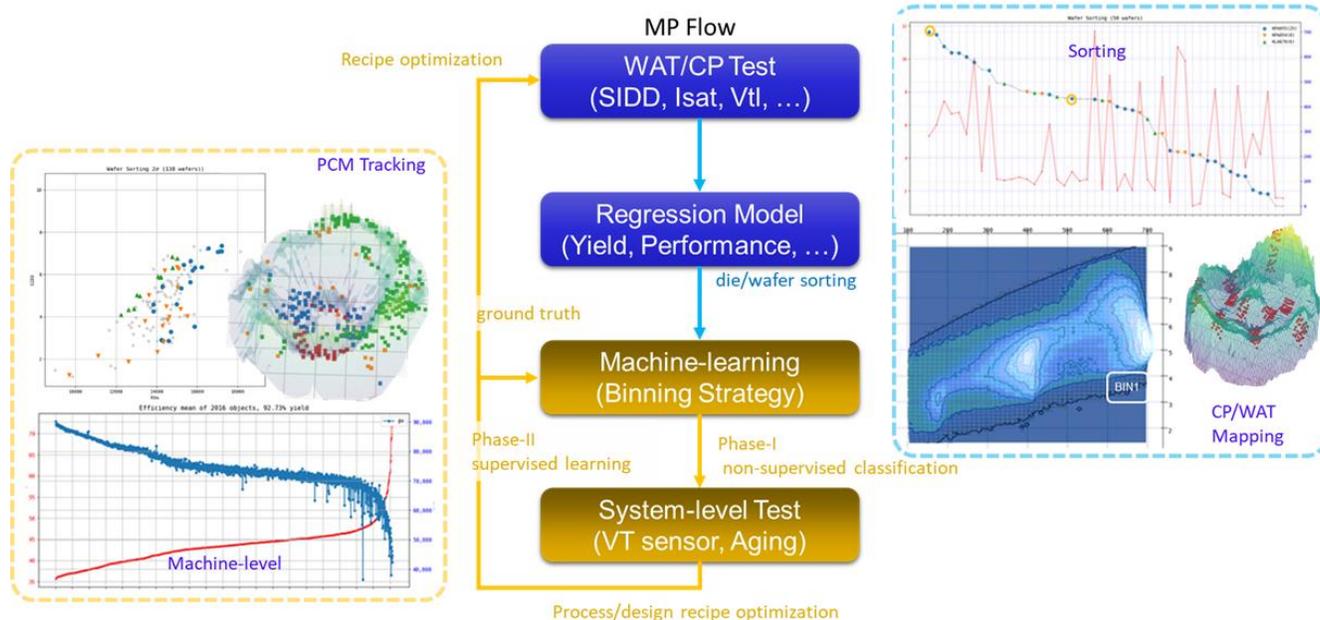


Fig. 6-15 Binning Strategy

6.4.3. 動態時序餘量警示與佈局

基於 RO 的感測方案通常具有微秒級的響應時間，這對於時序修正來說過於緩慢，無法直接反映資料路徑中的時序餘量。為了應對這一挑戰，基於 RAZOR 架構的動態時序餘量警示系統，作為晶片監測 IP 的進階功能，優化了設置警示 (SA) 和保持警示 (HA) 配置。該系統根據內部電壓與溫度變化，實現不同位置的實時時序餘量測量，並將結果編譯為高解析度位元，用於晶元時序監控，直接反映資料路徑中的設置和保持時間不確定性。其數據串行輸出可支持動態電壓與頻率調整 (AVS) 等可擴展設計，幫助追蹤時序簽核，避免過度設計，並識別最低運行頻率錯誤 (Min. Frequency)，支持實體設計中的時序驗證。

實務中，建議在實體設計階段的 P&R 後進行初步的 STA 和 IR 分析。對於 IR 風險較高區域，可選擇餘量較大的次關鍵時序路徑，將最後一級 DFF 換成 SA/HA，確保額外電容或繞線所帶來的懲罰不影響關鍵路徑的性能，如圖 Fig. 6-16 所示。

在先進技術節點下，時序簽核變得更具挑戰性，需仔細考量製程變異、溫度變化及靜態/動態 IR 熱點，避免過度設計導致晶片的 PPA (性能、功耗、面積) 惡化。傳統方法通常依賴在時鐘樹

中加入不確定性並對資料路徑設置時序降額，然而這可能導致過度設計。為解決此問題，將環形振盪器 (RO) 和模擬關鍵資料路徑嵌入晶片，並在 CP 和 WAT 測試中關聯晶片內部特徵，如漏電流/動態電流 (SIDD/DIDD) 和最大運行頻率 (Fmax)，可簡化時序簽核的追蹤，並引入設計技術協同優化 (DTCO) 以進行先進節點的電路設計。

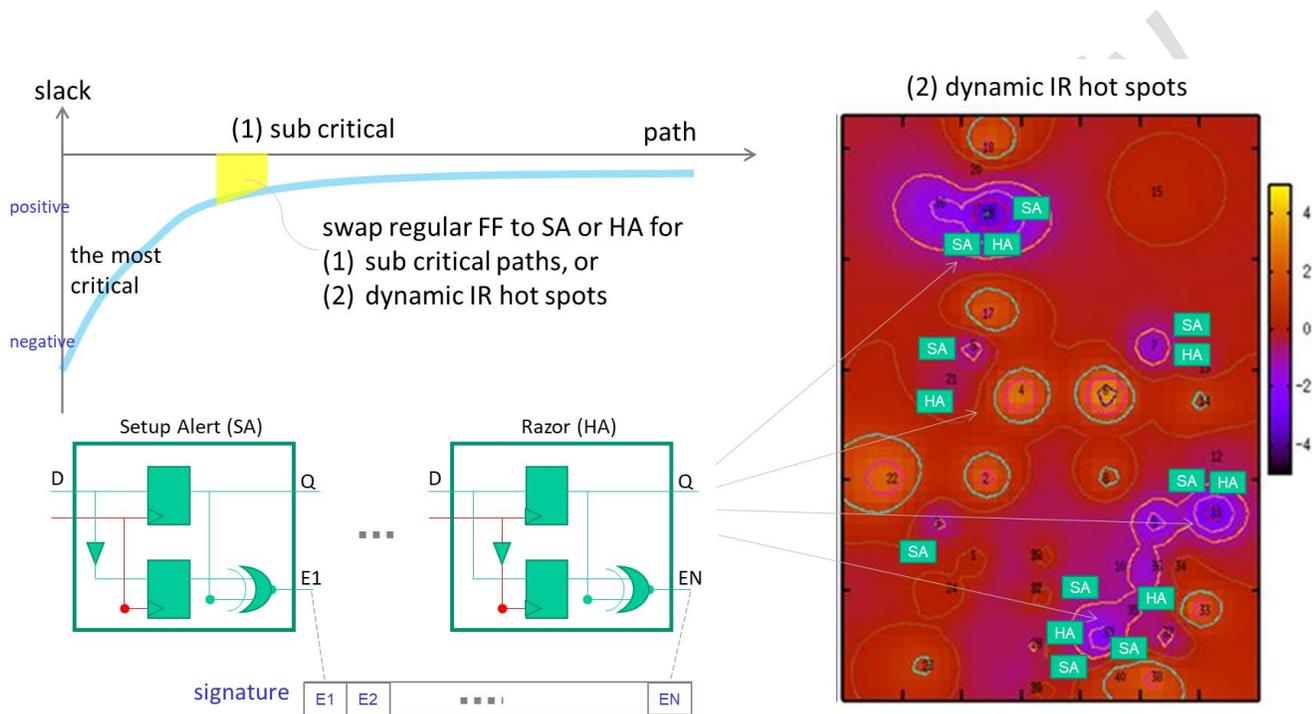


Fig. 6-16 Dynamic Slack Alert Integration Guideline

6.5. GRO 自動化工具與驗證流程

未來，晶片內預埋監測電路的概念將逐漸普及。然而，由於缺乏互通的標準規格，產業迫切需要一個可靠且便捷的解決方案，能夠自動生成基本的監測電路 IP 和測試程式樣本。結合 eFuse、WAT/CP/Aging 等測試所累積的標準化數據分析與機器學習，未來將出現更多可擴展的 EDA 應用場景，如晶片性能評等與電壓/頻率補償等，這些場景值得進一步規劃與開發。

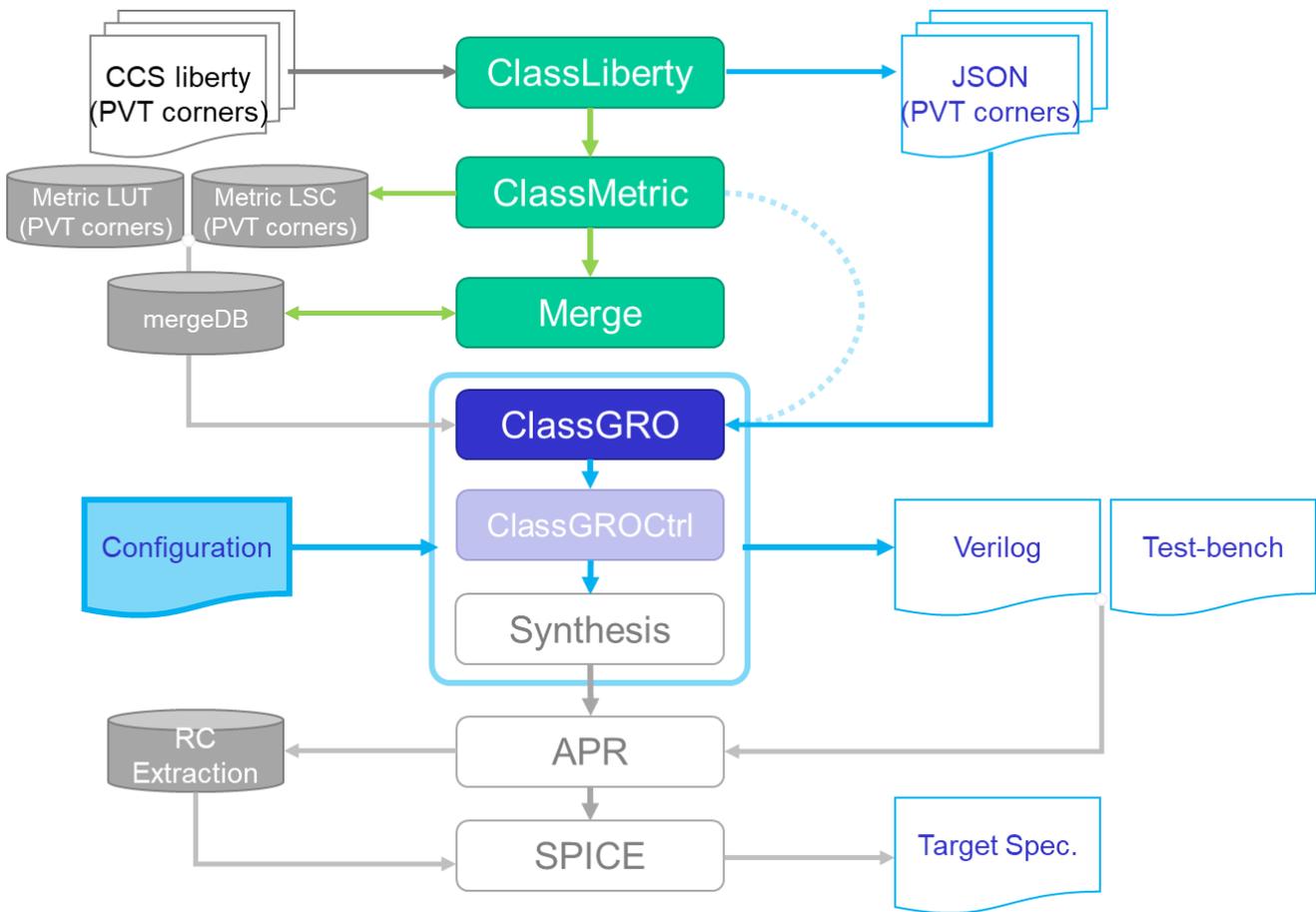


Fig. 6-17 GRO Automation Flow

圖 Fig. 6-17 為所提議之 EDA 開源工具 (<https://github.com/dipsci/DTCO/tree/main/GRO>) 及自動化流程。透過讀取配置檔案 (Configuration)，該檔案可定義多組目標 Delay-Line、頻率及最大計數值。工具將自動萃取 Liberty 資料，並獲取所有關鍵元件的時序、輸入輸出等資訊，根據目標設定自動生成 RO 監測器 RTL。

此外，該工具還提供自動化合成和測試環境生成，用於驗證電路的功能性和完整性，如圖 Fig. 6-18 所示。生成的 RO 監測器可以包含多個 Delay-Line，經過 Hardening 和 LVS 驗證後，萃取 RC 參數並進行 SPICE 仿真。仿真結果將作為後矽量測的 SPICE-silicon 比對校準參考，並保留以供後續對比使用。

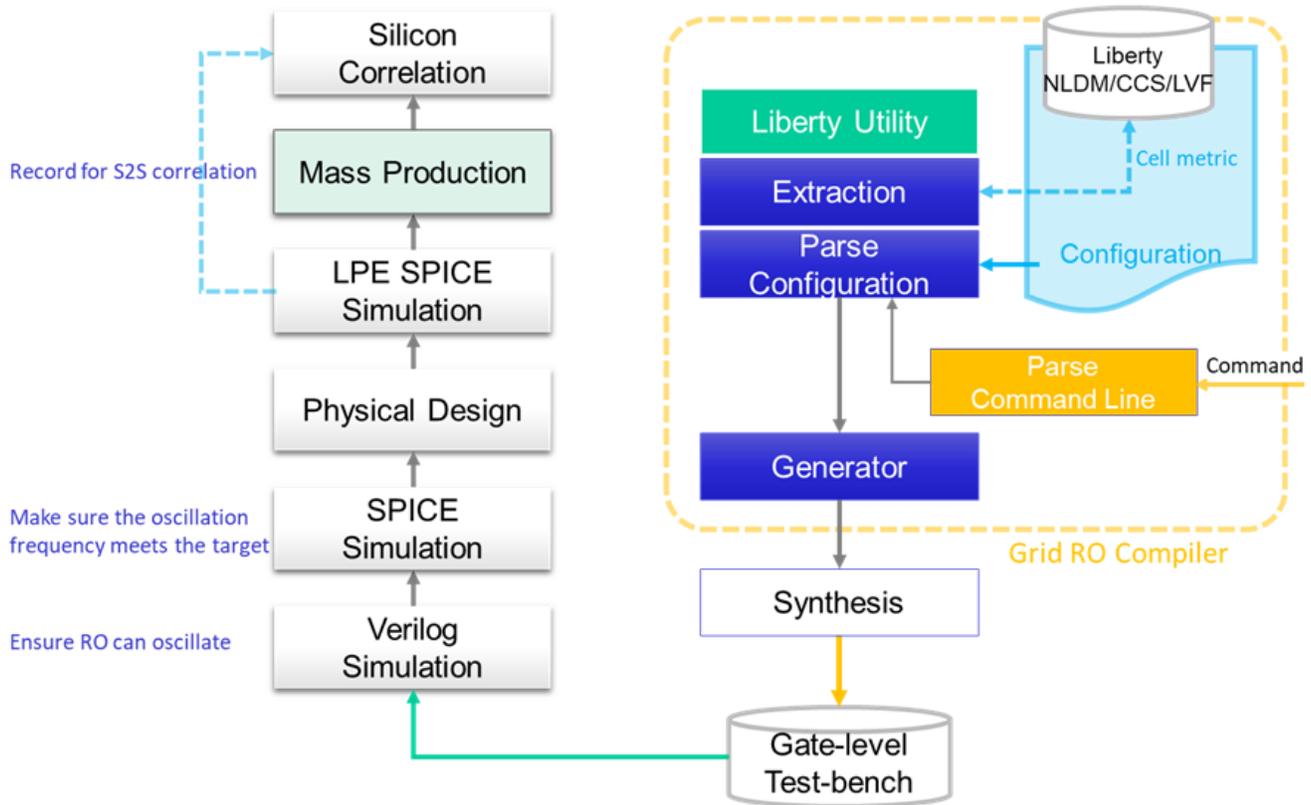


Fig. 6-18 GRO Verification and S2S Flow

第7章 數據分析與機器學習平台 (Copernic™)

7.1. 數據標準化與可視化實踐

DTCO.ML 是一個創新的 EDA 生態系統 (<https://github.com/dipsci/DTCO/tree/main/copernic>)，旨在提升晶片生產與測試過程中各階段的數據關聯性分析與視覺化能力。由於業界目前在晶片生產與測試過程中缺乏統一的數據格式與交互標準，我們需要預先定義一套可擴展的格式並開發相應的轉換工具，實現跨領域數據間的映射與串接。這不僅能提高數據處理效率，還能確保來自不同來源的測試數據能順利整合，進而支援後續的數據視覺化、機器學習及神經網絡等方法，發掘數據間的深層關聯，從而優化晶片的能效與產能，並實現更精確的製程與設計決策，如圖 Fig. 7-1 所示。

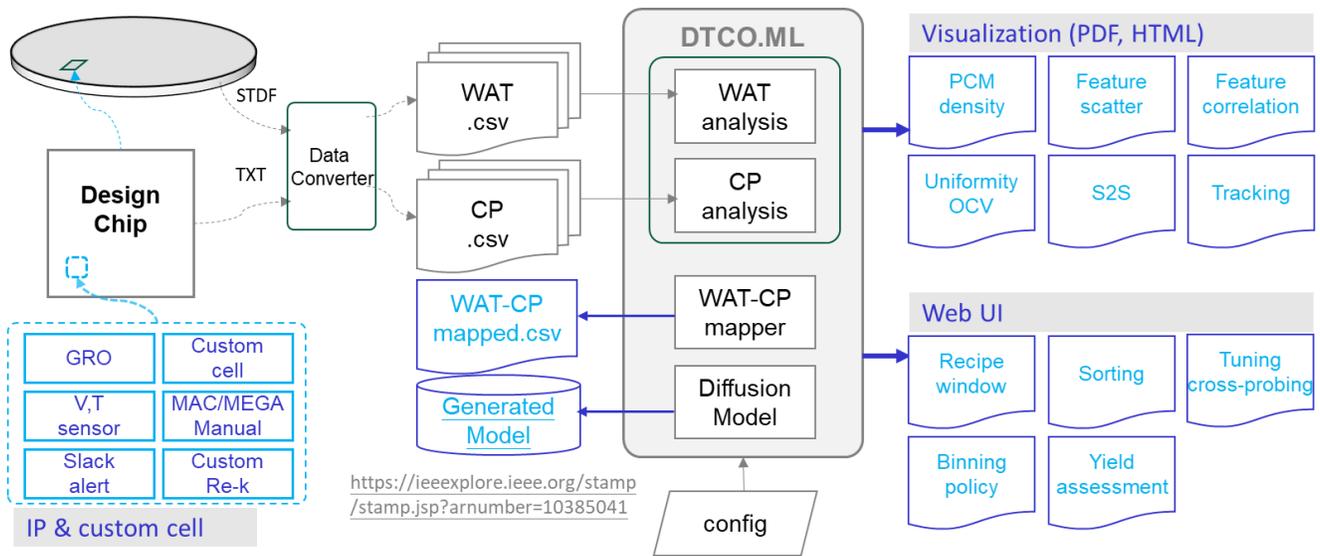


Fig. 7-1 DTCO.ML™ and Copernic™ Platform

為確保高質量測試數據，設計階段需進行前置規劃，例如整合 PVT Sensor，或透過 RO (Ring Oscillator) 搭載訂製電路 IP，無縫融入設計流程。此方式不僅能在量產過程中持續獲取製程與電路性能數據，還能为後續製程優化、时序分析与良率提升堅實基礎。

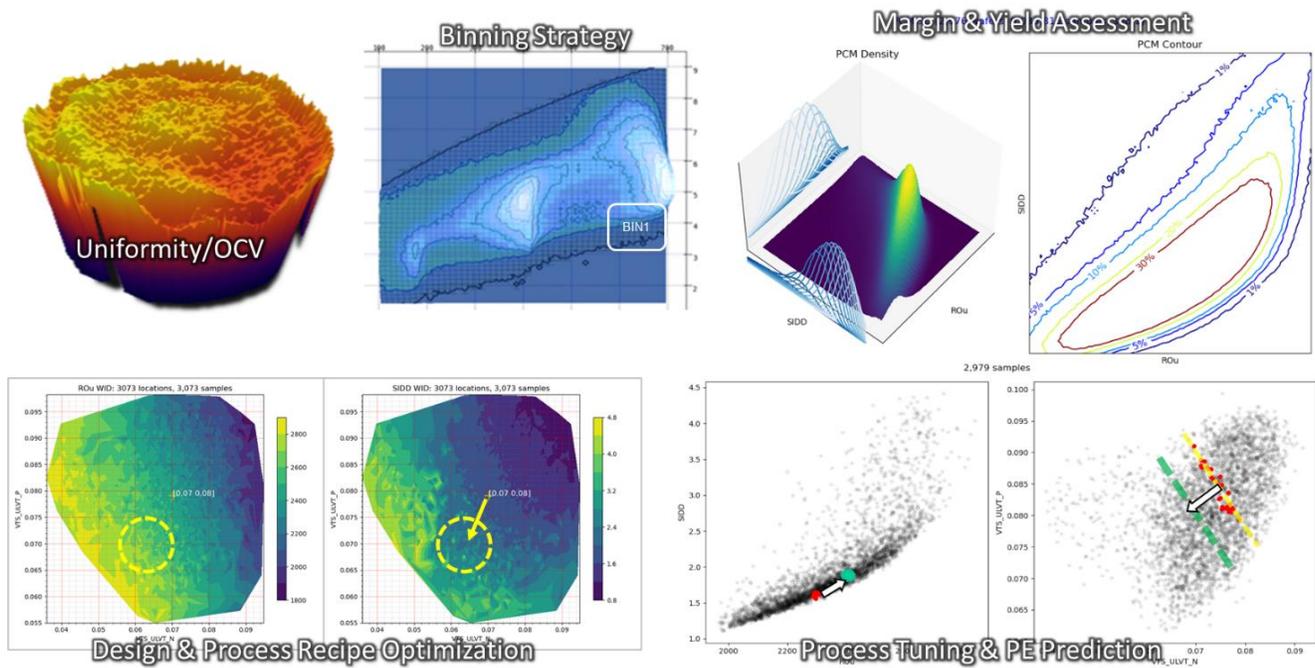


Fig. 7-2 Empowering DTCO.ML Applications

DTCO.ML/Copernic 平台提供直觀的可視化工具，用於追蹤與分析晶圓生產過程中的測試數據分佈與均勻性。平台通過統計與回歸分析來估算晶片內部的局部變異 (OCV)，優化設計餘量以減少過度設計。藉由跨資料維度的關聯分析，確定最佳設計參數與策略，並提供晶片性能分箱與電壓補償指標，提升良率與可靠性。數據可視化展示各物理特性分佈與趨勢，幫助不同領域的工程師更好地理解並迅速識別最佳製程參數範圍，如圖 Fig. 7-2 所示。

7.2. 多維資料的跨域映射

通過 Lot-Wafer-ID 實現跨資料維度的映射與標準化，WAT (Wafer Acceptance Test) 與 CP (Chip Performance) 數據，能揭示晶圓測試與晶片性能的深層關聯，如圖 Fig. 7-3。這種方法不僅能辨識不同製程條件下晶圓的表現差異，還能精確定位對晶片性能影響最大的製程變量，從而精準調整設計與製程參數，優化晶片品質與良率，並為製程改進提供關鍵數據支持。

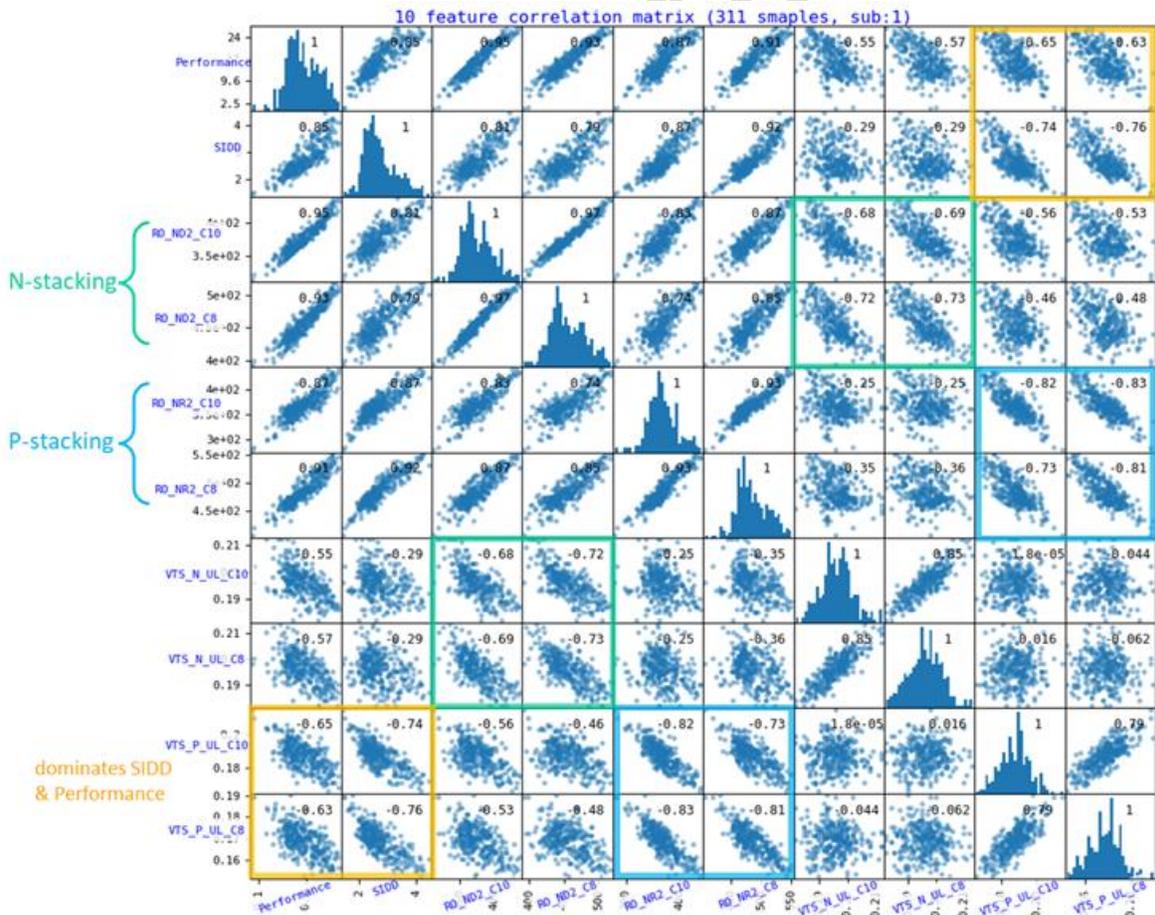


Fig. 7-3 WAT-CP Mapping and Correlation

7.3. 設計流程整合策略

DTCO.ML 設計流程包括 SPICE-Silicon 晶圓相關分析、製程變異理解、庫單元優化以及基於實際 WAT 參數分佈的元件庫時序 Re-K。這使我們能夠實施有效的防禦策略，提升晶片的產能和可靠性。此外，透過 WAT 和 CP/FT 映射、相關分析和建模，可以迅速識別最佳製程配方並完善分箱策略。

7.3.1. WAT-aware Timing Re-K

在當前製程技術中，真實製程參數與 SPICE 模型常存在偏差，如圖 Fig. 7-4 所示。特別是在非標準工作電壓下，關鍵時鐘元件的驅動能力往往會出現嚴重歪斜。僅依賴在較低或較高電壓條件下重新萃取元件時序 (Timing Re-K) 進行的實體設計，最終可能導致功能失敗。因此，分析 WAT 與 SPICE 模型之間的偏差對提取標準單元庫的時序特徵至關重要。基於兩者的相關分析，我們調整仿真校準、時序餘量及防禦策略，並根據晶片實際工作場景進行動態調整。掌握量產分佈趨勢並制定防守策略，將有助於提升產能與競爭力。

