Quantitative measurement of chemical composition in small alloy nanoparticles by STEM-EDS and Machine Learning

Murilo Henrique Martinez Moreira

▶ **To cite this version:**

Murilo Henrique Martinez Moreira. Quantitative measurement of chemical composition in small alloy nanoparticles by STEM-EDS and Machine Learning. Materials Science [cond-mat.mtrl-sci]. Université Claude Bernard - Lyon I; Universidade estadual de Campinas (Brésil), 2022. English. NNT : 2022LYO10038 . tel-04072449

HAL Id: tel-04072449

https://theses.hal.science/tel-04072449

Submitted on 18 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE de DOCTORAT DE
L'UNIVERSITE CLAUDE BERNARD LYON 1

École Doctorale N°52
Physique et Astrophysique

Discipline : Nanoscience, matériaux

Soutenue publiquement le 30/09/2022, par :

Murilo Henrique MARTINEZ MOREIRA

Quantitative measurement of chemical composition in small alloy nanoparticles by STEM-EDS and Machine Learning

---

Devant le jury composé de :

RIUL JR, Antonio, Dr, UNICAMP, Brésil.	Président
T. J. HELVOORT, Antonius, Dr, NTNU, Norvège	Rapporteur
E. ERNST, Wolfgang, Dr TU Graz, L'Autriche	Rapporteur
ZANCHET, Daniela, Dr, UNICAMP, Brésil	Examinatrice
AMANS, David, Dr, UCBL1, France	Examinateur
DIVITINI, Giorgio, Dr, IIT, Italie	Examinateur
HILLENKAMP, Matthias, Dr, UCBL1, France	Directeur de thèse

Universidade Estadual de Campinas and Université Claude Bernard Lyon 1
Instituto de Física Gleb Wataghin and Institut Lumière Matière

Murilo Henrique Martinez Moreira

Quantitative measurement of chemical composition in small alloy nanoparticles by STEM-EDS and Machine Learning.

Medição quantitativa da composição química de pequenas nanopartículas por STEM-EDS e Machine Learning.

CAMPINAS
2022

Murilo Henrique Martinez Moreira

Quantitative measurement of chemical composition in small alloy nanoparticles by STEM-EDS and Machine Learning

Medição quantitativa da composição química de pequenas nanopartículas por STEM-EDS e Machine Learning

Ph.D. thesis submitted to the Institute of Physics "Gleb Wataghin" (Brazil) of the University of Campinas and Institut Lumière Matière of the University Claude Bernard Lyon 1 (France) as part of the requirements for obtaining the title of Doctor of Science in the field of Applied Physics.

Tese de doutorado apresentada ao Instituto de Física "Gleb Wataghin" (Brasil) da Universidade Estadual de Campinas e Institut Lumière Matière da Université Claude Bernard Lyon 1 (França) como parte dos requisitos exigidos para a obtenção do título de Doutor em Ciências, na Área de Física Aplicada.

Advisors: Dr. Varlei Rodrigues (UNICAMP) / Dr. Matthias Hillenkamp (UCBL1)

This version of the thesis corresponds to the final version defended by the student Murilo Henrique Martinez Moreira and coordinated by Dr. Varlei Rodrigues and Dr. Matthias Hillenkamp. The thesis was defended in accordance with a cotutelle agreement between the two institutions from UNICAMP (Brazil) and UCBL1 (France)

CAMPINAS
2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Física Gleb Wataghin
Lucimeire de Oliveira Silva da Rocha - CRB 8/9174

M813q

Moreira, Murilo Henrique Martinez, 1993-
Quantitative measurement of chemical composition in small alloy nanoparticles by STEM-EDS and machine learning / Murilo Henrique Martinez Moreira. – Campinas, SP : [s.n.], 2022.

Orientadores: Varlei Rodrigues e Matthias Daniel Hillenkamp.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin.
Em cotutela com: Université Claude Bernard Lyon 1.

1. Nanopartículas. 2. Microscopia eletrônica de transmissão de varredura. 3. Espectroscopia de raio X dispersados em energia. 4. Análise de componentes principais. 5. Análise quantitativa. 6. Aprendizado de máquina. I. Rodrigues, Varlei, 1973-. II. Hillenkamp, Matthias Daniel, 1971-. III. Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. V. Título.

Informações Complementares

Título em outro idioma: Medição quantitativa da composição química de pequenas nanopartículas por STEM-EDS e machine learning
Palavras-chave em inglês:
Nanoparticles
Scanning transmission electron microscopy
Energy dispersive X-ray spectroscopy
Principal components analysis
Quantitative analysis
Machine learning
Área de concentração: Física Aplicada
Titulação: Doutor em Ciências
Banca examinadora:
Antonio Riul Júnior
Matthias Daniel Hillenkamp
Wolfgang Erhard Ernst
Antonius Theodorus Johannes Van Helvoort
Giorgio Divitini e Daniela Zanchet e David Amans
Data de defesa: 30-09-2022
Programa de Pós-Graduação: Física

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-9209-4993
- Currículo Lattes do autor: http://lattes.cnpq.br/9908625620454215

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DO ALUNO MURILO HENRIQUE MARTINEZ MOREIRA - RA 122806 APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 30/09/2022.

COMISSÃO JULGADORA:

- Prof. Dr. Antonio Riul Junior – Presidente (IFGW/UNICAMP)

- Dr. Matthias Daniel Hillenkamp (Institut Lumière Matière – Universitè Claude Bernard Lyon 1)

- Prof. Dr. Wolfgang Erhard Ernst (Graz University of Technology)

- Prof. Dr. Antonius van Helvoort (Norwegian University Of Science And Technology)

- Dr. Giorgio Divitini (Italian Institute of Technology)

- Profª. Dra. Daniela Zanchet (IQ/UNICAMP)

- Dr. David Amans (Institut Lumière Matière – Universitè Claude Bernard Lyon 1)

OBS.: Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

CAMPINAS

2022

Acknowledgements

To my parents Ana Aparecida Martinez Moreira and João Aparecido Moreira and my wife Angélica De Salvo Miotti, who always encouraged me in my work, supporting and allowing me to do my best.

To my brothers, family members and dear friends, who share my achievements.

To Prof. Dr. Varlei Rodrigues, Dr. Matthias Hillenkamp and Prof. Dr. Daniel Ugarte for their teachings and for always being present and willing to help me learn and grow professionally.

To the friends of GFNMN, for the moments of knowledge exchange and fun

To "Ministry of Foreign Affairs" and the "Campus France" Eiffel Excellence scholarship.

To "Ministry of Science, Technology and Innovation" and the "National Council for Scientific and Technological Development – CNPq"

To all those who somehow were present and contributed to the development of this work and my academic training.

Abstract

Bimetallic nanoparticles (BNPs) attract attention in both fundamental and applied sciences. Here the BNP properties depend not only on size, shape, and environment but also on their chemical composition and atomic ordering. One of the most intensely studied bimetallic systems is the gold-silver alloy NPs. In bulk, the elements are entirely miscible, but in nanoscale, the arrangement of the atomic species is still unsolved for small surfactant-free BNPs (Diam. $<$ 10 nm). The complete characterization of nanometric systems requires detailed knowledge of how the different chemical elements of the materials are distributed at the interfaces/surfaces (roughness, interdiffusion, etc.); and, finally, how these factors modify the electronic properties of the system. Henceforth, we propose the study of chemical composition and elemental distribution within AuAg BNPs produced by gas-phase aggregation. Here, to investigate the atomic organization and chemical composition of AuAg BNP as function of size, high-resolution Energy Dispersive X-ray Spectroscopy and Scanning Transmission Electron Microscopy (EDS-STEM) are used quantitatively. Thus, we aim to use the EDS-STEM technique and hyperspectral image (HSI) data treatment using machine learning to improve the quantitative chemical composition information extraction. EDS-STEM has gained significant improvement in its acquisition systems, making it possible to perform chemical analysis of nanometric objects, such as bimetallic nanoparticles (BNPs). These advances turn possible to extract quantitative information from individual small BNPs, opening the path for understanding chemical composition vs. size physico-chemical properties. This is a significant improvement concerning purely qualitative chemical mapping, as it is widely used in various communities. Proper quantification requires confidence intervals; therefore, we show a methodology to estimate uncertainty in chemical composition analysis after machine learning denoising by Principal Components Analysis (PCA), evaluating its detection and quantification limits in the application for elemental assessment in small BNPs. Thus, we quantify the chemical composition of AuAg BNPs and identify, for example, size-dependent composition effects hidden by Poisson noise. Finally, we propose using machine learning tools, such as PCA and Non-Negative Matrix Factorization (NMF), to study the unmixing of signals in HSIs. Here we aim to know the elemental distribution inside the AuAg BNPs. We show that we can measure a chemical gradient of 0.45 ± 0.02 to 0.62 ± 0.02 (Ag at. fraction) from the projected center of the BNP to its surface. The Ag enrichment towards the surface characterizes a core rich in Au and a shell rich in Ag. Consequently, to explain the radial Ag enrichment, we studied the effect of reactivity towards oxidation by analyzing carbon-protected and unprotected BNPs. Hence, due to the quantitative aspect of our analysis, we can know the number of atoms in different regions of the BNP. This allowed us to identify that carbon-protect and annealed samples show an increased level of alloying and only a small fraction of Ag enrichment in the surface, in contrast to the oxidized samples where the Ag atoms strongly tend to the BNP surface. Finally, with quantitative analysis methodologies employed in this thesis, the chemical composition characterization of small volumes can be improved to support theory and simulations of modeling in nanoscale physics and chemistry for both fundamental and applied studies.

Résumé

Les nanoparticules bimétalliques (BNP) attirent l'attention dans les sciences fondamentales et appliquées. Les propriétés des BNP dépendent non seulement de leur taille, de leur forme et de leur environnement, mais aussi de leur composition chimique et de leur ordonnancement atomique. Les NP d'alliage or-argent constituent l'un des systèmes bimétalliques les plus étudiés. Dans le volume, les éléments sont entièrement miscibles, mais à l'échelle nanométrique, la question de l'arrangement des espèces atomiques n'est toujours pas résolu pour les petites NPs (Diam. $<$ 10 nm). La caractérisation complète des systèmes nanométriques nécessite une connaissance détaillée de la façon dont les différents éléments chimiques des matériaux sont distribués aux interfaces/surfaces (rugosité, interdiffusion, etc.) ; et, finalement, comment ces facteurs modifient les propriétés électroniques du système. Ici, nous proposons l'étude de la composition chimique et de la distribution des éléments a l'interieur des BNP AuAg produits par agrégation en phase gazeuse. Pour étudier l'organisation atomique et la composition chimique des BNP d'AuAg en fonction de leur taille, nous utilisons la spectroscopie aux rayons X à dispersion d'énergie à haute résolution et la microscopie électronique à transmission à balayage (EDS-STEM) de manière quantitative. Nous souhaitons donc utiliser la technique EDS-STEM et le traitement des données d'images hyperspectrales (HSI) par apprentissage automatique pour améliorer l'extraction d'informations sur la composition chimique quantitative. La technique EDS-STEM a connu des améliorations significatives dans ses systèmes d'acquisition, rendant possible l'analyse chimique d'objets nanométriques, tels que les nanoparticules bimétalliques (BNP). Ces progrès permettent d'extraire des informations quantitatives de petites BNP individuelles, ouvrant la voie à la compréhension de la composition chimique par rapport aux propriétés physico-chimiques en fonction de la taille. Il s'agit d'une amélioration significative par rapport à la cartographie chimique purement qualitative, qui est largement utilisée dans diverses communautés. Une quantification correcte nécessite des intervalles de confiance; par conséquent, nous montrons une méthodologie pour estimer l'incertitude dans l'analyse de la composition chimique après un débruitage par apprentissage automatique en utilisant l'analyse en composantes principales (PCA), en évaluant ses limites de détection et de quantification dans l'application pour l'évaluation élémentaire des petits BNP. Ainsi, nous quantifions la composition chimique des BNP AuAg et identifions, par exemple, les effets de composition dépendant de la taille autrement cachés par le bruit de Poisson. Enfin, nous proposons d'utiliser des outils d'apprentissage automatique, tels que la PCA et la factorisation de matrices non négatives (NMF), pour étudier le démixage des signaux dans les HSI. Ici, nous cherchons à connaître la distribution élémentaire à l'intérieur des BNP AuAg. Nous montrons que nous pouvons mesurer un gradient chimique de 0.45 ± 0.02 à 0.62 ± 0.02 (fraction d'Ag at.) du centre projeté du BNP à sa surface. L'enrichissement en Ag vers la surface caractérise un cœur riche en Au et une coquille riche en Ag. Par conséquent, pour expliquer l'enrichissement radial en Ag, nous avons étudié l'effet de la réactivité à l'oxydation en analysant des BNP protégés et non protégés par du carbone. Ainsi, grâce à l'aspect quantitatif de notre analyse, nous pouvons connaître le nombre d'atomes dans les différentes régions du BNP. Cela nous a permis d'identifier que les échantillons protégés par du carbone et recuits montrent un niveau augmenté d'alliage et seulement une petite fraction d'enrichissement en Ag dans la surface, contrairement aux échantillons

oxydés où les atomes d'Ag tendent fortement vers la surface du BNP. Enfin, avec les méthodologies d'analyse quantitative employées dans cette thèse, la caractérisation de la composition chimique de petits volumes peut être améliorée pour soutenir la théorie et les simulations de modélisation en physique et chimie à l'échelle nanométrique pour des études fondamentales et appliquées.

,

Acronyms

ADF Annular Dark Field.

AEM Analytical Electron Microscopy.

BNP Bimetallic Nanoparticles.

BSS Blind Source Separation.

ED Electron Diffraction.

EDS Energy Dispersive X-ray Spectroscopy.

EELS Electron Energy Loss Spectroscopy.

EM Electron Microscopy.

HAADF High Angle Annular Dark Field.

HSI Hyperspectral Image.

ICA Independent Component Analysis.

LSPR Localized Surface Plasmon Ressonance.

NMF Non-Negative Matrix Factorization.

PCA Principal Components Analysis.

PDF Pair Distribution Function.

PED Precession Electron Diffraction.

SNR Signal-to-Noise Ratio.

SPM Scanning Probe Microscope.

STEM Scanning Transmission Electron Microscopy.

TEM Transmission Electron Microscopy.

Contents

1 Introduction **13**

2 Metallic Alloy Nanoparticles **16**
2.1 Physico-Chemical Properties 17
2.2 Nanoparticle Synthesis 22
2.3 Model Sample: Bimetallic AuAg Nanoparticles 25
2.4 Summary 28

3 Scanning Transmission Electron Microscopy **29**
3.1 Imaging in STEM 30
3.2 Analytical Electron Microscopy: Energy Dispersive X-ray Spectroscopy . 33
3.2.1 Characteristic X-ray generation and detection 34
3.2.2 Elemental mapping: Hyperspectral imaging 37
3.3 Summary 39

4 Quantitative EDS Chemical Analysis **40**
4.1 Cliff-Lorimer method 41
4.2 Chemical composition quantification of individual alloy nanoparticles . . 44
4.3 Summary 45

5 Machine Learning and Big data in Electron Microscopy: Advancing in chemical composition assesment **47**
5.1 Principal Component Analysis 49
5.1.1 Component Analysis in Hyperspectral EDS-STEM images: Denoising and feature extraction 53
5.1.2 Statistical Bias and Information Loss 56
5.2 Non-Negative Matrix Factorization 64
5.3 Partial conclusion 67

6 Exploiting Principal Component Analysis for denoising EM image spectra: Effects on quantitative EDS analysis **68**

7 Exploiting Principal Component Analysis for denoising EM image spectra: Uncertainty intervals **85**
7.1 Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part II, Uncertainty intervals 86

8 Chemical composition quantification within Bimetallic Nanoparticles: Measuring chemical gradients **97**
8.1 Elemental distribution of bimetallic AuAg nanoparticles 97
8.2 Effects of oxidation in the elemental distribution inside AuAg NPs 100
8.3 Non-Negative Matrix Factorization for AuAg NP studies: Measuring chemical gradients 109
8.4 Summary 123

9 Conclusions and Perspectives **124**

Bibliography **128**

A Appendix **141**

Chapter 1
Introduction

Bimetallic nanoparticles (BNPs) are systems formed by two metals, which can display different types of structures and adopt different morphologies and even chemical order [1, 2, 3]. By controlling the size and chemical composition, different physical and chemical properties can be obtained due to quantum, and surface-induced phenomena [4]. Hence, it is of great interest to establish reliable methods for the characterization of chemical composition and elemental distribution of a single small BNP (sub-10 nm) [5]. In this context, developing new tools with specific and well-adapted capabilities for nanosystems becomes essential. The study of individual nanosystems requires instruments that have a high spatial resolution. Electron microscopy (EM), as well as scanning probe microscopes (SPM), are specially adapted as they allow obtaining images of morphology, structural information, and even nanometer resolution spectroscopic information [6, 7]. Therefore, in this thesis, we wish to explore the measurement of the chemical composition of AuAg BNPs using Energy-dispersive X-ray Spectroscopy (EDS) combined with a Scanning Transmission Electron Microscope (STEM). This way, we expect to contribute in both EM and nanoalloy communities by the employ of quantitative EDS chemical analysis of small BNPs (Diam. $<$ 10 nm).

In chapter 2.1 we discuss the physico-chemical properties of the BNPs, especially the AuAg model. Also, we explain the methodology to synthesize our samples by the gas aggregation physical route, where the NPs are produced surfactant-free and deposited in the TEM substrate. In this chapter, we contextualize the need of reliable quantitative chemical analysis of BNPs as a function of size, where nanoscale effects play a fundamental role. Therefore, in chapter 3 we will introduce the STEM imaging and spectroscopies modes focusing in EDS elemental mapping details. Thus, we show our central methodology for quantifying the chemical composition of individual BNPs in chapter 4. In this chapter, we show how we select the pixels of interest (BNP's region) and process their spectra to result in the final quantification of the BNP's chemical composition by the Cliff-Lorimer method [8]. Here, we emphasize that proper quantitative analysis requires the establishment of confidence intervals (error bars). For that, we show the standard error propagation procedure for the obtention of these values.

Hence, we show that the principal difference between straightforward qualitative and quantitative analysis is that more X-ray counts must be measured. For EDS-STEM, due to the Poisson distribution nature of counting X-ray photons, we discuss the need to considerably increase the number of counts to attain better signal-to-noise ratios (SNR). We bring to the discussion the seminal work of Currie et al. [9, 10, 11]. It is shown that proper quantitative analysis requires uncertainty intervals and well-established criteria for detection and quantification limits, which are of utmost importance to increase the reliability of analytical studies.

Analytical electron microscopy (AEM) nowadays profits from the advances in data treatment, such as the use of unsupervised machine learning tools for Poisson noise reduction [12, 7] and feature extraction [13, 14]. Indeed, machine learning tools are nowadays being used in several different fields of nanoscience, allowing better information extraction and modeling of nanometric systems [15]. Electron microscopy is no different, and machine learning tends to become standard procedures in various applications [16]. In chapter 5 we introduce the basics concepts of unsupervised machine learning methods such as Principal Component Analysis (PCA) [17, 18] and Non-Negative Matrix Factorization [19]. In this chapter, we show the recent advances and application of these algorithms in the state-of-the-art of AEM data treatment, focusing on the applications of EDS-STEM for Poisson noise reduction (denoising) and blind source separation (BSS) [20, 21]. We discuss, supported by the literature of the last $\sim$5 years that these algorithms are very sensitive to noise, and we show their main drawbacks and pitfalls in the applications explored in this thesis. At this point, we show two works publicated during this thesis in chapters 6 and 7, where we explored PCA for denoising in EDS quantitative analysis of the AuAg BNPs. In the first chapter, we discuss statistical bias and information loss in the data set reconstruction by PCA, which induces artifacts and systematic error in our analysis. Then we finish proposing strategies to mitigate the bias in the reconstruction by the supporting of experiments and simulation interpretation. In the subsequent chapter, we show the proness of PCA as a denoising tool by the proposal of a new confidence interval afterward the data set reconstruction. Hence, by the studies in these two chapters, we expect to show that by the application of PCA denoising, we obtain improved results from the point of view of accuracy (chapter 6) and precision (chapter 7).

Moreover, motivated by the capabilities of measuring chemical composition quantitatively, we wish to contribute to the understanding of elemental distributions inside BNPs. Different chemical configurations (alloyed vs. segregated, partially segregated, chemically ordered, etc.) can change the optical, magnetic or catalytic properties considerably [3]. Furthermore, BNPs have been shown to not only permit the replacement of expensive or rare elements with cheaper and more abundant ones but that new physico-chemical properties can be achieved. AuAg nanoalloy is one o of the most studied bimetallic systems. Its BNPs have attracted attention for their catalytic [22], plasmonic [23] and photocatalytic [24] properties. In bulk, the two metals are miscible for the entire composition range. On the nanoscale, however, additional degrees of freedom have to be considered, namely the particle size and morphology, as well as possible segregation and/or chemical ordering due to thermodynamic and energetic effects [25] or environmental reactivity towards oxidation [26]. Although there are several experimental and theoretical works published on AuAg nanoparticles, the controversy about whether intrinsically one of the two metals preferentially segregates at the surface

or not is still not resolved. Therefore, in chapter 8, we discuss the elemental distribution of these systems and propose that quantitative EDS chemical analysis may provide crucial tools for understanding these questions. Besides, we propose the analysis of carbon-protected and annealed samples to try to separate effects from oxidation and natural BNP's conformation in the ground state.

Finally, in chapter 8, we study how the unsupervised machine learning applied to EDS-STEM data sets might help in the learning of not only how many atoms we have in the small AuAg BNPs, but also where they are located. For that, we propose using NMF for feature extraction in oxidized AuAg BNPs where AuAg@Ag core-shell might occur due to Ag segregation. Here, we investigate the detailed use of NMF quantitatively and finish the thesis with the establishment of reliable measurement of chemical gradients inside the BNPs, comparing both experimental and simulated results. For the last, we attended in the Appendix a publication prepared for immediate submission. The article concerns the chemical composition of the BNP's synthesized by hot atom sources. We show by the comparison of chemical composition experimentally measured, and Molecular Dynamics and Monte Carlo simulation that how plasma used as sputtering contributes to changes in the chemical composition of the BNP's due to temperature induce Ag loss. This work done in parallel was possible due to the quantitative EDS chemical analysis performed in this thesis.

Chapter 2

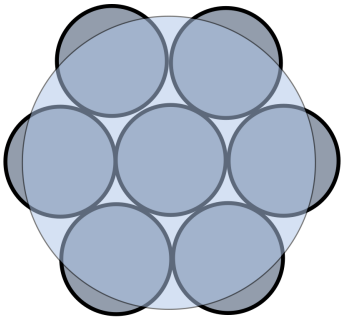
Metallic Alloy Nanoparticles

Nanoscience and Nanotechnology became popular around the 1970s and 1980s. However, nanosystems have been used for centuries to produce materials. For example, people of the Bronze Age (1200 - 1000 BC) in Frattesina di Rovigo in ancient Italy were already producing red-colored glass. This red color is attributed to the excitation of surface plasmons from copper nanoparticles located on the surface of the material, [27]. Today, the understanding of these systems is well established, and the investigation of their properties provides the fundamental knowledge for nanoscience and potential technologies, such as medical applications as, for example, hyperthermia [28] and sensor developments [29, 30].

Metal alloy nanoparticles are composed of miscible or partially miscible metal atoms, with diameters ranging from few to tens of nanometers. Their properties are defined by quantum and surface phenomena, possessing different properties depending on their size, shape, and chemical composition [31]. Nowadays, these systems can be produced in a variety of ways, with typical options being chemical synthesis in colloidal suspension [32], and physical synthesis methods in the vapor phase [33] using laser ablation or magnetron sputtering of a macroscopic metal target as the source of atoms [34]. Metal alloy nanoparticles attract massive interest in the technological community (ex., catalysis, magnetism, photonics, etc.)[1]. In this context, developing new tools with specific and well-adapted capabilities for the study of nanosystems becomes essential. The complete characterization of nanometric systems requires detailed knowledge: a) of the atomic arrangement; b) how the different chemical elements of the materials are redistributed at the interfaces/surfaces (roughness, interdiffusion, etc.); and finally, c) how the first two factors modify the electronic properties of the system. This chapter discusses bimetallic nanoparticles in general and important related scientific questions. We then introduce our model system (AuAg NPs), giving an overall idea of its synthesis and previous characterizations such as size distribution, ensemble chemical composition, and crystalline structure.

2.1 Physico-Chemical Properties

The Physico-chemical properties of nanosystems depend on size due to quantum and surface phenomena[35, 4]. In nanometric structures, the fraction of surface atoms is higher than for bulk materials, and this fraction varies with the size of the object, as illustrated by the figure 2.1. Estimated number for surface and volume atoms for structures ranging from 1 nm up to 100 nm in radius are presented below to illustrate how the surface fraction of atoms increases with decreasing system size.



Radius [nm]	n	nsurface	nsurface/n
1	125	100	0.8
2	10^3	400	0.4
10	10^5	10^4	0.08
100	10^8	10^6	0.008

Figure 2.1: Schematic representation of atomic cluster with most of its atoms at the surface and estimatives of the number of atoms in the volume and surface for different sizes.

This intrinsic property of surface/volume atom ratio for nanosystems leads to the occurrence of geometric effects, such as the relationship between melting temperature and nanoparticle size, as illustrated by the figure 2.2. In this example, the melting temperature of gold is considerably reduced for nanoparticles smaller than 5 nm with lower melting temperatures for smalles NPs [36, 37]. This occurs due to the lower coordination number of surface atoms which are more weakly bound to the material than bulk atoms, causing the melting temperature to decrease.

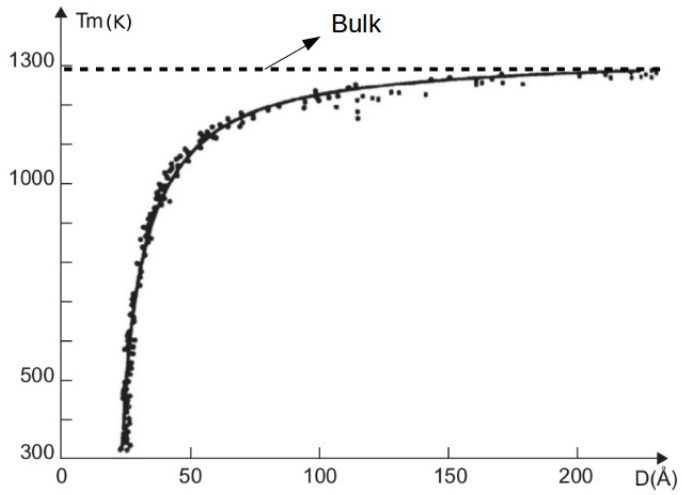


Figure 2.2: Melting points for Au nanoparticles as a function of diameter. Points represent experimental data, the dashed line shows the bulk melting point. The solid line is a curve fitted to the experimental data. Adapted from [36].

Adding different atomic species in these nanosystems adds an extra degrees of freedom, inducing new geometrical and morphological effects. Therefore, just as is done for macroscopic metallic alloys, combining the properties of the elements at the nanoscale makes it possible to explore new possibilities, mixing metallic elements and forming multi-metallic systems of various structures and morphologies.

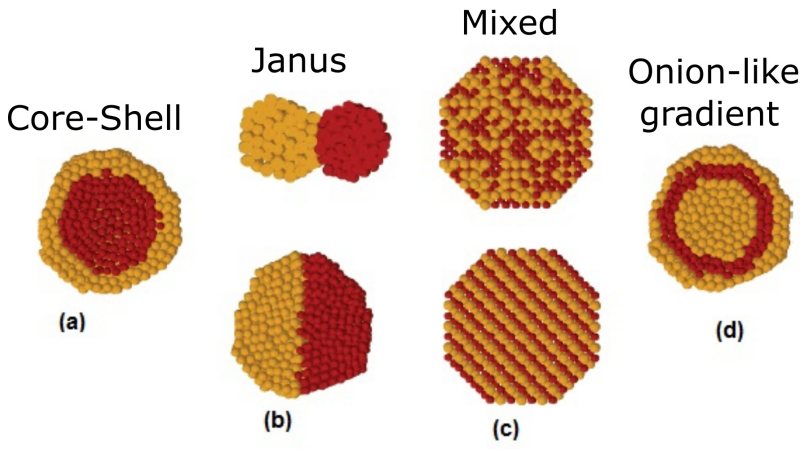


Figure 2.3: Schematic of several NP morphologies. Extracted from [1]

Assuming a binary system, for illustration purposes, figure 2.3 presents a schematic of different possibilities for the morphology of bimetallic nanoparticles: (a) Core-Shell particle, with one element composing the core and a second element composing the shell surrounding the core; (b) Janus particles (in reference to the two-headed Roman god),

where the elements are separated in two distinct phases and maybe completely segregated or forming a well-defined interface; (c) mixed alloy particles, where the elements are mixed, forming nanoalloys with or without chemical order; (d) intermediate structures, for example with alternating layers or a continuous gradient.

The same morphological patterns can be observed in systems with more than two elements. However, as the number of elements increases, the structure becomes more complex and difficult to interpret in terms of types of structures. They may even have different phases in the same object, mixing the different morphological patterns shown in the figure 2.3. Chen *et al.* [38] published a work in which multimetal nanoparticles composed of Ag, Au, Cu, Ni, and Co present at the same time mixed and Janus morphologies, forming heterostructures, as illustrated in figure 2.4. The nanoparticle at first presents, in electron microscopy imaging, a chemical contrast between two regions, revealing a Janus-type nanoparticle. However, using spatially resolved spectroscopy to generate elemental maps makes it possible to obtain complementary information indicating the formation of metallic alloys of AuAg, AuCu, and CoNi in three different domains of the nanoparticle.

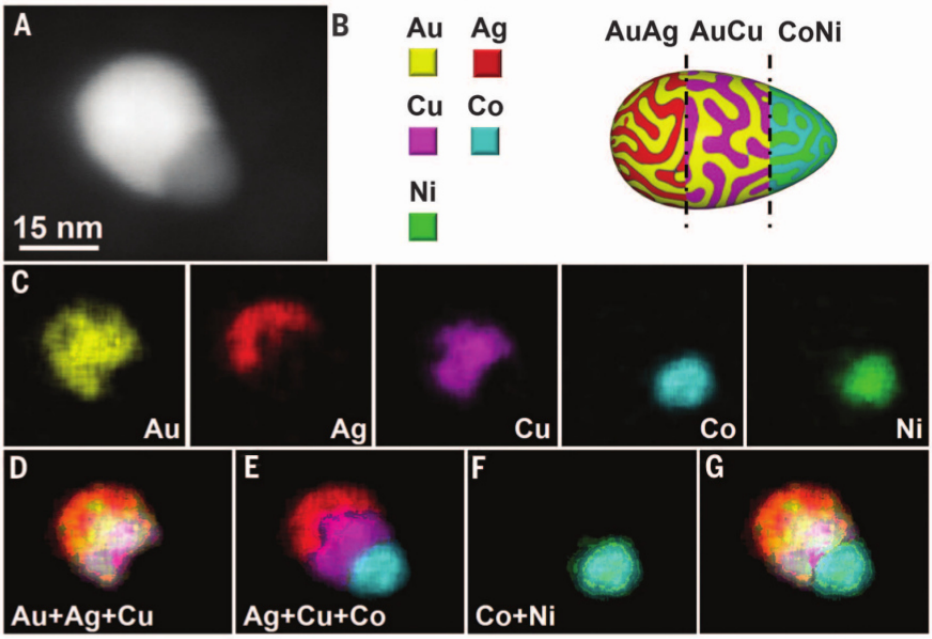


Figure 2.4: (A) Representative image of AgAuCuCoNi NP with 19% Au, 24% Ag, 28% Cu, 14% Co, 15% Ni). (B) Model of the NP measured in (A). The NP segregates in three different chemical phases of alloys: AgAu, AuCu e CoNi. (C a G) STEM-EDS elemental mapping of the NO in (A). (C) Elemental distribution of the metals into the NP. (D) Elemental mapping superposition of elements Au, Ag, e Cu. (E) Elemental mapping superposition of elements Ag, Cu, e Co. (F) Elemental mapping superposition of elements Co e Ni (G) Elemental mapping superposition of all elements. Extracted from [38].

In the same article, Chen [38] show that nanoalloy formation is dependent on chemical composition, whereas, for the case of AgAuCu nanoparticles, the nanoparticle modifies the metal alloy formation by enriching the system in Au as shown by figure 2.5.

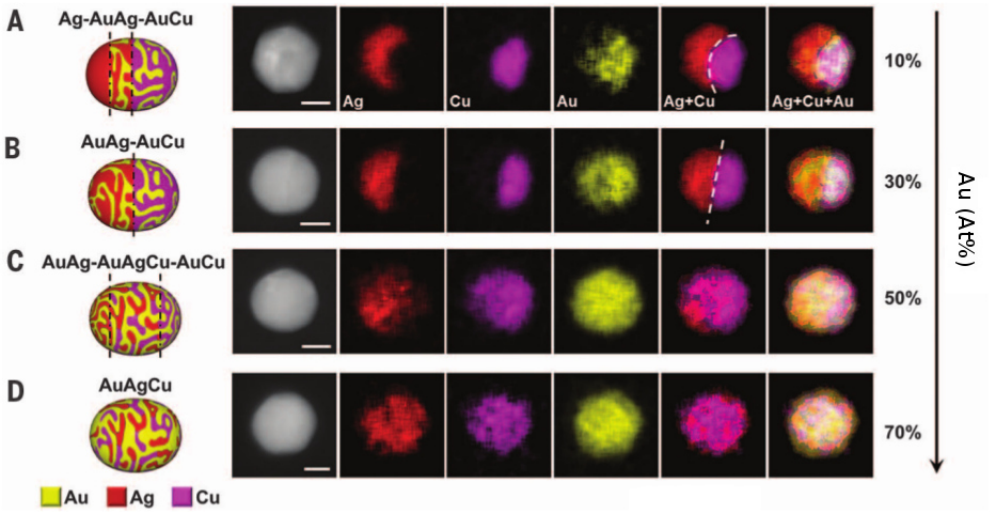


Figure 2.5: (A to D) NP schematic models, STEM-HAADF images, and STEM-EDS elemental maps of 4 representative NPs of AuAgCu, with the fraction of Au increasing from 10% to 70%. the dashed line in A and B indicates the position of the interface between the different chemical phases. The scale bar is 20 nm. Adapted from [38].

In figure 2.5, in electron microscopy images, the AgAuCu nanoparticles appear as a homogeneous system. However, by analyzing their elemental maps, it is possible to verify that with an atomic percentage of 10% of Au, the system separates int at least two distinct phases of Ag and AuCu. By gradually increasing the percentage of Au atoms, the miscibility increases proportionally until the formation of a complete AgAuCu nanoalloy on 70% Au. Spatially resolved chemical compositions can be obtained, as shown above, using spectral imaging [7], but even so, the atomic arrangement cannot be easily measured with this technique in such conditions. In recent developments of quantitative electron microscopy, atomically resolved imaging modes could be used to quantify, for example, atomic columns in nanoparticles [39] with the use of a combination of image simulation and atomic resolved images [40, 41]. Finally, Lyman [5] shows that expected miscibility gaps can be found when analyzing the quantitative chemical composition of BNPs as a the function of size, which remarkably illustrates the necessity of doing proper quantitative analysis to better characterize alloy nanoparticles.

Chemical composition and atomic arrangement play a fundamental role in the understading of the NP physics and chemistry. The characterization of these factors leads to the study of how they can modify the electronic properties of metallic nanoparticles, which is directly related to its physico-chemical properties.

An exciting example for probing the electronic structure of these nanoscale systems is the Localized Surface Plasmons Resonance (LSPR) [42, 43]. LSPRs are collective oscillations of free electrons in metallic nanoparticles, which occur due to the coupling of the electrons with the electric field of an incident electromagnetic radiation. As illustrated in figure 2.6, the electron charge density in resonance with the electric field forms an electric dipole in the surface of the NP. The coulomb interaction between positive nucleai and negative charges induces a restoring force oposit to the eletric field force. The electrons oscillate and collide with other electrons, ions or defects [42]. For dimensions less than the wavelength the plasmon is localized, as opposed to propagating surface

plasmons polaritons [44].

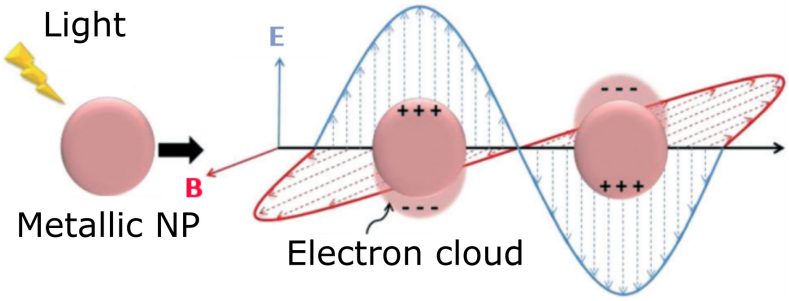


Figure 2.6: Schmeatic illustration of LSPR on metallic NPs. Adapted from [45].

Thus, different atomic species may lead to different absorption spectra. For example Ag, Au, and Cu nanoparticles with a size of 20 nm and spherical geometry, show absorption bands approximately 400 nm, 530 nm, and 580 nm, respectively, see figure 2.7. The spectral response of such nanoparticles is determined by the electronic structure which itself depends on the atomic and chemical structure. The LSPR of metallic nanoparticles leads to absorption in the UV-Visible spectral range which depends not only on size and shape of the particle but also in its electronic strcture. Although LSPRs are intrinsic to the NP properties such as size, composition and shape, they also depend of the environment [46]. Due to dielectric and quantum effects, the environment influences the plasmon peak position. Such systems require rigorous characterization and the support of theory and simulation to fully interpret and understand their physics.

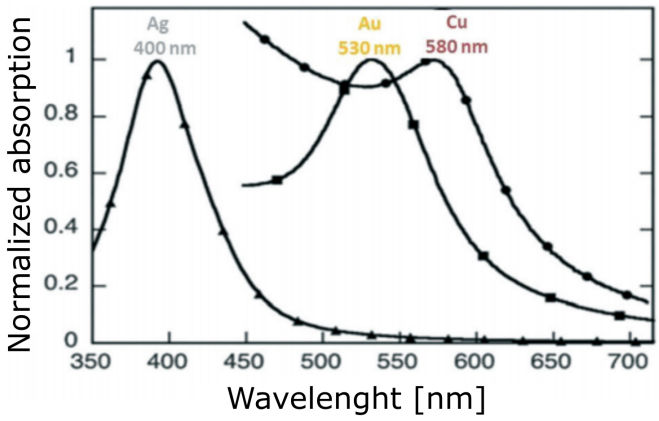


Figure 2.7: LSPR absorpion bands for different elements composing the NPs [47].

The absorption of light at different wavelengths for different elements shows that the properties of these nanosystems are sensitive to chemical composition. Thus, by mixing elements in the same nanoparticle, changes in the electronic strcture results in different physical and chemical propertie, as exemplified in the catalytic activity of metal alloy nanoparticles of 5 nm of RhAu and RhAg exemplified in figure 2.8.

Here, we discuss an example of catalitic activity in a chemical reaction. Hydrogenation occurs when a molecule is obtained by adding hydrogen to an unsaturated carbon

chain with double or triple bonds between carbons. In this scenario, RhAg and RhAu nanoparticles, depending on their chemical composition and when compared to nanoparticles composed of Rh alone, exhibit a higher conversion of molecules per active site per second (Turnover frequency), i.e., a higher catalytic activity in the hydrogenation reaction. The example strongly illustrates the necessity of precise chemical composition assessment for understading nanomaterials properties and its applications in the industry.

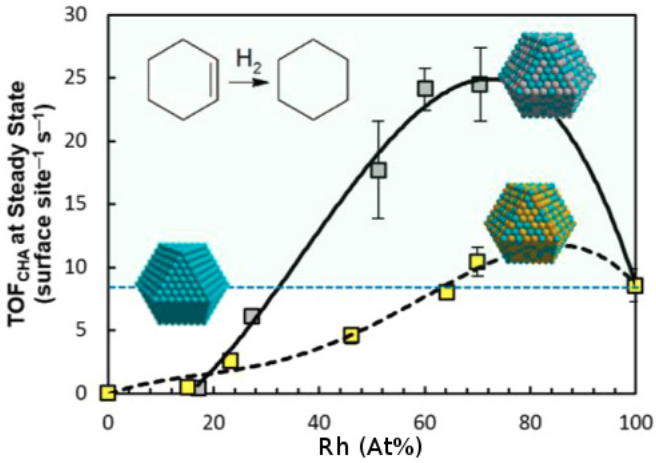


Figure 2.8: Catalytical activity of NPs in hydrogenation of cyclohexane under 25^{o}C as a function of atomic fraction of Rh for RhAg (grey) and RhAu (yellow) in contrast with pure Rh (dashed blue).Adapted from [48].

In order to experimentaly determine the chemical composition in very small nanoparticles ($\sim$ 6 nm), if possible quantitatively and with high spatial resolution ($\sim$ 0.5 nm), spectroscopic techniques based on electron microscopy can be used. Therefore, in this work, we wish to contribute to developing strategies to better quantify the atomic content and elemental distribution of bimetallic nanoparticles using spectral imaging. Thus, a model sample of bimetallic NPs with size and composition control is required. In this thesis BNPs produced by gas-aggregation were used [34].

2.2 Nanoparticle Synthesis

In order to contribute to the field of quantitative chemical analysis of small BNPs (Diam. $<$ 10 nm) a model sample with size and chemical composition controlled is required. In the metal cluster source used in this work, nanoparticles are synthesized by sputtering and subsequent aggregation of metals. This source of nanoparticles is inspired by the instruments developed by Haberland [49] and Hillenkamp [50]. Here atoms are sputtered from a metal target in a magnetron discharge in an inert He/Ar atmosphere. The metal atoms aggregate into clusters and nanoparticles before being transfered into vaccuum, forming a molecular beam. The magnetron discharge is of cylindrical symmetry [51] with twisted metal wires as target. The chemical composition of the atomic cloud, and thus of the BNPs is determined by the proper choice of wires.

The nanoparticle source, schematically illustrated in figure 2.9, consists of a magnetron sputtering head, an aggregation chamber, electrostatic lenses for ion beam control, a time

of flight mass spectrometer (TOF-MS) for size control and a deposition chamber for final sample preparation.

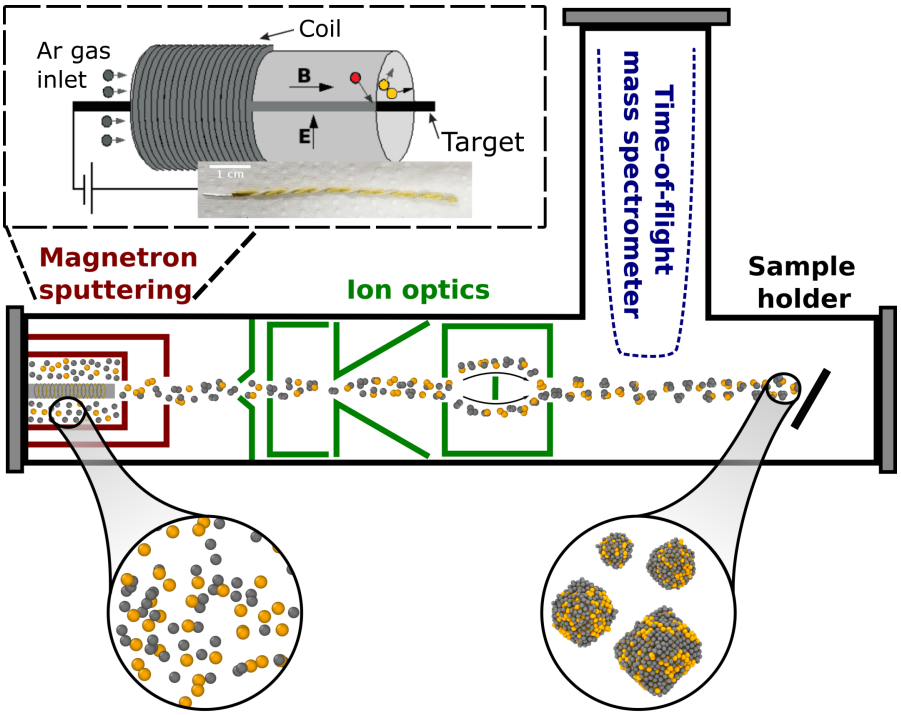


Figure 2.9: Schematics of the atomic cluster source used for nanoparticle synthesis using the gas-aggregation method. The inset shows the cyllindrical magnetron head [52] and the photograph of Au and Ag wires forming the target. The Ar gas is confined around the metal target by a magnetic field. Through collisions, the Ar ions remove metal atoms from the target [34].

In the figure above, details of the atom source are emphasized in the inset. The geometry of the sputtering setup illustrates the axial wires (Au and Ag in this case). The atomic cloud generated by the source of atoms is carried by a flow of gas (He and/or Ar), which, through three body collisions, promotes the aggregation of atoms into dimers. The dimers act as seeds for the nucleation of precursor clusters for the formation of nanoparticles [34]. The formation of these dimers only occurs due to the introduction of the inert gas flow. Their formation requires a third collision particle, He or Ar atoms, to remove the excess energy, allowing the atoms to form a stable dimer.

Starting from the dimers, further nucleation of atoms can occur, producing stable clusters. With the formation of these clusters can further aggregate in larger structures. Thus, the NPs grow into meta-stable structures, as illustrated in the figure 2.10, where the red spheres represent the atoms used for the formation of the atomic clusters and the green ones the residual carrier gas.

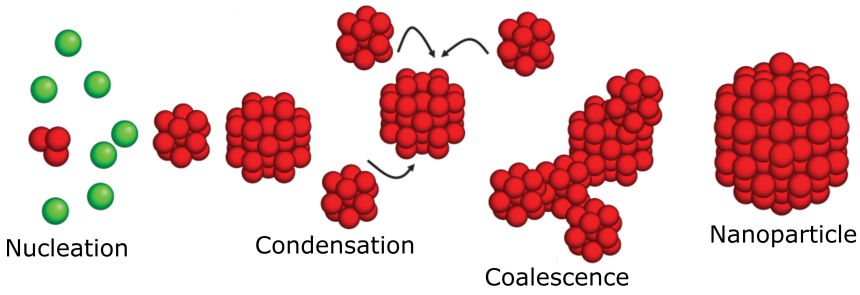


Figure 2.10: Schematic of NPs growing by gas aggregation. Adapted from [34].

For the condensation of the atoms to occur efficiently, implying aggregation without impurities, it is necessary to cool down the chamber.

Supersaturation induces condensation into clusters, following the minimization os Gibbs free energy. Even though the theory of formation of atomic clusters is still not fully described in theory, classical nucleation theory can provide a simplified analytical description [34].

Assuming two cases with gas with $N = n_1 + n_2$ atoms, where in the first case, the atoms are all in the vapor phase (equation (2.1)) and in a second case with n_1 atoms in the vapor phase and n_2 atoms forming a liquid droplet (equation (2.2)), it is possible to write the Gibbs free energy of the microcanonical ensemble for the two situations and evaluate its variation,

$$G_v = (n_1 + n_2)\mu_v(p, T) \tag{2.1}$$

$$G_g = n_1\mu_v(p, T) + n_2\mu_l(p, T) + 4\pi R^2\sigma \tag{2.2}$$

where $\mu_v(p, T)$ and $\mu_l(p, T)$ are the chemical potentials of the vapor phase and liquid phase, respectively, and $4\pi R^2\sigma$ is the surface energy. Thus, evaluating the Gibbs free energy change from the vapor phase to droplet formation, one can write:

$$\Delta G = G_g - G_v = 4\pi R^2\sigma + n_2(\mu_l - \mu_v) \tag{2.3}$$

Given that $\mu_v - \mu_l = k_B T ln(\phi_k)$ where $\phi_k = \frac{P_k}{P_v}$ with P_k and P_v being the supersaturation pressure and vapor pressure respectively, it follows that

$$\Delta G = 4\pi R^2 - n_2 k_B T ln(\phi_k) \tag{2.4}$$

Since n_2 can be written in terms of the volume of the droplet and the volume of the atoms V_a that make up the droplet, it follows that the change in Gibbs free energy is given by:

$$\Delta G = 4\pi R^2 - \frac{4\pi R^3}{3V_a} k_B T ln(\phi_k) \tag{2.5}$$

With the expression (2.5), it is possible to evaluate the variation of the Gibbs free energy for two different regimes, where $\phi_k < 1$ and $\phi_k > 1$, as follows in figure 2.11.

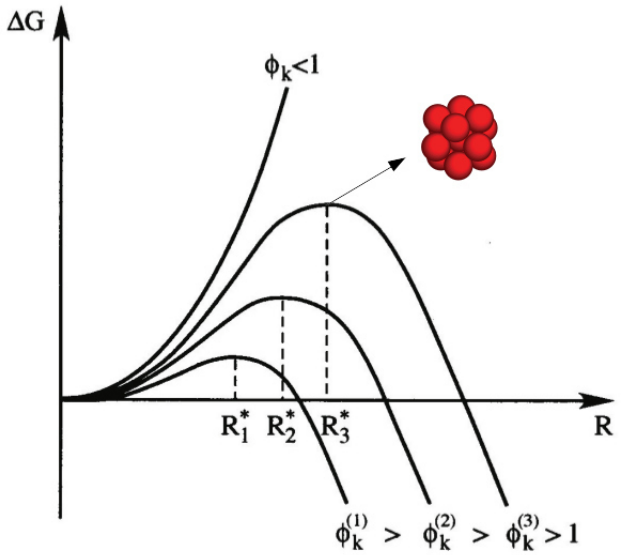


Figure 2.11: Variation of Gibbs free energy for the formation of atomic aggregates of radius R for different supersaturation ϕ_k. Extracted and adapted from [52].

In the regime $\phi_k < 1$, there is an increase in Gibbs free energy for all values of R, thus, aggregation is not stable in this regime. In the regime where $\phi_k > 1$, the Gibbs free energy presents a turning point for a critical radius R^*, i.e., after a certain size, the nanoparticle formation is stable and Gibbs free energy is further minimized.

Once the aggregation process is started, the longer the pulverized material remains in the aggregation region, the larger the cluster will become due to collisions with atoms and other clusters. This makes possible the control of the NP size by the length of and pressure in the aggregation chamber. These parameters, combined with the ion optics, define the size distribution of the particles produced.

We use twisted metal wires as a target for producing alloy nanoparticles. Hence, the atomic cloud used as a precursor for the nanoparticle growth consists of the elements from the wires. The inset in the source schematics 2.9 shows the geometry of the cylindrical magnetron head and a picture of twisted Au and Ag wires used as a typical target. Thereby, a model bimetallic sample with controlled chemical composition for our TEM studies can be designed. For this purpose, we chose AuAg alloys given their low reactivity with the environment and elevated atomic number, being the second vital characteristic of the system for most TEM-based techniques [7].

2.3 Model Sample: Bimetallic AuAg Nanoparticles

The optical properties of Au and Ag nanoparticles, being dependent on several physical and chemical characteristics, make these systems good options for applications in diverse fields such as sensing, imaging, biomedicine, and nanophotonics [53, 54, 55, 56]. Au is often more suited for biomedical applications because of its high biocompatibility and chemical stability [57]. On the other hand, Ag is preferred for optoelectronics, photovoltaics, and sensing [42, 58, 59, 60], because of its high plasmonic response. To use them for fundamental studies and possible applications, the development of nanotools for

the precise control of size, shape, crystalline structure, and of the chemical composition becomes essential. In this thesis, we use AuAg as a model system for quantitative analysis of the chemical composition using TEM (Transmission Electron Microscopy). Au and Ag their elevated atomic numbers generate high signals in most TEM-based imaging techniques [7]. More details will be given in Chapter 3. But first, it is important to explain in more detail the AuAg sample preparation and characterization.

The NP kinetic energy is controlled to achieve "soft landing" on the TEM grid (for a ∼4 nm in diam. NP, this corresponds to 0.05 eV/atom). The NP size distribution in the molecular beam can be measured in situ by time-of-flight mass spectrometry (TOF-MS). TEM images indicate that the NP size distribution follows a log-normal function (mean diameter ∼4 nm, with 3 nm in width, consistent with the in-situ TOF-MS characterization; see 2.12).

Au_xAg_{1-x} NPs samples with three different compositions have been used in the present study whose Au content in atomic fraction are obtained by EDS-TEM from an ensemble of NPs using a large open electron beam (several microns in diameter). More details about the chemical composition quantification will be given in Chapter 4.

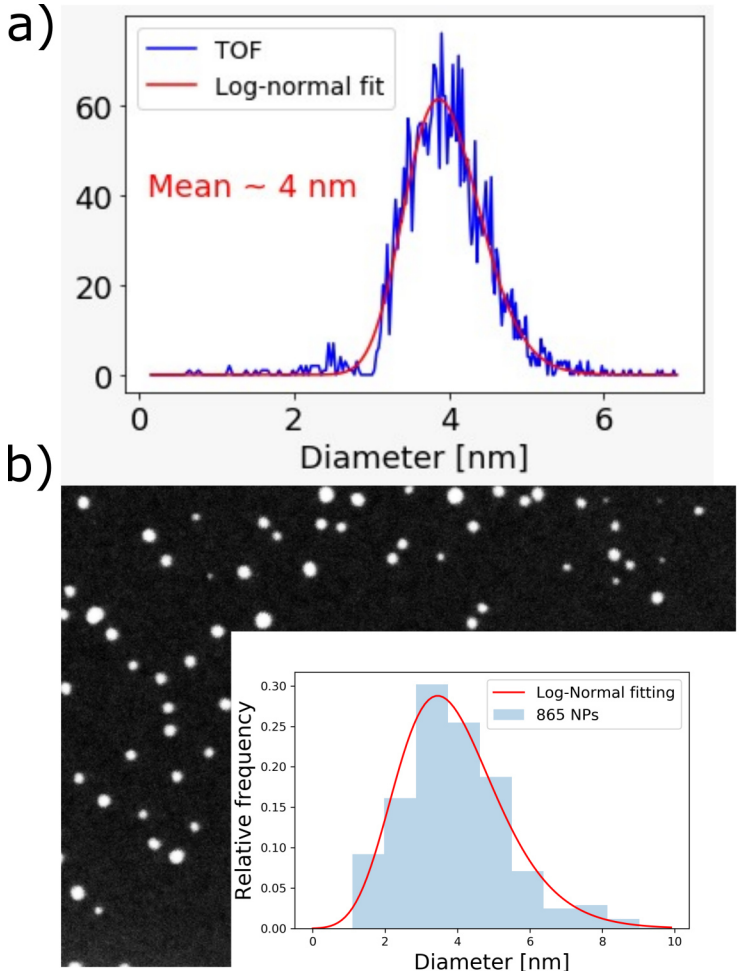


Figure 2.12: a) AuAg BNP size distribution measured in-situ with TOF-MS and b) size distribution histogram obtained from STEM-ADF images. The two size distributions are from two diferent samples but with the same mean diameter and prepared at similar conditions.

It is of utmost importance to keep in mind that the complete characterization of nanometric systems requires detailed knowledge of the atomic arrangement, of how the different chemical elements of the materials are redistributed, and finally, of how the size and chemical composition affects these factors. Motivated by the importance of exploring quantitative analysis to answer these questions, in collaboration with another doctorate student from the group, MSc. L. M. Correa, we used Precession Electron Diffraction. This technique allows us to exploit the full potential of the Pair Distribution Function for quantitative analysis [61].

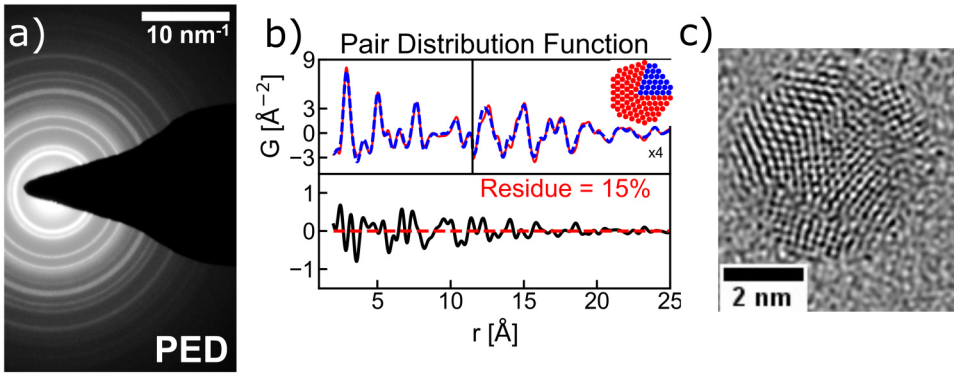


Figure 2.13: a) Precession electron diffraction (PED) pattern; b) Pair Distribution Function (PDF) calculated from PED pattern; and c) Representative High Resolution TEM (HRTEM) image of AuAg NPs. Adapted from [61].

With this method, we measure the average crystalline structure of our AuAg NPs. This approach makes use of the comparison between experiment and simulation of the PDF (Pair Distribution Function) derived from the electron diffracted patterns to check the differences between both, see figure 2.13. Using residual analysis in the PDF signal, we can test several different possible structures for the NPs like icosahedra, FCC, decahedra, etc. The refined model obtained for the samples produced in our setup indicates a mixed alloy system composed of 5-fold symmetry polycrystals of $\sim$ 3-4 nm composed of crystallites of $\sim$ 1.8 nm, forming a decahedra-like structure. Individual NP measures obtained from High-Resolution TEM indicate very polycrystalline NPs indeed. Therefore, the information derived from the PDF is consistent with the observed NPs in the sample. However, the information limit in this analysis does not allow us to extract an elemental distribution in the AuAg BNPs; different approaches such as EDS chemical analysis are necessary. Nonetheless, this example is interesting to show the quality of quantitative and statistically reliable analyses to answer a scientific question at the nanoscale. Likewise, spatially resolved spectroscopy for quantitative chemical analysis will be extensively studied and applied in individual AuAg NPs smaller than 10 nm, which is our model system to work in TEM studies.

2.4 Summary

This chapter introduced the multimetallic nanoalloys, most specifically the AuAg BNP system, which is the project's study sample.

BNPs systems are contextualized in the panorama of nanoscience and nanotechnology. Through the use of examples in plasmonics and catalysis, the contribution of the precise knowledge of size, morphology, structure, and chemical composition of BNP's electronic properties was explained.

The benchmark AuAg BNP synthesis and their characterization concerning size, structure, and composition tunning was described.

Finally, the need for spectroscopies at the nanoscale was proposed for the aim of the complete characterization of single BNP's chemical composition.

Chapter 3

Scanning Transmission Electron Microscopy

The study of individual nanosystems requires instruments that have a high spatial resolution. Scanning Transmission Electron Microscopy (STEM) and related techniques are specially adapted as they allow obtaining images of morphology, structural, and even nanometer resolution spectroscopic information [7, 6]. In the last 20 years, electron microscopy has received surprising improvements in its optical system with the optimization and commercialization of electronic lenses with spherical or chromatic aberration correctors and electron energy monochromators. These advances gives us the opportunity to obtain fine information with precision, accuracy and reproducibility in materials science. In this way, quantitative and statiscally representative analysis in electron microscopy becomes feasible and essential.

Concerning the measurement of chemical composition, as alloy nanoparticles (NPs) get smaller (in the ∼5 nm size regime), it is difficult to get reliable and reproducible information. In fact, sample stability under the electron beam becomes a critical issue, and chemical composition can gradually change due to the sputtering of lower atomic number atoms from the nanoparticle surface [62, 63, 64, 64, 65]. However, strategies to mitigate the radiation damage can be employed, and quantitative chemical composition characterization can be performed by Energy Dispersive X-ray Spectroscopy (EDS). Hence, the chemical composition dispersion can be assessed as a function of size, measuring individual NPs generating elemental maps by EDS-STEM.

This chapter aims at explaining relevant technical and scientific details related to STEM imaging and spectroscopy, particularly EDS, given the necessity of extracting localized elemental information from nanometric systems. In the AuAg nanoparticles explained in the last chapter, we saw the necessity to probe spatialy resolved chemical composition. Thus, since STEM is a powerful tool for analytical microscopy, we will broadly use this technique in this thesis.

3.1 Imaging in STEM

In STEM mode imaging, a convergent electron beam forming a probe, scans step-by-step (pixel-by-pixel) the specimen to be analyzed, like is ilustrated in figure 3.1 a).

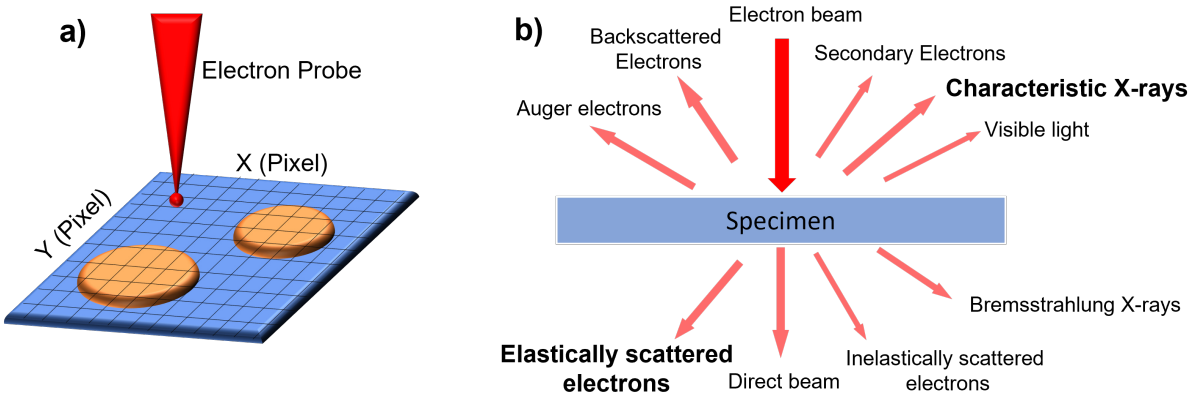


Figure 3.1: Scheme of a) electron probe scanning an area of the sample with object to be imaged and b) signals obtained due the electron probe interaction with the specimen.

The electron probe interacts with the specimen, generating several different types of signals (see 3.1 b)) like elastically scattered electrons (Dark-Field images) and characteristic X-rays (Elemental mapping). These signals are the two most important which will be used in the data acquisition of this thesis. Detecting those signals, pixel-by-pixel ,simultaneously with the probe scanning, makes it possible to create images with nanometric or even sub-Ångstrom lateral resolution [6].

In STEM, using the inelastically scattered and direct electrons, we generate a Bright-Field (BF) image, where the specimen's electronic cloud scatters the electrons. On the other hand, the Dark-Field (DF) images are generated by detecting the electrons scattered by the atomic nucleai of the specimen.

In DF image generation, the nucleus-spread beam follows approximately the Rutherford model. The angle at which the electrons are transmitted, therefore, depends of the atomic number ($I \propto Z^2$). The higher the atomic number, the greater the electron intensity [7, 6]. Hence, using an annular detector, as illustrated in figure 3.2, to detect the elastically scattered electrons, an annular dark field (ADF) image can be formed. The background is given by a low-intensity signal (fewer electrons coming to the detector). The sample region will have a more intense signal than the background, varying the contrast as a function of atomic number. In the figure 3.2, is shown a schematic of the formed electron probe being incident onto the specimen. The β_1 and β_2 angles represent the fraction of the transmitted beam used to detect the high angle elastically scattered electrons. When $\beta_1 > 50$ mrad and $\beta_1 < 200$ mrad, the imaging method is known as High Angle Annular Dark-Field (HAADF). In this condition, the image is formed in majority by electron thermal diffuse scattering (TDS) with the crystal lattice. this implies no contribution of diffraction effects, the image intensity depends only of the atomic number, the number of atoms and the mean square positions of the atoms vibrating in the lattice [66]. For simplicity, the image intensity is only dependent on the elements in the atomic column.

Due to its relatively simple image interpretation, this is the most popular imaging method in EM nowadays, the Z-Contrast imaging.

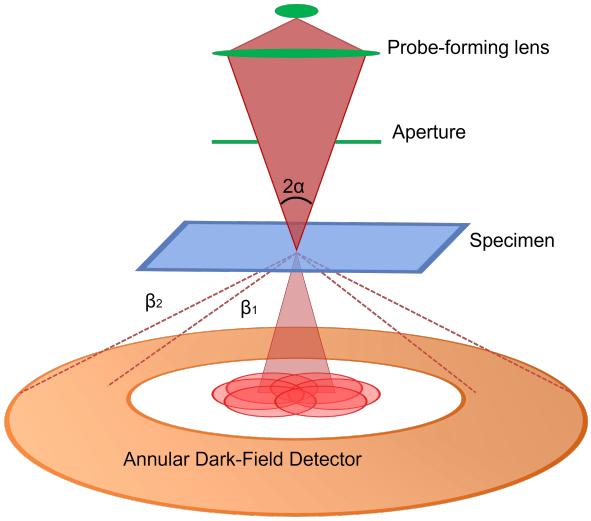


Figure 3.2: Schematic of STEM mode imaging. Condenser lenses form the convergent electron beam (probe) which is scattered in the sample in direction of an ADF detector.

Looking at the images in figure 3.3 it is possible to see that BF and DF images are complementary in intensity. In the BF image, there is bright field background and darker tones where the NPs are located. For DF images, the background is dark, and the NPs are bright. This is due to the nature of the elastic scattering and due to the angle of detection [7]. Both image methods can be used as reference for the spectroscopy data acquisition, but in this work we will always use the DF images for standardization and because this mode will be used for further high resolution images.

