



後摩爾定律時代的新興量子計算技術

退火計算

作者：于濂波、陳志宇、張慶瑞

作者簡介

于濂波是臺灣大學電機博士，目前任職於資策會智慧系統研究所。

陳志宇是臺灣大學物理博士，目前是中原大學電機資訊學院人工智慧應用學程助理教授。

張慶瑞是臺灣大學物理學系特聘教授。

退火計算 (annealing) 是一項專以解決大型、複雜、難解組合優化問題的新型態計算技術，主要包括量子退火和數位退火兩類技術，其中數位退火具有新計算模式、領域特化架構及橋接量子計算等特性，不僅是企業跨越數位與量子計算鴻溝的一項重要變革戰略，也是後摩爾定律時代一項極具商業價值的新型態高效計算 (HPC) 技術，有機會為臺灣資訊和半導體產業帶來新契機。本文將介紹退火計算關鍵軟、硬體技術發展現況。

一、概述

在後摩爾定律時代 (post-Moore's law era) 過去依賴計算堆棧「底部」半導體微縮、封裝、材料等技術來提升計算性能的方式可能將難以再有大幅成長和收益，並將逐漸轉移至「頂部」軟體、演算法、新計算模型 (例如量子計算) 和領域特化架構 (domain-specific architecture) 等創新，從而改變歷史發展趨勢 (圖 1)。

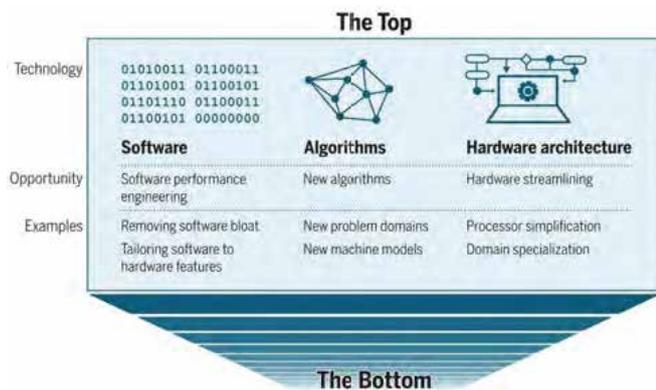


圖 1：後摩爾定律時代的計算性能提升趨勢。(C. Leiserson et al., Science 368, 2020)

另一方面，在量子計算實用化之前，基於「量子電腦特殊用途」——解決大型、複雜組合優化問題 (combinatorial optimization problem) 及「特殊用途量子電腦」——量子退火電腦 (quantum annealer) 或數位退火電腦 (digital annealer)，尤其是後者，兩項策略所形成的「量子啟發數位退火計算」已成為企業跨越數位與量子計算鴻溝的一項重要變革戰略 (圖 2)。此外，因其具有以下特性，故也成為後摩爾定律時代一項極具商業價值的新型態高效計算技術。

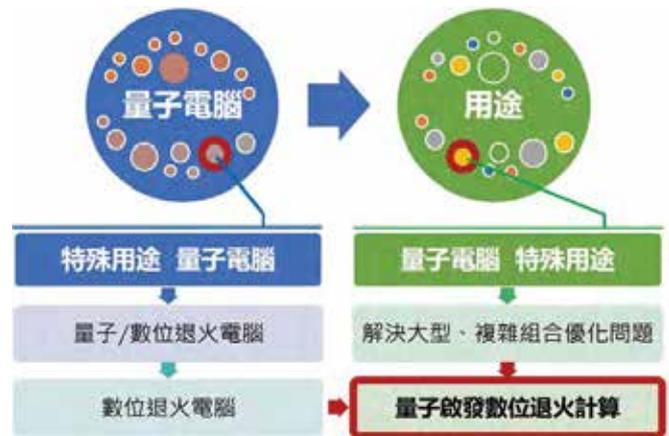


圖 2：銜接現在數位計算與未來量子計算鴻溝的策略。

新計算模式：運用量子力學原理的新數位計算模式。

領域特化架構：針對優化問題易辛模型 (Ising model) 特化的非傳統馮諾曼 (non-von Neumann architecture)。

架構橋接量子計算：採用量子啟發演算法，橋接 (bridging) 數位與量子計算。

臺灣資訊和半導體產業正可借力使力，布局數

位退火矽智財 (soft IP core) 或晶片 (annealing on chip)、數位退火專用加速器 (application-specific accelerator)、數位退火雲端服務等量子啟發計算新興市場。

