

基於機器學習之 kMLLS 聚類演算法及其在薄膜電子損失譜之應用

The Introduction to kMLLS Clustering and It's Applications to EELS Analysis of Thin Films

蔡任豐、張睦東、羅聖全、陳健群

Ren-Fong Cai, Mu-Tung Chang, Shen-Chuan Lo, Chien-Chun Chen

近年來，球面像差校正穿透式電子顯微鏡影像搭配能譜分析技術，已逐漸成為材料分析中不可或缺的工具。從能譜影像中，藉由適當的數據處理，我們除了原子特徵之外，更可以得到化學元素及電子結構等材料訊息。在本文中，我們介紹一種新穎的演算法—kMLLS 聚類，其結合了 k 平均聚類與複線性最小平方擬合的優點，使我們可以正確地萃取出終端材料能譜，並得到其分佈情形。利用 kMLLS 聚類，未來可應用在線上檢測，使研究人員即時得到更深入的材料訊息。

In recent years, combining a Cs-corrected scanning transmission electron microscope with an EDS and/or EELS detector has become an indispensable tool for material characterization. With a proper data processing, atomic structures, chemical compositions, and electronic configurations of materials can be resolved in a spectrum image. In this article, we introduce a novel algorithm - kMLLS clustering, which combines the advantages of k-means clustering and multiple linear least squares fitting, to accurately extract the spectra of the endmembers and the corresponding distribution from a spectrum image. kMLLS clustering has the great potential to the in-line application and provides significant insights into materials.

一、前言

球面像差校正穿透式電子顯微鏡 (cs-corrected scanning transmission electron microscopy) 搭配能量發散光譜儀 (energy dispersive spectroscopy, EDS) 以及能量損失光譜儀 (electron energy loss spectroscopy, EELS)，不僅可以將原子看得仔細，更可以得知其化學元素及電子組態等材料訊息^(1, 2)，已成為近年來不可或缺的分析工具之一。能譜影像 (spectrum image) 為

結合顯微影像與能譜分析之高維度數據，藉由紀錄 X、Y 座標及其能譜資訊，瞭解各化學成分或是鍵結在分析區域中的分佈情形。因此，一張能譜影像如同大數據一般，包含了許多的材料資訊。要如何從中提取出重要的訊息已成為顯微鏡學家所關注的熱門項目之一⁽³⁻⁵⁾。

機器學習 (machine learning) 近年來已成為了家喻戶曉的名詞，各行各業都想要利用它將以往所累積的海量數據中挖掘出金礦。圖 1 為傳統程式設計與機器學習之比較，傳統程式設計師設計一些規則 (演算法)，讓電腦程式可以依照其步驟進行數據處理。演算法的設計一定會存在著程式設計師的偏見，有些珍貴的訊息有可能會被忽略。在機器學習中，我們提供數據及答案讓電腦學習其中的規律，或是僅提供數據直接讓電腦幫我們整理出其數據的規則，前者稱為監督式學習 (supervised learning)，後者則為非監督式學習 (unsupervised learning)。在能譜分析中，我們通常無法提供數據的真實答案，因此僅能使用非監督式學習的方式分析能譜影像，使我們可以從中了解到數據中的寶貴資訊並降低因人為偏見所造成的訊息流失^(6,7)。

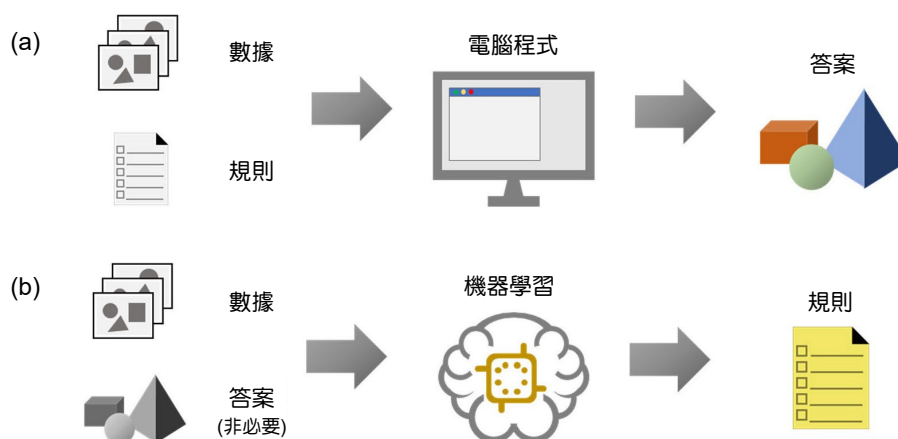


圖1. (a) 傳統電腦程式與 (b) 機器學習之比較。

在能譜影像分析中，我們關切的重點是了解各純物質或稱為終端材料 (endmember) 的分佈情形。然而在實驗中，我們所得到的數據多半為混合的能譜，該如何從中得到終端材料之能譜則是很重要的任務，混合能譜分離 (spectral unmixing) 技術就是再幫助我們解決此問題⁽⁸⁻¹¹⁾。已有許多文獻針對能譜分離進行研究，然而其仍有改善的空間，我們會在後文中一一探討。在本論文中，我們介紹一種基於機器學習之混合能譜分離演算法—kMLLS 聚類⁽¹²⁾，藉由此方法，我們將可以快速的將終端材料能譜萃取出來，並了解各終端材料真實的分佈情形。

二、混合能譜分解相關技術

1. 主成份分析

主成份分析 (principle component analysis, PCA) 是一種常見的統計學多變量分析方法，也是機器學習中常用的非監督式學習法，其主要應用在數據降維、資料壓縮以及特徵萃取等方面⁽¹³⁻¹⁵⁾。因此也有顯微鏡學家應用在從能譜影像中萃取出終端材料能譜，了解各材料的分

佈情形。藉由主成份分析，我們可以將一組能譜影像 $D((x, y), E)$ 分解為：

$$D((x, y), E) = T((x, y), n) \times P_{(E, n)}^T \quad (1)$$

其中 $P_{(E, n)}$ 稱作負荷矩陣 (loading matrices)，為數據共變異矩陣 (covariance matrix，即 $D^T D$) 之特徵矩陣 (eigen vector)，代表主成份的能譜訊息； $T((x, y), n)$ 為分數矩陣 (score matrices)，則記載著各成份的貢獻度； n 則為主成份的個數。一般來說， $P_{(E, n)}$ 代表終端材料能譜，但其能譜失去了元素特徵，因此不具有物理意義。幸運的是，我們可以選擇保留少數幾個主成份進行數據重組，降低能譜影像中的雜訊，這也是主成份分析在電子顯微鏡數據處理中最常見的應用。

