

Introduction

As technology advances, high throughput screening via automation¹ and hydrothermal continuous flow synthesis² leads to the increase in larger materials science datasets. Unlike Rietveld refinement, Principal Component Analysis offers a rapid interpretation of XRD patterns through production of relevant simulated patterns to create an underlying map to present the real data on. $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$ is used to test PCA on representing crystallite size and unit cell parameter, and composition.

Experimental

OriginLab was used for PCA calculations and data transformation

Hydrothermal synthesis of $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$

Composition: BaTiO_3 , $\text{Ba}_{0.5}\text{Sr}_{0.5}\text{TiO}_3$, SrTiO_3

Temperatures: 150 °C, 200 °C

Times: 2 hr, 4 hr, 6 hr, 24 hr

Simulation of $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$ XRD patterns

Unit cell parameters: 3.9 Å to 4.1 Å in steps of 0.02 Å.

Crystallite size: 2 nm to 20 nm in steps of 2 nm, 10 nm to 160 nm in steps of 15 nm.

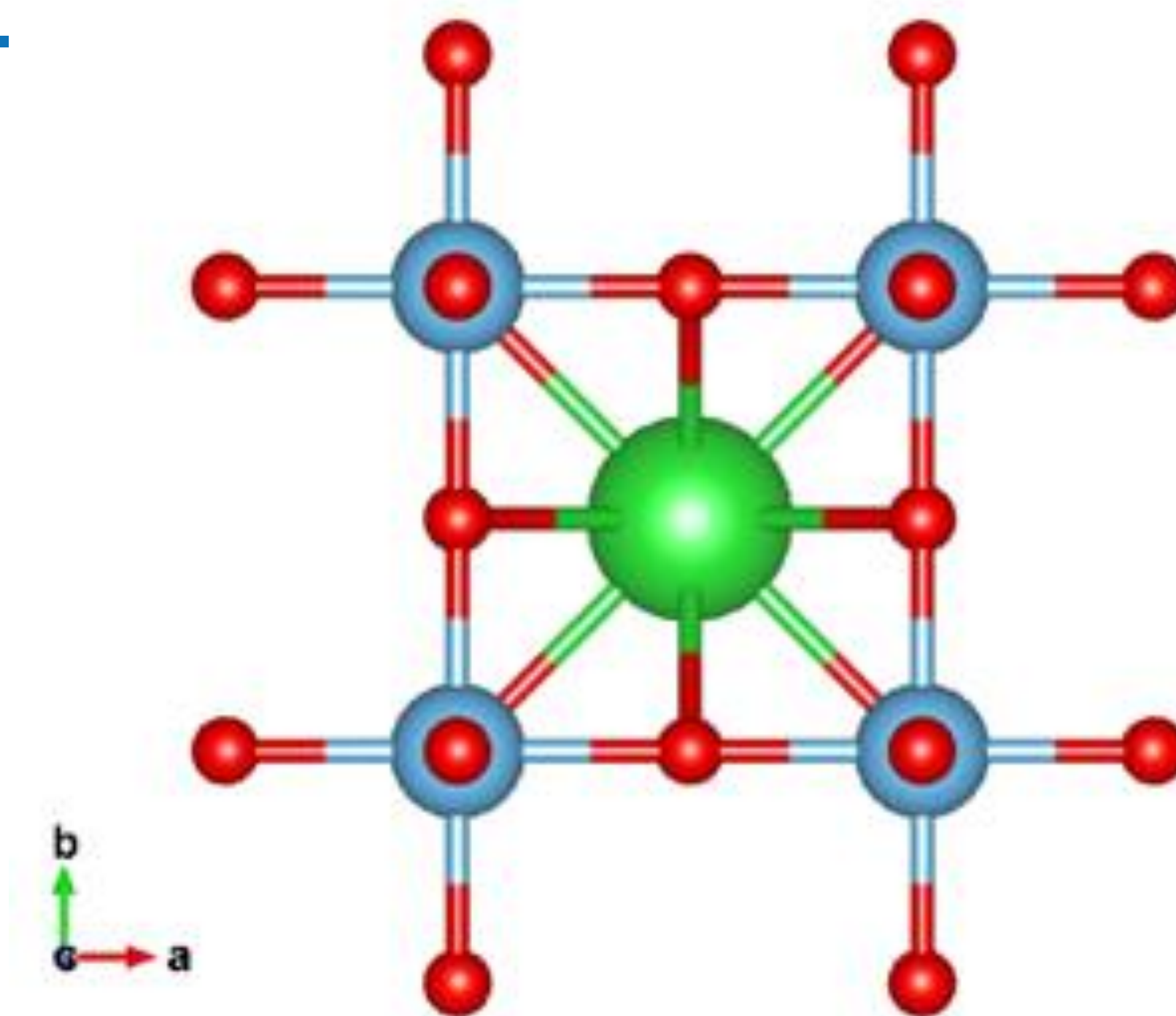


Fig. 1. Structure of BaSrTiO_3 .³

Simulated Dataset

XRD patterns broaden with change in crystallite size and shift in the 2θ direction with change of unit cell parameter (Fig 2 (a)). 71.6% of the information of the dataset is stored within PC1 (55.2%) and PC2 (16.4%) which correlate to crystallite size and unit cell parameter respectively (Fig 2, (b)). Using the score plot (Fig 2 (c)), verifies this relationship. Positive PC1 displaying larger sizes, negative with smaller sizes. PC2 represents larger unit cell parameter in the positive direction and smaller in the negative.

