機器學習應用於半導體先進製程三維多參數結構量測

周柏文/顧逸霞

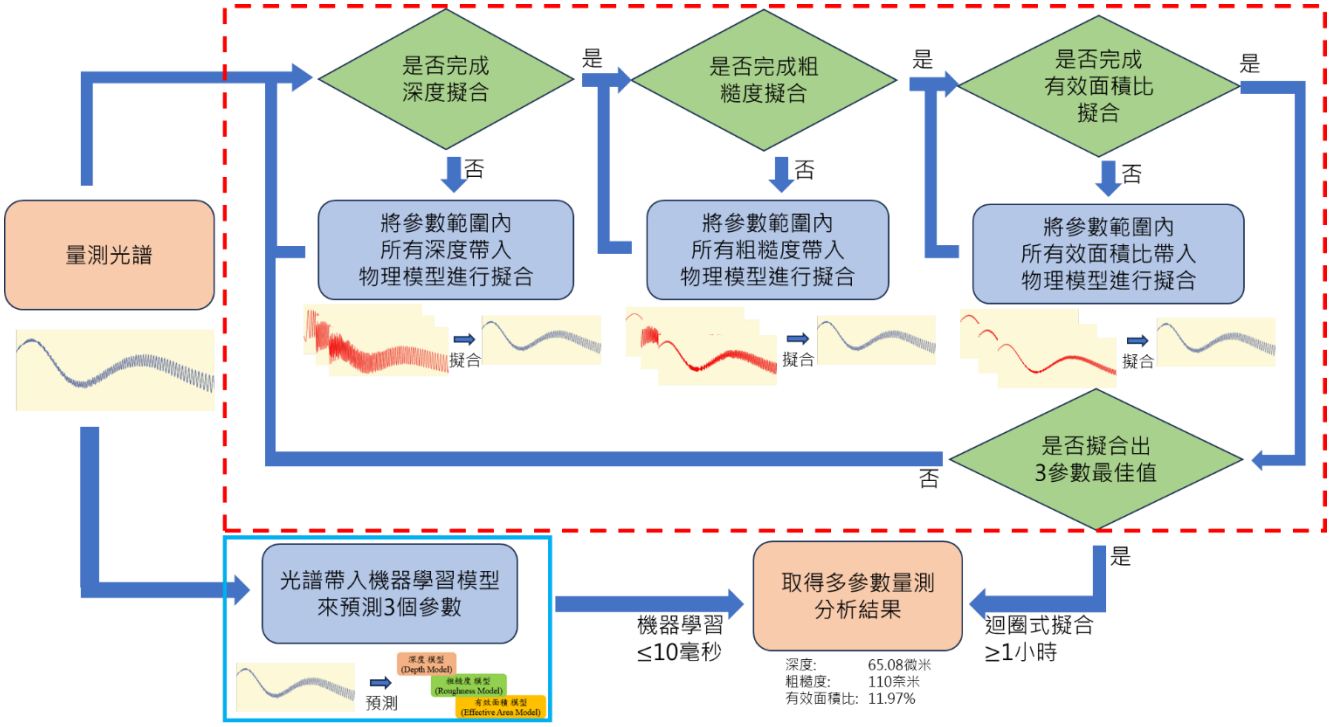
由於電晶體尺寸逐漸接近物理極限，因此近年來發展晶片立體堆疊設計，並採用矽通孔(TSV, through silicon via)核心製程技術成為延續摩爾定律的重要方法。本研究建立了多參數物理模型以進行量測光譜擬合，可用來分析矽通孔三維形貌參數，並已經由掃描式電子顯微鏡(Scanning Electron Microscope, SEM)切片影像驗證為非破壞性、高準確度的量測方法。但由於多參數擬合須多次迴圈逼近相對耗時，因此本研究結合機器學習方法，實驗結果顯示各參數的決定係數(Coefficient of determination ; R^2)高達0.99，且量測分析時間≤ 10毫秒。本研究證實機器學習的導入大幅提升量測速度，未來可應用在先進封裝製程三維結構的多參數量測。