2. 獨立成份分析與非負矩陣分解法

想像我們正在參加一場晚宴，宴會中人聲吵雜。此時有人正在與你交談，我們該如何把其他人的聲音排除，讓你能分辨清楚與你交談人的話語，這項工作就是獨立成分分析 (independent component analysis, ICA) 的開發構想^(11, 16)。在 EELS 能譜分析中，我們也可以藉由獨立成分分析將終端材料能譜訊息從數據中萃取出來。一般來說，能譜影像中的各能譜 S_n 可表示一系列的終端材料能譜 $R_i(E_{cm})$ 之線性組合，以數學可表示為：

$$S_n(E_{cm}) = \sum_{i=1}^p \beta_i R_i(E_{cm}) \quad (2)$$

獨立成分分析的主要任務就是找出一組能譜 $R_i(E_{cm})$ ，且符合各能譜之間皆線性獨立，這組能譜即為我們所想要的終端材料能譜。

非負矩陣分解法 (non-negative matrix factorization, NMF)，目標與獨立成份分析相似，著重於找到一組 $R_i(E_{cm})$ ，但會限制所有的 β_i 皆為正^(10, 17-19)。上述兩種方法皆可成功的在 EELS 能譜影像中萃取出終端材料能譜。然而，數據在分析前常搭配主成份分析進行能譜降噪 (denoising)，其過程也需要大量的計算資源。

3. 複線性最小平方擬合

複線性最小平方擬合 (multiple linear least squares fitting, MLLS) 在 EELS 能譜分析中，常用來分開兩個以上互相重疊能譜的技巧，其本質為最基本的機器學習基礎：線性迴歸 (linear regression)⁽²⁰⁻²²⁾。其表示式與公式 (2) 相同；但與獨立成分分析不同的是，在 MLLS 中，我們需要人工指定 $R_i(E_{cm})$ ，我們將其稱為參考能譜，亦即能譜影像中以人工方式辨別之從終端材料區域中擷取出的能譜。為了知道單一能譜 S_n 中各終端材料比例 β_i ，我們可將公式 (2) 寫成矩陣的格式：

$$S_n(E_{c1}, E_{c2}, \dots, E_{cm}) = [\beta_1, \dots, \beta_p] \begin{bmatrix} R_1(E_{c1}, E_{c2}, \dots, E_{cm}) \\ \vdots \\ R_p(E_{c1}, E_{c2}, \dots, E_{cm}) \end{bmatrix} \quad (3)$$

$R_p(E_{c1}, E_{c2}, \dots, E_{cm})$ 代表第 p 組參考能譜，能譜由 m 個頻道所紀錄。 β 的解可用矩陣運算求得：

$$\beta = S_n \times R^T \times (R \times R^T)^{-1} \quad (4)$$

我們把能譜影像中每一個能譜 S_n 皆進行以上的運算，即可得知各終端材料在每個分析位置的比例為何。

4. 群集分析

群集分析 (clustering analysis) 也是機器學習中的一種非監督式學習的技巧，目的在將相似之資料點進行分類，因此也有研究利用群集分析萃取終端材料能譜。分群的方式可依照數據點的質心、密度、分佈方式或是特性進行分類^(23, 24)。 k 平均聚類 (k -means clustering) 是一種以質心分群的手法，由於其方法簡單，因此有廣泛的應用。群集分析中，各數據點會以其群組中心表示。在 k 平均聚類中，群集的質心是由該群組中所有數據點的平均，因此應用在能譜分析時，雜訊會因為訊號平均的關係，使得能譜訊雜比 (signal-to-noise ratio) 得到改善^(23, 25, 26)。

k 平均聚類的操作步驟為：(1) 決定數據可以分為多少群，即 k 的數值。(2) 隨機指定 k 個數據點當作初始質心。(3) 各數據點依照其與各質心之相似度進行分群。(4) 計算各群的平均能譜作為新的質心。(5) 重複步驟 3 與 4，直到質心不再變化為止。步驟 3 中的相似度一般來說是計算數據點與所有質心的歐幾里德距離 (euclidean distance) 來定義的⁷，其數學表示如下：

$$similarity = \sqrt{\sum_m [S_m(E_{cm}) - R_p(E_{cm})]^2} \quad (5)$$

k 平均聚類已經被廣泛地應用在顯微影像的數據處理中，然而很少有文獻利用其技術進行能譜分析⁽²⁷⁻²⁹⁾。此外，在 k 平均聚類中，所有的數據點皆會依照最接近的質心進行分組，因此無法排除非純相的「雜訊」，因此終端材料的能譜還是很難被萃取出來。

三、kMLLS 聚類演算法開發

1. kMLLS 聚類演算法

經由前文的介紹，我們得知 k 平均聚類可以自動的從能譜影像中萃取出主要的材料成份，但其大多為混合材料之能譜。複線性最小平方擬合則可以有效的將混合能譜分解，但其關鍵在於需要人工的指定參考能譜才能進行運算。在本論文中，我們將介紹結合 k 平均聚類 (k -means clustering) 以及複線性最小平方擬合 (MLLS) 優點之 kMLLS 聚類 (即 k -means 聚類與 MLLS 之合成)，幫助我們自動地將能譜影像中的終端材料能譜萃取出來⁽¹²⁾。kMLLS 聚類演算法流程列於圖 2。kMLL 聚類的第一步為以 Elbow 法幫助我們決定在數據中有多少終端材料，即 k 的數量。第二個步驟為進行 k 平均聚類，各群組的質心能譜可當作初始的終端材料能譜。由於 Elbow 法僅能粗略的決定系統終端材料的個數，因此在第三步我們以暴力法排除非終端材料之能譜。接著將上述得到的初始終端材料能譜作為參考能譜，針對每個數據點進行複線性最小平方擬合。最後迭代進行檢查各係數是否有大於 100% 的資料點，若符合的話則將其平均，成為新的參考能譜。重複進行每個數據點的複線性最小平方擬合運算直到所有的係數都小於或等於 100%，則完成迭代步驟，最後的留下的參考能譜即為系統中的終端材料能譜。此外，分析之能譜多伴隨著雜訊存在，因此進行迭代步驟時會加入一容忍值，即各係數皆小於或等於 (100% + 容忍值) 才會完成迭代計算，以降低雜訊造成計算上的誤差。