二、退火計算原理與應用

退火計算靈感源自傳統冶金熱退火 (thermal annealing)，將金屬材料加熱到高於再結晶溫度，再緩慢控制降溫，造成材料微結構改變而最終趨向能量最低、最穩定的狀態——將優化問題成本函數 (cost function) 轉換為物理系統能量函數 (Hamiltonian)，接著在能量函數所定義的超大狀態空間 (configuration space)，運用退火原理跳脫局部最低能量狀態，搜尋全域最低或近似最低能量狀態 (圖 3)，最終轉換為原優化問題的全域最優解或最優近似解。

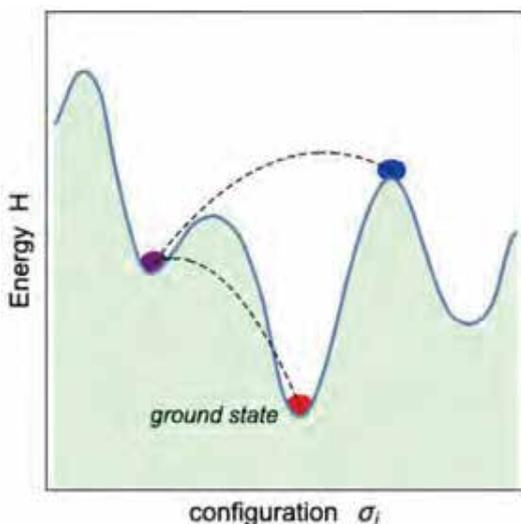


圖 3：退火計算搜尋最低能量狀態 (紅色位置) 示意圖。

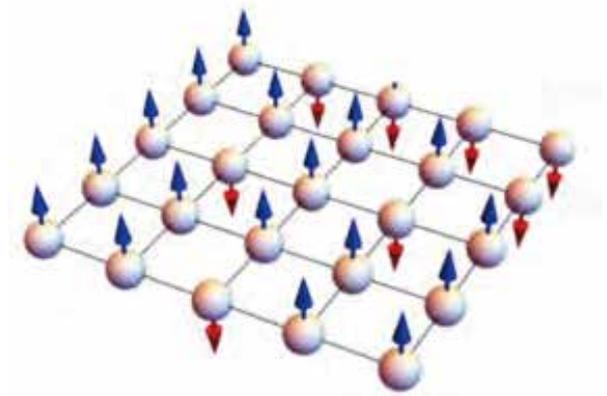


圖 4：二維易辛模型示意圖。

退火計算主要包括數位退火 (DA) 和量子退火 (QA) 兩類，簡述如下：

數位退火 (DA)：主要分為模擬退火 (simulated annealing, SA) 和模擬量子退火 (simulated quantum annealing, SQA) 2 類。SA 是一種尋找全域最優解或最優近似解的古典概率演算法，由柯克派崔克 (Scott Kirkpatrick) 等人於 1983 年提出^①，透過馬可夫鏈蒙地卡羅 (Markov Chain Monte Carlo) 方法模擬熱退火隨溫度變化的熱波動 (thermal fluctuation)，使對應優化問題的易辛模型物理系統得以跳脫局部最低能量狀態。所謂易辛模型是由一群相連「自旋」 (spin) 節點組成的數學模型 (圖 4)，自旋處於向上 (+1) 或向下 (-1) 2 種狀態之一，相鄰自旋的耦合強度 (coupling) 可因待解問題性質而調整。SQA 則是透過路徑積分蒙地卡羅模擬 (path-integral Monte

① 編註：Kirkpatrick, Scott, C. Daniel Gelatt, and Mario P. Vecchi, "Optimization by simulated annealing", *Science* Vol 220 Issue 4598 (1983): 671 ~ 680。

Carlo simulation) 或鈴木／截特分解法 (Suzuki-Trotter decomposition) 的量子蒙地卡羅模擬 (quantum Monte Carlo simulation) 等方法模擬量子退火，搜尋全域最優解或最優近似解。