一、 前言

(1) 研究緣起

摩爾定律是由Intel公司的Gordon Moore所提出，其核心是預測每單位面積上的電晶體數量倍增所需的時間。然而近年來由於電晶體尺寸逐漸接近物理極限，縮小電晶體的難度越來越高，因此開發晶片立體堆疊設計，並採用矽通孔核心製程技術成為延續摩爾定律的重要方法。由於矽通孔的高深寬比三維結構參數，包含孔深(depth)、孔底曲率(curvature)及粗糙度(roughness)，是影響後續填銅後形成立體導線的長度及電子訊號傳輸效能的關鍵，過去多採用掃描式電子顯微鏡進行矽通孔剖面檢測，昂貴、耗時且為破壞性。本研究近期建立了多參數物理模型來進行矽通孔量測光譜擬合分析，可用以分析矽通孔三維形貌參數，流程圖如圖一虛線所示。但由於建立多個物理模型較為複雜耗時，且由於各個參數對反射光譜波形特徵的影響具高度相關性，因此演算分析流程須進行多次迴圈式擬合，才能逐漸逼近最佳分析結果，約需耗時1-2小時。本研究期望開發一種快速、高準確度、非接觸性的量

測方法，可實際應用在複雜立體元件的多參數量測分析。



圖一 迴圈式擬合與機器學習流程圖

(2) 研究目的

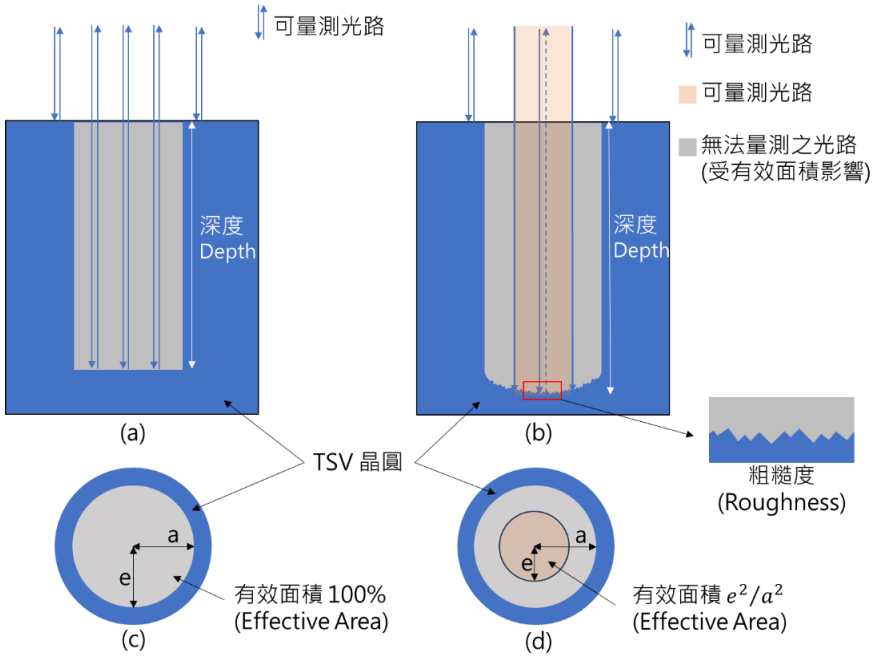
本研究目的是使用非破壞性的光學量測方法，搭配機器學習訓練模型來還原矽通孔三維形貌，並且提出優化模型超參數的方式，從而達到提升量測準確度與縮短模型訓練時間。首先本研究將矽通孔結構的三維參數，包含深度、粗糙度、有效面積比(effective area)，帶入物理模型計算出不同量測波長所對應的干涉光譜強度(intensity)，並且利用光譜強度標準化的作法將其換算成反射率，以取得大量可靠的數據用以發展機器學習模型，最終可應用訓練完成模型，用來即時預測矽通孔量測樣本微結構的深度、粗糙度、有效面積比三個參數，流程圖如圖一實線所示，並接著還原出三維立體結構形貌。本研究結合了原有光學量測方法，而由機器學習模型取代原本較為繁複耗時的擬合演算流程，加速得到多參數量測結果，使其能夠與先進的立體封裝製程整合，具備高準確度，又能即時回饋量測數據並優化製程的目的。