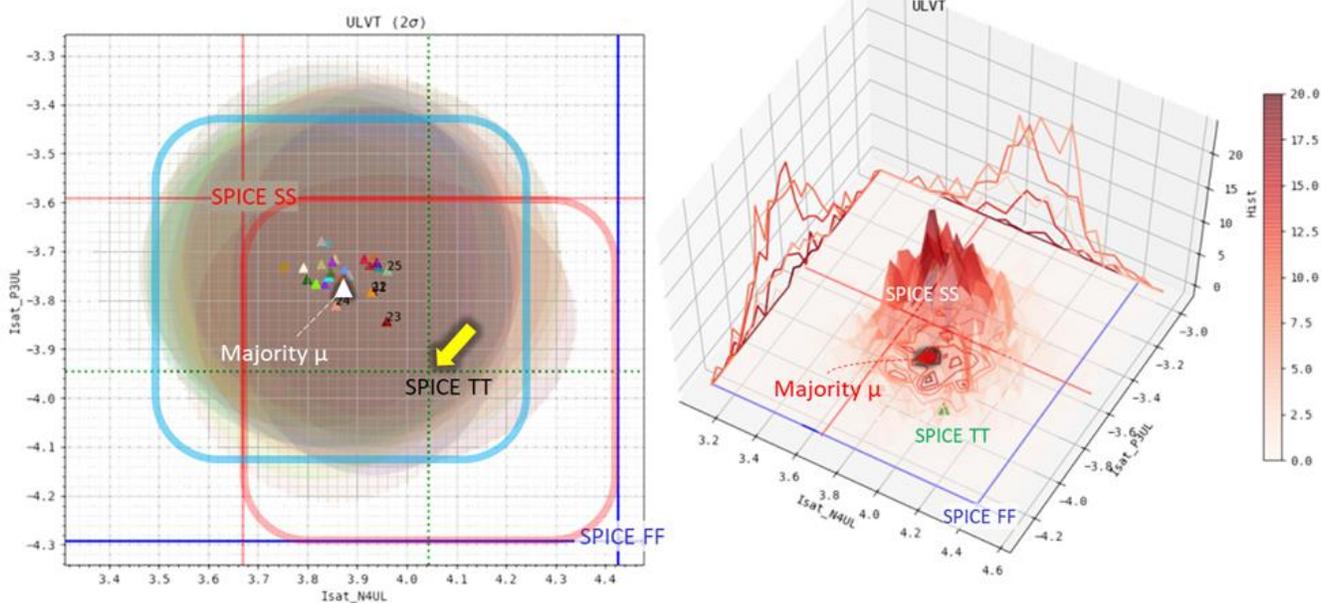


Fig. 7-4 WAT-aware SPICE-to-Silicon Correlation

7.3.2. WAT-CP 映射與相關性分析

WAT 的主要挑戰在於數據樣本不足，可能導致偏差結論。為改善結果，我們可透過特徵參數曲面回歸擴充數據量，並建立 WAT 與 CP/FT/SLT 測試數據的關聯分析。以 WAT 的 N/P 製程參數（如 V_{sat} 、 I_{sat} ）作為 XY 軸，CP/FT/SLT 測試特徵（如 RO、漏電流、良率、 F_{max} ）作為 Z 軸，構建特徵曲面輪廓（contour），直觀呈現 WAT 參數對晶片性能的影響。透過跨資料維度映射與高維數據降維，晶圓廠與晶片設計工程師能精確對齊製程與設計策略，確保優化方向一致，如圖 Fig. 7-5 所示。

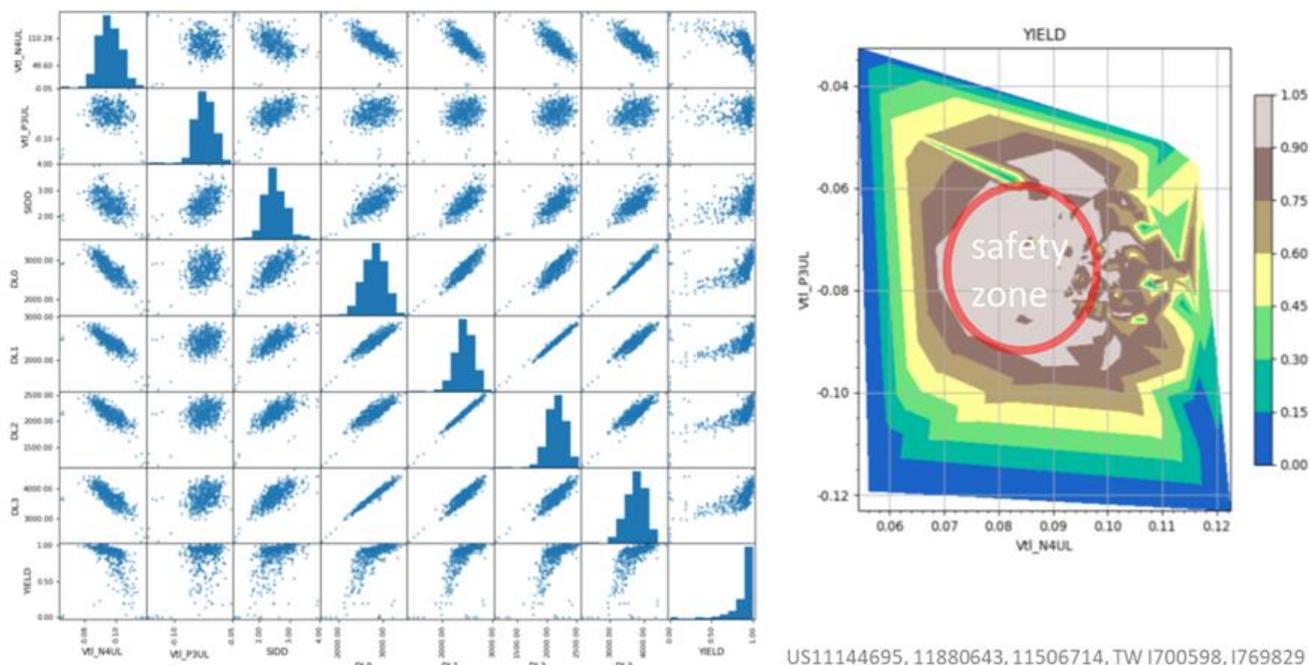


Fig. 7-5 WAT/CP Correlation and Process Window Optimization

7.3.3. OCV Analysis

在實務中，每片晶片只需一顆 RO，便可通過機器學習間接推算晶圓的均勻一致性、晶片內的等效電壓差及 OCV。我們可以在 CP 測試過程中計算 die-to-die 之間的 RO 差異。如圖 Fig. 7-6 所示，選取兩片晶圓上相同座標的 RO，其距離為一個 die，並逐步增加晶片間距離，計算相同座標下的 RO 差異。透過回歸分析，回推至距離為零（D0），即可估算 OCV，進一步評估片上變異性。

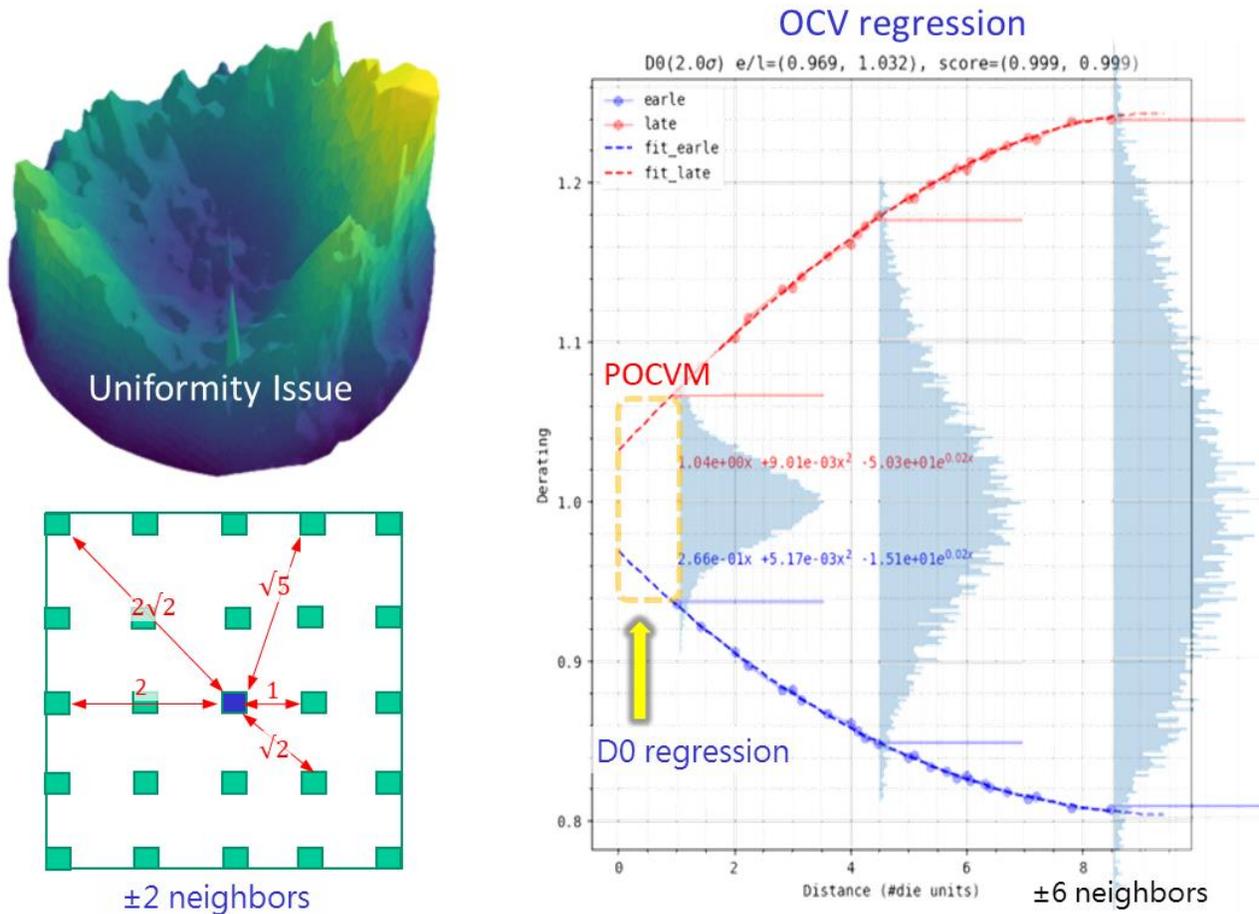


Fig. 7-6 On-chip Sensor and OCV Regression

7.4. 晶片內變異性(OCV) 分析與設計餘量優化

系統層級的「靜態」晶片效能梯度問題可能源於晶圓製造過程中光學、化學與機械效應綜合造成的表面不平坦性。如圖 Fig. 7-7 所示，每顆晶片內均佈有 6 個 Ring Oscillator (RO)，我們利用晶圓上所有晶片的 RO 數據構建特徵曲面，Z 軸對應 RO 頻率。局部放大顯示該區域內四顆相鄰晶片的 RO 頻率分布，其中淺粉色曲面代表 6 個 RO (以不同顏色標示) 頻率的平均值曲面。

該平均值曲面展現效能梯度，揭示晶片內部變異性 (On-chip Variation, OCV)，類似於大範圍的波動，影響晶片效能的均勻性 (此類非隨機變異無法相互抵消)。同時，曲面上疊加來自不同位置 RO 的局部波動，這些波動具有隨機性，進一步加劇了局部效能變異。

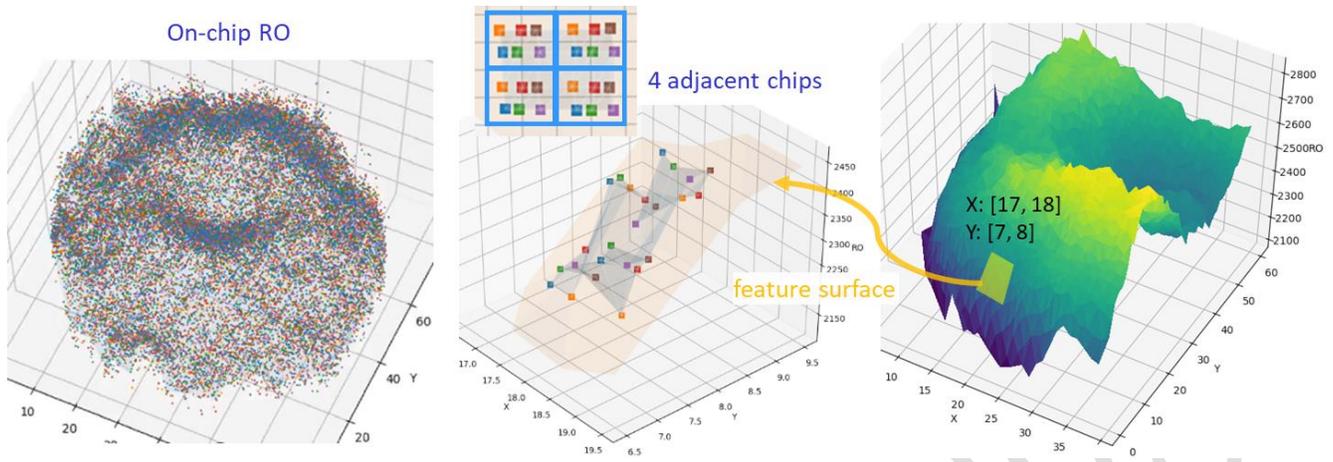


Fig. 7-7 RO Uniformity and and OCV Analysis

上述非隨機性會導致傳統的時序簽核工具失真，如圖 Fig. 7-8。局部電壓梯度效應需有效控制，避免過度消耗資源處理傳統全局 PVT corner。除非晶圓廠顯著改善製程均勻性，或晶片設計業者提供高成本的局部電壓源（LDO array），否則無法完全消除局域的電壓梯度。結合系統級 Binning 與電壓補償方案，建議適度放寬防守邊界（如 1.65 sigma），以避免過度保守的設計餘量影響能效與競爭力。

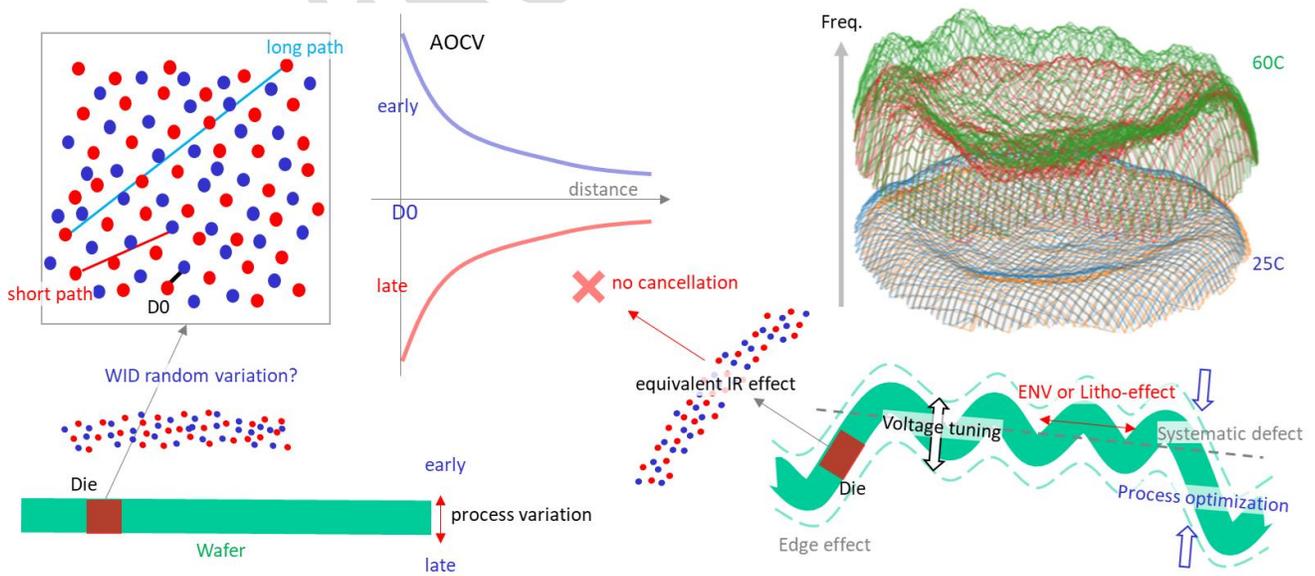


Fig. 7-8 OCV Derating and Cancellation

7.5. 後矽製程分析與優化 (Post-Silicon Analysis and Optimization)

透過 CP 的 XY 座標重建晶圓級物理參數的曲面，如圖 Fig. 7-9，我們通常會觀察到三種顯著的偏移和非隨機變異性：

粗粒諧波震盪 (Coarse-grained Harmonic)：晶圓物理或電性數據（如 SIDD）呈現多個低頻弦波組成的甜甜圈狀圖案，並伴隨沉積不均的傾斜現象。某些拋光步驟也會形成典型的 Polish Pattern，進一步加劇不均勻性，反映特定製程步驟導致的系統級缺陷，如圖 Fig. 7-10。

測試環境 (Test Environment)：類比元件如 V/T Sensor 內建抑制 PVT 變異的設計，凸顯出測試環境引起的偏差，如 Load Board 或 Probe Card 電路阻值不匹配，形成明顯的 Test Site Pattern。然而，由於 Leakage Uniformity 梯度較大，PVT 變異往往未能完全消除。RO 量測結果則融合多種系統級不平坦性，並與低頻諧波震盪疊加，呈現明顯的非隨機性，普遍存在於業界各種生產測試中，如圖 Fig. 7-11。

光刻效應 (Litho-effect)：分析物理位置與數據的關聯性顯示，光刻效應或測試版電阻不匹配不僅影響電性，還對元件與電路延遲 (Delay) 產生顯著影響。光罩曝光顯影 (Shot) 邊界經常出現規律的波峰狀 SIDD，或對應於元件延遲上的規律波谷。

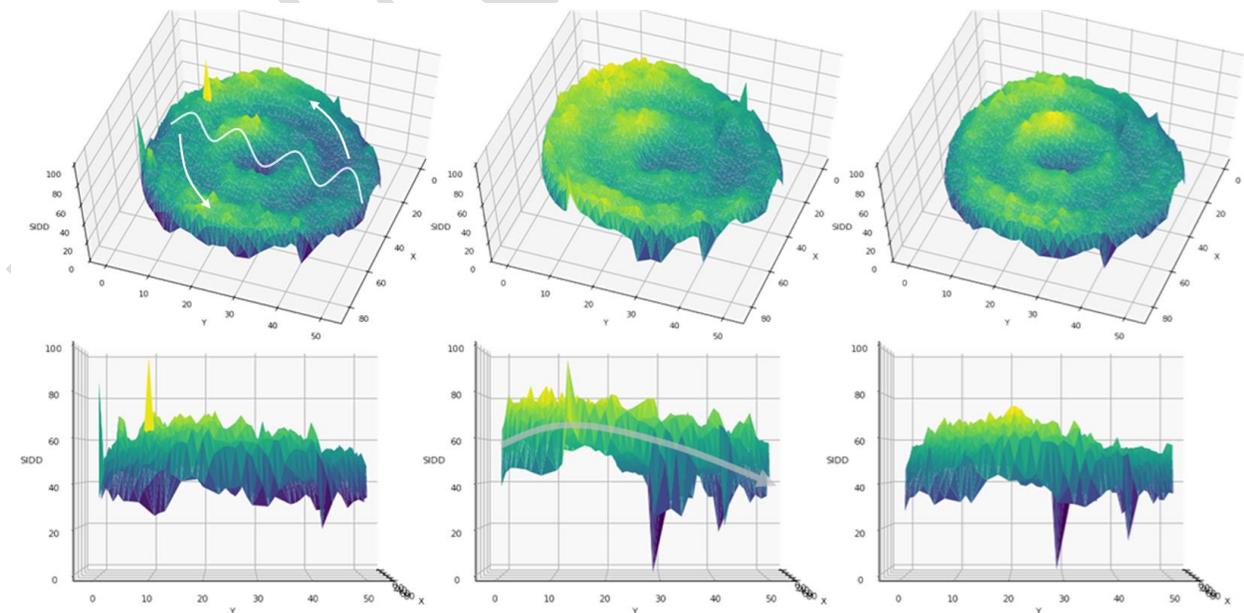


Fig. 7-9 Wafer-level Feature Surface

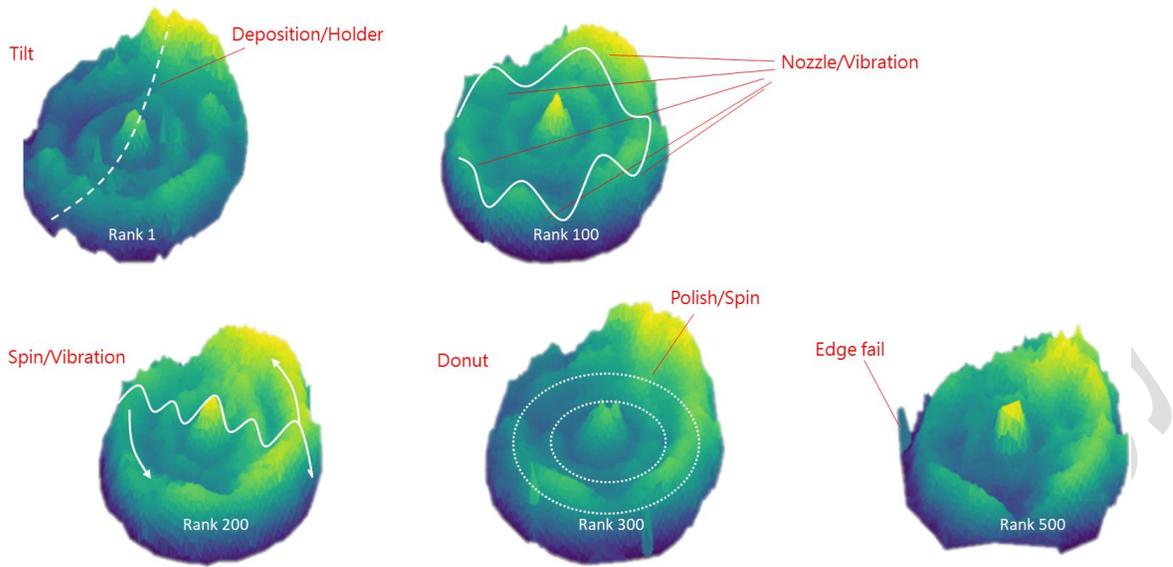


Fig. 7-10 Concentric Ripple Patterns

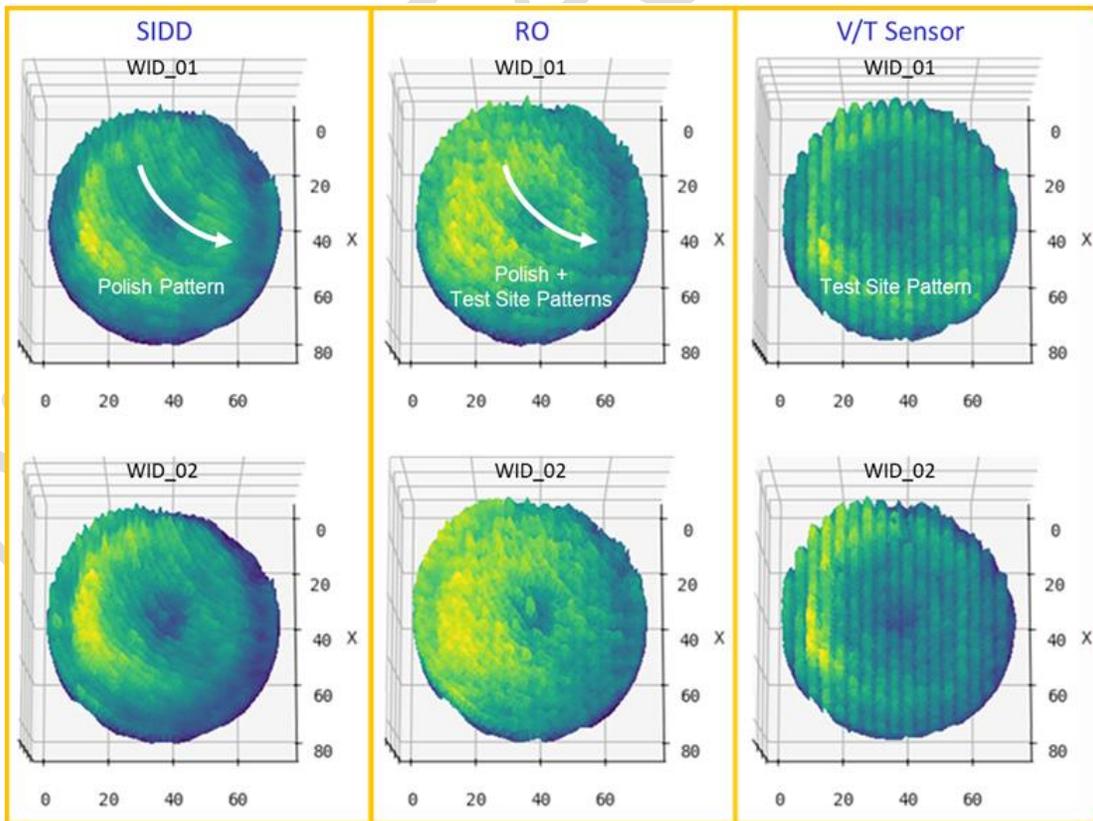


Fig. 7-11 Polish and Test Site Patterns

第8章 晶片性能評等策略與優化 (Binning-PG™)

8.1. 晶片評等與分箱策略對生產力的影響

分箱策略 (Binning Strategy) 在半導體產業中主要用來根據晶片的性能特徵將其分類，以確保晶片在不同工作條件下的穩定性與可靠性。這一策略對生產力的影響可以從以下幾個方面進行分析：

提高良品率與測試效率：精確的分箱策略能有效提高良品率，將晶片根據性能特徵進行分類與補償，並減少不合格品的損耗。同時，通過針對性測試，降低測試時間與成本。

優化生產線效率與穩定性：分箱策略有助於根據不同晶片的性能需求調整生產流程，從而提升生產過程的效率和穩定性，減少生產中斷。

提升產品質量與競爭力：精確的分類能確保每批晶片符合不同的性能標準，提高產品質量 and 一致性，進而提升可靠度和市場競爭力。同時，這種分類方式也能用於行銷策略，根據不同性能層級設定不同價格區間 (Price Bin)，從而最大化市場收益和盈利能力。

在評估多核心機器 (例如虛擬貨幣 BTC 挖礦機 [2]) 出貨是否達到最佳化時，可以利用 Binning 方法將機器根據計算性能進行分類。每個批次的機器經過分類後，將相似性能的機器歸為同一類，並根據每台機器的計算力 (Tera-hash rate, TH) 進行裝機測試，計算出每台機器的總計算能力。隨著出貨量的增加，所有機器的計算力將累積，形成累積計算力 (cumulate TH)。這一指標可以用來評估不同 Binning 方法 (Policy) 的效果，幫助確定哪種方法能夠更有效地分配機器性能，從而達到更高的總計算力。

如圖 Fig. 8-1 透過比較不同 Binning 方法所得到的累積計算力 (cumulate TH)，我們能夠衡量各種核心分配策略的效能，並選擇最佳的策略以達成機器出貨的最佳化。最佳化的目標是使每個批次的出貨計算力更為均衡，減少性能波動，從而提高整體的累積計算力，滿足市場需求並提升出貨效率。最終，這些方法將提升產品整體性能，實現極致能效，並最大化計算力與生產力。

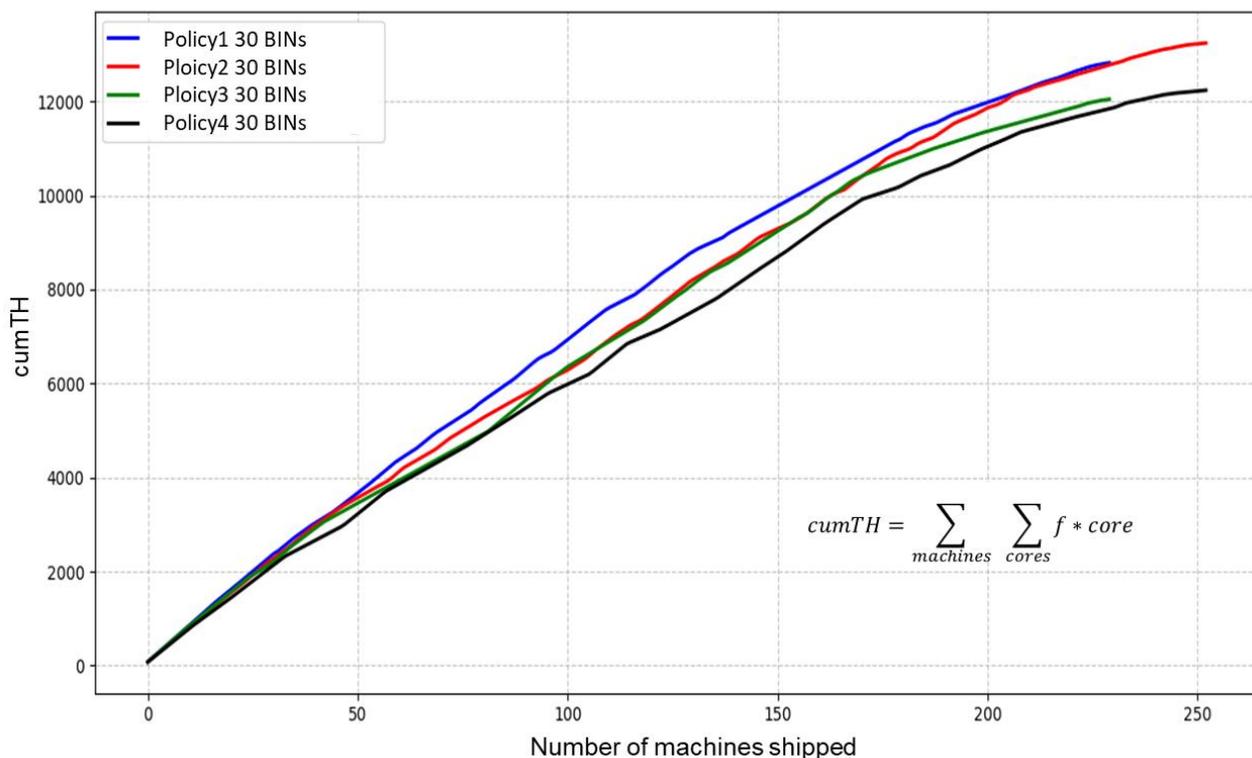


Fig. 8-1 Multi-core Machine Shipment and Cumulative Computing Power Evaluation

8.2. 晶片體質的分析與挑戰

傳統 Binning 方法通常使用類似於晶片性能（例如 Fmax）與 SIDD 漏電流的二維流水線分箱策略，如圖 Fig. 8-2。例如，先對 SIDD 進行粗粒度的晶片體質分割，再根據功能模式（Function Patterns）所能達到的性能進行細粒度分割。這種方法依賴大量工程師的時間進行測試程式編碼與反覆試驗，以確定分割條件及邊界值。然而，這種分群方式無法保證晶片在電壓或溫度敏感性上的一致性。如圖右側所示，即使晶片被歸為同一組，它們在其他維度指標上的表現可能存在顯著差異。