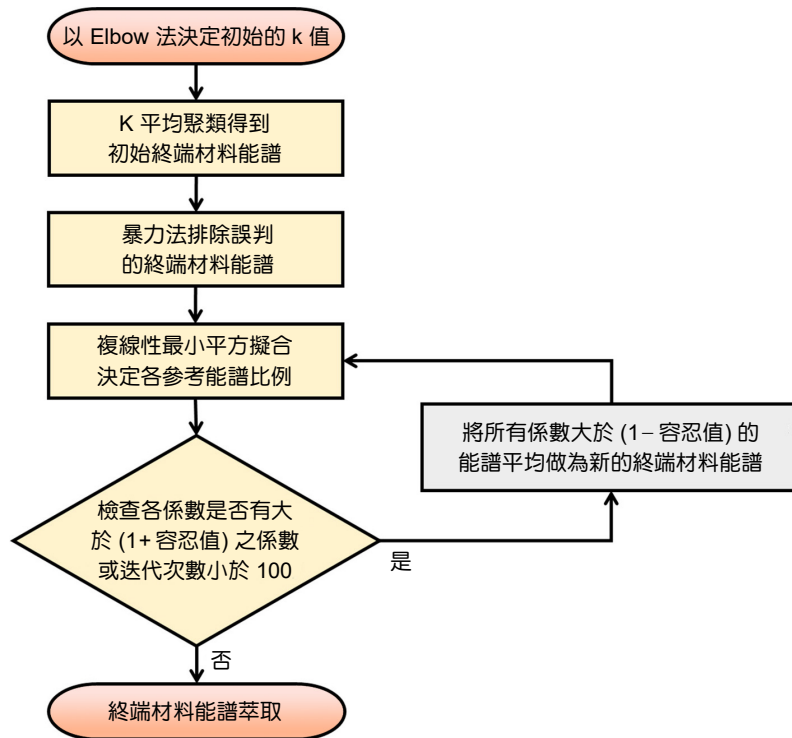


圖 2. kMLLS 聚類計算流程。

這邊再詳細解釋 Elbow 法以及暴力法的過程。在 k 平均聚類中，如何選出一個適當的群數 (即 k 的數量) 一直是研究人員所關注的問題。其中 Elbow 法為一種構想簡單且快速的方法⁽⁷⁾。Elbow 法的過程為分別將數據進行 $k = 1$ 至 $k = 10$ 的 k 平均聚類 (若是複雜系統，可以考慮計算更多的 k)，再分別計算不同 k 值中各數據點與所在群聚質心的平方距離總和，或稱平方誤差總和 (sum of squared error, SSE)。繪製出 k -SSE 曲線，曲線中的肘點處 (即 elbow 處，因此稱為 Elbow 法) 可當作初始的分群數。但在 Elbow 法中決定的群數，並非全是真正的終端材料。若是分析區域中混合區域比重過大時，仍會被視為一群，此時我們就要借助暴力法 (brute force algorithm)⁽³⁰⁾ 來將混合相的群組質心排除。

暴力法是一種很直觀，但是缺乏計算效能的技巧。其列舉出所有能譜組合的可能性，並逐一檢驗其是否為兩者或兩者群組質心所組成的，若是混合的能譜，則將從參考能譜中去除。暴力法是一種計算過程繁複的演算法，所幸在我們的演算法中，我們僅考慮各能譜是否為兩種或兩種以上所組成，整體的計算量並不會太多。舉例來說，若我們要檢查五個參考能譜中是否有混合能譜的計算量為： $C_3^5 \times 3 + C_4^5 \times 4 + C_5^5 \times 5 = 55$ 次計算，在現代一般使用電腦中計算時間僅為一眨眼功夫即完成檢查工作。

2. 演算法開發與實驗工具

為了驗證 kMLLS 聚類演算法之分群效果，我們以兩組模擬的能譜影像 (spectrum image) 以及一組實驗數據作為驗證。本論文中數據處理以及演算法開發的部份主要是以 Python 程式語言所撰寫，並利用 Python 中的 HyperSpy、Scikit-learn、Numpy 以及 Matplotlib 等函式庫進行訊號以及數據處理^(31, 32)。HyperSpy 函式庫提供了許多檔案讀寫的程式，可以

直接匯入一般電子顯微鏡數據常見的格式檔，如：dm3、dm4 以及 msa 等檔案格式。此外，HyperSpy 函式庫也提供了許多顯微鏡影像及能譜 (包含 EDS、EELS 等) 數據處理之函式，可供使用者直接進行數據操作處理。Scikit-learn 則是 python 語言中常用之機器學習函式庫，在本論文中，我們將使用裡面的 KMeans 函數進行 k 平均聚類之處理。文中提到的 PCA、ICA 以及 NMF 等演算法皆可在此函式庫中找到對應的函式進行數據處理分析。numpy 提供許多的數學函數，本文中將利用其產生亂數之函數，添加模擬數據的雜訊。最後，文章的圖片是以 matplotlib 函式庫所製作。

能譜模擬的部份則是使用 GATAN[®] DigitalMicrograph 軟體中的 EELS Advisor 套件進行模擬⁽³³⁾，其可以依照顯微鏡的加速電壓、光圈大小等物理參數快速的計算能譜強度分佈。然而，EELS Advisor 僅針對能譜強度進行模擬，若要研究能邊的精細結構則使用以密度泛函理論 (Density functional theory, DFT) 進行計算如：WIEN2k 等軟體進行模擬計算⁽³⁴⁾。

實驗數據為工研院材化所建置之冷場發射電子源 JEOL JEM-ARM200F 球面像差校正掃描穿透式電子顯微鏡，搭配 GATAN Quantum 965 EELS 偵測器。前者提供了絕佳的空間分辨力與較窄的能量發散電子光源，為 EELS 分析帶來了能量解析度的提昇，因此更可以從其電子結構了解材料的化學組態。後者則可以同時收集低能損失能譜與高能損失能譜，大大的提昇分析的方便度與可靠性。

四、kMLLS 聚類應用案例

1. 二元系統

首先我們先以一單一元素的鈹 (Be) 及化合物碳化硼 (B_4C) 之二元系統的線掃描能譜影像進行 kMLLS 聚類演算法測試，藉此了解其操作步驟及原理。圖 3(a) 顯示該系統，系統之左右兩端分別為純的鈹以及碳化硼，中間混合區域之能譜分別依照其比例線性疊加而成。圖 3(b) 顯示純相之 Be 與 B_4C 電子損失譜，其中 Be_K、B_K 以及 C_K 能量損失能邊 (energy loss edge) 分別位於 111、188 以及 284 電子伏特。為了可以更貼近實驗結果，因此在訊號中分別加入 10% 的卜瓦松雜訊 (Poisson noise)。雜訊之振幅大小是依照卜瓦松分佈函數 (Poisson distribution) 所計算⁽³⁵⁾，其公式入下：