二、 研究方法

(1) 矽通孔物理模型與參數影響

理想的矽通孔量測光路如圖二 (a)，孔底理想的反射有效面積比如圖二 (c)。但實際在量測矽通孔時受到三維多參數結構的影響，如圖二 (b)，由於矽通孔底部曲率的影

響，實際孔底可反射光面積如**圖二** (d)。通常對光譜造成劇烈影響的有三個主要參數，分別是矽通孔的深度、粗糙度、有效面積比。



圖二 (a) 理想矽通孔剖面圖。 (b) 實際矽通孔剖面圖。
(c) 理想矽通孔底部俯視圖。 (d) 實際矽通孔底部俯視圖

$$E_{via} = \left(\frac{e}{a}\right)\sqrt{(1-\alpha)e^{\left[-(4r_q\pi)^2/\lambda^2\right]}}E_0 r_{si}^{+} e^{\left[i2\pi\left(\frac{2d}{\lambda}\right)\right]} \quad (1)$$

孔深會使反射光譜震盪訊號的頻率改變，底部粗糙度會導致光譜訊號衰減，尤其是在短波長衰減更加嚴重，底部有效面積會使光譜訊號在整個量測光譜範圍整體縮減。公式如式(1)所示，其中E_{via}為孔底反射電場，e, a分別為孔底有效半徑與孔洞半徑。α與$1-\alpha$分別為照射光點內晶圓表面以及孔洞所占面積比。r_q為粗糙度，定義為表面高度與平均表面高度的均方根偏差(root-mean-square deviation；RMSD)。λ代表量測波長，E_0為入射光的電場，r_{si}^{+}為當材料為矽晶圓(silicon)時的菲涅耳(Fresnel)反射係數，d為矽通孔的深度[1]。

矽通孔結構量測理論模型是由晶圓表面反射訊號與矽通孔底部有效面積反射訊號所形成的干涉光譜訊號所組成，如**圖二**(b)所

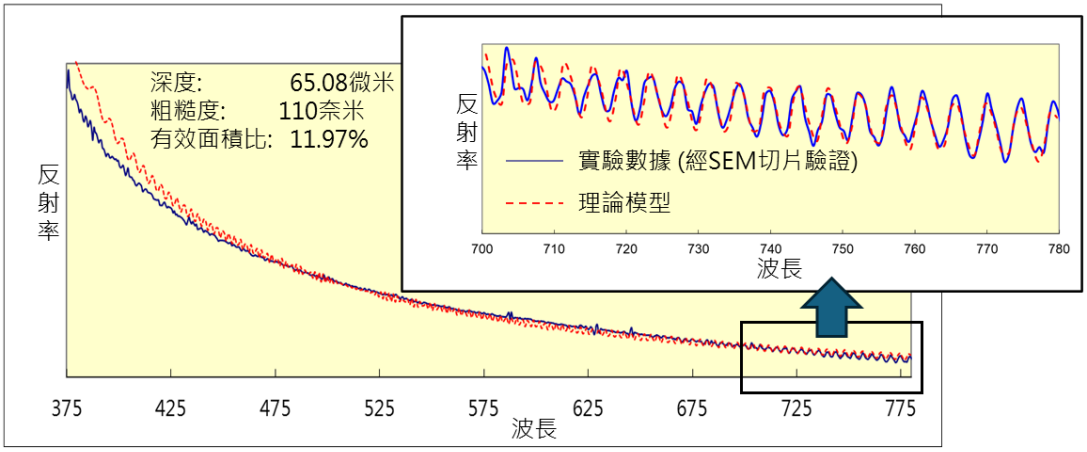
示。干涉光譜強度的模擬模型公式如式(2)所示，本研究帶入不同範圍的深度、粗糙度與有效面積比數值，面積比α，波長λ，以及電場E_0可計算出光譜中對應波長的光強數值，據此將可為機器學習模型建立大量的訓練數據集。

$$I = \alpha(r_{si}^{+}E_0)^2 + (1-\alpha)e^{\left[-(4r_q\pi)^2/\lambda^2\right]}(r_{si}^{+}E_0)^2\left(\frac{e}{a}\right)^2 + 2\sqrt{\alpha(1-\alpha)e^{\left[(4r_q\pi)^2/\lambda^2\right]}}(r_{si}^{+}E_0)^2\left(\frac{e}{a}\right)\cos\left(2\pi(2d/\lambda^2)\right) \quad (2)$$