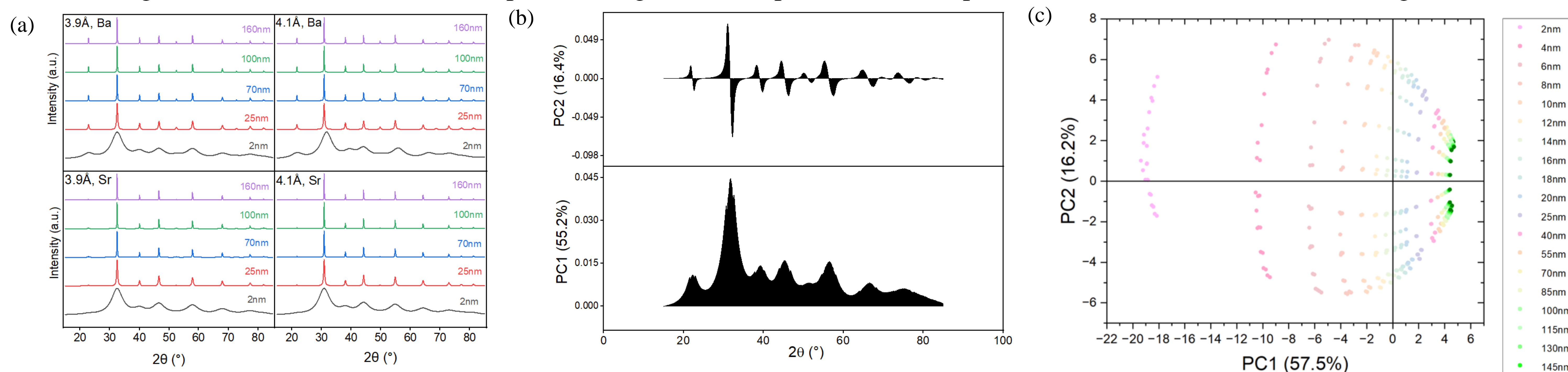


Fig. 2. (a) Simulated XRD patterns of $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$, (b) Loading plots of PC1 and PC2, from PCA, (c) Score plot of PC1 vs PC2 labelling crystallite size.

Real and Simulated Dataset

XRD patterns show a shift in 2θ with change of composition, this is consistent with associated SEM images (Fig 3 (a)). The respective loading plot (Fig 3 (b)) displays unit cell parameter as PC1 (60.1%), crystallite size as PC3 (2%), and an artifact from PCA as PC2 (35%). The combined score plot (Fig 3 (c)) allows the crystallite size for each real crystal to be identified. The same can be said when labelling unit cell parameters. Material composition has more influence on crystalline properties than time or temperature. Results are consistent when comparing to Rietveld refinement findings.

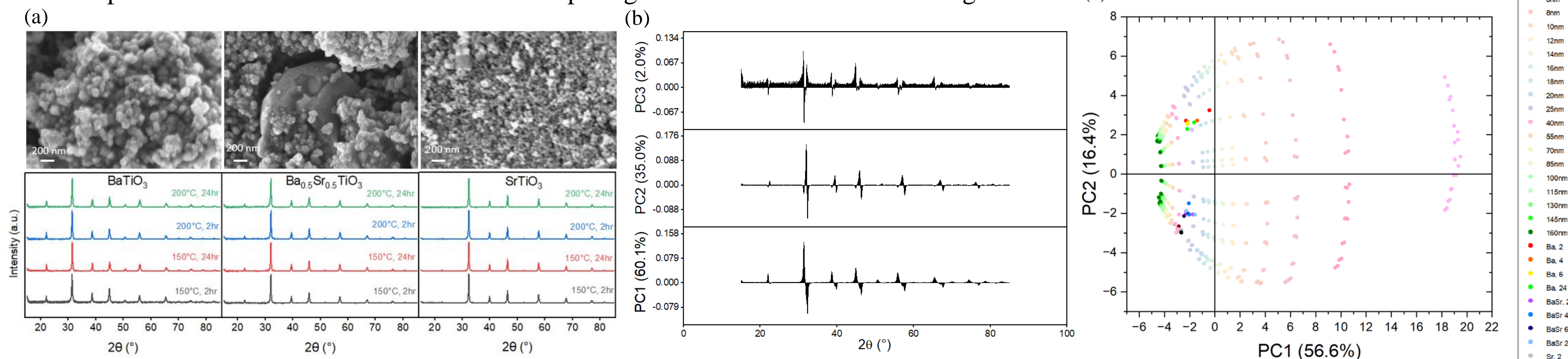


Fig. 3. (a) Real XRD patterns of $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$ with associated SEM Images, (b) Loading plot for real dataset, (c) PCA real and simulated dataset, labelling crystallite size.

Discussion & Conclusions

PCA offers a method of rapidly interpreting large datasets of XRD patterns, however it has limitations. Some crystalline properties, such as composition, cannot be distinguished. A large set of points gives way for the formation of artificial principal components with high variance. PCA trades accuracy for better visualisation and interpretation of data, hence it should not be employed solely, and instead to be heavily applied along-side other material property extraction techniques. Despite its constraints, the use of PCA on large XRD datasets can inevitably lead to the higher throughput of materials science discoveries.