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (6)$$

其中 λ 表示單位時間內會發生的機率。10% 之雜訊強度則是依照公式 (7) 所計算：

$$\frac{\sum |G(E) - F(E)|}{\sum F(E)} \times 100\% = 10\% \quad (7)$$

$G(E)$ 表示加入雜訊之訊號， $F(E)$ 則為原始之訊號。

分析的第一步會先進行系統之初始 k 值評估，圖 4(a) 為鈹和碳化硼之二元系統的線掃描能譜影像進行 Elbow 法評估結果，其結果顯示該系統之 $k = 3$ ，代表可將此能譜影像以 k 平均聚類區分為三群。其各群質心分別於圖 4(b) 至 (d) 所示。從結果可以很清楚的發現，質心 3 為質心 1 與質心 2 所混合之能譜，因此將會從參考能譜中排除，僅留下主要由碳化

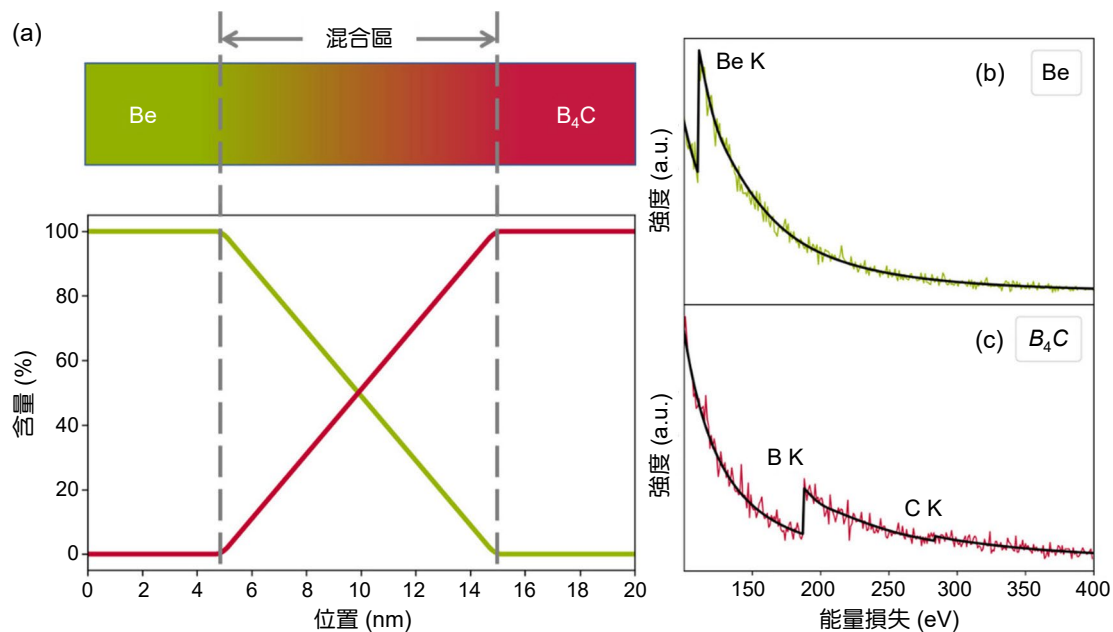


圖 3. 模擬二元線掃描能譜影像系統。(a) 系統示意圖及成份分佈。(b) 與 (c) 分別為模擬之鈹與碳化硼以及其雜訊混合能譜。

硼以及鈹所構成之質心 1 以及質心 2。將此兩質心當作參考能譜，進行複線性最小平方擬合可以得到各質心的含量分佈，其結果顯示於圖 4(e)。在純碳化硼與鈹之區域所得之平均含量分別為 $104.6 \pm 3.2\%$ 以及 $105.2 \pm 2.7\%$ ，顯示質心 2 與質心 1 仍皆為混合相之能譜，因此仍須進一步處理，將真正的終端能譜萃取出來。圖 4(f) 為以 kMLLS 聚類所得之質心 1 (圖 4(g)) 與質心 2 (圖 4(h)) 的含量分佈，其純相區之平均含量為 $101.5 \pm 3.0\%$ 以及 $101.6 \pm 2.5\%$ ，分別代表碳化硼與鈹。此結果顯示，利用 kMLLS 聚類所得到之質心較接近純相之終端能譜，但由於能譜訊號中雜訊的影響，因此仍有所誤差。

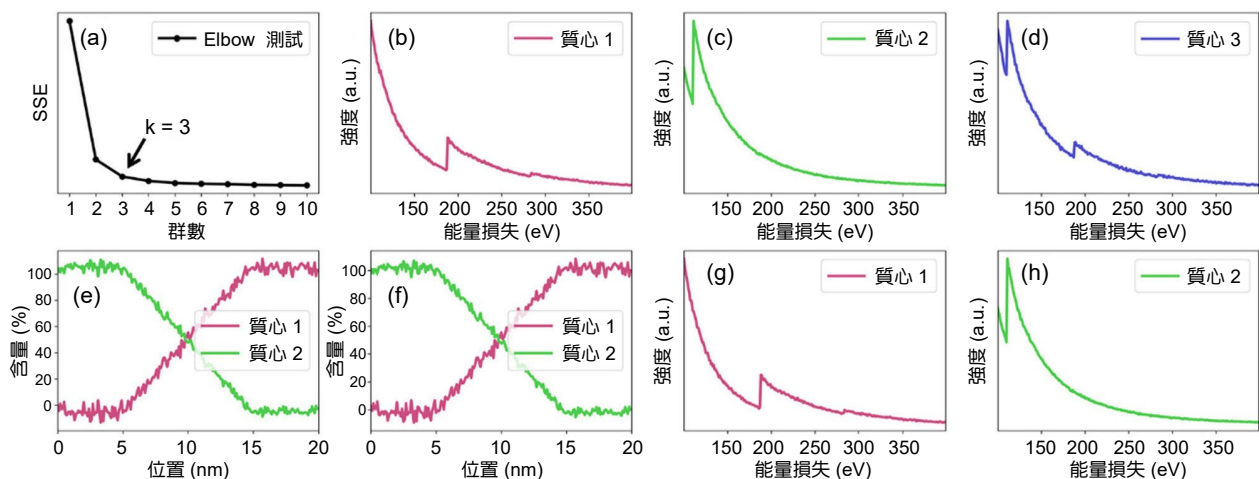


圖 4. (a) 以 Elbow 法求得系統之初始 k 值，結果顯示 $k=3$ 。(b) 至 (d) 分別為以 k 平均聚類之三個群組質心。(e) 及 (f) 分別為以 k 平均聚類以及 kMLLS 聚類所得之各質心含量分佈情形。(g) 與 (h) 分別為以 kMLLS 聚類所得到之終端材料能譜，質心 1 接近純碳化硼，質心 2 則接近純鈹。