由於實際量測時每一次入射光所產生的電場數值E_0並不相同，因此本研究將光譜訊號進行正規化(normalization)的前處理[2]，正規化如式(3)所示，其中E_{exp}為量測到的孔底反射與表面反射干涉能量，E_d為暗電流(dark current)能量，E_{ref}為晶圓表面時反射能量。

$$R = \frac{E_{exp} - E_d}{E_{ref} - E_d} \quad (3)$$

本研究的量測光譜與理論模型光譜比較如**圖三**所示，所量測的矽通孔孔徑約為5微米，週距(pitch)約為10微米，面積比α約為81 %，其中實線光譜為實際量測到的反射率光譜，虛線光譜則為理論模型擬合所產生的反射率光譜。經由理論模型擬合可以得到深度約為65.08微米，粗糙度約為110奈米，有效面積為11.97 %。過去採用掃描式電子顯微鏡切片影像驗證，可直接比對深度與粗糙度的結果，底部的有效面積則可透過橢圓模型理論來推算，如式(4)及式(5)推導。其中(e^2/a^2)為有效面積。Δd_{via}為深度解析度，d_{via}為矽通孔深度，$\Delta\lambda$為CCD的像素解析度，$\bar{\lambda}$為平均波長，a為矽通孔底部橢圓形貌的長軸，b為孔底橢圓形貌的短軸[1]。



圖三 理論模型套用實際參數與量測模組量測之波長與反射率比較圖

$$\Delta d_{via} = \frac{d_{via} * \Delta\lambda}{\bar{\lambda}} \tag{4}$$

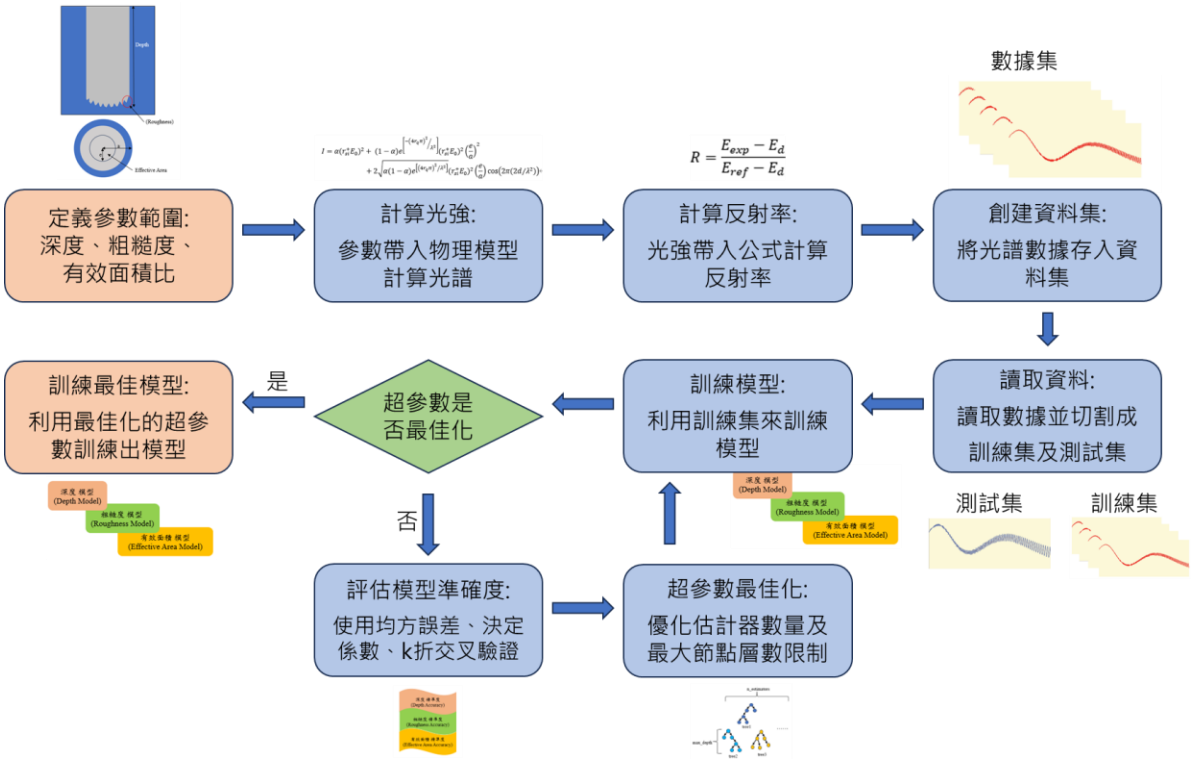
$$\frac{e^2}{a^2} + \frac{(-b + \Delta d_{via})^2}{b^2} = 1 \tag{5}$$