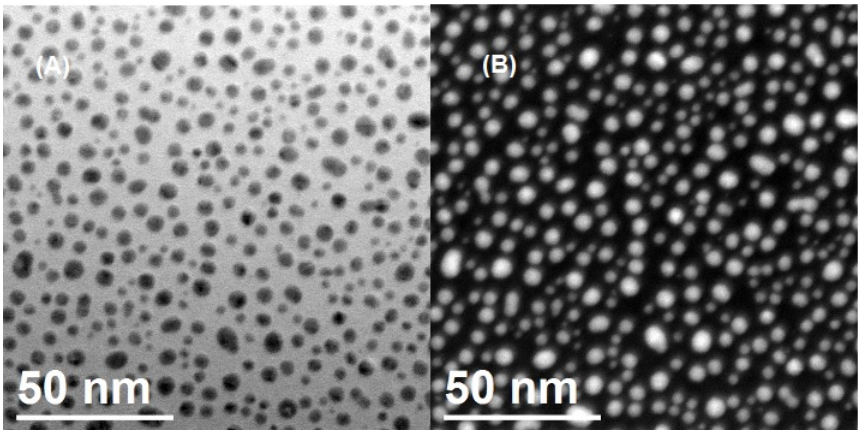


Figure 3.3: STEM images of Au nanoparticles obtained in JEM 2100F from LNNano. a) Bright-Field image and b) Dark-Field image.

In STEM-ADF/HAADF the imaging is spatially incoherent, more intensity simply means more atoms and higher atomic numbers. In this way we can express the image intensity as a convolution of the object wavefunction $\psi(\mathbf{r})$ and the probe $P(\mathbf{r})$ (point spread function), see equation (3.1).

$$I_{\mathrm{HAADF}}(\mathbf{r}) = |\psi(\mathbf{r})|^2 \otimes |P(\mathbf{r})|^2 \tag{3.1}$$

In this condition, the images are generated with less signal compared with typical ADF images; however, it is a condition where the image shows chemical contrast (Z-Contrast) with high spatial resolution since the probe size purely defines the resolution. This means that the image is not phase-dependent, and the brighter regions in the image are due to heavier elements [7]. In this way, it is possible to differentiate elements, qualitatively and directly, without the need of simulations or further complex interpretations, as can be seen in figure 3.4. In this example, the probe scans a GaAs thin sample, and the annular detector collects signals from the atomic columns. In the illustration, Ga is yellow and As red, and the dumbbell model can be verified in the image profile intensity due to the difference in Z (31 and 33).

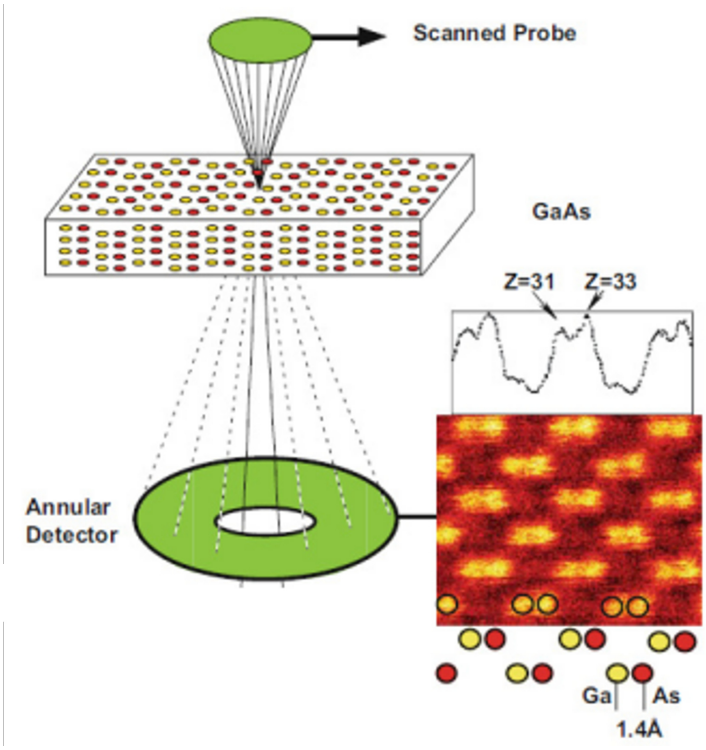


Figure 3.4: Schematic of the generation of a STEM-HAADF image of GaAs. The profile over the image ilustrates the sensibility to the atomic number. Yellow spheres are Ga and the red ones are As. Image extracted from [7].

In the scenario of AuAg BNP studies, an exciting application of Z-contrast is the analysis of Lasserus et al. [67], where high-resolution HAADF images are used to investigate the alloying process of core-shell NPs under the influence of temperature. They observe that the elemental diffusion in sub-5 nm NPs occurs for lower temperatures than in bulk materials. Analyzing the HAADF image intensities, they are also capable of determining the diffusion constant as a function of temperature. Despite the high quality of this type of analysis, even more in such complexes systems such as sub-5 nm NPs, quantitative measurements are required for fine and more robust analysis, such as the

investigation of atomically resolved chemical gradients in chemically disordered mixed BNPs.

Quantitative analysis nowadays is a reality due to the advances in EM instruments, and data treatment [40, 41]. Recent works show that proper calibration of the detector gain [40, 41] and the use of multislice simulations, considering frozen phonons and the spatial incoherence of the source in the model, permits to overcome the known Stobbs factor limitation (large constrast mismatch between experimental images and theoretical simulations) [68]. This enables the extraction of how many atoms are in the atomic columns with relatively good accuracy and precision through the comparison between simulated models and experimental images. In the bimetallic NPs and catalysts field, it is of interest not only to quantify the number of atoms but also the chemical composition of each column. Efforts are being made in this direction as well, whether on 2D atom counting [69] or 3D analysis (tomography) [70, 71, 72]. However, it is crucial to mention that a careful analysis is required having in view the radiation damage during the execution of these experiments, which can strongly limit the signal-to-noise-ratio (SNR) to detect atomic species. In this way, efforts are also being made on low dose experiments by High-Resolution TEM (HRTEM), where by means of focal series reconstruction [7] and Weak Phase Object Approximation (WPOA) on multislice simulation, the atomic column can be retrieved analyzing the reconstructed phase. Nonetheless, recent works show that the detection limit is only in the order of a few atoms [73], limiting the analysis to 2D materials.

It is essential to mention that all the imaging methods discussed above are not element-specific, and what elements are being observed can only be known due to previous knowledge of the sample's composition and geometry. Other spectroscopic techniques in microscopy, such as Energy Dispersive X-ray Spectroscopy (EDS) and Electron Energy Loss Spectroscopy (EELS), are used to assess the chemical signature of the sample directly, without the the need of previous knowledge and simulations.

To perform chemical analysis at the nanometric scale, maybe the most popular analytical electron microscopy (AEM) technique used is EDS-STEM because of the direct interpretation of chemical composition results. Despite the very known lack of statistics of the method for small systems, recent advances in the development of detectors with higher solid angles of detection (> 1 sR) [74, 75] makes possible the use of EDS-STEM to analyze with accuracy and precision the chemical composition of nanosystems faster (low acquisition times).

Therefore, to start our studies in quantitative analysis of chemical composition in binary nanoalloys, EDS-STEM is a logical AEM method choice, even more because of its simplicity. More details about this technique will be given in the next section.

3.2 Analytical Electron Microscopy: Energy Dispersive X-ray Spectroscopy

Analytical Electron Microscopy (AEM) refers to the spectroscopic techniques in EM that collect chemical composition information, qualitatively or quantitatively. In TEM, the two most prevalent techniques are Electron Energy Loss Spectroscopy (EELS) and Energy Dispersive X-ray Spectroscopy (EDS). EELS is typically suitable for analysis in

the range of 0 - 1000 eV and light elements (low atomic number Z). In contrast, EDS is suitable for heavy elements, and the analysis in the energy range higher than 1000 eV. Here in this thesis, we aim to contribute to the improvement of quantitative EDS chemical analysis of nanoalloys. Thus, AuAg NPs arise as a good choice for the model system because of their low reactivity with the environment and the high atomic number of both elements ($Z_{Au} = 96$ and $Z_{Ag} = 47$). With EDS, we can access the chemical signature of elements due to electronic transitions and subsequently characteristic X-ray emission as a response to radiation incidence [7].

3.2.1 Characteristic X-ray generation and detection

X-rays are electromagnetic waves, photons. They have higher energy than visible light, and roughly their energy is between 10 eV to 100 keV. X-rays can be generated in different manners, and most of them typically consist of the bombardment of a metallic target by electrons. The electron-nucleus interaction generates continuum or Bremsstrahlung X-rays, and electron-electron interaction generates characteristic X-rays. Henceforth, we will focus on the characteristic X-rays, where the chemical composition information is contained. Assuming the Bohr atomic model, the electrons are in energy-bound states, like orbitals or shells around the nucleus. When an electron with higher energy than the binding energy removes an electron from an inner shell, an electron from an outer shell can transit to the level of lower energy. A photon is released to conserve energy, with the same energy as the difference between the two states. To illustrate this effect, we can see a schematic in figure 3.5, where photons are emitted as a consequence of the electronic transitions. For different species of atoms, the energy state differences may change, and therefore, the value of the X-ray energy carries the elemental signature of the specimen.

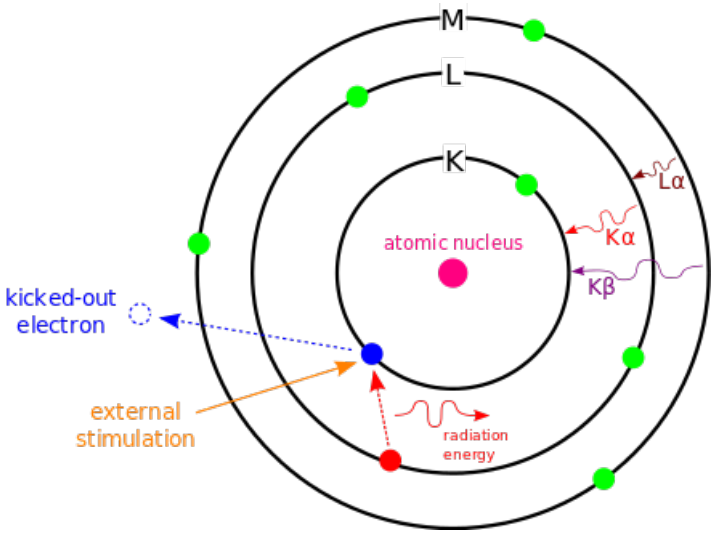


Figure 3.5: Illustration of the physical phenomenom of photon emission by electronic transition after ionization due external excitation. Iage extracted from the site http://www.wikiwand.com/en/Energy-dispersive_X-ray_spectroscopy 23/06/2022

Since the bound state energy of an electron is negative, the inner-shell electrons are more negative or lower in energy than the outer-shell electrons. The lowest shell is named

K, the second one is L, and so on. When an electron from the K shell is removed, an electron from an highest shell can transit to this vacancy emitting an X-ray photon. When the electronic transition occurs from the L shell to the K shell, we have a emission line described as K_α, from the M shell to the K shell would be K_β and so on for the L and M lines. These emission lines are characteristic for each element, and therefore, their emitted photons are characteristic, too.

After the atom ionization, only a fraction of them emits photons as a consequence of the electronic transitions. In C atoms, for example, only 0.26% of the recombination processes lead to X-ray photon emission [76]. In this case, the vast majority of processes involves electron emission, known as Auger electrons. The fraction of ionization processes that produces photons is the fluorescence yield ω. With increasing atomic number Z, the fluorescence yield increases, and therefore the Auger electron emission reduces, as can be seen in figure 3.6. In this way, to planEDS-related experiments, high Z elements are preferable to get more signal. Whence, our choice of AuAg as a model system is justified due to its high fluorescence yield of each element.

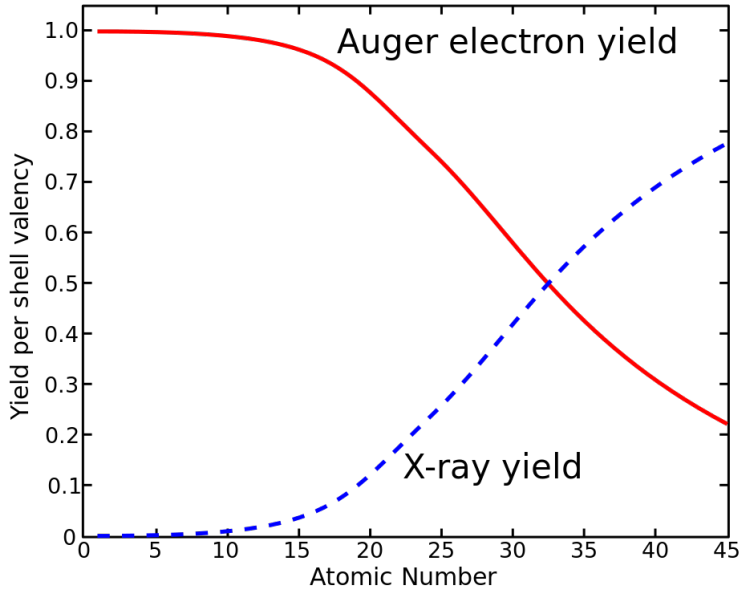


Figure 3.6: Emission profiles due to de-excitation of ionized atoms as a function of atomic number. Auger electrons in red and characteristic X-rays in blue. Extracted from https://en.wikipedia.org/wiki/File:Auger_Yield.svg, in 23/06/2022.

Counting (increasing the signal) is crucial for any analytical analysis; here is no different. The emitted X-rays must be counted/detected, and the detection in AEM is done using a Si or Ge-based detector, typically Si. In a conventional detector, the incident photons enter the material with moderate energy (few kV), exciting photoelectrons which generate electron-hole pairs as they go through the crystal. The electron-hole pairs are attracted and converted into electric charge using an anode and cathode to generate an electric field. Hence, the initial photons, in the end are transformed into an electric signal which is amplified by a FET (Field-Emission Transistor) circuit. The amplified electric signal is directly proportional to the number of photons detected, and

thus, photon counting becomes possible by proper signal processing. Nowadays, with the advancement of detection systems, the solid collection angle of photons is severely increased, being possible to collect >1 sR (compared with 0.1-0.3 sR for typical Si (Li) detector). This enormous advancement in detection systems is due to different factors (windowless detectors, for example), but one of the most important is the Silicon Drift Detector (SDD) technology. In the place of a simple anode-cathode detector generating a uniform electric field, a series of concentric metal electrodes generate an electric field steering the charges towards a small collector electrode in the center. This means that the capacitance in the collector is much lower, and therefore the noise is lower as well. In this way, the pre-amplifiers parameters can be optimized to get more count rates. SDD also are produced with more detection area and does not need to be cooled down by liquid nitrogen to reduce thermal noise. The use of thermoelectric systems cools down the equipment. The straightforward way of interpreting the detection system while measuring is that the analogic signal is converted to a digital one read by a computer which gives us the photon energy. The value read by the computer is incremented in the memory location (channel) corresponding to that energy. The channel widths are typically between 5 - 20 eV, and the multichannel detected signal is composed of 1024 - 4096 channels. Then the signal can be expressed as a one-dimensional histogram of countings for a specific energy position, forming our spectrum, as can be seen in figure 3.7. Where several elements are identified, such as Si from the detector, Cu from the TEM grid, Au and Ag from the AuAg sample analyzed.

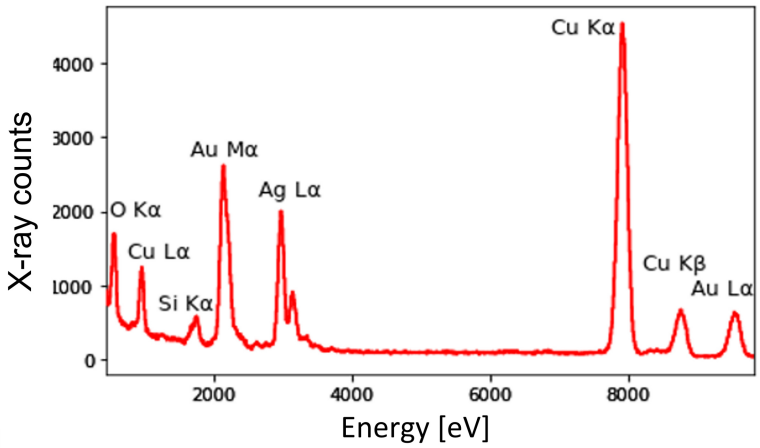


Figure 3.7: Representative EDS spectrum of AuAg NPs deposited on ultrathin carbon on a Cu grid.

The energy resolution of the EDS spectrum is typically 130 eV, much worse compared with other analytical techniques (EELS, for example). This resolution is mainly due to electronic noise related to the capacitance of the gate of FET amplifiers. Roughly increasing the processing time of the amplifier circuit decreases the width of the peak (improves the resolution). So it is possible to optimize this parameter to improve the counting rate to the detriment of energy resolution.

Counting independent events (photon emission) in time has the intrinsic property of following a Poisson distribution. Ergo, a significant characteristic to consider in EDS

chemical analysis and other EM-based techniques is the Poisson noise which is the noise related to the act of counting these events. In Poisson statistics, the variance ν is equal to the number of events (counts) I, thus, the standard deviation of the distribution is $\sigma = \sqrt{I}$. In most analytical analyses, whether qualitative or quantitative, Poisson noise is the most limiting factor in the detection limit [7, 9].

3.2.2 Elemental mapping: Hyperspectral imaging

Recently TEM has made huge progress considering electron optics, detector efficiency, automation, reproducibility, etc. [7, 77, 78]. Scanning transmission electron microscopes (STEMs) are capable of recording the so-called hyperspectral image (HSI, [6]), where an entire analytical spectrum or diffraction pattern can be registered at each image pixel. The HSI approach is currently used to generate huge 3D or 4D datasets, i.e, in X-ray energy-dispersive spectroscopy (EDS), electron energy loss spectroscopy (EELS), and cathodoluminescence (CL) or electron diffraction (ED). EDS-STEM is a technique with the power to obtain chemical composition information spatially localized in the nanometric scale. The method allocates an EDS spectrum (1D signal) to each pixel of a scanned image (2D signal), forming the HSI. Let's say we have an image with 100x100 pixels and 2048 channels, so we can say we have 100x100x2048 voxels where our X-ray counts are allocated in. A pictorial view of the dataset structure can be seen in figure 3.8.

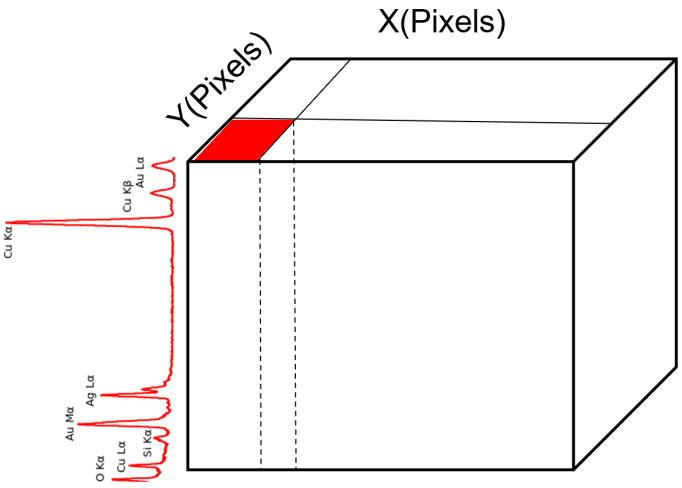


Figure 3.8: Schematic ilustration of the dataset architecture, a datacube. Each pixel contains a 1D signal (spectrum), forming a 3D dataset.

The data acquisition in EDS-STEM experiments can be very long, therefore, previous BF or DF images are used as reference for electron beam drift correction (relative to the sample). The drift correction is performed comparing fast acquired electron images using cross-correlation. This way, the elemental mapping can be generated with low spatial resolution loss. It is relevant to mention that newer TEMs show incredible stability and correction systems (drift less than 0.1 nm/min) which allows the chemical composition spatially resolved of such small systems.

For our first discussion about elemental mapping, it is essential to discuss the capacity of selecting the pixels and channels of interest in these HSI to generate images, or like in the case of EDS chemical analysis, the elemental maps. Considering, for example, the AuAg NPs sample, we can perform elemental mapping by selecting the channels corresponding to the peak of Ag Lα at 2.98 keV and Au Mα at 2.12 keV. The generated image filtered with these intensities will display spatially where are Ag and Au in the sample. This type of strategy immediately allows us to know the elemental distribution qualitatively in our BNPs (see figure 3.9). This way, we can learn that our particles seem alloyed and mixed with Ag and Au content distributed over every pixel, at least in a first view. We can also select all the intensities in the spectral range to generate our images, see figure 3.9. In this case, the image is noisier because all the background intensities contribute to each pixel's noise (Poisson noise = $\sqrt{I}$). The object to be analyzed must be well resolved and easily observable in a high-quality HSI, even with all the image spectra intensities integrated in each pixel, see figure 3.9 d). Typically standard acquisition and analysis software only show the beautiful filtered maps (with the elements of interest), which can be misleading for the operator or reader in judging the quality of the measurement. In our elemental mapping exposed above, although we can observe Au and Ag signals everywhere in the nanoparticles, we cannot know the amount of each element. Quantitative analysis is required to know wheter the atomic fraction in the whole BNP as like to know how the elements are distributed inside the BNP due to its possibility of having chemical gradients or radial enrichment of Ag towards the surface [25, 79, 80].

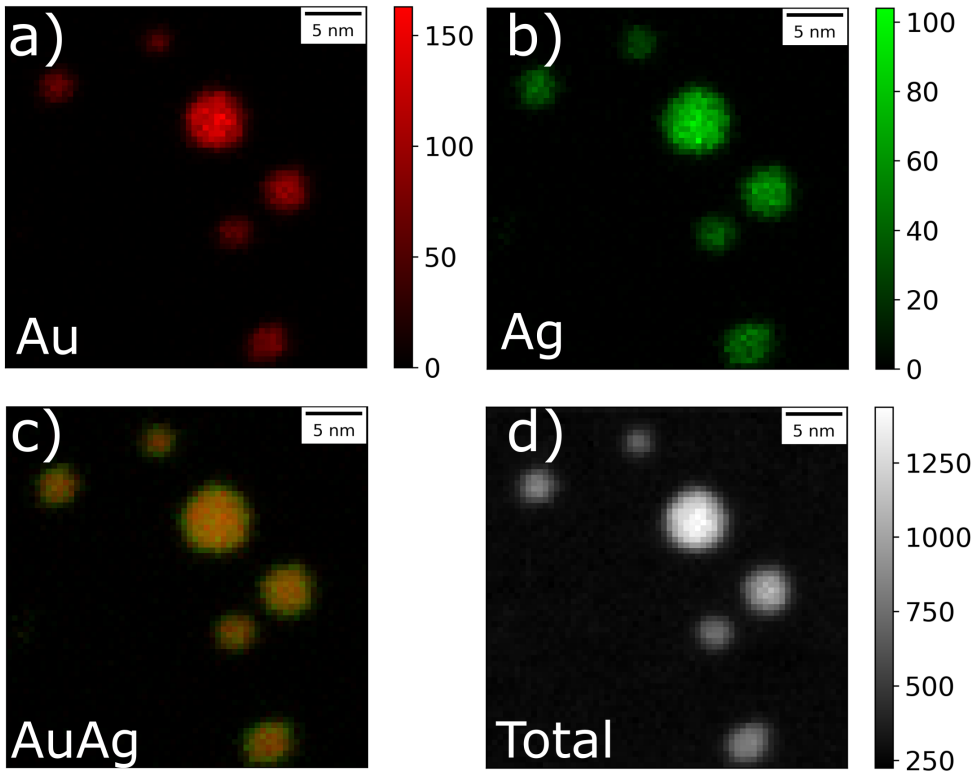


Figure 3.9: EDS-STEM elemental mapping results ofr AuAg NPS. a) Au, b) Ag and c) Au and Ag overlaied and d) all the intensisties of the HSI.

3.3 Summary

In this chapter, the basics of the HAADF-STEM and EDS-STEM were explained, which are the two experimental methods most used in this thesis. Here, the focus was mainly EDS-STEM elemental mapping because how the data sets are acquired and organized in hyperspectral images plays a fundamental role in the subsequent chapters. With EDS-STEM it was shown that localized chemical information can be obtained from small volumes such as bimetallic nanoparticles. However, the discussion was kept on a qualitative level but the advances of characterization nanomaterials on a quantitative approach requires the discussion of difficulties and possibilities.

Chapter 4

Quantitative EDS Chemical Analysis

Transmission Electron Microscopy results are very often qualitative and hardly accomplish the requirement of being statisticaly representative. In EDS-STEM, we can learn from our samples where a particular element is located in a specific position and also its amount. Proper chemical composition quantification requires, however, high SNR (signal-to-noise ratio) levels [9]. For example, in our experimental case, we have BNPs smaller than 10 nm, with only a few thousand atoms. Since the signal is proportional to the number of atoms, these systems are very challenging to measure quantitatively because of the fact that electron-induced ionization cross-section of an atom is very low ($\sim 10^2$ Barns). Additionaly, other factors such as fluorescence yield and electronic transition probability, influence the counting rate of X-ray photons. Together, these factors make the emission of photons difficult to detect for the low number of atoms.

Realizing that, we need to rationalize in terms of uncertainty intervals in order to be able to do a proper quantitative analysis. The difficult to count signal must be overcome to obtain significant confidence intervals,in order to measure the chemical composition distribution in the BNPs or the distribution inside them (chemical gradients). Due to Poisson noise, the one-sigma fractional counting uncertainty is estimated as $\sqrt{I}/I = 1/\sqrt{I}$, meaning that in order to reduce our uncertainty "...counts, counts, and more counts..." [7] are necessary. Therefore, small BNPs (Diam. $\sim$5 nm) are challenging for quantitative chemical analysis with high precision and accuracy, which is this thesis' main topic and contribution. Illustrating the problem, we can roughly assume 4000 atoms for a 5 nm NP. If we have a 5% atomic fraction of dispersion in the chemical composition of the whole population, this would mean that the difference between each NPs for 1σ is 200 atoms. Very good statistics is necessary to see this difference due to the Poisson nature of the counting problem. The seminal work of Currie [9, 10, 11, 7] shows that a quantitative analysis requires a significant increase of the signal level (approximately 15 times higher than the uncertainty interval) to attain the measurement of a quantity. In order to estimate the order of magnetude in our example above, thinking only in 1σ relative error criterium, we can roughly estimate the number of counts required to see this dispersion of 5% as $1/\sqrt{I} = 0.05$, which implies $I = 400$ counts per particle, which

for small systems is not readily achievable.

4.1 Cliff-Lorimer method

In the sequence, we will discuss more the quantitative elemental assessment of EDS-STEM measurements, explaining the path to extract chemical composition ratios from X-ray intensity ratios by the Cliff-Lorimer method [8]. It is important to remember that X-ray production is a two-step process. First the atom is ionized and then the excitation decays by photoemission The ionization probability of an atom is characterized by the ionization cross-section Q and the decay by the fluorescence yield ω. Also, in the X-ray intensity calculation, we must consider the probability of the transition occurring a of a specific emission line and the detector efficiency ϵ. Thus, for a certain element A, we can approximate the X-ray intensity detected as:

$$I_A = \frac{C_A Q_A \omega_A a_A \epsilon_A t}{M_A} \tag{4.1}$$

where M_A is the atomic mass of the element A, C_A its atomic fraction and t is the thickness of the material analyzed.

Now, considering the intensity ratio $\frac{I_A}{I_B}$ between two different A and B elements, we have:

$$\frac{I_A}{I_B} = \frac{C_A Q_A \omega_A a_A \epsilon_A M_B}{C_B Q_B \omega_B a_B \epsilon_B M_A} \tag{4.2}$$

$$\frac{C_A}{C_B} = \frac{(Q\omega a\epsilon)_B M_A}{Q\omega a\epsilon)_A M_B} \frac{I_A}{I_B} \tag{4.3}$$

$$\frac{C_A}{C_B} = K_{AB} \frac{I_A}{I_B} \tag{4.4}$$

where

$$K_{AB} = \frac{(Q\omega a\epsilon)_B M_A}{(Q\omega a\epsilon)_A M_B} \tag{4.5}$$

is the so-called Cliff-Lorimer constant, calibrated from a reference sample of the same elements to be analyzed but with a known chemical composition.

Now that we know how to convert X-ray intensities into a proportional chemical composition ratio, we must think about uncertainty intervals if we want to be appropriately quantitative. Using the propagation error formula, the uncertainty for the chemical composition ratio (R) can be expressed as a function of K_{AB}, I_A, and I_B uncertainties:

$$\left(\frac{\sigma_{\mathrm{R}}}{C_{\mathrm{A}}/C_{\mathrm{B}}}\right)^2 = \left(\frac{\sigma_{k_{\mathrm{AB}}}}{k_{\mathrm{AB}}}\right)^2 + \left(\frac{\sigma_{I_{\mathrm{A}}}}{I_{\mathrm{A}}}\right)^2 + \left(\frac{\sigma_{I_{\mathrm{B}}}}{I_{\mathrm{B}}}\right)^2 \tag{4.6}$$

Taking the peak background subtraction into account in the error calculation, we can rewrite the equation above as:

$$\left(\frac{\sigma_{\mathrm{R}}}{C_{\mathrm{A}}/C_{\mathrm{B}}}\right)^2 = \left(\frac{\sigma_{k_{\mathrm{AB}}}}{k_{\mathrm{AB}}}\right)^2 + \left(\frac{\sigma_{I_{\mathrm{A_b}}}}{I_{\mathrm{A_b}}}\right)^2 + \left(\frac{\sigma_{I_{\mathrm{B_b}}}}{I_{\mathrm{B_b}}}\right)^2 \tag{4.7}$$

where $\sigma_{I_{\mathrm{A_b}}}$ and $\sigma_{I_{\mathrm{B_b}}}$ are the uncertainities considering the background subtraction which can be calculated by error propagation as well, resulting in $\sigma_{I_{\mathrm{A_b}}} = \sqrt{\sigma_{I_{\mathrm{A}}}^2 + \sigma_{I_{\mathrm{b}}}^2}$ being I_b the background intensity.

Considering that sometimes it is preferable to express the chemical composition as an atomic fraction and not a ratio, we can use $C_A + C_B = 1$ and derive an expression for C_A in the function of the ratio. Consequently, an error propagation formula can be used, and both atomic fraction and uncertainty interval equations are:

$$C_B = \frac{1}{1 + C_A/C_B} \tag{4.8}$$

$$\sigma_{C_B} = \frac{\sigma_R}{(1 + C_A/C_B)^2} \tag{4.9}$$

This method can be used to quantify the chemical composition of various nanomaterials, not only nanoparticles, as we have been discussing in this thesis. Originally the method was proposed for thin films and can also be used in more complex systems made of three elements or more, such as semiconductor nanowires (NW) of InAsP [81]. In this work, the author verifies the content of III-V semiconductors in the NW and their diffusion into the Au NP growth catalyst for different synthesis parameters. In figure 4.1 we show a schematic of the heterostructure NW (Au, InP, InAs) and below a graph with the atomic fractions of In, As, P, and Au, determined with its respective error bars calculated by error propagation of the Cliff-Lorimer equation. In their analysis, doing a line scan in the NW, the authors verified that the NW is rich in In and As in its basis. Likewise, the other extreme of the scan shows more In and P and then a peak of As again in the NW and Au catalyst interface. The performance of EDS-STEM to analyze interfaces of complex systems like in this example quantitatively shows the proness of the Cliff-Lorimer method, despite its simplicity, to chemically characterize nanosystems.

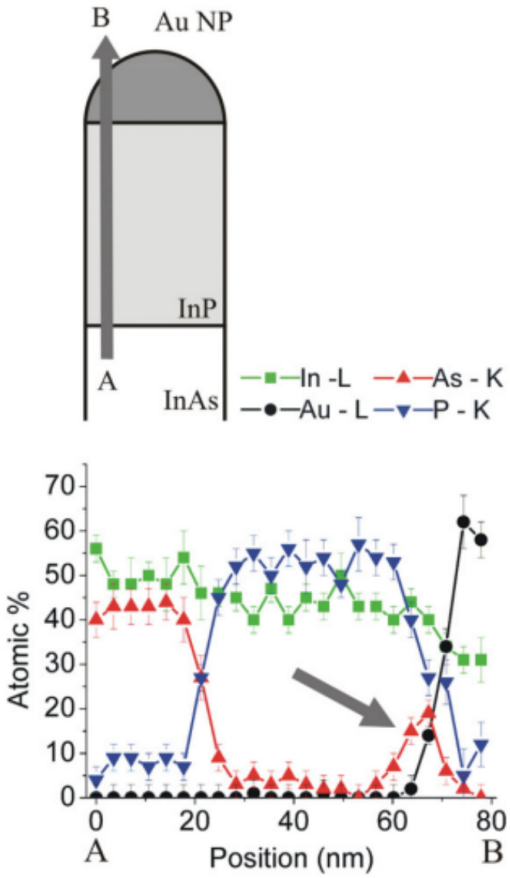


Figure 4.1: EDS line scans along the NW growth direction; a schematic diagram showing the NW analyzed region is shown in the upper region. EDS line scans along the arrow AB on a NW of the sample grown. Adapted from [81].

It is essential to mention that for this approach of chemical composition quantification, if thick objects are analyzed, the X-ray absorption must be considered [7]. However, everything we will analyze in this work is thin enough not to be concerned with absorption correction (less than 5% of X-rays are absorbed) in the Cliff-Lorimer formula. It is interesting to mention that most of the time, Cliff-Lorimer users work using the theoretical values of the K constant obtained in the data acquisition software. This is very dangerous for proper quantitative analysis. Since depending on the emission line, the error of the constant determination can be huge ($>$ than 20% for M family lines) [7]. The error in the chemical composition determination is much lower than being determined experimentally using a reference sample of known composition. Additionally, the software is typically not transparent regarding the methodology applied to calculate the constant, which can imply more errors in analytical studies.

Another well-known method for EDS chemical composition quantification is the ETA factor method, or Watanabe method [82]. Despite the simplicity of the Cliff-Lorimer method, finding a well-known standard reference sample is not always an easy task. In this context, pure elements can be explored to quantify the chemical composition of materials. The mass-thickness ρt of elements A and B can be written as a function of X-ray intensity I_i, atomic fraction C_i, electron dose D_e and the ETA-factor, were $i = A$

or B:

$$\rho t = \zeta_{\mathrm{A}} \frac{I_{\mathrm{A}}}{C_{\mathrm{A}} D_{\mathrm{e}}} \text{ and } \rho t = \zeta_{\mathrm{A}} \frac{I_{\mathrm{B}}}{C_{\mathrm{B}} D_{\mathrm{e}}} \tag{4.10}$$

Calibrating the ETA-factor with a certain element (A or B) in a reference sample of known thickness and density reference, it is possible to quantify the chemical composition ratio of a sample of elements A and B. It is essential to mention that the quantitative EDS analysis methodologies are in constant development and progression, aiming to improve precision, accuracy, and simplicity to increase the method's feasibility. A recent example is a quantification of ternary semiconductor heterostructures by a Watanabe-based method, where the feasibility of the methodology is increased by employing the use of internal references [83]. In this situation, using an iterative method to determine the ETA-factors from the internal reference, they show that the thickness and absorption effects are already included, simplifying the analysis. The most important drawback of the Watanabe-based methods is the need for a precisely determined electron dose during the measurement. Therefore, a Faraday cup or something similar is required to measure the probe current in-situ, adding a step in the process. Due to this reason and because we work with small NPs (very thin samples), we chose to use the standard Cliff-Lorimer method for our EDS chemical composition quantification of each BNP.

4.2 Chemical composition quantification of individual alloy nanoparticles

In order to increase the statistical reliability of our analysis, we developed a python-based software, and we used an open library for EM data treatment called Hyperspy [84]. Our routine aims to automate the chemical composition quantification of each alloy nanoparticle with all the necessary steps for the quantification: image calibration in energy and spatial dimensions, NP identification and counting, diameter measurement, spectra fitting, and background subtraction, integration, etc.

First, we perform a binning in the energy dimension. In this way, we increase the number of counts in each channel. With this approach, we increase the fitting quality of the EDS peaks, which reduces the sparsity (channels with low number of counts) of the dataset. Then, using the sci-kit image python library [85], we use segmentation and rotulation algorithms to identify and measure object properties in images generated by Au and Ag intensities. In sequence, using the rotulated image and addressing intensities from 0 to n number of particles (different colors), we can fit the NP size, as seen in figure 4.2. Thereby, we can select the pixel coordinates of interest in the raw dataset and sum all the spectra from these pixels, obtaining a representative spectrum for the entire NP. Performing a numerical fitting, background subtraction (included in error propagation), and peak integration of the Au and Ag peaks in the spectral region of interest, we extract the respective Au and Ag counts for their chemical composition quantification by the Cliff-Lorimer method. If the data acquisition parameters do not change very much, the procedure becomes automatic with a series of HSI as input. However, if any parameters change (ex: pixel size, beam current, dwell time, etc.), the masking threshold and segmentation criteria must be manually changed.

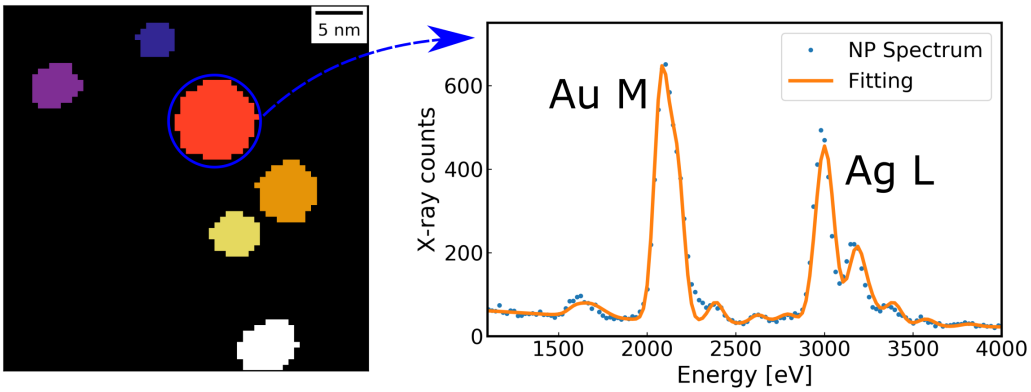


Figure 4.2: Segmented BNPs image and a representative EDS spectrum of a single NP. Blue circle is 1 pixel exaggerated to garantee all the signal extracted from the NP. Spectrum on the left results from the integrated spectra of all pixels inside the blue circle.

Here, we will always apply this methodology to extract the average chemical composition of each BNP to express the quantitative chemical information as a function of NP diameters. Through residual analysis, we can estimate the uncertainty for the diameter of each NP. By changing the model's size-adjusted radius and looking at the residues, we can estimate roughly our upper limit for size uncertainty as a single pixel. Additionally, using simulated datasets, we estimate an error in determining the diameter below 5%. Hence, choosing the upper limit of the uncertainty analysis explained above, we use one pixel as uncertainty for the diameter determination of NPs by our automatic algorithm. Considering this uncertainty, we always choose more pixels to integrate our region of interest to generate the spectrum. This way, we guarantee to measure all the chemical composition information from the BNP.

Besides, we can use the same data treatment strategy for spectra obtained through an open beam EDS-TEM measurement. In spite of the lack of spatial resolution of this operation mode, an ensemble average chemical composition of the sample can be obtained and used to verify if the chemical composition dispersions of single BNPs represent full disclosure of the sample's characterization. Furthermore, the chemical composition quantification may be improved even more through the application of advanced statistical and machine learning tools, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Non-negative Matrix Factorization (NMF), etc. In the next chapter we will discuss machine learning methodologies applied to electron microscopy HSI datasets.

4.3 Summary

This chapter addressed the importance of performing quantitative EDS chemical analysis.

The correct quantitative analysis requires confidence intervals. Therefore, we emphasized how the error bars in standard EDS analysis can be calculated.

Most STEM-EDS are limited to qualitative analysis due to the considerable increase

of work necessary for a quantitative interpretation. One example however, was given, showing the importance of quantitative analysis in the definition of interfaces in nanowires.

Here, it was also explained that the samples in this thesis are analyzed by the traditional Cliff-Lorimer method. But it was also highlighted the existence of other methods for quantification, such as the ETA-method.

The chapter finishes with the methodology used to identify the BNPs and quantify the chemical composition of each BNP from the data set.

Chapter 5

Machine Learning and Big data in Electron Microscopy: Advancing in chemical composition assesment

Electron microscopy is passing through a changeover in the way of data collection, treatment, and interpretation [16]. Recent advances in instrumentation and software technologies enable the acquisition of an enormous amount of spatially, energetic, and temporally resolved data. Likewise, advances in machine learning tools offer pathways to better visualize or represent EM information. Besides, machine learning tools are nowadays being used in several different fields of nanoscience, allowing better information extraction and interpretation, modeling of nanometric systems [15].

Probe forming the electron beam leads to the opportunity of localized data acquisition, registering 1D or 2D signals in each pixel of the scanned area, forming a hyperspectral (HSI) dataset, as explained before. A typical HSI contains 10^7 voxels considering 100x100 pixels and 1000 channels. The information contained in such datasets can be leveraged to better signal-to-noise ratios (SNR) and feature extraction [7, 12, 86]. Popular methods of machine learning tools applied to HSIs datasets are the decomposition algorithms (PCA, ICA, NMF, etc). The methodology, although simple, is compelling, but we must use it carefully. Introducing the idea of HSI decomposition, we can mathematically express the process by multiplication of matrices, where a matrix V is decomposed in two other matrices H and W:

$$V = H.W \tag{5.1}$$

Considering image spectra datasets, the matrix V is defined by m pixels (lines) and n channels (columns). Thus, H and W are $m \times k$ and $k \times n$ respectively, where typically $K << m$ is the number of components to decompose the dataset. Hence, the dataset can be expressed differently in terms of a new basis. Thereby, the aim is to calculate matrices H and W that composes V. Different methods can be used for this task, PCA or SVD (Singular Value Decomposition) being the most popular ones [17, 18]. In figure

5.1 we can see that the original datacube structure is decomposed into 1D signals (spectral components) and 2D signals (spatial components).

The spectral components are weighted by the pixel values of the spatial component so that the original dataset can be reconstructed. Each pixel weights the spectral components in a linear combination. We represent this concept with equation (5.1), where we have a matrix V which is the datacube written in terms of two other matrices H and W, for the spatial and spectral components. In figure 5.1, we illustrate the decomposition of the HSI data set into spectral components that can be used to reconstruct the data set again in a new basis.

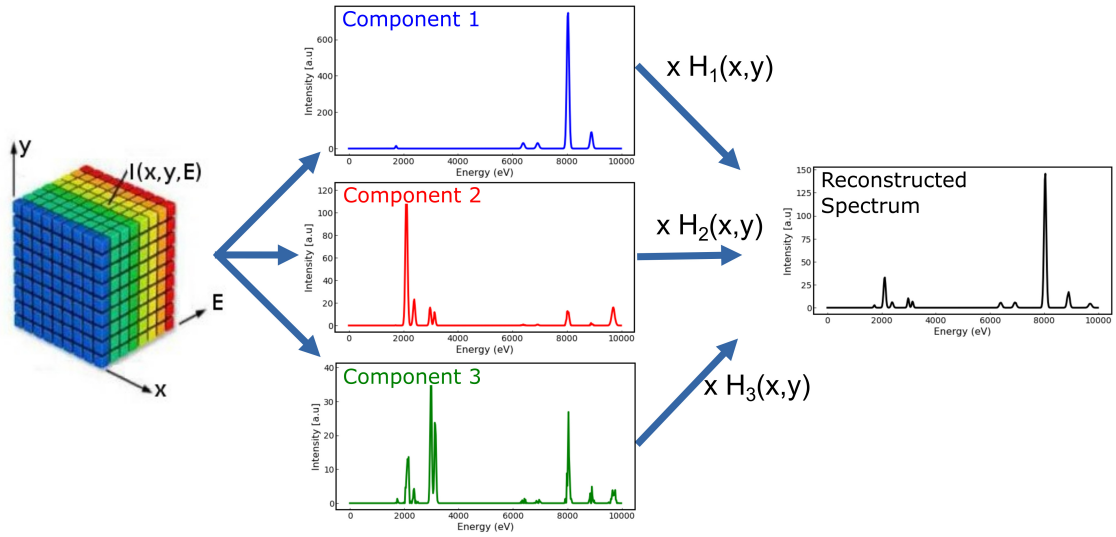


Figure 5.1: Schematic figure of a generic HSI decomposition procedure, where the original datacube is separated in eigenvector components that serves as a new basis for the reconstruction of the dataset. The spectral components (factors) are weighted by the intensities $H_i(x, y)$ of the spatial components (scores) resulting in a representative spectrum for the given pixel coordinates x and y.

The spectral components are weighted by the pixel values of the spatial component so that the original dataset can be reconstructed. Each pixel weights the spectral components in a linear combination. We represent this concept with equation (5.1), where we have a matrix V which is the datacube written in terms of two other matrices H and W, for the spatial and spectral components.

The intrinsic property of dimensional reduction of decomposition algorithms allows the reconstruction of the dataset with noise reduction [12, 87] or "denoising" as widely used in the EM community. This type of analysis is known as unsupervised machine learning since the model is built only with the data of interest itself, without a training dataset (supervised). The idea behind the method is that the voxels in the HSI can "learn" with each other expressing more representative information. As illustrated in figure 5.1, the HSI is decomposed into components, so-called loadings or factors, which are column vectors of the matrix W. The weights on the reconstruction are called scores, and the score matrix H contains weights multiplied by the factors to be summed in the linear combination in each element. In the subsequent section we will focus in the explanation of

Principal Component Analysis (PCA) which is the most popular decomposition algorithm and the one that we will constantly apply in this thesis.

5.1 Principal Component Analysis

Principal Component Analysis or PCA is a statistical method based on the reconstruction of a given data set on an orthogonal basis composed of principal components, which are the components that best represent the data set [17, 18]. This way, it is possible to reduce its dimensionality, preserving useful information and reducing noise. The method was first proposed by Pearson [88], and Hotelling[89] more than 100 years ago and is nowadays used to analyze and treat huge data sets, or as they are usually called, *big data* due to its intrinsic dimensional reduction property. PCA has been widely applied in many different fields of study, such as image processing [90], neuroscience [91] and precision agriculture [92], aiming whether better visualize or classify the information contained in data. The present study will use the method as a dimensionality reducer aiming at noise reduction of EDS-STEM hyperspectral data [7, 12, 87].

The PCA method searches for a new basis where the dataset can be reconstructed; that is, it seeks to rewrite the original dataset into a linear combination of vectors that best represents the information of interest contained in the data. The criteria established for the search for this new basis are: The maximization of variance, minimization of redundancy, and orthogonality of the generated basis, i.e., the method seeks that a first director vector is calculated from the direction of the most significant variance of the data set, then a second director vector is calculated, from the second direction of greatest variance, such that this is orthogonal to the first and so on consecutively. Thus, let $\mathbf{X}$ and $\mathbf{Y}$ be matrices $m \times n$ related by a linear transformation $\mathbf{P}$ where $\mathbf{X}$ represents the original data set and $\mathbf{Y}$ this same data set on a new basis.

$$\mathbf{PX} = \mathbf{Y} \tag{5.2}$$

PCA aims to eliminate or reduce redundant information in $\mathbf{Y}$. So, what is sought is that one variable covaries little with another variable. Thus the covariance matrix of Y, $\mathbf{Sy}$, must be optimized so that it has minimized covariances and maximized variances; that is, an optimized covariance matrix is a diagonal matrix since the non-diagonal terms are covariances and the diagonal terms are variances. The covariance matrix of $\mathbf{Y}$ is given by:

$$\mathbf{Sy} = \frac{1}{n-1}\mathbf{YY}^T \tag{5.3}$$

thus,

$$\frac{1}{n-1}\mathbf{YY}^T = \frac{1}{n-1}(\mathbf{PX})(\mathbf{PX})^T = \frac{1}{n-1}\mathbf{PX}X^T\mathbf{P}^T \tag{5.4}$$

$$\mathbf{Sy} = \frac{1}{n-1}\mathbf{PAP}^T \tag{5.5}$$

Where $\mathbf{A} = \mathbf{XX}^T$

Knowing that a symmetric $\mathbf{A}$ matrix is diagonalizable by a matrix of its orthonormal eigenvectors, one has that there exists $\mathbf{U}$ such that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ where $\mathbf{D}$ is a diagonal matrix and $\mathbf{U}$ is a matrix where the columns are eigenvectors of $\mathbf{A}$. Selecting $\mathbf{P}$ such that the row vectors $\mathbf{p}_i$, of the i-th rows of $\mathbf{P}$, are eigenvectors of $\mathbf{A}$, one has that $\mathbf{P} = \mathbf{U}^T$.

$$\frac{1}{n-1}\mathbf{P}\mathbf{A}\mathbf{P}^T = \frac{1}{n-1}\mathbf{P}\mathbf{P}^T\mathbf{D}\mathbf{P}\mathbf{P}^T \tag{5.6}$$

Since the inverse matrix of an orthogonal matrix is its transpose matrix, it follows that $\mathbf{P}^{-1} = \mathbf{P}^T$. Then,

$$\mathbf{S}\mathbf{y} = \frac{1}{n-1}\mathbf{P}\mathbf{P}^{-1}\mathbf{D}\mathbf{P}\mathbf{P}^{-1} \tag{5.7}$$

$$\mathbf{S}\mathbf{y} = \frac{1}{n-1}\mathbf{D} \tag{5.8}$$

Therefore, PCA explores the choice of a base of orthogonal eigenvectors based on variance maximization and redundancy minimization so that principal components of $\mathbf{X}$ are eigenvectors of $\mathbf{X}\mathbf{X}^T$, being these, the lines of $\mathbf{P}$ and the variances associated to these principal components, are the eigenvalues of the principal components. Thus, each component is associated with variance, and the higher the variance, the more representative the component's data set. This description can be found in more detail in [93].

The goal of using PCA on hyperspectral (HSI) EDS-STEM data is to decompose the original data into principal components that best represent it. These components can then reconstruct the original data set so that the relevant information is obtained and the spectral image noise is reduced. The reconstruction of data with less noise is done according to the number of principal components, and these are chosen according to the variance associated with these components. So, the variance proportion is analyzed to choose the number of components. This proportion is the measure of each variance contribution in data representation as, exposes the equation (5.9), where λ_i is the i-th variance and λ_j the sum variable which comprehends all p variances which represent the data set.

$$\frac{\lambda_i}{\sum_{j=1}^{p}\lambda_j} \tag{5.9}$$

Pre-processing data is a standard tread in machine learning and data mining procedures, which involves typically scaling the data or normalizing it. The HSI data generated in electron microscopy experiments such as EDS, EELS, or other methods that rely on "counting" are mainly dominated by Poisson noise, as discussed in chapter 4. Therefore, it is unsurprising that the noise cannot be evaluated equally in each pixel or channel. Since the Poisson noise scales with intensity, pixels with more counts contribute more to the noise. Thus, weighted data (or normalized) shows more interpretable results [94] and enhances the application of PCA for denoising due to the variance stabilization prior to the data decomposition. The scaled dataset matrix X_s has been calculated from $X_s = GXH$, where G and H are the spatial and spectral scaling factors applied to the X data matrix. In figure 5.2, we can see the effects of variance stabilization on the PCA

scree plots of the AuAg NPs dataset, showing that the Poisson normalization improves the "elbow" visualization on the graph, which allows us to select the principal components for the reconstruction.

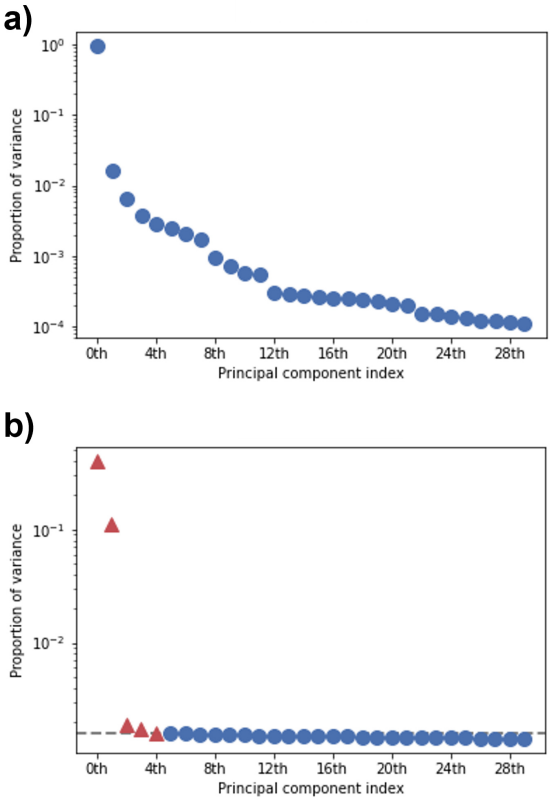


Figure 5.2: PCA scree plot of a) without and b) with Poisson normalization. The red triangles indicate the determined principal components by the "elbow" rule +1 component while the blue are the ones associated with noise and excluded from the reconstruction.

In this thesis, the machine learning algorithms used are from the python packet HyperSpy [84] which are implemented using the Scikit learn library [95]. In HyperSpy, the PCA is applied using SVD, a factorization methodology, to find the principal components, where the eigenvalues of the diagonalized covariance matrix are the squared singular values.

The determination of the number of principal components is crucial. Thus, Poisson normalization is of utmost importance to enhance the decomposition and establish a criterium to quantify the number of relevant components. In this way we are able to choose a non-arbitrary number of components associated with information and remove the rest of the reconstruction, optimizing de denoising process. Besides, we can evaluate the quality of our choice by the "elbow" method, reconstructing more or fewer components and analyzing the difference between the reconstructed data and the original data. Aiming a complementary method for the determination of PCs, Potapov [96] proposes that the information contained in the scores (weights on the 2D spatial components)

can be used to infer the number of components to be used in the data reconstruction. The author's idea is that the data cloud of the scores is anisotropic when information is contained in the components. Consequently, we can look to the score analysis (see figure 5.3) of our dataset and verify that the majority of the information can be found in the first 4-5 components. Combining the three methodologies, scree plots, inspection, and score plots, the chance of error due to truncation of the principal components is severely diminished in PCA reconstruction.

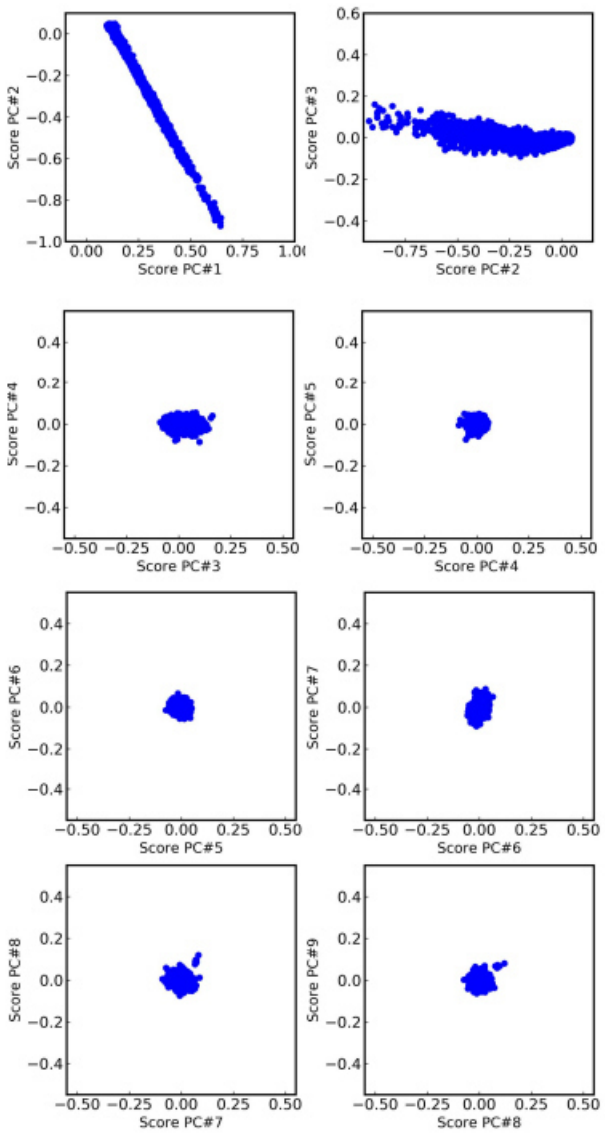


Figure 5.3: Cluster plot of scores for successive principal components obtained PCA processing a high SNR simulated EDS HSI of bimetallic NPs. Note that information-carrying component shows a structured anisotropic cloud (top), while noise components generate isotropic round clouds.

5.1.1 Component Analysis in Hyperspectral EDS-STEM images: Denoising and feature extraction

PCA is the most popular method for denoising EM data. The decomposed components are associated with different aspects of the dataset, some components carry more information than others, and some carry only noise. Selecting the components with information from the noisy ones, we can reconstruct the original dataset with the k principal components obtaining the Poisson noise reduction. The optimization maximizes variances in the diagonalization of the covariance matrix. Each maximized variance represents the amount of information associated with a component. Looking at the scree plot in figure 5.4, obtained for an experimental EDS-STEM data set of AuAg BNPs (see chapter 3), we identify as red the most relevant components of the decomposition; the principal components.

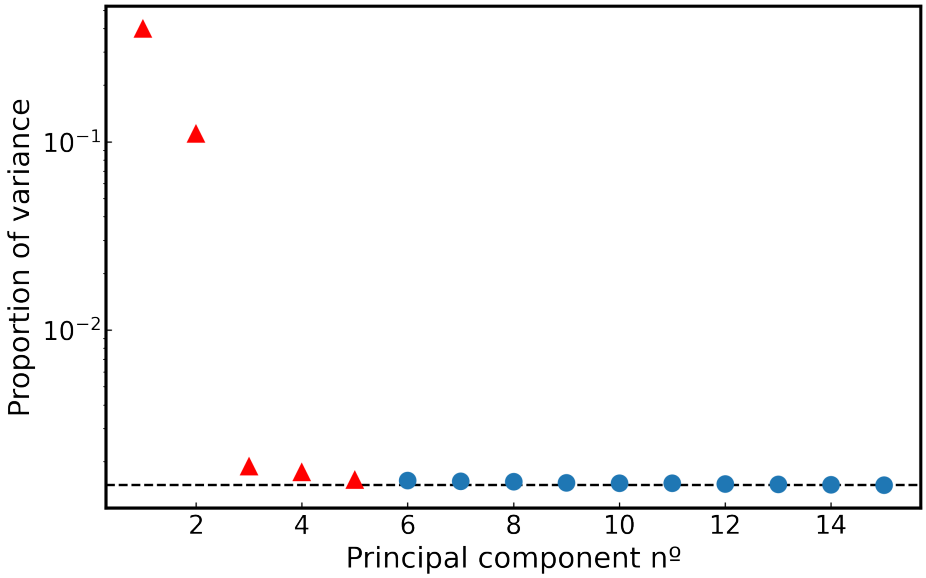


Figure 5.4: PCA scree plot of variances for each component. The red triangles indicate the principal components to be used for the reconstruction. The blue ondes are the components associated with noise excluded of the reconstruction. The dashed indicates the noise components level

Basically, with PCA what we seek is to determine components that best represent the original dataset. The principal components are used as an orthogonal basis for the reconstruction of the dataset, and the criterion that determines the relevant components are the variances associated with them. Thus, the variances in the figure 5.4 are values associated with the components that carry information representing the system. In this case, the first 3-5 variances that stand out are the variances that best represent the system, and the rest is redundancy associated with Poisson noise. In figure 5.5, we show a clear example of the proness of PCA method for denoising EDS-STEM image spectra.

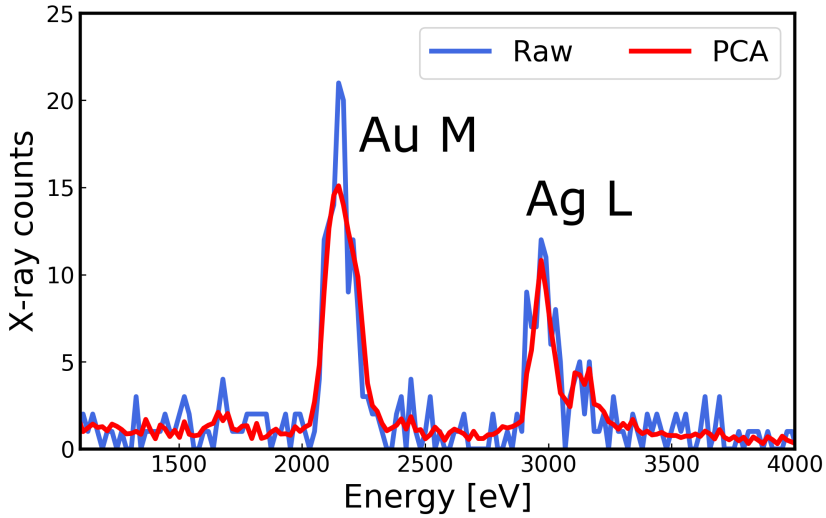


Figure 5.5: Single Pixel spectrum of raw (blue) and PCA denoised (red) EDS-STEM dataset.

The principal components can be used to reconstruct the original data set by reducing its noise. However, it is difficult to associate individual principal components with any physical meaning since the optimization allows negative counts, and the numerical algorithm has no constraints for physical reliability [97]. To overcome this difficulty, other similar methods for decomposition analysis exist. Non-Negative Matrix Factorization that forces matrix $H > 0$ and $W > 0$ is one of them. Another example is ICA that explores a linear independent but non-orthogonal basis for more physically reliable information extraction. In this case, the machine learning tool method is applied after PCA as blind source separation [20, 21]. With this methodology, Rossouw et al. [13] show that using ICA that information from different NP chemical phases is associated with each component and, therefore, can be unmixed, as can be seen in figure 5.6. Core-Shell NPs are analyzed in this work, and the background, core, and shell information are split into different components with high precision in the chemical composition assessment, showing that by unmixing signals, quantitative 3D information can be evaluated from 2D projected images without the use of complex tomography experiments.

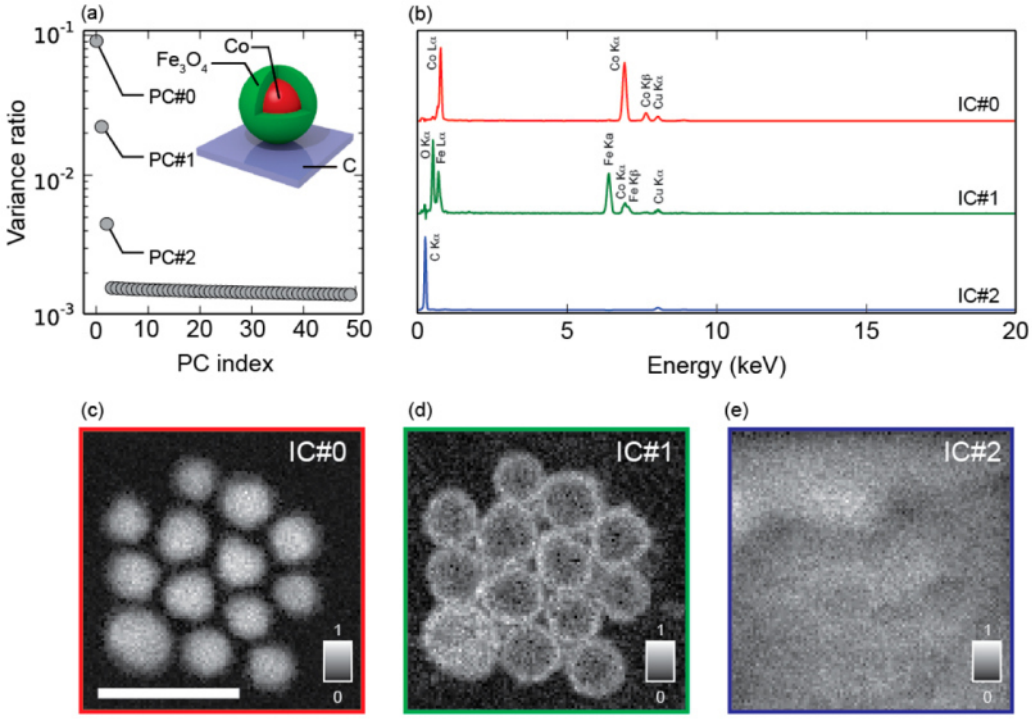


Figure 5.6: Phase identification results by PCA and ICA in spectral imaging of *core-shell* $Co@Fe_3O_4$ nanoparticles. (a) Proportion of variances of 50 components determined by PCA, (b) independent components calculated from the PCA principal components, containing the X-ray lines for the elements present. Maps of the independent components (c-e) revealing that (c) IC#0 corresponding to the cores, (d) IC#1 to the shells, and (e) IC#2 to the carbon background of the microscopy grid. 50 nm scale and normalized color bar. Adapted from [13].

In this case, the first three variances that stand out are the variances that best represent the system. Next, ICA uses the result of the PCA optimization to assign physical meaning to the generated components. Then, it select the relevant signals from the spectral image, that are shown in Figure 5.6 (b) as spectra with characteristic lines of the elements belonging to the sample IC#0, IC#1 and IC#2. The scores associated with each spectra of the independent components, are the signals of the three distinct phases, revealing with spatial resolution, the maps in figure 5.6 (c-e). Therefore, the method allows the automatic selection of the intensities corresponding to the different phases, without any bias, because it does not require any information a *priori* of the elements belonging to the system, using as the only input parameter the number of principal components obtained by PCA.