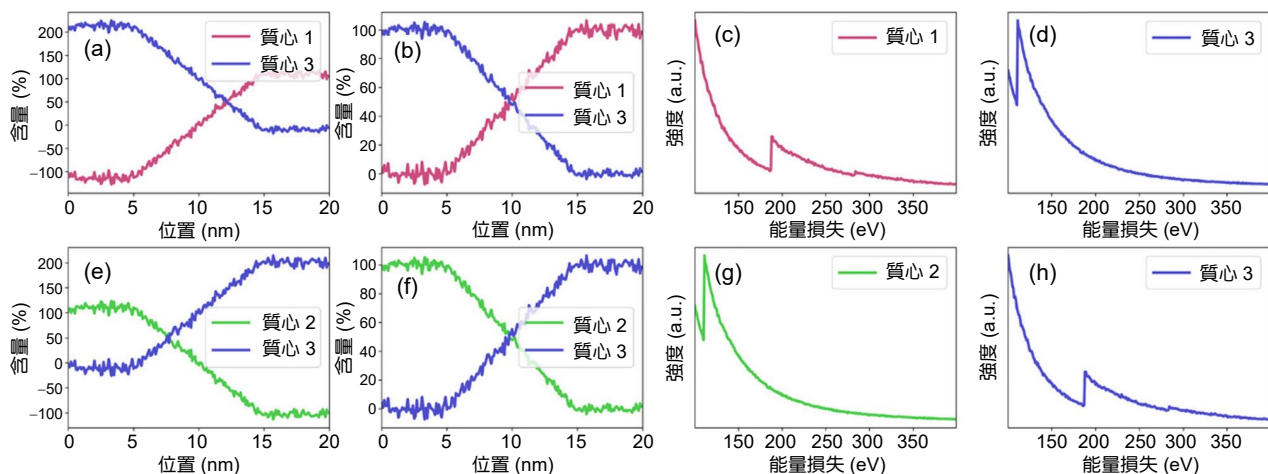


圖 5. (a) 與 (b) 分別以 k 平均聚類與 kMLLS 聚類所計算之質心 1 與質心 3 含量分佈。(c) 及 (d) 分別為以 kMLLS 聚類所萃取之終端材料能譜。(e) 與 (f) 分別以 k 平均聚類與 kMLLS 聚類所計算之質心 2 與質心 3 含量分佈。(g) 及 (h) 分別為以 kMLLS 聚類所萃取之終端材料能譜。

在上面的過程中，我們藉由暴力法將混合相的能譜去除，留下兩個較接近純相的能譜作為參考能譜進行終端材料能譜的萃取。為了測試 kMLLS 聚類演算法的可靠性，在這邊，我們故意將上述排除的混合相能譜 (即質心 3) 作為參考能譜之一，分別與質心 1 與質心 2 進行終端能譜萃取，以測試 kMLLS 聚類之終端能譜萃取能力。圖 5(a) 至 (d) 分別為 k 平均聚類與 kMLLS 聚類得到的質心 1 和 3 之含量分佈圖與以 kMLLS 聚類萃取出的終端材料能譜。由於質心 3 為混合相之能譜，因此造成 k 平均聚類各質心含量比例變化劇烈。利用 kMLLS 聚類萃取出終端材料能譜後，其含量分佈趨勢則與圖 4(f) 相同。而質心 3 也從混合相能譜轉變為接近純相之鉍能譜。若以質心 2 與質心 3 當作參考能譜進行兩聚類比較，kMLLS 聚類可以得到相似的結果，混合相的質心 3 則轉變為接近純相的碳化硼能譜，其結果如圖 5(e) 至 (h) 所示。由此測試可知，kMLLS 聚類可以有效的從混合能譜將真實的終端材料萃取出來。此外，在本案例中，我們故意使用由兩元素所組成之化合物進行測試，不管是 k 平均聚類或是 kMLLS 聚類所萃取出來的結果皆顯示為化合物而非個別元素，顯示其演算法確實能萃取出分析數據中真正的終端材料。

2. 三元系統

在矽基半導體材料 EELS 分析中，常用之閘極金屬鎢與高介電材料鉭之能邊常與矽嚴重重疊，因此我們設計一三元結構：矽化鉭－鉭－鎢，來測試 kMLLS 聚類之在訊號重疊的萃取能力。各材料之分佈與各純相能譜顯示於圖 6，Ta_M、Si_K 以及 W_M 之能量損失能邊分別位於 1735、1839 以及 1809 電子伏特，三種元素之能邊皆為延伸之吸收特徵，因此大大的提昇 EELS 分析難度。

首先我們先以 Elbow 法決定初始的 k 值，結果顯示其肘點落在 3 的位置，此結果與我們設計之系統相符。接著進行 k 平均聚類將此能譜影像之所有數據點分為三群，各質心以及分佈如圖 7 所示，在 0 至 5 奈米處，質心 2 約略為 110%，質心 3 則約略為 -10%，明顯與常理不符 (因在實際材料中不會有負含量)。結果顯示儘管 Elbow 法結果與系統終端材料數