量子退火 (QA)：受到 SA 引入熱波動的啟發，QA 引入量子波動 (quantum fluctuation)，透過量子穿隧 (quantum tunneling) 也可以尋找全域最低能量狀態。其概念由阿波羅尼 (Bruno Apolloni) 等人於 1988 年提出，之後西森秀稔 (Hidetoshi Nishimori) 和門脅正史 (Tadashi Kadowaki) 在 1998 年論文中首次將其模型製定為目前的形式。同一時期，法爾希 (Edward Farhi) 等人也在 2001 年提出了一篇與 QA 密切相關的絕熱量子計算 (adiabatic quantum computing, AQC) 論文。

三、退火計算軟、硬體關鍵技術

退火計算軟、硬體關鍵技術主要包括 QUBO ^② 解法器 (QUBO solver) 和退火處理器 (annealer)，說明如下：

1. QUBO 解法器

QUBO 解法器架構和代表性新創公司如圖 5 所示，主要軟體元件概述如下：

QUBO/ 易辛模型制定 (formulation)：將成本函數編碼成對應的易辛模型能量函數 (Hamiltonian) 或對等 QUBO 模型函數 ^③。QUBO 模型具備與退火計算處理器架構匹配、許多優化問題可以映射 (mapping) 到它、易於被分解處理再合併求解等

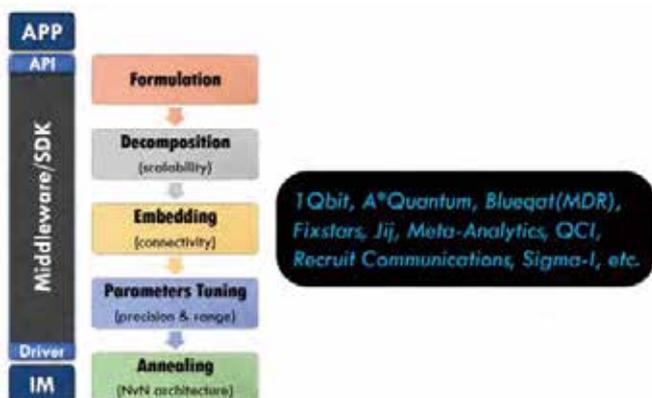


圖 5：(左) QUBO 解法器架構和 (右) 主要新創公司。

優點，因此很適合作為發展自動化工具的中間表示形式 (intermediate representation)。

QUBO 分解：受限於 (量子) 位元數有限及連接性 (connectivity) 不足的退火處理器，大型優化問題的 QUBO 必須分解成小到足以嵌入，並盡可能大到足以滿足高精度的解。

QUBO 嵌入 (embedding)：將待解問題的 QUBO 邏輯圖映射到退火處理器架構實體圖的一個次圖 (graph minor) ^④。QUBO 一般都採用完整圖嵌入 (complete graph embedding) 方法，由於目前退火處理器的連接性不足，因此嵌入會產生硬體資源利用低的問題。例如具備 5,000+ qubits Pegasus 架構的 D-Wave Advantage 量子退火電腦最多只能嵌入 185 節點的 clique (完整圖的節點集合)。

^② 註：quadratic unconstrained binary optimization，二次無約束二進制優化。

^③ 註：QUBO 變數值為 0/1，易辛變數值為 1/-1 的變數；易辛變數 z_i 與 QUBO 變數 x_i 關係式為 $z_i = 2x_i - 1$ 。

^④ 註：如果圖 G 可以透過刪除邊和頂點或收縮邊得到無向圖 (undirected graph) H，則稱 H 為 G 的次圖。

Who	Algo.	ASIC	FPGA	GPU/CPU/VE
	SA (DA)	<ul style="list-style-type: none"> • 8.192/100K spins • Fully-coupled 		
	SQA (MA)	<ul style="list-style-type: none"> • 144K(9x16K)spins • King's graph,40-nm 	<ul style="list-style-type: none"> • 25 FPGAs • King's graph 	<ul style="list-style-type: none"> • 256,000 spins • King's graph
	a/b/d SB (SBM)		<ul style="list-style-type: none"> • 4,096 spins/8-FPGA • Fully-coupled 	<ul style="list-style-type: none"> • 100,000 spins • Fully-coupled
	SA			<ul style="list-style-type: none"> • VE-based • (details unknown)
	SA (QLM)			<ul style="list-style-type: none"> • GPU-based • (details unknown)
	SA (SCA)	<ul style="list-style-type: none"> • 512 spins • Fully-coupled,65-nm 		
	SQA		<ul style="list-style-type: none"> • 8,192 spins, 32 Trotters • Fully-coupled(OpenCL) 	<ul style="list-style-type: none"> • 32,768 spins, 32 Trotters • Fully-coupled