By establishing this method, in Rossouw's [13] work, the signals extracted by ICA corresponding to the different phases of the sample, could be used to separate signals and quantify the chemical composition of a *Core-Shell* $FePt@Fe_3O_4$ nanoparticle sample. The chemical composition of the FePt core was quantified from the IC#0 component, which corresponds to the core signal, independent of the Fe_3O_4 shell signal present in the IC#1 component. To verify the quality of the quantification using the independent components, the chemical composition of FePt measured through the IC#0 component, is compared

with the quantification of the chemical composition of the cores before adding the shell, as follows in figure 5.7. The dots are the measurements made for the FePt nanoparticles before adding the Fe_3O_4 shells. The dashed lines represent the values obtained by using ICA for signal separation.

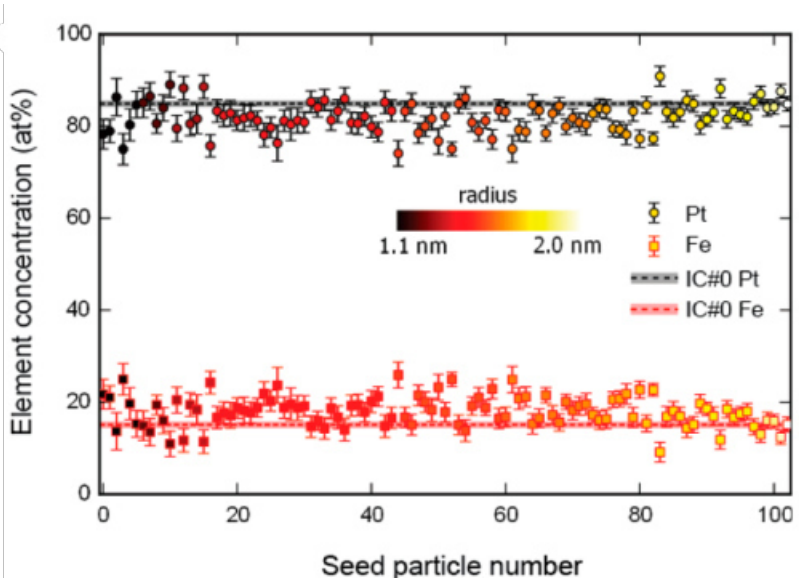


Figure 5.7: Comparison of the quantification of the FePt nanoparticle composition prior to the addition of the Fe_3O_4 shell, with the quantification by the IC#0 component after the addition of the shell. Adapted from [13]

Given the examples for SNR improvement or hidden information extraction of the datasets we discussed above, it is possible to see that machine learning tools are already very well suitable for HSI data processing. Henceforth, the methodologies may be, in the future, standard procedures to analyze image spectra such as EDS and EELS and also in 4DSTEM for imaging or electron diffraction, opening new paths for the development of quantitative analysis of nanomaterials and materials science.

In this context, we may wonder what we can learn by applying these procedures to the chemical analysis of our alloy BNPs: Can we perform denoising by machine learning tools to improve precision in our chemical composition quantification? Can we extract hidden information from our datasets that cannot immediately be seen? What is the limit of detection after the data processing? Is it possible to establish confidence intervals for the analysis? Henceforward, in this thesis, we will try to propose answers and solutions for these questions, analyzing the EDS-STEM dataset of AuAg nanoparticles. We will use PCA and NMF for whether noise reduction or blind source separation. Hence, in the next section, we will describe both methods in more detail.

5.1.2 Statistical Bias and Information Loss

The statistical bias of an estimator is the difference between the expected value and the true value of a parameter or variable being estimated. Supposing a statistical model parameterized by θ with joint probability distribution $P_\theta(x)$ for x being an observable

data. We can build an estimator θ of θ^* true value (hereafter * indicates true values). Thus, the bias $\hat{b}$ is defined as follows:

$$\hat{b}_\theta = E_x(\theta) - \theta^* \tag{5.10}$$

This way, an unbiased estimator is given when the equation (5.10) is zero or close to zero.

In PCA reconstruction, there are two main sources of errors. The truncation of principal components, which is a wrong choice of the number of components, underestimates the real number and lets relevant information out of the reconstruction. The second and more dangerous is the error in the decomposition of the factors (principal components) in the presence of noise [98, 99, 100, 101]. The user-friendly profile of PCA may erroneously indicate a fault-free procedure. Nonetheless, severely biased reconstructions (i.e., derived values different from true noiseless ones) may arise from low SNR datasets. Hence, the PCA denoising must be carefully applied to analysis where precision and accuracy are crucial. Despite the widespread use of PCA processing [102, 87, 13], the achievement of successful unbiased denoising is strongly dependent on appropriately designed and executed EDS HSI experiments with enough signal to minimize the factor decomposition error, which may induce artifacts due to biased results [86, 103, 104, 105].

Cueva et al. [86] show one of the first observations of strongly biased results after PCA denoising on electron microscopy HSI data, whereby some EELS peaks are strongly shifted, and even unrealistic signals of interfaces rise after the denoising procedure. In sequence, Lichtert and Verbeeck [104] investigate the statistical consequences of applying PCA as denoising in simulated hexagonal boron nitride atomic resolution EELS data. The authors simulated data consists of stacking four hexaboride nitride monoatomic layers, and in positions 1 and 2, in figure 5.8 extra N and B atoms are added, respectively. Looking at the decomposed components in the low and high noise situation, it is possible to see that the third component is associated with the difference in the atomic content between the atomic columns 1 and 2. This information easily appears in the low noise simulation, however, it vanishes in the high noise situation. The noise overcomes the information, and therefore the reconstructed data is severely biased due to the absence of the third composition in the PCA reconstruction. The authors propose that the information contained in the component appears only if the variance associated with this component is higher than a particular value.

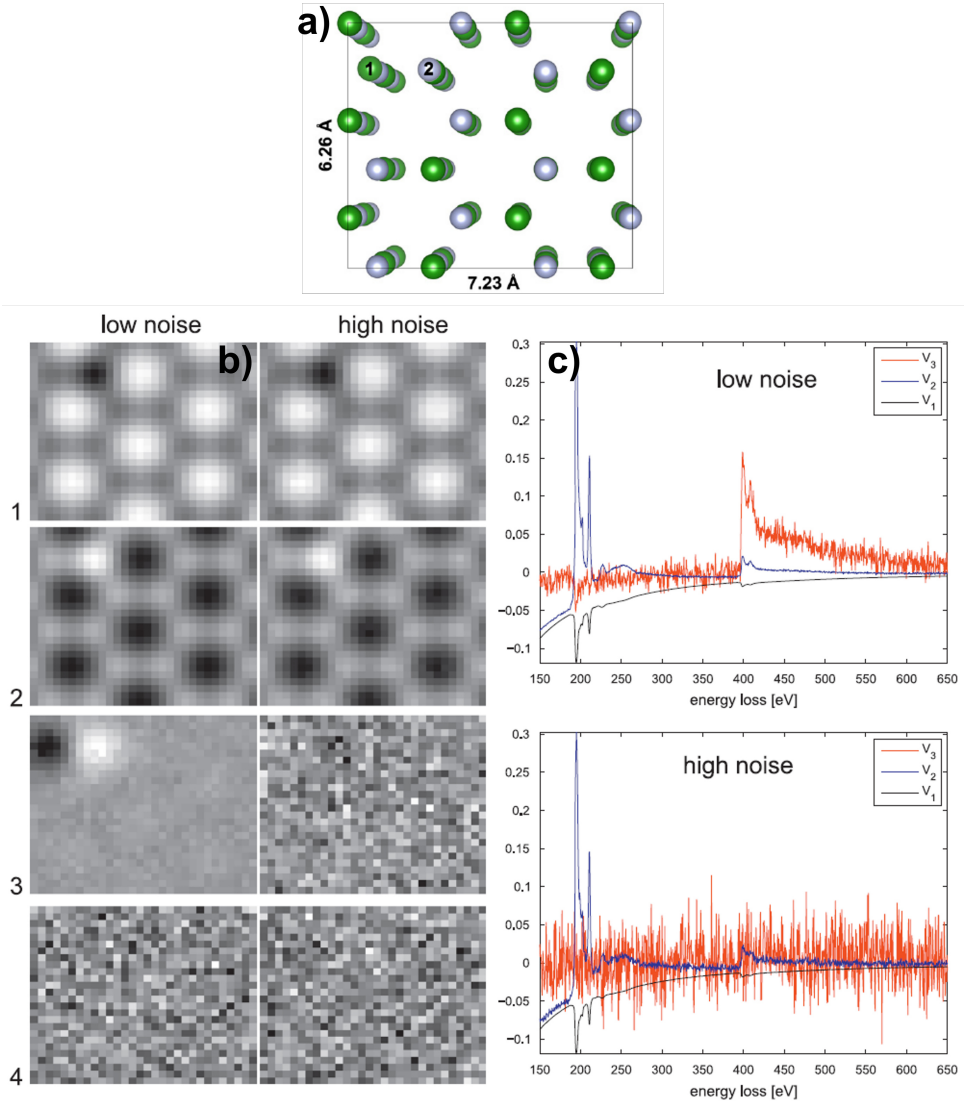


Figure 5.8: a) Model utilized to build the HSI simulation. b) Score and c) factors for the first four components in the low an d high noise situation. In the low noise information can be retrieved in the third component while the information disappears in the high noise situation.

Furthermore, in estimation theory, the Crámer-Rao inequality is a lower bound for the variance of an unbiased estimator of a parameter. The variance of such an estimator must be as high as the inverse of Fisher information, which measures the amount of information associated with a parameter. Hence, similarly to the "elbow" method in the scree plot, the authors propose that a principal component contains relevant information when the inequality $var(A_j^i) \geq F_{jj}^{-1}$ is satisfied, where F is the Fisher information matrix and the A_j^i terms are from the model parameter $\theta = (A_1^i, ..., A_j^i)$, see [104] for more details. Therefore, it can be said that reconstructing a dataset with noise overcoming the component's information raises errors in the process, similarly to neglecting the component (truncation error) in the reconstruction. The truncation reduces the noise drastically, but as a consequence, it may induce artifacts due to the biased singular values

(or variances). Lichter and Verbeeck wrote the bias based on the results of Faber et al. [100, 101] as a function of the number of pixels m, number of channels n, a homoscedastic noise (the same across all values) σ^2 and θ_i^* the true singular value in the decomposition.

$$\hat{b}_{\theta i} = \frac{m+n}{2}\frac{\sigma^2}{\theta_i^*} \tag{5.11}$$

The equation (5.11) clearly expresses the intuitive idea that the bias is proportional to the noise. However, the idea of increasing bias by increasing the dataset does not correspond to reality. Spiegelbeg and Rusz [103] propose that the bias should be $\tilde{b}_{\theta i} = \hat{b}_{\theta i}/nm$ and shows through simulated HSI data the effect of increasing m and n in the reconstruction, as can be seen in figure 5.9.

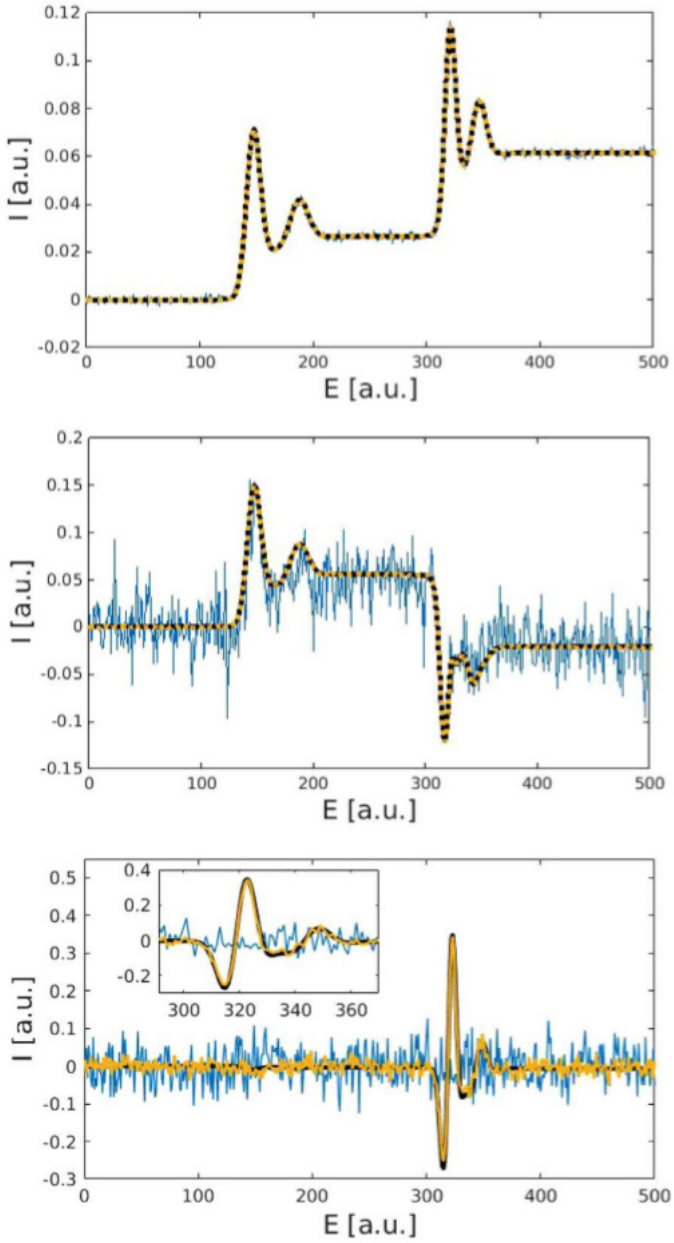


Figure 5.9: Reconstructed data with the first three principal components (top to bottom) of simulated hyperspectral data. In black with fewer pixels and no noise, in blue with fewer pixels and noise, and in yellow with more pixels and noise [103].

Similarly to figure 5.8 the information in a noisy component becomes closer to the true one, but now the noise level is constant, and what changes (drastically) is the sampling (number of pixels) in the analysis. Here the authors published results where two sets of hyperspectral data were used, the first with a reduced number of pixels and the second with a 2^{11} times larger number of pixels, but keeping the same number of channels in energy. In these simulated data, the same level of Gaussian noise was added, and the PCA method was employed to evaluate the decomposition of the data set into principal components. It can be seen that for the first two components, the three simulated conditions match, although the second component shows a noisier component

for the condition with fewer pixels (in blue). On the other hand for the third component, for the condition with fewer pixels, the signal vanishes when compared to the curve of the data set (in black), which generates an error in the reconstruction of the data and a bias. For the data set with noise but with many more pixels (in yellow), a better approximation of the two conditions (true and noisy) in the third component is verified.

In sequence, Potapov [105] discuss that alternatively to the statistical model of Faber for the description of the variances after PCA decomposition, a spiked covariance model of homoscedastic data proposed by Nadler [98, 99] can be used. Through simulations, Potapov shows that Nadler's model converges to Faber's one only for high m/n and describes much better the system for smaller and realistic m/n. The eigenvalues λ_i of the principal component can be calculated according to Nadler's theory as:

$$\begin{aligned} \lambda &= \left(1+\sqrt{\tfrac{n}{m}}\right)^2 \sigma^2, && \text{if } \tfrac{m}{n} < \left(\tfrac{\sigma^2}{\lambda^*}\right)^2 \\ \lambda &= \left(1+\tfrac{n}{m}\tfrac{\sigma^2}{\lambda^*}\right)(\lambda^*+\sigma^2), && \text{if } \tfrac{m}{n} \geq \left(\tfrac{\sigma^2}{\lambda^*}\right)^2 \end{aligned} \tag{5.12}$$

The interesting aspect of the Nadler's theory is that an information loss threshold is defined. Where the eigenvalue carries information only if (5.13) is satisfied.

$$\frac{m}{n} \geq \left(\frac{\sigma^2}{\lambda^*}\right)^2 \tag{5.13}$$

A phase transition in the information loss as a function of variance noise allows the proposition of a bias estimator. Thus, we can write information loss estimator rearranging (5.13):

$$E = \left(\frac{\sigma^2}{\lambda^*}\right)^2 \frac{m}{n} \tag{5.14}$$

Furthermore, in a subsequent publication, Potapov and Lubk [96] show the analysis of experimental EDS-STEM data of a microelectronic device and respective simulations, as can be seen in figure 5.10. In this case, the authors aim for the proper application of PCA denoising evaluating the information retrieval for an unbiased reconstruction. The layer's chemical content shows the variety that makes such a device a suitable model object for the PCA studies proposed. Here the authors aim to get an optimal number of components in the reconstruction and claim that in a system with multiple chemical phases, multiple principal components associated with these variations are expected. However, the number of components can typically be lower due to noise corruption, which may lead to biased results.

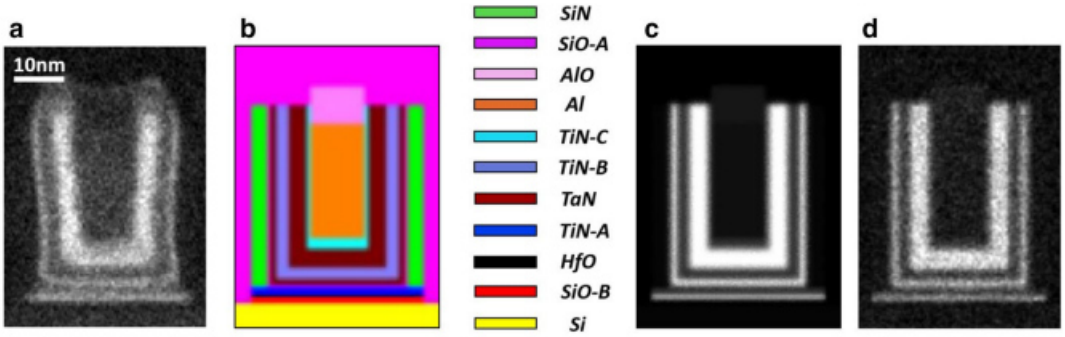


Figure 5.10: HSI dataset of CMOS microeletronic device. a) EDS-STEM acquired data, b) model proposed for the simulation, c) HSI simulated without noise and d) corrupted by Poisson noise.

They propose that, in contrast to EELS, where the enormous background may lead to artifacts after the reconstruction, the lack of signal in each channel of EDS, henceforth defined as sparsity, leads to not optimal performance of PCA as well. The authors propose that by applying previous data treatment to the data, such as weighting and filtering, the PCA reconstruction is enhanced, and the qualitative results may be less biased. In figure 5.11 the results of experimental data analyzed are shown, and it is possible to observe in the last case (filtered and weighted) that the information retrieval is enhanced and the intensity maps of different chemical phases are recovered with precision.

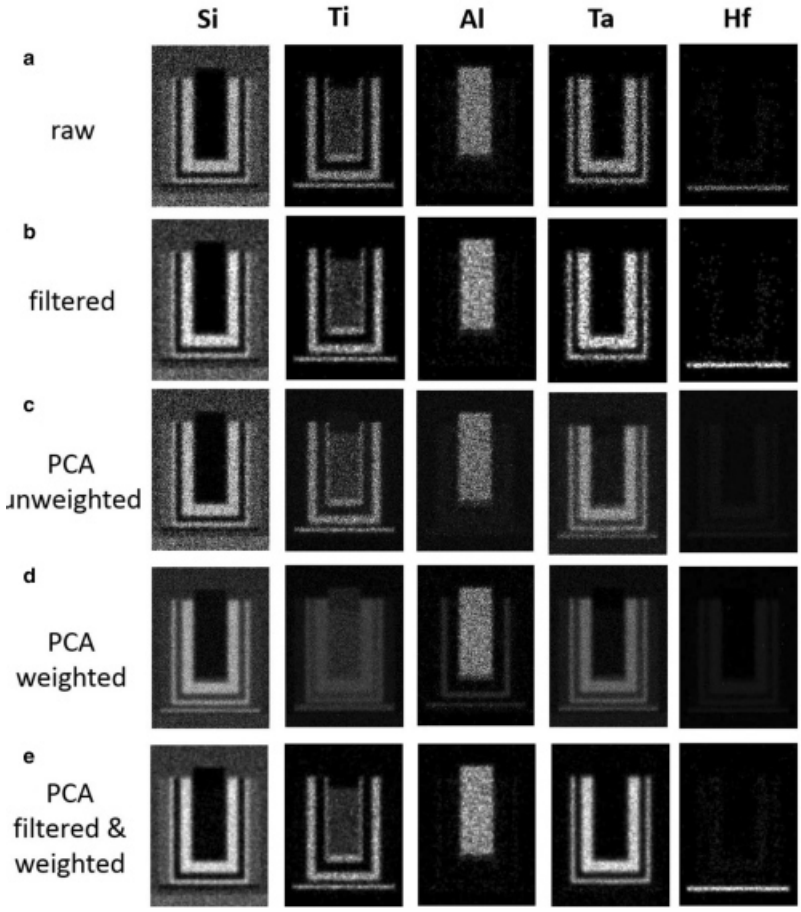


Figure 5.11: Results for the simulated HSI data. Elemental maps of a) raw, b filtered, c) PCA unweighted, d) PCA weighted and e) PCA filtered and weighted, Si, Ti, Al, Ta and Hf intensities. The last case e) where filtering and wieghted PCA are applied is the one that fully recovers all the denoised and unbiased information.

This analysis evaluates the principal components by the information loss threshold discussed above, arguing that the information in a certain component cannot be retrievable, even theoretically, if λ^* is lower than $\sqrt{\frac{n}{m}}\sigma^2$. Their results for each component show that information retrievability occurs only in 7 components. This number is later confirmed by the score cloud analysis explained in the last section. PCA denoising works well when applied to a dataset of experiments properly designed for achieving information retrieval. In an ideal situation, the components would also be used to extract information, as seen in the scores of figure 5.11. However, as already discussed before, the interpretability of the PCA components is not always an easy task, principally in the factors where the x-ray peaks can be highly unrealistic. Although PCA is a robust decomposition method, another method can be exploited for the task of unmixing signals. To unmix signals is a goal that can be achieved by making the components in the compositions more realistic and physically interpretable. Therefore, NMF rises as a possible tool for this task.

5.2 Non-Negative Matrix Factorization

Similar to PCA, NMF is a decomposition method for unsupervised machine learning. Nevertheless, the non-negativity of its scores and factors ($H > 0$ and $W > 0$ for $V = H.W$ where V is the data matrix), plays a significant role in finding hidden information in the dataset by blind source separation [106] in spectral imaging [19]. The method herein retrieves more physically interpretable results on the scores and factors, which is relevant for our studies where x-rays are always positive counts. It is important to properly unmix signals and interpret them in favor of chemical composition quantitative analysis.

One of the first works to be famous for using and interpreting NMF was in face recognition [107]. The authors discuss in the article the image properties and how humans and machines can perceive it, showing that while other decomposition methods split the information in a holistic and not intuitive way, NMF separates parts of the object being decomposed (in this case a face, see figure 5.12). In the figure a), the eigenfaces, i.e., the eigenvectors, are an average description of the face in the first component, and the other describes the differences, while in the NMF case represented by figure b), all attributes of the face are separated into defined objects which compose the face. It is reasonable, thereby, rationalize in terms of EDS HSI datasets where the features extracted by NMF are not a simple numerical optimization but are more realistic in the sense that physical or chemical significance can be attributed to the decomposed results more easily than in PCA [97]. It is relevant to address also that by reconstructing the dataset on the basis of the NMF components, the tool can work as a denoising procedure such as PCA [108] but probably less prone to noise reduction since no components are excluded, but optimized.

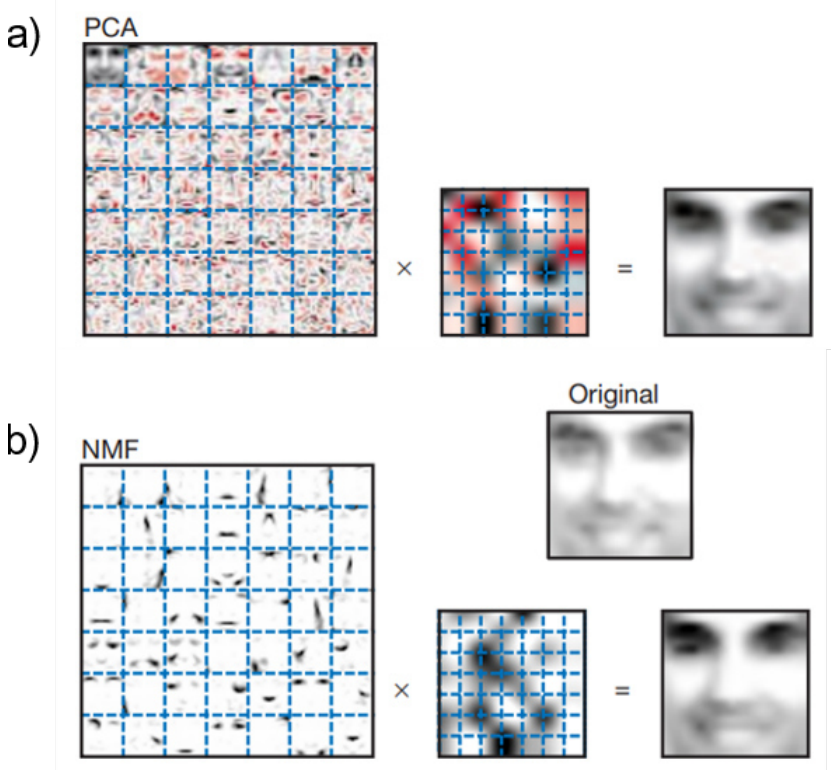


Figure 5.12: NMF learns resembling part objects of faces, whereas principal components analysis (PCA) learn holistic representations. Positive values are illustrated with black pixels and negative values with red pixels. The coefficients of the linear superposition (scores) are shown next to each montage

In the last years, NMF has been widely used for unmixing signals in electron microscopy HSIs in many research fields [109, 110, 111]. Applying to EDS quantitative chemical composition analysis, Jany et al. [14] is an exciting example of the method's utility in quantifying the chemical composition of unmixed signals. In their work, the authors analyze Scanning Electron Microscopy EDS data, in which the challenge is to quantify nanoobjects on a substrate since the beam depth is very long and signal from the object of interest and substrate are entirely mixed. Here, $AuIn_2$ nanowires are onto an InSb substrate. PCA is first applied to help infer the number of components to decompose the dataset in the NMF analysis. Thus, by the scree plot inspection, the dataset is decomposed into 3 NMF components: substrate signal, background signal, and nanowires, as shown in figure 5.13. Consequently, the chemical composition of the nanowires and the substrate are quantified from the second and third NMF components. Even though the authors do not discuss implementing error bars for this type of analysis, the values obtained are confirmed by Monte Carlos simulations and complementary cross-sectional measurements, where the mixing does not happen. Hence, even with the substrate and the nanowires sharing In, the signal is properly unmixed for further analysis showing that the method can work for EDS quantitative information extraction. ICA and NMF are also compared since both methods are the favorites for blind source separation applications. What their analysis reveals, in this case, is that NMF allows better information retrieval than ICA for that specific application.

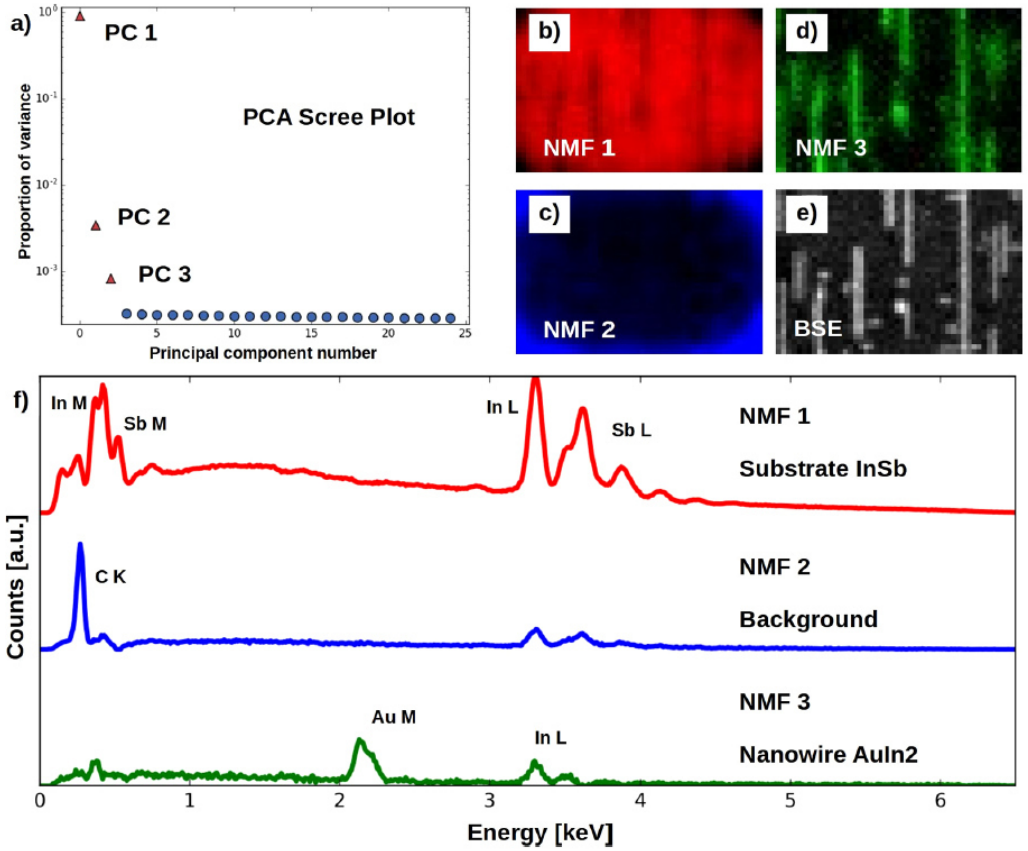


Figure 5.13: Results of the BSS by NMF and PCA of EDX spectrum image of AuIn2 nanowires on InSb. (a) PCA scree plot of first 25 principal components; 3 components exhibit significantly higher variance. Non-negative matrix factorization component maps (bd) and corresponding component spectra containing X-ray lines of the elements (f). Backscattered electrons image of the nowires (e). The BSS by NMF separates very good the signal from AuIn2 nanowires (NMF 3) from InSb substrate (NMF 1) and from background (NMF 2). Figure extracted from [14].

A vital discussion to address in unmixing signal procedures by the factorization methods is if the algorithm is convex, i.e., if the solutions converge [106]. Uesugi et al.[112] are one of the first in the electron microscopy community to properly discuss the convergence of NMF in 4DSTEM (2D electron diffraction in each pixel) applications. First, the authors compared the performance of PCA and NMF, showing that the PCA allows negative counts, the optimization generates unrealistic electron diffraction patterns in the principal components, and this issue is solved by employing NMF analysis. Thus, all the components generated carry direct crystallographic information. Nonetheless, this work's natural and important contribution is that it shows, as can be seen in figure 5.14, a statistical analysis of the number of components selected in the decomposition by NMF. Several reconstructions of the dataset are performed for a given number of components k. Increasing k and doing a mean squared error analysis between original data and reconstructed data, Uesugi et al. show that although the algorithm for NMF is not convex, the convergence is relatively good for a low number of components. Therefore, the quality of the convergence relies on the number of components selected to solve the

problem.

Guiding by the PCA scree plot is a convenient way of choosing the number of components to perform NMF. However, if the system is too complex, many components are expected, and therefore, the results may not converge, showing that applying a powerful statistical method requires careful analysis.

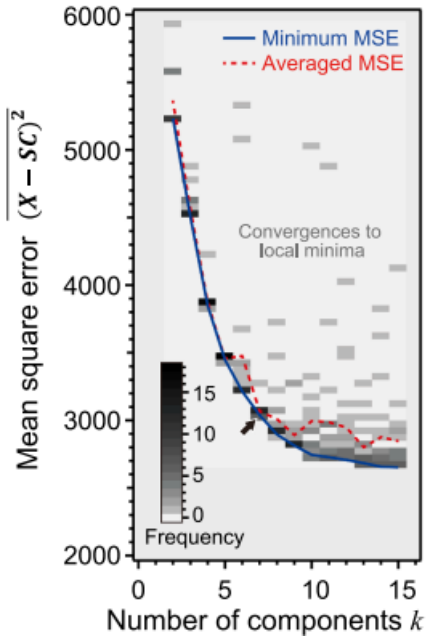


Figure 5.14: Mean square error as a function of number of components. Gray squares denote the frequency of the convergences, as shown by the inset brightness bar. The blue solid line and red broken line respectively indicate the minimum and averaged MSEs for each number of components.

5.3 Partial conclusion

We learned so far that EDS-STEM can be used for quantitative EDS chemical analysis and that we can use this tool to investigate elemental distributions on BNPs. In this chapter, we describe machine learning tools exploring the HSI data structure for noise reduction and feature extraction to optimize the EDS analysis of our data sets. We show we can use PCA for extreme noise reduction; however, the methodology is not fault-free, and we have to investigate if biased reconstructed data are generating artifacts in the final results. Despite the incredible advantages of exploiting of unsupervised machine learning tools such as PCA and NMF, we must consider its effects in quantitative analysis. The next chapter discusses in detail these aspects of our AuAg NP model system.

Chapter 6

Exploiting Principal Component Analysis for denoising EM image spectra: Effects on quantitative EDS analysis

This chapter contains the publication:

Murilo Moreira, Matthias Hillenkamp, Giorgio Divitini, Luiz H. G. Tizei, Caterina Ducati, Monica A. Cotta, Varlei Rodrigues and Daniel Ugarte

Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part I, Reducing Reconstruction Bias

Microscopy and Microanalysis, 2022, 28(2), 338-349. doi:10.1017/S1431927621013933

In this chapter, we show our main results, published as an article in the Microscopy and Microanalysis journal. Despite the wide use of PCA denoising of EDS-STEM data sets in the electron microscopy community, the procedure is not fault-free. Care must be taken with its application to noisy data sets, where the lack of information beyond noise may induce statistical bias in the generation of the principal components. These noise corrupted components are used to reconstruct the data set for Poisson noise reduction purposes. Thus, systematic effects can be observed in the quantitative analysis of AuAg BNP chemical compositions. Therefore, we propose a strategic use of experiments and simulations to derive the information loss estimators from them that can evaluate the quality of the reconstructions. Additionally, we expect to show ways to solve the bias issue for improving EDS chemical analysis in small nanoalloys.

Microscopy and Microanalysis (2022), 1–12
doi:10.1017/S1431927621013933

Original Article

Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part I, Reducing Reconstruction Bias

Murilo Moreira[1], Matthias Hillenkamp[1,2], Giorgio Divitini[3], Luiz H. G. Tizei[4], Caterina Ducati[3], Monica A. Cotta[1], Varlei Rodrigues[1] and Daniel Ugarte[1]*

[1]Instituto de Física "Gleb Wataghin", Universidade Estadual de Campinas-UNICAMP, Campinas, SP 13083-859, Brazil; [2]Institute of Light and Matter, University Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5306, Villeurbanne F-69622, France; [3]Department of Materials Science and Metallurgy, University of Cambridge, Cambridge CB3 0FS, UK and [4]Laboratoire de Physique des Solides, Université Paris-Saclay, CNRS, Orsay 91405, France

Abstract

Scanning transmission electron microscopy is a crucial tool for nanoscience, achieving sub-nanometric spatial resolution in both image and spectroscopic studies. This generates large datasets that cannot be analyzed without computational assistance. The so-called machine learning procedures can exploit redundancies and find hidden correlations. Principal component analysis (PCA) is the most popular approach to denoise data by reducing data dimensionality and extracting meaningful information; however, there are many open questions on the accuracy of reconstructions. We have used experiments and simulations to analyze the effect of PCA on quantitative chemical analysis of binary alloy (AuAg) nanoparticles using energy-dispersive X-ray spectroscopy. Our results demonstrate that it is possible to obtain very good fidelity of chemical composition distribution when the signal-to-noise ratio exceeds a certain minimal level. Accurate denoising derives from a complex interplay between redundancy (data matrix size), counting noise, and noiseless data intensity variance (associated with sample chemical composition dispersion). We have suggested several quantitative bias estimators and noise evaluation procedures to help in the analysis and design of experiments. This work demonstrates the high potential of PCA denoising, but it also highlights the limitations and pitfalls that need to be avoided to minimize artifacts and perform reliable quantification.

Key words: denoising, energy-dispersive X-ray spectroscopy (EDS), nanoparticles, principal components analysis (PCA), quantitative chemical analysis

(Received 19 October 2021; revised 21 October 2021; accepted 16 December 2021)

Introduction

Nanotechnology exploits the unique properties of nanoparticles (NPs) in many different fields such as catalysis, magnetism, and plasmonics (Heiz & Landman, 2007; Odom & Schatz, 2011; Binns, 2014). The electronic properties of bimetallic NPs show a complex behavior in relation to size and elemental composition; for example, several suggested applications exploit the optimization of either physical properties, such as surface plasmons in some optical devices, or chemical reactivity/selectivity in catalysis. The morphological and structural characterization of an NP sample requires the measurement of crystalline structure, shape (rod, sphere, wire, etc.), crystal habit (possible faceting), and size distribution (mean diameter and size dispersion). In the case of multielemental nanosystems [i.e., nanoalloys (Alloyeau et al., 2012; Ferrando, 2016)], we must determine the chemical composition (mean one and distribution as a function of size) and possibly the occurrence of chemical inhomogeneity within the particles [core–shell segregation, Janus distribution, compositional gradients, etc. (Lyman et al., 1995; Mukherjee et al., 2012, 2014)].

Transmission electron microscopy (TEM) has made huge progress in the last decades, considering electron optics, detector efficiency, reproducibility, automation, etc. (Williams & Carter, 2009, 2016; Hawkes & Spence, 2019). Scanning transmission electron microscopes (STEMs) are capable of recording the so-called hyperspectral image (HSI; Pennycook & Nellist, 2011), where an entire analytical spectrum or diffraction pattern can be registered at each image pixel. The HSI approach is currently used to generate huge 3D or 4D datasets [X-ray energy-dispersive spectroscopy (EDS), electron energy loss spectroscopy (EELS), cathodoluminescence (CL), electron diffraction (ED), scattering distribution, etc.] (Thomas et al., 2015).

A typical HSI datacube may routinely contain 10^7 voxels (100×100 pixels, each containing a spectrum of 1,000 channels). The information contained in such massive arrays can be leveraged to a better signal-to-noise ratio (SNR) by applying blind source separation algorithms (Cueva et al., 2012; Williams &

*Corresponding author: Daniel Ugarte, E-mail: dmugarte@ifi.unicamp.br

Cite this article: Moreira M, Hillenkamp M, Divitini G, Tizei LHG, Ducati C, Cotta MA, Rodrigues V, Ugarte D (2022) Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part I, Reducing Reconstruction Bias. *Microsc Microanal.* doi:10.1017/S1431927621013933

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Microscopy Society of America

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1431927621013933

Carter, 2016; Brown et al., 2020). By exploiting redundancies, machine learning (ML) methods are also able to efficiently classify data by finding hidden correlations. The most popular algorithm is principal component analysis (PCA; Jolliffe, 2002; Jolliffe & Cadima, 2016) that provides a number of orthogonal eigenvectors (referred to as "loadings" or "components") encoding the information contained in data variance. The reconstruction data with a few PCA components allows the extraction of meaningful information and reduction of noise (usually called "denoising") (Titchmarsh & Dumbill, 1996; Titchmarsh, 1999; Keenan & Kotula, 2004; Kotula & Keenan, 2006; Lichtert & Verbeeck, 2013; Kotula & Van Benthem, 2015; Potapov & Lubk, 2019). EDS data is particularly well suited for PCA processing, because signal is constituted by well-defined peaks on a low varying background (Lichtert & Verbeeck, 2013). In contrast, EELS display absorption edges and background which may show changing complex shapes as a function of chemistry and sample thickness, so PCA application may be rather complex (Cueva et al., 2012; Lichtert & Verbeeck, 2013).

EDS chemical mapping has become one of the most popular TEM-related experiments because the high counting capabilities of silicon drift detector (SDD) has strongly reduced total acquisition time (Schlossmacher et al., 2010; Watanabe et al., 2010). PCA has become a popular easy-to-use tool to denoise and improve qualitatively EDS elemental maps. In contrast, it has scarcely been used for quantitative chemical EDS-STEM analysis (Burke et al., 2006), where it is essential to confirm the PCA reconstruction fidelity (accuracy) and calculate reliable confidence intervals. The user-friendly profile of PCA may erroneously indicate a fault-free procedure; indeed, seriously biased reconstructions (i.e., derived values differ from true noiseless ones) may arise from low SNR datasets. The achievement of successful unbiased denoising is strongly dependent on appropriately designed and executed EDS HSI experiments. Despite the widespread use of PCA processing (Burke et al., 2006; Parish & Brewer, 2010; Rossouw et al., 2015), there are still many open questions on noise-related effects and on how to infer if reconstructions are biased.

Alloy NP samples are expected to display a size dispersion and elemental composition variation, making their detailed characterization an extremely challenging task. In this work, we have explored the application of PCA denoising in quantitative analysis of binary alloy NP using EDS HSI; this issue represents a rich, complex ensemble of yet open questions. This investigation requires the scrutiny of the two different aspects associated with any measurement: accuracy and precision. The first one requires the minimization of PCA reconstruction bias; the second topic deals with the potential improvement of uncertainty interval due to PCA capacity of using the redundancy of the whole dataset. We will tackle these two different points using both experiments and simulations, and the results will be described in a series of two papers (Part I and Part II, respectively). For EDS studies, the main contribution arises from counting noise (leading to PCA random noise bias), and the present manuscript reports a thorough analysis of the minimal level that the SNR must exceed to guarantee that PCA reconstruction carries the information of true NP composition variation in the sample. We have analyzed the applicability of several suggested PCA bias estimators (Faber et al., 1995*a*, 1995*b*; Malinowski, 2002; Nadler, 2008, 2009) and suggested some criteria to predict the reliability of derived chemical composition distributions.

Materials and Methods

Nanoparticle Synthesis

Bimetallic (Au_xAg_{1-x}) NPs have been produced using a homemade gas aggregation cluster source where a cylindrical magnetron is used to sputter atoms from a central target made of twisted Au and Ag wires (de Sá et al., 2014). A series of electrostatic lenses are used to reduce NP kinetic energy to achieve a "soft landing" on the TEM grid (for a ~4 nm in diam. NP, this corresponds to ~0.05 eV/atom). The NP size distribution contained in the molecular beam can be followed *in situ* by the time-of-flight mass spectrometer (TOF-MS). TEM images indicate that the NP size distribution follows a log-normal function (mean diameter ~4 nm, width ~3 nm in width; see Supplementary Fig. S1). Three different bimetallic nanoparticle samples (A, B, C) have been used in the present study whose Au content in atomic fraction is (0.73 ± 0.01), (0.55 ± 0.02), and (0.48 ± 0.01), respectively. The chemical composition has been measured on an ensemble of NPs using an open parallel TEM beam with a long counting time to increase X-ray total intensity while minimizing radiation damage (dose ~10 $e^-/Å^2$).

Electron Microscopy: Data Acquisition and Processing

We have used several STEM microscopes to compare and optimize experimental configurations and beam energy values: JEM-2100F (SDD 30 mm^2, LNNANO-Campinas-Brazil), JEM-F200 (two SDD, 0.8 sR, JEOL), FEI-Tecnai G20F (Si(Li) 30 mm^2, LCE-UFSCAR-Brazil), TitanThemis (Super X Quad SDD, 0.8 sR, LNNANO-Campinas-Brazil). Microscopes were set up for ~0.5 nm pixel size, probe diameter ~0.7–1.0 nm, and a dwell time of ~200 ms per pixel for the sake of comparison between instruments. We have used different TEM instruments and, for all experiments, specimens were mounted in a low background Be holder. On average, each HSI image contains about 5–8 particles, then several HSIs are necessary to gather statistical meaningful description of the NP sample.

Measured EDS spectra have been binned to get a total of 512 energy channels, in order to increase SNR for PCA processing. Automatic recognition of NPs on the chemical maps (binarization and segmentation) has been performed using Scikit-image Python library (van der Walt et al., 2014). We have calculated the total X-ray characteristic signal (Au-Mα and Ag-Lα peaks) from individual particles by adding signal from the pixels located inside each NP region, because it is well known that addition maintains the Poisson nature of the sum.

The quantitative analysis of the NP chemical composition followed the Cliff-Lorimer approach [Equation (1), C_x and I_x are the atomic fraction and the X-ray intensity of element x, respectively (Cliff & Lorimer, 1975)]. The Cliff-Lorimer factors (k_{AB}) were derived experimentally using a thin film of known composition ($k_{AuAg} = 0.93 \pm 0.01$ for data from Experiment I and $K_{AuAg} = 0.85 \pm 0.01$ for Experiment 3, simulations used $k_{AgAu} = 1$). The film was generated using a multilayer thin film calibrated using a quartz balance to control the atomic ratio through the relative film thickness.

$$\frac{C_{Au}}{C_{Ag}} = k_{AuAg}\frac{I_{Au}}{I_{Ag}}. \quad (1)$$

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

The calculation of chemical composition error bars has taken into account all EDS quantitative analysis steps applied on a fitted curved on the experimental EDS spectra (background removal, X-ray peak integration, and composition determination). All EDS and HSI PCA processing steps have been performed using the open-source Hyperspy Python library (de la Peña et al., 2017).

We have performed experiments at 80 kV to increase electron ionization cross-section considering relativistic estimations (Zaluzec, 2019). It is important to consider that recent microscopes (as the AC microscope used here) show very good performance at low voltages. However, in practice, the reduction from the nominal microscope voltage (300/200 keV) to lower values (100–80 keV) may reduce electron gun efficiency and increase the effect of TEM optical aberrations. Consequently, the probe current may decrease significantly, producing an unwanted consequence of, effectively, reducing EDS counts. It is important to verify properly the microscope performance at a lower voltage, because the optimal setup is strongly dependent on the particular TEM instrument. In addition, lower voltages (80–100 keV) should be preferred for lower atomic number samples in order to reduce knock-on damage (preferential damage mechanism for lower atomic number samples (Braidy et al., 2008; Egerton et al., 2010; Egerton, 2012, 2019). We would like to note that all experiments discussed in this study have been acquired using the same dwell time (200 ms) for the sake of comparison between different experimental conditions.

Our experiments accumulated several frames for the final HSI (Jones et al., 2018), what minimizes the dose rate and extends the sample useful life. Instead of taking a single image with dwell time 200 ms, we have taken a series of images of the same region (e.g., 10 scans at 20 ms dwell time), and subsequently, added the individual frames. An important issue related to alloy NPs analysis is guarantying that composition is not modified along measurement time by radiation damage. As Ag atoms should be sputtered more easily than the Au ones, we would expect a gold enrichment with time (Egerton, 2019). To analyze this issue, we have taken a long series of 20 ms dwell time scans and followed the chemical composition evolution. This allowed us to verify that no detectable composition changes occur for the chosen dwell time in this work (200 ms).

At present, high spatial resolution chemical mapping is routinely performed with good reliability and reasonably short acquisition times, but spectra from individual pixels typically have insufficient counts for proper statistical quantification. According to the basic procedures of EDS microanalysis, a characteristic peak occurrence can be identified if the signal is three times (3×) the background noise (σ_{BGD}). This criterion, however, only applies to detection and is inadequate for quantitative EDS chemical analysis. The seminal work of Currie (Currie, 1968, 1999; Belter et al., 2014; Williams & Carter, 2016) shows that a quantitative analysis requires a significant increase of the signal level (approximately 15 times higher than the uncertainty interval) to attain the measurement of a quantity with a 10% relative error. This increases enormously the signal required, in particular when the detection suffers from shot noise, given that in Poisson statistics intensity uncertainty is the square root of the intensity itself, $\sigma_I = \sqrt{I}$.

PCA Processing

PCA takes as basic input a 2D matrix D of dimension m x n; m represents the number of pixels that have been acquired and n represents the number of channels in the spectra (Jolliffe, 2002). This matrix description shows that PCA does not consider any specific information about spatial location of the pixels in the experimental image. When a spectral correlation exists between pixels it will show up as a significant weight (score) for a particular PCA component (loading).

PCA processing provides a number of orthogonal eigenvectors encoding the information contained in data variance. The eigenvalues (or singular values) associated with each component express their information content and are usually displayed in the so-called scree-plot (Titchmarsh & Dumbill, 1996; Titchmarsh, 1999; Jolliffe, 2002; Jolliffe & Cadima, 2016). In our experiments, the scree-plots show a clear and regular profile with a well-defined kink, where the number of components showing data variance visibly higher than noise can be easily identified (see example in Fig. 1).

The principal source of noise in an EDS HSI is counting noise of Poisson nature; then before PCA processing, we have performed the variance stabilization using a data scaling procedure (Keenan & Kotula, 2004; Kotula & Keenan, 2006; Kotula & Van Benthem, 2015). The scaled dataset matrix D_s has been calculated from $D_s = GDH$, where G is the spatial scaling factor, D is the raw data, and H is the spectral scaling factor [as implemented in Hyperspy (de la Peña et al., 2017)]. Recently, it has been

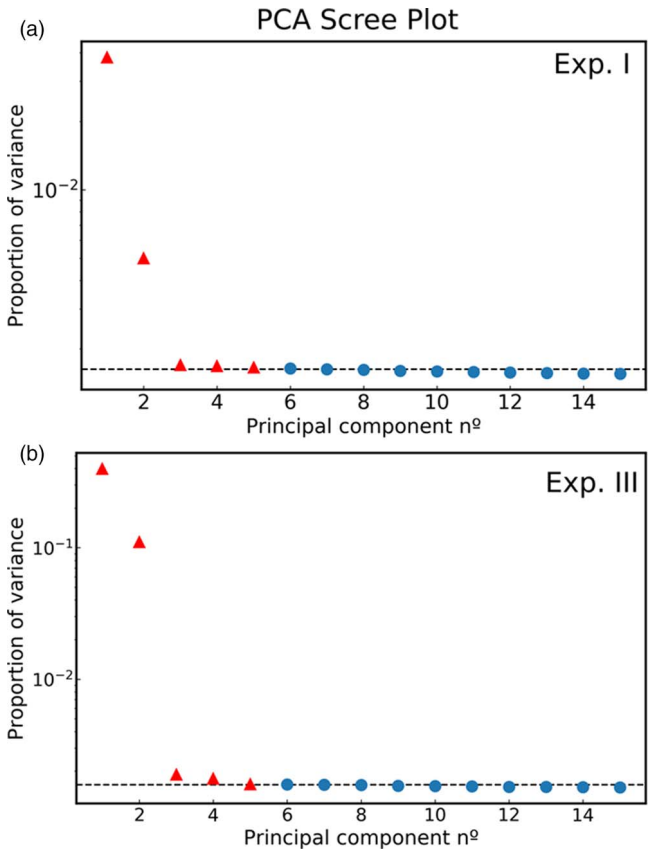


Fig. 1. Comparison of scree-plots derived from EDS measurements on AuAg alloy nanoparticles experiments acquired using a noncorrected FEG TEM/STEM (**a**) and a modern aberration-corrected microscope equipped with multiple EDS detectors (**b**). Reconstruction derived from experiments associated with (**a**) showed a strong random noise bias, while HSI described by (**b**) showed a good PCA reconstruction (see text for explanations). Note that the variance of the third principal component is slightly above noise component in (**b**). Triangular points represent the components used for data reconstruction in this work.

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

suggested that for a very sparse data matrix obtained from NP samples (Kotula & Van Benthem, 2015), better multivariate decomposition results are obtained if the scaling only considers the spectral scaling factor ($D_s = DH$). These authors analyzed a rather high sparse measurement (99.8%). Our experiments have been acquired with a pixel area of 0.5×0.5 nm^2 = 0.25 nm^2, considering a particle of 6 nm in diameter, its area is about 28 nm^2, which should be covered by approximatively 110 pixels. This simple calculation shows that the sparsity level of in about 85–90% (scans of 64×64 pixels containing about 5–8 NPs, about 80–90% [1−(660/4,096)]∼0.84 or 84%)). As the sparsity of our data is not so extremely low, we have used the conventional complete spatial and spectral scaling in this work; the possible influence of changing the scaling procedure will be analyzed in future work.

In this report, the analyzed samples are very close to ideal ones: binary alloy particles (with a certain composition variation) distributed on a uniform thin substrate (a–C). As a consequence, the information content of each PCA principal component can be understood reasonably well; simulations of high SNR HSI indicate three dominant loadings without significant information mixing (see Fig. 2): (a) first component associated with TEM support grid and another EDS spurious X-ray sources (Fe from magnetic lenses, Cu from grid, etc.), what is due to the fact that, in our experiments and simulations, most of pixels (∼85%) do not contain any NP EDS signal; (b) second component containing information on the mean NP EDS spectrum and mean chemical composition defined by the Au and Ag counts ratio of this component ($I_{Au,PC\#2}/I_{Ag,PC\#2}$); and, finally, (c) third component accounting for variation of composition among different particles (Potapov & Lubk, 2019). We must note that the third component represents an unphysical spectrum (Potapov, 2016), because the EDS peaks (Au-M and Ag-L edges) show up with opposite signs for the two elements (positive/negative, see Fig. 2c), as mathematically required to induce composition variation (or different I_{Au}/I_{Ag} ratio values).

As a rule of thumb, some components beyond the scree-plot elbow (two to three additional components) are included in the reconstruction in order to include some remaining information carried by the first components associated with noise. All reconstructions in this work have included five components to render easy comparison between different experiments and simulations of binary alloy NPs (see Fig. 1).

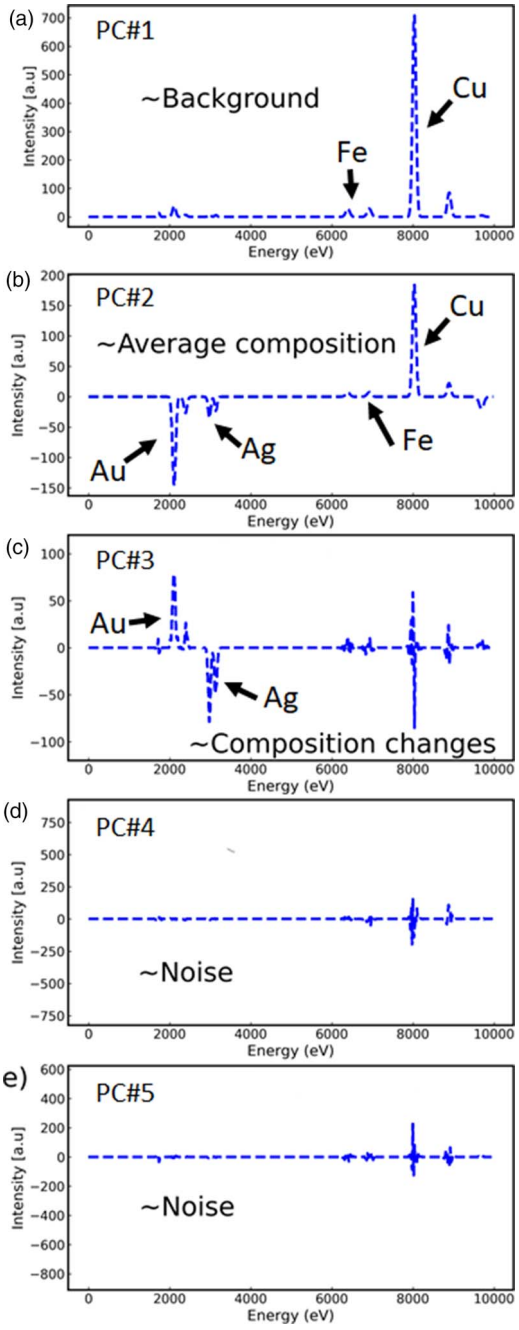


Fig. 2. Principal component profiles derived from a high SNR simulated HIS of bimetallic AuAg nanoparticles; for this ideal case, it is possible to make a clear identification of carried information. The loading in (**b**) includes Au and Ag peaks with the same signal (both negative, allowing the calculation of the NP average NP chemical composition); in contrast, the component in (**c**) shows Au and Ag peaks with opposite signal (negative–positive), then it is possible to induce concentration variations in the PCA reconstruction.

Simulated Datasets

Simulations offer irrefutable advantage against an experiment because the number, nature, and strength of latent factors can be fully controlled and the related effects followed in detail. Also, as the "true" object is precisely known, it is possible to compare the same object under situations of different noise levels.

It is critical that model spectra reproduce key features associated with measurements (background, energy calibration, and magnitude of the typical signals) which strongly influence PCA output. We must be aware that most of simulations do not include complex factors such as plural scattering (critical issue for EELS) or EDS sum/escape peaks, etc. In fact, several of these issues are consciously neglected because they might severely complicate interpretation between information content of the dataset and the PCA output).

Our simulations consider binary alloy NPs on a TEM grid and they include all other elemental EDS peaks related to the Cu grid and sample environment (Fe, Ni, etc.); the intensities of these peaks were scaled in relation to Au and Ag ones in agreement with our experimental data. After a noise-less spectra is calculated, random Poisson noise was added to data in order to yield a realistic model of experiments.

Once the total counts per NP to be considered in the simulation is defined, these counts are distributed inside pixels forming a circle (NP projection) in the simulated HIS. It is important to emphasize that counts distribution inside the NP associated pixels

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

considers the spherical shape of the NPs, such that counts and SNR diminish gradually from pixels located from center to surface (thickness variation along electron path). As a final step, we add the random Poisson noise to the voxels X-ray intensity. This issue is essential for a faithful description of experiments, as it induces variation in the overall EDS SNR for each NP pixel, which will strongly influence the PCA output.

Results

Figure 3 shows an example of a typical EDS HSI data from AuAg nanoparticles taken in an uncorrected 200 kV FEG TEM/STEM (described in the text as Experiment I); the spectrum in Figure 2a represents a good illustration of counting level that can be achieved. On average, we observe $I_{Au} \sim 380$ total counts for a ~6 nm NP, resulting in an intensity relative error of approximately $\Delta I/I = 5\%$ considering Poisson noise ($\sigma_I = \sqrt{I} \sim 19$ cts).

For these experiments, we have decided to apply the PCA processing to each HSI individually, because, in many cases, the HSIs were acquired several weeks apart. Grouping several HSIs may be useful to exploit additional redundancy, but noise components may vary between images. For example, a thicker substrate or the change of grid contribution due to a scan located closer to a grid bar may induce the appearance of additional meaningless principal components. These issues were not significant in our experiments (usually a square scan, ~30 nm side), but a careful analysis of EDS intensity showed that the beam current of the used microscope (more than 10 years old) was varying significantly. Then, each scan showed a slightly different SNR, which hindered their grouping into a big dataset and, analyzed by a single PCA processing step. For data acquired with a modern AC, TEM/STEM stability and reproducibility were not an issue and the dataset could be built by the stacking of several (64 × 64 pixels) HSI.

After PCA reconstruction, the automatic numerical recognition of smaller NPs on the chemical maps has been more efficient; this is even more pronounced for very small particles, 2–3 nm in diameter (Figs. 3e, 3f). PCA treatment generates a scree-plot where two components display much higher variance than noise (see Fig. 1); as expected, these two loadings carry information about the sample support and the average NP spectrum, respectively. In Figure 4a, raw and PCA-derived composition estimations (C_{Au}) of single NPs are distributed around the expected mean composition value. The analysis of composition distributions histograms at the right side of Figure 4a yields a standard deviation (STD) of $\Sigma_{C_{Au,den}} \sim 0.022$ versus $\Sigma_{C_{Au,raw}} \sim 0.05$ (hereafter, symbol Σ will be used to describe the calculated STD of a distribution $\Sigma^2_{STD} = \sum (x - \bar{x})^2/(n-1)$). The width of the calculated concentration distribution obtained after PCA denoising is much narrower than derived from raw data. This comparison indicates a strongly biased PCA reconstruction: due to the rather low SNR, PCA processing has been unable to rank properly the information on intensity variations between NPs (expected to be contained in the third principal component, Fig. 2c; Potapov & Lubk, 2019). Then, the ratio (I_{Au}/I_{Ag}) and the chemical composition derived from it is almost completely fixed and determined by the second principal component (representing the average NP spectrum). This generates a strong correlation between I_{Au} and I_{Ag} intensities significantly narrows the chemical composition distribution calculated for the ensemble of NPs. It is important to note that PCA-treated points are displayed without error bars, because there is no accepted and well-defined procedure for calculating confidence intervals after denoising [this issue will be addressed in a forthcoming paper associated with Part II of this work (Moreira, 2021)].

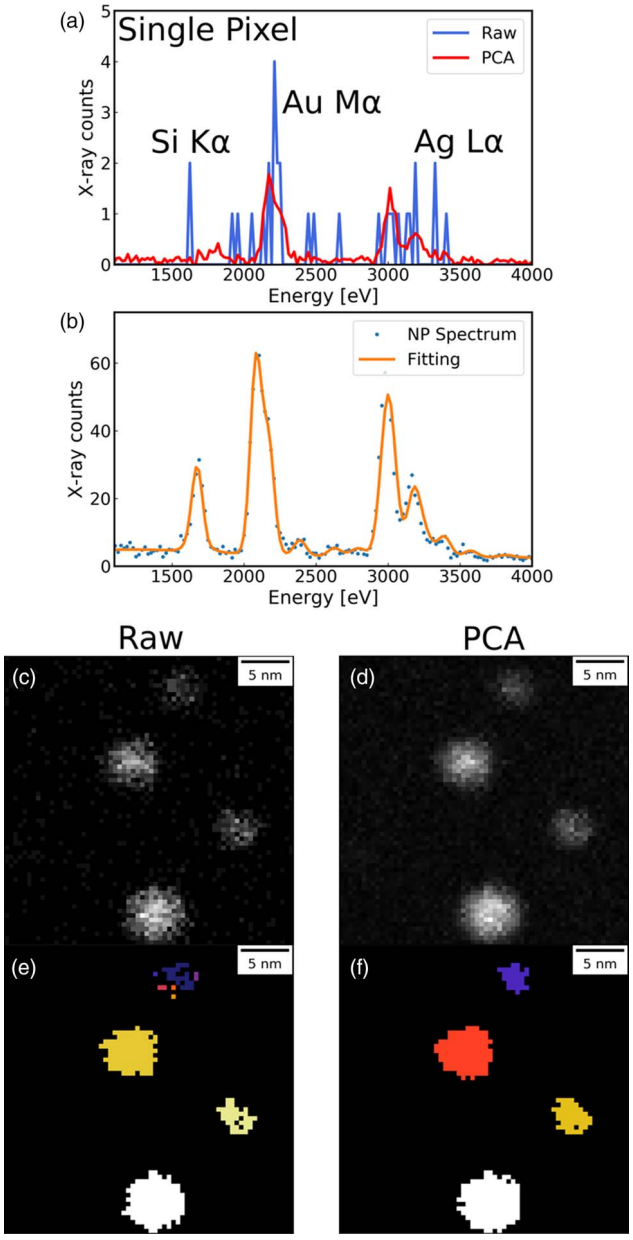


Fig. 3. Typical EDS HSI data from AuAg alloy nanoparticles (uncorrected FEG-STEM, 200 KV, 64 × 64 pixels scan). Comparison of raw and denoised EDS spectrum from a ~6 nm NP for a single pixel (**a**) and the entire NP (**b**) where the orange curve represents the final fit of EDS processing derived from hyperspy software used to integrate the peak intensity. (**c**,**d**) Raw and denoised Au chemical maps, respectively; the corresponding NP identification is displayed in (**e**) and (**f**), respectively (NP pixels are presented in different colors render easy the nanoparticles recognition). Note the increase in compactness of detected particles after PCA processing.

To test the above interpretation, we have built a dataset containing a bimodal distribution of chemical compositions by stacking a series of 4 HSI measured with the same microscope on two different NP samples (A-2x, B-2x, C_{Au}~0.50, and ~0.75 Au, respectively, hereafter noted Experiment II). Using raw data, chemical analysis can distinguish the separate composition distributions (see Fig. 4b, the calculated composition uncertainty bar was estimated to be $\sigma_{C_{Au}} \sim 0.04$ or 4 at%). After PCA treatment,

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

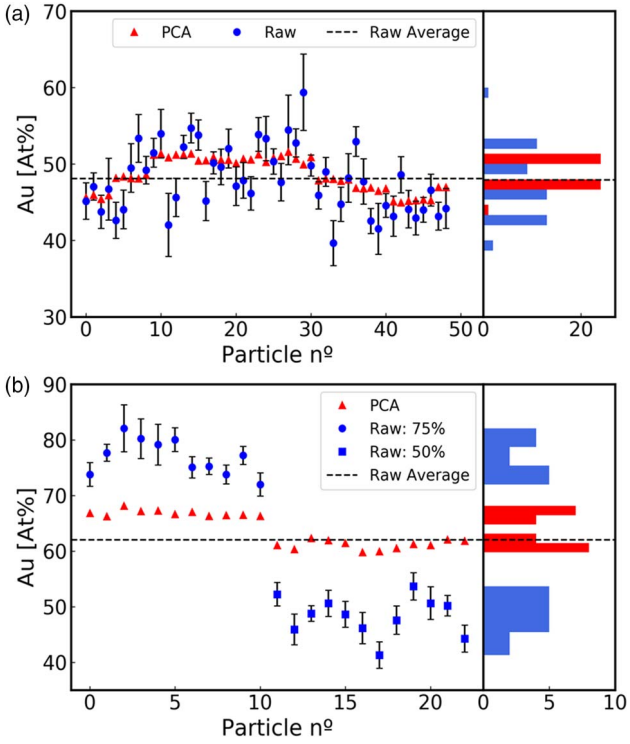


Fig. 4. Quantitative chemical composition analysis of alloy NPs. (**a**) Au atomic concentration derived from Exp. I using either raw or PCA denoised HIS (dashed line indicates average composition). (**b**) Au atomic concentration from Exp. II, including a bimodal chemical composition distribution (atomic concentration C_{Au}~0.5 and ~0.75). Resulting chemical composition distributions are plotted at the right. The two populations are easily identified from raw data, while denoising generates a strongly biased result and NP composition is close to the average of the ensemble of NPs.