(1) 反射式光譜量測

我們使用商用的薄膜量測儀[3]來量測對矽通孔樣本來進行量測，以實際量測的結果來驗證使用理論模型數據所完成的機器學習模型。矽通孔樣本所取得的干涉光譜通常會呈現在直流訊號上存在著規律性的振幅振盪。

(2) 機器學習架構介紹

本研究建構機器學習模型訓練的整體架構如**圖四**所示，由數據前處理及模型訓練過程兩個部分所構成，使用的程式語言為python 3.10.9，模型為隨機森林回歸(random forest regressor)，此模型是由複數決策樹(decision tree)所組成，當數據輸入模型時每一個決策樹會輸出一個預測的數值，最後由這些數值的平均值來決定模型輸出值。而選擇隨機森林回歸主要有以下優勢：第一，模型較簡易，相較於神經網路而言架構上較簡單。第二，訓練速度快，單一訓練模型時間只要 ≤ 700秒。第三，預測速度快，單一樣品的預測只需用時 ≤ 10毫秒。第四，透過結合許多決策樹的結果，使其具備抗過擬合(anti-overfitting)的能力。



圖四 機器學習模型流程圖

(3.1) 數據前處理

(1) 定義參數範圍

第一步是藉由式(2)來產生特定合理範圍的數據，由於在計算時只要給定深度、粗糙度、有效面積比、波長λ及光斑下孔洞與晶圓表面面積比α，入射表面電場E_0即可算出光強，因此可透過六大參數來產生數據。

(2) 模擬理論光譜

由於本文著重三維結構，因此模型輸出主要為深度、粗糙度、有效面積比三參數，而其餘參數可作為輸入特徵。因此會選定合理範圍的深度、粗糙度、有效面積比，並透過量測模組來確定其餘參數，便可產生機器學習的數據。

(3) 計算反射率

考量到不同量測模組所使用的照射光源、量測干涉光譜的光譜儀、及各項光路元件會有機差，在這一步驟會將模擬光譜從能量利用式(3)轉換成反射率，這樣一來即使使用不同量測模組也不會影響準確度。

(4) 創建資料集

在產生大量的數據後，需要將該數據的波長與反射率之關係圖以特徵的形式以及對應的深度、粗糙度、有效面積比以輸出的形式存儲下來，在實驗中將數據以資料庫的方式將其儲存下來。經過實驗證實，存入約