實際上影響晶片性能的參數通常不僅限於二維或三維，如圖 Fig. 8-3 所示。相同的晶片在不同電壓與溫度條件下，其性能曲面往往不是簡單直觀的。例如，在固定電壓下，某些工作頻率的晶片表現可能相似（如圖中粉紅色的割線，V-F 曲線顯示相似的性能趨勢），但當改變溫度條件時，這些晶片的性能可能會出現急劇的、懸崖式的變化。

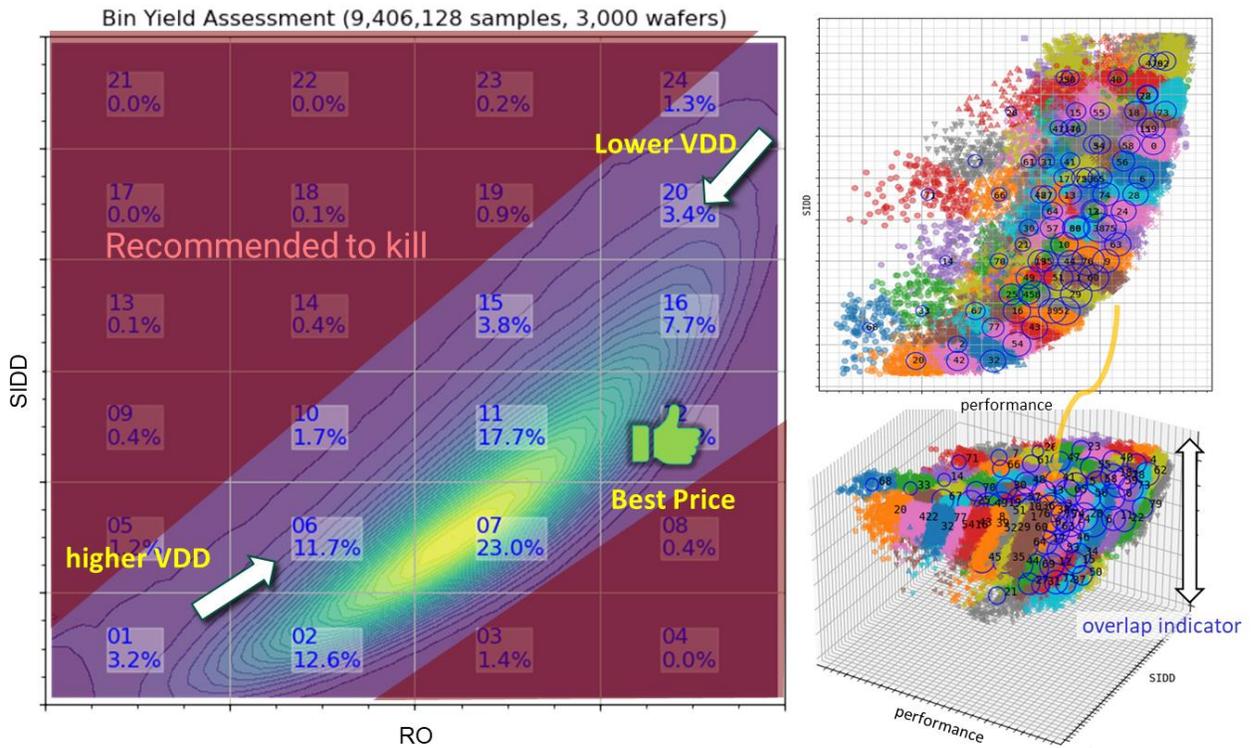


Fig. 8-2 Traditional 2D Pipeline Binning and Potential Issues

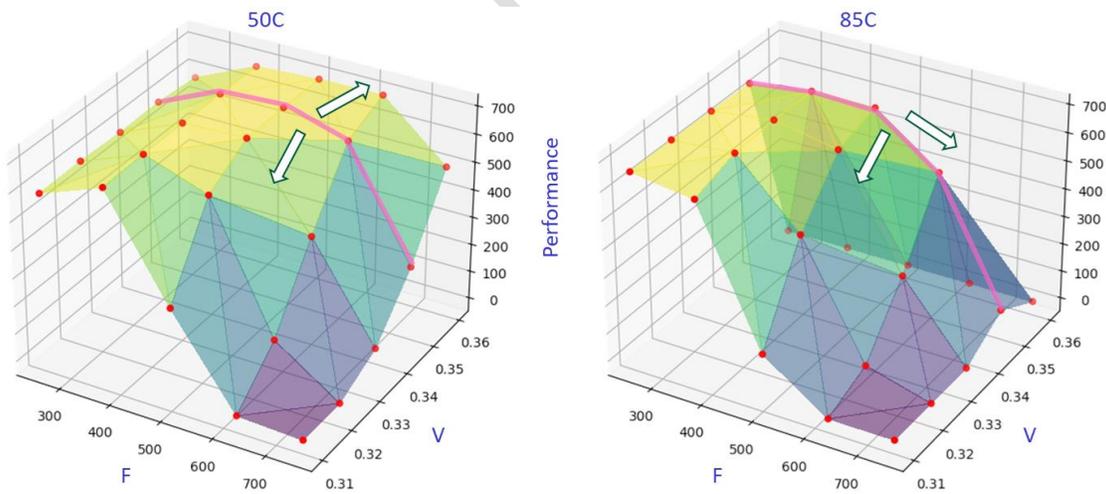


Fig. 8-3 V-T Sweep and Performance Analysis

為了克服晶片性能分群中的高維資料挑戰，K-means 聚類演算法經常被用於將數據分為預定數量的群組，並通過最小化群組內變異來進行分類。然而，在整機層面 (machine-level) 的實際運行中，該方法存在一些不足，其主要問題包括以下兩點：

特徵權重與重要性忽視：傳統分群演算法，例如 K-means，將所有維度視為同等重要，無法自動識別不同特徵對分群的重要性差異。這需要人工經驗的介入，通過特徵擴增 (feature augmentation) 或對各特徵進行權重調整等特徵工程，才能使聚類結果更具鑑別性。這種依賴人工的過程不僅增加了工作量，也可能導致分群準確性下降。如圖 Fig. 8-4 基於 6 維資料，透過權重的調整使 K-means 在處理晶片特徵時，有效考慮 SIDD 權重變化的影響。

多維特徵的離散性問題：晶片性能數據中的多維特徵往往是離散的，並且可能形成多個截然不同的高維流形。在這種情況下，不同流形的積分值可能在離散集合上相同，導致高維數據無法被有效區分，進而缺乏鑑別度。這種情況在處理多維離散特徵的數據時尤其顯著。例如 K-means 假設所有群組內的數據是相似的，因此在面對晶片中微小差異或異常時，它無法精確識別，尤其是在處理結構上或對 PVT 敏感性有細微差異的晶片時。

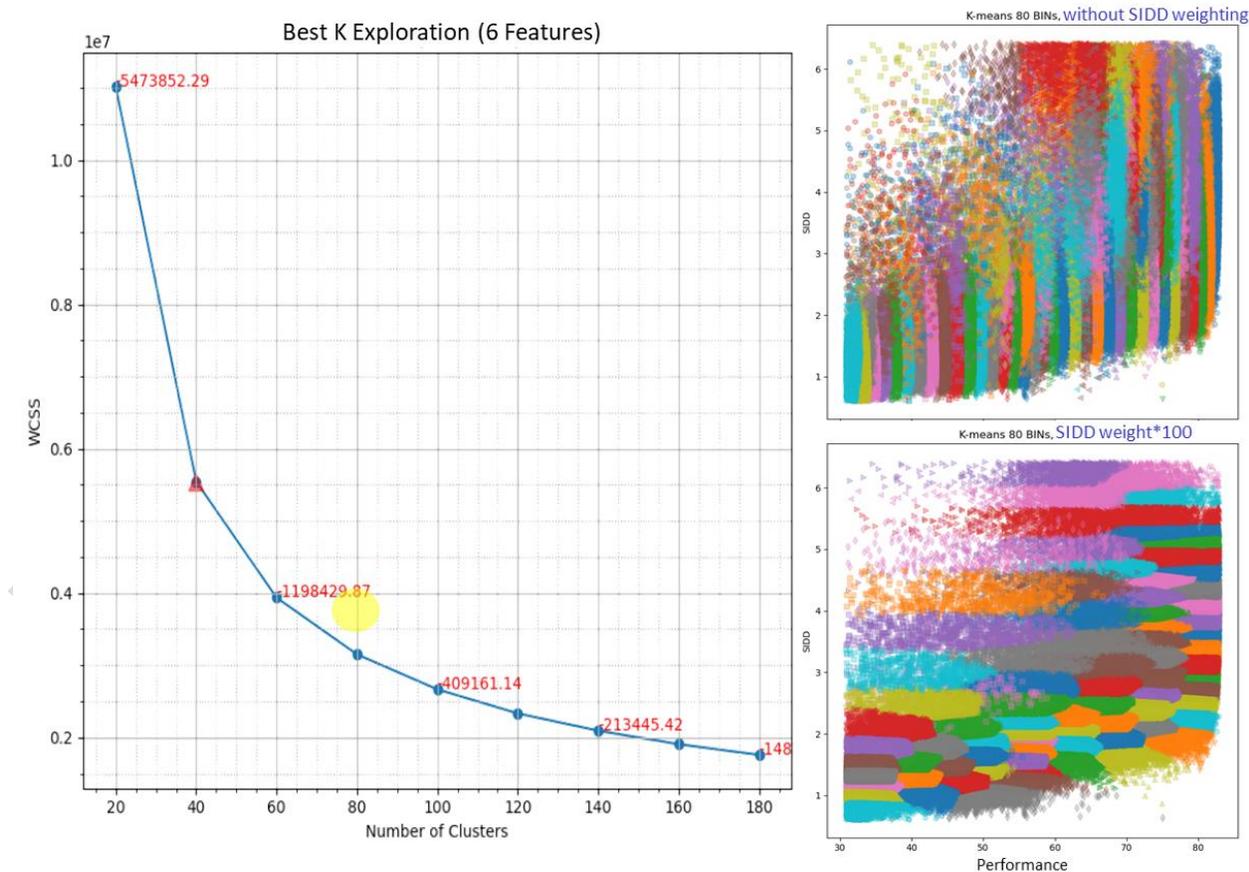


Fig. 8-4 Best-K Exploration

傳統方法中，我們通常在固定電壓下選用一組頻率與性能 (Performance) 的配置作為 K-means 分群的特徵 (如圖 Fig. 8-3 的粉紅色割線)。然而，這種方式在整機層面上的區分效果並不理想，尤其在處理高維離散特徵時表現不足。為了改善這一問題，我們可以利用現有離散網格點的方法來構建插值曲面，如圖 Fig. 8-5。通過插值曲面提升網格點的解析度，使得特徵空間的連續性與表徵能力顯著增強。同時，對 SIDD 特徵進行適當的權重調整，提升其在分群過程中的影響力。這種方式有效解決了高維離散特徵處理中的辨識度不足問題。

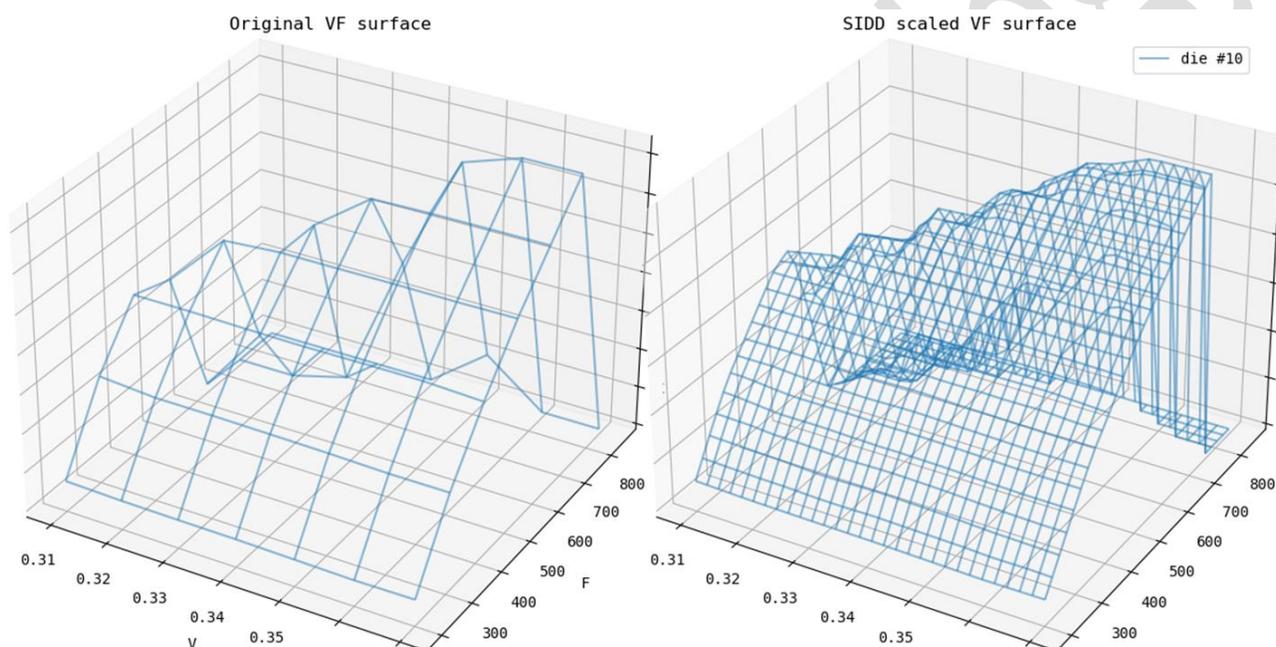


Fig. 8-5 V-F Grid Interpolation

在實務應用中，我們可以透過特徵擴增技術 (例如插值)，將原本僅有 6 維的 CP 測試項特徵空間 (如 RO、SIDD 及 4 個離散的性能指標) 擴展至數十維進行分群，同時確保晶片性能評估的準確性。此改進顯著提升了分群效果，為整機層面的性能區分提供更可靠的基礎。

8.3. 分箱策略生成 (Binning Policy Generation, Binning-PG™)

Binning 策略生成是一種系統化的過程，旨在根據晶片的性能特徵將其合理分類，以確保每個晶片在不同工作條件下的表現都能符合預期。在生成有效的分箱策略時，通常採取以下步驟：

數據收集與預處理：首先收集來自不同晶片的多維性能數據（如工作頻率、功率消耗、SIDD 漏電流等），並對數據進行清洗、去除異常值和標準化處理，確保數據的一致性與可靠性。

特徵提取與聚類分析：分析數據中的關鍵特徵，並使用聚類算法（如 K-means、DBSCAN 等）將晶片根據其性能特徵劃分為不同組，確定潛在的性能模式和分箱邊界。

策略調整與優化：根據生產需求和性能標準，對初步分群結果進行優化，調整每個分箱的範圍，確保每個分箱內的晶片符合所要求的性能目標。

策略驗證與實施：進行小規模生產測試或模擬驗證，根據結果進行調整，並在驗證成功後將策略應用於大規模生產，確保其穩定性與有效性。

在 Binning 策略的初期，通常依賴初始的 WAT、CP/FT 等特徵進行非監督式學習分群。這樣的分群目的在於將相同體質且對 PVT 反應一致的晶片歸為同一類。此過程屬於非監督學習（non-supervised learning），並且可能會結合部分人為介入（Human-in-the-Loop, HITL）的主動學習（active learning），以更精確地進行分群。在策略驗證與實施的過程中，隨著量產過程中積累的數據與觀察，這些初步的分群結果會逐漸轉化為真實標籤（ground truth），並最終為監督式學習（supervised learning）提供支持，從而進一步優化 Binning 策略。

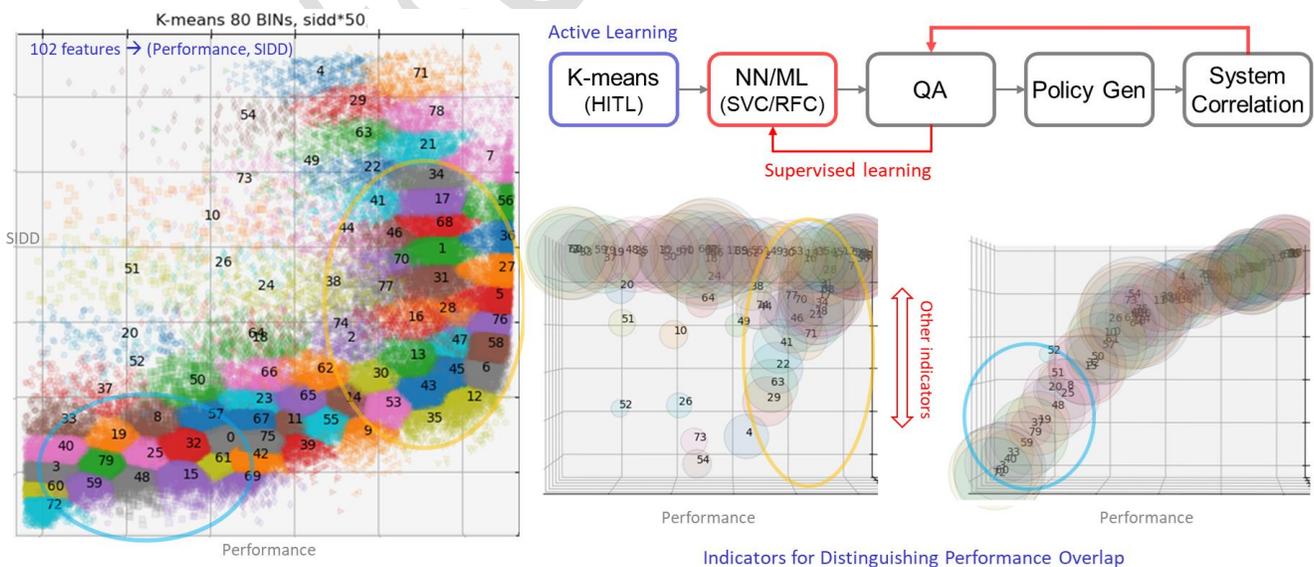


Fig. 8-6 Binning Policy Generation

8.4. 策略生成自動化與優化

Binning 策略自動化的主要目標是通過精確的數據分析和算法應用，實現晶片分類過程的自動化與高效化，從而提高生產力和產品質量。具體目標包括：

性能評估與資格指標確立：通過綜合評估指標（如 SIDD、 $\sum F_{max}$ 、 $cumHash = \sum F_{max} * pass_core$ 、SIDD-std、 $cumHash-std$ ），對每個晶片的性能進行量化分析，確保分箱過程能準確地反映晶片在不同條件下的表現，並建立清晰的合格標準，如圖 Fig. 8-7。

使用分群演算法與優化：利用混和 K-means 聚類算法自動化分群過程，根據晶片性能的相似性進行高效分類，實現更好的群內同質性。這不僅提升了分箱的精確度，還能減少人工干預，提高整體流程的自動化和穩定性。

建立有效的工作流方法論：根據成功項目的經驗，設計 CP/FT 測試建議流程，為後續項目提供一致的分箱策略指導。這能幫助在不同項目間快速適應並應用最佳的分箱策略，實現長期的生產效率提升。

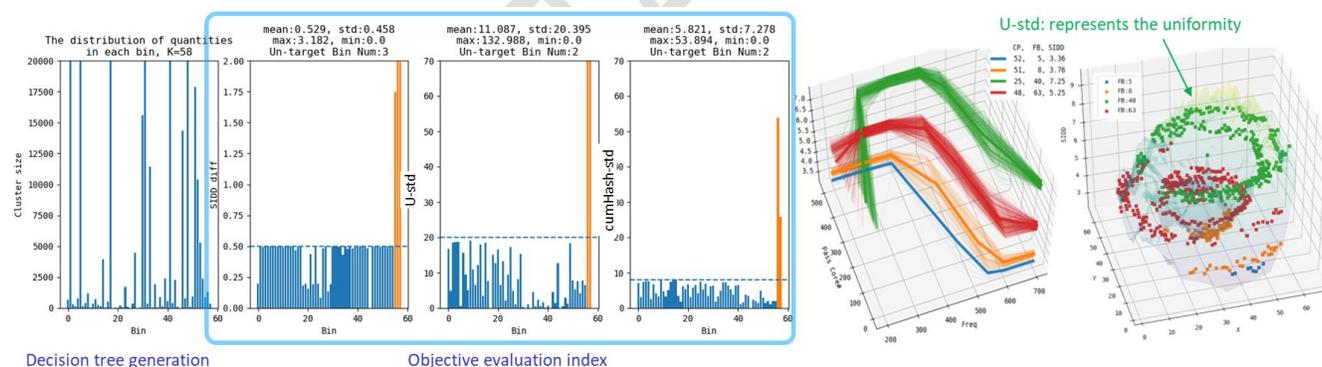


Fig. 8-7 Key indicators for Chip Classification

如圖 Fig. 8-8 所示，為提升晶片分群準確性與效率，採用了結合 K-means 聚類與人類介入（Human-in-the-Loop, HITL）的混合策略。首先，透過 HITL 技術進行特徵擴增與工程（如權重調整、插值曲面構建），生成細粒度的分群結果（Soft Bin），並作為基準（Ground Truth）支持主動學習（Active Learning）。此步驟在初期階段靈活調整分群標準，為後續模型訓練奠定基礎。

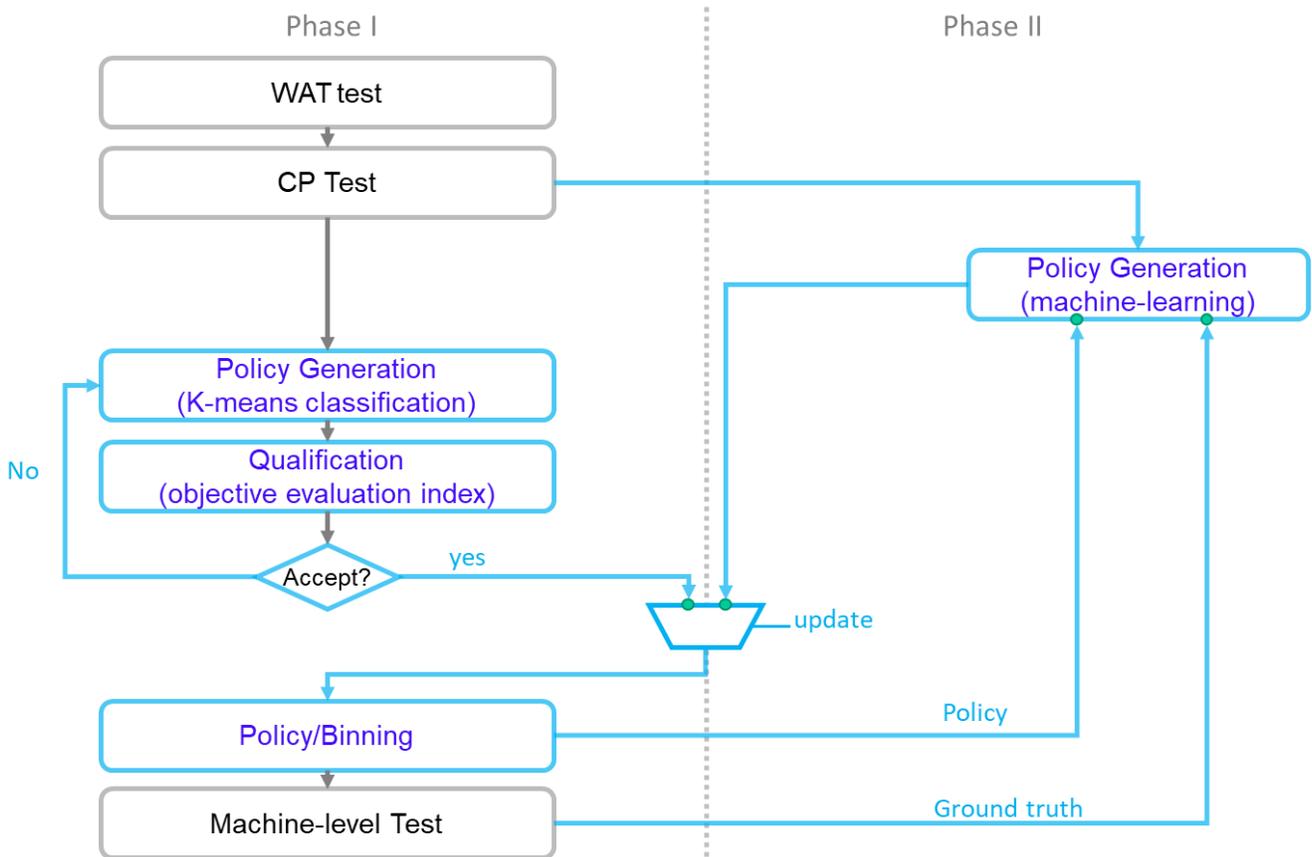


Fig. 8-8 Binning Policy Optimization Flow

接著，利用 **Ground Truth** 數據，採用監督式學習技術（如 SVM、隨機森林、神經網絡）訓練分類器，生成精確的分箱策略。此過程的目標是優化分群結果，減少 **Bin** 數量，提升分群效能與市場價值，並有效平衡分群準確性與生產成本。這一方法實現了策略設計的自動化與優化，縮短迭代週期並提高晶片製造的靈活性與效率。

8.5. 晶片內自分級應用 (On-chip Self-binning)

晶片內自分級（On-chip Self-Binning, OCSB）是一種創新技術，通過在晶片內集成自我測試與性能評估功能，使晶片能在生產過程中自動分類。該技術提高了分級效率，降低了測試成本，並減少了對外部測試設備的依賴。其核心方法包括：

內建自分級 (Built-in Self-Binning)：晶片內集成自測模組，能在不同電壓、頻率與溫度條件下，自動測試性能參數（如 RO 工作頻率、等效電壓、動態時序餘量等），快速準確獲取性能數據。

降低測試成本與資源需求：基於內建自測數據，晶片能即時分析性能與餘量，並根據預定標準自動確定分級，減少對外部測試設備與人工干預的依賴，顯著提高生產效率與自動化水平。

即時性能評估與動態調適能力：結合即時監測電路與 AI 推論引擎，晶片可根據實時運行環境動態調整分級與補償，提升可靠度與市場競爭力。

OCSB 技術結合神經網路，利用晶片內的記憶單元和 MAC 單元進行性能推論，實現自動化分箱。通過在量產前訓練神經網路學習性能特徵與分箱標準間的非線性關係，並將結果儲存於晶片內部，CP 測試階段可即時生成分箱結果。此技術能精確捕捉多維性能間的關係，縮短測試時間，減少外部設備依賴，提高生產效率，並根據市場需求靈活調整，從而降低成本並顯著提升生產力與產品質量，如圖 Fig. 8-9。

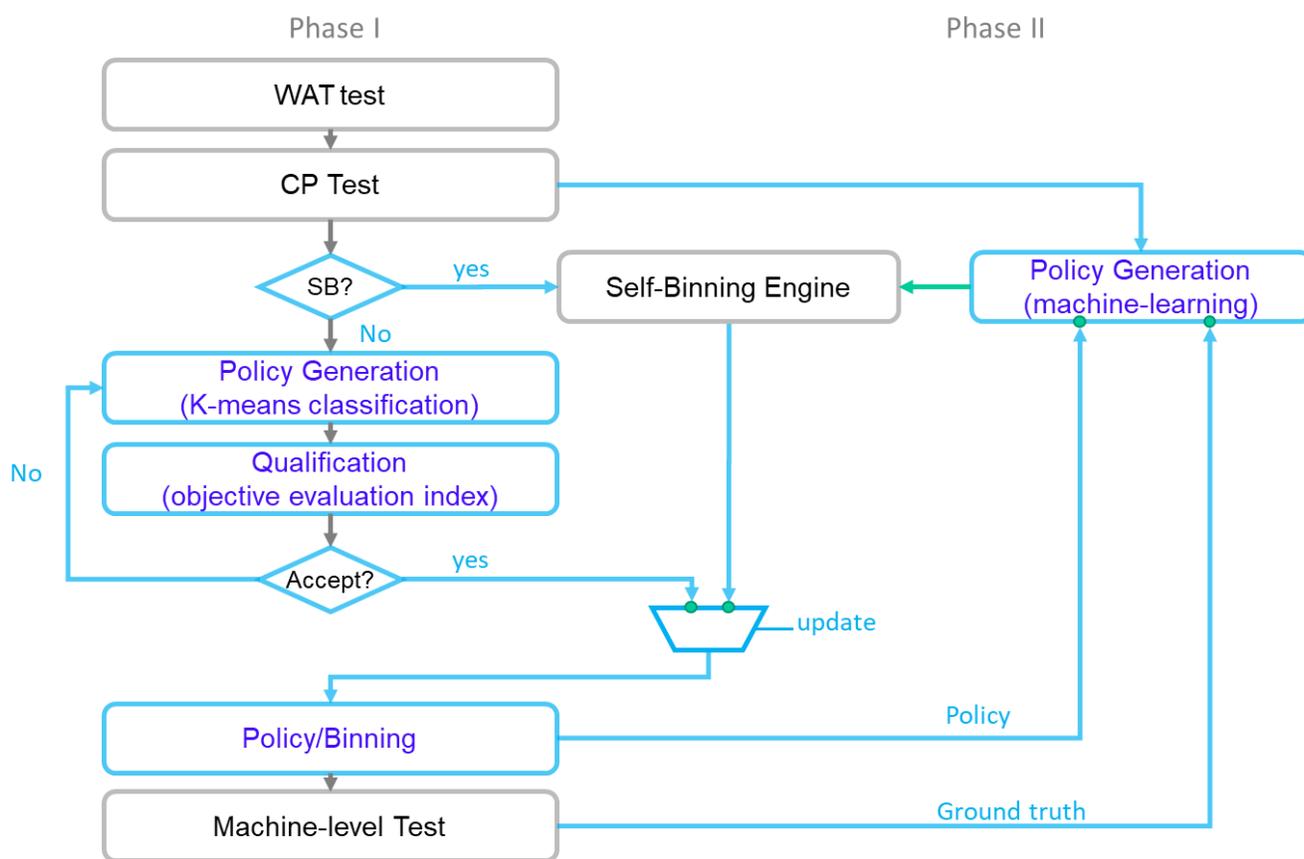


Fig. 8-9 On-chip Self-binning (OCSB)