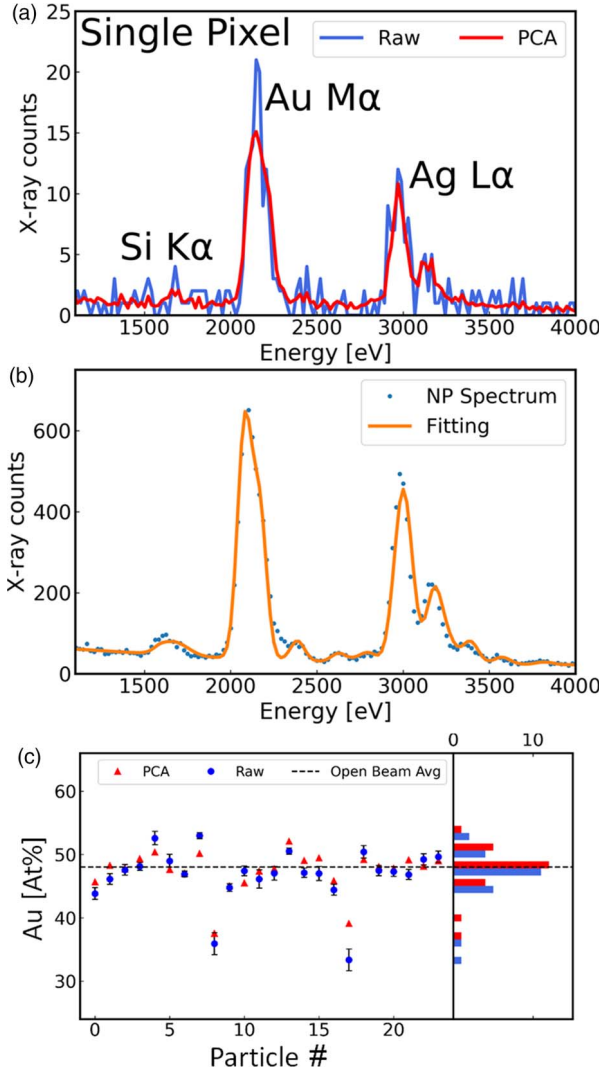


Fig. 5. Typical EDS HSI data from AuAg alloy nanoparticles obtained for experiment III (AC STEM, 4 SDD detectors, 80 kV, 64 × 64 pixels scan). Comparison of raw and denoised EDS spectrum from a ~6 nm NP for a single pixel (**a**) and the entire NP (**b**) where the orange curve represents the final fit of EDS processing derived from hyperSpy software used to integrate the peak intensity. (**c**) Comparison of chemical composition calculations derived using raw and denoised data; resulting chemical composition distribution is plotted at the right.

the chemical composition distribution from Exp. II appears significantly different from raw data: it is quite narrow and centered around the average composition of the two NP populations. The results derived from Exp. I and Exp. II indicate that this SNR level merely allows the reliable determination of *average* ensemble chemical composition, and information on individual NP intensity remains indistinguishable from noise (variations of I_{Au}/I_{Ag} should carried by the third principal component for a high SNR dataset, see Fig. 4). This represents a clear example of *random noise* PCA bias, showing that insufficient counting may yield unrealistic results. If this PCA reconstructed dataset is analyzed with hubris, serious misinterpretations may follow about the chemical homogeneity of the analyzed sample.

The logical solution to make further progress is to increase the SNR by collecting more X-ray counts. Reducing the incident beam energy from 200 to 80 keV increases the ionization cross-section (and consequently the EDS signal) by a factor of 1.6× considering relativistic estimations (Zaluzec, 2019). We have performed experiments at 80 keV using an aberration-corrected (AC) STEM microscope equipped with four EDS detectors (named Experiment III). A significant counting improvement has been attained, leading to a total integrated count of $I_{Au} = 1$, 950 counts for a 6 nm NP (see Fig. 5). Considering noise, the obtained ~5-fold gain in counts, just leads to a twofold decrease of intensity relative error $\Delta I/I \sim 2\%$; which clearly exemplifies how difficult and complex is to increase SNR in data subjected to Poisson statistics. For this experiment, the raw dataset was built by grouping 4× acquisitions of 64 × 64 HIS. PCA treatment of Experiment III shows a substantial improvement of the reconstruction; Figure 5c shows reasonably good agreement between chemical composition derived from raw and denoised data, indicating a strong reduction of *random noise* bias. The majority of measured NPs show similar composition within experimental confidence level (the measured composition distribution show similar widths as evidenced by the calculated standard deviations $\Sigma_{C_{Au,raw}} = 0.043$ and $\Sigma_{C_{Au,den}} = 0.031$). This amelioration is mainly associated with the better retrievability of a third PCA component carrying information on individual NP intensity variations due to chemical composition variability (note the slight improvement of the third component ranking in Fig. 1b). Only few NPs display significant differences with raw data, suggesting the occurrence of some biasing; in fact, these points correspond to the smallest detected NPs (i.e., with smaller count value per NP, ex. $I_{Au} \sim$ 400 cts, diameter ~3 nm).

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

Discussion

Experiment III provides evidence that it is essential to obtain measurements with counting level (or SNR) exceeding a certain minimal value to allow the minimization of *random noise* bias when analyzing alloy NP samples displaying a chemical composition variation. This allows the reliable ranking of physically significant PCA components and produces good fidelity denoised PCA reconstruction (Fig. 5c).

The comparison of Experiments I and III shows that PCA may allow reliable denoising, but it cannot recover information that is not actually carried by the combination of data SNR and redundancy (dataset size); this property has been named "EDS retrievability limit" (Potapov & Lubk, 2019). The observed bias is associated with inefficient retrieval of the "true" NP intensity variations, an information carried by the third principal component (see Figs. 1, 2).

To advance further, it is important to develop criteria/expressions to gather predictions of potential PCA random noise biasing effects directly from the measured data (counts value, data matrix size, etc.). Looking carefully at our results, we may note that the comparison of composition histograms of raw and denoised data reveals the different quality of data reconstructions from Exp. I and Exp. III (Figs. 4a, 5c). Then, it is interesting to suggest that the ratio of calculated composition distribution STDs or widths ($R_w = \Sigma_{C_{\mathrm{Au,den}}}/\Sigma_{C_{\mathrm{Au,raw}}}$) represents a good assessment of reconstruction fidelity and a potential indicator of the occurrence of random noise biasing.

Many studies have addressed the effect of random measurement noise on PCA bias (Faber et al., 1995*a*, 1995*b*; Malinowski, 2002; Nadler, 2008, 2009) when noise is assumed to be uncorrelated and homoscedastic (all its random variables have the same finite variance, σ_H^2). In contrast, our experiments involve Poisson noise, where the variance is dependent on the value of the particular voxel intensity ($\sigma^2 = I$). Then, the straightforward application of several bias indicators must be performed with extreme care to predict potential biasing (i.e., displacement from "true solutions"). Additional complications to estimate random noise variance arise from the high sparsity of EDS HSI of NPs (Keenan & Kotula, 2004) (many spectral energy channels include zero or just a few counts). For example, only 10–15% of pixels in our 64 × 64 scans carry X-ray counts from the particles. This leads to an average counting value per voxel ($I_{\mathrm{ave,voxel}}$) that may be <<1 by several orders of magnitude. Then, an estimation of data Poisson noise from mean voxel intensity ($\sigma = \sqrt{I_{\mathrm{ave,voxel}}}$) is nonfunctional, because noise becomes always larger than the measurement itself. From another angle, the variance stabilization through data scaling of EDS spectra before PCA processing (Keenan & Kotula, 2004; Kotula & Keenan, 2006; Kotula & Van Benthem, 2015) may induce additional significant error if sparsity is present.

Several studies (Verbeeck & Van Aert, 2004; Cueva et al., 2012; Lichtert & Verbeeck, 2013; Potapov, 2017; Spiegelberg & Rusz, 2017) have analyzed PCA eigenvalues bias using derivations of Faber et al. (1995*a*, 1995*b*), where the eigenvalue shift contains two terms. The first one represents an additive contribution directly related to random noise variance, and the second term depends on the number of pixels m and spectral channels n. Furthermore, combining Fisher's information concept and the Cramérs-Rao inequality, Verbeeck et al. (2004) and Lichtert & Verbeeck (2013) have been able to estimate a lower bound for systematical errors occurring in PCA reconstructed EELS data. This estimation requires the evaluation of a homoscedastic-like variance from the average count of the HSI voxels; as mentioned above, the sparsity of EDS HSIs strongly limits the applicability of these approaches.

In this work, we have explored the bias estimator derived from the so-called *spiked covariance model* (Nadler, 2008, 2009). This author has revealed the existence of an abrupt information-loss phase transition as a function of noise variance, leading to a well-defined bias estimator [E_{IL}, equation (2)]; a bias-free PCA reconstruction is always achieved when the estimator is less that a threshold value (T_{IL}). This bias estimator displays a simple mathematical expression describing the interaction between dataset size (n,m) and the ratio of experimental noise (variance σ^2) and *noise-free data variance* (v^{*2}, hereafter * will indicate that the variable is the *true noise-free data value*). It is important to emphasize that "true" parameters are seldom accessible to the experimentalist, so the elaboration of predictions is somewhat challenging. In the limit n,m tending to infinity, the threshold becomes $T_{\mathrm{IL}} = 1$ and the criterion becomes $E_{\mathrm{IL}} < 1$. The behavior of estimator has already been explored both in EELS and EDS HSI processing (Lichtert & Verbeeck, 2013; Potapov, 2016; Potapov & Lubk, 2019).

$$E_{\mathrm{IL}} = \frac{n}{m}\left(\frac{\sigma^2}{v^{*2}}\right)^2. \tag{2}$$

For our EDS experiments, the calculation of the E_{IL} using the average count per voxel yields very high values (10^7–10^{10}). It is essential to keep in mind that the interpretation of bias estimators must be analyzed in the context of the particular experimental constrains and targeted scientific problem. The EDS determination of NP elemental compositions requires the total number of counts per particle (I_{NP}), instead of the value of the individual spectral channels (or voxels) themselves. It is, therefore, crucial to identify a procedure to correctly evaluate the ratio of counting noise and *true data* variances (σ^2/v^{*2}) considering how it affects the information we are looking for (the assessment of composition variation between NPs), overcoming issues introduced by the intrinsic sparsity of the data.

For example, let's consider a NP generating a Au-M line total counts of $I_{\mathrm{Au,NP}}$; we will assume that these counts are homogeneously distributed over the number of pixels (m_{NP}) inside the particle region (for a scanning step 0.5 nm, this correspond to ~110 pixels for a 6 nm NP). Then, we distribute the estimated pixel counts ($I_{\mathrm{NP}}/m_{\mathrm{NP}}$) homogeneously on the spectral channels (n_{Peak}) under the EDS characteristic peak (for EDS peak ~200 eV wide and a spectrum channel width of 20 eV, we obtain $n_{\mathrm{Peak}} = 10$). The Poisson noise variance for voxels under the EDS peak is $\sigma^2_{I_{\mathrm{Au,vox}}} = I_{\mathrm{Au,NP}}/(m_{\mathrm{NP}} n_{\mathrm{Peak}})$. In fact, even taking this simple estimation of voxel intensity, it yields a value <1 for Exp. I, indicating strong sparsity. Only for the count values in Exp. III, the average total cts/voxel results slightly >1 and the counting error Poisson statistic (square root) can now be applied, always resulting in noise < measurement.

An essential parameter to evaluate the E_{IL} bias estimator is the noiseless data variance, so we must estimate how chemical composition dispersion influences the *noise-free* voxel X-ray count values. Let's start with a rather simple model case, where the NP ensemble displays a Gaussian distribution of chemical composition (mean $C_{\mathrm{Au,0}}$, variance $\Sigma^2_{C_{\mathrm{Au}}}$). To go further, we must deduce how to convert the Au composition standard deviation $\Sigma_{C_{\mathrm{Au}}}$ into

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

Table 1. Experimental Counts and Bias Estimators Evaluated for the Experiments I–III, Considering 6-nm NP, and a "True" Composition Dispersion $\nu^*_{C_{Au}}$=0.043.

	$I_{Au,NP}$ (cts)	$I_{Au,vox}$ (cts)	Bias (Qualitative)	$\sigma_{I_{Au,vox}}$ (cts)	$\nu^{\dagger}_{I_{Au,vox}}$ (cts)	Fidelity (R_w)	Information Loss (E_{IL})
Experiment I	380	0.35	Strong	0.59	0.029	0.44	19,100
Experiment II	293	0.37	Strong	0.61	0.079	0.20	1,460
Experiment III	1,950	1.77	Moderate/Low	1.33	0.15	0.73	180

a characteristic EDS intensity variation or the number of Au counts ($\Delta I_{Au,NP}$).

Chemical compositions are derived from EDS intensity using the Cliff-Lorimer expression in [equation (1)] (Cliff & Lorimer, 1975). Considering the normalization of a binary AuAg alloy compositions ($C_{Au} + C_{Ag} = 1$), we can relate C_{Au} to the EDS signal from Au atom:

$$C_{Au}\left(1 + \frac{I_{Ag}}{k_{AuAg} I_{Au}}\right) = 1. \quad (3)$$

As our experimental values for the Cliff-Lorimer factor have been determined to be very close to 1, we will assume that $k_{AuAg} = 1$, or in other terms that Au and Ag generate the same quantity of characteristic photons per atom. This approximation renders the calculation simpler for a binary alloy; if we consider that all particles have the same number of atoms, a change in the number of atoms of one element, reduces (or increases) its characteristic EDS peak by the same amount of count that increase (or reduces) the other element peak. In other words, the total number of characteristic X-ray counts generated by the NP is constant [$I_{Tot,NP} = (I_{Au} + I_{Ag}) = cnst$]. For $k_{AuAg} = 1$, equation (3) may be rewritten as $C_{Au}(I_{Au} + I_{Ag}) = I_{Au}$ or $C_{Ag}(I_{Au} + I_{Ag}) = I_{Ag}$. Then, we can deduce that

$$I_{Au} = C_{Au}(I_{Tot,NP}) \text{ and } I_{Ag} = C_{Ag}(I_{Tot,NP}). \quad (4)$$

Let's assume an NP population of average chemical composition $C_{Au,0}$ (showing total counts $I_{Au,0}$ and ratio ($I_{Au,0}/I_{Ag,0}$) at this concentration) and an STD $\Sigma_{C_{Au}}$. We can derive how many counts ($\Delta I_{Au,\sigma}$) must be added to $I_{Au,0}$ in order to get an NP with a new concentration $C_{Au,\sigma} = C_{Au,0} + \Sigma_{C_{Au}}$. Due to the direct proportionality relation of equation (4), a composition change of $\Sigma_{C_{Au}}$ is induced from an intensity change of $\Delta I_{Au} = I_{Tot,NP}\Sigma_{C_{Au}}$.

Next, we will use these approximations using experimental values; for example, let's take $C_{Au,0} = 0.5$ (50% Au atoms content), then $C_{Au,0}/C_{Ag,0} = 1$ (or $I_{Au,0}/I_{Ag,0} = 1$). A 6-nm diameter NP from Experiment I, should show a total EDS signal ($I_{Au} + I_{Ag}$) = $I_{Tot,NP} = 2\ I_{Au,0} = 760$ cts. In our STEM HSI, a particle of 6 nm in diameter contains ~110 pixels. Then, each image pixel in the NP image must contain on average $I_{Au,pixel}$ ~380/110~3 cts/pixel. These counts are distributed in the EDS peak, which we can consider with ~200 eV at the base (or 10 energy channels, if $\Delta E = 20$ eV); then, we finally arrive at $I_{Au,voxel} \sim 380/(110 \times 10) \sim 0.34$ cts per voxel in average, considering all pixels and energy channels in the HSI region containing the NP. Finally, the Poisson noise of each voxel is obtained by taking the square root $\sigma_{voxel,Poisson} = \sqrt{I_{voxel}} \sim 0.6$.

Assuming a "true" chemical composition distribution with standard deviation $\Sigma^*_{C_{Au}}$=0.04, then $\Delta I_{Au,NP} = I_{Tot,NP}\sigma_{C_{Au}} = 760 \times 0.04 = 30$ cts for a single nanoparticle. This value will be considered the "true" sample composition dispersion expressed in "true" *counts* variation per NP, $\nu^*_{I_{Au,NP}}$. Subsequently, we distribute these counts between the NP pixels and energy channels under EDS peak to get "true" *noise-free* intensity variations in the sample $\nu^{\dagger}_{I_{Au,vox}} = \Delta I_{Au,NP}/(m_{NP} n_{Peak})$ (as made above for NP intensity). This yields the "true" voxel intensity STD associated with the composition dispersion in the NP ensemble [$\nu^{\dagger}_{Au,vox} \sim 30/(110 \times 10) \sim 0.029$ cts].

These noise and *true noise-free* variances estimations ($\sigma^2_{I_{Au,vox}}$ and $\nu^{\dagger 2}_{I_{Au,vox}}$) yield a robust assessment of the effect of counting noise and how chemical composition dispersion influences the true noiseless data variance of EDS intensity, and they can be used to evaluate the information loss criterion and analyze PCA bias in our experimental studies. Hereafter, we will use these noise values in all evaluations of information-loss bias estimators and redundancy parameters (m, n) will be derived from the HSI matrix size. Taking as basis our best experimental results, we will consider the dispersion measured from Exp. III (our best measurement) as the "true" distribution concentration variance ($\nu^*_{C_{Au}} = \Sigma_{C_{Au},Exp.III} = 0.043$, Fig. 5c).

Table 1 shows main parameters associated with the evaluation of counting levels (per NP and per voxel), the ratio of raw and denoised compositions distribution widths (R_w), and E_{IL} bias estimator for our experiments. Although Experiment III shows a counting noise ($\sigma_{I_{Au,vox}} \sim 1.3$) which is about 10 times the "true" standard deviation ($\nu^{\dagger}_{I_{Au,vox}}$) due to composition dispersion, PCA exploits redundancy and it is able to obtain a PCA reconstruction conserving reasonably well the raw composition distribution (see Fidelity column, R_w).

It is very encouraging to see that denoising has preserved the small EDS signal variations necessary to describe concentration dispersion after PCA processing of Experiment III. Nonetheless, this contradicts the value of the information loss criterion value ($E_{IL} = 180$), which is much higher than the expected information loss threshold value $T_{IL} = 1$ (Nadler, 2008). In order to deepen our understanding of bias, we will analyze PCA reconstructions from a series of simulated HSI, where it is possible to consider a series of conditions in such a way that different SNR situations can be distinguished by comparing different NP "true" composition dispersions (variances $\nu^{*2}_{C_{Au}}$) versus total X-ray counting.

Firstly, we will consider a bimetallic NP system with a Gaussian chemical composition distribution (noiseless composition standard deviation of $\nu^*_{C_{Au}} = 0.02$). Figure 6 compares the distribution of chemical compositions (C_{Au}) for simulated HSI using raw and PCA reconstructed data. These simulations show two distinct behaviors; a strong bias occurs in the first case (low fidelity reconstruction, $I_{Au,NP} = 1{,}780$ cts, ~Experiment III counts) where the calculated composition distributions is much narrower after denoising ($\Sigma_{C_{Au},den} \sim 0.0002$, see histogram at Fig. 6a, right plot). The second case ($I_{Au,NP} = 7{,}100$ cts) seems to yield a much better output, although some narrowing of the chemical composition distribution of the NP ensemble is still observable

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

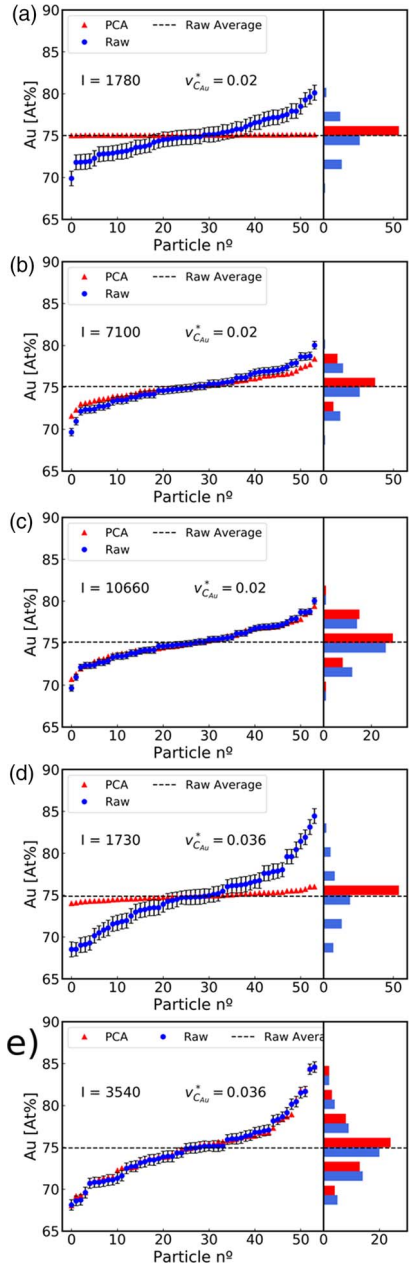


Fig. 6. Effect of denoising on the assessment of chemical composition distribution deduced from simulated datasets including different "true" composition dispersion and counting levels. (**a–c**) NP sample of average composition $C_{Au} = 0.75$ and distribution width $v^*_{C_{Au},STD} = 0.02$. (**d,e**) NP sample of average composition $C_{Au,1} = 0.75$ and a wider distribution $v^*_{C_{Au},STD} = 0.036$ (see text for explanations). The NPs are displayed following the increasing Au concentration along the horizontal axis, allowing an easy visualization of the composition distribution (histograms are displayed along the vertical direction on the right).

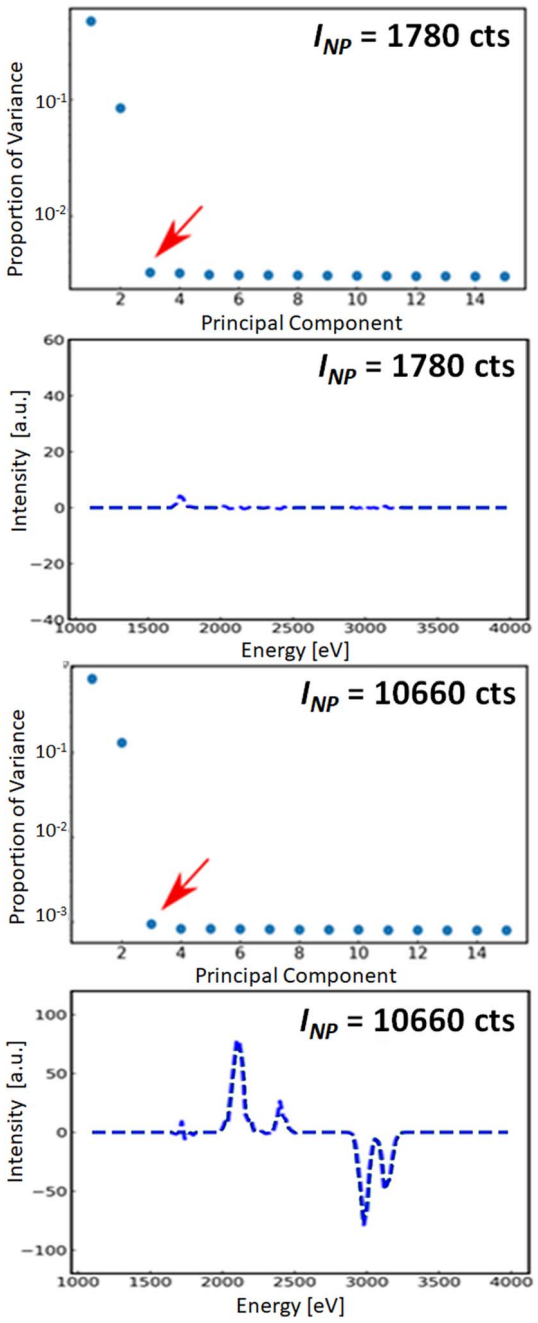


Fig. 7. Scree-plot and third principal component profile derived from the simulated HIS shown in Figures 6a and 6c. Note that for low counting level (1,780 cts, (**a**)) the profile of principal component #3 (arrowed) does not carry any information of chemical composition variation (characteristic peaks from Au and Ag are not present (**c**)). In contrast, the low bias (or higher cts) reconstruction of simulation from Figure 6c is associated with a third principal component with well-formed Au and Ag peaks of opposite sign (**d**) in order to induce changes of the I_{Au}/I_{Ag} ratio that determines NP chemical composition.

($\Sigma_{C_{Au},den} = 0.014$, Fig. 6b), indicating that this counting level is still insufficient to fully retrieve composition dispersion information. A very good recovery of the true composition distribution width [$\Sigma_{Au,den} = 0.019$, Fig. 6c requires a further increase of Au counts per NP (10,660 cts, ~5-fold increase on Exp. III counts)]. It is interesting to note the scree-plots from simulations in Figures 6a and 6c (high- and low-biased reconstructions) do not apparently show significant differences (Figs. 7a, 7b). For the strongly biased simulation displayed in Figure 6a ($I_{Au,NP} \sim 1{,}780$ cts), the third principal component is mostly flat in the region of Au and Ag EDS signal (Fig. 7c), generating no information of intensity variation associated with composition dispersion. In contrast, the low bias PCA reconstruction (Fig. 6c) shows that the high SNR ($I_{Au,NP} \sim 11{,}160$ cts) generates a third component (displaying well-defined Au and Ag EDS peaks of opposite sign). It is surprising that this very good retrieval of the third component is

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

associated with a singular value in the scree-plot which is very close to the noise-related components (see scree-plot in Fig. 7b)

By increasing the noiseless composition distribution width to $\nu^*_{C_{Au}} = 0.036$ (Figs. 6d, 6e), we can analyze bias occurrence as a function of SNR and true composition dispersion. As expected, it is possible to get a high-quality reconstruction with much less counts, due to the larger "true" *variance* of intensity associated to the much wider composition variability. In analogy to the precedent case, a counting level of $I_{Au,NP} = 1{,}730$ cts produces significantly low fidelity and biased reconstruction ($\Sigma_{C_{Au},den} = 0.005$), but just a twofold increase of counts ($I_{Au,NP} = 3{,}540$ cts) allows an excellent recovery of underlying "true" variability (i.e., accuracy) of NP composition distribution ($\Sigma_{C_{Au},den} = 0.035$). This total counts value is experimentally achievable, as it just represents less than a twofold increase of Exp. III counts.

Finally, we have also used simulations considering a bimodal true chemical composition distribution. We have considered two particle populations with composition averages around $C_{Au,1} = 0.75$ and $C_{Au,2} = 0.65$, each with $\nu^*_{C_{Au},STD} = 0.02$ (Fig. 8). A strong bias is observed for rather low counting ($I_{Au,NP} =$ 650 cts, Fig. 8a) and the existence of two populations remains undetected, the composition distribution is narrowly packed around the average elemental composition (equivalent to Experiment II). After increasing counting ($I_{Au,NP} = 1{,}650$ cts), PCA reconstruction reveals the existence of two populations with a clear indication of certain concentration distribution for each one. A further fourfold increase in counts ($I_{Au,NP} =$ 6,600 cts) yields a reliable chemical composition assessment after denoising (see Fig. 8c). These very good results are obtained at lower counting levels than before (Fig. 6), because the sample exhibits a quite wide "effective true" dispersion for the sample considered as a whole ($\nu^*_{C_{Au},STD} = 0.053$, wrapping the bimodal distribution). Briefly, simulations indicate that PCA reconstructions may attain an excellent accuracy for chemical composition distribution estimation (shown in Figs. 6c, 6e, 8c), when the SNR exceeds a certain threshold value.

At this point, we must carefully look at the PCA calculations (ex. scree-plot) to see if it can gather additional information of bias occurrence that can guide the analysis of experimental studies. Our simulations suggest that even for a very light increase in the ranking of the third principal component in relation to noise component seems to increase significantly the accuracy of reconstructions (see scree-plots in Figs. 1b, 7d). As all PCA reconstructions have included five components, we will assume that the fifth principal component represents just noise (see scree-plot in Figs. 1, 7). Then, the ratio of singular values of third and fifth components ($R_\lambda = \lambda_3/\lambda_5$) may represent an interesting quantitative guideline to identify if PCA processing has been able to rank correctly the third component and detect chemical composition variability.

Table 2 displays the different bias estimators, discussed in this article, (R_w, R_λ, and E_{IL}) calculated for our simulations in Figure 6. The cases displaying the low bias reconstructions (Figs. 6c, 6e) are correctly identified by R_w and R_λ showing values around 0.9 and 1.14, respectively. Also, the information-loss estimator E_{IL} shows the lowest values for the best PCA reconstruction, however the numerical values (~400) is much higher than that of the theoretical threshold value = 1 for an abrupt information loss event. Then, it is yet unclear if it can be used to predict the fidelity of NP composition assessment, even more considering that its evaluation requires the knowledge of noiseless data variance. For example, our estimation of E_{IL} from Exp. III yield a

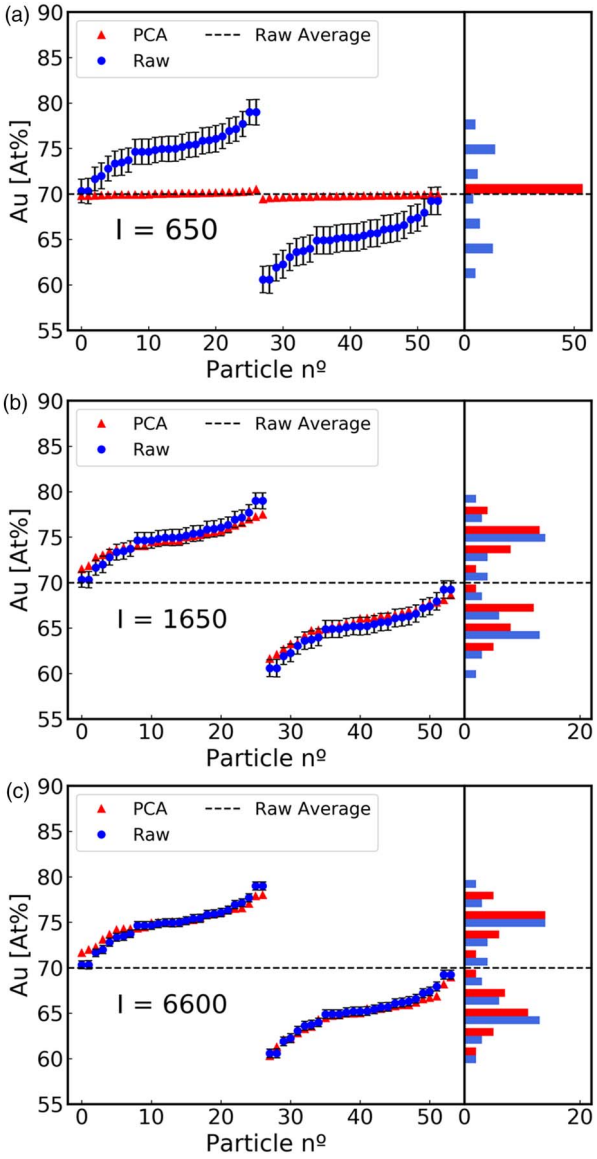


Fig. 8. Effect of denoising on the assessment of chemical composition distribution deduced from simulated datasets including bimodal "true" distributions of chemical composition (population 1: $C_{Au,1} = 0.75$, population 2: $C_{Au,2} = 0.65$, for both the distribution width is $\nu^*_{C_{Au},STD} = 0.02$). The calculated chemical composition distribution is plotted at the right (see text for explanations).

smaller value ($E_{IL} \sim 180$) than for very high signal SNR simulations ($E_{IL} \sim 400$); this is unexpected and it is probably due to our too optimistic guess of the "true" concentration variance (extracted from Fig. 5c). Both R_w and R_λ are readily accessible from measurements and data treatment without previous knowledge of "true" composition distribution, then their very good correlation with bias minimization suggest that they may be very useful for practical applications.

Figure 9 shows a plot of R_λ versus R_w for datasets (experiment and simulations) considered in this work. All cases displaying low bias are at the right top corner of the graph, allowing a direct assessment of PCA denoising performance. Then, for sample of binary alloy NPs with approximatively a Gaussian composition distribution, we suggest a criterion for obtaining a low bias PCA reconstructions: $R_\lambda \geq 1.15$ and $R_w \geq 0.9$; it is important to mention that both R_λ and R_w must exceed the minimal values.

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

Table 2. Bias Estimators Evaluated for the Simulated HSIs Shown in Figure 6.

	$\nu_{C_{Au}}$	$I_{Au,NP}$ (cts)	Bias (Qualitative)	R_w	R_λ	Information Loss E_{IL}
Simulation 1 (Fig. 6a)	0.02	1,780	Strong	0.01	1.043	15,400
Simulation 2 (Fig. 6b)	0.02	7,100	Moderate	0.70	1.035	970
Simulation 3 (Fig. 6c)	0.02	10,660	Low	0.90	1.138	430
Simulation 4 (Fig. 6d)	0.036	1,730	Strong	0.13	1.010	1,550
Simulation 5 (Fig. 6e)	0.036	3,540	Low	0.97	1.149	370

Different true concentration dispersion and counting levels are compared (ensemble of 54× 6-nm NPs, number spectral channels $n = 500$; image size $m = 64 \times 384 = 24{,}576$ pixels, scanning step 0.5 nm).

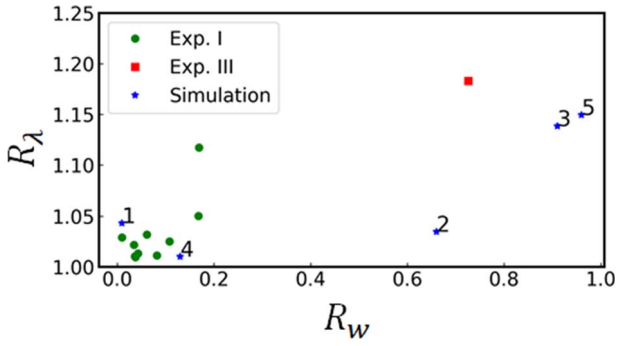


Fig. 9. Plot showing the correlation of proposed bias estimators R_w and R_λ for experiments I and III and simulations displayed in Figure 6 (index 1–5 indicated simulation describes in Figs. 6a–6f, respectively).

If the sample composition distribution is more complex and measurements are rather noisy (ex. bimodal distribution as in Exp. II and simulations of Fig. 7), the meaningful information may be distributed among more principal components. A strong random noise bias is immediately revealed by a low R_w; however, a proper understanding of information content carried by components may require additional simulations and a careful analysis of scree-plots and loading profiles before to define a different R_λ criterion.

Finally, we would like to provide a tool to roughly estimate EDS counts generated by NPs of different sizes and incident energy and detector geometries (remember that we have always used 200 ms dwell time, beam current ~400–500 pA). We have compiled the values of EDS counts per NP as a function of particle diameter for different instruments (uncorrected 200 kV and AC corrected 80 kV experiments) versus NP diameter; the experimental points were fitted with a third-order polynomial (see Supplementary Fig. S2). We expect that this plot may be exploited to allow a rough estimation of expected SNR for metal nanoparticle EDS experiments.

In this work, we have mainly addressed random noise bias, and not discussed *underfactoring bias* (reconstruction using an insufficient number of principal components) (Faber et al., 1995*a*, 1995*b*; Malinowski, 2002; Nadler, 2008, 2009), because the information content of each PCA principal component can be understood reasonably well for our sample. We would like to emphasize that Potapov & Lubk have reported an interesting tool to identify if the PCA component contains information or just carry noise (Potapov & Lubk, 2019) by plotting the scores for successive principal components: noisy ones generate isotropic round clouds, but information-carrying ones show a structured anisotropic cloud (see example in Supplementary Fig. S3).

Conclusion

Given the outstanding progress of TEMs instruments and associated detectors, it is essential to exploit their capabilities to target "quantitative" interpretation for image, diffraction, or spectroscopic data. PCA data processing is a very powerful tool to improve the SNR of hyperspectral images and extract all underlying information contained in the data. Our high spatial EDS experiments and simulations show that if the SNR exceeds a minimal threshold value, it is possible to avoid random noise biasing in PCA reconstructions and, then get chemical composition distribution with very good fidelity of true values (i.e., accuracy). In fact, the experimental requirements represent a compromise between dataset size (redundancy), counting level, and sample true composition variance.

We have suggested a few quantitative estimators that may provide information about random noise bias occurrence, and they can be derived for PCA processing output without knowledge of "true" noiseless sample properties. However, numerical simulations may be necessary to analyze the particular nanostructured system, in order to design properly data acquisition parameters and support the criteria used for the interpretation. It is important to emphasize that the physical behavior may vary dramatically between NP systems, so it is essential that simulations reproduce experiments as close as possible.

A natural question arises: how should one analyze measurements of bimodal distributions or three-element metal alloys? Which of the components will carry information on local differences? It is not easy to answer this question with a general applicable recipe for all kinds of samples. It is important to mention factors such as vibrations, scan noise or filament current oscillation may generate additional principal components in the scree-plot, and chemical information can become scattered into several components, so one must be prepared to carefully examine the scree-plots and information carried by principal component profiles.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S1431927621013933.

Funding. D.U. acknowledges financial support from the Brazilian Agencies FAPESP (No. 2014/01045-0), CNPq (402571/2016-9 and No. 306513/2017-0) and FAEPEX-UNICAMP. V.R. thanks funding from FAPESP (2007/01722-9) and CNPq (555647/2006-4 and 577046/2008-0). M.M. thanks funding from CNPq (No. 162541/2018-0). Access to the FEG-TEM/STEM from the Brazilian Nanotechnology National Laboratory is acknowledged (LNNANO, grant No. ME - 22329). M.A.C. acknowledges financial support from FAPESP (Nos. 2013/02300-1 and 2013/10957-0) and CNPq (No. 479486/ 2012-3). L.H.G.T., G.D., and C.D. acknowledge funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 312483 (ESTEEM2). M.H. acknowledges financial support from FAPESP (No. 2016/12807-4).

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

Conflict of interest. The authors declare no competing financial interest.

References

Alloyeau D, Mottet C & Ricolleau C (2012). *Nanoalloys: Synthesis, Structure and Properties*. London: Springer-Verlag.

Belter M, Sajnóg A & Barałkiewicz D (2014). Over a century of detection and quantification capabilities in analytical chemistry – Historical overview and trends. *Talanta* **129**, 606–616.

Binns C (2014). *Nanomagnetism: Fundamentals and Applications*, **vol. 6**. Amsterdam, Netherlands: Elsevier.

Braidy N, Jakubek ZJ, Simard B & Botton GA (2008). Quantitative energy dispersive X-ray microanalysis of electron beam-sensitive alloyed nanoparticles. *Microsc Microanal* **14**, 166–175.

Brown KA, Brittman S, Maccaferri N, Jariwala D & Celano U (2020). Machine learning in nanoscience: Big data at small scales. *Nano Lett* **20**, 2–10.

Burke MG, Watanabe M, Williams DB & Hyde JM (2006). Quantitative characterization of nanoprecipitates in irradiated low-alloy steels: Advances in the application of FEG-STEM quantitative microanalysis to real materials. *J Mater Sci* **41**, 4512–4522.

Cliff G & Lorimer GW (1975). The quantitative analysis of thin specimens. *J Microsc* **103**, 203–207.

Cueva P, Hovden R, Mundy JA, Xin HL & Muller DA (2012). Data processing for atomic resolution electron energy loss spectroscopy. *Microsc Microanal* **18**, 667–675.

Currie LA (1968). Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal Chem* **40**, 586–593.

Currie LA (1999). Detection and quantification limits: Origins and historical overview. *Anal Chim Acta* **391**, 127–134.

de la Peña F, Ostasevicius T, Fauske VT, Burdet P, Jokubauskas P, Nord M, Prestat E, Sarahan M, MacArthur KE, Johnstone DN, Taillon J, Caron J, Furnival T, Eljarrat A, Mazzucco S, Migunov V, Aarholt T, Walls M, Winkler F, Martineau B, Donval G, Hoglund ER, Alxneit I, Hjorth I, Zagonel LF, Garmannslund A, Gohlke C, Iyengar I & Chang H-W (2017). hyperspy/hyperspy: HyperSpy 1.3. doi:10.5281/ZENODO.583693.

de Sá ADT, Oiko VTA, di Domenicantonio G & Rodrigues V (2014). New experimental setup for metallic clusters production based on hollow cylindrical magnetron sputtering. *J Vac Sci Technol B* **32**, 061804.

Egerton RF (2012). Mechanisms of radiation damage in beam-sensitive specimens, for TEM accelerating voltages between 10 and 300 kV. *Microsc Res Tech* **75**, 1550–1556.

Egerton RF (2019). Radiation damage to organic and inorganic specimens in the TEM. *Micron* **119**, 72–87.

Egerton RF, McLeod R, Wang F & Malac M (2010). Basic questions related to electron-induced sputtering in the TEM. *Ultramicroscopy* **110**, 991–997.

Faber NM, Meinders MJ, Geladi P, Sjiistrijm M, Buydens LMC & Kateman G (1995*a*). Random error bias in principal component analysis. Part I. Derivation of theoretical predictions. *Anal Chim Acta* **304**, 257–271.

Faber NM, Meinders MJ, Geladi P, Sjiistrijm M, Buydens LMC & Kateman G (1995*b*). Random error bias in principal component analysis. Part II. Application of theoretical predictions to multivariate problems. *Anal Chim Acta* **304**, 273–283.

Ferrando R (2016). *Structure and Properties of Nanoalloys in Frontiers of Nanoscience*. Amsterdam: Elsevier. pp. 2–337.

Hawkes P & Spence JCH (2019). *Springer Handbook of Microscopy*. Switzerland: Springer.

Heiz U & Landman UE (2007). *Nanocatalysis*. Heidelberg, Berlin: Springer.

Jolliffe IT (2002). *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag.

Jolliffe IT & Cadima J (2016). Principal component analysis: A review and recent developments. *Philos Trans A* **374**, 20150202.

Jones L, Varambhia A, Beanland R, Kepaptsoglou D, Griffiths I, Ishizuka A, Azough F, Freer R, Ishizuka K, Cherns D, Ramasse QM, Lozano-Perez S & Nellist PD (2018). Managing dose-, damage- and data-rates in multi-frame spectrum-imaging. *Microscopy* **67**, 98–113.

Keenan MR & Kotula PG (2004). Accounting for poisson noise in the multivariate analysis of TOF-SIMS spectrum images. *Surf Interface Anal* **36**, 203–212.

Kotula PG & Keenan MR (2006). Application of multivariate statistical analysis to STEM X-ray spectral images: Interfacial analysis in microelectronics. *Microsc Microanal* **12**, 538–544.

Kotula PG & Van Benthem MH (2015). Revisiting noise scaling for multivariate statistical analysis. *Microsc Microanal* **21**, 1423–1424.

Lichtert S & Verbeeck J (2013). Statistical consequences of applying a PCA filter on EELS spectrum images. *Ultramicroscopy* **125**, 35–42.

Lyman CE, Lakis RE & Stenger Jr HG (1995). X-ray emission spectrometry of phase separation in PtRh nanoparticles for nitric oxide reduction. *Ultramicroscopy* **58**, 25–34.

Malinowski ER (2002). *Factor Analysis in Chemistry*, 3rd ed. New York: Wiley.

Moreira M, Hillenkamp M, Divitini G, Tizei LHG, Ducati C, Cotta MA, Rodrigues V & Ugarte D (2021). Improving quantitative EDS chemical analysis of alloy nanoparticles by PCA denoising: Part II, uncertainty intervals. Submitted.

Mukherjee P, Manchanda P, Kumar P, Zhou L, Kramer MJ, Kashyap A, Skomski R, Sellmyer D & Shield JE (2014). Size-induced chemical and magnetic ordering in individual Fe–Au nanoparticles. *ACS Nano* **8**, 8113–8120.

Mukherjee P, Zhang Y, Kramer MJ, Lewis LH & Shield JE (2012). $L1_0$ structure formation in slow-cooled Fe–Au nanoclusters. *Appl Phys Lett* **100**, 211911.

Nadler B (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann Stat* **36**, 2791–2817.

Nadler B (2009). Discussion. *J Am Stat Assoc* **104**, 694–697.

Odom TW & Schatz GC (2011). Introduction to plasmonics. *Chem Rev* **111**, 3667–3668.

Parish CM & Brewer LN (2010). Multivariate statistics applications in phase analysis of STEM-EDS spectrum images. *Ultramicroscopy* **110**, 134–143.

Pennycook SJ & Nellist PD (2011). *Scanning Transmission Electron Microscopy*. New York: Springer.

Potapov P (2016). Why principal component analysis of STEM spectrum images results in abstract, uninterpretable loadings? *Ultramicroscopy* **160**, 197–212.

Potapov P (2017). On the loss of information in PCA of spectrum-images. *Ultramicroscopy* **182**, 191–194.

Potapov P & Lubk A (2019). Optimal principal component analysis of STEM XEDS spectrum images. *Adv Struct Chem Imag* **5**, 4.

Rossouw D, Burdet P, de la Peña F, Ducati C, Knappett BR, Wheatley AEH & Midgley PA (2015). Multicomponent signal unmixing from nanoheterostructures: Overcoming the traditional challenges of nanoscale X-ray analysis via machine learning. *Nano Lett* **15**, 2716–2720.

Schlossmacher P, Klenov DO & Freitag B (2010). Enhanced detection sensitivity with a new windowless XEDS system for AEM based on silicon drift detector technology. *Microsc Today* **18**, 14–20.

Spiegelberg J & Rusz J (2017). Can we use PCA to detect small signals in noisy data? *Ultramicroscopy* **172**, 40–46.

Thomas JR, Leary RK, Eggeman AS & Midgley PA (2015). The rapidly changing face of electron microscopy. *Chem Phys Lett* **631–632**, 103–113.

Titchmarsh JM (1999). EDX spectrum modelling and multivariate analysis of sub-nanometer segregation. *Micron* **30**, 159–171.

Titchmarsh JM & Dumbill S (1996). Multivariate statistical analysis of FEG STEM EDX spectra. *J Microsc* **184**, 195–207.

van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E & Yu T (2014). The scikit-image contributors scikit-image: Image processing in python. *PeerJ* **2**, e453.

Verbeeck J & Van Aert S (2004). Model based quantification of EELS spectra. *Ultramicroscopy* **101**, 207–224.

Watanabe M, Kanno M & Okunishi E (2010). Atomic resolution elemental mapping by EELS and XEDS in aberration corrected STEM. *Jeol News* **45**, 8–12.

Williams DB & Carter CB (2009). *Transmission Electron Microscopy Part 1: Basics*, 2nd ed. New York: Springer.

Williams DB & Carter CB (2016). *Transmission Electron Microscopy: Diffraction, Imaging, and Spectroscopy*. New York: Springer.

Zaluzec NJ (2019). Improving the sensitivity of X-ray microanalysis in the analytical electron microscope. *Ultramicroscopy* **203**, 163–169.

Downloaded from https://www.cambridge.org/core. UNICAMP, on 17 Jan 2022 at 13:15:19, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms.
https://doi.org/10.1017/S1431927621013933

Improving Quantitative EDS Chemical Analysis of alloy Nanoparticles by PCA Denoising: Part I, reducing reconstruction bias.

Murilo Moreira[1], Matthias Hillenkamp[1,2], Giorgio Divitini[3], Luiz H. G. Tizei[4], Caterina Ducati[3], Monica A. Cotta[1],Varlei Rodrigues[1,*], Daniel Ugarte[1,*]

1. *Instituto de Fisica "Gleb Wataghin", Universidade Estadual de Campinas-UNICAMP, 13083-859, Campinas - SP, Brazil*
2. *Institute of Light and Matter, Université Claude Bernard Lyon 1, CNRS, UMR5306, F-69622 Villeurbanne, France*
3. *Department of Materials Science and Metallurgy, University of Cambridge, Cambridge CB3 0FS, United Kingdom*
4. *Laboratoire de Physique des Solides, Université Paris-Saclay, CNRS, 91405 Orsay, France*

***Correspondence and requests for materials should be addressed to:**
D.U. (email: dmugarte@ifi.unicamp.br) or to V.R. (email: varlei@ifi.unicamp.br)

Figure S1.

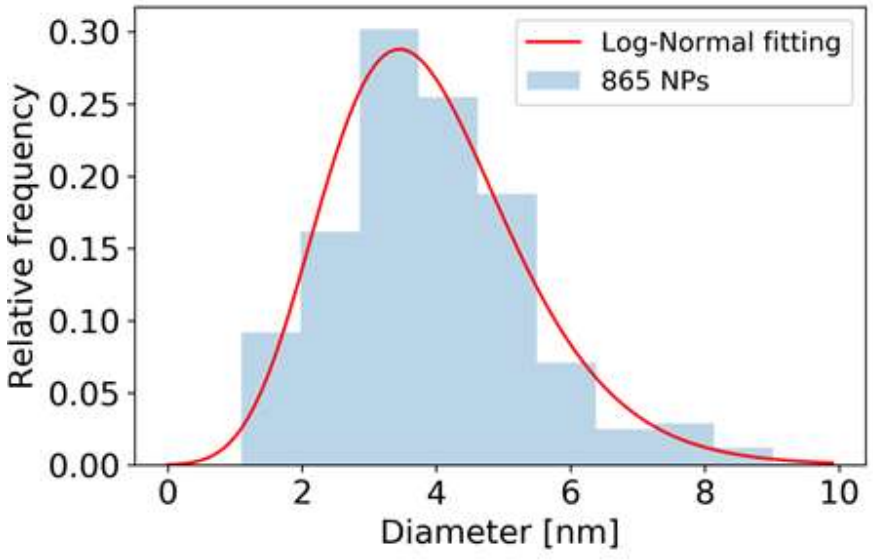


Figure S1. Typical size distribution of the alloy NP used in Experiment I, II and III as measured by TEM. The continuous curve represents a log-normal description of the distribution.

Figure S2.

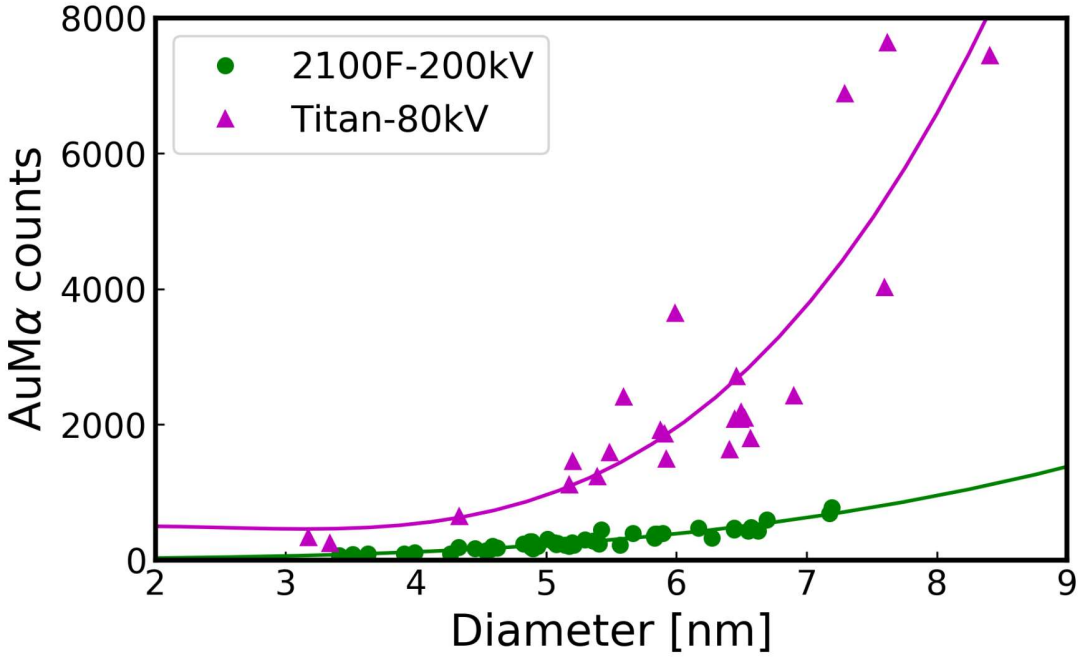


Figure S2. Integrated total counts per NP at the Au L_α peak (200 ms dwell time), as a function of NP diameter for different instruments. This allows a quick estimation of expected counts and Poisson noise for nanoparticle. A the mathematical expression describing the general tendency of the plots has been obtained by fitting the data. The observed counts as a function of NP diameter (D) can be estimated the equation below, derived from Figure S2.

for Titan-80kV, $I(D) = 38.3\ D^3 - 287.4\ D^2 + 670.7\ D$

for JEM2100F-200kV, $I(D) = 2.4\ D^3 - 6.9\ D^2 + 18.6\ D$

Figure S3.

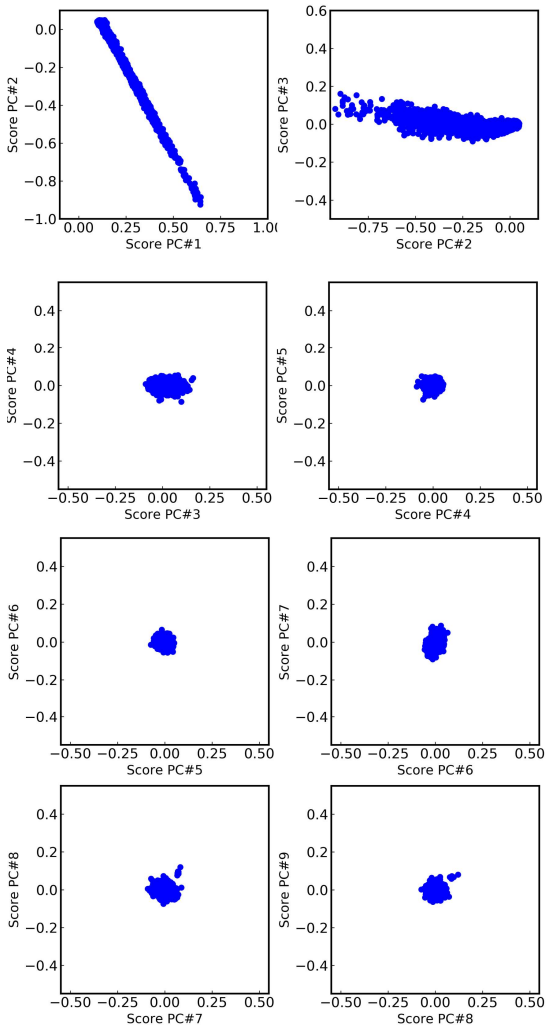


Figure S3. Cluster plot of scores for successive principal components obtained PCA processing a high SNR simulated EDS HSI of bimetallic NPs (Potapov, P., Lubk, A. (2019). Optimal principal component analysis of STEM XEDS spectrum images. *Adv. Struct. Chem. Imag.* ***5***, 4). Note that information-carrying component shows a structured anisotropic cloud (top), while noise components generate isotropic round clouds.

RightsLink®

Home | Help | Live Chat | Sign in | Create Account

Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part I, Reducing Reconstruction Bias

Author:
Murilo Moreira, Matthias Hillenkamp, Giorgio Divitini, Luiz H. G. Tizei, Caterina Ducati, Monica A. Cotta, Varlei Rodrigues, Daniel Ugarte

Publication: Microscopy and Microanalysis

Publisher: Cambridge University Press

Date: Jan 3, 2022

Copyright © Copyright © The Author(s), 2022. Published by Cambridge University Press on behalf of the Microscopy Society of America

License Not Required

Permission is granted at no cost for use of content in a Master's Thesis and/or Doctoral Dissertation. If you intend to distribute or sell your Master's Thesis/Doctoral Dissertation to the general public through print or website publication, please return to the previous page and select 'Republish in a Book/Journal' or 'Post on intranet/password-protected website' to complete your request.

BACK CLOSE

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Chapter 7

Exploiting Principal Component Analysis for denoising EM image spectra: Uncertainty intervals

This chapter contains the publication:
Murilo Moreira, Matthias Hillenkamp, Giorgio Divitini, Luiz H. G. Tizei, Caterina Ducati, Monica A. Cotta, Varlei Rodrigues and Daniel Ugarte
Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part I, Reducing Reconstruction Bias
Microscopy and Microanalysis, 2022, 28(2), 338-349. doi:10.1017/S1431927621013933

In the last chapter, we have demonstrated by using simulations and experiments that the effects of statistical bias in the PCA reconstruction can be mitigated by increasing signal-to-noise ratios. We used quantitative estimators exploring the information loss threshold, redundancy parameter, variance of the noise, and the variance of the information we want to measure. We showed that the variance associated with the chemical composition dispersion is mainly associated with the third component. When the components are corrupted by noise, it is required to satisfy certain conditions to extract quantitative and reliable information after PCA denoising. The third component contains the finest chemical composition information (the dispersion of single measurements). Even though the first two components associated with the average information can be easily recovered, this is not true for the third. These observations are consistent with the quantification limit for correct information retrieval. Even more in raw non-denoised data sets, where the dispersion follows the Poisson uncertainty directly, and the true information (size-dependent chemical composition) is hidden by the noisy distribution. After these observations, we may wonder if PCA is denoising our data set. Indeed this is predicted by the models used in the quantitative estimators proposed in the last chapter. However, here we wish to propose experiments and simulations to estimate the quantitative measurements' uncertainty reduction and demonstrate that

PCA denoising improves the EDS chemical composition quantification of nanoalloys.

7.1 Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part II, Uncertainty intervals

Microscopy and Microanalysis (2022), **28**, 723–731
doi:10.1017/S1431927622000551

Microscopy AND Microanalysis

Original Article

Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part II. Uncertainty Intervals

Murilo Moreira[1], Matthias Hillenkamp[1,2], Giorgio Divitini[3,4], Luiz H. G. Tizei[5], Caterina Ducati[3], Monica A. Cotta[1], Varlei Rodrigues[1] and Daniel Ugarte[1]*

[1]Instituto de Fisica "Gleb Wataghin", Universidade Estadual de Campinas-UNICAMP, 13083-859 Campinas, SP, Brazil; [2]Institute of Light and Matter, Université Claude Bernard Lyon 1, CNRS, UMR5306, F-69622 Villeurbanne, France; [3]Department of Materials Science and Metallurgy, University of Cambridge, Cambridge CB3 0FS, UK; [4]Electron Spectroscopy and Nanoscopy Group, Istituto Italiano di Tecnologia, via Morego 30, Genoa, Italy and [5]Laboratoire de Physique des Solides, Université Paris-Saclay, CNRS, 91405 Orsay, France

Abstract

Analytical studies of nanoparticles (NPs) are frequently based on huge datasets derived from hyperspectral images acquired using scanning transmission electron microscopy. These large datasets require machine learning computational tools to reduce dimensionality and extract relevant information. Principal component analysis (PCA) is a commonly used procedure to reconstruct information and generate a denoised dataset; however, several open questions remain regarding the accuracy and precision of reconstructions. Here, we use experiments and simulations to test the effect of PCA processing on data obtained from AuAg alloy NPs a few nanometers wide with different compositions. This study aims to address the reliability of chemical quantification after PCA processing. Our results show that the PCA treatment mitigates the contribution of Poisson noise and leads to better quantification, indicating that denoised results may be reliable from the point of view of both uncertainty and accuracy for properly planned experiments. However, the initial data need to be of sufficient quality: these results can only be obtained if the signal-to-noise ratio of input data exceeds a minimal value to avoid the occurrence of random noise bias in the PCA reconstructions.

Key words: denoising, energy-dispersive x-ray spectroscopy (EDS), nanoparticles, principal components analysis (PCA), quantitative chemical analysis

(Received 22 November 2021; revised 31 January 2022; accepted 31 March 2022)

Introduction

The electronic properties of metal alloy nanoparticles (NPs) show a complex behavior in relation to atomic or chemical structure. This is exploited for the optimization of either physical properties such as surface plasmons in plasmonic devices, or chemical reactivity/selectivity in catalysis (Heiz & Landman, 2007; Odom & Schatz, 2011; Alloyeau et al., 2012; Binns, 2014; Ferrando, 2016). The properties of NPs result from the interplay between chemical (composition) and physical (surface/volume ratio or NP size) features, so that a precise characterization requires the estimation of morphology and chemical composition as a function of size (Lyman et al., 1995).

The strong scientific interest in nanostructured materials has pushed transmission electron microscopy (TEM) to excellent performance, not only for spatial resolution and signal detection efficiency, but also for reproducibility and automation (Williams & Carter, 2009; Pennycook & Nellist, 2011; Carter & Williams, 2016; Hawkes & Spence, 2019). At present, scanning transmission electron microscopes (STEMs) are able to simultaneously record a complete analytical spectrum or a diffraction pattern at each image pixel, generating huge 3D or 4D data matrices containing 10^7–10^9 elements (usually denoted hyperspectral images, HSI). The full exploration of these outstanding analytical capabilities requires researchers to target "quantitative" TEM data interpretation for imaging, diffraction, or spectroscopic studies. This will allow a much deeper and robust approach to refine the physico-chemical understanding of nanomaterials. While quantitative results require significantly more effort in experimental optimization and data processing compared to qualitative studies, we must keep in mind that any data value (measured or calculated) without (a stated) uncertainty is of minor interest (Bevington & Robinson, 2003; Drosg, 2009; Hughes & Hase, 2010). The only approach to obtain meaningful, reproducible measurements is to follow rigorous error analysis, taking both precision and accuracy into account.

Maximizing information from such big data ensembles (HSIs) requires advanced statistical algorithms, such as machine learning (ML) tools, which exploit redundancies and can find non-apparent correlations in the data (Brown et al., 2020); in this way, they can extract information that would be inaccessible by traditional methods (Cueva et al., 2012; Carter & Williams, 2016; Brown et al., 2020). Principal component analysis (PCA; Jolliffe, 2002) is recognized as the most popular unsupervised

*Corresponding author: Daniel Ugarte, E-mail: dmugarte@ifi.unicamp.br

Cite this article: Moreira M, Hillenkamp M, Divitini G, Tizei LHG, Ducati C, Cotta MA, Rodrigues V, Ugarte D (2022) Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part II. Uncertainty Intervals. *Microsc Microanal* **28**, 723–731. doi:10.1017/S1431927622000551

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Microscopy Society of America

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

ML tool to reduce dimensionality. In favorable conditions, PCA decompositions can extract meaningful information and reduce the effects of noise. This procedure, usually described as "denoising", generates a new dataset with a higher signal-to-noise ratio (SNR) and values closer to the "true" ones. However, an essential point that is sometimes overlooked is that a PCA reconstruction of a noisy dataset may be strongly biased, that is, quite different from the "true" noiseless data (Faber et al., 1995a, 1995b; Malinowski, 2002; Nadler, 2008, 2009).

In the accompanying Part I of this work (Moreira et al., 2022), we have addressed the accuracy of energy-dispersive X-ray spectrometryy (EDS) chemical composition assessment from binary alloy NPs after PCA processing. To obtain an accurate denoised reconstruction, one requires adequate levels for experimental noise, redundancy (dataset size $m \times n$, where m is the number of pixels and n the number of spectral channels), and variance of the "noise-free" sample properties (Faber et al., 1995a, 1995b; Malinowski, 2002; Nadler, 2008, 2009). The next step to a complete error analysis is the evaluation of uncertainty intervals, which is addressed in the present manuscript. An intuitive understanding of noisy signals would suggest that "denoising" must lead to a reduction in the error bars for the outputs. In this paper, we set out to investigate this issue, evaluating the uncertainty intervals for quantitative EDS analysis after the application of PCA denoising. We will focus on STEM-EDS chemical analysis of few-nm-diameter bimetallic (AuAg) NPs using both experiments and simulations. Our results show that if the input SNR exceeds a certain minimal value, it is possible to avoid random noise bias and reduce uncertainty levels in quantitative EDS chemical analysis.

Materials and Methods

NP Synthesis

Binary alloy (Au_xAg_{1-x}) NPs have been generated using a home-made gas aggregation source (de Sá et al., 2014); in-situ time-of-flight mass spectrometry (TOF-MS) has been used to measure NP mass distribution (size distribution follows a log-normal function, ~4 nm mean average diameter and 3 nm in width; Moreira et al., 2022). The particles were deposited in a "soft landing" mode on the TEM grid (kinetic energy per atom ~0.05 eV for a 4 nm NP). The average Au composition (atomic fraction) of the sample is $C_{Au} = (0.48 \pm 0.01)$, which is obtained by EDS-TEM from an ensemble of NPs using a large open electron beam (several microns in diameter).

Electron Microscopy: Data Acquisition and Processing

We have used different STEMs to compare EDS results from different experimental configurations (all instruments were equipped with Silicon Drift Detectors—SDD): JEM-2100 F (0.39 sR, LNNano, Campinas, Brazil); Titan Themis (Super X Quad, 0.8 sR, LNNANO, Campinas, Brazil); and FEI Tecnai Osiris (SuperX 4, 0.9 sR, DMSM, University of Cambridge, Cambridge, UK). The EDS HSIs were acquired using a 0.5 nm pixel size and a dwell time of ~200 ms per pixel; the image size was set to 64×64 pixels to guarantee the operation of automatic drift correction when realigning scan regions (JEM-2100 F). In order to minimize the dose rate, a series of images of the same region were acquired and, subsequently, individual frames were added (10 scans at 20 ms dwell time) (Jones et al., 2018). For all experiments, we have used a low-background Be sample holder.

The raw EDS spectra were binned to 512 energy channels to increase the SNR for subsequent PCA calculation, leading to an EDS channel width of 20 eV. The Au-Mα and Ag-Lα EDS counts from individual particles have been obtained by summing the counts from pixels located inside each NP region. Since, in our experiments, the pixel size is 0.5 nm, a particle size of 6 nm in diameter (area ~28 nm^2) includes approximately 110 pixels. Each 64×64 pixels scans contained about 5–8 NPs and then the sparsity level is about 85–90%.