表格 1：數位退火處理器發展現況（僅部分列舉）。SA：simulated annealing，SQA：simulated quantum annealing，a/b/d SBM：simulated bifurcation machine，SCA：stochastic cellular automata，MA：momentum annealing，VE：vector engine，QLM：quantum learning machine。

2. 退火處理器

退火處理器主要包括量子退火處理器和數位退火處理器兩類。前者商轉最成功的是加拿大 D-Wave 公司，已發展至第 5 代 5,000+ qubits Pegasus 架構 D-Wave Advantage 量子退火電腦。日本電氣公司（NEC）則與奧地利 ParityQC 合作，預計 2023 年推出大型 QA 處理器。後者採用傳統數位科技，目前則是百家爭鳴，尤其日本在數位技術發展一直處於全球領先地位，歐洲則以 Atos 最為積極，目前商轉最成功的是日本富士通（Fujitsu），已發展至第 3 代 100,000 位元全連結（fully-connected）架構 DAU（digital annealing unit）。其他還有同樣基於易辛模型設計的退火處理器，例如使用光學參數震盪器建構光學神經網路系統的光學退火處理器（optical annealer），代表產品為日本 NTT 和史丹福大學的 CIM（Coherent Ising Machine），以及利用數值分析方法模擬分枝現象（bifurcation phenomena）的加速器，代表產品為日本東芝（Toshiba）的

SBM（Simulated Bifurcation Machine）。在中國方面，北京玻色量子科技公司宣稱近期即可推出百萬位元相干量子人工智慧協同處理器（coherent quantum AI coprocessor），也加入競爭行列。

現階段 QA 處理器與通用型量子處理器一樣，受限於量子位元數、錯誤率、稀疏連結性等問題，實務應用上仍面臨較多挑戰，而 DA 處理器採用傳統 multi-core CPU/GPU/FPGA/ASIC 等半導體處理器技術，具有位元數易擴充、零錯誤率、高連結性以及符合數據中心溫度或室溫要求等優點。尤其是 ASIC/ FPGA 可針對不同應用需求客製化設計或重新配置硬體架構，成為高成本效益的應用加速器或協同處理器，除可運用在雲端計算環境，也可運用在受到操作環境或資源限制的邊緣計算（edge computing）環境中，更有助於探索更廣泛的應用。此外，目前在 DA 發展的演算法、軟體或應用也可輕易地移植到未來的 QA 上，因此短期 DA 更具產業實用價值。茲列舉部分具代表性的研發單位產品（含 SBM）現況如表格 1。

四、臺灣大學與日本富士通數位退火應用研究案

日本富士通自 2020 年 9 月至 2021 年 2 月提供臺大團隊免費使用 8192-bit DAU 數位退火雲端服務，進行 COVID-19 相關優化議題的應用研究，茲說明如下：

1. 醫護人員排班問題

排班或排程問題 (scheduling problem) 研究如何用最少成本 (人數、趟次、時間、費用等) 讓有限資源發揮最大效益，應用範圍不限醫護人員，舉凡產線人員、士官兵、貨運司機等皆適用，主要差別在於應用場景及限制條件不同。針對 8 位護理人員，3 項限制條件，DAU 計算出來 2 週班表如圖 6。

day \ nurse	1	2	3	4	5	6	7	8	9	10	11	12	13	14
a														
b														
c														
d														
e														
f														
g														
h														

圖 6：醫護人員 2 週班表。(圖片來源：臺大學生卓建宏)

2. 醫療資源分配問題

當重大疾病 (例如 COVID-19) 爆發時，如何投放有限醫療資源 (例如個人防護裝備、疫苗) 達到阻止疾病傳播最大程度是非常重要的分配優化問題。以疫苗資源分配模擬醫院為例，DAU 計算的投放網絡如圖 7，其中藍

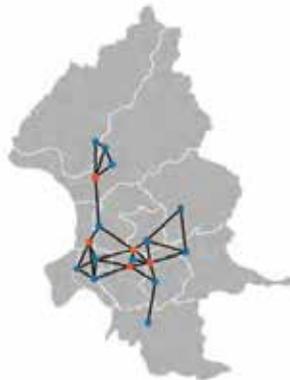


圖 7：模擬醫院投放網絡。(圖片來源：臺大學生鍾皓宇)