第三部分：DTCO.GenAI™ - 生成式 AI 驅動的晶片設計創新

本章節將探討生成式 AI (Generative AI) 在半導體與晶片開發領域的應用潛力，涵蓋生成對抗網絡 (GAN)、擴散模型 (Diffusion Model)、高標準差 (high- σ) Monte-Carlo 仿真加速決策及 WAT 超解析度技術的實現與應用 [3][4][5]。

第9章 生成式 AI 與 DTCO 融合 (DTCO.GenAI™)

9.1. 傳統建模方法的局限性分析

傳統的模型在建立時常將事件簡化為高斯分佈，卻忽略了真實晶片中各個物理量的分佈通常呈現偏態 (skewed-normal) 或對數正態分佈 (log-normal)。同時，這些模型也經常忽略了向量在高維空間中的相互關聯性。如圖 Fig. 9-1 所示，當擲骰子時，即使每個維度的特徵在單獨投影時符合母群體分佈，但在高維空間中組合起來的分佈可能會失去彼此之間的關聯性。

真實數據通常集中於高維空間中的低維流形，若假設多維特徵彼此獨立，隨著維度增加，生成的高維數據分佈可能變得不真實。解決此問題的關鍵在於考慮特徵間的相依性，並採用更適合高維建模的技術。

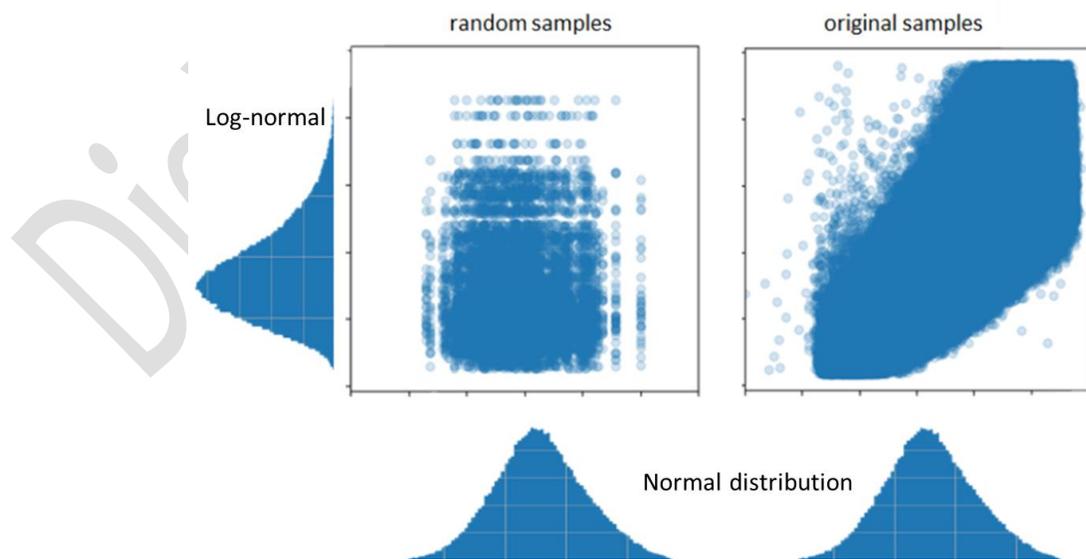


Fig. 9-1 Interdependencies of Physical Quantities

在現實情況中，假設特徵分佈彼此獨立的簡化假設可能無法準確反映晶片製造過程中的各種特性與相互關係。因此，為了更真實地模擬晶片行為，需要考慮非高斯分佈、偏態分佈或對數正態分佈的特徵，並充分考慮高維空間中的相互關聯性。

由於晶圓製造過程中的製程參數眾多，製程參數與晶圓或晶片級測試資料之間的關係變得非常複雜，傳統方法難以進行建模和分析。如圖 Fig. 9-2 所示，即使我們完全了解晶片層級的特徵分佈和相互關係，如果缺少晶片的座標資訊，數據於晶圓層級的分佈也將失去真實製程平坦度 (process uniformity) 的特徵。這是目前模擬分析建模中普遍存在的問題。事實上，特徵向量在高維空間中的分佈缺乏真實性，這導致量產數據與模擬數據之間存在巨大偏差。

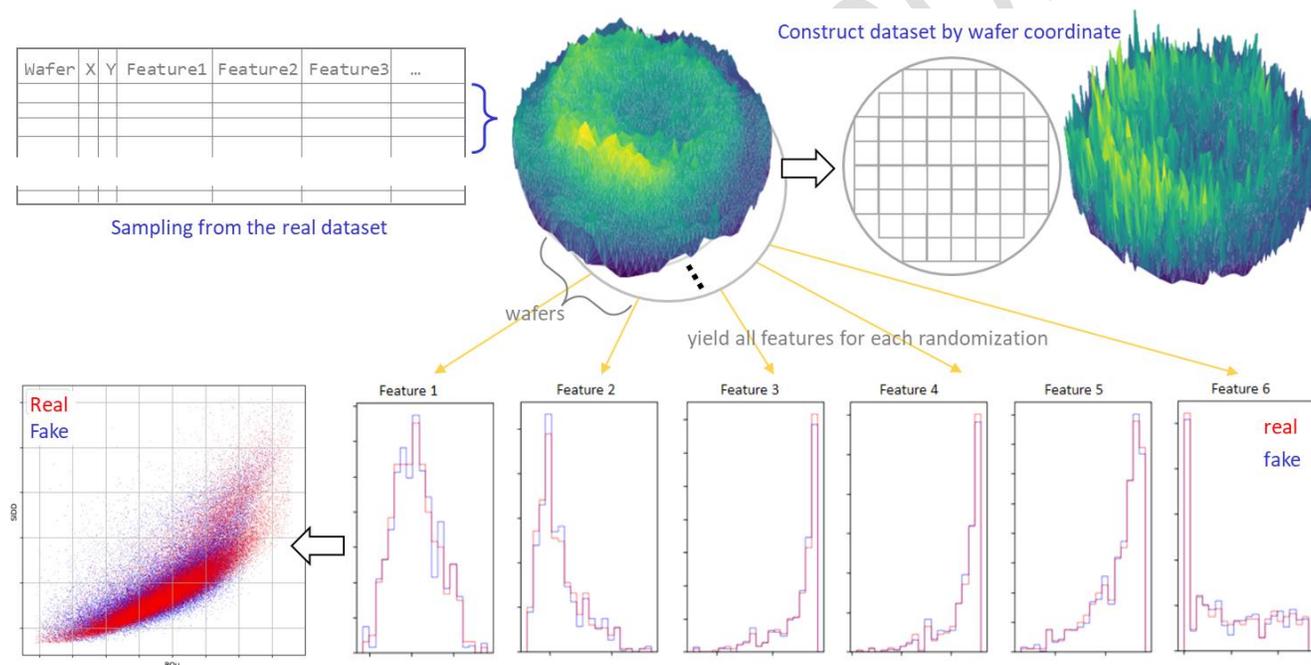


Fig. 9-2 Dissipation of Feature Spatial Continuity

這偏差限制了我們對晶片製造過程的理解和模擬準確性，因此我們需要開發更先進的建模和分析方法，以充分考慮製程參數的複雜性和晶圓座標資訊的重要性。這將有助於更準確地模擬晶片的特性和製程，並提供更可靠的量產數據和模擬結果。

9.2. 按圖索驥：Multivariate Normal Distribution

引入特徵之間的相關性（例如 GMM 的協方差矩陣）可以有效捕捉高維特徵間的相依性，從而生成更真實的資料分佈。當維度提高且模型能夠掌握更多特徵之間的相關性時，生成分佈將更接近真實數據的分佈。

假設我們能取得一個製程對晶片性能的簡單關聯性分析報告，該報告可能是量產前初期數據量還不是很足的樣貌，例如只有 TT 與 skew wafer 在內總測試數量約 420 筆晶片的 6 片晶圓數據。

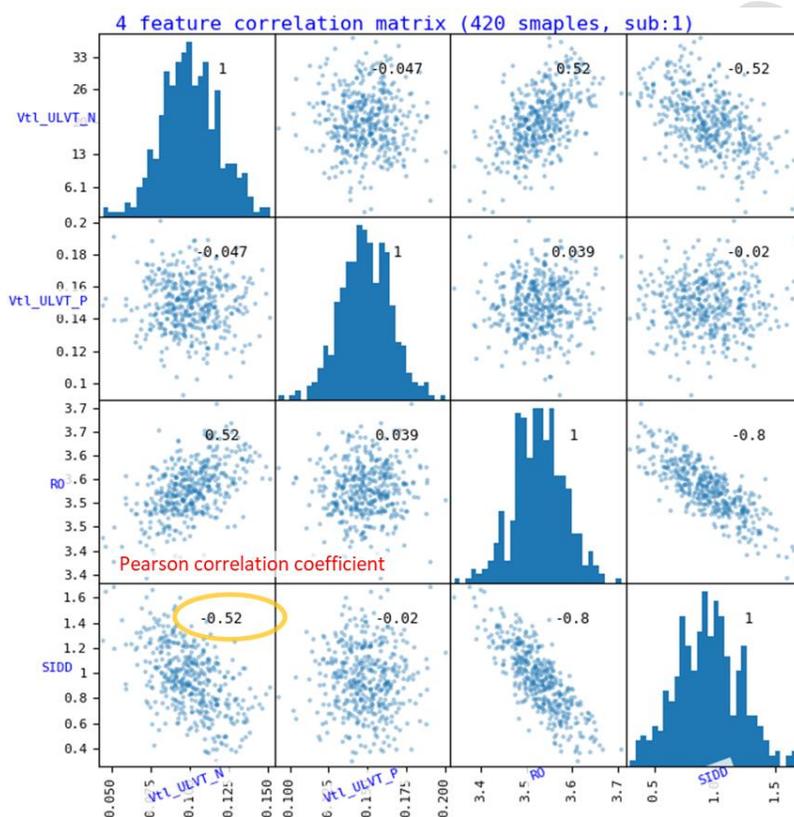


Fig. 9-3 Multivariate Feature Correlation

如圖 Fig. 9-3 所示，這是一個典型的關聯性矩陣，用於將 WAT site ID 與 CP 數據進行映射，其中數字表示 Pearson 關聯係數（Correlation Coefficient）。基於此矩陣，我們可利用多元常態分布（Multivariate Normal Distribution, MVN）快速構建高維空間的近似模型，生成密集數據以增強趨勢判斷、邊界輪廓描述及決策信心。如圖 Fig. 9-4，通過 MVN 隨機生成 10,000 筆符合該關聯矩陣分佈的數據。

Example 9-1 Multivariate Normal Distribution

```

import numpy as np
import pandas as pd

μ = [0.1, 0.15, 3.53, 0.95] # features mean
σ = np.array([0.11, 0.1, 0.37, 1.4])/6 # min-max 6 sigma (Gaussian)
σ1,σ2,σ3,σ4 = σ

# Pearson coefficient based on the scatter matrix
ρ = [np.array([-0.047, 0.52, -0.52]),
     np.array([ 0.039, -0.02]),
     np.array([-0.8])]

(ρ12,ρ13,ρ14), (ρ23,ρ24), (ρ34,) = ρ

# covariance matrix
Σ = [[σ1**2,      ρ12*σ1*σ2, ρ13*σ1*σ3, ρ14*σ1*σ4],
     [ρ12*σ1*σ2, σ2**2,      ρ23*σ2*σ3, ρ24*σ2*σ4],
     [ρ13*σ1*σ3, ρ23*σ2*σ3, σ3**2,      ρ34*σ3*σ4],
     [ρ14*σ1*σ4, ρ24*σ2*σ4, ρ34*σ3*σ4, σ4**2]]

data = np.random.multivariate_normal(μ, Σ, 10000) # generate 10K samples

# convert to Dataframe
dt = pd.DataFrame(data, columns=['Vt1_ULVT_N', 'Vt1_ULVT_P', 'RO', 'SIDD'])

```

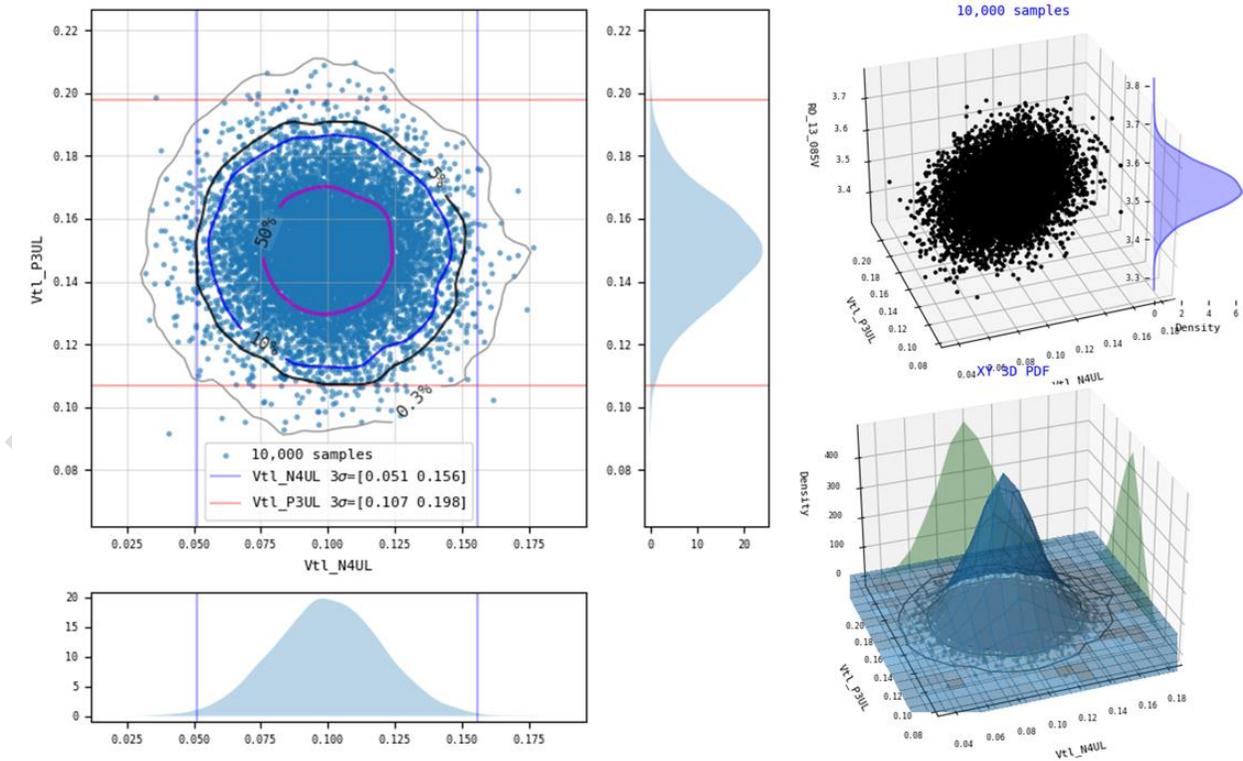


Fig. 9-4 Feature Distribution and CDF Contours

9.3. 虛擬晶片數據在 DTCO 中的實際應用 (DTCO.VS)

設計技術協同優化 (DTCO) 在先進技術節點上日益重要，但設計師與 EDA 工具開發者常缺乏可靠的製程參數與測試數據。在此情況下，虛擬矽晶數據成為關鍵，提供高質量且符合實際分佈的數據，同時避免機密性問題。藉由分析晶片的物理特徵，例如漏電流和不同頻率下的性能表現，設計師能有效識別缺陷根源與優化方向。大量生成的矽數據有助改進晶片性能評估策略，提升產品競爭力並優化晶圓整體性能。

圖 Fig. 9-5 展示了基於生成的大量虛擬晶片數據，通過這些數據的積累，我們能夠提高信任水平，並在逼近真實情況下，發現良率與設計餘量之間的妥協空間，從而為未來的設計方案與產能優化提供具體的指導方針。

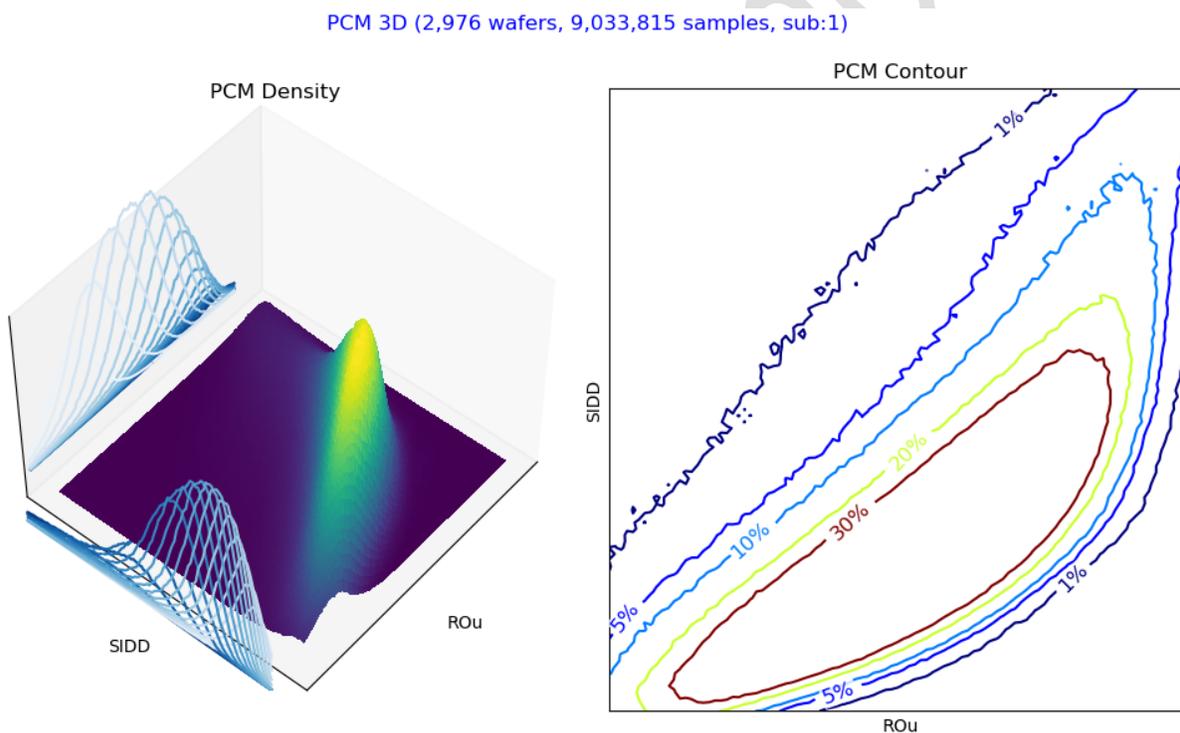


Fig. 9-5 CDF Contour

圖 Fig. 9-6 展示了一種常見的量產階段二維數據分箱策略 (Binning Strategy) 應用。透過大量接近真實情境的虛擬晶片數據，我們能精確預測產品性能分級的數量與補償，從而提升產品優化效率、提高產能並降低成本。同時，透過系統化策略確保品質的穩定性、可靠性與競爭力。

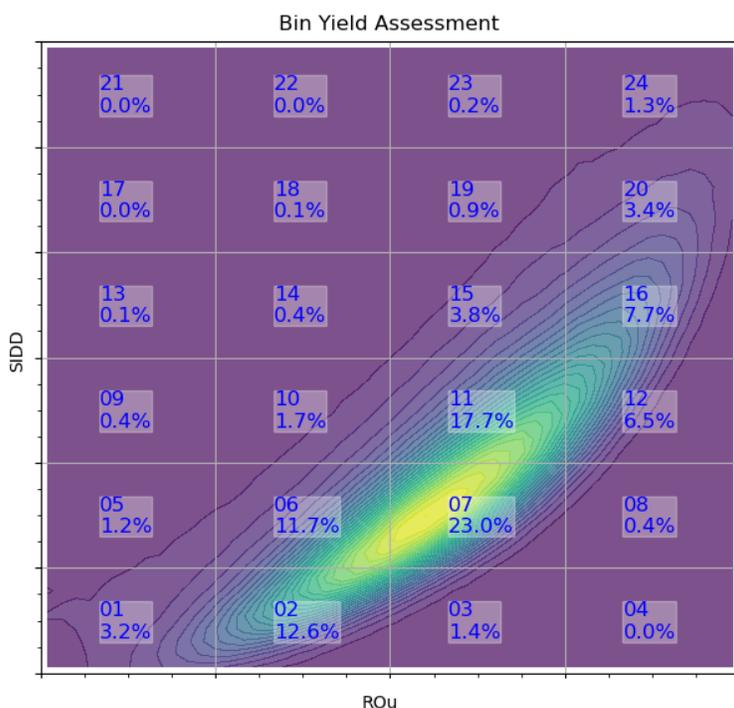


Fig. 9-6 Speed and SIDD Binning Strategy

圖 Fig. 9-7 展示了一種跨維度特徵輪廓 (Feature Contour) 的應用，其中 WAT N/P 製程參數作為 XY 軸基底，CP 性能與功耗數據作為 Z 軸。這類方法如同靜態的文章翻譯專家，能夠促進製程與晶片設計領域的專業術語互相理解。透過虛擬晶片數據的輔助，製程與設計工程師可在不同設計目標之間尋求平衡，實現優化的製程與性能。透過跨資料維度與關聯性分析，例如將晶圓層級的製程參數作為基底，我們能夠提供具體的製程調整策略與指導方針，以提升晶片性能的同時維持相同的體質 (如 SIDD 與 RO 頻率)。

此外，生成模型所產生的跨維度數據關聯性，為 DTCO 的實踐應用提供了極具潛力的發展方向。如圖 Fig. 9-8 所示，我們觀察到晶圓層級製程參數 (如 N-MOS 與 P-MOS 的臨界電壓 V_{th}) 與晶片層級的速度 (F_{max}) 與漏電流 (SIDD) 之間存在顯著相關性，且 N/P V_{th} 製程參數的變異性顯示出負斜率係數與截距。這一方法如同動態的口語翻譯專家，透過跨資料維度的關聯性分析與數據互動 (Cross-probing)，使我們能夠更有效地在目標能效與製程配方之間尋求最佳妥協與優化方向。

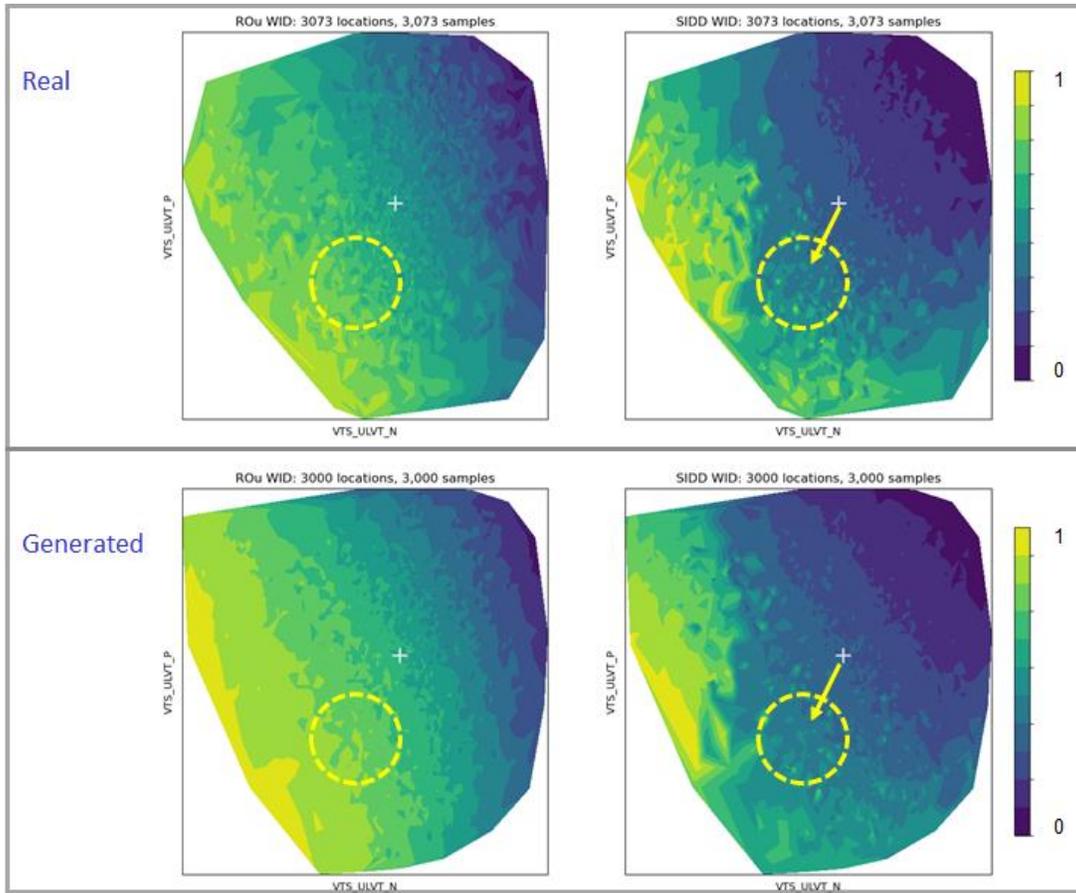


Fig. 9-7 Process Recipe Window Optimization

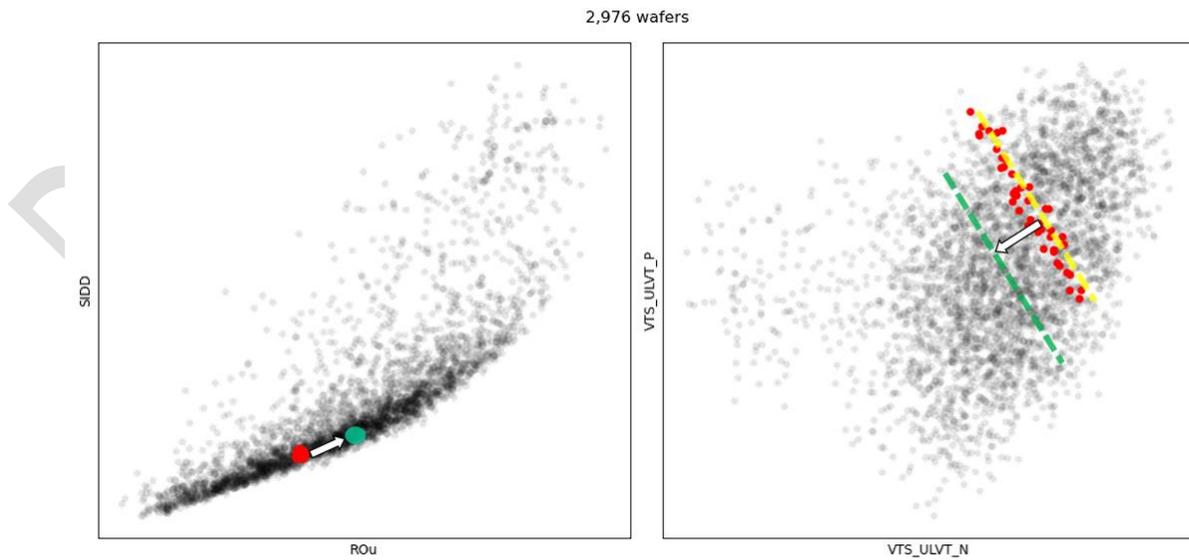


Fig. 9-8 Multi-Feature Cross-Probing

第10章 DTCO.VS 虛擬晶圓數據生成技術 (Virtual Silicon)

本隨著矽晶數據量的激增，機器學習技術在提升晶片性能方面發揮著重要作用。深入分析晶圓內部關鍵特性的分佈，有助於精確預測製程參數與晶片行為。然而，由於商業機密的限制，實際矽晶數據的獲取與共享面臨挑戰。本章提出了一種高品質的矽晶數據生成方法，結合晶片測試 (CP) 與晶圓驗收測試 (WAT) 數據，通過多項指標驗證其有效性，實現更完整的數據集，支持 DTCO 的協同優化。

生成式人工智慧 (Generative AI) 被探索用於學習晶片測試與晶圓驗收測試特徵，解決真實矽晶數據收集的難題。透過生成模型準確捕捉物理特性，設計人員能夠模擬量產測試過程，擬定分箱補償策略，進一步優化設計並克服數據傳輸與共享的挑戰。