200,000筆數據時只需花費約16分鐘，讀取數據只花費約半分鐘。

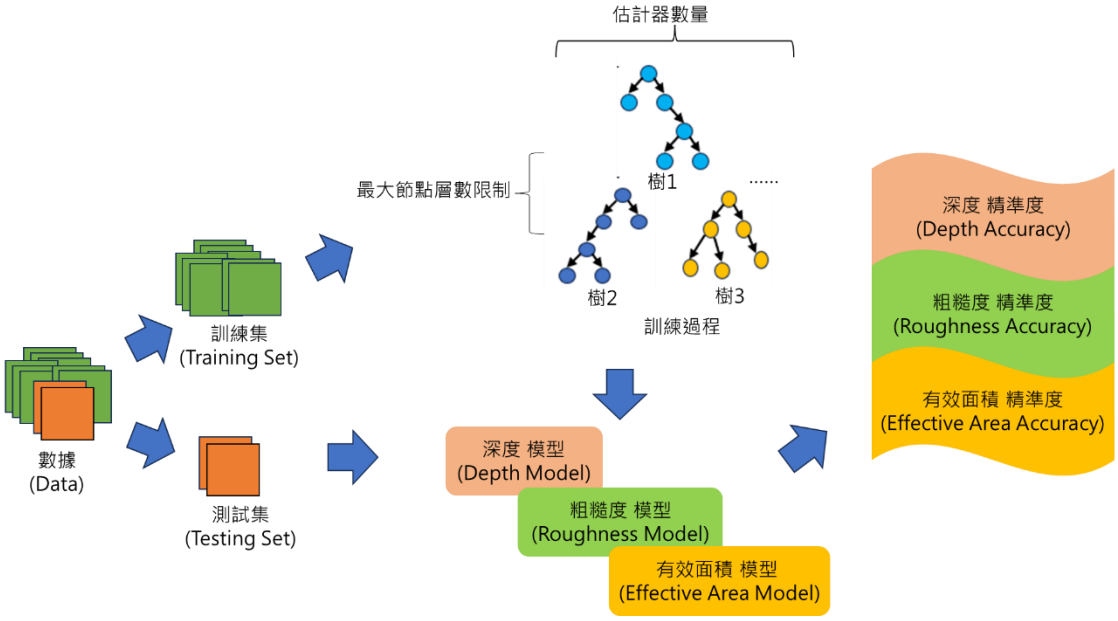
(3.2) 模型訓練過程

(1) 讀取資料

從資料庫中讀取反射率光譜、深度、粗糙度、有效面積比四個參數，並將數據切割成訓練集及測試集，比例為8:2。

(2) 訓練模型

機器學習的系統架構如**圖五**所示。由於輸出為3種不同的參數，因此在訓練時就會為3個參數各訓練一個模型。模型的輸入特徵將會是反射率光譜，輸出參數是深度、粗糙度、有效面積比其中之一。最終會產生3個不同的模型分別預測3個輸出參數，並透過測試集計算3個模型的準確度。



圖五 機器學習的系統架構

(3) 評估模型準確度

為了增加模型的可靠性，採用3種不同的方式來評估模型的準確度。

A. 均方誤差 (mean squared error, MSE)

均方誤差的功能是計算模型的預測輸出與測試集數據的誤差。公式如式(6)所示，其中n為樣本數，y_i為第i個測試集上的數據實際值，$\hat{y}_i$為對應第i個數據的預測值。MSE越小代表輸出越接近測試集數據，即模型準確度越高。

$$MSE = \frac{1}{n}\Sigma(y_i - \hat{y}_i)^2 \quad (6)$$

B. 決定係數

R^2的功能是比較訓練出來的模型與直接使用平均值當作模型的準確度。R^2公式如式(7)所示，其中$\bar{y}_i$為測試集上的數據其輸出平均值。通常R^2會介於0到1之間，$R^2 = 1$代表訓練出來的模型的輸出誤差為0，因此目的是使模型R^2的數值盡量接近1。

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \tag{7}$$

C. K 折交叉驗證(k-fold cross-validation)

本方法會將整個數據集切割成k個相同大小的子數據集，接下來會訓練出k個模型，每次訓練會讓其中一個子數據集作為測試集，剩下的則會作為訓練集，直到所有的子數據集都輪流作為測試集後，會將所有模型的MSE與R^2記錄下來，最後計算出其平均值。使用k折交叉驗證後能更全面的評估模型的準確度，減少因隨機分割數據造成的偏差。