The quantitative analysis of the individual NP chemical composition followed the Cliff–Lorimer approach (Cliff & Lorimer, 1975) using experimentally measured K_{AB} factors [equation (1), C_y and $I_{y,NP}$, atomic percentage and x-ray intensity of element y per NP, respectively].

$$\frac{C_A}{C_B} = K_{AB} \frac{I_{A,NP}}{I_{B,NP}}. \quad (1)$$

Cliff–Lorimer factors (K_{AB}) have been measured using a thin film of known composition. A calibrated multilayer thin film was evaporated onto a substrate using a quartz balance to control the atomic ratio through film thickness. The resulting values were as follows: $K_{AuAg} = 0.93 \pm 0.01$ for data acquired in the uncorrected instrument (JEM-2100 F) and $K_{AuAg} = 0.85 \pm 0.01$ for the AC instrument including four EDS detectors (Titan Themis).

Estimations of chemical composition and their error bars have taken into account all EDS quantitative analysis steps applied on a fitted curve of the experimental EDS spectra (background removal, x-ray peak integration, and composition determination); we have considered Poisson noise for all EDS count measurements. All EDS and HSI PCA processing steps have been performed using the open-source Hyperspy Python library (de la Peña et al., 2021).

Simulated Datasets

Synthetic datasets must be a faithful description of an actual experiment, including an appropriate level of noise for each pixel. As a first step, we have selected a total number of counts per NP; we then distributed these counts inside the NP projection (a circular region). It is important to emphasize that the simulated HSI intensity distribution considers the spherical shape of the NPs. This includes a thickness variation along the electron path, so that counts are highest in the center and decrease toward surface pixels. The AuAg NPs were assumed to lie on a uniform a-C film and spurious peaks from the sample environment (C, Fe, Ni, etc.) were also incorporated in the EDS spectra to reproduce experimental data as closely as possible using rescaled spurious counts from experiments. After a noise-less spectra was built, random Poisson noise was added to each energy channel of the EDS spectra in order to yield a realistic simulation of experimental HSIs. For the sake of simplicity, our simulations have employed $K_{AuAg} = 1$.

PCA Processing

PCA provides a number of orthogonal components encoding the information contained in the data variance within a dataset. The component eigenvalues represent their information content and

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

their values are usually compared in the so-called scree-plot (Titchmarsh, 1999; Jolliffe, 2002; Jolliffe & Cadima, 2016).

Before PCA processing, the dataset was rescaled to induce variance stabilization (Keenan & Kotula, 2004; Kotula & Keenan, 2006; Kotula & Van Benthem, 2015). The rescaled matrix D_s is derived from $D_s = GDH$, where G is the spatial factor, D is raw data, and H is the spectral factor (as implemented in Hyperspy package; de la Peña et al., 2021). All EDS data reconstructions in this work have included five PCA components to standardize the comparison between experiments and simulations (data processing in this paper follows the accompanying Part I of this pair of papers; Moreira et al., 2022).

The samples being analyzed are close to ideal: isolated NPs distributed on a uniform, thin carbon film, but with a set of different chemical compositions. The individual composition of each particle is the main information that we would like to assess and improve after PCA denoising. In an ideal high SNR dataset, the main EDS information should be carried by three dominant components (see Part I of this study; Moreira et al., 2022): (a) PC#1, TEM support grid and other EDS spurious x-ray sources (Fe from lenses and Cu from grid); (b) PC#2, mean NP EDS spectrum carrying the average chemical composition (that can be obtained through the Au and Ag peak counts $I_{Au,PC\#2}/I_{Ag,PC\#2}$ from this component); and (c) PC#3, modeling elemental differences between particles (Potapov & Lubk, 2019, see detailed analysis in Moreira et al., 2022). This third component (PC#3) shows an unphysical spectroscopic profile (Potapov, 2016), because the Ag and Au EDS peaks occur with opposite signs for the two elements (one positive, the other negative). This results from the need to model the anti-correlation between the two spectral features: when combined with PC#2 (mean composition), the addition or subtraction of PC#3 is equivalent to the substitution of one element with the other, describing Au- or Ag-rich particles.

Results

Different instrumental configurations result in a variety of incident electron probe characteristics and x-ray detection, yielding datasets with different SNRs. The results of chemical composition measurements from an uncorrected instrument are shown in Figure 1a (hereafter named Experiment E1, realized at 200 kV), where the total Au-Mα intensity per particle is around $I_{Au,NP}$ ~380 counts for a 6-nm-diameter NP. This counting level yields a typical concentration uncertainty interval of about $\sigma_{C_{Au,raw}} = 0.04$. The comparison of NP compositions (C_{Au}) derived from raw or denoised HSIs is shown in Figure 1a; this plot shows C_{Au} as a function of NP diameter to display potential size-effects on composition. The measured NP composition using raw data is centered at the expected average value (C_{Au} ~0.5), and the standard deviation of the distribution (STD) is $\Sigma_{C_{Au,raw}} = 0.05$ (hereafter Σ will be used to describe the STD of a distribution, $\Sigma = \sqrt{\sum_i (x_i - \bar{x})^2/(n-1)}$). With this precision, no clear size-dependent trend can be observed or proven.

The compositions derived from PCA-reconstructed data are still roughly distributed around the mean value but display a much narrower distribution ($\Sigma_{C_{Au,den}} = 0.022$). This may suggest the occurrence of significant random noise-related bias (Moreira et al., 2022). The occurrence of *random noise* PCA bias is associated with insufficient counting; it generates unrealistic reconstructions where the derived values differ from "true" noiseless ones. A

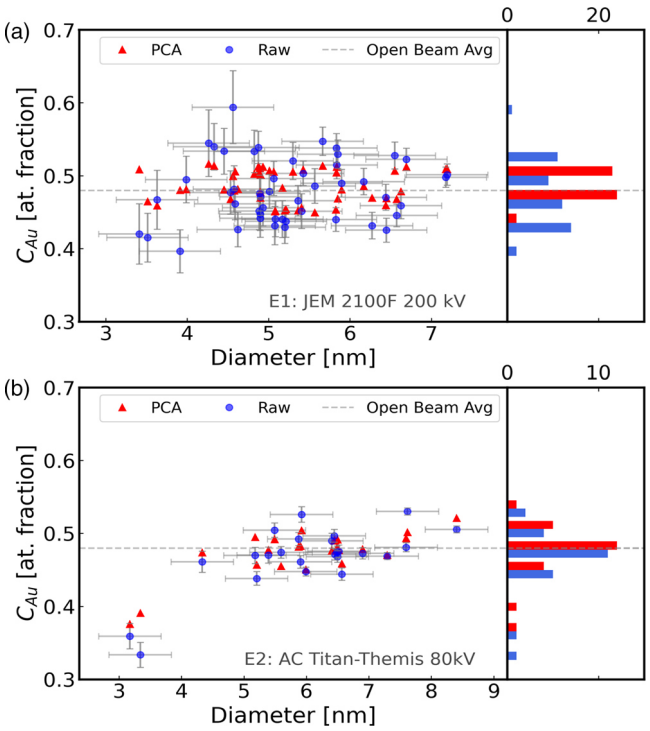


Fig. 1. Quantitative chemical composition analysis of alloy NPs using either raw or PCA denoised EDS data. (**a**) In Exp. E1, dashed line indicates average composition (JEM-2100 F); the plot shows a total of ~50 NP results compiled from 10 HSIs (64 × 64) denoised individually. (**b**) In Exp. E2, dashed line indicates average composition (Titan Themis, 64 × 256 image size generated by the stack of 4× scans 64 × 64 pixels). The resulting chemical composition histograms are plotted at the right.

biased PCA reconstruction may lead to a wrong quantification of the chemical composition of the analyzed sample. PCA-treated data have been displayed without error bars, because there is no accepted and well-defined procedure for calculating uncertainty intervals after denoising (this issue will be discussed in the next sections). In fact, a simple analysis reveals that the input dataset displays an insufficient SNR, and in this case, PCA processing has not been capable of correctly ranking the PC#3, which carries information on compositional variations between NPs. Then, the x-ray intensity ratio ($I_{Au,NP}/I_{Ag,NP}$) is dominated by the PC#2, describing the average EDS spectrum. Therefore, I_{Au} and I_{Ag} are strongly correlated and the ratio becomes close to fixed (and therefore the chemical compositions that follow). This explains why the composition distribution appears significantly narrower after denoising (see histograms at the right of Fig. 1a).

To improve counting statistics and data redundancy (dataset size), we have performed additional experiments using an aberration-corrected (AC) STEM microscope equipped with four EDS detectors (hereafter named Experiment E2). Multiple EDS detectors help to increase counting speed and efficiency; the signal has been further increased by reducing the beam energy to 80 keV, where the ionization cross-section raises by a factor of 1.6× (Goldstein et al., 2017; Zaluzec, 2019). A significant signal improvement has been obtained ($I_{Au,NP}$ = 1,950 counts for a 6-nm-diameter NP, compared to 380 counts in E1). The NP composition distribution (Fig. 1b) shows a good agreement between raw and denoised data (within uncertainty intervals). This counting level results in a PCA reconstruction of EDS HSIs, which agrees with chemical composition estimations (distribution

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

widths are $\Sigma_{C_{\mathrm{Au,raw}}} = 0.043$ and $\Sigma_{C_{\mathrm{Au,den}}} = 0.031$ for raw and PCA-denoised data, respectively; Moreira et al., 2022). For smaller NPs, the PCA-derived composition is statistically different from the raw estimates, moving closer to the average composition. The recorded intensity for these smaller NPs (~3 nm in diameter) is rather low ($I_{\mathrm{Au}} \sim 400$), suggesting that random noise bias behavior may play a role for smaller particles. The improved quality of PCA reconstruction from Experiment E2 suggests a potential size-dependent composition: Au depletion for smaller particles (Fig. 1b).

In order to confirm this size-dependent effect, we must provide data whose statistics are more robust and quantitative. Firstly, we must increase the number of particles included in the study. This can be achieved by performing more experiments with a counting level at least similar to Experiment E2 (ideally higher, to guarantee a sufficient signal level for small particles). A more challenging second step is to reduce uncertainty intervals, in a process that requires effort in several directions. We can increase the counting level to reduce Poisson noise if the sample can withstand an increased dose, although this might affect the very quantity we are looking to measure. Alternatively, we can exploit the ML techniques to increase data SNR. In any case, we are confronted with an unavoidable fact: the validation of a physical size-dependent effect requires a concrete quantitative evaluation of uncertainty. Many dimensional reduction approaches have become very popular (PCA; Jolliffe, 2002): Non-Negative Matrix Factorization (NMF; Lin, 2007), and Independent component Analysis (ICA; Jutten & Herault, 1991); unfortunately, they are mostly used in a qualitative manner, and the evolution of uncertainty intervals is not estimated after data treatment. The following sections will present a discussion of this essential issue for PCA-reconstructed EDS data.

PCA Processing of an Ideally Homogeneous NP Ensemble

From a fundamental point of view, counting noise originates from uncorrelated (random) radiative processes in the time domain, and we can assess the total uncertainty by repeating the measurements in identical conditions and analyzing the reproducibility through the measurement distribution width (Bevington & Robinson, 2003; Drosg, 2009; Hughes & Hase, 2010). This concept can be applied to chemical composition estimations, and in particular, when chemical compositions are derived from PCA-denoised EDS data, we just need to measure the same system multiple times and the uncertainty interval is the standard deviation (Σ_{STD}) of the data cloud.

To reproduce this idealized experiment, we will need "The perfect NP sample": all particles must have exactly the same size and elemental composition. This situation is quite challenging (if not impossible) to attain experimentally, but rather easy to implement as a simulation. In addition, synthetic datasets allow to vary the total Au counts per NP, so that it is possible to compare PCA denoising for different given levels of input SNR. Using simulation, we will be able to compare conventional EDS processing and PCA-denoised results.

We have performed simulations considering a HSI (64×384 pixels) containing 54 NPs, 6 nm in diameter with added Poisson noise (noiseless $I_{\mathrm{Au,NP}} = 1{,}025$ cts, see Fig. 2a). The "true" average composition was chosen to be $C^*_{\mathrm{Au}} = 0.75$ (hereafter * will indicate that the variable is the "true noise-free" *data value*) because this sets a more challenging situation compared to a Au/Ag 1:1 concentration, as I_{Au} and I_{Ag} are significantly

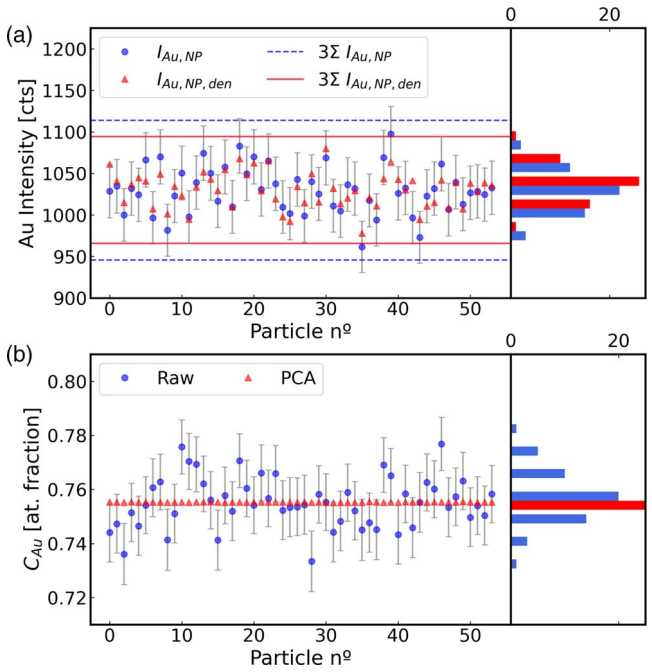


Fig. 2. Analysis of denoising effects on the assessment of chemical composition deduced from simulated data. Simulated HSI includes 54 identical NPs (6 nm in diameter, total count per NP 1,025 cts, C^*_{Au}=0.75, $\nu*_{C\ \mathrm{Au,STD}} = 0$). (**a**) Plot of the Au intensity for each NP; note that after denoising, there is still a clear variation of intensity between particles (dashed and continuous horizontal lines indicate the full width of the intensity distribution (mean value ~3 STD) for raw and denoised data, respectively). (**b**) Plot of the calculated Au concentration for each NP; note that after denoising, Au concentrations are narrowly concentrated around the average value.

different. For this counting level, a manual EDS analysis of an individual NP yields an uncertainty bar $\sigma_{C_{\mathrm{Au}}} = 0.01$; averaging for the ensemble of NPs, we obtain a precision interval of approximately 7× smaller ($\sigma_{C_{\mathrm{Au,Ave}}} = \sigma_{C_{\mathrm{Au}}}/\sqrt{54} = 0.0016$, 54 identical NPs).

The NP chemical composition distribution after PCA treatment (Fig. 2b) is narrowly distributed around the expected value. In fact, the width (STD) of the composition distribution after denoising ($\Sigma_{C_{\mathrm{Au,den}}}$) is about 10× smaller than the uncertainty level deduced in a conventional way by averaging on the 54 particles ($\sigma_{C_{\mathrm{Au,Ave}}}$). From another point of view, this may suggest that it may be too optimistic to use the composition distribution STD as uncertainty interval after PCA denoising.

For this homogeneous sample, the scree-plot shows only two relevant principal components above noise: PC#1, representing background, and PC#2, displaying the average NP spectrum. The second component (PC#2) contains all the information on the characteristic EDS signal for the NPs and, in particular, a well-defined $I_{\mathrm{Au}}/I_{\mathrm{Ag}}$ ratio. In other words, this means that $I_{\mathrm{Au,NP}}$ and $I_{\mathrm{Ag,NP}}$ intensities have become almost fully correlated variables. This explains why the particles show such a narrow composition dispersion after denoising (Fig. 2b). Although an almost constant $I_{\mathrm{Au,NP}}/I_{\mathrm{Ag,NP}}$ ratio dominates the EDS signal, the elemental x-ray line intensities for individual NPs show a certain dispersion (Fig. 2a), stemming from a variability of PCA scores among pixels. A simple analysis of the simulated HSI indicates that the STD of Au intensity distribution ($I_{\mathrm{Au,NP}}$) in raw data ($\Sigma_{I_{\mathrm{Au,NP}}} = 28$ cts) was close to expected Poisson noise ($\sigma_{I_{\mathrm{Au,NP,Poi}}} = \sqrt{1,025} = 32$ cts, only noise source considered in these simulations). After PCA denoising, the intensity variability of $I_{\mathrm{Au,NP}}$ reduces and the

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

STD becomes $\Sigma_{I_{\mathrm{Au,NP,den}}} = 21$ cts. We must conclude that PCA processing has not been able to recover "noise-free" intensity values, and some counting noise still remains in the PCA-reconstructed HSI. This remaining intensity variability could be considered a potential quantifiable uncertainty interval for "denoised" intensities per NP.

Considering the STDs of denoised x-ray counts ($\Sigma_{I_{\mathrm{Au,NP,den}}}$ and $\Sigma_{I_{\mathrm{Ag,\ NP,\ den}}}$) as intensity error bars, we can tentatively use them to estimate the post-PCA composition precision. For the sake of simplicity, we will assume that the Cliff–Lorimer factor is ~1 and shows negligible uncertainty ($\sigma_{K_{\mathrm{Au,Ag}}} = 0$). It is important to emphasize that, after denoising, $I_{\mathrm{Au,NP,den}}$ and $I_{\mathrm{Ag,NP,den}}$ have become fully correlated. These two intensity values must be considered dependent input parameters, and we must calculate error bars using a cross-derivative term between variables (Bevington & Robinson, 2003; Drosg, 2009; Hughes & Hase, 2010; Ritchie, 2020):

$$\sigma_{C_{\mathrm{A}}} = \sqrt{\left(\frac{\partial C_{\mathrm{A}}}{\partial I_{\mathrm{A}}}\right)^2 \sigma_{I_{\mathrm{A}}}^2 + \left(\frac{\partial C_{\mathrm{A}}}{\partial I_{\mathrm{B}}}\right)^2 \sigma_{I_{\mathrm{B}}}^2 + 2\frac{\partial C_{\mathrm{A}}}{\partial I_{\mathrm{A}}}\frac{\partial C_{\mathrm{A}}}{\partial I_{\mathrm{B}}}\ \sigma_{I_{\mathrm{A}}}\ \sigma_{I_{\mathrm{B}}}\rho_{\mathrm{AB}}}\ , \tag{2}$$

where ρ_{AB} describes correlation coefficient between intensities from elements A and B ($\rho_{\mathrm{AB}} \in [-1, 1]$).

As Au and Ag intensities are fully correlated (the measurement of one tells us the other), we will take $\rho_{\mathrm{Au,Ag}} = 1$. Using this criterion, the composition uncertainty after denoising results $\sigma_{C_{\mathrm{Au,den}}} = 0.0006$, that is around 2.8× smaller than the average error bar calculated by traditional methods on raw data and averaging the measurement for 54 NP $\sigma_{C_{\mathrm{Au,Ave}}} = 0.0016$. The fact that the uncertainty interval after PCA processing results smaller than a standard averaging of NP concentration may appear contradictory. In fact, when applying PCA to a dataset matrix of dimension $(m \times n)$ with homoscedastic noise σ_{H}, the maximum noise reduction should be that PCA can attain is $\sigma_{\mathrm{PCA,den}} \sim \sigma/\sqrt{\mathrm{nm}}$, as estimated by Nadler (2009). In other words, the maximum expected denoising performance of PCA is the averaging of all voxels in the dataset. In this work, we are analyzing NP chemical composition; this is a physical quantity, which is derived from a reasonably large group of pixels (~100). Thus, the 2.8× smaller error bar derived from denoised data does not contradict established knowledge on PCA processing (Nadler, 2009).

Having studied synthetic HSI data, it is essential to perform a similar analysis on actual experimental measurements. It is hardly possible to chemically or physically synthesize a binary metallic alloy NP sample with neither composition variation nor size dispersion. We have chosen a different sample that allows us to emulate an ensemble of identical particles. We have taken a III–V semiconductor nanowire made of indium phosphide (InP, with minute Ga content) grown by the VLS (vapour–liquid–solid) catalytic method (Fig. 3a) (Tizei et al., 2010). As the wire shape shows some tapering, we have selected a wire region close to the tip in order to avoid diameter variations that may induce additional spectroscopic effects associated with sample thickness (such as absorption). We have used a region of the wire as chemical standard using the In-L$_\alpha$ and P-K_α x-ray lines (see Fig. 3), giving a Cliff–Lorimer factor of $K_{\mathrm{In},P} = 0.63 \pm 0.01$. We have then taken a different wire region with constant thickness and considered 20 axial wire positions (AWPs) separated by a step of 5 nm. The aim is to emulate an ensemble of 20 identical NPs. To ensure a good SNR we summed the signal in the direction perpendicular to the axis (11 pixels per AWP), like we summed pixels across a NP in previous examples.

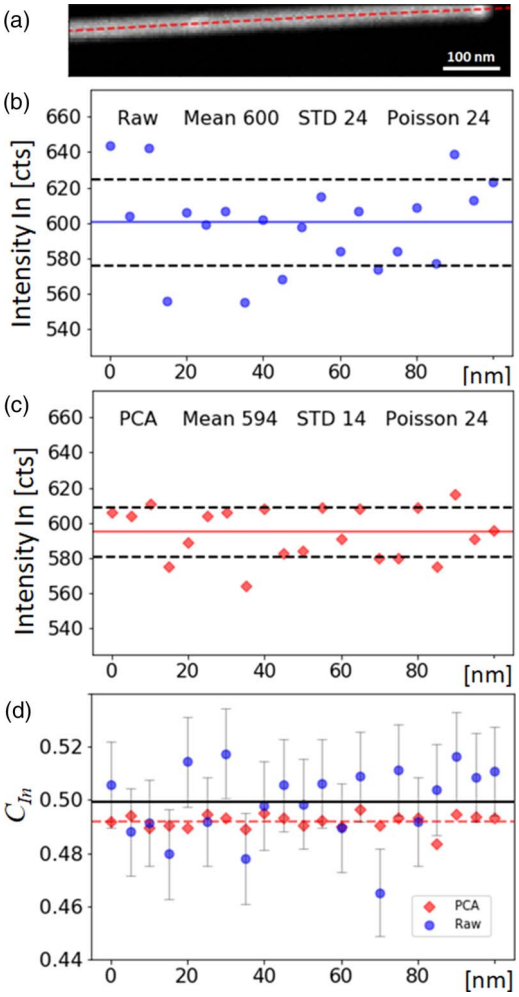


Fig. 3. EDS characterization of a semiconductor InP nanowire. (**a**) STEM annular dark field image of the nanowire. (**b**) Indium intensity for 20 AWP positions (raw data) along the nanowires axis emulating an ensemble of 20 identical NPs; note that the STD of the intensity distribution fits reasonably well the calculated poisson noise ($\sigma = \sqrt{I}$). (**c**) Denoised data: indium EDS counts from the 20 AWPs; note that after denoising, the EDS intensity distribution shows a dispersion, which is narrower than the calculated one using Poisson noise. This indicates that the PCA reconstruction has not been able to recover noise-free intensity values and some counting noise still remains in the denoised HSI. (**d**) Comparison of measured indium concentration for the 20 axial NW points derived from raw and PCA denoised HSI; note that denoised data are narrowly clustered around their mean value. In (**b**) and (**c**), continuous and dashed horizontal lines indicate the intensity distribution mean and (mean value ~ STD), respectively, so that the narrowing of intensity variability becomes easily identifiable by eye.

Raw data on the AWPs yield on average 600 and 391 x-ray counts for In and P, respectively (Fig. 3b). The observed intensity STDs ($\Sigma_{I_{\mathrm{In,AWP}}} \sim 24$ and $\Sigma_{I_{P,\mathrm{AWP}}} \sim 19$ cts) agree very well with expected Poisson noise values ($\sigma_{I_{\mathrm{In,Poi}}} = \sqrt{600} \sim 24$ cts and $\sigma_{I_{P,\mathrm{Poi}}} = \sqrt{391} \sim 20$ cts). The distribution of indium concentration (C_{In}) derived from raw data shows a distribution of mean $C_{\mathrm{In,Ave}} = 0.50$ and STD $\Sigma_{C_{\mathrm{In,As,raw}}} \sim 0.014$ (in good agreement with an estimated error bar of $\sigma_{C_{\mathrm{In}}} \sim 0.016$ for each AWP, Fig. 3d).

As observed for simulations displayed in Figure 2, the composition distribution is much narrower after denoising (Fig. 3d; mean $C_{\mathrm{In,Ave,den}} = 0.49$ and STD $\Sigma_{C_{In,\mathrm{den}}} \sim 0.0028$). In addition, intensities from each element show a dispersion reduction

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

compared to raw data ($\Sigma_{I_{\mathrm{In},\mathrm{AWP},\mathrm{den}}} \sim$ 14 cts, $\Sigma_{I_{\mathrm{P},\mathrm{AWP},\mathrm{den}}} \sim$ 7 cts, see Fig. 3). Using these last values as intensity uncertainty intervals and considering the full correlation of In and P x-ray intensities, the chemical composition error bar ($\sigma_{C_{\mathrm{In,den}}} = 0.0042$) is about 3× smaller after denoising than using raw data.

Briefly, experimental data have shown a behavior in very good agreement with simulated HSI of identical NPs. This suggests that PCA denoising may allow a partial reduction of counting noise (Fig. 3) and a potential reduction of uncertainty intervals.

PCA: Information Detection Limit and Estimation of Uncertainty Intervals

ML algorithms represent statistical methods capable of retrieving as much information as possible from available data by exploiting redundancies to find hidden correlations. After dimensionality reduction, bias describes the differences between reconstructed data and "true" values (Faber et al., 1995a, 1995b; Malinowski, 2002; Nadler, 2008, 2009). In fact, the SNR of input data determines how much of "true noiseless" properties of the sample are actually captured by experiments. It is essential to consider that ML algorithms cannot reveal what has not been detected due to low SNR (or in other terms, ML cannot create information).

In spite of using ML processing, it will always be essential to work out experiment details (number of pixels, detector channels, counting level, and noise sources) to guarantee the reliable detection of the physical effect being targeted (Moreira et al., 2022). In addition, simulations will also be necessary to support the physical interpretation of ML results. In fact, the association of simulations and experiments (as realized in the previous section) represents the recommended approach in the scientific and technological path to the data-driven next-generation TEM (Spurgeon et al., 2020). In this work, we would like to gather the chemical composition distribution from an ensemble of binary alloy NPs and we aim to detect the possible occurrence of a size-dependent effect on composition distribution. The increase of information content of the input data can be mainly reached by higher counting levels and secondly by increasing redundancy.

Most studies addressing the effect of noise on PCA reconstruction bias have suggested possible bias indicators based on models considering uncorrelated homoscedastic noise (Faber et al., 1995a, 1995b; Malinowski, 2002; Nadler, 2008, 2009). Conversely, our experiments involves Poisson noise, and we must account that PCA is not frequently applied on original data, but also on scaled datasets after variance stabilization (Keenan & Kotula, 2004; Kotula & Keenan, 2006; Kotula & Van Benthem, 2015). Then, we must emphasize that the application of these bias indicators must be performed with utmost precaution. Within the so-called *spiked covariance model,* Nadler has developed an information loss estimator [E_{IL}, see equation (3)] related to a sharp information loss-phase transition (Nadler, 2008, 2009). A bias-free PCA reconstruction is always achieved when E_{IL} is below an information loss threshold [T_{IL}, see equation (4)]. The estimator expression is simple, and it points out a relation between dataset dimensions (m pixels and n spectral channels), measurement noise variance σ^2, and "noise-free" data variance v^{*2}. The use of this indicator is somewhat limited, because the true noise-free data variance parameter is hardly accessible in practice. In the limit that data matrix size ($m \times n$) tends to infinity, the threshold becomes $T_{\mathrm{IL}} = 1$ (Nadler, 2008, 2009). The application of this estimator has already been explored in both EELS and EDS HSI processing (Lichtert & Verbeeck, 2013; Potapov, 2017; Potapov & Lubk, 2019).

$$E_{\mathrm{IL}} = \frac{n}{m}\left(\frac{\sigma^2}{v^{*2}}\right)^2, \tag{3}$$

$$E_{\mathrm{IL}} \leq T_{\mathrm{IL}}. \tag{4}$$

Potapov & Lubk (2019) have rewritten equation (3) to define a "retrievability criterion": a minimal standard deviation (S_{I}^*) that a "true" latent factor must show to be effectively detected by the PCA processing [equation (5)]. The retrievability of principal components was tested for the case of chemical maps from microelectronics devices [for this application, information loss threshold (T_{IL}) was assumed to be equal to 1 in equation (5)].

$$S_{\mathrm{I}}^* > \sqrt{\frac{n}{m}}\frac{1}{\sqrt[4]{T_{\mathrm{IL}}}}\sigma. \tag{5}$$

The deterministic sharp phase transition of information loss arises in the limit for infinite values of m and n. For finite values of (m, n), we must expect a continuous increasing probability in the relation between variance and limiting eigenvector instead of a sharp transition (Nadler, 2008, 2009). Nadler has reported a heuristic demonstration of this fact (Nadler, 2009), where the estimator was found with the same functional dependence but considering a multiplicative factor of >1 for the threshold relation.

From a numerical point of view, PCA processing of EDS data is quite performant, because x-ray signal is constituted by nicely shaped peaks on a very low and almost non-varying background (Potapov, 2017). This profile guarantees an easy identification of relevant EDS peaks but at the same time generates high sparsity data. In fact, most spectral energy channels in an EDS HSI include zero signal or just a few counts, then the average counting value per voxel ($I_{\mathrm{ave,voxel}}$) may be $\ll 1$ by several orders of magnitude. This renders extremely difficult to use the simple square root of intensity ($\sigma = \sqrt{I_{\mathrm{ave,voxel}}}$) as an effective and reliable HSI noise level due to Poisson statistics.

In Part I of this series of works (Moreira et al., 2022), we have suggested to replace the use of average intensity per voxel of the whole dataset as the noise variance of the EDS measurements. Instead, we have proposed to overcome the intrinsic sparsity of EDS data by calculating the noise variance taking into account just the region of the EDS spectrum carrying chemical information (the characteristic x-ray peaks). In this way, the experimental noise estimation takes into account the focus of our study, which is the assessment of individual NP composition. It is important to consider that chemical composition is an information directly related to the total characteristic counts per NP ($I_{\mathrm{Au,NP}}$ considering Au signal). These counts are firstly distributed over the number of pixels (m_{NP}) inside the particle region and, secondly, on the spectral channels (n_{Peak}) under the EDS peak. The noise variance for voxels inside the EDS peak of interest can be approximately calculated as follows:

$$(\sigma_{I_{\mathrm{Au,vox}}})^2 = I_{\mathrm{Au,vox}} = I_{\mathrm{Au,NP}}/(m_{\mathrm{NP}} n_{\mathrm{Peak}}), \tag{6}$$

assuming Poisson statistics and a homogeneous distribution of counts on NP pixels and EDS peak spectral channels.

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

The next parameter to estimate is the "noiseless" data variance (v^{*2}) for each component carrying chemical information. On the basis of our previous studies of binary NPs (Moreira et al., 2022), we must consider two PCA components, one providing the mean NP chemical concentration (PC#2) and a second one associated with composition variation (PC#3). To start, we will assume a "true" total number of counts per NP, which is associated with "true" mean Au concentration ($C_{Au,0}$). Following the approximations described above, the "noiseless variance" for PC#2 component is (hereafter noted by the symbol †):

$$v^{\dagger\,2}_{Au,PC\#2} = I_{Au,0,NP}/(m_{NP}n_{Peak}), \tag{7}$$

where $I_{Au,0,NP}$ represents the expected "true noiseless" intensity per NP.

At this point, we must estimate how the "true" chemical composition dispersion influences the *noise-free* voxel values. Let's start with a rather simple model, where the NP ensemble displays a Gaussian distribution of chemical composition (mean $C_{Au,0}$, variance $\Sigma^2_{C_{Au,0}}$). We can derive how many counts ($\Delta I_{Au,\sigma}$) must be added to $I_{Au,0}$ in order to get a NP with a new concentration $C_{Au,\sigma} = C_{Au,0} + \Sigma_{C_{Au,0}}$. Without losing the general applicability of our approach, we must note that the Cliff–Lorimer factor is assumed to be $k_{Au,Ag} \sim 1$ for our experiments. Then, it is easy to show that $C_{Au} = I_{Au,NP}/I_{Tot,NP}$, where $I_{Tot,NP} = (I_{Au,NP} + I_{Ag,NP})$. In addition, the total count per NP $I_{Tot,NP}$ is a constant for NPs of identical size (see Moreira et al., 2022).

Due to this direct proportionality between composition and elemental counts, a composition change of value $\Sigma_{C_{Au}}$ requires an intensity change of $\Delta I_{Au} = I_{Tot,NP}\,\Sigma_{C_{Au}}$. Distributing these counts between NP pixels and spectral channel, we can derive the true standard deviation of intensity distribution associated with chemical composition distribution variability, which is carried by the third PCA component PC#3. Then, we can estimate the "true" STD associated with PC#3 using:

$$v^{\dagger}_{Au,PC\#3} = I_{Au,Tot,NP}\,\Sigma_{C_{Au}}/(m_{NP}n_{Peak}). \tag{8}$$

(Note that the estimation of the "true" variance associated with PC#3 requires squaring this value).

In Part I of this series of works (Moreira et al., 2022), we have reported that denoised data reconstructions (both from experiments and simulations) may show very good fidelity (i.e. accuracy) to "true" noiseless chemical composition distribution. This PCA processing is usually associated with an information loss bias estimator $E_{IL} \sim 300$ (input parameters, $m = 24{,}576$, $n = 500$, $m_{NP} = 110$, $n_{Peak} = 10$, $I_{Au,0,NP} = 3540$, $C_{Au,0} = 0.75$, $\Sigma_{C_{Au}} = 0.036$, see examples in simulations #3 and #5 in Moreira et al. (2022)). This somewhat confirms Nadler's conclusion that for a finite data matrix (Nadler, 2009), it may be too stringent to use as threshold for information loss criteria as $T_{IL} = 1$. Keeping this idea in mind, it may be interesting to take 300 as a reasonable value for the threshold (T_{IL}) for the particular physical problem analyzed here (determination of chemical composition), although this value is much higher than frequently used in literature (Potapov, 2017; Potapov & Lubk, 2019).

We may rewrite the retrivability criteria [equation (5)] proposed by Potapov & Lubk (2019) as the minimal "true" data STD for detection of a specific PCA loading as a function of input data measurement noise σ_M as:

$$S^*_I > f_{red}\ f_{Th}\ \sigma_M, \tag{9}$$

where we have explicitly isolated the factors that may modify the detectability of latent factors. These factors are associated with redundancy:

$$f_{red} = \sqrt[4]{n/m} \tag{10}$$

and the information loss threshold value (Nadler, 2008, 2009):

$$f_{Th} = 1/\sqrt[4]{T_{IL}}. \tag{11}$$

Considering that f_{red} and f_{Th} are both <1, the expression in equation (9) indicates how PCA processing will exploit redundancy and allow the detection of latent components whose "true" variability is below measurement noise. Due to the fourth power root, the retrievability improvement by increasing redundancy [equation (10)] or threshold value [equation (11)] is rather moderate. For example, taking $T_{IL} \sim 300$, as suggested by Moreira et al. (2022) for the study of binary alloy NPs, the retrievability value S^*_I is just changed by a factor of ~1/4 ($0.24 = 1/\sqrt[4]{300}$).

Nadler (2009) has deduced that for a finite dataset size, the information loss probability should be gradual without a sharp phase transition and without a define threshold value $T_{IL} = 1$. Then, we suggest to use the well-known 3-sigma criterion as a detection decision value (Currie, 1968, 1999; Williams & Carter, 2009; Belter et al., 2014). This leads to the following estimation of the detection decision value D_I:

$$D_I = F_{det} S^*_I, \tag{12}$$

where $F_{det} = 3$ and represents multiplicative 3 factor of the 3-sigma criterion. Considering explicitly experimental noise, the final expression for D_I results:

$$D_I = (F_{det} f_{red} f_{Th})\ \sigma_M. \tag{13}$$

Any latent factor whose "noise-free" intensity variability is below D_I will not be detected and properly ranked by PCA processing. Therefore, the related information will be scattered in the ensemble of loadings related to noise. When we reconstruct the dataset considering all relevant principal components (selected from the scree-plot), the PCA reconstruction will include components whose true noiseless standard deviation must be at least equal or larger than the detection value ($\sigma_I \geq D_I$).

For accurate unbiased PCA reconstruction, D_I can be thought as an approximation for the lower bound intensity error bar for denoised intensity values ($\sigma_{I,den} = D_I$). We may suggest rewriting equation (13) as an estimation of uncertainty interval $\sigma_{I,den}$ of the PCA reconstruction:

$$\sigma_{I,den} = R_{PCA}\ \sigma_M, \tag{14}$$

where R_{PCA} represent the effective PCA contribution to reduce measurement noise and to increase precision ($R_{PCA} = F_{det} f_{red} f_{Th}$).

We can further use the proposed intensity error $\sigma_{I,den}$ in subsequent uncertainty calculations using conventional error propagation procedures (Bevington & Robinson, 2003; Drosg, 2009;

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

Table 1. Analysis of Bias Estimators and Uncertainty Intervals for Datasets That Have Shown a Good Fidelity in the PCA-Denoised Reconstruction.

Experiment/Simulation	"True" $\nu^*_{C_{Au}}$	$I_{Au,NP}$ (cts)	Bias (qualitative)	Inf. Loss E_{IL}	$\sigma_{I_{Au},NP}$ (Poisson)	R_{PCA}	$\frac{\sigma_{C_{Au},raw}}{\sigma_{C_{Au},den}}$
Experiment E2	0.043[a]	1,950	Moderate	180	44	0.30	2.3
Simulation 1[a]	0.02	10,660	Low	430	103	0.27	2.7
Simulation 2[a]	0.036	3,540	Low	370	60	0.27	2.7

Information loss threshold assumed as $T_{IL} = 300$.
[a]Simulations 1 and 2 reported in related publication Part 1 (reported as # 3 and# 5 in Moreira et al. (2022), respectively).

Hughes & Hase, 2010). At this point, it is essential to keep in mind that, after PCA denoising, EDS intensities are not independent variables. In particular, for binary alloy NP displaying a certain composition variability, PC#3 should contribute with intensities of Au and Ag, which are fully anticorrelated. The PC#3 loading shows these EDS peaks with the opposite sign (see Moreira et al., 2022), then we must use $\rho_{AuAg} = -1$ when applying equation (2) to calculate uncertainty for this case.

The aim of our study is the estimation of elemental composition of individual NP, an information related to total characteristic counts per NP total intensity (I_{NP}) and its uncertainty $\sigma_{I,NP}$ (mainly associated with Poisson statistics and background subtraction). Table 1 explores the application of error bar calculation using the suggested after-denoising intensity uncertainty [$\sigma_{I,NP,den}$ from equation (14)] to Experiment E2 and high SNR simulations from the Part I of this series of articles (Moreira et al., 2022).

Briefly, we may conclude that PCA denoising has contributed positively to increase precision; considering the values of our experiments and simulations, the reduction factor for intensity noise R_{PCA} lays in the range of ~0.3, leading approximately to composition uncertainty reduction of the same order, than for the monodispersed case analyzed above (~2.3- to 2.8-fold decrease).

It is essential to remember that this uncertainty analysis can only be applied to datasets where the occurrence of random noise bias has been minimized and the accuracy of reconstruction has been thoroughly analyzed. The evaluation of PCA reconstruction fidelity for the three cases included in Table 1 has been described in detail in Part I of this work (Moreira et al., 2022).

Conclusions

PCA data processing is a very powerful tool to reduce dimensionality and increase the SNR of HSIs, and it is frequently used to improve the quality of TEM/STEM EDS "chemical maps". Our study has shown that the PCA "denoising" procedure can also be applied to upgrade EDS quantitative chemical analysis of binary alloy NPs. In fact, the correct application of PCA may reduce the contribution of Poisson noise in EDS spectroscopy of binary systems, leading to a narrowing of composition uncertainty intervals. However, PCA users must be aware that this is not a general output, and it is essential to ensure that the SNR exceeds a minimal value to minimize random noise bias to an acceptable level in the PCA-reconstructed data (guaranteeing accuracy). Briefly, PCA processing can represent a potential practical route to obtain denoised results that are reliable from the point of view of accuracy and showing improved precision if input data are acquired with a suitable SNR.

Concerning the new knowledge gathered for NP research generated in the gas aggregation source, the developments described in this work are very promising. The improvement in the quantitative analysis precision by using PCA denoising may allow better EDS studies to identify issues such as size-effects or elemental distribution inside particles (core-shell and chemical gradients).

Acknowledgments. D.U. acknowledges financial support from the Brazilian Agencies FAPESP (No. 2014/01045-0), CNPq (402571/2016-9 & No. 306513/2017-0), and FAEPEX-UNICAMP (2632/17). V.R. thanks funding from FAPESP (2007/01722-9) and CNPq (555647/2006-4 and 577046/2008-0). M.M. thanks funding from CNPq (No. 162541/2018-0). Access to the FEG-TEM/STEM from the Brazilian Nanotechnology National Laboratory is acknowledged (LNNANO, grant No. ME-22329). M.A.C. acknowledges financial support from FAPESP (Nos. 2013/02300-1 and 2013/10957-0) and CNPq (No. 479486/ 2012-3). L.H.G.T., G.D., and C.D. acknowledge funding from the European Union Seventh Framework Program (FP7/2007-2013) under Grant Agreement No. 312483 (ESTEEM2). M.H. acknowledges financial support from FAPESP (No. 2016/12807-4).

Conflict of Interest. The authors declare no competing financial interest.

References

Alloyeau D, Mottet C & Ricolleau C (2012). *Nanoalloys: Synthesis, Structure and Properties*. London: Springer-Verlag.

Belter M, Sajnóg A & Barałkiewicz D (2014). Over a century of detection and quantification capabilities in analytical chemistry – historical overview and trends. *Talanta* **129**, 606–616.

Bevington PR & Robinson DK (2003). *Data Reduction and Error Analysis for the Physical Science*, 3rd ed. New York: MacGraw-Hill.

Binns C (2014). *Nanomagnetism: Fundamentals and Applications*, vol. 6. Amsterdam, Netherlands: Elsevier.

Brown KA, Brittman S, Maccaferri N, Jariwala D & Celano U (2020). Machine learning in nanoscience: Big data at small scales. *Nano Lett* **20**, 2–10.

Carter CB & Williams DB (2016). *Transmission Electron Microscopy: Diffraction, Imaging, and Spectroscopy*. New York: Springer.

Cliff G & Lorimer GW (1975). The quantitative analysis of thin specimens. *J Microsc* **103**, 203–207.

Cueva P, Hovden R, Mundy JA, Xin HL & Muller DA (2012). Data processing for atomic resolution electron energy loss spectroscopy. *Microsc Microanal* **18**, 667–675.

Currie LA (1968). Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal Chem* **40**, 586–593.

Currie LA (1999). Detection and quantification limits: Origins and historical overview. *Anal Chim Acta* **391**, 127–134.

de la Peña F, Prestat E, Fauske VT, Burdet P, Lähnemann, Furnival T, Jokubauskas P, Nord M, Ostasevicius T, MacArthur KE, Johnstone DN, Sarahan M, Aarholt T, Taillon J, pquinn-dls, Migunov V, Eljarrat A, Caron J, Poon T, Mazzucco S, Francis C, Martineau B, actions-user, Somnath S, Slater T, Tappy N, Walls M, Cautaerts N, Winkler F & DENSmerijn (2021). hyperspy/hyperspy: Release v1.6.5 (v1.6.5). Zenodo. doi:10.5281/zenodo.5608741.

de Sá ADT, Oiko VTA, di Domenicantonio G & Rodrigues V (2014). New experimental setup for metallic clusters production based on hollow cylindrical magnetron sputtering. *J Vac Sci Technol B* **32**, 061804.

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

Drosg M (2009). *Dealing with Uncertainties: A Guide to Error Analysis*, 2nd ed. Berlin: Springer.

Faber NM, Meinders MJ, Geladi P, Sjiistrijm M, Buydens LMC & Kateman G (1995a). Random error bias in principal component analysis. Part I. Derivation of theoretical predictions. *Anal Chim Acta* **304**, 257–271.

Faber NM, Meinders MJ, Geladi P, Sjiistrijm M, Buydens LMC & Kateman G (1995b). Random error bias in principal component analysis. Part II. Application of theoretical predictions to multivariate problems. *Anal Chim Acta* **304**, 273–283.

Ferrando R (2016). *Structure and Properties of Nanoalloys in Frontiers of Nanoscience*, pp. 2–337. Amsterdam: Elsevier.

Goldstein JI, Newbury DE, Michael JR, Ritchie NWM, Henry J, Scott HJ & Joy DC (2017). *Scanning Electron Microscopy and X-Ray Microanalysis*, 4th ed. New York: Springer.

Hawkes P & Spence JCH (2019). *Springer Handbook of Microscopy*. Switzerland: Springer.

Heiz U & Landman UE (2007). *Nanocatalysis*. Heidelberg, Berlin: Springer.

Hughes IG & Hase TPA (2010). *Mesurements and Their Uncertainties, A Practical Guide to Modern Error Analysis*. Oxford: Oxford University Press.

Jolliffe IT (2002). *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag.

Jolliffe IT & Cadima J (2016). Principal component analysis: A review and recent developments. *Philos Trans A* **374**, 20150202.

Jones L, Varambhia A, Beanland R, Kepaptsoglou D, Griffiths I, Ishizuka A, Azough F, Freer R, Ishizuka K, Cherns D, Ramasse QM, Lozano-Perez S & Nellist PD (2018). Managing dose-, damage- and data-rates in multi-frame spectrum-imaging. *Microscopy* **67**, 98–113.

Jutten C & Herault J (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process* **24**, 1–10.

Keenan MR & Kotula PG (2004). Accounting for poisson noise in the multivariate analysis of TOF-SIMS spectrum images. *Surf Interface Anal* **36**, 203–212.

Kotula PG & Keenan MR (2006). Application of multivariate statistical analysis to STEM xray spectral images: Interfacial analysis in microelectronics. *Microsc Microanal* **12**, 538–544.

Kotula PG & Van Benthem MH (2015). Revisiting noise scaling for multivariate statistical analysis. *Microsc Microanal* **21**, 1423–1424.

Lichtert S & Verbeeck J (2013). Statistical consequences of applying a PCA filter on EELS spectrum images. *Ultramicroscopy* **125**, 35–42.

Lin C-J (2007). Projected gradient methods for nonnegative matrix factorization. *Neural Comput* **19**, 2756–2779.

Lyman CE, Lakis RE & Stenger Jr. HG (1995). X-ray emission spectrometry of phase separation in PtRh nanoparticles for nitric oxide reduction. *Ultramicroscopy* **58**, 25–34.

Malinowski ER (2002). *Factor Analysis in Chemistry*, 3rd ed. New York: Wiley.

Moreira M, Hillenkamp M, Divitini G, Tizei LHG, Ducati C, Cotta MA, Rodrigues V & Ugarte D (2022). Improving quantitative EDS chemical analysis of alloy nanoparticles by PCA denoising: Part I, reducing reconstruction bias. *Microsc Microanal.* doi:10.1017/S1431927621013933.

Nadler B (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann Stat* **36**, 2791–2817.

Nadler B (2009). Discussion. *J Am Stat Assoc* **104**, 694–697.

Odom TW & Schatz GC (2011). Introduction to plasmonics. *Chem Rev* **111**, 3667–3668.

Pennycook SJ & Nellist PD (2011). *Scanning Transmission Electron Microscopy*. New York: Springer.

Potapov P (2016). Why principal component analysis of STEM spectrum images results in abstract, uninterpretable loadings? *Ultramicroscopy* **160**, 197–212.

Potapov P (2017). On the loss of information in PCA of spectrum-images. *Ultramicroscopy* **182**, 191–194.

Potapov P & Lubk A (2019). Optimal principal component analysis of STEM XEDS spectrum images. *Adv Struct Chem Imaging* **5**, 4.

Ritchie NWM (2020). Embracing uncertainty: Modeling the standard uncertainty in electron probe microanalysis—Part I. *Microsc Microanal* **26**, 469–483.

Spurgeon SR, Ophus C, Jones L, Petford-Long A, Kalinin SV, Olszta MJ, Dunin-Borkowski RE, Salmon N, Hattar K, Yang W-CD, Sharma R, Du Y, Chiaramonti A, Zheng H, Buck EC, Kovarik L, Penn RL, Li D, Zhang X, Murayama M & Taheri ML (2020). Towards data-driven next-generation transmission electron microscopy. *Nat Mater* **20**, 274–279.

Titchmarsh JM (1999). EDX spectrum modelling and multivariate analysis of sub-nanometer segregation. *Micron* **30**, 159–171.

Tizei LHG, Chiaramonte T, Cotta MA & Ugarte D (2010). Characterization of interface abruptness and material properties in catalytically Grown III-V nanowires: Exploiting plasmon chemical shift. *Nanotechnology* **21**, 295701.

Williams DB & Carter CB (2009). *Transmission Electron Microscopy Part 1: Basics*, 2nd ed. New York: Springer.

Zaluzec NJ (2019). Improving the sensitivity of X-ray microanalysis in the analytical electron microscope. *Ultramicroscopy* **203**, 163–169.

https://doi.org/10.1017/S1431927622000551 Published online by Cambridge University Press

CCC RightsLink

Home | Help | Email Support | Sign in | Create Account

Improving Quantitative EDS Chemical Analysis of Alloy Nanoparticles by PCA Denoising: Part II. Uncertainty Intervals

Author: Murilo Moreira, Matthias Hillenkamp, Giorgio Divitini, Luiz H. G. Tizei, Caterina Ducati, Monica A. Cotta, Varlei Rodrigues, Daniel Ugarte

Publication: Microscopy and Microanalysis
Publisher: Cambridge University Press
Date: Apr 18, 2022

Copyright © Copyright © The Author(s), 2022. Published by Cambridge University Press on behalf of the Microscopy Society of America

License Not Required

Permission is granted at no cost for use of content in a Master's Thesis and/or Doctoral Dissertation. If you intend to distribute or sell your Master's Thesis/Doctoral Dissertation to the general public through print or website publication, please return to the previous page and select 'Republish in a Book/Journal' or 'Post on intranet/password-protected website' to complete your request.

BACK CLOSE

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Chapter 8

Chemical composition quantification within Bimetallic Nanoparticles: Measuring chemical gradients

Energy Dispersive X-ray Spectroscopy (EDS) combined with Scanning Transmission Electron Microscopy offers the possibility of extracting quantitative chemical information localized in the nanoscale [7]. However, as with any experimental methodology, noise and properties of the measurement may hide information in the data sets. In this work, we wish to discuss the elemental distribution inside AuAg BNPs. Moreover, we wish apply and discuss the utility of machine learning approaches such as PCA and NMF to leverage information with a better signal-to-noise ratio (SNR) from EDS data sets. The data architecture plays a fundamental role in applying such methodologies. We have the so-called hyperspectral images (HSI) in an EDS-STEM data set, as already explained in the last chapters. Using these decomposition methods opens the door for blind source separation (BSS) approaches where the signal in the data can be entirely unmixed and hidden features can be extracted without human bias. Naturally, elemental distributions or chemical phases can be identified by those methods [13, 14] and therefore, we propose using these algorithms to process AuAg data sets to contribute to the understanding of how the atoms are distributed within the nanoparticles. In addition, we will discuss the physics and chemistry of elemental distribution inside the AuAg BNPs grown by physical routes. Here, we will try to clarify questions related with where the atomic species, in average, are located within the BNPs protected from air exposure effects.

8.1 Elemental distribution of bimetallic AuAg nanoparticles

Bimetallic nanoparticles (BNPs) have attracted increasing attention in both fundamental as well as applied sciences in the last decades [1, 2]. Here the nanoparticle

properties depend not only on size, shape, and environment but also their chemical composition and ordering. Different chemical configurations (alloyed vs. segregated, chemically ordered, etc.) can considerably change the optical, magnetic, or catalytic properties. One of the most intensely studied bimetallic systems is the gold-silver alloy. Its nanoparticles have attracted a lot of attention for their catalytic [22], plasmonic [23] and photocatalytic [24] properties. In bulk, the two metals are completely miscible across the entire composition range, and their atomic radii, crystal structure, valence, and electronegativity are very similar. On the nanoscale, however, additional degrees of freedom have to be considered, namely the particle size and morphology, as well as possible segregation and/or chemical ordering [1]. Despite the high number of experimental and theoretical articles published on AgAu nanoparticles, the controversy about whether intrinsically one of the two metals preferentially segregates at the surface or not is still not resolved, neither on the experimental nor on the theoretical side. Without claim of completeness, we cite just a few of the more recent works in order to illustrate the discrepancies. Concerning theoretical descriptions of AgAu nanoparticles, several approaches can be differentiated, some conceptually top-down (macroscopic thermodynamic description with nanoscale corrections), others rather bottom-up, starting from an atomistic description. A detailed nano-thermodynamic description, used to derive phase diagrams for different crystallographic structures, predicts silver segregation at the nanoparticle surface [25]. This work established rules to predict segregation based on melting temperatures and surface energies. Similar surface segregation of Ag was found using crystallographic Wulff constructions [113]. Many scientists have described the intermediate size range between roughly 2 and 10 nm diameter with atomistic Monte-Carlo or Molecular Dynamics approaches, some of them combining the two. Whereas sometimes complete alloying is found [114, 115], other authors report a more or less pronounced Ag surface segregation [116, 117]. As a special case should be considered here, simulations predicting alternating layers such as onions [118], or a gold subsurface layer beneath a silver surface [79]. The situation is even more complex for the smallest clusters, as described by Density Functional Theory. Here crystalline structure and magic numbers tend to dominate other effects, and charge transfer is claimed to play an important role. Some articles report preferential Ag segregation in the shell [119], in the core [120, 121] or random alloying [122]. Such calculations are also used to rationalize experimentally observed Au surface segregation in Au-rich BNPs [123].

As far as experimental studies are concerned, the overall situation is no less contradictory. In general, two complementary techniques for BNPs fabrication have to be distinguished, physical and chemical methods. Most of the reported work considers wet-chemically prepared, and functionalized AgAu BNPs [124]. Here all sorts of structures, alloyed vs. Ag@Au vs. Au@Ag, can be produced and even interconverted [125, 126, 127, 128, 129, 130, 131]. These extensive works show that the choice of synthesis sequence and especially of the surfactant dominates the presence of one or the other metal at the interface [132, 133]. For the physical techniques we can cite sequential pickup in helium droplets [67, 134], laser ablation and condensation [23, 123] or magnetron sputtering/gas aggregation [26]. Here the two main reasons for apparent contradictory structural observation are metastability and reactivity. Depending on the fabrication mechanism, the BNPs may be kinetically trapped in metastable structures.

Examples are Ag@Au and Au@Ag core-shell BNPs grown in He droplets [67] or random alloy structures in rapidly quenched CoPt BNPs, which can be thermally relaxed into ordered L_{10} structures [135]. So even though several studies hint at complete alloying for surfactant-free AgAu BNPs, this mixing might be due to the rapid quenching of the metal vapor. The second process to be considered for such physically prepared and thus surfactant-free BNPs is reactivity. Bare metallic AgAu BNPs rapidly react with oxygen upon exposure to air, leading to segregation of Ag at the surface and, eventually, to a complete separation of silver-oxide from the remaining bare Au cores [26]. So, all in all, the question of whether AuAg on the nanometer scale, relaxed to the ground state and free from dominating environmental influences, intrinsically form alloys or segregates to a more or less pronounced degree is still unresolved, both from the experimental as well as from the theoretical point of view.

Here we want to contribute to this discussion in two ways. The first is the implementation of quantitative analysis and machine learning signal improvement by EDS-STEM, leading to the precise knowledge of not only where are the atoms but how many of them are in which region. This is a significant improvement with respect to purely qualitative chemical mapping, as it is widely used in various communities.

In figure 8.1, Guisbiers et al. [25] show a clear example of Ag segregation for a 9 nm AuAg BNP using both HAADF and EDS images. The authors argue that the Ag atoms go to the surface due to their minimum surface energy. However, their qualitative methodology does not allow us to learn the amount of segregation, and it is not very statistically representative since in electron microscopy, very often, single objects are analyzed to describe the entire sample. For example, in a second BNP analyzed by HAADF and EDS in the same work, the Ag segregation is much less clearly visible. To improve the statistical quality of chemical composition analysis of BNPs, the quantitative approach becomes essential. Secondly, if successful, we wish to extend our studies to carbon-protected and annealed BNPs. Here we combine, on the one hand, the fabrication of surfactant-free but protected AuAg BNPs, annealed in order to provoke structural and chemical relaxation with, on the other hand, extensive sample characterization based on the quantitative chemical analysis of and inside individual BNPs.

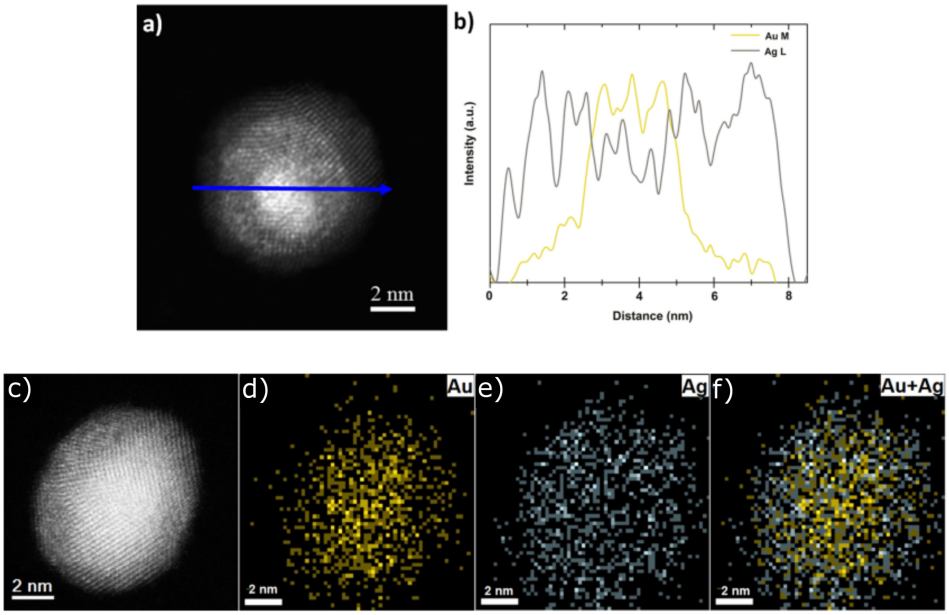


Figure 8.1: a) HAADF image of a clear core-shell AuAg BNP and b) its respective EDS profile with Au intensity in yellow and Ag intensity in grey. c) Another example of AuAg NP, but now the core-shell pattern is not evident. EDS results for the mixed BNP, with Au in yellow whereas Ag in gray. d) Au map, e) Ag map, f) Au an Ag maps overlayed.

8.2 Effects of oxidation in the elemental distribution inside AuAg NPs

In section 8.1, we introduced that physically synthesized AuAg NPs show different elemental distribution patterns according to metastability or chemical reactivity. In the case of reactivity towards oxidation, it is known that Ag severely segregates over time. Therefore, differences in the chemical composition in the core and the shell of the AuAg BNPs are expected. However, the degree of Ag segregation is not easy to predict and increases with time [26]. In figure 8.2, we show a High Angle Annular Dark Field (HAADF-STEM) image in atomic resolution of our AuAg BNPs, suggesting the existence of Ag oxide shells around it. The amorphous shell region of lower contrast suggests the core-shell AuAg@Ag_xO morphology for the BNPs. Looking at a big BNP (Diam. ∼6 nm), we can see a round NP with facets and a complex crystalline structure in the core, with a lower contrast shell of approximately 1 nm of thickness. Smaller BNPs are also shown in the corners of the image, on the bottom left, a small BNP of ∼ 2 nm of diameter clearly shows the facets and crystalline structure. The same amorphous region around the NP can be observed, and the surface fraction here is even more important, revealing that the oxidation may be even more relevant for the smallest BNPs.

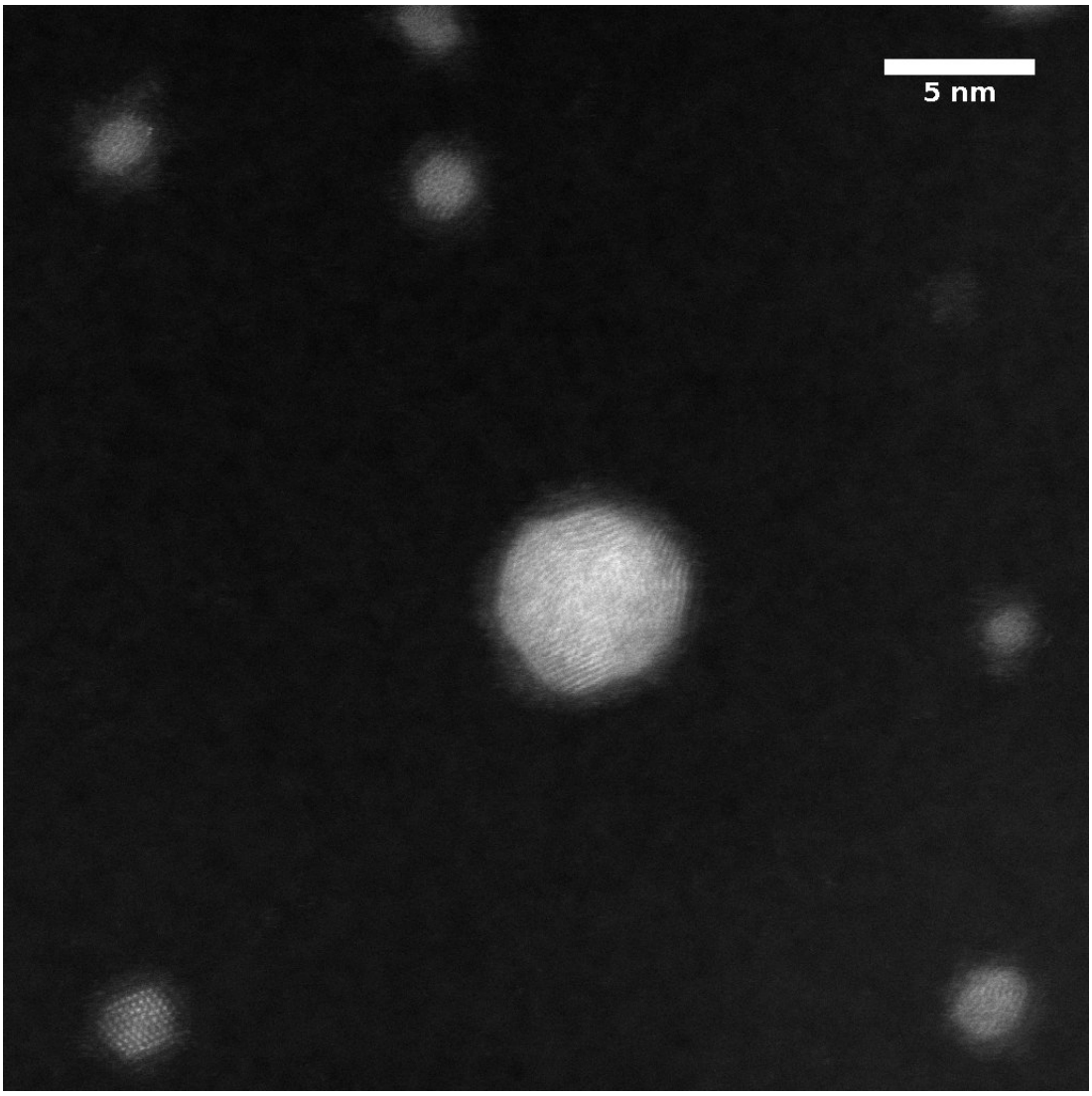


Figure 8.2: HAADF-STEM image of few AuAg BNPs with a silver oxide shell around a possibly alloyed core.

HAADF-STEM images offer the possibility to make images with Z-Contrast. However, spectroscopy in this context is a much more helpful tool to determine if the core and shell regions contain pure Au/Ag or mixed patterns. Hence, we can use Energy Dispersive X-ray Spectroscopy in Scanning Transmission Electron Microscopy (EDS-STEM) to learn where the atoms are and their amount inside the BNPs. This can be done by selecting, for example, pixels from different regions of the image and analyzing them quantitatively.

We used a Titan-Themis operated in STEM mode to acquire Dark-Field and EDS hyperspectral images (HSIs) under different experimental conditions as compared to the HAADF-STEM in figure 8.2. The detector used was a Silicon Drift Detector — (SDD) Super X Quad, 0.8 sR.(LNNANO, Campinas, Brazil). The acquisition parameters were optimized to generate reliable statistics. In order to reduce beam-induced modification of the sample due to knock-on damage, and thereby composition changes [62], the acceleration voltage was set to 80 kV [65]. The probe diameter was between 0.7 - 1.0 nm, beam current between 300-500 pA, and dwell time between 200-400 ms, with a

pixel size of 0.5 nm. The quantitative analysis of the EDS data followed the Cliff-Lorimer approach described in chapter 4, where the calibration constant was determined as $K_{AuAg} = 0.85 \pm 0.01$.