本章節將介紹基於生成對抗網絡 (GAN) 和擴散模型 (Diffusion) 的晶片與晶圓測試數據建模方法，並探討晶片虛擬化在機密資訊封裝、壓縮與跨領域交付的應用。透過這些方法，設計人員能夠生成更真實的設計範例，進行設計與製程協同優化 (DTCO)，從而提升產能並實現更高的能效比設計。

10.1. 資料集準備

本文示範的數據集涵蓋來自 3,984 片晶圓，總計約 1,200 萬筆晶片數據。經處理後，移除超出 3σ 範圍的離群值和缺失數據，同時特意保留部分具有製程平坦性缺陷的晶片作為模型訓練樣本。為確保晶片技術與晶圓製程數據的機密性，範例中的圖表與數值皆正規化至 -1 至 1。數據包含晶片座標資訊，特徵維度為 14，並轉換為多通道影像格式，使整體數據集尺寸達 $65 \times 66 \times 14$ (高度 \times 寬度 \times 特徵數量)。

表 TABLE I 描述了數據集中 14 個特徵的原始含義，其中 6 個來自晶片測試 (CP)，另 8 個則源自晶圓驗收測試 (WAT)。每個特徵皆對應特定的物理屬性，例如，CP1 反映晶片的漏電流，CP2 代表晶片速度。其他與頻率相關的 4 個特徵 (CP3、CP4、CP5 和 CP6) 則分別描述晶片在 300MHz、400MHz、500MHz 和 600MHz 工作頻率下的功能準確性。功能準確性定義為多核心電路在特定頻率下能夠正確輸出的核心數量。這些特徵提供了對晶片行為的深入洞察，幫助設計人員進行設計餘量分析與優化。

TABLE 1: Features of the Dataset

Feature	Description	Unit
CP1	Leakage current	μA
CP2	Chip speed	Hz
CP3	Functional accuracy at 300MHz	%
CP4	Functional accuracy at 400MHz	%
CP5	Functional accuracy at 500MHz	%
CP6	Functional accuracy at 600MHz	%
WAT1	Gate threshold voltage of low threshold NMOS	V
WAT2	Gate threshold voltage of low threshold PMOS	V
WAT3	Gate threshold voltage of ultra-low threshold NMOS	V
WAT4	Gate threshold voltage of ultra-low threshold PMOS	V
WAT5	Drain current of the low threshold NMOS	mA
WAT6	Drain current of the low threshold PMOS	mA
WAT7	Drain current of the ultra-low threshold NMOS	mA
WAT8	Drain current of the ultra-low threshold PMOS	mA

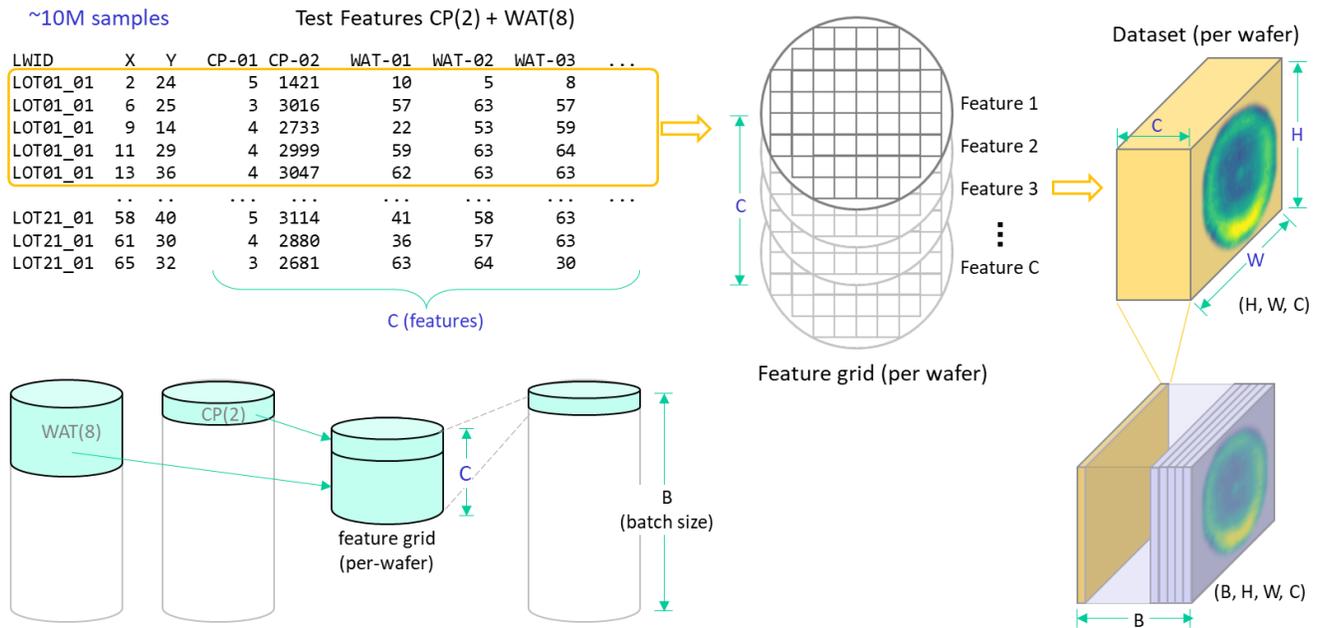


Fig. 10-1 Dataset Conversion

如圖 Fig. 10-1 所示，圖中展示了將晶圓數據轉換為多通道影像格式的過程。我們將原始 CP 和 WAT 測試數據轉換為二維影像，並設置多個特徵維度（參數 C）。參數 C 的大小與模型規模直接相關，且訓練時間隨 C 呈非線性增長。圖例中的數據集包含 8 個 WAT 測項 2 個 CP 測項，共

計 10 維 ($C=10$)。需要注意的是，晶圓本質上為圓形，而本文中數據的表示方式為矩形，因此位於晶圓邊界之外的數據會透過 Mask 過濾，以確保分析的數據範圍與準確性。

我們可進一步通過微小角度旋轉變換來擴增訓練集，這對於量產前期資料稀缺時尤為有效。這種方法不僅增加了資料的多樣性，還提升了模型的穩定性。通過模擬晶圓製造過程中可能出現的旋轉缺陷和製程參數分佈，如圖 Fig. 10-2 所示，我們能夠準確捕捉晶片製造過程中的關鍵特徵，從而提高了模型的訓練效果。

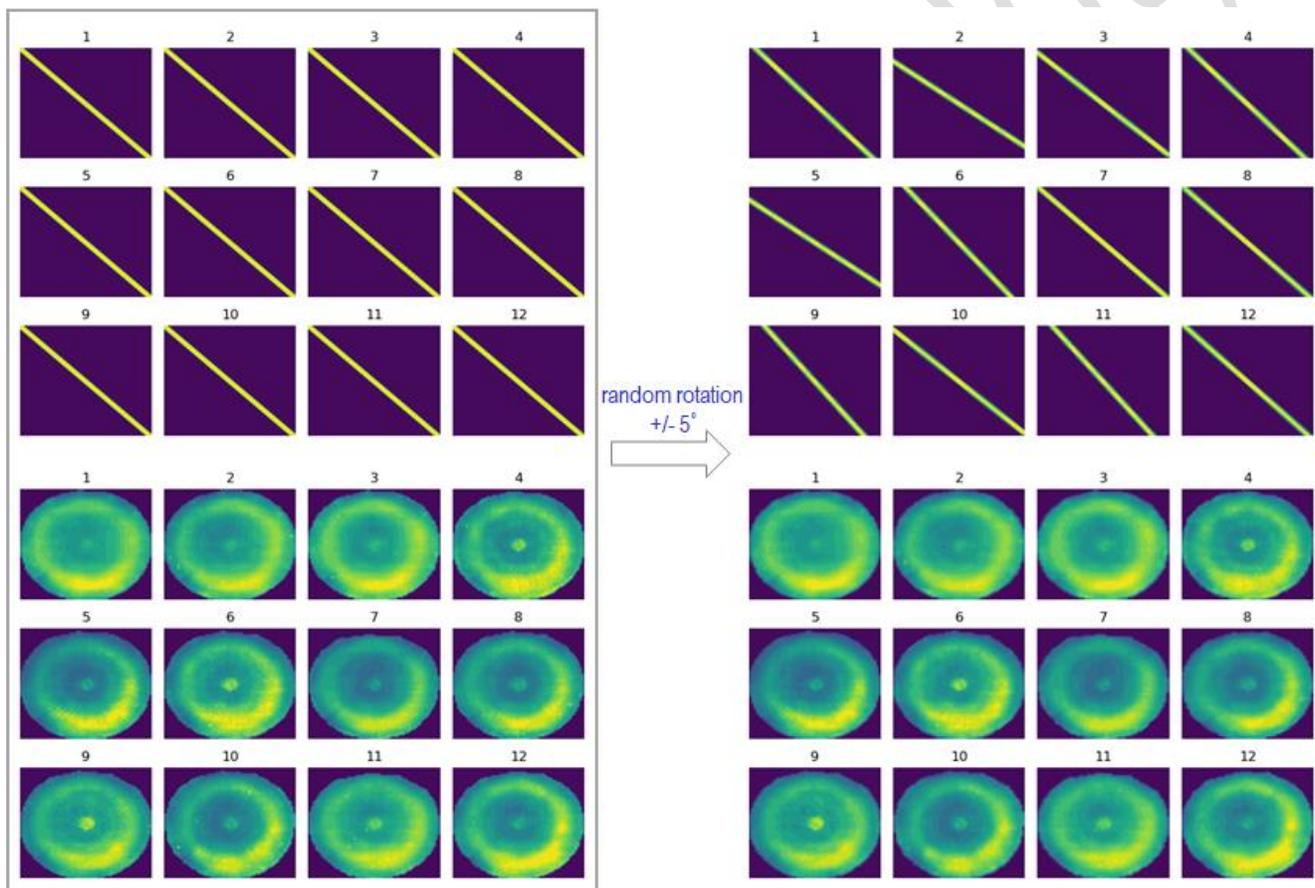


Fig. 10-2 Data Augmentation

10.2. 基於生成對抗網路的虛擬晶片數據 (GAN-based Virtual Silicon, GAN-VS)

本章說明如何利用 GAN 模型對晶片製造過程中的多維測試數據進行建模，該數據涵蓋晶片性能、晶圓製程特徵及潛在缺陷，以精準捕捉製程缺陷與參數一致性 (uniformity) 特徵。

10.2.1. GAN 模型

如 Fig. 10-3 所示，模型的生成器由多層卷積層組成，並通過 Tanh 層輸出多維影像，用於模擬晶片圖像。鑑別器也包含多層卷積層，並以 Sigmoid 層作為輸出，用於判斷晶片數據的真偽。兩者協同運作，有效生成高質量的晶片數據。

在訓練過程中，我們採用 BCE Loss 並透過梯度下降法最小化生成晶片與真實晶片之間的差異。為提升訓練效率與穩定性，模型引入批次正規化 (Batch Normalization) 與 LeakyReLU 激活函數。我們使用 Adam 優化器，學習率初始設為 0.001，並於每 100 個 epochs 衰減 0.9，批次大小為 20，訓練總計 10,000 個 epochs。最終，該 GAN 模型成功生成逼真的晶片數據，包括晶片位置、晶圓層級平坦度及製程缺陷，為晶片製造過程的模擬與分析提供了可靠的數據基礎。

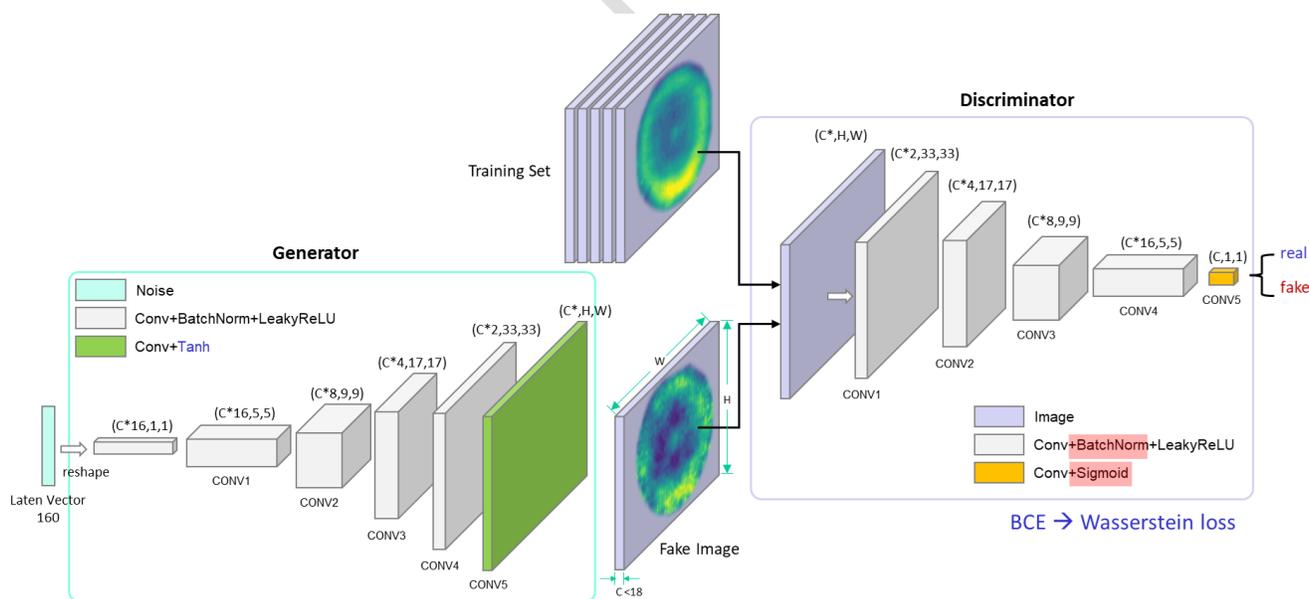


Fig. 10-3 GAN Modeling

10.2.2. GAN 模型性能評估

接著，以兩兩特徵組合作為坐標軸，將生成樣本與真實樣本的 2D 散點圖繪製出來。理想情況下，每一組數據的散點分佈應高度重疊，以確保生成樣本在單一特徵的分佈及多特徵的聯合分佈上與真實數據相匹配，如圖 Fig. 10-4。

除了進行資料視覺化外，我們還使用了定量指標來評估生成的晶片資料的質量。例如，我們使用 Jensen-Shannon 散度指標 (JS Divergence) 比較生成的晶片資料與真實晶片資料之間機率分的分佈相似度，並使用 KDE 指標來衡量各特徵機率分佈之間的數值差異以確保我們對生成晶片資料的可靠性和準確性。

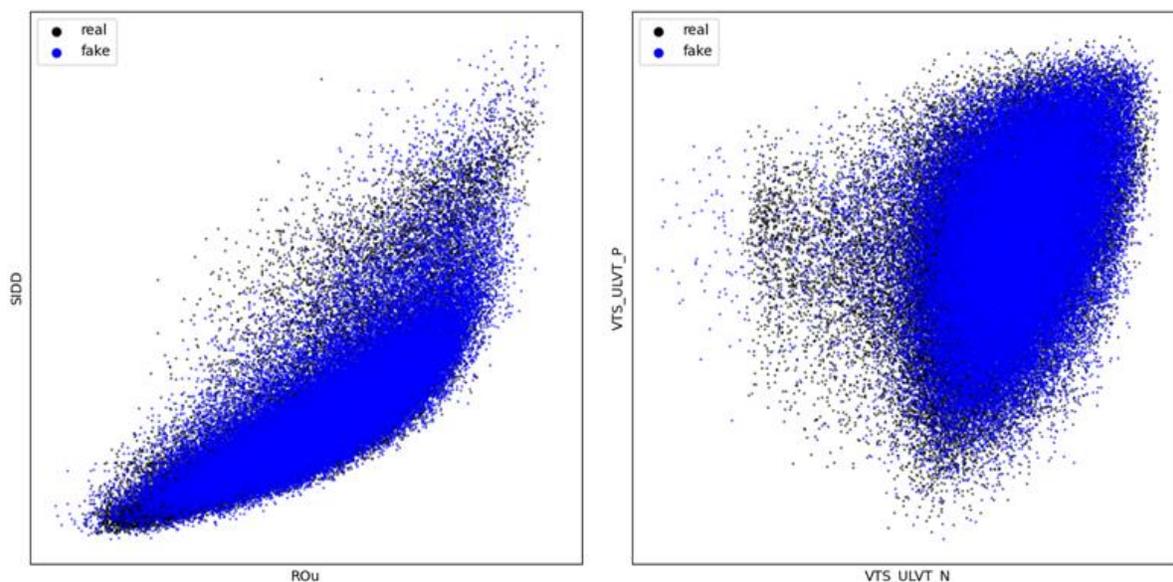


Fig. 10-4 Feature Scatter Plot for GAN Model Similarity

在物理特徵層面，我們通過比較真實數據與 GAN 模型生成樣本在各特徵上的分佈來評估其相似性。我們使用 Jensen-Shannon (JS) 散度來衡量兩個分佈的相似性，較低的 JS 散度表示生成樣本與真實樣本的相似度較高。真實分佈 P 和生成分佈 Q 之間的 JS 散度定義如下：

$$D_{JS}(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M) \quad (1)$$

其中，Kullback-Leibler 散度 $D_{KL}(P||M)$ 可由以下公式計算：

$$D_{KL}(P||M) = \sum_{x \in X} P(x) \log \frac{P(x)}{M(x)} \quad (2)$$

而 M 表示 P 和 Q 的混合分佈，定義為：

$$M = \frac{1}{2} (P + Q) \quad (3)$$

由於兩個分佈的 JS 散度範圍從 0 到 1， P 和 Q 的 JS 散度相似度定義為：

$$Similarity = 1 - D_{JS}(P||Q) \quad (4)$$

透過 JS 散度相似度分析 ($Similarity=1-JS$)，我們基於不含 CP3-CP6 的 10 維數據集發現，模型生成的虛擬晶片數據在各維度的機率分佈與真實數據高度吻合，相似度範圍介於 0.98 至 1.0 之間。此外，對於不自然的異常數據集，模型展現了優異的抗躁能力和泛化性能，具體結果請參見圖 Fig. 10-5。

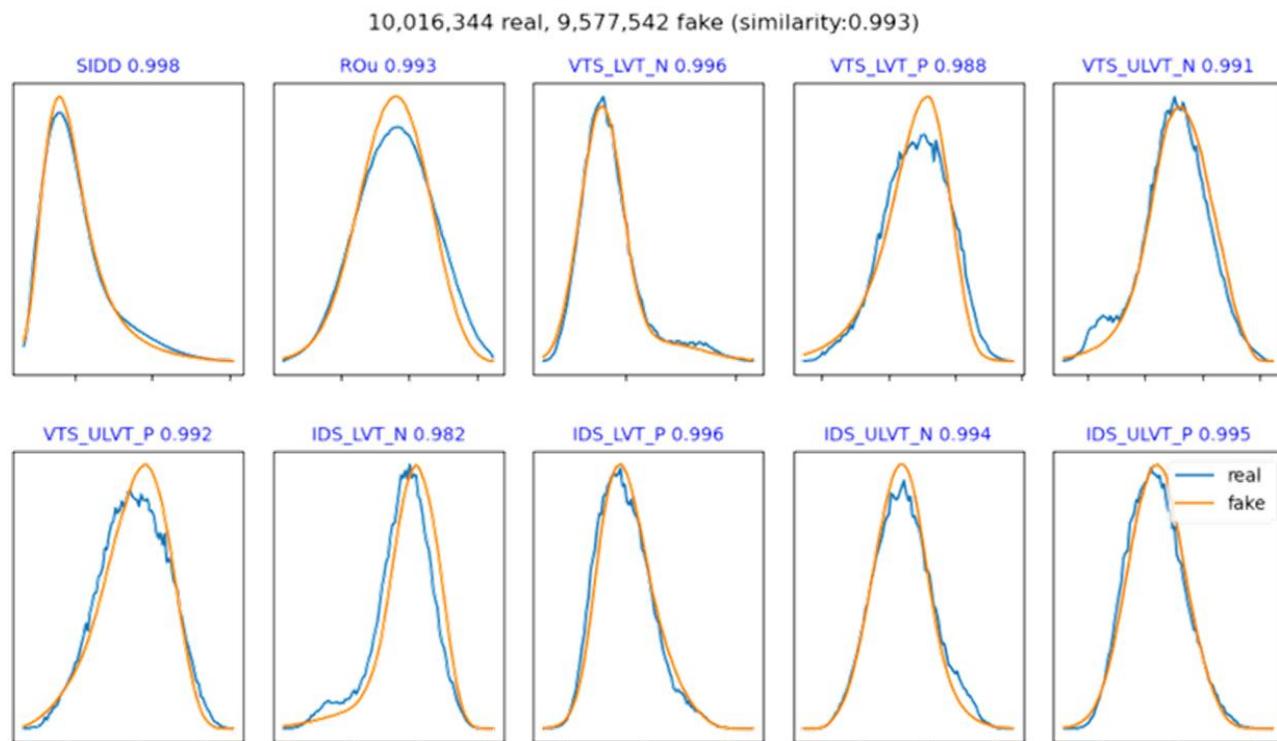


Fig. 10-5 Feature Distribution Similarity (C=10)

如圖 Fig. 10-6 所示，實驗結果表明，GAN 模型生成的高維晶片數據在任意兩特徵之間的散佈圖與真實晶片數據非常相似，並且成功學習了製程調整過程中的變異性，這些變異性在圖中以黃、綠、藍色圓圈標註。此外，在高維空間中，任意兩個特徵或更高維度的組合及其聯合分佈保持與原始樣本空間一致的關聯性。

為了進一步提升生成模型的品質，我們可以選擇性地省略一些在實際量產中不重要的數據，例如前期刻意的製程偏斜晶圓 (skew wafers)，以避免模型學習到早期階段製程參數調整過程中的異常值。與傳統數據建模方法對真實數據的分析相比，GAN 模型能夠更有效地學習晶片製造過程中的非線性關係與特徵表示，捕捉微小細節並更精確地擬合真實數據的分佈。進一步的分析顯示，GAN 模型成功學習並捕捉了晶片製造過程中的參數分佈、晶圓層級結構的平坦性一致性等細節信息，並實現了更高的準確率，反映出更真實的晶圓缺陷，如圖 Fig. 10-7。

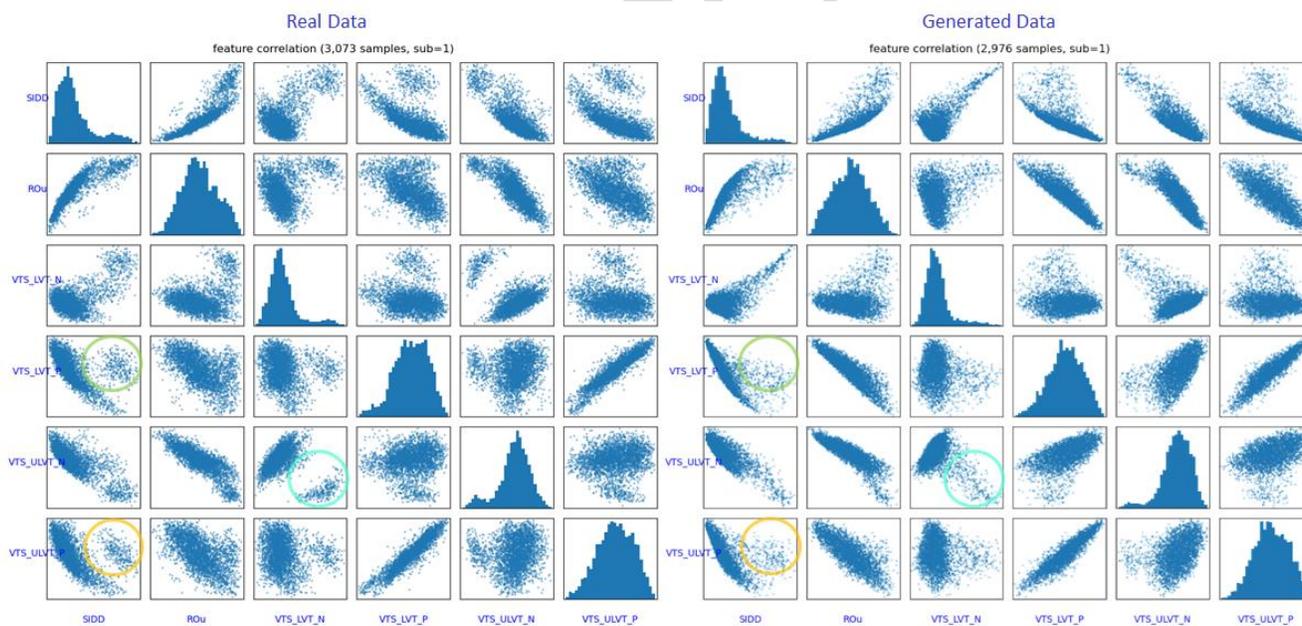


Fig. 10-6 Feature Correlation Matrix between Generated and Real Silicon

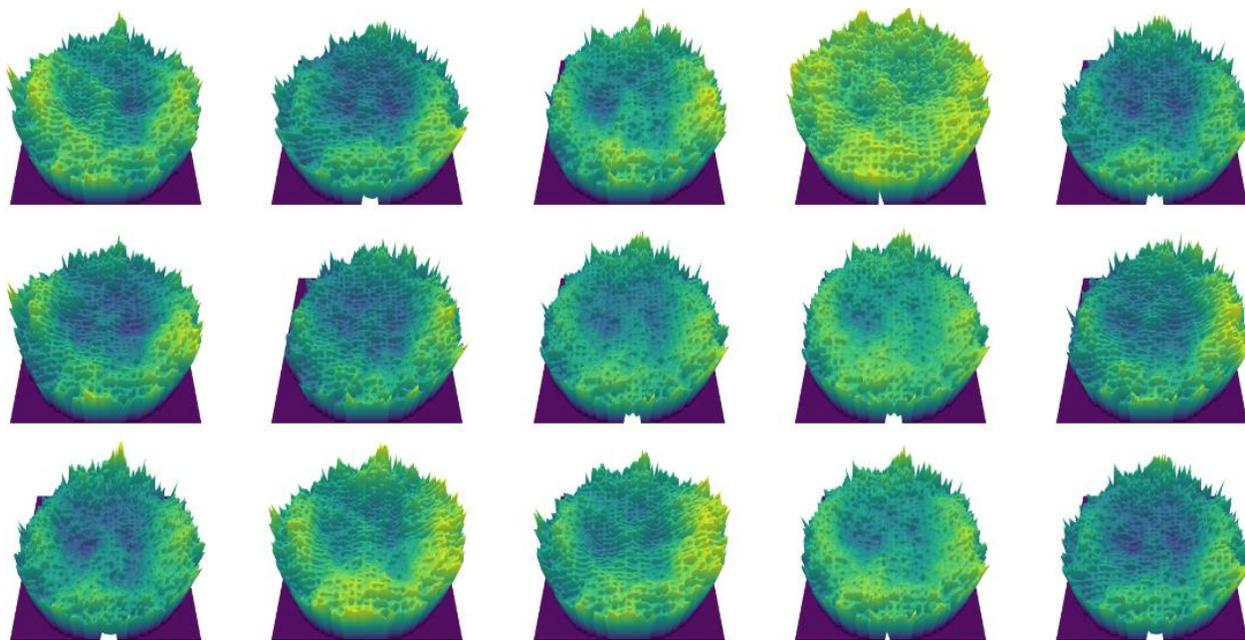


Fig. 10-7 Wafer-level Feature Uniformity of Generated Silicon

然而，基於 GAN 的方法常常面臨收斂困難和不穩定的問題。儘管改進版的 WGAN（基於 Wasserstein Loss，移除 BatchNorm 和 Sigmoid 層，如圖 Fig. 10-3）能提升性能，但實驗結果顯示，GAN 在處理具有多個峰值的數據分佈時仍存在挑戰。此外，晶片功能準確性（如 CP3-CP6 等頻率依賴特徵）的分佈通常不是高斯分佈，可能呈現雙峰、skew-normal、log 或 cosh 形式，而多核心晶片的性能受操作頻率影響較大，這使得 GAN 難以有效處理此類任務。下一章節將介紹如何使用擴散模型來解決這些問題。

10.3. 基於擴散模型的虛擬晶片數據 (Diffusion Model-based Virtual Silicon, DM-VS)

本章節簡要介紹了基於 DDPM (Denoising Diffusion Probabilistic Models) 去噪擴散概率模型的方法，該方法能夠生成更高質量的矽晶圓數據，並有效解決了 GAN 模型在處理多核心晶片性能相關特徵方面的局限性。通過 JS 散度和 Fréchet Inception Distance (FID) 指標評估數據分佈和質量，實驗結果表明，擴散模型能夠準確提取矽晶圓數據中的多重特徵分佈，並生成大量矽晶樣本，從而支持更深入的分析並加速 DTCO 進程。在 14 維數據集中，相較於 GAN，擴散模型生成的虛擬晶圓在數據分佈上更貼近真實數據，JS 散度相似度提升至 0.987，FID 降至 6.28。

10.3.1. 去噪擴散概率模型 (Denoising Diffusion Probabilistic Model)