(4) 超參數最佳化

隨機森林模型中有兩個超參數的數值是影響模型準確度最大的因素，其中一個超參數是估計器數量(n_estimators)，代表的是訓練模型的過程中會產生多少決策樹來計算結果。另一個是超參數是最大節點層數限制(max_depth)，該參數會決定決策樹最多具有多少層節點。由於超參數是在訓練模型完成前給定的數值，而每一種模型其最優數值都不相同，因此必須針對模型優化超參數的設定。其中可靠性最高的做法是使用窮舉法，但時間及計算成本過於龐大，因此在實驗中使用了貪婪演算法(greedy algorithm)來優化這兩個超參數。

本研究實驗中，首先將估計器數量設定為 10，最大節點層數限制設定範圍10~200，再逐一訓練模組找出可以訓練出R^2最高的10個最大節點層數限制。接著將最大節點層數限制設定為10，以同樣方式找出R^2最高的10個估計器數量。最後將10個結果最好的估　計器數量與最大節點層數限制共100個組合，以5折交叉驗證的方式來訓練出最好的模型。經過實驗，使用貪婪法花費的時間相較窮舉法快了41倍以上，節省了大量的時間成本。

表一為實驗最佳化後的模型參數，在物理模型參數上，雖然量測模組擷取的光譜範圍為 375 奈米到 780 奈米，但訓練模型時僅納入光譜範圍在 450 奈米到 750 奈米之間，這是由於該波長段的反射率較大的緣故。在

決定深度的範圍時，因為量測的樣本其深寬比約為 9~12，且實驗中量測的矽通孔主要為孔徑 5 微米，所以將模型的深度設定在 45 微米到 70 微米之間。而粗糙度與有效面積在設定時參照過去掃描式電子顯微鏡切片分析的數值約為 110 奈米與 11.97 %，因此將粗糙度範圍設定在 110 ± 40 奈米[4]，且有效面積範圍設定在 $12 \pm 7\%$。由此決定了物理模型的輸入光譜範圍與輸出參數範圍。而在優化隨機森林的超參數時，透過使用貪婪法與 5 折交叉驗證，將深度、粗糙度、有效面積比模型的估計器數量分別設定在 199、161 與 184，且將最大節點層數限制分別設定在 34、80 與 144 時的模型的精確度最高，由此決定了最佳化模型的超參數。

表一 最佳化參數設定

物理模型參數				
	波長 (奈米)	深度 (微米)	粗糙度 (奈米)	有效面積比 (%)
實際值	450	45	70	5
預測值	750	70	150	19
間距值	1	0.1	1	1
數據量	$301 * 406620$			
隨機森林超參數				
超參數	深度		粗糙度	有效面積比
估計器數量	199		161	184
最大節點層數限制	34		80	144

三、 結果與討論

(1) 模型參數設定與對理論數據的實驗結果

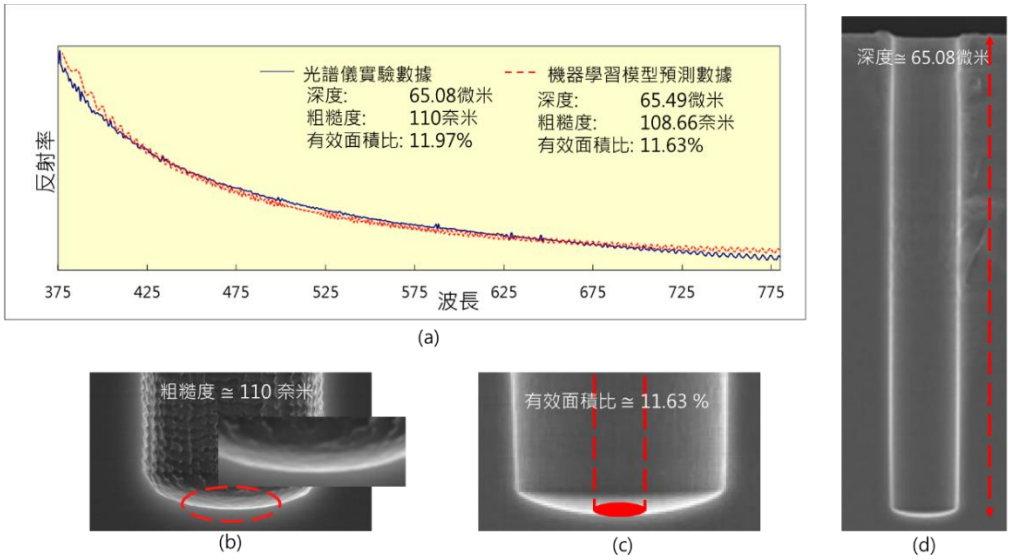
表二為實驗中測試集套用最佳化模型的結果，在實驗中訓練集為325,296筆數據，而測試集為81,324筆數據。其中深度測試集MSE為5.22，R^2為0.91，粗糙度測試集的MSE為56.05，R^2為0.9，而有效面積比測試集的MSE約為0.06，R^2為0.99。