In order to verify if EDS-STEM can quantitatively measure the Ag segregation, we can select a region of interest in the HSI, defining a shell and a core for the BNP, as illustrated by figure 8.3. With $r = (2/3)R$ as the radius of the core and shell thickness of $(1/3)R$, we integrate the pixels from the shell and obtain a representative EDS spectrum for this region. We chose this amount of pixels in order to optimize the signal-to-noise ratio and spatial resolution relation obtaining statistically reliable information from the BNP shell. Processing the spectrum as described in chapter 4, we obtain the graph of figure 8.3, where 37 NPs are quantified to obtain an average value of 0.62 ± 0.01 in Ag atomic fraction in the shell. The average overall NP composition is given by the dashed line showing that, indeed, Ag enrichment can be detected and quantified in the outer shell region of the BNPs. Although we have 2D projections of 3D objects in the EDS-STEM measurements, we can quantify the projected region of the defined shell, confirming that our Ag segregated model suggested above is reasonable.

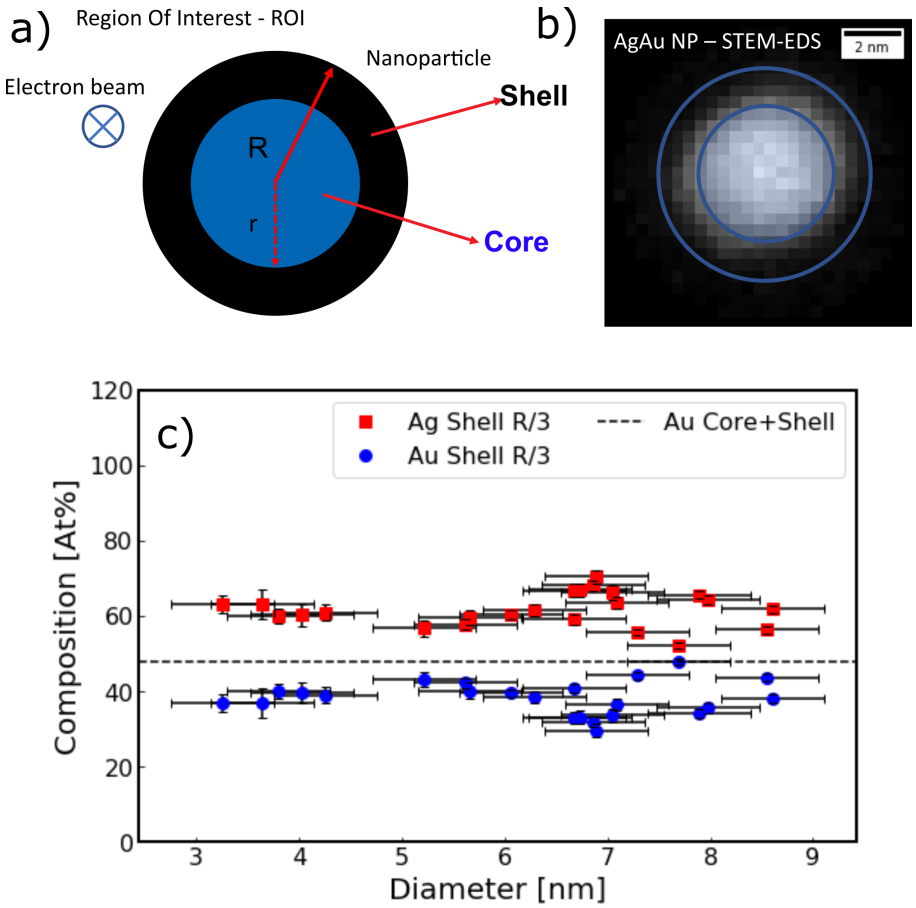


Figure 8.3: a) Schematic of the projected image of the BNP. The blue region defines the core, the black the shell, and the cross mark illustrates the direction of the electron beam entering the page. b) Representative AuAg EDS image with the ROI delimited. c) Results for the shell ROI chemical composition quantification.

Here we can come back to the discussion from the section 8.1, where several authors claim different elemental distributions for AuAg BNPs. Some authors show Ag segregation as thermodynamic and energetic consequences [25, 79], but two important are not discussed in their work. First of all, oxidation must be taken in account as it leads to Ag enrichemnt in the shell as we see in our images consitently with the litarature [26]. Secondly, the alloying of these systems through heat treatment also plays a fundamental role in the elemental distribution of BNP's physics. Using helium sequential droplet sources, Ag@Au and Au@Ag BNPs can be produced [67, 134]. This way, Lasserus et al.[67] reported that by annealing AuAg core-shell BNPs the elements favor mixing, and even a diffusion constant as a function of temperature is determined. Before this work, the inter atomic diffusion of Au/Ag in BNPs in this size range was never observed. This implies that metastable structure must be taken in account. Thus, in order to study the intrinsic mixing behavior of Au/Ag on the nanoscale, it becomes fundamental to protect our BNPs from oxidation and perform heat treatment in the samples in order to relax them to the ground state prior to the EDS chemical analysis.

An efficient way to protect metal nanoparticles against oxidation is by depositing a thin carbon layer (few nm) which retards the oxygen diffusion towards the AuAg NPs [135]. At the moment of the present study, the hollow cylindrical magnetron sputtering source allowed us to produce only oxidation unprotected NPs. To prepare BNPs in ultra high vacuum and to protect them in-situ with a thin carbon film, we fabricated new samples using a different source. We used a laser vapor source [23] where the BNPs are produced by gas phase aggregation, but here a laser is used to evaporate metals in a plasma. This way, we grew AuAg BNPs of 50:50 Au/Ag atomic fraction. The BNPs were deposited fragmentation-free on TEM grids, and we evaporated carbon filaments to produce the thin carbon film ($\sim$ 7 nm of thickness). We cannot, however, discard the effects of structural metastability of the BNPs as they are rapidly quenched in a supersonic expansion, stopping the growth. Their structure can be "frozen" without necessarily being in its global minimum energy conformation, as seen in figure 8.4. In this situation, most NPs show complex ramified crystallography [136] with disordered grains in a shape-like similar to beans. It is essential to note that the outer shells are not observed in this case, which reveals that the BNPs are at least much less oxidized compared to the BNPs from figure 8.2.

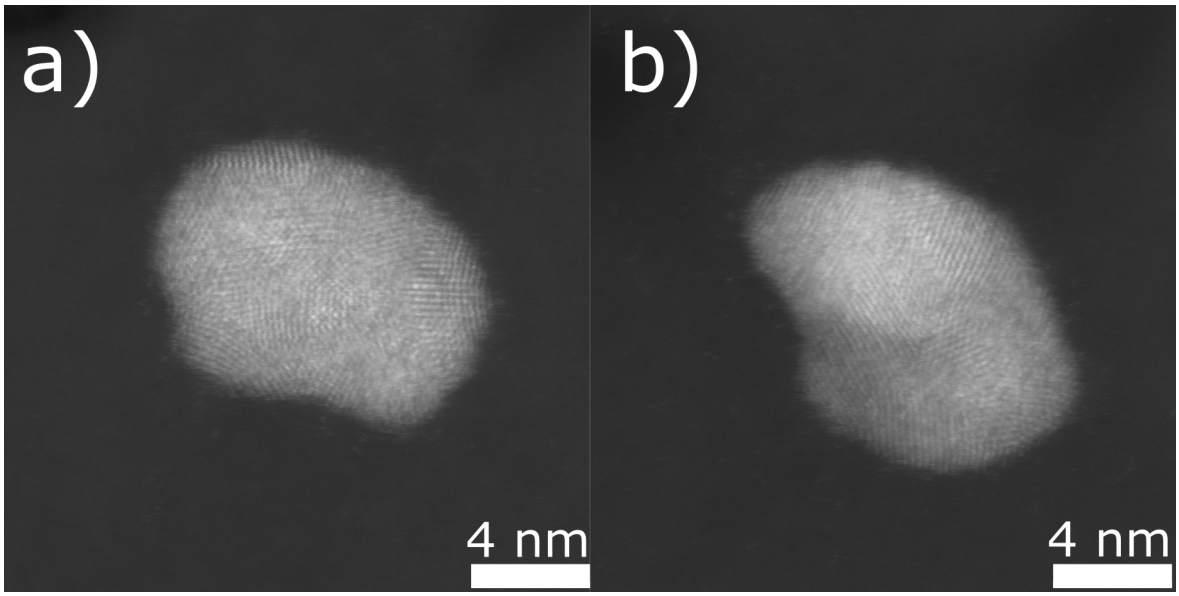


Figure 8.4: HAADF-STEM images of carbon proteced AuAg BNPs.

Considering the metastability, we first analyzed the BNPs carbon-protected as prepared and then annealing at of 400^{o}C for 2h. After annealing, the sample cools down slowly over 2h, relaxing the system to its fundamental atomic arrangement/structure [135]. To estimate the annealing temperature, we used the work of Lasserus et al.[67], choosing a temperature where they observed mixed AuAg BNPs of 5 nm. Some images of the our annealead BNPs can be seen in figure 8.5, evidencing that the particles are much more spherical and better crystalized than the as-prepared ones. However, the BNPs show is this preliminary results are not in the same size range, the annealed ones are more or less 5-7 nm in diameter, while the ones in figure 8.2 are about 8-9 nm. The different sizes may prejudice the comparison between them, and more images of the non-annealed sample are required for the completeness of the analysis.

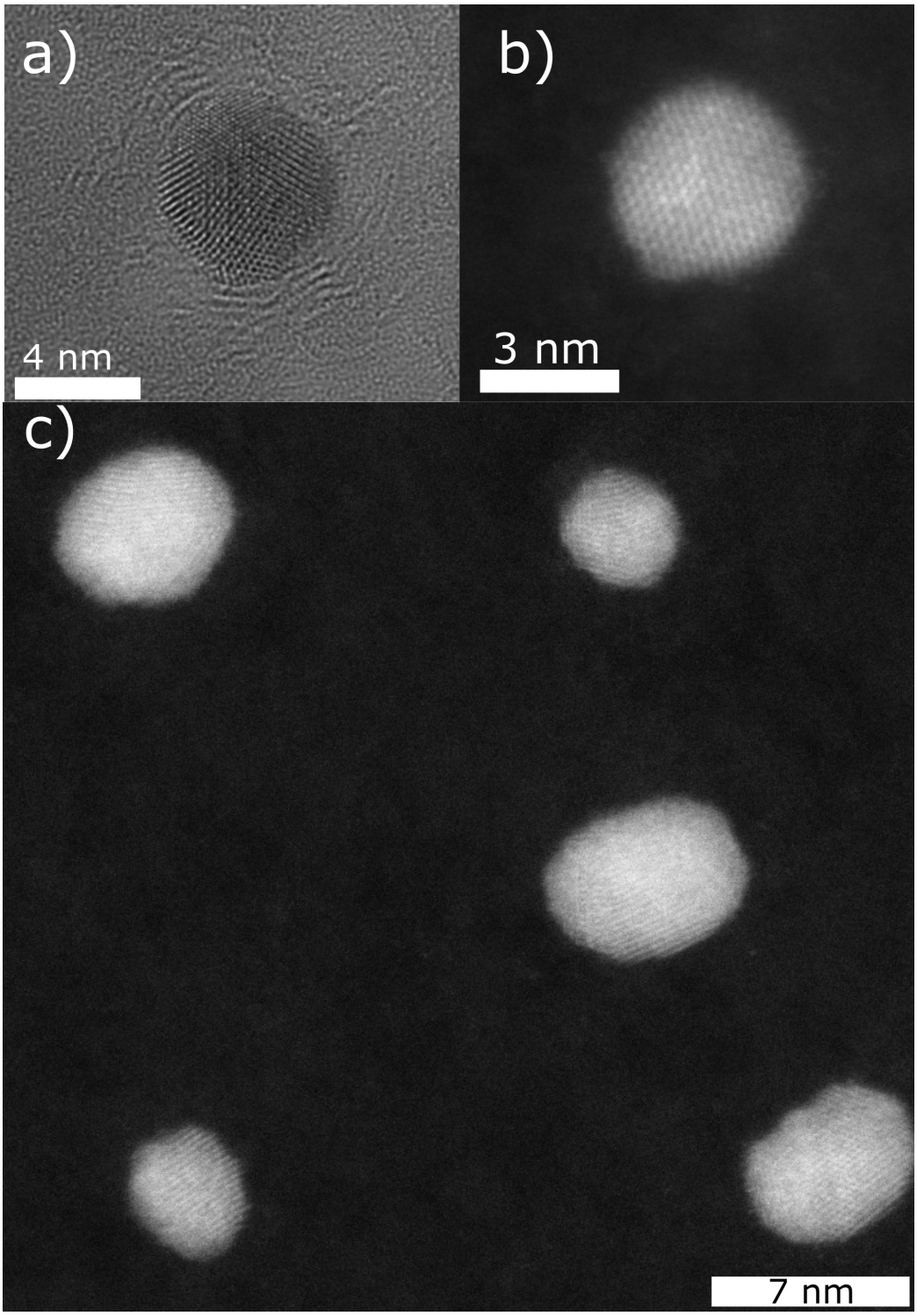


Figure 8.5: HRTEM and HAADF images of annealed carbon-protected AuAg BNPs showing rounder shapes.

Returning to the investigation of the elemental distribution in the protected samples, we quantify the chemical composition of individual BNPs through EDS-STEM measurements. The acquisition parameters were optimized again for two different microscopes, the first one being a JEOL NeoArm at Hubert Curien Laboratory, St.

Etienne University, France. The probe diameter was between 0.3 - 0.5 nm, and beam current $\sim$ 1nA. NPs have been measured with a pixel size between 0.2 - 0.5 nm and dwell time between 50 - 200 ms, resulting in integrated Au counts at 2,000 - 10,000 for 6 nm NPs after acquiring for tens of minutes. The quantitative analysis of the EDS data followed the Cliff-Lorimer approach described in chapter 4, where the calibration constant was determined as $K_{AuAg} = 0.69 \pm 0.01$. The second microscope was again at LNNano-Campinas, Brazil. For this one, the pixel size was 0.1 nm, the probe size estimated as $\sim$ 0.2 - 0.3 nm, beam current of 100-200 pA and 200 ms of dwell time, 128x128 pixels yielding 20,000 Au counts for 6 nm NP in 50 minutes. The data were processed using HyperSpy [84]. We do not detect compositional changes beyond the uncertainty level during the measurement, due to the choice of acceleration voltage (80 kV) [62]. No significant Ag loss is expected for the acceleration energy of the measurement since it is lower than the sputtering threshold energy of the two elements [64].

Consequently, for the mean chemical composition of the BNPs we obtained the value of 0.510$\pm$0.002 Au atomic fraction, which corresponds to the expected 0.5 nominal value of the metal target used in the cluster source. The compositon of individual BNPs, in function of size, can be observed in figure 8.6. Selecting a region of interest for the core and another for the shell, we can integrate all the pixels and generate a representative spectrum for the two regions just like it was done for the oxidized sample. Defining $r = (2/3)R$ and R, for the two regions, we can study the chemical composition of the shell region and compare it with the full NPs value.

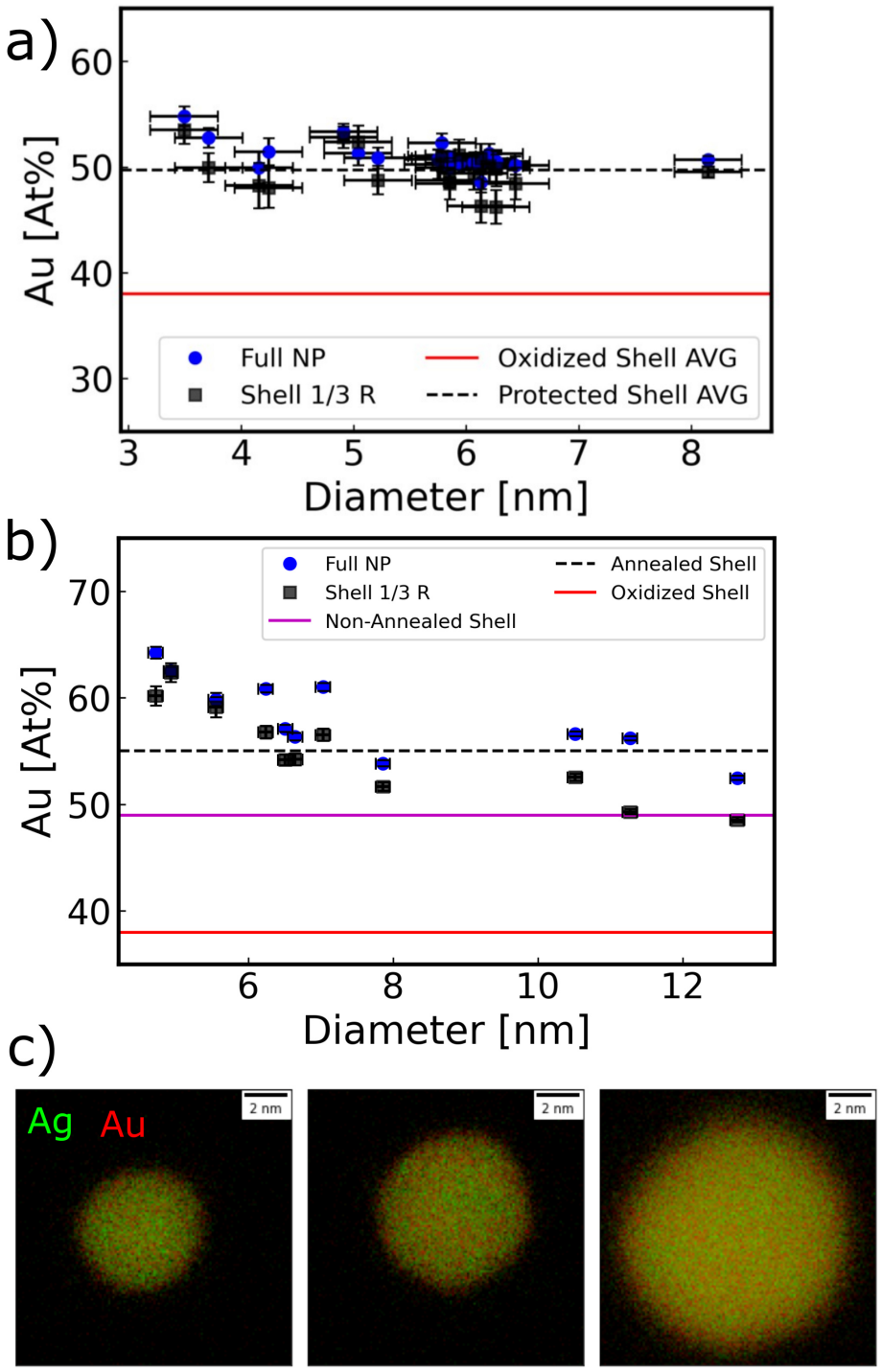


Figure 8.6: Quantitative analysis of AuAg BNPs comparing the full chemical composition and the value of its shell. Results for samples with a) carbon protection and as-prepared and b) carbon protection and annealed at 400ºC. C) EDS elemental mapping of three AuAg BNPs of representative sizes in the sample. In red and green, Au and Ag intensities, respectively.

In figure 8.6 a) and b), we show the chemical composition quantified for the full BNPs and the shell in a carbon-protected sample, before and after annealing, respectively. The

results obtained are put in contrast with a red line which represents the average value for the oxidized samples of figure 8.3, analyzed similarly. Thus, it is visible that when the AuAg NPs are carbon protected, the Ag segregation towards the surface is strongly reduced, and the system exhibits a high level of miscibility. In figure 8.6 c) we show AuAg maps of the annealed sample, which consistently reveals the miscibility of the elements. However, after the heat treatment, we observed Ag loss in the small BNPs (Diam. < 6 nm), which means that the annealing temperature was probably too high, and we may have ejected Ag atoms. This was already observed in the literature, where Co leaves small CoPt NPs, increasing the size of others by Ostwald ripening effect [137]. In our case the BNPs are carbon protected; therefore, the Ag atom movement towards other BNPs is not favored. Selecting, bigger BNPs (~ 8 nm) we find a chemical composition closest to the average expected value. For the big NPs, Ag atoms evaporation is reduced, because the melting point is reduced with BNP size [36]. The high miscibility observed for the big annealed AuAg NPs suggests that the system ground state favors the miscibility of the atoms with only a few percent of Ag enrichment in the shell. However, these results are preliminary and and further work is required. For example, we must analyze new samples with a lower annealing temperature, where Ag loss is no longer observable. For that, we propose, in the continuity of this work, to produce AuAg BNPs with controlled size distribution to optimize the annealing time and temperature for big (Diam. > 5 nm) and small (Diam. < 5 nm) BNPs separately. This strategy should increase the quality of the analysis since the optimum temperatures for distinct size regimes might be different.

The results obtained here show that despite the need for new samples annealed under appropriate temperatures, we found very high miscibility in relaxed AuAg BNPs. The average Ag atomic fraction in the shell for carbon-protected samples is 0.503 ± 0.003 while the full NP value is 0.490 ± 0.002, illustrating a strongly mixed system. In the carbon-protected and annealed sample, the average result obtained for the full NPs is 0.417 ± 0.001 and for the shell is 0.450 ± 0.002. Comparing both samples, despite of the Ag loss of the BNPs, it is possible to see a slight increase in Ag atoms in the defined shell. Thus, by performing these quantifications, we expect to contribute to understanding the elemental distribution inside AuAg BNPs synthesized by the physical route.

The whole reasoning above neglects the crystalyne structure on the BNPs, we can see however, even with our preliminary data (HAADF-STEM images) that shape and crystalinity change upon annealing. It is important to mention that, for the evaluation of the shape/morphology and structure of single NPs, the HAADF images are not a statistically sufficient representation of the whole sample. In order to evaluate the impact of crystalline structure on the chemical order, the proper technique would be electron diffraction and the use of Pair Distribution Functions (PDF) to obtain average and statistically reliable information on the samples [61]. Nonetheless, this experiment is beyond the scope of this thesis (measurement of elemental distributions by EDS chemical analysis).

Moreover, in the context of this thesis, we immediately may wonder how the advanced statistical methods of machine learning can help us to measure elemental distribution. In chapter 6 and 7 we show the advantages and drawbacks of using Principal Component Analysis (PCA) as a denoising method for quantitative EDS chemical analysis. Now, for analyzing chemical phases and elemental distribution inside BNPs, we will use Non-Negative Matrix Factorization (NMF) as unmixing signal tool. This is relevant to try to

find physical/chemical denoised information in the decomposed components, where this method is more advantageous than PCA (see chapter 5). Accordingly, we may be able to extract core and shell-related information quantitatively without the need to determine handly these two regions in the projected 2D HSIs, without previously assuming the morphology of the chemical phases.

8.3 Non-Negative Matrix Factorization for AuAg NP studies: Measuring chemical gradients

The AuAg BNPs, although one of the simplest systems, exhibit different elemental distribution patterns in different situations, such as discussed in section 8.1. In this context, the challenge of characterizing these systems quantitatively with high precision and accuracy is essential to properly learn the fine details of the sample. As we already discussed in chapter 5, Principal Component Analysis (PCA) is one of the most popular machine learning tools applied to electron microscopy (EM) data sets. Its powerful denoising properties are shown in chapter 6 and the validation of the methodology according to confidence interval definitions is discussed in chapter 7. However, as discussed in chapter 5, PCA does not offer physically reliable components. Other methodologies arise as complementary options to extract information of sample features from the components. Using Non-Negative Matrix Factorization (NMF), we propose using blind source separation by unsupervised machine learning [13]. This method works as a way of extracting hidden correlated information of the HSIs (hyperspectral images) of AuAg NPs, obtained by Energy Dispersive X-ray Spectroscopy and Scanning Transmission Electron Microscopy (EDS-STEM). In NMF, similarly to PCA, the data set is decomposed into orthogonal components; however, the components are always positive. These constraints allow us to retrieve more reliable and interpretable information in the components [14]. The main drawback of using NMF is that the algorithms typically used are not convex, which means that in some situations, the solution of the factorization is not convergent. In the chapter 5 we discuss the work of Uesugi et al. [112] where the authors show that NMF does not converge for a high number of components as output. Therefore, we used our experimental data and simulations to run the algorithm several times, and we obtained the same results in each of them. This allowed us to verify that the methodology is robust for the number of components we needed to extract information from the EDS-HSI of AuAg BNPs.

Due to the known elemental segregation in oxidized AuAg BNPs, we chose the data set from experiment III in chapter 6, to start the analysis with NMF. We have used a Titan-Themis operated in STEM mode to acquire Dark-Field images and EDS HSIs. The detector used was a Silicon Drift Detector—SDD) Super X Quad, 0.8 sR.(LNNANO, Campinas, Brazil); the EDS HSIs were acquired using a 0.5 nm pixel size and dwell time of 200 ms per pixel; the image size was set to 64×64 pixels. In order to minimize the dose rate, a series of images of the same region was acquired, and, subsequently, individual frames were added (10 scans at 20 ms dwell time) [138]. For all experiments, we have used a low-background Be sample holder. The raw EDS spectra were binned to 512 energy channels to increase the signal-to-noise ratio (SNR) for subsequent PCA/NMF calculation, leading to an EDS channel width of 20 eV. The Au-Mα and Ag-Lα EDS

counts from individual particles have been obtained by summing the counts from pixels inside each NP region. Each 64 × 64 pixel scan contained about 5–8 NPs, and we stacked 4 HSI together to increase the quality of decomposition methodologies by increasing the number of pixels [103]. The quantitative analysis of the individual NP chemical composition followed the Cliff–Lorimer [8] approach using the experimentally measured K factor ($K_{AuAg} = 0.85 \pm 0.01$) described in chapter 4.

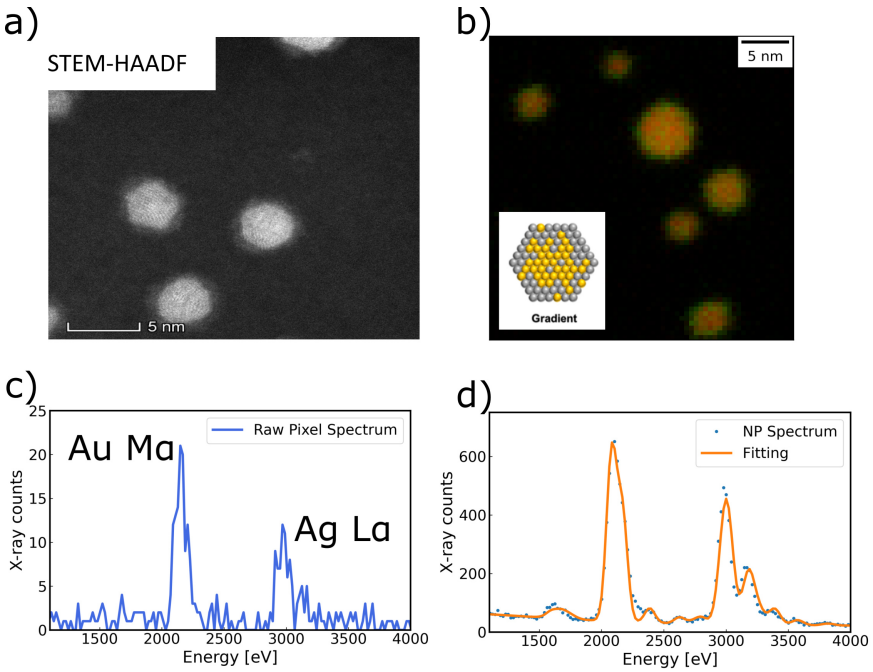


Figure 8.7: a) HAADF-STEM image of AuAg oxidized NPs. b) EDS-STEM of AuAg BNPs and inset figure of the model suggested by the qualitative measurement. c) Spectrum from a single pixel in the NP. d) Integrated spectrum for a single NP and the respective fitting.

Recapitulating from the last section, as can be seen in High Angle Annular Dark Field Images (HAADF), the NPs systematically show a ring of lower contrast around them (see figure 8.7). This is associated with oxidation of the BNPs, since these are not protected from the atmosphere during their transfer to the microscope. Sequentially, in the figure, we show the elemental mapping of AuAg NPs, which reveals Au and Ag atoms in every pixel of the NPs, suggesting a mixed alloy morphology for our model system. So, on the one hand, we have a core-shell-like morphology and, on the other, a mixed system. We also show in figure c) a single pixel spectrum and in d) an individual BNP spectrum to evidence the quality of the data set. This is crucial to demonstrate that, as described in chapter 6 and 7, a minimum signal-to-noise ratio (SNR) must be ensured for the level of information we want to extract from the HSI (chemical heterogeneities inside small BNPs).

It has already been observed that AuAg NPs oxidize, and, as a consequence, Ag atoms segregate at the surface [26]. Due to the short time between NP synthesis and microscopy, the segregation is only enough to form a core-shell structure and not a completely segregated Janus one. Analyzing Z-Contrast images, we see the core-shell-like

morphology of the BNPs. Nevertheless, we cannot quickly learn the elemetanl distribution precisely from this type of data since we cannot obtain the chemical signatures of the two phases with this technique. The most logical elemental distribution, thinking in atomic mobility and diffusion, would be the gradual enrichment of Ag atoms towards the surface, as illustrated in the inset of figure 8.7, and this is consistent with the qualitative elemental mapping.

Despite the fact that we have 2D projections of 3D objects in the EDS-STEM measurements, we can quantify the projected region of the shell defined by hand (see figure 8.3 in section 8.2), confirming that our model of Ag segregation suggested above is reasonable. Besides, we can improve our analysis using NMF as BSS for unmixing signals. This way, we expect to see the core and the shell separated in both spectral and spatial domains, as suggested in [13, 14]. First, we apply PCA to decompose the data set and study the proportion of variances, following the elbow rule. We also studied the Potapov diagrams [96](see chapter 5) to determine the number of components that correctly describe the data set. This pre-processing is crucial for the NMF analysis because the number of components is an input given to the algorithm. Knowing the proper number of components is a challenge for the analysis, and the number of components describing the data set must be minimized to guarantee the convergence of the decomposition's solution [112]. We started with 5 components following the PCA scree plot analysis, but, by inspection, we verified three as an optimum number to unmix all the information. More components do not change the results as the last components separate only noise and background signals such as the Si Kα line. In figure 8.8, we can see the PCA and NMF spectral components illustrating the simpler physical interpretation of the NMF compared with PCA.

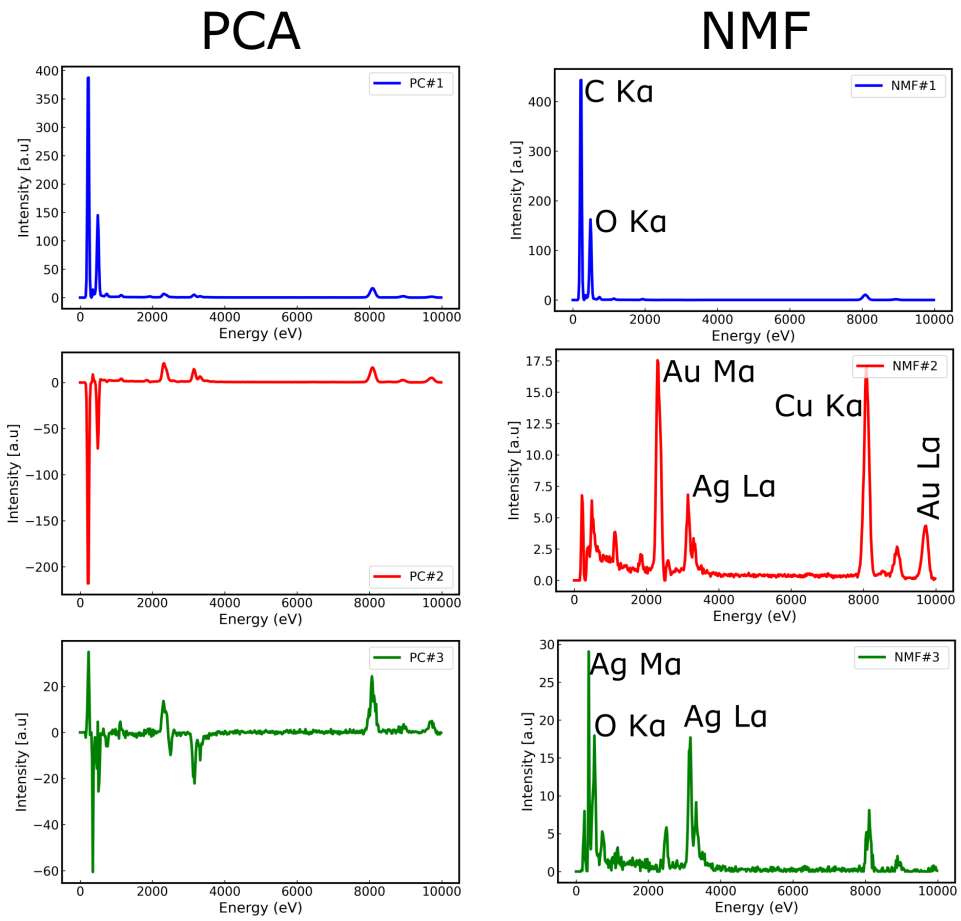


Figure 8.8: Components of PCA in the left column and of NMF in the right. PCA allows negative signals while NMF allows only positive ones. The characteristic X-ray peaks relevant for the analysis are labeled in the NMF components and are valid for the PCA too.

In PCA, we separated the information with unrealistic negative counts. The first component can be associated with the background information, the second with the average chemical composition, and the third with the variations of the chemical composition resulting in negative and positive Au and Ag peaks (see chapter 6 for more details). The NMF first component again shows the background content, the second only has Au and Ag content, and the third only Ag content. This suggests that the third component describes the variations in chemical composition, just as in PCA. Here we can perform some mathematical reasoning to understand how the information is split into the components for PCA and NMF. In both cases, for simplicity, in each pixel i, we can imagine the I_{Au} and I_{Ag} signal composition as a system of two linear combinations. Assuming that the whole BNP information is contained in components 2 and 3:

$$I_{Au,i} = s_{2,i} * I^2_{Au,Aver} + s_{3,i} * I^3_{Au,Var} \tag{8.1}$$

$$I_{Ag,i} = s_{2,i} * I^2_{Ag,Aver} + s_{3,i} * I^3_{Ag,Var} \tag{8.2}$$

where the $s_{2,i}$ and $s_{3,i}$ are the i-th pixel intensities of the score matrices of the

NMF#2 and NMF#3 respectively. $I^2_{Au,Aver}$ and $I^3_{Au,Var}$ are the intensities of the average and variation of the Au signal, associated with the second and third components, respectively. Thus, the intensity of gold and silver in each i-th pixel is built by two "pieces" of information. An average chemical composition comes from the NMF#2, and the distribution/variation in composition comes from the NMF#3. This logic is similar to the concept observed for PCA in chapter 6, where the third component, with oscillating Au and Ag peaks, is responsible for the information associated with the sample's chemical composition dispersion. Indeed, the information is the same, and how it is organized in the components changes, with NMF forcing everything to be positive with the second component accounting for all the Au in the entire particle and the according minimum content of Ag. The third component then describes the additional Ag content and thereby, their relative fluctuations. This changes how the different decomposition methods find the correlated information and split them into components.

Looking at the scores (image of the weights of the linear combinations, see figure 8.9), it is possible to visualize that the first component is spatially associated with the background, the second with a round NP shape, and the third with a larger round shape. This is consistent with the model of a core-shell-like NP, where the core would be the mixed AuAg nanoalloy and the shell Ag oxide. Along the same line, Rossouw et al. [13] discuss the same unmixing strategy by BSS applied to core-shell NPs (see chapter 5 for details), which influences our interpretation of separating a core and shell information. However, in their case, the core and shell regions are much better defined and do not show interdiffusion of elements, which simplifies the analysis of splitting the information in core-shell morphology. As we will continue discussing, the AuAg system is a more complex situation for blind source separation to resolve, and the the shell, suggested by the HAADF images and by the NMF#3, can be either composed of an AuAg alloy or of pure Ag, forming completely segregated chemical phase. In the first case, the elemental distribution would form a gradual Ag segregation generating a chemical gradient inside the BNPs. To verify which of the two possibilities we have for our system of study, in the following discussion, we will explore standard qualitative profile analysis and new quantitative approaches to measure the elemental distribution of these AuAg BNPs. Even though standard analysis of the spatial and spectral NMF components might suggest a core-shell (AuAg@Ag) structure, imagining the two extreme scenarios of core-shell vs mixed structures, we expect to find the real elemental distribution within AuAg BNP somewhere in between.

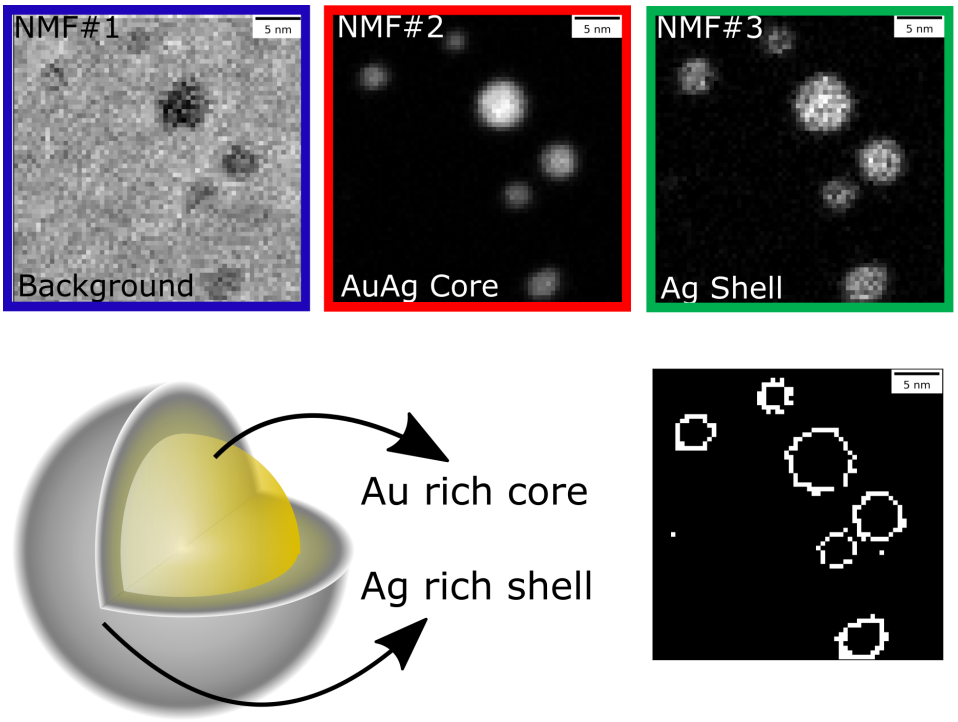


Figure 8.9: Scores of the NMF decomposition ilustrating the spatial localized information unmixed from the HSI. NMF#1 is associated with background, NMF#2 is associated with AuAg average composition and NMF#3 to an Ag excess in the shape of a shell. The model of the system with an Au rich core and a Ag rich shell is shown below. In the lower right the difference of binarized NMF#3 and NMF#2 acordding to Otsu thresholding algorithm [85, 139]

The figure above shows a model of the BNPs with an Au-rich core and an Ag-rich shell. The illustration shows that we consistently have more Au in the core and more Ag in the shell with the spatial and spectral components observations. In addition, we use Otsu thresholding [85, 139] to obtain binary images of the NMF#2 and NMF#3 scores. Hence, the two images are subtracted, exposing the spatially related shell information. Here, we obtain a shell thickness of 1 to 2 pixels, which means approximately 1 nm. In the HAADF-STEM images, the same thickness is measured for the amorphous shell around the BNPs. Quantifying the chemical composition from the experimental NMF#2 and NMF#3 spectra separated, we obtain 0.33 of Ag in the so-called "core" and 0.98 of Ag in the "shell". However, it is still unclear what these values represent, since we are not yet able to associate the NMF components with well defined chemical phases of a core-shell structure.

In order to obtain more information about the elemental distribution we continue our analysis by means of standard profile analysis of a representative BNP from the EDS data set (see figure 8.10). In the raw profiles 8.10 b), we can see that the shape of Ag and Au signals are almost the same when normalized and we have no clear evidence of Ag segregation towards the surface. Using NMF to improve the information retrieval we obtain profiles with a clearer concentration of Ag at the surface as seen in figure c). Therefore, comparing the raw profiles with the NMF#2 and NMF#3, we can verify that whether NMF increases the sensitivity for shell detection or the Ag segregation is not

relevant enough to form a pure Ag shell in the third component. These improved profiles are, however, very different from simulated curves for a perfect core-shell segregation as can be seen in figure d). The simulation was performed for a 6 nm NP with the same SNR of the experiment ($I_{Au} = 2000$ for each NP). Here, two phases were considered in the model, with a AuAg alloy in the core and pure Ag in the shell.

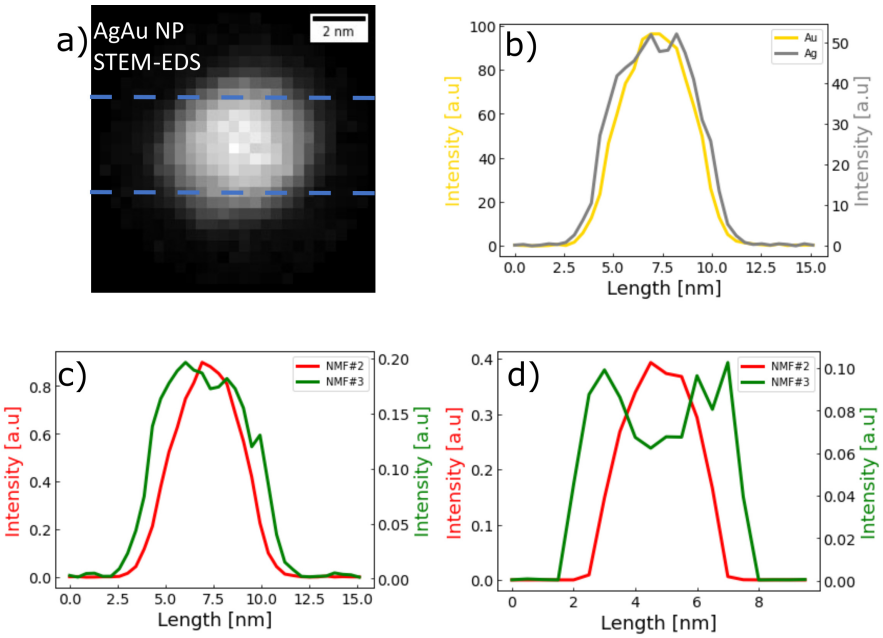


Figure 8.10: a) EDS-STEM image of AuAg NP, b) Au and Ag intensities profile of the pixels between the blue dashed line on a). c) Profile of the scores (NMF#2 and NMF#3) obtained using the same coordinates of b). d) Simulated profiles of perfect core-shell (AuAg@Ag) BNPs using the scores as in see in c).

As discussed above, following the ideal BSS applied on core-shell NPs, we should expect the core information in the second component and the Ag shell component in the third, as is easily observable in the simulated profile in figure 8.10 d). However, in the experiment, we still have to consider that the pure Ag shell may not be thick enough to be distinguished in our measurement with the spatial resolution in the experiment. Another possibility to explain the mismatch between experiment and simulation is the case where the shell is composed of an Ag-rich alloy phase and not a pure segregation. Therefore, the spatial resolution and SNR of our measurement are not enough to easily distinguish the two situations hitherto discussed only qualitatively. A deeper quantitative analysis of the profiles is required to progress in the chemical analysis and answer these questions undoubtedly.

Thus, we use the Cliff-Lorimer method to determine the chemical composition of the profiles. Aiming to quantify the intensity profiles of reconstructed NMF components, we first fit every pixel spectrum and subtract the background. Consequently, we use the intensity integrated from these peaks to form the elemental maps. However, despite the high quality of the data, the intensity is still low for the level of information and precision we must achieve. Consequently, to optimize the SNR of the analysis, we used an azimuthal integration, as the Otsu thresholding shown in figure 8.9 proves our educated

guess of spherical symmetry to be reasonable. Thus, according to figure 8.11, we select rings whithin the NPs with 1 pixel thickness and integrate all the pixels intensities for each ring.

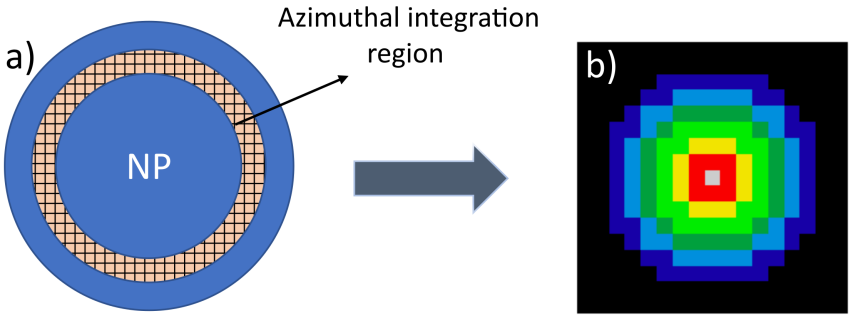


Figure 8.11: Scheme of azimuthal integration of intensities from the BNP. In a) the NP with a ring ilustrating the azimuthal integration in the pixels of interest. In b) the pixelated version of the NP with 1 pixel thick rings. The rings are drawn in different colors to better visualize the regions used to extract signal for the profile quantitative analysis.

Next we chose three representative BNPs indicated in figure 8.12. To obtain a representative profile for each NP, exploring the azimuthal symmetry, we average the intensity in each ring in the HSI with a stepsize of 0.5 nm. In sequence, the chemical composition of the profile is quantified following Cliff-Lorimer method, and error bars are estimated for each point in the graph of figure 8.13. It is important to mention that there is not yet a robust methodology for error bar determination in NMF because the spiked covariance model we proposed in chapter 7 is only valid for PCA. For that reason, we used error propagation of Poisson uncertainty, such as in the raw situation, but considering the intensities anticorrelated due to the NMF decomposition.

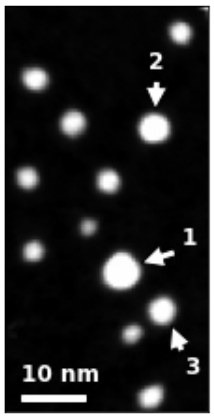


Figure 8.12: ADF image of AuAg BNPs. The numbers indicate the three NPs for which the chemical gradient was determined by EDS-STEM.

The quantitative chemical analysis of the profiles of these three NPs can be seen in figure 8.13, where the first line of graphs corresponds to the results of NP number 1 indicated in the DF image. The second and third lines correspond to the other NPs, respectively indicated in the figure 8.12 . The first column gives us the integrated azimuthal intensities, the second column the profile obtained from the averaged pixels, and finally, the Ag atomic fraction profile for each BNP in the third column. Here, we show that, it is indeed possible to measure chemical gradient within BNPs, providing values for the Ag enrichment radially, pixel by pixel, with relatively reasonable confidence intervals ($\sigma_{Ag} \sim 0.02$).

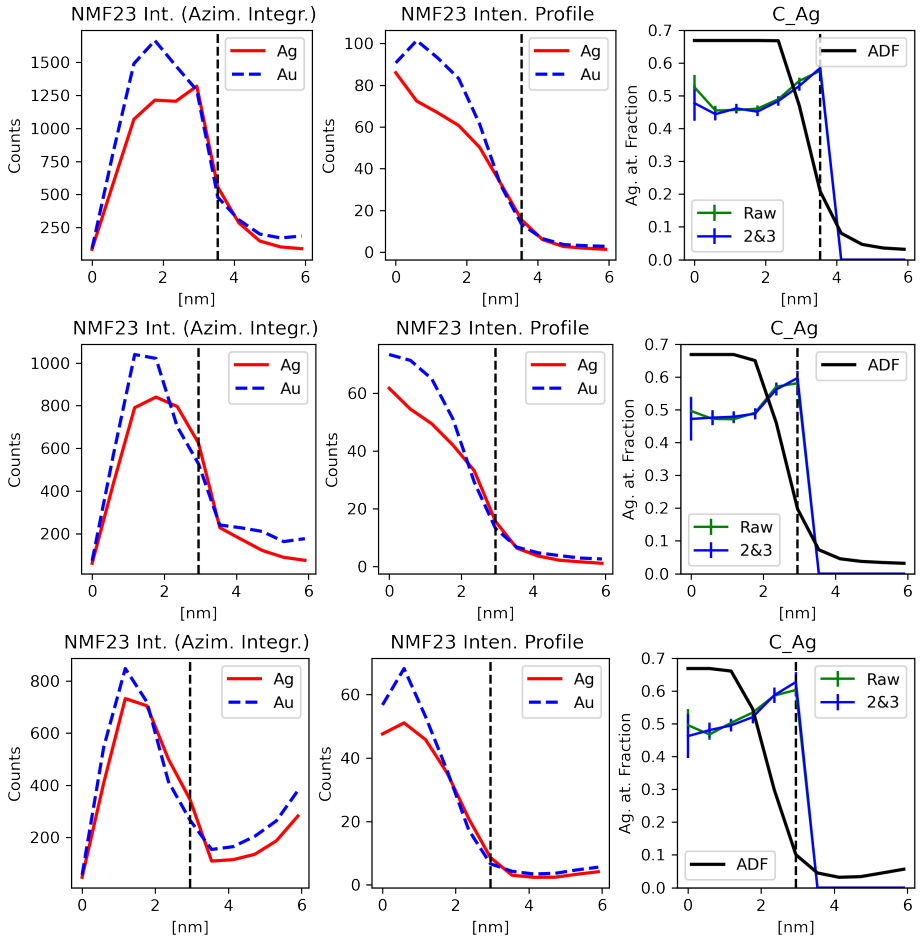


Figure 8.13: Results of the quantitative chemical gradient analysis. BNP azimuthal and profile intensities, excluding the NMF#1 from the reconstruction, are shown in the first an second columns, respectively. In the third collumn the Ag atomic fraction quantification obtained by the Cliff-Lorimer method is displayed. The x-axis shows the radial distance from the center of the BNP.

We have exposed in the graph the raw quantification and the one obtained by the reconstruction of the second and third NMF components. In the atomic fraction, the raw and NMF2&3 curves are similar, with small differences mainly in the center and in the surface of the BNPs, where the statistic is poorest. Even though we cannot unmix the information inside the alloyed BNPs, we can unmix the background from NPs signal.

Moreover, with NMF reconstruction of the relevant components (2 and 3), we learn that the information obtained from the reconstructed curves is not simply fluctuations and carries true information associated with detectable variances. On that account, the atomic fraction measured pixel by pixel is reliable, and error bars are overestimated because we are not considering the denoising aspect of the analysis in the uncertainty estimation. Additionally, excluding the first component, we increase the quality of the background definition, which acts as a pre-processing for background subtraction, increasing the quality of the results by considering only the information of interest in the reconstruction. This would not be possible with PCA, where the optimization leads to some AuAg NP's signal in more components, including the first, where the background is ideally associated. Consequently, in this analysis, we can observe that the chemical gradient can be measured and systematically indicates, as predicted, Ag enrichment towards the BNP surface. The atomic fraction varies from 0.45 ± 0.02 in the center to 0.62 ± 0.02 at the surface. Note that the images are 2D projections of the BNPs and therefore, the pixels in the center of the particle contain information from the shell. The values for the Ag atomic fractions in the center are thus upper limits. All the signal beyond the BNP edge, indicated by the straight line, is below the intensity detection criterium (3x error bar) [9, 10]. Therefore, no spatial distinction of the NMF2 and NMF3 is detectable with statistical relevance, since Au signal can be found as well whithin the Ag-rich surface.

To help in the interpretation of the chemical gradient measurement, a simulation that properly represents the data set must be used. In figure 8.10 we used a core-shell simulated data set assuming that the AuAg BNPs were well-defined core-shell structures (AuAg@Ag). However, here, we must build another simulation to try to extract the chemical gradients from it using the data treatment prosed in this thesis and therefore show the robustness of our method. An adequate simulation in this case would be at least two AuAg alloy distinct phases, with Au-rich in the core and Ag-rich in the shell. To build this data and interpret the chemical gradient measurement, we generated an HSI of 64 AuAg BNPs of 6 nm and 0.5 of average atomic fraction, similarly to the simulations in chapters 6 and 7. We simulated the Ag enrichment towards the surface using quantified experimental parameters exposed in figure 8.13 as input for the model. We used 0.33 of Ag atomic fraction in the core and 0.62 in the shell. To build the simulation, we first simulated core-shell (Au-rich alloy core and Ag-rich alloy shell) BNPs and calculated the 3D sphere's volume projected onto a 2D image. We converted this projected volume in atoms/X-ray intensity and allocated these values in each pixel. Therefore, the intensity decreases from the center of the BNPs to the borders, just like in the experiments, since the X-ray density is conserved. Hence, we used uniform distributions with core and shell Ag atomic fractions as probabilistic weights to simulate the dispersion of chemical composition in each pixel, obtaining the chemical gradient from the center to the borders. It is of utmost importance to underline that Poisson noise is applied to the simulated HSI (2,000 Au counts for each NP). As in the experimental data processing, we first perform PCA to analyze the scree plot. In figure 8.14 we can see that, according to the elbow rule, we have 3-5 principal components in the decomposition of the simulated data set. Then, choosing three components (minimum number to fully describe the system) for the NMF analysis again, we can compare the PCA and NMF components similarly as in figure 8.8.

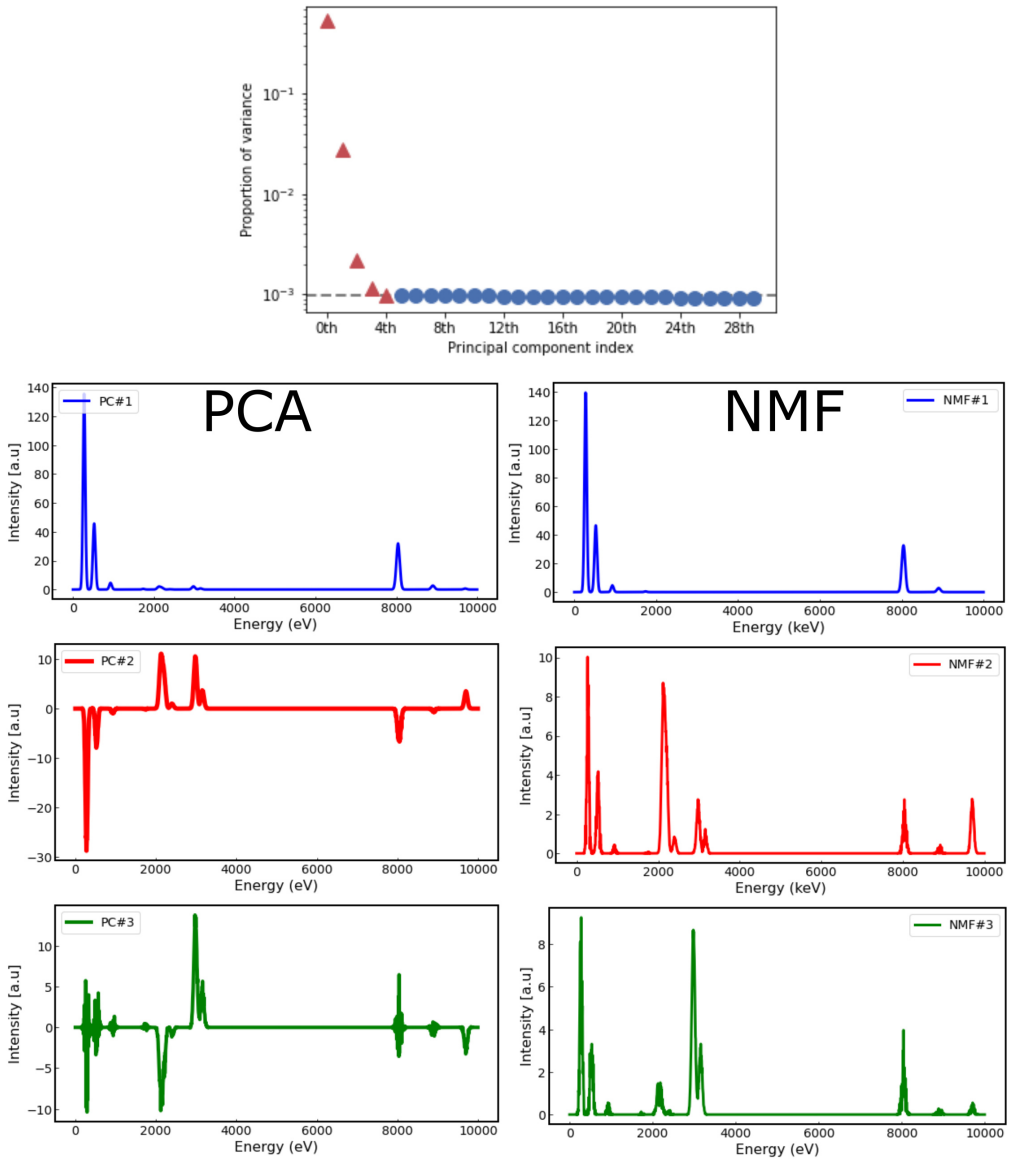


Figure 8.14: Simulated HSI results: Components derived by PCA in the left column and by NMF in the right column. PCA allows negative signals while NMF allows only positives.

Here we can see that the simulated components in both PCA and NMF agree with the measured HSI components. Observing the score images in figure 8.15, we can see that the correlated spatial information in the 3 NMF components agrees with the experiment derived from measured HSI. This is important because here in this simulation we are sure that our BNPs are not simple core-shell structures but a structure with an Ag fraction increasing from the core to the shell with simulated local fluctuations. In the first simulation which was discussed in this section, we considered a completely segregated Ag shell forming an AuAg@Ag morphology (see figure 8.10 d)). In this case, the signal unmixing generates the pure Ag shell in the third component. However, the latest simulated system with alloy in the core (Au-rich) and also in the shell (Ag-rich), shows the same spatial and spectral patterns as it would be for the AuAg@Ag core-shell

system, if analyzed carelessly. If the investigation relies only on experiments without the support of simulations, this misleading information may induce wrong conclusions about the chemical phases of the system analyzed. So the combinations of quantitative profile analysis and simulations is required to unravell the fine details of the chemical heterogeneities inside BNPs. Indeed, the unmixing of signals from inside small BNPs shows itself as a challenging and complex issue. But as suggested before, we can at least unmix the signal from the background and the signal from the BNPs without any doubts. In figure 8.15 we also show reconstructed images using first the NMF#1 alone, reconstructing the background of the HSI. Then, we also reconstruct the NMF23 image with the information from the BNPs without background and with the chemical composition information denoised. In the reconstructed BNPs, the quality of BNP definition is increased, and the process of the method to separate the relevant information from the background is evident.

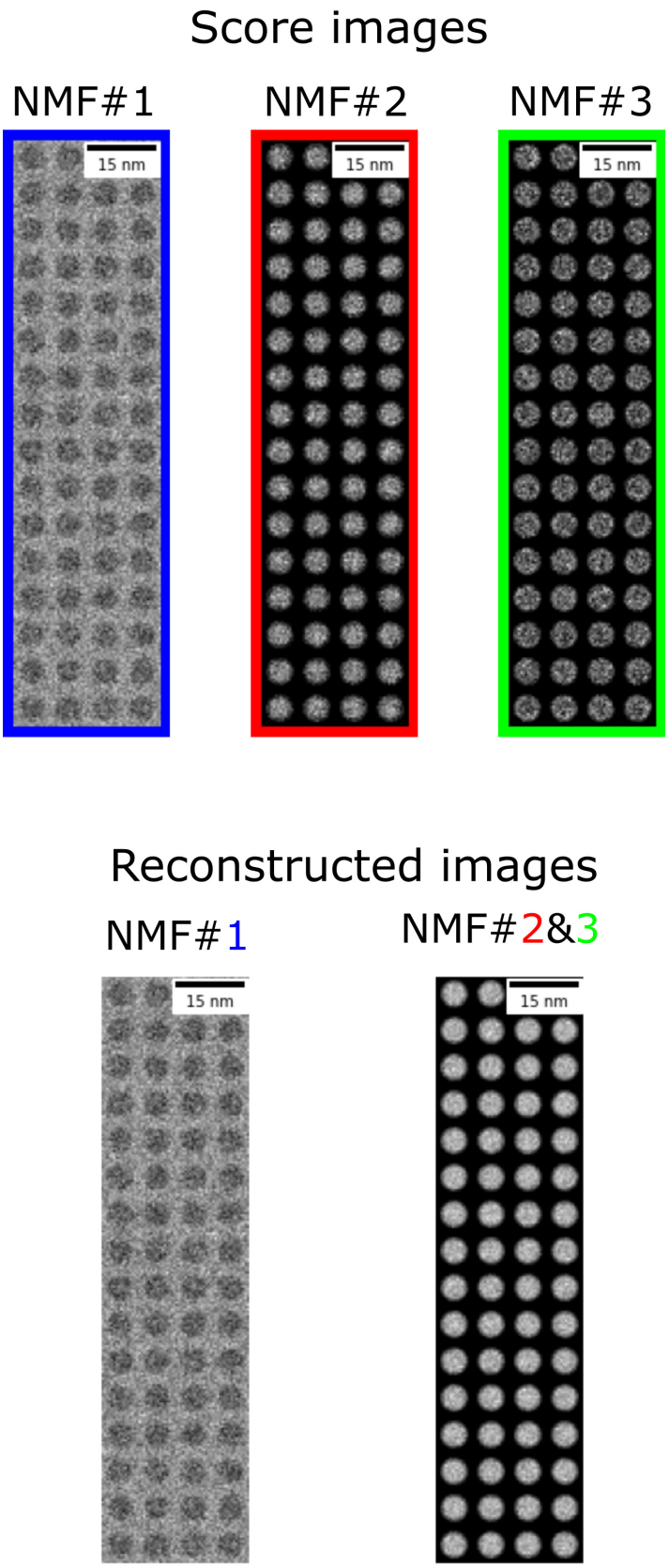


Figure 8.15: Simulated score images of the three NMF components. The firs is associated with background, the second with a core AuAg alloy phase, the third with broader signal similar to a shell with more Ag than Au. Below, the reconstructed images of the unmixed background and BNPs.

Sequentially, we employ the same methodology used for the reconstructed experimental data. We fit every the Au and Ag peaks in every pixel of every image using a linear combination of Gaussians and the background by a polynomial. With background subtraction, we integrate the fitted peaks and generate the elemental maps of Au and Ag separately. Hence, we apply the azimuthal integration in the pixels of a representative NP, and we obtain the intensity profile and its respective Ags atomic fraction quantified as can be seen in figure 8.16.

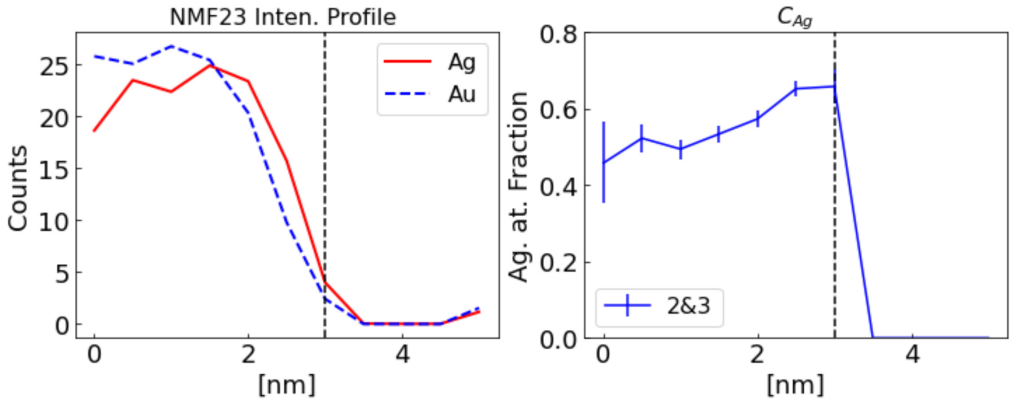


Figure 8.16: Azimuthal integration results for the simulated AuAg BNP. In the left, the intensity profile generated averaging pixels azimuthally every 0.5 nm (1 pixel). In the right corresponding chemical composition quantification following the Cliff-Lorimer method. The x-axis shows the radial distance from the center of the BNP.

In the intensity profile, the red Ag line is more intense in the surface than the Au blue dashed line, which is consistent with our experimental observations. Nevertheless, here, the most exciting aspect of the simulation is the quantitative analysis because we have control of the values expected. Thus, the Ag atomic fraction profile demonstrates that the chemical gradients are correctly measured and that our methodology is robust for this end.

Although we learned in chapters 6 and 7 how to use simulations to estimate information loss/retrieval parameters to qualify the quality of our reconstructions with PCA, the same approach is not so simple with NMF because there is no statistical model such as the spiked covariance model (see chapter 5 for details) valid for NMF. However, it might be possible to explore PCA assuming that if the information is retrievable for PCA, it must also be for NMF. Thus, in the continuity of this work, we will try to establish information loss estimators for PCA in order to establish the detection limits for NMF.

Looking at the experimental and simulated results, although within the method's limit, we can quantify the chemical gradients inside AuAg BNPs, showing Ag enrichment radially in small BNPs of 5-8 nm. This breakthrough result offers the opportunity to investigate the elemental distribution quantitatively by simple EDS-STEM analysis. From the point of view of fundamental nanoscale chemistry and physics, this methodology may offer quantitatively established parameters for further research and investigations in BNPs properties both from theory/modeling or applications, where the knowledge of precise elemental distribution is essential. At this point, we may return to the discussion of the physics and chemistry of the elemental distribution inside BNPs, supported by

our advanced data treatment for chemical gradients quantification. For example, we argued in section 8.2 that the formation of Ag oxide might induce Ag segregation and we showed in the sequence that by protecting the BNPs we can find more alloyed AuAg BNPs. Therefore, the oxidation-protected NPs can be analyzed to verify their chemical profile by evaluating the elemental distribution inside AuAg BNPs. These BNPs, after relaxation obtained through thermal treatment, would be very close to their ground state. In this case, would we expect alloying, Au/Ag segregation, or chemical gradients/order? What we observed is that a small fraction of Ag (a few percent) segregates at the surface (see figure 8.6). Now, with the tools developed in this chapter, we can go further in the analysis and investigate the chemical gradients inside this more challenging system where the amount of Ag segregation is reduced, and the chemical gradient is more difficult to detect by the characterization technique employed in this thesis (EDS-STEM). In the continuity of this work, we will analyze the data sets from section 8.2 applying NMF and the azimuthal integration to obtain the quantitative chemical profile of these BNPS. Despite the fact that the information is much more challenging to be obtained, the SNR for these data sets is much better for a reduced electron probe diameter (0.2-0.3 nm) which can make it possible to obtain the finest chemical information. However, this spatial resolution may be already within the limit of the methodology since, due to electron channeling effects and delocalization a quantitative analysis close to the atomic scale might not be possible at all for EDS-STEM [140, 141, 142].