去噪擴散概率模型 (DDPM) 是一種生成模型，透過逐步去除噪聲，將隨機噪聲樣本精細化為高品質的生成數據。擴散模型在文本與圖像生成等領域取得卓越成果，能夠產生細節清晰且高度擬真的結果。DDPM 的核心概念是逆轉擴散過程，透過逐步修正從正態分佈抽取的樣本，使其最終呈現出與真實數據相符的分佈，從而實現高品質的數據生成。圖 Fig. 10-8 展示了 DDPM 的運作過程，包括前向與反向過程，各經歷 T 個步驟。

在前向過程中，DDPM 將給定數據的分佈逐步轉換為更簡單的分佈（通常為正態分佈），這可視為一個馬爾可夫鏈，透過逐步添加高斯噪聲來完成。具體而言，初始晶圓樣本 w_0 在每個步驟中加入噪聲，最終轉化為純高斯噪聲樣本 w_T 。這些噪聲的添加方式由神經網絡學習，以便在反向過程中進行準確預測。

反向過程則逐步逆轉前向過程，從噪聲樣本中恢復原始數據分佈。具體來說，生成晶圓的過程始於噪聲樣本 w_T ，並將其與時間步長 $t=T-1$ 一起輸入神經網絡，預測當前步驟中加入的噪聲。隨後，從 w_T 減去預測出的噪聲，以獲得 w_{T-1} 。這一過程重複 T 次，最終將噪聲樣本 w_T 轉化為高品質的晶圓樣本 w_0 。

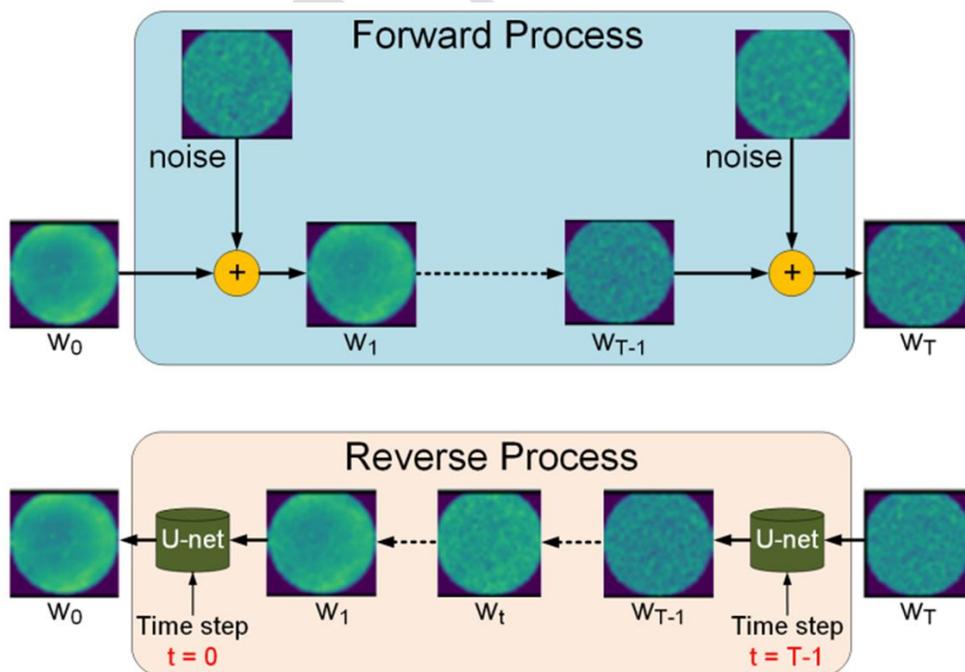


Fig. 10-8 Forward and Reverse Processes of the Diffusion Model

U-Net 常用於圖像分割與生成，其結構如圖 Fig. 10-9 所示。輸入首先通過一個卷積層，將通道數擴展至 16。模型包含兩個下採樣塊，逐步增加通道數至 64，然後透過兩個上採樣塊恢復解析度，同時將通道數減少回 16。最後，經過一個卷積層生成具有所需維度的輸出。圖 Fig. 10-10 進一步展示了 ResNet 區塊的細節。

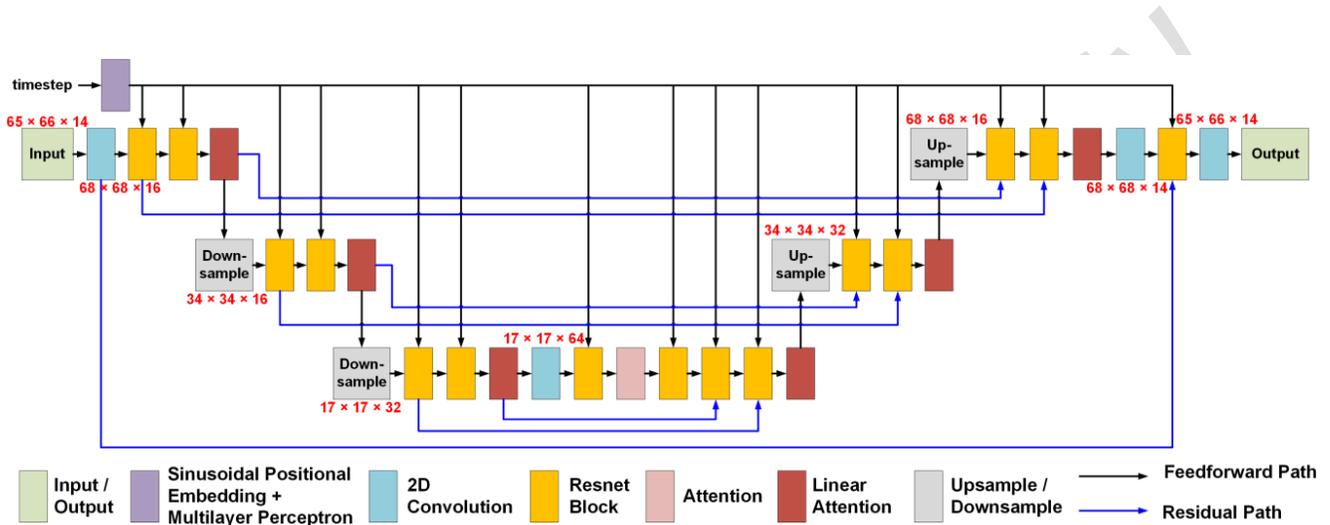


Fig. 10-9 U-Net of Diffusion Model

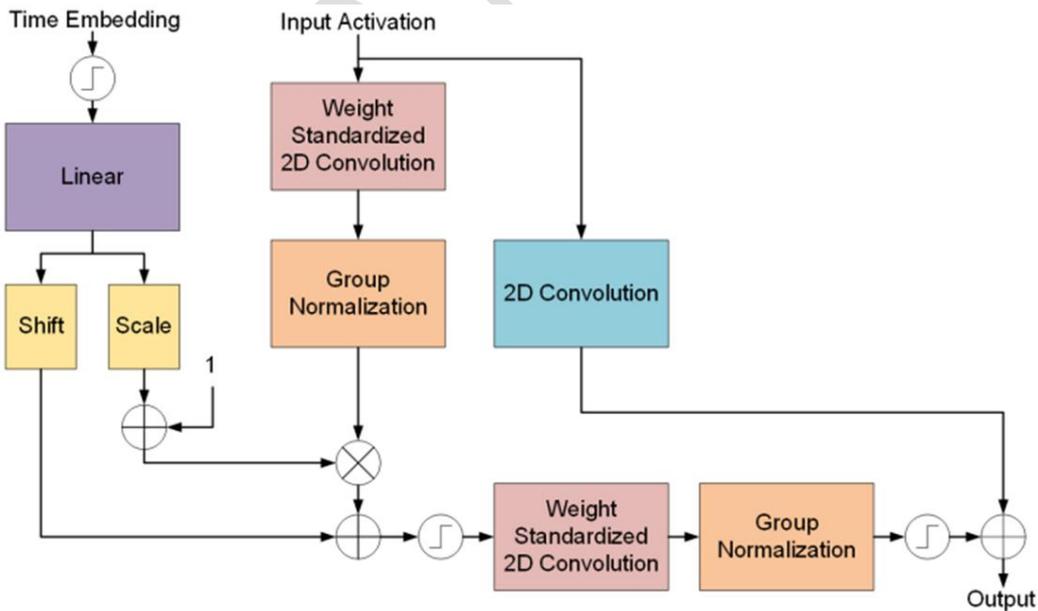


Fig. 10-10 Illustration of a ResNet Block

10.3.2. Diffusion 模型性能評估

本章節除了在晶片層級 (chip-level) 進行分析外，還引入了晶圓層級 (wafer-level) 的評估指標。與晶片層級不同，晶圓層級專注於來自同一晶圓的晶片，使我們能夠更有效地評估擴散模型是否成功捕捉訓練數據的複雜性，並提供更客觀的方式來比較擴散模型與生成對抗網絡 (GAN) 生成的數據。

在圖像生成任務中，常用的 Frechet Inception Distance (FID) 被用來衡量生成數據與真實數據之間的質量與多樣性。較低的 FID 表示生成數據更接近真實數據，反映出更高的生成品質。FID 計算過程涉及將真實與生成數據傳遞至預訓練的 Inception-v3 模型，並比較其特徵分佈。圖 Fig. 10-11 展示了晶圓數據轉換為 FID 計算所需維度的流程。

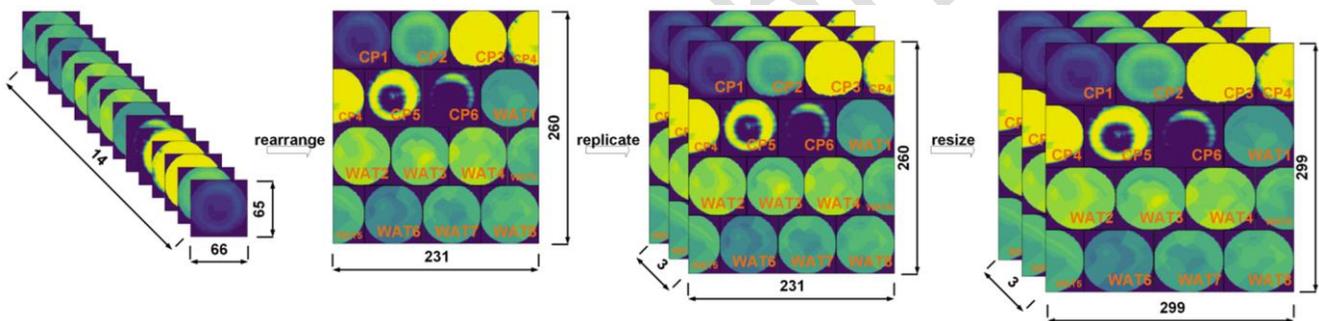


Fig. 10-11 Data Shape Transformation for FID Evaluation

首先，將尺寸為 $65 \times 66 \times 14$ 的晶圓樣本重新排列為 $231 \times 260 \times 1$ 的尺寸，然後複製到 3 個通道，最後使用雙線性插值將其調整為 $299 \times 299 \times 3$ 的尺寸。通過將數據傳遞通過 Inception-v3 網絡並提取最後池化層的激活向量，獲得晶圓樣本的向量表示。當真實數據和生成數據轉換為向量表示後，分別計算真實數據和生成數據的均值和協方差矩陣。設真實數據和生成數據的均值為 μ_1 和 μ_2 ，協方差矩陣為 C_1 和 C_2 ，則 FID 的計算公式為：

$$FID = \|\mu_1 - \mu_2\|_2^2 + \text{Tr}(C_1 + C_2 - 2(C_1 C_2)^{\frac{1}{2}}) \quad (5)$$

其中，符號 $\|\cdot\|_2$ 表示 L2 範數或歐氏範數， $\text{Tr}(\cdot)$ 表示矩陣的跡 (Trace)，為矩陣主對角線元素的總和。

- Die-Level Analysis

在晶片層面，散點圖用來比較真實數據和生成數據之間多個特徵的聯合分佈。圖 Fig. 10-12 顯示了擴散模型生成的四對特徵的散點圖。這四種情況下的生成數據點覆蓋了與真實數據點相似的区域，散點圖顯示了真實晶片和生成晶片在聯合分佈上的高度相似性。

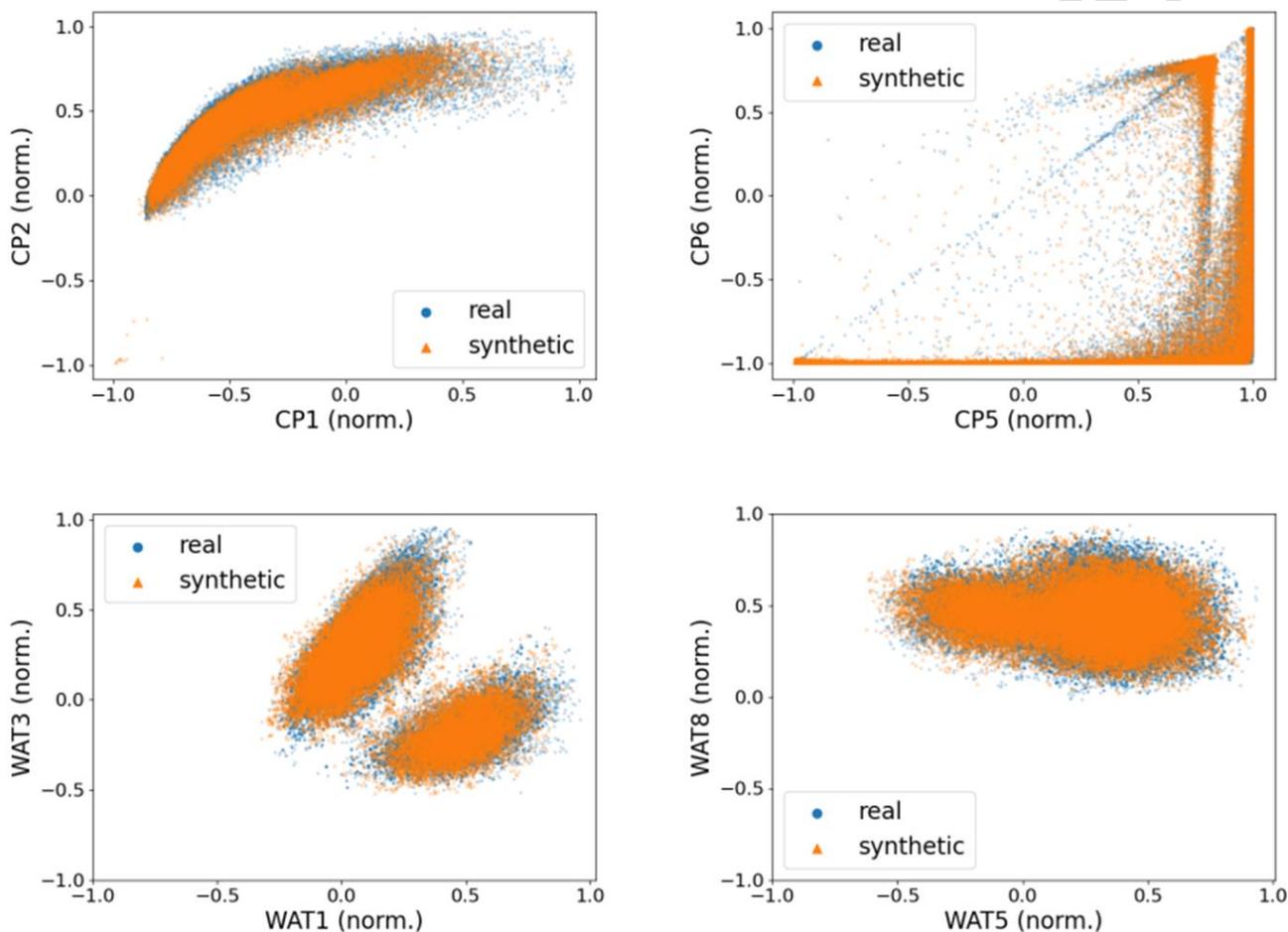


Fig. 10-12 Feature Scatter Plot for Diffusion Model Similarity

圖 Fig. 10-13 展示了所有特徵值範圍內的分佈情況，每個圖表上方的數字為該特徵的 JS 散度相似度。圖 Fig. 10-13 (a) 為 GAN 生成的分佈，可見其在特徵 CP3、CP4、CP5 和 CP6 上表現不佳。這是因為這些特徵的分佈具有多個峰值，且峰值高度的差異與位置的錯誤匹配，導致 JS 散度相似度較低。此外，GAN 未能準確呈現這些特徵分佈中的顯著峰值。

相比之下，圖 Fig. 10-13 (b) 顯示了擴散模型生成的分佈，展現了其在捕捉真實分佈方面的優勢。對於特徵 CP3、CP4、CP5 和 CP6，擴散模型能準確再現真實峰值的位置與高度。特別是針對具有特殊分佈（如 log 或 cosh 分佈）的多核心頻率性能特徵（CP3–CP6），GAN 的平均 JS 散度相似度為 0.963，而擴散模型顯著提升至 0.987，表明其對複雜分佈的表現更加優越。

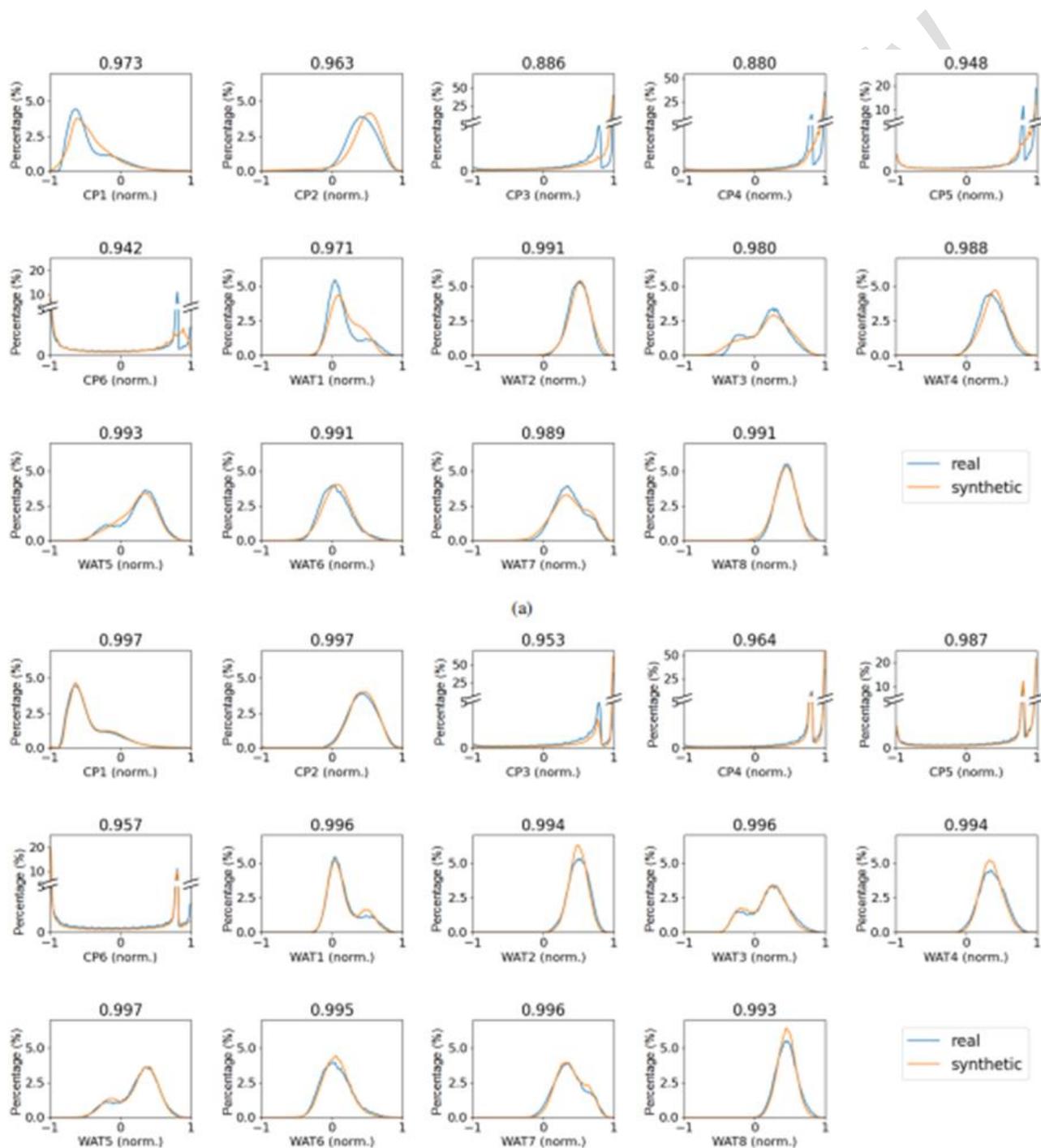


Fig. 10-13 Feature PDFs of GAN (a) and Diffusion Model (b)

- Wafer-Level Analysis

在晶圓層級，我們著重生成晶圓與真實晶圓在特定方向上相鄰晶片之間的差異，以評估生成數據的空間變化是否與真實數據一致。透過計算所有晶圓上的平均差異值，我們可以量化真實數據與生成數據的相似程度。

為簡化分析，這裡僅展示沿晶圓水平方向的切線，並檢查該方向上相鄰晶片的平均差異。圖 Fig. 10-14 顯示了由 GAN 和擴散模型生成的 CP1 與 WAT1 的平均差異分析結果，其中 x 軸代表水平方向上的 x 坐標，y 軸則表示平均差異。淺藍色區域標示出真實數據平均差異範圍內的一個標準差區間，以便直觀比較生成數據與真實數據的吻合程度。

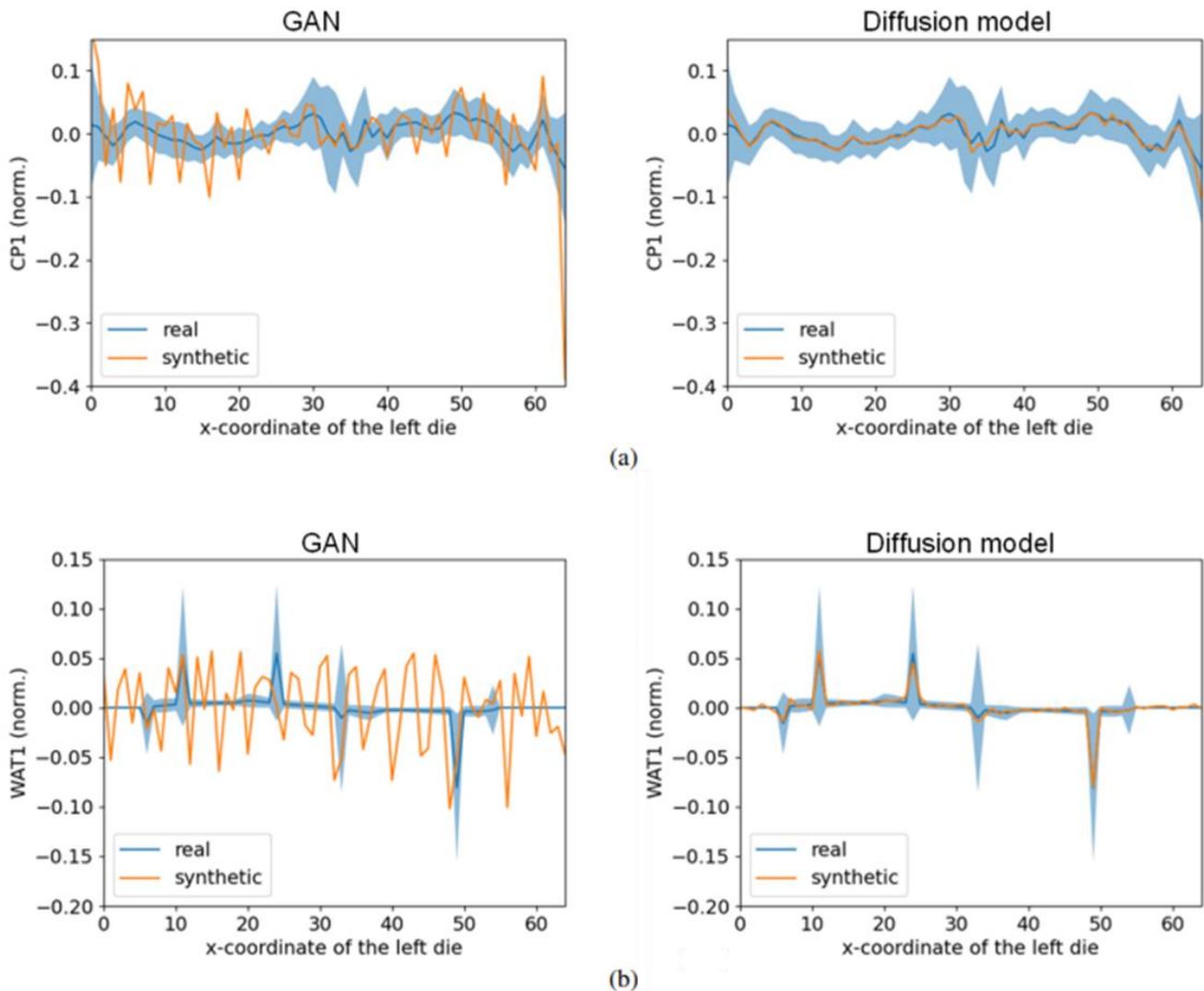


Fig. 10-14 Average Difference Analysis for (a) CP1 and (b) WAT1

圖 Fig. 10-14 (a) 顯示了 CP1 的平均差異。對於基於 GAN 的方法，生成數據的平均差異經常位於一個標準差範圍之外。此外，平均差異的劇烈波動表明 GAN 生成的生成數據存在不準確的模式。相反，對於基於擴散模型的方法，平均差異始終保持在一個標準差範圍內。圖 Fig. 10-14 (b) 中對於 WAT1 的結果類似，基於 GAN 的方法生成的生成數據的平均差異顯示出顯著的波動。雖然真實的 WAT1 的平均差異在某些位置顯示出峰值，但擴散模型仍能夠提取這些峰值。由擴散模型生成的平均差異始終保持在真實數據的平均差異的一個標準差範圍內。由擴散模型生成的生成晶圓在水平與垂直方向上的空間變化與真實晶圓非常相似。

表 TABLE II 進一步量化比較了擴散模型與 GAN 生成的虛擬晶圓的 FID (Frechet Inception Distance)，並與真實數據的 FID 進行對比。真實數據的 FID 參考值為 1.39，此數值透過將晶圓數據分為兩個等量部分計算得出。GAN 生成的 FID 為 55.13，顯示其在學習多峰值分佈特徵方面存在困難。相比之下，擴散模型生成的 FID 僅為 6.28，接近真實數據，展現出顯著的生成質量提升。整體而言，GAN 雖能在部分特徵上表現良好，但對於多峰值分佈特徵的捕捉能力有限。擴散模型更準確地模擬了真實數據的多特徵分佈，生成的晶圓樣本在晶片分佈和水平割線差異模式上均與真實數據高度一致，顯示出其生成能力的優勢。

TABLE II: Quality Comparison of Generated Data

Metric	GAN	Diffusion model
Average JS divergence similarity	0.963	0.987
FID	55.13	6.28

第11章 Generative AI 驅動的晶片能效優化與建模

11.1. WAT 超分辨率 (WAT Super Resolution, WAT-SR)

在製程參數決策中，傳統 WAT 採樣因點位稀疏，難以全面反映製程整體特性，影響決策的可靠性與信心。數據樣本不足可能導致決策偏誤，進一步降低參數調整的準確性。從高解析度的 CP 資料顯示，即使在同一曝光範圍內（例如 5×7 shot），晶片的特性仍可能存在顯著差異，如圖 Fig. 11-1。

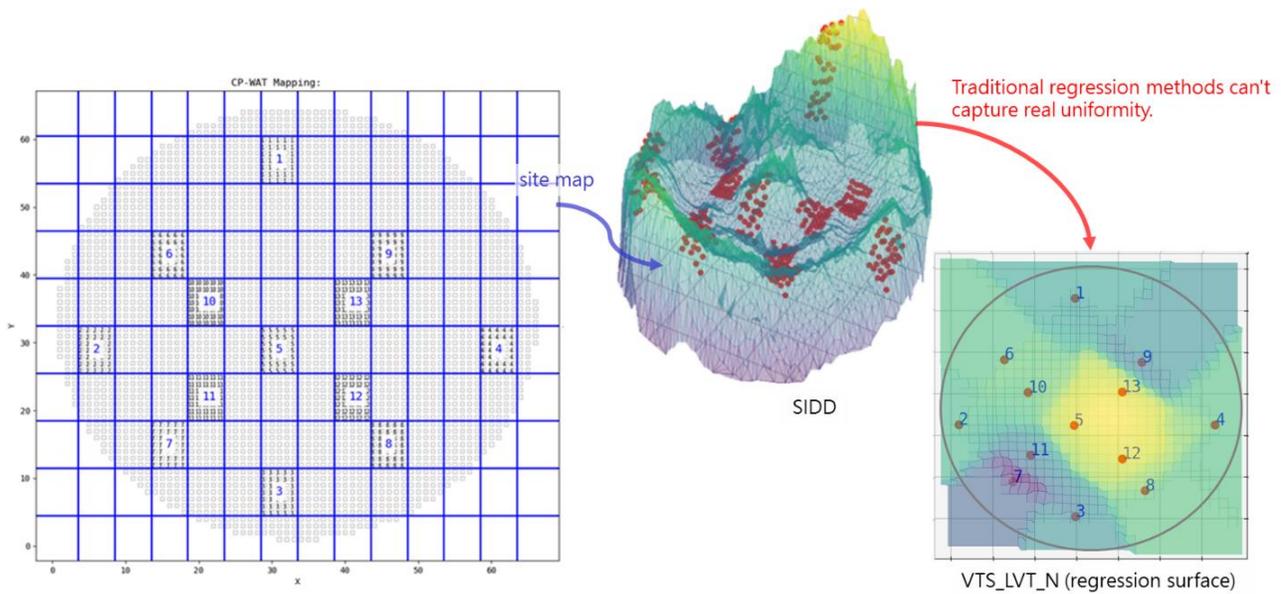


Fig. 11-1 WAT Sampling Issue

以單一 Lot (25 片晶圓) 的平均數據為例，隨著 WAT 採樣點數增加 (例如 80 點 full-map)，WAT 製程參數與 CP 高解析特徵曲面的均勻性分佈將逐漸清晰並趨於一致，如圖 Fig. 11-2 所示。若僅依賴量產時的 WAT 稀疏採樣 (例如 13 點或更少) 甚至進行線性內插，將難以準確捕捉晶圓表面的真實均勻性特徵。因此，應審慎評估採樣策略，確保數據量充足，以提升製程控制的精確度並強化決策可靠性。