色點為醫院位置，紅色點代表優先投放醫院，醫院間連結為假想。

3. 食物遞送問題

食物遞送是古典路徑優化問題 (routing problem) 的衍生問題，其限制條件包括外送員車子載貨量有限、遞送過程須先去餐廳再去顧客家，以及外送員不需再返回出發點。圖 8 優化路徑 0-1-2-3-4 (8.0 單位長) 比另一含橘色箭頭路徑 0-1-3-2-4 (8.5 單位長) 短。

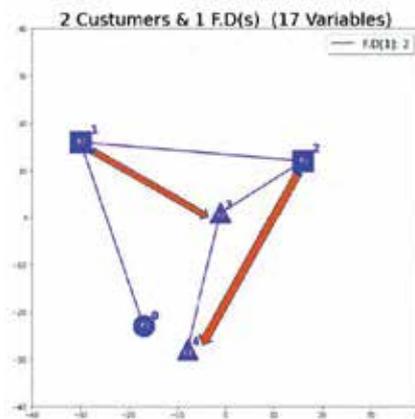


圖 8：有順序限制的路徑優化圖。(圖片來源：臺大學生蘇政維)

五、臺灣在量子啟發計算的機會與挑戰

量子計算已成為國際新軍備競賽，然國內仍處萌芽階段。近期在臺灣量子電腦暨資訊科技協會 (TAQCIT) 及各方人士積極參與推動下，已逐漸吸引電子、半導體、石化、金融、生醫等企業從被動參與到主動洽詢，並促成國際數位退火領導業

者日本富士通在國內探尋策略夥伴，迎來契機。目前國內投入量子啟發計算研究的法人單位有資訊工業策進會（資策會）與工業技術研究院（工研院），大學則有臺灣大學，中原大學、長庚大學等。臺灣大學（電機系、資工系）自 108 年底起結合資策會執行科技部為期 3 年的專案計畫，開始探索 DA 軟、硬體技術。另外，如前述，臺灣大學（物理系）自 109 年下半年起已與日本富士通合作探索優化應用，資策會也自 110 年起加入，並進一步結合跨領域（物理系、電機系、資工系）資源研發 DA 處理器。中原大學則透過與前述團隊合作，研究 DA 軟體並規劃研製 FPGA DA 處理器。長庚大學則希望加入類比退火線路的研究。目前學研單位已逐漸形成一個量子啟發計算暨應用聯盟（圖 9）。



圖 9：臺灣量子啟發計算暨應用聯盟。

量子啟發數位退火技術可應用在雲端和邊緣計算環境，解決金融、化學、人工智慧、先進製造、運輸物流等領域的大型、複雜、難解的組合優化問題，加上具有新計算模式、領域特化架構及橋接量子計算等特性，不僅是企業跨越數位與量子計算鴻

溝的一項重要變革戰略，也是後摩爾定律時代一項極具商業價值的新型態高效計算技術。臺灣資訊和半導體產業正可借力使力，布局以下新興市場：

- **數位退火矽智財或晶片**：應用於 IC 設計、IC 設計服務、矽智財、電子設計自動化（EDA）等半導體公司。
- **數位退火專用加速器**：專用型優化功能伺服器（appliance）或領域特化架構加速器／裝置供應商或系統商。
- **數位退火雲端服務**：提供各領域用戶導入量子最佳化應用方案。

臺灣量子電腦暨資訊科技協會規劃後續與臺灣半導體協會（TSIA）合作，希望透過協會及會員廠商參與，共同為臺灣半導體與量子計算產業的未來努力，培養半導體產業進入量子計算的人才，並在量子啟發計算新興市場找到適當切入點與制高優勢。◎

本文出處

本文原刊登於《台灣半導體產業協會簡訊》NO. 97 July 第 2 ~ 7 頁。本刊感謝作者與《台灣半導體產業協會簡訊》同意轉載。

延伸閱讀

- ▶ 陳奕廷〈機器學習的量子力學方法〉，科學 Online- 國立臺灣大學 <https://highscope.ch.ntu.edu.tw/wordpress/?p=78924>
- ▶ 于濂波〈量子啟發數位退火計算技術及應用〉，資策會 <https://www.youtube.com/watch?v=zLJY02QD-YQ>