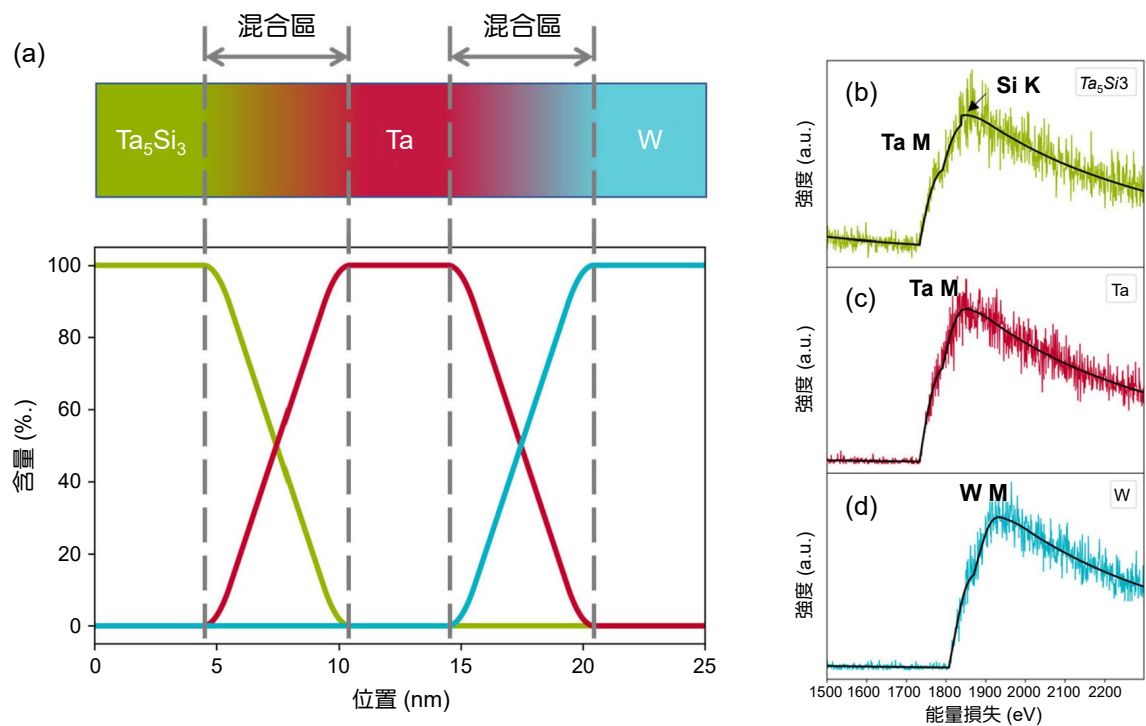


圖 6. 模擬三元線掃描能譜影像系統。(a) 系統示意圖及成份分佈。(b) 至 (d) 分別為模擬之純相矽化鉭、純鉭與純鎢以及其雜訊混合能譜。

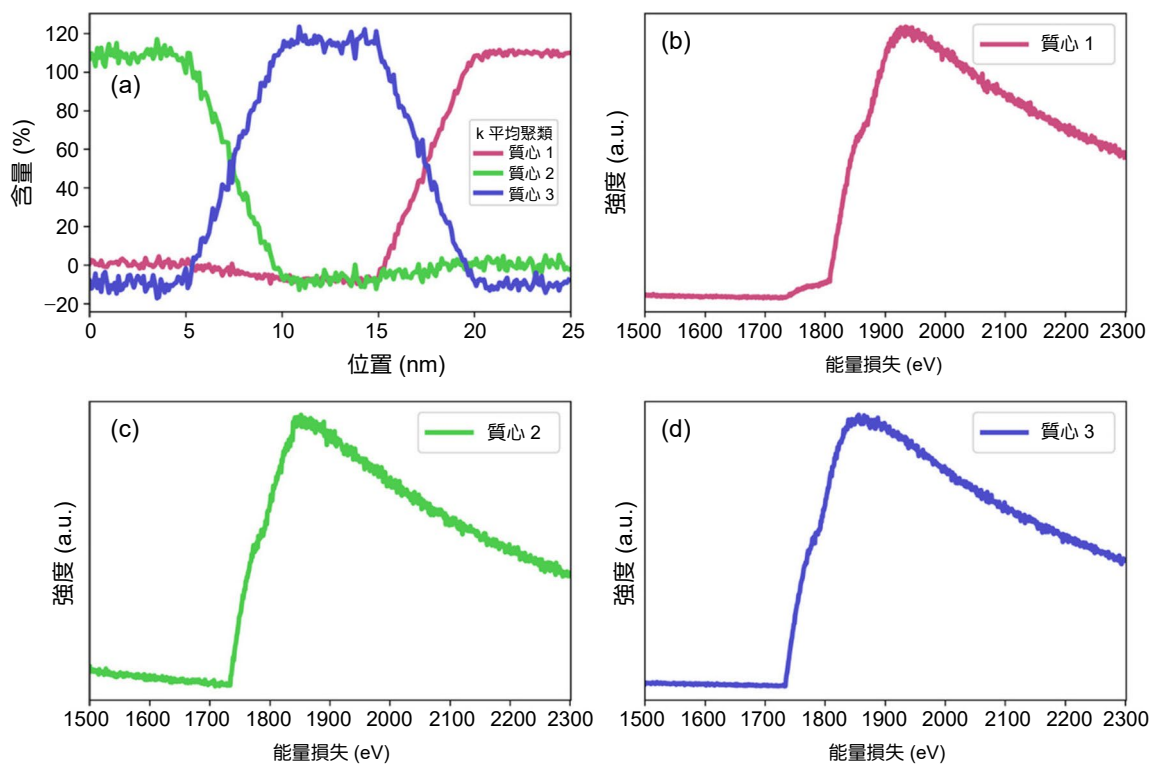


圖 7. k 平均聚類之各質心含量分佈與能譜。

目一致，但各質心皆為混合相能譜。其原因為 k 平均聚類中，每個能譜 (包含混合區) 皆會分配至各群組中，質心為各群組之能譜平均結果，故必為混合相能譜，因此造成含量計算上的誤差。

圖 8(a) 為 kMLLS 聚類各質心含量分佈，從結果可看出各質心成分約介於 0 至 100%，其分佈與設計之結果相符。因此我們可以了解，儘管各材料之間能譜互相重疊情形嚴重，其終端材料能譜仍可被正確的萃取出來。圖 8(b) 至 (d) 為 kMLLS 聚類萃取之中端能譜，分別為接近純相之鎢、鈦矽化鉬以及鉬。