Finally, we propose that complementary methodologies, such as electron diffraction PDF and quantitative HAADF-STEM, can be applied to go further in the analysis and learn about the chemical gradients/ordering on the atomic scale. The results found here for EDS can be used as support for these characterization methodologies, supplying parameters for simulations and refinements.

8.4 Summary

This chapter aimed to analyze the elemental distribution inside the AuAg BNPs quantitatively. Hence, open questions related to AuAg BNPs were discussed notably whether Au or Ag segregate in different AuAg BNPs according to different articles published in the last years.

In sequence, Ag segregation was observed in the elemental distribution in oxidized BNPs. Then, carbon-protected samples were prepared and annealed to check for an increase in alloying. While the as prepared particle showed complete miscibility some Ag segregation was observe for the annealed one. However, some Ag loss was measured after annealing, suggesting that the thermal treatment was too aggressive and the temperature must be reduced.

Finally, NMF decomposition is explored in the oxidized benchmark AuAg BNPs for elemental distribution quantification. It is shown that the BNPs are not the expected well-defined core-shell systems. However, surprisingly, it is shown that chemical gradients can be measured by quantifying integrated azimuthal intensities forming the chemical profile of the Ag atomic fraction segregation. The NMF denoising aspect of the analysis shows that the chemical gradients quantified are information and not noise fluctuations. Simulations are used to corroborate the quantitative measurement.

Chapter 9
Conclusions and Perspectives

The work developed in this thesis was mainly motivated by the need to perform quantitative approaches in Scanning Transmission Electron Microscopy (STEM) to study bimetallic nanoparticles (BNPs). The physico-chemical properties of nanometric systems, such as BNPs, are directly related to their atomic arrangement, morphology, size, and chemical composition. Therefore, understanding such objects requires knowledge of these characteristics and their influence on the BNP's electronic properties. Despite the helpful employ of qualitative analysis typically used in various communities, the quantitative approach provides much more detailed and rich information about their nature. In science, we should be quantitative in our analysis and conclusions whenever possible. Because doing so, it is possible to exclude models and correctly interpret our observations. Nowadays, quantitative analysis in electron microscopy has become possible due to the advances in its software and instrumentation in the last ~ 20 years. Additionally, machine learning rises in nanoscience and TEM data treatment as a powerful tool to help design experiments and extract hidden features from the data sets. Consequently, with this thesis, we aimed to contribute to this field by employing machine learning tools for quantitative chemical composition analysis of individual BNPs.

In chapter 2.1 we discussed the Physico-chemical properties of BNPs, such as their exciting characteristics in plasmonics and catalysis, where the size and chemical composition plays a fundamental role. In sequence, we showed that we synthesized small AuAg BNPs (Diam. < 10 nm) as a benchmark system for the improvement of quantitative analysis by Energy Dispervise X-ray Spectroscopy (EDS). We used gas aggregation approaches to physically prepare the samples surfactant-free, free of the influence of passivation that intrinsically changes the system's properties.

In the literature the need for the chemical composition quantification of small BNPs evaluating it in the function of size, where nanoscale effects show themselves as a playground for scientific investigators. The BNP's Physico-chemical properties are structural, size, and compositional dependent. Thus, the correct characterization of both ensemble and individual BNPs is crucial for understanding fundamental nanoscience and technological applications. Therefore, we showed in chapter 3 the use of high spatial

resolution spectroscopies and microscopies as crucial tools for measuring such systems. Sequentially we explain the use of EDS for these purposes showing the basics of the technique, its advantages, limitations, and technical details, focusing on the opportunity of quickly being quantitative with this methodology. EDS is one of the most popular techniques used in TEM, and due to its simplicity, the method is overused qualitatively in simple elemental mapping by various nanoscience communities. When it is considered, the quantitative analysis is typically resumed to the export of results from the data acquisition software without careful data treatment (spectral fitting, background subtraction, etc.) and rigorous error bar determination. Due to the lack of thoroughness in the quantitative analysis, even more in such a small system where high signal-to-noise ratios (SNRs) are difficult to be attained, not surprisingly, the quantifications do not always show realistic values. Then, this issue might lead to serious disturbs in modeling and interpreting general and applied physics and chemistry results.

In chapter 4, aiming to quantify with high precision and accuracy AuAg BNP chemical compositions, we used the well-established Cliff-Lorimer method for each BNP measured in the elemental maps. We developed our python software for the data treatment based on the HyperSpy library. Here, we prepared a semi-automatic data processing routine, where we extracted the information of each BNP, the spectrum fitting, background subtraction, integration, and uncertainty bar calculation. Then, in chapter 5, to improve the quantification of the chemical composition through elemental mapping, we used machine learning tools, such as Principal Component Analysis (PCA), for denoising the data set. Therefore, we discussed the use of unsupervised machine learning and the state-of-the-art of its use in the EDS scenario. The architecture of the data set, the so-called hyperspectral images (HSIs), are explored to represent each pixel/channel's intensities better. This way, the information can be reconstructed in a new basis of vectors, the so-called principal components, reducing the noise. In the subsequent chapter 6 and 7 supported by two publications in "Microscopy and Microanalysis", we emphasize the need for well-determined limits of detection and quantification for information extraction. In this context, we verified the advantages and problems of using this machine-learning filter. Even though the Poisson noise is severely reduced, if the SNR of the data set is not good enough for proper quantification (too far from the quantification limit), the denoising procedure generates biased results, consistent with previous work reported in the literature. Motivated by these observations, we modeled and proposed the detection and quantification limits of the methodology by studying statistical and information theory models. We estimate information loss parameters related to information variance, noise, number of pixels, and channels. Consequently, we could estimate confidence intervals for quantifying the chemical composition after the PCA denoising from these estimators. Comparing the confidence intervals obtained by simple error propagation of raw data and those from the denoised data, we could show that the measurements' uncertainty is reduced due to the Poisson noise reduction. Therefore, we show that by increasing the number of counts in our measurements and applying PCA denoising, we obtain improved results from the point of view of accuracy and precision.

In sequence, motivated by the open questions related to the elemental distribution inside AuAg BNPs, we profited from the opportunity of having control in the BNPs synthesis, data acquisition, and processing to try to go further in our analysis. Thus, in chapter 8, we proceeded with our studies on the AuAg system to contribute to learning

how many atoms we have in the BNPs and where they are located. Our results in raw EDS data showed Ag segregation in the shell of the BNP, and High Angle Annular Dark Field (HAADF) images confirmed the existence of a possible silver oxide-rich shell around them. Immediately we discussed that the Ag segregation might be due to the BNP oxidation, where according to the literature, the Ag mobility towards the surface is increased due to the formation of silver oxide. The suggested core-shell morphology, however, is very discreet in the EDS chemical analysis and cannot be easily verified in typical qualitative maps. The quantitative approach showed an Ag atomic fraction of 0.62±0.02 in the shell for 0.48±0.01 of the average full BNP's chemical composition. Thus, we analyzed an AuAg BNP sample with oxidation protection by the evaporation of a few nm of carbon above it. In this case, the Ag segregation severely decreased, consistent with the hypothesis of Ag-induced segregation due to particle oxidation. To guarantee the reasoning of the BNP alloying in the ground state, we annealed the carbon-protected samples and afterward relaxed them to ambient temperature. In this condition, according to our qualitative and quantitative observation with EDS-STEM chemical analysis, we confirmed the alloying of the BNPs. But still, a few percent of Ag can be measured on the shell. Furthermore, to improve the quality of our results, we used unsupervised machine learning in the oxidized BNPs to try to extract hidden information in the data set. Motivated by the literature, where blind source separation is used to unmix EDS signals from nanosystems, we used Non-Negative Matrix Factorization (NMF) to study how the information contained in the data set is separated into the components. This way, we verified three main relevant components of the analysis, consistent with the model of three different chemical phases (background, core, and shell). However, we learned that the unmixing property on complex core-shell-like systems could not be correctly applied to the autonomous quantitative analysis of this system. The main issue, in this case, is that the shell and core interface are not well defined due to the interdiffusion of Ag and Au atoms. Therefore, no abrupt core and shell phase transition exists to be easily detected by the algorithm. The pattern observed for our BNPs is Ag enrichment towards the surface. In the three NMF components, we extracted the background in the first and an alloyed AuAg chemical phase in the second. Finally, in the score image of the third, we found a pattern that can be recognized erroneously as the BNP shell. Here, the signal contained in the third component is responsible for modulating the variability in chemical composition. We verified by the Ag atomic fraction profile of the BNP reconstructed image that the two chemical phases could not be distinguished considering the detection limit criterium; therefore, the shell detected in the decomposition represents the Ag enrichment, not a pure Ag shell. With this description and supported by simulations, we showed that an alloy forms the AuAg BNPs system with more Ag in the shell and more Au in the core, but still a mixed system. By analyzing the BNPs profile quantitatively, we measured chemical gradients inside those BNPs, which was a breakthrough and not an expected result for EDS-STEM analysis. Supported by the denoising properties of NMF reconstruction, we could verify that the chemical gradient measures are indeed physical information and not noise fluctuations. Hence, we obtained the Ag atomic fraction increasing from 0.45±0.02 in the BNP's center to 0.62±0.02 in the surface.

In the continuity of the work developed in this thesis, we expect to employ NMF to investigate the carbon-protected samples where the alloying level is much more challenging to explore. Using NMF and chemical gradient quantification, we will evaluate

this sample's elemental distributions and compare the results quantitatively with the oxidized and segregated case. We also expect to analyze the sample aging and compare the evolution of the chemical gradients over time, which may allow us to estimate a segregation rate parameter for such systems. To completely understand the elemental distribution inside the AuAg BNPs, we will optimize the sample's annealing process (NP size, annealing time, temperature, etc.) to obtain relaxed BNPs in the ground state without the loss of Ag atoms. We also expect to use the support of simulations to correctly estimate the limit of detection and quantification in the analysis after NMF reconstruction. Finally, we have as the central perspective to apply this procedure to fully characterize the elemental distribution of AuAg BNPs, separating the external influences on BNP's conformation and finishing the measurement of chemical gradients by EDS-STEM. At this point, we have to add the possibility of continuing the work by using the microscopy techniques more suitable for the questions related to the atomic scale information. In this scenario, EDS fails to be quantitative due to electron beam delocalization, channeling, and the lack of statistics (X-ray counting). It is encouraging, however, for HAADF-STEM quantitative analysis, that EDS-STEM measured chemical gradients, even within the method's limit.

Motivated by our analytical and quantitative studies developed for EDS-STEM, one could use HAADF-STEM combined with detector calibration and rigorous modeling by multislice simulations. In this way, one can quantify the atomic columns of the AuAg BNPs and, therefore, obtain the chemical gradients at the atomic scale. Thus, this would lead the analytical study of BNPs to a new level, leveraging the extraction of fine details in the chemical composition of finite nanoscale systems. Additionally, one can develop new studies using in-situ and environmental TEM methodologies, studying chemical gradients, alloying, and segregation with the systems evolving in time and even under reactive environments like in redox (oxidation-reduction) reactions. Among other advanced methodologies being developed for TEM studies, Pair Distribution Function (PDF) analysis applied to electron diffraction datasets can also provide average information on the structure and atomic arrangement of BNP samples. In this sense, complementary approaches such as local spectroscopies, imaging, and PDF may fully describe quantitatively nanoscale systems' characteristics from the point of view of both ensemble and individual BNP. Knowing that the complete characterization of nanometric systems requires detailed knowledge of how the different chemical elements of the materials are distributed at the interfaces/surfaces (roughness, interdiffusion, etc.), the present work and the perspective presented above, construct through TEM studies opportunities to the scientific community to fully characterize BNP systems and understand its physico-chemical properties in both fundamental and applied sciences as well.

Bibliography

[1] R. Ferrando, J. Jellinek, and R. L. Johnston, "Nanoalloys: From theory to applications of alloy clusters and nanoparticles," 2008.

[2] D. Alloyeau, C. Mottet, and C. Ricolleau, eds., *Nanoalloys: synthesis, structure and properties*. Engineering Materials, London: Springer London, 2012.

[3] *Structure and properties of nanoalloys*, vol. 10 of *Frontiers of Nanoscience*. Elsevier, 2016.

[4] R. L. Johnston, *Atomic and molecular clusters*. CRC Press, apr 2002.

[5] C. Lyman, R. Lakis, and H. Stenger, "X-ray emission spectrometry of phase separation in PtRh nanoparticles for nitric oxide reduction," *Ultramicroscopy*, vol. 58, pp. 25–34, apr 1995.

[6] S. J. Pennycook and P. D. Nellist, eds., *Scanning transmission electron microscopy*. New York, NY: Springer New York, 2011.

[7] C. B. Carter and D. B. Williams, eds., *Transmission Electron Microscopy*. Cham: Springer International Publishing, 2016.

[8] G. Cliff and G. W. Lorimer, "The quantitative analysis of thin specimens," *Journal of Microscopy*, vol. 103, no. 2, pp. 203–207, 1975.

[9] L. A. Currie, "Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry," *Analytical Chemistry*, 1968.

[10] L. A. Currie, "Detection and quantification limits: origins and historical overview1Adapted from the proceedings of the 1996 joint statistical meetings (american statistical association, 1997). original title: and future of detection and quantification limits". contribution of the national institute of standards and technology; not subject to copyright.1," *Analytica Chimica Acta*, vol. 391, pp. 127–134, may 1999.

[11] M. Belter, A. Sajnóg, and D. Barałkiewicz, "Over a century of detection and quantification capabilities in analytical chemistry–historical overview and trends.," *Talanta*, vol. 129, pp. 606–616, nov 2014.

[12] P. G. Kotula, M. R. Keenan, and J. R. Michael, "Automated analysis of SEM X-ray spectral images: A powerful new microanalysis tool," *Microscopy and Microanalysis*, vol. 9, no. 1, pp. 1–17, 2003.

[13] D. Rossouw, P. Burdet, F. De La Peña, C. Ducati, B. R. Knappett, A. E. Wheatley, and P. A. Midgley, "Multicomponent Signal Unmixing from Nanoheterostructures: Overcoming the Traditional Challenges of Nanoscale X-ray Analysis via Machine Learning," 2015.

[14] B. R. Jany, A. Janas, and F. Krok, "Retrieving the Quantitative Chemical Information at Nanoscale from Scanning Electron Microscope Energy Dispersive X-ray Measurements by Machine Learning," *Nano Letters*, vol. 17, no. 11, pp. 6520–6525, 2017.

[15] K. A. Brown, S. Brittman, N. Maccaferri, D. Jariwala, and U. Celano, "Machine learning in nanoscience: big data at small scales.," *Nano Letters*, vol. 20, pp. 2–10, jan 2020.

[16] S. R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S. V. Kalinin, M. J. Olszta, R. E. Dunin-Borkowski, N. Salmon, K. Hattar, W.-C. D. Yang, R. Sharma, Y. Du, A. Chiaramonti, H. Zheng, E. C. Buck, L. Kovarik, R. L. Penn, D. Li, X. Zhang, M. Murayama, and M. L. Taheri, "Towards data-driven next-generation transmission electron microscopy.," *Nature Materials*, vol. 20, pp. 274–279, mar 2021.

[17] I. Jolliffe, *Principal Component Analysis*. Springer Series in Statistics, New York: Springer-Verlag, 2nd ed., 2002.

[18] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments.," *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, vol. 374, p. 20150202, apr 2016.

[19] M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori, and T. Tanji, "Sparse modeling of EELS and EDX spectral imaging data by nonnegative matrix factorization.," *Ultramicroscopy*, vol. 170, pp. 43–59, nov 2016.

[20] C. G. Puntonet and A. Prieto, eds., *Independent Component Analysis and Blind Signal Separation: Fifth International Conference, ICA 2004, Granada, Spain, September 22-24, 2004. Proceedings*, vol. 3195 of *Lecture notes in computer science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[21] *Handbook of blind source separation*. Elsevier, 2010.

[22] J.-H. Liu, A.-Q. Wang, Y.-S. Chi, H.-P. Lin, and C.-Y. Mou, "Synergistic effect in an auag alloy nanocatalyst: CO oxidation," *The Journal of Physical Chemistry B*, vol. 109, pp. 40–43, jan 2005.

[23] M. Gaudry, J. Lermé, E. Cottancin, M. Pellarin, J. L. Vialle, M. Broyer, B. Prével, M. Treilleux, and P. Mélinon, "Optical properties ofclusters embedded in alumina: Evolution with size and stoichiometry," *Physical Review B*, vol. 64, p. 085407, aug 2001.

[24] J. L. Wang, R. A. Ando, and P. H. C. Camargo, "Investigating the plasmon-mediated catalytic activity of AgAu nanoparticles as a function of composition: Are two metals better than one?," *ACS Catalysis*, vol. 4, pp. 3815–3819, nov 2014.

[25] G. Guisbiers, R. Mendoza-Cruz, L. Bazán-Díaz, J. J. Velázquez-Salazar, R. Mendoza-Perez, J. A. Robledo-Torres, J. L. Rodriguez-Lopez, J. M. Montejano-Carrizales, R. L. Whetten, and M. José-Yacamán, "Electrum, the gold-silver alloy, from the bulk scale to the nanoscale: Synthesis, properties, and segregation rules," *ACS Nano*, vol. 10, no. 1, pp. 188–198, 2016.

[26] D. Belić, R. L. Chantry, Z. Y. Li, and S. A. Brown, "Ag-au nanoclusters: Structure and phase segregation," *Applied physics letters*, vol. 99, p. 171914, oct 2011.

[27] I. Angelini, G. Artioli, P. Bellintani, V. Diella, M. Gemmi, A. Polla, and A. Rossi, "Chemical analyses of Bronze Age glasses from Frattesina di Rovigo, Northern Italy," *Journal of Archaeological Science*, vol. 31, no. 8, 2004.

[28] Z. Hedayatnasab, F. Abnisa, and W. M. A. W. Daud, "Review on magnetic nanoparticles for magnetic nanofluid hyperthermia application," *Materials and Design*, vol. 123, pp. 174–196, 2017.

[29] L. A. Mercante, V. P. Scagion, A. Pavinatto, R. C. Sanfelice, L. H. Mattoso, and D. S. Correa, "Electronic tongue based on nanostructured hybrid films of gold nanoparticles and phthalocyanines for milk analysis," *Journal of Nanomaterials*, 2015.

[30] A. Piriya V.S., P. Joseph, K. Daniel S.C.G., S. Lakshmanan, T. Kinoshita, and S. Muthusamy, "Colorimetric sensors for rapid detection of various analytes," *Materials Science and Engineering: C*, vol. 78, pp. 1231–1245, 2017.

[31] M. Gaudry, E. Cottancin, M. Pellarin, J. Lermé, L. Arnaud, J. R. Huntzinger, J. L. Vialle, M. Broyer, J. L. Rousset, M. Treilleux, and P. Mélinon, "Size and composition dependence in the optical properties of mixed (transition metal/noble metal) embedded clusters," *Physical Review B - Condensed Matter and Materials Physics*, vol. 67, 2003.

[32] J. Hühn, C. Carrillo-Carrion, M. G. Soliman, C. Pfeiffer, D. Valdeperez, A. Masood, I. Chakraborty, L. Zhu, M. Gallego, Z. Yue, M. Carril, N. Feliu, A. Escudero, A. M. Alkilany, B. Pelaz, P. del Pino, and W. J. Parak, "Selected Standard Protocols for the Synthesis, Phase Transfer, and Characterization of Inorganic Colloidal Nanoparticles," *Chemistry of Materials*, vol. 29, no. 1, p. 399–461, 2017.

[33] M. T. Swihart, "Vapor-phase synthesis of nanoparticles," *Current Opinion in Colloid and Interface Science*, vol. 8, pp. 127–133, 2003.

[34] Y. Huttel, *Gas-Phase Synthesis of Nanoparticles*, vol. 1. Wiley-VCH Verlag GmbH Co. KGaA, 1 ed., 3 2017.

[35] H. Haberland, V. I. Goldanskii, F. P. Schäfer, J. P. Toennies, and H. K. V. Lotsch, eds., *Clusters of atoms and molecules*, vol. 52 of *Springer series in chemical physics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994.

[36] P. Buffat and J. P. Borel, "Size effect on the melting temperature of gold particles," *Physical Review A*, vol. 13, 1976.

[37] T. Castro, R. Reifenberger, E. Choi, and R. P. Andres, "Size-dependent melting temperature of individual nanometer-sized metallic clusters," *Physical Review B*, vol. 42, no. 13, pp. 8548–8556, 1990.

[38] P.-C. Chen, X. Liu, J. L. Hedrick, Z. Xie, S. Wang, Q.-Y. Lin, M. C. Hersam, V. P. Dravid, C. A. Mirkin, M. R. Buck, R. E. Schaak, C. Chen, Y. Kang, Z. Huo, Z. Zhu, W. Huang, H. L. Xin, J. D. Snyder, D. Li, J. A. Herron, M. Mavrikakis, M. Chi, K. L. More, Y. Li, N. M. Markovic, G. A. Somorjai, P. Yang, V. R. Stamenkovic, G. Chen, Y. Zhao, G. Fu, P. N. Duchesne, L. Gu, Y. Zheng, X. Weng, M. Chen, P. Zhang, C. W. Pao, J. F. Lee, N. Zheng, L. Gan, C. Cui, M. Heggen, F. Dionigi, S. Rudi, P. Strasser, M. R. Buck, J. F. Bondi, R. E. Schaak, C. G. Read, T. R. Gordon, J. M. Hodges, R. E. Schaak, W. Sang, T. Zheng, Y. Wang, X. Li, X. Zhao, J. Zeng, J. G. Hou, Z. Nie, W. Li, M. Seo, S. Xu, E. Kumacheva, A. K. Ganguli, A. Ganguly, S. Vaidya, C. George, A. Genovese, A. Casu, M. Prato, M. Povia, L. Manna, T. Montanari, A. Demortière, R. D. Schaller, T. Li, S. Chattopadhyay, G. Krylova, T. Shibata, P. C. d. S. Claro, C. E. Rowland, J. T. Miller, R. Cook, B. Lee, E. V. Shevchenko, C. Xu, B. Wang, S. Sun, I. Schick, S. Lorenz, D. Gehrig, A. M. Schilmann, H. Bauer, M. Panthöfer, K. Fischer, D. Strand, F. Laquai, W. Tremel, J. C. Hill, A. T. Landers, J. A. Switzer, X. Huang, Z. Zhao, L. Cao, Y. Chen, E. Zhu, Z. Lin, M. Li, A. Yan, A. Zettl, Y. M. Wang, X. Duan, T. Mueller, Y. Huang, L. Zhang, L. T. Roling, X. Wang, M. Vara, M. Chi, J. Liu, S. I. Choi, J. Park, J. A. Herron, Z. Xie, M. Mavrikakis, Y. Xia, M. B. Cortie, A. M. McDonagh, D. Rodriguez-Fernandez, J. Langer, M. Henriksen-Lacey, L. M. Liz-Marzan, E. Reddington, A. Sapienza, B. Gurau, R. Viswanathan, S. Sarangapani, E. S. Smotkin, T. E. Mallouk, J. Chai, F. Huo, Z. Zheng, L. R. Giam, W. Shim, C. A. Mirkin, L. R. Giam, S. He, N. E. Horwitz, D. J. Eichelsdoerfer, J. Chai, Z. Zheng, D. Kim, W. Shim, C. A. Mirkin, G. Liu, D. J. Eichelsdoerfer, B. Rasin, Y. Zhou, K. A. Brown, X. Liao, C. A. Mirkin, P.-C. Chen, G. Liu, Y. Zhou, K. A. Brown, N. Chernyak, J. L. Hedrick, S. He, Z. Xie, Q. Y. Lin, V. P. Dravid, S. A. O'Neill-Slawecki, C. A. Mirkin, F. Huo, Z. Zheng, G. Zheng, L. R. Giam, H. Zhang, C. A. Mirkin, H. Yu, M. Chen, P. M. Rice, S. X. Wang, R. L. White, S. Sun, T. Mokari, E. Rothenberg, I. Popov, R. Costi, U. Banin, R. D. Piner, J. Zhu, F. Xu, S. Hong, C. A. Mirkin, M. Sankar, N. Dimitratos, P. J. Miedziak, P. P. Wells, C. J. Kiely, G. J. Hutchings, S. G. Kwon, G. Krylova, P. J. Phillips, R. F. Klie, S. Chattopadhyay, T. Shibata, E. E. Bunel, Y. Liu, V. B. Prakapenka, B. Lee, E. V. Shevchenko, G. Guisbiers, S. Mejia-Rosales, S. Khanal, F. Ruiz-Zepeda, R. L. Whetten, M. José-Yacaman, F. Calvo, J. S. Wu, A. M. Kim, R. Bleher, B. D. Myers, R. G. Marvin, H. Inada, K. Nakamura, X. F. Zhang, E. Roth, S. Y. Li, T. K. Woodruff, T. V. O'Halloran, and V. P. Dravid, "Polyelemental nanoparticle libraries.," *Science (New York, N.Y.)*, vol. 352, no. 6293, pp. 1565–9, 2016.

[39] R. Esparza, A. Santoveña, A. Ruíz-Baltazar, A. Angeles-Pascual, D. Bahena, J. Maya-Cornejo, J. Ledesma-García, and R. Pérez, "Study of ptpd bimetallic nanoparticles for fuel cell applications," *Materials Research*, vol. 20, pp. 1193–1200, may 2017.

[40] J. M. LeBeau, S. D. Findlay, L. J. Allen, and S. Stemmer, "Quantitative atomic resolution scanning transmission electron microscopy.," *Physical Review Letters*, vol. 100, p. 206101, may 2008.

[41] J. M. Lebeau and S. Stemmer, "Experimental quantification of annular dark-field images in scanning transmission electron microscopy.," *Ultramicroscopy*, vol. 108, pp. 1653–1658, nov 2008.

[42] U. Kreibig and M. Vollmer, *Optical properties of metal clusters*, vol. 25 of *Springer series in materials science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995.

[43] E. Cottancin, G. Celep, J. Lermé, M. Pellarin, J. R. Huntzinger, J. L. Vialle, and M. Broyer, "Optical properties of noble metal clusters as a function of the size: Comparison between experiments and a semi-quantal theory," *Theoretical chemistry accounts*, vol. 116, pp. 514–523, sep 2006.

[44] V. Amendola, R. Pilot, M. Frasconi, O. M. Maragò, and A. M. Iatì, "Surface plasmon resonance in gold nanoparticles: A review," *Journal of Physics: Condensed Matter*, vol. 29, no. 20, p. 203002, 2017.

[45] S. Peiris, J. McMurtrie, and H. Y. Zhu, "Metal nanoparticle photocatalysts: Emerging processes for green organic synthesis," *Catalysis Science and Technology*, 2016.

[46] A. Campos, N. Troc, E. Cottancin, M. Pellarin, H.-C. Weissker, J. Lermé, M. Kociak, and M. Hillenkamp, "Plasmonic quantum size effects in silver nanoparticles are dominated by interfaces and local environments," *Nature physics*, vol. 15, pp. 275–280, nov 2018.

[47] J. C. Scaiano, K. G. Stamplecoskie, and G. L. Hallett-Tapley, "Photochemical Norrish type i reaction as a tool for metal nanoparticle synthesis: Importance of proton coupled electron transfer," *Chemical Communications*, vol. 40, no. 48, p. 4798–4808, 2012.

[48] S. García, L. Zhang, G. W. Piburn, G. Henkelman, and S. M. Humphrey, "Microwave synthesis of classically immiscible rhodium-silver and rhodium-gold alloy nanoparticles: Highly active hydrogenation catalysts," *ACS Nano*, vol. 8, no. 11, p. 1512.

[49] H. Haberland, M. Mall, M. Moseler, Y. Qiang, T. Reiners, and Y. Thurner, "Filling of micron-sized contact holes with copper by energetic cluster impact," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 12, no. 5, pp. 2925–2930, 1994.

[50] M. Hillenkamp, G. Di Domenicantonio, and C. Felix, "Monodispersed metal clusters in solid matrices: A new experimental setup," *Review of Scientific Instruments*, vol. 77, no. 2, 2006.

[51] A. D. T. de Sá, V. T. Abrao Oiko, G. di Domenicantonio, and V. Rodrigues, "New experimental setup for metallic clusters production based on hollow cylindrical magnetron sputtering," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 32, no. 6, p. 061804, 2014.

[52] A. D. Tavares, *Nano-agregados metálicos: Produção e propriedades magnéticas*. PhD thesis, 2013.

[53] A. K. Samal, L. Polavarapu, S. Rodal-Cedeira, L. M. Liz-Marzán, J. Pérez-Juste, and I. Pastoriza-Santos, "Size tunable au@ag core-shell nanoparticles: synthesis and surface-enhanced raman scattering properties.," *Langmuir: the ACS Journal of Surfaces and Colloids*, vol. 29, pp. 15076–15082, dec 2013.

[54] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, and R. P. Van Duyne, "Biosensing with plasmonic nanosensors.," *Nature Materials*, vol. 7, pp. 442–453, jun 2008.

[55] P. K. Jain, X. Huang, I. H. El-Sayed, and M. A. El-Sayed, "Noble metals on the nanoscale: optical and photothermal properties and some applications in imaging, sensing, biology, and medicine.," *Accounts of Chemical Research*, vol. 41, pp. 1578–1586, dec 2008.

[56] K. Mao, Z. Zhou, S. Han, X. Zhou, J. Hu, X. Li, and Z. Yang, "A novel biosensor based on au@ag core-shell nanoparticles for sensitive detection of methylamphetamine with surface enhanced raman scattering.," *Talanta*, vol. 190, pp. 263–268, dec 2018.

[57] E. C. Dreaden, A. M. Alkilany, X. Huang, C. J. Murphy, and M. A. El-Sayed, "The golden age: gold nanoparticles for biomedicine.," *Chemical Society Reviews*, vol. 41, pp. 2740–2779, apr 2012.

[58] T. Fleetham, J.-Y. Choi, H. W. Choi, T. Alford, D. S. Jeong, T. S. Lee, W. S. Lee, K.-S. Lee, J. Li, and I. Kim, "Photocurrent enhancements of organic solar cells by altering dewetting of plasmonic ag nanoparticles.," *Scientific Reports*, vol. 5, p. 14250, sep 2015.

[59] H. Choi, S.-J. Ko, Y. Choi, P. Joo, T. Kim, B. R. Lee, J.-W. Jung, H. J. Choi, M. Cha, J.-R. Jeong, I.-W. Hwang, M. H. Song, B.-S. Kim, and J. Y. Kim, "Versatile surface plasmon resonance of carbon-dot-supported silver nanoparticles in polymer optoelectronic devices," *Nature photonics*, vol. 7, pp. 732–738, jul 2013.

[60] R. C. Hensel, M. Moreira, A. Riul, O. N. Oliveira, V. Rodrigues, and M. Hillenkamp, "Monitoring the dispersion and agglomeration of silver nanoparticles in polymer thin films using localized surface plasmons and ferrell plasmons," *Applied physics letters*, vol. 116, p. 103105, mar 2020.

[61] L. M. Corrêa, M. Moreira, V. Rodrigues, and D. Ugarte, "Quantitative structural analysis of auag nanoparticles using a pair distribution function based on precession electron diffraction: implications for catalysis," *ACS Applied Nano Materials*, oct 2021.

[62] N. Braidy, Z. J. Jakubek, B. Simard, and G. A. Botton, "Quantitative energy dispersive X-ray microanalysis of electron beam-sensitive alloyed nanoparticles," *Microscopy and Microanalysis*, vol. 14, no. 2, pp. 166–75, 2008.

[63] R. F. Egerton, P. Li, and M. Malac, "Radiation damage in the TEM and SEM," in *Micron*, vol. 35, pp. 399–409, 2004.

[64] R. F. Egerton, R. McLeod, F. Wang, and M. Malac, "Basic questions related to electron-induced sputtering in the TEM," *Ultramicroscopy*, vol. 110, p. 991–997, 2010.

[65] R. F. Egerton, "Mechanisms of radiation damage in beam-sensitive specimens, for TEM accelerating voltages between 10 and 300 kV," *Microscopy Research and Technique*, vol. 75, no. 11, pp. 1550–1556, 2012.

[66] D. Van Dyck, "Persistent misconceptions about incoherence in electron microscopy.," *Ultramicroscopy*, vol. 111, pp. 894–900, jun 2011.

[67] M. Lasserus, M. Schnedlitz, D. Knez, R. Messner, A. Schiffmann, F. Lackner, A. W. Hauser, F. Hofer, and W. E. Ernst, "Thermally induced alloying processes in a bimetallic system at the nanoscale: AgAu sub-5 nm core-shell particles studied at atomic resolution," *Nanoscale*, vol. 10, no. 4, 2018.

[68] M. Hÿtch and W. Stobbs, "Quantitative comparison of high resolution TEM images with image simulations," *Ultramicroscopy*, vol. 53, pp. 191–203, mar 1994.

[69] A. De Backer, G. T. Martinez, K. E. MacArthur, L. Jones, A. Béché, P. D. Nellist, and S. Van Aert, "Dose limited reliability of quantitative annular dark field scanning transmission electron microscopy for nano-particle atom-counting.," *Ultramicroscopy*, vol. 151, pp. 56–61, apr 2015.

[70] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo, "Three-dimensional atomic imaging of crystalline nanoparticles.," *Nature*, vol. 470, pp. 374–377, feb 2011.

[71] Y. Yang, C.-C. Chen, M. C. Scott, C. Ophus, R. Xu, A. Pryor, L. Wu, F. Sun, W. Theis, J. Zhou, M. Eisenbach, P. R. C. Kent, R. F. Sabirianov, H. Zeng, P. Ercius, and J. Miao, "Deciphering chemical order/disorder and material properties at the single-atom level.," *Nature*, vol. 542, pp. 75–79, feb 2017.

[72] J. Zhou, Y. Yang, Y. Yang, D. S. Kim, A. Yuan, X. Tian, C. Ophus, F. Sun, A. K. Schmid, M. Nathanson, H. Heinz, Q. An, H. Zeng, P. Ercius, and J. Miao, "Observing crystal nucleation in four dimensions using atomic electron tomography.," *Nature*, vol. 570, pp. 500–503, jun 2019.

[73] F.-R. Chen, D. Van Dyck, C. Kisielowski, L. P. Hansen, B. Barton, and S. Helveg, "Probing atom dynamics of excited co-mo-s nanocrystals in 3D.," *Nature Communications*, vol. 12, p. 5007, aug 2021.

[74] D. E. Newbury, "X-ray spectrometry and spectrum image mapping at output count rates above 100 kHz with a silicon drift detector on a scanning electron microscope.," *Scanning*, vol. 27, no. 5, pp. 227–239, 2005.

[75] N. J. Zaluzec, "Quantitative assessment and measurement of x-ray detector performance and solid angle in the analytical electron microscope.," *Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, vol. 28, pp. 83–95, feb 2022.

[76] J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. Ritchie, J. H. J. Scott, and D. C. Joy, *Scanning Electron Microscopy and X-Ray Microanalysis*. New York, NY: Springer New York, 2018.

[77] B. D. Willians, C. B. Carter, D. B. Williams, and C. B. Carter, *Transmission Electron Microscopy: a textbook for Materials Science*, vol. V1-V4. 2009.

[78] P. W. Hawkes and J. C. H. Spence, eds., *Springer handbook of microscopy*. Springer Handbooks, Cham: Springer International Publishing, 2019.

[79] L. Deng, W. Hu, H. Deng, S. Xiao, and J. Tang, "Au–ag bimetallic nanoparticles: Surface segregation and atomic-scale structure," *The Journal of Physical Chemistry C*, vol. 115, pp. 11355–11363, jun 2011.

[80] G. Liu, C. Zhang, J. Wu, and C. A. Mirkin, "Using Scanning-Probe Block Copolymer Lithography and Electron Microscopy to Track Shape Evolution in Multimetallic Nanoclusters," *ACS Nano*, vol. 9, no. 12, p. 12137–12145, 2015.

[81] L. H. G. Tizei, T. Chiaramonte, D. Ugarte, and M. A. Cotta, "III-v semiconductor nanowire growth: does arsenic diffuse through the metal nanoparticle catalyst?," *Nanotechnology*, vol. 20, p. 275604, jul 2009.

[82] M. Watanabe, D. W. Ackland, A. Burrows, C. J. Kiely, D. B. Williams, O. L. Krivanek, N. Dellby, M. F. Murfitt, and Z. Szilagyi, "Improvements in the X-ray analytical capabilities of a scanning transmission electron microscope by spherical-aberration correction," in *Microscopy and Microanalysis*, vol. 12, pp. 515–526, 2006.

[83] J. S. Nilsen and A. T. J. van Helvoort, "Composition analysis by STEM-EDX of ternary semiconductors by internal references.," *Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, vol. 28, pp. 61–69, feb 2022.

[84] F. De La Peña, T. Ostasevicius, V. T. Fauske, P. Burdet, P. Jokubauskas, M. Nord, M. Sarahan, E. Prestat, D. N. Johnstone, J. Taillon, J. Caron, T. Furnival, K. E. Macarthur, A. Eljarrat, S. Mazzucco, V. Migunov, T. Aarholt, M. Walls, F. Winkler, G. Donval, B. Martineau, A. Garmannslund, L.-F. Zagonel, and I. Iyengar, "Electron Microscopy (Big and Small) Data Analysis With the Open

Source Software Package HyperSpy," *Microscopy and Microanalysis*, vol. 23, no. S1, pp. 214–215, 2017.

[85] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, 6 2014.

[86] P. Cueva, R. Hovden, J. A. Mundy, H. L. Xin, and D. A. Muller, "Data processing for atomic resolution electron energy loss spectroscopy," in *Microscopy and Microanalysis*, vol. 18, pp. 667–675, 2012.

[87] C. M. Parish and L. N. Brewer, "Multivariate statistics applications in phase analysis of STEM-EDS spectrum images," *Ultramicroscopy*, vol. 110, no. 2, pp. 134–143, 2010.

[88] K. P. F.R.S., "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[89] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," vol. 24, no. 6, pp. 417–441, 1933.

[90] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, no. 4, pp. 1531–1549, 2010.

[91] A. Caprihan, G. D. Pearlson, and V. D. Calhoun, "Application of principal component analysis to distinguish patients with schizophrenia from healthy controls based on fractional anisotropy measurements," *NeuroImage*, vol. 42, pp. 675–682, 2008.

[92] M. L. Braunger, F. M. Shimizu, M. J. M. Jimenez, L. R. Amaral, M. H. d. O. Piazzetta, L. Gobbi, P. S. G. Magalhães, V. Rodrigues, O. N. Oliveira, and A. Riul, "Microfluidic electronic tongue applied to soil analysis," *Chemosensors*, vol. 5, no. 2, 2017.

[93] J. Shlens, "A tutorial on principal component analysis derivation, discussion and singular value decomposition," 2003.

[94] M. R. Keenan and P. G. Kotula, "Accounting for poisson noise in the multivariate analysis of ToF-SIMS spectrum images," *Surface and Interface Analysis*, vol. 36, pp. 203–212, mar 2004.

[95] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[96] P. Potapov and A. Lubk, "Optimal principal component analysis of STEM XEDS spectrum images.," *Advanced structural and chemical imaging*, vol. 5, p. 4, apr 2019.

[97] P. Potapov, "Why principal component analysis of STEM spectrum-images results in "abstract", uninterpretable loadings?," *Ultramicroscopy*, vol. 160, pp. 197–212, jan 2016.

[98] B. Nadler, "Finite sample approximation results for principal component analysis: A matrix perturbation approach," *The Annals of Statistics*, vol. 36, pp. 2791–2817, dec 2008.

[99] B. Nadler, "Discussion," *Journal of the American Statistical Association*, vol. 104, pp. 694–697, jun 2009.

[100] N. Faber, M. Meinders, P. Geladi, M. Sjöström, L. Buydens, and G. Kateman, "Random error bias in principal component analysis. part i. derivation of theoretical predictions," *Analytica Chimica Acta*, vol. 304, pp. 257–271, apr 1995.

[101] N. Faber, M. Meinders, P. Geladi, M. Sjöström, L. Buydens, and G. Kateman, "Random error bias in principal component analysis. part II. application of theoretical predictions to multivariate problems," *Analytica Chimica Acta*, vol. 304, pp. 273–283, apr 1995.

[102] M. G. Burke, M. Watanabe, D. B. Williams, and J. M. Hyde, "Quantitative characterization of nanoprecipitates in irradiated low-alloy steels: advances in the application of FEG-STEM quantitative microanalysis to real materials," *Journal of materials science*, vol. 41, pp. 4512–4522, jul 2006.

[103] J. Spiegelberg and J. Rusz, "Can we use PCA to detect small signals in noisy data?," *Ultramicroscopy*, vol. 172, pp. 40–46, 2017.

[104] S. Lichtert and J. Verbeeck, "Statistical consequences of applying a PCA noise filter on EELS spectrum images.," *Ultramicroscopy*, vol. 125, pp. 35–42, feb 2013.

[105] P. Potapov, "On the loss of information in PCA of spectrum-images.," *Ultramicroscopy*, vol. 182, pp. 191–194, nov 2017.

[106] A. Mirzal, "NMF versus ICA for blind source separation," *Advances in data analysis and classification*, vol. 11, pp. 25–48, mar 2017.

[107] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization.," *Nature*, vol. 401, pp. 788–791, oct 1999.

[108] A. R. Persson, M. Tornberg, R. Sjökvist, and D. Jacobsson, "Time-resolved compositional mapping during in situ TEM studies.," *Ultramicroscopy*, vol. 222, p. 113193, mar 2021.

[109] O. Nicoletti, F. de la Peña, R. K. Leary, D. J. Holland, C. Ducati, and P. A. Midgley, "Three-dimensional imaging of localized surface plasmon resonances of metal nanoparticles.," *Nature*, vol. 502, pp. 80–84, oct 2013.

[110] B. H. Martineau, D. N. Johnstone, A. T. J. van Helvoort, P. A. Midgley, and A. S. Eggeman, "Unsupervised machine learning applied to scanning precession electron diffraction data," *Advanced structural and chemical imaging*, vol. 5, p. 3, dec 2019.

[111] X. Li, G. Haberfehlner, U. Hohenester, O. Stéphan, G. Kothleitner, and M. Kociak, "Three-dimensional vectorial imaging of surface phonon polaritons.," *Science*, vol. 371, pp. 1364–1367, mar 2021.

[112] F. Uesugi, S. Koshiya, J. Kikkawa, T. Nagai, K. Mitsuishi, and K. Kimoto, "Non-negative matrix factorization for mining big data obtained using four-dimensional scanning transmission electron microscopy.," *Ultramicroscopy*, vol. 221, p. 113168, feb 2021.

[113] E. Ringe, R. P. Van Duyne, and L. D. Marks, "Wulff construction for alloy nanoparticles.," *Nano Letters*, vol. 11, pp. 3399–3403, aug 2011.

[114] F. Calvo, E. Cottancin, and M. Broyer, "Segregation, core alloying, and shape transitions in bimetallic nanoclusters: Monte carlo simulations," *Physical Review B*, vol. 77, p. 121406, mar 2008.

[115] K. Jia, M. Y. Khaywah, Y. Li, J. L. Bijeon, P. M. Adam, R. Déturche, B. Guelorget, M. François, G. Louarn, and R. E. Ionescu, "Strong improvements of localized surface plasmon resonance sensitivity by using au/ag bimetallic nanostructures modified with polydopamine films.," *ACS Applied Materials & Interfaces*, vol. 6, pp. 219–227, jan 2014.

[116] N. Eom, M. E. Messing, J. Johansson, and K. Deppert, "General trends in core-shell preferences for bimetallic nanoparticles.," *ACS Nano*, vol. 15, pp. 8883–8895, may 2021.

[117] E. y. El koraychy, D. Nelli, C. Roncaglia, C. Minnai, and R. Ferrando, "Growth of size-matched nanoalloys – a comparison of AuAg and PtPd," *The European Physical Journal Applied Physics*, vol. 97, p. 28, 2022.

[118] D. Cheng, W. Wang, S. Huang, and D. Cao, "Atomistic modeling of multishell onion-ring bimetallic nanowires and clusters," *The Journal of Physical Chemistry C*, vol. 112, pp. 4855–4860, apr 2008.

[119] F. Chen and R. L. Johnston, "Energetic, electronic, and thermal effects on structural properties of ag-au nanoalloys.," *ACS Nano*, vol. 2, pp. 165–175, jan 2008.

[120] D. Rapetti and R. Ferrando, "Density functional theory global optimization of chemical ordering in AgAu nanoalloys," *Journal of alloys and compounds*, vol. 779, pp. 582–589, mar 2019.

[121] A. Ghosh, S. Datta, and T. Saha-Dasgupta, "Understanding the trend in core–shell preferences for bimetallic nanoclusters: A machine learning approach," *The Journal of Physical Chemistry C*, vol. 126, pp. 6847–6853, apr 2022.

[122] M. Cerbelaud, R. Ferrando, G. Barcaro, and A. Fortunelli, "Optimization of chemical ordering in AgAu nanoalloys.," *Physical Chemistry Chemical Physics*, vol. 13, pp. 10232–10240, jun 2011.

[123] T.-W. Liao, A. Yadav, K.-J. Hu, J. van der Tol, S. Cosentino, F. D'Acapito, R. E. Palmer, C. Lenardi, R. Ferrando, D. Grandjean, and P. Lievens, "Unravelling the nucleation mechanism of bimetallic nanoparticles with composition-tunable core-shell arrangement.," *Nanoscale*, vol. 10, pp. 6684–6694, apr 2018.

[124] M. B. Cortie and A. M. McDonagh, "Synthesis and optical properties of hybrid and alloy plasmonic nanoparticles," *Chemical Reviews*, vol. 111, no. 6, p. 3713–3735, 2011.

[125] T. Shibata, B. A. Bunker, Z. Zhang, D. Meisel, C. F. Vardeman, and J. D. Gezelter, "Size-dependent spontaneous alloying of Au-Ag nanoparticles," *Journal of the American Chemical Society*, vol. 124, no. 40, p. 11989–11996, 2002.

[126] J. P. Wilcoxon and B. L. Abrams, "Synthesis, structure and properties of metal nanoclusters," 2006.

[127] H. Chen, R. Liu, L.-Y. Jang, J.-F. Lee, and S. Hu, "Characterization of core–shell type and alloy ag/au bimetallic clusters by using extended x-ray absorption fine structure spectroscopy," *Chemical physics letters*, vol. 421, pp. 118–123, apr 2006.

[128] Z. Y. Li, J. P. Wilcoxon, F. Yin, Y. Chen, R. E. Palmer, and R. L. Johnston, "Structures and optical properties of 4-5 nm bimetallic AgAu nanoparticles," *Faraday Discussions*, vol. 138, pp. 363–373, 2008.

[129] K. Kim, K. L. Kim, J.-Y. Choi, H. B. Lee, and K. S. Shin, "Surface enrichment of ag atoms in au/ag alloy nanoparticles revealed by surface-enhanced raman scattering of 2,6-dimethylphenyl isocyanide," *The Journal of Physical Chemistry C*, vol. 114, pp. 3448–3453, mar 2010.

[130] K. D. Gilroy, A. Ruditskiy, H.-C. Peng, D. Qin, and Y. Xia, "Bimetallic nanocrystals: syntheses, properties, and applications.," *Chemical Reviews*, vol. 116, pp. 10414–10472, sep 2016.

[131] I. J. Godfrey, A. J. Dent, I. P. Parkin, S. Maenosono, and G. Sankar, "Structure of gold–silver nanoparticles," *The Journal of Physical Chemistry C*, vol. 121, pp. 1957–1963, jan 2017.

[132] C. M. Gonzalez, Y. Liu, and J. C. Scaiano, "Photochemical strategies for the facile synthesis of goldsilver alloy and coreshell bimetallic nanoparticles," *The Journal of Physical Chemistry C*, vol. 113, pp. 11861–11867, jul 2009.

[133] C. Zhu, J. Zeng, J. Tao, M. C. Johnson, I. Schmidt-Krey, L. Blubaugh, Y. Zhu, Z. Gu, and Y. Xia, "Kinetically controlled overgrowth of ag or au on pd nanocrystal seeds: from hybrid dimers to nonconcentric and concentric bimetallic nanocrystals.," *Journal of the American Chemical Society*, vol. 134, pp. 15822–15831, sep 2012.

[134] G. Haberfehlner, P. Thaler, D. Knez, A. Volk, F. Hofer, W. E. Ernst, and G. Kothleitner, "Formation of bimetallic clusters in superfluid helium nanodroplets analysed by atomic resolution electron tomography.," *Nature Communications*, vol. 6, p. 8779, oct 2015.

[135] F. Tournus, N. Blanc, A. Tamion, V. Dupuis, and T. Epicier, "Coalescence-free l10 ordering of embedded CoPt nanoparticles," *Journal of applied physics*, vol. 109, p. 07B722, apr 2011.

[136] R. Alayan, L. Arnaud, M. Broyer, E. Cottancin, J. Lermé, J. L. Vialle, and M. Pellarin, "Morphology and growth of metal clusters in the gas phase: A transition from spherical to ramified structures," *Physical Review B*, vol. 73, p. 125444, mar 2006.

[137] D. Alloyeau, G. Prévot, Y. Le Bouar, T. Oikawa, C. Langlois, A. Loiseau, and C. Ricolleau, "Ostwald ripening in nanoalloys: when thermodynamics drives a size-dependent particle composition.," *Physical Review Letters*, vol. 105, p. 255901, dec 2010.

[138] L. Jones, A. Varambhia, R. Beanland, D. Kepaptsoglou, I. Griffiths, A. Ishizuka, F. Azough, R. Freer, K. Ishizuka, D. Cherns, Q. M. Ramasse, S. Lozano-Perez, and P. D. Nellist, "Managing dose-, damage- and data-rates in multi-frame spectrum-imaging," *Microscopy*, vol. 67, no. 1, pp. 98–113, 2018.

[139] X. Xu, S. Xu, L. Jin, and E. Song, "Characteristic analysis of otsu threshold and its applications," *Pattern recognition letters*, vol. 32, pp. 956–961, may 2011.

[140] S. R. Spurgeon, Y. Du, and S. A. Chambers, "Measurement error in atomic-scale scanning transmission electron microscopy-energy-dispersive x-ray spectroscopy (STEM-EDS) mapping of a model oxide interface.," *Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, vol. 23, pp. 513–517, jun 2017.

[141] K. E. MacArthur, H. G. Brown, S. D. Findlay, and L. J. Allen, "Probing the effect of electron channelling on atomic resolution energy dispersive x-ray quantification.," *Ultramicroscopy*, vol. 182, pp. 264–275, nov 2017.

[142] P. Lu, J. M. Moya, R. Yuan, and J. M. Zuo, "Studies of x-ray localization and thickness dependence in atomic-scale elemental mapping by STEM energy-dispersive x-ray spectroscopy using single-frame scanning method.," *Ultramicroscopy*, vol. 186, pp. 23–29, 2018.

Appendix A

Appendix

Influence of cluster sources on bimetallic nanoparticles' chemical composition and growth mechanisms

Murilo Moreira,[†,‡] Levi C. Felix,[†,¶] Emmanuel Cottancin,[‡] Michel Pellarin,[‡] Daniel Ugarte,[†] Matthias Hillenkamp,[†,‡] Douglas S. Galvao,[†] and Varlei Rodrigues[*,†]

†*'Gleb Wataghin' Institute of Physics, State University of Campinas, Campinas,SP, 13083-970, Brazil*

‡*Institute of Light and Matter, University Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5306, Villeurbanne F-69622, France*

¶*These authors contributed equally to this work*

E-mail: varlei@ifi.unicamp.br

Abstract

Metal nanoparticles are omnipresent in today's applied and fundamental research. Both wet-chemical as well as physical procedures for their fabrication are well-established, where the latter is of particular interest as they supply surfactant-free particles. The particle growth has been investigated for several decades, but due to its complexity, involving kinetic and dynamic processes on various length and time scales, often only phenomenological rules of thumb are available. In this study, we report on bimetallic AgAu nanoparticles and demonstrate how the additional degree of freedom of the chemical composition can be used to derive information about how and where the particles grow, depending on two different cluster source types (hollow magnetron vs. laser

vaporization). The chemical composition is quantified on the single-particle level using electron-induced x-ray spectroscopy (STEM-EDS) and shows significant differences for the two fabrication routes. Based on Molecular Dynamics and Monte-Carlo simulations, we derive that for hollow cylindrical sources both the mean particle size and the chemical composition are determined within the plasma region, where particles not only grow but also evaporate low-boiling silver. The comparably large plasma plays a decisive role here, as opposed to planar magnetron or laser vaporization sources, where no such evaporation is observed. These results shed light into the complex cluster growth and help understand and optimizing nanoscale fabrication processes.

Introduction

Metallic nanoparticles have been used in numerous applied and fundamental studies, involving different research fields, amongst others, optics,[1] catalysis,[2] photocatalysis,[3] electronics,[4] magnetism,[5] spintronics[6,7] and sensing.[8] Two main nanoparticle fabrication approaches can be differentiated: wet chemistry and physical techniques. The latter has the advantage of providing surfactant-free nanoparticles; among them, gas-phase synthesis is the most frequently used at present.[9] Despite a large amount of research on such metal nanoparticles, the detailed kinetics and dynamics of the aggregation processes still need to be fully understood, and classical nucleation theory can only provide a qualitative understanding.[10] One of the most used techniques since cluster science's early days is the laser vaporization source (LVS), as shown schematically in Figure 1a.[11] Here a laser hits the metal target rod and ignites a confined plasma of the order of a mm^3 for a very short time, typically several nanoseconds. Next, an inert gas quenches the hot plasma, and the metal vapor condenses into clusters and nanoparticles before rapidly cooling by a supersonic expansion.[12] This is a typical example of cluster fabrication by successive steps: metal vapor generation, supersaturation and condensation, and finally, termination of growth through dilution in the molecular beam.

A second popular method for nanoparticle fabrication is based on magnetron sputtering cluster sources (MSCS), which have been extensively investigated over the last years.[10,13–23] The general interpretation of cluster formation (also by some of the authors of this article) was that metal atoms are sputtered into the gas phase in the discharge and aggregate into clusters and nanoparticles by nucleation, successive atom addition and finally cluster coalescence all along the path between the discharge and the exit orifice of the aggregation region (cf. Figure 1b). This picture was supported by the simple and reproducible observation that increasing the distance between the magnetron head, and exit orifice monotonously increases the size of produced particles. Thus the particle growth was supposed to occur along the same general lines as described above for the LVS, despite the geometric and temporal differences. The magnetron discharge forms a torus of several cm in diameter and some mm in thickness, operated in almost all cases in continuous mode (dc). When used with bimetallic targets, both types of sources, LVS and MSCS, have produced nanoparticles with the same chemical composition as the targets.[17,24]

However, more recent works have started challenging this description for the magnetron source. The authors of[25,26] have monitored the aggregation of Cu and Ag into nanoparticles inside the aggregation tube by in situ X-ray scattering and optical spectroscopy of the silver surface plasmon resonance. They show that, contrary to the general interpretation, the nanoparticles are, in fact, formed in the very same region of the sputter target and not throughout the flow toward the end of the aggregation region. On the contrary, they demonstrated that the nanoparticles are significantly bigger in the vicinity of the discharge than after leaving the discharge area. The authors attribute this behavior to "trapping regions" in the center of the plasma torus and several centimeters above the sputter target, where inhomogeneities in the viscous flow of the inert gas lead to reduced velocities. However, the initial assumption is maintained, where particles can only grow bigger but not decrease in size inside the source. This means that if larger particles are present close to the discharge, they must be trapped or expelled from the formed beam somewhere.

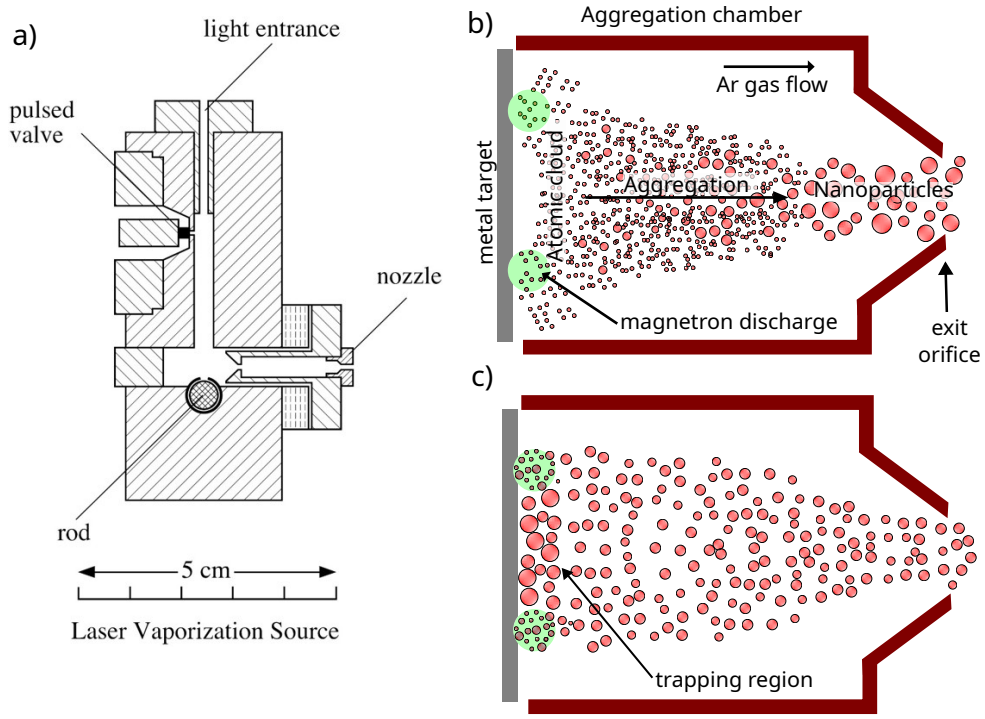


Figure 1: a) Scheme of a laser vaporization source. A laser pulse hits a rod target, eroding it and producing a localized plasma of $\sim$1 mm^3 for a very short time. The plasma is quenched by a pulse of inert gas that condensates the metal vapor into nanoparticles. (b-c) Schematic of a magnetron sputtering cluster source with its round metal target, usually 1 or 2 inches in diameter. The magnetron discharge forms a torus and erodes the target, producing the metal vapor that is used to form the nanoparticles. (b) The standard picture of nanoparticle formation starts from the atomic metal vapor, followed by the nucleation, successive atom addition, and finally, cluster coalescence all the way between the discharge and the exit orifice of the aggregation chamber, usually a few hundred millimeters long. (c) Recent works by Kousal et al.[25,26] propose that the nanoparticles are formed very close to the sputter target and not all along the aggregation chamber. Also, they demonstrate that trapping regions exist in the center of the plasma torus, where the biggest nanoparticles are trapped.

In this article, we contrast these observations with experiments performed using a magnetron sputtering cluster source in the hollow cylindrical configuration (HC-MSCS),[21] see Figure 2a-c. In the HC-MSCS, the target is a metal wire or several twisted wires of the same or different metals on the axis of a cylinder. An Argon plasma is formed inside the cylinder (anode) to erode the axial target (cathode), forming a cloud of metal used to generate the nanoparticles. In this setup, the plasma occupies a much larger volume than in the regular planar magnetron geometry (25 mm diameter and 50 mm long). Here we study bimetallic AuAg nanoparticles and exploit the additional degree of freedom of chemical composition to obtain more information about the aggregation process in our magnetron cluster source. In particular, we observe significant discrepancies between the chemical composition of the sputter target and the fabricated bimetallic nanoparticles (BNPs). This observation seems contradictory to >25 years of research using laser vaporization sources (LVS)[27] in one of the participating groups, where the average chemical composition of the investigated BNPs was typically determined to be very close to that of the target rod.[24,28,29] Furthermore, also for a planar magnetron source, a chemical composition identical to that of the sputter target has been reported.[17] We thus propose to exploit the obvious dependence of cluster fabrication on the exact design of the source to obtain more information about the underlying physical processes. Our results show that for the HC-MSCS geometry, aggregation and evaporation/fragmentation processes occur in the plasma region, where collisions heat the BNPs, modifying their relative composition through evaporation.

Results and Discussions

Experimental

We have studied three different BNP samples produced by the HC-MSCS: A, B, and C using three different targets: 1/1, 1/2, and 1/3 Au/Ag twisted wires, respectively (see Methods section for details). To correctly estimate an atomic cloud composition, we used the sputter

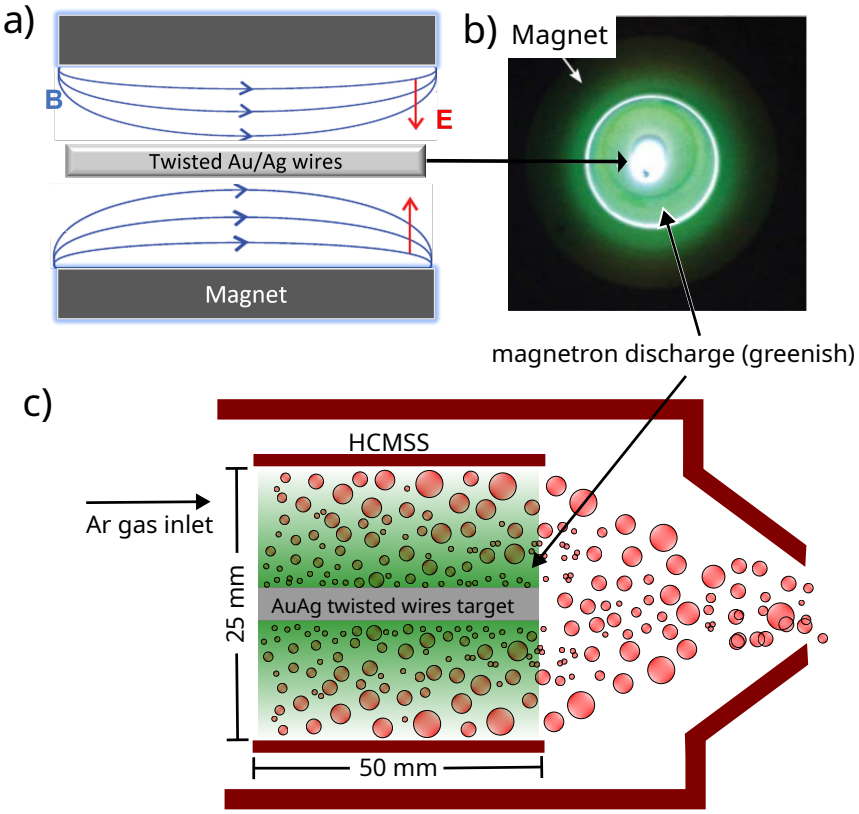


Figure 2: (a) Schematic of the hollow cylindrical magnetron sputtering (side view). The target (cathode) is in the center, and the anode corresponds to the whole cylindrical structure around it. Magnets are located in the cylindrical walls, and soft iron at the magnet extremities is used to ensure the closure of the magnetic flux lines. (b) Photo of the front view with plasma. Note the greenish color around the target, typically seen in silver plasma. (c) Schematic of the HC-MSCS mechanism of nanoparticle growth. The nanoparticles are formed in the hot plasma region (green). Large particles start growing at the beginning of the plasma region (furthest away from the exit orifice); they spend more time in the plasma than small particles formed at the end of the plasma.

yield of Au and Ag as 0.20 and 0.27 atoms/ion, respectively,[30] taking the plasma energy in the sputtering as 100 eV. Table 1 summarizes the estimated aggregation cloud Au/Ag content, henceforth called cloud composition. We note that Au and Ag abundance is slightly different from the initial target Au/Ag ratio due to the sputter yield correction. The uncertainty of the relative sputter yield dominates the estimated error in the atomic cloud composition.

Table 1: Au atomic fraction of the different BNP syntheses. Samples A, B, and C were prepared using the HC-MSCS, while sample D was produced using a laser vaporization BNP source. BNP'chemical composition quantification was obtained by Energy Dispersive X-ray Spectroscopy (EDS).

	Sample A	Sample B	Sample C	Sample D
Metal target	0.50 ± 0.01	0.33 ± 0.01	0.25 ± 0.01	0.50 ± 0.01
Atomic cloud	0.42 ± 0.03	0.27 ± 0.03	0.20 ± 0.03	0.50 ± 0.01
BNPs ensemble	0.727 ± 0.003	0.510 ± 0.005	0.500 ± 0.003	0.510 ± 0.002

To quantitatively evaluate the chemical composition of individual BNPs, we use the Scanning Transmission Electron Microscopy spatial resolution associated with the X-ray energy-dispersive spectroscopy (STEM-EDS). Quantitative chemical analyses were performed using the Cliff-Lorimer method and machine learning procedures[31–35] (See Supporting Information for more details). Initially, the chemical composition was measured for an ensemble of nanoparticles using an open parallel TEM beam with a long counting time to increase the total X-ray intensity while minimizing radiation damage (dose $\sim$ 10 $e^-/Å^2$). Table 1 presents the average experimental chemical composition of the BNPs. Unexpectedly, we observed very different values compared to the atomic cloud composition for the samples (A, B, C) prepared using the HC-MSCS. The composition of the three samples is always richer in Au. The HC-MSCS thus does not produce BNP samples with a chemical composition on average equal to the composition of the atom cloud used for their synthesis.