表二 機器學習對測試集模型參數與結果

	深度	粗糙度	有效面積比
MSE	5.22	56.05	0.06
R^2	0.91	0.9	0.99

(2) 量測模組套用機器學習模型之預測結果

將實際量測矽通孔時得到的光譜套用至機器學習訓練出來的模型並分別進行3個參數的驗證，**表三**為實驗中的一組代表性量測結果。圖六(a)為實際光譜與模型預測結果之光譜比較圖。量測孔徑約為5微米，pitch約為10微米，面積比α約為0.31。經掃描式電子顯微鏡切片驗證的深度實際值為65.08微米，如圖六 (d)所示，模型預測的數值為65.49微米，誤差值為 -0.41微米。粗糙度經掃描式電子顯微鏡切片驗證的實際數值為110奈米，如圖六 (b)所示，模型預測的數值為108.66奈米，誤差值為1.34 nm。有效面積實際數值經掃描式電子顯微鏡切片為11.97 %，如圖六 (c)所示，模型預測的數值為11.63 %，誤差值為0.34 %。



圖六 (a) 矽通孔實際量測結果與模型預測結果之光譜比較圖。

(b) (c) (d) 矽通孔經掃描式電子顯微鏡切片後粗糙度、 有效面積比 、深度量測結果。

表三 量測光譜套用機器學習模型之預測結果

	深度 (微米)	粗糙度 (奈米)	有效面積比 (%)
實際值	65.08	110	11.97
預測值	65.49	108.66	11.63
誤差值	-0.41	1.34	0.34

結論

本研究證實機器學習模型導入反射式光譜干涉量測應用，除了在矽通孔三維結構多參數量測時有非常高的準確度，同時取代了傳統理論模型迴圈式擬合演算法，更排除了多參數之間交互影響的痛點，大幅提升量測速度。由於現在製程複雜度越來越高，建構物理模型的成本技術與越來越高，而審慎訓練完成所建立的機器學習模型，就能在不需複雜演算流程的情況下，藉由機器學習一步到位快速獲得量測結果，大幅降低了開發及維護成本。機器學習除了矽通孔三維形貌量測的成功案例之外，亦可應用於先進製程複雜多參數的各種元件量測。未來會在訓練模型發展的同時，加入一定程度的雜訊及環境變因，使得模型更可以適用於多樣大量的實驗條件。

參考文獻

1. Ku, Y.-S. (2014). Spectral reflectometry for metrology of three-dimensional through-silicon vias. Journal of Micro/Nanolithography, MEMS, and MOEMS, 13(1), 011209. https://doi.org/10.1117/1.jmm.13.1.011209
2. Cheng, X., Tang, Y., Yang, K., & Han, C. (2022). Deep learning for thin film thickness measurement in spectroscopic reflectometry. IEEE Photonics Technology Letters, 34(18), 969–972. https://doi.org/10.1109/lpt.2022.3195997
3. Ku, Y.-S., & Yang, F. S. (2010). Reflectometer-based metrology for high-aspect ratio via measurement. Optics Express, 18(7), 7269. https://doi.org/10.1364/oe.18.007269

4. Ku, Y.-S., Huang, K. C., & Hsu, W. (2011). Characterization of high density through silicon vias with spectral reflectometry. Optics Express, 19(7), 5993. https://doi.org/10.1364/oe.19.005993

作者簡介

周柏文 / 量測中心

顧逸霞 / 量測中心 / 資深正研究員