此外，傳統的多項式回歸分析方法難以精確描繪晶圓層級的製程參數分佈，因為其處理高維度且具空間相關性的非線性變異能力有限，如圖 Fig. 11-3 所示。為提升數據品質與決策效能，應採用機器學習或生成模型等更創新的方法。

圖 Fig. 11-4 說明，大多數空間域信號可以通過頻率域中的少數關鍵諧波 (harmonic) 進行疊加近似。例如，利用 3 至 4 個不同協波週期、半徑和向位移，可以有效近似晶圓製程中的平坦一致性以及系統級缺陷，如邊緣效應、火山錐或甜甜圈形狀的同心圓不均勻性等。WAT-SR (WAT Super-Resolution) 技術旨在從晶圓的稀疏採樣點出發，通過神經網絡提升 WAT 解析度，進而捕捉晶圓層級真實的 WAT 均勻一致性分佈。

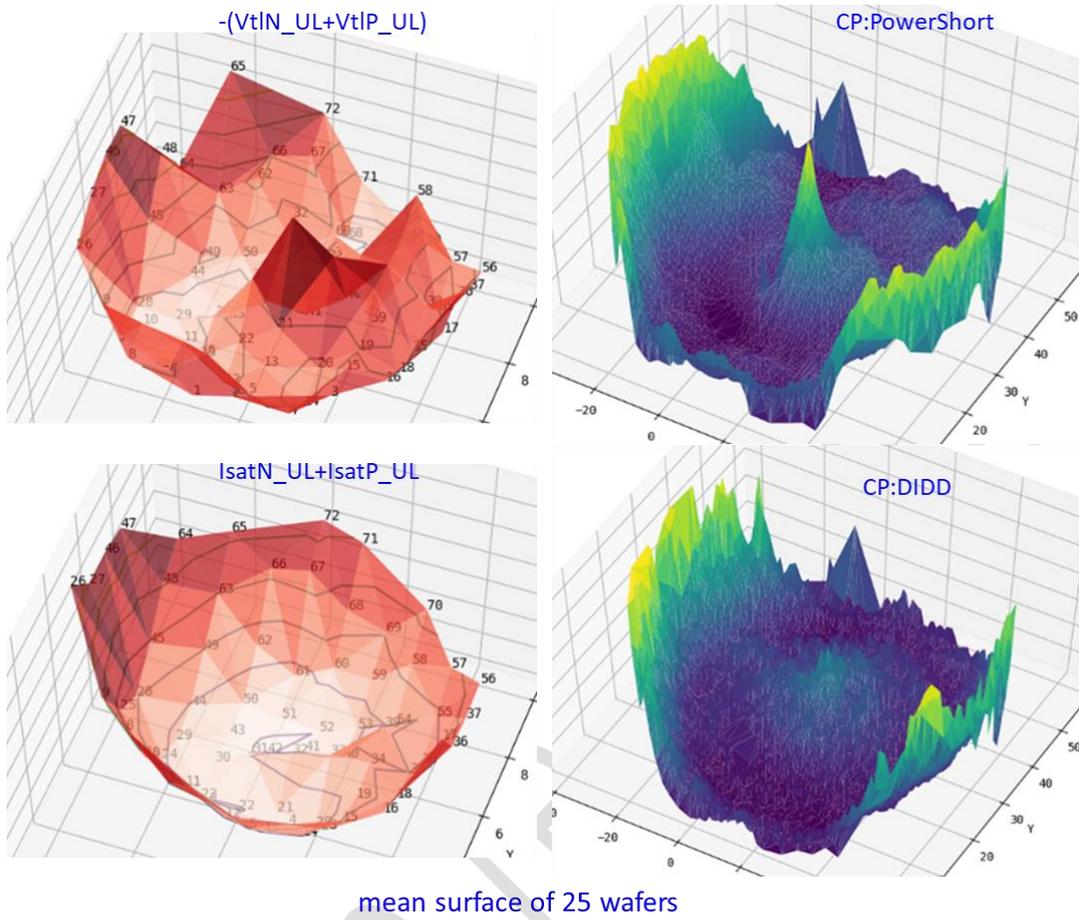


Fig. 11-2 Full Map Uniformity Comparison of WAT and CP

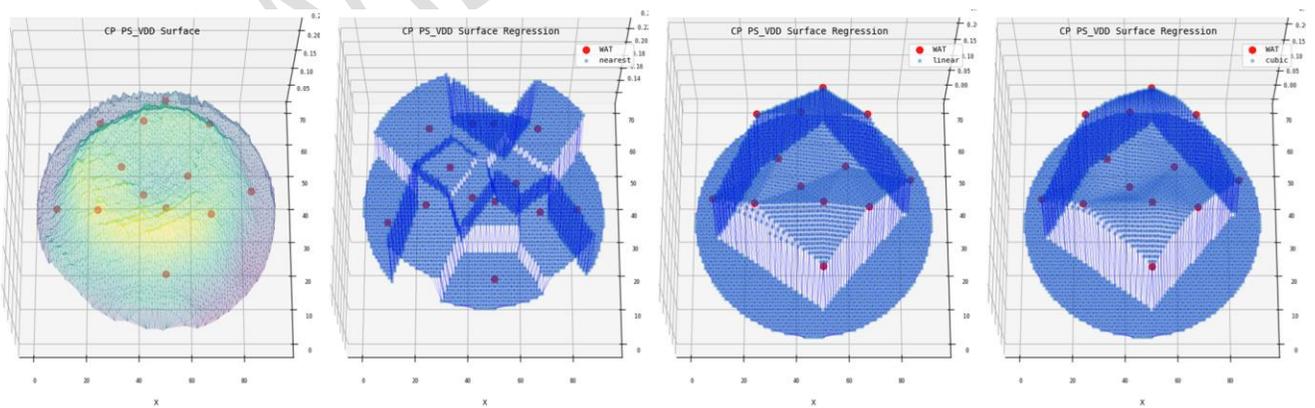


Fig. 11-3 Traditional Regression Models Miss Structural Information

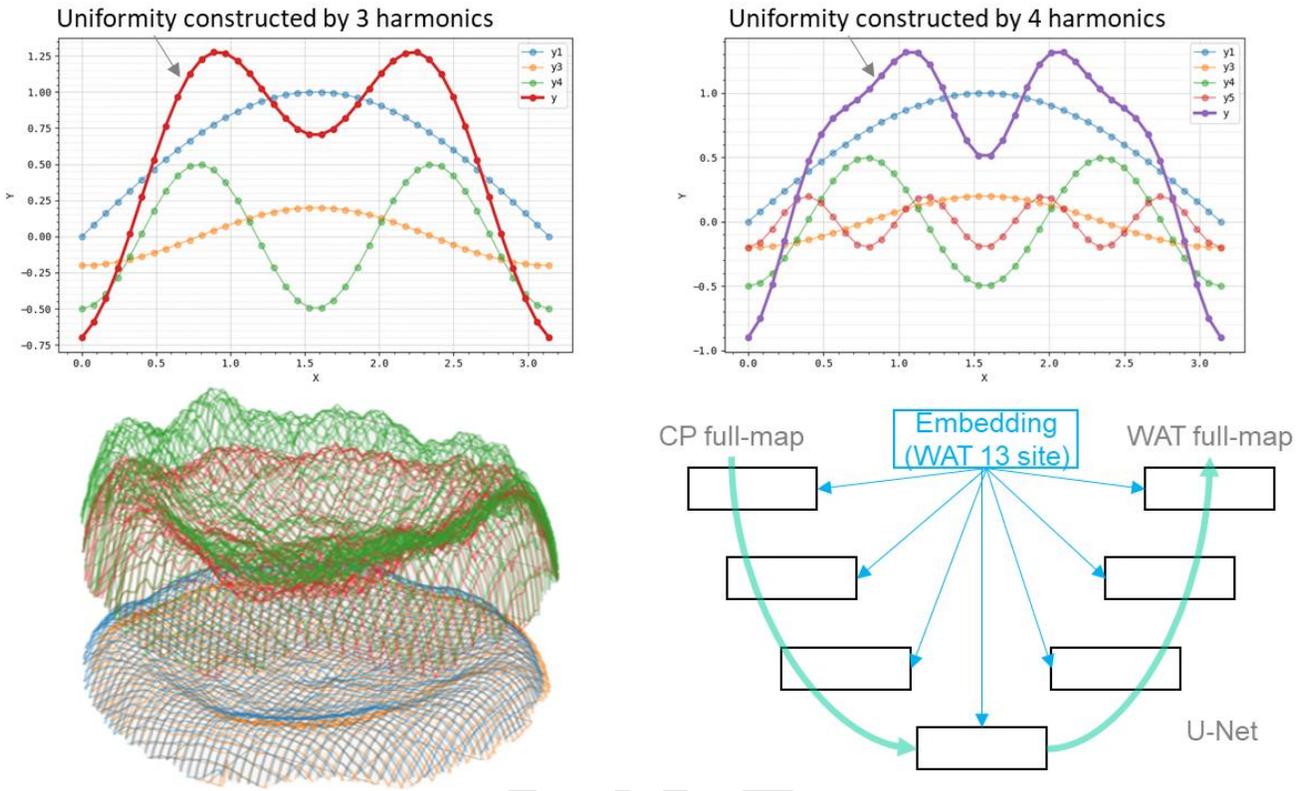


Fig. 11-4 Harmonic in Frequency Domain

核心思想是利用 U-Net 將稀疏的 WAT 數據（例如 13 個 SITE）通過 MLP 生成嵌入向量（Embedding），如圖 Fig. 11-5 所示，並與 CP 高解析度數據（如 RO 或 SIDD，65x66 晶片數量）共同作為訓練樣本輸入，生成與原始數據相似的高解析度 WAT 數據。此訓練過程不僅能精確重建晶圓的均勻一致性，還能有效捕捉晶圓製程中的缺陷結構特徵。

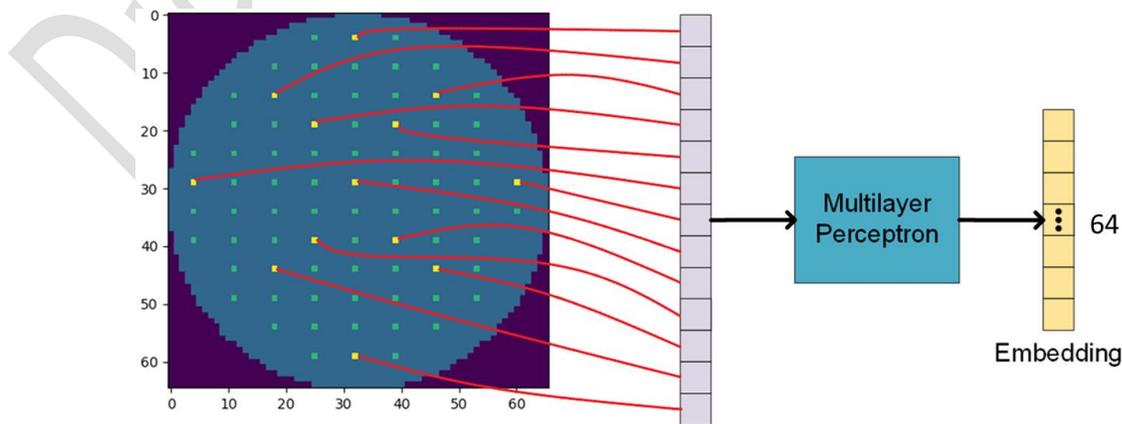


Fig. 11-5 Embedding 13 Sparse WAT Points Using an MLP

具體來說，U-Net 在全局結構推斷與局部細節保留之間達到良好的平衡，如圖 Fig. 11-6 所示。透過潛在空間（Latent Space）提供全局製程模式，並利用跳躍連接（Skip Connection）補充局部細節，使模型能夠同時考慮全局結構特徵與低階空間細節。例如，潛在空間捕捉到 CP 的全局製程特徵，但可能會丟失部分精確的位置信息。此時，跳躍連接能夠將編碼器（Encoder）中的低階特徵直接傳遞至解碼器（Decoder），有效補充這些細節，進而提升模型對局部結構與空間細節的理解。

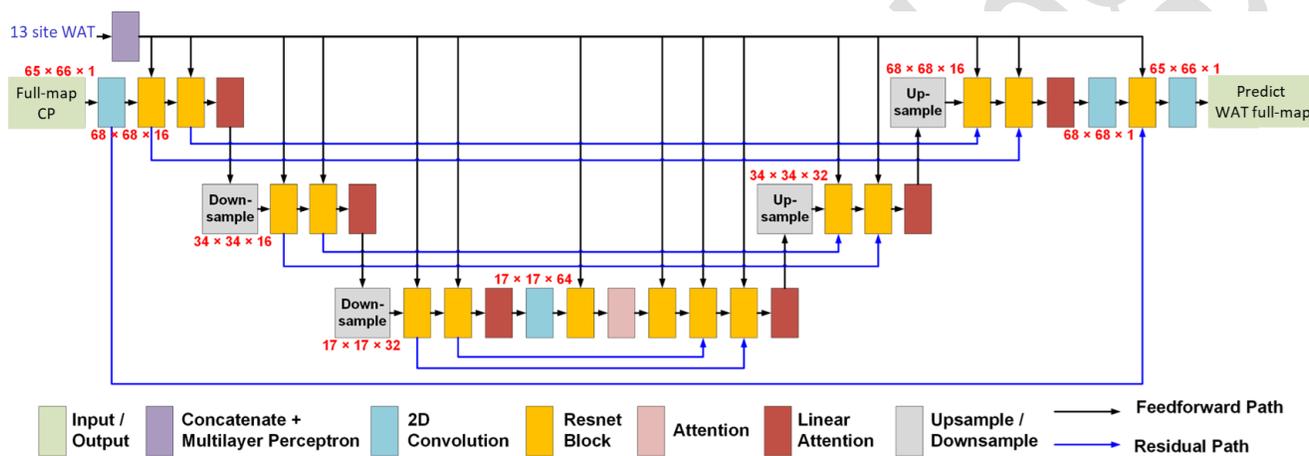


Fig. 11-6 U-Net of WAT-SR

卷積神經網絡通過逐層壓縮和特徵提取，不僅能在空間域中實現多頻率信息的分解與合成，還能有效學習高維空間中的多變異性（Multi-variance）。此外，將稀疏的 WAT 作為嵌入編碼向量（Embedding），並與 CP 高解析度資料結合，有助於充分挖掘訓練樣本中的結構特徵與細節信息，進一步提升模型表現。

模型透過高解析度的 CP 資料學習數據的內在分佈與結構，從而更準確地反映晶圓的均勻一致性分佈，並將稀疏的 WAT 資料提升至高解析度。這不僅提高了數據解析度，還能精確捕捉晶圓表面的真實特徵，進而增強製程參數控制的準確性與決策的可靠性。傳統上，由於稀疏資料所帶來的結構性問題難以捕捉，這一挑戰得以有效解決，使模型對晶圓均勻性的預測更加精確，並提高決策的信心。完整流程如圖 Fig. 11-7 所示。

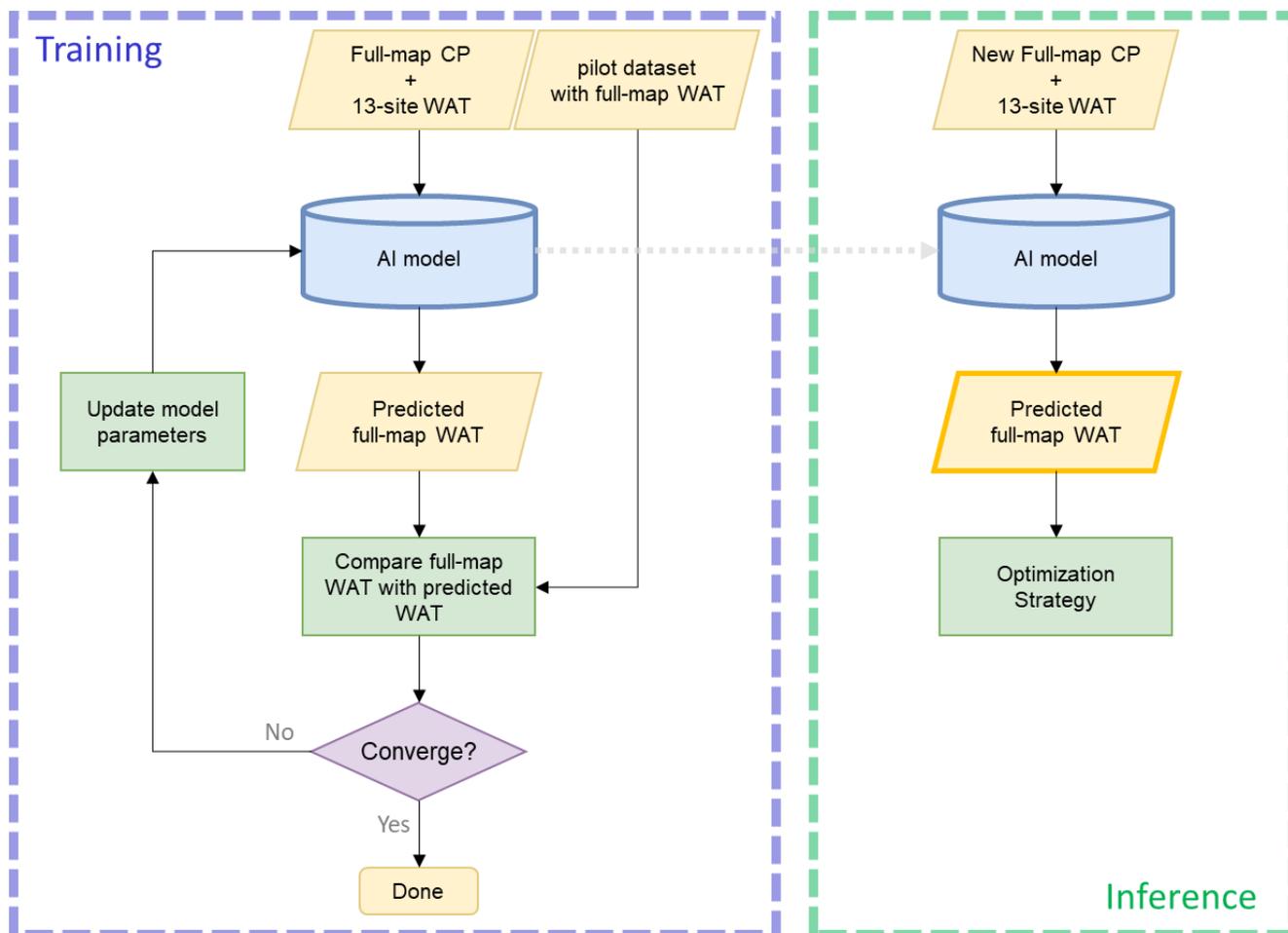


Fig. 11-7 WAT Super Resolution (WAT-SR) Flow

11.2. 高效矽製程偏移建模 (High-Efficiency SPICE-Silicon Bias Modeling, He-SSBM)

11.2.1. One-shot SPICE-Silicon N/P Correlation 的設計原理

RO 設置於晶片內，通常包含多條由不同元件組成的 Delay-Line (DL)。透過量產數據分析，我們可以間接評估與 SPICE 模型的差異，如圖 Fig. 11-8 所示。例如，當製程偏向 FS (NMOS 偏快，PMOS 偏慢) 時，P-stacking 元件的 DL 會導致其後矽量測頻率低於 SPICE 預期，而 N-stacking 元件的 DL 則會高於 SPICE 預期。根據多條 DL 的後矽量測值或目標值，我們可以推導出 SPICE N/P 參數調整的方向，從而使元件的時序重新萃取 Re-K 值，更接近實際的矽晶行為。

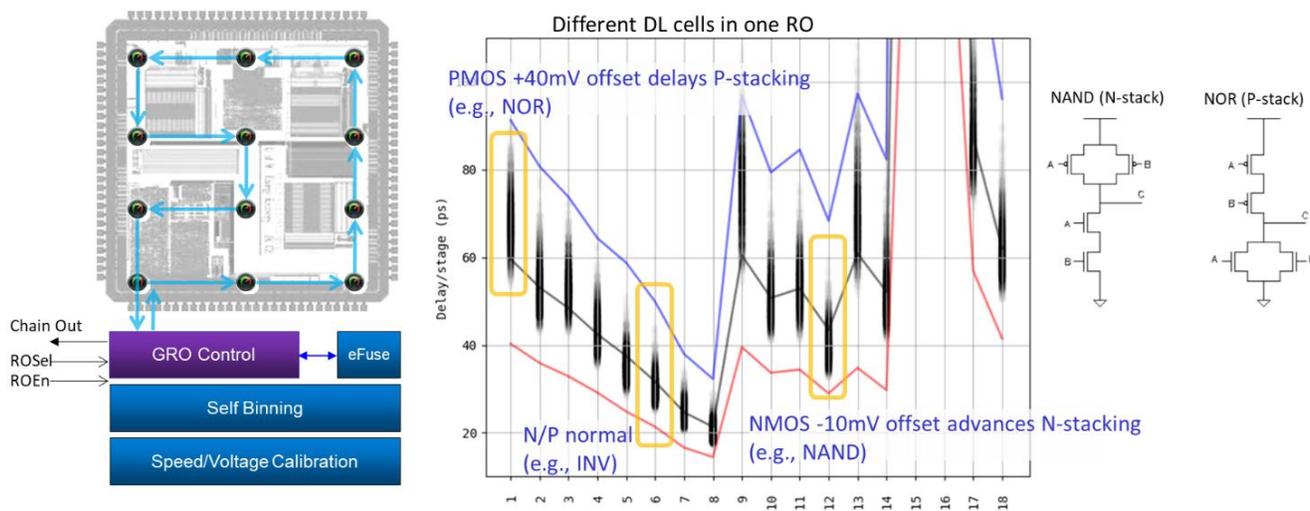


Fig. 11-8 RO Integration and S2S Correlation

11.2.2. 設計與防守策略優化

在晶片實體設計階段，我們會對每條 RO 的 Delay-Line (DL) 進行 SPICE 模擬，涵蓋 NOR (P-stack)、NAND (N-stack) 及 INV (正常模式下 N/P 平衡)。在晶片量產過程中，透過 RO 測量，我們便能夠了解其性能與 SPICE 模型之間的偏差，如表 TABLE III 所示。然而，如何將這些偏差映射回 SPICE 中的 N/P 偏移，以便校準時序並重新萃取 (Re-K)，從而實現更接近實際晶片數據的防守策略，並進一步提升晶片的性能與競爭力呢？

TABLE III Measured Target Silicon Data

	Type	DL1	DL2	DL3	DL4	DL5	DL6	DL7	DL8	DL9	DL10
Silicon1	Delay	4.068	3.138	3.862	1.924	2.497	1.337	2.947	1.430	1.922	1.026
	Power	145.000	548.300	163.100	81.720	577.900	301.500	146.300	107.200	513.700	393.200
Silicon2	Type	DL1	DL2	DL3	DL4	DL5	DL6	DL7	DL8	DL9	DL10
	Delay	6.437	4.971	5.772	3.218	3.790	2.251	4.310	2.395	2.850	1.720
	Power	230.400	872.300	244.900	137.100	884.200	511.800	216.200	181.700	770.100	670.600

圖 Fig. 11-9 展示了通過 RO 調整 N/P 偏移的核心概念，其具體流程如下：首先，利用 SPICE 模擬獲取目標數據 (如 RO 和 SIDD)，並基於 N/P 變異進行離散網格點仿真分析。接著，構建回歸模型，以精確刻畫大規模樣本的概率密度分佈，從而估算出與晶片平均值匹配的 N/P 偏移值

(需注意，可能存在多個滿足條件的解)。隨後，採用相同方法對 NAND、NOR 和 INV 等所有 DL 進行共通解搜尋，以確定具有最大交集的 N/P 製程配方。最後，將調整結果與量產 (MP) 階段的 WAT 數據進行關聯分析，進一步驗證其有效性。

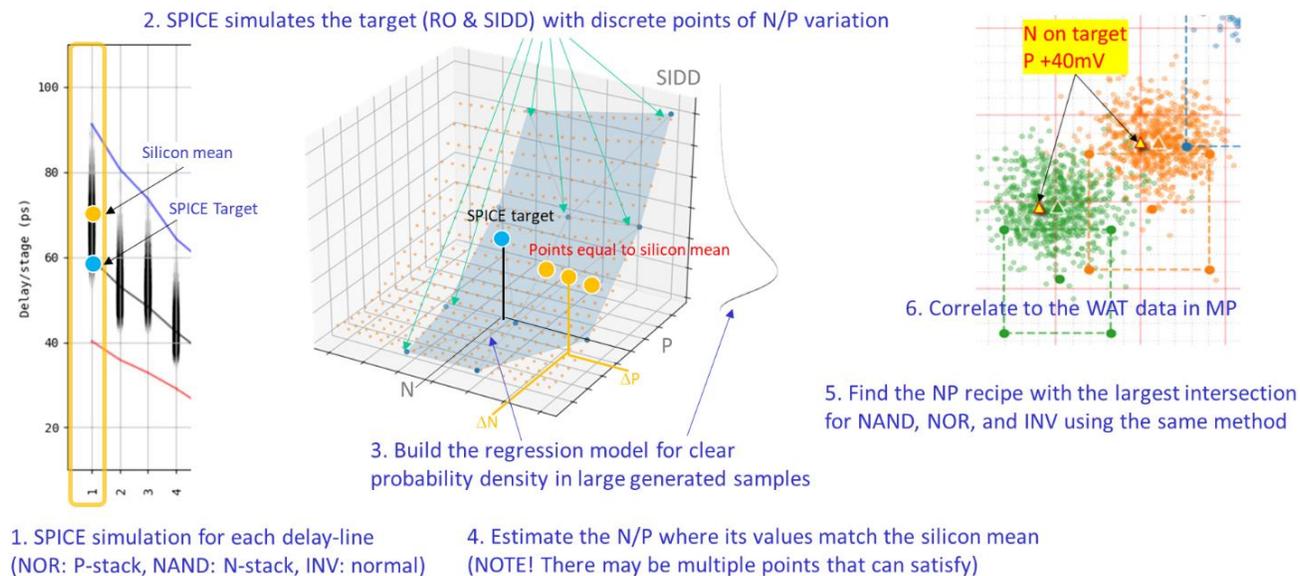


Fig. 11-9 S2S and N/P Bias Prediction

為了更高效地實現該核心思路並提升 N/P 偏移估算的準確性，我們可以進一步採用 MLP 模型進行訓練，以優化回歸模型的表現並提高偏移校正的精度，如圖 Fig. 11-10。

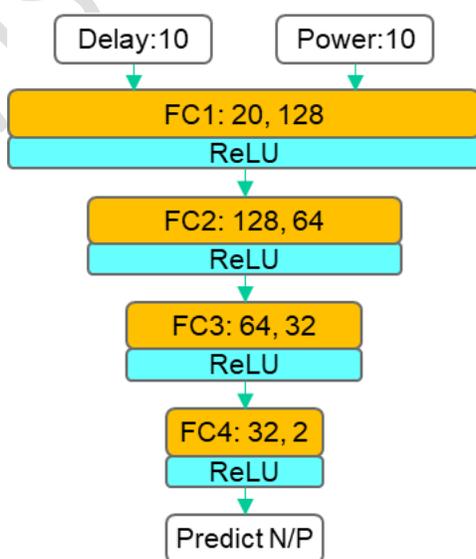


Fig. 11-10 MLP for N/P Bias Prediction

具體而言，我們首先對每條 DL 的少量離散數據進行曲面回歸。例如，表 TABLE IV 顯示了 9 組不同 N/P 偏移調整參數的 SPICE 仿真結果，對應於 3×3 N/P 網格下的不同延遲值。接著，利用這些離散網格數據構建回歸模型，並進一步插值擴展至更高解析度（如 100×100 點），以生成作為訓練集的大量樣本。