In order to further investigate the BNP composition distribution, we have evaluated the composition of a set of individual BNPs. Figure 3a) shows as an example the STEM-EDS elemental map of five nanoparticles and Figure 3b) presents a representative EDS spectrum for an individual nanoparticle of 6 nm. The EDS data are then analyzed using

machine learning procedures, described in detail in previous works.[34,35] Figure 3c) shows the composition of 37 BNPs from sample C. In this sample, the chemical composition distribution is size-dependent, with BNPs of a diameter larger than 5 nm being narrowly dispersed around the average value. The smaller ones are relatively poor in Au compared to the average value. It can be clearly observed that the BNP composition does not reflect the cloud composition and that the composition shows some size dependence.

To compare these BNPs composition features with similar BNPs produced in a different experimental setup, we used a laser vaporization BNP source (LVS).[27] The BNP sample from the LVS is thereafter called sample D, and its composition information is also presented in Table 1. In sample D, the chemical composition measured for the BNPs is directly related to the cloud composition. This feature can also be seen in Figure 4, where the mean value of the dispersion in Au atomic fraction is $\sim 1\%$ close to the expected nominal composition, with a dispersion of $\sim 2\%$ in the BNP composition distribution. Additionally, no chemical composition size-dependent is observed for the measured diameter range.

Simulations

In order to elucidate the role of the different physical processes occurring in the source on the Au/Ag content of the resultant nanoparticles, we resort to molecular simulation techniques. Simulation techniques have been extensively used to study BNP's energetic stability,[36,37] phase transitions[38] and core-shell segregation,[39–43] as well as nanoparticle coalescence[44–46] and growth.[47–49] It is important to emphasize that these studies usually consider the behavior of a single nanoparticle at the end of their simulations. In order to get correspondence between the composition of atomic gas and the average composition of an ensemble of BNPs, we need an approach that allows the analysis of many NP growth processes in a way that the statistical behavior can be inferred.

In this work, we simulated two distinct situations describing (i) BNP growth and aggregation from an atomic cloud with a given Au/Ag composition and; (ii) the evaporation of

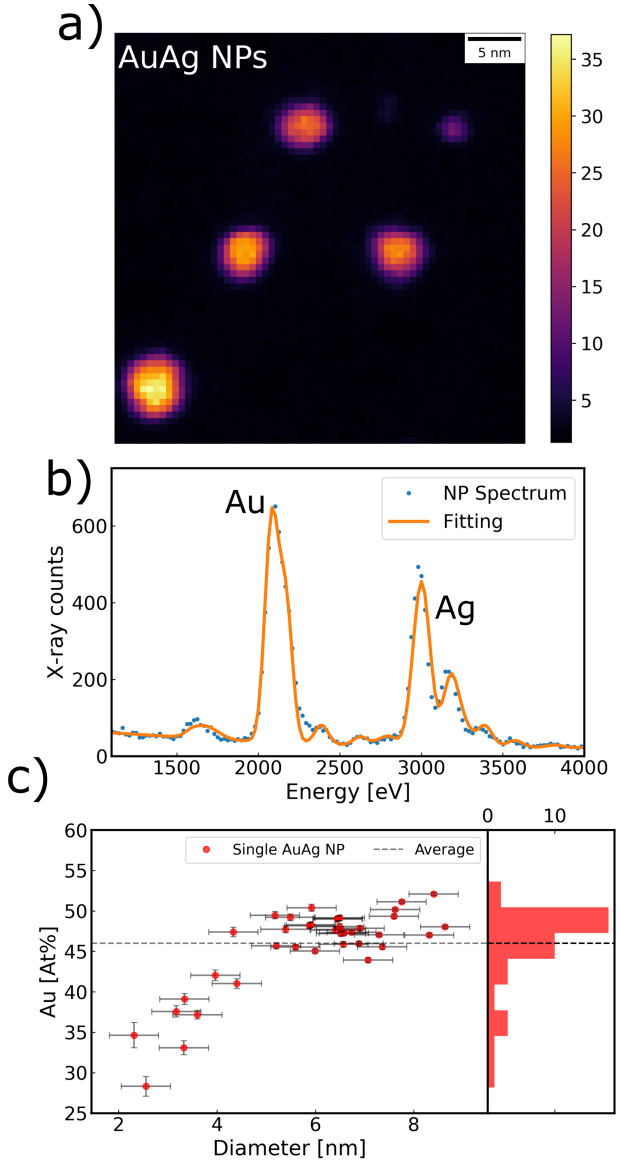


Figure 3: STEM-EDS measurement of AuAg BNPs. a) Elemental mapping obtained using the characteristic x-ray intensities from Au Mα and Ag Lα emission lines. b) Representative EDS spectrum for a single BNP and its fitting to integrate the peak intensity. c) Chemical composition for AuAg BNPs as a function of diameter from sample C. The red points are the composition quantification for individual BNPs, and the dashed line is the average value measured using open beam TEM-EDS. The $K_{AuAg} = 0.85 \pm 0.01$ Cliff-Lorimer constant was calibrated from a reference sample of known chemical composition. Note that all BNPs are richer in Au than the sputter target (25%).

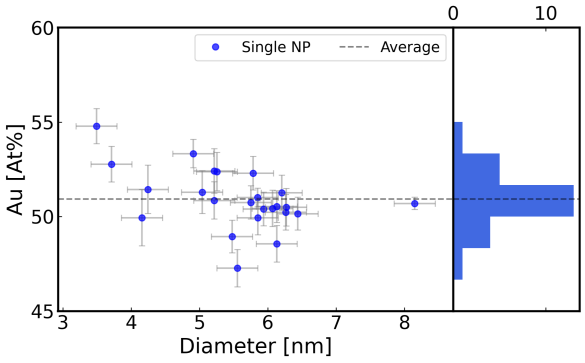


Figure 4: Chemical composition quantification of individual AuAg BNPs measured by STEM-EDS. BNPs are produced by laser vaporization source, where the BNPs are cold compared with the HC-MSCS setup. The nominal composition of the metal target (50:50) is reproduced in the measured BNPs. The $K_{AuAg} = 0.69 \pm 0.01$ Cliff-Lorimer constant was calibrated from a reference sample of known chemical composition.

a single BNP initially at a high temperature. By simulating the first case, we can confirm that BNPs produced solely by atom (or cluster) aggregation possess average compositions consistent with the Au/Ag content of the initial atomic cloud. In this way, a different process must explain the unexpected average compositions of BNPs produced by HC-MSCS. Next, we simulated the second situation, where we have a hot BNP that loses atoms due to its elevated temperature (escape/evaporation). For the latter case, we considered only a single BNP to avoid unnecessary complexities and analyzed the changes in composition resulting only from high temperatures.

The first simulation case, henceforth referred to as BNP growth, was performed by considering a system composed of N_g (N_s) atoms of gold (silver) randomly distributed inside a cubic simulation box, as shown in Figure 5 (a). The relative amount of gold and silver was such that we have Au compositions of 10%, 25%, 40%, and 50% (for more details, see the Methods section). We used Molecular Dynamics (MD) to initially simulate a hot gas of gold and silver atoms at a temperature of $T = 2000$ K. Then, the gas was quenched to $T = 300$ K, where the system loses energy very rapidly and forms a set of small atomic clusters (Figure 5 (b)). After this, we let the system thermalize at the same 300 K for a long time, and the system still consisted of small clusters made of a few atoms. The timescale of nanometer-

sized particle growth is computer-demanding by molecular dynamics simulations because there must be many low-energy collisions where atoms are incorporated in larger clusters and, subsequently, in a larger nanoparticle. In other words, the growth of BNPs from small atomic clusters are rare events that MD simulations cannot reproduce unless the simulations run for a long time. In fact, some previous works[50–52] have applied MD for complete BNPs growth, but for simulations time on the order of μs. One way to overcome these limitations and accelerate the growth process is to use Force-Bias Monte Carlo (FBMC) to aggregate small clusters into nanoparticles. This hybrid approach that uses MD and FBMC to simulate a process that has two stages on different time scales has already been employed in previous works.[53–55] Also, this same scheme was used to study core-shell segregation trends in BNPs.[56] To our knowledge, there are no works in literature that use these two techniques combined to accelerate the growth of many BNPs in a single simulation. Examples of the resulting BNPs grown by an FBMC run are shown in Figure 5 (c), where we computed each individual nanoparticle's Au/Ag composition using cluster analysis. All detailed information and simulation parameters are given in the Methods section. The composition of each individual BNP is shown as a point in Figure 5 (d), where dashed horizontal lines represent the average compositions. It can be seen that the average BNP composition is close to the initial gas composition. Nominal values and standard deviations for these quantities are reported in the caption of Figure 5. These results indicate that the discrepancy in the final nanoparticle compositions in the experiments arises from a different reason rather than just the chemical composition of the sputtered atoms from the target alone. The larger number of points for smaller Au percentages occurs because they have a larger number of atoms in the initial gas (explained in the Methods section) and, thus, can form more nanoparticles at the end of the simulation.

To describe high-temperature effects on Au/Ag BNPs, we considered a reasonably large nanoparticle of 4000 atoms (diameter of $\sim$ 5.0 nm) with three different Au compositions (25%, 50%, and 75%). Then, we carried out MD runs in an NVE (or microcanonical) en-

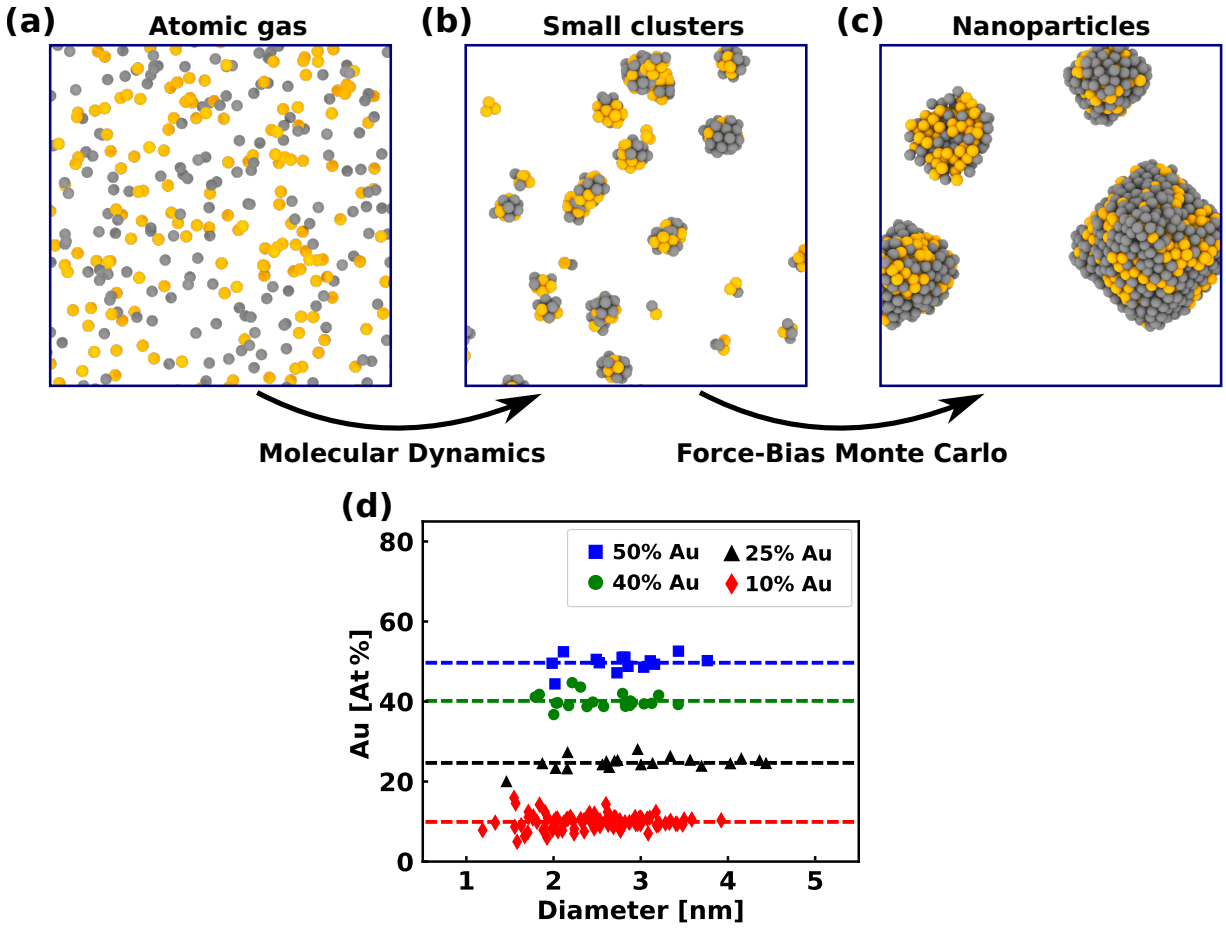


Figure 5: Hybrid MD/FBMC approach to study Au-Ag nanoparticle aggregation. (a) An initial atomic Au/Ag gas evolves to (b) a set of small clusters using MD. Then, FBMC is implemented to mimic a longer-timescale process of (c) nanometer-sized particle growth. For the sake of simplicity, in (a)-(c), we just show a piece of the whole simulation box. The chemical composition of individual nanoparticles is given by points with shapes indicated by the legend. The average composition (with their corresponding standard deviation) of each are $50 \pm 2\%$ (for 50% in Au), $40 \pm 2\%$ (for 40% in Au), $25 \pm 2\%$ (for 25% in Au) and $10 \pm 2\%$ (for 10% in Au).

semble and monitored the temperature change of the BNP as time evolved. For more details regarding the types of ensembles used in MD simulations, see the Supporting Information. We observe that atoms escape from the BNP due to its high temperature. Since the micro-canonical ensemble describes an isolated system inside the simulation box, where the energy is (on average) conserved, any temperature changes of the BNP are solely due to the evaporation of atoms. This process is known as evaporative cooling. For the evaporative cooling simulations, we need a BNP hot enough such that atoms can escape from the BNP due to the high temperature. On the other hand, we could not start from the temperature where the evaporation of the BNP would take place abruptly. With this in mind, the choice of the initial temperature should be intermediate between the melting point of silver and that of gold in such a nanoscale environment to stimulate the escape of atoms at the beginning of the simulation followed by a decreased rate of evaporation as the time evolves. In order to have a range of initial temperatures, we performed a thermal ramp from 100 K to 5000 K on the BNPs of each composition. We monitored the instantaneous atomic composition of each BNP until it completely evaporated. This gives us an insight into the range of temperatures where the BNPs start to evaporate. Thus, by observing the composition as a function of the temperature from the ramp simulations (See Supporting Information for more details), we observed that in the range between $\sim 2500 - 3500$ K we have silver atoms starting to escape the BNP, indicating that the evaporation of silver in this system lies in this temperature range. We then decided to choose 3000 K as the initial temperature. The melting temperature of nano-sized metallic particles can be difficult to choose as it can be determined by either energy or self-diffusion coefficient,[57] and, thus, we chose the evaporation point as the initial temperature. The chosen nanoparticle size was such that it is in the range of size-independent composition, as shown in Figure 3 (c). Initially, the nanoparticle starts to lose atoms due to its elevated temperature. This leads to a temperature decrease as a consequence of losing atoms with higher kinetic energies (cooling by evaporation), as can be seen in Figure 6 (a). Such decay is similar to the temperature evolution observed in water

droplets under evaporative cooling.[58] This Au-rich nanoparticle slower decay rate is expected since gold has a higher melting point than silver, which helps to stabilize nanoparticles with a larger gold content.[59,60] It is important to emphasize that the nanoscale behavior of metals is very distinct from that of bulk where, for instance, the evaporation points of nanoscale metals are considerably lower than that of their bulk counterparts.[59,61]

As the time evolves in the evaporative cooling simulations, only Ag atoms escape from the nanoparticle (independently of its chemical composition), as shown in Figure 6 (b). Similar to the reasons explained above for the greater stability of Au-rich nanoparticles, the weaker metallic bonding of silver contributes to this significant difference in evaporation rates between the two elements. This indicates that the higher Au compositions of the nanoparticles reported in Table 1 are a consequence of the evaporation of silver atoms from each nanoparticle before the sample holder deposition.

The role of core-shell segregation in the evaporation process is important to account for surface effects on the selective evaporation of Ag. Figure 6 (c) illustrates the core and shell definitions used in this work. We analyzed the core and shell composition as a function of the simulation time, shown in Figure 6 (d). We can see that Au content increases in the core and decreases in the shell. Since none of the gold atoms evaporated during the time span in our simulations, we can infer that there is a migration of gold from shell to core as it has a higher surface energy and, more importantly, a higher bulk cohesive energy, which was recently found to be of fundamental importance in determining core-shell segregation trends in bimetallic nanoparticles.[56] Both core and shell Ag moieties decrease with time, showing that even silver atoms in the core migrate to the surface, where they escape (are ejected) from the nanoparticle. This higher tendency of Ag evaporation is dominant, and it holds even for Au-rich nanoparticles

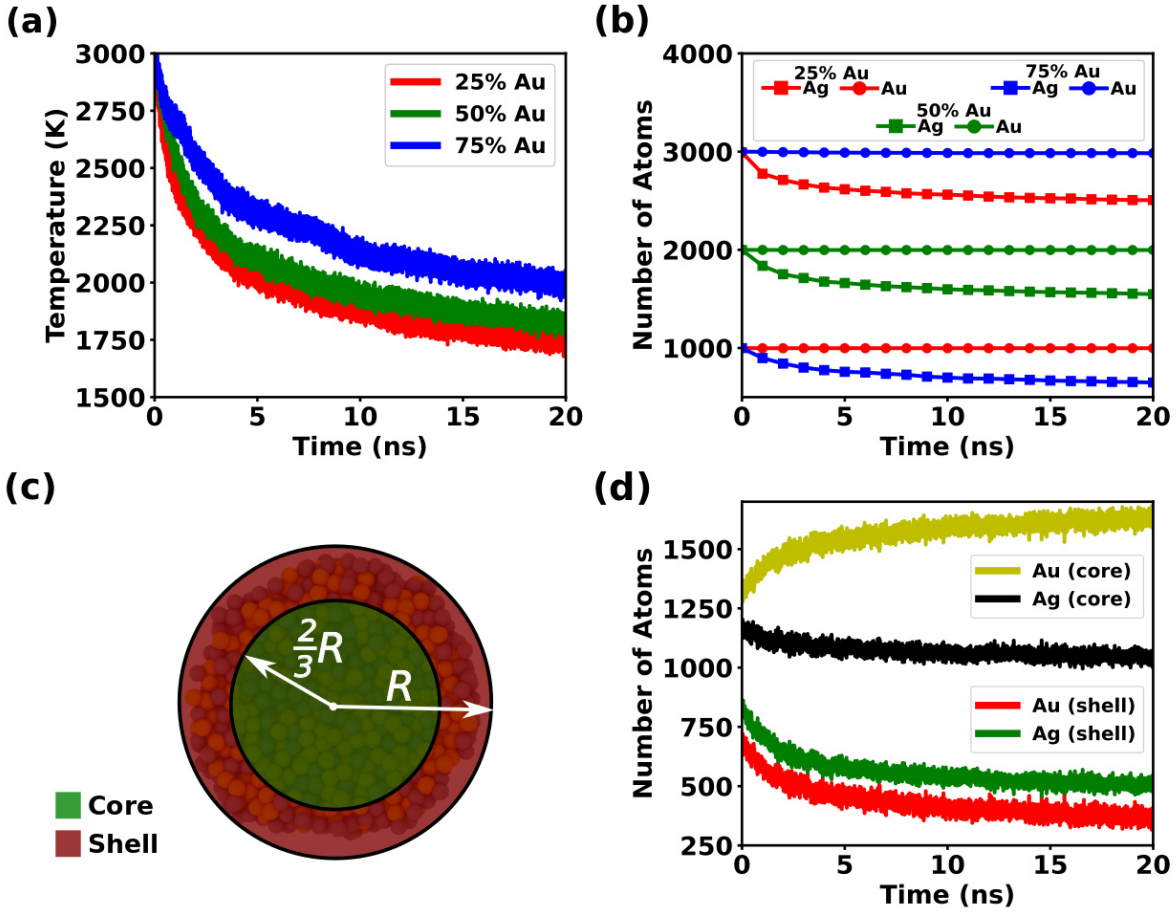


Figure 6: Evaporative cooling simulations for nanoparticles with 25%, 50%, and 75% in Au. (a) Time evolution of temperature decay as a consequence of atoms that escape from the nanoparticle initially at 3000 K. (b) The number of atoms of Au and Ag as a function of time, where squares denote silver and circles denote gold content. The same color refers to simulations of a given chemical composition. (c) Definition of core and shell, where the core is a $2R/3$-radius sphere concentric with the nanoparticle and the $R/3$-thickness spherical shell is the outside region representing the shell. (b) Core-shell resolved Au-Ag composition of a 50% gold nanoparticle, where the colors are indicated in the legend.

Discussion

Our simulations describe the two main processes in forming BNPs in cluster sources and allow us to estimate their importance in the two cases discussed here. Firstly, nucleation and coalescence, on the one hand, lead to particle growth and heating due to released binding energy. They are the most important at comparably low temperatures, where supersaturation and slow evaporation kinetics favor the sticking of atoms. Secondly, at higher temperatures, on the other hand, close to or above the size-dependent evaporation temperature of the nanoparticle material, atoms are evaporated, and the particle size decreases, i.e., while coalescence increases the particle temperature, evaporative cooling decreases it.

In the case of a hollow cylindrical magnetron source, the observed deviations in the chemical composition somehow represent a counterintuitive situation. Typically, nanoparticle growth only depends on the atomic vapor density and composition, in which the latter reflects the grown BNP average composition. Only the presence of a high-temperature environment changes the relative composition through evaporation. This evaporation process happens simultaneously with nucleation inside the HC-MSCS. Thus, we cannot conceptually separate high-temperature BNP's evaporation and simultaneous lower-temperature condensation.

Our simulation results and the observed deviations in the BNPs' chemical composition indicate that, in the HC-MSCS, all condensation and evaporation processes occur in the plasma region, where collisions with accelerated ions heat the BNPs. Even though the plasma removes atoms from the target by ion collisions and not thermally (so-called cold plasma), high temperatures in the metal vapor are achievable. Furthermore, the BNPs are heated by the binding energy of added atoms during aggregation. We find a stable kinetic state only in the region close to the target, where the particle temperature and residence time are defined by the various source parameters, such as gas flow conditions and discharge power, among others. We cannot obtain information about the exact temperature inside the plasma, and it might, at some point, be even higher than the evaporation temperature of

gold, leading to increased evaporation of both metals from possibly present large particles. However, the important point is that, eventually, the particles reach a temperature between the two evaporation points, the crucial regime in which the chemical composition can be modified. After leaving the plasma region, the nanoparticle temperature decreases through collisions with the carrier gas, and it rapidly cools below the evaporation temperature, and evaporative cooling and composition variation stop. This effect is enhanced because the melting temperature increases with increasing particle size.

The interpretation of the nanoparticle growth process proposed above is supported by the size dependence on the chemical composition. The smallest BNPs stay only a short time in the growth region, i.e., they start growing near the end of the plasma region, here, the end of the cylinder within which the plasma is confined, cf. figure 2c). They are, thus, less heated and lose fewer silver atoms, with the smallest generated clusters being close to the target composition.

These considerations lead us to redefine the concept of particle growth in magnetron sources. Contrary to the widespread acceptance, atomic addition throughout the trajectory towards the exit orifice seems negligible. The cluster size is largely defined in the plasma region close to the sputtering target, as already observed by Kousal et al.[25,26] The fact that an increase in the distance between the magnetron sputter and the exit of the aggregation chamber increases the particle size must then be linked to other parameters, notably the gas flow conditions, which affect both the residence time within the plasma and the transport through the aggregation tube.[26] The interpretation of cluster growth occurring exclusively in the plasma region is confirmed by the observation that metal redeposition principally takes place within the plasma tube and only to a very small extent onto the walls of the condensation tube.

However, we must remark that the exact geometry of the source, especially the plasma, is of fundamental importance in this discussion. While planar magnetron sources generate cluster beams without apparent electronic shell closures or odd-even oscillations,[15] they are

observed in the cylindrical setup discussed here,[21] similarly to sources based on high-energy sputtering in vacuum.[62] This observation indicates that the generated clusters are eventually hotter in the cylindrical setup than in the planar one, reflecting the differences in time the atoms spent inside the plasma. Furthermore, in laser vaporization sources, the growth mechanisms must also be in a cold environment, as reflected by the chemical composition of the BNPs. We interpret nucleation and coalescence as the determining mechanism; evaporative cooling does not seem to play a significant role. This is a reasonable assumption considering that the laser pulses are only of the order of 5 ns and the plasma is thus rapidly quenched to temperatures well below the boiling temperatures of the metals, contrary to magnetrons, where the dc plasma continuously heats the vapor and the clusters throughout their fabrication. Two more mechanisms for nanoparticle heating must also be discussed. First, in magnetron cluster sources, some plasma often opens between the cathode and the aggregation tube, forming a large "afterburner" leading to significantly increased ionization yields. In our cylindrical setup experiments, we have made sure to confine the plasma between the cathode and anode, thereby minimizing potential heating throughout the aggregation tube. Second, suppose the cluster ions are accelerated between the aggregation tube exit and the following ion optics (a skimmer). In that case, the collisions in this high-pressure region can easily lead to cluster heating and fragmentation, as evidenced by the appearance of odd-even oscillations in the cluster abundance. For the experiments discussed here, we have kept the potential difference between the tube and skimmer small enough to avoid significant acceleration-induced heating.

Finally, we would like to comment on three other possible explanations for the chemical composition deviation in our AuAg BNPs. First, we can consider that silver and gold could be sputtered at different solid angles, as has been observed before for pulsed laser desorption under high vacuum conditions.[63] However, in our case, the viscous carrier gas flow inside the hollow magnetron mixes up all the plasma components (gas, metal atoms, and nanoparticles) and sweeps them toward the plasma exit. A second possible explanation

could be a preferential sputtering of gold vs. silver, leading to gold-enriched BNPs. We can rule out this, as Ag's sputter yield is higher than Au's. Furthermore, this should lead to the enrichment of one metal over the other in the target rod. BNPs fabricated at the beginning and end of the target lifetime, however, showed no significant difference in chemical compositions. Lastly, different diffusion rates for Ag and Au in the gas flow between the plasma region and the aperture of the condensation tube could also lead to changes in the chemical composition if the lighter silver atoms migrate more quickly towards the aggregation tube walls and out of the beam. We, however, believe this effect to be minor, if at all present. In the hollow magnetron geometry, the vast majority of metal that does not condensate into clusters is redeposited onto the walls of the anode, i.e., inside the plasma region. This observation can be seen as another indication of the cluster growth happening solely within the plasma. But even in the more open planar magnetron geometry, no such increased diffusion has been seen, as evidenced by the averaged chemical composition of AgAu BNPs.[17]

On the basis of these considerations, we can now relate our results to the recent literature. In,[25,26] the authors suggest that large nanoparticles are generated in the plasma region of their magnetron source but cannot leave the source due to hydrodynamic trapping. We propose that a second mechanism should be considered, which is the evaporative cooling described above. Additional to hydrodynamical trapping, the authors may have measured a steady state large diameter population inside the plasma, which rapidly decreases in size in the outer region of the plasma, where the particles are still heated but no longer grow.

Conclusions

Although the study of metallic nanoparticles in the gas phase is an area with decades of development, the detailed knowledge of the nanoparticle growth process still presents fundamental questions. There are several routes for the formation of nanoparticles, from the way the metal cloud is generated to the strategy for its aggregation. However, all these processes

are always quite complex, and it is challenging to infer the different phenomena that occur there directly. Our chemical composition analysis of gas aggregated BNP using a hollow cylindrical magnetron sputter source cannot be understood in the framework of classical nucleation theory[64] or even considering more recent descriptions[25,26] of the planar magnetron geometry. In our case, the composition differs from that of the sputter target, which is impossible to rationalize simply in terms of the Au/Ag fraction in the initial atomic gas generated by sputtering, as shown by atomistic simulations of an MD/FBMC hybrid approach. On the contrary, evaporative cooling simulations using MD show that the selective evaporation of silver is responsible for the Au enrichment in BNPs synthesized by HC-MSCSs. Thus we can conclude that the nanoparticle growth must occur in a high-temperature region, where the plasma heats the newly formed BNPs. Further experiments are planned to verify our interpretation, e.g., using AuCu as BNP material. These two metals have similar evaporation temperatures, so we expect much less deviation between the target and BNP chemical composition.

Furthermore, these results shed new light on our knowledge of the process of forming nanoparticles in the gas phase, particularly when using dc magnetron sputtering. This study indicates that, in the HC-MSCS, NPs undergo annealing during their formation, which influences the structural characteristic of the particles, producing crystalline and compact ground-state atomic arrangements. This fact could be exploited for the in-flight fabrication of relaxed structures. Conversely, larger clusters obtained by LVS may be ramified.[65] We can furthermore continue based on these conclusions with quantitative structural analyses using a Pair Distribution Function derived from electron diffraction.[66] Also, our interpretation can be verified by studying the CoPt system, which has been demonstrated to display chemical order upon annealing.[24] In addition, we have gained a deeper understanding of how to model and control the parameters of hollow cylindrical magnetron sources to prepare high-quality model samples with chemical composition control, opening up new possibilities for studies and applications in essential areas such as catalysis, for example.

Methods

Nanoparticle Synthesis

Magnetron sputtering source:[21] Bimetallic (Au_xAg_{1-x}) NPs have been produced using a homemade gas aggregation cluster source as illustrated in Figure 1, where a cylindrical magnetron is used to sputter atoms from a central target made of twisted Au and Ag wires. A series of electrostatic lenses are used to reduce the BNP kinetic energy to achieve a "soft landing" on the TEM grid (for a ∼4 nm in diam. NP, this corresponds to ∼0.05 eV/atom). The NP size distribution contained in the molecular beam can be followed in situ by time-of-flight mass spectrometry (TOF-MS). TEM images indicate that the NP size distribution follows a log-normal function (mean diameter ∼4 nm, width ∼3 nm; see Supplementary Figure 8). Three different bimetallic nanoparticle samples (A, B, C) were used in the present study whose Au content is established by a proper choice of Au:Ag twisted wires (1:1, 1:2, and 1:3, respectively). Sample D was prepared using a laser vaporization source:[11] A doubled Nd:YAG laser (532 nm) is focused onto the surface of an $Ag_{50}Au_{50}$ target rod, and the generated plasma is quenched in a continuous He flow. Nanoparticles form in the volume of several mm^3 above the target before being swept into a vacuum through a 1 mm nozzle, creating a supersonic expansion in which the nanoparticles are extremely and efficiently cooled. The charged fraction of the nanoparticle beam is then mass-selected in a quadrupole deviator[27] before being deposited fragmentation-free onto TEM grids.

Electron Microscopy: Data Acquisition and Processing

In order to measure the average chemical composition of samples A, B, and C, Energy Dispersive X-ray Spectroscopy (EDS) was performed in an ensemble of NPs, using an open parallel TEM beam with a long counting time to increase X-ray total intensity, while minimizing radiation damage (dose ∼10 e-/Å^2). This measurement was performed in a JEM 2100F. The quantitative analysis of the EDS data followed the Cliff-Lorimer approach.[32]

Despite the quantitative precision and accuracy of recovering the average composition with this method, spatially resolved information is still required to quantify the chemical composition of each individual nanoparticle. The acquisition parameters were optimized to generate reliable statistics and do not produce beam-induced modification of the sample, measuring at 80 kV to reduce knock-on damage,[67] and composition changes.[68] The probe diameter was between 0.7 - 1.0 nm, the beam current between 300-500 pA, and the dwell time between 200-400 ms, with a pixel size of 0.5 nm. The results were processed and treated using machine learning approaches to improve quantitative EDS chemical analysis, as described in our previous works.[34,35] This measurement was performed in a JEM 2100F, and a Titan-Themis AC corrected from LNNano-Brazil. For sample D, the average chemical composition was calculated from an ensemble of 24 NPs measured individually by STEM-EDS. Here, the error is derived from standard error propagation rules. For sample D, the probe diameter was between 0.3 - 0.5 nm, beam current $\sim$ 1nA. NPs were measured with pixels size between 0.2 - 0.5 nm and dwell time between 50 - 200 ms, obtaining Au counts in the interval of 2,000 - 8,000 for 6 nm NPs. This measurement was performed in a JEOL NeoArm at Hubert Curien Laboratory at St. Etienne University, France. Data were processed using HyperSpy,[33] an open-source electron microscopy python package.

Molecular Dynamics Simulations

All MD simulations were performed with the open-source Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS).[69] Ag/Au atomic interactions for all simulations were modeled by the Embedded-Atom Method (EAM), with parameters obtained from Zhou *et al.*[70] The dimensions of the simulation box are $L_x = L_y = L_z = 50$ nm. The initial atomic gas compositions were chosen such that the Au content is kept fixed (5000 atoms). The number of Ag atoms for each case is 45000 (10% in Au), 15000 (25% in Au), 7500 (40% in Au), and 5000 (50% in Au). The dynamics of any system simulated through MD is described in a given ensemble that constrains some chosen thermodynamical variables. For instance,

a system described by a microcanonical ensemble is simulated by an NVE dynamics, where the total number of atoms N, the simulation box volume V, and the total energy E are not modified by the integration of the equations of motion. Similarly, an NVT dynamics lead the system to a desired temperature T.

The initial gas configuration was initially thermalized using an NVE ensemble for 100 ps. Then, the gas was cooled down to $T = 300$ K during 1 ns with the addition of a Langevin thermostat and a temperature damping of $10^4 dt$ that is a hundred times larger than the minimum recommended $100dt$ in order to assure high-quality results, where $dt = 0.1$ fs is the timestep used in all simulations in this work. After this, we let the system thermalize at the same 300 K during 2 ns, where a system with few-atom clusters was reached, and no further growth of nanometer-sized particles was observed.

We simulated the evaporative cooling of Au-Ag nanoparticles with 25%, 50%, and 75% of Au. Detail on how to build our model nanoparticles are given in the Supporting Information. First, we linearly increased the temperature of each nanoparticle from 100 K to 3000 K during 100 ps in an NVT dynamics. Then, we performed NVE molecular dynamics on the hot nanoparticles for 20 ns and observed the temperature decay as each system lost atoms, and their Ag-Au content was mapped as a function of the simulation time.

Force-Bias Monte Carlo

Force-Bias Monte Carlo (FBMC) was performed for 5×10^{11} steps at a target temperature of 300 K. FBMC runs are also implemented in LAMMPS. The cluster analysis was provided by the Open Visualization Tool (OVITO).[71] From the generated geometry files, we were able to obtain the number of atoms in each cluster. The nanoparticle diameters ($D = 2R$) were estimated from $R = R_W N^{1/3}$, where R_W is the Wigner radius (~ 0.16 for both Ag and Au) and N is the total number of atoms of each cluster.

Supporting Information Available

Simulations of thermal evaporation

To study the evaporation dynamics of Au and Ag in a bimetallic nanoparticle, we build a 40.9 Å×40.9 Å×40.9 Å cubic nanoalloy containing 4000 Ag atoms (See Figure 7 (a) at 100 K). Then, we considered three cases where Au ones replaced 25%, 50%, and 75% of the atoms. A temperature ramp using a Nosé-Hoover thermostat was applied, where the temperature of the system is increased from 100 K to 5000 K, linearly on average, as shown in Figure 7 (b). As the nanoalloy is heated, it assumes a spherical shape (snapshots shown in Figure 7 (a)). We observed the individual evaporation of Ag and Au atoms for all three relative percentages of each element, as can be seen in Figure 7 (c). From temperatures slightly higher than 3000 K, the nanoparticles of all compositions begin to evaporate. We chose this value to be the initial temperature in the evaporative cooling simulations described in the main manuscript. We also carried out the same procedure for a 12.27 Å×12.27 Å×12.27 Å cubic nanoalloy containing 100 atoms in total to determine how the evaporation of Au and Ag change for smaller sizes, as shown in Figure 7 (d). In addition, for the case where there is 50% of Au, we also simulated the thermal evaporation process for nanoparticles with 500 and 1000 atoms to evaluate the size dependence of Ag and Au dissociation from the original particle (Figure 7 (e)). The time evolution of the chemical composition of nanoparticles was analyzed using the Visual Molecular Dynamics (VMD) software through TCL scripting.

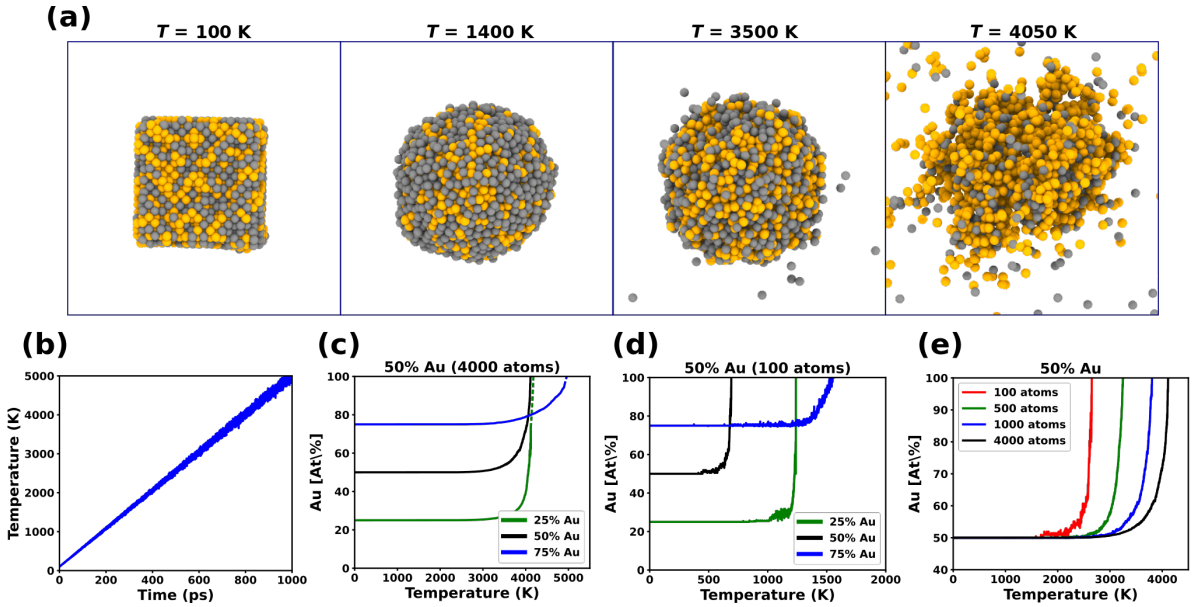


Figure 7: Thermal evaporation of an Au-Ag nanoparticle. (a) Structural evolution for successive indicated temperatures of a cubic nanoalloy containing a total of 4000 atoms. The relative amount of gold and silver is such that we have the following three cases: 25%, 50%, and 75% in Au. (b) Nosé-Hoover thermostat set the time evolution of the temperature of the system with an approximately linear profile. Au's composition as a function of temperature for nanoparticles with (c) 4000 atoms and (d) 100 atoms. (e) Size dependence of the gold composition for nanoparticles initially with 50% Au.

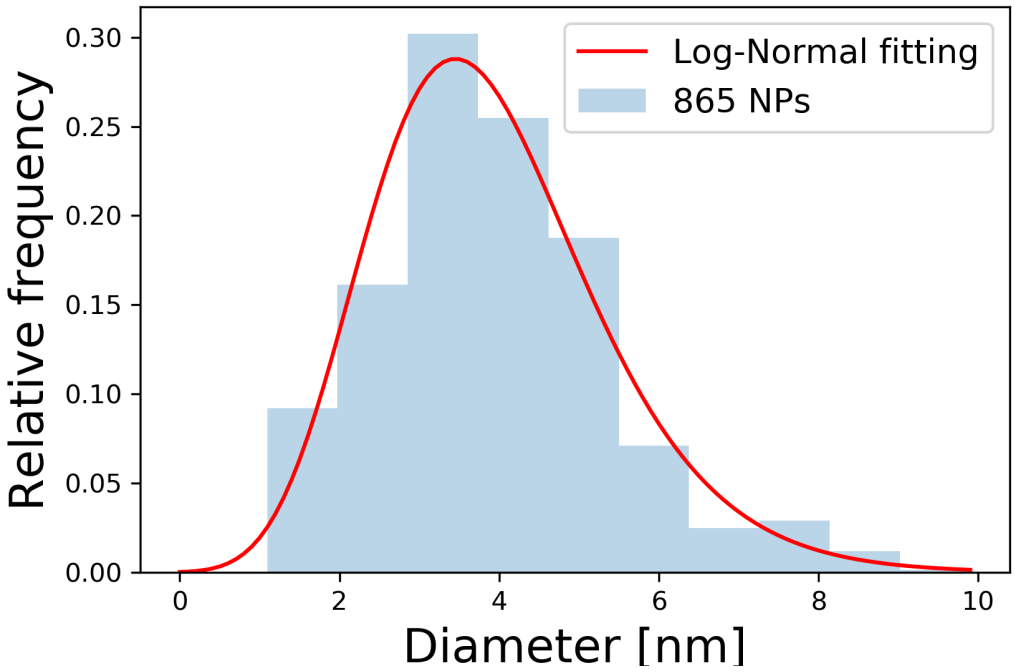


Figure 8: Histogram of AuAg NPs diameter distribution measured by TEM. The log-normal curve is adjusted on the histogram with a peak 4 nm and FWHM of 3 nm.

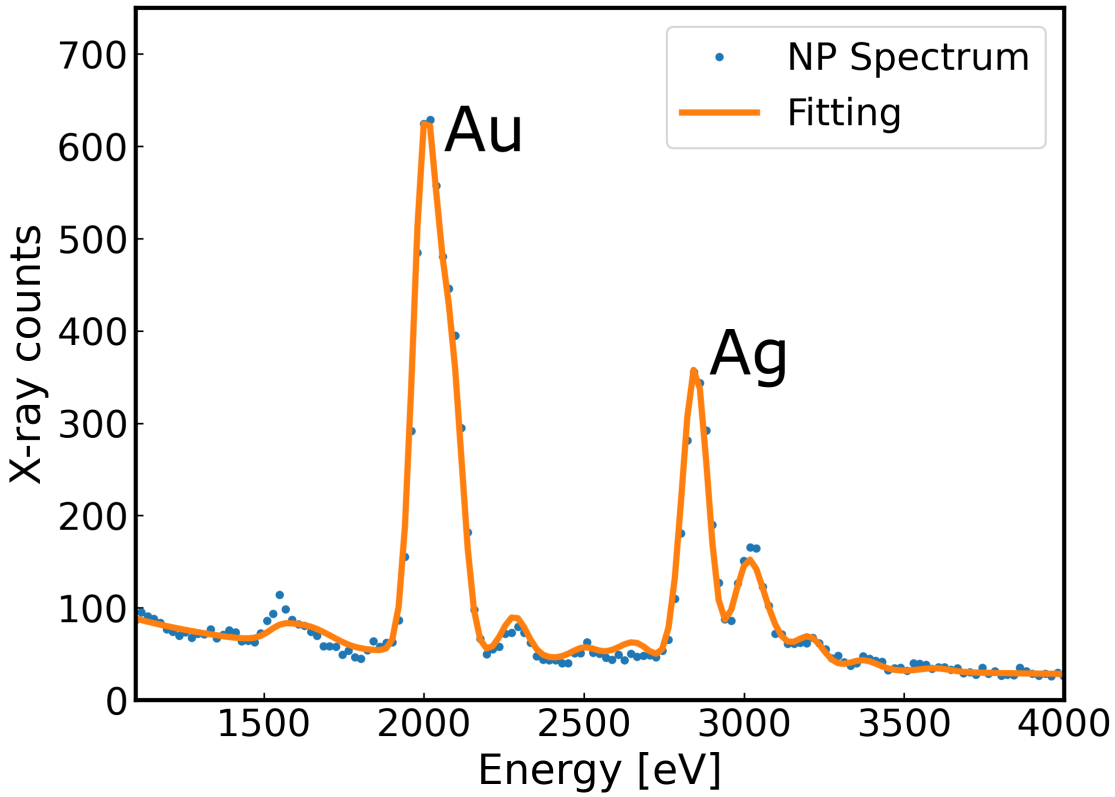


Figure 9: Representative EDS spectrum for a single BNP produced by the LVS and its fitting to integrate the peak intensity.

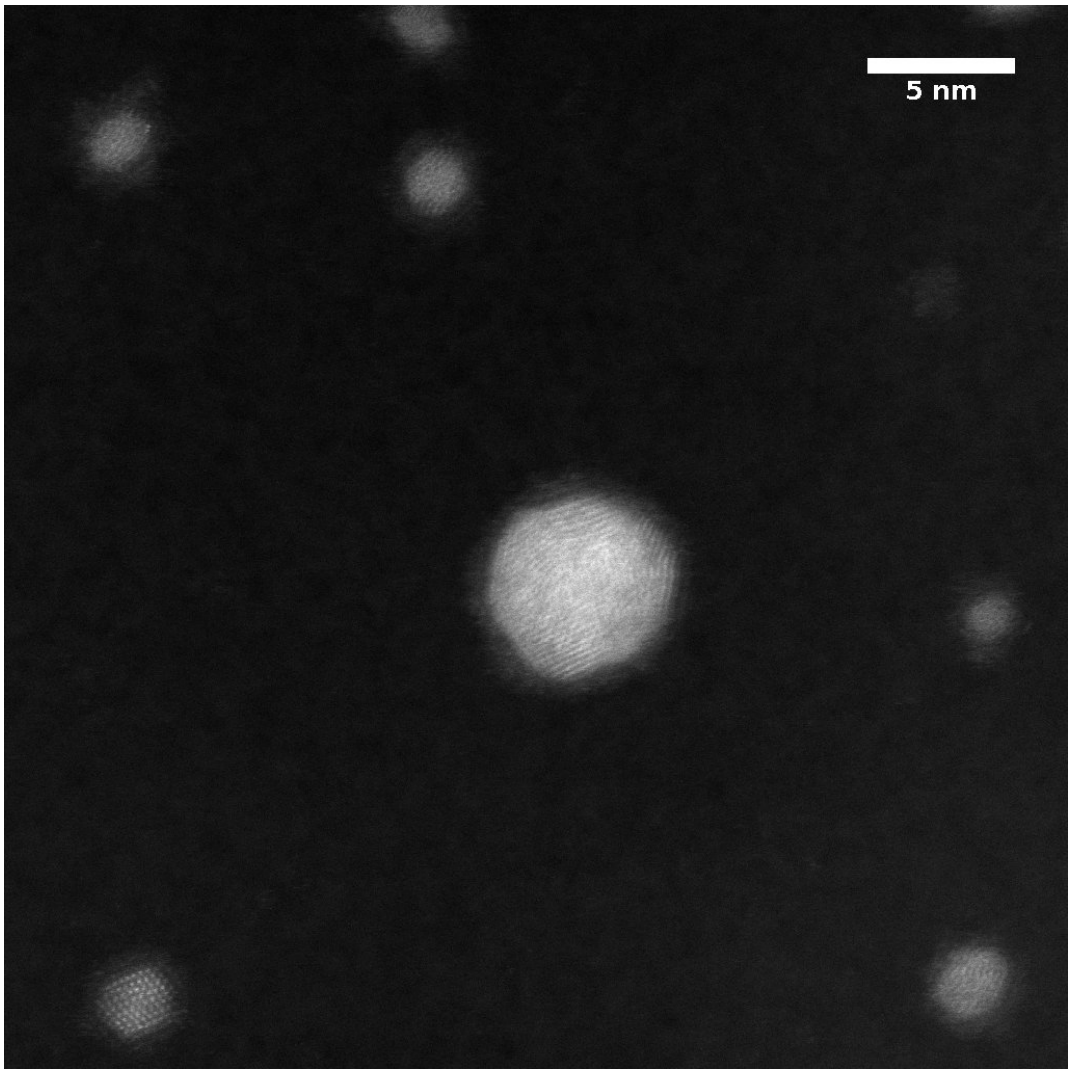


Figure 10: HAADF STEM image of AuAg BNPs produced by HCMSS.

References

(1) Campos, A.; Troc, N.; Cottancin, E.; Pellarin, M.; Weissker, H.-C.; Lermé, J.; Kociak, M.; Hillenkamp, M. Plasmonic Quantum Size Effects in Silver Nanoparticles are Dominated by Interfaces and Local Environments. *Nat. Phys.* **2019**, *15*, 275–280.

(2) Heiz, U., Landman, U. E., Eds. *Nanocatalysis*; Springer, 2007.

(3) Aslam, U.; Rao, V. G.; Chavez, S.; Linic, S. Catalytic conversion of solar to chemical energy on plasmonic metal nanostructures. *Nature Catalysis* **2018**, *1*, 656.

(4) Mirigliano, M.; Milani, P. Electrical conduction in nanogranular cluster-assembled metallic films. *Adv. Phys. X* **2021**, *6*, 1908847.

(5) Oyarzún, S.; Tamion, A.; Tournus, F.; Dupuis, V.; Hillenkamp, M. Size effects in the magnetic anisotropy of embedded cobalt nanoparticles: from shape to surface. *Sci. Rep.* **2015**, *5*, 14749.

(6) Serrano-Guisan, S.; Di Domenicantonio, G.; Abid, M.; Abid, J.-P.; Hillenkamp, M.; Gravier, L.; Ansermet, J.-P.; Félix, C. Enhanced magnetic field sensitivity of spin-dependent transport in cluster-assembled metallic nanostructures. *Nat. Mater.* **2006**, *5*, 730.

(7) Hai, P. N.; Ohya, S.; Tanaka, M. Long spin-relaxation time in a single metal nanoparticle. *Nat. Nanotechnol.* **2010**, *5*, 593–596.

(8) Hensel, R. C.; Braunger, M. L.; Oliveira, B.; Shimizu, F. M.; Oliveira, O. N.; Hillenkamp, M.; Riul, A.; Rodrigues, V. Controlled Incorporation of Silver Nanoparticles into Layer-by-Layer Polymer Films for Reusable Electronic Tongues. *ACS Appl. Nano Mater.* **2021**, *4*, 14231.

(9) Huttel, Y., Ed. *Gas-Phase Synthesis of Nanoparticles*; WILEY-VCH Verlag, 2017.

(10) Haberland, H. *Gas-Phase Synthesis of Nanoparticles*; John Wiley & Sons, Ltd, 2017; Chapter 1, pp 1–21.

(11) Perez, A. et al. Functional nanostructures from clusters. *Int. J. Nanotechnol.* **2010**, *7*, 523–574.

(12) De Toro, J. A.; Normile, P. S.; Binns, C. *Gas-Phase Synthesis of Nanoparticles*; John Wiley & Sons, Ltd, 2017; Chapter 3, pp 39–55.

(13) Haberland, H.; Mall, M.; Moseler, M.; Qiang, Y.; Reiners, T.; Thurner, Y. Filling of micron-sized contact holes with copper by energetic cluster impact. *J. Vac. Sci. Technol. A* **1994**, *12*, 2925–2930.

(14) Palmer, R. E.; Pratontep, S.; Boyen, H. G. Nanostructured surfaces from size-selected clusters. *Nat. Mater.* **2003**, *2*, 443–448.

(15) Hillenkamp, M.; Di Domenicantonio, G.; Félix, C. Monodispersed metal clusters in solid matrices: A new experimental setup. *Rev. Sci. Instrum.* **2006**, *77*, 025104.

(16) Smirnov, B. M.; Shyjumon, I.; Hippler, R. Flow of nanosize cluster-containing plasma in a magnetron discharge. *Phys. Rev. E* **2007**, *75*, 066402.

(17) Belić, D.; Chantry, R. L.; Li, Z. Y.; Brown, S. A. Ag-Au nanoclusters: Structure and phase segregation. *Appl. Phys. Lett.* **2011**, *99*, 171914.

(18) Hartmann, H.; Popok, V. N.; Barke, I.; von Oeynhausen, V.; Meiwes-Broer, K.-H. Design and capabilities of an experimental setup based on magnetron sputtering for formation and deposition of size-selected metal clusters on ultra-clean surfaces. *Rev. Sci. Instrum.* **2012**, *83*, 073304.

(19) Zhang, C.; Tsunoyama, H.; Akatsuka, H.; Sekiya, H.; Nagase, T.; Nakajima, A. Advanced Nanocluster Ion Source Based on High-Power Impulse Magnetron Sputtering

and Time-Resolved Measurements of Nanocluster Formation. *J. Phys. Chem. A* **2013**, *117*, 10211–10217.

(20) Pilch, I.; Söderström, D.; Brenning, N.; Helmersson, U. Size-controlled growth of nanoparticles in a highly ionized pulsed plasma. *Appl. Phys. Lett.* **2013**, *102*, 033108.

(21) de Sá, A. D. T.; Oiko, V. T. A.; di Domenicantonio, G.; Rodrigues, V. New experimental setup for metallic clusters production based on hollow cylindrical magnetron sputtering. *J. Vac. Sci. Technol. B* **2014**, *32*, 061804.

(22) Khojasteh, M.; Kresin, V. V. Influence of source parameters on the growth of metal nanoparticles by sputter-gas-aggregation. *Appl. Nanosci.* **2017**, *7*, 875–883.

(23) Sanzone, G.; Yin, J.; Sun, H. Scaling up of cluster beam deposition technology for catalysis application. *Front. Chem. Sci. Eng.* **2021**, *15*, 1360–1379.

(24) Tournus, F.; Blanc, N.; Tamion, A.; Hillenkamp, M.; Dupuis, V. Dispersion of magnetic anisotropy in size-selected CoPt clusters. *Phys. Rev. B* **2010**, *81*, 220405.

(25) Kousal, J.; Shelemin, A.; Schwartzkopf, M.; Polonskyi, O.; Hanus, J.; Solar, P.; Vaidulych, M.; Nikitin, D.; Pleskunov, P.; Krtous, Z.; Strunskus, T.; Faupel, F.; Roth, S.; Biederman, H.; Choukourov, A. Magnetron-sputtered copper nanoparticles: lost in gas aggregation and found by in situ X-ray scattering. *Nanoscale* **2018**, *10*, 18275.

(26) Drewes, J.; Ali-Ogly, S.; Strunskus, T.; Polonskyi, O.; Biederman, H.; Faupel, F.; Vahl, A. Impact of argon flow and pressure on the trapping behavior of nanoparticles inside a gas aggregation source. *Plasma Process. Polym.* **2022**, *19*, 2100125.

(27) Alayan, R.; Arnaud, L.; Bourgey, A.; Broyer, M.; Cottancin, E.; Huntzinger, J. R.; Lermé, J.; Vialle, J. L.; Pellarin, M.; Guiraud, G. Application of a static quadrupole deviator to the deposition of size-selected cluster ions from a laser vaporization source. *Rev. Sci. Instrum.* **2004**, *75*, 2461.

(28) Rousset, J. L.; Cadrot, A. M.; Cadete Santos Aires, F. J.; Renouprez, A.; Mélinon, P.; Perez, A.; Pellarin, M.; Vialle, J. L.; Broyer, M. Study of bimetallic Pd–Pt clusters in both free and supported phases. *J. Chem. Phys.* **1995**, *102*, 8574.

(29) Dupuis, V.; Robert, A.; Hillion, A.; Khadra, G.; Blanc, N.; Roy, D. L.; Tournus, F.; Albin, C.; Boisron, O.; Tamion, A. Cubic chemically ordered FeRh and FeCo nanomagnets prepared by mass-selected low-energy cluster-beam deposition: a comparative study. *Beilstein J. Nanotechnol.* **2016**, *7*, 1850.

(30) Laegreid, N.; Wehner, G. K. Sputtering Yields of Metals for Ar+ and Ne+ Ions with Energies from 50 to 600 ev. *Journal of applied physics* **1961**, *32*, 365–369.

(31) Carter, C. B., Williams, D. B., Eds. *Transmission Electron Microscopy*; Springer International Publishing: Cham, 2016.

(32) Cliff, G.; Lorimer, G. W. The quantitative analysis of thin specimens. *Journal of Microscopy* **1975**, *103*, 203–207.

(33) De La Peña, F. et al. Electron Microscopy (Big and Small) Data Analysis With the Open Source Software Package HyperSpy. *Microscopy and Microanalysis* **2017**, *23*, 214–215.

(34) Moreira, M.; Hillenkamp, M.; Divitini, G.; Tizei, L. H. G.; Ducati, C.; Cotta, M. A.; Rodrigues, V.; Ugarte, D. Improving quantitative EDS chemical analysis of alloy nanoparticles by PCA denoising: part I, reducing reconstruction bias. *Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada* **2022**, 1–12.

(35) Moreira, M.; Hillenkamp, M.; Divitini, G.; Tizei, L. H. G.; Ducati, C.; Cotta, M. A.; Rodrigues, V.; Ugarte, D. Improving quantitative EDS chemical analysis of alloy nanoparticles by PCA denoising: part II. uncertainty intervals. *Microscopy and microanalysis*

: the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada **2022**, 1–9.

(36) Chen, F.; Johnston, R. L. Energetic, electronic, and thermal effects on structural properties of Ag-Au nanoalloys. *ACS Nano* **2008**, *2*, 165–175.

(37) Liao, T.-W.; Yadav, A.; Hu, K.-J.; van der Tol, J.; Cosentino, S.; D'Acapito, F.; Palmer, R. E.; Lenardi, C.; Ferrando, R.; Grandjean, D.; Lievens, P. Unravelling the nucleation mechanism of bimetallic nanoparticles with composition-tunable core-shell arrangement. *Nanoscale* **2018**, *10*, 6684–6694.

(38) Qi, W. H.; Lee, S. T. Phase stability, melting, and alloy formation of au-ag bimetallic nanoparticles. *The Journal of Physical Chemistry C* **2010**, *114*, 9580–9587.

(39) Curley, B. C.; Rossi, G.; Ferrando, R.; Johnston, R. L. Theoretical study of structure and segregation in 38-atom Ag-Au nanoalloys. *The European Physical Journal D* **2007**, *43*, 53–56.

(40) Calvo, F.; Cottancin, E.; Broyer, M. Segregation, core alloying, and shape transitions in bimetallic nanoclusters: Monte Carlo simulations. *Physical Review B* **2008**, *77*, 121406.

(41) Cerbelaud, M.; Ferrando, R.; Barcaro, G.; Fortunelli, A. Optimization of chemical ordering in AgAu nanoalloys. *Physical Chemistry Chemical Physics* **2011**, *13*, 10232–10240.

(42) Deng, L.; Hu, W.; Deng, H.; Xiao, S.; Tang, J. Au–Ag Bimetallic Nanoparticles: Surface Segregation and Atomic-Scale Structure. *The Journal of Physical Chemistry C* **2011**, *115*, 11355–11363.

(43) Li, L.; Li, X.; Duan, Z.; Meyer, R. J.; Carr, R.; Raman, S.; Koziol, L.; Henkelman, G. Adaptive kinetic Monte Carlo simulations of surface segregation in PdAu nanoparticles. *Nanoscale* **2019**, *11*, 10524–10535.

(44) Faccin, G. M.; San-Miguel, M. A.; Andres, J.; Longo, E.; da Silva, E. Z. Computational modeling for the ag nanoparticle coalescence process: A case of surface plasmon resonance. *The Journal of Physical Chemistry C* **2017**, *121*, 7030–7036.

(45) Abbaspour, M.; Akbarzadeh, H.; Salemi, S.; Lotfi, S. Investigation of possible formation of au@m (M = cu, ir, pt, and rh) core–shell nanoclusters in a condensation–coalescence process using molecular dynamics simulations. *Industrial & engineering chemistry research* **2018**, *57*, 14837–14845.

(46) Grammatikopoulos, P.; Sowwan, M.; Kioseoglou, J. Computational modeling of nanoparticle coalescence. *Advanced Theory and Simulations* **2019**, 1900013.

(47) Harjunmaa, A.; Nordlund, K. Molecular dynamics simulations of Si/Ge cluster condensation. *Computational Materials Science* **2009**, *47*, 456–459.

(48) Turner, C. H.; Lei, Y.; Bao, Y. Modeling the atomistic growth behavior of gold nanoparticles in solution. *Nanoscale* **2016**, *8*, 9354–9365.

(49) Punjani, K.; Bhimalapuram, P. Study of shape changes during nanoparticle growth using Kinetic Monte Carlo simulation: a case study on gold nanoparticles. *Journal of Chemical Sciences* **2021**, *133*, 101.

(50) Akbarzadeh, H.; Shamkhali, A. N.; Mehrjouei, E. Ag-Au bimetallic nanoclusters formed from a homogeneous gas phase: a new thermodynamic expression confirmed by molecular dynamics simulation. *Physical Chemistry Chemical Physics* **2017**, *19*, 3763–3769.

(51) Vernieres, J.; Steinhauer, S.; Zhao, J.; Chapelle, A.; Menini, P.; Dufour, N.; Diaz, R. E.; Nordlund, K.; Djurabekova, F.; Grammatikopoulos, P.; Sowwan, M. Gas Phase Synthesis of Multifunctional Fe-Based Nanocubes. *Advanced functional materials* **2017**, *27*, 1605328.

(52) Grammatikopoulos, P. Atomistic modeling of the nucleation and growth of pure and hybrid nanoparticles by cluster beam deposition. *Current opinion in chemical engineering* **2019**, *23*, 164–173.

(53) Rao, M.; Berne, B. J. On the force bias Monte Carlo simulation of simple liquids. *The Journal of Chemical Physics* **1979**, *71*, 129–132.

(54) Mees, M. J.; Pourtois, G.; Neyts, E. C.; Thijsse, B. J.; Stesmans, A. Uniform-acceptance force-bias Monte Carlo method with time scale to study solid-state diffusion. *Physical Review B* **2012**, *85*, 134301.

(55) Neyts, E. C.; Bogaerts, A. Combining molecular dynamics with Monte Carlo simulations: implementations and applications. *Theoretical chemistry accounts* **2013**, *132*, 1320.

(56) Eom, N.; Messing, M. E.; Johansson, J.; Deppert, K. General Trends in Core-Shell Preferences for Bimetallic Nanoparticles. *ACS Nano* **2021**, *15*, 8883–8895.

(57) Ercolessi, F.; Andreoni, W.; Tosatti, E. Melting of small gold particles: Mechanism and size effects. *Physical Review Letters* **1991**, *66*, 911–914.

(58) Schlesinger, D.; Sellberg, J. A.; Nilsson, A.; Pettersson, L. G. M. Evaporative cooling of microscopic water droplets in vacuo: Molecular dynamics simulations and kinetic gas theory. *The Journal of Chemical Physics* **2016**, *144*, 124502.

(59) Guisbiers, G.; Mendoza-Cruz, R.; Bazán-Díaz, L.; Velázquez-Salazar, J. J.; Mendoza-Perez, R.; Robledo-Torres, J. A.; Rodriguez-Lopez, J.-L.; Montejano-Carrizales, J. M.; Whetten, R. L.; José-Yacamán, M. Electrum, the Gold-Silver Alloy, from the Bulk Scale to the Nanoscale: Synthesis, Properties, and Segregation Rules. *ACS Nano* **2016**, *10*, 188–198.

(60) Okamoto, H.; Massalski, T. B. The Ag-Au (Silver-Gold) system. *Bulletin of Alloy Phase Diagrams* **1983**, *4*, 30–38.

(61) Lasserus, M.; Schnedlitz, M.; Knez, D.; Messner, R.; Schiffmann, A.; Lackner, F.; Hauser, A. W.; Hofer, F.; Ernst, W. E. Thermally induced alloying processes in a bimetallic system at the nanoscale: AgAu sub-5 nm core–shell particles studied at atomic resolution. *Nanoscale* **2018**, *10*, 2017–2024.

(62) Conus, F.; Lau, J. T.; Rodrigues, V.; Félix, C. High sensitivity absorption measurement of small metal clusters embedded in an argon matrix. *Rev. Sci. Instrum.* **2006**, *77*, 113103.

(63) Urbassek, H.; Sibold, D. Gas-phase segregation effects in pulsed laser desorption from binary targets. *Phys. Rev. Lett.* **1993**, *70*, 1886.

(64) Mélinon, P. *Gas-Phase Synthesis of Nanoparticles*; John Wiley & Sons, Ltd, 2017; Chapter 2, pp 23–38.

(65) Alayan, R.; Arnaud, L.; Broyer, M.; Cottancin, E.; Lermé, J.; Vialle, J. L.; Pellarin, M. Morphology and growth of metal clusters in the gas phase: A transition from spherical to ramified structures. *Phys. Rev. B* **2006**, *73*, 125444.

(66) Corrêa, L. M.; Moreira, M.; Rodrigues, V.; Ugarte, D. Quantitative Structural Analysis of AuAg Nanoparticles Using a Pair Distribution Function Based on Precession Electron Diffraction: Implications for Catalysis. *ACS Applied Nano Materials* **2021**, *4*, 12541–12551.

(67) Egerton, R. F. Choice of operating voltage for a transmission electron microscope. *Ultramicroscopy* **2014**, *145*, 85–93.

(68) Braidy, N.; Jakubek, Z. J.; Simard, B.; Botton, G. A. Quantitative energy dispersive X-ray microanalysis of electron beam-sensitive alloyed nanoparticles. *Microscopy*

and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada **2008**, *14*, 166–175.

(69) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **1995**, *117*, 1–19.

(70) Zhou, X. W.; Johnson, R. A.; Wadley, H. N. G. Misfit-energy-increasing dislocations in vapor-deposited CoFe/NiFe multilayers. *Physical Review B* **2004**, *69*, 144113.

(71) Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering* **2010**, *18*, 015012.