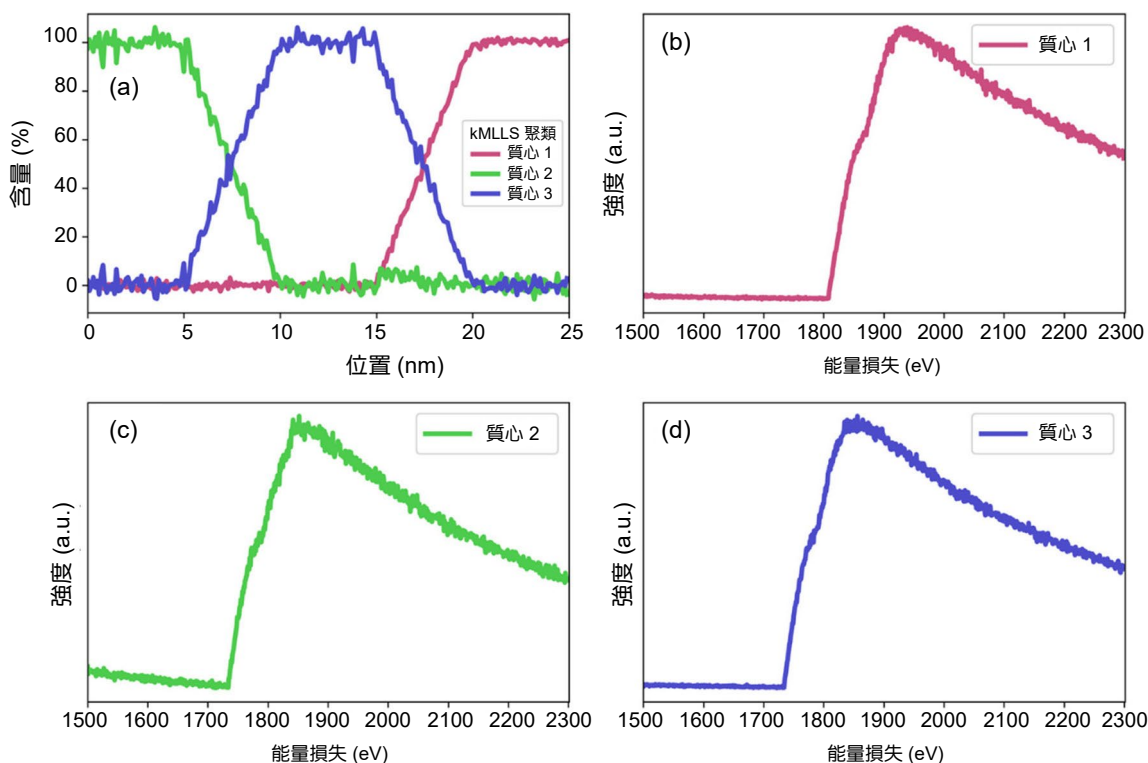


圖 8. kMLLS 聚類之各質心含量分佈與能譜。

3. ONO 薄膜分析

最後我們以實驗數據來驗證 kMLLS 聚類演算法之可行性。圖 9 為半導體常見的 ONO 薄膜之電子能量損失光譜儀線掃描能譜影像。其掃描電子源能量為 200 千伏，並同時收集零損失 (zero loss) 與核損失 (core loss) 能譜，以便進行每個分析位置之能量位置校正。如前所述， k 平均聚類與 kMLLS 聚類中能譜相似度是由各能譜與所在群組之質心的歐幾里德距離所定義的。然而在電子能量損失能譜中，其訊號強度會隨著能量呈指數衰減。因此在此分析中，我們僅分析主要元素之能量區間，以避免背景對相似度計算所造成的影響，除此之外我們未對實驗數據進行其他訊號處理。圖 9(a) 左側為 ONO 薄膜的環暗場影像 (annular dark-field image, 簡稱 ADF 影像) 與其對應位置之能譜影像。首先我們以 Elbow 法得到初始之 k 值為 2，結果如圖 9(b) 所示。直接以 k 平均聚類所得到之各質心能譜分佈結果顯示於圖 9(c)，從圖中可發現質心 2 在中間區域之含量約為 110%，質心 1 則約為 -10%。進一步以 kMLLS 聚類進行萃取，各質心之分佈列於圖 9(d)，各質心成份皆落在約 0—100% 之

間。為了了解分析過程中質心的變化，圖 9(e) 與 (f) 分別顯示質心 1 與質心 2 之 k 平均聚類與 kMLLS 聚類所萃取出之質心前後比較。從圖中可觀察到在 k 平均聚類中，如箭頭所指之處，其萃取出之質心仍有少部份的混合相存在。相對的，kMLLS 聚類則較接近純相之能譜，經後續人工鑑定可知質心 1 為氧化層，質心 2 則為氮化層。由此結果可知，藉由 kMLLS 聚類可以有效的排除混合相的能譜，萃取出系統中真實的終端材料。