TABLE IV SPICE Simulation Grid with N/P Bias

Delay	N	P	DL1	DL2	DL3	DL4	DL5	DL6	DL7	DL8	DL9	DL10
	0	0	0	2.741	2.104	2.732	1.234	1.742	0.853	2.131	0.916	1.373
15	0	0	2.524	1.930	2.314	1.191	1.476	0.821	1.727	0.879	1.108	0.628
15	15	15	3.103	2.372	2.592	1.535	1.665	1.063	1.870	1.138	1.199	0.812
0	-15	0	2.178	1.669	2.343	0.964	1.483	0.662	1.899	0.707	1.228	0.505
0	15	0	3.404	2.615	3.103	1.622	1.995	1.123	2.317	1.198	1.497	0.858
-15	0	0	2.955	2.272	3.139	1.321	2.002	0.912	2.535	0.970	1.650	0.695
-15	-15	0	2.018	1.543	2.024	0.915	1.281	0.628	1.573	0.673	1.010	0.481
-15	15	0	3.691	2.844	3.630	1.708	2.335	1.184	2.812	1.259	1.830	0.903
-15	-15	0	2.335	1.791	2.655	1.014	1.678	0.696	2.241	0.741	1.458	0.530

Power	N	P	DL1	DL2	DL3	DL4	DL5	DL6	DL7	DL8	DL9	DL10
	0	0	0	97.460	367.100	114.800	52.210	399.700	190.700	104.900	67.910	363.300
15	0	0	90.100	337.400	97.050	49.900	337.700	183.500	85.250	65.350	292.400	238.000
15	15	15	111.000	415.700	108.500	64.640	379.700	238.600	91.960	85.050	315.100	309.400
0	-15	0	77.330	290.000	98.690	40.030	341.100	146.400	94.450	52.000	327.100	189.600
0	15	0	121.500	457.500	130.400	68.070	457.500	251.400	114.600	89.220	396.000	326.000
-15	0	0	104.900	395.300	132.400	55.000	461.600	202.100	126.300	71.480	440.700	261.500
-15	-15	0	71.840	268.900	85.040	38.150	293.700	139.600	77.860	49.750	267.300	181.100
-15	15	0	131.500	496.400	153.000	71.490	537.700	264.000	139.800	93.340	487.200	341.900
-15	-15	0	82.700	310.300	112.000	41.970	387.200	153.400	112.000	54.310	390.700	198.100

最後檢驗目標與推論結果，若多個目標之間無法存在 N/P 交集，則問題無解，如圖 Fig. 11-11。為了涵蓋超出 N/P 變異標準差範圍的解，我們可能需要將原始值域擴展至 ±50mV 甚至更大。

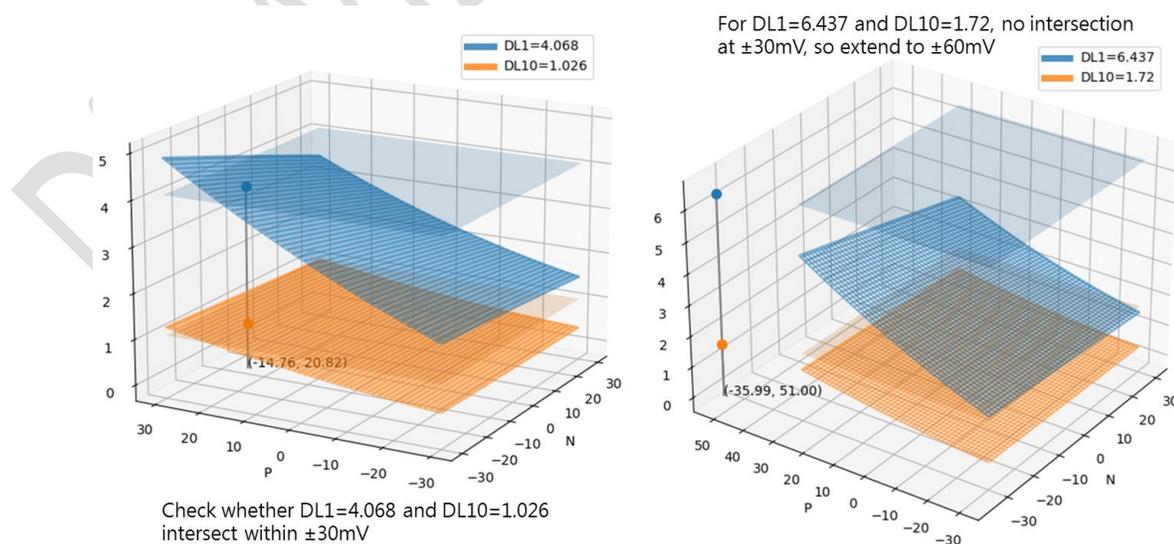


Fig. 11-11 Search Range Exploration and Data Augmentation

在一個包含 10 條 DL (DL1-DL10) 的 RO 實例中，我們可以通過更廣泛的回歸曲面檢查 DL1-DL10 的目標值 (z 軸平面)，以確認是否存在可能的 N/P 配方最佳解 (即回歸曲面與目標值相交)。如圖 Fig. 11-12 所示，在原始數據集中，目標延遲值的 $\pm 30\text{mV}$ 範圍內未能找到交點，因此需要進行數據擴增。為了完成這一目標任務，數據範圍可能需要擴展至至少 $\pm 60\text{mV}$ 。

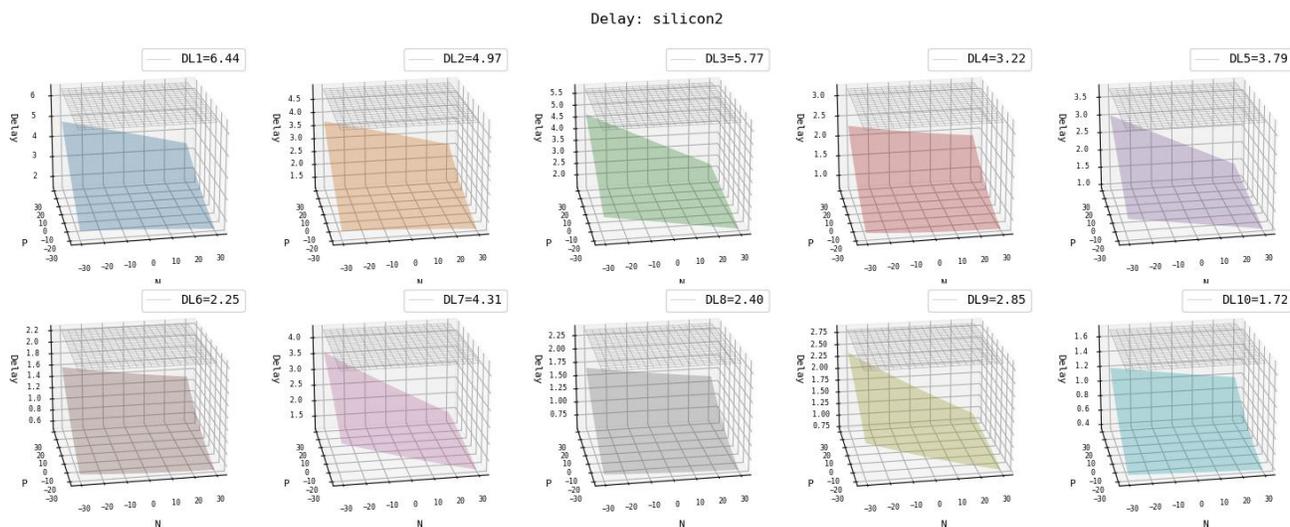


Fig. 11-12 Target N/P Feasibility Assessment

圖 Fig. 11-13 顯示針對表 TABLE III 中目標 Silicon2 的推論結果，呈現同時滿足 10 條延遲線延遲與功耗條件的 N/P Bias 最佳解。

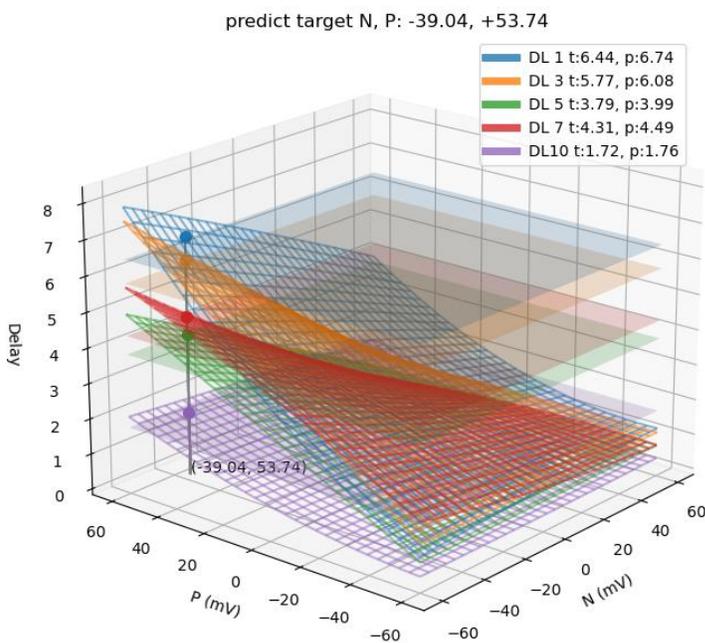


Fig. 11-13 One-shot SPICE-Silicon N/P Bias Prediction

11.3. 高效蒙地卡羅仿真近似 (High-Fidelity Generative Monte Approximation, H Σ -GMA)

11.3.1. 傳統 Monte Carlo 方法的限制

傳統 SPICE Monte Carlo 模擬在尋找高決策邊界時，面臨以下挑戰，導致過程極為耗時且困難：

- **樣本稀疏性**：高 sigma 區域屬於極低概率事件，隨機取樣難以有效覆蓋，導致樣本不足或分佈不均，無法提供充分的統計信息來準確估計高 sigma 邊界。
- **計算資源需求**：由於 Monte Carlo 仿真的收斂速度慢，而高 sigma 分析需要極大量樣本來降低統計誤差，導致計算資源消耗巨大，模擬時間顯著增加，特別是在大規模電路中。
- **參數相關性與非線性效應**：在高 sigma 區域，參數間的相關性與非線性行為更加顯著，傳統模擬方法難以捕捉這些複雜特性，進一步增加結果的不確定性。

11.3.2. 生成式神經網路的創新應用

當高維超曲面（網絡 θ ）投影至低維空間（如 3D）時，可能因拓撲變形或資訊遺失導致連續性破壞。如圖 Fig. 11-14 所示，降維後的圓錐超曲面可能失去原有的幾何結構。然而，若數據內在關係（如單調性）得以保留，cVAE 或 GAN 等生成模型可學習更平滑且具可解釋性的潛在分佈。

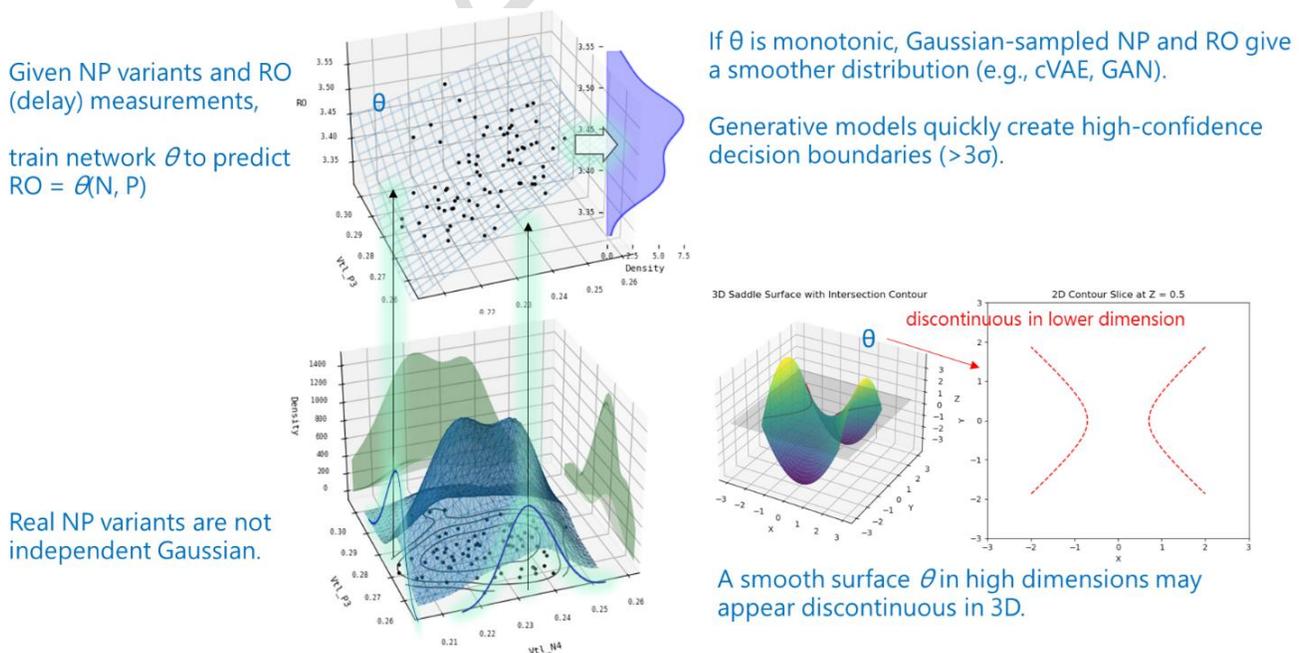


Fig. 11-14 Dimensionality Reduction and Smoothness Preservation

在給定 N/P 變異和 RO (延遲) 測量值的情況下，我們可以訓練網絡 θ 來預測 $RO=\theta(N, P)$ 。然而，實際的 N/P 變異並非獨立的高斯分佈。若 θ 是單調的，則高斯 N/P 取樣能夠預測出與原始分佈相似的平滑 RO 分佈，這一特性在像 cVAE 和 GAN 等生成模型中得到了體現。生成式模型的優勢在於，它們能快速創建高置信度的決策邊界 ($>3\sigma$)，從而顯著提升預測準確性與穩定性。

Example 11-1 XOR2D1 Power Modeling and KDE Density

```
# Model the power distribution of XOR2D1 using a two-component Gaussian mixture.
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import gaussian_kde, norm

def sigma_percentage(self, sigma):
    '''return left and right % boundary based on the specified sigma value'''
    return stats.norm.cdf((-sigma, sigma))*100

def featureKDE(v, res=100):
    kde = gaussian_kde(v)
    t = np.linspace(min(v)-v.std(),max(v)+v.std(),res)
    p = kde(t)
    return t,p

low, high = 4.4e-6, 4.8e-6 # Data range
mean1, std1, weight1 = 4.57e-6, 3e-8, 0.4 # low-density Gaussian
mean2, std2, weight2 = 4.63e-6, 3e-8, 0.6 # high-density Gaussian

# Generate samples for each Gaussian
n_samples = 1000
n_samples1 = int(n_samples * weight1)
n_samples2 = n_samples - n_samples1

# Combine and clip the samples within the specified range
np.random.seed(42)
samples1 = np.random.normal(mean1, std1, n_samples1)
samples2 = np.random.normal(mean2, std2, n_samples2)
samples = np.concatenate([samples1, samples2])
```

Example 11-2 Gaussian Mixture Model (GMM)

```
# Gaussian Mixture Model (GMM) Sample Generator
from sklearn.mixture import GaussianMixture

def GMM(v, n_samples=1000, n_components=2, scale=1e2):
    gmm = GaussianMixture(n_components=n_components, covariance_type='full', random_state=0)
    gmm.fit(v*scale)
    dv, _ = gmm.sample(n_samples=n_samples) # generated samples
    return dv/scale

fake = GMM(samples.reshape(-1,1), n_samples=len(samples)*50, scale=1e8).reshape(-1)

# statistical quantization
q1 = np.quantile(samples, q=sigma_percentage(3.0)/100)
q2 = np.quantile(fake, q=sigma_percentage(4.5)/100)
t1, p1 = featureKDE(samples)
t2, p2 = featureKDE(fake)
```

```

# visualization
plt.figure(figsize=(8,5))
plt.title('XOR2D1 Power Distribution')
plt.hist(samples, bins=50, density=True, alpha=0.4, color='skyblue', label=f'Histogram
({len(samples):,})')
plt.hist(fake, bins=50, density=True, alpha=0.4, color='orange', label=f'Fake ({len(fake):,})')
plt.plot(t1, p1, c='k', ls='--', lw=2, alpha=0.5, label='PDF (real)')
plt.plot(t2, p2, c='b', lw=3, alpha=0.5, label='PDF (fake)')
plt.axvline(q1[0], c='r', alpha=0.3)
plt.axvline(q1[1], c='r', alpha=0.3, label=f'Real 3$\sigma$: {q1[0]:.2e}, {q1[1]:.2e}')
plt.axvline(q2[0], c='b', alpha=0.3)
plt.axvline(q2[1], c='b', alpha=0.3, label=f'GMM 4.5$\sigma$: {q2[0]:.2e}, {q2[1]:.2e}')
plt.xlabel('Value')
plt.ylabel('Density')
plt.grid(which='major',linestyle='-',zorder=0, alpha=0.5)
plt.grid(which='minor',linestyle=':',zorder=0, alpha=0.5)
plt.minorticks_on()
plt.legend()
plt.tight_layout()

```

Example 11-1 使用兩個高斯分佈模擬 XOR2D1 單元在 1,000 次 Monte Carlo 仿真後的功耗數據分佈。Example 11-2 則透過高斯混合模型 (Gaussian Mixture Model, GMM) 對該數據進行建模，並生成 50,000 筆樣本。程式設置了兩個高斯分佈的均值、標準差與權重，並根據這些參數生成樣本數據。觀察 XOR2D1 原始功耗分佈及其 3σ 邊界，並與模擬數據的 4.5σ 邊界進行比較，再利用核密度估計 (KDE) 方法生成平滑的機率密度函數 (PDF)，以便進行視覺化呈現與分析。

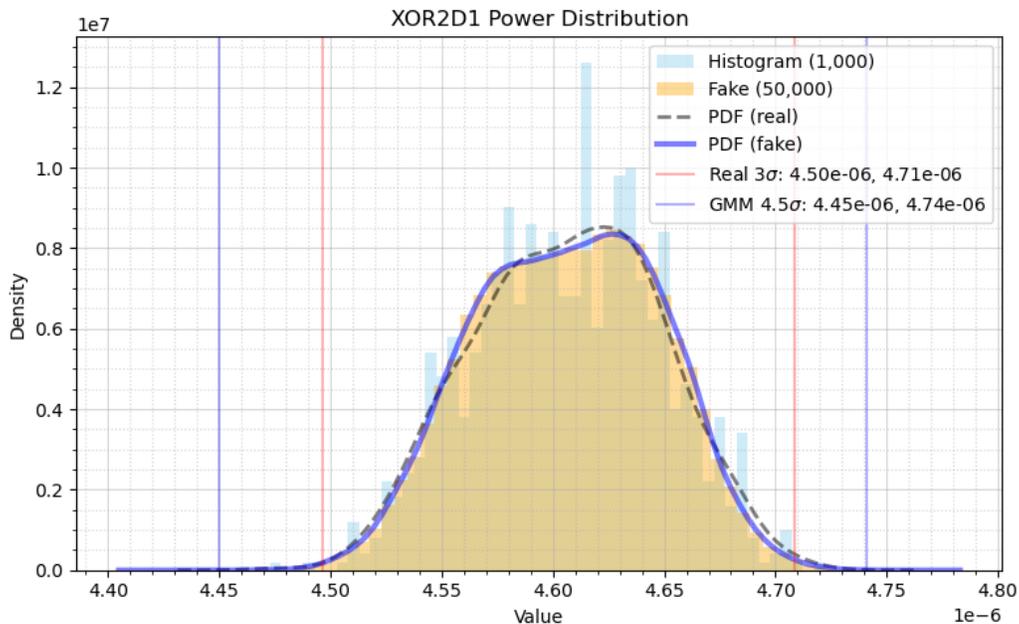


Fig. 11-15 XOR2D1 Power Probability Density Distribution

第12章 結論與展望

如圖 Fig. 12-1 所示，我們以 U-Net 結構類比半導體產業的生產流程與優化機制。在終端產品層面，目標是提升產能與產品競爭力，並透過數據積累的正向反饋來持續優化晶片設計，以實現極致能效。結合晶片監測與機器學習，DTCO.ML 採用跨數據維度分析的方法，構建高效的產能與能效優化框架，已成為半導體領域的重要研究方向。

然而，現有監測 IP 與測試數據仍缺乏標準化，跨領域數據的獲取與交互存在障礙，傳統建模方法容易偏離真實情況，而稀疏測試數據則可能導致決策偏差。此外，高信任度的仿真模擬雖然能提高準確性，卻往往耗時且需要進一步優化。在此背景下，生成式 AI 的引入有望為半導體領域帶來突破性進展，DTCO.GenAI 將成為推動整體解決方案升級的重要發展方向。

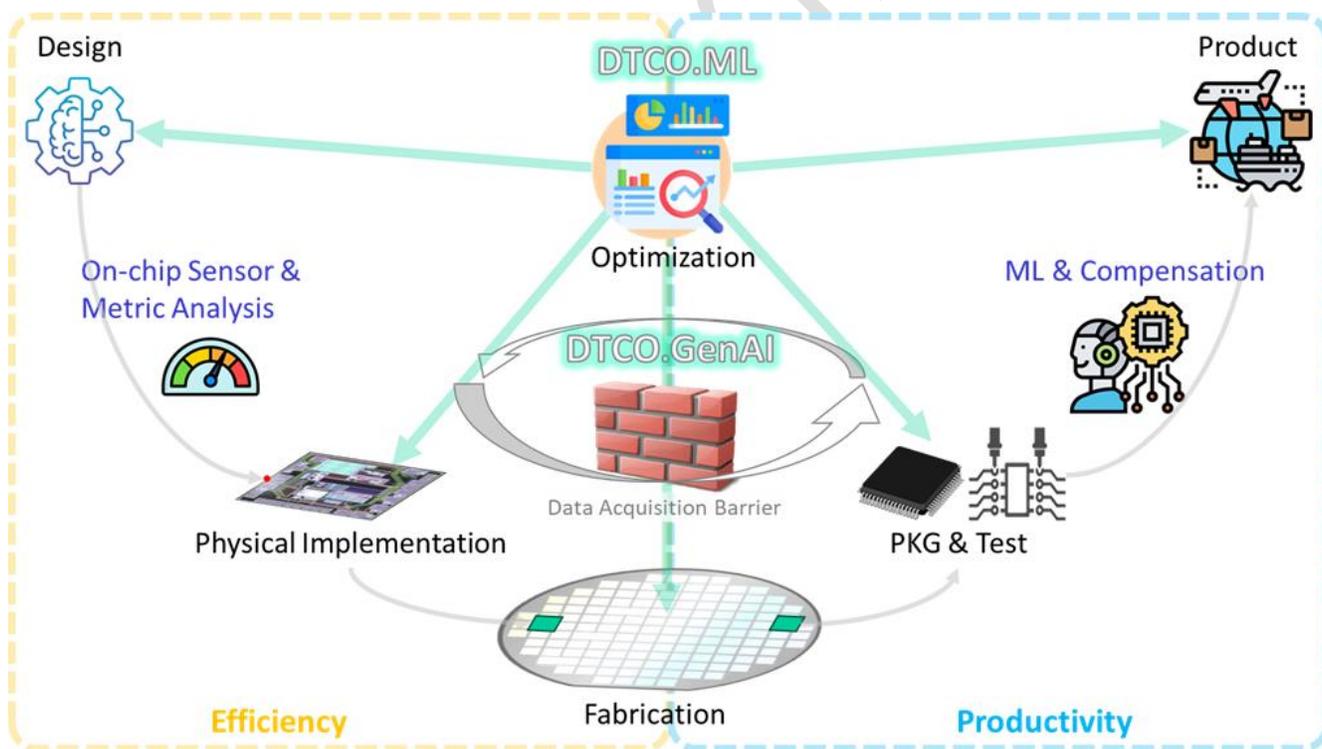


Fig. 12-1 半導體產業與晶片設計創新

12.1. 機器學習與 AI 賦能 DTCO : 革新晶片設計與製程優化 (DTCO.ML™)

機器學習融合 DTCO，可精確指導製程配方，降低測試成本並提升產品品質。其主要應用包括：

- **製程參數與設計配方優化**：利用機器學習模型，分析製程參數對晶片性能的影響，並提供最佳的參數組合建議。
- **元件設計與實體設計流程優化**：建立設計參數與製程條件之間的數學模型，實現設計與製程的動態調整，縮短開發週期。
- **分箱策略生成與優化**：基於歷史數據訓練的預測模型精準預測晶片性能分佈，並自動生成與優化分箱策略。系統級補償策略顯著提升產能、可靠性與競爭力。

12.2. 生成式 AI 驅動優化的未來趨勢 (DTCO.GenAI™)

機器學習模型透過大量晶片數據訓練，解析製程參數、性能變異性與生產效率之間的關聯。虛擬晶片數據 (Virtual Silicon Data) 涵蓋元件性能、電性特徵與製程分佈，已成為 DTCO 的關鍵工具。其主要優勢包括：

- **克服資料獲取障礙**：虛擬晶片反映真實製程與量產測試數據，透過生成模型壓縮數據並保障資料機密性，從而有效解決資料獲取問題。
- **落實跨領域協作**：提供標準化、可共享的數據平台，促進晶片設計與製程團隊的高效協作。
- **提升產品優化能力**：利用虛擬數據進行性能預測與製程模擬，實現快速迭代與優化。

12.3. 創新 EDA 開發與未來展望

AI 與機器學習正重塑 DTCO，提升效率與精確度，為智能化 EDA 工具奠定基礎。虛擬晶片生成技術結合 GAN 與擴散模型，精準建模微觀數據分佈與系統級變異性，優化設計餘量與晶片能效。同時，生成式 AI 推動 WAT 數據超解析度技術，加速 SPICE Monte Carlo 近似運算，縮短驗證時間並提升決策信心。這些突破正驅動新一代 DTCO EDA 工具，提升生產效率、降低成本並優化品質，推動半導體邁向更智能、高效的未來。

附錄

開源資源表列

DTCO Framework

<https://github.com/dipsci/DTCO>

<https://pypi.org/project/DTCO/0.1.4>

Google Colab: libMetric

<https://colab.research.google.com/drive/1KIJpLU4oZM4ITJ8XgJ7e947N6QyIlp9H>

https://colab.research.google.com/drive/16Y2aNTqC_v2vtJCqwqgpMn0LQk_eLxWA

<https://colab.research.google.com/drive/1pj0fnW09y2Jh7XkOHn5fK60U7oN04Vus>

Google Colab: Generated Model (Virtual Silicon)

<https://colab.research.google.com/drive/1oags2cgVHDtQ2UECVjbFYKxL8RWvxjmq>

<https://colab.research.google.com/drive/1JNai0036dDg2silpaOwmtf6WveUa1z-k>

參考文獻表列

[1] LBR: Design Dependent Mega Cell Methodology for Area and Power Optimization, DAC, 2020

[2] Micro-Architecture Optimization for Low-Power Bitcoin Mining ASICs, VLSI-DAT, 2019

[3] Multi-Feature Data Generation for Design Technology Co-Optimization A Study on WAT and CP, FC, 2023

[4] Application of Generative Adversarial Networks for Virtual Silicon Data Generation and Design-Technology Co-Optimization: A Study on WAT and CP, IEEE-Access, 2024

[5] A Diffusion Model-Based Methodology for Multi-Feature Wafer Data Generation, IEEE-TSM, 2025

專業術語表

A

AOCV : Advanced On-Chip Variation
AVS : Adaptive Voltage Scaling

B

BCE: Binary Cross Entropy
Binning : Chip classification by electrical metrics
BIST : Built-In Self-Test
BTC : Bitcoin

C

CCS : Composite Current Source (Liberty)
CDF : Cumulative Distribution Function
CG : Clock Gating
CP: Chip Probe
CPO : Co-Packaged Optics
CTS : Clock Tree Synthesis
cVAE : Conditional Variational Autoencoder

D

DDPM : Denoising Diffusion Probabilistic Model
DFF : D Flip-Flop
DFT : Design for Test
DL : Delay Line
DTCO : Design-Technology Co-Optimization
DUE : Device Under Extraction
DVFS : Dynamic Voltage and Frequency Scaling

E

EDA : Electronic Design Automation
ERA : Early Rail Analysis

F

FID : Frechet Inception Distance
FT : Final Test

G

GAN : Generative Adversarial Network
GenAI : Generative AI
GMM : Gaussian Mixture Model
GRO : Grid RO

H

HITL : Human-in-the-Loop

I

Isat : WAT Saturation Current

J

JS : Jensen-Shannon Divergence

K

KDE : Kernel Density Estimation

L

LEO : Low Earth Orbit Satellite
LDO : Low Dropout Regulator
LS : Least Squares Regression
LSC : LS Coefficient
LVF : Liberty Variation Format (Liberty)
LVS : Layout Versus Schematic

M

MBIST : Memory Built-In Self-Test
MCU : Microcontroller Unit
ML : Machine Learning
MVN : Multivariate Normal Distribution

N

NLDM : Non-Linear Device Model (Liberty)
N/P : N-type vs. P-type Semiconductor

O

OCM : On-chip Monitors
OCSB : On-chip Self-Binning
OCV : On-chip Variation

P

PDF : Probability Density Function
PL/PG : Pulse-Latch / Pulse Generator
PPA : Power, Performance, and Area
PVT : Process, Voltage, Temperature

R

Re-K : Re-characterization
RO : Ring Oscillator

S

S2S : SPICE-to-Silicon
SIDD : Static IDD, CP Leakage Current
SPICE : Simulation Program with Integrated Circuit Emphasis
SLT: System-level Test
STA : Static Timing Analysis

U

UID : Unique Identifier
ULE : Ultra-low Energy

V

Vsat : WAT Saturation Voltage
Vtl : WAT Threshold Voltage Low

W

WAT: Wafer Acceptance Test
WAT-SR : WAT Super Resolution
WID : Within Die