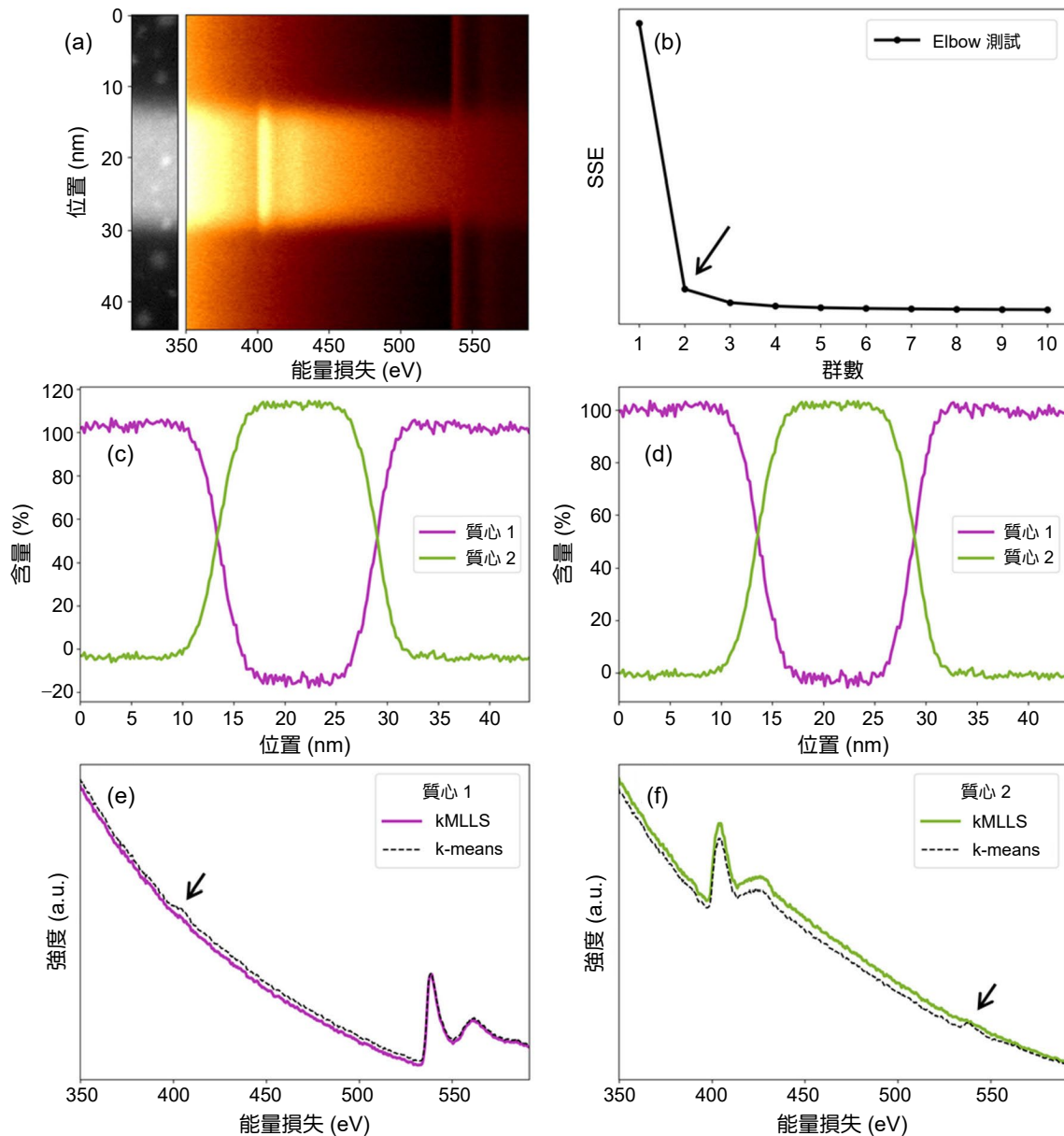


圖 9. (a) ONO 薄膜之環暗場影像 (左) 以及能量損失能譜影像 (右)。(b) Elbow 測試結果，結果顯示 $k=2$ 。(c) 與 (d) 依序為利用 k 平均聚類與 kMLLS 聚類之各質心分佈結果。(e) 與 (f) 分別為質心 1 與質心 2 之 k 平均聚類與 kMLLS 聚類比較。箭頭處指出在 k 平均聚類中的質心仍有少部份的混合成份在其中。

五、結論

在能譜分析中，如何有效且快速的從數據中萃取終端材料能譜是很重要的課題，而傳統方法仍有待改善。在本文中，我們提出了一種基於機器學習所開發之 kMLLS 聚類演算法，其結合 k 平均聚類以及複線性最小平方擬合之優點。過程中不需要手動設定參考能譜，且可以準確的萃取出終端材料能譜。為了驗證演算法之可靠性，我們以兩組模擬數據以及一組實驗數據進行測試，結果顯示利用 kMLLS 聚類演算法，可以有效的排除混合相，得到接近純相的能譜。此外，若各終端材料之間能譜嚴重重疊，我們仍可以有效的將其分離，得到準確的終端材料能譜。由於 kMLLS 聚類可以自動的將數據中的終端材料萃取出來，因此未來可以應用於線上分析，即時提供重要材料資訊予分析人員。

參考文獻

1. S. J. Pennycook, P. D. Nellist, *Scanning Transmission Electron Microscopy*, New York :Springer, (2011).
2. R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*, Boston: Springer, (2011).
3. J. A. Hunt, et al, *Ultramicroscopy* **38**, 47 (1991).
4. C. Jeanguillaume, et al, *Ultramicroscopy* **28**, 252 (1989).
5. V. C. Angadi, et al., *Journal of Microscopy*, **262**, 157 (2016).
6. F. Chollet, *Deep Learning with Python*, New York: Manning Publications, (2017).
7. S. Raschka, V. Mirjalili, *Python Machine Learning*. (Birmingham: Packt Publishing Ltd., (2017).
8. L. Miao, et al., *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 765 (2007).
9. N. Keshava, et al., *IEEE Signal Processing Magazine*, **19**, 44 (2002).
10. M. Arngren, et al., *Journal of Signal Processing Systems*, **65**, 479 (2011).
11. D. Rossouw, et al., *Nano Letters*, **15**, 2716 (2015).
12. R.-F. Cai et al., *New Journal of Physics*, doi:10.1088/1367-2630/ab7a89., (2020)
13. J. M. Titchmarsh, et al., *Journal of Microscopy*, **184**, 195 (1996).
14. P. Potapov, et al., *Advanced Structural and Chemical Imaging*, **5**, 4 (2019).
15. P. Cueva, et al., *Microscopy and Microanalysis*, **18**, 667 (2012).
16. F. dela Pena, et al., *Ultramicroscopy*, **111**, 169 (2011).
17. D. Lee, et al., *Nature*, **401**, 788 (1999).
18. N. Gillis, arXiv:1401.5226 [stat.ML], 1-25 (2014).
19. N. Wang, et al., *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **6**, 554 (2013).
20. R. Arenal, et al., *Ultramicroscopy*, **109**, 32 (2008).
21. H. Shuman, et al., *Ultramicroscopy*, **21**, 23 (1987).
22. R. D. Leapman, et al., *Ultramicroscopy*, **26**, 393 (1988).
23. P. Torruella, et al., *Ultramicroscopy*, **185**, 42 (2018).
24. P. M. Voyles, *Current Opinion in Solid State and Materials Science*, **21**, 141 (2017).
25. A. K. Jain, *Pattern Recognition Letters*, **31**, 651 (2010).
26. D. A. Shapiro, et al, *Nature Photonics*, **8**, 765 (2014).
27. V. Kumar, *Journal of Structural Biology*, **145**, 41 (2004).
28. S. Jesse, et al., *Scientific Reports*, **6**, 26348 (2016).
29. Z. Yang, et al., *Structure*, **20**, 237 (2012).
30. T. H.Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, C. *Introduction to algorithms*, The MIT Press, (2009).
31. Peña, F. dela et al., doi:10.5281/ZENODO.3396791, (2019).
32. F. Pedregosa, et al., *Journal of Machine Learning Research*, **12**, 2825 (2011).
33. N. K. Menon, et al., *Microscopy and Microanalysis*, **8**, 203 (2002).
34. P. Blaha, et al., *The Journal of Chemical Physics*, **152**, 074101 (2020).
35. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, (2007).

作者簡介

蔡任豐先生現為國立清華大學工程與系統科學研究所博士生及工業技術研究院材料與化工研究所副研究員。

Ren-Fong Cai is currently a Ph.D. student in Department of Engineering and System Science at National Tsing Hua University and an associate researcher at Material and Chemical Research Laboratories, ITRI.

張睦東先生為國立清華大學材料科學工程研究所博士，現為工業技術研究院材料與化工研究所研究員。

Mu-Tung Chang received his Ph.D. in Department of Materials Science and Engineering from National Tsing Hua University. He is currently a researcher at Material and Chemical Research Laboratories, ITRI.

羅聖全先生為國立清華大學工程與系統科學研究所博士，現為工業技術研究院材料與化工研究所副組長及正研究員。

Shen-Chuan Lo received his Ph.D. in Department of Engineering and System Science from National Tsing Hua University. He is currently a deputy division director and principal researcher at Material and Chemical Research Laboratories, ITRI.

陳健群先生為加州大學洛杉磯分校物理博士。目前任職於國立清華大學工程與系統科學系，並為國家同步輻射研究中心合聘助理研究員。

Chien-Chun Chen received his Ph.D. in Physics from California University, Los Angeles (UCLA), USA. He is currently an assistant professor in Department of Engineering and System Science at National Tsing Hua University and he is an assistant researcher of National Synchrotron Radiation Research Center.